High-Quality Symmetric Wyner–Ziv Coding Scheme for Low-Motion Videos

Journal of Electronic Imaging

SPIEDigitalLibrary.org/jei

High-quality symmetric Wyner–Ziv coding scheme for low-motion videos

Xianfeng Ou
Enrico Masala
Linbo Qing
Xiaohai He

SPIE

IS&T
imaging.org

High-quality symmetric Wyner–Ziv coding scheme for low-motion videos

Xianfeng Ou,[a] Enrico Masala,[b,*] Linbo Qing,[a] and Xiaohai He[a,*]
[a]Sichuan University, College of Electronics and Information Engineering, No. 24 South Section 1, Yihuan Road, Chengdu, 610064 China
[b]Politecnico di Torino, Control and Computer Engineering Department, corso Duca degli Abruzzi 24, Torino, 10129 Italy

Abstract. Traditional Wyner–Ziv video coding (WZVC) structures require either intra (Key) or Wyner–Ziv (WZ) coding of frames. Unfortunately, keeping the video quality approximately constant implies drastic bit-rate fluctuations because consecutive frames of different types (Key or WZ) present significantly different compression performances. Moreover, certain scenarios severely limit rate fluctuation. This work proposes a WZVC scheme with low bit-rate fluctuations based on a symmetric coding structure. First, this work investigates the performance of a generic nonasymmetric distributed source coding structure, showing that the low-density parity-check accumulate channel decoding method is best suited. This is used as a basis to design a symmetric WZVC scheme in which every input video frame is divided into four parallel subframes through subsampling, and then the subframes are encoded by using a symmetric method. Compared with the traditional asymmetric WZVC scheme, the proposed scheme can achieve higher bit-rate stability over time, which is a great advantage to guarantee a reliable transmission in many wireless communication application environments in which bit-rate fluctuations are strongly constrained. Simulation results show the effectiveness of the proposed symmetric WZVC scheme in maintaining a steady bit rate and quality, as well as a quality comparison with the traditional WZVC scheme.
 [DOI: 10.1117/1.JEI.23.6.061112]

Keywords: nonasymmetric source coding; Wyner–Ziv video quality; symmetric Wyner–Ziv video coding; bit-rate stability; subsampling.

Paper 14154SS received Mar. 27, 2014; revised manuscript received Aug. 19, 2014; accepted for publication Aug. 28, 2014; published online Sep. 15, 2014.

1 Introduction

Traditional video coding standards, such as H.264/AVC,[1–3] present high encoder complexity due to the attempt to exploit redundancies at the encoder. However, it is possible to shift the complexity at the decoder by means of the so-called distributed video coding (DVC)[4,5] paradigm. The theoretical foundations of DVC were established in the 1970s by Slepian–Wolf[6] for lossless distributed source coding (DSC) and Wyner–Ziv[7] for the lossy case. Different from traditional video coding standards, DVC, also referred to as Wyner–Ziv video coding (WZVC), presents low encoder complexity since complicated processes are shifted from the encoder to the decoder side. There, redundancy is used to generate side information (SI) which then assists in optimizing the decoding performance. The low encoding complexity characteristic makes DVC suitable for a number of potential applications, for instance, those in which terminals have limited power supply, small storage and computation capacity, such as video surveillance, field training, intelligent transportation system, wireless sensor networks, etc.

While DVC techniques are interesting in principle, their application in practical scenarios requires overcoming significant difficulties. Decoding complexity is one of the main issues to be solved. This partially stems from the use of two alternative coding schemes for video frames: the traditional intracoding method (used for Key frames) and the Wyner–Ziv method (used for WZ frames). The former presents limited encoding and decoding complexity and it is often implemented by relying on the intracoding mode of some standards, such as H.264/AVC. The latter presents very limited encoding complexity (that is the main advantage of DVC) but decoding is very complex. Clearly, different coding modes have significantly different compression performances: Key frames compression rate is the lowest one and WZ is the highest one. This fact can cause significant problems in a real scenario due to the drastic bit-rate fluctuations, which may negatively affect the performance of the transmission system, leading to temporary network congestion or bandwidth waste. This is particularly true if the network has been provisioned on the basis of the peak bandwidth value. Additionally, the use of two different types of coded frames, namely Key and WZ, typically requires anticipating Key frame decoding since this is needed to generate the SI necessary for WZ frame decoding. This fact negatively affects the possibility to parallelize the decoding process.

The requirement for two types of encoded frames, namely Key and WZ, is a direct consequence of the underlying coding paradigm based on asymmetric DSC. In fact, the term asymmetric indicates that sources use different encoding and decoding methods. This situation has two main drawbacks. First, the compression performance is different for the various sources, which implies that the complexity and power consumption of the various encoding methods are different. Second, channel bandwidth occupation is unbalanced among sources. In addition, in asymmetric DSC, the decoding process cannot be easily parallelized because it is necessary to first decode a subset of sources (e.g., the Key frames),

*Address all correspondence to: Enrico Masala and Xiaohai He, E-mail: enrico.masala@polito.it and hxh@scu.edu.cn

0091-3286/2014/$25.00 © 2014 SPIE and IS&T

which are then used to generate SI necessary to decode the other sources (e.g., WZ frames). To overcome this limitation, nonasymmetric systems have been proposed, in which all sources use a similar encoding structure. These schemes encompass both completely symmetric DSC structures and others in which it is possible to change the amount of symmetry among the encoding processes of the sources, so that fine-tuning of the encoding rate, complexity, and power consumption of the encoding process is possible for each source. Moreover, an interesting advantage of nonasymmetric systems is the possibility to decode sources in parallel.

A few nonasymmetric DSC schemes have been proposed in the literature. For instance, Pradhan and Ramchandran[8] proposed a method to implement nonasymmetric DSC based on the DISCUS algorithm.[9] Partial information of the two sources is encoded using independent, entropy coding based methods and transmitted directly to the decoder side, whereas the other information is encoded using DSC methods. At the decoder side, the two received sources were decoded by making use of the mutual information that is directly transmitted. Sartipi and Fekri[10] and Garcia-Frias and Cabarcas[11] proposed similar nonsymmetric DSC schemes. At the encoder side, the two sources X and Y are encoded through systematic codes to generate, as output, encoded information which contains two parts: the information bits and the parity bits. At the decoder side, the information bits of X and the parity bits of Y are combined to generate the SI of X; then the parity bits of X are used to complete channel decoding and reconstruction of X. The difference of the two symmetric schemes lies in the channel codes, Sartipi and Fekri using low-density parity-check (LDPC) and Garcia-Frias and Cabarcas using Turbo. The schemes presented in Refs. 10 differ for the designs of LDPC and Turbo encoders, which are based on the source correlation. In fact, when the source correlation changes, the puncture matrix should also be changed in order to adjust the rate distribution of the sources. The reconstruction of Y uses a similar process. Schonberg et al.[12] proposed a nonasymmetric DSC scheme by using a channel code split method. At the encoder side, generation matrices are produced by extracting subsets from the main LDPC generation matrix. Then, different sources use different generation matrices to implement a rate adjustable system, while at the decoder side, the sequence computed as the difference between the two sources (the diversity sequence) is decoded first. Zarasoa et al.[13] proposed a nonsymmetric DSC scheme based on syndrome LDPC. At the encoder side, sources X and Y are encoded independently using LDPC; then the system outputs equal length syndrome code words in parallel. The scheme in Ref. 13 designed and used different encode matrices according to the correlation of the sources. At the same time, in order to ascertain the location of the information bits, k linear dependent column vectors are extracted from a parity matrix of size $(n-k)\times n$. When the parity matrices change the encoder complexity, the information bits increase, too. At the decoder side, the diversity sequence is generated using XOR operations on the syndrome, and then the diversity sequence is used to decode the sources. In the context of nonasymmetric DSC methods, this paper investigates the performance of a rate-adjustable nonasymmetric DSC scheme based on source splitting method, investigating, in particular, the performance for different channel decoding methods. The results of this analysis are then used to design a new WZVC scheme as detailed later.

Several research works tried to apply the idea of symmetric source coding for the case of video compression and communication. Most notably, the multiple description coding (MDC) paradigm[14] relies on the possibility to generate independent descriptions of the same media source with the following properties: each description can be decoded independently and if an additional description is available at the decoder, the quality of the decoded media increases. However, in certain situations, such as in the case of packet lossy channels, descriptions might not be entirely received, causing drift between the decoder and encoder. Therefore, error robustness strategies must be employed. DVC schemes have shown to be useful in this context. For instance, Ref. 15 showed the relation between predictive MDC and WZ SI computation, determining the achievable rate-distortion region as well as the operational rate-distortion performance under certain restrictions. Results are provided for first-order Gauss–Markov sources. A similar system is proposed in Ref. 16 where two SI streams are considered instead of one. The advantage of using SI in MDC is to eliminate the prediction loop.[17] In addition, the possibility to use, at the decoder, the SI from the other side allows one to avoid quality degradation in case of losses and to maintain symmetry between the two descriptions. These characteristics have also been shown in Ref. 18, where cross-decoding increases performance for the case of two correlated memory-less Gaussian sources. Reference 19 proposes a novel DVC scheme that relies on traditional MDC, that is, when one channel does not work, the lost DVC description is efficiently estimated by the one received from the other channel. An alternative but conceptually similar scheme based on discrete cosine transform zero padding is presented in Ref. 20.

Other aspects of DVC have also been subject of investigation, such as the SI computation. Reference 21 presents a new model to estimate the correlation noise, which is an important element for SI generation. Reference 22 focuses on low delay by adopting an SI iterative refinement approach and SI extrapolation-based techniques. Improving the SI by means of combining global and local motion estimation is another possibility,[23] which is particularly useful for sequences with high global motion. The utility of SI has also been shown in the context of scalable coding,[24] where it is exploited at both the base and the enhancement layers. Finally, combining DVC techniques designed for the pixel and the transform domain has been shown to yield significant advantages in terms of compression gains and increased robustness against channel losses.[25]

Reference 26 uses subsampled spatial information instead of interpolated temporal information, which clearly provides a better performance for the condition of high-motion video sequences and a larger group of pictures (GOP). This technique is arguably the closest to the one presented in this paper, but, unfortunately, the lack of technical details in the work does not allow a throughout comparison except for a single value final performance result.

The main objective of this work is to propose a high-quality symmetric WZVC scheme based on subsampling of the source information (i.e., video frames) and to evaluate its performance in terms of bit rate and video quality. The

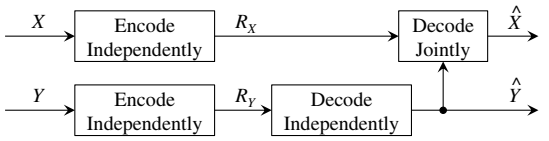


Fig. 1 The basic functional block diagram of the typical asymmetric distributed source coding (DSC).

video coding scheme builds on a nonasymmetric DSC scheme whose best configuration is evaluated in the first part of the paper.

The rest of this paper is organized as follows. Section 2 presents the theoretical background. A nonasymmetric DSC structure is introduced in Sec. 3, including the encoder and decoder, and its performance analysis using different channel codes. In Sec. 4, we propose a symmetric WZVC scheme based on the nonasymmetric DSC structure introduced in Sec. 3. The simulation results of both the nonasymmetric DSC structure and the symmetric WZVC scheme are presented in detail in Sec. 5. Finally, Sec. 6 draws the conclusions.

2 Theoretical Background

Figure 1 shows the basic functional block diagram of the typical asymmetric DSC.

According to the DSC theorems,[6,7] in order to decode and reconstruct sequences X and Y at the decoder, the rates should satisfy the following constraints:

$$R_X \geq H(X|Y), \quad R_Y \geq H(Y|X), \quad R_X + R_Y \geq H(X,Y), \tag{1}$$

where $H(X|Y)$ and $H(Y|X)$ are the conditional entropies and $H(X,Y)$ is the joint entropy. Equation (1) determines the reachable rate of DSC, which is shown in Fig. 2. The broken line A-B-C-D visually shows the rate limit of DSC.

Different coding schemes are possible at the encoder side, each one corresponding to a different segment of the rate limit. For instance, in the asymmetric DSC, the two input sources use different coding methods. A source can be decoded directly and then the result is used to generate the SI for the other one at the decoder side. In this case, the theoretical rate limit is either the B or the C point, and the achievable rate region is the light shadow area shown in Fig. 2. Another case is the nonasymmetric DSC, in which the two sources are encoded independently at the encoder side. Then, at the decoder side, the decoded information of both sources is used to generate the SI of the other one. The key idea is that they mutually assist each other in decoding. In this case,

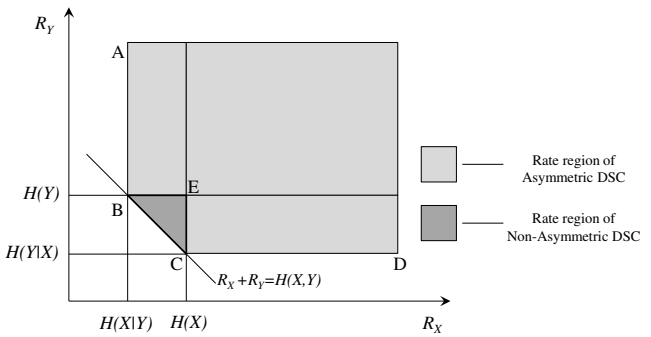


Fig. 2 The reachable rate region of DSC.

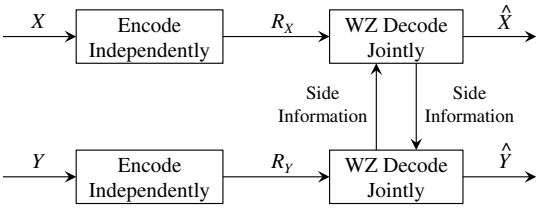


Fig. 3 Nonasymmetric model.

the theoretical rate limit is line B-C and the desired rate region is the triangular region formed by B, C, and E.

3 Design of a Nonasymmetric DSC

The basic functional block diagram of a nonasymmetric model is shown in Fig. 3.

Sources X and Y are encoded independently but decoded jointly. Partial decoded information of Y is used to assist in decoding X, and vice versa. This paper investigates the performance of a nonasymmetric DSC scheme using different channel codes and low-density parity-check accumulate (LDPCA) in particular. This scheme will be introduced in the following section in detail.

3.1 *Encoder of the Nonasymmetric DSC*

Figure 4 shows the encoder of the nonasymmetric DSC structure based on source splitting. Sources X and Y are two independent sequences with the same length. The first n bits of source X are encoded using standard entropy coding. The result of this operation is indicated by I_X and directly transmitted to the decoder. At the same time, the remaining m bits are encoded using LDPCA (Ref. 27) to obtain the syndrome S_X that is transmitted to the decoder. For source Y, the operations are similar except for the amount of bits coded using the standard entropy coding and LDPCA. In particular, for Y, the first n bits are coded using LDPCA to get syndrome S_Y, whereas the other m bits of source Y are encoded using the standard entropy coding. The result is indicated by I_Y and it is directly transmitted to the decoder. Note that the sum of n and m is equal to the length of X and Y. Changing the values of n and m allows achieving different points in the rate region B-C-E in Fig. 2.

3.2 *Decoder of the Nonasymmetric DSC*

In the nonasymmetric DSC structure, shown in Fig. 5, the received I'_X and I'_Y are decoded and then combined together to generate a sequence, which is used to assist in decoding the received syndromes S'_X and S'_Y, respectively. This last step allows one to reconstruct the final values $\hat{X}$ and $\hat{Y}$. Note that in this process, the decoding results of I'_X and I'_Y are both used as SI for sources Y and X, respectively.

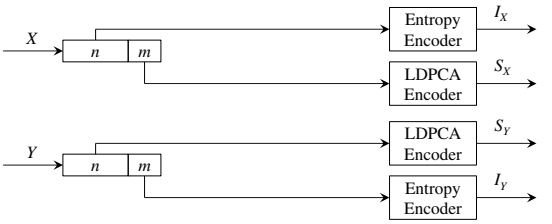


Fig. 4 The encoder of the nonasymmetric DSC.

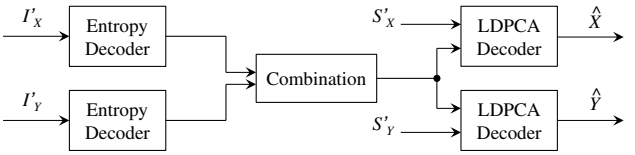


Fig. 5 The decoder of the proposed nonasymmetric DSC.

3.3 *Performance Analysis of the Nonasymmetric DSC*

In the scheme, parts of the information bits of the sources are transmitted to the decoder directly after entropy coding. The amount of information bits can be selected without constraints. At the decoder, the information bits are used as SI since they are present in both sources. Therefore, both sources can be decoded in parallel. Also, by changing the ratio in the channel decoding process, it is possible to adjust the rate of the source in this nonasymmetric DSC structure.

Considering the two structures, an asymmetric DSC and a nonasymmetric DSC one, for comparison purposes, we assume that both use the same channel codes (n, k). Thus, k is the length of the input sources and $n-k$ is the length of the output to be transmitted after channel coding. The probabilities of 0 and 1 in the input sources X and Y are assumed to be 0.5.

In the asymmetric DSC structure, X is encoded independently; hence, the output rate is

$$R_X \geq H(X) = 1. \tag{2}$$

After DSC, the output rate of Y is

$$R_Y \geq H(Y|X) = H(X,Y) - H(X). \tag{3}$$

If there is no puncturing of the encoder structure,

$$R_{Y\max} = (n-k)/k. \tag{4}$$

If no changes occur in source X, the correlation of X and Y is determined by source Y, and the maximum joint entropy $H(X,Y)$ of asymmetric DSC is

$$H(X,Y)_{\max\_\text{asymmetric}} = H(X) + R_{Y\max} = 1 + (n-k)/k. \tag{5}$$

For a certain channel code (n, k), the compression efficiency is

$$(n-k)/k = H(X,Y)_{\max\_\text{asymmetric}} - 1. \tag{6}$$

In the nonasymmetric DSC structure, instead, there are $a \cdot k$ and $b \cdot k$ bits $(a, b > 0, a + b = 1)$ using the entropy coding method, which are transmitted to the decoder side. The values of a and b determine the rate distribution. The other $a'(n-k)$ and $b'(n-k)$ output bits are transmitted after DSC encoding (a', b' represent the puncturing ratio). The output rate of nonasymmetric DSC is

$$\begin{aligned} R_{X,Y} &= R_X + R_Y = [a\cdot k + a'(n-k)]/k + [b\cdot k + b'(n-k)]/k \\ &= 1 + (a'+b')(n-k)/k. \end{aligned} \tag{7}$$

It is known that, in order to achieve lossless decoding, the encoding rate R_X and R_Y should satisfy the following constraint:

$$R_X + R_Y \geq H(X,Y), \tag{8}$$

where $H(X,Y)$ is the joint entropy of the two correlated sources X and Y. If there is no puncturing ($b' = 1$, $a' = 1$), the maximum output of nonasymmetric DSC is

$$H(X,Y)_{\max\_\text{nonasymmetric}} = R_{(X,Y)\max} = 1 + 2\cdot(n-k)/k. \tag{9}$$

We can compute the compression efficiency as

$$(n-k)/k = [H(X,Y)_{\max\_\text{nonasymmetric}} - 1]/2. \tag{10}$$

The previous analyses show the difference between the maximum joint entropy for asymmetric DSC [given by Eq. (5)] and nonasymmetric DSC [given by Eq. (9)]. For the same rate, the maximum joint entropy of the nonasymmetric DSC is higher, that is to say, asymmetric DSC has a better performance than nonasymmetric DSC. Equations (6) and (10) show the compression efficiency in terms of the maximum joint entropy for both schemes. The value of Eq. (10) grows more slowly than the one for Eq. (6). It is known from information theory principles that the joint entropy $H(X,Y)$ of X and Y depends on the correlation, and generally speaking, the lower the correlation, the higher is the value of $H(X,Y)$. Therefore, for a certain channel code with a given encoding ability, the nonasymmetric DSC is more suitable than asymmetric DSC for the case in which the sources have lower correlation.

4 Proposed Symmetric WZVC Scheme

This work proposes a symmetric WZVC scheme based on the nonasymmetric DSC structure introduced in Sec. 3 to design a system with very low rate fluctuation and capable of decoding frames in parallel. For simplicity, in the practical video coding case presented in this work, we adopt a fixed and equal split of the encoding rate which does not change, different from what would be possible in theory as pointed out in the previous section.

4.1 *Encoder of the Proposed Symmetric WZVC Scheme*

Every input video frame is divided into four subframes by means of subsampling, as shown in Fig. 6. Every circle represents the pixel location of the input video frame and the circles are denoted by four different numbers. Therefore, the four subframes are generated by taking the pixels denoted by the same number.[28]

Figure 7 shows the encoder of the proposed symmetric WZVC scheme. We assume that X and Y represent the odd frames and even frames of a video sequence, respectively. Any input frame X is divided into four subframes, which are named X_1, X_2, X_3, and X_4. X_1 and X_4 are encoded using H.264 intracoding and their coding rate is indicated by R_{IX_1} and R_{IX_4}, respectively; X_2 and X_3 are encoded using Wyner–Ziv coding. In more detail, the Wyner–Ziv coding

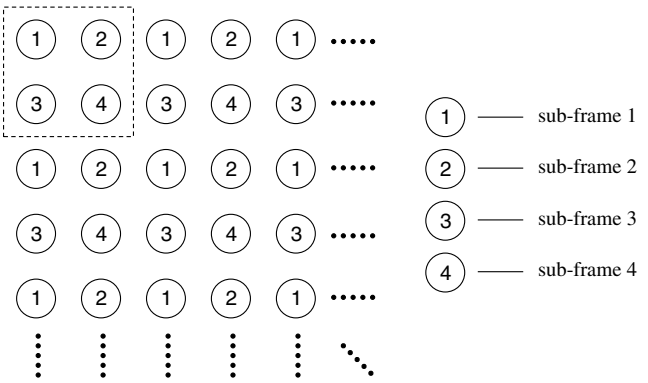


Fig. 6 Four subframes extracted by means of subsampling.

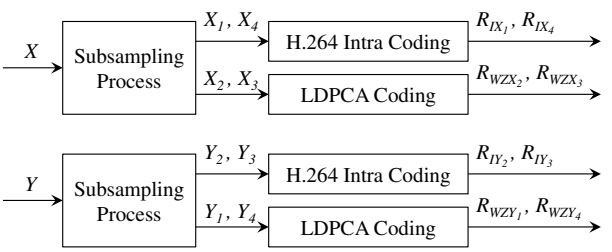


Fig. 7 The encoder of the proposed symmetric Wyner–Ziv video coding (WZVC) scheme.

scheme uses the LDPCA channel code and the syndromes are transmitted on request from the decoder. The coding rates of X_2 and X_3 are indicated by R_{WZX_2} and R_{WZX_3}, respectively.

The coding scheme of the next frame Y is analogous to the one of X shown in Fig. 7, but in this case, H.264 intracoding is used for Y_2 and Y_3 (as opposed to Y_1 and Y_4), whereas Y_1 and Y_4 are coded using the Wyner–Ziv coding method. Therefore, the coding method of the subframes of X and Y is symmetric.

4.2 *Decoder of the Proposed Symmetric WZVC Scheme*

At the decoder, SI is a predicted version of the current input video frame and it is one of the most important factors for achieving a good performance in WZVC. Figure 8 shows the proposed symmetric WZVC decoding scheme.

The decoding process works as follows. To decode X, first the received streams R'_{IX_1} and R'_{IX_4} are decoded to reconstruct the intracoded subframes, indicated by X'_1 and X'_4 in the scheme, corresponding to X_1 and X_4, respectively. The reconstructed intrainformation of X is stored in the reference subframes buffer and used to generate intraframes SI, $\mathrm{SI}_{\mathrm{IWZX}_2}$ and $\mathrm{SI}_{\mathrm{IWZX}_3}$, useful for the received streams, R'_{WZX_2} and R'_{WZX_3}. The subframes previously stored in the buffer are also used to generate temporal SI, $\mathrm{SI}_{\mathrm{TY}_1}$ and $\mathrm{SI}_{\mathrm{TY}_4}$, for the received streams, R'_{WZY_1} and R'_{WZY_4}, according to Ref. 29 and assuming rigid motion. However, when the actual motion is not rigid, errors in the temporal SI may become much more apparent. Thus, it is necessary to exploit the intraframe correlation among the subframes by means of a fusion operation to compute the SI finally used to decode the received streams R'_{WZY_1} and R'_{WZY_4}. Several schemes for information fusion have been proposed, as explained in Ref. 30, which analyzed SI fusion schemes and their performance. In this work, the following operation is used to generate the SI to assist the decoding of the WZ streams R'_{WZX_2} and R'_{WZX_3} of the current frame. The intraframe SI is computed as a linear combination of the available intraframe information, according to the following formulas:

$$\begin{aligned}\mathrm{SI}_{\mathrm{IWZX}_2}[t,(x,y)] &= \frac{1}{2}X'_1[t,(x+mvx_1,y)] \\ &\quad + \frac{1}{2}X'_4[t,(x,y+mvy_4)] \\ \mathrm{SI}_{\mathrm{IWZX}_3}[t,(x,y)] &= \frac{1}{2}X'_1[t,(x,y+mvy_1)] \\ &\quad + \frac{1}{2}X'_4[t,(x+mvx_4,y)],\end{aligned} \tag{11}$$

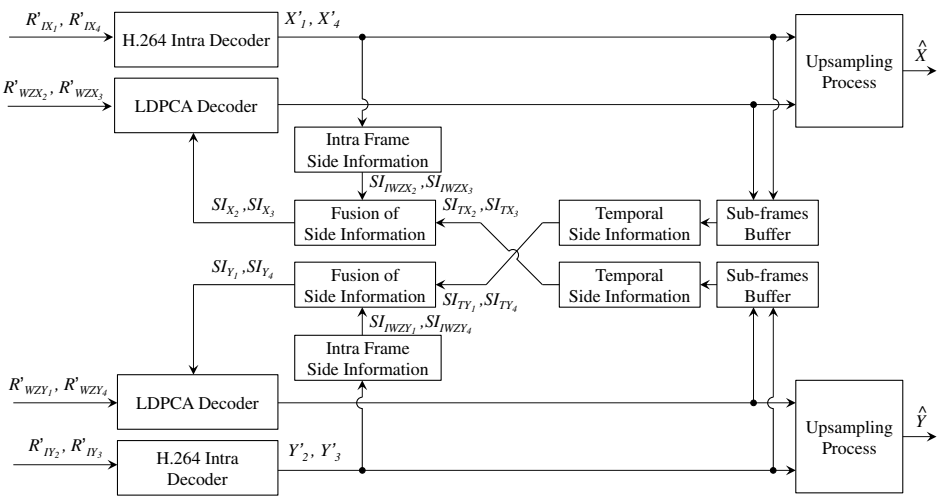


Fig. 8 Decoder of the proposed symmetric WZVC scheme.

where t is the time of the current frame that needs to be decoded and X'_1 and X'_4 represent the reconstructed subframes corresponding to X_1 and X_4, respectively. The symbols mvx_1 and mvy_1 represent the two components of the motion vector of the macro block in key subframe X'_1 calculated by using reference key subframe X'_4, and vice versa for mvx_4 and mvy_4. However, we assume that there is no need to shift when the SI is based on intrainformation that originally had the same component in the coordinates. For instance, when computing SI for subframe X_2, the y component is not shifted since the pixels of subframe X_1 have the same y in Fig. 6. The same applies for the x component when computing the SI for subframe X_2 from subframe X_4. Also, note that the purpose of Eq. (11) is to cope with the disparity caused by pixel shifts due to the subsampling process. Thus, the search range employed for the motion vector computation is much smaller than the one normally employed for the traditional motion compensated interpolation (MCI). Therefore, the computational load caused by Eq. (11) is much lower than the one of traditional MCI employed for temporal SI estimation.

We define the residual errors of the temporal and intra-SI as follows:

$$\begin{aligned}
E_{\mathrm{TX}_2}[t,(x,y)] &= |Y'_2[t_f,(x+mvx_{2f},y+mvy_{2f})] - Y'_2[t_b,(x+mvx_{2b},y+mvy_{2b})]| \\
E_{\mathrm{TX}_3}[t,(x,y)] &= |Y'_3[t_f,(x+mvx_{3f},y+mvy_{3f})] - Y'_3[t_b,[x+mvx_{3b},y+mvy_{3b})]| \\
E_{\mathrm{IWZX}_2}[t,(x,y)] &= |X'_1[t,(x+mvx_1,y)] - X'_4[t,(x,y+mvy_4)]| \\
E_{\mathrm{IWZX}_3}[t,(x,y)] &= |X'_1[t,(x,y+mvy_1)] - X'_4[t,(x+mvx_4,y)]|,
\end{aligned} \tag{12}$$

where t has the same meaning as in Eq. (11), t_f and t_b are the times of the forward and backward key subframes with respect to the current frame being decoded, mvx_{2f} and mvy_{3f} are the forward motion vectors, and mvx_{2b} and mvy_{3b} are the backward ones. Finally, an SI fusion scheme is adopted to fuse the two types of SI, i.e., temporal ($\mathrm{SI}_{\mathrm{TX}_2}$, $\mathrm{SI}_{\mathrm{TX}_3}$) and intra ($\mathrm{SI}_{\mathrm{IWZX}_2}$, $\mathrm{SI}_{\mathrm{IWZX}_3}$). The final SI of subframes X_2 and X_3, which is used by the LDPCA decoder, is given by

$$SI_{X_i}(x,y) = \begin{cases} \mathrm{SI}_{\mathrm{TX}_i}(x,y), & E_{\mathrm{TX}_i}(x,y) < \min(E_{\mathrm{IWZX}_i}) \\ \frac{E_{\mathrm{TX}_i}(x,y)}{E_{\mathrm{TX}_i}(x,y)+E_{\mathrm{IWZX}_i}(x,y)}\mathrm{SI}_{\mathrm{IWZX}_i}(x,y) + \frac{E_{\mathrm{IWZX}_i}(x,y)}{E_{\mathrm{TX}_i}(x,y)+E_{\mathrm{IWZX}_i}(x,y)}\mathrm{SI}_{\mathrm{TX}_i}(x,y), & \text{otherwise} \end{cases}, \tag{13}$$

where $i = 2$ or 3 for X_2 and X_3, and the dependency on t has been removed to simplify the notation.

Typically, the temporal SI has the highest accuracy. Therefore, when the residual error of the temporal SI at coordinate (x,y) is lower than the minimum of the residual errors of the intra-SI, the temporal SI is considered as the final SI. Otherwise, the temporal and intra-SI are fused together using a linear combination to generate the final SI. When all the four subframes are decoded, they are merged by using a simple upsampling process to get the reconstructed frame $\hat{X}$. The decoding of Y proceeds similarly to the one of X. Finally, note that the correlation between subsampled frame images decreases as the temporal difference between frames increases. Therefore, this method performs best when the encoding mode of the subframes with the same index is alternated over time.

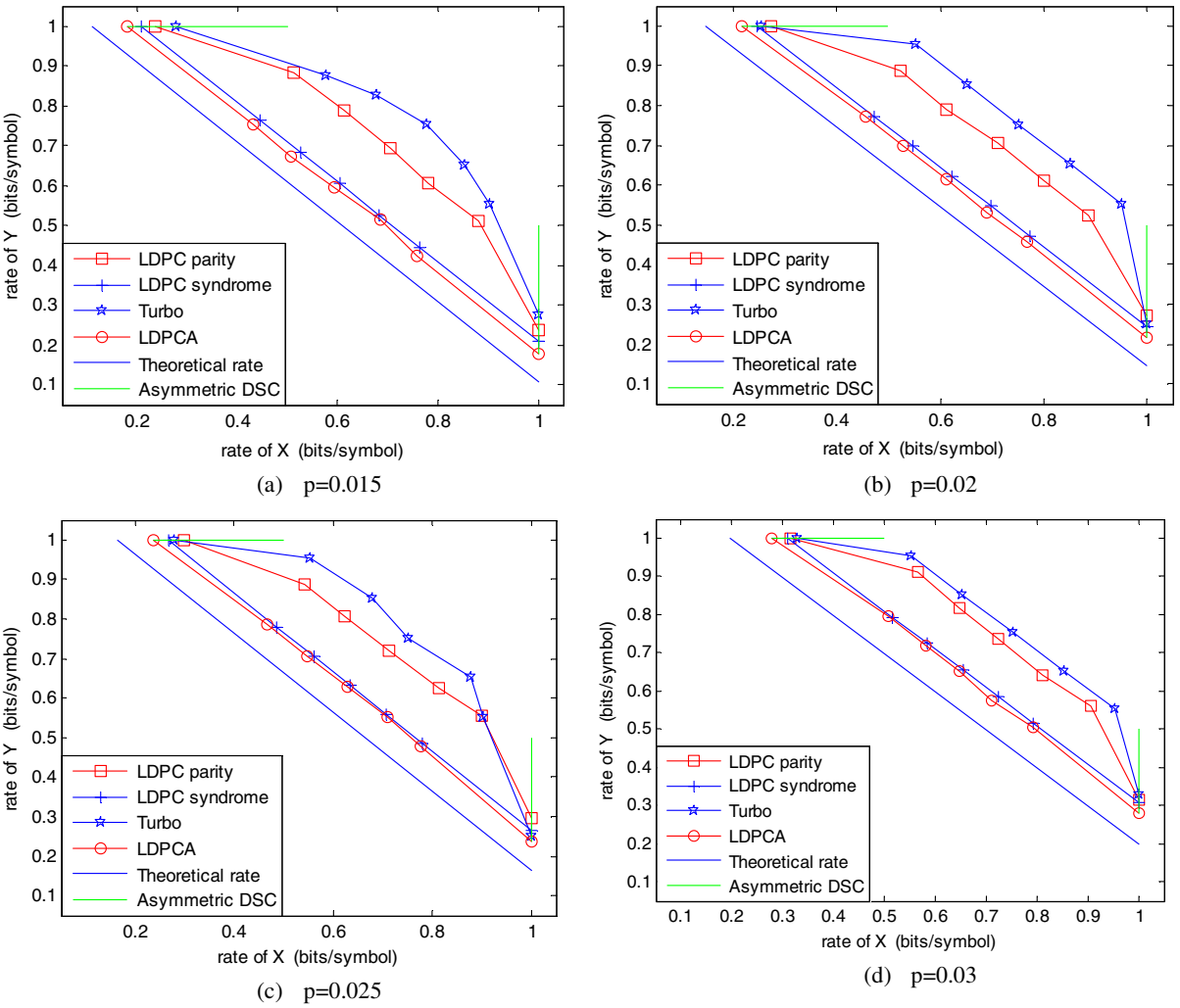
5 Simulation Results

5.1 *Performance of the Nonasymmetric DSC*

The performance of the nonasymmetric DSC scheme that uses the LDPCA channel decoding method is compared with the ones that use LDPC parity bits, LDPC syndrome, and Turbo. In the experiments, random sequences have been used as the two input sources X and Y. Their length is 6336 bits, the probability of 0 and 1 is 0.5, and the decoding error ratio has been assumed to be $\leq 10^{-5}$. The crossover probability p is used to indicate the correlation of the sources X and Y. In this work, the considered values of p have been 0.015, 0.02, 0.025, and 0.03. To quantify symmetry, we define the parameter s, which is derived from n and m, as $s = n/(n+m)$ for source X. The following values have been used as s: 0, 0.3, 0.4, 0.5, 0.6, 0.7, and 1. Each one corresponds to a distinct point on the line in Fig. 9. For every testing value, the final results are the average of 50 repetitions of the experiment.

Figure 9 shows the output rates of the nonasymmetric scheme investigated in this work when four different channel decoding methods are used, as well as the theoretical rate already shown as segment B-C in Fig. 2. Clearly, the output rates of the nonasymmetric schemes are much more flexible than the one of the asymmetric scheme, which would be fixed at 1 bit/symbol. However, the performance still presents a certain gap compared with the asymmetric system, especially when the correlations of the two sources are low (i.e., high crossover probability) and the performance of the channel code is not sufficiently good. In other words, in this case, the information bits transmitted directly without channel encoding in the nonasymmetric scheme cannot decrease the amount of channel coded information. Therefore, the total amount of information to transmit is higher and the performance of the nonasymmetric scheme would consequently decline.

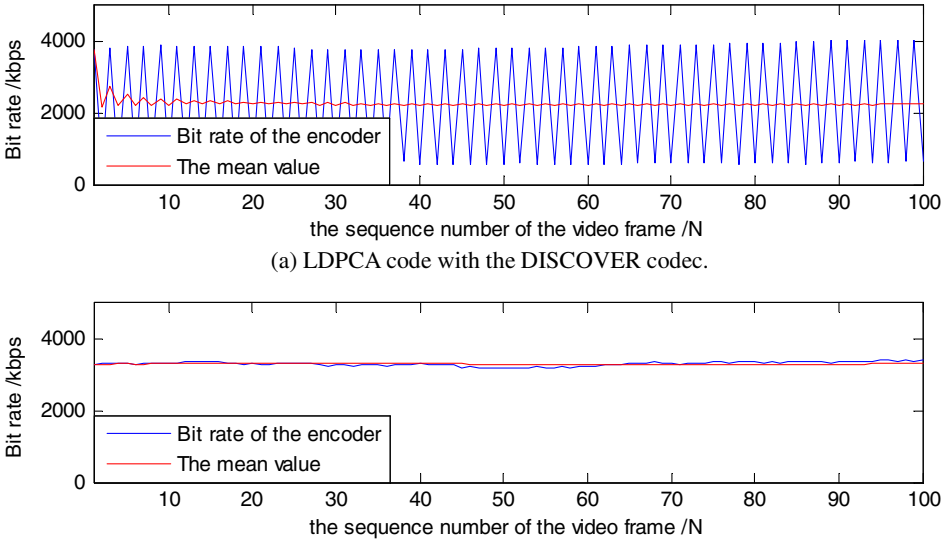
The performance of the system using the LDPC code based on syndromes is much better than that of the system using the LDPC or Turbo code based on parity bits. As



(a) p=0.015 (b) p=0.02

(c) p=0.025 (d) p=0.03

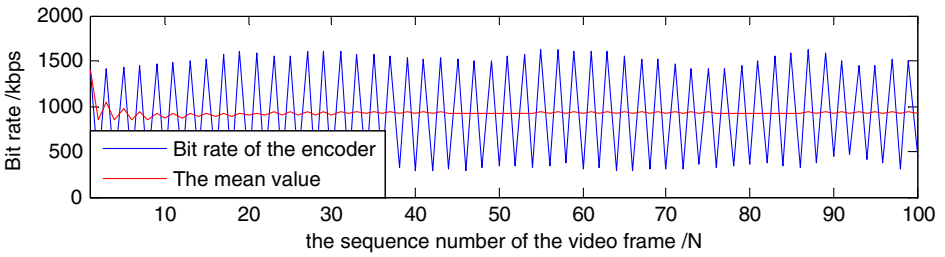
Fig. 9 Rate region for different DSC schemes.



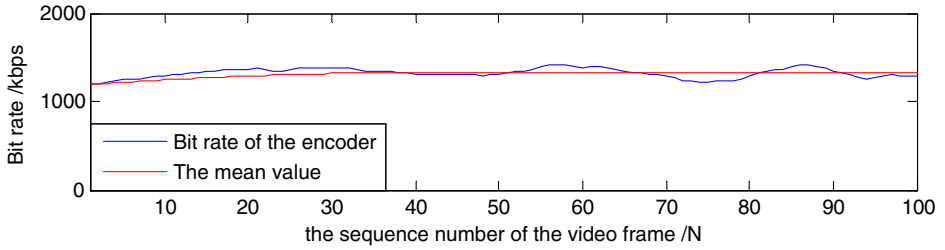
(a) LDPCA code with the DISCOVER codec.

(b) Proposed symmetric WZVC scheme.

Fig. 10 Encoding rate for the mobile sequence.



(a) LDPCA code with the DISCOVER codec.



(b) Proposed symmetric WZVC scheme.

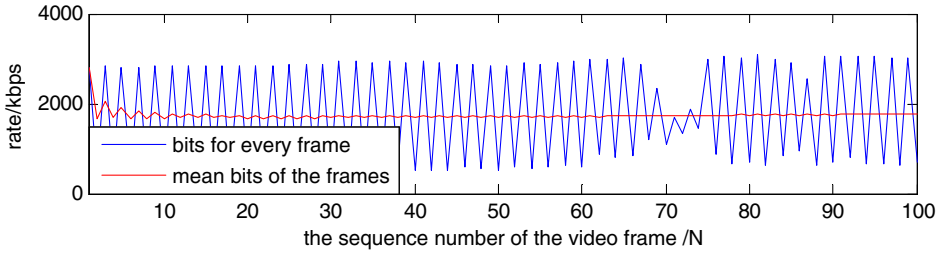
Fig. 11 Encoding rate for the foreman sequence.

shown in Fig. 9, the nonasymmetric DSC using LDPCA presents the best performance. Hence, the LDPCA has been selected to be used in the proposed symmetric WZVC scheme.

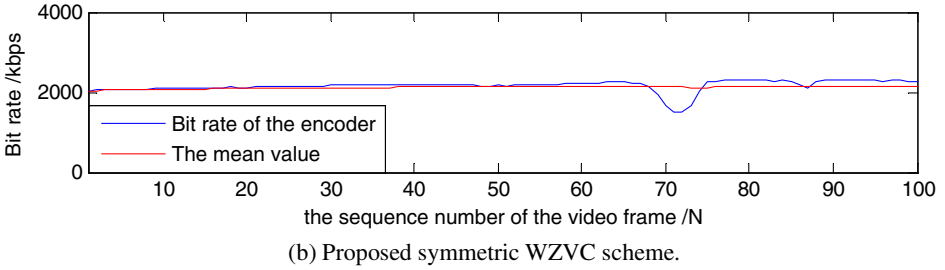
5.2 *Performance of the Proposed Symmetric WZVC*

This section demonstrates the effectiveness of the proposed symmetric WZVC by simulation results and shows a comparison with the exiting typical asymmetric DVC (DISCOVER codec[31]). Four low-motion video sequences with different motion characteristics have been used: mobile, foreman, coastguard, and container. Within the first 100 frames, the mobile sequence shows complicated movements, including vertical movement of the wall calendar, a ball rolling, and the toy train moving. For the foreman sequence, movements are mainly due to the foreman's head and the camera movements, while most of the background is still. In the coastguard sequence, movements are relatively drastic, including the yacht and ship's fast movement, also the moving water and background. As for the container sequence, movements are mainly due to the steamship, the yacht, and the flag, while the background does not present objects in movement. All of the four video sequences are in CIF format (352×288), 100 frames with a GOP size equal to 2 when coded with the DISCOVER codec. The temporal resolution is 15 frames per second and only the luminance component of the video sequences has been encoded.

Figures 10 to 13 show the simulation results when the proposed symmetric WZVC and the DISCOVER codec present similar SI and reconstruction quality. Both the instantaneous encoder bit rate and the mean value are shown. The

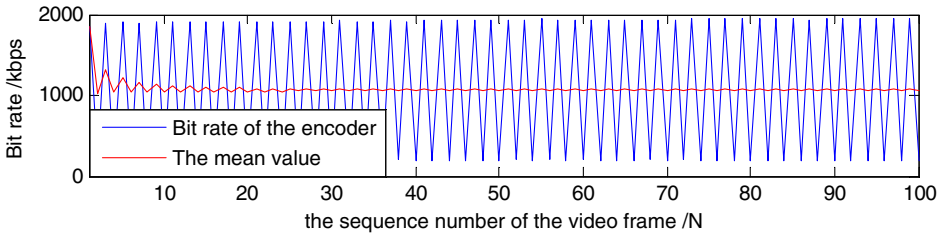


(a) LDPCA code with the DISCOVER codec.

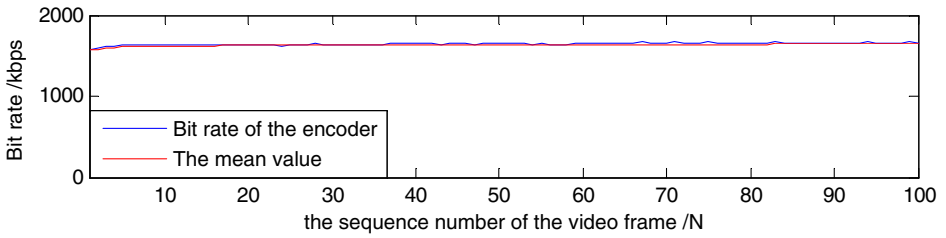


(b) Proposed symmetric WZVC scheme.

Fig. 12 Encoding rate for the coastguard sequence.

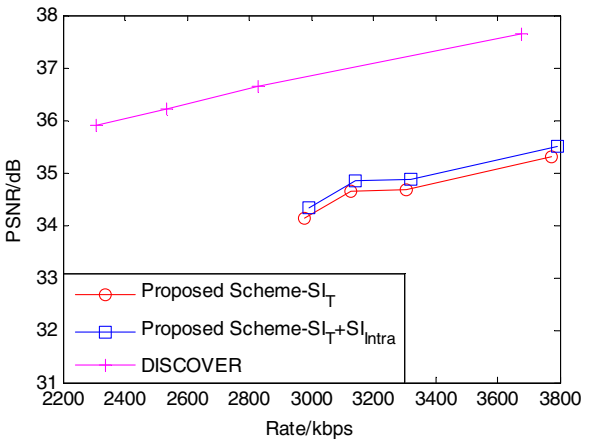


(a) LDPCA code with the DISCOVER codec.

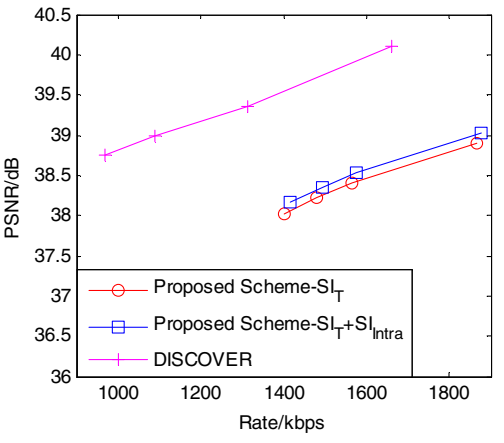


(b) Proposed symmetric WZVC scheme.

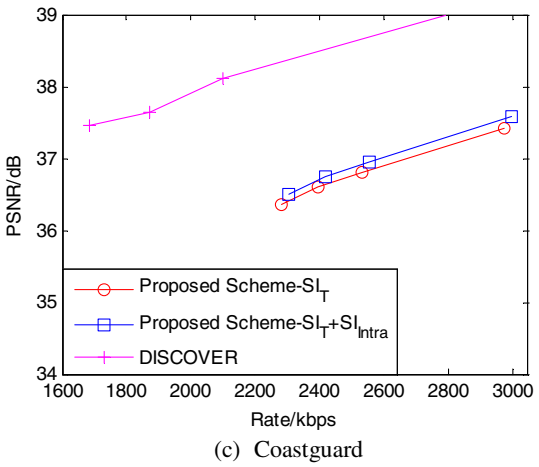
Fig. 13 Encoding rate for the container sequence.



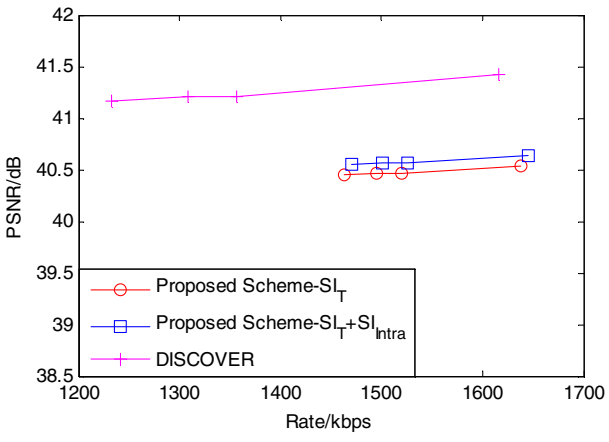
(a) Mobile



(b) Foreman



(c) Coastguard



(d) Container

Fig. 14 Rate-distortion performance comparison using peak signal-to-noise ratio. The proposed scheme shows the performance for two cases: with side information fusion ($SI_T + SI_{Intra}$) and with temporal side information only (SI_T).

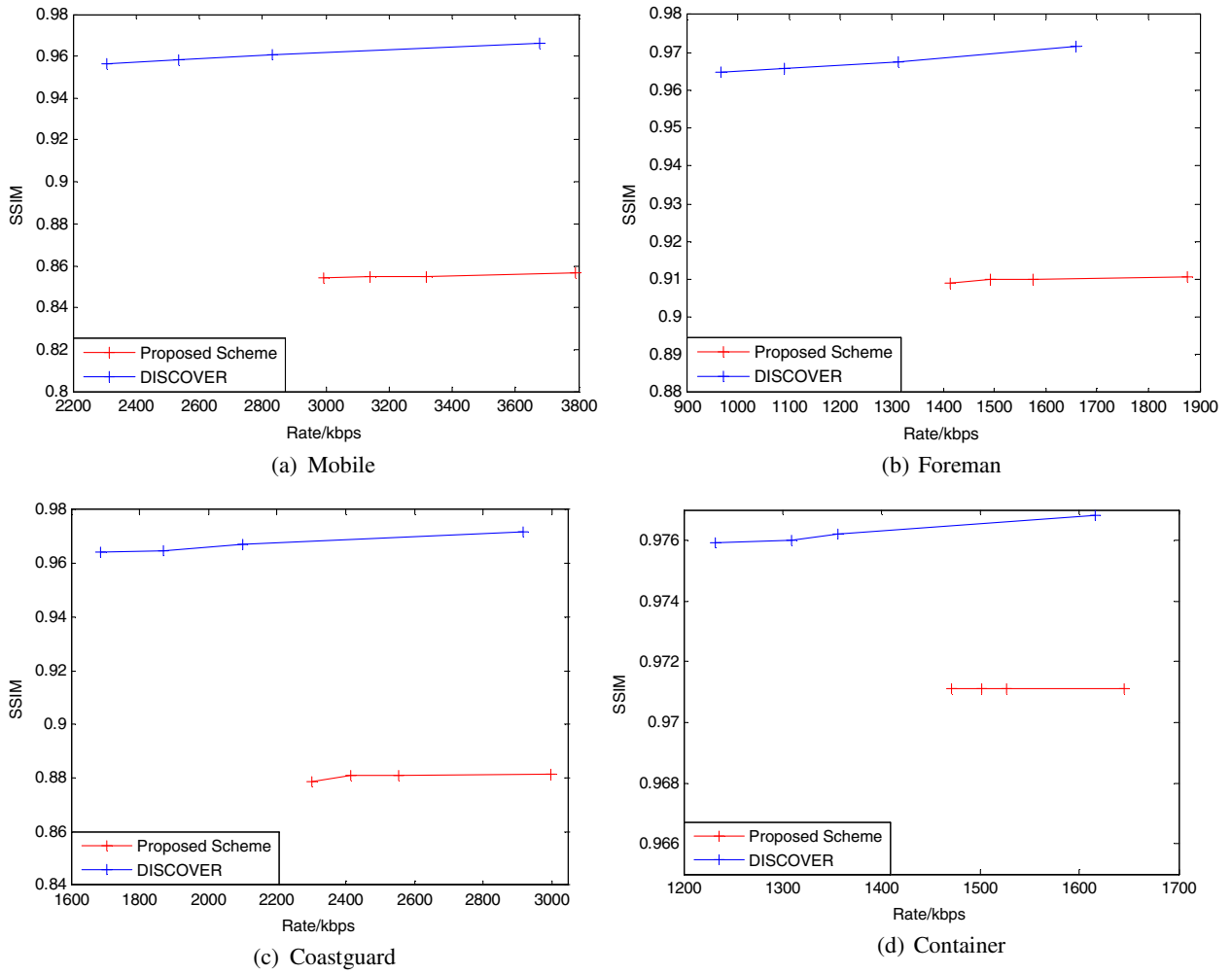


Fig. 15 Structural similarity (SSIM) performance comparison.

mean value is computed as the average since the first frame. If, for instance, the mean value needs to be computed at frame $N = 40$, the result will be the average of the instantaneous bit rates of the encoder over the first 40 frames. In the asymmetric DVC system (DISCOVER codec structure), the rates of Wyner–Ziv frames and Key frames are strongly different; thus, the output rate of the whole system varies dramatically over time, which might negatively affect the performance over a transmission channel. In the proposed symmetric WZVC system, instead, the output rates are very stable over time, potentially benefiting the channel usage efficiency. In terms of additional decoding complexity, experimental results show that the decoding process is only slightly more complicated than the one of the DISCOVER codec. Thus, the decoding time only slightly increases with respect to the DISCOVER case. Although smoothness could also be achieved by means of buffering, the proposed solution has the advantage that the additional encoding complexity is low. Moreover, the different coding schemes allow for parallel decoding under certain conditions, which is typically difficult to be applied to traditional schemes, such as the DISCOVER codec.

The following results will show how the achieved bit rate smoothness affects quality. First, the quality of the reconstructed video sequences is shown in Fig. 14 by means of rate-distortion graphs using the PSNR as a quality measure. Two sets of results are shown: using SI fusion ($SI_T + SI_{Intra}$) and using temporal SI only (SI_T). In the first case, the performance is better but decoding parallelism cannot be achieved, since intracoded subframes and the WZ-coded subframes cannot be decoded simultaneously. If only the temporal SI is used, the intracoded subframes and the WZ-coded subframes can be decoded simultaneously; therefore, parallelism can be achieved. However, the performance penalty for this case with respect to the previous one is limited, ranging from 0.1 to 0.2 dB peak signal-to-noise ratio (PSNR). Therefore, it is possible to choose to either use the SI fusion or not depending on the scenario without significantly influencing the performance.

Comparing these results with the DISCOVER one, as expected, the much smoother output rate comes at a cost: the compression performance of the symmetric WZVC scheme is lower than that of the DISCOVER codec for the same video quality. This can be attributed to the fact that the subsampling process reduces the quality of the SI at the decoder since subframes have obviously less high-frequency content (i.e., image information and details are lost) compared to the original frames. As it can be surmised from the figure, the gap depends on the characteristics of the video sequence. For instance, for limited motion sequences, such

as container, the gap is small, about 0.6 dB PSNR. The gap increases, as expected, for higher-motion sequence, such as mobile, up to 2 dB PSNR.

Moreover, in Fig. 15, the structural similarity quality metric[32] is shown to give a better understanding of the perceptual quality of the tested video sequences. Clearly, the quality gap varies depending on the video content, ranging from <0.01 for the container sequence up to 0.1 for the case of the mobile sequence, i.e., the most difficult one to encode. Also, with the proposed method, the lowest quality is ∼0.85, again for the case of the mobile sequence, notoriously difficult to encode using WZVC schemes. However, for scenarios in which more static content is present, the quality can achieve values as high as 0.97, as for the case of the container sequence.

Finally, note that we compare the performance with the DISCOVER codec configured to use a GOP size equal to 2. This seems to be fair since in the proposed system half of the data is coded using H.264 intra and half with WZ, as for the DISCOVER codec. If reference schemes have a larger GOP size, Ref. 26 suggests that a subsampling scheme can present even better performance.

6 Conclusion

In this paper, we investigated the performance of a nonasymmetric DSC structure when different channel decoding methods are used. Simulation results showed significant performance gain when the LDPCA channel decoding method based on source splitting is used, compared to other channel decoding methods. Using this DSC scheme as a basis, a symmetric WZVC scheme was proposed. In this scheme, an SI fusion algorithm is designed to exploit to the maximum extent both the intraframe SI and the temporal SI to maximize the reconstruction quality. Simulation results showed the main advantage of the proposed scheme, i.e., that the output rate is very stable, which is an important feature for some types of transmission channels. Moreover, the scheme allows parallel decoding implementations. Concerning the compression performance, there is still a gap between our proposed symmetric WZVC and the typical asymmetric DVC system, which is the cost for the much smoother output rate. Further work is under way to improve the performance of the symmetric WZVC system as a whole, for instance, by taking advantage of pixels coming from adjacent blocks in the computation of SI at the receiver and by using a more effective SI generating algorithm to improve the overall system performance.

Acknowledgments

This work was supported by Doctor Station Foundation of National Ministry of Education of China (20110181120009), the National Natural Science Foundation of China (61201388), and the EU Seventh Framework Programme FP7-PEOPLE-IRSES (247083).

References

1. A. Tamhankar and K. R. Rao, "An overview of H.264/MPEG-4 part 10," in *Proc. of 4th EURASIP Conf. on Video/Image Processing and Multimedia Communications*, pp. 1–51, European Association for Signal Processing (EURASIP), Zagreb, Croatia (2003).
2. T. Wiegand, G. J. Sullivan, and G. Bjøntegaard, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.* **13**(7), 560–576 (2003).
3. J. Ostermann et al., "Video coding with H.264/AVC: tools, performance, and complexity," *IEEE Circuits Syst. Mag.* **4**(1), 7–28 (2004).
4. R. Puri and K. Ramchandran, "PRISM: a new robust video coding architecture based on distributed compression principles," presented at *Proc. of the Allerton Conf. on Communication, Control, and Computing*, Urbana-Champaign, IL (2002).
5. A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. of the 36th Asilomar Conf. on Signals, Systems and Computers*, Vol. 1, pp. 240–244 (2002).
6. D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory* **19**(4), 471–480 (1973).
7. A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory* **22**(1), 1–10 (1976).
8. S. S. Pradhan and K. Ramchandran, "Distributed source coding: symmetric rates and applications to sensor networks," in *Proc. of the Data Compression Conf.*, Snowbird, UT, pp. 363–372, IEEE, Piscataway, NJ (2000).
9. S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Trans. Inf. Theory* **49**(3), 626–643 (2003).
10. M. Sartipi and F. Fekri, "Distributed source coding in wireless sensor networks using LDPC codes: the entire Slepian-Wolf rate region," in *Proc. of the IEEE Wireless Communications and Networking Conf.*, New Orleans, LA, Vol. 4, pp. 1939–1944, IEEE, Piscataway, NJ (2005).
11. J. Garcia-Frias and F. Cabarcas, "Approaching the Slepian-Wolf boundary using practical channel codes," *Signal Process.* **86**(11), 3096–3101 (2006).
12. D. Schonberg, K. Ramchandran, and S. S. Pradhan, "Distributed code constructions for the entire Slepian-Wolf rate region for arbitrarily correlated sources," in *Proc. of the Data Compression Conf.*, Snowbird, UT, pp. 292–301, IEEE, Piscataway, NJ (2004).
13. V. Toto-Zarasoa, A. Roumy, and C. Guillemot, "Rate-adaptive codes for the entire Slepian-Wolf region and arbitrarily correlated sources," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Las Vegas, NV, pp. 2965–2968, IEEE, Piscataway, NJ (2008).
14. V. K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Process. Mag.* **18**(5), 74–93 (2001).
15. A. Jagmohan and N. Ahuja, "Wyner-Ziv encoded predictive multiple descriptions," in *Proc. of Data Compression Conf.*, Snowbird, UT, pp. 213–222, IEEE, Piscataway, NJ (2003).
16. J. Wang et al., "Multiple descriptions in the Wyner-Ziv setting," in *Proc. of IEEE Int. Symp. on Information Theory*, Seattle, WA, pp. 1584–1588, IEEE, Piscataway, NJ (2006).
17. Y. Fan et al., "A novel multiple description video codec based on Slepian-Wolf coding," in *Proc. of Data Compression Conf.*, Snowbird, UT, p. 515, IEEE, Piscataway, NJ (2008).
18. O. Crave, C. Guillemot, and B. Pesquet-Popescu, "Multiple description source coding with side information," presented at *Proc. of 16th European Signal Processing Conf.*, Lausanne, Switzerland (August 2008).
19. H. Ma et al., "Distributed video coding based on multiple description," in *Proc. of 5th Information Assurance and Security*, Xian, China, pp. 294–297, IEEE, Piscataway, NJ (2009).
20. A. Wang, Y. Zhao, and H. Bai, "Robust multiple description distributed video coding using optimized zero-padding," *Sci. China F: Inf. Sci.* **52**(2), 206–214 (2009).
21. T. Maugey et al., "Using an exponential power model for Wyner Ziv video coding," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Dallas, TX, pp. 2338–2341, IEEE, Piscataway, NJ (2010).
22. A. Tomé and F. Pereira, "Low delay distributed video coding with refined side information," *Signal Process.: Image Commun.* **26**(4–5), 220–235 (2011).
23. A. Abou-Elailah et al., "Fusion of global and local motion estimation for distributed video coding," *IEEE Trans. Circuits Syst. Video Technol.* **23**(1), 158–172 (2013).
24. X. HoangVan, J. Ascenso, and F. Pereira, "Improving scalable video coding performance with decoder side information," in *Proc. of Picture Coding Symp.*, San Jose, CA, pp. 229–232, IEEE, Piscataway, NJ (2013).
25. S. Milani and G. Calvagno, "Distributed video coding based on lossy syndromes generated in hybrid pixel/transform domain," *Signal Process.: Image Commun.* **28**(6), 553–568 (2013).
26. Y.-C. Shen, J.-C. Luo, and J.-L. Wu, "Subsampling input based side information creation in Wyner-Ziv video coding," in *Proc. of Data Compression Conf.*, Snowbird, UT, p. 519, IEEE, Piscataway, NJ (2013).
27. D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Process.* **86**(11), 3123–3130 (2006).
28. X. Ou et al., "Multiple descriptions Wyner-Ziv video coding," *Optik* **125**(5), 1650–1656 (2014).
29. J. Ascenso and F. Pereira, "Advanced side information creation techniques and framework for Wyner-Ziv video coding," *J. Vis. Commun. Image Represent.* **19**(8), 600–613 (2008).

30. T. Maugey et al., "Fusion schemes for multiview distributed video coding," presented at *Proc. of 17th European Signal Processing Conf.*, Glasgow, Scotland (24–28 August 2009).
31. X. Artigas, J. Ascenso, and M. Dalai, "The DISCOVER codec: architecture, techniques and evaluation," presented at *Proc. of Picture Coding Symp.*, Lisbon (November 2007).
32. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).

Xianfeng Ou received his BS degree in electronic information science and technology and MS degree in communication and information system from Xinjiang University, Xinjiang, China, in 2006 and 2009, respectively. He is currently working toward his PhD degree in the College of Electronics and Information Engineering at Sichuan University, Chengdu, China. He was a visiting researcher at the Internet Media Group, Politecnico di Torino, Turin, Italy, from January to April 2014, working on distributed video coding and transmission. His main research interests include image and video coding technologies.

Enrico Masala received his MSc degree (summa cum laude) and his PhD degree in computer engineering from the Politecnico di Torino, Turin, Italy, in 1999 and 2004, respectively. In 2003, he was a visiting researcher at the Signal Compression Laboratory, University of California, Santa Barbara, California, USA, where he worked on joint source channel coding algorithms for video transmission. He is currently an assistant professor in the Control and Computer Engineering Department at the Politecnico di Torino. He coauthored over 40 scientific publications in multimedia communications-related topics. His main research interests include processing, coding, and robust transmission of multimedia signals (especially video) over packet networks.

Linbo Qing received his PhD degree in communication and information systems from Sichuan University, Chengdu, China, in 2008. From February to August 2011, he was a visiting researcher at the Agder Mobility Laboratory, University of Agder, Norway, where he worked on data aggregation and security in mobile communications. From September 2011 to February 2012, he was a visiting researcher at the Internet Media Group, Politecnico di Torino, Turin, Italy, where he worked on distributed video coding and transmission. He is currently an associate professor in the College of Electronics and Information Engineering at Sichuan University. He coauthored over 30 scientific publications in image processing related topics. His main research interests include image processing, video coding and transmission, and information theory.

Xiaohai He received his PhD degree in biomedical engineering from Sichuan University, Chengdu, China, in 2002. He is a currently a professor in the College of Electronics and Information Engineering at Sichuan University. He is an editor of both the *Journal of Information and Electronic Engineering* and the *Journal of Data Acquisition & Processing*. His research interests include image processing, pattern recognition, and image communication. He is a senior member of the Chinese Institute of Electronics.