

# Witsenhausen–Wyner Video Coding

Mei Guo, Zixiang Xiong, *Fellow, IEEE*, Feng Wu, *Senior Member, IEEE*, Debin Zhao,  
Xiangyang Ji, and Wen Gao, *Fellow, IEEE*

**Abstract**—Inspired by Witsenhausen and Wyner’s 1980 (now expired) patent on “interframe coder for video signals,” this paper presents a Witsenhausen–Wyner video codec, where the motion-compensated previously decoded video frame is used at the decoder as side information for joint decoding. Specifically, we replace predictive Inter coding in H.264/AVC by the syndrome-based coding scheme of Witsenhausen and Wyner, while keeping the Intra and Skip modes of H.264/AVC unchanged. We employ forward motion estimation at the encoder and send the motion vectors to help generate side information at the decoder, since our focus is not on low-complexity encoding. We also examine the tradeoff between the motion vector resolution and coding efficiency. Within the Witsenhausen–Wyner coding mode, we optimize the decision between syndrome coding and entropy coding among different discrete cosine transform (DCT) bands and among different bit-planes within each DCT coefficient. Extensive simulations of video transmission over wireless networks show that Witsenhausen–Wyner video coding is more robust against channel errors than H.264/AVC. The price paid for enhanced error-resilience with Witsenhausen–Wyner coding is a small loss in compression efficiency.

**Index Terms**—Distributed video coding, error resilience, syndrome coding.

## I. INTRODUCTION

FTER TWO decades of research on video compression, we now have a series of international standards such as MPEG-2 [1] and H.264/AVC [2] that are widely used in applications [3] such as satellite TV, DVDs, and video telephony. As portable devices such as camera phones and digital video

Manuscript received May 24, 2010; revised September 6, 2010; accepted October 21, 2010. Date of publication March 17, 2011; date of current version August 3, 2011. This work was supported by the China Scholarship Program, by the NSFC, under Grants 60736043 and 60828002, by the China 973 Program, under Grant 2009CB320905, and by the Qatar National Research Fund. This paper was recommended by Associate Editor R. Bernardini.

M. Guo was with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China. She is currently with MediaTek, Inc., Beijing 100080, China (e-mail: mguo@jdl.ac.cn).

Z. Xiong is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: zx@ece.tamu.edu).

F. Wu is with Microsoft Research Asia, Beijing 100080, China (e-mail: fengwu@microsoft.com).

D. Zhao is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: dbzhao@jdl.ac.cn).

X. Ji is with the Department of Automation, Tsinghua University, Beijing 100080, China (e-mail: xyji@mail.tsinghua.edu.cn).

W. Gao is with the Institute of Digital Media, Beijing University, Beijing 100080, China (e-mail: wgao@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2011.2129910

cameras are penetrating deeper in people’s lives, wireless video applications are becoming more and more popular. The challenge in these applications is to deliver satisfactory video quality when the compressed video is transported over wireless channels, which are unstable and noisy. To meet this challenge, the key issue is error-resilient video coding.

In standard video coding, consecutive frames are encoded jointly to achieve maximum coding efficiency. Denote  $X$  and  $Y$  as the current frame to be encoded and the motion-compensated previously decoded frame, respectively. With joint encoding, instead of compressing  $X$  directly, the motion-compensated residual, that is  $D = X - Y$ , is encoded to save bit rate. When the transmission channel is noiseless,  $D$  is perfectly reconstructed at the decoder and added to the prediction  $Y$ , so  $X$  is recovered perfectly at the decoder. The latest H.264/AVC standard allows advanced options such as variable block-size motion compensation, quarter-pel motion compensation, and directional spatial prediction to reduce the energy in  $D$ . Furthermore, context-adaptive variable-length coding or context-adaptive arithmetic coding (CABAC) can be employed for efficient compression of  $D$ .

Whereas joint encoding is most beneficial in terms of coding efficiency, it however renders the compressed bit-stream highly vulnerable to channel errors. This is because when either  $D$  or  $Y$  is reconstructed erroneously at the decoder,  $X = Y + D$  will be in error. In addition, decoding errors in the current frame  $X$  will propagate (till the end of the group of pictures), leading to error-drifting. Among the three coding modes, namely, Inter coding, Intra coding, and Skip, in H.264/AVC, Inter coding is chiefly responsible for error-drifting due to exploitation of temporal redundancy. On the other hand, Intra coding is the most error-robust due to its independent nature; consequently its coding efficiency is the lowest. Finally, the Skip mode is less sensitive to channel errors than the Inter mode because its reconstructed  $X$  depends only on  $Y$ . Hence, if the Inter mode is disabled, i.e., only the Intra and Skip modes are allowed in H.264/AVC, error resilience of H.264/AVC coding will be significantly enhanced—at the expense of coding efficiency.

A natural question to ask is whether we can have a coding scheme that offers the middle ground—with coding efficiency approaching that of Inter coding while being error resilient (as Intra coding). Distributed source coding (DSC), based on the paradigm of separate encoding and joint decoding, offers an alternative solution to error-robust video compression for networked multimedia. Slepian and Wolf [4] laid the theoretical foundation of DSC by showing that the rate of near-

lossless source coding with decoder side information only is the same as that of joint encoding (with side information at both the encoder and decoder). For lossy source coding with decoder side information, Wyner and Ziv [5] established the theoretical rate-distortion function.

For Slepian–Wolf code design, the idea of using linear codes was first illustrated in Wyner’s 1974 paper [6] and implemented recently using turbo or low-density parity-check (LDPC) codes. For Wyner–Ziv coding, most practical designs are based on the framework of quantization followed by Slepian–Wolf compression. In the quadratic Gaussian case, state-of-the-art trellis-coded quantization in conjunction with turbo/LDPC code based Slepian–Wolf coding have been shown to be able to approach the Wyner–Ziv rate-distortion limit.

Because of channel decoding (e.g., nearest neighbor decoding) techniques employed in the Slepian–Wolf decoder, a Slepian–Wolf coded bitstream is inherently error-robust. Wyner–Ziv video coding thus has the advantage of increased error-robustness over standard H.264/AVC when transporting video over noisy channels. This is one of the reasons behind many recent research interests in distributed video coding (see [7]–[29]).

The idea of video coding based on DSC principles was first proposed by Witsenhausen and Wyner in their 1980 U.S. patent on an “inter-frame coder for video signals” [30]. Despite the flurry of recent research activities on distributed video coding (DVC), the number of references to Witsenhausen and Wyner’s patent on Google scholar is only 18 (at the time of this writing) and the patent itself expired without actually being implemented.

In this paper, we first take a closer look at the patent and see how it is related to recent papers (e.g., [8], [29], [31]). We then propose a Witsenhausen–Wyner video coding (WWVC) scheme that replaces Inter coding in H.264/AVC by WWVC, while keeping the Intra and Skip modes of H.264/AVC unchanged. WWVC encodes each block itself instead of its motion-compensated residual, and the motion-compensated prediction (or side information) is only involved in making the encoding decisions among Intra, WWVC and Skip modes. We employ forward motion estimation at the encoder and send the motion vectors to help generate side information at the decoder for joint decoding. We also examine the tradeoff between the motion vector resolution and coding efficiency.

Within the WWVC mode, syndrome coding [6], [30] is employed to compress the quantized DCT coefficients of the current frame. The syndrome former [30] divides all quantized DCT coefficients into cosets of an error correction code (ECC) and outputs the index, namely the syndrome, of the coset that each quantized DCT coefficient belongs to. The syndrome decoder chooses the most probable quantized DCT coefficient from the coset indicated by the received syndrome with the help of decoder side information. In our proposed WWVC scheme, syndrome coding consists of two steps. First, nested scalar quantization (NSQ) employs a 1-D nested lattice [32] to determine the cosets and produces the nested quantization indices of the DCT coefficients. Then, bit-planes of the NSQ

indices are compressed by LDPC code based syndrome coding. Due to the variation of source correlation within the DCT domain and across bit-planes, the decision between syndrome coding and entropy coding is further optimized.

We compare WWVC and H.264/AVC bitstream transmission using a wireless channel simulator of RTP/IP [33] over 3GPP [34] networks from Qualcomm, Inc., San Diego, CA. Extensive simulations show that WWVC is more robust against channel errors than H.264/AVC. This error resilience with WWVC is achieved at a small loss of coding efficiency when compared to H.264/AVC under noiseless channel conditions. Moreover, compared to H.264/AVC-IntraSkip (i.e., Intra and Skip modes only in H.264/AVC), WWVC performs better in both noiseless and noisy channels.

The rest of this paper is organized as follows. Section I ends with a survey of related works. In Section II, we review Witsenhausen and Wyner’s patent and see how it is related to recent works. Section III describes our proposed WWVC scheme and Section IV explains the mode decision process between syndrome coding and entropy coding for each DCT band and for different bit-planes of each DCT coefficient. Simulation results of WWVC for both noiseless and noisy channels, as well as comparisons with those of H.264/AVC with different settings, are given in Section V. Section VI concludes this paper.

#### A. Related Works

Based on the theory in [4] and [5], many DVC techniques have been proposed (e.g., [7]–[28]) in recent years. In Wyner–Ziv coding (WZC), namely lossy source coding with side information at the decoder, the side information could be any correlated signals available at the decoder. In [7]–[28], the side information is derived from the neighboring decoded frames. Girod *et al.* proposed a DVC scheme with a low complexity encoder, which outputs parity bits to achieve compression, while the decoder counts on additional bits from the encoder through a feedback channel to adjust the bit rate [7]. The delay caused by the feedback requests prevents it from being used in real-time applications. Since the side information is generated with extrapolation or interpolation in [7], the size of GOP (i.e. group of pictures) is restricted and more intra-coded frames are required. A DSC-based video codec, which targets at flexible allocation of complexity between the encoder and decoder or error robustness, was presented in [8], [27], and [28]. In it, the transform coefficients and bit-planes of NSQ indices are encoded with a designated coding method (i.e., entropy or syndrome coding) without considering the correlation statistics of each frequency band or bit-plane. Moreover, when motion estimation is performed at the encoder, the depth of NSQ used for each individual DCT coefficient depends on the specific residual between the original coefficient and its side information, additional overhead indicating the depth information for each coefficient must be inserted into the bitstream. If the decoder utilizes cyclic redundancy check (CRC) to estimate the motion vectors and generate the side information, the coding performance suffers from fake side information blocks chosen by fixed-length CRC checksums that have finite error correction capability.

It is well known that side information estimation and correlation modeling play important roles in DVC. Methods based on interpolation [9], 3-D-models [10], universal prediction [11], motion search with some hash information [12], unsupervised learning of disparity [13], and iterative refinement [14], [15] have been proposed to produce side information at the decoder. The tradeoff between the rate-distortion performance and the complexity of motion estimation at the decoder was addressed in [16]. Since the original frame is absent from the motion estimation at the decoder, the quality of side information is much worse than that achieved by motion estimation at the encoder. The lowered correlation leads to a gap in coding efficiency between DVC with low complexity encoding and standard H264/AVC. Furthermore, in order to exploit the spatial correlation, DVC is typically implemented in the DCT [18], [20] or wavelet domain [19]. Correlation modeling in the transform domain was studied in [21]. Besides, channel codes such as turbo code [22], [23], LDPC codes [24], [31], and trellis code [25], [26] have been employed for DVC so that the bit rate can approach the conditional entropy of the video source given the side information.

Compared to the above DVC works, our proposed WWVC scheme has the following novelties. First, the resolution of motion vectors and the motion search range are selected based on the available computational power of the encoder. Second, the depth of NSQ is determined by the correlation statistics instead of the specific residual values. Thus, no overhead indicating each depth is needed. Third, both channel code based syndrome coding and entropy coding are adopted to compress the NSQ indices regardless of the channel condition. The choice between the two coding methods depends on the temporal and statistical correlation and the target coding efficiency. To sum up, in our proposed WWVC scheme, both the coding parameters and the coding modes are determined by the correlation among video frames.

Besides compression of a single video sequence with side information obtained from neighboring decoded frames, WZC has also been effectively employed in other video applications. Xu *et al.* proposed to utilize WZC to generate the enhancement layer in a layered video coding scheme, in which the side information is the reconstruction of the H.26L coded base layer [35], [36]. In [37], a Wyner-Ziv coded bitstream is generated to compensate for the quality gap when switching from one bitstream to another with different bit-rates, using the reconstructed video prior to switching as decoder side information. When WZC is applied to coding of multiterminal video sources (e.g., [38], [39]), the side information is extracted from neighboring terminals at the decoder. However, Witsenhausen and Wyner's patent, as discussed below, focused on DVC of a single video sequence.

## II. WITSENHAUSEN AND WYNER'S PATENT

### A. Inter-Frame Coder for Video Signals

In March 1980, Witsenhausen and Wyner issued a U.S. patent on an “inter-frame coder for video signals” [30], whose main idea is illustrated in Fig. 1, which is reproduced from [30]. Assume the video signal source is binary, and the current

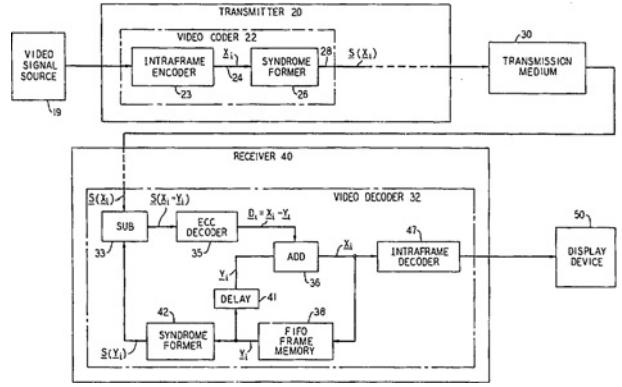


Fig. 1. “Inter-frame coder for video signals” in Witsenhausen and Wyner’s patent.

frame  $X$  and the previous frame  $Y$  are correlated, e.g., with i.i.d  $X$  and  $Y$  satisfying the relationship  $X = Y + D$ , with  $D$  representing the residual between  $X$  and  $Y$ . Each frame is divided into  $n$ -bit blocks, whose indices are omitted here. The syndrome former at the video encoder outputs  $S(X) = TX$ , which is the syndrome of  $X$ , where  $X$  is a block of the current frame and  $T$  denotes the  $(n-k) \times n$  parity-check matrix of an  $(n, k)$  ECC for the binary channel with  $X$  and the corresponding block in the previous frame  $Y$  being its input and output, respectively. Note that the correlation between  $X$  and  $Y$  should be taken into account when selecting the ECC. For each encoder input block  $X$  of length  $n$ , the output of the encoder  $S(X)$  has length  $n-k$ . Since only the syndrome  $S(X)$  is transmitted to the decoder, the encoder achieves a compression ratio of  $\frac{n}{n-k}$ . After receiving the transmitted syndrome  $S(X)$ , the video decoder first computes  $S(X) - S(Y) = T(X - Y)$ , since it already has the previous frame  $Y$ , before applying hard-decision decoding to recover  $D = X - Y$  noiselessly by picking  $D$  as the leader of the ECC coset indexed by  $S(X) - S(Y)$ . Finally,  $D = X - Y$  is added to  $Y$  to form the decoded  $X$ .

For general  $M$ -ary video sources, the patent suggests the use of  $M$ -ary ECCs or the above binary syndrome-based scheme for the most significant bit-plane in conjunction with conventional coding schemes for the remaining bit-planes. We shall take the latter approach in our work.

### B. Relationship Between Witsenhausen and Wyner’s Decoder and the Slepian–Wolf Decoder

In this subsection, we take a look at how Witsenhausen and Wyner’s patent relates to Slepian–Wolf coding (SWC) [4]. The syndrome-based encoding step in Witsenhausen and Wyner’s patent is the same as that in SWC [31]. The first decoding step in the patent, forming  $TD = T(X - Y)$ , could easily be used in SWC [31] and decode for  $D$  instead of  $X$  directly, as done in SWC [31]. The “noisy” vector in this case would be an all-zero vector instead of  $Y$ . The result of the decoding algorithm should then be added to  $Y$  to form the most likely  $X$ , similarly to the patent.

For example, assume the source  $X$  and the side information  $Y$  are equiprobable binary triplets with  $X, Y \in \{0, 1\}^3$  and they

differ at most in one position. The Slepian–Wolf encoder partitions the set of all possible outcomes of  $X$  into four cosets  $C_{00}$ ,  $C_{01}$ ,  $C_{10}$ , and  $C_{11}$  with  $C_{00} = \{000, 111\}$ ,  $C_{01} = \{001, 110\}$ ,  $C_{10} = \{010, 101\}$ , and  $C_{11} = \{100, 011\}$ , so that the two elements in any coset have Hamming distance  $d_H = 3$ . The encoder uses two bits to send the syndrome  $S$  of coset  $C_S$  that  $X$  belongs to. The joint decoder picks the  $X$  in coset  $C_S$  with  $d_H(X, Y) \leq 1$ . For instance, if  $X = [001]$  and  $Y = [101]$ , then  $X$  belongs to  $C_{01}$  and  $S(X) = [01]$  is sent. At the receiver, the decoder chooses the symbol  $[001]$  that is closest to the side information  $Y = [101]$  from  $C_{01}$ . Hence, with SWC it is possible to send  $H(X|Y) = 2$  bits instead of  $H(X) = 3$  bits for  $X$  and decode it losslessly at the joint receiver, where  $H(X)$  denotes the self entropy of  $X$  and  $H(X|Y)$  the conditional entropy of  $X$  given  $Y$ .

In Witsenhausen and Wyner's patent, the encoder is the same as that in SWC, which outputs syndrome  $S(X) = [01]$ . At the decoder, the side information  $Y = [101]$  is first fed into a syndrome former to obtain the syndrome  $S(Y) = [10]$ , since  $Y$  appears in the coset  $C_{10}$ . Then, the decoder calculates  $S(D) = S(X) - S(Y) = [11]$ . The coset leader  $[100]$  is selected from  $C_{11}$  as  $D$ . Finally,  $D$  is added to  $Y$  and get  $X = Y + D = [101] + [100] = [001]$ , which is identical to the result of Slepian–Wolf decoding.

Nevertheless, the coding scheme in Witsenhausen and Wyner's patent differs from SWC [31] in terms of the decoding algorithm. Hard-input decoding is used in the patent, whereas soft-input iterative message-passing decoding is used in [31]. This could be because at the time the patent was written syndromes were mainly used for decoding of block codes (hard-input) while soft-input decoding was only considered possible for convolutional codes through the Viterbi algorithm. However, the hard-decision Witsenhausen–Wyner decoder is optimum under the binary symmetric channel (BSC) model, meaning that in the setup of [31] it will give the same performance as the message-passing algorithm. The reason is that, in BSC model, the soft-input probabilities to the decoder only have two levels (more likely and less likely) and the decoding with the minimum Hamming weight is thus the maximum-likelihood approach. Of course, this will not be the case when the correlation model is not the BSC in which the soft-input to the decoder has more levels, e.g., different bits have different cross-over probabilities or there is an additive white Gaussian noise channel model. Since Witsenhausen and Wyner first conceived the idea of video compression based on distributed source coding principles and in some cases the decoder in their patent is optimal and acquires the same result as state-of-the-art Slepian–Wolf decoders, we advocate the use of Witsenhausen–Wyner coding when the previous video frame is adopted to generate the side information at the Slepian–Wolf/Wyner–Ziv decoder [7], [8]. We think of this as a tangible way of giving due credits to Witsenhausen and Wyner.

### III. PROPOSED WITSENHAUSEN--WYNER VIDEO CODER

We propose a Witsenhausen–Wyner video codec for robust video transmission over noisy channels. Our proposed encoder

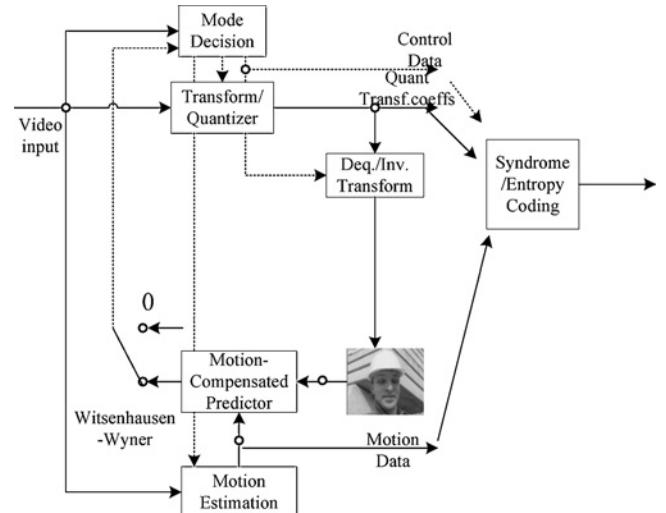


Fig. 2. Proposed Witsenhausen–Wyner video encoder. “0” in the figure signifies the Intra mode.

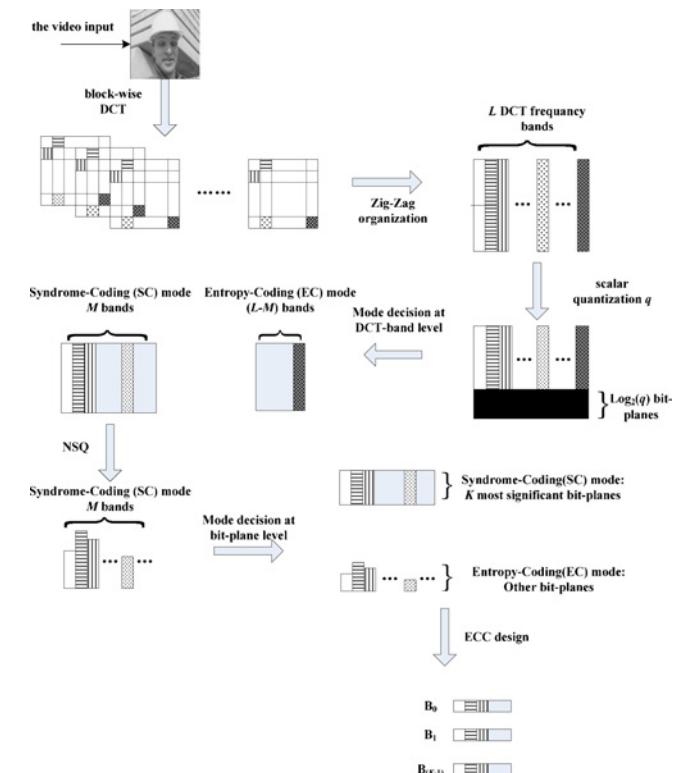


Fig. 3. Proposed encoding process of WWVC mode.

and decoder are illustrated in Figs. 2 and 4, respectively. It can be seen that the main difference from H.264/AVC coding is the replacement of Inter coding by WWVC, which achieves Inter-like coding efficiency while being insensitive to channel errors. Our proposed encoder does not directly compress the residual between the original video input and the side information. It only enlists the help of the side information in deciding among WWVC, Intra, and Skip modes for each original video block and in classifying the WWVC blocks. The mode decisions and the class indices, both included in “Control Data,” are

entropy coded and sent to the decoder. The side information is generated from the previously reconstructed frame with motion compensation at the encoder. The motion vector resolution is adaptively selected from full-, half-, and quarter-pixel based on the available computational resource of the transmitter. The complexity of interpolating the previous frame and motion search can be saved at the expense of degraded quality for the side information. The resulting motion vectors, signified as “Motion Data,” are entropy encoded as well before being sent to produce side information at the decoder. The Intra and Skip modes of H.264/AVC coding remain in our WWVC scheme, since they are inherently robust against channel errors.

#### A. Proposed Witsenhausen–Wyner Video Encoder

Our Witsenhausen–Wyner video encoder exploits the correlation in video signals at four levels. First, each video block is classified into Intra, WWVC, and Skip modes based on the correlation between itself and its motion compensated prediction (or side information). Second, within the WWVC mode, blocks with similar correlation statistics are further classified, and the coding parameters of each class are tuned to the correlation statistics. Third, depending on the correlation statistics, either syndrome coding or entropy coding is chosen for each DCT frequency band. Fourth, owing to the variation of correlation from the most to the least significant bit-plane, a decision on syndrome or entropy coding is also made at the bit-plane level for NSQ indices.

As illustrated in Fig. 3, the WWVC blocks go through DCT, NSQ, LDPC based syndrome coding and entropy coding in turn. We discuss each step in the sequel, with optimal decision between syndrome and entropy coding detailed in Section IV.

1) *Transform*: In order to remove the spatial redundancy, distributed coding is carried out in the transform domain. The block-based DCT is applied to both the original frame and its side information. In each video frame, the transform coefficients belonging to the same frequency band are grouped together. Denote  $X_i$  and  $Y_i$  ( $0 \leq i < L$ ) as the coefficients in the  $i$ th frequency band of the original and the side information frames. There are  $L$  frequency bands in total which are ordered in zig-zag manner starting from the DC (zero-frequency) band. In Fig. 3, the column corresponding to each band is filled with a specific pattern.

2) *Coding Mode Decision*: The mode and block-size decisions from H.264/AVC are used in WWVC with the only change of relabeling Inter as WWVC. A simple version of the rate-distortion optimization approach in H.264/AVC is employed, which calculates the rate-distortion cost of the predicted residual coefficients. The WWVC mode can have the block size from  $16 \times 16$  to  $8 \times 8$ . As for  $16 \times 16$ ,  $16 \times 8$ , and  $8 \times 16$  blocks, each of them will be divided into several non-overlapped  $8 \times 8$  blocks and coded one by one. Furthermore, all the  $8 \times 8$  blocks of WWVC mode are classified according to how they correlate with their side information. The correlation is calculated in terms of the mean squared error (MSE) between the transform coefficients of the WWVC blocks and their corresponding side information blocks. For each class and each DCT band, a set of coding parameters including  $N_{i,j}$  denoted as the depth of NSQ involved in syndrome coding and

the weighting factors at the joint decoder, is determined from the correlation statistics, where  $i$  and  $j$  represent the DCT band index and the class index, respectively. Details on determining  $N_{i,j}$  will be given shortly with the understanding that the class index  $j$  needs to be transmitted to the decoder.

3) *Scalar Quantization*: All DCT coefficients are scalar quantized with uniform step size  $q$ , resulting in quantized symbols

$$X_i^q = \text{int}\left(\frac{X_i}{q}\right) \quad (1)$$

where  $\text{int}(\cdot)$  denotes the function that returns the closest integer of a real number. Note that scalar quantization of  $X_i$  can be viewed as an operation that removes its  $\text{int}(\log_2(q))$  least significant bit-planes, which are marked in black in Fig. 3.

4) *Syndrome Encoding*: If the coding mode is WWVC, syndrome coding replaces entropy coding. Due to the varying correlation statistics, our coding method is tailored to each DCT band, with  $M$  low frequency bands being classified into syndrome coding mode (SC-mode) while the remaining  $(L - M)$  DCT bands into entropy coding mode (EC-mode), as shown in Fig. 3. Given a SC-mode coefficient  $X_{i,j}^q$  that is in the  $i$ th DCT band and taken from a block of class  $j$ , an ECC with minimum distance

$$d_{\min} = 2^{N_{i,j}} \quad (2)$$

is applied, resulting in NSQ index

$$B_{i,j} = X_{i,j}^q \bmod (2^{N_{i,j}}). \quad (3)$$

When

$$|Y_{i,j} - X_{i,j}^q \times q| < \frac{d_{\min}}{2} \quad (4)$$

the decoder will be able to identify the correct  $X_{i,j}^q$  by choosing the codeword closest to the corresponding side information coefficient  $Y_{i,j}$  in the coset identified by the index  $B_{i,j}$ . However, when

$$|Y_{i,j} - X_{i,j}^q \times q| \geq \frac{d_{\min}}{2} \quad (5)$$

the decoder is likely to select an incorrect codeword. Thus, besides the distortion introduced by scalar quantization, there exists distortion caused by the ECC.

Since the uniform quantization step size  $q$  is set as in H.264/AVC coding, the NSQ parameter  $N_{i,j}$  is determined based on the correlation between the source signal and the side information so that the distortion caused by the ECC can be minimized. For each class and each DCT band

$$N_{i,j} = \text{int}\left(\log_2\left(\frac{\gamma \times \sigma_{i,j}}{q}\right)\right) \quad (6)$$

where  $\sigma_{i,j}$  denotes the root MSE between  $X_{i,j}$  and  $Y_{i,j}$ . The more correlated  $X_{i,j}$  and  $Y_{i,j}$  are, the more information is expected to be inferred from  $Y_{i,j}$  about  $X_{i,j}$  at the decoder, leading to a smaller  $N_{i,j}$ . The weighting factor  $\gamma$  in (6) roughly describes the temporal correlation of video sequence and can be tuned based on the motion speed of video sequence. Furthermore, when the transmission is noisy, the degraded side

information leads to larger  $\sigma_{i,j}$  values. Instead of changing all  $\sigma_{i,j}$  values, we look into the variation and adjust  $\gamma$  accordingly.

The resulting NSQ indices  $B_{i,j}$  are expressed with  $N_{i,j}$  bit-planes in a top-down manner. In order to increase the code length of ECC and the coding efficiency accordingly, the NSQ indices from different DCT bands and the blocks of different classes are grouped together. As shown in Fig. 3, due to the correlation of each bit-plane with its corresponding side information,  $K$  most significant bit-planes  $E_0 E_1 \cdots E_{K-1}$  will fall into the SC-mode while the remaining bit-planes belong to EC-mode. Owing to the complexity and memory requirements of  $M$ -ary ECCs, we adopt multilevel LDPC codes to compress  $E_0 E_1 \cdots E_{K-1}$  using the syndrome-based approach of [31].

Each bit-plane in SC-mode is individually encoded with one LDPC code. Only the syndrome bits are transmitted to the decoder (to achieve compression). The rate of syndrome coding for  $E_i$  ( $0 \leq i \leq K-1$ ) depends on the conditional entropy  $H(E_i|Y, E_0, E_1, \dots, E_{i-1})$  [40], [41], which represents the minimum rate needed for lossless recovery of  $E_i$  given previously decoded bit-planes  $E_0 \cdots E_{i-1}$  and the side information  $Y$  at the decoder. The subscripts of  $Y$  are omitted here, since there is no need to distinguish the DCT band and the class. In our irregular LDPC code designs, the code degree distribution polynomials are optimized using density evolution [42] under Gaussian approximation. The bipartite graph for the irregular LDPC code, which determines the sparse parity check matrix, is then randomly constructed based on the optimized code degree distribution polynomials.

5) *Entropy Coding:* Due to weakened correlation with the side information, the  $(L - M)$  high frequency coefficients of EC-mode and  $(N_{i,j} - K)$  EC-mode bit-planes extracted from SC-mode coefficients are compressed with CABAC in H.264/AVC coding.

#### B. Proposed Witsenhausen–Wyner Video Decoder

Fig. 4 illustrates the proposed Witsenhausen–Wyner video decoder. At the receiver, Intra and Skip blocks are decoded as in H.264/AVC. The previously reconstructed frame is used to generate the side information frame together with the received motion vectors. After the same size DCT as at the encoder, the coefficients of each band  $Y_i$  ( $0 \leq i < L$ ) are produced.

1) *Entropy Decoding:* The entropy coded bit-stream is decompressed directly to restore the EC-mode coefficients and bit-planes. The  $(N_{i,j} - K)$  least significant bit-planes are reproduced and used to reconstruct the NSQ indices  $B_{i,j}$  together with the syndrome decoded  $K$  most significant bit-planes. The intra coded high frequency coefficients of EC-mode are decompressed to obtain the quantized symbols  $X_{i,j}^q$ .

2) *Syndrome Decoding:* As discussed in Section II, the hard-decision decoder in Witsenhausen and Wyner's patent is not in general optimal, we thus employ the soft-decision decoder used in [31]. Since we use multilevel LDPC codes instead of  $M$ -ary ECCs to compress SC-mode bit-planes, given a bit-plane  $E_i$ , the received syndrome bits are jointly decoded together with the side information  $Y$  and the previously decoded bit-planes. The message-passing algorithm [43] is employed for iterative LDPC decoding, in which the received syndrome bits correspond to the check nodes on the bipartite

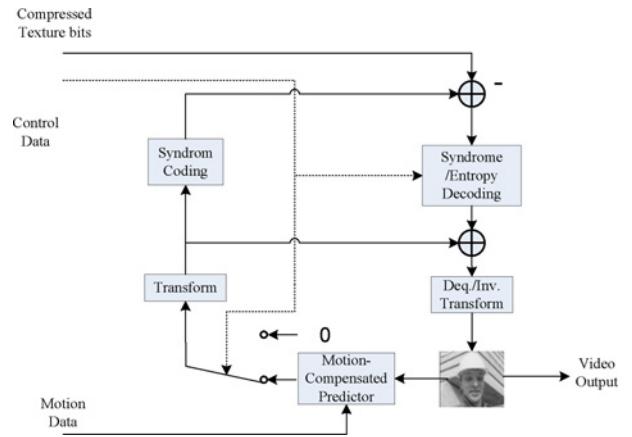


Fig. 4. Proposed Witsenhausen–Wyner video decoder. “0” in the figure signifies the Intra mode.

graph. The log-likelihood ratio (LLR), which describes the *a prior* probability about the candidate decoding output can be calculated with the side information and previously decoded bit-planes as

$$LLR = \log \frac{p(E_i = 0 | E'_0, \dots, E'_{i-1}, Y)}{p(E_i = 1 | E'_0, \dots, E'_{i-1}, Y)} \quad (7)$$

where  $p(\cdot|\cdot)$  denotes the conditional probability function and  $E'_i$  the decoded  $E_i$ . According to the chain rule

$$\begin{aligned} H(E_0 E_1 \cdots E_{K-1} | Y) &= H(E_0 | Y) + H(E_1 | E_0, Y) \\ &\quad + \cdots + H(E_{K-1} | E_0, \dots, E_{K-2}, Y). \end{aligned} \quad (8)$$

Thus, there is no rate loss caused by using multilevel LDPC codes.

3) *Dequantization:* The optimal reconstructed coefficient  $X'_{i,j}$  is estimated based on the quantization indices  $X_{i,j}^q$  and the side information  $Y_{i,j}$  as

$$X'_{i,j} = \frac{\omega_{i,j}^s \times (X_{i,j}^q \times q) + \omega_{i,j}^q \times Y_{i,j}}{\omega_{i,j}^s + \omega_{i,j}^q}. \quad (9)$$

The weighting factor  $\omega_{i,j}^s$  is just the  $\sigma_{i,j}^2$ , and  $\omega_{i,j}^q$  calculated from the MSE between  $X_{i,j}$  and  $(X_{i,j}^q \times q)$ .

Finally, the inverse DCT is applied to the reconstructed coefficients to obtain the reconstructed video output.

#### IV. OPTIMAL DECISION BETWEEN SYNDROME CODING AND ENTROPY CODING

In predictive coding such as H.264/AVC, the temporal redundancy is mainly removed by computing the motion-compensated residual. Entropy coding is used to further eliminate the statistical redundancies in the residual. However, in WWVC, the blocks are coded independently of the side information, so the encoder needs to select between syndrome and entropy coding to minimize the bit rate. When the channel is error-prone, we assume that the previous frame  $F_{t-1}$  is completely lost when the temporal correlation of frame  $F_t$  is computed. The reconstructed frame  $F'_{t-1}$  is extracted from the previous reconstruction  $F'_{t-2}$  and used to model the correlation of  $F_t$ . Without loss of generality, we assume that the joint

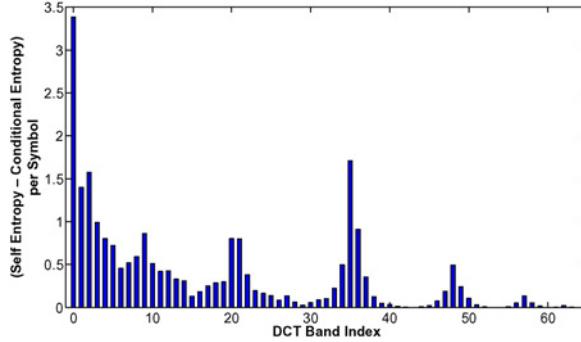


Fig. 5.  $H(X_i^q) - H(X_i^q|Y_i)$  for all DCT bands.

distribution of the original DCT coefficients  $X$  and the side information  $Y$  is jointly Gaussian, where  $X = Y + D$ . The residual  $D$  is zero-Gaussian and independent of  $Y$ . Given the distortion  $Z$ , the ideal bit rate of WZC denoted as  $R_{WZ}(Z)$  only depends on  $\sigma$  with

$$R_{WZ}(Z) = \max\left(\frac{1}{2} \log_2 \frac{\sigma^2}{Z}, 0\right). \quad (10)$$

$\sigma$  is the root MSE between  $X$  and  $Y$ .

#### A. Decision at DCT-Band Level

Since the temporal correlation tends to decrease from low-frequency to high-frequency bands, we need to determine how to code each DCT band individually. Toward this end, we look at the statistics in terms of the self-entropy of  $H(X_i^q)$  and the conditional entropy  $H(X_i^q|Y_i)$  for each DCT band. Since conditioning does not increase entropy,  $H(X_i^q|Y_i) \leq H(X_i^q)$  in general. We thus use syndrome coding as the default coding mode, but switch to entropy coding if the difference between  $H(X_i^q)$  and  $H(X_i^q|Y_i)$  is very small. An example of  $H(X_i^q) - H(X_i^q|Y_i)$  for all  $8 \times 8$  DCT bands is plotted in Fig. 5. Accordingly,  $M$  low-frequency bands are classified into the SC-mode and the remaining  $(L - M)$  bands belong to the EC-mode. When the transmission channel is noisy, the deteriorated side information magnified  $\sigma$  by a factor of  $\eta$ , which implies that an additional bit rate of

$$\Delta R_{WZ} = \log_2(\eta) \quad (11)$$

is needed. Therefore, a smaller  $M$  has to be adopted.

#### B. Decision at Bit-Plane Level

As for the  $M$  low-frequency bands, the NSQ indices are represented from the top to bottom bit-planes as  $E_0 E_1 \dots E_{N-1}$ . Given a bit-plane  $E_i$ , depending on the difference between the self-entropy  $H(E_i)$  and the conditional entropy  $H(E_i|Y, E_0, \dots, E_{i-1})$ , which are the minimum bit rates for entropy coding and syndrome coding, respectively, a decision between SC-mode and EC-mode is made. Two examples of  $H(E_i) - H(E_i|Y, E_0, \dots, E_{i-1})$  are shown in Fig. 6. For each bit-plane, a larger difference implies that syndrome coding saves more bit rate over entropy coding. Accordingly,  $K$  most significant bit-planes are syndrome coded, and the remaining bit-planes are entropy coded. In noisy networks, the depth of

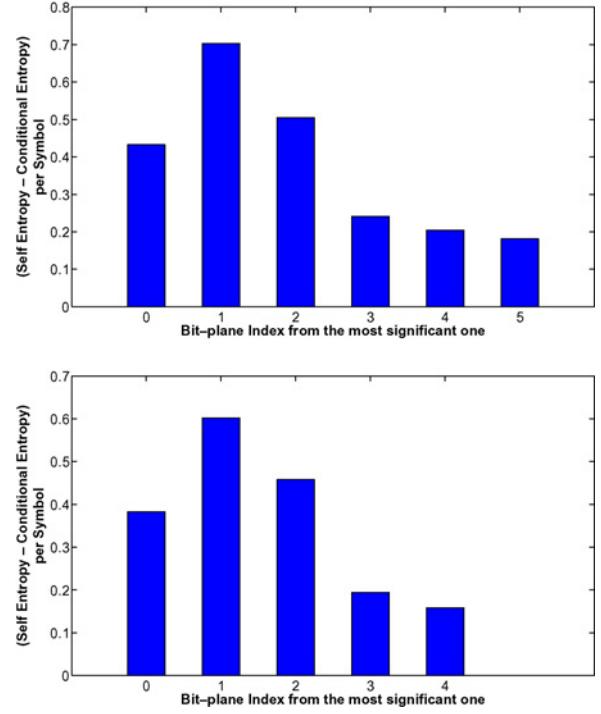


Fig. 6. With noiseless transmission,  $H(E_i) - H(E_i|Y, E_0, E_1, \dots, E_{i-1})$  for all bit-planes taken from NSQ indices which contain six bit-planes (upper) and five bit-planes (lower), respectively.

NSQ  $N_{i,j}$  is enlarged by the increased  $\sigma$  according to (6). The bit-planes are also divided into SC-mode and EC-mode according to the differences of two entropy values, which are exemplified in Fig. 7.

## V. EXPERIMENTAL RESULTS

Our proposed WWVC scheme aims at achieving error resilience when the compressed video bitstream is transmitted over noisy wireless networks. We evaluate the performance of our WWVC scheme under both noiseless and noisy channel conditions, and compare WWVC with H.264/AVC, H.264/AVC-IntraSkip which disables Inter mode and only allows Intra and Skip modes, H.264/AVC-IntraRefresh which forces 20% macroblocks to be intra coded, and H.264/AVC-Flexible macroblock ordering (FMO). Since the motion vector resolution is selected according to the available computational resource at the encoder in WWVC, we generate results of WWVC, H.264/AVC, H.264/AVC-IntraRefresh, and H.264/AVC-FMO with full-, half-, and quarter-pixel motion estimation. When the channel is noiseless, the coding efficiency is given in terms of PSNR versus bit-rate; when the transmission is noisy, the average PSNR values of the reconstructions achieved at the decoder with different packet loss rates are adopted to assess error robustness of various schemes. Moreover, the encoding time of WWVC and H.264/AVC is also given.

#### A. Coding Efficiency in Noiseless Channels

We evaluate the coding efficiency of WWVC in this subsection. For simplicity but without loss of generality,  $16 \times 16$ ,

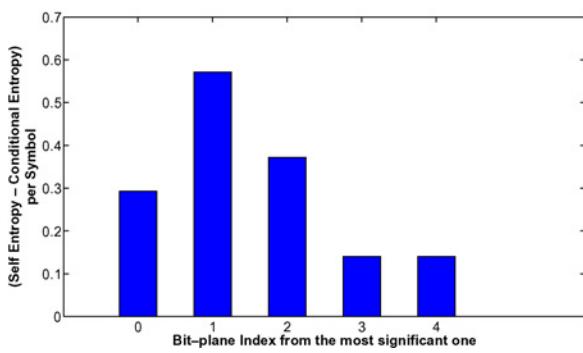
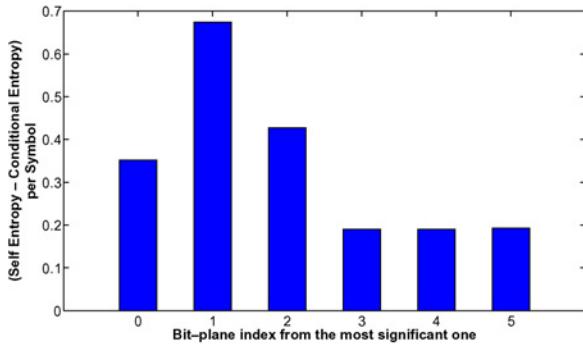


Fig. 7. With noisy transmission,  $H(E_i) - H(E_i|Y, E_0, E_1, \dots, E_{i-1})$  for all bit-planes taken from NSQ indices which contain six bit-planes (upper) and five bit-planes (lower), respectively.

TABLE I  
CLASSIFICATION OF THE WWVC-MODE BLOCKS

Class Index	<i>Football</i>	<i>Mobile</i>	<i>Susie</i>
0	$[0, 70.5]$	$[0, 62.5]$	$[0, 23]$
1	$[70.5, 142)$	$[62.5, 230)$	$[23, +\infty)$
2	$[142, 254)$	$[230, 613.5)$	—
3	$[254, 425)$	$[613.5, +\infty)$	—
4	$[425, 687.5)$	—	—
5	$[687.5, 1096)$	—	—
6	$[1096, 1861.5)$	—	—
7	$[1861.5, +\infty)$	—	—

$16 \times 8$ ,  $8 \times 16$  block-wise motion estimation and  $8 \times 8$  block-wise DCT are used. CABAC is chosen for entropy coding in H.264/AVC with different settings. The EC-mode coefficients of the WWVC-mode blocks are also coded with CABAC. Each coding slice contains one frame. Data partition is enabled in all schemes and each bitstream is divided into header, intra and inter (or WWVC) parts. Each of them is coded and delivered individually. The first 15 frames from the *Football* ( $720 \times 486$ , 15 f/s), *Mobile* ( $720 \times 576$ , 15 f/s), and *Susie* ( $720 \times 486$ , 15 f/s) sequences are used in our simulations.

Since these sequences have different amount of motion, we use different settings in our simulations accordingly. For example, the WWVC-mode blocks of *Football* are classified into eight classes while four and two classes are used for *Mobile* and *Susie*, respectively. The thresholds for classification in terms of the MSE between the transform coefficients of WWVC blocks and those of their corresponding side information are listed in Table I. Thirty-six low-frequency bands are included in the SC-mode for *Football* and all

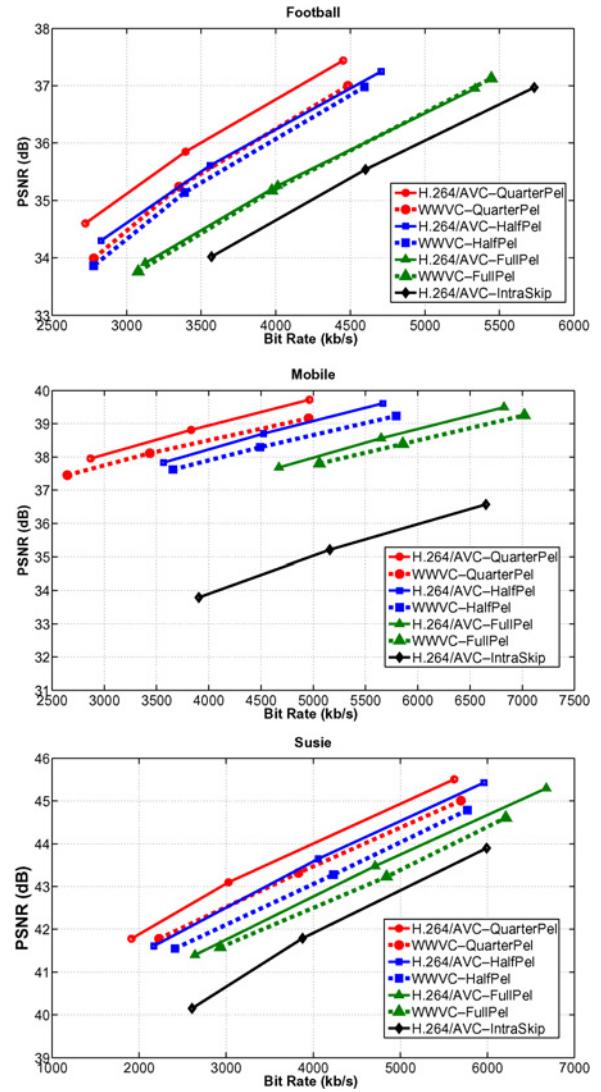


Fig. 8. Coding efficiency of the proposed WWVC scheme, H.264/AVC coding scheme, and H.264/AVC-IntraSkip coding scheme.

coefficients belong to the SC-mode for *Mobile* and *Susie*. For all three sequences, the three most significant bit-planes are fed into the irregular LDPC based syndrome encoder while the remaining bit-planes are entropy coded. Hence, three binary LDPC codes are designed for each sequence, with all SC-mode coefficients in one frame are grouped together for syndrome coding.

Each LDPC code is designed as follows. First, the bit-rate of syndrome bits for the bit-plane  $E_i$  is selected based on the conditional entropy  $H(E_i|E_0, E_1, \dots, E_{i-1}, Y)$ . A small redundancy is added so that the error bit probability after ECC decoding is under  $10^{-6}$ . The corresponding LDPC rate is equal to one minus the chosen bit rate of syndrome bits. Second, the degree distribution polynomials of the LDPC codes are optimized using Gaussian approximation. We adopt the LDPC codes given in [44]. Finally, the bipartite graph is generated randomly. The block lengths for *Football*, *Mobile*, and *Susie* are roughly  $1.0 \times 10^5$ ,  $1.8 \times 10^5$ , and  $1.5 \times 10^5$  bits, respectively.

The coding efficiency in terms of PSNR versus bit-rate is given in Fig. 8. With the same motion estimation, the WWVC

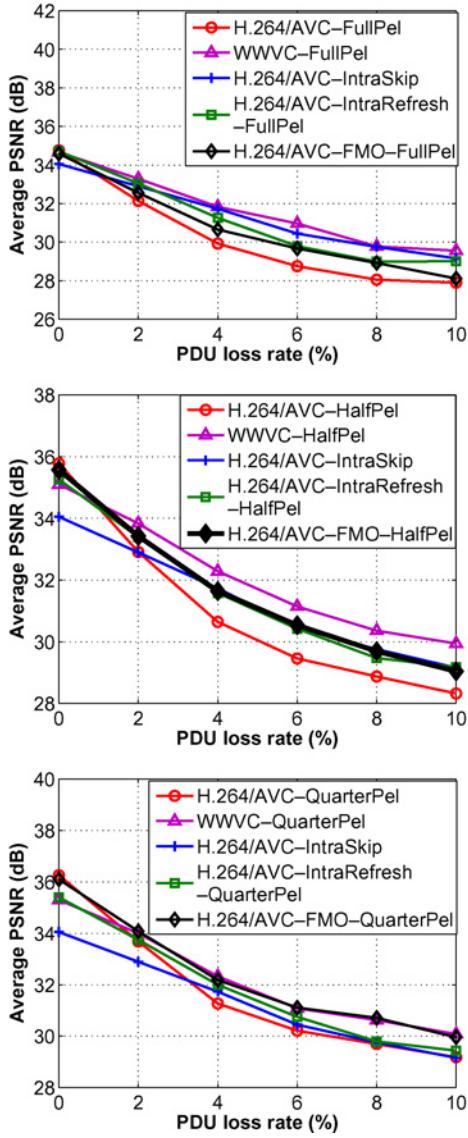


Fig. 9. Error robustness of the proposed WWVC scheme, H.264/AVC coding scheme, H.264-IntraSkip coding scheme, H.264/AVC-FMO coding scheme, and H.264/AVC-IntraRefresh coding scheme tested with *Football* sequence.

curve is a little bit lower than that of H.264/AVC. It can be observed that both H.264/AVC and WWVC suffer from the reduced complexity of motion estimation. However, the gap between H.264/AVC and WWVC is narrowed when the resolution of motion estimation is reduced. Since the Skip mode takes advantage of motion vectors, less accurate motion vectors tend to produce fewer Skip blocks and more Inter (or WWVC) blocks. The increased code length of LDPC codes improves the coding efficiency of syndrome coding. Besides, H.264/AVC-IntraSkip scheme gives the worst coding efficiency because much less temporal redundancy is removed.

#### B. Error Robustness in Noisy Channels

We adopt the wireless channel simulator for RTP/IP [33] over 3GPP [34] from Qualcomm, Inc., to test the error robustness of WWVC and H.264/AVC with different settings. The simulator transmits the real-time transport protocol (RTP)

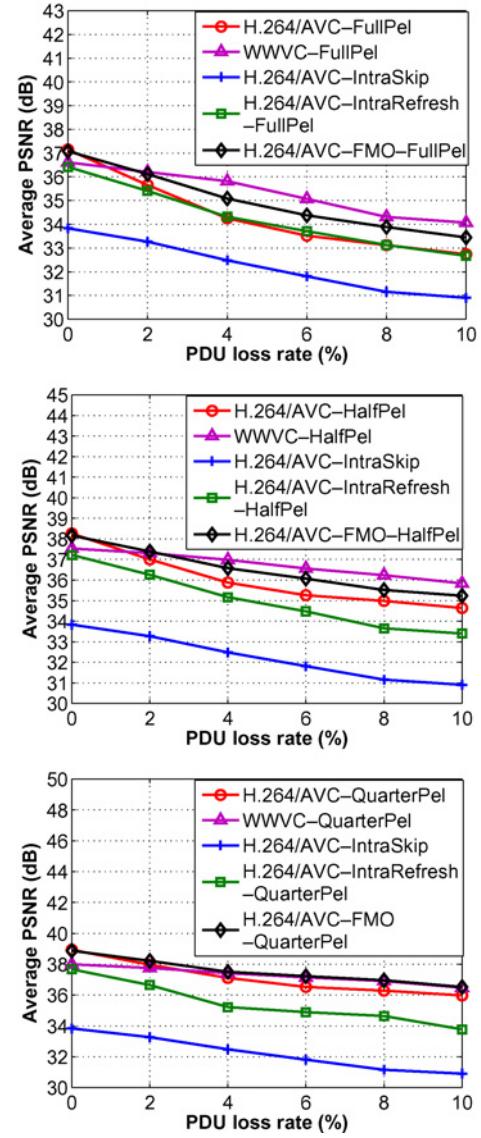


Fig. 10. Error robustness of the proposed WWVC scheme, H.264/AVC coding scheme, H.264-IntraSkip coding scheme, H.264/AVC-FMO coding scheme, and H.264/AVC-IntraRefresh coding scheme tested with *Mobile* sequence.

stream with a radio channel. The transmission rate is 64 kb/s in our simulation. Each RTP packet is fragmented into equal-size protocol data units (PDU). In the simulation each PDU contains 640 bytes. Channel errors are randomly introduced to the PDUs. If all the PDUs belonging to one packet are received no later than the maximum end-to-end delay, the packet is considered successfully received by the decoder. The Qualcomm simulator also provides FEC simulation with Reed-Solomon (RS) code.

As mentioned before, all generated bitstreams are partitioned into header, intra, and inter (or WWVC) parts. In all simulations, the header part is assumed to be error free. When the packet loss randomly happens to the other two parts, a simple error concealment scheme, which extracts the reconstruction from the previously decoded frame with the motion vectors included in the header, is utilized. We assume

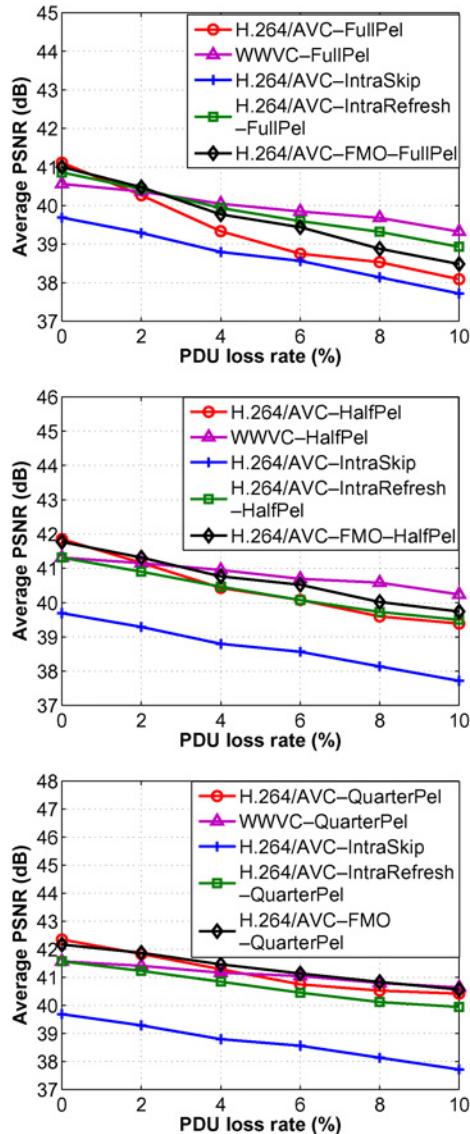


Fig. 11. Error robustness of the proposed WWVC scheme, H.264/AVC coding scheme, H.264-IntraSkip coding scheme, H.264/AVC-FMO coding scheme, and H.264/AVC-IntraRefresh coding scheme tested with *Susie* sequence.

ideal error detection. When large errors in previous frame are detected, the LLR in (7) is set to be zero. It means that when the previous frame is not reliable, no prior knowledge from it is utilized.

For fair comparisons, we use the same bit rate to generate the bitstreams for different coding schemes. The bit rates for *Football*, *Mobile*, and *Susie* are 3780, 4280, and 2400 kb/s, respectively. We add 25% FEC overhead to each bitstream, i.e., 20% of the overall transmission rate is used for RS-based FEC. Thus, the overall transmission rate for *Football*, *Mobile*, and *Susie* are 4725, 5163, and 3000 kb/s, respectively.

Due to the latency requirement and packet fragmentation during transmission, there are still residual RTP packet losses at the decoder. For *Football* sequence, the residual RTP packet loss rates corresponding to 2%, 4%, 6%, 8%, and 10% PDU loss rates are 0.56%, 1.39%, 2.47%, 3.65%, and 4.97%, respectively. For *Mobile* sequence, the corresponding residual

TABLE II  
ENCODING TIME (IN S) OF WWVC AND H.264/AVC

Scheme	QuarterPel	HalfPel	FullPel
H.264/AVC	127.33	121.42	111.49
WWVC	155.36	152.34	143.76

RTP packet loss rates are 0.55%, 1.34%, 2.39%, 3.58%, and 5.00%. For *Susie* sequence, the residual RTP packet loss rates are 0.65%, 1.45%, 2.25%, 3.53%, and 4.89%, respectively. For each PDU loss rate, 100 runs are simulated.

Figs. 9–11 depict the average PSNR versus PDU loss rate achieved with different video coding schemes. It can be observed from each figure that the curves corresponding to WWVC and H.264/AVC have a cross point. WWVC becomes more advantageous over H.264/AVC as the channel gets worse. No matter whether the channel is noisy or noiseless, H.264/AVC-IntraSkip is inferior to WWVC. FMO and IntraFresh are adopted in H.264/AVC for error robustness. We thus include them in our simulations. In H.264/AVC-IntraRefresh, 20% macroblocks are forced to be intra-coded in each frame. H.264/AVC-IntraRefresh in fact offers a tradeoff between H.264/AVC and H.264/AVC-IntraSkip in terms of coding efficiency and error robustness. In our FMO simulations, we use two slices with the “Dispersed” map. Experimental results show that WWVC is better than FMO with full-pixel and half-pixel motion estimation and is on par with FMO at quarter-pixel motion estimation.

The encoding time for H.264/AVC(WWVC)-QuarterPel, H.264/AVC(WWVC)-HalfPel, and H.264/AVC(WWVC)-FullPel is given in Table II. The bitstreams of *Susie* shown in Fig. 11 are used in these comparisons. The same motion estimation and mode decision algorithms are used in both WWVC and H.264/AVC. The added complexity with WWVC comes from float-precision DCT transform, NSQ, and frame-based syndrome coding with LDPC codes.

## VI. CONCLUSION

Inspired by Witsenhausen and Wyner’s 1980 patent, we proposed the first WWVC scheme that targets at robust video transmission over noisy channels. The main idea is to replace predictive Inter coding in H.264/AVC by WWVC, which gives close to Inter coding efficiency while being error robust. The reason is that we employed forward motion estimation at the encoder and sent the motion vectors to help generate side information at the decoder for WWVC. In addition, within WWVC, we optimized the decision between syndrome coding and entropy coding among different DCT bands and among different bit-planes within each DCT coefficient. Simulation results in terms of both coding efficiency and error resilience (in various networks conditions) are given, indicating that WWVC achieves better performance than H.264/AVC and H.264/AVC-IntraSkip over noisy networks. The price paid for enhanced error-resilience with WWVC is a small loss in compression efficiency.

## REFERENCES

- [1] B. Haskell, A. Puri, and A. Netravali, *Digital Video: An Introduction to MPEG-2*. New York: Chapman and Hall, 1996.
- [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. 1–19, Jul. 2003.
- [3] D. Marpe, T. Wiegand, and G. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *IEEE Commun. Mag.*, vol. 44, no. 8, pp. 134–143, Aug. 2006.
- [4] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [5] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [6] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. 20, no. 1, pp. 2–10, Jan. 1974.
- [7] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 1–12, Jan. 2005.
- [8] R. Puri, A. Majurndar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 1–13, Oct. 2007.
- [9] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *Proc. 5th Eurasip Conf. Speech Image Process. Multimedia Commun. Services*, Jul. 2005 [Online]. Available: [http://amalia.img.lx.it.pt/~cicb/papers/EURASIP'05\\_ImprovingFI.pdf](http://amalia.img.lx.it.pt/~cicb/papers/EURASIP'05_ImprovingFI.pdf)
- [10] M. Maitre, C. Guillemot, and L. Morin, "3-D model-based frame interpolation for distributed video coding of static scenes," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1246–1257, May 2007.
- [11] Z. Li, L. Liu, and E. Delp, "Wyner-Ziv video coding with universal prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 11, pp. 1430–1436, Nov. 2006.
- [12] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver," in *Proc. ICIP*, vol. 5. Oct. 2004, pp. 3097–3100.
- [13] D. Varodayan, D. Chen, M. Flierl, and B. Girod, "Wyner-Ziv coding of video with unsupervised motion vector learning," *Signal Process. Image Commun.*, vol. 23, pp. 369–378, Jun. 2008.
- [14] B. Macchiavello, D. Mukherjee, and R. de Queiroz, "Iterative side-information generation in a mixed resolution Wyner-Ziv framework," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 10, pp. 1409–1423, Oct. 2009.
- [15] R. Martins, C. Brites, J. Ascenso, and F. Pereira, "Refining side information for improved transform domain Wyner-Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1327–1341, Sep. 2009.
- [16] Z. Li, L. Liu, and E. Delp, "Rate distortion analysis of motion side estimation in Wyner-Ziv video coding," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 98–113, Jan. 2007.
- [17] L. Liu, Z. Li, and E. Delp, "Efficient and low-complexity surveillance video compression using backward-channel aware Wyner-Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 453–465, Apr. 2009.
- [18] A. Aaron, S. D. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *Proc. SPIE VCIP*, vol. 5308. Jan. 2004, pp. 520–528.
- [19] X. Guo, Y. Lu, F. Wu, and W. Gao, "Distributed video coding using wavelet," in *Proc. ISCAS*, 2006, pp. 5427–5430.
- [20] D. Rebollo-Monedero, S. Rane, and B. Girod, "Wyner-Ziv quantization and transform coding of noisy sources at high rates," in *Proc. 38th Asilomar Conf. Signals Syst. Comput.*, vol. 2. Nov. 2004, pp. 2084–2088.
- [21] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, pp. 1177–1190, Sep. 2008.
- [22] A. Liveris, Z. Xiong, and C. Georghiades, "A distributed source coding technique for correlated images using turbo codes," *IEEE Commun. Lett.*, vol. 6, no. 9, pp. 379–381, Sep. 2002.
- [23] S. Yasakethu, W. Weerakkody, W. Fernando, F. Pereira, and A. Kondoz, "An improved decoding algorithm for DVC over multipath error prone wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 10, pp. 1543–1548, Oct. 2009.
- [24] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 376–387, Feb. 2009.
- [25] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.
- [26] A. Avudainayagam, J. M. Shea, and D. Wu, "Hyper-Trellis decoding of pixel-domain Wyner-Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 557–568, May 2008.
- [27] S. Milani, J. Wang, and K. Ramchandran, "Achieving H.264-like compression efficiency with distributed video coding," in *Proc. SPIE VCIP*, vol. 6508. Jan. 2007, p. 65082Z.
- [28] J. Wang, V. Prabhakaran, and K. Ramchandran, "Syndrome-based robust video transmission over networks with bursty losses," in *Proc. ICIP*, Oct. 2006, pp. 741–744.
- [29] A. Aaron, D. Varodayan, and B. Girod, "Wyner-Ziv residual coding of video," in *Proc. PCS*, Apr. 2006 [Online]. Available: <http://www.stanford.edu/~bgirod/pdfs/AaronPCS06.pdf>
- [30] H. Witsenhausen and A. Wyner, "Interframe coder for video signals," U.S. Patent 4 191 970, Mar. 1980.
- [31] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, Oct. 2002.
- [32] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.
- [33] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RFC 1889: Rtp: A transport protocol for real time applications," Jan. 1996.
- [34] P. Chaudhury, W. Mohr, and S. Onoe, "The 3GPP proposal for IMT-2000," *IEEE Commun. Mag.*, vol. 37, no. 12, pp. 72–81, Dec. 1999.
- [35] Q. Xu and Z. Xiong, "Layered Wyner-Ziv video coding," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3791–3803, Dec. 2006.
- [36] Q. Xu, V. Stankovic, and Z. Xiong, "Wyner-Ziv video compression and fountain codes for receiver-driven layered multicast," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 901–906, Jul. 2007.
- [37] M. Guo, Y. Lu, F. Wu, D. Zhao, and W. Gao, "Wyner-Ziv switching scheme for multiple bit-rate video streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 569–581, May 2008.
- [38] Y. Yang, V. Stankovic, Z. Xiong, and W. Zhao, "Two-terminal video coding," *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 534–551, Mar. 2009.
- [39] X. Guo, Y. Lu, F. Wu, D. Zhao, and W. Gao, "Wyner-Ziv based multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 6, pp. 713–724, Jun. 2008.
- [40] S. Cheng and Z. Xiong, "Successive refinement for the Wyner-Ziv problem and layered code design," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3269–3281, Aug. 2005.
- [41] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.
- [42] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
- [43] S. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 657–670, Feb. 2001.
- [44] A. Amraoui. (2001). *LdpcOpt* [Online]. Available: <http://ipgdemos.epfl.ch/ldpcopt>



**Mei Guo** received the B.S. degree in computer science from the Beijing University of Posts and Telecommunications, Beijing, China, in 2004, and the M.S. and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2007 and 2010, respectively.

From 2005 to 2007, she was with Microsoft Research Asia, Beijing, as an Intern. She was a Visiting Student with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, from 2008 to 2010. She is currently with MediaTek, Inc., Beijing, China. Her current research interests include distributed video coding, video compression, and video streaming.



**Xixiang Xiong** (S'91–M'96–SM'02–F'07) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1996.

From 1995 to 1997, he was with Princeton University, Princeton, NJ, first as a Visiting Student and then as a Research Associate. From 1997 to 1999, he was with the University of Hawaii, Honolulu. Since 1999, he has been with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, where he is currently a Professor.

He was with Microsoft Research, Redmond, WA, in 1998 and 1999. His sabbatical leave was with Stanford University, Stanford, CA, in 2010. His current research interests include network information theory, code designs and applications, networked multimedia, and genomic signal processing.

Dr. Xiong received the National Science Foundation Career Award in 1999, the Army Research Office Young Investigator Award in 2000, and the Office of Naval Research Young Investigator Award in 2001. He received the 2006 IEEE Signal Processing Magazine Best Paper Award. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 1999 to 2005, the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2002 to 2005, the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2006, and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS (PART B) from 2005 to 2009. He is currently an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was the Publications Chair of ICASSP 2007, the Technical Program Committee Co-Chair of ITW 2007, and the Tutorial Chair of ISIT 2010.



**Feng Wu** (M'99–SM'06) received the B.S. degree in electrical engineering from Xidian University, Xi'an, China, in 1992, and the M.S. and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1996 and 1999, respectively.

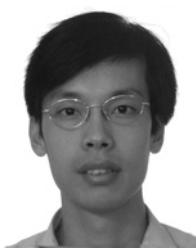
He joined Microsoft Research China, Beijing, China, as an Associate Researcher in 1999. He has been a Researcher with Microsoft Research Asia since 2001. He has been an active contributor to ISO/MPEG and ITU-T standards. Some techniques have been adopted by MPEG-4 FGS, H.264/MPEG-4 AVC, and the coming H.264 SVC standard. He served as the Chairman of the China AVS Video Group from 2002 to 2004 and led the efforts on developing the China AVS Video Standard 1.0. He has authored or co-authored over 100 conference and journal papers. He has about 30 U.S. patents granted or pending in video and image coding. His current research interests include image and video representation, media compression and communication, computer vision, and graphics.



**Debin Zhao** received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1985, 1988, and 1998, respectively.

He is currently a Professor with the Department of Computer Science, School of Computer Science and Technology, Harbin Institute of Technology. He has published over 200 technical articles in refereed journals and conference proceedings in the areas of image and video coding, video processing, video streaming and transmission, and pattern recognition.

Dr. Zhao was the recipient of three National Science and Technology Progress Awards of China (Second Prize).



**Xiangyang Ji** received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1999 and 2001, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Beijing, China, in 2008.

He is currently a Lecturer with the Department of Automation, Tsinghua University, Beijing. He has authored or co-authored over 50 conference and journal papers. His current research interests include video/image coding, video streaming, and

multimedia processing.



**Wen Gao** (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan.

He is currently a Professor of computer science with the Institute of Digital Media, Beijing University, Beijing, China. Before joining Beijing University, he was a Full Professor of computer science with the Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and with the Chinese Academy of Sciences (CAS), Beijing, from 1996 to 2005. With CAS, he served as a Professor, the

Managing Director of the Institute of Computing Technology, the Executive Vice President of the Graduate School of CAS, and the Vice President of the University of Science and Technology of China. He has published extensively, including four books and over 500 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics.

Dr. Gao is the Editor-in-Chief of the *Journal of Computers* (a journal of the China Computer Federation), an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, an Associate Editor of the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, an Area Editor of the EURASIP Journal of Image Communications, and an Editor of the *Journal of Visual Communication and Image Representation*. He chaired a number of prestigious international conferences on multimedia and video signal processing, and also served on the advisory and technical committees of numerous professional organizations.