Perceptual Depth Quality in Distorted Stereoscopic Images

Jiheng Wang, *Member, IEEE*, Shiqi Wang, *Member, IEEE*, Kede Ma, *Student Member, IEEE*, and Zhou Wang, *Fellow, IEEE*

***Abstract*— Subjective and objective measurement of the perceptual quality of depth information in symmetrically and asymmetrically distorted stereoscopic images is a fundamentally important issue in stereoscopic 3D imaging that has not been deeply investigated. Here, we first carry out a subjective test following the traditional absolute category rating protocol widely used in general image quality assessment research. We find this approach problematic, because monocular cues and the spatial quality of images have strong impact on the depth quality scores given by subjects, making it difficult to single out the actual contributions of stereoscopic cues in depth perception. To overcome this problem, we carry out a novel subjective study where depth effect is synthesized at different depth levels before various types and levels of symmetric and asymmetric distortions are applied. Instead of following the traditional approach, we ask subjects to identify and label depth polarizations, and a depth perception difficulty index (DPDI) is developed based on the percentage of correct and incorrect subject judgements. We find this approach highly effective at quantifying depth perception induced by stereo cues and observe a number of interesting effects regarding image content dependency, distortion-type dependence, and the impact of symmetric versus asymmetric distortions. Furthermore, we propose a novel computational model for DPDI prediction. Our results show that the proposed model, without explicitly identifying image distortion types, leads to highly promising DPDI prediction performance. We believe that these are useful steps toward building a comprehensive understanding on 3D quality-of-experience of stereoscopic images.**

***Index Terms*— Depth perception, depth quality, stereoscopic image, 3D image, image quality assessment, quality-of-experience, asymmetric distortion, depth polarization.**

I. Introduction

AUTOMATICALLY assessing the quality of 3D visual experience is a challenging problem [1]–[3], especially due to the sophistication and interaction between multiple 3D visual cues including image quality, depth quality and visual comfort [4]–[6]. Recent progress on subjective and objective studies of 3D image quality assessment (IQA) is promising [7]–[9] but the understanding of 3D depth quality remains limited. Depth quality is no doubt an essential aspect of human quality-of-experience (QoE) when viewing stereoscopic 3D images.

Existing studies on the topic appear to be unconclusive, limited, and sometimes conflicting. In [10], it was reported that the perceived depth performance cannot always be predicted from displaying image geometry alone, while other system factors, such as software drivers, electronic interfaces, and individual participant differences, may also play significant roles. Similarly, in [11], it was pointed out that an appropriate choice of camera and display system setup can eliminate some stereoscopic distortions that affect the perceived depth. In [4], subjective studies showed that the perceived depth increases when increasing the camera-base distance. This has been further explored in [12], where it was found that the perceived depth quantity always increases but the perceived depth quality may decrease with the increasing level of binocular depth. In [4] and [13], it was suggested that the perceived depth may need to be considered independently from the perceived 3D image quality. The results in [4] showed that increasing the level of JPEG compression has no clear effect on the perceived depth however a negative effect on image quality, which is generally consistent with the results given in [13] for different levels of blurring. On the other hand, in [12], subjective studies suggested that 3D image quality is not sensitive to variations in the degree of the binocular depth, which is agreed by [14], where the perceived image quality exhibits less correlation with the perceived depth.

Other studies pointed out the perceptual depth as an important component in the holistic 3D QoE. In [15], a blurring filter, where the level of blur depends on the depth of the area where it is applied, is used to enhance the viewing experience. In [16], subjective studies revealed that humans tend to prefer DCT compressed stereopairs over the monoscopic single-views even though the blocking artifacts are annoying, which has been partially verified in [17], where it was found that the strength of this preference depends on the quality range being investigated and there also exists a content dependency that this preference could be flipped. In [18], the depth range of color-plus-depth has been optimized to increase visual comfort for stereoscopic 3D displays. In [19], depth naturalness was shown to be a useful ingredient in the assessment of 3D video QoE. Similarly, in [20], the added value of depth naturalness has been verified for pristine and blurred stereoscopic images. In [21], stimuli with various stereo depth and image quality were evaluated subjectively in terms of naturalness, viewing

Manuscript received June 13, 2016; revised November 10, 2016; accepted December 6, 2016. Date of publication December 20, 2016; date of current version January 20, 2017. Preliminary results on the subjective testing part of this work were presented at the IEEE International Workshop on Multimedia Signal Processing, Xiamen, China, Oct. 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Stefan Winkler.

J. Wang, K. Ma, and Z. Wang are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: j237wang@uwaterloo.ca; k29ma@uwaterloo.ca; zhou.wang@uwaterloo.ca).

S. Wang is with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore 637553 (email: wangshiqi@ntu.edu.sg).

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.2016.2642791

1057-7149 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

experience, image quality, and depth perception, and the experimental results suggested that the overall 3D QoE is approximately 75% determined by image quality and 25% by the perceived depth. In [22], Chen *et al.* showed that subjective evaluation of depth quality has a low correlation with that of 3D image quality and verified that the overall 3D QoE can be predicted using a single linear model from 3D image quality and depth quality.

Several objective models have been proposed to automatically predict the perceived depth quality and subsequently to predict 3D quality by combining depth quality and 2D image quality. In [23], peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [24], and video quality metric (VQM) [25] were employed to predict the perceived depth quality, and PSNR and SSIM appear to have slightly better performance. In [26] and [27], disparity maps between left- and right-views were estimated, followed by 2D quality assessment of disparity quality using SSIM and C4 [28], which was subsequently combined with 2D image quality to produce an overall 3D image quality score. The results suggested that C4 outperforms SSIM on evaluating stereoscopic image pairs and disparity maps on IRCCyN/IVC 3D Image Database [26] and also showed that the 3D-IQA performance of SSIM can be improved when adding depth quality. You *et al.* [29] evaluated stereopairs as well as disparity maps with respect to ten well-known 2D-IQA models, i.e., PSNR, SSIM, multi-scale SSIM (MS-SSIM) [30], universal quality index (UQI) [31], visual information fidelity (VIF) [32], visual signal-to-noise ratio (VSNR) [33], etc. The results suggested that an improved performance can be achieved when stereo image quality and depth quality are combined appropriately. Similarly, Yang *et al.* [34], [35] proposed a 3D-IQA algorithm based on the average PSNR of left- and right-views and the absolute difference with respect to the disparity map. In [36], Zhu *et al.* proposed a 3D video quality assessment (VQA) model by considering depth perception, and the experimental results showed that the proposed human vision system (HVS) based model performs better than PSNR.

Nevertheless, in [37]–[39], comparative studies showed that none of these 3D-IQA/VQA models, with depth information involved, perform better than or in most cases, even as good as, direct averaging 2D-IQA measures of both views. In particular, in [37], it was shown that averaging PSNR, SSIM, MS-SSIM, UQI, and VIF measurements of left- and right-views performs equally well or better than the advanced 3D-IQA models [26], [29], [34], [36] on LIVE 3D Image Quality Database Phase I. Similar results were also observed in [38], where averaging SSIM and MS-SSIM measurements of both views outperformed advanced 3D-IQA models [26], [29] on LIVE 3D Image Quality Database Phase II. In [39], it was reported that directly averaging MS-SSIM outperformed 3D-IQA models [26], [29] on Ningbo University 3D Image Quality Assessment Database. All these observations suggest that the progress on how to automatically predict depth quality and how to combine 3D image quality and depth quality remains limited. This lack of successful objective QoE methods for 3D visual experience has limited their applications in the development of 3D imaging applications and services.

TABLE I
DESCRIPTION OF VISUAL EXPERIENCE CRITERIA

Criterion	Description
2DIQ	The image content quality
3DIQ	The image content quality without considering depth and comfortness
DQ	The amount, naturalness and clearness of depth perception experience
VC	The comfortness when viewing stereoscopic images
3DQoE	The overall 3D viewing experience

In this work, we carry out two subjective experiments on depth quality. The first one adopts a traditional absolute category rating (ACR) [40] protocol widely used in general IQA research. We find this approach problematic in this scenario because monocular cues and the spatial quality of images have strong impact on the depth quality scores given by subjects, making it difficult to single out the actual contributions of stereoscopic cues in depth perception. To overcome this problem, we conduct the second subjective study where depth effect is synthesized at different depth levels before various types and levels of symmetric and asymmetric distortions are applied. Instead of following the traditional approach, we ask subjects to identify and label depth polarizations, and a Depth Perception Difficulty Index (DPDI) is developed based on the percentage of correct and incorrect subject judgements. We find the second approach highly effective at quantifying depth perception induced by stereo cues. We then carry out a series of analysis to investigate the impact of image content, distortion type, and distortion symmetricity on perceived depth quality. Furthermore, we propose a novel computational model for DPDI prediction. Our results show that the proposed model, without explicitly identifying image distortion types, leads to highly promising DPDI prediction performance.

II. SUBJECTIVE STUDY I

A. Image Database

The WATERLOO-IVC 3D Image Quality Database Phase I [41], [42] was created from 6 pristine stereoscopic image pairs and their corresponding single-view images as shown in Fig. 1. Each single-view image was altered by three types of distortions: additive white Gaussian noise contamination, Gaussian blur, and JPEG compression, and each distortion type had four distortion levels. The single-view images are employed to generate distorted stereopairs, either symmetrically or asymmetrically. There are totally 78 single-view images and 330 stereoscopic images in the database. Following previous works [4], [12], [22], the subjects were asked to evaluate four aspects of their 3D viewing experience, including the perception of 3D image quality (3DIQ), depth quality (DQ), visual comfort (VC), and overall 3D quality of experience (3DQoE). The detailed descriptions of each aspects of visual experience including 2D image quality (2DIQ) are elaborated in Table I. More comprehensive descriptions are in [41] and [42]. In this paper, we focus on depth quality only, which refers to the amount, naturalness and clearness of depth perception experience.

(a) (b) (c) (d) (e) (f)

Fig. 1. The pristine images in Waterloo-IVC 3D Image Database Phase I. Only the right-views are shown here. (a) Art. (b) Books. (c) Dolls. (d) Moebius. (e) Laundry. (f) Reindeer.

TABLE II
VIEWING CONDITIONS OF THE SUBJECTIVE TEST

Parameter	Value	Parameter	Value
Subjects Per Monitor	1	Screen Resolution	1920 × 1080
Screen Diameter	27.00”	Viewing Distance	45.00”
Screen Width	23.53”	Viewing Angle	29.3°
Screen Height	13.24”	Pixels Per Degree	65.5 pixels

B. Subjective Test

The subjective test was conducted in the Lab for Image and Vision Computing at University of Waterloo. The test environment has no reflecting ceiling walls and floor, and was not insulated by any external audible and visual pollution. An ASUS 27" VG278H 3D LED monitor with NVIDIA 3D Vision™2 active shutter glasses is used for the test. The default viewing distance was 3.5 times of the screen height. In the actual experiment, two subjects (both male) preferred a larger viewing distance of about 4 times of the screen height and were allowed to make such adjustment before the test. The details of viewing conditions are given in Table II. Twenty-four naïve subjects, 14 males and 10 females aged from 22 to 45, participated in the study. A 3D vision test (Random dot stereo test) was conducted first to verify their ability to view stereoscopic 3D content. Three of them (1 male, 2 females) failed the vision test and did not continue with the subsequent experiment. As a result, a total of twenty-one subjects proceeded to the formal test.

One-pass experiment (where a subject gives 3DIQ, DQ, VC, and 3DQoE scores to each stereoscopic image in one trial) may cause significant visual fatigue of the human subjects within a short period of time. To avoid this problem, we resorted to a multi-pass approach [12] in the formal test, where within each pass, the subject gives one of the four scores. In addition, there is a 2DIQ sub-test for single-view images. Fig. 2 shows the detailed procedure of our formal subjective test. We followed the ACR protocol and the subjects were asked to rate the 2D or 3D visual experience criteria of each image between 0 and 10 pts.

For both 2DIQ and 3DIQ sub-tests, we use three types of images in the training phase: pristine images, moderately distorted images, and highly-distorted images. The subjects were told to give scores at the high end (close to 10 pts) to the pristine images, at the mid-range to the moderately distorted images, and at the low end (close to 0 pts) to the highly-distorted images.

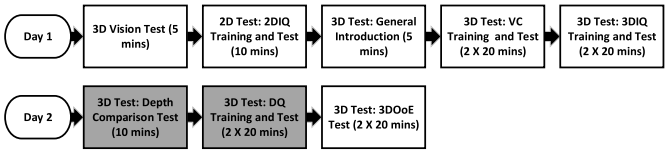


Fig. 2. The procedure of the subjective test in Subjective Study I.

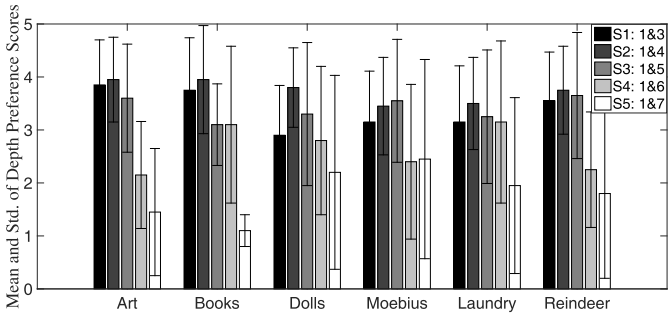


Fig. 3. Means and standard deviations of depth preference scores in the depth comparison test.

For DQ sub-test, a self-training process was employed to help the subjects establishing their own rating strategies with the help of the depth comparison test (stimuli with the same source image similar to what are used in the formal test but different depth levels were presented to help the subjects establish the concept on the amount of depth), and subjects were introduced to build their own rating strategies.

The motivation of introducing a depth comparison test is to help human subjects understand the amount of depth perception for each pristine stereopairs from their own preference and thus let them focus on evaluating the depth quality degraded by different distortions in the following depth quality test. The six pristine stereopairs from WATERLOO-IVC 3D image database were utilized in this test. For each pristine stereopair, a single-view image (view 1) was firstly displayed to help the subjects get familiar with image content and then five different stereopairs with an increasing amount of depth were presented, which are S1 (view 1 and view 3), S2 (view 1 and view 4), S3 (view 1 and view 5), S4 (view 1 and view 6), and S5 (view 1 and view 7). Subjects were allowed to compare these six stereopairs back and forth and then to rank them based on their own preference for depth perception. Some subjects favored S5 with the largest amount of depth while others preferred the mid-level S3 as they felt the 3D objects presented in S5 come too close to their faces. The depth preference score is assigned from 1 to 5 pts, for which 1 represents the least preferred and 5 the most preferred. The means and standard deviations of depth preference scores are shown in Fig. 3, where we observe high variations between subject scores, suggesting diverse subject opinions in depth preference.

Previous works reported that the perception of depth quality are both highly content and texture dependent [43] and subject dependent [12], [22]. Therefore, it is not desirable to over-educate the subjects to use the same given rating strategy. Thus after the depth comparison test, the 3D pristine stereopairs were first presented and the subjects were instructed to give high scores (close to 10 pts) to such images, and the 2D pristine images (with no depth from stereo cues) were presented and the subjects were instructed to give low scores (close to 0 pts). Next, stereopairs of different types/levels of distortions were presented and the subjects were asked to practice by giving their ratings on depth quality between 0 and 10 pts. During this process, the instructor also repeated the definition of depth quality and emphasized that there is not necessarily any correlation between depth quality and the type/level of distortions.

Most stimuli were shown once in each sub-test. However, there were 6 repetitions for single-view images and 12 repetitions for stereopairs, which means that for each subject, her/his first 6 single-view images and first 12 stereopairs were shown twice. The order of stimuli was randomized and the consecutive testing single-view images or stereopairs were from different source images. The 2DIQ sub-test, including 84 testing single-view images with 6 repetitions, was finished under 10 minutes. For 3DIQ , DQ , VC, and 3DQoE sub-tests, 342 testing stereopairs with 12 repetitions were partitioned into two sessions and each single session (171 stereopairs) was finished in 15 to 20 minutes. Sufficient relaxation periods (5 minutes or more) were given between sessions. Moreover, we found that repeatedly switching between viewing 3D images and grading on a piece of paper or a computer screen is a tiring experience. To overcome this problem, we asked the subject to speak out a score, and a customized graphical user interface on another computer screen was used by the instructor to record the score. All these efforts were intended to reduce visual fatigue and discomfort of the subjects.

C. Observations and Discussions

Following the previous work [24], the raw 2DIQ, 3DIQ, and DQ scores given by each subject were converted to Z-scores [44], respectively. Then the entire data sets were rescaled to fill the range from 1 to 100 and the mean opinion scores (MOS) for 2DIQ, 3DIQ, and DQ, i.e., MOS 2DIQ, MOS 3DIQ, and MOS DQ, were computed. The detailed observations and analysis of the relationship between MOS 2DIQ and MOS 3DIQ and how to predict the image content quality of a stereoscopic 3D image from that of the 2D single-view images can be found in [8] and [42]. Here we focus on the depth quality part, i.e., individual DQ scores and MOS DQ.

For each stereopair, the standard deviation of Z-scores represents the degree of variation and the means of these standard deviations are 12.00 for 3DIQ scores and 20.01 for DQ scores, respectively, indicating large variations in DQ scores. Table III reports Pearson's linear correlation coefficient (PLCC), Spearman's rank-order correlation coefficient (SRCC), and Kendall's rank-order correlation coefficient (KRCC) between individual 3DIQ/DQ scores and

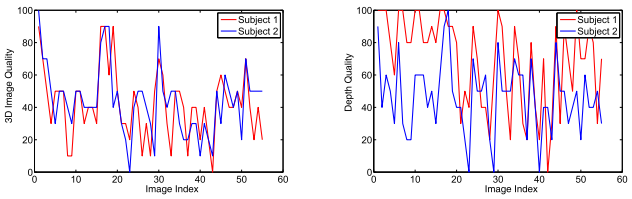


Fig. 4. Comparison of two subjects' 3DIQ and DQ scores on the Art stereopairs. (a) 3D Image Quality. (b) Depth Quality.

TABLE III
MEANS AND STANDARD DEVIATIONS OF CORRELATIONS BETWEEN INDIVIDUAL SCORES AND MOS

	PLCC		SRCC		KRCC	
Criterion	Mean	Std.	Mean	Std.	Mean	Std.
3DIQ	0.8045	0.0658	0.7762	0.0726	0.6315	0.0695
DQ	0.7414	0.1410	0.7298	0.1443	0.5838	0.1259

MOS 3DIQ/DQ scores, which reflect the degree of agreement of 3DIQ/DQ scores among the subjects. PLCC is adopted to evaluate prediction accuracy [45] and SRCC and KRCC are employed to assess prediction monotonicity [45]. Higher PLCC, SRCC, and KRCC indicate better consistency with the average human opinions of quality. PLCC is usually computed after a nonlinear mapping between the subjective and objective scores and the results may be sensitive to the choice of the mapping function. SRCC and KRCC are nonparametric rank order-based correlation metrics, independent of any monotonic nonlinear mapping between subjective and objective scores but do not explicitly estimate the accuracy of quality prediction.

From Table III, it can be observed that individual DQ scores show less correlation with MOS compared with individual 3DIQ scores. To further understand this, Fig. 4 shows a comparison of two subjects' 3DIQ and DQ scores on the Art stereopairs. It can be observed that these two subjects exhibit general agreement on 3DIQ scores but behave drastically differently in giving DQ scores.

Thus our preliminary analysis shows that there is a large variation between subjects on depth quality scores as different people may have very different perception and/or opinions about perceptual depth quality. The rest of this section will focus on the relationship between DQ scores and the 3DIQ scores.

Fig. 5 shows the scatter plots of MOS 3DIQ vs. averaging MOS 2DIQ of left- and right-views and MOS 3DIQ vs. MOS DQ. Fig. 5 (a) suggests that there exists a strong distortion type dependent prediction bias when predicting quality of asymmetrically distorted stereoscopic images from single-views [41], [42]. Specifically, for noise contamination and JPEG compression, average prediction overestimates 3DIQ (or 3DIQ is more affected by the poorer quality view), while for blur, average prediction often underestimates 3DIQ (or 3DIQ is more affected by the better quality view).

From Fig. 5 (b), it can be observed that human opinions on 3DIQ and 3D DQ are highly correlated. This is somewhat surprising because 3DIQ and DQ are two different perceptual attributes and the stimuli were generated to cover all combinations between picture qualities and stereo depths.

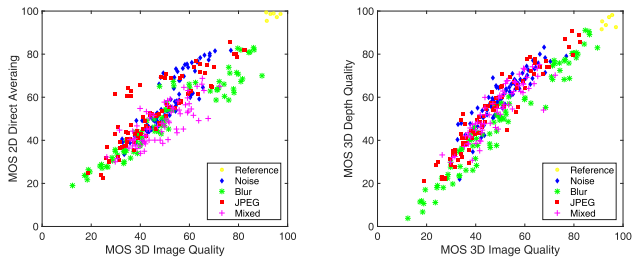


Fig. 5. Relationships between (MOS 3DIQ and MOS 2DIQ) and (MOS 3DIQ and MOS DQ) in Subjective Study I. (a) MOS 3DIQ vs. MOS 2DIQ. (b) MOS 3DIQ vs. MOS DQ.

Through more careful observations of the data and discussions with the subjects who did the experiment, we found two explanations. First, psychologically humans have the tendency to give high DQ scores whenever the 3DIQ is good and vice versa, and the strength of such a tendency varies between subjects. Second, humans interpret depth information using many physiological and psychological cues [46], including not only binocular cues such as stereopsis, but also monocular cues such as retinal image size, linear perspective, texture gradient, overlapping, aerial perspective, and shadowing and shading [47], [48]. In the real world, humans automatically use all available depth cues to determine distances between objects but most often rely on psychological monocular cues. Therefore, the DQ scores obtained in the current study are a combined result from many monocular and binocular cues, and it becomes difficult to gauge the role of stereopsis.

However, what we are interested in the current study is to measure how much stereo information can help with depth perception. Based on the explanations above, in traditional ways of subjective testing like the current one, many depth cues are mixed together and the results are further altered by the spatial quality of the image, making it difficult to quantify the real contributions of using stereoscopic images in depth perception. This inspires us to design a novel depth perception test, which will be presented in the next section.

III. Subjective Study II

A. Image Database

We created a new Waterloo-IVC 3D Depth Database from 6 pristine texture images (Bark, Brick, Flowers, Food, Grass, and Water) as shown in Fig. 6. All images were collected from the VisTex Database at MIT Media Laboratory [49]. A stereogram can be build by duplicating the image, selecting a region in one image, and shifting this region horizontally by a small amount in the other one. The region seems to virtually fly in front of the screen, or be behind the screen if the two views are swapped. In our experiment, this horizontal shifting is controlled by six different levels of Gaussian surfaces with different heights and different widths, which were obtained by translating and scaling Gaussian profiles. Depth 1 and Depth 6 denote the lowest and highest depths, respectively, and were selected to ensure a good perceptual separation. Thus each texture image was used to generate 6 stereopairs with

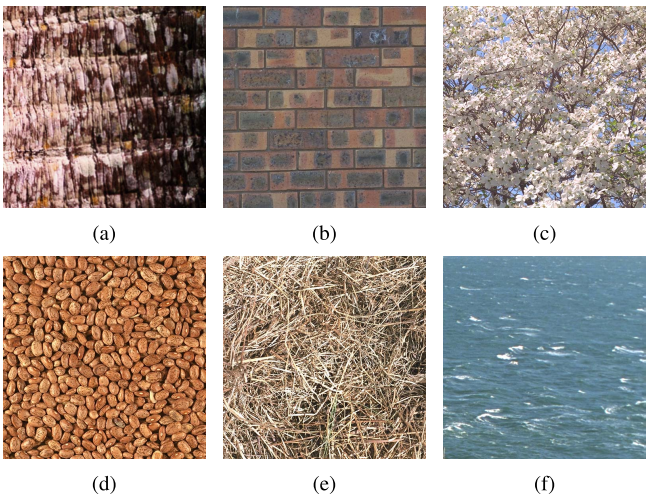


Fig. 6. The texture images used in Subjective Study II. (a) Bark. (b) Flower. (c) Food. (d) Grass. (e) Water.

TABLE IV
VALUE RANGES OF CONTROL PARAMETERS TO GENERATE IMAGE DISTORTIONS

Distortion	Control Parameter	Range
Noise	Variance of Gaussian	[0.11 1.12]
Blur	Variance of Gaussian	[1.50 11.00]
JPEG	Quality parameter	[1 10]

different depth levels. By switching left- and right-views, the hidden depth could be perceived towards inside or outside and we denote them as inner stereopairs and outer stereopairs, respectively. As such, for each texture image, we have 12 pristine stereopairs with different depth polarizations and depth levels. In addition, one flat stereopair without any hidden depth information is also included.

Each pristine stereopair (inner, outer, and flat) was altered by three types of distortions: additive white Gaussian noise contamination, Gaussian blur, and JPEG compression. Each distortion type had four distortion levels as reported in Table IV, where the distortion control parameters were decided to ensure a good perceptual separation. The distortions were simulated either symmetrically or asymmetrically. Symmetrically distorted stereopairs have the same distortion type and level on both views while asymmetrically distorted ones have the distortion on one view only. Altogether, there are 72 pristine stereoscopic images and 1728 distorted stereoscopic images (864 symmetrical and 864 asymmetrical distortions) in the database. In terms of the depth polarity, there are 684 inner stereopairs, 684 outer stereopairs, and 432 flat stereopairs. An example of the procedure of generating a symmetrically blurred stereopair is shown in Fig. 7.

For each image, we provide the subjects with four available choices to respond, i.e., inner, outer, flat, and unable to decide. The motivation of introducing the last choice is that for some distorted stereopairs, the subjects can perceive the existence of depth information but feel difficult to make confident judgements on depth polarity.

There are three important features of the current database that distinguish it from others. First, the depth information

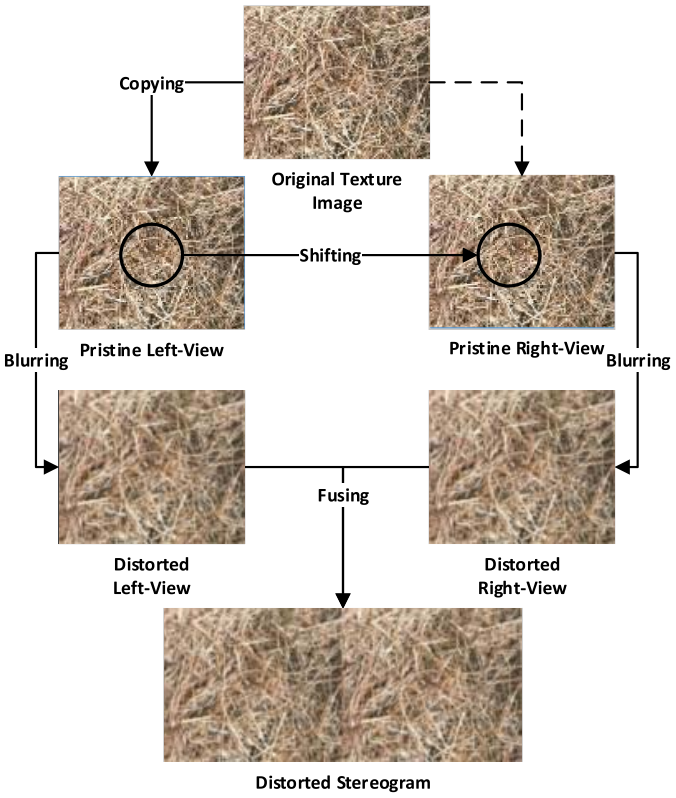


Fig. 7. Procedure of generating a symmetrically blurred stereoscopic image in Subjective Study II.

embedded in each stereopair is independent of its 2D scene contents, such that subjects can only make use of stereo cues to identify depth change and judge the polarity of depth. Second, the database contains distorted stereopairs from various distortion types, allowing us to compare the impacts of different distortions on depth perception. Third, the current database contains both symmetrically and asymmetrically distorted stereopairs, which allows us to directly examine the impact of asymmetric distortions on depth perception. This may also help us better understand what are the key factors that affect depth quality in stereoscopic images.

B. Subjective Test

The subjective test was conducted in the Lab for Image and Vision Computing at University of Waterloo with the same test environment, the same 3D display system, and the same viewing conditions as described in Section II. Thus here we only describe some important differences from Subjective Study I. Twenty-two naive subjects, 11 males and 11 females aged from 21 to 34, participated in the study and no one failed the vision test. As a result, a total of twenty-two subjects proceeded to the formal test. The training process is fairly straightforward. Twelve stereopairs with different depth configurations including polarities and levels were presented to the subjects. Subjects were asked to speak out their judgements for these training stereopairs as an exercise. Then a multi-stimulus method was adopted to obtain subjective judgements for all test stereopairs. Each stimulus contains six stereopairs with the same depth level and the same image content but different depth polarity or image distortion. All stimuli were shown once and the order of stimuli was randomized. 75 stimuli were evaluated in one session and each session was controlled to be within 20 minutes. Similarly, subjects only needed to speak out their judgements and an instructor was responsible for recording subjective results.

We observe a significant variation between subjects' behaviors, i.e., they exhibit different levels of ability to identify depth polarizations and show different preferences for inner or outer images, which is expected as humans exhibit a wide variety of stereo-acuity and stereo-sense [50]. The rest of this section focuses on the impact of depth level, depth polarity, image content and image distortion. More detailed analysis of the other aspects of the subjective data will be reported in future publications.

TABLE V
AVERAGE DPDI VALUES OF DIFFERENT DEPTH LEVELS

Depth Levels	Inner	Outer	Outer − Inner	All
Level 1	0.9196	0.9146	-0.0050	0.9171
Level 2	0.7605	0.7883	0.0278	0.7744
Level 3	0.5829	0.6721	0.0892	0.6275
Level 4	0.4095	0.5732	0.1637	0.4914
Level 5	0.3409	0.5008	0.1599	0.4209
Level 6	0.2811	0.4474	0.1663	0.3643

C. Depth Perception Difficulty Index (DPDI)

For each test image, there are 3 possible ground-truth polarity answers - inner, outer, and flat. Meanwhile, pooling the subjective judgements on the image leads us to four percentage values, denoted by $\{P_{\text{in}}, P_{\text{out}}, P_{\text{flat}}, P_{\text{unable}}\}$, corresponding to the percentages of subject judgements of inner, outer, flat, and unable to decide, respectively, and $P_{\text{in}} + P_{\text{out}} + P_{\text{flat}} + P_{\text{unable}} = 1$. Given these values, we define a novel measure named Depth Perception Difficulty Index (DPDI), which indicates how difficult it is for an average subject to correctly perceive the depth information in the image. Specially, if the ground-truth is an inner image, we define

$$\begin{aligned}\text{DPDI} &= 1 - \max\{0, P_{\text{in}} - P_{\text{out}}\} \\ &= \min\{1, P_{\text{flat}} + P_{\text{unable}} + 2P_{\text{out}}\}. \end{aligned} \tag{1}$$

Similarly, for an outer image

$$\begin{aligned}\text{DPDI} &= 1 - \max\{0, P_{\text{out}} - P_{\text{in}}\} \\ &= \min\{1, P_{\text{flat}} + P_{\text{unable}} + 2P_{\text{in}}\}. \end{aligned} \tag{2}$$

This DPDI is bounded between 0 and 1. The values of DPDI in some extreme cases are as follows: when we have $\{1, 0, 0, 0\}$ for inner images or $\{0, 1, 0, 0\}$ for outer images, DPDI equals 0; when we have $\{0.25, 0.25, 0.25, 0.25\}$, which is equivalent to the case of random guess, DPDI equals 1.

D. Analysis and Key Observations

Table V shows the mean DPDI values for different depth levels for the cases of all images, inner images, and outer images. Unsurprisingly, DPDI drops with increasing depth in each test group. A much more interesting observation here is

TABLE VI
AVERAGE DPDI VALUES OF DIFFERENT IMAGE CONTENTS

Image Contents	Inner	Outer	Outer − Inner	All
Bark	0.4831	0.5793	0.0962	0.5339
Brick	0.7562	0.9226	0.1664	0.8209
Flower	0.4232	0.4985	0.0753	0.4545
Food	0.4948	0.6007	0.1059	0.5448
Grass	0.2646	0.4315	0.1669	0.3620
Water	0.8712	0.8846	0.0134	0.8794

TABLE VII
AVERAGE DPDI VALUES OF DIFFERENT DISTORTION TYPES AND LEVELS

Distortions	All	Level 1	Level 2	Level 3	Level 4
Noise Sym.	0.7986	0.6275	0.7412	0.8838	0.9419
Noise Asym.	0.6504	0.5215	0.6477	0.7058	0.7500
Blur Sym.	0.5470	0.4962	0.4912	0.5492	0.6515
Blur Asym.	0.5431	0.3902	0.4975	0.6048	0.6528
JPEG Sym.	0.5660	0.4444	0.5278	0.6187	0.6730
JPEG Asym.	0.5473	0.4470	0.5265	0.5871	0.6679

that with a given level of depth, inner images generally have lower DPDI values and the difference in mean DPDI values between inner and outer images increase with the level of depth. This indicates that it is easier for humans to perceive depth information when objects appear to be behind the screen than in the opposite case.

Table VI reports the mean DPDI values for different background image contents. First, it appears that DPDI is highly image content dependent as it varies significantly across content. In general, DPDI decreases with the increase of high-frequency details, which is consistent with the previous vision research [51] that stereo gain is higher for the high spatial-frequency system than the low spatial-frequency system. Second, although inner images always have higher DPDI values, the gap between inner and outer images is image content dependent.

Table VII shows the mean DPDI values of different distortion types and levels. First, across distortion types, noise contamination has more impact on depth perception than JPEG compression and Gaussian blur. Second, more interestingly, although the cases of symmetric distortions double the total amount of distortions than asymmetric distortions (because the same level of distortions is added to both views), the DPDI gap between asymmetric and symmetric distortions is distortion type dependent. The gaps in the case of noise contamination is much higher than those of Gaussian blur and JPEG compression. The point worth noting is that adding blur or JPEG compression to one view of stereopair results in similar difficulty in depth perception as adding the same level of distortion to both views. This is quite different from the distortion type dependency in 3D image quality perception, as shown in Fig. 5 (a). It is interesting to note that some of our new observations are somehow implicitly consistent with previous vision studies [52], [53]. For example, in [53], Hess *et al.* found that stereoacuity was reduced when one view was severely blurred by filtering off high spatial frequencies and loss of acuity was much less severe when both views are blurred.

E. Impact of Eye Dominance

Eye dominance is a common visual phenomenon, referring to the tendency to prefer the input from one eye to the other, depending on the human subject [54]. When studying visual quality of asymmetrically distorted images, it is important to understand if eye dominance plays a significant role in the subjective test results. For this purpose, we carried out a separate analysis on the impact of eye dominance in the depth perception of asymmetrically distorted stereoscopic images. The side of the dominant eye under static conditions was checked first by Rosenbach's test [55]. This test examines which eye determines the position of a finger when the subject is asked to point to an object. Among twenty subjects who finished the formal test Subjective Study II, ten subjects (6 males, 4 females) had a dominant left eye, and the others (5 males, 7 females) are right-eye dominant.

The DPDI for each image in Waterloo-IVC 3D Depth Database were computed for left-eye dominant subjects and right-eye dominant subjects, denoted as $\mathrm{DPDI_L}$ and $\mathrm{DPDI_R}$, respectively. We employed the one-sample t-test to obtain a test decision for the null hypothesis that the difference between $\mathrm{DPDI_L}$ and $\mathrm{DPDI_R}$, i.e., $\mathrm{DPDI_D} = \mathrm{DPDI_L} - \mathrm{DPDI_R}$, comes from a normal distribution of zero-mean and unknown variance. The alternative hypothesis is that the population distribution does not have a mean equaling zero. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise. The returned p-values for symmetric and asymmetric images are 0.3448 and 0.3048, respectively, thus the null hypothesis cannot be rejected at the 5% significance level, which indicates that the impact of eye dominance in the perception of depth quality of asymmetrically distorted stereoscopic images is not significant.

It is worth noting that in [8] we found that the eye dominance effect does not have strong impact on the perceived image content quality of stereoscopic images. Our two observations are consistent with the "stimulus" view of rivalry that is widely accepted in the field of visual neuroscience [56]. A comprehensive review and discussion on "stimulus" rivalry versus "eye" rivalry can be found in [56] and [57].

IV. OBJECTIVE STUDY: PREDICTION OF DEPTH PERCEPTION DIFFICULTY INDEX

A. DPDI Prediction Model

We opt to use a multiple-stage approach in the design of an objective DPDI predictor. The first stage aims to predict the DPDI for different depth levels $H_{(L)evel}$ and image contents $H_{(C)ontent}$, while in the second stage, a patch-structure representation is developed to predict the DPDI for different distortion types and levels $H_{(D)istortion}$. Finally, these components are combined to yield an overall DPDI prediction model.

In Section III-D, DPDI is found to decrease with the depth level monotonically. Here we look for an efficient approach to predict DPDI values of different levels using stereo matching, which is an active research area in computer vision over the last few decades [58]. Specially, given a stereopair of $\mathbf{x}_l$ and $\mathbf{x}_r$ for the left-view and right-view reference images, respectively, we first estimate the disparity map $\mathbf{D}_{lr}$, which is simply done

TABLE VIII
DPDI, H_L VALUES, AND $\mu_{\mathbf{D}_{lr}}$ OF DIFFERENT DEPTH LEVELS

Depth Levels	DPDI	H_L	$\mu_{\mathbf{D}_{lr}}$
Depth 1	0.9171	0.8398	0.0063
Depth 2	0.7744	0.7970	0.0319
Depth 3	0.6275	0.7010	0.1006
Depth 4	0.4914	0.5550	0.2507
Depth 5	0.4209	0.3977	0.5358
Depth 6	0.3643	0.2662	1.0329

TABLE IX
DPDI, H_C VALUES, AND ENERGIES OF DIFFERENT IMAGE CONTENTS

Image Content	DPDI	H_C	Energy E
Bark	0.5339	0.5766	561
Brick	0.8209	0.7939	99
Flower	0.4545	0.5036	1,404
Food	0.5448	0.5149	1,199
Grass	0.3620	0.4440	3,719
Water	0.8794	0.8208	85

by using MATLAB®'s utility `disparityMap` [59]. Our experiment shows that the estimations are highly accurate, allowing us to design a simple approach to predict how DPDI changes with depth levels. We denote $\mu_{\mathbf{D}_{lr}}$ as the mean of disparity values and apply a nonlinear mapping on $\mu_{\mathbf{D}_{lr}}$ to predict the DPDI values of different depth levels

$$H_L = \frac{\alpha}{(\mu_{\mathbf{D}_{lr}})^{\beta} + \gamma}, \tag{3}$$

where the best parameters are found to be $\alpha = 0.4$, $\beta = 1$ and $\gamma = 0.47$ and H_L values of different depth levels are reported in Table VIII.

In Section III-D, we find that DPDI is highly image content dependent as it varies significantly across content. In general, DPDI decreases with the increase of high-frequency details or energy. We measure the energy by computing the local variances at each spatial location, i.e., the variances of local image patches extracted around each spatial location, for which an 11×11 circular-symmetric Gaussian weighting function $\mathbf{w} = \{w_i | i = 1, 2, \cdots, N\}$ with standard deviation of 1.5 samples, normalized to unit sum ($\sum_{i=1}^{N} w_i = 1$), is employed. The mean of local variances is used to measure the energy E. Empirically, we observe that E shows a high dependency with DPDI for different image contents at different levels of complexity. The relationship can be well accounted for by the following nonlinear mapping

$$H_C = \frac{\tau}{\log(E^{\lambda})}, \tag{4}$$

where the best parameters are found to be $\tau = 21.9$ and $\lambda = 6$ and H_C values of different depth image contents are reported in Table IX.

Any image patch can be represented in a unique and adaptive way by three conceptually independent components: mean intensity, signal strength, and signal structure [60]. This novel representation has been found to be useful in IQA of multi-exposure image fusion [61] and contrast changed images [62]. In this work, we show that this representation can well explain the distortion type dependency observations we described in Section III-D.

TABLE X
MEAN VALUES OF $\Delta\theta$ FOR DIFFERENT DISTORTION TYPES AND LEVELS

Distortion	Noise	Blur	JPEG
Level 1	83.27°	24.05°	62.87°
Level 2	81.98°	19.00°	58.02°
Level 3	80.95°	15.44°	50.74°
Level 4	79.96°	12.71°	48.59°

Given a $\sqrt{N} \times \sqrt{N}$ local image patch $\mathbf{x}$ that is represented as an N-dimensional vector, we decompose it by

$$\begin{aligned} \mathbf{x} &= \mu_{\mathbf{x}} + ||\mathbf{x} - \mu_{\mathbf{x}}|| \cdot \frac{\mathbf{x} - \mu_{\mathbf{x}}}{||\mathbf{x} - \mu_{\mathbf{x}}||} \\ &= c_1^{\mathbf{x}} \cdot \mathbf{v}_1^{\mathbf{x}} + c_2^{\mathbf{x}} \cdot \mathbf{v}_2^{\mathbf{x}}, \end{aligned} \tag{5}$$

where $||\cdot||$ denotes the l^2 norm of a vector, $\mu_{\mathbf{x}}$ is the mean intensity of the patch. $\mathbf{x}$ is now represented as a linear combination of two unit-length vectors,

$$\mathbf{v}_1^{\mathbf{x}} = \frac{1}{\sqrt{N}} \cdot \mathbf{1} \quad \text{and} \quad \mathbf{v}_2^{\mathbf{x}} = \frac{\mathbf{x} - \mu_{\mathbf{x}}}{||\mathbf{x} - \mu_{\mathbf{x}}||}, \tag{6}$$

each associated with a coefficient

$$c_1^{\mathbf{x}} = \sqrt{N}\mu_{\mathbf{x}} \quad \text{and} \quad c_2^{\mathbf{x}} = ||\mathbf{x} - \mu_{\mathbf{x}}||. \tag{7}$$

Here $\mathbf{1}$ denotes a column vector with all entries equaling 1. Since $\mathbf{v}_1^{\mathbf{x}}$ is fixed, each source patch $\mathbf{x}$ can be uniquely represented by three components $c_1^{\mathbf{x}}$, $c_2^{\mathbf{x}}$ and the unit-length vector $\mathbf{v}_2^{\mathbf{x}}$, which denote the mean intensity, signal strength and signal structure, respectively. The representation or decomposition is adaptive, where the basis $\mathbf{v}_2^{\mathbf{x}}$ points to a specific direction in the signal space and is adapted to the input signal.

Now assume $\mathbf{x}$ and $\mathbf{y}$ are the co-located patches in the reference and distorted images, respectively. Then from Eq. (5), we have $\mathbf{v}_2^{\mathbf{x}}$ and $\mathbf{v}_2^{\mathbf{y}}$, which represent the signal structures of the reference and distorted images, respectively. We denote the angle between $\mathbf{v}_2^{\mathbf{x}}$ and the structural distortion vector $(\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}})$ as $\Delta\theta$. Then $\cos\Delta\theta$ can be computed as

$$\cos\Delta\theta = \frac{|\mathbf{v}_2^{\mathbf{x}} \cdot (\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}})|}{||\mathbf{v}_2^{\mathbf{x}}||||\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}}||}, \tag{8}$$

and $\Delta\theta$ can be subsequently obtained through an arc-cosine function. Note that $\Delta\theta$ is the angle between two orientations and thus has a dynamic range between 0 and $\frac{\pi}{2}$. Table X reports the mean values of $\Delta\theta$ for each distortion type and level. Interestingly, the results show a strong distortion type dependency of $\Delta\theta$. In particular, for noise contaminated image, $\Delta\theta$ is close to $\frac{\pi}{2}$ (90°); for blurred image, $\Delta\theta$ is below $\frac{\pi}{6}$ (30°); and for JPEG compressed image, $\Delta\theta$ typically lies between $\frac{\pi}{4}$ (45°) and $\frac{\pi}{3}$ (60°).

Some qualitative explanations of this phenomenon are as follows. When left- and right-views are both noise contaminated, the distortion vectors $\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}}$ are orthogonal to the original vectors $\mathbf{v}_2^{\mathbf{x}}$, thus the original necessary information used to establish stereoscopic cues is affected by independent noise only. In this case, the impact of distortion on the depth quality is additive. As such, the gap of DPDI between noise added to one-view and two-views is much higher than those of Gaussian blur and JPEG compression because twice amount of noise is added. On the other hand, when left- and right-views

are either blurred or JPEG compressed, the distortion vectors $\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}}$ can be decomposed into two orthogonal components, one of which aligns with and the other is orthogonal to the original vector. The original necessary information used to establish stereoscopic cues is affected by not only the relative strength of these two components, but also the consistency of such relative strengths on the left- and right-views. When deterministic distortions such as blur or JPEG compression are applied equally to both views, high consistency is expected. In this case, the impact of blurriness or compression artifacts on the depth quality is more dependent on the lower quality view with more structural distortions. As such, the gap of DPDI between one-view and two-views is reduced.

The above analysis shows that this patch-structure representation provides useful cues to account for the distortion type dependency we observed in Section III-D. This inspires us to develop an objective model to automatically predict DPDI for different distortion types and levels.

Let $(\mathbf{x}_l, \mathbf{y}_l)$ and $(\mathbf{x}_r, \mathbf{y}_r)$ be the co-located patches in the reference and distorted left- and right-views images, respectively. Let d_l and d_r denote the local distortion measures for $(\mathbf{x}_l, \mathbf{y}_l)$ and $(\mathbf{x}_r, \mathbf{y}_r)$, respectively. We characterize the local measure of DPDI by

$$H_D = \left(|d_l|^p + |d_r|^p\right)^{\frac{1}{p}}, \tag{9}$$

where $p \geq 0$ is an exponent parameter. With various choices of the value of p, this general formulation leads to a family of combination rules with different physical meanings. The larger the p value, the more emphasis is put on the patches that have relatively larger distortion between left- and right-view. Specifically, $p = 1$ corresponds to length-weighted averaging; $p = 2$ corresponds to energy-weighted averaging; and $p = \infty$ corresponds to picking the patch that has the larger distortion.

It remains to determine the value of p. Instead of fixing p to be a constant for each distortion types and levels, here we propose an automatic approach that chooses p at each spatial location adaptively. From Eq. (5), we have $(\mathbf{v}_2^{\mathbf{x}_l}, \mathbf{v}_2^{\mathbf{y}_l})$ and $(\mathbf{v}_2^{\mathbf{x}_r}, \mathbf{v}_2^{\mathbf{y}_r})$ for left-view and right-view, respectively. We denote $\Delta\theta_l$ as the angle between $\mathbf{v}_2^{\mathbf{x}_l}$ and $\mathbf{v}_2^{\mathbf{y}_l} - \mathbf{v}_2^{\mathbf{x}_l}$ and $\Delta\theta_r$ as the angle between $\mathbf{v}_2^{\mathbf{x}_r}$ and $\mathbf{v}_2^{\mathbf{y}_r} - \mathbf{v}_2^{\mathbf{x}_r}$. $\cos\Delta\theta_l$ and $\cos\Delta\theta_r$ are computed using Eq. (8), then p is determined by

$$p = (1 + \cos\Delta\theta_l + \cos\Delta\theta_r)^2. \tag{10}$$

In particular, when both views are noise contaminated, $\Delta\theta_l$ and $\Delta\theta_r$ are close to $\frac{\pi}{2}$, $\cos\Delta\theta_l$ and $\cos\Delta\theta_r$ are close to 0, and thus p is close to 1, then we have

$$H_D \simeq (|d_l| + |d_r|); \tag{11}$$

when only one view (e.g. left-view) is noise contaminated and the other one (e.g. right-view) is pristine, $\Delta\theta_r$ is 0 and $\cos\Delta\theta_r$ is 1, p goes relatively larger, then we have

$$H_D \simeq \max\{|d_l|, |d_r|\}. \tag{12}$$

When both views are blurred or JPEG compressed, $\Delta\theta_l$ and $\Delta\theta_r$ are close to 0 or around $\frac{\pi}{4}$, $\cos\Delta\theta_l$ and $\cos\Delta\theta_r$ are close to 1 (for higher JPEG compression levels), thus p is relatively large; when only one view (e.g. left-view) is blurred or JPEG compressed and the other one (e.g. right-view) is pristine, $\Delta\theta_r$ is 0 and $\cos\Delta\theta_r$ is 1, and thus p is also a larger number. In both cases, we have

$$H_D \simeq \max\{|d_l|, |d_r|\}. \tag{13}$$

As such, the value of p is automatically determined, without recognizing the distortion types explicitly.

Once the value of p is determined at each spatial location, the local H_D measure is computed using Eq. (9). The global H_D measure is the average of the local H_D across all spatial locations. Finally, the three components, H_L, H_C and H_D are combined to yield an overall DPDI prediction

$$H = H_L \cdot H_C \cdot H_D. \tag{14}$$

B. Validation

We use the new Waterloo-IVC 3D Depth database to test the proposed DPDI prediction model. First, DPDI predictions from H_D only are computed for each depth level and each image content. PLCC, SRCC, and KRCC between the observed and the predicted DPDI values are reported in Table XI, where the results are summarized as the average performance for each image content. The direct averaging (Ave.) method corresponds to the case of $p = 1$ in Eq. (9), while in the adaptive-p (Adpt.) method the value of p is adaptively determined using Eq. (10) to Eq. (13). PSNR, SSIM, MS-SSIM, information content weighted SSIM (IW-SSIM) [63], Feature SIMilarity (FSIM) [64], and VIF are employed to create the base single-view distortion measurements d, where we let $d = 50 - \text{PSNR}$ and $d = 1 - \text{SSIM}$, MS-SSIM, IW-SSIM, FSIM or VIF. For fairness, a global approach to compute H_D is adopted, i.e., an average spatial pooling on $\cos\Delta\theta_l$ and $\cos\Delta\theta_r$ is applied to the left- and right-view, respectively, and thus p is determined globally. From Table XI, it can be observed that the adaptive-p model outperforms the direct averaging method in almost all cases. In addition, MS-SSIM, IW-SSIM, and VIF achieve larger improvements than PSNR, SSIM, and FSIM. Considering the performance and computational complexity, MS-SSIM is chosen as the distortion measurement method in the subsequent tests.

Table XII shows PLCC, SRCC, and KRCC results for DPDI predictions from all individuals and combinations of H_L, H_C, and H_D for all stereopairs and each distortion type. Note that H_L and H_C are pre-determined using Eqs. (3) and (4) in Section IV-A, but are completely independent of the following tests with H_D and their combinations. It can be seen that DPDI prediction performance from H_L only and H_C only are similar, which indicates that depth level and image content are about equally important to DPDI estimation, and their combination, not surprisingly, provides a relatively better DPDI prediction performance. It can also be observed that DPDI predictions from H_D only are not as good as those from H_L only or H_C only, even though the adaptive-p method reduces the prediction bias. When H_L, H_C, and their combination are combined with H_D, significant improvements are obtained, but in the case of using the adaptive-p method only and not in the case of direct averaging.

TABLE XI

PERFORMANCE COMPARISON OF DPDI ESTIMATIONS USING DIFFERENT BASE 2D DISTORTION MEASURES. AVE.: DIRECT AVERAGING; ADPT.: PROPOSED ADAPTIVE-p MODEL

PLCC														
	Bark		Brick		Flower		Food		Grass		Water		Average	
2D-IQA	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.
PSNR	0.3770	0.5700	0.1643	0.3330	0.5028	0.7621	0.2673	0.4848	0.3961	0.7177	0.2358	0.3168	0.3239	0.5307
SSIM	0.5272	0.6458	0.2937	0.4457	0.5182	0.7420	0.3423	0.5054	0.4638	0.6152	0.2167	0.3275	0.3937	0.5469
MS-SSIM	0.6046	0.7365	0.4139	0.6013	0.6435	0.7647	0.4364	0.5575	0.5932	0.6964	0.4426	0.5519	0.5224	**0.6514**
IW-SSIM	0.5921	0.7061	0.5521	0.6230	0.6296	0.7079	0.3647	0.5236	0.5849	0.6639	0.5575	0.5645	0.5468	**0.6315**
FSIM	0.5668	0.6716	0.4116	0.5065	0.4448	0.4221	0.2210	0.4970	0.5283	0.5244	0.4324	0.4958	0.4342	0.5196
VIF	0.5717	0.6308	0.4566	0.5259	0.5397	0.7874	0.2897	0.4684	0.5307	0.7805	0.5434	0.5728	0.4886	**0.6277**
SRCC														
2D-IQA	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.
PSNR	0.3925	0.5563	0.1781	0.3029	0.5787	0.7104	0.2834	0.4101	0.5131	0.5942	0.2140	0.1862	0.3600	0.4600
SSIM	0.4654	0.6142	0.1753	0.2984	0.4771	0.6037	0.2566	0.3501	0.3319	0.4016	0.2784	0.2253	0.3308	0.4156
MS-SSIM	0.5588	0.6587	0.3277	0.4830	0.5785	0.6310	0.3013	0.3679	0.4881	0.5001	0.3924	0.4207	0.4411	**0.5102**
IW-SSIM	0.5133	0.6204	0.3673	0.4840	0.5552	0.6000	0.2637	0.3254	0.4868	0.4978	0.5094	0.5443	0.4493	**0.5120**
FSIM	0.4804	0.6271	0.2706	0.4383	0.2764	0.3080	0.1434	0.2643	0.2518	0.1785	0.3322	0.3523	0.2925	0.3614
VIF	0.4293	0.5706	0.3228	0.4563	0.5325	0.6505	0.2508	0.3372	0.4404	0.5255	0.5068	0.5837	0.4138	**0.5206**
KRCC														
2D-IQA	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.
PSNR	0.2969	0.4216	0.1364	0.2184	0.4389	0.5546	0.2104	0.2985	0.3839	0.4468	0.1606	0.1411	0.2712	0.3468
SSIM	0.3470	0.4710	0.1360	0.2165	0.3519	0.4563	0.1834	0.2530	0.2497	0.2988	0.2165	0.1790	0.2474	0.3124
MS-SSIM	0.4157	0.5017	0.2391	0.3550	0.4328	0.4804	0.2232	0.2725	0.3596	0.3740	0.2955	0.3099	0.3276	**0.3822**
IW-SSIM	0.3779	0.4666	0.2652	0.3559	0.4108	0.4553	0.1912	0.2405	0.3584	0.3712	0.3903	0.4138	0.3323	**0.3839**
FSIM	0.3476	0.4716	0.2009	0.3189	0.2189	0.2390	0.1084	0.1925	0.1930	0.1558	0.2480	0.2588	0.2195	0.2728
VIF	0.3196	0.4215	0.2436	0.3259	0.3996	0.4995	0.1865	0.2520	0.3298	0.3947	0.3945	0.4528	0.3123	**0.3911**

TABLE XII

PERFORMANCE COMPARISON OF DPDI ESTIMATIONS USING DIFFERENT COMBINATIONS OF PREDICTION COMPONENTS

	All			Noise			Blur			JPEG		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
H_L	0.5773	0.6160	0.4576	0.6253	0.6686	0.5034	0.6281	0.6676	0.5051	0.4988	0.5752	0.4291
H_C	0.5452	0.5331	0.4087	0.3433	0.3117	0.2338	0.6200	0.6168	0.4816	0.6832	0.6663	0.5153
$H_L + H_C$	0.7278	0.7393	0.5680	0.6840	0.6972	0.5307	0.7995	0.8072	0.6374	0.7543	0.7616	0.5929
H_D for $p = 1$	0.3189	0.3441	0.2390	0.4102	0.3508	0.2471	0.1913	0.1748	0.1232	0.5022	0.4899	0.3427
H_D for adaptive p	0.4077	0.3917	0.2716	0.4040	0.3931	0.2797	0.2313	0.2307	0.1627	0.6112	0.5853	0.4178
$H_L + H_D$ for $p = 1$	0.5885	0.5873	0.4212	0.6933	0.6853	0.5054	0.4885	0.4952	0.3475	0.6722	0.6755	0.4902
$H_L + H_D$ for adaptive p	0.6555	0.6583	0.4810	0.7203	0.7216	0.5403	0.6160	0.6169	0.4424	0.7366	0.7421	0.5534
$H_C + H_D$ for $p = 1$	0.5173	0.5249	0.3724	0.5266	0.5101	0.3691	0.4392	0.4165	0.2959	0.6326	0.6262	0.4493
$H_C + H_D$ for adaptive p	0.6258	0.6299	0.4534	0.5489	0.5297	0.3873	0.6455	0.6262	0.4580	0.7151	0.6878	0.5090
$H_L + H_C + H_D$ for $p = 1$	0.7104	0.7160	0.5327	0.8046	0.7861	0.6072	0.6698	0.6758	0.4959	0.7713	0.7761	0.5844
$H_L + H_C + H_D$ for adaptive p	0.7970	0.8018	0.6190	0.8118	0.7964	0.6199	0.8205	0.8248	0.6452	0.8415	0.8404	0.6577

TABLE XIII

PERFORMANCE COMPARISON OF DPDI ESTIMATIONS FOR DIFFERENT DEPTH LEVELS

Depth Levels	Level 1			Level 2			Level 3		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
$H_L + H_C + H_D$ for $p = 1$	0.4830	0.5664	0.4342	0.6630	0.6470	0.4805	0.6786	0.6714	0.4959
$H_L + H_C + H_D$ for adaptive p	0.7294	0.6696	0.5280	0.8045	0.8182	0.6398	0.8140	0.8197	0.6379
Depth Levels	Level 4			Level 5			Level 6		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
$H_L + H_C + H_D$ for $p = 1$	0.6435	0.6218	0.4485	0.6564	0.6403	0.4586	0.6258	0.5980	0.4292
$H_L + H_C + H_D$ for adaptive p	0.8201	0.7873	0.5914	0.8288	0.7956	0.5926	0.8224	0.7765	0.5791

The best prediction happens in the case that all DPDI prediction components are included. For all images and each distortion type, the proposed method, when combined with MS-SSIM as the base 2D distortion measure, without attempting to recognize the distortion types or giving any specific treatment for any specific distortion type, leads to highly promising DPDI prediction performance.

Moreover, Tables XIII and XIV report PLCC, SRCC, and KRCC values of the overall DPDI predictions for different depth levels and different image contents, respectively. Interestingly, it can be observed that improvements are most pronounced at the middle depth levels (Level 2 and Level 3) or at the textural contents of middle complexity (Bark and Flower), which indicates that the impact of symmetric and asymmetric distortions on the perception of depth is more significant in these "middle" cases.

V. DISCUSSIONS

A. Comparison Between Subjective Studies I and II

A main issue with the traditional subjective testing approaches such as that used in our Subjective Study I (Section II) is the difficulty in singling out the contribution of stereo cues in depth perception, and the subjective scores collected through such experiments show strong correlations between 3D image quality and depth quality scores, even though they are substantially different perceptual attributes. The second subjective testing method introduced in our

TABLE XIV
PERFORMANCE COMPARISON OF DPDI ESTIMATIONS FOR DIFFERENT IMAGE CONTENTS

Image Contents	Bark			Brick			Flower		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
$H_L + H_C + H_D$ for $p = 1$	0.7009	0.7026	0.5190	0.5747	0.5741	0.4198	0.7190	0.7248	0.5392
$H_L + H_C + H_D$ for adaptive p	0.8288	0.8267	0.6382	0.6299	0.6620	0.4933	0.8146	0.8025	0.6169
Image Contents	Food			Grass			Water		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
$H_L + H_C + H_D$ for $p = 1$	0.6104	0.6035	0.4384	0.6529	0.6302	0.4610	0.6745	0.5991	0.4589
$H_L + H_C + H_D$ for adaptive p	0.7645	0.7572	0.5664	0.7490	0.7073	0.5312	0.6711	0.6274	0.4930

TABLE XV
CORRELATIONS BETWEEN 3DIQ AND DPDI PREDICTIONS

Depth Levels	PLCC	SRCC	KRCC
Level 1	0.5680	0.5512	0.4020
Level 2	0.5673	0.5496	0.4014
Level 3	0.5669	0.5509	0.4011
Level 4	0.5498	0.5274	0.3842
Level 5	0.5521	0.5202	0.3782
Level 6	0.4989	0.4476	0.3333
Average	0.5505	0.5245	0.3834

Subjective Study II (Section III) is an attempt to overcome this problem. To observe it more closely, for each stereopair on the new Waterloo-IVC 3D Depth Database, we estimate its 3DIQ using the binocular rivalry-inspired weighting method presented in [8] and its DPDI using the proposed DPDI prediction model, respectively. Table XV shows the PLCC, SRCC and KRCC values between the predicted 3DIQ and DPDI. It is important to note that the correlations between 3DIQ and DPDI predictions are relatively low, which is quite different from the observations in the first subjective test we discussed in Section II-C. This result suggests that our new subjective testing approach is able to provide more independent information on the depth perception aspect of 3D visual perception. A comprehensive comparison between Subjective Study I and Subjective Study II is provided in Table XVI.

B. Perceptually-Driven Asymmetrical Stereoscopic Video Coding

In Section II-C, we described that there exists a strong distortion type dependency with 3D image quality [8]. Then in Section III-D, a different distortion type dependency in depth perception has been discovered. The discovery of such a distortion type dependency in depth perception not only has scientific values in understanding depth perception in the HVS, but is also desirable in the practice of 3D video compression and transmission. The distortions involved in 3D video coding/communication are not only compression artifacts. The practical encoder/decoder also needs to decide on whether inloop/out-of-loop deblocking filters need to be turned on, and whether mixed-resolutions of the left/right-views should be used. Mixed-resolution coding, asymmetric transform-domain quantization coding, and postprocessing techniques (deblocking or blurring) can be employed individually or collectively. Previously in [65]–[67], the extent of the downsampling ratio that can be applied to a low quality view without a noticeable degradation on the 3D quality has been investigated. In [67], symmetric stereoscopic video coding, asymmetric quantization coding and mixed-resolution coding have been compared and the results suggested that mixed-resolution coding achieves the best coding efficiency. In [68], different levels of Gaussian blurring are applied after asymmetric quantization and a significant bit rate reduction has been achieved for this joint asymmetric compression and postprocessing method. However, here our new observations indicate that asymmetric compression and asymmetric blurring will influence the perceived 3D depth quality, i.e., adding blur or JPEG compression to one view of stereopair has similar effect in depth perception as adding the same level of distortion to both views. This is quite different from the distortion type dependency in 3D image quality perception. Therefore, the current study suggests that mixed-resolution coding, asymmetric transform-domain quantization coding, and postprocessing schemes need to be carefully reexamined and redesigned to maintain a good tradeoff between perceptual 3D image quality and depth quality. One possible solution is that a threshold on H may be used as a constraint in the process of asymmetrical bit allocation, ensuring that the quality of depth perception will not be severely affected. In the rest of this section, we will demonstrate how to use our current findings to guide the asymmetric transform-domain quantization coding with low-pass postprocessing filtering.

In [68], we found that the prediction of stereoscopic 3D video quality can be calculated by a weighted average of the left- and right-view video quality

$$Q^{3D} = w_l Q_l^{2D} + w_r Q_r^{2D}, \tag{15}$$

where w_l and w_r are determined by the relative energy of the two views [8].

The fundamental issue in stereoscopic video compression is to obtain the best tradeoff between the total rate of both left- and right-views and the perceived distortion. With the distortion model in Eq. (15), such a rate distortion optimization (RDO) problem can be expressed as

$$\max \left\{ w_l Q_l^{2D} + w_r Q_r^{2D} \right\} \text{ subject to } R_l + R_r \leq R_c, \tag{16}$$

where $w_l + w_r = 1$, and R_l and R_r represent the bit rates of the left- and right-views, respectively. The major difference between stereoscopic and monoscopic video coding lies in the bit allocation between the two views for maximal stereoscopic quality.

Recall the distortion model in Eq. (15), which can be rewritten as

$$\begin{aligned} Q^{3D} &= w_l Q_l^{2D} + (1 - w_l) Q_r^{2D} \\ &= \underbrace{\frac{Q_l^{2D} + Q_r^{2D}}{2}}_{Quality\ Average} + \underbrace{(w_l - \frac{1}{2})(Q_l^{2D} - Q_r^{2D})}_{Quality\ Divergence}, \end{aligned} \tag{17}$$

TABLE XVI
COMPARISON BETWEEN SUBJECTIVE STUDY I AND SUBJECTIVE STUDY II

	Subjective Study I - Traditional	**Subjective Study II - New**
Image Database	Waterloo-IVC 3D Image Database: Noise, Blur, and JPEG symmetric and asymmetric distortions	Waterloo-IVC 3D Depth Database: Noise, Blur, and JPEG symmetric and asymmetric distortions
Visual Experience Criteria	Depth Quality: The amount, naturalness and clearness of depth perception experience	Identify and label depth polarizations for inner images, outer images, and flat images
Depth Cues	Monocular and binocular depth cues together	Binocular depth cues only
Subjective Response	ACR with an 11-grade numerical categorial scale	Inner, outer, flat, and unable to decide
Subjective Scores	MOS DQ: Mean opinion scores for Depth Quality	DPDI: Depth Perception Difficulty Index
Relations to 3D Image Quality (3DIQ)	Very high, Refer to Fig. 5	Relatively low, Refer to Table XV
Distortion Type Dependency	3DIQ(DQ): Noise and JPEG, $\rightarrow$ lower quality view; Blur, $\rightarrow$ higher quality view	DPDI: Noise: Additive; Blur and JPEG, $\rightarrow$ lower quality view

where without loss of generality, left-view is denoted as the higher quality view.

This suggests that the quality measure is composed of two terms, which are the quality average and quality divergence of the two views. If we ignore the quality divergence term by setting ($w_l = w_r = 0.5$) and assume that the two views are independent of each other, borrowing bits from one view to the other would not be wise because the performance gain for the high quality view would not be able to compensate for the loss in the low quality view. This suggests that the quality divergence term plays a crucial role in asymmetrical coding.

The underlying principle in the quality measure Eq. (15) is that the view with higher energy dominates the final 3D visual quality. Therefore, it becomes natural to allocate more bits to one view and perform low-pass filtering to the other when the bit budget cannot support both views to be coded at the high quality level. As a result, the divergence term in Eq. (17) is increased at the expense of the quality dropping on the average quality. Although blurring artifacts already exist in the compressed video, the low-pass postprocessing filtering is still necessary as the blocking artifacts are also very significant in low bit rate coding. After low-pass filtering, the view with lower quality has smaller energy. Consequently, the overall quality approaches the high quality view, leading to a significant improvement on the final stereoscopic quality.

It is worth noting that though the quality divergence term is maximized when more coding bits are allocated to the high quality view, the average quality drops. Therefore, optimal asymmetrical bit allocation does not necessarily mean that all bits should be allocated to the high quality view. Finding the best tradeoff is desirable. More importantly, it is recognized that introducing blur artifacts will lead to degradations on depth quality, which motivates us to add a threshold on H to control the maximum allowed DPDI as another constraint. This can be used to determine the low-pass postprocessing filtering level.

Given the inherent disparity exhibited in the original stereopairs, the relative DPDI change is computed as

$$\frac{H}{H_{\mathrm{L}}} = H_{\mathrm{C}} \cdot H_{\mathrm{D}}, \tag{18}$$

where H_C is computed from Eq. (4) and H_D is estimated based on the quality of the left- and right-view Q_l^{2D} and Q_r^{2D}. For the compression artifacts induced by HEVC and the blurriness induced by low-pass postprocessing filtering, based on Eq. (13), we estimate H_D by

$$H_D \simeq \max\left\{1 - Q_l^{2D}, 1 - Q_r^{2D}\right\} = 1 - Q_r^{2D}, \tag{19}$$

where without loss of generality, Q_l^{2D} and Q_r^{2D} is assumed to be bounded between 0 and 1.

Finally, a threshold ΔH_{th} is imposed on the relative depth variation to ensure that a reasonable depth quality is preserved

$$\frac{H}{H_{\mathrm{L}}} = H_{\mathrm{C}} \cdot H_{\mathrm{D}} < \Delta H_{th}. \tag{20}$$

It should also be noted that when the overall bit budget is high enough to support both views coded at high quality, the postprocessing becomes unnecessary. Consequently, before bit allocation, the bit budget is examined in terms of bits/pixel to determine whether the proposed scheme should be performed.

VI. CONCLUSIONS

We have carried out two subjective studies on depth perception of stereoscopic 3D images. The first one follows a traditional framework where subjects are asked to rate depth quality directly on distorted stereopairs. The second one uses a novel approach, where the stimuli are synthesized independent of the background image content and the subjects are asked to identify depth changes and label the polarities of depth. Our analysis shows that the second approach is much more effective at singling out the contributions of stereo cues in depth perception, through which we have several interesting findings regarding distortion type dependency, image content dependency, and the impact of symmetric and asymmetric distortions on the perception of depth. Furthermore, we propose a novel computational model for DPDI prediction. Our results show that the proposed model, without explicitly identifying image distortion types, leads to highly promising DPDI prediction performance. We believe these findings provide useful insights in the future development of comprehensive 3D QoE models for stereoscopic images, which have great potentials in real-world applications such as asymmetric compression of stereoscopic 3D videos.

REFERENCES

[1] C.-C. Su, A. K. Moorthy, and A. C. Bovik, "Visual quality assessment of stereoscopic image and video: Challenges, advances, and future trends," in *Visual Signal Quality Assessment*. 2015, pp. 185–212.

[2] M. Barkowsky *et al.*, "Subjective and objective visual quality assessment in the context of stereoscopic 3D-TV," in *3D-TV System With Depth-Image-Based Rendering*. New York, NY, USA: Springer, 2013, pp. 413–437.

[3] L. M. J. Meesters, W. A. Ijsselsteijn, and P. J. H. Seuntiëns, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 381–391, Mar. 2004.

[4] P. J. H. Seuntiëns, "Visual experience of 3D TV," Ph.D. dissertation, Faculty Technol. Manage., Eindhoven Univ. Technol., Eindhoven, The Netherlands, 2006.

[5] *Subjective Assessment Methods for 3D Video Quality*, document ITU-T P.915, Mar. 2016. [Online]. Available: https://www.itu.int/rec/T-REC-P.915/en

[6] *Information and Guidelines for Assessing and Minimizing Visual Discomfort and Visual Fatigue From 3D Video*, document P.916, Mar. 2016. [Online]. Available: https://www.itu.int/rec/T-REC-P.916/en

[7] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, and Q. Dai, "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2971–2983, Oct. 2015.

[8] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.

[9] F. Shao, W. Tian, W. Lin, G. Jiang, and Q. Dai, "Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2059–2074, May 2016.

[10] N. S. Holliman, B. Froner, and S. P. Liversedge, "An application driven comparison of depth perception on desktop 3D displays," *Proc. SPIE*, vol. 6490, p. 64900H, Jan. 2007.

[11] A. J. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," *Proc. SPIE*, vol. 1915, pp. 36–48, Sep. 1993.

[12] W. Chen, F. Jérôme, M. Barkowsky, and P. Le Callet, "Exploration of quality of experience of stereoscopic images: Binocular depth," in *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron*, Scottsdale, AZ, USA, Jan. 2012, pp. 116–121.

[13] R. G. Kaptein, A. Kuijsters, M. T. M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx, "Performance evaluation of 3D-TV systems," *Proc. SPIE*, vol. 6808, p. 680819, Jan. 2008.

[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. Zwicker, S. Yea, A. Vetro, C. Forlines, W. Matusik, and H. Pfister, "Display pre-filtering for multi-view video compression," in *Proc. IEEE Int. Conf. Multimedia Expo*, Sep. 2007, pp. 1046–1053.

[16] A. Schertz, "Source coding of stereoscopic television pictures," in *Proc. IEEE Int. Conf. Image Process.*, Maastricht, The Netherlands, Apr. 1992, pp. 462–464.

[17] P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet, "Perceptual preference of S3D over 2D for HDTV in dependence of video quality and depth," in *Proc. IVMSP Workshop, 3D Image/Video Technol. Appl.*, 2013, pp. 1–4.

[18] F. Shao, Q. Jiang, R. Fu, M. Yu, and G. Jiang, "Optimizing visual comfort for stereoscopic 3D display based on color-plus-depth signals," *Opt. Exp.*, vol. 24, no. 11, pp. 11640–11653, May 2016.

[19] V. Kulyk, S. Tavakoli, M. Folkesson, K. Brunnström, K. Wang, and N. Garcia, "3D video quality assessment with multi-scale subjective method," in *Proc. Int. Workshop Quality Multimedia Exper.*, Klagenfurt, Austria, Jul. 2013, pp. 106–111.

[20] W. A. IJsselsteijn, H. D. Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 225–233, Mar. 2000.

[21] M. Lambooij, W. IJsselsteijn, D. G. Bouwhuis, and I. Heynderickx, "Evaluation of stereoscopic images: Beyond 2D quality," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 432–444, Jun. 2011.

[22] 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. Interpretation*, Santa Fe, NM, USA, Apr. 2012, pp. 173–176.

[23] 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.

[24] 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.

[25] 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.

[26] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, p. 659024, Oct. 2008.

[27] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 389–392.

[28] 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. Barcelona, Spain, Sep. 2003, pp. 185–188.

[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. Quality Metrics Consum. Electron*, Scottsdale, AZ, USA, Jan. 2010, pp. 61–66.

[30] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2003, pp. 1398–1402.

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

[32] 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.

[33] 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.

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

[35] J. Yang, C. Hou, R. Xu, and J. Lei, "New metric for stereo image quality assessment based on HVS," *Int. J. Imag. Syst. Technol.*, vol. 20, no. 4, pp. 301–307, Nov. 2010.

[36] 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.

[37] 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, Dec. 2013.

[38] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379–3391, Sep. 2013.

[39] 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.

[40] K. Wang, M. Barkowsky, K. Brunnstrom, M. Sjöström, R. Cousseau, and P. Le Callet, "Perceived 3D TV transmission quality assessment: Multi-laboratory results using absolute category rating on quality of experience scale," *IEEE Trans. Broadcast.*, vol. 58, no. 4, pp. 544–557, Dec. 2012.

[41] J. Wang and Z. Wang, "Perceptual quality of asymmetrically distorted stereoscopic images: The role of image distortion types," in *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron.*, Chandler, AZ, USA, Jan. 2014, pp. 1–6.

[42] J. Wang, K. Zeng, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic images from single views," in *Proc. IEEE Int. Conf. Multimedia Expo*, Chengdu, China, Jul. 2014, pp. 1–6.

[43] 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. Perception*, vol. 3, no. 2, pp. 95–109, Apr. 2006.

[44] A. M. van Dijk, J.-B. Martens, and A. B. Watson, "Quality asessment of coded images using numerical category scaling," *Proc. SPIE*, vol. 2451, pp. 90–101, Feb. 1995.

[45] VQEG. (Apr. 2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. [Online]. Available: http://www.vqeg.org

[46] M. Mehrabi, E. M. Peek, B. C. Wuensche, and C. Lutteroth, "Making 3D work: A classification of visual depth cues, 3D display technologies and their applications," in *Proc. 14th Austral. User Interface Conf.*, 2013, pp. 91–100.

[47] D. F. McAllister, *Stereo Computer Graphics and Other True 3D Technologies*. Princeton, NJ, USA: Princeton Univ. Press, 1993.

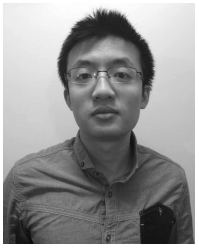
[48] T. Okoshi, *Three-Dimensional Imaging Techniques*. Amsterdam, The Netherlands: Elsevier, 2012.

[49] *The Vision Texture Database*. (Dec. 2002). [Online]. Available: http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html

[50] C. M. Zaroff, M. Knutelska, and T. E. Frumkes, "Variation in stereoacuity: Normative description, fixation disparity, and the roles of aging and gender," *Invest. Ophthalmol. Vis. Sci.*, vol. 44, no. 2, pp. 891–900, 2003.
[51] C. M. Schor and I. Wood, "Disparity range for local stereopsis as a function of luminance spatial frequency," *Vis. Res.*, vol. 23, no. 12, pp. 1649–1654, 1983.
[52] R. T. Goodwin and P. E. Romano, "Stereoacuity degradation by experimental and real monocular and binocular amblyopia," *Invest. Ophthalmol. Vis. Sci.*, vol. 26, no. 7, pp. 917–923, 1985.
[53] R. F. Hess, C. H. Liu, and Y. Z. Wang, "Differential binocular input and local stereopsis," *Vis. Res.*, vol. 43, no. 22, pp. 2303–2313, 2003.
[54] A. Z. Khan and J. D. Crawford, "Ocular dominance reverses as a function of horizontal gaze angle," *Vis. Res.*, vol. 41, no. 14, pp. 1743–1748, Jun. 2001.
[55] O. Rosenbach, "On monocular prevalence in binocular vision," *Med. Wochenschrift*, vol. 50, pp. 1290–1292, 1903.
[56] R. Blake, "A primer on binocular rivalry, including current controversies," *Brain Mind*, vol. 2, no. 1, pp. 5–38, Apr. 2001.
[57] A. P. Mapp, H. Ono, and R. Barbeito, "What does the dominant eye dominate? A brief and somewhat contentious review," *Perception Psychophys.*, vol. 65, no. 2, pp. 310–317, Feb. 2003.
[58] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.
[59] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. San Diego, CA, USA, Jun. 2005, pp. 807–814.
[60] K. Ma and Z. Wang, "Multi-exposure image fusion: A patch-wise approach," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, QC, Canada, Sep. 2015, pp. 1–5.
[61] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
[62] S. Wang, K. Ma, H. Yeganeh, Z. Wang, and W. Lin, "A patch-structure representation method for quality assessment of contrast changed images," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
[63] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
[64] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[65] L. Stelmach, W. J. Tam, D. Meegan, and A. Vincent, "Stereo image quality: Effects of mixed spatio-temporal resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 188–193, Mar. 2000.
[66] H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video*, Potsdam, Germany, May 2009, pp. 1–4.
[67] P. Aflaki, M. M. Hannuksela, and M. Gabbouj, "Subjective quality assessment of asymmetric stereoscopic 3D video," *Signal, Image Video Process.*, vol. 9, no. 2, pp. 331–345, Mar. 2015.
[68] J. Wang, S. Wang, and Z. Wang, "Quality prediction of asymmetrically compressed stereoscopic videos," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, QC, Canada, Sep. 2015, pp. 1–5.

Jiheng Wang (S'11–M'17) received the M.Math. degree in statistics-computing and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, ON, Canada, in 2011 and 2016, respectively. In 2013, he was with the Video Compression Research Group, Blackberry, Waterloo. He is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering from the University of Waterloo. His current research interests include 3D image and video quality assessment, perceptual 2D and 3D video coding, biomedical signal processing, statistical learning, and dimensionality reduction.

Shiqi Wang (M'15) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008, and the Ph.D. degree in computer application technology from Peking University in 2014. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. He is currently a Research Fellow with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. His research interests lie in image and image/video coding, processing, quality assessment, and analysis.



Kede Ma (S'13) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. His research interests lie in perceptual image processing and computational photography.

Zhou Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree from The University of Texas at Austin in 2001. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He has authored over 100 publications in these fields with over 30 000 citations (Google Scholar). His research interests include image processing, coding, and quality assessment, computational vision and pattern analysis, multimedia communications, and biomedical signal processing.

Dr. Wang served as a member of the IEEE Multimedia Signal Processing Technical Committee (2013–2015), an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–2014), the *Pattern Recognition* since 2006, and the IEEE SIGNAL PROCESSING LETTERS (2006–2010), and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2013–2014 and 2007–2009). He serves as a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING since 2015 and an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY since 2016. He is a fellow of Canadian Academy of Engineering. He was a recipient of the 2015 Primetime Engineering Emmy Award, the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Magazine Best Paper Award, the 2009 IEEE Signal Processing Society Best Paper Award, the 2009 Ontario Early Researcher Award, and the ICIP 2008 IBM Best Student Paper Award (as senior author).