Rate-Distortion Optimized Low-Delay 3D Video Communications

Enrico Masala

Control and Computed Engineering Department
Politecnico di Torino
corso Duca degli Abruzzi, 24 — I-10129 Torino, Italy
masala@polito.it

***Abstract*—This paper focuses on the rate-distortion optimization of low-delay 3D video communications based on the latest H.264/MVC video coding standard. The first part of the work proposes a new low-complexity model for distortion estimation suitable for low-delay stereoscopic video communication scenarios such as 3D videoconferencing. The distortion introduced by the loss of a given frame is investigated and a model is designed in order to accurately estimate the impact that the loss of each frame would have on future frames. The model is then employed in a rate-distortion optimized framework for video communications over a generic QoS-enabled network. Simulations results show consistent performance gains, up to 1.7 dB PSNR, with respect to a traditional a priori technique based on frame dependency information only. Moreover, the performance is shown to be consistently close to the one of the prescient technique that has perfect knowledge of the distortion characteristics of future frames.**

I. Introduction

Videocommunications with depth perception are increasingly gaining popularity and are expected to become an important share of video applications in the next few years. One of the most commonly used technique to achieve depth perception is to send separate video information to each eye, since this has been demonstrated to be a sufficient condition to perceive depth [1]. Video coding standards are addressing the issue of stereoscopic video by either including new profiles in existing codecs, such as the Stereo High Profile in H.264/AVC [2] or substantially modifying state-of-the-art video coding standards to support stereoscopic or even multiview video, such as in the recently developed H.264/MVC [3].

In order to achieve a good compression ratio, such coding algorithms exploit the correlation among different views at the same time instant by means of disparity compensation, in analogy with traditional video codecs which perform efficient differential encoding between subsequent frames by means of motion compensation. On the one hand, disparity compensation allows better coding efficiency, but on the other hand such additional dependencies between views must be accounted for in the case of transmission over packet lossy channels since errors may propagate across views.

MMSP'10, October 4-6, 2010, Saint-Malo, France.
978-1-4244-8112-5/10/\$26.00 ©*2010 IEEE.*

Moreover, measuring the quality of a stereoscopic video sequence is not as straightforward as in the case of monoscopic video. Many factors, in fact, should be taken into account in addition to the distortion introduced in the reconstructed frames at the decoder, such as, for instance, the quality of the depth perception and the fatigue caused by the stereoscopic video in the observer. Those factors are difficult to model and to incorporate in objective quality measures. Therefore, similarly to monoscopic video, one of the most commonly used objective quality estimation technique computes the quality of the stereoscopic video as a combination of the qualities of the two views. In particular, in the case of video containing artifacts, it has been proposed to average the quality of the left and right views [4], which is the approach employed in this work.

Once a suitable quality measure for stereoscopic video is identified, the optimization of stereoscopic video communications becomes possible. For instance, in the case of pre-encoded stereoscopic video sequences, a computationally-intensive analysis can be carried out to characterize the rate-distortion functions of the compressed video data in case of losses. The resulting data might then be used to optimize the transmission within a rate-distortion optimization framework such as the one proposed in [5]. The work in [6] indeed performs an offline parameter estimation to analytically model the rate-distortion function of the sequences to be transmitted, and a similar approach is followed to estimate the performance of the channel coder and the distortion due to losses. The work then uses such information to minimize the end-to-end distortion in a streaming scenario and to obtain optimal encoder and channel coding rates.

However, few efforts have been devoted to the case of low-delay stereoscopic video communication scenarios, which require low-complexity algorithms and do not allow pre-computation since frames are available only at encoding time. The scenario is important since it is commonly encountered in interactive communication applications such as videoconferencing, and it is expected that stereoscopy and depth perception in general could contribute towards the goals of the so called telepresence. For monoscopic video, the work in [7] proposes the use of a simple model to estimate the distortion that would be caused by the loss of the frame being encoded in future frames until the next resynchronization point, e.g., an I-type frame.

In this work we build on [7] and propose a distortion estimation model suitable for a stereoscopic low-delay video communication scenario that can estimate the expected distortion at the decoder as a function of a given transmission policy. We specifically address the case of a two view H.264/MVC encoded video which heavily relies on disparity compensation to improve compression. The proposed model can estimate the effect of the loss of the currently encoded frame on the future frames of both the left and right video sequences, using a low-complexity approach that makes use of information which is already available at the encoder. The performance of the proposed model is tested by simulations to assess its performance within a rate distortion framework aimed at transmission over a generic QoS-enabled network. For comparison purposes, both a traditional a priori approach relying on frame dependency information only and a prescient algorithm which has access to future frame information have also been considered.

The paper is organized as follows. Section II reviews the H.264/MVC encoding standard and its applicability for the low-delay stereoscopic video communication scenario. Then, Section III introduces the proposed distortion model. The simulation setup is described in Section IV, followed by the results in Section V. Conclusions are drawn in Section VI.

II. LOW-DELAY H.264/MVC STEREOSCOPIC CODING

The H.264/MVC extension [3] of the H.264/AVC standard adds efficient support for multiview video coding. It heavily relies on disparity compensation among different views to achieve a high coding efficiency, and it also supports complex prediction structures among different views. Each frame in a view can reference, in addition to past and future frames as in traditional encoders, also frames in one or two views that represent the scene at the same time instant from a different viewpoint. This approach allows to reuse multiple-prediction coding schemes already found in traditional encoders for, e.g., B-type pictures.

Even if the H.264/MVC extension allows complex prediction structures involving a high number of dependencies between frames, in the case of low-delay video communication only simpler structures can be used, as in traditional video encoding schemes using P-type frames only. In the case of stereoscopic video, only two views are involved. The first view, e.g., the left one, is coded independently, using an IPPP... coding pattern as in traditional schemes, while the second view can take advantage of the information already encoded in the first one. Thus, for each non-intra frame in the second view, motion compensation exploits the previous coded frame in the same view, while disparity compensation uses the already encoded corresponding frame in the first view. The dependency scheme is depicted in Figure 1.

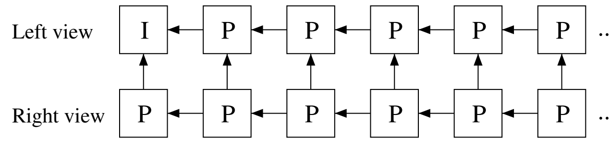

Fig. 1. Encoding dependencies for a stereoscopic low delay scenario.

This structure allows for low-delay coding and decoding, since there is no need to wait for future frames to be coded before processing the current left and right frames. However, the increased number of dependencies with respect to the monoscopic case implies that decoding errors in the first view propagate not only in the future frames in the same view, but also in the other view. On the contrary, in case losses happen in the second view, distortion is limited to future frames belonging to this view. Moreover, in this case, correctly decoded frames in the first view might help in quickly recovering from the loss. The amount of error propagation in future frames, however, depends on how much the encoder employed disparity compensation to code frames rather than motion compensation from previously encoded frames in the same view.

III. DISTORTION ESTIMATION MODEL FOR STEREOSCOPIC VIDEO

In order to optimize multimedia communications, a reliable technique to minimize the expected distortion at the decoder is needed. Such a technique should be able to consider the distortion introduced by the concealment, as well as the instantaneous characteristics of the compressed multimedia signals, e.g., the coding dependencies between frames due to motion and disparity compensation to estimate error propagation. In principle, it would be possible to compute an exact estimate of the expected distortion at the decoder by simulating each possible pattern of loss events, computing the distortion in the reconstructed multimedia signal at the decoder and weighting it by the probability that such a pattern occurs. However, this is infeasible since the number of patterns grows exponentially with the number of transmitted units, i.e., packets, and all patterns of loss events should be simulated because of the differential encoding between frames which is the main cause of error propagation. Moreover, in a low-delay transmission scenario, which is the main focus of this work, such a computation is impossible since future frames still have to be encoded when an estimate of the expected distortion value must be available.

In order to reduce the complexity of the problem, first we consider one frame at a time, and we estimate the total distortion d_i introduced by the loss of a single frame i by simulating the decoder behavior in case of loss. Frames belonging to different views are considered separately. A distortion is introduced in the concealed video frame, as well as in all frames which are dependent on either the concealed one, directly and indirectly. When an independently encoded frame, e.g., an I-type frame, is encountered, i.e., at the beginning of a new group of pictures (GOP), error propagation stops.

Therefore, it is possible to compute the exact distortion caused by the loss of a single frame in future frames of both views until the beginning of the next GOP. If we assume that distortion contributions when more than one frame is lost are additive, the number of decoding simulations needed to compute the distortion caused by the loss of each frame equals the number of frames. This assumption seems reasonable if the frame loss rate is limited, since it implicitly assumes that the data used for concealment are error-free. In the case of a simple frame copy concealment technique, this is equivalent to assume that the previous frame is error-free. Experimental studies and results will show that such an approximation is justified by the good performance of the proposed techniques.

Since an I-type frame stops error propagation, we consider one GOP at a time. Let the total distortion for a given loss pattern X composed by N events be

$$D = \sum_{j=1}^{N} x_j d_j \tag{1}$$

where $X = (x_1, ..., x_N)$ is a vector containing the outcome x_j of N transmission events, whose value is equal to zero if frame j is correctly received, and one otherwise. N is the number of frames in a given GOP. Eq. (1) can be restated for the case of stereoscopic video by separating the distortion for the left and right views:

$$D = \sum_{i=1}^{N} x_{L,i} \frac{1}{2} \left(d_{L,i}^{(L)} + d_{R,i}^{(L)} \right) + \sum_{i=1}^{N} x_{R,i} \frac{1}{2} \left(d_{L,i}^{(R)} + d_{R,i}^{(R)} \right) \tag{2}$$

where i is a given time instant in the GOP, subscripts indicate the view, i.e., left or right, to which the events and the distortion values refer, and superscripts indicate in which view (L) or (R) the loss was introduced to compute the distortion values. Clearly, since the left view does not depend on the right view, as shown in Figure 1, losses in the right view do not introduce distortion in the first view, i.e., $d_{L,i}^{(R)} = 0$.

In case of low delay scenarios, however, it is not possible to compute any of the values $d_{v,i}^{(l)}$ as they are defined above, since they include all the contributions to distortion propagation in future frames which are, of course, not yet available at the encoder. The only quantity that can be easily computed at the encoder at a given time instant are the distortions introduced in the left and right frames at time instant i, that we refer to as $\hat{d}_{v,i}^{(l)}$. Incidentally, this also reduces computational requirements with respect to the previous case which would require to decode all future frames until the end of the GOP.

Figure 2 shows an example of the distortion in the current frame and in each of the future frames when the first frame of a given GOP is lost. The trend shown in the figure is similar for all the video sequences and GOPs considered in this work. The chosen sequences are representative of videos with different characteristics, as detailed in Section IV. The error propagation over time if a loss happen in the main (left) view is similar for both views, whereas the distortion due to error propagation in case of loss in the right view rapidly decrease because disparity

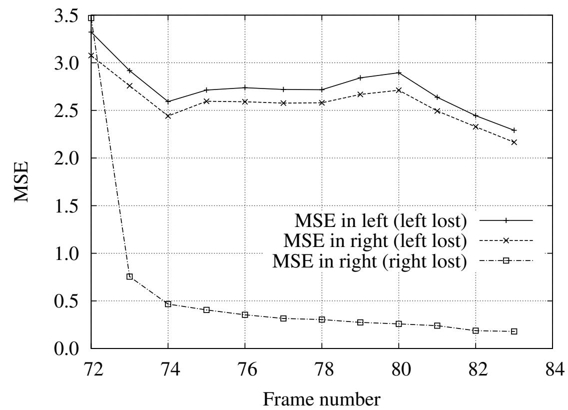

Fig. 2. Distortion (MSE) introduced in the left and right view of a GOP if the first left or right frame of the GOP is lost. *Interview* sequence.

compensation from correctly decoded frames in the left view helps in quickly reducing the distortion. Therefore, we propose to use the following model in order to estimate the distortion in future frames, thus being able to compute an estimate of $d_{v,i}^{(l)}$, referred to as $\tilde{d}_{v,i}^{(l)}$, on the basis of $\hat{d}_{v,i}^{(l)}$ values only, without waiting for future frames, making the technique suitable for very low delay scenarios. The $\tilde{d}_{v,i}^{(l)}$ values are computed as

$$\begin{cases} \tilde{d}_{L,i}^{(L)} &= \hat{d}_{L,i}^{(L)} + n_G f_L \hat{d}_{L,i}^{(L)} \\ \tilde{d}_{R,i}^{(L)} &= \hat{d}_{R,i}^{(L)} + n_G f_L \hat{d}_{R,i}^{(L)} \\ \tilde{d}_{L,i}^{(R)} &= 0 \\ \tilde{d}_{R,i}^{(R)} &= \hat{d}_{R,i}^{(R)} + n_G f_R \hat{d}_{R,i}^{(R)} \end{cases} \tag{3}$$

where n_G is the number of frames needed to reach, from the time instant $i+1$, the end of the GOP, and f_L and f_R are two tunable parameters of the model, which have been determined experimentally.

One hundred random frame loss patterns have been generated for different loss probabilities, ranging from 5% to 20%, and the distortion for each considered video sequence has been computed by actual decoding as well as by means of the proposed model, using different values for f_L and f_R. Experimentally, it has been found that the values $f_L = 1.0$ and $f_R = 0.15$ yield, on average, a good distortion estimate, close to the actual distortion. Note that the distortion has been computed, for each video sequence, as the mean squared error (MSE) with respect to the error-free decoded video, so that frames which are not affected by error propagation do not contribute to the total distortion.

The complexity of the proposed model is limited, since $\hat{d}_{v,i}^{(l)}$ values can be computed in $O(P)$, P being the number of pixels, using the decoded frames present in the decoder memory for motion and disparity compensation purposes, and the complexity implied by Eq. (3) is $O(M)$, M being the number of macroblocks.

At this point, Eq. (3) allows to formulate the classical multimedia quality optimization problem over a generic unreliable communication channel, since it can be used to compute an estimate of the expected distortion at the decoder. The problem can be formulated as choosing the transmission policy Π which solves the equation

$$\min_{\Pi} E[D(\Pi)] \tag{4}$$

subject to all the transmission constraints, and in particular channel constraints such as maximum rate. If a given transmission policy $\Pi = (\pi_1, ..., \pi_N)$ assigns a residual loss probability $p_j(\pi_j)$ to each frame j to be transmitted with transmission mode π_j, in the case of stereoscopic video transmission the problem can be formulated as

$$\min_{\Pi} \sum_{i=1}^{N} p_{L,i}(\pi_{L,i}) \frac{1}{2} \left(\tilde{d}_{L,i}^{(L)} + \tilde{d}_{R,i}^{(L)} \right) + \sum_{i=1}^{N} p_{R,i}(\pi_{R,i}) \frac{1}{2} \tilde{d}_{R,i}^{(R)} \tag{5}$$

where the subscripts and superscripts have the same meaning as in Eq. (2), N is the number of frames in one view of one GOP and $\Pi = (\pi_{L,1}, ..., \pi_{L,N}, \pi_{R,1}, ..., \pi_{R,N})$ is the transmission policy for each frame in the GOP.

IV. Simulation Setup

In order to test how the proposed distortion model performs in a practical case, we consider a simple QoS-enabled network which offers two service levels: perfect protection against losses and best-effort with packet loss probability p. This could model, for instance, a DiffServ network [8] which offers both a standard best-effort service and a fee-based loss-free service. For simplicity's sake we assume that each frame is put into one packet. Let $r_{L,i}$ and $r_{R,i}$ be the size of the left and right frames at time instant i, respectively, and R_{max} the maximum bandwidth that the transmitter is willing to send as loss-free. Let $E[D(\Pi)]$ and $R(\Pi)$ be defined as

$$\begin{aligned} E[D(\Pi)] &= \sum_{i=1}^{N} (1 - \pi_{L,i}) p \frac{1}{2} \left(\tilde{d}_{L,i}^{(L)} + \tilde{d}_{R,i}^{(L)} \right) \\ &+ \sum_{i=1}^{N} (1 - \pi_{R,i}) p \frac{1}{2} \tilde{d}_{R,i}^{(R)} \end{aligned} \tag{6}$$

$$R(\Pi) = \sum_{i=1}^{N} \pi_{L,i} r_{L,i} + \pi_{R,i} r_{R,i} \tag{7}$$

where each $\pi_{v,i}$ is either zero or one depending on the service (best effort or loss-free) used for transmitting the frame i belonging to view v. The transmission optimization problem can be formulated as

$$\begin{aligned} &\min_{\Pi} E[D(\Pi)] \\ \text{s. t.:} \quad &R(\Pi) < R_{max}. \end{aligned} \tag{8}$$

Eq. (8) can also be used if we are interested in minimizing the distortion with respect to the original uncompressed video sequence, since the total distortion can be approximated as the sum of the encoding distortion and the distortion due to channel losses [6]. The results section will show that such an approximation is reasonable and it yields good results.

R_{max} can be expressed either as an absolute bandwidth value or as a fraction s of the total video bandwidth, as done in this work. In particular, for each GOP, we compute $R_{max} = s \cdot R_{estGOP}$, where R_{estGOP} is an estimate of the total size R_{GOP} of the current GOP, which is not known by the encoder in a low-delay scenario until the GOP has been fully coded and transmitted. For this reason, we use an estimate of the GOP size based on the previous size of the GOP, i.e., $R_{estGOP} = R_{prevGOP}$.

This work considers three different strategies to assign packets to the two network services. First, as a comparison term, a random assignment is performed, while trying to fulfill the constraint on the maximum fraction s of bandwidth to assign to the loss-free service. Since in a low-delay stereoscopic video transmission scenario a decision must be taken for the left and right frames every time a new couple of frames is produced by the encoder, frames are sent using the loss-free service with uniform probability equal to s.

A second technique implements an "a priori" approach which considers frame dependency information. The first couple of frames (both left and right) of each GOP are sent using the loss-free service until the number of sent bytes does not exceed the maximum bandwidth constraint $s \cdot R_{estGOP}$, where R_{estGOP} is estimated as previously described. The underlying idea is that frames, such as the first ones in each GOP, that might propagate errors to a large number of frames must be protected better than the others in the GOP. Since both techniques do not guarantee to send exactly $s \cdot R_{GOP}$ bytes using the loss-free service, because the exact size R_{GOP} of the current GOP is not known while transmitting it, the remaining (or excess) bytes are added (or subtracted) from the bandwidth constraint for the loss-free service in the next GOP. In this way the fraction of bytes sent using the loss-free service tends to s despite the potentially unreliable estimate of each GOP size.

A third technique exploits the distortion information computed as described in Section III by means of a rate-distortion optimization framework. For each GOP, the constrained minimization problem in Eq. (8) can be recasted into an unconstrained one using a Lagrangian multiplier λ:

$$\min_{\Pi} D(\Pi) + \lambda R(\Pi). \tag{9}$$

Finding the λ value which satisfied the constraint in Eq. (8) would require the knowledge of the rate and distortion information for all the frames in a GOP, which is not possible in the low-delay scenario considered in this work. Therefore, an alternative approach is used. A value of λ is chosen for each GOP, then, given that λ value, the minimization problem can be solved for each left and right frame at time i independently from the other frames thus determining each $\pi_{v,i}$ independently in $O(1)$. Hence, each frame is sent using either the loss-free or the best-effort service immediately after encoding, depending on the resulting $\pi_{v,i}$ value. At the end of the GOP, the total amount of loss-free bandwidth is computed and compared with the fraction s of the total GOP size. The λ value is then adjusted before transmitting next GOP to increase or decrease the amount of loss-free bandwidth depending on the previous over- or under-utilization of the fraction s of loss-free bandwidth. The approach is similar to the one

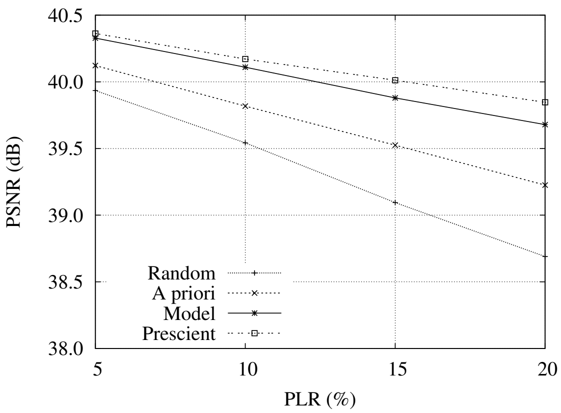

Fig. 3. PSNR as a function of the PLR for the *interview* video sequence, 70% sent as loss-free service.

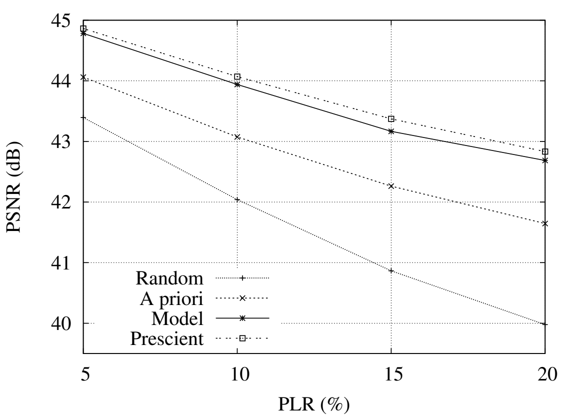

Fig. 4. PSNR as a function of the PLR for the *elephantsdream* video sequence, 70% sent as loss-free service.

used by rate-distortion optimized rate control algorithms in video encoders, where only the λ value is adjusted in order to meet the rate constraints without explicitly enforcing an absolute rate constraint. The λ update rule is the same as the one employed in [9]. Thus, the total complexity of the rate-distortion optimized transmission system is $O(P)$, P being the number of pixels.

Various test sequences, different in content type and resolution, are used in this work. The *ballet* sequence is obtained by the homonymous Microsoft multiview sequence [10] by using the data captured by camera #4 and #5 as the left and right views. The *elephantsdream* sequence is obtained by selecting a segment from the freely available homonymous computer-animated movie [11]. In this case, the stereoscopic sequence was generated by rendering the selected scene from two different point of views, slightly left and right of the original camera position. Finally, the *interview* sequence is obtained by applying a depth-image based rendering (DIBR) technique [12] to sequence [13], that only provides video plus depth-range information. The DIBR technique is able to compute a left and right view by shifting the pixels of the original monocular video of an amount proportional to the depth of each pixel. These video sequences have been chosen in order to represent different types of stereoscopic content originating from widely different techniques.

TABLE I
CHARACTERISTICS OF THE VIDEO SEQUENCES.

Sequence	Frames	Bitrate (kbit/s)	PSNR (dB)
Interview (704×576)	241	1033 (left)	40.50 (left)
		572 (right)	40.49 (right)
		1605 (total)	40.49 (avg.)
Elephantsdream (960×512)	89	659 (left)	45.39 (left)
		334 (right)	45.41 (right)
		993 (total)	45.40 (avg.)
Ballet (512×384)	97	374 (left)	41.53 (left)
		259 (right)	41.05 (right)
		633 (total)	41.29 (avg.)

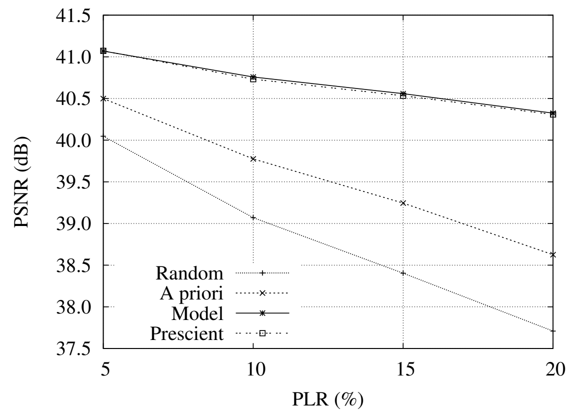

Fig. 5. PSNR as a function of the PLR for the *ballet* video sequence, 70% sent as loss-free service.

Video has been encoded using the H.264/MVC test model software v. 8.0 [14], with the frame dependency scheme shown in Figure 1. The quantization parameter (QP) has been kept constant equal to 26 for all the frames. Every twelve frames in the main (left) view a new I-type frame is inserted. Table I summarizes the characteristics of the video sequences including bitrates and encoding PSNR. Different fractions s of loss-free bandwidth as well as frame loss rates have been tested. For each condition, values are averaged over 30 channel realizations.

V. RESULTS

Figure 3, 4 and 5 show the performance of the random, a priori and proposed model-based rate-distortion optimized techniques as a function of the packet loss rate (PLR) when the fraction s of loss-free bandwidth is equal to 0.70. The rate-distortion optimized technique shows gains ranging from 0.2 up to about 1.7 dB PSNR with respect to the a priori technique, depending on the considered video sequence and the packet loss rate. The gain is even more pronounced, as expected, with respect to the random technique. The gain increases if the packet loss rate increases, which confirms that the distortion

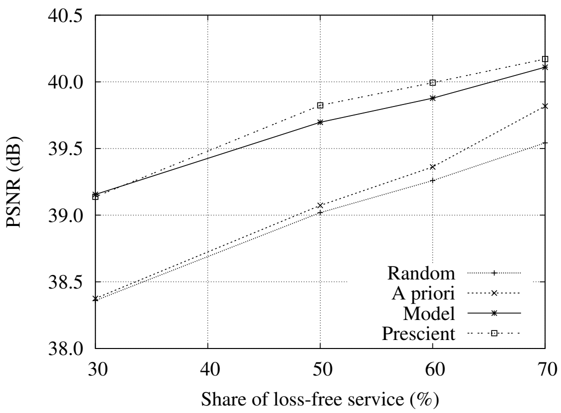

Fig. 6. PSNR as a function of the share of loss-free service for the *interview* video sequence, 10% PLR.

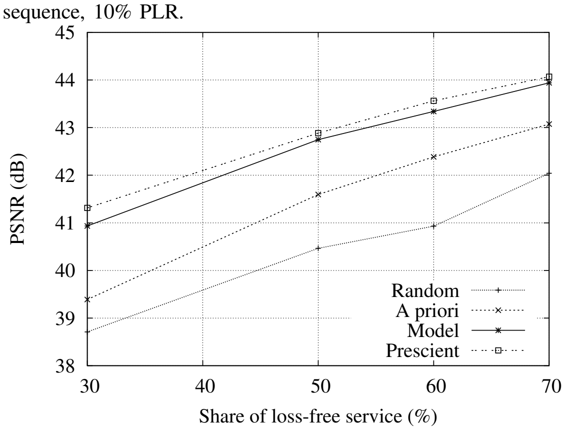

Fig. 7. PSNR as a function of the share of loss-free service for the *elephantsdream* video sequence, 10% PLR.

additivity assumption do not negatively affect the performance, at least up to 20% PLR.

Moreover, those figures also show the performance that can be achieved by fully decoding the whole GOP to estimate the distortion in future frames, i.e., using the $d_{v,i}^{(l)}$ values instead of the $\tilde{d}_{v,i}^{(l)}$ ones in the rate-distortion optimized technique. The technique, named "prescient" due to the need for future information to perform distortion computation, is useful for establishing an upper bound on the performance of the proposed model-based technique. The figures show that the proposed model-based rate-distortion optimized technique is always very close to this upper bound.

Figure 6 and 7 show the performance of the three techniques as well as the upper bound as a function of the fraction of loss-free bandwidth when the packet loss rate is 10%. The rate-distortion optimized technique shows gains ranging from 0.3 to about 1.6 dB PSNR with respect to the a priori technique. The gain, as expected, increases as the share of loss-free service decreases, since in this case the possibility of taking into account the distortion that is potentially introduced in future frames due to the loss of each frame is a great advantage.

VI. Conclusion

In this work a new low-complexity distortion estimation model for stereoscopic video suitable for low-delay video communication scenarios has been proposed. The distortion introduced by the loss of a given frame has been investigated and a model has been designed in order to accurately estimate the impact that the loss of each frame would have on future frames in both the left and right views. The model only relies on information that can be easily computed by means of encoder side information. Simulations results using the model in a rate-distortion optimized framework for video communications over a generic QoS-enabled network show consistent performance gains with respect to a traditional prioritization technique based on frame dependency information only, as well as a very small gap from the upper bound, i.e., the performance of a prescient technique that has perfect knowledge of the distortion characteristics of future frames. Future work will be devoted to extend the model to the multiview case.

Acknowledgment

The author wishes to thank Andrea Marquet for providing the rendering of the stereoscopic version of the part of the *elephantsdream* video sequence used in this work.

References

[1] A. Puri, R.V. Kollarits, and B.G. Haskell, "Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4," *Signal Process. Image Comm.*, vol. 10, pp. 201–234, Jul. 1997.

[2] ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services," *ITU-T*, Mar. 2009.

[3] Y. Chen, Y.-K. Wang, K. Ugur, M.M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," EURASIP Journal on Advances in Signal Processing, 2009, Article ID 786015, 13 pages.

[4] H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," in *Proc. of 3DTV Conference*, Potsdam, Germany, May 2009.

[5] P.A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390–404, Apr. 2006.

[6] A. Serdar Tan, A. Aksay, G.B. Akar, and E. Arikan, "Rate-distortion optimization for stereoscopic video streaming with unequal error protection," EURASIP Journal on Advances in Signal Processing, 2009, Article ID 632545, 14 pages.

[7] F. De Vito, D. Quaglia, and J.C. De Martin, "Model-based distortion estimation for perceptual classification of video packets," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, Siena, Italy, Sep. 2004, pp. 79–82.

[8] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," RFC 2475, Dec. 1998.

[9] R. Zhang, S.L. Regunatha, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communic.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[10] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. on Graphics*, pp. 600–608, Aug. 2004.

[11] Orange Open Movie Project studio. (2006) "Elephantsdream" video sequence, time range 2:54-2:58, scene 03_19. [Online]. Available: http://www.elephantsdream.org

[12] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. of SPIE*, San Jose, CA, Jun. 2004.

[13] "Interview" video sequence. [Online]. Available: ftp://ftp.tnt.uni-hannover.de/pub/3dav/3DAV_Demos/FHG_HHI

[14] (2010, Mar.) JMVC reference software v. 8.0. [Online]. Available: jvtuser@garcon.ient.rwth-aachen.de:/cvs/jvt co -r JMVC_8_0 jmvc