IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 3, 2018, accepted December 1, 2018, date of publication January 4, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2887276

A Survey of Stereoscopic 3D Just Noticeable Difference Models

YU FAN[1,2], (Member, IEEE), MOHAMED-CHAKER LARABI[1], (Senior Member, IEEE), FAOUZI ALAYA CHEIKH[2], (Senior Member, IEEE), AND CHRISTINE FERNANDEZ-MALOIGNE[1], (Member, IEEE)

[1]XLIM UMR CNRS 7252, University of Poitiers, 86000 Poitiers, France
[2]Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

Corresponding author: Mohamed-Chaker Larabi (chaker.larabi@ieee.org)

This work was supported by the Research Council of Norway through the Project HyPerCept under Grant 221073 and by the region Nouvelle Aquitaine through the Doctoral Allocation Program.

ABSTRACT Just noticeable difference (JND) for stereoscopic 3D content reflects the maximum tolerable distortion; it corresponds to the visibility threshold of the asymmetric distortions in the left and right contents. The 3D-JND models can be used to improve the efficiency of the 3D compression or the 3D quality assessment. Compared to 2D-JND models, the 3D-JND models appeared recently and the related literature is rather limited. In this paper, we give a deep and comprehensive study of the pixel-based 3D-JND models. To our best knowledge, this is the first review on 3D-JND models. Each model is briefly described by giving its rationale and main components in addition to providing exhaustive information about the targeted application, the pros, and cons. Moreover, we present the characteristics of the human visual system presented in these models. In addition, we analyze and compare the 3D-JND models thoroughly using qualitative and quantitative performance evaluation based on Middlebury stereo datasets. Besides, we measure the JND thresholds of the asymmetric distortion based on psychophysical experiments and compare these experimental results to the estimates from the 3D-JND models in order to evaluate the accuracy of each model.

INDEX TERMS Human visual system, just noticeable difference (JND), 3D compression, 3D-JND models, 3D quality assessment.

I. INTRODUCTION

The digital era has allowed simplifying the spread of Stereoscopic 3D (S3D) technologies in different application domains (*e.g.,* 3D-Cinema, 3D-TV) in recent decades. The most important aspect is that S3D can provide viewers with favorable immersion and natural sensation thanks to both binocular and monocular depth cues. However, there is a noticeable decrease in the attractiveness of S3D technology during the last few years. This is due to the complexity of such a content and the undesirable effect that it may generate from a perceptual point of view. S3D brought many technical challenges in the field of image and video processing linked to quality assessment, enhancement, and compression. Specifically, the main challenges are evaluating and optimizing the S3D imaging system with respect to storage capacity and quality of the user's experience (QoE).

To do so, it is important to understand and explore the different perceptual processes of the human visual system (HVS). For decades, the scientific community has exhaustively studied two-dimensional (2D) perception. Several properties and models of the HVS have been successfully exploited like the just noticeable difference (JND) models [1], [2]. These models refer to thresholds, depending on luminance, contrast, and temporal/spatial frequency of the local regions in the image, beyond which a distortion is visible. In other words, a given distortion cannot be perceived by the HVS if it is lower than the JND threshold. Therefore, JND models have been widely applied in visual signal processing, especially in compression and image processing [3], [4].

Over the last decades, numerous 2D-JND models have been developed either in transform domain [5]–[9], or in pixel domain [10]–[16]. Comprehensive reviews on 2D-JND models have been recently done in [17] and [18]. 2D-JND models are generally proposed based on specific characteristics of monocular vision, which does not fit with the complexity of 3D perception requiring specific

models accounting for both monocular and binocular depth cues.

To date, only a few 3D-JND models have been proposed because of the complex processes to be modeled [19]–[25]. Additional research efforts are undoubtedly needed to achieve a more accurate and efficient modeling that can effectively improve the performance of S3D applications (*e.g.,* compression, quality assessment, watermarking...). To the best of our knowledge, no review exists for the comparison of 3D-JND models in the framework of image quality assessment (IQA) [26].

In this paper, we propose a comprehensive survey of 3D-JND models. Since most of the existing 3D-JND models are computed in the pixel domain, we focus this survey on pixel-based 3D-JND models. Each model is briefly described by giving its rationale and main components in addition to providing exhaustive information about the targeted applications, the pros, and cons. The paper also provides a brief review of visual masking effects considered in these models. Furthermore, we present a thorough comparative analysis between the 3D-JND models using qualitative and quantitative performance evaluation. This study aims at comparing the distortion masking ability of the 3D-JND models using the widely used Middlebury stereo datasets [27]–[29], and evaluating the accuracy of these models using psychophysical experiments.

In summary, the major contributions of this paper include:

- An exhaustive review of the 3D-JND models;
- Creation of a dataset composed of asymmetrically distorted S3D images using 2D texture images from ETHZ Synthesizability Dataset [30];
- An extensive experimental comparison with qualitative and quantitative performance evaluation of the 3D-JND models.

The remainder of the paper is organized as follows. In Section II, we describe the main visual characteristics largely employed by 3D-JND models. Section III reviews the existing 3D-JND models separately. In Section IV, the models are thoroughly analyzed and compared. Section V presents the experimental results on the performance comparison of the 3D-JND models in terms of masking ability and accuracy using two different datasets. This paper ends with some conclusions and discussion of open issues in Section VI.

II. VISUAL CHARACTERISTICS FOR 3D-JND MODELS

Over the last decade, HVS has been studied based on physiological and psychophysical experiments [31]. HVS models are widely used in image/video processing [32], since such models can simplify and mimic the behaviors of the so complex HVS system. For instance, 3D-JND models, aiming to determine whether the distortion is undetectable by the HVS in a given block, can be used to improve the coding efficiency (CE) for S3D image/video. Therefore, understanding and studying the HVS mechanisms of the HVS are critical for developing a more reliable 3D-JND model. In general, 3D-JND models from the literature account for the HVS sensitivity and VM effects.

In this section, we explain the factors affecting the HVS sensitivity related to S3D content. Most 3D-JND models are developed by combining some of the factors including spatial contrast sensitivity, luminance adaptation, contrast masking, binocular masking, temporal masking, and depth masking.

A. SPATIAL AND TEMPORAL CONTRAST SENSITIVITY

The luminance contrast sensitivity (CS) of the HVS describes the ability to perceive the various frequencies of stimuli with different luminance contrasts [33], [34]. This sensitivity for a given target can be determined by measuring the minimum contrast necessary for an observer to detect the target. Accordingly, the CS depends on the spatial frequency of the visual stimuli [35]. Several psychophysical experiments measured the CS by determining the minimum contrast to make a sine-grating of a given spatial frequency visible in an image [36], [37]. The relationship between the CS and the spatial frequency of the grating in the image is typically modeled by the achromatic contrast sensitivity function (CSF) having a band-pass behavior [38], [39]. In addition to achromatic CSF models, chromatic CSF models having a low-pass behavior were proposed in [40] and [41]. Moreover, some spatial-temporal CSF models have taken the temporal CS into account [42], [43]. Recently, Rousson *et al.* [44], [45] proposed a CSF for observing stereoscopic content on S3D display. Moreover, some 2D-JND models were developed using CSF [6]–[8].

B. MASKING EFFECTS

The JND thresholds for S3D content depend not only on the spatial-temporal CS but also on visual masking effects (MEs). The latter characteristics are often used in 3D-JND modeling. The visual masking (VM), a perceptual phenomenon, describes the visibility reduction (masking effect) of one stimulus (*e.g.,* the target) to human eyes in the presence of another (the masker) where these stimuli are coincident in space and simultaneous in time [46]. For 2D content, the masking effect (ME) is modeled by using spatial frequency [47], orientation [48], motion (commonly in video) [49] of both image signals. For S3D, the disparity/depth should be considered in VM [50].

1) LUMINANCE ADAPTATION

According to [51], the HVS has the ability to quickly adjust to the levels of background light in order to distinguish objects. This ability is known as luminance adaptation (LA). It is related to background luminance masking (LM). As described in [5], human eyes are more sensitive to luminance variation/contrast than absolute luminance intensity. In other words, the salience of an object in an image could be more influenced by the difference between its luminance and the luminance of its adjacent background than by its own absolute luminance. LA allows adjusting the sensitivity of the HVS in response to the relative luminance variations. LA can

be measured in an increment threshold experiment [51] that describes the just-noticeable luminance difference of a stimulus as a function of the background luminance intensity. The experimental results showed that the ratio between the just-noticeable luminance difference and background luminance, known as Weber's fraction [6], [52], is approximatively constant for a wide range of luminance intensities. The luminance contrast LC_w can be defined as:

$$LC_w = \frac{\Delta L}{L_{bg}}, \quad \Delta L = \left| L - L_{bg} \right|, \tag{1}$$

where L is the luminance of a test stimulus, and L_{bg} is the surrounding background luminance. For the scene with L_{bg} of high levels, LC_w remains nearly constant as L_{bg} increases. LC_w is considered in this case as Weber fraction. On the other hand, LC_w increases when L_{bg} decreases in the case of low background luminance. This describes a high visibility threshold of luminance contrast in dark regions of the scene. Chou and Li [5] and Yang *et al.* [10] estimated LA (*i.e.,* visibility threshold of LM) of an image pixel in the pixel domain as follows:

$$LA_{CY}(i,j) = \begin{cases} c_1 \times (1 - \sqrt{\frac{\overline{L_{bg}}(i,j)}{127}}) + c_3, & if\ \overline{L_{bg}}(i,j) \le 127 \\ c_2 \times (\overline{L_{bg}}(i,j) - 127) + c_3, & otherwise, \end{cases} \tag{2}$$

where c_1, c_2, and c_3 are constants, and are set to 17, 3/128, and 3 respectively for a viewing distance of six times the targeted image height [5], [10]. It should be noted that the value of c_1 is proportional to the viewing distance. $\overline{L_{bg}}(i,j)$ is the average background luminance at pixel of coordinate (i,j), and is computed by:

$$\overline{L_{bg}}(i,j) = \frac{1}{32} \sum_{x=1}^{5} \sum_{y=1}^{5} I(i-3+x, j-3+y) \times B(x,y), \tag{3}$$

where $I(i,j)$ is the luminance intensity at pixel (i,j), and the kernel of low-pass filter B is represented as:

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 0 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{4}$$

In addition to the method described above, Zhao *et al.* [20] measured the visibility threshold of LA based on psychophysical experiments. It was conducted using binocular patterns (corresponding to S3D images) that are asymmetrically distorted by noise, as shown in Fig. 1(a). During reading, the para-fovea could process the information within 5° of visual angle of its fixation point, while the fovea processes the information located within 2° around the fixation point [53], [54]. The fovea and para-fovea in human eye contribute jointly to the perception of a fixated region and its

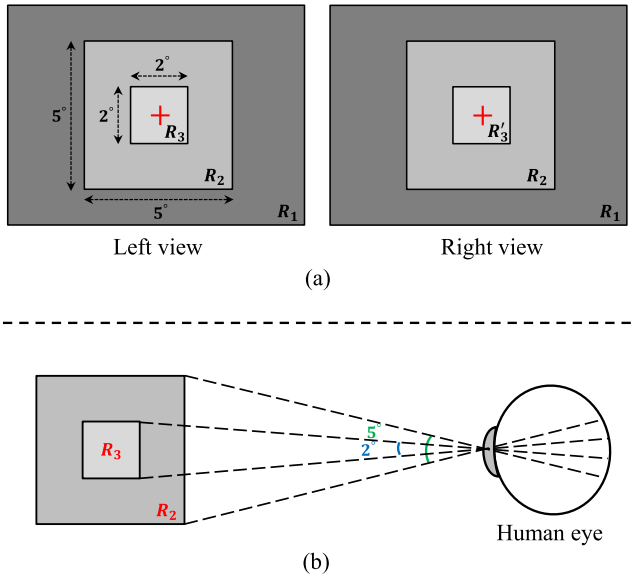


FIGURE 1. (a) Binocular patterns used for modeling luminance adaptation (LA), (b) Schematic illustration of regions covered by the fovea and the para-fovea. Note that R_1, R_2, and R_3/R'_3 correspond to three regions of the retinal image: the peri-fovea, the para-fovea covered by a square with 5° × 5° of visual angle, and the fovea covered by 2° × 2° square. In this case, the level of average background luminance corresponds to that of background luminance since the squares are uniform. The luminance intensity of R_1 is set to 112. The luminance in R_2 represents the background luminance L_{bg}. The luminance levels of R_3 and R'_3 are different, and are represented as $L_{bg} \pm n_l$ and $L_{bg} \pm n_r$, respectively, where n_l and n_r denote the amplitude of the bipolar patterns noise injected in the left and right views respectively.

surrounding regions perceived in the range of para-fovea. As shown in Fig. 1(b), the visual stimulus in an image is modeled as 2° × 2° square (called R_3) corresponding the fovea, and 5° × 5° square (called R_2) covering the para-fovea. The aim of the psychophysical experiment of Fig. 1(a) is to determine the amplitude of the maximum noise LA_{Z_r} (*i.e.,* visibility threshold of LA) injected in one view (*e.g.,* right view) without evoking binocularly perceptible difference due to LM, under a background luminance L_{bg_l} in this view (*e.g.,* left view), for a given noise with amplitude n_l injected in other view (*e.g.,* left view). More specifically, given the background luminance L_{bg_r} and noise amplitude n_l in the left view, an observer adjusted the noise amplitude in the right view n_r to binocularly detect the just noticeable noise. The amplitude of the aforementioned noise on a given pixel in the right image, namely $LA_{Z_r}(i,j)$, is computed by:

$$LA_{Z_r}(i,j) = A_{max}(L_{bg_l}(i,j+d)) \times \left[1 - (\frac{n_l(i,j+d)}{A_{max}(L_{bg_l}(i,j+d))})^{\gamma} \right]^{\frac{1}{\gamma}}, \tag{5}$$

where λ, set to 1.25, allows adjusting the noise influence in left view. d is the disparity value at pixel (i,j) corresponding to the horizontal shift of the pixel between right to left view. It should be noted that $L_{bg_r}(i,j)$ corresponds to the average background luminance of the pixel (i,j), which is determined by averaging the luminance intensity in the 5 × 5 window.

$A_{max}(L_{bg_l})$ denotes the visibility threshold of right view LA if there is no noise in the left view. $A_{max}(L_{bg})$ is calculated by the following formula:

$$A_{max}(i,j) = \begin{cases} a \times \left[L_{bg}(i,j)^2 - 96L_{bg}(i,j)\right] + 8, \\ \quad if\ 0 \le L_{bg}(i,j) < 48 \\ b \times \left[L_{bg}(i,j)^2 - 32L_{bg}(i,j)\right] + 1.7, \\ \quad if\ 48 \le L_{bg}(i,j) \le 255 \end{cases} \tag{6}$$

where $a = 2.7 \times 10^{-3}, b = 1.0 \times 10^{-4}$. $LA_r(i,j)$ becomes maximum, namely $A_{max}(i,j+d)$ when $n_l(i,j+d) = 0$.

2) CONTRAST MASKING

Contrast masking (CM) describes the VM effects in presence of two or more stimuli if these stimuli are of similar or same contrast/spatial non-uniformity (*e.g.,* spatial frequency, orientation) [55]. CM is also known as spatial masking. CM explains the fact that the presence of one stimulus reduces the ability of a subject to detect a targeted stimulus. For instance, HVS could tolerate more noises in textured regions than smooth regions since the spatial frequencies in noise and textured regions are similar.

According to previous studies, the visibility threshold of CM can be defined as a function of the average background luminance $\overline{L_{bg}}(i,j)$ and the amplitude of luminance edge (namely, edge height) $Eh(i,j)$, which refers to the contrast degree. For a viewing distance of six times of the targeted image height, Chou and Li [5] computed the visibility threshold related to contrast masking $CM_C(i,j)$ as follows:

$$CM_C(i,j) = 0.01\overline{L_{bg}}(i,j) \times [0.01G_m(i,j) - 1] + 0.115G_m(i,j) + c_4, \tag{7}$$

where c_4 adjusts the average amplitude of $CM_C(i,j)$, and is set to 0.5 in [5]. $G_m(i,j)$ denotes the maximum gradient at pixel (i,j) over four directions, and is computed as follows:

$$G_m(i,j) = \max_{s=1,2,3,4} \{grad_s(i,j)\}, \tag{8}$$

with

$$grad_s(i,j) = \frac{1}{16}\sum_{x=1}^{5}\sum_{y=1}^{5} I(i-3+x, j-3+y) \times g_s(x,y), \tag{9}$$

where $g_s(x,y)$ are kernels corresponding to four directional hight-pass filters. These four kernels are defined in equation 10 and 11:

$$g_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 8 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -3 & -8 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$
$$g_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 8 & 3 & 0 & 0 \\ 1 & 3 & 0 & -3 & -1 \\ 0 & 0 & -3 & -8 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \tag{10}$$

$$g_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 8 & 0 \\ -1 & -3 & 0 & 3 & 1 \\ 0 & -8 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix},$$
$$g_4 = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 8 & 0 & -8 & 1 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix} \tag{11}$$

Since HVS is more sensitive to the distortion around edge regions than that in textured regions, CM in edge and textured regions should be considered separately. Yang *et al.* [10] found that Chou and Li approach overestimates the visibility threshold of CM for edge regions. Thus, they used the Canny detector to decrease the thresholds for edge regions, and divided CM into texture masking (TxM) and edge masking (EM). Note that we focus only on the luminance component here. For a viewing distance of six times the targeted image height, Yang *et al.* [10] calculated the visibility threshold of $CM_Y(i,j)$ by:

$$CM_Y(i,j) = 0.117 \times W_{ed} \times G_m(i,j), \tag{12}$$

where $G_m(i,j)$ describes the maximal weighted average of gradients for the pixel (i,j). W_{ed} denotes the edge-related weight of the pixel (i,j), and is defined as:

$$W_{ed}(i,j) = Ed(i,j) * h_{lp}, \tag{13}$$

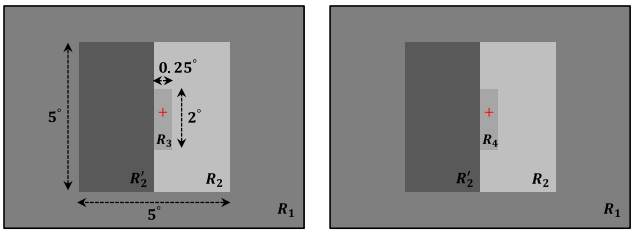
where Ed is the edge map estimated by Canny's detector [56] with a threshold of 0.5. The symbol $*$ represents the convolution operator, and h_{lp} is a $k \times k$ Gaussian low-pass filter having σ as a standard deviation. In [10], σ and k are set to 0.8 and 7, respectively.

Similarly, Liu *et al.* [12] employed the image decomposition method [57] to decompose the targeted image into structural and textural regions that lead to EM with Canny's detector and TxM, respectively. Therefore the visibility thresholds of CM due to edge and texture are described by:

$$CM_L(i,j) = 0.117 \times (w_e \cdot CM_e(i,j) + w_t \cdot CM_t(i,j)), \tag{14}$$

where $w_e = 1$ and $w_t = 3$ are the weights for edge masking (CM_e) and texture masking (CM_t) respectively. This means that the CM effect is stronger in textured regions than edge regions. CM estimation proposed by Chou and Li (see Eqs. 7, 8 and 9) was used to calculate CM_e and CM_t for both structural and textural images, respectively.

Zhao *et al.* [20] estimated the visibility threshold of CM using binocular patterns shown in Fig. 2 Similar to the psychophysical experiment used in LA, the subjects are asked to focus on R_3 and R_4, and adjust the luminance of the noise in the right view, n_r until the noise around the edges is binocularly just detected, L_{bg} and n_l being fixed. The just noticeable noise pair $\{n_l,\ n_l\}$ are then recorded. Zhao *et al.* [20] conducted several experiments to determine different noise pairs $\{n_l,\ n_l\}$ under different L_{bg} or n_l. Thus, the visibility



Left view Right view

FIGURE 2. Binocular patterns used in the experiment for modeling contrast masking. Note that R_1, R_2/R_2', and R_3/R_4 correspond to three regions of a human retinal image: the peri-fovea, the para-fovea covered by a square with 5° × 5° of visual angle, and the slice in fovea region with 2° height and 0.25° width. The luminance intensity of R_1 is set to 112. The background in patterns consists of 2 regions: R_2 with luminance L_{bg} and R_2' with luminance of $L_{bg} - Eh$, where Eh represents the edge height. The luminance levels of R_3 and R_4 are different, and equal to $L_{bg} \pm n_l$ and $L_{bg} \pm n_r$, respectively. n_l and n_r denote the amplitude of the bipolar patterns noise injected in the left and right views, respectively.

threshold due to CM of the right view, which depends on L_{bg_l} and the left image I_l, is expressed by:

$$CM_{Z_r}(i,j) = A_{max}(L_{bg_l}(i,j+d)) + F(L_{bg_l}(i,j+d)) \times Eh(I_l(i,j+d)), \quad (15)$$

where d is the disparity, A_{max} is estimated by Eq. 6, and I_l denotes the left image. F is a fitting function according to the average background luminance of one view, L_{bg_l}, is experimentally described as:

$$F(i,j) = -10^{-6} \times \left[0.7L_{bg}(i,j)^2 + 32L_{bg}(i,j)\right] + 0.07. \quad (16)$$

The edge height $Eh(i,j)$ of one pixel in image I is calculated using the following formula:

$$Eh(i,j) = \sqrt{E_h^2(i,j)) + E_v^2(i,j))}, \quad (17)$$

where

$$E_k(i,j) = \frac{1}{24}\sum_{h=1}^{5}\sum_{v=1}^{5} I(i-3+h, j-3+v) \times G_k(h,v), \quad k = h, v, \quad (18)$$

$$G_h = \begin{bmatrix} -1 & -2 & 0 & 2 & 1 \\ -2 & -3 & 0 & 3 & 2 \\ -3 & -5 & 0 & 5 & 3 \\ -2 & -3 & 0 & 3 & 2 \\ -1 & -2 & 0 & 2 & 1 \end{bmatrix},$$

$$G_v = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 3 & 5 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -3 & -5 & -3 & -2 \\ -1 & -2 & -3 & -2 & -1 \end{bmatrix} \quad (19)$$

3) BINOCULAR MASKING

Binocular masking (BM) describes the interocular interaction/masking in the case of two dissimilar stimuli presented to both eyes [58], [59]. The limited distortion in one view is influenced/masked by the other so that the two views can be successfully fused to a 3D image. This visual phenomenon is known as the binocular fusion (BF) [60]. The BM reveals that the HVS can tolerate a certain limited asymmetric distortion in one view that does not impair 3D perception. For instance, the subject perceives a stereo pair where the blur is introduced in the right image while the left image is kept unchanged. The fused 3D image is slightly blurred since the blur effect is reduced by the left image. Zhao *et al.* [20] modeled the BM using LA and CM as described previously. In addition, Qi *et al.* [23] conducted a psychophysical experiment similar to one of Fig. 1 in order to determine the visibility threshold of the right view $BM_r(i,j)$ relative to left one due to BM, which is approximately described as:

$$BM_r(i,j) = \begin{cases} 15 \times (1 - \sqrt{\frac{\overline{L_{bg_l}}(i,j)}{127}}) + 5.08, \\ \quad if\ \overline{L_{bg_l}}(i,j) \le 127 \\ 0.04 \times (\overline{L_{bg_l}}(i,j) - 127) + 5.08, \\ \quad otherwise, \end{cases} \quad (20)$$

where $\overline{L_{bg_l}}$ is the average background luminance that is calculated by using Eq. 3 and 4. The BM described above is similar to the LA shown in Eq. 2, but the visibility threshold of one view is calculated based on the luminance intensity of the other view.

4) TEMPORAL MASKING

The visual MEs mentioned above are dedicated to images, while the one discussed here focuses on video. Based on the free energy principle, HVS adaptively conceals the disorder tendency information in a continued movement scene, and tries to focus on the definite content of the input image [61]. This phenomenon can be modeled as the temporal masking (TM) caused by temporal discontinuities in intensity, such as motion when watching a video [62], [63]. Yang *et al.* [64] indicated that TM is proportional to motion. Inspired by Chou and Chen [65], Zhou *et al.* [22] estimated the visibility threshold of TM using the temporal JND (TJND) model described as follows:

$$TJND_Z(i,j,t) = \begin{cases} max\left\{\tau, \frac{H}{2} \times e^{\frac{-0.15}{2\pi} \times [\Delta(i,j,t)+255]} + \tau\right\}, \\ \quad if\ \Delta(i,j,t) \le 0 \\ max\left\{\tau, \frac{K}{2} \times e^{\frac{-0.15}{2\pi} \times [255-\Delta(i,j,t)]} + \tau\right\}, \\ \quad otherwise, \end{cases} \quad (21)$$

where

$$\Delta(i,j,t) = \frac{I(i,j,t) - I(i,j,t-1) + \overline{L_{bg}}(i,j,t) - \overline{L_{bg}}(i,j,t-1)}{2}, \quad (22)$$

$TJND_Z(i,j,t)$ is the TJND threshold of a pixel (i,j) of a given frame of multi-view plus depth video. $I(i,j,t)$ and $\overline{L_{bg}}(i,j,t)$ denote the luminance and the average background

luminance of the pixel (i, j) respectively. τ, H, and K are set to 8, 3.2, and 0.8, respectively. $\Delta(i, j, t)$ represents the luminance difference of the inter-frame. Larger $\Delta(i, j, t)$ values result in higher TM thresholds. $H > K$ reveals that the changes from high to low luminance can bring more TM than the changes from low to high luminance. Similarity, Qi *et al.* [23] estimated TM with the following formula:

$$TJND_Q(i, j, t) = max\left\{f_1(i, j, t), f_2(i, j, t)\right\}, \tag{23}$$

where

$$f_1(i, j, t) = max\left\{abs(CM_C(i, j, t) - CM_C(i, j, t-1)), \Delta\overline{CM_C}\right\}, \tag{24}$$

$$f_2(i, j, t) = max\left\{abs(LA(i, j, t) - LA(i, j, t-1)), \Delta\overline{LA}\right\}, \tag{25}$$

$CM_C(i, j, t)$ and $LA(i, j, t)$ are the visibility thresholds of CM and LA at pixel (i, j) in the frame t ($t \geq 2$), respectively. Eqs. 2 and 7 were used to calculate the $CM^C(i, j, t)$ and $LA(i, j, t)$ respectively. $\Delta\overline{CM_C}$ and $\Delta\overline{LA}$ denote respectively the average difference between two adjacent frames of all $\overline{CM_C}$ and $\overline{LA}$ of the whole video:

$$\Delta\overline{ME} = \frac{1}{N}\sum_{t=2}^{N}\left[ME(i, j, t) - ME(i, j, t-1)\right], \tag{26}$$

where ME represents LA or CM_C, and N is the number of frames. TM thresholds for the left and right views are calculated separately.

5) DEPTH MASKING

In addition to 2D VM effects, binocular depth masking (DM) have been studied by De Silva *et al.* [19], [66], who demonstrated that the subject cannot perceive sufficiently small depth changes on the scene. Moreover, the studies in indicated that the quality of the S3D video (with color plus depth representation) hardly changes with the compression of the depth map [67], [68]. In this circumstance, De Silva *et al.* [19], [66] first derived the visibility threshold relative to DM, which is known as the just noticeable difference in depth (JNDD). As described in [19], the JNDD threshold is mainly dependent on the viewing distance and the displayed depth level of the image. Based on the existing psychophysical models, a mathematical JNDD model for real-world viewing scenarios is defined as follows:

$$JNDD = 10^{[0.94\times\log_{10}(v)-2.25]} + K_w \times |dp|, \tag{27}$$

where K_w is the Weber constant and experimentally set to 0.03. dp is the simulated depth level with meter unit, while v denotes the distance between the subject's eyes and the fixation point *i.e.,* the screen.

As shown in Fig. 3(a), JNDD is estimated according to v and K_w. In fact, the JNDD in Eq. 25 can split in two parts: first, the visibility thresholds $JNDD_{d=0}$ when the simulated depth is equal to zero. Then the visibility thresholds $JNDD_{|d|>0}$ in

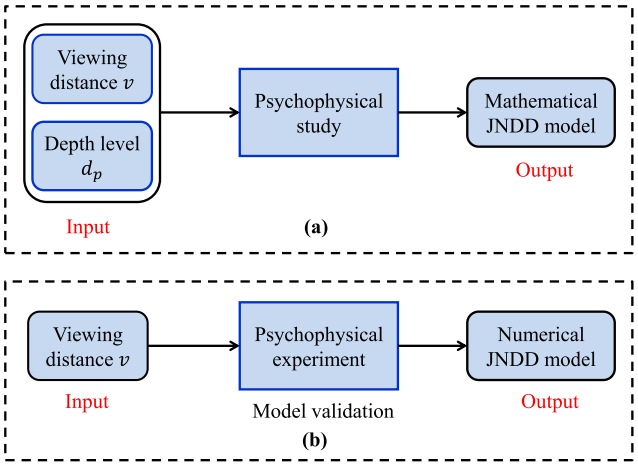


FIGURE 3. Frameworks of the JNDD model for: (a) real-world 3D perception, and (b) S3D display.

the case of nonzero disparity. $JNDD_{d=0}$ and $JNDD_{|d|>0}$ are described by :

$$JNDD_{d=0} = 10^{[0.94\times\log_{10}(v)-2.25]} \tag{28}$$

and

$$JNDD_{|d|>0} = 0.03 * |dp|. \tag{29}$$

According to Eqs. 25, 26, and 27, the JNDD curves corresponding to $JNDD$, $JNDD_{d=0}$, $JNDD_{|d|>0}$ are depicted in Fig. 4. One can notice that the linear summation between green and blue curves derive the red curve (JNDD thresholds).

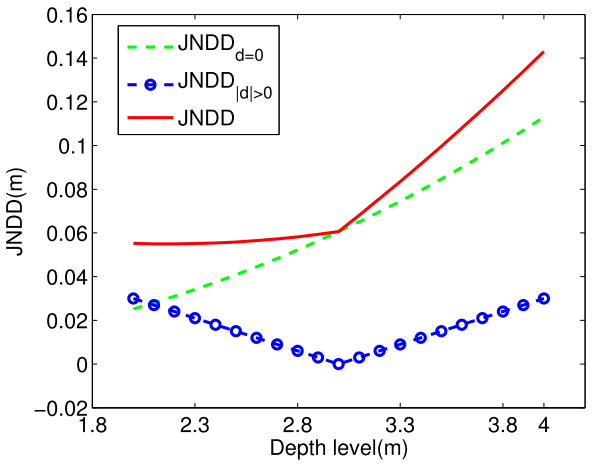


FIGURE 4. JNDD thresholds $JNDD$, $JNDD_{d=0}$ and $JNDD_{|d|>0}$ according to simulated distance for real-world 3D perception. The viewing distance is set to 3 m.

III. 3D-JND MODELS

In this section, we give a brief introduction of the existing 3D-JND models. Specifically, each model is described with its framework as well as its mathematical expression. It should be noted that all 3D-JND presented models measure

the achromatic JND thresholds. In other words, only JND thresholds of the luminance component of the color image are taken into account.

A. JNDD MODEL

A JNDD model is addressed firstly in [19], [66], and [68], which indicates that a human subject could not perceive depth changes below the JNDD threshold. The visibility thresholds due to DM (described in Section II-B.5) could not be applied for S3D displays. This is due to the fact that the viewing distance rarely changes when a subject watches a S3D image/video on a S3D display. Therefore, De Silva *et al.* [19], [66] ignored the viewing distance, and only considered the depth level in JNDD estimation (see Fig. 3(b)). They conducted a psychophysical experiment to validate the $JNDD_{|d|>0}$ (as shown in Fig. 4), and to measure the JNDD thresholds on a S3D display using 2D-plus-depth videos.

As described in [19] and [66], the simulated depth level is 8 bits, where 0 and 255 denote the farthest and the nearest positions apart from the subject, respectively. Objects on the display with a depth value of 128 have zero disparity. The plane with zero disparity, called zero parallax plane, is the co-planar with the display plane. In the psychophysical experiment, two identical (left and right) objects were first displayed at the same depth level, namely initial depth level, and then the depth level of one object is changed gradually. The subjects were asked to inform about depth changes between the two objects when perceived. Various initial depth levels of the two objects have been investigated and the final threshold is obtained by averaging the JNDD values of all subjects. By analyzing the JNDD values according to different initial depth levels, the JNDD threshold $JNDD_{num}$, for a given initial depth value dp_i, is modeled as follows:

$$JNDD_{num}(i,j) = \begin{cases} 21, & if\ 0 \le dp_i(i,j) < 64 \\ 19, & if\ 64 \le dp_i(i,j) < 128 \\ 18, & if\ 128 \le dp_i(i,j) < 192 \\ 20, & if\ 192 \le dp_i(i,j) < 225 \end{cases} \tag{30}$$

where $dp_i(i,j)$ is the depth value (in pixels) of the original depth map at the pixel coordinate (i,j). The JNDD thresholds in Eq. 28 correspond to the symmetrical shape of the $JNDD_{|d|>0}$ in Fig. 4 except the zero disparity level (128). Moreover, according to the experimental results, the expert's subjects are more sensitive to depth changes than the non-expert ones.

B. BJND MODEL

Meanwhile, another 3D-JND model, namely binocular JND (BJND) was proposed by Zhao *et al.* [20]. It reveals the threshold in inter-difference between the left and right views that human can recognize. The BJND model investigates the properties of the binocular vision in response to asymmetric noise in a stereo pair based on the VM effects of the HVS. These considered in this model consists of LA (see Section II-B.1) and CM (see Section II-B.2). Eqs. 5 and 15 are used to calculate the visibility thresholds related to LA (LA_{Z_r}) and CM (CM_{Z_r}), respectively. Fig. 5 illustrates the framework of calculating the BJND thresholds of one view of the stereopair. It is worth noting that there are two (left and right) BJND thresholds for each stereopair, since the BJND of one view indicates the maximum distortions that can be introduced in this view without evoking binocularly visible differences, given the distortions in the corresponding pixels of the other view. Like this, BJND of the left or right view $BJND_{l|r}$ is defined by:

$$\begin{aligned} &BJND_{l|r}(i,j) \\ &= BJND_{l|r}(L_{bg_{r|l}}(i,j-d_{l|r}), Eh_{r|l}(i,j-d_{l|r}), n_{r|l}(i,j-d_{l|r})) \\ &= CM_Z(L_{bg_{r|l}}(i,j-d_{l|r}), Eh_{r|l}(i,j-d_{l|r})) \\ &\quad \times \left[1 - \left(\frac{n_{r|l}(i,j-d_{l|r})}{CM_Z(L_{bg_{r|l}}(i,j-d_{l|r}), Eh_{r|l}(i,j-d_{l|r}))}\right)^{\gamma}\right]^{\frac{1}{\gamma}} \end{aligned} \tag{31}$$

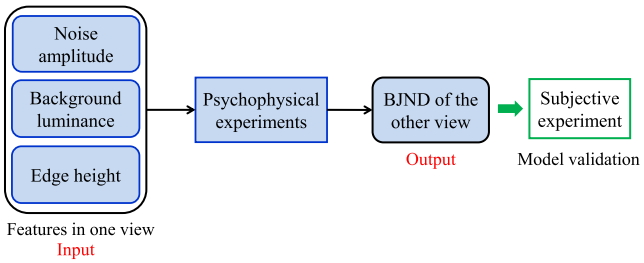


FIGURE 5. Framework for calculating the BJND of a single view of a stereo pair.

where $l \mid r$ represents left or right, and d is the horizontal disparity value at pixel (i,j). The disparity values of the left view (d_l) are positive while those of the right view (d_r) are negative. $L_{bg}(i,j)$ indicates the average background luminance at pixel (i,j) that is estimated by averaging the luminance intensity in the 5×5 surrounding region. $Eh(i,j)$ refers to the edge height that is estimated using Eqs. 17, 18, and 19. CM_Z denotes the visibility thresholds of the CM computed by Eq. 15. $n(i,j)$ is the luminance difference between the original and distorted images at pixel (i,j), *e.g.*, (noise amplitude). Note that $0 \le n_{r|l} \le CM_{Z_{r|l}}$, and BJND of one view $BJND_{l|r}$ can be reduced to $CM_{Z_{r|l}}$ if there is no noise in the other view. The BJND model was validated by means of subjective experiments [20]. The experimental results showed that human perceives the noise when viewing the stereo images if and only if this noise in one view is higher than the BJND value.

C. JJND MODEL

In addition to LA and CM, the binocular depth cue is proposed to be considered for the design of this 3D-JND model. Since monocular and binocular cells in V1 area have different receptive fields [69], it is reasonable to calculate the JND thresholds for monocular and binocular regions

separately. The monocular region in one view refers to 1) the pixels not having corresponding pixels in the other view due to the occlusion effect; or/and 2) disparity-shifted pixels within image borders. Thus the monocular region is known as an occluded region (OR) or non-corresponding region (NCR), and the OR/NCR is only seen by one eye. In contrast, the binocular region in one view is called non-occluded region (NOR) or corresponding region (CR), and the NOR/CR can be perceived by both eyes correspondingly.

Accordingly, Li *et al.* [21] estimated the JND thresholds of both OR and NOR, and thus proposed the joint JND (JJND) model based on the idea that a human subject has different perceptions of objects with different depths. Unlike the JNDD and BJND models, the JJND model was developed with a 2D-JND model, namely non-linear additively masking model (NAMM) [10], which accounts for LA and CM. As shown in Fig. 6, the JND thresholds of one image (*e.g.,* left image) are calculated using NAMM. This JND threshold $JND_{Y_l}(i,j)$ of a pixel (i,j) in the left image is defined by:

$$JND_{Y_l}(i,j) = LA_{CY}(i,j) + CM_Y(i,j) - C \times min\{LA_{CY}(i,j), CM_Y(i,j)\}, \quad (32)$$

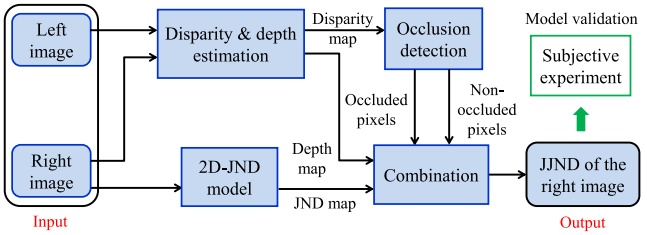


FIGURE 6. Framework for calculating the JJND of the right view of a stereo pair.

where C is a constant used to adjust the inter-effect between $LA_{CY}(i,j)$ and $CM_Y(i,j)$. The latter are calculated using Eqs. 2 and 12, respectively. C is within the [0, 1] range, and set to 0.3 in [10].

For the other image (*e.g.,* right image), disparity estimation is firstly performed in order to classify the image pixels into two classes: occluded and non-occluded pixels [70], [71]. The OR, often appearing at the objects' edges or the image borders, represents very strong monocular clues and any distortion in this region is easy to be noticed compared to NOR. Besides, the depth map is derived according to the disparity map and viewing distance. Based on the aforementioned classification, the JJND of the right view is proportional to its 2D-JND thresholds estimated by NAMM, where the coefficients are defined as 1) a fixed value α_{OR} for occluded pixels; and 2) depth-dependent value β_{NOR} for non-occluded pixels (see [21]). The JJND of the right image is formalized as follows:

$$JJND_r(i,j) = \begin{cases} \alpha_{or} \times JND(i,j), & if\ (i,j) \in OR \\ \beta_{dp} \times JND(i,j), & otherwise, \end{cases} \quad (33)$$

where $JND(i,j)$ is the visibility threshold for right image at pixel (i,j). $\alpha_{or} = 0.8$ is used to limit the JND thresholds for OR. The effectiveness of the JJND model was demonstrated using subjective quality evaluations. Specifically, the qualities of the noise-injected S3D images are compared between using JJND and 2D-JND [10]. The experimental results showed that the S3D images receiving JJND noise tolerate more noise than with 2D-JND, in the case of nearly same perceptual quality.

D. MJND MODEL

Even though the JJND accounts for binocular depth cues, the reliability of this model can be reduced for the stereopairs with low average depth value or uniform depth map. To avoid this constraint, Zhou *et al.* [22] designed a JND in the Multi-view case (MJND) by combining spatial JND (SPJND), TJND and depth JND (DPJND). As shown in Fig. 7, the MJND model is defined as:

$$MJND(i,j,t) = [SPJND(i,j,t)]^{w1} \times [TJND_Z(i,j,t)]^{w2} \times [DPJND(i,j,t)]^{w3}, \quad (34)$$

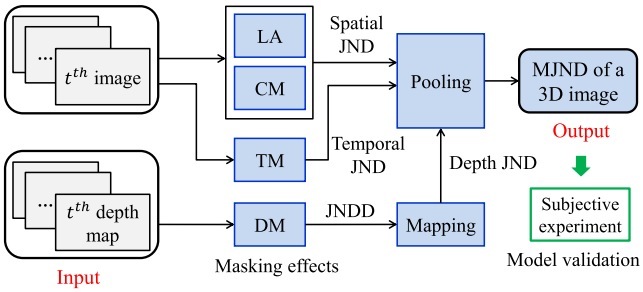


FIGURE 7. Framework for calculating the MJND of one image from multi-view videos.

where $MJND(i,j,t)$ is the JND threshold at pixel (i,j) at the t^{th} 3D frame (image plus depth map). w_1, w_2 and w_3, are used to control the contribution of SPJND, TJND and DPJND respectively, are set to 1. *SPJND* denotes the JND thresholds for both LA and CM, and is calculated using a 2D-JND model [5] defined as follows:

$$SPJND(i,j,t) = max\{LA_{CY}(i,j,t), CM_C(i,j,t)\}, \quad (35)$$

where $LA_{CY}(i,j,t)$ and $CM_C(i,j,t)$ are estimated based on Eqs. 2 and 7. In Eq. 2, c_1, c_2, and c_3 are set to 14, 3/128, and 2, respectively. c_4 in Eq. 7 is set to 1/4. In addition, $TJND_Z(i,j,t)$ is determined using Eq. 21. Zhou *et al.* [22] estimated the DPJND thresholds based on the JNDD model proposed in [66]. Thus, the $DPJND(i,j,t)$ is defined by:

$$DPJND(i,j,t) = 1 + \frac{JNDD_{num}(i,j,t)}{256}, \quad (36)$$

where $JNDD_{num}$ denotes the numerical JND thresholds computed by Eq. 30. The performance of MJND was validated based on subjective experiments. Compared to using the spatial-temporal JND (STJND) [22] or the foveated JND (FJND) [13], the noise-injected 3D video distorted using

MJND can tolerate much more noise for the same perceptual quality. Furthermore, the multi-view coding (MVC) [72] using the MJND model achieves better perceptual quality than using the joint multi-view model [73] for the same bit rate.

Inspired by the MJND, Liu *et al.* [74] proposed a new multi-view JNDD (MJNDD) model used to improve the joint multi-view video coding (JMVC). The MJNDD model combines STJND with an adapted JNDD model, which segment the texture frame into background regions (BR) and foreground regions (FR). Recently, Shi *et al.* [75] developed a new 3D-JND model, which considers the depth information and visual saliency in addition to LA, CM, and TM.

E. SJND MODEL

Qi *et al.* [23], [76] developed the stereo JND (SJND) model for 3D video with the stereo interleaving format [77] (*i.e.,* left and right frames). The SJND model takes into account both intra-view and inter-view MEs in addition to LA and CM. The intra-view masking includes BM, whereas inter-view masking refers to TM.

As shown in Fig. 8, for one of the left and right frames, the visibility thresholds for intra-view ME (namely $TJND_L/TJND_R$) are determined by integrating LA, CM, and TM. $TJND_L/TJND_R$ is calculated according to Eq. 23. For a pair of stereoscopic frames, the stereo TJND ($TJND_s$) is computed as follows:

$$TJND_s(i,j,t) = \frac{3}{8} \times [TJND_L(i,j,t)] + \frac{5}{8} \times [TJND_R(i,j,t)], \tag{37}$$

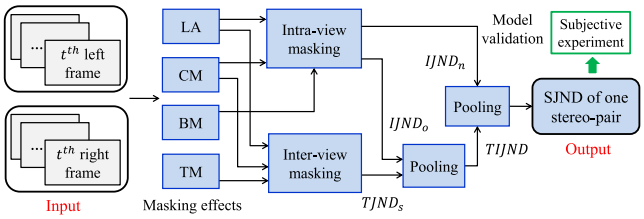


FIGURE 8. Framework for calculating the SJND of a stereo pair.

The weights for left and right views are used to determine the asymmetry between views [78]. Besides, the views are decomposed into NOR and OR involved respectively in the binocular fusion (BF) [60] and the binocular rivalry (BR) [79]. The human brain can fuse the left and right views into a single mental image when the stimuli in both views are similar. However, if the stimuli are sufficiently different, our brain fails to merge both views resulting in BR phenomena. To model the BM, different intra-view JND (IJND) thresholds are computed based on left and right views according to OR and NOR. The occluded pixels appear on the edge of foreground objects. Therefore, The IJND for non-occluded pixels $IJND_o$ only accounts for CM, and is defined as:

$$IJND_o(i,j,t) = r(t) \times CM_l(i,j,t) + [1 - r(t)] \times CM_r(i,j,t), \tag{38}$$

where $r(t)$ is a random value in the range [0, 1]. Since OR is detected for a random moment, $r(t)$ varies according to time. $CM_{l/r}$ thresholds are calculated based on Eq. 7. $IJND_o$ is not based on experiments measuring BM effect, thus $IJND_o$ and $TJND_s$ above should be combined to consider the VM effects of both inter-frame and intra-frame. Accordingly, the new model called *TIJND* is described as follows:

$$ITJND(i,j,t) = w_t \times TJND(i,j,t) + w_b \times IJND_o(i,j,t), \tag{39}$$

where w_t and w_b are the weights used to balance the importance of inter-frame and intra-frame JNDs, respectively. Since BM appears less on OR than NOR, BM effect should be considered less than TM for NOR. Thus w_t and w_b are set to 0.9 and 0.1, respectively. In contrast, LM, CM, and BM are taken into account for NOR. The visibility threshold of the intra-view masking for a non-occluded pixel, namely $IJND_n$ is represented by:

$$IJND_n(i,j,t) = max\{f_1(i,j,t), f_2(i,j,t), BM(i,j,t)\}, \quad n \in l, r, \tag{40}$$

where f_1 and f_2 are calculated based on Eqs. 25 and 25, respectively. $BM(i,j,t)$ refers to the luminance visibility of one view relative to the other view in the t^{th} frame of the video. By using Eq. 20, $BM(i,j,t)$ can be determined. $IJND_n$ of a stereo pair is computed by averaging the $IJND_n$ values for left and right views ($IJND_l$ and $IJND_r$). By integrating $IJND_n$ with *ITJND*, the SJND threshold of a stereo pair is defined as:

$$SJND(i,j,t) = [TIJND(i,j,t)]^{\mu} \times [IJND_n(i,j,t)]^{(1-\mu)}, \tag{41}$$

where μ manages the tradeoff between *TIJND* and $IJND_n$, and is set to 0.5 in [23] or 0.6 in [76]. The effectiveness of SJND for stereoscopic video quality assessment (SVQA) was demonstrated thanks to subjective experiments.

F. HJND MODEL

It has been demonstrated that depth perception is influenced not only by depth intensity (DI) but also by depth contrast (DC). In light of this, Zhong *et al.* [80] first proposed a 3D image JND model combining 2D-JND with depth saliency taking DI and DC into account. Moreover, the serious geometric distortion (GD) in synthesized views attracts visual attention leading to smaller JND thresholds. Therefore, based on their previous work and a 2D-JND model [10], Zhong *et al.* [24] recently developed a hybrid JND (HJND) model, which considers GD in addition to DI and DC.

HVS is more sensitive to closer objects than deeper ones, and the regions with inconsecutive depth or higher DC attract more attention. Thus, depth saliency is influenced by DI, depth intensity contrast and depth orientation contrast. Based on these considerations, a depth saliency model [81] was used to quantify the combined action of DI and DC for 3D video. For the n^{th} and $(n+1)^{th}$ views, the disparity map can be estimated by the stereo matching algorithm. In order to obtain

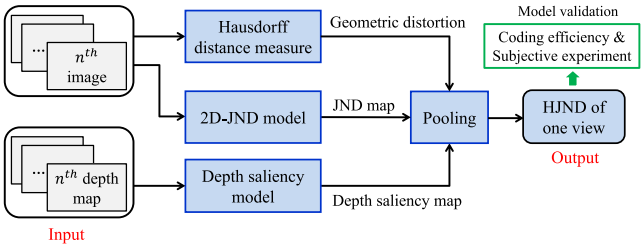


FIGURE 9. Framework for calculating HJND for a single view.

the depth map, the first step is to translate the disparity value *disp* into depth value *dp* by:

$$dp(i,j) = \frac{b \times f}{disp(i,j)}, \quad disp(i,j) \neq 0, \tag{42}$$

where b and f denote the baseline distance between the adjacent cameras and focal length of the camera, respectively. $disp(i,j)$ is the disparity of the pixel at coordinate (i,j). The intersection of two adjacent cameras creates a zero disparity plane, and this zero disparity corresponds to the 3D display. Deep objects refer to positive disparity, while the pop-out objects have a negative disparity. Next, the depth value is quantized as an 8 bits value, where 0 means the farthest object and 255 denotes the nearest one. Nearer objects are obviously the most salient to the observers. Thereby the depth value $dp(i,j)$ is mapped to the range [0, 255] through the non-linear quantization, defined as follows:

$$dp_m(i,j) = \left\lfloor 255 \times \frac{dp_{min}}{dp(i,j)} \times \frac{dp_{max} - dp(i,j)}{dp_{max} - dp_{min}} + 0.5 \right\rfloor, \tag{43}$$

where $\lfloor v \rfloor$ denotes the integer less than or equal to v. dp_{max} and dp_{min} represent the maximum and minimum values of depth, respectively. dp_m is the depth map used to determine depth saliency map S_d. The detail of the S_d estimation is described in [24]. The GD in synthetic views, created by the depth image-based rendering (DIBR) technique [82], [83], is related to the quality of the distorted depth map, and measured by the Hausdorff distance [84]. The latter calculates the geometric distance between the surfaces of the synthesized view and that of the original one. Besides, Yang *et al.* [10] proposed a 2D-JND model expressed by Eq. 32. Combining the depth saliency map S_d, GD image G, and the 2D-JND map JND_Y, the HJND threshold of one view can be calculated as follows:

$$HJND(i,j) = \varepsilon \times JND_Y(i,j) \times \omega^{N(S_d(i,j) \cdot G(i,j))}, \tag{44}$$

where the parameters ε and ω are empirically set to 1.4 and 0.15, respectively. The symbol $N(\cdot)$ represents a unity-based normalization function that brings all values into the range [0, 1]. To validate the effectiveness of the HJND model, it was integrated into the MVC encoding framework to remove the perceptual redundancy. Compared to the standard JMVC scheme and the joint multi-view video plus depth scheme using JJND, the JMVC using HJND can save more bit-budget while providing a better perceptual quality.

G. DJND MODEL

As described in the HJND model, the HVS is more sensitive to nearby objects than far away objects in the scene. In the real world, the focused areas have higher resolution on the retina while the other areas are blurred by the HVS [85], namely depth of focus (DOF) blur effect [86]. However, conventional 3D displays cannot reproduce the DOF blur effect. In this case, the viewer focuses on the whole scene, which does not correspond to human depth perception. Moreover, this behavior may result in visual fatigue. The described above 3D-JND models have not considered the DOF blur effect. Since FR are more sensitive by HVS than BR, the JND thresholds of FR and BR should be calculated differently. Thereby, Xue *et al.* [25] proposed a disparity-based JND (DJND) model by combining LA, CM with disparity information used to simulate the DOF blur effect.

Fig. 10 shows the framework for calculating the DJND of the left view of a stereo pair. First, the visibility thresholds of LA of the left view (LJND) is estimated according to Eq. 2. In order to distinguish thresholds for FR and BR, the LJND is filtered by a Gaussian low-pass filter simulating the DOF blur effect. The standard deviation of this filter $\sigma(i,j)$ is adaptively calculated based on average disparity values of the disparity image as follows:

$$\sigma(i,j) = \left[\psi + e^{-\alpha \times N(\overline{Disp(i,j)}) - \beta}\right]^2, \tag{45}$$

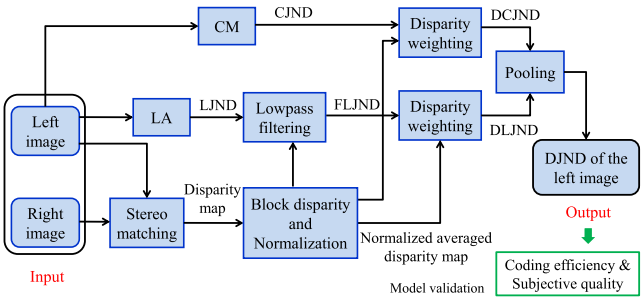


FIGURE 10. Framework for calculating the DJND of the left views.

with

$$\overline{Disp(i,j)} = \frac{1}{25} \sum_{x=-2}^{2} \sum_{y=-2}^{2} Disp(i+x, j+y), \tag{46}$$

where $N(\overline{Disp(i,j)})$ denotes the normalized average disparity value of a 5×5 block centered at pixel (i,j). The constants α and β are set to 10 and 0.6, respectively. ψ is a constant and defined as 0.117. Then, the filtered LJND is calculated by:

$$FLJND(i,j) = \frac{1}{G} \sum_{x=-2}^{2} \sum_{y=-2}^{2} \left\{ \left[e^{-\frac{x^2+y^2}{2 \times \varepsilon^2(i,j)}} \right] \times LJND(i+x, j+y) \right\}, \tag{47}$$

where G is a bidimensional Gaussian function. Next, the disparity information is used to weight the FLJND by the following negative exponential function:

$$DLJND(i,j) = e^{-2\times\overline{Disp(i,j)}} \times FLJND(i,j) + \delta, \quad (48)$$

where δ is a constant and set to 3. The region with a larger disparity (*e.g.,* FR) has lower DLJND thresholds than that with smaller disparity (*e.g.,* BR). Besides, the disparity-based CJND is estimated by:

$$DCJND(i,j) = e^{-2\times\overline{Disp(i,j)}} \times CJND(i,j), \quad (49)$$

$CJND(i,j)$ corresponds to CM_Y that is calculated based on Eq. 12. Finally, the DJND of the left view is obtained by combining DLJND with DCJND using NAMM as follows:

$$DJND(i,j) = DLJND(i,j) + DCJND(i,j) - \phi \times min\{DLJND(i,j), DCJND(i,j)\}, \quad (50)$$

where ϕ is used to adjust the overlapping effect of LA and CM, and set to 0.3. Similar to HJND model, the DJND is applied to MVC in order to evaluate its performance. The DJND model was validated based on two aspects: 1) the DJND-based MVC outperforms the conventional JMVC in terms of both subjective quality of FR and visual comfort in 3D videos, 2) compared with the conventional JMVC and JMVC using 2D-JND [64], the DJND-based MVC saves more coding bit-budget without the degradation of the perceived quality. This is because MVC using DJND preserves the details in the salient regions and it reduces the redundancies in the other regions.

In addition to the previously mentioned 3D-JND models, Zhang *et al.* [87] proposed a foveated stereoscopic JND model and then applied it to improve the 3D video CE. Moreover, Wang *et al.* [88] developed a 3D just noticeable distortion model for asymmetrical coding. Recently, Du *et al.* [89] studied the effect of texture complexity on the JND threshold for asymmetrically encoding S3D images based on subjective experiments.

IV. COMPARISON OF 3D-JND MODELS

In this section, we compare the previously described 3D-JND models by analyzing various aspects. The overall comparison between these models is given in Table 1. The summary of some important notations and abbreviations used in Table 1 is given in Table 2.

For each 3D-JND model, the Inputs, the MEs, the 3D content format and the process of model validation have been presented in the previous section. Thereby, we mainly compare in this section these models in terms of their complexity, pros, and cons, as well as applications.

A. COMPLEXITY

To compare the complexity between 3D-JND models, we evaluated not only the MEs and features considered in each model, but also the computational runtime of each model for S3D images. It is worth noting that JJND has been discarded from this evaluation because of its dependency on psychophysical conditions that cannot be controlled here. In order to calculate the runtime of the 3D-JND model, we employed four S3D images with LR images format from the open Middlebury stereo database [90]. This set contains "Teddy" with resolution 450 × 375 [91], "Art" and "Moebius" with resolution 463 × 370 [27], [28], and "Baby2" with resolution 1240×1110 [27], [28]. The ground-truth disparities in this database were used for this evaluation. JND thresholds were estimated using the right view for all 3D-JND models except SJND because of its definition. Besides, the TM effect has not been considered in MJND and SJND for S3D images As both MJND and SJND were designed for S3D videos, it is not fair to compare other 3D-JND models with MJND and SJND including TM effect in terms of computational runtime. The runtime (in second) per image for each 3D-JND model is shown in Table 3. Considering runtime, MEs and features (in Table 1), the complexity for each 3D-JND model is reported in Table 1 using stars. The greater the number of stars is, the higher its complexity is, and vice versa.

Note that the experiments are performed by using MATLAB code on a computer (Inter Core i7-2630 QM Processor at 2.00 GHz, 4GB RAM, Windows 7). As shown in Table 3, HJND consumes the longest time among all models due to the process of DIBR and estimation of GD per block. Even though DJND, JJND and MJND accounted for disparity/depth information, DJND is lower than two other models due to the consideration of DOF blur effect. In addition, MJND and SJND use more MEs than, they are faster than BJND. This is due to the fact that MJND and SJND were designed based on a conventional 2D-JND model, and BJND was developed integrating the noise amplitude with LA and CM.

B. PROS AND CONS

In this section, we assess the 3D-JND models in terms of their pros and cons. The JNDD has been designed thanks to psychophysical experiments on stereoscopic 3D displays. This model can be extended to various types of S3D displays [66], such as auto-stereoscopic display and passive stereoscopic display [92]. However, it can only measure the visibility threshold with limited depth levels, not satisfying the desired depth range for real application. For instance, JNDD is not suitable for estimating the tolerable depth difference in virtual view rendering [93], [94]. Furthermore, this model is only compatible with the 2D-plus-depth representation of 3D content, and its accuracy depends on the quality of the depth image. Hence, a depth image with poor quality may lead to inaccurate JND thresholds.

Compared to the JNDD, BJND is closer to human binocular perception. Moreover, it can use 2D/color-plus-depth and LR formats. However, this model was designed based on PE using binocular patterns with zero disparity. In other words, BJND ignored the effect of disparity of the visual stimuli on visibility thresholds, which makes it less suitable for real-life

TABLE 1. Comparison between the described 3D-JND models.

	JNDD	BJND	JJND	MJND	SJND	HJND	DJND
Inputs	VD, DpM	LCs NAM	LC, DsM, DpM	LCs, DpM, TI	LCs, DsM, TI	LCs, DpM	LC, DsM
VM & Features	DM	LA, CM	LA, CM, DI	LA, CM, TM, DM	LA, CM, TM, BM	LA, CM, DI, DC, GD	LA, CM, DI, DOF
3D format	2D + depth	LR images	LR images	MVD	LR frames	DIBR, MVD	LR images
Model validation	Theoretical results vs. results derived from PE	Noise detection probability in S3D images	Comparison with 2D-JND in terms of SQ	MJND-based MVC vs. JMVM-based MVC in terms of CE and PQ	A metric using SJND vs. SVQA metrics in terms of SQ	MVC with HJND vs. MVC with JJND vs. JMVC in terms of CE and SQ	MVC with DJND vs. MVC with 2D-JND vs. JMVC in terms of CE,PQ and VC
Complexity	_____	***	*	*	**	*****	**
Pros	Extension for various 3D displays	Suitable to several 3D formats	Binocular vision properties	Multiple MEs	Multiple MEs	Considering DC and GD	Several 3D formats, VC improvement
Cons	Limit 3D format, influence of depth image quality	Disparity effect ignoring, SMA impact on JND accuracy	Accuracy decrease for low or uniform disparities, lack of comparison with 3D-JND models	Accuracy decrease for large depth range	Difficult to design the PE for model validation, many parameters in the model	Highly depending on DIBR techniques, specially designed for MVD format	Accuracy decrease for 3D image with small depth difference between FR and BR
Application	Depth sensation enhancement, 3D QoE enhancement, 3D video coding, S3DW	Sharpness /contrast enhancement, 3D video coding, SIQA, S3DW, S3D IR	S3DW	MVC and 3D-HEVC	SVQA	MVC	MVC

stereoscopic images. To avoid this constraint, Kim *et al.* [95] conducted PE to measure the binocular visibility thresholds with different disparities under various amplitudes of the asymmetric noises and background luminance levels. However, they have not studied the impact of the disparity on JND estimation for CM. In addition, the disparity estimation error issue from stereo matching algorithm may decrease the reliability of the BJND estimation. Finally, BJND did not explore the visibility threshold for different types of asymmetric noises (*e.g.,* Gaussian/Poisson noise).

JJND model copes with the issue of disparity ignoring in BJND. This model mimics BF and BR by computing different JND thresholds for OR and NOR, separately. However, the performance of JJND can be reduced for a pair of S3D images having uniform disparity maps or/and low disparity *i.e.,* weak depth perception. Even though it was reported that JJND is more effective than 2D-JND models, the authors did not make any comparison with other 3D-JND models.

MJND and SJND are the most reliable among these 3D-JND models since they take into account both 2D and 3D MEs so that they completely model the stereoscopic HVS characteristics. Since depth values in MJND vary in a very small range, the accuracy of the model may be decreased for

TABLE 2. Important notations and abbreviations used in Table 1.

BM	binocular masking	LR	left and right
BR	background regions	MVC	multi-view video coding
CE	coding efficiency	MVD	multi-view video plus depth
CM	contrast masking	NAM	noise amplitude map
DC	depth contrast	PE	psychophysical experiment
DI	depth intensity	PQ	perceived quality
DIBR	depth image-based rendering	QA	quality assessment
DM	depth masking	SMA	stereo matching algorithm
DpM	depth map	SIQA	stereoscopic image QA
DsM	disparity map	SQ	subjective quality
FR	foreground regions	SVQA	stereoscopic video QA
GD	geometric distortion	S3DW	S3D watermarking
IR	image retargeting	TI	temporal information
JMVC	joint multi-view video coding	TM	temporal masking
JMVM	joint multi-view video model	VC	visual comfort
LA	luminance adaptation	VD	viewing distance
LC	luminance component	VM	visual masking

TABLE 3. Computational runtime (in second) of the described 3D-JND models.

	BJND	JJND	MJND	SJND	HJND	DJND
Teddy	1.41	0.15	0.04	0.30	9.68	0.60
Art	1.33	0.16	0.04	0.31	10.02	0.60
Moebius	1.37	0.17	0.04	0.29	9.51	0.58
Baby2	11.74	1.09	0.35	2.28	76.05	4.82
Average	**3.96**	**0.39**	**0.12**	**0.80**	**26.32**	**1.65**

S3D images with a larger depth range. For SJND, a subjective validation is difficult because there are several factors from different MEs. Tunning the parameters is somewhat complicated and may result in very different results, in addition to the necessary adjustment to the used dataset. As described previously for SJND, the NOR leads to BF, whereas OR leads to BR. In fact, BR can occur on NOR when a large inter-difference exists between left view non-occluded pixels and the corresponding pixels in the right view. The relationship between BF and BR should be better explored to model the human binocular vision.

In contrast to JJND, HJND has taken DC into account in addition to DI. Considering GD makes this model more reliable. However, HJND using GD is specifically developed for multi-view video plus depth (MVD) format, and the estimation based on LR views format may not be correct. The accuracy of this model is highly depending on the rendered images obtained using DIBR.

DJND can estimate the visibility thresholds for S3D video with LR or MVD formats. As reported by the authors, using this model in MVC can increase the VC in the S3D display. However, DJND is less effective for S3D images with small depth difference between FR and BR. In other words, this model performs well if FR and BR have large depth difference.

C. APPLICATIONS EMBEDDING 3D-JND MODELS

In order to improve the compression efficiency of 3D videos, De Silva *et al.* [66] proposed a depth map preprocessing algorithm based on JNDD to remove depth details that are imperceptible by viewers. Similarity, Ding *et al.* [96] recently developed a depth map preprocessing method using JNDD to improve the 3D extension of the high efficiency video coding (HEVC) standard. Bai *et al.* [97] applied JNDD in H.265/HEVC for color image coding by adjusting the quantization parameter (QP). JNDD has also been employed in depth sensation enhancement [98]–[100]. The principle is to increase the depth difference between objects such that it exceeds the JNDD. In addition, Lee *et al.* [101] proposed a stereoscopic watermarking method for DIBR using JNDD. More recently, it has been used in visual presence measurement [102] and 3D QoE (*e.g,* VC and depth sensation) enhancement [103].

Over the past few years, the BJND has been applied in several domains. First, Jung *et al.* [98], [99] applied it in sharpness enhancement of S3D images, and the reliability of BJND has been evaluated by considering the accuracy of the stereo matching algorithm. Second, Sdiri *et al.* [104] recently proposed a contrast enhancement method for stereo endoscopic images combining both local image activity and depth information with BJND. The latter was used to control the inter-view enhancement and avoid visual fatigue. Second, BJND was used in 3D video coding or compression. For instance, Fezza *et al.* [105] proposed a non-uniform asymmetric coding method for S3D video based on BJND and depth level. This method employs BJND to measure the minimum distortion in one view that generates 3D perceptual difference, and then uses depth information to adjust the resolution. Meanwhile, Zhu *et al.* [106] developed a fast mode decision approach using BJND to improve the efficiency of MVC. For S3D compression, a new macroblock level rate

control method based on BJND model has been proposed in [107]. The visual perception factor measured by BJND was used to adjust the macroblock level bit allocation. From a different perspective, BJND was used in several works related to stereoscopic image quality assessment [108]–[113]. The main idea is to use the 2D-JND and BJND to model the visual sensitivity for OR and NOR respectively, and then monocular/binocular visual sensitivity is employed to weight image quality [114]. Besides, Zhou *et al.* [115] proposed a S3D watermarking scheme based on BJND with the aim to guide the watermark embedding. Finally, Shao *et al.* [116] recently carried out a seam carving method for S3D image retargeting combining the 3D visual attention model with BJND.

Wang *et al.* [117] developed a S3D watermarking method using JJND. This method validated the authenticity and integrity of stereoscopic images by localizing the tampered regions. MJND model has been used in order to improve the efficiency of 3D-HEVC [75] and MVD video coding [74]. SJND was used for SVQA, whereas HJND and DJND have been applied to improve 3D CE for MVD. To date, there is no application in other domains for these three models since they have been proposed recently.

V. EXPERIMENTAL RESULTS

In this section, extensive experiments are carried out to compare the performance of the described 3D-JND models. On the one hand, we evaluate the performance using Middlebury stereo database [90] consisting of real-world S3D images. On the other hand, the accuracy estimation of each 3D-JND model for S3D images is compared using psychophysical experiments.

A. PERFORMANCE EVALUATION ON THE MIDDLEBURY STEREO DATABASE

To compare the efficiency of previously described 3D-JND models, we performed an experimental quantitative evaluation as well as a qualitative demonstration using the Middlebury stereo database. As shown in Fig. 11, twenty S3D images from 2005 stereo datasets [27], [28], 2006 stereo datasets [27], [28] and 2014 stereo datasets [29] were chosen for the experimental evaluation. We used the stereo pairs with a full-size resolution from three datasets. The used resolution ranges from 1342 × 1100 to 1390 × 1100 in 2005 stereo datasets, 1240 × 1100 to 1372 × 1100 in 2006 stereo dataset and 2632 × 1988 to 2964 × 2000 in 2014 stereo dataset. These images have been selected based on the number and the ''textureness'' of the objects in FR and BR.

1) QUANTITATIVE EVALUATION AND COMPARISON

Inspired by Wang *et al.* [15], we propose to evaluate the distortion masking ability as a performance of the 3D-JND models. The distortion tolerance ability (DTA) is estimated in terms of energy of the JND map of one view as follows:

$$DTA = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} [JND_{3D}(i,j)]^2 , \qquad (51)$$

FIGURE 11. Right views of the S3D image set from Middlebury databases.

where *DTA* denotes the JND energy of the 3D-JND map (*i.e.*, JND_{3D}) of the left/right view. H and W are the image height and width respectively. In fact, the *DTA* value corresponds to the mean square error (MSE) between original and test images with a maximal degradation. To compute *DTA*, the 3D-JND map of the right image is considered as JND_{3D} for the whole 3D-JND models except SJND. The latter calculates the JND thresholds of the stereo pair [23], [76].

Table 4 shows the distortion tolerance ability of different 3D-JND models. It can be observed that HJND and JJND achieve the best and second-best performance in terms of distortion tolerance ability among all models. This is mainly due to the fact that HJND and JJND thresholds depend highly on disparity/depth values, having a great effect on distortion masking. Higher average disparity value results in higher HJND/JJND energy that corresponds to stronger distortion masking ability. Even though SJND does not take into account disparity/depth information, its masking ability for is close to JJND thanks to the one of consideration of both left and right views. DJND shows lower masking ability than JJND even though both models are developed based on the same 2D-JND model [10]. This could be explained by the DOF blur effect considered in DJND which reduces JND thresholds of FR. MJND takes DM effect into account, where small depth change implies low JND energy. BJND yields the worst performance in terms of distortion masking ability, because the BJND ignores the disparity/depth for 3D visibility thresholds.

Furthermore, we calculated the average disparity level from the ground-truth disparity maps of the Middlebury stereo database, and then revealed the relationship between

TABLE 4. Distortion tolerance ability comparison of 3D-JND models. The best result for each image is highlighted in **boldface**, while the second-best result is shown in *italic*.

Image name	BJND	JJND	MJND	SJND	HJND	DJND
Art	8.143	54.323	29.995	*59.179*	**73.609**	35.411
Books	15.363	36.425	19.592	*40.224*	**42.300**	21.467
Dolls	8.704	*59.187*	30.260	58.713	**77.235**	25.969
Laundry	9.972	35.744	18.896	*42.610*	**46.444**	24.755
Moebius	7.758	*50.469*	21.003	44.669	**54.910**	30.571
Reindeer	7.163	*104.748*	44.623	83.879	**116.702**	57.102
Aloe	15.987	32.127	14.755	*34.038*	33.030	**35.944**
Baby2	16.241	32.188	15.620	*32.573*	**38.140**	30.362
Flowers	14.109	*58.438*	25.986	47.643	**70.535**	56.364
Jadeplant	8.969	*86.622*	38.710	70.508	**104.081**	67.885
Bowling2	12.087	25.378	14.061	*32.072*	**36.324**	20.740
Cloth1	13.013	22.342	10.152	**29.494**	*28.968*	19.281
Midd2	13.928	38.469	19.957	*41.160*	**47.906**	34.007
Plastic	13.138	31.493	17.015	*35.990*	**41.717**	19.425
Rocks2	10.142	32.962	15.286	*35.557*	**38.859**	26.620
Wood1	8.073	26.441	11.283	*28.221*	**31.835**	24.291
Motorcycle	11.068	58.599	30.579	*58.742*	**80.058**	37.431
Piano	9.305	*82.391*	39.199	75.196	**102.741**	44.118
Pipes	9.159	*87.985*	44.078	80.267	**115.784**	58.727
Playroom	10.701	*75.550*	38.174	68.923	**100.328**	42.382
Average	11.151	*51.594*	24.961	49.982	**64.075**	35.643

the average disparity level and the JND energy. Higher the average disparity level lead to stronger distortion tolerance ability for same luminance intensity and luminance contrast. For instance, "Aloe" with an average disparity of 72.44 has lower 3D-JND energy than "Jadeplant" with an average disparity of 270.98. It is worth noting that BJND energy of "Aloe" stereo pair is higher than that of "Jadeplant" stereo pair due to the lack of consideration of DM. Fig. 12 shows that the horizontal shift/disparity between left and right images in "Jadeplant" is larger compared to "Aloe".

Besides, Fig. 13 depicts the plots of the average 3D-JND energies and of the average disparity levels. It can be observed that the average JND energy is approximatively proportional to the average disparity value. More specifically, the visibility threshold of the distortion in S3D image increases as the disparity amplitude increases in the case of similar luminance intensity and luminance contrast. This is consistent with the conclusion drawn in [95].

As the distortion in edge regions is more sensitive to HVS than non-edge regions, the 3D-JND model yielding high JND thresholds for edge region is efficient for 3D compression. Therefore, we further explore the relative strength of 3D-JND for edge pixels using different 3D-JND models. Firstly, we divide image pixels into two regions: edge regions R_E and non-edge regions R_{NE}. To achieve this, we use the method proposed in [12]. This method can accurately detect edge pixels and deal with the issue of the confusion between textural and edge regions. Then, the 3D-JND map JND_{3D} of the right view is estimated based on the 3D-JND model. Finally, the relative strength of the 3D-JND for edge

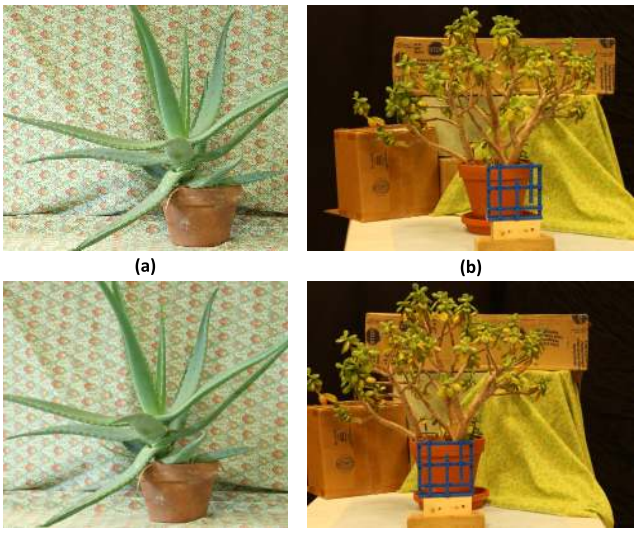


FIGURE 12. "Aloe" and "Jadeplant" stereo pairs. (a) left view of "Aloe", (b) left view of "Jadeplant", (c) right view of "Aloe", (d) right view of "Jadeplant".

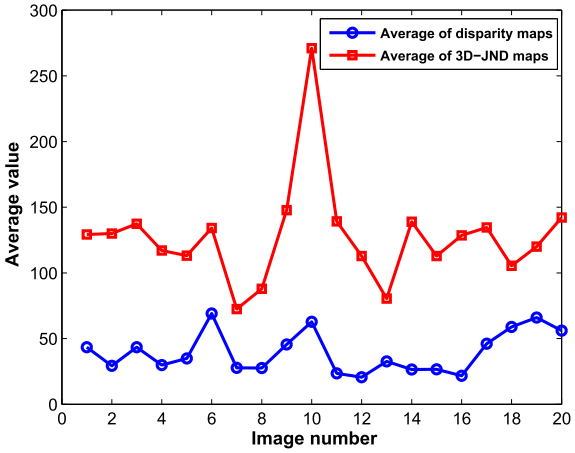


FIGURE 13. Plots of the average 3D-JND energies and of the disparity values.

pixels r_E, the percentage of distortion (*e.g.*, noise) injected to the edge regions, is calculated by:

$$r_E = \frac{\frac{1}{N_E}\sum_{p\in R_E} JND_{3D}(p)}{\frac{1}{N_E}\sum_{p\in R_E} JND_{3D}(p) + \frac{1}{N_{NE}}\sum_{p\in R_{NE}} JND_{3D}(p)}. \quad (52)$$

Higher r_E corresponds to higher distortions for edge regions given the same level of 3D-JND thresholds. The r_E values for the different 3D-JND models in the right view are given in Table 5. Here we assume that left view is not distorted in order to create the asymmetrically distorted stereopairs. The relative strength of the 3D-JND in the left view can also be calculated in a similar manner. MJND performs best in terms of distortion masking ability for edge pixels. BJND achieves better performance in contrast to JJND and HJND since it estimates higher CM around edge regions than non-edge regions. SJND performs quite similarly to BJND.

TABLE 5. The relative strength of 3D-JND for edge pixels using different 3D-JND models. The values below are expressed as a percentage, higher value means that the 3D-JND model can mask more noise in edge regions. The best result for each image is highlighted in **boldface**, while the second-best result is shown in *italic*.

Image name	BJND	JJND	MJND	SJND	HJND	DJND
Art	**59.343**	53.491	*56.426*	55.360	53.162	50.386
Books	54.530	52.493	**68.602**	*62.891*	53.350	51.021
Dolls	*56.214*	53.512	**56.535**	54.808	54.075	51.982
Laundry	*58.902*	52.481	**64.035**	58.737	51.692	50.434
Moebius	**59.497**	54.046	*57.658*	55.862	54.590	50.328
Reindeer	**59.778**	50.138	*54.207*	53.940	49.802	50.027
Aloe	52.462	50.449	**57.332**	*54.708*	52.132	49.692
Baby2	49.147	52.627	**65.314**	*60.776*	53.665	53.501
Flowers	**55.662**	45.248	49.940	*51.773*	48.547	47.326
Jadeplant	**56.801**	46.578	53.009	*54.100*	49.051	45.046
Bowling2	43.966	57.414	61.788	*58.174*	**59.840**	53.918
Cloth1	53.397	*56.081*	53.435	52.073	**56.343**	51.415
Midd2	58.133	55.330	**65.161**	*62.359*	57.017	47.398
Plastic	58.511	51.891	**62.762**	*60.678*	49.736	50.392
Rocks2	55.439	55.398	**60.924**	*57.257*	55.938	53.048
Wood1	**60.501**	50.204	*55.684*	53.933	48.340	49.071
Motorcycle	**60.386**	51.409	*57.389*	56.032	51.728	49.328
Piano	**56.209**	45.537	51.902	*52.472*	47.363	47.779
Pipes	**61.785**	42.814	*52.680*	52.157	44.178	42.735
Playroom	**58.743**	46.054	*53.271*	52.826	46.164	46.261
Average	*56.470*	51.160	**57.903**	56.046	51.836	49.554

In fact, MJND and SJND based on 2D-JND estimated by Chou and Li [5] have higher masking ability for edge regions than HJND, JJND, and DJND relying on the 2D-JND model of Yang *et al.* [10]. This is due to the fact that Yang's 2D-JND model estimates lower CM thresholds for edge pixels, whereas Chou's 2D-JND model considers that the CM thresholds for edge and texture regions are the same. In general, by considering the results in both Table 4 and 5, SJND results in the highest distortion ability among all the 3D-JND models.

2) QUALITATIVE EVALUATION AND COMPARISON

In this section, we provide a qualitative comparison of the six 3D-JND models based on the analysis of the JND profile/map of a stereo pair. The JND thresholds of the right image of a stereo pair were computed using the different 3D-JND models. Dark and bright regions of the JND map indicate regions having low and high JND visibility thresholds, respectively. For a test stereo pair, we used a method proposed in [70] to detect occluded pixels of the left and right videos. Moreover, the hole in disparity/depth maps was filled by using an efficient algorithm proposed by Jain *et al.* [118] in order to obtain the accurate 3D-JND thresholds. This algorithm can accurately fill the holes of the disparity map based on both color and disparity information of the stereo pair.

The 3D-JND maps of "Art", "Plastic", and "Piano" are shown in Fig. 15, 17 and 19, respectively. For "Art", the four circles with different disparity values indicate different

FIGURE 14. "Art" stereo pair. From left to right: left and right views (top), disparity map with holes filling of the right view and occlusion map of the right view. Occluded regions appear in black.

FIGURE 15. JND profiles of "Art" stereo pair obtained using different 3D-JND models. (a) BJND map, (b) JJND map, (c) MJND map, (d) SJND map, (e) HJND map, (f) DJND map.

JND thresholds for all 3D-JND models except BJND. The farther the circle is from the observer, the brighter the circle is, and the higher the JND thresholds are. In other words, the farthest circle has the highest distortion masking ability among all circles. This is especially highlighted in the DJND map (see Fig. 15(f)), because this model distinguishes the FR

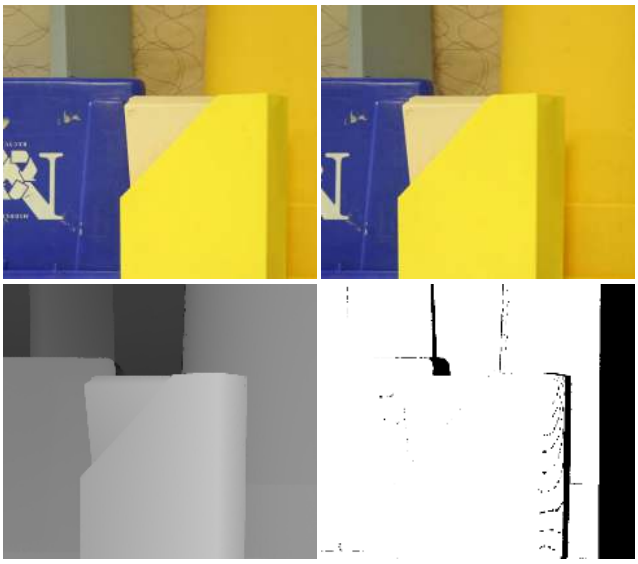


FIGURE 16. "Plastic" stereo pair. From left to right: left and right views (top), disparity map with holes filling of the right view and occlusion map of the right view. Occluded regions appear in black.

from the BR with different visibility thresholds. As shown in Fig. 15(f), JND thresholds of the overall FR are lower than those of the overall background ones. One can also notice that the leftmost object of the DJND map is brighter than that of the HJND map even if DJND and HJND are developed on the top of Yang's 2D-JND model [10].

The comparison between Fig. 17(e) and Fig. 17(f), focusing on the middle yellow object in Fig. 16, also demonstrates the decrease of the DJND thresholds in FR. As shown in Fig. 15(b) and (e), the Yang's 2D-JND model based JJND and HJND maps are similar. However, the JJND map shows a ghosting effect (*e.g.,* around the sculpture) on the occluded pixels (see Fig. 14). This is due to the fact that JJND distinguishes the visibility of the occluded and non-occluded regions in terms of visibility thresholds. The BJND map also shows the ghosting effect on the occluded pixels in Fig. 15(a). It should be noted that the black bands of the BJND maps (Fig. 15(a), Fig. 17(a) and Fig. 19(a)) represent "unknown JND thresholds". The width of this band depends on the maximal ground-truth disparity. In addition, the edge around the middle yellow object in Fig. 17(b) is darker than the one in Fig. 17(e). HJND model not only depends on DI, but also on DC. The latter corresponding to depth variation around the edge of this object (see Fig. 16), and attracting more visual attention, results in a decrease of the distortion masking ability. This conclusion can be demonstrated by referring to the variation of the JND thresholds around the edge of the lampshade in Fig. 19(e).

For BJND, the fact that all circles have quite similar JND thresholds in Fig. 15(a) demonstrates that depth cues have a limited influence on its thresholds. This is indeed consistent with ignoring the binocular disparity in the design of the model. Furthermore, its profiles (Fig. 15(a), Fig. 17(a) and Fig. 19(a)) exhibit higher JND thresholds at edges than

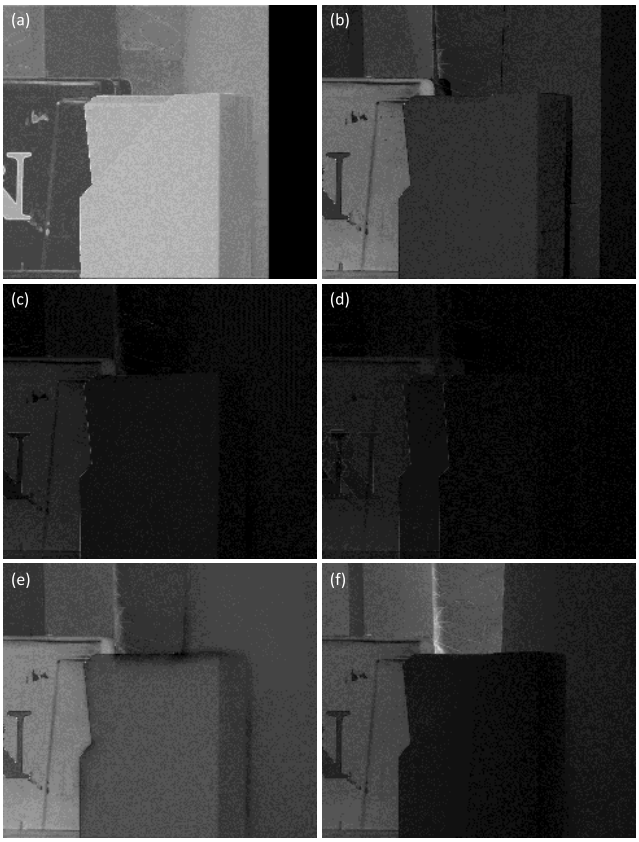


FIGURE 17. JND profiles of "Plastic" stereo pair obtained using different 3D-JND models. (a) BJND map, (b) JJND map, (c) MJND map, (d) SJND map, (e) HJND map, (f) DJND map.

FIGURE 18. "Piano" stereo pair. From left to right: left and right views (top), disparity map with holes filling of the right view and occlusion map of the right view. Occluded regions appear in black.

textures of certain objects, such as the circles in Fig. 15(a) and the music book in Fig. 17(a).

In addition, the edges around the sculpture in Fig. 15(c) and (d) appear slightly brighter than the ones in Fig. 15(b) and (f), and the edges around the middle yellow object in Fig. 17(c) and (d) have slightly higher JND thresholds than the ones in Fig. 17(b) and (f). This can be

FIGURE 19. JND profiles of "Piano" stereo pair obtained using different 3D-JND models. (a) BJND map, (b) JJND map, (c) MJND map, (d) SJND map, (e) HJND map, (f) DJND map.

explained by the fact that MJND and SJND estimate lower CM thresholds for edge regions than texture regions whereas the same CM thresholds are estimated for edge and texture regions in JJND and DJND. The ghosting effect illustrated in Fig. 15(d) indicates that the SJND map is estimated using the left and right views.

B. PERFORMANCE EVALUATION BASED ON PSYCHOPHYSICAL EXPERIMENTS

To evaluate the accuracy of each 3D-JND model, we compare the estimated JND thresholds with the JND thresholds obtained thanks to the psychophysical experiments. We first present the generation of the synthesized 3D images containing textures collected from the ETHZ dataset [30]. Then, we describe the experimental setup used in the subjective measurement of the visibility threshold of the asymmetric distortion. In addition, we further explain how to estimate the visibility thresholds of the synthesized 3D images using previously described 3D-JND models. Finally, we evaluate the 3D-JND models' accuracy by comparing their estimated JND results with the JND data from the psychophysical experiments.

1) SELECTION OF THE TEXTURE IMAGES

In order to generate 3D images to be used in psychophysical experiments, we selected the texture images from ETHZ Synthesizability dataset [30] as the patches of the 3D images. The main idea is to synthesize the 3D images consisting of different texture images. In addition, we explore the relationship between 3D-JND thresholds, the "textureness" of the texture image provided in ETHZ dataset [30] and the average of the 2D-JND thresholds. The latter is calculated by averaging the JND values obtained based on the 2D-JND model described in [12]. The "textureness" score indicates the texture strength of the image. As shown in Fig. 20, the higher the "textureness" score is, the more the image is textured. Fig. 21 shows the textures selection used in our psychophysical experiments. Firstly, the number of texture images is chosen so as to fit in 7 classes according to their "textureness" score and the average of 2D-JND thresholds. Finally, we randomly select one texture image from each dataset. The 7 texture images shown in Fig. 20 are further used to synthesize the 3D images. It is worth noting that these texture images were converted to grayscale for the following experiments.

Image number	1	2	3	4	5	6	7
2D-JNDs average	4.45	5.96	6.97	9.84	10.69	12.12	14.04
Textureness	0.070	0.245	0.435	0.478	0.715	0.803	0.974

FIGURE 20. Seven texture images used to synthesize the S3D stereo pairs. The average of the 2D-JND thresholds and "textureness" values of the texture images are given for each image.

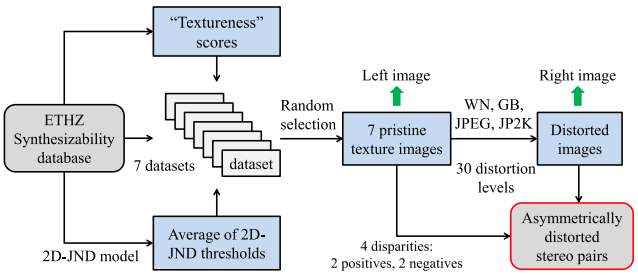


FIGURE 21. A process of the selection of the textures used in psychophysical experiments.

2) STIMULI

To determine the visibility thresholds for the different types of distortion in our psychophysical experiment, we synthesize the asymmetrically distorted stereo pairs based on the 7 pristine texture image shown in Fig. 20. Fig. 22 illustrates an example of the 3D images presented in the psychophysical experiment. The stimulus consists of reference (left) and distorted (right) stereo pairs. The right image of the stereo pair was altered by four types of distortions, including white Gaussian noise (WN), Gaussian blur (GB), JPEG, and JPEG 2000 (JP2K), respectively. Each distortion type was applied using thirty distortion levels, where the control parameters of these distortions indicated in Table 6 were decided to ensure that the subject detects the just noticeable distortion not too early and not too late on the 3D display using stereo glasses. More specifically, the standard deviation σ_{WN} of the WN was used to control the distortion level on the intensity image. The intensity image was filtered using a rotationally symmetric 2D Gaussian kernel of size 7×7 with standard deviation σ_{GB} for GB distortion. The control parameter of the JPEG compression was the quality compression level QC_{JPEG} that determines the amount of information

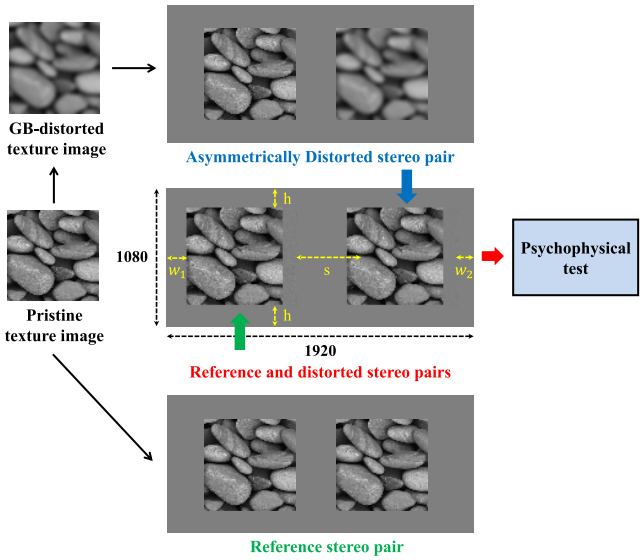


FIGURE 22. An example illustration of the visual stimuli.

TABLE 6. Increment step and value ranges of control parameters for distortion simulation on the right texture view of the stereo pair.

Distortion	Control parameter	Increment step	Range
WN	Standard deviation of the Gaussian distribution	0.8	[0.8, 24]
GB	Standard deviation of the Gaussian low-pass filter	0.1	[0.1, 3]
JPEG	Quality compression level	0.3	[0.3, 9]
JP2K	Quantization value	−0.1	[9.9, 7]

that is lost during compression of the MATLAB functions "im2jpeg.m" and "jpeg2im.m" provided in [119]. Similarly, the JP2K compression was simulated using the MATLAB function "im2jpeg2k.m" and "jpeg2k2im.m" with $n = 5$ and quantization value ranging from 9.9 to 7.

Fig. 22 shows the visual stimuli presented in our psychophysical experiments. They consist of texture images with a resolution of 300 × 300 corresponding to a visual angle of 2.86° × 2.86° with the experimental condition described in Table 7, and a uniform background with an intensity equal to 128. In addition, we created the stimuli with different disparities in order to investigate how the binocular disparity interacts with the detection of the just noticeable distortion. More specifically, the threshold of stereoscopic acuity is approximately 2.3 minutes of arc (arcmin) [120]. In order to easily perceive 3D effects for texture image and to avoid visual fatigue effect during the experiment, we have chosen disparity values well above 2.3 arcmin with both positive and negative parallax: ±26 and ±43 as indicated in Table 7. The positive disparity corresponds to "inside" 3D effect and the negative one to "outside" 3D effect. In contrast to the experiment described in [95], the reference stereo pair has been used in our experimental design. This is because we considered not only the noise distortion but also the blurring

TABLE 7. Viewing conditions and visual stimulus attributes of the psychophysical test. The positive and negative values correspond to the "inside" and "outside" 3D effects for texture images.

Parameter		Value	
Disparity of the texture image	arcmin	±26	±52
	pixel	±43	±86
	degree	±0.44	±0.87
	radian	±0.0076	±0.0151
Screen (m)	width	1.015	
	height	0.57	
Screen resolution (px)	horizontal	1920	
	vertical	1080	
Viewing distance (m)		1.5	
Viewing angle (degree)		37.38	
Background luminance (px)		128	
Ambient illumination (lx)		65	

and compression artifacts. As a result, a 3D image with a resolution of 1920 × 1080 presented in the subjective test is composed of reference and distorted stereo pairs with the parameters reported on Table 7. h and s are set to 390 and 200 pixels, respectively. s is the distance between the reference and the distorted stereo pairs. $s = 200$ ensures that subjects can move its eyes but not the head to detect the stimuli during the experiment. Given the disparity value $Disp$ (in pixel) of a stereo pair, w_2 was defined as $560 - Disp$, and w_1 was set to 560. In sum, a total number of 112 visual stimuli (7 texture images × 4 distortion types × 4 disparities) were presented to the subjects during the psychophysical test. For each stimulus, the level of the asymmetric distortion was increased gradually until the subject binocularly just detects the distortion.

3) SUBJECT

The psychophysical experiment was conducted in the XLIM Laboratory at the University of Poitiers. Eighteen subjects, 13 males and 5 females with age ranging from 25 to 35, participated in this experiment. These subjects are composed of 9 naive participants and 9 expert participants who work in the domain of the image processing/computer vision. Each subject undergoes acuity and stereoscopic acuity test. All subjects have the visual acuity around 1.29 with normal or corrected vision, measured by Freiburg Visual Acuity Test (FrACT) with "Landolt C" setting and 1.2 m of viewing distance. Additionally, they have a stereoscopic acuity more than 70 seconds of arc, checked by the RANDOT stereo test. The subjects who used visual correction in daily life were asked to keep it during the experiment.

4) APPARATUS

The psychophysical experiment was conducted in a diffuse lighting and noise-isolated room designed especially for

subject test. The ambient illumination of the room was set to 65 lux measured by a lux-meter. A 3D display and polarized 3D glasses were used during the experiment. The display is a calibrated Hyundai TriDef S465D with 60 Hz progressive scanning at a resolution of 1920 × 1080, and a display area of 1.015 meters width and height 0.57 meter. It can work with 2D and S3D modes. The brightness of this display was set to 50% of the maximum. The viewing distance between the subject and the 3D display was set to 1.5 meters, which is recommended by the user's guide of this display and approximatively three times the height of the display. The detail of the experimental setup is given in Table 7.

5) EXPERIMENT PROCEDURE

After the visual screening, the subject was informed about the objective of this experiment, and instructed on how to report the results by using the keyboard. He/She was asked to wear the 3D glasses during the whole experiment. The distorted S3D images with different distortion types and levels were presented to each subject in order to get familiar with the experiment. Once subjects confirmed their understanding of the experiment process, the experiment started.

During the experiment, the subjects compared two 3D images, and checked whether the distortion is just noticeable. The subjects pressed "space" key to continue to increase the distortion level the if the previous level is considered as invisible. Otherwise, the subjects pressed the "enter" key to report the JND result for this stimulus. To provide sufficient time to judge the just noticeable distortion, the exposure time of a stimulus is not limited to subjects. Each image of the test sequence related to the stimulus was exposed and followed by a neutral grey image with 128 intensity to avoid visual memory. By pressing the enter key, this ended the current sequence and a message was presented to remind the subject to move to the next stimulus. After 56 visual stimuli, the subjects were asked to take a break of 25 minutes to avoid visual fatigue. For each subject, the experiment was stopped immediately when he/she started feeling visual fatigue. The subjects can move their head freely during the test. For each visual stimulus, we assumed that the measures corresponding to the JND threshold ($(DL_{JND})_n$) from all subjects follows a Gaussian distribution. The experimental DL_{JND} value should be within the interval $[0.95DL_{JND}, 1.05DL_{JND}]$, namely confidence interval, where DL_{JND} is the distortion level corresponding to the maximum value of the histogram. To obtain more accurate JND thresholds, $0.95DL_{JND}$ was selected to show the just unnoticeable distortion in the image, whereas $1.05DL_{JND}$ was selected to show the just noticeable distortion in the image. The estimation of the experimental 3D-JND values is described in Fig. 23(a).

6) JND MAPS ESTIMATION USING PSYCHOPHYSICAL DATA

In order to compare the performance of the 3D-JND models, we first estimated totally the 27 JND maps (7 texture images × 4 disparities) for each model based on the reference 3D images. It should be noted that each synthesized 3D image has only one disparity and no occluded regions. Fig. 24 shows an example of the JND maps of a synthesized 3D image estimated by different 3D-JND models. In addition, the MJND thresholds shown in Fig. 24(c) are higher than JJND thresholds shown in Fig. 24(b) around the edges.

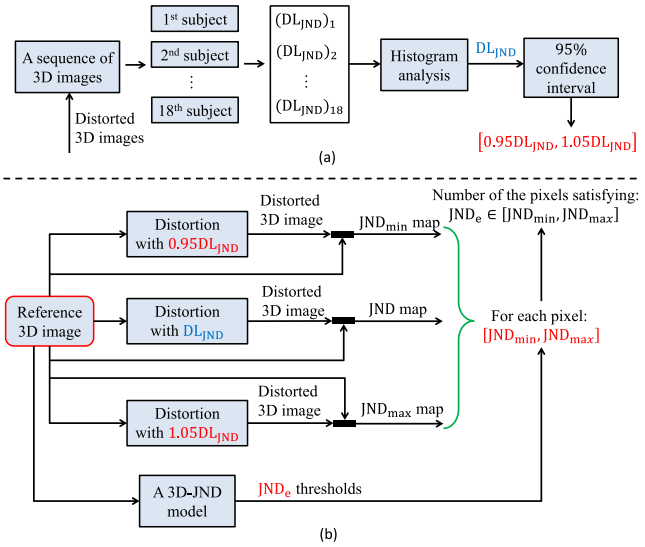


FIGURE 23. Framework of the accuracy evaluation of the 3D-JND model. (a) Estimation of the 3D-JND interval based on psychophysical results, (b) Accuracy evaluation of the 3D-JND model using the 3D-JND interval.

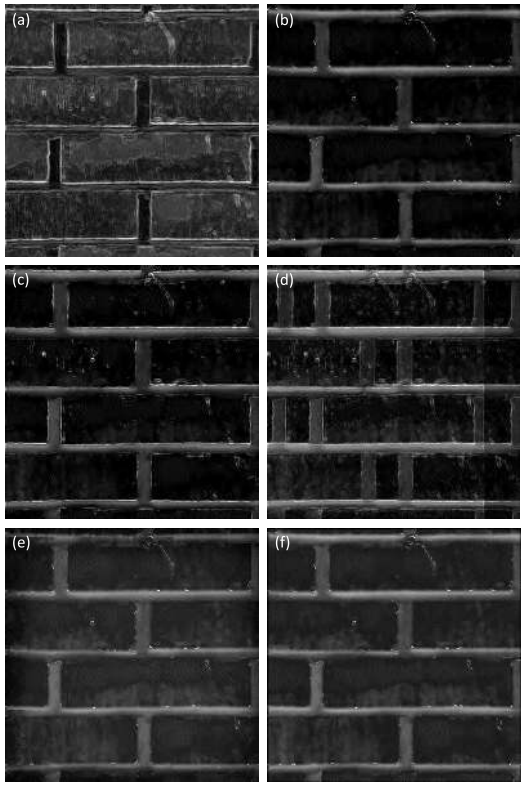


FIGURE 24. An example illustration of the JND maps of a synthesized 3D image estimated by different 3D-JND models. (a) BJND map, (b) JJND map, (c) MJND map, (d) SJND map, (e) HJND map, (f) DJND map.

7) JND MAPS ESTIMATION BASED ON 3D-JND MODELS

As shown in Fig. 23(b), we generated three distorted 3D images using $0.95DL_{JND}$, DL_{JND} and $1.05DL_{JND}$. Then, we computed the difference maps between reference and distorted images, and considered these difference maps as the JND maps. Next, we determined a JND interval namely $[JND_{min}, JND_{max}]$ for each pixel of the reference 3D image. Meanwhile, we estimated the JND map using a 3D-JND model. Finally, we checked whether the estimated JND value of each pixel is included in $[JND_{min}, JND_{max}]$, and computed the number of pixels in the JND map included in their corresponding intervals. The greater the number is, the more the 3D-JND model's accuracy is.

8) RESULTS AND DISCUSSION

In this section, we investigate the effects of distortion type and disparity on the experimental visibility threshold of asymmetric distortions. Besides, the comparison of the estimation accuracy between the 3D-JND models is described. Table 8 shows the average distortion level corresponding to 3D-JND thresholds for different types of distortion according to different texture images. The "textureness" scores of these texture images are reported in Fig. 20. The just noticeable distortion level for WN is proportional to the "textureness" of the texture image. The results for JPEG is similar to those for WN except for images 6 and 7. This is because the image 6, with its low coarseness, can mask more JPEG artifact (*i.e.,* blockiness) than image 7. In contrast to WN, the just noticeable blur (JNB) for GB decreases as the "textureness" score increases. This is due to the fact that images with high texture strength have a low ability to mask the blurring. However, the JNB of image 6 is higher than that of image 5 because of the high contrast of the latter. Accordingly, the blur is easier to be detected in image 5 than in image 6. In general, the just noticeable distortion level is proportional to the "textureness" score for additive distortions as shown in the 6^{th} column of Table 8. For subtractive distortion (GB+JP2K), the just noticeable blur level is inversely proportional to the "textureness" score for the texture images from image 3 to 5.

TABLE 8. Psychophysical distortion level corresponding to the 3D-JND thresholds according to different texture images. The maximal value of each column is highlighted in boldface, while the minimal value is shown in *italic*.

Texture images number	Distortion level corresponding to 3D-JND threshold					
	WN	**GB**	**JPEG**	**JP2K**	**WN + JPEG**	**GB + JP2K**
1	*3.60*	0.95	*1.95*	*0.75*	*2.78*	*0.85*
2	6.40	0.90	3.00	1.05	4.70	0.95
3	5.20	0.85	3.15	1.50	4.18	1.18
4	8.80	1.05	3.90	1.20	6.35	1.12
5	9.00	*0.70*	**4.95**	0.85	6.98	0.78
6	10.60	**1.30**	3.75	1.75	7.17	**1.53**
7	**11.60**	0.85	3.45	**1.80**	**7.53**	1.33

Additionally, we computed the average distortion level corresponding to 3D-JND thresholds according to four disparities in order to explore its effect on asymmetric distortion level threshold. As shown in Table 9, the distortion level corresponding to 3D-JND threshold increases as the absolute disparity value increases for all distortion types. In addition, the distortion level thresholds of disparity $+26$ are generally higher than those of disparity -26 for WN, GB, and JPEG. We can draw the same conclusion when comparing the results of disparity *i.e.,* ± 52 for WN, GB, and JP2K. The 3D image with positive disparity ("inside" 3D effect) is farther than that with negative disparity ("outside" 3D effect) from the subject. The larger the distance between the 3D image and the subject is, the less the distortion is visible, thus the higher the JND threshold is. As a result, the asymmetric distortion level threshold for the image with positive disparity is higher in comparison to the one with negative disparity for the same disparity magnitude. In general, the visibility threshold of the asymmetric distortion is proportional to the disparity magnitude under the same background luminance and luminance contrast. This conclusion is in agreement with the observations in [95]. Furthermore, higher depth values in the 3D image may make the asymmetric distortion more tolerable by the HVS.

TABLE 9. Psychophysical distortion levels corresponding to the 3D-JND thresholds according to different disparities for each type of distortion.

Disparity (arcmin)	Distortion level corresponding to 3D-JND thresholds			
	WN	**GB**	**JPEG**	**JP2K**
-26	11.60	0.85	3.45	1.75
-52	13.40	1.30	6.00	1.90
$+26$	13.20	1.15	4.20	1.50
$+52$	15.60	1.90	6.00	1.95

Based on the psychophysical experiment results mentioned previously, we evaluated and compared the performance of the 3D-JND models in terms of estimation accuracy given in Table 10. Generally, SJND performs the best among all 3D-JND models, while BJND ranks second. HJND has the lowest estimation accuracy within all 3D-JND models. For GB, BJND achieves better performance than SJND. Conversely, SJND performs better than BJND for

TABLE 10. Estimation accuracy (%) comparison of the 3D-JND models according to different distortion types. The larger the value is, the higher the estimation accuracy is. The best result for each distortion type is highlighted in boldface, while the second-best result is shown in *italic*.

Distortion type	BJND	JJND	SJND	HJND	DJND	MJND
WN	2.43	**3.19**	3.02	*3.16*	2.91	2.70
GB	**3.28**	3.05	*3.21*	3.07	3.10	3.16
JPEG	*34.77*	30.08	**36.57**	31.45	32.64	32.24
JP2K	*31.33*	29.85	**32.78**	30.40	30.54	31.01
Average	*17.95*	16.62	**18.89**	17.02	17.30	17.28

JPEG and JP2K. The estimation accuracies for JPEG and JP2K are generally higher in contrast to WN and GB for all 3D-JND models. This is due to the fact that blockiness is easier to notice by HVS than noise and blur. For WN, HJND and JJND based on Chou's 2D-JND model [5] perform better than SJND and MJND based on Yang's 2D-JND model [10]. For WN, the edge region is more sensitive than the texture region, thus the visibility threshold of the edge region should be lower than that of the texture region. Yang's model estimates lower CM thresholds for edge regions than texture regions whereas the same CM thresholds are estimated for edge and texture regions in Chou's model. Therefore, the 3D-JND models based on Yang's model (*i.e.,* HJND and JJND) are more accurate than those based on Chou's model (*i.e.,* SJND and MJND). In contrast, BJND, SJND and MJND perform better than the DJND, HJND, and JJND for GB. For JPEG and JP2K, SJND and BJND achieve higher estimation accuracy than the other models, and JJND performs the worst.

Table 11 shows the estimation accuracy comparison of the 3D-JND models according to four disparities. The results in this table demonstrate that SJND and BJND deliver the best and the second-best performance compared to the other models. DJND and MJND are quite similar in terms of estimation accuracy, because both of them account for depth information. JJND generally performs worse than other models. In addition, the comparison between the results for disparity ± 52 and disparity ± 26 indicates that the larger the disparity magnitude is, the more accurate the 3D-JND models are.

TABLE 11. Estimation accuracy (%) comparison of the 3D-JND models according to four disparities. The larger the value is, the higher the estimation accuracy is. The best result for each distortion type is highlighted in boldface, while the second-best result is shown in *italic*.

Disparity (arcmin)	BJND	JJND	SJND	HJND	DJND	MJND
−26	*15.38*	14.63	**16.47**	14.74	15.19	14.99
−52	*15.86*	15.36	**16.68**	15.40	15.14	15.49
+26	*16.93*	15.02	**17.74**	15.69	16.23	16.15
+52	*18.30*	16.81	**19.13**	17.42	17.48	17.49
Average	*16.62*	15.45	**17.51**	15.81	16.01	16.03

Table 12 shows the estimation accuracy of each 3D-JND model according to different texture images. SJND and BJND outperform all the other models, and thus achieve the best and second-best performance, respectively. DJND is similar to MJND in terms of estimation accuracy. The accuracy of the 3D-JND models for texture image 4 is the highest among 7 texture images. This is mainly due to the fact that highly uniform or textured images (*e.g.,* image 1 or 7) may result in a decrease of the JND estimation accuracy.

It can be noticed that the estimation accuracies of the 3D-JND models for texture 5 and 7 are generally lower compared to the results of other texture images. As shown in Fig. 20, the image 5 has a large coarseness whereas image 7 has a large average contrast. Therefore, the detection of the visibility threshold of the asymmetric distortion in these two images based on psychophysical experiments is error-prone. In summary, results in Table 10, 11 and 12 demonstrate that SJND and BJND outperform the other 3D-JND models in terms of estimation accuracy. This is mainly due to the fact that SJND model accounts for various MEs of both monocular and binocular vision, which undoubtedly correspond better to the human quality judgment. BJND achieves slightly lower accuracy than SJND because it ignores the effect of binocular disparity in the development of this model.

TABLE 12. Estimation accuracy (%) comparison of the 3D-JND models according to different texture images. The larger the value is, higher the estimation accuracy is. The best result for each distortion type is highlighted in boldface, while the second-best result is shown in *italic*.

Texture image	BJND	JJND	SJND	HJND	DJND	MJND
1	**15.88**	15.52	*15.77*	15.74	15.18	15.15
2	*18.40*	17.15	**18.97**	17.45	17.55	17.18
3	**17.56**	14.33	*17.52*	14.83	16.10	15.16
4	**22.17**	14.92	*21.80*	15.82	18.60	17.13
5	12.99	12.91	**14.02**	12.87	12.51	*13.41*
6	19.33	19.47	**21.84**	20.14	19.97	*21.05*
7	12.41	*16.18*	15.08	**16.22**	14.72	15.50
Average	*16.96*	15.78	**17.86**	16.15	16.38	16.37

VI. CONCLUSION

In this paper, we presented a comprehensive review of pixel-based 3D-JND models. The visual characteristics of the HVS considered in these models have been specifically introduced. In addition, these models have been briefly described by giving their rationale and main components in addition to their application, pros, and cons. Besides, we performed an extensive experimental evaluation using Middlebury stereo database with a qualitative demonstration and a performance comparison between these models. Finally, we thoroughly compared the estimation accuracy of the 3D-JND models by using subjective results from our psychophysical experiments. Our study on 3D-JND models allow determining the important characteristics, that will help in the design of a more accurate and efficient 3D-JND model to be used in 3D quality assessment and compression.

REFERENCES

[1] B. A. Wandell, *Foundations of Vision*. Sunderland, MA, USA: Sinauer Associates, 1995.
[2] M. K. Stern and J. H. Johnson, "Just noticeable difference," *The Corsini Encyclopedia of Psychology*. Hoboken, NJ, USA: Wiley, 2010.
[3] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, no. 10, pp. 1385–1422, Oct. 1993.
[4] H. R. Wu, A. R. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *Proc. IEEE*, vol. 101, no. 9, pp. 2025–2043, Sep. 2013.
[5] C.-H. Chou and Y.-C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, Dec. 1995.

[6] I. Hontsch and L. J. Karam, "Adaptive image coding with perceptual distortion control," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 213–222, Mar. 2002.

[7] X. H. Zhang, W. S. Lin, and P. Xue, "Improved estimation for just-noticeable visual distortion," *Signal Process.*, vol. 85, no. 4, pp. 795–808, 2005.

[8] X. Zhang, W. Lin, and P. Xue, "Just-noticeable difference estimation with pixels in images," *J. Vis. Commun. Image Represent.*, vol. 19, no. 1, pp. 30–41, Jan. 2008.

[9] Z. Wei and K. N. Ngan, "Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 337–346, Mar. 2009.

[10] X. K. Yang, W. S. Ling, Z. K. Lu, E. P. Ong, and S. S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Process., Image Commun.*, vol. 20, no. 7, pp. 662–680, Aug. 2005.

[11] Y. Jia, W. Lin, and A. A. Kassim, "Estimating just-noticeable distortion for video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 820–829, Jul. 2006.

[12] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.

[13] Z. Chen and C. Guillemot, "Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 806–819, Jun. 2010.

[14] J. Wu, G. Shi, W. Lin, A. Liu, and F. Qi, "Just noticeable difference estimation for images with free-energy principle," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1705–1710, Nov. 2013.

[15] S. Wang, L. Ma, Y. Fang, W. Lin, S. Ma, and W. Gao, "Just noticeable difference estimation for screen content images," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3838–3851, May 2016.

[16] J. Wu, G. Shi, W. Lin, and C. C. J. Kuo, "Enhanced just noticeable difference model with visual regularity consideration," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1581–1585.

[17] Z. Chen and H. Liu, "JND modeling: Approaches and applications," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Aug. 2014, pp. 827–830.

[18] E.-L. Tan and W.-S. Gan, "Perceptual image coding with discrete cosine transform," in *SpringerBriefs in Electrical and Computer Engineering*. Singapore: Springer, 2015, pp. 21–41.

[19] D. V. S. X. De Silva, W. A. C. Fernando, S. T. Worrall, S. L. P. Yasakethu, and A. M. Kondoz, "Just noticeable difference in depth model for stereoscopic 3D displays," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2010, pp. 1219–1224.

[20] 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.

[21] X. Li, Y. Wang, D. Zhao, T. Jiang, and N. Zhang, "Joint just noticeable difference model based on depth perception for stereoscopic images," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process. (VCIP)*, Nov. 2011, pp. 1–4.

[22] L. Zhou *et al.*, "A new just-noticeable-distortion model combined with the depth information and its application in multi-view video coding," in *Proc. 8th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process. (IIH-MSP)*, Jul. 2012, pp. 246–251.

[23] F. Qi, T. Jiang, X. Fan, S. Ma, and D. Zhao, "Stereoscopic video quality assessment based on stereo just-noticeable difference model," in *Proc. 20th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 34–38.

[24] R. Zhong, R. Hu, Z. Wang, and S. Wang, "3D hybrid just noticeable distortion modeling for depth image-based rendering," *Multimedia Tools Appl.*, vol. 74, no. 23, pp. 10457–10478, 2015.

[25] F. Xue, C. Jung, and J. Kim, "Disparity-based just-noticeable-difference model for perceptual stereoscopic video coding using depth of focus blur effect," *Displays*, vol. 42, pp. 43–50, Apr. 2016.

[26] Y. Fan, M.-C. Larabi, F. A. Cheikh, and C. Fernandez-Maloigne, "On the performance of 3D just noticeable difference models," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1017–1021.

[27] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[28] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[29] D. Scharstein *et al.*, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Proc. German Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2014, pp. 31–42.

[30] D. Dai, H. Riemenschneider, and L. Van Gool, "The synthesizability of texture examples," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3027–3034.

[31] R. H. Wurtz and E. R. Kandel, "Central visual pathways," *Princ. Neural Sci.*, vol. 4, pp. 523–545, Jan. 2000.

[32] S. Winkler, "Vision models and quality metrics for image processing applications," M.S. thesis, Dept. Electr., École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2000.

[33] A. B. Watson, H. B. Barlow, and J. G. Robson, "What does the eye see best?" *Nature*, vol. 302, no. 5907, pp. 419–422, 1983.

[34] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Vis. Comput. Graphics*, vol. 3, no. 4, pp. 291–306, Oct./Dec. 1997.

[35] L. Barghout-Stein, "On differences between peripheral and foveal pattern masking," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, CA, USA, 1999.

[36] O. H. Schade, "Optical and photoelectric analog of the eye," *J. Opt. Soc. Amer.*, vol. 46, no. 9, pp. 721–739, Sep. 1956.

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

[38] A. B. Watson, "Visual detection of spatial contrast patterns: Evaluation of five simple models," *Opt. Express*, vol. 6, no. 1, pp. 12–33, 2000.

[39] A. B. Watson and A. J. Ahumada, "A standard model for foveal detection of spatial contrast," *J. Vis.*, vol. 5, no. 9, p. 6, 2005.

[40] K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *J. Physiol.*, vol. 359, pp. 381–400, Feb. 1985.

[41] K. Momose and M. Saito, "Determination of the chromatic contrast sensitivity using sweep VEP technique," in *Proc. 2nd Joint 24th Annu. Conf. Annu. Fall Meeting Biomed. Eng. Soc. Eng. Med. Biol. (EMBS/BMES)*, vol. 3, Oct. 2002, pp. 2145–2146.

[42] J. G. Robson, "Spatial and temporal contrast-sensitivity functions of the visual system," *J. Opt. Soc. Amer.*, vol. 56, no. 8, pp. 1141–1142, 1966.

[43] G. M. Young, R. B. Goldstein, E. Peli, and L. E. Arend, "Contrast sensitivity to patch stimuli: Effects of spatial bandwidth and temporal presentation," *Spatial Vis.*, vol. 7, no. 1, pp. 1–14, 1993.

[44] J. Rousson, J. Haar, L. Platiša, B. Piepers, T. R. Kimpe, and W. Philips, "Subjective contrast sensitivity function assessment in stereoscopic viewing of Gabor patches," *Proc. SPIE*, vol. 9391, p. 93910O, Mar. 2015.

[45] J. Rousson *et al.*, "Contrast sensitivity function in stereoscopic viewing of Gabor patches on a medical polarized three-dimensional stereoscopic display," *J. Electron. Imag.*, vol. 25, no. 2, pp. 15–25, 2016.

[46] M. J. Nadenau, J. Reichel, and M. Kunt, "Performance comparison of masking models based on a new psychovisual test method with natural scenery stimuli," *Signal Process., Image Commun.*, vol. 17, no. 10, pp. 807–823, 2002.

[47] J. M. Foley, "Human luminance pattern-vision mechanisms: Masking experiments require a new model," *J. Opt. Soc. Amer. A*, vol. 11, no. 6, pp. 1710–1719, 1994.

[48] F. W. Campbell and J. J. Kulikowski, "Orientational selectivity of the human visual system," *J. Physiol.*, vol. 187, no. 2, pp. 437–445, 1966.

[49] L. K. Choi and A. C. Bovik, "Flicker sensitive motion tuned video quality assessment," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation (SSIAI)*, Mar. 2016, pp. 29–32.

[50] D. Pająk *et al.*, "Perceptual depth compression for stereo applications," *Comput. Graph. Forum*, vol. 33, no. 2, pp. 195–204, 2014.

[51] H. Davson, *Physiology of the Eye*. Amsterdam, The Netherlands: Elsevier, 2012.

[52] H. B. Barlow, "Increment thresholds at low intensities considered as signal/noise discriminations," *J. Physiol.*, vol. 136, no. 3, pp. 469–488, 1957.

[53] R. Engbert, A. Longtin, and R. Kliegl, "A dynamical model of saccade generation in reading based on spatially distributed lexical processing," *Vis. Res.*, vol. 42, no. 5, pp. 621–636, 2002.

[54] J. Hyönä, "Foveal and parafoveal processing during reading," in *Oxford Handbook of Eye Movements*. Oxford, U.K.: Oxford Univ. Press, 2011, pp. 819–838.

[55] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1471, Dec. 1980.

[56] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[57] W. Yin, D. Goldfarb, and S. Osher, "Total variation based image cartoon-texture decomposition," Columbia Univ., New York, NY, USA, CORC Rep. TR-2005-01, 2005.
[58] S. P. McKee, M. J. Bravo, H. S. Smallman, and G. E. Legge, "The 'uniqueness constraint' and binocular masking," *Perception*, vol. 24, no. 1, pp. 49–65, 1995.
[59] T. S. Meese, M. A. Georgeson, and D. H. Baker, "Interocular masking and summation indicate two stages of divisive contrast gain control," in *Proc. 28th Eur. Conf. Vis. Perception*, 2005, pp. 42–43.
[60] J. M. Wolfe, "Stereopsis and binocular rivalry," *Psychol. Rev.*, vol. 93, no. 3, pp. 269–282, 1986.
[61] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.
[62] A. Seyler and Z. Budrikis, "Detail perception after scene changes in television image presentations," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 1, pp. 31–43, Jan. 1965.
[63] B. G. Breitmeyer and H. Ogmen, "Recent models and findings in visual backward masking: A comparison, review, and update," *Perception Psychophysics*, vol. 62, no. 8, pp. 1572–1595, 2000.
[64] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 742–752, Jun. 2005.
[65] C.-H. Chou and C.-W. Chen, "A perceptually optimized 3-D subband codec for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 143–156, Apr. 1996.
[66] D. V. S. X. De Silva, E. Ekmekcioglu, W. A. C. Fernando, and S. T. Worrall, "Display dependent preprocessing of depth maps based on just noticeable depth difference modeling," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 335–351, Apr. 2011.
[67] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, S. Villette, and A. M. 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.
[68] D. V. S. X. De Silva, W. A. C. Fernando, G. Nur, E. Ekmekcioglu, and S. T. Worrall, "3D video assessment with just noticeable difference in depth evaluation," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 4013–4016.
[69] R. Bensalma and M.-C. Larabi, "Binocular energy estimation based on properties of the human visual system," *Cognit. Comput.*, vol. 5, no. 4, pp. 589–609, 2013.
[70] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.
[71] J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum, "Symmetric stereo matching for occlusion handling," in *Proc. IEEE Conf. Soc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 399–406.
[72] M. Flierl and B. Girod, "Multiview video compression," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 66–76, Nov. 2007.
[73] Y. Chen, M. M. Hannuksela, L. Zhu, A. Hallapuro, M. Gabbouj, and H. Li, "Coding techniques in multiview video coding and joint multiview video model," in *Proc. Picture Coding Symp.*, May 2009, pp. 1–4.
[74] C. Liu, P. An, Y. Zuo, and Z. Zhang, "Applications of just-noticeable depth difference model in joint multiview video plus depth coding," *Proc. SPIE*, vol. 9273, p. 92732R, Oct. 2014.
[75] Y. Shi, Y. Wang, and Y. Wang, "A perceptual preprocess method for 3D-HEVC," *Proc. SPIE*, vol. 9622, p. 962219, Aug. 2015.
[76] F. Qi, D. Zhao, X. Fan, and T. Jiang, "Stereoscopic video quality assessment based on visual attention and just-noticeable difference models," *Signal, Image Video Process.*, vol. 10, no. 4, pp. 737–744, 2016.
[77] A. Vetro, "Representation and coding formats for stereo and multiview video," in *Intelligent Multimedia Communication: Techniques and Applications*. Berlin, Germany: Springer, 2010, pp. 51–73.
[78] P. Aflaki, M. M. Hannuksela, J. Häkkinen, P. Lindroos, and M. Gabbouj, "Subjective study on compressed asymmetric stereoscopic video," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 4021–4024.
[79] R. Blake and N. K. Logothetis, "Visual competition," *Nature Rev. Neurosci.*, vol. 3, no. 1, pp. 13–21, 2002.
[80] R. Zhong *et al.*, "Just noticeable difference for 3D images with depth saliency," in *Proc. Pacific-Rim Conf. Multimedia*. Berlin, Germany: Springer, 2012, pp. 414–423.
[81] Y. Zhang, G. Jiang, M. Yu, K. Chen, and Q. Dai, "Stereoscopic visual attention-based regional bit allocation optimization for multiview video coding," *EURASIP J. Adv. Signal Process.*, vol. 2010, p. 848713, Dec. 2010.
[82] H. Shum and S. B. Kang, "Review of image-based rendering techniques," *Proc. SPIE*, vol. 4067, pp. 2–13, May 2000.
[83] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *Proc. SPIE*, vol. 5291, pp. 93–104, May 2004.
[84] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, "MESH: Measuring errors between surfaces using the Hausdorff distance," in *Proc. ICME*, Aug. 2002, pp. 705–708.
[85] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.*, vol. 8, no. 3, p. 33, 2008.
[86] Y. J. Jung, H. Sohn, S.-I. Lee, F. Speranza, and Y. M. Ro, "Visual importance- and discomfort region-selective low-pass filtering for reducing visual discomfort in stereoscopic displays," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1408–1421, Aug. 2013.
[87] L. Zhang, Q. Peng, Q.-H. Wang, and X. Wu, "Stereoscopic perceptual video coding based on just-noticeable-distortion profile," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 572–581, Jun. 2011.
[88] X. Wang *et al.*, "Visibility threshold of compressed stereoscopic image: Effects of asymmetrical coding," *Imag. Sci. J.*, vol. 61, no. 2, pp. 172–182, 2013.
[89] B. Du *et al.*, "Novel visibility threshold model for asymmetrically distorted stereoscopic images," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process. (VCIP)*, Nov. 2016, pp. 1–4.
[90] D. Scharstein and R. Szeliski. (2002). *Middlebury Stereo Vision Page*. [Online]. Available: http://www.middlebury.edu/stereo
[91] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2003, pp. 195–202.
[92] V. De Silva, A. Fernando, S. Worrall, H. K. Arachchi, and A. Kondoz, "Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-D displays," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 498–506, Jun. 2011.
[93] Y. Zhang, S. Kwong, L. Xu, S. Hu, G. Jiang, and C.-C. J. Kuo, "Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3497–3512, Sep. 2013.
[94] Z. Peng, F. Chen, G. Jiang, M. Yu, F. Shao, and Y.-S. Ho, "Depth video spatial and temporal correlation enhancement algorithm based on just noticeable rendering distortion model," *J. Vis. Commun. Image Represent.*, vol. 33, pp. 309–322, Nov. 2015.
[95] H. G. Kim, S.-I. Lee, and Y. M. Ro, "Experimental investigation of the effect of binocular disparity on the visibility threshold of asymmetric noise in stereoscopic viewing," *Opt. Express*, vol. 24, no. 17, pp. 19607–19615, 2016.
[96] H. Ding, Z. Li, and D. Li, "Depth map pre-processing algorithm for compression based on 3D-HEVC scheme," in *Proc. 16th IEEE Int. Conf. Commun. Technol. (ICCT)*, Oct. 2015, pp. 290–294.
[97] Y. Bai, Y. Zhang, and Z. Li, "3D video coding using just noticeable depth difference based on H. 265/HEVC," in *Proc. 11th Int. Conf. Comput. Intell. Secur. (CIS)*, Dec. 2015, pp. 142–145.
[98] S.-W. Jung and S.-J. Ko, "Depth sensation enhancement using the just noticeable depth difference," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3624–3637, Aug. 2012.
[99] S.-W. Jung, "A modified model of the just noticeable depth difference and its application to depth sensation enhancement," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3892–3903, Oct. 2013.
[100] J. Lei, C. Zhang, Y. Fang, Z. Gu, N. Ling, and C. Hou, "Depth sensation enhancement for multiple virtual view rendering," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 457–469, 2015.
[101] J.-W. Lee, H.-D. Kim, H.-Y. Choi, S.-H. Choi, and H.-K. Lee, "Stereoscopic watermarking by horizontal noise mean shifting," *Proc. SPIE*, vol. 8303, p. 830307, Feb. 2012.
[102] H. Oh and S. Lee, "Visual presence: Viewing geometry visual information of UHD S3D entertainment," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3358–3371, Jul. 2016.

[103] F. Shao, W. Lin, Z. Li, G. Jiang, and Q. Dai, "Toward simultaneous visual comfort and depth sensation optimization for stereoscopic 3-D experience," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4521–4533, Dec. 2017.
[104] B. Sdiri, A. Beghdadi, F. A. Cheikh, M. Pedersen, and O. J. Elle, "An adaptive contrast enhancement method for stereo endoscopic images combining binocular just noticeable difference model and depth information," *Electron. Imag.*, vol. 2016, no. 13, pp. 1–7, 2016.
[105] S. A. Fezza, M.-C. Larabi, and K. M. Faraoun, "Asymmetric coding of stereoscopic 3D based on perceptual significance," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5656–5660.
[106] Y. Zhu, M. Yu, X. Jin, Z. Peng, F. Shao, and G. Jiang, "Fast mode decision algorithm for multiview video coding based on binocular just noticeable difference," *J. Comput.*, vol. 9, no. 10, pp. 2428–2434, 2014.
[107] G. Zhu *et al.*, "A novel macroblock level rate control method for stereo video coding," *Sci. World J.*, vol. 2014, Mar. 2014, Art. no. 136854.
[108] S. A. Fezza, M.-C. Larabi, and K. M. Faraoun, "Stereoscopic image quality metric based on local entropy and binocular just noticeable difference," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2002–2006.
[109] W. Hachicha, A. Beghdadi, and F. A. Cheikh, "Stereo image quality assessment using a binocular just noticeable difference model," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 113–117.
[110] F. Shao, G.-Y. Jiang, M. Yu, F. Li, Z. Peng, and R. Fu, "Binocular energy response based quality assessment of stereoscopic images," *Digit. Signal Process.*, vol. 29, no. 1, pp. 45–53, Mar. 2014.
[111] W. Zhou, G. Jiang, M. Yu, F. Shao, and Z. Peng, "PMFS: A perceptual modulated feature similarity metric for stereoscopic image quality assessment," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 1003–1006, Aug. 2014.
[112] Y. Cao, W. Hong, and L. Yu, "Full-reference perceptual quality assessment for stereoscopic images based on primary visual processing mechanism," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2016, pp. 1–6.
[113] Y. Lin, J. Yang, L. Wen, Q. Meng, Z. Lv, and H. Song, "Quality index for stereoscopic images by jointly evaluating cyclopean amplitude and cyclopean phase," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 89–101, Feb. 2017.
[114] X. Wang, S. Kwong, Y. Zhang, and Y. Zhang, "Considering binocular spatial sensitivity in stereoscopic image quality assessment," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process. (VCIP)*, Nov. 2011, pp. 1–4.
[115] W. Zhou, L. Yu, Z. Wang, M. Wu, T. Luo, and L. Sun, "Binocular visual characteristics based fragile watermarking scheme for tamper detection in stereoscopic images," *AEU-Int. J. Electron. Commun.*, vol. 70, no. 1, pp. 77–84, 2016.
[116] F. Shao, W. Lin, W. Lin, G. Jiang, M. Yu, and R. Fu, "Stereoscopic visual attention guided seam carving for stereoscopic image retargeting," *J. Display Technol.*, vol. 12, no. 1, pp. 22–30, Jan. 2016.
[117] J. Wang, M. Yu, T. Luo, F. Shao, Z. Peng, and G. Jiang, "Joint just noticeable distortion based stereo image watermarking method with self-recovery," *Future Inf. Eng.*, vol. 49, p. 51, Mar. 2014.
[118] A. K. Jain, L. C. Tran, R. Khoshabeh, and T. Q. Nguyen, "Efficient stereo-to-multiview synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 889–892.
[119] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*. New York, NY, USA: McGraw-Hill, 2010.
[120] B. E. Coutant and G. Westheimer, "Population distribution of stereoscopic ability," *Ophthalmic Physiol. Opt.*, vol. 13, no. 1, pp. 3–7, 1993.

YU FAN received the M.Sc. degree in telecommunications, multimedia and automation from the University of Poitiers, France, in 2013. He is currently pursuing the Ph.D. degree in signal and image processing with the Norwegian University of Science and Technology and the University of Poitiers. His research interests include visual quality assessment and binocular perception modeling.

MOHAMED-CHAKER LARABI (M'05–SM'07) received the Ph.D. degree from the University of Poitiers, in 2002. He is currently an Associate Professor with the University of Poitiers. He is the Deputy Scientific Director of the GdR-ISIS (French research group on signal and image processing). He has published over 150 papers. His actual scientific interests include the quality of experience and bio-inspired processing/coding/optimization of images and videos, 2D, 3D, HDR, and 360. He is a member of MPEG and JPEG committees, CIE, and IS&T. He has been elected to serve as a member for the IEEE SPS IVMSP and MMSP technical committees. He served as the Chair for the JPEG Advanced Image Coding and the Test and Quality Subgroup. He acted as the French Head of Delegation for several years. He played several roles at different conferences. He was the Program Chair of the EUVIP 2011 and 2018, the Plenary Chair of the EUVIP 2013, the Chair of the EI Image Quality and System Performance, from 2014 to 2016, and the Short Courses Co-Chair of the EI, from 2016 to 2018. He was the Special Sessions Co-Chair of the ICIP 2016 and the Publicity Chair of the ICIP 2017. He serves as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the Springer SIVP, the SPIE/IS&T JEI, the IEEE ACCESS, the Elsevier JVCI, and the Elsevier SPIC. He is a regular reviewer for several international conferences and journals. He participated at several national and international projects. He supervised more than 15 Ph.D. students.

FAOUZI ALAYA CHEIKH (M'10–SM'13) received the Ph.D. degree in information technology from the Tampere University of Technology, Tampere, Finland, in 2004. He has been a Researcher with the Signal Processing Algorithm Group, Tampere University of Technology, since 1994. Since 2006, he has been an Associate Professor with the Department of Computer Science and Media Technology, Gjøvik University College, Norway. Since 2016, he has been with the Norwegian University of Science and Technology. He teaches courses on image and video processing and analysis and media security. His research interests include e-learning, 3D imaging, image and video processing and analysis, video-based navigation, biometrics, pattern recognition, embedded systems, and content-based image retrieval. In these areas, he has published over 100 peer-reviewed journal and conference papers, and has supervised four Post-doctoral researchers, five Ph.D. students, and a number of M.Sc. dissertation projects. He is currently a Co-Supervisor of five Ph.D. students. He has been involved in several European and national projects, among them: ESPRIT, NOBLESS, COST 211Quat, HyPerCept, IQ-Med, and H2020 ITN HiPerNav. He is a member of NOBIM and Forskerforbundet (The Norwegian Association of Researchers). He is on the Editorial Board of the *IET Image Processing Journal*, the *Journal of Advanced Robotics and Automation*, and on the technical committees of several international conferences. He is an expert reviewer for a number of scientific journals and conferences related to the field of his research.

CHRISTINE FERNANDEZ-MALOIGNE is currently a Vice-Rector of Poitiers University, where she is in charge of international relations and the Director of the CNRS Research Federation (MIRES), which gathers 560 researchers in the southwest of France, in the area of mathematics, image processing, computer graphic, computer science, and communication systems. Her research activities are focused on color imaging, including fundamental researches about introduction of human visual system models in multiscale color image processes as well as practical applications for biomedical, patrimonial, and audio-visual digital contents. She is currently an Appointed Member of the National Council of the French Universities, the Secretary of Division 8, Image Technologies, of the CIE (International Commission of Lighting), and the Deputy Editor-in-Chief of JOSA A.