No-Reference Quality Assessment of Natural Stereopairs

Ming-Jun Chen, Lawrence K. Cormack, and Alan C. Bovik, *Fellow, IEEE*

***Abstract*— We develop a no-reference binocular image quality assessment model that operates on static stereoscopic images. The model deploys 2D and 3D features extracted from stereopairs to assess the perceptual quality they present when viewed stereoscopically. Both symmetric- and asymmetric-distorted stereopairs are handled by accounting for binocular rivalry using a classic linear rivalry model. The NSS features are used to train a support vector machine model to predict the quality of a tested stereopair. The model is tested on the LIVE 3D Image Quality Database, which includes both symmetric- and asymmetric-distorted stereoscopic 3D images. The experimental results show that our proposed model significantly outperforms the conventional 2D full-reference QA algorithms applied to stereopairs, as well as the 3D full-reference IQA algorithms on asymmetrically distorted stereopairs.**

***Index Terms*— Binocular rivalry, 3D image quality, stereoscopic quality assessment, no-reference QA.**

I. Introduction

THE QUANTITY of digital 3D videos and images available for human consumption has increased dramatically in the last few years. According to statistics gathered by the Motion Picture Association of America (MPAA), half of all moviegoers saw at least one 3D movie in 2011, while those under 25 years old saw more than twice that number [1]. To meet this demand, the number of 3D movies has been increasing by at least 50% annually over the past few years [1], [2]. This increase has not been limited to the theatre screen: 3D television broadcasts [3] and 3D cameras, including camera phones, have become commonplace. As such, consumer 3D content can be expected to become increasingly available in the near future, presenting challenges to efforts to maintain and improve the quality of experience (QoE) of visual content.

Understanding how to monitor the integrity of 2D and 3D visual signals throughout computer networks has become a critical question. Being able to provide visual quality assurance via the ability to automatically assess the quality of visual media delivered to the client is both demanding and increasingly urgent. Thus, the development of objective visual quality assessment models of images and videos has been a busy and fruitful area of work [4]. However, while great advances have been made on modeling regular (non-stereoscopic) image and video quality [5], [6], progress on the question of 3D image quality has been limited [7].

Research on visual quality assessment can be divided into three categories based on the amount of information which is accessible to the algorithm [4]. *Full-reference* (FR) models require the original pristine content to be able to assess, for comparison, the quality of the tested content. *Reduced-reference* (RR) approaches operate under the assumption that some small fraction (at least) of information about the original content is available. This fractional information could range from a few parameters extracted from the pristine content to extra side data (such as a watermark) imposed on the tested content. Finally, *no-reference* (NR) quality assessment algorithms operate on the tested content without any information extracted from the corresponding pristine content. Since pristine reference versions of visual signals transmitted over networks are rarely available, NR QA algorithms are potentially much more feasible solutions. Indeed, 2D NR QA algorithms [5], [8]–[11] have been developed that deliver competitive performance relative to 2D FR QA algorithms. This is not the case with stereoscopic 3D NR QA models.

The problem of stereoscopic 3D image quality assessment is much more complex than that of 2D quality assessment. A number of important issues arise with the additional dimension of 3D content (depth or disparity). First, an observer may experience binocular rivalry[1] while stereoscopically viewing 3D content, which may affect the perceived 3D quality [12]–[15]. Further, the perceived quality of the depth sensation (e.g., as a function of the rendering algorithm [16]) likely involves interactions between depth quality, 2D quality, and 3D quality [17], [18]. Finally, the important factors of visual discomfort and fatigue arising from incorrect stereography [19], [20] can negatively affect the experience of viewing a stereoscopic image. From among these various factors,we attempt to isolate and focus on predicting the 3D perceptual quality of stereopairs that may have been subjected to distortion but without considering issues related to visual comfort.

Even without considering visual comfort, it has proved quite challenging to develop 3D FR QA models. Moorthy *et al.* [7] tested a number of 2D and 3D QA algorithms against the distorted 3D content of the Phase I dataset of the LIVE 3D Image Quality Database, and found that the tested 3D FR QA algorithms (those that utilized depth or disparity in

Manuscript received September 26, 2012; revised March 15, 2013 and May 19, 2013; accepted May 29, 2013. Date of publication June 10, 2013; date of current version July 9, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Guest Editor Anthony Vetro.

M.-J. Chen and A. C. Bovik are with the Department of Electrical and Computer Engineering, University of Texas, Austin, TX 78712 USA (e-mail: mjchen@utexas.edu; bovik@ece.utexas.edu).

L. K. Cormack is with Center for Perceptual Systems and the Department of Psychology, University of Texas, Austin, TX 78712 USA (e-mail: cormack@psy.utexas.edu).

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.2013.2267393

[1]We use the term in a broader sense than is used by people who study biological vision

same manner) did not perform better than high performance 2D FR QA algorithms applied to the stereo imagepair when cast against recorded human opinions on the stereoscopically viewed 3D images.

The perceptual effects of distortions arising from compression channel errors, noise, and camera artifacts are becoming quite important as 3D capture, transmission, and display technology have begun to penetrate the wireless and mobile markets. We do allow for the possibility of binocular rivalry arising from distortion asymmetries. The result of our modeling effort is a 3D NR IQA algorithm that predicts the quality of stereo images whether the distortion is symmetric or not (rivalrous). The algorithm extracts both 2D and 3D natural statistical features from a stereopair to be quality assessed. These features are used to first classify a distorted stereopair as either symmetrically or asymmetrically distorted. Then the same set of features is used to evaluate the quality of the stereopair.

This paper is organized as follows. Section II reviews previous 3D image QA models that only focus on predicting picture quality of 3D images. In Section III, we develop our model, describing in detail the feature extraction and feature-based quality prediction mechanisms. Section IV describes the LIVE 3D Image Quality Database and details the experimental results. Finally, Section V concludes the paper with a discussion of ideas for future work.

II. Previous Work

While just three categories (FR, RR, and NR) are commonly used to distinguish 2D QA models, we will also further divide 3D QA models into two classes that may come from any of the usual three categories. The first class (Class 1) [21]–[23] are 2D-based 3D QA models which do not utilize computed or otherwise measured depth/disparity information from the stereopairs. Among Class 1 models, the methods in [21], [22] conduct 2D FR QA on the left and right views independently, then combine (by various means) the two scores into predicted 3D quality scores. Gorley, *et al.* [23] compute quality scores on matched feature points delivered by SIFT [24] and RANSAC [25] applied to the two views.

The second class (Class 2) of models include some kind of disparity information in the overall 3D QA process. Among Class 2 3D QA algorithms, Benoit, *et al.* [26] proposed a FR 3D QA algorithm that computes quality scores between the left reference and left distorted view, the right reference and right distorted view, and between computed reference and distorted disparity maps. The quality scores are computed using the 2D FR QA indices C4 [27] and SSIM [28], then combined to produce a final predicted 3D QA score. Their results suggest that disparity information might improve 3D QA performance when SSIM is applied to the disparity data, but they also claimed that the 2D C4 algorithm performs better than applying SSIM on both stereo image pair and the disparity map. They observe that the disparity estimation algorithm can affect the performance of 3D QA. You, *et al.* [29] further extended the idea of predicting the 3D quality of a stereopair by applying 2D QA algorithms on the stereopair and on its disparity map. They applied a large pool of FR 2D QA algorithms on the stereopairs and disparity maps, and concluded that applying SSIM on stereopairs and mean-absolute-difference (MAD) on the estimated disparity map yields good performance in predicting the 3D quality of stereo images. Unlike the work reported by Benoit, *et al.*, their SSIM-based 3D QA algorithm significantly outperformed all 2D FR QA models on their dataset. Using a similar framework, Zhu, *et al.* [30] proposed a 3D QA model based on a 2D QA algorithm. Similarly, Yang, *et al.* [31] proposed a FR 3D QA algorithm based on the average PSNR of the two images in the stereopair and the absolute difference between the left and right view. Their algorithm therefore did not deploy a stereo matching stage to compute depth or disparity. However, none of these studies benefited by a large scale study of algorithm QA performance against human opinion. The study in [7], which employed thirty-two subjects who viewed two hundred distorted stereopairs affected by five different distortions, found the remarkable result that none of the 3D QA algorithms developed at the time of the study delivered QA prediction performance better than or in most cases, even as good as, 2D QA algorithms applied to the individual stereo image pairs.

Maalouf, *et al.* [32] proposed to perform the task of 3D QA on a Cyclopean image, which they define as the average of the left image and the disparity-compensated right image. Of course, the Cyclopean image is not an average of the left and right views but instead, is a 3D percept consisting of luminance and color patterns superimposed on a 3D images [33], [34]. Bensalma, *et al.* [35] proposed a 3D QA algorithm that measures the difference of binocular energy between the reference and tested stereopairs, and thus considers the potential influence of binocularity on perceived 3D quality. Chen, *et al.* [36] proposed a 3D FR QA framework that models the influence of binocular effects, and claim that modelling binocular effect can yield better performance when predicting the subjective 3D quality of stereopairs. Further positive developments would likely benefit from a better understanding of the relationship between perceived distortions and binocular rivalry.

Only a small amount of work has targeted the development of RR and NR 3D QA models. An RR 3D QA algorithm was proposed by Hewage, *et al.* [37]. In their model, edges are computed from the depth map, then the PSNR between the reference and test edge maps are analyzed to predict 3D quality. Akhter, *et al.* [38] proposed a NR 3D QA algorithm which extracts features from stereopairs and an estimated disparity map. A logistic regression model is used to predict 3D quality scores from these features.[2]

III. 3D NR QA Model Design

A flowchart of our proposed model is shown in Fig. 1. Since disparity is estimated and used to conduct QA, it is a Class 2

[2]There are other NR 3D QA algorithms that specifically deal with depth-image-based-rendering (DIBR) generated 3D images/videos, but here we only discuss the algorithms that operate on natural stereo content.

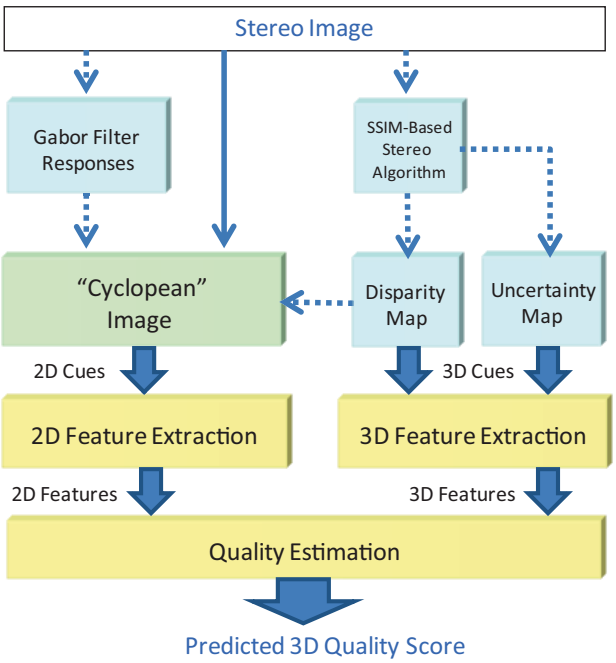

Fig. 1. Flowchart of proposed 3D NR QA model.

3D NR IQA model. Given a stereo imagepair, an estimated disparity map is generated by a SSIM-based stereo algorithm, while a set of multi-scale Gabor filter responses are generated on the stereo images using a filter bank. A Cyclopean Image is then synthesized from the stereo image pair, the estimated disparity map, and the Gabor filter responses. 2D features are then extracted from the synthesized Cyclopean Image, while 3D features are independently extracted from the estimated disparity map and an uncertainty map that is also produced by the stereo matching algorithm. Finally, all of the extracted 2D and 3D features are fed into a quality estimation module which predicts the perceived 3D quality of each tested stereo imagepair.

An extension of the linear model proposed by Levelt [34] is used to synthesize the Cyclopean Image from a stereo image pair. First, a disparity map is estimated from a test stereo pair using a very simple SSIM-based dense stereo matching algorithm. The algorithm operates by search for disparities yield the best SSIM match between left and right image patches, with ties broken by selecting the lower disparity solution. This estimated disparity map is then used to create a disparity-compensated right view image. The Gabor filter responses are then extracted from the left view image and the disparity-compensated right view image. Finally, the Cyclopean image is calculated as a weighted summation of the left and disparity-compensated right views, where the weights are computed from the Gabor filter responses. The details of this process can be found in [36]. Since the contribution of this work is the method of selecting and extracting features and the way they are used to conduct NR 3D QA, we only focus on explaining these later parts.

A. 2D Feature Extraction

Research on natural scene statistics (NSS) has clearly demonstrated that images of natural scenes belong to a small set of the space of all possible signals and that they obey predictable statistical laws [39]. Successful 2D NR QA algorithms [5], [8], [9] based on the statistics of natural scenes (and the fact that human perception has adapted to these statistics over the eons) have achieved comparable QA prediction performance as high performance FR QA models [28], [40]. Although images of real-world scenes may vary greatly in their luminance and color distributions, by pre-processing images in biologically relevant way, e.g., by processes of predictive coding [41] and divisive normalization [42], yields transformed images obeying a regular parametric statistical model [39], [43]. Ruderman [39] showed that images processed via a simple local mean subtraction and divisive variance normalization produces nearly decorrelated luminances obeying a Gaussian-like distribution. This model closely mimics the classical front-end center surround model with adaptive gain control. Using these kind of NSS features, Mittal *et al.* [8] developed a highly competitive 2D NR IQA model, called BRISQUE.

We apply similar pre-processing on the synthesized Cyclopean Image:

$$M(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{1}$$

where i, j are spatial indices, μ and σ are the local sample mean and weighted standard deviation computed by a local window, and C is a constant that ensures stability. In our implementation, we use an 11×11 Gaussian weighting matrix with a sigma of 3.67 pixels to compute μ and σ, and fixed $C = 0.01$.

Following [8] we model the coefficients in equation (1) of the possibly distorted cyclopean image as following a generalized Gaussian distribution(GGD):

$$f_x(x; \mu, \sigma^2, \gamma) = ae^{-[b|x-\mu|]^\gamma} \tag{2}$$

where μ, σ^2 and γ are the mean, variance, and shape-parameter of the distribution,

$$a = \frac{b\gamma}{2\Gamma(1/\gamma)} \tag{3}$$

$$b = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}} \tag{4}$$

and $\Gamma(\cdot)$ is the gamma function:

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt, \; x > 0. \tag{5}$$

The parameters (σ and γ) are estimated using the method used in [44]. The skewness and the kurtosis of these coefficients are also estimated.

B. 3D Feature Extraction

Features based on natural scene statistics have been shown to be effective, robust tools for predicting the quality of natural images. The success of NSS-based features is built on the fact that pristine natural scenes tend to follow certain regular statistical laws. Towards further following this philosophy when modeling 3D images, we also build into our QA model features derived from 3D natural scene statistic models.

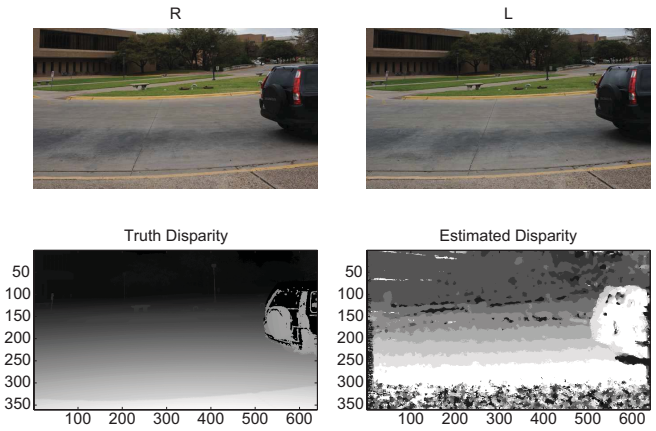

Fig. 2. A stereopair with ground truth disparity and estimated disparity. Top left: Right view of the stereo image. Top right: Left view of the stereo image. Bottom left: Ground truth disparity. Bottom right: Estimated disparity.

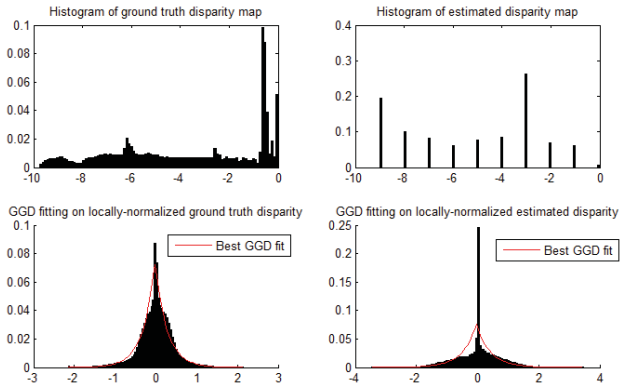

Fig. 3. Top left: Histogram of ground truth disparity map. Top right: Histogram of the estimated disparity map. Bottom left: Histogram of the local-normalized ground truth disparity map GGD fit overlaid. Bottom right: Histogram of the estimated disparity map with GGD fit overlaid.

Fig. 4. A undistorted natural stereopair (free-fuse to view).

As compared to the extensive body of literature on 2D natural luminance statistics, studies on the statistics of disparity and depth have been quite limited. One possible explanation for this dearth is that acquiring accurate disparity data is a much more difficult task than capturing 2D imaging data.

There has also been only a small amount of reported work on modeling of natural 3D statistics. Huang [45] first studied the statistics of range images, using range data measured by a laser range-finder. They begin with the assumption that natural range maps follow the random collage model, i.e. that a range image can be partitioned into disjoint smooth surfaces separated by discontinuities. Yang and Purves [46] further studied the statistics of range data, which are acquired by a laser scanner, and found that their range data is quite rough and has a property of anisotropy. If a viewing model is defined, the range data can be transformed into disparity data. To study the statistics of disparity, Hibbard [47] and Liu *et al.* [48] model the fixation distance of a virtual subject. An essential difference between their work is that ground truth range data (measured by a laser scanner) is analyzed in Liu's work, while Hibbard used a random collage sphere model to synthesize range data. Both groups found that the distribution of disparity follows a Laplacian shape. In the following, we discuss how to bring to bear 2D and 3D NSS models on the 3D QA problem.

To conduct the task of no-reference quality assessment on a stereo image pair, it is assumed that only the stereopair is available, without any reference data, including ground truth disparity. Thus, the only accessible 3D feature is estimated disparity from a stereo matching algorithm. Here, we use a simple SSIM-based stereo matching algorithm to estimate a disparity map. Therefore, it is worth discussing the difference between ground truth disparity and estimated disparity. Figure 2 shows a stereopair with ground truth disparity and estimated disparity map. This stereopair was captured using a parallel-camera set-up with a laser scanner that captures ground truth range data. The ground truth disparity map is directly converted from the range data since the capture model is known. In Fig. 2, one can clearly see that there are many estimated errors, especially towards the bottom sections of the image. The errors are produced by the complex, repetitive texture of the sidewalk, which the simple low-complexity stereo-algorithm doesn't handle well. Moreover, the slanting foreground surface plane is smoothly captured in the ground truth disparity map while the estimated map shows a ladder-like appearance map due to the integer pixel precision of the stereo algorithm. Figure 3 depicts the histogram of a ground truth and an estimated disparity map both before and after local mean removal and divisive normalization as in (1). The top left and top right of Fig. 3, suggest that no known model distribution could be used to consistently fit them. However, following the normalization process, the ground truth disparity distribution takes a Gaussian-like shape, while the estimated disparity distribution is much more peaky and heavily tailed. However, both are zero-mean symmetric and can be modelled as following a GGD. As before, we take the following as features: GGD parameters, standard deviation, skewness, and kurtosis.

These 3D NSS features can be effectively used in the process of distinguishing a pristine stereo image pair from a distorted version of it. We used five common distortion types to impair the stereo image data: white noise, blur, JPEG compression, JPEG2000 (JP2K) compression, and a Fast-Fading (FF) model based on the Rayleigh fading channel. Figure 5 shows histograms of the estimated disparities of stereo images distorted by these models. The left plot in figure 5 shows the distributions of symmetrically distorted stereopairs, while the right plot shows the distributions of asymmetrically distorted stereopairs. Different types of distortions were applied on an undistorted natural stereopair (Figure 4) to demonstrate how the statistics of distinctly

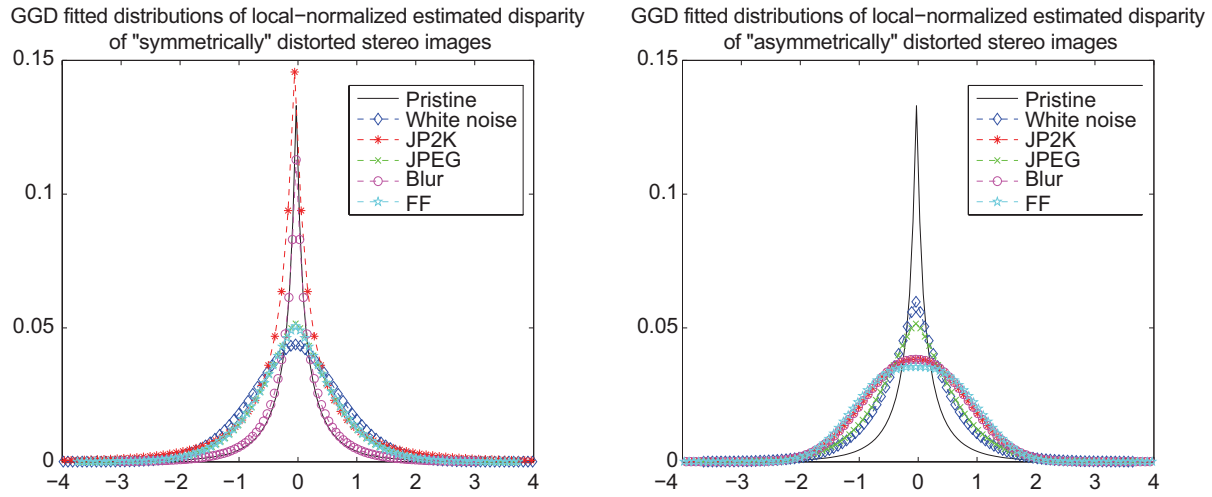

Fig. 5. Left: Fitted disparity distributions for a natural undistorted stereopair and its "symmetrically" distorted versions. Right: Fitted disparity distributions for a natural undistorted stereopair and its "asymmetrically" distorted versions.

distorted stereopairs tend to differ from one another. It may be seen that symmetrical distortions create peakier distributions whereas asymmetrical distortions result in Gaussian-like distributions. Different distortion types cause feature distributions that have different shape (kurtosis) and spread.

Other than the estimated disparity, the uncertainty produced by the SSIM-based stereo matching algorithm is a useful feature for the task of 3D NR QA. The uncertainty is defined as

$$Uncertainty(l,r) = 1 - \frac{(2\mu_l\mu_r + C_1)(2\sigma_{lr} + C_2)}{(\mu_l^2 + \mu_r^2 + C_1)(\sigma_l^2 + \sigma_r^2 + C_2)} \tag{6}$$

where l is the left-view image and r is the disparity-compensated right-view image of a stereopair. The uncertainty reflects the degree of similarity (or lack thereof) between the corresponding pixels of a stereopair. We have observed that the histograms of noise-free natural stereopairs captured using a paralleled-camera setting present a very positive skew distribution. This may be understood by observing that the stereo-matching algorithm generally finds good matches (low-uncertainty) at most places, while relatively rare occluded or ambiguous flat or textured areas may cause sparse errors in the results of the stereo matching algorithm (high-uncertainty), contributing weight to the tail of the uncertainty distribution. Figure 6 demonstrates this observation. The bottom right plate of Fig. 6 shows that most regions of the image have a low uncertainty, while higher uncertainty values are observed around the sky and trees. To model this observation, we fit a log-normal distribution to the histogram of the uncertainty map. The probability density function of a log-normal distribution is defined as

$$f_X(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp - \frac{(\ln x - \mu)^2}{2\sigma^2} \tag{7}$$

where μ is the location parameter and σ is the scale parameter. A maximum likelihood method is used to estimated μ and σ for a given histogram of uncertainties.

The histogram of uncertainty also varies as a stereopair is distorted. Figure 7 shows the uncertainty distribution of

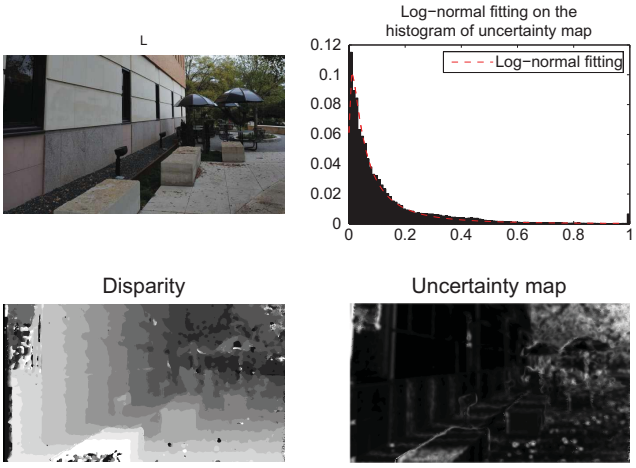

Fig. 6. Top left: Left view of a stereopair. Top right: Histogram of the uncertainty map and the best log-normal fit. Bottom left: The estimated disparity map. Bottom right: The uncertainty map produced by the stereo matching algorithm.

a natural stereopair distorted by white noise, blur, FF, JPEG compression and JP2K compression. As depicted in Fig. 7, the uncertainty distribution predictably changes with distortion type and the way a stereopair is distorted (symmetrically or asymmetrically). For symmetrically distorted stereopairs, a Gaussian blur distortion, JP2K compression or packet loss in both views tends to suppress details in both images in the stereopair, so uncertainties in the disparity estimation are reduced, yielding a peakier distribution of uncertainties. Conversely, white noise and JPEG distortion increase the uncertainty of stereo matching and reduce the peakiness of the uncertainty distribution. For asymmetrically distorted stereopairs, the uncertainty generally increases because of the unmatched distortions, except for Gaussian blur distortion.

To summarize, the 3D features used for 3D NR QA prediction are the GGD fitting parameters (mu, $sigma$), the standard deviation, skewness, and kurtosis of the local-normalized estimated disparity map, and the best-fit log-normal parameters (μ and σ), skewness, and kurtosis of the uncertainty map.

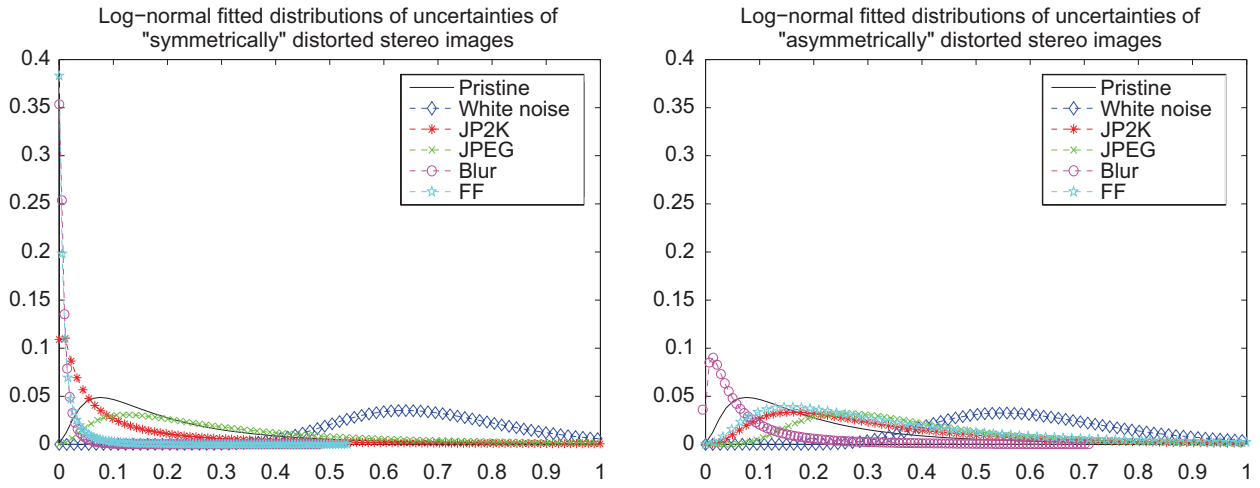

Fig. 7. Left: Fitted uncertainty distributions for a natural undistorted stereopair and its "symmetrically" distorted versions. Right: Fitted uncertainty distributions for a natural undistorted stereopair and its "asymmetrically" distorted versions.

By showing the fitted histograms of a natural stereopair and distorted versions of it, we visually demonstrate how these features tend to vary, and they can be used as a source of features that predict perceived 3D quality. These observations are supported by the results obtained on independent training and testing datasets as explained in the experimental section.

C. Quality Estimation

A two-stage QA framework is used to predict the quality of a test stereopair. This follows the framework introduced in [49] and elaborated in the 2D IQA DIIVINE index [5]. In their model, a probabilistic support vector classifier is applied first to decide the most likely distortion type afflicting the stereopair. A support vector regressor (SVR) is then used to assess the perceptual distortion severity. However, unlike DIIVINE, the classifier in our 3D NR IQA model is designed to decide whether a stereo pair is symmetrically or asymmetrically distorted, without predicting the distortion type. This is important since asymmetrically distorted stereopairs may create binocularly rivalrous 3D experiences, and may yield different extracted 3D features than symmetrically distorted stereopairs. In the human study on distorted stereopairs that we conducted [15], we found that the perceived quality of a asymmetrically distorted stereopair is not accurately predicted by the simple average quality of the stereo views, although the quality of symmetrically distorted stereopairs might be accurately predicted in this manner. The same feature vector is used for classification and regression. After the classification process is complete, the predicted quality score is computed as the dot product of the distortion probability vector and the vector of symmetric/asymmetric quality scores.

IV. Experimental Results

We utilized the LIVE 3D Image Quality Database to verify the performance of our proposed 3D NR IQA model. Although

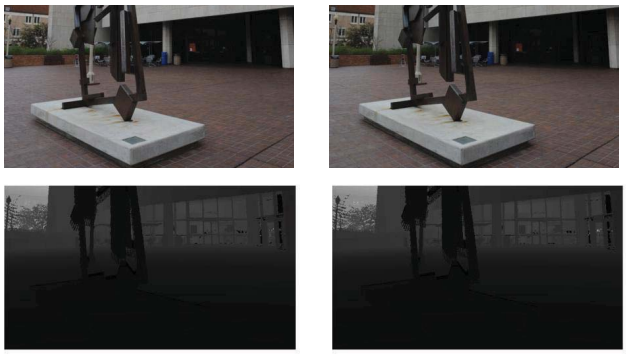

Fig. 8. A stereo image (free-fuse the left and right images) and ground truth disparity maps.

TABLE I
Range of Parameter Values for Distortion Simulation

Distortion	Control Parameter	Phase I Range	Phase II Range
WN	Variance of Gaussian	[0.01 0.15]	[0.001 0.5]
Blur	Variance of Gaussian	[0.01 20]	[0.5 30]
JP2K	Bit-rate	[0.05 3.15]	[0.04 0.5]
JPEG	Quality parameter	[10 50]	[8 50]
FF	Channel signal-to-noise ratio	[12 20]	[15 30]

part of this database is publicly available [7] (Phase I, consisting of symmetric distortions), a second phase has only recently been created.

A. LIVE 3D Image Quality Database

This database was constructed in two phases (phase I contains symmetrically distorted stimuli while phase II has both symmetrically and asymmetrically distorted stimuli). Phase I and phase II are actually different and complementary datasets. Phase I [7] has 20 pristine stereopairs and 365 distorted stereopairs, while phase II has 8 pristine sterepairs and 360 distorted stereopairs. The details of the dataset and of the human studies that were conducted on them to subjectively annotate the stimuli are described in the following.

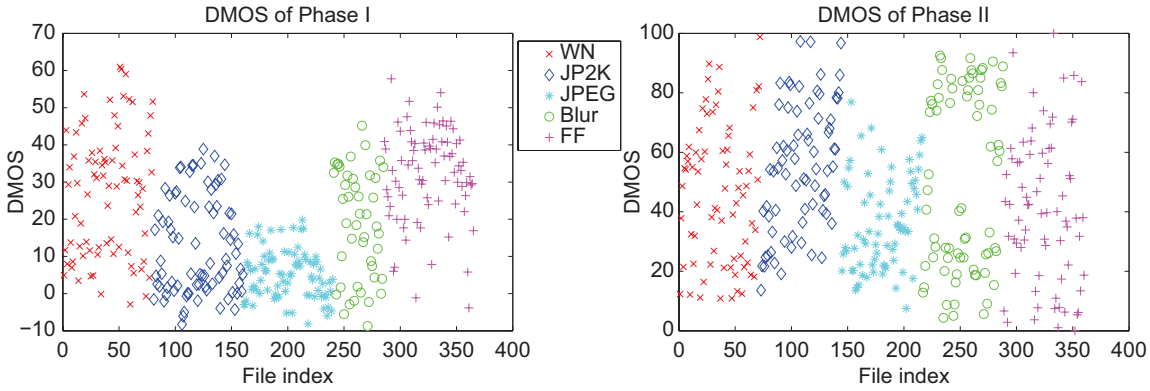

Fig. 9. Left: DMOS of LIVE 3D Image Quality Database Phase I. Right: DMOS of LIVE 3D Image Quality Database Phase II.

1) Source Images: The pristine stereo images used in both phases are stereo images co-registered with range data measured by a high-performance range scanner (RIEGL VZ-400 [50]) obtained by a Nikon D700 digital camera. The stereo image pairs were shot using a 65 mm baseline. The sizes of the images are 640×360 pixels. Figure 8 shows a stereopair and its associated ground truth depth map. For further details on the data acquisition, see [7].

2) Participants: In both phases, each subject reported normal or corrected normal vision and no acuity or color test was deemed necessary. However, a Randot stereo test was used to pre-screen participants for normal stereo vision in phase II. Phase I utilized thirty-two participators with a male-majority population. In phase II, six females and twenty-seven males participated in the experiment, aged between 22 and 42 years.

3) Display Setting: Phase I was conducted with a iZ3D 22" 3D monitor with passive polarized 3D glasses, while phase II was conducted using a Panasonic 58" 3D TV (TC-P58VT25) with active shutter glasses. The viewing distance was four times the screen height in both cases.

4) Stimuli: Both phases used five types of distortions: compression using the JPEG and JPEG2000 compression standards, additive white Gaussian noise, Gaussian blur and a fast-fading model based on the Rayleigh fading channel. The degradation of stimuli was varied by controlling parameters within a pre-defined range, as reported in Table I. The ranges of control parameters were decided beforehand to ensure that the distortions varied from almost invisible to severely distorted with a good overall perceptual separation between distortion levels throughout. Due to the different viewing environments, the range of distortions are also different in the two experimental phases.

The phase I dataset contains only symmetrically distorted stereo images (80 each for JP2K, JPEG, WN, and FF; 45 for Blur) while the phase II dataset had both symmetrically and asymmetrically distorted stereo images (72 images for each distortion type). The 'symmetrically' distorted stereopair means that the same 'amount' of distortion was created for the left and right image, while the 'asymmetrically' distorted stereopair has a different 'amount' of distortion in the two views. In the phase II dataset, for each distortion type, every reference stereopair was processed to create three symmetric distorted stereopairs and six asymmetric distorted stereopairs.

5) Procedure: A single stimulus continuous quality scale (SSCQS) [51] experiment with hidden reference was conducted in both phases. Both studies used continuous scales labelled by equally spaced adjective terms: bad, poor, fair, good, and excellent, i.e. a Likert scale. Both studies were divided into 2 sessions; each of less than 30 minutes to minimize subject fatigue. A training session was also conducted before the beginning of each study to help familiarize participants with the GUI.

6) Subjective Quality Scores: Difference opinion scores (DOS) were obtaining by subtracting the ratings that the subject gave each reference stimuli from the ratings that the subject gave to the corresponding test distorted stimuli. The remaining subjective scores were then normalized to Z-scores, then averaged across subjects to produce difference mean opinion scores (DMOS). A subject rejection process suggested by [51] was performed on the phase II dataset and two out of thirty-three were rejected. Figure 9 shows the distribution of DMOS of the database. The DMOS distributions of phase I and phase II are quite different. In the phase I dataset, the DMOS given to WN and FF distorted stimuli varied from -10 to 60, the DMOS given to JP2K and Blur distorted stimuli have a range between -10 and 40, and the DMOS given to JPEG have a significantly narrower range from -10 to 20 indicating less perceptual distortion overall and smaller differences in perceived severity. Similarly, JP2K and Blur was generally less visible than WN and FF in the phase I dataset. However, in the phase II dataset, only the JPEG distorted stimuli were less visible than other distortion types, while WN, JP2K, Blur, and FF distortions were generally rated as falling within a similar quality range. The DMOS scores of both symmetric and asymmetric stimuli are plotted in Fig. 10. From the plot, it is apparent that the different DMOS ranges were not caused by the symmetry (or lack of) of the distortion.

B. Classification Accuracy

The first layer of our proposed 3D NR IQA model is the symmetric vs. asymmetric distortion classifier. We used the LIBSVM package [52] to perform classification. To assess the performance of the classifier, we performed 1000 iterations of the train-test process. At each iteration, we randomly picked

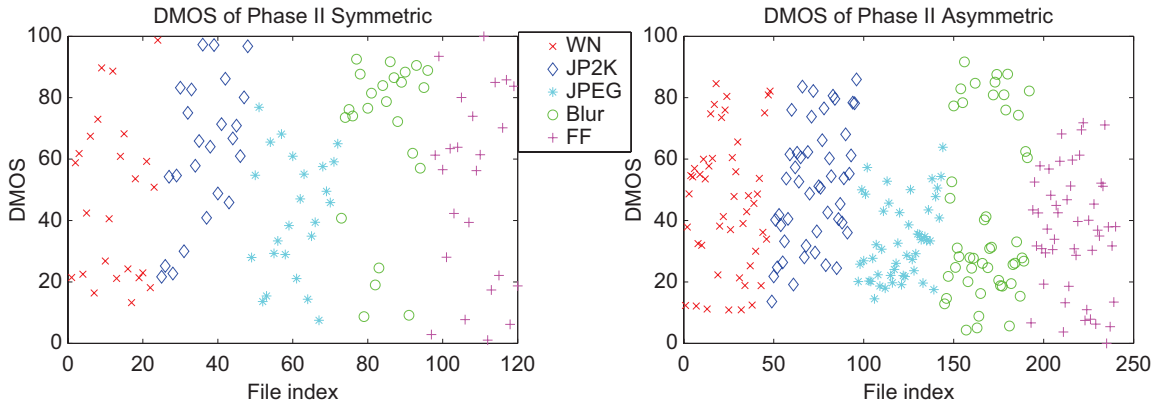

Fig. 10. Left: DMOS of Phase II symmetric distorted stimuli. Right: DMOS of Phase II asymmetric distorted stimuli.

TABLE II
COMPARISON OF 2D IQA ALGORITHMS: SROCC AGAINST DMOS ON THE LIVE PHASE I 3D IQA DATASET

	WN	JP2K	JPEG	Blur	FF	All
PSNR	0.932	0.799	0.121	0.902	0.587	0.834
SSIM	0.938	0.858	0.436	0.879	0.586	0.876
MS-SSIM	0.942	0.892	0.613	0.926	0.723	0.926
BRISQUE	0.940	0.812	0.569	0.860	0.784	0.901
Our Model	0.919	0.863	0.617	0.878	0.652	0.891

Italicized algorithms are NR IQA algorithms, all others are FR IQA algorithms.

TABLE III
COMPARISON OF 2D IQA ALGORITHMS: LCC AGAINST DMOS ON THE LIVE PHASE I 3D IQA DATASET

	WN	JP2K	JPEG	Blur	FF	All
PSNR	0.935	0.785	0.219	0.916	0.703	0.834
SSIM	0.939	0.865	0.487	0.919	0.721	0.872
MS-SSIM	0.940	0.919	0.686	0.944	0.802	0.926
BRISQUE	0.941	0.847	0.615	0.926	0.853	0.910
Our Model	0.917	0.907	0.695	0.917	0.735	0.895

Italicized algorithms are NR IQA algorithms, all others are FR IQA algorithms.

80% of the dataset as training data and the remaining 20% to test. The mean classification was 82.07% with standard deviation 2.88.

C. Performance

1) Phase I dataset: Since the proposed algorithm algorithm requires training, 1000 iterations of the train-test process was used. At each iteration, the phase I dataset was randomly divided into 80% training and 20% test across 1000 iterations. The performance was measured using Spearman's Rank Ordered Correlation Coefficient (SROCC), (Pearson's) linear correlation coefficient (LCC), the root-mean-squared error (RMSE) and outlier ratio (OR)[3] between the predicted scores and the DMOS. LCC, RMSE and OR were computed after logistic regression as described in [53]. Higher SROCC and LCC values indicate good correlation (monotonicity and accuracy) with human quality judgments, while lower values of RMSE and OR indicate better performance.

We compared the performance of our 3D NR IQA model with several 2D FR and NR IQA (Class 1) models: PSNR, SSIM [28], MS-SSIM [54], and BRISQUE [8]. SSIM and MS-SSIM are FR IQA algorithms, while BRISQUE is a high performance NR QA algorithm. For all 2D QA algorithms, the predicted quality of a stereopair is taken to be the average quality predicted from the left and right views. The performance numbers are shown in Tables II–IV. Our proposed model performs as well as BRISQUE, but a little less well than MS-SSIM. As we found in the human study [15], depth seems to have little influence on the perceived picture quality

TABLE IV
COMPARISON OF 2D IQA ALGORITHMS: RMSE AGAINST DMOS ON THE LIVE PHASE I 3D IQA DATASET

	WN	JP2K	JPEG	Blur	FF	All
PSNR	5.896	8.022	6.381	5.820	8.843	9.036
SSIM	5.740	6.497	5.712	5.715	8.613	8.013
MS-SSIM	5.672	5.111	4.758	4.794	7.419	6.181
BRISQUE	5.640	6.894	5.158	5.473	6.491	6.793
Our Model	6.433	5.402	4.523	5.898	8.322	7.247

Italicized algorithms are NR IQA algorithms, all others are FR IQA algorithms.

TABLE V
COMPARISON OF 3D IQA MODELS: SROCC AGAINST DMOS ON THE PHASE I DATASET

	WN	JP2K	JPEG	Blur	FF	All
Benoit [26]	0.930	0.910	0.603	0.931	0.699	0.899
You [29]	0.940	0.860	0.439	0.882	0.588	0.878
Gorley [23]	0.741	0.015	0.569	0.750	0.366	0.142
Cyclopean MS-SSIM [36]	0.948	0.888	0.53	0.925	0.707	0.916
Hewage [37]	0.940	0.856	0.500	0.690	0.545	0.814
Akhter [38]	0.914	0.866	0.675	0.555	0.640	0.383
Our Model	0.919	0.863	0.617	0.878	0.652	0.891

Italicized algorithms are NR IQA algorithms, all others are RR or FR IQA algorithms.

[3]Because the standard deviations of DMOS scores of the phase I dataset are not available, we didn't report OR numbers on the phase I dataset.

TABLE VI
COMPARISON OF 3D IQA MODELS: LCC AGAINST DMOS ON THE PHASE I DATASET

	WN	JP2K	JPEG	Blur	FF	All
Benoit [26]	0.925	0.939	0.640	0.948	0.747	0.902
You [29]	0.941	0.877	0.487	0.919	0.730	0.881
Gorley [23]	0.796	0.485	0.312	0.852	0.364	0.451
Cyclopean MS-SSIM [36]	0.942	0.912	0.603	0.942	0.776	0.917
Hewage [37]	0.895	0.904	0.530	0.798	0.669	0.830
Akhter [38]	0.904	0.905	0.729	0.617	0.503	0.626
Our Model	0.917	0.907	0.695	0.917	0.735	0.895

Italicized algorithms are NR IQA algorithms, all others are RR or FR IQA algorithms.

TABLE VII
COMPARISON OF 3D IQA MODELS: RMSE AGAINST DMOS ON THE PHASE I DATASET

	WN	JP2K	JPEG	Blur	FF	All
Benoit [26]	6.307	4.426	5.022	4.571	8.257	7.061
You [29]	5.621	6.206	5.709	5.679	8.492	7.746
Gorley [23]	10.197	11.323	6.211	7.562	11.569	14.635
Cyclopean MS-SSIM [36]	5.581	5.320	5.216	4.822	7.837	6.533
Hewage [37]	7.405	5.530	5.543	8.748	9.226	9.139
Akhter [38]	7.092	5.483	4.273	11.387	9.332	14.827
Our Model	6.433	5.402	4.523	5.898	8.322	7.247

Italicized algorithms are NR IQA algorithms, all others are RR or FR IQA algorithms.

TABLE VIII
COMPARISON OF 2D IQA MODELS: SROCC AGAINST DMOS ON THE PHASE II DATASET

	WN	JP2K	JPEG	Blur	FF	All
PSNR	0.919	0.597	0.491	0.690	0.730	0.665
SSIM	0.922	0.704	0.678	0.838	0.834	0.792
MS-SSIM	0.946	0.798	0.847	0.801	0.833	0.777
BRISQUE	0.846	0.593	0.769	0.862	0.935	0.770
Our Model	0.950	0.867	0.867	0.900	0.933	0.880

Italicized algorithms are NR IQA algorithms, others are RR or FR IQA algorithms.

TABLE IX
COMPARISON OF 2D IQA MODELS: LCC AGAINST DMOS ON THE PHASE II DATASET

	WN	JP2K	JPEG	Blur	FF	All
PSNR	0.919	0.597	0.491	0.690	0.730	0.665
SSIM	0.922	0.704	0.678	0.838	0.834	0.792
MS-SSIM	0.946	0.798	0.847	0.801	0.833	0.777
BRISQUE	0.846	0.593	0.769	0.862	0.935	0.770
Our Model	0.950	0.867	0.867	0.900	0.933	0.880

Italicized algorithms are NR IQA algorithms, others are RR or FR IQA algorithms.

of distorted stereopairs (we have found no depth masking phenomena that affects perceived picture quality), although other aspects of binocular fusion, such as introduced rivalries, are important. Since there is little or no binocular rivalry from the distortions present in the stimuli in the Phase I dataset, we do not expect any significant improvement in performance under our 3D NR IQA model. In addition, the performance of our model is significantly lower on the JPEG distorted stimuli as compared to other distorted stimuli, but the same holds far all of the other QA algorithms. As shown in Fig. 9, the JPEG distorted stimuli represent a more difficult challenge because their qualities are less perceptually separated.

We also studied the relative performance of 3D IQA (Class 2) algorithms. The FR 3D IQA algorithms compared include a PSNR-based 3D stereo IQA algorithm proposed by Gorley and Holliman [23], a SSIM-based stereo IQA model proposed by Benoit, *et al.* [26], Cyclopean MS-SSIM [36] which considers the influence of binocular rivalry on perceived 3D quality, and a 3D QA algorithm proposed by You, *et al.*, who applied a variety of 2D FR IQA algorithms on steropairs and disparity maps to combine the predicted quality scores from stereopairs and disparity maps into predicted 3D quality score in a variety of ways. We report their best result on our database (a SSIM-based algorithm). The RR 3D IQA algorithm proposed by Hewage, *et al.* [37] and the NR 3D IQA algorithm proposed by Akhter, *et al.* [38] are also included. We used a SSIM-based stereo-matching algorithm to generate disparity maps for these 3D IQA models. Their performance in terms of SROCC, LCC, and RMSE are reported in Tables V–VII. These tables show that Cyclopean MS-SSIM has the best performance among all compared 3D IQA algorithms although its performance is not significantly different than the performance of 2D MS-SSIM. The results also show that our NR algorithm outperforms most of the 3D IQA algorithms, except for Cyclopean MS-SSIM and the FR models proposed by Benoit, *et al.* when dealing with symmetrically distorted stereopairs. The RR IQA model [37] performs slightly worse than 2D PSNR, while the NR IQA model proposed by Akhter, *et al.* [38] performs significantly worse than 2D PSNR. As shown in the table, 2D MS-SSIM showed the best performance on symmetrically distorted stereopairs, outperforming most 3D QA algorithms, except for Cyclopean MS-SSIM.

2) Phase II Dataset: Binocular rivalry is the main factor that affects the perceived 3D quality of symmetrically distorted stereopairs [15]. On the Phase II dataset, 1000 iterations of train-test process were again used. We report the median result of the 1000 runs. The same set of 2D and 3D IQA algorithms was tested on the phase II dataset. The performance numbers are reported in Tables VIII–XV. As shown in Tables VIII–XI, the performance of our model is significantly better than all of the 2D IQA models. Breaking down performance by distortion type, the improvement relative to different 2D QA models are observed for all distortion types, except for WN. This observation is reasonable, since there is no binocular suppression observed in WN distorted stereopairs. The perceived quality of a WN distorted stereopair is about the average of the qualities of the left and right view [15].

Tables XII–XV shows the results against the mixed dataset of all 3D IQA algorithms. Our model delivers the best performance compared to most other models. The FR Cyclopean MS-SSIM yields an insignificant difference in performance. However, all of the others delivered significantly lower per-

TABLE X
COMPARISON OF 2D IQA MODELS: RMSE AGAINST DMOS ON THE PHASE II DATASET

	WN	JP2K	JPEG	Blur	FF	All
PSNR	4.269	7.674	6.514	9.865	7.456	8.275
SSIM	3.988	6.783	5.572	7.370	5.910	6.741
MS-SSIM	3.334	5.621	3.792	8.397	6.212	7.025
BRISQUE	5.731	7.193	4.448	4.323	4.206	7.038
Our model	3.513	4.298	3.342	4.725	4.180	5.102

Italicized algorithms are NR IQA algorithms, others are RR or FR IQA algorithms.

TABLE XI
COMPARISON OF 2D IQA MODELS: OR AGAINST DMOS ON THE PHASE II DATASET

	WN	JP2K	JPEG	Blur	FF	All
PSNR	0	0	0	0.083	0.014	0.017
SSIM	0	0	0	0.042	0	0.006
MS-SSIM	0	0	0	0.069	0	0.017
BRISQUE	0	0	0	0	0	0.003
Our model	0	0	0	0	0	0

Italicized algorithms are NR IQA algorithms, others are RR or FR IQA algorithms.

TABLE XII
COMPARISON OF 3D IQA MODELS: SROCC AGAINST DMOS ON THE PHASE II DATASET

	WN	JP2K	JPEG	Blur	FF	All
Benoit	0.923	0.751	0.867	0.455	0.773	0.728
You	0.909	0.894	0.795	0.813	0.891	0.786
Gorley	0.875	0.110	0.027	0.770	0.601	0.146
Cyclopean MS-SSIM	0.940	0.814	0.843	0.908	0.884	0.889
Hewage	0.880	0.598	0.736	0.028	0.684	0.501
Akhter	0.714	0.724	0.649	0.682	0.559	0.543
Our model	0.950	0.867	0.867	0.900	0.933	0.880

Italicized algorithms are NR IQA algorithms, others are RR or FR IQA algorithms.

TABLE XIII
COMPARISON OF 3D IQA MODELS: LCC AGAINST DMOS ON THE PHASE II DATASET

	WN	JP2K	JPEG	Blur	FF	All
Benoit	0.926	0.784	0.853	0.535	0.807	0.748
You	0.912	0.905	0.830	0.784	0.915	0.800
Gorley	0.874	0.372	0.322	0.934	0.706	0.515
Cyclopean MS-SSIM	0.957	0.834	0.862	0.963	0.901	0.9
Hewage	0.891	0.664	0.734	0.450	0.746	0.558
Akhter	0.722	0.776	0.786	0.795	0.674	0.568
Our model	0.947	0.899	0.901	0.941	0.932	0.895

Italicized algorithms are NR IQA algorithms, others are RR or FR IQA algorithms.

formance than these two models. Among individual distortion types, our model performed either the best or at parity with the best for all distortion types. Compared with the other 3D NR IQA algorithms [38], our model performed significantly better on the entire dataset and for each distortion type.

TABLE XIV
COMPARISON OF 3D IQA MODELS: RMSE AGAINST DMOS ON THE PHASE II DATASET

	WN	JP2K	JPEG	Blur	FF	All
Benoit	4.028	6.096	3.787	11.763	6.894	7.490
You	4.396	4.186	4.086	8.649	4.649	6.772
Gorley	5.202	9.113	6.940	4.988	8.155	9.675
Cyclopean MS-SSIM	3.368	5.562	3.865	3.747	4.966	4.987
Hewage	10.713	7.343	4.976	12.436	7.667	9.364
Akhter	7.416	6.189	4.535	8.450	8.505	9.294
Our model	3.513	4.298	3.342	4.725	4.180	5.102

Italicized algorithms are NR IQA algorithms, others are RR or FR IQA algorithms.

TABLE XV
COMPARISON OF 3D IQA MODELS: OR ON THE PHASE II DATASET

	WN	JP2K	JPEG	Blur	FF	All
Benoit	0	0	0	0.125	0.014	0.028
You	0	0	0	0.042	0	0.008
Gorley	0	0.028	0	0	0.028	0.044
Cyclopean MS-SSIM	0	0	0	0	0	0
Hewage	0	0	0	0.111	0.056	0.064
Akhter	0	0	0	0.056	0.069	0.039
Our model	0	0	0	0	0	0

Italicized algorithms are NR IQA algorithms, others are RR or FR IQA algorithms.

TABLE XVI
BREAK DOWN OF PERFORMANCE ON SYMMETRICALLY AND ASYMMETRICALLY DISTORTED STIMULI IN THE PHASE II DATASET. SROCC NUMBERS ARE REPORTED

	Symmetric	Asymmetric
2D PSNR	0.776	0.587
2D SSIM	0.828	0.733
2D MS-SSIM	0.912	0.684
2D BRISQUE	0.849	0.667
Benoit	0.860	0.671
You	0.914	0.701
Gorley	0.383	0.056
Cyclopean MS-SSIM	0.923	0.842
Hewage	0.656	0.496
Akhter	0.420	0.517
Our model	0.918	0.834

Italicized algorithms are NR IQA algorithm, others are RR or FR IQA algorithms.

We also studied the performance of the tested algorithms broken down by the way they are distorted (symmetrically or asymmetrically). Table XVI shows the performances of the 2D and 3D IQA algorithms. As one can see, our model performs as well as 2D MS-SSIM, You's algorithm, and Cyclopean MS-SSIM on symmetrically distorted stereo 3D images. When dealing with asymmetrically distorted stereo 3D images, our model significantly outperforms all other both 2D and 3D IQA algorithms, except Cyclopean MS-SSIM, which also models binocular rivalry.

Making comparisons across the phase I and phase II datasets, the FR Cyclopean MS-SSIM model and our new

TABLE XVII
TEST ACROSS DATASETS: SROCC AGAINST DMOS OF THE PHASE I DATASET

	WN	JP2K	JPEG	Blur	FF	All
Train with Phase II dataset	0.826	0.849	0.626	0.882	0.423	0.865
1000 iterations on I dataset	0.919	0.863	0.617	0.878	0.652	0.891

3D NR IQA model show competitive performance on the symmetric dataset and outperform all other 2D and 3D IQA algorithms on the mixed dataset. The 3D FR QA model proposed by You, *et al.* [29] performed either equally well as SSIM on both datasets. The SSIM-based 3D FR model proposed by Benoit performed as well as SSIM on the symmetric dataset, but significantly worse than SSIM on the mixed dataset. The others 3D IQA models performed worse than PSNR on both datasets.

To further verify the performance of our model, we also report performance across datasets. Since only the phase II dataset included both symmetrically and asymmetrically distorted stereopairs, we trained our model on the phase II dataset and tested on the phase I dataset. The result is reported in Table XVII. Across datasets, our model performs equally well on WN, JP2K, JPEG, and the blur distorted stereopairs, but the performance was lower on FF distorted stereopairs. The overall performance drops slightly due to the the performance lost on the WN and FF distorted stimuli.

Lastly, we also tested our algorithm on another 3D image quality database created under significantly different conditions. The MICT stereo image database [38][4] has 480 JPEG distorted stereo images, which include both asymmetrically and symmetrically JPEG distorted stereo images, and 10 pristine stereo image. However, a double stimulus impairment scale(DSIS) protocol and a discrete scale was used in the subjective study. Subjects were asked to assess the annoyance their experienced when viewing each distorted stereo image-pair against the simultaneously displayed reference image by choosing a rating among the following five options: 5=Imperceptible, 4=Perceptible but not annoying, 3=Slightly annoying, 2=Annoying and 1=Very annoying. The display they used was a 10-inch auto-stereoscopic display, but the viewing distance was not provided.

The performance numbers (SROCC, LCC, MSE, and OR) of 2D and 3D QA models on the MICT database are shown in Table XVIII. The performance of our model reported on the MICT database was computed by a 10-fold cross-validation train-test procedure [55]. From Table XVIII, it is clear that the FR models MS-SSIM and C4 deliver the best performance among all the compared QA models on the MICT dataset, even though the majority of the distorted stereo images in the MICT dataset are distorted asymmetrically. The performance of our proposed NR model does not quite match that of MS-SSIM, but is quite competitive with Benoit's 3D FR QA algorithm.

[4]We thank the authors of [38] for kindly providing access to their dataset. We obtained 13 reference stereo images and corresponding 624 distorted stereo images. Following the instructions given by the authors, we used only 10 reference images and 480 distorted stereo images in our experiments.

TABLE XVIII
PERFORMANCE NUMBERS TESTED AGAINST MICT DATABASE

	SROCC	LCC	MSE	OR
2D PSNR	0.586	0.554	0.971	0.255
2D SSIM	0.846	0.862	0.591	0.098
2D MS-SSIM	0.935	0.935	0.415	0.051
C4	0.921	0.927	0.438	0.059
Benoit*	0.902	0.910	0.483	0.071
You*	0.857	0.864	0.586	0.090
Cyclopean MS-SSIM*	0.862	0.864	0.587	0.084
*Akhter**	0.785	0.795	0.708	0.133
*Our model**	0.904	0.913	0.431	0.053

Italicized algorithms are NR IQA algorithm, others are FR IQA algorithms, and 3D IQA algorithms are marked with an asterisk.

Akhter's algorithm[5] did not perform well and did not match the reported numbers in [38].

V. CONCLUSION

We proposed a no-reference stereoscopic 3D image quality assessment algorithm based on 2D and 3D natural scene statistics. The resulting algorithm utilizes statistical features previously proposed for 2D NR algorithms and binocular rivalry modelled by 3D FR IQA algorithms. When there was no binocular rivalry, our algorithm performs equally well as the state-of-the-art 2D NR IQA algorithm. Compared with 3D IQA algorithms, our algorithm significantly outperformed 3D NR QA algorithms and delivered competitive performance relative to high performance 3D FR IQA algorithms.

In the future, we think that extending this framework to be able to predict the quality of depth image base rendered (DIBR) 3D images is an important direction. DIBR generated 3D images may have distortions caused by hole-filling algorithms, 3D warping algorithms, and errors from depth estimation. The challenge of IQA models for DIBR generated 3D images is not limited to visible distortions. Unnaturalness of synthesized 3D stereopairs may contribute to visual discomfort, which is more difficult to quantify than image quality.

REFERENCES

[1] (2011). *Theatrical Market Statistics*, MPAA. Washington, DC, USA [Online]. Available: http://www.mpaa.org/resources/5bec4ac9-a95e-443b-987b-bff6fb5455a9.pdf
[2] (2005). *List of 3D Movies* [Online]. Available: http://en.wikipedia.org/wiki/List_of_3-D_films
[3] *ESPN 3D Broadcasting Schedule*, ESPN. Bristol, CT, USA [Online]. Available: http://espn.go.com/3d/schedule.html
[4] A. C. Bovik and Z. Wang, *Modern Image Quality Assessment*. Morgan, CA, USA: Morgan & Claypool, 2006.
[5] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[6] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[5]We used the parameters reported in [38] without optimizing the parameters of Akhter's algorithm on the dataset.

[7] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, 2012.

[8] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 75–78, Feb. 2012.

[9] M. A. Saad and A. C. Bovik, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[10] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 305–312.

[11] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.

[12] P. Seuntiens, L. Meesters, and W. Ijsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Trans. Appl. Percept.*, vol. 3, no. 2, pp. 95–109, Apr. 2006.

[13] D. V. Meegan, L. B. Stelmach, and W. J. Tam, "Unequal weighting of monocular inputs in binocular combination: Implications for the compression of stereoscopic imagery," *J. Experim. Psychol., Appl.*, vol. 7, no. 2, pp. 143–53, Nov. 2001.

[14] W. J. Tam, L. B. Stelmach, and P. J. Corriveau, "Psychovisual aspects of viewing stereoscopic video sequences," *Proc. SPIE*, vol. 3295, pp. 226–235, Jan. 1998.

[15] M.-J. Chen, A. C. Bovik, and L. K. Cormack, "Study on distortion conspicuity in stereoscopically viewed 3D images," in *Proc. IEEE 10th IVMSP Workshop*, Jun. 2011, pp. 24–29.

[16] M.-J. Chen, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Optimization the 3D image display using the stereoacuity function," in *Proc. 19th IEEE Int. Conf. Image Process.*, Oct. 2012, pp. 617–620.

[17] M.-J. Chen, D.-K. Kwon, and A. C. Bovik, "Study of subject agreement on stereoscopic video quality," in *Proc. IEEE Southwest Symp. Image Anal. Intervent.*, Apr. 2012, pp. 173–176.

[18] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "Quality of experience model for 3DTV," *Proc. SPIE*, vol. 8288, pp. 1–9, Mar. 2012.

[19] M. T. M. Lambooij, W. A. Ijsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: A review," *Proc. SPIE*, vol. 6490, p. 17, Apr. 2007.

[20] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: Visual comfort," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 335–346, Apr. 2011.

[21] 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.

[22] 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, Jul. 2008.

[23] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE*, vol. 6803, p. 5, Feb. 2008.

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

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

[26] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, pp. 1–13, Jan. 2009.

[27] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 2003, pp. 185–193.

[28] 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.

[29] 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. Int. Workshop Video Process. Qaulity Metrics Consum. Elect.*, 2010, pp. 1–6.

[30] Z. Zhu and Y. Wang, "Perceptual distortion metric for stereo video quality evaluation," *WSEAS Trans. Signal Process.*, vol. 5, no. 7, pp. 241–250, Jul. 2009.

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

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

[33] B. Julesz, *Foundations of Cyclopean Perception.*. Chicago, IL, USA: Univ. Chicago Press, 1971.

[34] W. J. M. Levelt, *On Binocular Rivalry*. The Hague, The Netherlands: Mouton, 1968.

[35] 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, Feb. 2012.

[36] M.-J. Chen, D.-K. Su, C.-C. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 1–5.

[37] C. Hewage and M. Martini, "Reduced-reference quality metric for 3D depth map transmission," in *Proc. Transmiss. Display 3D Video 3DTV Conf., True Vis., Capture*, Jun. 2010, pp. 1–4.

[38] R. Akhter, J. Baltes, Z. M. Parvez Sazzad, and Y. Horita, "No-reference stereoscopic image quality assessment," *Proc. SPIE*, vol. 7524, p. 75240T, Feb. 2010.

[39] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[40] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[41] R. Granit, *Receptors and Sensory Perception*. New Haven, CT, USA: Yale Univ. Press, 1955.

[42] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neurosci.*, vol. 9, no. 2, pp. 181–197, 1992.

[43] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vis.*, vol. 18, no. 1, pp. 17–33, 2003.

[44] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in sbuband decomposition of video," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.

[45] J. Huang, A. B. Lee, and D. Mumford, "Statistics of range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2000, pp. 324–331.

[46] Z. Yang and D. Purves, "Image/source statistics of surfaces in natural scenes," *Netw., Comput. Neural Syst.*, vol. 14, no. 3, pp. 371–390, 2003.

[47] P. Hibbard, "A statistical model of binocular disparity," *Vis. Recognit.*, vol. 15, no. 2, pp. 149–165, 2007.

[48] Y. Liu, A. C. Bovik, and L. K. Cormack, "Disparity statistics in natural scenes," *J. Vis.*, vol. 8, no. 11, pp. 1–14, Aug. 2008.

[49] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[50] (2011, Oct.). *Riegl VZ-400 3D Terrestrial Laser Scanner*, RIEGL. Troy, MI, USA [Online]. Available: http://rieglusa.com/products/terrestrial/vz-400/index.shtml

[51] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R Assembly and International Telecommunication Union, Geneva, Switzerland, 2003.

[52] C. C. Chang and C. J. Lin, "LIBSVM: A library for suport vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2001.

[53] H. Sheikh, M. Sabir, and A. 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.

[54] 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.

[55] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, vol. 2. 1995, pp. 1137–1143.

Ming-Jun Chen received the B.S. and M.S. degrees from the Computer Science and Engineering Department, National Cheng-Kung University, Tainan, Taiwan, in 2001 and 2003, respectively. He joined CyberLink, as a Software Engineer, specializing in audio and video processing. He received the Ph.D. degree from the University of Texas at Austin, Austin, TX, USA, in 2012.

His current research interests include 2-D/3-D quality assessment, video/audio content analysis, and visual perception of stereo 3-D content. He serves as an amateur stereographer who was consulted for creating 3-D stereo content.

Lawrence K. Cormack is a Professor of psychology and neuroscience with the University of Texas at Austin, Austin, TX, USA, and is an active member of the Center for Perceptual Systems. He received the B.S. degree (Hons.) in psychology from the University of Florida, Gainesville, FL, USA, in 1986, and the Ph.D. degree in physiological optics from the University of California, Berkeley, CA, USA, in 1992.

His current research interests include how the brain processes the motion of objects through the 3-D environment. This topic is of interest because it is of utmost behavioral importance, it is not well understood, and yet it is a synthesis of two topics that are reasonably well understood (the binocular perception of depth, and the perception of 2-D motion). His other interest include natural scene statistics and their relation to the evolution of the visual system, particularly with regard to depth and motion processing.

Dr. Cormack teaches mostly graduate statistics.

Alan C. Bovik (S'80–M'84–SM'89–F'96) is the Curry/Cullen Trust Endowed Chair Professor with the University of Texas at Austin, Austin, TX, USA, where he is the Director of the Laboratory for Image and Video Engineering. He is a Faculty Member with the Department of Electrical and Computer Engineering, Center for Perceptual Systems, Institute for Neuroscience. His current research interests include image and video processing, computational vision, and visual perception. He has published more than 650 technical articles. He holds two U.S. patents. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009).

He was the SPIE/IS&T Imaging Scientist of the Year for 2011. He has received a number of major awards from the IEEE Signal Processing Society, including the Best Paper Award in 2009, the Education Award in 2007, the Technical Achievement Award in 2005, and the Meritorious Service Award in 1998. He received the Hocott Award for Distinguished Engineering Research from the University of Texas at Austin, the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana in 2008, the IEEE Third Millennium Medal in 2000, and two Journal Paper Awards from the International Pattern Recognition Society in 1988 and 1993. He is a fellow of the Optical Society of America, the Society of Photo-Optical and Instrumentation Engineers, and the American Institute of Medical and Biomedical Engineering. He has been involved in numerous professional society activities and positions, including the Board of Governors of the IEEE Signal Processing Society from 1996 to 1998, a co-founder and an Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002, the Editorial Board for PROCEEDINGS OF THE IEEE from 1998 to 2004, Series Editor for *Image, Video, and Multimedia Processing* (Morgan and Claypool Publishing Company), since 2003, and a Founding General Chairman for First IEEE International Conference on Image Processing, Austin, TX, in 1994.

Dr. Bovik is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.