IMPROVING TURBO CODEC INTEGRATION IN PIXEL-DOMAIN DISTRIBUTED VIDEO CODING

Marco Dalai, Riccardo Leonardi

Department of Electronics for Automation,
University of Brescia, Italy
name.surname@ing.unibs.it

Fernando Pereira

Instituto Superior Técnico,
Instituto de Telecomunicações,
Lisbon, Portugal
fp@lx.it.pt

ABSTRACT

The field of Distributed Video Coding (DVC) theory has received a lot of attention in recent years and effective encoding techniques have been proposed. In the present work the framework of pixel domain Wyner-Ziv coding of video frames is considered, following the scheme proposed in [1]. Some key frames are supposed to be available at the decoder while other frames are Wyner-Ziv encoded using turbo codes; at the decoder motion compensated interpolation between key frames is performed in order to construct the side information for the Wyner-Ziv frame decoding. In this paper an improved model for the correlation noise between the side information frame and the original one is proposed. It is shown that modeling nonstationary nature of the noise leads to substantial gain in rate-distortion performance. Also the memory of the noise is considered and the importance of placing an interleaver before the turbo code is shown as well.

1. INTRODUCTION

The new emerging field of Distributed Video Coding has been receiving more and more attention in recent years, being a completely new approach to the ever important problem of video coding. The main idea underlying DVC, whose theoretical foundations relies on Slepian-Wolf' ([2]) and Wyner-Ziv' ([3]) information theory famous work in the '70s, is the possibility of performing video compression by exploiting typical time redundancy of the video sequence in the decoding phase. From a practical perspective this means that instead of doing motion compensation at the encoder, the task is moved to the decoder. We can thus say that within this approach the computational expensive analysis required for video compression can be moved from the encoder to the decoder, reversing the computational complexity allocation with respect to classic video coding paradigm based on motion compensation.

In this context, some practical scheme for DVC have been recently proposed by different groups of researchers (see [4, 1]). For the scheme proposed in [1] in particular some contributions have been given in the literature that focus on improving the performance of the Wyner-Ziv coding by improving the quality of the constructed side information (see for example [5, 6]). Only few attention (see [7]) has been paid instead to the problem of better modeling the correlation between side information and original data in order to improve the channel codes performance. In this work, focusing on the approach proposed in [1] the problem of a finding a good model for the correlation between side information and original data is considered. In particular, the main objective of the paper is to propose a good model for the correlation noise between an original video frame and a prediction obtained by motion compensated interpolation between the adjacent frames. So, we are here only interested in the very basic situation where every odd-indexed frame is supposed to be available at the decoder while even-indexed frames are Wyner-Ziv encoded. This particular assumption is better clarified in the next section, where a brief discussion on the video codec is given. In section 3 and 4, the main contribution of the paper is presented while in section 5 we show some experimental results supporting the theoretical discussion of the paper.

2. VIDEO CODEC ARCHITECTURE

In this paper the video coding architecture firstly proposed in [1] is considered. The starting point for the work to be presented here is an implementation of the cited architecture with some important modification performed by IST group, for which the reader is referred to [5] and [6]. In this section we aim at giving just a quick description of the codec architecture and we refer the reader to [1, 5, 6] for a more detailed discussion.

In the considered architecture every odd-indexed frame X_{2n+1} (key frame) of the video sequence is supposed to be available at the decoder and we focus on the problem of Wyner-Ziv encoding of even indexed frames X_{2n}. For these frames a bit-plane based encoding approach is considered (see Fig. 1); the gray level values are uniformly quantized and the bit-planes are fed one by one into a turbo encoder. A systematic turbo code is used in order to extract parity bits from each

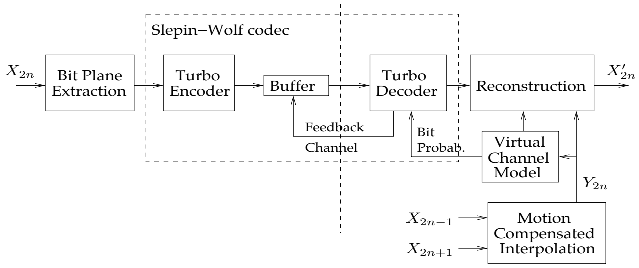


Fig. 1. Architecture of the considered codec

bitplane which are then passed to the decoder.

In the decoding phase, for every even frame X_{2n} an estimation Y_{2n} is constructed by applying motion compensated interpolation between the two adjacent frames X_{2n-1} and X_{2n+1}. The parity bits output from the turbo encoder are then used in order to correct the estimation Y_{2n} and extract a better reconstruction X'_{2n} of the original frame. The codec artchitecture above described is shown in more details in Fig. 1. In this paper we assume, as in [1, 5], that the key frame are losslessly available at the decoder. This hypothesis is not admissible in a practical video coder but the effects of quantization on the key frames can be considered of second importance for the study presented in this paper.

Of the whole architecture shown in Figure 1 it is important to consider here the virtual channel model block and the turbo codec part. In the virtual channel block the correlation model between the side information Y_{2n} and the original frame X_{2n} is used in order to compute bit probabilities to be fed into the turbo decoder where Soft-Input Soft-Output decoders are used (see Fig. 2). This bit probabilities computation is detailed described in the next section where the non stationary noise model is proposed.

3. VIRTUAL CHANNEL MODEL

In the first part of this section the state of the art model for the correlation noise is explained while in the second part the proposed nonstationary model is explained.

3.1. From side information to bit probabilities

Let us call $X_{2n}(r,c)$ the pixel value in the r-th row and c-th column of the $2n$-th frame. The side information Y_{2n} is constructed at the decoder by motion compensated interpolation between frames X_{2n-1} and X_{2n+1}. This means that for every (r,c) point an estimation $Y_{2n}(r,c)$ of the value $X_{2n}(r,c)$ is computed as

$$Y_{2n}(r,c) = \frac{X_{2n-1}(r-v_x, c-v_y) + X_{2n+1}(r+v_x, c+v_y)}{2} \tag{1}$$

where v_x and v_y are (halves of) the estimated motion vector components. We are not interested here in how v_x and v_y may be computed, the reader is referred to [5] for an important contribution in this direction. Once the whole side information frame Y_{2n} has been constructed, it is used for the bit probabilities evaluation. This means that for every point (r,c) the value of $Y_{2n}(r,c)$ is used in order to evaluate the probability of every bit of $X_{2n}(r,c)$ being 1 or 0. What is done in the literature (see [1, 5]) is to consider that the virtual noise between X_{2n} and Y_{2n} has a laplacian distribution with zero mean and estimated standard deviation $1/\alpha$. Hence, for every possible value of the amplitude x, the probability that $X_{2n}(r,c)$ is equal to x is evaluated as[1]

$$p[X_{2n}(r,c) = x] = \frac{1}{2}\alpha \exp\left(-\alpha|x - Y_{2n}(r,c)|\right) \tag{2}$$

Let then $X^i_{2n}(r,c)$ be the i-th bit of the value $X_{2n}(r,c)$ and let Z_i be the set of x values that have i-th bit equal to zero; then for every i we compute

$$p[X^i_{2n}(r,c) = 0] = \sum_{x \in Z_i} p[X_{2n}(r,c) = x] \tag{3}$$

In this way, for a given i bitplane we can compute the probabilities $p^i_0(r,c) = p[X^i_{2n}(r,c) = 0]$ for all the values of r and c, and we consider this probabilities as a priori probabilities to be input to the turbo decoder.

It is important to note that in the above presentation, the α parameter is assumed to be fixed for all (r,c) positions. If we look at bit probabilities as "confidence levels" assigned to the bits, the fact that the α parameter is fixed for different positions means that the side information Y_{2n} is considered to have the same confidence in all points. In other words there is an implicit assumption that the quality of the side information is constant along the frame, without considering motion effects. In the following section we show that it is possible instead to use a nonstationary noise model is order to exploit the motion differences in different areas of the frame.

3.2. Non-stationary implementation

As explained in the previous section, using a laplacian distribution model with a fixed parameters corresponds to giving the same confidence to the side information in every point of the frame. It is not difficult anyway to realize that the quality of the side information is very different from point to point depending on the motion, on occlusions and so on. So, a good model for the noise between X_{2n} and Y_{2n} should take into account this space varying nature and a possible simple model consists in considering the virtual noise to have nonstationary laplacian distribution. In other words we let the α parameter vary from point to point and we thus indicate it with $\alpha(r,c)$. The effect of this choice at the turbo decoding level is that the well predicted values are considered to be more reliable by

[1] Actually the amplitude of the Laplacian must be rescaled in order to have total probability be equal to 1, as the amplitude values x are typically clipped between 0 and 255.

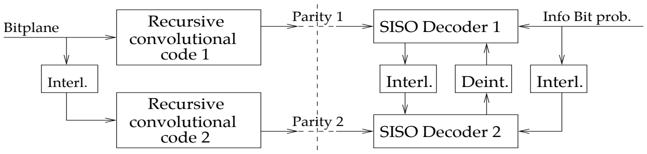


Fig. 2. Turbo codec scheme.

the decoder and it is then easier to correct errors where the discrepancy between X_{2n} and Y_{2n} is actually higher.

In this setting, the important point is then how to set the value of $\alpha(r, c)$ depending on the information we have on the side information confidence in the (r, c) point. A simple yet effective approach we have considered in this work consists on using expression (1) in order to set the value of $\alpha(r, c)$. In fact, for any (r, c) point, other than the value of the obtained side information $Y_{2n}(r, c)$, it is very important to consider the two values $X_{2n-1}(r - v_x, c - v_y)$ and $X_{2n+1}(r + v_x, c + v_y)$ from which $Y_{2n}(r, c)$ is obtained as an average. It is clear in fact that the more those two values differ and the less confidence we should give to their average. So, we should use an expression for $\alpha(r, c)$ in such a way that $\alpha(r, c)$ decreases when the value

$$\Delta(r, c) = |X_{2n-1}(r - v_x, c - v_y) - X_{2n+1}(r + v_x, c + v_y)| \tag{4}$$

increases and viceversa[2].

A possible expression for the $\alpha(r, c)$ values, which has shown to give good empirical results, is the following:

$$\alpha(r, c) = \frac{\beta}{\gamma + \Delta(r, c)}. \tag{5}$$

where α and β are estimated parameters. As for the α parameter, the values of β and γ depend on the type of sequence. We will briefly discuss this fact in Section 5 where experimental results are given.

The main point here, apart from considering particular expressions for the $\alpha(r, c)$ values, is to have a nonstationary model of the noise. In the above expression (5) for $\alpha(r, c)$ we have only used the value of $\Delta(r, c)$ but it is clear that further information may be used, as for example the value of the motion vector of the block containing the point (r, c). Moreover the best choice for the $\alpha(r, c)$ parameter may also depend on the values of $\Delta(i, j)$ for (i, j) in a neighborhood of (r, c) and not just in the very same point.

4. PRE-INTERLEAVING

In the previous section we have presented a possible way of handling the nonstationary nature of the correlation noise. As we said, the main benefit obtained from such an approach is that it improves the turbo decoding process by giving more reliability to better predicted pixels and low reliability to badly predicted ones.

Anyway, another important characteristic of the correlation noise is the memory property. As the correlation noise is mainly due to motion, in fact, it is important to note that the $\alpha(r, c)$ parameter will have most of the times high values on some areas where there is high motion and occlusions and low values where there is low motion (or anyway efficient motion compensation). This implies that, for a generic bitplane, the bits associated to high motion areas are very “noisy” while bits associated to low motion areas have low virtual noise.

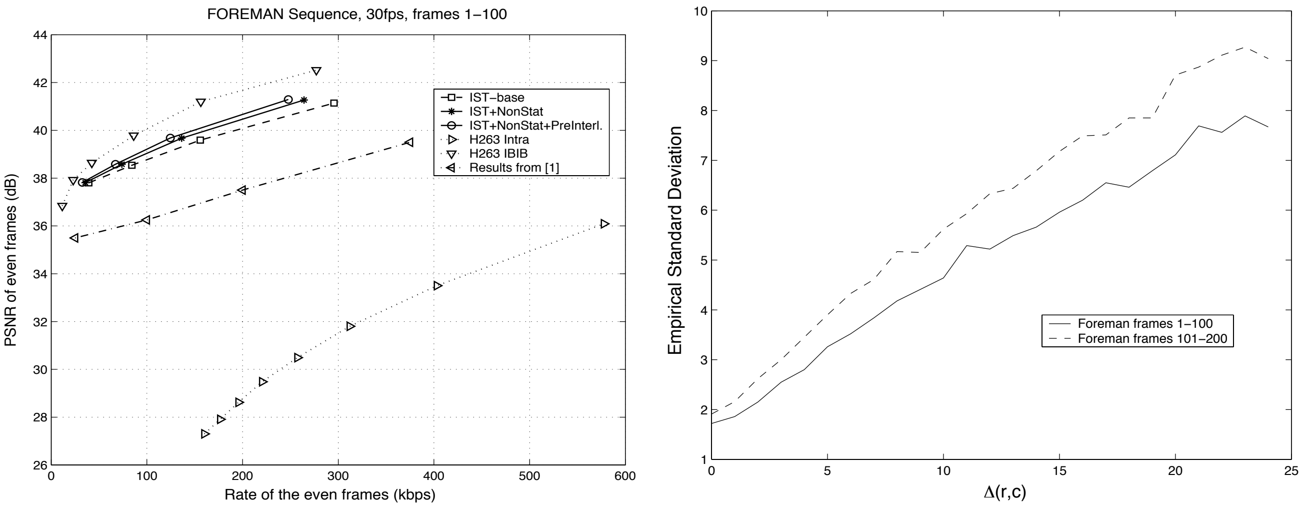
It is therefore important to note that high motion or occlusion areas leads to sort of burst errors in the bitplane. So, if the turbo encoder is fed with bits read row by row from the frame bitplane, the first of the two SISO decoders inside the turbo decoder (see Fig. 2) is faced with the problem of correcting sequences of consecutive very noisy bits. So, due to the fact that the used codes are recursive convolutional codes, in order to correct this noisy areas a high number of parity bits is required from the first decoder. But in the considered scheme the bitrate is managed by using rate compatible puncturing, as explained in [1]. It is then not possible to simply increase the number of parity bits associated to noisy areas, and additional requested parity bits are “spread” all over the frame. This implies that in order to have a sufficient number of parity bits in noisy areas we must have much more parity bits available, most of which are wasted in low noise areas.

Note that this problem does not affect the second SISO decoder inside the turbodecoder, as the interleaver positioned before this second decoder cancels the burst effect spreading noisy bits far apart in the bitstream. So, a very simple but important step in the use of turbo codes in this framework relies on placing an interleaver also before the first encoder. This is only a very simple block in the considered architecture but it gives substantial improvements in the rate distortion performance.

5. EXPERIMENTAL RESULTS

In this section some experimental results are given. In Figure 3(a) the rate distortion performance comparison for the foreman sequence is shown, where the improvements given by non stationary model and by the pre-interleaver are visible. For this sequence we have set $\alpha = 0.37$ for the stationary model. For the non stationary model we set $\gamma = 10$ and we set β so as to have an average standard deviation equal to the stationary model. In this way we are sure that the shown results are only due to the non stationary model and not on some different a priori assumptions. In Fig. 3(b) the empirical standard deviation of the correlation noise conditioned to the value of $\Delta(r, c)$ is shown. It is clearly visible that the value of the correlation noise is strongly correlated with the value of $\Delta(r, c)$. It is also important to note that the first frames

[2] Remember that the standard deviation is $1/\alpha$.



(a) Rate-distortion performance on foreman sequence, 30fps, frames 1-100.

(b) Empirical standard deviation conditioned to $\Delta(r, c)$ on foreman.

Fig. 3. Experimental results on Foreman sequence.

1-100 of foreman sequence contain less motion than frames 101-200; it is then possible to see that the standard deviation of the noise (even when conditioned to $\Delta(r, c)$) depends on the motion level in of the sequence.

6. CONCLUSION AND FUTURE WORK

In this paper a turbo code based DVC architecture has been considered. In this context a new model for the correlation noise between original frames and side information has been proposed and two different contributions have been given in order to improve the performance of turbo codes in the Wyner-Ziv coding of video frames. In this paper only the case of lossless key frame has been considered. Future work will be focusing on the study of key frames quantization effects on the overall discussion.

7. ACKNOWLEDGMENTS

The work presented was developed within DISCOVER (Distributed Coding for Video Services), a Future and Emerging Technologies (FET) project, funded by the European Commission. DISCOVER software started from the IST_WZ software developed at the Image Group from Instituto Superior Tcnico (IST) of Lisbon by Catarina Brites, João Ascenso and Fernando Pereira. We thank Catarina Brites, João Ascenso, Xavi Artigas and Sven Klomp for many helpful discussion on the DISCOVER software.

8. REFERENCES

[1] A. Aaron, R. Zhang, and B. Girod, “Wyner-ziv coding for motion video,” *Asilomar Conference on Signals, Systems and Computers*, November 2002.

[2] S. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. on Information Theory*, vol. 19, no. 4, July 1973.

[3] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. on Information Theory*, vol. 22, no. 1, January 1976.

[4] R. Puri and K. Ramchandran, “PRISM: A new robust video coding architecture based on distributed compression principles,” *40th Allerton Conference on Communication, Control and Computing*, October 2002.

[5] J. Ascenso, C. Brites, and F. Pereira, “Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding,” *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, June 2005.

[6] J. Ascenso, C. Brites, and F. Pereira, “Motion compensated refinement for low complexity pixel based distributed video coding,” *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, September 2005.

[7] A. Trapanese, M. Tagliasacchi, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, “Improved correlation noise statistics modeling in frame-based pixel domain wyner-ziv video coding,” *International Workshop VLBV*, September 2005.