Joint Just Noticeable Difference Model Based on Depth Perception for Stereoscopic Images

Xiaoming Li[1], Yue Wang[2], Debin Zhao[1]
[1]Department of Computer Science and Technology
Harbin Institute of Technology
Harbin, 150001, China
[2]Graduate University
Chinese Academy of Sciences
Beijing, 100080, China
{xmli, wangyue, dbzhao}@jdl.ac.cn

Tingting Jiang[3], Nan Zhang[4]
[3] National Engineering Lab. for Video Technology
Key Lab of Machine Perception (MOE), School of EECS
Peking University
Beijing, 100871 China
ttjiang@pku.edu.cn
[4]Capital Medical University
Beijing, 100069, China
zhangnan@ccmu.edu.cn

***Abstract*—Just noticeable difference (JND) model can reflect the least perceptible distortion from images, including 2D images and stereoscopic images. As we know, for the perception of human visual system (HVS), stereoscopic images have quite different characteristics from 2D images, since stereoscopic images contain not only planar information, but also depth information. This paper proposes a joint JND (JJND) model based on depth perception for stereoscopic images. Firstly, disparity estimation is performed in order to decompose the image into the occlusion region and the non-overlapped region. Then, different JND thresholds are applied on different regions, according to the depth information of the region, which can be derived from the disparity of the region. Experimental results verified our model's validity for stereoscopic images.**

***Index Terms*—** Just noticeable difference, stereoscopic images, human visual system, disparity

I. INTRODUCTION

Stereoscopic images have been applied in many areas, from entertainment programs, such as 3D movie and stereoscopic TV, to professional applications, including medical and space area. In consequence, much work has been done for stereoscopic images and videos. ISO/IEC JTC 1 Moving Picture Experts Group (MPEG) established a 3DAV (3D Audio-Visual) group in 2001 [1][2], and Joint Video Team (JVT) of ITU-T Video Coding Experts Group (VCEG) and MPEG developed a new standard for multiview video coding (MVC), which promoted the application and the development of stereoscopic image techniques. In order to improve the video coding performance, the redundant information in stereoscopic image pair should be removed because it cannot be perceived by human. Meanwhile, human have different perceiving ability of stereoscopic images compared to 2D images. Therefore, it is meaningful to develop a novel JND model for stereoscopic images.

Conventional 2D JND models such as [3] and [4] can be extended to stereoscopic images assessment directly. [3] advised a JND estimator in the image domain, namely nonlinear additivity model for masking(NAMM), integrating spatial masking factors, luminance adaptation and texture masking. As an enhanced model of NAMM, [4] proposed a new contrast masking estimation algorithm, which splits the image into the structural image and the textural image, and applies different weights for the two images.

Both of the two methods mentioned above are reasonable for the isolated right or left view of the stereoscopic images respectively. However, there are some differences between the 2D images and the stereoscopic images. On one hand, much image content appears in both the right and the left views, so that the distortion in one view can be compensated or masked by the other view, which is called binocular suppression theory [5][6]. Considering stereoscopic images' unique properties, such as binocular combination and rivalry, [7] proposed a binocular JND (BJND) model based on psychophysical experiments. On the other hand, stereoscopic images can bring people depth perception, which is different from the conventional 2D images. [8] derived a mathematical model to explain the just noticeable difference in depth, which inspired our work. However, we aimed at least perceptible distortion from images with the depth factor.

We believe that HVS has different perceptions to the objects with different depths. Based on this idea, this paper proposes a joint JND model based on depth perception for stereoscopic images. Firstly, disparity estimation is performed. After image matching by the generated disparity maps, the image is decomposed into two components, the occlusion region and the non-overlapped region. For the non-overlapped region, the depth of the object is derived from the disparity value. Then different JND models for the occlusion region and the non-overlapped region are built, where objects with different depths have different JND thresholds in the

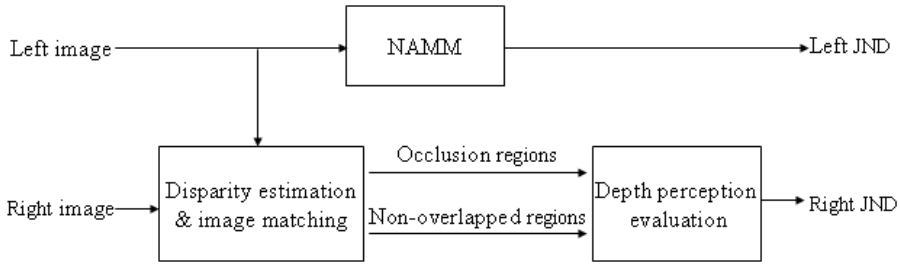

Figure 1. The framework of the proposed joint JND model.

non-overlapped image.

The rest of this paper is organized as follows. Section 2 describes the proposed JND model based on depth perception in detail. Section 3 presents the experimental results, and conclusions are given in Section 4.

II. DEPTH PERCEPTION BASED JOINT JND MODEL

For the conventional 2D images, Yang *et al.* proposed a classic JND model, named NAMM, in which background luminance adaptation and texture masking are two factors that determine the JND threshold of the image. The formulation of NAMM is described as follows [3]:

$$JND^{Y}(x,y) = T^{l}(x,y) + T^{t}(x,y) - C \times \min\{T^{l}(x,y), T^{t}(x,y)\}$$

$$T^{l}(x,y) = \begin{cases} 17\left(1 - \sqrt{\frac{\bar{I}_Y(x,y)}{127}}\right) + 3, & \text{if } \bar{I}_Y(x,y) \le 127 \\ \frac{3}{128}(\bar{I}_Y(x,y) - 127) + 3, & \text{otherwise} \end{cases} \quad (1)$$

$$T^{t}(x,y) = \beta_\theta \times G(x,y) \times W(x,y)$$

where T^l and T^t are the visibility thresholds determined by the background luminance adaptation factor and the texture masking factor, and C is a constant in [0,1], representing the overlapping effect between T^l and T^t. $\bar{I}_Y(x,y)$ is the average luminance of the region centered around (x, y). $G(x, y)$ and $W(x, y)$ denote the maximal weighted average of the gradient around the pixel at (x, y) and the edge-related weight of the pixel respectively, and β_θ is a control parameter, which is set 0.117 in this paper. The detailed descriptions can be found in [3].

It is reasonable that for stereoscopic images, the factors that determine the 2D images can work partly, but the human visual perception characters, especially the depth perception of stereoscopic images should also be considered for the JND model. Therefore, in this section, a joint JND model is presented in order to reflect the depth perception, as shown in Fig. 1. Usually, stereoscopic images comprise two viewpoints, the right image and the left image, and in some formats one of them is a synthesized image. In some applications, the right image is the enhancement layer to the left image. So in this paper, the JND threshold of the left image is calculated as Yang *et al.*'s method. For the right view, disparity estimation and image matching are performed firstly, in order to decompose the image into the occlusion region and the non-overlapped region, and then the two regions are distinguished by different weights according to the different depth perceptions.

A. Disparity estimation and image matching

Disparity maps obtained by disparity estimation are used for image matching and depth deriving. Many computer vision algorithms such as [9] can be used to generate the disparity maps. In this paper, we adopt the pixel based weighted block matching method, which is an extended script of [10]. For a pixel (u, v) in the right view, the disparity value $D_{lr}(u, v)$ is obtained as follows:

$$D_{lr}(u,v) = \arg\min_{(s,t)\in S}(\mathrm{Reg}(s,t) + \mathrm{WSAD}(s,t))$$

$$\mathrm{Reg}(s,t) = \gamma(|D_{lr}(u-1,v) - (s,t)| + |D_{lr}(u,v-1) - (s,t)|) \quad (2)$$

$$\mathrm{WSAD}(s,t) = \sum_{\substack{-M\le i\le M \\ -N\le j\le N}} W(i,j)\,|I_r(u+i,v+j) - I_l(u+s+i,v+t+j)|$$

where S is the search range, which is set as 64 x 6 rectangle region centered around the coordinate (0, 0). Reg(s, t) is the regularity item, which is used to smooth the disparity map, and γ is set 0.55 in this paper. WSAD means weighted summary absolute difference of blocks in the right and the left image. The size of the matching block is $(2\times M+1)\times(2\times N+1)$, and the value of weight matrix $W(i, j)$ is determined by the distance from the pixel (i, j) to the block's center. And similarly, the disparity of the left image is represented as D_{rl}. Next, we use cross checking method [11] to decompose the image into the occlusion region and the non-overlapped region, where the occlusion region is defined as the set of pixels in the right image whose corresponding pixels can not be found in the left view, and formally,

$$R_{occlusion\_r}(u,v) = \begin{cases} 0 \text{ if } |(D_{rl}(u',v') + (u',v')) - (u,v)| < \delta \\ 1 \text{ otherwise} \end{cases}, \quad (3)$$

$$(u',v') = (u,v) + D_{lr}(u,v)$$

where $R_{occlusion\_r}(u, v)=1$ means the pixel (u, v) in the right image is in the occlusion region, and otherwise in the non-

overlapped region. δ is a toleration parameter for matching inaccuracy, and is set as 3 in this paper.

B. Depth perception based joint JND model

Based on the idea that human has different visual perception for the objects with different depths, the JND thresholds of the occlusion region and the non-overlapped region are distinguished by the proposed joint JND model. Because the occlusion region often appears at the edges of objects with different depths, which brings human stronger depth perception and the minimum noticeable distortion in this region should be lower than that in the isolated images. On the other hand, for the non-overlapped region, the distortion of the region in the right image can be composed from the corresponding region in the left image, which means the masking effect of this region is stronger than that in the isolated images. In addition, considering the human experience that the objects with larger depth are less attractive than the nearby objects, the masking effect of the non-overlapped region should be higher when the corresponding depth becomes larger. Based on the above observations, our proposed joint JND model of the right image is expressed as:

$$JJND(x,y)=\begin{cases} JND^{Y}(x,y)\cdot\alpha & \text{if } R_{occlusion}(x,y)=1 \\ JND^{Y}(x,y)\cdot\beta(x,y) & \text{otherwise} \end{cases}, \quad (4)$$

where $JND^{Y}(x, y)$ means the JND value of (x, y) obtained by Yang *et al.'s* method. α is the depth perception parameter which is set as 0.8 in this paper. $\beta(x, y)$ is the joint masking effect parameter, which is determined by the depth of (x, y), $d(x, y)$. Considering the limitation of human perception ability to the depth, we divide the depths appearing in the un-overlapped region into 5 depth levels uniformly according to their values, and assign $\beta(x, y)$ from the parameter set **JMEPS** (joint masking effect parameter set) according to $d(x, y)$, instead of building a projection relation from $d(x, y)$ to $\beta(x, y)$. Concretely, $\beta(x, y)$ is assigned as JMEPS[i], if $d(x, y)$ belongs to the ith depth level. In this paper **JMEPS** is set as {1.1, 1.2, 1.3, 1.4, 1.5}. Since depth can be derived from the disparity of the corresponding pixels in the image [12][13], the disparity maps obtained before are reused to determine pixels' depth level.

III. Experimental Results

In order to evaluate our model, we compare our model to model in [3]. Firstly, noise was injected into the right and the left images as follow:

$$I^{JND}(x,y)=I(x,y)+s_{rand}(x,y)JND^{m}(x,y), \quad (5)$$

where $I(x, y)$ represents the illumination of pixel (x, y), the value of $s_{rand}(x, y)$ is either +1 or -1 randomly, and $JND^{m}(x, y)$ means the JND threshold calculated by [3] or by JJND method. Ten subjects were asked to compare the quality of the stereoscopic images distorted by two models with shutter glasses. We used VSCOE23 LCD with NVIDA GeForce GTS 450 display card. The stereoscopic 3D experience was rendered by NVIDIA's 3D Vision, including wireless glasses and an IR emitter. Test images are shown in Table I. Because the real resolution of the display is 1680x1050, the images such as *AltMoabit*, *BookArrival* and *DoorFlowers* [14] with resolution 1024x768 were expanded to 1680x1050 with black margin, while other images, *Poznan_CarPark* and *Poznan_Street* [15], were cut to 1680x1050 by margin pruning. The stereoscopic images from the same scene injected with noise by different model were displayed in turn randomly, and subjects gave scores by ITU-R BT.500-11 standard [16], as shown in Table II. In this table, a negative value means the second stereo image is better. Because subjects do not know whether the stereo image is processed by [3] or by JJND, the average scores should adjusted by the play turn, which are list in Table III. In addition, we performed another experiment of comparison between the 2D right images injected with the same noise as before. The experimental results including average scores and the detection probabilities (DP) [4][7] are given in Table III, with the PSNR of each image, in which DP reflects the probability that subjects can detect the difference of the two injected method. In the "Score" columns, a negative value means the image (stereoscopic or 2D) injected with noise by JJND model has higher subjective quality than that injected by [3].

From Table III it can been seen that the stereoscopic images injected by two model have nearly the same

TABLE I. Test Images for Comparison

Image	Resolution	View point	
		L	***R***
AltMoabit	1024x768	10	9
BookArrival	1024x768	10	9
DoorFlowers	1024x768	4	3
Poznan_CarPark	1920x1088	4	3
Poznan_Street	1920x1088	4	3

TABLE II. Comparison Scale for Subjective Quality

Score	perceptible differences
-3	Much worse
-2	Worse
-1	Slightly worse
0	The same
+1	Slightly better
+2	Better
+3	Much better

TABLE III. Comparison Results of Subjective Quality Between [3] and JJND

Image	PSNR(dB)			Stereoscopic		2D	
	L	***R***		***Score***	***DP***	***Score***	***DP***
		[3]	***JJND***				
AltMoabit	32.87	32.81	31.29	0.5	70	0.6	90
BookArrival	31.19	31.30	30.45	0.0	50	-0.1	60
DoorFlowers	31.64	31.62	30.80	-0.2	40	0.3	70
Poznan_ CarPark	32.42	32.34	31.25	-0.3	70	1.0	80
Poznan_ Street	34.48	34.44	33.47	0.1	70	0.6	60

subjective quality averagely, while the PSNRs of the right images have nearly 1 dB difference between the two models, which means the JJND model is more effective than [3] for stereoscopic images. The subjective quality evaluation of 2D images proves the difference of subjective quality existing in right images and it can be concluded that the visible differences in isolated images is invisible in stereoscopic images. Furthermore, the values of DP verify the experimental results. However, Table III also shows that for image *AltMoabit*, JJND does not work so well. It is because the average disparity value of *AltMoabit* is the lowest among all the test images, that is, the weakest depth perception existing in this image. So, the JJND model does not entertain such image well since it is based on the depth perception.

IV. CONCLUSIONS

This paper proposed a depth perception based joint JND model for stereoscopic images, based on the idea that human has different visual perception for the objects with different depths. Firstly, disparity estimation is performed in order to decompose the image into the occlusion region and the non-overlapped region. Then, different JND models are proposed for different regions, according to the depth information of the region, which can be derived from the disparity of the region. After subjective quality evaluation experiment for stereoscopic images and 2D images, our model was verified to be valid to stereoscopic images. In the future, the subject quality assessment metric for stereoscopic images and videos should also be exploited

V. ACKNOWLEDGEMENTS

This work was partly supported by the National Science Foundation of China, 60736043， the Major State Basic Research Development Program of China, 973 Program 2009CB320905 and National Science Foundation (60833013).

REFERENCES

[1] “Report on 3DAV exploration,” ISO/IEC JTC/SC29/WG11, N5878, July 2003.

[2] M. Jose, “MPEG 3DAV AhG activities report,” 65th MPEG Meeting, Trondheim, Norway (2003).

[3] X. K. Yang, W. Lin, Z. K. Lu, E. P. Ong, and S. S. Yao, “Just noticeable distortion model and its applications in video coding,” *Signal Processing: Image Commun.*, vol. 20, no. 7, pp. 662–680, 2005.

[4] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, “Just noticeable difference for image with decomposition model for separating edge and texture regions,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 20, No. 11, pp. 1648–1652, Nov. 2010.

[5] L. Stelmach and W. J. Tam, ” Stereoscopic Image Coding: Effect of Disparate Image-Quality in Left- and Right-Eye Views,” *Signal Processing: Image Communication*, Vol. 14, pp. 111-117, 1998.

[6] L. Stelmach, W.J. Tam; D. Meegan, and A. Vincent, “Stereo image quality: effects of mixed spatio-temporal resolution,” *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 10, No. 2, pp. 188-193, Mar 2000.

[7] Y. Zhao, Z. Chen, C. Zhu, Y.Tan, and L. Yu, “Binocular just noticeable-difference model for stereoscopic images,” *IEEE Signal Processing Letters*, Vol. 18, No. 1, pp. 19-22, Jan 2011.

[8] 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,” *IEEE ICME 2010*, pp. 1219-1224, Jul. 2010.

[9] J. Sun, Y. Li, S. Kang, and H. Shum, “Symmetric stereo matching for occlusion handling,” *CVPR*, Vol. II: 399– 406, 2005.

[10] K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, and Y. Yashima, “Multiveiw video coding using view interpolation and color correction,” *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 17, No. 11, pp. 1436-1449, Nov 2007.

[11] G. Egnal and R. Wildes. “Detecting binocular halfocclusions: empirical comparisons of five approaches,” *PAMI*, 24(8):1127–1133, 2002.

[12] X. Li, D. Zhao, X. Ji, Q. Wang, and W. Gao, “A fast inter frame prediction algorithm for multiview video coding,” in *IEEE International Conference on Image Processing*, Sep. 2007.

[13] R. Hartley, A. Zisserman, “Multiple View Geometry in Computer Vision,” Cambridge University Press, UK, 2003.

[14] “HHI Test Material for 3D Video,” ISO/IEC JTC1/SC29/WG11, M15413, April 2008.

[15] “Poznań Multiview Video Test Sequences and Camera Parameters,”. ISO/IEC JTC1/SC29/WG11, M17050, 2009

[16] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU, Document ITU-R BT.500-11, Geneva, Switzerland, 2002.