Comfortable Disparity Modification for Existing DIBR Systems

HUNG-MING WANG, HSI-CHUN TSENG AND KE-YING LIAO
Department of Electrical Engineering
Institute of Computer and Communication Engineering
National Cheng Kung University
Tainan, 701 Taiwan
E-mail: ming@video5.ee.ncku.edu.tw

In recent years, 3DTV displays have become more popular in the consumer market. However, the visual discomfort is still the common problem when audiences view the stereoscopic 3D video for a long time period. Most of the discomfort comes from the excessive or improper disparity. To ease this problem, the comfortable disparity modification for existing depth-image-based rendering (DIBR) system is proposed in this paper. The proposed method contains several main parts such as degree-dependent algorithm, disparity range determination, and zero-parallax region based disparity remapping. To achieve customized 3D perception, a novel 3D adjustment mechanism is also proposed. Experimental results reveal that the proposed method can adjust the disparity effectively, which will be a great help for existing DIBR systems to generate comfortable 3D videos.

Keywords: 3D video, visual discomfort, comfortable disparity modification, DIBR, 3D perception

1. INTRODUCTION

The developments of television (TV) terminals are from black-and-white TV, color TV, high-definition (HD) TV, to three-dimensional (3D) TV. The developing trend reflects the desires of human visual experiences becoming high quality and realistic 3D presentation now. To generate the 3D contents, one common way is to use the original 2D video associated with its depth information to generate the desired stereo (or multiview) video, which is the so-called depth-image-based rendering (DIBR) technique [1]. So far, there were many research papers published to improve the performance of DIBR system [2, 3]. However, most of them concentrate on how to prevent artifacts in the synthesized view but not discuss the visual discomfort in DIBR systems.

For the 3D perception, the human visual system (HVS) relies on a lot of cues for estimating distance, depth, and shape of objects located in the three dimensional space [4]. Visual cues can be classified into monocular and binocular cues. For monocular visual system, the image of object from an eye or a light sensor provides the information of occlusion, linear perspective, texture gradient and relative size. For binocular visual system, we can perceive the world in the three dimensions. The small difference between the viewpoints of our two eyes is called binocular disparity [5, 6] which is a form of depth perception constructed by the human brain. And two different views are fused to a three dimensional object through our brain. To investigate the binocular vision, some just no-
Noticeable difference (JND) models based on depth perception are also proposed [7-9]. The JND model can help to uncover the depth details that are not perceived by human, which can be used for 3D compression, 3D content processing, and 3D quality assessment.

Naturally, the focal lengths (accommodation) and viewing angles (convergence) of human vision will adjust automatically to make the objects of interest fall on the zero-parallax setting (ZPS) plane. However, when watching the 3D movies, our eyes fixate (convergence) on the virtual 3D-object but focus (accommodation) on the screen. In other words, the accommodation distance is constant but convergence distance varies with the disparity (as shown in Fig. 1), which leads to the accommodation-convergence conflict. This decoupling may potentially be the primary reason of visual discomfort [10].

Besides the accommodation-convergence conflict, excessive binocular disparity is also another reason for causing visual discomfort. According to Panum’s fusional area [11, 12], there is a fusion limit for human eyes. There are many factors to affect limits of Panum’s fusional area, including stimulus size, spatial frequency, temporal modulation of disparity information, exposure duration, continuous features, temporal effects, the amount of luminance, and individual differences. The limits increase with larger moving objects and decrease with smaller, detailed and stationary objects. Hence, there is a rule-of-thumb for Panum’s fusional area, which is defined within the one degree of disparity as shown in Fig. 2. For the stereoscopic image, if the disparity between two corresponding points is too excessive, e.g. beyond Panum’s fusional area, the human eyes could not

---

Fig. 1. The change of accommodation and convergence distances when (a) watching natural 3D object and (b) watching virtual 3D object.

Fig. 2. The diagram of Panum’s fusion area.
fuse the corresponding points in different views correctly. Thus, it will produce double images effect in our vision such that it will cause the visual discomfort.

To solve the above problems, a comfortable disparity modification mechanism is proposed in this paper. The rest of this paper is organized as follows. First, we will give a brief overview of the proposed system and then address the detailed functionalities in Section 2. Experimental results to demonstrate the effectiveness of the proposed algorithm are shown in Section 3. Finally, we conclude the paper in Section 4.

2. THE PROPOSED DISPARITY MODIFICATION SYSTEM

Fig. 3 shows the flowchart of the proposed disparity modification system. The system contains four main parts such as display parameter consideration, degree-dependent algorithm, disparity range determination and depth to disparity remapping. The inputs are color frames with the associated depth information. We first adjust disparity map according to display parameters such as screen size, resolution, and viewing distance. There are different limitations of human vision with different 3D displays. To achieve the best perception, we have to know the parameters in advance. To ensure that disparity range is not over Panum’s fusional area to avoid excessive binocular disparity, we calculate the maximum distances outside and inside the screen and convert the original depth to the disparity map. The converted disparity map is then further adjusted by the proposed zero-parallax region based disparity remapping to avoid accommodation-convergence conflict. In the zero-parallax region based disparity remapping process, we will detect the interesting regions in the videos to determine the suitable ZPS plane. To be customized, we also provide an extra functionality to adjust disparity range by viewers. Viewers can adjust the disparity by themselves to reach the desired 3D effect under the designed disparity range.

2.1 Display Parameters Consideration

It is noted that pixel value in the depth map is only a reference value which indicate
the relative distance. The depth map is only the intermediate information to generate the virtual views from original view. To generate virtual views through DIBR technique, depth map is usually converted to disparity map for warping procedure. The common formula from depth to disparity is given below:

\[ D(i, j) = \frac{d(i, j) - d_0}{S} \]  

(1)

where \( D(i, j) \) is the disparity value at coordinate \((i, j)\) in disparity map, \( d(i, j) \) is the depth value at the coordinate \((i, j)\) in depth map. \( d_0 \) is the shift factor which is used for zero-parallax setting (ZPS) and \( S \) is the scale factor which determines level of disparity. If we choose appropriate \( d_0 \) and \( S \), we can convert depth to disparity and generate the corresponding 3D contents.

However, the optimal values \( d_0 \) and \( S \) will vary with different displays since display parameters such as the best viewing distance, screen size, aspect ratio, horizontal resolution and screen width are different. For instance, the optimal solutions \( d_0 \) and \( S \) for 27” display might not be suitable for 55” display. So we need to consider display parameters to help us find the optimal conversion.

2.2 Degree-dependent Algorithm

We consider display parameters to calculate acceptable disparity range to avoid excessive binocular disparity which makes people feel discomfort. Although the Panum’s fusional area is defined within the one degree of disparity range, it only works when the left and right view are perfect. Unfortunately, it is difficult to reach this situation when DIBR techniques are adopted to generate the left and right views. The actual degree \( \alpha \) should depend on the quality of synthesized views. For simplicity, we empirically set \( \alpha \) to 0.5 degree in our system. To decide the disparity range, we need to obtain the maximum distances (cm) \( l_{out} \) and \( l_{in} \) for outside and inside the screen individually, as shown in Fig. 4.

![Fig. 4. The geometric representation of viewing condition.](image)
The value of $l_{out}$ and $l_{in}$ can be calculated by the following equations

$$l_{out} = l - \frac{a}{2 \tan \left( \frac{2 \theta + \alpha}{2} \right)} ,$$

$$l_{in} = \frac{a}{2 \tan \left( \frac{2 \theta - \alpha}{2} \right)} - l ,$$

where $\theta = \tan^{-1} \left( \frac{a}{2l} \right)$ is the viewing angle and $a$ is the average pupil distance set to 6.5 cm in this paper.

Table 1 shows the limitations for comfortable viewing distances at different viewing distances in perfect situation. The parameter $\alpha$ is set to 0.5 in the calculation. We can find that comfortable viewing distance become smaller when viewers are close to the screen, and become larger when viewers are far from the screen.

<table>
<thead>
<tr>
<th>Viewing distance $l$ (cm)</th>
<th>$l_{out}$ (cm)</th>
<th>$l_{in}$ (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>100</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>200</td>
<td>43</td>
<td>73</td>
</tr>
<tr>
<td>300</td>
<td>87</td>
<td>202</td>
</tr>
<tr>
<td>500</td>
<td>200</td>
<td>1000</td>
</tr>
</tbody>
</table>

2.3 Disparity Range Determination

After getting $l_{in}$ and $l_{out}$ from degree-dependent algorithm, we then convert them to two corresponding disparities (pixels), which are $D_{in}$ and $D_{out}$. The relationship of $D_{in}$ and $D_{out}$ with display parameters are shown in Fig. 4, which could be calculated as follows

$$D_{out} = \frac{al_{out}}{l_{out}-l} \frac{W_p}{W_c} ,$$

$$D_{in} = \frac{al_{in}}{l+l_{in}} \frac{W_p}{W_c} ,$$

where $W_p$ and $W_c$ represent the screen width in pixels and centimeters, separately. In (4) and (5), it is noted that the signs of $D_{in}$ and $D_{out}$ are opposite since the warping directions are reversed.

2.4 Zero-parallax Region based Disparity Remapping

If we have two maximum disparities inside and outside the screen, we can limit the
disparity range to avoid excessive binocular disparity. However, to avoid accommodation-convergence conflict, we should set the depth value of interesting region to be zero-parallax (i.e., on the screen). To achieve this, we can set the shift factor $d_0$ in (1) to the depth value of interesting region. To compute the desired shift factor in real time, the computation should be carefully considered. It is noted that viewers usually focus on actors and actresses in the movies when they are existed. Although this assumption might not be always true, it still works in most cases. Based on the above observation with the practical issue, we adopt face detection to help find the interesting region and determine the value of shift factor. Here, we do not adopt other complex salient region detection methods [13-15] are due to the consideration of computation time. The flowchart of shift factor determination is shown in Fig. 5.

![Flowchart of shift factor determination](image)

The inputs are color and depth frames with 960×540 resolution. We think that important actors or actresses are usually in the central area of color frame. They occupy a high proportion in the central area. So we only retain central area to reduce the computation. As for face detection, there are already many researches work on this area [16-19]. However, most of them occupy a lot of computation which cannot be used in real-time applications. Viola et al. [20] propose the concept of integral image based on Ada Boost which is to train facial detection by cascade of classifiers to build a facial detection system in real time. For the real-time issue, we simply adopt [17] but modify some parameters such as scale factor and minimum object size to speed up the process of face detection. In our experiments, the scale factor is set to 1.4 and the minimum object size is set to 70×70.

After face detection, we can find the central positions of the faces in color frame. The central positions are found by finding the center of different rectangle windows. By using the central positions in the depth frame, we can find the largest depth value of positions and set it to be the current shift factor. Here, we choose the largest depth value as the shift factor is because that people in the movie are usually the foreground part. It is
noted that large variation of shift factor will makes people feel discomfort. To deal with this problem, we only increase or decrease a small step from the previous value of shift factor instead of substituting it directly. The update equation is shown as follows

\[ S_t = S_{t-1} + \text{sign}(S_{curr} - S_{t-1}) \cdot \epsilon, \]

where \( S_t \) is the updated shift factor in time \( t \), \( S_{t-1} \) is the updated shift factor in previous time, \( S_{curr} \) is the current calculated shift factor in time \( t \), and \( \text{Sign}(\cdot) \) denotes the sign function, whose output is equal to one when \( x \) is larger than zero and equal to minus one when \( x \) is smaller than zero. In Eq. (6), \( \epsilon \) is the pre-determined step size for changing shift factor.

In general, we can use Eq. (1) to convert depth to disparity, a new depth to disparity remapping is proposed according to \( D_{out}, D_{in} \) and \( d_m \). We use \( D_{out} \) and \( D_{in} \) to limit disparity range to solve the excessive binocular disparity problem. And shift factor is set on the region of visual focus to overcome the accommodation-convergence conflict. Fig. 6 shows a simple depiction to illustrate remapping method. The depth values of 0, shift factor, and 255 are remapped to \(-D_{in}\), 0 and \(-D_{out}\) disparity values, respectively. It is noted that since \( D_{out}\) and \( D_{in}\) are opposite, we inverse them to make it more straightforward. The remapping formula is then described as

\[ D(i, j) = \begin{cases} 
-1 \cdot \frac{d(i, j) \cdot D_{out}}{d_m + D_{in}}, & \text{if } d(i, j) \leq d_m \\
\frac{(d(i, j) - d_m) \cdot D_{out}}{(255 - d_m)}, & \text{otherwise}
\end{cases} 
\]

where \( d(i, j) \) is the depth value at the coordinate \((i, j)\) in depth map, \( d_m \) is the shift factor which is set to be zero parallax, and \( D_{out} \) and \( D_{in} \) are the maximum disparities outside the screen and inside the screen.

2.5 Manual Adjustment

After depth to disparity remapping, we can get adjusted disparity map by automatic procedure. However, the perception of visual discomfort is varied with different people. It is better to allow the parameters adjustable by viewers when they feel visual discomfort or want more 3D effects. In common 3DTVs, only the qualitative adjustment is available, which means viewers can only increase or decrease the 3D effect. By considering the display parameters, the proposed system can further make the viewers to adjust the 3D range more precise and quantitative (in centimeter) via adjusting \( l_{in} \) and \( l_{out} \). For
human vision, the adjusted disparity range is still restricted in the safe range. The safe range for $l_{in}$ and $l_{out}$ are constrained by $L'_1$ and $L'_2$, which are shown in Fig. 7. The adjusted $L'_1$ and $L'_2$ are described as follows:

$$L'_1 = \{r_{out} \ l_{out} \ | \ 0 \leq r_{out} \leq k_1\}, \quad (8)$$

$$L'_2 = \{r_{in} \ l_{in} \ | \ 0 \leq r_{out} \leq k_2\}, \quad (9)$$

where $l_{out}$ and $l_{in}$ are the same as we mentioned in the degree-dependent algorithm with the controlled parameters $k_1$ and $k_2$ for disparity adjustment. After we have $L'_1$ and $L'_2$, we then can calculate the corresponding $D_{out}$ and $D_{in}$ according to Eqs. (4) and (5). Empirically, we set the values of $k_1$ and $k_2$ to 1.5 in the proposed system.

Besides the adjustment for maximum inside and outside distances, we also provide another function for viewing distance adjustment. Although there could be the best viewing distance suggested for different 3D displays, people might still move from the screen when viewing 3DTVs. We adjust viewing distance from the screen by parameter $L'$, which is shown in Fig. 8. The formula for adjusting $L'$ is

$$L' = \{l' \ | \ q_{min}l \leq l' \leq q_{max}l\}, \quad (10)$$

where $l$ is the best viewing distance for a particular 3D display while $q_{min}$ and $q_{max}$ denote the minimum and maximum adjusted coefficients, respectively. Empirically, the coefficient $q_{min}$ is 0.9 and $q_{max}$ is 1.1 in our experiments.
In Fig. 8, $L'$ and $L''$ means the available farthest and nearest adjustment distance between the viewer and the screen. After adjusting the viewing distance, two maximum distances and corresponding disparities are also needed to be updated.

### 2.6 Multi-view Rendering

After we get color image and depth map, 3D contents are produced by DIBR technique. Now, we briefly explain the main concepts of the DIBR system. A typical procedure can be illustrated by the block diagram which is shown in Fig. 9. The system includes three important parts, pre-processing of depth map, 3D image warping and hole-filling. Pre-processing is applied to generate a smooth depth map for reducing the sharp horizontal transitions, which will cause the big holes in the synthesized view after 3D image warping. After image warping, if there are still holes caused by occlusion or inaccurate depth map, the hole-filling process will be performed to recover missing pixels.

![Block diagram of DIBR system.](image)

In our experiments, after obtaining color frame and adjusted disparity map, we will first treat the original color from as the left view and perform 3D image warping to generate the right view. Afterward, the hole-filling will be applied to recover the missing pixels. For simplicity, we recover the hole pixels by copying to the nearest non-hole pixel in the right side (background part).

### 3. EXPERIMENTAL RESULTS

The proposed disparity modification systems are implemented by using Microsoft Visual Studio 2008. It is tested on Intel(R) Core(TM)2 Duo CPU E8400 @ 3.00GHz with 2GB memory in Windows 7 32-bit OS. The implemented computer is equipped with NVIDIA GeForce GTS 450 GPU. The test sequences are Philips Espresso sequence downloaded from Philips WOWvx website and MPEG test sequences [21] with Newspaper, Undo Dancer, and Poznan Hall. The detected results for different scenes are shown in Fig. 10. In each scene, we show the color image of the first frame, its corresponding depth map, results after face detection, and the detected zero-parallax regions, respectively. The results of zero-parallax region are presented by retaining the regions with the value of determined shift factor. From Fig. 10, we can see that we can successfully find the faces and find the reasonable ZPS plane.

The comparisons between the estimated shift factors $d_m$ and ideal $d_m$ for Philips Espresso sequence are shown in Fig. 11. The red line demonstrates the ideal $d_m$, which is acquired by human eyes and the blue line demonstrates the estimated shift factor $d_m$, which is extracted by our proposed method. The ideal shift factor is acquired by three
people with their visual attention areas. If there are differences between their results, the majority will be chosen. From Fig. 11, we can see that the proposed method is suitable for obvious actors or actresses in the scenes. It is known that the large variation of shift factor makes people feel discomfort. So we use small variation to update the shift factor. This makes the curve of shift factor gradually changes without an abrupt drop.

We use *anaglyph* images to demonstrate the proposed disparity modification through DIBR technique. There are two parameters, shift factor and scale, in our DIBR technique. However, the optimal values of shift factor and scale vary with different scenes and displays. By the proposed disparity modification, the shift factor and comfort disparity range are calculated automatically according to environmental display parameters. The proposed method is compared with constant shift factor and scale. It proves that we can find suitable stereoscopic effect by the proposed method instead of trying different parameters step by step based on different displays.

The experimental results with and without disparity modification in *anaglyph* images for Philips Espresso sequence (scenes 1~4) and MPEG test sequences (scenes 5~7) are compared in Fig. 12. For comparisons, we also set the shift factor to 128 and scale to 5, 10 and 20. From Fig. 12, we can find that the disparity of the interesting regions (e.g., actor’s or actress’s face) is reduced by applying the proposed disparity modification.

<table>
<thead>
<tr>
<th>Color Image</th>
<th>Depth Map</th>
<th>Face Detection</th>
<th>Zero-parallax Region</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Color Image" /></td>
<td><img src="image2.png" alt="Depth Map" /></td>
<td><img src="image3.png" alt="Face Detection" /></td>
<td><img src="image4.png" alt="Zero-parallax Region" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Color Image" /></td>
<td><img src="image6.png" alt="Depth Map" /></td>
<td><img src="image7.png" alt="Face Detection" /></td>
<td><img src="image8.png" alt="Zero-parallax Region" /></td>
</tr>
<tr>
<td><img src="image9.png" alt="Color Image" /></td>
<td><img src="image10.png" alt="Depth Map" /></td>
<td><img src="image11.png" alt="Face Detection" /></td>
<td><img src="image12.png" alt="Zero-parallax Region" /></td>
</tr>
</tbody>
</table>

Fig. 10. The detected results for different scenes.
COMFORTABLE DISPARITY MODIFICATION FOR EXISTING DIBR SYSTEMS

(1) Shift factor=128. Scale=5                      (2) Shift factor=128. Scale=10

Fig. 12. Comparisons results with and without the proposed disparity modification.

Color Image | Depth Map | Face Detection | Zero-parallax Region
---|---|---|---

Fig. 10. (Cont’d) The detected results for different scenes.

Fig. 11. Comparisons of the estimated \(d_m\) and ideal \(d_{eq}\).

(1) Shift factor=128. Scale=5                      (2) Shift factor=128. Scale=10

Fig. 12. Comparisons results with and without the proposed disparity modification.
Fig. 12. (Cont’d) Comparisons results with and without the proposed disparity modification.

(a) Scene 1

(b) Scene 2

(c) Scene 3

Fig. 12. (Cont’d) Comparisons results with and without the proposed disparity modification.
Fig. 12. (Cont’d) Comparisons results with and without the proposed disparity modification.

(d) Scene 4

(1) Shift factor=128. Scale=5
(2) Shift factor=128. Scale=10

(3) Shift factor=128. Scale=20

(4) Proposed disparity modification

(5) Shift factor=128. Scale=5
(6) Shift factor=128. Scale=10

(7) Shift factor=128. Scale=20

(8) Proposed disparity modification

(e) Scene 5

(9) Shift factor=128. Scale=5
(10) Shift factor=128. Scale=10

Fig. 12. (Cont’d) Comparisons results with and without the proposed disparity modification.
Fig. 12. (Cont’d) Comparisons results with and without the proposed disparity modification.

4. CONCLUSIONS

In this paper, a comfortable disparity modification mechanism is proposed. The proposed system is designed to deal with the visual discomfort caused by accommodation-convergence conflict and excessive binocular disparity. Experimental results show that the disparity can be adjusted suitably, which can help existing DIBR technique generate more comfortable 3D videos.

REFERENCES

3. L.-M. Po, S. Zhang, X. Xu, and Y. Zhu, “A new multidirectional extrapolation hole-
Hung-Ming Wang (王泓民) was born in Tainan, Taiwan, in 1981. He received the B.S. and Ph.D. degrees in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan, in 2003 and 2009, respectively. He is currently an Assistant Research Fellow of Center for Tomorrow Ubiquitous Cloud and Hypermedia Services in the NCKU. His major research interests include image processing, video coding, multimedia communication, and 2D to 3D conversion.

Hsi-Chun Tseng (曾喜駿) was born in Taoyuan, Taiwan, in 1987. He received the B.S. degree in Communication Engineering from National Central University, Taoyuan, Taiwan, in 2011, and M.S. degree in computer and communication engineering from NCKU in 2013. He is currently working in FocalTech Systems Corporation in Hsinchu. His major research interests include digital image processing and 3D video processing.

Ke-Ying Liao (廖恪應) was born in Taipei, Taiwan, in 1980. He received the B.S. degree in Engineering Science from National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan, in 2002, M.S. degree in Electrical Engineering from National Cheng Kung University (NCKU), Tainan, Taiwan, in 2004, and Ph.D. degree in Institute of Computer and Communication Engineering from NCKU, Tainan, Taiwan, in 2010, respectively. His major research interests include SoC, image processing, transcoding and video coding. Currently, he is a Postdoctoral Fellow of Center for Tomorrow Ubiquitous Cloud and Hypermedia Services in the NCKU.