Vis Comput (2009) 25: 69–82
DOI 10.1007/s00371-008-0279-z

ORIGINAL ARTICLE

Distributed Video Coding using Turbo Trellis Coded Modulation

J.L. Martinez · W.A.R.J. Weerakkody · W.A.C. Fernando · G. Fernandez-Escribano · H. Kalva · A. Garrido

Published online: 28 August 2008
© Springer-Verlag 2008

Abstract Distributed Video Coding (DVC) has been proposed for increasingly new application domains. This rise is apparently motivated by the very attractive features of its flexibility for building very low cost video encoders and the very high built-in error resilience when applied over noisy communication channels. Yet, the compression efficiency of DVC is notably lagging behind the state-of-the-art in video coding and compression, H.264/AVC in particular. In this context, a novel coding solution for DVC is presented in this paper, which promises to improve its rate-distortion (RD) performance towards the state-of-the-art. Here, Turbo Trellis Coded Modulation (TTCM), with its attractive coding gain in channel coding, is utilized and its resultant impact in both pixel domain and transform domain DVC framework is discussed herein. Simulations have shown a significant gain in the RD performance when compared with the state-of-the-art Turbo coding based DVC implementations.

Keywords Algebraic triangulation · Partition of unity implicits · Orthogonal polynomials

J.L. Martinez (✉) · G. Fernandez-Escribano · A. Garrido
Instituto de Investigacion en Informatica, University of Castilla-La Mancha, Campus Universitario s/n., 02071 Albacete, Spain
e-mail: joseluismm@dsi.uclm.es

G. Fernandez-Escribano
e-mail: gerardo@dsi.uclm.es

A. Garrido
e-mail: antonio@dsi.uclm.es

W.A.R.J. Weerakkody · W.A.C. Fernando
Center for Communications Research, University of Surrey, Guildford GU2 7XH, UK

W.A.R.J. Weerakkody
e-mail: r.weerakkody@surrey.ac.uk

W.A.C. Fernando
e-mail: w.fernando@surrey.ac.uk

H. Kalva
Department of Computer Science and Engineering, Florida Atlantic University, 777 Glades Road, Boca Raton, 33431 FL, USA
e-mail: hari@cse.fau.edu

1 Introduction

Video coding technologies have evolved significantly during the past decade, dominated by the work on ISO/IEC MPEG and ITU-T H.26x standards based techniques. These approaches are characterized by a highly complex video encoder structure and a significantly less complex decoder structure as demanded by many popular applications involving one-to-many topologies. Video capturing and encoding was conventionally limited compared to the largely spread viewers (decoders) of the mostly broadcast type services. However, this concept is increasingly challenged by the emerging and progressively more popular new paradigm of applications involving more encoders than decoders, such as multimedia wireless sensor networks, wireless video surveillance, disposable video cameras, medical applications, and mobile camera phones. For all the applications mentioned above, we need to have a low complex encoder probably at the expense of a high complexity in the decoder. The major task of exploiting the source redundancies to achieve the video compression is accordingly placed in the decoder. The DVC encoder thus performs a computationally very inexpensive operation enabling a drastically low cost implementation of the signal processor in DVC based video cameras. Matching the rate-distortion (RD) performance of the standardized MPEG and H.26x based systems is a challenge ahead of the researchers promoting DVC.

The Slepian–Wolf theorem [1] and Wyner–Ziv model [2] were the basis for the DVC. Current practical schemes developed for DVC are based, in general, on the following principles: the video frames are organized into two types; Key frames and Wyner–Ziv frames, while the key frames are encoded with a conventional intraframe codec, the frames between them are Wyner–Ziv encoded. At the decoder, the side information is obtained using previously decoded key frames and a motion interpolation scheme, responsible for obtaining the most accurate representation of the original frame. The rate–distortion performance of the DVC codec has a large bearing on the quality of the side information stream since the parity input to the turbo coding based DVC decoder suffers a significant loss in information content due to the puncturing process carried out as a mechanism for achieving video compression. The quality of the side information stream is typically determined by its closeness to the corresponding original Wyner–Ziv frame used at the DVC encoder.

Based on this theoretical framework, several DVC codecs have been proposed recently [3, 4] in the pixel domain using traditional turbo coding. The distributed video coding paradigm may be also applied in the transform domain. The RD performance of the DVC pixel domain codecs can be further improved by using a transform coding tool with the same purpose as in traditional video coding, i.e., to exploit spatial correlations between neighboring sample values and to compact the block energy into as few transform coefficients as possible. Several proposals have been reported in the literature aiming to implement different transform coding tools, using the well-known Discrete Cosine Transform (DCT) [5, 6]. However, there still exists a significant gap between the RD performance of these proposals and the standardized MPEG and H.26x based systems.

In this work, we present an improved transform domain Wyner–Ziv video codec based on Turbo Trellis Coded Modulation (TTCM) with significant rate–distortion (RD) gains compared to the state-of-the-art results available in the literature. As in the conventional turbo based Wyner–Ziv encoder, transform quantized coefficients are applied to the TTCM encoder, and parity bits are generated from both constituent encoders. However, TTCM symbols are not generated at the encoder since they are not sent to the decoder. Parity bits produced by the TTCM encoder are stored in a buffer and transmitted to the decoder upon request. TTCM symbols are generated at the decoder and these symbols are passed to the TTCM decoder for demodulation.

This paper is organized as follows. Section 2 presents a description of theoretical background of DVC. Section 3 introduces a review of the state-of-the-art of different channel coding techniques and architectures in DVC proposed in the literature. Section 4 presents our channel coding approach and its implementation in DVC architecture. In Sect. 5, some experimental results are presented to evaluate the performance of our proposal with different scenarios. Finally, conclusions are presented in Sect. 6.

2 DVC theoretical background

In video coding, as standardized by MPEG or the ITU-T H.26x recommendations, the encoder exploits the statistics of the source signal. However, efficient compression can also be achieved by exploiting source statistics (partially or completely) at the decoder only. This is consequence of information-theoretic bounds established in the 1970s by Slepian and Wolf for distributed lossless coding [1], and by Wyner and Ziv for lossy coding with side information [2]. The next two sections describe these two theorems which represent the mathematical background of DVC.

2.1 Slepian–Wolf coding

Distributed compression refers to the coding of two or more dependent random sequences where each one is coded independently. Let X and Y be two *independent identically distributed* (i.i.d.) binary correlated sources to be losslessly encoded. In joint coding where the encoder has access to both signals, it is well-known from Shannon's theory that the minimum lossless rate for X and Y is given by the joint entropy $H(X, Y)$. The Slepian–Wolf [1] theorem established that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and their correlation is known to both the encoder and the decoder defined by (1):

$$\begin{aligned} &R_X + R_Y \geq H(X, Y), \\ &R_X \geq H(X \mid Y),\ R_Y \geq H(Y \mid X), \end{aligned} \tag{1}$$

where $H(X \mid Y)$ and $H(Y \mid X)$ denote the conditional entropies between two sources. Surprisingly, the sum of rates $R_x + R_y$ can achieve the joint entropy $H(X, Y)$, just as for joint coding. Let us consider the particular case of distributed source coding that occurs when Y is available at the decoder and has been coded separately (with respect to X) with a rate equal to $R_Y = H(Y)$ (depicted in Fig. 1). On the other hand, according to Slepian–Wolf Theorem, the source X can be coded lossless with a rate asymptotically equal to $H(X \mid Y)$. Therefore, the minimum total rate for the two sources is $H(X \mid Y) + H(Y) = H(X, Y)$. This rate, ideally, would be the same as the rate used where the encoder has access to both signals.

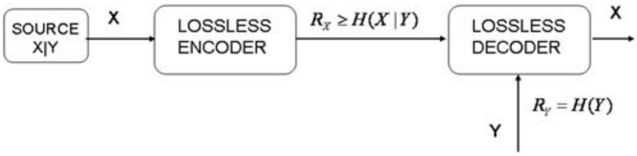


Fig. 1 Compression of a sequence X with side information Y available at the decoder only

A particular case of distributed source coding is Distributed Video Coding where the source is a video sequence. In this kind of architecture, Y may be provided to the decoder using a traditional coding solution such as the MPEG-X or H.26x intra coding standard. The dilemma would be then how to encode X in order to reach optimal performance in terms of rate. The goal is to build a system that is able to approach the rate point $(H(X \mid Y), H(Y))$ [1].

2.2 Wyner–Ziv coding

Wyner and Ziv extended the Slepian's and Wolf's work to establish information-theoretic bounds for lossy compression with *side information* (SI) at the decoder. Wyner and Ziv considered the problem of coding of two correlated sources X and Y with respect to a fidelity criterion [2]. They have established the rate–distortion (RD) function $R_{*X|Y}(D)$ for the case where the SI Y is perfectly known to the decoder only. For a given target distortion D, $R_{*X|Y}(D)$, in general, verifies$R_{X|Y}(D) \leq R_{*X|Y}(D) \leq R_X(D)$, where $R_{X|Y}(D)$ is the rate required to encode X if Y is available to both the encoder and the decoder, and R_X is the minimal rate for encoding X without SI. Wyner and Ziv have shown that, for correlated Gaussian sources and a mean square error distortion measure, there is no rate loss with respect to joint coding and joint decoding of the two sources, i.e., $R_{*X|Y}(D) = R_{X|Y}(D)$ [2].

3 Related work

Although the mathematical background of DVC was proposed in the 1970s, only recently emerging applications have motivated practical attempts. The correlation between X and the side information Y is modeled as a virtual channel, where Y is regarded as a noisy version of X. Channel capacity achieving codes, block codes [3], *turbo codes* (TC) [7], or *Low Density Parity Check* (LDPC) codes [8] have been able to achieved the rate point depicted in the Slepian–Wolf theorem (see Sect. 2). The compression of X is achieved by transmitting only a binary index [3], or parity bits [7, 8]. The decoder corrects the virtual channel noise, and thus estimates X given the received parity bits, or index, and the SI Y regarded as a noisy version of the codeword systematic bits. The test results show that channel coding performed after quantization allows approaching the theoretical bounds, i.e., the Slepian–Wolf and the Wyner–Ziv limits. The following sections describe possible solutions for channel coding in DVC (Sect. 3.2) and how one can get this performance in practical architectures (Sect. 3.1) of Wyner–Ziv video approach in both pixel and transform domain.

3.1 Architectures of DVC

3.1.1 Pixel domain Wyner–Ziv video coding

Figure 2 illustrates the more common architecture of the video codec based on pixel domain; this video codec solution follows the same architecture as the one proposed by Aaron et al. in [4]. In a nutshell, the coding process is as follows: the video frames are organized into key frames and Wyner–Ziv frames. The key frames are traditionally intraframe coded. The Wyner–Ziv frame pixel values are quantized using a 2^M-level uniform scalar quantizer; in this case, $2^M \in \{2, 4, 8, 16\}$. Over the resulting quantized symbol stream, bitplane extraction is performed. Each bitplane is then independently turbo encoded, starting with the most significant one. The parity bits produced by the turbo encoder are stored in the buffer and transmitted in small amounts upon decoder request via the feedback channel; the systematic bits are discarded. At the decoder, the frame interpolation module is used to generate the side information frame, an estimate of the WZ frame, based on previously decoded frames, X_{2-1} and X_{2+1} (see Fig. 2). The side information is used by an iterative turbo decoder to obtain the decoded quantized symbol stream. The turbo decoder is constituted by two soft-input soft-output (SISO) decoders; each SISO decoder is implemented using the *Maximum A Posteriori* (MAP) algorithm. It is assumed that the decoder has ideal error detection capabilities regarding the current bitplane error probability, the decoder requests for more parity bits from the encoder via feedback channel; otherwise, the current bitplane turbo decoding task is considered successful and another bitplane starts being turbo decoded.

3.1.2 Transform domain Wyner–Ziv video coding

In a non-distributed source coding scenario, transform coding is another source coding technique used to reduce the transmission rate. Typically, the energy of a frame is stored only in a few significant coefficients which need to be transmitted, reducing the bit rate: the remaining coefficients do not offer a major impact into the reproduced image quality. In DVC, transform domain tools have been also introduced in order to exploit the spatial redundancies exhibited in a video frame in a similar way of traditional video coding. Several proposals have been reported in the literature aiming

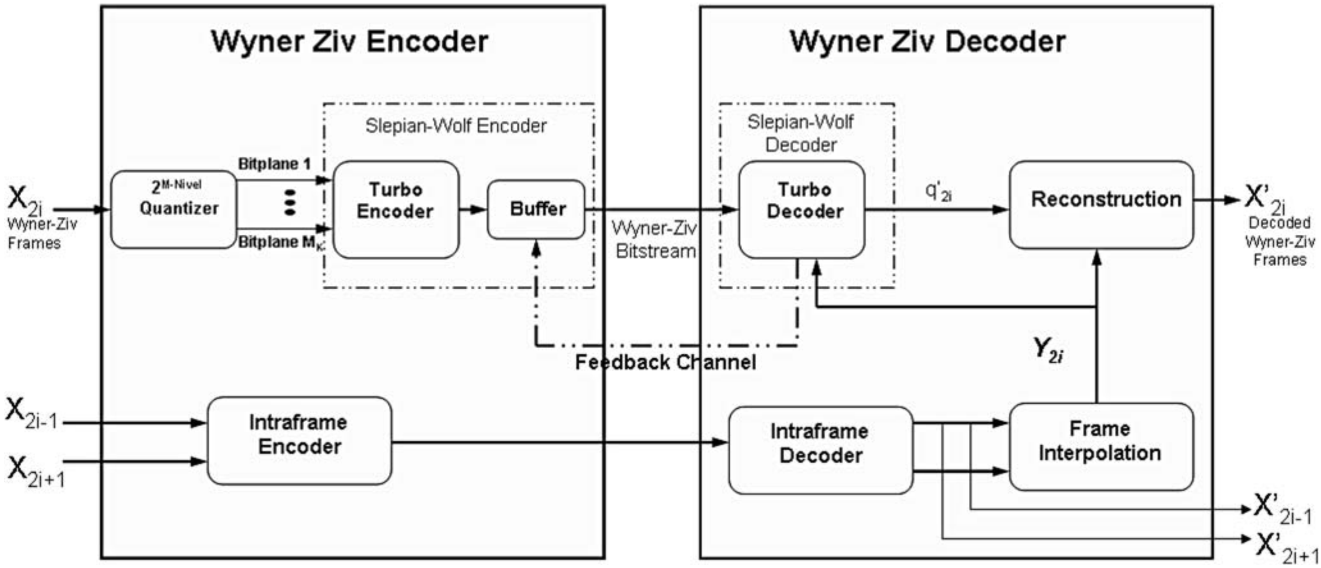


Fig. 2 Traditional pixel domain architecture

to implement different transform coding tools. In the following paragraphs, some of the most prominent ones which use DCT and DWT are introduced.

The first architecture of DVC working in transform domain was proposed by Aaron et al. in [5] and it is based on turbo codes. In this system, the DCT is applied before to quantization and each band of coefficients is encoded independently using a turbo coder. The decoder uses previously reconstructed frames to generate side information to conditionally decode the Wyner–Ziv frames. Simulation results show significant gains above DCT-based intraframe coding and improvements over the pixel-domain Wyner–Ziv video coder [5].

Brites et al. extended the Aaron et al. work [6] proposing an improved transform domain Wyner–Ziv video codec using the integer block-based transform defined in the H.264/MPEG-4 AVC standard, quantizer with a symmetrical interval around zero for AC coefficients, a quantization step size adjusted to the transform coefficient bands dynamic range, and advanced frame interpolation for side information generation.

On the other hand, there exist more highlighted architectures in the literature which make use of the transform tool to improve the performance to the pixel domain and which are based on Discrete Wavelet Transform instead of DCT. It has been proved that DWT can overcome the 'block-effect' brought by block-wise DCT and achieve better coding performance in image coding. Wang et al. [9] proposed a DVC paradigm based on lattice vector quantization in wavelet domain. In this scheme, the authors use a fine and a coarse lattice vector quantizer to wavelet coefficients, and the difference of two lattice quantizer is coded by turbo encoder which is different from the one given in [5, 6] based on scalar quantization. At the decoder, side information is gradually updated by motion-compensated refinement. Bernardini et al. [10] have proposed another wavelet domain distributed coder for video which allows scalability and does not require a feedback channel. Efficient distributed coding is obtained by processing the wavelet transform with a suitable folding function and compressing the folded coefficients with a wavelet coder. At the receiver side, the authors use the statistical properties between similar frames to recover the compressed frame. Experimental results show that the proposed scheme has good performance when compared with similar asymmetric video compression schemes.

3.2 Channel coding techniques for DVC

Turbo coding is a channel coding technique widely appreciated for use in DVC. A turbo encoder is formed by parallel concatenation of two *recursive systematic convolutional* (RSC) encoders separated by an interleaver, as illustrated in Fig. 3.

The construction of an RSC encoder is determined by the generator polynomial which takes the form: $G(D) = [1, g2(D)/g1(D)]$, where $g1(D)$ and $g2(D)$ are feedforward and feedback polynomials, respectively. Figure 4 illustrates an example of RSC encoder with the generator polynomial $(1, 13/15)$ in octal form.

In the context of DVC, the turbo decoder plays the key role of correcting the errors in the side information stream, which is considered to resemble a Laplacian noise model when compared with the original Wyner–Ziv frame. The parity bit stream received from the encoder is used in the turbo decoder for achieving the above purpose. Turbo coding was proposed by Berrou et al. in 1993 for channel coding

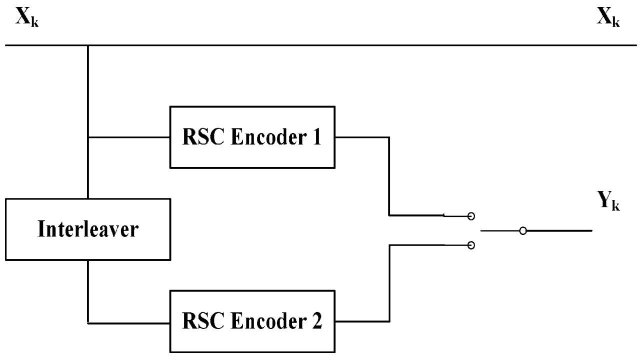


Fig. 3 Turbo Encoder

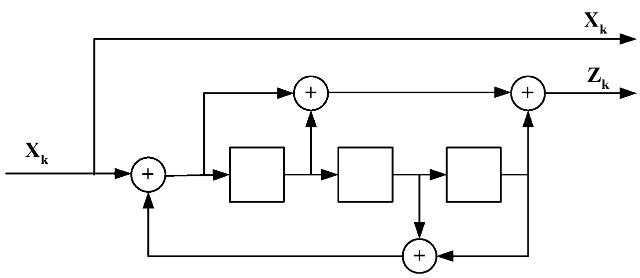


Fig. 4 Example of an RSC Encoder

in communications [11]. This concept has been successfully adopted for DVC. The structure of the turbo decoder is illustrated in Fig. 5.

Soft channel outputs containing received parity bits (L_{cykl1}) from the first encoder and the systematic bits (side information—L_{cyks}) is fed into SISO Decoder 1. In case of rate compatible punctured turbo (RCPT) codes, parity bits are punctured, thus on the receiver side, zeros are inserted into the punctured positions. At the first iteration, there is no a priori information about the sent bits, thus log likelihood ratio (LLR) $L(uk)$ is initialized to 0. After the first iteration of SISO Decoder 1, the LLR $Le(uk)$ of bits are interleaved and become the a priori information for SISO Decoder 2. The inputs of SISO Decoder 2 consist of interleaved version of systematic bits (from side information) (L_{cyks}), punctured parity bits from Encoder 2 (L_{cykl2}), and a priory information $L(uk)$ that is derived from the other constituent decoder in the previous iteration. Here, $L(uk)$ is an additional information that helps the Turbo decoder to converge. SISO Decoder 2 then produces a posteriori information $L(uk \mid y)$. The extrinsic information yielded from this is then de-interleaved and becomes a priori information for the next iteration. The iterative turbo coding usually converges and saturates in 4 to 8 iterations. Finally, a hard decision decoding is performed to extract the binary output of the decoder.

4 Our approach

This section shows our approach, i.e., *Turbo Trellis Coded Modulation* (TTCM) (Sect. 4.1) adapted to DVC architectures and its practical applications which use it in both pixel and transform domain (Sect. 4.2).

4.1 TTCM codes for DVC

Turbo Trellis Coded Modulation [12] is a joint coding and modulation technique that has a similar structure to the well-known turbo codes, but involves *Trellis Coded Modulation* (TCM) [13] as the component codes. TCM utilizes a set partitioning based signal labeling mechanism in order to maximize the protection of the unprotected bits by maximizing the Euclidean distance of those bits in the signal constellation. It is reported that TTCM can achieve a given *bit error rate* (BER) on a noisy communication channel at a lower *Signal to noise ratio* (SNR) compared to Turbo coding [13] due to the higher coding gain.

4.1.1 TTCM encoding

In the conventional implementation of TTCM for communications channels, the symbols are generated by combining a number of data bits with additional parity bits introduced for protecting the data bits. For example, 1 data bit and 1 parity bit would be enclosed into one *Quadrature Phase Shift Keying* (QPSK) symbol, as illustrated in Fig. 6.

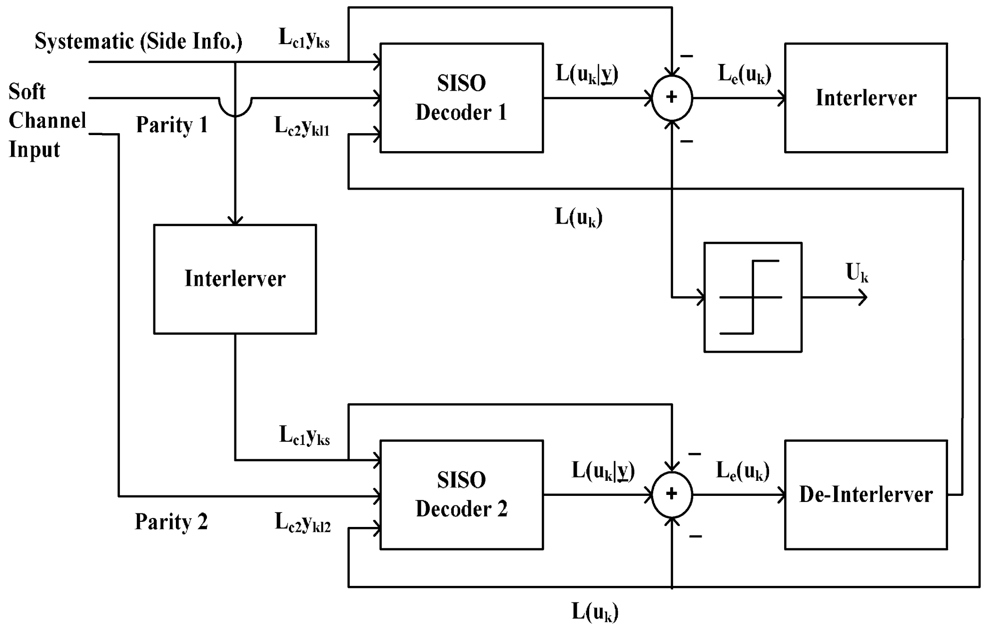
The set partition for this case is given in Fig. 7, where the data bit to be protected would be placed at the second level with a higher Euclidean distance (d_1) between two constellation points compared to first level (d_0). When higher order modulation is encountered (e.g., 16-QAM) with possibly not all bits being protected by parity, the unprotected bits would be placed at the highest Euclidean distance points on the signal constellation.

In the proposed solution for the DVC implementation based on TTCM, the process of generating the TTCM symbols is distributed between the encoder and decoder since in Wyner–Ziv coding only the parity bits are transmitted from the encoder to the decoder. The original data bits are discarded after generating the parity bit stream. At the decoder, the side information estimated by the motion compensation of the key frames is used to generate the symbols, in combination with the parity bits received from encoder. This process is illustrated in Fig. 8. These symbols are then passed into the TTCM decoder.

4.1.2 TTCM Decoding

The TTCM decoder incorporates a non-binary symbol based MAP algorithm [13]. The schematic of the TTCM decoder

Fig. 5 Turbo Decoder



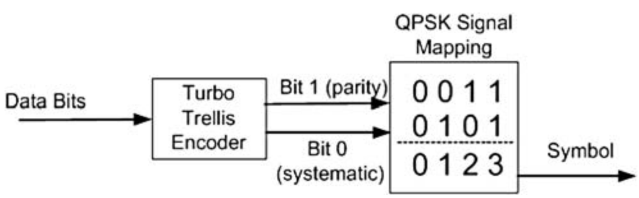


Fig. 6 TTCM symbols with QPSK modulation

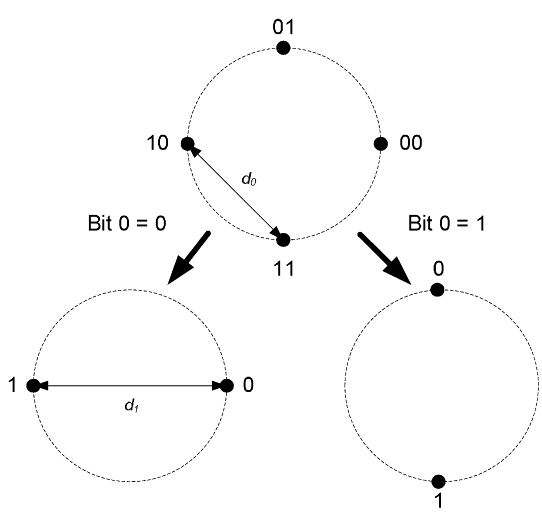


Fig. 7 QPSK set partitioning

is presented in Fig. 9. Since the parity bit stream is punctured at the encoder for shrinking the bit rate, the symbol mapping needs to be adapted by identifying the punctured bit positions. This purpose is served by the de-puncturer module placed before the symbol-by-symbol MAP algorithm. The hard decision decoding is performed after a number of soft iterations of the TTCM decoder as shown in the Fig. 9.

4.2 Proposed transform domain Wyner-Ziv codec

The majority of the architectures of DVC in transform Domain available in the literature have been carried out using Turbo Codes. Both DCT and DWT have been used, but always keeping Turbo Codes as the channel coding technique. On the other hand, Sect. 4.1 describes TTCM as an extension of turbo codes and which has been demonstrated in our previous works based on pixel domain [11] to be able to improve the PSNR at the same bitrate with less memory compared to the turbo coded pixel domain. Within this framework, in this work we continue to improve the performance of our pixel domain TTCM based architecture into a further paradigm, i.e., transform domain TTCM based DVC codec.

Figure 10 shows the block diagram of the proposed video codec implementation in transform domain. There are some differences between our solution and the previous solutions proposed in the literature (see Sect. 3), particularly in channel codes employed, the behavior of Slepian–Wolf codec, DCT and quantizer modules. In short, the video sequence is divided into WZ frames and key frames which are coded with H.264 AVC [14] Intra Coding. Over each WZ frame, X_{2i}, a 4×4 block-based DCT as defined by the H.264 AVC video standard [14] is applied. The DCT coefficients are then grouped together, according to their position, forming DCT coefficients bands. After the transform coding operation, each DCT band is uniformly quantized with 2^{M_k} levels, where the number of levels, 2^{M_k}, varies depending on each band. The turbo decoder–encoder structure is similar to the one used in pixel domain and has been depicted

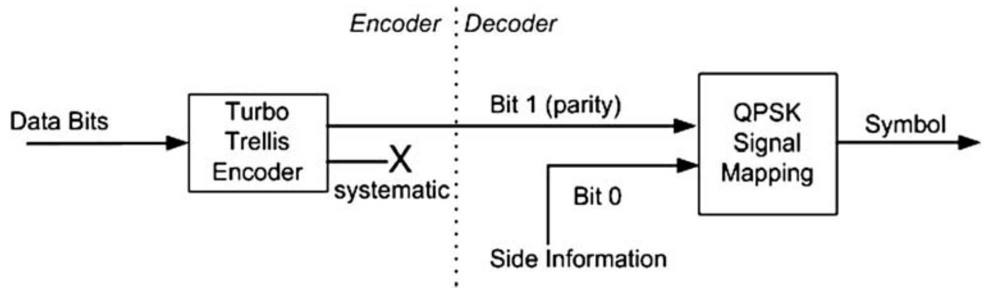


Fig. 8 Symbol mapping at the Wyner–Ziv decoder

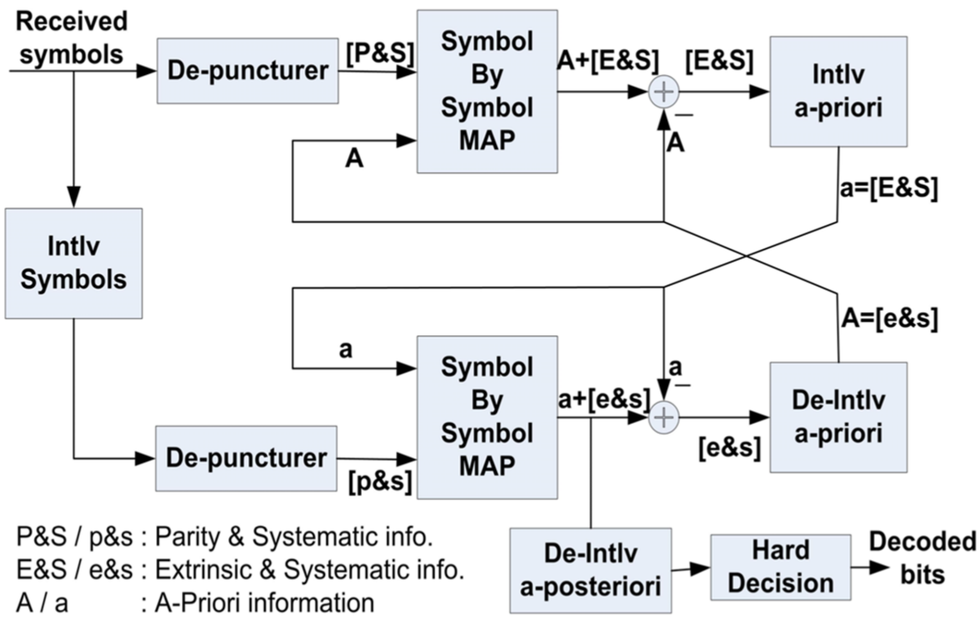


Fig. 9 Schematic diagram of the TTCM decoder

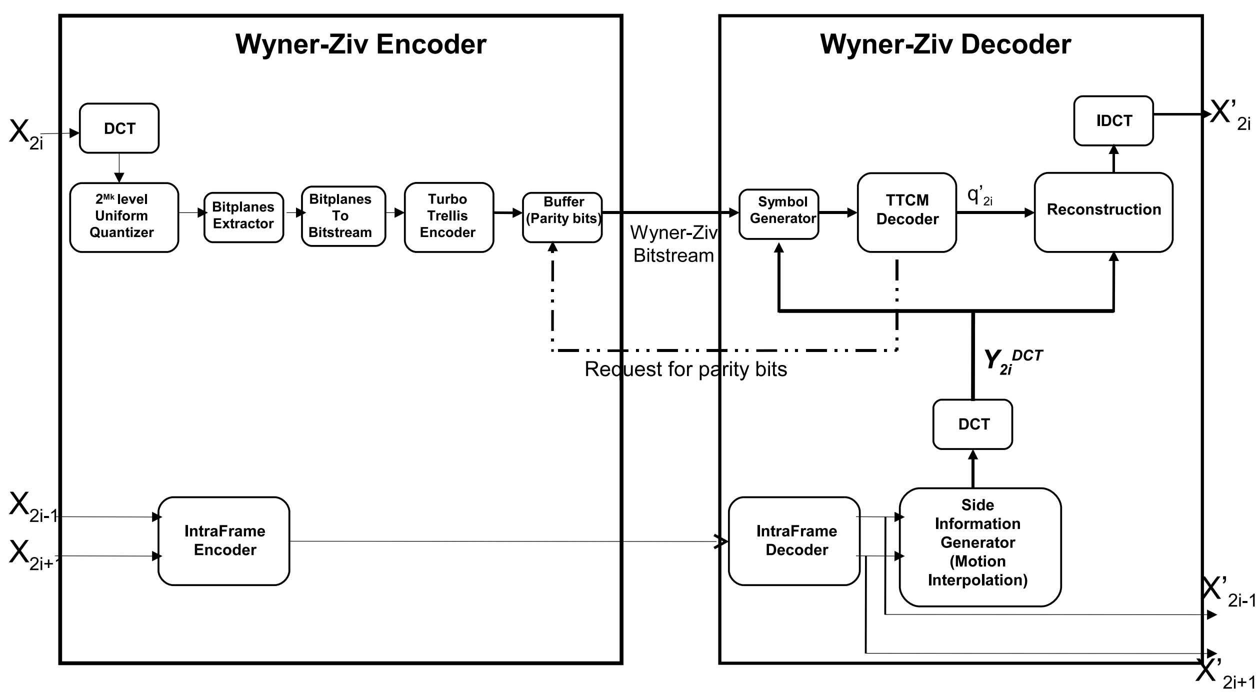


Fig. 10 Proposed architecture

in Sect. 3.1.2. The redundant information for each block is stored in a buffer and sent in small amounts upon decoder request. The decoder performs frame Motion Compensated Temporal Interpolation (MCTI) using previous and next adjacent frames in order to get an estimate of the WZ frame. The residual statistic between the WZ frame (X_{2i}) and its side information (Y_{2i}) is assumed to be modeled by a Laplacian distribution and the alpha parameter is estimated offline for entire sequence in a frame level. In future works we will estimate this hard information separately for each band dynamically using an on-line process. The decoded quantized symbol stream associated to each block can be obtained through an iterative turbo decoding procedure similar to the one explained in Sect. 4.1. The decoder has an ideal error detection capability in order to determine if the decoding block is considered successful in a similar way to the rest of architectures available in the literature. The reconstruction function generates the reconstructed symbol from the side information and q'_{2i} to reconstruct each DCT band of the frame. After all, a block-based IDCT is performed and the reconstructed frame X'_{2i} is obtained. In the next sections we are going to describe in depth each of the most important modules of our architecture.

4.2.1 *Discrete cosine transform*

The transform employed in our architecture relies on the 4×4 block-based DCT as defined by H.264 AVC [12]; the transform is applied to all 4×4 non-overlapping blocks of the X_{2i} frame, from left to right and from top to bottom. The major characteristic of this transform is that it is carried out using integer arithmetic (keeping the philosophy of low-complexity encoder), i.e., all the operations can be executed using only additions, subtractions, and shifts without accuracy loss. Moreover, all operations can be performed using 16-bit arithmetic, reducing the computational complexity. At the decoder, the DCT is also applied over the side information, so TTCM decoder uses it to obtain the reconstructed symbol; and it is necessary that both WZ frame and side information are in the same domain (frequency domain).

4.2.2 *Quantization*

Once DCT is applied, the coefficients are passed to the quantization process. The amplitude of the AC coefficients, within a 4×4 DCT coefficients block, tends to be higher for the coefficients close to DC coefficient and decrease as the coefficients approach the higher spatial frequencies. Since *Human Visual System* (HVS) is more sensitive to lower spatial frequencies, the DCT coefficients representing lower spatial frequencies are quantized using low quantization step sizes, i.e., with a higher number of quantization intervals (levels).

By letting the decoder know, for each X_{2i} frame, the dynamic range of each DCT coefficients band (i.e., in what range the DCT coefficients vary) instead of using a fixed value, it is possible to have a quantization interval width adjusted to the dynamic range of each band similar to [6]. Besides, we also need the number of quantization levels. Figure 11 indicates the number of quantization levels associated to each DCT coefficients band. These matrices are used to achieve the different RD performance point in the evaluation. In Fig. 11, the coefficients bands assigned value 0 means that there is no rate spent for this band and the corresponding band in the side information will be replaced in the reconstructed frame. The matrices are also used in the literature [5, 6]. In fact, five of them are extracted from [6] and the last one has been introduced by the authors in order to test our approach for higher bit rates. In our solution, we use the sign bit of the DC coefficient band to enhance the precision since all DC coefficients are always positive.

4.2.3 *TTCM encoder/decoder*

After quantizing the DCT coefficients bands, the quantized symbols (represented by integer values) are converted into a binary stream. The bits of same significance in the quantized symbols (e.g., the most significant bit) are grouped together forming the corresponding bitplane array. In [5, 6], these arrays are independently turbo encoded, i.e., the turbo encoder works on different bands independently. In our solution, the input-length of turbo encoder was studied and we observed better performance, working block by block.

Next, the Slepian–Wolf encoder based on TTCM codes incorporates the bit plane extractor and then the Turbo Trellis Encoder (depicted in Sect. 4.1). Each rate $1/2$ component encoder of our implementation has a constraint length $K = M + 1 = 4$ and a generator polynomial of [1102] in octal form. A pseudo-random interleaver is used in front of the second constituent encoder. Only the parity bit sequence thus generated is retained in the parity buffers and the systematic bits are discarded. The side information together with the parity bits passed from the encoder, upon request, form the PSK symbols to be processed in the TTCM decoder. A multilevel set partitioning is done with the constellation mapping of the TCM symbols in order to maintain the maximum Euclidean distance between the information bits. Where parity bits are not available due to puncturing being effective, the symbol automatically reduces to a lower modulation level. In the implementation under discussion, a combination of 4 PSK and Binary-PSK is used based on the availability of the parity bits for the constellation mapping.

5 Performance evaluation

In order to evaluate the rate–distortion performance of the proposed architecture in this paper, four test sequences were

Fig. 11 Quantization matrices associated to our transform domain DVC codec

16	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(a)

32	16	8	4
16	18	4	0
8	4	0	0
4	0	0	0

(b)

128	64	32	16
64	32	16	8
32	16	8	4
16	8	4	0

(c)

32	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(d)

64	32	16	8
32	16	8	4
16	8	4	4
8	4	4	0

(e)

256	128	64	32
128	64	32	16
64	32	16	8
32	16	8	0

(f)

Table 1 Main characteristics of the video test sequences

Video Sequence Name	Container	Foreman	Hall Monitor	Mother & Daughter
Total Number of Frames (evaluated)	300	300	300	300
Spatial resolution	QCIF	QCIF	QCIF	QCIF
Temporal Resolution	30 fps	30 fps	30 fps	30 fps

considered. Table 1 provides a brief description of the main characteristic of each test sequence.

We split up this section into three different scenarios which correspond to the next sections. The first one includes the lossless coding performance where the key frames are considered available in the decoder without compression. It is an impracticable solution, but in this framework we can focus on the performance just of the WZ frames and do not take into account the effect of the degradable quality of key frames. In the second scenario, we include the most reliable architecture introducing key frames coding (lossy coding scenario) and showing the performance of the whole architecture. Finally, in the last one, we show a comparative performance of our architecture with some of the state-of-the-art techniques proposed in the literature.

In this entire scenario, we have considered only the luminance data in the RD performance. The puncturing pattern is varied dynamically inside the codec, in order to generate an optimum decoded output with respected to the quantized input fed to the encoder. Further, it is noted that our implementation of the encoder uses a memory length of 3, whereas the codec proposed in [6] uses length 4. Shorter memory length means reduced computational complexity. The maximum allowable decoding iterations number is 6, in other solutions [6] it was considered 18 to allow the turbo decoder to converge. Less iteration means less complexity at the decoder. When we compare Pixel Domain architecture with the Transform Domain, the only differences between both are the inclusion of DCT and the corresponding changing of the quantization (more significant bitplanes in pixel domain instead of quantization matrices in transform domain) and, when we compare Turbo Codes with TTCM, the only change in the architecture is the encoder/decoder, the rest of modules and attributes inside them are the same.

5.1 Lossless coding

This scenario shows the performance of our architectures based on TTCM codes compared to the traditional ones based on turbo codes when the key frames are available at the decoder without any compression, which is well-known in the literature such as lossless coding scenario. Figure 12 shows the RD performance where only the WZ frames for rate and PSNR calculations have been considered. These results show the improvement of TTCM codes for both Pixel and Transform Domain. For a better comparative performance analysis of rate–distortion function, we also show the average PSNR difference (ΔPSNR) and the average bitrate difference (ΔBitrate) in the Table 2. The PSNR and bitrate differences are calculated according to the numerical averages between the RD-curves derived from the proposed transform domain DVC codec and the different codecs, respectively. The detailed procedures in calculating these differences can be found in the JVT document authored by Bjontegaard [15]. Note that PSNR and bit-rate differences should be regarded as equivalent, i.e., there is either the decrease in PSNR or the increase in bit-rate, but not both at the same time.

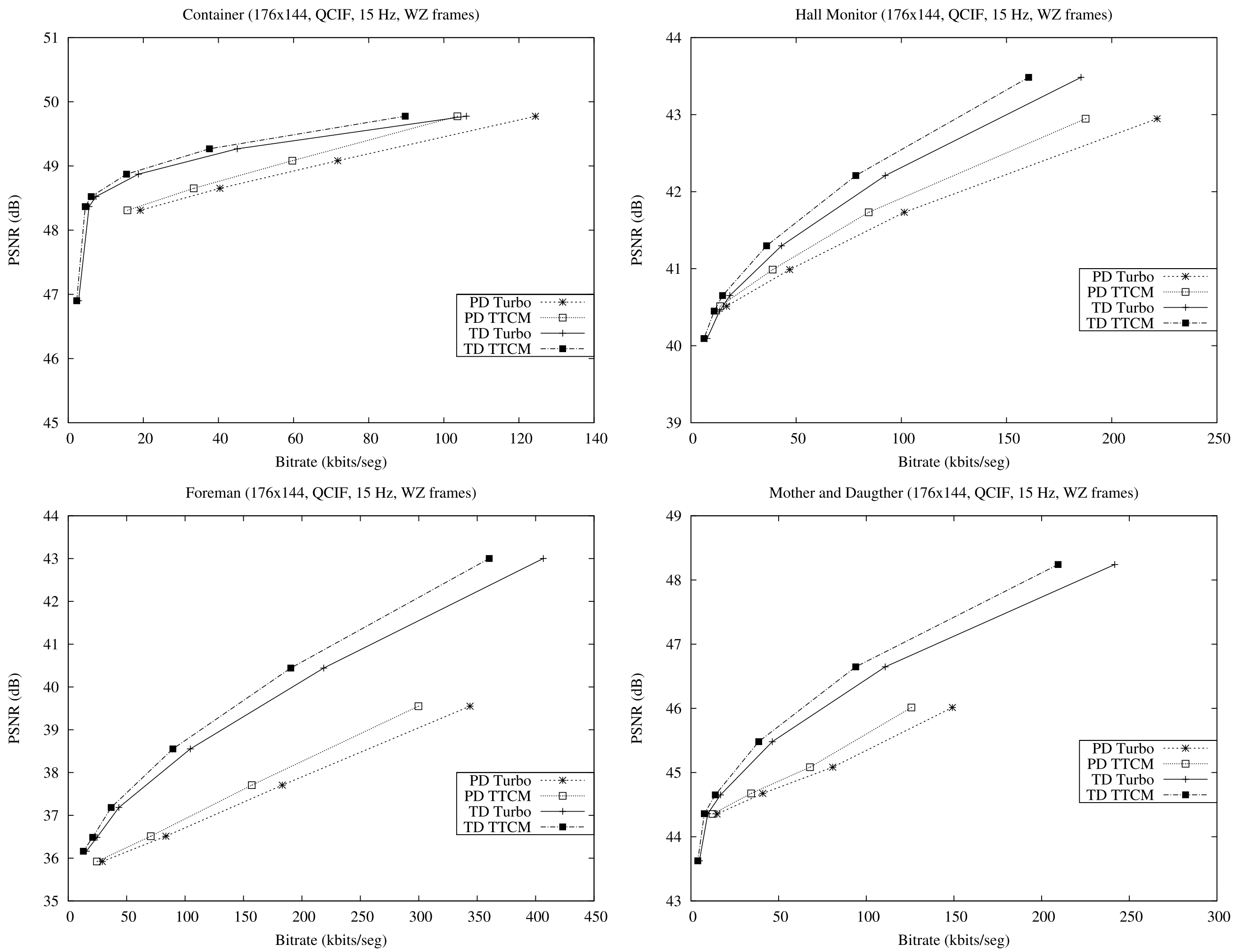


Fig. 12 RD performance for the lossless coding scenario

Table 2 Quantitative comparative performance (transform domain proposed vs. other architectures both for pixel and transform, and TC and TTCM. Lossless coding of test sequences)

	Container		Foreman		Hall Monitor		Mother & Daughter	
	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)
PDTC	−261.62	0.418	−143.46	1.0	−92.6	0.459	−45.02	0.590
PD TTCM	−28.07	0.262	−124.88	0.863	−33.21	0.321	−82.47	0.515
TD TC	−20.22	0.07	−15.36	0.163	−10.3	0.036	−17.75	0.136

5.2 Lossy coding

Lossy coding scenario is referred in the literature when the key frames of the sequence have been taken into account in the work of the WZ frames. This scenario is more realistic than the previous one, and these results can be useful to generalize a better performance of the proposed architecture for all possible situations. The key frames have been coded using H.264 AVC Intra, where the QP factor has been chosen following the criterion to keep the same quality of WZ frames as Key frames. Tables 3 and 4 show theses QP selections for both kinds of architectures: pixel and transform domains. Figure 13 shows the RD performance for the lossy key frames coding for the sequences under study and, the average PSNR and bit-rate differences are shown in Table 5 in a similar way to Table 2.

Table 3 QP selection for H.264 AVC [14] for key frames. Architecture in pixel domain

Video Sequence Name	Container	Foreman	Hall Monitor	Mother & Daughter
QP 4	12	24	21	17
QP 5	13	27	22	19
QP 6	13	28	24	19
QP 7	14	29	24	20

Table 4 QP selection for H.264 AVC [14] for key frames. Architecture in transform domain

Video Sequence Name	Container	Foreman	Hall Monitor	Mother & Daughter
(a)	12	20	20	14
(b)	12	23	22	17
(c)	13	26	24	18
(d)	13	27	24	19
(e)	14	28	24	19
(f)	15	29	25	20

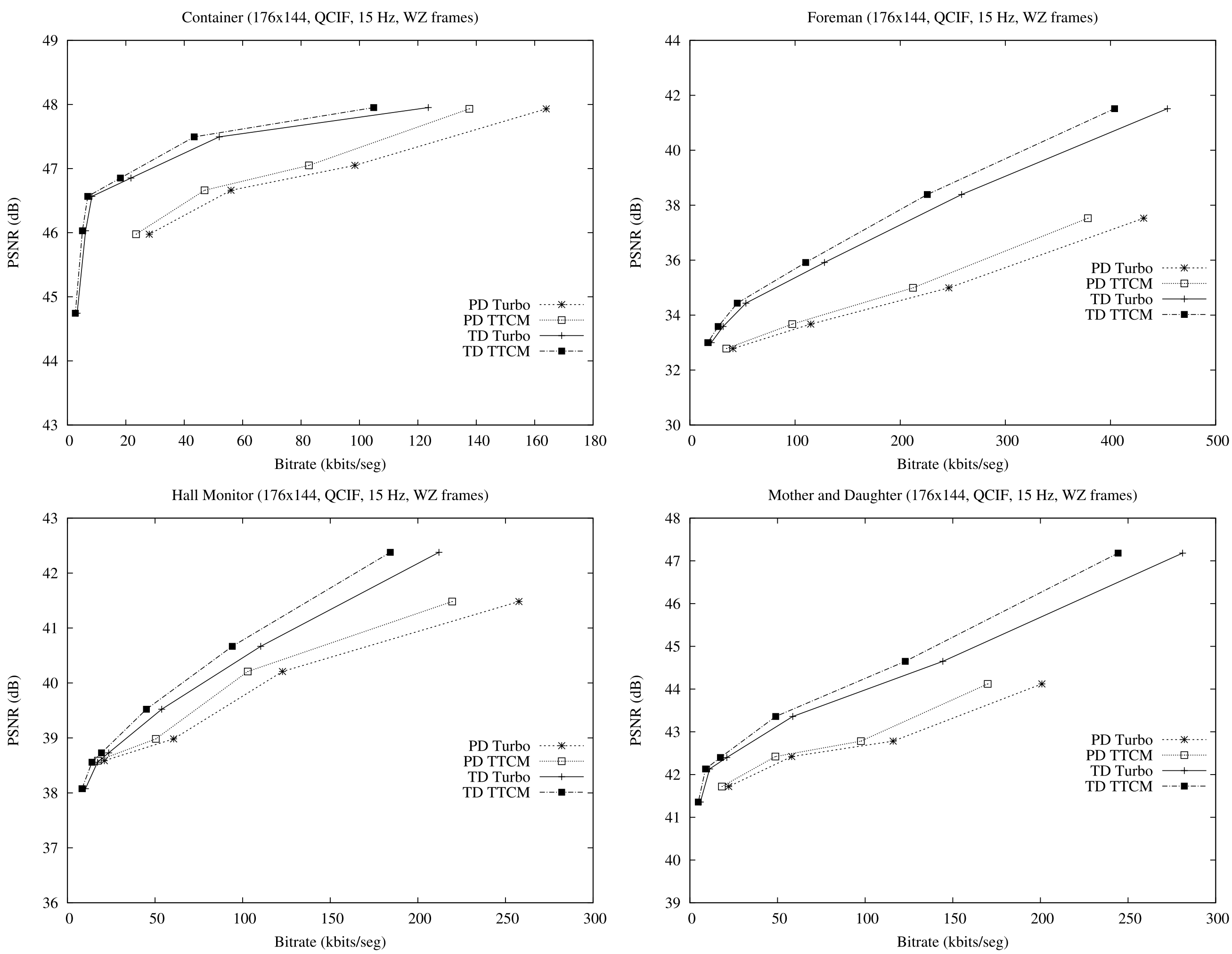


Fig. 13 RD performance for the lossy coding scenario

Table 5 Quantitative comparative performance (transform domain proposed vs. other architectures both for pixel and transform, and TC and TTCM. Lossy coding of test sequences)

	Container		Foreman		Hall Monitor		Mother & Daughter	
	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)
PDTC	–	0.772	−566.24	1.787	−113.45	0.805	−119.65	0.797
PD TTCM	–	0.710	−179.19	1.532	−52.52	0.6	−1393.0	0.723
TD TC	−82.71	0.154	−16.89	0.328	−25.51	0.348	−17.18	0.150

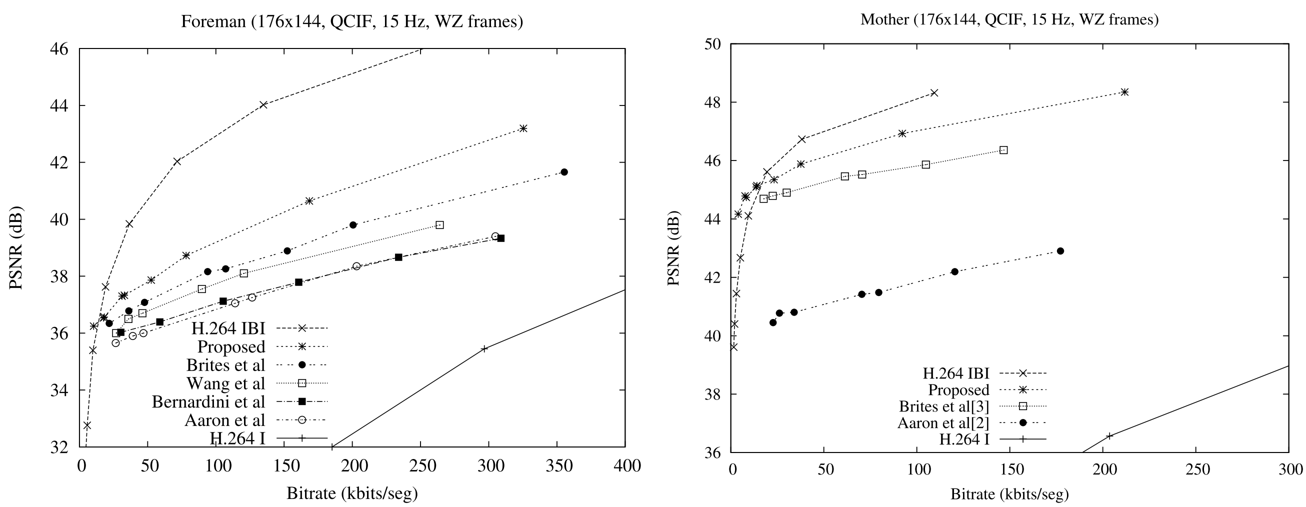


Fig. 14 Proposed vs. H.264, Aaron, Brites, Wang, and Bernardini

5.3 Comparison with other known results

In this section, a comparative performance evaluation in terms of the RD performance results of the proposed Transform Domain Distributed Video Codec is carried out with other transform domain architectures proposed in the literature, namely by Aaron, Brites, Wang, and Bernardini, described in Sect. 3.2. In order to compare our results with other architectures depicted in Sect. 3.2, we have employed the side information generation proposed in [6] and we were working in lossless key frames coding scenario. Figure 14 shows the results for the first 101 frames of the Foreman and Mother & Daughter sequences. We also included the H.264 results for comparison.

Table 6 shows quantitative results of a better comparative performance of rate–distortion function, in terms of the average PSNR difference (ΔPSNR) and the average bit-rate difference (ΔBitrate). As seen in Table 6, the usage of our transform domain DVC codec outperforms the other codecs with significantly lower bit-rate.

Table 6 Quantitative Comparative Performance (Transform Domain Proposed vs. H.264, Aaron, Brites, Wang, and Bernardini Architectures. Foreman and Mother & Daughter Sequences)

	Foreman		Mother & Daughter	
	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)
H.264 IBI	23.29	−1.572	41.02	−0.555
Brites et al.	−118.9	0.579	−117.81	0.522
Wang et al.	−48.08	0.77	–	–
Bernardini et al.	−133.67	1.052	–	–
Aaron et al.	−97.21	1.095	–	3.237
H.264 I	−217.24	3.891	–	–

6 Conclusion

This paper presented the utilization of TTCM in DVC in contrast to the more conventional Turbo coding based implementations, with its attractive high coding gain evident in communications. The discussion was concentrated on modifying the transform domain DVC framework to use TTCM.

The generation of TTCM symbols was distributed between the DVC encoder and decoder to suit the Wyner–Ziv coding framework. The simulation results show that the proposed TTCM based codecs can significantly outperform the state-of-the-art DVC codecs which are based on turbo codes.

References

1. Slepian, D., Wolf, J.K.: Noiseless coding of correlated information sources. IEEE Trans. Inf. Theory **19**, 471–480 (1973)
2. Wyner, D., Ziv, J.: The rate–distortion function for source coding with side information at the decoder. IEEE Trans. Inf. Theory **22**, 1–10 (1976)
3. Pradhan, S.S., Ramchandran, K.: Distributed source coding using syndromes (DISCUS): Design and construction. In: Proc. IEEE Data Compression Conference, Snowbird, UT, pp. 158–167, March 1999
4. Aaron, A., Zhang, R., Girod, B.: Wyner–Ziv coding for motion video. In: Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, November 2002
5. Aaron, A., Rane, S., Setton, E., Girod, B.: Transform-domain Wyner–Ziv codec for video. In: Proc. Visual Communications and Image Processing, San Jose, CA, January 2004
6. Brites, C., Ascenso, J., Pereira, F.: Improving transform domain Wyner–Ziv video coding performance. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, France (2006)
7. Aaron, A., Girod, B.: Compression with side information using turbo codes. In: Proc. IEEE Data Compression Conference, pp. 252–261, April 2002
8. Liveris, A.D., Xiong, Z., Georghiades, C.N.: Compression of binary sources with side information at the decoder using LDPC codes. IEEE Commun. Lett. **6**, 440–442 (2002)
9. Wang, A., Zhao, Y., Wei, L.: Wavelet-domain distributed video coding with motion-compensated refinement. In: Proc. of IEEE Int. Conference on Image Processing, Atlanta, USA (2006)
10. Bernardini, R., Rinaldo, R., Zontone, P., Alfonso, D., Vitali, A.: Wavelet domain distributed coding for video. In: Proc. of IEEE International Conference on Image Processing, Atlanta, USA (2006)
11. Weerakkody, W.A.R.J., Fernando, W.A.C., Adikari, A.B.B., Rajatheva, R.M.A.P.: Distributed video coding of Wyner–Ziv frames using Turbo Trellis coded modulation. In: Proceedings of International Conference on Image Processing, ICIP 2006, Atlanta, USA (2006)
12. Robertson, P., Wörz, T.: Bandwidth-efficient Turbo Trellis coded modulation using punctured component codes. IEEE J. Sel. Areas Commun. **16**, 206–218 (1998)
13. Ungerböck, G.: Channel coding with multilevel/phase signals. IEEE Trans. Inf. Theory **IT-28**, 55–67 (1982)
14. ISO/IEC International Standard 14496-10:2003: Information Technology—Coding of Audio. Visual Objects—Part 10: Advanced Video Coding
15. Bjontegaard, G.: Calculation of average PSNR differences between RD-curves. In: 13th VCEG-M33 Meeting, Austin, TX, April 2001

J.L. Martinez received his B.Sc. and M.S. degrees in Computer Science from the University of Castilla-La Mancha, Albacete, Spain in 2005 and 2007 respectively. In 2005, he joined the Department of Computer Engineering at the University of Castilla-La Mancha, where he is currently a researcher of Computer Architecture and Technology. He is completing his Ph.D. in the *Instituto de Investigación en Informática* (I3A) in Albacete, Spain. His research interests include Distributed Video Coding (DVC), video quality metrics, multimedia standards, video transcoding. He has also been a visiting researcher at the Florida Atlantic University, Boca Raton (USA). He has over 15 publications in these areas in international refereed journals and conference proceedings. He is a member of the IEEE.

W.A.R.J. Weerakkody received the B.Sc. (Hons) in Electronic and Telecommunication Engineering from the University of Moratuwa, Sri Lanka in 2000. Since then he has five years industry experience in Wireless Communications. Currently he is a research student completing his Ph.D. in the Centre for Communication Systems Research (CCSR) in the University of Surrey, UK. His research interests are in distributed video coding and wireless video communications. He has more than 30 publications in these areas in international refereed journals and conference proceedings. He is a member of the IEEE and IET.

W.A.C. Fernando received the B.Sc. Engineering degree (First class) in Electronic and Telecommunications Engineering from the University of Moratuwa, Sri Lanka in 1995 and the MEng degree (Distinction) in Telecommunications from Asian Institute of Technology (AIT), Bangkok, Thailand in 1997. He completed his Ph.D. at the Department of Electrical and Electronic Engineering, University of Bristol, UK in February 2001. Currently, he is a senior lecturer in signal processing at the University of Surrey, UK. Prior to that, he was a senior lecturer in Brunel University, UK and an assistant professor in AIT. His current research interests include Distribute Video Coding (DVC), 3D video coding, intelligent video encoding for wireless communications, OFDM and CDMA for wireless channels, channel coding and modulation schemes for wireless channels. He has published more than 165 international papers on these areas. He is a senior member of IEEE and a fellow of the HEA, UK. He is also a member of the EPSRC College.

G. Fernandez-Escribano received his M.Sc. degree in Computer Engineering, in 2003, and the Ph.D. degree from the University of Castilla-La Mancha, Spain, in 2007. In 2004, he joined the Department of Computer Engineering at the University of Castilla-La Mancha, where he is currently a researcher of Computer Architecture and Technology. His research interests include multimedia standards, video transcoding, video compression, video transmission, and machine learning mechanisms. He has also been a visiting researcher at the Florida Atlantic University, Boca Raton (USA), and at the Friedrich Alexander Universität, Erlangen-Nuremberg (Germany).

H. Kalva joined the Department of Computer Science and Engineering at Florida Atlantic University as an Assistant Professor in August 2003. Prior to that he was a consultant with Mitsubishi Electric Research Labs, Cambridge, MA. He was a co-founder and the Vice President of Engineering of Flavor Software, a New York company founded in 1999, that developed MPEG-4 based solutions for the media and entertainment industry. Dr. Kalva is an expert on digital audio-visual communications systems with over 12 years of experience in multimedia research, development, and standardization. He has made key contributions to the MPEG-4 Systems standard and also contributed to the DAVIC standards development. His research interests include pervasive media delivery, content adaptation, video transcoding, video compression, and communication. He has over 50 published papers and five patents (11 pending) to his credit. He is the author of one book and co-author of five book-chapters. Dr. Kalva received a Ph.D. and an M.Phil. in Electrical Engineering from Columbia University in 2000 and 1999, respectively. He received an M.S. in Computer Engineering from Florida Atlantic University in 1994, and a B. Tech. in Electronics and Communications Engineering from N.B.K.R. Institute of Science and Technology, S.V. University, Tirupati, India in 1991.

A. Garrido received the degree in physics (electronics and computer science) and the Ph.D. degrees from the University of Granada, Spain, in 1986 and University of Valencia, Spain, in 1991, respectively. In 1986, he joined the Department of Computer Engineering at the University of Castilla-La Mancha, where he is currently a Full Professor of Computer Architecture and Technology and Dean of the Escuela Politecnica Superior de Albacete (School of Computer Engineering). His research interests include high performance networks, telemedicine, video compression, and video transmission. He has published over 60 papers in international journals conferences on performance evaluation of parallel computer and communications systems and compression and transmission in high-speed networks. He has led several research projects in telemedicine, computer networks and advanced computer system architectures.