3D-MAD: A Full Reference Stereoscopic Image Quality Estimator Based on Binocular Lightness and Contrast Perception

Yi Zhang and Damon M. Chandler, *Senior Member, IEEE*

***Abstract*— Algorithms for a stereoscopic image quality assessment (IQA) aim to estimate the qualities of 3D images in a manner that agrees with human judgments. The modern stereoscopic IQA algorithms often apply 2D IQA algorithms on stereoscopic views, disparity maps, and/or cyclopean images, to yield an overall quality estimate based on the properties of the human visual system. This paper presents an extension of our previous 2D most apparent distortion (MAD) algorithm to a 3D version (3D-MAD) to evaluate 3D image quality. The 3D-MAD operates via two main stages, which estimate perceived quality degradation due to 1) distortion of the monocular views and 2) distortion of the cyclopean view. In the first stage, the conventional MAD algorithm is applied on the two monocular views, and then the combined binocular quality is estimated via a weighted sum of the two estimates, where the weights are determined based on a block-based contrast measure. In the second stage, intermediate maps corresponding to the lightness distance and the pixel-based contrast are generated based on a multipathway contrast gain-control model. Then, the cyclopean view quality is estimated by measuring the statistical-difference-based features obtained from the reference stereopair and the distorted stereopair, respectively. Finally, the estimates obtained from the two stages are combined to yield an overall quality score of the stereoscopic image. Tests on various 3D image quality databases demonstrate that our algorithm significantly improves upon many other state-of-the-art 2D/3D IQA algorithms.**

***Index Terms*— Full reference, stereoscopic image quality, contrast gain-control, binocular combination.**

I. Introduction

A. Background

THE rapid development of digital image processing technology has encouraged tremendous research in the field of 2D image quality assessment (IQA), and noticeable progress has been made on developing various full reference, reduced reference, and no reference 2D IQA algorithms (see [2]–[4] for reviews). Stereoscopic IQA, however, has been relatively less explored, partially due to the difficulty in viewing commercially available 3D content, as well as issues surrounding visual discomfort and fatigue after long-term stereoscopic viewing [5]. However, in recent years, 3D media has increased in popularity and availability, and applications may range from entertainment (e.g., 3D television, 3D cinema, and 3D games) to more specialized purposes (e.g., 3D remote education, 3D robot navigation, and 3D medical imagery). As a result, there is a need for IQA algorithms that can assess the quality of stereoscopic/3D imagery.

A stereoscopic image contains two slightly different views, each of which is projected separately onto the human retina. When a stereoscopic image is viewed by a human subject, the human visual system (HVS) does not examine the left view and right view individually; instead, complex binocular fusion and rivalry processes between the two views operate to yield a merged 3D mental view. This merged mental view depends not only on the disparities of individual stimulus elements, but also on the geometrical relation of different parts of the pattern presented to each eye [6]. Therefore, the quality of a 3D image is affected not only by the degradation level of each individual left and right view, but also by the experience of binocular visual perception. A 3D QA algorithm, however, cannot access this merged single view. Thus, assessing the perceived quality of a stereoscopic image can be quite challenging if only the two stereoscopic views are available.

The 3D QA task can be easier when the left and right views are equally distorted (also called *symmetrically* distorted stimuli). One technique is to simply apply 2D QA methods on both stereoscopic views independently and then combine the two scores to yield an overall quality estimate (e.g., [7]–[10]). The other technique is to employ both the stereopairs and the depth/disparity information (e.g., [11], [12]), assuming that the HVS also uses binocular depth/disparity information to judge the 3D image quality. However, two important points need to be considered. First, since the ground truth depth/disparity maps are not always available, these models can operate based only on the estimated maps. Thus, the accuracy of the depth/disparity estimate algorithm employed may substantially affect the performance. Second, the depth/disparity information may have no correlation with the 3D perceived quality. This latter point has been confirmed by Kaptein *et al.* [13] who ran subjective experiments using blurred images with the same objects at different depths; Kaptein *et al.* found that the depth level in a 3D display does not influence

Manuscript received October 3, 2014; revised February 3, 2015, May 13, 2015, and June 26, 2015; accepted June 27, 2015. Date of publication July 14, 2015; date of current version July 30, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dacheng Tao.

Y. Zhang is with the School of Electrical and Computer Engineering, Oklahoma State University–Stillwater, Stillwater, OK 74078 USA (e-mail: yi.zhang@okstate.edu).

D. M. Chandler is with the School of Electrical and Computer Engineering, Oklahoma State University–Stillwater, Stillwater, OK 74078 USA, and also with the Department of Electrical and Electronic Engineering, Shizuoka University, Shizuoka 432-8561, Japan (e-mail: chandler.damon.michael@shizuoka.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2456414

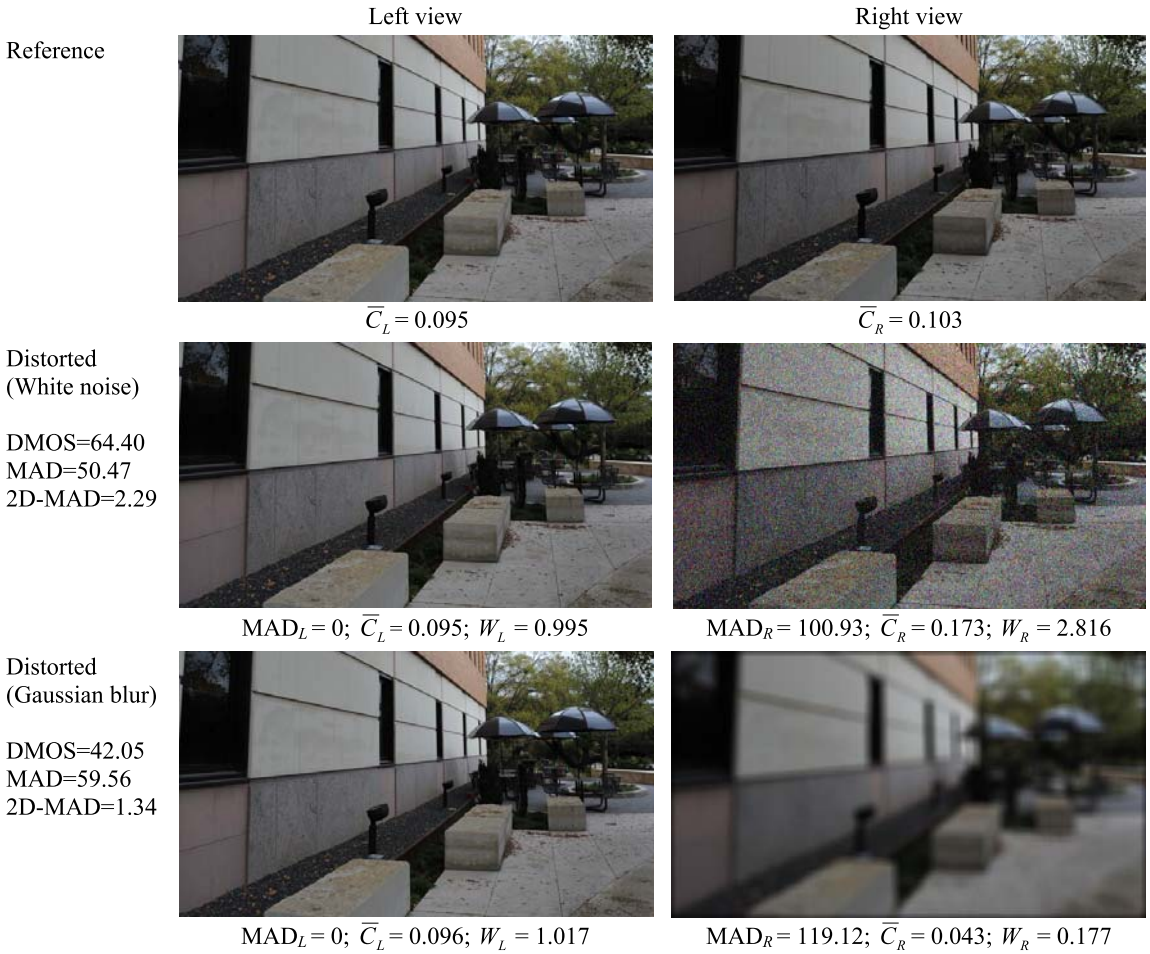


Fig. 1. Two demonstrative asymmetrically distorted stereoscopic images from the LIVE 3D image database [28] that have high quality on the left view and low quality on the right view. Note that $\bar{C}_L$ and $\bar{C}_R$ represent the averaged block-based contrast value for the left and right views, respectively; W_L and W_R denote their corresponding weights computed through Eqs. (4) and (5). "MAD" denotes a quality score computed by averaging MAD_L and MAD_R. "2D-MAD" is computed through Eq. (6).

the perceived image quality. Also in [14], Goldmann and Ebrahimi have demonstrated a similar shortcoming when applying 2D QA algorithms for 3D QA.

Despite these difficulties, the two aforementioned techniques can still achieve fairly good quality predictions of symmetrically distorted stereoscopic images. However, if the two stereoscopic views contain different amounts and/or types of distortions (also called *asymmetrically* distorted stimuli), these algorithms perform much less effectively. Asymmetrically distorted stimuli can make the stereoscopic IQA problem much more challenging primarily because the merged mental view changes depending on the distributions of the distortion. To demonstrate, Figure 1 shows two distorted stereoscopic images that contain similar high-quality images for the left view, and low-quality images for the right view. As reported by subjects, for the white noise image, they perceive a merged view that looks more like the right-view image, whereas for the Gaussian blurred image, the perceived merged view looks more like the left-view image. Thus, effectively modeling the HVS's binocular perceptual mechanisms when asymmetrically distorted views are presented to the two eyes is essential to enhance the performance of a 3D QA algorithm.

According to [15], binocular vision is a complex visual process that involves both binocular fusion and binocular rivalry that coexist at the same point in space and time. Binocular rivalry inevitably occurs whenever visual stimuli are presented, even if the two stimuli are identical. In regards to the precedence of the two processes, Blake and Boothroyd [16] stated that

> "The binocular visual system first seeks to establish correspondence between image features contained in the two monocular views. Failure to establish such correspondence leads to binocular suppression, an effective means for eliminating diplopia and/or confusion. When correspondence is established, however, binocular fusion takes precedence over the suppression. Only those monocular features with no interocular counterpart participate in the rivalry process [16]."

To study these two properties of binocular vision, other vision studies (e.g., [17]–[27]) have used unnatural stimuli (or simulated data) with highly controllable parameters, and their results can indeed offer some basic insights. However, it remains unclear how the results of such controlled studies

should be used for 3D QA because the findings have never tested on real natural images.

For QA of stereoscopic images that contain asymmetrically distorted stimuli, a recent technique has been employed, which attempts to model the 3D QA behavior of the HVS based on binocular fusion and rivalry properties. Approaches based on this technique often apply 2D QA methods on the so-called *cyclopean view*, a single mental image of a scene created in the brain by combining the two images received from the two eyes. These approaches assume that the HVS perceives 2D degradation after they are combined in the cyclopean view, and thus it is meaningful to measure 2D artifacts on the cyclopean view as an estimate of the 3D quality degradation level. In [29], the cyclopean image was defined as the average of the left image and the disparity-compensated right image. More recently in [30], the cyclopean image was synthesized by using a linear model proposed in [31], in which the coefficients of the linear model are computed based on the local Gabor filter energy. Although a better prediction performance was observed compared with the two aforementioned techniques, which partially demonstrates the effectiveness of the cyclopean 3D QA model, the performance on symmetrically distorted stereopairs is actually not as high as expected (see Table III). Also, issues concerning how the cyclopean image is formulated in the brain, and how the HVS judges quality based on this cyclopean image, remain to be further explored.

Beyond these aforementioned techniques, no strong baseline or core principle has been reported to point out the nature of the stereoscopic IQA problem, and these techniques are only roughly distinguished from each other based on the type and/or the amount of information they extract from 3D images. In fact, some recently developed 2D IQA algorithms can achieve impressive predictive performance on most stereoscopic image databases that contain symmetrically distorted stimuli when these algorithms are only applied on the two stereoscopic views and then averaged (see Table III). Thus, we argue that a core principle of 3D QA algorithm design should be to improve the algorithm performance on both symmetrically and asymmetrically distorted stereoscopic images. Under this principle, we further argue that a grand challenge confronted by current 3D QA research is to more accurately model the QA behavior of the HVS in binocular vision especially when asymmetrically distorted stimuli are presented, which is also the most fundamental driving force of this work. Unfortunately, as noted by Moorthy *et al.* in [28], no current 3D QA algorithms has achieved prediction performance better than (or in most cases, even as good as) 2D QA algorithms applied to the individual stereopairs.

B. Motivation

Based on the aforementioned points, the main motivation of our proposed approach is to develop a technique that more effectively models the binocular rivalry and fusion behaviors of the HVS, which occur during QA of stereoscopic images. Towards this end, we adopt in our work two main strategies: (1) a contrast-weighting strategy, and (2) a cyclopean-feature-image strategy.

The first strategy, motivated by both the theoretical analysis in [31]–[34] and the practical 3D-viewing experiences reported by subjects, aims at balancing the roles of each stereoscopic view in determining the overall 3D image quality. According to [31]–[34], the monocular stimulus with higher contrast often receives enhanced predominance and shorter duration of suppression in human visual perception. If the contrasts of the stimuli presented to both eyes are equally high, then each stimulus remains suppressed for shorter periods of time and, consequently, the two stimuli alternate in dominance more rapidly. Thus, we believe that the quality of a monocular scene with higher contrast can play a more dominant role in determining the HVS's judgment of 3D image quality.

As demonstrated in Figure 1, the two distorted stereoscopic images contain similar quality degradation levels in each of the two corresponding views. However, the differential mean opinion scores (DMOS values) indicate that the stereoscopic image containing the Gaussian blurred right view has better perceived quality than the stereoscopic image containing the white noise in the right view. Please note that, in general, the relative perceived quality between an image containing Gaussian blurring vs. white noise will also depend on the degree of distortion in each image. The reason for these subjective ratings is that, for the white noise image, the HVS judges quality based mainly on the right view which has higher contrast because of the noise; while for the Gaussian blurred image, the left view plays a more important role, since the right view has relatively lower contrast due to the blurring. Thus, the proposed method of combining the quality estimates of both views via contrast-based weighting can predict the qualities of these two images quite well, whereas the conventional approach that only averages the two quality scores cannot.

The second strategy, motivated by the recent multi-pathway contrast gain-control model (MCM) [35], [36] and the cyclopean-image-based 3D QA approach proposed in [29] and [30], is employed to more effectively model the binocular fusion and binocular rivalry properties of the HVS especially for asymmetrically distorted stimuli. As mentioned in Section I-A, the HVS judges stereoscopic image quality based on a merged 3D mental view; this fact makes the 3D QA task much more challenging for an algorithm. The *cyclopean image* can be used as a replacement of this merged view; however, the complex QA behavior of the HVS in binocular vision makes the simple linear model [30] less effective in cyclopean image computation. Thus, we employ the MCM to account for the three interocular contrast gain-control mechanisms to better model the binocular perception of image quality in the HVS.

To incorporate the MCM in 3D QA work, we propose a feature-based MCM [see Eq. (12)]. Specifically, we propose two different types of quality-related features (lightness distance and pixel-based contrast) and build three types of corresponding cyclopean feature images[1] (cyclopean images of the global/local lightness distance and pixel-based contrast)

[1]Here, we use the term "cyclopean feature image" to distinguish from the term "cyclopean image" that deals only with the image luminance.

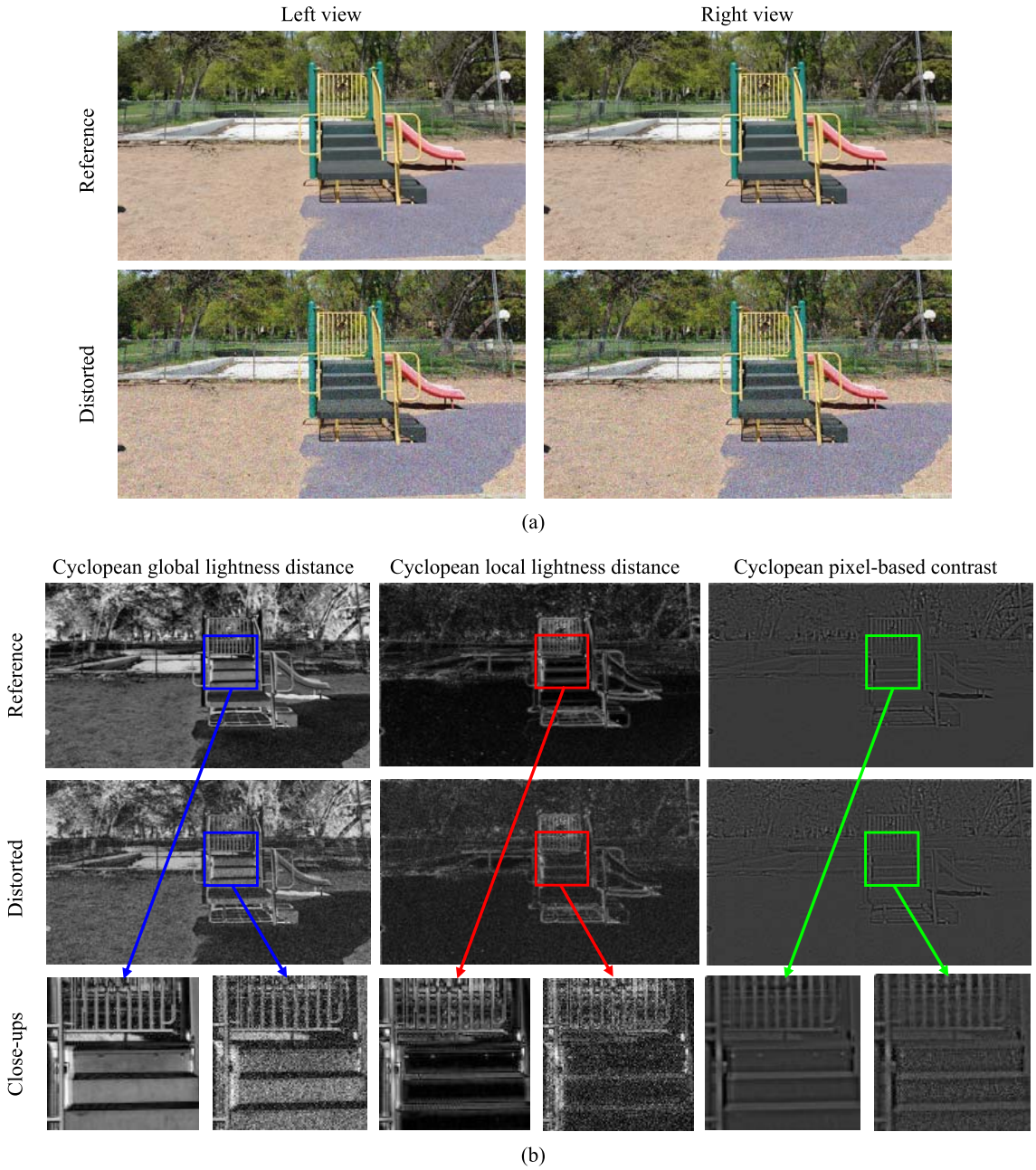


Fig. 2. Demonstrative cyclopean feature images corresponding to the lightness distance and pixel-based contrast (b) computed from the reference and distorted stereopairs (a) in the the LIVE 3D image database [28]. The close-ups show some dissimilar regions between the cyclopean feature images.

by using the MCM, which is also improved by incorporating new parameters to adaptively indicate the different dominant roles of the two eyes in 3D vision. We then quantify distortions contained in these computed cyclopean feature images as a measurement of the 3D image quality by using the statistical-difference-based features, which are combined via support vector machine learning. As shown in Figure 2, the three cyclopean feature images [Figure 2(b)] computed from the reference and distorted stereopairs [Figure 2(a)] exhibit considerable dissimilarities between the reference and distorted stereoscopic images, and some dissimilar regions are clearly visible in the close-ups.

There are two reasons to employ these features [see Eqs. (7)-(9)] instead of the absolute luminance values to estimate image quality. First, the MCM provides only a model for how stimulus contrast is combined in binocular vision; thus, it cannot be used directly with a (luminance) image. Also, it is important to note that the MCM was designed only for modeling unnatural stimuli with highly controllable parameters, and not for natural images. Thus, to extend the MCM to work for 3D QA for natural images, we utilize the lightness distance and pixel-based contrast feature maps as various representations of the perceived contrast.

Second, it is evident that no current models can fully model binocular vision. Hence, it is more reasonable to use multiple features to comprehensively measure the distortions, just as the fact that human use multiple cues to judge image quality. This has been demonstrated in Table I, in which we compared

TABLE I
PERFORMANCE COMPARISON OF USING FEATURES VS. ABSOLUTE LUMINANCE TO PREDICT STEREOSCOPIC IMAGE QUALITY. NOTE THAT FOR BOTH OPTIONS, WE USE THE SAME MCM PARAMETERS TO COMPUTE THE CYCLOPEAN IMAGES (DENOTED BY "CYC-LUMINANCE") AND CYCLOPEAN FEATURE IMAGES (DENOTED BY "CYC-MAD"), AND THE SAME 2D LIVE IMAGE DATABASE [37] FOR MODEL TRAINING

	LIVE3D Phase I		LIVE3D Phase II	
Approach	CC	SROCC	CC	SROCC
Cyc-MAD	0.944	0.937	0.921	0.915
Cyc-luminance	0.944	0.937	0.894	0.885

the quality predictive performance of using the proposed three features vs. the absolute luminance. (In Table I, we used the same MCM parameters to compute the cyclopean image and the cyclopean feature images, and we also used the same 2D LIVE image database [37] for model training. For a fair comparison, we do not consider the 2D-MAD quality estimate on stereopairs, and only use the quality degradation measurements of the cyclopean images/feature images as the overall quality estimate of the stereoscopic image.) As shown in Table I, the two options can achieve almost equal performance when testing on the LIVE 3D Phase I database [28] which contains symmetrically distorted stimuli. However, when testing on the LIVE 3D Phase II database [28], which also contains asymmetrically distorted stimuli, the proposed strategy that employs the three quality-related features can achieve much better performance than that which employs the luminance only. This result also demonstrates the effectiveness of the proposed features and models for QA of asymmetrically distorted stereoscopic images.

C. Proposal and Contributions

Motivated by the points in Section I-B, in this paper, we extend our previous 2D MAD algorithm [1] to address the 3D IQA problem via two main stages corresponding to the two proposed strategies. The first stage estimates perceived quality degradation due to 2D monocular scene artifacts. This stage operates by first directly applying the MAD algorithm on two separated stereoscopic views (left and right views) of the reference and distorted images, respectively. Then, the left- and right-image quality scores are linearly combined using weights that are computed as the average of the normalized block-based contrast value of each stereopair. The second stage of the algorithm estimates perceived quality degradation due to cyclopean view artifacts. This stage operates by measuring local statistical differences between the reference and distorted cyclopean lightness-distance and pixel-based-contrast feature images. All of these cyclopean feature images are disparity-compensated and are synthesized by using stereopairs following a multipathway contrast gain-control model. Finally, the quality measurements obtained from the two stages are combined to yield an overall estimate of the perceived 3D quality degradation. As we will demonstrate, these two stages together allow 3D-MAD to achieve good predictive performance on both symmetrically and asymmetrically distorted stereoscopic images across various databases.

The main contributions of this work are as follows. First, we tackle the aforementioned challenge confronted by current 3D QA research by presenting two strategies to enhance the quality predictive performance on asymmetrically distorted stereoscopic images: the contrast-weighting strategy and the cyclopean-feature-image strategy, both of which allow the proposed method to take into account different contributions of the left and right views towards the overall 3D quality. As we will demonstrate, the former strategy improves upon the conventional averaged-approach [29] (when comparing 2D-MAD results with MAD), and the latter strategy improves upon the previous linear cyclopean image model [31] (when comparing Cyc-MAD results with Chen's method [30]), and the combined strategies allows 3D-MAD to achieve better performance on most 3D image databases. Also, the training process employed in Section III-B3 has demonstrated a way to predict 3D image quality by referring to the distortion information of 2D images. This contribution suggests that if appropriate features and models are utilized, the 2D image and 3D image can have some properties in common in the feature space. Finally, to the best of our knowledge and based on the test databases used in our work, 3D-MAD is currently the first 3D full reference (FR) IQA algorithm that can achieve competitive or even better performance than 2D FR IQA methods applied on stereopairs.

This paper is organized as follows: Section II provides a brief review of current 3D QA algorithms. In Section III, we describe details of the proposed 3D-MAD algorithm. In Section IV, we analyze and discuss the performance of the proposed algorithm on various 3D image quality databases. General conclusions are presented in Section V.

II. PREVIOUS WORK

In this section, we provide a brief review of current 3D QA algorithms. As mentioned in Section I, a stereoscopic image may contain either symmetrically or asymmetrically distorted stimuli. The complex QA behavior of the HVS in viewing symmetrically and asymmetrically distorted stereopairs has given rise to a variety of simple or sophisticated 3D QA techniques. Based on the type and/or the amount of information extracted from stereoscopic images, these techniques approximately fall into one of the three categories: (1) those that apply 2D QA methods separately on each view of the stereoscopic image, and then combine the two scores into an overall quality score; (2) those that take into account the depth information and apply 2D QA methods on both stereopairs and the estimated disparity map; and (3) those that take into account the binocular visual properties of HVS and apply 2D QA methods on the cyclopean images and/or stereopairs.

A. Stereopair-Based IQA

As stated in Section I, the most straightforward way to estimate the quality of a 3D image is to directly apply 2D QA algorithms on each individual view, and then combine the scores into one quality measurement.

For instance, Campisi *et al.* [10] applied four IQA algorithms (SSIM [38], UQI [39], C4 [40], and RRIQA [41]) on stereopairs and combined quality scores via three approaches: "average" approach, "main eye" approach, and "visual acuity" approach. Gorley and Holliman [9] proposed a "Stereo Band Limited Contrast" metric to estimate 3D image quality based on matched points of stereopairs delivered by SIFT [42] and RANSAC [43]. Yang *et al.* [44] estimated 3D image quality based on the average PSNR of stereopairs, and also based on the absolute difference in pixel values between the left and right view. Also, in [7] and [8], three widely accepted 2D QA algorithms (PSNR [45], SSIM [38], and VQM [46]) were investigated on their correlation with perceived quality of 3D video.

B. Algorithms Based on Stereopair and Depth Information

Another technique commonly used in 3D QA is to apply 2D QA methods on estimated disparity maps and to possibly combine this information with 2D QA estimates of the stereopairs. For instance, Benoit *et al.* [12] applied 2D QA methods (SSIM [38], C4 [40]) on the left and right views independently, and the 3D image quality was computed by combining the two quality scores with an estimate of disparity map distortion. You *et al.* [11] investigated the capabilities of some 2D IQA algorithms for stereoscopic IQA and found that applying SSIM on stereopairs combined with a mean absolute difference to compute the disparity map distortion always yields the best performance within all possible combinations considered in the paper. Hwang and Wu [47] designed a visual attention and depth-assisted stereo image quality model, which consists of three main components: a stereo attention predictor, a depth variation predictor, and a stereo distortion predictor. Xing *et al.* [48] proposed a perceptual quality estimator which operates based on three main factors (crosstalk level, camera baseline, and scene content) contributing to crosstalk perception in evaluating quality levels of stereoscopic presentations; the quality score of this approach is computed as the average of the SSIM-weighted disparity map.

There are also some researchers who have employed stereo depth information for stereo video quality assessment (3D VQA). Although these works address video quality assessment, they still demonstrate a correlation between depth information and visual quality. For example, Ha and Kim [49] used the disparity information (estimated via a horizontal block matching algorithm [50]) to evaluate stereoscopic video quality by taking into account four factors: temporal variance, disparity variation in intra-frames, disparity variation in inter-frames, and the disparity distribution of frame boundary areas. Tikanmaki *et al.* [51] proposed a 3D VQA algorithm by applying PSNR [45] and VSSIM [52] on color and depth sequences. Hewage *et al.* [8] proposed a reduced-reference quality assessment approach for 3D depth map transmission using extracted edge information and PSNR [45] to assess video quality. Boev *et al.* [53] proposed a compound FR 3D VQA algorithm composed of a monoscopic quality component and stereoscopic quality component. The former assesses the monoscopic perceived distortions and the latter assesses the perceived degradation of the perceptual disparity maps as well as of the cyclopean images.

C. Algorithms Based on Binocular Fusion and Rivalry

A third type of 3D QA technique has been recently developed, which incorporates information regarding various binocular artifacts, visual masking effects, and monocular visual properties, all of which may affect the quality of a 3D image. In [54], Ko *et al.* introduced a binocular perception model called the structural distortion parameter (SDP) to control the three components of SSIM to obtain an overall quality index. Similar work can be seen in [55]. In [56], Wang *et al.* presented a 3D IQA algorithm based on binocular spatial sensitivity (BSS), which considers binocular fusion and rivalry properties. The BSS module is implemented in corresponding and non-corresponding regions according to the disparity map generated by disparity matching. In [57], Qi *et al.* suggested that binocular vision is sensitive to the structure of low-level features and the discrepancy between the two views of the stereoscopic image pair. Thus, the phase congruency and saliency maps are employed to compose a feature map, and then a local matching function is employed to weight the discrepancy between the two feature maps to estimate image quality.

Other 3D QA algorithms combine two views into one cyclopean view and the quality of the stereoscopic image is determined by analyzing this integrated image. For example, Bensalma and Larabi [58] developed a "Binocular Energy Quality Metric," which first simulates the binocular visual signal by modeling simple and complex cells; the quality is then estimated based on the difference of the associated binocular energy. Maalouf and Larabi [29] applied a multi-spectral wavelet decomposition to the cyclopean image, and then used the contrast sensitivity coefficients of the cyclopean image, together with disparity maps, for quality assessment. Chen *et al.* [30] applied MS-SSIM [59] on the cyclopean image which was generated by using a linear model, where the model weights are computed from the normalized Gabor filter magnitude responses. Lin and Wu [60] utilized the binocular combination and the binocular frequency integration as the bases for measuring the quality of a stereoscopic image. Shao *et al.* [61] measured the binocular energy-response similarity between the original and distorted 3D image by using components of binocular energy and binocular just noticeable difference (BJND), and then the two evaluation results are nonlinearly combined to estimate the 3D image quality. Also, in [62], Shao *et al.* classified the stereoscopic images into non-corresponding, binocular fusion, and binocular suppression regions; these regions are then evaluated separately via the BJND to estimate the quality.

In the following sections, we describe our dual-strategy-based 3D IQA algorithm, 3D-MAD, which employs adaptive left/right weighting, cyclopean feature images, and machine learning to estimate stereoscopic image quality.

III. Algorithm

The proposed 3D-MAD algorithm is based on the assumption that the overall quality of a 3D scene is

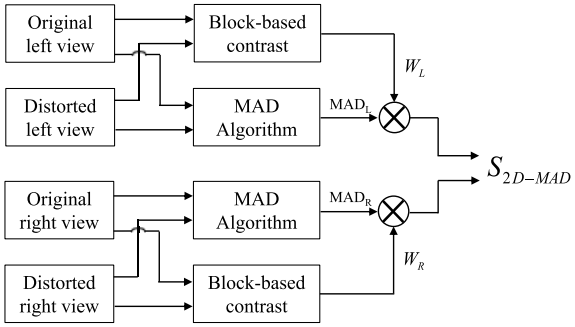


Fig. 3. Block diagram of the first stage of our algorithm, which applies 2D MAD on stereopairs.

a combination of 2D monocular scene quality and 3D binocular-based cyclopean scene quality. Thus, 3D-MAD operates via two main stages: (1) 2D-MAD-based quality estimate on stereopairs, and (2) MCM-based quality estimate on cyclopean feature images. Then, the two quality estimates are combined through a geometric mean to yield a single value that represents the overall perceived quality degradation of the stereoscopic image. The following subsections provide details for each stage.

A. 2D-MAD on Stereopairs

In the 2D-MAD-based QA stage, the conventional MAD algorithm [1] is applied on the stereopairs (i.e., the left and right view images) to estimate the perceived distortion corresponding to each monocular view, but with different trained parameters. Then, the overall 2D-MAD quality is computed as a weighted sum of both stereopair distortion measures, where the weights are computed based on the normalized block-based contrast. The block diagram of this stage is shown in Figure 3.

The MAD algorithm consists of two separate strategies: (1) a detection-based strategy, which computes the perceived distortion due to visual detection (denoted by d_{detect}) and (2) an appearance-based strategy, which computes the perceived distortion due to visual appearance degradation (denoted by d_{appear}). The final MAD quality measurement is a weighted geometric mean given by

$$\text{MAD} = (d_{detect})^{\alpha} \times \left(d_{appear}\right)^{1-\alpha}, \tag{1}$$

where $\alpha \in [0, 1]$ serves to adaptively combine the two strategies based on the overall level of distortion:

$$\alpha = \frac{1}{1 + \beta_1 \times (d_{detect})^{\beta_2}}. \tag{2}$$

In the original MAD algorithm, two parameters ($\beta_1 = 0.467$ and $\beta_2 = 0.130$) and five scale-specific weights ($w_s = 0.5$, 0.75, 1, 5, and 6) were all obtained after training on the A57 image database [63] (see [1] for a complete description of the MAD algorithm). However, the A57 image database contains only 57 images (3 reference images and 54 distorted images), which is a small dataset that might be insufficient for training. Thus, to obtain more reasonable estimations of these parameters as well as to further improve the algorithm performance, we train on the 2D LIVE image database [37], which yields the following trained parameters: $\beta_1 = 0.369$, $\beta_2 = 0.153$, w_s = 0.75, 1.25, 0.5, 4, and 7 for $s \in \{1, 2, 3, 4, 5\}$, respectively.

With these newly-trained parameters, we apply MAD to each view of the stereoscopic image by assuming that the 3D quality is influenced by 2D content, and that the local masking which comes from the two visual paths is also observed. As we advocated in Section I, the perceived quality of a 3D image is determined largely by the quality of the monocular view with higher contrast. If the contrasts perceived by both eyes are the same, then we assume that the quality of both visual signals are of equal significance. Therefore, we weight each MAD score by using block-based contrast maps.

Specifically, a block-based contrast map for an image is computed in the lightness domain by first dividing the image into blocks of 16×16 pixels (with 75% overlap between neighboring blocks), and then measuring the root mean squared contrast of each block.

$$C(b) = \tilde{\sigma}(b)/\mu(b) \tag{3}$$

where $\mu(b)$ represents the average luminance value of block b, and $\tilde{\sigma}(b)$ represents the minimum standard deviation among the four 8×8 subblock within b (see Appendix A in [1]).

For both the reference and distorted stereopairs, we first compute their block-based contrast maps denoted as C_L^{ref}, C_R^{ref}, C_L^{dst}, and C_R^{dst}, respectively. Here, "ref" denotes the reference stereoscopic image; "dst" denotes the distorted stereoscopic image; "L" and "R" denote the left and right view of each stereopair. The normalized weights of the distorted stereopairs are then computed via

$$W_L = (\bar{C}_L^{dst}/\bar{C}_L^{ref})^{\gamma} \tag{4}$$

$$W_R = (\bar{C}_R^{dst}/\bar{C}_R^{ref})^{\gamma} \tag{5}$$

where $\bar{C}$ is the mean value of C, and $\gamma = 2$ is a factor that is aimed to emphasize higher contrast values (see the online supplement at http://vision.okstate.edu/3dmad/ for experiment results based on different γ values.). Note that the computed block-based contrast may vary even for the reference stereopairs; thus, we normalized the contrast of each distorted view by dividing by the contrast of each corresponding reference view to ensure that the weight difference for a stereopair is produced by only the distortion, and not by the image content.

Finally, the 2D-MAD score is given by

$$\text{S}_{2D\text{-}MAD} = \frac{W_L \cdot \exp(\frac{\text{MAD}_L}{100}) + W_R \cdot \exp(\frac{\text{MAD}_R}{100})}{W_L + W_R}, \tag{6}$$

where MAD_L and MAD_R denote the MAD quality scores of the left and right views of the distorted stereoscopic image, respectively. The division by 100 simply brings the two values (MAD_L and MAD_R) into an approximate range of [0, 1].

B. MCM-Based Cyclopean IQA

In the second stage, the appearance-based strategy (which has been used in [1]) is applied to compute the statistical difference between the reference and distorted stereoscopic images by using the cyclopean lightness distance and pixel-based contrast maps. These cyclopean feature images are synthesized by using stereopairs and an

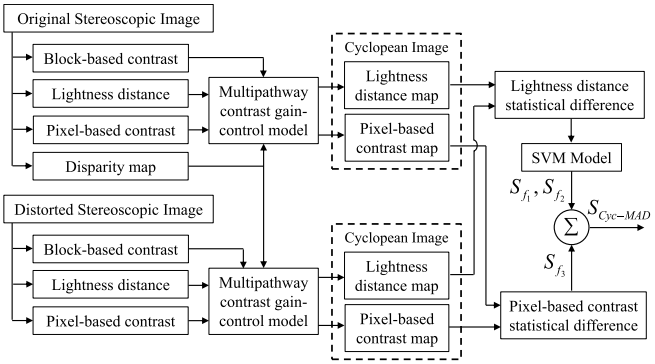


Fig. 4. Block diagram of MCM-based cyclopean IQA.

estimated disparity map based on a multipathway contrast gain-control model [35], [36]. Figure 4 shows a block diagram of this stage.

1) Feature Maps: We propose to use the lightness distance and pixel-based contrast as our algorithm's raw inputs for cyclopean-based quality assessment. In [64], we demonstrated that lightness distance can be a useful feature for detecting main subjects in photos. Here, we use maps of both local and global versions of that feature for estimating quality. The global lightness distance measures how much the lightness of each pixel differs from the average lightness of the whole image. The local lightness distance measures how much the lightness of each pixel differs from the average lightness of a local area around that pixel. As a third feature map, we use the pixel-based contrast; larger pixel-based contrast values often indicate edges or sharper regions in an image. These three features are computed via

$$f_1(x,y) = \left|L^*(x,y) - \bar{L}_I^*\right|, \tag{7}$$

$$f_2(x,y) = \left|L^*(x,y) - \bar{L}_B^*(x,y)\right|, \tag{8}$$

$$f_3(x,y) = \frac{L^*(x,y)}{\bar{L}_B^*(x,y) + K}, \tag{9}$$

where L^* denotes the lightness component in the Commission Internationale de i' Eclairage (CIE) 1976 (L^*, a^*, b^*) color space (CIELAB); $\bar{L}_I^*$ denotes the average lightness value for the whole image I; $\bar{L}_B^*(x,y)$ denotes the average lightness value of a 9×9 block centered around pixel (x,y); and $K = 0.001$ is a small constant that prevents division by zero.

2) MCM-Based Cyclopean Feature Maps: Given the lightness distance and pixel-based contrast maps for each view of the reference and distorted stereoscopic image, we build disparity-compensated cyclopean maps corresponding to these two types of features. First, we compute a disparity map of the reference stereopairs (denoted by D) by using the segment-based stereo-matching approach proposed in [65]. Then, the cyclopean feature image is synthesized based on the MCM [35], [36], which assumes that each eye exerts contrast gain control not only on the other eye's *visual* signal, but also on the incoming *gain control* signal from the other eye, with both effects in proportion to the eye's own total contrast energy.

Specifically, MCM employs three gain-control mechanisms: (1) attenuation of signal contrast in the non-dominant eye; (2) stronger direct inhibition from the dominant eye; and (3) stronger indirect inhibition from the dominant eye to the gain control signal coming from the non-dominant eye. Let C_L and C_R denote the scene contrast presented to the left and right eye, respectively. By assuming that the left eye is the dominant eye, after contrast gain control, the signal strengths perceived by the two eyes are modeled as:

$$C'_L = C_L \frac{1}{1+\frac{\epsilon_R}{1+\beta\epsilon_L}}, \quad C'_R = \eta C_R \frac{1}{1+\frac{\alpha\epsilon_L}{1+\epsilon_R}} \tag{10}$$

where $\epsilon_L = \rho C_L^{\gamma_1}$ and $\epsilon_R = \rho(\eta C_R)^{\gamma_1}$. The parameter ρ is the contrast gain-control efficiency; η is used to model contrast attenuation in right eye; α and β are used to model the stronger inhibition to the right eye from the left eye; and γ_1 is the transducer nonlinearity in the gain-control pathway. According to [36], the perceived contrast of the cyclopean image is given by

$$C = \left[\left(C_L \frac{1}{1+\frac{\epsilon_R}{1+\beta\epsilon_L}}\right)^{\gamma_2} + \left(\eta C_R \frac{1}{1+\frac{\alpha\epsilon_L}{1+\epsilon_R}}\right)^{\gamma_2}\right]^{\frac{1}{\gamma_2}}, \tag{11}$$

where γ_2 is the transducer nonlinearity for the power summation used in the binocular contrast combination.

Motivated by [36], we employ an MCM-based technique to compute the cyclopean feature images based on two assumptions. First, we assume that both eyes will contribute equally when the reference stereopairs are viewed. Second, when the distorted stereopairs are viewed, the eye that receives the signal with the higher contrast will dominate. Based on these two assumptions, we compute the perceived cyclopean lightness distance and pixel-based contrast images for each pixel (x,y) by using Eq. (12), as shown at the bottom of this page, where $C_{f_i}(i = 1,2,3)$ denotes the computed cyclopean lightness distance and pixel-based contrast image; $d_{x,y} = D(x,y)$ denotes an estimated disparity index in D; $\epsilon_L(x,y) = \rho[\eta_L C_L(x,y)]^{\gamma_1}$ and $\epsilon_R(x,y) = \rho[\eta_R C_R(x,y)]^{\gamma_1}$; and C_L and C_R denote the block-based contrast value computed via Eq. (3) for the left and right view of a stereo image, respectively.

We set the following parameter values by referring to the experiment results provided in [66]: $\gamma_1 = 1.5$; $\rho = 10$. For the reference stereopairs, we set $\eta_L = \eta_R = 1$, which indicates

$$C_{f_i}(x,y) = \frac{\left[\left(\eta_L f_{i,L}(x - d_{x,y}, y)\frac{1}{1+\frac{\epsilon_R(x-d_{x,y},y,y)}{1+\beta\epsilon_L(x-d_{x,y},y,y)}}\right)^{\gamma_2} + \left(\eta_R f_{i,R}(x,y)\frac{1}{1+\frac{\alpha\epsilon_L(x,y)}{1+\epsilon_R(x,y)}}\right)^{\gamma_2}\right]^{\frac{1}{\gamma_2}}}{\left[\left(\frac{\eta_L}{1+\frac{\epsilon_R(x-d_{x,y},y)}{1+\beta\epsilon_L(x-d_{x,y},y)}}\right)^{\gamma_2} + \left(\frac{\eta_R}{1+\frac{\alpha\epsilon_L(x,y)}{1+\epsilon_R(x,y)}}\right)^{\gamma_2}\right]^{\frac{1}{\gamma_2}}}, \quad (i = 1,2,3) \tag{12}$$

that there is no contrast attenuation coming from the other (non-dominant) eye. For the distorted stereopairs, the dominant eye is first determined by comparing the averaged block-based contrast of the two views (i.e., $\bar{C}_L$ and $\bar{C}_R$). If the left eye dominates (i.e., $\bar{C}_L > \bar{C}_R$), then we set $\eta_L = 1$ and $\eta_R = 0.9$. Otherwise, we set $\eta_L = 0.9$ and $\eta_R = 1$ to indicate the dominant role of the right eye. Note that the value of 0.9 was also selected by referring to [66]. Additionally, for simplicity we set $\gamma_2 = 0.5$ and $\alpha = \beta = 1$, which correspond to supra-linear summation and equal inhibition between the two eyes. We have found that changing γ_2 does not affect the algorithm performance significantly (see the online supplement at http://vision.okstate.edu/3dmad/).

3) Compute Statistical Differences: Based on the cyclopean feature images computed from the lightness distance and pixel-based contrast maps, we estimate quality by using local statistical differences between the cyclopean feature images for the reference and distorted stereopairs by the local standard deviation, skewness, and kurtosis. This approach of comparing local statistics was shown in [1] to be highly effective for quantifying the appearances of suprathreshold distortions. Here, we propose an alternative approach to map these statistical differences to image quality.

Specifically, in [1], quality is assessed based on a statistical difference map, which is computed by linearly combining the three statistics of the log-Gabor subbands at five scales ($s \in \{1, 2, 3, 4, 5\}$) and four orientations ($o \in \{1, 2, 3, 4\}$):

$$\eta(b) = \sum_{s=1}^{5}\sum_{o=1}^{4} w_s[|\sigma_{s,o}^{ref}(b) - \sigma_{s,o}^{dst}(b)| + 2|\varsigma_{s,o}^{ref}(b) - \varsigma_{s,o}^{dst}(b)| + |\kappa_{s,o}^{ref}(b) - \kappa_{s,o}^{dst}(b)|], \quad (13)$$

where $\sigma_{s,o}(b)$, $\varsigma_{s,o}(b)$, and $\kappa_{s,o}(b)$ denote, respectively, the standard deviation, skewness, and kurtosis computed for a 16×16 block b (with 75% overlap between blocks) in the log-Gabor subband corresponding to scale s and orientation o. The scale-specific weights ($w_s = 0.5$, 0.75, 1, 5, and 6) were obtained through training on the A57 database [63], which also accounts for the HVS's preference for coarse scales over fine scales. Finally, the statistical difference map is collapsed into a single scalar that represents the perceived quality degradation of the image.

Instead of using the Eq. (13), we propose to train the three statistical difference values by using support vector machine (SVM) learning. Specifically, we decompose the cyclopean feature images of the reference and distorted stereopairs by using log-Gabor filters at five scales and four orientations, and then compute three statistical difference maps for each scale. This first step is given by

$$\eta_{\phi_s}(b) = \sum_{o=1}^{4} \left|\phi_{s,o}^{ref}(b) - \phi_{s,o}^{dst}(b)\right|, \quad (14)$$

where ϕ represents the three statistics (σ, ς, and κ).

Next, we collapse each statistical difference map into a single scalar, which is given by

$$\bar{\phi}_s = \left[\frac{1}{B_{\phi_s}}\sum_{b}\eta_{\phi_s}(b)^2\right]^{\frac{1}{2}}, \quad (15)$$

TABLE II

PERFORMANCE COMPARISON OF USING DIFFERENT FEATURES/FEATURE COMBINATIONS TO PREDICT STEREOSCOPIC IMAGE QUALITY. NOTE THAT FOR THIS COMPARISON, WE DO NOT CONSIDER THE 2D-MAD QUALITY ESTIMATE ON STEREOPAIRS, AND ONLY USE THE QUALITY DEGRADATION MEASUREMENTS OF THE CORRESPONDING CYCLOPEAN FEATURE IMAGES AS THE FINAL QUALITY ESTIMATE OF THE STEREOSCOPIC IMAGE

	LIVE3D Phase I		LIVE3D Phase II	
Cyclopean features	CC	SROCC	CC	SROCC
Global lightness distance (f_1)	0.944	0.938	0.888	0.877
Local lightness distance (f_2)	0.937	0.927	0.893	0.889
Lightness distance ($f_1 + f_2$)	0.948	0.939	0.899	0.891
Pixel-based contrast (f_3)	0.909	0.903	0.886	0.879
All ($f_1 + f_2 + f_3$)	0.944	0.937	0.921	0.915

where B_{ϕ_s} denotes the total number of blocks within one statistical difference map (ϕ) at a specific scale s ($s \in \{1, 2, 3, 4, 5\}$). Thus, for one distorted cyclopean feature image, we compute a vector of 15 statistical-difference-based features denoted by

$$\mathbf{v} = [\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3, \bar{\sigma}_4, \bar{\sigma}_5, \bar{\varsigma}_1, \bar{\varsigma}_2, \bar{\varsigma}_3, \bar{\varsigma}_4, \bar{\varsigma}_5, \bar{\kappa}_1, \bar{\kappa}_2, \bar{\kappa}_3, \bar{\kappa}_4, \bar{\kappa}_5]. \quad (16)$$

Given the feature vectors extracted from the cyclopean feature images, we then map each vector into a quality score by using the trained SVM regression models. Specifically, we trained our models on the 2D LIVE image quality database [37], which contains 779 distorted images with five types of distortions: Gaussian blurring, additive white noise, JPEG compression, JPEG2000 compression, and fast fading. For each 2D image, we extracted the same type of features. The only difference is that these statistical-difference-based features were extracted from feature maps corresponding to one single reference/distorted image, and not the cyclopean images. We used the LIBSVM package [67] to implement the training. To improve predictive performance, optimal radial basis function kernel parameters were used for the SVM regression.

Note that the SVM-based approach is applied only to the lightness distance (local and global) cyclopean images, resulting into two computed feature vectors, v_{f_1} and v_{f_2}; the two corresponding quality degradation scores are denoted by S_{f_1} and S_{f_2}, respectively. The reason for using only the lightness distance maps is based on several observations. First, the lightness distance map of the two views can be combined (via the MCM model) to effectively predict 3D image quality when a symmetrically-distorted 3D image is viewed. By training on a 2D image database, it is assumed that the given 3D image has symmetrically distorted stimuli, and therefore its quality is equally determined by the quality of each stereoscopic view. Thus, training the lightness distance statistical features on LIVE [37] works quite well for those images that contain symmetrically distorted stimuli, but it is less effective on asymmetrically distorted images, because the overall quality of these images are more determined by the visual signal that has higher contrast (as shown in Table II). Thus, as mentioned previously, we additionally employ

the pixel-based contrast statistical features to improve the performance of 3D-MAD on asymmetrically distorted images. In this case, we have found that the 2D LIVE image database is less effective at training these contrast features, possibly due to the different characteristics that the HVS uses when judging the quality of a 2D image vs. an asymmetrically distorted 3D image.

In this work, we adopt a similar strategy as in [1] to solve this problem. Specifically, we first decompose the cyclopean pixel-based contrast images by using the middle radial frequency (third scale) log-Gabor filters, and then compute the statistical difference map via:

$$\eta_{f_3}(b) = \sum_{o=1}^{4}[|\sigma_{3,o}^{ref}(b) - \sigma_{3,o}^{dst}(b)| + 2|\varsigma_{3,o}^{ref}(b) - \varsigma_{3,o}^{dst}(b)| + |\kappa_{3,o}^{ref}(b) - \kappa_{3,o}^{dst}(b)|], \quad (17)$$

where σ, ς, and κ denote the three statistics. The third scale of the log-Gabor subbands are utilized because we found that distortions in the cyclopean pixel-based contrast images are more representative and distinguishable in the middle-level frequency range. The final quality degradation score of perceived contrast (denoted by S_{f_3}) is then given by

$$S_{f_3} = \left[\frac{1}{B}\sum_{b}\eta_{f_3}(b)^2\right]^{\frac{1}{2}}, \quad (18)$$

where B denotes the total number of blocks within the pixel-based contrast statistical difference map.

Finally, based on all three scores (S_{f_1}, S_{f_2}, S_{f_3}), we compute the MCM-based cyclopean feature image quality degradation score (denoted by $\mathrm{S}_{Cyc\text{-}MAD}$) as

$$\mathrm{S}_{Cyc\text{-}MAD} = \frac{S_{f_1} + S_{f_2}}{10} + S_{f_3}. \quad (19)$$

Note that S_{f_1} and S_{f_2} are obtained through training on the 2D LIVE image database (that has DMOS values approximately between 0 and 100) and then testing using the cyclopean feature images. Thus, both values approximately range from 0 to 100. The S_{f_3} value is obtained directly by analyzing the cyclopean pixel-based contrast images and approximately ranges from 0 to 10. The division by 10 in Eq. (19) was chosen simply to bring the three values (S_{f_1}, S_{f_2}, and S_{f_3}) on approximately the same scale.

C. 3D-MAD Quality Estimate

Given both the 2D-MAD-based score from the stereopairs and the MCM-based score from the three cyclopean feature images, the final stage of 3D-MAD is to combine these scores into an overall perceived distortion estimate (denoted by $\mathrm{S}_{3D\text{-}MAD}$) for the stereoscopic image. Specifically, $\mathrm{S}_{3D\text{-}MAD}$ is computed as a product of $\mathrm{S}_{2D\text{-}MAD}$ and $\mathrm{S}_{Cyc\text{-}MAD}$, which is given by

$$\mathrm{S}_{3D\text{-}MAD} = \mathrm{S}_{2D\text{-}MAD} \times \mathrm{S}_{Cyc\text{-}MAD}. \quad (20)$$

Smaller values of $\mathrm{S}_{3D\text{-}MAD}$ denote predictions of better stereoscopic image quality. A value of $\mathrm{S}_{3D\text{-}MAD} = 0$ indicates that the distorted stereoscopic image is equal in quality to the reference image.

IV. Results and Analysis

In this section, we analyze 3D-MAD's ability to predict stereoscopic image quality. For this task, we test 3D-MAD on three publicly available 3D image databases: (1) the LIVE 3D image database [28], (2) the MCL-3D image database [68], and (3) the IRCCyN/IVC 3D image database [12].

A. Stereoscopic Image Quality Databases

The LIVE 3D image quality database [28], from the University of Texas at Austin, USA, was constructed in two phases. Phase I contains 20 original images and 365 symmetrically distorted stereopairs corresponding to five distortion types: Gaussian blurring (GBLUR), additive white noise (WN), JPEG compression (JPEG), JPEG2000 compression (JP2K), and simulated packet-loss of transmitted JPEG2000-compressed images, which is also known as fast fading (FF). Phase II contains eight original images and 360 distorted stereopairs, with the same five distortion types. For each distortion type, every original stereopair was processed to create three symmetric distorted stereopairs and six asymmetric distorted stereopairs. Both phases contain the ground truth depth/disparity information for every reference image and the associated DMOS values for distorted images.

The MCL-3D image quality database [68], from the University of Southern California, USA, contains 693 symmetrically distorted stereopairs and associated mean opinion score (MOS) values. One-third of the images in the database are of resolution 1024×728, and two-thirds are of resolution 1920×1080. Nine image-plus-depth sources were first selected, and a depth-image-based rendering technique was used to render stereopairs. Four levels of distortions were applied to either the texture image or the depth image before stereoscopic image rendering, which include Gaussian blur, additive white noise, down-sampling blur (SBLUR), JPEG compression, JPEG2000 compression and transmission error (TERROR). The distortion caused by imperfect rendering was also examined.

The IRCCyN/IVC 3D image database [12], from the University of Nantes, France, contains six original images, 15 symmetrically distorted versions of each original image, and subjective ratings of quality for each distorted image (DMOS value) obtained by averaging 14 subjective scores. Three types of distortions are included: JPEG compression, JPEG2000 compression, and blurring (BLUR); each distortion type has five different degradation levels. Note that there are two types of blurred images: Gaussian filtered images and re-sampled images (see [12]). The database contains 96 stereopairs, but without depth and disparity information.

B. Algorithms and Performance Measures

We compared 3D-MAD with seven other 3D IQA algorithms: an SSIM-based stereo IQA model proposed by Benoit *et al.* [12], an SDP-based method proposed by Ko *et al.* [54], a BSS weighted method proposed by Wang *et al.* [56], the cyclopean MS-SSIM proposed by Chen *et al.* [30], the frequency-integrated PSNR (FI-PSNR) proposed by Lin and Wu [60], the BJND-based method

TABLE III

OVERALL PERFORMANCES OF 3D-MAD AND OTHER 2D/3D IQA ALGORITHMS ON THE LIVE, MCL, AND IRCCyN/IVC 3D IMAGE DATABASES. ITALICIZED ENTRIES DENOTE 2D ALGORITHMS. RESULTS OF THE BEST-PERFORMING 3D IQA ALGORITHM ARE BOLDED

		SSIM	*MS-SSIM*	*VIF*	*MAD*	You	Benoit	Wang	Ko	Chen	Lin	Shao	2D-MAD	Cyc-MAD	3D-MAD
CC	LIVE (Phase I)	*0.872*	*0.926*	*0.925*	*0.942*	0.895	0.915	0.888	0.910	0.917	0.864	0.932	0.945	0.944	**0.951**
	LIVE (Phase II)	*0.801*	*0.778*	*0.840*	*0.854*	0.729	0.812	0.817	0.760	0.906	0.658	0.836	0.901	0.921	**0.927**
	MCL	*0.883*	*0.915*	*0.856*	*0.931*	0.837	0.871	0.890	0.893	0.881	0.870	0.848	**0.942**	0.912	0.934
	IRCCyN/IVC	*0.736*	*0.802*	*0.798*	*0.802*	0.745	0.820	0.725	0.742	0.689	0.695	0.716	0.824	**0.878**	0.864
	Average	*0.851*	*0.877*	*0.866*	*0.907*	0.819	0.864	0.862	0.855	0.884	0.806	0.858	0.925	0.920	**0.932**
SROCC	LIVE (Phase I)	*0.876*	*0.922*	*0.920*	*0.939*	0.896	0.911	0.890	0.907	0.916	0.856	0.927	0.942	0.937	**0.944**
	LIVE (Phase II)	*0.792*	*0.772*	*0.817*	*0.842*	0.681	0.806	0.805	0.756	0.901	0.638	0.819	0.890	0.915	**0.924**
	MCL	*0.893*	*0.925*	*0.858*	*0.931*	0.851	0.874	0.901	0.903	0.884	0.869	0.853	**0.942**	0.917	0.939
	IRCCyN/IVC	*0.691*	*0.716*	*0.706*	*0.772*	0.743	0.791	0.678	0.662	0.673	0.636	0.648	0.811	**0.866**	0.853
	Average	*0.852*	*0.874*	*0.854*	*0.902*	0.814	0.862	0.861	0.853	0.883	0.794	0.851	0.921	0.919	**0.931**
RMSE	LIVE (Phase I)	*8.033*	*6.193*	*6.228*	*5.498*	7.312	6.633	7.536	6.804	6.550	8.242	5.941	5.365	5.406	**5.052**
	LIVE (Phase II)	*6.757*	*7.096*	*6.132*	*5.869*	7.727	6.582	6.502	7.341	4.767	8.496	6.196	4.889	4.393	**4.220**
	MCL	*1.221*	*1.051*	*1.345*	*0.953*	1.424	1.277	1.185	1.172	1.230	1.285	1.380	**0.874**	1.065	0.930
	IRCCyN/IVC	*14.936*	*13.172*	*13.299*	*13.176*	14.719	12.639	15.191	14.792	15.998	15.860	15.408	12.503	**10.578**	11.118
	Average	*5.126*	*4.567*	*4.476*	*4.049*	5.262	4.618	4.940	4.933	4.336	5.692	4.566	3.698	3.552	**3.395**

proposed by Shao *et al.* [62], and a 3D IQA approach proposed by You *et al.* [11], in which the quality of each stereoscopic image was computed based on various combinations of 2D quality estimates on stereopairs and disparity maps. Note that for the methods by You *et al.* and by Benoit *et al.*, the combinations leading to the best performances are presented. Specifically, for You *et al.*'s method, the quality of each stereopair was estimated by SSIM [38] and the quality of the disparity map was estimated via a mean absolute difference; the overall stereoscopic image quality was then computed via a "global combination," in which the optimum parameters were determined by the Levenberg-Marquardt algorithm. For Benoit *et al.*'s method, a local SSIM map weighted by the Euclidean-distance-based local disparity distortion measure was employed to estimate the quality of each stereoscopic view, and the final quality estimate was the average of the two views. For both algorithms, the same stereo matching algorithm [69] was used to create the disparity maps.

We also list the results of four 2D IQA algorithms for reference: SSIM [38], MS-SSIM [59], VIF [70], and MAD [1] (the original 2D version). For these 2D IQA algorithms, the predicted quality of a stereoscopic image was taken to be the average quality predicted from the left and right views.

Before evaluating the performance of a particular quality assessment method on a particular database, we applied a logistic transform to bring the prediction values on the same scales as the DMOS values. The logistic transform recommended by Video Quality Experts Group [71] is a four-parameter sigmoid given by

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp(-\frac{x-\tau_3}{|\tau_4|})} + \tau_2, \tag{21}$$

where x denotes the raw predicted score, and where τ_1, τ_2, τ_3, and τ_4 are free parameters selected to provide the best fit of the predicted scores to the MOS/DMOS values. Three criteria were used to measure the prediction monotonicity and prediction accuracy of each algorithm: (1) the Spearman Rank-Order Correlation Coefficient (SROCC), (2) the Pearson Linear Correlation Coefficient (CC), and (3) the Root Mean Square Error (RMSE) after non-linear regression. Note that the logistic transform in Eq. (21) will affect only CC and RMSE, not SROCC.

C. Overall Performance

The overall testing results on the LIVE, MCL, and IRCCyN/IVC 3D image databases are shown in Table III in terms of CC, SROCC, and RMSE. Also included are the testing results of the first and the second stage of 3D-MAD (denoted by "2D-MAD" and "Cyc-MAD," respectively) for comparison. Italicized entries denote 2D IQA algorithms. The results of the best-performing 3D IQA algorithm in each case are bolded. As recommended in [68], we tested 648 stereoscopic images in the MCL-3D database, which cover six types of distortions applied to the texture image, the depth image, and both, respectively. Also note that the MCL-3D image pairs are of large size, which may require different parameter settings in the disparity estimation algorithm (compared with those for LIVE and IRCCyN/IVC). Thus, we rescaled[2] the images to 1/2 (for 1024×728 images) or 1/3 (for 1920×1080 images) of the original size before testing in order to use consistent parameter settings, as well as to save computational time.

From Table III, observe that 3D-MAD outperforms all the other 2D/3D IQA algorithms in terms of all three performance criteria on all databases. Specifically, on the LIVE Phase I and MCL-3D databases, which contain symmetrically distorted stimuli, all IQA algorithms perform quite well, and there seems to be less performance difference between 2D and 3D methods. However, these 2D-based IQA algorithms generally perform worse on the LIVE Phase II database, which partially contains asymmetrically distorted stimuli. The reason for this performance drop is that these algorithms estimate the quality of each stereopair without taking into account the depth/disparity information. To demonstrate this latter assertion, Table IV shows the CC, SROCC, and RMSE values of all these algorithms tested on separated subsets of symmetrically and asymmetrically distorted images in the LIVE Phase II database. Observe from Table IV that 3D-MAD improves upon the performance of the original MAD algorithm, most notably on asymmetrically distorted images.

In comparison, the 3D-based IQA algorithms take into account the stereo depth/disparity information,

[2]In this paper, we used bicubic interpolation to rescale the images; the output pixel value is a weighted average of pixels in the nearest 4×4 neighborhood.

TABLE IV
PERFORMANCE OF 3D-MAD AND OTHER IQA ALGORITHMS ON SYMMETRICALLY AND ASYMMETRICALLY DISTORTED STIMULI IN THE LIVE PHASE II DATABASE. ITALICIZED ENTRIES DENOTE 2D ALGORITHMS. RESULTS OF THE BEST-PERFORMING 3D IQA ALGORITHM ARE BOLDED

	Symmetric			Asymmetric		
Metric	CC	SROCC	RMSE	CC	SROCC	RMSE
SSIM [38]	*0.839*	*0.825*	*6.950*	*0.760*	*0.736*	*6.658*
MS-SSIM [59]	*0.927*	*0.915*	*5.990*	*0.694*	*0.677*	*7.589*
VIF [70]	*0.920*	*0.914*	*5.291*	*0.771*	*0.731*	*6.513*
MAD [1]	*0.936*	*0.934*	*4.687*	*0.783*	*0.756*	*6.379*
You [11]	0.911	0.898	7.128	0.659	0.604	8.009
Benoit [12]	0.921	0.910	5.712	0.746	0.732	6.976
Wang [56]	0.868	0.861	6.364	0.768	0.736	6.570
Ko [54]	0.896	0.886	6.558	0.679	0.669	7.703
Chen [30]	0.937	0.925	4.429	0.875	0.854	4.927
Lin [60]	0.816	0.811	8.467	0.588	0.549	8.510
Shao [62]	0.931	0.936	5.187	0.762	0.739	6.644
3D-MAD	**0.954**	**0.947**	**3.900**	**0.903**	**0.895**	**4.371**

and/or binocular HVS properties, and thus these algorithms should achieve better performance than 2D-based IQA methods. However, the performances of the methods by You *et al.* and Benoit *et al.* depend largely on the accuracy of the stereo-matching algorithm used to compute the disparity maps, and they are also influenced by the fact that 2D QA of the disparity maps may not always coincide with human judgments of quality on stereoscopic images. Although Wang *et al.*, Ko *et al.*, and Lin *et al.* used properties of binocular vision, their methods are still based on weighting two quality scores obtained by applying 2D QA methods (e.g., SSIM and PSNR) on both views. Shao *et al.*'s method is also based on combining the quality scores obtained from the left and right views, but the combination strategy varies according to different image regions (non-corresponding region, binocular suppression region, and binocular fusion region). Since more efficient features are employed, this method performs quite well on the LIVE Phase I database, but the performance is not that high for the other databases. Chen *et al.*'s method models binocular rivalry and estimates quality based on a cyclopean image, and thus achieves better performance than the other six algorithms on the LIVE Phase II database. On the IRCCyN/IVC 3D image database, Benoit *et al.*'s method and 3D-MAD perform best.

Compared with 2D-MAD and Cyc-MAD, observe that the combined stages either improve or balance the performance of each individual stage across different 3D image databases. For example, the combined stages help increase the overall performance on the LIVE 3D image database, but only achieve a performance balance on the MCL and IRCCyN/IVC 3D databases. For MCL, 2D-MAD performs better than Cyc-MAD. For IRCCyN/IVC, 2D-MAD and all other 2D/3D QA algorithms considered here perform less effectively (see Section IV-E for a detailed explanation). Overall, 3D-MAD still performs competitively well, and both stages are required to achieve a reasonable performance balance when all three databases are considered.

The last rows of the CC, SROCC, and RMSE results in Table III show the average CC, SROCC, and RMSE,

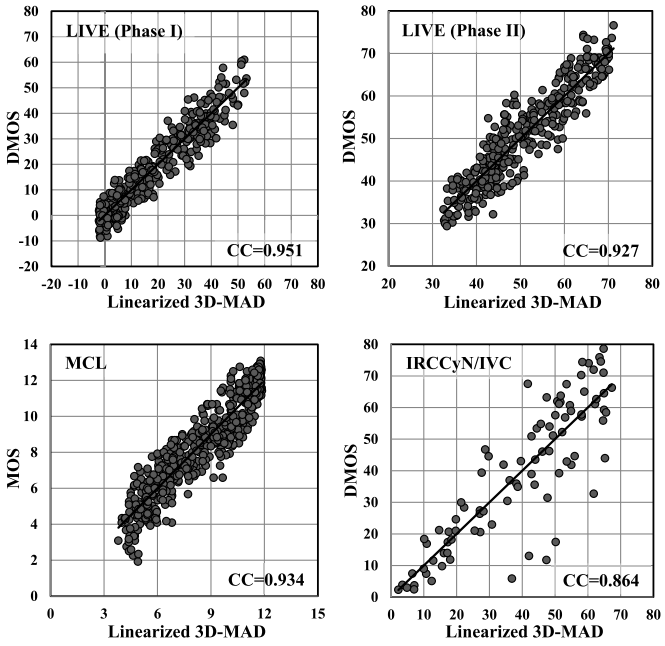


Fig. 5. Scatter plots of objective scores predicted by 3D-MAD algorithm after logistic transform versus subjective scores on different image databases.

where the averages are weighted by the number of distorted images tested in each database. Also, shown in Figure 5 are scatter-plots of logistic-transformed 3D-MAD quality predictions vs. subjective ratings (MOS/DMOS) on different databases. In all graphs, the y-axis denotes the subjective ratings of the perceived distortions and the x-axis denotes the predicted quality value transformed via Eq. (21). Despite the presence of some outliers, the plots are generally heteroscedastic. In summary, when looking at the overall performance across databases, 3D-MAD has a better average performance than other 2D/3D IQA methods.

D. Statistical Significance

A statistical significance test was performed by using an F-test to quantify whether the numerical difference between the IQA algorithms' performances are statistically significant. The test statistic is the ratio of two algorithms' residual variances (errors in predictions), denoted by $F = \sigma_A^2/\sigma_B^2$. A smaller residual variance indicates a better prediction. Values of $F > F_{critical}$ (or $F < 1/F_{critical}$) indicate that at a given confidence level, method A has significantly larger (or smaller) residuals than method B, where $F_{critical}$ is computed based on the number of residuals and the confidence level. Note that if the residuals are not Gaussian, then the significance test is often inconclusive. In this paper, a formal test using Jarque-Bera (JB) statistic [72] is used to measure the Gaussianity of the residuals. A smaller value of the JB statistic denotes less deviation from Gaussianity, and vice versa.

Table V shows the overall statistical performance of each 3D IQA algorithm on the LIVE, MCL, and IRCCyN/IVC 3D image databases. Each entry is the ratio of the residual variance of the algorithm in the row to the algorithm in the column. Bold entries denote that the algorithm in the row has a statistically smaller residual variance than the algorithm in the

TABLE V

STATISTICAL SIGNIFICANCE RELATIONSHIPS (RATIO OF RESIDUAL VARIANCE) BETWEEN 3D IQA ALGORITHMS ON THE LIVE, MCL, AND IRCCyN/IVC 3D IMAGE DATABASES. SEE TEXT FOR DETAILS

	You	Benoit	Wang	Ko	Chen	Lin	Shao	3D-MAD	You	Benoit	Wang	Ko	Chen	Lin	Shao	3D-MAD
	LIVE (Phase I)								LIVE (Phase II)							
You	–	*1.215*	0.941	1.155	*1.246*	**0.787**	*1.515*	*2.095*	–	*1.378*	*1.412*	1.108	*2.627*	**0.827**	*1.555*	*3.352*
Benoit	**0.823**	–	**0.775**	0.950	1.025	**0.648**	*1.246*	*1.723*	**0.726**	–	1.025	**0.804**	*1.906*	**0.600**	1.128	*2.432*
Wang	1.062	*1.291*	–	*1.227*	*1.324*	**0.836**	*1.609*	*2.225*	**0.708**	0.976	–	**0.784**	*1.860*	**0.586**	1.101	*2.373*
Ko	0.866	1.052	**0.815**	–	1.079	**0.681**	*1.312*	*1.814*	0.903	*1.244*	*1.275*	–	*2.372*	**0.747**	*1.404*	*3.026*
Chen	**0.802**	0.975	**0.755**	0.927	–	**0.632**	*1.216*	*1.681*	**0.381**	**0.525**	**0.538**	**0.422**	–	**0.315**	**0.592**	*1.276*
Lin	*1.271*	*1.544*	*1.196*	*1.467*	*1.583*	–	*1.925*	*2.661*	*1.209*	*1.666*	*1.707*	*1.339*	*3.176*	–	*1.880*	*4.053*
Shao	**0.660**	**0.802**	**0.621**	**0.762**	**0.823**	**0.519**	–	*1.383*	**0.643**	0.886	0.908	**0.712**	*1.690*	**0.532**	–	*2.156*
3D-MAD	**0.477**	**0.580**	**0.449**	**0.551**	**0.595**	**0.376**	**0.723**	–	**0.298**	**0.411**	**0.421**	**0.330**	**0.784**	**0.247**	**0.464**	–
JBSTAT	*3.3*	8.2	19.4	6.5	45.6	*1.9*	*2.0*	*1.2*	9.2	45.2	8.3	30.5	31.5	8.8	6.0	*0.4*
	MCL								IRCCyN/IVC							
You	–	*1.243*	*1.444*	*1.477*	*1.342*	*1.229*	1.066	*2.346*	–	1.356	0.939	0.990	0.847	0.861	0.913	*1.753*
Benoit	**0.804**	–	*1.162*	*1.188*	1.079	0.989	**0.857**	*1.887*	0.737	–	**0.692**	0.730	**0.624**	**0.635**	**0.673**	1.292
Wang	**0.692**	**0.861**	–	1.022	0.929	**0.851**	**0.738**	*1.624*	1.065	*1.445*	–	1.055	0.902	0.917	0.972	*1.867*
Ko	**0.677**	**0.842**	0.978	–	0.909	**0.832**	**0.722**	*1.589*	1.010	1.370	0.948	–	0.855	0.870	0.922	*1.770*
Chen	**0.745**	0.926	1.076	1.100	–	0.916	**0.794**	*1.748*	1.181	*1.602*	1.109	1.170	–	1.017	1.078	*2.070*
Lin	**0.814**	1.012	*1.175*	*1.201*	1.092	–	**0.867**	*1.909*	1.161	*1.575*	1.090	1.150	0.983	–	1.060	*2.035*
Shao	0.938	*1.166*	*1.355*	*1.385*	*1.259*	*1.153*	–	*2.201*	1.096	*1.486*	1.029	1.085	0.928	0.944	–	*1.921*
3D-MAD	**0.426**	**0.530**	**0.616**	**0.629**	**0.572**	**0.524**	**0.454**	–	**0.571**	0.774	**0.536**	**0.565**	**0.483**	**0.491**	**0.521**	–
JBSTAT	6.8	7.7	21.1	7.3	28.0	*3.8*	10.3	*5.4*	*2.0*	47.2	6.1	19.2	*3.2*	7.9	10.4	26.7

TABLE VI

SROCC VALUES OF 3D-MAD AND 2D/3D IQA ALGORITHMS ON DIFFERENT TYPES OF DISTORTION ON THE LIVE, MCL, AND IRCCyN/IVC 3D DATABASES. ITALICIZED ENTRIES DENOTE 2D IQA ALGORITHMS. RESULTS OF THE BEST-PERFORMING 3D IQA ALGORITHM ARE BOLDED

		SSIM	*MS-SSIM*	*VIF*	*MAD*	You	Benoit	Wang	Ko	Chen	Lin	Shao	3D-MAD
LIVE (Phase I)	JP2K	*0.857*	*0.898*	*0.902*	*0.925*	0.884	0.887	0.870	0.891	0.896	0.839	0.883	**0.916**
	JPEG	*0.436*	*0.599*	*0.582*	*0.736*	0.547	0.565	0.445	0.527	0.558	0.207	0.599	**0.700**
	WN	*0.938*	*0.942*	*0.932*	*0.950*	0.929	0.939	0.939	0.933	0.948	0.928	0.930	**0.950**
	GBLUR	*0.879*	*0.928*	*0.931*	*0.954*	0.910	0.911	0.918	0.941	0.926	0.935	0.910	**0.942**
	FF	*0.586*	*0.735*	*0.804*	*0.772*	0.629	0.683	0.654	0.756	0.688	0.658	0.793	**0.833**
LIVE (Phase II)	JP2K	*0.703*	*0.817*	*0.826*	*0.869*	0.834	0.842	0.727	**0.902**	0.833	0.719	0.788	0.895
	JPEG	*0.679*	*0.827*	*0.778*	*0.839*	0.755	0.839	0.694	0.728	0.840	0.613	0.745	**0.866**
	WN	*0.920*	*0.947*	*0.820*	*0.885*	0.878	0.926	0.934	0.900	**0.955**	0.907	0.807	0.952
	GBLUR	*0.836*	*0.801*	*0.950*	*0.924*	0.275	0.766	0.882	0.836	0.910	0.711	0.939	**0.942**
	FF	*0.835*	*0.830*	*0.934*	*0.918*	0.740	0.862	0.865	0.811	0.889	0.701	**0.935**	0.922
MCL	JP2K	*0.900*	*0.928*	*0.933*	*0.930*	0.919	0.927	0.904	0.897	0.902	0.919	0.862	**0.930**
	JPEG	*0.885*	*0.895*	*0.895*	*0.885*	0.803	0.823	0.888	0.898	0.857	0.882	0.888	**0.899**
	WN	*0.912*	*0.928*	*0.944*	*0.952*	0.855	0.886	0.915	0.919	0.889	0.859	0.890	**0.950**
	GBLUR	*0.912*	*0.936*	*0.952*	*0.953*	0.881	0.916	0.913	0.920	0.893	0.895	0.873	**0.970**
	SBLUR	*0.905*	*0.938*	*0.954*	*0.972*	0.891	0.944	0.918	0.930	0.920	0.954	0.859	**0.972**
	TERROR	*0.891*	*0.908*	*0.789*	*0.892*	0.871	0.842	0.874	0.890	**0.903**	0.815	0.889	**0.903**
IRCCyN/IVC	JP2K	*0.875*	*0.941*	*0.924*	*0.965*	0.820	0.947	0.910	0.857	0.823	0.885	0.802	**0.972**
	JPEG	*0.832*	*0.968*	*0.956*	*0.915*	0.842	0.916	0.855	0.863	0.775	0.782	0.896	**0.918**
	BLUR	*0.408*	*0.431*	*0.321*	*0.535*	0.556	0.496	0.470	0.417	0.459	0.467	0.321	**0.789**

column with confidence greater than 95%. Italicized entries denote that the algorithm in the row has statistically greater residual variance with the same confidence. Plain text entries denote that there is statistically no difference between the residuals of the two predictions. Also contained in Table V are the JB statistic measures of Gaussianity. Italicized JB entries denote that the residuals can be deemed Gaussian with 95% confidence. Larger values of the JB statistic denote larger deviations from Gaussianity.

As shown in Table V, 3D-MAD is statistically the best-performing algorithm on the LIVE Phase I, Phase II, and MCL-3D databases. On the IRCCyN/IVC database, 3D-MAD has statistically the same performance as Benoit *et al.*'s method, which might due to a small number of images being tested, but is still statistically better than the others. Also, note that on the LIVE Phase I, Phase II, and MCL-3D databases, 3D-MAD has more Gaussian residuals than most other algorithms as denoted by the JB statistic. However, on the IRCCyN/IVC database, 3D-MAD has a relatively higher JB statistic due to the several outliers (see Figure 5), and thus the fact that it can still achieve a lower residual variance with these outliers is noteworthy.

E. Performance on Individual Distortion Types

We also report the performance of 3D-MAD and other 2D/3D IQA algorithms on subsets of the LIVE, MCL, and IRCCyN/IVC 3D image databases corresponding to each individual distortion type. Test results of the five distortion types in LIVE, the six in MCL, and the three in IRCCyN/IVC are presented in Table VI in terms of SROCC values. The CC and RMSE values follow similar trends (see the online supplement at http://vision.okstate.edu/3dmad/). For the CC and RMSE evaluation, the same logistic transform was used and the performance for each distortion type was evaluated based on extracting the corresponding transformed scores previously computed when the whole database was considered.

TABLE VII
RUNTIME REQUIREMENTS (SECONDS/IMAGE) FOR 3D-MAD AND OTHER 3D IQA METHODS ON DIFFERENT IMAGE SIZES

	128×128	256×256	512×512	768×768	1024×1024
You [11]	0.54±0.03	1.72±0.04	7.25±0.25	14.37±0.48	28.53±0.46
Benoit [12]	0.55±0.04	1.72±0.04	7.31±0.12	14.54±0.12	28.91±0.38
Wang [56]	0.32±0.01	1.76±0.02	9.55±0.10	22.74±0.18	51.38±0.50
Ko [54]	0.03±0.00	0.06±0.00	0.18±0.00	0.34±0.00	0.61±0.02
Chen [30]	3.86±0.12	10.29±0.17	37.82±0.24	80.72±0.36	145.57±0.73
Lin [60]	0.06±0.02	0.15±0.01	0.68±0.01	1.61±0.03	2.92±0.06
Shao [62]	4.91±0.51	20.19±2.48	96.21±8.71	257.81±25.80	514.64±77.79
3D-MAD-I	0.25±0.01	0.83±0.02	4.02±0.03	9.12±0.03	17.86±0.13
3D-MAD-II	0.63±0.01	2.69±0.04	13.04±0.08	31.16±0.18	58.01±0.30
3D-MAD	0.88±0.02	3.52±0.04	17.06±0.09	40.29±0.19	75.89±0.31

Bold entries denote the best-performing 3D IQA algorithm for each distortion type on each database. Italicized entries denote the 2D IQA algorithms.

As shown in Table VI, 3D-MAD provides better predictions on most distortion types in comparison to other 3D IQA algorithms. Compared with the 2D IQA algorithm results (especially the original MAD), however, we observe that 3D-MAD does not always perform better (and can even be slightly worse) on certain distortion types (e.g., JPEG and JPEG2000 compression). This interesting observation, together with results in Tables III and IV, might indicate that the simple strategy of averaging the quality estimates for the two stereoscopic views works for evaluating degradation levels of certain distortion types applied symmetrically to each stereopair. However, this strategy has difficulty in dealing with asymmetrically distorted stimuli, and also has difficulty in bringing all quality measures on the same scale when viewing the database as a whole. Also, note that all 2D/3D IQA algorithms considered here seem to have relatively worse quality predictions on blurred images in the IRCCyN/IVC database. The reason is that these IQA algorithms are insensitive to the resampled images, which is also considered as a blur distortion in the IRCCyN/IVC database. In summary, when looking at the performance on individual distortion types, 3D-MAD is still the best overall choice to estimate stereoscopic image quality.

F. Computational Analysis

To determine the computational runtime of each stage in the 3D-MAD algorithm (denoted by 3D-MAD-I and 3D-MAD-II, respectively), we measured the runtime required to compute the qualities of 45 images with different sizes (128×128, 256×256, 512×512, 768×768, and 1024×1024). We also compared the overall computational time of 3D-MAD with other 3D IQA algorithms on these images. Note that all of these tests were performed by running unoptimized MATLAB code on a modern desktop computer (AMD Phenom II X4 965 Processor at 3.39 GHz, 4 GB RAM, Windows 7 Pro 64-bit). The results are shown in Table VII in terms of average seconds per image followed by the standard deviation.

As shown in Table VII, the second stage of 3D-MAD consumes a much longer time than the first stage. This is due to the fact that in the second stage, the disparity map of the reference stereopairs is computed and six cyclopean feature images are synthesized. Compared with other 3D IQA algorithms, 3D-MAD is faster than Chen *et al.*'s and Shao *et al.*'s approaches, but is slower than the others, which can be attributed to the more complex analysis on stereoscopic images. However, considering 3D-MAD's improved predictive performance, we believe that the time cost is justified.

V. CONCLUSION

This paper presented an algorithm to evaluate stereoscopic image quality, called 3D-MAD, which extends our previous MAD algorithm from 2D to 3D. The proposed algorithm operates via two main stages. In the first stage, the original MAD algorithm is employed to assess the quality of each stereoscopic view (left view and right view) of a stereoscopic image, and then the two obtained quality measurements are linearly combined through two averaged block-based contrast values (computed from each channel) to quantify the stereopair image quality degradation. In the second stage, disparity-compensated cyclopean feature images are synthesized from the lightness distance maps and the pixel-based contrast map by using a multipathway contrast gain-control model. Accordingly, three cyclopean feature images are generated from both the reference and distorted stereopairs, respectively. These distorted cyclopean feature images are then compared against the reference cyclopean feature images, and the computed statistical difference maps/features are employed to predict the cyclopean feature image quality degradation. Finally, quality estimates from both the stereopair images and cyclopean feature images are combined to yield an overall 3D image quality prediction. We acknowledge that there are many parameters available in the underlying models, which could be a drawback to the implementation of the proposed method. However, most of these parameters are predetermined (fixed) by the prior work, and there are only four empirical values we have selected. Experimental results on various databases demonstrated that 3D-MAD can achieve better performance than many other 2D/3D IQA algorithms.

REFERENCES

[1] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011006-1–011006-21, 2010.
[2] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
[3] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[4] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, Nov. 2013, Art. ID 905685.
[5] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imag. Sci. Technol.*, vol. 53, no. 3, pp. 30201-1–30201-14, 2009.
[6] O. J. Braddick, "Binocular single vision and perceptual processing," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 204, no. 1157, pp. 503–512, 1979.
[7] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality analysis for 3D video using 2D video quality models," *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1969–1976, Nov. 2008.
[8] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, and A. M. Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2-D video," *Electron. Lett.*, vol. 44, no. 16, pp. 963–965, 2008.

[9] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE*, vol. 6803, pp. 680305-1–680305-12, Feb. 2008.

[10] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," in *Proc. 15th Eur. Signal Process. Conf.*, Poznań, Poland, Sep. 2007, pp. 2110–2114.

[11] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. 5th Int. Workshop Video Process. Quality Metrics Consum. Electron.*, 2010, pp. 1–6.

[12] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, Jan. 2009, Art. ID 659024.

[13] R. G. Kaptein, A. Kuijsters, M. T. M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx, "Performance evaluation of 3D-TV systems," *Proc. SPIE, Image Quality Syst. Perform. V*, vol. 6808, pp. 680819-1–680819-11, Jan. 2008.

[14] L. Goldmann and T. Ebrahimi, "3D quality is more than just the sum of 2D and depth," in *Proc. IEEE Int. Workshop Hot Topics 3D*, 2010, pp. 1–2.

[15] J. M. Wolfe, "Stereopsis and binocular rivalry," *Psychol. Rev.*, vol. 93, no. 3, pp. 269–282, 1986.

[16] R. Blake and K. Boothroyd, "The precedence of binocular fusion over binocular rivalry," *Perception Psychophys.*, vol. 37, no. 2, pp. 114–124, 1985.

[17] J. M. W. Levelt, "The alternation process in binocular rivalry," *Brit. J. Psychol.*, vol. 57, nos. 3–4, pp. 225–238, 1966.

[18] R. Fox and F. Rasche, "Binocular rivalry and reciprocal inhibition," *Perception Psychophys.*, vol. 5, no. 4, pp. 215–217, 1969.

[19] M. Hollins, "The effect of contrast on the completeness of binocular rivalry suppression," *Perception Psychophys.*, vol. 27, no. 6, pp. 550–556, 1980.

[20] T. J. Andrews and D. Purves, "Similarities in normal and binocularly rivalrous viewing," *Proc. Nat. Acad. Sci. USA*, vol. 94, no. 18, pp. 9905–9908, 1997.

[21] R. P. O'Shea, R. Blake, and J. M. Wolfe, "Binocular rivalry and fusion under scotopic luminances," *Perception*, vol. 23, no. 7, pp. 771–784, 1994.

[22] V. A. Nguyen, A. W. Freeman, and D. Alais, "Increasing depth of binocular rivalry suppression along two visual pathways," *Vis. Res.*, vol. 43, no. 19, pp. 2003–2008, 2003.

[23] S. Takase, S. Yukumatsu, and K. Bingushi, "Local binocular fusion is involved in global binocular rivalry," *Vis. Res.*, vol. 48, no. 17, pp. 1798–1803, 2008.

[24] M. Vergeer and R. van Lier, "Feature-based activation and suppression during binocular rivalry," *Vis. Res.*, vol. 50, no. 8, pp. 743–749, 2010.

[25] J. Brascamp, H. Sohn, S.-H. Lee, and R. Blake, "A monocular contribution to stimulus rivalry," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 21, pp. 8337–8344, 2013.

[26] A. Bartels and N. K. Logothetis, "Binocular rivalry: A time dependence of eye and stimulus contributions," *J. Vis.*, vol. 10, no. 12, p. 3, 2010.

[27] M. A. Georgeson and S. A. Wallis, "Binocular fusion, suppression and diplopia for blurred edges," *Ophthalmic Physiol. Opt.*, vol. 34, no. 2, pp. 163–185, 2014.

[28] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 870–883, 2012.

[29] A. Maalouf and M.-C. Larabi, "CYCLOP: A stereo color image quality assessment metric," in *Proc. IEEE Int. Conf. Acoust., Speech, Singal Process.*, May 2011, pp. 1161–1164.

[30] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Process., Image Commun.*, vol. 28, no. 9, pp. 1143–1155, 2013.

[31] W. J. M. Levelt, "On binocular rivalry," Ph.D. dissertation, Nat. Defence Res. Org. TNO, Inst. Perception RVO-TNO, Soesterberg, The Netherlands, 1965.

[32] R. Blake, "A primer on binocular rivalry, including current controversies," *Brain Mind*, vol. 2, no. 1, pp. 5–38, 2001.

[33] T. J. Mueller and R. Blake, "A fresh look at the temporal dynamics of binocular rivalry," *Biol. Cybern.*, vol. 61, no. 3, pp. 223–232, 1989.

[34] C. J. H. Bossink, P. F. M. Stalmeier, and C. M. M. de Weert, "A test of Levelt's second proposition for binocular rivalry," *Vis. Res.*, vol. 33, no. 10, pp. 1413–1419, 1993.

[35] C.-B. Huang, J. Zhou, Y. Zhou, and Z.-L. Lu, "Contrast and phase combination in binocular vision," *PLoS One*, vol. 5, no. 12, p. e15075, 2010.

[36] C.-B. Huang, J. Zhou, Z.-L. Lu, and Y. Zhou, "Deficient binocular combination reveals mechanisms of anisometropic amblyopia: Signal attenuation and interocular inhibition," *J. Vis.*, vol. 11, no. 6, p. 4, 2011.

[37] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack. *Image and Video Quality Assessment Research at LIVE*. [Online]. Available: http://live.ece.utexas.edu/research/quality/

[38] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[39] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[40] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *Proc. Int. Conf. Image Process.*, vol. 3. Sep. 2003, pp. III-185–III-188.

[41] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE, Human Vis. Electron. Imag. X*, vol. 5666, pp. 149–159, Mar. 2005.

[42] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2. Sep. 1999, pp. 1150–1157.

[43] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[44] J. Yang, C. Hou, Z. Zhang Y. Zhou, and J. Guo, "Objective quality assessment method of stereo images," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss., Display 3D Video*, May 2009, pp. 1–4.

[45] *Objective Video Quality Measurement Using a Peak-Signal-to-Noise-Ratio (PSNR) Full Reference Technique*, document ANSI T1.TR.74-2001, 2001.

[46] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[47] J.-J. Hwang and H. R. Wu, "Stereo image quality assessment using visual attention and distortion predictors," *Trans. Internet Inf. Syst.*, vol. 5, no. 9, pp. 1613–1631, 2011.

[48] L. Xing, J. You, T. Ebrahimi, and A. Perkis, "A perceptual quality metric for stereoscopic crosstalk perception," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4033–4036.

[49] K. Ha and M. Kim, "A perceptual quality assessment metric using temporal complexity and disparity information for stereoscopic video," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2525–2528.

[50] Z. M. P. Sazzad, S. Yamanaka, and Y. Horita, "Continuous stereoscopic video quality evaluation," *Proc. SPIE, Stereosc. Displays Appl. XXI*, vol. 7524, pp. 75241E-1–75241E-12, Feb. 2010.

[51] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller, "Quality assessment of 3D video in rate allocation experiments," in *Proc. IEEE Int. Symp. Consum. Electron.*, Apr. 2008, pp. 1–4.

[52] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121–132, 2004.

[53] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and G. B. Akar, "Towards compound stereo-video quality metric: A specific encoder-based framework," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, 2006, pp. 218–222.

[54] H. Ko, C.-S. Kim, S. Y. Choi, and C.-C. J. Kuo, "3D image quality index using SDP-based binocular perception model," in *Proc. 11th IEEE Workshop 3D Image/Video Technol. Appl.*, Jun. 2013, pp. 1–4.

[55] S. Ryu, D. H. Kim, and K. Sohn, "Stereoscopic image quality metric based on binocular perception model," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 609–612.

[56] X. Wang, S. Kwong, and Y. Zhang, "Considering binocular spatial sensitivity in stereoscopic image quality assessment," in *Proc. IEEE Vis. Commun. Image Process.*, Nov. 2011, pp. 1–4.

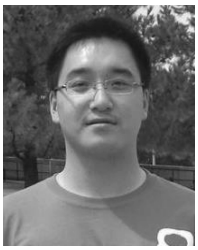
[57] F. Qi, T. Jiang, S. Ma, and D. Zhao, "Quality of experience assessment for stereoscopic images," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2012, pp. 1712–1715.

[58] R. Bensalma and M.-C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Syst. Signal Process.*, vol. 24, no. 2, pp. 281–316, 2013.

[59] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Conf. Rec. 37th Asilomar Conf. Signals, Syst., Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.

[60] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1527–1542, Apr. 2014.

[61] F. Shao, G.-Y. Jiang, M. Yu, F. Li, Z. Peng, and R. Fu, "Binocular energy response based quality assessment of stereoscopic images," *Digit. Singal Process.*, vol. 29, pp. 45–53, Jun. 2014.
[62] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.
[63] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
[64] C. Vu and D. Chandler, "Main subject detection via adaptive feature refinement," *J. Electron. Imag.*, vol. 20, no. 1, p. 013011, 2011.
[65] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Proc. 18th Int. Conf. Pattern Recognit.*, vol. 3. 2006, pp. 15–18.
[66] F. Hou, C.-B. Huang, J. Liang, Y. Zhou, and Z.-L. Lu, "Contrast gain-control in stereo depth and cyclopean contrast perception," *J. Vis.*, vol. 13, no. 8, p. 3, 2013.
[67] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2001, Art. ID 27.
[68] R. Song, H. Ko, and C. C. J. Kuo. (2014). "MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source." [Online]. Available: http://arxiv.org/abs/1405.1403
[69] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 41–54, 2006.
[70] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[71] VQEG. (Aug. 2003). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II*. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx
[72] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Electron. Lett.*, vol. 6, no. 3, pp. 255–259, 1980.



Yi Zhang received the B.S. and M.S. degrees from Northwestern Polytechnical University, Xi'an, China, in 2008 and 2011, respectively, and the Ph.D. degree from Oklahoma State University, Stillwater, OK, in 2015, all in electrical engineering. His research interests include 2D/3D image processing, machine learning, pattern recognition, and computer vision.

Damon M. Chandler received the B.S. degree in biomedical engineering from Johns Hopkins University, Baltimore, MD, in 1998, and the M.Eng., M.S., and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY, USA, in 2000, 2003, and 2005, respectively. From 2005 to 2006, he was a Post-Doctoral Research Associate with the Department of Psychology, Cornell University. From 2006 to 2015, he was on the faculty of the School of Electrical and Computer Engineering, Oklahoma State University. He is currently an Associate Professor with the Department of Electrical and Electronic Engineering, Shizuoka University, Japan. His research interests include image processing, data compression, computational vision, natural scene statistics, and visual perception.