

# QoE Enhancement for Stereoscopic 3D Video Quality Based on Depth and Color Transmission over IP Networks: A Review

Wahiba Ismaiel<sup>a\*</sup>, Niema Osman<sup>b</sup>, Is-Haka Mkwawa<sup>c</sup>

<sup>a,b</sup>*Sudan University, Khartoum and +249183, Sudan*

<sup>c</sup>*University of Plymouth, Plymouth, UK*

<sup>a</sup>*Email: whiba2008123@yahoo.com*

<sup>b</sup>*Email: niema.osman@gmail.com*

<sup>c</sup>*Email: Is-Haka.Mkwawa@Plymouth.ac.uk*

## Abstract

In this review paper we focus on the enhancement of Quality of Experience (QoE) for stereoscopic 3D video based on depth information. We focus on stereoscopic video format because it takes less bandwidth than other format when 3D video is transmitted over an error channel but it is easily affected by the network parameters such as packets loss, delay and jitter. The packet loss on 3D video has more impact in the depth information than other 3D video factors such as comfort, motion, disparity and discomfort. The packet loss on depth information causes undesired effect on color and depth maps. Therefore, in order to minimize quality degradation, the application of frame loss concealment technique is preferred. This technique is expected to improve the QoE for end users. In this paper we will also review 3D video factors and their challenges, methods of measuring the QOE, algorithms used for packets loss recovery.

**Keywords:** 3D video; Stereoscopic; QoE; Color-plus-Depth; error concealment.

---

\*Corresponding author.

## **1. Introduction**

Multimedia communication field plays important role in the network technology revolution and future applications. Multimedia applications are used in several fields such as business, education, entertainment, home and public places. These applications add additional challenges especially when they are transmitted over IP network because they contain large volume of information.

In recent years three-dimensional video (3DV) demand has increased in multimedia applications. 3DV plays a great role in services such as 3D cinema, 3D game and free viewpoint. Stereoscopic color-plus-depth is now more popular compared to the Left and Right views, therefore, devices for 3D stereoscopic are manufactured in a large numbers .It is a preferred for stereoscopic 3D video transmission because it requires less bandwidth [1].

This study presents Quality of Experience (QoE) for 3D video because it is an important aspect for end users. When transmitting video services over an IP network, challenges such as packet loss, limited bandwidth and traffic variation are inevitable and they affect 3D video quality.

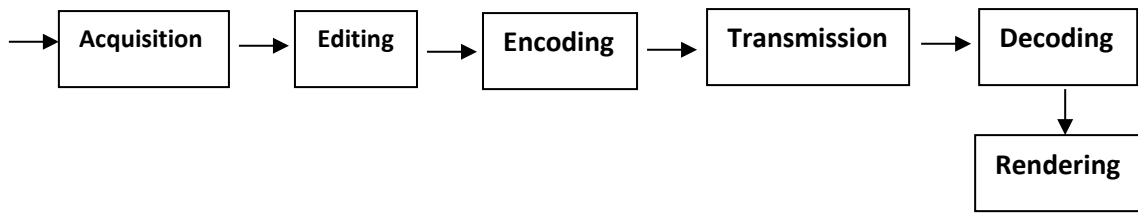
This survey focuses on how to improve stereoscopic quality caused by packets loss. We discuss some studies that enhance stereoscopic transmission in order to achieve better quality with concealment methods. This paper also reviews 3D video subjective and objective quality assessments. Some recommended methodologies to assess subjective quality stereoscopic video sequences are ITU-T BT.2021[2] and ITU-R BT 1483[3].

This paper is organized as follows: Section 2 provides presents an overview of 3D video. Section 3 introduces 3D video transmission, transmission protocols, transmission challenges and error resilience tools. Section 4 describes stereoscopic format. Section 5 presents 3D video quality by addressing subjective and objective methods. Section 6 discusses state of the art for frame concealment for 3D video stereoscopic color-plus-depth. The last section concludes the review.

## **2. 3D video**

3D video displays from left and right views which add sensation of the depth by adding the depth dimension in order to solve complexity in 2D video viewing and some features such as (free view point and depth-controlled object insertion) [4].

3D video technique provides us with new services and applications today such as 3D Game, 3DTV, 3D animal picture books, sport performance analysis, 3D surgery monitoring, 3D archive of traditional dances [5], video personal recording, 3D telepresence, 3D telemedicine, 3D surveillance and free point television (FTV) [6]. 3D video processing chain starts from 3D video production to display it on user devices. Figure 1 shows the processing steps [7].



**Figure 1:** 3D video processing chain

- Acquisition is the first phase in 3D video communication chain which includes two stages, capturing and post-processing. The capturing stage requires setting up 2 or 3 cameras according to different 3D video formats. Multi camera system can be used but will face challenges such as temporal synchronization, geometrical calibration and color balance between the individual cameras [8]. Other things that can be taken into account when setting up camera are position and angle of camera, camera parameters and format [7]. The post-processing stage contributes to making color balancing between cameras by enabling the camera views to be aligned [1].
- The second phase is editing 3D video stream, this can use different types of editing operation such as color, texture, shape, and motion editing [5].
- The third phase is encoding, this phase requires compression techniques to represent 3D video in a compressed form without any loss of raw video information because 3D video needs to be packetized and transmitted over a limited bandwidth. In this operation, representation format has an effect on all phases in 3D video communication chain [5,8].
- The fourth phase is 3D video transmission, the transport format is important in transmitting 3D video content because 3D video faces different transmission challenges over the IP network. The presence of depth maps and multiple views affects the video reconstruction and end user QoE. They are also making different quality between left and right views [1,9].
- The fifth phase is decoding, it converts signals into raw 3D video for the left and right views, but it faces some problems such as block noise and reduced spatial resolution which influence the quality. This phase uses frame sequential format in order to avoid reduced spatial resolution problems [7].
- The sixth phase is rendering, this is the last phase in processing 3D video chain. In this phase user receives 3D video content. The transport format data must be converted into display format [8]. The display format is very important for 3D video output on the device, in addition to other requirements of displaying 3D video which are display type, glasses, monitor size, luminance and viewing distance/angle [7]. On the other hand, it causes a number of physiological cues such as binocular disparity, motion parallax, and ocular convergence in order to produce depth sensations in human visual system (HVS) [1].

### 3. 3D video transmission

The delivery of 3D video content has become one of the most important issues in recent years. Therefore, 3D video sequences sent over network channels would require mechanisms and concealment strategies to achieve end to end acceptable quality.

### **3.1 3D video transmissions over the IP Network**

Transmission of 3D video over the IP network uses protocol stack RTP/UDP/IP encapsulated video streams. In recent years, modern 3D video applications use the newest protocol rather than UDP which is named as Congestion Control Protocol (DCCP). DCCP has congestion control method that computes transmission rate by using TCP throughput equation. DCCP is suitable for services which have large size of data [10]. Authors in [11-12] have discussed DCCP in detail.

### **3.2 Challenges of transmitting 3D over IP networks**

The following part discusses some of the challenges of transmitting 3D video over IP networks.

- Packet losses: are caused by network congestion or faulty in wire and wireless link [1,9].
- Bit error or Burst error: it happens by noisy and multi-path propagation. Multi-path propagation has more effects in mobile environment [9]
- Temporal domain de-synchronization aspect: it is caused by delay in one view, this also causes missing frame and discomfort in 3D viewing [1]
- Error on reconstructed video propagation: this error propagates from one frame to another due to prediction mechanism.
- Error concealment method: error concealment mechanism is a very important aspect for 3D video transmission in order to mitigate error propagation. Error concealment techniques are limited to errors resulting from compression and codec [9].

### **3.3 Error resilience tools**

Four types of error resilience tools are considered as follows [9],

- Localization: It removes or decreases the spatial or temporal dependency between frames or slices in order to stop the propagation of errors.
- Data partitioning: Frames can be merged or fragmented according to their importance if error in low priority data happens at the encoder side, while at decoder side the corresponding packets are removed in order get better group of highest priority received without errors.
- Redundant coding: This tool provides flexibility at the decoder side by encapsulating the additional data with bit streams which results from duplicating primary slices as redundancy slices or Multiple Description Coding (MDC) or Reversible Variable Length Coding (RVLC).
- Concealment driven coding: This tool can be used at the encoder and decoder; it will recover errors by utilizing the additional received information about general behavior of the lost data.

## **4. Stereoscopic format**

Stereoscopic 3D (S3D) video is more popular in presenting 3D video because it is simple. It consists of two views, left and right view which are received by left and right eyes of an observer, it depends on stereopsis

principle. It also has another representation called color-plus-depth, which has some advantages such as flexibility, low encoding cost and backwards compatibility compared to left-right stereoscopic video. S3D has some advantages compared to other 3D video representations. These advantages are shown below,

- Easy data acquisition
- More accuracy
- It requires less storage and bandwidth
- It is suited for applications which contain passive systems based on polarized glasses or active systems based on liquid crystal shutter glasses [13-14].

## 5. 3D video QoE

This section explains some concepts of QoE such as QoE assessment, factors which have impact on 3D video quality.

### 5.1. QoE Assessment

Quality of Experience is an important issue for assessing 3D videos when transmitted over IP network. QoE means user perceptual of the performance of services [15]. The ITU-T recommendation P.10/G.100 defines QoE as “The overall acceptability of an applications or services, as perceived subjectively by the end-user. QoE includes the complete end-to-end system effects (client, terminal, network and services infrastructure) and overall acceptability may be influenced by user expectations and context”. Figure 2 shows QoE aspects which are technical factors that consist of the network performance, device characteristics and content representation format, the social and psychological factors containing prior experience, content preference, user expectations and environmental factors [16-17].

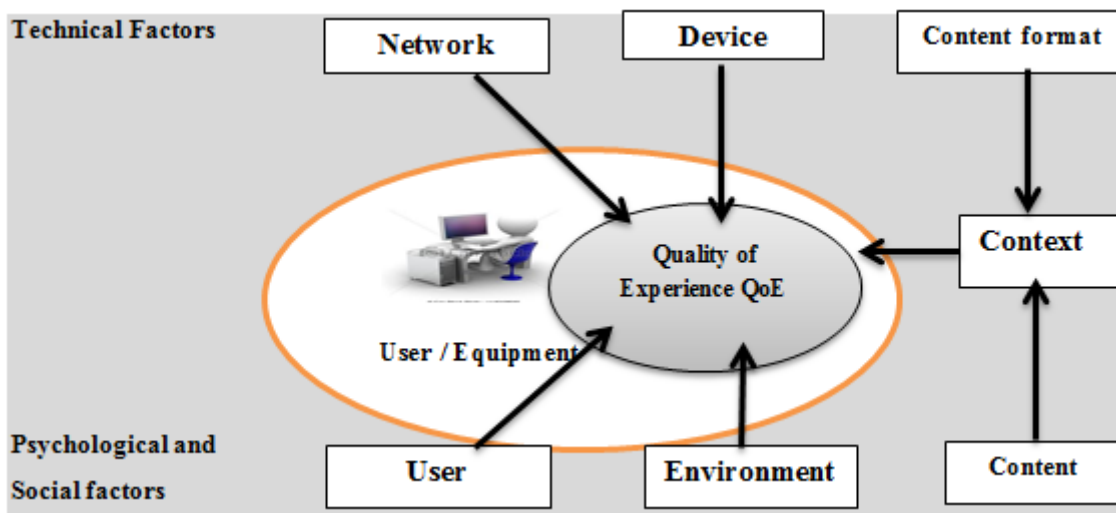


Figure 2: QoE aspects

QoE is also defined by The European Telecommunications Standards Institute (ETSI) as "A measure of user

performance based on both objective and subjective psychological measures of using an ICT service or product" [18]. QoE has two methods to assess the video sequences which are subjective and objective methods. Video Quality Experts Group (VQMG) is a basic group of validation and measuring objective perceptual quality metrics. Their working results appear through International Telecommunication Union (ITU) standardization which is designed for the standard definition of television and for multimedia applications [19].

### **5.1.1 The subjective method**

The ITU-T recommendation P.910 [20] recommended and contributed on assessing the video quality by providing four types of methods,

- **Absolute Category Rating (ACR):** Test sequences are shown one at a time and are rated independently on category scale. The time duration is 10 seconds and the scale is ranging from 5 to 1 (5 Excellent, 4 Good, 3 Fair, 2 Poor and 1 Bad).
- **Absolute Category Rating with Hidden Reference (ACR-HR):** Test sequences are presented one at a time and are rated separately on category scale. In this method, viewers do not notice the presence of the source sequence. The scale ranges from 5 to 1 (5 Excellent, 4 Good, 3 Fair, 2 Poor and 1 Bad).
- **Degradation Category Rating (DCR):** Test sequences are shown in pairs, one is source reference and the second is degraded sequence (also known as the Double Stimulus Impairment Scale Method (DSIS)). This method has five levels scale for rating the impairment (5 imperceptible, 4 perceptible but not annoying, 3 slightly annoying, 2 annoying and 1 very annoying). This method also uses explicit references.
- **Pair Comparison (PC):** Test sequences are shown in pairs consisting of the same sequences and the general pair combined when the system under test are equal to  $n(n-1)$ . The scale in this method has seven levels for rating the impairment which are: -3 Much worse, -2 Worse, -1 Slightly worse, 0 The same, 1 Slightly better, 2 Better, 3 Much better [21].

In all the methods above, the average quality result from viewers is recorded by using one of two ways, Mean Opinion Score (MOS) and the Difference Mean Opinion Score (DMOS). The DMOS means that the mean of differential subjective scores (mean and the standard deviation are used).

### **5.1.2. The objective method**

This was classified into three methods, Full Reference (FR), Reduce Reference (RR) and No Reference (NR). The ITU-R Recommendation ITU-T J.247 (08/2008) represents the FR model which means comparing the original video sequence with degraded video sequence [22]. However, recommendation ITU-T J.246 (08/2008) defines RR by comparing the extracted features of original video sequence with degraded video sequence [23], but the NR does not need the original sequence, it only needs to assess the degraded sequence [24]. Objective assessment methods are classified into five categories according to the type of input data which is used to assess video quality.

1. **Media-layer model:** It uses speech or video to give QoE
2. **Parametric packet-layer model :** It depends on packet-header information to give QoE, it doesn't require media signal
3. **Parametric planning model:** It takes QoE for networks and terminals depending on prior information
4. **Bit stream layer model:** It takes information of encoded bit stream and information of packet-layer to achieve QoE
5. **Hybrid model:** It predicts QoE by merging other models

## 5.2. 3D video QoE

Three aspects affect QoE on stereoscopic system such as visual comfort, depth quality and picture quality. ITU-R recommendation BT.2021 defined QoE 3D video by using some factors such as visual discomfort, depth perception and visual quality. Some researchers have different opinion for psychological impairment of stereoscopic imaging when assessed in general concepts such as naturalness and sense of presence [2]. The 3D video QoE factors were considered in [25] as general factors which have impact on 3D video services and are divided into three types:

- Content production factors which include (camera performance, video shooting conditions, effects of the picture frame and magnitude/intensity of convergence change).
- Video compression which include (transmission error, coding algorithm and format)
- Delivery factors and viewing environment factors which include (viewing distance, viewing angle and display scheme).

### 5.2.1. Subjective 3D video methodology



There are standards and recommendations that define subjective quality assessment of 3D video such as International Tele-communication Union Radio-communication Standards Section (ITU-R) BT 1438, ITU-R BT 500-1, and ITU Telecommunication Standards Sector (ITU-T) P.910.

The (ITU-R) BT 1438 evaluated subjective quality according to factors that assess stereoscopic television system which are: resolution, color rendition, motion portrayal, overall quality, sharpness and depth [3]. Subjective methodology can be used to assess stereoscopic picture and it is recommended in ITU-R BT 500-12. All these methods relate to the depth quality, picture quality and visual comfort of stereoscopic imaging systems. These methods are explained in the Table 1 in terms of depth quality.

### 5.2.2. Objective 3D video Methodologies

Objective assessment of 3D video is not simple compared to 2D video. 3D video has additional factors such as (depth perceptual and image quality), therefore, in several researches, conventional metrics were used to assess 3D video such as PSNR, VQM and SSIM while other researchers developed 2D video metrics. New methods are proposed in Section (5.3.2).

**Table 1:** Subjective methods for the assessment of depth quality

Mode of presentation	Sequence duration	Binary scale	Discrete scale	Continuous scale
Single-stimulus (SS) methods : In this method just degraded sequence appears and the subjects are asked to assess the quality	~10 s		5 Excellent 4 Good 3 Fair 2 Poor 1 Bad	
Double-stimulus impairment scale (DSIS) method or European Broadcasting Union(EBU): It displays multiple sequences containing source reference and degraded <b>sequence and viewers are asked to assess the quality</b>			5 Imperceptible 4 Perceptible, but not annoying 3 Slightly annoying 2 Annoying 1 Very annoying	
Double-stimulus continuous quality scale (DSCQS) method: It is used to assess the quality in terms of sharpness and overall depth of stereoscopic and monoscopic image streams [Rec BT.1438], in this method volunteers are showed a pair sequence twice then they rate the quality of each sequence in the pair at the second presentation.	~10 s			
Stimulus-comparison (SC) methods: Is the comparison of two degraded sequences against each other.	~10 s	A vs. B	-3 Much worse -2 Worse -1 Slightly worse 0 The same 1 Slightly better	



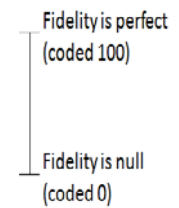
2 Better

3 Much better

Single-stimulus continuous quality evaluation (SSCQE) method: This method is using slider mechanism with an associated quality score and shows the degraded sequence only; the subjects rate the quality sequence. ~3-5 min



Simultaneous double stimulus for continuous evaluation (SDSCE) method: It shows the reference sequence and degraded sequence together, subjects rate the quality of the sequence but it requires viewers' attention on the left and right squence.



### 5.3. Subjective and objective evaluation for stereoscopic depth information

This section discusses several researches which are using methodologies for 2D video and other new methodologies to assess subjective and objective quality for stereoscopic depth information.

#### 5.3.1. Subjective for stereoscopic depth information

This section explains methods developed to assess subjective quality of 3D video. Some methods use 2D video quality assessment and others were only developed for 3D video

Lebreton and his colleagues [26] evaluated the quality of distortion of depth perceptual 3D video stereoscopic which is caused by coding or transmission. In this study, the assessed depth included two aspects, depth quality and depth perception. Authors explained subjective experiment and proposed new objective model to assess the quality based on depth information. It also measured the interaction between binocular and monocular depth features and depth perception. Subjective tests environment was based on standardized ITU-R BT.500-12 recommendation. The experiment consisted of 64 source reference single (SRCS) of database content, which has different types of scenes, different values of (image quality, depth quality and visual discomfort). All sequences are full HD stereoscopic videos, high quality, slow or fast motion, each view has resolution 1920x1080, frame rate of 25, each sequence has 10 seconds length, viewing distance was set at 3H and the number of viewers was 24. Authors used ACR for quality rating; the depth was rated on five grades which were denoted to "very high", "high", "medium", "low" or "very low." While the evaluation of visual comfort was labeled as "much more", "more", "as", "less", "much less" and "comfortable than watching 2D video". The subjective tests took 50 minutes including training session. The results illustrated high correlation with the HVS.

Umar and his colleagues [27] designed subjective method for 3D video human perception by using depth information. Authors designed questionnaires to gather data and assist in reaching some information about requirements, expectation, viewing comfort, needs, perception, preference, satisfaction of 3D user and overall 3D video quality. Authors used two datasets in the experiment. First dataset from at Ecole Polytechnique Federale de Lausanne EPFL [53], it has resolution of 720x576 pixels, encoded using 3D MDC-SIMI with QP=5, sequence of 10 seconds long. They included stereoscopic images and 250 frames at 25 fps. The second dataset from visual media research group at Microsoft research [54] with resolution 1024x768 pixels with 15 fps, 150 frames and duration of 10 seconds. But authors didn't specify what codec was used to encode the sequences and what 3D video format was used. Authors used 4x6 images for training, 3x6 for testing and 2x6 for validation and the subjects who participated in the assessment were 25 volunteers aged between 15 and 50 years old. The average age was 23 years old. There were 18 males and 17 females, all were non experts, however, authors didn't explain their educational levels and eyes condition. In the experiment, stereoscopic shutter glass, the screen resolution of 1152x900 pixels, frequency of 60 Hz x 60Hz, 2cm of viewing distance, assessment size of 360 x 270 mm, optical path length of 320 mm, the horizontal FOV of 0.8 rad and 17 inches laptop monitors were used for display. Volunteers were asked to evaluate the overall 3D video quality and overall perception (0-100) values according to ITU-R Recommendation (Excellent, Good, Fair, Don't Know, bad), (High impact, moderate impact, less impact, Don't know, No impact), (Very important, Important, Less important, Don't know, Not important) and (Strongly Agree, Agree, Disagree, Don't know, Strongly disagree). The experiment took 3 hours and each questionnaire took around 10 minutes. The authors computed the MOS using equation (1), where N denoted number of volunteers,  $m_{ij}$  is score by volunteer i of test condition j, and estimated mean value by using confidence interval is given in equation (2).

$$MOS_j = \frac{1}{N} \sum_{i=1}^N m_{ij} \quad (1)$$

$$C_{ij} = \left( \left( \frac{t(1-\alpha)}{2}, N \right), \frac{\sigma_j}{\sqrt{N}} \right) \quad (2)$$

where  $\frac{t(1-\alpha)}{2}, N$  is the *t-value* corresponding to a two-tailed t-Student distribution with N-1 degree of freedom and a desired significance level  $\alpha$ . The results showed the confidence interval of 0.05, this value corresponds to a degree of significance of 95%, and the results ensure the subjective test was well done. Authors used dataset from public dataset for 3D video. The subjective tests were based on ITU-Recommendation.

Mysirlidis et al., [28] measured the impact of asymmetric packet losses on scalable Multiple Description Coding (MDC) and delivery of 3D video to the receiver over two different paths, MDC combined with Scalable Video Coding (SVC) to create two layers using Medium Gain Scale (MGS). MDC creates two sub video with half of the original spatial resolution from color plus depth format. The result illustrated that the MDC scheme performs better by using a single description with less packet loss than using double description. The subjective tests were based on Double Stimulus Continuous Quality Scale (DSCQS). However authors didn't give any details of subjects such as the number of viewers, viewers' gender and age. They also didn't show the bit rates used for transmission.

Kim et al., [29] presented a new approach which was called multimodal Interactive Continuous Scoring of quality (MICSQ) which they claimed to be better than the traditional method (Single Stimulus Continuous Quality Evaluation (SSCQE)). The aim of MICSQ was to decrease deviations between viewing and evaluation, and to facilitate visual comfort for the end user in order to assess the quality of the 3D video and improve the subjective evaluation. This was achieved by conducting several related experiments to illustrate the usefulness of the new method such as high reliability, multi user assessment, speed and direction of depth motion, comfortable viewing Zone (CVZ) and Artificial Depth of Field (DOF) naturalness. MICSQ used stereoscopic format and subjective score was set to 0-10 by using two ways. The first is between assessment tools (interaction devices) which are used to display 3D using wireless network protocol. The second is between viewers (human interaction) and tablet, single or multi viewers can be assessed at the same time. The proposed methodology results were better for human concentration and visual effort than traditional methods (SSCQE) in terms of disparity and motion. The authors found relationship between visual comfort for motion and disparity and the relationship between DOF and naturalness.

### ***5.3.2. Objective for stereoscopic depth information***

Hossein et al., [30] presented new Reduce Reference (RR) metric of stereoscopic color plus depth for 3D video transmission. The idea was to merge the extracted edge information with spatial neighboring information to construct the proposed Grey level Co-occurrence Matrix (GLCM) metric. This metric is more efficient because it used GLCM in real time stereoscopic quality assessment, it requires less bandwidth (2 Kbps). Consequently, 725 videos were generated at 25 fps, the experiment used 29 references and subjective tests based on ITU guideline for experiment. Results showed that the proposed metric highly correlated with subjective quality score. However, the proposed metric needs more computation in all frames for both original and compressed video.

Hewage and his colleagues [31] proposed a Reduced-Reference (RR) quality metric for compression depth maps associated with color plus depth 3D video according to edge detection which was divided into two sections. The first is gradient explore edge detection based on the minimum and the maximum in the first derivative of the image. The second is Laplacian at the location of zeros in the second derivative edges detection. All RR and Full-Reference (FR) use PSNR metric, the result showed that at the minimum of Quantization Parameter (QP), the RR is close to the FR but RR is not accurate at higher QP. The strength of the proposed method is that it takes less bandwidth than FR. However, PSNR is not good for QoE and does not give information about the depth perception. If subjective tests were done, they would reveal more realistic results.

Chaminda and his colleagues [32] developed RR quality metric for color plus depth 3D video transmission. This method depends on the use of extracted edge information in the depth map. It used RR D-PSNR Metric to map the depth, and for color extracted edge information from the corresponding color image near the edge, it used RR C-PSNR. The strength of the method is that it reduces overhead (by decreasing bit rate) and bandwidth utilized compared with Full Reference (FR) method. Edge information can be used to describe the basic in front object. However, the authors used the bitrates without clarifying what bit rates range has been utilized.

Joveluro and his colleagues [33] provided the accurate two dimensional (2D) objectives metric related with the Human Visual System (HVS) for evaluation 3D video quality. Using stereoscopic format which encodes the color and depth image by the Content Adaptive Binary Arithmetic Coding (CABAC) scheme in the Joint Scalable Video Model (JSVM) and representing left and right views by using Depth-Image-Base-Rendering (DIBR) before evaluating 2D metrics. The study powerfully presents the subjective method based on standard ITU-R WP6Q using SAMVSQ. However, the experiment didn't determine the viewing distance. The proposed metric Perceptual Quality Metric (PQM) model has strong correlation when comparing their outcome with the result of Video Quality Metric (VQM) and its design is based on human visual system (HVS). This metric can be used for 2D and 3D quality. However, it is based only for stereoscopic measure in luminance component color and depth.

Politis and his colleagues [34] suggested a metric to assess compression and packet loss. Two objective metrics were proposed, Video Quality Metric (VQM) base on ANSI T1.801.03 for perceived quality of stereoscopic. However, it didn't assess Disparity factor. The metric VQMComp is taking weighted maximum and minimum VQMs of the stereoscopic views to assess compression quality using equation (3). VQM left denotes left view and VQM Right denotes right view. The authors got the best weight from merging the minimum and the maximum weights of left and right views. The second metric VQM<sub>pl</sub> measures packet loss using equation (4), in this equation authors selected three weight factors which are A, w and z. The value of A is 0.322 and w =z=0.072. These values are the result of linear regression of equation (4). The factors w and z are based on the Unequal Error Protection (UEP) strategy, the purpose of that is to decrease the impact on one of the two views according to the following rule,

- w =z, in case of using High Priority to I-frames
- w<z, in case of using High Priority to Left-view

The subjective tests based on standardization by using DSCQS method and RTP packets were created by using Single Network Abstraction (NAL) encoded by CABAC in JSVM. The results show VQM<sub>comp</sub> metric is close to the subjective scores and performed better than the average VQM and VQM<sub>pl</sub>

$$VQM_{Comp} = 0.375. \max(VQM_{left}, + VQM_{right}) + 0.125. \min(VQM_{left}, + VQM_{right}) \quad (3)$$

$$VQM_{PL} = A \left[ \frac{w. VQM_l + z. VQM_r}{2} \right] \text{ and } A \in (0,1) \quad (w, z) > 0 \quad (4)$$

Yi Han and his colleagues [35] presented No Reference objective metric Video Quality Metric (NVQM) to 3D video stereoscopic real-time evaluation. It is based on the model in ITU\_T G.1070 for 2D video assessment to evaluate the packet loss and various bit-rates. The model contains five sets of coefficient for MPEG-4 and ITUT H.264 and authors utilized one of the five set for MPEG-4 to build their model. The proposed metric was

compared with two FR metrics which are VQM and SSIM. This model used fixed frame rate explained by the equation (5). The coefficient  $a_1$  denoted the different MOS of the perceived 3D video quality,  $a_2$  is a fixed coefficient,  $P_{plv}$  is the packet loss rate and  $Br_v$  indicates bit rate. However, the coefficients from  $a_1$  to  $a_5$  are utilized from subjective test results. In the experiment, authors used different five video sequence with different motion (low, medium and high) and different packet loss such as (0%,0.5% ,1%,2%,3%,4%,5%,6%,7%,10% ).

$$V_{3Dq} = a_1 + a_2 * e^{-\frac{P_{plv}}{a_3+a_4 * e^{-\frac{Br_v}{a_5}}}} \tag{5}$$

They used fixed frame rate of 18 fps and low and high bit-rate of 2 and 4 Mbps, respectively, with resolution of 1920x1080 pixel. Viewers used shutter glasses, 5mx5m quite room, distance between screen and viewers 1m and video duration from 6 to 14 seconds. Volunteers assessed quality ranging from 1 to 5 (bad, poor, fair, good and excellent). The subjective results illustrated that the MOS values were better at high bit-rate than low bit-rate and when PLR was less than 4%. However, when PLR was from 0 to 1%, the MOS values for both low and high bit-rate decreased by more than 1.0 until 2% PLR and after 4% PLR it decreased to 1.5. The proposed model achieved better result when compared with SSIM, VQM and G.1070 (c.f., Table 2).

**Table 2:** Proposed method results compared to other metric

	SSIM	VQM	G.1070	NVQM
<b>High bit-rate</b>	0.8712	0.7885	0.9059	0.9736
<b>Low bit-rate</b>	0.7474	0.8892	0.8606 ,	0.8961

The proposed model (NVQM) achieves high accuracy when compared to SSIM, it has 11% and 19% for high and low bit-rate and it has 23% for high bit-rate compared to VQM, and 5% with G.1070. However, authors did not explain if it can be used for stereoscopic color plus depth.

Han and his colleagues [36] developed a metric called extended No-reference 3D Video Quality Metric (eNVQM) based on packet loss rate, 3D video bit rate and 3D video frame rate. It aimed to assess stereoscopic 3D video quality including color and depth information. This model used three input parameters. It used most of the coefficients of the ITU-T G.1070 similar to 2D video. Authors used equations (6-8). In equation (6) the  $I_{coding}^{3D}$  includes frame rate and bit-rate  $t$  when packet loss ( $P_{plv}$ ) is 0%.

$$V_q^{3D} = 1 + I_{coding}^{3D} e^{-\frac{P_{plv}}{D_{plv}^{3D}}} \tag{6}$$

$$I_{coding}^{3D} = a_1 \ln Fr_v + a_2 \ln(a_3 + Br_v) \tag{7}$$

$$D_{Pplv}^{D3} = a_5 + a_6 * e^{-\frac{Fr_v}{a_1}} + a_8 * e^{-\frac{Br_v}{a_9}} \tag{8}$$

For quality of color and depth authors used equations (6) and (8) denoted for  $v_{color}^{D3}$  and  $v_{Depth}^{D3}$  and two set of coefficients  $A = \{a_1, a_2, \dots, a_{12}\}$  and the  $A_{color}$  and  $A_{depth}$ . However,  $a_1$  denoted the impact of frame rate and  $a_2$

denoted the impact of bit-rate when there is no packet loss. All coefficients from  $a_1$  to  $a_9$  are based on three parameters, video format, codec type and display size. In the experiment authors used frame rates of 18 fps and 11 fps, the bit rates with high (4 Mbps), medium (3 Mbps), and low 2 Mbps. The video was encoded into MPEG-4 SVC. Authors used different duration from 6 to 12 second. In this experiment different PLRs were used (0% ,0.1% , 0.5% , 1% , 2%, 3%, 4% ,5%, 6% ,8% ,10% ). The subjective test contain 10 set of video sequences chosen in order of video content, packet loss, bit rate and frame rate. A set included 33 video sequences with the resolution of 1920x1080 pixels. 40 volunteers assessed the quality over all 3D experience, 3D depth experience, eye comfort, 3D enjoyable level and 3D effect enhancement level buy using MOS values. The experimental results showed that at packet loss 0%, 1% and 3% MOS values were higher when bit rate and frame rate were increased and increasing bit rate had more effect than increased frame rate. The study was based on standard ITU-T G.1070 and compared to two previous studies. It achieved better result of 0.944 than SSIM 0.911 and VQM 0.932 in accuracy.

Sawsen and his colleagues [37] developed a new NR metric called MD-QA for evaluating stereoscopic 3D video by taking features from stereoscopic frames containing motion vector lengths and depth map information. Depth information was generated by utilizing disparity (the horizontal distance value of each image pixel), but disparity for stereoscopic image is defined as the horizontal pixel between two similar pixels. The proposed metric MD-QA used H264/AVC MPEG1 and MPEG2. The proposed model was assessed by three FR metrics, SSIM, PSNR and VQM, it was compared to MOS score in the first step, then the second step extracted disparity map between the left and right views of each frame using sum of absolute differences method (SAD) in order to generate the depth information by using equation (9). Motion vectors map were also generated by using block matching method in order to contribute into 3D video quality. After that the author calculated weighting value of depth and motion vector using non-linear regression function. In the final step the author used Pearson Linear Correlation Coefficient PLCC benchmark to measure the proposed metric performance depending on the weighted values of depth and motion vectors features by equation (10)

$$Z_p = \frac{2fl}{d} = \frac{fB}{d} \quad (9)$$

$$Q_f = (w)M_f + (1-w)D_f \quad (10)$$

The database was used from EPFL and the authors used six videos, each has 5 scenes and frame rate of 25 fps, resolution of 1920x1080 pixel and duration was 10 second. 20 subjects evaluated the videos. The proposed model MD-QA was evaluated and correlated with MOS, the model achieved better results than others.

Tselios and his colleagues [38] suggested a QoE objective model for 3D video according to network parameters. The model proposed PSNR metric to assess the perceptual quality of stereoscopic video of the two diminutions (2D), left and right views measured at the receiver. This paper used Double stimulus Continuous Quality Scale (DSCQS) subjective method, and the measured environmental luminance was based on the recommendations described in ITU-R BT.500. RTP packets were created according to standard (Single NAL Unit, Single-Timing Aggregation Packet). The study found that there is a strong correlation between objectivity and subjective quality. However, authors did not justify the use of different resolutions 1280x1024 for degraded and 640x480

for objective and the same quantization parameters (QP=24) for each left and right views with different bit rates.

## **6. Frame concealment for 3D stereoscopic based depth information**

This section includes three aspects, the first aspect discusses methods and algorithms for frame concealment due to packet loss of stereoscopic 3D video transmitted over IP network, the second aspect considers the preliminary results and the third aspect describes the tools.

### **6.1. Frame concealment and algorithms**

Carreira and his colleagues [39] considered three different algorithms which are named as Frame-Freeze (FF), Double-Freeze (DF) and Base-Copy (BC). In FF, the last frame that was received in the auxiliary view is copied to the temporal instant of the lost frame in the same view. DF is the base view where a frame temporally co-located within the last one that was received in the auxiliary view is copied into both views at the missing time instant. The BC copies the temporally co-located frame from the base view into the auxiliary view at the missing time instant. All concealment methods in this work are used to assess the subjective impact in temporary loss of depth information, in addition each of them have three different forms named as (1X2 means every other frame is lost, 2X3 means one of 2 consecutive frames are lost in each group of 3 and 3X4 consider that 3 of 4 frames are lost). These losses happen in two ways. First is (1 x 50) loss which occurs in the second half of the sequence 50%. Second is (1 x 100) loss frame which happens for the whole sequence 100%. In this study the base view is received without losing whereas the auxiliary view is prone to loss, therefore, only depth information is impacted. The experimental results of subjective test explain that all methods give higher MOS values in 1x50 patterns than in 1x100, also the FF method gives higher average MOS values than other methods in case the depth information loss decreases, and it decreases when percentage of loss increases. However, the concealment methods assessed in this paper use regular frame loss and therefore, the results might not be suitable for the random frame loss.

Pinto and his colleagues [40] includes different frame loss methods which were mentioned in [39], i.e., FF, DF and BC. The DF method corresponds to freezing the last 3D image received without errors and BC method transforms 3D to 2D because both views consist of properly the same frames during the error time. However, this study assumes that the corresponding frame is lost whenever a packet is lost. The research in [40] aims to evaluate the effect of packet loss on the perceived depth whereby packet loss occurs more randomly than the previous study. Markov model examined two models of the packet loss in order to generate the packet loss forms. The author conducted subjective test using ACR methodology to find out the relationship between 3D quality and content features, frame concealment methods and packet loss statistics. The experimental results showed all concealment methods achieved similar MOS with very small error, but when error is increased, the BC gives the best MOS results compared to the FF and DF methods except for the highest disparity sequence, whereas 3D quality with low perceived depth is less annoying than high perceived depth. However, this work was not suitable to recover color because it focused only on concealing depth perception.

Tae-Young and his colleagues [41] split loss frames into four error forms. Table.3 explains how to recover the

error pattern. The authors encoded stereoscopic V+D of color and depth individually. Disparity Compensated Prediction (DCP) and intra-prediction for predicting the right color and depth frames from the left frame were used, whereas, using Motion Compensated Prediction (MCP) for predicting the left frames of color and depth temporally from previous frames. The proposed method assumed the first frame of GOP is received correctly. The performance of the proposed method assessed in the case without loss of the first frame, at frame loss rate of 10% the color and depth in the left and right views are randomly lost. The strength of the proposed method was based on error location. The simulation results showed the proposed algorithm which uses PSNR for objective test gave better results than four traditional methods (DBMA, MV sharing, PMVE +DFDP and PMVE + SCA-P). If the loss of color and depth happened at the same time, the proposed algorithm conceals by copying the temporally previous frames to replace the lost frames.

**Table 3:** Error pattern and concealment methods

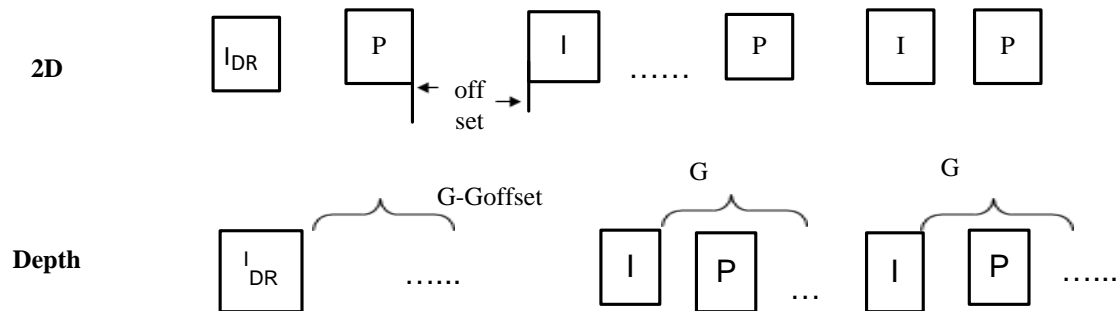
Error pattern	Left-color loss	Left–depth loss	Right-color loss	Right-depth loss
<b>Concealment Methods</b>	Using Motion Vectors(MV) for the temporal error concealment	Using MV for the corresponding left color frame methods	First select matching pixels between left and right views using 3D warping scheme then utilize the high interconnection between the matching pixels	Using MV for the right and left color frame

Yang and his colleagues [42] presented slice loss method for intra frame 3D video called a depth-assisted error concealment method of 2D+depth video sequence. The authors used motion information directly from depth without considering the similarities in the motion information by using the MVs from the depth sequence, but in the case of the intra mode, the MV information cannot be used. The authors supposed that intra frames of the depth sequence are a temporal offset which is presented in Figure 3. The GoP denotes the size of G and GoP offset as G offset ( $0 < G_{offset} < G$ ). The frame in the depth sequence which matches the intra frame was presented as the inter-coded 2D sequence. The proposed method used H.264/AVC codec reference software (JM16.2). In this study a slice represented a Macro Block (MB) row. The experiment used different resolution such as (1024x768), (720x540), (1280x960) and (1920x1080) and different PLRs (2%, 5%, 10%, 15% and 20%). The proposed depth-assisted method was compared against three other methods.

- First is the Spatial Interpolation or (spatial) method, in this method the linear interpolating takes the nearest pixels in the top and bottom neighbor of MBs to recover each pixel in the lost slice.
- The second method is the Zero Motion Estimation or copy method which replaces the co-located MBs of the previous decoded frame to MBs of the lost slice.
- The third method is the Decoder Motion Estimation combined with BMA (MCEC), which applied the motion estimation at the decoder for the 6 neighbor MBs as MV candidates to conceal each MB of the lost slice in current frame and to recover the lost region by Boundary Matching Algorithm (BMA).



The authors exploited the similarity of motion information between 2D sequence and the depth sequence and availability of the MVs of the corresponding frame in the depth sequence to conceal lost packet in the 2D sequence. For representing the depth sequence at the 2D+depth, authors took the depth from 3D scene in order to assist with 2D sequence. In the experiment the authors put in mind the average value of all slice loss in all intra frames. After testing each intra slice separately, the first (Data Reuse) IDR frame of a sequence is not taken into account. This method is that it decreases the calculation at the decoder, the output of bit rate is stable and The SNRY achieved better result than other three methods because the proposed method used 9MV while MCEC using 6 MV.



**Figure 3:** Organization of 2D and depth sequence

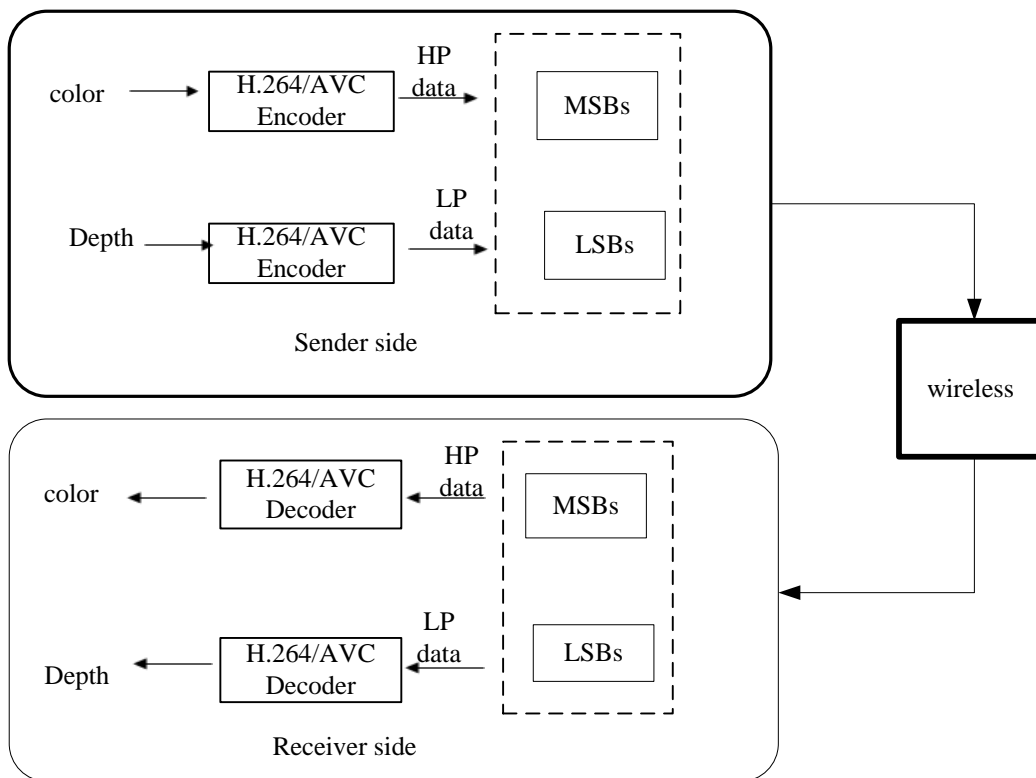
Ekmekcioglu and his colleagues [43] proposed algorithm that depended on the similarity in motion features of the color and depth map by calculating only one set of MV for each macro block in the color, then duplicating it for concealing the depth map loss frame. The authors assumed independent transmission, encoding and decoding of color-plus-depth. Therefore, there is no decoded data exchange between color and depth map. Packet loss in color utilizes the correct MV information from the co-located macro block in depth map to conceal the basic MV information for the loss color frame. Also if both color and depth are lost simultaneously the algorithm will apply motion copy concealment. In this work both depth map and color are encoded with H.264 reference software with joint model version15 (JM v15). The experiment was conducted under 3% packet loss and simulation results showed the stereoscopic quality calculated by the average PSNR value of the rendered left and rendered right video. The strength of the proposed method was in the motion vector set at the encoder side that is only calculated one time, so it decreases processing time at the encoder side. It also provides a redundant set of color video motion vector to enhance the motion vector. Furthermore, it improves average quality of rendering left and right video at color and depth combination. The error concealment performance was compared with reference system and the decoded quality changed from 1.5 dB to 6.5 dB. However it is not good for quality of depth map because only color quality was improved. The method t requires subjective quality assessment.

Alajel and his colleagues [44] presented Unequal Error Protection (UEP) scheme which relies on Hierarchical Quadrature Amplitude Modulation (HQAM) for the color plus depth maps stereoscopic transmitted over wireless channel. Figure 4 shows the UEP scheme, it focuses on the unique feature of color plus depth map. It takes into account unequal importance of the color and depth map video by using by hierarchical 16-QAM.

HQAM splits data stream into two sub-streams, the most sensitive bits are known as the high priority HP, while the others bits are known as the low priority (LP). The proposed scheme gave color more protection by assigning it with HP. It denoted most significant bits (MSBs) whereas the depth has (LP) which denoted less significant bits (LSBs). In this research two sequences are used, one is with the low motion while the other is with the high motion encoded at spatial resolution of 720x576, the color and depth maps are reconstructed by using DIBR equation (11) and (12). The results achieve better quality up to 5dB when compare to conventional Equal Error Protection (EEP) method. In the experiment, the average PSNR was used calculate quality of left and right views. The experimental results showed the depth map decreases from 13.98 dB to 11.49 dB and color quality increases from 14.09dB to 18.79dB. The advantage of this research is that it protected high priority data without increasing bits. However, it requires subjective tests to validate objective results.

$$XL = x + \frac{a_x.t_c}{2} \left( \frac{1}{z} - \frac{1}{z_c} \right) \tag{11}$$

$$XR = x - \frac{a_x.t_c}{2} \left( \frac{1}{z} - \frac{1}{z_c} \right) \tag{12}$$



**Figure 4:** The proposed UEP Method

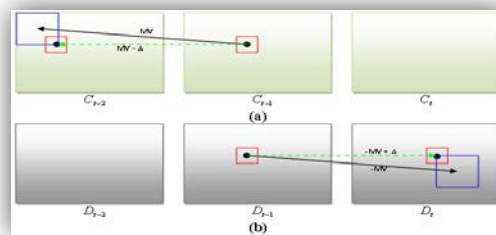
Li and his colleagues [45] considered error concealment method in color plus depth dual format, the proposal conceals the lost information of color or depth on receiving correct information of color or depth. It encoded each color and depth by using H.264/AVC separately. The proposed algorithm has three cases of the loss frame.

- The color was lost while the corresponding depth frame was received correctly

- The depth was lost while the corresponding color frame was received correctly
- The color and depth are lost

The concealment of color loss has two parts, first part is pre-processing of the MVs in the former color frame (t-1) which contains MV estimation for intra-coded MB and MV refinement for all MBs, in this part authors computed the MV of neighboring non-intra blocks by using unit of each intra-coded MB 4x4 part. It used the corresponding depth information which corresponds to neighboring 4x4 blocks of large depth difference to the current 4x4 block to get more accurate MVs.

The second part is extrapolation of color MVs from previous frame (t-1) which is helped by depth information in current frame (Depth-assisted MV extrapolation). In this part authors split MV extrapolation into two processing. The first processing depends on the MVs, which takes from the PMVE (Pixel-based Motion Vector Extrapolation) method to give a Color Error Concealed Base Frame (CEC\_BF). The second one is CEC\_BF, this process classified pixel into R (Reliable), SR (Slightly Reliable) and, UR (Unreliable). SR and UR tolerate the different refinements.

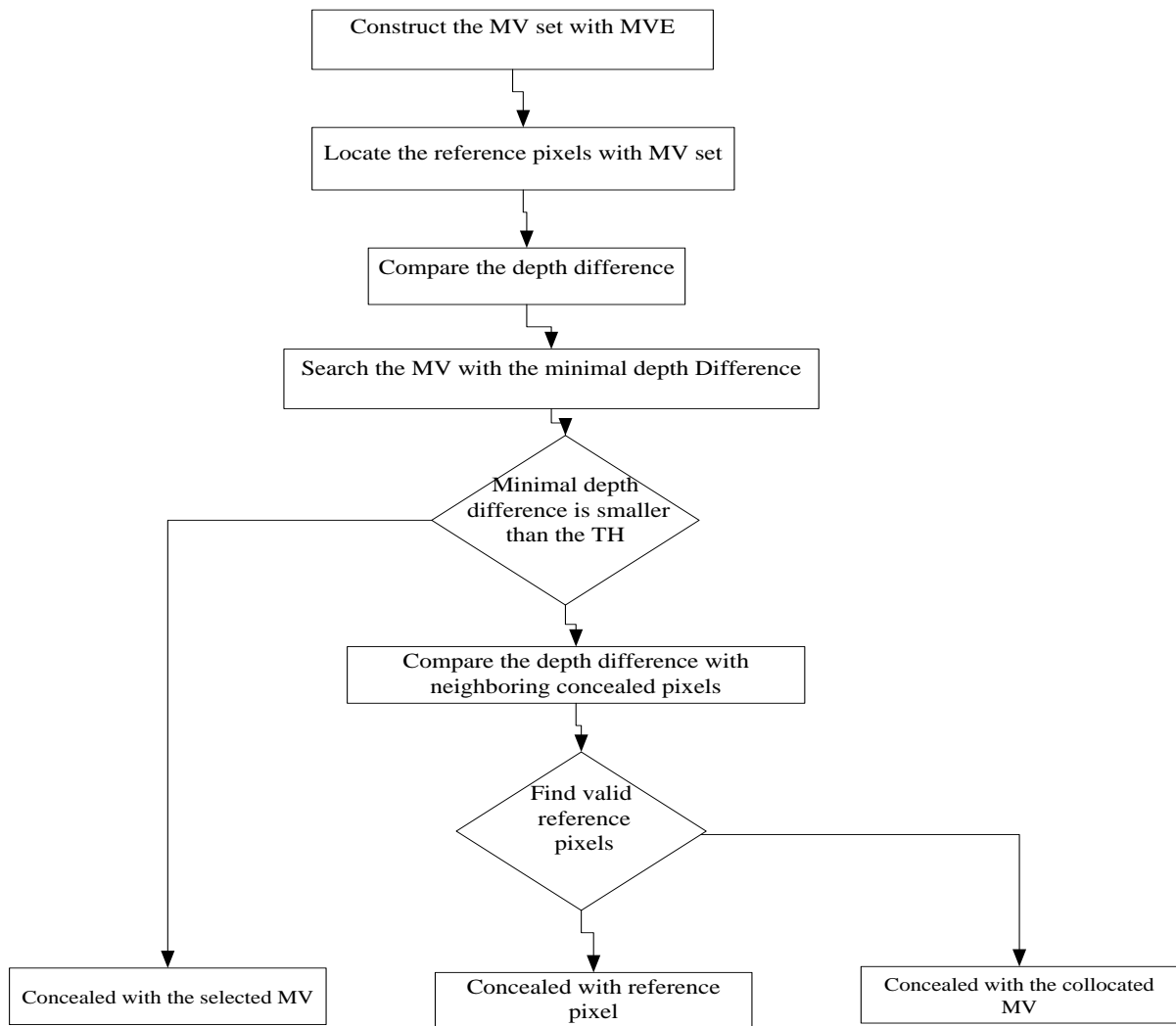


**Figure 5:** The MV sharing and BMA (Boundary Matching Algorithm)

Depth loss in the proposed method depends on MV sharing and Boundary Matching Algorithm (BMA) (Figure 5), where the subscript t-1 is the frame index, (i,j) and (x,y) are pixel positions,  $C_{t-1}(i,j)$  means the color value,  $D_{t-1}(i,j)$  means the depth value,  $MV_{t-1,x}^c(x,y)$ ,  $MV_{t-1,y}^c(x,y)$  is the color MV, and  $B_{(x,y)}^B$  is a template window around (x,y). BMA updates by estimating MVs in units of 4x4 blocks. The experimental results demonstrate average PSNR under different packet loss rates (5%, 10% and 15%), PSNR achieved up to 0.68 dB in color and up to 2.42 dB in depth. However authors do not explain how to recover frame if both color and depth are lost. This work also requires subject assessment.

Bo Yan and his colleagues [46] considered an efficient frame concealment algorithm for depth image based on 3D video transmission. It used MV extrapolation (MVE) for the 2D video and contained the Hybrid MVE (HMVE) and Pixel Motion Vector Extrapolation (PMVE). HMVE used MV of the pixel and MV of the block. This method split pixel of error frame into three types, pixels that are covered by at least one extrapolated 4x4 blocks, pixels that are not covered by any of the extrapolated 4x4 blocks and pixels that are not covered by any of the extrapolated 4x4 blocks. The algorithm is detailed in Figure 6. The simulation encoded sequences by using H.264/AVC codec. The subjective results showed that the proposed algorithm achieve accurate motions,

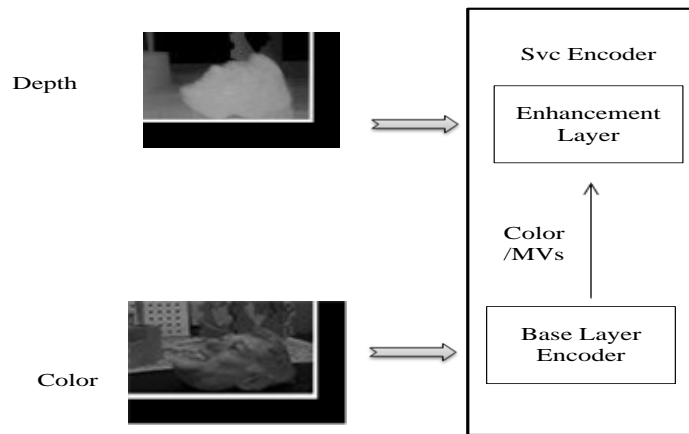
however it is not suitable for color concealment.



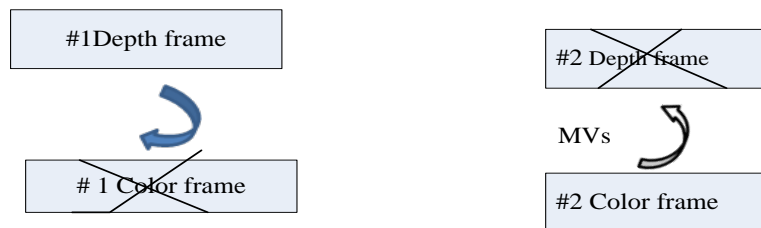
**Figure 6:** Procedures of proposed method

Hewage and his colleagues [48] presented concealment method for color and depth map based on stereoscopic which uses the motion correlation of color and depth map image sequence to recover frame losses during transmission. In this method MV sharing scheme is using the Scalable Video Coding (SVC) layered encoding architecture as show in Figure 7. The color is encoded at base layer whereas depth at enhancement layer. Figure 8 shows how to retrieve lost MVs of color frame using corresponding depth which is received correctly, during the depth map encoding. The first mode is motion compensation. It is using the enhancement layer for predicting the current depth image MBs by reusing the MVs of the corresponding color image. It is faster than traditional motion compensation modes however, it does not do motion estimation while performing depth encoding. The second mode is macro block MB skip for increasing the compression efficiency. An uncorrupted corresponding color is used to recover loss of MVs depth frame. The recovered MVs from corresponding view are used to predict the current frame. If both corresponding color and depth frame are lost then traditional frame concealment algorithm is used to conceal the lost frames. Two sequences of color and depth map are used in experiment, one sequence is more complex with camera motion and multiple objects, while the other sequence

is captured with a static camera and a stationary background .The average SVC encoder bit rate is set to 2Mbps. The depth bit rate is kept below 25% of its corresponding color video bit rate, the author mentioned that in some area for low correlation between motion of color and depth map , the proposed MV sharing method imposes a bit balance during depth map coding. The overhead is added due to the depth information and requires to be kept at a smaller percentage of its corresponding color video. For generating bit-stream using separate MV estimation, the authors assume the I-frame for color and depth are received without loss. The corrupted frames are recovered using frame coping which is decoded from previous frame then copied to current frame position.It also enhanced the perception of depth in SSV and performed better than both frame copying methods based on separate motion estimation and shared MVs because an error free decoding as residual information is not used during reconstruction streams. Despite that it achieved lower results than separate MVs (Table 4); the authors set average SVC encoder bit rate at 2Mbps without any justification. The proposed concealment method is required to be measured with subjective method.



**Figure 7:** The MV sharing architecture based on SVC



**Figure 8:** Frame concealment using MVs

**Table 4:** The image quality

	Orbi		Interview	
	Color PSNR/dB	Depth PSNR/dB	Color PSNR/dB	Depth PSNR/dB
Separate MVs	40.13	38.10	41.05	42.34
Proposed MV sharing scheme	39.69	37.25	41.05	40.27

## **7. Discussion and future work**

This paper focused on the issues of improving the transmission of stereoscopic color-plus-depth based on error concealment for better QoE. This paper was split into two parts.

The first part discussed the concept of 3D video communication chain, 3D transmission challenges, error resilience tools, stereoscopic format and QoE for 3D video. This part also reviewed the methods and methodologies of subjective objective metrics evaluation. It discussed subjective evaluation of stereoscopic depth information metrics such as ACR, DSCQS, MICSQ and SAMVSQ and the objective evaluation of stereoscopic depth information metrics such as PQM, VQM, VQMPL, VQMcomp, NVQM, ENVQM, PSNR, and MD-QA .

The second part investigated the algorithms and methods for concealing and recovering the lost information of color and depth maps. Some researches worked on using motion correlation to recover frame losses during transmission while others worked on similarity motion, others used both macro block skip and motion estimation. Other researches use data partition by proposed multiple description coding (MDC) and Unequal Error Protection (UEP). Other studies used pixel value to propose decision making and pixel motion vector extrapolation. Different methods and algorithms were found to contribute on concealment color and depth such as FF, DF, BC, DCP, MCP, HMVE and combine the depth offset and decoder motion estimation.

Finally, the paper found that most of the state of the art error concealment methods for stereoscopic color-plus-depth focused on the depth maps recovering. Other 3D video factors such as comfort, motion, disparity and discomfort should be investigated further in order to derive and quantify their impact on QoE.

Packet losses can be caused by queuing congestion and error channels. Both packet losses due to queuing delay and error channels can be predicted before they occur. A good model to predict packet losses can be effective to mitigate the end users QoE by using alternative techniques such as bitrate adaptation and Forward Error Channels (FEC). In this context, frame concealment methods would be the last resort if other techniques have failed. The combination of these techniques is envisaged to improve the end users QoE just before and after packet losses have occurred. The future work will propose QoE objective models and metrics in order to measure the quality for the stereoscopic color plus depth transmission, especially for NR and RR metrics due to the lack of special metrics for 3D video. The investigation on how to conceal slice loss and mitigate the impact of other network parameters such as delay in order to achieve better quality will be carried out.

## **8. Conclusions and Recommendations**

This paper has reviewed the state of the art research on the enhancement of Quality of Experience for stereoscopic 3D video based on depth information. The main focus was the application of frame loss concealment techniques in order to mitigate the impact of packet losses on the quality of experience for end users. This review paper also discussed the concept of 3D video communication chain, 3D transmission challenges, error resilience tools, stereoscopic format and QoE for 3D video with its subjective objective metrics evaluation. It was found that most of the state of the art error concealment methods for stereoscopic color plus

depth concentrated on the depth maps recovering. However, other 3D video factors such as comfort, motion, disparity and discomfort should be investigated further in order to derive and quantify their impact on the QoE. I recommend researchers in this field to make the effort to develop and propose algorithms and schemes to solve the problem of packet loss in stereoscopic 3D video because it has the most effect in transmission sequences so at the end to satisfy the end user. Nowadays social networking programs are increasing and user use to send and receive video for this purpose we must develop plans to solve the problem to improve QoE further more to propose efficient methods which can be used in other 3D video format such as multiple view video MVV.

## References

- [1] Chaminda Hewage. "The transmission aspect of 3D video" In 3D processing and transmission fundamentals, 1st edition ,2014, PP. 63- 68 .
- [2] Rec BT.2021, "Subjective methods for the assessment of stereoscopic 3DTV systems", 2008.
- [3] Recommendation REC. ITU-R BT.1438, "subjective assessment of stereoscopic television pictures", 2000.
- [4] Smolic, A., Mueller, K. , Merkle, P. and Vetro, A. , "Development of a new MPEG Standard for Advanced 3D Video Applications," , Image and Signal Processing and Analysis, ISPA Proceedings of 6th International Symposium on 2009 , PP. 400 – 407.
- [5] Takashi Matsuyama, Shohei Nobahara, Takeshi takai, Tony Tung . 3D video and Its Application. , Springer-Verlag : London ,2012,PP.2-6.
- [6] Bruno Zatt, Muhammad Shafique, Sergio Bampi, Jorg Herkee, Energy. 3D video coding for Embedded Devices , Efficient Algorithms & Architectures. Springer Science+Business Media : New York, 2013,PP.2-3.
- [7] Kazuhisa Yamagishi, Taichi Kawano, and Kimiko Kawashima, "QoE Assessment Methodologies for 3D Video Services." NTT Technical Review, Vol. 11 No. 5, PP.1-5, May 2013
- [8]. Philipp MERKLE, Karsten MÜLLER, and Thomas WIEGAND, "3D video: Acquisition, Coding and Display." IEEE Transactions on Consumer Electronics, vol.56, no.2, pp.946-950, 2010.
- [9] Anil Fernando, Stewart T, Worall and Erhan Ekmekcioğlu. (2013,AUG,16). 3DTV Processing and transmission of 3D video signals. (First Edition). [on-line]:Available: [www.onlinelibrary.wiley.com/book/10.1002/9781118705711](http://www.onlinelibrary.wiley.com/book/10.1002/9781118705711)[Feb, 23, 2016].
- [10] Guan-Ming Su, Yu-chi Lai, Andres Kwasinski and Haohong Wang. (2012,Dec,12).3D visual Communications.(first edition) .[on-line]:Available:[www.onlinelibrary.wiley.com/book/10.1002/9781118358023](http://www.onlinelibrary.wiley.com/book/10.1002/9781118358023) [JUL, 3, 2016].

- [11] E.Kohler, M.Handley and S.Floyd,, “ Datagram Congestion Control Protocol (DCCP) ” ,Internet Engineering Task Force RFC 4340”, Mar .2006.
- [12] S.Floyd, E.Kohler and Padhye, “Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC)”, Internet Engineering Task Force RFC 4342, Mar .2006
- [13] Frederic dufoax Beatrice Pesquet-Popescu, & Marco cagnazzo. (2013, APR, 22). Emerging technologies for 3D video creation, coding, transmission and Rendering.(first edition).[on-line].Available: <http://onlinelibrary.wiley.com/book/10.1002/9781118583593> [Feb, 24, 2016].
- [14] Ekmekcioglu, Guildford, Fernando, Worrall, S.T, Kondo, A.M., “ Utilization of motion similarity in Color-plus-Depth 3D video for improved error resiliency, ”, published Information and Automation for Sustainability (ICIAFs), 5th International Conference on 17-19 Dec. 2010 , PP:475 – 478.
- [15] Renaud, ManLi,DavidSoldani , “ QoE management in UMTS Cellular Systems.”, Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd , Vol:2 ,PP:981-985.
- [16] Amendment 1 to ITU-T Recommendation P.10/G.100 (2006). “Vocabulary for performance and quality of service”, was agreed on 25 January 2007 by ITU-T Study Group 12 (2005-2008)
- [17] Chaminda T. E. R. Hewage and Maria G. Martini, “Quality of Experience for 3D video streaming.” , IEEE Journals & Magazines of Communications Magazine, Vol.51, No.5,PP: 101 - 107 ,2013.
- [18] ETSI TR 102 643 V1.0.2 (2010-01) , “ Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services” , 2010.
- [19] Kjell Brunnström, David Hands, Filippo Speranza and Arthur Webster, “VQEG Validation and ITU Standardization of Objective Perceptual Video Quality Metrics.”, IEEE Signal processing magazine, Vol. 26, PP .96-10 , 2009.
- [20] Recommendation ITU-T P.910, “Subjective video quality assessment methods for multimedia applications”, 2008.
- [21] Recommendation ITU-T P.913 (01/2014) ,“ Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment”, 2014.
- [22] Recommendation ITU-T J.247 , “objective perceptual multimedia video quality measurement in the presence of a full reference” , 2008.
- [23] Recommendation ITU-T J.246, “perceptual visual quality measurement techniques for multimedia services over digital cable television Networks in the presence of a reduced bandwidth reference”, 2008.



- [24] D. Hands, O. V. Barriac, and F. Telecom, "Standardization activities in the ITU for a QoE assessment of IPTV." *IEEE Communications Magazine*, vol. 46, pp. 78–84, 2008.
- [25] NNT group. "Communication services QoE Assessment Research portal site @NNT." Internet: [http://www.ntt.co.jp/qos/qoe/eng/technology/3d/01\\_1.html](http://www.ntt.co.jp/qos/qoe/eng/technology/3d/01_1.html) , Feb.5.2016 [Jule.3.2016].
- [26] Pierre Lebreton, Alexander Raake, Member, Marcus Barkowsky, Member, and Patrick Le Callet, Member, "Evaluating Depth Perception of 3D Stereoscopic Videos.", Vol.6, No.6,PP.710- 720, 2012
- [27] A.S Amar, R.M.Swash and A.H.Sadka "Subjective Quality Assessment of 3D Videos,", The Falls Resort and Conference Centre, Livingstone, Zambia ,September 2011,PP.13 – 15.
- [28] Mysirlidis, C, Dagiuklas, T, Politis, I, Ekmekcioglu, E, Dogan, S, Kotsopoulos, S, "Quality evaluation of 3D video using color-plus-depth & MDC over IP networks," , 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2014, PP.1- 4.
- [29] Taewan Kim, Jiwoo Kang, Sanghoon Lee and Alan C. Bovik, " Multimodal Interactive Continuous Scoring of Subjective 3D Video Quality of Experience ," , Vol.16, PP.387- 402 , 2014
- [30] Hossein Malek mohamadi, W. A. C. Fernando and A. M. Kondoz, " A New Reduced Reference Objective Quality Metric for Stereoscopic Video," , GC12 Workshop: Quality of Experience for Multimedia Communications , 2012. PP.1325-1328.
- [31] Hewage and Maria G. Martini , "Reduced-Reference Quality Evaluation For Compressed Depth Maps Associated With Color Plus Depth 3d Video," *Image Processing (ICIP), IEEE International Conference on*, 2010,PP.4017- 4020
- [32] Chaminda T. E. R. Hewage and Maria G. Martini, "Reduced-Reference Quality Assessment for 3D Video Compression and Transmission.", *consumer electronics IEEE Transaction*, Vol.57, PP.1185-1193, 2011.
- [33] P. Joveluro, H. Malek mohamadi, W.A.C Fernando, and A.M. Kondoz , "Perceptual Video Quality Metric for 3D Video Quality Assessment," , 3DTV- conference: the true vision- caption, transmission and display of 3D video (3DTV-CON), 2010, PP.1-4.
- [34] Politis, V. De Silva, S. Dogan, C. De Alwis , T. Dagiuklas, and A. M. Kondoz , " Objective Measurement of Transmission Losses and Compression Artefacts in Stereoscopic 3D Video," , *Communications Workshops (ICC), IEEE International Conference on* , 2013,PP.667-671.
- [35] Yi Han, Zhenhui Yuan, Gabriel-Miro Muntean, " No Reference Objective Quality Metric for Stereoscopic 3D Video," , *Broadband Multimedia System and broadcasting (BMSB)* ,2014 , PP.1-6.
- [36] Yi Han, Zhenhui Yuan, Gabriel-Miro Muntean, "Extended No Reference Objective Quality Metric for

- Stereoscopic 3D Video,” Communication Workshop (ICCW), IEEE International Conference on , 2015, PP.1729-1734.
- [37] Sawsen Abdulhadi Mahmood , Rana Fareed Ghani , “Objective Quality Assessment of 3D Stereoscopic Video Based on Motion Vectors and Depth Map Features,” Computer Science and Electronic Engineering Conference (CEEC) , 2015, PP.179-183
- [38] Tselios, Athanasios Kordelas, Tasos Dagiuklas, and Andreas Papadaki , “A model of network related QoE for 3D video, ”, Globecom Workshops (GC Wkshps)EEE, 2012, PP.1335-1340.
- [39] J. Carreira, L. Pinto, N. Rodrigues, S. Faria, P. Assuncao “Subjective Assessment Of Frame Loss Concealment Methods In 3D Video,” Picture Coding Symposium (PCS) , 3,2010 ,PP. 182–185.
- [40] L.Pinto,J.Carreira , S.Faria, N.Rodrigues,P.Assuncao “Subjective Quality Factors In Packet 3D video,” in Quality of Multimedia Experience (QoMEX) Third International Workshop,Sept.2011,PP. 149–154 .
- [41] Tae-Young Chung, Sanghoon Sull, and Chang-Su Kim, “Frame Loss Concealment for Stereoscopic Video Plus Depth Sequences.”, IEEE Trans. on Consumer Electronics, vol. 57, PP., 336-344 , 2011.
- [42] Meng Yang, Xuguang Lan, Nanning Zheng, Cosman.P, “ Depth-Assisted Temporal Error Concealment for Intra Frame Slices in 3-D Video.” IEEE Trans. on Broadcasting , vol. 60, PP. 385- 393,2014.
- [43] E.Ekmekcioglu, “ Utilization of Motion Similarity in Color-plus-Depth 3D video for Improve Error Resiliency,” , in Information and Automation for Sustainability (ICIAFs), on. 17–19.Dec. 2010, PP. 475– 47.
- [44] Khalid Mohamed Alajel, Wei Xiang, and Yafeng Wang,“ Unequal Error Protection Scheme Based Hierarchical 16-QAM for 3-D Video Transmission.” , IEEE Trans. on Consumer Electronics , vol. 58,PP.731–738., 2012.
- [45] Wen-Nung Lie , Guan-Hua Lin “Error Concealment for 3D Video Transmission,” , IEEE International Symposium on Circuits and Systems (ISCAS), May. 2013,PP. 2856– 2559.
- [46] Bo Yan, Senior Member, IEEE, and Jie Zhou “Efficient Frame Concealment for Depth Image-Based 3-D Video Transmission.”, IEEE/Journal Trans. on Multimedia, vol. 4 ,PP. 936-941, 2014.
- [47] Chaminda T.E.R. Hewage, Stewart T. Worrall, Safak Dogan, and Ahmet M. Kondoz “A novel frame concealment method for depth maps using corresponding color motion vectors, ”, vectors,on 3DTV,2008, PP. 149–152.
- [48] C.T.E.R. Hewage, S. Worrall, S. Dogan, and A.M. Kondoz “Frame concealment algorithm for stereoscopic video using motion vector sharing,” , IEEE on Multimedia and Expo, 2008 , PP. 485–488.