Engineering Science and Technology, an International Journal 17 (2014) 105–115

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: http://www.elsevier.com/locate/jestch

Full length article

3D video bit rate adaptation decision taking using ambient illumination context

G. Nur Yilmaz [a,*], H.K. Arachchi [b], S. Dogan [b], A. Kondoz [b]

[a] Kirikkale University, Electrical and Electronics Engineering Department, Kirikkale, Turkey
[b] I-Lab Multimedia Communications Research, Centre for Vision, Speech, and Signal Processing, Faculty of Engineering & Physical Sciences, University of Surrey, Guildford GU2 7XH, Surrey, UK

ARTICLE INFO

Article history:
Received 7 December 2013
Received in revised form
7 April 2014
Accepted 21 April 2014
Available online 20 May 2014

Index terms:
3D video quality perception
Adaptation decision taking
Video adaptation
Ambient illumination
Depth intensity
Luminance contrast
Motion
Structural feature

Keywords:
3D video adaptation decision taking
3D video coding

ABSTRACT

3-Dimensional (3D) video adaptation decision taking is an open field in which not many researchers have carried out investigations yet compared to 3D video display, coding, etc. Moreover, utilizing ambient illumination as an environmental context for 3D video adaptation decision taking has particularly not been studied in literature to date. In this paper, a user perception model, which is based on determining perception characteristics of a user for a 3D video content viewed under a particular ambient illumination condition, is proposed. Using the proposed model, a 3D video bit rate adaptation decision taking technique is developed to determine the adapted bit rate for the 3D video content to maintain 3D video quality perception by considering the ambient illumination condition changes. Experimental results demonstrate that the proposed technique is capable of exploiting the changes in ambient illumination level to use network resources more efficiently without sacrificing the 3D video quality perception.

1. Introduction

The stereoscopic viewing ability of humans has always been the driving force behind the efforts for bringing 3-Dimensional (3D)-enabled technologies (e.g., 3D video capture, representation, coding, etc) to reality [1]. Although several developments in these technologies have been accomplished to date, there are still many areas that need to be improved through vigorous research. 3D video adaptation decision taking is one of the important areas that require in-depth investigations for enabling next generation pervasive 3D media environments.

The process of transforming a 3D video content into another version to satisfy a set of constraints (e.g., terminal capabilities, network and natural environment conditions, etc) is called 3D video adaptation. 3D video adaptation decision taking is a process of deciding the most important parameters to utilize in the adaptation operations. The overall target of the 3D video adaptation is to maximize user experience in terms of 3D perceptual quality [2,3]. Hence, it is necessary to determine key contextual and content related factors that can affect perceived 3D video quality, so as to lead the 3D adaptation decision taking techniques to assist the 3D video adaptation for achieving its target. Using ambient illumination condition as contextual information is an interesting research topic, which is yet to be thoroughly investigated for 3D video adaptation decision taking purposes. If 3D video adaptation decision taking techniques are developed using ambient illumination context, a wide range of new 3D video applications maximizing user experience can be allowed by Future Media Internet. Technologies related to these applications range from 3D display to rendering, etc. Ambient illumination context can be gathered through light sensors placed on user devices to collect information about the level of brightness in the consumption environment.

In this paper, ambient illumination is exploited as a context to develop a 3D bit rate adaptation decision taking technique as an extension of the studies in Refs. [4,5] by adding depth information to the 2D video contents and using more content related contextual information in adaptation decision taking operations. Due to its

[*] Corresponding author.
E-mail addresses: gnur@kku.edu.tr, gokceem@gmail.com (G. Nur Yilmaz), h.kodikaraarachchi@surrey.ac.uk (H.K. Arachchi), s.dogan@surrey.ac.uk (S. Dogan), a.kondoz@surrey.ac.uk (A. Kondoz).
Peer review under responsibility of Karabuk University.

http://dx.doi.org/10.1016/j.jestch.2014.04.003
2215-0986/Copyright © 2014, Karabuk University. Production and hosting by Elsevier B.V. All rights reserved.

flexibility and compatibility with the existing coding and transmission technologies color texture plus depth map based 3D video is used as the representation format while developing this technique [5].

The sensitivity of Human Visual System (HVS) for perceiving visual artifacts in a 3D video content corresponds to the amount of light the eyes capture from the viewing environment and iris' adaptation of its size considering the amount of light captured. The sensitivity of HVS for detecting depth perception associated cues (e.g., sharpness, shadows, reflections, contrast, etc) that enhance depth perception in a 3D video content changes according to the change in the ambient illumination. The sensitivity of HVS for perceiving overall quality of a 3D video content is related to how the combination of video quality and depth is perceived under a specific ambient illumination condition as also discussed in Refs. [5–8].

Subjective experiments are conducted to monitor the effect of ambient illumination changes on 3D video perception (i.e., video quality, depth, and overall 3D video quality perception) in this paper. Thus, video quality refers to the perceived quality of the color texture sequence regardless of the depth perception. Depth perception refers to the perceived quality of the depth of a 3D video regardless of the perceived quality of the color texture sequence. 3D perceptual quality refers to the overall perceived quality of 3D video (i.e., both the perceived video quality and depth perception).

A user perception model is developed by utilizing the knowledge gained through the results of the subjective experiments. Using this model, the perception characteristics of a user towards 3D video content viewed under a particular ambient illumination condition can be determined. Motion, structural feature, and luminance contrast of color texture and depth intensity of depth map of a 3D video are utilized as content related contextual factors in the user perception model. The quantitative values of the determined contextual factors in 3D video contents are measured using metrics proposed in the paper. A 3D video bit rate adaptation decision taking technique is developed to determine bit rate for adapting a 3D content to maintain overall 3D video quality perception when ambient illumination condition of the viewing environment changes using the user perception model. A significant amount of network bandwidth can be saved by adapting the 3D video content with the determined bit rate, and the bandwidth can be exploited to serve other users' needs in shared network environments. Thus, the shared network bandwidths can be distributed more efficiently in these environments.

Terminal functionalities for increasing or decreasing the brightness of the display could be used for maintaining 3D video perception by enhancing the content visibility under ambient illumination condition changes. However, it should be noted that the focus in this paper is not to improve the content visibility, but to distribute shared network resources more effectively while maintaining the overall 3D video quality perception by exploiting the ambient illumination changes.

The rest of the paper is organized as follows. In Section 2, the effects of ambient illumination on video quality, depth, and overall 3D video quality perception are discussed. The experimental set-up for the subjective assessments conducted to evaluate these effects, and the results of these assessments are also introduced in Section 2. In Section 3, the proposed user perception model and the metrics developed to measure the content related contextual factors of this model are presented. The proposed adaptation decision taking technique is described in Section 4. In Section 5, the results of the adaptation decision taking experiments are discussed and assessed. Finally, Section 6 concludes the paper.

2. Effects of ambient illumination on 3D video

Ambient illumination of the viewing environment has different effects on the perception of video quality, depth, and overall video quality of 3D video contents. The sensitivity of HVS towards perceiving finer details in a 3D video content changes according to the amount of light captured by the eyes and the adaptation of iris' size according to this amount. While viewing the 3D video content in a dark room, most of the light captured by the eyes comes from the device used to display the content. In this situation, the iris enlarges to let more light in coming from the device. Thus, the details in the 3D content become more distinguishable to the eye. When the 3D video clip is viewed in a bright environment, the eyes capture the ambient light from the display device, light bulbs, windows, reflections of the walls and objects in the room, etc. In this situation, the size of the iris decreases to control the amount of light taken in. Hence, only a small amount of light is captured from the device. As a result, the fine details in the content are less visible to the HVS [5,8].

The sensitivity of HVS for detecting sharpness, shadows, reflections, contrast, etc in a 3D scene, all of which are essential cues to enhance depth perception while viewing a 3D video, decreases when the amount of ambient illumination in the environment increases [7]. The sensitivity of HVS to perceive overall 3D video quality corresponds to how the combination of video quality and depth perception is perceived under a particular ambient illumination condition [5–8]. Subjective experiments conducted to monitor the effect of ambient illumination on video quality, depth, and overall video quality perception of 3D video contents. The experimental set-up for these experiments will be introduced in the following sub-section. Then, the results of the video quality, depth, and overall 3D video quality perception assessments will be discussed in different sub-sections.

2.1. Experimental set-up for the subjective assessments

2.1.1. Stimuli

Nine different test sequences namely: Butterfly, Couples, Ice, Windmill, Advertisement, Chess, Eagle, Football, and Interview were utilized in the subjective experiments. These sequences contain the frames ranging from: 0–150, 60–210, 0–150, 50–200, 30–180, 0–150, 0–150, 0–150, and 0–150 of their respective entire sequences. The test sequences were of High Definition (HD) resolution (i.e., 1920 × 1080 pixels) at 25 fps. The Joint Scalable Video Model (JSVM) reference software version 9.13.1 was used to encode the test sequences with three quality enhancement layers [9]. Medium Grain Scalability (MGS) [10] was utilized as the quality scalability support. Four different channel bandwidths (i.e., 256, 512, 1024, and 1536 kbps) were selected as target source bit rates. 80% of the source bit rate was allocated for the color sequences and the remaining bit rate (i.e., 20%) was allocated for the depth map sequences in the experiments [3]. Different constant QP sets were used to encode the base and quality enhancement layers of color and depth map sequences to match the bit rates of the encoded video sequences to the target bit rates, and the best matching sets were used for the experiments. It should be noted that rate-control is not used while determining the best QPs to encode the sequences. The best matching QPs were determined empirically.

2.1.2. Test methodology and procedure

The effects of the ambient illumination on video quality, depth, and overall 3D video quality perception were assessed in four different ambient illumination conditions (i.e., 5, 52, 116, and 192 lux), created by the self-contained media laboratory facilities of I-Lab, University of Surrey. 5 lux corresponds to a dark condition, while 192 lux indicates a bright light environment. These conditions were measured using a Gretag Macbeth Eye-One Display 2 device [11]. The subjective tests were conducted for each viewer to assess all of the test sequences separately, which were randomly ordered for each environment condition to prevent any prejudice. The Double Stimulus Impairment Scale (DSIS) methodology, as described in International Telecommunication Union-Recommendation (ITU-R) BT-500.11 [12], was used throughout the subjective tests. A 42″-Philips multi-view auto-stereoscopic display, which has a resolution of 1920 × 1080 pixels, was used to display the sequences in the experiments. The 3D video sequences were presented in such a way that they filled up the display in the experiments.

The video quality and depth perception assessment experiments were conducted together. The overall 3D perceptual quality was assessed in different experiments. In the video quality and depth perception experiments, the subjects were asked to assess the video quality and depth perception individually by comparing the impaired videos with the reference ones. For the overall 3D video quality perception experiments, the subjects were asked to give rates to the impaired videos with respect to the reference ones considering the overall 3D video quality perception. Following the experiments, the Mean Opinion Scores (MOSs) [12,13] obtained from all of the subjects were computed. A score of 5 in the assessment scale means the impaired video has no video quality, depth, or overall 3D video quality perception degradations compared to the reference, while a score of 1 presents high perception degradations.

2.1.3. Viewers

18 viewers (6 females and 12 males) participated in the video quality and depth perception assessment experiments. After the outliers were screened using the outlier detection method introduced in Ref. [12] and removed, the MOS scores of 16 viewers (5 females and 11 males) were utilized for the experiments. 18 viewers (6 females and 12 males) participated in the overall 3D perceptual quality assessment experiments. The MOS scores of 16 viewers (5 females and 11 males) were calculated for the experiments after the outliers were detected and removed. All of the attendees were non-expert viewers, whose ages ranged from 20 to 35. Their eye acuity was tested against Snellen eye chart and the stereo vision was tested with the TNO stereo test. All of the viewers had a visual acuity of >0.7 and stereo vision of 60 s of arc. Furthermore, their color vision was verified with the Ishihara test, and all viewers were reported to have good color vision [13].

2.2. Video quality and depth perception assessment experiment results

Fig. 1 illustrates the video quality and depth perception experiment results for a selected test sequence (i.e., Windmill). As observed from Fig. 1(a), which present the video quality results for the sequence, the lowest subjective scores are demonstrated in the 5 lux environment regardless of the varying bit rate. When the ambient illumination of the viewing environment increases (i.e., from 5 lux to 52 lux; to 116 lux; and to 192 lux), the subjective scores given by the viewers also increase. The reason is as above-mentioned the compression artifacts in the impaired 3D video sequence are more visible to the viewers' eyes when the ambient illumination is low (i.e., 5 lux), and they start becoming less and less distinguishable to their eyes when the ambient illumination increases.

As can be seen from Fig. 1(b), the depth perception results demonstrate a reverse subjective score pattern. In the 5 lux environment, the highest subjective scores of perceived depth are monitored regardless of the varying bit rate. When the ambient illumination increases (i.e., from 5 lux to 52 lux; to 116 lux; and to 192 lux), the subjective scores given by the viewers reduce. The reason behind this is that the visibility of depth perception related cues (e.g., contrast, shadows, sharpness, etc) in both the reference and impaired 3D video contents increase when they are viewed in a dark environment. Nevertheless, the depth perception related cues become less visible to the eye under bright environment due to the reflection of the ambient light coming from the windows, objects, light bulbs, etc to the viewing display. As a result, in bright environment, the impaired video content is evaluated with respect to an original 3D video content presenting low depth perception. A more detailed discussion about the video quality and depth perception experiments for all of the test sequences can be found in our previous study in Ref. [8].

2.3. Overall 3D video quality perception assessment experiment results

The results of the overall 3D video quality perception assessment experiments are illustrated in Fig. 2 for all the test sequences utilized in the experiments. As seen from the figures, the overall 3D video quality perception is for the highest level at different ambient illumination conditions for the test sequences. For instance, for the Butterfly sequence (i.e., Fig. 2(a)), the MOS results obtained in the 116 lux ambient illumination condition present higher values than the MOS results reported in the other ambient illumination conditions. The MOS results obtained in the 5, 192, and 52 lux ambient illumination conditions respectively follow those in the 116 lux ambient illumination condition. However, as can be observed from Fig. 1(b), the MOS results recorded in the 52 lux ambient illumination condition outperform the MOS results obtained in the other

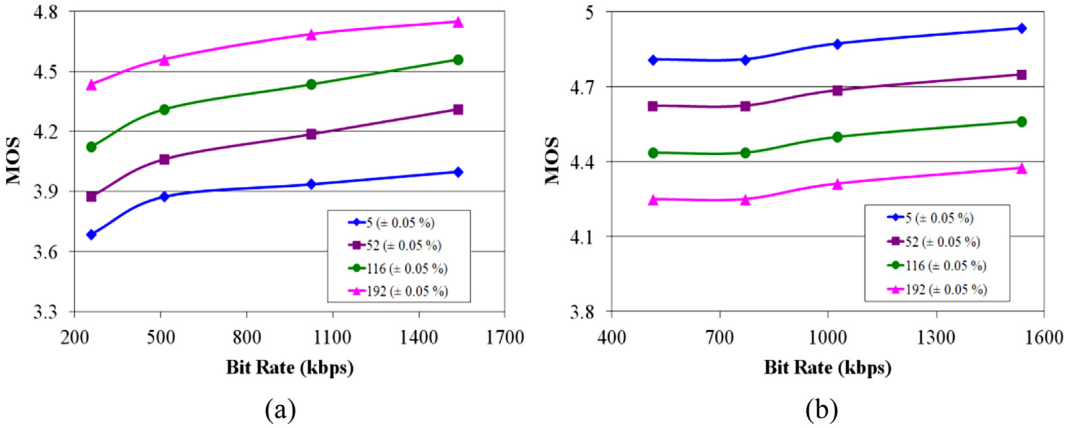
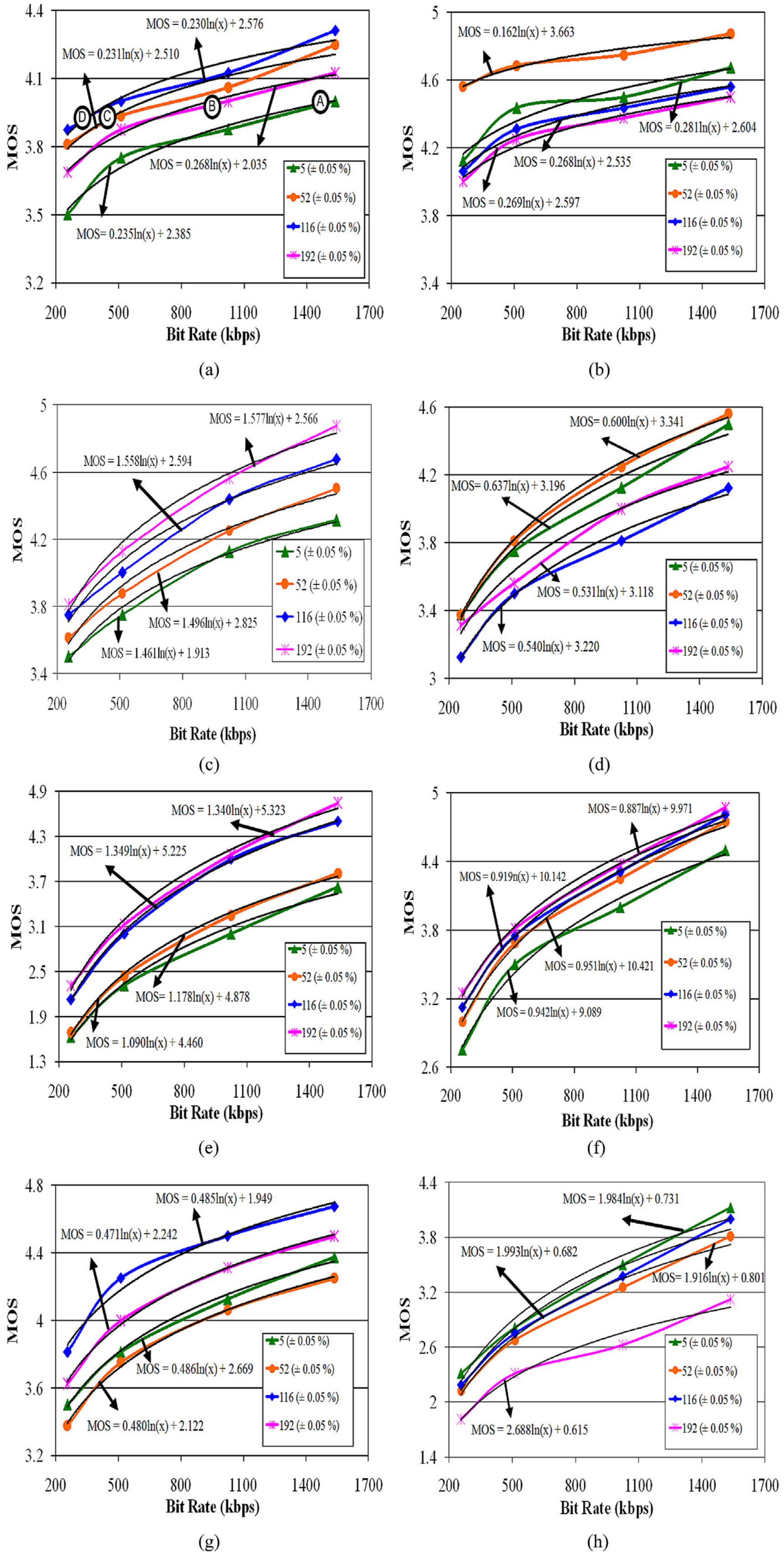
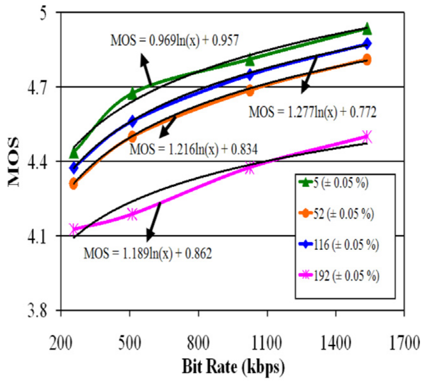


Fig. 1. MOS scores of (a) video quality and (b) depth perception experiments conducted under different ambient illumination conditions (i.e., 5, 52, 116, and 192 lux) for the Windmill sequence.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. MOS scores of subjective experiments for (a) Butterfly (b) Couples (c) Ice (d) Windmill (e) Advertisement (f) Chess (g) Eagle (h) Football (i) Interview sequences.



(i)

Fig. 2. (*continued*).

ambient illumination conditions for the Couples sequence. The MOS results obtained in the 5, 116, and 192 lux ambient illumination conditions respectively follow the ones in the 52 lux ambient illumination condition.

The reason why the level of overall 3D video quality perception varies depending on the ambient illumination condition for the test sequences is related to the degree of presence of natural depth effect in the color texture of the 3D video. Natural depth effect refers to sensing depth in a scene due to relative distance of objects and background, relative size of objects in the scene, etc. Moreover, according to the severity of natural depth effect in the color texture scene, the importance of the video quality and depth perception related cues to enhance the level of overall 3D video quality perception in an ambient illumination condition changes. For instance, if the color texture of the 3D video presents a high natural depth effect, the viewers tend to overlook ambient illumination's effect on depth perception but instead focus on its effect on video quality in a bright environment. Accordingly, the highest overall 3D video quality perception is achieved since the visual artifacts in the color texture sequence cannot be distinguished in this environment (e.g., Fig. 2(c), (e), and (f)). Nevertheless, if the color texture sequence presents a low natural depth effect, the depth perception related cues become more effective than the video quality related cues. Thus, the lowest overall 3D video quality perception is achieved since the depth perception related cues are at the lowest visibility to eye in the bright environment (e.g., Fig. 2(b), (h), and (i)).

As can be viewed from Fig. 2(a), which presents the overall 3D video quality perception assessment results for the Butterfly sequence, the points marked as A, B, C, and D, which are respectively on the 5, 52, 192, and 116 lux curves, result in similar MOSs despite corresponding to different bit rates. Similar behaviors can be realized for the other points on the other curves. It can be revealed after this observations that when the ambient illumination of the viewing environment changes from low to high or low to high (e.g., from 5 to 192 or 192 to 52 or 52 to 116 lux, etc), the overall 3D video quality perception of an input sequence is not compromised by viewing it at a lower bit rate.

3. Modeling user perception

In order to model the perception of users for 3D video contents viewed under particular ambient illumination conditions, firstly, mathematical function of every curve in the graphs are determined, as also illustrated in Fig. 2. It is observed that generic forms of the functions present the following pattern:

$$\mathrm{MOS} = W\ln(B) + Z \tag{1}$$

where, B is the bit rate (kbps), W and Z are two constants. (1) agrees with the fact that the MOS can be represented as a function of bit rate as also reported in Refs. [5,14]. As the second step, it is proposed to devise generic functions for W and Z using content related contexts that can affect their values and ambient illumination context. In this way, it is possible to derive a user perception model for predicting overall 3D video quality perception (i.e., MOS) for a given 3D video sequence encoded at B bit rate and consumed under a given ambient illumination condition.

The "Experimental" column of Table 1 presents the values of the W and Z that approximate the experimental results shown in Fig. 2. As seen from the table, when the ambient illumination changes the values of the constants also change. For example, the W values for the Butterfly sequence are 0.235, 0.231, 0.230, and 0.268 when it is viewed under 5, 52, 116, and 192 lux ambient illumination conditions, respectively. Similar observations can also be made for the Z values. Accordingly, it can be concluded that one of the factors that the values of the constants depend on is ambient illumination. Ambient illumination is referred to as I in the table. However, as observed from the table, the values of the constants are content dependent for the same ambient illumination condition. For instance, the Butterfly and Couples sequences have 0.235 and 0.281 W values, respectively, for the 5 lux ambient illumination condition. Similar findings can be observed for the Z values. This clearly indicates that W and Z are not only ambient illumination dependent but they also depend on content related contexts. The content related contexts affecting the values of W and Z will be explained in the following sub-section.

3.1. *Content related contexts of the user perception model*

It is observed from Table 1 that the values of the constants vary for the video sequences presenting different motion activity characteristics. For instance, for the 52 lux ambient illumination condition, the Interview sequence, which presents low motion, and the Football sequence, which has high motion, have 1.216 and 1.916 W values, respectively. Thus, a metric to measure the motion activity of a color texture sequence is proposed. This proposed metric will be discussed in the following sub-section.

3.1.1. *Motion*

The motion activity of a color texture sequence is measured using its motion intensity in this paper. The optical flow algorithm of pyramidal Lucas and Kanade [15] is used for the motion intensity measurements. Motion information is not distributed to all parts of an image. Therefore, the prominent points which can be taken into account in the optical flow measurements should be determined. Shi and Tomasi algorithm [15] which selects the corners of the objects as the prominent feature points is used in the optical flow measurements in this research study. After the prominent points are selected, they are tracked from frame to frame in a video sequence by the pyramidal Lucas and Kanade algorithm. Let MV(x, y) be the motion vector of a feature point having x and y direction components, the motion intensity of a frame of a sequence is calculated as in (2):

$$\Pi(i) = \sum_{d=1}^{\mathrm{NoP}} \left|\mathrm{MV}_d(x_i, y_i)\right| \tag{2}$$

where, $\Pi(i)$ is the motion intensity of the ith frame of a sequence, d and NoP are a feature point and the number of feature points in

Table 1
Experimental values of *W* and *Z* and their predicted values using their generic functions.

Seq.	*M*	*C*	*L*	*D*	*I*	Experimental		Predicted	
						W	*Z*	*W*	*Z*
Butterfly	0.117	0.020	2.394	2120.45	5	0.235	2.385	0.274	1.403
					52	0.231	2.510	0.237	1.614
					116	0.230	2.576	0.242	1.603
					192	0.268	2.035	0.243	1.585
Couples	0.110	0.039	1.777	1891.69	5	0.281	2.104	0.342	1.460
					52	0.162	2.263	0.264	1.796
					116	0.268	2.035	0.267	1.782
					192	0.269	2.097	0.270	1.762
Ice	0.219	0.009	35.328	3385.02	5	1.461	1.913	1.300	2.657
					52	1.496	2.825	1.233	3.058
					116	1.558	2.594	1.243	3.035
					192	1.577	2.566	1.256	3.002
Wind.	0.182	0.129	32.808	5122.35	5	0.637	3.096	0.516	0.373
					52	0.600	3.341	0.515	0.429
					116	0.540	3.220	0.486	0.427
					192	0.531	3.118	0.490	0.421
Adv.	0.493	0.134	54.237	5044.35	5	1.090	4.460	1.486	5.057
					52	1.178	4.878	1.374	5.821
					116	1.340	5.323	1.387	5.778
					192	1.349	5.225	1.400	5.713
Chess	0.312	0.141	49.009	3916.73	5	0.942	9.089	1.371	8.926
					52	0.951	10.421	1.229	10.275
					116	0.919	10.142	1.394	10.099
					192	0.887	9.971	1.211	10.085
Eagle	0.207	0.083	69.121	3377.64	5	0.486	2.669	1.477	2.807
					52	0.480	2.122	1.365	3.232
					116	0.485	1.949	1.438	3.208
					192	0.471	2.242	1.393	3.172
Football	0.411	0.155	21.514	898.63	5	1.984	0.731	3.166	0.566
					52	1.916	0.801	3.088	0.651
					116	1.993	0.682	3.112	0.647
					192	2.688	0.615	3.145	0.639
Int.	0.088	0.120	29.545	1380.38	5	0.969	0.957	1.444	0.633
					52	1.216	0.834	1.254	0.728
					116	1.277	0.772	1.262	0.723
					192	1.189	0.862	1.276	0.715

the frame, respectively. $MV_d(x_i, y_i)$ is the motion vector of the *i*th frame at feature point *d*.

The motion intensity is measured in terms of pixels. It should also be noted that the motion intensity of a frame is directly proportional to spatial resolution of the frame and inversely proportional to the temporal resolution of the video sequence. Therefore, the normalized average motion intensity over a given video sequence can be calculated as follows:

$$M = \frac{\sum_{i=1}^{\mathrm{NoF}} \Pi(i)}{\mathrm{NoF}} \cdot \frac{F}{S} \tag{3}$$

where, *M* is the motion value in a color texture sequence. NoF is the number of frames in the sequence. *F* and *S* are the frame rate and spatial resolutions of the sequence, respectively.

Using this metric, the motion values of the sequences used in the subjective assessments were measured. Resulting motion values are shown in the "*M*" column of Table 1. As can be realized from the table, the values of the constants, *W* and *Z*, are different for the video sequences that have different motion values. Nevertheless, even though the motion values of the Butterfly and Couples sequences are very similar (i.e., 0.117 and 0.110), the values of the constants are different at the same ambient illumination condition. This clearly shows that motion is not the only content related context that has an effect on the values of the constants, *W* and *Z*.

It has been envisaged that structural feature is another content related factor that has an influence on determining the values. The reason why structural feature is also important for perceiving overall 3D video perception for determining the constant values is that the HVS is sensitive to extract structural information from a scene rather than the errors in the scene [16,17]. Accordingly, the HVS perceives the structural distortion in the objects and background in 3D video contents as quality related visual artifacts. The change in the video quality perception under different ambient illumination conditions corresponds to the visibility of quality related visual artifacts in those conditions. The next sub-section elaborates on how the structural feature is measured in this paper.

3.1.2. *Structural feature*

Contours, which characterize the boundaries of the objects in video frames, are used to represent the structural feature of the visual scenes in this paper. The Canny edge detection algorithm [16,17] was used to determine the contours in the frames without suppressing the pixels that are considered as edges by setting them to 1. To develop the structural feature algorithm, the number of pixels that are set to 1 is counted in every frame of a video sequence [18]. The total value is then normalized using the NoF and *S* to provide consistency across different video sequences as follows:

$$C = \frac{\sum_{i=1}^{\mathrm{NoF}} \delta(i)}{\mathrm{NoF}S} \tag{4}$$

where, *C* is the measured structural feature value in a color texture sequence, $\delta(i)$ is the number of edge pixels in the *i*th frame of the sequence. The measured *C*s of the video sequences are also illustrated the "*C*" column of Table 1.

Both the motion and structural feature are content related factors that have an effect on the video perception. However, content related factors associated with depth perception also have an influence on the perceived overall 3D video quality. Therefore, the important content related factors associated with depth perception should be determined to predict the values of the constants. Luminance contrast [19,20] is envisaged as a depth perception related factor that has an effect on the values. It is an important factor, as it presents varying levels of contrast between the objects and background in a 3D visual scene, which indicate different depth levels. When the contrast in a color texture sequence increases, the depth perception also increases [21,22]. The metric discussed in the next sub-section is proposed for the luminance contrast measurements in this research study.

3.1.3. Luminance contrast

The luminance contrast of a color texture sequence is measured using *Median Absolute Deviation* (*MAD*) in this paper. *MAD* is a measure of statistical dispersion of a set of data [23]. *MAD* is utilized to measure the contrast in a color texture sequence because it computes the distance from the median not the difference from the mean of the data. In this way, it is more suitable to measure the luminance contrast rather than the other statistical methods [24,25]. The *MAD* of a frame of a color texture sequence is measured as follows:

$$MAD(i) = \sum_{k=1}^{S} |t_k - med(t_i)| \tag{5}$$

where, $MAD(i)$ is the luminance contrast of the ith frame of a color texture sequence. t represents each luminance value and $med(t)$ is the median of the luminance values in the frame. The *MAD* computed for each frame is then integrated together to determine the *MAD* across the color texture sequences. The calculated *MAD*s are normalized with NoF and *S* for providing consistent measurement among different video sequences.

Accordingly, the metric presented below is devised for luminance contrast measurements in color texture sequences:

$$L = \frac{\sum_{i=1}^{\mathrm{NoF}} MAD(i)}{\mathrm{NoFS}} \tag{6}$$

where, *L* is the luminance contrast of a color texture sequence.

The luminance contrasts of the sequences utilized in the subjective assessments are also presented in the "*L*" column of Table 1. As can be observed from the table, even though the Ice and Windmill sequences have relatively close *L* values (i.e., 35.328 and 32.808), they have different *W* and *Z* values. Therefore, it is clear that luminance contrast is not the only depth perception related factor that has an effect on calculating the *W* and *Z* values.

Depth intensity is considered as another depth perception associated factor for determining the *W* and *Z* values. Each pixel in the depth map frame of color plus depth representation format of 3D video contents has an associated pixel in the color texture frame. The pixels in the depth map determine the distance of the associated color texture pixel to the viewer. They take gray values ranging from 0 to 255. 0 represents the furthest away pixel from the viewer, while 255 corresponds to the closest pixel to the viewer in a 3D scene [3]. The variation in the pixel depth values in the depth maps corresponds to depth intensity in this paper. Depth intensity is an important depth perception related factor since the pixel depth values aid in perceiving the distance to the objects and background of a 3D video content by the HVS. Depth intensity of a depth map is measured with the proposed metric discussed in the following sub-section.

3.1.4. Depth intensity

Depth intensity is measured by applying standard deviation to the pixel depth values in depth map frames. The reason behind using the standard deviation for the measurement of depth intensity is that it is the measure of the dispersion or variability of a set of values around the mean or arithmetic average of that set [23]. Thus, if the depth map has high variability of the pixel depth values in the depth map frames, the standard deviation of the pixel depth values is expected to be high. The standard deviation in a depth map frame is measured as follows:

$$\mathrm{SD}(i) = \sqrt{\frac{\sum_{k=1}^{S} (x_k(i) - \mu(i))^2}{S}} \tag{7}$$

where, SD(i) is the standard deviation of the ith frame of a depth map frame. x and μ are the pixel depth values and mean of the pixel depth values in the depth map frame, respectively. *S* is the number of pixels in the depth map frame (i.e., width × height of the depth map frame), which is equal to the *S* of the associated color texture frame. Subsequently, the average depth variation over a given depth map sequence, *D*, can be calculated as follows:

$$D = \frac{\sum_{i=1}^{\mathrm{NoF}} \mathrm{SD}(i)}{\mathrm{NoF}} \tag{8}$$

where, NoF is the number of frames in the depth map which is the same as the number of frames in the corresponding color texture sequence. The depth intensities of the video sequences used in the subjective assessments were measured using (8), and the measurement results were also presented in "*D*" column of Table 1.

The observations from the subjective tests indicate that the generic functions of *W* and *Z* need to consider five factors: the ambient illumination condition, the motion, the structural feature, and the luminance contrast of the input color texture sequence, and the depth intensity of the input depth map sequence to devise the user perception model.

3.2. Generic functions for the constants

In order to devise generic functions for *W* and *Z* (i.e., the constants) using the contextual factors of the user perception model, the graphs representing the *M* vs *W*, *C* vs *W*, *L* vs *W*, *D* vs *W*, *I* vs *W*, *M* vs *Z*, *C* vs *Z*, *L* vs *Z*, *D* vs *Z*, and *I* vs *Z* were plotted, as presented in Fig. 3. Then, curve fitting functions [26] were utilized to approximate the relationships between all of these pairs, as also illustrated in the figure. The constants in the functions are shown with *a*, *b*, *c*, *d*, e, *f*, *g*, *h*, *j*, *k*, *l*, *m*, *n*, *o*, *p*, *r*, *s*, *t*, *u*, *v*, *y*, υ, *w*, *z*, α, β, ε, ø, λ, π, τ, ψ, ω θ,Ώ, Γ, Π, T, ∞, *, ×, ℘, κ, Φ, and ∩ in the figure.

As the third step, a set of numerical constants were introduced to each of the functions to calculate the *W* and *Z* values best correlating with the *W* and *Z* values obtained experimentally. as discussed in Ref. [5] Subsequently, the functions of *W* and *Z* were integrated together to devise the generic functions of them as follows:

$$W = (f(M)f(C)f(D)f(L)f(I)), \quad Z = (g(M)g(C)g(D)g(L)g(I)) \tag{9}$$

The *W* and *Z* values predicted utilizing (9) are also shown in the "Predicted" column of Table 1 for the test sequences. The

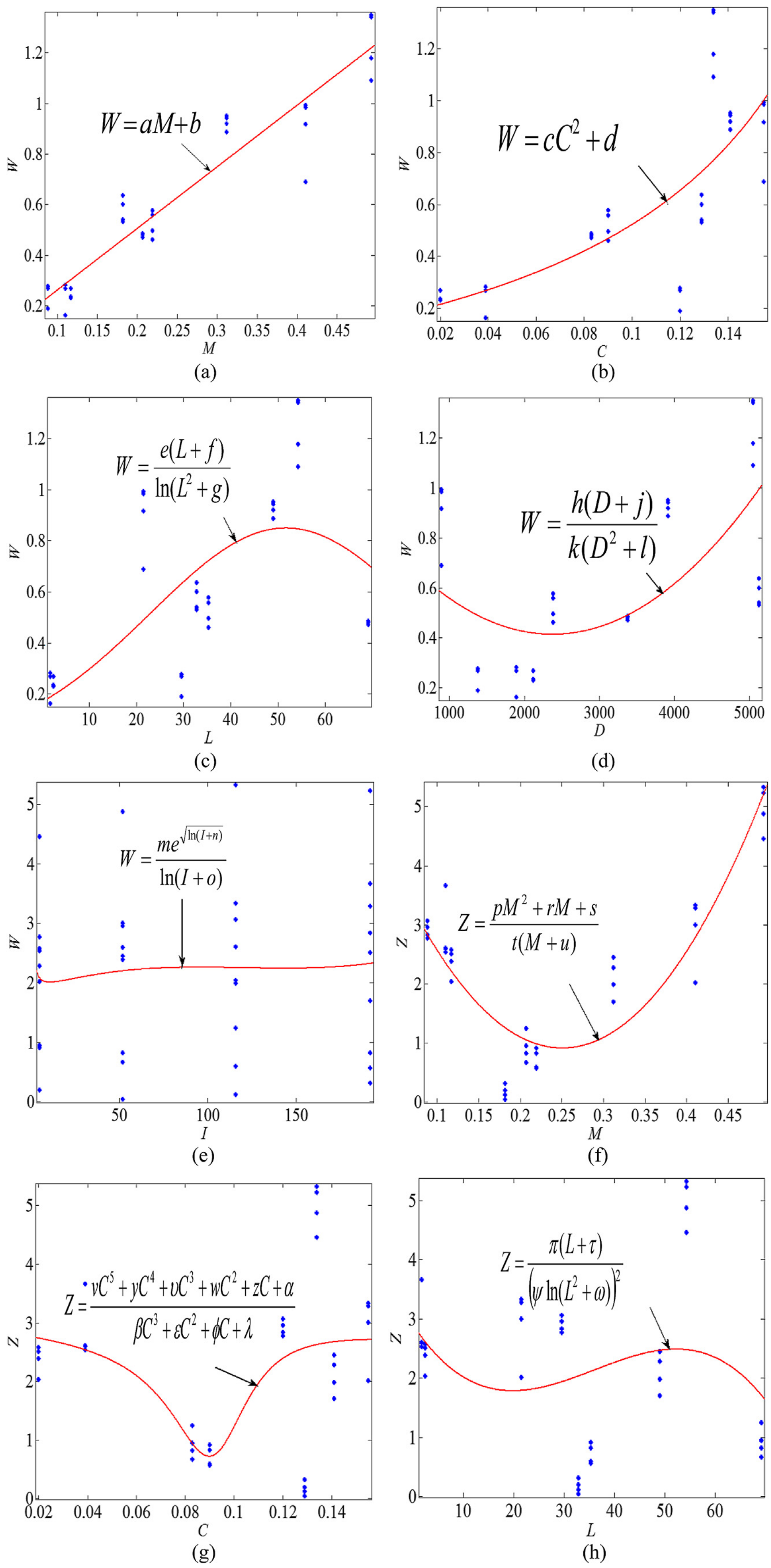
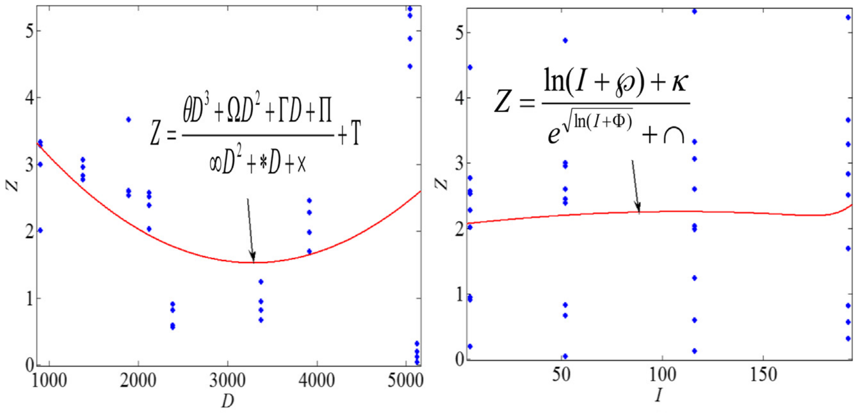


Fig. 3. *W* vs (a) *M* (b) *C* (c) *L* (d) *D* (e) *I*; and *Z* vs (f) *M* (g) *C* (h) *L* (i) *D* (j) *I*.



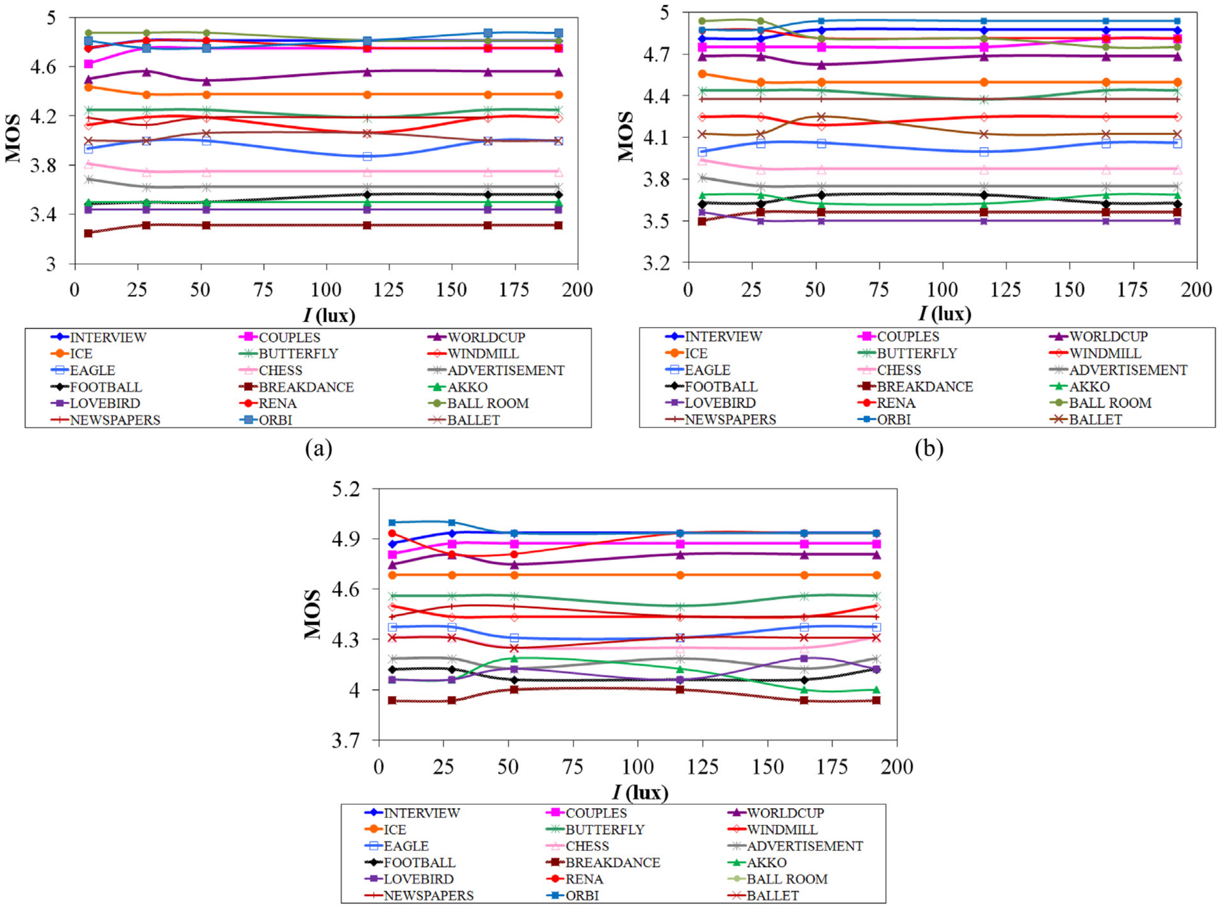
(i) (j)

Fig. 3. (*continued*).

correlation coefficients [27] between the values observed using the overall subjective experiments and the ones predicted for *W* and *Z* are determined as 96% and 93%, respectively. It should be noted that even though the values of the experimental and predicted values seem to be large for some sequences, the most important factor to correlate the experimented and predicted *W* and *Z* values is the pattern that these values present.

4. Proposed adaptation decision taking technique

In this section, the derivation of the proposed adaptation decision taking technique is discussed with the help of a use-case scenario. Let us consider a user is watching an input 3D video sequence in environment *X* (e.g., a dim room at home), which has I_X lux, using mobile, tablet PC, etc. While continuing



(a) (b)

(c)

Fig. 4. MOS scores at (a) 384 (b) 512 and (c) 768 kbps under different ambient illumination conditions.

Table 2
Resulting adapted bit rates using the proposed technique.

Sequence	I_Y (lux)	B_Y (kbps)		
		$B_X = 384$	$B_X = 512$	$B_X = 1024$
Butterfly	28	385.680	531.446	1172.466
	52	398.907	556.289	1239.628
	116	368.870	510.879	1119.744
	164	417.428	576.590	1255.666
	192	387.683	536.216	1171.488
Couples	28	414.072	598.540	1454.262
	52	426.940	619.730	1521.013
	116	420.321	607.574	1476.210
	164	435.603	624.584	1488.219
	192	423.250	609.308	1465.892
Ice	28	370.175	500.105	1032.437
	52	383.045	518.758	1077.258
	116	371.966	502.525	1037.432
	164	360.546	484.782	989.379
	192	359.176	483.737	991.181
Windmill	28	496.840	674.301	1407.451
	52	348.226	464.547	930.278
	116	495.819	672.915	1404.558
	164	549.377	748.456	1576.667
	192	477.131	645.940	1340.179
Advertisement	28	367.736	502.046	1062.956
	52	357.470	487.920	1032.483
	116	348.956	474.913	997.923
	164	353.019	478.032	992.386
	192	346.195	469.809	980.415
Chess	28	323.481	450.835	1003.190
	52	254.670	351.025	760.528
	116	249.976	299.015	693.477
	164	354.519	490.070	1019.221
	192	323.498	448.026	981.908
Eagle	28	457.890	624.656	1320.150
	52	458.018	625.257	1323.596
	116	341.233	458.527	934.387
	164	417.228	564.670	1170.695
	192	422.768	573.536	1195.944
Football	28	293.833	450.321	963.562
	52	433.891	582.723	1185.944
	116	414.571	555.511	1124.385
	164	376.789	502.416	1004.981
	192	390.155	521.192	1047.145
Interview	28	549.787	748.503	1574.174
	52	545.286	742.192	1559.976
	116	526.014	714.562	1494.847
	164	481.827	651.035	1346.066
	192	494.157	669.035	1388.316
Break Dance	28	497.313	675.764	1414.641
	52	493.077	669.916	1401.936
	116	474.385	643.038	1338.242
	164	437.661	590.130	1212.591
	192	447.358	604.310	1247.213
World Cup	28	459.695	619.969	1274.547
	52	364.828	447.714	970.126
	116	439.024	590.732	1207.795
	164	404.751	542.176	1096.529
	192	412.746	553.587	1123.037
Akko	28	398.647	582.235	1338.511
	52	409.356	604.723	1405.684
	116	404.251	584.625	1368.425
	164	418.823	608.462	1372.357
	192	407.336	592.451	1347.852
Lovebird	28	386.127	568.436	1421.393
	52	397.458	588.682	1496.537
	116	388.593	571.385	1452.742
	164	405.646	596.835	1461.274
	192	394.742	579.724	1443.842
Rena	28	289.569	408.672	981.428
	52	242.596	325.153	727.385
	116	227.845	268.547	658.352
	164	326.657	458.493	984.594
	192	294.341	426.583	951.624
Ball Room	28	377.325	558.936	1124.578
	52	385.125	574.647	1185.364
	116	392.462	555.663	1143.683
	164	398.536	587.382	1158.342
	192	388.452	558.486	1137.842
Newspapers	28	532.614	731.826	1559.124
	52	532.457	729.126	1542.347
	116	513.248	699.442	1482.732
	164	468.215	647.032	1332.468
	192	481.132	652.473	1368.914
Orbi	28	479.617	658.231	1388.152
	52	331.164	448.332	1014.984
	116	472.812	651.815	1391.448
	164	538.177	731.158	1498.231
	192	459.128	597.846	1326.125
Ballet	28	307.137	414.732	986.172
	52	248.121	337.154	743.246
	116	234.674	283.963	678.614
	164	338.418	475.524	1004.363
	192	306.782	431.223	967.835

to watch the sequence, the user goes into environment Y (e.g., a bright room at home), which has I_Y lux. In order to maintain the overall 3D video perception, the input 3D video sequence can be adapted to a lower or higher bit rate depending on the I_X, I_Y, and the characteristics of the content viewed (e.g., motion, depth intensity, etc).

It has been taken into consideration according to the scenario that the perceived overall 3D video quality should be maintained regardless of the difference in the ambient illumination of the viewing environment, as follows:

$$MOS_X = MOS_Y \tag{10}$$

where, MOS_X and MOS_Y present the perceived overall 3D quality of input video sequence viewed in I_X lux and adapted video sequence viewed in I_Y lux, respectively. Using (1) and (10), the proposed adaptation decision taking technique is devised to solve the following equation:

$$W_X \ln(B_X) + Z_X = W_Y \ln(B_Y) + Z_Y \tag{11}$$

$$B_Y = e^{\frac{W_X \ln(B_X) + Z_X - Z_Y}{W_Y}} \tag{12}$$

where, W_X and Z_X are the constants for the 3D video sequence to be adapted (i.e., the input 3D video sequence), and W_Y and Z_Y are the constants for the adapted 3D video sequence. Equation (12) can be exploited to calculate the output bit rate (B_Y) for adapting an input 3D video sequence from a given input bit rate (B_X) under a specified amount of ambient illumination change without compromising the perceived overall 3D video quality.

5. Results and discussion

In this section, the results of the adaptation decision taking experiments to validate the efficiency of the proposed adaptation decision taking technique are discussed. The experiments were conducted considering the use-case scenario discussed in Section 4 for all of the test sequences and 9 additional test sequences (i.e., Break Dance, World Cup, Akko, Lovebird, Rena, Ball Room, Newspapers, Orbi, and Ballet) were used for the experiments. In order to maintain a common reference, ambient illumination for all of the adaptation experiments I_X was kept as 5 lux, whereas five different I_Ys (i.e., 28, 52, 116, 164, and 192 lux) were utilized in the experiments. Three different B_Xs (i.e., 384, 512, and 1024 kbps) were used for the input sequences during the experiments. The resultant B_Ys utilizing the B_Xs are shown in Table 2. 80% of the resultant B_Ys was allocated for the color texture sequences whereas 20% of the remaining bit rate was allocated for the depth maps. For example, as observed from the Butterfly sequence $B_Y = 385.680$ kbps is calculated from $B_X = 384$ kbps. 80% of the B_Y, which is 308.544 kbps, is used to encode the color texture sequences and 20% of the B_Y, which is 77.136 kbps, is utilized to encode the depth map sequences. The bit rates allocated for the color texture and depth map sequences for the remaining results in Table 2 can be computed in the same way.

As seen from the figure, the resultant B_Ys get values higher and lower than the B_Xs depending on the ambient illumination condition values (I_X and I_Y) and 3D video content characteristics (i.e., motion, structural feature, luminance contrast, and depth intensity). Thus, the resultant B_Y values on their own cannot reflect the performance efficiency of the proposed model. The opinion of the real human observers towards the 3D video contents having the resultant B_Ys should be investigated under different I_Ys to understand the performance of the resultant B_Ys. Therefore, further subjective experiments were conducted to validate the efficiency of the resultant B_Ys towards maintaining the overall 3D perceptual quality regardless of the change from I_X to I_Ys. The adapted 3D video sequences were presented to the users under the aforementioned conditions during the experiments. The DSIS method was utilized in the experiments as recommended by the ITU-R BT-500.11 [12]. 19 (9 females and 10 males) viewers participated in the experiments, and after the outliers were detected, the MOS scores were calculated for 16 (8 females and 8 males) subjects in the experiments, which is in compliance with the ITU-R recommendation.

The results of the subjective tests are illustrated in Fig. 4. As observed from the figure, the 3D perceptual qualities of the video sequences slightly vary (~up to 5%) under changing ambient illumination conditions. This observation proves the efficiency of the proposed technique to determine the output bit rate of the 3D video sequence considering its input bit rate under different ambient illumination conditions.

6. Conclusion

In this paper, a user perception model, which relies on determining the perception characteristics of a user for 3D video content viewed in a particular ambient illumination condition, has been proposed. Motion, structural feature, luminance contrast, and depth intensity characteristics of 3D video, which have acted as the primary content related contexts associated with 3D video perception, is exploited to devise this model. A 3D video bit rate adaptation decision taking technique has been developed by exploiting the user perception model. Further subjective assessments have verified that the proposed technique is efficient to determine adequate bit rate to adapt 3D video content while maintaining overall 3D video quality perception under ambient illumination changes. In our future studies, other contexts and content related characteristics that can be utilized in 3D video bit rate adaptation decision taking will be considered to enhance the application of the developed technique in various scenarios.

References

[1] Y. Liu, S. Ci, J. Liu, Y. Qi, Integrating stereoscopic image transcoding with retargeting for mobile streaming, in: Visual Communications and Image Processing Conference 2012 (VCIP 2012), San Diego, CA, USA, 27–30 Nov. 2012.
[2] M.B. Kim, J. Nam, W. Baek, J. Son, J. Hong, The adaptation of 3D stereoscopic video in MPEG-21 DIA, Elsevier Signal Proc. – Image Com. Spec. Issue Mul. Ad. 18 (Sep. 2003) 685–697.
[3] Y. Liu, S. Ci, H. Tang, Y. Ye, J. Liu, QoE-oriented 3D video transcoding for mobile streaming, ACM Transac. Multimed. Comput. Commun. Appl. (TOMCCAP) 8 (3) (2012) article 42.
[4] G. Nur, H. Kodikara Arachchi, S. Dogan, A.M. Kondoz, Ambient illumination as a context for video bit rate adaptation decision taking, IEEE TCSVT 20 (12) (Dec. 2010) 1887–1891.
[5] G. Nur, H. Kodikara Arachchi, S. Dogan, A. Kondoz, Advanced adaptation techniques for improved video perception, IEEE Trans, Circuits Syst. Video Technol. 22 (February 2012) 225–240.
[6] G.J. Burton, S. Nagshineh, K.H. Ruddock, Processing by the human visual system of the light and dark contrast components of the retinal image, Springer Biol. Cybern. 27 (4) (1977) 189–197.
[7] T.R. Robinson, Light intensity and depth perception, Am. J. Psychol. 7 (4) (Jul. 1896) 518–532.
[8] G. Nur, S. Dogan, H. Kodikara Arachchi, A.M. Kondoz, Assessing the effects of ambient illumination change in usage environment on 3D video perception for user centric media access and consumption, in: 2nd International ICST Conference on User Centric Media, Palma de Mallorca, Spain, 1–3 Sep. 2010.
[9] JSVM 9.13.1. CVS Server [Online]. Available Telnet: garcon.ient.rwth-aachen.de:/cvs/jvt.
[10] T.C. Thang, J.W. Kang, J.-J. Yoo, J.-G. Kim, Multilayer adaptation for MGS-based SVC bitstream, in: Proc. of the 16th ACM Multimedia, Vancouver, British Columbia, Canada, 26–31 Oct. 2008.
[11] Gretag Macbeth eye-one display 2 [Online]. Available: http://www.xrite.com.
[12] ITU-R BT.500–513, Methodology for the Subjective Assessment of the Quality of Television Pictures, 2012.
[13] ITU-R BT.1438, Methodology for the Subjective Assessment of Stereoscopic Television Pictures, 2000.
[14] G. Cermak, M. Pinson, S. Wolf, The relationship among video quality, screen resolution, and bit rate, IEEE Trans. Broadcast. 57 (2) (Jun. 2011).
[15] D.J. Fleet, Y. Wiess, Optical flow estimation in Paragios, in: Handbook of Math. Models in Comp. Vis., Springer, 2006, p. 239.
[16] J. Shi, C. Tomasi, Good features to track, in: IEEE Conf. On Com. Vis. and Pat. Recog., Seattle, USA, Jun. 1994.
[17] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Proc. (2004) 600–612.
[18] C. Grigorescu, N. Petkov, M.A. Westenberg, Contour and boundary detection improved by surround suppression of texture edges, Image Vis. Comput. 22 (2004) 609–622.
[19] J. Malik, S. Belongie, T. Leung, J. Shi, Contour and texture analysis for image sequenceation, Int. J. Comput. Vis. 1 (2001) 7–27.
[20] R.A. Frazor, W.S. Geisler, Local luminance and contrast in natural images, Elsevier Vis. Res. J. 46 (10) (May. 2006) 1585–1598.
[21] W.S. Geisler, Visual perception and the statistical properties of natural scenes, Rev. Psychol. 59 (2008) 167–192.
[22] S. Ichihara, N. Kitagawa, H. Akutsu, Contrast and depth perception: effects of texture contrast and area contrast, Perception 36 (5) (2007) 686–695.
[23] V. Jones, Mean Direction and Mean Absolute Deviation, ASTM Standards and Engineering Digital Library, Jan. 2009.
[24] J.L. Devore, Probability and Statistics for Engineering and the Sciences, Duxbury, 1995.
[25] P. Hall, M.P. Wand, On the minimization of absolute distance in kernel density estimation, Stat. Probab. Lett. 6 (Apr. 1988) 311–314.
[26] Mathworks, Curve Fitting Toolbox, 2013 (Online). Available: http://www.mathworks.com/access/helpdesk/help/toolbox/curvefit/.
[27] Wolfram Mathworld, Correlation Coefficient, 2013 (Online). Available: http://mathworld.wolfram.com/CorrelationCoefficient.html.