

3D Visual Attention for Stereoscopic Image Quality Assessment

Qiuping Jiang, Fenfang Duan, Feng Shao*

Faculty of Information Science and Engineering, Ningbo University, 315211, Ningbo, China

Email: shaofeng@nbu.edu.cn

Abstract—Depth perception is one of the most important characteristic in three-dimensional (3D) images different from traditional two-dimensional (2D) images. Therefore, 3D visual attention will be advantageous to improve 3D visual experience and particularly depth perception. In this paper, we propose a 3D visual attention model for stereoscopic image quality assessment task. The proposed model is constructed based on 2D saliency model, center bias, depth cue (foreground cue and background). Different combination and modulation means of the 3D visual attention model for quality assessment are investigated. The experimental results show that compared with other schemes, the proposed 3D visual attention-based pooling scheme can achieve higher consistency with the subjective assessment of stereoscopic images.

Index Terms—3D visual attention, stereoscopic image quality assessment, saliency model, center bias, depth cue

I. INTRODUCTION

Three-dimensional (3D) technologies have received wide attention as a result of great push from the industry and academia^[1,2]. The necessity for designing perceptual 3D image quality assessment (3D-IQA) approach is increasingly important^[3], since such perceptual issues in 3D are hardly considered in the traditional 2D image quality assessment (2D-IQA)^[4]. Following the research of 2D-IQA, 3D-IQA approaches can fall into two categories: subjective assessment and objective assessment. Some publicly available 3D databases were provided, such as LIVE 3D image database^[5], EPFL 3D image database^[6], etc, by adding different types of stimuli (e.g., distortion or camera distance) on left and right images.

In objective assessment, the term ‘quality of experience (QoE)’ should be considered to capture the various factors that contribute to the overall visual experience of the 3D visual signal^[7]. In contrast to the 2D case, QoE of 3D involves not only evaluating 2D image quality, but also additional aspects of quality, e.g., depth perception, visual comfort, and other visual experience. In some cases, the latter factors will be much important. The depth processing is mainly based on identifying relative objective positions and the bottom-up processing^[8]. Therefore, visual attention can be useful to improve 3D QoE and particularly depth perception^[9]. Recently, visual attention for 2D-IQA was explored by integrating visual attention into quality metrics to improve prediction performance, according to the principle of more weight to the distortions appearing on

the saliency areas^[10-11]. However, how to derive the 3D visual attention information for quality assessment is still an open issue.

Many computational models have been proposed to predict the visual attention through saliency map by adopting bottom-up and top-down mechanisms^[12-13]. Compared with 2D services, 3D much emphasizes the visual perception of viewers, while depth cues are the most important features on the understanding of 3D/stereoscopic visual perception. Therefore, it is interesting to investigate whether depth cues can improve the performance of saliency model. Lang *et al.* quantitatively assessed the contribution of depth cues in visual attention in 3D scenes by analyzing the visual saliency information from eye tracking datasets^[14]. Li *et al.* proposed a human-brain-inspired framework for the fusion of the depth cues by analyzing the reliability of the depth cues locally^[15]. Patapova *et al.* learned probabilistic models of various 2D and 3D saliency cues and fused them into a final saliency map^[16]. Zhang *et al.* proposed a bottom-up visual attention model for stereoscopic content by using the depth map as an additional cue, and used a depth-based fusion with the spatial and motion saliency map^[17]. Park *et al.* proposed a 3D visual attention model by combining bottom-up and top-down models for comfortable 3D viewing^[18]. However, the impact of depth cues on stereoscopic perception is not taken into account.

From another perspective, visual perceptual properties were another important clue in 3D-IQA. Maalouf *et al.* computed the cyclopean image from left and right images to simulate the brain perception, and used contrast sensitivity coefficients as the basis of evaluation^[19]. Gorley *et al.* proposed a Stereo Band Limited Contrast (SBLC) algorithm to evaluate the stereoscopic image quality, which accounts for the HVS sensitivity to contrast and luminance changes at regions of high spatial frequencies^[20]. Hwang *et al.* proposed a metric by considering the impact of visual attention, depth variation and stereo distortion prediction, to detect visually significant distortions based on human visual properties^[21]. Wang *et al.* proposed a metric by considering the binocular spatial sensitivity to reflect the binocular fusion and suppression properties^[22]. Bensalma *et al.* proposed a Binocular Energy Quality Metric (BEQM) by modeling the simple cells responsible for the local spatial frequency analysis and the complex cells responsible for the generation of the binocular energy^[23].

However, these methods are simple extensions of the monocular visual properties into the binocular vision, and how these monocular visual properties affect the binocular vision is still not accounted.

In this paper, we examine the 3D visual attention deployment for quality assessment task. More specifically, the goal is to determine whether a significant relationship exists between visual attention and quality assessment. The main contributions of this work are as follows: 1) We construct a 3D visual attention model by applying 2D saliency model, center bias and depth cue (foreground and background); 2) We investigate various combination means of 3D visual attention and different modulation means for quality assessment. The rest of the paper is organized as follows. Section II presents the computational models of 3D visual attention, which helps to motivate the method. Section III presents the proposed quality assessment method. The experimental results are analyzed in Section IV, and finally conclusions are drawn in Section V.

II. COMPUTATIONAL MODELS OF 3D VISUAL ATTENTION

The mechanism of visual attention is an important conception when explaining HVS. Recent researches have found that visual attention is closely relevant to the image perceptual quality^[24]. For 2D image, many bottom-up and top-down visual saliency models have been proposed by investigating various perceptual characteristics^[25]. While for 3D/stereoscopic images, they have additional characteristics, e.g., depth perception, that makes them different with 2D images. In this paper, we apply 2D saliency model, center bias and depth bias (foreground and background cues), to create a 3D visual attention model. The proposed model, S_{3D} , is defined as follows

$$S_{3D} = \{S_{2D}, CB, FM, BM\} \quad (1)$$

where S_{2D} is the 2D saliency model, CB is the center bias information, FM and BM are the foreground and background information for 3D viewing (disparity used for depth perception in this regard), respectively. Fig.1 shows the concept diagram of the proposed 3D visual attention model.

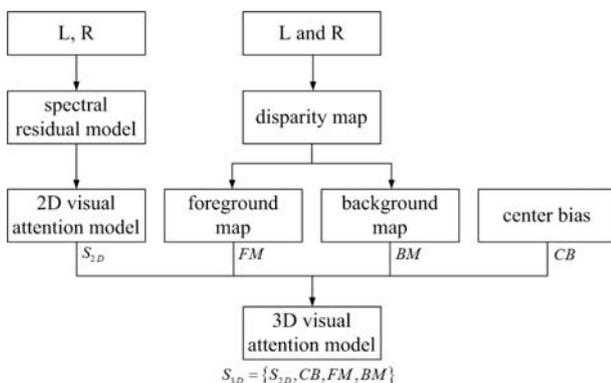


Figure 1. Flow-chart of the proposed 3D visual attention model.

2.1 Model of 2D Saliency

The measure of visual importance of different regions plays an important role in evaluating the image quality. In this paper, we employ the spectral residual model^[26] for saliency calculation. Given an image, we first apply Fourier transform to obtain the amplitude spectrum $A(f)$ and phase spectrum $R(f)$. Then, the spectral residual $P(f)$ can be generated based on the log-spectrum representation $L(f) = \log(A(f))$ according to

$$R(f) = L(f) - L_a(f) \quad (2)$$

where $L_a(f)$ denotes the averaged spectrum, which is obtained by convolving the log-spectrum $L(f)$ with a median filter. Then, the saliency map is obtained by convolving a 2D spectrum map with a 2D Gaussian function

$$SM(x, y) = g(x, y) * (\zeta^{-1}(\exp(R(f) + jP(f))))^2 \quad (3)$$

where $g(x, y)$ is a Gaussian function, ζ^{-1} denotes the inverse Fourier transform.

However, the distortion in image usually hinders accurate saliency feature detection, e.g., some distorted regions will become more salient. To distinguish the effects of the same saliency regions in the original and the distorted images, we define the final 2D saliency map as follows

$$S_{2D}(x, y) = \max(SM_{org}(x, y), SM_{dis}(x, y)) \quad (4)$$

where SM_{org} and SM_{dis} are the saliency maps from the original and the distorted images, respectively,

2.2 Model of the Center Bias

Based on the results of the literatures and subjective experiments, the viewers tend to focus on the central fixation location than other locations^[27]. That is, pixels located near to the center may provide more information than the other pixels, thus be coming more salient. Even though 3D fixation map can be constructed by eye tracking equipment^[28], center bias property is still prominent in 3D viewing. In this paper, central bias is modeled by 2D Gaussian with the strong central fixation distribution on the center and then spreads to the neighbors. The center fixation of the visual field is fixed on the center of the input image

$$(x_c, y_c) = \left(\frac{W_{im}}{2}, \frac{H_{im}}{2} \right) \quad (5)$$

where W_{im} and H_{im} are the width and height of the image. Then, we apply a Gaussian convolution at the center fixation, and calculate the distance to the center for each pixel by

$$CB(x, y) = \exp \left\{ - \left(\frac{(x - x_c)^2}{2\sigma_x^2} + \frac{(y - y_c)^2}{2\sigma_y^2} \right) \right\} \quad (6)$$

where (x_c, y_c) is the center of the image, σ_x^2 and σ_y^2 are the variance along the two directions respectively. In the experiment, σ_x^2 is set to $0.5W_{im}$ and σ_y^2 is set to $0.5H_{im}$.

2.3 Model of the Depth

Stereoscopic contents provide additional depth cues (e.g., occlusion, binocular disparity, accommodation, etc) that are used by humans in the understanding of visual attention in 3D scenes. However, how and to what extent

these depth cues affect the visual attention is still an open issue. In this work, we only consider the case of depth from binocular disparity, since this factor is particularly important in binocular vision. Intuitively, the closes objects will attract more attention than the farthest one. The binocular disparity is estimated from the left and right images by the disparity estimation method^[29]. In order to characterize depth cue, disparity map is first transformed through a sigmoid function. Then, foreground map (FM) and background map (BM) are separated from the disparity map by comparing a given threshold T_1 , i.e., the pixels values higher or smaller than a given threshold are classified into the foreground or background. The combination of the foreground and background maps is call depth model S_{depth} , which is computed as follows:

$$S_{depth} = \begin{cases} FM, & \text{if } f(x,y) > T_1 \\ BM, & \text{otherwise} \end{cases} \quad (7)$$

Here, the threshold T_1 is set to the half the maximum and minimum disparity values through a sigmoid function. Fig.2 shows an example of the result for foreground and background separation.

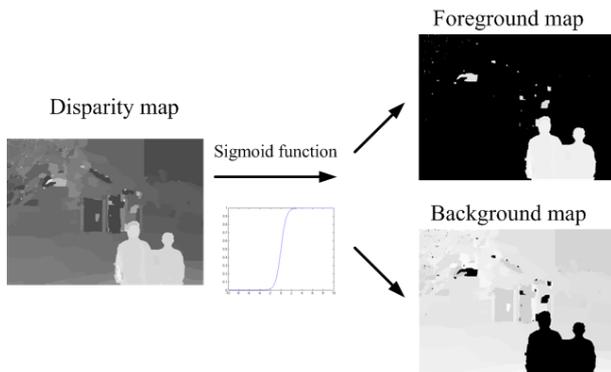


Figure 2. Example of the result for foreground and background separation.

2.4 Proposed 3D visual Attention Model

Since it is hard to simulate the process of combination of various saliency cues in binocular vision, we investigate summation combination to obtain a final saliency map. The final 3D saliency map S_{3D} is equal to the sum of individual cues:

$$S_{3D}(x,y) = \omega_1 \cdot S_{2D}(x,y) + \omega_2 \cdot CB(x,y) + \omega_3 \cdot FM(x,y) + \omega_4 \cdot BM(x,y) \quad (8)$$

where $\sum \omega_i = 1$ and we set the weights adaptively by taking the relative importance of each cue into account. Other combination means will be discussed in the next subsection 4.4. In this work, we train these weighting parameters by optimizing the evaluation results between the objective and subjective scores.

III. PROPOSED QUALITY ASSESSMENT METHOD

3.1 Cyclopean Image Formation

In binocular vision, the images from left and right eyes are combined into a single image, i.e., cyclopean image,

by matching the local regions in the left and right images. However, the quality of stereoscopic image is not a simple combination of the qualities of left and right images. For example, the high-quality view will suppress the low-quality one based on suppression theory of binocular vision^[30]. Therefore, in order to better account for these binocular visual characteristics, we use the energy of the Gabor filter responses to simulate the cyclopean image by locally weighting a stereoscopic pair, because simple cell in the primary visual cortex can be well-modeled using Gabor filter response^[31]. The simulated cyclopean image is expressed as^[32]

$$CI(x,y) = \omega_L \cdot I_L(x,y) + \omega_R \cdot I_R(x-d_L(x,y),y) \quad (9)$$

where $\omega_L = \frac{GE_L(x,y)}{GE_L(x,y) + GE_R(x-d_L(x,y),y)}$,

$\omega_R = \frac{GE_R(x-d_L(x,y),y)}{GE_L(x,y) + GE_R(x-d_L(x,y),y)}$, GE_L and GE_R are

the magnitudes of left and right images respectively summed on all scales and orientations, and $d_L(x,y)$ denote the disparity vector. Considering that disparity estimation algorithm cannot operate well in the distortion images, we use the original disparity to calculate the cyclopean image for the distorted stereoscopic images.

Fig.3 shows the constructed cyclopean images from the original and Gaussian Blur distorted stereoscopic images. As observed from the figures, artificial contours will appear among the boundaries due to occlusion/disocclusion in disparity matching. For example, the background behind the foreground in the left image will be disoccluded in the right image. As a result, monocular vision will occur in these regions. Therefore, in order to accurately account for binocular vision, these regions are excluded in quality assessment, and a full-reference assessment metric is applied on the remaining regions of the original and distorted cyclopean images.



Figure 3. The constructed cyclopean images.

3.2 Proposed Quality Assessment Method

The 3D saliency map $S_{3D}(x,y)$, created by the above computational models, can be used as a modulation function for IQA pooling. Thus, we define saliency modulated the IQA as follows

$$Q = \frac{\sum_{x,y} m_k(x,y) \cdot IQA(x,y)}{\sum_{x,y} m_k(x,y)} \quad (10)$$

where $m_k(x,y)$ is the value of saliency map, $IQA(x,y)$ is the value of local distortion map. In this work, for simplicity, we use Structural SIMilarity (SSIM)^[33] for

TABLE.1.
PERFORMANCE COMPARISON OF DIFFERENT WEIGHTING SCHEMES.

	Weight	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
JPEG	PLCC	0.9479	0.9499	0.9475	0.9376	0.9588	0.9464	0.9491	0.9455
	SROCC	0.9499	0.9508	0.9512	0.9486	0.9579	0.9541	0.9599	0.9570
	KROCC	0.8055	0.8075	0.8115	0.8135	0.8297	0.8196	0.8297	0.8257
	RMSE	4.4218	4.3372	4.4363	4.8248	3.9425	4.4834	4.3700	4.5201
JPEG2000	PLCC	0.8988	0.9549	0.9125	0.9325	0.9116	0.9474	0.9384	0.9555
	SROCC	0.9252	0.9651	0.9431	0.9466	0.9260	0.9572	0.9536	0.9596
	KROCC	0.7657	0.8364	0.7939	0.8020	0.7697	0.8162	0.8061	0.8202
	RMSE	4.8968	3.3175	4.5702	4.0329	4.5917	3.5755	3.8599	3.2950
Gaussian Blur	PLCC	0.9550	0.9789	0.9625	0.9647	0.9472	0.9702	0.9554	0.9663
	SROCC	0.9561	0.9741	0.9632	0.9550	0.9608	0.9646	0.9538	0.9603
	KROCC	0.8317	0.8722	0.8479	0.8297	0.8540	0.8479	0.8196	0.8378
	RMSE	5.8732	4.0434	5.3727	5.2151	6.3519	4.800	5.8479	5.1005
White Noise	PLCC	0.9672	0.9596	0.9653	0.9690	0.9583	0.9652	0.9604	0.9571
	SROCC	0.9648	0.9584	0.9680	0.9735	0.9555	0.9698	0.9673	0.9650
	KROCC	0.8364	0.8283	0.8424	0.8606	0.8202	0.8505	0.8465	0.8424
	RMSE	3.8085	4.2137	3.9104	3.6991	4.2838	3.9214	4.1746	4.3396
H.264	PLCC	0.9231	0.9698	0.9350	0.9448	0.9291	0.9588	0.9419	0.9611
	SROCC	0.9382	0.9542	0.9435	0.9401	0.9281	0.9527	0.9495	0.9536
	KROCC	0.7863	0.8213	0.8045	0.7905	0.7765	0.8269	0.8073	0.8227
	RMSE	4.9097	3.1128	4.5287	4.1833	4.7217	3.6269	4.2885	3.5256
All	PLCC	0.9165	0.9202	0.9227	0.9267	0.9168	0.9301	0.9233	0.9277
	SROCC	0.9264	0.9298	0.9291	0.9307	0.9283	0.9347	0.9304	0.9329
	KROCC	0.7543	0.7679	0.7626	0.7640	0.7619	0.7751	0.7633	0.7725
	RMSE	6.5149	6.3765	6.2771	6.1195	6.5023	5.9830	6.2569	6.0789

measuring the IQA. Besides, in order to demonstrate the impact of saliency map on IQA pooling, we design the following modulation methods by using different saliency maps

$$\begin{cases}
 m_1(x, y) = 1 \\
 m_2(x, y) = S_{2D}(x, y) \\
 m_3(x, y) = (S_{2D}(x, y) + CB(x, y)) / 2 \\
 m_4(x, y) = (S_{2D}(x, y) + FM(x, y)) / 2 \\
 m_5(x, y) = (S_{2D}(x, y) + BM(x, y)) / 2 \\
 m_6(x, y) = S_{3D}(x, y) \\
 m_7(x, y) = S_{3Db}(x, y) \\
 m_8(x, y) = S_{3Dn}(x, y)
 \end{cases} \quad (11)$$

Where $S_{3D}(x, y)$ is the proposed 3D saliency map, detailed explaining about $S_{3D}(x, y)$ are presented in section 4.2. SM_{3Db} is a binarized 3D saliency map,

$$SM_{3Db}(x, y) = \begin{cases} 1, & \text{if } SM_{3D}(x, y) > T_2 \\ 0, & \text{otherwise} \end{cases},$$

$SM_{3Dn}(x, y)$ is a labeled 3D saliency map, $SM_{3Dn}(x, y) = \begin{cases} SM_{3D}(x, y), & \text{if } SM_{3D}(x, y) > T_2 \\ 0, & \text{otherwise} \end{cases}$, and threshold T_2 equals to 14 in the experiment.

IV. EXPERIMENTAL RESULTS

4.1. Stereoscopic Image Quality Database

In the experiment, we have used the database presented in [34]. Twenty-six non-expert adult viewers were participated in the subjective evaluation of the database. According to Double Stimulus Continuous Quality Scale (DSCQS) testing method described in ITU-R recommendation BT.500-11, the subjective ratings for the distorted stereoscopic images were obtained on a scale of 0-100. The database includes 12 original stereoscopic image pairs, from which 312 distorted stereoscopic images are generated with five types of distortion: JPEG, JPEG2000, Gaussian Blur, White Noise and H.264. The symmetric distortions are added on left and right images. More specifically, there are 60, 60, 60,

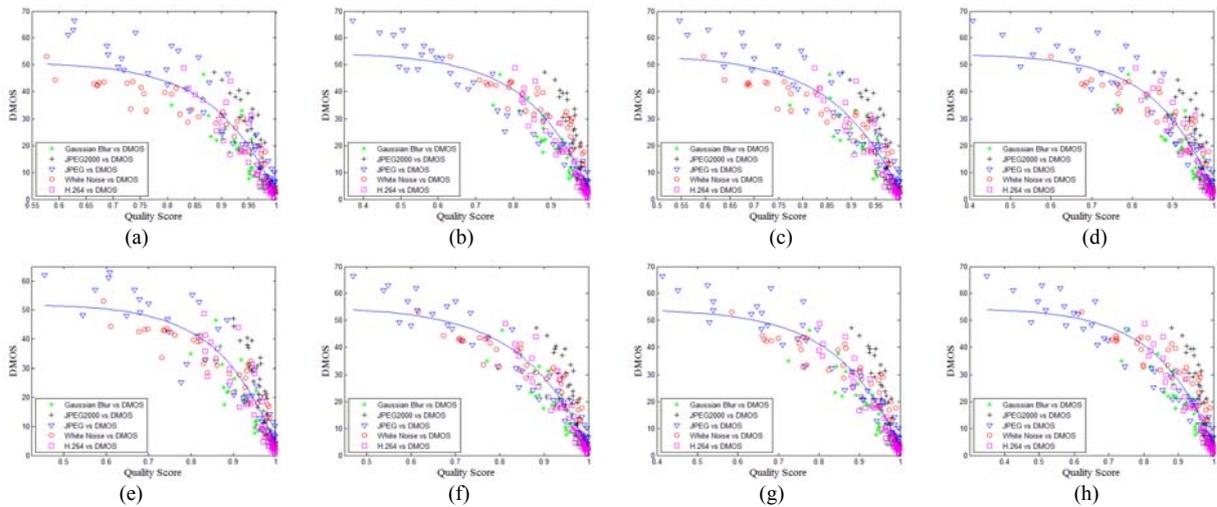


Figure 4. The scatter plots of different weighting schemes: (a) weight m_1 ; (b) weight m_2 ; (c) weight m_3 ; (d) weight m_4 ; (e) weight m_5 ; (f) weight m_6 ; (g) weight m_7 ; (h) weight m_8 .

60 and 72 distorted stereoscopic images in the database with JPEG, JPEG2000, Gaussian Blur, White Noise and H.264 distortions, respectively; there are different distortion levels for each distortion type. The corresponding differential mean opinion score (DMOS) values are provided.

4.2. Performance Determination

To obtain the relationship between the objective scores and the subjective scores, we use the nonlinear regression with four-parameter logistic function by

$$DMOS_p = \frac{\beta_1 - \beta_2}{1 + \exp(-(x - \beta_3) / \beta_4)} + \beta_2 \quad (12)$$

where β_1 , β_2 , β_3 and β_4 are determined by using the subjective scores and the objective scores.

Four commonly used performance indicators are employed to further evaluate the metric: Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), Kendall rank-order correlation coefficient (KROCC), and root mean squared error (RMSE), between the objective scores after nonlinear regression and the subject scores. Among these four criteria, SROCC and KROCC are employed to assess prediction monotonicity, and PLCC and RMSE are used to evaluate prediction accuracy. For a perfect match between the objective and subjective scores, PLCC=SROCC=KROCC=1 and RMSE=0.

In the proposed scheme, we determine the parameters $\omega_i (i=1,2,3,4)$ in Eq.(8) by training to optimize the PLCC values between the objective and subjective scores. In the experiments, we select a subset of the database to train the parameters. For simplicity, the parameters are chosen by linear regression optimization. The parameter determination results are $\omega_1=0.800$, $\omega_2=0.005$, $\omega_3=0.190$ and $\omega_4=0.005$. It is obvious that 2D saliency map component is more important than other components. In the following experiments, the proposed metric is tested on the remaining test sequences in the database (in this way, we avoid same sequences for

training and testing). Thus, totally 234 distorted stereoscopic images are adopted in the evaluation.

4.3. Overall Assessment Performance

The values of PLCC, SROCC, KROCC and RMSE of each distortion type with the database are listed in Table.1. From the table, some observations are given. Firstly, by considering the saliency map modulation, the evaluated results can be significantly improved (from m_2 to m_8), and the best results are obtained by using weighting m_6 for different saliency cues, which suggest that a simple saliency-based pooling is not a good solution to improve the visual quality prediction. Secondly, the 2D saliency weight m_2 seems to be more effective for independent distortion type, but the overall evaluated results are lower than other schemes (e.g., weight with m_2 , m_4 , m_6 , m_7 , m_8). Thirdly, by comparing the evaluated results with weighting m_4 and m_5 , it is obvious that foreground cue of disparity map will dominate the depth perception in quality assessment, which conform to the fact that this phenomenon widely exists in stereoscopic vision. Also, the overall performance of partial region weights (m_7 and m_8) will be decreased even though it may be effective for some individual distortion types. The scatter plots of different weighting schemes are shown in Fig.4.

4.4. Impact of Different Combination Schemes

To demonstrate the impact of different combination of cues, we design the following two schemes for comparison, denoted by Scheme-A and Scheme-B. For Scheme-A, we use multiplication combination method. Noted that, in order to avoid being zero after multiplication, a small constant is added to the foreground and background cues, and all cues are normalized before multiplication. For Scheme-B, we set $\omega_i=0.25$ in Eq.(8). The results of PLCC and SROCC are presented in Table.2. From the tables, we can see that the overall evaluation performance can be gradually promoted by properly weighting the importance of each

TABLE.2.
PLCC AND SROCC COMPARISON FOR DIFFERENT COMBINATION METHODS.

	PLCC			SROCC		
	Scheme-A	Scheme-B	Proposed	Scheme-A	Scheme-B	Proposed
JPEG	0.8672	0.9477	0.9474	0.8962	0.9547	0.9572
JP2K	0.9490	0.9071	0.9464	0.9557	0.9399	0.9541
GB	0.9720	0.9595	0.9702	0.9666	0.9581	0.9646
WN	0.9358	0.9667	0.9652	0.9391	0.9652	0.9698
H.264	0.9559	0.9296	0.9588	0.9163	0.9378	0.9527
All	0.8622	0.9203	0.9301	0.8925	0.9281	0.9347

cue, and the performances for most of distortion types are also promoted.

V. CONCLUSIONS

This paper presents a new three-dimensional (3D) visual attention model for stereoscopic image quality assessment. The prominent advantage of the proposed method is that we construct the 3D visual attention model by applying two-dimensional (2D) saliency model, center bias, depth cue (foreground and background), and investigate various combination and modulation means for quality assessment task. It can be observed from the experimental results show that the proposed method can achieve much higher consistency with the subjective assessments. In this research, we only take the depth cue from binocular disparity into account. In the future work, more comprehensive study of various depth cues on visual saliency is needed, and more importantly, visual comfort factor should be fully considered in the 3D visual attention model.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of China (grant 61271021), and the Natural Science Foundation of Ningbo (grant 2012A610039).

REFERENCES

- [1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang, "Three-dimensional video postproduction and processing," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 607-625, April 2011.
- [2] P. Geng, H. Jiang, and Z. Zhang, "A video denoising method with 3D surfacelet transform based on block matching and grouping," *Journal of Computers*, vol. 7, no. 5, pp. 1065-1066, 2012.
- [3] H. T. Quan, P. Le Callet, and M. Barkowsky, "Video quality assessment: from 2D to 3D – challenges and future trends," in *Proc. of IEEE International Conference on Image Processing*, pp. 4025-4028, Hong Kong, Sep. 2010.
- [4] Y. Han, J. Zhang, and B. Chang, "Novel fused image quality measures based on structural similarity," *Journal of Computers*, vol. 7, no. 3, pp. 563-566, 2012.
- [5] A. K. Moorthy, C. C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, DOI: 10.1016/j.image.2012.08.004, 2012.
- [6] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," in *Proc. of SPIE: Three-Dimensional Image Processing and Applications*, vol. 7526, no. 75260S, Feb. 2010.
- [7] M. Lambooi, W. IJsselsteijn, D. G. Bouwhuis, and I. Heynderickx, "Evaluation of stereoscopic images: beyond 2D quality," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 432-444, June 2011.
- [8] J. Gautier, and O. Le Meur, "A time-dependent saliency model combination center and depth biases for 2D and 3D viewing conditions," *Cognitive Computation*, vol. 4, no. 2, pp. 141-156, June 2012.
- [9] Q. H. Thu, M. Barkowsky, and P. Le Callet, "The importance of visual attention in improving the 3D-TV viewing experience: Overview and new perspectives," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 421-431, June 2011.
- [10] O. Le Meur, A. Ninassi, P. Le Callet and D. Barba, "Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 547-558, Aug. 2010.
- [11] J. You, A. Perkis, M. M. Hannuksela, and M. Gabbouj, "Perceptual quality assessment based on visual attention analysis," in *Proc. of the 17th ACM international conference on Multimedia*, pp. 561-564, Beijing, China, Oct. 2009.
- [12] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.
- [13] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model the bottom-up visual attention," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802-817, May 2006.
- [14] C. Lang, T. V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, and S. Yan, "Depth matters: influence of depth cues on visual saliency," in *Proc. of the European Conference on Computer Vision*, pp. 101-115, Firenze, Italy, Oct. 2012.
- [15] C. T. Li, Y. C. Lai, C. Wu, S. F. Tsai, T. C. Chen, S. Y. Chien, and L. G. Chen, "Brain-inspired framework for fusion of multiple depth cues," *IEEE Trans. on Circuits and Systems for Video Technology*, DOI: 10.1109/TCSVT.2012.2223874, 2012.
- [16] E. Potapova, M. Zillich, and M. Vincze, "Learning what matters: combining probabilistic models of 2D and 3D saliency cues," in *Proc. of International Conferences on Computer Vision Systems*, pp. 132-142, Sophia Antipolis, France, Sep. 2011.

- [17] Y. Zhang, G. Y. Jiang, M. Yu, Y. Yang, Z. J. Peng, and K. Chen, "Depth perceptual region-of-interest based multiview video coding," *Journal of Visual Communication and Image Representation*, vol. 21, no. 5-6, pp. 498-512, June 2010.
- [18] Y. Park, B. Lee, W. Cheong, and N. Hur, "Stereoscopic 3D visual attention model considering comfortable viewing," in *Proc. of IET Conference on Image Processing*, pp. 1-5, London, UK, July 2012.
- [19] A. Maalouf, and M. C. Larabi, "CYCLOP: A stereo color image quality assessment metric," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1161-1164, Prague, Czech Republic, May 2011.
- [20] P. Gorley, and N. Holliman, "Stereoscopic image quality metrics and compression," in *Proc. of SPIE: Stereoscopic Displays and Applications XIX*, San Jose, CA, USA, vol. 6803, no. 680305, 2008.
- [21] J. J. Hwang and H. R. Wu, "Stereo image quality assessment using visual attention and distortion predictors", *KSII Trans. on Internet and Information Systems*, vol.5, no.9, pp.1613-1631, September 2011.
- [22] X. Wang, S. Kwong, and Y. Zhang, "Considering binocular spatial sensitivity in stereoscopic image quality assessment," in *Proc. of IEEE Visual Communications and Image Processing*, Tainan City, Taiwan, Nov. 2011.
- [23] R. Bensalma, and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems and Signal Processing*, DOI: 10.1007/s11045-012-0178-3, 2012.
- [24] U. Engelke, H. Kaprykowsky, H. J. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 50-59, Nov. 2011.
- [25] H. Wang, Y. Dang, H. Ke, and X. Liu, "A method of target region detection based on multi-channel weighted visual attention," *Journal of Computers*, vol. 8, no. 10, pp. 2478-2482, 2013.
- [26] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1-8, Minneapolis, Minnesota, USA, June 2007.
- [27] W. Yang, Y. Y. Tang, B. Fang, Z. Shang, and Y. Lin, "Visual saliency detection with center shift," *Neurocomputing*, vol. 103, no. 1, pp. 63-74, March 2013.
- [28] J. Wang, P. Le Callet, S. Tourancheau, V. Ricordel, and M. Perreira Da Silva, "Study of depth bias of observers in free viewing of still stereoscopic synthetic stimuli," *Journal of Eye Movement Research*, vol. 5, no. 5, pp. 1-11, Sep. 2012.
- [29] V. Kolmogorov, and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. of IEEE International on Computer Vision*, vol. 2, pp. 508-515, Vancouver, Canada, Jul. 2001.
- [30] F. Shao, G. Y. Jiang, M. Yu, K. Chen, and Y. S. Ho, "Asymmetric coding of multi-view video plus depth based 3D video for view rendering," *IEEE Trans. on Multimedia*, vol. 14, no. 1, pp. 157-167, Feb. 2012.
- [31] H. Liu, and I. Heynderickx, "Visual attention in objective image quality assessment based on eye-tracking data," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971-982, July 2011.
- [32] M. J. Chen, C. C. Su, D. K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," in *Proc. of Forty-Sixth Annual Asilomar Conference on Signals, Systems, and Computers*, Monterey, California, Nov. 2012.
- [33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.
- [34] J. Zhou, G. Jiang, X. Mao, M. Yu, F. Shao, Z. Peng, and Y. Zhang, "Subjective quality analyses of stereoscopic images in 3DTV system," in *Proc. of IEEE Visual Communications and Image Processing*, Tainan City, Taiwan, Nov. 2011.

Qiuping Jiang received his B.S.degree from Department of Information and Engineering, China Jiliang University, Hangzhou, China, in 2012. He is now working towards his M.S. degree in Ningbo University, Ningbo, China. His current research interests include image and video processing and stereoscopic image quality assessment.

Fenfang Duan received her B.S.degree from Department of Information and Engineering, East China Institute of Technology, Nanchang, China, in 2011. She is now working towards her M.S. degree in Ningbo University, Ningbo, China. Her current research interests include image and video processing and stereoscopic image quality assessment.

Feng Shao received his B.S. and Ph.D degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in Electronic Science and Technology. He is currently an associate professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests include video coding, image processing and perception, etc.