SUBJECTIVE QUALITY FACTORS IN PACKET 3D VIDEO

L.Pinto, J. Carreira, S. Faria, N. Rodrigues, P. Assuncao

{luispinto25, jfmcarreira}@gmail.com; {sergio.faria, nuno.rodrigues, pedro.assuncao}@ipleiria.pt

Polytechnic Institute of Leiria/ESTG, Institute of Telecommunications
Leiria, Portugal

ABSTRACT

This paper presents an experimental study on subjective quality of 3D video under packet loss conditions. A priority network is assumed, such that the base view of a stereoscopic video stream is delivered through a guaranteed channel and the auxiliary view is subject to packet losses over a low priority channel. Due to the packetisation scheme, one frame is lost whenever a packet is lost. Three frame concealment methods are proposed to deal with frame losses and seven different 3DTV sequences are used. The subjective evaluation study was carried out for different packet loss ratios and concealment methods, using shutter glasses with synchronized display. The experimental results show that sequence disparity, type of loss concealment and packet loss pattern should be jointly considered as relevant quality factors in 3D video over packet networks.

***Index Terms*—** 3D packet video, subjective quality, 3D IP TV

1. INTRODUCTION

The forthcoming 3D video services and applications based on the multiview video coding (MVC) extension of the H.264/MPEG-4 AVC standard [1], are expected to be delivered over diverse network infrastructures as happens nowadays with classic 2D video. Regardless of the underlying transmission technology (e.g., optical fiber, satellite, broadband wireless), packet based transport over IP protocols is foreseen as natural evolution for 3D video services over the future media networks [2]. In such 3D media communications environment, the occurrence of packet loss is inevitable and the consequences are well known. In video transmission, lost packets may result in missing frames at the receiving terminal which in turn lead to error propagation over several other frames due to coding dependencies.

In the case of 3D MVC video streams, the effect of such coding dependencies can be much worse than in 2D video. In addition to the temporal dependency within the same Group of Pictures (GOP) in both views, in 3D MVC streams there is an additional level of dependency between views. This higher level of dependency leads to catastrophic 3D distortion effects when frames of the base view are lost. Another possible cause for losing frames in 3D coded video can be content adaptation engines operating along communication networks [3], though such process can be controlled to limit error propagation. In recent work the impact of packet loss on perceived 3D quality has been evaluated for the video+depth format [4,5]. However, since the video+depth format does not suffer from the same type of dependency as MVC and the type of loss concealment is quite different, the impact of lost packets in perceived quality is not necessarily the same.

Frame loss concealment methods previously used in 2D methods were recently extended into 3D video [6,7] in order to cope with packet loss in transport networks. In [8] the authors evaluate the subjective impact of basic frame loss concealment methods under regular frame skipping. However, the results are not relevant for the case of random bursty losses, as those occurring in the most common packet networks.

In this paper, a subjective evaluation study is presented with the aim of finding relevant subjective quality factors that affect the user quality of experience (QoE) in 3D video delivery services over lossy packet networks. The 3D perceived quality is evaluated, based on the influence of packet loss probability combined with variable loss burst length and different types of frame loss concealment. This study is only focused on the influence of packet loss in depth perception, thus high quality images are used. The results obtained through subjective testing reveal that the significance of such factors to the overall user perception of 3D quality exhibits some level of interdependency. For instance, for a certain packet loss probability, there is no frame loss concealment method better than the others. Instead, the best concealment method is suggested to be dependent on both the loss probability and the error burst size.

The paper is organized as follows. The next section describes the subjective evaluation framework. Section 3 describes the frame loss concealment methods used to recover the lost frames and section 4 presents the experimental setup used to carry out the subjective testing. Section 5 presents the results and related discussion. Finally, conclusions and remarks are presented in section 6.

2. SUBJECTIVE EVALUATION FRAMEWORK

The framework used to evaluate the subjective quality obtained in packet 3D video uses two models to simulate the network loss behavior and the concealment methods used at the decoder. In the packetisation stage, the maximum packet size is assumed to be used. Then, each single packet is capable of carrying either complete frames or most of their coded data. Therefore, in this model each lost packet leads to losing an entire frame. As a consequence the receiver has to implement a loss concealment strategy for the whole frame.

2.1 Packet Loss Model

The random packet loss process used in this paper is based on a two-state Markov model. A finite state Markov chain, characterized by a good (G) and a bad (B) state and transition probabilities representing packet arrivals, was implemented to generate the packet loss patterns used to in video quality evaluation studies (Figure 1).

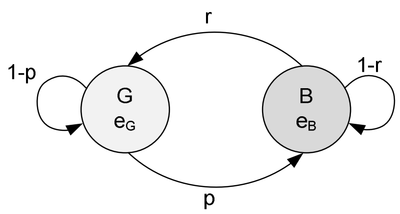

Fig. 1 – Two-State Markov Model

This is based on the Gilbert-Elliot model where errors occur with independent probabilities in each state [9]. The simplified Gilbert model was used in this paper where the error rates are 0 and 1 for the *good* and *bad* states, respectively. In the simplified Gilbert model of Fig. 1, *p* represents the probability of a transition from the *good state* to the *bad state* and *r* represents the opposite. The *good state* corresponds to a successful reception of a packet whilst the *bad state* implies an error in the reception of a packet. The output of this Markov model is a packet error trace file, where the number '1' means a lost packet and '0' a successfully received one. The parameters used to achieve the desired overall error probabilities in trace files were set in the experimental process.

2.2 Frame Loss Concealment Methods

As mentioned before, in this study whenever a packet is lost, it is assumed that the corresponding frame is lost. Therefore, the lost frames must be recovered in order to reconstruct the periodic presentation and then to evaluate the subjective impact in the perceived 3D quality. Since the aim is to evaluate the impact of packet loss on perceived depth, it is assumed that the user is guaranteed the minimum experience of 2D video. Therefore, the effect of losing normal 2D viewing does not exist and only depth perception is affected by frame losses. In this subjective study, a lossless base view provides a guaranteed 2D video sequence while the auxiliary view is the only one prone to losses. Thus, when frames of the auxiliary view are discarded, only depth information is affected while 2D video quality is maintained error-free.

Different frame loss concealment methods were used to evaluate the quality degradation experienced by users in presence of temporary loss of depth perception. The first concealment method is named Frame-Freeze (FF) and consists in copying the last received frame of the auxiliary view into the subsequent temporal instants corresponding to the lost frames in the same view. In the second method, named Double-Freeze (DF), both the previous received base and auxiliary views are respectively copied into the same corresponding views at the missing temporal instants. This concealment strategy corresponds to freeze the last 3D image received without errors. In the third method, named Base-Copy (BC), the temporally co-located frame from the base view is copied into the auxiliary view to fill the corresponding missing frame at the same time instant. The BC method implements a switching from 3D to 2D because during the error period, both views are comprised of exactly the same frames.

3. SUBJECTIVE QUALITY EVALUATION

This subjective quality evaluation study was conducted with the objective of finding out relevant relationships between perceive 3D quality and content characteristics, packet loss statistics and frame loss concealment methods. Then such relationships should translate into subjective factors that should be taken into account in practical implementations.

Seven stereoscopic sequences with diverse characteristics were subject to packet losses in the auxiliary view where each lost packet implies a lost frame. The test environment and the assessment methodology used in the subjective evaluation process were carefully designed in order to meet most standard requirements. The three frame concealment methods described above were used for different error probabilities and average loss burst.

3.1 3D Video Sequences

The seven 3D video sequences have duration of 12 seconds and resolution of 1024×768@25Hz. No audio track was used in these subjective tests. Among the various types of sequences chosen for this study, each of them exhibits different features, regarding the pictorial content, the motion complexity and depth/disparity range of values.

The main characteristics of the test sequences are presented in Table 1. The disparity shown in the table (Disp) was obtained as a normalized average sum of the absolute pixels displacement between both views of the stereo sequence.

Name	Content description	Disp.
Champ. Tower	A woman stands next to a cup pyramid and handles one of the cups; moderate motion and high depth.	21.5
Panton.	Two clowns moving around near a box; moderate motion and moderate/high depth.	17.7
Kendo	Two men practice kendo with spectators on the back; moving camera; high motion and moderate/high depth.	16.7
Balloons	A man entering on a big balloon while moving; moving camera; intense/complex motion and moderate/high depth.	16.0
Jungle	A toy plane shaking while another toy moves under it; moderate motion and moderate depth.	13.7
Uli	Two men talking; low motion and low depth.	10.4
Dog	A woman playing with a dog in front of a curtain; moderate motion and low depth.	8.8

Table 1 – 3D Video Sequences used for Subjective Evaluation

Two of these sequences (Uli and Jungle) were randomly chosen to be used in the demonstration period described in section 3.3. Thus only the other five sequences were used in the actual subjective evaluation study.

3.2 Test Environment

The subjective quality assessment experiments were conducted in a test room conforming to the viewing conditions defined in BT.500 [10]. The background lighting conditions used daylight color temperature with an intensity of 200 Lux. The room is also equipped with an adequate sound insulation system and it has white walls. A workstation was used to reproduce all the test sequences as well as to register the subjective quality results. A 22-inch computer LCD monitor was used to display the test sequences and the viewer was positioned at a convenient distance as specified below. The viewers could adjust the default viewing distance The LCD monitor used to display the stereoscopic test sequences has a display rate of 120Hz, native resolution of 1680×1050 pixels and a dot pitch of 0,282 mm. The viewers wearing liquid crystal shutter glasses were seated in an arm tray chair and each sequence was evaluated immediately after observation by using a software interface and saving the result on a local database.

3.3 Subjective Assessment Methodology

The Single-Stimulus Absolute Category Rating (ACR) method was used in these experiments [11]. The processed video sequences (i.e., with concealed frames) were divided into three groups, to avoid exposing the viewers to a session that would last more than half an hour [10]. Each viewer evaluated only the sequences of one group, presented in random order, but each sequence was evaluated by the same number of viewers. Each test sequence was played for 12 seconds and afterwards the viewer was asked to judge the 3D perception using a 5 grade scale: 1 (Bad) to 5 (Excellent). The voting period for each sequence was not time-limited. Then the observer was asked to click on a button to proceed to the next sequence. The viewers chair was placed at a distance of eight times the physical height of the picture from the screen, according to the PVD table in [10], but the participants were allowed to adjust their position to the most comfortable viewing distance.

Each evaluation session was comprised of one continuous test session, divided into four parts as shown in Fig. 2. In the adaptation period, the observers stereopsis was first assessed using a random dot stereogram test. This dynamic stereopsis test is suggested on ITU recommendation for stereoscopic subjective assessments [12]. All participants reported to have a normal acuity and all passed the random dot stereogram test. Furthermore, some synthetic and natural 3D sequences were presented to all observers, prior to the testing period, in order to allow them to adjust to stereoscopic viewing.

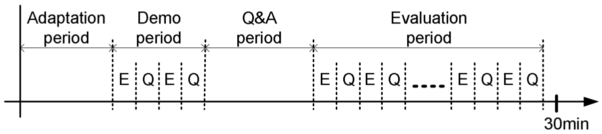

Fig. 2 – Time pattern of a subjective evaluation session

In the demonstration period, two sequences with the temporal artifacts used in our study were shown to each viewer, in order to avoid the influence of any surprising effect in perceptual evaluation and also to practice the assessment method. During this period the viewer simulates the evaluation period by experiencing the 3D viewing (E) and the quality evaluation process (Q). Note that the main purpose of the demonstration period is to familiarize users with the whole system setup and evaluation procedures.

Therefore the actual sequences used in this period are not expected to have any influence in the subsequent evaluation period. A period for questions and answers (Q&A) was then allowed, during which the viewer could pose any questions regarding the test. Finally, in the assessment period the participants observed each sequence (i.e., E periods) and rated the corresponding quality (i.e., Q periods), following the methodology previously described.

3.4 Users

A total of 36 male and 12 female viewers participated in the subjective evaluation, aged from 15 to 55 with an average age of 26.4 years old. Each of the three processed video groups was evaluated by a random group of 16 observers. Before start running the tests the instructions were provided to users in written form and then they could ask any questions regarding to the evaluation process. The total number of viewers conforms to the minimum value of 15 specified by the international recommendations [10]. Participants were randomly recruited among a general public mainly composed by students and faculty staff and relatives. The participants did not have any previous experience in a subjective video quality assessment experience. Furthermore, they did not possess any expertise in video processing and quality assessment.

4. EXPERIMENTAL RESULTS

To find out the relevant factors with subjective impact in the perceived quality of packet 3D video, five stereoscopic video sequences of those described in section 3.1 were subject to different packet loss ratios (PLR). Then missing frames were reconstructed using the concealment methods described in section 2.2. The statistics of the loss patterns generated by the Markov model presented in section 2.1 are shown in Table 2.

	PL Burst Parameters			
PLR (%)	N	Min	Max	Avg
46	11	2	22	7.6
38	10	1	14	6.9
29	7	1	33	6.7
18	4	4	15	8.0
6	2	5	6	5.5

Table 2 – 3D Video Sequences used for Subjective Evaluation

The packet loss burst (PL Burst) is characterized by the number of bursts (N) occurring in the sequence and their minimum (Min), maximum (Max) and average (Avg) burst length. While the average burst length is kept roughly the same for all PLRs, the number of loss bursts and their size variability is quite different. The longer burst (size=33) was generated for PLR=29% and the shorter (size=6) was for PLR=6%.

Taking into account the combinations resulting from the frame loss concealment methods and packet loss probabilities, a total of 80 test sequences were generated and used in the subjective tests. The relevant factors found in this subjective study are discussed in the following subsections.

4.1 Quality Factor Analysis - concealment methods

In this subsection the results regarding the viewers' Mean Opinion Score (MOS) versus the Packet Loss Ratio (PLR) are analysed. Figure 3 shows theses results for each of the five sequences used in the experiments.

In the Figure, it is worthwhile to notice, that for very small error probabilities, all concealment methods present a similar MOS. However, as the number of errors increases (i.e., higher PLR), the BC method rapidly presents the best MOS result, for all sequences, except for the highest disparity sequence *Champagne Tower*. Note that the BC method tends to compromise the stereo effect by switching to 2D viewing during the error period, but maintains the motion information, by copying it from the base view. The results presented in Figure 3 demonstrate that, for low to moderate disparity sequences, the users prefer to lose 3D perception than to endure motion related artifacts, like frame freezing.

One may also notice that in this set of experiments users tend to give higher scores to those sequences that present 30% of missing packets. This can be explained by looking carefully at the error pattern generated in this case. These particular sequences were subjected to longer packet error bursts, which resulted in longer error periods, but also longer periods without errors. This factor not only benefits the BC concealment method because it results in less transitions from 3D to 2D and *vice-versa*, but also the other methods, because the absence of errors during long periods of time. Further studies, with longer sequences, may be required to conclude about this particular factor, i.e., the influence of the length of the error-free period between long periods of error.

4.2 Quality Factor Analysis - disparity/depth

The general viewers' opinion can also be related with other factors. One of the relevant factors in 3D video is the sequence disparity. Previous studies have shown that in the absence of errors, viewers tend to give higher MOS to high disparity sequences [8]. Figure 4 represents the average MOS, for all concealment methods, as a function of the disparity of the corresponding test sequence (as given in Table 1).

Nevertheless, one may also notice that, for high disparity values, the MOS of the sequence processed with BC severely decays. This is because users tend to give higher significance to errors in sequences where the 3D quality effect (i.e., depth quality perception) is subjectively higher.

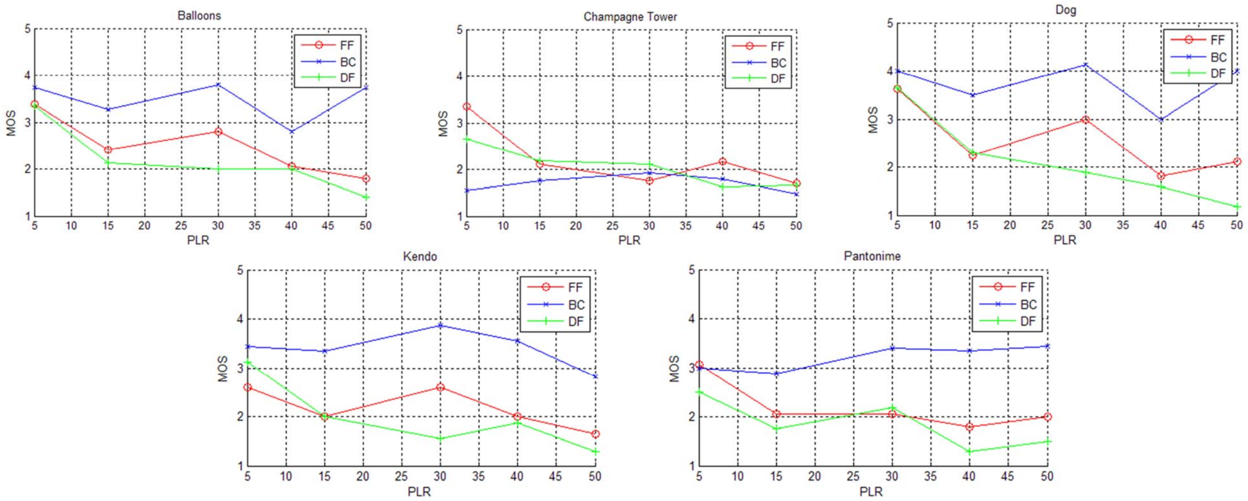

Fig. 3 – MOS as a function of the PLR for each of the concealment methods and test sequences

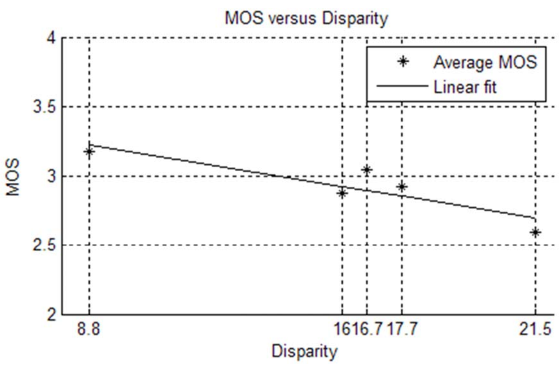

Fig. 4 – MOS as a function of sequence disparity

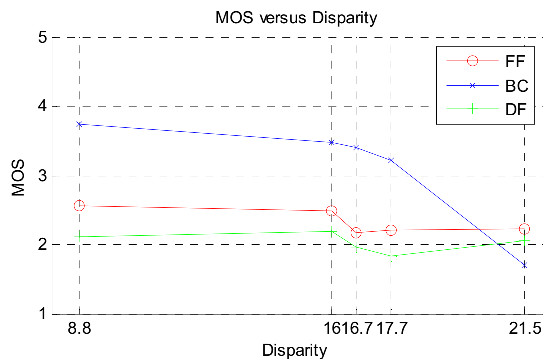

Fig. 5 – MOS as a function of sequence disparity, for each concealment method

In Figure 4, one may observe that, in the presence of random errors, viewers tend to change their initial evaluation (i.e. higher score given to higher disparity sequences) and give higher scores to low disparity sequences. This means that viewers better withstand 3D errors that have smaller stereoscopic visual impact. These results show that for high disparity sequences, the stereoscopic errors are more noticeable, penalizing the MOS Figure 5 presents a more detailed analysis. The MOS is represented as a function of the disparity of each sequence, for each of the concealment methods implemented in this study. Again, one may observe the advantage of the BC concealment method which is able to outperform the FF and DF schemes for low to medium disparity sequences. In other words, losing 3D quality in sequences with low perceived depth is less annoying than in sequences with high perceived depth.

4.3 Overall results

Figure 6 presents the MOS obtained for each concealment method, for all the test sequences, as a function of the packet loss ratio (PLR). One may notice a clear advantage of the BC error concealment scheme, for all error probability values. Figure 6 also shows that, for increasing values of packet error probability, the results of the FF and DF methods tend to decay. On the contrary, the Mean

Opinion Score for BC concealed sequences tends to maintain a more stable value. This fact further demonstrates the advantage of using BC for error concealment in the tested scenarios. A common observation in all results is that MOS is not monotonic with PLR. As pointed out in section 4.1 this is due to the burst length distribution is these particular set of experiments. Note that the concealment methods used in this subjective study are basic schemes based on the replacement of one entire frame when one frame is missing at the receiver. Therefore, the absolute MOS obtained with better frame concealment methods (e.g., using spatio-temporal and interview information) are expected to be better than those obtained in this study. However, their relative behavior and the main factors presented in this paper should remain valid for other concealment methods because the perceptual experience is the same, though this effect might be smoothed using better concealment methods.
Therefore these experimental results can be considered the lowest ground truth measure of the expected performance of more sophisticated receivers. Similar results and conclusions were recently found in [13], which also validates the results of this paper.

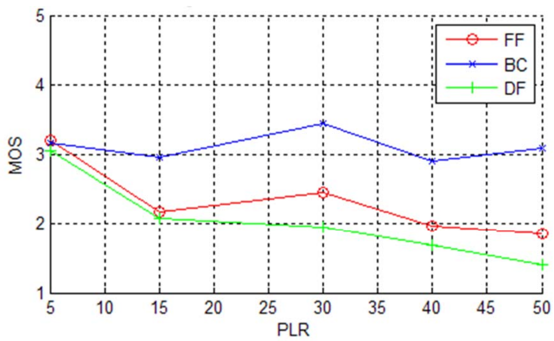

Fig. 6 –MOS as a function of the PLR

Acknowledgements:
This work was supported by IPLeiria (Project P3DTV) and FCT (Project PTDC/EEA-TEL/114487/2009).

5. CONCLUSIONS

A subjective evaluation study was carried out and its main results were presented and discussed. In the presence of missing frames at the receiver, this paper provides relevant conclusions about two main factors influencing the 3D quality experienced by users: the disparity of the original sequence and the concealment method used at the receiver. An error concealment method that switches 3D viewing to 2D was found to be acceptable for all cases of missing frames in the auxiliary view, particularly for long error bursts. Moreover, it was found that users' opinion about 3D quality is dependent on the disparity of the sequence. The higher the disparity the worse the MOS, in presence of missing frames. In general, users penalize those sequences that originally provide better 3D experience. Overall, the results and conclusions presented in this paper provide relevant insight to deal with packet loss in 3D video streaming.

6. REFERENCES

[1] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services", ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2010.

[2] Gozde B. Akar, A. Murat Tekalp, Christoph Fehn, M. Reha Civanlar, Transport Methods in 3DTV—A Survey, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No 11, pp. 1622-1630, Nov. 2007.

[3] Aksay, A., Pehlivan, S., Kurutepe, E., Bilen, C., Ozcelebi, T., Akar, Gozde, Bozdagi, Civanlar, M. Reha, Tekalp, A. Murat, End-to-end stereoscopic video streaming with content-adaptive rate and format control, Signal Processing: Image Communication, Vol. 22, Issue 2, Feb. 2007.

[4] Chaminda, T.E.R. Hewage, Stewart T. Worrall, Safak Dogan, Stephane Villete, Ahmet M. Kondoz, Quality Evaluation of Color Plus Depth Map-Based Stereoscopic Video, IEEE Journal of Selected Topics in Signal Processing, Vol. 3, No 2, pp. 304-318, April 2009.

[5] Hewage, C.T.E.R.; Worrall, S.; Dogan, S.; Kodikaraarachchi, H.; Kondoz, A.M.; , "Stereoscopic TV over IP,"4th European Conference on Visual Media Production, (IETCVMP), pp.1-7, 27-28 Nov. 2007.

[6] Shujie Liu; Ying Chen; Ye-Kui Wang; Gabbouj, M.; Hannuksela, M.M.; Houqiang Li; , "Frame loss error concealment for multiview video coding,"IEEE International Symposium on Circuits and Systems (ISCAS), pp.3470-3473, 18-21 May 2008

[7] Stankiewicz, O.; Wegner, K.; Domański, M.; , "Error concealment for MVC and 3D video coding," Picture Coding Symposium (PCS), pp.498-501, 8-10 Dec. 2010.

[8] Carreira, J.; Pinto, L.; Rodrigues, N.; Faria, S.; Assuncao, P.; , "Subjective assessment of frame loss concealment methods in 3D video," Picture Coding Symposium (PCS), pp.182-185, 8-10 Dec. 2010.

[9] Hohlfeld, O.; Geib, R.; Hasslinger, G.; , "Packet Loss in Real-Time Services: Markovian Models Generating QoE Impairments,", 16th International Workshop on Quality of Service (IWQoS), pp.239-248, 2-4 June 2008.

[10] ITU-R, Recommendation BT.500-12, Methodology for the subjective assessment of the quality of television pictures, 2009.

[11] ITU-T, Recommendation P.910, Subjective video quality assessment methods for multimedia applications, 2008.

[12] ITU-R, Recommendation BT.1438, Subjective assessment of stereoscopic television pictures, 2000.

[13] Barkowsky, M.; Wang, K; Cousseau, R.; Brunnström, K.; Olsson, R.; Le Callet, P.; , "Subjective quality assessment of error concealment strategies for 3DTV in the presence of asymmetric transmission errors,"18th International Packet Video Workshop (PV), pp.193-200, 13-14 Dec. 2010.