

Perceptual Distortion Modeling for Side-by-Side 3D Video Delivery

César Díaz, Jesús Gutiérrez, Julián Cabrera, Fernando Jaureguizar, and Narciso García

Abstract—A frame-level distortion model based on perceptual features of the human visual system is proposed to improve the performance of unequal error protection strategies and provide better quality of experience to users in Side-by-Side 3D video delivery systems

I. INTRODUCTION

To allow compatibility with deployed HDTV infrastructure, the first phase DVB 3DTV delivery system enables the use of frame compatible formats to distribute stereoscopic 3D content [1]. In this sense, the Side-by-Side (SbS) format has become a very popular choice.

In this paper, we focus on AVC-encoded SbS video streaming over packet-switched networks. In this type of communication channels, packet losses may severely impair the transmitted video sequences, resulting in a heavy degradation of the end user Quality of Experience (QoE). Hence, error control mechanisms need to be introduced [2].

As different parts of the encoded sequence are of varying importance to the overall quality perception due to error propagation, unequal error protection (UEP) schemes are usually introduced to face the unreliability of the channel. To find the best protection policy to follow, most UEP schemes resort to distortion models to estimate the unequal importance of video data regarding the quality degradation that their loss might cause. Data are then ranked and protected accordingly.

Simple frame-level distortion models usually just turn to frame types for data prioritization [3]. So, in the case of non-hierarchical structures, I-frames are always ranked the most relevant, followed by P-frames. Finally, B-frames are equally considered of a minor importance. However, B-frames are not equally important from a perceptual point of view, as the loss of different frames unequally affect QoE. Especially for stereoscopic video, the loss of different frames cause dissimilar levels of binocular rivalry and visual discomfort [4].

Thus, we study the use of perceptual features to derive the distortion model with the aim of maximizing the end user QoE. These features are based on a Just Noticeable Distortion (JND) model, since they can provide information about the maximum distortion that the signal can suffer without being perceived [5]. These models have been extensively used in the field of audio and video compression and also applied to quality metrics, thanks to their direct relation to the perceptual behavior of the human perception [6].

II. SBS-FORMAT-AWARE PERCEPTUAL CLASSIFICATION

To take into account in the protection system the effects on the user QoE of the degradations caused by packet losses, a distortion model incorporating perceptual features of the human visual system (HVS) is considered. The proposed approach consists of the computation of some perceptual indicators that are based on the JND model for transmission errors presented in our previous work [4].

In particular, a value is computed for each frame of the video to be delivered (i.e., undistorted) representing its perceptual importance, which can be used in the distortion model to protect the more relevant frames. This value is obtained from the sensitivity thresholds calculated for each macroblock (MB), taking the greater masking effect between the contrast and the texture features, after weighting them with a temporal threshold. The component related to intensity contrast is based on the higher sensitivity of the HVS to luminance variations with respect to the environment of a stimulus, than to the absolute luminance of the stimulus, as stated by the Weber-Fechner law. In fact, a piecewise approximation of the sensitivity threshold was used [6]. Moreover, the temporal masking factor is modeled as a function of the luminance difference between corresponding MB of contiguous frames [5]. Finally, the threshold related to the local activity of the environment of the analyzed MB is computed from the difference between the maximum and the minimum mean luminance of the neighboring MBs [6]. In addition, to reflect the effects related to the fusion of both stereo views, corresponding MBs in each stereo view are used in the computation of the luminance and texture thresholds.

The most appropriate value to represent the perceptual importance of the frame is selected considering the specific characteristics of the scenario. In our case, the error concealment technique employed when packet losses occur is based on the substitution of the whole frame by the previous one. Therefore, the maximum value among all of those computed for each MB of the frame is chosen, since it will reflect better the differences between adjacent frames.

III. FRAME-LEVEL DISTORTION MODEL

The proposed model consists of two steps. In the first one, frames are prioritized in a coarse fashion just regarding the GOP coding structure and the effects of error propagation, i.e. the more frames depend on a certain reference frame for decoding, the more distorting its loss might be and, hence, the more relevant this reference frame is considered.

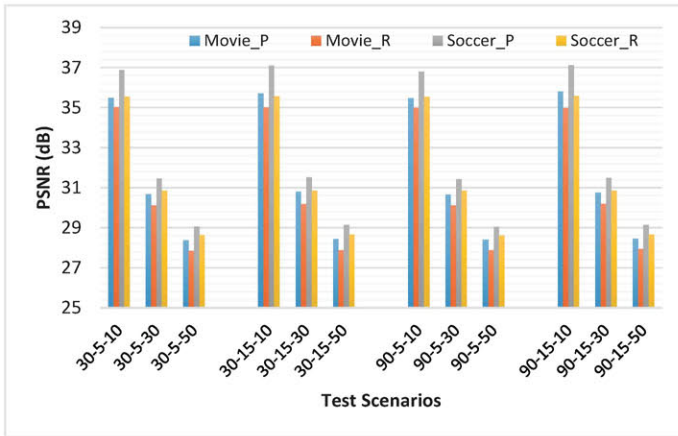


Fig. 1. Results obtained using PSNR

In the second step the proposed SbS-format-aware perceptual classification comes into play. The frame sorting resulting from the first step is refined by ranking the non-reference frames considering the values obtained after applying the approach described above.

The reason for only applying our proposal to non-reference frames is that the coding structure always masks the intrinsic perceptual values of the frames. Even if a reference frame is unimportant from a perceptual perspective, losing this frame implies introducing a greater distortion than that caused by the loss of a perceptually more important non-reference frame. Nevertheless, despite only operating on non-reference frames, there still exists room for improving the end user QoE, as simulation results show.

IV. SIMULATION AND RESULTS

To evaluate the performance of the proposed distortion model, a set of scenarios have been considered.

First, regarding the test videos employed, they were selected with the intention of using content similar to that usually watched by people in their homes, so a segment of a movie and a soccer match were chosen. The format of the sequences was SbS and 1080p for the movie and 720p for the soccer match, both of them with 25fps and a duration of one minute. These sequences were encoded using the four different non-hierarchical GOP coding structures defined by the combinations of two GOP sizes (30 and 90 frames) and two distances between two anchor frames (5 and 15 frames). Just one slice was used per frame.

To analyze the implications of using the proposed perceptual frame classification, we have compared it to a second frame arrangement in which the non-reference frames are ranked randomly. The two options are labelled in the figures with the extensions “_P” and “_R”, respectively.

Finally, the comparison was carried out dropping different percentages of B-frames (10%, 30%, and 50%). The less important B-frames regarding each ranking were discarded. Then an error concealment method consisting in copying the previous frame was applied to obtain the distorted sequences.

The resulting sequences were evaluated using Peak-Signal-to-Noise-Ratio (PSNR) and Video Quality Metric (VQM) [7],

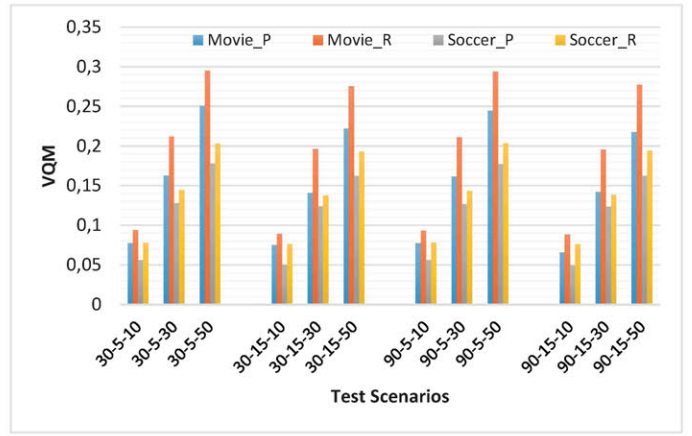


Fig. 2. Results obtained using VQM

which is a standardized full-reference video quality metric that can provide results more correlated with subjective assessments. As it can be seen, using the perceptual indicators in the distortion model provides better results with both metrics in all the test cases (it is worth noting that VQM provides quality estimations in the range from 0 to 1, the former meaning that there is no degradation).

V. CONCLUSIONS AND FUTURE WORK

In this paper, we use HVS perceptual features to derive a frame-level distortion model to better protect SbS 3D video distribution. Results show the unequal importance of B-frames, according to the quality degradation that their loss cause, and how our distortion model takes advantage of this fact to allow delivering better QoE to end users.

Ongoing work is first focused on carrying out subjective tests to more strongly confirm these results. Additionally, a new version of the distortion model is being created to offer intra-frame-level data ranking, which jointly with slicing error concealment, is expected to better exploit JND characteristics.

REFERENCES

- [1] DVB, “Frame compatible plano-stereoscopic 3DTV,” DVB Document A154, Feb. 2011.
- [2] A.C. Begen, “Error Control for IPTV over xDSL Networks,” *IEEE Consumer Communications and Networking Conference, 2008*, pp. 632–637, 10-12 Jan. 2008.
- [3] C. Díaz, J. Cabrera, F. Jaureguizar, and N. García, “A video-aware FEC-based unequal loss protection system for video streaming over RTP,” *IEEE Trans. on Consumer Electronics*, vol. 57, no. 2, pp. 523–531, May 2011.
- [4] J. Gutiérrez, P. Pérez, F. Jaureguizar, J. Cabrera, and N. García, “Monitoring packet loss impact in IPTV and 3DTV receivers,” *IEEE International Conference on Consumer Electronics*, pp. 174–175, Jan. 2012.
- [5] X. Yang, W.S. Ling, Z. Lu, E. Ong, and S. Yao, “Just noticeable distortion model and its applications in video coding,” *Signal Processing: Image Comm.*, vol. 20, no. 7, pp. 662–680, Aug. 2005.
- [6] W. Lin, L. Dong, and P. Xue, “Visual distortion gauge based on discrimination of noticeable contrast changes,” *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 15, no. 7, pp. 900–909, Jul. 2005.
- [7] ITU, “Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference,” *Recommendation ITU-T J.144*, Mar. 2004.