MOTION-BASED SIDE-INFORMATION GENERATION FOR A SCALABLE WYNER-ZIV VIDEO CODER

B. Macchiavello, R. L. de Queiroz

Universidade de Brasilia
Departameneto de Engenharia Eletrica
Brasilia, Brasil
bruno@image.unb.br, queiroz@ieee.org

D. Mukherjee

HP-Labs
Hewlett Packard Laboratories
Palo Alto, California, USA
debargha.mukherjee@hp.com

ABSTRACT

A motion-based side-information generation scheme with semi super-resolution for a scalable Wyner-Ziv coder framework is introduced. It is known that the performance of any Wyner-Ziv coder is heavily dependent on the efficiency of the side-information generation. We propose an iterative block based scheme to generate a semi super-resolution frame using the past and future reference frames which should be coded at full-resolution. To enable this side-information generation the framework should allow for low encoding complexity, reducing the spatial resolution only in the non-reference frames. The enhancement layer is produced using a residual frame of the reduced resolution encoded frame. The decoder first decodes the low resolution base layer and iteratively generates the side-information, along with channel decoding, to obtain a higher quality version of the decoded frame. Results of the implementation of the framework using the motion-based side-information in the H.263+ and H.264 standards are presented.

Index Terms— Distributed video coding, side-information generation, reversed-complexity, scalable video coding.

1. INTRODUCTION

The paradigm of distributed source coding (DSC), which has its roots in the theory of coding correlated sources developed by Slepian and Wolf [1] for the lossless case and Wyner and Ziv [2] for the lossy case, has recently became the focus of different kinds of video coding schemes [3]–[9]. A good review on DSC applied to video coding can be found in [10]. DSC is promising in creating reversed complexity codecs for power constrained (hand-held) devices. Unlike regular broadcast oriented video codecs with high encoding complexity and low decoding complexity, reversed complexity codecs have low encoding complexity but high decoding complexity. In prior work [4], [5] the reversed complexity scenario is addressed, proposing encoding methods requiring no motion estimation at the encoder. Other work [8] addresses spatio-temporal scalability using distributed coding, also enabling complexity reduction within its framework. A layered Wyner-Ziv (WZ) codec with similar results to the MPEG-4/H.26L FGS has been presented [6].

The framework used here [11] can be implemented as an optional coding mode in any existing video codec standards [12], [13]. The encoding complexity is reduced by lower resolution encoding, while the residue is WZ encoded by cosets. We propose a side-information (SI) generation method to be used in this framework. Other works, have also explore spatial reduction [14] and used the low resolution encoded frames to perform motion estimation and compensation to generate the SI. Here, the main goal is to use a mixed resolution approach that can potentially yield a better rate-distortion performance by enabling the proposed method of SI generation. Since, the reference frames are encoded at full resolution the motion-based side-information will be at what we call semi super-resolution. In related works [15],[16], a rate distortion analysis of motion side estimators and a study of using subpixel and multireference motion search methods to improve WZ efficiency has been addressed. In our method the decoded WZ frame is generated iteratively. The number of iterations is important, because the more iterations the more complex the decoder is and the better the SI generation. Some other studies [17] also used mode-aided motion compensation to generate SI. There, the key frames are coded with H.264/AVC intra coding mode. At the WZ frames for each block they evaluate a amount of candidate blocks in order to find the best candidate as the SI. Then, a error correction code (ECC) decoding is applied, no feedback channel is required in the codec. Here, also, no feedback is required, and the cosets are used for error correction. However, the used framework allows for normal coding of the key frames (I,P or B).

2. THE FRAMEWORK

The scalable framework is based on non-reference WZ frames (NRWZ) [11]. The complexity reduction is applied only to these frames. Since the reference frames are coded as I, P or B regular frames, there will not be any drifting error.

2.1. NRWZ Encoder Architecture

The NRWZ coding mode, shown in Fig. 1, will first decimate all the frames in the reference lists as well as the current frame by a factor of $2^n \times 2^n$. Then, the low-resolution (LR) current frame is encoded. Note that the syntax element object list for reference frames are also transformed into a form that is appropriate for reduced resolution encoding. The enhancement layer, or WZ layer, is computed as the difference between the full resolution original frame and the interpolated reconstruction form of the LR coded frame.

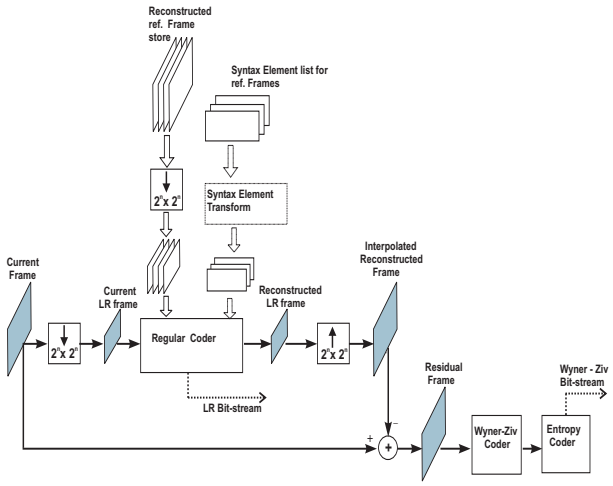


Fig. 1. Architecture for the NRWZ coding mode

2.2. NRWZ Decoder Architecture

The decoder architecture is presented in Fig.2. First, the LR image is decoded and interpolated. The process of enhancement begins with the generation of the SI. The interpolated decoded frame and the reference frames are used to create what we call the semi super-resolution frame, because except for the current frame, the other frames used are already at higher resolution. Once the semi super-resolution frame is generated, it is subtracted by the interpolated LR decoded frame. The resulting residual frame is the actual SI frame to be used for channel decoding. The channel decoder decodes the WZ bit-stream layer with the SI residual frame acting as a noisy version of the original residual frame transmitted.

2.3. Wyner-Ziv Coder

The model assumes that the coefficients are Laplacian distributed with standard deviation σ_x. First, the block transform of the residual frame are quantized. That is, if x is a coefficient of the residual image, its quantized value will be x_q based on a quantization step-size QP, possibly with a dead zone. The WZ codec does not transmit the residual coefficient

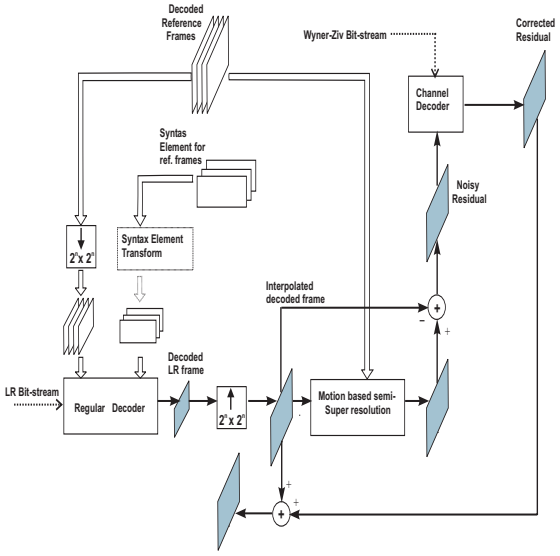


Fig. 2. Architecture for the NRWZ decoder

but cosets. We use cosets of odd modulus, with indices centered at 0. The transmitted coset index $c = \psi(x_q, M)$ of order M is computed as:

$$\psi(x_q, M) = \begin{cases} (x_q) - M\lfloor x_q/M \rfloor, & (x_q) - M\lfloor x_q/M \rfloor < M/2 \\ (x_q) - M\lfloor x_q/M \rfloor - M, & (x_q) - M\lfloor x_q/M \rfloor \geq M/2 \end{cases} \quad (1)$$

If quantization bin x_q corresponds to $[x_l(q), x_h(q)]$. Then, the probability of the bin $x_q \in \Psi_q$, and the probability of a coset index $c \in \Psi_c$ are given by:

$$p(x_q) = \int_{x_l(q)}^{x_h(q)} f_x(x)dx \quad (2)$$

$$p(c) = \sum_{x_q \in \Psi_q, \psi(x_q, M)=c} p(x_q), \quad (3)$$

where $f_x(x)$ is the pdf of x. Since the distribution $p(c)$ is symmetric for odd M, has zero as its mode, and decays with increasing magnitude, the entropy coder that already exists in the regular codec can be reused for c. However, the architecture allows to use a specifically designed entropy coder. In the WZ framework, the parameters QP and M are different for each coefficient x_{ij}. They are chosen based on an estimate of the noise level between the SI block and the original. In this work, a combination of the number of bits spent to code the corresponding residual block in the LR layer and an edge activity measure are used to estimate the noise level. In our model if y is the SI available only for decoding, then $y = x + z$, where z is Gaussian distributed with variance σ_z. At the decoder, the minimum MSE reconstruction function $\hat{x}_{yc}(y, c)$ based on y and received as c, is given by

$$\hat{x}_{yc}(y, c) = \frac{\sum_{x_q \in \Psi_q, \psi(x_q, M)=c} \int_{x_l(q)}^{x_h(q)} x f_{x|y}(x, y)dx}{\sum_{x_q \in \Psi_q, \psi(x_q, M)=c} \int_{x_l(q)}^{x_h(q)} f_{x|y}(x, y)dx}, \quad (4)$$

In decoding, the same information (bit rate at the LR layer and edge activity) not only yield the QP and M parameters applied during encoding, but they also map the model parameters $\{\sigma_x, \sigma_z\}$ (estimated offline) to be used in the computation above. However, the exact computation of Eq. 4 is difficult. The use of approximations on pre-computed tables can be useful. Note that y belongs to SI generated in the decoder which is unquantized and that a maximum value for QP and a minimum value for M were established.

3. SIDE-INFORMATION GENERATION

Our proposed method computes iteratively the semi super-resolution frame followed by channel decoding in multiple passes. Let the interpolated LR reconstructed frame be called F_0. Let $SS(F, FS)$ denote the semi super-resolution operation to yield a higher resolution version F^{HR} of F based on the frames stored FS. Also, let $CD(R, b_{WZ})$ denote the channel decoding operation yielding a corrected version of the residual frame based on noisy version R using the WZ layer bit-stream b_{WZ}. Then, iterative decoding comprises the following steps for $i = 0, 1, 2, , N$:

$$F_i^{HR} = SS(F_i, FS) \quad (5)$$

$$F_{i+1} = CD(F_i^{HR} - F_0, b_{WZ}) + F_0. \quad (6)$$

In order to generate the semi super-resolution, first the past and future references frames in the frame store are low-pass filtered. Next, a block-matching algorithm is applied using the interpolated decoded frame. The block-matching algorithm works as follows. For every 8×8 block in the interpolated decoded frame, the best sub-pixel motion vectors in the past and future filtered frames are computed. If the corresponding best predictor blocks are denoted as B_p and B_f in the past and future filtered frames respectively, several predictor candidates are calculated as $\alpha B_p + (1 - \alpha)B_f$, where α assumes values between 0 and 1 (for our implementation we use $\alpha \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$). If the NRWZ frame is used as a P frame, only the B_p predictors are used. Then, if the sum of absolute differences (SAD) of the best predictor of a particular macroblock is lower than a threshold T, then that macroblock is replaced by its best predictor, but with the compensation now conducted from unfiltered past and future frames. Otherwise, no action is taken. When all blocks in F_i have been processed, the updated frame is referred to as F_i^{HR} in Eq. 5 and 6.

From iteration to iteration three things are changed: the strength of the low-pass filter is gradually reduced, the grid of the block matching is offset and the threshold T is also gradually reduced. In practice, the low pass filtering operation for the reference frames is eliminated after one or two iterations as the frame becomes more and more accurate. The grid for block matching is offset from iteration to iteration to smooth out the blockiness and to add spatial coherence. For example, the shifts used in four passes can be (0, 0), (4, 0), (0, 4) and (4, 4). And, finally, the threshold T is also gradually reduced so that fewer blocks are changed in later iterations. This process is shown in Fig. 3.

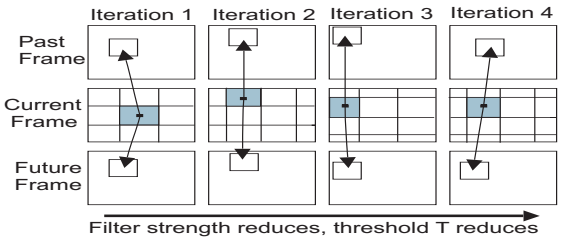
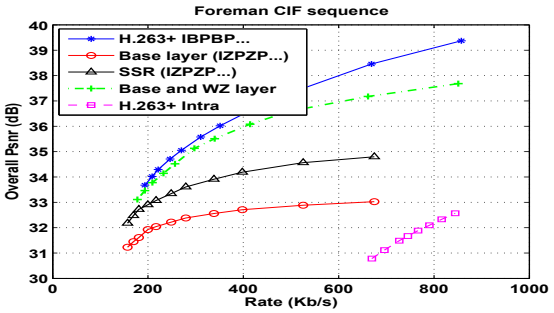
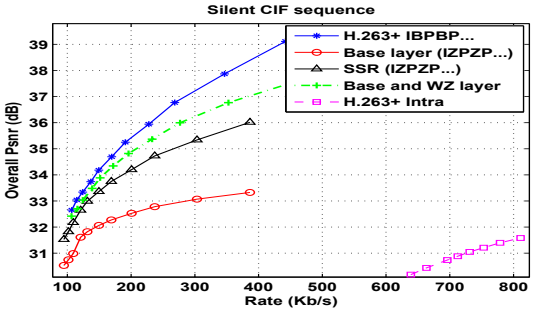


Fig. 3. Semi super-resolution for non-reference WZ frames. Filtered frames are used for motion estimation, but unfiltered frames are used for motion compensation.



(a)



(b)

Fig. 4. R-D results. Comparing H.263+ in IBPBP.. mode, H.263+ in intra mode, the LR base layer, the reversed complexity mode (base and WZ layer) after 4 iterations and the LR layer with the SSR frame after 1 iteration: (a) section of Foreman sequence; (b) section of Silent sequence

4. RESULTS

The framework was implemented using as the regular coder two different video coding standards: H.263+ and H.264/AVC. In both cases, the B-frames of the regular codecs are replaced by NRWZ-B frames. The decimation factor used was 2×2 (quarter resolution). In the H.264 version the CAVLC entropy

coder was used. The coding performance of a reversed complexity codec operating in $IZPZPZ...$ mode with Z frames indicating NRWZ-B frames, is compared against regular codec operating in $IBPBPB...$ mode. If we assume the encoding complexity of a LR frame to be roughly $1/4$ the complexity of a full resolution frame, then ignoring the additional complexity due to Wyner-Ziv coding, the reversed-complexity codec will work at half the complexity of the regular codec. As can be observed in Fig. 4, the reversed-complexity mode applied to the H.263+ standard, using the base and WZ layer after four iterations, at some rates has a performance less than 1 dB lower than the regular H.263+. This is mainly due to the accurate SI generation. The results of the base layer along with the semi super-resolution frame (SSR) without channel decoding and after only one iteration are also presented. For both test sequences it can be seen that the semi super-resolution frame increases the quality in comparison with the base layer.

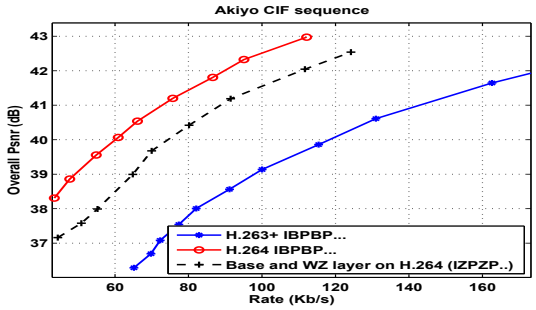


Fig. 5. R-D results on H.264. Comparing H.264 and H.263+ in IBPBP mode with the reversed complexity mode for H.264 after 4 iterations for Akiyo CIF sequence.

Finally, in Fig. 5, the performance of the H.264/AVC version is presented. As can be seen, the reversed complexity mode, operating with four iterations to create the SI, works better than the regular H.263+ codec. However, the gap between the regular coding mode and the reversed-complexity mode is larger than the gap in the H.263+ version. This, apparently, is because we need to improve the interpolation process of the LR frame for the SI generation. Since the H.264 coder generates a better quality high resolution frame, the differences between the decimated frames and the high resolution frames are more significant.

5. CONCLUSION

In this paper, we have introduced a motion-base semi super-resolution side-information generation mode in a Wyner-Ziv scalable framework. The generation of the side-information is done iteratively and allows better results in the Wyner-Ziv decoder, since this component is critical to the robustness of the distributed video codec. The results have shown the importance of an accurate side-information generation, because it allows to obtain a competitive reversed-complexity coding mode. It must also be mentioned that while it is unlikely that a different channel coding technique would lead to any sizeable gains, there could be improvements that come from novel channel codes designed specifically for non-uniformly distributed sources.

6. REFERENCES

[1] J. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans on Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul 1973.

[2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans on Inf. Theory*, vol. 2, no. 1, pp. 1–10, Jan 1976.

[3] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (discus): design and construction," *In Proc. IEEE Data Comp.n Conf.*, pp. 158–167, 1999.

[4] A. Aaron, R. Zhang, and B. Girod, "Transform-domain Wyner-Ziv codec for video," *In Proc. SPIE Visual Com. and Img. Proc.*, vol. 5308, pp. 520–528, January 2004.

[5] R. Puri and K. Ramchandram, "Prism: A new robust video coding architecture based on distributed compression principles," *Allerton Conf. on Com., Cont. and Comp.*, 2002.

[6] Q. Xu and Z. Xiong, "Layered WynerZiv video coding," *IEEE Trans on Img. Proc.*, vol. 15, no. 12, pp. 3791–3809, Dec 2006.

[7] H. Wang, N. M. Cheung, and A. Ortega, "A framework for adaptive scalable video coding using Wyner-Ziv techniques," *EURASIP Journal on Applied Signal Proc.*, pp. 1–18, 2006.

[8] M. Tagliasacchi, A. Majumdar, and K. Ramchandram, "A distributed-source-coding based robust spatio-temporal scalable vedo codec," *Picture Coding Symposium*, December 2004.

[9] X. Wang and M. T. Orchard, "Desing of trellis codes for source coding with side infotmation at the decoder," *In Proc. of IEEE Data Compression Conf.*, pp. 361–370, 2001.

[10] B. Girod, A.M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, Jan 2005.

[11] D. Mukherjee, "A robust reversed complexity Wyner-Ziv video codec introducing sign-modulated codes," *HP Labs Tech. Report, HPL-2006-80*, May 2006.

[12] T. Weigand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/avc video coding standard," 2003.

[13] G. Cote, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit-rates," *IEEE Trans. Circuits Syst. Video Technology*, vol. 8, no. 7, pp. 849–866, Nov. 1998.

[14] M. Wu, G. Hua, and C. W. Chen, "Syndrome-based lightweight video coding for mobile wireless application," *In Proc. Int. Conf. on Multimedia and Expo*, pp. 2013–2016, 2006.

[15] Z. Li and E. J. Delp, "Wyner-Ziv video side estimator: conventional motion search methods revisited," *In Proc. IEEE Int. Conf. on Img. Proc.*, pp. 825–828, 2005.

[16] Z. Li, L. Liu, and E. J. Delp, "Rate Distortion Analysis of Motion Side Estimation in WynerZiv Video Coding," *IEEE Trans on Img. Proc.*, vol. 16, no. 1, pp. 98–113, Jan 2007.

[17] L. W. Kang and C. S. Lu, "Wyner-Ziv video coding with coding mode-aided motion compensation," *In Proc. IEEE Int. Conf. on Img. Proc.*, pp. 237–240, 2006.