Subtitle Region Selection of S3D Images in Consideration of Visual Discomfort and Viewing Habit

77

GUANGHUI YUE, Shenzhen University, China
CHUNPING HOU, Tianjin University, China
TIANWEI ZHOU, Shenzhen University, China

Subtitles, serving as a linguistic approximation of the visual content, are an essential element in stereoscopic advertisement and the film industry. Due to the vergence accommodation conflict, the stereoscopic 3D (S3D) subtitle inevitably causes visual discomfort. To meet the viewing experience, the subtitle region should be carefully arranged. Unfortunately, very few works have been dedicated to this area. In this article, we propose a method for S3D subtitle region selection in consideration of visual discomfort and viewing habit. First, we divide the disparity map into multiple depth layers according to the disparity value. The preferential processed depth layer is determined by considering the disparity value of the foremost object. Second, the optimal region and coarse disparity value for S3D subtitle insertion are chosen by convolving the selective depth layer with the mean filter. Specifically, the viewing habit is considered during the region selection. Finally, after region selection, the disparity value of the subtitle is further modified by using the just noticeable depth difference (JNDD) model. Given that there is no public database reported for the evaluation of S3D subtitle insertion, we collect 120 S3D images as the test platform. Both objective and subjective experiments are conducted to evaluate the comfort degree of the inserted subtitle. Experimental results demonstrate that the proposed method can obtain promising performance in improving the viewing experience of the inserted subtitle.

CCS Concepts: • **General and reference** → **Cross-computing tools and techniques**; • **Evaluation**; • **Networks** → Network reliability;

Additional Key Words and Phrases: Stereoscopic 3D, visual discomfort, subtitle, horizontal disparity, evaluation, just noticeable depth difference (JNDD)

ACM Reference format:
Guanghui Yue, Chunping Hou, and Tianwei Zhou. 2019. Subtitle Region Selection of S3D Images in Consideration of Visual Discomfort and Viewing Habit. *ACM Trans. Multimedia Comput. Commun. Appl.* 15, 3, Article 77 (August 2019), 16 pages.
https://doi.org/10.1145/3325197

This work was supported by the National Natural Science Foundation of China under Grants (Nos. 61731003, 61520106002, 61871274, 61801305, and 81571758), National Natural Science Foundation of Guangdong Province (Nos. 2017A030313377 and 2016A030313047), Shenzhen Peacock Plan (No.KQTD2016053112051497), and Shenzhen Key Basic Research Project (Nos. JCYJ20170818142347251 and JCYJ20170818094109846).
Authors' addresses: G. Yue is with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China; email: guanghuiyue.doctor@gmail.com; C. Hou (corresponding author), Tianjin University, Weijin Road, Tianjin, 300072, China; email: hcp@tju.edu.cn; T. Zhou (corresponding author) is with the College of Management, Shenzhen University, Shenzhen 518060, China; email: tianweizhoudr@gmail.com.

1551-6857/2019/08-ART77 $15.00
https://doi.org/10.1145/3325197

1 INTRODUCTION

Recently, stereoscopic 3D (S3D) images and videos have been widely prevailed and successfully applied in many applications, e.g., movies, advertisements, and computer games [11, 38, 39]. However, there is still a long way to widespread promotion of S3D services due to issues of visual discomfort, display adaptation, and content postproduction [5, 17, 31, 32, 41]. How to solve these issues is indeed an urgent task for both academia and industry. Currently, many works have been carried out on visual discomfort prediction as well as visual comfort enhancement of S3D images [4, 7, 13, 22, 41].

In general, visual discomfort can be induced by a variety of factors, such as geometrical distortions, vergence-accommodation (VA) conflict, object motion, and so on [16, 18, 35, 40]. The horizontal disparity,[1] which induces VA conflict in the S3D display system, is regarded as the main reason for generating visual discomfort. To date, many visual discomfort prediction methods have been proposed via analysis of the disparity. For example, the disparity gradient, as the limiting factor for the binocular fusion, was considered as the objective metric to assess the visual discomfort [16]. In Ref. [25], an effective predictor was proposed by considering accommodation and vergence conflicts. To indicate the accommodation conflict, the local 3D bandwidth was defined and used. To formulate the vergence conflict, some features were extracted from the disparity map. Park et al. [26] extracted 13 exemplar tuning curves of neural activities of the middle temporal area and disparity statistics as features to predict the visual discomfort. Kim et al. [18] simplified the processing procedure of the visual path from eyes to the whole brain when watching S3D images as a transfer function and creatively proposed a unique temporal visual discomfort prediction model.

Generally speaking, film producers are keen to place the object outside of the screen by arranging a large disparity to provide a strong sense of immersion. However, visual discomfort would happen if the disparity is too large. To cope with this problem, researchers have designed many postproduction techniques for improving visual comfort [14, 29]. Among them, disparity remapping is a direct way. Unfortunately, unnaturalness, such as occlusion and disparity inversion, may occur due to simple linear remapping. By contrast, the nonlinear disparity remapping schemes work better. Lang et al. [21] designed a nonlinear disparity remapping scheme to enhance the perception of salient objects. Sohn et al. [34] applied both linear and nonlinear disparity mapping algorithms to further improve the visual comfort. Specifically, a linear disparity remapping was performed to adjust the disparity range to the target range, and a nonlinear disparity remapping was employed to selectively adjust the disparity of problematic local disparity ranges. Besides, researchers also used the visual discomfort map to guide the operation. Shao et al. [30] relaxed the visual discomfort by automatically adjusting the disparity range according to optimal zero disparity plane, which was guided by a visual discomfort map, as required.

In video sequences or images, the subtitle is always inserted in front of the foremost object in scenes, and its region directly affects the visual discomfort [36]. However, there is no related work on visual dicomfort prediction or comfort enhancement of the subtitle. In this article, we propose a pioneering subtitle region selection method of S3D images. The main contributions of this article are listed as follows: (1) The proposed method optimizes the subtitle region in consideration of visual discomfort and viewing habit via analysis of the foremost object and the just notice difference depth (JNDD) model. It enhances the visual comfort of S3D subtitles and does not affect the immersion sense of original image contents. (2) We collect 120 S3D images with various contents for the S3D subtitle insertion. These images can be classified into the comfortable part and the uncomfortable part. For images in the comfortable part, their disparity values are within the comfort zone

[1]In this article, for simplicity, disparity denotes the horizontal disparity if there is no special explanation.

(i.e., Percival's Zone of Comfort (PZC)), while those in the uncomfortable part are beyond PZC. To the best of our knowledge, this is the first test platform for performance evaluation of the S3D subtitle insertion. (3) Experimental results from objective and subjective evaluations demonstrate that the proposed method can reduce the visual discomfort caused by the subtitle and enhance the viewing perception while retaining the sense of immersion produced by original image contents.

The remaining sections of this article are organized as follows. Section 2 provides some background knowledge and formulates the motivation. Section 3 describes the proposed region selection method of S3D subtitles in detail. Comprehensive evaluations are given and discussed in Section 4. Finally, the conclusions are drawn in Section 5.

2 RELATED WORK AND MOTIVATION

In this section, we first briefly introduce the comfort zone and JNDD model. Then, we describe some relevant works about the S3D subtitle comfort evaluation. Finally, the motivation of this article is highlighted.

2.1 Comfort Zone and JNDD Model

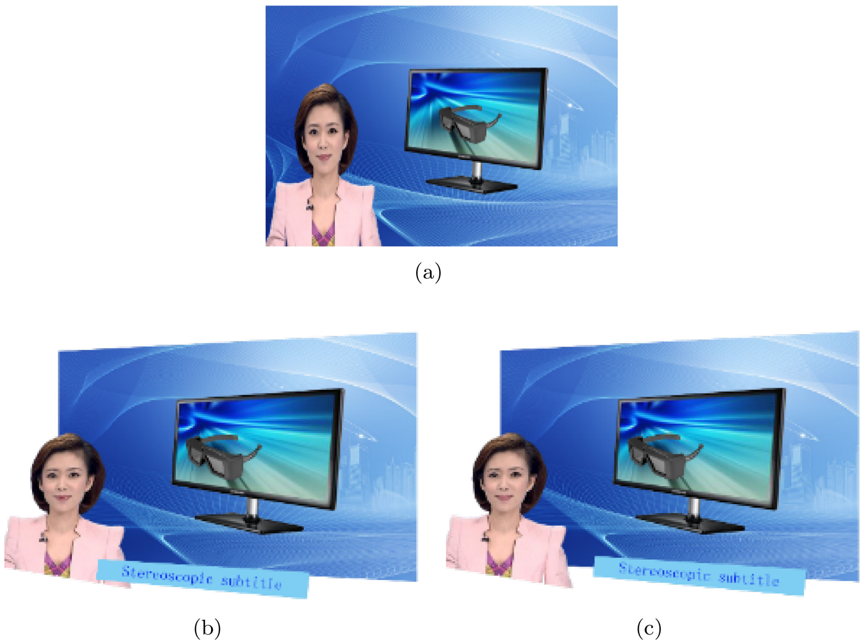
In general, if the disparity does not lie within the comfort zone, the so-called PZC, observers will perceive visual discomfort [20]. Therefore, the PZC can be treated as an essential element to guide the object's position in the S3D scene. In the literature [6], the investigation about how sensitive the users are in perceiving the depth changes in a S3D scene was conducted through careful subjective experiments. To be specific, a synthetic image sequence with two of the same objects (i.e., a car) was presented to the viewer. Initially, the two cars were placed at the same depth level relative to the screen. Then, the depth of the right one was gradually changed and the viewer gave the response when he/she perceived the movement of the right car. Once the movement was sensed, the associated depth level difference between the two cars was recorded as the sensed depth changed. By changing the initial depth level of two objects, one could finally obtain the depth change under different depth levels. The depth information can be transformed into binocular disparity. After rejecting the unqualified data, a JNDD model was built based on the remaining results. According to Weber's law, a larger stimulus difference is required in case of a larger initial stimulus [24, 33]. The just noticeable depth difference is proportional to the distance between the perceived object and screen, and inversely proportional to the square of the viewing distance. In this study, we use the quantized JNDD model instead of the continuous one for the simplification [6]. Formally, it can be formulated as:

$$D_{JNDD}(D_v) = \begin{cases} 21 & , if \quad 0 \le D_v \le 64 \\ 19 & , if \quad 64 \le D_v \le 128 \\ 18 & , if \quad 128 \le D_v \le 192 \\ 20 & , if \quad 192 \le D_v \le 255 \end{cases}, \tag{1}$$

where D_v is the depth level. By definition, the JNDD represents the minimal depth difference between objects perceived by a human visual system (HVS) in a S3D scene.

2.2 Subtitle Comfort Evaluation

With the popularity of S3D media, people gradually notice the visual discomfort problem caused by subtitles. Up to now, some basic studies have been conducted on issues of S3D subtitle comfort evaluation. For example, Wu et al. [37] conducted subjective experiments to investigate the impact of S3D subtitles on visual comfort. Experimental results showed that subjects felt more discomfort when watching subtitles with uncrossed disparities than that with crossed disparities. Through subjective experiments, Lambooij et al. [19] revealed that almost all contents in the S3D scene could cause visual discomfort. The subtitle, regarded as the secondary salient region, could help to



(a) (b) (c)

Fig. 1. Subtitles insertion in 3DTV. (a) Original 3D scene. (b) Subtitle inserted in front of the foremost object. (c) Subtitle inserted manually.

reduce the visual discomfort induced by other contents. This indicates that the subtitle's position is very essential. Wan et al. [36] conducted a subjective experiment to explore the approximation disparity value for placing subtitles in the S3D video. Experimental results demonstrated that, for a better viewing experience, the subtitle's disparity value should not exceed two pixels than that of the foremost object in the S3D scene. In summary, all these works investigate the visual discomfort caused by the S3D subtitle on the surface via subjective experiments.

2.3 Motivations

Compared with the 2D subtitle, the S3D subtitle should guarantee the visual comfort and avoid the visual unnaturalness caused by depth occlusion. One simple idea is inserting an S3D subtitle in front of the foremost object at the bottom of the image. In such case, the subtitle may shadow the foremost object and its disparity value is larger than the foremost object, causing more visual discomfort than the object as well as visual unnaturalness. Given that observers pay more attention to the foremost objects [8], and the subtitle is regarded as the second salient information in the image [28], both the foremost object and subtitle should be arranged to mutually exclusive positions. Currently, to confirm visual comfort and naturalness, the S3D subtitle is manually inserted after film editing or advertisement synthesis. To be specific, a film producer manually inserts an S3D subtitle in an optimal region, where it produces stronger immersive perception with less visual discomfort and avoids collision with other objects. An intuitive example is given in Figure 1, where the original 3D scene[2] is depicted in Figure 1(a). Figure 1(b) describes the effect after simple subtitle insertion, while Figure 1(c) shows the effect with manual subtitle insertion. It is clear

[2]The image is generated by combining some images downloaded from the Internet and the copyright belongs to their rightful owners. The authors do not claim ownership. No copyright infringement is intended.

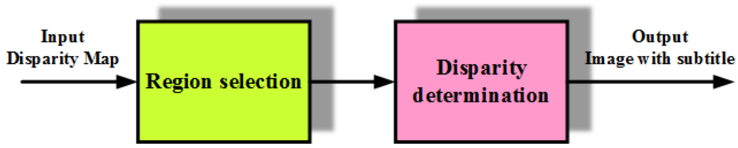


Fig. 2. Framework of the proposed method.

that the disparity value of Figure 1(c) is relatively smaller than the foremost object without occlusion with it. However, manual subtitle insertion is time-consuming, expensive, and has become a possible obstacle toward large-scale promotions of S3D productions. Therefore, providing an automatic region selection method for S3D subtitle is an urgent but challenging task due to the lack of reference and comparison metrics. To fill this void, we propose a simple but effective method for visual discomfort amelioration of S3D subtitles without affecting the viewing effect of original image content.

3 SUBTITLE REGION SELECTION METHOD

Figure 2 depicts the framework of the proposed method. The proposed method mainly contains two stages, i.e., region selection and disparity value determination.

3.1 Region Selection

The main goal of this article is to choose an optimal region for discomfort reduction of the S3D subtitle without changing image contents. This work is carried out on the disparity map due to that disparity is the main reason for the VA conflict as discussed above. However, it is very cumbersome to operate directly on the entire disparity map. Since the natural scene is composed of independent objects with different depth layers, the disparity map can be divided into multiple layers according to the disparity range [23]. The depth can be converted into disparity as described in Section 3.2. Figure 3 gives a simple schematic of viewing S3D images. The green and orange lines denote the different depth layers, respectively. According to the position, the scene located on the display plane has zero disparity. Whereas, that located in front of the display plane has negative (crossed) disparity. In this study, we separate the disparity map into multiple layers (the disparity range is 6 pixels in each layer), and determine the position and disparity value by analyzing the depth layer.

As illustrated in Figure 3, there are multiple depth layers. Therefore, the selective order about which depth layer is processed first directly determines the computational complexity and the final position of the subtitle. Generally speaking, we hope that the S3D subtitle should have a small disparity value, and thus induce less visual discomfort. In such a situation, placing the subtitle on the zero disparity plane is the best. However, for a S3D scene, viewers are more interested in the foremost object [9]. Meanwhile, the subtitle serves as the second most interesting element [19]. That is to say, viewers pay attention to both elements when viewing the S3D scene. If these two elements have too large a depth difference, the viewer will sense a large disparity change, which induces visual discomfort. The larger the difference is, the stronger the discomfort will be. As stated in the Introduction, reediting the S3D scene is contrary to the initial purpose that produces a stronger sense of immersion. In this study, we try to select the best position for inserting a subtitle without editing the S3D content. Specifically, the subtitle is arranged with a similar depth layer as that of the foremost object. In this case, the depth difference between the foremost object and subtitle is the smallest, thereby causing minimal visual discomfort. According to the disparity value of the foremost object, it can be divided into two situations. One is that disparity value of the foremost object exceeds the PZC boundary (i.e., -1°). In this situation, the foremost object easily induces severe visual discomfort. If the subtitle is inserted with the similar depth layer, it will also

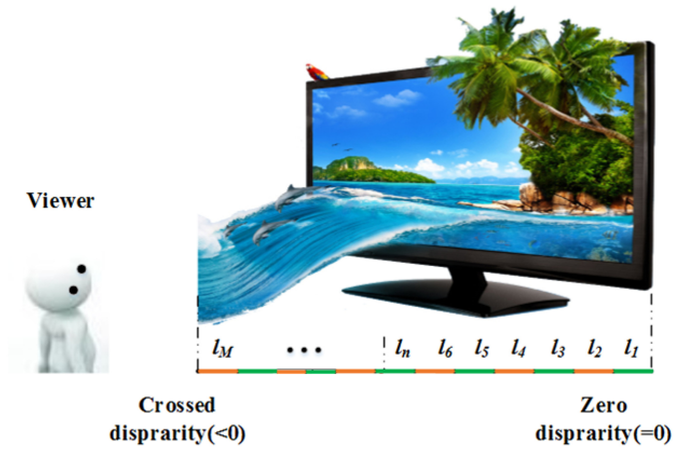


Fig. 3. Simple schematic of viewing S3D images. The green and orange lines denote the different depth layers, respectively.

cause severe discomfort. To ensure the visual immersion and avoid the large disparity difference between it and the foremost object, we directly arrange it on the PZC boundary plane. Another situation is that the disparity value of the foremost object is within the PZC range. Naturally, the subtitle could be directly inserted with the similar depth layer as that of the foremost object. After determining the coarse depth layer, our next task is to select the optimal position for the subtitle. To be specific, for the depth layer, we first convert it to the binary one:

$$B_d = \begin{cases} 0 & , if \;\; d \le T_1 \\ 1 & , otherwise \end{cases}, \tag{2}$$

where B_d is the binary pattern of the depth layer, d is the disparity value in depth layer. T_1 is the threshold for computing the binary map. In this study, we set $T_1 = d_m - 6$ empirically, where d_m[3] is the minimum disparity value in the depth layer. Obviously, d_m is the minimum value of the disparity map in the subtitle region if the foremost object is within PZC range. By Equation (2), the region with disparity value beyond T_1 is full of zero, and the other regions are full of one. Clearly, the region full of one can be selected as the candidate region for subtitle insertion, and the inserted subtitle should avoid depth collision with objects in the region full of zero.

Figure 4 gives an intuitive example about this operation. Figure 4(a) is the left view of the stereoscopic pair (the subtitle region is labeled in the red rectangle[4]), and Figure 4(b) is the associated disparity map of the labeled region. Then, the disparity map is divided into multiple layers according to the disparity value. Figure 4(d) is the binary pattern of one disparity layer in Figure 4(c). Note that the white and black appearances denote the regions within and beyond the threshold, respectively. In Figure 4(d), the white area is chosen as the candidate region for the subtitle insertion. As seen, the binarization approach can achieve good segmentation results for candidate regions. Then, the problem is transformed to select an optimal position for the subtitle insertion in white regions. It is found that multiple equal-size regions as the subtitle text may appear in white regions.

[3]Actually, the subtitle is usually inserted outside of the screen and its disparity value is accordingly negative. Here, the disparity takes the pixel as the unit.
[4]In this study, we take the region (ranging from the image bottom to $\frac{1}{4}$ image height and the whole image width) as the subtitle region. In practical situations, this region can be changed based on the user's penchant.

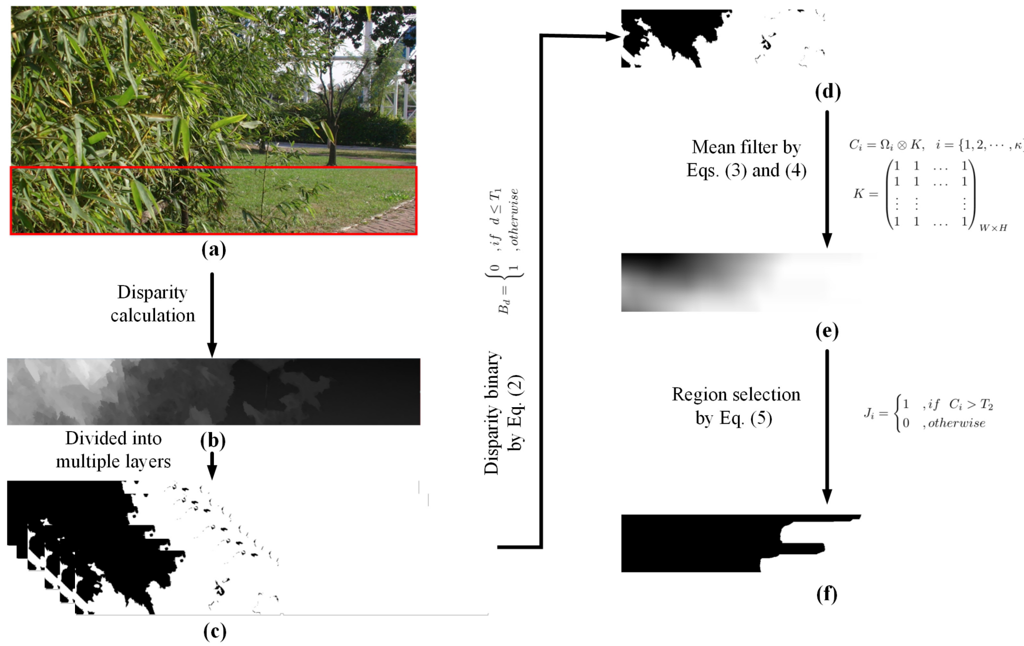


Fig. 4. Framework of the S3D subtitle region selection.

To determine the suitable region for subtitle insertion, for a pixel, its surrounding region Ω_i with the same size as the subtitle text is first processed by the mean filter:

$$C_i = \Omega_i \otimes K, \quad i = \{1, 2, \ldots, \kappa\} \tag{3}$$

with

$$K = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix}_{W \times H}, \tag{4}$$

where κ is the number of pixels, K is the kernel of the mean filter, which is equal to the size of the subtitle text with width W and height H. Theoretically, the value of the white region should still be one after filtering, while that of the region across the black and white regions is less than one. In other words, the filtered value reflects the feasibility of subtitle insertion in this region. Specifically, whether the surrounding region of the i-th pixel is suitable for subtitle insertion can be determined by:

$$J_i = \begin{cases} 1 & ,if\ \ C_i > T_2 \\ 0 & ,otherwise \end{cases}, \tag{5}$$

where 1 denotes that this region is suitable for subtitle insertion. T_2 is the threshold to control the size of the optional area. Ideally, T_2 should be arranged as 1. In this article, we set it as 0.98 empirically since the calculated disparity map is not flawless. Figures 4(d)–(f) depict the visualization using the above operators. By filtering the binary pattern (i.e., Figure 4(d)), the edges between the black region and the candidate region are diffused, as shown in Figure 4(e). With this characteristic, we can identify the edge where the subtitle is not suitable to insert. To this end, the ultimate region for subtitle insertion (as shown in the white area of Figure 4(f)) is finally determined through the threshold processing. Considering the viewing habit, the optimal position is obtained as the median value of coordinates in the ultimate region.

3.2 Disparity Determination

Through region selection, we have determined the optimal position for the S3D subtitle insertion. Our next task is to determine the disparity value of the inserted subtitle. According to the region selection procedure, it can be observed that the subtitle is inserted with the similar disparity value as that of the considered depth layer. Hence, we basically set the disparity value of the subtitle as T_1 (T_1 is set as -1° when the foremost object is beyond the PZC). Our next consideration is how to further reduce the visual discomfort without loss of depth perception. In this study, we apply the JNDD model [15] to guide the disparity value of the subtitle. Since the JNDD model is operated on a depth map, which is not realistic to be directly obtained via a commonly used stereo camera, the depth map should be calculated primarily. Fortunately, there are a lot of mature stereo matching algorithms. We can apply these algorithms to obtain disparity map, d, then, transform it into a depth map. Formally, the relationship between d and real word depth z is expressed as:

$$z = \frac{B \cdot f}{d}, \tag{6}$$

where B denotes the baseline and f is the focus length of the capture camera. Next, the value of normalized depth map, D_n, can be expressed as:

$$D_n = 255 \cdot \left(\frac{1}{z} - \frac{1}{z_{max}}\right) \bigg/ \left(\frac{1}{z_{min}} - \frac{1}{z_{max}}\right), \tag{7}$$

where z_{min} and z_{max} denote the original minimum and maximum depth value of the real word scene, respectively. Through the integration of Equations (6)–(7), the relationship between D_n and d can be further expressed as:

$$\begin{aligned} D_n &= 255 \cdot \left(\frac{d}{B \cdot f} - \frac{d_{max}}{B \cdot f}\right) \bigg/ \left(\frac{d_{min}}{B \cdot f} - \frac{d_{max}}{B \cdot f}\right) \\ &= \frac{255 \cdot d}{d_{max} - d_{min}} - \frac{255 \cdot d_{min}}{d_{max} - d_{min}} \end{aligned} \tag{8}$$

where d_{min} and d_{max} represent the minimum and maximum disparity value in the disparity map. It can be intuitively observed from Equation (8) that the depth value is proportional to the disparity value.

With the binocular disparity in hand, we can apply Equation (1) to further determine the disparity value of the inserted subtitle. More specifically, the disparity value of the S3D subtitle can be slightly larger (because the disparity value is negative, the visual comfort is enhanced when the value becomes larger) than that of the basically determined value T_1 within the JNDD range. With such an operation, the visual comfort of a subtitle can be improved without affecting the perception of depth. Figure 5 shows the effect of the proposed method. In the figures, A denotes the foremost object and B is the subtitle text in the image content. Figure 5(a) depicts the basic method that inserts the subtitle in front of the foremost object. In such a case, a subtitle is the most discomforting factor in image. As for the proposed method, we first judge whether the disparity value of the foremost object is within PZC. If so, the disparity value of B is arranged with larger values than that of A without loss of depth perception; that is, the depth difference between it and A should be within JNDD range, as shown in Figure 5(b). To this end, the JNDD model works effectively and the subtitle's comfort can be improved. Otherwise, when A exceeds PZC too much, the disparity value of the subtitle text is still beyond PZC, even using the JNDD model as shown in Figure 5(c).

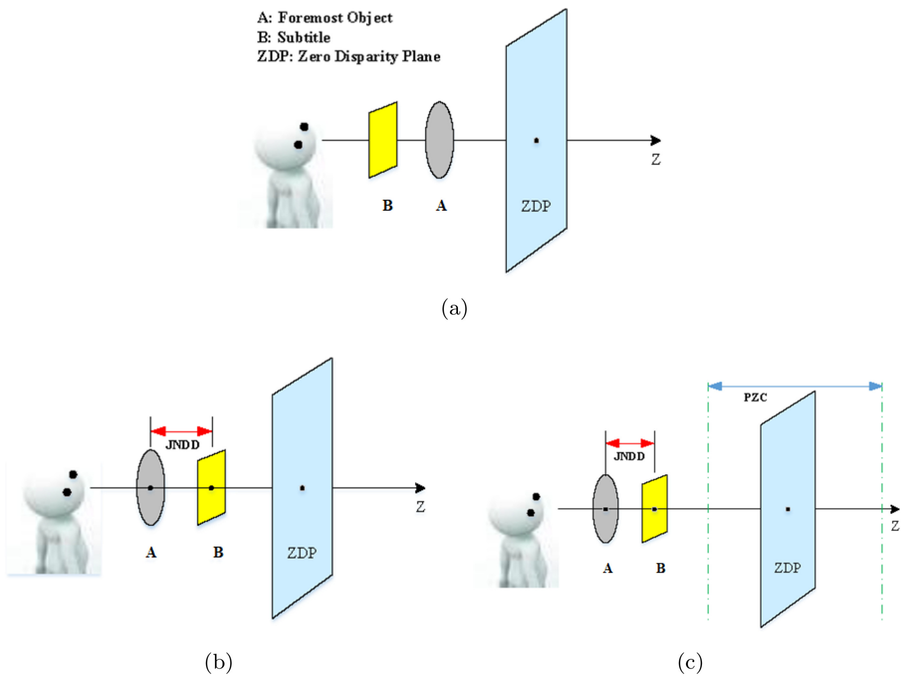


Fig. 5. Illustration of different scenarios for the subtitle insertion. (a) The subtitle is inserted in front of the foremost object. (b) The subtitle is inserted behind the foremost object within the JNDD range. (c) The subtitle is inserted behind the foremost object within the JNDD range, but exceeds the PZC.

In such a case, we directly insert the subtitle text inside the PZC boundary without consideration of JNDD to reduce visual discomfort and maintain the depth perception of the subtitle text.

4 EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental Setup

To create a platform for the proposed method, we collect 120 stereoscopic image pairs with highly diverse disparity values. All the images are captured by a JVC GS-TD1BAC stereo camera with a resolution of 1920 × 1080. Figure 6 gives some example images. During image acquisition, the camera is placed on the horizontal platform to alleviate vertical disparity. It is worth noting that several S3D image databases, e.g., Jung's database [16] and École Polytechnique Fédérale de Lausanne (EPFL) database [10], have been built before this work. However, our work possesses innovative significance from the following perspectives. First, the visual discomfort of subtitles is more competent with image disparity range beyond PZC. Existing S3D image databases are specifically used for the image discomfort evaluation. The disparity ranges of most images in these databases are within PZC. Therefore, these images are unsuitable for the subtitle discomfort evaluation. Second, on the contrary, the collected images mostly exceed PZC with more negative disparities, as depicted in Figure 7. If the disparity range of image exceeds PZC, using the basic subtitle inserted method, i.e., placing the subtitle in front of the foremost object, may induce an uncomfortable viewing experience. In such a case, it is conducive to verify the effectiveness of a visual discomfort algorithm. Since the proposed method is operated on the disparity map, the accuracy of the disparity map has a direct effect on the performance of the experiment. In this article, we take the optical flow software [1] as an effective tool to obtain the more accurate disparity map.

Fig. 6. Example images from the 120 S3D images. To better present in the text, only the left view is given.

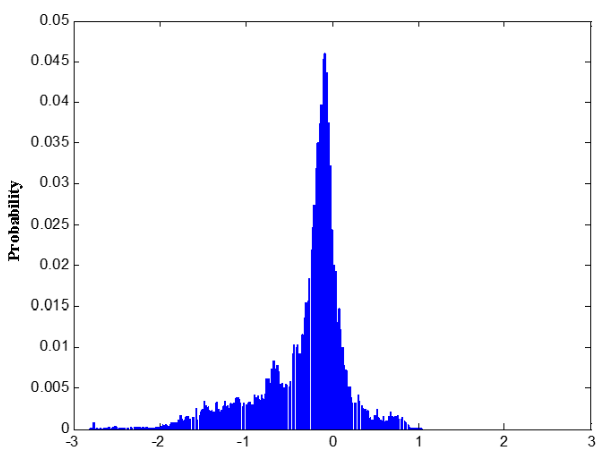


Fig. 7. Horizontal disparity distribution over all 120 image pairs (in deg. units). As can be seen, the distribution is nearly one-side, with more negative disparities.

To the best of our knowledge, no related work has been reported for visual discomfort melioration. Thus, we take the basic method, which always directly inserts subtitles in front of the foremost object, as a benchmark for comparison. After the subtitle insertion, how to evaluate the visual discomfort is a thorny problem as existing objective visual discomfort prediction metrics are specifically designed for S3D images rather than subtitles [12, 13]. Given that the visual discomfort of the S3D subtitle is induced by VA conflict, we can draw lessons from the S3D image discomfort evaluation methods, which only utilize a single disparity value. With this consideration, an objective discomfort evaluation metric, i.e., Perrin's comfort function [27], is employed to evaluate the algorithm's performance. Perrin's comfort function takes spatial frequency and disparity value as variables. It can be expressed as:

$$C(d, s) = \alpha \cdot (d - d_0 - k \cdot s^{k'}), \tag{9}$$

(a) (b)

Fig. 8. Examples of the anaglyph images processed by (a) the proposed method and (b) the basic method to insert the S3D subtitle. The disparity values of the subtitle region in the image is $[-1.5^\circ, 0.5^\circ]$.

where s is the spatial frequency of images. α, d_0, k, and k' are the constants with values of -0.01, 18.9, 221.1, and 0.74, respectively [27]. At first glimpse, it seems unsuitable for the subtitle comfort evaluation. Since the subtitle has the same texture with small spatial disparity values, its spatial frequency can be set as a constant. In this article, we set it as 1 cpd. Therefore, Perrin's comfort function can be employed for S3D subtitle discomfort evaluation.

We also conduct a subjective test to further verify the effectiveness of the proposed method. Twenty subjects with normal stereo vision, aged from 20 to 26, participated in the test. The experimental environment is strictly in accordance with the recommendations ITU-R BT.500-11 [2] and ITU-R 1438 [3]. All the S3D image pairs in the database are processed by two methods to insert the subtitle. That is, a total of 240 stereoscopic image pairs are included in test sequences. During the experiment, the image pair is displayed on a 23-inch 3D monitor, LG 2343P, with resolution 1920 $\times$ 1080. The generated image pairs with subtitle inserted by two metrics are presented to subjects successively in random order. Then, subjects are required to evaluate the comfort degree between two image pairs: "1: better," "0: equal," and "-1: worse." Besides, they also need to judge which one accords with a viewing habit using the same evaluation standard. To avoid the accumulated fatigue, viewers are requested to have a rest every 5 minutes. After the subjective experiment, we eliminate the abnormal data via the Grubbs algorithm.

4.2 Visualization Results

To better present the effectiveness of the proposed method, the subtitle-inserted S3D images using two methods are demonstrated in Figure 8. We illustrate this issue from two perspectives, visual comfort and viewing habit. The selected S3D image is seriously uncomfortable with the disparity range $[-1.5^\circ, 0.5^\circ]$. By the basic method, the disparity value of the subtitle is too large and exceeds PZC, leading to serious visual discomfort. On the contrary, the disparity value of the subtitle is smaller than the one processed by the proposed method. All the anaglyph images are presented in form of a red-green format. Readers can view them with a red-green glass. If there is no glass, readers can also judge the disparity degree by observing the degree of shadow between red and green images. The bigger shadow corresponds to a larger disparity value (here, we indicate the absolute value). To facilitate visual examination, the subtitle text regions are marked by red rectangular boxes, as depicted in Figure 8(a) and (b). It can be intuitively observed that the subtitle's disparity reduces greatly. In addition, using the basic method, the subtitle is directly inserted in

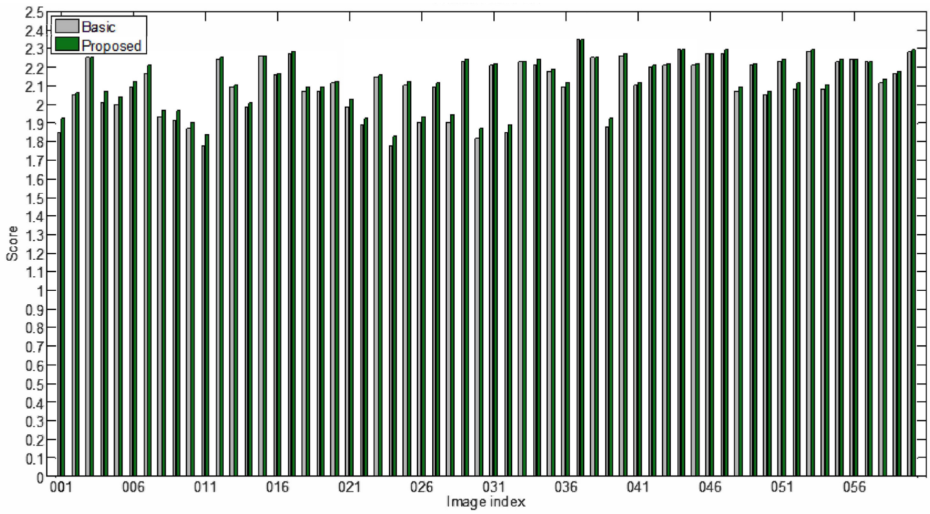
(a) (b)

Fig. 9. Examples of the anaglyph images processed by (a) the proposed method and (b) the basic method to insert the S3D subtitle. The disparity values of the subtitle region in the image is $[-0.5^\circ, 1^\circ]$.

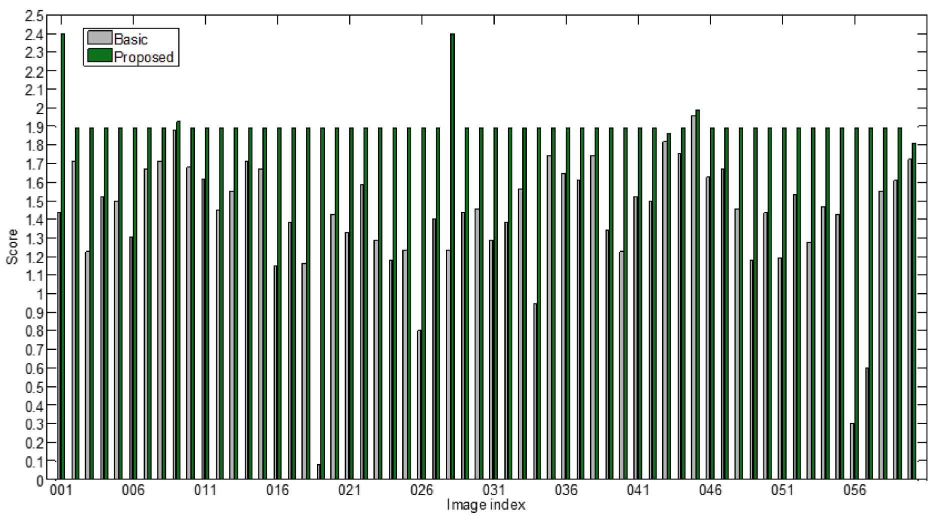
the middle of the subtitle text region and interrupts the views from the plush toy. Extensive studies have proved that the object with a larger disparity value possesses high saliency [8]. That is, observers pay more attention to the foremost object. Unfortunately, the subtitle may interpret the views from the foremost object using the basic method due to the uncertainty of scene content. Through the procession of the proposed method, the position of subtitle is selected as the center of comfort region where disparity value is small, as shown in Figure 8(b). In such a case, both subtitle and the foremost object don't affect their viewing. In addition, we also give an example with the image, which disparity ranges from -0.5° to 1°. As shown in Figure 9, the proposed method coarsely possesses a similar result as the basic one. Actually, the disparity value of the proposed method is still slightly larger than that of the basic one. This is because the effect of the JNDD model. In summary, the proposed method improves the visual comfort of the inserted subtitle in both the situations when the foremost object is within and beyond PZC.

4.3 Results of Objective Evaluation

In this section, we further demonstrate the effectiveness of the proposed method on S3D subtitle comfort improvement by analyzing the results of the objective comfort evaluation method. To reflect the different performances on different scenes, in the perception of visual comfort, the experimental results are depicted individually for the comfortable part and the uncomfortable part, as shown in Figure 10. For images in the comfortable part, the disparity values are within PZC, while those in the uncomfortable part are beyond PZC. In the figure, the x-axis denotes the image index, while the y-axis denotes the corresponding computed comfort score. Note that a larger score corresponds to larger comfort. Figure 10(a) gives the results on the comfort portion. As seen, there is no obvious difference between two subtitle insertion methods. The proposed method obtains a slightly higher comfort score than the basic one. On the contrary, our method has a great advantage on visual comfort improvement with respect to the uncomfortable portion, as described in Figure 10(b). As expected, compared with the basic method, the comfort score is tremendously increased using the proposed one. The reasonable explanation of this phenomenon is illustrated as follows. Our method contains two stages. The first stage is to find an optimal region, where the disparity value is small, for subtitle insertion. The second stage is to determine the disparity value of subtitle by utilizing the JNDD model. Hence, the visual comfort has little improvement when the disparity range of the original image is within PZC, as only JNDD works to improve the visual comfort. However, this rule is broken down when the disparity range exceeds PZC. In such case,



(a)



(b)

Fig. 10. Objective evaluation of visual comfort results on: (a) the comfortable part and (b) the uncomfortable part.

the subtitle's disparity is assigned with the value that closes to the PZC boundary. As a result, the visual comfort is improved. Overall, our approach can obtain a good performance on S3D subtitle visual comfort improvement.

4.4 Results of Subjective Evaluation

In the previous section, we have verified the effectiveness of the proposed method by the objective evaluation method. It is worthwhile to comprehensively validate the performance through an extra subjective evaluation. As the proposed method considers both the visual comfort improvement and viewing perception enhancement, we illustrate the experimental results from two perspectives.

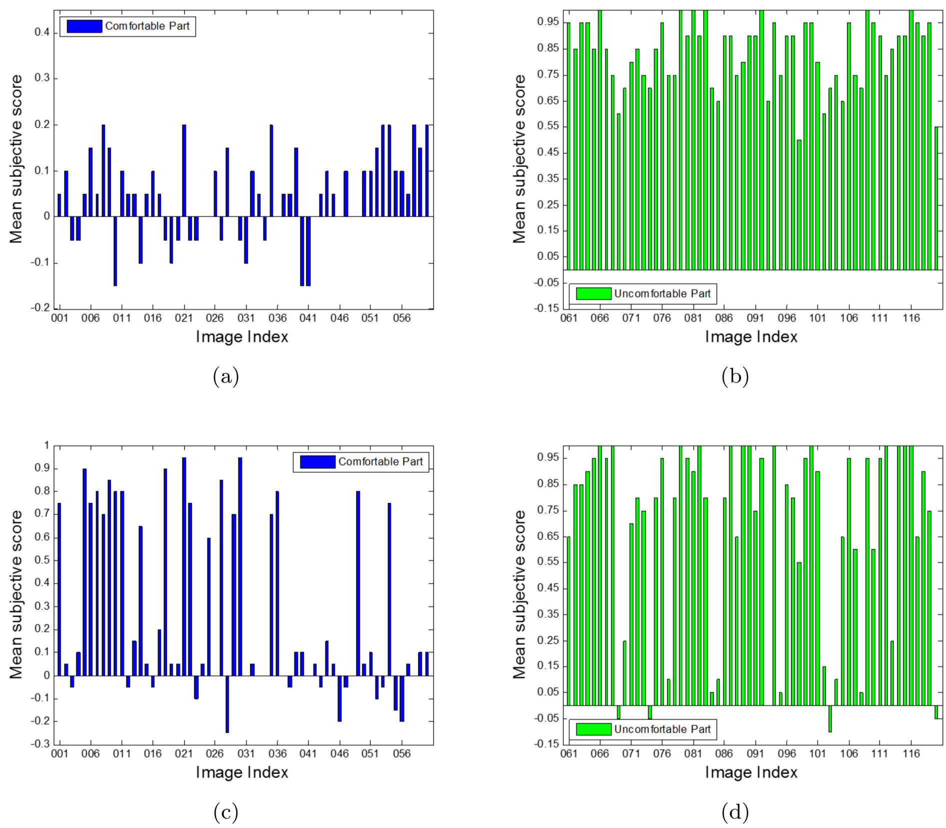


Fig. 11. Subjective evaluation of visual comfort results on: (a) the comfortable part and (b) the uncomfortable part. Subjective evaluation of visual perception results on: (c) the comfortable part and (d) the uncomfortable part.

Similar to the results in the above section, Figure 11(a) and (b) describe the subjective evaluation results in terms of visual comfort for the comfortable part and uncomfortable part, respectively. In the figure, the x-axis denotes the image index, while the y-axis denotes the corresponding mean values of all the subjects' scores. Through analysis, we find that observers possess similar performance on the comfortable part. Note that the small fluctuation is normal and has no influence on the conclusion. The reason can be attributed to that there is only JNDD between the disparity values of the two methods in view of the comfortable portion, and thus, observers have a hard time distinguishing them well. However, the results have changed dramatically with respect to the uncomfortable part. Observers produce stronger perception of comfort on the subtitle-inserted image processed by the proposed method. Toward this end, we can confidently declare that the proposed method plays a positive role in 3D subtitle visual discomfort amelioration.

Figure 11(c) and (d) describe the subjective evaluation results in terms of viewing perception for the comfortable part and uncomfortable part, respectively. As can be seen from the figures, the proposed method outperforms the basic method on most images in both parts. This observation validates the fact that our method can enhance the viewing perception. Besides, meticulous readers can find that some images, e.g., images 2, 3, and 4, have little difference under the two methods. Through analysis, we find that the disparity map is smooth in the bottom of these images. As a result, the subtitles' positions are just similar between the proposed method and the basic method.

Overall, the proposed method can improve visual comfort and enhance viewing perception synchronously.

5 CONCLUSION

Although the stereoscopic 3D display has attracted attention for many years, the visual comfort problem induced by the stereoscopic 3D subtitle has not been an object of concern yet. In this article, we make a pioneering attempt to design a region selection technique for the subtitle of S3D images and collect a number of S3D images as the test platform. The proposed method contains two stages, i.e., region selection and disparity determination, to ensure the visual comfort improvement as well as viewing perception enhancement. First, we select the optimal region for subtitle insertion by considering the visual perception effect and visual comfort via analysis of depth layers. Then, the disparity value is determined in consideration of the JNDD model. Finally, we conduct both objective and subjective experiments to validate the effectiveness of the proposed method. Experimental results show that the proposed method can effectively increase the utility of the S3D subtitle in the aspect of visual comfort improvement and viewing habit enhancement. To a certain extent, the proposed method may offer a certain guiding significance for producing stereo propagating posters and advertisements. In the future, we will extend this work toward solving a subtitle insertion problem in stereo videos.

REFERENCES

[1] Kurt Akeley, Simon J. Watt, Ahna Reza Girshick, and Martin S. Banks. 2004. A stereo display prototype with multiple focal distances. *ACM Trans. Graphics* 23, 3 (2004), 804–813.

[2] Rec. ITU-R BT-500.11. 2002. Methodology for the subjective assessment of the quality of television pictures. (2002).

[3] Rec. ITU-R BT.1438. 2000. Methodology for the subjective assessment of the quality of television pictures. (2000).

[4] Zaiqing Chen, Junsheng Shi, Xiaoqiao Huang, Lijun Yun, and Yonghan Tai. 2014. Visual comfort modeling for disparity in 3D contents based on Weber–Fechner's Law. *J. Disp. Technol.* 10, 12 (2014), 1001–1009.

[5] Chung-Hua Chu. 2015. Visual comfort for stereoscopic 3D by using motion sensors on 3D mobile devices. *ACM Trans. Multimedia Comput. Commun. Appl.* 12, 1s (2015), 14.

[6] D. Varuna SX De Silva, Erhan Ekmekcioglu, Warnakulasuriya Anil Chandana Fernando, and Stewart T. Worrall. 2011. Display dependent preprocessing of depth maps based on just noticeable depth difference modeling. *IEEE J. Sel. Top. Signal Process.* 5, 2 (2011), 335–351.

[7] Song-Pei Du, Belen Masia, Shi-Min Hu, and Diego Gutierrez. 2013. A metric of visual comfort for stereoscopic motion. *ACM Trans. Graph.* 32, 6 (2013), 222.

[8] Yuming Fang, Junle Wang, Manish Narwaria, Patrick Le Callet, and Weisi Lin. 2014. Saliency detection for stereoscopic images. *IEEE Trans. Image Process.* 23, 6 (2014), 2625–2636.

[9] Yuming Fang, Chi Zhang, Jing Li, Jianjun Lei, Matthieu Perreira Da Silva, and Patrick Le Callet. 2017. Visual attention modeling for stereoscopic video: A benchmark and computational model. *IEEE Trans. Image Process.* 26, 10 (2017), 4684–4696.

[10] Lutz Goldmann, Francesca De Simone, and Touradj Ebrahimi. 2010. Impact of acquisition distortion on the quality of stereoscopic images. In *Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics.*

[11] Nicolas S. Holliman, Neil A. Dodgson, Gregg E. Favalora, and Lachlan Pockett. 2011. Three-dimensional displays: A review and applications analysis. *IEEE Trans. Broadcast.* 57, 2 (2011), 362–371.

[12] Qiuping Jiang, Feng Shao, Gangyi Jiang, Mei Yu, and Zongju Peng. 2017. Visual comfort assessment for stereoscopic images based on sparse coding with multi-scale dictionaries. *Neurocomput.* 252 (2017), 77–86.

[13] Qiuping Jiang, Feng Shao, Weisi Lin, and Gangyi Jiang. 2016. On predicting visual comfort of stereoscopic images: A learning to rank based approach. *IEEE Signal Process. Lett.* 23, 2 (2016), 302–306.

[14] Cheolkon Jung, Lihui Cao, Hongmin Liu, and Joongkyu Kim. 2015. Visual comfort enhancement in stereoscopic 3D images using saliency-adaptive nonlinear disparity mapping. *Disp.* 40 (2015), 17–23.

[15] Seung-Won Jung and Sung-Jea Ko. 2012. Depth sensation enhancement using the just noticeable depth difference. *IEEE Trans. Image Process.* 21, 8 (2012), 3624–3637.

[16] Yong Ju Jung, Hosik Sohn, Seong-Il Lee, Hyun Wook Park, and Yong Man Ro. 2013. Predicting visual discomfort of stereoscopic images using human attention model. *IEEE Trans. Circuits Syst. Video Technol.* 23, 12 (2013), 2077–2082.

[17] Bochao Kan, Yan Zhao, and Shigang Wang. 2018. Objective visual comfort evaluation method based on disparity information and motion for stereoscopic video. *Opt. Express* 26, 9 (2018), 11418–11437.

[18] Taewan Kim, Sanghoon Lee, and Alan Conrad Bovik. 2015. Transfer function model of physiological mechanisms underlying temporal visual discomfort experienced when viewing stereoscopic 3D images. *IEEE Trans. Image Process.* 24, 11 (2015), 4335–4347.

[19] M. Lambooij, M.J. Murdoch, Wijnand A. IJsselsteijn, and Ingrid Heynderickx. 2013. The impact of video characteristics and subtitles on visual comfort of 3D TV. *Disp.* 34, 1 (2013), 8–16.

[20] Marc T.M. Lambooij, W.A. IJsselsteijn, and Ingrid Heynderickx. 2007. Visual discomfort in stereoscopic dsplays: A review. *Proceedings of SPIE, 2007 vol. 6490.*

[21] Manuel Lang, Alexander Hornung, Oliver Wang, Steven Poulakos, Aljoscha Smolic, and Markus Gross. 2010. Nonlinear disparity mapping for stereoscopic 3D. In *ACM Trans. Graph.* 29, 4 (2010), 75:1–10.

[22] Jianjun Lei, Cuicui Zhang, Yuming Fang, Zhouye Gu, Nam Ling, and Chunping Hou. 2015. Depth sensation enhancement for multiple virtual view rendering. *IEEE Trans. Multimedia* 17, 4 (2015), 457–469.

[23] Xingang Liu, Kai Kang, and Yinbo Liu. 2016. Stereoscopic image quality assessment based on depth and texture information. *IEEE Syst. J.* 2016, DOI: 10.1109/JSYST.2015.2478119.

[24] H. Oh and S. Lee. 2016. Visual presence: Viewing geometry visual information of UHD S3D entertainment. *IEEE Trans. Image Process.* 25, 7 (2016), 3358–3371.

[25] Jincheol Park, Sanghoon Lee, and Alan Conrad Bovik. 2014. 3D visual discomfort prediction: Vergence, foveation, and the physiological optics of accommodation. *IEEE J. Sel. Top. Signal Process.* 3, 8 (2014), 415–427.

[26] Jincheol Park, Heeseok Oh, Sanghoon Lee, and Alan Conrad Bovik. 2015. 3D visual discomfort predictor: Analysis of disparity and neural activity statistics. *IEEE Trans. Image Process.* 24, 3 (2015), 1101–1114.

[27] Jérôme Perrin, Philippe Fuchs, Corinne Roumes, and Francois Perret. 1998. Improvement of stereoscopic comfort through control of the disparity and of the spatial frequency content. In *Aerospace/Defense Sensing and Controls.* International Society for Optics and Photonics, 124–134.

[28] Jenny Read. 2005. Early computational processing in binocular vision and depth perception. *Prog. Biophys. Mol. Biol.* 87, 1 (2005), 77–108.

[29] Feng Shao, Qiuping Jiang, Randi Fu, Mei Yu, and Gangyi Jiang. 2016. Optimizing visual comfort for stereoscopic 3D display based on color-plus-depth signals. *Opt. Express* 24, 11 (2016), 11640–11653.

[30] Feng Shao, Zhutuan Li, Qiuping Jiang, Gangyi Jiang, Mei Yu, and Zongju Peng. 2015. Visual discomfort relaxation for stereoscopic 3D images by adjusting zero-disparity plane for projection. *Displ.* 39 (2015), 125–132.

[31] Feng Shao, Wenchong Lin, Randi Fu, Mei Yu, and Gangyi Jiang. 2017. Optimizing multiview video plus depth retargeting technique for stereoscopic 3D displays. *Opt. Express* 25, 11 (2017), 12478–12492.

[32] Feng Shao, Weisi Lin, Zhutuan Li, Gangyi Jiang, and Qionghai Dai. 2017. Toward simultaneous visual comfort and depth sensation optimization for stereoscopic 3-D experience. *IEEE Trans. Cybern.* 47, 12 (2017), 4521–4533.

[33] V. De Silva, A. Fernando, S. Worrall, H. K. Arachchi, and A. Kondoz. 2011. Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-D displays. *IEEE Trans. Multimedia* 13, 3 (2011), 498–506.

[34] Hosik Sohn, Yong Ju Jung, Seong-Il Lee, Filippo Speranza, and Yong Man Ro. 2014. Visual comfort amelioration technique for stereoscopic images: Disparity remapping to mitigate global and local discomfort causes. *IEEE Trans. Circuits Syst. Video Technol.* 24, 5 (2014), 745–758.

[35] Wa James Tam, Filippo Speranza, Sumio Yano, Koichi Shimono, and Hiroshi Ono. 2011. Stereoscopic 3D-TV: Visual comfort. *IEEE Trans. Broadcast.* 57, 2 (2011), 335–346.

[36] Shuai Wan, Bo Chang, and Fuzheng Yang. 2013. Viewing experience of 3D movie with subtitles: Where to put subtitles in a 3D movie? In *5th International Workshop on Quality of Multimedia Experience (QoMEX'13).* IEEE, 170–175.

[37] Yan Wu, Qi Li, Jingjing Yang, Xiujun Li, and Ning Gao. 2016. Study on the impact of the depth of 3D subtitles on visual comfort. In *3rd International Conference on Information Science and Control Engineering (ICISCE'16).* IEEE, 659–663.

[38] Guanghui Yue, Chunping Hou, Ke Gu, Tianwei Zhou, and Guangtao Zhai. 2019. Combining local and global measures for DIBR-synthesized image quality evaluation. *IEEE Trans. Image Process.* 28, 4 (2019), 2075–2088.

[39] Guanghui Yue, Chunping Hou, Qiuping Jiang, and Yang Yang. 2018. Blind stereoscopic 3D image quality assessment via analysis of naturalness, structure, and binocular asymmetry. *Signal Process.* 150 (2018), 204–214.

[40] Guanghui Yue, Chunping Hou, Kaining Lu, Dandan Feng, and Yao Li. 2016. Subjective visual comfort assessment based on fusion time for depth information. In *2016 11th International Conference on Computer Science & Education (ICCSE).* IEEE, 733–737.

[41] Li Zhang, Ya-Qin Zhang, Jing-Shang Zhang, Liang Xu, and Jost B Jonas. 2013. Visual fatigue and discomfort after stereoscopic display viewing. *Acta Ophthalmol.* 91, 2 (2013).

Received July 2018; revised February 2019; accepted April 2019