Evaluation of visual comfort for stereoscopic video based on region segmentation

Shigang Wang Xiaoyu Wang Yuanzhi Lv

Abstract

In order to accurately reflect the human eye's visual comfort in the process of viewing stereoscopic video, a method based on region segmentation is proposed to assess the visual comfort of stereoscopic video. First, adaptively segment and extract the interested area of the human eye from the sequence of stereoscopic video. Then, combined with depth perception theory and space technology, model the characteristics of visual comfort. Finally, complete the assessment of stereoscopic video comfort. The experiment results show that the method of assessing the stereoscopic video visual comfort can simulate the perception of human eye's viewing the region directly, so they have important reference value for the study of visual comfort.

Key words: Region segmentation Visual comfort Depth perception Disparity 3D displays

1 Introduction

With the vigorous development of multimedia technology and Internet, the stereoscopic video technology is developed rapidly. Compared with 2D video,3D video not only bring us more realistic visual experience, but also give us more immersive feeling. However,

Xiaoyu Wang (✉)
Jilin province Changchun, nanhu road no.5372, 130012, china
e-mail:tianjiaoyu2009@163.com

This study was supported by the national high technology development program (863 Program) (2012AA011505), the national natural science foundation (U0935001) and Doctoral university special research foundation (20130203110005)

3D display is imperfect, the contents stereoscopic video are not compatible, and viewing of the limitations of the environment make people feel visual discomfort in watching stereoscopic video (headache, dry eyes, eye fatigue)[1]-[2].

Commonly viewers watching the stereoscopic video use the concept of visual comfort to reflect the subjective comfortable degree of feeling，embodied in the reflection of eye fatigue and physical conditions after viewing the stereoscopic video a certain time. Researches show that, the human eye for image consistency requirements is very high, and if the difference of image presented to the human eye exceeds a certain range, the human eye will not be able to fuse into a single stereoscopic image, generating uncomfortable symptoms when watching the stereoscopic images. If stereoscopic video content is comfortable or not directly relates to the viewer’s visual perception and physical health, which has become one of the main factors of stereo products rapidly and widely recognized by the users[3].

Currently, the stereoscopic image quality evaluation standard can be roughly divided into two categories: the measuring method based on visual perception and the measuring method based on visual interest. The opinion of image quality evaluation method based on visual interest is: when one is viewing an image, only the details of one partial region can distinguished, and other regions can’t be identified in the image at the same time, so the vision is only interested in the region which brightness or texture changes significantly. The objective embodied in the human vision has certain selectivity. So the image can be divided into visual regions of interest (ROI, Region of interest)[4] and non-ROI, and according to the degree of interest ROI is set for the weighted value. The so-called visual region of interest is the regions in the image that change significantly. Because of the significant change, people will pay more attention to watch it carefully, while the texture area which spatial frequency is approximate or the brightness area which is uniform smooth is ignored. The whole image quality often depends on the quality of the ROI, so the degradation of the non-ROI is less affected.

Combined with current situation of the development of visual comfort, this paper presents a comfort evaluation method of stereoscopic video based on region segmentation, and establishes an objective quality evaluation model in line with the results of human subjective perception.

2 Evaluation System Framework

As shown in the following diagram, this algorithm is based on image segmentation and depth perception[5] as the basic framework, and combined with spatial combining technology[6], establishes a visual comfort model in accordance with human visual properties. When the human eye is viewing an image in which there is a marked difference in brightness and scenery, people will focus on the bright area. Considering this characteristic, this paper introduces a significant region extraction method, GBVS (the graph-based visual saliency) algorithm[7].The algorithm can mark out the saliency region which attracts human attention clearly and accurately. First, the system disposes current frame vision and gets the standard disparity of the view. Then, by applying the GBVS algorithm to deal with the left and right view respectively, get a significant regional distribution and by combining the standard disparity map with the saliency map generate an image. For the saliency area, the segmentation can be extracted from that image (i.e. the area which attracts more attention and could cause large deviation of the experimental results). Using the high-order detection method to deal with two adjacent frame views, while suppressing the noise, can get the motion area of two views[8]. Extract the segmentation region and calculate their respective disparity depth. References of the space technology can get a relatively stable value of the disparity d, which substitute into the visual comfort model and obtain the results of visual comfort[9]-[10].Flow chart of the algorithm is shown in Figure 1.

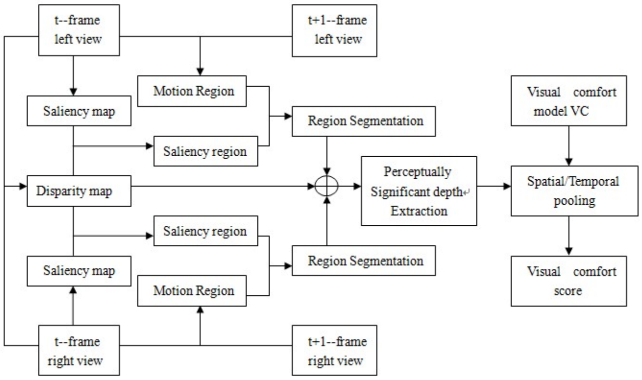

Figure 1: Algorithm framework

3 Evaluation Model

People have feelings of headache, dry eyes, visual fatigue and other symptoms when they are watching 3D stereoscopic video, so it is necessary to compare the observations of different state on different conditions. Experimental results show that the observational results of the curve shape with no adverse symptoms are almost the same as the experimental results in the symptoms of headache, dry eyes, visual fatigue , thus hypothesize that the visual comfort meet certain functional relationship when human views stereoscopic video. According to logical relationship based on the results of the subjective evaluation experiments with disparity d, the mathematical expression of stereo video visual comfort is determined: $VC = c_1 d^3 + c_2 d^2 + c_3 d + c_4$

where VC is the score of visual comfort, c is characteristic parameters, d is disparity depth.

3.1 Feature Region Selection

Considering the particularity of the human eye attention mechanism, when observing video they will pay more attention to the significant areas which have bright colors and motion region, so this article will only discuss two kinds of characteristic region foresaid.

For the selection of the feature region, this paper presents motion areas that consist of the moving object within significant regions and others in the original view in the sport area. The rest of the background object is classified as significant regional background part. Based on the conception mentioned above, simulate the human visual attention mechanism.

First, segment and extract of the motion area. The so-called motion area is the region which has frame difference in the adjacent image. Rapid displacement of the object, may result in the corresponding noise interference. Therefore, in order to restrain noise and extract motion area clearly, this paper uses high-order statistic detection method which considers the frame difference of motion region as non-Gaussian signal, and considers noise and static region in background as a zero mean Gaussian signal. In this way, the motion region detection process can be modeled as a non-Gaussian signal extracted from the Gaussian signal.

Then, the segment and extract significant region. Salient region extraction by GBVS algorithm can cause the interest of the viewers and display the content of the image. These regions contain more information, which plays an important role in the analysis and evaluation of image. For the selection of salient region, commonly used method is based on human visual attention mechanism, through the mathematical model to simulate human's attention mechanism to achieve significant area of adaptive extraction. Extracted significant area where the color is highlighted and the content is rich can perform characteristic of the image well.

3.2 Evaluation Model

Human eyes mostly concentrate on the motion region and static part of a significant area, but human eyes often pay more attention to moving objects than static part of a significant area. So in the paper, the motion region and static part of the significant areas are given different weights that are assigned 0.8, 0.2. Combined with the previously mentioned visual comfort model, we can get a new visual comfort score model which simulates human vision characteristic:

$VC = f(\overline{d_M}, \overline{d_S}) = w_1(a_1\overline{d_M}^3 + a_2\overline{d_M}^2 + a_3\overline{d_M} + a_4) + w_2(b_1\overline{d_S}^3 + b_2\overline{d_S}^2 + b_3\overline{d_S} + b_4)$ where VC is the score of visual comfort, d_M is the disparity of the motion region, d_S is the disparity of the significant region in the background section, w_1, w_2 are assigned 0.8, 0.2, a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4 are the corresponding characteristic parameters.

4 Subjective Evaluation Experiment

4.1 Equipment and Viewing Environment

The experiment is conducted in NVIDIA 3d Vision display platform, the configuration of a computer is Intel Core^{TM}2 Duo 3 GHz processor, GeForce GTX 240 graphics card. Projection equipment adopts Sony VPL-HW30ES stereoscopic video projector, subjects view stereoscopic video standing in the distance of 3m behind the stereo screen by wearing the Sony active shutter glasses, including the stereo screen is set 2m.

4.2 Observers

The experiment select a total of 16 subjects whose binocular visions are normal (the corrected visual acuity of the naked eye or wear glasses can reach 1.0, no eye disease that have effect on the experimental data): the age of 20 to 35 years old; 12 males, 4 females; 10 with stereo technical background, 6 people do not have; subjects wearing active shutter glasses had no adverse reaction after experiments.

4.3 Rated Standard and Experimental Sources

According to the criterion for subjective evaluation of the quality of TV image issued by ITU (2002)[11], as shown in figure 2, the comfort of stereoscopic images is divided into five levels evaluation, which rated the scoring accuracy of 0.1.

Stereo image comfort level of subjective evaluation

Score	Describe	Comfort level
5	Could not detect the three-dimensional image Comfort deterioration	very good
4	Stereo image comfort minor changes, but does not obstruct the view	Good
3	Stereo image comfort significantly worse and watching slightly hinder	Ordinary
2	Could feel the stereo image comfort becomes quite bad and watching hinder	Bad
1	Stereo image comfort become extremely bad, very serious impediment to watch	very bad

Figure 2: The standard of subjective evaluation

The experiment will adopt the standard sequence issued by International Video Organization as the stereo video sequence, utilize the related software for processing standard sequence into the left and right format stereo video and play the stereo video through the SSP stereoscopic player.

4.4 Experimental Data

The experiments detect stereo video sequence, so that the participants are subjective to judge the score of the motion region and the static part of the significant region. Under the condition of different disparity depth, the subjective scores are shown in the coordinate system. The scores form a smooth curve. Based on the curve obtain the parameters of the evaluation model which simulates the human vision characteristic.

According to the definition of the horizontal parallax and the principle of stereoscopic perception we can obtain the disparity d of corresponding target area. The logical relationship of the horizontal parallax and disparity depth has shown in the form below:

$$d = \frac{Lp}{p - e}$$

where L, e are known quantities set, p is the horizontal parallax of the adjacent frames views in the case that the vertical parallax is ignored.

5 Experimental Results and Conclusions

The figure below shows the subjects' visual comfort under the condition of different depth of the parallax[12]. We indicate the degree of visual comfort through comfort level rating, which can improve the visual comfort model mentioned in the previous content:

$$VC = f(\overline{d_M}, \overline{d_S}) = 0.8 * \left(55.902 * \overline{d_M}^3 + 45.56 * \overline{d_M}^2 + 9.188 * \overline{d_M} + 4.763\right) + 0.2 * (6.537 * \overline{d_S}^3 + 5.038 * \overline{d_S}^2 + 5.026 * \overline{d_S} + 4.868)$$

In order to check the reliability of the model, the paper once again ask the subjects to watch and score 6 period of current hot showing 3D movies. As shown in figure 3, experimental results show that the calculating comfort scores of comfort model obtained by this paper are the same as the subjective comfort scores given by the subjects. Thus it can be seen, comfort model mentioned in the article has a good practicability for the rapid development of the evaluation of visual comfort for the stereoscopic video.

Performance comparison of visual comfort

Film Name	$\overline{d}_M$	$\overline{d}_S$	VC	Subjective Rating
Ice Age	-0.403	-0.374	4.520	4.5
Avatar	-0.301	-0.180	4.504	4.5
Alice In Wonderland	-0.588	-0.392	3.678	4.4
Sea Rex	-0.416	-0.152	4.691	4.7
TransformersIII	-0.096	-0.837	3.507	4.2
Sea World	-0.052	-0.205	4.319	4.3

Figure 3: Performance comparison of visual comfort

This paper assumes that the disparity and the visual comfort of the 3D contents have a strong correlation, and on the basis establishes a three-dimensional visual comfort evaluation model based on the region segmentation, in which the significant region is divided into moving regions and static parts. By the subjective evaluation, we can obtain the subjects' visual comfort level when viewing the stereo video in different depths of parallax. At the same time, compare the stereo visual comfort ratings with the subjective comfort scores by watching. The results confirm the reliability and usefulness of the comfort evaluation model. The scoring results in the third section video and the fifth section video exist certain differences, the likely cause is that the video resolution ratio is low and clarity is poor which affects the subjective rating of subjects.

6. References

[1] Donghyun Kim, and Kwanghoon Sohn, “Visual fatigue prediction for stereoscopic image, Circuits and systems for Video Technology,” IEEE Transaction, Feb, 2011,pp. 231-236.

[2] Marc Lambooij, Wijnand I Jsselsteijn, et al., “Stereoscopic displays and visual comfort: a review,” Einhoven University of Technology Einhoven University of Technology Philips Research, SPIE Newsroom 2, April, 2007.

[3] Frank L. Kooi, and Alexander Toet, “Visual comfort of binocular and 3D displays,” Display, 2004, vol. 25, no. 2-3, pp. 99-108.

[4] Ma M.Y, and He X.-L, “Research on ROI extraction of color image,” Intelligent Control and Automation (WCICA), World Congress on July, 2010, pp. 724-728.

[5] HosikSohn, and Yong Ju Jung, “Attention model-based visual comfort assessment for stereoscopic depth perception,” Digital Signal Processing (DSP), 2011 17th International Conference on July, 2011, pp. 1-6.

[6] Zhou Wang, and Xinli Shang, “Spatial pooling strategies for perceptual image quality assessment,” Image Processing, 2006 IEEE International Conference, Oct, 2006, pp. 2945-2948.

[7] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” Advances in Neural Information Processing Systems, 2007.

[8] Qian Zhou, “Study on Object Segmentation in Stereo Images and Stereo Videos,” asia-pacific workshop on visual information processing, Nov, 2006.

[9] Donghyun Kim, and Sunghwan Choi, “Visual Comfort Enhancement for Stereoscopic Video Based on Binocular Fusion Characteristics, Circuits and Systems for Video Technology,” IEEE Transaction on March, 2013, pp. 482-487.

[10] Ye Bi, Jun Zhou, “A visual comfort metric for stereoscopic 3D video based on SMDE approach,” Signal Processing, Communication and Computing (ICSPCC), 2012 IEEE International Conference on 12-15 Aug. 2012, pp. 72-77.

[11] ITU Recommendation BT.1438, Subjective assessment of stereoscopic television pictures [S], 2000.

[12] SEUNTIENS P J H, visual experience of 3D TV [D], Eindhoven: Eindhoven University, 2006.