
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Modeling user perception of 3D video based on ambient illumination context for enhanced user centric media access and consumption

PLEASE CITE THE PUBLISHED VERSION

<http://dx.doi.org/10.1007/s11042-011-0824-z>

PUBLISHER

© Springer Verlag

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Nur, Gokce, Hemantha Kodikara Arachchi, Safak Dogan, and Ahmet Kondo. 2019. "Modeling User Perception of 3D Video Based on Ambient Illumination Context for Enhanced User Centric Media Access and Consumption". figshare. <https://hdl.handle.net/2134/24242>.

Modeling User Perception of 3D Video Based on Ambient Illumination Context for Enhanced User Centric Media Access and Consumption

G. Nur, H. Kodikara Arachchi, S. Dogan, and A. M. Kondo

*I-Lab Multimedia Communications Research, Centre for Vision, Speech and Signal Processing,
Faculty of Engineering and Physical Sciences,
University of Surrey, Guildford GU2 7XH, Surrey, UK
{G.Nur, H.Kodikaraarachchi, S.Dogan, A.Kondo}@surrey.ac.uk*

Abstract— For enjoying 3D video to its full extent, it is imperative that access and consumption of it is user centric, which in turn ensures improved 3D video perception. Several important factors including video characteristics, users' preferences, contexts prevailing in various usage environments, etc have influences on 3D video perception. Thus, to assist efficient provision of user centric media, user perception of 3D video should be modeled considering the factors affecting perception. Considering ambient illumination context to model 3D video perception is an interesting research topic, which has not been particularly investigated in literature. This context is taken into account while modeling video quality and depth perception of 3D video in this paper. For the video quality perception model: motion and structural feature characteristics of color texture sequences; and for the depth perception model: luminance contrast of color texture and depth intensity of depth map sequences of 3D video are used as primary content related factors in the paper. Results derived using the video quality and depth perception models demonstrate that these models can efficiently predict user perception of 3D video considering the ambient illumination context in user centric media access and consumption environments.

Index Terms— 3D video, 3D user perception model, ambient illumination context, depth perception, user centric 3D media access and consumption, video quality perception.

1. INTRODUCTION

3-Dimensional (3D) video distribution and access in heterogeneous usage environments have posed significant research and technology development challenges for delivering 3D video to a wide range of users. These challenges have been exacerbated not only by the existence of different networking infrastructures, diverse user terminals, and numerous media content representations, but also by the users themselves and their various preferences and high lev-

els of expectations. In turn, this has led the research efforts to focus on user centricity while providing media services rather than conventional purely technology orientated service provisions.

Although the development of 3D video related technologies such as capture, representation, coding, transmission, rendering, etc are evolving promptly, the influence of these technologies on 3D perceptual quality to enrich proliferation of user centric 3D video applications into the consumer market has not been thoroughly investigated to date. One of the reasons for this is the lack of available 3D video perception models to predict the user perception towards 3D video. Even though, there are commonly used 2D objective quality assessment models available in literature, which correlate with Human Visual System (HVS), these models cannot provide accurate video quality and depth perception assessments for 3D video due to the multi-dimensional nature of it (e.g., video quality, depth perception, naturalness, etc) [1]. Therefore, reliably assessing 3D video perception can only be performed using subjective quality assessment techniques currently, which are conducted with real human observers. However, subjective quality assessment techniques are expensive in terms of time and effort. Thus, 3D user perception models should be developed to assist the advancement of 3D media technologies that support user centric access and consumption of 3D video, which in turn will ensure enhanced 3D user perception.

User perception of 3D video is influenced by several factors including 3D video characteristics, usage environment contexts, etc. Such factors should firstly be determined to develop models for predicting 3D user perception accurately. The effects of varying ambient illumination context in the usage environment surrounding the user on the use and experience of video quality and depth perception of 3D video have been reported in our previous study [2]. Extending on this work, video quality and depth perception of 3D video are modeled considering ambient illumination context in this paper.

Color-plus-depth map 3D video representation is employed for the research carried out in the paper since it has many advantages compared to the left and right 3D video representation, and has also been highly exploited in research and standardization activities to date [1].

In order to model user perception of 3D video considering ambient illumination context, the video quality and depth perception of users towards 3D video are monitored under different ambient illumination conditions through a series of subjective experiments initially. Afterwards, video quality and depth perception models are developed by exploiting the results of the subjective experiments conducted. Motion and structural feature of color texture se-

quences are exploited as the primary content related contexts in the video quality perception model. Luminance contrast of color texture and depth variation of depth map sequences are considered as the content related contexts while developing the depth perception model.

The rest of the paper is organized as follows. Assessment of 3D video quality and depth perception using the existing quality assessment methods is discussed in Section 2. Section 3 introduces the experimental set-up for the subjective experiments and discusses the results of these experiments. The developed video quality and depth perception models are defined in Section 4. Section 5 discusses the video quality and depth perception assessment results predicted using the developed models. Moreover, the video quality and depth perception results of the subjective experiments are compared with those of the predicted ones using the developed models in this section. The paper is concluded in Section 6.

2. 3D VIDEO QUALITY ASSESSMENT USING THE EXISTING QUALITY ASSESSMENT METHODS

Even though a significant amount of work has been carried out to develop 2D objective quality assessment techniques to measure perceptual quality of 2D video contents, same effort has not been performed for developing 3D objective quality assessment techniques to date. There are three commonly objective quality assessment models for 2D video in literature namely: Peak Signal-to-Noise Ratio (PSNR), Video Quality Metric (VQM), and Structural SIMilarity (SSIM).

PSNR is a widely used 2D objective quality assessment technique that is measured on logarithmic scale. It relies on the Mean Squared Error (MSE) between an original and a distorted video frame. The MSE calculation is as follows [3]:

$$MSE = \frac{\sum_{i=1}^M \sum_{j=1}^N [g(i,j) - G(i,j)]^2}{M \cdot N} \quad (1)$$

where $g(i,j)$ is the original signal at pixel (i,j) , $G(i,j)$ is the degraded signal and $M \times N$ is the frame size, where M and N refer to the width and height of a frame, respectively. The resultant (i.e., MSE value) is a single number in decibels (dBs). For n bits per frame, PSNR is calculated as follows [4]:

$$PSNR = 20 \cdot \log_{10} \left(\frac{2^n - 1}{\sqrt{MSE}} \right) \quad (2)$$

PSNR has low computational complexity and thus produces quick results. However, it has poor correlation with real human perception [5]. Limitation of PSNR has led the researchers to develop quality evaluation techniques that approximate the responses of real human observers. VQM is an objective quality measurement technique that provides video quality assessment, which is highly correlated to HVS. VQM comprises two phases for video quality assessments, namely: Calibration and Calculation. Calibration phase involves spatial alignment, valid region estimation, gain-offset calculation, and temporal alignment of the distorted video contents. Jerkiness, color distortion, blurriness, global noise, and block distortion of the calibrated video content are measured compared to the original one within the VQM's Calculation phase. As a result, both the spatial impairments (e.g., blurriness) and temporal impairments (e.g., jerkiness) are determined at the same time for the quality evaluations using VQM. Due to its good correlation with HVS, VQM has been standardized by the American National Standards Institute (ANSI) in 2004, and added as a normative video quality estimation method in the International Telecommunication Union-Recommendation (ITU-R) BT. 500-11 [5] [7] [8]. Compared to PSNR, however, VQM has higher computational complexity [5].

SSIM is another objective quality measurement technique, which evaluates the structural distortion of a distorted video compared to the original one. SSIM is formed based on the fact that HVS is specialized in extracting structural information from a visual scene rather than the errors in the scene [9].

The equation for SSIM is shown below:

$$SSIM = \frac{(2\bar{x}\bar{y} + C_1)(2\sigma_{xy} + C_2)}{(\bar{x}^2 + \bar{y}^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

where, $x = \{x_i | i = 1, 2, \dots, N\}$ is the original signal, $y = \{y_i | i = 1, 2, \dots, N\}$ is the distorted signal, and i is the pixel index. \bar{x} , \bar{y} , σ_x^2 , σ_y^2 and σ_{xy}^2 are the mean of x , mean of y , the variance of x , the variance of y and the covariance of x and y , respectively. C_1 and C_2 are constants. The SSIM metric takes values between 0 and 1. 0 corresponds to the lowest quality whereas 1 corresponds to the highest quality. SSIM has less computational complexity than VQM [9].

Due to the multi-dimensional nature of 3D video (e.g., video quality, depth perception, naturalness, etc), these commonly used 2D objective quality assessment techniques cannot provide 100% accurate measurements for video quality and depth perception of 3D video [1]. Currently, assessing 3D video quality in a reliable way can only be per-

formed using subjective quality assessment techniques conducted with real human observers. Nevertheless, subjectively assessing video quality is highly time and effort consuming. Therefore, 3D user perception models should be developed to enrich the progression of user centric 3D related video technologies to yield enhanced 3D video perception of users. Moreover, content-related characteristics and usage environment contexts, which have significant effects on 3D video perception, should also be considered while developing these models to enrich the enhanced perception.

3. SUBJECTIVE ASSESSMENT OF 3D VIDEO QUALITY AND DEPTH PERCEPTION UNDER DIFFERENT AMBIENT ILLUMINATION CONDITIONS

In order to understand how 3D video quality and depth are perceived under different ambient illumination conditions, a set of subjective experiments were conducted in a controlled environment. These experiments were conducted according to the ITU-R BT. 500-11 [8] guidelines except for those related to the ambient illumination condition. In this section, first of all, the experimental set-up for the subjective assessments is explained. Afterwards, the results are discussed.

3.1. *Experimental Set-up*

For the experiments, publicly available *Interview*, *Chess*, *Windmill*, *Ice*, *Advertisement*, and *Eagle* test sequences were used. The color texture and depth map sequences of the 3D video clips 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 [10] was used to encode the sequences. Four different channel bandwidths (i.e., 512, 768, 1024, and 1536 kbps) were selected as target bit rates. 80% of the target bit rate was allocated for the color sequences and the remaining bit rate (i.e., 20%) was allocated for the depth map sequences while performing the experiments [11]. These sequences were displayed on a 42" Philips multi-view auto stereoscopic display, which has a resolution of 1920×1080 pixels.

The effects of the ambient illumination on perceptual quality and depth perception were assessed under four different ambient illumination conditions (i.e., 5, 52, 116, and 192 lux), created in 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 [12]. 16 volunteers (5 females and 11 males) participated in the experiments. They were all 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 them surpassed 0.7 eye acuity and 60 seconds of arc stereo vision levels, respectively. Furthermore, their color vision was verified with the Ishihara test, and all viewers were reported to have good color vision [1]. The subjective tests were conducted with each viewer to assess all of the test sequences individually, which were randomly ordered for each environment condition to avoid any potential prejudices. The subjects were asked to assess both the video quality and depth perception by comparing the impaired video sequences with the reference ones. Following the experiments, the Mean Opinion Scores (MOSs) [8] obtained from all of the viewers were computed. A score of 5 in the MOS assessment scale means the impaired video has the same perceptual quality or depth perception as the reference one, while a score of 1 means very annoying presentation. The tests lasted 20 minutes on average, including the initial training session.

3.2. *Subjective Assessment Results*

The results of the video quality and depth perception assessment experiments are presented in this sub-section.

3.2.1. *Video Quality Perception Results*

Fig. 1 illustrates the bit rate versus MOS results reported on viewers' video quality perception assessment experiments. As can be observed from the results, the video sequences viewed under different ambient illumination conditions have demonstrated an increasing perceptual quality rating pattern as the amount of light in the environment increases across all of the bit rate range that was considered in the experiments. Here, the perceived video quality presents the lowest subjective scores in the 5 lux environment compared to those in the other environments 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 also increase. The sensitivity of HVS towards perceiving finer details in a 3D video sequence changes according to the amount of light the eyes capture from the viewing environment and the iris' adaptation of its size corresponding to this amount.

While viewing the video sequence in a dark room, almost all the light captured by the eyes is originated at the device used to display the content. In this situation, the iris enlarges to let more light in that comes from the device. Thus, the details in the content become more distinguishable to the eye. When the 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, because excessive light energy would harm the eye. 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.

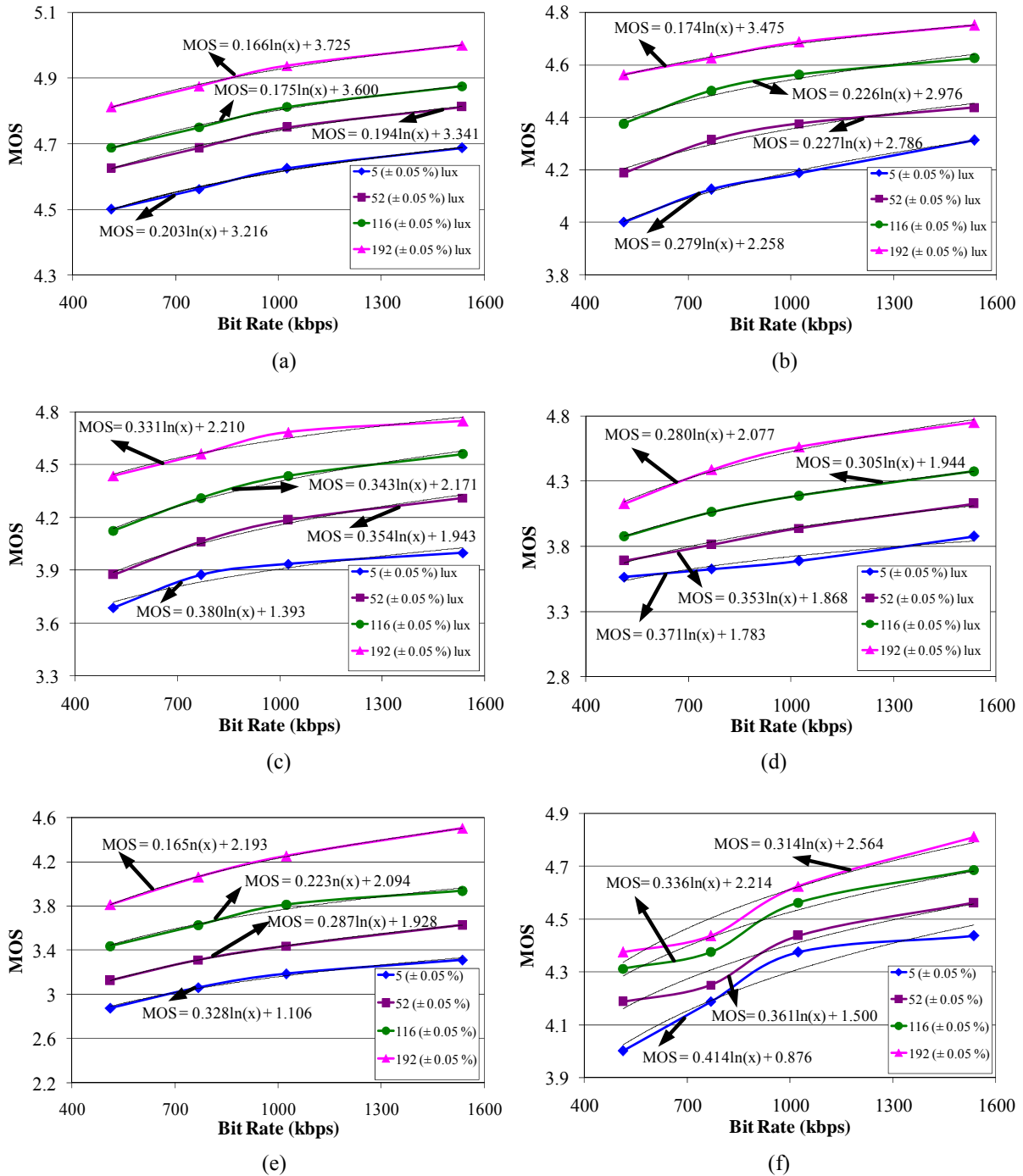
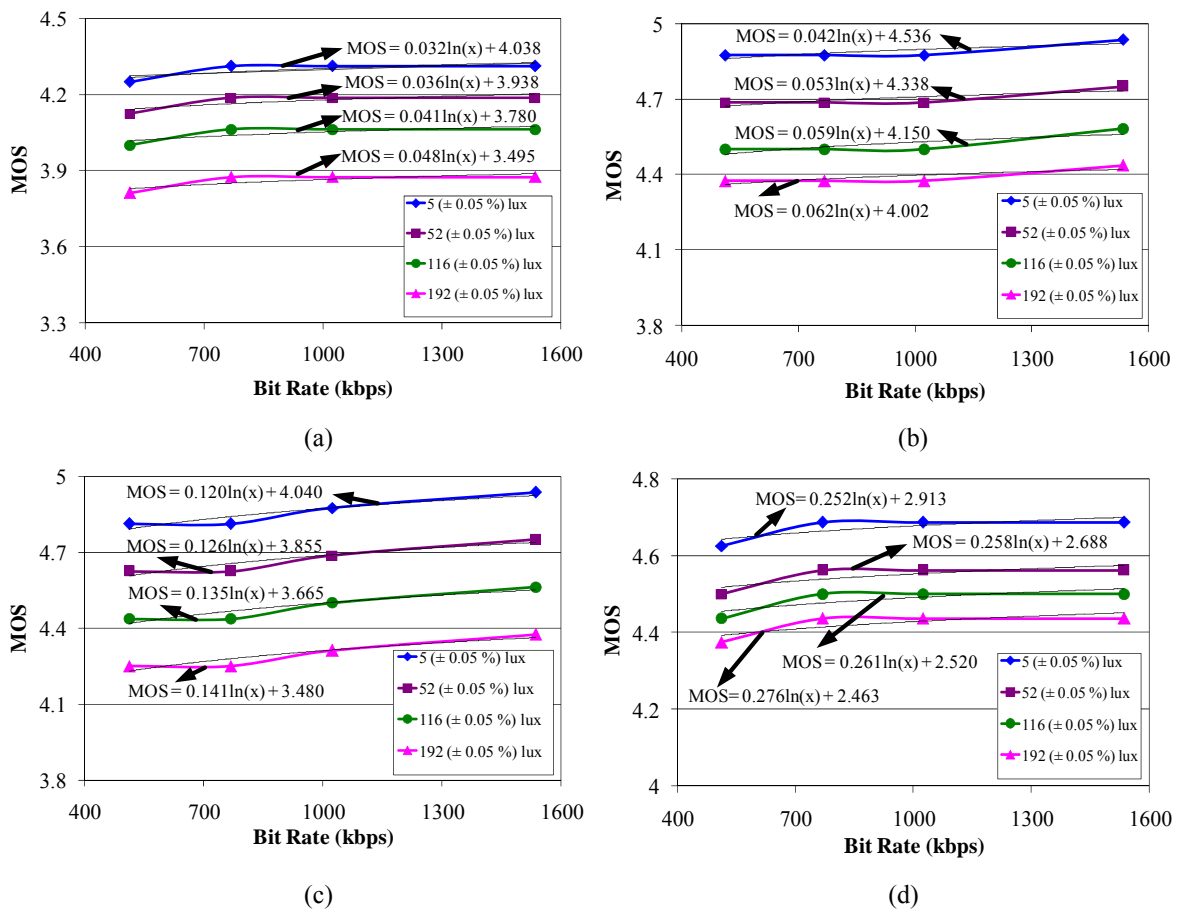


Fig. 1. The (a) *Interview* (b) *Chess* (c) *Windmill* (d) *Ice* (e) *Advertisement* (f) *Eagle* sequence bit rate versus MOS under different ambient illumination conditions (5, 52, 116, and 192 lux) based on video quality perception assessments

3.2.2. Depth Perception Results

The depth perception assessment results are shown in Fig. 2. As can be realized from the results, the video sequences viewed under different ambient illumination conditions have demonstrated a decreasing depth perception rating pattern as the amount of light in the environment increases across all of the bit rate range. Indeed, the perceived depth presents the highest subjective scores in the 5 lux environment when compared with those in the other environments 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. These results reveal that the lower the amount of ambient illumination in the 3D video access and consumption environment the better the depth perception is. This can be explained as follows: the sensitivity of HVS towards detecting sharpness, shadows, reflections, contrast, etc in the visual content, all of which are essential cues to enhance depth perception of 3D video, decreases due to the increase in the ambient illumination [13].



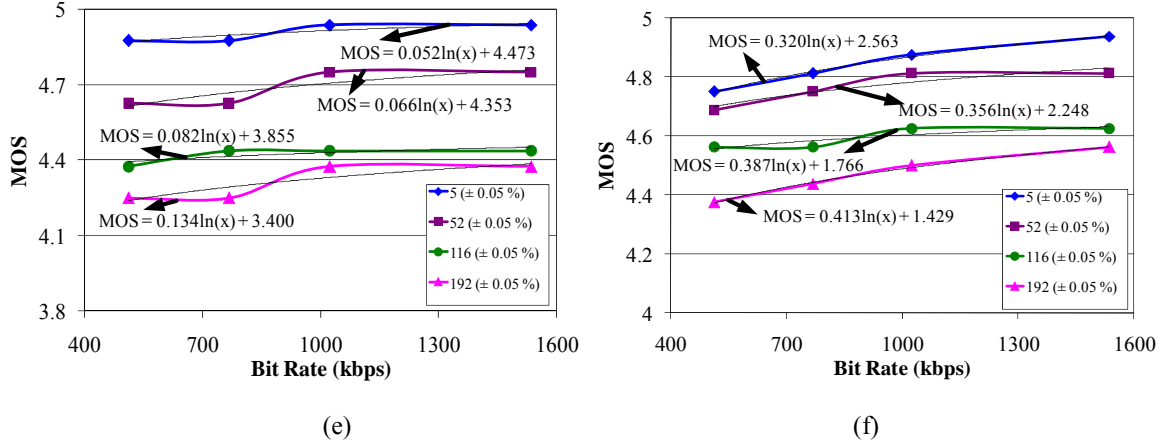


Fig. 2. The (a) *Interview* (b) *Chess* (c) *Windmill* (d) *Ice* (e) *Advertisement* (f) *Eagle* sequence bit rate versus MOS under different ambient illumination conditions (5, 52, 116, and 192 lux) based on depth perception assessments

4. USER PERCEPTION MODEL

In this section, the video quality and depth perception models developed based on the video quality and depth perception results of the subjective experiments are discussed in the following sub-section.

4.1. Video Quality Perception Model

In order to model video quality perception of users towards 3D video contents viewed under different ambient illumination conditions, the mathematical function of every curve in Fig. 1 was determined, as also shown in the figure. It has been observed from the mathematical functions that they all present the following pattern:

$$MOS = P \ln(B) + R \quad (4)$$

where, P and R are two constants and B is the bit rate. We proposed to devise generic functions for P and R using content related contexts that can affect their values and ambient illumination context. In this way, it is possible to derive a video quality perception model for predicting the 3D video quality perception for a given video sequence encoded at B bit rate and consumed under a known ambient illumination condition.

The ‘‘Experimental values’’ column of Table I presents the values of P and R that approximate the experimental results shown in Fig. 1. As seen from the table, when the ambient illumination changes the values of the constants also differ. For example, the P values for the *Interview* sequence are 0.203, 0.194, 0.175, and 0.166 when it is viewed under 5, 52, 116, and 192 lux ambient illumination conditions, respectively. Similar observations can also be made

for the R values. Accordingly, it can be argued that one of the contexts that the values of the constants depend on is ambient illumination. As can also be observed from the table, the values of the constants are video content dependant for the same ambient illumination condition. For instance, the *Chess* and *Windmill* sequences have 0.279 and 0.380 values, respectively, for the 5 lux ambient illumination condition. Similar findings can also be observed for the R values. This clearly indicates that P and R are not only ambient illumination dependant but also they depend on content related contexts. The content related contexts affecting the values of the constants (i.e., P and R) will be explained in the following section.

TABLE I
EXPERIMENTAL VALUES OF P AND R AND THEIR PREDICTED VALUES USING THE VIDEO QUALITY PERCEPTION MODEL
FOR THE VIDEO SEQUENCES USED IN THE SUBJECTIVE EXPERIMENTS

Sequence	Ambient Illumination	Experimental		Predicted	
		P	R	P	R
<i>Interview</i>	5	0.203	3.216	0.301	2.566
	52	0.194	3.341	0.262	2.962
	116	0.175	3.600	0.221	3.274
	192	0.166	3.725	0.202	3.502
<i>Chess</i>	5	0.279	2.258	0.302	2.292
	52	0.227	2.786	0.236	2.781
	116	0.226	2.976	0.224	3.041
	192	0.174	3.475	0.213	3.171
<i>Windmill</i>	5	0.380	1.393	0.415	1.116
	52	0.354	1.943	0.404	1.308
	116	0.343	2.171	0.391	1.611
	192	0.331	2.210	0.366	1.798
<i>Ice</i>	5	0.371	1.783	0.484	0.653
	52	0.353	1.868	0.462	0.979
	116	0.305	1.944	0.391	1.482
	192	0.280	2.077	0.356	1.774
<i>Advertisement</i>	5	0.328	1.106	0.192	1.991
	52	0.287	1.928	0.187	2.134
	116	0.223	2.094	0.172	2.261
	192	0.165	2.193	0.165	2.508
<i>Eagle</i>	5	0.414	0.876	0.334	1.923
	52	0.361	1.500	0.306	2.262
	116	0.336	2.214	0.288	2.426
	192	0.314	2.564	0.280	2.517

4.1.1. Content Related Contexts of the Video Quality Perception Model

Observing that the values of the constants alter for video sequences presenting different motion activity characteristics a metric to measure the motion activity of a color texture sequence is proposed. This proposed metric is discussed in the following sub-section.

4.1.2. Motion

The motion activity of a color texture sequence is measured using its motion intensity in this paper. The optical flow algorithm of the pyramidal implementation of the Lucas and Kanade technique [14] is used for motion intensity measurements in the paper. Motion information is not evenly 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 implementation of the Lucas and Kanade technique [14]. 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:

$$\Pi(i) = \sum_{d=1}^{NoP} |MV_d(x_i, y_i)| \quad (5)$$

where, $\Pi(i)$ is the motion intensity of the i^{th} frame of a sequence. d and NoP are a feature point and the number of feature points in the frame, respectively. $MV_d(x_i, y_i)$ is the motion vector of the i^{th} frame at the 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:

$$M = \frac{\sum_{i=1}^{NoF} \Pi(i)}{NoF} \cdot \frac{F}{S} \quad (6)$$

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, M , of each sequence used in the subjective assessments was measured. Resulting M values are shown in Table II. As can be realized from Table I and II, the values of the constants, P and R , are different for the video sequences having different motion values, M_s . However, even though the motion values, M_s , of the *Ice* and *Eagle* sequences are very similar (i.e., 0.219 and 0.207), the values of the constants are different at the same ambient illumination conditions. This clearly shows that motion is not the only content related context that has an effect on the values of the constants, P and R .

It has been envisaged that structural feature is another content related factor that has an influence on determin-

ing the values. The reason why structural feature is also important for perceiving 3D video quality so as 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 [9]. 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 3D video quality perception under different ambient illumination conditions corresponds to the visibility of visual artifacts in those conditions. The following sub-section elaborates how the structural feature is measured in this paper.

4.1.3. 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. Canny edge detection algorithm [16] was used to determine the contours in the frames without suppressing the pixels that are considered as edges by setting them to 1 [17]. 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] [19]. The total value is then normalized using the NoF and S to provide consistency across different video sequences as follows:

$$C = \frac{\sum_{i=1}^{\text{NoF}} \delta(i)}{\text{NoF } S} \quad (7)$$

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 structural feature, C , values of the video sequences are also illustrated in Table II.

TABLE II
THE M AND C VALUES OF THE VIDEO SEQUENCES USED IN THE SUBJECTIVE EXPERIMENTS

quences	M	C
<i>Interview</i>	0.088	0.120
<i>Chess</i>	0.312	0.141
<i>Windmill</i>	0.182	0.129
<i>Ice</i>	0.219	0.009
<i>Advertisement</i>	0.493	0.134
<i>Eagle</i>	0.207	0.083

4.1.4. Generic Functions of the Constants of the Video Quality Perception Model

In order to devise generic functions for P and R using the contextual factors affecting their values, the graphs representing the M vs P , C vs P , I vs P , M vs R , C vs R , and I vs R were plotted, as presented in Fig. 3. Then, curve fitting functions [20] 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 $a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, t, u, v, y, z, \psi, \lambda, \Delta, \Phi, \Psi, \Pi, \theta$,

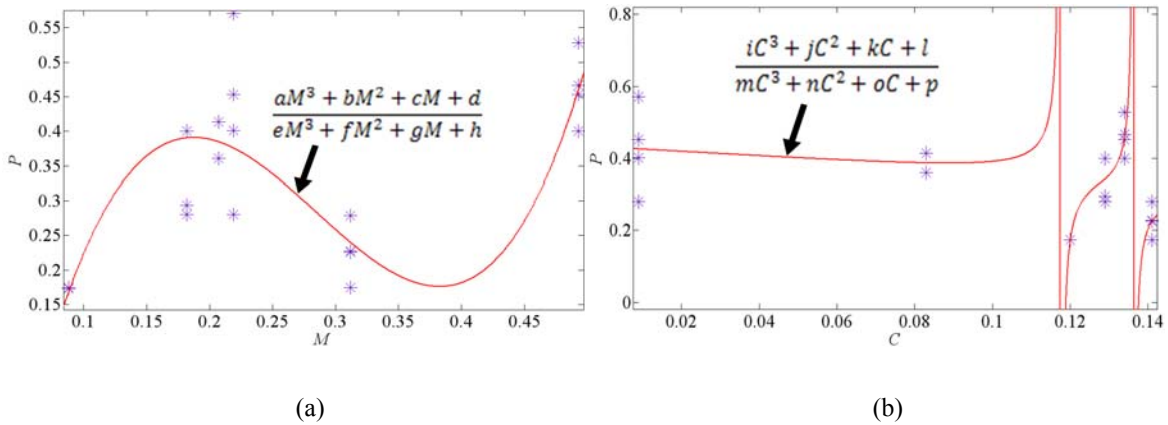
3, α , β , φ , \varkappa , ι , and ω in the figure. As the third step, a set of numerical constants were introduced to each of the functions to calculate the P and R values that best correlate with the P and R values obtained experimentally. Accordingly, the functions of P and R are devised as follows:

$$\begin{aligned}
 P &\propto f(M) = \left(\frac{0.02M^3 + 7M^2 + 0.3M + 2}{425M^3 + 4M^2 + 5M + 0.01} \right) \\
 P &\propto f(C) = \left(\frac{0.01C^3 + 0.25C^2 + 0.02C + 0.1}{0.2C^3 + 4C^2 + 0.3C + 0.01} \right) \\
 P &\propto f(I) = \left(\frac{1410^{-4}}{910^{-3}(\ln I + \frac{1}{I + 0.1})} \right) \\
 R &\propto g(M) = \left(\frac{2M^3 + 0.3M^2 + M + 0.135}{0.3M^4 + 0.03M^3 + 1.2M^2 + 0.5M + 0.35} \right) \\
 R &\propto g(C) = \left(\frac{0.75C^2 + 2C + 1}{4C^2 + C + 2} \right) \\
 R &\propto g(I) = (2.2 \ln I + 1.8)
 \end{aligned} \tag{8}$$

Subsequently, the functions of P and R were integrated together to devise their generic functions as follows:

$$P = (f(M)f(C)f(I)); R = (g(M)g(C)g(I)) \tag{9}$$

Using (9), P and R are predicted and the predicted values are shown in the ‘‘Predicted values’’ column of Table I.



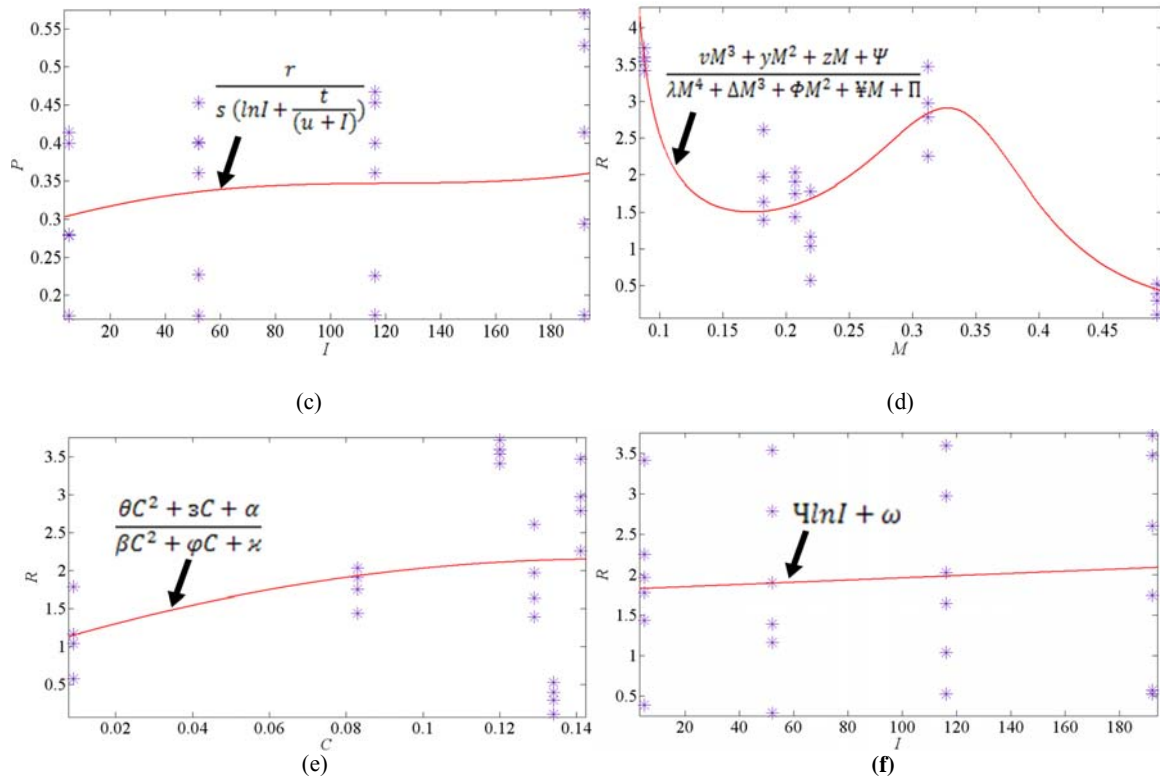


Fig. 3. P versus (a) M (b) C (c) I ; and R versus (d) M (e) C (f) I

4.2. Depth Perception Model

In order to devise the depth perception model, the mathematical functions of the curves in Fig. 2 are determined. It has been observed from these functions that they also present the pattern introduced in (4). Thus, (4) is used to devise depth perception model in this research study, yet with a slight modification. The modification is that V and Y are assigned as constants for depth perception modeling instead of P and R of (4), respectively. The V and Y values of the curves in Fig. 2 are presented in the “Experimental values” column of Table III. As can be observed from the table, one of the contexts affecting the values of the constants is ambient illumination since the values of the constants change for different ambient illumination conditions for the video sequences. As can also be seen from the table, the values of the constants are different even for the similar ambient illumination conditions. Therefore, it is concluded that ambient illumination is not only the context influencing the values of the constants, and thus the content related contexts are similarly noted to have an effect on determining the constant values, as discussed in the following subsection.

TABLE III
EXPERIMENTAL VALUES OF V AND Y AND THEIR PREDICTED VALUES USING THE DEPTH PERCEPTION MODEL
FOR THE VIDEO SEQUENCES USED IN THE SUBJECTIVE EXPERIMENTS

Sequence	<i>Ambient Illumination</i>	Experimental		Predicted	
		V	Y	V	Y
<i>Interview</i>	5	0.032	4.038	0.032	4.054
	52	0.036	3.938	0.035	4.036
	116	0.041	3.780	0.036	4.031
	192	0.048	3.495	0.037	4.027
<i>Chess</i>	5	0.042	4.536	0.055	4.321
	52	0.053	4.338	0.062	4.117
	116	0.059	4.150	0.063	4.049
	192	0.062	4.002	0.064	4.007
<i>Windmill</i>	5	0.120	4.040	0.212	3.251
	52	0.126	3.855	0.238	3.105
	116	0.135	3.665	0.241	3.056
	192	0.141	3.480	0.242	3.026
<i>Ice</i>	5	0.252	2.913	0.297	2.618
	52	0.258	2.688	0.301	2.538
	116	0.261	2.520	0.303	2.512
	192	0.276	2.463	0.304	2.495
<i>Advertisement</i>	5	0.052	4.473	0.196	3.533
	52	0.066	4.353	0.221	3.137
	116	0.082	3.855	0.222	3.005
	192	0.134	3.400	0.223	2.922
<i>Eagle</i>	5	0.320	2.563	0.266	2.972
	52	0.356	2.248	0.299	2.653
	116	0.387	1.766	0.302	2.546
	192	0.413	1.429	0.303	2.280

4.2.1. Content Related Contexts of the Depth Perception Model

Luminance contrast [21] [22] 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 [23]. The metric discussed in the next sub-section is proposed for the luminance contrast measurements in this research work.

4.2.2. 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 [24], and 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. Thus, it is more suitable to measure the luminance contrast rather than using other statistical methods [25] [26]. The *MAD* of a frame of a color texture sequence is measured as follows:

$$MAD(i) = \sum_{k=1}^S |t_k - med(t_i)| \quad (10)$$

where, $MAD(i)$ is the luminance contrast of the i^{th} 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 $MADs$ 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:

$$L = \frac{\sum_{i=1}^{NoF} MAD(i)}{NoF S} \quad (11)$$

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

The luminance contrast, L , values of the sequences utilized in the subjective assessments are also presented in Table IV. As can be observed from Tables III and IV, even though the *Ice* and *Windmill* sequences have similar luminance contrast, L , values (i.e., 35.328 and 32.808), they have different V and Y values. Therefore, it is clear that luminance contrast is not the only depth perception related factor that has an effect on computing the V and Y values.

Depth variation is considered as another depth perception related factor for determining the V and Y 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 grey 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 during presentation [27]. The variation in the pixel depth values in the depth maps corresponds to depth variation in this paper. Depth variation is another 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 variation of a depth map is measured with the proposed metric discussed in the following sub-section.

4.2.3. Depth Variance

Depth variation is measured by applying standard deviation [25] to the pixel depth values in depth map frames in this research study. The reason behind using the standard deviation for the measurement of depth variation 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 [27]. 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:

$$SD(i) = \sqrt{\frac{\sum_{k=1}^S (x_k(i) - \mu(i))^2}{S}} \quad (12)$$

where, $SD(i)$ is the standard deviation of the i^{th} 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 \times 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:

$$D = \frac{\sum_{i=1}^{NoF} SD(i)}{NoF} \quad (13)$$

where, NoF is the number of frames in the depth map sequence which is the same as the number of frames in the corresponding color texture sequence. The depth variances of the video sequences used in the subjective assessments were measured using (13), and the measurement results were also presented in Table IV.

TABLE IV
THE L AND D VALUES OF THE VIDEO SEQUENCES USED IN THE SUBJECTIVE EXPERIMENTS

Sequences	L	D
<i>Interview</i>	29.545	1380.38
<i>Chess</i>	49.009	3916.73
<i>Windmill</i>	32.808	5122.35
<i>Ice</i>	35.328	3385.02
<i>Advertisement</i>	54.237	5044.35
<i>Eagle</i>	69.121	3377.64

4.2.4. Generic Functions of the Constants of the Depth Perception Model

Similar to the video perception modeling, the graphs representing the M vs V , C vs V , I vs V , M vs Y , C vs Y , and I vs Y were plotted, as presented in Fig. 4, to devise generic functions for V and Y using the contextual factors influencing depth perception of users. The curve fitting functions [20] used to approximate the relationships between all of these pairs are also shown in the figure. The constants in the functions are presented with $\epsilon, \vartheta, \beta, \mathfrak{g}, \partial, \mathfrak{h}, \mathfrak{A}, a, \mathfrak{b}, d, \mathfrak{f}, \wp, \mathfrak{h}, i, j, \mathfrak{k}, \Xi, \sigma, \rho, \ell, m, n, q, r, s, t, u, v, w, x, y, \mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{d}, \mathfrak{e}, \mathfrak{f}, \mathfrak{g}, \mathfrak{h},$ and \mathfrak{i} in the figure. Subsequently, a set of numerical constants were introduced to each of the functions to calculate the predicted V and Y values that best correlate with the V and Y values obtained experimentally. In this way, the following functions are devised:

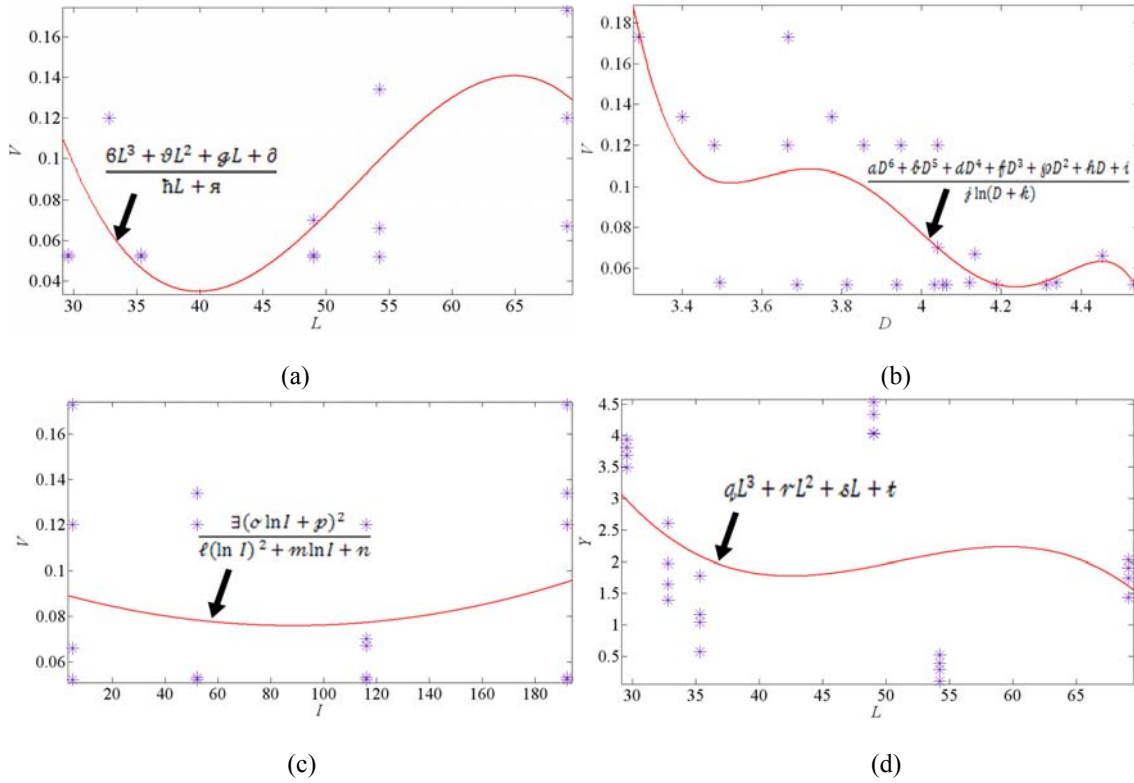
$$\begin{aligned}
V &\propto h(L) = \left(\frac{10^{-8}L^3 + 210^{-4}L^2 + L + 0.1}{0.15L + 0.01} \right) \\
V &\propto h(D) = \left(\frac{710^{-8}D^6 + 2510^{-4}D^5 + 510^{-4}D^4 + 210^{-2}D^3 + 410^{-3}D^2 + D + 0.07}{\ln(D + 0.1)} \right) \\
V &\propto h(I) = \left(\frac{210^{-16}(\ln I + 210^{-6})^2}{6510^{-4}(\ln I)^2 + 10^{-3}\ln I + 2510^{-3}} \right) \\
Y &\propto q(L) = (1210^{-4}L^3 + 0.45L^2 + L + 2) \\
Y &\propto q(D) = \left(\frac{0.2710^{-16}D^5 + 1}{510^{-5}D^4 + D^3 + 2D^2 + D} \right) \\
Y &\propto q(I) = \left(\left(\frac{4510^5}{2e^{\sqrt{0.03\ln(I+0.01)+1}} + 3110^{-4}} \right) - 410^5 \right)
\end{aligned} \tag{14}$$

Similar to the video perception model, the functions of V and Y were integrated together to devise the generic functions as follows:

$$V = (h(L)h(D)h(I)), \quad Y = (q(L)q(D)q(I)) \tag{15}$$

Using (15), V and Y are predicted and the predicted values are presented in the ‘‘Predicted values’’ column of Table

III.



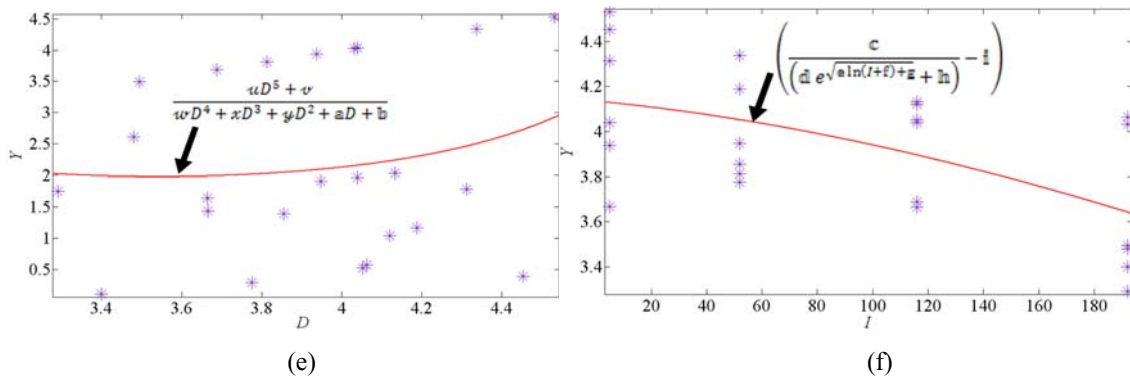


Fig. 4. V versus (a) L (b) D (c) I ; and Y versus (d) L (e) D (f) I

5. RESULTS AND DISCUSSION

The performance assessment results of the developed video quality and depth perception models are discussed in this section for the video sequences used in the subjective experiments (i.e., *Interview*, *Chess*, *Windmill*, *Ice*, *Advertisement*, and *Eagle*) and for two additional video sequences (i.e., *Butterfly* and *Couples*). The additional test sequences were utilized in the experiments to verify the validity of the proposed models on different sequences other than those used during the training and development phases. The M , C , L , and D values of the additional video sequences are presented in Table V. Moreover, their predicted values of P and R are computed using (9) as shown in Table VI.

TABLE V
THE M , C , L , AND D VALUES OF THE *BUTTERFLY* AND *COUPLES* SEQUENCES

Sequence	M	C	L	D
<i>Butterfly</i>	0.117	0.020	2.394	2120.45
<i>Couples</i>	0.110	0.039	1.777	1891.69

TABLE VI
PREDICTED VALUES OF P AND R USING THE VIDEO QUALITY PERCEPTION MODEL
FOR THE *BUTTERFLY* AND *COUPLES* SEQUENCES

Sequence	Ambient Illumination	Predicted	
		P	R
<i>Butterfly</i>	5	0.418	1.743
	52	0.381	2.108
	116	0.351	2.371
	192	0.289	2.812
<i>Couples</i>	5	0.397	1.921
	52	0.371	2.186
	116	0.366	2.261
	192	0.331	2.558

Similar to the *Interview*, *Chess*, *Windmill*, *Ice*, *Advertisement*, and *Eagle* sequences, subjective experiments were conducted for the *Butterfly* and *Couples* sequences using the experimental set-up discussed in Section 3.1. The

MOS results of the video quality related subjective experiments and those predicted using the developed video quality perception model are illustrated for all of the sequences in Table VII. Absolute error percentages (i.e., $|\text{Error}|$ (%)) between the experimental and predicted MOS results for the 3D video sequences are also presented in the table. As can be observed from the results, the absolute error percentages vary for different 3D video sequences. An average absolute error percentage of 2.47% can be achieved using the model for predicting video quality perception of users considering ambient illumination context of the viewing environment.

The results presented in the table also show that the absolute error percentages are different for changing ambient illumination conditions. The average absolute error percentage results (i.e., the average $|\text{Error}|$ (%)) calculated for each ambient illumination condition using the *Interview*, *Chess*, *Windmill*, *Ice*, *Advertisement*, *Eagle*, *Butterfly*, and *Couples* sequences are shown in Table VIII. These results indicate that the average deviation of the predicted MOS score is less than 3.19% when the proposed video quality perception model is used.

To further demonstrate the efficiency of the proposed video quality perception model, the average absolute error percentages dedicated to the additional video sequences (i.e., *Butterfly* and *Couples*) considering each of the ambient illumination condition are presented in a separate table (Table IX). As observed from this table, a maximum of 1.51% average absolute error percentage can be achieved using the video quality perception model.

TABLE VII
THE MOS RESULTS OF THE VIDEO QUALITY RELATED SUBJECTIVE EXPERIMENTS AND VIDEO QUALITY PERCEPTION MODEL

Ambient Illumination (lux)	Bit Rate (kbps)	Seq.	Experimental MOS	Predicted MOS	$ \text{Error} $ (%)	Seq.	Experimental MOS	Predicted MOS	$ \text{Error} $ (%)
5	512	<i>Interview</i>	4.500	4.443	1.27	<i>Ice</i>	3.562	3.672	4.49
	768		4.562	4.565	0.07		3.625	3.868	8.08
	1024		4.625	4.652	0.58		3.687	4.007	10.04
	1536		4.687	4.774	1.86		3.875	4.203	9.75
52	512		4.625	4.596	0.63		3.687	3.861	4.72
	768		4.687	4.702	0.32		3.812	4.048	6.19
	1024		4.750	4.777	0.57		3.937	4.181	6.20
	1536		4.812	4.884	1.50		4.125	4.368	5.89
116	512		4.687	4.658	0.62		3.875	3.921	1.19
	768		4.750	4.749	0.02		4.062	4.079	0.42
	1024		4.812	4.805	0.14		4.187	4.192	0.12
	1536		4.875	4.902	0.55		4.375	4.370	0.11
192	512		4.812	4.762	1.04		4.125	4.044	1.96
	768		4.875	4.843	0.66		4.387	4.188	4.54
	1024		4.937	4.902	0.71		4.562	4.291	5.94
	1536		5.000	4.983	0.34		4.750	4.335	8.74

5	512	Chess	4.000	4.175	4.38	Advertisement	2.875	3.158	9.84
	768		4.125	4.298	4.19		3.062	3.246	6.01
	1024		4.187	4.385	4.73		3.187	3.301	3.58
	1536		4.312	4.507	4.52		3.312	3.379	2.02
52	512		4.187	4.253	1.58		3.125	3.300	5.60
	768		4.312	4.348	0.83		3.312	3.376	1.93
	1024		4.375	4.416	0.94		3.437	3.430	0.20
	1536		4.437	4.512	1.69		3.625	3.505	3.31
116	512		4.375	4.438	1.44		3.437	3.383	1.57
	768		4.500	4.529	0.64		3.625	3.453	4.74
	1024		4.562	4.593	0.68		3.812	3.503	8.11
	1536		4.625	4.684	1.28		3.937	3.572	9.27
192	512		4.562	4.499	1.38		3.812	3.587	5.90
	768		4.625	4.585	0.86		4.062	3.654	10.04
	1024		4.687	4.647	0.85		4.250	3.701	12.92
	1536		4.750	4.733	0.36		4.250	3.768	11.34
5	512	Windmill	3.687	3.734	1.27	Eagle	4.000	4.006	0.15
	768		3.875	3.902	0.70		4.187	4.141	1.10
	1024		3.937	4.022	2.16		4.375	4.237	3.15
	1536		4.000	4.190	4.75		4.437	4.373	1.44
52	512		3.875	3.838	0.95		4.187	4.171	0.38
	768		4.062	4.001	1.50		4.250	4.294	1.04
	1024		4.187	4.118	1.65		4.437	4.382	1.24
	1536		4.312	4.281	0.72		4.562	4.506	1.23
116	512		4.125	4.160	0.85		4.312	4.222	2.09
	768		4.312	4.318	0.14		4.375	4.339	0.82
	1024		4.437	4.431	0.14		4.562	4.422	3.07
	1536		4.562	4.589	0.59		4.687	4.538	3.18
192	512		4.437	4.290	3.31		4.375	4.263	2.56
	768		4.562	4.438	2.72		4.437	4.377	1.35
	1024		4.687	4.543	3.07		4.625	4.457	3.63
	1536		4.750	4.692	1.22		4.812	4.571	5.01
5	512		4.375	4.350	0.57		4.437	4.397	0.90
	768		4.562	4.519	0.94		4.500	4.558	1.29
	1024		4.625	4.640	0.32		4.562	4.672	2.41
	1536		4.687	4.809	2.60		4.687	4.833	3.11
52	512	4.437	4.484	1.06	4.562	4.501	1.34		
	768	4.625	4.638	0.28	4.625	4.651	0.56		
	1024	4.687	4.748	1.30	4.687	4.757	1.49		
	1536	4.750	4.903	3.22	4.812	4.907	1.97		
116	512	4.500	4.561	1.36	4.625	4.544	1.75		
	768	4.687	4.702	0.32	4.750	4.692	1.22		
	1024	4.750	4.803	1.12	4.812	4.797	0.31		
	1536	4.812	4.945	2.76	4.937	4.945	0.16		
192	512	4.583	4.614	0.68	4.812	4.623	3.93		
	768	4.750	4.731	0.40	4.875	4.757	2.42		
	1024	4.812	4.815	0.06	4.937	4.852	1.72		
	1536	4.875	4.932	1.17	4.937	4.986	0.99		
5	512	Butterfly	4.375	4.350	0.57	Couples	4.437	4.397	0.90
	768		4.562	4.519	0.94		4.500	4.558	1.29
	1024		4.625	4.640	0.32		4.562	4.672	2.41
	1536		4.687	4.809	2.60		4.687	4.833	3.11
52	512		4.437	4.484	1.06		4.562	4.501	1.34
	768		4.625	4.638	0.28		4.625	4.651	0.56
	1024		4.687	4.748	1.30		4.687	4.757	1.49
	1536		4.750	4.903	3.22		4.812	4.907	1.97
116	512		4.500	4.561	1.36		4.625	4.544	1.75
	768		4.687	4.702	0.32		4.750	4.692	1.22
	1024		4.750	4.803	1.12		4.812	4.797	0.31
	1536		4.812	4.945	2.76		4.937	4.945	0.16
192	512		4.583	4.614	0.68		4.812	4.623	3.93
	768		4.750	4.731	0.40		4.875	4.757	2.42
	1024		4.812	4.815	0.06		4.937	4.852	1.72
	1536		4.875	4.932	1.17		4.937	4.986	0.99

TABLE VIII
THE AVERAGE |ERROR| % FOR THE AMBIENT ILLUMINATION CONDITIONS CONSIDERING THE VIDEO QUALITY PERCEPTION MODEL FOR THE *INTERVIEW*, *CHESS*, *WINDMILL*, *ICE*, *ADVERTISEMENT*, *EAGLE*, *BUTTERFLY*, AND *COUPLES* SEQUENCES

Ambient Illumination (lux)	Average Error (%)
5	3.19
52	1.94
116	1.58
192	3.18

TABLE IX
THE AVERAGE |ERROR| % FOR THE AMBIENT ILLUMINATION CONDITIONS CONSIDERING THE VIDEO QUALITY PERCEPTION MODEL FOR THE *BUTTERFLY* AND *COUPLES* SEQUENCES

Ambient Illumination (lux)	Average Error (%)
5	1.51
52	1.40
116	1.12
192	1.42

In order to show the performance assessment results of the depth perception model for the *Butterfly* and *Couples* sequences, the predicted values of V and Y are calculated for these sequences using (15), and the calculated values are presented in Table X.

TABLE X
PREDICTED VALUES OF V AND Y USING THE DEPTH PERCEPTION MODEL FOR THE *BUTTERFLY* AND *COUPLES* SEQUENCES

Sequence	Ambient Illumination	Predicted	
		V	Y
<i>Butterfly</i>	5	0.264	2.977
	52	0.271	2.783
	116	0.273	2.782
	192	0.274	2.781
<i>Couples</i>	5	0.151	3.289
	52	0.16	3.168
	116	0.162	2.956
	192	0.163	2.735

The MOS results of the depth perception related subjective experiments and those predicted by the developed depth perception model, as well as the absolute error percentages between the experimental and predicted MOS results are shown in Table XI for the *Interview*, *Chess*, *Windmill*, *Ice*, *Advertisement*, *Eagle*, *Butterfly*, and *Couples* sequences. As can be observed from these results, similar to the video quality perception related findings, the absolute error percentages change for each of the 3D video sequences in the table. An average absolute error percentage of 2.82% can be attained using the depth perception model.

TABLE XI
THE MOS RESULTS OF THE DEPTH RELATED SUBJECTIVE EXPERIMENTS AND DEPTH PERCEPTION MODEL

Ambient Illumination (lux)	Bit Rate (kbps)	Seq.	Experimental MOS	Predicted MOS	Error (%)	Seq.	Experimental MOS	Predicted MOS	Error (%)		
5	512	Interview	4.250	4.253	1.07	Ice	4.625	4.471	3.33		
	768		4.312	4.266	0.86		4.687	4.591	2.05		
	1024		4.312	4.275	0.56		4.687	4.676	0.23		
	1536		4.312	4.288	3.13		4.687	4.796	2.33		
52	512		4.125	4.254	1.93		4.500	4.415	1.89		
	768		4.187	4.268	2.17		4.562	4.537	0.55		
	1024		4.187	4.278	2.51		4.562	4.624	1.36		
	1536		4.187	4.292	6.38		4.562	4.746	4.03		
116	512		4.000	4.255	5.12		4.437	4.402	0.79		
	768		4.062	4.270	5.39		4.500	4.524	0.53		
	1024		4.062	4.281	5.74		4.500	4.612	2.49		
	1536		4.062	4.295	11.67		4.500	4.734	5.20		
192	512		3.812	4.257	10.25		4.375	4.391	0.37		
	768		3.875	4.272	10.53		4.437	4.514	1.74		
	1024		3.875	4.283	10.92		4.437	4.602	3.72		
	1536		3.875	4.298	6.36		4.437	4.725	6.49		
5	512	Chess	4.875	4.565	3.88	Advertisement	4.875	4.755	2.46		
	768		4.875	4.686	3.55		4.875	4.835	0.82		
	1024		4.875	4.702	4.31		4.937	4.891	0.93		
	1536		4.937	4.724	3.93		4.937	4.907	0.61		
52	512		4.687	4.503	3.39		4.625	4.515	2.38		
	768		4.687	4.528	3.01		4.625	4.605	0.43		
	1024		4.687	4.546	3.77		4.750	4.668	1.73		
	1536		4.750	4.571	1.29		4.750	4.758	0.17		
116	512		4.500	4.442	0.73		4.375	4.389	0.32		
	768		4.500	4.467	0.33		4.437	4.479	0.95		
	1024		4.500	4.485	1.57		4.437	4.543	2.39		
	1536		4.583	4.511	0.25		4.437	4.633	4.42		
192	512		4.375	4.364	0.25		4.250	4.313	1.48		
	768		4.375	4.386	0.62		4.250	4.403	3.60		
	1024		4.375	4.402	0.29		4.375	4.467	2.10		
	1536		4.437	4.424	4.97		4.375	4.557	4.16		
5	512		Windmill	4.812	4.573		3.18	Eagle	4.750	4.631	2.51
	768			4.812	4.659		3.18		4.812	4.739	1.52
	1024			4.875	4.720		2.65		4.875	4.815	1.23
	1536			4.937	4.806		0.78		4.937	4.873	1.30
52	512	4.625		4.589	1.32	4.687	4.518		3.61		
	768	4.625		4.686	1.43	4.750	4.639		2.34		
	1024	4.687		4.754	2.13	4.812	4.725		1.81		
	1536	4.750		4.851	2.75	4.812	4.846		0.71		
116	512	4.437		4.559	4.94	4.562	4.429		2.92		
	768	4.437		4.656	5.02	4.562	4.552		0.22		
	1024	4.500		4.726	5.72	4.625	4.639		0.30		
	1536	4.562		4.823	6.71	4.625	4.761		2.94		
192	512	4.250		4.535	9.01	4.375	4.170		4.69		
	768	4.250		4.633	9.07	4.437	4.292		3.27		
	1024	4.312		4.703	9.74	4.500	4.380		2.67		
	1536	4.375		4.801	5.17	4.562	4.503		1.29		

5	512	<i>Butterfly</i>	4.875	4.623	2.95	<i>Couples</i>	4.250	4.231	0.45
	768		4.875	4.731	2.63		4.312	4.292	0.46
	1024		4.937	4.807	0.49		4.375	4.335	0.91
	1536		4.937	4.913	6.04		4.500	4.397	2.29
52	512		4.750	4.463	3.73		4.062	4.166	2.56
	768		4.750	4.573	3.35		4.125	4.231	2.57
	1024		4.812	4.651	1.06		4.187	4.276	2.13
	1536		4.812	4.761	2.75		4.312	4.342	0.70
116	512		4.612	4.485	0.35		3.875	3.966	2.35
	768		4.612	4.596	0.28		3.937	4.032	2.41
	1024		4.687	4.674	2.07		4.000	4.078	1.95
	1536		4.687	4.784	1.19		4.125	4.144	0.46
192	512		4.437	4.490	3.70		3.625	3.752	3.50
	768		4.437	4.601	4.00		3.750	3.818	1.81
	1024		4.500	4.680	6.47		3.812	3.865	1.39
	1536		4.500	4.791	1.07		3.937	3.931	0.15

The results of the table present that the absolute error percentages are also different for each of the varying ambient illumination conditions. The average absolute error percentage results computed for each ambient illumination condition are shown in Table XII. The results in this table show that the average deviation of the predicted MOS score is less than 4.21% when the proposed depth perception model is used.

Similar to the video quality perception model, absolute average error percentages are calculated for the *Butterfly* and *Couples* sequences (i.e., the additional video sequences) to emphasize the efficiency of the depth perception model as shown in Table XIII. As can be noted from the table, a maximum of 2.76% absolute error percentage can be attained by utilizing the depth perception model.

TABLE XII
THE AVERAGE $|\text{ERROR}| \%$ FOR THE AMBIENT ILLUMINATION CONDITIONS CONSIDERING THE DEPTH PERCEPTION MODEL FOR THE *INTERVIEW*, *CHESS*, *WINDMILL*, *ICE*, *ADVERTISEMENT*, *EAGLE*, *BUTTERFLY*, AND *COUPLES* SEQUENCES

Ambient Illumination (lux)	Average Error (%)
5	2.08
52	2.25
116	2.74
192	4.21

TABLE XIII
THE AVERAGE $|\text{ERROR}|$ % FOR THE AMBIENT ILLUMINATION CONDITIONS CONSIDERING
THE DEPTH PERCEPTION MODEL FOR THE *BUTTERFLY* AND *COUPLES* SEQUENCES

Ambient Illumination (lux)	Average $ \text{Error} $ (%)
5	2.03
52	2.36
116	1.38
192	2.76

6. CONCLUSION

Video quality and depth perception of users towards 3D video viewed under different ambient illumination conditions are modeled in this paper. Motion and structural feature of color texture sequence of 3D video are used as content related contexts while modeling video quality perception of users. Luminance contrast of color texture and depth variation of depth map sequences act as primary content related contexts in the depth perception model. The results demonstrate that both the video quality and depth perception models can be efficiently utilized to predict the user perception of video quality and depth of 3D video. The video quality and depth perception models proposed in this research study can accelerate the advancement of user centric 3D video related technologies into the 3D consumer market, which in turn will ensure enhanced video quality and depth perception of users, particularly in user centric media access and consumption scenarios.

REFERENCES

- [1] C.T.E.R. Hewage, S.T. Worrall, S. Dogan, S. Villette, and A. M. Kondo, "Quality Evaluation of Color Plus Depth Map Based Stereoscopic Video," *IEEE Journal of Selected Topics in Signal Processing: Visual Media Quality Assessment*, vol. 3, no. 2, pp. 304-318, Apr. 2009.
- [2] G. Nur, S. Dogan, H. Kodikara Arachchi, and A. M. Kondo, "Assessing the Effects of Ambient Illumination Change in Usage Environment on 3D Video Perception for User Centric Media Access and Consumption," *2nd International ICST Conference on User Centric Media*, Palma de Mallorca, Spain, 1-3 Sep. 2010.
- [3] B. Girod, "What's Wrong with Mean-Squared Error," *Digital Images and Human Vision*, A. B. Watson Ed., Chapter 15, pp. 207-220, the MIT press, 1993.
- [4] M. Ghanbari, "Standard Codecs: Image Compression to Advanced Video Coding," The Institution of Electrical Engineers, London, UK, 2003.
- [5] Q. Huynh-Thu and M. Ghanbari, "Scope of Validity of PSNR in Image/Video Quality Assessment," *IET Electronics Letters*, vol. 44, no. 13, pp. 800-801, Jun. 2008.
- [6] H. Pinson and S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Transactions on Broadcasting*, vol. 50, no.3, pp. 312-322, Sep. 2004.
- [7] S. Wolf and M. Pinson, "VQM Software and Measurement Techniques," *National Telecommunications and Information Administration Report 02-392*, Jun. 2002.
- [8] International Telecommunication Union (ITU) Radio Communication Sector: 'Methodology for the Subjective Assessment of the Quality of Television Pictures', ITU-R BT.500-11, 2002.
- [9] Z. Wang, L. Lu, and A. C. Bovik, "Video Quality Assessment based on Structural Distortion Measurement," *Proc. of Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121-132, Feb. 2004.
- [10] JSVM 9.13.1, CVS Server, garcon.ient.rwth-aachen.de/cvs/jv

- [11] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller, "Quality Assessment of 3D Video in Rate Allocation Experiments," *IEEE Symposium on Consumer Electronics*, Apr. 2008.
- [12] Gretag Macbeth Eye-One Display 2, <http://www.xrite.com>
- [13] T. R. Robinson, "Light Intensity and Depth Perception," *American Journal on Psychology*, vol. 7, no. 4, pp. 518-532, Jul. 1896.
- [14] D. J. Fleet, and Y. Weiss, "Optical Flow Estimation in Paragios," *Handbook of Math. Models in Comp. Vis.*, Springer, 2006.
- [15] J. Shi and C. Tomasi, "Good Features to Track," *IEEE Conf. on Com. Vis. and Pat. Recog.*. Seattle, USA, Jun. 1994.
- [16] C. Grigorescu, N. Petkov, and M. A. Westenberg, "Contour and Boundary Detection Improved by Surround Suppression of Texture Edges," *Image and Vision Computing*, vol. 22, no. 8, pp. 609-622, 2004.
- [17] G. Papari, P. Campisi, N. Petkov, and A. Neri, "A Multiscale Approach to Contour Detection by Texture Suppression," *SPIE Im.Proc.: Alg. and Syst.*, vol. 6064A, pp. 107-118, San Jose, CA, USA, 16-18 Jan. 2006.
- [18] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and Texture Analysis for Image Segmentation," *Int. Journ. Comput. Vision*, vol. 1, no. 43, pp. 7-27, Jun. 2001.
- [19] J. F. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, Jun. 1986.
- [20] <http://www.mathworks.com/access/helpdesk/help/toolbox/curvefit/>
- [21] W. S. Geisler, "Visual Perception and the Statistical Properties of Natural Scenes," *Annual Review of Psychology*, no. 59, vol. 26, pp.167-192, Sep. 2008.
- [22] R. A. Frazor and W. S. Geisler, "Local Luminance and Contrast in Natural Images," *Elsevier Vision Research Journ.*, no. 46, vol. 59, pp. 1585-1598, Jan. 2006.
- [23] S. Ichihara, N. Kitagawa, and H. Akutsu, "Contrast and Depth perception: Effects of Texture Contrast and Area Contrast," *Perception*, vol. 36, no. 5, pp. 686-695, May. 2007.
- [24] V. Jones, "Mean Direction and Mean Absolute Deviation," *ASTM Standards and Engineering Digital Library*, Jan. 2009.
- [25] J. L. Devore, "Probability and Statistics for Engineering and the Sciences", Duxbury, 1995.
- [26] P. Hall and M. P. Wand, "On the Minimization of Absolute Distance in Kernel Density Estimation," *Statistics and Probability Letters*, no 6, pp. 311-314, Apr. 1988.
- [27] G. Nur, S. Dogan, H. Kodikara Arachchi, and A. M. Kondoz, "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 Jun. 2010.