



# Real-time low-bitrate multimedia communication for smart spaces and wireless sensor networks

K.R. Vijayanagar J. Kim

Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3301 South Dearborn Street, Siegel Hall Suite 103, Chicago, IL 60616, USA  
E-mail: kvijayan@iit.edu

**Abstract:** Smart spaces is a scenario that is ideal for the application of wireless communication technology to benefit non-expert users. A prime example of such an application would be low-cost and real-time video surveillance of hallways or video conferencing. Traditional solutions for these scenarios use state-of-the-art video codecs like H.264/AVC in order to achieve the low bitrate which is essential in wireless communication. On the flip side, such encoders are computationally complex and need expensive hardware to run on thus making them inefficient choices for being deployed in clusters. Recent research has resulted in distributed video coding (DVC) being proposed as a solution for applications that have limited battery resources and low hardware complexity, thus necessitating a low-complexity encoder. It is now a popular topic in the research community and the past years have seen several implementations. However, current DVC solutions use iteratively decodable channel codes like low-density parity check – accumulate (LDPCA) codes or Turbo codes that have large latencies. In order to make real-time communication possible, the proposed architecture makes efficient use of skip blocks to reduce the bitrate, eliminates the iterative decoding nature of the Wyner-Ziv (WZ) channel and uses a simple data-hiding-based compression algorithm. This drastically cuts down on the time complexity of the decoding procedure while still maintaining a rate–distortion performance better than that of H.264/AVC intra-coding and other current DVC solutions.

## 1 Introduction

Distributed video coding (DVC) is widely known in the research community nowadays and can be concisely described as the reverse paradigm of conventional video coding [1, 2]. Conventional video codecs employ a very complex encoder and have simple decoders like in the case of H.264/AVC, MPEG4, H.263+. Such architectures are ideal in a scenario where a video is encoded once but decoded several times (e.g. broadcasting a video over the Internet). Since the decoding process is simple, relatively low-cost and low-complexity hardware is needed at the decoder. However, there are new wireless video applications which need a low-complexity encoder because of resource constraints. Examples for such scenarios keeping in mind the smart spaces concept would be video conferencing between different rooms of a house/building, real-time monitoring of a hallway and access control to a building where one would like to see and communicate with the person requesting access. Solutions for such applications do exist but are mostly wired solutions that make it cumbersome to move around or are high bitrate and high delay.

DVC is a promising solution for the above-mentioned scenarios because it aims at shifting the coding complexity from the encoder to the decoder. It is popularly referred to as the reverse paradigm of conventional coding, where the entire complexity is moved from the encoder to the decoder. Recently, several practical implementations have been

proposed, including the Stanford codec [3, 4], DISCOVER codec [5, 6], SPI-DVC [7], PRISM codec [8] etc. Keeping in mind DVC's potential usage for monitoring/surveillance and low-cost live video conferencing, we propose an architecture that is block based and unidirectional in nature. The proposed architecture follows a block-based approach but deviates from the fixed alternating block pattern as dictated by H.264/AVC's dispersed FMO mode. Here, blocks can be classified into key, Wyner-Ziv (WZ), skip. This classification process, as the results show, is useful in reducing the bitrate and improving rate–distortion (RD) performance. Also, the traditional low-density parity check – accumulate (LDPCA)/Turbo codes used in almost all current DVC literature is not used and in its place a very low-complexity and high-speed compression algorithm that greatly speeds up encoding/decoding times is used. The architecture has two modes of operation where one can enable/disable rate control at the encoder depending on the complexity of the available encoder hardware. The feedback channel from the decoder to the encoder is removed since the WZ decoding is not iterative making the proposed architecture unidirectional.

As we have mentioned before, the system is a low-delay DVC codec implying that encoding and decoding take place in real time or faster than currently available DVC solutions. A low-delay DVC system [9] has been defined as one that does not use the future frame for prediction and the proposed architecture adheres to this definition.

Another significant improvement in the proposed system is the reduced complexity of the decoder. Since DVC has been termed as the reverse paradigm of conventional coding, it is natural that the decoder is highly complex and its operations are time consuming. However, this is not suitable for low-delay applications and the proposed codec overcomes this problem by simplifying the decoder's operations and thus achieving fast encoding–decoding times.

The organisation of the rest of the paper is as follows. In Section 2, we discuss the proposed architecture in detail. In Section 3, we provide simulation results and finally the conclusion.

## 2 Architecture

### 2.1 Encoder

The proposed architecture is shown in Fig. 1. At the encoder an incoming frame is classified as either a key (if it is the first frame of a group of pictures (GOP)) or a WZ frame (remaining frames of the GOP). A key frame is transmitted entirely to the encoder after being compressed using the H.264/AVC encoder operating in the main profile and in the Intra mode. This is popularly referred to as transmitting the key information losslessly to the decoder in DVC literature.

**2.1.1 Block classification:** The WZ frames are further subjected to a block classification scheme where the frame is divided into non-overlapping blocks of size  $N \times N$  and each block is classified into a key, WZ or a skip block. A key or a WZ block is defined as a block that shows sufficient difference (i.e. greater than a pre-defined threshold) when compared with its co-located block in the previous frame. A skip block is one that is almost identical to its co-located block in the previous frame. Such a block need not be transmitted to the decoder; rather it can be substituted with the co-located block from the previous frame at the decoder. The advantage in doing so is that the bitrate is greatly reduced when there is sufficient correlation between successive frames of a sequence.

Now that it is clear as to which blocks are classified as skip block, we explain the procedure of classifying a block as key or WZ. At first, after the frame is divided into non-overlapping blocks of size  $N \times N$ , the blocks are classified as key and WZ blocks alternatively. This is similar to the checker-board pattern employed in H.264/AVC's Dispersed

FMO mode. This alternating pattern is carried over at the temporal level too, with the  $n$ th frame having its first block as a key block and the  $(n + 1)$ th frame starting with a WZ block. After this initial classification, we compare each block with its co-located block in the previous frame. The comparison is done by computing the boundary sum of absolute differences (BSAD) between the two co-located blocks. BSAD is a modification of the SAD distortion metric and it takes only the boundary values of a block into consideration for the SAD computation as shown in Fig. 2.

If we divide a quarter common intermediate format (QCIF) frame ( $176 \times 144$  pixels) into  $4 \times 4$  blocks then 1584 SAD comparisons have to be performed between the blocks of the current frame with those of the previous frame. This will amount to  $1584 \times 16$  subtraction and  $1584 \times 15$  addition operations. In a wireless communication scenario where battery power is limited, efforts to cut down on repetitive computations can go a long way in conserving battery resources and in the proposed encoder, one area that has repetitive computations is the classification unit. One way to reduce the number of computations is to use only the boundary pixels for SAD calculation as shown in Fig. 2. In our experiments, it is seen that this method gives fairly accurate results. The number of subtraction operations reduces to  $1584 \times 12$  and the number of addition operations reduces to  $1584 \times 11$ , which is a drop of almost 24% in the number of computations. For a classification purpose, we see that the BSAD more than suffices with an error percentage of around 2–3%. This error percentage translates to a negligible distortion and a very small increase in bitrate. This is an effort to reduce the number of

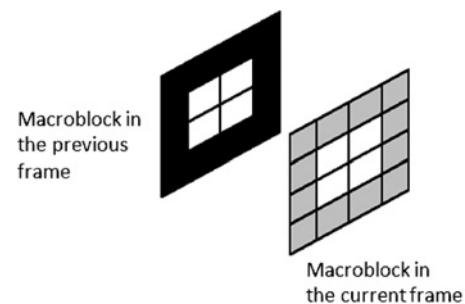


Fig. 2 Computation of BSAD

Only the boundary pixels of a  $4 \times 4$  block are taken to compute the SAD between two co-located blocks

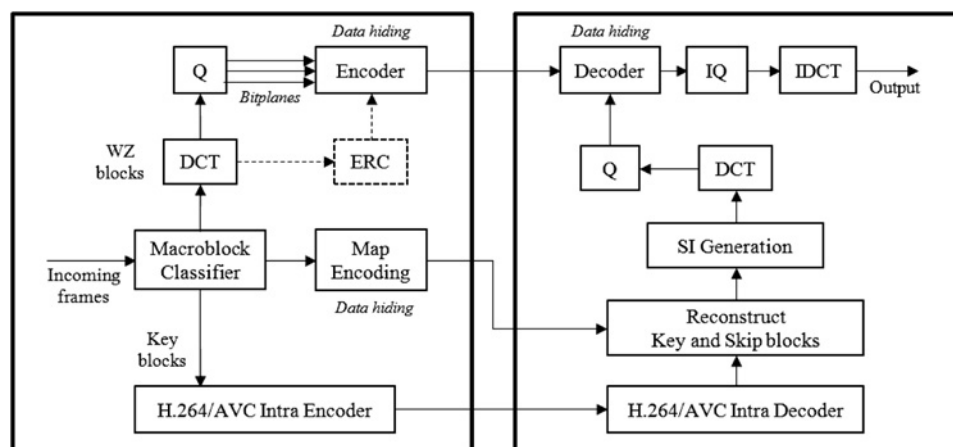


Fig. 1 Proposed architecture

computations performed and thus save power and conserve battery life. The classification is done by using

$$\text{Classification} = \begin{cases} \text{skip} & \text{if BSAD} \leq T \\ \text{key/WZ} & \text{if BSAD} > T \end{cases} \quad (1)$$

Hence, if the normalised BSAD value is less than a pre-defined threshold ( $T$ ), then the block is classified as a skip block. Otherwise, it retains its classification as either a key/WZ block. The threshold  $T$  is decided empirically after trial and error. Since initially we assumed an alternating checker-board pattern, the classification using (1) ensures that every WZ block has either key/skip blocks as part of its four-neighbourhood.

**2.1.2 Encoding of key/WZ blocks:** After classification, the key blocks are encoded using H.264/AVC's encoder in Intra mode and the skip blocks are not transmitted. The WZ blocks are transformed, quantised using a uniform scalar quantiser specified in [5]. The quantisation indices are grouped into frequency bands where each band contains the quantisation indices belonging to the same frequency from every WZ block and then each frequency band is separated into bitplanes which are compressed using an efficient and low-complexity data-hiding scheme.

The block classification map is binary in nature needing only one bit to specify a skip or key/WZ block and is compressed losslessly using a data-hiding scheme which is explained later in this paper. Additionally, if the frame is a key frame, only 1 bit needs to be transmitted to inform the decoder of this decision instead of transmitting the entire classification map, thus lowering the bitrate.

**2.1.3 Encoder rate control:** In this section, we describe the encoder rate control (ERC) scheme used in the proposed codec. The main idea for having an ERC module is to reduce the bitrate while maintaining the distortion or in other words the quality of the final image. This would translate to sending the least amount of WZ information while improving the refinement of the side information (SI) at the decoder. To make such a decision, the encoder would have to know the quality of the SI at the decoder. In the absence of the feedback channel, the encoder has to recreate the SI but it must do so in a way that is low in complexity to avoid increasing encoder complexity. In the proposed architecture, this is easily achieved since the number of WZ blocks to be estimated is quite small. However, we still use a simple method to recreate the SI at the encoder which is spatial error concealment (SEC) as discussed in Section 2.2.1. SEC is very simple and involves a weighted averaging process which gives a fairly good estimate of the SI developed at the decoder. After the estimated SI (ESI) is developed, it is transformed and quantised using a uniform scalar and the quantisation indices are grouped into frequency bands. Simultaneously, the original WZ blocks are also transformed, quantised using the same uniform scalar quantiser used on the ESI and grouped into frequency bands.

The difference between the quantisation indices of any particular frequency band of the original WZ blocks and the ESI blocks is computed as follows

$$\max\_diff_k = \max(Q_{\text{original}(k,l)} - Q_{\text{ESI}(k,l)}) \quad (2)$$

where  $k$  indicates the frequency band and  $l$  is a quantisation

index in the  $k$ th frequency band. The range of  $l$  is dependent on the number of WZ blocks. This difference is used as a rough estimate of the difference in the quality of the SI and the original. The result  $\max\_diff_k$  is then expressed in binary notation and this gives us an indication of the number of bitplanes to be transmitted to the decoder to ensure successful decoding. For example, if the maximum difference between the quantisation indices of a particular band is 4 which is  $(100)_b$ , it means that last three bitplanes have to be transmitted for all the quantisation indices of that band.

In the absence of ERC, the encoder transmits the DC band and the first three AC bands in zig-zag scan order completely. It is evident that the ERC module will reduce the bitrate, albeit increasing the complexity of the encoder by a small amount. If we consider the quantisation matrices defined by [5], then we can see that for  $Q_8$ , eight bitplanes are needed to represent the quantisation indices of the DC band, seven bitplanes for the first and second AC band and six for the third AC band. However, by applying the proposed ERC method, it is seen that in most cases the DC band need not be transmitted owing to an efficient SI generation process. Apart from this, on an average four bitplanes need to be transmitted for the remaining AC bands. This results in a substantial bitrate savings while not affecting the performance too much.

**2.1.4 Data-hiding-based compression:** In most DVC solutions, LDPCA or Turbo codes are used to transmit the WZ bits needed to refine the SI at the decoder. This is done as follows. First, the WZ information (frame or blocks depending on the architecture) is transformed using two-dimensional discrete cosine transform (DCT). The resulting transform coefficients are quantised using a uniform scalar quantiser and the quantisation indices are separated into bitplanes. Each bitplane is encoded using a channel code like LDPCA or Turbo code. Only a few of the parity bits obtained after encoding are transmitted to the decoder. At the decoder, the SI is subjected to the same steps performed on the original WZ blocks at the encoder, that is, the SI is transformed, quantised and separated into bitplanes. This is then converted into soft information by assuming the difference between the SI and the original frame to be modelled as a Laplacian PDF. Using this soft information and the parity bits received from the encoder, channel decoding is performed (LDPCA or Turbo). If the decoding fails, then more parity bits are requested from the decoder and this iterative decoding process continues till the decoding is deemed successful. It is evident that unless the SI is accurate or of high quality and the modelling of the virtual channel is accurate, the decoding procedure will need a lot of iterations and this will result in large latencies. This is a deterrent to real-time communication and must be addressed.

Another point of consideration is the block length of the input to LDPCA/Turbo codes. It is known that channel codes generally perform better as the code length increases, whereas in the proposed codec the aim has been to take full advantage of inter-frame correlation in order to skip the maximum number of blocks thus reducing the bitrate. Also, the nature of the classification scheme makes it impossible to know the number of blocks that will be classified as WZ blocks beforehand. This will in turn mean that the LDPCA/Turbo codes have to be block-length independent which will increase their complexity or will necessitate the storage of multiple tables/trees or padding the bitplanes with zeros in order to achieve a predefined block length.

To avoid the above-mentioned shortcomings, the proposed architecture uses a data-hiding-based compression scheme for the WZ blocks since it can handle any source length and involves only a comparison operation between two bits at a time and eliminates the need for storing large matrices thus improving the speed of execution and reducing hardware costs. The algorithm used in the proposed architecture is from [10], which describes a lossless compression method that is best suited for sources that have skewed probabilities. In the algorithm, an incoming bitstring  $X$  is divided into two substrings  $U$  and  $V$  based on a simple formula

$$|U| = \frac{|X|}{1 + p_0} \quad (3)$$

where  $|U|$  is the length of  $U$ ,  $|X|$  is the length of  $X$  and  $p_0$  is the probability of zeros in the string assuming that  $p_0 > p_1$ . The first  $|U|$  bits of  $X$  are copied into  $U$  and the remaining into  $V$ . Then based on a simple comparison operation between  $U$  and  $V$ , two new substrings  $L$  and  $M$  are created, which are enough to recover the entire original sequence. Thus, the payload consists only of  $L$  and  $M$  which is usually 50–60% the length of the original string. In the proposed codec, the data-hiding scheme is used to encode the WZ blocks and also to compress the block classification map thus achieving a high rate of compression while keeping complexity and time delay to an absolute minimum.

## 2.2 Decoder

**2.2.1 Reconstruction of key and skip blocks:** The first step in the decoding process is to decode the compressed block classification map. This map is important as it contains the location of key, WZ and skip blocks. After decoding the classification map, the H.264/AVC Intra-encoded bitstream is decoded using the H.264/AVC Intra decoder and the key blocks are reconstructed. Since skip blocks are those blocks that do not vary significantly from their co-located blocks in the previous frame, we copy the co-located blocks from the previously fully reconstructed frame into the skip block locations as dictated by the classification map. After completing these two operations, we are left with a partial frame consisting of only key and skip blocks with the WZ blocks needing to be reconstructed. The process of reconstructing the WZ blocks is referred to as SI generation and this is explained in the next section.

**2.2.2 Side information generation:** SI generation, as we mentioned is the process of reconstructing the WZ blocks using temporal and spatial information, available at the decoder. Being a low-delay DVC, the decoder cannot wait for the future frame to arrive before starting the SI generation process and so it must use only the previously reconstructed frame for temporal information. Since the proposed DVC method is block-based architecture and it uses the unique block classification scheme, the decoder is ensured that every WZ block is surrounded by either a key or a skip block in its four-neighbourhood, thus enabling better SI generation because of the presence of more spatially correlated information.

Given the temporal and spatial information, SI generation reduces to the task of error concealment (EC) where one assumes that the WZ blocks are blocks that were lost during

transmission and now have to be concealed using well-known EC techniques. For this purpose, we use a mode-selection-based EC [11] that selects between temporal EC (TEC) and SEC based on a cost criterion. At first, using boundary matching-based motion estimation, a matching block is found for the WZ block in the previous frame. After the match is found, we calculate the temporal activity (TA) given by

$$TA = E[(x - x^*)^2] \quad (4)$$

where  $x$  is boundary pixels of the WZ block in the current frame and  $x^*$  is the boundary pixels of the matching block in the previous frame. Next, we find the spatial activity (SA) given by

$$SA = E[(x - \mu)^2] \quad (5)$$

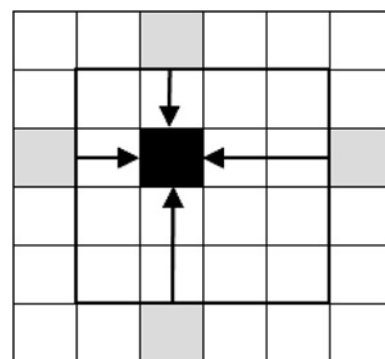
where  $x$  is boundary pixels of the WZ block in the current frame and  $\mu$  is the mean of the boundary pixels of the WZ block in the current frame. After computing SA and TA, we perform the mode selection given by

$$\text{If } (TA < SA), \text{ apply TEC; else apply SEC} \quad (6)$$

As mentioned earlier, if TEC is chosen we use the matching block from the previous frame to compensate for the WZ block. However, if SEC is chosen, we use bilinear EC to conceal the WZ block. Bilinear EC uses the pixels from the neighbouring blocks to recreate the missing pixels using a weighted averaging procedure. Let  $i$  and  $j$  represent the vertical and horizontal coordinates of the WZ block and  $(0 \leq i < 4)$  and  $(0 \leq j < 4)$  assuming that we are using  $4 \times 4$  blocks. Let  $T(j)$  and  $B(j)$  be the pixels on the top and bottom boundaries of the lost block and  $L(i)$  and  $R(i)$  be the pixels on the left and right boundaries. These are shown as the grey pixels in Fig. 3.

If  $mb_s$  is the lost pixel which is to be estimated, then the replacement pixel is obtained using (7) and (8) where the weights are inversely proportional to their distance from the lost pixel

$$mb_s = \frac{w_T(i)T(j) + w_B(i)B(j) + w_L(j)L(i) + w_R(j)R(i)}{w_T(i) + w_B(i) + w_L(j) + w_R(j)} \quad (7)$$



**Fig. 3** Lost pixel (black) with the four nearest pixels from neighbouring blocks used to conceal it

where the weights  $w_T(i)$ ,  $w_B(i)$ ,  $w_L(j)$ ,  $w_R(j)$  are defined as

$$\begin{aligned} w_T(i) &= 4 - i \\ w_B(i) &= i + 1 \\ w_L(j) &= 4 - j \\ w_R(j) &= j + 1 \end{aligned} \quad (8)$$

After the EC process, we use overlapped block motion compensation (OBMC) for refining the SI [11]. This algorithm uses the motion vectors of the four-neighbourhood of each WZ block to further refine the estimated pixels. This concept has been used earlier in H.263+ to obtain almost 1–1.5 dB improvement in the quality of the estimates. This is particularly true when there is high temporal correlation between frames.

After applying OBMC, the SI generation process concludes and we now have an estimate of the original WZ blocks at the decoder. To visualise the decoding process till this point, Fig. 10 can be seen. In Fig. 10, we demonstrate the decoding process for four video sequences, namely – Hall Monitor, Mother–Daughter, Miss America and Akiyo. These video sequences are characteristic of low motion scenarios which will be encountered in practical applications like video conferencing and monitoring of hallways/surveillance. In each row, we show the original frame received at the encoder, the key blocks reconstructed at the decoder, partial frame consisting of key and skip blocks and the frame with WZ blocks reconstructed using SI generation methods explained earlier.

It is interesting to note that hardly 30–40 blocks are classified as WZ blocks in each frame thus reducing the time taken by the decoder to reconstruct the WZ blocks. This reduction in latency is an important factor in achieving real-time communication. Also, upon close examination, it can be deduced that blocks lying on naturally occurring edges in the image or blocks that exhibit significant difference from their collocated blocks in the previous frame (especially blocks around the facial regions when a person is speaking) or noisy regions are chosen as WZ blocks and the remaining blocks are skipped. This simple classification translates to a large savings in terms of bitrate and power because only a few blocks have to be processed by the H.264/AVC encoder and decoder thus saving power or battery life and in turn resulting in a shorter bitstream.

**2.2.3 Refining the SI:** At the encoder, the WZ blocks are transformed using 2D DCT and then quantised using a uniform scalar quantiser. The quantisation indices are grouped into frequency bands and each band is separated into bitplanes. The bitplanes are compressed using the data-hiding-based compression scheme and transmitted to the decoder. If ERC is enabled then a decision is made regarding the number of bitplanes to be sent for each band (or a pre-defined number of frequency bands). Otherwise, all the bitplanes of the DC and the first three AC bands (in zig-zag order) are compressed and sent to the decoder. At the decoder, the SI is transformed, quantised, grouped in frequency bands and each band is separated into bitplanes. Using the information received from the encoder, the bitplanes are refined using the data-hiding-based compression scheme. If a bit-plane or a band is not sent by the encoder, then the corresponding bit-plane or band of the SI is retained as it is. The proposed refinement scheme is very simple and does not increase the system latency.

**2.2.4 Final frame reconstruction:** After the SI's quantisation indices have been refined, first the quantisation indices are subjected to inverse quantisation to obtain DCT coefficients and then the coefficients are subjected to inverse DCT operations to obtain the final refined pixel values. These are then used to reconstruct the final frame.

### 3 Simulation results

#### 3.1 Test conditions

In this section, we discuss the simulation results and demonstrate the performance of the proposed architecture in comparison with other codecs. For comparison purposes of the performance with other DVC architectures, we take the current state-of-the-art frame-based DVC solution – the DISCOVER codec, the MLWZ codec and the best-performing block-based architecture SPI-DVC. For comparison with conventional coding, we take H.264/AVC operating in the Intra mode and in the no-motion mode and also H.263+ operating in the Intra mode as standards. The comparison chosen is standard and common to most current DVC publications. Since the proposed codec is aimed at low-motion scenarios like monitoring of hallways and video conferencing, we chose the standard test video sequences like Hall Monitor, Akiyo, Mother–Daughter and Miss America which all exhibit such characteristics. Each of the test sequences are of QCIF resolution ( $176 \times 144$  pixels) and sampled at a rate of 15 frames/s. Only the luminance component is taken into consideration for the RD performance and the bitrate consumed for the encoding the classification map is also considered in the results.

#### 3.2 Efficiency of block classification

An integral part of the proposed architecture is the classification step at the encoder. By making maximum use of the temporal correlation between frames, the codec reduces the complexity of the encoder and the decoder and also improves the RD performance. Since inter-frame correlation is quite high in low-motion scenarios, skipping of blocks has the potential of reducing the bitrate drastically. To demonstrate this we use the Hall Monitor sequence and the first 12 frames of the sequence are analysed with a GOP size of 2. In our implementation, the block size is fixed at  $4 \times 4$  and the frame size is  $176 \times 144$  pixels (QCIF) giving a total of 1584 blocks/frame. Table 1 shows the percentage of blocks skipped in every WZ frame. Recollecting the process of classification, the blocks of a WZ frame are compared with their co-located blocks in the previous frame and compared with a pre-defined threshold  $T$  which is taken as 0.15 in our implementation. If the BSAD for a particular block is less than  $T$ , then it is skipped.

**Table 1** Number of blocks skipped per frame

Frame number	Number of blocks skipped	Percentage skipped
2	1558	98.35
4	1550	97.85
6	1550	97.85
8	1562	98.61
10	1546	97.6
12	1537	97.00

It can be seen that nearly 95% of the blocks in each of the frames of the sequence is skipped thus reducing the bitrate and the computational burden on the encoder and decoder, in turn translating to saving in power and battery life.

### 3.3 RD performance for low-motion sequences

To evaluate the codec in low-motion scenarios we use the Hall Monitor sequence as a test sequence. We compare the RD performance with the DISCOVER codec [5], MLWZ codec [12], SPIDVC [7] (for GOP = 2) and with H.263+ (Intra), H.264/AVC (Intra) and H.264/AVC (no motion). The Hall Monitor sequence used is of QCIF format sampled at 15 frame/s. and the comparison is made at GOP sizes of 2 and 4 for the first 75 frames. The RD performance is shown in Figs. 4 and 5. Results for SPIDVC are not available for GOP = 4. The performance of the proposed codec surpasses that of H.263+ (Intra) and H.264/AVC (Intra) but is behind that of H.264/AVC (no motion) like all other current DVC solutions. In comparison with other DVC architectures, its performance is better than that of DISCOVER and MLWZ at most bitrates and surpasses that of SPIDVC which is considered to be a state-of-the-art block-based DVC codec. The performance of the proposed codec is good considering the fact that DISCOVER, MLWZ SPIDVC use bi-directional motion estimation and have a feedback channel making them more complex.

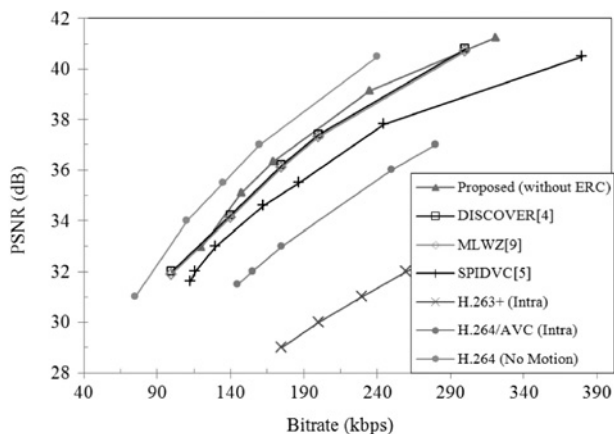


Fig. 4 RD performance for Hall Monitor at GOP = 2

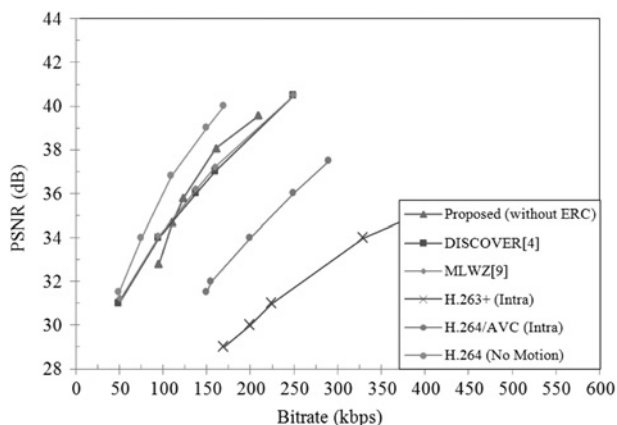


Fig. 5 RD performance for Hall Monitor at GOP = 4

### 3.4 RD performance for head-and-shoulder sequences

As we mentioned earlier, an important application of DVC is in the area of video conferencing or access control, where one can see and communicate with the person seeking access. Such scenarios typically fall under the head-and-

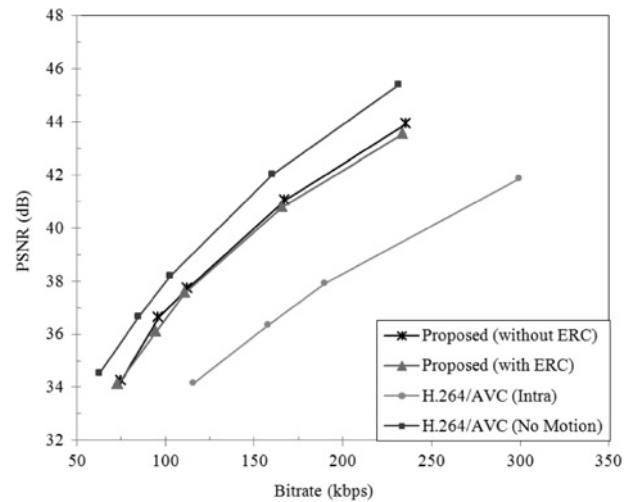


Fig. 6 RD performance for Akiyo at GOP = 2

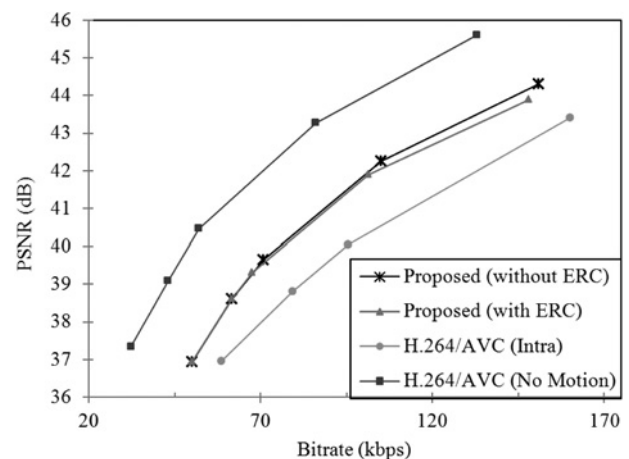


Fig. 7 RD performance for Miss America at GOP = 2

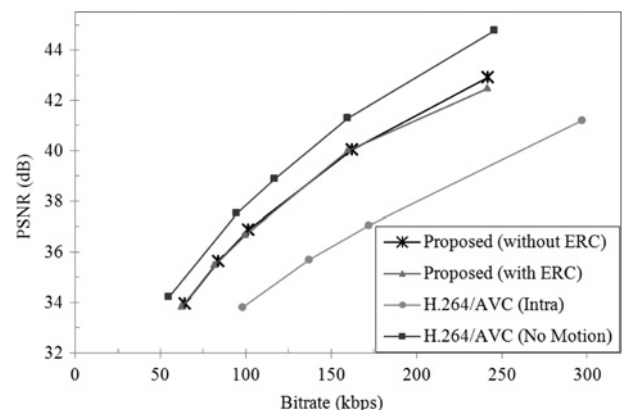


Fig. 8 RD performance for Mother-Daughter at GOP = 2

shoulder-type video sequences and to demonstrate the RD performance of the proposed codec we compare the performance of the proposed codec with H.264/AVC (Intra) and H.264/AVC (no motion) for the Mother–Daughter, Akiyo and Miss America sequences. The sequences are of QCIF size and sampled at 15 frames/s. The first 50 frames are considered and bitrates for the classification map and for luminance components are considered. The results are shown in Figs. 6–8. It is clearly seen that the proposed codec outperforms H.264/AVC (Intra) at all bitrates but lags behind the performance of H.264/AVC (no motion) like most current DVC solutions.

### 3.5 Decoding complexity comparison

We compare the time taken by the proposed codec to encode and decode a frame with other codecs. Since the aim is to facilitate real-time communication, the codec should be able to encode and decode a frame quickly.

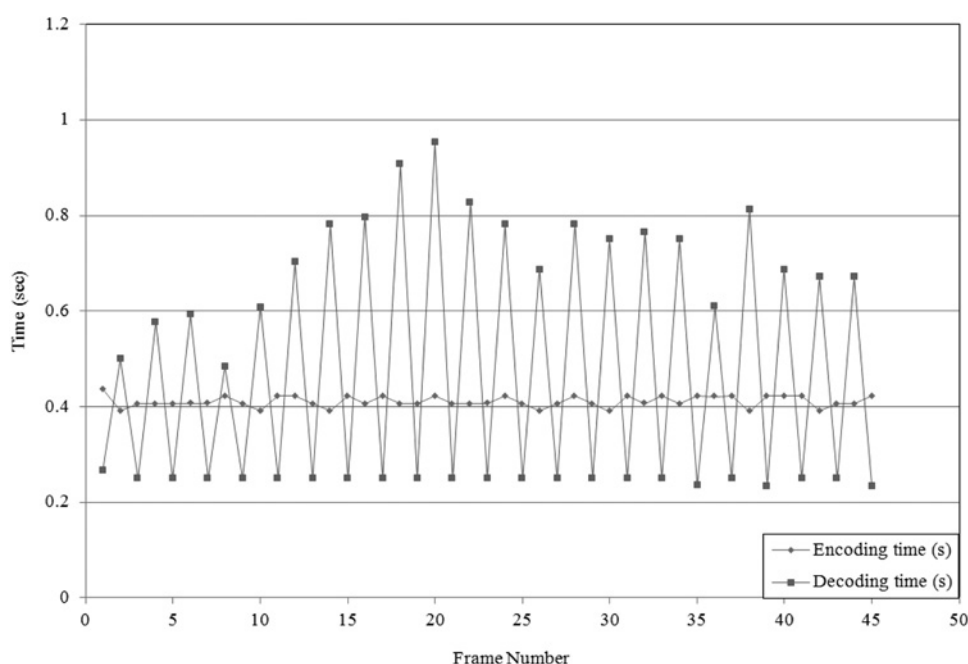
Conventional codecs take a long time to encode a frame but decode very quickly. On the other hand, DVC solutions like DISCOVER take very little time to encode but a lot of time to decode a frame. This is because of the iterative nature of the WZ channel decoding. In the proposed codec, there is no feedback channel and the iteratively decodable channel codes are replaced with a simple data-hiding-based compression scheme. This greatly reduces the time needed to decode a frame. We compare the proposed codec's decoding performance with that of DISCOVER codec by referring to the performance evaluation made public by the DISCOVER group at [13]. The DISCOVER group used a x86 machine with a dual-core Pentium D processor at 3.4 GHz with 2 GB of RAM and we have used a x86 machine with a dual-core Core 2(Duo) processor at 2.80 GHz and 3 GB of RAM. The decoding time comparison can be seen in Table 2. It can be seen that to decode the entire Hall Monitor sequence at QP = 24 and

**Table 2** Time complexity comparison

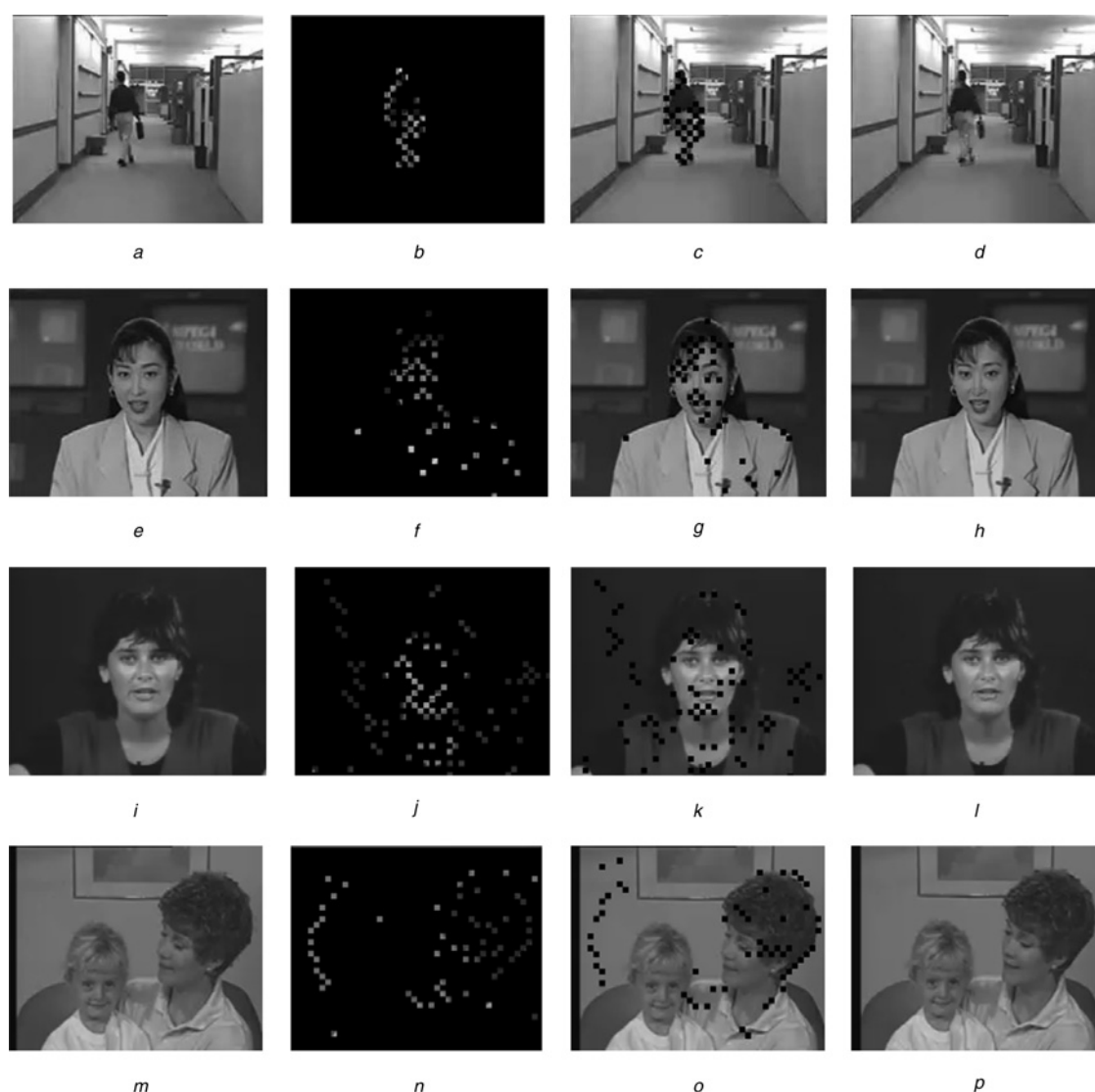
QP	DISCOVER (s) time needed to decode	Proposed (s) time needed to encode and decode
36	391.14	134.86
31	732	136.39
24	1210.47	144.18

GOP = 2, the DISCOVER codec takes roughly 1210 s to decode the Hall Monitor sequence whereas the proposed codec takes hardly 145 s to perform both encoding and decoding. There is a very large improvement in decoding speed with respect to DISCOVER while still achieving better RD performance.

We can also see the encoding and decoding times for 45 frames of the Hall Monitor sequence with ERC in Fig. 9. It can be seen that both the encoding and decoding take less than a second to complete. It can be seen that every alternate frame has a decoding time of roughly 0.275 s whereas the other frame have a higher decoding time. This is because the first frame of every GOP is coded using H.264/AVC which is known to have very high decoding speed. The next frame being a WZ frame has to undergo the process of SI generation and refinement and thus takes more time to be decoded. Encoding is done fairly quickly and both the key and WZ encoder take approximately the same amount of time to encode the information. However, in comparison with other DVC architectures, the proposed architecture is much faster at the same RD performance. Surely, with dedicated hardware, optimised programming and with hardware–software co-design the performance can be improved and can prove to be a viable solution for smart spaces and wireless multimedia sensor networks.



**Fig. 9** Encoding and decoding time for the first 45 frames of Hall Monitor with GOP = 2 and QP = 20



**Fig. 10** Reconstruction of a single frame: first column contains the original frame; second column shows the key blocks reconstructed at the decoder; third column shows the partial frame consisting of key and skip blocks; fourth column shows the frame after WZ blocks have been reconstructed

a–d Hall Monitor

e–h Akiyo

i–l Miss America

m–p Mother–Daughter

## 4 Conclusion

In this paper, we propose a novel low-delay and low-bitrate DVC architecture which is suited for application in the smart spaces concept and wireless multimedia sensor networks. The codec is block based, unidirectional, and makes efficient usage of skip blocks to take advantage of the temporal and spatial correlation in low-motion scenarios. The performance of the proposed codec surpasses that of current DVC solutions and also conventional state-of-the-art codecs like H.264/AVC and H.263+ in the Intra mode. By using data-hiding-based compression scheme, the codec is able to achieve real-time encoding–decoding and very low latency. With more optimised programming, the encoding–decoding speed can be increased. A very simple ERC scheme is also proposed that aims at reducing the bitrate while not affecting the distortion. Although there is a drop in performance at very high peak signal-to-noise ratio

(PSNR) values, the ERC scheme works satisfactorily at lower PSNR values and succeeds in reducing the number of bitplanes to be transmitted, thus conserving battery resources. Future work involves extending this scheme to higher-motion scenarios and optimising the code.

## 5 References

- 1 Slepian, D., Wolf, J.: ‘Noiseless coding of correlated information sources’, *IEEE Trans. Inf. Theory*, 1973, **19**, (4), pp. 471–480
- 2 Wyner, A., Ziv, A.: ‘The rate-distortion function for source coding with side information at the decoder’, *IEEE Trans. Inf. Theory*, 1976, **22**, (1), pp. 1–10
- 3 Girod, B., Aaron, A.M., Rane, S., Rebollo-Monedero, D.: ‘Distributed video coding’, *Proc. IEEE*, 2005, **93**, (1), pp. 71–83
- 4 Varodayan, D., Aaron, A., Girod, B.: ‘Rate-Adaptive distributed source coding using low-density parity-check code’. Conf. Record 39 Asilomar on Signals, Systems and Computers, November 2005, vol. 28, pp. 1203–1207



- 5 Brites, C., Ascenso, J., Pereira, F.: 'Improving transform domain Wyner-Ziv video coding performance'. IEEE Int. Conf. on Acoustics, Speech and Signal Process, (ICASSP 2006), May 2006, vol. 2, pp. 525–528
- 6 Brites, C., Pereira, F.: 'Encoder rate control for transform domain Wyner-Ziv video coding'. IEEE Int. Conf. on Image Processing (ICIP 2007), October 2007, vol. 2, pp. II-5–II-8
- 7 Anantrasirichai, N., Agrafiotis, D., Bull, D.: 'Enhanced spatially interleaved DVC using diversity and selective feedback'. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, (ICASSP 2009), April 2009, pp. 717–720
- 8 Puri, R., Majumdar, A., Ramchandran, K.: 'PRISM: a video coding paradigm with motion estimation at the decoder', *IEEE Trans. Image Process.*, 2007, **16**, (10), pp. 2436–2448
- 9 Liu, H., Li, Y., Liu, X., Ma, M., Zhao, D., Gao, W.: 'Improved low delay distributed video coding'. Picture Coding Symp. (PCS 2009), August 2009, pp. 1–4
- 10 Kim, H.J.: 'A new lossless data compression method'. IEEE Int. Conf. on Multimedia and Exposition, (ICME 2009), June 2009, pp. 1740–1743
- 11 Agrafiotis, D., Bull, D.R., Canagarajah, C.N.: 'Enhanced error concealment with mode selection', *IEEE Trans. Circuits Syst. Video Technol.*, 2006, **16**, (8), pp. 960–997
- 12 Martins, R., Brites, C., Ascenso, J., Pereira, F.: 'Statistical motion learning for improved transform domain Wyner-Ziv video coding', *IET Image Process.*, 2010, **4**, (1), pp. 28–41
- 13 [http://www.img.lx.it.pt/~discover/dec\\_time\\_complexity\\_gop248.html](http://www.img.lx.it.pt/~discover/dec_time_complexity_gop248.html)

Copyright of IET Communications is the property of Institution of Engineering & Technology and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.