International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:7, 2013

Cartoon Effect and Ambient Illumination Based Depth Perception Assessment of 3D Video

G. Nur

***Abstract*—** Monitored 3-Dimensional (3D) video experience can be utilized as "feedback information" to fine tune the service parameters for providing a better service to the demanding 3D service customers. The 3D video experience which includes both video quality and depth perception is influenced by several contextual and content related factors (e.g., ambient illumination condition, content characteristics, etc) due to the complex nature of the 3D video. Therefore, effective factors on this experience should be utilized while assessing it. In this paper, structural information of the depth map sequences of the 3D video is considered as content related factor effective on the depth perception assessment. Cartoon-like filter is utilized to abstract the significant depth levels in the depth map sequences to determine the structural information. Moreover, subjective experiments are conducted using 3D videos associated with cartoon-like depth map sequences to investigate the effectiveness of ambient illumination condition, which is a contextual factor, on depth perception. Using the knowledge gained through this study, 3D video experience metrics can be developed to deliver better service to the 3D video service users.

***Keywords*—**3D Video, Ambient Illumination, Cartoon Effect, Depth Perception.

I. Introduction

To speed up the advancement of 3-Dimensional (3D) video services in consumer electronics market, the effect of these technologies on user perception should be investigated. This investigation can then be utilized as feedback information to fine tune the service parameters. However, to be able to fine tune the system parameters in the best way as possible, 3D video experience should be assessed by considering the effective factors on this experience.

In the case of color plus depth map 3D video representation, the depth map sequence is exploited to shape the position of the associated color sequence in 3D space. Therefore, the depth perception is a significant factor for rendering at the receiver side [1]. Assessing the depth perception of users has a complex nature due to the contextual and content related factors effective on this perception. Therefore, predicting the depth perception requires a deep investigation study using these factors.

In this paper, the effects of structural information, which is a content related factor, and ambient illumination condition, which is a contextual factor, on the depth perception are investigated.

Even though the objects and background of a depth map sequence may be comprised of individual pixel depth values, the basic structures of these objects and background can be determined using the significant depth levels [2]. Thus, the structural information is associated with the significant depth levels in the depth map sequences in this paper. In order to determine the significant depth levels, the depth map sequences are cartoon-like affected. Bilateral filter [3] is employed to provide the cartoon-like effect in this study.

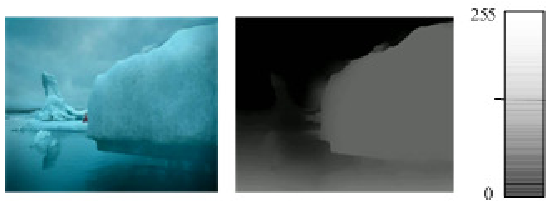
In order to determine the effect of ambient illumination condition on depth perception of cartoon-like affected depth map sequences, subjective experiments are carried out.

The rest of the paper is organized as follows. Section II discusses the color-plus-depth map based 3D video. The cartoon-like effect is described in Section III. In Section IV, the subjective experiments are presented. The results are discussed in Section V. Finally, Section VI concludes the paper.

G. Nur is with Electrical and Electronics Engineering Department, Kirikkale University, Kirikkale, Turkey (phone: +90 318 4242-1218; e-mail: gnur@kku.edu.tr).

II. Color-plus-depth-map based 3D video representation

A snapshot of color texture image plus associated depth map representation of Ice sequence captured using a 3D depth-range camera is shown in Fig. 1. As observed from the figure, each pixel in the depth map has an associated pixel in the color image. The pixels in the depth map, which take grey values ranging from 0 to 255, determine the distance of the associated color image pixel to the viewer. 0 corresponds to the furthest away pixel from the viewer, while 255 represents the closest pixel to the viewer in a 3D scene [4].



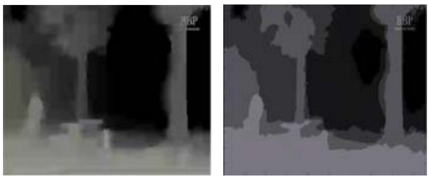
(a) (b)

Fig. 1. The Ice sequence: (a) color (b) associated depth map

In order to render left and right views considering color and per-pixel depth maps, Depth-Image-Based Rendering (DIBR) technique is exploited. The basic principle of the DIBR technique relies on calculating screen parallax, which represents pixel shifting to generate different images for the left and right eye of the viewer. These images together are interpreted by human brain to sense the depth effect [4].

III. CARTOON-LIKE EFFECT

The cartoon-like effect aims at lowering the perceptual and cognitive effort to sense the depth in a depth map sequence. Using this effect, meaningful structural information in the depth map sequence is abstracted and extra information that is not necessary to perceive the depth is excluded [3]. A compressed depth map sequence of the Farm sequence and its cartoon-like affected depth map sequence is illustrated in Fig. 2. The number of pixel depth values in the compressed depth map sequence is much higher than that of the abstract compressed depth map sequence.



(a) (b)

Fig. 2. The depth map sequence of the Farm sequence: (a) compressed (b) cartoon-like affected

Bilateral filter, which is based on Gaussian filtering method, is exploited for cartoon-like abstraction in this study. The bilateral filter aims at preserving the edges and smoothing the image with noise reduction. The bilateral filter uses Gaussian weighting coefficients to replace each pixel value with a weighted average of neighborhood pixel values. The researcher using the bilateral filter determines these coefficients [3].

For a pixel *a*, the bilateral filtered *Fa* is calculated as;

$$F_a = \sum_{b\in\beta} w_{ab} I_b \,/ \sum_{b\in\beta} w_{ab} \quad (1)$$

where, I_b is at the intensity of pixel b in the kernel neighborhood β. The weighting coefficient at pixel *b* is calculated as follows;

$$w_{ab} = c(a,b)s(a,b) \quad (2)$$

where, *c(a,b)* and *s(a,b)* are clossiness and similarity kernel filters, respectively. These kernel filters are computed as;

$$c(a,b) = e^{\left(\frac{-1}{2}(a-b)^2/\sigma_c^2\right)} \quad (3)$$

$$s(a,b) = e^{\left(\frac{-1}{2}(I_a-I_b)^2/\sigma_s^2\right)} \quad (4)$$

IV. SUBJECTIVE EXPERIMENTS

In this section, the subjective experiments performed using 3D videos comprised of compressed color texture sequences associated with compressed and compressed-cartoon-like affected depth map sequences. Five 3D video sequences namely; Ice, Eagle, Chess, Farm and Advertisement are used in the subjective experiments. The Joint Scalable Video Model (JSVM) reference software version 9.13.1 [5] is exploited to encode these 3D video sequences at four different bit rates (i.e., 512, 768, 1024, and 1536 kbps).

Double Stimulus Impairment Scale (DSIS) method is used for the experiments as described in the International Telecommunication Union-Recommendation (ITU-R) BT-500.11 standard [6]. Thus, the video sequences are shown one after other: the first video is the reference and the second video is the compressed or compressed-cartoon-like affected one.

Four different ambient illumination conditions (i.e., 5, 53, 195, and 293 lux), which are measured using a Gretag Macbeth Eye-One Display 2 device [7], are utilized to conduct the experiments. 5 lux corresponds to a dark condition, while 293 lux indicates a bright light environment. During the experiments, first of all, the video sequences are presented to the viewers under the 5 lux ambient illumination condition according to the DSIS method. After the orders of the video sequences are changed to hinder prejudice, the same experiment is conducted under the 53 lux ambient illumination condition. Then, similar procedure is applied for the remaining ambient illumination conditions.

An evaluation scale ranging from 1 to 5 is utilized to rate the sequences throughout the experiments. A score of 1 represents the lowest video quality and a score of 5 represents the best video quality for the impaired video sequence compared to the original one.

The 3D video sequences are presented on a 42” Philips multi-view auto-stereoscopic display, which has a resolution of 1920 × 1080 pixels, to the viewers. 18 viewers (7 females and 11 males) volunteered in the experiments, which is in compliance with the recommendation in [6]. They are all non-expert viewers, whose ages ranged from 19 to 37. After the experiments, the outliers are eliminated. Hence, the Mean Opinion Scores (MOSs) and confidence intervals [6] are calculated using 16 participants.

The thumbnails of the five 3D video sequences used in the subjective experiments are depicted in Fig. 3.

(a)

(b)

(c)

(d)

(e)

Fig. 3. Color texture and associated depth map of the (a) Ice (b) Eagle (c) Chess (d) Farm (e) Advertisement sequences

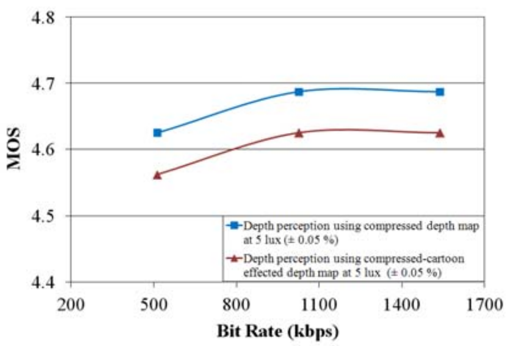
V. Results and Discussion

In this section, the results of the subjective experiments are analyzed in detail. Figs. 3-4 illustrate the bit rate versus MOS results reported on the depth perception evaluations under different ambient illumination conditions for the compressed and compressed-cartoon-like Ice and Eagle sequences, respectively.

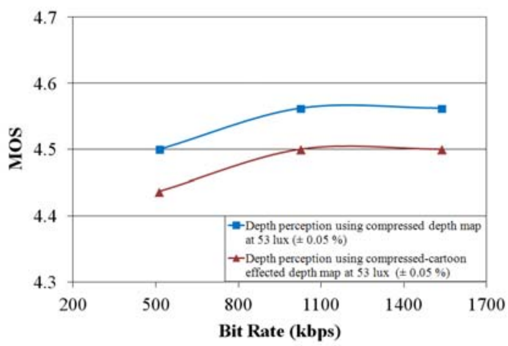
As can be observed from the figures, as the amount of light in the environment increases, a decreasing depth perception pattern is demonstrated across all of the bit rate range for both the compressed and compressed-cartoon-like sequences. Moreover, the highest MOS scores are presented in the 5 lux environment regardless of the varying bit rate. The MOS scores decreases when the ambient illumination increases (i.e., from 5 lux to 52 lux; to 116 lux; and to 192 lux). This is due to the fact that depth perception cues like sharpness, shadows, reflections, contrast differences, etc in the 3D sequences which contribute to perceive the depth better cannot be detected when the ambient illumination increases.

Moreover, as seen from the figures, the compressed sequences present slightly higher (~0.015%) or similar MOS scores compared to the compressed-cartoon-like sequences. The remaining test sequences also present similar patterns.

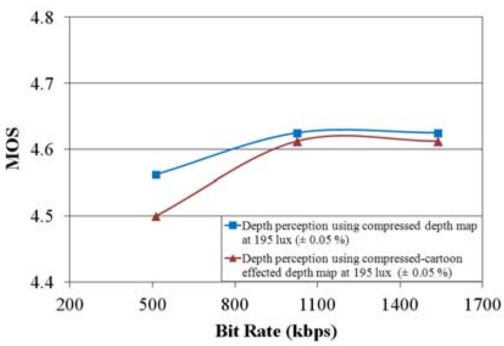
This clearly indicates that the cartoon-like affected depth map sequences can be utilized instead of the compressed video sequences for rendering with a low overhead.



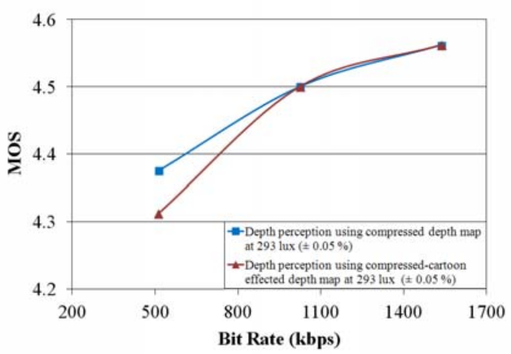
(a)



(b)



(c)



(d)

Fig. 4. The depth perception MOS scores for the cartoon-like Ice sequence under (a) 5 (b) 53 (c) 195 (d) 293 lux ambient illumination conditions

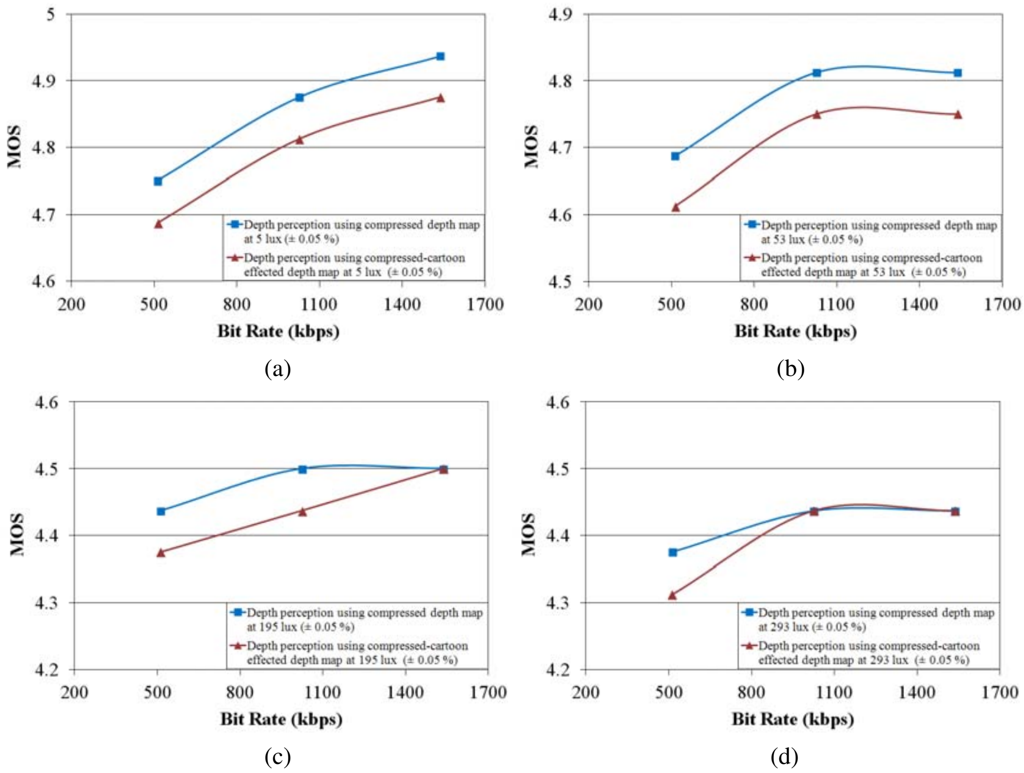


Fig. 4. The depth perception MOS scores for the cartoon-like Eagle sequence under (a) 5 (b) 53 (c) 195 (d) 293 lux ambient illumination conditions

VI. Conclusion

In this paper, the effects of the cartoon-like abstraction and ambient illumination condition in the usage environment on the depth perception of 3D video sequences have been investigated. It has been concluded after the investigations that the ambient illumination condition is effective on depth perception of both the compressed and compressed-cartoon-like affected sequences with a similar pattern. These observations provide a notable set of findings for developing 3D video experience metrics in future studies.

References

[1] T.E.R. Chaminda and M. G. Martini, "Edge-based Reduced-Reference Quality Metric for 3-D Video Compression and Transmission," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no.5, Sep. 2012.

[2] Nur G. and Bozdagi Akar G., "An Abstraction Based Reduced Reference Depth Perception Metric for 3D Video," *International Conference on Image Processing (ICIP)*, Orlando, Florida, USA, 30 September-3 October 2012.

[3] Tomasi C. and Manduchi R., "Bilateral Filtering for Gray and Color Images," *IEEE International Conference on Computer Vision*, Bombay, India, 1998.

[4] Nur G., Dogan S., Kodikara Arachchi H., and Kondoz A. M., "Impact of Depth Map Spatial Resolution on 3D Video Quality and Depth Perception," IEEE 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, Tampere, Finland, 7-9 June 2010.

[5] JSVM 9.13.1. CVS Server [Online]. Available Telnet: garcon.ient.rwth aachen.de:/cvs/jvt

[6] Methodology for the Subjective Assessment of the Quality of Television Pictures, ITU-R BT.500–11, 2002.

[7] Gretag Macbeth Eye-One Display 2, http://www.xrite.com

Gokce Nur received her B.Sc. (Hons.) degree in Electrical and Electronics Engineering from Sakarya University, Turkey in 2005. She worked as a Research Fellow in Kirikkale University, Turkey, during 2005-2007. She received her Ph.D. degree from the I-Lab Multimedia Communications Research group in the Centre for Vision, Speech and Signal Processing of the University of Surrey, UK in 2011. She is currently an Assistant Professor Dr in Electrical and Electronics Engineering Department, Kirikkale University, Turkey. Her research interests include 2D/3D video adaptation, coding, and QoE. Dr Nur is a member of IEEE and IEEE Signal Processing Society.