WYNER-ZIV VIDEO CODING BASED ON SET PARTITIONING IN HIERARCHICAL TREE

Xun Guo[1,], Yan Lu[2], Feng Wu[2], Wen Gao[3], Shipeng Li[2]*

[1]School of Computer Science, Harbin Institute of Technology, Harbin, 150001, China
[2]Microsoft Research Asia, Beijing, 100080, China
[3]School of Electronic Engineering, Peking University, Beijing, 100080, China

ABSTRACT

In this paper, we propose a Wyner-Ziv video coding scheme based on set-partitioning in hierarchical trees (SPIHT) which can utilizing not only the spatial and temporal correlations but also the high-order statistical correlations. Wyner-Ziv theory on source coding with side information is employed as the basic coding principle, which makes the independent encoding and joint decoding become possible. In the proposed scheme, wavelet transform is first used to de-correlate the spatial dependency of a Wyner-Ziv frame. Then the quantized transform coefficients are organized by using magnitude with a set partitioning sorting algorithm. The ordered bit planes are coded using the Wyner-Ziv coding based on turbo codes. At the decoder, side information generated by motion compensated interpolation is used to conditionally decode the Wyner-Ziv frame. Since the high order statistical correlation is used, the proposed algorithm owns advantages over the traditional pixel-domain and transform-domain Wyner-ziv video coding schemes.

1. INTRODUCTION

In traditional video coding schemes, asymmetric complexity exits in the encoder and decoder. The motion compensated prediction makes the encoder much more time-consuming than the decoder. 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 complexity becomes a big burden. As we know, the correlation exploration modules such as motion estimation dominate the encoding complexity in traditional video coding. Is there any way to shift the correlation exploration from encoder to decoder, i.e., 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 dependently 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 Wyner-Ziv coding techniques have been proposed for distributed source coding (DSC) [3], and also for distributed video coding (DVC). In [4], Pradhan and Ramchandran proposed a DVC framework based on syndrome of codeword co-set. In [5] and [6], Aaron and Girod proposed a frame based DVC scheme using turbo codes. In these schemes, input video frames are classified into two categories, namely intra frame and Wyner-Ziv frame, which are inherently coded with intra mode and inter mode, respectively. The main goal of these researches is to make the coding performance closer to the conditional entropy to the maximum extend.

Basically, there are mainly three kinds of correlations to be utilized in the video coding: temporal correlation, spatial correlation and statistical correlation. In the existing DVC systems, the temporal correlation is usually utilized at the decoder by, for example, generating the side information frame from the neighboring intra-coded frames. The spatial correlation within Wyner-Ziv frames is usually utilized by performing DCT [6] or wavelet transform [7]. 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 schemes due to the difficulty of the alignment between the source and the side information signals. As we know, statistical correlation always plays an important role in the entropy coding. Thus, if we want to further improve the coding efficiency of Wyner-Ziv frames, this is an important issue that we should consider.

Therefore, in this paper, we propose a novel DVC scheme which can exploit the high-order statistical correlation among the transform coefficients using tree structure and bit plane coding. Following the basic idea proposed in [5], the Wyner-Ziv frame is encoded using turbo codes and decoded with side information generated from the intra frames by using motion compensated prediction. Wavelet transform is used in the Wyner-Ziv

*This work was done when the author was with Microsoft Research Asia as an intern.

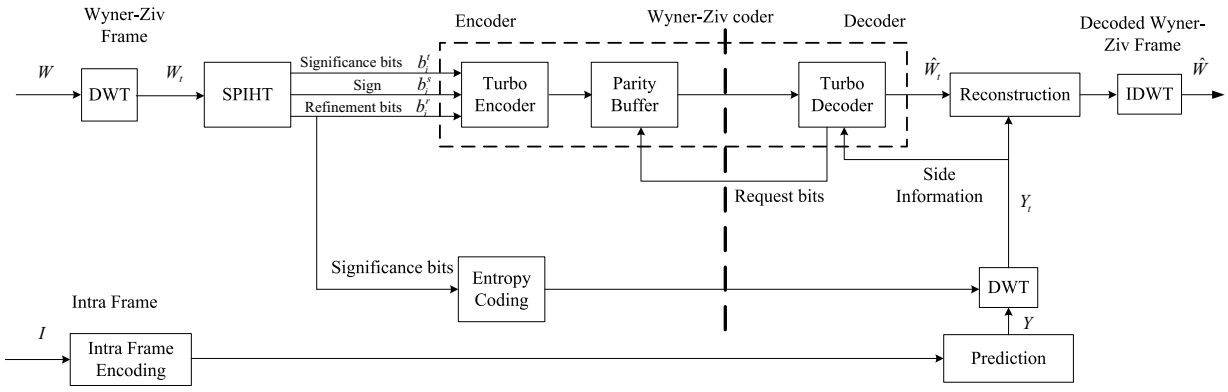

Figure 1. Diagram of proposed wavelet based Wyner-Ziv video coding scheme using turbo codes

frame coding for the purpose of exploiting the spatial correlation. To further utilize the high-order statistical correlation, the organization of transform coefficients based on SPIHT [8] is also incorporated in the proposed scheme. In particular, the transformed coefficients are split into different bit planes, named significance bits, sign bits and refinement bits. These bit planes are coded. Since Turbo codes are employed, only the punctured parity bits are transmitted to the decoder.

The remainder of this paper is organized as follows: Section 2 presents the whole DVC scheme employed by this paper including overall analysis and key techniques. Section 3 gives the experimental results. Conclusions are drawn in section 4.

2. PROPOSED WYNER-ZIV VIDEO CODING

2.1. Wyner-Ziv Video Coding Architecture

In frame-based DVC scheme, temporal correlations are exploited by using side information through channel coding method and spatial correlations are exploited by using transform. But no method for exploiting high-order statistical correlations, i.e. entropy coding, is considered. This is mainly because that after entropy coding, the transformed coefficients will be transferred into variable length codes and the map between Wyner-Ziv frame and side information will be damaged. Also, the distribution model of the virtual correlation channel can not be simulated at the decoder and the decoding efficiency of the turbo code will decrease largely.

To tackle this problem, we use the discrete wavelet transform (DWT) to de-correlate the spatial correlation of Wyner-Ziv frame. The advantages of using DWT in Wyner-Ziv coding are as follows. The self-similarity for co-located scales can be utilized to further exploit the correlations between coefficients. That is, in a set of scales generated by DWT, each coefficient in a given scale can be related to a set of coefficients in the next finer scale of similar orientation. This hierarchical structure leads to a fact that if a coefficient in coarse scale is smaller than a threshold in a given bit plane, its descendants in the finer scales are very likely smaller than the threshold. Some algorithms which utilize this characteristic, e.g. SPIHT, have been used in image coding long time ago and achieved promising performance.

Based on the above analysis, we employ the SPIHT to reorder the transformed coefficients before turbo coding. This is not just a straightforward application. There are two reasons that make this algorithm fit to DVC scenario. Firstly, the process of this algorithm corresponds to partially entropy coding. Secondly, the distribution model of bit planes generated by this algorithm can be easily established and the map between Wyner-Ziv frame and side information will not be disarranged.

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.

As shown in Figure 1, at encoder, a DWT is applied to the Wyner-Ziv frame W to generate coefficient set W_t. Then, W_t is reordered using a set partition process similar to zero tree generation. In this process, the coefficients are mapped into different bit planes b_i^t, b_i^s and b_i^r, which indicate significance bits, sign bits and refinement bits in bit plane i respectively. Significance bits are the most important information which indicates the ordered structure of bit plane i. These bits are coded with intra coding or inter coding and transmitted into the decoder. The core of the Wyner-Ziv encoder is a rate-compatible punctured turbo code (RCPT) [9]. The turbo code consists of two identical constituent convolutional codes. Only parity bits are stored and transmitted.

At decoder, the Wyner-Ziv frame is inter decoded using side information Y which is the prediction of W

generated from adjacent intra frames. After applying DWT on side information Y, decoder can extract the coefficients corresponding to those of W using the information of significance bits and form coefficient set Y_t. Then, Y_t 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.

2.2. Bitplane Encoding

In this process, three kinds of bits will be generated: significance bits, sign bits and refinement bits. The basic encoding order is as follows: significance bits will be encoded and transmitted with intra coding or Wyner-Ziv coding firstly. Then, sign bits and refinement bits are Wyner-Ziv coded using the information of significance bits. We denote a coefficient as x whose descendents set is $D(x)$, $S_k(x)$ represents the significance value of x in bitplane k and $B_k(x)$ represents the k th bitplane of x. Then, for bitplane n, if $2^n \le |B_n(x)| \le 2^{n+1}$, x is significant at bitplane n and $S_k(x)$ is 1. We can see that significance bits always represent the true values of coefficient bit planes.

Thus, besides spatial correlations between $S_k(x)$ and $S_k(D(x))$, temporal correlations between $S_k(x)$ and $S_k(y)$ also exists. Therefore, we use both intra coding and Wyner-Ziv coding method for significance bits. For the first two bitplanes, the temporal correlations are strong and the tree structures Wyner-Ziv frame and side information are similar. Therefore, we use significance bits from side information to predict the significance bits of Wyner-Ziv frame. However, as the bitplane level increases, the temporal correlations become weak. So intra coding is used in this case. For sign bits, we also use Wyner-Ziv method but not ouput them directly. We assume distribution between W and Y is a Laplacian model, $f(d) = \frac{\alpha}{2} e^{-\alpha|d|}$, where d is the difference between corresponding coefficients in W and Y.

Before SPIHT, deadzone method is used to decrease the different probability of signs between W_t and Y_t. Then, the sign bits can be coded with Wyner-Ziv coding and decoded together with significance bits. For example, if a sign of a significant coefficient is 1 then it can be taken as 0x11 when decoding and use the distribution model as a true value.

2.3. Bitplane Decoding

Log-MAP algorithm is used in this paper and for a coefficient bit of W, denoted as b_i,

$$L(b_i) = \log \frac{P(b_i=0|y_i)}{P(b_i=1|y_i)} \qquad (1)$$

When decoding, the decoder will judge the value of $L(b_i)$, if $L(b_i) > 0$, $b_i = 0$, or $b_i = 1$. For practical implementation of turbo codes, (1) can be expressed as,

$$L(b_i) = \log\left(\frac{\sum_{\chi_0} \alpha_{i-1}(s')\gamma_i(s',s)\beta(s)}{\sum_{\chi_1} \alpha_{i-1}(s')\gamma_i(s',s)\beta(s)}\right), \qquad (2)$$

where χ_0 is the set of all transitions from state s' to s with input 0, and χ_1 is similarly defined. And,

$$\gamma_i(s',s) = P(b_i)P(b_i^{'} \mid y_i)P(c_i \mid p_i), \qquad (3)$$

where $b_i^{'}$ is the evaluation of b_i, and c_i is the evaluated output of parity bits p_i. Considering the previous decoded bitplanes of b_i, $P(b_i^{'}|y_i)$in (3) should be modified into $P(b_i^{'}|y_i, b_{i-1}, b_{i-2}, \ldots)$. and can be calculated as

$$P(b_i^{'} \mid y_i) = \frac{\alpha}{2} e^{-\alpha\left|d_{b_i^{'}}\right|}, \text{ with}$$
$$d_{b_i^{'}} = (V(b_i) + m_i I(b_i^{'}) + offset) - I(y_j) \qquad (4)$$

where m_i represents the magnitude of ith bitplane and $V(b_i)$ represents the value of the previous decoded bitplanes of b_i. $I(b_i^{'})$ indicates the possible value of b_i, which is equal to 1 or 0. y_j is the coefficient of side information corresponding to b_i. *offset* is a estimated value used to compensate the lower part of b_i, because the lower bitpane of b_i is still not decoded now. The value of *offset* is according to the distribution parameter and the bitplane i. After current bitplane is decoded, it will be used to help decoding the next bitplane.

3. EXPERIMENTAL RESULTS

In order to verify the coding efficiency of proposed algorithm, we implemented above Wyner-Ziv video coding scheme. Results of two sequences, foreman, Mother and Akiyo, in QCIF format 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 IWIW, where I represents intra frame and W represents Wyner-Ziv frame. Intra frames are encoded with H.263+ intra coding and decoded independently at decoder to generate side information. A symmetric motion estimation based algorithm is used to interpolate side information. Obviously, in some areas with non-linear motion, the side information may not be accurate and the Wyner-Ziv bits can be used to compensate the loss. We approximate the parameters of the Laplacian model by fitting the difference between reconstructed intra frame and side information, and each bit plane may have different value. Through observing the performance of different parameters, we found that the model is not very sensitive to the parameters in a limited range, just the same as the claim in [5]. Thus, the performance will not decrease much if the approximation does not equal to real value.

Figure 2 gives the coding performance for luminance component of the three sequences and only Wyner-Ziv frames are illustrated. To show the results more obviously,

we use consistent side information in each video sequence. And the PSNR of the side information for these sequences are 31.86 dB, 36.13 dB and 37.11 dB, respectively. There are four curves in each figure. The curve of "263+ I frame" indicates the results of intra coding using H.263+ for the Wyner-Ziv frames and is taken as the benchmark. Results of pixel domain and wavelet transform domain are also given, denoted as "Pixel domain" and "Wavelet", respectively. The result of pixel domain is achieved using the method proposed [5] and is also taken as a benchmark of Wyner-Ziv video coding.

As shown in Figure 2, 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 SPIHT coding method outperforms the other wavelet domain methods up to 0.7 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.

4. CONCLUSIONS

In this paper, we have presented a Wyner-Ziv video coding scheme based on wavelet and SPIHT, in which Wyner-Ziv theory on source coding with side information is employed as the basic coding principle. Wavelet transform and SPIHT are used to exploit spatial and high-order statistical correlations. Without increasing encoder complexity too much, proposed scheme achieves promising performance compared to the results without or with little entropy coding.

5. 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. S. Pradhan and Kannan Ramchandran, "Distributed Source Coding Using Syndromes (DISCUS): design and construction," *IEEE Transactions on Information Theory*, vol.49, pp.626-643, Mar. 2003

[4] R. Puri and K. Ramchandran, "PRISM: a new robust video coding architecture based on distributed compression principles," *Proc. of 40 th 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] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv codec for video," Proc. Visual Communications and Image Processing, VCIP-2004, San Jose, CA, Jan. 2004.

[7] X. Guo, Y. Lu, F. Wu, W. Gao and S. Li, "Distributed multi-view video coding," *Proc. Visual Communications and Image Processing*, VCIP-2006, San Jose, CA, Jan, 2006.

[8] A. Said and W. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on CSVT*, vol. 6, No. 3, pp.243-250, June, 1996.

[9] 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.

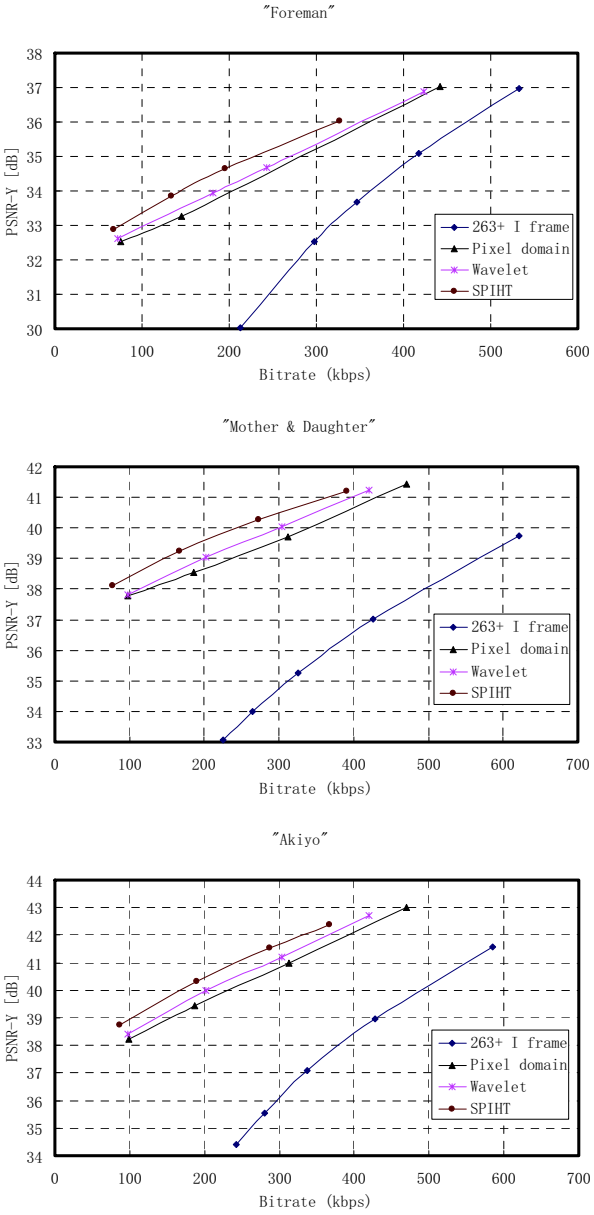

Figure 2. Simulated results for foreman, mother and akiyo