Rate Allocation for Transform Domain Wyner-Ziv Video Coding without Feedback

Tao Sheng\(^1\), Guogang Hua\(^2\), Hongxing Guo\(^1\), Jingli Zhou\(^1\), Chang Wen Chen\(^3\)

\(^1\)College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China
\(^2\)Qualcomm Inc., 5775 Morehouse Drive, San Diego, CA 92121 USA
\(^3\)Department of Computer Science and Engineering, University at Buffalo, State University of New York Buffalo, NY, 14260, USA

taosheng@acm.org, ghua@qualcomm.com, {wtguohx, jlzhou}@hust.edu.cn, chencw@buffalo.edu

ABSTRACT

In this paper, we propose a new rate allocation algorithm for transform domain Wyner-Ziv video coding (WZVC) without feedback. In contrast to conventional video coding, Wyner-Ziv video coding aims to design simple intra-frame encoding and complex inter-frame decoding based on the Slepian-Wolf and Wyner-Ziv distributed source coding theorems. To allocate proper number of bits to each frame, most existing Wyner-Ziv video coding solutions need a feedback channel (FC) at the decoder. However, in many video coding applications, the FC is not allowed. Moreover, the FC will introduce latency and an increase of decoder complexity because several iterative decoding operations may be needed to decode the data to achieve target video quality. The proposed algorithm predicts the number of bits for each Wyner-Ziv frame at the encoder as a function of the coding mode and the quantization parameters. Such predictions will not significantly increase the complexity at the encoder. However, the prediction will be able to properly select the best mode and quantization parameter for encoding each Wyner-Ziv frame. Experimental results show that the proposed algorithms is able to achieve good encoder rate allocation while still maintains consistent coding efficiency. Comparing to the WZVC coder with FC, this new WZVC coder without FC induces only a small loss in Rate-Distortion performance.

Categories and Subject Descriptors
I.4 [Image Processing and Computer Vision]: Compression

General Terms
Algorithms, Experimentation

Keywords
Rate Allocation, Wyner-Ziv Video Coding, Feedback Channel.

1. INTRODUCTION

Current video compression standards are based on inter-frame predictive coding in order to exploit temporal correlation between successive frames. Since predictive coding makes use of motion estimation, the video encoder is typically several times more complex than the decoder even when efficient fast motion search is employed.

However, there are numerous contemporary video applications, such as wireless video surveillance and multimedia sensor networks, that require low-complexity encoders. Wyner-Ziv video coding is a new type of lossy video compression method with side information available to the decoder only. From the Slepian-Wolf theory [2] for lossless distributed source coding and Wyner-Ziv theory [1] for lossy distributed source coding, we know that such a coding system with intra-frame encoder and inter-frame decoder is able to approach the efficiency of the conventional predictive coding.

Wyner-Ziv video coders can be classified into transform domain coders and pixel domain coders [3]. In this paper, we focus on transform domain WZVC since current WZVC is able to offer better rate distortion (RD) performances.

One of the most challenging tasks in WZVC is in the allocation of proper number of bits to encoding each video frame. This is mainly because the encoder often has no access to the motion compensation information at the decoder. Moreover, since WZVC is implemented via channel codes such as turbo codes [4] and low-density parity-check (LDPC) codes [5], small variations in the number of bits allocated to each frame may cause substantial degradation in RD performance. Most practical Wyner-Ziv video coders [3][6] resolve this problem by using a feedback channel (FC), which allows the decoder to request additional bits from the encoder until meeting the error probability requirement. In this way, an optimal rate allocation can be achieved. However, such an FC solution introduces latency and increases decoder complexity because multiple groups of bitstreams are requested and the corresponding multiple decoding needs to be implemented. For broadcasting and some unidirectional streaming applications, the FC becomes impossible. In these scenarios, the FC-driven rate allocation solution becomes unacceptable.

A rate allocation algorithm for pixel-domain distributed video coders without feedback channel was presented in [7]. However, this algorithm does not have sufficient efficiency for video sequences with large motion since it assumes that when the allocated rate is an underestimation of the proper rate, the decoded frames with an excessive number of errors will be discarded and the frame is reconstructed from previously decoded frames available to decoder.

In contrast, the encoder side rate allocation algorithm will need to estimate the bit-rate for each Wyner-Ziv frame at the encoder and delivers them to the decoder. Thus, the latency in such a system is small and the decoder complexity is reduced since only one decoding operation is needed to correct the side information. However, for WZVC, since the encoder has no access to the side information.
information, two key questions need to be answered when designing rate allocation at encoder in WZVC scenario. The first question is how to optimally choose coding modes and the quantization parameters when the encoder is aware of the increase in encoder complexity this operation may introduce. The second question is how to allocate proper bit-rate at the encoder in order to achieve the same RD performance obtained with FC-driven decoder rate allocation. Recently, Roca et al proposed a rate control method to select the best mode and quantization for each video frame for pixel domain WZVC with an FC [8].

In this paper, we present an efficient rate allocation algorithm for transform domain Wyner-Ziv video coders without FC. By adopting a rate-distortion model, the proposed algorithm allows WZVC to select the coding mode and the quantization parameters of each Wyner-Ziv frame in order to provide accurate estimates of coding rate and compression efficiency. The estimated rates are quite close to the ones obtained by an FC-driven algorithm.

The rest of this paper is organized as follows. Section II provides an introduction to transform domain WZVC. Section III presents the formulation of the problem and the proposed rate allocation algorithm for WZVC without FC. Section IV shows the experimental results and finally Section V concludes this paper with discussion and future topics.

2. Transform domain wyner-ziv video coding

In this section, we review the fundamentals of the transform domain WZVC. As shown in Figure 1, the transform domain WZVC schemes that have been developed recently include a general architecture: intra encoding and inter decoding.

![Figure 1. Block diagram of the transform domain WZVC.](image)

The frames are organized into key frames (K-frames) and Wyner-Ziv frames (WZ-frames). The K-frames are encoded using the Wyner-Ziv paradigm. At the Wyner-Ziv encoder, the blockwise 8×8 Discrete Cosine Transform (DCT) is applied to each WZ-frame. The DCT coefficients of the entire frame are grouped together in DCT bands. Subsequently, each DCT band is quantized and bit-planes are extracted and sent to the LDPC encoder. To achieve compression effect, the systematic bits are then discarded and only parts of the syndrome bits (punctured from the original syndrome bits) generated by LDPC encoder for each bit-plane are sent to the decoder. These syndrome bits are transmitted in order to correct the errors in the side information. They are estimated for each bit-plane of each DCT band by the proposed encoder rate allocation module. At the decoder, the correlation between the video frames is used to construct an estimation of the current frame based on previously decoded frames and such estimation can be viewed as a noisy version of the original frame.

For a group of pictures (GOP) length of 2, \( \hat{S}_p \) and \( \hat{S}_s \) correspond to the previous and the next temporally adjacent key frames, \( S_p \) and \( S_s \), after being decoded but GOP may be longer. An 8×8 DCT and quantization is carried out over \( \hat{R} \). Then, the LDPC decoder combines the syndrome bits received and the estimation of the current frame. After all DCT bands are obtained by grouping the bit-planes together to form the symbol streams, an 8×8 inverse DCT and inverse quantization are performed. The reconstructed current frame is obtained at the end.

3. The encoder side rate allocation algorithm

In Wyner-Ziv video coders, the optimum rate \( R \) is the minimum rate to correctly decode the received syndrome bits and the reconstructed side information. A rate that is higher than \( R \) will not result in a reduction in distortion, but will cause unnecessary bit-rate expansion. On the other hand, a rate estimated at the encoder that is lower than \( R \) may cause large number of errors in the channel decoding of a block. This is because of the threshold effect of the channel codes we adopt in the Slepian-Wolf coder.

To achieve the optimal solution, most existing Wyner-Ziv video coders use a FC-driven rate allocation algorithm and rate-adaptive channel codes [5] at the decoder. In this scenario, the channel encoder generates all the syndrome bits for the blocks to be encoded, saves these bits in a buffer, and divides them into syndrome bit sets. To determine the adequate number of syndrome bits to send for each block, the encoder first transmits one part of syndrome bits from the buffer. If the decoder detects that the error probability is above the threshold \( Z \), it requests additional syndrome bits from the buffer through the FC. This transmission-request process is repeated until the error probability is less than the threshold \( Z \).

Although the FC-driven rate allocation solutions allow the system to obtain an optimal rate, the FC is not available in numerous applications where communication from the decoder to the encoder is not possible. In this case, it is desirable to propose an appropriate rate allocation algorithm at the encoder for WZVC. In order to optimally allocate the rate for each block of the WZ-frames, the proposed method estimates the minimum required number of syndrome bits at the encoder for a given error probability. An important aspect of the proposed approach is to avoid underestimation of the optimal number of parity bits. Indeed, if the rate is underestimated, the LDPC decoding of the blocks will not be error-free and this will lead to substantial increase in distortion. As shown in Figure 2, the proposed encoder rate allocation algorithm comprises three modules: (1) low-complexity estimation of reference frame module, (2) encoding rate estimation of WZ-frames module, and (3) coding mode and quantization parameter selection module. The following section describes these modules in more detail.

![Figure 2. Proposed encoder rate allocation framework](image)
3.1 Low-Complexity Estimation of Reference Frame
As in [8], we assume K-frames are the odd frames and WZ-frames are the even frames of a given video sequence. The first step in the proposed encoder rate allocation algorithm is to obtain an accurate estimation of the reference frame by exploiting the temporal correlation in video. The accuracy of the estimated reference frame will significantly influence the WZVC rate distortion performance. Unfortunately, complex frame interpolation schemes cannot be performed at the Wyner-Ziv encoder since the encoder complexity should be kept low.

We propose in this paper an efficient low-complexity estimation approach as described in the following. There are three types of frame interpolation that are adopted in the proposed framework. We divide the WZ-frame into 8×8 blocks. In the first and second schemes, the pixel value of each reference block is equal to the corresponding value in the preceding block and succeeding block as described in (1) (2). In the third scheme, there is a reference block that is estimated by the bilinear interpolation between the preceding block and succeeding block as described in (3):

\[
\hat{R}(x,y) = S_k(x,y) \quad (1) \\
\hat{R}(x,y) = S_k(x,y) \quad (2) \\
\hat{R}(x,y) = \frac{1}{2}[S_k(x,y) + S_{k+1}(x,y)] \quad (3)
\]

where \(\hat{R}(x,y)\) represents the pixel value of the reference block at the \((x,y)\) spatial location. Let us denote \(D_k\) as a random variable representing the difference between pixel values of the original block \(K\) and the reference block \(K\):

\[
D_k = \sum_{y=0}^{8} \sum_{x=0}^{8} |R_k(x,y) - S_k(x,y)| \quad (4)
\]

According to the value of \(D_k\), we can select the best interpolation method. The index of selected method will be transmitted to the decoder for each block. Considering the encoder complexity in this paper, we describe only the three simple approaches. Similar schemes have also been developed with more complex methods and can be adopted for this case.

3.2 Encoding Rate Estimation of WZ-frames
In order to allocate the adequate bit-rate for each WZ-frame, we need to predict the bit error probabilities \(P_e\) of each block \(K\) at the encoder. The error probabilities estimation technique is inspired by [7]. Let \(D\) be the difference between pixel values of the original frame \(S\) and the corresponding pixel values of its reference frame \(R\). As in [8], the difference \(D\) is assumed to follow a Laplacian distribution with a probability density function \(f(d) = \frac{\alpha}{\sigma} e^{-\alpha |d|} \), where \(\alpha = \frac{\sqrt{2}}{\sigma}\) and \(\sigma\) is the standard deviation of the difference \(D\), respectively. In practice, since pixel values can only be in the interval [0, 255]. Therefore, \(D\) is a discrete random variable that can only take integer values \(d\) in the interval [-255, 255]. Hence, we can derive the probability function for each block \(K\) as follows:

\[
P(D = d) = \int_{d-0.5}^{d+0.5} f_0(t) dt, \quad \text{except for } d = -255 \text{ and } d = 255 \text{ where the integration intervals of (5) are } (-\infty, 254.5) \text{ and } (254.5, +\infty), \text{ respectively.}
\]

Since every block of a WZ-frame \(S\) is extracted to the bit-plane, the virtual channel is modeled as a binary symmetric channel (BSC). To calculate \(P_e\), we first estimate \(\sigma^2\) at the encoder so the estimate should be very simple in order to avoid significant increase in encoder complexity. For each block \(K\), we set \(\sigma\) to be the square of the average of \(D_k\). In general, the resulting \(\sigma^2\) is an overestimate of the real \(\sigma^2\) since it is expected the motion compensation performed at the decoder will be more accurate than the simple reference information at the encoder. This overestimation will avoid the mass errors occurring in the LDPC decoder.

Let \(X_i\) and \(Y_i\) denote the transmitted and received bit in the \(K\)-th bit-plane, respectively. The total error probability for the corresponding bit-plane is:

\[
P_e = P(X_i = 1, Y_i = 0) + P(X_i = 0, Y_i = 1) \quad (6)
\]

Take into account the symmetry of the error distribution,

\[
P_e = 2P(X_i = 1, Y_i = 0) \quad (7)
\]

By using (5), (8) together with the variance estimate \(\sigma^2\), we obtain the \(P_e\). At the encoder, once \(P_e\) is estimated, we can choose the adequate syndrome bits generated by LDPC encoder that allows LDPC decoder correctly decoding the received syndrome bits.

3.3 Coding Mode and Quantization Parameter Selection
In this section, we propose a rate-distortion model to allow the transform domain WZVC to select the optimal coding mode and quantization parameter. When the correlation noise has a large variance in blocks of a video sequence, conventional intra-coding can be more efficient than Wyner-Ziv coding, because the error probability may go beyond the designed LDPC decoding capability, which will cause mass visible distortion. For these blocks, we allow WZ-frames to be encoded in intra mode. Furthermore, in portions of the video with very little motion, the distortion of the interpolated frames at the decoder can be very close to the distortion of their \(k\)-frames. Therefore, it would be unnecessary to send any syndrome bits in skip mode. Finally, when a WZ-frame is encoded using neither the intra mode nor the skip mode, parts of syndrome bits must be transmitted. We refer to this coding mode as WZ mode. As discussed in the last section, \(D_k\) represents the difference between pixel values of the original block and the reference block. When \(D_k\) is more than a threshold, we determine this block to be encoded in intra mode. Then, the rate distortion model, as described in the following in (9) and (10) [8] will select the mode (skip, WZ) of each WZ-frame and the quantization step size of all the WZ-frames.

\[
im = \text{arg min}_{i=0} \left( S_i, \frac{I_m}{QP} \right) \quad (9)
\]
where $I'_t$ is the best coding mode, $D_{\text{wz}}(S, I_t | QP)$ is the distortion when the quantization size $QP$ has been used in the encoding of block $K$. For a given sequence and WZ-frames rate estimation scheme, $D_{\text{wz}}(S, I_t | QP)$ depends on only the quality parameter $QP$. Because once $QP$ is set, the distortion can be determined as well. Suppose $R'$ is the target bit-rate obtained by encoding with $QP$. Thus, the quantization parameter $QP$ could be determined by minimizing $D_{\text{wz}}(S, I_t | QP)$ subjected to the bandwidth limitation $[R_{\text{wz}}(S, I_t | Q) \leq R']$. Therefore, we can select an adequate quantization parameter. $R_{\text{wz}}(S, I_t | Q)$ is the allocated bits in coding mode $I_t$. $R_{\text{wz}}(S, I_t | Q)$ is estimated by using the results obtained in last section. After the steps are repeated for all blocks, the optimum coding mode and quantization parameter are selected. The chosen coding mode and quantization parameter determines the number of bits allocated for each block.

### 4. Experimental results

In this section, we investigate through experiments the accuracy of the proposed encoder rate allocation algorithm as it applies to transform domain Wyner-Ziv video coder (TDWZ) without FC. We then compare the proposed algorithm with the rate allocation algorithms provided by the same TDWZ coder with FC (FCWZ).

The TDWZ video coder used in the experiments is described in detail in Section II, integrated with the encoder rate allocation algorithm. The K-frames are intra-coded using H.264/AVC test model coder JM8.6 with initial quantization parameter $QP = 28$.

The number of original bits participating in the LDPC code is fixed to $N = 6336$. We use the same LDPC codes as specified in [5]. We test the video sequences ‘Carphone’, ‘Football’ with the QCIF (176×144) resolution. The threshold $Z$ for TDWZ and FCWZ is set to $1/N$, where $N$ is the number of pixels in each frame.

Table I show the percentages of frames with different rates between the TDWZ and FCWZ. Note that in both sequences with little motion and high motion, we allocate an appropriate rate compared to the FC-driven rate allocation solution, since the intra mode is used to encode the high motion regions. In Figure 3, we show the rate-distortion curves of ‘Carphone’, ‘Football’ for the TDWZ coder, and we compare them with the corresponding rate-distortion curves when an optimal rate is allocated by FC. We observe that the loss in rate-distortion performance of the TDWZ coder compared to the FCWZ coder is very small.

### 5. Conclusion

In this paper, we have presented a rate allocation algorithm for transform domain Wyner-Ziv video coding, which enables us to avoid the use of feedback channel in the Wyner-Ziv video coder. The proposed algorithm estimates the appropriate number of bits for each WZ-frame at the encoder without significantly increasing the encoder complexity. Comparing the results of the proposed rate allocation algorithm and the solutions based on the feedback channel, we observe that the loss in rate-distortion performance due to the absence of an FC is acceptable. Hence, the proposed algorithm can be very useful in the application scenarios when an FC is not allowed. In the future, we plan to integrate the encoder rate allocation solution with future full HDTV broadcasting video encoding based on distributed source coding.

### 6. REFERENCES


