CHROMATIC VARIATIONS ON 3D VIDEO AND QOE

*Daniel Piedade**, *Marco V. Bernardo*†*, *Paulo T. Fiadeiro*†, *António M. G. Pinheiro*†, *Manuela Pereira**

†Remote Sensing Unit/Optics Center, *Instituto de Telecomunicações
Universidade da Beira Interior
Rua Marquês d'Ávila e Bolama - 6201-001 Covilhã, Portugal
daniel.alpiedade@gmail.com, {mbernardo, pinheiro, mpereira, fiadeiro}@ubi.pt

ABSTRACT

In this paper a study on the perceived quality that results of chromatic variations in 3D video is reported. The testing videos were represented in the CIE 1976 ($L^*a^*b^*$) color space, and their colors were initially subdivided into clusters based on their similarity. Predefined chromatic errors were applied to these color clusters. These videos were shown to subjects that were asked to rank their quality based on the colors naturalness. The Mean Opinion Scores were computed and the sensibility to chromatic changes on 3D video was quantified. Moreover, attention maps were obtained and a short study on the changes of the visual saliency in the presence of these chromatic variations is also reported.

***Index Terms*—** Quality of Experience, 3D Video, Mean Opinion Score, Visual attention

1. INTRODUCTION

The analysis of possible impairments influence on the Quality of Experience (QoE) is of the utmost importance for the creation of valid QoE models. Subjective testing is very time consuming and cost demanding. As a consequence, it is difficult to rely exclusively on its estimation to validate the different models used on every available multimedia devices and systems. The subjective evaluation of 2D visual quality according to standardized methods has been the subject of study in recent years. Thus, subjective assessment of audio and visual quality is considered to be the most accurate method reflecting the human perception [1]. To made these assessments, the International Telecommunication Union (ITU) has issued several recommendations, including the widely used ITU-R BT.500 [2]. Perceptual evaluation of video quality, standardized in ITU-T J.247 [3], defines the chrominance indicator for spatial distortion analysis of the visual quality perceptual evaluation, together with the luminance indicator (based on luminance difference of edge images between the reference and distorted videos), and two temporal variability indicators for omitted and introduced components.

In this work the influence of chromatic variations on the perceptual quality of 3D videos is studied. It is noteworthy that chromatic variations exist in most of the available displays, which are also dependent of the local calibrations. Moreover, the Human Visual System (HVS) reveal a very low sensitivity to chromatic variations, and that fact is also used by the typical video streaming encoding mechanisms that typically use chromatic components subsampling. In [4], a study on the influence of chromatic variations on the perceived quality is reported. In this work, a similar approach is used to induce chromatic errors to the 3D video stream. The existence of two extra dimensions, motion and depth, will create an even lower sensitivity to the induced chromatic variation.

2. METHOD

To create the database of videos with chromatic variations the video sequences were converted to CIE 1976 ($L^*a^*b^*$) color space and then divided into color clusters [4]. Finally, for each color cluster, an error variation with a predefined magnitude and random direction was applied to each pixel.

To avoid the effects of spatial artifacts, the colors of the images were subdivided into clusters applying the K-Means algorithm [5] to the video shot key frames. The key frames definition is presented in section 2.2 and the cluster procedure is presented in section 2.3. The chromatic variations were applied by adding a ΔE^*_{ab} error to every pixel of each cluster using a set of fixed chromatic errors, but applying different random directions (see section 2.4).

Considering that the HVS is more sensitive to luminance and due to display color gamut limitations [6], the lightness L^* of the resulting image always remained unchanged for all the image pixels. Nevertheless, due to the color gamut of the display device some pixel colors could not be represented. So, these colors were displayed by clipping them to the closest color of the display gamut surface in the CIE 1976 ($L^*a^*b^*$) color space. The number of clipped pixels was in general less than 2%.

2.1. CIELAB Color Space Conversion

The videos were converted to CIE 1976 ($L^*a^*b^*$) color space before applying the chromatic error. This color space also known as CIELAB, is device independent, partially uniform and based on the HVS [7–10]. For that, the 3D videos were first converted from RGB to sRGB [11].

Then, in a second step the videos were converted from sRGB to XYZ assuming the D65 reference white [11]. Furthermore, the videos were converted from XYZ to CIELAB. This transformation requires the chromatic coordinates of a reference white, being used the reference white of the display. The determination of these chromatic coordinates (X_W, Y_W, Z_W) was achieved using a spectroradiometer (SpectraColorimeter, PR-650, Photo Research Inc., Chatsworth, CA) [12].

2.2. Key Frames Selection

The clusters were defined using the complete set of all key frames. The key frames selected for each video sequence were the following:

- Basket : $[297, 298, 299, 300, 301, 302, 303, 304]$;
- Hall : $[36]$;
- Barrier gate : $[385]$;
- Umbrella : $[176, 176, 233, 239, 240]$;

The Basket sequence has a fast camera motion that requires a larger number of key frames for a proper color clustering representation.

2.3. Clusters Definition

The K-Means algorithm [4] was used for the definition of the color clusters. Initially K was tested with values varying from 4 to 9. With the higher values of K spatial artifacts tend to appear. However, values smaller than $K = 4$ are not sufficient to represent the different chromatic regions. The value of $K = 4$ was used to the remaining of the experiment since this value results in a good division of chromatic regions.

The pixels in the remaining frames were assigned to the cluster with the closer centroid, using the minimal euclidean distance. With this procedure, each frame is divided into 4 clusters defined by the centroids of the key frames clustering. Moreover, since consecutive frames define their clusters based in the same centroids spatial artifacts were avoided.

2.4. Application of the Chromatic Variation

After defining the clusters for all the frames, a chromatic error was applied to each pixel of each cluster. Hence, each pixel suffered a chromatic variation of ΔE^*_{ab} with a predefined magnitude and random direction (α) defined for each cluster. Since 4 clusters were defined, 4 random directions were randomly chosen ($\alpha_n, n = 1, 2, 3, 4$). The error applied to each pixel is defined by equation (2).

$$\begin{aligned} L_i^{*'} &= L_i^* + (\Delta E^*_{ab} \cos(\theta)) \\ a_i^{*'} &= a_i^* + (\Delta E^*_{ab} \cos(\alpha) \sin(\theta)) \\ b_i^{*'} &= b_i^* + (\Delta E^*_{ab} \sin(\alpha) \sin(\theta)) \end{aligned} \quad (1)$$

In this case $\theta = 90^o$ and α_n belonging to $[0^o, 360^o]$. Hence, only chromatic variations are allowed and $L_i^{*'} = L_i^*$ (no luminance changes occur). This procedure guaranteed that groups of similar colors were changed in the same direction. The evaluation provided will depend exclusively on the color changes, which is the aim of the study, because color artifacts were reduced to unperceived levels. Three different videos for each chosen sequences were generated and for each predefined magnitude error to cover a larger number of directions in the CIELAB color space.

3. SUBJECTIVE QUALITY ASSESSMENT EXPERIMENT

3.1. Laboratory

The experiment for subjective quality evaluation were conducted at the Optics Center of the "Universidade da Beira Interior" (UBI), which is compliant with the recommendations for subjective evaluation of visual data issued by ITU-R BT-500 [2]. The videos used in this study were displayed on a CRT color monitor (Sony, GDM-F520) with 20-inch visible part. The videos used in this study present a spatial resolution of 1152×648. A black band was applied in the top and in the bottom resulting in a final resolution of 1152×864. With this procedure videos will have the same resolution as the display, providing a direct relation with the data recovered from the eye tracker. The display had a resolution with 1152×864, with 85 Hz frame rate and 77.10 Hz scan rate.

The videos were observed by the subjects seated in line with the display center, and a viewing distance of 1 m. The head was centered in the eye tracker equipment, both horizontally and vertically. The room illumination was only provided by the display and by two projectors Kaiser RB1 used to illuminate the eye tracker. These projectors used a halogenic lamp with 650 watts each.

3.2. Dataset

The original sequences used in this experiment were created by the University of Nantes and the University Polytechnic of Madrid. This database is named Nantes-Madrid-3D-Stereoscopic-V1, NAMA3DS1-COSPAD1 database[1]. This 3D video database includes uncompressed and encoded videos with very high bitrate. It is composed of Full-HD

---

[1] http://www.irccyn.ec- nantes.fr

(a) Basket - Sequence (b) Hall - Sequence (c) Barrier gate - Sequence (d) Umbrella - Sequence

Fig. 1. Example of a single frame for each sequence used in the experience.

stereo sequences at 25 frames per second, captured using a semi-professional 3D camera.

The provided sequences are summarized in Fig. 1 which shows a frame of each sequence. Detailed information on each sequence, including a description, the shooting conditions, the camera parameters, and depth information are presented in [13].

The original database has ten different sequences. Six were selected to be used in this experiment, from which four sequences were used for the assessment and two sequences for the trial period.

Moreover, two indoor and two outdoor sequences were selected for the assessment. One of the indoor sequences, presents fast camera motion and scene rotation (Basket sequence represented in Fig. 1.(a)). The other presents only slow motion (Hall sequence represented in Fig. 1.(b)). The two outdoor sequences present respectively slow camera motion (Barrier gate sequence represented in Fig. 1.(c)) and the localized motion (Umbrella sequence represented in Fig. 1.(d)).

3.3. Subjects

The panel of volunteers comprised 25 subjects, all within the range of $16 - 33$ years old, with average age of 23 and a sample standard deviation of 3.26. In this set, 64% were male and 36% were female. All of them were naive as to the aim of the experiment. The subjects were submitted to color, stereopsis and ocular domination tests. For color vision assessment was used the Ishihara color test, the Randot Stereo tests for stereopsis and the Miles test for the ocular domination. The ocular domination was performed to detect the dominant eye required for the correct use of the eye tracker. Informed consent was obtained from all participants and the research was conducted according to the Declaration of Helsinki [14].

3.4. Procedure

The Absolute Category Rating with Hidden Reference (ACR -HR) standard test methodology [15] was chosen for the experiments. In this methodology a sequence of videos is presented to the subjects, one at each time. The sequences are evaluated independently with a classification scale. A quality continuous scale with 11 points was chosen [16]. The reference sequence is included in the test and classified as all the others. From the obtained results can be computed the Mean Opinion Scores (MOS). Using the MOS of the reference sequence was also computed the Differential Mean Opinion Scores (DMOS). To minimize the contextual effects, the videos were presented in a random order different to each subject.

Prior to the beginning of the test, an experiment protocol was shown to the subjects. The test was composed by two parts: a first one, called Training Period (using the trial sequences), where the subject was allowed to be familiarized with the procedure, followed by the Test Session. This trial period was the same for all the subjects.

The complete test had an average duration of 30 minutes. Longer test sessions are not advisably because the subjects tiredness would influence the final results and also the attention points detected with the eye tracker.

During the Test Session the subject watched each sequence followed by the respective quality evaluation. For the evaluation, a keyboard placed in front of the subject was used. They could select a grade between 0 (low quality/artificial) and 10 (high quality/ natural) in a simulated continuous scale (100 levels). Only the answers given by the subject during the Test Session were considered for the results of this study.

4. ANALYSIS OF SUBJECTIVE RESULTS

4.1. Distribution Analysis

The ANOVA analysis was used for the statistical analysis. From the ANOVA tests can be concluded that the results are statistically significant, since the P value is always smaller than 0.05 and the F value is always higher than 3.354. Several authors argue about the influence of removing the outliers detected in the results. The criteria defined in ITU-R BT.500 [2] for outliers removal was applied and no subject was considered an outlier.

4.2. Mean Opinion Score

Statistical measures were computed to describe the MOS distribution across the subjects, that was computed for each

Source	SS	df	MS	F	Prob > F
	Basket sequence				
Groups	76.77	3	25.59	5.70	8.6e-4
Error	1103.62	246	4.48		
Total	1180.40	249			
	Gate barrier sequence				
Groups	129.39	3	43.13	11.28	5.8e-07
Error	940.10	246	3.82		
Total	1069.5	249			
	Hall sequence				
Groups	154.64	3	51.54	7.46	8.4e-05
Error	1699.1	246	6.90		
Total	1853.8	249			
	Umbrella sequence				
Groups	193.05	3	64.35	12.44	1.3e-7
Error	1271.9	246	5.17		
Total	1464.9	249			

Table 1. ANOVA

video sequence as,

$$\text{MOS} = \frac{\sum_{i=1}^{N} s_i}{N} \tag{2}$$

where N is the number of subjects and s_i is the score given by the subject i to each test video sequence.

Figure 2(a) represents the MOS as a function of the chromatic error ΔE^*_{ab} for the tested video sequences and the respective confidence interval.

The MOS quality scores stay relatively high for $\Delta E^*_{ab} = 6$ units. In some situations a larger MOS value than the obtained for the reference video sequence results. In those cases the induced chromatic variations led to more saturated colors, which might produce a better sense of quality. In general, for videos with an error $\Delta E^*_{ab} > 6$ was noted a chromatic degradation resulting in a MOS decrease as the error increases.

From the analysis of the Fig. 2(a) can be noted that the confidence intervals are small, revealing a similarity between the assessment of the different subjects. Moreover these values stay stable for the different errors.

4.3. Differential Mean Opinion Score

The DMOS was computed comparing the MOS from the stimulus (MOS$_i$) with the MOS of the reference (MOS$_{ref}$) as presented in equation (3).

$$\text{DMOS}_i = \text{MOS}_i - \text{MOS}_{ref} + 10 \tag{3}$$

As recommended in [17], the DMOS was computed as a function of the chromatic error and also of the stimulus. The results are shown in Fig. 2(b) and Fig. 2(c). A DMOS value larger than 10, correspond to video sequences with a larger MOS than the reference. Videos with $\Delta E^*_{ab} = 18$ present the lower values for the quality assessment of the reference, showing that high chromatic errors become perceptible and somehow annoying. Figure 2(c) represents the DMOS as a function of the induced chromatic error. As expected, the subjective evaluation decays for larger chromatic variations.

4.4. Attention Maps

A High-Speed Video Eye Tracker (HS-VET) was used in the experiments, for the acquisition of the attention maps. The used device is precise and presents low noise levels, without frame suppression. Moreover, this device represents the visual stimulus calibrated in a display with time precision and has a reliable synchronization mechanism with external devices for data recovering. The locations that are focus of attention when a 3D video is presented to an subject are acquired for each frame.

Several attention maps were recorded during the subjective testing experiments. Only the maps of subjects with more than 50% of tracking were considered for analysis. Figure 3 represents the attention maps of the selected subjects. Each color in the maps corresponds to a different subject. The maps represent the attention of the second third of the video. The represented frame is the first frame of the second third of the video.

Can be noticed that for all chromatic variations the attention is located in the central part of the image, that is where motion occurs. The same happened with the other video sequences, where the attention was mostly located were the motion is occurring. In fact, the absence of attention in most of the locations and the high attention given to motion shall be the reason that causes the lack of sensitivity to the induced chromatic errors.

5. CONCLUSIONS

As it is well known on multimedia technology, chromatic deviations have little influence on the perception of quality. In this work, the influence of this type of deviations on 3D video was quantified and studied. As a practical consequence multimedia manufacturers and providers may have a more clear figure of how sensitive subjects are to the chromaticity quality level.

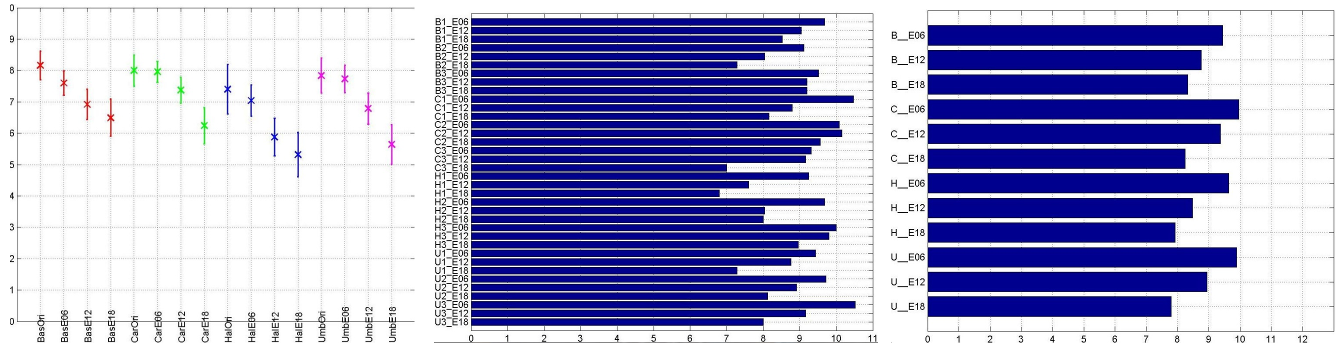
From this study, was concluded that subjects have a large tolerance to any kind of global chromatic variation on 3D video. Moreover, from the attention maps of the different videos can be concluded that the motion and also the depth tend to cause most of the attention. Hence, those features tend to mask any chromatic variation.

As future work, this evaluation will be compared with the results obtained for 2D Video.

REFERENCES

[1] ITU, "ITU-T Recommendation P.911: Subjective audiovisual quality assessment methods for multimedia application," Tech. Rep., ITU - Telecom. Standard. Sector, December 1998.

[2] ITU, "ITU-R Recommendation BT 500-12: Methodology for the subjective assessment of the quality of tele-

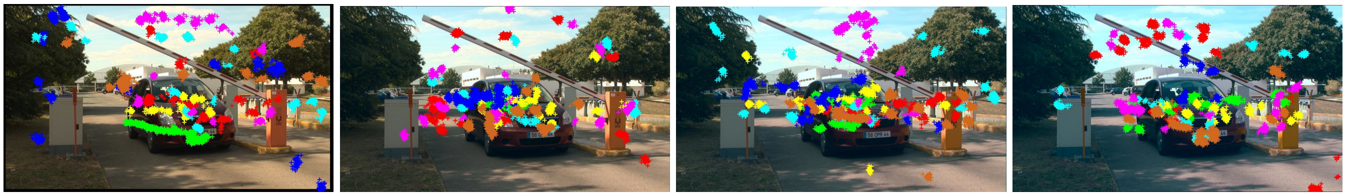


(a) MOS and confidence interval for each ΔE^*_{ab}.

(b) DMOS for each stimulus

(c) DMOS for each ΔE^*_{ab}

Fig. 2. Subjective results



(a) $\Delta E^*_{ab} = 0$

(b) $\Delta E^*_{ab} = 6$

(c) $\Delta E^*_{ab} = 12$

(d) $\Delta E^*_{ab} = 18$

Fig. 3. Attention maps of the second third of the "Barrier gate" video sequence for different ΔE^*_{ab}.

vision pictures," Tech. Rep., ITU - Radiocom. Sector, September 2009.

[3] ITU, "ITU-T Recommendation J.247: Objective perceptual multimedia video quality measurement in the presence of a full reference," Tech. Rep., ITU - Telecom. Standard. Sector, August 2008.

[4] Marco V. Bernardo, Antonio M. G. Pinheiro, Manuela Pereira, and Paulo T. Fiadeiro, "A study on the user perception to color variations," in *Proceedings of the 20th ACM international conference on Multimedia*, Nara, Japan, 2012, pp. 1009–1012.

[5] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall PTR, 2011.

[6] W. Wintringham, "Color television and colorimetry," *Proc. of the IRE*, vol. 39, no. 10, pp. 1135 –1172, October 1951.

[7] CIE, "Colorimetry official recommendation of the international commission on illumination," CIE publication 15.2, CIE Central Bureau, 1986.

[8] T. Smith and J Guild, "The C.I.E. colorimetric standards and their use," *Transactions of the Optical Society*, vol. 33, no. 3, pp. 73–134, 1931.

[9] R. Hunt, *The Reproduction of Colour*, John Wiley & Sons, New York, 3 edition, March 1975.

[10] G Wyszecki and W Stiles, *Color Science*, John Wiley & Sons, New York, 3 edition, March 1967.

[11] IEC 61966-2-1, "Colour measurement and management in multimedia systems and equipment - part 2-1: Default RGB colour space - sRGB," Tech. Rep., International Electrotechnical Commission, 1999.

[12] D. Brainard, "Calibration of a computer controlled color monitor," *Color Research & Application*, vol. 14, no. 1, pp. 23–34, 1989.

[13] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricordel, P. Le Calle, J. Gutierrez, and N. Garcia, "Nama3ds1-cospad1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3d stereoscopic sequences," in *IRCCyN-IVC*, July 2012, pp. 109–114.

[14] WMA, "World Medical Association Declaration of Helsinki - Ethical principles for medical research involving human subjects," 2009.

[15] ITU, "ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications," Tech. Rep., ITU - Telecom. Standard. Sector, April 2008.

[16] D. M. Rouse, R. Pepion, P. Le Callet, and S. S. Hemami, "Tradeoffs in subjective testing methods for image and video quality assessment," in *Proc. SPIE 7527, Human Vision and Electronic Imaging XV*, February 2010.

[17] VQEG, "Final report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment," Phase II VQEG, March 2000.