Scalable Wyner-Ziv Video Coding with Adaptive Bit-Plane Representation

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

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

[2]Microsoft Research Asia, Beijing, 100080, China

ABSTRACT

Scalable Wyner-Ziv video coding is desirable, especially for mobile devices with limited computational resources and bandwidth. However, the conventional bit-plane representation used in hybrid video coding does not work well in the scenario of Wyner-Ziv coding. As we know, the bit-plane representation is closely related to the quantization. In this paper, a new bit-plane representation with optimal quantization at any bit-plane is proposed. In particular, for the DCT-domain Wyner-Ziv video coding, since the distribution of DCT coefficients and the conditional distribution given side information can be modeled with symmetric *Laplacian* functions, a simplified adaptive bit-plane representation is proposed without pre-knowing the *Laplacian* distribution parameters. Based on the proposed bit-plane representation, a DCT-domain scalable Wyner-Ziv video coding scheme is proposed, in which the encoding and the bit-stream can be flexibly truncated according to the available computational resources and the bandwidth. No coding performance loss is introduced due to truncation.

Keywords: Distributed video coding, Wyner-Ziv coding, bit-plane representation, optimal quantization

1. INTRODUCTION

With the popularity of various portable multimedia devices, the demand of real-time visual communication over wireless networks is increasing rapidly. However, the processor workloads to process the video material become a bottleneck. Distributed video coding (DVC) is committed to the compression of video with limited computational resources [1]. Different from hybrid video coding for example H.264/MPEG-4 AVC (Advanced Video Coding) standard [2], DVC technologies encode the source and the side information separately. The mutual correlation is exploited at the decoder side, so the task of motion compensation is shifted to the decoder. A low-complexity encoder is thus achieved.

The development of DVC technologies can be traced back to the Slepian-Wolf theorem [3] and the Wyner-Ziv theorem [4]. Slepian-Wolf theorem suggests that two correlated signals can be separately encoded with the same rate as that of joint encoding as long as the collaborative decoders are employed [3]. As for lossy compression, when the encoder does not know the side information, a bounded rate loss is incurred [4] [5]. In the past, a number of practical DVC technologies have been developed [6] ~ [10]. Similar to conventional video coding, temporal, spatial and statistical correlations are utilized in DVC. The temporal correlation in a video is usually utilized by generating the side information frame at the decoder with interpolation [1], hash-based motion estimation [6] or CRC-based motion estimation [7]. The spatial redundancy in a frame is removed by applying Wyner-Ziv coding into transform domains such as DCT [8] or wavelet [9]. Finally, the channel coding techniques such as turbo code [1], LDPC code [10] and syndrome coding [7] are used to exploit the statistical correlation. The Wyner-Ziv bit stream is robust to the packet or frame loss caused by channel transmission errors, since it is generated independently of the side information [1]. Thanks to the separate encoding, Wyner-Ziv coding has been adopted to compensate the mismatch between encoder and decoder existing in hybrid video coding. A layered video coder, which produces the base layer with hybrid video coding for example H.26L and the enhancement layer with Wyner-Ziv coding based on nested scalar quantization, is proposed in [11]. In [12], Wyner-Ziv coding is utilized to produce the switching stream for multiple bit-rate video streaming.

Besides low complexity encoding, scalability is also desirable for Wyner-Ziv coding, especially for mobile devices with limited computational resources and bandwidth. When the computational resources are not enough, the encoder may stop at any coding status. Moreover, when the available bandwidth is not enough, the produced bit stream may also be cut to a certain length. In any case, it is desirable to have the optimal rate-distortion performance given a rate.

Intuitively, bit-plane representation provides an inherent solution to scalable coding. However, the conventional bit-plane representation used in hybrid video coding does not work well in the scenario of scalable Wyner-Ziv coding. In hybrid video coding, the residue between source and side information is entropy coded directly, so the bit-plane representation could be designed to achieve optimum rate-distortion according to the distribution of quantized residue signal. In Wyner-Ziv coding, the source signal is encoded individually but decoded depending on the side information and the mutual correlation. As we know, the bit-plane representation of source signal is closely related to the quantization. In [13], the optimal quantization of Wyner-Ziv coding has been studied for the general source, which shows that the quantization in Wyner-Ziv coding may not be identical to the joint coding. Therefore, in scalable Wyner-Ziv coding, the bit-plane representation should also be re-designed by considering the optimal quantization.

In this paper, the bit-plane representation with optimal quantization at any bit-plane is considered in terms of scalable Wyner-Ziv coding. Actually, the bit-plane based coding is quite similar to the encoding with different quantization step sizes. The bit-plane representation should be associated with the partition method of quantization bins at each bit-plane. The bit-planes are adaptively produced according to the joint distribution of the source and the side information. A method which can achieve bit-plane representation with optimum quantization at any bit-plane is generally presented in terms of ideal *Laplacian* signal. For the DCT-domain Wyner-Ziv video coding, since the distribution of DCT coefficients in source frame and the distribution of difference between the source and the side information can be modeled as *Laplacian* functions, a simplified adaptive bit-plane representation is proposed without pre-knowing the distribution model parameters. Based on the proposed adaptive bit-plane representation, a scalable Wyner-Ziv video coding scheme is proposed, in which the encoding and the bit-stream can be truncated according to the available computational resources and the bandwidth. No coding performance loss is introduced due to the truncation.

The rest of this paper is organized as follows. Section 2 analyzes the problem of traditional bit-plane representation and presents the proposed representation with optimal quantization for scalable Wyner-Ziv coding. Section 3 describes the proposed scalable Wyner-Ziv video coding. Section 4 gives some experimental results. Finally, this paper is concluded in Section 5.

2. ADAPTIVE BIT-PLANE REPRESENTATION

	A(-11)	B(75)	C(-6)	D(-28)	E(2)	
B_0	-	+	-	-	+	
B_1	0	1	0	0	0	
B_2	0	0	0	0	0	m
B_3	0	0	0	1	0	
B_4	1	1	0	1	0	n
B_5	0	0	1	1	0	
B_6	1	1	1	0	1	
B_7	1	1	0	0	0	

Figure 1: Illustration of the traditional bit-plane representation.

In this section, we will analyze the problem of traditional bit-plane representation at first. Then, a bit-plane representation to achieve optimum quantization at any bit-plane level is generally presented in terms of ideal *Laplacian*

source signal. Particularly, for DCT-domain Wyner-Ziv video coding, a simplified adaptive bit-plane representation is proposed without the knowledge of distribution parameters.

2.1 Analysis of Rate-Distortion Penalty with Traditional Bit-plane Representation

After the uniform scalar quantization of zero-mean *Gaussian* or *Laplacian* signal, a sequence of quantization indices with signs is generated. We take the sign bit as an example to explain the rate penalty. In bit-plane based hybrid video coding, the residue between source and side information is directly entropy coded. The sign bit is put immediately before the first significant bit of the magnitude. However, in DVC, the source is encoded individually without knowing the side information. Since the decoder cannot identify the first significant bit of the magnitude with the bit-plane representation in hybrid coding, the encoder has to code the sign bits either before or after the coding of all other bit-planes, which is denoted as traditional bit-plane representation.

Figure 1 shows a set of symbols with their traditional bit-plane representation. Let B_0 represent the bit-plane of sign bits. The coding is performed in a top-down manner. For simplicity, we can analyze the rate penalty in terms of sign bits. When all bit-planes are encoded and transmitted, the rate-distortion performance will not change too much compared with the non-scalable coding. However, when the encoding is stopped or the bit-stream is truncated at a certain bit-plane (e.g., B_2 in Figure 1), the transmitted sign bits contribute little to the distortion reduction but increase the rate, if the corresponding magnitudes are zero (e.g., symbols A, C, D and E in Figure 1). It should be noted that the sign bits will not be transmitted if these magnitudes are pre-quantized to zero at B_2 in scalable Wyner-Ziv coding.

2.2 Proposed Adaptive Bit-plane Representation

The above example shows that traditional bit-plane representation results in rate-distortion penalty in scalable Wyner-Ziv coding. Actually, the bit-plane representation is closely related to the quantization. The bit-plane representation divides the range of source from 2 to 2^k bins when 2^k-level uniform pre-quantization is adopted. At each bit-plane level, the source is partitioned into uniform bins. The size of bin decreases from the most significant bit-plane to the least significant one. A bin achieved at a certain bit-plane B_{t-1} will be halved at the next bit-plane B_t. In Wyner-Ziv coding, different ways can be used to divide one bin into two. The design of bit-plane representation should be associated with optimal quantization. Actually, it is unnecessary to make a bin at a certain bit-plane always cover a continuous range, because the final reconstruction is also determined with knowledge of side information. As depicted in Figure 2, one bin from a to b is divided into two bins with identical size. One is composed of [a, l] and [r, b] while the other one is [l, r). r which can be any value from $(a+b)/2$ to b, is specified in order to achieve optimum rate-distortion performance.
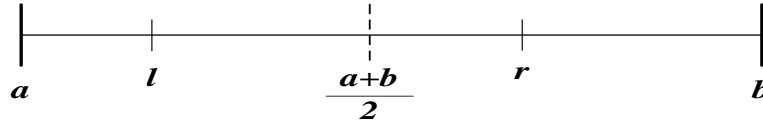

Figure 2: Bin partition at bit-plane level.

Denote X as the source signal and Y as the side information. x and y are referred to as their elements at the same instant, respectively. Assuming X and Y are quantized in the same way, let Q_x and Q_y be the discrete variables of the quantized source signal and the quantized side information respectively. When the ideal coding in which the encoding bit rate can approach the entropy bound can be achieved, R denotes the needed bits for each element. In lossy coding, R is equal to the conditional entropy $H(Q_x/Y)$. The Wyner-Ziv decoder estimates the quantized version of the source signal based on the side information and their mutual correlation, which is always modeled as a *Laplacian* function in Wyner-Ziv video coding. If y is located at the same bin as x, the estimation is accurate and no Wyner-Ziv bits are needed. However, if x and y belong to different bins, Wyner-Ziv bits are required to be delivered for the purpose of correcting the errors at the decoder. Without loss of generality, a variable V_q is introduced and defined as follows

$$V_q = \begin{cases} 1 & q_x = q_y \\ 0 & q_x \neq q_y \end{cases} \tag{1},$$

where q_x and q_y are the quantized elements x and y respectively. Thus,

$$R = H(Q_x | Y) = H(V_q) = P_0 \log\left(\frac{1}{P_0}\right) + P_1 \log\left(\frac{1}{P_1}\right) \tag{2}.$$

P_0 and P_1 represent the probability that V_q takes the value of 0 and 1 as follows

$$P_0 = \sum P(q_x \neq q_y) \quad \text{and} \quad P_1 = \sum P(q_x = q_y) \tag{3},$$

Consequently, R can be minimized by maximizing P_1, namely, the sum of probabilities that x and y are located in the identical bin. With regard to the distortion D, no matter which quantization method is utilized there exists $D \approx q^2/12$ when the source is encoded with high bit rate. q denotes the quantization step size. While the source is encoded with low bit rate, the reconstructed value is achieved to minimize the distortion based on both the side information and the quantization bin.

According to the above criteria, the optimum partition method can be derived from the joint distribution of the source signal and the side information. Here, we take the *Laplacian* signal as an example. Assume that both the distribution of side information Y and the conditional probability distribution of source X given Y have *Laplacian* functions, the parameters of which are u_y and u_e respectively. At a certain bitplane, the bin to be partitioned is $[-N, +N]$. Two bins with identical size are generated. One includes the single cell $[-N+r, r)$, while the other one may contain two discontinuous cells $[-N, -N+r)$ and $[r, +N)$. Thus,

$$P_1 = \int_{-N}^{-N+r} \int_{-N}^{-N+r} p(x,y)dxdy + \int_{-N+r}^{r} \int_{-N+r}^{r} p(x,y)dxdy + \int_{r}^{N} \int_{r}^{N} p(x,y)dxdy$$

$$= 1 - \exp(-u_y N)$$

$$+ \frac{1}{4} \frac{u_y}{u_y - u_e} \left\{ \begin{array}{l} \exp(-u_y N)\left[\exp(u_y r) - \exp(u_y r - u_e r)\right] \\ + \exp(-u_e N)\left[-\exp(u_e r) - \exp(-u_y r + u_e r)\right] \\ + \exp(-u_y r) - \exp(-u_e r) + 2\exp(-u_y N) \end{array} \right\} \tag{4}.$$

$$+ \frac{1}{4} \frac{u_y}{u_y + u_e} \left\{ \begin{array}{l} \exp(-u_y N - u_e N)\left[\exp(u_y r) + \exp(u_e r)\right] \\ + \exp(-u_y N)\left[\exp(-u_e r) - \exp(u_y r)\right] \\ + \exp(-u_e N)\left[\exp(-u_y r) - \exp(u_e r)\right] - \exp(-u_e r) - \exp(-u_y r) \end{array} \right\}$$

r, existing in $[0, N)$, is determined by maximizing P_1. Obviously, r relies on u_e, u_y and N. When $u_e = u_y$,

$$r = \arg\max(P_1) = \frac{N}{2} \tag{5},$$

which is illustrated in Figure 3(a). Otherwise, at several most significant bit-plane levels where N is large, the maximum P_1 will appear at the point of $N/2$ as depicted in Figure 3(b). As N is decreased, the point at $N/2$ will decline more rapidly than the point at zero as shown in Figure 3(c) and Figure 3(d).

There are some other cases where the bin to be partitioned exist on the left or the right of the zero point, for example $[a, a+2N]$ with $a > 0$ and $[b-2N, b]$ with $b < 0$. We take the first case as an example. One of the two divided bins may be composed of two discontinuous cells $[a, -N+r)$ and $[r, a+2N)$ while the other one contains one single cell $[-N+r, r)$. Hence,

$$P_1 = \int_a^{-N+r} \int_a^{-N+r} p(x,y)dxdy + \int_{-N+r}^r \int_{-N+r}^r p(x,y)dxdy + \int_r^{a+2N} \int_r^{a+2N} p(x,y)dxdy$$

$$= \frac{1}{2}\exp(-u_y a) - \frac{1}{2}\exp(-u_y a - 2u_y N)$$

$$+ \frac{1}{4}\frac{u_y}{u_y+u_e}\left\{ \begin{array}{l} \exp[u_e(a+N-r)+u_y(N-r)] + \exp[u_e(-a-2N+r)+u_y(-a-2N)] \\ +\exp(-u_e N-u_y r)-\exp[u_y(N-r)]-\exp(-u_y a)-\exp(-u_y r) \end{array}\right\}$$

$$+ \frac{1}{4}\frac{u_y}{u_y-u_e}\left\{ \begin{array}{l} -\exp[u_e(N-r+a)-u_y a]-\exp[u_e(r-a-2N)-u_y r] \\ -\exp[-u_e N+u_y(N-r)]+\exp[u_y(N-r)]+\exp[-u_y(a+2N)]+\exp(-u_y r) \end{array}\right\}$$

$$(6).$$

r lies in $[a+N, a+2N)$ and is specified by maximizing P_1. Here, r relies on u_e, u_y, a and N. But the maximum P_1 will always appear at the point of $a+N$, as illustrated in Figure 4. In a word, for the symmetric case, the partition at several most significant bit-plane levels should make the part around the zero point in one bin and two discontinuous parts on both sides in one bin. In the asymmetric case, each bin just needs to be partitioned at the middle point.
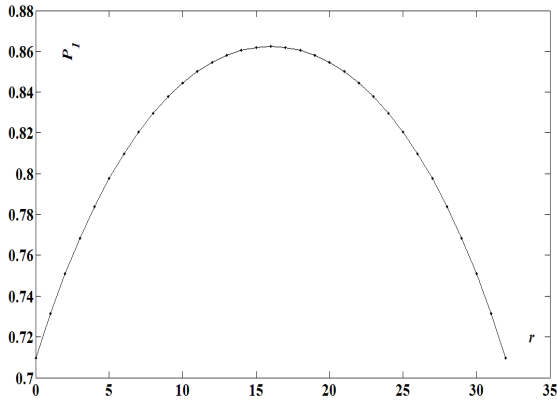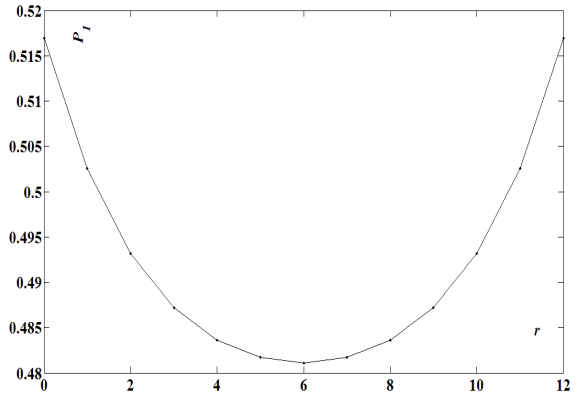

(a) $u_e = u_y = 0.1$, $N = 32$
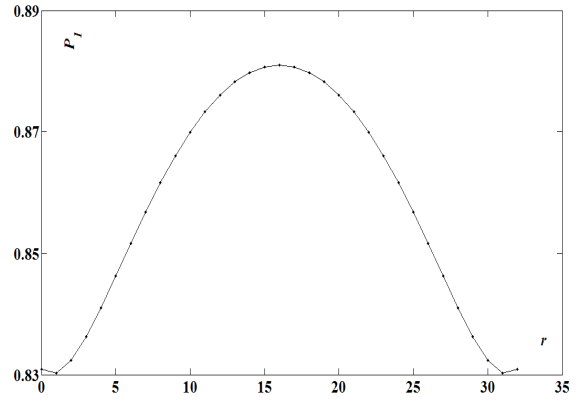
(b) $u_y = 0.1$, $u_e = 0.32$, $N = 32$
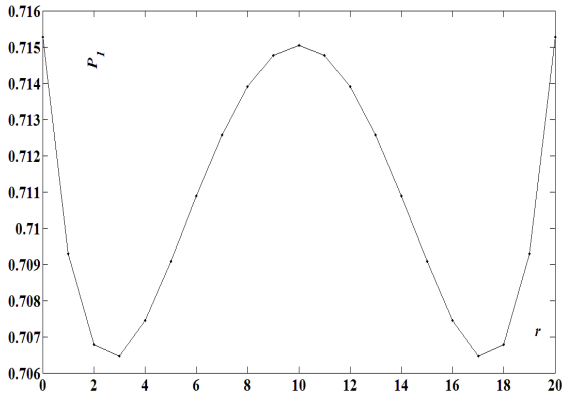
(c) $u_y = 0.1$, $u_e = 0.32$, $N = 20$

(d) $u_y = 0.1$, $u_e = 0.32$, $N = 12$

Figure 3: The selection of r to maximize P_1 for symmetric case.

(a) $u_e = u_y = 0.1$, $a = 16$, $N = 16$

(b) $u_y = 0.1$, $u_e = 0.32$, $a = 16$, $N = 4$

(c) $u_y = 0.1$, $u_e = 0.32$, $a = 16$, $N = 8$
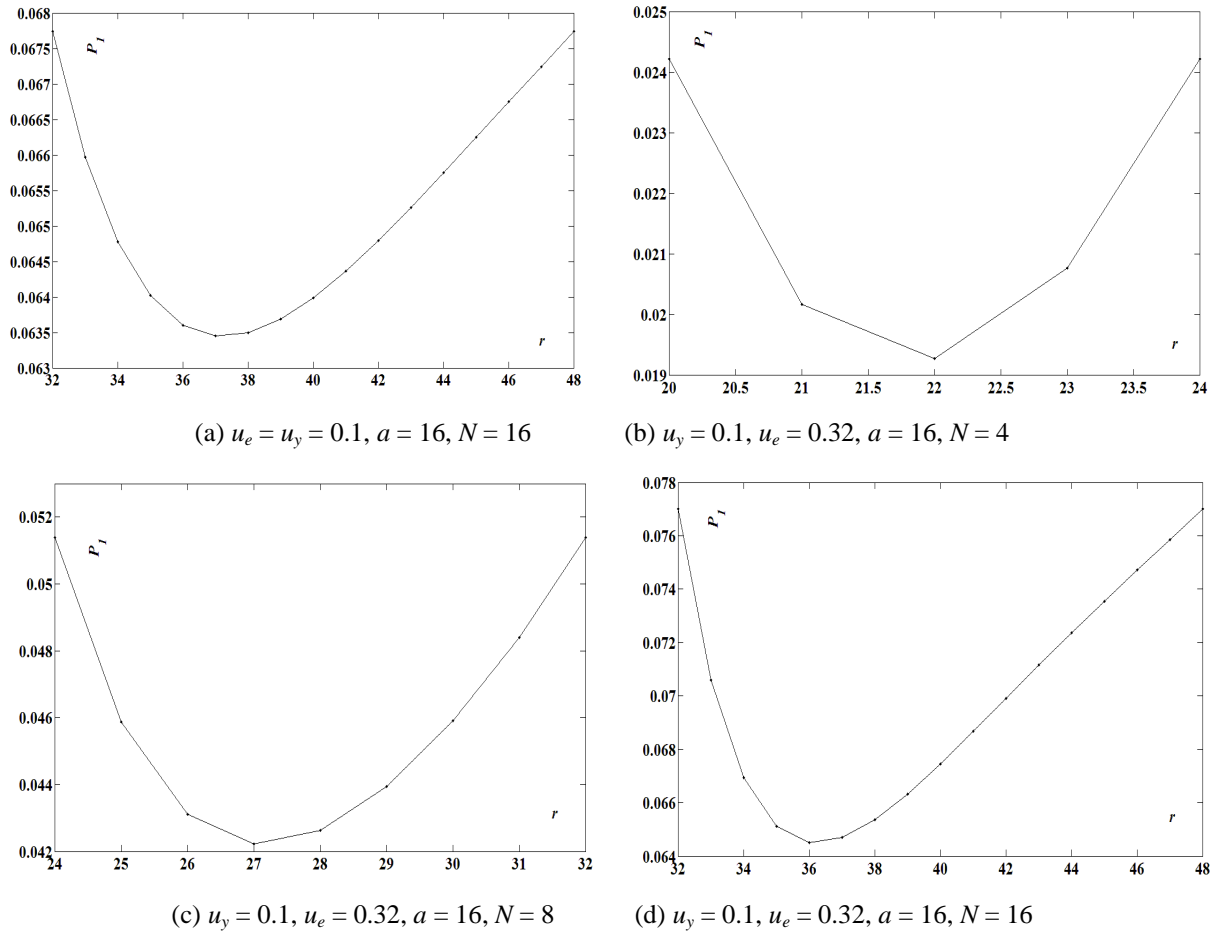
(d) $u_y = 0.1$, $u_e = 0.32$, $a = 16$, $N = 16$

Figure 4: The selection of r to maximize P_1 for asymmetric case.
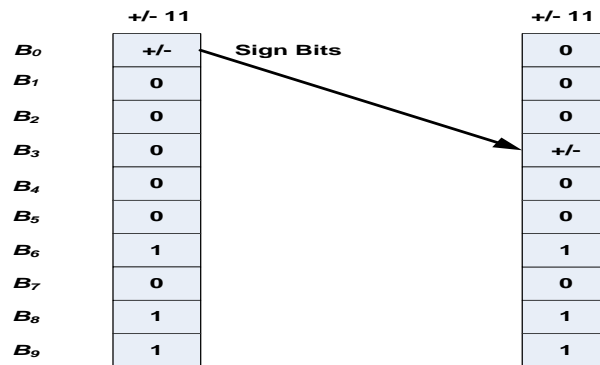

Figure 5: Bit-plane representation: traditional bit-plane representation (left); proposed bit-plane representation(right).

In Wyner-Ziv coding, the correlation is explored at the decoder. At each bit-plane level, the decoder estimates the partitioned bins of source signal based on the side information. The bit-plane representation should be consistent with the optimum partition method. The traditional bit-plane representation puts the sign bit at the most significant bit-plane as depicted in Figure 5. In this way, the range of the source is first partitioned at the middle point as shown in Figure 6(a). The most significant bit-plane, namely the sign bit, will distinguish positive symbols from negative ones. According to our previous discussion, the division at the middle point incurs larger rate compared to the division depicted in Figure 6(b). Actually, with the statistical distribution parameters u_y and u_e, the optimum partition method can be calculated at

each bit-plane level. Different bit-plane levels have different a and N. There are two choices for the bin around the zero point depending on N. At several most significant bit planes where the peak of P_1 appears at the point of r equal to $N/2$, each symbol should be first distinguished as significant one or non-significant one, as depicted in Figure 6(b). Then, as shown in Figure 6(c), the significant set is further classified into positive significant set and negative significant set with the sign bit. And the non-significant set is partitioned recursively as illustrated in Figure 6(d). In top-down manner, as N decreases, the peak of P_1 moves from the point of r equal to $N/2$ to the one of r equal to zero corresponding to Figure 3(c) and (d). At a certain bit-plane level B_k which has optimum r equal to zero, the non-significant set should be partitioned into positive set and negative set at the zero point in order to minimize rate. Accordingly, the sign bit should be located at bit-plane B_k, as illustrated in Figure 5. As for the positive and negative significant sets located on one side, they are divided in the middle.
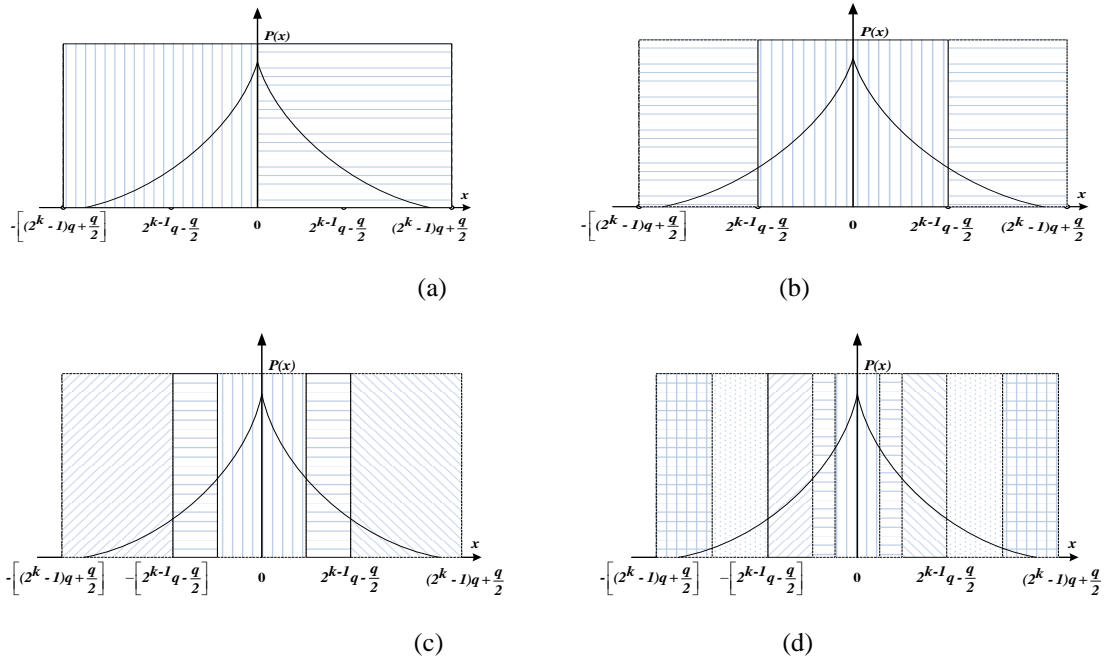

Figure 6: The partition of the quantization bins. (a) B_0 (traditional bit-plane representation); (b) B_0 (optimum); (c) B_1 (optimum); (d) B_2 (optimum). Each filling pattern of line indicates one quantization bin.

2.3 Simplified Adaptive Bit-plane Representation for Scalable Wyner-Ziv Video Coding
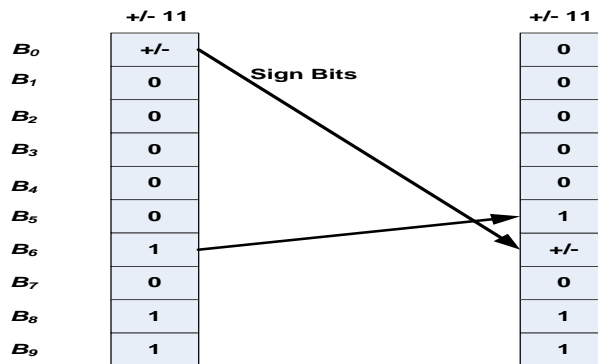

Figure 7: A simplified adaptive bit-plane representation.

There are many other cases where the distribution of the side information is other than some general functions such as *Laplacian* or *Gaussian*, for example an image or a video frame in pixel domain. The maximum P_1 may be achieved with different values of *r* depending on the joint distribution of the source and the side information. Besides the generated bit-planes, some additional information that indicates the position of division needs to be delivered. When DCT transform is applied to an image or a frame from video data, the coefficients of side information always have *Laplacian* distribution [14]. In Wyner-Ziv video coding, the side information is not available at the encoder, so the conditional distribution cannot be known definitely. But the distribution of difference between source and side information can be modeled with a *Laplacian* function. We propose a simplified adaptive bit-plane representation for scalable Wyner-Ziv video coding. As illustrated in Figure 7, the first non-zero bit is shifted left by one bit and followed by the inserted sign bit.

Given one coefficient, the sign bit will be encoded and transmitted after it is identified as a significant coefficient. If the encoding will stop at bit-plane B_t, only the sign bits of the coefficients that have been determined as significant ones with the bit-planes from B_0 to B_{t-1} are delivered. The coefficients that are just distinguished as significant ones at bit-plane B_t are still considered as non-significant ones, since there will be no chance to send the information of their signs. At the decoder, all the coefficients are assumed to be non-significant ones before decoding. After decoding one bit-plane, some coefficients are known to be significant ones. Their next bit-plane will indicate the sign information. The non-significant coefficients are still to be distinguished as significant ones or non-significant ones. After all the bit-planes are decoded, all non-zero quantized coefficients belong to the significant set.

We take the second frame from News sequence at QCIF format as an example. The side information is generated by interpolating two neighboring reconstructed frames with the symmetric motion estimation algorithm. The quality of the side information in terms of PSNR is 32.07dB. Table 1 gives the probabilities that the coefficients of the source and the corresponding ones of the side information exist in the same bin. The conditional entropy $H(B_0, B_1, \cdots, B_t / Y)$ at each bit-plane level *t* is calculated as

$$H(B_0, B_1, \cdots, B_t \mid Y) = \sum P(B_0, B_1, \cdots, B_t, y) \log \frac{1}{P(B_0, B_1, \cdots, B_t, y)} \qquad (7).$$

It can be observed from Table 1 that larger P_1 leads to less conditional entropy. Except B_5, P_1 achieved by the proposed bit-plane representation is larger than that achieved by the traditional manner. Hence, the bit rate is reduced. The overall coding efficiency will be shown in the Section 4.

Table 1. P_1 and $H(B_0, B_1, \cdots, B_t/Y)$ at each bit-plane level with two bit-plane representation methods.

Bitplane	Proposed Method		Traditional Method	
t	P_1	H	P_1	H
0	0.9994	0.0016	0.9462	0.2586
1	0.9977	0.0054	0.9456	0.2602
2	0.9942	0.0141	0.9438	0.2643
3	0.9803	0.0575	0.9386	0.2771
4	0.9402	0.2329	0.9205	0.3289
5	0.8346	0.6470	0.8788	0.5050
6	0.7981	0.8156	0.7981	0.8156

3. PROPOSED SCALABLE WYNER-ZIV VIDEO CODING

Figure 8 shows the proposed scalable Wyner-Ziv video coding framework, which is similar to that in [1]. However, the proposed adaptive bit-plane representation is employed to enhance the coding efficiency. In each group of pictures (GOP), the first frame is compressed with intra coding. And other frames are Wyner-Ziv coded. We mainly focus on the Wyner-Ziv coding in this paper. After 4x4 DCT transform, the coefficients of the source frame *X* are scalar quantized into 2^k uniform bins with 2^k-level quantizer. The bit-planes of the quantized coefficients *I* are arranged with our proposed representation, which has been introduced in the Section 2. The generated bit-planes denoted as $B_0, B_1, \cdots, B_{k-1}$, are fed into turbo code based Slepian-Wolf encoder in top-down manner. Each bit-plane is encoded individually. The

complexity controller informs the Slepian-Wolf encoder whether the encoding of the current bit-plane can be finished according to the available computational resources and/or the bandwidth.
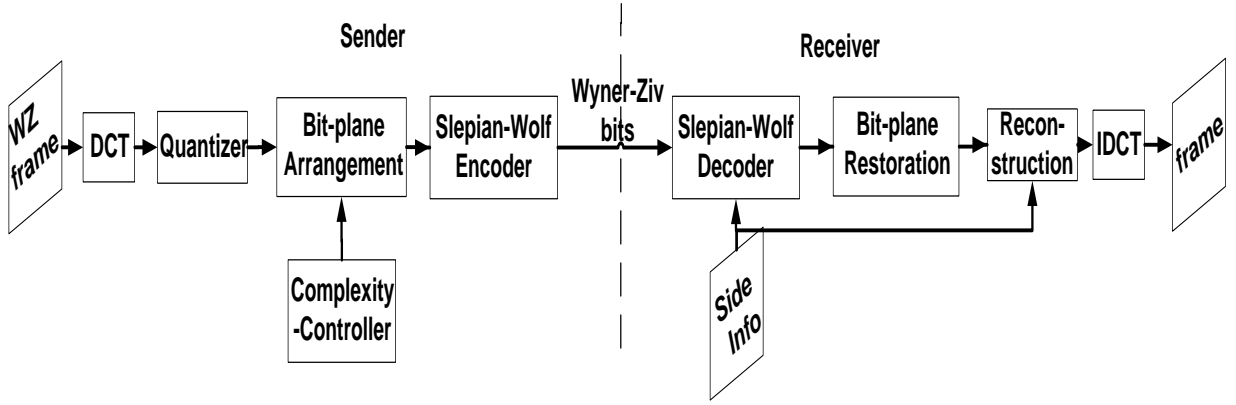

Figure 8: Bit-plane based scalable Wyner-Ziv video coding.

At the decoder side, the side information Y is created by interpolating the adjacent reconstructed frames with symmetric motion estimation. Although the generated side information may not be accurate in some non-linear areas, the Wyner-Ziv bits will be used to compensate for the loss. After the same 4x4 DCT transform, the coefficients of the side information are fed into Slepian-Wolf decoder in conjunction with the received Wyner-Ziv bits. The decoding of one bit-plane is based on the judgement of posterior probability (*PP*). Given a possible value j equal to zero or one, *PP* is expressed as

$$PP = \sum_{\chi_j} \alpha_{i-1}(s')\gamma_i(s',s)\beta_i(s)$$
, (8)

where χ_j is the set of all transitions from state s' to s with the input j. The probability functions $\alpha_i(s)$ and $\beta_i(s)$, which are the probabilities of approaching the state s at the instant i from the starting state and the ending state respectively, can be recursively calculated from the probability $\gamma_i(s', s)$ as introduced in [15].

Given one bit-plane, the decoding exploits both the correlation with the side information and with the previously decoded bit-planes. Assuming B_t is currently being decoded, the transitional probability is represented as

$$\gamma_i(s',s) = P(j)P(j \mid y_i, B_0, B_1, \cdots, B_{t-1})P(u_i \mid p_i)$$
, (9)

where u_i is the output parity bit of the transition from state s' to s with the input j. y_i and p_i represent the corresponding side information and the received parity bit. The conditional probability $P(j \mid y_i, B_0, B_1, \ldots, B_{t-1})$ can be calculated numerically as the probability of the difference between the estimated coefficient and the side information. The distribution of difference can be well fitted by a *Laplacian* function. The estimated value of current coefficient is chosen from the bin which is specified by possible value j of current bit-plane together with the previously decoded bit-planes from B_0 to B_{t-1}. The assignment of the partitioned bins at a certain bit-plane level relies on the bit-plane arrangement method.

The decoded bit-planes are organized to restore the quantized coefficients I. The quantized coefficients and the side information are then used to reconstruct the coefficient X' as $E(X'|Y,I)$. After the inverse DCT transform, the source frame is reproduced at the decoder side. If only the Wyner-Ziv bits generated from the first few bit-planes are received, the reconstruction value will be selected in the specified bin with the knowledge of the side information. The distortion is bounded by the size of the partitioned bin achieved by the decoded bit-planes.

4. EXPERIMENTAL RESULTS

The proposed scalable Wyner-Ziv coding approach has been tested on a number of video sequences. The results of Foreman, Akiyo, News and Salesman (100 frames at QCIF format) are shown in Figure 9. We set the GOP structure as IWIW, and code the intra frames with H.264. Symmetric motion estimation is used to generate the side information frames. The same side information frames for each sequence are used to test the different coding schemes. The average

PSNRs of side information frames for Foreman, Akiyo, News and Salesman are 29.30dB, 29.79dB, 29.90dB and 29.48dB, respectively. Only the results of Wyner-Ziv frames are included in Figure 9.

First, the non-scalable Wyner-Ziv coding with pre-quantization is evaluated. The curves of TRA_NO_TRUNC and PRO_NO_TRUNC indicate the coding with the traditional and the proposed bit-plane representation, respectively. The quantization step sizes are equal to 4, 8, 16 and 32 correspondingly. It can be observed that both schemes have similar reconstruction quality with the identical pre-quantization step size, but the proposed approach results in lower rate. In theory, the rates from the two approaches with the same pre-quantization step should be the same if ideal Slepian-Wolf coding can be performed. The superior performance of the proposed approach lies in that it improves the correlation between the source at each bit-plane and the side information.

Then, the scalable Wyner-Ziv coding with bit-stream truncation is evaluated. Similarly, the curves of TRA_TRUNC and PRO_TRUNC indicate the coding with the traditional and the proposed bit-plane representation, respectively. The scalar uniform quantization step size equal to 4 is adopted. The performance, when 0, 1, 2, and 3 bit-planes are discarded due to limited resources, is given respectively. It can be observed that the performance of traditional bit-plane representation falls greatly when the truncation happens. Thanks to the proposed adaptive bit-plane representation, there is almost no rate-distortion penalty compared with non-scalable coding. There are many paths from the point achieved by the side information to the high bit rate end where all the bit-planes are encoded and transmitted to the decoder. It can be observed from Figure 9 that our scheme produces a better path compared to the scheme adopting the traditional bit-plane representation. In our scheme, when the truncation happens at the first several bit-plane levels, lower bit rate is used to correct the serious errors in the side information. In summary, the proposed bit-plane representation with optimal quantization at each bit-plane works well for both scalable and non- scalable Wyner-Ziv video coding.
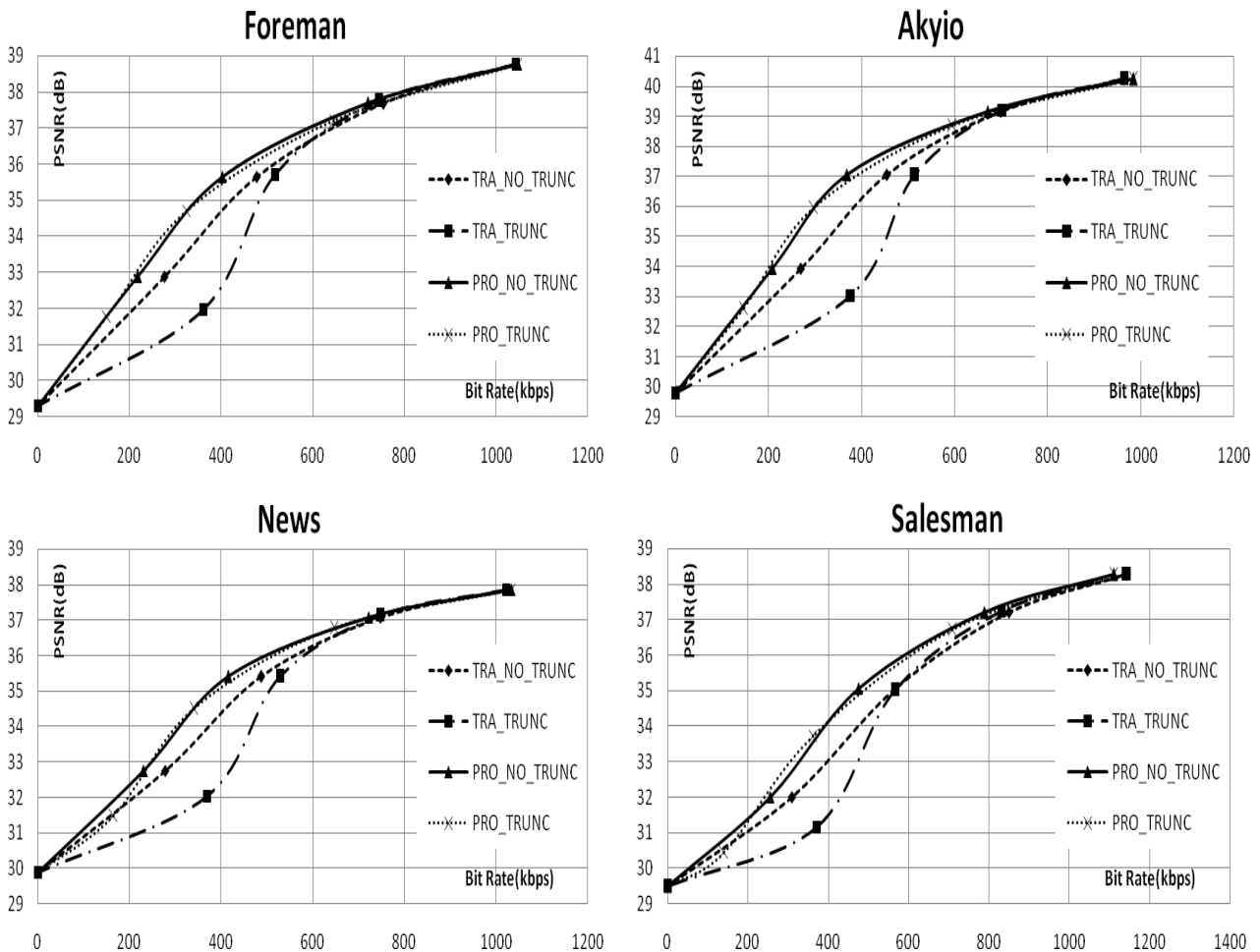

Figure 9: The coding performance of luminance component.

5. CONCLUSIONS

In this paper, we have proposed a scalable Wyner-Ziv video coding scheme with adaptive bit-plane representation. The encoding process or the bit-stream can be truncated at a certain bit-plane according to the available computational resources and/or the bandwidth. The optimal bit-plane representation is designed by considering the quantization issues in Wyner-Ziv video coding. For ideal source, the bit-plane representation can be determined with the distribution of side information and the conditional distribution of source given side information. Furthermore, a simplified method is proposed to be used in DCT-domain Wyner-Ziv video coding scheme.

6. REFERENCES

1. B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding", *Proc. IEEE*, vol. 93, Jan. 2005.

2. T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, Jul. 2003.

3. J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471-480, Jul. 1973.

4. A. D. Wyner, and J. Ziv, "The Rate-distortion function for source coding with side information at the decoder", *IEEE Trans. Inform. Theory*, vol. IT-22, Jan. 1976.

5. R. Zamir, "The rate loss in the Wyner-Ziv problem", *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 2073-2084, Nov. 1996.

6. A. Aaron, S. Rane, and B. Girod, "Wyner–Ziv video coding with hash-based motion compensation at the receiver", *Proc. IEEE Int. Conf. on Image Processing*, vol. 5, pp. 3097- 3100, Oct.2004.

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

8. A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video", *Visual Communications and Image Processing, Proc. of SPIE VCIP*, Jan. 2004.

9. X. Guo, Y. Lu, F. Wu, W. Gao, "Distributed video coding using wavelet", *Proc. IEEE Int. Symposium on Circuits and Systems*, pp. 5427-5430, May 2006.

10. Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes", *Proc. Asilomar Conf. Signals, Systems and Computers*, pp. 825-829, Nov. 2003.

11. Q. Xu, and Z. Xiong, "Layered Wyner-Ziv video coding", *IEEE Trans. Image Processing*, vol. 16, no. 12, pp. 3791-3803, Dec. 2006.

12. M. Guo, Y. Lu, F. Wu, D. Zhao, W. Gao, "Practical Wyner-Ziv switching scheme for multiple bit-rate video streaming", *Proc. IEEE Int. Conf. on Image Processing*, pp. 249-252, 2006.

13. D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding", *Proc. IEEE Data Compression Conf.* , pp. 13-22, Mar. 2003.

14. R. Reininger and J. Gibson, "Distributions of the two-dimensional DCT coefficients for images", *IEEE Trans. Communication,* vol. 31, no. 6, pp 835-839, Jun. 1983.

15. C. Berrou, A, Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-Codes," *Proc. IEEE Int. Conf. on Communications*, pp. 1064-1070, May 1993.