

RESEARCH

Open Access

Distributed video coding supporting hierarchical GOP structures with transmitted motion vectors

Kyung-Yeon Min¹, Woong Lim¹, Junghak Nam¹, Donggyu Sim^{1*} and Ivan V Bajić²

Abstract

In this paper, we propose a new distributed video coding (DVC) method, with hierarchical group of picture (GOP) structure. Coding gain of DVC can be significantly improved by enlarging GOP size for slow-moving frames. The proposed DVC decoder estimates a side information (SI) frame and transmits motion vectors (MVs) of the SI to the proposed encoder. Using the received MVs from the decoder, the proposed encoder can generate a predicted SI (PSI), which is the same as the SI in the decoder, and estimate the quality of PSI with minimal computational complexity. The proposed method decides the best coding mode among key, Wyner-Ziv (WZ), and skip modes, by estimating rate-distortion costs. Based on the selected best coding mode, the best GOP size can be automatically determined. As the GOP size is adaptively decided depending on the SI quality, entropy and parity bits can be effectively consumed. Experimental results show that the proposed algorithm is around 0.80 dB better in Bjøntegaard delta (BD) bitrate than an existing conventional DVC system.

Keywords: Video coding; Distributed video coding; Side information; GOP; RD competition

1 Introduction

As many portable multimedia devices have been developed, such as mobile phones, electronic pads, and laptops, many people enjoy using them to take videos, transmit them to friends or web-sites such as YouTube and Facebook, and in turn, view them. These days, video sensor networks are also used to monitor very large outdoor areas for environment surveillance and safety. Therefore, demands for low cost and powerful encoders are continuously increasing. However, conventional video coding standards, such as MPEG-x and H.26x, cannot satisfy these requirements, because those encoders have high computational complexity, while their decoders require low complexity. Distributed video coding (DVC) methods have been researched to meet these requirements. DVC technology is based on migration of computational complexity from encoders to decoders and can achieve coding gain with regard to prediction on the decoder side.

DVC was developed as a new video-coding paradigm derived from Slepian-Wolf information theory [1]. They proved that the DVC can perform encoding by disregarding

correlation between two input signals and the coding performance of the decoder side by exploiting the correlation can come close to the efficiency of the conventional coding systems that employs the correlation at the encoder side. Wyner-Ziv [2] presented the extended work to show information theoretic bounds for lossy compression by side information at the decoder. Based on the Wyner-Ziv theory, several lossy DVC approaches which do not perform motion estimation have been proposed in order to reduce computational complexity of the DVC encoder [3-6]. To reduce temporal redundancy, motion estimation is performed on the DVC decoder side, not in the encoder. For DVC based on the Wyner-Ziv approach, the original input frames are coded by two different modes [3-6]. One mode is to code with the conventional intra coding technique and the coding mode is called the key-frame mode. The other mode is performed by a channel coder after pre-processing, and the coding mode is called the Wyner-Ziv (WZ) mode. While the outputs of the channel coder are parity bits and original data, only a part of parity bits are sent to the DVC decoder for compression performance. The reconstructed WZ frame is reconstructed by a channel decoder with the transmitted parity bits for a side information (SI) frame. The SI frame is generated the same as possible as the original frame with the reconstructed

* Correspondence: dgsim@kw.ac.kr

¹Department of Computer Engineering, Kwangwoon University, Wolgye-dong, Seoul 447-1, Nowon-gu, South Korea

Full list of author information is available at the end of the article

key frames in the decoder when the size of the group of picture (GOP) is small, e.g., the size is equal to 2. The error of SI frame is assumed to be transmission error caused by a variable channel. SI frame is regarded as the predicted frame of the original frame (WZ frame), degraded by channel errors. Therefore, the errors are corrected by a channel decoder. A low density parity check accumulate (LDPCA) coder and Turbo coder are often used for DVC systems [7-12].

In general, conventional codecs, such as h.26x and MPEG-x, set GOP size from 8 to 30, as the increasing number of intra-frame degrades compression rate. However, a lot of the conventional DVC systems set the GOP size to the minimum of two, because performance of DVC is directly related to accuracies of SI frames. Accuracies of SI frames cannot be known at both the encoder and decoder sides, and accuracies of SI frames are generally the best with GOP size set as two. In addition, since accuracies of SI frames vary, depending on the features of a sequence, they cannot be correctly predicted. However, SI frames are generated well for slow-moving cases. For these cases, GOP size can be prolonged, to reduce the entropy bits and/or parity bits. Therefore, some conventional DVC algorithms are proposed to predict accuracies of SI frames on the encoder side and increase GOP size [13,14]. Since the purpose of DVC is to reduce computational complexity of the encoder, they should generate a predicted SI (PSI) with low delay, though an SI frame which is generated using several motion estimation algorithms and filters, for high quality of an SI on the decoder side. Therefore, although the encoder generates a PSI frame by using the estimated MVs and key frames in its own, the PSI frame is not the same as the SI frame on the decoder side. As a result, the estimated accuracies of PSI frames are different from those of SI frames and lead to a decrease in compression performance.

The proposed DVC performs motion estimation at the decoder side, and the estimated motion vectors (MV) are transmitted to the corresponding DVC encoder. The proposed encoder can generate a PSI that is identical to the SI frame of the decoder side with minimal computation load, because motion compensation is performed with the received MVs and reference key frames. Therefore, the proposed encoder can correctly estimate the quality of the SI frames. Based on the accuracies of the SI frames, the best coding mode is selected based on rate-distortion (RD) optimization; thus, the GOP size can be adaptively and hierarchically set. In this paper, each frame is coded as one among key, WZ, and skip modes. In order to assess the RD cost of the key mode with minimum computational complexity, the proposed method estimates it with a weighted linear interpolation of RD costs of neighboring key frames. Distortion of a

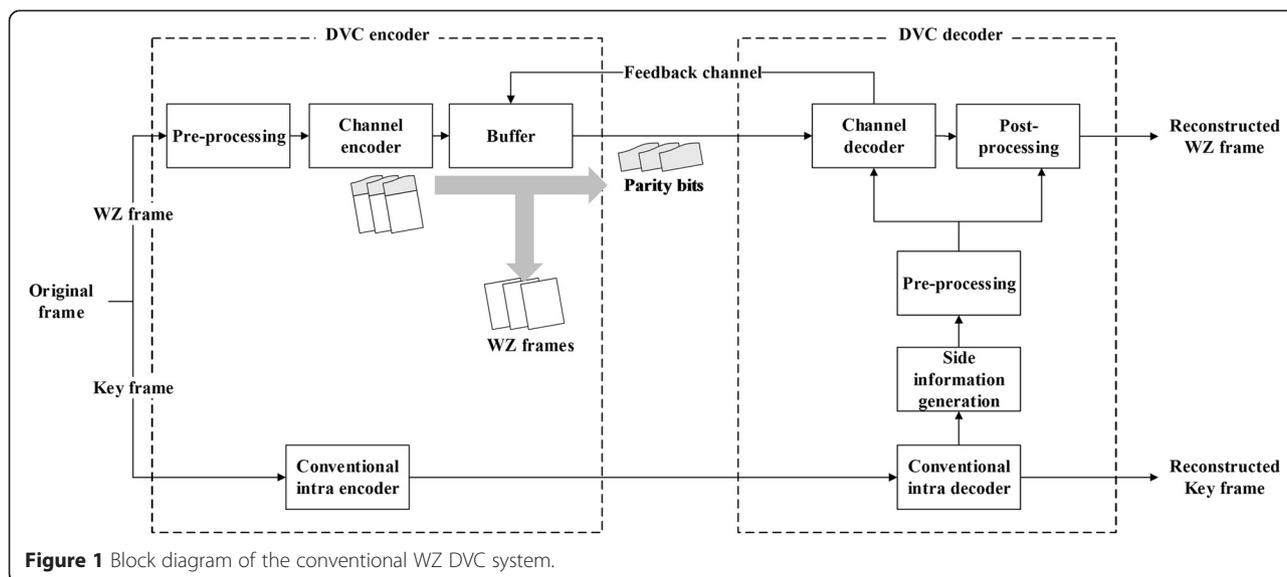
frame coded by WZ mode can be estimated with the original frame, and the PSI frame and rates of the frame can be estimated with the number of errors. Therefore, the proposed method assesses the RD cost of WZ mode with the compensated frame. The RD cost of skip mode can be estimated with the PSI frame on the encoder side. Based on RD competition, the proposed method can select the best coding mode and GOP size prior to actual encoding. The RD competition estimates rates and distortions for each coding modes in advance. Therefore, the proposed method improves coding performance by enlarging the GOP size for slow-moving frames. Note that the WZ frame is coded in frequency domain with LDPCA.

The rest of this paper is organized as follows. Section 'Conventional DVC algorithms' introduces several conventional DVC algorithms. Section 'Proposed DVC for hierarchical GOP structure' presents details of the proposed method. In Section 'Experimental results,' experimental results are given and discussed. Finally, Section 'Conclusions' concludes this paper and gives further work items.

2 Conventional DVC algorithms

DVC is a new video-coding paradigm that allows us to shift complexity from an encoder to a decoder, for distribution of computation complexity. While the conventional video codecs employ motion estimation at the encoder side, motion estimation can be performed on the decoder side of DVC. Therefore, the DVC encoder can be suitable for portable devices, unlikely conventional encoders. Figure 1 shows the block diagram of conventional WZ DVC systems [2-12,15-37]. Original input frames are divided into two types according to coding methods, as shown in Figure 1. Input frames are coded by key mode or WZ mode. Key mode is the same as the conventional intra coding method, such as H.264/AVC intra mode. WZ frames are coded with three main modules: pre-processing which can be regarded as transform and/or quantization, channel coding, and key frame coder. For DVC, motion estimation is conducted on the decoder side, and a predicted frame is generated. The estimated frame is called SI. The SI has some prediction errors, and the errors can be corrected with transmitted parity bits from a channel encoder. For the channel coder, LDPCA and Turbo channel coders are generally used in the DVC system [32-37].

The quality of SI frames directly impacts the performance of DVC systems, since the required number of parity bits is proportional to the errors of SIs. Most conventional DVC systems set their key frame interval to minimum, since it is much easier to estimate accurate SI frames with closer key frames. However, because the amount of motion activities varies over time, even in a sequence, the



quality of SI frames also varies from frame to frame, and sequence to sequence. Table 1 shows the average peak signal-to-noise ratio (PSNR) of SI frames in terms of the key frame intervals, for six sequences. To obtain the PSNR of SI, we employed the existing algorithm [23]. Figure 2 shows a PSNR graph of SI frames over time in terms of the key frame intervals for the ‘Race’ sequence. As shown in Figure 2, PSNR values of the 90th to 100th SI frames with GOP size of 8 are slightly better than those with GOP size of 4. Thus, the RD performance of DVC can be improved with the adaptive GOP size depending on frames characteristics. However, the encoder should estimate the quality of the SI, to determine the best size of GOP. It is not easy to estimate the quality of SI frames at the encoder side, due to non-availability of SI. Therefore, several conventional algorithms were proposed to assess the quality of SI frames [7-9,24-27]. Since the purpose of the DVC encoder is to encode videos with low computation complexity, the conventional algorithms generate PSIs with minimum computational complexity, by repetition, temporal linear interpolation, or rough block matching algorithms [7-9,24-27]. Ahmad et al. proposed a DVC supporting adaptive GOP size, so that the GOP size is

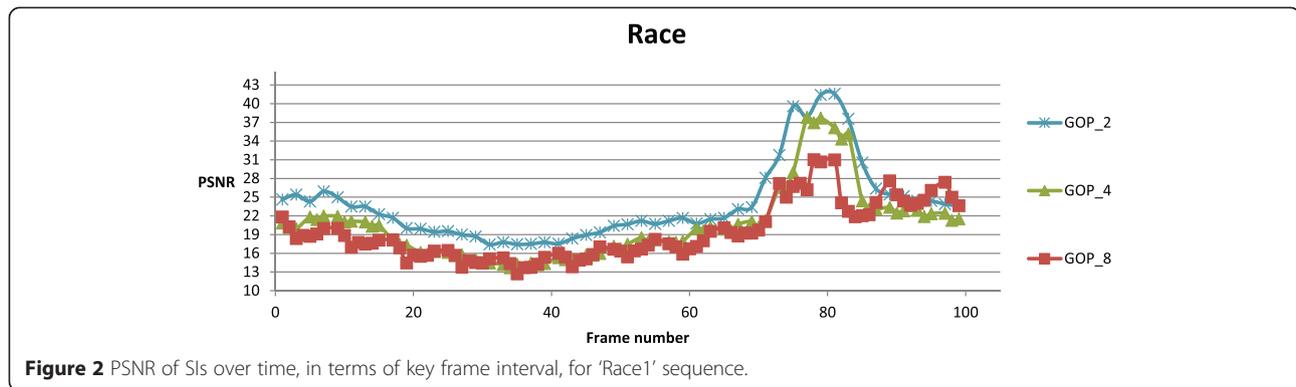
determined by the rate of the previous WZ frame at the encoder side [25]. They can improve performance of DVC with additional minimum complexity of the encoder. However, the method of SI frame generation at an encoder is different from that of a decoder. Even though the estimated rates are correct, the RD for the current frame might not be appropriate for the consecutive future frames, because of scene change and/or moving objects. For higher performance, Yaacoub et al. proposed a hierarchical decision of the GOP size, based on RD competition [26]. Two successive frames, the first and the last frames, for a given GOP size are coded as key frames. A target frame in between two given frames is coded as either WZ or key frame, based on RD competition. This procedure is hierarchically repeated for decision of the best GOP size. However, the encoder requires high computational complexity for GOP size decision in hierarchically evaluating the RD costs of key and WZ modes. The predicted SI at the encoder side is different from that at the decoder side, thus, the RD cost of the WZ mode is not the same to the original one. As a result, the amount of parity bits would not be accurate to correct the errors of SI frames.

Table 1 Average PSNRs of SI frames in terms of key frame intervals

Test sequence	Key frame interval						
	2	3	4	5	6	7	8
Akko	33.81	31.19	29.24	27.59	25.79	25.67	24.67
Ballroom	28.97	26.44	24.32	22.79	21.70	21.85	21.11
Flamenco2	31.35	28.96	27.44	26.40	25.55	25.53	24.91
Race1	23.85	22.07	20.46	19.83	18.84	20.94	19.53

3 Proposed DVC for hierarchical GOP structure

Since videos generally include not only fast-motion but also slow-motion performance of DVC, systems can be improved by adaptively modifying the GOP size. When the accuracies of SI frames are quite high, the encoder is not likely to send parity bits, and can reduce the number of key frames. Therefore, the proposed method adaptively modifies hierarchical GOP size based on RD competition, for compression performance of DVC systems.



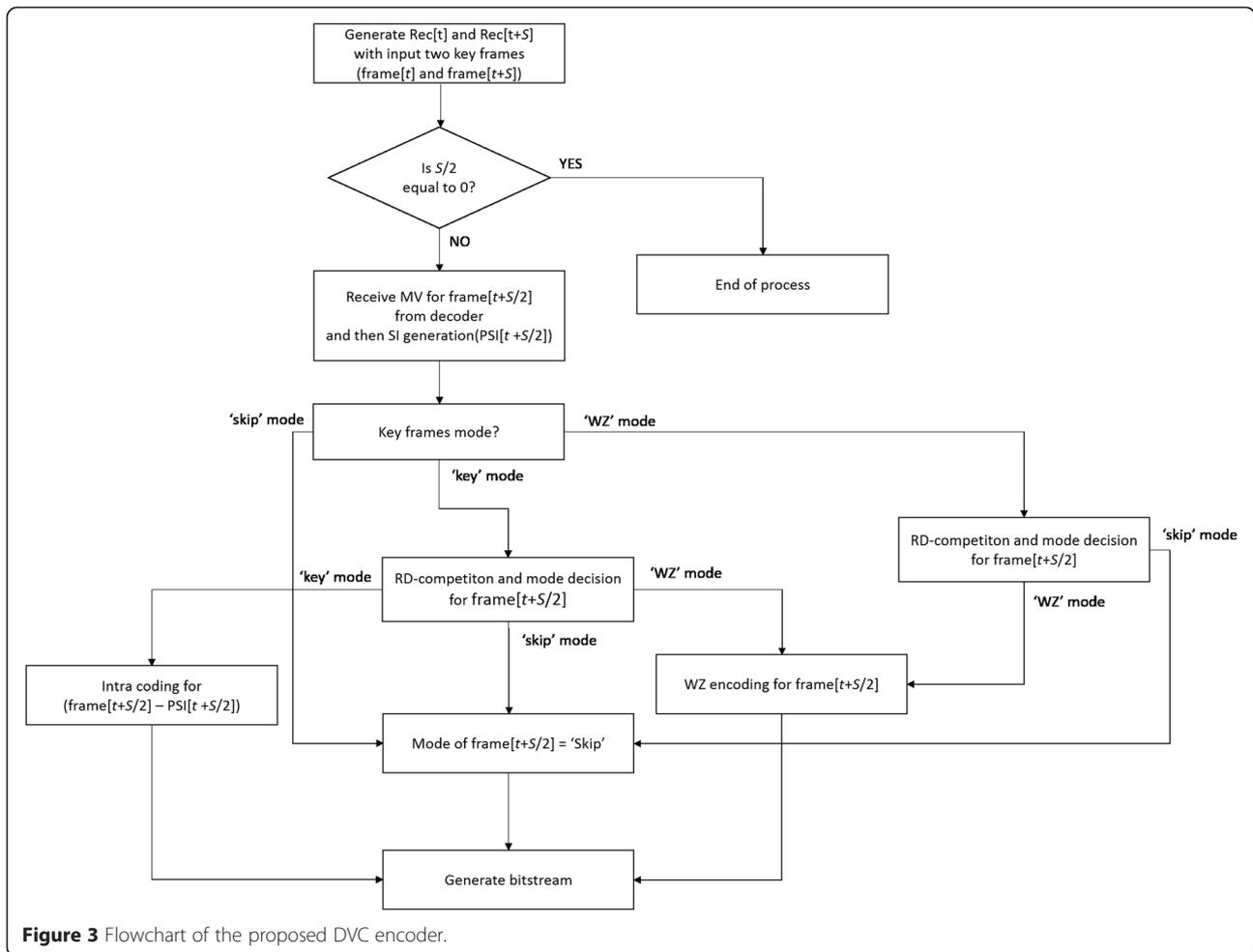
3.1 Proposed DVC encoder and decoder supporting hierarchical GOP structure

The proposed method sets the initial GOP size (S) and encodes the first frame at t and the last frame at $t + S$ in a GOP range as the key mode. Coded bitstreams of the key frames are sent and reconstructed in the decoder side. The SI frame at $t + S/2$ that is located at the interposition between two key frames is generated with the key frames, and then MVs are estimated from the SI frame at $t + S/2$ to the key frames, and the compressed MVs in a lossless mode are sent to the encoder side. As the encoder generates a PSI frame at $t + S/2$ with the received MVs and the key frames, the PSI can be the same as the SI in the decoder, without high computational complexity. Based on the PSI frame, the proposed encoder assesses the RD costs of key, WZ, and skip modes and selects one of them. The frame at $t + S/2$ is coded by the selected mode, and its associated data are sent to the decoder side. Hierarchically, the SI frame at $t + S/4$ ($t + 3S/4$) is generated with the frame t and the frame at $t + S/2$ (frame at $t + S/2$ and frame at $t + S$).

Figure 3 shows the flowchart of the proposed DVC encoder. The first frame (frame[t]) and the last frame (frame[$t + S$]) are coded as a key-frame coding and H.264/AVC intra-frame coding is employed for our implementation. The other frames between two key frames are hierarchically and recursively coded, as shown in the flowchart. Given two reconstructed frames, the encoder receives the MVM (MVM[$t + S/2$]) that includes MVs and their compensation directions for all the blocks of the target frame (frame[$t + S/2$]). With the received MVs, the proposed DVC encoder can generate the PSI (PSI[$t + S/2$]), and then decide RD competition candidates, according to the modes of the two input frames. When either input frame (Rec[t] and Rec[$t + S$]) is coded by 'Skip' mode, the target frame (frame[$t + S/2$]) is coded by skip mode, and only one 'Skip' indication bit is transmitted. In addition, recursive processing is not required, since all the frames between the two input frames (Rec[t] and Rec[$t + S$]) are coded by skip mode.

If either input frames of the reconstructed frames (Rec[t] and Rec[$t + S$]) is coded by 'WZ' mode, RD competition candidates for the target frame are skip and WZ modes. Based on the RD competition, the frame is encoded by a selected mode, and the associated indication bit is sent to the decoder side. When both input frames are coded by the key mode, the candidates are key, WZ, and skip modes for RD competition. The target frame is encoded by one of the candidate modes, and the mode bits are sent to the decoder side. This recursive encoding keeps running, until all the frames between two key frames are coded. Note that the WZ frames are coded in frequency domain with the LDPCA channel coder. Because the proposed DVC encoder requires motion compensation with the received motion vectors, the encoder computing time could slightly increase. However, the proposed hierarchical DVC encoder can skip several frames in a GOP structure. As a result, we found that the proposed DVC encoder complexity is almost same to the conventional DVC encoders. In our evaluation, the proposed algorithm is up to 5% slower than the DISCOVER coder, depending on sequences.

Figure 4 shows the flowchart of the proposed DVC decoder. First, two key frames are decoded by the corresponding intra decoder with a received bitstream. The other frames are hierarchically and recursively decoded in the proposed method. With two reconstructed frames (Rec[t] and Rec[$t + S$]), an SI frame (SI[$t + S/2$]) and MVM (MVM[$t + S/2$]) are generated in the proposed method, as shown in the flowchart. If two input frames are decoded by the skip mode, all the frames between them will be regarded as skip mode, and reconstructed by the SI generation algorithm, without any additional data. Otherwise, the SI frame is decoded depending on following syntax elements. MVM (MVM[$t + S/2$]) is not sent to the encoder for all the frames between two input frames. When neither of two input frames is coded by the skip mode, MVM (MVM[$t + S/2$]) should be sent to the encoder and the target frames are decoded depending on the coding mode. Note that the motion vectors



are predicted with median filtering of the neighboring motion vectors and motion vector difference is coded with the Exp-Golomb code for better compression.

3.2 SI frame generation and PSI frames compensation

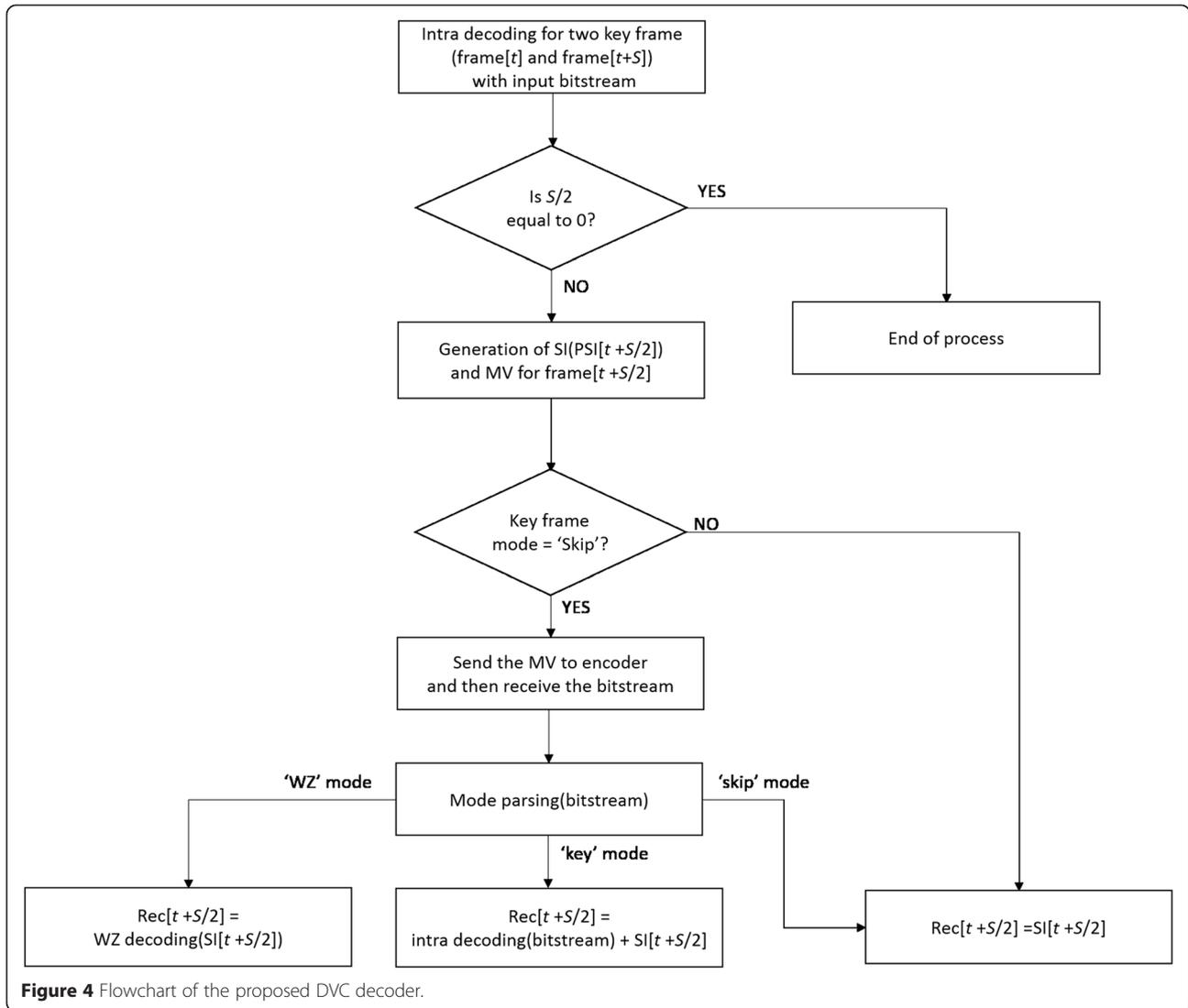
The quality of an SI frame directly impacts on the performance of a DVC system. For the proposed algorithm, it is also important to generate a PSI frame that is the same as the SI frame, for proper decisions of coding modes on the encoder side. In order to make sure that PSI frames are the same as SI frames, SI frames in the proposed algorithm are generated with the existing two-stage algorithm [38]. In the first stage, an initial SI (ISI) frame is estimated with key frames [23] for a target frame, however; any SI frame generation (SIG) algorithms with a gap-filling algorithm can be employed. At the second stage, the proposed DVC decoder performs motion estimation from the ISI frame to the neighboring key frames, and then the final SI frame is reconstructed with the key frames and the estimated MVs [38]. The motion vectors are sent to the decoder side. Regardless of the first stage motion estimation algorithm, we can

guarantee that the SI of the decoder side can be reconstructed with the transmitted MVs and related data at the encoder side, because the motion vectors are defined from the target frame to key frames. In addition, the proposed DVC encoder does not require any hole-filling algorithms and blending of two overlapped blocks. As a result, the SI frame can be generated with minimum computational load. Note that the first stage motion estimation is conducted with a conventional algorithm, based on adaptive search range for DVC [23].

Figure 5 shows the flowchart of the proposed ISI algorithm. The proposed method performs hierarchical ME in descending order of block variance values with adaptive search range. Search range is adaptively determined according to MVs of neighboring blocks. Then, the hole regions are compensated with uni-directional ME and a linear interpolation.

3.3 RD competition

Conventional codecs such as H.264/AVC calculate RD cost for all cases and select the best mode having the



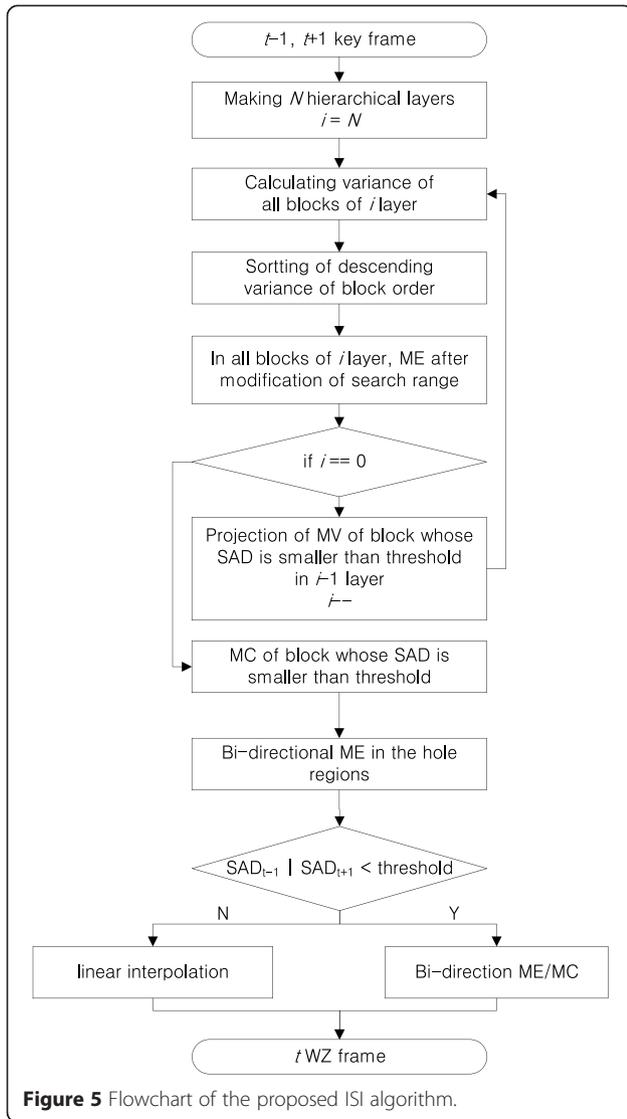
smallest RD cost for the best RD performance. The RD cost is defined by:

$$RD = \lambda \cdot rate + distortion \tag{1}$$

where λ is a scaling factor. Conventional encoders, such as H.264/AVC, conduct pre-encoding and calculate rates and distortions for multiple modes. Then they select the best coding mode jointly having minimum rate and distortion. Therefore, they require high computational complexity, although they can select the best coding mode. However, since the purpose of the proposed DVC is to encode videos with low computational complexity, conventional methods to compute RD costs are not suitable for the best mode selection in the DVC encoder.

The key feature of DVC codecs is to encode videos with low computational complexity. In the conventional

codecs, RD competition generally increases RD performance with high computational complexity. However, conventional DVC encoders do not employ RD competition, due to the computational complexity and non-availability of the reconstructed frames. In the proposed DVC, we employ RD competition for high RD performance with minimum computational complexity. In addition, to reduce encoding computational time, the proposed method determines the candidate modes to conduct RD competition among key, WZ, and skip modes, depending on the coding mode of a previous coded frame. As input frames are hierarchically coded in the proposed method, we can predict which modes are suitable for the consecutive frames, based on the coding mode of the previous frame. In the proposed algorithm, approximate RD costs are computed. To perform RD competition, the proposed method estimates RD costs of the selected candidates



depending on the conditions, as shown in the encoder flowchart. However, we need to note that the quality of a frame coded by WZ (skip) mode could be reasonably good, even when the rate is 0. The distortion of the frame is the same as that of the associated SI frame, because SI frames are generated from reference key frames without any explicit data. Therefore, WZ or skip mode is likely to be selected with low rate and high distortion by RD competition. However, a video quality that is too low is not suitable for commercial video applications. For quality control, the best mode is decided not only by RD competition but also by a quality threshold.

When an accurate objective visual quality and its bitrate are known, we can perform accurate RD competition. For the competition, actual encoding and decoding should be conducted at the encoder side. However, DVC-based encoders have the philosophy of

low complexity at the encoder side. In the proposed algorithm, the PSI can be reconstructed with the received motion vectors; thus, it is helpful to estimate more accurate objective visual quality. However, we cannot reconstruct the decoded frames at the encoder side due to low complexity constraint. Nevertheless, the proposed algorithm is better than the exiting algorithms with better prediction in estimating approximated RD competition.

3.3.1 Approximate RD cost of key frame mode

In this work, we propose an approximate RD cost of key frame modes before actual encoding for low computational complexity. Since the conventional RD costs require high computational complexity, DVC encoders cannot conduct pre-encoding of all the modes for RD competition. In addition, it is hard to correctly estimate rate and distortion of a frame before actual encoding. Figure 6 shows the actual data rate and distortion values for 100 intra-coded frames of the ‘Akko’ sequence. However, we found that rate and distortion are likely to slowly change for a short period of time, as shown in Figure 6a,b. Note that the key frames are coded by H.264/AVC intra coding. We assume that one rate for a key frame can be estimated by a linear relationship of those of adjacent key frames for low complexity. Thus, we propose the approximately estimated RD value with weighted linear interpolation by:

$$R_{L,t}^K = w_i \cdot R_{L,\alpha}^K + w_j \cdot R_{L,j}^K \quad (2)$$

$$D_{L,t}^K = w_i \cdot D_{L,\alpha}^K + w_j \cdot D_{L,j}^K \quad (3)$$

where R , D , and L denote rate of a key mode, its distortion, and hierarchical layer, respectively. l , α , and i are layer indexes ($\alpha < l$ and $i > l$). t , β , and j are display indexes ($\beta < t$ and $j > t$). The weight values (w_i , w_j) are determined by distance ratio between the key frames to the target frame with the condition, $w_i + w_j = 0$. The RD cost of the key mode is estimated by:

$$RD_{L,t}^K = \lambda_K \cdot R_{L,\alpha}^K + D_{L,j}^K \quad (4)$$

where λ_K is a scaling factor between rate and distortion.

3.3.2 Approximate RD cost of WZ mode

WZ frames are reconstructed from the SI frame with error correction via a channel decoder. Since channel decoding operation is one of the main sources of computational load, it is not proper to perform the channel decoding on the encoder side for low complexity encoding. Therefore, the proposed method estimates approximate RD costs, by predicting the reconstructed WZ

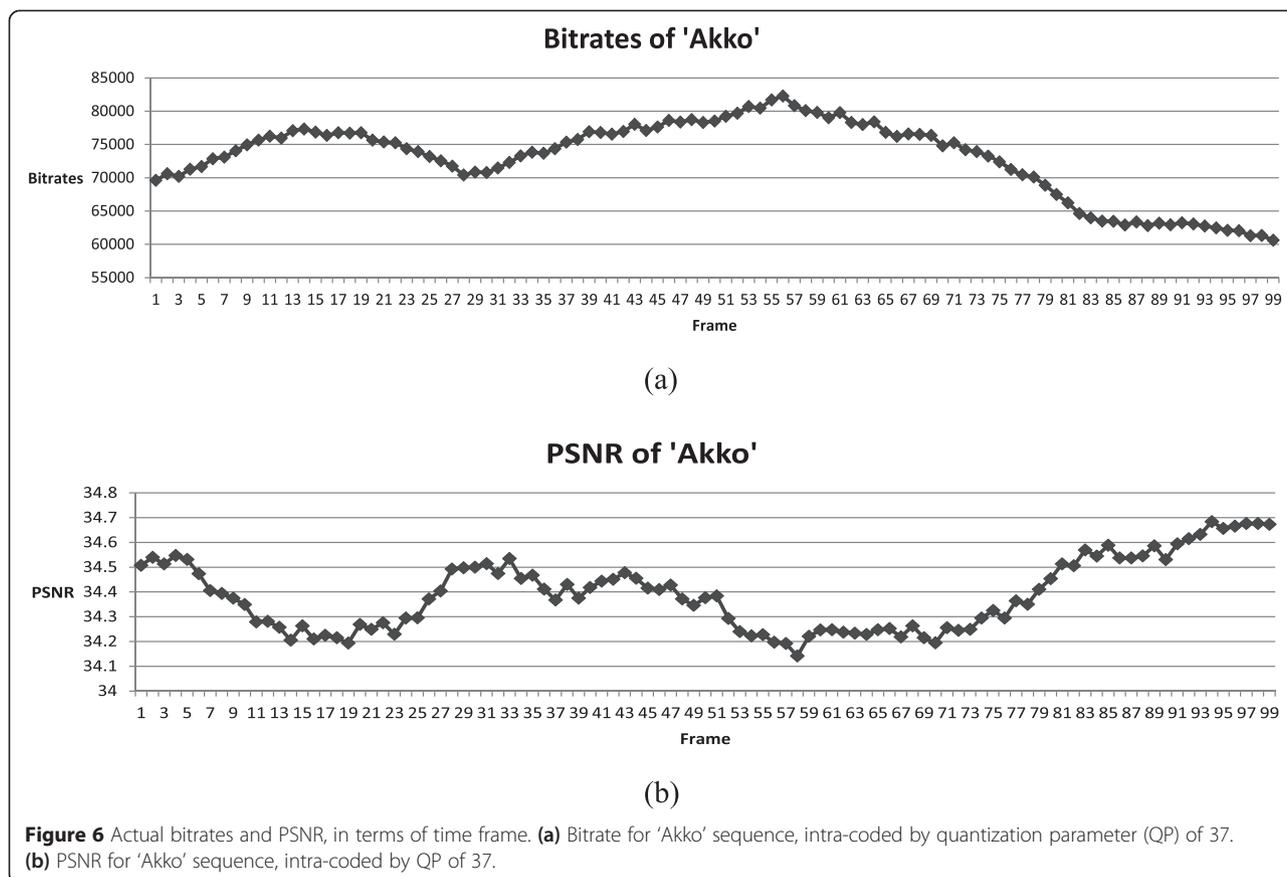


Figure 6 Actual bitrates and PSNR, in terms of time frame. **(a)** Bitrate for 'Akko' sequence, intra-coded by quantization parameter (QP) of 37. **(b)** PSNR for 'Akko' sequence, intra-coded by QP of 37.

frame with low computation complexity, before actual WZ encoding.

Conventional DVC encoders quantize discrete cosine transform (DCT) coefficients and generate parity bits for pre-determined frequency components in a DCT block, in order to correct all of the errors in the pre-determined regions. Therefore, we can assume that the pre-determined DCT regions of a WZ frame have no errors, and other regions have errors as much as those of the corresponding parts of an SI frame, as shown in Figure 7. Figure 7a,b shows blocks of an SI frame and a reconstructed WZ frame in the DCT domain, respectively. In Figure 7a, the light gray region is the pre-determined region by a quantization, and the region will be coded by a channel coder. Therefore, after conducting of channel decoding, the light gray part will be corrected. The white region in Figure 7b represents the no-error region, because all errors are corrected by a channel decoder with the received parity bits from an encoder. This means that the part is the same as the corresponding part of an original frame. The dark-gray parts that exist in Figure 7a,b depict the regions not to be corrected by a channel coder for coding efficiency. Therefore, the region of an SI frame is the same as the corresponding part of a WZ frame. Therefore, the proposed method can

generate the predicted reconstructed WZ (PRW) frame without high computational complexity, as the white part and dark-gray part are reconstructed by the original frame and the PSI, respectively, with minimal complexity.

With the PRW frame, the proposed method can predict the approximate rate and distortion of the WZ mode without high computational complexity. In the proposed method, the rate of the PRW frame is predicted with the correction capability of a channel coder and error rates of the PRW frame. The correction capability of a channel coder is generally related to the quantity of parity bits, and the amount of parity bits depends on the number of errors, as shown in Figure 8. Figure 8 shows the number of bits to correct the error in the triangular pyramid of the DCT bitplane in terms of the error rate of the PSI for several test sequences. Therefore, as the proposed method accounts for the number of errors of a PSI frame, the rate of the WZ mode can be approximately assessed. For each block, the demand bit is determined based on the rate of the WZ mode. Therefore, as the proposed method accounts for the number of errors of a PSI frame, the rate of the WZ mode can be approximately assessed. Since error correction by a channel coder is conducted in the frequency domain, the original and PSI frames are transformed with DCT by:

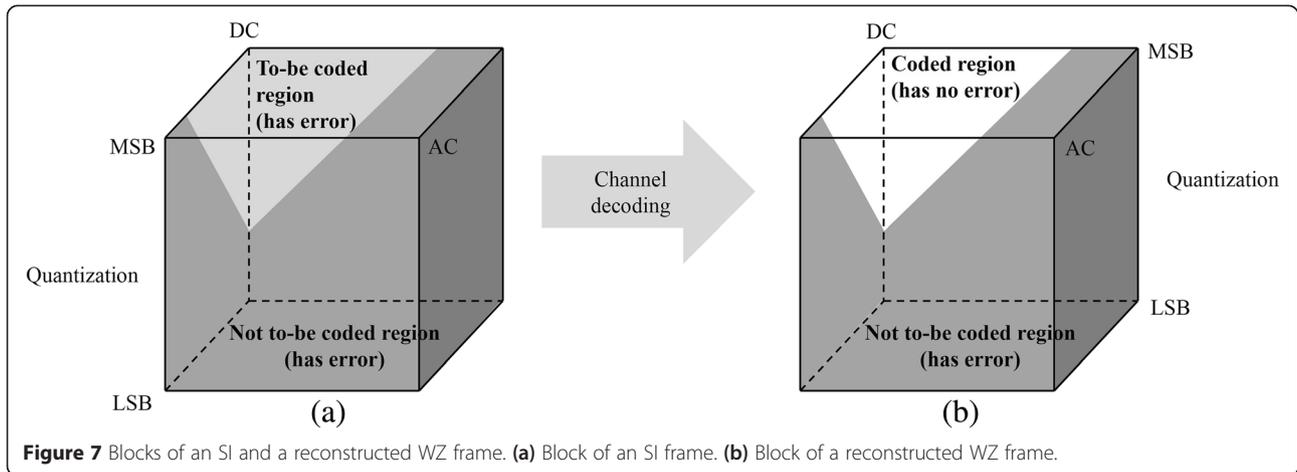


Figure 7 Blocks of an SI and a reconstructed WZ frame. (a) Block of an SI frame. (b) Block of a reconstructed WZ frame.

$$PSI_{L,t}^{DCT} = FD(PSI_{L,t}) \tag{5}$$

$$ORI_{L,t}^{DCT} = FD(ORI_{L,t}) \tag{6}$$

where FD is a function of forward DCT transform. After the transformation, the proposed method calculates the amount of errors of a PSI frame by:

$$NE_{L,t} = \sum_b^B \sum_c^C \sum_p^P |ORI_{L,t}^{DCT}[b, c, p] - PSI_{L,t}^{DCT}[b, c, p]| \tag{7}$$

where (b, c, p) indicates block, coefficient, and bitplane parameters, respectively. (B, C) means the number of blocks in a frame and coefficients in a block, respectively. P represents the number of bitplanes. The error rate $(E_{L,t})$ of the frame is estimated by:

$$E_{L,t} = \frac{NE_{L,t}}{NT_{L,t}} \times 100 \tag{8}$$

where $NT_{L,t}$ indicates the number of total bits in a frame. Based on the computed error rate, the proposed

method estimates the approximate rate of the WZ mode, as shown in Figure 8. In order to compute the distortion of the PRW frame, we need to generate the PRW frame. With the transformed original and PSI frames, the PRW frame is compensated for by:

$$PRW_{L,t}^{DCT}[b, c, p] = f(p)ORI_{L,t}^{DCT}[b, c, p] + (1-f(p))PSI_{L,t}^{DCT}[b, c, p] \tag{9}$$

$$f(p) = \begin{cases} 1 & p < P_Q \\ 0 & \text{otherwise} \end{cases}$$

Note that P_Q is a quantization parameter in defined in bitplane. In order to assess a distortion, the sum of squared error (SSE) is computed after inverse transformation and is denoted by:

$$PRW_{L,t} = ID(PRW_{L,t}^{DCT}) \tag{10}$$

$$D_{L,t}^W = \sum_{x=0}^X \sum_{y=0}^Y (ORI_{L,t}[x, y] - PRW_{L,t}[x, y])^2 \tag{11}$$

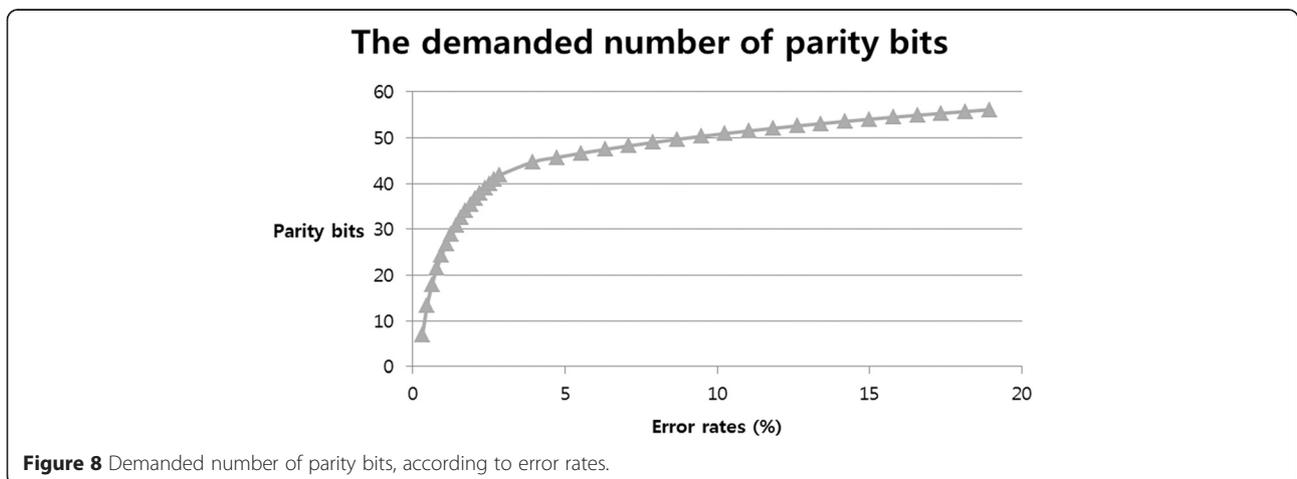


Figure 8 Demanded number of parity bits, according to error rates.

where ID represents a function of inverse DCT. With the estimated rate and distortion, the proposed method computes the RD cost of the WZ mode by:

$$RD_{L,t}^W = \lambda_K(1 + R_{L_{ij}}^W) + D_{L_{ij}}^W \quad (12)$$

When a frame is selected to encode by WZ mode, the proposed method sends as many parity bits by accounting for the number of DCT blocks and the number of the demanded bit for the block, as shown in Figure 8. Therefore, the proposed method does not need feedback iteration and reduces time delay. If the demanded number of parity bits for error correction is different from the computed value, the performance of the proposed method could decrease. Once a frame is to be coded as the WZ mode, each block is evaluated whether its quality is enough good or not. Parity bits for the well-predicted one with the PSI could be not sent to the decoder side. For other blocks, the proper amount of bits given by Figure 8 is supposed to be sent.

3.3.3 Approximate RD cost of skip mode

For the proposed skip mode, the proposed method estimates the RD values with the PSI frames that are generated in the previous step. Since a skip indication bit is sent to the decoder side, the rate is one bit for skip mode. The distortion is calculated with the SSE between an original and PSI frames, instead of SI frames, and the distortion is represented by:

$$D_{L,t}^S = \sum_{x=0}^X \sum_{y=0}^Y (\text{ORI}_{L,t}[x,y] - \text{PSI}_{L,t}[x,y])^2 \quad (13)$$

For the skip RD cost, the proposed method computes a skip RD cost with a distortion of skip by:

$$RD_{L,t}^S = 1 \cdot \lambda_K + D_{L_{ij}}^S \quad (14)$$

Note that λ_K is empirically computed with six sequences ('Akko', 'Ballroom', 'Exit', 'Flamenco2', 'Race1', and 'Rena' sequences). The parameter is set to (1,518, 3,824, 9,636, 24,281, and 61,185) as a function of QPs (33, 37, 41, 45, and 49).

4 Experimental results

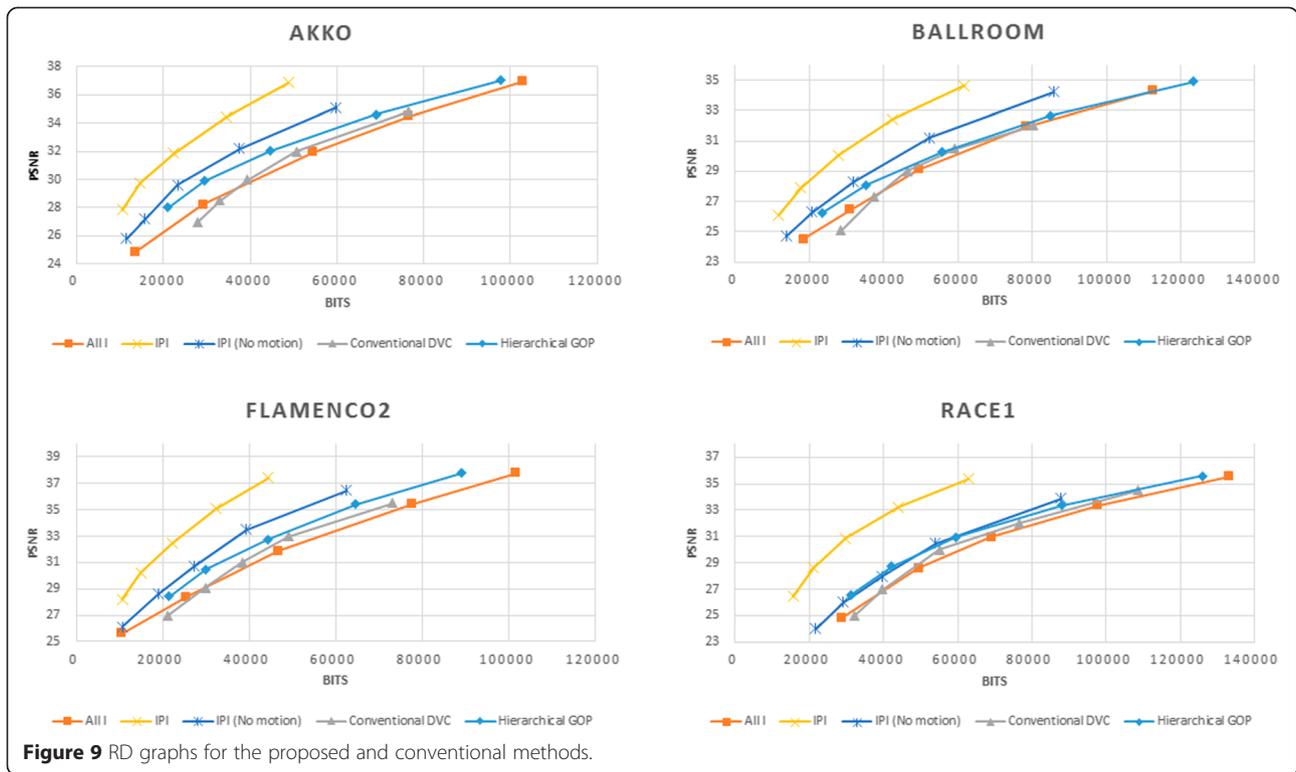
For performance evaluation of the proposed algorithm, the RD performance of the proposed and conventional algorithms was evaluated. Four test sequences ('Akko', 'Ballroom', 'Flamenco2', and 'Race1') were used with the format and size of 4:0:0 YUV and 640 × 480, respectively. Key frames were coded using JM 17.2, and five QP points (33, 37, 41, 45, and 49) were used. Note that the

'Akko', 'Ballroom', 'Flamenco2', and 'Race1' sequences consist of 300, 250, 250, and 250 frames, respectively. The conventional algorithm employs every other frame as the key frame, while the number of key frames are determined depending on GOP size. The SI frames are reconstructed based on an adaptive search range [24] and an LDPCA channel coder with a matrix length of 6,336 [5].

Table 2 shows errors of the estimated rate and PSNR of the proposed algorithm in terms of GOP size for four test sequences. The accuracies are shown in differences of rates and PSNRs for all the frames between the key frames given by intra periods of 8, 4, and 2. The proposed algorithm has three estimators for three modes. The figures in the table are average errors for the three modes of the proposed algorithm. For the case of GOP size of two, the proposed accuracy is quite high. As the GOP increases, estimated errors become larger. As shown in the table, although computed rates and PSNRs for 'GOP8' are less accurate than those for 'GOP2', we can say that the overall accuracy is high, with minimal computational load. In any cases, accuracies of the computed values are less than 2%. 'Ballroom' and 'Race1'

Table 2 Estimated errors of rates and PSNRs with the proposed algorithm in terms of GOP sizes

Sequence	QP	Rate (%)			PSNR (%)		
		Average rate difference			Average PSNR difference		
		GOP8	GOP4	GOP2	GOP8	GOP4	GOP2
Akko	33	0.725	0.328	0.219	0.078	0.063	0.033
	37	0.814	0.423	0.257	0.088	0.072	0.035
	41	0.993	0.519	0.310	0.115	0.076	0.046
	45	0.939	0.654	0.390	0.151	0.111	0.055
	49	1.057	0.696	0.399	0.168	0.124	0.061
Ballroom	33	1.027	0.434	0.222	0.111	0.065	0.025
	37	1.116	0.526	0.326	0.132	0.081	0.036
	41	1.186	0.609	0.313	0.155	0.100	0.046
	45	1.308	0.696	0.405	0.163	0.109	0.045
	49	1.156	0.622	0.360	0.216	0.155	0.065
Flamenco2	33	0.596	0.376	0.202	0.099	0.077	0.036
	37	0.509	0.373	0.245	0.100	0.074	0.031
	41	0.683	0.518	0.294	0.123	0.102	0.055
	45	1.075	0.692	0.405	0.128	0.106	0.049
	49	0.857	0.628	0.362	0.212	0.168	0.084
Race1	33	0.910	0.361	0.168	0.095	0.067	0.025
	37	0.788	0.411	0.214	0.120	0.079	0.032
	41	0.940	0.465	0.265	0.152	0.107	0.040
	45	0.986	0.629	0.344	0.175	0.115	0.054
	49	1.270	0.684	0.404	0.282	0.153	0.083
Average		0.944	0.577	0.320	0.143	0.104	0.049



sequences have relatively high motion activity over the other sequences. Thus, the estimated errors are obtained to be somehow larger.

Figure 9 shows RD graphs of the existing [24] and the proposed methods. ‘All I’ means that all the frames in a video are coded by intra mode. ‘IPI’ indicates that even frames are coded by inter mode, and odd frames are encoded by intra mode. ‘IPI(No motion)’ represents that even frames are coded by intra mode, and the other frames are coded by inter mode with zero motion vectors. ‘Conventional DVC’ means that each sequence is coded by DISCOVER [29]. The accuracies of SI frames are likely to be high, when motion is linear and/or slow in a video. The proposed method can exactly compute the SI quality without high computational complexity, using the PSI frames in the proposed DVC encoder. Therefore, we can reduce the number of key frames to increase RD performance. This means that the proposed method yields higher performance than the existing DVC and the all intra-coded cases, as shown in Figure 9. For the existing DVC cases, a lot of parity bits are required to correct the error in SI frames for large motion pictures. However, the proposed method evaluates the accuracies of SI frames, and it can prevent excessive parity bits, by balancing the rate and distortion at the encoder side. In addition, the proposed method influences the current coding mode on the decision of coding mode of nest frames for low computational complexity.

Since the proposed method hierarchically determines coding modes, the best coding mode for each frame is effectively decided with low computational complexity. Experimental results show that the proposed algorithm yields BD bitrate reduction of around -11.42% on the top of the existing DVC system. However, the DVC-based coders do not outperform the existing video coders based on hybrid transform coding such as H.264/AVC and others. DVC is employed to improve predicted frames by channel coders. Any channel coders cannot perfectly guarantee error corrections, however; they can statically achieve error correction. The conventional video coders (H.264/AVC, HEVC, and other international standards) guarantee matching between encoder and decoder sides, while DVC-based coders cannot in general guarantee the encoder-decoder matching. The encoder cannot know the exact decoded pictures and vice versa. Thus, at any sides, we cannot make sure whether the correction is proper or not. Parity bits from channel coders could make errors corrected; however, they could also make corrected prediction wrong. That is one of critical reasons that DVC-based coders have difficulty in outperforming the H.264/AVC, HEVC, and so on.

Figure 10 shows the bitrates and PSNR of the proposed algorithm with respect to frame index for ‘Ballroom’ sequence. The bitrates and PSNR were obtained with QP of 33. As shown in the figure, high bitrates are observed periodically with GOP size of 8 due to intra-frame coding.

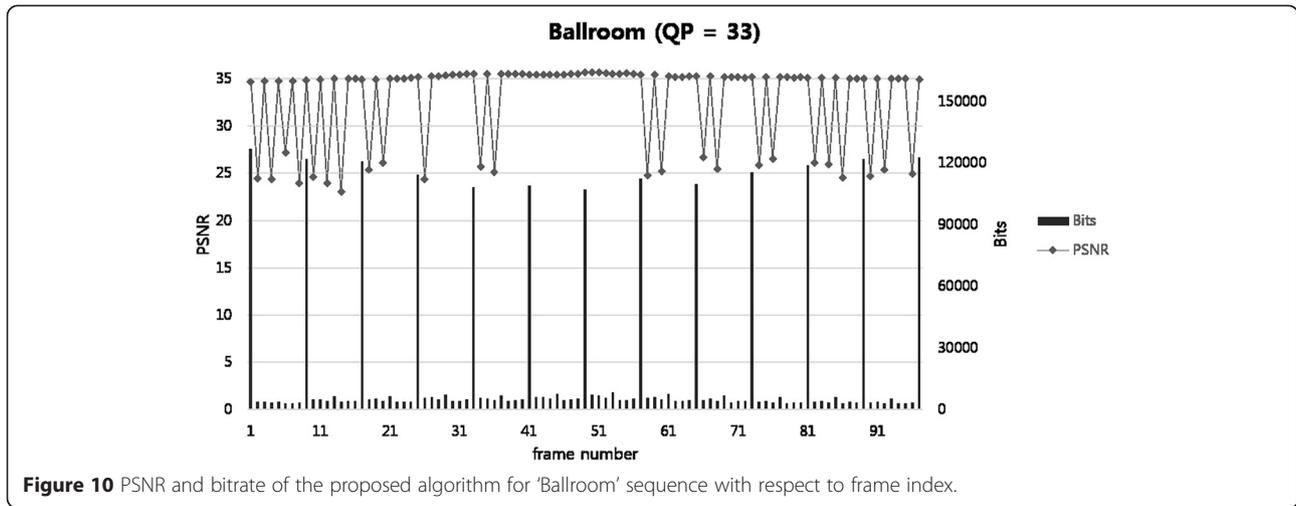


Figure 10 PSNR and bitrate of the proposed algorithm for 'Ballroom' sequence with respect to frame index.

In between two intra frames, low bitrates are seen for parity bits, motion vectors, and additional syntaxes. High quality coding is achieved for the frames between 40th and 55th indexes because motion activity is relatively low. For several frames, prediction accuracy is not so high thus error correction does not work well because the error rates exceed the correction rate of the channel coder.

Due to the proposed hierarchical GOP structure, a delay should be involved, like conventional video coding having GOPs. Figure 11 shows a delay diagram of the proposed method. The first and last frames for the initial GOP structure are encoded and sent to the decoder side in parallel. The decoder reconstructs the frames with the received bitstreams. With the reconstructed frames, the SI frame located in between the frames is generated, and the motion vectors for the SI frame are transmitted to the encoder.

The encoder generates the PSI with the received motion vectors and performs the proposed RD competition. With the decided best mode, the proposed encoder codes the target frame. This processing is repeated until half of the initial GOP structure has been reached. The processing delay of the proposed method can be denoted by:

$$\text{Delay} = \frac{S}{2}(\alpha + \varepsilon) + (S-1)\beta + \left(\frac{S}{2}-1\right)(\gamma + \delta + \zeta) \quad (15)$$

where S means the time interval for the initial GOP size ($=8$). α , ε , β , γ , δ , and ζ indicate encoding, decoding, transmission, PSI generation, RD competition, and SI generation times, respectively. Through the experiment for estimation of the delay, the encoding, decoding, transmission, PSI generation, RD competition, and SI

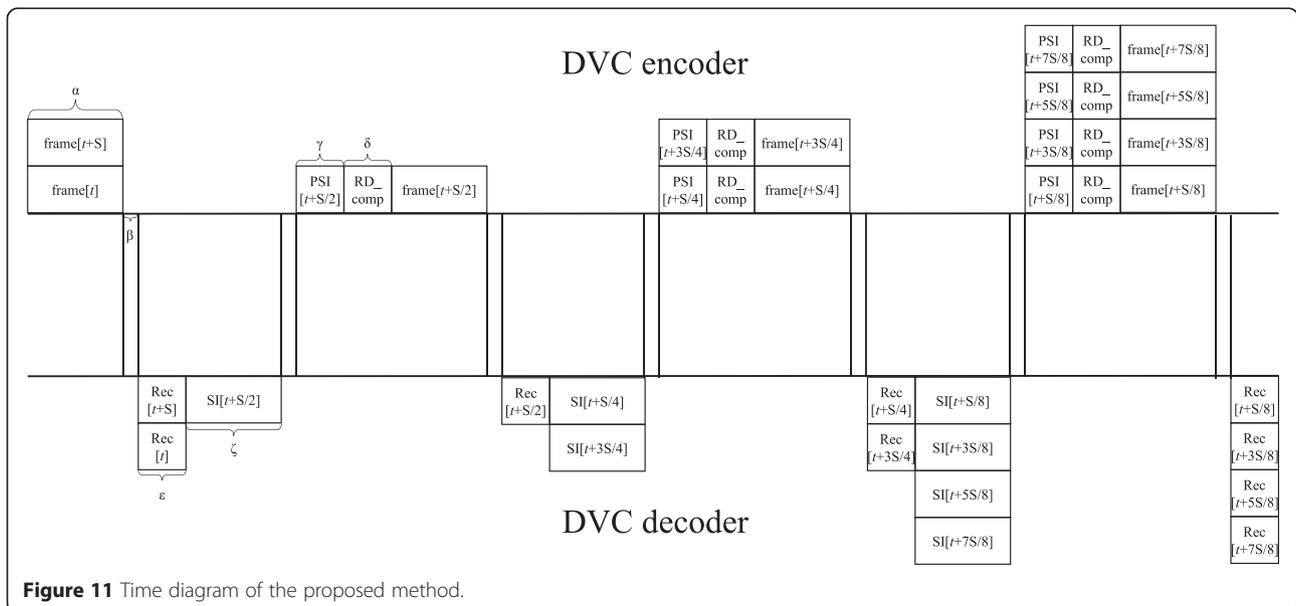


Figure 11 Time diagram of the proposed method.

generation time are 600, 14, 5, 14, 400, and 3,500 ms, respectively. We found in the experiment that the proposed DVC requires 14,233 ms for a GOP structure. Note that the proposed system was implemented on Intel i5 (2,53 GHz) with 4 GB over Window 7. The proposed feedback-based DVC requires the delay, and it makes the proposed algorithm applied for high frame-rate video applications. However, the proposed algorithm is considered to be a trade-off between no-feedback DVC and iterative feedback DVC algorithms. Note that RD performance of the proposed algorithm is better than the no-feedback algorithms. This evaluation and assessment would be not practical for practical scenarios and conditions. The network delay can vary depending on traffics. In addition, we employed JM reference encoding software and SIG having large computational complexity in the evaluation. Hardwired logics or fast computing platforms can be employed to implement practical applications based on the proposed DVC system. Extensive further research should be performed for practical applications and services in the future.

5 Conclusions

In this paper, a new adaptive distributed video coder has been proposed, with hierarchical GOP structure. In the proposed algorithm, the PSI can be reconstructed in the encoder, using reference key frames and MVs without motion estimation. Therefore, we can estimate the exact accuracies of SI frames with the PSI frames. With the PSI frames, the proposed method performs RD competition and selects the best coding mode. Based on the decided coding mode, the best GOP structure is automatically decided in the proposed method. As the proposed method reduces the number of key frames when a video has little and/or linear motion, the performance of the proposed method improves. In addition, the proposed method reduces the number of WZ frames, if large motion between consecutive frames occurs, because an SI frame of low accuracy requires many parity bits for error correction. Therefore, the proposed method has higher performance than the several existing methods. However, the proposed method requires high computational complexity, according to the initial GOP size. For further work, we would optimize the encoding, decoding, and SI generation modules, for reduction of the delay.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This research was partly supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A2A1A11052210) and the MSIP (Ministry of Science, ICT & Future Planning), Republic of Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2014-H0301-14-1018).

Author details

¹Department of Computer Engineering, Kwangwoon University, Wolgye-dong, Seoul 447-1, Nowon-gu, South Korea. ²School of Engineering Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada.

Received: 14 May 2014 Accepted: 29 April 2015

Published online: 10 May 2015

References

1. D Slepian, J Wolf, Noiseless coding of correlated information sources. *IEEE Trans Inf Theory* **19**(4), 471–480 (1973)
2. A Wyner, J Ziv, The rate-distortion function for source coding with side information at the decoder. *IEEE Trans Inf Theory* **22**(1), 1–10 (1976)
3. J Micallef, JR Farrugia, C Debono, *Low-density parity-check codes for asymmetric distributed source coding*. Paper presented at the 2010 1st IEEE International Conference on Information Theory and Information Security (IEEE, Beijing, China, 2010)
4. Q Linbo, H Xiaohai, L Rui, D Xiewei, *Application of punctured turbo codes in distributed video coding*. Paper presented at the 2007 4th IEEE International Conference on Image and Graphics (IEEE, Sichuan, China, 2007)
5. A Aaron, R Zhang, B Girod, *Wyner-Ziv coding of motion video*. Paper presented at the 2002 37th Asilomar Conference on Signals and Systems (IEEE, Grove, CA, 2002)
6. D Varodayan, A Aaron, B Girod, Rate-adaptive codes for distributed source coding. *EURASIP Signal Process J Spec Sect Distributed Source Coding* **86**(11), 3123–3130 (2006)
7. C Brites, F Pereira, *Encoder rate control for transform domain Wyner-Ziv video coding*. Paper presented at the 2007 14th IEEE International Conference on Image Processing (IEEE, San Antonio, TX, 2007)
8. F Zhai, IJ Fair, Techniques for early stopping and error detection in turbo decoding. *IEEE Trans Commun* **51**(10), 1617–1623 (2003)
9. WJ Chien, LJ Karam, GP Arousleman, *Rate-distortion based selective decoding for pixel-domain distributed video coding*. Paper presented at the 2008 15th IEEE International Conference on Image Processing (IEEE, San Diego, CA, 2008)
10. J Skorupa, J Slowack, S Mys, P Lambert, R Van de Walle, C Grecos, *Stopping criterions for turbo coding in a Wyner-Ziv video codec*. Paper presented at the 2009 27th IEEE Picture Coding Symposium (IEEE, Chicago, IL, 2009)
11. JL Martinez, C Holder, GE Fernandez, H Kalva, F Quiles, *DVC using a half-feedback based approach*. Paper presented at the 2008 9th IEEE International Conference on Multimedia and Expo (IEEE, Hannover, Germany, 2008)
12. B Du, H Shen, *Encoder rate control for pixel-domain distributed video coding without feedback channel*. Paper presented at the 2009 3rd IEEE International Conference on Multimedia and Ubiquitous Engineering (IEEE, Qingdao, China, 2009)
13. C Yaacoub, J Farah, B Pesquet-Popescu, *Content adaptive gop size control with feedback channel suppression in distributed video coding*. Paper presented at the 2009 16th IEEE International Conference on Image Processing (IEEE, Cairo, Egypt, 2009)
14. J Ascenso, C Brites, F Pereira, *Content adaptive Wyner-Ziv video coding driven by motion activity*. Paper presented at the 2006 13th IEEE International Conference on Image Processing (IEEE, Atlanta, GA, 2006)
15. M Morbee, J Prades-Nebot, A Pizurica, W Philips, *Rate allocation algorithm for pixel-domain distributed video coding without feedback channel*. Paper presented at the 2007 32nd IEEE International Conference on Acoustic, Speech, and Signal Processing (IEEE, Honolulu, HI, 2007)
16. J Kubasov, K Lajnef, C Guillemot, *A hybrid encoder/decoder rate control for a Wyner-Ziv video codec with a feedback channel*. Paper presented at the 2007 9th IEEE Workshop on Multimedia Signal Processing (IEEE, Crete, Greece, 2007)
17. WJ Chien, LJ Karam, GP Arousleman, *Block-adaptive Wyner-Ziv coding for transform-domain distributed video coding*. Paper presented at the 2007 32nd IEEE International Conference on Acoustic, Speech, and Signal Processing (IEEE, Honolulu, HI, 2007)
18. L Limin, L Zhen, EJ Delp, *Backward channel aware Wyner-Ziv video coding*. Paper presented at the 2006 13th IEEE International Conference on Image Processing (IEEE, Atlanta, GA, 2006)
19. W Jia, W Xiaolin, Y Songyu, S Jun, New results on multiple descriptions in the Wyner-Ziv setting. *IEEE Trans Inf Theory* **55**(4), 1708–1710 (2009)
20. R Liu, Z Yue, C Chen, Side information generation based on hierarchical motion estimation in distributed video coding. *J Aeronaut* **22**(2), 167–173 (2009)

21. Y Shuiming, M Ouaret, F Dufaux, T Ebrahimi, *Improved side information generation with iterative decoding and frame interpolation for distributed video coding*. Paper presented at the 2008 15th IEEE International Conference on Image Processing (IEEE, San Diego, CA, 2008)
22. H Xin, S Forchhammer, *Improved side information generation for distributed video coding*. Paper presented at the 2008 10th IEEE Workshop on Multimedia Signal Processing (IEEE, Cairns, Australia, 2008)
23. KY Min, SN Park, DG Sim, *Side information generation using adaptive search range for distributed video coding*. Paper presented at the 2009 11th IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (IEEE, B.C., Canada, 2009)
24. KY Min, SN Park, JH Nam, DG Sim, SH Kim, *Distributed video coding based on adaptive block quantization using received motion vectors*. *KICS J* **35**(2), 172–181 (2010)
25. I Ahmad, Z Ahmad, I Abou-Faycal, *Delay-efficient GOP size control algorithm in Wyner-Ziv video coding*. Paper presented at the 2009 7th IEEE International Symposium on Signal Processing and Information Technology (IEEE, Ajman, UAE, 2009)
26. C Yaacoub, J Farah, B Pesquet-Popescu, *New adaptive algorithms for GOP size control with return channel suppression in Wyner-Ziv video coding*. *Int J Digit Multimedia Broadcasting* **2009**, 319021 (2009)
27. G Huchet, W Demin, *Distributed video coding without channel codes*. Paper presented at the 2010 3rd IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (IEEE, Shanghai, China, 2010)
28. JL Martinez, G Fernandez-Escribano, H Kalva, WARJ Weerakkody, *Feedback-free DVC architecture using machine learning*. Paper presented at the 2008 15th IEEE International Conference on Image Processing (IEEE, San Diego, CA, 2008)
29. X Artigas, J Ascenso, M Dalai, S Klomp, D Kubasov, M Ouaret, *The discover codec, architecture, techniques and evaluation*. Paper presented at the 2007 IEEE Picture Coding Symposium (IEEE, Lisbon, Portugal, 2007)
30. M Jang, JW Kang, and SH Kim, *A design of rate-adaptive LDPC codes for distributed source coding using PEG algorithm*. Paper presented at the 2010 IEEE Military Communications Conference, San Joes, CA, 31 October-3 November 2010
31. SY Shin, M Jang, JW Kang, SH Kim, *New distributed source coding scheme based on LDPC codes with source revealing rate-adaptation*. Paper presented at the 2011 12th IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (IEEE, Victoria, Canada, 2011)
32. CK Kim, DY Suh, *Channel adaptive rate control for loss resiliency of distributed video coding*. Paper presented at the 2010 International Conference on Electronics, Information, and Communication (IEEK, Cebu, Philippine, 2010)
33. JA Park, DY Suh, GH Park, *Distributed video coding with multiple side information sets*. *IEICE Trans Inf Syst* **E93-D**(3), 654–657 (2010)
34. JY Lee, CW Seo, DG Sim, JK Han, *Efficient ME/MD schemes for Wyner-Ziv codec to VC-1 transcoder*. Paper presented at the 2011 International Technical Conference on Circuits/Systems, Computers and Communications (IEEK, Gyeongju, Korea, 2011)
35. SY Shim, JK Han, J Bae, *Adaptive reconstruction scheme using neighbour pixels in PDWZ coding*. *Electron Lett* **46**(9), 626–628 (2010)
36. R Oh, JB Park, BW Jeon, *Fast implementation of Wyner-Ziv video codec using GPGPU*, in *Symposium on IEEE BMSB*, 2010, pp. 1–5
37. X Van Hoang, BW Jeon, *Flexible complexity control solution for transform domain Wyner-Ziv video coding*. *IEEE Trans Broadcasting* **58**(2), 209–220 (2012)
38. KY Min, DG Sim, *Adaptive distributed video coding with motion vectors through a back channel*. *EURASIP J Image Video Process* **22**, 1–12 (2013)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
