♦ Providing 3D Video Services: The Challenge From 2D to 3DTV Quality of Experience

José María Cubero, Jesús Gutiérrez, Pablo Pérez, Enrique Estalayo, Julián Cabrera, Fernando Jaureguizar, and Narciso García

Recently, three-dimensional (3D) video has decisively burst onto the entertainment industry scene, and has arrived in households even before the standardization process has been completed. 3D television (3DTV) adoption and deployment can be seen as a major leap in television history, similar to previous transitions from black and white (B&W) to color, from analog to digital television (TV), and from standard definition to high definition. In this paper, we analyze current 3D video technology trends in order to define a taxonomy of the availability and possible introduction of 3D-based services. We also propose an audiovisual network services architecture which provides a smooth transition from two-dimensional (2D) to 3DTV in an Internet Protocol (IP)-based scenario. Based on subjective assessment tests, we also analyze those factors which will influence the quality of experience in those 3D video services, focusing on effects of both coding and transmission errors. In addition, examples of the application of the architecture and results of assessment tests are provided.

Introduction

Both the entertainment industry and the research community are focused on the decisive introduction and evolution of three-dimensional (3D) video in entertainment media. The possibilities of this novelty technology in terms of immersiveness and enhancement of the user experience are generating great expectations in the sector, as new market and business opportunities are foreseen.

Several possible 3D scenarios can be materialized in the short to medium term to offer the end user a 3D experience. In this paper we identify these scenarios, and analyze the technology-associated trends in the whole end-to-end chain, from content capture to content display, including coding and representation formats, and available delivery standards.

Taking an Internet Protocol (IP)-based scenario as the starting point, we propose an architecture for the delivery of 3D content, identifying the new functionalities required for managing 3D content. Moreover, 3D content information is organized and aggregated in different ways, depending on the technology involved in the implementation of each particular scenario.

Panel 1. Abbreviations, Acronyms, and Terms

2D—Two dimensional
3D—Three dimensional
3DTV—3D television
3DVC—3D video coding
AUD—Access unit delimiter
AVC—Advanced Video Coding
B&W—Black and white
DVB—Digital Video Broadcasting
DVD—Digital video disc
GOP—Group of pictures
HAS—HTTP Adaptive Streaming
HD—High-definition
HDMI—High-definition multimedia interface
HDTV—High-definition television
HEVC—High Efficiency Video Coding
HLS—HTTP live streaming
HTTP—Hypertext Transfer Protocol
HVS—Human visual system
IETF—Internet Engineering Task Force
IP—Internet Protocol
IPTV—Internet Protocol television
ITU—International Telecommunication Union
ITU-R—ITU Radiocommunication Sector
LDI—Layered depth image
MOS—Mean opinion score
MPEG—Moving Picture Experts Group
MVC—Multiview video coding
MVD—Multiview video plus depth
NALU—Network abstraction layer unit
OTT—Over-the-top
PSNR—Peak signal-to-noise ratio
QoE—Quality of experience
RTP—Real Time Transport Protocol
S3D—Stereoscopic 3D
SbS—Side-by-side
SDTV—Standard-definition television
SEI—Supplemental enhancement information
SMPTE—Society of Motion Picture and Television Engineers
SNR—Signal-to-noise ratio
SSCQE—Single Stimulus Continuous Quality Evaluation
STB—Set-top box
SVC—Scalable video coding
TaB—Top-and-bottom
TOF—Time-of-flight
TV—Television
ULP—Unequal loss protection
V+D—View plus depth
VQEG—Video Quality Experts Group

Thus, the end user experience can be improved by taking advantage of the new solutions made available through content-aware processing appliances and services hosted at either the head end or in the delivery network. Such solutions have already been successfully demonstrated in two-dimensional (2D) video [27].

In this sense, the characterization of the quality of experience (QoE) of the end user depends highly on the particular technologies adopted for the representation, coding, and delivery of 3D content. In this paper, we also present the results of preliminary work on how losses during content compression or content distribution impact user perception.

3D Scenarios

Providing the depth perception necessary to enhance the audiovisual content experience is a pressing issue in both the entertainment industry and the research community. Interests in this area have been reinforced by the positive feedback received from the consumer market in 3D cinemas and 3D home entertainment scenarios. This positive feedback has been a consequence of a mature technology, especially related to displays, that provides excellent quality and an enhanced video experience. In this sense, the research community and standardization bodies are also working in different technological areas in order to provide a common reference framework that guarantees the interoperability of these newly proposed solutions. As examples, we can mention the recent extension to H.264/Advanced Video Coding (AVC), multiview video coding (MVC) [35], to handle the coding of several views in a multiview scenario; 3D video coding (3DVC) and the work in High Efficiency Video Coding (HEVC) being performed by a Joint Collaborative Team on Video Coding with video coding

experts from the Moving Picture Experts Group (MPEG) and International Telecommunication Union (ITU); or the Digital Video Broadcasting (DVB), Frame Compatible Plano-Stereoscopic 3DTV (DVB-3DTV) BlueBook [8].

There is no unique way to fulfill this 3D experience, but different approaches exist, ranging from the use of monocular cameras in "shape from motion" techniques, to complex caves for virtual reality applications that display complete surrounding audiovisual information.

In this paper, we consider the provisioning of depth information in a television (TV) environment. Therefore, we focus on those approaches, and also on the technologies involved which target the enhancement of the user experience to a 3D perception of the visual content in a TV scenario. Two different scenarios can be distinguished for handling depth information:

- *Video signal* scenario, which includes only video signals coming from different cameras, comprising classical stereo video with two views, and multiview video with more than two views. In particular, stereoscopic video has taken advantage of the developments in technology, especially in the field of displays, and has currently been adopted by 3D cinemas, by home entertainment solutions, as Blu-ray Disc* 3D, and even by some pioneering 3DTV broadcasting operators such as Sky 3D [4].
- *Depth-enhanced* scenario, which considers the combination of video signals with a representation of the geometry of the scene. This representation is later used in the rendering of virtual views that are used to feed the 3D displays. It includes video plus depth (V+D), multiview video plus depth (MVD), and layered depth images (LDI).

Both scenarios impose several requirements and constraints on the whole chain: content acquisition, format representation and coding, content distribution, and rendering and display. These relations are displayed in **Figure 1**, particularized for the simplest case for each scenario: stereoscopic video (targeted to stereoscopic displays), and view plus depth (targeted to autostereoscopic displays). The former case is also split into two, depending on whether the coding and transport chain is or is not compatible with existing high-definition television (HDTV) distribution technology. In the following sections, we provide a brief analysis of the state of the art in all these areas.

Content Acquisition

There is currently a wide range of solutions to capture 3D content, from methods using a single camera, to those involving several cameras arranged in an array. On one hand, the video acquired by a single camera is 2D, but can be converted to 3D using monocular depth cues obtained from the 2D scene. This conversion can be achieved in real time to make the 3D content available at the end user device just prior to content rendering, or in situations where a non-heavy camera is required (e.g., with the moving zenithal camera in sports or other live events). Nevertheless, the results obtained are non-optimal due to the flatness, the occlusions in the edges of the rendered objects, and the consequent unrealistic feeling. In addition, offline 2D to 3D conversion obtains better results thanks to the supervision of an operator, resulting in semi-automatic processing. This last situation is of interest for getting 3D content from conventional 2D footage. On the other hand, stereoscopic cameras, either a single camera with two lenses or two identical cameras arranged in a rig, directly acquire 3D content as stereoscopic video which is suitable for current 3DTV transmissions, but also can be used to provide depth to the scene. Another way to obtain depth is by using time-of-flight (TOF) cameras, or other type of rangefinders. These devices enrich the 2D view from a single camera with 3D information. A more complex acquisition system is a camera array arrangement which acquires multiple 2D views of the same scene, but from slightly different points of view. The result is high quality 3D video content known as multiview video, but this solution is very complex, as each camera must be calibrated and rectified in relation to one of the cameras of the array. Finally, the 3D information from the multiview arrangement can be enriched by adding a TOF device to every camera in the array.

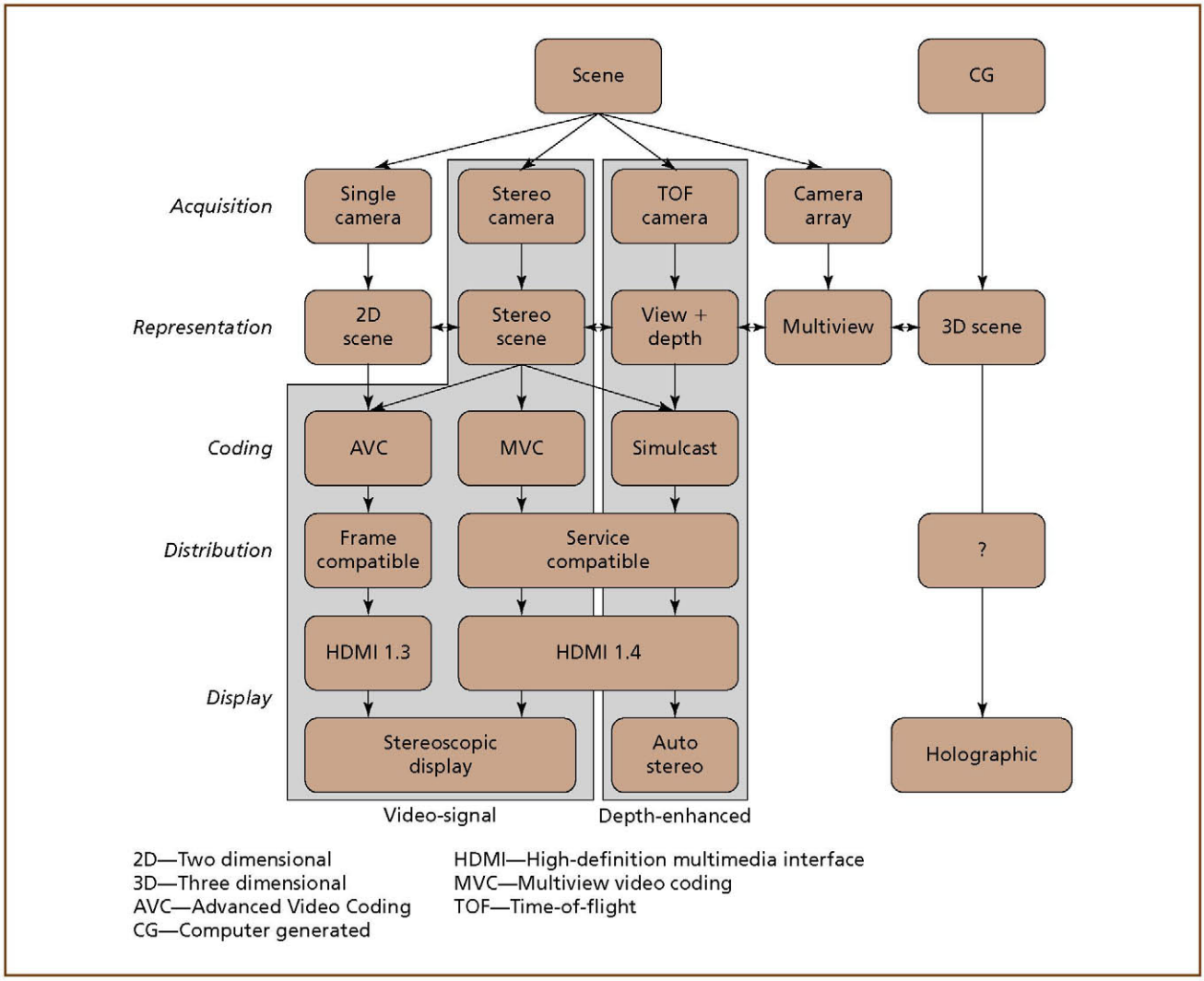


Figure 1.
Scenarios for 3D video services.

Each camera setup is able to obtain the input image format required for each scenario. These scenarios can be, up to a point, interchangeable because depth information can be measured by TOF cameras, but also can be estimated from monocular, stereo, or multiview video, allowing the movement between adjacent scenarios because they share the same kind of information (see Figure 1).

Finally, a new scenario regarding computer-generated images can also be considered. In this case, a 3D model of the objects in the scene is used to paste natural views, and thus render virtual views from any point of view. This is the case for animation movies and other types of synthetic content.

Representation Formats

Frame representation in the video signal scenario is quite straightforward: Each view is represented by a single frame. This implies that, for each time instant, several parallel images have to be represented and transmitted.

Aiming at compatibility with current 2D technology (frame compatible scenario described by DVB [8]), and looking for a reduced bit rate, there are several possible representation formats based on spatial and/or temporal multiplexing which result in a single output sequence:

- *Temporal multiplexing.* The output sequence is made up of alternating frames from each camera.

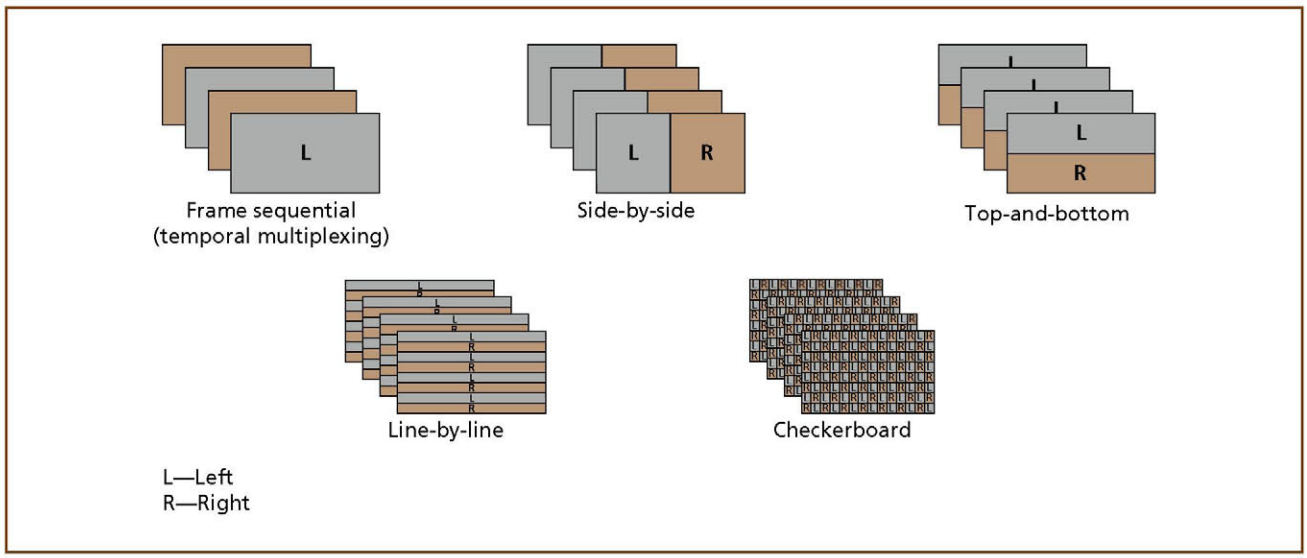


Figure 2.
Frame compatible multiplexing formats.

Thus, the resulting sequence doubles the frame rate of its components so it requires a higher bit rate. In this approach, the spatial and temporal resolution of the input views is maintained.

- *Spatial multiplexing*. The frames from each camera are subsampled and combined into a common frame that can be handled by any conventional 2D processing or transmission equipment. Each resulting frame maintains the original size of the input frames (see **Figure 2**).
  - *Side-by-side (SbS)*. Both input frames are downsized to half their original width.
  - *Top-and-bottom (TaB)*. The horizontal resolution of both input frames is halved.
  - *Line-by-line*. The output frame is built via line-by-line interlacing from both input frames. Since the output frame maintains the size of the input frames, half of the lines of each input frame are discarded.
  - *Checkerboard*. Output frames are composed in a checkerboard fashion from the input images. In this case, half of the pixels on each line are discarded following a checkerboard pattern.

For depth-enhanced scenarios, the activity in many research projects [9, 10, 31], and the activities in standardization bodies like the Moving Picture Experts Group, are focused on representing video signals from multiple views and information on the distance of the objects in the scene to one or more cameras. These format types are known as video plus depth. A depth map is a data format that represents the distance of the scene objects with respect to a camera. Generally, this information is represented with a depth value per pixel of the video signal. The depth values are scaled and quantized to form a monochromatic image in which, generally, the value of each pixel image represents the inverse depth value.

V+D representations range from a single view together with its depth map, to a more complex scenario where several views plus their depth maps are considered: multiview video plus depth. In the latter case, layered depth images [29, 41] are an alternative representation to the depth sequences. LDI aggregates depth information from different cameras into one single, multi-layered representation. As a result, each of the pixels of the LDI image contains information about a visible pixel as well as hidden ones. This data format may be useful in view synthesis algorithms for handling occlusion.

Coding Format

The frame-compatible video signal scenario as standardized in [8] has the simplest coding requirements: since there is a single output video sequence,

any video codec can be used to encode it. Nevertheless, additional signaling is required to specify the interleaving scheme chosen. In H.264/AVC [17], the use of supplemental enhancement information (SEI) has been standardized in order to identify both views and the interleaving scheme that has been applied. As a drawback, this solution requires that the decoder supports the use of SEIs in order to undo the multiplexing of the views.

When more independent views, depth planes, or layers are added, then it is necessary to use coding formats which handle more than one picture simultaneously. A first possible approach is using simulcast, which consists of encoding each of the video and/or depth sequences of the considered scenario (stereoscopic, multiview, V+D, or MVD) independently, using an H.264/AVC encoder [12]. This approach is very efficient in terms of complexity and latency, since it works with efficient compression tools and maintains the compatibility with existing compression technology. On the contrary, its main drawback is that the scheme does not consider the nature of the new content to increase the efficiency of the compression: it does not exploit the interview redundancy of stereoscopic video or multiview video, nor are H.264/AVC encoding tools optimized for depth map sequences. Moreover, additional synchronization and signaling mechanisms may be required to properly render the 3D content.

The multiview coding amendment of the H.264/AVC standard [35] provides encoding tools that exploit redundancy of multiview video. MVC was designed to cope with those scenarios where camera arrays are used, and is able to encode multiple views more efficiently than the simulcast approach (see **Figure 3**). Nevertheless, the requirement that the main view be compatible with H.264/AVC limited the development of new interview compression tools, as well as compression efficiency. In a later revision, the stereo high profile was included in an attempt to customize MVC tools to stereoscopic video. This profile has been adopted by the Blu-ray 3D standard.

Recently, a standardization process for a new 3D video coding standard was initiated within MPEG [13]. The scope of 3DVC is to provide a new data format for both stereoscopic systems and multiview systems. Currently, the working framework consists of a multicamera scenario with a limited number of views, plus depth information. Depth data can be represented in either a depth map or LDI format. Nevertheless, this activity is in its initial phase, and the organization has released a call for proposals on 3D video coding technology [14]. This document describes the requirements, and testing environment and procedures for the proponents of technology. In it, three different categories are considered, according to compatibility with existing coding technology:

- AVC-compatible solutions that consider AVC, a well-known and widespread coding standard;
- HEVC-compatible solutions as a consequence of promising results in terms of the rate distortion efficiency of HEVC; and

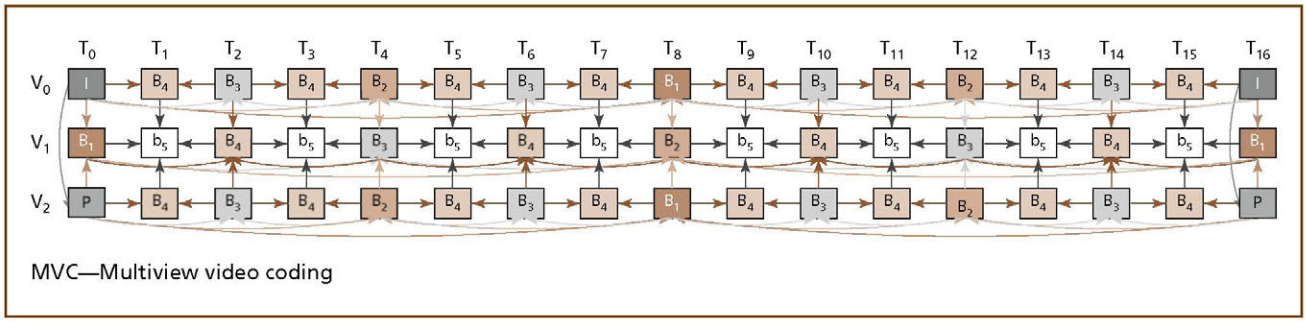


Figure 3.
MVC prediction structure.

- Unconstrained solutions which allow the proposal of new coding schemes not bounded to existing standards.

Proposed solutions will be evaluated both objectively and subjectively, in the latter case considering the use of stereoscopic displays and auto-stereoscopic displays.

Distribution

Content providers and technology developers are encouraging the definition of a common 3DTV standard in order to avoid a battle of formats, to ensure consumers that they will be able to view the 3D content they purchase, and to provide 3D home solutions for all pockets.

Early efforts on 3D standardization were carried out by MPEG and the Society of Motion Picture and Television Engineers (SMPTE), but other organizations such as the DVB Project, ITU, and the DVD Forum, have created their own investigation groups, and have already offered to collaborate to reach a common solution. As a result of this standardization impulse, new and against-the-clock recommendations for stereoscopic 3DTV services are being depicted, for example DVB-3DTV [8], which goes from the *frame compatible* to the *service compatible* delivery system specification.

A delivery system specified for frame compatible 3DTV services enables service providers to utilize their existing HDTV infrastructures to deliver 3DTV services in a video signal scenario with stereoscopic 3D content (S3D), i.e., only two views. This scenario is compatible with the 3DTV-capable displays already available on the market, and can share the infrastructure with other HDTV services that are already deployed.

In a further situation, 3DTV service compatible systems will require a new set-top box (STB) and a new display, but will offer a normal 2D high-definition (HD) resolution (horizontal, vertical, and diagonal) per eye. These systems will support multiview video signal and depth-enhanced scenarios.

Displays

The perception of 3D in the human vision system (HVS) is based on several monocular and binocular cues. The most commonly used mechanism to generate 3D video on current flat displays is the binocular parallax. This implies that the left view image and the right view image (the stereo pair) have to be sent to their respective eyes to generate the perception of 3D by the user. Flat displays for 3D systems can be classified into two main technologies: stereoscopic displays that require the viewer to wear glasses (passive polarized or active shutter), and auto-stereoscopic displays that do not need additional devices to separate the stereo pair.

Passive stereoscopic displays use spatial multiplexing to present the left and right eye images simultaneously in time, interleaving both images line-by-line. A polarization filter film over the surface of the display polarizes the horizontal lines with alternate polarization, so an observer watching the display with the appropriate pair of passive polarized glasses will see each image in the corresponding eye. The 3D content shown in this kind of display is equivalent to an interlaced image, assigning the information for each eye to alternate lines. So, the vertical resolution per eye is halved.

Active stereoscopic displays use time multiplexing to present the stereo pair alternate in time. The observer must wear active shutter glasses (with liquid crystal) synchronized with the display to alternatively darken one eye, while allowing the other to receive its corresponding image. The spatial resolution is conserved for the left and right eye images, so it is necessary to at least double the frame rate (120 Hz or even 240 Hz are common for these displays) to reduce the annoying flickering caused by blocking the light to each eye alternatively.

Currently, there are two auto-stereoscopic technologies which do not use glasses to route the corresponding view to each eye: lenticular and parallax barrier. The aim in both cases is to send each view to a different spatial region in front of the display. When the observer is located in the right place, the left and the right eyes will receive the corresponding images to compose a stereo pair. Many displays allow several views of the same scene simultaneously (e.g., nine views to compose eight stereo pairs), turning them into auto-multiscopic displays. Due to that, the spatial resolution is reduced in a proportional way.

Different display technologies are the main support for the different scenarios defined in the value

chain: stereoscopic displays require two different views as input, while autostereoscopic displays typically render their different views based on a depth-enhanced input.

This input source is usually injected using high-definition multimedia interface (HDMI), an audio and video interface for transmitting uncompressed digital data. HDMI version 1.4 defines input/output protocols that allow 3D display and source devices to communicate through the cable link with resolutions up to 1080p in 3D. It supports several stereoscopic display methods such as frame packing (a full resolution top-bottom format), field alternative, line alternative, SbS half, SbS full, left V+D and V+D + graphics + graphics depth. Previous HDMI version 1.3 compatible televisions are not able to handle all 3D formats specified for version 1.4, except SbS and TaB frame compatible formats, those currently being used in a stereoscopic video signal scenario.

Processing 3D Video in the Network

The new approaches to encode and represent 3D stereo video content also propose new schemes to organize the information. Taking this into account, a content-aware process can provide advanced features to video services, such as protecting the most relevant information by adding redundancy or providing hierarchical retransmission mechanisms; grouping the same kind of information; or prioritizing the information to allow smart discarding of transmitted information packets.

We propose an architecture of audiovisual network services able to provide a smooth transition from 2D to 3DTV in an IP-based scenario. This architecture allows 3D video pre-processing to provide content protection through unequal loss protection (ULP) and packet prioritization, and 3D video adaptation and delivery to permit 2D to 3D migration and to deliver an adapted (or even adaptive) stream to heterogeneous clients, jumping from 3D to another scenario.

The following subsections set out an approach to 3D content delivery problems, and propose a solution built over a traditional architecture for IP-video delivery. This solution identifies the new functionalities required for delivery of 3D content by selecting the most feasible technologies from those depicted in the previous sections.

Architecture

To process 3D video in the network, some new elements have to be layered on top of the common video delivery architecture. There are two reasons supporting this addition: on the one hand, the video coding structure is more complex, both for video signal and depth-enhanced scenarios. On the other, it is likely that several different scenarios have to coexist, thus making it necessary to have points where it is possible to jump from one scenario to another. In fact, this architecture should support a first scenario such as the one depicted in the DVB frame compatible specification (DVB-3DTV [8]), and could allow the evolution to a second service compatible scenario (see Figure 1), intended to be standardized in a second phase of DVB.

A simplified vision of this architecture is shown in **Figure 4** and can be explained as follows:

- The input of the system is a set of coded representations of the same scene, in one of the coding formats which have been described previously. All of these representations must be synchronized.
- This input is preprocessed to create a structured stream with all the information.
- This stream is distributed to the network edge using the common mechanisms (multicast, broadcast, and content delivery network).
- The network edge contains elements (delivery nodes, in Figure 4) which are able to adapt the video stream to the requirements of the end client.

Video pre-processing. The video pre-processing system must be able to segment the video input so that different views, depth planes, and other coding structure elements can be easily identified and separated downwards. This kind of network preprocessing has already been successfully performed in 2D video, to separate video frames from audio in an MPEG-2 transport stream over Real Time Transport Protocol (RTP) [27]. In this process, an extension is added to the RTP header in order to signalize the type of elementary stream information included within the

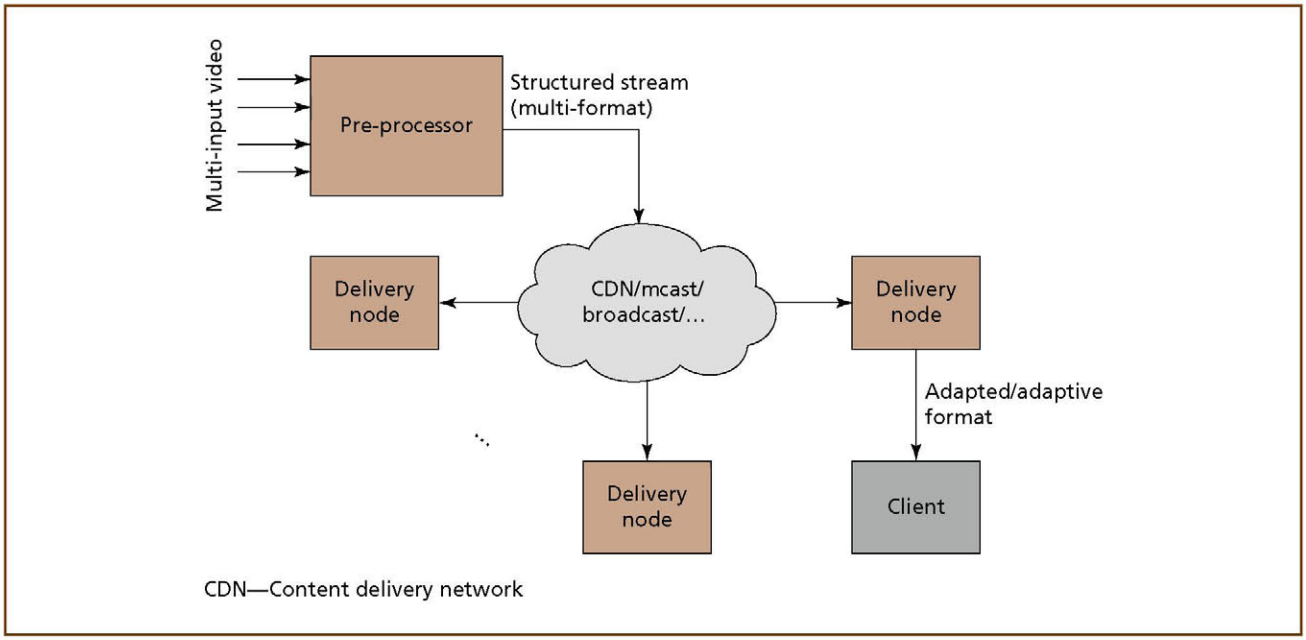


Figure 4.
Network architecture.

packet. We propose adding a new extension which adds information about the layer being coded.

It is worth noting that, with the exception of frame compatible video, it is always possible to define a main layer (or view): the base MVC view, the view in V+D, the top image in LDI, or an arbitrary view in simulcast. In all of these cases, in retaining the main layer, it is possible to recover a correct 2D stream. Hence, all of them can be treated in a homogeneous way from the RTP point of view, thus significantly simplifying the complexity of network processing.

Video pre-processing should also be able to provide synchronization points where it is possible to jump from one representation to another in a smooth way (provided that the streams are encoded in a way that allows it).

Video adaptation and delivery. Once the video has been appropriately pre-processed, its delivery to the end user is only a matter of selecting the right flows to send. In most cases, this selection should be doable by filtering out all packets that do not belong to the end client profile. However, there are inevitably situations where the correct representation format is not available, either because the client profile has not been taken into account when creating the original stream, or because advanced services (such as free viewpoint video, where the user can dynamically select his viewpoint of the scene) are being provided. In such cases, an additional layer of processing has to be added in the pixel domain, as depicted in **Figure 5**. For scalability

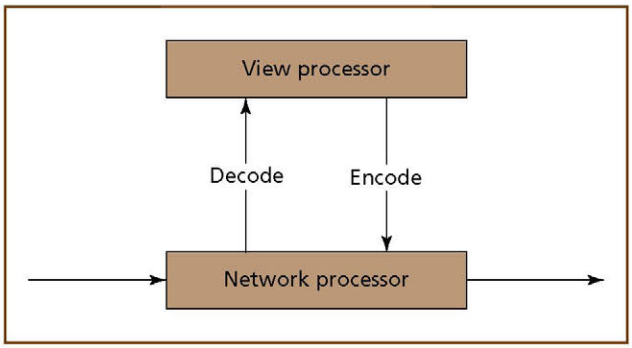


Figure 5.
Double-layer processing.

reasons, this pixel-based processing on the network edge should be avoided whenever possible.

When the delivery technology supports it, adaptation to the capabilities of client devices can also take into account network issues, and then become an adaptive delivery, which is a current market trend in upcoming over-the-top (OTT) scenarios and applications, where users connect directly to an unmanaged network to access content. This is the case with HTTP Adaptive Streaming (HAS), where different bit rate profiles can be generated by just adding or removing views from the stream. A 2D proposal already pushed to standardization is HTTP live streaming (HLS) [25], available as an Internet draft at the Internet Engineering Task Force (IETF). In that proposal, the content is encoded at several bit rates, and then segmented and encapsulated for delivery. A playlist file is used as an index by the client to select the proper version of segment, depending on its processing and network capabilities.

Unequal Loss Protection

The primary aim of video pre-processing is to determine the main layer (or view) and maintain it, allowing the recovery of a correct 2D stream. Unequal loss protection schemes are used when transmission resources are limited, and the introduction of a second stream may involve going over the available bit rate, which can be used in this case to protect the main layer.

ULP strategies are usually used to decide which part of the data should be protected and how, so that resource availability is not exceeded and the overall quality after decoding is kept as high as possible [24]. Different ULP techniques have been proposed in the literature. They usually differ in two main aspects, which influence the computational cost of the scheme: the scope of the decisions made, and the level at which the analysis of video data is performed. The first aspect refers to the structure of the data: a set of packets in a stream, a set of macroblocks in a frame, a frame, or a video layer. The second alludes to the units within the encoded video stream, whose features are analyzed to perform the prioritization: macroblock ranking, frame classification, and video scalability exploitation [5, 6, 11]. In general, the finer the granularity of evaluation, the more computationally costly the technique. Once data are accessed and analyzed, most of the techniques raise cost minimization problems whose solutions determine the behavior of the scheme, that is, which protection policies to follow. The cost function to be minimized is typically based on a model of the distortion that affects the video when a portion of the information is lost [30, 37].

In 3D video transmission scenarios, some new ULP approaches dealing with the protection of multiview video streaming (two-view and multiview stereoscopic video, and free viewpoint video) have been proposed [1]. These strategies are based on the ULP schemes already mentioned but adjusted to the video stream structure characteristics of the specific 3D video stream [23, 32].

Packet Prioritization

We propose the application of a ULP scheme adjusted to the coding and representation formats of the different scenarios. This scheme will be based on a packet prioritization by establishing several priority levels. These can be defined inside a view or layer (reducing the problem to one already solved [26]), and between the different views. The priorities can be used by delivery nodes (see Figure 4) to discard information, by removing a view or a frame, changing from 3D to 2D, discarding frames, reducing the frame rate, or a tradeoff of all of these.

Different groups of information can be identified in this stereoscopic video content, such as frame views. The identification of these views, depth planes, and other coding structure elements depends on the coding and representation formats. The main point will be to determine what is more important, and what should be prioritized.

Coding formats. The H.264 standard [17] has been extended to support MVC and scalable video coding (SVC). The following information is used to assign a higher priority to the container packets.

A new type of network abstraction layer unit (NALU) has been added. This type contains a NALU extended header and carries the information of the non-main layers. The NALU's header has up to four bytes instead of one, including the layer or view information, depending on the encoding technique, SVC or MVC, and a priority code [28, 36].

A NALU prefix type has been added (type 14), which is placed before the AVC NALU slice, and carries the three extended header bytes. Another NALU (type 19) extends SewParamSet to support MVC/SVC. All access units are composed by the whole set of views or layers in a given time t, then there is only one access unit delimiter (AUD) in that instant t, carrying all the views or layers.

There are two different MVC profiles, both applicable to 3D content. The first one supports two views (stereo), while the other supports N views, but only progressive scan. SVC has three scalability types: temporal (already available for AVC), resolution, and signal-to-noise ratio (SNR), which can be used jointly.

Representation format. Several ways of representing stereo video content in an asymmetric fashion have been identified: in a single AVC stream (by using SEIs), two different multicast streams, and MVC or V+D simulcast. In all of them, one view is identified as the primary one, which should be prioritized over the rest. Any view can be prioritized as the main one in the case of symmetric video (AVC or simulcast).

Migrating From 2D to 3D

One important advantage of the architecture proposed is that it provides a way to perform a smooth migration of 2D to 3D video. Considering a live RTP multicast video network, such as the typical Internet Protocol television (IPTV) deployment, the only requirement is that either the delivery nodes or the client elements are able to filter out the information which does not belong to the main view. For instance, different multicast groups or ports could be used to process the different views, so that a legacy STB would only receive the legacy 2D video, while more modern clients would know where to search for the rest of the video data.

The proposed architecture also provides support for the possible transitions among different and coexisting 3D transmission and distribution scenarios, as well as the evolution from frame compatible to service compatible scenarios (as described in DVB [8]).

Quality of Experience in 3DTV

As people are the final users of 3D services, the performance of the aforementioned techniques for processing and delivering 3D video content should be evaluated to take the user experience into account. Several factors influence the viewing experience of end users, which are covered under the term quality of experience. For example, subjective factors of the viewers, such as interests or experiences using multimedia applications, can influence their opinion of quality. In addition, many factors related to the environment where the audiovisual content is observed could condition the perceived quality, including lightning, or the display used. Additional factors include video quality, audio quality, the synchronization between both, and distortions that content could suffer in the distribution chain to households [38]. Furthermore, when dealing with 3D content, new aspects influencing the viewing experience of end users have to be considered in the evaluation of QoE, such as depth perception, naturalness, sense of presence, or visual discomfort.

The significant influence of subjective factors in the QoE perceived by viewers makes subjective assessment tests the most accurate methods for conducting an evaluation. In such experiments, the audiovisual content to be evaluated is shown to a number of observers who rate it according to a set test methodology. Several studies have been presented in the literature analyzing the subjective quality factors related to conventional video, which led to a standard evaluation methodology proposed by the ITU [16]. In contrast, the evaluation of the QoE associated with 3D video is the focus of many current studies, since new factors may be considered to obtain reliable measurement techniques [7]. These types of experiments provide a better understanding of the HVS, and the results obtained establish a basis for developing automatic estimators of video quality. These estimators are very useful for many practical applications, like video quality monitoring in broadcasting systems. Therefore, both for monoscopic and 3D video, very active research is being done on these techniques [18, 22], most outstanding being the guidance and standardization activity of the Video Quality Experts Group (VQEG).

To evaluate the performance of a 3D video delivery system, it is necessary to know the possible artifacts

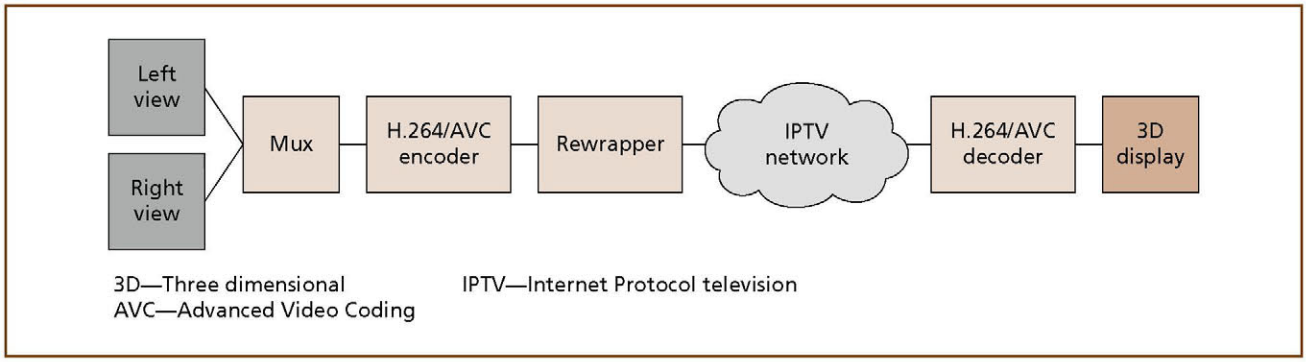


Figure 6.
Broadcasting scenario considered in the subjective test.

that could be introduced along the processing and transmission chain. Although multimedia quality could be degraded in any phase of this chain, the next subsections focus on the analysis of those effects related to 3D content delivery, which are caused by errors during coding and transmission, and will vary depending on the selected scenario. The study of the impact of these distortions on video quality will provide the required knowledge to optimize the algorithms for transmission control, like ULP, which works in the video processing architecture described in the previous section.

Therefore, a subjective experiment was carried out to evaluate the impact of transmission errors in an SbS video broadcasting scenario, as shown in **Figure 6**. This stereoscopic video signal scenario is considered to be the first approach to deliver 3D video to home environments, and it has recently begun to be used.

Effect of Coding Artifacts

The compression of the data when video content is encoded could cause some degradation in quality. For example, with conventional monoscopic video, typical coding artifacts were blocking effects, blurriness, or ringing [38]. Although these distortions also appear when 3D video is encoded, their impact on the perceived QoE could be different, since not only could the quality of the sequence be distorted, but also the stereoscopic perception. Therefore, several studies have been proposed in the literature analyzing the effects of coding degradation on QoE. For example, dealing with stereo views, the effects of coding artifacts could be more annoying if each view is affected differently, making the fusion of both images difficult for the HVS. For instance, it has been shown that when blocking effects affect each view differently, binocular rivalry could be produced [2]. Some other studies analyzed the effects of coding V+D formats, showing that the HVS is less sensitive to distortions affecting the depth map than those degrading the quality of video color [34]. However, a coarse quantization of the depth map can produce a *cardboard effect*, which is the discrete division of the depth into various planes, causing the depth of the scene to appear unnatural.

The effects of the aforementioned distortions, and the appearance of specific artifacts related to 3D video coding techniques, like cross distortions between views, should be better analyzed by means of subjective tests. Subjective experiments provide perceptual information on the quality of the encoded content, in contrast to the peak signal-to-noise ratio (PSNR) commonly used to estimate the quality of encoded images and videos. This fact makes subjective tests the most accurate methodology to evaluate the performance of coding algorithms, as has been shown with the use of such assessment techniques in the standardization of 3D content coding techniques like SVC [28], MVC [36], and the novel 3DVC [14] currently in development.

Effect of Transmission Errors

Packet loss and jitter are the main transmission errors that could appear in IP networks used for delivering

multimedia applications. For monoscopic video, the effects of these errors on the perceived quality have been extensively studied by means of subjective experiments, and also by developing objective metrics [3, 22]. However, the results of those studies cannot be directly assumed for 3D video, since the impact of these degradations are highly dependent on the scheme used for encoding the content, and on the new perceptual factors involved in the stereoscopic vision of 3D sequences. Therefore, new quality assessment experiments are being carried out.

Only a few research works have been presented in the literature analyzing the impact of transmission errors on the quality of 3DTV. One example can be found in [2], where Barkowsky et al. carried out a subjective study of the effects of different patterns of packet losses in a simulcast scenario. In addition, various error concealment techniques for 3DTV were analyzed. Another study was proposed by Yasakethu et al. [39], where transmission errors were simulated in 3D video broadcasts, in which S3D and V+D videos were asymmetrically encoded using SVC.

Since little research has been done analyzing the perceptual effects of transmission errors in frame compatible scenarios, we carried out subjective assessment tests considering the broadcasting of SbS 3D video. Typical transmission errors, the properties of which are described in **Table I**, were considered to evaluate their impact on quality perceived by end users. In addition, effects caused by a decrease of the quality of service in the network, like frame rate and bit rate drops, were also considered.

The experiments were carried out in a lab under the conditions recommended by the ITU Radiocommunication Sector (ITU-R) in BT.500-11 [16] and ITU-R BT.1438 [15], equipped with a 42″ stereoscopic television with resolution of 1920 × 1080. For visualizing 3D video, the observers were placed at a distance of three times the height of the TV, and they were required to wear active shutter glasses. A total number of 19 viewers participated in the experiments. A monoscopic and stereoscopic version of the sequences described in **Table II** were used in the experiments. The sequences had a duration of five minutes, and were encoded with H.264/AVC, with a group of pictures (GOP) length of 24 frames, and a structure IBBBP, with a bit rate of 8 Mb/s for HD resolution, and 4 Mb/s for standard-definition television (SDTV). In the case of 3D sequences, the left and right views were first multiplexed SbS into one single frame.

The methodology used in the test was based on the Single Stimulus Continuous Quality Evaluation (SSCQE) [16]. Single stimulus methods stay as close as possible to typical viewing conditions in real home environments, since rating is not done by comparing

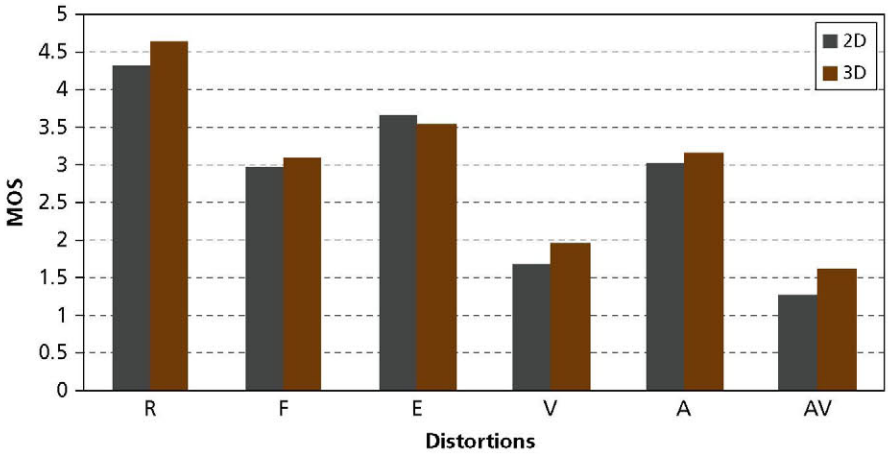
Table I. Considered distortions used in subjective tests of transmission errors.

Error type	Description
R	Bit rate drops.
F	Frame rate drops.
E	Video losses producing macroblocking. The losses could affect different fractions of the frames, and various lengths were considered.
V	Video freeze of different duration.
A	Audio losses of different duration.
AV	Video freeze combined with audio loss

Table II. Test sequences used in subjective tests of transmission errors.

Sequence	Format	Content
1	1920 × 1080p 24 fps	Movie. Some slow segments with dialog. Some other segments with fast camera movement.
2	720 × 576p 25 fps	Documentary. Slow action. Some segments with camera panning. Only music as soundtrack.

fps—Frames per second



2D—Two dimensional
3D—Three dimensional
3DTV—3D television
A—Audio losses of different duration
AV—Video freeze combined with audio losses
E—Video losses producing macroblocks
F—Frame rate drops
MOS—Mean opinion score
R—Bit rate drops
V—Video freeze of different duration

Figure 7.
Results of the impact of transmission errors in 2D and 3DTV.

the distorted content with a reference. Moreover, they are the most appropriate methods to evaluate the quality of the performance of broadcasting systems, since the evaluation is done in a continuous way. Therefore, the sequences were divided into segments of 6 seconds, and transmission errors were randomly introduced in alternate segments. This way, after a distorted segment, the observers could rate the impact of the artifact during the 6 seconds of the following segment, which had a number printed indicating the corresponding square to fill in a questionnaire. The five grade-scale impairment scale proposed in [16] was used. The same procedure was conducted for the monoscopic and stereoscopic versions of the sequences, which were rated in 30-minute sessions preceded by an explanation of the experiment. A maximum of four observers participated in each session.

The results of the experiments were processed, and the mean opinion scores (MOS) were obtained for each distortion described in Table I. A comparison of the results obtained for both monoscopic and stereoscopic videos is shown in **Figure 7**. As similar results were obtained for both 2D and 3D content, a more exhaustive analysis of significance was carried out, showing that the differences between the results for 2D and 3D are statistically significant in the cases of bit rate drops (R), video losses (E), video freezing (V), and audio losses (A). Slightly better results were obtained for 3D video for all the considered distortions, except for the video losses (E). This is due to the fact that SbS frames are encoded as a conventional frame using H.264/AVC, and when a video packet is lost, some macroblocks cannot be decoded. In this case, the decoder's error concealment algorithm substitutes corresponding macroblocks from a previous frame. The macroblocks affected by the packet loss could be placed in different positions of the SbS frame, thereby distorting the left and right views differently, as shown in **Figure 8**. This could cause binocular rivalry when each stereoscopic view is displayed, indicating that the corresponding regions of each view are very different. Then, the HVS could not properly fuse both images, and visual discomfort is usually felt. In contrast, the other distortions analyzed

Figure 8.
Effect of video losses on side-by-side video.

in the experiments affect both views in nearly the same way, without altering the stereoscopy. Therefore, better results were obtained for 3D video, probably due to the added value of depth perception.

It is also worth noting that audio losses (A) could be more annoying than video losses (E), while the worst degradations are obtained by combining video freeze and audio losses (AV). Finally, in situations when the performance of the network decreases, a reduction of the bit rate (R) is less annoying than reducing the frame rate (F).

Knowledge of the impact of these effects on perceived quality will allow the improvement of the transmission control techniques used in 3DTV distribution.

Other Factors Concerning the Viewing Experience

As new HVS mechanisms impact the visualization of stereoscopic content, new factors affect the QoE perceived by users in comparison to viewing conventional video. For example, in addition to image quality, additional aspects like depth perception, naturalness, and sense of presence are usually considered. Some studies have been presented in the literature analyzing the effects of these factors on the QoE. For example, a subjective assessment study is presented in [19] evaluating the perceived depth and naturalness in 3DTV systems. In addition, in the subjective tests we carried out as described in the previous section, users were also asked to evaluate the naturalness and sense of presence experienced viewing 3D content. The results showed that these factors are highly dependent on the video content, the quality of the production of depth perception, and the display technology.

The research on 3D video technology has shown the huge importance of the visual discomfort commonly perceived by users of 3D content. In fact, this factor is of major significance in that it is slowing down the expected success of 3D technology, and entertainment applications should not cause any discomfort. This effect was also analyzed in our subjective assessment tests, since the observers evaluated the discomfort felt during the visualization of the 3D test sequences. The results showed that more than 15 percent experienced headache or dizziness, while more than half of the observers felt some type of discomfort.

A number of studies, including [21], have been carried out to analyze the causes of visual discomfort

in 3DTV and how to minimize it. They have shown that the main cause of visual discomfort related to viewing 3D content is the conflict between convergence and accommodation, since the eyes converge in the virtual planes in which the objects are represented, while the point of accommodation is on the screen. The difference between the points of accommodation and convergence does not take place in normal HVS stereoscopic vision in the real world, and thus the compensation carried out by the brain results in discomfort. In addition, a higher level of visual discomfort could be felt when the sequences contained scenes with a high degree of activity and camera motion, as reported by some observers in our tests. Therefore, the visual discomfort caused by the accommodation-convergence conflict could potentially be reduced by careful production of 3D content. However, this cannot always be achieved, especially today when 3D cameras have arrived on the consumer market, and anyone can capture their own 3D videos. Therefore, it will be interesting to observe the possibility of adjustment to the level of disparity according to the display, the viewing distance, and user preferences. In the case of autostereoscopic displays, this could be done by adapting the generation of the different virtual views, while for stereoscopic displays, some techniques have been proposed based on shifting the image [40], or creating virtual views from the real stereo pair [20].

Some of the proposals for adjusting the disparity of 3D content allow users to interact with the TV to set the parameters according to their preference and viewing conditions. This is one of the functionalities that 3DTV could provide to increase user interaction, which plays an important role in the viewing experience. In this aspect, the maximum exponent of user interaction that 3DTV could provide is the sense of immersion in the displayed 3D space. This could be achieved through the use of free viewpoint TV technology, which allows users to navigate inside a 3D scene, changing the viewpoints. Thus, users are able to interact with the displayed 3D scene, selecting different viewpoints and increasing their sense of presence. Intensive research efforts are being carried out in this field to make this technology feasible, and provide this attractive service to users, not only with computer graphics, but also with real content [33].

At the end of our subjective tests, many observers reported some other factors related to the display technology used for viewing 3DTV that influenced their viewing experience, and which have also been discussed in the literature [7]. For example, many observers reported the inconvenience of wearing glasses for watching 3D content, especially in home environments, where it is particularly unnatural and uncomfortable. This is one of the major drawbacks that 3D video technology faces in the consumer market, since households currently use stereoscopic displays. Moreover, both passive and active glasses cause a significant loss of luminance, which is usually reported by observers as an annoying effect when viewing 3D content in stereoscopic displays. Furthermore, in the case of displays based on the use of active shutter glasses, the room illumination conditions are critical, since annoying flickering effects could be perceived by the observers. Another important aspect related to the display technology is the crosstalk, caused by a deficient separation of the different views displayed in the TV, which is not only annoying, but can produce visual discomfort [21].

Finally, after analyzing all the factors affecting the QoE of 3DTV users, it can be expected that users would prefer to watch 3D content over that of conventional video. However, a lot of work remains to be done to achieve high performance in relation to the aforementioned factors, in order to provide a significant added value with 3D technology. In fact, the observers who participated in our subjective experiments had to express their preference between the monoscopic and the SbS stereoscopic versions of the test sequences. The results showed that 53 percent of the viewers preferred the 3D version of source 1, while only 21 percent preferred the stereoscopic version of source 2. This implies that viewers will switch to 3D technology only when the content is properly produced, and it adds performance to conventional video without causing discomfort.

Conclusions

This paper describes current trends in 3D video technology considering the main actors in the end-to-end

chain: content acquisition and representation formats, encoding, distribution, and displays. Taking this description as a reference, our objective has been to define a taxonomy of the availability and possible introduction of 3D-based services. In this sense, we have also proposed an audiovisual network services architecture which provides a smooth transition from 2D to 3DTV in an IP-based scenario. The proposed architecture integrates both pre-processing and delivery adaptation services functionalities that support the 2D to 3DTV transition, as well as any future evolution between different 3D scenarios such as migration from the standardized frame compatible distribution scenario (DVB 3DTV) to a service compatible scenario.

On the other hand, pre-processing and delivery adaptation services can play a major role in the development of protection mechanisms for the distribution of 3D content. We have proposed an unequal loss protection scheme that takes advantage of the pre-processing stage, where encoded 3D content is re-packetized according to the characteristics of its representation and coding format. Analyzing the importance of the content of the packets and its impact on the quality of the decoded 3D content, several priority levels can be defined and used in the ULP scheme.

Finally, we analyzed the factors which influence the QoE in those 3D video services, focusing both on effects of coding and on transmission errors. Subjective tests have been used to assess the level of impact of the different effects of transmission errors in the QoE perceived by the users of frame compatible 3DTV systems. In general, the results of the experiment showed that the effects of transmission errors in 3DTV are less annoying than those in conventional IPTV, possibly due to the added value provided by depth perception. On the other hand, video packet losses can cause significant degradation in SbS 3D video, making the fusion of the stereo views difficult for the HVS. This is one of the causes of visual discomfort inherently produced in 3DTV by the conflict between convergence and accommodation. Viewer discomfort is a crucial factor that must be improved to achieve definitive success for 3D video technology. Other factors related to the 3D viewing experience were also described, like naturalness, sense of presence, and interactivity.

Acknowledgements

The authors thank Álvaro Villegas for his valuable contributions to this work.

*Trademarks

Blu-ray Disc is a registered trademark of Blu-ray Disc Association.

References

[1] G. B. Akar, A. M. Tekalp, C. Fehn, and M. R. Civanlar, "Transport Methods in 3DTV—A Survey," IEEE Trans. Circuits Syst. Video Technol., 17:11 (2007), 1622–1630.

[2] M. Barkowsky, K. Wang, R. Cousseau, K. Brunnström, R. Olsson, and P. Le Callet, "Subjective Quality Assessment of Error Concealment Strategies for 3DTV in the Presence of Asymmetric Transmission Errors," Proc. 18th Internat. Packet Video Workshop (PV '10) (Hong Kong, Chn., 2010), pp. 193–200.

[3] F. Boulos, B. Parrein, P. Le Callet, and D. S. Hands, "Perceptual Effects of Packet Loss on H.264/AVC Encoded Videos," Proc. 4th Internat. Workshop on Video Process. and Quality Metrics for Consumer Electron. (VPQM '09) (Scottsdale, AZ, 2009).

[4] British Sky Broadcasting (BSkyB), <www.sky.com>.

[5] B. Cavusoglu, D. Schonfeld, R. Ansari, and D. K. Bal, "Real-Time Low-Complexity Adaptive Approach for Enhanced QoS and Error Resilience in MPEG-2 Video Transport over RTP Networks," IEEE Trans. Circuits Syst. Video Technol., 15:12 (2005), 1604–1614.

[6] Y. C. Chang, S. W. Lee, and R. Komyia, "A Fast Forward Error Correction Allocation Algorithm for Unequal Error Protection of Video Transmission over Wireless Channels," IEEE Trans. Consumer Electron., 54:3 (2008), 1066–1073.

[7] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New Requirements of Subjective Video Quality Assessment Methodologies for 3DTV," Proc. 5th Internat. Workshop on Video Process. and Quality Metrics for Consumer Electron. (VPQM '10) (Scottsdale, AZ, 2010).

[8] DVB Project, "Digital Video Broadcasting (DVB), Frame Compatible Plano-Stereoscopic 3DTV (DVB-3DTV)," DVB BlueBook doc. A154, Feb. 2011.

[9] European Commission, Fifth Framework Programme (FP5), Community Research and Development Information Service (CORDIS), "Advanced Three-Dimensional Television System Technologies (ATTEST)," Information Society Technologies (IST) Programme 2002–2004, <http://cordis.europa.eu/search/index.cfm>.

[10] European Commission, Seventh Framework Programme (FP7), "3D4YOU—Content Generation and Delivery for 3D Television," Information and Communication Technologies (ICT) Work Programme 2007–2008, <http://www.3d4you.eu>.

[11] H. Ha and C. Yim, "Layer-Weighted Unequal Error Protection for Scalable Video Coding Extension of H.264/AVC," IEEE Trans. Consumer Electron., 54:2 (2008), 736–744.

[12] International Organization for Standardization and International Electrotechnical Commission, "Report of the Subjective Quality Evaluation for MVC Call for Evidence," ISO/ IEC JTC1/SC29/WG11, Doc. N6999, Jan. 2005.

[13] International Organization for Standardization and International Electrotechnical Commission, "Applications and Requirements on 3D Video Coding," ISO/IEC JTC1/SC29/WG11, Doc. N10570, Apr. 2009.

[14] International Organization for Standardization and International Electrotechnical Commission, "Call for Proposals on 3D Video Coding Technology," ISO/IEC JTC1/SC29/WG11, Doc. N12036, Mar. 2011.

[15] International Telecommunication Union, Radiocommunication Sector, "Subjective Assessment of Stereoscopic Television Pictures," ITU-R Rec. BT.1438, Mar. 2000, <http://www.itu.int>.

[16] International Telecommunication Union, Radiocommunication Sector, "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU-R Rec. BT.500-11, June 2002, <http://www.itu.int>.

[17] International Telecommunication Union, Telecommunication Standardization Sector, "Advanced Video Coding for Generic Audiovisual Services," ITU-T Rec. H.264, Mar. 2010, <http://www.itu.int>.

[18] P. Joveluro, H. Malekmohamadi, W. A. C. Fernando, and A. M. Kondoz, "Perceptual Video Quality Metric for 3D Video Quality Assessment," Proc. 3DTV Conf.: The True Vision—Capture, Transmission and Display of 3D Video (3DTV-CON '10) (Tampere, Fin., 2010), pp. 1–4.

[19] R. G. Kaptein, A. Kuijsters, M. T. M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx, "Performance Evaluation of 3D-TV Systems," Proc. IS&T/SPIE Annual Symp. on Electron. Imaging: Sci. and Technol. (EI '08) (San Jose, CA, 2008), Conf. on Image Quality and Syst. Perform. V, SPIE vol. 6808, session 10, paper 6808 19.

[20] D. Kim and K. Sohn, "Depth Adjustment for Stereoscopic Image Using Visual Fatigue Prediction and Depth-Based View Synthesis," Proc. IEEE Internat. Conf. on Multimedia and Expo (ICME '10) (Singapore, Sgp., 2010), pp. 956–961.

[21] M. T. M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx, "Visual Discomfort in Stereoscopic Displays: A Review," Proc. IS&T/SPIE Annual Symp. on Electron. Imaging: Sci. and Technol. (EI '07) (San Jose, CA, 2007), Conf. on Stereoscopic Displays and Virtual Reality Syst. XIV, SPIE vol. 6490, paper 64900I.

[22] T.-L. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. C. Cosman, and A. R. Reibman, "A Versatile Model for Packet Loss Visibility and Its Application to Packet Prioritization," IEEE Trans. Image Process., 19:3 (2010), 722–735.

[23] A. Manta, "Multiview Imaging and 3D TV. A Survey," Delft University of Technology, Information and Communication Theory Group, Jan. 2008.

[24] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal Loss Protection: Graceful Degradation of Image Quality over Packet Erasure Channels Through Forward Error Correction," IEEE J. Select. Areas Commun., 18:6 (2000), 819–828.

[25] R. Pantos (ed.) and W. May, "HTTP Live Streaming," IETF Internet Draft, Mar. 31, 2001, <http://tools.ietf.org/html/draft-pantos-http-live-streaming-06>.

[26] P. Pérez and N. García, "Lightweight Multimedia Packet Prioritization Model for Unequal Error Protection," IEEE Trans. Consumer Electron., 57:1 (2011), 132–138.

[27] D. C. Robinson and A. Villegas Nunez, "Intelligent Wrapping of Video Content to Lighten Downstream Processing of Video Streams," European Patent EP2071850 (2009).

[28] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," IEEE Trans. Circuits Syst. Video Technol., 17:9 (2007), 1103–1120.

[29] J. Shade, S. Gortler, L.-W. He, and R. Szeliski, "Layered Depth Images," Proc. 25th Internat. Conf. on Comput. Graphics and Interactive Techniques (SIGGRAPH '98) (Orlando, FL, 1998), pp. 231–242.

[30] Y. Shi, C. Wu, and J. Du, "A Novel Unequal Loss Protection Approach for Scalable Video Streaming over Wireless Networks," IEEE Trans. Consumer Electron., 53:2 (2007), 363–368.

[31] Spain, Ministry of Science and Innovation, Centre for Technological Industrial Development (CDTI), Ingenio 2010, National Strategic Consortiums for Technological Research (CENIT), "VISION Project," <http://www.cenit-vision.org>.

[32] A. S. Tan, A. Aksay, C. Bilen, G. B. Akar, and E. Arikan, "Error Resilient Layered Stereoscopic Video Streaming," Proc. 3DTV Conf.: The True Vision—Capture, Transmission and Display of 3D Video (3DTV-CON '07) (Kos Island, Grc., 2007), pp. 1–4.

[33] M. Tanimoto, "FTV (Free-Viewpoint TV)," Proc. IEEE Internat. Conf. on Image Process. (ICIP '10) (Hong Kong, Chn., 2010), pp. 2393–2396.

[34] A. Tikanmäki, A. Gotchev, A. Smolic, and K. Miller, "Quality Assessment of 3D Video in Rate Allocation Experiments," Proc. 12th IEEE Internat. Symp. on Consumer Electron. (ISCE '08) (Vilamoura, Prt., 2008), pp. 1–4.

[35] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y.-K. Wang, "Joint Draft 8.0 on Multiview Video Coding," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Doc. JVT-AB204, July 2008.

[36] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," Proc. IEEE, 99:4 (2011), 626–642.

[37] W.-C. Wen, H.-F. Hsiao, and J.-Y. Yu, "Dynamic FEC-Distortion Optimization for H.264 Scalable Video Streaming," Proc. 9th IEEE Workshop on Multimedia Signal Process. (MMSP '07) (Chania, Crete, Grc., 2007),pp. 147–150.

[38] S. Winkler, Digital Video Quality: Vision Models and Metrics, John Wiley & Sons, Chichester, West Sussex, Hoboken, NJ, 2005.

[39] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality Analysis for 3D Video Using 2D Video Quality Models," IEEE Trans. Consumer Electron., 54:4 (2008), 1969–1976.

[40] C. Yuan, H. Pan, and S. Daly, "Stereoscopic 3D Content Depth Tuning Guided by Human Visual Models," Proc. 49th Internat. SID Symp., Seminar, and Exhibition (SID Display Week '11) (Los Angeles, CA, 2011), pp. 3–6.

[41] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-Quality Video View Interpolation Using a Layered Representation," ACM Trans. Graphics, 23:3 (2004), 600–606.

(Manuscript approved October 2011)

JOSÉ MARÍA CUBERO is an IPTV and multimedia engineer in Alcatel-Lucent's Multimedia Integration Practice, in the Video & Networking Competence Center in Madrid, Spain. He holds an ingeniero de telecomunicación degree (five-year telecommunications engineering program) from the Universidad Politécnica de Madrid (UPM), Spain. Mr. Cubero is on the Alcatel-Lucent 5910 Video Services Appliance (VSA) product team, and also has a key role in several public co-funded national and European research and development (R&D) projects. His professional and research interests are in the area of multimedia, particularly advanced video services, video processing, and S3D content delivery.

JESÚS GUTIÉRREZ holds a telecommunication engineering degree (five-year engineering program) from the Universidad Politécnica de Valencia (UPV), Spain. He is a member of the Grupo de Tratamiento de Imágenes (Image Processing Group) at the Universidad Politécnica de Madrid (UPM), Spain, where he is currently working toward a Ph.D. degree in telecommunication. His research interests are in the area of subjective and objective multimedia quality of experience evaluation and 3D video processing.

PABLO PÉREZ is an IPTV solution system architect in the Multimedia Integration Practice at Alcatel-Lucent's Network and Systems Integration division in Madrid, Spain. He is the technical lead for the design of the Alcatel-Lucent 5910 Video Services Appliance (VSA)

product, and also has a key role in several public co-funded research and development (R&D) projects. Mr. Pérez received the ingeniero de telecomunicación degree (five-year telecommunications engineering program) from the Universidad Politécnica de Madrid (UPM), Spain and was first in his year. His professional and research interests are in the area of multimedia quality of experience modeling and management, as well as video services enabling.

ENRIQUE ESTALAYO was an IPTV development engineer in Alcatel-Lucent's Multimedia Integration Practice when this paper was written, and worked in the Video & Networking Competence Center in Madrid, Spain. He was part of the Alcatel-Lucent 5910 Video Services Appliance (VSA) product core team, and also actively participated in several public co-funded R&D projects. Mr. Estalayo received the 5-year telecommunications engineering degree at Universidad Politécnica de Madrid, Spain. His professional interests are focused in the areas of enhanced video services, computer vision, and 3D graphics.

JULIÁN CABRERA holds a telecommunication engineering degree, and a Ph.D. degree in telecommunication, both from the Universidad Politécnica de Madrid (UPM), Spain. In addition, he is a member of the UPM Image Processing Group. Dr. Cabrera was a Ph.D. scholar in the Information Technology and Telecommunication Programs of the Spanish National Research Plan, as well as a member of the UPM faculty, and is currently an associate professor of signal theory and communications. His professional interests include image and video coding, and design and development of multimedia communications systems, focusing on multiview video coding, 3D video coding, and video transmission over variable rate channels. He has been actively involved in European projects (Acts, Telematics, and IST), and national projects in Spain.

FERNANDO JAUREGUIZAR holds a six-year telecommunication engineering degree, and a Ph.D. degree in telecommunication, both from the Universidad Politécnica de Madrid (UPM), Spain. In addition, he is a member of the Image Processing Group at UPM. Dr. Jaureguizar is a member of the faculty of the E.T.S. Ingenieros de Telecomunicación at UPM, and an associate professor of signal theory and communications at the Department of Signals, Systems, and Communications. His professional interests include digital image processing, video coding, 3DTV, and design and development of multimedia communications systems. He has been actively involved in European projects (Eureka, Acts, and IST), and national projects in Spain.

NARCISO GARCÍA holds an ingeniero de telecomunicación degree (five year engineering program) (Spanish National Graduation Award), and a doctor ingeniero de telecomunicación degree (Ph.D. in communications) (Doctoral Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Spain. He is a member of the faculty at UPM, where he is currently a professor of signal theory and communications, and also leads the Grupo de Tratamiento de Imágenes (Image Processing Group). Dr. García served as coordinator for the Spanish Evaluation Agency from 1990 to 1992, and as an evaluator, reviewer, and auditor of European programs since 1990. His professional and research interests are in the areas of digital image and video compression, and computer vision. ♦