

BITPLANE INTRA CODING WITH DECODER-SIDE MODE DECISION IN DISTRIBUTED VIDEO CODING

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

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

Nikos Deligiannis,
Adrian Munteanu

Vrije Universiteit Brussel,
ETRO Department – IBBT
Pleinlaan 2, B-1050 Brussels, Belgium

ABSTRACT

While distributed video coding (DVC) has emerged as a new video coding paradigm, the compression performance of current systems is still low compared to conventional solutions such as H.264/AVC. While the latter uses many coding modes and an efficient mode decision strategy for choosing the best mode, in DVC, only a limited number of modes has been developed so far. Since encoder-side mode decision in DVC increases encoder's complexity, in this paper, we introduce decoder-side mode decision choosing between bitplane WZ coding and bitplane intra coding. This strategy proves to be efficient, delivering rate gains up to 22% over DISCOVER, without increasing the complexity at the encoder.

Index Terms— Distributed video coding, mode decision, Wyner-Ziv coding

1. INTRODUCTION

Video compression is typically performed by exploiting redundancies in the frame sequence. The temporal redundancies are usually exploited through a process called *motion estimation*. While this highly-complex task is performed at the encoder in conventional solutions such as MPEG-2 and H.264/AVC, this task is assigned to the decoder in distributed video coding (DVC). As a result, the complexity of the encoder is low compared to the complexity of the decoder, which facilitates a range of applications including visual sensor networks and networked camcorders [1].

The most frequently adopted DVC architecture is the one proposed by Aaron et al. [2, 3]. In this system, the frame sequence is split into key frames and Wyner-Ziv (WZ) frames. Key frames are intra coded, i.e. without using other frames as references. For each WZ frame, a prediction is generated at the decoder using block-based motion estimation techniques. This decoder-side prediction is called the *side information*. Next, the encoder sends error correcting information calculated on the original frame, allowing the decoder to correct the

side information. This basic architecture has been extended by numerous researchers through the years. Some interesting side information generation techniques have been proposed in the context of DISCOVER [4], as well as by other researchers [5, 6, 7]. Information about the correlation between the side information at the decoder and the original at the encoder, is needed at the decoder when using high-performance channel coding techniques (e.g. LDPC or turbo coding) for correcting the side information. Online techniques for estimating this correlation have been proposed (e.g. [8, 9, 10]).

Despite these and many other contributions, the compression efficiency of DVC systems is still not comparable to state-of-the-art conventional video compression with encoder-side motion estimation, such as H.264/AVC. One of the possible reasons for this performance gap is the lack of an effective mode decision strategy. While a large number of possible predictors are generated in H.264/AVC, only a limited number of coding modes are currently used in DVC.

Most DVC techniques apply mode decision at the encoder. Typically, the encoder generates an estimation of the side information using low-complexity techniques, and mode decision is performed by comparing the original frame to this estimation. In PRISM, a block-based approach is followed, and blocks are classified as skip, intra or WZ, and coded accordingly. Mys et al. [11] use similar ideas to introduce a block-based skip mode in the popular architecture by Aaron et al. [3]. Ascenso et al. [12] use fast motion estimation techniques for encoder-side mode decision between intra mode and WZ mode. Intra blocks are coarsely quantized and used at the decoder to improve the quality of the side information.

One of the drawbacks of encoder mode decision in DVC is that there is a trade-off between encoder complexity and compression performance, since the best results will be achieved if the encoder has an accurate estimation of the side information. Therefore, in this paper, an intra mode is defined for which mode decision is performed at the decoder instead of at the encoder. As a result, the complexity of the encoder remains fairly the same.

After a description of the codec architecture in Sect. 2, the results of offline experiments are presented in Sect. 3. These experiments are then used to define a mode decision strategy (Sect. 4), which shows to be very effective (Sect. 5). Conclusions are provided in Sect. 6.

2. CODEC ARCHITECTURE

The codec architecture proposed in this paper is depicted in Fig. 1. This architecture is based on the DVC architecture developed at Stanford [3], with some important improvements adopted from more recent work (as will be discussed). The main novelty in this paper is the addition of a bitplane intra coding mode with decoder-side mode decision. The motivation for this extension is based on offline analysis, as described in Sect. 3. However, to understand the context of this analysis, we first discuss the operation of the codec in detail.

At the encoder, the frame sequence is split into key frames and WZ frames. Key frames are intra coded using H.264/AVC intra coding. WZ frames are partitioned into non-overlapping blocks of size 4-by-4, and each block is transformed using a discrete cosine transform (DCT). Coefficients at identical positions are grouped into coefficient bands, resulting in a total of 16 coefficient bands per color component. Next, each coefficient band is quantized by a quantizer having 2^{M_k} levels. Bits at corresponding positions (e.g. all most significant bits) are grouped into so-called bitplanes, resulting into M_k bitplanes per coefficient band. Each bitplane is then coded using one out of two options, for which information is received from the decoder. If the intra mode is selected by the decoder (1), a binary arithmetic coder is used to code the bitplane. This arithmetic coder applies an adaptive context model that is initialized to a uniform distribution before the bitplane is coded. The coded bitplane is then sent to the decoder. On the other hand, the decoder can choose to use a conventional turbo coding strategy (2), as used in many Stanford-based systems. In this mode (referred to as the WZ mode), parity bits are calculated by a turbo coder, and sent to the decoder in portions upon request.

At the decoder, coded key frames are decoded into I' , using H.264/AVC intra decoding. For each WZ frame, a prediction Y is generated using the techniques implemented in the DISCOVER codec [4]. Next, the decoder decides the coding mode for each bitplane, and sends this information to the encoder, as discussed further on. Bitplanes coded using the arithmetic coder (1) are decoded without using the side information or any other information, delivering the original bitplane as a result. In the second case (2), the turbo decoder requests parity bits from the encoder until reliable decoding can be performed. The criterion for reliable decoding has been developed in our previous work [13]. In addition, the techniques used for online correlation noise estimation are described in [9]. Finally, if all bitplanes are decoded (using either mode), they are grouped and the reconstructed coef-

ficients are calculated through centroid reconstruction [14]. The decoded frame W' is then obtained after applying the inverse DCT.

Sending modes from decoder to encoder introduces a small overhead. In our system, the 16 luma coefficient bands are each quantized to minimum three and maximum eight bits. This means that between 48 and 128 bitplane modes need to be communicated. Using one bit to signal WZ or intra, the overhead for a GOP of size 4 at 30 fps (as used in the experimental results) is between 1.08 and 2.88 kbps. This corresponds to a negligible bit rate overhead below 0.5%.

3. INTRODUCING A NON-WZ MODE

The motivation for using an additional bitplane mode is based on offline analysis. Note that the decoded bitplane in intra mode is always equal to the original bitplane, while only high reliability is obtained for the WZ mode. However, in practice the difference in distortion between both modes is negligible, so that only bit rate is considered in our analysis.

Offline tests have been performed. In a first pass, the arithmetic coding option is used exclusively. For each frame, and for each bitplane, the coding rate is stored in a file. In a second pass, the WZ coding option is used exclusively, and bit rates are stored in a similar fashion. To illustrate the evolution of these bit rates as a function of the frame index, as an example, results are generated for the least significant bitplane of the DC coefficient, for the first 129 frames of the Table Tennis sequence (CIF resolution, 30 fps). A GOP of size 4 is used, and a hierarchical coding structure is applied. This means that the sequence $I_1 - W_1 - W_2 - W_3 - I_2 - \dots$ is decoded in the following order: $I_1 - I_2 - W_2 - W_1 - W_3 - \dots$. Hence, two layers of WZ frames can be identified, depending on the distance between the reference frames used for generating the side information. The first layer contains frames W_2, W_6 , etc., which use the decoded I frames for generating the side information. For the second layer, containing frames W_1, W_3, W_5, W_7 , etc., the reference frames are much closer (e.g. decoded frames I'_1 and W'_2 for W_1). The DC coefficient is quantized to 5 bits and mapped using a gray mapping. The intra quantization parameter is chosen so that intra frames and WZ frames have similar quality.

The results are depicted in Fig. 2, with separate graphs for each layer of WZ frames. An interesting observation is that the WZ coding mode is not always the best option. Instead, in some cases it appears to be more advantageous to use an intra coding mode, and code and decode without using the side information at all. This is due to inaccuracies in both the side information as well as the correlation noise estimation.

4. BITPLANE-LEVEL MODE DECISION

In this paper, mode decision is performed at the decoder instead of at the encoder. This is in favor of the DVC paradigm

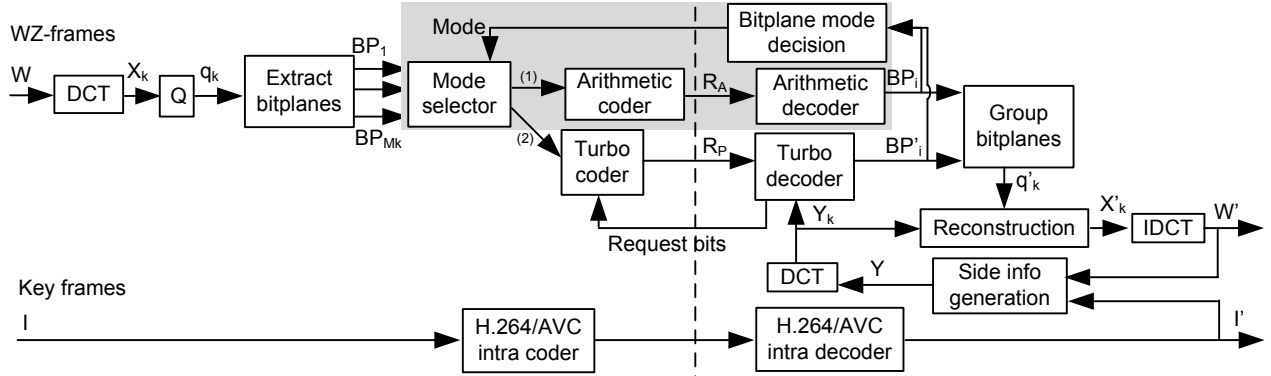


Fig. 1. The proposed DVC architecture, featuring an additional bitplane intra coding mode with decoder-side mode decision.

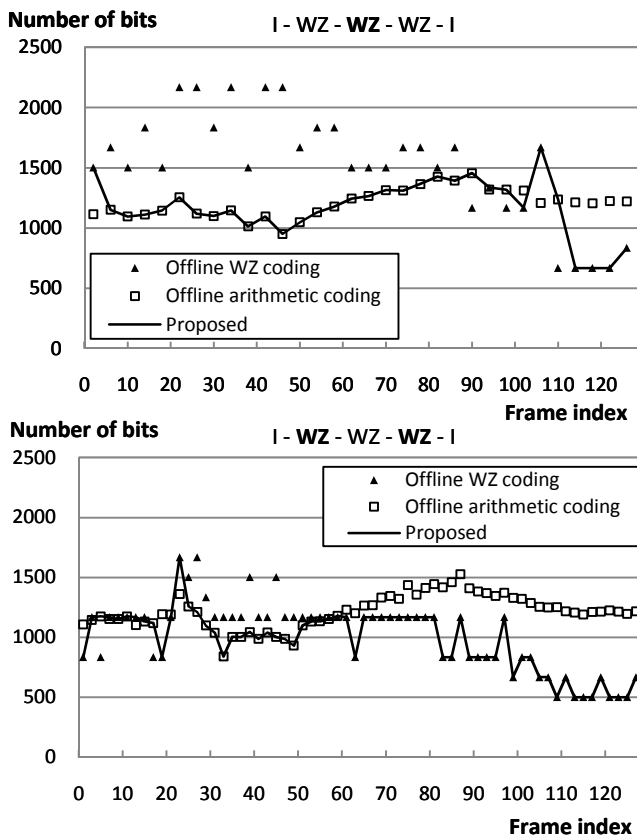


Fig. 2. Bitplane intra coding versus bitplane WZ coding, for the least significant bitplane of the DC, for two different layers of WZ frames. The proposed online mode decision illustrates significant gains over using only the WZ mode.

featuring simple encoders. From the results in Fig. 2, strong temporal correlation can be observed: if the WZ coding mode is best for a bitplane k in frame F , then chances are high that it will be best for bitplane k in the next WZ frame (in the same layer of WZ frames). However, since only one mode is used for coding a particular bitplane in F , the bit rate for the other

mode needs to be estimated. To this extent, the following procedure is used at the decoder to estimate the bit rate:

- if bitplane k was WZ coded, the decoded bitplane is coded using the arithmetic coder,
- if bitplane k was intra coded, a turbo coding/decoding strategy is adopted.

Hence, in both cases, an estimation of the bit rate is obtained through coding of the decoded bitplane, which is available at the decoder. This simple strategy is effective, as illustrated by the actual rate spent in the proposed system, switching between bitplane intra coding and bitplane WZ coding (Fig. 2).

5. RESULTS

Tests have been conducted on several sequences in CIF resolution at 30 fps. A GOP of size 4 is used and only the luma component is coded, to allow comparing with the DISCOVER codec. As a reference, results have been generated using the system described in this paper, but excluding the intra mode. This configuration is labeled “previous work”.

The results in Fig. 3 indicate that adding a bitplane intra mode increases performance significantly. Average Bjøntegaard [15] rate gains over DISCOVER can be observed of 22% for Foreman, and 14% for Coastguard. The gain is the largest at high rates, where the correlation between the side information bitplanes and the original bitplanes is poor.

Good results are obtained for other sequences as well (not depicted due to space constraints), leading to average rate gains over DISCOVER of 14% for Mother and Daughter, and 8% for Table Tennis.

6. CONCLUSIONS

A technique for decoder-side mode decision has been introduced, choosing between coding a bitplane using a WZ approach or an intra approach. The results indicate that DVC

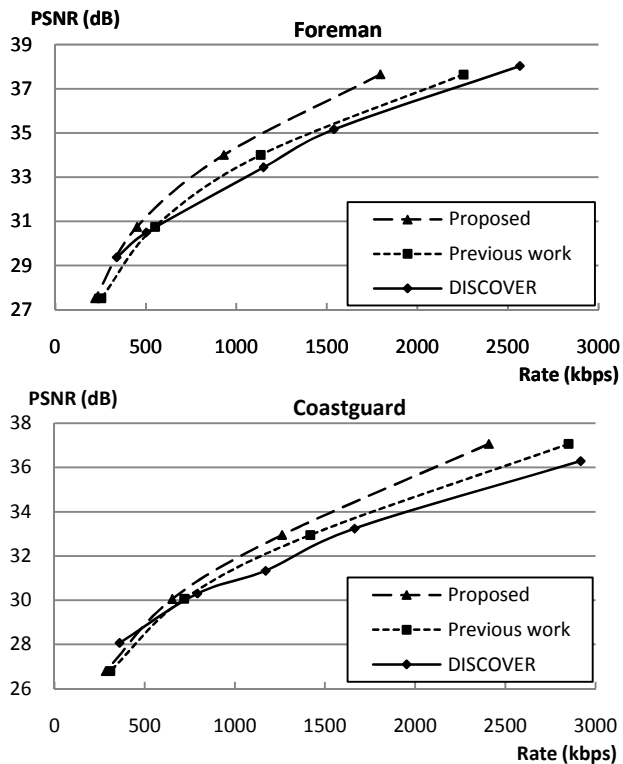


Fig. 3. Compared to DISCOVER, our method achieves average rate gains of 22% for Foreman and 14% for Coastguard.

system performance can be improved by using different coding modes, without the need for side information estimation at the encoder.

7. REFERENCES

- [1] Fernando Pereira, Luis Torres, Christine Guillemot, Touradj Ebrahimi, Riccardo Leonardi, and Sven Klomp, "Distibuted video coding: Selecting the most promising application scenarios," *Signal Processing: Image Communication*, pp. 339–352, 2008.
- [2] Anne Aaron, Rui Zhang, and Bernd Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conference on Signals, Systems and Computers*, November 2002, vol. 1, pp. 240–244.
- [3] Anne Aaron, Shantanu Rane, Eric Setton, and Bernd Girod, "Transform-domain Wyner-Ziv codec for video," in *Proc. SPIE Visual Communications and Image Processing*, January 2004, vol. 5308, pp. 520–528.
- [4] 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)*, November 2007.
- [5] Denis Kubasov and Christine Guillemot, "Mesh-based motion-compensated interpolation for side information extraction in Distributed Video Coding," in *Proc. IEEE International Conference on Image Processing (ICIP)*, October 2006.
- [6] S. Ye, M. Ouaret, F. Dufaux, and T. Ebrahimi, "Improved side information generation for distributed video coding by exploiting spatial and temporal correlations," *EURASIP Journal on Image and Video Processing*, vol. 2009, 2009, Article ID 683510.
- [7] Xiaopeng Fan, Oscar C. Au, Ngai Man Cheung, Yan Chen, and Jiantao Zhou, "Successive refinement based Wyner-Ziv video compression," *Signal Processing: Image Communication*, 2009, doi:10.1016/j.image.2009.09.004.
- [8] 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, September 2008.
- [9] Jürgen Slowack, Stefaan Mys, Jozef Škorupa, Peter Lambert, Christos Grecos, and Rik Van de Walle, "Accounting for quantization noise in online correlation noise estimation for distributed video coding," in *Proc. Picture Coding Symposium (PCS)*, May 2009.
- [10] Xin Huang and Søren Forchhammer, "Improved virtual channel noise model for transform domain Wyner-Ziv video coding," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009, pp. 921–924.
- [11] Stefaan Mys, Jürgen Slowack, Jozef Škorupa, Peter Lambert, and Rik Van de Walle, "Introducing skip mode in distributed video coding," *Signal Processing: Image Communication*, vol. 24, no. 3, pp. 200–213, 2009.
- [12] J. Ascenso and F. Pereira, "Low complexity intra mode selection for efficient distributed video coding," in *Proc. International Conference on Multimedia and Expo (ICME)*, July 2009, pp. 101–104.
- [13] Jozef Škorupa, Jürgen Slowack, Stefaan Mys, Peter Lambert, Christos Grecos, and Rik Van de Walle, "Stopping criterions for turbo coding in a Wyner-Ziv video codec," in *Proc. Picture Coding Symposium (PCS)*, May 2009.
- [14] D. Kubasov, J. Nayak, and C. Guillemot, "Optimal reconstruction in Wyner-Ziv video coding with multiple side information," in *IEEE MultiMedia Signal Processing Workshop*, October 2007, pp. 183–186.
- [15] Gisle Bjøntegaard, "Calculation of average PSNR differences between RD-curves," Tech. Rep., VCEG, April 2002, Contribution VCEG-M33.