

ANALYSIS OF ERROR PROPAGATION DUE TO FRAME LOSSES IN A DISTRIBUTED VIDEO CODING SYSTEM

Thomas Maugey¹, Thomas André¹, Béatrice Pesquet-Popescu¹, and Joumana Farah²

¹ TELECOM ParisTech, CNRS LTCI, Signal and Image Processing Department
46 rue Barrault, 75634 Paris Cedex 13, France
Email: {maugey, thandre, pesquet}@telecom-paristech.fr

² Faculty of Sciences and Computer Engineering, Holy-Spirit University of Kaslik
Jounieh, Lebanon
joumanafarah@usek.edu.lb

ABSTRACT

In this paper, we propose a theoretical model for the error propagation phenomenon generated by a frame loss in a distributed video coding framework. Using rate-distortion functions, we analyze the impact of a frame loss on the average distortion of a group of pictures depending on the position of the lost frame within the GOP, as well as the level of motion in each frame and the quantization errors in the key frames and the Wyner-Ziv frames. This theoretical analysis is further confirmed by a practical implementation of the DVC framework using different configurations of frame losses.

1. INTRODUCTION

Distributed Video Coding (DVC) is a promising paradigm in video coding that allows moving the computation burden from the encoder to the decoder by performing intraframe coding at the encoder and inter-frame decoding at the decoder. Many applications can benefit from it, such as video compression on mobile devices, multi-sensor surveillance systems, etc. DVC is based on two fundamental results of information theory stated by Slepian and Wolf for lossless coding [1] and extended by Wyner and Ziv for lossy compression [2]. Theoretically, the performances of DVC can attain the ones of classical predictive codecs (H.263, H.264,...). In reality, the performances are still below, even though DVC performs better than intra coding in most cases.

In DVC, the frames are separated in two subsets, key frames (KFs) and Wyner-Ziv frames (WZFs), thus yielding two correlated sources which are then separately encoded and jointly decoded. In general, the group of pictures (GOP) has the following structure [3]: one KF followed by n WZFs. Very often, n is fixed, but some studies have proposed to vary n in accordance with the sequence content [4], in such a way to optimize the system performances. In this paper, the adopted DVC scheme is similar to the one in [3]: KFs are encoded and decoded using a classical intra codec, such as H.264 intra. However, the processing of WZFs is different. First, a transform is applied (mainly an integer-to-integer 4×4 DCT), since the performances are better in the transform domain than in the pixel domain [3, 5]. After quantization, the WZFs are channel-encoded using a powerful channel code, such as turbo-codes [6] in the adopted scheme, or LDPC codes in other works [7]. At the decoder, the KFs are used to build the side information needed in order to turbo-decode and reconstruct the WZFs. In such an inter-frame decoding process, the performances are greatly dependant on the side

information quality. Moreover, since this side information is constructed using previously decoded frames, it becomes important to study the behavior of DVC in case of frame loss. A video coder behavior can in general be modeled using the rate distortion function, which directly shows the codec performances. In [8], a rate distortion analysis is used to estimate the motion interpolation error using a Kalman filter framework, in order to optimize the GOP size. In this paper, we experimentally estimate the motion estimation errors and propose an analytical model for error propagation inside a GOP, by taking into account the quantization errors, the motion interpolation errors, as well as the position of the lost frame. Based on the model introduced in Section 2, this paper theoretically studies frame losses in a DVC system in Section 3. In Section 4, experimental results are presented to verify the theoretical study. Finally, conclusions and future work are drawn in Section 5.

2. PROPOSED MODEL

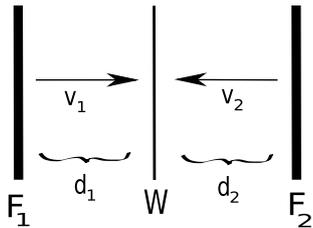
Before calculating the rate-distortion functions of interest, a model is proposed. The hypotheses and the expressions obtained are described in the following. In a DVC decoding scheme, a WZF is decoded thanks to the side information computed based, in general, on two reference frames. The first purpose of this study is to find an expression of the variance of the estimation error of this WZF. Let us denote by F_1 (resp. F_2) the first (resp. second) reference frame at a distance of d_1 (resp. d_2) from W (as illustrated in Fig.1(a)). \tilde{F}_1 (resp. \tilde{F}_2) is the quantized version of F_1 (resp. F_2), and \bar{F}_1 (resp. \bar{F}_2) is the quantized motion compensated reference frame (Fig.1(b)). We assume that the side information for W is a linear combination of the reconstructed reference frames:

$$\bar{F}_{SI} = k_1 \bar{F}_1 + k_2 \bar{F}_2, \text{ where } k_1 = \frac{d_2}{d_1 + d_2}, k_2 = \frac{d_1}{d_1 + d_2}. \quad (1)$$

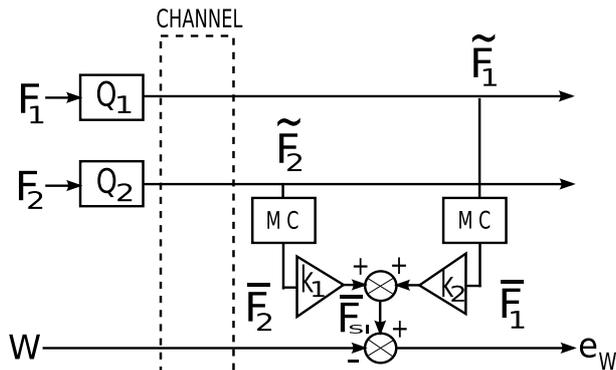
Let e_W be the error between the original frame W and its side information \bar{F}_{SI} : $e_W = W - \bar{F}_{SI}$. We verified through simulations that e_W has zero mean.

We denote by $\mathbf{p} = (m, n)$ the vector corresponding to the pixel in line n and column m . We notice that $\bar{F}_1(\mathbf{p}) = \tilde{F}_1(\mathbf{p} - \mathbf{v}_1)$ and $\bar{F}_2(\mathbf{p}) = \tilde{F}_2(\mathbf{p} - \mathbf{v}_2)$.

Let us determine the expression of $\sigma_{e_W}^2$, the variance of the



(a) Motion compensated interpolation of the WZ frame. d_1 and d_2 are the distances between the WZF and the reference frames, while \mathbf{v}_1 and \mathbf{v}_2 are the motion vectors extrapolated from \mathbf{v} .



(b) Side information generation scheme.

Figure 1: F_1 and F_2 are used to generate the side information used to estimate and decode W .

estimation error of the frame W :

$$\begin{aligned} \sigma_{eW}^2 &= E \left\{ (W(\mathbf{p}) - \bar{F}_{SI}(\mathbf{p}))^2 \right\} \\ &= E \left\{ (W(\mathbf{p}) - k_1 \bar{F}_1(\mathbf{p}) - k_2 \bar{F}_2(\mathbf{p}))^2 \right\} \\ &= E \left\{ (W(\mathbf{p}) - k_1 \tilde{F}_1(\mathbf{p} - \mathbf{v}_1) - k_2 \tilde{F}_2(\mathbf{p} - \mathbf{v}_2))^2 \right\} \\ &= E \left\{ (W(\mathbf{p}) - k_1 F_1(\mathbf{p} - \mathbf{v}_1) - k_2 F_2(\mathbf{p} - \mathbf{v}_2) \right. \\ &\quad \left. + k_1 (F_1(\mathbf{p} - \mathbf{v}_1) - \tilde{F}_1(\mathbf{p} - \mathbf{v}_1)) \right. \\ &\quad \left. + k_2 (F_2(\mathbf{p} - \mathbf{v}_2) - \tilde{F}_2(\mathbf{p} - \mathbf{v}_2)))^2 \right\}. \end{aligned} \quad (2)$$

Line (2) corresponds to the estimation error with respect to the original reference frames. Lines (3) and (4) correspond to the quantization errors of the reference frames. We assume that these three errors are independent. This is a reasonable assumption at high bitrates, since the first one only depends on the sequence characteristics, while the last two are given by the quantification scheme. Therefore, we can write:

$$\begin{aligned} \sigma_{eW}^2 &= E \left\{ (W(\mathbf{p}) - k_1 F_1(\mathbf{p} - \mathbf{v}_1) - k_2 F_2(\mathbf{p} - \mathbf{v}_2))^2 \right\} \\ &\quad + k_1^2 E \left\{ (F_1(\mathbf{p} - \mathbf{v}_1) - \tilde{F}_1(\mathbf{p} - \mathbf{v}_1))^2 \right\} \\ &\quad + k_2^2 E \left\{ (F_2(\mathbf{p} - \mathbf{v}_2) - \tilde{F}_2(\mathbf{p} - \mathbf{v}_2))^2 \right\}. \end{aligned}$$

Assuming that the motion vectors \mathbf{v}_i are the same when estimated with the quantized or non-quantized reference

frames, we denote by M_{d_1, d_2} the estimation error, when the motion estimation is performed using the original frames. It depends on the two distances d_1 and d_2 . We also denote by D_{F_1} and D_{F_2} the distortions due to the quantization of F_1 and F_2 . Therefore, the expression of σ_{eW}^2 reads:

$$\sigma_{eW}^2 = M_{d_1, d_2} + k_1^2 D_{F_1} + k_2^2 D_{F_2}. \quad (5)$$

We finally obtain a simple expression of the estimation error variance of W in which the distortions of the reference frames are separated from the estimation error using original reference frames. This will simplify the forthcoming calculations and allows a simple recursive analysis of the error propagation in a GOP.

3. THEORETICAL ANALYSIS OF FRAME LOSSES

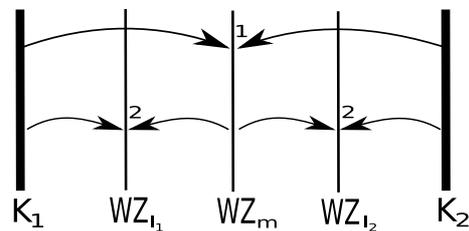


Figure 2: GOP structure and decoding strategy. (1) and (2) indicate the decoding operations order.

In this section, we propose to theoretically study the outcomes of a frame loss in a distributed video coding system. In order to simplify the study, the GOP size is fixed. Since a GOP size of two frames does not present intermediate WZFs in the decoding process, it is of no particular interest. Therefore, the adopted GOP size is 4, i.e. one KF followed by three WZFs, as shown in Fig 2. The decoding process for a GOP of four frames is as follows. First the middle WZ frame, denoted by WZ_m , is decoded thanks to a side information generated using the two KFs, K_1 and K_2 . Then, the lateral frames WZ_{l_1} and WZ_{l_2} are decoded using the reference frames and the decoded frame WZ_m . In [9], we prove that this decoding strategy is optimal between all possible schemes. It has also been empirically used in [6]. We notice that the three kinds of frames play a different role in this decoding process. The expression of the average distortion in a GOP is: $D_T = \frac{1}{4}(D_K + D_{WZ_{l_1}} + D_{WZ_m} + D_{WZ_{l_2}})$. We recall that the general rate distortion function for a frame X can be approximated, at high bitrate, by

$$D_X = \alpha \sigma_X^2 2^{-2R_X}, \quad (6)$$

where R_X is the allocated rate in bits per pixel, σ_X^2 the original variance of the frame X , and α a constant depending on the source distribution.

Case of a lossless transmission: In this case, the distortion of the KF is simply due to quantization:

$$D_K = \alpha_K \sigma_K^2 2^{-2R_K}. \quad (7)$$

As for the frame WZ_m , since $k_1 = k_2 = \frac{1}{2}$, its distortion can

be written using (5):

$$\begin{aligned} D_{WZ_m} &= \alpha_{WZ_m} \sigma_{e_{WZ_m}}^2 2^{-2R_{WZ_m}} \\ &= \alpha_{WZ_m} \left(M_{2,2} + \frac{1}{2} D_K \right) 2^{-2R_{WZ_m}}. \end{aligned} \quad (8)$$

Similarly, the distortion of the frames WZ_{l_i} , for $i \in \{1, 2\}$, is:

$$\begin{aligned} D_{WZ_{l_i}} &= \alpha_{WZ_{l_i}} \sigma_{e_{WZ_{l_i}}}^2 2^{-2R_{WZ_{l_i}}} \\ &= \alpha_{WZ_{l_i}} \left(M_{1,1} + \frac{1}{4} D_K + \frac{1}{4} D_{WZ_m} \right) 2^{-2R_{WZ_{l_i}}}. \end{aligned} \quad (9)$$

We obtain the average distortion of a GOP using:

$$D_T = \frac{1}{4} (D_K + D_{WZ_m} + 2D_{WZ_{l_i}}). \quad (10)$$

Loss of parity bits for WZ_{l_1} : in the case the parity bits used to decode the frame WZ_{l_1} are lost, its estimation error can not be corrected. Thus, we have $R_{WZ_{l_1}} = 0$. The distortion of the lost frame WZ_{l_1} is then:

$$D_{WZ_{l_1}}^* = \alpha_{WZ_{l_1}} \left(M_{1,1} + \frac{1}{4} D_K + \frac{1}{4} D_{WZ_m} \right). \quad (11)$$

The distortion of the KF, as well as WZ_m and WZ_{l_2} , remain unchanged and are expressed as in (7), (8) and (9).

The average GOP distortion becomes:

$$D_T^{WZ_{l_1}} = \frac{1}{4} (D_K + D_{WZ_m} + D_{WZ_{l_1}}^* + D_{WZ_{l_2}}). \quad (12)$$

Loss of parity bits for WZ_m : in this case, the distortion of the KF is as in (7), and the distortion of the WZ_m frame is:

$$D_{WZ_m}^* = \alpha_{WZ_m} \left(M_{2,2} + \frac{1}{2} D_K \right), \text{ since } R_{WZ_m} = 0. \quad (13)$$

Therefore, the distortion of the WZ_{l_i} frames, for $i \in \{1, 2\}$, becomes:

$$D_{WZ_{l_i}} = \alpha_{WZ_{l_i}} \left(M_{1,1} + \frac{1}{4} D_K + \frac{1}{4} D_{WZ_m}^* \right) 2^{-2R_{WZ_{l_i}}}. \quad (14)$$

We have the following average distortion of a GOP:

$$D_T^{WZ_m} = \frac{1}{4} (D_K + D_{WZ_m}^* + 2D_{WZ_{l_i}}). \quad (15)$$

Loss of a Key Frame: When K_1 is lost (and similarly for K_2), before decoding the corresponding GOP, this frame needs to be estimated using other KFs supposed to be well received (the two located at a distance of 4 frames before and after the current lost KF). Therefore, the distortion of the KF is:

$$\begin{aligned} D_K^* &= \alpha_K^* \sigma_{e_K}^2 2^{-2R_K}, \text{ with } R_K = 0 \\ D_K^* &= \alpha_K^* \left(M_{4,4} + \frac{1}{2} D_K \right). \end{aligned} \quad (16)$$

Thus, the distortion of the WZ_m frame will be:

$$D_{WZ_m} = \alpha_{WZ_m} \left(M_{2,2} + \frac{1}{4} D_K^* + \frac{1}{4} D_K \right) 2^{-2R_{WZ_m}}. \quad (17)$$

and the distortion of the WZ_{l_1} and WZ_{l_2} frames modify accordingly:

$$D_{WZ_{l_1}} = \alpha_{WZ_{l_1}} \left(M_{1,1} + \frac{1}{4} D_K^* + \frac{1}{4} D_{WZ_m} \right) 2^{-2R_{WZ_{l_1}}} \quad (18)$$

$$D_{WZ_{l_2}} = \alpha_{WZ_{l_2}} \left(M_{1,1} + \frac{1}{4} D_{WZ_m} + \frac{1}{4} D_K \right) 2^{-2R_{WZ_{l_2}}}. \quad (19)$$

We have the following average GOP distortion in this case:

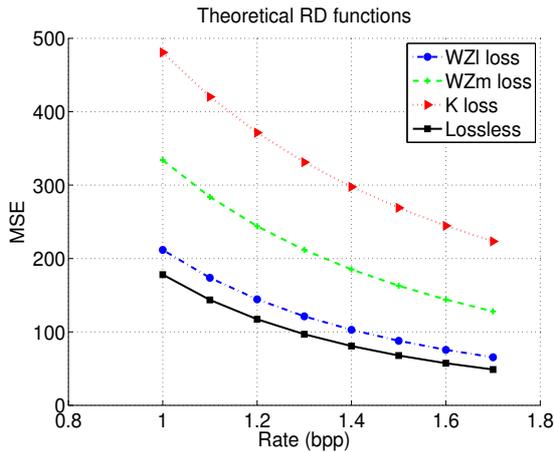
$$D_T^K = \frac{1}{4} (D_K^* + D_{WZ_m} + D_{WZ_{l_1}} + D_{WZ_{l_2}}). \quad (20)$$

As explained in Sec.1, in this paper, the motion interpolation errors ($M_{1,1}$, $M_{2,2}$, $M_{4,4}$) are experimentally estimated. These errors, as well as σ_K^2 , have been estimated with the test sequences ‘‘Foreman’’ (QCIF, 30 fps, 200 frames) and ‘‘Coastguard’’ (QCIF, 30 fps, 150 frames). The estimation of α_i coefficients is based on a detailed rate distortion analysis presented in [10]. For the KFs, the hypotheses are: Gaussian distribution, high bitrate, while for WZFs we considered a Laplacian distribution. Note also that, in this reference, we deduce rate-distortion models for theoretical sources and low bitrates. However, these are less practical to exploit, so here we keep with the classical high bitrate rate-distortion model. Moreover, we experimentally established that the rates for the four frames must be different in order to have a uniform decoding quality in a GOP: if we consider a rate R in bpp for the KF, the rate for the WZ_m frame is taken $R/2$ and for the WZ_{l_i} as $R/4$. These ratios were adopted for the theoretical plots (Fig.3) where we present the average rate in bpp.

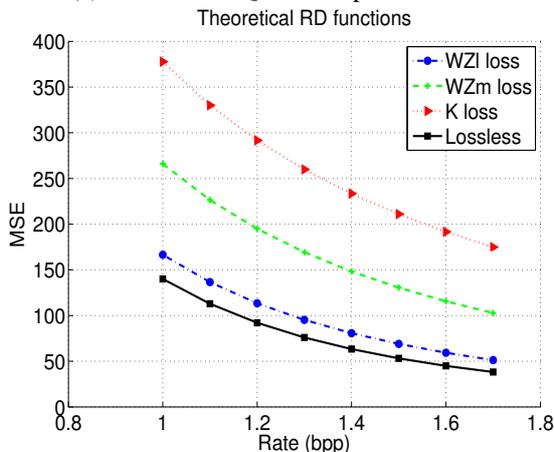
Before commenting these theoretical plots, it is important to recall that the previous calculation was conducted under the assumption of high bitrate coding. First, in (2), we assumed that the motion vectors, \mathbf{v}_1 and \mathbf{v}_2 , are the same when estimated from quantized reference frames or from original reference frames. This hypothesis is verified only at high bitrate. Then, in order to obtain the expression in (5), it was assumed that the estimation error and the quantization errors were independent, which is again verified at high bitrate. Finally, we used in (6) an approximation for the rate distortion function which only holds at high bitrate. For all these reasons, the values of the theoretical rate distortion function are bigger than expected and we only present the curves at high bitrate (above 1 bpp). However, these plots still allow an interesting comparison between the different curves. In Fig.3, we notice that the plots clearly highlight the importance of the error propagation phenomenon. Indeed, for both video sequences, the loss of a KF propagates over the entire GOP and leads to a much higher distortion than in the case of a WZ_m loss, which in turn induces a more important distortion than that caused by a WZ_{l_i} frame loss. These theoretical results thus illustrate the fact that an error occurred in a K or WZ_m frame will spread over the other frames when using that biased frame as a reference frame.

4. EXPERIMENTAL RESULTS

In this section, we propose a comparative analysis of the experimental rate-distortion functions in the same frame loss conditions as those considered in the theoretical study in Section 3, in order to compare them to the theoretical results presented in Fig 3. For this purpose, a Wyner-Ziv video



(a) "Foreman", QCIF, 30 fps, 200 frames.

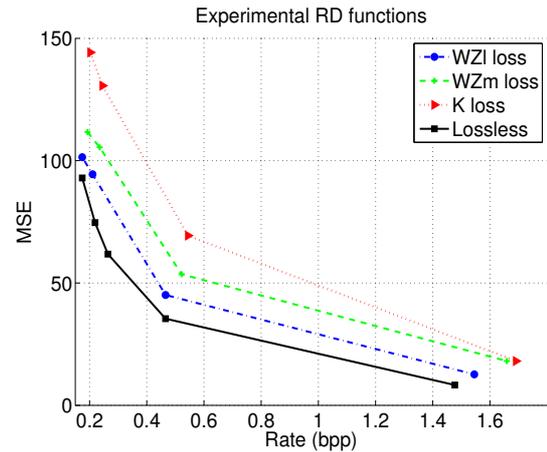


(b) "Coastguard", QCIF, 30 fps, 150 frames.

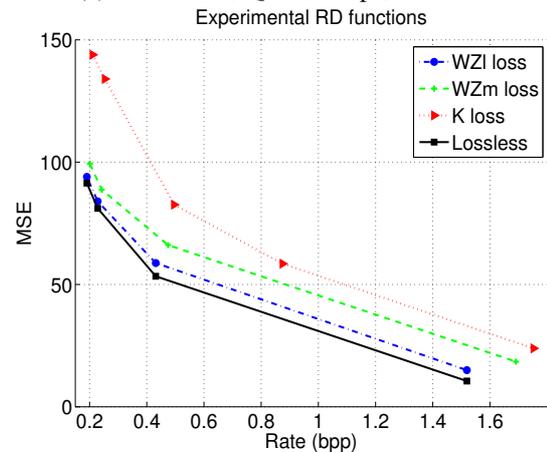
Figure 3: Theoretical rate-distortion functions, corresponding to the lossless case and to the three loss situations, for (a) "Foreman" and (b) "Coastguard" sequences.

codec was built following the general coding scheme [11] presented in Section 1. For side information generation, a motion-based interpolation proposed by Pereira in [12] was used. Experiments were run on the same test video sequences "Foreman" and "Coastguard". The results presented in Fig 4 correspond, at each bitrate, to the average distortion of the entire sequence. For each loss type, every GOP in the sequence is affected by the loss (e.g., for a WZ_m or WZ_l loss, one over four frames in the sequence are lost). If the lost frame is a WZF, its parity bits are transmitted but cannot be exploited by the decoder. If the lost frame is a KF, the frame is estimated at the decoder using the two closest KF.

Two main remarks can be done regarding these experimental plots. First, we are able to see in the obtained curves the error propagation caused by a frame loss. Indeed, the experiments show that if a frame is used to generate the side information for other WZFs, its loss will deeply affect the decoding performances. The second remark concerns the similarity between the theoretical and experimental plots. Indeed, the theoretical plots have predicted the relative importance of the frame losses (K, WZ_m, WZ_l) at high bitrate. One can see in the experiments that this prediction is also true at low bitrate. The proposed theoretical approach can thus be used in simi-



(a) "Foreman", QCIF, 30 fps, 200 frames.



(b) "Coastguard", QCIF, 30 fps, 150 frames.

Figure 4: Experimental rate-distortion functions, corresponding to the lossless case and to the three loss situations, for (a) "Foreman" and (b) "Coastguard" sequences.

lar situations in order to improve the decoding performances.

Moreover, we present another experimental result which analyses the evolution of the decoder behavior through time and compares the case of lossless transmission to the case where the transmission is randomly affected by frame losses (Fig.5). In such a decoding scheme, it is interesting to study the side information evolution linked with the rate per frame evolution. Indeed, the final PSNR of each frame is almost equal for a lossy or a lossless transmission, since the rate for a WZF will increase in order to correct the errors using the parity bits. Then, if the estimation error is bigger, the requested parity bits will be more numerous, but the decoded frame will have almost the same PSNR. In Fig.5 (up), we present the evolution of the side information quality. Since the notion of side information does not exist for the KFs, we represent the PSNR of the decoded KFs. In Fig.5 (bottom), the evolution of the transmitted rate per frame is presented. The experiments were run on the "Coastguard" (QCIF, 30 fps) sequence with the first 97 frames. For the lossy transmission (plain plots), the frame losses occurred randomly, i.e. in a GOP, the lost frame could be K, WZ_m, WZ_l or none of them. The vertical lines represent the moments when the

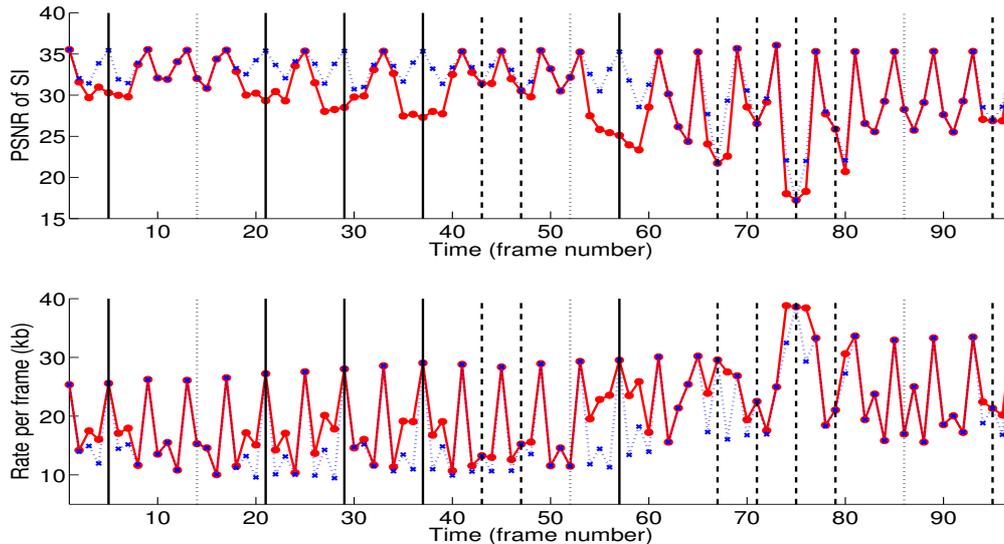


Figure 5: Evolution of the side information PSNR (up) and of the rate per frame (bottom) through time. The dotted curves correspond to a lossless transmission and the plain curves correspond to the case where the transmission is randomly affected by frame losses. The KF losses (resp. WZ_m and WZ_l) are represented by vertical plain lines (resp. dashed and dotted lines).

losses occurred (plain lines for K losses, dashed lines for WZ_m losses, and dotted lines for WZ_l losses). One can notice that the rates for KFs and WZFs do not exactly correspond to the ratios indicated in Sec. 3. Indeed, they have been established from statistics taking into account a larger number of frames. The obtained curves confirm the previous remarks on the relative importance of the frame losses (K , WZ_m , WZ_l). Indeed, we can see that a K loss affects the 6 other frames around it, i.e. their SI PSNR is lower and their rate per frame is bigger. Besides, the loss in SI PSNR and the increase in the requested data rate are more important for the closest neighbors than the rest of the GOP. This proves that the error propagation influence due to frame loss decays with time (in both directions). On the other side, a WZ_m loss affects only two frames around it, whereas a WZ_l loss does not affect any other frame. In fact, a WZ_l loss is not visible on the presented curves because only the reconstruction is affected in this case and it does not concern the transmission rate or the SI PSNR.

5. CONCLUSION

In this paper, we provided an analytical model for the rate-distortion behavior of frame inter-dependencies in a DVC framework. This allowed us to study the impact of a frame loss, depending on its role in the GOP (KF , WZ_l , WZ_m). Experimental results confirmed this analysis on various video sequences and error loss settings. Moreover, the methodology is interesting in itself because it proposes a recursive approach which permits a relatively simple rate-distortion analysis. In future work, we will look forward how to extend this analysis to low bitrates.

REFERENCES

- [1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. Vol. 19, pp. pp. 471–480, July 1973.
- [2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. on Information Theory*, vol. Vol. 22, pp. pp. 1–11, Jan. 1976.
- [3] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE*, vol. Vol. 93, no. 71, pp. pp. 71 – 83, Jan. 2005.
- [4] J. Ascenso, C. Brites, and F. Pereira, "Content adaptive Wyner-Ziv video coding driven by motion activity," in *IEEE ICIP*, Atlanta, GA, USA, July 2006, pp. 605–608.
- [5] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," *SPIE Visual Communications and Image Processing Conference*, vol. Vol. 5308, pp. pp. 520–528, Jan. 2004, San Jose, CA.
- [6] A. Aaron, E. Setton, and B. Girod, "Towards practical Wyner-Ziv coding of video," *Image Processing, IEEE ICIP*, vol. 3, pp. 869–72, 14–17 Sept. 2003, Barcelona, Spain.
- [7] Q. Xu and Z. Xiong, "Layered Wyner-Ziv video coding," *IEEE Trans. Image Processing*, vol. Vol. 15, no. 12, pp. pp. 3791–803, Jan. 2006.
- [8] M. Tagliasacchi, L. Frigerio, and S. Tubaro, "Rate-distortion analysis of motion-compensated interpolation at the decoder in distributed video coding," *IEEE Signal Processing Letters*, vol. Vol. 14, pp. pp. 625–628, Sept. 2007.
- [9] T. Maugey and B. Pesquet-Popescu, "Side information estimation and new symmetric schemes for multi-view distributed video coding," *submitted*, 2007.
- [10] A. Fraysse, B. Pesquet-Popescu, and J.C. Pesquet, "On the uniform quantization of a class of sparse source," *submitted*, 2007.
- [11] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner-Ziv video coding performance," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.
- [12] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Smolenice, Slovak Republic, June 2005.