

SUBJECTIVE STUDY OF ADAPTIVE STREAMING STRATEGIES FOR 3DTV

Jesús Gutiérrez , Pablo Pérez , Fernando Jaureguizar , Julián Cabrera , and Narciso García

ABSTRACT

Although the delivery of 3D video services to households is nowadays a reality thanks to frame-compatible formats, many efforts are being made to obtain efficient methods to transmit 3D content offering a high quality of experience to the end users. In this paper, a stereoscopic video streaming scenario is considered, and the perceptual impact of various strategies applicable to adaptive streaming situations are compared. Specifically, the mechanisms are based on switching between copies of the content with different coding qualities, on discarding frames of the sequence, on switching from 3D to 2D, and on using asymmetric coding of the stereo views. In addition, when video freezes happen, the possibility of keeping the end-to-end latency or maintaining the continuity of the video are considered. These aspects were evaluated carrying out a subjective assessment test, considering also visual discomfort issues, using a methodology designed to keep as far as possible domestic viewing conditions.

Index Terms— 3DTV, Quality of Experience, adaptive streaming, subjective evaluation, visual discomfort

1. INTRODUCTION

Nowadays, the offer of video services over packet-based networks is even greater than in classical broadcasting systems like satellite, cable, or terrestrial television. One of the main causes of this fact is the facility to access those services from many different scenarios than the traditional domestic environment, thanks to the possibility of using other devices, in addition to typical television sets, such as personal computers, laptops, tablets, or mobile phones.

In this type of networks the development of traffic management tools has been a crucial aspect due to the varying transmission conditions and scenarios of the distribution systems. These techniques are aimed to obtain a flexible and efficient data transmission to cope with adverse situations in the network. In fact, when video services are delivered, this is especially a critical aspect to guarantee a satisfactory visual experience to the end users, whose quality expectations are more and more demanding.

Therefore, some techniques have been proposed to face those events that reduce the network performances, trying to keep the quality level as high as possible. For instance, packet prioritization models have been developed to cope with network losses, taking into account that video packets have information that could influence differently the user quality of experience (QoE) [1]. Other examples are the techniques applied when the quality of service (QoS) of the network drops, and the amount of data to be transmitted should be reduced. For example, discarding frames of the video sequence or

transmitting lower-quality versions of the video content. This latter method is also applicable in cases in which the same content should be delivered to devices with different technologies. For example, mobile devices do not provide full-HD resolution as TV sets do, so it is not necessary to transmit the same information to both systems.

Although the research on adaptive streaming mechanisms for conventional video is still an open issue, the arrival of 3D television (3DTV) services broadcasted to households has motivated the investigation on similar methods to control the efficient transmission of 3D content. Currently, stereoscopic video is transmitted in frame-compatible format, which means that the same infrastructure (i.e., encoder, network, and decoder) as for classical video is used. This approach has allowed the rapid introduction of 3DTV services as a quick response of the broadcasting industry to the good reception of 3D movies at cinemas by the public. However, this format has the main drawback that the video resolution is reduced to multiplex the stereo views into one single typical video stream. Therefore, many efforts are being done to develop coding techniques to make possible the efficient transmission of high-quality 3D content, reducing as much as possible the data to be transmitted, especially taking advantage of the similarity between the stereo views. It is worth noting the work carried out towards the elaboration of coding standards applicable to 3D video, such as multi-view coding (MVC), scalable video coding (SVC), or 3D video coding (3DVC) currently under development. Taking this into account, the first approaches for 3DTV adaptive streaming could be based on strategies developed for conventional video, since frame-compatible transmission is used. However, new possibilities should be analyzed taking into account the specific properties of stereoscopic content.

Moreover, since video services are addressed to people, to identify the most appropriate mechanism to apply when adverse events take place in the network, extensive studies of the effects of these strategies on the user QoE should be done by means of subjective tests.

Therefore, in this paper we present a subjective study comparing different strategies to deliver 3D video content in varying network conditions. In Section 2, the considered mechanisms are detailed. In Section 3 the subjective experiment carried out to compare the perceptual effects of the streaming techniques is described. The obtained results are presented in Section 4, and finally, in Section 5 some conclusions are exposed.

2. ADAPTIVE STREAMING STRATEGIES FOR 3DTV

Taking into account the existing adaptive streaming strategies for conventional video and the approaches for transmitting 3D content, four specific mechanisms have been considered in the current work. These techniques allow the creation of versions of a video sequence with lower bitrate to be transmitted in cases when the QoS of the network is reduced, with the objective of affecting as less as possible

the QoE perceived by the end users.

The first strategy is called *denting*, and consists of dropping some frames of the video sequence. It is one of the most simple methods and has been already applied in IPTV scenarios [2]. Usually, the effect of this technique is a reduction on the framerate of the displayed video, which could degrade drastically the visual experience of the viewers.

Another technique that is recently attaining more importance is the adaptive streaming of the same content with different coding quality depending on the network conditions [3]. In this case, the original content should be encoded with various bitrates, and then, when the transmitted data should be reduced, the system switches to a lower-quality version. This version should be delivered correctly and displayed smoothly at the end user side. The switching from one version to another should be done in pre-fixed points of the video in order to maintain the continuity of the sequence and in a transparent way to the user. Therefore, the copies of the video are divided in chunks, which can be independently decoded. The effect on the QoE of the viewer should only be a drop in the coding quality of the video.

While the previous strategies have been already applied in broadcasting scenarios of conventional video, the research on adaptive streaming approaches for 3DTV services is taking its initial steps, and the first approaches are being proposed. For example, Gürlér *et al.* proposed a mechanism based on the concept of asymmetric coding using SVC [4]. This idea consists of encoding one of the stereo views with lower quality than the other. It is supported on the binocular suppression theory, which states that the degradations on one stereo view could be masked by the other view if the structural information of the image is not affected. Therefore, this method has been considered in the comparison of the present work.

The last strategy considered is somehow the equivalence for 3DTV of the adaptive streaming using various versions of different coding quality. In this case a stereoscopic and a monoscopic version of the same content should be available. Then, when the network could not withstand the data load needed for 3DTV, it is possible to switch to the conventional version of the video reducing the bitrate to transmit. This technique has been already analyzed in the context of error concealment techniques for 3DTV [5].

In addition, in video streaming services a common situation is that the receiver video buffer turns empty, causing a video freeze. Usually, when the playback could continue, the end-to-end latency is maintained by restarting the playback from the frame corresponding to the actual instant. However, in some cases, it is possible to continue the playback from the subsequent frame to the last displayed (and frozen), keeping the continuity of the video with the drawback of increasing the end-to-end delay. This could be especially important in the streaming of live content. Besides, the study of the impact of freezes of different durations could be helpful in the selection of the chunks for the adaptive streaming strategies.

3. SUBJECTIVE EXPERIMENT

A comparison of the perceptual effects of the described strategies on the QoE of the end users was done by means of a subjective assessment test according to the following settings.

3.1. Evaluation Methodology

The subjective assessment tests should be carried out according to certain conditions that should be defined taking into account the specific purpose of the experiments, in order to obtain conclusive re-

sults. In the current study the objective is to know the impact on the viewer QoE of various situations that take place in video transmission scenarios. Therefore, the subjective experiment should respect as far as possible home viewing conditions. This aspect determines several factors of the design of the assessment methodology. For example, the sequences used in the tests should be similar to those usually watched by people in their homes, to attract the attention of the observers that would be interested in the content itself and not focused on detecting impairments. Moreover, the sequences should be shown to the observers following a single stimulus procedure, which means that no unimpaired reference is presented to them to compare with the test video, in order to not distance the test from home environment scenarios. Finally, the evaluation should be carried out in a nearly continuous way, since the effects of transmission events are highly dependent on the instant when they occur, and they are not stationary.

Considering these aspects, the most common standard methodologies should be discarded, since they were generally designed to evaluate the performance of coding techniques and the effects of the impairments introduced by them. Then short sequences (e.g., ten seconds) are used with content focused on testing the capabilities of the encoders, for example with colorful or high-motion sequences. The closest standard approach to the desired methodology is the Single Stimulus Continuous Quality Evaluation (SSCQE) [6]. However in the current study it is not needed a strictly continuous evaluation, since there are no continuous quality variations. In fact, the idea is to use a similar evaluation as in the Absolute Category Rating (ACR) [7] based on segments. However, in that case, various short sequences are evaluated during grey segments, which is far from domestic viewing conditions.

Therefore, a particular test methodology was designed to evaluate the effects of transmissions artifacts on the end users QoE. This procedure has been successfully used in [8], where a detailed explanation could be found. One of the main aspects of the methodology is to use long sequences with common content of TV services, instead of several short sequences. The test sequences are divided in segments and the degradations under study are inserted in alternate segments. Then, each degradation is evaluated in the following segment which is unimpaired, using the five-degree impairment scale recommended in [6]. This way, the observers could provide their scores without losing any degradation, and the continuity of the whole video is maintained, which helps to keep the attraction of the subjects.

3.2. Test material

Following the design condition regarding the test sequences mentioned before, two stereoscopic video were selected with contents like those typically watched in real domestic situations. The main properties of the test sequences used in the subjective experiment are described in Table 1. The test sequences have a duration of five minutes, and according to the methodology previously described, they were divided in segments of six seconds. They have been encoded using the standard H.264/AVC, with a reference bitrate of 12Mbps.

Source	Format	Content
SRC1	1920x1080p 24fps	Movie containing segments with slow motion and dialogs, and others with fast motion activity.
SRC2	1280x720p 50fps	Football match with commentaries as soundtrack.

Table 1. Test sequences

Code	Description
R1	Bitrate drop of both views to the half of the reference
R2	Bitrate drop of both views to a quarter of the reference
F1	Framerate drop to the half of the reference
F2	Framerate drop to a quarter of the reference
D1	Switch to 2D coded with the half of the reference
D2	Switch to 2D coded with a quarter of the reference
AS1	Asymmetric coding. Left view encoded with 3/8 of the reference bitrate and right view with 1/8
AS2	Asymmetric coding. Left view encoded with 3/16 of the reference bitrate and right view with 1/16
CS1	Video freeze of 2 seconds and skip
CS2	Video freeze of 4 seconds and skip
CC1	Video freeze of 2 seconds and continue
CC2	Video freeze of 4 seconds and continue

Table 2. Considered strategy patterns

In the current work, to compare the aforementioned strategies related to 3D content streaming, two situations of bitrate drops were considered: one to the half of the reference bitrate (i.e., 6Mbps) and other to a quarter of the reference (i.e., 3Mbps). Therefore, two patterns for each mechanism were simulated corresponding to these two cases to make an equitable comparison. The description of all the patterns is detailed in Table 2. Also, the different durations of the video freezes considered for the two playback possibilities described before are detailed.

3.3. Environment and Equipment

The subjective tests were carried out at the Universidad Politécnic de Madrid, in a lab especially fitted-out to satisfy the recommendations given in ITU-R BT.500 [6] for assessments in home environments. The display used for presenting the sequences was a 42" Panasonic stereoscopic television, with resolution of 1920x1080 and aspect ratio 16:9. To visualize the 3D sequences, the observers wore the active shutter glasses associated to this TV set. The subjects were positioned at a viewing distance of 3 times the height of the TV set, following the indications of the manufacturer.

3.4. Observers

In the subjective experiment participated 36 observers, all of them having normal visual and stereoscopic acuity, and color vision. The ages of the subjects were ranged between 20 and 44 years old, with an average age of 25.

4. RESULTS

The evaluations collected from the observers that participated in the subjective test were processed to obtain the Mean Opinion Scores (MOS) and the 95% confidence intervals of the values. In addition, a screening of the results was done to discard possible random evaluations. The obtained results are presented in the following figures.

Specifically, in Figure 1 a comparison of the subjective ratings obtained for the considered adaptive streaming strategies for both test videos are shown. As it can be seen, in both video sources and all the mechanisms, in the case of more severe bitrate reduction the degradations caused are more annoying for the observers. Focusing on one of the test sequences and on one of the considered network bitrate drops, for example SRC1 and the reduction to the half of the original bitrate (i.e., strategies represented with codes ending in

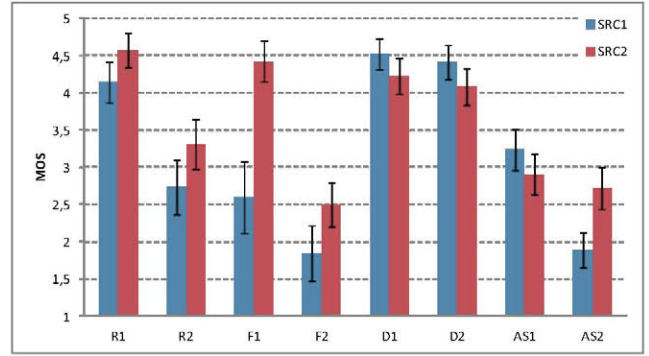


Fig. 1. Results for the considered strategies for each source

“1”)), it is possible to analyze which of the strategies could be preferred from the end user point of view. As it is shown, the impact of switching from 3D to 2D is less annoying than the rest of the mechanisms (D1), followed by applying the same bitrate reduction to both views (R1). In comparison to this case, the severe bitrate reductions applied to the right view in the case of asymmetric coding (AS1) causes a worse performance of this strategy. Finally, the option of discarding frames produces a framerate reduction which is very annoying, since the natural flow of the video is corrupted. Regarding this last strategy, it is worth noting that in the case of SRC2 the annoyance is significantly reduced, due to the fact that the original framerate of the sequence is 50fps. In this case, the performance of this mechanism could be as good as those provided by switching to 2D or reducing the bitrate similarly for both views.

To further analyze the performances of the considered strategies with the aim of identifying the most appropriate, the ratings provided by the observers for both test sequences were grouped, obtaining global evaluations. Since both test videos have different framerates, the effects of that strategy have not been considered in the global comparison. However, as aforementioned, it could only be a good option when the original framerate is high enough to make the reductions imperceptible. For example, as it occurs with SRC2, reducing to the half a framerate of 50fps entails a resulting framerate of 25fps, which is sufficient to not perceive video discontinuities. As it can be seen in Figure 2, switching to 2D and reducing the bitrate by symmetric coding of the views are the best options when the bitrate drop in the network is not too severe. However, this latter option does not provide a good performance in the case of a reduction to a quarter of the original bitrate. In this situation, the best option is to transmit a monoscopic version of the video with lower coding quality than the

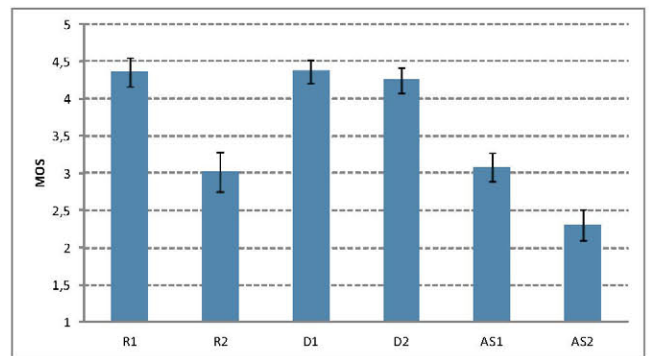


Fig. 2. Global results for the considered strategies

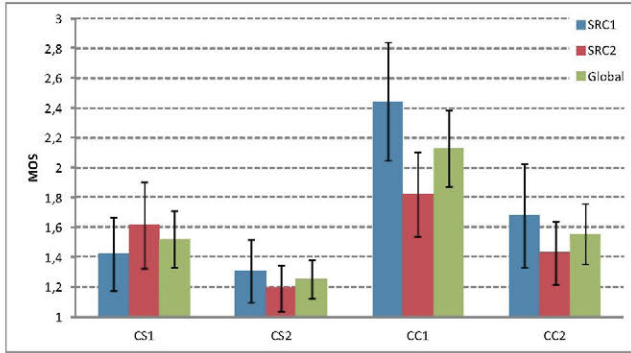


Fig. 3. Results for the video freeze events

original, since as the results show, the annoyance of the transition from 3D is limited. Nevertheless, it is obvious that this effect could be worse depending on the disparity and 3D perception related to the visualized sequence.

Finally, in Figure 3 the results obtained for the analyzed video freeze patterns are shown. Specifically, the bars show the MOS values for each test sequence and for the global evaluation considering the results for both sequences. Although the overlap of some confidence intervals of the values corresponding to the two video sources could indicate that the differences are not significant, the important comparison is mainly between global results for the two strategies and a certain duration. As expected, the longer the duration of the freeze, the worse the degradation of the users QoE is. Besides, the option of continuing after the video freeze from the last displayed frame keeping the continuity of the video is preferred. However, these aspects should be further studied since in the case of streaming live events the increase of the end-to-end latency could be a critical factor. For example, in the case of a football match, it could happen that the viewer sees a goal with a certain delay respect to the reality. On the other hand, when prerecorded content is streamed, this option could be the best for the user visual experience.

In addition to these results, the subjects were asked to indicate if the effects of the event evaluated in a certain segment caused difficulties in convergence or in the fusion of the stereo views. As explained in [8], the evaluation of the degradations was done using a questionnaire where the observers should write a cross in the corresponding box for the evaluated segment. Then, in the case that they have experienced vision difficulties, they had to write just a line, instead of a cross. The results showed that more than a 60% of the observers noticed the transitions from 3D to 2D, and nearly the half of the observers experienced binocular difficulties in the segments with asymmetric coding. Also, nearly a 30% of the observers felt similar issues with severe symmetric reductions of the coding quality, and annoyance with framerate drops.

5. CONCLUSION

Lately, service providers have started offering broadcasted 3D video to end users. However, many efforts are being made focused on the research on efficient methods for encoding and transmitting high-quality 3D video. Furthermore, the importance that packet-based networks have achieved in the delivery of video services motivate the development of that type techniques for IP-based systems. In addition to new coding algorithms and transmission formats for 3DTV, an important issue concerning efficient transmission is the development of adaptive streaming strategies to be applied when the network

could not provide the sufficient performance to transmit high-quality 3D content. Therefore, in this paper a comparative study among different strategies for adaptive streaming of 3D video is presented. The comparison is done from subjective evaluations of the perceptual effects of the strategies. In addition, the effects of the considered impairments on the visual comfort of the observers were analyzed.

Two of the considered approaches have been already used in the transmission of conventional video over packet-based networks. The first one consists of discarding frames to reduce the overall bitrate, while the second is based on the adaptive selection of versions with different coding quality of the sequence to deliver. In addition to these techniques, two specific approaches for 3D video were also analyzed. The first one consists on switching from 3D to a 2D version of the video, while the second one is based on the asymmetric coding of the stereo views. The results show that switching to a monoscopic version of the video sequence is normally the best option.

Furthermore, two possibilities to be applied when the network impairments cause video freeze were considered. The first one consists of continuing the playback with the frame corresponding to the actual instant keeping the end-to-end latency, while the second continues from the last displayed frame, maintaining the continuity of the video. The subjective results show that this latter approach is usually preferred by the users, but further studies should be carried out considering the effects of the consequent delay in streamed live events, since their effects on the users visual experience could be critical.

6. REFERENCES

- [1] P. Pérez and N. García, "Lightweight multimedia packet prioritization model for unequal error protection," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 132–138, Feb. 2011.
- [2] A. Dehili, "Evolution of access networks to support video quality requirements," *Proc. Telecommunications Network Strategy and Planning Symposium*, New Delhi, India, pp. 1–3, Nov. 2006.
- [3] T. Stockhammer, "Dynamic adaptive streaming over HTTP - Standards and design principles," *Proc. ACM Conference on Multimedia Systems*, New York, USA, pp. 133–143, Feb. 2011.
- [4] C. G. Gürlér, K. T. Bağci, and A. M. Tekalp, "Adaptive stereoscopic 3D video streaming," *Proc. IEEE International Conference on Image Processing*, Hong Kong, China, pp. 2409–2412, Sept. 2010.
- [5] 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. International Packet Video Workshop*, Hong Kong, China, Dec. 2010.
- [6] ITU, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation ITU-R BT.500-12*, Sept. 2009.
- [7] ITU, "Subjective video quality assessment methods for multimedia applications," *ITU-T Recommendation P.910*, Apr. 2008.
- [8] J. Gutiérrez, P. Pérez, F. Jaureguizar, J. Cabrera, and N. García, "Subjective assessment of the impact of transmission errors in 3DTV compared to HDTV," *Proc. IEEE 3DTV Conference*, Antalya, Turkey, pp. 1–4, May 2011.