https://doi.org/10.2352/ISSN.2470-1173.2017.12.IQSP-243


# Towards Foveated Just Noticeable Difference Modeling for Virtual Reality

***Yuqiao Deng, Yingxue Zhang, Daiqin Yang, Zhenzhong Chen; School of Remote Sensing and Information Engineering, Wuhan University; Wuhan, Hubei, China***

## Abstract

*For the virtual reality (VR) applications, it is important to quantify the just-noticeable-difference (JND) profile of the image such that immersive experience could be provided. In this paper, we first propose a test plan of subjective quality assessment for VR image, then a foveated JND (FJND) model that is capable of modeling the basic vision properties to quantify the correlation between eccentricity and conventional JND. The psychophysical experiments discover the relationship among the relevant factors and quantify the foveated JND that could be used for immersive VR applications.*

## Introduction

Just-noticeable-difference (JND) describes the minimum amount that stimulus intensity must be adjusted to produce a noticeable variation in sensory experience [1]. Due to the sensitivity of human eye and the masking properties, the changes below the JND threshold around a pixel cannot be sensed by the human visual system (HVS). [2]. Under this circumstance, the JND model can be utilized to improve the video transmission by ignoring the distortions below the threshold. Therefore, the model is widely utilized in video and image compression, video retargeting, and other related domains to improve transmission quality and efficiency [3, 4].

Conventional JND measures the visibility threshold under the assumption that visual acuity remains constant over the whole image. Physiologically, the HVS is not uniformly sensitive in that the fovea on the retina owns the highest density of cones, which leads to a highest resolution of perception. The density of cone and retinal ganglion cells decreased gradually as the eccentricity increases, which results in descending visual acuity. Therefore, the foveated just-noticeable-difference (FJND) model is proposed in [5, 6] to improve the JND model.

As related hardware and technology develop, virtual reality (VR) draws more attention recently. It provides immersive viewing experience for viewers with the head-mounted devices (HMD-s) displaying stereoscopic, panoramic videos, 3D, high resolution audio and natural interactions. Concerning the features of the immersive video, conventional video comprehension method may not do well in this category. So we call for new perception of immersive scene and its impact to human vision system.

VR provides simulated and immersive environment, experience and interaction for users by displaying immersive videos or images in head-mounted displays(HMDs), which differs greatly from the normal images. Therefore, a targeted test plan is needed for quality assessment of VR contents.

In this paper, we first propose a test plan of subjective quality assessment for VR images. With the plan, we conduct experiments to explore the relationship between eccentricity and contrast masking effect and build a refined FJND model. Besides, using the virtual reality HMD rather than the monitor to display the virtual scene ideally solves the problem of fixing subjects' distance to the screen and therefore obtains more precise results in measuring the FJND threshold.

The rest of the paper is organized as follows. Section II describes the subjective test plan for VR. Section III verifies the FJND model by psychological experiments using VR HMD.

## Test Plan of Subjective Quality Assessment for VR Images

VR provides immersive environment, experience and interaction for users by displaying immersive scenes in head-mounted displays(HMDs). To evaluate and compare the quality of the VR images generated in different conditions, a standard test plan for subjective quality assessment is necessary. In this section, we describe a targeted test plan for subjective assessment of VR image quality based on [18].

### *Subjective Test Environment*

Differing from the normal quality assessment, the VR image should be displayed with the VR HMDs, e.g., Oculus rift, HTC VIVE. With an HMD, the rating has to be completed by interactions through the screen of the HMD and the stimuli should be displayed in a single-stimulus way. Since the assessment experiments with different aims require varied levels of rating precision, the rating scale can be either five-grade [16] or one-hundred-grade [19] according to the certain aim of the test, which can be qualitatively divided into five degrees, i.e., "Bad","Poor","Fair","Good", and "Excellent".

### *Display Control*

For VR applications, panoramic images shown in the VR H-MDs will be displayed in the sphere format, which means that the contents exist in all directions. It makes up the ultimate characteristic of VR images in the subjective assessment test. Therefore, we suggest that the subjects could move their heads freely to view the images, which guarantees a thorough viewing of the content.

### *Subject Screening*

Besides the common test conditions for subject screening described by [16, 17] such as :

- Normal far vision and near vision with or without correction,
- Normal color vision,

- The subjects have not participated in any subjective image or video quality assessment within a short period of time [17],

the following issues should be screened on vision acuity before the test:

- Normal stereopsis for 3D images
- No severe sickness with VR HMDs

### Subjective Assessment Procedure

The subjective assessment procedure is shown in Fig. 1, which consists of three phases: The instruction session, training session and test session(s). Due to the influence of viewing fatigue or sickness, breaks should be properly arranged throughout the entire procedure.

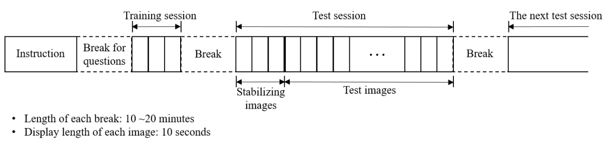


***Figure 1.*** *the subjective assessment procedure.*

Before the assessment procedure starts, a detailed instruction on the test should be given to subjects. The aim of the instruction is to make the subjects know clearly what to do and how to operate in different stages during the test and thus ensure a valid process. After the instruction, all the questions from the subjects should be answered to avoid any misunderstanding during the test.

For subjective image quality assessment, the subjects have watching experience in their daily life. They can judge the quality on the basis of their experience. But for VR, non-expert subjects hardly have the experience on VR viewing, so that an extra training session will be necessary.

During the training session, a set of representative panoramic images should be displayed to the subjects for voting. The content of the training images must be different from the test ones. The quality of training images should cover the entire range of test quality. After the training session, the subjects are expected to become acquainted with the assessment process and quality range of the test.

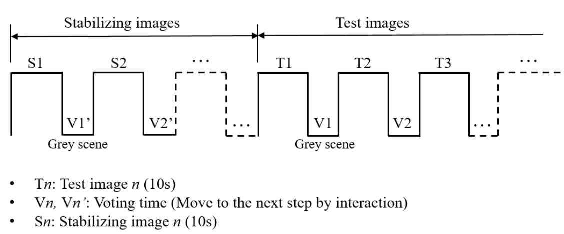


***Figure 2.*** *Test session presentation structure.*

As is shown in Fig. 2, at the beginning of the test session, three to five stabilizing images ($S_i$) should be presented to stablize the sujects rating [16]. The rating of these images will not be included in the final result. During the main part of the session, a test image ($T_i$) will be presented for same time, e.g., 10 seconds, then a grey scene will be displayed with an obvious sign to remind the subjects to vote for the image ($V_i$). Once the score for an image is determined, it is not allowed to be changed and the next test image will then be presented till the test session ends. It is noted that due to the usage of HMD, the scores should be displayed in HMD and the subject need to interact with the input device such as mouse to control the scores.

## Foveated JND Modeling for VR

### Background information

When an image is displayed to us, the visual stimulus projected on the fovea where exists highest density of sensor cones will be perceived at the highest resolution. The density of cone and retinal ganglion cells decreased gradually as the eccentricity increases, which results in descending visual acuity [5]. Thus a contrast masking model incorporating the foveation effect will lead to a better realization for HVS to discriminate image differences. To explore the relationship between visual acuity and eccentricity, we conduct the experiment in laboratory to obtain the visibility threshold for further analysis.

The experiments for FJND should be conducted on the premise that the subjects must always fix on the center of the pattern, which guarantees a constant eccentricity for each pattern during test. Therefore, the image pattern should be adapted to the changed view point when the subject has head movement.

### Experimental Setup

#### Apparatus

In our experiment, instead of displaying images on monitors screen, we build a virtual scene to get more precise results. And we choose Oculus Rift DK2 to display our virtual scene. Presenting the immersive scene with a VR HMD makes it easier to fix subject distance and more reliable.

Oculus VR (Irvine, CA) introduced an immersive 3-D virtual reality head-mounted display (HMD), the Oculus Rift, in late 2012. Oculus Rift Development Kit 2 from Oculus VR provides a higher-resolution (960 * 1080 per eye), lower-persistence OLED display and higher refresh rate. With Oculus Rift DK2, we can build a virtual experimental scene with better immersion experience, which will make the model more convincing.

#### Data Transformation

Since CRT monitors were used in the experiments of earlier JND models, whose Gamma characteristic differs from that of OLED monitors we use here, we implement corresponding mapping function [14] to calibrate digital luminance of different type of monitors. The calculated noise threshold (in digital form) is first mapped to physical luminance and then converted to the digital luminance level of the LCD display.

### Modeling Contrast Masking and Foveation Effect

Based on the aforementioned background, an image pattern exploring the correlation of contrast masking effect with retina eccentricity is designed, following the previous work [5]. The test image with predefined contrast $eh$, background $bg$ and eccentricity $e$, is shown at a fixed depth level. As is shown in Fig. 3, the pattern consists of a circular area whose radius is calculated with viewing distance and eccentricity, and some small squares around the circular boundary. The luminance outside and inside the circular boundary is set to $bg$ and $bg - eh$, respectively, to guarantee a constant contrast on the boundary. Therefore, within the small square, the luminance of pixels outside the boundary is either $bg+A$ or $bg-A$, while that of pixels inside the boundary is either $bg-eh+A$ or $bg-eh-A$. In both cases the luminance is limited within the minimum and maximum luminance of 0 and 255. The center of the pattern is designed as a fixation point by setting its luminance to 0 or 255 to help the subjects fixing their attention.

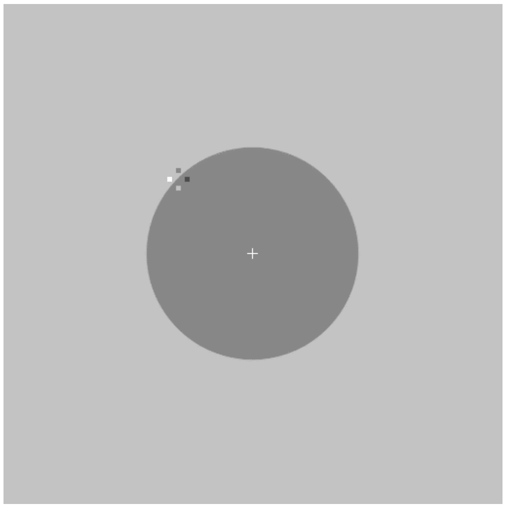

***Figure 3.*** *Image pattern used in experiments exploring contrast masking effect.*

To obtain the visibility threshold distribution in immersive scene due to eccentricity and contrast masking, the test is conducted at a fixed depth level. At the beginning of the test, the noise is invisible to subjects, which means setting the amplitude to zero. With the attention fixing on the center point, the subjects are first asked to increase the amplitude by 1 each time until the noise area becomes visible and the corresponding amplitude will be recorded as $A_1$. The amplitude is then set to $A_1+\Delta A$ being slightly higher than $A_1$. The subjects conduct a reversed procedure, i.e., decreasing the amplitude from $A_1+\Delta A$ till the noise becomes invisible and the corresponding amplitude will be recorded as $A_2$. The procedure, as shown in Fig. 4, could be repeated several times to obtain sufficient sample results $A_i$. The noise visibility threshold under the experimental scene will be represented by the average of $A_i$.

As shown in Fig. 5, the visibility threshold increases with contrast increasing. This could be explained by existing Spatial JND model, which shows a linear relationship between visibility threshold and contrast. The visibility threshold increases more rapidly as eccentricity increases. These results indicate the sharp drop of visual acuity due to the increasing distance between image pixel and fixation point.

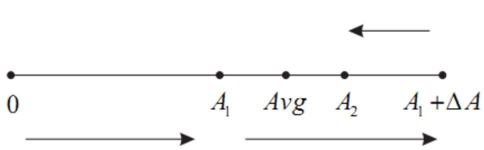


***Figure 4.*** *The procedure of the test. $Avg$ shows the average value of $A_i$.*

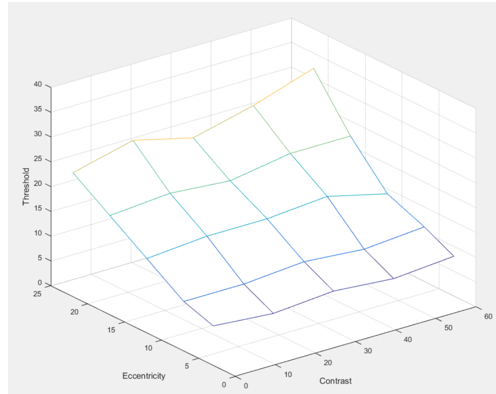


***Figure 5.*** *Experimental results due to contrast and eccentricity.*

## Foveated JND Model

Based on the correlation, we build a model to estimate the relationship between contrast masking effect and foveation effect in immersive scene quantitatively. To estimate the contrast masking effect, a Spatial JND model is adopted in this paper. The contrast masking effect is described in [7] as:

$$f_1(x,y) = [0.01B(x,y)+11.5][0.01G(x,y)-1]-12 \tag{1}$$

where $f_1$ estimates the spatial masking effect, $B(x,y)$ and $G(x,y)$ represent the background and contrast of point $(x,y)$. The retinal eccentricity of a certain point $P(x,y)$ around the fixation point $(x_f,y_f)$ can be calculated as:

$$e(v,P) = \arctan\left(\frac{d}{v}\right) \tag{2}$$

$$d = \sqrt{(x-y_f)^2-(y-y_f)^2} \tag{3}$$

where $v$ represents the viewing distance. The contrast sensitivity is measured as a function of eccentricity [8]:

$$CS(f,e) = \frac{1}{CT(f,e)} \tag{4}$$

$$CT(f,e) = CT_0 \exp\left(\chi \cdot f \cdot \frac{e+e_2}{e_2}\right) \tag{5}$$

where $CT(f,e)$ denotes contrast threshold. $f$ is the spatial frequency (cycles/degree), $e$ is the retinal eccentricity (degree). The model parameters $CT_0$, $\chi$, $e_2$ are set to $1/64$, 0.106, and 2.3, respectively, representing the minimum contrast threshold, the spatial frequency decay constant and the half-resolution eccentricity constant. To obtain the normalized contrast sensitivity:

$$S_f = \frac{f_m(v,P)}{f_m(v,0)} \tag{6}$$

We define the cutoff frequency for any given location $P$ as:

$$f_m(v,P) = \min(f_c(e(v,P)), f_d(v)) \tag{7}$$

where the cutoff frequency $f_c$ can be obtained by setting contrast threshold $CT$ to zero:

$$f_c(e) = \frac{e_2 \ln\left(\frac{1}{CT_0}\right)}{\chi(e+e_2)} \tag{8}$$

The display cutoff frequency $f_d(v)$ is set to half of the display resolution $r$ [9]:

$$f_d(v) = \frac{r}{2} \approx \frac{1}{2} \cdot \frac{\pi v}{180} \tag{9}$$

Obtaining $S_f$, the contrast masking effect and foveation effect is modeled using $S_f$ and $eh$. For each visibility threshold, the corresponding masking effect $f_1$ is first calculated with its background and contrast. The ratio between experimental threshold and the corresponding masking effect is calculated to measure the foveation effect. Concerning that the ratio varies under different background, the mean ratio is introduced here and fitted as a function of contrast and eccentricity. The foveation model of contrast masking effect is defined as:

$$m_1(S_f, eh) = (\frac{1+a}{S_f+a})^{(\log(\frac{eh}{255}+1)+b)^2+c} \tag{10}$$

where $S_f$ is the normalized contrast sensitivity, $eh$ measures the contrast of each pixel in the image. $a$, $b$ and $c$ are set to -0.086, 0.078 and 0.204, respectively.

Finally, we define the foveated contrast masking effect model on the basis of the established models above:

$$F_1 = f_1 \cdot m_1 \tag{11}$$

$f_1$ measures the contrast masking effect (1), while $m_1$ the foveation effect of contrast masking (10).

With the proposed model, the perceptible distortion threshold when the visual attention is fixed on a certain location of image can be measured quantitatively.

## Conclusion and Discussion

In this paper, we propose a targeted test plan of subjective quality assessment for VR image. Based on the plan, a refined FJND model for VR is developed by psychological experiments with VR HMDs. The improved JND model is utilized to model the relationship between eccentricity and contrast masking effects. The psychophysical experiments discover the relationship among foveation effect and contrast masking effect. With contrast or eccentricity increasing, the visibility threshold also increases. We believe that the correlation discovered in this paper attaches much importance to the immersive VR and could be further used for related applications, e.g., VR image quality assessment, compression and transmission.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 61471273), the National High-tech R&D Program of China (863 Program, 2015AA015903), and the Natural Science Foundation of Hubei Province of China (No. 2015CFA053).

## References

[1] USD Internet Sensation & Perception Laboratory, "Webers law of just noticeable differences," http://apps.usd.edu/coglab/WebersLaw.html.

[2] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385–1422, 1993.

[3] C. K. Chiang, S. F. Wang, Y. L. Chen, and S. H. Lai, "Fast jnd-based video carving with gpu acceleration for real-time video retargeting," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 19, no. 11, pp. 1588–1597, 2009.

[4] J. Kim, S. H. Bae, and M. Kim, "An hevc-compliant perceptual video coding scheme based on jnd models for variable block-sized transform kernels," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 25, no. 11, pp. 1–1, 2015.

[5] Z. Chen and C. Guillemot, "Perceptually-friendly h.264/avc video coding based on foveated just-noticeable-distortion model," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 20, no. 6, pp. 806–819, 2010.

[6] Y. Li, H. Liu, and Z. Chen, "Perceptually-lossless image coding based on foveated-jnd and h.265/hevc," *Journal of Visual Communication & Image Representation*, vol. 40, pp. 600–610, 2016.

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

[8] W. S. Geisler, "Real-time foveated multiresolution system for low-bandwidth video communication," *Proceedings of SPIE*, vol. 3299, pp. 294–305, 1998.

[9] Z. Wang, L. Lu, and A. C. Bovik, "Foveation scalable video coding with automatic fixation selection," *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 243–254, 2003.

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

[11] 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 Transactions on Circuits & Systems for Video Technology*, vol. 20, no. 11, pp. 1648–1652, 2010.

[12] C. H. Chou, Y. C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 5, no. 6, pp. 467–476, 1996.

[13] S. H. Schwartz, "Visual perception: a clinical orientation," *Medical*, 1994.

[14] Y. Zhao, Z. Chen, C. Zhu, Y. Tan, "Binocular Just-Noticeable-Difference Model for Stereoscopic Images," *Signal Processing Letters IEEE*, vol. 18, no. 1, pp. 19–22, 2011.

[15] C. H. Chou, C. W. Chen, "A perceptually optimized 3-D subband codec for video communication over wireless channels," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 6, no. 2, pp. 143–156, 1996.

[16] Int. Telecommun. Union, Methodology for the Subjective Assessment of the Quality of Television Pictures ITU-R Recommendation BT.500-13, Tech. Rep., 2012.

[17] Video Quality Experts Group (VQEG), "Test plan for evaluation of video quality models for use with stereoscopic three-dimensional television content," 2012.

[18] Y. Zhang, Y. Wang, Z. Liu, Z. Chen et al., "Test plan for subjective assessment of VR video quality," *Video Quality Experts Group (VQEG)*, 2016.

[19] H. R. Sheikh, M. F. Sabir, A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.