Quality Evaluation of Colour plus Depth Map Based Stereoscopic Video

Chaminda T. E. R. Hewage, *Student Member, IEEE*, Stewart T. Worrall, *Member, IEEE*, Safak Dogan,
Member, IEEE, Stephane Villette, *Member, IEEE*, and Ahmet M. Kondo, *Member, IEEE*

Abstract—In the near future, many conventional video applications are likely to be replaced by immersive video to provide a sense of “being there”. This transition is facilitated by the recent advancement of 3-D (3-Dimensional) capture, coding, transmission, and display technologies. Stereoscopic video is the simplest form of 3-D video available in the literature. “Colour plus depth map” based stereoscopic video has attracted significant attention, as it can reduce storage and bandwidth requirements for the transmission of stereoscopic content over communication channels. However, quality assessment of coded video sequences can currently only be performed reliably using expensive and inconvenient subjective tests. To enable researchers to optimize 3-D video systems in a timely fashion, it is essential that reliable objective measures are found. This paper investigates the correlation between subjective and objective evaluation of colour plus depth video. The investigation is conducted for different compression ratios, and different video sequences. Transmission over IP (Internet Protocol) is also investigated. Subjective tests are performed to determine the image quality and depth perception of a range of differently coded video sequences, with packet loss rates ranging from 0% to 20%. The subjective results are used to determine more accurate objective quality assessment metrics for 3-D colour plus depth video.

Index Terms—Stereoscopic video, Scalable video coding, Depth-Image-Based-Rendering, Objective quality metric

I. INTRODUCTION

STEREOSCOPIC video has the potential to provide high quality immersive experiences for a variety of applications in the home, in the workplace, and in public spaces. Recent advances, in acquisition technologies and displays, mean that the provision of services such as 3DTV is now technically feasible. However, it is currently difficult to evaluate the quality of

Manuscript received April 30, 2008. This work was developed within VISNET II, a European Network of Excellence-NoE (<http://www.visnet-noe.org>), funded under the European Commission IST FP6 programme.

Chaminda T.E.R. Hewage, Stewart T. Worrall, Safak Dogan, Stephane Villette and Ahmet M. Kondo are with I-LAB, Centre for Communication Systems Research, Faculty of Engineering and Physics, University of Surrey, Guildford, GU2 7XH, UK. (phone: +44-(0)1483-684723; fax: +44-(0)1483-686011; e-mail: e.thushara@surrey.ac.uk).

such applications without resorting to full subjective test campaigns. This is because very little work has been carried out to find objective measurement techniques for 3-D video quality.

The existing 2D objective quality measures of individual colour and depth map sequences may not represent the true image quality as perceived by the human viewers. Although some researchers currently use Peak Signal to Noise Ratio (PSNR) for assessing 3-D video, the limitations of PSNR for 2-D (2-Dimensional) video have been demonstrated in tests by the Video Quality Experts Group (VQEG) [1]. These limitations are likely to be similar for the assessment of image quality in 3-D video. In addition, PSNR does not give any information about depth perception. Therefore, explorative studies based on appreciation oriented methods are required to obtain a true understanding of 3-D perception. The multidimensional attributes (e.g., blocking and noise) associated with 2-D video can not be used in measuring the perceptually important qualities of 3-D video, such as overall image quality, sharpness, naturalness and depth perception. Therefore, 3-D quality models which account for depth reproduction in immersive video are required.

In this paper subjective testing using human viewers is used to measure the effect of compression artefacts and frame losses on the perceived quality of the reconstructed stereoscopic video. The perceptual attributes namely overall image quality and depth perception are used to measure the response from human observers. The relationship between the perceived overall image quality and the perceived depth are also discussed in this paper. Finally, the perceived attributes of reconstructed stereoscopic video are predicted using different objective quality models. The results help to answer important questions concerning the correlation between objective and subjective measurements of 3-D video quality, and also concerning the best methods for measuring the quality of 3-D video objectively, for error free and error prone channels.

The investigations focus on the colour plus depth representation of stereoscopic video, which been widely used in research and standardization activities [2]-[4] due to its flexible representation and compatibility with existing coding and transmission technologies. This format can be more efficiently compressed, compared to the left and right view format for stereoscopic video [5].

This paper is split into five sections. In the background section (Section II), previous research on symmetric/asymmetric coding, and on measuring the quality of colour plus depth 3-D video is discussed. Furthermore, the stereoscopic video coding architecture used for encoding colour and depth video in this paper is described in Section II. Section III presents the methods of carrying out subjective tests for both symmetric/asymmetric coding configurations and different packet loss scenarios. The subjective test procedure and results are analyzed in section IV. Section V concludes the paper.

II. BACKGROUND

This section gives an overview of the coding methods used in the quality assessments, and on available methods for

assessment of 2-D and 3-D video. It begins with a discussion of the colour plus depth representation of stereoscopic video, and how it can be compressed using existing video codecs. Next, previous research on symmetric and asymmetric coding of stereoscopic video sequences is discussed as a method of reducing the storage and bandwidth requirements of the applications. Finally, the subjective and objective quality measures which can be used in quantifying the perception of 3-D video are also described.

A. Colour plus Depth 3-D Video Compression

Depth images may be generated using special range cameras, or by processing the output from multiple camera views of a scene [6]-[8]. Depth maps typically have the same spatial resolution as colour video (see Fig. 1), and only contain a single plane of image data. The per-pixel depth value determines the position of the associated colour texture in the 3-D space. The image warping technique known as Depth-Image-Based Rendering (DIBR) is used to render the desired binocular viewpoint image [9]. The reconstructed quality of stereoscopic video using this technique is heavily dependent on the availability of accurate depth maps for rendering novel views [10].

Fig. 1. The *Interview* sequence: (a) Colour image, (b) Per-pixel depth image. The depth maps are normalized to a near clipping plane Z_{near} and a far clipping plane Z_{far}

The storage and bandwidth requirements of stereoscopic video communication applications are higher compared to monoscopic video applications. Hence, efficient compression techniques are required for medium bitrate communication channels. A survey of 3-D video coding algorithms for different 3-D representations is presented in [11]. The traditional 2-D compression techniques could be used to encode both colour and depth map sequences simultaneously [12]. Fehn *et al* show in [13] that the depth maps obtained using 3-D/depth-range cameras can be efficiently compressed using the MPEG-4 and H.264/AVC video coding standards. Furthermore, a comparison of colour and depth map encoding architectures based on MPEG-4 MAC (Multiple Auxiliary Component), H.264/AVC and the scalable extension of H.264/AVC is discussed in [14].

The layered coding architecture of scalable video coding methods can be effectively used in encoding stereoscopic video. In layered coding approaches, the colour and depth/disparity image sequences are coded as the base and enhancement layers

respectively (see Fig. 2). The backward compatibility of this layered architecture allows the enhancement of existing 2-D video applications with stereoscopic video applications. Furthermore, the scalability of this configuration can be used to exploit the asymmetric coding of stereoscopic video content (e.g., low spatial/quality/temporal resolution colour/depth image sequences) to overcome the storage and bandwidth limitations.

Fig. 2. The Stereoscopic video coding architecture based on SVC

In this paper, the layered coding configuration, based on the scalable extension of H.264/AVC is chosen to encode colour and depth map video [15]. As the scalable extension of H.264/AVC incorporates superior H.264/AVC video coding algorithms [16], this configuration has high potential to achieve good stereo quality at reduced bitrates. Furthermore, it is shown in [13] that H.264/AVC outperforms other MPEG video coding standards that are currently being considered for depth map coding. This will be further supported by the range of scalability options such as spatial, temporal, and quality scalabilities available with the scalable extension of H.264/AVC. A performance analysis of a backward compatible SVC encoding configuration for ‘colour plus depth’ based SSV is discussed in [14].

A further issue for 3-D video quality assessment is the effects from transmission over networks. The transmission of high resolution 3-D video using broadcast methods such as DVB (Digital Video Broadcasting) and IP based streaming techniques has been discussed in [17], [18]. Stereoscopic video streaming over networks (e.g., IP core network) may lead to degraded quality due to packet losses. These random packet losses may result in losing the entire frame at the receiver/intermediate network node. The effect of frame losses on the perceived quality of reconstructed stereoscopic video may be worse compared to that of conventional video (i.e., 2-D video) applications. Hence, this paper explores the image quality of reconstructed binocular video over IP core network using objective and subjective quality measures. In addition to that several objective quality models are used to predict the perceived image quality and depth perception of stereoscopic video rendered over IP.

B. Symmetric and Asymmetric Coding of Stereoscopic Video

Mixed-resolution coding of stereoscopic video is based on the theories of binocular suppression. This concept is widely used for left and right view based binocular video [19], [20]. Mixed-resolution coding assumes that the high quality view (either left or right) of a stereo pair will dominate the perceptual quality of the reconstructed 3-D video. This method has been used to reduce the storage and bandwidth requirements for stereo video applications [21], [22]. The symmetric/asymmetric coding of colour and depth map based stereoscopic video is also can be used in reducing the storage and bandwidth requirements. But this has not been well exploited in current research.

The spatial down-sampling of depth map sequences for low bitrate video applications is discussed in [23]. This study shows that the spatial down-sampling prior to H.264/AVC encoding achieves higher image qualities for both colour and depth map video at low bitrates compared to full resolution colour and depth video sequences at the same bitrate. The bitrate savings for stereoscopic TV applications using coarsely quantized depth maps are studied in [24]. In the above mentioned studies, subjective evaluations using human observers have not been performed to obtain the perceptual image quality ratings. The ROI and metadata based unequal bitrate allocation schemes for depth maps are discussed in [10] and [25] respectively. Furthermore, Ekmekcioglu *et al* proposes a bitrate adaptive down-sampling method for coding multi-view with depth information [26].

In this paper, the effect of symmetric/asymmetric coding on the perception of reconstructed 3-D video at reduced bitrates will be investigated. Symmetric and asymmetric coding differ in the way quality levels are assigned during coding. In symmetric coding, both image sequences are coded into the same spatial/temporal/quality levels. Whereas in asymmetric coding one image sequence (i.e., either colour or depth map video) is coded into different quality levels compared to the other image sequence (e.g., one image sequence is coarsely quantized compared to the other sequence). The compression artifacts of both colour and depth maps will affect the overall image quality and depth perception. Hence, this study enables us to find the perceptual bounds of symmetric/asymmetric coding combinations of colour and depth video.

C. 3-D Quality Measurements

The effect of different artifacts introduced by image capture, coding and delivery methods on the perceived quality of video are diverse in nature. The objective measures of rendered left and right views may not account for the perceptual factors such as depth reproduction, stereoscopic impairments, and visual comfort. Therefore, subjective evaluations which incorporate human observers need to be carried out to get a true representation of human 3-D perception. Subjective evaluation methods are widely used in assessing the perceptual attributes of conventional 2-D video and stereoscopic video [27], [28]. The appreciation oriented psychophysical scaling paradigms can be used in assessing emerging 3-D technologies such as stereoscopic TV and stereoscopic video streaming [29]. ITU-Recommendation BT.1438 [30] describes the methodology of carrying out appreciation

oriented subjective evaluation tests for stereoscopic television applications. In the explorative study described in this paper, subjective tests are performed in order to obtain human feedback on the effect of different combinations of symmetric/asymmetric coding levels and different packet loss rates (PLRs). The feedback provides information on the perceived quality of colour and depth map based stereoscopic video. The perceived quality of the reconstructed binocular video is measured using two image attributes, namely perceived overall image quality and depth.

The image quality models define the relationship between the physical parameter (e.g., coding and delivery method) of the system and the perceived quality of 3-D video. The objective quality metrics which incorporate perceptual attributes for conventional 2-D video are well exploited in research [27], [31], [32]. However, the development of quality models for 3-D video may be difficult as there are a number of perceptual attributes (e.g., naturalness, presence, stereo impairments, comfort, perceived depth, *etc.*) associated with 3-D perception. Seuntjens *et al* describes in [33] that JPEG coding of stereoscopic video has an effect on overall image quality, but no effect on perceived depth. Similarly in [34] the added value of depth is not taken into account when assessing the perceived image quality of MPEG-2 coded stereo sequences. However, a positive relationship between depth and perceived image quality for uncompressed stereoscopic images is discussed in [35]. An objective quality metric for free-viewpoint video production in a studio environment is described in [36].

In this paper, the correlation between different attributes of 3-D perception (i.e., between the depth perception and overall image quality) is derived for different symmetric/asymmetric coding combinations and different PLRs. Furthermore, different objective quality measures are used in this paper to predict the perceived attributes of stereoscopic video. Therefore, the subjective image quality ratings are mapped against the candidate objective measures such as PSNR and SSIM (structural similarity) using a logistic function.

Objective measures, such as Mean Average Difference (MAD) and PSNR, are widely used in measuring 2-D video quality. But the use of 2-D quality metrics in measuring stereoscopic video quality may not be straightforward, because the perceptual attributes (e.g., perceived depth) of 3-D are different compared to conventional 2D-video. However, the objective measures of individual left and right views (i.e., generated using a stereo camera pair) could be used as a candidate measure of perceived stereoscopic video quality. In the case of colour and depth map video it may be accurate enough to take individual objective measures of colour and depth map video as a measure of perceived stereoscopic video quality as depth map together with its texture video renders two left and right views at the display. On the other hand, the objective measures of the rendered left and right views using the DIBR technique can be used as a candidate measure of 3-D scene perception (i.e., the rendered left and right views using the coded colour and depth map video will be compared against the synthesized left and right video using original colour and depth map sequences.). Therefore, in this experiment the objective measures of colour, depth, and rendered

left and right sequences are used to predict the perceived image quality and depth perception of stereoscopic video.

In this paper, in addition to the PSNR quality metric, two other objective measures, namely SSIM and Video Quality Metric (VQM) are used. The SSIM is based on the fact that the human visual system is capable of extracting structural information from a scene [37]. This measure accounts for structural degradation of a scene. If $x = \{x_i \mid i = 1, 2, \dots, N\}$ is the original signal and $y = \{y_i \mid i = 1, 2, \dots, N\}$ is the distorted signal the structural similarity index can be calculated using (1) as follows,

$$SSIM = \frac{(2\bar{x}\bar{y} + C_1)(2\sigma_{xy} + C_2)}{(\bar{x}^2 + \bar{y}^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (1)$$

In this equation $\bar{x}, \bar{y}, \sigma_x, \sigma_y, \sigma_{xy}$ are the estimates of the mean of x , mean of y , the variance of x , the variance of y and the covariance of x and y . C_1 and C_2 are constants. The value of SSIM is between -1 and 1 and gets the best value of 1 if $x_i = y_i$ for all values of i . The quality index is applied to every image using a sliding window with 11x11 circular-symmetric Gaussian weighting functions for which the quality index is calculated and the total index of the image is the average of all the quality indexes of the image.

In order to derive a relationship between the structural information of individual colour/depth map sequences and the perception of depth, the SSIM ratings are used as an objective measure. Existing objective measures do not necessarily accurately model perceptual quality due to their lack of correlations with the human visual system (HVS) characteristics [38]. A number of objective quality models, which are highly correlated to subjective measures, can be found in the literature, such as VQM [31] and Moving Picture Quality Metric (MPQM) [32]. In this paper VQM, developed by the Institute of Telecommunication Sciences (ITS) and American National Standard Institute (ANSI), is used to find a relationship between the objective measures (i.e., VQM scores) of the colour video/rendered left and right views and the perceived quality attributes of stereoscopic video. VQM measures the perceptual effects of video impairments including blurring, jerky/unnatural motion, global noise, block distortion and colour distortion, and combines them in to a single metric. This model for estimating video quality and its associated calibration techniques are independently evaluated by the Video Quality Experts Group (VQEG) and has proved to have high correlation with subjective quality assessment [1]. For more information about the VQM calibrations and quality evaluation techniques readers should refer to [39,40].

III. METHOD

This section describes the methodology employed in carrying out subjective tests using colour and depth map sequences, and the process of mapping subjective quality to objective quality metrics. A detailed description of the test sequences, the SVC

coding parameters, 3-D display equipment and display calibration parameters, the simulated stereoscopic video transmission system over IP, and the objective/subjective score mappings are presented.

A. Test Sequences

Four colour and depth map based stereoscopic video sequences namely *Orbi*, *Interview*, *Break dance* and *Ballet* are used in this experiment. Fig. 3 shows frames from the original scenes used in the experiment.

Fig. 3. Original Test sequences a) *Orbi*, b) *Interview*, c) *Break-dance*, and d) *Ballet*

The *Orbi* and *Interview* sequences have been captured using a depth/range camera, where each colour image sequence and its associated per-pixel depth map are captured simultaneously. *Orbi* is a very complex sequence with camera motion and multiple objects, whereas *Interview* is a sequence captured with a static camera and featuring a stationary background. The resolution of these two sequences is 720x576 pixels which is the resolution of standard definition (SD) TV. The original frame rate of these sequences is 25 frames/s and 250 frames (i.e., 10 seconds long sequences) are used in the experiment.

The rest of the sequences (i.e., *Break dance* and *Ballet*) are obtained from the multi-view image sequences generated by the Interactive Visual Media group at Microsoft Research [8]. The fourth camera view and the corresponding depth map computed from stereo are used in this experiment [8]. The *Break dance* sequence contains a highly dynamic break dancer in the foreground and a number of supporters with limited motion in the background. In contrast to the *Break dance* test sequence, *Ballet* occupies a stationary observer in the foreground and a Ballet dancer operating behind the foreground observer. Both sequences are captured using a stationary camera. The resolution and original frame rate of these two sequences are 1024x768 and 15 frames/s respectively. The 10 seconds long (i.e., 150 frames) *Break dance* and *Ballet* sequences are used in this experiment.

B. Stereoscopic video coding

The JSVM (Joint Scalable Video Model) reference software version 9.4, which is developed by JVT (Joint Video Team), is used to encode each test sequence at different compression rates. The basic encoding parameters are: IPPP... sequence format, a single reference frame, and Content Adaptive Binary Arithmetic Coding (CABAC). However, during the evaluation of stereoscopic video performance over IP network, IPPP...IPPP... sequence format is utilized instead of IPPP... sequence format to minimize error propagation caused by frame losses. Hence, I-frames are included during the coding of colour and depth image sequences every 5 seconds in the video (i.e., an I-frame is coded every 125 frames of Orbi and Interview sequences, whereas an I-frame is inserted every 75 frames when coding Break dance and Ballet sequences).

The Quantization parameters (QPs) of the base and enhancement layer encoders are varied to obtain symmetrically/asymmetrically coded colour and depth image sequences. Highly compressed video is obtained using higher QP values whereas lower QP values are used to obtain good quality video at higher bitrates. In symmetric coding the same QP value is used at the base and enhancement layers to obtain symmetrically compressed colour and depth video sequences. On the other hand in asymmetric coding, two different QP values are used to encode colour and depth video sequences (i.e., incrementing the QP value of either colour or depth video while coding the other sequence using a constant QP value). In order to highlight the visibility of coding artifacts, the QP sizes 30, 35, 40, 45, and 50 are selected for both colour and depth image sequences. However, in the performance analysis of stereoscopic video over IP, the QP value of 30 is used in encoding both colour and depth image sequences.

For all encoding configurations, the average bitrate requirements and resultant PSNR of individual colour and depth map sequences are shown in Table 1. The PSNR of individual components (i.e., colour and depth map) encoded using QP 30 for both colour and depth image sequences show good image qualities at average bitrates of 1054 kbps, 736 kbps, 1243 kbps, and 739 kbps for Orbi, Interview, Break dance, and Ballet sequences respectively. As this paper investigates the quality of

stereoscopic video applications at moderate bitrates, the colour and depth image sequences coded using a QP value 30 is considered as the benchmark of this experiment.

TABLE 1. Image quality and bitrate of SVC coded colour and depth video using the QP value 30

<i>Video Sequence</i>	<i>Orbi</i>	<i>Interview</i>	<i>Break dance</i>	<i>Ballet</i>
<i>Colour/PSNR (dB)</i>	37.709	37.112	38.149	39.310
<i>Depth/PSNR (dB)</i>	40.006	42.314	44.389	44.840
<i>Average Bitrate/(kbps)</i>	1054.9	736.95	1243.2	739.03

C. Equipment

The 42" Philips multi-view auto-stereoscopic display is used in the experiment to display the stereoscopic material. This uses lenticular lens technology to separate left and right views [41]. The input video format of the display is colour and depth map video sequences which are arranged side-by-side. Before displaying 3-D content on the display the supplied colour and depth image sequences are converted into left and right image sequences using the DIBR technique. This display supports 3-D viewing without using viewing aids (e.g., shutter glasses) and screen parallax enabling users to look around the scene objects.

The resolution and the aspect ratio of the 3-D display are 1920x1080 and 16:9 respectively. The optics are optimized for a viewing distance of 3m. Hence, the viewing distance for the observers is set to 3m. This viewing distance is in compliance with the Preferred Viewing Distance (PVD) of the ITU-R BT.500-11 recommendation [42], which specifies the methodology for subjective assessment of the quality of television pictures. The basic principles of this standard have been used in ITU-Recommendations for subjective assessment of stereoscopic television pictures [30]. The 3-D display is calibrated using a GretagMacbeth Eye-One Display 2 calibration device. Peak luminance of the display is 200 cd/m². The measured environmental illumination is 180 lux, which is slightly below the ITU-R BT.500-11 recommended value for home environments (i.e., 200 lux) [42]. The background luminance of the wall behind the monitor is 20 lux. The ratio of luminance of inactive screen to peak luminance is less than 0.02. These environmental luminance measures remained the same for all test sessions, as the lighting conditions of the test room are kept constant.

The original, SVC coded colour and depth image sequences are initially combined into side-by-side video sequences. These side-by-side image sequences are then input to the 3-D display to allow subjects to evaluate the overall image quality and depth perception.

D. Stereoscopic video streaming over IP

IP error patterns, generated specifically for Internet video experiments, are used to corrupt the SVC bit-streams at different PLRs [43]. Each coded video frame (i.e., either colour or depth map frame) is encapsulated in a single Network Abstraction Layer (NAL) unit, and is fragmented into packets of 1400 bytes, which is below the Internet's Maximum Transfer Unit (MTU) size of 1520 bytes. The first fragment contains the important header data of the frame NAL unit. Hence, if the first fragment of a frame is lost, it is assumed that the whole frame is lost, as the header data is unrecoverable without adding redundant information. Furthermore, it is assumed that the sequence headers and the first I frame of both colour and depth image sequences are received uncorrupted. The corrupted bit-streams are later decoded using the JSVM decoder. The JSVM frame copying method is used to conceal the missing frames of both the colour and the depth image sequences. To average the results, the objective image quality metrics (i.e., PSNR) are averaged over 3000 frames using different starting positions in the error traces. For the subjective evaluation test, the clips are selected that feature PSNR values close to the calculated average PSNR for that test condition for both colour and depth sequences.

E. Subjective/Objective Quality Measures

During the subjective evaluation tests, observers are asked to rate the video sequences according to the Double Stimulus Continuous Quality Scale (DSCQS) method, as described in the International Telecommunication Union's Recommendation on subjective quality assessment [42]. The perceived overall image quality and perceived depth are rated on a quality categorical scale from 1 to 5 corresponding with 1 for bad image quality/depth perception and 5 for excellent image quality/depth perception. Furthermore, the scale is labeled with the adjective terms namely [bad], [poor], [fair], [good], and [excellent] according to the ITU-R BT.500-11 recommendations [42].

During the analysis of results, the difference in subjective ratings for the impaired image sequence and the original image sequence is calculated. Then the difference is scaled into a linear opinion score scale, which ranges from 0 (excellent) to 100 (bad). The 0 represents the perceived image quality same as the original image sequence and 100 represents the worst perceived image quality compared to the original video sequence. Then the mean opinion score (MOS) for each test sequence is obtained after averaging the opinion scores over all subjects.

PSNR, SSIM, and VQM are used as objective quality metrics. Each metric is applied using three different methods:

- Application to the colour video only (e.g., Y-PSNR of just the colour video);
- Application to the depth video only;
- Application to the left and right views rendered using the DIBR technique.

The DIBR technique uses (2) to render the stereoscopic left and right views:

$$P_{pix} = -x_B \frac{N_{pix}}{D} \left[\frac{m}{255} (k_{near} + k_{far}) - k_{far} \right] \quad (2)$$

Where, N_{pix} and x_B are the number of horizontal pixels of the display and eye separation respectively. The depth value of the image is represented by the N -bit value m . k_{near} and k_{far} specify the range of the depth information respectively behind and in front of the picture, relative to the screen width N_{pix} . The viewing distance is represented by the parameter value D .

The VQM, developed by the Institute for Telecommunication Sciences (ITS), is used to obtain the VQM ratings for the colour and rendered left and right views video sequences [31]. The VQM ratings are mapped to a DSCQS based MOS scale, which ranges from 0 (excellent) to 100 (bad).

The relationship between the perceived quality (i.e., MOS) and other objective measures of distortion is approximated by a symmetrical logistic function, as described in ITU-R BT.500-11 [42]. The symmetrical logistic function is used because all the distortion parameters (i.e., objective quality metrics) are measured in related units (e.g., PSNR/dB). The normalization of opinion scores and the symmetrical logistic function are given (3) and (4) respectively.

$$p = (\bar{u} - u_{min}) / (u_{max} - u_{min}) \quad (3)$$

with: p : normalized opinion score u_{min} : minimum score available on the u -scale for the worst quality

$\bar{u}$: mean scores u_{max} : maximum score available on the u -scale for the best quality

$$p = 1 / [1 + \exp (D - D_M) G] \quad (4)$$

where, p is the normalized opinion score, D is the distortion parameter, D_M and G are constants and G may be positive or negative.

IV. EXPERIMENT 1: SYMMETRIC/ASYMMETRIC CODING OF COLOUR AND ASSOCIATED DEPTH MAP VIDEO

The perceptual quality of symmetrically/asymmetrically coded stereoscopic video content is analyzed in this section. The relationships between different perceptual attributes and between subjective and objective measurements of SVC coded colour plus depth video is also presented. First, the methods for carrying out subjective tests and coding combinations are presented. Finally the results are analyzed and conclusions are drawn.

A. METHOD

Observers

Thirty two non-expert observers (ten female observers and twenty two male observers) volunteered to participate in the experiment. They are divided into two groups in order to assess one attribute per group. The attributes assessed in this experiment are perceived image quality and perceived depth. The observers are mostly research students and staff with a

technical background. Their ages range from 20 to 40. Fourteen observers have prior experience with 3-D video content using different viewing aids such as shutter glasses and red-blue anaglyph glasses. All participants have a visual acuity of ≥ 1 (as tested with the Snellen chart), good stereo vision < 60 seconds of arc (as tested with the TNO stereo test), and good colour vision (as tested with the Ishihara test).

Stimuli

The stimulus set contains thirteen SVC coded video sequences and the original, uncompressed version of each scene is used as the reference in the evaluation test. Stereo image pairs are obtained by symmetric or asymmetric coding of colour and depth image sequences. The coding combinations used to reduce the average bitrate requirements are 30_30, 30_35, 30_40, 30_45, 30_50, 35_35, 40_40, 45_45, 50_50, 35_30, 40_30, 45_30, and 50_30. The first and second numbers respectively refer to the colour and depth map QPs used during SVC encoding. All combinations of SVC coded colour and depth video sequences are presented twice. Therefore, in total four test sequences, one repetition and 13 coding combinations are used. This resulted in a stimulus set of $4 \times 2 \times 13 = 104$ video sequences.

B. Procedure

A set of 104 stereoscopic video sequences is randomized and presented sequentially. The overall image quality and depth perception are rated by a different group of 16 observers and each observer rates one attribute. The stimulus set is judged for each attribute in two sub-sessions (i.e., 52 stimuli for each sub-session) with a small break in between. Each sub-session lasts approximately 25 minutes. In order to evaluate the negative effects of 3-D viewing for a long period of time, a symptom (i.e., general discomfort, fatigue, headache, eye-strain, difficulty focusing and blurred vision) check list is maintained for each subject. Each symptom is rated as *none*, *slight*, *moderate* and *severe* by all subjects before and after completing a session. Before starting the test observers are asked to read the instructions explaining the test procedure and the attributes they have to rate. Thereafter, a trial session is conducted before the actual test in order to get acquainted with the 3-D display and the range of compression levels associated with the stimulus set.

C. Results

Perceived overall image quality and depth

Fig. 4 shows the MOS scores for the perceived overall image quality, depth perception, the colour image sequence measured using VQM, and the average of the rendered left and right video measured using VQM. The x-axis represents the SVC QP parameters for the symmetric and asymmetric coding combination of colour and depth video sequences in increasing bitrate to the right. The y-axis represents the MOS scores from 0 (excellent) to 100 (bad). Furthermore, standard errors for perceived image quality/depth and 5% error margin for VQM scores are also presented.

Fig. 4. The MOS scores for the perceived overall image quality (Subj. Q), depth perception (Subj. D), colour video measured using VQM, and the average of the rendered left and right views measured using VQM: (a) Orbi; (b) Interview; (c) Break dance; (d) Ballet.

The subjective quality scores for image quality (i.e., MOS) increase with increasing bitrates for all of the sequences. But there are visible quality dips when the coding configuration contains high QP values for the colour video component (e.g., 50_50, 45_45, 50_30, and 45_30). These quality dips prevent the formation of a straightforward relationship between perceived overall image quality and average bitrates. However, medium QP values (i.e., 35 and 40) for colour video coding together with any QP value for depth video can be effectively used in reducing the bandwidth while achieving good perception of overall image quality and depth.

The VQM score of the colour video and average VQM score of the rendered left and right video using SVC coded colour and depth map sequences demonstrate similar trends compared to the perceptual image quality scores for all the sequences, but with an offset in terms of the MOS score. This indicates that viewing the video in 3-D has increased the perception of image quality for the SVC coded stereoscopic video, compared to 2-D video quality predicted by VQM. In addition to that, the VQM ratings of the colour and individual left and right video sequences could be used as a good model for prediction of the perceived image quality and depth.

The depth perception shows a similar trend to image quality perception, but on average has lower MOS scores compared to the overall image quality results. Therefore, depth perception is not significantly affected by using different QP values for depth map coding. However, the artifacts (i.e., blocking, blurring) introduced to the colour video during image coding heavily influences depth perception (e.g., the depth perception is low for the 35_30, 40_30, 45_30, and 50_30 coding combinations where depth is coded using a constant QP and colour with different QPs). Furthermore, it can be deduced from Fig.4 that there is high correlation between the overall image quality and the depth perception. Fig. 5 shows the scatter plot for perceived image quality and depth perception attributes for each coding combination. The relationship between these two attributes (i.e., perceived overall image quality and depth) is approximated by a first order linear function. According to Fig. 5 the correlation coefficient of determination between perceived image quality and depth is 0.92. Thus, the perception of depth is positively influenced by the perceived image quality for SVC coded colour and depth sequences.

Fig. 5. Scatter plot showing the correlation between image quality and depth perception

Fig. 6 and Fig. 7 show the perceived overall image quality and depth perception respectively for different coding configurations with fixed QP for colour image coding. The coding configurations which have a constant QP for the colour image, feature good image quality, irrespective of the QP value for the depth. The x-axis represents the asymmetric coding

combination of colour and depth video sequences in increasing bitrate to the right. According to Fig. 6 depth map coding using different QPs does not have any effect on perceived overall image quality at all bitrates. On the other hand there is a slight improvement in depth perception with the increase of overall average bitrates (see fig. 7). Thus, the average bitrate requirements for stereoscopic video applications can be reduced using high QPs in depth map coding without affecting the perception of overall image quality, even though there is a slight degradation in depth perception.

Fig. 6. The perceived overall image quality (MOS) for *Orbi*, *Interview*, *Break dance*, and *Ballet* sequences using different QPs for depth map coding.

Fig. 7. The perceived depth (MOS) for *Orbi*, *Interview*, *Break dance*, and *Ballet* sequences using different QPs for depth map coding.

The relationship between the perceived quality (i.e., MOS) and other objective measures of distortion are important in optimizing the video system (e.g., encoding algorithm and media delivery method) for a given condition, without having to

conduct any subjective evaluation tests. Hence, the relationships between the objective quality models and perceived 3-D quality attributes for SVC coded colour and depth sequences are analyzed. The quality assessment models used for prediction are PSNR/SSIM/VQM scores of the coded colour video, PSNR/SSIM of the coded depth map sequence, and the average PSNR/SSIM/VQM of the rendered left and right views. The relationship is approximated by a symmetrical logistic function, as described in (4).

Fig. 8 shows the scatter plot of perceived quality (MOS) versus model prediction for each model. Each plot of Fig. 8 contains model predictions for both perceived overall image quality and perceived depth. This demonstrates that all image quality models except the objective metrics associated with the depth map sequence perform better in predicting the perceived quality of SVC coded stereoscopic video. The quantitative measures for each prediction model are presented in Table 2. This shows that the VQM scores of the colour video sequence is a good model for predicting the perceived image quality. Whereas the average VQM of the rendered left and right view shows high correlation to the perceived depth of SVC coded stereoscopic image sequences. The PSNR/SSIM scores of depth map sequences have failed in predicting the perceived image quality and depth of SVC coded stereoscopic video due to the influence of the coded colour image quality in perceiving depth. This is further reflected by the high correlation of objective quality models based on colour video on depth perception.

(c)

(d)

(e)

(f)

(g)

(h)

Fig. 8. Scatter plots of subjective MOS (i.e., MOS for image quality and depth perception) versus model prediction. Each sample point represents one test image sequence. (a) Colour PSNR, (b) Depth PSNR, (c) Average PSNR of the rendered left and right video, (d) Colour VQM, (e) Average VQM of the rendered left and right video, (f) Colour SSIM, (g) Depth SSIM, and (h) Average SSIM of the rendered left and right video

TABLE 2: Performance of Image Quality Assessment Models. CC: Correlation Coefficient; RMSE: Route Mean Squared Error; SSE: Sum of Squares due to Error

<i>Objective Quality Model</i>	<i>Overall Image Quality</i>			<i>Depth Perception</i>		
	<i>CC</i>	<i>RMSE</i>	<i>SSE</i>	<i>CC</i>	<i>RMSE</i>	<i>SSE</i>
Colour PSNR	0.8100	0.0924	0.4265	0.8192	0.0529	0.1401
Depth PSNR	0.0043	0.1242	0.7714	0.0000	0.2119	2.2450
Average PSNR of the Rendered Left and Right views	0.7356	0.1090	0.5936	0.7768	0.0588	0.1729
Colour VQM	0.8904	0.0701	0.2460	0.8421	0.0490	0.1224
Average VQM of the Rendered Left and Right	0.8764	0.0745	0.2774	0.8432	0.0493	0.1215
Colour SSIM	0.7618	0.1034	0.5347	0.7274	0.0650	0.2112
Depth SSIM	0.0015	0.2117	2.2410	0.0188	0.1233	0.7602
Average SSIM of the Rendered Left and Right Views	0.7528	0.1054	0.5550	0.7195	0.0659	0.2173

V. EXPERIMENT 2: PERFORMANCE OF COLOUR AND ASSOCIATED DEPTH MAP VIDEO OVER IP

The effect of channel errors (i.e., frame losses) on perceived quality of stereoscopic video is analyzed in this section. The SVC coded colour plus depth map video is corrupted using the error patterns representing an IP core network. The resultant quality of the synthesized binocular video is rated by subjects for perceived overall image quality and depth. Furthermore, the relationship between the perceived quality over IP network and objective quality measures such as PSNR and SSIM are discussed.

A. Method

Observers

Thirty non-expert observers (eleven female observers and nineteen male observers) volunteered to participate in the experiment. The observers are mostly research students and staff with a technical background. Their ages range from 20 to 35. Nine observers have prior experience with different formats of 3-D video content. All participants have a visual acuity of ≥ 1 (as tested with the Snellen chart), good stereo vision < 60 seconds of arc (as tested with the TNO stereo test), and good colour vision (as tested with the Ishihara test). The attributes, namely perceived overall image quality and perceived depth are assessed using two groups (i.e., 15 observers for each group) of observers and each observer rates only one attribute.

Stimuli

Four colour and depth map based stereoscopic video sequences namely *Orbi*, *Interview*, *Break dance* and *Ballet* are used in this experiment.

The stimulus set contains five scene versions which are decoded after corrupting the SVC bit-streams using five different error patterns (i.e., 0%, 3%, 5%, 10%, and 20% PLRs). The original, uncompressed version of each scene is used as the reference in the evaluation test. The symmetric coding configuration 30_30 (i.e., both colour and depth map are coded with QP 30) is used to encode each test sequence. The average bitrates of the coded sequences are 1073 kbps, 763 kbps, 1251 kbps, and 748 kbps, for *Orbi*, *Interview*, *Break dance*, and *Ballet* sequences respectively.

All combinations of corrupted, SVC coded colour and depth video sequences are presented twice to the viewer. Therefore, in total four test sequences, one repetition and five PLRs are used. This resulted in stimulus set of $4 \times 2 \times 5 = 40$ video sequences.

B. Procedure

A set of 40 stereoscopic video sequences is randomized and presented sequentially. The subjective test for each attribute lasts for approximately 20 minutes. Before starting, the test observers are asked to read the instructions explaining the test procedure and the attributes they had to rate. Thereafter, a trial session is conducted before the actual test in order to get acquainted with the 3-D display and the range of different PLRs associated with the stimulus set.

C. Results

Perceived overall image quality and depth

Fig. 9 shows the MOS scores for perceived overall image quality, depth perception, the colour image sequence using VQM, and the average of the rendered left and right video using VQM, with different packet loss rates. The standard errors for perceived image quality/depth and 5% confidence interval for VQM scores are also presented. The x-axis and y-axis respectively represent different PLRs applied for SVC coded colour and depth sequences and MOS scores from 0 (excellent) to 100 (bad).

As expected, the perception of image quality and depth degrades with increasing PLRs for all test sequences. However, the MOS scores for depth perception are better compared to the perceived image quality at all bitrates. Therefore, the loss of colour and depth frames has significantly affected the perceived overall image quality, whereas depth perception is not severely affected. The VQM ratings of the corrupted colour image sequence and the rendered left and right video using corrupted colour and depth video sequences follow the same trend as the subjective ratings for image quality and depth. But, both VQM ratings are higher compared to the MOS of the overall image quality. This shows that viewing the video in 3-D (i.e., the added value of depth in 3-D viewing) improves the perceived image quality at all PLRs, compared to 2-D video quality predicted using VQM. Nevertheless, the results indicate that both VQM ratings can be used to predict the user perception of overall image quality and depth with reasonable accuracy.

According to Fig. 9, both perceived image quality and depth follows the same trend at all PLRs. Fig. 10 shows the scatter plot between perceived image quality and depth perception for all the test sequences at different PLRs. The correlation coefficient of determination is 0.88 when approximated by a first order linear polynomial. Thus, increase in perception of image quality could be attributed to an increase in depth perception in a scenario where colour and depth video is transmitted over a packet loss network.

(a)

(b)

(c)

(d)

Fig. 9. The MOS scores for the perceived overall image quality, perceived depth, the colour video measured using VQM, and the average of the rendered left and right views measured using VQM under different packet loss rates: (a) *Orbi*, (b) *Interview*, (c) *Break dance*, and (d) *Ballet* sequences.

Fig. 10. Scatter plot between image quality and depth perception

The objective quality measures could be used in predicting the perceived attributes of stereoscopic video over packet loss networks. Hence, the relationships between the subjective ratings and objective quality metrics (i.e., colour PSNR/SSIM/VQM and the average PSNR/SSIM/VQM of the rendered left and right video using the corrupted SVC bitstream) are derived. As depth perception is hindered by the perceived image quality (see Fig. 10) the individual objective measures of depth image sequence are not used for model prediction. Fig. 11 shows the relationships between the perceived image quality/depth and different image quality models. Each data point on the scatter plots of Fig. 11 represents one corrupted/uncorrupted image

sequence. The relationship is mapped using the symmetrical logistic function described in (4). According to Fig. 11, all considered objective models except the SSIM metric demonstrate good model prediction for overall image quality and depth perception of stereoscopic video over IP. The structural consistency measure using SSIM has not performed well under the given conditions. This may be due to the frame concealment technique (i.e., dropped frames are concealed using the previously decoded frames) used in this experiment to prevent error propagation. The quantitative measures on the performance of the quality assessment models are shown in Table 3. The average PSNR of the rendered left and right video shows good model prediction for both image quality and depth perception. The colour PSNR performance is the same as the prediction level achieved by the average PSNR of the rendered left and right video. Furthermore, both VQM ratings show high correlation for image perception but less correlation for depth perception. However, the prediction for depth perception is better with the average VQM of the rendered left and right views compared to colour VQM.

(a)

(b)

(c)

(d)

Fig. 11. Scatter plots of subjective MOS (i.e., MOS for image quality and depth perception) versus model prediction. Each sample point represents one test image sequence. (a) Colour PSNR, (b) Average PSNR of the rendered left and right video, (c) Colour VQM, (d) Average VQM of the rendered left and right views, (e) Colour SSIM, and (f) Average SSIM of the rendered left and right video.

TABLE 3: Performance of Image Quality Assessment Models. CC: Correlation Coefficient; RMSE: Route Mean Squared Error; SSE: Sum of Squares due to Error

Objective Quality Model	Overall Image Quality			Depth Perception		
	CC	RMSE	SSE	CC	RMSE	SSE
Colour PSNR	0.7992	0.0802	0.1158	0.8854	0.0496	0.04426
Average PSNR of Rendered Left and Right views	0.8054	0.07890	0.1123	0.8884	0.0489	0.04310
Colour VQM	0.7848	0.0830	0.1241	0.7918	0.0668	0.08039
Average VQM of Rendered Left and Right	0.8045	0.0792	0.1128	0.8303	0.0603	0.06554
Colour SSIM	0.4926	0.1275	0.2927	0.6183	0.0905	0.1474
Average SSIM of Rendered Left and Right Views	0.5099	0.1253	0.2827	0.6310	0.0890	0.14250

VI. CONCLUSION

This paper addresses the lack of good 3-D video quality assessment metrics, which could hinder the development and introduction of new 3-D video services. Subjective results are analyzed to demonstrate the most effective metrics for determining the quality of compressed 3-D color plus depth video. Further tests are carried out to establish the effectiveness of the metrics in packet loss scenarios.

The results show that the output from the VQM objective metric can be mapped so that it correlates strongly with both the overall viewer perception of image quality and depth perception. It is noted that the raw MOS scores from VQM are consistently worse than viewer ratings. This suggests that the addition of the depth increases viewer satisfaction significantly, masking the effects of compression artifacts in the color video. A further important result is that assessing the quality of the depth component on its own, using any objective metric, provides no useful information on viewer ratings of image quality or depth perception. The artifacts introduced to the color video hinder the depth perception of stereoscopic video sequences. Hence, there is a high correlation between perceived image quality and perceived depth.

In packet loss simulations, it can be seen that depth perception is unaffected by frame losses in most cases. It is also clear the VQM objective metric can be mapped to provide a good prediction of viewer MOS scores for image quality and depth perception.

Overall, the results show that the predicted image quality of the color video and the DIBR rendered left and right views using VQM is a good model prediction for both the perceived overall image quality and depth of 3-D video. This is true for both error free and packet loss scenarios. This technique correlates strongly with viewer subjective tests, and is certainly no worse a metric than PSNR for 2-D video. This suggests that, while subjective test results should remain the “gold standard”, the use of objective quality assessment metrics as described in this paper is an acceptable compromise for the 3-D video research community.

Further work is suggested to extend the results reported in this paper. In particular, it would be interesting to carry out the same tests using different 3-D displays to fully determine the effect of the displays on viewer MOS scores for image quality and depth perception.

ACKNOWLEDGMENT

The authors would like to thank Dr. Hemantha Kodikaraarachchi and Duncan Aitken for insightful comments, Erhan Ekmekcioglu for assisting in display calibration and Interactive Visual Media Group at Microsoft Research for providing *Break dance* and *Ballet* sequences.

REFERENCES

- [1] Video Quality Experts Group (VQEG), "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II", *VQEG*, 2003. Available: <http://www.vqeg.org>.
- [2] *Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information*, ISO/IEC JTC1/SC29/WG11, Jan. 2007, Doc. N8768, Marrakesh, Morocco.
- [3] C. Fehn, R. De La Barre, and S. Pastoor, "Interactive 3-D TV: Concepts and Key Technologies", *Proceedings of the IEEE*, vol. 94, pp. 524-538, 2006.
- [4] B. P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability", *Signal Processing: Image Communication: Special issue on three-dimensional video and television*, vol. 22, pp. 217-234, February 2007 (available online).
- [5] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Trans-mission for a New Approach on 3D-TV", *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, pp. 93-104, January 2004.
- [6] M. Kawakita, K. Lizuka, T. Aida, H. Kikuchi, H. Fujikake, J. Yonai, and K. Takizawa, "AXI-vision camera (Real-time distance-mapping camera)", *Appl. Opt.*, vol. 39, pp. 3931-3939, 2000.
- [7] G. Iddan and G. Yahav, "3D imaging in the studio (and elsewhere)", *Proc. SPIE*, vol. 4298, pp. 48-55, 2001.
- [8] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation", *ACM SIGGRAPH and ACM Trans. on Graphics*, Los Angeles, CA, Aug. 2004, pp. 600-608.
- [9] C. Fehn, "3D-TV Approach using Depth-Image-Based Rendering (DIBR)", *In Proceedings of Visualization, Imaging, and Image Processing (VIIP'03)*, 2003, pp. 482-487.
- [10] R. Krishnamurthy, B.-B. Chai, H. Tao, and S. Sethuraman, "Compression and transmission of depth maps for image-based rendering", *in Proc. ICIP*, Thessaloniki, Greece, 2001, pp. 828-831.
- [11] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, and A. Koz, "Coding Algorithms for 3DTV-A Survey", *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, pp. 1606-1621, Nov. 2007.
- [12] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo images sequences using joint 3D motion/disparity compensation", *IEEE Trans. Video Technol.*, vol. VT-11, Mar. 1973.
- [13] C. Fehn, K. Hopf, and Q. Quante, "Key Technologies for an Advanced 3D-TV System", *In Proceedings of SPIE Three-Dimensional TV, Video and Display III*, 2004, pp. 66-80.

- [14] C.T.E.R. Hewage, H.A. Karim, S. Worrall, S. Dogan, and A.M. Kondoz, "Comparison of Stereo Video Coding Support in MPEG-4 MAC, H.264/AVC and H.264/SVC", *In Proceedings of 4th IET conference on Visual Media Engineering*, London, UK, 2007.
- [15] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz and M. Wien, "Joint draft 8 of SVC amendment", *ISO/IECJTC1/SC29/WG11 and ITU-T SG16 Q.69 (JVT-U201)*, 2006.
- [16] T. Wiegand and G. J. Sullivan, "Overview of the H.264/AVC Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, July 2003.
- [17] G.B. Akar, A. M. Tekalp, C. Fehn, and M.R. Civanlar, "Transport methods in 3DTV-A survey", *IEEE Transactions on Circuits and Systems for Video Technology*, vol.17, pp.1622-1630, November 2007.
- [18] M. Johanson, "Stereoscopic video transmission over the internet". *In Proceedings of the IEEE Workshop on Internet Applications*, San Jose, CA, USA, 2001, pp.12-19.
- [19] M.G. Perkins, "Data compression of stereopairs", *IEEE Transactions on Communications*, vol.40, pp.684-696, April 1992.
- [20] D.V. Meegan, L.B. Stelmach, and W.J. Tam, "Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery", *Journal of Experimental Psychology: Applied*, vol.7, pp. 143-153, January 2001.
- [21] P. Seuntiens, L. Meesters, and W. Ijsselstein, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation", *ACM Transactions on Applied Perception (TAP)*, vol. 3, pp. 95-109, 2006.
- [22] L. Stelmach, W.J. Tam, D. Meegan, A. Vincent. "Stereo Image Quality: Effects of Mixed Spatio-Temporal Resolution", *IEEE Transactions on circuits and systems for video technology*, vol. 10, pp. 188-193, 2000.
- [23] H. A. Karim, A. H. Sadka, and A. M. Kondoz, "Reduced Resolution Depth Compression for 3-D Video over Wireless Networks", *Proceedings of the IEE 2nd International Workshop on Signal Processing for Wireless Communications*, King's College London, UK, June 2004.
- [24] C.T.E.R. Hewage, S. Worrall, S. Dogan, H. Kodikara Arachchi, and A.M. Kondoz, "Stereoscopic TV over IP", *In Proceedings of the 4th IET European Conference on Visual Media Production*, 2007.
- [25] O. Grau, S. Minelly, and G. Thomas, "Applications of depth metadata," *in Proc. IBC*, Sept. 2001.
- [26] E. Ekmekcioglu, S. Worrall, A. M. Kondoz, "Bit-Rate Adaptive Down-sampling For The Coding Of Multi-View Video With Depth Information", *3DTV-Conference 2008, The True Vision Capture, Transmission And Display Of 3D Video*, Istanbul, Turkey, May 2008, Accepted for publication.

- [27] S. Bech, R. Hamberg, M. Nijenhuis, C. Teunissen, H. L. de Jong, P. Houben, and S. K. Pramanik, "The RaPID perceptual image description method (RaPID)", *Proc. SPIE*, vol. 2657, pp. 317-328, 1996.
- [28] L.B. Stelmach, W.J. Tam, and D.V. Meegan, "Perceptual basis of stereoscopic video", *Proc. SPIE: Special issue on Stereoscopic Displays and Virtual Reality Systems VI*, vol. 3639, pp. 260-265, May 1999.
- [29] L.M.J. Meesters, W.A. IJsselsteijn, and P.J.H. Seuntjens, "Survey of perceptual quality issues in three-dimensional television systems", *Proceedings of the SPIE*, vol. 5006, pp. 313-326, 2003.
- [30] International Telecommunication Union/ITU Radio communication Sector, "Subjective Assessment of Stereoscopic Television Pictures", *ITU-R BT.1438*, Jan 2000.
- [31] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, pp. 312-322, Sept. 2004. Available: <http://www.its.bldrdoc.gov/pub/n3/video/index.php>.
- [32] C. J. B. Lambrecht and O. Verscheure. "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System", *Proc. SPIE*, vol. 2668, pp. 450-461, March 1996.
- [33] P. Seuntiëns, L. Meesters, and W. IJsselsteijn, "Perceptual evaluation of JPEG coded stereoscopic images", *Proc. SPIE*, vol. 5006, pp. 215-226, 2003.
- [34] W. Tam, L. Stelmach, and P. Corriveau, "Psychovisual aspects of viewing stereoscopic video sequences", *Proc. SPIE*, vol. 3295, pp. 226-235, 1998.
- [35] W. IJsselsteijn, H. de Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 225-233, Mar. 2000.
- [36] J. Starck, J. Kilner, and A. Hilton, "objective quality assessment in free-viewpoint video production", *3DTV 2008 conference*, Istanbul, Turkey, 2008.
- [37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*, vol. 13, pp. 600-612, April 2004.
- [38] B. Girod, "What's wrong with Mean-Squared Error", *Digital Images and Human Vision*, A. B. Watson Ed., Chapter 15, pp. 207-220, MIT press, 1993.
- [39] A. P. Hekstra et al., "PVQM- a perceptual video quality measure", *Signal Processing: Image Communication*, vol. 17, pp. 781-798, 2002.
- [40] ITS Video quality research, Available: www.its.bldrdoc.gov/n3/video/VQM_software.php, 2007.
- [41] Philips, "Philips 3D solutions", 2006. Available: <http://www.philips.com/3dsolutions>.

[42] International Telecommunication Union/ITU Radio communication Sector, "Methodology for the subjective assessment of the quality of television pictures", *ITU-R BT.500-11*, Jan 2002.

[43] W. Stephan. "Error patterns for Internet video experiments", *ITU-T SG16 Document Q15-I-16-R1*, 1999.

Chaminda T.E.R. Hewage received the B.Sc. (Hons.) degree in Electrical and Information Engineering from the University of Ruhuna, Galle, Sri Lanka. During 2004-2005, he worked as a Telecommunication Engineer in the field of Data Communication. Currently, he studies for his Ph.D. at University of Surrey, Guildford, UK. His research aims at providing QoS support for 3-D video communication application. He also works on VISNET II Network of Excellence (NoE) of IST FP6

programme in the area of robust immersive media communications. He is a member of IET and a student member of IEEE. He was awarded a gold medal by University of Ruhuna, Sri Lanka, for his achievements in Engineering discipline at the general convocation held in 2004.

S. Worrall received the MEng Electronic Engineering degree from the University of Nottingham in 1998. He joined the Centre for Communication Systems Research (CCSR), at the University of Surrey, in 1998 as a research student, and obtained his PhD in 2001. From 2001-2003, he continued to work within CCSR as a research fellow, and was appointed as a lecturer in multimedia communications in 2003. His research interests include error robust video coding, multiple description coding, video

compression, transmission of video over wireless networks, and multi-view video coding. He is a member of the IEEE and the IET.

Safak Dogan received received his B.Sc. (Hons.) degree in Electronics and Communications Engineering from the Technical University of Istanbul, Turkey, in 1995, the M.Sc. degree in Satellite Communications Engineering and the Ph.D. degree in Multimedia Communications Engineering, both from the University of Surrey, Guildford, UK, in 1996 and in 2001, respectively. He has been with the I-Lab Multimedia Technologies and DSP Research Group of the Centre for Communication Systems

Research (CCSR) at the University of Surrey since 2001. His primary research interests are video adaptation, context-based media content adaptation, video transcoding, multimedia communication networks, low bit rate video communications, error resilience for next generation mobile-wireless video transmissions, and multi-party audiovisual collaboration/conferencing

technologies. Dr Dogan has authored several research papers published in various peer-refereed international conferences and prestigious journals. He has acted as a referee for the Research Council of Canada, and has also been serving as a reviewer to a number of international journals and conferences in his research areas. Moreover, he has contributed to a technical book with a highly specialized video transcoding chapter. He is a member of the IEEE and one of the leading members of the European Commission Future Media Internet-Task Force (formerly known as the Networked Media-Task Force).

Stéphane Villette received the engineering degree from the Ecole Nationale Supérieure des Telecommunications (ENST) in Paris in 1996. He joined the Centre for Communication Systems Research (CCSR), at the University of Surrey, in 1996, became a research fellow in 1997 and obtained his PhD in 2001. His main interests are in signal processing, in particular speech and audio coding and processing, and 3D audio/video. He and his team have been involved in the development of low bit rate

speech coding algorithms, and developed candidate systems to several international standardization efforts. He is a member of the IEEE.

A.M. Kondo received the B.Sc. (Hons.) degree in engineering, the M.Sc. degree in telematics, and the Ph.D. in communication in 1983, 1984, and 1986, respectively. He became a Lecturer in 1988, a Reader in 1995, and then in 1996, a Professor in Multimedia Communication Systems and deputy director of Centre for Communication Systems Research (CCSR), University of Surrey, Guildford, U.K.

He has over 250 publications, including a book on low bit-rate speech coding and several book chapters. He has graduated more than 40 Ph.D. students in the areas of speech/image and signal processing and wireless multimedia communications, and has been a consultant for major wireless media terminal developers and manufacturers. Prof Kondo has been awarded several prizes, the most significant of which are The Royal Television Societies' Communications Innovation Award and The IEE Benefactors Premium Award. He has been on the Refereeing College for EPSRC and on the Canadian Research Councils. He is a member of the IEEE and the IET.B.Eng.