Low Complexity Video Encoder using Stanford WZ video coding architecture with Iterative SOVA decoder.

Naveen Kandwal, Prof. Indr Jeet Rajput
Hasmukh Goswami College of Engineering

***Abstract*- We address the problem of encoding the video from conventional video encoders with the emerging technologies limitations of limited resources and power supplies .In this paper, we worked on Distributed Video Coding (DVC) which is an emerging coding scheme that employs principles of source coding with side information (SI) at the decoder. The DVC encoding system has low complexity because of the elimination of motion estimation which uses most of the resources for prediction and necessary implementation functions. In this paper, we present a comparison between the DVC encoder and conventional video encoder in terms of complexity and time consumption. And have found that with the help of iterative decoder the output result can be improved.**

***Index Terms*- DVC, MPEG-2, H.264, DCT, ME, MV, WZ, SOVA,**

I. INTRODUCTION

Modern video coding systems, such as MPEG-2 and H.264/MPEG-4 AVC, are based on a hybrid compression scheme consisting of spatial-temporal prediction and blockwise Discrete Cosine Transform (DCT). The most computationally expensive operation involved in the encoding process is motion estimation (ME), which produces a prediction of the current video frame, in terms of motion vectors (MVs) and previously decoded frames. ME, followed by motion compensation, is one of the most effective methods for exploiting temporal redundancy, but usually has high complexity. In these video coding systems, the encoder is typically one to two orders of magnitude more complex than the decoder, especially due to the ME process that is performed at the encoder. This suits well downlink oriented applications such as video broadcasting, in which a low complexity decoder is important since the video is encoded once and then decoded by many users. However, today we see a shift towards producing and sharing videos, especially for real time application that rely on an upstream model. Examples are video conferencing over wireless/cellular networks, video surveillance and many more. The clients, often mobile, that capture and encode the video, have low-power and limited resources, in contrast to a central server, which is usually powerful. A novel video coding

As shown in Figure 1, Source X and Source Y are two statically correlated sources. When source X is encoded conventionally, paradigm, known as Distributed Video Coding (DVC), has emerged in the last decade. This paradigm employs principles of There exist several DVC solutions, based on these theorems, which try to exploit the video data correlation mostly at the decoder. Examples are PRISM [3], Stanford [4] and DISCOVER [5] systems. Also with the modified decoder the encoded data is recovered with less number of changes. The decoder we are working on is SOVA decoder with iteration which increases the quality of picture produced at the decoder side. Although with this new era of iteration decoder the bitplane of the image need not to perform same number of iterations in all the bit planes. As it was observed in the research that with the increasing biplane depth from 1 to 8 in gray scale, the changes in the bit planes are decreasing plus within the decoding system the increment in iteration for the depth bit planes recovers the better quality of picture rather than the upper planes.

II. DVC THEORY AND FRAMEWORK

DVC is consequence of Distributed Source Coding [1], and base on the lossy result obtained by Wyner and Ziv [2]. There are two primary schemes of DVC, since the scheme called Prism given by Rohit Puri and Kannan Ramchandran [3], focus on the robustness while transmitting and it is relatively complex, this paper takes the DVC scheme raised by Bernd Girod of Stanford University[6] into consideration. A software realization and optimization of this scheme called DISCOVER [7].

2.1 DVC THEORY

Distributed Source Coding is to encode two or more correlated sources separately, then decode them jointly [6].

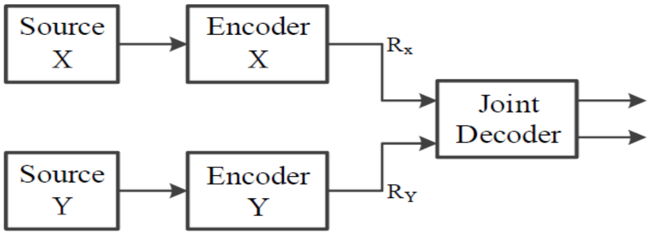

Figure 1: Distributed Source Coding scheme

It can be decoded lossless while $Rx >= H(X)$, $H(X)$ means the entropy of Source X. When X and Y are decoded jointly, they can be decoded lossless when:

$RX + RY >= H(X, Y)$

$RX >= H (X|Y)$

lossy source coding with side information at the decoder, also known as Wyner-Ziv (WZ) coding. These principles rely on the seminal information theory theorems by Slepian and Wolf [1] (for the lossless case) and by Wyner and Ziv [2] (for the lossy case).

At the decoder, one frame is decoded independently, which is called Key Frame, and the other one is decoded dependently on the reconstruction of key frame.

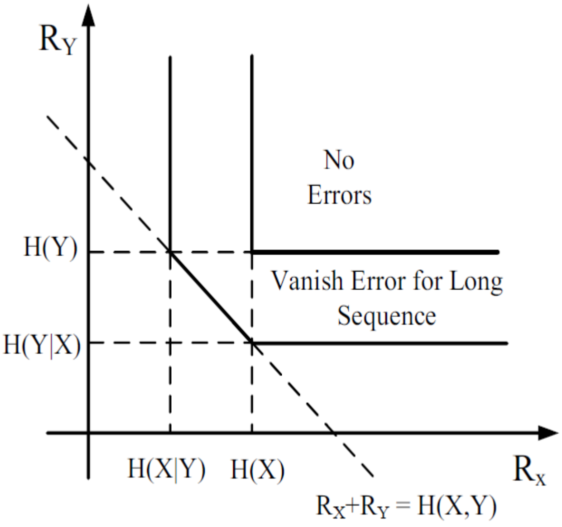

Figure 2: Available rate region for correlated source encoding independently in DSC [6].

2.2 DVC FRAMEWORK

A low complexity DVC paradigm with feed-back channel is raised by Bernd Girod in 2005 [6]. After that, some European researchers come out a complete DVC framework base on the Stanford paradigm, which is called DISCOVER [5]. It solves a lot of problems in realization of Stanford paradigm, for example, the dependence of original video source at the decoder. The structure of Discover is shown in Figure 3.

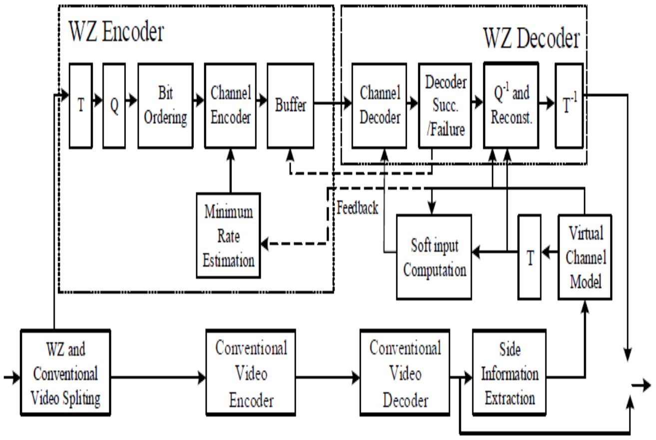

Figure 3: Block Paradigm of DISCOVER Structure [5]

RY >= H (Y|X)

H (X|Y) is the conditional entropy of X when Y is known. The region of rate is shown in Figure 2. Since H(X, Y) =< H(X) + H(Y),

the rate it needs to decode X and Y jointly and lossless is low than that to decode them separately when X and Y are correlated. Therefore it achieves information compression.

domain. In the typical pixel domain, the first frame in each groups of picture (GOP) is called key frame with traditional intra frame video encoding, the remaining frames in the same GOP are called WZ frame which applies with distributed encoding principle. In the encoder, WZ frame divided into three functions: quantization, channel coding, and buffer. In quantization function, uniform quantization and split into bit planes. In channel coding function, turbo code and Low Density Parity Check Accumulate (LDPCA) could be used. Here, channel coding was deployed for rate adaptive error correction of frames estimated at decoder, by sending parity bits. In the decoder, the neighboring of decoded key frames could be used as reference frame, which is as the side information (SI). SI generation is between the neighboring of decoded key frames with interpolation for generated reference WZ frames, and then requests the feedback channel doing error correction to reconstruct correct WZ frames. Next, the final decoded WZ frames are generated through inverse quantization (IQ). Finally, decoded key frames and decoded WZ frames were recombined together to complete whole video decoding. Pixel domain provides a simple encoder to reduce encoder complexity, and however performance is below expectations. Thus, additional simple encoding function to increase performance, transform domain, is one of the methods. Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) are good for block coding and frame coding, and other transform algorithms could be used. Transform function is added before quantization in the encoder, and the transformed coefficients are then quantized before splitting into bit planes. Decoder is also subject to inverse transform (IT) after IQ, depicted in Fig. 4

.

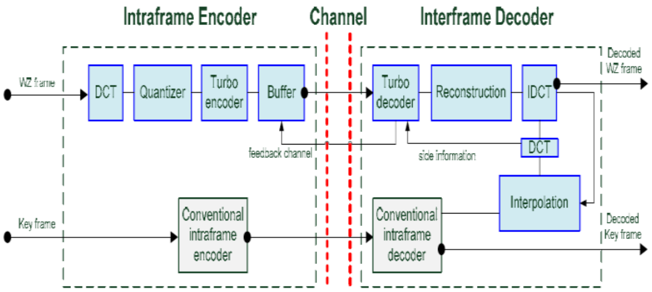

Figure 4: Stanford WZ video coding architecture

III. RELATED WORK

Although the mathematical background of DVC was proposed in

The DVC framework is made up of 2 codecs, one is call conventional video codec and the other is called WZ codec. Conventional video codec is used to encode and decode key frames, usually is H.264-Intra codec, JPEG codec and so on. WZ codec is the video codec that uses DVC theory to encoder video sequence. Since it is based on the rate-distortion result given by Wyner and Ziv, the encoder is called WZ encoder that is short for Wyner-Ziv encoder. The WZ video codec [6] was first proposed by Stanford in 2002, and four domains were suggested for research direction. Pixel domain, joint decoding and motion estimation, rate control, and then extended to transform

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 and how one can get this performance in practical architectures 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

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 [11]. 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 2M-level uniform scalar quantizer; in this case, $2M \in \{2, 4, 8, 16\}$. 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, X2−1 and X2+1. 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

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 [8], turbo codes (TC) [9], or Low Density Parity Check (LDPC) codes [10] have been able to achieved the rate point depicted in the Slepian–Wolf theorem. The compression of X is achieved by transmitting only a binary index [8], or parity bits [9,10]. 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

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 [12]. Brites et al. extended the Aaron et al. work [13] 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. [14] 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 [12,13] based on scalar quantization. At the decoder, side information is gradually updated by motion-compensated refinement. Bernardini et al. [15] 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)

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 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 [12] 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

encoders separated by an interleaver. 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.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 in communications. This concept has been successfully adopted for DVC. The structure of the turbo decoder with Soft channel outputs containing received parity bits (Lcykl1) from the first encoder and the systematic bits (side information—Lcyks) 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) (Lcyks), punctured parity bits from Encoder 2 (Lcykl2), 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 | 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.

IV. OUR APPROACH

This section shows our approach, i.e., Turbo Coded Modulation (TCM) (Sect. 4.1) adapted to DVC architectures and its practical applications which use it in pixel domain. 4.1 TCM codes for DVC Turbo Trellis Coded Modulation [12] is a joint coding and modulation technique that has a similar structure to the DISCOVER architecture. 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 TCM 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 [16] due to the higher coding gain.

4.1.1 TCM encoding

In the conventional implementation of TCM for communications

4.1.4 CHANNEL ENCODER

Finally, bit streams are input into channel encoder, which is verified to be closer to Shannon limit of channel than Turbo Codes. Turbo encoder consists of recursive systematic convolutional encoder. The systematic bits are transformed into the encoded sequence the parity bits are stored in the buffer and are sent to the decoder over BPSK modulation.

4.1.5 TCM Decoding

The TCM decoder incorporates a non-binary symbol based SOVA algorithm [16]. 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 SOVA algorithm. The hard decision decoding is performed after a number of soft iterations of the TCM decoder. 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 (X2i) and its side information (Y2i) is assumed to be modeled by a Laplacian distribution and the alpha parameter is estimated offline for entire sequence in a frame level. 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 channels, the symbols are generated by combining a number of data bits with additional parity bits protecting the data bits. For example, 1 data bit and 2 parity bit would be enclosed. 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.

4.1.2 FRAME SPLITTER

Firstly, all of the frames to be encoded are separated into groups depending on the parameter of group of pictures (GOP) size. The first frame of every group is encoded by conventional video encoder and other frames, which we call them WZ frames, will be encoded though WZ encoder.

4.1.3 BIT ORDERING

After Frame splitting, all bits of the WZ Frame are extracted in bit planes and are reshaped into a stream of bits. Then the encoder will scan every bit of the numbers in the same band and based on the rate of the encoder i.e., 1/3 each bit generate three bits one systematic bit and two parity bits. The priority bits are stored in the buffer and are used for sending it to the decoder via channel.

V. SIMULATION RESULTS

This paper uses two frames to get the iterative result, mentioned in figure 5. The frames with 100X100 pixel @15fps, with a GOP size of 2, which mean every group includes one key frame and one WZ frame.The encoding time is obtained on a CPU@2.50GHz, Intel(R) Core(TM) i5-3210M.

5.1 ENCODING COMPLEXITY ON SOFTWARE

In Table 1, it is easily observed that WZ encoder is about 10 times faster than H.264 intra encoder. And the encoding of H.264 is much more complex than WZ encoder and H.264 intra encoder. H.264 No motion encoder for H.264 inter encoder since it doesn’t search the motion vector and only encoder the residual between P-frame and I-frame. Therefore it is much faster than H.264 inter encoder, and is similar to H.264 intra encoder. WZ encoder presents a large advantage in encoding time and it encodes video sequence in real time. Compared to conventional video encoder, WZ encoder is more suitable for low processing ability video sensors.

Table 1: Encoding time for Different Codes [18]

Video Codec	Encoding Time (ms)
H.264 Intra	103
H.264 (No Motion)	430

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.

4.1.6 ITERATIVE SOVA DECODER

SOVA is a modified Viterbi algorithm with additional output values associated with the original decoded bit sequence. It was formerly used in serial concatenated coding scheme. At the decoding end, firstly the key frames are decoded by conventional decoder. Side-Information computation block generates Side Information (SI) though Bi-directional Motion Estimation and Motion Compensated Interpolation (MCI) [17]. Then distribution of the residual between SI and original WZ frame is estimated in Virtual Channel Model block [13].The SI is transformed and quantized to prepare for the reconstruction. Then the SI is converted into soft-input information to the channel decoder by the estimated parameter of virtual channel. Channel decoder corrects the errors in the SI using a part of parity though an iterative algorithm with the conditional probability P (WZ| SI) which is obtained from previously decoded bits and SI. If channel decoding successes, decoded stream is put into reconstruction block.

Table 2: Result of Changes within the WZ Frame, Key Frame and Output Generated Frame.

		KEY FRAME	OUTPUT (ITERATIONS)			
			1	2	3	4
		CHANGES				
BIT PLANE	PLANE1	3140	3145	3173	3149	3156
	PLANE2	3002	2990	2995	2995	2995
	PLANE3	2210	2186	2160	2121	2109
	PLANE4	1492	1479	1461	1442	1423
	PLANE5	1000	956	936	933	952
	PLANE6	783	711	710	709	709
	PLANE7	307	236	239	239	239
	PLANE8	196	111	105	88	82

VI. CONCLUSION AND FUTURE WORK

This paper introduces the original architecture of DVC, an outstanding implementation of DVC on software. Finally, this paper gives comparison of DVC architecture and Conventional video codec complexity on software. The results of experiment show that DVC has competitive rate-distortion performance to conventional video codec, and has amazing reduction in complexity on the encoding time. The bit plane at the depth on gray scale level image show better improvement in increasing the

WZ Encoder	8.36

5.3 ITERATION RESULT AT DECODER

Each bit plane is encoded and decode separately. Comparing the bitplane iteration results in table 2.The depth bit plane is having lower number of changes compare to the upper bit plane along with that it was seen that the depth bitplanes shows the betterment in results while increasing the iterations on the other side the upper bit plane does not any change in the betterment with the increasing of iterations. Hence multiple iterations has to be done at the lower bit plane while for the upper bit plane less number of iteration will do the work.

(A) (B) (C)

Figure 5: DVC encoder decoder results with (a) Frame 1 the WZ Frame (b) Frame 2 Key Frame used for side Information generation (c) Frame 1 Regenerated with the help of SOVA iterative decoder.

REFERENCES

[1] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. IEEE Transactions on Information Theory, 19(4):471–480, July 1973.

[2] A. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. IEEE Transactions on Information Theory, 22(1):1–10, Jan. 1976.

[3] R. Puri, A. Majumdar, and K. Ramchandran. PRISM: A video coding paradigm with motion estimation at the decoder. IEEE Transactions on Image Processing, 16(10):2436 –2448, Oct. 2007.

[4] A. Aaron et al. Transform-domain Wyner-Ziv codec for video. In SPIE Conference Series, volume 5308, pages 520–528, 2004.

[5] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret. The DISCOVER codec: Architecture, Techniques and Evaluation. In Picture Coding Symposium (PCS’07), 2007.

[6] Girod, B., Aaron, A. M., Rane, S., & Rebollo-Monedero, D. (2005). Distributed video coding. Proceedings of the IEEE, 93(1), 71-83.

[7] Artigas, X., Ascenso, J., Dalai, M., Klomp, S.,

iteration rather than at the upper level bit plane. Hence the iteration should be controlled for every bit plane individually so that the better quality of picture can be obtained.

Moreover, since the decoder is very complex and not real-time, parallel and speeding up methods may also be taken into consideration.

Kubasov, D., & Ouaret, M. (2007, November). The DISCOVER codec: architecture, techniques and evaluation. In Picture Coding Symposium (Vol. 17, No. 9, pp. 1103-1120). Lisbon, Portugal.

[8] 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

[9] Aaron, A., Girod, B.: Compression with side information using turbo codes. In: Proc. IEEE Data Compression Conference, pp. 252–261, April 2002

[10] 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)

[11] 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

[12] 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

[13] 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)

[14] 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)

[15] 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)

[16] Ungerböck, G.: Channel coding with multilevel/phase signals.IEEE Trans. Inf. Theory IT-28, 55–67 (1982)

[17] Ascenso, J., Brites, C., & Pereira, F. (2005, July). Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding. In 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services (pp. 1-6). Smolenice, Slovak Republic.

[18] Xiao Yang, Fei Qiao, Huazhong Yang: Hardware Realization of Low Complexity Video Encoder within Distributed Video Coding Framework. In: ISBN: 978-0-9891305-6-1 (2014)