Document downloaded from:

http://hdl.handle.net/10251/159210

This paper must be cited as:

Canovas Solbes, A.; Taha, M.; Lloret, J.; Tomás Gironés, J. (2019). A cognitive network management system to improve QoE in stereoscopic IPTV service. International Journal of Communication Systems. 32(12):1-23. https://doi.org/10.1002/dac.3992

The final publication is available at

https://doi.org/10.1002/dac.3992

Copyright John Wiley & Sons

Additional Information

This is the peer reviewed version of the following article: Cánovas, A, Taha, M, Lloret, J, Tomas, J. A cognitive network management system to improve QoE in stereoscopic IPTV service. Int J Commun Syst. 2019; 32:e3992, which has been published in final form at https://doi.org/10.1002/dac.3992. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

A Cognitive Network Management System to improve QoE in Stereoscopic IPTV Service

Alejandro Cánovas[1*], Miran Taha[1, 2], Jaime Lloret[1], Jesús Tomas[1]
[1]Integrated Management Coastal Research Institute, Universitat Politècnica de València, Gandia, Spain
[2]Computer science department, University of Sulaimani, Kurdistan region –Iraq
alcasol@posgrado.upv.es, miab2@doctor.upv.es, jlloret@dcom.upv.es, jtomas@dcom.upv.es

* Corresponding Author: Jaime Lloret

Abstract

New Internet Protocol Television (IPTV) services are including new technologies such as Stereoscopic TV and three dimensions (3D) HDTV. As well, increase ubiquitous networking and promoting in smart devices has led to high demand Internet Protocol Television (IPTV) (over networks. Stereoscopic content is required higher data flow to support these emerging TV services and there are higher requirements at the network layer to provide good quality of service and quality of experience to the end users in delivering stereoscopic IPTV. In this paper, we propose a new concept of cognitive network management algorithm and protocol based on 3D coding techniques for delivering of stereoscopic IPTV service. The proposed approach explains how the management algorithm observes the network performance to guarantee the quality of the stereoscopic IPTV services, by measuring the performance of quality of service (QoS) parameters (delay, jitter, and packets loss) and quality of experience (QoE) metrics such as Peak Signal-to-Noise Ratio (PSNR), Moving Image Videography (MIV), and Mean Opinion Score (MOS). Those parameters are monitored in order to take appropriate codification decision for IPTV service provider. Moreover, the codification decision uses K-mean classification to select the better codification for the end-users. Therefore, both kinds of 3D coding formats such as Stereo Video Coding (SVC) format and 2D+Z Coding format (2D-plus-Depth) are selected in our experiments. As a result, our proposal successfully ensures the appropriate quality of service and quality of experience to the end users when the service of stereoscopic IPTV is being delivered.

KEYWORDS

3D Coding Formats; Cognitive Network IPTV; QoS; QoE

1. INTRODUCTION

IPTV is becoming a common service for many IP service providers, especially, live television and Video on Demand (VOD). Currently, IPTV networks are not only designed to transmit regular video. Because of this, several systems are developed to provide higher quality to the end user [1]. Moreover, adaptive IPTV services as mentioned in [2] is used to enhance the IPTV service on the Internet. IPTV can distinguish three types of quality parameters: Quality of Service (QoS), Quality of Perception (QoP) and Quality of Experience (QoE) [3]. Nevertheless, current efforts are not only focused on typical video transmission architectures. There are numerous researchers introducing new techniques and methods over other types of IPTV architectures (such as peer to peer networks in [4]). Another challenges in IPTV was how the content of stereoscopic video can be delivered to the end-user, which is become very popular for 3D vision films [5]. The stereoscopic video provides the users with a sense of depth perception by showing two frames to each eye simultaneously. In stereoscopic service, there is two challenges are faced to the IPTV service: first, challenge of analyzing of the stereoscopic service when the service is delivered over IP communication. Second, challenge of improving the process of the transmitting the stereoscopic video.

The quality perception of the delivered stereoscopic 3D video is the key to take into account when assessing the QoE. Taking ETSI TR 102 643 [3] as a reference, the QoE as "A measure of user performance based on both objective and subjective psychological measures of using an ICT (Information and Communication Technology) service or product". The parameters that affect the QoE are different according to the type of application. Some examples can be found in [6] for VoIP and Web [7]. Thus, one of the goals is to determine the factors that affect the QoE in each stage of the stereoscopic 3DTV delivery system. According to [8], a 3DTV integrate system is defined by these stages: First, the 3D/stereoscopic content generation, second, the 3D/stereoscopic representation quality, coding, transmission, decoding, synthesis, rendering and display. Thus, the authors paid special attention to coding, transmission, and display stages. Moreover, introduced transcoding the 3D coding format, deliver to end-users and playback on client sides and the effect these parameters can present to the QoE. Video codification is needed because the bandwidth is limited. This process introduces errors that may affect the video quality. The techniques of 3D video coding is pointed in the technical report of [9], the technical report was explained the video coding process of different 3DTV system such as Multiple Video Coding (MVC), Multiple Description Coding (MDC) and 3D mesh Compression (3DMC). The report showed that, different schemes of coding are used for IPTV but the coding techniques are not curial in IPTV networks as explain in [10], the IP network is extremely significant in the transmission of the IPTV service stream. The network parameters that mainly affect the received stereoscopic 3D video quality during the transmission process are the delay and packet loss. Therefore, there are several studies for 2D but bare researches on the stereoscopic 3D. In the 3D case, which is presented in [8], a transmission distortion is perceived differently for one than for both views. A degradation in one view or a temporal misalignment between the left and the right view leads to binocular rivalry. This binocular rivalry strongly degrades the QoE of end-users, as it exhibits visual discomfort, which might lead to a headache or nausea. The service is required new systems and components to transmit stereoscopic 3D properly and more parameters must be taken into account in order to have the minimum degradation of the QoE. In addition, adaptive streaming is becoming one of the main ways to improve the QoE of the end users during watching 3D videos [11], the approach which by providing an exponential relationship between human decisions and the same decisions expressed as a difference of objective metrics.

In this paper, the important parameters of the network service are focused to bring forth the high possibility deliver of 3DTV service to end-user, which is providing better QoE to the users. In order to provide the proposal, two kind of 3D video coding techniques are chosen to design the algorithm for managing the network. The proposed system observes activity of he QoS parameters such as delay, jitter, and packet loss and QoE parameters such as PSNR and MIV when Stereo Video format and 2D+Z format are used in the process. Although jitter is not a relevant parameter due to the usual use of buffering techniques in almost all multimedia players and IPTV setup boxes, it is very helpful to study it in order to know the best configuration for the multimedia network design test bench.

The structure of the paper is arranged as follows. Section II provides the works that related to 3D video coding and IPTV transmission. Section III explains the general concepts of 3D coding techniques and the evaluation parameters utilized in our experiments. Section IV shows the network management algorithm, which guarantees the QoE for delivering stereoscopic IPTV. Section V details the video controller decision algorithm. The training phase and tagging process are explained in Section VI. Section VII shows the performance evaluation of the cognitive network management algorithm. Finally, Section VIII provides the conclusion and future works.

2. RELATED WORK

In this section, we present some related works on 3D and stereoscopic coding techniques, then giving details on the algorithms and the protocols for stereoscopic 3D video delivering in networks.

Prominent methods are reviewed for depth stereoscopic video and coding color, Hewage et al. [12], the coding efficiencies are proposed for the video depth and encoding color, which is based on the standard of H.264/Stereo Video Coding (SVC) video-coding. Performance of the experiments is compared to implementation factors and coding efficiency with H.264/AVC video coding standards and MPEG-4 MAC (Multiple Auxiliary Component). They concluded that the configuration based on H.264/SVC performed similar to H.264/AVC and outperformed the MPEG-4 MAC.

The work presented in [13] gives an overview of the techniques that ISO's Moving Pictures Experts Group, which defined in the MPEG-2 and MPEG-4 standards for encoding stereoscopic video. The new technique presented in [13] was based on the capture from multiple cameras; it allows a very simple synthesis of different viewpoints by disparity-compensated projection. Therefore, the results showed that the feature of viewpoint adaptation towards a video object could be accomplished with a low-complexity scheme, while high quality preserved.

Furthermore, Chung-hua et al. in [14] pointed out the critical issue deteriorates the 3D viewing experiences on the 3D mobile devices to improve visual comfort on mobile devices, an efficient and effective algorithm to stabilize the stereoscopic images and video for the 3D mobile devices in [14]. The proposed algorithm rectified the video jitter. Moreover, they used the stabilization method adopts matrix operations to speed up the process and reach high quality in the stereo videos.

Authors of [15] Kiana et al. described the challenges of stereoscopic content when the videos are shown on TV sets, desktop monitors, and mobile devices. The problem is addressed by proposing a new system structure for streaming the stereoscopic content. The aspect of their work provided a computationally efficient depth adjustment technique, which can automatically optimize quality of experience for videos of field of sports such as soccer, football, and tennis. Additionally, the proposed method enables depth personalization to permit end-users to adjust the amount of depth according to their preferences.

Petrovic et al. [16] presented a prototype to implement 3D-video streaming system using an IP network. They delivered the multi-perspective transmission as an on-demand layered streaming problem and implemented a stereoscopic streaming prototype using standard transport protocols and compression techniques. Authors of [17] addressed the continuous navigation requirement of 3D-video systems. In [18], the algorithm to provide the best QoE to the end user is proposed. The algorithm is built by considering the evaluation performed when four free codecs are used in stereoscopic videos. El-Yamany et al. [19] performed the assessment of video quality that aims to evaluate the distortions initiated by depth map compression and the view synthesis procedure in MVD coding systems. They described that many of the existing researches overlooked are essential to the reliability of quality evaluation. As a conclusion, they recommended using average luma PSNR of the synthesized prospect resulting from compressed depth views and uncompressed texture views.

Young-il Kim et al. [20] suggested the 3D image transmission technology in the mobile IPTV environment. They tested the performance of their proposed method in terms of QoS/QoE parameter of users. According to the simulation results presented in this work, the proposed "Separated 3D transmission" algorithm, support good quality image service in terms of the user’s QoE/QoS parameter. Young-uk Chung [21] presented a novel selective frame discard method to address the problem of network bandwidth mismatch. The proposed method considers the relationship between the 2D video and the depth map in its decision to discard overdue frames. These frames are selected after additional consideration of the playback deadline and the inter-frame dependency relationship within a group of picture. The simulation results show that the proposed method enhances the quality of 3D video streaming even in bad network conditions.

In [22], authors presented the prototype for real-time video amassing application, transmission, and high-resolution 3D display. Array of hardware-synchronized cameras is used to take multiple perspective views of the scene. Therefore, the system is developed to a fully distributed architecture with clusters on the sender and the receiver. According to the results, the proposed system architecture is flexible sufficient to enable a broad range of testing on 3D TV.

Finally, the work is presented in [23] describes a framework of the application of stereoscopic TV transmission over IP networks. Authors are deeply investigated on the technology of stereoscopic video in the networks. They aimed to decrease the usage of available bandwidth of the networks where the video is delivered to end-users while keeping the depth perception at acceptable levels. Therefore the asymmetric coding technique is proposed for depth and color video, which is used the scalable extension of H.264/AVC.

Therefore, the approach presents in this paper is contrary from other approaches discussed above; we employ a cognitive model to enhance QoE by providing proactive analyze of QoS metrics and QoE parameters for delivering Stereoscopic IPTV to end-users, the model bases on using classification approach to classify 3D stereoscopic coding video streaming for IPTV service.

3. STEREOSCOPIC (3D) CODING AND EVALTION PARAMTERES

In this section, we give details the main conceptions of the 3D coding formats, which utilize in our experiments then the evaluation parameters to measure the received stereoscopic 3D video as described follows.

3.1. 2D+Z Coding format

2D+Z [24] is a 3D coding format. It uses two images: one for the color component and one for the depth map. The depth map image is in grey scale. A depth value is assigned to every pixel of the 2D color image, which is used to render a 3D image. Figure 1 shows an example of the 2D+Z format. The main feature of this codification reduced the bandwidth used regarding the stereo codification. It happens because it only uses an image with all the information and another with the depth in grey scale. Depth range is limited by minimum and maximum distance to the camera. To compress the 2D+Z format video sequence H.264 Auxiliary Picture Syntax is used as compression standard. The final 3D images can be displayed on a special TV like 3DWOW [25].

3.2. Stereo Video Coding Format

The stereo video coding format is collected by two video sequences of a given target. One of the video sequences correlates to the left view, and the second video sequence correlates to the right view. This is absolutely what human vision does. Human enable to view by using both eyes, generating a left view and a right view of our environment. Our brain is in charge of creating 3D images. The video coding techniques based on this format take into consider the similarity between left and right image in order to reduce the data. The fact of using two images (left and right) in order to obtain the 3D image implies to double the bandwidth needed when stereoscopic videos will be delivered. This stereoscopic content is represented in left/right, half width resolution. An example of a stereo video coding format is shown in figure 2. We used H.264/AVC Multi-View Video Coding (MVC) standard for the stereo video-coding format. It's encoded using two views and uses inter-picture (temporal) and interview prediction. The stereoscopic raw video made by joining the left view and the right view. In order to code and decode both videos, joint multi-view coding video (JMCV) software is used [26]. Basically, their difference has been that one of them is formed by two stereoscopic views, and the other one by one view plus the depth information respect to the other view. In this sense, the second case has higher complexity than the first one because the system has to estimate the depth information.

3.3. Stereoscopic (3D) Evaluation parameters

Once the video is received, the displays are the interface between the system and the human. Next, we will discuss the main ergonomic parameters that may affect the QoE [27]. A 3D exhibit is a multi-view apparatus. Multi-view systems are generally identified for providing superior reproduction of 3D images, since the image becomes different with the observer's point of view related to the screen. In order to heighten the illusion of depth in stereoscopic images the number of views can increase, so the image can be perceived from several positions. Neither the quality of the 2D video features nor the metrics are valid for stereoscopic 3D. The perception of the quality of the stereoscopic 3D video depends on the sensations such as sharpness, depth perception, eyestrain, naturalness or presence [28]. The attributes that effect to these sensations are [29]: cross-talk between views, key-stone distortions, depth plane curvature, puppet theater effect, cardboard effect, shear distortion, picket-fence effect and image flipping. In [30] other effects are introduced such as ghosting and staircase, flicker is introduced in [31] and resolution is introduced in [32].

Fig.1. Example of the 2D+Z format

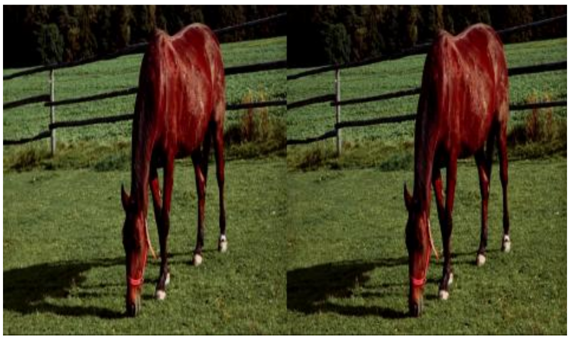

Fig.2. Example of stereo coding format

There are other features of the screens 3D/ST, such as the interface between the visualization system and the human, which may affect the QoE. Classification the type of 3D/ST screens is used. ST/3D screens can be classified as stereoscopic and auto stereoscopic, between others [8]. In stereoscopic screens, users have to use glasses. However, they are not needed in auto stereoscopic. Stereoscopic screeners can be temporal multiplexed or spatial multiplexing. Temporal multiplexed screeners use active shutter glasses (SG) and spatial multiplexed use passive polarized glasses. Volumetric 3D screens are included in auto stereoscopic 3D. They form a visual representation of the objects in 3 dimensions. After this classification, some factors related to the stereoscopic 3D screeners can impact the QoE describes as follows. These factors are loss of resolution, viewing area, distortions and glasses:

- Loss of resolution: A loss of resolution is given when we increase the number of views. The number of pixels that can be placed in a liquid crystal display is limited.
- Viewing Area: The images on common 3D screens, designed having from 62 to 65 mm of view width, may seem wrong and uncomfortable if the user does not see the image from the front and from a safe distance. It happens because the eye may detect 2D images in some areas of the screen.
- Stereoscopic video systems: seeks to exhibit a true 3D view of the real scene. In this process, the distortions can be occurred. These distortions can modify the observer’s perception of the depicted scene or even reduce the QoE of the stereoscopic.
- Another important facet is the necessitation of 3D glasses. These glasses are uncomfortable for the end users. A user is accustomed to watch TV without any special glasses and the use of these glasses for a 3D view is a constraint.

Finally, we describe the QoE assessment methods used in 3D/ST. There are objective and subjective methods. The quality of a video sequence can be evaluated using objective methods with metrics such as: Peak Signal-to-Noise Ratio (PSNR), Universal Quality Index (UQI) [33], PSNR-HVS [34], Single-scale Structural SIMilarity (SSIM) [35], Visual Signal-to-Noise Ratio (VSNR) [36], Weighted Signal-to-Noise Ratio (WSNR), Visual Information Fidelity (VIF) [37], Information Fidelity Criterion (IFC) [38] and Noise quality measure (NQM) [39]. One of the most used metrics is PSNR. However, as it has been demonstrated in [40] and in other related works, PSNR and QoE have low correlation. In this work, we use measures based on MOS, MIV calculates the percentage of frames with a MOS which worse than the reference. This is the encoded video without flow through the network transmission.

There are many works that use subjective evaluation methods as an alternative to the objective ones. 2D video subjective evaluation methods are quite standardized, but in 3D there are few works. The most common assessment methods are described in [8]: Double Stimulus Impairment Scale (DSIS), of the ITUR BT.500-13 (Rec 2012b), Double Stimulus Continuous Quality Scale (DSCQS), of the ITUR BT.500-13 (Rec 2012b), and the Absolute Category Rating (ACR), of the ITU-T P.910 (ITU-T 2008). The difference between the first and the second method is that the first method shows to the evaluators the reference video and then the distorted one, and then it is evaluated, while in the second method videos are shown randomly. The difference between the third method and the other two is that in this last one the reference video is not shown. Only the distorted video is evaluated.

In [41], there is a similar classification, which is only based on the stimulus. In this case, the evaluation methods are divided into Double Stimulus (DSIS), Single Stimulus (SS) and Stimulus

Comparison (SC). A part of the aforementioned methods in [8], two evaluation methods are added, Kaptein and Engeldrum, both specific for 3D evaluation. One of them is exclusively based on 3D quality features or attributes and the other one taking also account 2D features or attributes. In [30], authors only take into account Absolute categorical Rating (ACR) and Paired Comparisons (CP) subjective evaluation methods. In addition, the work presented in [28] includes an alternative method to DSIS, SS, and SC, which the method is called single-Stimulus-Continuous-Quality Evaluation as defined by (SSCQE). This method is based on a continuous gathering of quality judgments in time video sequences with variable time. Authors in their work are used Double Stimulus Continuous Quality Scale (DSCQS) methods to evaluate the video quality of end-users. Therefore, Table 1 and 2 show the main parameters of stereoscopic 3D evaluation that are used in our approach.

Table 1. Stereoscopic 3D video parameters

Video format	Codification	Quantization (QS)
Stereo Video	MVC	2,4 & 8
2D+Z	H.264	2,4 & 8

Table 2. Intelligent classification system parameters

Input metrics	Output metrics
Jitter, delay, packet loss, PSNR & MIV	3D video coding

4. ARCHITECTURE AND ALGORITHM TO IMPROVE STEREOSCOPIC VIDEO DELIVERY

In this section, we give detail the network architecture and the algorithm for delivering stereoscopic video over IP networks by considering QoS metrics and QoE parameters.

The most critical feature in a video transmission is that the video transmission should not be interrupted in order to have an adequate video quality. This reason makes to avoid interruption when the video is playing and it is more important than having lower video quality. However, for 2D video, the tradeoff between spatial quality and temporal quality is an open question and is often determined by the property of the content. The 3D specific visual artifacts further complicate this problem. Other important transmission features that may hugely affect the video quality are the delay and lost packets. Figure 3 explains the network architecture to implement the algorithm for delivering stereoscopic video over IP networks. Firstly, the protocols are used for transmitting video includes RTP/RTCP over UDP. This type of protocols can help to know the video sequences of the video flows sent by the video-streaming server. The architecture needs a video-streaming server to deliver the media flows to the end devices and decode the media flow. The video controller manages all delivery processes; it sends information to monitor the network and to manage the video-streaming server. Therefore, in order to design our algorithm, for this purpose we consider four phases: Initial phase, control phase, adaptation phase and improvement phases.

4.1. Initial Phase

The end user sends a request to the aggregation router of the distribution network in order to receive the stereoscopic content, an open flow capable router is used [42]. The router receives the request. It will accept or reject this request according its available resources, a regular situation is its acceptance, but if there are not enough available resources, the request is rejected. If there is a rejection, the router will send a message to the end user indicating “try it again later”. When the router accepts the request sent by the final user, it forwards this request with its available resource values to the video controller. According to the received parameters, the video controller selects which is the best coding type to be used by the video-streaming server. Then, video controller sends a message to video streaming server with the selected coding parameters. Finally, the Video Streaming Server sends to the final device (set-top box, computer, etc.) the appropriate stereoscopic video coding and the limits to send alarms (max. delay, max. jitter, max. lost rate, etc.) to the router. The initial phase protocol procedure is shown in figure 4.

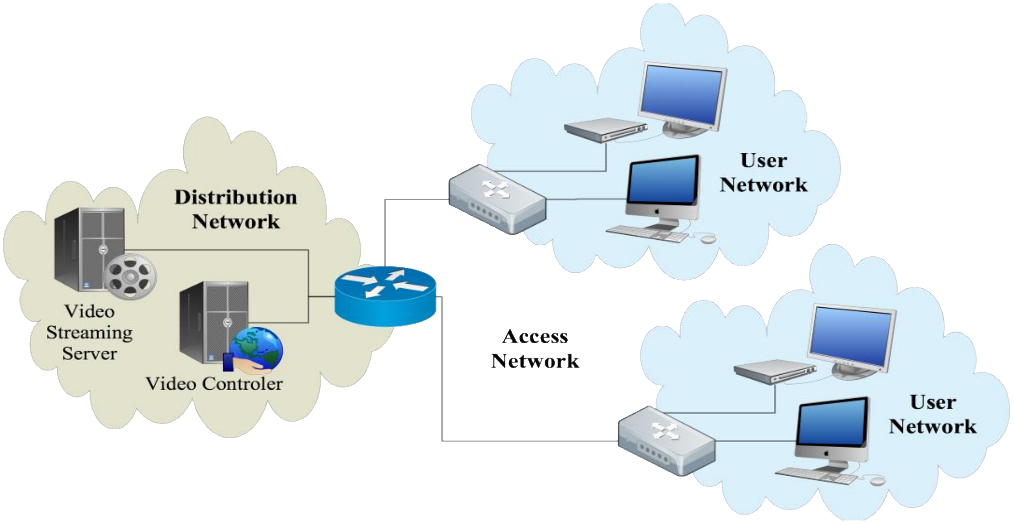

Fig. 3. Architecture of stereoscopic video delivery.

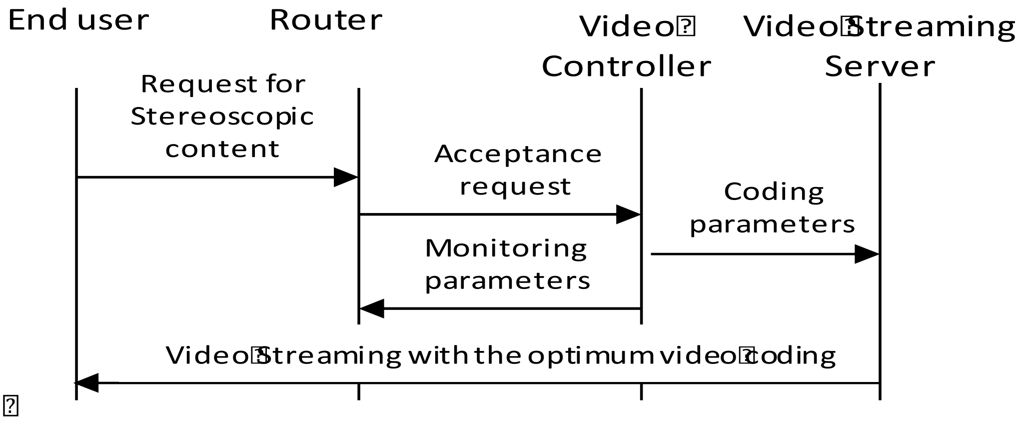

Fig.4. The protocol for the initial phase.

4.2. Control Phase

In this phase, the router estimates for each path the QoS parameters. The router and set-top boxes send alarms to the video controller when the jitter or delay of the received packets or the number of lost packets does not guarantee the video quality (as shown in previous works [43]). The controller will not be activated when it receives only one alarm from the router. In order to the controller is active, the control must receive three consecutive messages that contain the alarms with different value of the parameters.

4.3. Adaptation Phase

The video controller has to receive three consecutive alarms to start this phase. The video controller sends a message to the video-streaming server letting it know which the best coding configuration is for the current behavior of the network. The best coding configuration will depend on the jitter, delay, packet loss, and PSNR and MIV values and will be defined by the intelligent video controller decision algorithm. Then, the video-streaming server sends the appropriate coded video to the end user. Therefore, the adaptation phase protocol procedure is shown in figure 5.

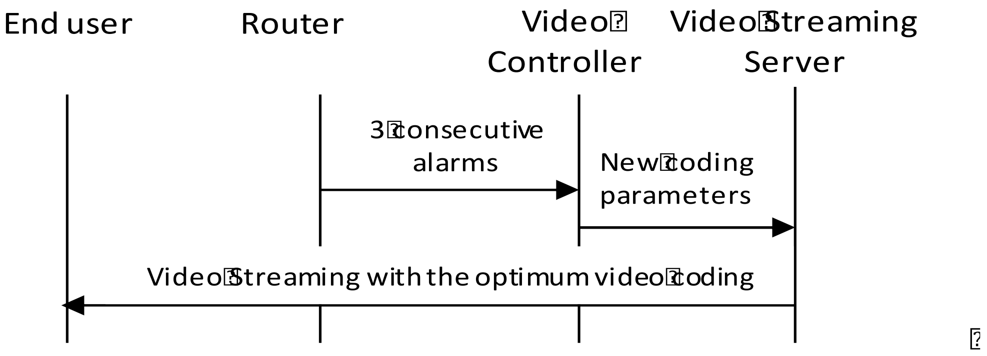

Fig. 5. The protocol for adaptation phase.

4.4. Improvement Phase

The parameters of QoS, determined by the delay and the packet loss, these metrics are affected the QoS and the QoE [43], so when the delay or packet loss rises a certain threshold an alarm is sent to the video controller. End user device (set-top-box, Personal Computer, tablets and etc.) is in charge to gather these values every predefined time. When any of them raise their threshold, an alarm is sent to come back to the initial state of the transmission. When the video controller receives an alarm, it requests a burst with a Group of Pictures (GOP) and the values of jitter, delay and packet loss to the device of the end user. In order to avoid having the burst affected by the network conditions, it is used the Low Latency Queuing (LLQ) mechanism. LLQ let us provide the maximum priority to the burst with the GOP. Once this information is received, video controller compares the original video fragment with the burst sent by the end user. From this process, PNSR and MIV parameters are estimated. Then, using these parameters together with jitter, delay and packet loss, video controller evaluates the best coding technique for improving the transmission. If a change is appropriated, the video controller indicates the new stereoscopic coding to the video-streaming server.

The stereoscopic IPTV delivery algorithm is summarized in figure 6. An end user requests a stereoscopic IPTV channel. Then, the devices in the access network will accept the connection only if they have enough available resources to support the connection, otherwise, the request will be refused. When a connection is accepted the checking process starts. It analyzes the main parameters of the video streaming in order to check its performance. The parameters that are going to be taken into account in the system studies are: jitter, delay and packet loss [43]. If the end user or the router detects that the network conditions do not guarantee the required QoS, then, an alarm is sent to the network manager. If this alarm is repeated three or more times during thirty seconds, then the network manager will request a burst of images to the end user. Requested packets will be marked at the video streaming transmitter and will be classified as a maximum priority by using LLQ. The receiver will forward the same marked packets to the video controller. This process is possible because it is a RTP transmission. Video controller will analyze the burst parameters: jitter, delay, packet loss, PSNR and MIV. According to these parameters, the video controller will take some smart decisions. It selects the adequate video streaming and provides the information to the video controller in order to take the appropriate decision for stereoscopic video streaming. The described process is included in the adaptation phase. The parameters that will be configured at this phase are video encoding and Quality Scale (QS). Our algorithm determines the optimum video codification, based on the aforementioned parameters. Finally, the video-streaming server delivers stereoscopic videos to the end users. Since the whole context is stereoscopic, the proposed algorithm is specific to stereoscopy. End user is a stereoscopic video user, video streaming server only transmits stereoscopic video and the video controller ensures quality of the received video at end user side taking the appropriate decisions. Alarms are mainly based on jitter, delay and packet loss parameters. The proposed algorithm comes from previous works [28][44] but it has been adapted to stereoscopic video and it includes the artificial intelligence in the decision algorithms.

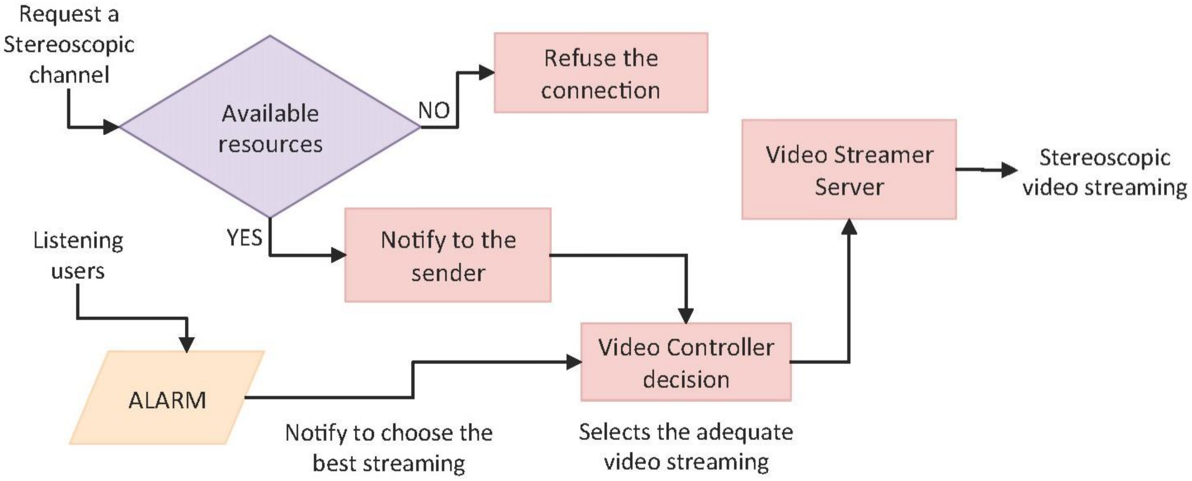

Fig. 6. Delivery algorithm.

5. VIDEO CONTROLLER DECISION ALGORITHM

Video controller decision algorithm uses a stereoscopic video coding classifier, which selects the best video coding according to some QoS network parameters, certain QoE parameters and class tags. It has the following stages: data preprocessing stage, training stage (based on a cluster algorithm) and classification algorithm.

5.1. Processing stage

The sequences are used to build the corpus, which included 6 videos, three with 2D+Z Coding Format and three with stereo video coding format. Each one was codified with QS of 2, 4 and 8 [45] and we have divided them in video clips of 60 seconds. Typical features (such as used codec, compression profile, average bitrate, fame size, frames per second (FPS), etc.) are attached to the video clips. This procedure let us to have 180 video clips to build our corpus, which is divided in 2 types of 3D stereoscopic formats for each quantized scale. All video clips had 25 fps. In 2D+Z format, and variable bitrate with average values of 5282kb/s, 2438kb/s y 926kb/s for QS=2, 4 and 8 respectively. In Stereo Video format, the bitrate is variable with average values of 9650kb/s, 4399kb/s and 1464kb/s for QS=2, 4 and 8 respectively. We build a corpus in the preprocessing stage based on supervised learning. Videos have been divided in several fragments to provide more samples. Fragments (and their size) were selected depending on the detection of changes produced in results of uniform and exponential jitter. Each coding has a class tag for each fragment. By transmitting each video in different jitter, delay and packet loss environment conditions, the best coding for each environment is tested, based on the results of the subjective evaluation of the QoE in MOS values and using DSCQS method. Jitter, delay and packet loss values, together with the end user PSNR and MIV, are used to build the characteristics vector. The coding that gives the best results in those conditions will provide the class tag of the characteristics vector. Then, we can build the corpus by using the characteristics vectors and the class tags. This corpus will be later used in the training phase. The corpus is formed principally by characteristic vectors and class tags. Class tags correspond to each 3D video coding technique while the characteristic vectors are formed by QoS parameter values such as jitter, delay, packet loss, PSNR and MIV. These data let us build X=$\{\mathrm{x}_1, \mathrm{x}_2 \ldots \mathrm{x}_5\}$ vectors, where each xi values is a QoS parameter value. Each sample is tagged with a class tag of one of the 3D coding format, 2D+Z and Stereo Video, and its QS. They are represented by K= $\{\mathrm{k}_1, \mathrm{k}_2 \ldots \mathrm{k}_6\}$. For the tagging, we gather the samples which results provide higher MOS. Thus, the obtained class tag is determined by the MOS value. Only the results with the highest MOS values will be used to guarantee the selection of the best coding format of each experiment. The information of the rest of samples will not be used to create the corpus. After these experiments, a characteristics vector for each codification is obtained and for each video clip. Then, a script will estimate which one of both 3D coding format has obtained highest MOS value for each video clip. This vector and its corresponding tag is saved in our corpus. This process is repeated during all experiments in order to build the corpus. We have gathered the corpus from a previous work [44].

5.2. Training stage

In the training phase, we use the corpus data to train the classifier. The set of samples that make up this corpus is given by $\mathrm{X} = \{\mathrm{X}_\mathrm{i}\}$, Where $\mathrm{i} = 1, \ldots, \mathrm{n}$. The classifier is based on the K-means algorithm [46], which uses Gaussian mixture model. The algorithm classifies the learning samples in K clusters, defined as $\mathrm{C} = \{\mathrm{C}_\mathrm{k}\}, \text{where } \mathrm{k} = 1, \ldots, \mathrm{K}$, The K-mean algorithm finds a partition such that it minimizes the quadratic error between the empirical cluster mean (μ_k) and the point of the cluster (C_k). The quadratic error defined in eq. (1). Moreover, it takes into account all clusters in eq. (2). Since the objective of the algorithm is to minimize the quadratic error then we will look for those values that minimize this error and define in eq. (3). Where, μ_k is the mean value of C_k samples, and X_i is each sample.

$$Y(C_k) = \sum_{X_i \in C_k} \left|\left|X_i - \mu_k\right|\right|^2 \qquad (1)$$

$$Y(C) = \sum_{k=1}^{K} \sum_{X_i \in C_k} \left|\left|X_i - \mu_k\right|\right|^2 \qquad (2)$$

$$argmin \sum_{k=1}^{K} \sum_{X_i \in C_k} \left| |X_i - \mu_k| \right|^2 \qquad (3)$$

The algorithm to perform the training task is described as follows: 1. Initiate the number of clusters with k=6. 2. Initiate the algorithm by selecting six random training samples that is determined the first centroids. 3. Estimation of the distance of every sample to the centroids and their minimum distance groups. Then, the mean value of the new clusters is estimated and the new centroids are obtained. 4. Step three is repeated until the system converges. The convergence is achieved when there are not new assignations or when a maximum number of iterations are defined. This algorithm lets us split the samples in 6 clusters, one for each 3D coding formats and QS. The clusters will be represented by a centroid. When the k-means algorithm is applied, we obtain a centroids matrix, which will be the base of our video controller classifier that will be used in the classification stage. The centroids matrix is given by:

$$C_{6,5} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,5} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,5} \\ \vdots & \vdots & \vdots & \vdots \\ c_{6,1} & c_{6,2} & \cdots & c_{6,5} \end{bmatrix} \qquad (4)$$

5.3. Classification Stage

After reading the data to be classified (video flow parameters), which will be represented as characteristics vector, taking into account the previously obtained centroids matrix, the distance between the vector and every centroid is estimated. The minimum distance will determine the closest centroid to this video flow. Every centroid represents a class and every class represents a 3D coding technique. The selected 3D coding format technique is determined by each class tag, which is corresponding to a centroid. Moreover, to estimate the distance between the sample and each of the centroids, and Euclidean distance. Therefore, given the sample can be $\mathrm{s} = (\mathrm{s}_1, \mathrm{s}_2, \dots, \mathrm{s}_\mathrm{n})$. Where the centroid $\mathrm{c}_{\mathrm{k,n}} = (\mathrm{c}_{\mathrm{k},1}, \mathrm{c}_{\mathrm{k},2}, \dots, \mathrm{c}_{\mathrm{k,n}})$, the distance between both vectors can be defined in eq.(5). Thus, the distance with respect to each of the centroids has been modeled in eq.(6). Moreover, The goal is to find that centroid that minimized the distance between the sample and centroid as described in eq. (7).

$$\begin{aligned} \mathrm{d}(\mathrm{s}, \mathrm{c}_{\mathrm{k,n}}) &= \sqrt{(\mathrm{s}_1 - \mathrm{c}_{\mathrm{k},1})^2 + (\mathrm{s}_2 - \mathrm{c}_{\mathrm{k},2})^2 + \cdots + (\mathrm{s}_\mathrm{n} - \mathrm{c}_{\mathrm{k,n}})^2} \\ &= \sqrt{\textstyle\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{s}_\mathrm{i} - \mathrm{c}_{\mathrm{k,i}})^2} \qquad (5) \end{aligned}$$

$$\mathrm{f}(\mathrm{c}_1, \mathrm{c}_2, \dots, \mathrm{c}_\mathrm{k}) = \sqrt{\textstyle\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{s}_\mathrm{i} - \mathrm{c}_{\mathrm{k,i}})^2} \qquad (6)$$

$$\underset{\mathrm{c}_\mathrm{k} \in \mathrm{C}}{\mathrm{argmin}} \ \mathrm{f}(\mathrm{c}_1, \mathrm{c}_2, \dots, \mathrm{c}_\mathrm{k}) \qquad (7)$$

The resulting centroid show that the codec is better suited to network and user conditions. The video controller decision algorithm is explained in figure 7. In short, video controller has a matrix formed by centroids. These centroids represent different encoding options under different network environments and objective estimates of the QoE. The system has representative values of jitter, delay, packet loss, PSNR and MIV of each analyzed coding format and QS values (2D+Z and Stereo with QS values of 2, 4 and 8). Real time classification process is next. During video streaming transmission, the video controller will be listening in order to receive alarms. When an alarm is received, it requests the user a burst using LLQ from the video sequence and the network parameters for this sequence. Video controller will compare this sequence with the original sequence and will estimate PSNR and MIV values. Then, it will join this information with the network parameter in order to build the characteristics vector. Video controller will estimate the distance to each centroid using this vector. The shortest distance determines the optimal coding (taking into account the received network characteristics vector). When the network conditions get worse, an alarm is activated and the received bitrate is requested to the end user in order to estimate the PSNR and MIV.

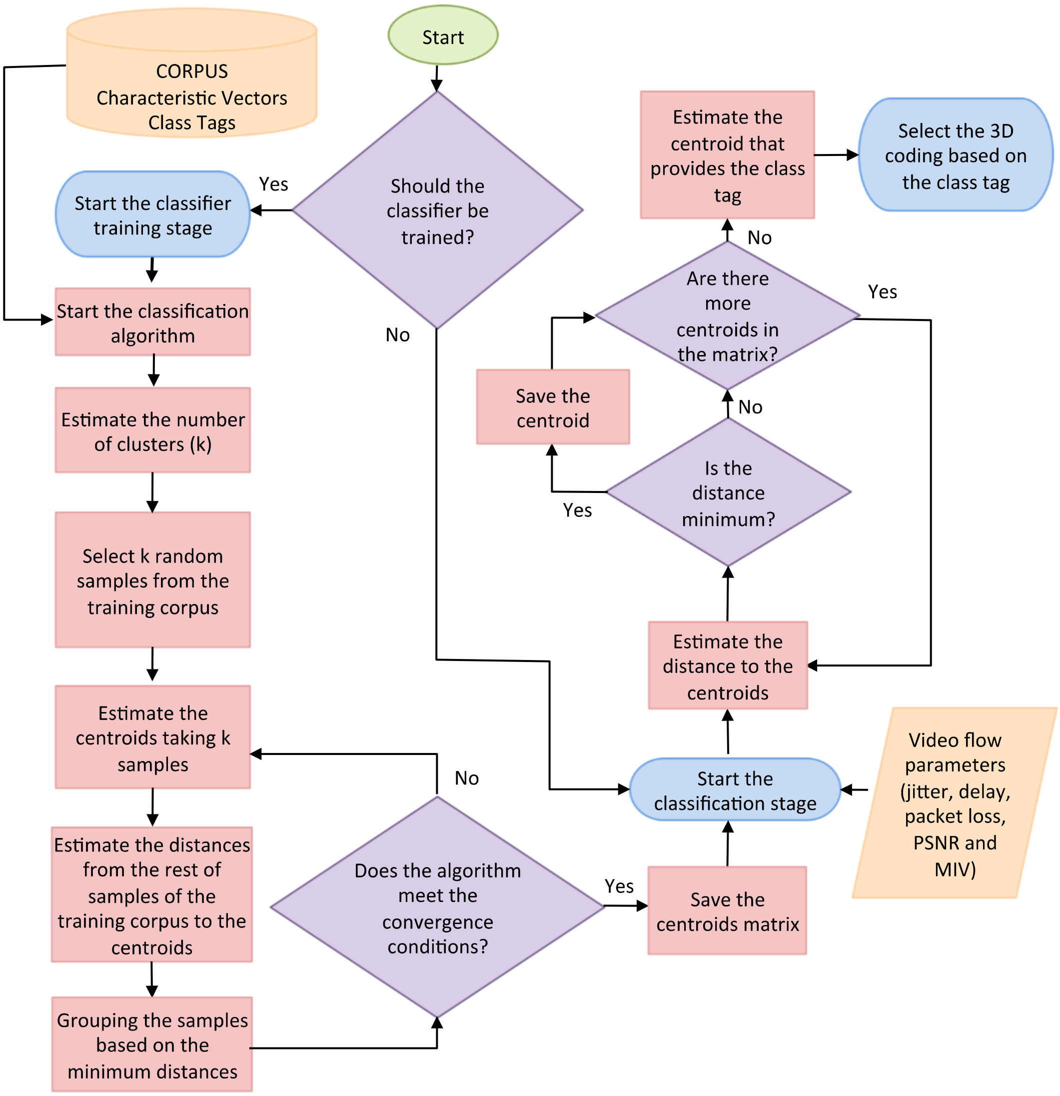

Fig. 7. Decision algorithm.

6. TRAINING PHASE AND TAGGING PROCESS

In this section, we show the tests performed to train the system. It includes the delay, jitter, packet loss, PSNR and MIV tests for two 3D video coding techniques using three types of QS= 2, 4 and 8. After coding the video fragments with different QS values, the observations show that the superior QS8 values do not provide enough quality.

In order to achieve our goal, an IP network is provided and processed 3 raw video sequences for Stereo Video format and 3 more for 2D+Z format for all QSs. We used RTP for video delivery. In order to assess the QoS parameters and know which coding parameters must be proposed to the network management algorithm for 3D IPTV optimum delivery, the analysis process is divided into several stages. First, we created different packet loss for each video sequence. Second, both uniform and exponential jitter, and plus delay with the same range is added to the phase. They are described in detail in [47]. First we varied the jitter in the network for each coded video. Although jitter is not a relevant parameter due to the use of buffering techniques (generally used in almost all multimedia players and IPTV setup boxes), it is very helpful to study it in order to know the best configuration for the test bench. The jitter was uniform and exponential with a constant delay of 30 ms. Uniform jitter value varies following eq. 8.

$$F(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \alpha < 0 < \beta \\ 0 & x < \alpha \quad or \quad x > \beta \end{cases} \qquad (8)$$

We chose α=1 and β=20 because we think that it is the most appropriate range to avoid too spread values. The function only takes values between alpha and beta. In addition, an exponential jitter is simulated for whose probability of the density function is given by eq. 9. In this case, λ=0.1.

$$F(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \quad or \quad x = 0 \\ 0 & x < 0 \end{cases} \qquad (9)$$

In order to introduce different network behaviors we used Netdisturb [48]. It let us vary the delay based on different jitter behaviors such as exponential or uniform jitter. Next subsections provide delay, jitter, packet loss, PSNR, MIV and MOS mean values when our algorithm selects both 3D coding techniques in different cases such as uniform and exponential jitter, and 0.1% and 0.01% of packet loss (this allows us to compare each decision). Instantaneous values of some part of this test are published in [44].

6.1. Training Test Bench and Experiments

Our test bench topology is shown in figure 8. The end user is connected to a set top box, which is connected to a Personal Computer acting as a router. We describe the process to create the 2D+Z and stereo video sequences used in this experiment. First, three raw video sequences (Left view, Right view and 2D depth view) with 4:2:0 YUV format to build the 2D+Z and Stereo video sequence are used. They were downloaded from [49]. 2D+Z sequences are compressed using H.264 Auxiliary Picture Syntax. Stereo Video sequences have been coded using H.264/AVC Multi-View Video Coding (MVC) using JMVC (Joint Multi-view Video Coding) software [26] both with different quantization scales (QS2, QS4, QS8) and with the following features: 1,080x720 progressive scanning (720) at 25 fps. The number of frames shown in the figures of this section will depend on the detail needed for our purposes. Therefore there are graphs of 140, 180 or 200 frames, depending on where the transmission changes are produced. Table 3 explained the comparison of different format.

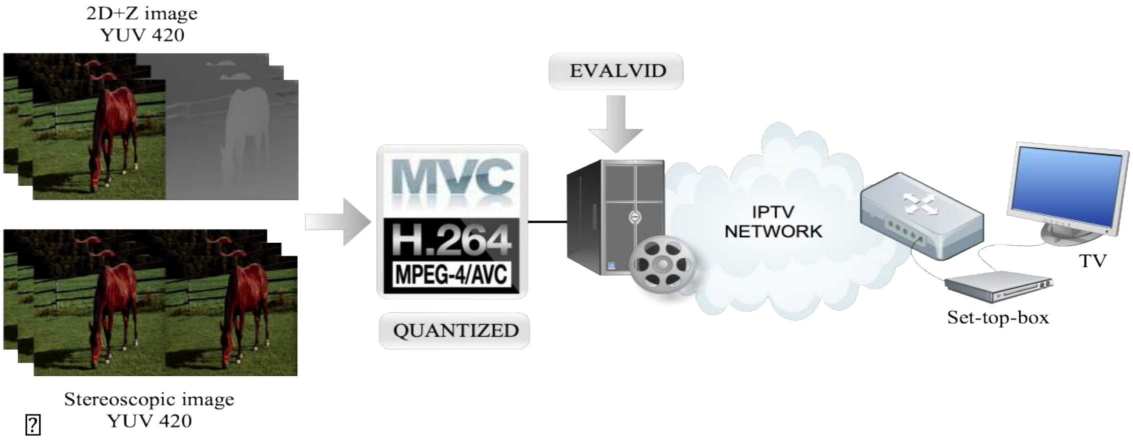

Fig. 8. Encoding process and test bench.

Table 3. Comparison of different formats in kb/s.

Video sequences	Size QS2	Size QS4	Size QS8
Stereo Video	9650	4399	1464
2D+Z	5282	2438	926

Both videos have been encapsulated to MP4 using FFMPEG [50]. Table 3 shows the KB/s of each used for both 3D video formats. It is important to note that a Stereo Video format does not present a great reduction when compared with 2D+Z, which fluctuates between 42% and 50% less than a Stereo Video. This is because half of the coded information in a 2D+Z sequence corresponds to the depth of the sequence in gray scale, which has high spatial and temporal redundancy. It is very useful when it is needed to reduce the amount of the video stream to be sent. VQM [51] was proposed to provide an objective measurement for perceived video quality for 2D. It measures the perceptual effects of video impairments including blurring, jerky/unnatural motion, global noise, block distortion and color distortion, and combines them into a single metric. Following the indications provided in [52], the video objective evaluation metric VQM can be used to evaluate 3D stereoscopic QoE although it does not have very high correlation. VQM takes the original video and the processed video as input and provides the output by using a linear combination of the quality parameters. To estimate VQM, Video Quality Estimator is used. Figure 9 shows the VQM using QS2, QS4, and QS8. QS closer to zero means higher quality coded video. In this case, the obser can see that the videos with QS2 have a good quality. In both encoding systems, the VQM obtained is lower than 1. When we use 2D+Z with a quantizer equal to 4, the VQM is still less than 1, but it is above of the coding systems with QS2. Stereo Video coding with QS4 and QS8 is used, the video quality is very poor. This also occurs with 2D+Z with a QS8. The worst coding system is Stereo Video coding with QS8, because its quality is very low and it does not have the best coding gain.

6.2. Jitter test

In figures 10, 11 and 12 we evaluate the behavior of the jitter when there is both uniform jitter and exponential jitter for 2D+Z and Stereo Video coding. In Figure 10, we analyze the jitter received when the video is encoded with QS2 for both 2D+Z and Stereo Video coding. Between the 10th and 125th frame the algorithm selects 2D+Z when there is uniform jitter because it has lower values. It has an average values close to 0.005 ms. in the test, 2D+Z with exponential jitter is the one with worst jitter values. At the end of the graph the uniform jitter provides better jitter for both 3D coding techniques than exponential jitter. The jitter received by the end user using stereoscopic videos with QS4 is shown in Figure 11. For almost all frames both 3D coding techniques are good when there is an exponential jitter. But when there is a uniform jitter, after 10th frame, Stereo Video coding performs better jitter. After 47th frame both 3D coding techniques in exponential jitter and Stereo Video coding with uniform jitter have similar jitter values. Figure 12 provides the jitter received by the end user for stereoscopic videos with QS8. We can see that from 0 to the 45th frame the best option is 2D+Z when exponential jitter is used. From the 85th frame to the end all cases perform similar jitter values (although the lowest ones are those with exponential jitter). Analyzing all figures of the jitter test, the jitter is lower when QS increases.

6.3. Delay Test

Figures 13, 14 and 15 show the delay received by the end users when the network has uniform and exponential jitter. In Figure 13, we evaluate both 3D coding techniques using QS2. We can see that 2D+Z coding technique provides lower delay in both types of jitter starting from the 12th frame to the end of the graph. The worst value is obtained for Stereo Video coding in the uniform jitter case. Figure 14 shows the delay received by the end user for stereoscopic videos with QS4. All case show similar values, but the lowest ones are provided in the exponential jitter case. From 0 to the 65th frame the lowest delay values are given in the Stereo Video coding for the exponential jitter, while from the 65th frame to the end the lowest delay is given by 2D+Z in for the exponential jitter. The delay received by the end user for stereoscopic video with QS8 is shown in Figure 15. From 0 to the 47th frame, the lowest value is given by 2D+Z in the exponential jitter case, while in the uniform jitter, both 3D coding techniques perform quite similar behavior. After the 47th frame the lowest value is provided by the Stereo Video coding in the exponential jitter, but all the others have close values. When we have a network that introduces jitter (uniform or exponential), the 2D+Z encoding technique introduces less delay and less fluctuations in video transmission. However, QS8 provides lower delay values than QS4 and QS2.

6.4. Packet Loss Test

Figure 16, 17 and 18 show the packet loss received by the end users when there is between 0.1% and 0.01% of packet loss in the network for uniform jitter and exponential jitter. In Figure 16, we see the packet loss received by the end user for both 3D coding techniques using QS2. 2D+Z coding technique provides lower packet loss in both cases. From 0 to approximately the 50th frame, there is lower loss percentage for exponential jitter. While, from 50th frame to the end uniform jitter have lower loss percentage. Figure 17 shows the packet loss for both 3D coding techniques using QS4. Both have lower packet loss percentage when uniform jitter. Generally, exponential jitter provides less packet loss at the end user side. It is emphasized after the 40thframe.

6.5. PSNR Test

The PSNR (Peak Signal-to-Noise Ratio) received by the end user is analyzed in figures 19, 20 and 21. In figure 19 we see the received PSNR (in dB) when the video delivered is sent using QS2. In both cases, uniform and exponential jitter, 2D+Z coding is the chosen technique that provides higher PSNR, so it is preferred. The best case is given for 2D+Z in exponential jitter and the worst case is given for Stereo Video in uniform jitter. In figure 20 we observe a similar behavior than in Figure 19, but in this case the average PSNR is lower because a higher quantizer is used, so video quality sent from the video-streaming server is lower (although the PSNR value is still higher than 35 dB when we used 2D+Z). 2D+Z is the preferred coding technique for both uniform and exponential coding PSNR received (dB) by the end user for stereoscopic videos with QS8 is shown in figure 21. The observations show t that 2D+Z coding is always better in both uniform and exponential jitter. But when we compare them in terms of the type of jitter, we observe that before the 105th frame, exponential jitter provide better results for almost always. In figures 19, 20 and 21, we can observe that the uniform jitter introduces a worse performance of the received PSNR by the end-user.

6.6. MIV Test

MIV (Moving Image Videography) is the maximum percentage of frames with a MOS worse than original. Figures 22, 23 and 24 show the MIV for QS2, QS4 and QS8. Figure 22 shows MIV received by the end user for both 3D coding techniques with QS2. As shown that 2D+Z provides better results in both cases, uniform and exponential jitter. Figure 23 shows the MIV received by the end user for both 3D coding techniques using QS4. 2D+Z is the best 3D coding technique in both uniform and exponential jitter. Figure 24 shows the same behavior than in the previous graphs for QS2 and QS4. 2D+Z-coding technique provides better results.

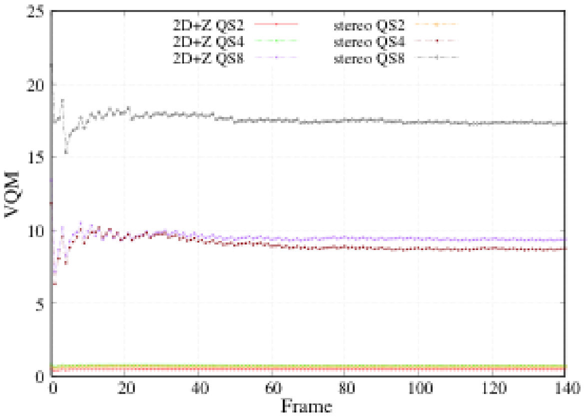

Fig.9. Objective measurement.

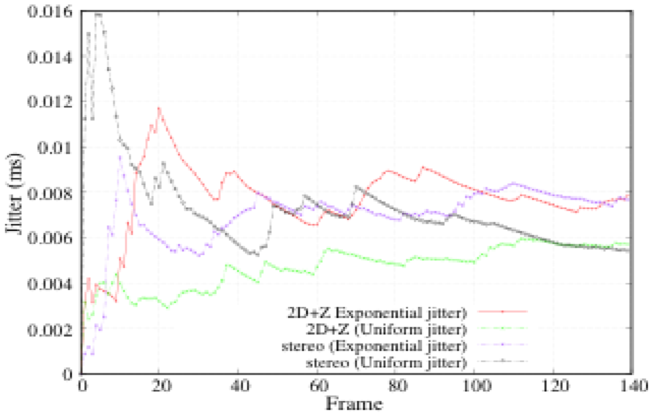

Fig.10. Jitter received by end-user with QS2

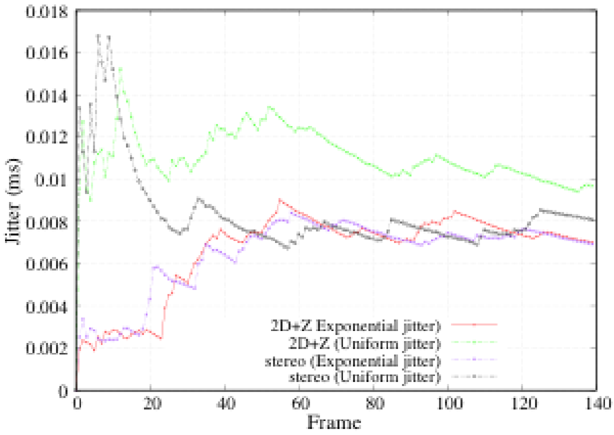

Fig.11. Jitter received by end-user with QS4

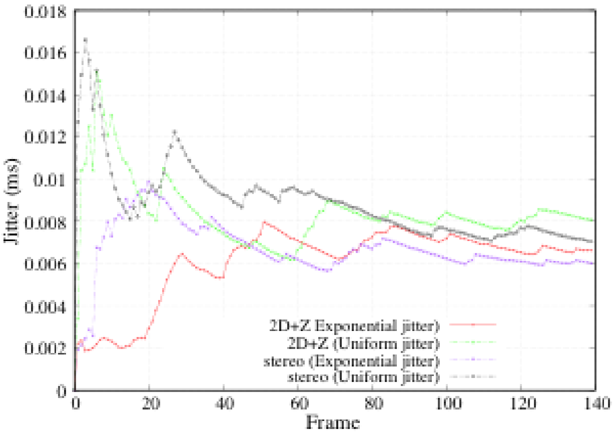

Fig.12. Jitter received by end-user with QS8

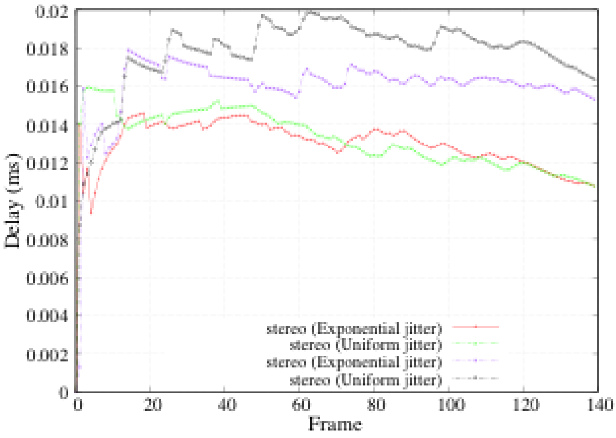

Fig.13. Delay received by end-user with QS2

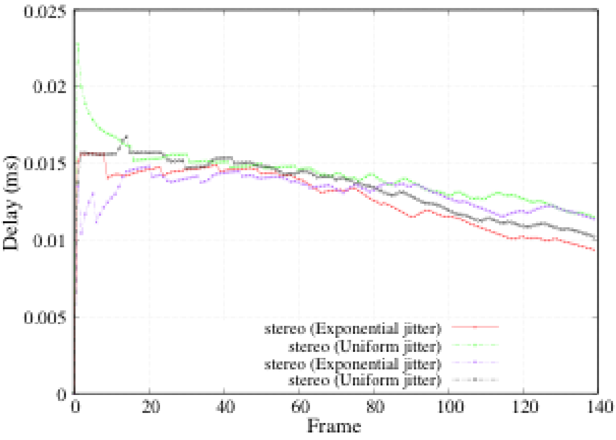

Fig.14. Delay received by end-user with QS4

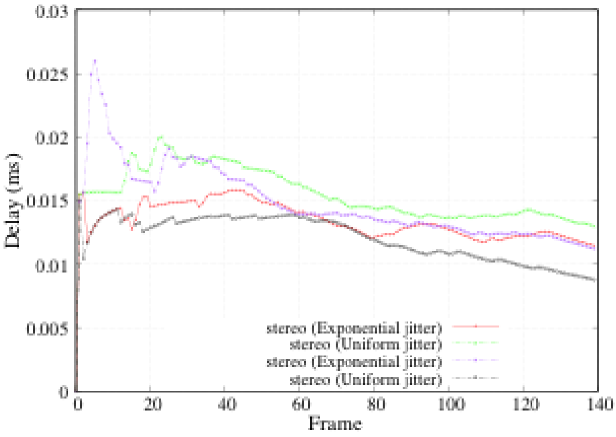

Fig.15. Delay received by end-user with QS8

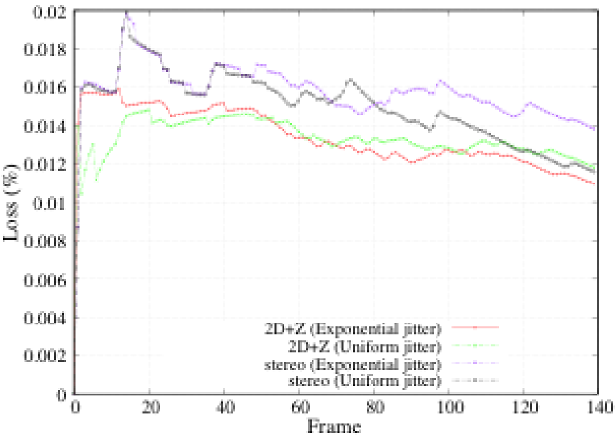

Fig.16. Packet loss by end-user with QS2

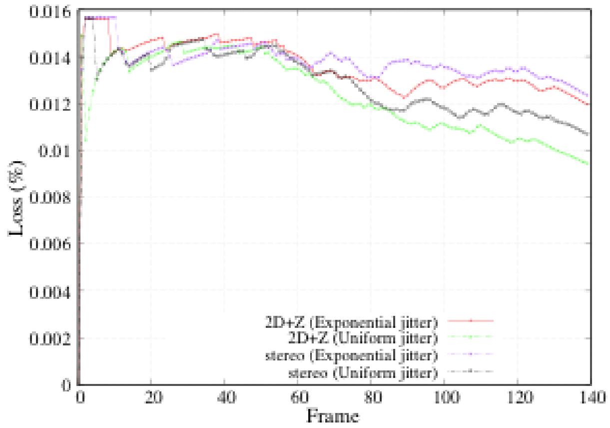

Fig.17. Packet loss by end-user with QS4

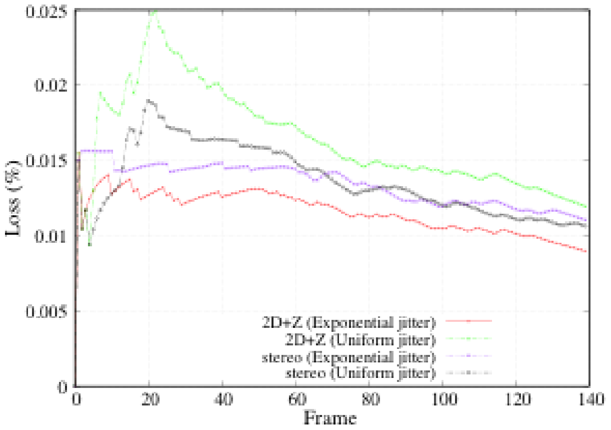

Fig.18. Packet loss by end-user with QS8

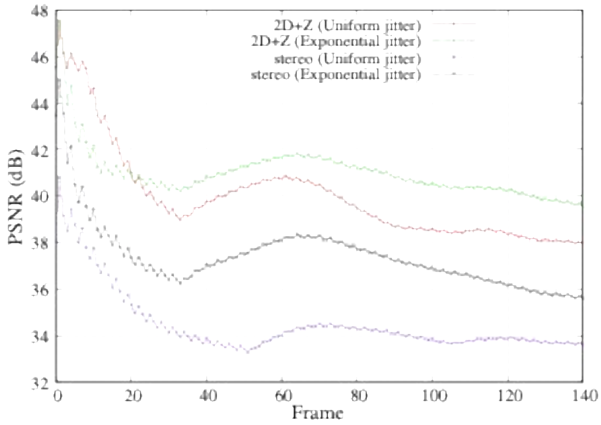

Fig.19. PSNR by end-user with QS2

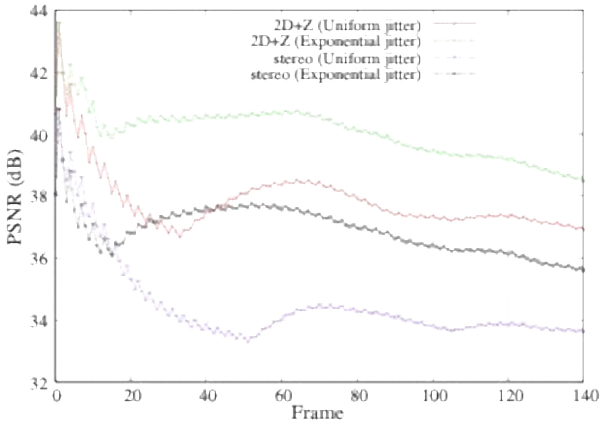

Fig.20. PSNR by end-user with QS4

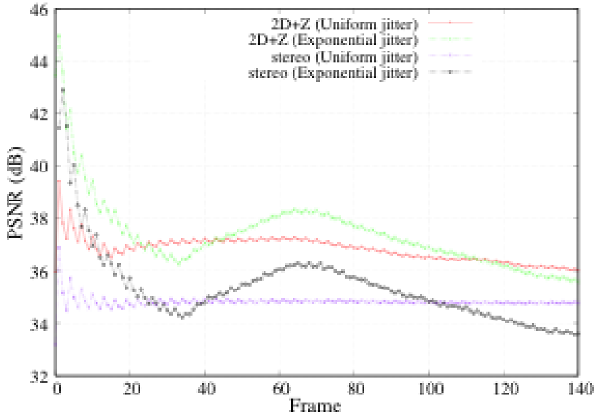

Fig.21. PSNR by end-user with QS8

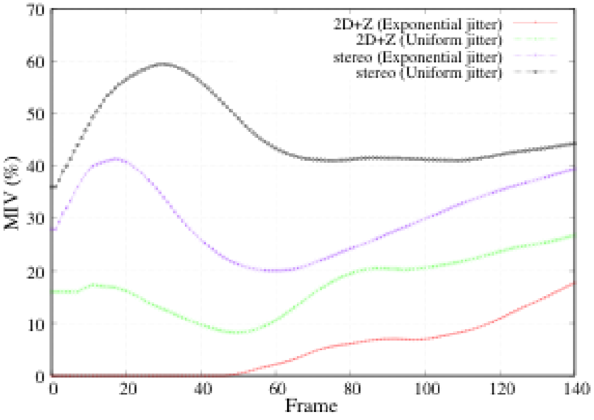

Fig.22. MIV by end-user with QS2

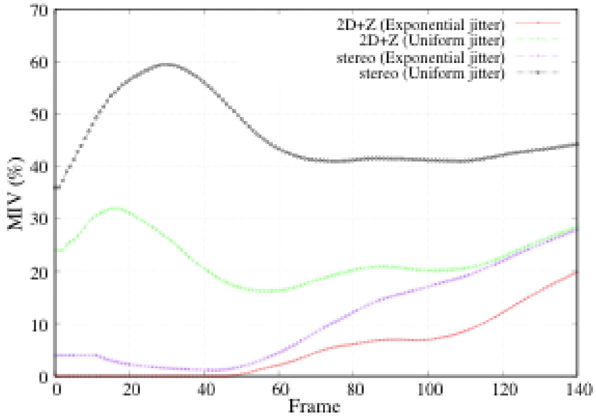

Fig.23. MIV by end-user with QS4

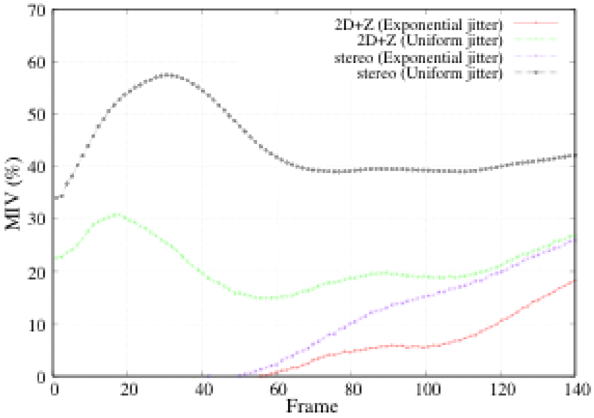

Fig.24. MIV by end-user with QS8

6.7. Subject evaluation of the QoE and samples tagging

In this subsection, we present the evaluation of the QoE by the end users for each video clip. In order to perform the evaluation, we selected five people, which are inexpert in image processing. The method used in our evaluation is Double Stimulus Continuous Quality Scale (DSCQS), where, each inexpert sees both the reference and the distorted video clips then a continuous scale from 0 to 5 is used to measure the perceived quality, which is divided into five levels: Bad, Poor, Fair, Good and Excellent. The evaluation parameters taken into account by the reviewers for the 3D stereoscopic videos have been: image quality, sharpness, naturalness and depth perception. For the quality image, the parameters taken into account have been the same than for 2D (blur, ghosting, pixilation, etc.). This provides a value for each attribute of each 3D stereoscopic video clip. In this study each attribute has the same weight. The final QoE value is the mean value in the MOS scale. The quality of the distorted video will be conditioned by the quality of the reference video. Table 4 shows an example of the class tags in exponential jitter (EJ) and uniform jitter (UJ) for 2D and stereo with 0.1% of packet loss in the transmission. Once we have seen all cases for each video and for each network condition, we will use as a class tag that one which codification corresponds with the closest MOS value perceived subjectively. In order to train the decision algorithm, the system saves the class tag with the jitter, delay, packet loss, and PSNR and MIV features. Figure 25 shows a typical visual error given in video transmission. This particular case belongs to a transmission with uniform jitter, QS2, and 0.1% of average losses. This error in the quality image is called "Blur" effect. It is a perception effect related with 3D stereoscopic videos. We want to highlight that the values obtained for the perceived depth have been low. But they were also low for reference video clips as shown in the table. Thus, they are not low because of transmission problems. We observed that the sharpness and naturalness are highly affected when the quality of the image decreases.

7. STEREOSCOPIC VIDEO DELIVERY PERFORMANCE TEST

In this section we provide the test performed to check the proper running of the proposed algorithm. First, using the samples the system starts the training process. Then, it builds the centroids matrix, which is used by the controller jointly with a characteristics vector. The characteristics vector is depended on the network conditions and the GOP burst of the distorted video when the alarm is sent. Finally, the algorithm estimates the best codification to guarantee the QoE. In order to perform the test, we transmit six video clips in different network conditions, in terms of jitter, delay and packet loss, to the training ones. The QoE is obtained by the evaluation of several voluntaries, which analyze subjectively the video clips received when there is an alarm. This let us know the best codification for each case and the results of the classification algorithm. The results of delivering video stereoscopic are shown in figure 26. The classification algorithm is quite accurate when the best codifications are selected for each case (in terms of MOS and network conditions).

Fig. 25. Perceived visual errors: Left image belongs to the reference video, Right image belong to the distorted.

Table 4. Perceived quality by the end-users.

		QS2			QS4			QS8		
		Clip1	Clip2	Clip3	Clip1	Clip2	Clip3	Clip1	Clip2	Clip3
2D	UJ/Loss 0.1%	3.71	4.15	3.75	4.12	4.09	3.80	4.12	4.02	3.91
	EJ/ Loss 0.1%	4.03	4.11	4.52	4.49	3.97	3.95	4.00	2.98	3.78
Stereo	UJ/ Loss 0.1%	3.53	3.76	3.63	4.02	3.70	3.64	4.21	3.63	3.67
	EJ/ Loss 0.1%	4.41	4.34	3.57	4.32	3.89	3.77	3.83	3.82	3.53

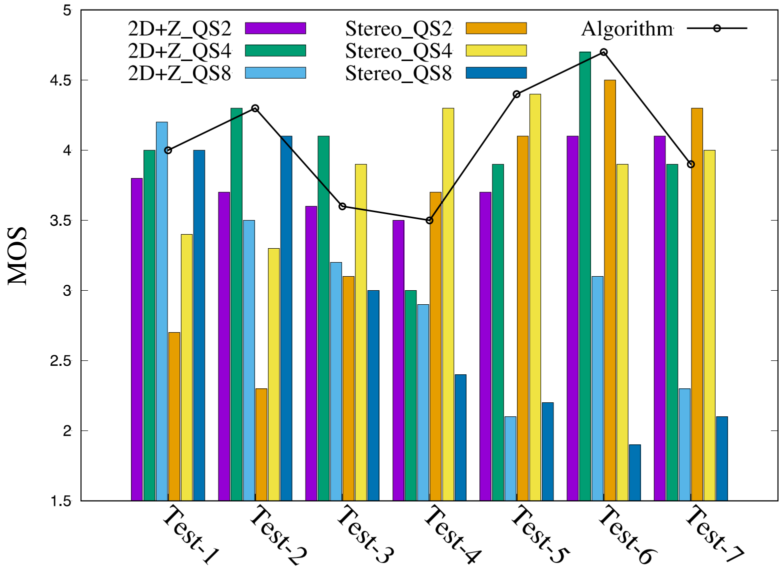

Fig. 26. Test results.

8. CONCLUSIONS AND FUTURE WORKS

In this paper, analyze 2D+Z and Stereo Video formats for stereoscopic videos are studied. We propose a cognitive algorithm and protocol to enhance stereoscopic video delivery based on the QoS network parameters therefore, several types of 3D coding formats are used to deliver the best stereoscopic video separately. We compare all possible cases for both uniform and exponential jitter and for both 0.1 and 0.01 packet losses. Moreover, the delay, PSNR, and MIV are included in the training corpus.

Subjective methods are high cost, so accurate objective methods for QoE estimation are needed. The tests performed for QoE estimation have provided good results although we used PSNR and MIV (related literature indicate that they do not have good correlation with QoE when 3D stereoscopic is delivered). In this analysis we can observe that each technique performs better than the other depending on each QS. Generally in QS2, when there is uniform jitter in the network, 2D+Z is selected in the algorithm, because it provides better results. If PSNR and MIV were only considered, 2D+Z would be always selected. Stereo Video coding is always selected in QS4 when there is uniform jitter in the network and in some cases of QS8. Moreover, we can say that the Stereo Video coding technique is the least swayed by the type of jitter. From the figures of the jitter test, we can observe it in the exponential jitter, lower jitter values is obtained when QS increases. But we have not observed any relationship between the types of 3D coding technique and jitter. Thus depending on the type of jitter, QS or packet loss, the algorithm will select different a 3D coding technique.

The development and integration of the proposed system in cognitive networks can have an important applicability: provides better QoE to end-user in delivering Stereoscopic IPTV. Moreover, The cognitive networks designed by taking the characteristic of 3D Coding Formats and the important input parameters to the system. From these parameters, the proposed system estimated the QoE and thus improves the delivery of Stereoscopic IPTV.

Our future work will focus on adding some other stereoscopic video codecs to provide experiments by using our algorithm and to test its performance when multiple types of data are being delivered in the IPTV network. Also we will reduce size of the characteristics vector by adding Principal Component Analysis (PCA) procedures, to improve the classification results. Furthermore, more sophisticated methods will be included such as direct Gaussian mixtures or neural networks classification methods. We will use virtualized network emulators as utilized by [53] to provide testing of stereoscopic IPTV videos in virtualized environment too. Moreover, we will also analyze which 3D stereoscopic factors affect to the quality of the QoE and include new features in the characteristics vector, such as PQM (Perceptual Quality Metric). Finally, we will conduct to consider all entry parameters to the system as well as those parameters related to perception of quality from the user that have the same weight. We aim to provide an interesting aspect that models the results better is to consider them with different weights.

ACKNOWLEDGMENT

This work has been partially supported by the “Ministerio de Ciencia e Innovación”, through the “Plan Nacional de I+D+i 2008-2011” in the “Subprograma de Proyectos de Investigación Fundamental”, project TEC2011-27516, and by the Polytechnic University of Valencia, though the PAID- 15-11 multidisciplinary projects.

REFERENCES

[1] R. A. Santos Cruz, M. Serafim Nunes, L. Menezes, J. Domingues. “IPTV architecture for an IMS environment with dynamic QoS adaptation.”, Multimedia Tools and Applications, Vol.53, no.3, May 2010, pp.557-589.

[2] J. Arnaud, D. Négru, M. Sidibé, J. Pauty and H. Koumaras. “Adaptive IPTV services based on a novel IP Multimedia Subsystem.”, Multimedia Tools and Applications, Vol. 55, no. 2, Sept. 2010, pp. 333-352.

[3] ETSI TR 102 643, Human Factors; Quality of Experience (QoE) Requirements for Real-time Communication Services, Sophia Antipolis, France, 2009. Accessed July 1, 2017. [Online]. Available: https://www.scribd.com/document/203104445/ETSI-TR-102-643.

[4] Naor, Moni, and Udi Wieder. "Novel architectures for P2P applications: the continuous-discrete approach." ACM Transactions on Algorithms (TALG) ,3, no. 3 (2007): pp.34.

[5] H. Kalva, L. Christodoulou, Mayron, O. Marques, B. Furht. “Challenges and Opportunities in Video Coding for 3-D TV”., IEEE Int. Conf. on Multimedia & Expo (ICME) 2006, Toronto, Canada, Jul. 9-12, July 2006, pp. 1689 – 1692.

[6] Van Den Ende, N., De Hesselle, H. and Meesters, L., 2007, May. Towards content-aware coding: User study. In European Conference on Interactive Television (pp. 185-194). Springer, Berlin, Heidelberg..

[7] I. Taboada, J.O. Fajardo, F. Liberal, Performance Analysis of Scheduling Algorithms for Web QoE Optimization in Wireless Networks, Network Protocols and Algorithms, vol.4, no. 4. 2012. pp.27-43.

[8] K. Wang, “Stereoscopic 3D Video Quality of Experience : impact of coding, transmission and display technologies.” [Thesis]. Sundsvall: Mid Sweden University; June 2013. Mid Sweden University licentiate thesis. Accessed July 1, 2017. [Online]. http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A625187.

[9] A. Smolic. D30.3 Technical Report #3 on 3D Coding Techniques. Project Number 511568. 3DTV research (2008). Accessed July 1, 2017. [Online]. http://www.3dtvresearch.org/publicDocs/techReports08/D30.3_Public.pdf.

[10] A. M. Tekalp. D32.3 Technical Report # 3 on Transport Methods for 3DTV.Project No 511568. 3DTV research (2008). Accessed July 1, 2017. [Online]. http://www.3dtv-research.org/publicDocs/techReports08/D32.3_Public.pdf

[11] A. Álvarez, L. Pozueco, S. Cabrero, X. García Pañeda, R. García, D. Melendi, G. Díaz Orueta, “Adaptive Streaming: A subjective catalog to assess the performance of objective QoE metrics”, Network Protocols and Algorithms, Vol. 6, no. 2, 2014, pp. 123-136.

[12] C.T.E.R. Hewage, H.A. Karim, S. Worrall,S.Dogan, and A.M. Kondoz. “Comparison of Stereo Video Coding Support in MPEG-4 MAC, H.264/AVC and H.264/SVC.”, Visual Information Engineering-VIE07, London, UK, July 2007.

[13] J. R. Ohm. Stereo/multiview encoding using the MPEG family of standards. In Electronic Imaging'99, San Diego, CA, January 1999.

[14] Chu, Chung-Hua. "Visual comfort for stereoscopic 3D by using motion sensors on 3D mobile devices." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Vol.12, no. 1s, (Oct. 2015), pp. 14..

[15] Kiana Calagari, Tarek Elgamal, Khaled Diab, Krzysztof Templin, Piotr Didyk, Wojciech Matusik, and Mohamed Hefeeda. "Depth personalization and streaming of stereoscopic sports videos." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 12, no. 3, (March 2016), pp. 41..

[16] G. Petrovic, D. Farin and Peter HN de With. “Toward 3D-IPTV: design and implementation of a stereoscopic and multiple-perspective video streaming system.” Proc. of the SPIE 2008, International Society for Optics and Photonics, Vol. 6803, Feb. 2008.

[17] G. Saygili, C.G. Gurler, A.M. Tekalp, “3D display dependent quality evaluation and rate allocation using scalable video coding.” 16th IEEE Int. Conf. on Image Processing 2009, 7-10 Nov. 2009, pp.717-720.

[18] J. Lloret, M. Garcia, A. Canovas, C. Turro, “A stereoscopic video transmission algorithm for an IPTV network based on empirical data.”, Int.J.of Communication Systems, Vol. 24, no.10, Oct. 2011, pp. 1298–1329

[19] A. N. El-Yamany, K. Ugur, M. M. Hannuksela, M. Gabbouj, Evaluation of depth compression and view synthesis distortions in multiview-video-plus-depth coding systems. Proceedings of 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON 2010), Tampere, Finland, June 7-9, 2010.

[20] Young-il Kim, G.Khishigjargal, Soon Young Song, Yong Su Lee, Park Dae Geun, Chun Sun Sim, Park Young Soo, Yeonjoon Chung, Cheol-Hye Cho, Won Ryu, Seong Young Seo, Suk Chan Kim, N. Chuluunbandi, "3D image transmission technology for mobile IPTV," International Conference on Information and Communication Technology Convergence (ICTC2010), 17-19 Nov. 2010, pp.280-283.

[21] Young-uk Chung. Symmetrical Frame Discard Method for 3D video over IP Networks. IEEE Transactions on Consumer Eletronics, vol. 56, no. 4, November 2010, pp. 2790-2796.

[22] W. Matusik and H. Pfister. 3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. ACM Trans. Graph. Vol. 23, no. 3, August 2004, pp. 814-824.

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

[24] K.-J. Oh, M. Kim, J. Yoon, J. Kim, I. Park, S. Lee, C. Lee, J. Heo, S.-B. Lee, P.-K. Park, S.T. Na, M.-H. Hyun, J. Kim, H. Byun, H. Kim, Y.-S Ho. “Multi-View Video and Multi-Channel Audio Broadcasting System.” 3DTV Conference, Kos Island, Greece, May 2007, pp. 1-4

[25] Philips 3D Solutions, Accessed July 1, 2017. [Online] http://www.business-sites.philips.com/3dsolutions/home/index.page).

[26] A. Hallapuro, M. Hannuksela, J. Lainema, M. Salminaa, K. Ugur, K. Willner, Y. Wang, Y. Chen, “3D Video & The MVC Standard”, Nokia Research Center, May 2009. Accessed July 1, 2017. [Online] http://research.nokia.com/page/4988.

[27] P. Brooks and B. Hestnes, User Measures of Quality of Experience: Why Being Objective and Quantitative is Important, IEEE Network, vol. 24, no. 2, Mar.–Apr. 2010, pp. 8–13.

[28] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiens, “A survey of perceptual evaluations and requirements of three dimensional TV”, IEEE Trans. Circuits Syst. Video Technol. Vol. 14, 2004, pp.381–391.

[29] A. Benoit, P.L. Callet, P. Campisi, R. Cousseau, “Quality assessment of stereoscopic images,” EURASIP Journal Image and Video Processing, Vol. 6590, 2008, doi:10.1155/2008/659024

[30] E. Bosc, R. Pépion, P. L. Callet, M. Köppel, P. Ndjiki-Nya, M. Pressigout, L. Morin. “Towards a new quality metric for 3-D synthesized view assessment.” IEEE J Sel Topics Signal Process, vol. 5, no. 7, 2011, pp.1332–1343.

[31] A. B. Watson, A. J. Ahumada. “64.3: Flicker Visibility: A Perceptual Metric for Display Flicker.”, SID Symposium Digest of Technical Papers. V. 42, I. 1, pp. 957–959, June 2011.

[32] Ingrid Heynderickx, Ronald Kaptein, “Perception of detail in 3D images.”, Proc. SPIE, vol. 7242. 2009.

[33] Z. Wang and A. C. Bovik, “A universal image quality index,” IEEE Signal Process. Lett., vol. 9, no. 3, pp. 81–84, Mar. 2002.

[34] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, “New full-reference quality metrics based on HVS,” in Proc. 2nd Int. Workshop Video Process. Quality Metrics, Scottsdale, AZ, 2006.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, April 2004.

[36] D. M. Chandler and S. S. Hemami, “VSNR: A wavelet-based visual signal-to-noise ratio for natural images,” IEEE Trans. Image Process., vol. 16, no. 9, , Sep. 2007, pp. 2284–2298.

[37] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” IEEE Trans. Image Process., vol. 15, no. 2, , Feb. 2006, pp. 430–444.

[38] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An information fi- delity criterion for image quality assessment using natural scene sta- tistics”, IEEE Trans. Image Process., v. 14, n. 12, , Dec. 2005, pp. 2117–2128.

[39] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, “Image quality assessment based on a degradation model,” IEEE Trans. Image Process., vol. 9, no. 4, , Sep. 2002, pp. 636–650.

[40] Q. Huynoh-Thu and M. Ghanbari, “Scope of validity of PSNR in image/video quality assessment,” Electronics Letters, v. 44, n. 13, 2008.

[41] P. Seuntiëns, "Visual experience of 3D TV," doctoral thesis, Eindhoven University of Technology, Eindhoven, the Netherlands. June 2006.

[42] Open Networking Foundation. Software-defined Networking: The New Norm for Networks. ONF White Paper. Accessed June 1, 2017. [Online]. At: https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn newnorm.pdf

[43] M Garcia, A Canovas, M Edo, J Lloret, A QoE management system for ubiquitous IPTV devices, 2009 Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM '09), October 11-16, 2009 – Sliema, Malta.

[44] J. Lloret, M. Atenas, A. Canovas, M. Garcia, “A Network Management Algorithm Based on 3D Coding Techniques for Stereoscopic IPTV Delivery,” IEEE Globecomm 2011, Houston, TX, USA, 5-9 Dec. 2011, pp. 24-28.

[45] A. Canovas, M. Garcia, J. Lloret, M. Atenas R. Rizo, Improving IPTV QoE taking the suitable MPEG-2/MPEG-4 Quantizer based on Jitter, Delay and lost packets measurements, Int. Journal On Advances in Telecommunications, vol. 3, no. 3&4, Dec. 2010, pp. 129 - 139.

[46] J. MacQueen, “Some methods for classification and analysis of multivariate observations" 5th Berkeley Symp. on Mathematical Statistics and Probability, vol. 1, pp. 281-296, Berkeley, CA, 1965.

[47] International Telecommunication Union: ITU-T Rec. G.114 (05/2003). One-way transmission time, Accessed July 1, 2017.

[48] Netdisturb Website. Accessed July 1, 2017. [Online]: http://www.zti-telecom.com/EN/NetDisturb.html.

[49] Mobile 3DTV Solideyesight, Video Plus Depth. Accessed July 1, 2017. [Online]: http://sp.cs.tut.fi/mobile3dtv/video-plus-depth/

[50] Accessed July 1, 2017. [Online]: FFmpeg Webstite, At http://www.ffmpeg.org

[51] M. Pinson and S. Wolf. “A New Standardized Method for Objectively Measuring Video Quality,” IEEE Transactions on Broadcasting, Vol. 50, no. 3, , September 2004, pp. 312-322

[52] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, S. Villette, A. M. Kondoz, Quality evaluation of color plus depth map-based stereoscopic video, IEEE J. of Sel. Topics in Signal Processing, vol. 3, no. 2, 2009, pp. 304-318.

[53] Taha, M., 2016, June. A Novel CDN Testbed for Fast Deploying HTTP Adaptive Video Streaming. In Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications (pp. 65-71). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).