2009 11th IEEE International Symposium on Multimedia

Frequency Band Coding Mode Selection for Key Frames of Wyner-Ziv Video Coding

Ghazaleh R. Esmaili and Pamela C. Cosman
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA, 92093-0407
gesmaili@ucsd.edu, pcosman@ucsd.edu

Abstract

In most Wyner-Ziv video coding approaches, the temporal correlation of key frames is not exploited, since they are simply intra encoded and decoded. In previous work, by using the previously decoded key frames as the side information, we proposed to divide the frequency bands of each block into two classes. Wyner-Ziv coding was used for the low frequency bands of each block, while high frequency bands were intra coded. In this paper, we improve this approach with an efficient coding mode selection technique. Frequency bands are grouped as low and high bands and an appropriate method of coding is selected for them based on the correlation characteristics of each frame with the past. This method does not add complexity to the encoder. Simulation results show that using the proposed method, one can achieve up to 4 dB improvement over prior work.

1. Introduction

The Slepian-Wolf [9] and Wyner-Ziv [10] theorems prove that distributed compression of correlated sources can be as efficient as joint compression. Wyner-Ziv video coding founded on these two theorems is a promising solution to some recent applications such as sensor networks, video surveillance, and mobile camera phones which require many simple and low cost encoders.In this scheme, interframe dependency is exploited at the decoder, unlike traditional video coding standards such as MPEG-x and H.26x where similarities among adjacent frames are exploited at the encoder by using motion compensation techniques. Therefore, the Wyner-Ziv encoder has much lower complexity.

The first implementations of Distributed Video Coding were presented in [8] and [2]. In [8], Puri and Ramchandran introduced a syndrome-based video coding scheme with flexible sharing of computational complexity between the encoder and the decoder and no feedback was required. It was upgraded to a more practical solution in [7]. In [2], Aaron and Girod proposed a feedback-required Wyner-Ziv video coding algorithm based on Turbo codes in the pixel domain. This technique was extended to the transform domain in [1] to exploit spatial correlation between neighboring pixels, thus achieving better performance. In [4], Brites and Ascenso outperformed [1] by adjusting the quantization step size and applying an advanced frame interpolation for side information generation. Later on, in [3] and [11], enhanced frame interpolation techniques were proposed to achieve better performance. In [12], Jinrong et al. proposed a transform domain classification method to differentiate low motion blocks from high motion blocks to exploit additional video statistics. In [6], an iterative algorithm is proposed to find which blocks should be considered for intra coding and which for Wyner-Ziv coding. In this method, complexity of the encoder is increased compared to simply intra coding by adding a buffer, generating some form of side information at the encoder, processing the iterative algorithm and compressing mode information bits.

In most existing Wyner-Ziv schemes, frames are grouped into two different classes: Wyner-Ziv and key frames. In a GOP size of 2, key frames occur every other frame, and Wyner-Ziv frames are the frames in between. Key frames are intra coded using conventional video coding. At the decoder, they are decoded and used to generate side information that is statistically correlated with the Wyner-Ziv frame in between. If the correlation between the Wyner-Ziv frame being encoded and the side information is high, then fewer bits need to be sent from the encoder to the decoder to have a reliable decoding. The statistical dependence between two consecutive key frames is not as high as the statistical dependence between a Wyner-Ziv frame and its corresponding generated side information, because in the latter case the side information comes from frames that are only one

978-0-7695-3890-7/09 $26.00 © 2009 IEEE
DOI 10.1109/ISM.2009.88

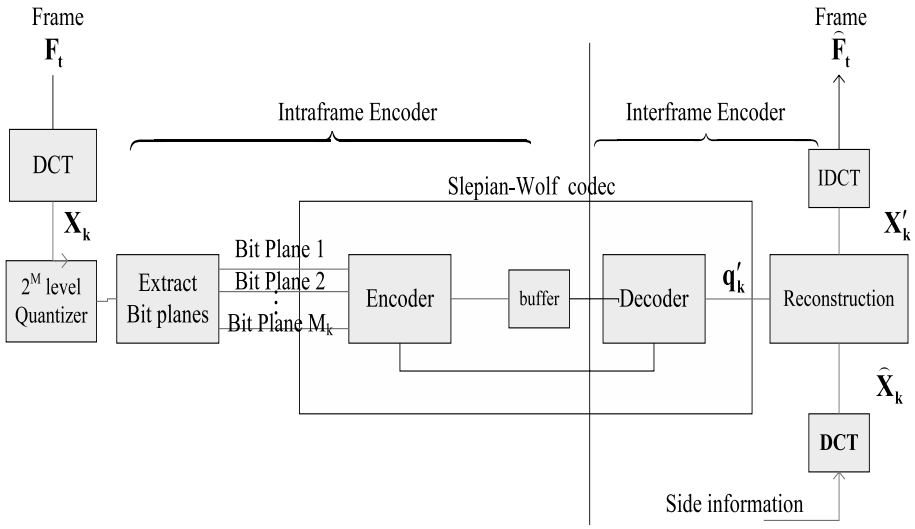

Figure 1. Transform domain Wyner-Ziv video codec

frame-time distant, and comes from frames on both sides temporally. Nonetheless, extending the Wyner-Ziv coding method to key frames as well can help to exploit temporal correlation and improve rate-distortion performance.

In previous work [5], we proposed two different coding methods for key frames to exploit their inter-frame correlation: block classification and frequency band classification. In the block classification method, based on the mean square error between a given block in a key frame and the co-located block in the reconstructed previous key frame, blocks were divided into three classes: Wyner-Ziv, Intra and Skip. This method works very well for relatively low motion sequences where many blocks are classified as Skip or Wyner-Ziv. In the frequency band classification method, Wyner-Ziv coding was used for low frequency bands which are usually highly correlated, and Intra coding was applied for high frequency bands. Adding no complexity to the encoder and using existing modules of the system made frequency band classification more attractive. But the inter-frame correlation of high frequency bands, which also tends to be high when side information is a good estimation of the source, was not exploited.

In this paper, without adding any complexity to the encoder, we propose a coding mode selection technique which tries to select the proper coding method based on the correlation characteristics of the low and high frequency bands of each frame to the past. In this method, the decoder decides the coding mode.

The rest of this paper is organized as follows. In Section 2, different coding modes and the proposed codec with mode decision are described in detail. In Section 3, the performance of the proposed method is evaluated, and conclusions are made in Section 4.

2. The proposed method

Since Wyner-Ziv coding and Intra coding modules are already part of existing Wyner-Ziv codecs, switching between Wyner-Ziv and intra coding to exploit correlation between consecutive key frames does not add any complexity to the encoder as long as the switching decision is done at the decoder. In the frequency band classification method in [5], frequency bands were divided into low and high frequency classes, and Wyner-Ziv coding and Intra coding were assigned to these classes, respectively. If the side information is not a sufficiently accurate estimation of the source, Wyner-Ziv coding can do worse than intra. Since the temporal correlation of low frequency bands is usually high, Wyner-Ziv coding usually outperforms Intra coding.

For high frequency bands, on the other hand, it is often the case that Intra outperforms Wyner-Ziv coding because the temporal correlation is fairly low. So, we need to estimate the accuracy of the side information in order to decide the right coding method for high frequency bands. Since we want to avoid increasing the complexity of the encoder, we leave this decision to the decoder. The decoder compares the received low frequency components of the current key frame with their corresponding side information to decide if Wyner-Ziv coding is the right choice for encoding the rest of the frequency bands. The main idea is measuring the distortion between source and side information of the low frequency bands at the decoder in order to estimate the accuracy of the side information for high frequency bands. Once the decision is made, the decoder lets the encoder know by sending a single bit through the feedback channel which is already a part of the system. In this section, after describing different coding modes, we represent our mode selection

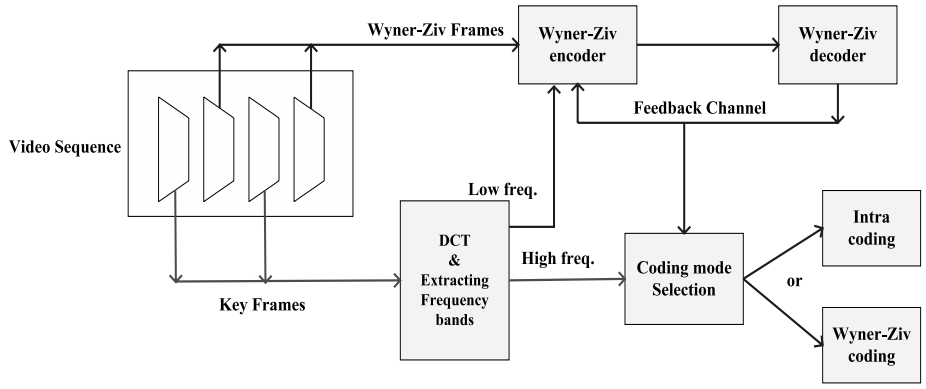

Figure 2. Proposed video codec with frequency band coding mode selection for key frames

scheme.

2.1. Wyner-Ziv coding

Fig. 1 shows the transform domain Wyner-Ziv video codec architecture. A vector X_k is formed by grouping together the k^{th} DCT coefficient from all blocks. The coefficients of X_k are quantized to form quantized symbols q_k. After representing the quantized values q_k in binary form, bit planes are extracted and blocked together to form M_k bit plane vectors. Each bit-plane vector is then fed to the Slepian-Wolf encoder. $\hat{X}_k$ is generated by grouping the k^{th} transform coefficients of the side information. When Wyner-Ziv coding is used for key frames, the side information is the previous decoded key frame. The decoder and reconstruction blocks assume a Laplacian distribution to model the statistical dependency between X_k and $\hat{X}_k$. For each coefficient band, the Slepian-Wolf decoder successively decodes bit-planes beginning from the most significant bit-plane. The received bits of the bit-plane and the side information $\hat{X}_k$ are the decoder tools to decode the current bit-plane. If the decoder can not meet the desired bit error rate, it asks for additional bits through feedback. At the end, the reconstructed coefficient band X'_k is calculated as $E(X_k|q_k, \hat{X}_k)$.

2.2. Intra coding

For the Intra mode, the quantized DCT coefficients are arranged in a zigzag order to maximize the length of zero runs. The codeword represents the run-length of zeros before a non-zero coefficient and the size of that coefficient. A 2-dimensional Huffman code (Run, Size) is used because there is a strong correlation between the size of a coefficient and the expected run of zeros which precedes it. Small coefficients usually follow long runs and larger coefficients tend to follow shorter runs. In our simulation Huffman and run length coding tables are borrowed from the JPEG standard.

2.3. Coding mode selection

Fig. 2 shows our proposed codec applying coding mode selection for key frames. Wyner-Ziv coding is applied to low frequency bands. To separate different frequency bands of the key frame to be encoded, first the DCT is applied. To form the k^{th} frequency band, the k^{th} DCT coefficients from all blocks are grouped to form vector X_k. Low frequency bands are encoded and decoded by Wyner-Ziv coding. The previously decoded key frame is used as the side information. To provide corresponding side information for each frequency band, the DCT is applied on the previously reconstructed key frame and the k^{th} DCT coefficient from all blocks are grouped to form vector $\hat{X}_k$. Once the decoder receives and decodes all low bands, the distortion between these reconstructed bands and their corresponding side information is calculated as $E = \frac{1}{K \times L} \sum_{k \epsilon Low\, freq.} \sum_{l=1}^{L} (X'_k(l) - \hat{X}_k(l))^2$ where X'_k denotes the reconstructed X_k at the decoder. L is the number of elements in each frequency band which can be found as $\frac{M \times N}{bsize}$ where M and N represent the number of pixels in each dimension of a frame and $bsize$ denotes the size of the DCT. K is the total number of frequency bands. If E is less than a threshold T, the side information is fairly accurate, so that Wyner-Ziv coding is chosen for high frequency bands. Otherwise, Intra coding is applied for them. The decoder sends a single bit per frame through the feedback channel to indicate the selection. The added effect of sending a single bit per frame through the feedback channel on the latency of the system is negligible, since in traditional Wyner-Ziv coding, feedback bits might be sent for each bit plane to request more accumulated syndrome to meet the desired bit error rate.

3. Simulation results

Fig. 3 (a), (b), (c) and (d) show the results for the first 99 frames of the $Foreman$, $Mother\text{-}daughter$, $Carphone$

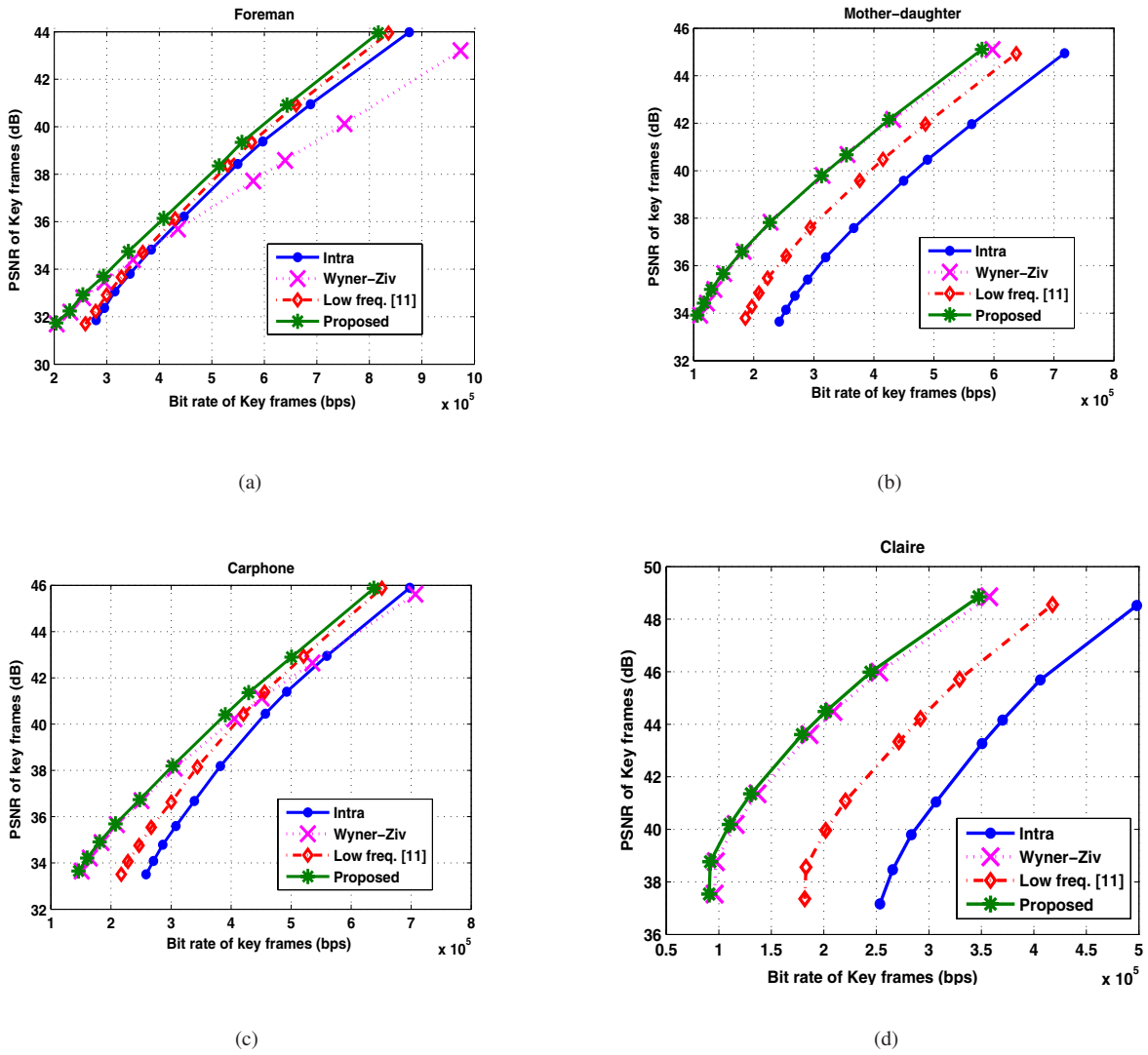

Figure 3. PSNR vs. Rate of four different coding methods for key frames

and $Claire$ QCIF sequences at 30 frames per second. For all plots, only the rate distortion performance of the luminance of key frames is included. Odd frames are considered as key frames. The bit rate of key frames is calculated as $\frac{B\times R}{F}$ where B is the total number of sent bits for all key frames, F is the total number of frames and R is the frame rate. So, the bit rate of key frames in our case is calculated as $\frac{B\times 30}{99}$. With a 4×4 DCT, $f(1,1)$, $f(1,2)$ and $f(2,1)$ are considered low frequency bands and the rest are considered high frequency bands.

In our previous work [5], we used different quantization methods for Wyner-Ziv and Intra modes but we tried to provide the same level of quality. Here, to have the same quality for both Wyner-Ziv and Intra modes $Q(x)$ is used to quantize DCT coefficients where $Q(a_{i,j}) = round\,(\frac{a_{i,j}}{QP\times c_{i,j}})$, $a_{i,j}$ is the unquantized coefficient at position (i,j), $c_{i,j}$ is the element of the quantization matrix at position (i,j) and QP is the quantization parameter. The quantization matrix applied in our simulation is the initializing quantization matrix borrowed from H.264 JM 9.6:

$$C = \begin{bmatrix} 6 & 12 & 19 & 26 \\ 12 & 19 & 26 & 31 \\ 19 & 26 & 31 & 35 \\ 26 & 31 & 35 & 39 \end{bmatrix}$$

In our simulation $QP\epsilon\{0.4, 0.65, 0.85, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$. For each of these quantization parameters, a threshold value is set. We tried different values between 50 and 1800 with step sizes 20 to 100 for several video sequences at different quantization parameters. The value of the step

size depends on the quantization parameter, with bigger step sizes for bigger quantization parameters. Threshold values $T = [200, 270, 290, 400, 500, 740, 860, 1200, 1500, 1800]$ corresponding to quantization parameters $QP = [0.4, 0.65, 0.85, 1, 1.5, 2, 2.5, 3, 3.5, 4]$, were chosen as they work well for different sequences with different characteristics.

The result is compared with applying just Intra mode for all frequency bands ("Intra"), which is used in most existing methods. We also compare against the proposed frequency band classification method in [5] called "Low freq." and applying just Wyner-Ziv coding for all frequency bands called "Wyner-Ziv" As shown, the proposed method results in up to 9 dB improvement over "Intra" and up to 4 dB improvement over "Low freq". The gain for $Claire$ and $Mother\text{-}daughter$ is greater since they are relatively low motion sequences in comparison with $Foreman$ and $Carphone$. That means high frequencies of more frames are chosen for Wyner-Ziv coding. Also, the average distortion E of such frames in $Mother\text{-}daughter$ and $Claire$ is lower than for frames chosen for Wyner-Ziv coding in $Foreman$ and $Carphone$. Less distortion means more accurate side information and therefore saving more bits by applying Wyner-Ziv over Intra.

4. Conclusion

In this paper, we presented a coding method for key frames which attempts to exploit their temporal correlation in order to improve the overall coding performance. The proposed coding method for key frames always assigns Wyner-Ziv coding method to low frequency bands and chooses between Intra and Wyner-Ziv coding for high frequency bands of each frame. The coding mode selection has been designed such that no complexity is added to the encoder. Applying the proposed method results in up to 9 dB improvement over the Intra method and 4 dB improvement over the frequency band classification method in [5].

References

[1] A. Aaron, S. Rane, and B. Girod. Transform-domain Wyner-Ziv codec for video. *VCIP*, San Jose, January 2004.

[2] A. Aaron, R. Zhang, and B. Girod. Wyner-Ziv coding of motion video. *Proc. Asilomar Conference on Signals and Systems*, Nov. 2002.

[3] S. Argyropoulos, N. Thomosy, N. Boulgourisz, and M. Strintzis. Adaptive frame interpolation for Wyner-Ziv video coding. *IEEE 9th workshop on Multimedia signal processing*, pages 159–162, Oct. 2007.

[4] C. Brites, J. Ascenso, and F. Pereira. Improving transform domain Wyner-Ziv video coding performance. *IEEE ICASSP*, 2, May 2006.

[5] G. Esmaili and P. Cosman. Low complexity spatio-temporal key frame encoding for Wyner-Ziv video coding. *Proc. Data Compression Conference*, 2009.

[6] L. Liu, D. He, A. Jagmohan, L. Lu, and E. Delp. A low complexity iterative mode selection algorithm for Wyner-Ziv video compression. *Proc. IEEE ICIP*, pages 1136–1139, Oct. 2008.

[7] R. Puri, A. Majumdar, and K. Ramchandran. Prism: A video coding paradigm with motion estimation at the decoder. *IEEE Trans. on Image Processing*, 16:2436–2447, Oct. 2007.

[8] R. Puri and K. Ramchandran. Prism: A new robust video coding architecture based on distributed compression principles. *Proc. Allerton Conference on Communication, Control, and Computing*, Oct. 2002.

[9] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. on Information Theory*, IT-19, no. 4:471–480, July 1973.

[10] A. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. on Information Theory*, IT-22, no. 1:1–10, Jan. 1973.

[11] S. Ye, M. Ouaret, F. Dufaux, and T. Ebrahimi. Improved side information generation with iterative decoding and frame interpolation for distributed video coding. *Proc. ICIP*, 2008.

[12] J. Zhang, H. Li, Q. Liu, and C. W. Chen. A transform domain classification based Wyner-Ziv video codec. *IEEE International Conference on Multimedia and Expo*, pages 144–147, July 2007.