INTRA-WZ QUANTIZATION MISMATCH IN DISTRIBUTED VIDEO CODING

Jürgen Slowack, Jozef Škorupa, Peter Lambert, and Rik Van de Walle

Ghent University – IBBT ELIS Dept., Multimedia Lab Gaston Crommenlaan 8 bus 201 B-9050 Ledeberg-Ghent, Belgium

ABSTRACT

During the past decade, Distributed Video Coding (DVC) has emerged as a new video coding paradigm, shifting the complexity from the encoder- to the decoder-side. This paper addresses a problem of current DVC architectures that has not been studied in the literature so far, that is, the mismatch between the intra and Wyner-Ziv (WZ) quantization processes. Due to this mismatch, WZ rate is spent even for spatial regions that are accurately approximated by the side-information. As a solution, this paper proposes side-information generation using selective unidirectional motion compensation from temporally adjacent WZ frames. Experimental results show that the proposed approach yields promising WZ rate gains of up to 7% relative to the conventional method.

Index Terms— Distributed Video Coding, Wyner-Ziv coding, quantization

1. INTRODUCTION

Distributed Video Coding (DVC) has received a lot of research attention in the past decade, being a radically new way of performing video compression. Instead of exploiting the similarities between frames at the encoder, in DVC these similarities are exploited at the decoder. This leads to a reversed complexity distribution, with a simple encoder but complex decoder as opposed to a complex encoder but simple decoder in conventional systems, e.g., H.264/AVC.

Most of the current DVC systems described in the literature are based on the architecture initially proposed by Aaron et al. [1]. In such systems, the frame sequence is partitioned into key frames and Wyner-Ziv (WZ) frames. The key frames are coded without using other frames as references (e.g., through H.264/AVC intra coding). For each WZ frame, the decoder generates a prediction – called side information – using already decoded frames as references. Next, errors in the side information are corrected via error correcting information sent by the encoder (such as turbo or LDPC codes).

Nikos Deligiannis, Adrian Munteanu

Vrije Universiteit Brussel – IBBT ETRO Department, Pleinlaan 1 B-1050 Brussels, Belgium

Many researchers have made contributions focusing on different aspects of DVC. In particular, in the context of side information generation, contributions include spatial smoothing and subpixel refinement [2], mesh-based techniques [3], temporal non-uniform interpolation [4], and partial decoding and refinement [5].

Other research directions followed in the context of DVC focus on estimating the correlation between the side information (available at the decoder) and the original (available at the encoder) [6, 7], and on eliminating the feedback channel [8, 9].

In this paper we identify an additional problem in DVC that has not been studied in the literature so far, i.e., the mismatch between intra and WZ quantization. A detailed description of this problem is provided in Section 2. Next, in Section 3, we provide a solution to this problem through selective unidirectional motion compensation. The results in Section 4 indicate bit rate improvements up to 7% for the WZ frames. Finally, conclusions and a discussion of future work are provided in Section 5.

2. PROBLEM ANALYSIS

The quantization process in the context of intra coding is quite different from its counterpart in the context of WZ coding. First of all, the nature and range of the values that are provided as inputs to the quantizer are different. For example, in H.264/AVC intra coding the encoder first generates a spatial prediction of the current block of pixels, and subsequently transforms and quantizes the residual between the current block and its prediction. In WZ coding, the procedure is in general much simpler, i.e., no prediction is generated at the encoder and the pixel values are transformed and quantized directly. Apart from the fact that the nature of the input values is different between the intra and the WZ quantizer, both quantizers typically have different bin widths, and different characteristics (e.g., having a deadzone or not). At the decoder, the reverse quantization steps are different as well; for example, optimal minimum mean square error reconstruction



Fig. 1. Consider a value X (e.g., a transform coefficient) in a temporally non varying region. This value could be intra reconstructed as X'_I in I'_1 and I'_5 . However, when using X'_I as part of the side information for W_3 , rate will be spent inefficiently to obtain the correct WZ bin Q_k containing X.

is performed in WZ decoding which is not the case in intra decoding.

In conventional DVC architectures the common approach is to generate side-information using MCI techniques based on previously decoded information, be it intra and/or WZ frames. In temporally not varying regions, although MCI provides efficient prediction, due to the mis-alignment between the quantizers in the intra and WZ frames, the resulting sideinformation and the WZ samples to be predicted do not necessarily fall in the same quantization interval. As a consequence, unnecessary WZ rate is spent for these regions.

Let us consider for example a GOP of size four, i.e., I_1 – $W_2 - W_3 - W_4 - I_5$, decoded as $I'_1 - W'_2 - W'_3 - W'_4 - I'_5$. Following a hierarchical procedure, the intra frames I_1 and I_5 will be coded first and sent to the decoder where they are reconstructed as I'_1 and I'_5 , respectively. Next, the decoder will generate side information Y_3 for W_3 , using I'_1 and I'_5 as references. However, even if a certain spatial region in Y_3 is temporally non varying, due to quantization mismatch, the corresponding intra reconstructed regions in I'_1 and I'_5 might contain values that lie outside the correct WZ quantization bins in W_3 . Hence, despite of accurate interpolation, rate will be spent inefficiently by the channel decoder in trying to resolve the quantization mismatch and retrieve the correct WZ bins. This effect is illustrated in Figure 1. Note that Figure 1 is a simplified representation, showing only differences in alignment and bin width of the quantizers, ignoring differences with respect to transformation or other preprocessing steps.

Also for the other WZ frames in the GOP, rate will be spent inefficiently since the decoder will use at least one decoded I frame as a reference frame for generating the side information through bidirectional motion compensation, namely, I'_1 and W'_3 for generating Y_2 , and W'_3 and I'_5 for generating Y_4 .

3. SOLUTION

To resolve the mismatch and improve compression, several strategies could be followed. A first approach could be to adjust the intra coding process, increasing the probability that the intra reconstructed values lie within the correct WZ bins.

However, these adjustments will most likely lower the compression performance of the I frames, and it could therefore be difficult to realize overall compression gain.

In this paper we propose an alternative technique as solution to the quantizer mistmatch problem. Our method focuses on WZ frames for which side information is generated using one intra decoded frame and one WZ decoded frame as references. The (de)coding procedure for the remaining WZ frames is left unmodified. Hence, for a GOP of size four, this will impact W_2 and W_4 but not W_3 . The technique affects the decoder only.

3.1. Main algorithm

The solution in this paper is based on the fact that the WZ quantizer in our system [10] remains the same for the entire sequence (as in most DVC systems). As a result, in general, there is no quantizer mismatch between WZ frames¹, only between I frames and WZ frames. Therefore, when generating side information using one intra decoded reference frame and one WZ decoded reference frame, the decoder can prefer using the WZ decoded reference frame only, to avoid intra-WZ quantization mismatch. The possible benefit of a such a selective unidirectional approach is illustrated by Figure 2.

Our solution is accomplished at the final step of the side information generation process², i.e., when the decoder has obtained a bidirectional motion vector for each block, and is about to perform bidirectional motion compensation by calculating the average between a past reference block B_P and a future reference block B_F . In specific, in addition to creating side information Y through bidirectional motion compensation, we also create a version Y^* where some of the blocks are obtained via unidirectional motion compensation from the WZ reference frame only. In other words, instead of using the average of B_P and B_F , for some blocks in Y_4^* we will use B_P only, while for some blocks in Y_2^* we will use B_F .

The decision between unidirectional and bidirectional motion compensation for Y^* is based on a comparison between B_P and B_F . If the difference between both blocks is considered due to quantization noise only, the block is assumed to be temporally non-varying and so unidirectional compensation from the WZ reference frame is performed.

In specific, the evaluation is performed as follows. As explained in our previous work [10], an estimation of the quantization noise is generated at the decoder. To this end, denote σ_Q the standard deviation of the decoder-estimated pixel-domain quantization noise. Then, unidirectional motion compensation is performed in case a sufficient number of T corresponding pixels in B_P and B_F have an absolute difference not greater than $2\sigma_Q$. In other words, unidirectional

¹This is not the case for regions with motion, as described later on.

²The techniques for side information generation in our system are similar to the techniques applied in DISCOVER [2].



Fig. 2. Highlighted blocks indicate where the DC coefficient will need to be corrected, in case – (left) I'_1 , (right) W'_3 , or (center) the conventional MCI side info Y_2 – is used as side information for decoding W_2 . Despite the fact that Y_2 is fairly accurate, the background is not entirely accurately predicted, mainly due to quantization mismatch caused by interpolation using I'_1 . Hence, instead, the decoder could follow a unidirectional approach, using only information from W'_3 to create the background in Y_2 .

compensation is performed if:

$$\sum_{x,y} c_{B_F,B_F}(x,y) \ge T,\tag{1}$$

$$c_{B_P,B_F}(x,y) = \begin{cases} 1 & \text{if } |B_P(x,y) - B_F(x,y)| \le 2\sigma_Q \\ 0 & \text{otherwise} \end{cases},$$

where x and y denote the horizontal and vertical offset within B_P and B_F . This evaluation is performed on 4x4 blocks (so $0 \le x, y \le 4$). The threshold T is set to 12 in this work, meaning that at least 12 pixels out of 16 in one block should be comparable in order to perform unidirectional motion compensation. In the opposite case, bidirectional motion compensation is performed instead.

After turbo decoding using Y^* , the coefficients are reconstructed through centroid reconstruction.

3.2. Further refinements

The method proposed here is still further refined, due to two reasons. A first reason is that the WZ reconstructed values are only guaranteed to lie within the correct bins for a 4x4 grid of reconstructed blocks, and not for any intermediary blocks within the grid. Therefore, we introduce an additional constraint for unidirectional motion compensation, namely that the unidirectional motion vector to the WZ reference frame should be of the form (4i, 4j), with $i, j \in \{0, 1, 2, ...\}$. If this is not the case, bidirectional motion compensation is performed instead.

A second problem is that inaccuracies in the estimation of the side information, different noise types (such as camera noise), or changing lighting conditions can decrease the performance of the proposed algorithm. Therefore, the method is made adaptive. After decoding, the decoder analyzes the block positions for which unidirectional motion compensation has been performed. For each transformation coefficient in these blocks, the decoder determines the quantization bin Q_{bid} containing the bidirectionally compensated side information (provided by Y), as well as the bin Q_{uni} of the unidirectionally compensated alternative (provided by Y^*). With the decoded (and hence correct) bin denoted Q_{dec} , the decoder determines the fraction of coefficients for which $Q_{uni} = Q_{dec}$, out of the set of coefficients for which $Q_{uni} \neq Q_{bid}$. If this fraction is at least 70%, the algorithm well, and so it will be applied to the following WZ frame. Otherwise, the following WZ frame will be decoded using Y instead of Y^* . Regardless of whether the proposed method has been applied or not for turbo decoding, the evaluation as described here is always performed.

4. RESULTS

The techniques proposed in this paper have been implemented in our codec described in [7] with the extension explained in [10]. This codec is similar to DISCOVER but it implements improved methods for correlation noise modeling.

Results are provided in Table 1 for different sequences in CIF resolution at 30 fps, for a GOP of size 4. Despite the fact that our technique only applies to two out of three WZ frames in each GOP, still, we are able to achieve overall WZ rate gains up to 7%, as well as total gains up to 3% (key frames + WZ frames). Our technique works particularly well in case the quantization noise is high. For fine quantization, small temporal variations (e.g., caused by camera noise, or slightly changing lighting conditions) become relatively more important. These temporal variations are not compensated for by our technique, which explains why there is less gain for fine quantization levels.

When comparing the results for the different sequences,

	Mother and Daughter				Silent				Foreman			
	Rate	PSNR	ΔWZ	Δ	Rate	PSNR	ΔWZ	Δ	Rate	PSNR	ΔWZ	Δ
Q1	1029	41.9	0%	0%	1975	38.3	0%	0%	2636	37.8	0%	0%
Q2	593	39.1	0%	0%	1091	35.1	2%	1%	1326	34.0	0%	0%
Q3	340	36.5	6%	3%	601	32.4	5%	2%	669	30.8	0%	0%
Q4	196	33.8	7%	3%	310	29.8	7%	3%	333	27.6	2%	1%

Table 1. Results for different sequences in CIF resolution at 30 fps, coded with a hierarchical GOP of length 4. The rows indicate results from fine to coarse quantization (i.e., Q1 to Q4). For each sequence, each first column lists the total bit rate in kpbs (key + WZ frames) of all color components, while each second column lists the luma PSNR in dB (for key + WZ frames). The third column (labeled Δ WZ) indicates the gain in bit rate for the WZ frames only, while the fourth column (labeled Δ) provides the overall bit rate gain. Gains are relative to our previous work [10].

we observe that the highest gains are achieved for Mother and Daughter, and Silent, which are sequences with no global motion and reasonably large spatially non-varying backgrounds. For Foreman, less gain was expected, due to the presence of global motion, and the fact that our technique only compensates in case the unidirectional motion vector is of the form (4i, 4j), with $i, j \in \{0, 1, 2, ...\}$.

Compared to H.264/AVC, the proposed system outperforms H.264/AVC intra coding with average (Bjøntegaard delta) quality improvements of 3.7 dB for Mother and Daughter, 3.3 dB for Silent, and 0.3 dB for Foreman. On the other hand – similar to other DVC systems in the literature – there is a quality gap with H.264/AVC inter coding of 1.8 dB for Mother and Daughter, 1.9 dB for Silent, and 3.8 dB for Foreman.

5. CONCLUSIONS AND FUTURE WORK

This paper was a first study on quantization mismatch within DVC, and further research is required in this area. For example, since our technique applies to certain WZ frames in a GOP, it would be interesting to investigate how to obtain gains for the other WZ frames as well. In addition, one could study how to adjust the intra and WZ coding process so that quantization mismatch is completely or partially avoided.

6. REFERENCES

- A. Aaron, S. Rane, E. Setton, and B. Girod, "Transformdomain Wyner-Ziv codec for video," in *Proc. SPIE Vi*sual Communications and Image Processing, Jan. 2004, vol. 5308, pp. 520–528.
- [2] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, techniques and evaluation," in *Proc. Picture Coding Symposium (PCS)*, Nov. 2007.
- [3] D. Kubasov and C. Guillemot, "Mesh-based motioncompensated interpolation for side information extraction in Distributed Video Coding," in *Proc. IEEE Inter-*1000 July 2010 - 2010

national Conference on Image Processing (ICIP), Oct. 2006.

- [4] G. Petrazzuoli, M. Cagnazzo, and B. Pesquet-Popescu, "High order motion interpolation for side information improvement in DVC," in *Proc. IEEE International Conference on Acoustics Speech and Signal Processing* (*ICASSP*), Mar. 2010, pp. 2342 –2345.
- [5] R. Martins, C. Brites, J. Ascenso, and F. Pereira, "Refining side information for improved transform domain Wyner-Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1327–1341, Sep. 2009.
- [6] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, pp. 1177–1190, Sep. 2008.
- [7] J. Škorupa, J. Slowack, S. Mys, N. Deligiannis, J. De Cock, P. Lambert, A. Munteanu, and R. Van de Walle, "Exploiting quantization and spatial correlation in virtual-noise modeling for distributed video coding," *Signal Processing: Image Communication*, vol. 25, no. 9, pp. 674 – 686, 2010.
- [8] T. Sheng, X. Zhu, G. Hua, H. Guo, J. Zhou, and C. W. Chen, "Feedback-free rate-allocation scheme for transform domain Wyner-Ziv video coding," *Multimedia Systems*, Jan. 2010.
- [9] C. Yaacoub, J. Farah, and B. Pesquet-Popescu, "Content adaptive gop size control with feedback channel suppression in distributed video coding," in *IEEE International Conference on Image Processing (ICIP)*, Nov. 2009.
- [10] J. Škorupa, J. De Cock, J. Slowack, S. Mys, N. Deligiannis, P. Lambert, A. Munteanu, and R. Van de Walle, "Correlation modeling with decoder-side quantization distortion estimation for distributed video coding," in *Proc. Picture Coding Symposium (PCS)*, Dec. 2010.