A simple encoder scheme for distributed residual video coding.

HU, C., ZHAO, Y., YU, L., JIANG, Y. and XIONG, Y.

2020

A simple encoder scheme for distributed residual video coding

Chunyun Hu[1] · Yafan Zhao[2] · Long Yu[1] · Yang Jiang[3] · Yunhui Xiong[4]

Received: 11 January 2019 / Revised: 24 December 2019 / Accepted: 28 February 2020 /

Abstract

Rate-Distortion (RD) performance of Distributed Video Coding (DVC) is considerably less than that of conventional predictive video coding. In order to reduce the performance gap, many methods and techniques have been proposed to improve the coding efficiency of DVC with increased system complexity, especially techniques employed at the encoder such as encoder mode decisions, optimal quantization, hash methods etc., no doubt increase the complexity of the encoder. However, low complexity encoder is a widely desired feature of DVC. In order to improve the coding efficiency while maintaining low complexity encoder, this paper focuses on Distributed Residual Video Coding (DRVC) architecture and proposes a simple encoder scheme. The main contributions of this paper are as follows: 1) propose a bit plane block based method combined with bit plane re-arrangement to improve the dependency between source and Side Information (SI), and meanwhile, to reduce the amount of data to be channel encoded 2) present a simple iterative dead-zone quantizer with 3 levels in order to adjust quantization from coarse to fine. The simulation results show that the proposed scheme outperforms DISCOVER scheme for low to medium motion video sequences in terms of RD performance, and maintains a low complexity encoder at the same time.

Keywords Distributed residual video coding (DRVC) Bit plane block based · Low complexity encoder

1 Introduction

Distributed Video Coding (DVC), based on Slepian-Wolf [27] and Wyner-Ziv [34] theorems, is an emerging video coding paradigm and it shifts the computational complexity and storage burdens from encoder to decoder. It encodes video frames independently and decodes them jointly, which is a promising scheme for applications with limited resources [25], such as sensor networks, wireless video surveillance, etc. However, RD performance of DVC is still considerably less than that of conventional motion-compensation based video

✉ Yunhui Xiong
yhxiong@scut.edu.cn

Extended author information available on the last page of the article.

coding such as H.264/AVC. As indicated in [16], the mostly adopted DVC framework DISCOVER consistently outperforms H.264/AVC Intra-coding (i.e., all the frames are Intra coded) in terms of RD performance, except for scenes with complex motion. Nevertheless, the RD performance of DISCOVER remains generally inferior to that of a full H.264/AVC codec. The gap of the coding performance is due to a number of reasons: sub-optimality of channel coding tools, inaccuracies in the correlation noise model, and bad SI qualities. To further enhance the RD performance of the DVC framework, many methods and techniques have been proposed at the cost of the system complexity, especially techniques used at the encoder such as encoder mode decisions, optimal quantization, hash methods etc., no doubt increase the complexity of the encoder.

Encoder Block Mode Decision (EBMD) is a useful method for improving the coding efficiency and many literatures [9, 11, 13, 17, 22, 24, 32] have focused on this method. Authors in [22] presented a structure of Temporal Group of Blocks (TGOB) at the encoder, which assesses the spatial-temporal properties of each image block to determine suitable modes dynamically. Experimental results show that achieved RD gains depend on video content. This method is not suitable, and is more complex for videos of high dynamics. In [11, 17], in order to reduce the computational complexity, the mode selection algorithm uses the Sum of Absolute Difference (SAD) between the blocks as an indication of the temporal coherence solution.A bit plane Motion Estimation (ME) algorithm is proposed in [13] and the residual error of ME is used in the selection of the coding mode for each block. In [24], SI is required to be generated at the encoder and an iterative algorithm is proposed to select the mode dynamically to improve the accuracy of mode decision. In [9, 32], Lagrange RD cost function of each mode is calculated for each block and the block mode with the minimum cost is chosen as the best mode. Although the above EBMD algorithms improve the RD performance considerably, the complexity of the encoder is also increased.

SI plays an important role in DVC. This is because higher compression efficiency and lower bitrates could be achieved with an increased dependence of SI. Since the decoder does not have any information about current WZ frame, the hash can be sent at the encoder as auxiliary information to assist SI generation [1, 2, 8, 14, 15, 20, 26, 30]. The common hash function is CRCs [26]. Both low frequency DCT coefficients and high frequency DCT coefficients are used for hash codes in [2] and [1]. Adaptive hash-based approach was proposed in [8] which selects different number of low frequency coefficients as the hash code for blocks of a frame to achieve optimal RD performance. Other information, such as the b most significant bit planes that are entropy coded [14, 30] and a down-sampled WZ frame that is intra coded [15], can be used as hash information. The detailed analyses of hash-based motion estimation were presented in [20]. Most experimental results show that hash-based methods benefit video with medium to high bitrates than that with low bitrates. This indicates that more hash code is required for better SI. Therefore, the performance of the hash-based DVC is enhanced at the cost of increased number of hash code and the complexity of encoding.

Quantization is another important technique employed at the encoder. In most of the existing DVC schemes, scalar quantization is often used due to its simplicity. Scalar quantization scheme does not take into account of the DCT coefficient distribution in a block or the characteristics of the video sequences, therefore it cannot achieve optimal coding performance. Adaptive quantization methods have been proposed in many DVC schemes [12, 28, 31, 33, 35]. Literature [35] first analyzed three types of adaptive quantization methods including frame level adaptive quantization [12], sub-band level adaptive quantization

[28] and overall adaptive quantization [31], and then proposed a perception-based adaptive quantization scheme. The scheme is very complex because SI needs to be generated at the encoder and model of Perceptual Distortion Probability (PDP) is developed to estimate the perceptual distortion of SI and to derive the target perceptual distortion. Three components (i.e. quality of SI frame, perceptual features and RD optimization) are integrated with the estimated perceptual distortion of SI and target perceptual distortion to determine the optimal quantization matrix adaptively and iteratively. An optimal entropy-constrained non-uniform scalar quantizer was proposed for pixel domain DVC in [33] that is also complex. First, an estimation of the rate and distortion model based on the conditional probability density function is adopted at the encoder. Then, a rate-distortion optimization function is derived. A modified Lloyd-Max algorithm with a novel quantization partition updating algorithm is used to optimize the RD function. Experimental results in both literatures show that the proposed quantization schemes improve the RD performance, but it also increases the complexity of the encoder due to the heavy computational load and the complexity of the developed algorithms.

In addition to EBMD, hash and quantization schemes, there are other technologies employed at the encoder to improve the coding efficiency. Literature [6] proposed an interpolation side information algorithm that uses the SIFI algorithm at the encoder to obtain global motion vector. The proposed algorithm can improve the quality of SI effectively. However, due to the feature-point matching process at the encoder, the complexity of the encoder is increased significantly. Literature [23] proposed a DVC scheme based on the Human Visual System (HVS), in which any changes below the Just-Noticeable-Difference (JND) distortion threshold can hardly be perceived. In order to employ the JND model, a simple SI is generated at the encoder. Experimental results demonstrate that the proposed algorithm saves the bitrates significantly. The additional encoding complexity are mainly contributed by the generation of SI and the calculation of JND values.

The tradeoff between the RD performance and the encoder complexity is still a challenge in DVC. In order to improve the coding efficiency, and meanwhile to maintain a low complex encoder, a simple encoder scheme for DRVC has been proposed in this paper. Our main contributions include:

(1) propose a Bit Plane Block Based (BPBB) method combined with Bit Plane Rearrangement (BPRA) to improve the dependency between source and SI, and meanwhile to reduce the amount of data to be channel encoded. In view of the principle of DVC, the compression efficiency comes from the correlation between source and SI, that is, high correlation results in high compression efficiency. In this paper, based on the analysis of the statistical distribution of residual pixel values, BPBB is proposed to divide each bit plane into non-overlapping 4×4 blocks and each block is classified as 0-Block or 1-Block according to a simple criterion. BPRA is proposed to remove the bits that are not required to be channel encoded.

(2) present an iterative and 3-level dead-zone quantizer. The proposed quantizer has only 3 quantization levels and is very simple. In order to reduce the quantization errors and thus adjust quantization from coarse to fine, the quantized interval will be narrowed at each iteration. Another benefit of the iterative quantization is that the reconstructed frame at last iteration can be used as the refined SI for the current frame, and thereby this can improve the coding efficiency.

The remainder of this paper is structured as follows. Section 2 introduces the related studies on DRVC. Section 3 presents the proposed simple encoder scheme for DRVC in

details. In Section 4, experimental results are shown and discussed. Finally, we conclude the paper in Section 5.

2 Related work about DRVC

The DVC architectures have been developed by researchers in Stanford University, which mainly includes pixel-domain DVC (PDDVC) [4], transform-domain DVC (TDDVC) [3], and distributed residual video coding (DRVC) [5]. This paper focuses on DRVC, which compresses residual frames by PDDVC codec Fig. 1 illustrates the basic architecture of DRVC. At the encoder, a video sequence is divided into key frames K and WZ frames W. For key frames K, the H.264/AVC intra-encoding and decoding are implemented. For WZ frames W, residual frames $R = W - W_{re}$ are obtained and Wyner-Ziv encoded, where W_{re} is a simple estimation to W and is accessible at both encoder and decoder side. After encoding, the parity bits are stored in the buffer and transmitted to the decoder upon the request. At the decoder, a more accurate estimation to W, donated as Y, is generated from previously decoded key frames by motion compensated frame interpolation (MCFI) [7]. $R' = Y - W_{re}$ is Wyner-Ziv decoded as the decoder SI for R. Finally, the reconstruction $\hat{W} = W_{re} + \hat{R}$ is achieved.

In [5], DRVC has been proved that it has the same performance as the TDDVC due to the exploitation of temporal correlation, and it also has less complex than TDDVC because of no DCT transformation. In our previous work [18, 19], we proposed an efficient Encoder Rate Control (ERC) solution and a novel EBMD for DRVC. The proposed EBMD only depends on the values of residual pixels without measurement of block difference, compression rate or distortion function used in the existing EBMD. The proposed ERC is at frame level instead of bit plane level, resulting in fewer computational load and lower latency. The simulation results show that our scheme outperforms DISCOVER and the state-of-the-art ERC solution for video sequences with low motion, and has competitive RD performance for other video sequences. Literature [10] presented two types of DRVC based system. The first one is DRVC based on low-quality reference (LQR) hash, in which the decoded LQR hash is used as W_{re} and the residual R is decomposed by Discrete Wavelet Transform (DWT) followed by SW-SPIHT (Slepian-Wolf Set Partitioning In Hierarchical Trees) coding. The experimental results show this scheme achieves better RD performance than

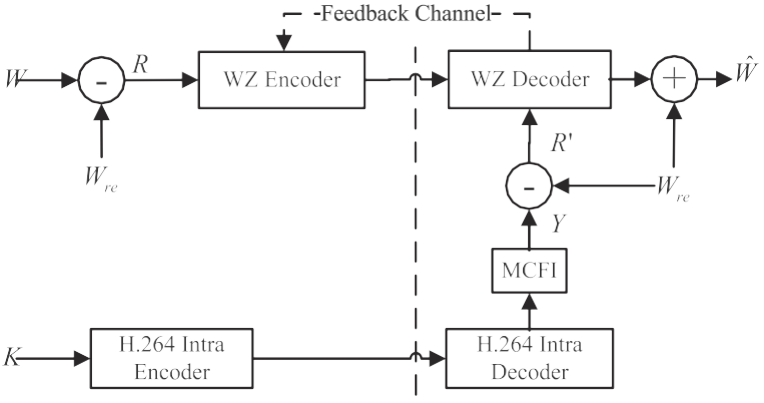


Fig. 1 Basic architecture of DRVC

the existing schemes due to use of the residual coding and the efficient LQR hash-based motion compensation. The other one is DRVC combined with the SW-SPIHT coding and the intra mode decision technique, in which the reference frame W_{re} is obtained by the weighted average interpolation of the previous and next decoded key frames. The residual R is decomposed by DWT,and the resulting coefficients are classified to different modes. The experimental results show this hybrid DVC obtains up to 3 dB improvement for Hall Monitor sequence and up to 0.9 dB improvement for Foreman when compared with DISCOVER. However, the DRVC based systems in [10] still have high complexity at the encoder due to the considerable computational loads of DWT, LQR hash, SW-SPIHT and mode decision.

3 Proposed DRVC scheme

3.1 Architecture of the proposedscheme

In this section, the proposed DRVC scheme is described in Fig. 2. It has the following new features when compared to the basic DRVC described in Section 2.

1) At the encoder, a dead-zone quantizer with 3 levels is used to quantize residual frame R. The quantization process can be applied iteratively if high quality frames are required.
2) Two bit planes denoted as BP_1 and BP_2 are extracted from the quantized indexes that are binary presented.
3) A bit plane block based(BPBB) method is proposed to divide each bit plane into a number of 4×4 blocks and the blocks are classified as 0-Block or 1-Block. Two bit planes, which include the block information and are denoted as BP_1^{block} and BP_2^{block} respectively, are fed into LDPCA encoder.
4) A bit plane re-arrangement(BPRA) scheme is proposed to remove the bits that are not required to be channel encoded. The re-arranged bit planes denoted as BP_1^{ra} and BP_2^{ra} respectively are fed into LDPCA encoder.

   The LDPCA encoder in 3) is the same as that in 4), and only one LDPCA encoder is required at the encoder. But for sake of clarification, two LDPCA encoders are presented in Fig. 2.
5) At the decoder, R' is processed in the same way as R at the encoder and the bit planes are decoded one by one. The decoded BP_1^{block} and BP_2^{block} are used to refine decoder SI, so a simple bit plane refinement (BP refine) is proposed. After LDPCA decoding,

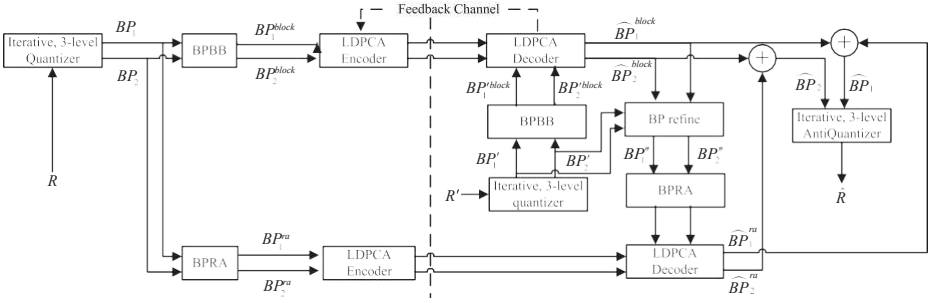


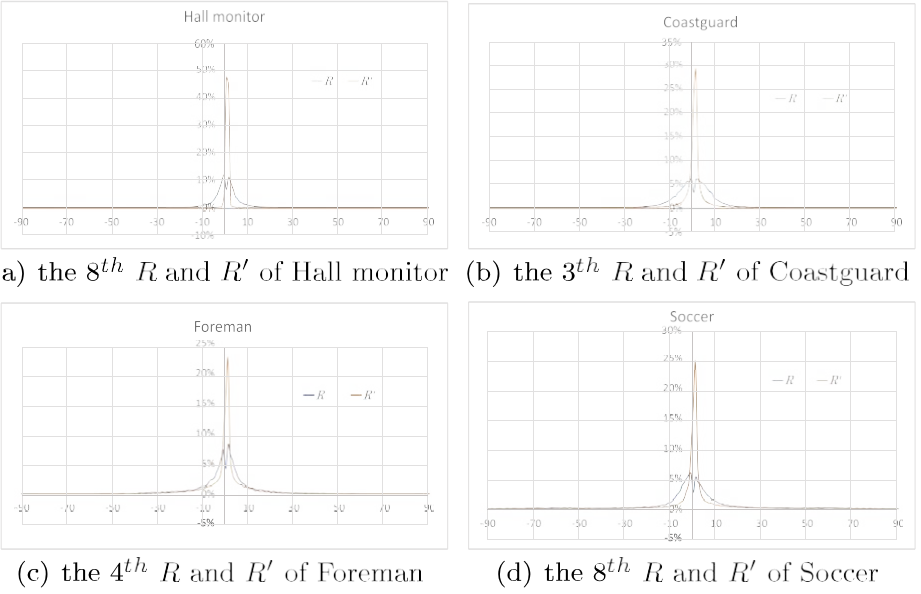
Fig. 2 Proposed DRVC architecture

the decoded $BP_i^{block}(i = 1, 2)$ and $BP_i^{ra}(i = 1, 2)$ denoted as $\hat{BP}_i^{block}$ and $\hat{BP}_i^{ra}$ respectively, are combined to obtain the decoded $BP_i(i = 1, 2)$, namely $\hat{BP}_i$. Finally, the inverse quantization [29] is implemented to reconstruct the residual frames $\hat{R}$.

3.2 Analysis of the distribution of residual pixels

The motivation of the proposed DRVC framework is based on the analysis of the statistical distribution of residual pixels in R and R' frames. Figure 3 illustrates the probability distribution curves of residual pixels in some of R and R' frames extracted from Hall Monitor, Foreman, Coastguard, and Soccer video sequences. It can be seen that both curves are steep around 0, indicating that most of the residual pixels have small magnitudes. Furthermore, the distribution curve of residual pixels in R' is sharper than that in R , that is, R' has more residual pixels concentrated on 0. This is because R' can be regarded as the motion-compensated errors from (1), (2) and (3). Equation (1) is used to calculate W_{re} , which is the average interpolation of the previous and next decoded key frames (namely $\hat{K}_{pre}$ and $\hat{K}_{nex}$). Equation (2) is used to calculate Y where $mv = (mv_x, mv_y)$ is the estimated motion vector. Equation (3) is used to calculate R'. From the three equations, (4) is derived to denote R', where $\hat{K}_{pre}(x + mv_x, y + mv_y) - \hat{K}_{pre}(x, y)$ and $\hat{K}_{nex}(x - mv_x, y - mv_y) - \hat{K}_{nex}(x, y)$ are called motion-compensated error of the previous and next decoded key frame respectively. Since there is little change of background and foreground information between one frame and its compensated reference frame, the motion-compensated errors tend to be small. This results in that the majority of the cases are $R' = Y - W_{re} \approx 0$.

Based on the statistical distributions of residual pixels analyzed above, the key techniques in the proposed DRVC scheme are designed and detailed in Sections 3.3, 3.4, 3.5 and 3.6.



(a) the 8^{th} R and R' of Hall monitor (b) the 3^{th} R and R' of Coastguard

(c) the 4^{th} R and R' of Foreman (d) the 8^{th} R and R' of Soccer

Fig. 3 Probability distribution curves of residual pixels in some of R and R'

For simplification, the probability distribution of residual pixels in R and R' are denoted as the distribution of R and R' respectively.

$$W_{re} = \frac{1}{2}(\hat{K}_{pre}(x, y) + \hat{K}_{nex}(x, y)) \tag{1}$$

$$Y = \frac{1}{2}[\hat{K}_{pre}(x + mv_x, y + mv_y) + \hat{K}_{nex}(x - mv_x, y - mv_y)] \tag{2}$$

$$R' = Y - W_{re} \tag{3}$$

$$\begin{aligned} R' &= \frac{1}{2}[\hat{K}_{pre}(x + mv_x, y + mv_y) + \hat{K}_{nex}(x - mv_x, y - mv_y)] \\ &\quad - \frac{1}{2}(\hat{K}_{pre}(x, y) + \hat{K}_{nex}(x, y)) \\ &= \frac{1}{2}\{[\hat{K}_{pre}(x + mv_x, y + mv_y) - \hat{K}_{pre}(x, y)] \\ &\quad + [\hat{K}_{nex}(x - mv_x, y - mv_y)] - \hat{K}_{nex}(x, y)]\} \end{aligned} \tag{4}$$

3.3 Iterative, 3-level quantizer

Quantization has a significant impact on the coding performance of DRVC because it is performed on the original residual pixel values directly. If the distributions of R and R' in one quantization partition mismatch too much, the bit error probability between R and R' will be large and more parity bits will be used. Therefore, coarse quantization can save bitrates. This is why a 3-level quantizer is proposed. Since the values of residual pixels range from -255 to 255, a dead zone quantizer is used and the quantization partition bin is designed as $[-255 \quad -l \quad l \quad 255]$. Comparing the distribution curves of R and R' in Fig. 3, it can be expected that the larger the quantized interval $2l$ is, the less the mismatches between R and R' are. Thus, more parity bits can be saved. Although coarse quantization saves bitrates, it leads to an increase of distortions in reconstructed frames. In order to obtain high quality frames, the quantization process can work iteratively. The quantized interval $2l$ is narrowed down at each iteration to help with the quantization from coarse to fine. Furthermore, the reconstructed frame at last iteration can be used as the refined SI for the current frame, and thereby this can improve the coding efficiency.

3.4 Bit plane block based (BPBB) module

Given that the quantization partition bin is $[-255 \quad -l \quad l \quad 255]$ and the corresponding quantized indexes are binary presented by 10, 00 and 01 respectively, two bit planes denoted as BP_1 and BP_2 are extracted. The proposed BPBB module divides each bit plane into non-overlapping 4×4 blocks and each block is classified as 0-Block or 1-Block according to a simple criterion, that is, a block that has 16 bits of 0 is 0-Block and otherwise it is defined as 1-Block. There are only two types of blocks so only one bit is used to encode 0-Block as 0 and 1-Block as 1. Then two new bit planes, BP_1^{block} and BP_2^{block} that indicate the block type information, are obtained and fed into LDPCA encoder. An explicit description of the BPBB process is demonstrated in Fig. 4b.

3.5 Bit plane re-arrangement (BPRA) module

After BPBB process, there are four types of combinations for blocks in BP_1 and their co-located blocks in BP_2, that is, (0-Block, 0-Block), (0-Block,1-Block), (1-Block,0-Block)

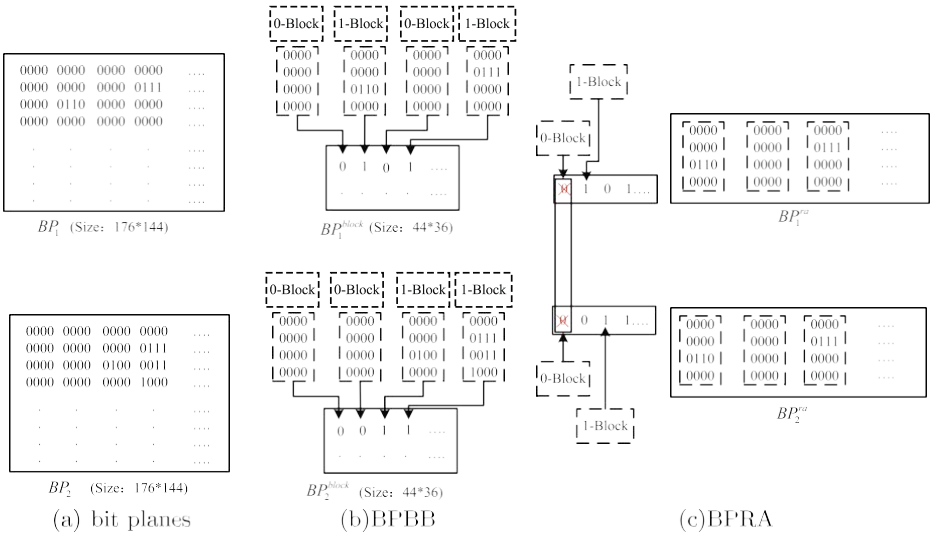


Fig. 4 Process of BPBB and BPRA

and (1-Block,1-Block). The BPRA module is designed to remove the bits belonging to the combination of (0-Block, 0-Block) in both BP_1 and BP_2 . These bits of 0 do not need to be channel encoded and transmitted. This is because the block information will be known at the decoder side after BP_1^{block} and BP_2^{block} are decoded correctly, and the removed bits in (0-Block, 0-Block) can be recovered by 0s. After BPRA, the re-arranged bit planes, namely BP_1^{ra} and BP_2^{ra}, are fed into LDPCA encoder. The process of the BPRA is shown in Fig. 4c.

3.6 BP refinement

When BP_1^{block} and BP_2^{block} are decoded correctly, the block information is known at the decoder. Using 0-blocks in BP_1^{block} and BP_2^{block} can help refining BP_1' and BP_2', which are the bit planes extracted from R'. If one block is 0-Block, its corresponding 16 bits in BP_1' or BP_2' can be corrected by 0s. Then the refined BP_1' and BP_2', denoted as BP_1'' and BP_2'' are used to help decoding $BP_1'^{a}$ and $BP_2'^{ra}$ respectively.

4 Experiments and analyses

In order to evaluate the performance of the proposed DRVC scheme, an extensive simulation has been carried out on four test video sequences, namely Hall Monitor, Foreman, Coastguard, and Soccer with QCIF resolution and the frame rate of 15Hz (i.e., 149 frames for Soccer, Foreman, Coastguard respectively, and 165 frames for Hall Monitor). The GOP of 2 is used which is mostly adopted in literatures. Odd frames are KEY frames encoded by H.264/AVC Intra mode with QP parameter equal to 20, 25, 27, 29, 30, and 34 respectively. Even frames are WZ frames used to obtain residual frames. The value of parameter l and the number of iterations of the 3-level dead-zone quantizer are selected so that the average

Table 1 Conditional probability matrix between BP_1 and $BP_1^{\downarrow}$

x \ y	0	1
0	0.9999	0.0769
1	0.0001	0.9231

quality of WZ frames is similar to that of Key frames. For lower frame quality, the quantization is implemented once and l is chosen as 96. For higher frame quality, the quantization is implemented twice and the corresponding values of l are 96 and 64 respectively. The thresholds of 96 and 64 are obtained empirically. For LDPCA codec [21], LDPCA with 396 nodes is adopted in the proposed work. In the case of the bitstream length less than 396, the padding zeros are applied.

4.1 Efficiency of BPBB

The compression efficiency of DVC depends on the correlation between source and SI strongly. The higher the correlation is, the lower the compressed bitrates are. The proposed DRVC scheme is based on the bit plane level and the bit plane correlation between source and SI can be measured by the binary conditional probability $P(X|Y)$. X and Y are binary source and there are four cases of the conditional probabilities, namely $P(X=0|Y=0)$, $P(X=1|Y=0)$, $P(X=0|Y=1)$ and $P(X=1|Y=1)$. The sum of $P(X=0|Y=0)$ and $P(X=1|Y=1)$, denoted as SoP, is used to measure the correlation between the bit plane at the encoder and the bit plane at the decoder. The bigger the sum is, the higher the dependency is. For example, Table 1 shows a matrix of $P(X|Y)$ that are the binary conditional probabilities between BP_1 and $BP_1^{\downarrow}$. Table 2 shows another matrix of $P(X|Y)$ that are the binary conditional probabilities between BP_2 and $BP_2^{\downarrow}$. The SoPs in Tables 1 and 2 are 1.923 (0.9999+0.9231) and 1.8887 (0.9998+0.8889), respectively. It clearly demonstrates that the correction between BP_1 and $BP_1^{\downarrow}$ is better than that between BP_2 and $BP_2^{\downarrow}$, and therefore the number of bits sent for decoding BP_1 is less than that for decoding BP_2.

The proposed BPBB method can improve the dependency between source and SI. In order to test the hypothesis, the correlations between $BP_i(i=1,2)$ and $BP_i^{\downarrow}(i=1,2)$ and the correlations between $BP_i^{block}(i=1,2)$ and $BP_i^{\downarrow block}(i=1,2)$ of all the residual frames for four test video sequences are calculated. These are the correlations before and after using BPBB. Figure 5 illustrates the comparisons of these SoPs. Since the quantization can be applied up to twice depending on the frame quality, the comparisons of SoPs with different values of l are presented. It is clear that most points are above the diagonal line regardless of the values of l, indicating that the correlation between $BP_i^{block}(i=1,2)$ and $BP_i^{\downarrow block}(i=1,2)$ are commonly higher than that between $BP_i(i=1,2)$ and $BP_i^{\downarrow}(i=1,2)$. Figure 5 demonstrates that the dependency between source and SI have been improved by using BPBB method. Furthermore, the size of $BP_i^{block}(i=1,2)$ is only one sixteenth of the size of $BP_i(i=1,2)$. Due to the increase of the correlation and the

Table 2 Conditional probability matrix between BP_2 and $BP_2^{\downarrow}$

x \ y	0	1
0	0.9998	0.1111
1	0.0002	0.8889

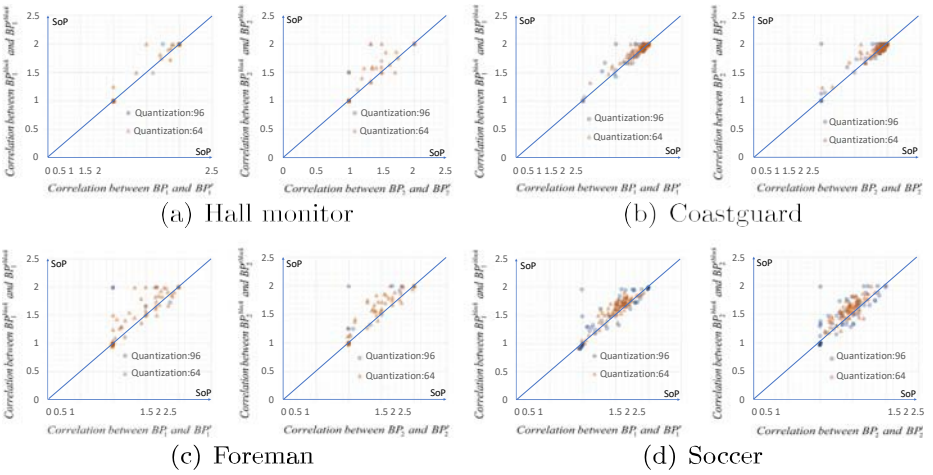


Fig. 5 Comparisons of SoPs for video sequences with different values of l

decrease of the size, it can be predicted that the bitrates send for decoding BP_i^{block} will be reduced significantly, This demonstrates that the proposed BPBB is very effective in bitrates reduction. In addition, Fig. 5 also reveals that the efficiency of BPBB correlates to the numbers of the points which are under the diagonal line. The larger the number of the points under the diagonal line is, the lower the efficiency is. Among the four videos, the efficiency of BPBB for Hall Monitor is the highest while that for Soccer is the lowest. This is because Hall Monitor is a low motion sequence and has many 0-Blocks at both encoder and decoder sides after using BPBB, and vice versa for Soccer sequence.

4.2 Efficiency of BPRA

As described in Section 3, BPRA is used to remove the bits of (0-Block, 0-Block) combination in both BP_1 and BP_2 and thus to reduce the amount of data to be channel encoded. The percentage of the removed bits is calculated by using (5).

$$Rduce = \frac{[Number\ of\ (0-Block, 0-Block)] \times 16}{176 \times 144} \times 100\% \tag{5}$$

Figure 6 shows the average reducing percentages of all R for Hall monitor, Coastguard, Foreman and Soccer with different values of l. It can be seen that the percentages range from 85% to nearly 100%. This indicates that a great number of bits have been saved, and therefore the BPBA scheme is effective. Furthermore, the percentages of the removed bits with l=64 are less than that with l=96. It is because the number of 0-Blocks at the encoder with l=64 is less than that with l=96, the details of which are analyzed in Section 4.4.

4.3 Efficiency of BP refinement

0-Blocks are obtained from the decoded $BP_i^{block}(i\ \ 1,\ 2)$ and then are used to refine $BP'_i(i\ \ 1,\ 2)$ at the decoder by setting the corresponding 16 bits with 0s. In order to evaluate the efficiency of the BP refinement, the relative improvement ratio of the bit error probability of each bit plane is defined by (6) where ρ_i' and ρ_i'' are the bit error probability

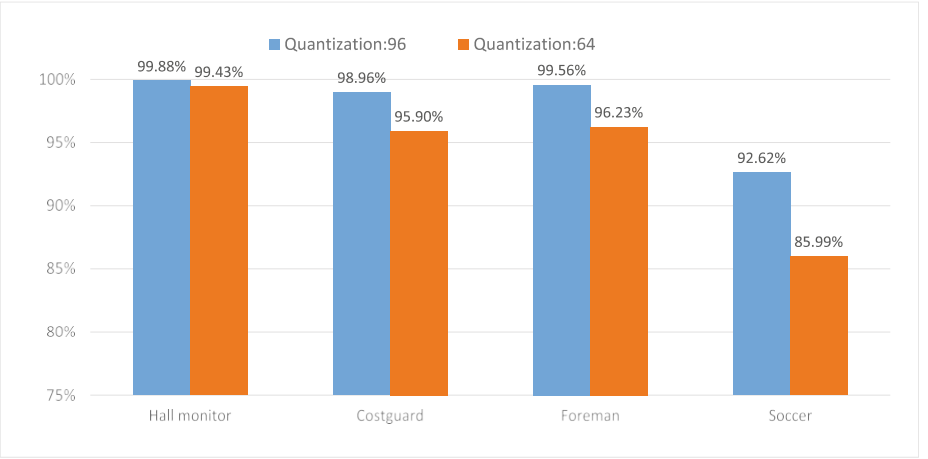


Fig. 6 Average reducing percentages for video sequences with different values of l

of the $i^{th}(i = 1, 2)$ bit plane before and after using BP refinement. ρ_i^{I} and ρ_i^{II} are obtained by (7) and (8) respectively, where $b_{i,j}$,$b_{i,j}^{\text{I}}$ and $b_{i,j}^{\text{II}}(i = 1, 2; j = 1 \cdots 176 \times 144)$ denote the binary bits in BP_i, BP_i^{I}, BP_i^{II} respectively and the symbol $\oplus$ denotes the binary XOR operator.

$$\rho_i = \left(\frac{\rho_i^{\text{I}} - \rho_i^{\text{II}}}{\rho_i^{\text{I}}} \right) \times 100\% \tag{6}$$

$$\rho_i^{\text{I}} = \frac{(b_{i,j} \oplus b_{i,j}^{\text{I}})}{176 \times 144} \times 100\% \tag{7}$$

$$\rho_i^{\text{II}} = \frac{(b_{i,j} \oplus b_{i,j}^{\text{II}})}{176 \times 144} \times 100\% \tag{8}$$

Figure 7 shows the average ρ_i of all the residual frames for Hall monitor, Coastguard, Foreman and Soccer. It can be seen that the maximum ratio is only 12% and the efficiency is not significant. That is because the principle of BP refinement is using the 0-Blocks at encoder to rectify the co-located 1-Blcok at decoder. As analyzed in Section 4.4, the number of 1-Blocks at the decoder is relatively low, and the number of 1-Blocks that need to be corrected is lower. Therefore, the refinement is limited.

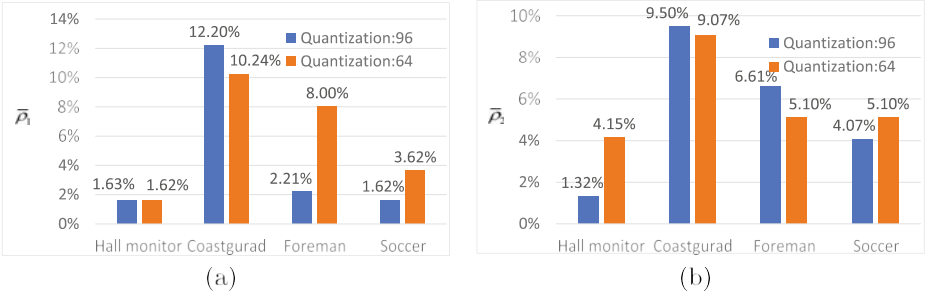


Fig. 7 Relative improving ratio of the bit error probability of **a** the first plane **b** the second plane

4.4 Analysis of the iterative quantization

In this work, the quantization process is carried out iteratively according to the reconstructed frame quality. For the low frame quality, the quantization is applied once and the corresponding l is 96. For the high frame quality, the quantization is applied twice and the according values of l are 96 and 64 for the first and second iteration respectively. Since the residual pixels within the interval (-ll) are encoded as 00, it can be drawn that more 0-Blocks are obtained at both sides if more residual pixels fall into this interval after quantization and BPBB, and there are certainly more 0-Blocks for l=96 than that for l=64. Furthermore, based on the specific distribution of R and $R^{\prime}$ in Fig. 3, it is clear that $R^{\prime}$ has more 0-Blocks than R. This is because more residual pixels in $R^{\prime}$ are concentrated around 0, indicating more residual pixels falling into the interval (-l l). Figure 8 shows the average percentages of 0-Blocks in R and that in $R^{\prime}$ for the four videos sequences with different value of l, which demonstrates the above analyses. From Fig. 8, the highest percentage is up to 100% for Hall Monitor and the lowest percentage is 92.51% for Soccer.

As discussed in Section 3.3, if the distribution of R and $R^{\prime}$ in one quantization partition mismatch too much, the correlation between source and SI is weak and the coding efficiency is low. With respect to the proposed scheme, since the 0-Blocks are in majority at both encoder and decoder sides, the mismatches in one quantization partition can be treated as the difference of the number of 0-Blocks at both sides. The bigger the difference is, the weaker the correlation between the source and SI is. Figure 8 shows that the difference becomes bigger when l changes from 96 to 64 for all four test sequences and thus the correlation becomes weaker. For example, for Soccer sequence the difference of the number of 0-Blocks at both sides changes from 2.29% (99.76%-97.47%) to 3.83% (96.34%-92.51%) when l changes from 96 to 64. Therefore, it can be concluded that reducing l can help with quantization from coarse to fine but will reduce the efficiency of the proposed scheme at the same time.

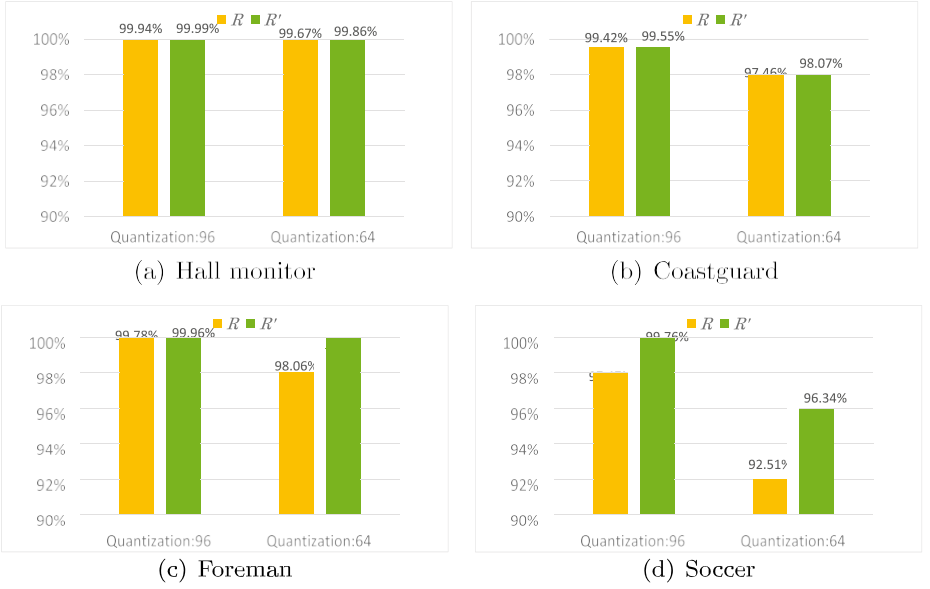


Fig. 8 Average percentages of 0-Blocks in R and $R^{\prime}$ for video sequences with different values of l

4.5 RD performance

Figure 9 compares RD performance of the proposed DRVC scheme, DISCOVER and our previous work [19]. Only luminance component is considered in the calculation of the RD performance. DISCOVER has become a benchmark for DVC research due to its good performance. The simulation results of DISCOVER are obtained from [16]. Our previous work [19] proposed a relatively simple encoder,which outperforms other DRVC systems in terms of encoder complexity [19].

1) Comparation with DISCOVER. From Fig. 9, it can be seen that the proposed DRVC scheme performs better than DISCOVER for Hall Monitor and Coastguard, and obtains up to 1.5dB gains for Hall Monitor sequence. Figure 9 also shows RD performance gap for Foreman at high bitrate and Soccer at all bitrate range. This is because the coding efficiency of the proposed scheme is based on the difference of the number of 0-Blocks at both encoder and decoder sides. For Hall Monitor, as shown in Fig. 8, the differences are 0.05%(l=96) and 0.19%(l=64), the lowest differences among all the test sequences. Thereby the efficiency of BPBB and BPAR for Hall Monitor are the highest, which can be seen from Figs. 5 and 6. Soccer has the most complex motion in all test sequences, and the differences are 2.29%(l=96) and 3.83%(l=64) , the highest among all test sequences. Thereby, the efficiency of BPBB and BPRA for Soccer is the lowest. The differences for Coastguard (0.13% and 0.59%) are smaller than that for Foreman (0.18% and 1.3%), so the performance of Coastguard is better than that of Foreman. With regard to the encoder complexity, DISCOVER is higher than the proposed work. DISCOVER has introduced DCT and several new modules at the encoder to enhance the overall performance, such as the adaptive GOP selection using a hierarchical clustering algorithm, the encoder rate-control mechanisms computing the minimum rate

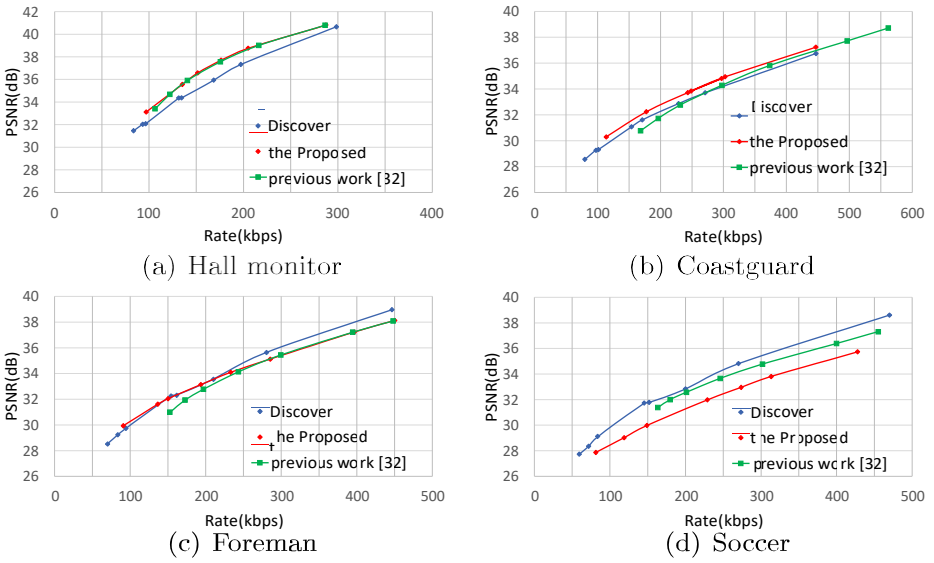


Fig. 9 RD performance of video sequences

for each bit plane of each coefficient band, and the CRC sum of the encoded bit plane. While in the proposed DRVC scheme, only two simple modules named BPBB and BPRA are employed at encoder and there are no heavy computations and complex process. Furthermore, the proposed DRVC scheme performs well for sequences with low motion, which is well suitable for wireless video surveillance applications where typical video contents have low motion scenes.

2) Comparation with our previous work [19]. In order to improve the coding efficiency, we presented an efficient encoder rate control solution combined with an encoder block mode decision for DRVC in [19]. The results in [19] show our previous work outperforms DISCOVER and the state-of-the-art rate control solution in terms of RD performance for video sequences with low motion. It also has competitive RD performance for other types of video sequences. From Fig. 9, the proposed DRVC scheme is better than [19] for Coastguard, similar to [19] for Hall Monitor and better than [19] at low bitrates for Forman. It can be concluded that the proposed work provides a RD performance quite close to [19] except for Soccer. In [19], there are three modules at the encoder side, that is, block mode decision module, scrambling module and rate control module. While the proposed DRVC scheme has two simple modules, BPBB and BPRA, at the encoder. [19] requires low level of computations while the computational requirement in the proposed DRVC scheme is negligible.

5 Conclusion

Low complexity encoder is most widely cited advantage of DVC. In order to improve the coding efficiency while maintaining a low complex encoder, a simple encoder scheme for DRVC system is proposed in this paper. There are two modules employed at the encoder. One is BPBB that divides each bit plane into blocks and defined them as 0-Blocks or 1-Blocks according to a simple criterion. The other is BPRA that removes bits of the (0-Block, 0-Block) combination. The coding efficiency of the proposed scheme is enhanced based on the dependency improved by BPBB and the amount of data reduced by BPRA. In addition, there is a 3-level quantizer that is implemented at encoder iteratively. If high quality frame is required, the quantized interval is narrowed down and the quantization process repeats. The preceding analyses and experimental results show that there are no heavy computations and complex process at the encoder. The RD performance of the proposed scheme outperforms that of DISCOVER for low to medium motion video sequences, and it is also similar to our previous work [19] except for the sequences with highly irregular motion.

Acknowledgment The research activities that have been described in this paper were funded by China Scholarship Council (CSC NO. 201708440523) and National Natural Science Foundation of China(NSFC NO. 51978271).

References

1. Aaron A, Girod B (2004) Wyner-Ziv video coding with low encoder complexity. Picture Coding Symposium, pp 429–433
2. Aaron A, Rane S, Girod B (2004) Wyner-Ziv video coding with hash-based motion compensation at the receiver. In: 2004 international conference on image processing, pp 3097–3100
3. Aaron A, Rane S, Setton E, Girod B (2004) Transform-domain Wyner-Ziv codec for video. In: Visual communications and image processing 2004, pp 520–528

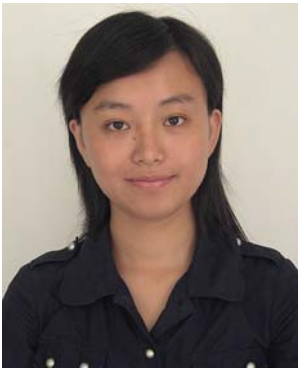
4. Aaron A, Rui Z, Girod B (2002) Wyner-Ziv coding of motion video. In: Conference record of the 36th Asilomar conference on signals, systems and computers, pp 240–244
5. Aaron A, Varodayan DP, Girod B (2006) Wyner-Ziv residual coding of video. In: Picture coding symposium 2006, pp 28–32
6. Abou-Elailah A, Dufaux F, Farah J, Cagnazzo M, Pesquet-Popescu B (2013) Fusion of global and local motion estimation for distributed video coding. IEEE Trans Circ Syst Vid Technol 23(1):158–172
7. Ascenso J, Brites C, Pereira F (2005) Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding. In: 5th EURASIP conference on speech and image processing, multimedia communications and services, pp 1–6
8. Ascenso J, Pereira F (2007) Adaptive hash based side information exploitation for efficient Wyner-Ziv video coding. In: 2007 IEEE international conference on image processing, pp. 29–32
9. Ascenso J, Pereira F (2009) Low complexity intra mode selection for efficient distributed video coding. In: 2009 IEEE international conference on multimedia and expo, pp 101–104
10. Bai H, Wang A, Zhao Y, Pan JS, Abraham A (2011) Distributed multiple description coding: Principles, algorithms and systems. In: Chapter 5:Algorithms of DVC. Springer Publishing Company, New York, pp 128–142
11. Chiang JC, Chen KL, Chou CJ, Lee CM, Lie WN (2010) Block-based distributed video coding with variable block modes. In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems, pp 125–128
12. Chien WJ, Karam LJ (2009) Transform-domain distributed video coding with rate-distortion-based adaptive quantisation. IET Image Process 3(6):340–354
13. Clerckx T, Munteanu A, Cornelis J, Schelkens P (2007) Distributed video coding with shared encoder/decoder complexity. In: 2007 IEEE international conference on image processing, pp 417–420
14. Deligiannis N, Munteanu A, Clerckx T, Cornelis J, Schelkens P (2009) Overlapped block motion estimation and probabilistic compensation with application in distributed video coding. IEEE Signal Process Lett 16(9):743–746
15. Deligiannis N, Verbist F, Iossifides AC, Slowack J, Van de Walle R, Schelkens P, Munteanu A (2012) Wyner-Ziv video coding for wireless lightweight multimedia applications. EURASIP J Wirel Commun Netw 106:1–20
16. The DISCOVER EU-project (2005) DISCOVER-distributed coding for video services. http://www.discoverdvc.org
17. HoangVan X, Jeon B (2012) Flexible complexity control solution for transform domain Wyner-Ziv video coding. IEEE Trans Broadcast 58(2):209–220
18. Hu CY, Hu BJ (2016) Encoder rate control algorithm based on scrambling with pseudo-random code for distributed residual coding of video. Acta Electronica Sinica 44(6):1490–1495. (in Chinese)
19. Hu CY, Hu BJ, Tu WQ, Xiong YH (2018) A low-complexity and efficient encoder rate control solution for distributed residual video coding. Multimed Tools Appl 77(5):5713–5735
20. Jia Y, Wang Y, Song R, Li J (2015) Decoder side information generation techniques in Wyner-Ziv video coding: A review. Multimed Tools Appl 74(6):1777–1803
21. Kubasov D, Nayak J, Guillemot C (2007) Optimal reconstruction in Wyner-Ziv video coding with multiple side information. In: IEEE 9th workshop on multimedia signal processing, pp 183–186
22. Lee CM, Chiang ZH, Tsai DC, Lie WN (2013) Distributed video coding with block mode decision to reduce temporal flickering. EURASIP J Adv Signal Process 177:1–13
23. Li Y, Zhao D, Ma S, Gao W (2009) Distributed video coding based on the human visual system. IEEE Signal Process Lett 16(11):985–988
24. Liu L, He Dk, 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, pp 1136–1139
25. Ma T, Hempel M, Peng D, Sharif H (2013) A survey of energy-efficient compression and communication techniques for multimedia in resource constrained systems. IEEE Commun Surv Tutor 15(3):963–972
26. Puri R, Majumdar A, Ramchandran K (2007) PRISM: A video coding paradigm with motion estimation at the decoder. IEEE Trans Image Process 16(10):2436–2448
27. Slepian D, Wolf J (1973) Noiseless coding of correlated information sources. IEEE Trans Inf Theory 19(4):471–480
28. Sofke S, Pereira F, Muller E (2009) Dynamic quality control for transform domain wyner-ziv video coding. EURASIP J Image Video Processing 978581:1–15
29. Varodayan D, Aaron A, Girod B (2006) Rate-adaptive codes for distributed source coding. Signal Process 86(11):3123–3130
30. Verbist F, Deligiannis N, Jacobs M, Barbarien J, Schelkens P, Munteanu A, Cornelis J (2013) Probabilistic motion-compensated prediction in distributed video coding. Multimed Tools Appl 66(3):405–430

31. Wang S, Rehman A, Wang Z, Ma S, Gao W (2013) Perceptual video coding based on SSIM-inspired divisive normalization. IEEE Trans Image Process 22(4):1418–1429
32. Wang Y, Wu C (2010) A block based Wyner-Ziv video codec. In: 2010 3rd international congress on image and signal processing, pp 1–5
33. Wu B, Zhang N, Ma S, Zhao D, Gao W (2014) Optimal entropy-constrained non-uniform scalar quantizer design for low bit-rate pixel domain DVC. Multimed Tools Appl 70(3):1799–1824
34. Wyner AD, Ziv J (1976) The rate-distortion function for source coding with side information at the decoder. IEEE Trans Inf Theory 22(1):1–10
35. Zhang L, Peng Q, Wu X (2017) Perception-based adaptive quantization for transform-domain Wyner-Ziv video coding. Multimed Tools Appl 76(15):16699–16725

Chunyun Hu was born in Jiangxi, China. She received the B.Sc. from Nanchang University, Jiangxi, China and her M.Sc and Ph.D. from South China University of Technology, Guangzhou, China. Currently, she is a lecturer in college of Electronic Engineering, South China Agricultural University. From January 2018 to January 2019, she was an academic visitor in Robert Gordon University. Her research interests include image and video processing, distributed source coding, and routing problem.

Yafan Zhao received the B.Sc. from Nanjing University of Science and Technology, China and her M.Sc. and Ph.D. from The Robert Gordon University U.K. Currently, she is a senior lecturer in the School of Engineering, The Robert Gordon University. Dr. Zhao is the author of over 30 research articles in journals, conferences, industry standard contributions and patent applications. She is also an experienced consultant in standard-based codec design and implementations and a patent expert. Her areas of expertise include video compression and standards, perceptual visual quality, computervision and multimedia communication.

Long Yu was born in Anhui, China. He received the Ph.D. degree from South China Agriculture University, Guangzhou, China, in 2010. In 2013, he was a visiting scholar at Washington State University in the United States for one year. His research interests include image processing and compress, embedded system develop.



Yang Jiang is a lecturer and PhD supervisor in computing science and digital media in Robert Gordon University. Her major research and teaching experience are on Computer Graphics, computer animation and gaming, 3D visualisation and immersive technologies (including virtual, augmented and mixed realities). As a key researcher in RGU User Experience Lab, in collaboration with artists and creative media industry, she has explored research on 3D visualisations and immersive technologies to: (i) create interesting animation characters and scenes with new mathematics models; (ii) reconstruct events from videos to enrich computer animations; (iii) 3D visualisations for GIS; and (iv) new game and user-centred design in HCI.

Yunhui Xiong an Associate Professor in 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. His research interests are image and video processing, distributed source coding, computer graphics, geometry modeling and processing, 3D reconstruction and 3D printing.

Affiliations

Chunyun Hu[1] · Yafan Zhao[2] · Long Yu[1] · Yang Jiang[3] · Yunhui Xiong[4]

Chunyun Hu
hcy2182@scau.edu.cn

Yafan Zhao
y.zhao@rgu.ac.uk

Long Yu
yulong@scau.edu.cn

Yang Jiang
y.jiang2@rgu.ac.uk

[1] College of Electronic Engineering, South China Agricultural University, Guangzhou, China

[2] School of Engineering, Robert Gordon University, Aberdeen, UK

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

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