Quality Assessment of Stereoscopic 3D Image Compression by Binocular Integration Behaviors

Yu-Hsun Lin, *Student Member, IEEE*, and Ja-Ling Wu, *Fellow, IEEE*

Abstract—The objective approaches of 3D image quality assessment play a key role for the development of compression standards and various 3D multimedia applications. The quality assessment of 3D images faces more new challenges, such as asymmetric stereo compression, depth perception, and virtual view synthesis, than its 2D counterparts. In addition, the widely used 2D image quality metrics (e.g., PSNR and SSIM) cannot be directly applied to deal with these newly introduced challenges. This statement can be verified by the low correlation between the computed objective measures and the subjectively measured mean opinion scores (MOSs), when 3D images are the tested targets. In order to meet these newly introduced challenges, in this paper, besides traditional 2D image metrics, the binocular integration behaviors—the binocular combination and the binocular frequency integration, are utilized as the bases for measuring the quality of stereoscopic 3D images. The effectiveness of the proposed metrics is verified by conducting subjective evaluations on publicly available stereoscopic image databases. Experimental results show that significant consistency could be reached between the measured MOS and the proposed metrics, in which the correlation coefficient between them can go up to 0.88. Furthermore, we found that the proposed metrics can also address the quality assessment of the synthesized color-plus-depth 3D images well. Therefore, it is our belief that the binocular integration behaviors are important factors in the development of objective quality assessment for 3D images.

Index Terms—Stereo 3D image, HEVC, color plus depth 3D images, binocular frequency integration, PSNR, SSIM, MS-SSIM, VIF, VSNR, WSNR, UQI, frequency integrated metric (FI-metric), quality assessment, human visual system (HVS), stereopsis, binocular vision, ventral stream, dorsal stream.

I. INTRODUCTION

THE idiom, "A picture is worth a thousand words", has demonstrated the importance of visual information for human perceptions. Most perceptible information is represented visually in our daily life (e.g. paintings, magazines, images, videos, and 3D graphics), which enables numerous visual applications. The evaluation of the quality of various visual applications has initiated the research of image quality assessment (IQA). Generally, IQA can be classified into the categories of subjective and objective quality assessments. Since human visual system (HVS) is the final receiver of the visual information, subjective IQA is the ultimate evaluation of an image. However, there are noticeable variations among people who have different physiological and psychological status. In order to achieve a reliable subjective evaluation, appropriate procedures of subjective quality assessment had been standardized in ITU-R BT.500 [1].

Although subjective IQA provides the ultimate perceptual quality evaluation, the associated high cost and complexity handicap its value in real applications. In order to address this issue, computational objective IQA has long been an active research area since the last decade. The most widely used objective quality metric is the mean-squared error (MSE), which in turn derived the major similarity evaluation criteria (i.e. PSNR) for the image/video compression standards. However, there are lots of research work [2]–[5] demonstrated that images with the same level MSE could have noticeable different visual perceptions. Therefore, numerous objective metrics [2], [3], [6]–[8] have been studied to fill up the gaps between the objective metrics and the subjective evaluations. The study [9] utilized multi-metric fusion approach to address the 2D image IQA and showed the state-of-the-art status obtainable so far. For further detail comparison and survey of objective IQA, the studies [10]–[12], the VQEG report [13] and the references therein are highly recommended.

Owing to the booming up of 3D movies and the advances in display devices, 3D image is becoming the new research target for IQA [14]–[20], [39]. The quality assessment of anaglyph 3D images were addressed in [21] and the work [22] dealt with the quality assessment of multi-camera applications (e.g. panorama images). One direct arisen question is the applicability of existing 2D objective metrics to the 3D images [23], [24]. Notice that the study [24] focused on the quality assessment of color-plus-depth 3D images. All these research work showed that 2D objective metrics can well predict the quality of 3D images (e.g. picture quality and depth quality) only when they are **symmetric** (that is, PSNR's of the two-eye images are nearly equal to each other).

Compression plays a vital role for promoting the 3D image/ video since its huge data volume makes 3D image/video requiring more storage space and higher transmission bandwidth than that of its 2D counterpart. Fig. 1 illustrates that we can classify the types of stereo image compression according to the relation between the quantization parameters (QPs) of

Y.-H. Lin is with the Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei 106, Taiwan (e-mail: lymanblue@gmail.com).

J.-L. Wu is with the Department of Computer Science and Information Engineering, Communication and Multimedia Laboratory, Taipei 106, Taiwan (e-mail: wjl@cmlab.csie.ntu.edu.tw).
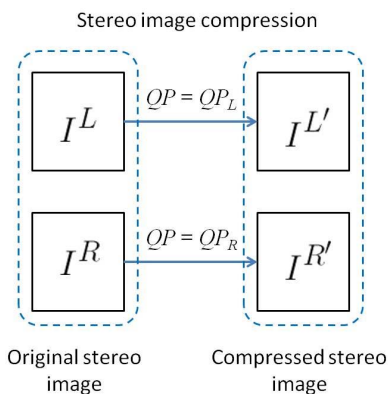
Fig. 1. The types of 3D stereo image/video compression can be classified according to the relation between the quantization parameters (QP's) of the left-eye image (QP_L) and the right-eye image (QP_R).

the left-eye image (QP_L) and the right-eye image (QP_R). That is, the types of stereo image compression can be classified as:

- Symmetric-Stereo compression: $QP_L \cong QP_R$.
- Asymmetric-Stereo compression: $QP_L \neq QP_R$.

Although [24] carried out asymmetric compression to the color and the depth images (i.e. $QP_{color} \neq QP_{depth}$), the quality of the final synthesized stereo image is close to **symmetric**. There are subjective evaluations [25], [26] of the Symmetric- and Asymmetric-Stereo compressions; however, only empirical observations are described without proposing an associated quality assessment metric. The work [27] performed Asymmetric-Stereo compression, also on the basis of some experimental observations, and found that the image encoded at lower quality in the asymmetric compressed stereo image should go beyond certain PSNR threshold for providing good overall quality; otherwise, the perceived quality of the compressed stereo image will be severely degraded. The existing 2D quality assessment metrics can predict well for the Symmetric-Stereo compression, however, the prediction results are not well addressed for the Asymmetric counterpart by the same metrics. This work aims to fill-up this gap by proposing a quality assessment metric which is applicable to both the Symmetric-Stereo and the Asymmetric-Stereo compressions.

There are numerous physiological discoveries of binocular vision where we focus on the binocular visual behaviors that describe the visual inputs integration process. For simplicity, these physiological discoveries of binocular vision are denoted as binocular integration behaviors which consist of binocular combination and binocular frequency integration behaviors. In order to overcome the challenges of 3D image IQA, we integrate the binocular integration behaviors into the existing 2D objective metrics for evaluating the quality of 3D images. We denote the integrated quality assessment metrics as the Frequency-Integrated metrics (FI-metrics). Fig. 2 and Table I demonstrate an example of the high correlation between the proposed FI-metrics and the reported MOS, for a given symmetrically/asymmetrically compressed stereo image, as compared with the normal averaged quality metrics of the left-eye and the right-eye images. Since the success of objective 3D IQA can bring new thoughts for the existing 3D visual applications and the development of

3D video compression standards [28]–[32], the significance and/or contribution of our work is quite clear.

The rest of this paper is organized as follows. Section II introduces previous research works of 3D IQA models. Section III briefly surveys the related physiological research works on HVS. Section IV describes the proposed computational framework of FI-metrics. Section V presents the experimental results. The discussions of the proposed framework and the conclusion of this work are given in Section VI.

II. PREVIOUS WORK OF 3D IQA MODELS

There are emerging research works of 3D IQA model for 3D stereoscopic images/videos. Based on the utilized information, we can classify the 3D IQA models into two categories:

- Color Information only
  The study [33] computed quality scores on the SIFT-matched feature points. A multiple channel model is adopted to estimate the stereo image quality [34]. The perceptual distortion of a stereo video is computed in discrete wavelet transform (DWT) domain [35]. The research work [36] proposed a reduced-reference (RR) quality assessment of stereo images based on the extracted edge information. The Gabor response of binocular vision was modeled in the study [37] to measure the quality of stereo images. A learning-based quality metric [38], StSD (Stereoscopic Structural Distortion), was proposed to address the quality assessment for symmetric/asymmetric video compressions (e.g. H.264, HEVC) which is the state-of-the-art 3D IQA model for 3D video compression.

- Color plus Disparity Information
  The quality of color images and the disparity maps are integrated to evaluate the overall 3D perceptual quality [39]–[41]. The research work [42] proposed an RR 3D IQA model to evaluate the quality of stereo images based on eigenvalues/eigenvectors analyses. The no-reference (NR) 3D IQA was addressed by studies [43], [44] where the performance of research work [44] is comparable or even better than some full-reference (FR) 3D IQA models. The study [45] addressed binocular rivalry issues by modeling the binocular suppression behaviors which gives the state-of-the-art FR 3D IQA model.

Accurate disparity information estimation requires high computational cost and the computed disparity map quality depends on the adopted stereo matching algorithm and the stereo image content. For real-time 3D video compression applications (e.g. 3D video camera), accurate disparity map estimation will be an extra computational burden.

Therefore, an accurate 3D IQA model based only on color information is a must for the development of real-time 3D video compression applications. Our work incorporates related binocular integration behaviors to enhance the ability to evaluate 3D image quality by color information.

III. RELATED STUDIES OF BINOCULAR VISUAL SYSTEM

Since the new challenges of 3D IQA come mainly from the interactions between the two eyes, a better understanding of the physiological studies [46]–[49] of binocular vision is

Fig. 2. The stereo image kendo is (a) symmetrically and (b) asymmetrically compressed by HEVC. The subjective evaluation (MOS) has 11 quality levels (0 for the lowest quality and 10 for the highest quality). (a) QP = 40 for both the left-eye and the right-eye images, where the corresponding MOS = 6.01. (b) QP = 30 for the left-eye image and QP = 45 for the right-eye image, where the corresponding subjective quality MOS = 5.52. The corresponding objective evaluation of different quality metrics are listed in Table I. For the best viewing (i.e. free binocular fusion) of this example, the PDF version is suggested.

TABLE I
OBJECTIVE QUALITY MEASUREMENT RESULTS OF FIG. 2

Avg-Metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
Figure 2(a) MOS_1=6.07	37.45	31.87	0.761	0.952	0.979	0.998	34.52
Figure 2(b) MOS_2=5.52	38.64	34.11	0.783	0.953	0.978	0.998	36.44
Agree with $MOS_1 > MOS_2$	0	0	0	0	**1**	0	0
FI-Metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
Figure 2(a) MOS_1=6.07	47.04	42.18	0.934	0.995	0.996	0.841	43.37
Figure 2(b) MOS_2=5.52	45.45	46.92	0.928	0.994	0.995	0.855	41.77
Agree with $MOS_1 > MOS_2$	**1**	0	**1**	**1**	**1**	0	**1**

beneficial to the development of effective computational models for 3D images. We briefly revisit the findings of binocular vision, which are related to this work, in the following.

A. Two Visual Pathways of Binocular Visual System

Fig. 3 illustrates a simplified two visual pathways representation of binocular visual system. There are two visual pathways for neural processing of visual information in visual cortex, **dorsal stream** and **ventral stream** [50].

- **Dorsal stream ("Where" pathway)**
  The dorsal stream starts from V1 (primary visual cortex), goes through V2, V3 to V5 area. The functions of dorsal stream are about visual information guided actions.

- **Ventral stream ("What" pathway)**
  The ventral stream begins from V1 area, goes through V2, V3 to V4 area. The perception and recognition visual behaviors occur in ventral stream.

The detail functions of each visual area (i.e. V1, V2, V3, V4 and V5) are not described in this work. We kindly suggest readers refer to the binocular vision book [50] for further detail descriptions of each visual area.

Since the perception function occurs in ventral stream, we will briefly describe (1) the visual response of two eyes, (2) the binocular combination behaviors in the early stage of visual pathway and (3) the visual information representations and
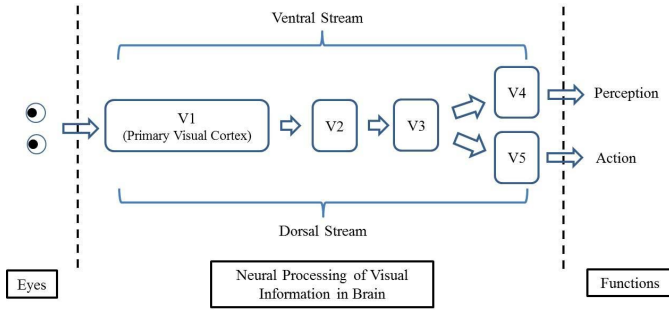
Fig. 3. A simplified illustration of ventral stream and dorsal stream for neural processing of visual information in brain.

integrations from V1 to V4. These binocular physiological behaviors will lay the foundations for the proposed quality assessment framework.

B. Visual Response of HVS

The visual response (or receptive field) of HVS can be approximated by a set of Difference-of-Gaussian (DOG) models [49], [51], [52] where the number of DOG models ranges from 2 to 4. The DOG model of an image I can be formulated as

$$(G(\sigma) - G(k\sigma)) * I, \tag{1}$$

where σ is the standard deviation of Gaussian response, $G(\cdot)$, and k is the so-called space constant and $k = 1.6$ has been derived in the studies [53]–[55] for optimal approximation when the visual response is a Laplacian of Gaussian.

C. Neural Mechanisms of Binocular Combination Behaviors

Since asymmetric compression is one of our quality assessment targets, more related physiological findings about asymmetric binocular brightness combination are needed to address this challenge. The binocular masking behavior is also addressed since masking behavior plays an important role in quality assessment of 2D images/videos.

1) The Binocular Masking Behavior: The research work [56] utilized the luminance and contrast masking to derive the just-noticeable-difference (JND) model for stereoscopic images. The concept of binocular JND has been integrated into the 3D IQA study [57] where a training process is required for fixing some parameters. However, the binocular contrast masking behavior is a complex process [58] as compared with the monocular counterpart. The binocular vision will have different behaviors among different stimuli [59]–[62]. That is, binocular vision can improve the **detection** ability (e.g. binocular JND), but the binocular **discrimination** will converge to the monocular counterpart at moderate pedestal contrast and above [59]–[61].

2) The Binocular Brightness Combination: The physiological research findings [63] about binocular combination for matched stereo images are briefly revisited in this sub-section. The binocular combination of the asymmetric brightness or contrast to each eye has been studied for decades, there are some interesting observations and the biological models reported in the literatures [63]–[65]. The perceived binocular

brightness (or image) I^B can be described as a function, f_B, of the left-eye and the right-eye brightness (or image). That is

$$I^B = f_B(I^L, I^R). \tag{2}$$

Behaviors of Binocular Combination

There are some interesting behaviors of binocular brightness combination when the input brightness is asymmetric in both eyes (e.g. Fechner's paradox, cyclopean perception). Cyclopean perception means that we will have single perceptual image when we perceive 3D images/videos by two eyes. These behaviors play the role of constraints for selecting the plausible biological models of binocular combination.

- Fechner's Paradox
  This binocular combination behaivor describes the phenomenon that a bright light to one eye may appear less bright when a dim light is shown to the other eye. That is,

$$f_B(I, 0) > f_B(I, \delta) \text{ or} \tag{3a}$$
$$f_B(0, I) > f_B(\delta, I), \tag{3b}$$

  where I is a bright light or an image and δ is a small amount of brightness. This binocular behavior is called paradox because the relation between the energy of input brightness and the energy of the perceived brightness does not have a positive correlation which is discussed in [63] for binocular combination model.

- Cyclopean Perception
  The cyclopean perception was suggested as a constraint in the research work [65] for modeling the binocular combination behavior, that is

$$f_B(I, I) \cong f_B(I, 0) \cong f_B(0, I). \tag{4}$$

  This cyclopean perception constraint is a common vision behavior in our daily life where we can perceive similar brightness for the same scene even when we close one of the two eyes.

Models of Binocular Brightness Combination

There are numerous biological models (e.g. eye-weighting model, quadratic summation model, vector summation model, and neural network model) [63] proposed to describe the binocular combination behavior. Due to page limit, we can only briefly describe the related biological models in the following, the detailed comparisons among them can be found in the study [63] and the references there in.

- Eye-Weighting Model
  This model describes the binocular combination as

$$f_B(I^L, I^R) = w_1 \cdot I^L + w_2 \cdot I^R, \tag{5}$$

  where w_1 and w_2 are the weights of the left-eye and the right-eye, respectively. There are variant models (e.g. auto-correlation model) incorporated with eye-weighting mechanisms. These models are simple but cannot explain Fencher's paradox and the cyclopean perception.

- Quadratic Summation Model
  The quadratic summation of input brightness is treated as the perceived brightness in this model, that is,

$$f_B(I^L, I^R) = \sqrt{(I^L)^2 + (I^R)^2}. \tag{6}$$

This model can explain some binocular combination behaviors (e.g. the input brightness is symmetric for both eyes), however, this model cannot explain Fencher's paradox and the cyclopean perception.

- Vector Summation Model

  The vector summation model suggests the following binocular combination,

$$f_B(I^L, I^R) = \frac{\varepsilon_L^2 + \varepsilon_R^2}{\varepsilon_L + \varepsilon_R}, \tag{7}$$

  where ε_L and ε_R are the monocular brightness flux signals of the left-eye and the right-eye, respectively. This model explains Fencher's paradox and the cyclopean perception well.

- Neural Network Model

  The neural network model computes the neural responses of each eye (e.g. excitation and inhibition) for the binocular combination. That is

$$f_B(I^L, I^R) = N_L + N_R, \tag{8}$$

  where N_L and N_R are the neural responses of the left-eye and the right-eye, respectively. This model can explain Fencher's paradox, however, the cyclopean perception may not be well-explained depending on the adopted neural response modeling [65].

- Gain-Control Theory Model

  This is one of the latest biological models for binocular combination [65] which models the binocular combination as

$$\begin{aligned} f_B(I^L, I^R) &= \left(\frac{1 + E_L}{1 + E_L + E_R}\right)I^L + \left(\frac{1 + E_R}{1 + E_L + E_R}\right)I^R \\ &= g^L \cdot I^L + g^R \cdot I^R, \end{aligned} \tag{9}$$

  where E_L and E_R are the sums of energy over all the frequency channels for the left-eye and the right-eye, respectively.

  The Gain-Control model accurately describes an early stage of binocular combination [65] and explains both Fencher's paradox and the cyclopean perception well.

D. Visual Information Processing from V1 to V4

The study [66] discovered that the depth discrimination (or disparity detector) occurs in V1 region by observing the neuron activities on cats. Further experiments about the visual system behaviors for normal RDS (Random Dot Stereogram) and anti-correlated RDS were conducted on monkeys [67]. The study [67] examined the neuron activities on monkeys for both the normal RDS and the anti-correlated RDS and suggested that "the disparity energy signals are integrated across spatial frequency channels for generating a representation of stereoscopic depth in V4."

As suggested by one of the reviewers, and also for clearness, we denote these binocular vision related physiological discoveries as binocular frequency integration (BFI) behaviors.

IV. THE PROPOSED COMPUTATIONAL FRAMEWORK

Based on the physiological discoveries from binocular visual system, the multiple-frequency-channel representation
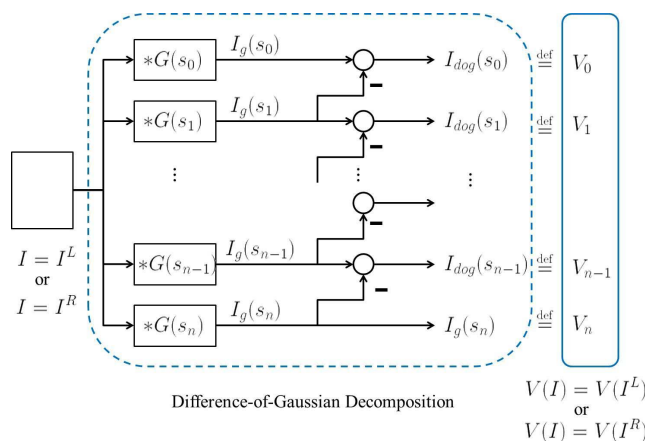

Fig. 4. The flow of Difference-of-Gaussian decomposition for a given image I (where $I = I^L$ or $I = I^R$).

will be a plausible way to address the quality assessment of stereoscopic image in an unified framework.

The Gain-Control theory model and BFI behaviors, reviewed in Section III, explain the neural mechanisms of binocular vision for the symmetric and the asymmetric brightness input signals, well. Therefore, these neural mechanisms have to be taken into consideration when we are addressing the distortion of 3D stereo images. The details of the proposed framework will be described in the following sub-sections.

A. Difference-of-Gaussian Decomposition of Stereo Images

Each view of the input stereo image is fed into a DOG filter bank which is built based on the physiological findings revisited in Section 2. Fig. 4 illustrates the flow of DOG decomposition for an image I, where I is either the left-eye image or the right-eye image (i.e. $I = I^L$ or $I = I^R$).

In order to obtain the DOG decomposition of an image I, we define a sequence of standard deviations s_i as

$$s_i = \begin{cases} 0, & i = 0 \\ 1, & i = 1 \\ k s_{i-1}, & i > 1, (k = 1.6). \end{cases} \tag{10}$$

Let $V(I)$ be the vector form representation of the DOG-bank-model of I, that is

$$\begin{aligned} V(I) &= \left[I_{dog}(s_0), \ldots, I_{dog}(s_{n-1}), I_g(s_n)\right] \\ &\stackrel{\text{def}}{=} \left[V_0, \ldots, V_{n-1}, V_n\right]. \end{aligned} \tag{11}$$

where $I_g(s_n)$ (i.e. V_n) can be regarded as the low-frequency component and $I_{dog}(s_0), \ldots, I_{dog}(s_{n-1})$ (i.e. $V_0, \ldots, V_{n-1}$) represent different spatial frequency band components of I.

B. The Binocular Combination and Frequency Integration

Following (11), the original left-eye image I^L and right-eye image I^R can respectively be represented by the DOG-bank-model as

$$V(I^L) = \left[V_0^L, \ldots, V_{n-1}^L, V_n^L\right], \tag{12}$$

and

$$V(I^R) = \left[V_0^R, \ldots, V_{n-1}^R, V_n^R\right]. \tag{13}$$

We derive the gain g_i for each DOG band based on the Gain-Control theory model given in Eq. (9). The energy of a DOG frequency band can be computed as:

$$E(V_i^L) = \sum_{q \in V_i^L} q^2, \quad (14)$$

where q is the pixel value in the DOG frequency band V_i^L.

Therefore, the gain g_i^L of the DOG frequency band V_i^L is defined as:

$$g_i^L = \frac{1 + E(V_i^L)}{1 + E_L + E_R}, \quad (15)$$

where $E_L = \sum_j (E(V_j^L))$ and $E_R = \sum_j (E(V_j^R))$.

Based on the pre-described BFI process in the ventral stream, we can derive the Frequency-Integrated metric (FI-metric) for computing the visual distortion of a 3D stereo image as:

$$\text{FI-metric} = \sum \{g_i^L D(V_i^L, V_i^{L'}) + g_i^R D(V_i^R, V_i^{R'})\}, \quad (16)$$

where $D(.)$ represents one of the 2D quality assessment metrics (e.g. PSNR, SSIM, ..., etc.).

For a given FI-metric D_{FI}, the coincidence axiom must be followed, that is:

$$D_{FI}(I_1, I_2) = 0 \text{ if and only if } I_1 = I_2. \quad (17)$$

The distorted stereo image may have $E(V_i^{L'}) = 0$ in some DOG band which leads to the gain $g_i' = 0$ and violates the axiom presented in Eq.(17). Therefore, we compute the gain g_i based on the original stereo image instead of the gain g_i' from the distorted stereo image.

C. Frequency Integrated Quality Metrics (FI-Metrics)

PSNR, WSNR [68], VSNR [7], UQI [2], SSIM [3], MS-SSIM (Multi-Scale SSIM) [4] and VIF (Visual Information Fidelity) [8] are chosen as our test quality metrics because they are the most widely used and the state-of-the-art metrics considered in various applications.

The FI-MSE between the original left-eye image I^L and the distorted left-eye image $I^{L'}$ is defined as

$$\text{FI-MSE}_L = \sum_{i=0}^{n} g_i^L \text{MSE}(V_i^L, V_i^{L'}), \quad (18)$$

which is the weighted summation of MSE's for each one of the DOG frequency bands and FI-MSE$_R$ is defined likewise.

Then, FI-PSNR can be derived from FI-MSE's of the left-eye and the right-eye images by

$$\text{FI-PSNR}_{L,R} = 10 \log_{10}\left(\frac{(2^B - 1)^2}{[\text{FI-MSE}_L + \text{FI-MSE}_R]}\right), \quad (19)$$

where B represents the required number of bits of one pixel value (e.g. $B = 8$ for the maximum pixel value equals to 255).

Similarly, FI-WSNR is defined likewise, where MSE is replaed by contrast sensitivity function (CSF)-weighted MSE among the frequency bands. In general, we can define the corresponding FI-metric of each 2D metric on the basis of Eq. (16).

V. EXPERIMENTAL RESULT

We will describe the subject data collection, which is denoted as CML database, in this work first. Then, we demonstrate various performance evaluations of the proposed FI-metrics on the CML database in the following subsections. We also apply the proposed FI-metrics to an available 3D symmetric/asymmetric video compression database (denoted as StSD 3D video database) [38], as a comparison. The performance evaluations of FI-metrics for different distortion types of stereo images are also conducted on LIVE 3D image database phase II [44], [45], as a benchmarking.

A. Subjective Data Collection of CML Database

- **Subjective Test Method**
  Our subjective evaluation follows the Double Stimulus Impairment Scale (DSIS) test method with 11 quality levels (0 for the lowest quality and 10 for the highest one) which was suggested by MPEG call for proposals on 3D video coding [28] (CfP-on-3DVC, for short).

- **Test Data**
  There are 7 test data sets in total from MPEG test sequences in CfP-on-3DVC, where the left-eye images are shown in Fig. 5. Since this work focuses mainly on the stereo image quality assessment, the first frames of each one of the video sequences are chosen as the test stimuli.

- **View Synthesis Procedure**
  The View Synthesis Reference Software (VSRS) was developed by MPEG for virtual view images generation [69]. VSRS utilizes 3D warping technique for virtual view synthesis. The relation between the intensity value $I_z(p)$ in depth map I_z and the corresponding disparity value $d(p)$ is expressed as:

$$I_z(p) = 255 \cdot \frac{d(p) - d_{\min}}{d_{\max} - d_{\min}}, \quad (20)$$

  where $d_{\max}$ and $d_{\min}$ represent the maximum and the minimum disparity values, respectively [69]. The related camera parameters and 3D information are provided in CfP-on-3DVC. Further details of VSRS operations can be found in the related MPEG document [69]. We will briefly describe the two related configurations of VSRS in the following:
  2-view configuration
  In this configuration, VSRS will synthesize one virtual view between the left-eye image and the right-eye image based on the input stereo images and the corresponding depth map of each view (i.e. 2 color images plus 2 depth maps). The newly synthesized virtual view image will become the new left-eye image of the stereo image. That is, the newly synthesized stereo image will be represented as "[the virtual-view image, the right-eye image]." Therefore, the synthesized stereo image has smaller binocular disparity values as compared with that of the original stereo image.
  3-view configuration
  The multiple view synthesis scenario is addressed by the 3-view configuration of VSRS. The 3-view input data

(a) Balloons, 1024×768 pixels   (b) Kendo, 1024×768 pixels   (c) News, 1024×768 pixels

(d) Dancer, 1920×1088 pixels   (e) Street, 1920×1088 pixels   (f) Hall, 1920×1088 pixels   (g) Fly, 1920×1088 pixels

Fig. 5. The left-eye images of the test data. The 2-view configurations [left-view-cam, (virtual-view-cam), right-view-cam] in MPEG CfP on 3DVC [28] for each one of the test images are (a) Balloon [3, (4), 5], (b) Kendo [3, (4), 5], (c) News [4, (5), 6], (d) Dancer [2, (3), 5], (e) Street [4, (3.5), 3], (f) Hall [7, (6.5), 6] and (g) Fly [5, (4), 2].

consists of the left-eye image, the center image and the right-eye image and the corresponding depth maps (i.e. 3 color images plus 3 depth maps). The 3-view input data is utilized for generating multiple virtual view images for the multi-view auto-stereoscopic 3D display.

- **Symmetric-Stereo and Asymmetric-Stereo Compressions**

  The low complexity HEVC intra compression with reference software version HM-2.0 is used to compress the target stereo images, which is the anchor codec also suggested in MPEG CfP-on-3DVC. The adopted set of quantization parameters (QPs) is $QP_{color} = \{30, 35, 40, 45\}$ for color images which was suggested to do subjective evaluation under MPEG test conditions for 3DVC [70]. The corresponding QPs for the depth images are fixed to 39, 42, 45, and 48, as also suggested in [70].

  There are four QP settings for each color image, so the number of test stereo images are $4 \times 4 = 16$ for each sequence. In total, there are $16 \times 7 = 112$ test stereo images involved in the Symmetric-Stereo and the Asymmetric-Stereo compression experiments.

  For conducting the quality assessment of synthesized stereo image, the original stereo image is symmetrically compressed and the test rendering conditions follow the suggestions of 3DVC test conditions. There are four synthesized stereo images for each sequence, and therefore, $4 \times 7 = 28$ synthesized stereo images are involved in the experiment.

- **Subjective Test Dimensions**

  - *Picture Quality*
  Subjects were asked to evaluate the picture quality between the original stereo image and the distorted stereo image based on the image sharpness, clearness where the evaluation criteria of picture quality is the same as that defined in ITU-R BT.500-13 [1].

  - *Depth Quality*
  The depth quality assessment consists of two parts: **depth cues** and **disparity**. Subjects have to evaluate the depth quality based on the provided depth cues (e.g. linear perspective, gradients, and depth sensation from pictures) and disparity between the original stereo image and the distorted stereo image. The importance of depth cues and disparity are determined by subjects which will not be specified in the experiments.

  - *Picture/Depth Quality of Synthesized Stereo Image*
  The quality assessment (for both picture quality and depth quality) are conducted between the original stereo image and the synthesized stereo image. Subjects have to evaluate the picture quality and the depth quality for the synthesized stereo image as compared with the original stereo image.

- **The Subjects**

  There are two group of subjects with background knowledge of image processing where their ages range from 20 to 35. The first group consists of 31 subjects (27 male and 4 female observers) to participate in the subjective quality evaluation experiments for the 4 test data Balloons, Kendo, Dancer and Street. The second group consists of 32 subjects (23 male and 9 female observers) participated in the experiments for the remaining 3 test data News, Hall and Fly.

  The visual acuity of the participants are all normal or corrected to normal and they all can perceive the depth from random dot stereogram. The adopted 3D display is a $16''$ wide SONY VPCF237HW 3D Notebook with 1920×1080 pixels resolution and active shutter glasses-based 3D technology. The test environment follows the suggestion of ITU-R BT.2021 [71], which is a recommendation for 3DTV subjective evaluations, the viewing distance is about 3.1H (H denotes the height of display), where the distance between two adjacent pixels will correspond to 1 arc-minute angular disparity in the retina of viewers [71].

  The participants will take a break (e.g. 10 min in our experiment) after every 30 minutes quality assessment, as suggested in ITU-R BT.2021. The standard deviations of the raw subjective scores for each test data are listed in Table II.

TABLE II
THE STANDARD DEVIATION VALUES FOR EACH
TEST DATA IN DIFFERENT SCENARIOS

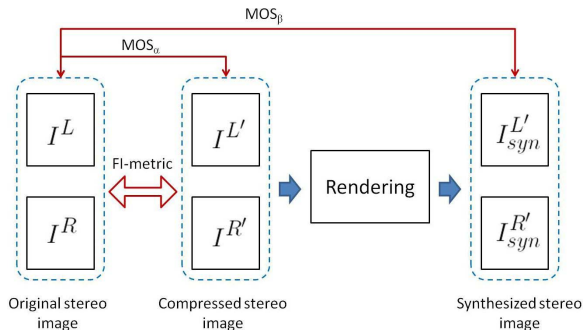| Standard deviation	Picture	Depth	Syn. Picture	Syn. Depth
Balloon	1.44	1.74	1.25	2.17
Kendo	1.30	1.65	1.34	2.53
Dancer	1.19	1.15	1.28	1.88
Street	1.38	1.33	1.23	1.95
News	1.80	1.60	1.98	2.19
Hall	1.77	1.76	1.63	2.10
Fly	1.72	1.64	1.57	1.86
All	1.51	1.55	1.47	2.1

Fig. 6. The proposed universal framework for measuring the quality of a 3D image, where MOS_α represents the direct full reference scenario for stereo IQA and MOS_β denotes the indirect IQA scenario for a synthesized stereo image.

B. The Number of Frequency bands for FI-metrics

We set $n = 4$ for DOG decomposition (i.e. 1 low-frequency band plus 4 DOG frequency bands) of a given image where the physiological study [52] suggested the appropriate number of DOG model is 4. In our pilot study of random dot stereogram with asymmetric filtering, there is no significant difference in subjective quality assessment for the subjects with different eye dominance.

C. The Performance Evaluation of FI-Metrics

There are two perceptual quality assessments (i.e. picture quality and depth quality assessments) for each stereo image and both the perceptual quality evaluations follow the descriptions of ITU-R BT.2021. The performance between the FI-metrics and the subjective MOS is evaluated based on the methods presented in ITU-R BT.500. That is, MOS are normalized from the range $[MOS_{min}, MOS_{max}]$ to $[0, 1]$. The relation between the normalized MOS and the objective FI-metrics is then approximated by a symmetrical logistic function [1]. The performance of the objective metrics are evaluated by the following 4 consistency evaluation metrics: Pearson's Linear correlation coefficient (CC), Spearman rank correlation coefficient (SRCC), outlier ratio (OR), and mean squared error (MSE). OR represents the number of the prediction error that is greater than $2e_{std}$, where e_{std} is the standard deviation of prediction differences between the subjective MOS and the regressed MOS based on the objective quality metric.

As shown in Fig. 6, there are two test scenarios for measuring the performance consistency between FI-metrics and the subjective MOS. The consistency between the subjective score MOS_α and FI-metric is used to evaluate the quality of the distorted stereo image directly. We compute the distortions in color images to access depth quality since the depth quality subjective evaluation relates to the depth cues provided by color information.

Nevertheless, the stereo image can be synthesized to generate another stereo image with different depth perception. The evaluation of the consistency between the subjective evaluation MOS_β and FI-metric is quite useful for the development of indirect quality assessment of 3D images. For example, in 3D video compression applications, the question about whether the rendering module should be integrated into the compression framework or not can be answered with the aid of this evaluation. Notice that the test data (i.e., the 2-view configuration) of MPEG CfP-on-3DVC [28] consist of 2 color images and 2 depth images, while 3DVC test conditions [70] of the synthesized stereo image include one virtual view image and one existing view image which are different to the two virtual view images scenario conducted in the study [24]. The synthesized stereo images in 3DVC test conditions have more asymmetric properties as compared with the test data adopted in [24]. We will show by experiments that the proposed computational framework, as shown in Fig. 6, can be applied to both direct and indirect quality assessments of stereo images well. Fig. 7 shows the three quality metrics (i.e. PSNR, SSIM, MS-SSIM) in which the quality assessment abilities for picture quality are improved statistically significant by the proposed FI-metrics. The detail experimental results for different compression scenarios are described in the following sub-sections.

D. The Quality Assessment of Symmetric-Stereo Compression

The averaged 2D quality metrics (i.e., averaging the quality measures of the left-eye and the right-eye images) are the most widely used quantitative method for evaluating the quality of a stereo image. Tables III and IV show the consistency of various objective quality assessment metrics with respect to the subjective MOS's, when the stereo compression is symmetric. The last rows of Tables III and IV show F-test results which demonstrate that the performance of FI-metrics is statistically significant different (i.e. 1 for better, -1 for worse) than that of the corresponding Avg-metrics in the picture quality and the depth quality, respectively. On one hand, as shown in Tables III and IV, some of the specific Avg-metrics are somewhat able to predict the picture quality of symmetrically compressed stereo images (e.g. the CC's of Avg. WSNR are higher than or equal to 0.9 for both picture quality and depth quality); on the other hand (also from the tables), FI-metrics improve the performance of PSNR, SSIM, and MS-SSIM statistically significant.

The decreasing performance consistency of FI-UQI metric comes from the fact that DOG bands contain many regions with zero intensity values. UQI metric will become unstable in these areas as stated in [3]. SSIM is an enhanced version of UQI which addressed these unstable areas, and therefore, FI-SSIM improves the performance consistency significantly.

(a) Avg-PSNR   (b) Avg-SSIM   (c) Avg-MS-SSIM
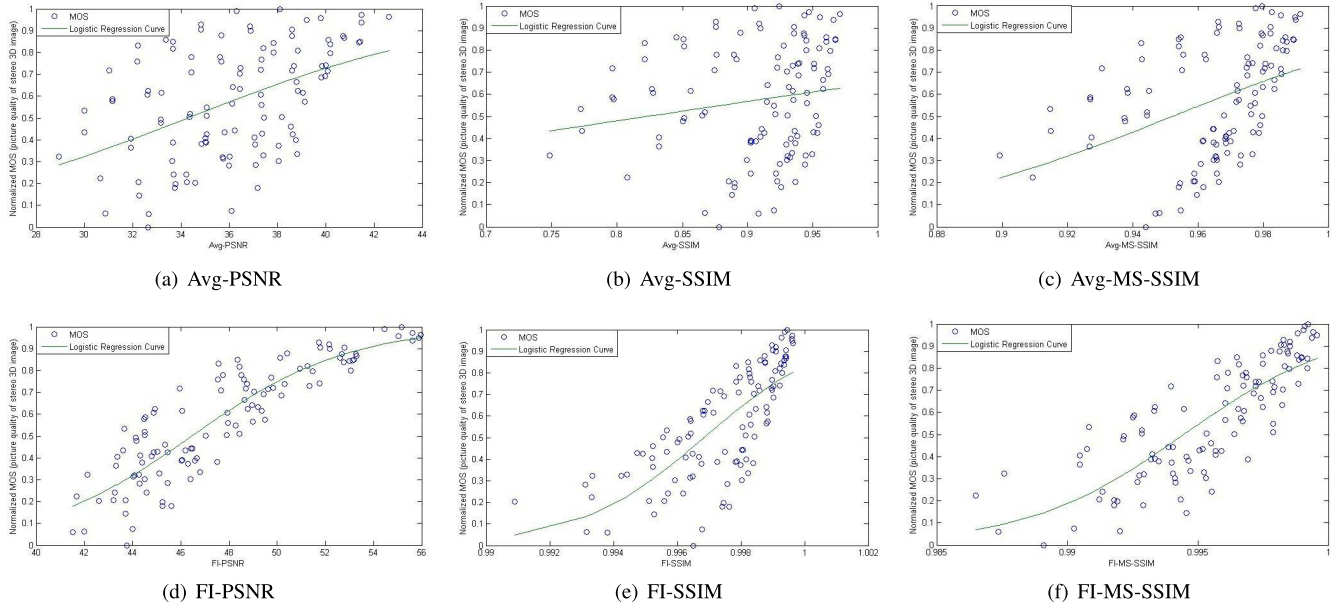
(d) FI-PSNR   (e) FI-SSIM   (f) FI-MS-SSIM

Fig. 7. The regression curves (a) Avg-PSNR, (b) Avg-SSIM, (c) Avg-MS-SSIM, (d) FI-PSNR, (e) FI-SSIM and (f) FI-MS-SSIM as compared with the normalized MOS of picture quality. These three FI-metrics improve the performance statistically significant as compared with that of the averaged counterparts.

TABLE III

THE PERFORMANCE EVALUATIONS BETWEEN THE OBJECTIVE METRICS AND THE CORRESPONDING
SUBJECTIVE MOS'S IN PICTURE QUALITY FOR SYMMETRIC-STEREO COMPRESSION

Avg-metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
CC	0.6738	0.7406	0.8848	0.3660	0.6300	0.7895	**0.9376**
SRCC	0.7001	0.7302	0.8741	0.4444	0.7083	0.8632	**0.9152**
MSE	0.0586	0.0484	0.0233	0.0929	0.0647	0.0407	**0.0130**
OR (%)	3.5714	3.5714	3.5714	0	0	3.5714	3.5714
e_{std}	0.2465	0.2241	0.1554	0.3104	0.2591	0.2055	**0.1160**
FI-metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
CC	**0.9372**	0.7991	0.9006	0.8338	0.9006	0.5305	0.9235
SRCC	**0.9113**	0.8046	0.8878	0.8987	0.9048	0.5632	0.8976
MSE	**0.0130**	0.0388	0.0203	0.0327	0.0203	0.0771	0.0158
OR (%)	7.1429	7.1429	3.5714	7.1429	3.5714	7.1429	7.1429
e_{std}	**0.1163**	0.2005	0.1450	0.1842	0.1450	0.2828	0.1280
F-test	**1**	0	0	**1**	**1**	0	0

TABLE IV

THE PERFORMANCE EVALUATIONS BETWEEN THE OBJECTIVE METRICS AND THE CORRESPONDING
SUBJECTIVE MOS'S IN DEPTH QUALITY FOR SYMMETRIC-STEREO COMPRESSION

Avg-metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
CC	0.6105	0.7397	0.8705	0.3286	0.5930	0.6916	**0.9103**
SRCC	0.6665	0.7051	0.8425	0.4007	0.6777	0.8589	**0.9013**
MSE	0.0560	0.0405	0.0216	0.0797	0.0579	0.0466	**0.0153**
OR (%)	3.5714	3.5714	7.1429	10.7143	7.1429	7.1429	**3.5714**
e_{std}	0.2411	0.2048	0.1498	0.2874	0.2451	0.2198	**0.1261**
FI-metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
CC	0.9073	0.8109	0.9142	0.8214	**0.9178**	0.5319	0.9135
SRCC	0.8994	0.7932	0.8707	0.9052	0.8986	0.5321	0.8969
MSE	0.0158	0.0306	0.0147	0.0291	**0.0141**	0.0640	0.0148
OR (%)	7.1429	7.1429	3.5714	7.1429	7.1429	7.1429	7.1429
e_{std}	0.1280	0.1781	0.1233	0.1736	**0.1208**	0.2577	0.1240
F-test	**1**	0	0	**1**	**1**	0	0

E. The Quality Assessment of Asymmetric-Stereo Compression

To evaluate the quality of stereo images involving Asymmetric-Stereo compression imposes great challenges for the traditional Avg-metrics. Tables V and VI demonstrate that the averaged 2D metrics performed poorly for predicting the quality of asymmetrically compressed stereo images. It is also clear, from Tables V and VI, the consistencies of FI-metrics with respect to the subjective MOS's are significantly better than those of Avg-metrics in both picture quality and depth

TABLE V

THE PERFORMANCE EVALUATIONS BETWEEN THE OBJECTIVE METRICS AND THE CORRESPONDING
SUBJECTIVE MOS'S IN PICTURE QUALITY FOR ASYMMETRIC-STEREO COMPRESSION

Avg-metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
CC	0.3438	0.4480	0.6785	0.0515	0.2948	0.6545	**0.8264**
SRCC	0.3323	0.4260	0.6757	0.1282	0.4046	0.7715	**0.8291**
MSE	0.0697	0.0632	0.0426	0.0788	0.0722	0.0452	**0.0251**
OR (%)	0	1.1905	1.1905	0	0	2.3810	1.1905
e_{std}	0.2656	0.2528	0.2077	0.2824	0.2702	0.2139	**0.1592**
FI-metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
CC	**0.8501**	0.5605	0.7254	0.7051	0.7623	0.2901	0.8381
SRCC	0.8142	0.5510	0.7300	0.7510	0.7794	0.3176	0.8368
MSE	**0.0219**	0.0542	0.0374	0.0397	0.0331	0.0724	0.0235
OR (%)	3.5714	4.7619	2.3810	5.9524	2.3810	1.1905	5.9524
e_{std}	**0.1489**	0.2342	0.1946	0.2005	0.1830	0.2706	0.1543
F-test	**1**	0	0	**1**	**1**	**-1**	0

TABLE VI

THE PERFORMANCE EVALUATIONS BETWEEN THE OBJECTIVE METRICS AND THE CORRESPONDING
SUBJECTIVE MOS'S IN DEPTH QUALITY FOR ASYMMETRIC-STEREO COMPRESSION

Avg-metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
CC	0.2391	0.4581	0.6150	0.0017	0.2554	0.5071	**0.6851**
SRCC	0.2465	0.4149	0.6238	0.0454	0.3576	0.6831	**0.7124**
MSE	0.0599	0.0502	0.0395	0.0636	0.0594	0.0472	**0.0337**
OR (%)	2.3810	3.5714	7.1429	2.3810	3.5714	5.9524	**1.1905**
e_{std}	0.2463	0.2255	0.2000	0.2536	0.2452	0.2186	**0.1848**
FI-metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
CC	**0.8044**	0.5738	0.7155	0.6638	0.7700	0.2874	0.7950
SRCC	**0.7856**	0.5532	0.7351	0.7436	0.8018	0.2870	0.8287
MSE	**0.0224**	0.0426	0.0310	0.0356	0.0259	0.0583	0.0234
OR (%)	4.7619	5.9524	7.1429	5.9524	2.3810	3.5714	7.1429
e_{std}	**0.1507**	0.2077	0.1772	0.1897	0.1619	0.2429	0.1539
F-test	**1**	0	0	**1**	**1**	0	0

quality (e.g. most of the CC's of FI-metrics are still higher than 0.7), even if the stereo images have been asymmetrically compressed. Some OR of FI-metrics are increased, however, this may come from the decreasing in e_{std} of FI-metrics, as compared with the cases of Avg-metrics.

F. The Quality Assessment of All Compression Combinations

Tables VII and VIII demonstrate the performances of FI-metrics for all compression combinations (i.e. symmetric-stereo and asymmetric-stereo compressions). The proposed FI-metrics improve the performances of PSNR, SSIM and MS-SSIM statistically significant. In general, FI-PSNR and FI-WSNR have the outstanding performance in various scenarios.

G. The Correlation between the Picture Quality Measure and the Depth Quality Measure

There is an interesting phenomenon worth mentioning: the measured depth quality is highly correlated to the measured picture quality (with SRCC higher than 0.92), where the corresponding regression curve is shown in Fig. 8. After further investigation, the fidelity of the semantic depth cue provided by the corresponding color image plays an important role in this high correlation characteristic. The subjects stated that the distorted semantic depth cues will decrease the depth perception even though the binocular disparity is the same. In other words, this highly correlated relation between color and
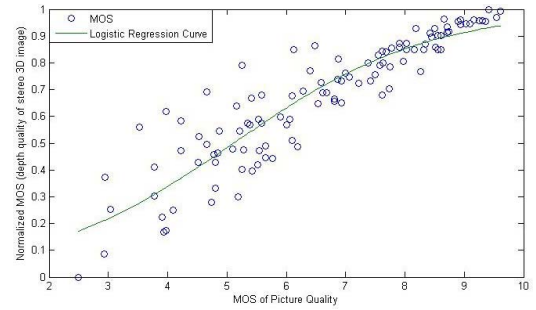
Fig. 8. The y-axis represents the normalized depth quality MOS and the x-axis denotes the picture quality MOS, where the logistic regression curve fits the MOS quite well. (CC = 0.9044, SRCC = 0.9271).

depth images may not be valid for the stereo images without semantic depth cues (e.g. with Random Dot Stereogram as input). This highly correlated relation between picture quality and depth quality may suggest that depth quality can be accessed by computing the distortion of color images. This phenomenon will be affected by rendering operations which are addressed in the following sub-sections.

H. The Indirect Quality Assessment of Synthesized Stereo Images

In this sub-section, let's focus on the indirect quality assessment of the synthesized stereo images in which the input stereo

TABLE VII

THE PERFORMANCE EVALUATIONS BETWEEN THE OBJECTIVE METRICS AND THE CORRESPONDING
SUBJECTIVE MOS'S IN PICTURE QUALITY FOR ALL COMPRESSION COMBINATIONS

Avg-metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
CC	0.4666	0.5522	0.7605	0.1650	0.4172	0.6794	**0.8644**
SRCC	0.4600	0.5264	0.7543	0.2409	0.5170	0.8135	**0.8637**
MSE	0.0496	0.0441	0.0267	0.0617	0.0524	0.0341	**0.0160**
OR (%)	0.8929	0	1.7857	2.6786	0	2.6786	1.7857
e_{std}	0.2237	0.2108	0.1642	0.2494	0.2298	0.1856	**0.1272**
FI-metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
CC	**0.8831**	0.6445	0.7927	0.7516	0.8143	0.3751	0.8679
SRCC	**0.8775**	0.6267	0.7988	0.8065	0.8379	0.3930	0.8731
MSE	**0.0140**	0.0371	0.0236	0.0276	0.0214	0.0545	0.0156
OR (%)	5.3571	6.2500	1.7857	5.3571	3.5714	1.7857	7.1429
e_{std}	**0.1187**	0.1934	0.1542	0.1668	0.1468	0.2344	0.1256
F-test	**1**	0	0	**1**	**1**	**-1**	0

TABLE VIII

THE PERFORMANCE EVALUATIONS BETWEEN THE OBJECTIVE METRICS AND THE CORRESPONDING
SUBJECTIVE MOS'S IN DEPTH QUALITY FOR ALL COMPRESSION COMBINATIONS

Avg-metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
CC	0.3798	0.5614	0.7193	0.1232	0.3851	0.5675	**0.7687**
SRCC	0.3896	0.5101	0.7171	0.1745	0.4821	0.7520	**0.7814**
MSE	0.0446	0.0357	0.0251	0.0513	0.0444	0.0353	**0.0213**
OR (%)	4.4643	4.4643	8.9286	3.5714	3.5714	6.2500	**3.5714**
e_{std}	0.2121	0.1898	0.1593	0.2275	0.2116	0.1888	**0.1467**
FI-metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
CC	**0.8440**	0.6565	0.7952	0.7263	0.8280	0.3766	0.8343
SRCC	0.8605	0.6263	0.8009	0.8023	0.8551	0.3680	0.8678
MSE	**0.0150**	0.0296	0.0192	0.0246	0.0164	0.0447	0.0158
OR (%)	**3.5714**	6.2500	7.1429	4.4643	4.4643	4.4643	5.3571
e_{std}	**0.1229**	0.1729	0.1391	0.1576	0.1286	0.2124	0.1264
F-test	**1**	0	0	**1**	**1**	0	0

images have been symmetrically HEVC compressed. Since this is an indirect quality assessment, as illustrated in Fig. 6, the consistency evaluation between FI-metric and MOS_β is our target. In order to analyze the effect of view synthesis process solely for IQA, only symmetrically compressed stereo images are taken into account, in this experiment.

Table IX shows the consistency evaluations of Avg-metrics for indirect picture quality assessments are higher than those of their direct counterparts in assessing the quality of the same asymmetrically compressed stereo images (see Table V). This is because the two-eye images of stereo input have nearly the same PSNRs in nature, so the traditional Avg-metrics worked well. However, the synthesized stereo images have certain asymmetric characteristics due to the rendering operations which make the performance evaluations of Avg-metrics not so good as those obtained by using FI-metrics (e.g. the CC of FI-PSNR is higher than 0.92, as shown in Table IX).

Table X illustrates the performance evaluations of indirect depth quality assessment of the synthesized stereo images. Fig. 9 indicates the correlation between the picture quality MOS and the depth quality MOS of the synthesized stereo image is, as expected, a bit lower (SRCC = 0.9068) than that of its direct assessment counterpart.

This experiment verified that the rendering operations for synthesizing stereo images will decrease the correlation between the picture quality and the depth quality. This fact suggests that rendering modules should be included into the
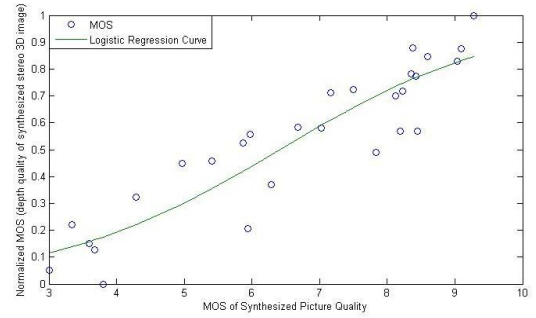
Fig. 9. The y-axis represents the normalized depth quality MOS of the synthesized stereo images and the x-axis denotes the picture quality MOS of the synthesized stereo images, once again, the logistic regression curve fits the MOS reasonably good (CC = 0.9082, SRCC = 0.9068).

compression framework if good depth quality perception is a must.

I. The 3D IQA Model Comparisons on CML database

StSD [38] is the state-of-the-art quality metric for 3D video compression which also deals with symmetric/asymmetric video compression by learning techniques. The released $StSD_{LC}$ metric is the low complexity version of StSD quality metric. The quality metric Cyclopean MS-SSIM [45] models binocular suppression behaviors by incorporating disparity information to address the binocular rivalry scenario which is one of the state-of-the-art FR 3D IQA models.

TABLE IX

THE PERFORMANCE EVALUATIONS BETWEEN THE OBJECTIVE METRICS AND CORRESPONDING
SUBJECTIVE MOS'S IN PICTURE QUALITY ASSESSMENT FOR THE SYNTHESIZED STEREO IMAGES

Avg-metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
CC	0.6598	0.6638	0.8654	0.3606	0.6274	0.8331	**0.9309**
SRCC	0.6409	0.6207	0.8068	0.3804	0.6513	0.8971	0.8960
MSE	0.0544	0.0539	0.0242	0.0838	0.0584	0.0297	**0.0129**
OR (%)	3.5714	3.5714	0	0	0	3.5714	7.1429
e_{std}	0.2375	0.2364	0.1583	0.2947	0.2461	0.1754	**0.1156**
FI-metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
CC	**0.9263**	0.7327	0.8834	0.8569	0.8961	0.4865	0.9122
SRCC	0.8555	0.7050	0.8139	0.8648	0.8347	0.4663	0.8648
MSE	**0.0137**	0.0446	0.0212	0.0256	0.0190	0.0735	0.0162
OR (%)	7.1429	7.1429	3.5714	7.1429	3.5714	0	0
e_{std}	**0.1191**	0.2151	0.1481	0.1629	0.1404	0.2761	0.1296
F-test	**1**	0	0	**1**	**1**	**-1**	0

TABLE X

THE PERFORMANCE EVALUATIONS BETWEEN THE OBJECTIVE METRICS AND THE CORRESPONDING
SUBJECTIVE MOS'S IN DEPTH QUALITY ASSESSMENT FOR THE SYNTHESIZED STEREO IMAGES

Avg-metrics	PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	WSNR
CC	0.4585	0.4050	0.6591	0.1214	0.3991	**0.8672**	0.7930
SRCC	0.4683	0.3925	0.6153	0.1325	0.4333	**0.9216**	0.7809
MSE	0.0569	0.0602	0.0407	0.0710	0.0606	**0.0180**	0.0267
OR (%)	0	3.5714	3.5714	3.5714	3.5714	3.5714	0
e_{std}	0.2429	0.2499	0.2055	0.2713	0.2506	**0.1365**	0.1666
FI-metrics	FI-PSNR	FI-VSNR	FI-VIF	FI-SSIM	FI-MS-SSIM	FI-UQI	FI-WSNR
CC	0.8059	0.5324	0.7034	**0.8161**	0.7475	0.2610	0.7751
SRCC	0.7415	0.5305	0.6435	**0.8307**	0.7029	0.2299	0.7420
MSE	0.0253	0.0516	0.0364	**0.0241**	0.0318	0.0671	0.0288
OR (%)	0	10.7143	7.1429	3.5714	3.5714	3.5714	0
e_{std}	0.1618	0.2314	0.1943	**0.1580**	0.1816	0.2639	0.1727
F-test	**1**	0	0	**1**	0	**-1**	0

TABLE XI

THE PERFORMANCE COMPARISON BETWEEN VARIOUS 3D IQA MODELS
ON CML DATABASE IN PICTURE QUALITY

Metrics	CC	SRCC	MSE	OR (%)	e_{std}
Cyclopean MS-SSIM	0.3638	0.4351	0.0550	**1.7857**	0.2356
$StSD_{LC}$	0.5858	0.5629	0.0417	3.5714	0.2050
FI-SSIM	0.7516	0.8065	0.0276	5.3571	0.1668
FI-MS-SSIM	0.8143	0.8379	0.0214	3.5714	0.1468
FI-PSNR	**0.8831**	**0.8775**	**0.0140**	5.3571	**0.1187**
FI-WSNR	0.8679	0.8731	0.0156	7.1429	0.1256

TABLE XII

THE PERFORMANCE COMPARISON BETWEEN VARIOUS 3D IQA MODELS
ON CML DATABASE IN DEPTH QUALITY

Metrics	CC	SRCC	MSE	OR (%)	e_{std}
Cyclopean MS-SSIM	0.2533	0.3154	0.0488	4.4643	0.2218
$StSD_{LC}$	0.4745	0.4443	0.0404	**3.5714**	0.2018
FI-SSIM	0.7263	0.8023	0.0246	4.4643	0.1576
FI-MS-SSIM	0.8280	0.8551	0.0164	4.4643	0.1286
FI-PSNR	**0.8440**	0.8605	**0.0150**	**3.5714**	**0.1229**
FI-WSNR	0.8343	**0.8678**	0.0158	5.3571	0.1264

TABLE XIII

THE PERFORMANCE COMPARISON BETWEEN VARIOUS 3D IQA MODELS
ON CML DATABASE IN SYNTHESIZED PICTURE QUALITY

Metrics	CC	SRCC	MSE	OR (%)	e_{std}
Cyclopean MS-SSIM	0.6603	0.6836	0.0543	3.5714	0.2373
$StSD_{LC}$	0.6656	0.6754	0.0537	**0**	0.2360
FI-SSIM	0.8569	**0.8648**	0.0256	7.1429	0.1629
FI-MS-SSIM	0.8961	0.8347	0.0190	3.5714	0.1404
FI-PSNR	**0.9263**	0.8555	**0.0137**	7.1429	**0.1191**
FI-WSNR	0.9122	**0.8648**	0.0162	**0**	0.1296

TABLE XIV

THE PERFORMANCE COMPARISON BETWEEN VARIOUS 3D IQA MODELS
ON CML DATABASE IN SYNTHESIZED DEPTH QUALITY

Metrics	CC	SRCC	MSE	OR (%)	e_{std}
Cyclopean MS-SSIM	0.4352	0.4776	0.0584	3.5714	0.2461
$StSD_{LC}$	0.5562	0.4765	0.0498	3.5714	0.2271
FI-SSIM	**0.8161**	**0.8307**	**0.0241**	3.5714	**0.1580**
FI-MS-SSIM	0.7425	0.7029	0.0318	3.5714	0.1816
FI-PSNR	0.8059	0.7415	0.0253	**0**	0.1618
FI-WSNR	0.7751	0.7420	0.0288	**0**	0.1727

We compare the performances of these two state-of-the-art 3D IQA models on our CML database where the performance evaluation results are listed in Tables XI, XII, XIII, and XIV. The original parameter $d_{\max}$ (maximum disparity) in Cyclopean MS-SSIM is set to 25 pixels for the test stereo image with image width 640 pixels in 3D LIVE 3D image database phase II. Therefore, we set the parameter $d_{\max}$ linearly proportional to the image width of the test data (i.e. $d_{\max} = 40$ for test data with 1024×768 pixels and $d_{\max} = 75$

for test data with 1920×1088 pixels). The proposed FI-metrics outperform $StSD_{LC}$ and Cyclopean MS-SSIM quality metrics in both picture quality and depth quality.

Table XV illustrates the performance evaluation of CC for each test data. The performance of $StSD_{LC}$ may be compromised due to the lack of temporal information since the original usage of $StSD_{LC}$ is for quality assessment of 3D video compression. The Cyclopean MS-SSIM provides

TABLE XV

CC FOR EACH TEST DATA OF EACH 3D IQA MODEL ON CML DATABASE IN PICTURE QUALITY

CC	Cyclopean MS-SSIM	$StSD_{LC}$	FI-SSIM	FI-MS-SSIM	FI-PSNR	FI-WSNR
Balloon	0.9055	0.8626	0.9795	0.9764	**0.9827**	0.9826
Kendo	0.9572	0.8699	**0.9905**	0.9835	0.9844	0.9843
Dancer	0.5928	0.7717	0.9835	**0.9837**	0.9710	0.9701
Street	0.8324	0.7979	**0.9868**	0.9866	0.9851	0.9849
News	0.9367	0.8442	0.9629	0.9605	**0.9671**	0.9669
Hall	0.8259	0.8526	0.9796	0.9795	**0.9821**	0.9820
Fly	0.7734	0.8435	0.9691	**0.9771**	0.9715	0.9712

TABLE XVI

TIME COMPLEXITY COMPARISON OF EACH 3D IQA MODEL

Time (sec)	$StSD_{LC}$	FI-metrics in Table XV	Cyclopean-MS-SSIM
1024×768	1.24	5.12 (Avg.)	167.28
1920×1088	2.01	13.77 (Avg.)	637.16

good performance results in some test data (e.g. Balloons, Kendo and News), but its performances are compromised for the stereo image contents which are difficult for doing disparity map estimation (e.g. Dancer and Fly). The performance of Cyclopean MS-SSIM would be improved by adopting a better stereo matching algorithm or an optimized parameter setting of d_{max}. However, these extra computational intensive fine-tune processes will be obstacles for the real-time 3D video compression applications. On the other hand, the proposed FI-metrics providing outstanding performance in each test data.

Table XVI shows the time complexity comparison among the 3D IQA models. These 3D IQA models are implemented by MATLAB and the platform is Intel Core 2 Quad Q9500 2.83 GHz with 4 GB memory. Due to the need of disparity map estimation, Cyclopean MS-SSIM requires more computational time as compared with other 3D IQA models.

J. Proposed Metrics on 3D Video Compression Database

The 3D video compression database [38] consists of symmetric/asymmetric video compression results by H.264 and HEVC codecs. There are training set and testing set for the development of learning-based StSD quality metric [38]. For comparison, we applied FI-SSIM, FI-MS-SSIM, FI-PSNR and FI-WSNR to the testing set of 3D video compression database. Table XVII demonstrates the performance results of Avg-metrics and the proposed FI-metrics. The performance of FI-PSNR is comparable[1] to that of the full version StSD quality metric reported in [38] (e.g. CC = 0.9591 and SRCC = 0.9713). This result demonstrates the proposed FI-PSNR is comparable to the state-of-the-art 3D video quality metric in performance measure.

K. Analysis

There are some issues should be considered when applying the proposed framework to a given 2D quality metric.

- **The effect of DOG bands**
  Since each quality metric has its own specific visual properties and background assumptions, a quality metric could not perform well at DOG frequency bands which are far from the original usages (e.g. UQI is unstable

in the regions with zero values and the original usage of VIF is for natural image only). The proposed DOG decomposition may be replaced by Gabor filter bank which is another popular physiological model [72] and the corresponding performance evaluation will be one of our future works.

- **The effect of the adopted 2D visual properties**
  Some 2D quality metrics (e.g. VSNR, WSNR) have already utilized some 2D visual properties for quality assessment. The visual behavior in binocular vision will have different effects on the adopted visual properties. For example, the binocular behavior described in Section III-C.1 states that the ability of binocular detection is improved whereas the binocular discrimination is converged to normal 2D masking. Therefore, the binocular viewing condition will have more effects on VSNR [7] (i.e. based on near threshold properties) as compared with WSNR [68] that adopts CSF-based (i.e. masking) visual property. The masking behavior is relatively stable as compared to other visual properties which explains why WSNR outperforms other 2D metrics.

- **Interplay between the 2D metrics and the framework**
  The performance of FI-metrics may be affected if there is a conflict between the adopted visual property of a 2D quality metric and the binocular viewing condition. Therefore, the quality metrics with fewer visual assumptions (i.e. PSNR, SSIM, MS-SSIM) or the quality metric with more stable visual properties in binocular viewing conditions (i.e. masking behavior in WSNR) would have much better performances. These results suggest that the integration of other existing 2D metrics and the proposed framework needs further investigation since the adopted visual behaviors in 2D metrics may have different actions in 3D viewing conditions, which may open up a new research direction for 3D IQA model development.

- **The gain values of the left-eye image and the right-eye image**
  The gain values of the left-eye image and the right-eye image are listed in Table XVIII. The Gain-Control model explains the binocular combination behaviors in various scenarios. In normal situation, the binocular combination may have the same gain value for each band of the two-

---

[1]Due to copyright issue, one of the testing set videos, Water Splash, is not released.

TABLE XVII
PERFORMANCE EVALUATION ON STSD 3D VIDEO COMPRESSION DATABASE

Metrics	Avg-SSIM	Avg-MS-SSIM	Avg-PSNR	Avg-WSNR	FI-SSIM	FI-MS-SSIM	FI-PSNR	FI-WSNR
CC	0.7248	0.9095	0.8275	0.8665	0.9430	0.9466	**0.9605**	0.9123
SRCC	0.7687	0.9300	0.8582	0.8781	0.9581	**0.9685**	0.9661	0.8891
MSE	0.0368	0.0134	0.0244	0.0193	0.0086	0.0080	**0.0060**	0.0130
OR (%)	**0**	5.8834	8.8235	5.8824	**0**	2.9412	8.8235	2.9412
e_{std}	0.1946	0.1174	0.1585	0.1409	0.0939	0.0910	**0.0786**	0.1156

TABLE XVIII
THE GAIN FOR EACH TEST DATA

Test Data	Gain	V_0	V_1	V_2	V_3	V_4
Balloons	g_i^L	0.0008	0.0001	0.0003	0.0005	0.5005
	g_i^R	0.0008	0.0002	0.0003	0.0006	0.4959
Kendo	g_i^L	0.001	0.0002	0.0003	0.0006	0.5044
	g_i^R	0.001	0.0002	0.0003	0.0006	0.4914
Dancer	g_i^L	0.0014	0.0002	0.0003	0.0004	0.4943
	g_i^R	0.0014	0.0002	0.0003	0.0005	0.5009
Street	g_i^L	0.0013	0.0003	0.0005	0.0008	0.4966
	g_i^R	0.0013	0.0003	0.0005	0.0008	0.4976
News	g_i^L	0.0007	0.0002	0.0003	0.0007	0.4938
	g_i^R	0.0009	0.0002	0.0004	0.0007	0.5022
Hall	g_i^L	0.0009	0.0002	0.0003	0.0005	0.4963
	g_i^R	0.0011	0.0002	0.0003	0.0005	0.4998
Fly	g_i^L	0.0011	0.0001	0.0002	0.0003	0.4983
	g_i^R	0.0011	0.0001	0.0002	0.0003	0.4983

TABLE XIX
CC RESULTS OF LIVE 3D DATASET PHASE 2

IQA Type	CC	WN	JP2K	JPEG	Blur	FF	ALL
Color	Gorley	0.874	0.372	0.322	0.934	0.706	0.515
Color	Hewage	0.891	0.664	0.734	0.450	0.746	0.558
Color	FI-SSIM	0.916	0.745	0.488	0.726	0.745	0.691
Color	FI-MS-SSIM	0.933	0.867	0.874	0.707	0.745	0.729
Color	FI-PSNR	0.915	0.747	0.617	0.739	0.739	0.662
Color	FI-WSNR	0.961	0.908	0.827	0.771	0.702	0.705
Color + Disparity	Benoit	0.926	0.784	0.853	0.535	0.807	0.748
Color + Disparity	You	0.912	0.905	0.830	0.784	0.915	0.800
Color + Disparity	Cyclopean MS-SSIM	0.957	0.834	0.862	0.963	0.901	0.900

TABLE XX
SRCC RESULTS OF LIVE 3D DATASET PHASE 2

IQA Type	SRCC	WN	JP2K	JPEG	Blur	FF	ALL
Color	Gorley	0.875	0.110	0.027	0.770	0.601	0.146
Color	Hewage	0.880	0.598	0.736	0.028	0.684	0.501
Color	FI-SSIM	0.909	0.700	0.564	0.739	0.735	0.680
Color	FI-MS-SSIM	0.929	0.849	0.858	0.746	0.709	0.712
Color	FI-PSNR	0.907	0.719	0.631	0.711	0.701	0.638
Color	FI-WSNR	0.955	0.901	0.807	0.757	0.684	0.684
Color + Disparity	Benoit	0.923	0.751	0.867	0.455	0.773	0.728
Color + Disparity	You	0.909	0.894	0.795	0.813	0.891	0.786
Color + Disparity	Cyclopean MS-SSIM	0.940	0.814	0.843	0.908	0.884	0.889

eyes (e.g. Fly). However, there are emerging 3D + 2DTV 3D display systems with asymmetric illuminations which resolved the issues of viewers without 3D glasses [73]. The transcoding applications for Asymmetric-Stereo compressed stereo images will also bring new challenges to the 3D IQA research. The applications of the proposed FI-metrics for these newly introduced challenging scenarios will be investigated in our future work.

L. Proposed Metrics for Different Distortion Types

For further testing the effectiveness of the proposed frequency integration behavior, we apply FI-SSIM, FI-MS-SSIM, FI-PSNR and FI-WSNR to another publicly available 3D subjective evaluation database [44], [45] (LIVE 3D IQA-Phase II). The database consists of 8 reference stereo images and 360 distorted stereo images with various symmetric/ asymmetric distortion types: JPEG, JPEG 2000 (JP2K), Gaussian blur (Blur), Gaussian noise (WN), and fast-fading (FF) based on the Rayleigh fading channel.

We compared the performances of various 3D IQA models with different distortion types in this database. The color information based 3D IQA models include: Gorley [33] and Hewage [36], while 3D IQA models based on color plus disparity information include: Benoit [39], You [41] and Cyclopean MS-SSIM [45]. The performance comparison results

(we only listed the results of CC and SRCC due to page limit) are reported in Tables XIX and XX.

The proposed FI-metrics outperform the 3D IQA models, if color information is used only. The performances of FI-metrics are comparable to or even better than that of the color plus disparity information 3D IQA models in compression-like distortions (e.g. WN, JPEG and JP2K). The distortion of Blur and FF will introduce binocular rivalry which requires the disparity information to suppress the response of rivalry regions.

These results demonstrate the proposed FI-MS-SSIM, FI-PSNR and FI-WSNR provide the most outstanding performances in compression-like distortions. In order to improve the performance of the proposed FI-metrics, addressing the binocular rivalry scenarios from distortion types will be one of our future works.

VI. Conclusion

The objective quality assessment of stereo images plays a key role for the development of 3D image/video compression standards and the success in enriching various 3D visual applications. The traditional 2D objective metrics are suitable only for addressing the stereo images with Symmetric-Stereo compression. However, the perfect symmetric quality scenario is very unusual to be found in practical stereo images which makes the traditional 2D objective metrics do not work well in 3D IQA. This work revisits the physiological findings in HVS and proposes a binocular integration behavior based computational framework for dealing with the challenges arisen in 3D IQA. Experimental results demonstrate that significant improvement in performance consistencies between the proposed FI-metrics and the corresponding subjective MOS's can be achieved for both picture quality and depth quality assessments, even if the stereo images are synthesized from another stereo images. These binocular behaviors shed a light on a new paradigm for the developments of IQA on 3D images/videos. This also provides new opportunities to investigate the new quality metrics for 3D image/video IQA. To investigate the effects of different types of distortions and seek for a more powerful 3D IQA computational framework will, of course, be the possible directions of our future work.

Acknowledgment

References

[1] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU Recommendation BT.500-13, Jan. 2012.

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

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

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

[5] Z. Wang and A. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[6] M. Sampat, Z. Wang, S. Gupta, A. Bovik, and M. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2385–2401, Nov. 2009.

[7] D. Chandler and 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.

[8] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[9] T.-J. Liu, W. Lin, and C.-C. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793–1807, May 2013.

[10] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[11] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[12] Y. Tong, H. Konik, F. Cheikh, and A. Tremeau, "Full reference image quality assessment based on saliency map analysis," *J. Imag. Sci.*, vol. 54, no. 3, pp. 30503-1–30503-14, 2010.

[13] *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*, document VQEG, Jun. 2000.

[14] Y. Liu, L. Cormack, and A. Bovik, "Statistical modeling of 3D natural scenes with application to Bayesian stereopsis," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2515–2530, Sep. 2011.

[15] L. Jin, A. Boev, A. Gotchev, and K. Egiazarian, "3D-DCT based perceptual quality assessment of stereo video," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2521–2524.

[16] L. Meesters, W. Ijsselsteijn, and P. Seuntiens, "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.

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

[18] Z. Mai, C. Doutre, P. Nasiopoulos, and R. K. Ward, "Rendering 3D high dynamic range images: Subjective evaluation of tone-mapping methods and preferred 3D image attributes," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 5, pp. 597–610, Sep. 2012.

[19] M. Mikkola, S. Jumisko-Pyykko, D. Strohmeier, A. Boev, and A. Gotchev, "Stereoscopic depth cues outperform monocular ones on autostereoscopic display," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 698–709, Oct. 2012.

[20] P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet, "Evaluating depth perception of 3D stereoscopic videos," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 710–720, Oct. 2012.

[21] M. Park, J. Luo, and A. C. Gallagher, "Toward assessing and improving the quality of stereo images," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 5, pp. 460–470, Sep. 2012.

[22] M. Solh and G. AlRegib, "MIQM: A multicamera image quality measure," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3902–3914, Sep. 2012.

[23] C. Hewage, S. Worrall, S. Dogan, and A. Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2D video," *Electron. Lett.*, vol. 44, no. 16, pp. 963–965, Jul. 2008.

[24] C. Hewage, S. Worrall, S. Dogan, S. Villette, and A. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 304–318, Apr. 2009.

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

[26] P. Aflaki, M. Hannuksela, J. Hakkinen, P. Lindroos, and M. Gabbouj, "Subjective study on compressed asymmetric stereoscopic video," in *Proc. IEEE Int. Conf. Image Prcoess.*, Sep. 2010, pp. 4021–4024.

[27] G. Saygili, C. Gurler, and A. Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 593–601, Jun. 2011.

[28] *Call for Proposals on 3D Video Coding Technology*, ISO/IEC Standard JTC1/SC29/WG11 MPEG2011/N12036, Mar. 2011.

[29] Y.-H. Lin and J.-L. Wu, "A depth information based fast mode decision algorithm for color plus depth-map 3D videos," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 542–550, Jun. 2011.

[30] Y.-H. Lin and J.-L. Wu, "A digital blind watermarking for depth-image-based rendering 3D images," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 602–611, Jun. 2011.

[31] Y.-H. Lin and J.-L. Wu, "Rendering lossless compression of depth image," in *Proc. Data Compress. Conf.*, Mar. 2011, p. 467.

[32] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, "Three dimensional scalable video adaptation via user-end perceptual quality assessment," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 719–727, Sep. 2008.

[33] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE Electron. Imag., Stereosc. Displays Appl. XIX*, vol. 6803, pp. 1–12, Feb. 2008.

[34] L. Shen, J. Yang, and Z. Zhang, "Stereo picture quality estimation based on a multiple channel HVS model," in *Proc. Int. Congr. Image Signal Process.*, 2009, pp. 1–4.

[35] 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.

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

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

[38] V. De Silva, H. Arachchi, E. Ekmekcioglu, and A. Kondoz, "Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3392–3404, Sep. 2013.

[39] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, no. 659024, pp. 1–13, Jan. 2009.

[40] 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*, 2009, pp. 1–4.

[41] 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*, 2010, pp. 1–6.

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

[43] R. Akhter, Z. M. P. Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," *Proc. SPIE, Stereosc. Displays Appl. XXI*, vol. 7524, pp. 75240T-1–75240T-12, Feb. 2010.

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

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

[46] R. Blake and H. Wilson, "Binocular vision," *Vis. Res.*, vol. 51, no. 7, pp. 754–770, 2011.

[47] S. Steinman, B. Steinman, and R. Garzia, *Foundations of Binocular Vision: A Clinical Perspective*, 1st ed. New York, NY, USA: McGraw-Hill, 2000.

[48] I. P. Howard and B. J. Rogers, *Seeing in Depth*. New York, NY, USA: Oxford Univ. Press, 2008.

[49] G. Mather, *Foundations of Sensation and Perception*. Oxon, U.K.: Psychology Press, 2008.

[50] D. Stidwill and R. Fletcher, *Normal Binocular Vision: Theory, Investigation and Practical Aspects*, 1st ed. New York, NY, USA: Wiley, 2010.

[51] H. R. Wilson, "Quantitative characterization of two types of line-spread function near the fovea," *Vis. Res.*, vol. 18, no. 8, pp. 971–981, 1978.

[52] H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," *Vis. Res.*, vol. 19, no. 1, pp. 19–32, 1979.

[53] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Royal Soc. Lond. B*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.

[54] D. Marr, *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. Cambridge, MA, USA: MIT Press, 2010.

[55] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Royal Soc. B, Biol. Sci.*, vol. 204, no. 1156, pp. 301–328, 1979.

[56] Y. Zhao, Z. Chen, C. Zhu, Y.-P. Tan, and L. Yu, "Binocular just-noticeable-difference model for stereoscopic images," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 19–22, Jan. 2011.

[57] 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.

[58] D. H. Baker and T. S. Meese, "Binocular contrast interactions: Dichoptic masking is not a single process," *Vis. Res.*, vol. 47, no. 24, pp. 3096–3107, 2007.

[59] G. E. Legge, "Binocular contrast summation—I. Detection and discrimination," *Vis. Res.*, vol. 24, no. 4, pp. 373–383, 1984.

[60] G. E. Legge, "Binocular contrast summation—II. Quadratic summation," *Vis. Res.*, vol. 24, no. 4, pp. 385–394, 1984.

[61] T. S. Meese, M. A. Georgeson, and D. H. Baker, "Binocular contrast vision at and above threshold," *J. Vis.*, vol. 6, no. 11, pp. 1224–1243, 2006.

[62] G. E. Legge, "Spatial frequency masking in human vision: Binocular interactions," *J. Opt. Soc. Amer.*, vol. 69, no. 6, pp. 838–847, 1979.

[63] S. Grossberg and F. Kelly, "Neural dynamics of binocular brightness perception," *Vis. Res.*, vol. 39, no. 22, pp. 3796–3816, 1999.

[64] D. H. Baker, S. A. Wallis, M. A. Georgeson, and T. S. Meese, "Nonlinearities in the binocular combination of luminance and contrast," *Vis. Res.*, vol. 56, pp. 1–9, Mar. 2012.

[65] J. Ding and G. Sperling, "A gain-control theory of binocular combination," *Proc. Nat. Acad. Sci. United States Amer.*, vol. 103, no. 4, pp. 1141–1146, 2006.

[66] I. Ohzawa, G. DeAngelis, and R. Freeman, "Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors," *Science*, vol. 249, no. 4972, pp. 1037–1041, 1990.

[67] H. Kumano, S. Tanabe, and I. Fujita, "Spatial frequency integration for binocular correspondence in macaque area V4," *J. Neurophysiol.*, vol. 99, no. 1, pp. 402–408, 2008.

[68] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 525–536, Jul. 1974.

[69] *Report on Experimental Framework for 3D Video Coding*, ISO/IEC Standard JTC1/SC29/WG11 MPEG2010/N11631, Oct. 2010.

[70] *Common Test Conditions for HEVC- and AVC-Based 3DV*, ISO/IEC Standard JTC1/SC29/WG11 MPEG2011/N12352, Dec. 2011.

[71] *Subjective Methods for the Assessment of Stereoscopic 3DTV Systems*, document ITU-R BT.2021, May 2012.

[72] A. B. Watson and A. J. Ahumada, "A standard model for foveal detection of spatial contrast," *J. Vis.*, vol. 5, no. 9, pp. 717–740, 2005.

[73] S. Scher, J. Liu, R. Vaish, P. Gunawardane, and J. Davis, "3D+2DTV: 3D displays with no ghosting for viewers without glasses," *ACM Trans. Graph.*, vol. 32, no. 3, p. 21, Jun. 2013.

Yu-Hsun Lin (S'12) is currently pursuing the Ph.D. degree with the Graduate Institute of Networking and Multimedia, National Taiwan University.

Ja-Ling Wu (SM'98–F'08) has been a Professor with the Department of Computer Science and Information Engineering, National Taiwan University (NTU), since 1996. From 2004 to 2007, he was the Head of the Graduate Institute of Networking and Multimedia, NTU. He was selected to be one of the lifetime Distinguished Professors of NTU, in 2006.