

# Multi-Hypothesis based Distributed Video Coding using LDPC Codes

Kiran Misra, Shirish Karande, Hayder Radha

Department of Electrical and Computer Engineering

2120, Engineering Building

Michigan State University

East Lansing, MI 48824 USA

{misrakir, karandes, radha}@msu.edu

## Abstract

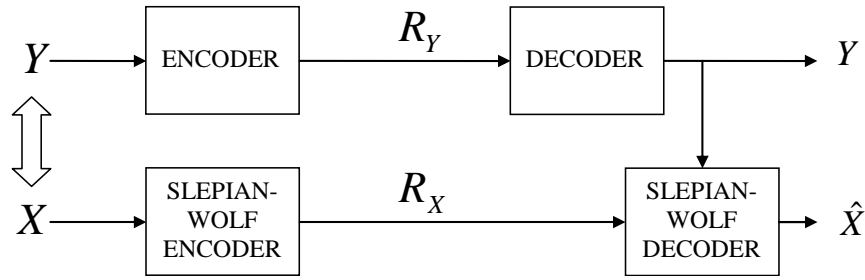
Conventional video coding paradigms are usually built on the assumption that a video stream will be encoded only once but decoded several times. Over the years, this has led to the development of encoders, which use complex motion-estimation algorithms to reduce the size of the video bitstream. However in applications like distributed sensor networks, cell-phones etc. the encoder has a limited amount of resources available to it and cannot employ complex motion-estimation algorithms. Intra-coded frames used in conventional video coding have low-complexity, however they need large bandwidth to be transmitted. This has led to the development of Wyner-Ziv video codecs [1] which are based on the principles of side-information based coding, first propounded in the 1970s by Slepian-Wolf and Wyner-Ziv in their seminal work [2] and [3] respectively. Wyner-Ziv video codecs perform intra-frame encoding at the transmitter, and inter-frame decoding at the receiver. This reduces the size of the intra-frame considerably, at the cost of increasing complexity at the decoder. For video transmission using cell-phones this increased decoder complexity can be absorbed by the base stations.

Current distributed video coding schemes achieve compression by using a single reference frame for side-information. This single reference frame could be the temporally adjacent picture frame, or on an interpolated frame constructed using temporally adjacent frames. This paper proposes a new scheme for decoding video using more than one video frame as reference. Unlike schemes which use an interpolated frame, our decoding scheme makes use of data directly from the reference frames, giving an improvement of up to 1 dB. The approach is based on the premise that, covered and uncovered region are better represented in the original side-information frames, compared to the interpolated frame. This is especially true for sequences which do not have low-motion content.

**Index Terms** – Channel Codes, distributed source coding, multi-hypothesis.

## 1. Introduction

In state of the art video coding schemes like MPEG-4 [4] or the latest H.264 [5] standard, the encoder uses highly complex motion estimation algorithms to achieve efficient video compression. The decoding process in comparison is low-complexity. However, the low power consumption requirement for cell-phones, distributed sensor networks etc. put severe restrictions on the complexity of the encoder. But these applications can afford high computation complexity at the decoder. This has led to the development of the inverted



**Figure 1: Slepian-Wolf Side-Information Based Decoding**

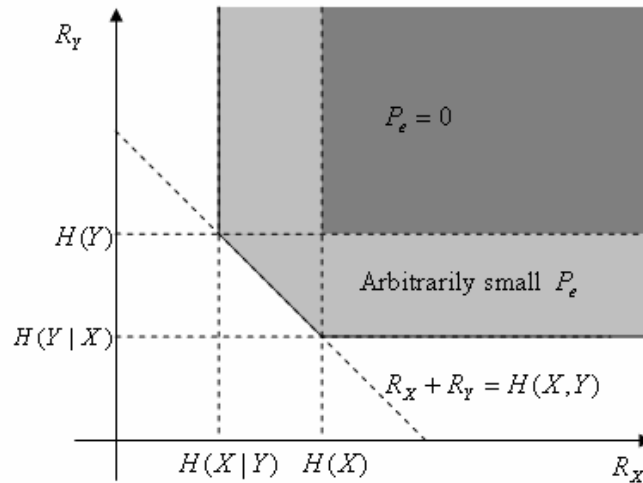
paradigm for Wyner-Ziv video coding where encoders are low-complexity and the decoders are highly complex. Pradhan et. al. first proposed the DISCUS [6] architecture which used syndrome-based encoding to perform Wyner-Ziv coding of video. García proposed a different model for binary sources in [7], where one of the correlated sources is treated as a corrupted version of the other, turbo codes were then used to correct the error. We use the later approach in our paper. Another important Distributed Video Coding framework is the PRISM [8] architecture proposed by Puri et. al.

The Slepian-Wolf theorem can be briefly described as follows: Assume there are two correlated sources  $X$  and  $Y$  as shown in Figure 1. Let  $Y$  be transmitted after being entropy-encoded at a rate  $R_Y \geq H(Y)$ ,  $H(Y)$  is the entropy of  $Y$ . If the Slepian-Wolf decoder is aware of the statistical relationship between the two correlated sources, then  $X$  can be Slepian-Wolf encoded and transmitted at a rate  $R_X \geq H(X|Y)$ , where  $H(X|Y)$  is the conditional entropy of  $X$  given  $Y$ . The decoder can then use the statistical information available to it and the side-information  $Y$  to reconstruct  $\hat{X}$  with an arbitrarily small probability of error  $P_e$ . The Slepian-Wolf theorem defines a region of achievable rates based on:

$$R_X + R_Y \geq H(X, Y) \quad (\text{I})$$

$$R_X \geq H(X|Y) \quad (\text{II})$$

$$R_Y \geq H(Y|X) \quad (\text{III})$$



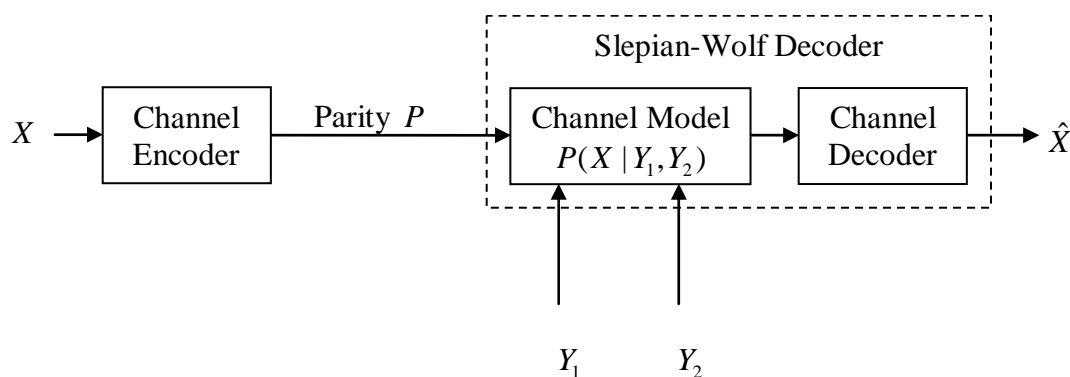
**Figure 2: Slepian-Wolf Achievable Rate Region [2]**

The rate-region defined by these constraints is shown in Figure 2. In video applications, a small distortion is acceptable and hence  $X$  can be transmitted at a rate  $R_Y$  which is greater than  $H(X|Y)$ . In video coding, the correlated sources are usually temporally adjacent frames. In the second approach discussed above, the Slepian-Wolf encoder channel codes  $X$  and transmits the parity information  $P$ . At the decoder,  $Y$  is treated as a noisy version of  $X$ . A channel decoder then uses  $P$  to correct the errors in  $Y$  to get  $\hat{X}$ . As the length of  $X$  increases the probability of  $\hat{X} \neq X$  diminishes.

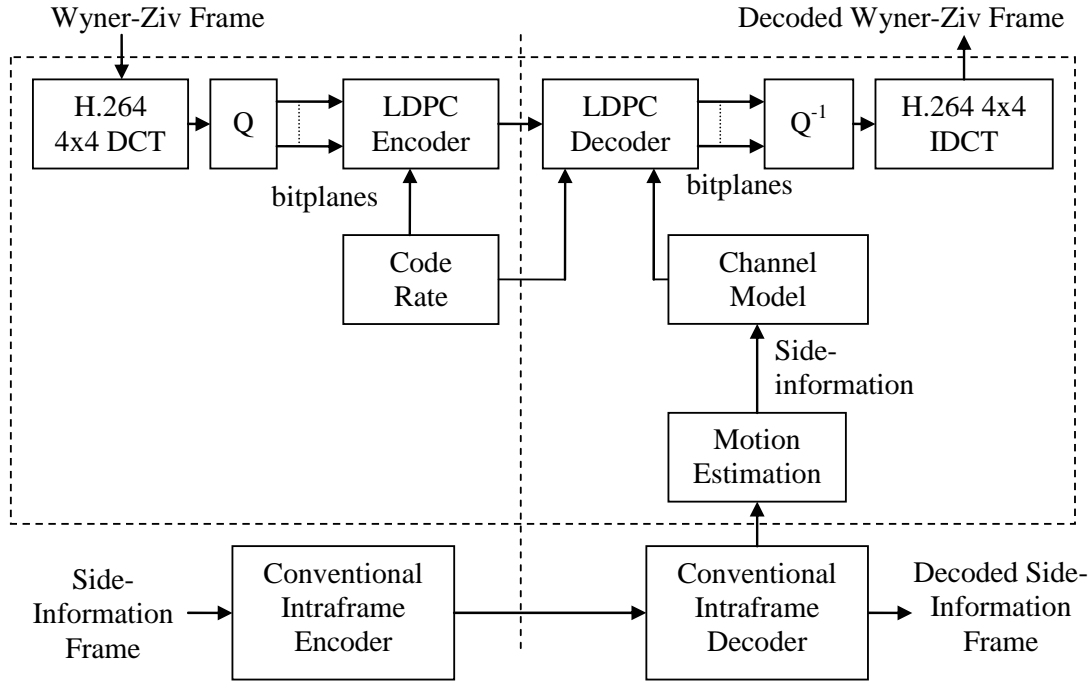
If more side-information frames are used for decoding the bit-rate can be further reduced. In this paper we propose a scheme which uses two side-information frames for decoding as opposed to a single interpolated side-information frame. In the following sections we discuss in more details the encoding-decoding algorithm used. Section 2 discusses the motivation for this approach; Section 3 outlines the details of the distributed video codec. In Section 4 we describe the experimental setup and discuss results. Section 5 summarizes the key conclusions of this work.

## 2. Motivation

If more than one side-information is available at the decoder, then the rate at which  $X$  needs to be transmitted can be further reduced. Let  $Y_1$  and  $Y_2$  be temporally adjacent two side-information frames then, rate at which  $X$  needs to be transmitted is  $R_x \geq H(X|Y_1, Y_2)$ . The interpolating schemes discussed in [9] averages the two side-information frames to obtain  $\hat{Y} = (Y_1 + Y_2)/2$ , and therefore restricts the rate to  $R_x \geq H(X|\hat{Y})$ . Random variables  $X$ ,  $(Y_1, Y_2)$  and  $\hat{Y}$ , form a Markov chain  $X \rightarrow (Y_1, Y_2) \rightarrow \hat{Y}$ , since  $X$  and  $\hat{Y}$  are conditionally independent given  $(Y_1, Y_2)$ . It follows from Shannon's data-processing theorem [10] that  $I(X; Y_1, Y_2) \geq I(X; \hat{Y})$ . As a matter of fact using the data processing theorem it can be shown that  $I(X; Y_1, Y_2) \geq I(X; g(Y_1, Y_2))$ , where  $g(Y_1, Y_2)$  could be a more elaborate weighting mechanism. This in turn implies that,  $H(X|Y_1, Y_2) \leq H(X|\hat{Y})$ , thus leading to a reduction in the number of bits required to transmit  $X$ . In this paper we show that the above theoretical implications can be realized in a Wyner-Ziv video codec. Specifically, we show that the reduction in bit-rate translates into a performance gain of up to 1 dB. These savings are especially pronounced for videos which have high motion content.



**Figure 3: Slepian-Wolf decoding using two side-information frames.**



**Figure 4: LDPC based Wyner-Ziv video codec**

As seen in Figure 3, the Slepian-Wolf encoder transmits the parity bits  $P$  to the decoder. The decoder in turn uses the channel model represented by  $P(X | Y_1, Y_2)$ , and the side-information frames  $Y_1$  and  $Y_2$  to determine an estimate  $\hat{X}$  of the transmitted frame  $X$ . Thus in an actual implementation, the performance can also depend on the accuracy of the channel model, better is the channel model higher will be the gain.

### 3. Wyner-Ziv Video Codec

In Figure 4 we see a block diagram representation of the Low-Density-Parity-Code (LDPC) based Wyner-Ziv video codec. The video sequence is first partitioned into even and odd frames. The odd frames are transmitted using conventional intra-frame coding, and the even frames are Wyner-Ziv coded. The odd frames are used as side-information at the decoder.

#### 3.1. Forward Transform and Quantization

The Wyner-Ziv frames are partitioned into 4x4 pixel blocks before an integer 4x4 DCT is applied to them. The DCT transform used is the same as the fast transform used in H.264 [11]. The transform coefficients with same frequency are grouped together to form coefficient bands. Each coefficient band is then quantized with  $2^{M_k}$  uniform quantization levels where  $k$  represents the coefficient band being quantized.  $M_k$  represents the number of bit-planes and can take on values  $\{0, 1, \dots, 8\}$ . A set of quantizers was designed using training based on several sequences. A distortion constraint was chosen, corresponding to the four points used in the simulation setup. This led to the choice of the four quantizers shown in Figure 5. A zero implies no bits are transmitted and the decoder copies  $(Y_1 + Y_2)/2$  to the reconstruction frame.

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

$$Q_0$$

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

$$Q_1$$

16	8	2	0
16	2	0	0
8	0	0	0
0	0	0	0

$$Q_2$$

8	2	0	0
4	0	0	0
0	0	0	0
0	0	0	0

$$Q_3$$

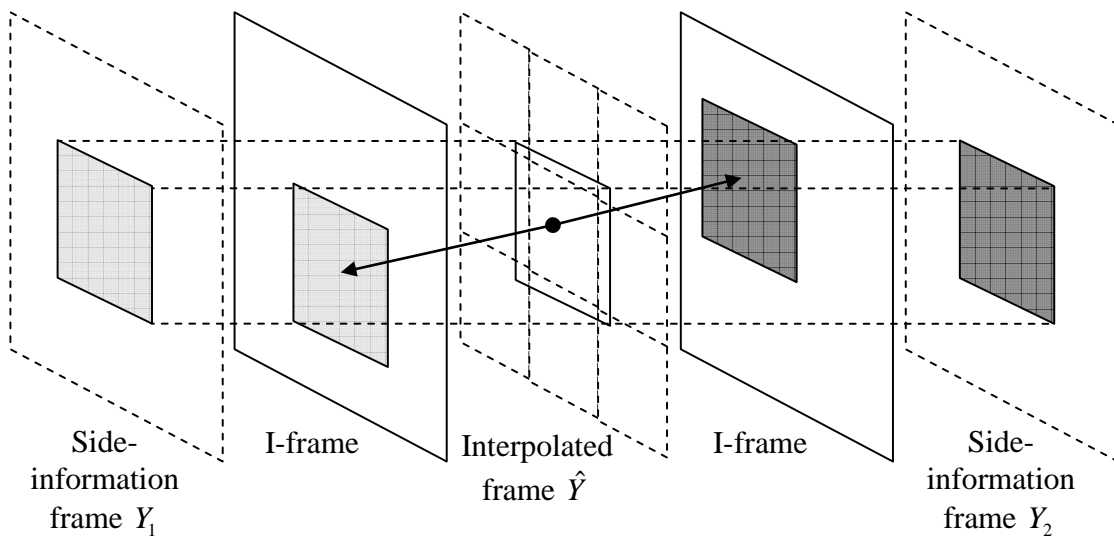
**Figure 5: Quantizers used in simulations**

### 3.2. LDPC Encoding

The quantized coefficient bands are binarized and transmitted, one bit-plane at a time, starting from the most significant bitplane (MSB). The bitplanes are fed as a vector to the LDPC encoder, which converts the input vector into a codeword. Parity bits are culled from the codeword and transmitted via a lossless channel. It is assumed that the encoder knows exactly how many parity bits need to be transmitted for successful reconstruction of  $\hat{X}$ . The simulation uses channel codes which have a regular distribution. The degree of each variable node (from node perspective) is three. An ensemble of codes, with channel code rate 0.5 to 0.97 (in steps of 0.01) were used. If the rate of the code required is less than 0.5 then the raw bits are transmitted directly to the decoder, since for rates less than 0.5, the length of parity would exceed the length of actual data. A modified version of Radford Neals package for LDPC encoding/decoding [12] was used for simulations. It can be argued that the lossless transmission of parity corresponds to Slepian-Wolf Encoding/Decoding. However, due to the lossy nature of quantization the overall codec is Wyner-Ziv.

### 3.3. Motion Estimation

At the decoder, the intra-coded frames are reconstructed. As shown in Figure 7, motion estimation is performed between the odd frames to determine the motion field. This is



**Figure 6: Symmetric Bidirectional Overlapped Block Motion Estimation**

comparable to B-frame [4] coding in conventional video codecs. The equivalent picture sequence is I-B-I where the I-frames are used for predicting the B-frames. The motion estimation algorithm performs bidirectional motion search using overlapped blocks [13]. It assumes symmetric motion vectors about the interpolated frame. The motion vector field obtained in such a manner may have low spatial coherence; this can be improved using weighted vector median filtering [14]. The weights are determined by the *Mean Square Error* (MSE) corresponding to each candidate motion vector. The motion smoothing scheme removes discontinuities in the motion field at the boundaries, it also removes outliers in the homogeneous region. The weighted median vector proposed in [15] is defined as:

$$\sum_{j=1}^N w_j \|x_{wvmf} - x_j\|_L \leq \sum_{j=1}^N w_j \|x_i - x_j\|_L \quad (\text{IV})$$

where  $x_i, i=1,2,\dots,N$  represent the motion vectors of adjacent blocks and the collocated block in the previously interpolated frame.  $x_{wvmf}$  is that candidate vector whose sum of (weighted) L-norm distances from the other vectors is the least. The weight  $w_j$  is determined as:

$$w_j = \frac{MSE(x_c, B)}{MSE(x_j, B)} \quad (\text{V})$$

where  $x_c$  represents the candidate vector of the current block  $B$ .

The motion vectors obtained after smoothing are assumed to be symmetrical. In interpolation schemes the overlapped blocks are averaged to obtain the corresponding block in the interpolated frame. In our scheme however, we do not average the overlapped blocks but place them in two different frames at collocated positions. This gives us two different side-information frames. We reason that, any difference in the predictor blocks of the adjacent frames, represent covered/uncovered regions. Averaging will lead to reduction in strength of these regions and therefore loss of information. Once the side-information frames are constructed the 4x4 DCT is performed on each block of the frame and coefficient bands are formed. The two coefficient bands along with the statistical information about the hypothetical noisy channel are used in obtaining estimate  $\hat{X}$ .

### 3.4. LDPC Decoding

The side-information coefficient bands are termed  $Y_1$  and  $Y_2$ . The LDPC channel decoder uses Belief-Propagation (BP) algorithm based on the soft-decoding approach detailed in [16][17][18] to retrieve the transmitted bits. The BP algorithm is based on confidence propagation through Bayesian Network [19]. The confidence level of each bit is set based on the channel model and the bit-received. It is termed log-likelihood ratio (LLR). The LLR of a bit  $c_i$  can therefore be setup as:

$$L(c_i) = \log \left( \frac{\Pr(X_i = 0 | Y_{i1} = y_{i1}, Y_{i2} = y_{i2})}{\Pr(X_i = 1 | Y_{i1} = y_{i1}, Y_{i2} = y_{i2})} \right) \quad (\text{VI})$$

where,  $i$  represents the  $i^{th}$  component of a vector (of length  $N$ ). The value of bit  $c_i$  depends not only on the side-information, but also on the bit-values of the previously decoded bit-planes [20], therefore the LLR can be further refined to:

$$L(c_i) = \log \left( \frac{\Pr(X_i^{(b)} = 0 | Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, X_i^{(b+1)}, X_i^{(b+2)}, \dots, X_i^{(m-1)})}{\Pr(X_i^{(b)} = 1 | Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, X_i^{(b+1)}, X_i^{(b+2)}, \dots, X_i^{(m-1)})} \right) \quad (\text{VII})$$

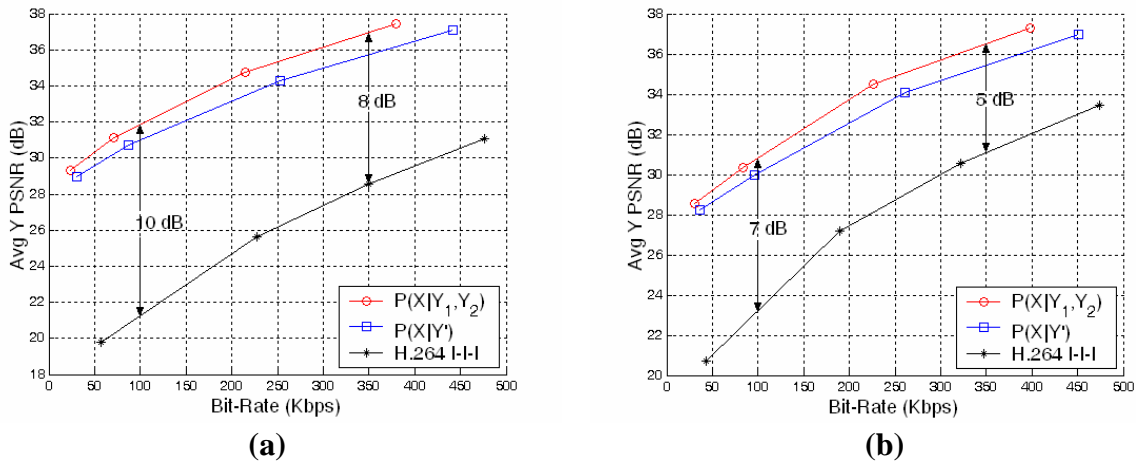
when,  $X_i$  contains  $m$  bit-planes ( $b$  represents the bitplane being decoded, the bitplanes are numbered 0 to  $m-1$ ). The BP algorithm performs Maximum-Likelihood decoding to determine an estimate  $\hat{X}$  of the transmitted vector [17]. The channel model was based on training over several sequences.

### 3.5. Inverse Quantizer and Inverse Transform

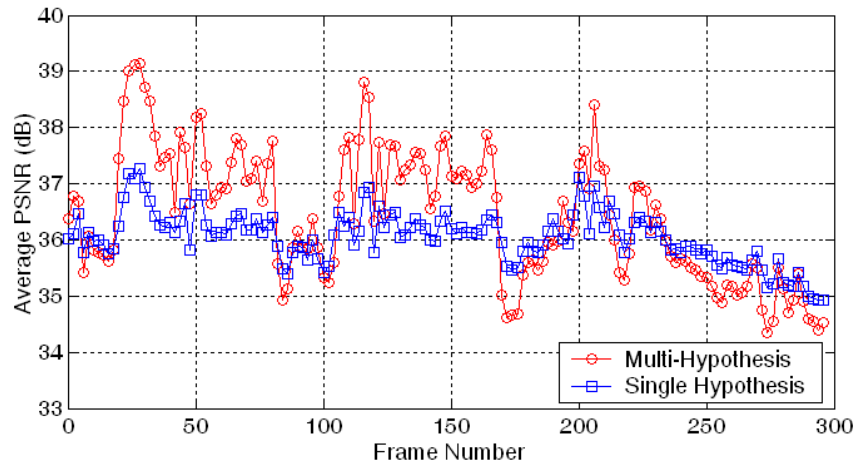
The transmitter may choose to transmit only a part of the bit-planes to conserve bandwidth; this would imply that the decoded quantization index can take a range of values, corresponding to the LSBs being set to all 0, and to all 1. This defines a continuous range of values  $X$  can take, spanning over several of the original quantization bins.  $\hat{X}$  is then chosen to be the centroid of the conditional probability distribution  $p(X | Y_1, Y_2)$ , over this range. This represents statistically the best choice to minimize MSE. Therefore,  $\hat{X} = E(X_q | q, Y_1, Y_2)$ , where  $X_q$  represents the range of  $X$  determined by  $q$ ; the bitplanes received so far. The inverse-quantizer is then used to obtain the reconstructed DCT coefficients. The coefficients are then passed to the Inverse Transform to obtain the final reconstruction of the transmitted frame.

## 4. Results

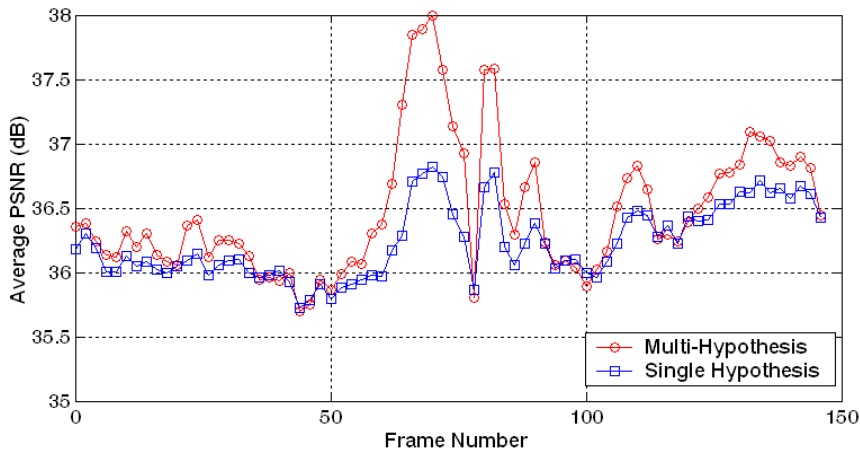
Extensive simulations were carried over a large number of sequences using the proposed video codec. However for brevity only results for Stefan-QCIF and Bus-QCIF are presented here. Figure 7 shows the performance plots of the proposed Wyner-Ziv codec, and the



**Figure 7: PSNR Vs. Bit-Rate plot for (a) Stefan-QCIF and (b) Bus-QCIF 15 fps**



(a)



(b)

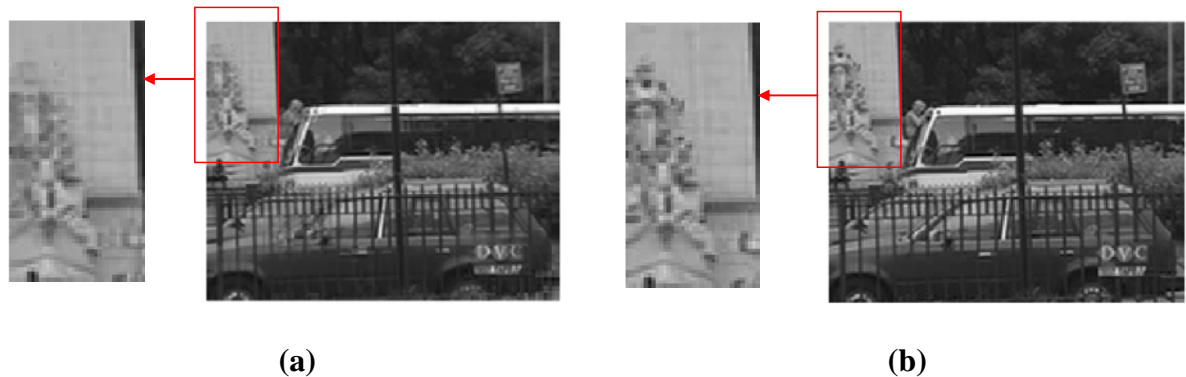
**Figure 8: PSNR level of Wyner-Ziv frames using the two schemes for**

**(a) Stefan QCIF and (b) Bus QCIF**

interpolated scheme. The plots show average Y PSNR (dB) Vs. Bit-rate (Kbps) of Y-frames, for both the sequences. The original sequences had a frame-rate of 30 fps, however, since only the even frames were Wyner-Ziv coded the results represent data for a frame rate of 15 fps. The scheme assumes same rate and quality for the odd frames and hence their effect on the PSNR plots is not considered. As can be seen, the PSNR of the proposed video codec is about 1 dB better than the performance of an interpolation based codec. A purely I-coded H.264 (JM9.6, baseline, UVLC), video stream performs 5-10 dB below the proposed codec. A subjective evaluation of the two schemes leads to the conclusion that the overall picture quality for the proposed codec is better. Figure 8 shows the (Wyner-Ziv) frame level average Y PSNR for the two sequences using quantizer  $Q_0$  (the fourth point in Figure 7), for comparing the two schemes. Multi-Hypothesis outperforms the single hypothesis case in most case. Figure 9 and Figure 10, show a sample frame for comparison, as seen, the proposed video codec has higher clarity at lower bit-rates. Simulations were also carried out for sequences with low-motion content like Mother and Daughter-QCIF, however the performance gains for such sequences were only modest and in the region of 0.25 dB. Nevertheless it is noteworthy that, in most band-width constrained applications the video sequence is usually decimated in time, this would increase the amount of motion in a sequence, thereby increasing the utility of the proposed codec.



**Figure 9: Stefan-QCIF, frame 36 of (a) Interpolation scheme (452 Kbps) and (b) Proposed codec (388Kbps)**



**Figure 10: Bus-QCIF, frame 16 of (a) Interpolation Scheme (461Kbps) and (b) Proposed codec (407Kbps)**

## 5. Conclusion

In this paper we propose a pure multi-hypothesis based Wyner-Ziv video codec. The proposed codec makes use of better channel representation to achieve lower bit rates. It was shown that performance gains of up to 1dB over interpolation based schemes can be attained. The theoretical justification for such a scheme lies in Shannon's Data Processing theorem. The proposed scheme can be extended even further to include more hypothesis, however prior experience in traditional video coding suggests that increasing the number of hypothesis need not give significant performance benefits, and comes at cost of increased complexity.

## Acknowledgements

The authors would like to thank KiMoon Lee for useful discussions on LDPC based source coding. We would also like to thank Keyur Desai for key insights in estimation theory.

## References

- [1] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, CA, Nov. 2002.

- [2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471-480, July 1973.
- [3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1-10, January 1976.
- [4] *ISO/IEC International Standard 14496-2:2001/Amd 2*, "Information Technology – Coding of Audiovisual Objects – Part 2: Visual, Amendment 2: Streaming Video Profile".
- [5] *ISO/IEC International Standard 14496-10:2003*, "Information Technology – Coding of Audiovisual Objects – Part 10: Advanced Video Coding".
- [6] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction", *Proc. IEEE Data Compression Conference (DCC)*, 1999.
- [7] J. García-Frías, "Compression of correlated binary sources using turbo codes," *IEEE Communications Letters*, vol. 5, no. 10, pp. 417-419, Oct. 2001.
- [8] R. Puri and K. Ramchandran, "PRISM: a new robust video coding architecture based on distributed compression principles," *Allerton Conference on Communication, Control and Computing*, Allerton IL, 2002.
- [9] A. Aaron, S. Rane, E. Setting, and B. Girod, "Transform-domain Wyner-Ziv codec for video," presented at the *SPIE Visual Communications and Image Processing Conf.* San Jose, CA, 2004.
- [10] T. M. Cover and J. A. Thomas, "Elements of Information Theory," *Wiley Series in Telecommunications*.
- [11] "H.264/MPEG-4 Part 10: Transform & Quantization" White Paper at <http://www.vcodex.com>
- [12] Radford Neal, "Software for Low Density Parity Check (LDPC) Codes," <http://www.cs.toronto.edu/~radford/ldpc.software.html>
- [13] M. T. Orchard, G. J. Sullivan, "Overlapped block motion compensation: an estimation-theoretic approach," *IEEE Tran. on Image Processing*, vol. 3, No. 5, pp. 693-699, Sept. 1994.
- [14] T. Viero, K. Oistamo and Y. Neuvo "Three-dimensional median related filters for color image sequence filtering," *IEEE Trans. Circ. Syst. Video Technol.*, Vol. 4, No. 2, pp. 129-142, 1994
- [15] J. Ascenso, C. Brites, F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak Republic, July 2005.
- [16] R.G. Gallager, "Low Density Parity Check Codes", Cambridge, MA: *MIT Press*, 1963.
- [17] W.E. Ryan, "An introduction to LDPC codes," in *CRC Handbook for Coding and Signal Processing for Recoding Systems (B. Vasic, ed.)*, CRC Press, 2004.
- [18] D. J. C. MacKay and R. M. Neal, "Near-Shannon-limit performance of low-density parity-check codes", *Electron. Lett.*, vol. 32, pp. 1645-1646, Aug. 1996.
- [19] K. Murphy, Y. Weiss, M. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, Stockholm, Sweden, 1999, pp. 467-475.
- [20] A. Sehgal, A. Jagmohan, N. Ahuja, "Scalable Video coding Using Wyner-Ziv Codes", *Picture Coding Symposium*, San Francisco, California, USA, December 2004.