AUTHOR(S): HU, C., HU, B., TU, W. and XIONG, Y.

TITLE: A low-complexity and efficient encoder rate control solution for distributed residual video coding.

YEAR: 2017

Publisher citation: HU, C., HU, B., TU, W. and XIONG, Y. 2018. A low-complexity and efficient encoder rate control solution for distributed residual video coding. Multimedia tools and applications [online], 77(5), pages 5713-5735. Available from: https://doi.org/10.1007/s11042-017-4484-5.

OpenAIR citation: HU, C., HU, B., TU, W. and XIONG, Y. 2018. A low-complexity and efficient encoder rate control solution for distributed residual video coding. Multimedia tools and applications, 77(5), pages 5713-5735. Held on OpenAIR [online]. Available from: https://openair.rgu.ac.uk.

JrnlID 11042_ArtID 4484_Proof#1 - 09/02/2017

Multimed Tools Appl
DOI 10.1007/s11042-017-4484-5

A low-complexity and efficient encoder rate control solution for distributed residual video coding

Chunyun Hu[1,2] · Binjie Hu[2] · Wanqing Tu[3] · Yunhui Xiong[4]

Received: 28 July 2016 / Revised: 25 December 2016 / Accepted: 6 February 2017

Abstract Existing encoder rate control (ERC) solutions have two technical limitations that prevent them from being widely used in real-world applications. One is that encoder side information (ESI) is required to be generated which increases the complexity at the encoder. The other is that rate estimation is performed at bit plane level which incurs computation overheads and latency when many bit planes exist. To achieve a low-complexity encoder, we propose a new ERC solution that combines an efficient encoder block mode decision (EBMD) for the distributed residual video coding (DRVC). The main contributions of this paper are as follows: 1) ESI is not required as our ERC is based on the analysis of the statistical characteristics of the decoder side information (DSI); 2) a simple EBMD is introduced which only employs the values of residual pixels at the encoder to classify blocks into Intra mode, Skip mode, and WZ mode; 3) an ERC solution using pseudo-random sequence scrambling is proposed to estimate rates for all WZ blocks at frame level instead of at bit plane level, i.e., only one rate is estimated; and 4) a quantization-index estimation

✉ Yunhui Xiong
yhxiong@scut.edu.cn

Chunyun Hu
hcy2182@scau.edu.cn

Binjie Hu
eebjiehu@scut.edu.cn

Wanqing Tu
w.tu@rgu.ac.uk

1 College of Electronic and Engineering, South China Agricultural University, Guangzhou, China

2 School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China

3 School of Computing Science and Digital Media, Robert Gordon University, Aberdeen, UK

4 School of Mathematics, South China University of Technology, Guangzhou, China

algorithm (QIEA) is proposed to solve the problem of rate underestimation. The simulation results show that the proposed solution is not only low complex but also efficient in both the block mode decision and the rate estimation. Also, as compared to DISCOVER system and the state-of-the-art ERC solution, our solution demonstrates a competitive rate-distortion(RD)performance. Due to maintain the low-complexity nature of the encoder and have good RD performance, we believe that our ERC solution is promising in practice.

Keywords Distributed residual video coding (DRVC) · Encoder rate control (ERC) · Encoder block mode decision (EBMD) · Low-complexity encoder · Pseudo-random sequence scrambling

1 Introduction

With the wide deployment of wireless networks and the technical advances in micro-electronics, there is a growing number of new video applications, such as wireless low-power video surveillance and video sensor networks, becoming popular. The traditional joint video encoding paradigms (e.g., H.264/AVC and MPEG-4) which put a significant burden on the encoder mainly due to the complex motion estimation techniques do not suit for the new applications because the encoders (normally sensor devices) are limited in power, memory and computational capabilities. The new applications could benefit from a codec with a low complexity encoder.Therefore, in the past decade, a coding paradigm called distributed video coding (DVC) which is famous for a low complexity encoder coupled with a more complex decoder has gained the attention of the scientific community.

The theoretical foundations of DVC are the SlepianCWolf [22] theory which is about the lossless distributed coding and the Wyner-Ziv [27] theory which is about the loss distributed coding. These theories suggest that the statistical redundancies in a (video) signal can be exploited at the decoder side with only a limited performance loss as compared to a system employing redundancies at the encoder. Under this suggestion, the motion-compensated prediction in DVC is shifted from the encoder to the decoder that facilitates the design of a simple encoder coupled with a complex decoder.The well-known DVC architectures have been developed by researchers in Stanford University, mainly including pixel-domain DVC (PDDVC) [1], transform-domain DVC (TDDVC) [2] and distributed residual video coding (DRVC) [3]. Our study in this paper will focus on DRVC.

During the past decade, the research hotspots in DVC are focused on improving coding efficiency, decreasing system latency, removing feedback channel, and maintaining error resilience. In order to improve the coding efficiency, well-known strategies such as side information refinement [12, 29], more accurate correlation noise model [24], more effective reconstruction [25, 30] and block mode decisions [4, 7, 8, 11, 13, 23, 26, 28] have been presented. In order to decrease the latency, low-delay DVC systems based on motion-compensated extrapolation [17, 21] and DVC systems using entropy coding without iterative channel codes have been presented [18, 19]. In order to remove the feedback channel, encoder rate control(ERC) solutions [5, 6, 9, 15, 16, 20] have been proposed. In order to maintain error resilience [14], multiple-description coding has been proposed to overcome transmission errors in video communications over error-prone networks.

This paper studies ERC problem without using a feedback channel (FC). In most existing DVC systems, FC is expected to allocate a proper bit rate for a certain target quality that is called FC-driven rate allocation or decoder rate control (DRC). However, FC is not only unavailable in many video applications but also results in additional latency and increasing decoding complexity due to several feedback-decoding iterations. To overcome these drawbacks, scholars have proposed ERC solutions without FC which however bear two limitations preventing them from being widely used in real-world applications. One is that the generation of encoder side information (ESI) increases the encoder complexity. Since efficient rate allocation relies on the quality of side information (SI) and SI is not available at the encoder, ESI is always required to be generated. The other limitation is that the rate estimation at the bit plane level causes computational complexity and latency when there are many bit planes.

Since the encoder has limited capability, the main objective of this paper is to find an ERC solution which can maintain the low-complexity nature of the encoder and therefore have good practical use. As compared to the existing ERC solutions, our solution presents four advantages.

- Our ERC solution does not need to generate ESI which benefits the encoder with low complexity. After the analysis of the side information at the decoder, we derive an assumption that the quantization version of the decoder residual frame is full of 0. Under this assumption, the correlation between the residual frames at the encoder and decoder is simply calculated by (8) which has nothing to do with ESI.
- A simple encoder block mode decision (EBMD) is introduced to improve the coding efficiency. Our EBMD only employs the values of residual pixels at the encoder to classify blocks into Intra mode, Skip mode, and WZ mode without any considerable computation.
- Our ERC solution is proposed to estimate the transmitting rate for all WZ blocks at frame level instead of at bit plane level, i.e., only one rate is estimated. The ERC solution is based on pseudo-random sequence scrambling which is used to scramble the residual pixels in WZ blocks at both the encoder and the decoder. When the residual pixels are scrambled, the error probabilities between the codewords at both the encoder and the decoder become homogeneous. So the transmitting rate can be the same.
- A quantization-index estimation algorithm (QIEA) is presented to solve the problem of rate underestimation. After inverse pseudo-random sequence scrambling, the failed decoded quantization indexes will be scattered among the decoded ones which are used to predict the former and then solve the problem of underestimation.

This paper is organized as follows. Section 2 introduces the related studies on ERC and EBMD. Section 3 presents our ERC solution in detail. In Section 4, experimental results are demonstrated and discussed. Finally, we conclude the paper in Section 5.

2 Related work

2.1 Recent work on ERC solutions

As it is known, there are two challenges in ERC solutions. One is to estimate the accurate statistical correlation between the source information and the side information. The other is to allocate the proper bit rate for each bit plane since most of the existing ERC solutions are implemented at bit plane level. Both the challenges are related to SI. As SI is not available at the encoder, ESI is always required to be generated. In [15, 16], the ERC technique estimates the rate based on the lookup tables obtained through a training stage. The table-based rate estimations are not dynamic and their efficiency strongly depends on the training video sequences. In [9], a fast block matching algorithm is used to generate ESI and the rate is estimated depending on the current bit plane error probability and the conditional entropy. The results show there is a small gap in RD performance when compared to the corresponding DRC scenario. In [20], ESI is generated by selecting the block among three candidates. A block with the minimum summation of the absolute difference (SAD)is selected as the block in ESI. The estimated rate is a linear model of the theory-bound rate. The experimental results indicate the performance of the proposed ERC is close to but still lower than that of DRC peer. In [5], Each frame is divided into two sub-frames: a key frame and a WZ frame. ESI is generated by taking the average of neighbor pixels in key frames. The RD performance is also lower than that of DRC solutions. In [6], the author proposed an efficient ERC solution which can be taken as a new benchmark. In the work, ESI is generated by a fast motion compensated interpolation and the parity bits for each bit plane is estimated by taking the inter bit plane correlation and the probability of errors into consideration. At the decoder, along with the correlation noise model updating technique and a novel soft reconstruction, a weighted overlapped block motion compensation technique is proposed to refine the side information. The experiments show that the ERC solution provides a promising result which equals to the RD performance of DISCOVER system. Even though the ERC in [6] is very powerful and efficient, its outperformance is at the cost of increasing the complexity at both the encoder and the decoder.

In a nutshell, ERC solutions always increase computational complexity and latency at the encoder due to the generation of ESI and the repeated rate estimation at bit planes. The RD performance of ERC scheme is always lower than that of the corresponding DRC scheme.

2.2 Recent work on ERMD

Block mode decision is a useful method to improve the coding efficiency. Several EBMD algorithms have been presented in DVC literatures. Intra mode and WZ mode are often introduced. In literature [7, 11] , the mode selection depends on the SAD between a block and its collocated block in the previous frame as an indication of the temporal coherence. If SAD is less than a certain threshold, WZ mode is chosen; otherwise, Intra mode is chosen. In [23], both spatial and temporal block coherence are taken into account by calculating the pixel variance of each block and the SAD, respectively. In [8], a DVC codec with three coding modes is presented: Intra, Inter, and WZ. At the encoder, a bit plane motion estimation (ME) algorithm is carried out and the ME residual error is used to select the coding mode for each block. In [13], an iterative algorithm is proposed to dynamically select either Intra mode or WZ mode for a DCT block. In order to make more accurate mode decision, ESI is required to be generated. In [4, 26], The block mode decision depends on a RD cost function which is composed of compression rate and distortion. The coding mode with the minimum cost is chosen for each block. In addition to the Intra and WZ mode, skip mode used in [28] is also introduced which can save the transmission data and therefore improve the RD performance.

Although the above EBMD algorithms can improve the coding efficiency, they undoubtedly incur computational complexity at the encoder due to the calculation of metrics such as SAD, compression rate, distortion function, etc. Furthermore, if there are some thresholds which should be pre-defined, the users usually have no clue to set them.

In this paper, DRVC system is studied and an ERC solution combining with an EBMD for DRVC (ERC-EBMD-DRVC) is proposed which maintains a low complexity encoder and hence should be promising in practice.

3 Proposed ERC-EBMD-DRVC solution

Table 1 lists the major symbols used in the paper. DRVC is a video coding architecture developed by Stanford University. Figure 1 illustrates the ERC-EBMD-DRVC architecture proposed in this paper. In the codec, a video sequence is divided into WZ frames (X_{2k}) and Key frames (X_{2k+1}) by setting $GOP = 2$. Key frames are encoded and decoded by H.264/AVC Intra. Once a past and a future Key frame are decoded, the reference frame (X_{re}) and the side information(Y_{2k}) for an intermediate WZ frame are generated. Then the residual frames (R at the encoder and R' at the decoder) are generated.

Table 1 The major symbols

X_{2k}	WZ frame
X_{2k-1}	past Key frame
X_{2k+1}	future Key frame
X_{re}	reference frame
$\hat{X}_{2k-1}$	decoded past Key frame
$\hat{X}_{2k+1}$	decoded future Key frame
R	residual frame at the encoder
R'	residual frame at the decoder
Y_{2k}	side information for X_{2k}
$\hat{R}$	reconstruction of R
$\hat{X}_{2k}$	reconstruction of X_{2k}
R'_q	quantization index of R'
R_{block}	the macro block in R
R'_{block}	the macro block in R'
p_i	the residual pixel in R_{block}
p'_i	the residual pixel in R'_{block}
R_{wz-f}	the frame composed of all WZ blocks in R
R'_{wz-f}	the frame composed of all WZ blocks in R'
S	a sequence composed of all the residual pixels in R_{wz-f}
S'	a sequence composed of all the residual pixels in R'_{wz-f}
$C_{wz,k}$	the kth codeword at the encoder
L	the code length of LDPC
$C'_{wz,k}$	the kth codeword at the decoder
$\rho_{wz,i}$	the error probability between $C_{wz,k}$ and $C'_{wz,k}$
ρ_{est}	the estimated error probability
ρ_{real}	the real error probability
$\rho^{bef}_{wz,i}$	the error probability calculated before scrambling
$\rho^{aft}_{wz,i}$	the error probability calculated after scrambling
ρ^{bef}	a set consisting of $\rho^{bef}_{wz,i}$
ρ^{aft}	a set consisting of $\rho^{aft}_{wz,i}$
v	the estimated rate
v'	the ideal rate
$\hat{R}^{dec}_q$	the decoded quantization index
$\hat{R}^{notdec}_q$	the failed decoded quantization index
R'_{q-esti}	the predicted value of $\hat{R}^{notdec}_q$

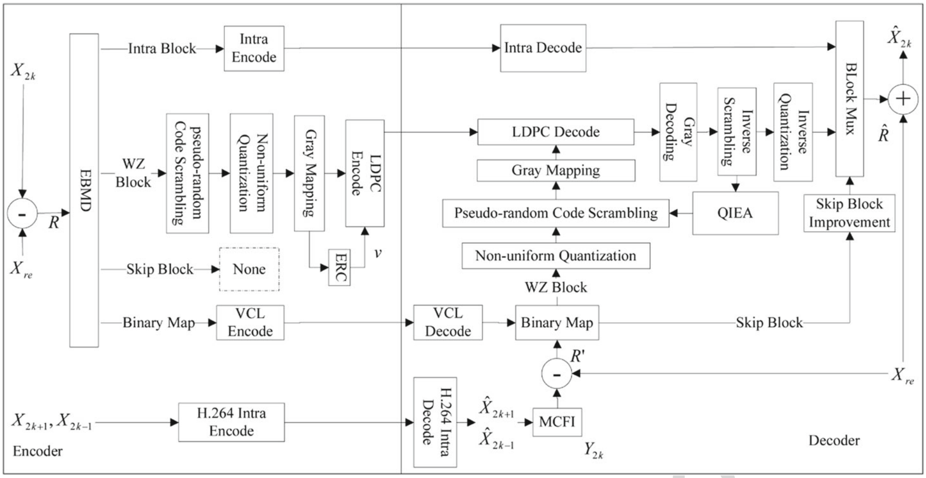

Fig. 1 System diagram of ERC-EBMD-DRVC

X_{re}, Y_{2k}, R, and R' are defined as (1), (2), (3), and (4), respectively.

$$X_{re} = (\hat{X}_{2k-1} + \hat{X}_{2k+1})/2\ , \tag{1}$$

$$Y_{2k} = \frac{1}{2}[\hat{X}_{2k-1}(x + mv_x, y + mv_y) + \hat{X}_{2k+1}(x - mv_x, y - mv_y)], \tag{2}$$

$$R = X_{2k} - X_{re}, \tag{3}$$

$$R' = Y_{2k} - X_{re}, \tag{4}$$

In formula (1), $\hat{X}_{2k-1}$ and $\hat{X}_{2k+1}$ are the decoded Key frames. In formula (2), $mv = (mv_x, mv_y)$ is the motion vector estimated by an algorithm called motion compensated frame interpolation (MCFI).

At the encoder, R is divided into non-overlapping 4×4 macro blocks. For each block the coding mode is determined by the EBMD module which support three modes: Intra, Skip, and WZ. For Intra blocks an approach similar to H.263+ Intra is used that the Intra blocks are transformed by a discrete cosine transform (DCT), scalar quantized and entropy coded. For Skip blocks they take no further part in the encoding process and are skipped without transmission. For WZ blocks, the encoder groups all of them into one frame in which the pixels are randomly scrambled, non-uniform quantized, and Gray encoded. Then codewords are fed to a LDPC coder and the amount of the parity bits transmitted to the decoder is estimated by the ERC module. Meanwhile, A binary mode decision map employing run-length coding is sent to the decoder.

At the decoder, according to the decoded mode decision map, the coding mode for each block in R' is the same as the coding mode for the co-located block in R. The former is called the side information block for the latter. Intra blocks are decoded by intra decoder. WZ blocks are decoded by correcting errors in their side information blocks using the received parity bits. If the parity bits are not enough to decode WZ blocks successfully, QIEA module is used to solve the problem of rate underestimation. The side information blocks marked as WZ mode are also randomly scrambled, non-uniform quantized,Gray encoded and then fed to a LDPC decoder. The decoded codewords are Gray decoded, inversely scrambled, and inversely quantized. Skip blocks are replaced by their side information blocks. If the quality of the side information blocks is not good enough, the decoder

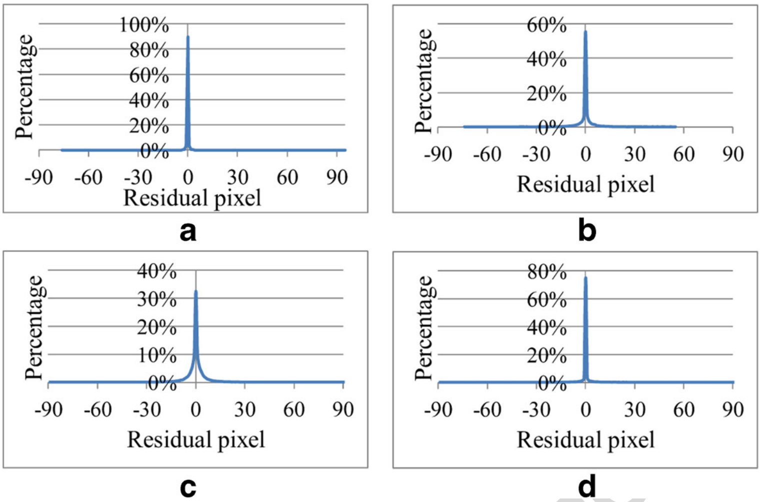

Fig. 2 Probability distribution of **a** the 13^{th} residual frame of Hall Monitor, **b** the 32^{th} residual frame of Foreman, **c** the 4^{th} residual frame of Coastguard, and **d** the 19^{th} residual frame of Soccer

can check and improve it. Finally, all decoded blocks are combined to form a decoded residual frame $\hat{R}$. Then a decoded WZ frame is obtained $\hat{X}_{2k} = \hat{R} + X_{re}$.

3.1 Analysis of R' and the quantization index R'_q

In DRVC system, R' is the side information for R. The probability distribution curves of residual pixels in any R' in Hall Monitor, Foreman, Coastguard, and Soccer videos are illustrated in Fig. 2. It shows that each curve is sharp near 0, meaning that the pixel values are centered at 0. By comparing (1), (2), and (4), we find that R' can be regarded as motion-compensated errors for a past and a future decoded Key frame. Since most background and foreground in a frame and its motion-compensated frame change a little information, the motion-compensated errors are very small, resulting in the case $R' = Y_{2k} - X_{re} \approx 0$ being in the majority.

Specific to the nonuniform distribution of R', nonuniform quantization is employed. Let the quantization intervals be $[-255, -31]$, $[-30, 30]$, and $[31, 255]$ where the threshold 30

Table 2 The percentage of R'_q

quantization interval	$[-255, -31]$	$[-30, 30]$	$[31, 255]$
quantization index R'_q	-1	0	1
Hall Monitor			
(the 13^{th} residual frame)	0.295928 %	99.45155 %	0.252525 %
Foreman			
(the 32^{th} residual frame)	0.323548 %	99.08854 %	0.58791 %
Coastguard			
(the 4^{th} residual frame)	0.591856 %	98.78078 %	0.627367 %
Soccer			
(the 19^{th} residual frame)	0.994318 %	97.16304 %	1.842645 %

is empirically obtained and let the corresponding quantization indexes (R'_q) be -1, 0, or 1. Table 2 gives the quantization results of the frames which are shown in Fig. 2. It is not hard to find out that since the case $R' = Y_{2k} - X_{re} \approx 0$ is in the majority, the residual pixels falling in the interval $[-30, 30]$ are up to 99 %. So, it can be assumed that $R'_q = 0$ is a 100 % case.

Here the statistic characteristic of R' can be summarized as follows. The values of the residual pixels concentrate near 0. After implementing nonuniform quantization, we can assume that $R'_q = 0$ accounts for 100 %.

3.2 Proposed encoder block mode decision

Based on the hypothesis that $R'_q = 0$ accounts for 100 %, a simple and efficient EBMD is proposed to classify each block in R into one of the three modes. The definitions and decision criteria are:

- The size of macro block is 4×4, totally 16 residual pixels.
- Let R_{block} and R'_{block} be the macro block in R and R' respectively. $PSNR(R_{block}, R'_{block})$ is defined as the peak signal to noise ratio (PSNR) of the macro block.
- Let p_i and p'_i be the residual pixel in R_{block} and R'_{block} respectively, where $p_i, p'_i \in [-255, 255], i = 1, 2, 3, ..., 16$.
- Intra block: It refers to the block whose correlation with the co-located side information block is weak. For Intra block, Intra codec is more effective than Wyner-Ziv codec. Given the specific hypothesis of $R'_q = 0$, the case $|R_q| = 1$ means that the correlation is weak. Therefore, a block with at least six p_i satisfying $|p_i| > 30$ is classified as an Intra block.
- Skip block: It refers to the block whose correlation with the co-located side information block is strong. It can be replaced by its side information block. Given the specific hypothesis of $R'_q = 0$, the case $R_q = 0$ means the correlation is strong. In order to obtain higher $PSNR(R_{block}, R'_{block})$, a block with all the p_i satisfying $|p_i| \leq 10$ is classified as a Skip block.
- WZ block: A block which is neither an Intra block nor a Skip block is classified as a WZ block.

The proposed EBMD is simple, only depending on the values of the residual pixels at the encoder without computing metrics such as SAD, compression rate, distortion function, etc. Figure 3a and b show the first residual frame of Foreman and the three kinds of blocks in it, respectively.

Our EBMD is based on the hypothesis that $R'_q = 0$ accounts for 100 %. However, this is not always true. In practice, when the occasional case $R'_q \neq 0$ occurs, it may result in wrong decisions. If R_{block} is wrongly classified as an Intra block, it will be reconstructed correctly by Intra decoding. If R_{block} is wrongly classified as a Skip block, it is unable to be replaced by the side information block. Because in this case, $R'_q = 0$ while $R'_q \neq 0$, the correlation of R_{block} and R'_{block} is not good enough. To solve this problem, (5) is used to check and improve the quality of R'_{block}. The improved R'_{block} satisfies that all the p'_i in R'_{block} are less than or equals to 10 that is consistent with the decision criterion for Skip blocks.

$$p'_i = \begin{cases} 10 & p'_i > 10 \\ -10 & p'_i < -10 \\ p'_i & |p'_i| \leq 10 \end{cases} \qquad p'_i \in R'_{block} \tag{5}$$

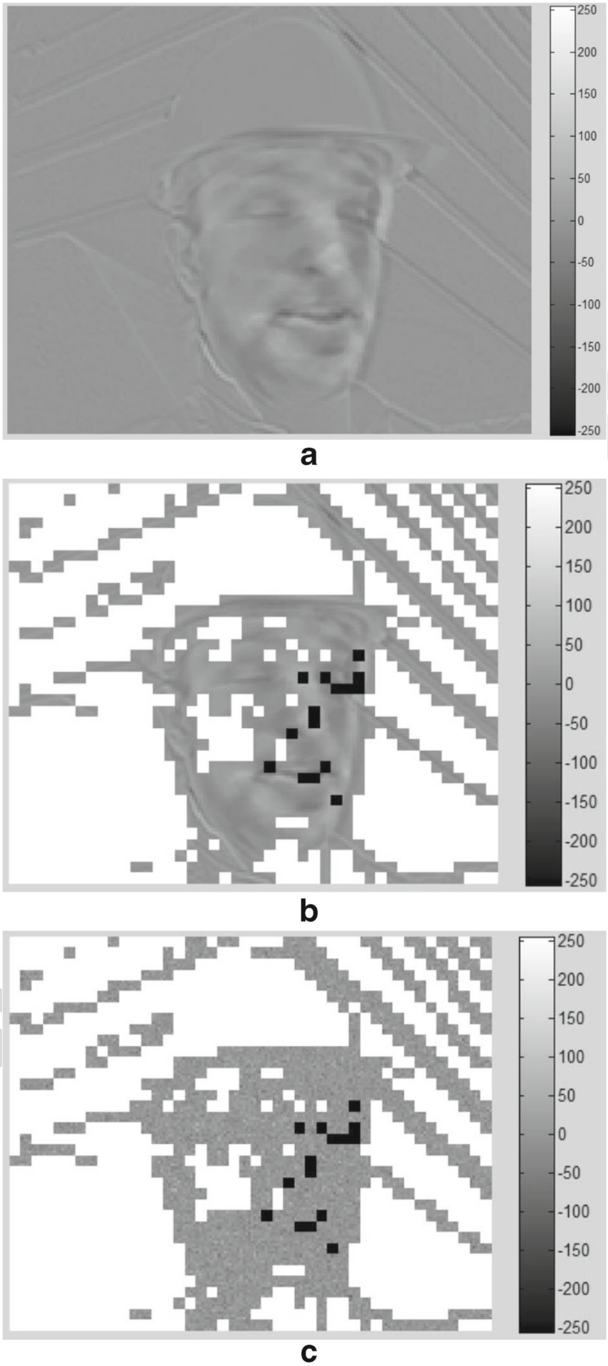

Fig. 3 Display of **a** the first residual frame of Foreman, **b** three kinds of blocks which use *white*, *black*, and *gray* to represent Skip blocks, Intra blocks, and WZ blocks respectively, **c** the residual frame with WZ blocks scrambled by pseudo-random sequence

3.3 Proposed encoder rate control based on pseudo-random sequence scrambling

For each residual frame R, the encoder groups all the WZ blocks into one frame called R_{wz-f}. In a similar way, the decoder groups all the side information blocks marked as WZ mode into another frame called R'_{wz-f} which is regarded as the SI for R_{wz-f}. The proposed ERC is used to estimate the rate for R_{wz-f} at the frame level, i.e., only one rate is estimated.

3.3.1 Pseudo-random sequence scrambling

Pseudo-random sequence is used to scramble the residual pixels in both R_{wz-f} and R'_{wz-f}. Figure 4 shows the process of scrambling that is assumed there are sixteen pixels in R_{wz-f}. The process is as follows. Firstly, the residual pixels in R_{wz-f} form a sequence S in column by column order. Secondly, a rand function generates a pseudo-random sequence S' with the same length of S. Then the random numbers in S' are sorted in ascending order and the corresponding index sequence is obtained. Finally, the pixels in S are sorted in the obtained index order and then a scrambled R_{wz-f} is achieved. Figure 3c shows the first residual frame in Foreman with the scrambled WZ blocks, i.e., the gray blocks are scrambled by pseudo-random sequence.

After scrambling, the differences between R_{wz-f} and R'_{wz-f} become homogeneous which can be testified by calculating the error probabilities between the codewords at both the encoder and the decoder. At the encoder, based on the LDPC codeword length L, the transmitted data are divided into k codewords. If the length of the last codeword is less than L, 0 is added. Let $C_{wz,1}$ $C_{wz,2}$ ... $C_{wz,k}$ be the k codewords at the encoder and let $C'_{wz,1}$ $C'_{wz,2}$ ... $C'_{wz,k}$ be the k codewords at the decoder. The error probability between the corresponding codewords at both the encoder and the decoder is calculated by (6)

$$\rho_{wz,i} = \sum(C_{wz,i} \oplus C'_{wz,i}) / L \ , \quad i = 1, 2 \cdots k, \tag{6}$$

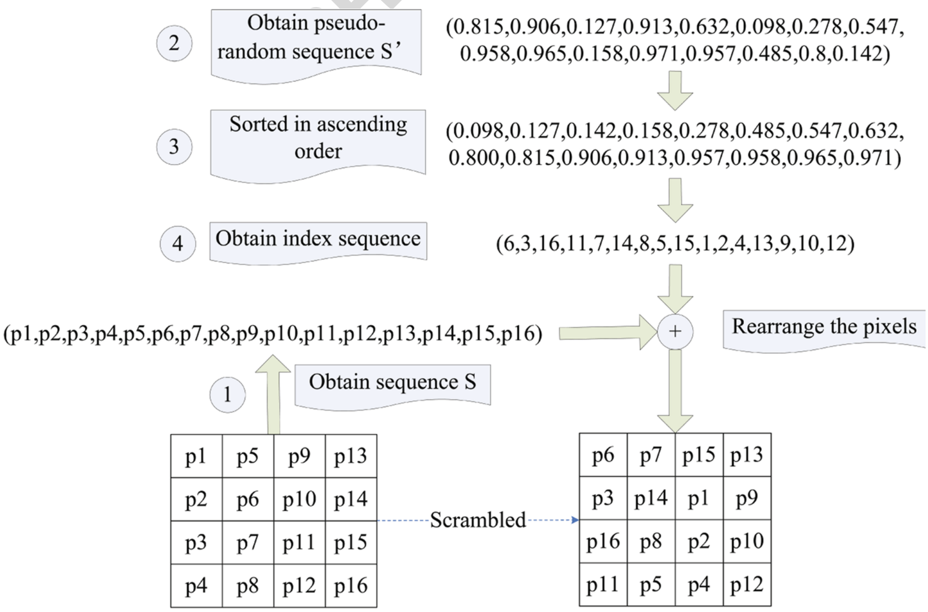

Fig. 4 Scrambled by pseudo-random sequence

where the symbol $\oplus$ denotes the binary XOR operator.

Let $\rho_{wz,i}^{bef}$ and $\rho_{wz,i}^{aft}$ be the error probability calculated before and after scrambling, respectively. Let ρ^{bef} be a set consisting of $\rho_{wz,i}^{bef}$ and ρ^{aft} be a set consisting of $\rho_{wz,i}^{aft}$, $i = 1, 2, ..., k$.

Figure 5 shows the comparison between ρ^{bef} and ρ^{aft} for any one frame in the test videos. It can be seen that the error probabilities changes from inhomogeneous to approximately homogeneous. Furthermore, we calculate the variance of each set. The bigger the variance is, the more inhomogeneous the error probabilities in the set are, and vice verse. Figure 6 depicts the variance of set ρ^{bef} and the variance of set ρ^{aft} for each R_{wz-f} in the test videos. It shows that the variance of set ρ^{bef} is larger than the one of set ρ^{aft}. The latter is approximately equal to 0 which means the error probabilities in set ρ^{aft} are approximately homogeneous. Figures 5 and 6 demonstrate that the scrambling is effective. According to (7) (see Section 3.3.2), we know that the error probability is proportional to the parity bit rate. When the error probabilities are approximately homogeneous, the parity bit rates can be the same, i.e., only one rate is estimated for the k codewords.

3.3.2 Rate estimation

As mentioned above, the error probabilities of k codewords are approximately homogeneous after pseudo-random sequence scrambling. We assume that $\rho = \rho_{wz,1} = \rho_{wz,2} \cdots = \rho_{wz,i}$ $(i = 1, 2 \cdots k)$. If we know the value of error probability ρ , the parity bit rate can be calculated by (7)

$$v = \mathrm{H}(\rho) + 0.03 = -\rho\log_2(\rho) - (1-\rho)\log_2(1-\rho) + 0.03 \tag{7}$$

$H(\rho)$ is the theory bound. In fact, the parity bit rate is always more than $H(\rho)$. Thus, we modify the rate formula into (7) where 0.03 is the empirical value.

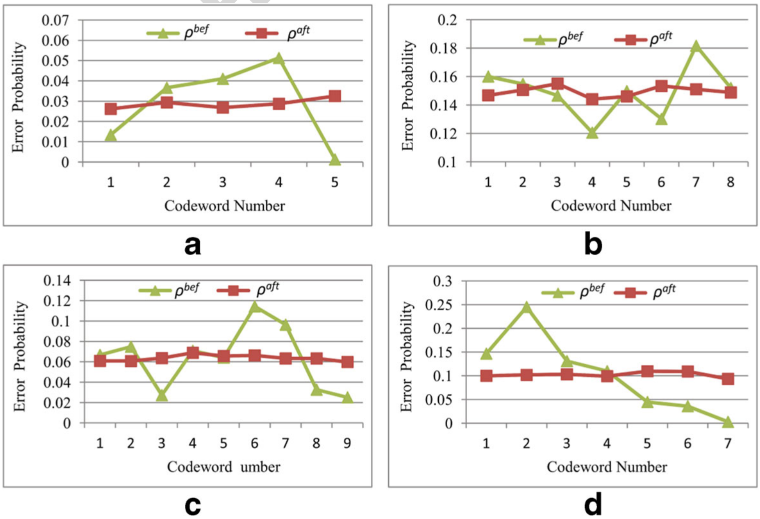

Fig. 5 Comparison of ρ before and after **a** the 6^{th} residual frame of Hall Monitor, **b** the 14^{th} residual frame of Foreman, **c** the 18^{th} residual frame of Coastguard, and **d** the 54^{th} residual frame of Soccer scrambled by pseudo-random sequence

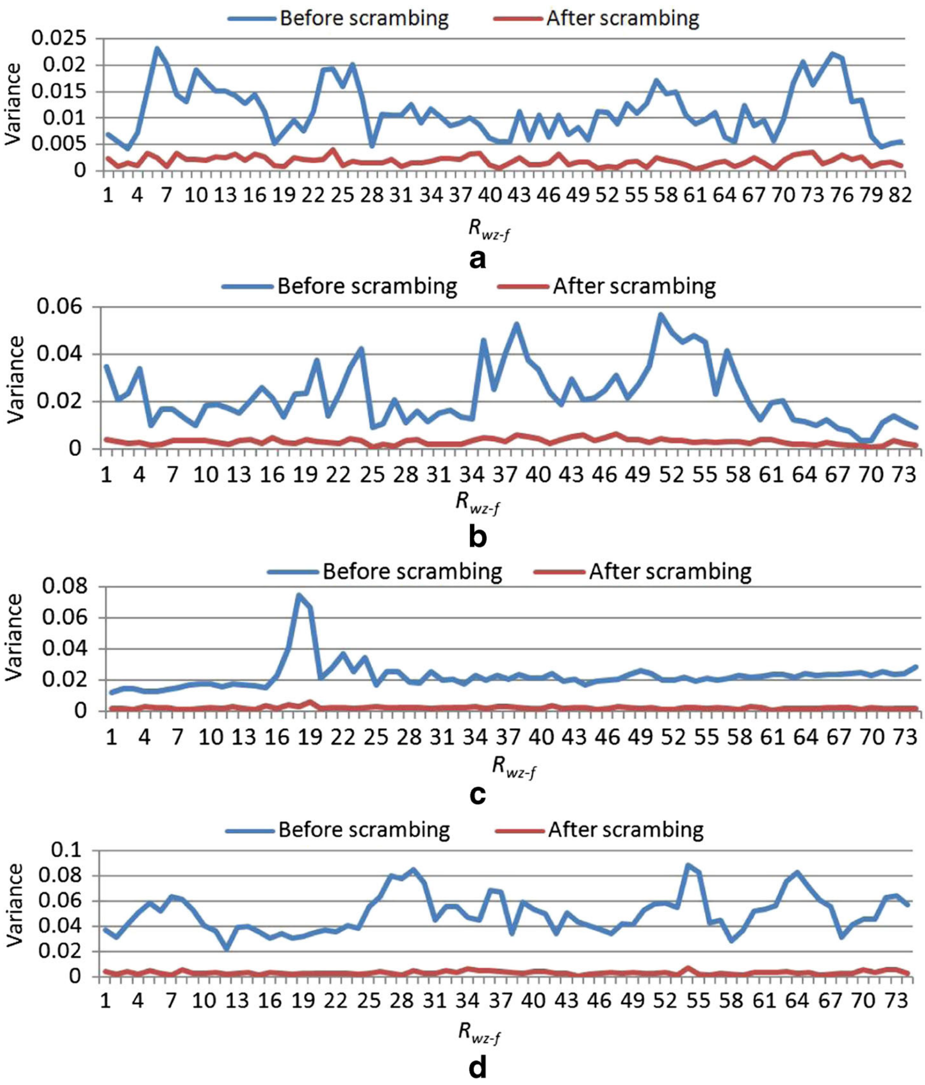

Fig. 6 Variance of set ρ^{bef} and variance of set ρ^{aft} for each R_{wz-f} in **a** Hall Monitor, **b** Foreman, **c** Coastguard, and **d** Soccer

How to compute ρ? From the analysis of the quantization version of R' in Section 3.1 , we can infer that the codewords at the decoder is full of 0s under the hypothesis that $R'_q = 0$ accounts for 100 %. Let $num(i)$ be the number of 1 in $C_{wz,i}$, e.g., if $C_{wz,i} = 0101001$, then $num(i) = 3$. Therefore, ρ is estimated by (8) and denoted as ρ_{est}.

$$\rho_{est} = \frac{\sum_{i=1}^{k} num(i)}{k * L} \tag{8}$$

As we can see that (8) has nothing to do with ESI. It just depends on the number of 1 in k codewords, so the encoder does not generate any ESI that maintains the low-complexity at the encoder.

3.4 Quantization index estimation algorithm (QIEA) based on pseudo-random sequence scrambling

As mentioned in Section 3.3.1, all the codewords are transmitted at the same rate v. If v is overestimated or well estimated, the codewords are decoded successfully which are used to obtain the decoded quantization indexes. Otherwise, the failed decoded codewords and the failed quantization indexes are obtained. Let $\hat{R}_q^{dec}$ and $\hat{R}_q^{notdec}$ be the decoded and failed decoded quantization index, respectively. After inverse scrambling, $\hat{R}_q^{notdec}$ will be scattered among $\hat{R}_q^{dec}$. So the adjacent $\hat{R}_q^{dec}$ can be used to predict $\hat{R}_q^{notdec}$. Let R'_{q-esti} be the predicted value of $\hat{R}_q^{notdec}$. If we use R'_{q-esti} to decode the failed codewords again, we can solve the problem of underestimation. Figure 7 shows how $\hat{R}_q^{notdec}$ scattered among $\hat{R}_q^{dec}$ after inverse scrambling ('$\surd$' represents $\hat{R}_q^{dec}$ and '$\times$' represents $\hat{R}_q^{notdec}$). More specifically, QIEA works as follows.

1. Obtain R'_q which is corresponding to $\hat{R}_q^{notdec}$ and all $\hat{R}_q^{dec}$ which are adjacent (referring to up, down, left, and right neighborhoods) to $\hat{R}_q^{notdec}$.
2. Calculate the absolute difference (AD) between R'_q and each $\hat{R}_q^{dec}$.
3. Assign $\hat{R}_q^{dec}$ with the minimum AD to R'_{q-esti}.
4. Use R'_{q-esti} instead of R'_q to decode the failed codeword again.
5. If the number of $\hat{R}_q^{notdec}$ is reduced, then go to step 1; otherwise, end the algorithm.

4 Experimental results and discussion

To evaluate the performance of the proposed ERC solution, we perform extensive simulations. In the simulations, four test video sequences, namely Hall Monitor, Foreman, Coastguard with QCIF resolution at 15Hz and Soccer with QCIF resolution at 30Hz are employed. The GOP is 2. Odd frame is Key frame encoded by H.264/AVC Intra for QP parameter equal to 16, 18, 20, 24, 27, 30, 32, and 34, respectively. Even frame is WZ frame. The reference frame and residual frame are the same as those described in the introduction section of the ERC-EBMD-DRVC architecture in 3. For WZ blocks, the non-uniform quantization mentioned in Section 3.1 is used and the codeword length of LDPC is 396.

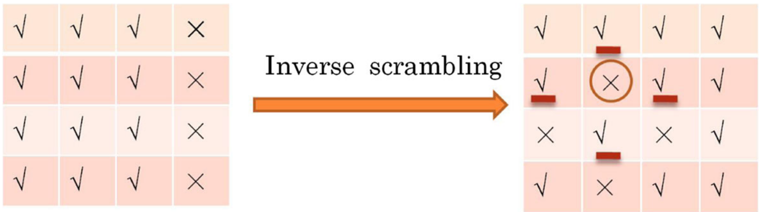

Fig. 7 Inversely scrambled by pseudo-random sequence

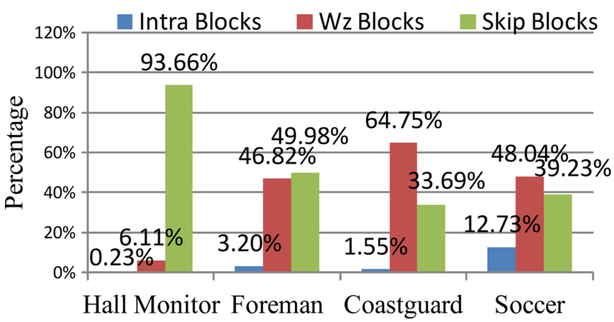

Fig. 8 The percentages of each mode in test videos

4.1 Efficiency of EBMD

Figure 8 shows the percentage of each mode in the test sequences where the QP is 24. It shows Skip blocks in Hall Monitor account for the largest proportion, reaching 93.66 %. Intra blocks in Soccer accounting for 12.73 % are more than those in other three videos. That means the lower the motion is, the more percentage the skip blocks account for and the higher the motion is, the more percentage the Intra blocks account for. Figure 9 shows the average PSNR (APSNR) for each kind of block which is shown in Fig. 8. It can be seen that for Intra blocks, the APSNR is mainly 14dB-17dB that means the correlations between Intra blocks and their side information blocks are weak and the Intra codec is appropriate. For WZ blocks, the APSNR is mainly 29 dB that means the qualities of their side information blocks are medium and WZ codec is appropriate. For Skip blocks, the APSNR is as high as 43.19 dB, usually over 39 dB, which means the correlations between Skip blocks and their side information blocks are strong. So Skip blocks can be replaced by their side information blocks. There is an exceptional case for Intra blocks and WZ blocks in Coastguard. As we can see, the APSNR of the two mode blocks in Coastguard are higher than those in other videos. It is because Coastguard is a video with well behaved motion. The quality of the side information generated by MCFI for Coastguard is good. We know that the quality and quantity of Skip blocks, WZ blocks, and Intra blocks are all contribute to the quality of SI. Figure 8 shows Skip blocks in Coastguard accounting for 33.69 % is the lowest percentage when compared with Skip blocks in other videos. So,in Coastguard, Skip blocks

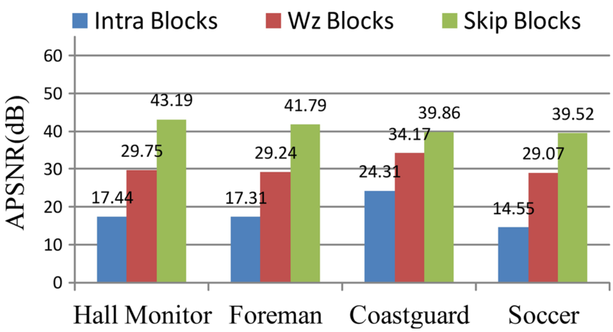

Fig. 9 The APSNR of each kind of block in our test videos

make relatively less contributions to the good quality of SI and therefore it results in Intra blocks and WZ blocks having relatively better SI blocks when compared with other videos.

Figure 10 shows the APSNR comparison before and after using (5) for the blocks which are wrongly classified as Skip mode. It can be seen that the gains are up to about 2.8dB, 2.1 dB, 1.2dB, and 1.4dB in Hall Monitor, Foreman, Coastguard, and Soccer, respectively. The results show (5) is simple and helpful.

In short, as we can see from the results of Figs. 8–10, the proposed EBMD is very simple and effective in mode decision for any degree motion video. Since the proposed EBMD is based on the hypothesis that $R_q' = 0$ accounts for 100 %, the satisfactory results also justify the hypothesis.

4.2 Efficiency of ERC

To testify the efficiency of our rate estimation method, we can compare the estimated error probability ρ_{est} with the real error probability ρ_{real} and then compare the estimated rate v with the ideal rate v' which is obtained in DRC scenario. Figure 11 shows the former comparison. It can be seen that the estimated ρ_{est} calculated by (8) are mostly close to ρ_{real} in Hall Monitor, Foreman and Soccer. But in Coastguard ρ_{est} is mostly higher than ρ_{real} . It is because that the quality of WZ blocks in Coastguard (about 34dB) is higher than the ones (about 29dB) in other three videos from Fig. 9, If (8) is effective to estimate the error probabilities for WZ blocks with lower APSRN, it is certainly overestimated for WZ blocks with higher APSRN. Figure 12 shows the latter comparison. There are three scenarios, namely, v higher than v' , v lower than v' and v equal to v' , which represent overestimation, underestimation, and well-estimation. Overestimation cannot reduce the distortion, but it causes unnecessary bit-rate expansion. Underestimation can induce errors and result in severe distortion. Figure 12 shows the comparison of v and v' for each frame in the tested videos. It can be seen that most of the points are closer to line $v = v'$ which means the difference between v and y is little. In the four videos, the overestimation in Coastguard is in the majority that because ρ_{est} is overestimated and so the rate.

4.3 Efficiency of QIEA

Figure 13 shows the successful decoding ratios (SDR) before and after adopting QIEA. The SDR refers to the number of the successful decoded codewords divided by the number of

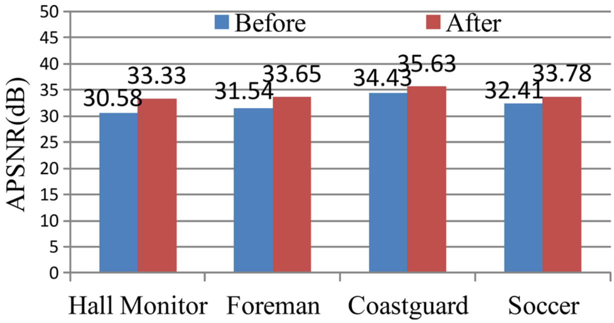

Fig. 10 APSNR comparison before and after using (5) for the blocks which are wrongly classified as Skip mode

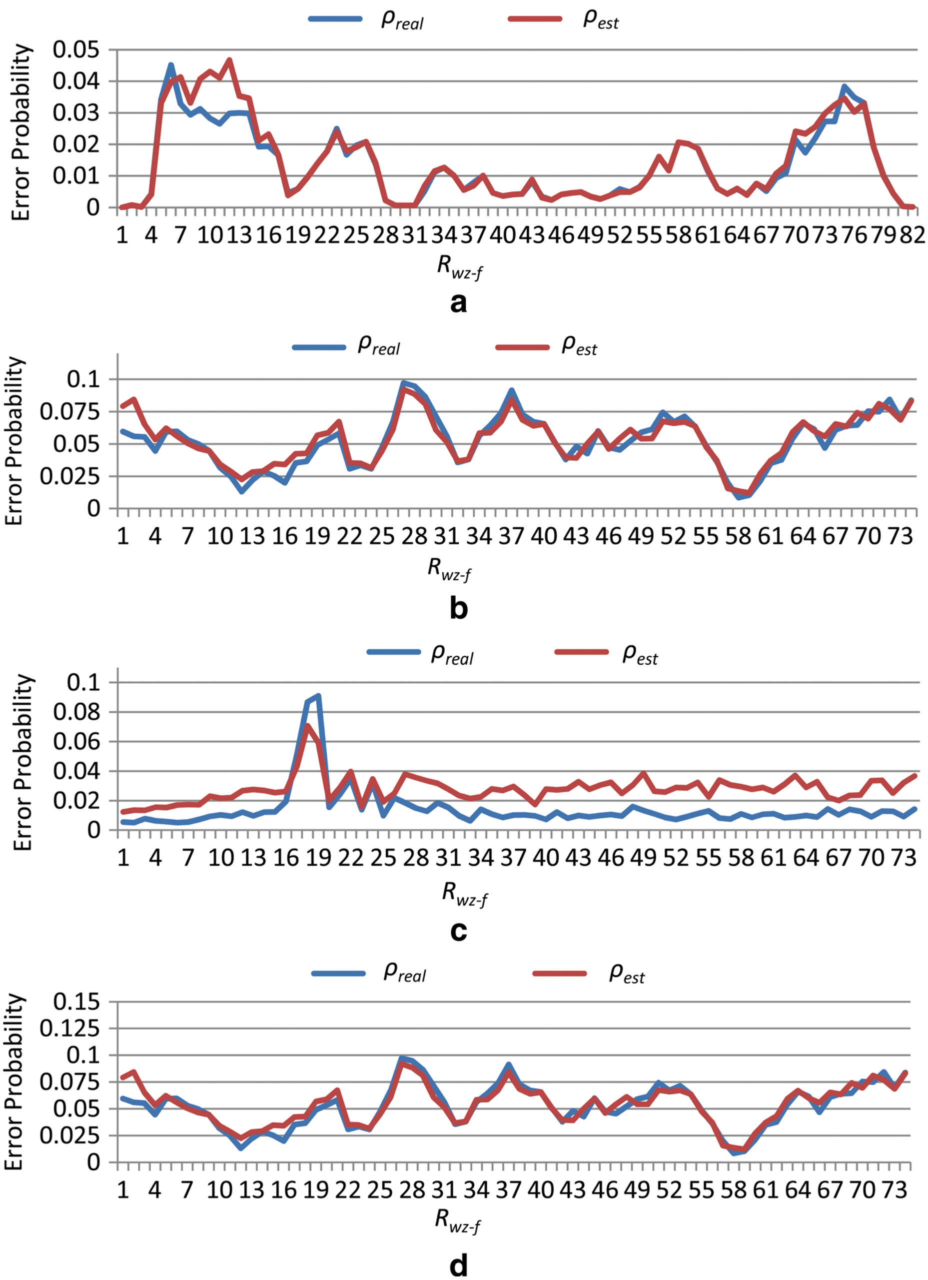

Fig. 11 Comparison of ρ_{est} and ρ_{real} in **a** Hall Monitor, **b** Foreman, **c** Coastguard, and d Soccer

all the codewords in a video. As seen from Fig. 13, for Coastguard, the SDR before using QIEA is up to 84 % which is the highest ratio in the four videos. The explanation is that the overestimation in Coastguard is in the majority which helps to improve the SDR. The SDRs for Hall Monitor, Foreman, and Soccer before using QIEA are 58.54 %, 51.06 %, and

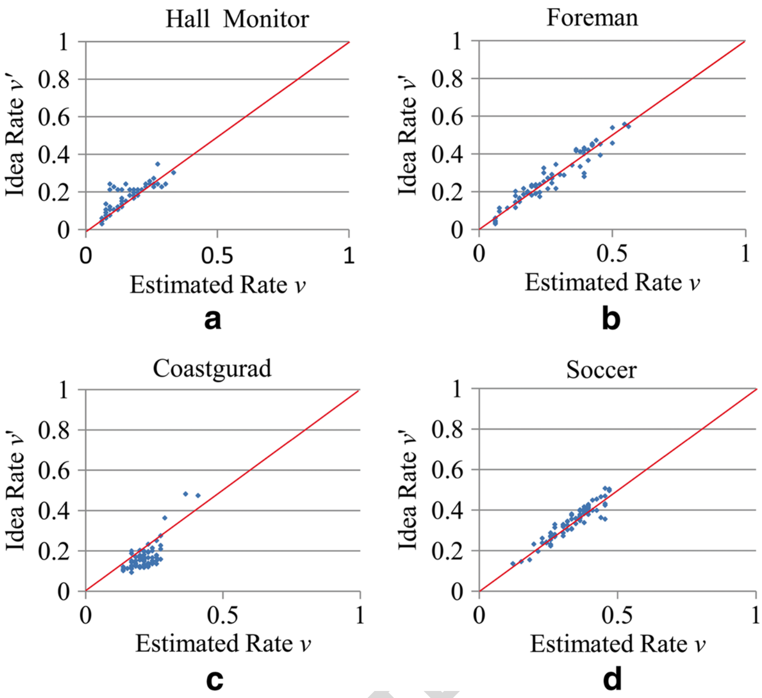

Fig. 12 Estimated v versus idea rate v' for **a** Hall Monitor video, **b** Foreman video, **c** Coastguard video, and **d** Soccer video

47.6 %, respectively. It means the higher the motion, the more difficult is to estimate the rate and the lower the SDR. After using QIEA, SDR is increased for all videos. Especially for Soccer, the SDR reaches 94.46 %. The results show our QIEA is efficient. There are two reasons. One is that after inverse scrambling, the failed decoded quantization indexes $\hat{R}_q^{notdec}$ are scattered among the successful decoded quantization indexes $\hat{R}_q^{dec}$ which can be used to predict $\hat{R}_q^{notdec}$. The other is the prediction is always accurate duo to the range of the quantization indexes is narrow and only three values, namely -1,0,1, are defined after non-uniform quantization. The SDR for Hall Monitor changes from 58.54 to 59.76 % which seems our QIEA is inefficient. As we know the efficiency of QIEA depends on it that

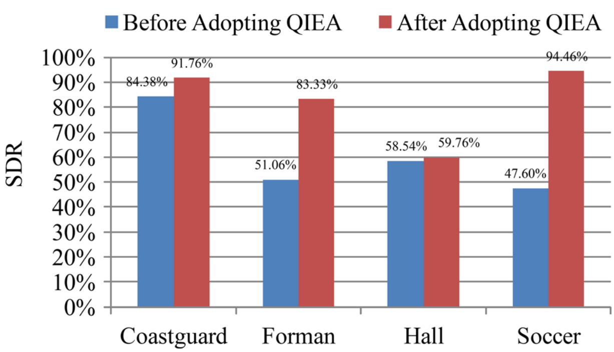

Fig. 13 Comparison of SDR of our test videos before and after adopting QIEA

$\hat{R}_q^{notdec}$ must be scattered among $\hat{R}_q^{dec}$ which indicates that the more the number of $\hat{R}_q^{dec}$, the better the performance of QIEA is. But in Hall Monitor, the percentage of WZ blocks is only 6.11 % in Fig. 8. Hence, the number of $\hat{R}_q^{dec}$ must be less than 6.11 %, so the effect is inconspicuous. Figure 14 shows the example for the improvement of the image quality before and after adopting QIEA. As we can seen there are some failed decoded blocks especially in the face in (a) which are successfully become decoded in (b) and the PSNR is improved by 5.1dB.

4.4 Analysis of the complexity

ERC-EBMD-DRVC is a simple video framework which is the basic residual video framework adding EBMD module, scrambling module, ERC module, and QIEA module. Although EBMD module, scrambling module, and ERC module are added at the encoder, they are all simple. EBMD only depends on the values of residual pixels to determine the block mode. Scrambling module generates a pseudo-random sequence and sorts it in order. ERC module without ESI estimates the rate at frame level. None of them brings heavy calculation and latency at the encoder. Although QIEA module is added at the decoder, it is also simple. Figure 15 shows the average iteration for QIEA with all QP values for the test videos. We can see the number of iteration does not exceed 3 that means the latency and complexity increased at the decoder are not severe.

4.5 RD performance of ERC-EBMD-DRVC

Figure 16 shows the RD performance of the proposed ERC-EBMD-DRVC solution for all the test videos, compared with that of DISCOVER [10] and TDDVC-ERC [6]. Only luminance component is taken into account.

1. Compared with DISCOVER, DISCOVER is currently considered as one of the best performing TDDVC system which is regarded as the benchmark for DRC scenario. The simulation results of DISCOVER come from [10]. Figure 16 shows that ERC-EBMD-DRVC performs better (the gain up to 1.5dB on average) for Hall Monitor video, the reason of which is that Skip blocks are in the majority and help to improve the RD performance. For Coastguard, ERC-EBMD-DRVC achieves RD performance similar to the one obtained by DISCOVER codec which is explained that WZ codec works well in Coastguard. As we can see from Figs. 8 and 9 that the percentage and APSNR

Fig. 14 Comparison of the image quality for the 148th frame of Soccer before and after adopting QIEA, **a** PSNR = 27.254 dB before adopting QIEA, **b** PSNR = 32.31 dB after adopting QIEA

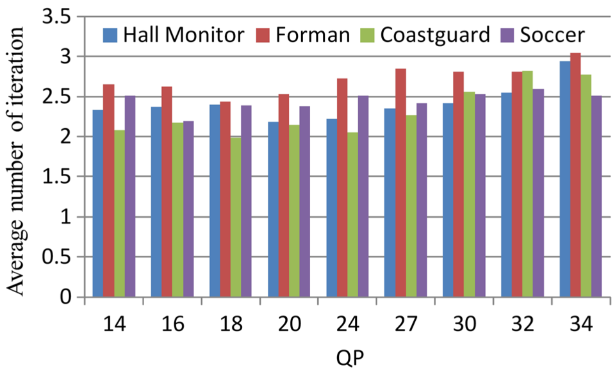

Fig. 15 Comparison of SDR of our test videos before and after adopting QIEA

of WZ blocks in Coastguard are 64.75 % and 34.17 dB respectively which means WZ blocks are not only in the majority but also with satisfied SI, so the parity bits needed to correct the errors in SI are not much which help to achieve good performance. For video sequences characterized by moderate to high motion such as Foreman and Soccer, the proposed ERC-EBMD-DRVC are close to DISCOVER at low bit rates and presents a very small RD performance gap at high rates. Because the rate estimation is a little difficult for such videos with unsatisfied SI. In a word, the proposed solution is very efficient especially for the videos with low and well behaved motion such as Hall Monitor and Coastguard.

2. Compared with TDDVC-ERC [6], TDDVC-ERC [6] is a very powerful and efficient ERC solution and becomes a new benchmark for other ERC solutions as it was mentioned in [6]. The simulation results of TDDVC-ERC come from [6] in which it

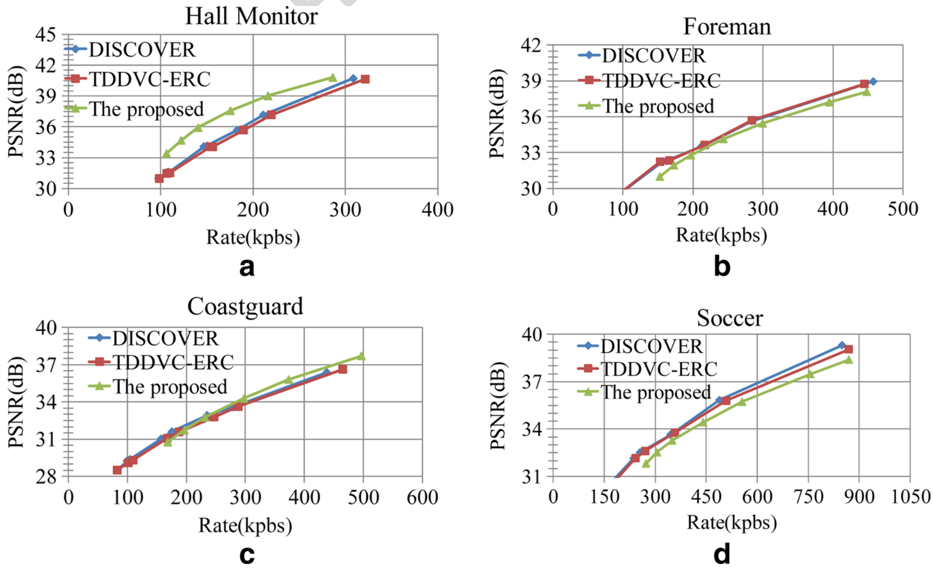

Fig. 16 RD performance of ERC-EBMD-DRVC solution and benchmarks

provides a RD performance quite close to the target RD performance obtained by DISCOVER. As we can see from Fig. 16, ERC-EBMD-DRVC outperforms TDDVC-ERC for the videos with low motion and achieves similar RD performance for other videos at low rates and presents a very small gap at high rates. The results indicate that the proposed ERC solution is competitive due to its good RD performance. TDDVC-ERC has a complex framework in which many improved techniques such as a novel weighted overlapped block motion compensation technique, a correlation noise model updating technique, and a novel soft reconstruction technique are used to improve the RD performance. Furthermore, in TDDVC-ERC, ESI is required to be generated and the rates are estimated at bit plane level which increase the complexity and incur computation overheads at the encoder. While in ERC-EBMD-DRVC, ESI is not required and the rates are estimated at frame level. Due to the competitive RD performance and simple framework, ERC-EBMD-DRVC is more practical than TDDVC-ERC solution.

5 Conclusion

This paper proposes an efficient ERC solution called ERC-EBMD-DRVC which maintains a low-complexity encoder. In order to improve the RD performance, the proposed ERC solution combines a simple yet efficient EBMD which only employs the values of residual pixels to decide the block coding mode. Based on the hypothesis that $R'_q = 0$ accounts for 100 %, the correlation between the residual frames at the encoder and decoder is simply estimated by (8) which has nothing to do with ESI. So the encoder does not need to generate ESI that greatly decreases the complexity at the encoder. Moreover, our rate estimation is at frame level that brings fewer computational load and latency than the ones brought by the rate estimation at bit plane level. The problem of rate underestimation is also solved using QIEA in this paper. The simulation result shows that ERC-EBMD-DRVC outperforms DISCOVER and the state-of-the-art ERC solution in the videos with low motion and has competitive RD performance for other videos with moderate or high motion. Furthermore, our scheme has simple framework. Although EBMD module, scrambling module, ERC module, and QIEA module are added, they are all simple. So ERC-EBMD-DRVC is a promising scheme.

Acknowledgment This work is supported by the Ministry of Industry and Information Technology of Things special fund([2014]351), the National Natural Science Foundation of China (No.61302055), and the Project sponsored by SRF for ROCS, SEM.

References

1. Aaron A, Zhang R, Girod B (2002) Wyner-Ziv coding of motion video. In: The thirty-sixth asilomar conference on signals systems and computers, institute of electrical and electronics engineers computer society, conference record of the asilomar conference on signals, systems and computers, vol 1, pp 240–244. doi:10.1109/ACSSC.2002.1197184
2. Aaron A, Rane S, Setton E, Girod B (2004) Transform-domain Wyner-Ziv codec for video. In: Visual communications and image processing 2004, SPIE, Proceedings of SPIE - The international society for optical engineering, vol 5308, pp 520–528. doi:10.1117/12.527204
3. Aaron A, Varodayan D, Girod B (2006) Wyner-Ziv residual coding of video. In: 25th PCS: picture coding symposium 2006, PCS2006, INRIA, 25th PCS Proceedings: picture coding symposium 2006, PCS2006, vol 2006, pp 1–5

4. Ascenso J, Pereira F (2009) Low complexity intra mode selection for efficient distributed video coding. In: 2009 IEEE international conference on multimedia and expo, IEEE, pp 101–104
5. Brites C, Pereira F (2007) Encoder rate control for transform domain Wyner-Ziv video coding. In: 2007 IEEE international conference on image processing, vol 2. IEEE, pp II–5–II–8
6. Brites C, Pereira F (2011) An efficient encoder rate control solution for transform domain Wyner-Ziv video coding. IEEE Trans Circuits Syst Video Technol 21(9):1278–1292
7. Chiang JC, Chen KL, Chou CJ, Lee CM, Lie WN (2010) Block-based distributed video coding with variable block modes
8. Clerckx T, Munteanu A, Cornelis J, Schelkens P (2007) Distributed video coding with shared encoder/decoder complexity. In: 14th IEEE international conference on image processing, ICIP 2007, inst. of elec. and elec. eng. computer society, proceedings - international conference on image processing, ICIP, vol 6, pp VI417–VI420. doi:10.1109/ICIP.2007.4379610
9. Du B, Shen H (2009) Encoder rate control for pixel-domain distributed video coding without feedback channel, IEEE
10. Group MSP (2007) http://www.img.lx.it.pt/~discover/home.html
11. HoangVan X, Jeon B (2012) Flexible complexity control solution for transform domain Wyner-Ziv video coding. IEEE Trans Broadcast 58(2):209–220
12. Ji W, Frossard P, Chen Y (2014) Exit-based side information refinement in Wyner-Ziv video coding. IEEE Trans Circuits Syst Video Technol 24(1):141–156
13. Liu L, He D, Jagmohan A, Lu L, Delp EJ (2008) A low-complexity iterative mode selection algorithm for Wyner-Ziv video compression. In: 2008 15th IEEE international conference on image processing, IEEE, pp 1136–1139
14. Liu W, Vijayanagar KR, Kim J (2013) Low-complexity distributed multiple description coding for wireless video sensor networks. IET Wireless Sens Syst 3(3):205–215
15. Morbée M, Prades-Nebot J, Roca A, Pižurica A, Philips W (2007) Improved pixel-based rate allocation for pixel-domain distributed video coders without feedback channel. In: International conference on advanced concepts for intelligent vision systems, Springer, pp 663–674
16. Morbee M, Prades-Nebot J, Pizurica A, Philips W (2007) Rate allocation algorithm for pixel-domain distributed video coding without feedback channel. In: 2007 IEEE international conference on acoustics, speech and signal processing, ICASSP '07, institute of electrical and electronics engineers inc., ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol 1, pp I521–I524. doi:10.1109/ICASSP.2007.366731
17. Natário L, Brites C, Ascenso J, Pereira F (2005) Extrapolating side information for low-delay pixel-domain distributed video coding. In: Visual content processing and representation: 9th international workshop, VLBV 2005, Springer, vol 3893, p 16
18. Qiu G, Gao S, Tu G (2013) Low-complexity low-delay distributed video coding. In: 2013 18th international conference on digital signal processing, DSP 2013, IEEE computer society, 2013 18th international conference on digital signal processing, DSP 2013. doi:10.1109/ICDSP.2013.6622739
19. Ruchet G, Wang D (2010) Distributed video coding without channel codes. In: 2010 IEEE international symposium on broadband multimedia systems and broadcasting, BMSB 2010, IEEE computer society, IEEE international symposium on broadband multimedia systems and broadcasting 2010, BMSB 2010 - final programme. doi:10.1109/ISBMSB.2010.5463126
20. Sheng T, Zhu X, Hua G, Guo H, Zhou J, Chen CW (2010) Feedback-free rate-allocation scheme for transform domain Wyner-Ziv video coding. Multimedia Systems 16(2):127–137
21. Skorupa J, Slowack J, Mys S, Deligiannis N, De Cock J, Lambert P, Grecos C, Munteanu A, Van de Walle R (2012) Efficient low-delay distributed video coding. IEEE Trans Circuits Syst Video Technol 22(4):530–544
22. Slepian D, Wolf J (1973) Noiseless coding of correlated information sources. IEEE Trans Inf Theory 19(4):471–480
23. Tagliasacchi M, Trapanese A, Tubaro S, Ascenso J, Brites C, Pereira F (2006) Intra mode decision based on spatio-temporal cues in pixel domain Wyner-Ziv video coding. In: 2006 IEEE international conference on acoustics, speech and signal processing, ICASSP 2006, institute of electrical and electronics engineers inc., ICASSP, IEEE international conference on acoustics, speech and signal processing - proceedings, vol 2, pp II57–II60
24. Tang ZH, Liang XY, Qin TF, Chang K (2015) Correlation noise modeling algorithm based on multiple probability distributions for distributed video coding. Tien Tzu Hsueh Pao/Acta Electron Sin 43(2):365–370. doi:10.3969/j.issn.0372-2112.2015.02.024
25. Van Luong H, Raket LL, Forchhammer S (2014) Re-estimation of motion and reconstruction for distributed video coding. IEEE Trans Image Process 23(7):2804–2819

26. Wang Y, Wu C (2010) A block based Wyner-Ziv video codec. In: 3rd international congress on image and signal processing (CISP), 2010, IEEE, vol 1, pp 1–5
27. Wyner A, Ziv J (1976) The rate-distortion function for source coding with side information at the decoder. IEEE Trans Inf Theory 22(1):1–10
28. Yang HP, Hsieh HC, Chang SH, Chen SJ (2015) An improved distributed video coding with low-complexity motion estimation at encoder. In: 28Th IEEE international system on chip conference, SOCC 2015, IEEE computer society, international system on chip conference, vol 2016, pp 111–114. doi:10.1109/SOCC.2015.7406923
29. Zhang D, Wu Y, Wan M (2014) Improved side information generation algorithm for Wyner-Ziv video coding. The Journal of China Universities of Posts and Telecommunications 21(1):109–115
30. Zhang Y, Xiong H, He Z, Yu S, Chen CW (2011) Reconstruction for distributed video coding: a context-adaptive markov random field approach. IEEE Trans Circuits Syst Video Technol 21(8):1100–1114

Chunyun Hu was born in Jiangxi, China. She received the B.S. degree from Nanchang University, Jiangxi, China, in 1998 and the M.S. degree from South China University of Technology, Guangzhou,China, in 2003, where she is currently working toward the Ph.D. degree. Her research interests include image and video processing, distributed source coding, and routing problem.

Binjie Hu was born in Shanxi, China. He received the M.S. degree from China Research Institute of Radiowave Propagation, Henan, China, in 1991 and the Ph.D. degree from the University of Electronic Science and Technology of China, Sichuan, China, in 1997. From 1997 to 1999, he was a Postdoctoral Fellow with South China University of Technology, Guangzhou, China. From 2001 to 2002, he was a Visiting Scholar at the Department of Electronic Engineering, City University of Hong Kong. In 2005, he was a Visiting Professor at the Universitë de Nantes, Nantes, France. He is currently a Full Professor at South China University of Technology. His current research interests include wireless communications, cognitive radios, microwave circuits, and antennas.

Wanqing Tu is a Reader (Associate Professor) in the School of Computing Science and Digital Media, Robert Gordon University, Aberdeen, the UK. Before joined RGU, she was a Senior Lecturer in Nottingham Trent University and Glyndwr University. Dr. Tu was an IRCSET Embark Initiative Postdoctoral Research Fellow at University College Cork, Ireland. She received her PhD from the City University of Hong Kong in 2006. Dr. Tu's research interests lie in next-generation multimedia communication networking, high-performance and secured IoT services, energy-efficient communications, etc. She is a Senior Member of IEEE.

Yunhui Xiong is an Associate Professor of School of Mathematics in South China University of Technology. He received Ph.D. degree in computer science engineering from the South China University of Technology in 2010.12. His research interests are image and video processing, distributed source coding, geometry modeling and processing, 3D printing and computer graphics.