VISUAL FATIGUE EVALUATION AND ENHANCEMENT FOR 2D-PLUS-DEPTH VIDEO

Jaeseob Choi, Donghyun Kim, Bumsub Ham, Sunghwan Choi and Kwanghoon Sohn

Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea
khsohn@yonsei.ac.kr

ABSTRACT

A 3D video is expected to be a representative technique of realistic system but still has some problems such as visual fatigue and headache. In this paper, we propose a visual fatigue evaluation algorithm to predict the degree of visual fatigue from a 2D-plus-depth video. Spatial and temporal characteristics of the depth video are main factors of visual fatigue for autostereoscopic displays. Using depth image directly, we estimate spatial and temporal complexities, depth position and scene movement of the 3D video. Then, the overall visual fatigue of the 3D video is evaluated to have higher correlation with subjective fatigue evaluation by a linear regression. Moreover we control the pixel value of depth image from the 3D video which may induce severe fatigue to make more comfortable 3D video. The results of proposed algorithm show a considerable correlation with subjective visual fatigue.

Index Terms— 3D video, 2D-plus-depth, visual fatigue

1. INTRODUCTION

Three dimensional TV is expected to be a representative technique of realistic broadcasting system. Although 3D image/video processing and 3D display technology have developed considerably, there had not been enough effort on considering a human visual characteristics and reducing visual fatigue. Many people agree that visual fatigue from viewing 3D image/video is a main bottleneck of 3D industry. Generally, visual fatigue is caused by contradiction of the convergence and accommodation, differences in properties of the left and right images and excessive spatial/temporal complexity [1-8]. Particularly, after the photosensitive seizure was occurred by incomplete TV program in Japan, a safety of video contents has received wide attention. Therefore a 3D video fatigue evaluation and a related standardization are essential to guarantee the safety of 3D video and facilitate 3D industry.

Many researches have been done to evaluate the visual fatigue of 3D image/video, particularly stereoscopic image/video. Nojiri *et al.* described the effects of the range of parallax distribution and temporal parallax changes using stereoscopic HDTV [1]. Yano *et al.* showed the relationship between convergence eye movement and the accommodation function, and evaluated the visual fatigue from viewing motion stereoscopic images [2]. They also measured the change of visual fatigue in accommodation before and after viewing images, and evaluated both degree of parallax and amount of motion in stereoscopic images [3]. Emoto *et al.* determined the proper index of visual fatigue for quantitative analysis and evaluated the capability that viewers can recover the visual fatigue [4][5]. In addition, Speranza *et al.* investigated the relationship between binocular disparity, object motion and visual comfort using computer-generated stereoscopic video [6]. In our previous work, Kim *et al.* proposed a depth image quality metric for stereoscopic and autostereoscopic video using depth range, vertical misalignment and temporal consistency [7]. Leon *et al.* showed that motion and complexity of the depth image have a strong influence on the acceptable depth quality in 3D videos using the autostereoscopic 3D display [8]. According to the previous studies, visual fatigue is occurred by the excessive parallax, fast motion of objects and the complex scene. However, most of visual fatigue evaluations are limited to subjective evaluation.

In this paper, we propose a visual fatigue evaluation algorithm for 2D-plus-depth video of autostereoscopic 3D display. The reason why we use an autostereoscopic 3D display is that a stereoscopic display has limited FOV (field of view) and inconvenience of wearing glasses. Moreover, depth image can be directly applied to several applications such as freeview TV, DIBR (depth image-based rendering), etc [9]. MPEG also focus on compression of 2D-plus-depth video [10]. We evaluate spatial and temporal characteristics to estimate the visual fatigue using depth image. After the visual fatigue evaluation, a 3D video which may induce severe fatigue is adjusted to comfort video by depth image control. The block diagram of the proposed visual fatigue evaluation is shown in Fig. 1.

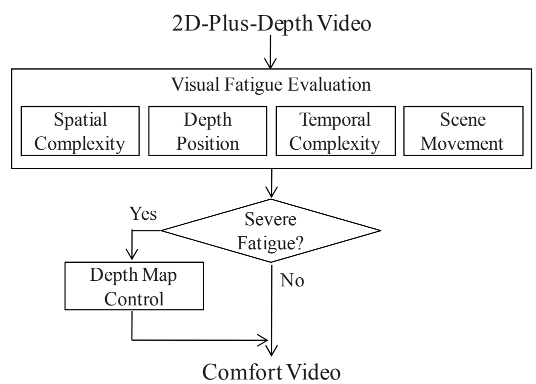


Fig. 1. Block diagram of proposed algorithm

2. 3D VIDEO VISUAL FATIGUE EVALUATION

Table 1. Physical meaning of visual fatigue evaluation factors

fatigue ⟷				**comfort**
complex scene (multiple objects)	-	Spatial complexity	-	simple scene (few objects)
outside comfort zone	-	Depth position	-	inside comfort zone
dynamic motion	-	Temporal complexity	-	static motion
high velocity	-	Scene movement	-	low velocity

We define 3D video characteristics - spatial complexity, depth position, temporal complexity and scene movement. Physical meanings of these factors are shown in Table 1. We assume that depth images are reliable and do not have noise or artifact.

2.1. Spatial complexity

The spatial complexity factor measures the complexity of scene composition. If multiple objects are distributed in many depth levels, it is difficult to focus on every object simultaneously. The spatial complexity is calculated by the depth pixel variation in a frame as in Eq. (1). An exponential function is used to normalize from 0 to 1. Moreover, we have divided the depth image into nine domain and assigned different weights as shown in Fig.2 because viewers tend to concentrate a center region [5].

$$C_{spatial} = \frac{1}{N}\sum_{i \in N} e^{-\left(\frac{\frac{1}{mn}\sum_{x \in m}\sum_{y \in n} w(x,y)\left(f_i(x,y) - E\left(f_i(x,y)\right)\right)^2}{\sigma_s^2}\right)} \tag{1}$$

0.95	1.0	0.95
1.0	1.2	1.0
0.95	1.0	0.95

Fig. 2. Weight of depth image

,where $C_{spatial}$ defines the spatial complexity, N is the number of frames, m, n are vertical/horizontal size of depth image, $f_i(x,y)$ is depth pixel value at (x,y), $E(f(x,y))$ is an average of depth pixel values in one frame and $w(x,y)$ is a weight function. $C_{spatial}$ becomes close to 0 when the depth composition is complicated, and becomes close to 1 when the depth composition is simple.

2.2. Depth position

A comfort zone is defined as the area which is closed to the display plane as shown in Fig. 3 [11]. Objects outside the comfort zone make difficult to synthesize a depth and generate crosstalk and quality degradation. This factor measures the average distance from the display plane as in Eq. (2).

$$A_{spatial} = \frac{1}{N}\sum_{i \in N} e^{-\left(\frac{E\left(f_i(x,y) - D_{display}\right)^2}{\rho_s^2}\right)} \tag{2}$$

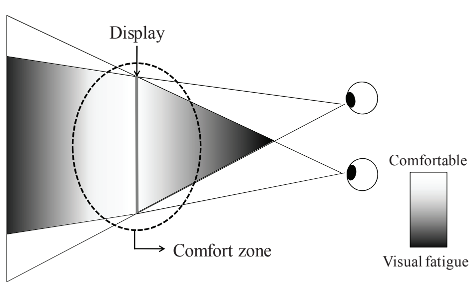


Fig. 3. Concept of comfort zone

,where $A_{spatial}$ defines the depth position and $D_{display}$ represents depth of display plane ($D_{display}$=128 in this experiment). $A_{spatial}$ becomes close to 0 when the depth position is far from the comfort zone, and becomes close to 1 when the depth position is inside the comfort zone.

2.3. Temporal complexity

The temporal complexity determines the amount of motion in a scene. Many objects which have different direction and velocity make difficult to synthesize depth in brain and cause visual fatigue. This factor is determined by variation of difference between consecutive depth frames. To calculate a temporal factor unaffected by spatial properties, range of depth image must be extended to 0~255 at first.

$$C_{temporal} = \frac{1}{N-1}\sum_{i \in N-1} e^{-\left(\frac{\frac{1}{mn}\sum_{x \in m}\sum_{y \in n} w(x,y)\left(d_i(x,y) - E\left(d_i(x,y)\right)\right)^2}{\sigma_t^2}\right)} \tag{3}$$

,where $C_{temporal}$ is the temporal complexity and $d_i(x,y)$ is different value between $f_i(x,y)$ and $f_{i+1}(x,y)$. $C_{temporal}$ becomes close to 0 when the scene includes dynamic motions, and becomes close to 1 when the scene includes static motions.

2.4. Scene movement

The scene movement measures the overall speed of the 3D video induced by camera movement and scene change. If a video is fast, it is hard to successively fuse the depth of video. This factor is calculated by an average of absolute different value between consecutive depth frames after the depth range is extended to 0~255.

$$A_{temporal} = \frac{1}{N-1}\sum_{i \in N-1} e^{-\left(\frac{E\left(\left|d_i(x,y)\right|\right)^2}{\rho_t^2}\right)} \tag{4}$$

,where $A_{temporal}$ means the scene movement. $A_{temporal}$ becomes close to 0 when the overall video speed is fast, and becomes close to 1 when the overall video speed is slow.

2.5. Overall visual fatigue measurement

We can predict the overall visual fatigue C of 3D video from a linear combination of previously defined proposed fatigue factors by Eq. (5).

$$C = \alpha C_{spatial} + \beta A_{spatial} + \gamma C_{temporal} + \delta A_{temporal} \tag{5}$$

Coefficients are calculated by least square method which maximizes the correlation between predicted visual fatigue and subjective evaluation.

3. VISUAL COMFORT ENHANCEMENT OF 3D VIDEO

An incomplete 3D video induces a fatigue and a headache. Therefore, the visual fatigue of 3D videos has to be minimized in filming or display step. Since the proposed algorithm use the depth image directly, we can make comfortable video easily by adjusting the depth images.

3.1. Discomfort 3D video detection

We classify the cause of the visual fatigue into three cases. The first is that a scene is too complicated to synthesize depths at once. When we focus on a specific object, other objects can be double-imaged or out-focused. In this case, $C_{spatial}$ factor is very small.

Second, if the movement of video is dynamic, viewers are difficult to synthesize depth of objects simultaneously. Because of insufficient time to feel a depth, viewers experience tiredness and lack of presence. In this case, $C_{temporal}$ and $A_{temporal}$ are small.

Moreover, while some objects are located outside the comfort zone, feeling the depth is unstable due to excessive parallax. In this case, $A_{spatial}$ is small. As shown in Fig. 4, a depth distribution should be placed in the comfort zone for less fatigue, although a 3D effect relatively decreased.

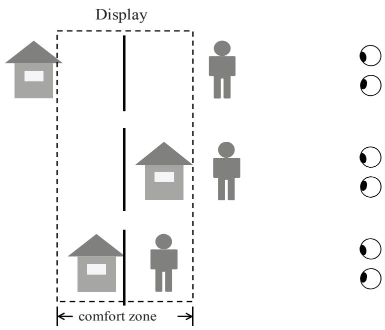


Fig. 4. Depth distribution : (top) small $C_{spatial}$, (middle) large $C_{spatial}$ and small $A_{spatial}$, (bottom) comfort video

3.2. Depth image adjustment

A depth image shows object segmentation accurately. Moreover, a real time depth image adjustment is possible with lower computational complexity. In this paper, when one of four factors is lower than 0.1 or the overall fatigue is lower than 2, we consider that these videos cause severe fatigue. Then, the depth image of severe fatigue video is mapped into comfort zone by Eq. (6).

$$D_{new} = \min_{new} + \frac{\max_{new} - \min_{new}}{\max - \min} \times D \tag{6}$$

,where D and D_{new} are depth value of fatigue and comfort video, max/min and max_{new}/min_{new} are maximum/minimum depth value of fatigue and comfort video, respectively.

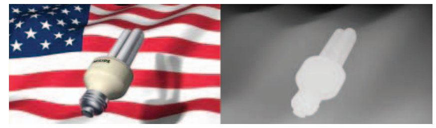
4. EXPERIMENTAL RESULTS

Visual fatigue evaluation is conducted by using 20 inch Philips WOWvx autostereoscopic 3D display with 1600x1200 resolutions. We utilize ten 2D-plus-depth videos which are offered by the Philips [12]. Each video has different characteristics such as the number of objects, depth levels and motion as shown in Fig. 5. In addition, we generated additional video sets scaled as shown in Fig.4 (middle, bottom) to evaluate the effect of depth position. Ten evaluators (eight males and two females) with an average age of 26.8, participated in this study. We used the absolute category rating (ACR) method to assess the visual fatigue by using one of the five levels of scales (not at all (5) – slight fatigue (4) – moderate fatigue (3) – fatigue (2) – severe fatigue (1)) [13].

The proposed algorithm is compared to the algorithm of [3] which measures the magnitude of parallax and motion using the block-matching method as in Eq. (7) and Eq. (8). An amount of parallax is estimated by a depth image because parallax and depth have same meaning.

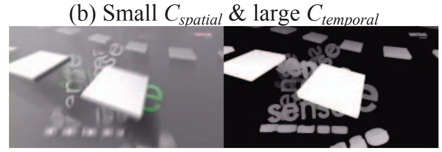
$$I(N) = \sum_{k<D_{display}-\alpha,\, k>D_{display}+\alpha} (k \times S(k)) \Big/ \sum_{D_{display}-\alpha<k<D_{display}+\alpha} (k \times S(k)) \tag{7}$$

$I(N)$ is the magnitude of parallax in frame N, and k is depth level (parallax). $S(k)$ is number of blocks in frame N with measured parallax of k, $D_{display}$=128 and α=48.

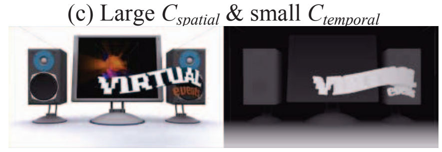


(a) Small $C_{spatial}$ & small $C_{temporal}$

(b) Small $C_{spatial}$ & large $C_{temporal}$



(c) Large $C_{spatial}$ & small $C_{temporal}$



(d) Large $C_{spatial}$ & large $C_{temporal}$

Fig. 5. Test videos

$$M(N) = \sum_{\beta 1<|m|<\beta 2} (m \times P(m)) \Big/ \sum_{0\le|m|<\beta 2} (m \times P(m)) \tag{8}$$

$M(N)$ is the magnitude of motion in frame N, n is an absolute value for the vector sum of the horizontal and vertical motions, $P(m)$ is number of blocks in frame N with measured motion of m, $\beta 1=3$, $\beta 2=sqrt(20^2+10^2)$. To compare with the proposed algorithm, we modified Eq. (7) and Eq. (8) into Eq. (9) and Eq. (10), and final visual fatigue of [3] is also evaluated by the least squared method.

$$I = \frac{1}{N}\sum_{i\in N} e^{-\left(\frac{I(i)}{\sigma_i^2}\right)}, \quad M = \frac{1}{N-1}\sum_{i\in N-1} e^{-\left(\frac{M(i)}{\sigma_m^2}\right)} \qquad (9)(10)$$

Fig. 6 represents the scatter plot of the proposed and compared algorithm with subjective evaluation. As shown in Fig. 6, the proposed algorithm is less scattered than that of [3]. We confirm that the proposed algorithm is better than the compared algorithm in terms of a correlation coefficient R^2 which can determine how scatterings are related to a represented linear optimum function. Correlation coefficient R^2 is calculated by Eq. (11). This result means that considering more properties of 3D video lead to accurate prediction of visual fatigue.

$$R^2 = \frac{\sum (y_i - \bar{y})^2 - \sum (y_i - y_{cal})^2}{\sum (y_i - \bar{y})^2} \tag{11}$$

,where y_i is a fatigue value, $\bar{y}$ is an average of fatigue value and y_{cal} is a optimal function.

Fig. 7 shows results of subjective fatigue evaluation and the proposed fatigue evaluation about thirty test video. Effects of different characteristic of each video are significant. *(*$C_{spatial}$ *:* $F(4,45)=4.82$, $p<0.01$, $A_{spatial}$ *:* $F(4,45)=3.60$, $p<0.05$, $C_{temporal}$ *:* $F(4,35)=8.66$, $p<0.01$, $A_{temporal}$ *:* $F(4,45)=3.60$, $p<0.05$*)*. It shows that the video inside comfort zone is more comfort than another case. Most of the results show a considerable correlation with subjective evaluation results except some cases. Video 4, 9 and 10 show lower correlation with the subjective evaluation result because depth values are not reflect depth image properties efficiently.

5. CONCLUSION

In this paper, we proposed the visual fatigue evaluation algorithm of the 2D-plus-depth video and enhanced the visual comfort. Visual fatigue evaluation factors are spatial/temporal complexity, depth position and scene movement. The proposed visual fatigue algorithm show reliable performance in comparison with the conventional algorithm. Moreover, we reduced visual fatigue by adjusting depth image with low computational complexity.

However, our algorithm has problem that it can be affected by a noise, artifact in the depth image and shape of objects. For the further work, we need to consider these factors to increase the performance of fatigue evaluation.

6. REFERENCES

[1] Y. Nojiro, H. Yamanoue, A. Hanazato, M. Emoto, and F. Okano, "Visual comfort/discomfort and visual fatigue caused by stereoscopic HDTV viewing," *Proc. of SPIE-IS&T Electronic Imaging* vol. 5291, 2004.

[2] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays 25,* pp.141-150, 2004.

[3] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaits, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays 23,* pp.191-201, 2002.

[4] M. Emoto, Y. Nojiri, and F.Okano, "Change in fusional vergence limit and its hysteresis after viewing stereoscopic TV," *Displays 25,* pp.67-76, 2004.

[5] Y. Nojiri, H. Yamanoue, S. Ide, S. Yano, and F. Okano, "Parallax distribution and visual comfort on stereoscopic HDTV," *IBC*, 2006.

[6] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, "Effect of disparity and motion on visual comfort of stereoscopic images," *Proc. of SPIE-IS&T Electronic Imaging* vol. 6055, 2005.

[7] D. Kim, D. Min, J. Oh, S. Jeon, and K. Sohn, "Depth image quality metric for three-dimensional video," *Proc. Of SPIE-IS&T, Electronic Imaging*, pp. 723719_1-713719_2, 2009.

[8] G. Leon, H. Kalva, and B. Furht, "3D video quality evaluation with depth quality variations*," IEEE 3DTV-CON'08*, 2008.

[9] D. Min, D. Kim, S. Yun, and K. Sohn, "2D/3D freeview video generation for 3DTV system," *Proc. IEEE ICIP*, pp. 1760-1763, 2008.

[10] ISO/IEC JTC 1/SC 29/WG 11, "Joint Video Specification (ITU-T Rec. H.264|ISO/IEC 14496-10 AVC)." JVT Document E146d34, Geneva, Switzerland, 2002.

[11] Bernard Mendiburu, *3D Movie Making,* Focal Press, 2009.

[12] http://www.wowvx.com/video.html

[13] "Subjective assessment of stereoscopic television pictures," ITU-R Recommendation BT.1438, 2000.

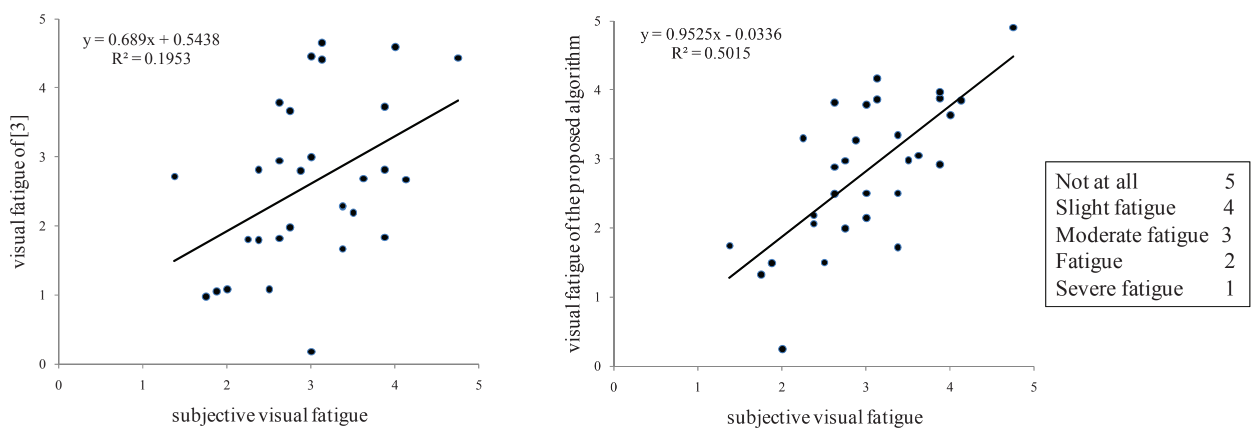


Fig. 6. (a) Scatter plot of visual fatigue of [3] and subjective visual fatigue
(b) Scatter plot of visual fatigue of the proposed algorithm and subjective visual fatigue

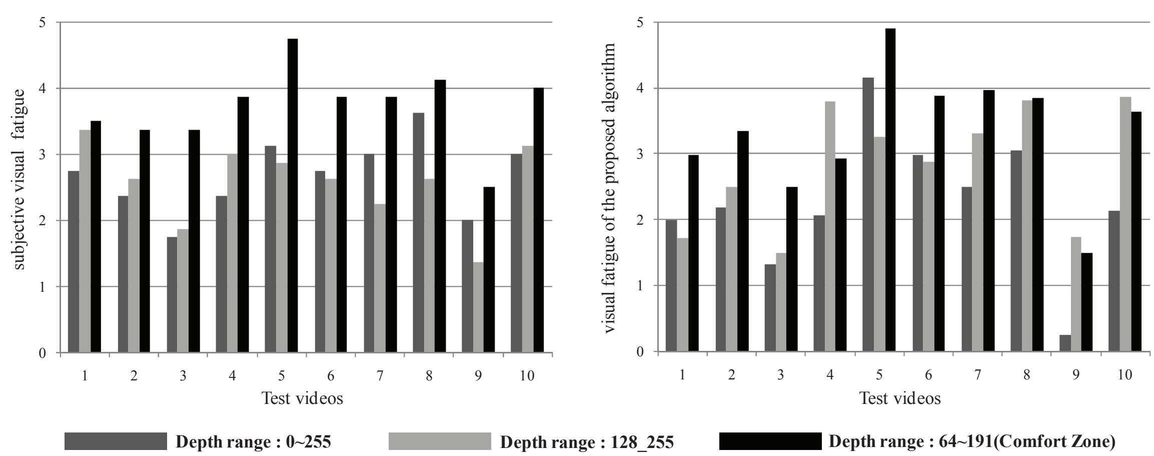


Fig.7. (a) subjective visual fatigue, (b) visual fatigue of the proposed algorithm