Quality assessment of multiply and singly distorted stereoscopic images via adaptive construction of cyclopean views

Yi Zhang [a,*], Damon M. Chandler [b], Xuanqin Mou [a]

[a] *School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, 710049, China*
[b] *College of Information Science and Engineering, Ritsumeikan University, Shiga, 525-8577, Japan*

ARTICLE INFO

ABSTRACT

A challenging problem confronted when designing a blind/no-reference (NR) stereoscopic image quality assessment (SIQA) algorithm is to simulate the quality assessment (QA) behavior of the human visual system (HVS) during binocular vision. An effective way to solve this problem is to estimate the quality of the merged single view created in the human brain which is also referred to as the cyclopean image. However, due to the difficulty in modeling the binocular fusion and rivalry properties of the HVS, obtaining effective cyclopean images for QA is non-trivial, and consequently previous NR SIQA algorithms either require the MOS/DMOS values of the distorted 3D images for training or ignore the quality analysis of the merged cyclopean view. In this paper, we focus on (1) constructing accurate and appropriate cyclopean views for QA of stereoscopic images by adaptively analyzing the distortion information of two monocular views, and (2) training NR SIQA models without requiring the assistance of the MOS/DMOS values in existing databases. Accordingly, we present an effective opinion-unaware SIQA algorithm called MUSIQUE-3D, which blindly assesses the quality of multiply and singly distorted stereoscopic images by analyzing quality degradations of both monocular and cyclopean views. The monocular view quality is estimated by an extended version of the MUSIQUE algorithm, and the cyclopean view quality is computed from the distortion parameter values predicted by a two-layer classification-regression model trained on a large 3D image dataset. Tests on various 3D image databases demonstrate the superiority of our method as compared with other state-of-the-art SIQA algorithms.

1. Introduction

Recently, the rapid development of virtual reality technology has provided users with the exciting visual experience of stereoscopic/3D content, and consequently various 3D services and applications such as 3D television, 3D video conferencing, 3D cinema, and 3D games, etc. have gained popularity. Normally, these 3D visual contents often go through multiple stages of processing (e.g., image acquisition, compression, transmission, reception, display, etc.) before ultimately being presented to the consumers, and for each stage, various types of distortions can be introduced, negatively impacting the quality of the user's 3D visual experience. Thus, there is a need for effective and reliable stereoscopic image quality assessment (SIQA) algorithms that can estimate the perceptual quality of the finally observed 3D scene.

Although noticeable progress has been made on developing various kinds of 2D image quality assessment (IQA) algorithms to assess the perceptual quality of 2D images (see [1,2] for reviews), designing effective SIQA algorithms to automatically assess the perceptual quality of a 3D scene is extremely challenging because many different factors including heterogeneous distortions, mismatched inter-view perceptions, excessive binocular disparity, and inappropriate depth-of-focus

can all lead to visual discomfort. Although visual discomfort is certainly an important factor that contributes to the overall 3D quality of experience, another very important factor is image distortion, which has been the focus of most existing works, and which is also the focus of this paper.

Even when considering only a single factor, predicting the quality of a 3D scene is still difficult, because the human visual system (HVS) perceives a 3D scene through two stereoscopic views and judges its quality based mainly on the merged single view created in the brain after complex binocular fusion and rivalry processes. Although such QA behaviors seem to be natural to humans, designing a 3D QA algorithm to mimic this process is non-trivial given that only two view images are available. The difficulty arises first from the inevitable occlusion and border area in the two view images, which is attributed to the two slightly different perspectives of the two cameras in capturing the image [3]. The difficulty also arises from the fact that binocular combinations under different distortion types should be considered. As claimed in [4], the higher quality view that contains sufficient information will help suppress the lower quality view

with information-loss distortion (e.g., blurring), while for information-additive distortion (e.g., blockiness), the lower quality view cannot be compensated [5]. Moreover, compared to the single-distortion scenario, the multiple-distortion scenario adds another level of difficulty for 3D QA: the algorithm must not only consider the joint effects of different distortions on the two views, but also consider the influence of these distortions on the binocular combination behaviors, because both monocular and binocular visions provide important information for the HVS to judge the 3D image quality.

Despite these difficulties, there have been a number of SIQA algorithms developed over the recent past, and consequently various kinds of features, models, and frameworks have been proposed. For no-reference (NR) SIQA, the most common approach is to learn regression models which map the quality-related features of the test image to associated quality scores. Features that have been used include natural-scene-statistics (NSS) features based on monocular/cyclopean images [6–8]; saliency-based binocular features [9]; univariate and bivariate statistical features [10]; binocular quality-aware features based on eye-weighting and contrast-gain-control models [11]; and those extracted by using deep neural networks (e.g., [4,12,13]) and autoencoders (e.g., [14]). These are often called "opinion-aware" approaches, because the regression models are trained on distorted images along with their human subjective rating scores. Consequently, their applicability is restricted by the limited number of existing 3D image quality databases. In comparison, much fewer "opinion-unaware" approaches have been presented. The main idea of this type of approach is to estimate quality differences between distorted images and pristine images by using the various quality-related features and measurements, such as the BRISQUE [15] features and the Mahalanobis distance measurement of multivariate Gaussian models adopted in [16], the amplitude/phase difference features and the visual-codebook-based quality-lookup method adopted in [17,18], the label consistent K-singular value decomposition classification framework adopted in [19], etc.

Although the aforementioned NR SIQA algorithms are effective, most of them were originally designed to work only for singly-distorted stereoscopic images (SDSIs), while in practice a stereoscopic image can be simultaneously contaminated by multiple distortions during the multiple stages of processing. With the recently developed multiply-distorted stereoscopic image quality database (i.e., the NBU-MDSID database [20,21]), the NR SIQA topic on multiply-distorted stereoscopic images (MDSIs) has begun to receive increased attention. However, due to the aforementioned difficulties introduced by multiple distortions, only a few related works have been reported. For example, Shao et al. [20] proposed a multi-model joint sparse representation framework based on learning modality specific dictionaries and projection matrices from singly-distorted images. Later, Shao et al. [21] proposed another multi-model sparse representation framework which uses a local phase and amplitude description for dictionary learning and employs a multi-stage pooling strategy for quality estimation. Jiang et al. [3] proposed a unified NR quality evaluator for SDSIs and MDSIs based on learning monocular and binocular local visual primitives in order to characterize the local receptive field properties of the visual cortex.

Indeed, these three algorithms pioneered the progress in the field of NR multiply-distorted SIQA; however, they all suffer from certain limitations. For example, both Shao's two sparse-representation-based methods follow the traditional SIQA pipeline that the quality of the two monocular views are evaluated separately, followed by a linear combination that collapses the two quality estimates into one final quality score. Without analyzing the quality degradation of the merged view perceived by the HVS, both algorithms are unable to interpret how the binocular fusion and rivalry behaviors operate when symmetrically and asymmetrically distorted stereopairs are viewed, and thus do not fully mimic the intrinsic mechanism of the HVS in judging the visual quality of stereoscopic scenes. Although Jiang's method takes into account binocular vision by incorporating a similar cyclopean framework as [22], the method operates in an "opinion-aware" manner, and thus is also restricted by the limited quantity of training data. Finally, all three algorithms consider only three distortion types (white noise, Gaussian blur, and JPEG compression) and their combinations, while another common distortion type, JPEG2000 compression, is not included.

By summarizing all existing SIQA approaches, we argue that a grand challenge for effective blind QA of MDSIs and SDSIs lies in solving three fundamental problems: (1) how to assess the quality of two monocular views corrupted by multiple distortions without requiring training the IQA models on human subjective rating scores; (2) how to construct accurate and appropriate cyclopean views to simulate the merged single view created in the human brain by using only stereopairs; and (3) how to combine quality estimates corresponding to monocular and binocular visions into one scalar to represent the overall 3D image quality. One promising approach to build IQA models without training on human subjective rating scores is our recently developed MUSIQUE algorithm [23], which decouples the QA task into two subtasks: (1) estimation of distortion parameters from input distorted images, and (2) estimation of quality from the estimated parameters. Motivated by this approach, as well as to overcome the potential limitations encountered by the current NR SIQA works, we present in this paper an effective opinion-unaware SIQA algorithm called MUSIQUE-3D to blindly assess the quality of both MDSIs and SDSIs.

Specifically, MUSIQUE-3D operates via three main stages as shown in Fig. 1. In the first stage, the quality of the two monocular views are estimated separately by an extension of the MUSIQUE algorithm to take into account four common distortion types (white noise, Gaussian blur, JPEG compression, and JPEG2000 compression) and their combinations. To this end, we present a more advanced classification framework trained on a large dataset of 2D images to distinguish among nine different distortion cases. We also present a more advanced quality-fusion strategy which adaptively addresses the joint effects of the four distortion types by considering the masking effect caused particularly by noise.

In the second stage, the quality of the cyclopean view is estimated based on modeling the most crucial properties of the HVS in 3D viewing. Specifically, intermediate maps corresponding to the luminance and the pixel-based contrast are generated based on an optical flow algorithm which is employed to compute a disparity map, and based on a quality-compensated multipathway contrast gain-control model (QC-MCM) which is employed to model the binocular fusion and rivalry behaviors of the HVS in viewing symmetrically and asymmetrically distorted 3D images. The cyclopean view quality is then estimated by a cyclopean IQA framework, which contains classification and regression models trained on a large dataset of 3D images to predict the distortion parameter values of the two intermediate maps (cyclopean luminance and contrast images).

In the final stage, the two quality estimates obtained from the two monocular views are combined, which is then incorporated by the cyclopean view quality obtained in the second stage to yield the overall quality estimate of the stereoscopic image. To this end, we propose a new combination strategy that adaptively merges the quality estimates of the left and right views based not only on the contrast of each view, but also on analyzing whether or not the lower quality view can be compensated by the other when the two view images share similar perceived contrast. As we will demonstrate, all stages together allow MUSIQUE-3D to achieve better/competitive quality predictive performance as compared with many other FR/NR IQA algorithms on various 3D image quality databases.
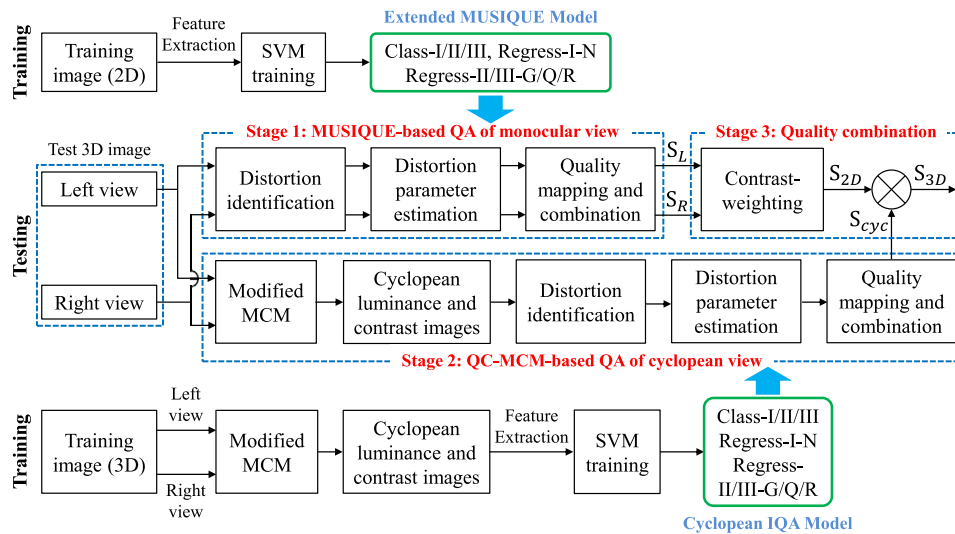
Fig. 1. A block diagram of the MUSIQUE-3D algorithm. Note that Stage 1 operates independently for the left and right views. Also note that the SVM approach can possibly be replaced by other machine learning models such as the back propagation (BP) neural network (please refer to Section 3.5 for more discussions).

The rest of the paper is organized as follows. Section 2 describes details of the proposed MUSIQUE-3D algorithm. Section 3 analyzes the performance of MUSIQUE-3D on various multiply and singly distorted stereoscopic image databases. General conclusions are presented in Section 4.

2. Algorithm

The proposed MUSIQUE-3D algorithm is based on the assumption that the overall perceptual quality of a 3D scene can be evaluated by combining the two monocular view qualities and the merged binocular view quality. Thus, MUSIQUE-3D operates via three main stages as mentioned in Section 1: (1) MUSIQUE-based QA of the two monocular views; (2) QC-MCM-based QA of the cyclopean view; and (3) combination of the monocular and cyclopean views' qualities to yield the final quality score of the stereoscopic image. We provide the algorithm details in the following subsections.

2.1. MUSIQUE on stereopairs

As reported in [23], the MUSIQUE algorithm operates based on a two-layer classification-regression framework, which is employed to predict distortion parameter values of three distortion types (white noise, Gaussian blur, and JPEG compression) by analyzing 218 distortion-sensitive features extracted from the image. To achieve the goal of distortion parameter estimation, the framework consists of three classification models and five regression models, all of which were trained on a large 2D image dataset generated by adding the three distortion types to the 125 pristine natural images in the Berkeley segmentation database [24]. By employing parametric functions to map the distortion parameter values to associated quality scores, and also by employing the most-apparent-distortion strategy to combine the different quality scores to a single value to represent the overall image quality, the MUSIQUE algorithm was shown to achieve reliable quality predictive performance on various multiply and singly distorted 2D image databases.

Recognizing that considering only three distortion types is insufficient for an NR SIQA algorithm, in the MUSIQUE-based QA stage of this work, we developed an extended version of MUSIQUE [23] for QA of the stereopairs (i.e., the left and right view images) also containing JPEG2000 compression artifacts. Specifically, we present a more advanced two-layer classification model (shown in Fig. 2) such

that four distortion types and their combinations are considered: white noise, Gaussian blur, JPEG compression, and JPEG2000 compression. Among the four distortion parameters corresponding to the four distortion types, three of them are the same as in MUSIQUE: (1) the standard deviation σ_G for generating the Gaussian blur (Gblur); (2) the compression quality factor Q for the JPEG compression (JPEG); and (3) the variance $\bar{\sigma}_N$ for generating the white noise (WN). The fourth distortion parameter is the compression ratio R for measuring the size of the JPEG2000 (JP2K) output bit stream. Note that in this work, we compute the JPEG2000 distortion parameter as $\bar{R} = \ln(1 + R)$ to balance the different ranges of R. As in original MUSIQUE, the same three steps are employed: (1) distortion identification; (2) distortion parameter estimation; and (3) quality mapping and combination. We provide details for each step in the following subsections.

2.1.1. Distortion identification

To predict the four distortion parameter values, a two-layer classification model is first applied to perform distortion identification. As shown in Fig. 2, the "Class-I" model in the first layer distinguishes three classes: (1) WN only; (2) WN + Gblur/JPEG/JP2K; and (3) Gblur/JPEG/JP2K only. Then, the two parallel sub-classification models in the second layer (i.e., "Class-II" and "Class-III") further distinguish the four sub-classes under the two different noise conditions (i.e., noise present or not present), and consequently nine distortion types are distinguished: (1) Gblur, (2) JPEG, (3) JP2K, (4) WN, (5) Gblur + JPEG, (6) Gblur + WN, (7) JPEG + WN, (8) JP2K + WN, and (9) Gblur + JPEG + WN.

It is important to note that the four distortion types will theoretically produce 11 ($C_4^2 + C_4^3 + C_4^4$) different distortion combinations. However, we do not consider images that simultaneously contain both JPEG and JPEG2000 distortions in this work, because images are often compressed by using only one standard in practice, which implies that the compression artifacts of JPEG and JPEG2000 rarely coexist in an image. Also note that in our model, we do not have training images with such distortion combinations as "Gblur + JP2K" or "WN + Gblur + JP2K". This is due to the fact that when Gaussian blur and JPEG2000 compression coexist, the two distortion types will interact with each other, making images look more blurred, and consequently it is difficult to determine the equivalent values of the Gaussian blur and JPEG2000 compression parameters of the training images generated by simultaneously adding the two distortions. However, this does not mean that the extended MUSIQUE algorithm performs less effectively
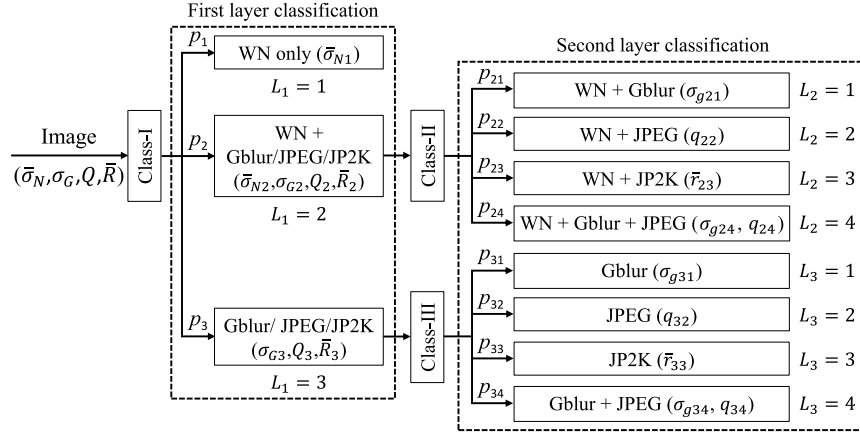
First layer classification

Fig. 2. Topological diagram of the two-layer classification model. Note that p_k, p_{2l}, and p_{3l} ($k = 1, 2, 3; l = 1, 2, 3, 4$) represent the class probabilities given by the three classification models, respectively; L_1, L_2, and L_3 are the corresponding classification labels; and the numbers 1, 2, 3, 4 represent the corresponding class. Also note that the symbols in parentheses represent the distortion parameters estimated for each corresponding class.

on these kinds of images. In fact, since the blur artifacts introduced by Gaussian blur and JPEG2000 compression are similar, the joint effect between the two distortion types will not affect the performances of corresponding regression models in predicting σ_G and $\bar{R}$ values, even if they were not trained on "Gblur + JP2K" and "WN + Gblur + JP2K" images.

To train the three classification models, a large dataset of distorted images was generated by adding the four distortion types to the 125 pristine natural images in the Berkeley segmentation database [24], as well as the left views of the 20 pristine stereoscopic images in the high-resolution stereo datasets [25], and then quality-sensitive features were extracted from these distorted images as the training data. Here, we summarize all of the features adopted by MUSIQUE-3D in Table 1, and we refer interested readers to [23] for more details about the feature extraction. A summary of the features required by each classification model in the first distortion identification stage is shown in Table 2. By using support vector machine (SVM) learning for training, each classification model has two types of output, the predicted label and the class probability, and both will be used by the corresponding regression models to predict the distortion parameter values.

2.1.2. Distortion parameter estimation

The next task is to estimate distortion parameters which will be used to assess image quality. Specifically, we adopt the same probability weighting strategy in [23], but employ different equations to compute the Gaussian blur, JPEG, and JPEG2000 compression parameters because more distortion types are classified in the second layer. To this end, we trained corresponding regression models via SVM on the large dataset of the 2D distorted images whose four distortion parameter values were saved when creating the data. A block diagram of this stage is shown in Fig. 3.

The noise estimation is almost the same as in [23]. Let "Regress-I-N" denote the two regression models trained on the "WN" and "WN + Gblur/JPEG/JP2K" images, respectively. Let $\bar{\sigma}_{N1}$ and $\bar{\sigma}_{N2}$ denote the two outputs of Regress-I-N. The noise distortion parameter (denoted by $\bar{\sigma}_N$) is computed by

$$\bar{\sigma}_N = \begin{cases} \left(p_1 \cdot \bar{\sigma}_{N1} + p_2 \cdot \bar{\sigma}_{N2}\right) / \left(p_1 + p_2\right), & L_1 = 1, 2 \\ p_1 \cdot \bar{\sigma}_{N1} + p_2 \cdot \bar{\sigma}_{N2}, & L_1 = 3 \end{cases} \quad (1)$$

where p_1 and p_2 are the classification probabilities given by the Class-I model; L_1 is the predicted classification label; and the numbers 1, 2, 3 represent the three corresponding classes (see Fig. 2).

To estimate the distortion parameters of the Gaussian blur, JPEG compression, and JPEG2000 compression, regression models trained on images contaminated by the three distortion types are required. Specifically, we train "Regress-II-G" and "Regress-III-G" models to predict the

Gaussian blur parameters σ_{gi1} and σ_{gi4}; we also train "Regress-II-Q" and "Regress-III-Q" models to predict the JPEG parameters q_{i2} and q_{i4}; and finally we train "Regress-II-R" and "Regress-III-R" models to predict the JPEG2000 parameters $\bar{r}_{i3}$. Note that distortion parameters estimated for each corresponding class are shown in Fig. 2. Then, the intermediate Gaussian blur, JPEG, and JPEG2000 parameters (denoted by σ_{Gi}, Q_i, and $\bar{R}_i$, respectively) are computed by

$$\sigma_{Gi} = \begin{cases} \dfrac{p_{i1} \cdot \sigma_{gi1} + p_{i4} \cdot \sigma_{gi4}}{p_{i1} + p_{i4}}, & L_i = 1, 4 \\ \dfrac{p_{i1} \cdot \sigma_{gi1} + p_{i4} \cdot \sigma_{gi4}}{p_{i1} + p_{i2} + p_{i4}}, & L_i = 2, 3 \end{cases} \quad (2)$$

$$Q_i = \begin{cases} \dfrac{p_{i2} \cdot q_{i2} + p_{i4} \cdot q_{i4}}{p_{i2} + p_{i4}}, & L_i = 2, 4 \\ (p_{i1} + p_{i3}) \cdot 80 + p_{i2} \cdot q_{i2} \\ \quad + p_{i4} \cdot q_{i4}, & L_i = 1, 3 \end{cases} \quad (3)$$

$$\bar{R}_i = \begin{cases} \bar{r}_{i3}, & L_i = 3 \\ \dfrac{p_{i2} \cdot \ln 2 + p_{i3} \cdot \bar{r}_{i3}}{p_{i2} + p_{i3}}, & L_i = 1, 2, 4 \end{cases} \quad (4)$$

where i takes the values of 2 or 3; p_{il} ($l = 1, 2, 3, 4$) denote the classification probabilities given by the Class-II/III models; L_i denote their predicted classification labels; and the numbers 1, 2, 3, 4 represent the four corresponding classes (see Fig. 2).

Note that by Eqs. (1)–(4), we assume that (1) the noise distortion parameter for a noise-free image is zero, (2) the blur parameter for a non-blurred image is zero, (3) the JPEG parameter for a non-JPEG-compressed image is 80, and (4) the JPEG2000 parameter for a non-JPEG2000-compressed image is one. As claimed in [23], the value of 80 was selected based upon the range of the JPEG parameters that have been used to generate the training data. The value of one indicates that a non-JPEG2000-compressed image has an equal size with its original. Although the same probability-weighting rule in the original MUSIQUE algorithm is adopted, a key difference occurs when computing the Gaussian blur and JPEG2000 compression parameters. For example in Eq. (2), the "WN + JP2K" and "JP2K" terms are not included in the weighted summation when L_i equal to 2 and 3 (i.e., the denominator does not contain p_{i3}). The reason is that the Gaussian blur parameters for non-blurred images (e.g., "WN + JP2K" and "JP2K") are not zero, owing to the fact that JPEG2000 compression can also introduce blur. The same principle is followed when computing the JPEG2000 parameter in Eq. (4), because the JPEG2000 parameters for non-JPEG2000-compressed images (e.g., "Gblur", "Gblur + JPEG",

Table 1
A summary of the bag of features extracted by MUSIQUE-3D algorithm.

Feature ID	Feature description
f_1–f_{16}	PS/LDS-based spatial-domain features (σ)
f_{17}–f_{32}	PS/LDS-based spatial-domain features (α)
f_{33}–f_{52}	Gradient-weighted LBP features
f_{53}–f_{68}	LDS-based log-Gabor-domain features
f_{69}–f_{92}	Relative phase features
f_{93}–f_{122}	Across-scale correlation features
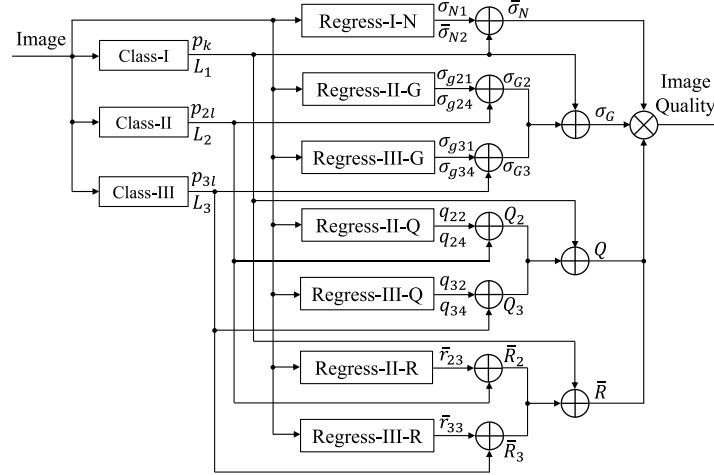
Fig. 3. Block diagram of estimating the four distortion parameters by using the two-layer classification model. The outputs of the three classification models p_k, p_{2l}, p_{3l}, and L_k ($k = 1, 2, 3; l = 1, 2, 3, 4$) represent the same meaning as in Fig. 2.

"WN + Gblur", and "WN + Gblur + JPEG") are not one. As we have found in our research, giving up these unknown terms in the weighted summation can significantly increase the accuracy in predicting the corresponding distortion parameter values.

The final estimates of the Gaussian blur, JPEG compression, and JPEG2000 compression parameters (denoted by σ_G, Q, and $\bar{R}$, respectively) are given by

$$\sigma_G = \begin{cases} p_2 \cdot \sigma_{G2} + p_3 \cdot \sigma_{G3}, & L_1 = 1 \\ \dfrac{p_2 \cdot (1-\omega) \cdot \sigma_{G2} + p_3 \cdot \omega \cdot \sigma_{G3}}{p_2 \cdot (1-\omega) + p_3 \cdot \omega}, & L_1 = 2 \\ \dfrac{p_2 \cdot \sigma_{G2} + p_3 \cdot \sigma_{G3}}{p_2 + p_3}, & L_1 = 3 \end{cases} \quad (5)$$

$$Q = \begin{cases} p_1 \cdot 80 + p_2 \cdot Q_2 + p_3 \cdot Q_3, & L_1 = 1 \\ \dfrac{p_2 \cdot (1-\omega) \cdot Q_2 + p_3 \cdot \omega \cdot Q_3}{p_2 \cdot (1-\omega) + p_3 \cdot \omega}, & L_1 = 2 \\ \dfrac{p_2 \cdot Q_2 + p_3 \cdot Q_3}{p_2 + p_3}, & L_1 = 3 \end{cases} \quad (6)$$

$$\bar{R} = \begin{cases} p_1 \cdot \ln 2 + p_2 \cdot \bar{R}_2 + p_3 \cdot \bar{R}_3, & L_1 = 1 \\ \dfrac{p_2 \cdot (1-\omega) \cdot \bar{R}_2 + p_3 \cdot \omega \cdot \bar{R}_3}{p_2 \cdot (1-\omega) + p_3 \cdot \omega}, & L_1 = 2 \\ \dfrac{p_2 \cdot \bar{R}_2 + p_3 \cdot \bar{R}_3}{p_2 + p_3}, & L_1 = 3 \end{cases} \quad (7)$$

where p_k ($k = 1, 2, 3$) denote the classification probabilities given by the Class-I model; L_1 and the three numbers 1,2,3 represent the same meanings as they are in Eq. (1); ω is a parameter determined by $\bar{\sigma}_N$ through a sigmoid function:

$$\omega = A/[1 + e^{t_1(\bar{\sigma}_N - t_2)}] + B, \quad (8)$$

where $t_1 = 6$ and $t_2 = 1.25$. Because ω is in the range of $[0, 1]$, we set $A = 1$ and $B = 0$. The motivation of employing a control variable ω in Eqs. (5)–(7), as well as the selections of t_1 and t_2 values in Eq. (8) are the same as that in original MUSIQUE. We refer interested readers to [23] for more detailed analysis and discussion.

To train the regression models (see Fig. 3), relevant distortion-sensitive features are required. A summary of the features employed by each regression model is shown in Table 2. As in [23], the NSS features employed for the Gaussian blur and JPEG2000 parameters estimation are extracted from the sharper regions of an image, which are defined as those image blocks corresponding to the 50% largest sharpness values within the sharpness map computed by the FISH$_{bb}$ algorithm [26]. By summarizing both the distortion identification and distortion parameter estimation stages, we observe that (1) most features required by the classification and regression models in MUSIQUE-3D are similar to those used in MUSIQUE; and (2) only part of the original MUSIQUE features are adopted in this work. For example, we do not use all of the features extracted from the denoised image and the Laplacian pyramid bands. As we have found, such a feature optimization can significantly increase the algorithm speed while still maintaining an equivalent QA performance.

2.1.3. Quality mapping and combination

The final step of the MUSIQUE-based QA stage is to map the four distortion parameter values to quality scores, which are then combined to yield the final quality estimate. To this end, we employ the same polynomial fitting method in [23] to map the four distortion parameters ($\bar{\sigma}_N$, σ_G, Q, and $\bar{R}$) to VIF [27] quality scores. Fig. 4 shows scatter plots of distortion parameter values (which correspond to the four subsets of the training images) versus VIF scores for four types of distortion: Gaussian blur, white noise, JPEG compression, and JPEG2000 compression.

Table 2
A summary of the required features for all classification and regression models in the MUSIQUE-based QA stage of MUSIQUE-3D. Note that "✓" indicates that the features are required by the corresponding models.

	f_1–f_{16}	f_{17}–f_{32}	f_{33}–f_{52}	f_{53}–f_{68}	f_{69}–f_{92}	f_{93}–f_{122}
Class-I	✓	✓	✓	✓		✓
Class-II	✓	✓			✓	✓
Class-III	✓	✓	✓		✓	
Regress-I-N	✓	✓		✓		✓
Regress-II-G	✓				✓	✓
Regress-III-G	✓	✓	✓		✓	
Regress-II-Q	✓	✓		✓	✓	✓
Regress-III-Q	✓	✓	✓	✓	✓	
Regress-II-R	✓				✓	✓
Regress-III-R	✓	✓	✓		✓	

As in [23], the shape of each scatter plot can be modeled by a four-parameter third-order polynomial curve (i.e., the red curves in Fig. 4) which is given by

$$y = \lambda_1 \cdot x^3 + \lambda_2 \cdot x^2 + \lambda_3 \cdot x + \lambda_4, \tag{9}$$

where λ_i ($i = 1, 2, 3, 4$) are the curve parameters. The fitted values of the four curve parameters for each distortion type are shown in Table 3.

Note that, although we added Gaussian blur to the pristine 2D/3D images at 12 different levels (see Section 3.1) to generate the training data, the curve parameter values computed based on only the eight different levels in [23] are also qualified for the quality-mapping task in this work. In fact, the more number of Gaussian blur levels will only lengthen the curve tail above the horizontal axis, and thus the modeling of the curve can be easily achieved by setting all mapped negative VIF scores (if encountered by using the four-parameter polynomial curve for quality mapping) to be zero instead of using a higher-order polynomial curve.

Let VIF_G, VIF_N, VIF_Q, and VIF_R denote the original mapped qualities [computed by using Eq. (9) with parameter values in Table 3] for Gaussian blur, white noise, JPEG compression, and JPEG2000 compression, respectively. The corrected values for the quality degradation (denoted by D_G, D_N, D_Q, and D_R, respectively) corresponding to the four distortion types are given by

$$\text{D}_G = 1 - \text{VIF}_G, \tag{10}$$

$$\text{D}_N = 1 - \left(\text{VIF}_N + \beta_1 \right), \tag{11}$$

$$\text{D}_Q = 1 - \left(\text{VIF}_Q + \beta_2 \right), \tag{12}$$

$$\text{D}_R = 1 - \text{VIF}_R, \tag{13}$$

where β_1 and β_2 denote the two offsets employed to make the mapped quality more reasonable across different distortion types; the subtraction from one aims to make the mapped quality as differential mean opinion scores.

With these four estimated quality degradation scores, the next step is to combine these scores into one scalar that represents the overall image quality. Let D_{GR} denote the maximum value of D_G and D_R. Then, the final image quality score, denoted by S, is computed by

$$\text{S} = \begin{cases} \text{D}_N, & L_1 = 1 \\ \text{D}_1^\gamma \times \text{D}_2^{1-\gamma}, & L_1 = 2 \\ \max\left(\text{D}_{GR}, \text{D}_Q \right) \times \rho^{\min\left(\text{D}_{GR}, \text{D}_Q \right)}, & L_1 = 3 \end{cases} \tag{14}$$

where L_1 and the numbers 1, 2, 3 denote the same meanings as in Eq. (1); $\rho = 1.15$ is an exponential factor which aims to model the impact of the second-most-apparent distortion; D_1 and D_2 are the two quality estimates computed respectively for two different noise levels. Specifically, D_1 is computed when images are assumed to be corrupted by only a slight amount of noise, in which case the previous most-apparent-distortion strategy in [23] is adopted. In contrast, D_2 is computed when images are assumed to be severely contaminated by the noise distortion, in which case the other distortion

artifacts (e.g., the blur/block/ringing artifacts introduced by Gaussian blur/JPEG/JPEG2000 compression) are masked due to the much increased local contrast of the image. Accordingly, the two quality estimates are given by

$$\text{D}_1 = d_1 \times \rho^{d_2}, \tag{15}$$

$$\text{D}_2 = \max\left(\text{D}_{GR} + \beta_3, \text{D}_N \right) \times \rho^{\min\left(\text{D}_{GR} + \beta_3, \text{D}_N \right)}, \tag{16}$$

where d_1 and d_2 denote the top two largest values of D_{GR}, D_Q, and D_N; β_3 denotes the masking effect caused by the noise distortion. To measure the overall quality degradation of an image simultaneously corrupted by noise and other distortion types, D_1 and D_2 are adaptively combined based on a control parameter γ, which is determined by $\bar{\sigma}_N$ through another sigmoid function that has the same form as Eq. (8). In this work, we set the following parameter values: $\beta_1 = \beta_2 = \beta_3 = -0.1$, $A = 1$, $B = 0$, $t_1 = 3$, and $t_2 = 0.5$ to help achieve the best performance across different 3D image quality databases.

2.2. QC-MCM-based cyclopean IQA

As mentioned in Section 1, the HVS judges the quality of a 3D scene based mainly on the merged single view created in the brain after binocular fusion and rivalry processes. Thus, in the second stage of MUSIQUE-3D, we first build a so-called *cyclopean view*[1] from the two monocular views, and then evaluate the cyclopean view quality by using the same approach in Section 2.1. In the following subsections, we provide algorithm details for each step.

2.2.1. Compute disparity map

To compute the cyclopean view, a disparity map which indicates the apparent pixel difference or motion between the two views of a stereoscopic image is required. In an FR setting, the disparity map is often computed from the reference stereopairs. However, in an NR setting, obtaining an accurate disparity map can be difficult because of the various types of distortion. Although a denoising algorithm can possibly be applied to reduce the error of the estimated disparity map, as shown in Fig. 5, we have found that the contribution of this process is rather minor in regards to improving the overall QA performance (see Section 3.5 for a demonstration). Thus, in this paper, we directly apply the optical flow algorithm presented in [28] (namely "SIFT flow") to stereopairs to obtain the disparity map.

2.2.2. QC-MCM-based cyclopean image

Given the estimated disparity map, we next build the disparity-compensated cyclopean view images to mimic the merged single view seen by humans based on a QC-MCM. Specifically, we build two cyclopean images, one based on luminance, and another one based on contrast, which is given by

$$\text{I}(x, y) = L(x, y) / [\bar{L}_B(x, y) + K] \tag{17}$$

---

[1] The term "cyclopean view" is usually defined as the merged single image of a scene created in the brain by combining the two images received from the two eyes.

Table 3
Values of polynomial fitting parameters for the four distortion types.

	λ_1	λ_2	λ_3	λ_4
Gblur	-4.6738×10^{-3}	7.4033×10^{-2}	-4.0485×10^{-1}	8.4124×10^{-1}
WN	-2.7938×10^{-3}	4.9839×10^{-2}	-3.4269×10^{-1}	9.2673×10^{-1}
JPEG	3.0303×10^{-6}	-4.7193×10^{-4}	2.9372×10^{-2}	3.0030×10^{-2}
JP2K	-3.9012×10^{-3}	1.0140×10^{-1}	-8.8495×10^{-1}	2.6059

where L denotes the luminance value; $\bar{L}_B(x, y)$ denotes the average luminance value of a 9×9 block centered around pixel (x, y); $K = 0.001$ is a small constant that prevents division by zero.

In our previous FR SIQA work [29], we argued that the quality of a monocular view with higher contrast will play a more dominant role in determining the HVS's judgment of the 3D image quality. In this paper, we further argue that given two monocular views with similar contrast but apparently different quality degradations, the view with greater quality degradation should weigh more on determining the overall 3D image quality. This statement is also in accord with findings in [5] that the low-quality view caused by information-additive distortion (e.g., blockiness) cannot be compensated by the high-quality view. To emphasize the dominant role of such information-additive distortion in stereoscopic viewing, we present a quality-compensated MCM as follows:

(1) If the two view images have significantly different quality ratings but similar averaged contrast values, then the view with a lower quality rating will be compensated by a larger weight in the cyclopean-image-build process;

(2) Given a stereopair satisfying the condition of (1), if the lower-quality view contains no/less noise and its estimated JPEG compression parameter Q is lower than a threshold Q_T, then the higher-quality view image is also compressed by using JPEG and the newly-compressed image will be the input for the computation of the cyclopean images.

The assumption is that when there is no obvious noise distortion being observed, the first modification alone is insufficient to describe the much dominant role of the view with massive blocking artifacts in the QA process (note that heavy noise can suppress the dominant role of the blocking artifact). Thus, the other higher-quality view image should be compensated by more blocking artifacts in order that the computed cyclopean image can get closer to the merged 3D view perceived by the HVS.

Specifically, we use the MUSIQUE scores obtained in the first stage of MUSIQUE-3D as the approximate quality estimates of the two monocular views, and their quality similarity is computed by $r = 2S_L S_R/(S_L^2 + S_R^2)$, where S_L and S_R denote the quality estimates of the left and right views, respectively. Then, the perceived cyclopean luminance and contrast images [denoted by C_i ($i = 1, 2$)] for each
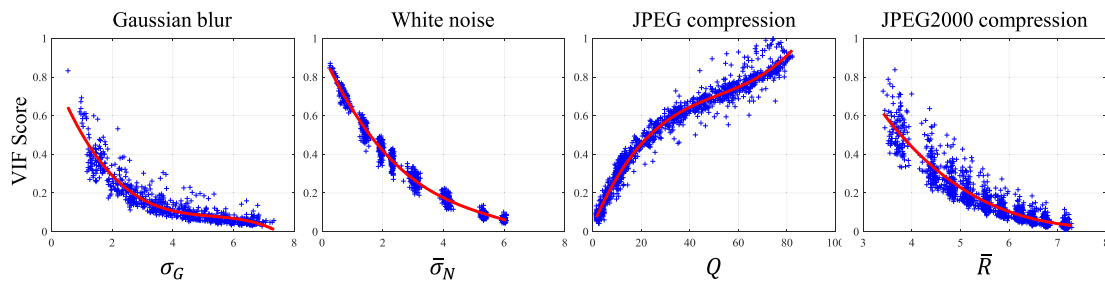
Fig. 4. Scatter plots of distortion parameter values versus VIF quality scores for images distorted by Gaussian blur, white noise, JPEG compression, and JPEG2000 compression, respectively. Note that for each scatter plot, the x axis represents the distortion parameter value and the y axis represents the corresponding VIF quality score computed from the distorted image regenerated by using that parameter value.
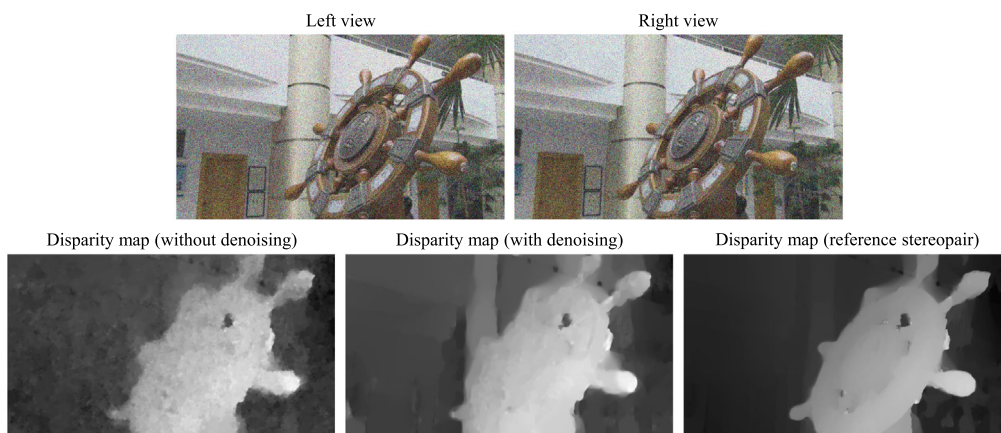
Fig. 5. Disparity maps computed for a reference stereopair as well as a multiply-distorted stereopair in the NBU-MDSID database [20,21] with and without denoising. Observe that when noise is present, the computed disparity map displays erroneous granule-shaped regions, whereas the denoised version captures the smooth, segment-based depth information.

$$C_i(x,y) = \frac{\left[\left(\eta_L I_{i,L}(x,y)\frac{E_L}{1+\frac{\epsilon_R(x,y)}{1+\beta\epsilon_L(x,y)}}\right)^{\gamma_2} + \left(\eta_R I_{i,R}(x-d_{x,y},y)\frac{E_R}{1+\frac{\alpha\epsilon_L(x-d_{x,y},y)}{1+\epsilon_R(x-d_{x,y},y)}}\right)^{\gamma_2}\right]^{\frac{1}{\gamma_2}}}{\left[\left(\frac{\eta_L E_L}{1+\frac{\epsilon_R(x,y)}{1+\beta\epsilon_L(x,y)}}\right)^{\gamma_2} + \left(\frac{\eta_R E_R}{1+\frac{\alpha\epsilon_L(x-d_{x,y},y)}{1+\epsilon_R(x-d_{x,y},y)}}\right)^{\gamma_2}\right]^{\frac{1}{\gamma_2}}}, (i=1,2) \tag{18}$$

pixel (x,y) are computed by Eq. (18), where $I_{i,L/R}$ $(i=1,2)$ denotes the luminance and contrast of the two view images; $d_{x,y} = D(x,y)$ denotes an estimated disparity index in D; E_L and E_R are the two compensation factors; all the other parameters (i.e., ϵ_L, ϵ_R, η_L, η_R, α, β, and γ_2) share the same meanings and values as they were previously defined in [29].

In this work, we define:

$$E_L = [s(\omega)]^{l(r)\cdot u(S_L - S_R)}, \tag{19}$$

$$E_R = [s(\omega)]^{l(r)\cdot u(S_R - S_L)}, \tag{20}$$

where $u(\cdot)$ is a step function; and, $l(\cdot)$ is a linear piecewise function which aims to control the weight compensation for the lower-quality view in building cyclopean images. We assume that $r < 0.8$ indicates an apparent quality difference between the two views, and $r > 0.9$ indicates considerable quality similarity. Accordingly, $l(r)$ is computed by

$$l(r) = \begin{cases} 1, & r < 0.8 \\ -10r + 9, & 0.8 \le r \le 0.9 \\ 0, & r > 0.9. \end{cases} \tag{21}$$

$s(\omega)$ is a sigmoid function which has the same form as Eq. (8). The function variable $\omega = 2\bar{C}_L\bar{C}_R/(\bar{C}_L^2 + \bar{C}_R^2)$ represents the contrast similarity between the two monocular views, where $\bar{C}_L$ and $\bar{C}_R$ are computed by

$$\bar{C}_{L/R} = \left[\frac{1}{N}\sum_{b=1}^{N} C_{L/R}(b)\right] \cdot \left[\frac{1}{N_T}\sum_{n_t=1}^{N_T} F_{L/R}(n_t)\right]. \tag{22}$$

Here, C_L and C_R denote the block-based contrast maps computed for the left and right views of a stereoscopic image by using the same approach in [29] (note that b is the block index and N is the total number of blocks); F_L and F_R denote the sharpness maps of the two views computed by using the FISH$_{bb}$ algorithm[2] [26]; n_t and N_T represent the index and total number of the blocks corresponding to the 1% largest sharpness values within the sharpness map. For $s(\omega)$, we set the following parameter values $A = 50$, $B = 1$, $t_1 = -20$, $t_2 = 0.75$ to provide a larger weight for the lower-quality view when ω is close to one.

As mentioned previously, given a noise-free or lightly-noised stereopair, the higher-quality view should be compensated by more block artifacts if the lower-quality view is deemed to be severely distorted by JPEG compression. Thus, for such stereopairs, we first simulate the JPEG compressed version of the higher-quality view by using a JPEG parameter Q given by

$$Q = \exp(\sqrt{\ln(1 + Q_L) \times \ln(1 + Q_R)}) - 1, \tag{23}$$

where Q_L and Q_R denote the JPEG parameters predicted for the left and right views respectively in the first stage of MUSIQUE-3D. Then, we use the original lower-quality view image and the newly-compressed view image to compute the cyclopean luminance and contrast images.

---

[2] Here in this work, we use the weight of $\{1, 2, 4\}$ (instead of $\{4, 2, 1\}$ in [26]) in combining the three per-level log-energy values in order that the energy of the middle-frequency bands are emphasized.

2.2.3. Cyclopean quality estimate

The quality estimation of the cyclopean images [defined in Eq. (18)] follows the same procedures in Section 2.1. Specifically, we trained another two-layer classification-regression framework to predict the four distortion parameters corresponding to the four distortion types in the cyclopean luminance and contrast images. The training data are the same quality-sensitive features employed in Section 2.1, but extracted from a large dataset of symmetrically-distorted 3D images generated by adding the four distortion types to the 50 pristine stereopairs among which 30 were obtained from the high-resolution stereo datasets [25] and the remaining 20 were captured by our own 3D camera. Note that the 3D training dataset contains only symmetrically-distorted stereopairs primarily for simplicity because the distortion parameters of a symmetrically-distorted stereopair are equal to the parameters of either of the two monocular views. Again, in order for these classification and regression models to work effectively, different features are required. A summary of the features required by each classification and regression models in the QC-MCM-based cyclopean IQA stage is shown in Table 4, in which "Luminance" and "Contrast" indicate that the corresponding features are extracted from the cyclopean luminance and cyclopean contrast images, respectively.

As shown in Table 4, the relative phase and across-scale correlation features (i.e., $f_{69}-f_{122}$) are only extracted from the cyclopean luminance image. This is due to the more abundant frequency components contained in the luminance map, which ultimately facilitate the phase and scale correlation analysis of the complex wavelet subbands. Specifically, the across-scale correlation features ($f_{93}-f_{122}$) describe the distribution of different frequency components on different wavelet scales and orientations, and thus are more sensitive to the noise. Conversely, the relative phase features ($f_{69}-f_{92}$), which describe the phase change for each individual wavelet subband, are more sensitive to the Gaussian blur, JPEG, and JPEG2000 compression. Consequently, $f_{93}-f_{122}$ are adopted whenever the training images contain noise distortion, and $f_{69}-f_{92}$ are adopted whenever the parameters of Gaussian blur/JPEG/JPEG2000 are predicted. In regards to the other features, we observe that the gradient-based LBP features ($f_{33}-f_{52}$) and the LDS-based log-Gabor domain features ($f_{53}-f_{68}$) are crucial and fundamental, and thus they are required by all classification and regression models. We also observe that the PS/LDS-based spatial domain features (f_1-f_{32}) are extracted from the cyclopean contrast image for noise estimation (same for $f_{33}-f_{68}$), and from the cyclopean luminance image to deal with the Gaussian blur/JPEG/JPEG2000 classification/regression task. This is due to the fact that noise is more observable in the contrast map which mostly contains the high-frequency components of an image. Similar to Section 2.1, the NSS features employed for the Gaussian blur and JPEG2000 parameters estimation (i.e., $f_{53}-f_{122}$ are not applied) in the cyclopean IQA stage are extracted from the sharper regions of the cyclopean luminance and contrast images, and the sharpness map is computed from the monocular view that has less quality degradation instead of the cyclopean images.

After predicting the four distortion parameters of the cyclopean luminance and contrast images, the final step of the cyclopean IQA stage is to map these parameter values to the corresponding VIF scores, which are then combined to yield the final cyclopean view quality. To this end, we employ the same approach in Section 2.1.3, and the only difference is the parameter setting. Specifically, we have found that the four curve parameters in Eq. (9), though originally

Table 4
A summary of the required features for all classification and regression models in the QC-MCM-based cyclopean IQA stage of MUSIQUE-3D. Note that "✓" indicates that the features are required by the corresponding models.

	Luminance						Contrast			
	f_1–f_{16}	f_{17}–f_{32}	f_{33}–f_{52}	f_{53}–f_{68}	f_{69}–f_{92}	f_{93}–f_{122}	f_1–f_{16}	f_{17}–f_{32}	f_{33}–f_{52}	f_{53}–f_{68}
Class-I						✓	✓	✓	✓	✓
Class-II	✓	✓	✓	✓	✓	✓			✓	✓
Class-III	✓	✓	✓	✓	✓				✓	✓
Regress-I-N						✓	✓	✓	✓	✓
Regress-II-G			✓	✓	✓	✓			✓	✓
Regress-III-G			✓	✓	✓				✓	✓
Regress-II-Q	✓	✓	✓	✓	✓	✓			✓	✓
Regress-III-Q	✓	✓	✓	✓	✓				✓	✓
Regress-II-R			✓	✓	✓	✓			✓	✓
Regress-III-R			✓	✓	✓				✓	✓

designed for 2D images, also work for cyclopean images. However, a cyclopean luminance/contrast image can be a bit more blurred due to inevitable computational errors in the disparity map. Thus, we set $D_{GR} = \max(D_G, D_R) - 0.05$ to take into account the unexpected blur distortion. In addition, we set $\beta_1 = 0$ in Eq. (11) and $D_N = D_N - 0.05$ in Eq. (16) to account for the increased noise strength in the cyclopean images as compared to each monocular view, and $\beta_3 = -0.2$ in Eq. (16) to account for the same masking effect caused by the noise distortion. All other parameters are assigned the same values as in Section 2.1.3.

2.3. MUSIQUE-3D quality estimate

Given the quality estimates obtained from the two monocular views and the cyclopean view (denoted by S_L, S_R, and S_{cyc}, respectively), the final stage of MUSIQUE-3D is to combine these scores into an overall perceived distortion estimate for the stereoscopic image. To this end, we first combine S_L and S_R based on a new contrast-weighting strategy which takes into account both the image's local contrast and the dominant role of the information-additive distortions. Accordingly, the combined quality estimate corresponding to the two monocular views (denoted by S_{2D}) is computed by

$$S_{2D} = \frac{\left[\frac{E_L P_L}{N} \sum_{b=1}^{N} C_L(b)\right] \cdot S_L + \left[\frac{E_R P_R}{N} \sum_{b=1}^{N} C_R(b)\right] \cdot S_R}{\left[\frac{E_L P_L}{N} \sum_{b=1}^{N} C_L(b)\right] + \left[\frac{E_R P_R}{N} \sum_{b=1}^{N} C_R(b)\right]}, \quad (24)$$

where E_L, E_R, C_L, and C_R share the same meanings and methods of computation as in Eqs. (19)–(22); P_L and P_R describe the weight compensation for the lower-quality view deemed to be severely distorted by JPEG compression. Accordingly, both P_L and P_R take values of one if the view with massive blocking artifacts does not take a dominant role in the QA process (i.e., the higher-quality view is not compensated by blocking artifact); otherwise, we define

$$P_L = \left[\frac{Q_T}{\min(Q_L, Q_R)}\right]^{(1-r) \cdot u(S_L - S_R)}, \quad (25)$$

$$P_R = \left[\frac{Q_T}{\min(Q_L, Q_R)}\right]^{(1-r) \cdot u(S_R - S_L)}, \quad (26)$$

where $Q_T = 15$ denotes the aforementioned threshold of the JPEG parameter, as we have observed that images compressed by using a JPEG parameter Q less than 15 will contain a large amount of blocking artifacts; $u(\cdot)$ is a step function; S_L, S_R, and r share the same meanings and methods of computation as in Eqs. (19)–(20); Q_L and Q_R share the same meanings as in Eq. (23).

The final quality estimate of the stereoscopic image (denoted by S_{3D}) is a scalar computed as a square product of S_{2D} and S_{cyc}, which is given by

$$S_{3D} = \sqrt{S_{2D} \times S_{cyc}}. \quad (27)$$

The assumption is that the HVS judges the quality of a 3D scene based on the merged cyclopean view created in the brain, whose appearance is influenced by the distortion in each 2D monocular view. Thus, the overall 3D image quality can be more precisely estimated by combining the two monocular view qualities with the merged cyclopean view quality. Although the exact combination strategy used by the HVS remains unknown, we have found that this simple geometric mean is an effective way to average the two scores while compensating for their different ranges. Note that smaller values of S_{3D} denote predictions of greater stereoscopic image quality.

3. Results

In this section, we analyze MUSIQUE-3D's ability to predict image quality by using various multiply and singly distorted stereoscopic image quality databases. We also compare the performance of MUSIQUE-3D with other FR and NR SIQA algorithms.

3.1. Training

As mentioned in Section 2, we trained MUSIQUE-3D's models on large datasets of both 2D and 3D images. For the 2D dataset, the distorted images were generated by adding the four distortion types (Gaussian blur, white noise, JPEG compression, and JPEG2000 compression) either singly or multiply to the 125 Berkeley segmentation database [24] images, as well as to the left views of the 20 high-resolution stereo datasets [25] images at various levels, which ultimately leads to a total of 33,940 distorted images. For the 3D dataset, the distorted images were generated by adding the same distortion types and each with the similar distortion levels as in the 2D dataset to the 50 pristine stereopairs, which ultimately leads to a total of 6610 symmetrically-distorted 3D images.

Specifically, for white noise, Gaussian blur, and JPEG compression, we used the same approach as in [23] to generate the training data. For JPEG2000 compression, we applied a discrete wavelet transform to obtain wavelet subbands coefficients, which were then quantized and entropy coded to form the output bit stream with compression ratio R. For the multiple distortions, we followed the same procedure as in [30]: the Gaussian blur is first introduced, followed by JPEG or JPEG2000 compression, and finally the white noise. Detailed information of the training data is shown in Table 5.

After generating the distorted images, their corresponding distortion types and parameter values (σ_G, $\bar{\sigma}_N$, Q, and $\bar{R}$) were saved for training the classification and regression models. First, three support vector classification (SVC) models (Class-I/II/III) were trained to decide the type of distortion that an image may contain as well as its probability. Let **p** denote the k-dimensional vector of probabilities, and its elements are denoted by $p(i)$ ($i = 1, 2, \ldots, k$). For any classification feature vector **x** and its corresponding label y, the goal is to estimate $p(i) = P(y = i|\mathbf{x})$. Based on the "one-against-one" approach [31] for multiclass classification, the pairwise class probabilities (i.e., $r_{ij} \approx P(y = i|y = i \text{ or } j, \mathbf{x})$) are estimated by using method proposed in [32]. After

Table 5
Details of the generated training dataset.

Distortion type	Distortion level (2D)	# 2D image	Distortion level (3D)	# 3D image
Gblur	12	1500	12	600
JPEG	10	1250	15	750
JP2K	10	1250	10	300
WN	8	1000	8	400
Gblur + JPEG	40	5000	40	800
Gblur + WN	40	5000	40	800
JPEG + WN	24	3000	24	480
JP2K + WN	24	3000	24	480
Gblur + JPEG + WN	100	12 500	100	2000

collecting all r_{ij} values, the method in [33] is employed to obtain $p(i)$ by solving the following optimization problem:

$$\min_{p} \quad \frac{1}{2} \sum_{i=1}^{k} \sum_{j:j\neq i} \left(r_{ji} p(i) - r_{ij} p(j) \right)^2 \tag{28}$$

$$\text{subject to} \quad p(i) \geq 0, \forall i, \quad \sum_{i=1}^{k} p(i) = 1.$$

More details for solving Eq. (28) are referred to [33].

Next, the ϵ-SVR [34] approach is employed to train seven support vector regression (SVR) models (Regress-I-N, Regress-II/III-G/Q/R) to predict the four distortion parameters in different cases. Specifically, given a set of training data $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)$ where $\mathbf{x}_l \in \mathbb{R}^d$ and $y \in \mathbb{R}$, ϵ-SVR aims to find a function, which takes the form

$$f(\mathbf{x}) = \langle w, \phi(\mathbf{x}) \rangle + b, \tag{29}$$

such that $f(\mathbf{x})$ has at most ϵ deviation from the desired target y for all the training data, and at the same time is as flat as possible. In Eq. (29), $\phi(\mathbf{x}) \in \mathbb{R}^{d'}$ ($d' > d$) is a non-linear projection, which maps the data from the original d-dimensional domain to a higher d'-dimensional feature space; $w \in \mathbb{R}^d$ is the weight vector, and $b \in \mathbb{R}$ is the bias. This can be done by solving the following optimization problem:

$$\min_{w,b,\xi,\xi^*} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l} \left(\xi_i + \xi_i^* \right)$$

$$\text{subject to} \quad \begin{cases} y_i - \langle w, \phi(\mathbf{x}_i) \rangle - b & \leq & \epsilon + \xi_i \\ \langle w, \phi(\mathbf{x}_i) \rangle + b - y_i & \leq & \epsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq & 0 \end{cases} \tag{30}$$

where ξ_i and ξ_i^* are the slack variables introduced to allow for small errors; $C > 0$ determines the trade-off between the flatness of f and the amount up to which deviations larger than ϵ are tolerated. By using Lagrange multipliers and a quadratic programming solver, the final solution of Eq. (30) is given by

$$f(\mathbf{x}) = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^* \right) \cdot K\left(\mathbf{x_i}, \mathbf{x} \right) + b, \tag{31}$$

where $K(\mathbf{x_i}, \mathbf{x}) = \phi(\mathbf{x_i})^T \phi(\mathbf{x})$ is a kernel function; α_i and α_i^* are Lagrange multipliers, which are both zeros if $\mathbf{x_i}$ does not contribute to the error function. More details for training ϵ-SVR models are referred to [35,36]. We use the LIBSVM package [35] to implement the training, and the optimal radial basis function kernel parameters to improve the predictive performance.

3.2. Testing

To evaluate the performance of MUSIQUE-3D in predicting the quality of MDSIs and SDSIs, five databases of subjective image quality were used: (1) the NBU-MDSID database [20,21], (2) the LIVE 3D database [37], (3) the WaterlooIVC 3D database [38], (4) the NBU 3D database [39], and (5) the IRCCyN/IVC 3D database [40]. Among the three databases that were constructed in two phases (NBU-MDSID, LIVE 3D, and WaterlooIVC 3D), Phase I of NBU-MDSID and LIVE 3D contain symmetrically-distorted images, and Phase II of NBU-MDSID and LIVE 3D contain asymmetrically-distorted images. For WaterlooIVC 3D, both phases contain symmetrically- and asymmetrically-distorted images. Among all the five databases, NBU-MDSID contains multiple distortions, and the other four contain only single distortions. For NBU-MDSID and IRCCyN/IVC 3D, all images in the two databases were used for testing. For the other three, only images with their two monocular views corrupted by the same four distortion types were selected for testing. Thus, the number of test images were 270 for NBU-MDSID Phase I (NBUMD-I), 300 for NBU-MDSID Phase II (NBUMD-II), 285 for LIVE 3D Phase I (LIVE-I), 288 for LIVE 3D Phase II (LIVE-II), 258 for WaterlooIVC 3D Phase I (WIVC-I), 340 for WaterlooIVC 3D Phase II (WIVC-II), 240 for NBU 3D, and 90 for IRCCyN/IVC 3D.

We compared MUSIQUE-3D with various FR and NR IQA methods for which code is publicly available. The four FR 3D methods were the cyclopean MS-SSIM proposed by Chen et al. [22], the frequency-integrated PSNR proposed by Lin et al. [41], the BJND-based method proposed by Shao et al. [39], and the stereoscopic omnidirectional image quality evaluator (SOIQE) proposed by Chen et al. [42]. The ten NR methods were: Saliency-guided binocular model proposed by Xu et al. [9], S3D INtegrated Quality (SINQ) Predictor [10], deep convolutional neural network (DCNN) [43], blind/referenceless image spatial quality evaluator (BRISQUE) [15], Laplacian of Gaussian based model (GM-LOG) [44], gradient weighted histogram of LBP model (GLBP) [45], quality-aware clustering model (QAC) [46], natural image quality evaluator (NIQE) [47], integrated local NIQE (ILNIQE) [48], six-step blind metric (SISBLIM) [49]. Among these NR methods, the three SIQA methods (Xu's method, SINQ, and DCNN) were trained on the LIVE 3D phase II database by referring to [9,10,43]. The three opinion-aware methods (BRISQUE, GM-LOG, and GLBP) were trained on the cyclopean images computed from the same 6610 symmetrically-distorted 3D images by using MCM [29], and their qualities were estimated by averaging the two VIF [27] scores predicted from the two views. For the four opinion-unaware NR methods (QAC, NIQE, ILNIQE, and SISBLIM), the predicted quality of a stereoscopic image was taken to be the weighted quality predicted from the left and right views, where the weights were determined based on a block-based contrast measure [29]. After applying a four-parameter logistic transform to bring the algorithm's raw predicted scores to the same scales of the MOS/DMOS values, four common criteria were used to measure the prediction monotonicity and prediction accuracy of each algorithm: (1) the Pearson Linear Correlation Coefficient (PLCC), (2) the Spearman Rank-Order Correlation Coefficient (SROCC), (3) the Kendall Rank-Order Correlation Coefficient (KROCC), and (4) the Root Mean Square Error (RMSE). Also, as advocated in [50,51], we use three additional criteria to take into account the uncertainty in subjective scores: (1) the Area Under the Curve (AUC) values for different vs. similar analysis (AUC-DS), (2) AUC values for better vs. worse analysis (AUC-BW), and (3) correct classification in better vs. worse analysis (CC-BW). Thus, seven criteria in total were used to measure the performance of each algorithm. Note that for databases that do not provide standard deviation of the subjective scores, we simply use 10% of the MOS/DMOS range. As we have found, this choice does not affect the relative rankings among different algorithms. Also note that the logistic transform will affect only PLCC and RMSE, not SROCC nor KROCC.

3.3. Performance on multiply distorted 3D image databases

In this section, we evaluate the quality predictive performances of MUSIQUE-3D and other FR/NR IQA algorithms on MDSIs. To this end, the NBU-MDSID Phase I and Phase II databases which contain distorted stereoscopic images simultaneously corrupted by three distortion types (Gaussian blur, JPEG compression, white noise) were used for testing. The overall testing results in terms of PLCC, SROCC, KROCC, RMSE, AUC-DC, AUC-BW, and CC-BW on the two phases are shown in Table 6, in which italicized entries denote the FR IQA algorithms. Also included are the testing results of the first and the second stage of MUSIQUE-3D (denoted by "MUSIQUE-2D" and "MUSIQUE-Cyc", respectively) for comparison. The results of the best-performing FR IQA algorithm are italicized and bolded, and the results of the best-performing NR IQA algorithm are bolded.

As shown in Table 6, MUSIQUE-3D achieves the best performance on both NBU-MDSID Phase I and Phase II databases as compared with other FR/NR IQA algorithms. Compared with MUSIQUE-2D and MUSIQUE-Cyc, we observe that the combined stages can improve the algorithm's overall performance. We also observe that the performance of MUSIQUE-Cyc is often not as good as MUSIQUE-2D (as shown in Table 7 also), and the reasons are twofold. First, the disparity map estimated based on distorted stereopairs is imperfect, and consequently the cyclopean images computed by using the map have some errors. Also, the QC-MCM employed in this work, though improved upon compared to the previous version in [29], does not fully model the binocular processing of the HVS. All of these potential errors are inevitably introduced in the second cyclopean IQA stage which ultimately decrease the QA performance on the cyclopean images. Second, the training data generated for the second stage contain only symmetrically-distorted stereopairs with the main purpose of easily collecting the ground-truth distortion parameters for model training, while database such as NBU-MDSID Phase II contains asymmetrically-distorted stereopairs. In other words, we are actually training symmetrically-distorted cyclopean IQA models to predict the quality of asymmetrically-distorted stereopairs, which inevitably brings prediction errors. Despite that, combining the cyclopean IQA stage is meaningful and necessary, especially when considering that humans judge 3D image quality based mainly on a merged cyclopean view instead of two separate monocular views. As demonstrated in the following section, combining the two stages can also greatly improve the algorithm performance on the LIVE 3D, WaterlooIVC 3D, and IRCCyN/IVC 3D databases.

3.4. Performance on singly distorted 3D image databases

Although MUSIQUE-3D was developed mainly for predicting the quality of MDSIs, we demonstrate that it also works for SDSIs. To this end, the subsets of four singly-distorted stereoscopic image quality databases corresponding to the individual distortion types of Gaussian blur, white noise, JPEG compression, and JPEG2000 compression were used for testing. The overall testing results of MUSIQUE-3D and other FR/NR IQA algorithms on the four databases are shown in Table 7, in which the averaged values of the PLCC, SROCC, KROCC, AUC-DC, AUC-BW, and CC-BW results were weighted by the actual number of distorted images tested in each database. Note that we do not compute the averaged RMSE value because different databases have different MOS/DMOS ranges. As in Section 3.3, the testing results of the two stages in MUSIQUE-3D are also listed in Table 7, which along with MUSIQUE-3D are denoted by "M-2D", "M-Cyc", and "M-3D", respectively, for brevity. Moreover, we additionally report the performances of these FR/NR IQA algorithms on the four databases corresponding to each individual distortion type. Test results of the four distortion types in LIVE and NBU, and three distortion types in WaterlooIVC and IRCCyN/IVC are presented in Table 8 in terms of SROCC values (additional test results are available in the online supplement to this paper at http://vision.eng.shizuoka.ac.jp/MUSIQUE3D/). For both tables, italicized entries denote the FR IQA algorithms. The results of the

best-performing FR IQA algorithm are italicized and bolded, and the results of the best-performing NR IQA algorithm are bolded. Note that Xu's method [9], SINQ [10], and DCNN [43] were trained on the LIVE 3D Phase II database, and thus their testing results on LIVE-II are not presented in Tables 7 and 8.

As shown in Table 7, MUSIQUE-3D achieves better results (on average) as compared to other FR/NR IQA algorithms on subsets of all four databases. Specifically, on the LIVE 3D, WaterlooIVC 3D, and IRCCyN/IVC 3D databases, MUSIQUE-3D is superior to all the other NR IQA algorithms considered, and even challenges the FR methods. On the NBU 3D database, however, the performance of MUSIQUE-3D drops slightly, which might be attributable to the relatively lower QA performance on the WN images (see Table 8), but is still very competitive and promising. Compared with its two individual stages, we observe that the combination can generally improve or balance the performance of each individual stage across different 3D image databases.

By comparing M-2D with QAC, NIQE, ILNIQE, and SISBLIM in Tables 6 and 7, we observe performance improvements on most stereoscopic image databases, which demonstrates the effectiveness of the proposed two-layer classification/regression models as well as the contrast-weighting strategy in predicting the quality of stereoscopic images. Also, the better performance of M-Cyc compared with BRISUQE, GM-LOG, and GLBP demonstrates the effectiveness of the proposed cyclopean IQA model. Note that although Xu's method, SINQ, and DCNN were trained on SDSIs with associated human opinion scores, their relatively weak performances on WaterlooIVC 3D, NBU 3D, and IRCCyN/IVC 3D indicate that training on one stereoscopic image quality database cannot always guarantee good performances on others because different databases may significantly vary in their distortion types and levels. Also note that most of the FR and NR approaches considered here perform less effectively on the IRCCyN/IVC 3D database. The reason is that these algorithms are insensitive to the resampled images, which are also considered as blur distortion in IRCCyN/IVC 3D.

Fig. 6 shows the scatter plots of logistic-transformed MUSIQUE-3D quality predictions versus subjective ratings (MOS or DMOS) on subsets of all five databases. Scatter plots corresponding to other SIQA algorithms can be referred to http://vision.eng.shizuoka.ac.jp/MUSIQUE3D/. Note that the x axis represents the predicted value after logistic transformation; the y axis represents the true DMOS value for the NBU-MDSID, LIVE 3D, NBU 3D, and ICCyN/IVC 3D databases, and true MOS value for the WaterlooIVC 3D database. Despite the presence of some outliers, the plots are generally heteroscedastic. Fig. 7 showcases the quality scores given by MUSIQUE-3D to sample distorted stereoscopic images from the LIVE 3D Phase I database. Note that S_L and S_R denote quality estimates given to the left and right views, respectively; S_{2D}, S_{cyc}, and S_{3D} denote quality estimates given by the first, second, and final stage of MUSIQUE-3D (we refer readers to http://vision.eng.shizuoka.ac.jp/MUSIQUE3D/ for more visual examples). In summary, when looking at the overall performance across all databases, MUSIQUE-3D demonstrates a better performance than other NR IQA methods.

3.5. Discussion

In this section, we explore alternative ways to potentially improve the performance of our algorithm, and at the same time discuss their value/efficacy. As mentioned in Section 2.2.1, a denoising algorithm can be applied to prevent the erroneous granule-shaped regions in the estimated disparity map. Here, we investigate the impacts of two denoising algorithms on the QA performance of MUSIQUE-3D. One is the block-matching and 3D filtering (BM3D) color-image denoising method (C-BM3D) [52], and the other one is the denoising convolutional neural network (DnCNN) [53]. We also investigate how the performance changes when different optical flow algorithms and classification/regression models are employed. Specifically, the SIFT flow

Table 6
Overall performances of MUSIQUE-3D and other FR/NR IQA algorithms on the NBU-MDSID Phase I and Phase II databases. Results of the best-performing FR algorithm are italicized and bolded, and results of the best-performing NR algorithm are bolded.

IQA metrics	NBU-MDSID Phase I							NBU-MDSID Phase II						
	PLCC	SROCC	KROCC	RMSE	AUC-DC	AUC-BW	CC-BW	PLCC	SROCC	KROCC	RMSE	AUC-DC	AUC-BW	CC-BW
Chen [22]	*0.915*	*0.899*	*0.716*	*3.860*	*0.757*	*0.960*	*0.891*	*0.835*	***0.822***	*0.634*	*6.610*	*0.711*	*0.924*	*0.852*
Lin [41]	***0.931***	***0.917***	***0.738***	***3.484***	***0.839***	***0.972***	***0.903***	*0.832*	*0.797*	*0.611*	*6.669*	*0.703*	*0.923*	*0.841*
Shao [39]	*0.827*	*0.807*	*0.608*	*5.377*	*0.681*	*0.915*	*0.832*	*0.732*	*0.721*	*0.534*	*12.020*	*0.633*	*0.879*	*0.798*
SOIQE [42]	*0.807*	*0.766*	*0.570*	*5.658*	*0.682*	*0.899*	*0.811*	***0.858***	***0.822***	***0.637***	***6.172***	***0.726***	***0.936***	***0.855***
Xu [9]	0.659	0.671	0.515	7.200	0.724	0.172	0.221	0.425	0.501	0.360	12.020	0.568	0.252	0.299
SINQ [10]	0.828	0.822	0.607	5.363	0.755	0.920	0.834	0.685	0.676	0.482	8.760	0.631	0.848	0.770
DCNN [43]	0.755	0.737	0.519	6.282	0.686	0.864	0.787	0.635	0.619	0.440	9.281	0.588	0.827	0.744
BRISQUE [15]	0.750	0.738	0.517	9.573	0.769	0.893	0.788	0.526	0.559	0.386	10.225	0.587	0.790	0.716
GM-LOG [44]	0.613	0.160	0.078	7.564	0.662	0.644	0.544	0.424	0.314	0.225	10.884	0.587	0.646	0.628
GLBP [45]	0.420	0.395	0.270	9.573	0.579	0.707	0.646	0.337	0.275	0.189	11.318	0.524	0.649	0.608
QAC [46]	0.845	0.855	0.651	5.121	0.793	0.934	0.860	0.724	0.740	0.548	8.294	0.633	0.872	0.807
NIQE [47]	0.926	0.901	0.718	3.608	0.796	0.967	0.894	0.849	0.840	0.649	6.349	0.709	0.925	0.863
ILNIQE [48]	0.708	0.592	0.427	6.765	0.660	0.824	0.734	0.716	0.682	0.488	8.391	0.618	0.859	0.772
SISBLIM [49]	0.599	0.706	0.631	7.663	0.567	0.913	0.846	0.479	0.587	0.530	10.552	0.613	0.845	0.797
MUSIQUE-2D	0.947	0.930	0.766	3.084	0.854	**0.981**	0.918	0.897	0.874	0.692	5.315	0.791	0.958	0.886
MUSIQUE-Cyc	0.916	0.904	0.717	3.845	0.819	0.965	0.894	0.837	0.831	0.636	6.572	0.732	0.931	0.855
MUSIQUE-3D	**0.948**	**0.934**	**0.770**	**3.047**	**0.856**	0.980	**0.921**	**0.900**	**0.884**	**0.704**	**5.251**	**0.803**	**0.960**	**0.892**
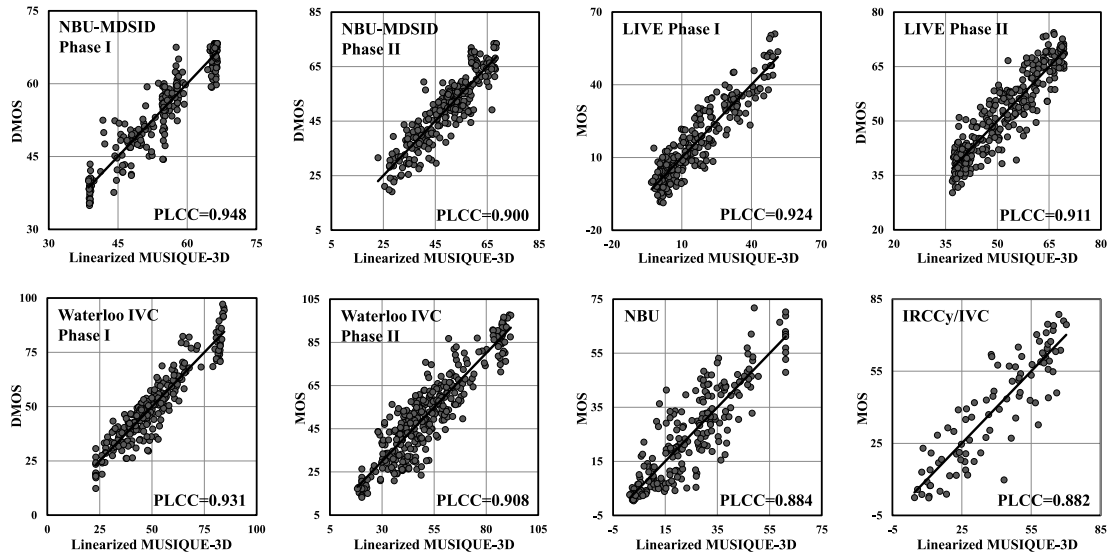
Fig. 6. Scatter plots of objective scores predicted by MUSIQUE-3D algorithm after logistic transform versus subjective scores on different 3D image databases.

algorithm is replaced by a scalable Lucas-Kanade optical flow (SLK flow) [54], and the SVM model is replaced by a five-layer back propagation (BP) neural network.[3] Accordingly, the following comparison versions were created:

- C-BM3D: stereopairs were first denoised by the C-BM3D algorithm before being used for disparity map estimation;
- DnCNN: stereopairs were first denoised by DnCNN before being used for disparity map estimation;
- SLK: SIFT flow was replaced by SLK flow;
- BP: SVM was replaced by BP neural network.

Note that each version was built by training on the same 33,940 regenerated 2D images and 6,610 regenerated 3D images, and all algorithm settings were the same as that used in training MUSIQUE-3D except the replaced parts. The testing was performed on the subsets of images corrupted by the four distortion types from five databases (see Section 3.2), and the results are shown in Table 9 in which the averaged

values were weighted by the actual number of distorted images tested in each database. Also shown in Table 9 are the testing results of the original MUSIQUE-3D algorithm for reference.

As observed in Table 9, MUSIQUE-3D achieves the best performance (on average) among its four variants. Specifically, by looking at the results of C-BM3D and DnCNN, we conclude that the employment of noise reduction can indeed benefit our algorithm on some databases (e.g., LIVE 3D Phase II). However, this performance improvement is less attractive when considering the added computational complexity; furthermore, the performance on other databases either drops or maintains the same level. For SLK and BP, we observe that the algorithm performance drops on most databases, and this fact implies that the performance of MUSIQUE-3D is dependent on the selected optical flow algorithm and machine learning model. Note that for C-BM3D, DnCNN, and SLK, their S_{2D} results do not change significantly, owning to the fact that both noise reduction and optical flow estimation belong to the second stage of MUSIQUE-3D. Also note that the S_{2D} and S_{cyc} results of using BP network drop considerably, suggesting that a more powerful and complicated network might be required to more accurately predict the distortion parameters. Thus, based on current test, the SVM model is a promising choice in this work due to its decent performance, simplicity, and easy implementation.

---

[3] In this work, the numbers of nodes for the three hidden layers are set to 40, 80, and 60, respectively, and we use the same network architecture to train all classification/regression models.

Table 7

Overall performances of MUSIQUE-3D and other FR/NR IQA algorithms on subsets of the LIVE 3D, WaterlooIVC 3D, NBU 3D, and IRCCyN/IVC 3D image databases. Results of the best-performing FR IQA algorithm are italicized and bolded, and results of the best-performing NR IQA algorithm are bolded.

		Chen	Lin	Shao	SOIQE	Xu	SINQ	DCNN	BRISQUE	GM-LOG	GLBP	QAC	NIQE	ILNIQE	SISBLIM	M-2D	M-Cyc	M-3D
PLCC	LIVE-I	0.930	0.872	**0.933**	0.799	0.742	0.901	0.866	0.718	0.774	0.795	0.893	0.858	0.897	0.820	0.916	0.908	**0.924**
	LIVE-II	**0.911**	0.655	0.819	0.824	–	–	–	0.778	0.750	0.767	0.846	0.832	0.784	0.641	0.891	0.880	**0.911**
	WIVC-I	0.711	0.688	**0.833**	0.830	0.157	0.419	0.667	0.695	0.729	0.739	0.741	0.776	0.731	0.894	0.909	0.911	**0.931**
	WIVC-II	0.573	0.578	0.737	**0.751**	0.035	0.356	0.640	0.648	0.673	0.691	0.702	0.668	0.653	0.802	0.876	0.894	**0.908**
	NBU 3D	0.880	**0.932**	0.918	0.874	0.461	0.574	0.786	0.791	0.799	0.760	0.800	0.817	**0.906**	0.682	0.886	0.861	0.884
	IVC 3D	0.689	0.697	0.719	**0.753**	0.587	0.452	0.762	0.576	0.317	0.485	0.693	0.535	0.672	0.574	0.842	0.803	**0.882**
	Average	0.785	0.731	**0.835**	0.808	–	–	–	0.713	0.715	0.732	0.788	0.770	0.779	0.757	0.892	0.886	**0.910**
SROCC	LIVE-I	0.903	0.830	**0.904**	0.870	0.738	0.849	0.885	0.684	0.751	0.757	**0.891**	0.825	0.870	0.786	0.888	0.871	**0.890**
	LIVE-II	**0.904**	0.639	0.797	0.815	–	–	–	0.757	0.723	0.727	0.828	0.820	0.764	0.528	0.887	0.865	**0.903**
	WIVC-I	0.626	0.611	**0.817**	0.783	0.069	0.259	0.472	0.591	0.653	0.624	0.547	0.601	0.696	0.872	0.897	0.884	**0.918**
	WIVC-II	0.489	0.479	**0.721**	0.717	0.183	0.150	0.530	0.523	0.622	0.611	0.541	0.515	0.618	0.778	0.849	0.862	**0.885**
	NBU 3D	0.887	**0.939**	0.924	0.885	0.420	0.638	0.826	0.830	0.798	0.796	0.823	0.840	**0.906**	0.636	0.884	0.862	0.885
	IVC 3D	**0.673**	0.636	0.649	0.653	0.551	0.386	0.761	0.543	0.228	0.431	0.684	0.417	0.643	0.609	0.844	0.798	**0.877**
	Average	0.745	0.682	**0.815**	0.799	–	–	–	0.660	0.676	0.682	0.717	0.693	0.755	0.715	0.877	0.864	**0.895**
KROCC	LIVE-I	0.726	0.637	**0.734**	0.681	0.503	0.657	0.685	0.474	0.552	0.543	0.704	0.625	0.676	0.589	0.703	0.684	**0.707**
	LIVE-II	**0.731**	0.480	0.605	0.651	–	–	–	0.551	0.528	0.531	0.630	0.612	0.566	0.395	0.702	0.667	**0.720**
	WIVC-I	0.469	0.458	**0.636**	0.596	0.062	0.172	0.319	0.429	0.482	0.449	0.393	0.456	0.523	0.706	0.723	0.711	**0.764**
	WIVC-II	0.346	0.344	**0.535**	0.527	0.017	0.110	0.369	0.367	0.453	0.444	0.385	0.366	0.451	0.596	0.662	0.681	**0.707**
	NBU 3D	0.701	**0.781**	0.757	0.699	0.291	0.382	0.621	0.633	0.601	0.604	0.620	0.639	0.730	0.451	0.702	0.677	0.704
	IVC 3D	0.502	0.483	0.498	**0.527**	0.380	0.265	0.564	0.383	0.161	0.249	0.498	0.289	0.461	0.437	0.639	0.608	**0.694**
	Average	0.579	0.523	**0.637**	0.619	–	–	–	0.477	0.497	0.494	0.538	0.517	0.573	0.542	0.693	0.679	**0.718**
RMSE	LIVE-I	5.723	7.620	**5.576**	15.532	10.424	6.746	7.774	10.816	9.832	9.430	6.988	7.979	6.858	8.895	6.226	6.520	**5.949**
	LIVE-II	**4.619**	8.452	6.409	6.338	–	–	–	7.025	7.401	7.169	5.969	6.200	6.942	8.583	5.076	5.321	**4.613**
	WIVC-I	11.936	12.319	**9.397**	9.457	16.950	15.405	12.641	12.200	11.608	11.427	11.395	10.704	11.583	7.606	7.073	6.990	**6.174**
	WIVC-II	16.627	16.556	13.705	**13.403**	20.286	18.962	15.593	15.453	15.017	14.678	14.459	15.101	15.377	12.116	9.777	9.080	**8.488**
	NBU 3D	8.322	**6.344**	6.933	8.517	15.545	14.344	10.826	10.719	10.528	11.381	10.503	10.092	7.427	12.818	8.117	8.911	**8.197**
	IVC 3D	15.998	15.822	15.326	**14.507**	17.866	19.676	14.290	18.032	20.926	19.288	15.909	18.644	16.331	18.064	11.917	13.153	**10.416**
AUC-DC	LIVE-I	0.805	0.738	**0.812**	0.759	0.686	0.781	0.782	0.612	0.616	0.645	0.767	0.726	0.763	0.711	0.777	0.778	**0.791**
	LIVE-II	**0.750**	0.621	0.681	0.705	–	–	–	0.667	0.658	0.654	0.705	0.709	0.657	0.627	0.753	0.734	**0.775**
	WIVC-I	0.626	0.632	**0.699**	0.663	0.559	0.516	0.615	0.649	0.662	0.654	0.620	0.615	0.631	0.755	0.778	0.778	**0.818**
	WIVC-II	0.554	0.582	**0.630**	0.628	0.552	0.531	0.611	0.640	0.638	0.641	0.621	0.577	0.606	0.676	0.735	0.747	**0.764**
	NBU 3D	0.739	**0.824**	0.797	0.747	0.543	0.556	0.674	0.709	0.669	0.681	0.678	0.712	**0.774**	0.616	0.748	0.732	0.756
	IVC 3D	0.601	0.599	0.603	**0.645**	0.562	0.532	0.644	0.558	0.503	0.537	0.594	0.539	0.575	0.590	0.700	0.699	**0.747**
	Average	0.684	0.667	**0.711**	0.694	–	–	–	0.648	0.639	0.647	0.672	0.656	0.675	0.672	0.754	0.750	**0.778**
AUC-BW	LIVE-I	0.977	0.949	**0.980**	0.958	0.891	0.961	0.965	0.842	0.898	0.903	0.967	0.943	0.961	0.920	0.972	0.964	**0.974**
	LIVE-II	**0.966**	0.820	0.923	0.922	–	–	–	0.893	0.885	0.882	0.938	0.931	0.903	0.811	0.962	0.951	**0.969**
	WIVC-I	0.831	0.830	**0.936**	0.918	0.493	0.608	0.786	0.826	0.858	0.853	0.795	0.821	0.884	0.963	0.970	0.969	**0.981**
	WIVC-II	0.737	0.757	**0.874**	0.873	0.570	0.581	0.798	0.796	0.831	0.834	0.780	0.754	0.833	0.910	0.948	0.955	**0.964**
	NBU 3D	0.958	**0.985**	0.979	0.956	0.709	0.805	0.936	0.939	0.930	0.928	0.938	0.946	**0.973**	0.854	0.966	0.958	0.967
	IVC 3D	**0.836**	0.811	0.804	0.782	0.799	0.707	0.896	0.774	0.615	0.688	0.852	0.696	0.839	0.828	0.937	0.917	**0.957**
	Average	0.884	0.858	**0.927**	0.914	–	–	–	0.850	0.862	0.866	0.878	0.863	0.902	0.888	0.961	0.957	**0.970**
CC-BW	LIVE-I	**0.920**	0.870	0.916	0.894	0.795	0.882	0.895	0.777	0.823	0.819	0.906	0.864	0.894	0.841	0.905	0.896	**0.907**
	LIVE-II	**0.910**	0.771	0.840	0.864	–	–	–	0.812	0.798	0.802	0.857	0.845	0.819	0.723	0.893	0.876	**0.903**
	WIVC-I	0.765	0.760	**0.862**	0.838	0.467	0.597	0.683	0.746	0.777	0.757	0.725	0.758	0.799	0.900	0.909	0.904	**0.930**
	WIVC-II	0.693	0.691	**0.797**	0.793	0.510	0.554	0.706	0.706	0.754	0.748	0.714	0.704	0.752	0.833	0.868	0.880	**0.893**
	NBU 3D	0.904	**0.948**	0.931	0.903	0.666	0.722	0.856	0.864	0.847	0.847	0.859	0.869	**0.918**	0.764	0.903	0.886	0.903
	IVC 3D	0.783	0.769	0.772	**0.789**	0.716	0.652	0.813	0.710	0.584	0.636	0.777	0.662	0.760	0.751	0.856	0.839	**0.884**
	Average	0.829	0.798	**0.859**	0.851	–	–	–	0.772	0.784	0.783	0.807	0.795	0.827	0.809	0.892	0.885	**0.905**

4. Conclusion

This paper presented an opinion-unaware NR algorithm for quality assessment of multiply and singly distorted stereoscopic images via adaptive construction of cyclopean views. Our method, called MUSIQUE-3D, operates under the principle that the quality degradation of a 3D image can be represented by the distortion parameters of both the monocular and cyclopean views, and can thereby lead to an improved estimate of quality. Accordingly, MUSIQUE-3D contains three main stages: (1) MUSIQUE-based quality assessment of the two monocular views; (2) QC-MCM-based quality assessment of the cyclopean view; and (3) quality combination of the monocular and cyclopean views to yield the final quality score. Based on the results presented in this paper, we can make the following conclusions: (1) In regards to the first stage, our results suggest that it is not only important to accurately predict the quality of each view separately, but also to properly combine them such that the combined quality estimate can be representative of the overall 3D image quality. (2) In regards

to the second stage, our results suggest that it is necessary to use a quality compensation mechanism such that the non-compensable quality-degradation effect caused by the information-additive distortions (e.g., JPEG compression) in binocular vision can be captured. (3) For the third stage of predicting the overall 3D image quality, we observe that the QA estimates of both previous stages have to be sufficiently veridical in order that the combined performance can be further enhanced.

Despite the effectiveness of the proposed MUSIQUE-3D algorithm, there are number of aspects which could benefit from future research. One line of future work would involve improving the accuracy of computing the disparity maps, or better still, to perform the QA task on stereoscopic images without using such maps. Other future work would involve exploiting new methods to extract more efficient features for quality prediction (e.g., using pre-trained deep neural networks) instead of the hand-crafted features employed in MUSIQUE-3D. Finally, the important QA behaviors/properties of the HVS during binocular vision could certainly benefit from future psychophysical research and

Table 8

SROCC values of MUSIQUE-3D and other FR/NR IQA algorithms on different types of distortion on the LIVE 3D, WaterlooIVC 3D, NBU 3D and IRCCyN/IVC 3D databases. Results of the best-performing FR IQA algorithm are italicized and bolded, and results of the best-performing NR IQA algorithm are bolded.

		Chen	Lin	Shao	SOIQE	Xu	SINQ	DCNN	BRISQUE	GM-LOG	GLBP	QAC	NIQE	ILNIQE	SISBLIM	MUSIQUE-3D
LIVE-I	WN	*0.948*	0.928	0.930	0.925	0.088	0.910	0.619	0.731	0.891	0.760	0.855	0.914	0.926	0.931	**0.935**
	JP2K	0.896	0.839	0.883	*0.902*	0.841	0.819	**0.919**	0.477	0.664	0.749	0.917	0.744	0.845	0.853	0.869
	JPEG	0.558	0.207	0.611	*0.618*	0.382	0.561	**0.730**	0.479	0.491	0.218	0.701	0.597	0.587	0.678	0.673
	Gblur	0.926	*0.935*	0.910	0.911	0.813	0.882	0.884	0.795	0.866	0.850	0.894	0.881	0.885	**0.901**	0.879
LIVE-II	WN	0.955	0.907	0.807	*0.963*	–	–	–	0.908	0.841	0.571	0.686	0.895	**0.948**	0.863	0.917
	JP2K	0.833	0.719	0.788	*0.915*	–	–	–	0.586	0.413	0.654	0.782	0.632	0.768	0.634	**0.860**
	JPEG	*0.840*	0.613	0.745	0.833	–	–	–	0.601	0.596	0.624	0.809	0.637	0.663	**0.871**	0.832
	Gblur	0.910	0.711	*0.939*	0.820	–	–	–	0.754	0.842	0.846	0.846	0.864	0.889	**0.900**	0.877
WIVC- I	WN	0.837	0.784	0.829	*0.912*	0.489	0.772	0.655	0.669	0.443	0.563	0.873	0.889	0.807	0.805	**0.898**
	JPEG	0.935	*0.960*	0.828	0.919	0.749	0.290	0.585	0.595	0.761	0.675	0.797	0.861	0.726	**0.876**	0.862
	Gblur	0.600	0.758	0.924	*0.932*	0.037	0.272	0.562	0.876	0.879	0.430	0.451	0.935	0.945	0.932	**0.959**
WIVC- II	WN	0.783	0.797	0.799	*0.893*	0.560	0.721	0.700	0.641	0.522	0.630	**0.903**	0.834	0.797	0.839	0.852
	JPEG	0.859	*0.903*	0.727	0.875	0.138	0.111	0.649	0.543	0.571	0.449	0.596	0.680	0.651	0.733	**0.822**
	Gblur	0.312	0.669	*0.912*	0.823	0.228	0.137	0.591	0.854	0.920	0.619	0.717	0.936	0.883	0.880	**0.937**
NBU 3D	WN	*0.925*	0.860	0.912	0.828	0.244	0.764	0.938	0.802	0.765	0.720	0.811	0.879	**0.906**	0.614	0.848
	JP2K	0.874	*0.947*	0.940	0.910	0.569	0.508	0.865	0.850	0.808	0.871	0.839	0.857	0.855	0.121	**0.886**
	JPEG	0.904	*0.939*	0.934	0.925	0.466	0.519	0.917	0.905	0.896	0.860	0.881	0.882	0.894	0.695	**0.951**
	Gblur	0.930	0.953	0.939	*0.966*	0.444	0.483	0.858	0.839	0.926	0.946	0.856	0.916	**0.951**	0.750	0.946
IVC 3D	JP2K	0.823	0.885	0.802	*0.958*	0.641	0.452	0.749	0.619	0.252	0.224	0.746	0.748	0.743	0.249	**0.877**
	JPEG	0.775	0.782	0.897	*0.938*	0.420	0.510	0.862	0.665	0.637	0.470	0.754	0.714	0.689	0.814	**0.956**
	Gblur	0.459	*0.467*	0.321	0.143	0.657	0.243	0.725	0.466	0.186	0.307	0.693	0.143	0.835	0.777	**0.847**

DMOS = -2.09; S_L = 0.3310, S_R = 0.3246, S_{2D} = 0.3277, S_{cyc} = 0.4422, S_{3D} = 0.3807

DMOS = 19.28; S_L = 0.8956, S_R = 0.9131, S_{2D} = 0.9044, S_{cyc} = 0.7889, S_{3D} = 0.8447

DMOS = 28.31; S_L = 0.9398, S_R = 0.9416, S_{2D} = 0.9407, S_{cyc} = 0.8519, S_{3D} = 0.8952

Fig. 7. Quality scores given by MUSIQUE-3D to sample distorted stereoscopic images from the LIVE 3D Phase I database.

associated models. Although the quality-compensated multipathway contrast gain-control model presented in this paper is effective, it is limited to four distortion types and their arbitrary combinations. Also, the two monocular view images considered in this work are assumed to be corrupted by the same distortion types/combinations, while in practice other cases can occur. In the future, more advanced cyclopean frameworks that can handle more/different distortion scenarios could be developed, thus giving rise to improved future SIQA algorithms.

CRediT authorship contribution statement

Yi Zhang: Conceptualization, Methodology, Software, Investigation, Writing - original draft. **Damon M. Chandler:** Software, Data

Table 9

PLCC, SROCC, and KROCC values tested by including a denoising operation, as well as using different optical flow algorithms and classification/regression models in the MUSIQUE-3D algorithm. Note that S_{2D}, S_{cyc}, and S_{3D} represent correlation coefficient values computed for quality scores given by the first, second, and final stage of the algorithm, respectively.

	Datasets	C-BM3D			DnCNN			SLK			BP			MUSIQUE-3D		
		S_{2D}	S_{cyc}	S_{3D}	S_{2D}	S_{cyc}	S_{3D}	S_{2D}	S_{cyc}	S_{3D}	S_{2D}	S_{cyc}	S_{3D}	S_{2D}	S_{cyc}	S_{3D}
PLCC	NBUMD-I	0.947	0.928	0.948	0.947	0.929	0.947	0.947	0.706	0.955	0.940	0.927	0.945	0.947	0.916	0.948
	NBUMD-II	0.897	0.830	0.889	0.897	0.822	0.881	0.897	0.815	0.916	0.886	0.816	0.870	0.897	0.837	0.900
	LIVE-I	0.916	0.906	0.923	0.916	0.903	0.922	0.916	0.878	0.917	0.902	0.852	0.888	0.916	0.908	0.924
	LIVE-II	0.891	0.890	0.916	0.891	0.891	0.917	0.892	0.866	0.901	0.852	0.780	0.840	0.891	0.880	0.911
	WIVC-I	0.909	0.911	0.930	0.909	0.903	0.926	0.910	0.898	0.928	0.854	0.888	0.904	0.909	0.911	0.931
	WIVC-II	0.876	0.889	0.905	0.876	0.876	0.900	0.878	0.896	0.912	0.833	0.855	0.859	0.876	0.894	0.908
	NBU 3D	0.886	0.864	0.884	0.886	0.871	0.889	0.886	0.736	0.826	0.862	0.824	0.855	0.886	0.861	0.884
	IVC 3D	0.842	0.814	0.884	0.842	0.809	0.881	0.842	0.712	0.846	0.842	0.746	0.803	0.842	0.803	0.882
	Average	0.900	0.885	0.912	0.900	0.881	0.910	0.900	0.826	0.906	0.873	0.844	0.876	0.900	0.883	0.914
SROCC	NBUMD-I	0.930	0.917	0.933	0.930	0.909	0.934	0.930	0.862	0.938	0.926	0.923	0.934	0.930	0.904	0.934
	NBUMD-II	0.874	0.824	0.869	0.874	0.798	0.855	0.874	0.761	0.888	0.857	0.812	0.852	0.874	0.831	0.884
	LIVE-I	0.888	0.870	0.891	0.888	0.867	0.889	0.888	0.820	0.882	0.863	0.794	0.842	0.888	0.871	0.890
	LIVE-II	0.887	0.875	0.907	0.887	0.877	0.906	0.888	0.855	0.895	0.845	0.779	0.838	0.887	0.865	0.903
	WIVC-I	0.898	0.879	0.917	0.897	0.865	0.906	0.901	0.871	0.914	0.781	0.860	0.878	0.897	0.884	0.918
	WIVC-II	0.849	0.851	0.878	0.849	0.829	0.868	0.852	0.886	0.897	0.760	0.815	0.818	0.849	0.862	0.885
	NBU 3D	0.884	0.864	0.887	0.884	0.871	0.890	0.884	0.736	0.828	0.862	0.830	0.855	0.884	0.862	0.885
	IVC 3D	0.844	0.809	0.878	0.844	0.800	0.877	0.845	0.688	0.836	0.843	0.743	0.795	0.844	0.798	0.877
	Average	0.884	0.865	0.895	0.884	0.855	0.891	0.885	0.824	0.890	0.840	0.825	0.855	0.884	0.864	0.898
KROCC	NBUMD-I	0.766	0.738	0.768	0.766	0.731	0.773	0.766	0.655	0.780	0.753	0.749	0.769	0.766	0.717	0.770
	NBUMD-II	0.691	0.627	0.686	0.691	0.606	0.675	0.693	0.566	0.710	0.678	0.612	0.667	0.692	0.636	0.704
	LIVE-I	0.703	0.681	0.706	0.703	0.679	0.705	0.703	0.630	0.698	0.668	0.591	0.647	0.703	0.684	0.707
	LIVE-II	0.702	0.681	0.726	0.702	0.682	0.725	0.705	0.652	0.705	0.646	0.580	0.642	0.702	0.667	0.720
	WIVC-I	0.724	0.707	0.763	0.723	0.692	0.749	0.727	0.697	0.757	0.600	0.674	0.703	0.723	0.711	0.764
	WIVC-II	0.663	0.670	0.699	0.662	0.647	0.687	0.667	0.706	0.723	0.572	0.623	0.627	0.662	0.681	0.707
	NBU 3D	0.702	0.683	0.707	0.702	0.690	0.711	0.702	0.551	0.641	0.676	0.644	0.671	0.702	0.677	0.704
	IVC 3D	0.639	0.623	0.691	0.639	0.617	0.690	0.640	0.524	0.642	0.657	0.533	0.593	0.639	0.608	0.694
	Average	0.702	0.680	0.719	0.702	0.670	0.715	0.704	0.634	0.714	0.654	0.632	0.669	0.702	0.678	0.723

curation, Visualization, Formal analysis, Writing - review & editing. **Xuanqin Mou:** Validation, Resources, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] W. Lin, C.-C.J. Kuo, Perceptual visual quality metrics: A survey, J. Vis. Commun. Image Represent. 22 (2011) 297–312.

[2] D.M. Chandler, Seven challenges in image quality assessment: past, present, and future research, ISRN Signal Process. 2013 (2013) 905685, 53 pages.

[3] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, Y.-S. Ho, Unified no-reference quality assessment of singly and multiply distorted stereoscopic images, IEEE Trans. Image Process. 28 (2019) 1866–1881.

[4] F. Shao, W. Tian, W.-S. Lin, G.-Y. Jiang, Q.-H. Dai, Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions, IEEE Trans. Image Process. 25 (2016) 2059–2074.

[5] P. Seuntiens, L. Meesters, W. Ijsselsteijn, Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric jpeg coding and camera separation, ACM Trans. Appl. Percept. (TAP) 3 (2006) 95–109.

[6] M.-J. Chen, L.K. Cormack, A.C. Bovik, No-reference quality assessment of natural stereopairs, IEEE Trans. Image Process. 22 (2013) 3379–3391.

[7] C.-C. Su, L.K. Cormack, A.C. Bovik, Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation, IEEE Trans. Image Process. 24 (2015) 1685–1699.

[8] J. Yang, Y. Zhao, Y. Zhu, H. Xu, W. Lu, Q. Meng, Blind assessment for stereo images considering binocular characteristics and deep perception map based on deep belief network, Inform. Sci. 474 (2019) 1–17.

[9] X. Xu, Y. Zhao, Y. Ding, No-reference stereoscopic image quality assessment based on saliency-guided binocular feature consolidation, Electron. Lett. 53 (2017) 1468–1470.

[10] L. Liu, B. Liu, C.-C. Su, H. Huang, A.C. Bovik, Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment, Signal Process., Image Commun. 58 (2017) 287–299.

[11] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, T. Luo, Blind quality estimator for 3D images based on binocular combination and extreme learning machine, Pattern Recognit. 71 (2017) 207–217.

[12] W. Zhang, C. Qu, L. Ma, J. Guan, R. Huang, Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network, Pattern Recognit. 59 (2016) 176–187.

[13] H. Oh, S. Ahn, J. Kim, S. Lee, Blind deep S3D image quality evaluation via local to global feature aggregation, IEEE Trans. Image Process. 26 (2017) 4923–4936.

[14] Q. Jiang, F. Shao, W. Lin, G. Jiang, Learning a referenceless stereopair quality engine with deep nonnegativity constrained sparse autoencoder, Pattern Recognit. 76 (2018) 242–255.

[15] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (2012) 4695–4708.

[16] W.-J. Zhou, L. Yu, W.-W. Qiu, T. Luo, Z.-P. Wang, M.-W. Wu, Utilizing binocular vision to facilitate completely blind 3D image quality measurement, Signal Process. (2016).

[17] F. Shao, W. Lin, S. Wang, G. Jiang, M. Yu, Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook, IEEE Trans. Image Process. 61 (2015) 154–165.

[18] F. Shao, W. Lin, S. Wang, G. Jiang, M. Yu, Q. Dai, Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images, IEEE Trans. Cybern. 46 (2016) 730–743.

[19] F. Shao, Z. Zhang, Q. Jiang, W. Lin, G. Jiang, Toward domain transfer for no-reference quality prediction of asymmetrically distorted stereoscopic images, IEEE Trans. Circuits Syst. Video Technol. 28 (2018) 573–585.

[20] F. Shao, W. Tian, W. Lin, G. Jiang, Q. Dai, Learning sparse representation for blind quality assessment of multiply distorted stereoscopic images, IEEE Trans. Multimed. 19 (2017) 1821–1836.

[21] F. Shao, Y. Gao, Q. Jiang, G. Jiang, Y. Ho, Multistage pooling for blind quality prediction of asymmetric multiply-distorted stereoscopic images, IEEE Trans. Multimed. 20 (2018) 2605–2619.

[22] M.-J. Chen, C.-C. Su, D.-K. Kwon, L.K. Cormack, A.C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process., Image Commun. 28 (2013) 1143–1155.

[23] Y. Zhang, D.M. Chandler, Opinion-unaware blind quality assessment of multiply and singly distorted images via distortion parameter estimation, IEEE Trans. Image Process. 27 (2018) 5433–5448.

[24] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: IEEE International Conference on Computer Vision (ICCV), Vol. 2, 2001, pp. 416–423.

[25] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, P. Westling, High-resolution stereo datasets with subpixel-accurate ground truth, in: German Conference on Pattern Recognition, Springer, 2014, pp. 31–42.

[26] P.V. Vu, D.M. Chandler, A fast wavelet-based algorithm for global and local image sharpness estimation, IEEE Signal Process. Lett. 19 (2012) 423–426.

[27] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2006) 430–444.

[28] C. Liu, Beyond Pixels: Exploring New Representations and Applications for Motion Analysis (Ph.D. thesis), Massachusetts Institute of Technology, 2009.

[29] Y. Zhang, D.M. Chandler, 3D-MAD: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception, IEEE Trans. Image Process. 24 (2015) 3810–3825.

[30] W. Sun, F. Zhou, Q. Liao, MDID: A multiply distorted image database for image quality assessment, Pattern Recognit. 61 (2017) 153–168.

[31] S. Knerr, L. Personnaz, G. Dreyfus, Single-layer learning revisited: a stepwise procedure for building and training a neural network, in: Neurocomputing, Springer, 1990, pp. 41–50.

[32] H.-T. Lin, C.-J. Lin, R.C. Weng, A note on platt's probabilistic outputs for support vector machines, Mach. Learn. 68 (2007) 267–276.

[33] T.-F. Wu, C.-J. Lin, R.C. Weng, Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res. 5 (2004) 975–1005.

[34] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.

[35] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (2011) 27.

[36] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Statist. Comput. 14 (2004) 199–222.

[37] A.K. Moorthy, C.-C. Su, A. Mittal, A.C. Bovik, Subjective evaluation of stereoscopic image quality, Signal Process., Image Commun. 28 (2012) 870–873.

[38] J. Wang, A. Rehman, K. Zeng, S. Wang, Z. Wang, Quality prediction of asymmetrically distorted stereoscopic 3D images, IEEE Trans. Image Process. 24 (2015) 3400–3414.

[39] F. Shao, W. Lin, S. Gu, G. Jiang, T. Srikanthan, Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics, IEEE Trans. Image Process. 22 (2013) 1940–1953.

[40] A. Benoit, P.L. Callet, P. Campisi, R. Cousseau, Quality assessment of stereoscopic images, EURASIP J. Image Video Process. 2008 (2009) 1–13.

[41] Y.-H. Lin, J.-L. Wu, Quality assessment of stereoscopic 3D image compression by binocular integration behaviors, IEEE Trans. Image Process. 23 (2014) 1527–1542.

[42] Z. Chen, J. Xu, C. Lin, W. Zhou, Stereoscopic omnidirectional image quality assessment based on predictive coding theory, IEEE J. Sel. Top. Sign. Proces. 14 (2020) 103–117.

[43] Y. Fang, J. Yan, X. Liu, J. Wang, Stereoscopic image quality assessment by deep convolutional neural network, J. Vis. Commun. Image Represent. 58 (2019) 400–406.

[44] W. Xue, X. Mou, L. Zhang, A.C. Bovik, X. Feng, Blind image quality assessment using joint statistics of gradient magnitude and laplacian features, IEEE Trans. Image Process. 23 (2014) 4850–4862.

[45] Q. Li, W. Lin, Y. Fang, No-reference quality assessment for multiply-distorted images in gradient domain, IEEE Signal Process. Lett. 23 (2016) 541–545.

[46] W. Xue, L. Zhang, X. Mou, Learning without human scores for blind image quality assessment, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 995–1002.

[47] A. Mittal, R. Soundararajan, A.C. Bovik, Making a completely blind image quality analyzer, IEEE Signal Process. Lett. 20 (2013) 209–212.

[48] L. Zhang, L. Zhang, A.C. Bovik, A feature-enriched completely blind image quality evaluator, IEEE Trans. Image Process. 24 (2015) 2579–2591.

[49] K. Gu, G. Zhai, X. Yang, W. Zhang, Hybrid no-reference quality metric for singly and multiply distorted images, IEEE Trans. Broadcast. 60 (2014) 555–567.

[50] A. Aldahdooh, E. Masala, O. Janssens, G. Van Wallendael, M. Barkowsky, P.L. Callet, Improved performance measures for video quality assessment algorithms using training and validation sets, IEEE Trans. Multimed. 21 (2019) 2026–2041.

[51] L. Krasula, K. Fliegel, P.L. Callet, M. Klima, On the accuracy of objective image and video quality models: New methodology for performance evaluation, 2016, pp. 1–6.

[52] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space, in: IEEE International Conference on Image Processing, Vol. 1, IEEE, 2007, pp. I–313.

[53] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, IEEE Trans. Image Process. 26 (2017) 3142–3155.

[54] S. Smets, T. Goedeme, M. Verhelst, Custom processor design for efficient, yet flexible lucas-kanade optical flow, 2016, pp. 138–145.