WYNER-ZIV VIDEO CODING WITH CODING MODE-AIDED MOTION COMPENSATION

Li-Wei Kang and Chun-Shien Lu*
Institute of Information Science, Academia Sinica, Taipei, Taiwan, ROC
Email: {lwkang, lcs}@iis.sinica.edu.tw

ABSTRACT
In distributed video coding, individual frames are encoded independently but decoded conditionally. The Wyner-Ziv theorem-based source coding with side information only available at the decoder states that an intraframe encoder with interframe decoder system can approach the efficiency of a conventional interframe encoder and decoder system. In this paper, a new block discrete cosine transform (DCT)-based Wyner-Ziv video codec with coding mode-aided motion compensation at the decoder is proposed. The key is that for each block, a large amount of candidate blocks are evaluated based on some criteria derived from Reed-Solomon (RS) decoding and best neighborhood matching to find the best candidate block as the side information. Another characteristic is that error correction code (ECC) decoding is proposed to participate in generating side information. Compared with some known Wyner-Ziv video coding systems, in the proposed video codec, no extra information should be transmitted and feedback channel is unnecessary. The coding performance of our method has been verified through simulations.

Index Terms—Video codecs, video coding, DCT, motion compensation, Reed-Solomon codes, run length codes, entropy codes

1. INTRODUCTION
Conventional video compression standards, such as MPEG-4, H.263, and H.264/AVC [1], usually perform interframe predictive coding exploiting the similarities among successive frames. Since predictive coding usually employs motion estimation, the encoder is typically 5 to 10 times more complex than the decoder [2]. However, with the advancement of emerging applications of wireless video sensor networks, current video coding paradigm is insufficient if some new requirements, such as restrictions on computational capability and memory of resource-limited video sensors, are considered. In fact, this calls for a low complexity video encoder.

Recently, distributed video coding becomes a new video coding paradigm, where individual frames are encoded independently (intraframe coding) but decoded conditionally (interframe decoding) [2]-[6]. Contrast to conventional video coding, distributed video coding shifts partial computational burden (e.g., motion estimation) from the encoder to the decoder, resulting in a new codec with light encoder and heavy decoder. Based on the Slepian-Wolf theorem [7] and Wyner-Ziv theorem [8], an intraframe encoder with interframe decoder system can approach the coding efficiency of an interframe encoder with interframe decoder system. Assume that $X$ and $Y$ are statistically dependent random processes, and the Wyner-Ziv theorem states distributed source coding with side information only available at the decoder states that an intraframe encoder with interframe decoder system can approach the efficiency of a conventional interframe encoder and decoder system. Assume that $X$ and $Y$ are two statistically dependent discrete signals, which are encoded independently but decoded jointly. Based on the Slepian-Wolf theorem [7], the achievable rate region for lossless coding is defined to be surrounded by $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$, and $R_X + R_Y \geq H(X,Y)$, where $R_X$ and $R_Y$ are the rates of encoding $X$ and $Y$, respectively, $H(X|Y)$ and $H(Y|X)$ are the conditional entropies of $X$ and $Y$, respectively, and $H(X,Y)$ is the joint entropy of $X$ and $Y$. On the other hand, the Wyner-Ziv theorem states distributed source coding with side information for lossy coding. Assume that $X$ and $Y$ are statistically dependent random processes, and $Y$ is known as the side information for encoding $X$. The conditional distortion function for $X$ will be unchanged no matter $Y$ is available only at the decoder, or both at the encoder and decoder.

In the literature, a few distributed video coding methods, usually called Wyner-Ziv video codec, have been proposed [2]-[6]. A general flowchart of Wyner-Ziv video codec is shown in Fig. 1. At the encoder, an input video sequence is divided into key frames and Wyner-Ziv frames. Each key frame is encoded using a conventional intraframe video codec, while each Wyner-Ziv frame is encoded using a distributed encoder to generate Wyner-Ziv bits. For example, in [4]-[5], a distributed encoder consists of the DCT, a scalar quantization followed by a turbo encoder to generate parity bits, which are then partly transmitted as the Wyner-Ziv bits. On the other hand, the encoder can transmit some extra information to the decoder optionally. For example, in [4]-[5], the encoder transmits the hash bits consisting of a small subset of the quantized DCT coefficients of some selected blocks in Wyner-Ziv frames. At the decoder, each key frame is decoded using the conventional intraframe decoder. For a Wyner-Ziv frame, it is decoded using the distributed decoder with the assistance of side information. Side information can be generated using any previous decoded frames and/or the extra information transmitted from the encoder. For example, in [3], the side information for a Wyner-Ziv frame is generated based on the interpolation of some previous decoded frames. In [4]-[5], for each block in a Wyner-Ziv frame, the side information is generated based on the hash bits (transmitted
from the encoder) to find the best matched block from the previous decoded frame by means of motion-compensation extrapolation. The generated side information, \( \hat{W} \), of a Wyner-Ziv frame, \( W \), can be viewed as the estimate of \( W \) and can cooperate with the distributed decoder to decode the Wyner-Ziv bits. In [3]-[5], if the received Wyner-Ziv bits (a subset of parity bits) are not enough for decoding, the decoder can request additional parity bits from the encoder via a feedback channel.

![Diagram](image_url)

**Fig. 1. Flowchart of a general Wyner-Ziv video codec.**

In this paper, a new block DCT-based Wyner-Ziv video codec with coding mode-aided motion compensation at the decoder is proposed. The major characteristics include (a) For each block, a large amount of candidate blocks are evaluated based on some criteria derived from Reed-Solomon (RS) decoding and best neighborhood matching to find the best candidate block as the side information; (b) Error correction code (ECC) decoding is applied to participate in generating side information; (c) No feedback channel is required in the proposed codec. Conversely, in previous works [3]-[5], the side information is generated without considering ECC decoding (turbo code is used). Hence, the generated side information may not be the best one for turbo decoding and additional parity bits may be requested from the encoder via a feedback channel.

2. PROPOSED WYNER-ZIV VIDEO CODEC

The proposed Wyner-Ziv codec consisting of an intraframe encoder and an interframe decoder is shown in Fig. 2.

![Diagram](image_url)

**Fig. 2. The proposed Wyner-Ziv video codec.**

2.1. Proposed Wyner-Ziv Video Encoder

At the encoder, an input video sequence is divided into several GOPs (group of pictures), in which a GOP consists of a key frame followed by several Wyner-Ziv frames. Each key frame \( (K) \) is encoded using the H.264/AVC intra encoder [1]. Each Wyner-Ziv frame is divided into several nonoverlapping \( N \times N \) blocks. First, the coding mode for each block in a Wyner-Ziv frame will be decided based on the estimated motion activity. Here, the original previous frame will be kept in a buffer. For each block, \( b_i \), in the current frame, the difference, \( d_i \), between the block and the co-located block in the previous frame is calculated. If \( d_i \leq T_1 \), the coding mode of \( b_i \) is declared to be skip mode. If \( T_2 < d_i \leq T_3 \), the coding mode of \( b_i \) is declared to be non-skip with RS coding mode. Otherwise, the coding mode of \( b_i \) is declared to be non-skip without RS coding mode. Here, \( T_1 \) and \( T_2 \) are two predefined positive thresholds and \( T_1 < T_2 \). The coding mode information for all the blocks in a Wyner-Ziv frame will be either encoded using the run-length coding followed by the entropy coding or hidden into the key frames (the first frame) of the next GOP. The extra overhead in the encoder is a buffer with the size the same as that of an uncompressed frame plus that of the coding mode information in a Wyner-Ziv frame.

In this study, the coding mode information for each Wyner-Ziv frame will be encoded using run-length coding followed by entropy coding. To further decrease the overhead induced by the coding mode information, the encoded coding mode information for the last Wyner-Ziv frame in a GOP will be hidden into the key frame of the next GOP. Here, a simple odd-even data hiding scheme is employed, which will not change the bitrates and the quality of a key frame significantly. In the context-based adaptive variable length coding (CAVLC) for a zig-zag ordered \( 4 \times 4 \) DCT block in H.264/AVC, some important encoded components are described as follows [1]. First, the component, Coeff_token, encodes the number of non-zero coefficients (e.g., TotalCoeff) and trailing ones (\( \pm 1 \)) (e.g., TrailingOnes). In H.264/AVC, \( 0 \leq \text{TrailingOnes} \leq 3 \) and more ones (except for the first three ones in the reverse zig-zag order) are encoded as “\( \text{Level} \)”. Second, Level (sign and magnitude) denotes the remaining non-zero coefficients except TrailingOnes. The third component, Total.zeros, encodes the total number of zeros after the first non-zero coefficient in the reverse zig-zag order. To maintain the rate-distortion (RD) performance for a block while performing data hiding, the key is that Coeff.token and Total.zeros should not be changed. Here, if the data bit to be hidden is \( a_{ij} \), the first Level \( L_j \) in the reverse zig-zag order in a \( 4 \times 4 \) DCT block is determined as

\[
L_j = \begin{cases} 
1 & \text{if } \{b_i, \text{mod} 2 = a_i\} \text{ and } \{\text{TrailingOne} < 1, \text{mod} 2 = a_i\} \\
2 & \text{if } \{b_i, \text{mod} 2 = a_i\} \text{ and } \{\text{TrailingOne} < 1, \text{mod} 2 = 1\} \\
3 & \text{if } \{b_i, \text{mod} 2 = a_i\} \text{ and } \{\text{TrailingOne} = 1, \text{mod} 2 = 1\} \\
4 & \text{otherwise}
\end{cases}
\]

That is, a coefficient (Level) for hiding one bit is either unchanged or increased/decreased by one and a \( 4 \times 4 \) block hides at most one bit. The data hiding process is performed.
while encoding a key frame. Finally, the output of the proposed H.264/AVC intra encoder with data hiding forms the key frame bits, as shown in the bottom of Fig. 2 (except that the first key frame does not perform data hiding).

On the other hand, for a Wyner-Ziv frame, each block will be encoded based on its coding mode. For a block with skip mode, no data will be encoded. For a block with non-skip mode, a block DCT will be performed followed by a scalar quantization to obtain a symbol block containing $N \times N$ DCT symbols. Four possible employed quantizers are shown in Fig. 3. For example, if the quantizer shown in Fig. 3(a) is used, the DC value will be quantized to a symbol (denoted by 6 bits) with at most 64 levels.

![Fig. 3. Four possible scalar quantizers.](image)

For a block with non-skip with RS coding mode, the most three important DCT symbols will be encoded directly. The remaining DCT symbols will be encoded using $(u, v)$ RS codes [9] to generate parity symbols, where $v$ is the number of DCT subbands having the same number of quantization level. Only RS parity symbols will be encoded. For example, if a 4×4 DCT block is quantized using the quantizer shown in Fig. 3(b), the most three important DCT symbols will be encoded directly with 6 bits, 4 bits, and 4 bits, respectively. The remaining 10 DCT symbols will be encoded using (12, 10) RS code to generate 2 parity symbols (2×3 bits). In this case, a block is totally encoded with 20 bits. On the other hand, for a block with non-skip without RS coding mode, all the DCT symbols will be encoded directly. Usually, this kind of blocks is few and coarser quantizers (e.g., Fig. 3(d)) are adopted.

The resulted encoded symbols (DCT symbols and RS parity symbols) for all the blocks with non-skip mode in a Wyner-Ziv frame result in the so-called Wyner-Ziv bits. Both the key frame bits and the Wyner-Ziv bits will be transmitted to the decoder. The computational complexity of the proposed encoder for a Wyner-Ziv frame is dominated by those of the DCT and RS encoding, and is similar to that of a conventional intraframe encoder consisting of the DCT and entropy coding.

2.2. Proposed Wyner-Ziv Video Decoder

At the decoder, a key frame excluding the first one will be decoded using the H.264/AVC intra decoder with hidden data extraction. Then, the hidden coding mode information for the last Wyner-Ziv frame in a GOP will be decoded.

For a Wyner-Ziv frame, the coding mode information will be decoded first and then all the blocks with skip mode will be reconstructed by assigning the co-located blocks of the previous reconstructed frame. On the other hand, for each block with non-skip mode, the proposed coding mode-aided motion compensation scheme is employed to find the corresponding side information. In addition, for a block with non-skip mode, some search windows within the previous reconstructed frames are formed so that each block in the search windows will be a candidate block. The DCT followed by the scalar quantization performed at the encoder will be applied to each candidate block. The reconstructed 8-connected neighboring blocks for each candidate block will be also extracted.

In our method, for a block, $b$, with non-skip with RS coding mode, each candidate block, $c$, in the search windows will be evaluated. The best candidate block satisfying the three criteria described below will be selected to be the side information for $b$:

(a) The difference between the most important three DCT symbols of $b$ and those of $c$ should be minimized;
(b) The number of incorrect RS-decoded symbols based on the parity symbols of $b$ and the DCT symbols (except the most important three symbols) of $c$ should be minimized;
(c) The difference between the 8-connected neighboring blocks of $b$ and those of $c$ should be minimized.

Here, the RS-decoding symbol error rate usually ranges from 1% to 10%. However, the most important three DCT symbols have been kept. Some error symbols in the other subbands are usually acceptable. An illustrated example is shown in Fig. 4, where the light-colored blocks are already reconstructed blocks. For the blocks in a Wyner-Ziv frame (excluding those with skip mode), they are reconstructed in a raster-scan order.

![Fig. 4. An example of motion compensation at decoder for a block, $b$, with non-skip with RS coding mode.](image)
$Y_r$ in the side information is within the coefficient interval denoted by $q_r$, the symbol, $q_r$, will be dequantized to $Y_r$; otherwise, the boundary of the quantization interval that is nearest to $Y_r$ is used to reconstruct $q_r$. That is, the reconstructed DCT coefficient, $X'_r$, is represented as $X'_r = E(X_r | q_r, Y_r)$. For a block with non-skip without RS coding mode, the same strategy is employed to reconstruct all the DCT coefficients. Finally, the inverse DCT is then applied to a reconstructed DCT block to obtain the pixel block. Here, the source data $X_r$ and its side information $Y_r$ have been shown to have a Laplacian residual distribution.

3. SIMULATION RESULTS

Several QCIF video sequences formatted with GOP size, 8, frame rate, 10fps, and encoded with several different bitrates were used to evaluate the proposed video codec. In this study, $4 \times 4$ DCT ($N = 4$) was used. Here, the bitrates were adjusted by changing the quantization parameter for key frames, by changing the quantizers for Wyner-Ziv frames, and by changing the number of blocks for each coding mode. $T_r$ and $T_q$ were empirically determined. The H.264/AVC intraframe coding and H.264/AVC interframe coding [1] were employed for comparisons with the proposed video codec. The RD performances for the Salesman and Hall Monitor sequences are shown in Fig. 5.

4. CONCLUSIONS

In this paper, a new Wyner-Ziv video codec with coding mode-aided motion compensation at decoder is proposed. Its characteristics include: (a) since each block is encoded based on its coding mode decided by its estimated motion activity, no extra information should be transmitted; (b) since a large amount of candidate blocks are evaluated based on some criteria to find the best candidate block as the side information for each block with non-skip mode at decoder, no feedback channel is required. The proposed video codec has shown to present significant gains over the H.264/AVC intraframe coding while having comparable encoding complexity. Unavoidably, there is still a performance gap from the H.264/AVC intraframe coding. For future researches, more precise side information generation scheme will be investigated to achieve better RD performance. The key to this goal is accurate generation of side information. We’ll plan to employ robust perceptual hashing (e.g., [10]) to design and reconstruct Wyner-Ziv frames at the decoder. On the other hand, the rate control, error resilience, and security issues deserve further studying.

REFERENCES