WZS: WYNER-ZIV SCALABLE PREDICTIVE VIDEO CODING

Huisheng Wang, Ngai-Man Cheung and Antonio Ortega

Integrated Media Systems Center and Department of Electrical Engineering
University of Southern California, Los Angeles, CA 90089-2564
E-mail: {huishenw, ncheung, ortega}@sipi.usc.edu

ABSTRACT

A novel Wyner-Ziv scalable (WZS) video coding approach is proposed to enhance MPEG-4 fine granularity scalability (FGS). Starting from the same base layer as MPEG-4 FGS, the proposed coder can achieve higher coding efficiency using Wyner-Ziv coding, by selectively exploiting the high quality reconstruction of the previous frame in the enhancement layer bitplane coding of the current frame. This creates a multi-layer Wyner-Ziv prediction "link", connecting the same bitplane level between successive frames, to provide more improved prediction as compared to MPEG-4 FGS, while keeping complexity reasonable at the encoder. Since the temporal correlation varies in time and space, a block-based adaptive mode selection algorithm is designed for each bitplane to switch between different coding modes. Experimental results show improved coding efficiency up to 2dB over FGS for video sequences with high temporal correlation.

1. INTRODUCTION

Scalable coding has become a desired functionality for error-resilient video transmission over heterogenous networks because it facilitates adapting to varying channel conditions, such as available bandwidth and packet loss rate. Predictive coding, in which previously reconstructed frames are used as a predictor for the current frame, is an important technique to remove temporal redundancy among successive frames. Efficient scalable coding becomes more difficult if predictive techniques are used because scalability leads to multiple possible reconstructions of each frame. In this situation either a single prediction is used, which leads to either drift or coding inefficiency, or a different prediction is obtained for each reconstructed version, which leads to added complexity.

The MPEG-4 committee has developed the fine granularity scalability (FGS) [1] profile that provides a scalable solution for video streaming. In MPEG-4 FGS, a video sequence is coded into a base layer and an enhancement layer (EL). The base layer uses a nonscalable codec, where only base layer information for previous frames can be utilized in the motion-compensated prediction (MCP) loop. The EL encodes for each frame the residual between base layer reconstruction and the original, using bit-plane coding of the DCT coefficients. Since MPEG-4 FGS does not exploit the EL information of previous frames in the MCP loop, arbitrary truncation of the EL bitstream for one frame will not introduce drift problems to succeeding frames, which makes the MPEG-4 FGS flexible in supporting streaming applications. However this also results in low coding efficiency, especially for sequences that exhibit high temporal correlation.

Rose and Regunathan [2] proposed a multiple-MCP-loop approach for general SNR scalability, in which the EL predictor is optimally estimated by considering all the available information from both base and enhancement layers. This type of closed-loop prediction (CLP) has the disadvantage of requiring the encoder to generate all possible decoded versions for each frame, so that each of them can be used to generate a prediction residue. Thus, the complexity is high at the encoder especially for multi-layer coding scenarios. Several alternative multi-layer techniques have also been proposed to exploit the further temporal correlation in the EL inside the FGS framework, such as PFGS [3] and AMC-FGS [4]. The common features of these techniques are to employ one or more additional MCP loops for P and B EL frames (or B frames only), for which a certain number of FGS bitplanes, M, are included in the EL MCP loop, to improve the coding efficiency. In this case prediction drift will occur within the FGS layers when fewer than M bitplanes are received. M is chosen by considering the trade-off between the coding efficiency and prediction drift. In summary, traditional CLP techniques suffer the inherent limitation that, in order to avoid drift, the same predictor has to be available at both encoder and decoder.

Based on the Wyner-Ziv framework [5], several video codecs using side information (SI) at the decoder have been proposed in the recent literature [6, 7]. These can be thought of as an intermediate step between "closing the prediction loop" and coding independently. In closed-loop prediction the encoder needs the exact value of the predictor to generate the residue. Instead, a Wyner-Ziv encoder only requires the **correlation structure** between the current signal and the predictor. Thus there is no need to generate the decoded signal at the encoder as long as the correlation structure is known, or can be found. Some of the recent work addresses the problem of scalable coding in this setting. Steinberg and Merhav [8] formulated the theoretical problem of successive refinement of information, originally proposed by Equitz and Cover [9], in the Wyner-Ziv setting. The achievable region is given, and the necessary and sufficient conditions are also provided for successive refinability in the sense that both stages can asymptotically achieve the Wyner-Ziv R-D function simultaneously. Xu and Xiong [10] proposed an MPEG-4 FGS-like scheme by treating a standard coded video as a base layer, and building the bit-plane enhancement layers using Wyner-Ziv coding with current base and lower layers as SI. We also proposed a Wyner-Ziv scalable predictive coding method in [11], using nested lattice quantization followed by multi-layer Slepian-Wolf coders (SWC) with layered side information. The approach was developed on a first-order DPCM model and has shown significant benefits in terms of the enhancement layer reconstruction.

Here, we extend our work in [11] to scalable video coding in the FGS framework, where our goal is to construct efficient ELs starting from a standard CLP base layer video coder like MPEG-4. Consider multiple layers EL_{i1}, EL_{i2}, ..., EL_{iL} for the ith frame as shown in Fig. 1. EL_{ij} is coded by exploiting all the information from $EL_{ik},\ k < j$ and $EL_{i-1,k},\ k \leq j$. Our approach can be seen as the Wyner-Ziv counterpart of the CLP-based estimation-theoretic (ET) approach in [2], where, in order to reduce the complexity, we do not explicitly construct multiple motion-compensation loops. The proposed approach differs from [10] in that we explore the remaining temporal correlation between the successive frames in the EL by Wyner-Ziv coding to achieve improved performance over FGS. Compared to proposed variations of FGS techniques [3, 4], our approach supports more flexible techniques to encode, bitplane by bitplane, the residue between base layer and original video. Specifically, several different modes can be used to encode each macroblock in any given bitplane, so that pure "intra-frame" bitplane refinement is used when a macroblock exhibits low temporal correlation, while previous side information is used when high temporal correlation exists.

The paper is organized as follows. In Section 2, we briefly review the ET approach, and then describe our proposed scalable coding techniques. In Section 3, we analyze the correlation structure for typical video sequences, which we explore in our coding algorithms. Simulation results on two video sequences are presented in Section 4 and show substantial improvement in video quality for sequences with high temporal correlation. Finally, conclusions and future research directions are given in Section 5.

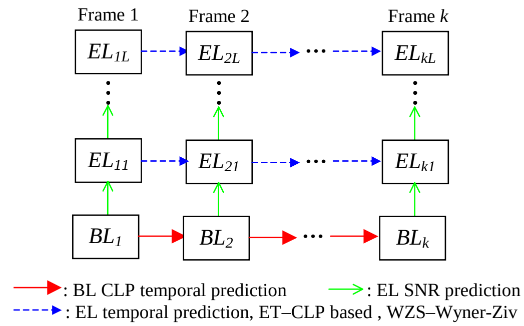


Fig. 1. Basic difference of prediction techniques used in ET and in proposed WZS. BL_i: the base layer of the ith frame. EL_{ij}: the jth EL of the ith frame, where the most significant EL bitplane is denoted by $j = 1$.

2. WZS VIDEO CODER DESIGN

Let us assume that the following information is available when EL_{ij} is being decoded: (1) $EL_{i-1,k},\ k \leq j$, and (2) all information from $EL_{ik},\ k < j$, including reconstruction, prediction mode, base layer motion vector for each Inter-mode macroblock, and the compressed residual. For simplicity, the base layer motion vectors are reused by all EL bitplanes.

2.1. Brief Review of ET Approach

The temporal evolution of DCT coefficients can be usually modelled by a first-order Markov process

$$x_k = \rho x_{k-1} + z_k,\ x_{k-1} \perp z_k \tag{1}$$

where x_k is a DCT coefficient in the current frame and x_{k-1} is the corresponding DCT coefficient after motion compensation in the previous frame. Let $\hat{x}_k^b$ and $\hat{x}_k^e$ be the base and enhancement layer reconstruction of x_k, respectively. Assume the base layer is already compressed and gives the information $x_k \in (a, b)$. The optimal base layer reconstruction $\hat{x}_k^b$ is then $E[x_k|\hat{x}_{k-1}^b,\ x_k \in (a, b)]$. In addition to the information provided by the base layer, the EL decoder has access to the EL reconstructed DCT coefficient $\hat{x}_{k-1}^e$ of the previous frame. Thus the optimal EL predictor is given by

$$\begin{aligned} \tilde{x}_k^e &= E[x_k|\hat{x}_{k-1}^e,\ x_k \in (a, b)] \\ &\approx \rho\hat{x}_{k-1}^e + E[z_k|z_k \in (a - \rho\hat{x}_{k-1}^e, b - \rho\hat{x}_{k-1}^e)] \end{aligned} \tag{2}$$

The EL encoder then quantizes the residual $r_k^e = x_k - \tilde{x}_k^e$. Let (c, d) be the quantization interval associated with r_k^e, i.e., $r_k^e \in (c, d)$, and let $e = \max(a, c + \tilde{x}_k^e)$ and $f = \min(b, d + \tilde{x}_k^e)$. The EL reconstruction $\hat{x}_k^e$ is $E[x_k|\hat{x}_{k-1}^e, x_k \in (e, f)]$.

The ET predictor in (2) can be simplified to either $\hat{x}_k^b$ or $\hat{x}_{k-1}^e$ under the following conditions: (1) $\tilde{x}_k^e \approx \hat{x}_k^b$ if correlation low, $\rho \approx 0$, or the total rate is approximately the same as the base-layer rate, i.e., $\hat{x}^e_{k-1} \approx \hat{x}^b_{k-1}$; (2) $\tilde{x}^e_k \approx \hat{x}^e_{k-1}$ for the case with high correlation and where the base-layer reproduction is much coarser than that of EL.

Since the statistics of the EL prediction residual r^e_k vary greatly depending on the exact value of (a, b), [2] developed two entropy codes for two different classes of r^e_k: one for the case when $0 \in (a, b)$, another for all the other cases.

2.2. WZS Prediction techniques for enhancement layers

The main disadvantage of the ET approach for multi-layer coding resides in its complexity, since multiple motion compensation loops are necessary for EL prediction coding. If the complexity at the encoder is limited, we cannot generate all possible reconstructions of the reference frame at the encoder. Under this constraint we investigate techniques to exploit the temporal correlation between neighboring frames at each EL. In this paper, we propose to use Wyner-Ziv coding (WZC) to replace the closed loop between the respective ELs of neighboring frames. The basic difference between multi-loop predictive techniques such as ET and the proposed WZS approach is illustrated in Fig. 1. For example, in the ET approach, in order to encode EL_{21} (first enhancement layer of the second frame), the exact reproduction of EL_{11} must be available at the encoder. Instead, exact knowledge of EL_{11} will not be necessary for the WZS approach as long as we are able to estimate the correlation between EL_{21} and EL_{11}.

The following discussion is in the context of two-layer coder, and it can be easily extended to multi-layer coding scenario. In the ET approach, the EL encoder quantizes the residual $r^e_k = x_k - \tilde{x}^e_k$ and sends it to the decoder. However, in our problem, the encoder can only access $\hat{x}^b_k$ while the decoder has access to both $\hat{x}^b_k$ and $\hat{x}^e_{k-1}$. Thus, the encoder does not have access to $\tilde{x}^e_k$ in order to avoid having to "close the loop" at each layer. We can rewrite r^e_k as

$$r^e_k = x_k - \tilde{x}^e_k = (x_k - \hat{x}^b_k) - (\tilde{x}^e_k - \hat{x}^b_k) \qquad (3)$$

To better show how this can be cast as a Wyner-Ziv coding problem, let $u_k = x_k - \hat{x}^b_k$ and $v_k = \tilde{x}^e_k - \hat{x}^b_k$. With this notation u_k plays the role of the input signal and v_k plays the role of SI available at the decoder only. Now the problem is how to design a method to estimate the correlation between u_k and v_k. In order to do this, we first propose a simplification of the optimal predictor $\tilde{x}^e_k$ and then discuss correlation estimation at the encoder.

In the FGS framework each coefficient at the i-th bitplane can take at most 3 different values: -1, 0 and 1, and the quantization interval of the ith bitplane is at least 1/4 of that of the $(i-1)$th bitplane. In this setting, the ET optimal predictor $\tilde{x}^e_k$ can be simplified to either $\hat{x}^e_{k-1}$ or $\hat{x}^b_k$, depending on whether the temporal correlation is strong (choose $\hat{x}^e_{k-1}$) or not (choose $\hat{x}^b_k$). If $\tilde{x}^e_k = \hat{x}^b_k$, then $v_k = 0$. We send u_k directly as FGS. If $\tilde{x}^e_k = \hat{x}^e_{k-1}$, we apply WZC to u_k with the estimated correlation between u_k and v_k.

One approach for correlation estimation is to construct an approximation of $\hat{x}^e_{k-1}$ at the encoder, and then use the approximation in the correlation calculation. We requantize the original frame x_{k-1} at the encoder to a quality level that approximates that achieved by EL reconstruction at the decoder. We denote this as $\bar{x}^e_{k-1}$, which is then used to approximate $\hat{x}^e_{k-1}$. Let $s_k = \bar{x}^e_{k-1} - \hat{x}^b_k$ be the approximation of v_k. We assume that $E[u_k - v_k] = 0$ and then estimate the variance $E[(u_k - v_k)^2] \approx E[(u_k - s_k)^2] = E[(x_k - \bar{x}^e_{k-1})^2]$, which will then be used in the WZ coder.

As mentioned above, we allow a different prediction mode to be selected on a macroblock (MB) by macroblock basis (allowing adaptation of the prediction mode for smaller units, such as blocks or DCT coefficients would be impractical). Our adaptive mode selection operates at the macroblock level, so that, as will be seen next, differences in temporal correlation in different MBs can be exploited. For all macroblocks to be encoded using WZC, the variance $E[(u_k - v_k)^2]$ is estimated as $E = \sum_{k \in MB} (x_k - \bar{x}^e_{k-1})^2$, and used to select a channel coder with an appropriate rate.

2.3. WZS System Architecture

Fig. 2 depicts the WZS encoder and decoder in the FGS framework. Let X_k, $\hat{X}^b_k$ and $\hat{X}^i_k$ be the original frame, its base layer reconstruction and the ith EL reconstruction, respectively, in the pixel domain. In bitplane operation of DCT coefficients, $x(l)$ represents the lth bitplane of x, while x^l indicates the value of x truncated to the l most significant bitplanes.

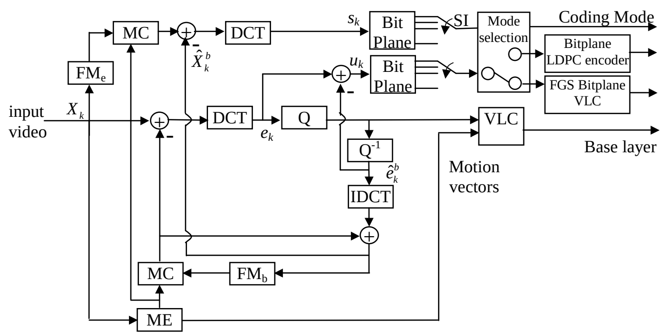
Encoding: In base layer, the DCT residual of the difference between the original image X_k and the motion compensated (MC) base layer reference is given by $e_k = T(X_k - MC_k[\hat{X}^b_{k-1}])$, where $T(.)$ is the DCT transform, and $MC_k[.]$ is the motion-compensated prediction of the kth frame given $\hat{X}^b_{k-1}$. The reconstructed MC residual after base layer quantization and dequantization is denoted by $\hat{e}^b_k$. Following the same discussion as Section 2.2,

$$u_k = e_k - \hat{e}^b_k = T(X_k - MC_k[\hat{X}^b_{k-1}]) - \hat{e}^b_k \qquad (4)$$

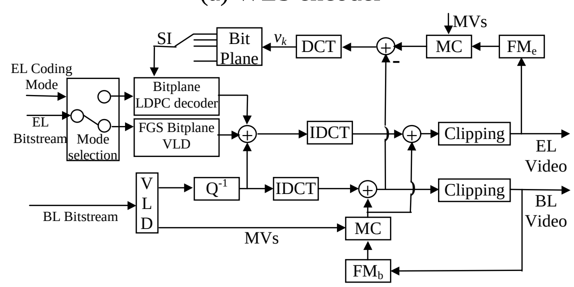
For bitplane coding, the reconstruction of u_k at the lth bitplane is exactly the truncation of u_k to the l most significant bitplanes, i.e., u^l_k. If u^{l-1}_k is selected as the predictor for $u_k(l)$, we can simply code $u_k(l)$ in the same way as FGS.

If we choose the EL information from the previous frame as the predictor, we need to approximate v_k at the encoder to estimate the correlation between u_k and v_k. The reference frame X_{k-1} is first motion-compensated, and then we subtract $\hat{X}^b_k$ to obtain the difference image. The DCT residual of this difference image is an approximation of v_k

$$s_k = T(MC_k[X_{k-1}] - \hat{X}^b_k) \qquad (5)$$



(a) WZS encoder



(b) WZS decoder

Fig. 2. The block diagram of WZS encoder and decoder. FM: frame memory, ME: motion estimation, MC: motion compensation.

To switch between these two cases for coding $u_k(l)$, we define the following parameters for each MB, counting the luminance component only.

$$\begin{aligned} E_{intra} &= \sum_{MB_i} (u_k^l - u_k^{l-1})^2 \\ E_{inter} &= \sum_{MB_i} (u_k^l - s_k^l)^2 \end{aligned} \tag{6}$$

Obviously, the larger E_{intra}, the more bits the FGS coding requires. Similarly E_{inter} gives an indication of the correlation level of the ith MB between u_k^l and s_k^l, which is an approximation of u_k and v_k at the lth bitplane. The mode selection algorithm is then defined as follows.

For the l th bitplane,

(1) If base layer MB mode = INTRA, then EL MB mode = FGS-MB, else EL MB mode = WZS-MB;

(2) If a DCT block contains all zeros, then block mode = **ALL-ZERO**, else go to (3);

(3) If EL MB mode = FGS-MB, then its 4 DCT block mode = **FGS**, else go to (4);

(4) If the SI s_k^l of this DCT block is exactly same as the signal u_k^l, then block mode = **WZS-SKIP**, else block mode = **WZS**.

No extra texture information is sent in either ALL-ZERO or WZS-SKIP modes. The ALL-ZERO mode already exists in the current MPEG-4 FGS. For a block coded in WZS-SKIP, the decoder just copies the corresponding block of the reference frame.[1].

For the WZS blocks, we apply two coding passes, significance pass and refinement pass, both of them using LDPC encoders to generate syndrome bits but at different rates for each case. A refinement coefficient in a bitplane is defined as a coefficient that has been coded as '1' in at least one of the most significant bitplanes. All other coefficients are called significant coefficient. The significance pass codes the absolute values 0/1 of all significant coefficients in the lth bitplane using an LDPC bitplane encoder with the corresponding SI s_k^l. Then, for all the 1s in the significance pass, their sign bits are coded with an appropriate LDPC encoder. The refinement coefficients are then coded. Different pass coding is employed due to their statistical difference between these two classes of coefficients. Note that the two coding passes follow the same motivation as the two entropy codes in ET, mentioned in Section 2.1. The channel rates can be easily estimated from the correlation between s_k^l and u_k^l. The coding mode information is sent to the decoder as well.

Decoding: Decoding of the X_k EL bitplanes proceeds by using the EL reconstruction of the previous frame $\hat{X}_{k-1}^e$ to form the SI for each bitplane. The syndrome bits received are used to decode the blocks in WZS mode. The procedure is the same as at the encoder, except that the original frame X_{k-1} is now replaced by the high quality reconstruction $\hat{X}_{k-1}^e$ to generate SI: $v_k = T(MC_k[\hat{X}_{k-1}^e] - \hat{X}_k^b)$. The decoder performs sequential decoding since each bitplane decoding requires the reconstruction from its most significant bitplanes.

Complexity analysis: The base layer structure remains unchanged from MPEG-4 FGS. An additional set of frame memory, MC and DCT modules is introduced at both the encoder and decoder. In comparison, the ET approach requires multiple MCP loops, each of which needs a separate set of frame memory, MC, DCT and IDCT modules. By considering the bitplane characteristic and the mode decisions from lower bitplanes, we may modify the mode decision rule in some conditions to avoid computing (6). For example, we do not use the WZS mode for the least significant bitplanes (there could be more than one), since they have low temporal correlation. If a MB has already been coded as a FGS-MB for a higher significant bitplane, it is unlikely to be coded in WZS-MB mode again. Therefore, the proposed approach can be implemented in reasonable complexity even for multiple layers.

[1]The WZS-SKIP mode may introduce some small errors due to the difference between the SI at the encoder and decoder.

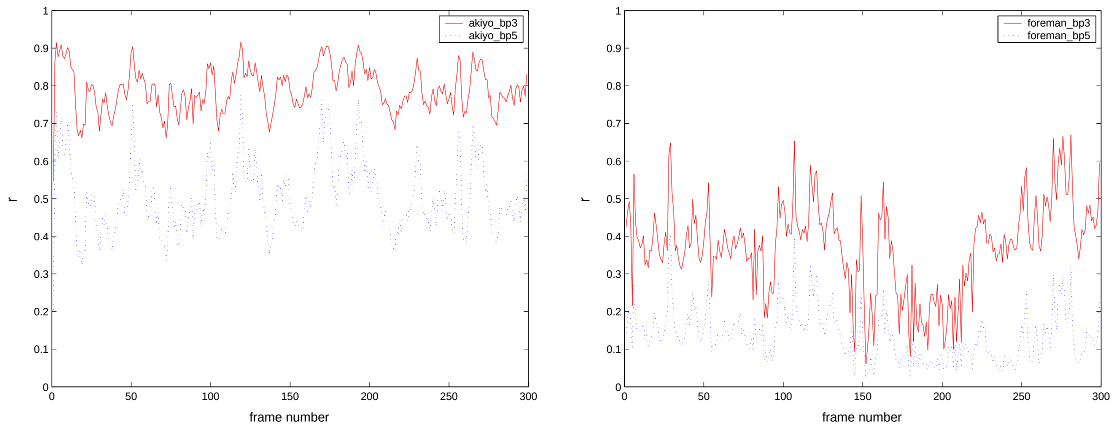


Fig. 3. Correlation coefficient for different frames and bitplanes for akiyo (left) and foreman (right) sequences. The base layer quantization parameter is 31.

3. CORRELATION ANALYSIS

Suppose the decoder has received the base layer and the $i-1$ most significance bitplanes of the current frame X_k. We define E_i as the information in the ith bitplane of X_k, and S_i as the information provided by the $i$the bitplane of motion-compensated prediction of X_k. They can be expressed as $E_i = \hat{X}_k^i - \hat{X}_k^{i-1}$ and $S_i = MC_k(\hat{X}_{k-1}^i) - \hat{X}_k^{i-1}$. Regarding E_i and S_i as two correlated random variables, we can compute their correlation coefficient[2], similar to what is proposed in [4], to measure the temporal correlation for each frame in the ith bitplane as a whole[3]

$$|r_i| = \frac{|\sum_{w=1}^{W}\sum_{h=1}^{H} E_i(w,h) S_i(w,h)|}{\sqrt{\sum_{w=1}^{W}\sum_{h=1}^{H}(E_i(w,h))^2 \sum_{w=1}^{W}\sum_{h=1}^{H}(S_i(w,h))^2}} \tag{7}$$

where W and H are the width and height of the frame respectively, $E_i(w,h)$ and $Y_i(w,h)$ are the mean-removed pixel values of E_i and S_i. Fig. 3 plots the values of r for two different bitplanes as function of frame number for various video sequences. In general, the correlation level decreases when the bitplane level gets higher. The small residuals coded in the higher EL layers mostly consist of noise. Meanwhile, the correlation between very low layers are not much useful as well in terms of coding since they only contain a few residuals with large magnitude and the main part of zero components can be well coded by FGS bitplane coding. It is also noted that different video sequences have large difference on their correlation level.

[2]Correlation coefficient r for two zero-mean random variables X and Y is defined as $r = \frac{E(XY)}{\sigma_X \sigma_Y}$. Here we regard every pixel in the image as a realization of the random variable.

[3]r_i is related to ρ in (1), but also takes into account the bitplane level.

4. EXPERIMENTAL RESULTS

We implement a WZS scalable video codec based on the Microsoft MPEG-4 FGS reference software. Several LDPC codes with different code rates and block sizes are developed for coding streams with different channel characteristics [12].

To evaluate the coding efficiency of the proposed WZS approach, initial experiments have been performed to compare WZS with MPEG-4 FGS and nonscalable coding. The results are given for two MPEG-4 test sequences *Akiyo* and *Foreman* at CIF-resolution, 30Hz, which have very different temporal correlation. For both sequences, the first frame is encoded as an I frame and all the other frames are encoded as P frames. Two different base layer quantization step sizes 31 and 8 are used to test the enhancement layer coding performance with different base layer rates. Fig. 4 shows the performance comparison between the proposed WZS, MPEG-4 FGS and nonscalable coder for various enhancement layer and base layer rates.

Note that the PSNR gain obtained by the proposed WZS approach depends greatly on the temporal correlation statistics of the video sequence. For the Akiyo sequence, which has higher temporal correlation as shown in Fig. 3, the PSNR gain of WZS is greater than that for the Foreman sequence. The proposed coder outperforms FGS up to 2dB for the sequences with high temporal correlation. Since the coding penalty of FGS for high-motion sequences, such as Foreman, over the nonscalable coding is already small, the coding gap between the proposed WZS coder and nonscalable coder may be very limited for almost all sequences. Another important point is that the coding gain of the WZS approach decreases if a higher quality base layer is used. That is because the temporal correlation between the successive frames is already well exploited by a high-quality

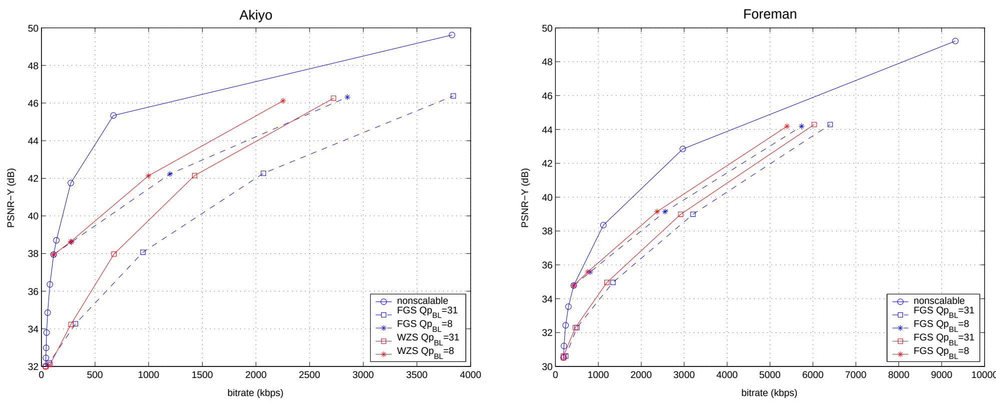


Fig. 4. Comparison between the proposed WZS, MPEG-4 FGS and nonscalable coding.

base layer. From Fig. 3, we can also see that the correlation degree drops with the least signifant bitplanes. These two observations are in agreement with analysis in Section 2.1

We also mention that we only use a small set of channel codes with several fixed code rates and block sizes in the current implementation, and this coarse level of rate control may result in some performance loss. A more flexible channel coder with easy rate adjustment is currently being investigated.

5. CONCLUSIONS AND FUTURE WORK

This paper presents a new practical scalable coding structure, based on the Wyner-Ziv principle, to improve quality in an FGS framework. The proposed coder achieves high coding efficiency by selectively exploiting the high quality reconstruction of the previous frame in the enhancement layer bitplane coding. An adaptive mode selection algorithm is proposed to switch between different coding modes based on the temporal correlation. Simulation results show much better performance over MPEG-4 FGS for sequences with high temporal correlation and limited improvement for high-motion sequences. A possible reason is due to the less accurate motion compensation prediction in the enhancement layer when sharing motion vectors with the base layer. It can be improved by exploring the flexibility at the decoder, an important benefit of the Wyner-Ziv coding, to refine the enhancement layer motion vectors by taking into account the received enhancement layer information from the previous frame. The current coder cannot really achieve the fine granularity scalability in that the LDPC coder can only decode the whole block at the bitplane boundary, and a more flexible coder should be developed, which is also mentioned in [10]. In addition, it is also interesting to evaluate the error resilience performance of the proposed coder. In principle, the Wyner-Ziv coding has more tolerance on noise introduced to the side information.

6. REFERENCES

[1] W. Li, "Overview of fine granularity scalalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 301 –317, Mar. 2001.

[2] K. Rose and S.L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Trans. Image Processing*, vol. 10, pp. 965–976, Jul. 2001.

[3] F. Wu, S. Li, and Y.Q. Zhang, "A framework foe efficient progressive fine granularity scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 332 –344, Mar. 2001.

[4] M. van der Schaar and H. Radha, "Adaptive motion-compensation fine-granular-scalability (AMC-FGS) for wireless video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 360 –371, Jun. 2002.

[5] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.

[6] R. Puri and K. Ramchandran, "PRISM: a new robust video coding architecture based on distributed compression principles," in *Proc. Allerton Conf. Communication, Control and Computing*, Oct. 2002.

[7] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE, Special Issue on Advances in Video Coding and Delivery*, 2004.

[8] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *Submitted to IEEE Trans. Inform. Theory*, Mar. 2003.

[9] W.H. Equitz and T.M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.

[10] Q. Xu and Z. Xiong, "Layered Wyner-Ziv video coding," in *Proc. VCIP'04*, Jan. 2004.

[11] H. Wang and A. Ortega, "Scalable predictive coding by nested quantization with layered side information," in *Proc. ICIP 2004*, Oct. 2004.

[12] http://www.cs.toronto.edu/ radford/ldpc.software.html.