Volume 28 Issue 8 September 2013 ISSN 0923-5965

ELSEVIER

SIGNAL PROCESSING
IMAGE
COMMUNICATION

Special Issue on Biologically Inspired Approaches for Visual Information Processing and Analysis
Guest Editors: Azeddine Beghdadi, Abdesselam Bouzerdoum,
Khan M. Iftekharuddin, Mohamed-Chaker Larabi

Theory, Techniques & Applications
A publication of the European Association to Signal Processing (EURASIP)

Available online at www.sciencedirect.com
ScienceDirect

EURASIP

Signal Processing: *Image Communication* 28 (2013) 870–883

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Subjective evaluation of stereoscopic image quality

Anush Krishna Moorthy *, Che-Chun Su, Anish Mittal, Alan Conrad Bovik

Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, USA

ARTICLE INFO

Available online 7 September 2012

Keywords:
Stereoscopic quality
3D quality
Quality assessment datatabase

ABSTRACT

Stereoscopic/3D image and video quality assessment (IQA/VQA) has become increasing relevant in today's world, owing to the amount of attention that has recently been focused on 3D/stereoscopic cinema, television, gaming, and mobile video. Understanding the quality of experience of human viewers as they watch 3D videos is a complex and multi-disciplinary problem. Toward this end we offer a holistic assessment of the issues that are encountered, survey the progress that has been made towards addressing these issues, discuss ongoing efforts to resolve them, and point up the future challenges that need to be focused on. Important tools in the study of the quality of 3D visual signals are databases of 3D image and video sets, distorted versions of these signals and the results of large-scale studies of human opinions of their quality. We explain the construction of one such tool, the LIVE 3D IQA database, which is the first publicly available 3D IQA database that incorporates 'true' depth information along with stereoscopic pairs and human opinion scores. We describe the creation of the database and analyze the performance of a variety of 2D and 3D quality models using the new database. The database as well as the algorithms evaluated are available for researchers in the field to use in order to enable objective comparisons of future algorithms. Finally, we broadly summarize the field of 3D QA focusing on key unresolved problems including stereoscopic distortions, 3D masking, and algorithm development.

1. Introduction, definitions and previous work

1.1. Introduction

The field of automatic quality assessment (QA) of 2D images and videos has seen tremendous activity in the past decade, with many successful algorithms being proposed [1–5]. The topic of QA of 3D images however, remains relatively un-explored. This is partially because until recently, commercially available 3D presentations were difficult to view (think red–green glasses) and were often synonymous with headaches and nausea, making their acceptance difficult. However, greatly improved capture and display technologies, along with tremendously successful commercial cinematic releases have put 3D back on the map. For example, in 2010, the total number of 3D movies that reached the silver screen was estimated to be thrice the number released in 2007 [6]. Apart from movies on the big screen, there is a glut of non-cinematic 3D content that is making its way to the consumer, especially over wireless networks such as 3D on mobile devices [7], 3D TVs, IPTV and 3D broadcast (e.g., ESPN 3D, Sony, Imax, Discovery, etc.). Further, given the expected future growth of video on mobile devices (as much as 50× over the next few years [8]), and mobile devices capable of producing and displaying stereoscopic content (for example the recently launched HTC EVO 3D [9]), non-cinematic 3D is becoming increasingly relevant. As Intel CEO P. Otellini stated at the 2010 Consumer Electronics Show (CES)—"3D . . . is the next thing that's poised to explode in the home".

Commercially at least, 3D content has begun to permeate everyday life. Unfortunately, 3D movies are not universally loved, indeed, many critics and artists have labeled 3D as unwatchable, predicting its eventual death [10,11]. The major reasons for this attitude include reports of 3D movies inducing nausea and headaches, distortions

* Corresponding author.
E-mail address: anushmoorthy@gmail.com (A.K. Moorthy).

0923-5965/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2012.08.004

[7], poor quality 'post-production' 3D, perceived 'dimness' [10,11] and so on.[1] Thus, even though 3D images seem to have a buzz around them today, our understanding of the many aspects of the 3D quality of experience is still lacking.

In the immersive 3D realm, the term 'quality of experience' is used to capture the wide gamut of factors that contribute to the overall palatability of the 3D visual signal. We will touch upon these factors, but our focus will be on 3D image quality assessment (IQA), or automatic measurement of the quality of *distorted* images relative to human subjective opinions of visual quality. While quality of experience is of interest, the growing non-cinematic nature of stereoscopic presentations implies that humans will view increasing amounts of compressed 3D streams that are transmitted over lossy networks such as IP or wireless [8]. The presence of distortions in stereoscopic content, either owing to the compression employed, or the transmission loss, will definitely degrade the viewing experience, and it is of immediate importance to understand how such degradations affect the palatability of the presentation. In addition, as vision scientists, we subscribe to the notion that when there are multiple complex factors (in this case, stereography, 3D display, geometry, and distortions) contributing to a perception problem (in this case, 3D QoE), all very poorly understood, it is best to attempt to isolate and study each factor before proceeding towards formulating an explanatory theory of the overall problem. In the study reported here, we focus on the perception of quality as it is affected by the image distortion, setting aside issues such as camera placement, and 3D display considerations.

Historically, QA algorithms are generally classified as (1) full-reference (FR), (2) reduced-reference (RR), and (3) no-reference (NR) algorithms. FR algorithms predict the quality of a distorted visual signal given the original reference signal. RR algorithms perform quality assessment on the distorted signal, given *incomplete* knowledge of the original reference signal. Finally, NR algorithms are required to gauge the quality of the distorted signal without *any* additional information about the reference.

Although these terms can be used when discussing 3D images, the definitions do not apply in quite the same way. This is because it is not possible to obtain access to either an original 3D signal *as it is perceived* or a distorted 3D signal *as it is perceived*! This follows since, while we can only access the left and right views of the scene (and possibly a depth/disparity map that has been independently computed or measured), we cannot access the 3D visuo-sensory experience – the *cyclopean image* – that the human re-creates in his/her brain. This is true for both 'original' and impaired cyclopean images and hence the problem is *double blind*.

Thus, the field of algorithmically assessing the 3D quality of experience and/or 3D quality is an extremely challenging one, making it a fertile ground for research. The complexity of the problem, coupled with our yet nascent understanding of 3D perception and of the increasing commercial shift toward 3D entertainment makes the area of 3D QA interesting, formidable and practically relevant. In the recent past, researchers have attempted to develop algorithms that are capable of predicting not only 3D quality but also 3D quality of experience. In order to develop successful 3D IQA algorithms, it is imperative to understand the human perception of 3D quality [15]. Here, we describe our recent efforts in creating a large-scale publicly available dataset of 3D reference and distorted images along with human/subjective opinion scores of the quality of these images.

The new LIVE 3D IQA database consists of left–right stereo image pairs accompanied by co-registered *precision* depth maps measured by a LIDAR-based range scanner, yielding valuable *ground truth* depth information. Since true depth is available, we envision that these images and range scans will be uniquely useful for 3D quality assessment studies, as well as for the development and benchmarking of 3D stereo vision estimation algorithms; supplementing the limited and dated Middlebury stereo database [16],[2] and for a variety of 3D vision science inquiries, such as studies of the statistics of stereoscopic images and distances in the real world [17,18]. Previous approaches to 3D QA have involved simple extensions of 2D QA along with some additional quality information gleaned from *computed* depth maps. As these depth/disparity maps are computed using an algorithm, their contribution to 3D QA is suspect, since algorithmic computation of disparity is still an open area of research. In order to ensure that 3D QA algorithms are not crippled by the approach adopted for disparity computation, this dataset provides the necessary tools for algorithm development, by not only providing high precision human opinion scores, but also true depth information from a range scanner.

Through the rest of this paper we summarize other such 3D quality assessment databases which have been used in the recent past to gauge the performance of 3D QA algorithms. We then describe in detail the LIVE 3D IQA database including capture, distortion simulation, subjective study and performance evaluation of 2D and 3D quality assessment algorithms. In the final segment, we attempt to foretell the future of visual quality assessment of 3D signals. We describe our own efforts at creating objective/algorithmic 3D quality assessment algorithms and explain a sample framework for FR 3D QA using a perceptual model. We describe research efforts that we believe are important in understanding 3D quality and hypothesize about possible future work in this area.

1.2. A primer on stereo creation and perception

Before we begin, however, it may be prudent to go through a quick primer on stereoscopic content creation and perception. Two calibrated cameras separated by a fixed distance are mounted on a rig and the pair of signals so acquired are referred to as a stereoscopic pair. As illustrated

[1] Not to mention the 4–10% of people that exhibit some degree of stereo deficiency, and hence do not fully appreciate stereo presentations [12–14].

[2] Too small to be statistically significant, and acquired using a much less precise range acquisition technology.

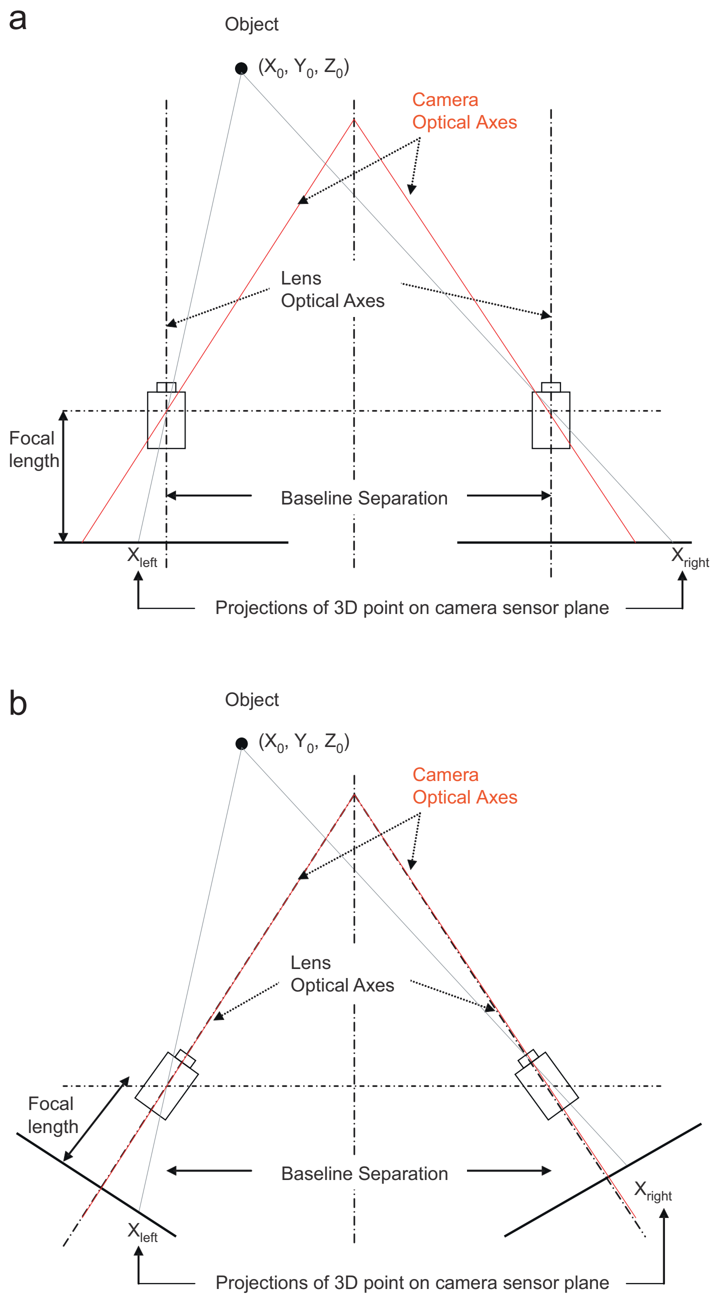

Fig. 1. Camera geometry to capture stereoscopic content: (a) parallel base-line configuration and (b) toe-in configuration.

in Fig. 1, the camera arrangement could be parallel baseline, i.e., the (lens) axes of the two cameras are parallel to each other, or a toe-in configuration in which the (lens) axes intersect.[3] In the study to be described, we opted to utilize a parallel optical axis camera geometry as the simples and most practical nominal assumption. Going forward, a 3D quality database that includes stereo pairs acquired under vergent conditions would be of great interest (for a very broad variety of vision studies, in fact), but this would require a deep stereographic study to select fixation and would probably preclude ground-truth acquisition. Such a database would be invaluable for understanding the effects of geometry on the physiology and psychology of stereoscopic viewing—important considerations towards optimizing viewer comfort.

Let us now turn to the display of 3D images. In order to create the perception of a 3D experience, the left–right pair is displayed such that the left image is seen only by the left eye, and the right image is seen only by the right. This is generally accomplished by using polarizing filters for each of the two projections such that each polarizer is orthogonal to the other.[4] This, coupled with matched

[3] Note that the choice of the baseline separation could be a function of the scene being imaged and the associated comfort when projected onto a 3D display. While an analysis of this is beyond the scope of this article, the interested reader is directed to [19,20].

[4] Recall that this was once done almost exclusively using two different color (red–green) images overlaid (called an anaglyph) and red–green glasses so that one image fell on each eye.

Table 1
Databases used by various researchers and their properties.

Database	Distortions	# of ref.	# of dist.	# of subjects	Public
Toyoma [24]	Symmetric/Asymmetric JPEG compression	10	490	24	No
Ningbo [25]	JPEG, JPEG2000, Gaussian Blur and white noise (right only, left pristine)	10	400	20	No
IRCCyN/IVC [26]	JPEG/JPEG2000 compressed images, Blur	6	90	17	Yes

polarized glasses creates a 3D perception. Since the polarizers are not perfect, some amount of the left image 'leaks' into the right and vice versa, leading to 'cross-talk'. Recently, active stereoscopic rendering, which greatly reduces crosstalk, has gained commercial acceptance. With this technology, the left and right images are flashed on the screen one after the other, and the appropriate eye is 'shut-off' by the glasses. Another, still evolving category called autostereoscopic displays, do not require any glasses at all, and consist of lenticular lenses or parallax barriers which redirect the image to different viewing regions [21]. The types of polarization (linear vs. circular) and the advantages and disadvantages of active vs. passive displays and the problems with autostereoscopic displays are discussed in detail in [21,22]. In our study, we used a polarized display and the stereoscopic images were captured using a parallel baseline setup. Our desire to capture ground truth data associated with the acquired stereo-pairs limited this work to studying the subjective quality of static stereo images only (an unsolved problem, in any case). While capturing ground truth temporal stereo data is feasible, it is not possible to capture it at high spatial resolutions as we have done here.

1.3. Previous work

There exist some databases that have previously been used to evaluate 3D quality and below we summarize them. We note that none of these databases include "true-depth" information from range-scanners.

The authors of [23] conducted two experiments to gauge visual quality on mobile stereoscopic devices. In the first, a single stimulus study,[5] the participants rated the quality of experience on a discrete unlabeled scale from 0 to 10 as well as the quality for viewing mobile 3D TV on a binary (yes/no) scale. In the first experiment, each evaluation was conducted in two different contexts, while in the second there were three different contexts.[6] The signals were encoded using a variety of video bit-rate/frame-rate/audio bit-rate combinations using the H.264 codec. Their results indicate that, even at high bit-rates, preference for 3D signals is below the level of 2D signals. The authors also analyzed verbal descriptions obtained from those that participated; various factors, such as ghosting, the need to focus, unpleasantness, and unease in viewing, were used to describe the 3D experience.

Other researchers have used databases to evaluate their algorithm performance and some of these databases have been made available for public use. In Table 1, we list these databases, the distortions considered and the number of images in each as well as their availability. In cases where the availability was unclear, we contacted the authors, where no replies were forthcoming, we categorized the database as unavailable for public use. A database has limited value unless it is made publicly available so that other researchers can make comparisons.

The LIVE 3D IQA database incorporates symmetric distortions and spans a wider gamut of distortions as compared to those listed in Table 1. Further, apart from DMOS, the database also provides researchers access with 'true' depth information obtained from a range scanner, which all of the above databases lack. Finally, the LIVE 3D IQA database is available freely for research purposes, so that objective comparison of algorithms can be undertaken.

2. LIVE 3D image quality assessment database

2.1. Database creation

Conducting a human study on the quality of displayed visual signals is a complex, multi-faceted task—especially when the signals represent 3D information. Recently, we conducted such a study as a service to the 3D QA community of researchers. It is always desirable to have available a diverse set of databases across which algorithmic performance may be analyzed. Here we describe how we went about creating this first phase of the LIVE 3D IQA database (future studies are planned).

2.1.1. Data acquisition

The image and range data used in this study were collected using an advanced terrestrial range scanner, the RIEGL VZ-400, with a co-registered 12.1 megapixel Nikon D700 digital camera mounted on top of it [27] (see Fig. 2). The RIEGL VZ-400 allows for a maximum scan angle range of $100^\circ(+60^\circ/-40^\circ)$, with a minimum angle step-width of 0.0024°. Scan speeds up to 120 lines/s can be achieved, with an angle measurement resolution of better than 0.0005° and a maximum measurement range of up to 500 m. The "ground truth" precision range data that we acquire in this way is a unique feature of the LIVE 3D IQA database.

The range scanner and the camera assembly was mounted on a specially designed stereoscopic plate that we constructed, which allowed for lateral displacement of the assembly. The stereoscopic plate is equipped with

[5] Meaning one image is shown to the subject at a time, as opposed to when images are shown in relation to the reference, such as a side-by-side pairwise; these are referred to as "double stimulus" comparisons.

[6] Context of use comprises of user characteristics, tasks, as well as technical, physical and social environments. Here they consisted of different environments such as a laboratory, home viewing, on the bus, and on a station.

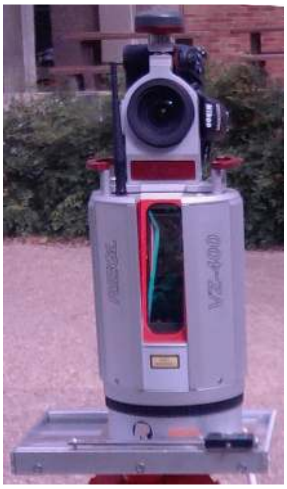

Fig. 2. The RIEGL VZ-400 terrestrial range scanner and co-registered Nikon D7 00 DSLR camera used to collect stereoscopic signals.

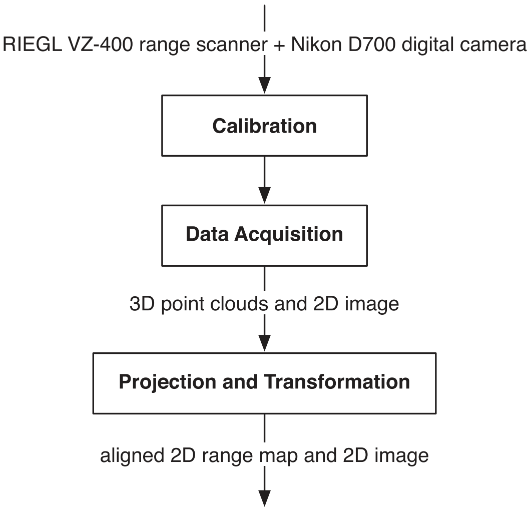
RIEGL VZ-400 range scanner + Nikon D700 digital camera
Calibration
Data Acquisition
3D point clouds and 2D image
Projection and Transformation
aligned 2D range map and 2D image

Fig. 3. A diagram representing the data acquisition flow.

a digital vernier scale in order to accurately measure lateral displacement. To acquire each stereoscopic image pair, an image-range map pair was first acquired at vernier reading 0 mm; the assembly was then displaced by ≈65 mm (adult inter-ocular distance) and another image-range map was then acquired. We note that while parallel baseline configurations for 3D acquisitions are acceptable, researchers are still trying to understand how to optimally capture stereoscopic content given the scene content, in order to minimize fatigue, discomfort, headache and other negative factors induced by improper geometry or stereography [19,20,28].

The two images form the stereoscopic pair, and the two range maps yield precision depth information of the scene being imaged. Having two range maps is quite useful, since occlusions and measurement errors in the range data may be corrected with the additional information. Manual calibration was performed prior to acquisition using the RIEGL RiScan Pro software [29], and the 3D point cloud and the 2D images were processed to obtain a stereoscopic pair (left–right) of high quality JPEG images at a resolution of 640 × 360, along with two range maps of resolution 640 × 360 for each scene. The procedure is described below (Fig. 3).

The acquired range data was exported from the range scanner as a point cloud with the three-dimensional coordinate and the range value, while the image data was stored in the digital camera as (high quality) JPEG files. Finally, to obtain the aligned 2D range map with the 2D image, the 3D point clouds were projected and transformed into the 2D range map by applying the pinhole camera model with lens distortion [30,31].

First, the three-dimensional coordinates of the point clouds were converted into the undistorted two-dimensional pixel coordinates

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{A} \cdot \mathbf{RT} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

$$\mathbf{A} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$\mathbf{RT} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x/z \\ y/z \end{bmatrix} \tag{4}$$

where $[X\ Y\ Z]^T$ is the three-dimensional coordinate of the point cloud, **A** is the camera's intrinsic matrix, **RT** is the joint rotation-translation matrix, and $[u\ v]^T$ is the undistorted two-dimensional pixel coordinate.

In the intrinsic matrix **A**, $[c_x\ c_y]^T$ is the coordinate of the principal point, which is usually at the image center, and (f_x, f_y) are the focal lengths along the x- and y-axes, all expressed in the unit of pixels.

The parameters in the joint rotation-translation matrix **RT** were computed from the manual calibration after mounting the digital camera onto the range scanner.

Since real lens usually have distortions, viz. radial and tangential, the distorted two-dimensional pixel coordinates were computed by transforming the undistorted two-dimensional pixel coordinates as follows:

$$\begin{aligned} u_d = u + u' f_x (k_1 r^2 + k_2 r^4 + k_3 r^6 + k_4 r^8) \\ + 2 f_x u' v' p_1 + p_2 f_x (r^2 + 2u'^2) \end{aligned} \tag{5}$$

$$\begin{aligned} v_d = v + v' f_y (k_1 r^2 + k_2 r^4 + k_3 r^6 + k_4 r^8) \\ + 2 f_y u' v' p_2 + p_1 f_y (r^2 + 2v'^2) \end{aligned} \tag{6}$$

$$u' = (u - c_x)/f_x \tag{7}$$

$$v' = (v - c_y)/f_y \tag{8}$$

$$r = u'^2 + v'^2 \tag{9}$$

where $[u_d\ v_d]^T$ is the distorted two-dimensional pixel coordinate, (k_1, k_2, k_3, k_4) are the radial distortion coefficients, and (p_1, p_2) are the tangential distortion coefficients.

After the distorted two-dimensional pixel coordinates of each point cloud were computed, the aligned 2D range map was obtained by filling the range value at each pixel location with the one at the closest distorted two-dimensional pixel coordinate. All scenes were imaged at a resolution of 2823 × 4256, and the range scanner set with an angular precision of 0.04, since a higher scanning resolution would lead to an increase in scanning time as well as an increase in the probability of inconsistency between the 3D point clouds and the corresponding 2D image in natural scenes. In addition, since the digital camera is mounted in portrait mode onto the range scanner, the field of view for the 3D point clouds needs to be adjusted to match the aspect ratio of the portrait image, resulting in 60° and 100° fields of views in the horizontal and vertical direction respectively. As a result, the resolution of the 3D point clouds from the range scanner is $\frac{60}{0.04} \times \frac{100}{0.04} = 1500 \times 2500$ (points), which is smaller than the image resolution captured by the digital camera. The range in depth that the scanner can measure depends on the operation mode, the sunlight, the weather, the targets' reflectivity (material), etc. During our data acquisition, we used the long-range mode, where the min. range is 1.5 m and the max. range is 280–600 m, depending on the reflectivity. The data type "double" was used to represent the ranges in MATLAB.

To provide accurately aligned 2D range maps and images while keeping their resolution as high as possible, the 3D point clouds were projected and transformed into a 2D range map with a resolution of 708 × 1064, while the original 2D image was also down-sampled to the same size. Inaccurate range values at boundary pixels in the natural scene were removed by cropping the aligned 2D range map and 2D image to a resolution of 640 × 360, which is appropriate for display and viewing using our setup. Although higher resolution display was possible, we decided that an intermediate display size would be preferable given the proliferation of large-format displays and the expected large-scale deployment of small format display devices. Finally, slight differences in contrast between the two views were resolved using a simple histogram matching approach. Note that since the image pairs were not captured at the same time, small variations (due to leaves, dust, birds, etc.) may have occurred between the two views. While the binocular compensation reduces many of these variations, one cannot guarantee that the two views demonstrate no variation. We have further tried to reduce these variations by collecting a large sample of images and pruning out those images which demonstrated large variations. Care was also taken during the capture process to image scenes at times when such variations would be minimized (for example, on a non-windy day).

Thus, for each scene imaged, a stereoscopic pair (left–right) of high quality JPEG images at a resolution of 640 × 360, and two 2D range maps of resolution 640 × 360 were obtained.

All of the stereoscopic data were collected from outdoor scenes. Fig. 4 shows some examples of the natural scenes that were obtained with the aligned 2D range map and 2D images. The natural scenes where the image and range data were collected include different parts of the

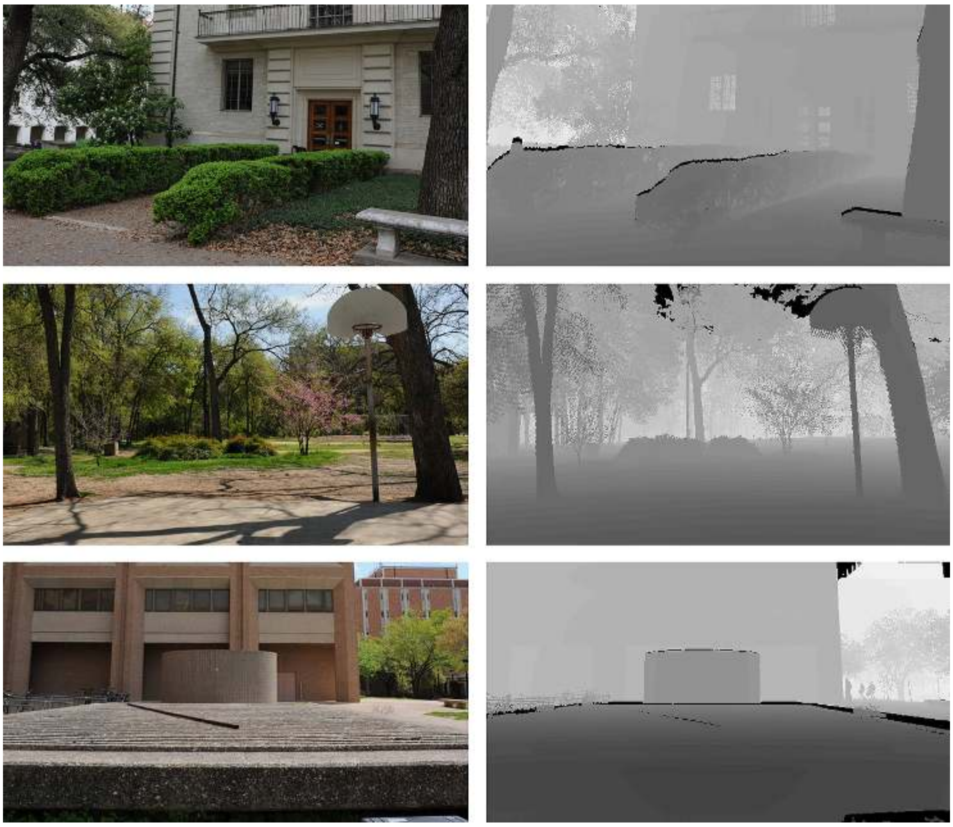

Fig. 4. Examples of the natural scenes, 2D images on the left and aligned 2D range maps on the right. Black regions indicate locations were range was not obtainable.

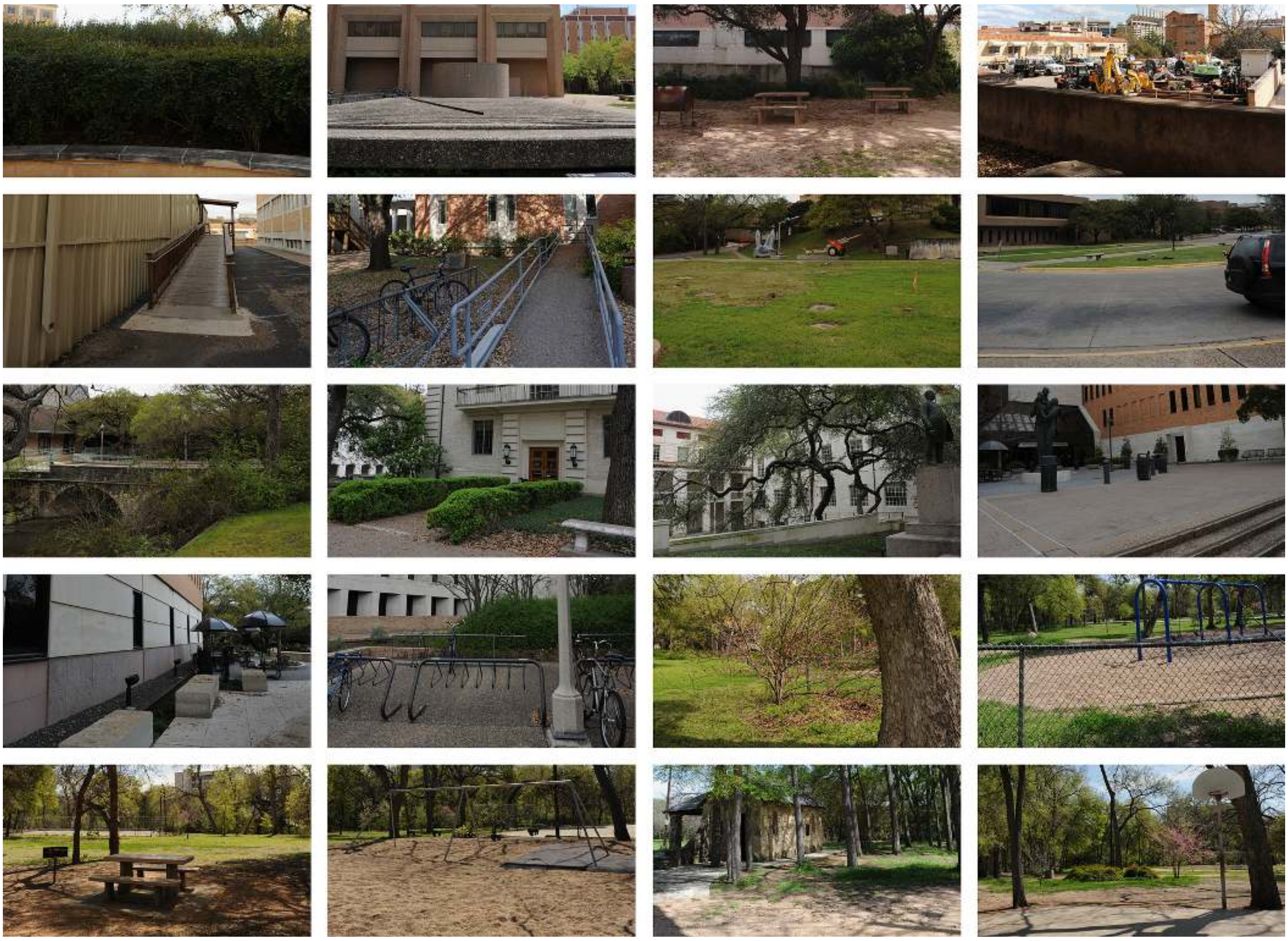

Fig. 5. The 20 reference images used in the subjective study. Shown here are only the left-views.

campus at The University of Texas at Austin, and the Eastwoods Park nearby. In Fig. 5, we plot each of the 20 reference images (left view) that were acquired and used in the subsequent study.

2.1.2. *Distortion simulation*

The distortions that we selected to use in this study mirror those in the popular LIVE IQA database [32]. The distortions that were simulated include 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. Degradation in visual quality for each of these distortions was achieved by varying a control parameter within a particular range; all of which are tabulated in Table 2. As an illustration, in Fig. 6, we show a stereoscopic pair of images from the LIVE 3D IQA database that has been distorted by fast-fading distortion; the reader is encouraged to free-fuse this pair in order to visualize how 2D distortions can affect 3D percepts.[7]

Table 2
Range of parameter values for distortion simulation.

Distortion	Control parameter	Range
JP2K	Bit-rate	[0.05 3.15]
JPEG	Quality parameter	[10 50]
WN	Variance of Gaussian	[0.01 1]
Blur	Variance of Gaussian	[0.01 15]
FF	Channel SNR	[12 20]

JPEG compression was simulated using MATLAB's JPEG compression utility, while JPEG2000 (JP2K) compression was simulated using the Kakadu encoder—the parameters varied were the 'quality' parameter and the bit-rate, respectively. Additive white Gaussian noise (WN) was simulated using the `imnoise` command in MATLAB, where Gaussian noise was applied equally across the R, G and B planes. Similarly, Gaussian blur was simulated by applying a Gaussian low-pass filter to each of the color planes. For both WN and Blur, the control parameter was the variance of the Gaussian. Fast-fading (FF) distortion consisted of a JP2K compressed image transmitted over a Rayleigh fading channel, with the channel Signal-to-Noise ratio (SNR) as the control parameter.

Since we are dealing with stereoscopic signals, distortions may be applied asymmetrically or symmetrically.

[7] For example, one of the subjects in our study (described in the text) questioned the wisdom of adding noise 'in-the-front' while the image was perfect 'at-the-back'. This would imply that noise does not destroy the depth-percept but still leads to some annoyance.

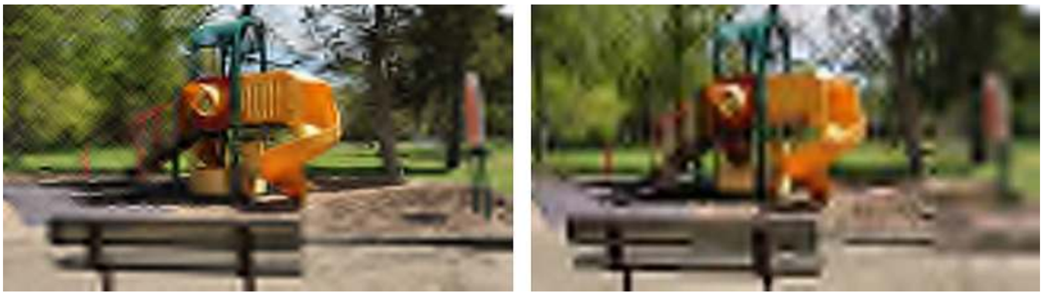

Fig. 6. Stereoscopic distorted pair from LIVE 3D IQA database. Free-fuse the left and right images to obtain a 3D percept.

The recent past has seen some research activity on asymmetric compression of stereoscopic signals [33,34], and asymmetric distortions and their effect on visual quality remain an interesting avenue for research. However, in this study, all distortions are symmetric. Specifically, the left and right images from each stereoscopic pair were distorted using the five different distortions above, where the 'amount' of each distortion remains the same for the left and right image.

A total of 20 reference images and 365 distorted images (80 each for JP2K, JPEG, WN and FF; 45 for Blur) were thus created and utilized for the subjective study.

2.2. Subjective study

A single-stimulus continuous quality evaluation (SSCQE) with hidden reference study [2] was conducted at the University of Texas at Austin (UT), over the course of two weeks. The subject pool consisted of 32 (mostly under-graduate) students from UT. The subjects were a mix of males and females, with a male-majority and were informally tested for stereo acuity. While a visual acuity test was not performed, a verbal confirmation of the same was obtained prior to the study. The study involved two sessions of viewing, each lasting less than 30 min, in order to minimize subject-fatigue [35]; the average testing time was approximately 22 min. An informal after-study feedback conducted indicated that the subjects were able to perceive stereoscopic signals well and that they did not experience any uneasiness or fatigue during the course of the study. Each image was displayed on the screen for 8 s. Each session began with a short training module in which the subject saw six stereoscopic signals chosen to span the range of distortions that the subject was about to view. The signals used for training differed from those in the actual study. The study consisted of the set of images shown in random order. The order was randomized for each subject as well as for each session. Care was taken to ensure that two consecutive sequences did not belong to the same reference, to minimize memory effects [35]. Images were displayed on a 22 in. IZ3D passive stereoscopic display with the screen resolution set at 800 × 600.

The study design was such that each image received ratings from 17 subjects, and the ratings that the subject gave the distorted signal were subtracted from the rating that the subject gave the corresponding reference signal to form a differential opinion score (DOS). A subject rejection procedure was then run as per recommendations [35] which rejected two subjects. The remaining subjective scores were then averaged across subjects to produce differential mean opinion scores (DMOS).

At this juncture it may be prudent to discuss our study design a bit further. Specifically, as independent vision scientists working in academia, we find it to be scientifically judicious to depart from "standardized" procedures at times such as those set forth by the ITU [35]. ITU recommendations demand rigid screening and experimental setups that are no longer relevant in this era (e.g., they were designed to formalize studies of quality assessment on CRT TVs, an environment where screen sizes, expected viewing distances and overall environment varied much less than today's variegated video experiences). It is our opinion that standards should be used to the extent for which they have been designed. The topic of subject screening is a good example. While psychometric studies ordinarily require great rigor in subject screening (of acuity in 2D and 3D, of color sense, and in this case, of stereo capability), unlike our non-QA work, we are relaxing our subject screening, since image display devices are being deployed in high diverse and dynamic environments, and we think that subjects should model the general populace as much as possible. In this study, we only tested stereo blindness since we wished to specifically explore the interplay between distortions and perceived depths.

Another important consideration is the number of participants in the study. There exist recommendations on this as well, and some researchers have studied the question of the maximum number of subjects to conduct meaningful studies. While QA studies are typically large (>20 subjects), we believe that a more important metric than subject count is the statistical confidence in the study, even if fewer subjects are used. The scores from the LIVE 3D IQA database, as we shall see, satisfy this requirement.

The LIVE 3D IQA database consists of 20 reference images, 5 distortion categories and a total of 365 distorted images along with the associated DMOS. A histogram of DMOS scores and a histogram of the standard errors is shown in Fig. 7. We note that these standard deviations are in line with previous studies of this nature for 2D images and videos [2,36]. Further, the DMOS distribution is uniform through a large portion of the scale indicating that the distortions in the LIVE 3D IQA database span a wide range of visual quality.

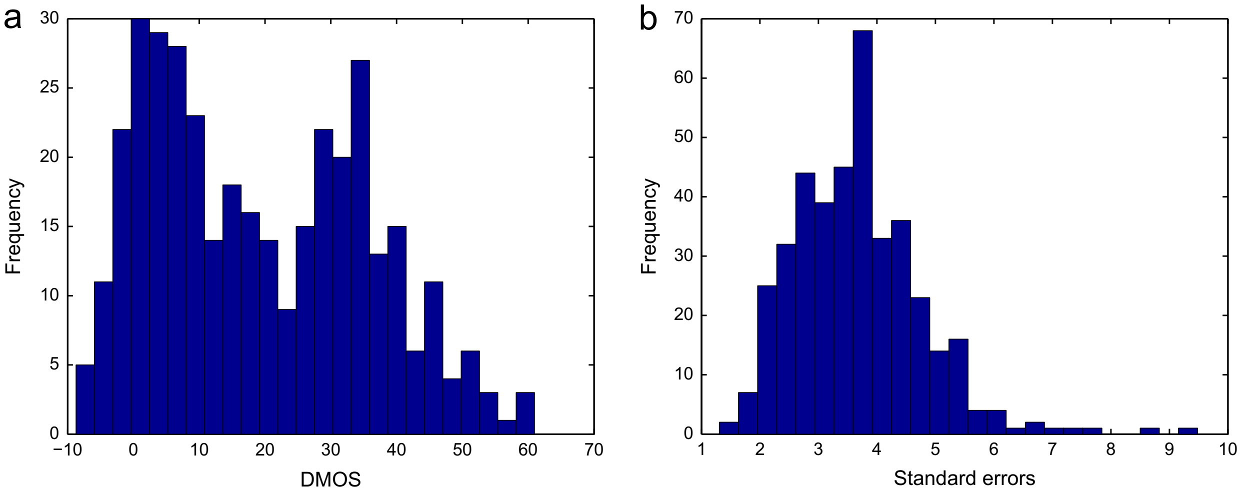

Fig. 7. Subjective study data from the LIVE 3D IQA database: (a) Histogram of DMOS scores and (b) Histogram of standard errors.

2.3. Algorithm performance evaluation

We evaluated the performance of a number of FR 3D IQA algorithms on the LIVE 3D IQA dataset. These algorithms were chosen based on their reported performance and availability. Further, in order to provide a baseline performance and to evaluate the efficacy of 2D algorithms for predicting 3D visual quality, we also evaluated a set of popular 2D FR IQA algorithms on the dataset. When using a 2D IQA algorithm, the algorithm was applied on the left and right images separately and the estimated quality scores averaged to produce single measure of 3D quality. In the case of 3D FR IQA algorithms, the algorithms were applied as described in the cited references. The 2D IQA algorithms used here are part of the MetrixMux toolbox, available at [37]. For 3D QA algorithms, the respective authors were contacted for code. We used the code provided by the authors in [26] and coded all the other algorithms ourselves in Matlab. The code for each of these algorithms is available as part of the LIVE 3D IQA database, which will be publicly accessible at the LIVE QA web site by the time this article appears.

Tables 3 and 4 list the 2D and 3D IQA algorithms evaluated in this study. To conserve space, we do not describe the 2D algorithms here, so the reader is referred to the cited literature.

The performance measures used are Spearman's rank ordered correlation coefficient (SROCC), the linear (Pearson's) rank ordered correlation coefficient (LCC) and the root-mean-squared error (RMSE) [32,3]. LCC and RMSE were computed after logistic regression through a non-linearity, as described in [32]. An SROCC and LCC value close to 1 indicates good correlation with human perception, while lower values of RMSE indicate better performance. Tables 5–7 list the performance of the various 2D IQA algorithms. The results of 3D IQA algorithms are listed in Tables 8–10.

We also performed a statistical significance analysis using the *t*-test between the residuals in prediction obtained from the non-linear regression process that was used to compute the linear correlation coefficient on the entire dataset [32,52]. The results are listed in Table 12. The algorithms in [48,49,51,50] are statistically worse than 2D PSNR, while that in [46] is statistically equivalent to 2D PSNR. All the other algorithms are statistically superior to 2D PSNR. UQI [41], the best performing algorithm on the dataset, is statistically superior to all algorithms, except MS-SSIM and WSNR, which are statistically equivalent to UQI.

The results in Table 5 differ from those for the same algorithms when used for 2D quality assessment [36]. For example, SSIM(MS) and VIF are top performers on the LIVE IQA database, while the performance of UQI is far worse. However, for 3D QA, UQI seems to outperform VIF and SSIM(MS), although the latter have good performance.

Table 3
List of FR 2D IQA algorithms evaluated in this study.

No.	Algorithm
1.	Peak Signal-to-Noise ratio (PSNR)
2.	Structural Similarity Index (SSIM) [38]
3.	Multi-scale Structural Similarity Index (SSIM (MS)) [39]
4.	Visual Signal-to-Noise ratio (VSNR) [40]
5.	Visual Information Fidelity (VIF) [36]
6.	Universal Quality Index (UQI) [41]
7.	Noise Quality Measure (NQM) [42]
8.	Weighted Signal-to-Noise ratio (WSNR) [43]
9.	C4 [44]
10.	Blind Image Quality Index (BIQI) [45]

Table 4
List of 3D IQA algorithms evaluated in this study. *Italics* indicates an NR (blind) algorithm.

No.	Algorithm
1	Benoit [26]
2	Hewage [46]
3	You [47]
4	Gorley [33]
5	Shen [48]
6	Yang [49]
7	Zhu [50]
8	*Akhter* [51]

Table 5
Performance of 2D IQA algorithms in predicting perceived 3D image quality: Spearman's Rank Ordered Correlation Coefficient (SROCC). *Italics* indicates an NR (blind) algorithm.

Algorithm	JP2K	JPEG	WN	Blur	FF	All
PSNR	0.7967	0.1311	0.9318	0.9016	0.5957	0.8370
SSIM	0.8572	0.4346	0.9395	0.8822	0.5849	0.8772
SSIM (MS)	0.8975	0.6019	**0.9439**	0.9262	0.7316	0.9237
VSNR	0.8313	0.4062	0.9049	0.8306	0.7283	0.8817
VIF	0.9018	0.5828	0.9325	0.9312	0.8037	0.9204
UQI	0.9101	**0.7371**	0.9272	0.9238	0.8322	**0.9381**
NQM	0.8619	0.5399	0.9237	0.9058	0.7509	0.9103
WSNR	0.8997	0.6132	0.9369	0.9291	0.7604	0.9255
C4	**0.9108**	0.6365	0.9425	0.9361	**0.8349**	0.9144
BIQI	*0.7727*	*0.4887*	*0.9277*	*0.8596*	*0.7067*	*0.8652*

Table 6
Performance of 2D IQA algorithms in predicting perceived 3D image quality: Linear Correlation Coefficient (LCC).

Algorithm	JP2K	JPEG	WN	Blur	FF	All
PSNR	0.7889	0.2311	0.9347	0.8937	0.7062	0.8251
SSIM	0.8650	0.4849	0.9374	0.9197	0.7212	0.8727
SSIM (MS)	0.9306	0.6712	**0.9474**	0.9461	0.8060	0.9302
VSNR	0.8898	0.4107	0.9111	0.8726	0.7867	0.8665
VIF	0.9361	0.6738	0.9273	0.9570	0.8542	0.9183
UQI	**0.9512**	**0.7727**	0.9273	0.9565	**0.8788**	**0.9424**
NQM	0.9159	0.5666	0.9252	0.9399	0.7878	0.9151
WSNR	0.9326	0.6763	0.9369	0.9443	0.8113	0.9260
C4	0.9378	0.6497	0.9359	**0.9649**	0.8754	0.9193
BIQI	*0.8203*	*0.6136*	*0.9323*	*0.8995*	*0.7762*	*0.8792*

Table 7
Performance of 2D IQA algorithms in predicting perceived 3D image quality: Root Mean-Squared-Error (RMSE).

Algorithm	JP2K	JPEG	WN	Blur	FF	All
PSNR	7.9587	6.3624	5.9145	6.5271	8.7971	9.2678
SSIM	6.4984	5.7191	5.7947	5.6814	8.6069	8.0059
SSIM (MS)	4.7417	4.8473	**5.3236**	4.6887	7.3553	6.0187
VSNR	5.9097	5.9623	6.8588	7.0782	7.6714	8.1868
VIF	4.5570	4.8319	6.2291	4.1986	6.4615	6.4903
UQI	**3.9983**	**4.1508**	6.2261	4.2222	**5.9297**	**5.4865**
NQM	5.1974	5.3895	6.3124	4.9439	7.6530	6.6134
WSNR	4.6740	4.8167	5.8148	4.7651	7.2644	6.1902
C4	4.4951	4.9708	5.8615	**3.8003**	6.0067	6.4530
BIQI	*7.4100*	*5.1636*	*6.0166*	*6.3238*	*7.8679*	*7.8119*

Based on the results in Tables 5 and 8, it is clear that for the set of distortions considered, the 2D IQA algorithms perform well in terms of correlation with human subjectivity, while the addition of disparity/depth in the 3D algorithms does not materially improve the performance (in agreement with a previous study [26]). Yet, our own experiences with distorted 3D images leads us to believe that disparity activity (e.g., caused by rapid changes in depth) may affect distortion visibility, viz., that the experience of depths and depth variations may render some distortions more or less visible [53]. Towards this end, we designed a "laboratory-only" algorithm that incorporates disparity activity computed from the ground truth depth data in the LIVE 3D IQA database, whose results we tabulate in Table 11. Of course, we could have employed a stereo matching algorithm, but the necessary characteristics and accuracy of such algorithms for this problem remain open questions.

In the "laboratory-only" algorithm, the stereoscopic pair is decomposed by a set of multi-scale oriented complex Gabor filters (both the reference and distorted image). The squared differences of left and right response amplitudes are divisively normalized (similar to [54]). The result is then also divisively normalized by ground-truth disparity. In practice, of course, disparity could be estimated from the stereo images. Table 8 shows these results (labeled as Gabor energy – masking). The performance of the model without any masking is also shown (labeled as Gabor energy – no masking).

The result suggests that incorporating a luminance masking model and a disparity masking model into the Gabor energy algorithm improves the performance. Our observations indicate that there is a definite interaction between disparity and luminance and that disparity activity may reduce luminance masking. This agrees with a recent study on visual masking which suggests that the presence of a target and a mask at different depths activates two separate pools of neurons, thereby reducing the masking ability of the target [55]. Our observations are bolstered by the performance of the 2D UQI. UQI is a preliminary version of the popular SSIM [41,38], and while it does account for contrast masking using a divisive normalization-based approach, SSIM incorporates a much better model for contrast masking—thereby making SSIM a better measure for 2D quality assessment. However, such strong contrast masking may not reflect human perception accurately when viewing 3D scenes [55].

Table 8
Performance of 3D IQA algorithms in predicting perceived 3D image quality: Spearman's Rank Ordered Correlation Coefficient (SROCC). *Italics* indicates an NR algorithm.

Algorithm	JP2K	JPEG	WN	Blur	FF	All
Benoit [26]	**0.9103**	0.6028	0.9292	**0.9308**	**0.6989**	**0.8992**
Hewage [46]	0.8558	0.5001	0.8963	0.6900	0.5447	0.8140
You [47]	0.8598	0.4388	**0.9395**	0.8822	0.5883	0.8789
Gorley [33]	0.4203	0.0152	0.7408	0.7498	0.3663	0.1419
Shen [48]	0.2133	0.2440	0.8917	0.6586	0.2665	0.0679
Yang [49]	0.1501	0.1328	0.8471	0.3266	0.1426	0.0785
Zhu [50]	0.7708	0.2929	0.4651	0.7935	0.4752	0.6388
Akhter [51]	*0.8657*	0.6754	*0.9137*	*0.5549*	*0.6393*	*0.3827*

Table 9
Performance of 3D IQA algorithms in predicting perceived 3D image quality: Linear Correlation Coefficient (LCC). *Italics* indicates an NR algorithm.

Algorithm	JP2K	JPEG	WN	Blur	FF	All
Benoit [26]	**0.9398**	0.6405	0.9253	**0.9488**	**0.7472**	**0.9025**
Hewage [46]	0.9043	0.5305	0.8955	0.7984	0.6698	0.8303
You [47]	0.8778	0.4874	**0.9412**	0.9198	0.7300	0.8814
Gorley [33]	0.4853	0.3124	0.7961	0.8527	0.3648	0.4511
Shen [48]	0.5039	0.3899	0.8988	0.6846	0.4830	0.5743
Yang [49]	0.2012	0.2738	0.8701	0.6261	0.2824	0.3909
Zhu [50]	0.8073	0.3790	0.5178	0.7770	0.5038	0.6263
Akhter [51]	*0.9059*	0.7294	*0.9047*	*0.6177*	*0.6603*	*0.4270*

Table 10
Performance of 3D IQA algorithms in predicting perceived 3D image quality: Root-mean-squared-error (RMSE). *Italics* indicates an NR algorithm.

Algorithm	JP2K	JPEG	WN	Blur	FF	All
Benoit [26]	**4.4266**	5.0220	6.3076	**4.5714**	**8.2578**	**7.0617**
Hewage [46]	5.5300	5.5431	7.4056	8.7480	9.2263	9.1393
You [47]	6.2066	5.7097	**5.6216**	5.6798	8.4923	7.7463
Gorley [33]	11.3237	6.2119	10.1979	7.5622	11.5691	14.6350
Shen [48]	12.2754	6.0216	7.2939	10.5547	10.8820	13.5473
Yang [49]	12.6979	6.2894	8.2002	12.1291	11.9462	15.2481
Zhu [50]	7.6813	6.0684	14.7201	9.1270	10.7362	12.7828
Akhter [51]	*5.4836*	4.4736	*7.0929*	*11.3872*	*9.3321*	*14.8274*

Table 11
Performance of a "laboratory-only" algorithm in predicting perceived 3D image quality: Spearman's Rank Ordered Correlation Coefficient (SROCC).

Algorithm	JP2K	JPEG	WN	Blur	FF	All
Gabor energy—no disparity activity	0.8877	0.5102	0.9326	0.9352	0.6846	0.9163
Gabor energy—disparity masking	0.9013	0.6620	0.9445	0.9389	0.7389	0.9336

These and other such insights [56] are significant for algorithm design and suggest that the simplistic approach that most of the 3D algorithms take may not be sufficient to predict perceptual quality.

3. Discussion and future work

Our analysis of algorithms in the area of 3D QA might lead one to believe that 2D algorithms are sufficient to gauge the perceptual quality of stereoscopic signals. It may be prudent to discuss the implication of our findings for research in 3D QA. The first thing to note is that almost all of the 3D QA algorithms are simple extensions of 2D QA algorithms with some additional 'features' extracted from depth (generally disparity differences). The way in which this disparity information is incorporated into these 3D QA algorithms is not yet based on any perceptual principles (such as disparity masking, which we do not really understand yet). While these observations may partially explain why the 3D QA algorithms are not appreciably better than their 2D counterparts, it does not explain why 2D algorithms do well on the LIVE 3D IQA database.

Amongst the distortions that we consider in the dataset, WN and Blur are global distortions and hence are less likely to affect the perception of depth. Predictably, 2D algorithms do extremely well in these two categories. For those distortions that create localized artifacts, however, 2D algorithm performance is below par—especially for the local blocking/blurring types of distortion caused by JPEG compression. This suggests a possible answer to our question of 2D algorithm performance. When assessing localized distortions that may lead to depth irregularities, 2D algorithms do not do well. Their performance is not bad however, and the reason for this is that stereoscopic quality is a complex function of monoscopic quality and irregularity in depth/disparity as we have discussed before and demonstrated elsewhere [53]. Since 2D algorithms account for the monoscopic component, their performance is not abysmal. The poor performance of 3D algorithms on these distortions is likely explained by the simplistic design of these methods, and our current poor understanding of how distortions affect the 3D sensory experience, and in particular how disparity and luminance perception interact. We are designing a series of psychophysical experiments to better understand exactly how luminance and disparity may mask one another (for an example of one such study, the reader is referred to [53]).

Note that the distortions in the LIVE 3D IQA database are *not* specifically stereoscopic. Some lead to stereoscopic errors

Table 12
Results of a *t*-test for statistical significance between pairs of 2D and 3D algorithms considered here. A value of ‘1’ indicates that the row is superior to the column algorithm, while a ‘ – 1’ indicates that the row is inferior statistically; a ‘0’ indicates that the row and column algorithms are statistically equal in the performance. Also listed in brackets is the associated *p*-value. A *p*-value of 0 is to be read as $p < 0.001$.

	PSNR	SSIM	SSIM (MS)	VNSR	VIF	UQI	NQM	WSNR	C4	*BIQI*	Benoit	Hewage	You	Gorley	Shen	Yang	Zhu	*Akhter*
PSNR	0 (0.50)	1 (0.00)	1 (0.00)	1 (0.02)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	0 (0.17)	1 (0.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
SSIM	– 1 (1.00)	0 (0.50)	1 (0.00)	0 (0.89)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	0 (0.21)	1 (0.00)	1 (0.04)	– 1 (0.98)	0 (0.25)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
MS – SSIM	– 1 (1.00)	– 1 (1.00)	0 (0.50)	– 1 (1.00)	0 (0.88)	0 (0.15)	0 (0.90)	0 (0.71)	– 1 (1.00)	– 1 (0.96)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
VSNR	– 1 (0.98)	0 (0.11)	1 (0.00)	0 (0.50)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.02)	1 (0.00)	1 (0.00)	0 (0.83)	1 (0.03)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
VIF	– 1 (1.00)	– 1 (1.00)	0 (0.12)	– 1 (1.00)	0 (0.50)	1 (0.01)	0 (0.55)	0 (0.27)	– 1 (0.99)	0 (0.68)	– 1 (0.96)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
UQI	– 1 (1.00)	– 1 (1.00)	0 (0.85)	– 1 (1.00)	– 1 (0.99)	0 (0.50)	– 1 (0.99)	0 (0.95)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
NQM	– 1 (1.00)	– 1 (1.00)	0 (0.10)	– 1 (1.00)	0 (0.45)	1 (0.01)	0 (0.50)	0 (0.23)	– 1 (0.99)	0 (0.63)	0 (0.94)	– 1 (1.00)	– 1 (0.99)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
WSNR	– 1 (1.00)	– 1 (1.00)	0 (0.29)	– 1 (1.00)	0 (0.73)	0 (0.05)	0 (0.77)	0 (0.50)	– 1 (1.00)	0 (0.87)	– 1 (0.99)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
BIQI	– 1 (1.00)	0 (0.79)	1 (0.00)	– 1 (0.98)	1 (0.01)	1 (0.00)	1 (0.01)	1 (0.00)	0 (0.50)	1 (0.02)	0 (0.20)	– 1 (1.00)	0 (0.56)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
C4	– 1 (1.00)	– 1 (1.00)	1 (0.04)	– 1 (1.00)	0 (0.32)	1 (0.00)	0 (0.37)	0 (0.13)	– 1 (0.98)	0 (0.50)	0 (0.91)	– 1 (1.00)	– 1 (0.99)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
Benoit	– 1 (1.00)	– 1 (0.96)	1 (0.00)	– 1 (1.00)	1 (0.04)	1 (0.00)	0 (0.06)	1 (0.01)	0 (0.80)	0 (0.09)	0 (0.50)	– 1 (1.00)	0 (0.85)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
Hewage	0 (0.83)	1 (0.02)	1 (0.00)	0 (0.17)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	0 (0.50)	1 (0.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
You	– 1 (1.00)	0 (0.75)	1 (0.00)	– 1 (0.97)	1 (0.00)	1 (0.00)	1 (0.01)	1 (0.00)	0 (0.44)	1 (0.01)	0 (0.15)	– 1 (1.00)	0 (0.50)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)	– 1 (1.00)
Gorley	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	0 (0.50)	1 (0.01)	0 (0.41)	1 (0.00)	0 (0.88)
Shen	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	– 1 (0.99)	0 (0.50)	– 1 (0.98)	0 (0.14)	– 1 (1.00)
Yang	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	0 (0.59)	1 (0.02)	0 (0.50)	1 (0.00)	0 (0.91)
Zhu	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	– 1 (1.00)	0 (0.86)	– 1 (1.00)	0 (0.50)	– 1 (1.00)
Akhter	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)	0 (0.12)	1 (0.00)	0 (0.09)	1 (0.00)	0 (0.50)

(e.g., JPEG), however, this is not a controlled phenomenon. At this juncture it is important to realize that a deep understanding of what one means by stereoscopic distortions, and how they are perceived is needed. Once such understanding is obtained, one needs a simulation process whereby turning a ‘knob’ will change the strength of stereoscopic distortions in a controlled manner for example: incorporating variable compression artifacts at known depths.

Once we are able to deepen our understanding of the perception of stereoscopic distortions, better algorithms will follow. How does the disparity affect luminance masking? Do large disparity variations modify the strength of luminance masking? Such an understanding will lead to a mathematical formulation of 3D masking.

Another tremendous avenue of inquiry is studies of 3D natural scene statistic models of disparity and disparity-luminance [17,18]. Understanding the statistics of disparity in natural scenes and how they might be used in 3D QA algorithms is a very fruitful avenue of research [57]. Such an approach has worked well for 2D NR IQA [45], and we believe that 3D NR QA will benefit from such an approach as well. For example, the 2D NR (blind) algorithm – BIQI [45] – beats the 2D FR PSNR and is as good as the 2D FR SSIM (see Table 8)!

Moving away from static scenes to 3D videos, important avenues of research are space–time 4D QA of stereo video, and the development of models of spatio-temporal-disparity masking. Perceptually motivated approaches such as those in [54] for FR 2D VQA will serve as useful inspirations in this process. For this problem, modeling natural 3D statistics as a function of motion and their relationship to perceptual quality will likely be quite useful.

The field of 3D QA remains an extremely interesting one, and there is tremendous scope for research in this area. Yet, there remain large gaps in our understanding of the perception of stereoscopic distortions and of appropriate statistical models for 3D natural scenes. A multi-pronged approach combining research and concepts from the visual sciences and image processing will hopefully lead to models that predict stereoscopic quality with high accuracy.

Before we conclude, it may be prudent to discuss the issue of stereoscopic displays. In this writeup we have avoided explicitly describing the effect of monitor choice on QA algorithm performance. While this is an important factor, we believe that as technology progresses, improved stereoscopic displays will eliminate many of the issues that current displays face. We also refer the reader to the special issue on 3D media and display in the IEEE Proceedings of April 2011 and the special issue on 3D-TV in the IEEE Transactions on Broadcasting of June 2011 which covers a range of other relevant issues including 2D-to-3D conversion and 3D TV broadcasting and distribution systems.

We have made a small journey through the interesting and multi-disciplinary field of subjective stereoscopic quality assessment. Hopefully, we have convinced the reader that the problem is exciting and needs solutions from researchers that incorporate not only techniques from quality assessment, but also those from vision science.

We have tried to cast this paper in the light of our own evolving beliefs and incomplete understanding of the 3D QA problem. We plan to continue studying the problem in "depth" and encourage the reader to join us in this venture.

Acknowledgment

This research was supported by the National Science Foundation under grant CCF-0728748 and by Intel and Cisco corporation under the VAWN program.

References

[1] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Transactions on Image Processing 15 (11) (2006) 3440–3451. http://dx.doi.org/10.1109/TIP.2006.881959.

[2] K. Seshadrinathan, R. Soundararajan, A.C. Bovik, L.K. Cormack, Study of subjective and objective quality assessment of video, IEEE Transactions on Image Processing 19 (2) (2010) 1427–1441.

[3] Video Quality Experts Group (VQEG). Final Report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment Phase I, 〈http://wwwitsbldrdocgov/vqeg/projects/frtv_phaseI〉, 2000.

[4] VQEG. Final Report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment Phase II, 〈http://wwwitsbldrdocgov/vqeg/projects/frtv_phaseII〉, 2003.

[5] A.K. Moorthy, K. Seshadrinathan, R. Soundararajan, A.C. Bovik, Wireless video quality assessment: a study of subjective scores and objective algorithms', IEEE Transactions on Circuits and Systems for Video Technology 20 (4) (2010) 513–516.

[6] List of 3d Movies 2010, 〈http://3dmovieslist.blogspot.com/〉.

[7] H. Lee, S. Cho, K. Yun, N. Hur, J. Kim, A backward-compatible, mobile personalized, 3DTV broadcasting system based on T-DMB, Three-Dimensional Television (2010) 11–28.

[8] CISCO Corp. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010–2015, 〈http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf〉, 2010.

[9] HTC. Htc Evo 3d Overview, 〈http://www.htc.com/www/product/evo3d/overview.html〉, 2011.

[10] R. Ebert, Why I Hate 3-d (and you should too), Newsweek, 〈http://www.newsweek.com/2010/04/30/why-i-hate-3-d-and-you-should-too.html〉, 2010.

[11] M. Kermode, Come in Number 3d your Time is up, BBC News, 〈http://www.bbc.co.uk/blogs/markkermode/2009/12/come_in_number_3d_your_time_is.html〉, 2009.

[12] W. Richards, Stereopsis and stereoblindness, Experimental Brain Research 10 (4) (1970) 380–388.

[13] Test TE. Six Million Brits Can't see in 3d, 〈http://www.mcvuk.com/news/39930/Six-million-Brits-cant-see-in-3D〉, 2010.

[14] W.J. Tam, F. Speranza, S. Yano, K. Shimono, H. Ono, Stereoscopic 3d-TV: visual comfort, IEEE Transactions on Broadcasting 57 (2) (2011) 335–346.

[15] L. Meesters, W. IJsselsteijn, P. Seuntiens, A survey of perceptual evaluations and requirements of three-dimensional TV, IEEE Transactions on Circuits and Systems for Video Technology 14 (3) (2004) 381–391.

[16] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, International Journal of Computer Vision 47 (1) (2002) 7–42.

[17] Y. Liu, A. Bovik, L. Cormack, Disparity statistics in natural scenes, Journal of Vision 8 (11) (2008).

[18] Y. Liu, L.K. Cormack, A.C. Bovik, Natural scene statistics at stereo fixations, in: Symposium on Eye Tracking Research and Applications, 2010.

[19] G. Sun, N. Holliman, Evaluating methods for controlling depth perception in stereoscopic cinematography, in: Proceedings of the SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 7237, 2009.

[20] B. Mendiburu, 3D Movie Making Stereoscopic Digital Cinema from Script to Screen, Elsevier, 2008.

[21] N. Holliman, N. Dodgson, G. Favalora, L. Pockett, Three-dimensional displays: a review and applications analysis, IEEE Transactions on Broadcasting 57 (2) (2011) 362–371.

[22] B. Javidi, F. Okano, Three-Dimensional Television, Video, and Display Technologies, Springer Verlag, 2002.

[23] S. Jumisko-Pyykko, T. Utriainen, User-centered quality of experience: is mobile 3D video good enough in the actual context of use, in: Proceedings of VPQM, 2010.

[24] Z.M.P. Sazzad, S. Yamanaka, Y. Kawayoke, Y. Horita, Stereoscopic image quality prediction, in: IEEE Quality of Multimedia Experience, 2009.

[25] X. Wang, M. Yu, Y. Yang, G. Jiang, Research on subjective stereoscopic image quality assessment, in: Proceedings of SPIE, vol. 7255, 2009, p. 725509.

[26] A. Benoit, P. Le Callet, P. Campisi, R. Cousseau, Quality assessment of stereoscopic images, EURASIP Journal on Image and Video Processing 2008 (2009) 1–13.

[27] RIEGL, RIEGL VZ-400 3D Terrestrial Laser Scanner, 〈http://rieglusa.com/products/terrestrial/vz-400/index.shtml〉, 2010.

[28] R. Neuman, Bolt 3d: a case study, in: Proceedings of the SPIRE, vol. 7237, 2009.

[29] RIEGL, RIEGL RiSCAN PRO Software for 3D Terrestrial Laser Scanner, 〈http://rieglusa.com/products/terrestrial/vz-400/software.shtml〉, 2010.

[30] RIEGL. RIEGL RiSCAN PRO Software Manual, 〈http://www.riegl.com/download/?nav=display&file=389〉, 2010.

[31] Intel Corporation. OpenCV: Camera Calibration and 3D Reconstruction, 〈http://opencv.willowgarage.com/documentation/cpp/camera_calibration_and_3d_reconstruction.html〈, 2010.

[32] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Transactions on Image Processing 15 (11) (2006) 3440–3451.

[33] P. Gorley, N. Holliman, Stereoscopic image quality metrics and compression, in: SPIE Conference on Stereoscopic Displays and Applications XIX, vol. 6803, 2008.

[34] P. Seuntiens, L.M.J. Meesters, W. Ijsselsteijn, Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation, ACM Transactions on Applied Perception (TAP) 3 (2) (2006) 109.

[35] Bt-500-11: Methodology for the Subjective Assessment of the Quality of Television Pictures, International Telecommunication Union, 2002.

[36] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Transactions on Image Processing 15 (2) (2006) 430–444.

[37] M.D. Gaubatz, D.M. Rouse, S.S. Hemami, Metrix mux, 〈http://foulard.ece.cornell.edu/gaubatz/metrix_mux〉, 2010.

[38] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error measurement to structural similarity, IEEE Signal Processing Letters 13 (4) (2004) 600–612.

[39] Z. Wang, L. Lu, A. Bovik, Foveation scalable video coding with automatic fixation selection, IEEE Transactions on Image Processing 12 (2) (2003) 243.

[40] D.M. Chandler, S.S. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, IEEE Transactions on Image Processing 16 (9) (2007) 2284–2298.

[41] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Processing Letters 9 (3) (2002) 81–84.

[42] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, A. Bovik, Image quality assessment based on a degradation model, IEEE Transactions on Image Processing 9 (4) (2002) 636–650.

[43] J. Mannos, D. Sakrison, The effects of a visual fidelity criterion on the encoding of images, IEEE Transactions on Information Theory 20 (4) (1974) 525–535.

[44] M. Carnec, P. Le Callet, D. Barba, An image quality assessment method based on perception of structural information, in: IEEE International Conference on Image Processing, vol. 3, IEEE, 2003, ISBN: 0780377508.

[45] A.K. Moorthy, A.C. Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Processing Letters 17 (2) (2010) 587–599.

[46] C.T.E.R. Hewage, M.G. Martini, Reduced-reference quality metric for 3D depth map transmission, in: IEEE International Conference on Image Processing, 2010.

[47] J. You, L. Xing, A. Perkis, X. Wang, Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis, in: Proceedings of the International Workshop on Video Processing and Quality Metrics, 2010.

[48] L. Shen, J. Yang, Z. Zhang, Stereo picture quality estimation based on a multiple channel HVS model, in: The Second International Congress on Image and Signal Processing, IEEE, 2009, pp. 1–4.

[49] J. Yang, C. Hou, Y. Zhou, Z. Zhang, J. Guo, Objective quality assessment method of stereo images, in: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009, pp. 1–4.

[50] Z. Zhu, Y. Wang, Perceptual distortion metric for stereo video quality evaluation, WSEAS Transactions on Signal Processing 5 (7) (2009) 241–250.

[51] R. Akhter, Z. Sazzad, Y. Horita, J. Baltes, No reference stereoscopic image quality assessment, in: Proceedings of SPIE, vol. 7524, 2010.

[52] D. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, CRC Press, 2004.

[53] M.J. Chen, A.C. Bovik, Study on distortion conspicuity in stereoscopically viewed 3D images, in: IEEE International Workshop on Image, Video and Multidimensional Signal Processing (IVMSP), 2011.

[54] K. Seshadrinathan, A.C. Bovik, Motion tuned spatio-temporal quality assessment of natural videos, IEEE Transactions on Image Processing 19 (2) (2010) 335–350.

[55] S.G. Wardle, J. Cass, K.R. Brooks, D. Alais, Breaking camouflage: binocular disparity reduces contrast masking in natural images, Journal of Vision 10 (14:38) (2010) 1–12.

[56] Y. Liu, L.K. Cormack, A.C. Bovik, Dichotomy between luminance and disparity features at binocular fixations, Journal of Vision 10 (12) (2010).

[57] Y. Liu, L. K. Cormack, A. C. Bovik, Natural Scenes With Application to Bayesian Stereopsis, IEEE Transactions on Image Processing 20, 9, September 2011, pp. 2515–2530.