IMPROVED SIDE INFORMATION GENERATION WITH ITERATIVE DECODING AND FRAME INTERPOLATION FOR DISTRIBUTED VIDEO CODING

Shuiming Ye, Mourad Ouaret, Frederic Dufaux, Touradj Ebrahimi

Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

ABSTRACT

Distributed Video Coding (DVC) is a new paradigm in video coding, which is receiving a lot of interests nowadays. Side Information (SI) generation is a key function in the DVC decoder, and plays a key-role in determining the performance of the codec. This paper proposes an improved side information generation scheme, which exploits both spatial and temporal correlations in the sequences. Partially decoded Wyner-Ziv (WZ) frames, based on initial SI by Motion Compensation Temporal Interpolation (MCTI), are exploited to improve the performance of the whole SI generation. In addition, an enhanced temporal frame interpolation is proposed, including motion vector refinement and smoothing, optimal compensation mode selection, and a new matching criterion for motion estimation. Simulation results show that the proposed scheme can achieve up to 2.3 dB improvement in Rate Distortion (RD) performance for video with high motion, when compared to state-of-the-art DVC.

Index Terms— Distributed Video Coding, Side Information, Motion Estimation, Motion Compensation, Motion Filtering

1. INTRODUCTION

Nowadays, the most popular digital video coding solutions are represented by the ISO/IEC MPEG and ITU-T H.26x standards, which rely on a highly complex encoder. However, in some emerging applications, such as wireless low-power video surveillance, multimedia sensor networks, and wireless PC cameras, low complexity encoding is required. Distributed Video Coding (DVC) [1], a new paradigm in coding which allows for very low complexity encoding, is well-suited for these applications.

We consider a DVC architecture which divides a video sequence into key frames and WZ frames. The key task to exploit source statistics is carried out in SI generation process to produce an estimation of the WZ frame being decoded. SI has a significant influence on the RD performance of DVC. Indeed, more accurate SI implies that fewer bits are requested from the encoder, so that the bitrate is reduced for the same quality. In common DVC codecs, the SI is obtained by MCTI from the previous and next key frames and utilizes the Block Matching Algorithm (BMA) for motion estimation. However, motion vectors from BMA are often not faithful to true object motions. Unlike classical video compression, it is more important to find true motion vectors for SI generation in DVC.

Several attempts have been made to improve SI generation. Spatial motion vector smoothing was proposed to improve the performance of bi-directional MCTI in [2, 3]. However, spatial motion smoothing is only effective at removing false vectors that occur as isolated impulsive noises. In [4], the authors proposed to use sub-pixel interpolation to improve motion estimation, but it is only based on the key frames at the decoder. Joint decoding and motion estimation was proposed to conduct more accurate motion estimation at the decoder, based on auxiliary information sent by the encoder, such as the Cyclic Redundancy Check (CRC) [5] and hash [6] bits. On the other hand, these approaches increase the complexity of the encoder. Further, iterative decoding and frame interpolation has also been proposed, such as motion vector refinement via bitplane refinement [7] and iterative MCTI techniques [8, 9], but with a high cost of several iterations of motion estimation and decoding.

This paper proposes a new SI generation scheme by exploiting spatio-temporal correlations at the decoder. It uses partially decoded WZ frame generated by the WZ decoder to improve SI generation. In other words, the proposed scheme is not only based on the key frames, but also on the WZ bits already decoded. Furthermore, an enhanced temporal frame interpolation is applied, including motion vector refinement and smoothing, optimal compensation mode selection, and a new matching criterion for motion estimation. Experimental results show that the proposed scheme significantly improves the quality of the SI and RD performance of DVC.

The paper is organized as follows. First, the DVC architecture used in this work is introduced in section 2. Then, the proposed SI generation scheme is presented in section 3. The simulation results are presented in section 4. Finally, section 5 concludes the paper.

2. DVC ARCHITECTURE

In this paper, we consider the Transform Domain Wyner-Ziv (TDWZ) DVC architecture from [3], as shown in Figure 1. A video sequence is divided into key frames and WZ frames. Hereafter, we consider a Group of Pictures (GOP) size of 2, namely the odd and even frames are key frames and WZ frames, respectively. Key frames are conventionally encoded using H.264/AVC Intra coding [10]. Conversely, for WZ frames, a DCT transform is firstly applied, and the resulting transform coefficients undergo quantization. The quantized values are then split into bitplanes which are turbo encoded. At the decoder, SI approximating the WZ frames is generated by MCTI of the decoded key frames. The SI is used in the turbo decoder, along with WZ parity bits requested from feedback channel, in order to reconstruct the decoded WZ frames. In this paper, the turbo decoder stops requesting more bits if the bitplane bit-error rate is below a given threshold equal to 10^{-3}.

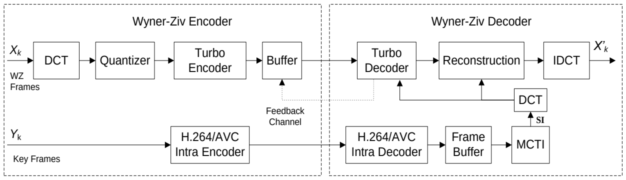


Figure 1: DVC architecture.

3. PROPOSED SI GENERATION SCHEME

The proposed SI generation scheme is illustrated in Figure 2. Firstly, the MCTI with spatial motion smoothing from [3] is used to compute motion vectors and to estimate the Initial Side Information (ISI) for the frame being decoded. Based on the ISI, the WZ decoder is first applied to generate a partially decoded WZ frame. The latter is used to detect suspicious motion vectors. These suspicious motion vectors are then refined by bi-directional motion estimation and smoothed using a spatial smoothing filter. A new matching criterion between the partially decoded WZ frame and the reference key frames is used. Then, optimal motion compensation mode selection is also conducted to further improve the SI. The final SI is constructed using motion compensation based on the refined motion vectors. Finally, the WZ decoder is run again to get the final decoded frame.

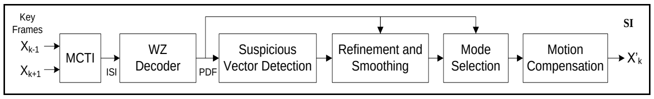


Figure 2: Proposed SI generation procedure.

3.1. Partially decoded WZ frame

One of the main novelties of the proposed approach is to improve the quality of the SI using the partially decoded WZ frame. Common MCTI techniques use only the previous and next key frames to generate the SI. However, once decoded, the WZ frame contains additional information from WZ bits. Therefore, by using this partially decoded WZ frame, spatio-temporal correlations between this frame and the key frames can be exploited to improve the quality of SI.

In this paper, the MCTI with motion vector filtering from [3] is first used to generate the ISI. Then the WZ decoder is run to generate the partially decoded WZ frame based on the ISI. The partially decoded frame is exploited in the following steps to improve SI, including suspicious vector detection, motion vector refinement and smoothing, and motion compensation mode selection.

3.2. Matching criterion

To exploit the spatio-temporal correlations based on the partially decoded WZ frame, a new matching criterion is used to evaluate the error in motion estimation. The proposed matching criterion is based on both the Mean Absolute Difference (*MAD*) and the Boundary Absolute Difference (*BAD*) [11]. The matching distortion (D_{ST}) between two frames F_1 and F_2 is defined as follows:

$$\begin{aligned} D_{ST}(F_1,F_2,MV) &= \alpha BAD(F_1,F_2,MV) \\ &\quad +(1-\alpha)MAD(F_1,F_2,MV) \end{aligned} \qquad (1)$$

where α is a weighting factor, and MV the estimated motion vector.

3.3. Suspicious vector detection

Generally, for most sequences, the majority of motion vectors estimated are close to the true motion. However, erroneous vectors may result in serious block artifacts if they are directly used in frame interpolation. In this paper, a threshold T is established to define the candidate blocks for further refinement based on the matching criterion D_{ST}. If an estimated MV satisfies the criteria defined in Eq.(2), it is considered to be a good estimation; otherwise, it is identified as a suspicious vector and will be further processed.

$$D_{ST}(F_{n-1},F_n^{'},MV)+D_{ST}(F_{n+1},F_n^{'},MV)<T \qquad (2)$$

where $F_{n\text{-}1}$ and F_{n+1} are the previous and next key frame, respectively, and $F_n^{'}$ the partially decoded WZ frame.

3.4. Motion vector refinement and smoothing

The spatio-temporal correlations between the partially decoded WZ frame and the key frames are exploited to refine and smooth the estimated motion vectors. More specifically, the motion vectors are re-estimated by bi-directional motion estimation using the matching criterion of Eq.(1) and partially decoded WZ frame. They are then filtered using motion vector spatial smoothing. This process generates a new estimation of the motion vector for the block to be interpolated.

It is observed that the motion vectors have sometimes low spatial coherence. A spatial motion smoothing filter is therefore used, similar to [2], but with the matching criterion of Eq.(1) and the partially decoded WZ frame. More precisely, a weighted median vector filter is used to maintain the motion field spatial coherence. This filter is adjusted by a set of weights controlling the smoothing strength. The weighted median vector filter is defined as

$$MV_F=\underset{MV_i}{\arg\min}\sum_{j=1}^{Num}w_j\left\|MV_i-MV_j\right\|_L,\ i\in[1,Num] \qquad (3)$$

where MV_1, …, MV_{Num} are the motion vectors of the corresponding nearest 8-neighbour blocks. MV_F is the motion vector output of the weighted vector median filter, which is chosen in order to minimize the sum of distances to the other Num-1 vectors.

The weights w_1, …, w_{Num} are calculated based on the new matching criterion and the partially decoded WZ frame

$$w_j=\frac{D_{ST}(F_{n-1},F_n^{'},MV_c)+D_{ST}(F_{n+1},F_n^{'},MV_c)}{D_{ST}(F_{n-1},F_n^{'},MV_j)+D_{ST}(F_{n+1},F_n^{'},MV_j)} \qquad (4)$$

where MV_c is the re-estimated vector for the block to be interpolated.

3.5. Optimal motion compensation mode selection

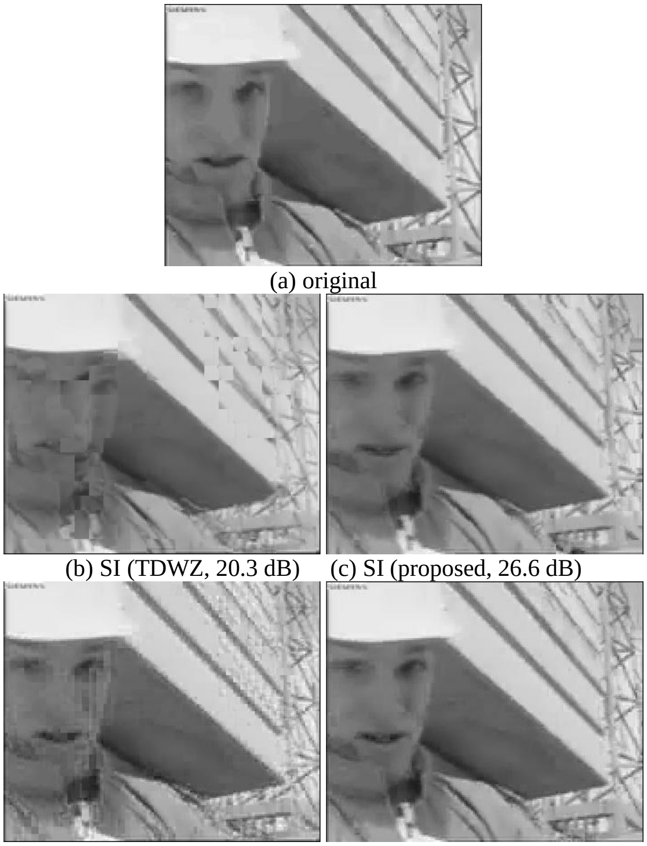
The objective of this step is to generate an optimal motion compensated estimate. In the DVC proposed in [3], while bi-directional prediction is shown to be effective, it is limited to motion-compensated average of the previous and the next key frames.

Based on the partially decoded WZ frame, the most similar block to the current block could be selected from three sources: the previous frame, the next frame, the bi-directional motion-compensated average of the previous and the next frame. Among these modes, the decision is performed according to the matching criterion defined in Eq.(1), and the one with the minimum matching error is retained.

Based on the refined motion vectors and the selected interpolation mode, motion compensation is applied to generate the final SI. While more iterations of decoding and SI generation could be used, our experiments show that this does not provide a significant improvement.

4. RESULTS AND DISCUSSIONS

The TDWZ DVC codec proposed in [3] is used in our experiments and only luminance data is coded. The video sequences *Foreman*, *Soccer* and *Hallmonitor*, are used in QCIF format and at 15 fps. The DVC codec is run for the first 149 frames. Eight RD points are computed per sequence. The results are compared to the SI generation in the TDWZ codec (TDWZ) [3]. The weight α in Eq.(1) and the threshold T in Eq.(2) are empirically set to 0.3 and 10, respectively.


(a) original

(b) SI (TDWZ, 20.3 dB) (c) SI (proposed, 26.6 dB)

(d) decoded (TDWZ, 27.5dB) (e) decoded (proposed, 29.8dB)

Figure 3: Visual result comparisons.

Figure 3 shows the visual results of SI and decoded frames for *Foreman*. The face and the building in the SI generated by the TDWZ contain block artifacts (Figure 3b). On the contrary, the SI generated by the proposed method (Figure 3c) is much better. The improvement in the SI also results in a better quality of the decoded WZ frame (2.3 dB improvement). There are much fewer block artifacts on the face and building in the decoded frame (Figure 3e) when compared to the proposed method by the TDWZ (Figure 3d).

Figure 4 shows the SI quality for *Foreman*. The proposed algorithm achieves up to 6.7 dB, and an average of 2.4 dB improvement, when compared to the SI in the TDWZ. The PSNR values of the decoded WZ frames are shown in Figure 5. Compared to the TDWZ, the proposed method achieves up to 2.2 dB, and an average of 0.6 dB improvement.

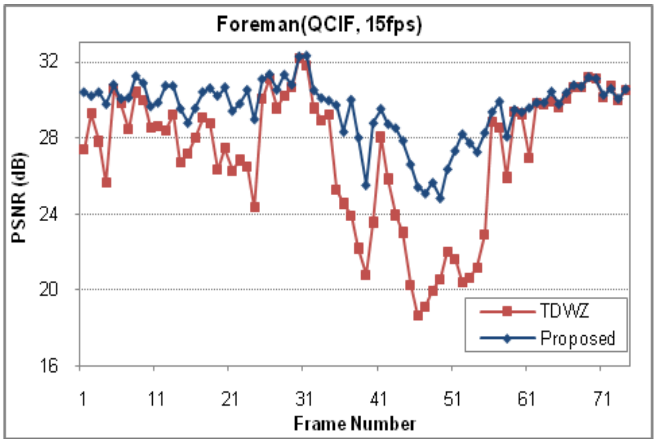


Figure 4: PSNR of SI for *Foreman* frames

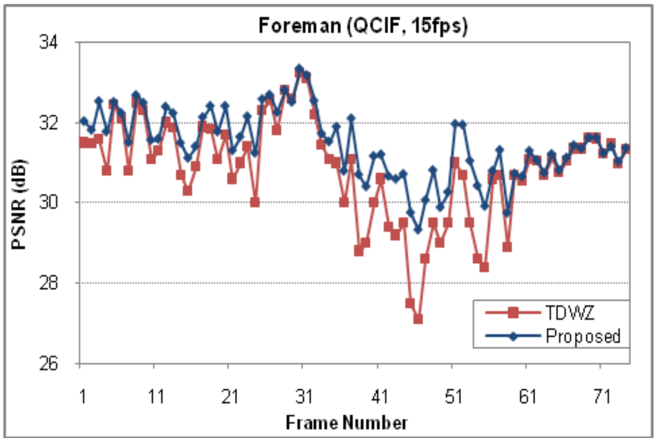


Figure 5: PSNR of decoded *Foreman* frames.

The RD performance of the proposed method for all the sequences is shown in Figure 6, Figure 7, and Figure 8, respectively. RD performance improvements over the TDWZ are observed for all sequences. For *Soccer* (Figure 7), the proposed method significantly improves the objective quality (up to 1.0 dB at QP6) with respect to TDWZ. When compared to H.264/AVC Intra, the performance is still inferior but the gap is brought down to a maximum of 2 dB, while it was around 3 dB for the TDWZ. For video with simple motion such as *Hallmonitor* (Figure 8), the performance of the proposed method is only slightly better than the TDWZ, which is already around 3 dB superior to H.264/AVC Intra. For *Foreman*, the introduced SI improves the performance over H.264/AVC Intra for low bitrates (Figure 6).

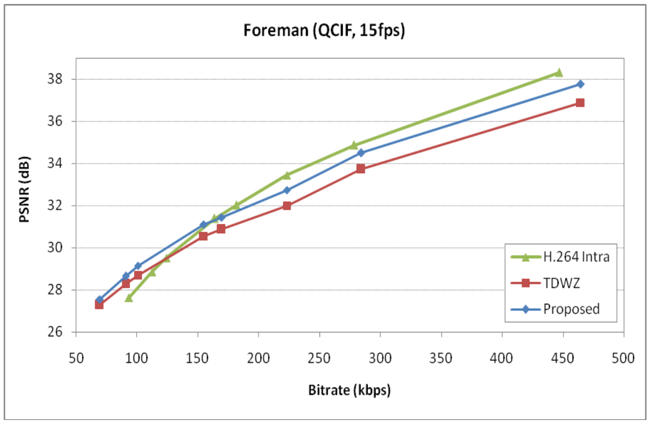


Figure 6: RD performance for sequence *Foreman*.

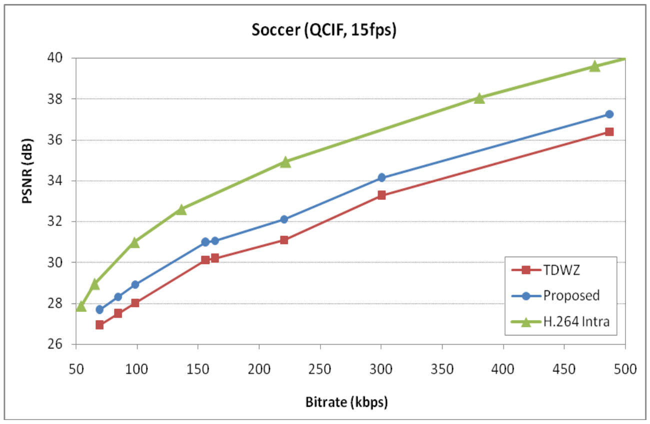


Figure 7: RD performance for sequence *Soccer*.

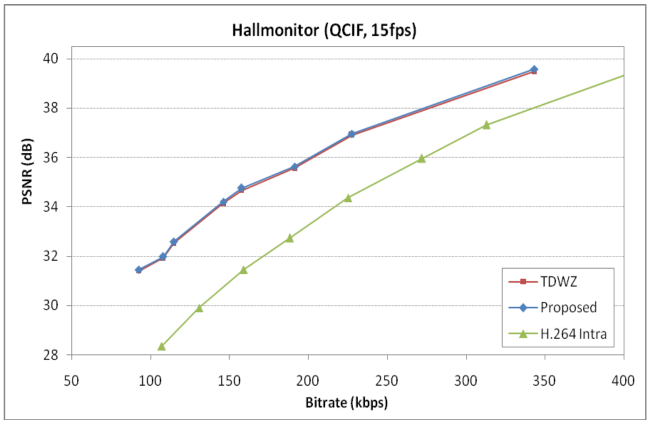


Figure 8: RD performance for sequence *Hallmonitor*.

5. CONCLUSIONS

A new side information generation scheme is proposed to improve the performance of DVC. The use of the partially decoded WZ frame improves the performance of SI generation by motion vector refinement, smoothing, and optimal compensation mode selection. Simulation results show a significant improvement of performance has been achieved by the proposed approach, compared to state-of-the-art DVC.

ACKNOWLEDGEMENT

This work was partially supported by the European Network of Excellence VISNET II (http://www.visnet-noe.org) IST Contract 1-038398, funded under the European Commission IST 6th Framework Program.

REFERENCES

[1] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed Video Coding”, *Procedings of the IEEE*, Vol. 93, No. 1, Jan. 2005, pp. 71-83.

[2] J. Ascenso, C. Brites, and F. Pereira, “Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding”, *EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak, July 2005.

[3] C. Brites, J. Ascenso, F. Pereira, “Improving Transform Domain Wyner-Ziv Video Coding Performance”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 2006.

[4] L. Wei, Y. Zhao, A. Wang, “Improved Side-Information in Distributed Video Coding”, *International Conference on Innovative Computing, Information and Control*, Beijing, China, Aug. 2006.

[5] R. Puri and K. Ramchandran, “PRISM: A ‘Reversed’ Multimedia Coding Paradigm”, *IEEE International Conference on Image Processing*, Barcelona, Spain, 2003.

[6] A. Aaron, S. Rane, and B. Girod, “Wyner-ziv Video Coding with Hash-based Motion Compensation at the Receiver”, *IEEE International Conference on Image Processing*, Singapore, Oct. 2004.

[7] 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*, Sardinia, Italy, Sept. 2005.

[8] X. Artigas and L. Torres, “Iterative Generation of Motion-compensated Side Information for Distributed Video Coding”, *IEEE International Conference on Image Processing*, Genova, Italy, Sept. 2005.

[9] W.A.R.J. Weerakkody, W.A.C. Fernando, J.L. Mart ńez, P.Cuenca, and F.Quiles, “An Iterative Refinement Technique for Side Information Generation in DVC”, *IEEE International Conference on Multimedia and Expo*, July 2007.

[10] T.Wiegand, G. J. Sullivan, G. Bj ǿntegaard, and A. Luthra, “Overview of the H.264/AVC Video Coding Standard”, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, July 2003, pp. 560-576.

[11] Y. Chen, O. C. Au, C.-W. Ho, and J. Zhou, “Spatio-Temporal Boundary Matching Algorithm for Temporal Error Concealment”, *IEEE International Symposium on Circuits and Systems*, Greece, May, 2006.