Journal of Engineering Science and Technology
Vol. 6, No. 5 (2011) 542 - 550

TRANSFORM DOMAIN SLICE BASED DISTRIBUTED VIDEO CODING

A. ELAMIN*, VARUN JEOTI, SAMIR BELHOUARI

Electrical Eng., Uinevrsiti Teknologi PETRONAS, Perak, Malaysia
*Corresponding Author: eaea33@itu.dk

Abstract

Distributed video coding depends heavily on the virtual channel model. Due to the limitations of the side information estimation one stationary model does not properly describe the virtual channel. In this work the correlation noise is modelled per slice to obtain location-specific correlation noise model. The resulting delay from the lengthy Slepian-Wolf (SW) codec input is also reduced by reducing the length of SW codec input. The proposed solution does not impose any extra complexity, it utilizes the existing resources. The results presented here support the proposed algorithm.

Keywords: Video Coding, Slepian-Wolf, Transfer Domain, Noise Model.

1. Introduction

Today's digital video coding paradigm represented by the ITU-T and MPEG standards mainly relies on a hybrid of block-based transform and inter-frame predictive coding approaches. In this coding framework, the encoder architecture has the task to exploit both the temporal and spatial redundancies present in the video sequence, which is a complex process and it requires a noticeable amount of resources (Power and memory), while the decoder remains a pure executer of the encoder "instructions". As a result, all standard video encoders have a much higher computational complexity than the decoder (typically five to ten times more complex) [1, 2]. Recently new emerging applications such as wireless low-power surveillance and multimedia sensor networks, wireless PC cameras and mobile camera phones have very different requirements than those of traditional video delivery systems. For some applications notably when there is a high number of encoders and only one decoder, e.g., surveillance, low cost encoder

devices are necessary. It is essential to have a low-power and low-complexity encoder device, possibly at the expense of a higher complexity decoder.

The low-complexity encoding could be achieved by moving some of the encoder tasks to the decoder part, specially the most complex motion estimation process. This useful hint is the consequence of information-theoretic principles established in the 1970s by Slepian and Wolf for distributed lossless coding [3], and by Wyner and Ziv for lossy coding [4] with some side information made available at decoder. Schemes that are based on these theorems are generally referred to as distributed source coding algorithms. The attractive idea of distributed source coding is that, in the case of joint decoding, two correlated sources X and Y can be compressed separately without the knowledge of the other source. It can still attain the same compression as if the other source was known. For the specific case where Y is available at the decoder and X and Y are jointly Gaussian, Wyner [4] proved that by using channel coding at the encoder the compression can still be achieved without Y being known at the encoder, since the availability of a correlated source is only necessary at the decoder. Designing practical video codec utilizing these principles in distributed video coding is still in its infancy, and researchers all over the world are eager to explore it and propose video coding systems.

One of the first practical WZ video coding solutions has been developed at Stanford University [1, 2, 5] works at the frame level. In this work it will be referred to by frame based solution. This solution has become the most popular WZ video codec design in the literature. The basic idea of this WZ video coding architecture is that the decoder, based on some previously and conventionally transmitted frames, the so-called *key frames*, creates the so-called SI which works as estimates for the other frames to code, the so-called *WZ frames*. The WZ frames are then encoded using a channel coding approach, e.g., with turbo codes or Low-Density Parity-Check (LDPC) codes [5, 6], to correct the 'estimation' errors in the corresponding decoder estimated side information frames. In this case, the encoding is performed assuming there is (high) correlation between the original WZ frames to encode and their associated SI frames at the decoder; higher correlation leads to more efficient encoding process.

In this paper, section two reviews the *state-of-the-art* transform domain WZ codec and presents the problems and issues related to the state-of-the-art the section also presents some related works that address the identified issues. Section three introduces the proposed slice based transform domain DVC. The slice composing method is presented in section four. The experimental work and its results are presented in section five. The current work and the future work are concluded in section six.

2. Literature Review

Figure 1 shows the state-of-art transform domain WZ video coding as presented in [1, 2, 7, 9]. The overall coding architecture works as follows: the video sequence is divided into WZ frames and key frames; the key frames are H263+ intra coded. Over each WZ frames a 4-by-4 DCT is applied. The DCT coefficients of the entire frame are grouped together in DCT bands. Each band is uniformly quantized and bitplanes are extracted and sent to the turbo encoder. The

turbo encoding process for a given DCT band starts with the most significant bitplane. Only a fraction of the parity information generated by the turbo encoder for each bitplane is sent to the decoder. The decoder the frame interpolation module is used to generate the side information Y_t; an estimate of the X_t frame, based on the previously decoded frames X_{t+1} and X_{t-1}. For a GOP length 2, these two frames are the key frames. A 4-by-4 DCT is then carried out over Y_t in order to obtain Y_t^{DCT}, an estimate of X_t^{DCT}. The residual statistics between corresponding X_t^{DCT} and Y_t^{DCT} is assumed to be modeled by a Laplacian distribution. The Laplacian parameters are estimated online for each DCT coefficient based on the residual between the frames X_{t+1}^{DCT} and X_{t-1}^{DCT} after motion compensation. Once the Y_t^{DCT} and the residual statistics for a given DCT band are know the decoded quantized symbol stream associated to that DCT band can be obtained through an iterative turbo decoding procedure. After turbo decoding the most significant bitplane of the DCT the turbo decoder proceeds in analogous way to the remaining bitplanes associated to that band. Once all the bitplane arrays of a given DCT band are turbo decoded the turbo decoder starts decoding the next DCT band. This procedure is repeated until all the DCT bands for which WZ bits are transmitted are turbo decoded. After turbo decoding the bitplanes associated to a given DCT band, these bitplanes are grouped together to form the decoded quantized symbol stream associated to that band. Once all the decoded quantized symbols are obtained it is possible to reconstruct the matrix of DCT coefficients $\hat{X}_t^{DCT}$. For some bands no WZ bits are transmitted; at the decoder those bands are replaced by the corresponding DCT bands of SI Y_t^{DCT}. The remaining DCT bands are obtained using the reconstruction function which bounds the error between DCT coefficients of X_t^{DCT} and $\hat{X}_t^{DCT}$ (also known as reconstruction distortion) to the quantizer coarseness.

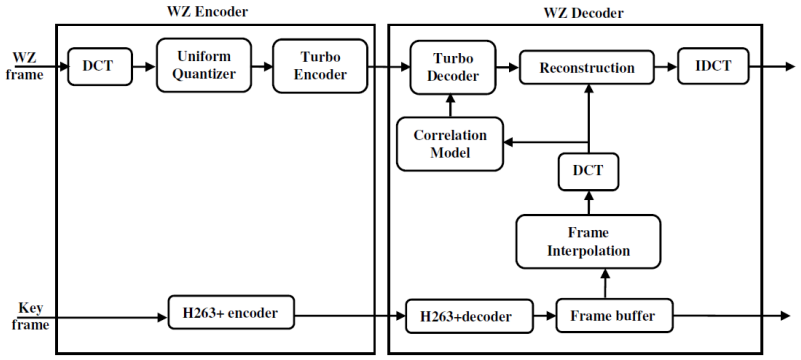

Fig. 1 State-of- art Transform Domain WZ Codec Architecture.

This section intends to investigate some issues in the *state-of-the-art* such as complexity of the decoding process and the accuracy of the correlation noise model and review the literature related to these issues addressed in this paper. The size of input block is assumed arbitrarily large because it represents all bits from the whole frame or one bit-plane at once [7-9]. Too large input block will produce

significant computation latency during the encoding and decoding process. In other words, the system will not be able to provide a timely WZ decoded output due to the vast Slepian-Wolf turbo coding and decoding delay. Frame based DVC systems make use of a rate-compatible-punctured codes (RCPT) codec as the module of the SW codec [10, 11] and the puncture window size is usually 8, resulting in 16 puncture code rates, which is referred to as rate control levels. The encoder "blindly" sends the parity bits according to the puncturing pattern determined by next lower coding rate. This causes the waste of transmission that some unnecessary parity bits are sent to help decoding some bits which are already correctly decoded. The decoding will still request more parity bits until enough parity bits are received. It can be seen that by dividing the WZ frame into several small parts for transmission, the transmission of parity bits is not aiming to decode the whole frame but a small part of the frame therefore reduces the possibility to transmit the unnecessary parity bits.

In traditional video coding, the Laplacian distribution is typically used to model the distribution of the motion-compensated residual DCT coefficients [12]. More accurate models can be found in literature, such as the generalized Gaussian distribution; however, the Laplacian distribution constitutes a good trade-off between model accuracy and complexity and, therefore, it is often chosen [7-9, 13]. For the same reason, most of the implementations of the frame based DVC assume stationary noise (Laplacian) and constant quality of the side information estimated at the decoder along the frame. In distributed video compression, the side information is actually estimated at the decoder side; the decoder has to make a prediction of the current video frame without actually knowing this frame. As a consequence there will be within the motion predicted frame co-exist regions where the motion estimation is successful (high correlation) and a region where the interpolation has (more or less) fails due to the occlusion. Occlusions create noise in motion estimate frame prediction with two important properties. Firstly, this noise is very location-specific; and always located at the edge of a moving object or the frame edge in the case of camera motion; Secondly, occlusion noise is hard to characterize [14]. Distributed source coding relies heavily on efficient error correcting codes; the performance of these codes depends greatly on the choice of the noise model that characterizes the virtual dependency channel [15]. It was concluded that the behavior of the virtual dependency channel is substantially more complicated than a simple BSC or AWGN channel model often assumed in communications systems that apply channel codes [16].

Several works has been proposed for frame-based in order to cope with above two drawbacks. In [18], in order to exploit the spatial variability of the correlation noise, the models' parameter is calculated at the block level; the block size considered is equal to the one used in the frame interpolation process (8-by- 8) in order to more easily match the frame interpolation errors; the resulting RD performance is outperformed the frame level based noise channel modeling. Non-stationary model to characterize the correlation noise is proposed in [14], where two models are used one to model the non-occluded regions and the other to model the occluded regions, the results shows that the performance of the SISO decoder, and therefore of the overall system, can be improved greatly by classifying the decoder-generated side-information into two (or more) reliability classes. It also shows that the channel model should be an accurate representation of the real behavior of the channel; otherwise the decoding performance will heavily degrade. The lengthy

block problem has also been addressed in [19, 20], by dividing the WZ frame X into M sub-images X_m, $m = \{1, 2, \ldots, M\}$each sub-image is independently encoded using a Turbo-code based Wyner-Ziv encoder. Therefore, the block size of the Turbo encoder decreases to $1/M$ of the frame size. In the proposed system [19], BAWZC-based DVC system, the puncturing rate is independently adjusted for each block and the BAWZC decoder has to perform to inform the WZC encoder which blocks needs more parity bits. The system proposed in [20] restricts the length of SW input to reduce the decoding complexity and the resulting delay. Both work assume the stationary noise and use the same correlation noise model to characterize the noise over all sub-images, which makes no point of the sub-image other than just reducing the lengthy block. They alleviate the problem of spreading the estimation errors all over the bitstream, by keeping the localized errors close to each other in the generated symbol stream. In the results of both works, it is observed that system performance is further enhanced.

3. Proposed DVC System

For a video sequence, the odd frames are the Key frames, and the even frames are the Wyner-Ziv frames. The Key frames can be intraframe encoded by using any conventional video codec and intraframe decoded at the decoder with the same codec. In the proposed DVC scheme, the Wyner-Ziv frames are intraframe encoded by using a slice Wyner-Ziv codec. However, they are interframe decoded by the proposed slice based decoder jointly with the side information which is generated from the corresponding key frames. The WZ frame is divided into slices based on a binary map generated by the decoder. The decoder composes the slice and generates binary map to help the encoder compose the corresponding slices from the original WZ frame. The decoder assesses the degree of success in the generation of side information frame blocks, the background blocks are grouped together to compose the first slice. The Occluded regions blocks grouped together to compose another slice; the remaining blocks are grouped together as another slice. The resulting slices will namely be the background slice, the simple motion slice and the occluded regions slice. The decoder generates the corresponding side information for each slice and transforms them into DCT domain. The DCT coefficients are grouped into sub bands, the correlation noise model for each sub band is estimated; the resulting model is more accurate than using the single model to describe the noise over DCT sub band formed to the entire frame and it is a location-specific model.

$$p\left(X_{slice}^{DCT} - Y_{slice}^{DCT}\right) = \frac{\alpha_{slice}^{DCT}}{2} e^{-\alpha_{slice}^{DCT} \times \left| X_{slice}^{DCT} - Y_{slice}^{DCT} \right|} \tag{1}$$

The feedback channel plays an important role in turbo-code based DVC systems [1, 2, 7, 19]. In the proposed system, the Slice based DVC decoder has to perform not only rate control but also the decoder uses the feedback channel to send the binary map to guide the encoder to compose the corresponding original slices. After WZ frame is divided into slices at the Slice based DVC encoder, a 4-by-4 DCT is applied to each slice, each sub band is then quantized using quantization scheme as proposed in [9], each quantized sub band represented as a set of bitplanes according to the number of the quantization levels of the sub band corresponding quantizer. The resulting binary sequence is then fed into a turbo encoder as a symbol stream.

After turbo encoding, all systematic bits are discarded and all parity bits are stored in a buffer. The parity bits are progressively requested by the SWZC decoder to perform the rate control process [1, 2, 7, 9].

4. Slice Composing

The decoder performs the blocks classification to slice the side information and help the encoder create the corresponding slices, to avoid add more complexity to the decoder the process of slicing the side information and generating the binary map is embedded within the process of frame interpolation. The motion estimation/compensation process is performed at the decoder in which the motion vectors forward/backward are estimated, the background slice is simply composed by zero forward/backward motion vectors blocks. The same information (forward/backward motion vectors) is used to detect the occluded regions blocks. The usual assumption in estimation of occlusions from two frames [21, 22], is excessive intensity matching (motion-compensated prediction) error observed; reference-frame pixels that disappear cannot be accurately matched in the target frame and thus induce significant errors. Let $F_1(x)$ denote intensity of the first frame of a sequence at spatial position x, and $F_2(x)$ – similar intensity in the second frame. If $d_f(x)$ denotes a forward motion (disparity) field anchored on the sampling grid of key frame #1 (reference) and pointing to the target key frame #2, while $d_b(x)$ denotes a backward motion field, then the corresponding motion-compensated prediction errors at x are:

$$\Delta_f|x| = F_1(x) - F_2\left(x + d_f(x)\right) \tag{2}$$

$$\Delta_b|x| = F_2(x) - F_1\left(x + d_f(x)\right) \tag{3}$$

The usual occlusion detection methods then declare a pixel in the reference frame as being occluded in the target frame if $\Delta_f > \varepsilon$ or frame #1 and $\Delta_b > \varepsilon$ for frame #2. Note that although newly exposed areas cannot be detected by this mechanism (pixels are not visible), effectively the occluded areas in frame #2 are in fact the newly-exposed areas for frame #1.

5. Experiments and Results

To test the performance of the proposed DVC scheme, the first 101 frames of the QCIF (144 x 176) Foreman and Hall Monitor sequences at 15 frames per second were coded. . The occlusion detection based on the photometry based method (equation 2, 3) is performed at the decoder during the side information creation, the threshold value for the occluded regions is obtained empirically and hardcoded as 253. The frame interpolation process is performed as in [5]. From Table 1, the length of the bitplane is reduced; as result the corrupted part of the original bitplane is isolated in different bitplanes (bitplanes of slice 2 and slice 3). Table 2 shows the correlation noise parameters for same sub bands in different slices, the different parameters correspond to the difference in the quality of the side information along the estimate frame. Figures 2 and 3 show the resulting rate-distortion performance of the proposed system for the Wyner-Ziv frames of the Foreman and Hall Monitor sequences, respectively.

Table 1. The Bitplane Length per Slices.

Video sequence	Slice 1 length of bitplane	Slice 2 length of bitplane	Slice 3 length of bitplane
Hall Monitor	1264	280	40
Foreman	824	412	348

Table 2. The Correlation Noise Model Parameter per Sub-band/Slice.

Foreman Sequence				Hall Monitor Sequence			
Slice Number	DC	AC1	AC2	Slice Number	DC	AC1	AC2
1	0.2267	0.5661	0.6705	1	0.2348	0.4318	0.4364
2	0.2301	0.4750	0.4629	2	0.2385	0.2627	0.3630
3	0.2090	0.3617	0.3608	3	0.2991	0.4329	0.3105

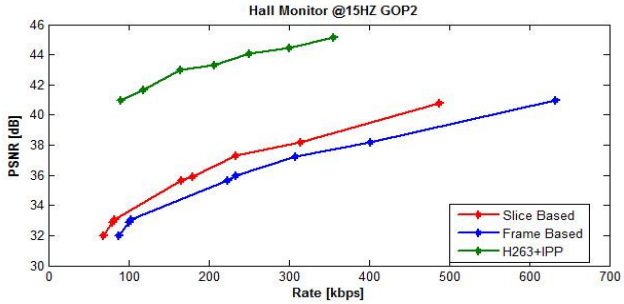

Fig. 2. PSNR Comparison of the Proposed System for the Hall Monitor Sequence.

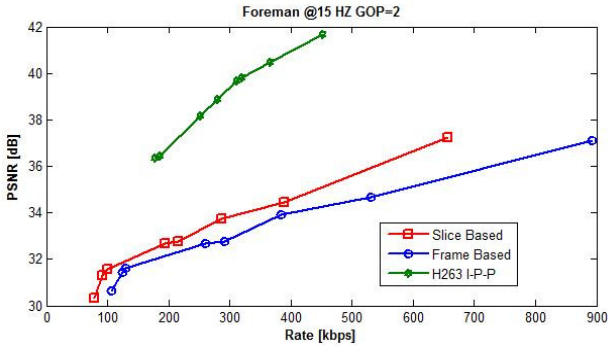

Fig. 3. PSNR Comparison of the Proposed System for the Foreman Sequence.

For comparison, Figs. 2 and 3 also show the corresponding rate distortion performance of the transform-domain DVC systems of [5] and of H.263 interframe coding (P frames). From Figs. 2 and 3, it can be seen that the proposed system (Slice based) RD performance is closer to H.263 interframe coding than existing DVC systems [5]. In addition, the proposed DVC system results in a 25% to 35% reduction in the average bitrate for the Foreman sequence, and in an

average bit-rate reduction of 21% to 30% for the Hall Monitor sequence, as compared to [5]. In the rate distortion performance the rate of the feedback is not considered as extra rate since it flows from the decoder to encoder. The same reconstruction method as in [5] is applied on both as the rate distortion figures show same quality PSNR for both solutions, with different rates.

6. Conclusions

The design of the slice structure optimizes usage of the existing resources. It is found that the utilization of slice structure in WZ video coding brings two advantages. On one hand, the input block size is reduced and non-stationary model properly characterizes the virtual channel. On the other hand, with slice structure, the encoder avoids sending unnecessary parity bits as a result the RD performance is improved as a whole. The gap between conventional interframe codec and the distributed video coding is reduced and further narrowing to this gap is possible by improving the side information estimation. The process of detecting the occluded regions does not increase the complexity of the decoder since it is performed within the frame interpolation process; communicating the slicing binary map utilizes the existing feedback channel; therefore the proposed slice based DVC does not impose extra complexity to the overall system. Future work is to devise advanced method to extract the background blocks and replace them with the key frames blocks.

References

1. Aaron, A.; Zhang, R.; and Girod, B. (2002). Wyner-Ziv coding of motion video. *The Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, 1, 240-244.
2. Aaron, A.; and Girod, B. (2004).Wyner-Ziv video coding with low-encoder complexity "Invited Paper". *In Proc. Picture Coding Symposium*, PCS-2004, San Francisco, CA.
3. Wyner, A.; and Ziv, J. (1976). The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information and Theory*, 22(1), 1-10.
4. Aaron, A.; and Girod, B. (2002). Compression with side information using turbo codes. *IEEE Conference on Data Compression*, 252-261.
5. Varodayan, D.; Aaron, A.; and Girod, B. (2005). Rate-adaptive distributed source coding using low-density parity-check codes. *Conference Record of the Thirty-ninth Asilomar Conference on Signals, Systems and Computers*, 1203-1207.
6. Rebollo-Monedero, D.; Aaron, A.; Girod, and B. (2003). Transforms for high-rate distributed source coding. *In Proceedings of 37th Asilomar Conference on Signals, Systems and Computers*, 1, 850-854.
7. Ascenso, J.; Brites, C.; and Pereira, F. (2005). Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding. In *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak Republic.

8. Aaron, A.; Rane, S.; Setton, E.; and Girod, B. (2004). Transform-domain Wyner-Ziv codec for video. *Proceedings of SPIE Visual Communications and Image Processing*, 520-528.
9. Bajcsy, J.; and Mitran, P. (2001). Coding for the Slepian-Wolf problem with turbo codes. *In IEEE Global Telecommunications Conference*, 2, 1400-1404.
10. Garcia-Frias, J. (2001). Compression of correlated binary sources using turbo codes. *In IEEE Communications Letters*, 5(10), 417-419.
11. Bellifemine, F.; Capellino, A.; Chimienti, A.; Picco, R.; and Ponti, R. (1992). Statistical analysis of the 2D-DCT coefficients of the differential signal for images. *Signal Process: Image Communication*, 4(6), 477-488.
12. Avudainayagam, A.; Shea, J.M.; and Wu, D. (2008). Hyper-Trellis decoding of pixel-domain Wyner-Ziv video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(5), 557–568.
13. Meyer, P.F.A.; Westerlaken, R.P.; Gunnewiek, R.K.; and Lagendijk, R.L. (2005). Distributed source coding of video with non-stationary side-information. *Proceedings of Conference on Visual Communications and Image Processing*, 857-866.
14. Westerlakena, R.P.; Gunnewiekb, R.K.; and Lagendijk, R.L. (2005). The role of the virtual channel in distributed source coding of video. *IEEE International Conference on Image Processing*, 1, I-581-4.
15. Westerlaken, R.P.; Borchert, S.; Gunnewiek, R.K.; and Lagendijk, R. (2006). Finding a near optimal dependency channel model for a ldpc-based Wyner-Ziv video compression scheme. *In 12th annual conference of the Advanced School for Computing and Imaging*, 546-463, Lommel, Belgium.
16. Borade, S.; Nakiboglu, B.; and Zheng, L. (2008). Some fundamental limits of unequal error protection. *IEEE International Symposium on In Information theory , ISIT 2008*, 2222-2226.
17. Brites, C.; and Pereira, F. (2008). Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding. *IEEE transactions on Circuit and Systems for Video Technology*, 18(9), 1177-1190.
18. Chien, W.J.; Karam, L.J.; and Abousleman, G.P. (2007). Block-adaptive Wyner-Ziv coding for transform-domain distributed video coding. *In 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing*, I-525 – I-528.
19. Xue, Z. (2009). *Research and developments of distributed video coding*. PhD Thesis, Brunel University, School of Engineering and Design.
20. Ince, S.; and Konrad, J. (2005). Geometery-based estimation of occlusion from video frame pairs. *In IEEE International Conference on Acoustics, Speech and Signal Processing*, 2, ii/933 - ii/936, Philadelphia, PA, USA.
21. Lim, K.P.; Das, A.; and Chong, M.N. (2002). Estimation of occlusion and dense motion fields in a bidirectional Bayesian framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 712-718.
22. Thoma, R.; and Bierling, M. (1989). Motion compensating interpolation considering covered and uncovered background. *Signal Processing: Image Communication*, 1, 191–212.