# Distributed Video Coding Using Wavelet

Xun Guo*[1], Yan Lu[2], Feng Wu[2], Wen Gao[1]
[1]School of Computer Sciences, Harbin Institute of Technology, Harbin, 150001, China
[2]Microsoft Research Asia, Beijing, 100080, China
[3]Institute of Computing Technology, Chinese Science Academy, Beijing, 100080, China

*Abstract*—**This paper proposes a distributed video coding scheme based on the zerotree entropy (ZTE) coding. Wyner-Ziv theory on source coding with side information is taken as the basic coding principle, which makes independent encoding and joint decoding possible. In this scheme, wavelet transform is used to exploit the spatial correlation of a Wyner-Ziv frame. The quantized wavelet coefficients are reorganized in terms of the zero tree structure so as to identify the significant and insignificant coefficients. The significance map is intra-codec and transmitted. In particular, the significant coefficients are independently encoded with turbo coder, and only the parity bits are transmitted. At the decoder, a predictive frame generated through motion-compensated prediction is used as the side information, with which the Wyner-Ziv frame can be conditionally decoded. Experimental results show that, compared to the traditional intra-frame coding and pixel-domain Wnyer-Ziv video coding, the proposed scheme can achieve a better coding performance, especially at low bit rates.**

## I. INTRODUCTION

In traditional video coding schemes, asymmetric complexity exits in encoder and decoder. The motion estimation modules dominating the encoding complexity are usually very time-consuming. However, in many applications such as sensor networks and multi-camera scenarios, the complex compression has to be done in the processors with low processing capabilities. In this case, the encoding complexity becomes a big burden. As we know, most of the encoding complexity comes from the correlation exploration modules such as motion estimation. Is there any way to shift the correlation exploration from the encoder to the decoder, so that we can separately encode each frame in a video, while the performance is still as good as jointly encoding? In theory, distributed source coding (DSC) can provide a solution to this problem.

Theory of Slepian-Wolf shows that even if correlated sources are encoded without getting information from each other, coding performance can be as good as dependent encoding if the compressed signals can be jointly decoded [1]. Wyner and Ziv have extended the theory to the lossy source coding with side information [2]. Recently, several practical Slepian-Wolf and Wyner-Ziv coding techniques have been proposed for video coding, known as distributed video coding (DVC). In [3] and [4], a DVC framework based on syndrome of codeword co-set has been proposed. In [5] and [6], a DVC scheme using turbo codes has been proposed as well. In these algorithms, input video frames are classified into two categories, including intra frames and Wyner-Ziv frames, corresponding to the intra mode and inter mode, respectively. The goal of these researches is to make the coding performance close to the conditional entropy.

Basically, there are mainly three kinds of correlations to be utilized in video coding: temporal, spatial and statistical correlations. In the existing DVC systems, the temporal correlation is usually utilized at the decoder, for example, by generating the side information frame from the neighboring intra-coded frames. The spatial correlation within Wyner-Ziv frames is usually utilized by performing DCT or wavelet transform. As for the statistical correlation, some channel coding algorithms such as turbo codes have been employed in the source coding with side information. However, the high-order statistical correlation is seldom considered in the existing DVC systems, because it is difficult to align the source and side information signals. As we know, statistical correlation always plays an important role in entropy coding. Thus, it is desirable to consider the utilization of high-order statistical correlations so as to further improve the coding efficiency of Wyner-Ziv frames.

Therefore, we propose a novel DVC scheme, in which the high-order statistical correlation among transform coefficients can be further utilized. Following the basic idea in [5], we encode the Wyner-Ziv frame using turbo codes and decode it using the side information generated from surrounding intra frames with the motion-compensated prediction. Wavelet transform is used in the Wyner-Ziv frame coding for the purpose of exploiting spatial correlation. In order to utilize the high-order statistical correlation, a coding method based on zero-tree entropy (ZTE) coding [7] is developed. The transform coefficients are quantized using scalar quantization and reorganized into wavelet trees. Then, the significant coefficients are coded and the punctured parity bits are transmitted. At the decoder, the corresponding coefficients of the side information are extracted according to the significance map and used to reconstruct the Wyner-Ziv frame.

---

*This work was done while Xun Guo was with Microsoft Research Asia as a research intern.

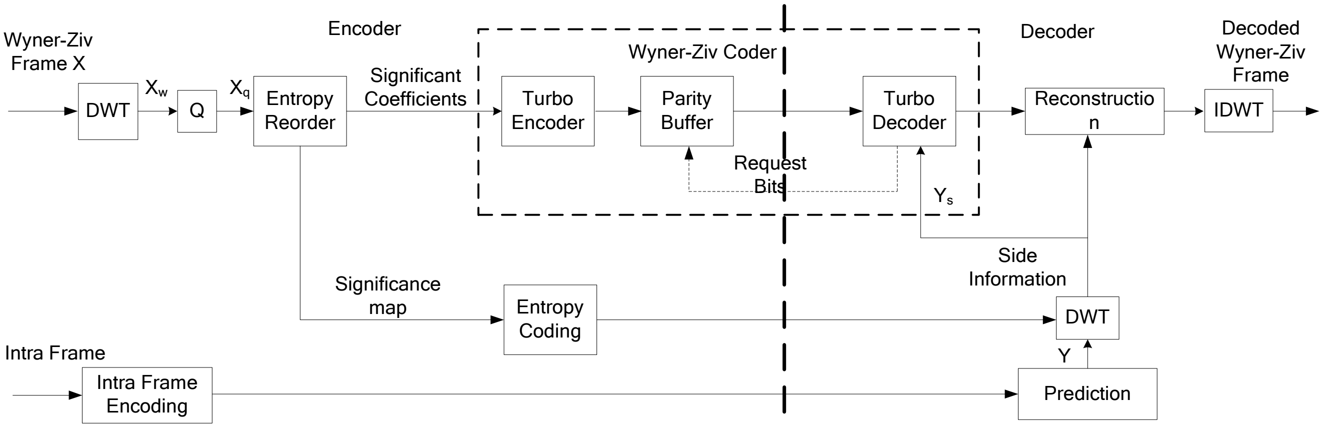

Figure 1. Diagram of proposed wavelet based Wyner-Ziv video coding scheme using turbo encoding and decoding.

The rest of this paper is organized as follows. Section II presents the proposed DVC scheme, including the coding architecture and the zerotree coding algorithm. Section III gives the experimental results and the analysis. Conclusions are drawn in section IV.

## II. PROPOSED WYNER-ZIV VIDEO CODING ALGORITHM

### A. *Proposed DVC Scheme*

Figure 1 shows the coding scheme of the proposed DVC system, in which frames of the input video sequence are classified into two categories: Intra frames and Wyner-Ziv frames. Intra frames are coded with the traditional DCT based intra coding method. Thus, the key point of the proposed DVC scheme is on how to efficiently compress the Wyner-Ziv frames. Our scheme uses wavelet transform to exploit the spatial correlation. The reason using wavelet is based on two observations: 1) the hierarchical structures of wavelet transform can decompose one frame into several scales from coarse to fine. This characteristic makes the spatial and quality scalability become possible and easy to be implemented. 2) These different scales have self-similarity which can be used to partially order the transformed coefficients and split the coefficients into significant or insignificant set. That is, in a hierarchical subband system, with the exception of the highest frequency subbands, every coefficient at a given scale can be related to a set of coefficients at the next finer scale of similar orientation. After quantization with some step size, if a coefficient in coarse scale is zero (insignificant), its descendants in the finer scales are very likely to be zero.

As shown in Figure 1, at encoder, a DWT is applied to the Wyner-Ziv frame X to generate coefficient set $X_w$. After quantization, the quantized coefficient set $X_q$ is reordered using a set partitioning process similar to the zero tree generation. In this process, the coefficients are mapped into significant or insignificant set. As a consequence, a tree structure can be achieved to indicate the positions of the significant coefficients. The significance map are coded with entropy coding and transmitted into decoder with intra coding mode. The significant coefficients are input into the Wyner-Ziv encoder. The core of the Wyner-Ziv encoder is a rate-compatible punctured turbo code (RCPT) [8]. The turbo code consists of two identical constituent convolutional codes. Only parity bits are stored and transmitted. A puncture schedule module will allocate proper bits to current Wyner-Ziv frame and transmit corresponding parity bits to the decoder. Currently, we use a feedback channel for request bits, which is the same as that in [5]. Since the feedback channel incurs high latency, it is desirable to have a puncture schedule at the encoder for the bit allocation. Frame complexity corresponding to the motion estimation accuracy can be used to estimate the possible symbol error rate, which remains the future research.

At decoder, the Wyner-Ziv frame is inter-decoded using side information Y which is the prediction of X generated from adjacent intra frames. After applying DWT on side information Y, decoder can extract the coefficients corresponding to the significant ones of X using the information of significance map and form coefficient set $Y_s$. Then, $Y_s$ is sent to the turbo decoder to decode the Wyner-Ziv frame together with the received parity bits. The decoder will successively decode the coefficients of a subband until an acceptable probability of bit error rate is achieved.

There are several methods to generate the side information: directly using previously reconstructed frame, extrapolation method and interpolation method. In order to get the best performance, we use the frame interpolation method based on motion compensation. In particular, the prediction of current Wyner-Ziv frame is generated from the forward and backward intra frames. This method is similar to the symmetric method in the prediction of traditional B frames. We assume that most of the motions in three successive frames are linear and the current motion vectors can be derived from the motion between adjacent two intra frames. Thus, motion compensation can be finished even if the current frame is absent.

The turbo decoder and reconstruction module assume a Laplacian residual distribution between X and Y. Let $d$ be the difference between corresponding coefficients in X and Y. Then the distribution of $d$ can be approximated as $f(d) = \frac{\alpha}{2} e^{-\alpha|d|}$ [5] with all the subbands grouping together. This characteristic is very useful to us, because sometimes we need to extract the significant coefficients from different subbands and code them together.

### *B. Zerotree Coding*

The Entropy reorder module in Figure 1 represents proposed zerotree based coding algorithm. In this paper, we use the coding concept of ZTE. Different from other embedded coding method such as EZW and SPIHT, proposed algorithm does not produce an embedded bitstream, which remains our future work. Quantization is explicit instead of implicit and can be performed distinctly from the zerotree growing process. Thus, more flexible adjustment can be easily done according to the content of some sequence. Uniform scalar quantization with deadzone is used to quantize the subbands coefficients into M levels. Note that different M values are selected for different subbands.

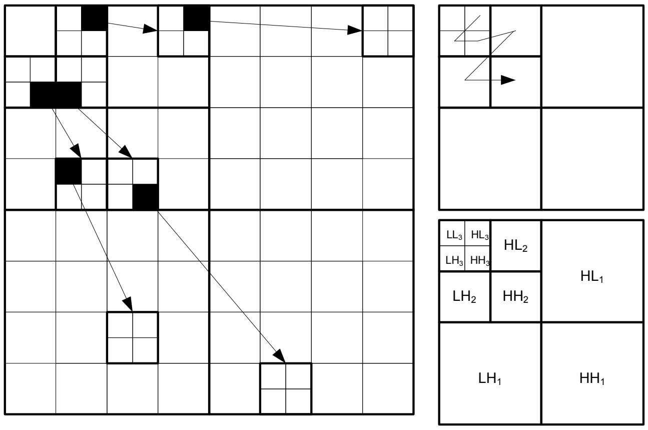

Figure 2. The relations between wavelet coefficients in different subbands (left) and the scan order using zerotree (upper right). The names of each subband are denoted in lower right figure.

Figure 2 gives the relations between wavelet coefficients in different subbands and the scan order when growing zerotrees. The blocks in the left one represents coefficients from different subbands and each blank block is a parent node. The arrow lines connect parent node and their children. Each parent node corresponds to four children nodes. The right two figures give the scan order example. Actually, proposed zerotree coding algorithm consists of two main parts:

1) Reorganizing quantized wavelet coefficients into tree structure. In this step, the transform coefficients have been previously quantized according to a pre-defined quantization matrix and the quantized subbands are scanned level by level. Parents are scanned before any of their children, but only after all neighboring parents have been scanned. A coefficient is significant if its amplitude is nonzero. If a coefficient is zero with all its children are zeros, the zerotree root symbol is used to signify it. And the isolated zero symbol signifies a coefficient which is zero but with at least one child nozero. Through this one pass scan process, significant coefficients of each subband can be found out and input into turbo codes. Thus, no bits need to be used upon the zeros. This is similar to the EOB used in block-wise DCT coding.

2) Encoding/decoding significant coefficients with turbo codes. In this step, significant coefficients of one subband are coded together. In order to achieve more flexible coding quality, the turbo codes consists of two 1/2 convolutional codes and the generator matrix $[\frac{1+D+D^3+D^4}{1+D^3+D^4}]$ is used to generate parity bits. At encoder, Coefficients are input into turbo codes bitplane by bitplane and sign information is implicitly indicated by the quantization index. In other words, the signs can also be predicted through the jointly decoding. At decoder, zero tree structure is intra decoded first. Then the corresponding coefficients of side information are extracted according to the tree structure. Thus, the received parity bits can be used to decode the coefficients together with side information.

For the *i*th bitplane of a coefficient $c_j$, denoted as $c_j^i$, and $c_j^{i'}$ represents estimated reconstruction value for $c_j^i$. The probability $P$ can be computed using the residual distribution model as follows:

$$P = \frac{\alpha}{2} e^{-\alpha \left| d_{C_j^i} \right|}, \text{ with } d_{C_j^i} = (m_i I(c_j^{i'}) + \mathit{offset}) - I(y_j)$$

where $m_i$ represents the magnitude of *i*th bitplane. $\mathrm{I}(c_j^{i'})$ indicates the possible value of $c_j^i$, which is equal to 1 or 0. $\mathrm{y}_\mathrm{j}$ is the coefficient of side information corresponding to $c_j$. *offset* is a estimated value used to compensate the lower part of $c_j$, because the lower bitpane of $c_j$ is still not decoded now. The value of *offset* is according to the distribution parameter and the quantization step size. After current bitplane is decoded, it will be used to help decoding the next bitplane.

## III. EXPERIMENTAL RESULTS

In order to verify the coding efficiency of proposed algorithm, we implemented the above Wyner-Ziv coding scheme. Foreman, Mother and Daughter and Akiyo sequences (QCIF, 30 fps) are used in the test. In each sequence, 200 frames are selected and Wyner-Ziv frame is set with the interval distance 1. Thus, we get the GOP structure as I-W-I-W, where I represents intra frame and W represents Wyner-Ziv frame. The intra frame is coded with traditional 263+ intra coding and can be decoded independently. Thus, every Wyner-Ziv can be predicted from its forward and backward intra frame. The symmetric motion estimation method is used to generate side information. Obviously, in some areas with non-linear motion, the prediction may not be accurate and the Wyner-Ziv bits can compensate the loss efficiently. We approximate the parameters of the Laplacian model by fitting the histogram of the residual between reconstructed intra frame and side information, and each subbands have different values. Through observing the performance of different parameters, we found that the model is not very sensitive when the differences of parameters are not large. So the performance will not decrease much if the approximation is not equal to real value.

Figure 4 gives the rate-distortion (RD) curves of the three sequences. Only the PSNR of luminance component and bitrate of Wyner-Ziv frames are illustrated. There are four curves in each figure. The curve of “263+I frames” indicates the results of 263+ intra coding for the Wyner-Ziv frames and is taken as the lower bound. Results of pixel domain and wavelet transform domain without zerotree are also shown, denoted as “pixel domain” and “wavelet domain”. The results of pixel domain are achieved using the scheme similar to [5]. 4/5 turbo codes is used and three quantization level (2, 4, 16) are shown.

To show the results more obviously, we use the consistent side information in each video sequence. The average PSNRs of the side information for the three sequences are 31.86 dB, 36.13 dB and 37.11 dB, respectively. From the curves, we can see that the results of Wyner-Ziv coding are much better than the intra coding. Particularly, up to 8 dB gain can be achieved compared to the all Intra-frame coding. In other words, the Wyner-Ziv coding method has exploited the temporal correlation efficiently. Compared to the pixel domain coding method, the wavelet domain method outperforms up to 0.6 dB. In other words, the spatial correlation has been be utilized by the wavelet transform. In addition, the proposed zerotree coding method outperforms the original wavelet domain method up to 0.5 dB. That is to say, the high-order statistical correlation has been utilized in the proposed method, which really benefits the overall coding performance of Wyner-Ziv coding.

## IV. CONCLUSIONS

In this paper, we have presented a distributed video coding scheme based on zerotree entropy coding, in which the Wyner-Ziv theory on source coding with side information is employed as the basic coding principle. Wavelet transform is used in Wyner-Ziv frames to exploit spatial correlation and high-order entropy is further utilized by grouping the wavelet transform coefficients into tree structure. Without increasing encoder complexity too much, proposed scheme achieves promising performance compared to the results without entropy coding.

## REFERENCES

[1] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, pp.471-480, July 1973.

[2] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, pp.1-10, Jan.1976.

[3] S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): design and construction,” *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626-643, Mar. 2003.

[4] R. Puri, and K. Ramchandran, “PRIMS: a new robust video coding architecture based on distributed compression principles,” *Proc.* of *40th Allerton Conference on Communication, Control, and Computing*, Allerton, Illinois, Oct. 2002.

[5] A. Aaron, R. Zhang and B. Girod, “Wyner-Ziv coding of motion video,” Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA, Nov. 2002

[6] B. Girod, A. Aaron, S. Rane and D. R. Monedero, “Distributed video coding,” *Proceedings of the IEEE*, Special Issue on Video Coding and Delivery, vol.93, no.1, pp.71-88, Jan. 2005.

[7] S. A. Martucci and I. Sodagar, “Zerotree entropy coding of wavelet coefficients for very low bit rate video,” ICIP1996, Lausanne, Swizerland, Sept. 1996.

[8] D. Rowitch and L. Milstein, “On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo codes,” *IEEE Transactions on Communications*, vol.48, no.6, pp.948-959, June 2000.

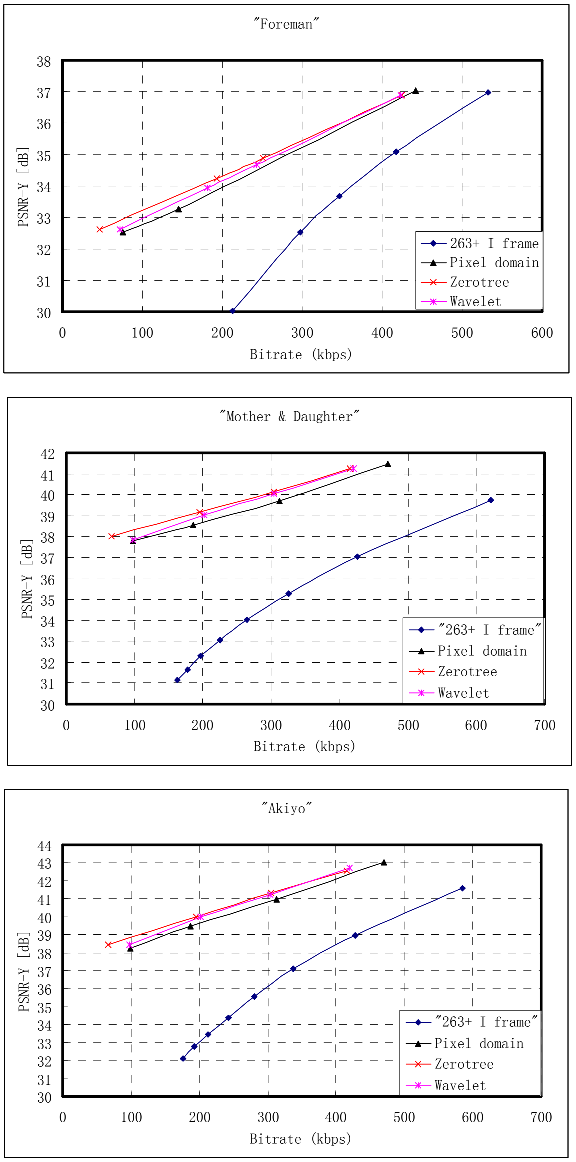

Figure 3. Experimental resuts for Foreman, Mother&Daughter and Akiyo sequences.