

Binocular Just-Noticeable-Difference Model for Stereoscopic Images

Yin Zhao, Zhenzhong Chen, *Member, IEEE*, Ce Zhu, Yap-Peng Tan, *Senior Member, IEEE*, and Lu Yu

Abstract—Conventional 2-D Just-Noticeable-Difference (JND) models measure the perceptible distortion of visual signal based on monocular vision properties by presenting a single image for both eyes. However, they are not applicable for stereoscopic displays in which a pair of stereoscopic images is presented to a viewer's left and right eyes, respectively. Some unique binocular vision properties, e.g., binocular combination and rivalry, need to be considered in the development of a JND model for stereoscopic images. In this letter, we propose a binocular JND (BJND) model based on psychophysical experiments which are conducted to model the basic binocular vision properties in response to asymmetric noises in a pair of stereoscopic images. The first experiment exploits the joint visibility thresholds according to the luminance masking effect and the binocular combination of noises. The second experiment examines the reduction of visual sensitivity in binocular vision due to the contrast masking effect. Based on these experiments, the developed BJND model measures the perceptible distortion of binocular vision for stereoscopic images. Subjective evaluations on stereoscopic images validate of the proposed BJND model.

Index Terms—Binocular just-noticeable-difference (BJND), contrast masking, luminance masking, stereoscopic images.

I. INTRODUCTION

THREE-dimensional video is becoming more and more prevalent today. When watching a 3-D video, a viewer receives two views of the scene, which probably contain different distortions due to asymmetric video processing in the process of capture, coding, and streaming. For instance, asymmetric video coding may result in different fidelity in different views, [1]. Just-noticeable-difference (JND) models for conventional 2-D video have been developed for determining the maximum waveform distortions that are imperceptible or at least tolerable to the human visual system (HVS) when viewing visual signal from a single image [2]–[4]. However, the asymmetric distortions in a stereo pair of 3-D video usually exist on the two views with different intensities, locations, and appearances. In this case, the conventional 2-D-JND models fail to indicate the visibility of

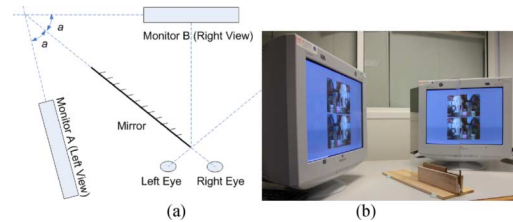


Fig. 1. Mirror-based stereoscopic display system. (a) Schematic illustration of the system. (b) Actual setup in the validation experiment.

asymmetric distortions (or noises). This calls for a new model to measure the visibility of such distortions in stereoscopic images.

Generally, the JND models are developed based on HVS features, e.g., the luminance and contrast masking effects. To build a binocular JND (BJND) model, we have conducted psychophysical experiments to model the two masking effects in binocular viewing. Supposing two stereoscopic images are represented with asymmetric noises at intensity levels A_1 and A_2 , respectively, the luminance masking experiment aims to quantify the visibility thresholds of the combined noises ($\{A_1, A_2\}$ that makes the binocular noises just noticeable) subject to uniform background luminance conditions. The contrast masking experiment attempts to measure the just perceptible noise levels due to the background contrast. Based on the two experiments, we propose a BJND model to measure the joint visibility of a pair of distortions in two views. The BJND model determines the minimum distortions in one view that evoke binocularly visible differences, given the background information and the distortions in the corresponding area of the other view.

The rest of the letter is organized as follows. In Section II, we introduce two psychophysical experiments to model the luminance and contrast masking effects for binocular vision, and then develop the BJND model. Section III provides experiments to and Section IV concludes the letter.

II. BINOCULAR JUST-NOTICEABLE-DIFFERENCE MODEL

A. Experimental Setup

A mirror-based stereoscopic display system was used to present stereoscopic images, as illustrated in Fig. 1(a). Compared with stereoscopic displays with shutter or polarized glasses, this mirror-based system is free from cross-talk effects and brightness attenuation of the glasses. In the psychophysical experiments in this section, we used two Philips 190B LCD monitors (color temperature 6000 K, ambient luminance 0.43 cd/m^2). As the 8-bit image format is prevalent, we use digital luminance levels throughout the letter which correspond to the physical luminance values provided in Fig. 3(d). We also limited the viewing distance to about 1.5 times of the monitor height, and calibrated the monitors' luminance to compensate the slight brightness decrease in the left view due to light reflection on the

Manuscript received August 18, 2010; revised October 10, 2010; accepted October 14, 2010. Date of publication October 28, 2010; date of current version November 22, 2010. This work was supported in part by LKY2/09 Start-Up Grant M58040018, Singapore Ministry of Education Academic Research Fund Tier 1 (AcRF Tier 1 RG7/09) and the National Basic Research Program of China (973) under Grant.2009CB320903. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alexander C. Loui.

Y. Zhao, Z. Chen, C. Zhu, and Y. P. Tan are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: zhaoyin@ntu.edu.sg; zzchen@ntu.edu.sg; eczhu@ntu.edu.sg; eypatn@ntu.edu.sg).

L. Yu is with the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (email: yul@zju.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2010.2090041

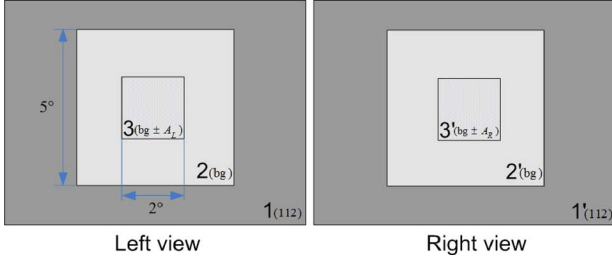


Fig. 2. Binocular patterns used in the luminance masking experiment.

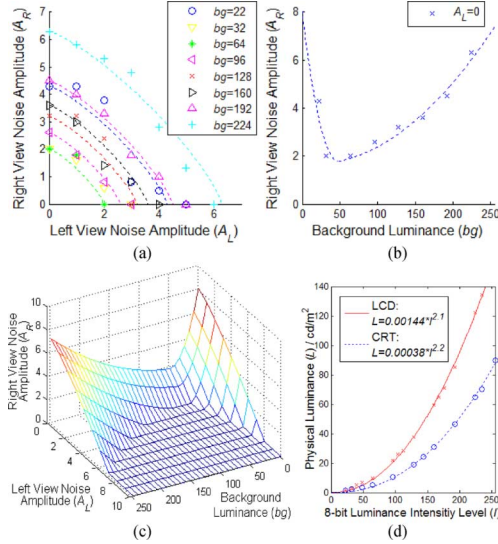


Fig. 3. (a) $\{A_L, A_R\}$ pairs under eight background luminance levels (the dash curves are fitting curves of (5) with $\lambda = 1.25$). (b) The A_R values when $A_L = 0$ under eight background luminance levels (the dash curve is the curve of fitting function (4)). (c) The BJND profile based on luminance masking effect as indicated by (6). (d) The mapping functions from 8-bit luminance level to physical luminance (cd/m^2) for the LCD display in the psychophysical experiments in Section II (red solid line) and for the CRT display in the validation test in Section III (blue dot line), respectively.

mirror. Five subjects (three male and two female with or corrected to normal vision) participated in the two psychophysical experiments elaborated in the following two subsections.

B. Modeling Luminance Masking Effect and Noise Combination

Subjects were presented with a pair of patterns shown in Fig. 2. In the center of a view, there was a $5^\circ \times 5^\circ$ (covering the parafovea) square (Region 2) of constant luminance bg , outside which the luminance level was fixed at a medium value (i.e., 112 in the experiment). The central $2^\circ \times 2^\circ$ (covering the fovea) was added with noise of random polarizations and fixed amplitude A_L (or A_R) in the left (or right) view (the noise patterns in two eyes are identical). Disparity between the patterns is zero, as we focus on the binocular combination property and assume the disparity (compensated by eye convergence) imposes minor impact on the combination. Given a left-view noise amplitude A_L , if the noise is binocularly invisible, subject n was required to adjust the right view noise amplitude $A_R(n)$ to make the pair of noises just noticeable based on the staircase method (reversal 3) [5] instead of that of ascending limits used in [3]. Each test took about 40 minutes (with a 30-second on-site rest every 5 minutes). All pairs of $\{A_L, A_R(n)\}$ that induced just noticeable binocular noise were recorded, and the average of

$A_R(n)$, denoted as A_R , was computed based on the majority rule as

$$A_R = \begin{cases} 0, & \text{if } p(A_R(n) = 0) > 0.75 \\ \text{avg}(A_R(n)), & \text{otherwise} \end{cases} \quad (1)$$

where $p(A_R(n) = 0)$ represents the average probability of a subject perceiving noise with $\{A_L, 0\}$ pair, and $\text{avg}(A_R(n))$ is the average of the individual thresholds. It means that if most subjects report $\{A_L, 0\}$ to be visible, the invisible judgment from the minority (less than 25%) is ignored. As some subjects had slightly higher thresholds due to their stricter threshold criterion, we adopted this operation to prevent the averaged threshold A_R from being nonzero when there are a minority of nonzero values of $A_R(n)$. Then, we collected $\{A_L, A_R\}$ pairs under eight background luminance levels (bg) from 22 to 224, as shown in Fig. 3(a). The cases for $bg = 0$ and 255 are not considered, because the monitor luminance calibration based on RGB mapping function limited the mutual luminance range of the two views from 16 to 240.

Based on prior vision studies on binocular vision [6], [7] and the symmetric performance of two eyes as shown in Fig. 3(a), we model the binocular noise combination with a more generic function in contrast with that in [6]

$$B(A_1, A_2) = \alpha \cdot A_1^\lambda + \beta \cdot A_2^\lambda, \quad \lambda > 0 \quad (2)$$

where α and β ($\alpha + \beta = 1$) are the weights for left and right eyes, and $B(A_1, A_2)$ denotes the perceived binocular noise amplitude based on the monocular noise amplitude A_1 and A_2 in the two views. The weights vary due to contour information [6] in the two views. In our experiment, the surrounding areas of the noise-injected regions in the two views were identical. Thus, we assumed $\alpha = \beta = 0.5$.

Let $Th(bg)$ denote the binocular threshold for perceptible distortion under background luminance levels (bg), we have

$$\begin{aligned} Th(bg) &= 0.5 \cdot (A_1^\lambda + A_{2\min}^\lambda(bg, A_1)) \\ &= 0.5 \cdot (0 + A_{2\min}^\lambda(bg, 0)) \end{aligned} \quad (3)$$

where $A_{2\min}(bg, A_1)$ is the minimum noise amplitude in one view that evokes perceptible difference under a certain bg for a given A_1 in the other view. Denote $A_{\text{limit}}(bg) = A_{2\min}(bg, 0)$ which is the upper limit of $A_{2\min}(bg, A_1)$ for $A_1 = 0$. Some sample values of $A_{\text{limit}}(bg)$ obtained from the experiment (where A_1 refers to the left view) are shown in Fig. 3(b), which can be fitted as

$$A_{\text{limit}}(bg) = \begin{cases} 0.0027 \cdot (bg^2 - 96 \cdot bg) + 8, & \text{if } 0 \leq bg < 48 \\ 0.0001 \cdot (bg^2 - 32 \cdot bg) + 1.7, & \text{if } 48 \leq bg \leq 255 \end{cases} \quad (4)$$

Equation (3) can be rewritten as

$$\left(\frac{A_1}{A_{\text{limit}}(bg)} \right)^\lambda + \left(\frac{A_{2\min}(bg, A_1)}{A_{\text{limit}}(bg)} \right)^\lambda = 1 \quad (5)$$

where $0 \leq A_1, A_{2\min}(bg, A_1) \leq A_{\text{limit}}(bg)$.

For a local region i (e.g., a pixel) in one view with background luminance level $bg(i)$ and noise amplitude $A_1(i)$, we can obtain the minimum perceptible noise amplitude in the corresponding region j (established by stereo matching) in the other view by

TABLE I
RESULTS OF FITTING FUNCTION WITH DIFFERENT λ VALUES

λ	0.5	0.75	1.0	1.25	1.5
RMSE	2.019	1.312	0.802	0.592	0.644
λ	1.75	2.0	3.0	4.0	5.0
RMSE	0.787	0.931	1.311	1.490	1.583

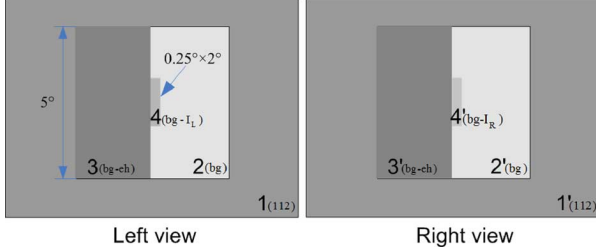


Fig. 4. Binocular patterns used in the contrast masking experiment.

(5). Thus, we define an initial BJND profile according to the luminance masking and noise combination:

$$\begin{aligned}
 BJND_L(bg(i), A_1(i)) &= A_{2\min}(bg(i), A_1(i)) \\
 &= A_{\text{limit}}(bg(i)) \cdot \left(1 - \left(\frac{A_1(i)}{A_{\text{limit}}(bg(i))}\right)^\lambda\right)^{1/\lambda}, \\
 0 \leq A_1(i) &\leq A_{\text{limit}}(bg(i))
 \end{aligned} \quad (6)$$

where $bg(i)$ is the average of pixel luminance of region i in one view. The value of λ can be obtained by minimizing the mean squared error between all subjective A_R values in the luminance masking experiment (as shown in Fig. 3(a)) and the predicted values according to the BJND model indicated by (6). The root mean squared error (RMSE) scores under different λ values are listed in Table I, which suggests the λ value for binocular noise combination is within the range of [1.0, 1.5]. In this letter, we use $\lambda = 1.25$ which tends to yield the best fitting performance. The initial luminance-masking-based BJND profile is shown as the surface in Fig. 3(c).

C. Modeling Contrast Masking Effect

Contrast masking effect suggests that visibility thresholds increase with contrast [8]. In [9], luminance edge is used to describe the contrast degree, and the results show visibility thresholds are elevated linearly by the edge height. We have designed an experiment to model the contrast masking effect with asymmetric distortions. In the center of the left view, there is a $5^\circ \times 5^\circ$ square divided evenly into two regions: region 2 with luminance of bg and region 3 with luminance of $(bg-eh)$, which presents a sharp edge of height eh as shown in Fig. 4. Outside the square, the luminance level is fixed at 112. A $0.25^\circ \times 2^\circ$ slice (region 4) is located along the edge, and the luminance of the slice is distorted to $(bg-I_L)$. The right view is arranged similarly, except that the luminance intensity in region 4' is $(bg-I_R)$.

Similar to the luminance masking experiment, given a I_L , a subject n is asked to adjust $I_R(n)$ based on staircase method (reversal 3) and report all $\{I_L, I_R(n)\}$ that make the edge distortion just visible. $I_R(n)$ are averaged to obtain the mean value I_R , as shown in Fig. 5(a). The average gradients (K) of the three fitting lines under the background luminance levels 96, 144, and 192 are 0.059, 0.053, and 0.038, respectively. The results suggest that visibility thresholds are generally elevated almost lin-

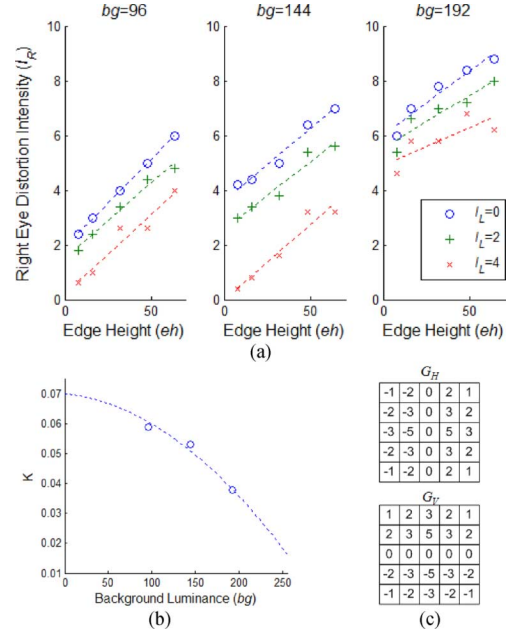


Fig. 5. (a) $\{I_L, I_R\}$ pairs for just visible distortions under different edge heights and background luminance levels. (b) Polynomial fitting function for the average gradient of the three fitting lines under each test background luminance level in (a). (c) Sobel operators for calculating horizontal and vertical edge heights.

early by the surrounding edge height, but the elevating effect decreases as the background luminance increases, which can be approximated by

$$A_{C,\text{limit}}(bg, eh) = A_{\text{limit}}(bg) + K(bg) \cdot eh \quad (7)$$

$$\begin{aligned}
 K(bg) &= -10^{-6} \cdot (0.7 \cdot bg^2 + 32 \cdot bg) + 0.07, \\
 bg &\in [0, 255]
 \end{aligned} \quad (8)$$

where $A_{C,\text{limit}}(bg, eh)$ is the elevated threshold due to the contrast masking effect, and $K(bg)$ is the fitting function for the elevating factor as shown in Fig. 5(b).

D. Binocular Just-Noticeable-Difference Model

By incorporating the models of the binocular combination of injected noises, luminance and contrast masking effects, we have developed a binocular JND ($BJND$) model by combining (6) and (7):

$$\begin{aligned}
 BJND(bg(i), eh(i), A_1(i)) &= A_{C,\text{limit}}(bg(i), eh(i)) \\
 &\cdot \left(1 - \left(\frac{A_1(i)}{A_{C,\text{limit}}(bg(i), eh(i))}\right)^\lambda\right)^{1/\lambda}
 \end{aligned} \quad (9)$$

where $0 \leq A_1(i) \leq A_{C,\text{limit}}(bg(i), eh(i))$, and the $eh(i)$ is calculated by 5×5 Sobel operators [Fig. 5(c)] at region i :

$$eh(i) = \sqrt{E_H^2(i) + E_V^2(i)} \quad (10)$$

$$\begin{aligned}
 E_k(i) &= \frac{1}{24} \sum_{h=1}^5 \sum_{v=1}^5 p(x_i - 3 + h, y_i - 3 + v) \cdot G_k(h, v), \\
 k &= H, V.
 \end{aligned} \quad (11)$$

The proposed model is an integration of luminance masking based BJND profile shown in (6) and the contrast masking model approximated by (7), where $A_{\text{limit}}(bg)$ in (6) is substituted by $A_{C,\text{limit}}(bg, eh)$ in (7). The proposed BJND model measures the perceptible distortion threshold of binocular

vision for stereoscopic images. The 2-D-JND model can be considered as a special case of the BJND model with symmetric threshold of $0.5^{1/\lambda} \times A_{C,\text{limit}}(bg, eh)$ in both views, i.e., let $(A_1/A_{\text{limit}}(bg))^\lambda = (A_{2\text{min}}(bg, A_1)/A_{\text{limit}}(bg))^\lambda = 0.5$ in (5).

III. EXPERIMENTAL RESULTS

Ten stereo images (with disparity maps) from [10] are chosen for the validity test of the proposed BJND model. The image resolutions are around 670×480 . Disparities of the stereo pairs are generally less than 20 pixels. Twenty subjects (six female and 14 male) participated in the subjective evaluation, following a procedure slightly different from the “adjectival categorical judgment methods” in the ITU-R BT.500-11 standard [11] which has been widely used in 2-D JND validation tests [4]. The reference and the test stereo pairs were vertically juxtaposed on the screen, where the top one was the reference pair consisting of two original images and the bottom one was the test pair comprising noise-injected images, as shown in Fig. 1(b). Vertical juxtaposition reduces visual discomfort compared with horizontal juxtaposition, since strong diplopia appears when subjects viewing the image margins. However, the luminance values of the top and bottom pairs are different on the LCD displays with Twisted Nematic (TN) panels. Therefore, we used two Sony G520 CRT displays in this experiment (color temperature 11700 K, ambient luminance 0.17 cd/m^2), and the viewing distance was about four times of the image height (≈ 1.5 monitor height). The digital luminance level was first mapped to the physical luminance [as shown in Fig. 3(d)] presented by the CRT display, and then converted to the 8-bit levels of the LCD display used in the BJND development experiments. The BJND profile was calculated based on the mapped luminance levels.

The noises were injected in two ways, following the 2-D-JND validation schemes in [3][4]. For noise type 1, the left view was divided into 16×16 blocks, and each block i was added with random noises with amplitude of $\alpha \times A_{C,\text{limit}}(bg(i), eh(i))$ (α was randomly selected from 0, 0.382, 0.574, 0.7 and 1). For noise type 2, the left view was noise free. The noise intensity of the corresponding pixels in the right view was determined as the *maximum* tolerable amplitude indicated by the proposed BJND model in (9). Correspondences in the two views were obtained based on the disparity map, assuming HVS matches objects in the two views in a way similar to the stereo matching. Due to occlusion or non-overlapped visual field between the two views, some regions did not have correspondences in the other view, and then we injected random noises of the *maximum* tolerable amplitude, i.e., $[A_{C,\text{limit}}(bg(i), eh(i))]$, in these regions.

Each juxtaposed comparison group was displayed for 8 s, followed by a 3-second gray image. During the break, the subjects were required to give quantitative scores for the test pair. If the two pairs have no difference, they should rank 0 (a strict criterion for unnoticeable difference); otherwise, using the continuous quality comparison scale listed in [4]. The individual scores for each group were averaged. The detection probability (DP, percentage of subjects perceiving difference) is also provided, for which 50% probability is a standard level for JND. The results show that the injected *maximum* tolerable noises based on the BJND model are nearly invisible. The PSNR results of the noise-injected stereo pairs are shown in Table II.

TABLE II
SUBJECTIVE SCORES FOR THE NOISE-INJECTED STEREOSCOPIC IMAGES

Images	Noise 1 - Random				Noise 2 - Right max				
	PSNR		Score		PSNR		Score		
	L	R	Bi	R	Bi	Mo	Bi	Mo	
Aloe	40.4	40.4	-0.05	35	37.2	-0.15	-0.15	35	20
Art	40.9	40.9	-0.05	35	38.0	0.00	-0.35	40	45
Books	40.4	40.2	-0.05	25	37.3	-0.05	-1.00	35	75
Bowling2	41.7	41.7	0.05	45	38.6	-0.05	-0.20	25	35
Cloth2	40.6	40.8	0.05	55	37.3	0.00	-0.25	40	55
Dolls	40.6	40.7	0.20	50	37.6	0.00	-0.40	30	55
Moebius	42.0	42.1	0.00	30	38.7	-0.20	-0.55	20	45
Monopoly	40.8	41.0	-0.10	25	37.8	-0.30	-0.40	30	50
Reindeer	40.9	40.9	0.30	30	37.8	-0.05	-0.15	25	15
Rocks2	41.6	41.7	0.15	35	38.4	0.10	-0.05	30	20

Note: “Bi” means watching left and right views via the mirror system (binocular vision in charge); “Mo” means watching the right view with both eyes (monocular vision in charge). “DP” denotes the detection probability.

In addition, we conducted another experiment, in which subjects viewed the right view distorted by noise type 2 with both eyes. In this case, most subjects reported that the injected noises were more perceptible, corresponding to generally lower scores and higher detection probabilities, as shown in the “Mo” columns of Table II. The results suggested that the injected noises determined by the BJND model were “monocularly” visible but binocularly invisible.

IV. CONCLUSION

In this letter, we propose a binocular just-noticeable-difference (BJND) model, which is developed based on two psychophysical experiments by modeling and incorporating the binocular combination of injected noises, luminance masking and contrast masking in binocular viewing. To the best of our knowledge, this is the first model to measure the perceptible distortion threshold of stereoscopic images for binocular vision. Subjective evaluations on stereoscopic images injected with noises associated with the BJND model demonstrate the validity of the proposed model. The impact of disparity on the BJND model will be considered in the future work.

REFERENCES

- [1] L. B. Stelmach and W. J. Tam, “Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views,” *Signal Process.: Image Commun.*, vol. 14, pp. 111–117, Nov. 1998.
- [2] N. Jayant, J. Johnston, and R. Safranek, “Signal compression based on models of human perception,” *Proc. IEEE*, vol. 81, pp. 1385–1422, Oct. 1993.
- [3] 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, pp. 143–156, Apr. 1996.
- [4] 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, pp. 806–819, Jun. 2010.
- [5] S. H. Schwartz, *Visual Perception: A Clinical Orientation*, Fourth ed. New York: McGraw-Hill Medical, 2009, ch. 11.
- [6] W. J. Levelt, “Binocular brightness averaging and contour information,” *Brit. J. Psychol.*, vol. 56, pp. 1–13, Feb. 1965.
- [7] A. I. Cogan, “Monocular sensitivity during binocular viewing,” *Vis. Res.*, vol. 22, pp. 1–16, 1982.
- [8] G. E. Legge and J. M. Foley, “Contrast masking in human vision,” *J. Opt. Soc. Amer.*, vol. 70, pp. 1458–1471, Dec. 1980.
- [9] A. N. Netravali and B. G. Haskell, “Visual Psychophysics,” in *Digital Pictures: Representation and Compression*. New York: Plenum, 1988, ch. 4.
- [10] *Middlebury Stereo Vision*, [Online]. Available: <http://vision.middlebury.edu/stereo>
- [11] Methodology for the Subjective Assessment of the Quality of Television Pictures ITU, Geneva, Switzerland, 2002, ITU-R Rec. BT.500-11, 6.2.4.1.
- [12] B. W. Keelan, “Just noticeable differences,” in *Handbook of Image Quality: Characterization and Prediction*. New York: Marcel Dekker, 2002, ch. 3.