Low Complexity Distributed Video Coding with Golay Codes

Aparna P. and Sumam David

***Abstract*—In this paper we present a low complexity video coder working on the principle of distributed source coding that combines the principle of channel coding with source coding. In this work the encoder complexity is shifted to the decoder to support uplink friendly video applications, simultaneously achieving the rate-distortion performance of the conventional predictive coding system. In this work concept of syndrome coding with Golay codes is adopted for compression. The simulation results presented in this paper reveals the superior performance of this distributed video coder over the Intraframe coders in terms of rate distortion performance, simultaneously achieving low complexity when compared to predictive coders.**

***Index Terms*—Syndrome coding, compression, uplink, distributed source coding.**

I. Introduction

Current popular video standards like ISO MPEG and ITU-H.26x have been successful in delivering the required compression and quality standards to support sophisticated multi- media applications. These standards have an interframe coding model that exploits source statistics at the encoder resulting in a very high compression performance at an acceptable quality level. The interframe coding model uses motion estimation and compensation algorithm at the encoder that removes the temporal redundancy between frames. The motion estimation and compensation process amounts to 80% of the encoder complexity and computational resources resulting in a bulky encoder with higher power consumption. This feature makes it inappropriate for uplink friendly applications like mobile video cameras, wireless video sensor networks, wireless surveillance etc. These wireless video applications demand a simple en- coder since power, size and the computational resources are of primary concern in the wireless scenario. Distributed Video Coding is a new coding paradigm that attempts to fulfill the requirement of wireless-video applications. In this scheme the complex motion search algorithm at the encoder is eliminated, but can be incorporated at the decoder. Distributed video coder thus exploits the source statistics at the decoder alone, interchanging the traditional balance of complex encoder and simple decoder. A video codec working on this principle is thus very promising for uplink friendly video applications. In such a coding system the encoder encodes each video frame separately with respect to the correlation statistics between itself and the side information. The decoder decodes the frames jointly using the side information available only at the decoder. This video paradigm is as opposed to the conventional coding system where the side information is available both at the encoder and decoder. Alternatively, we have the intraframe coding scheme (Motion-Jpeg) with a low complexity encoder but a poor rate-distortion performance as the temporal redundancy across the frames are not taken care. However it provides robustness against channel errors [1] as every frame is coded independent of others. All these factors lead to the development of a new coding paradigm called the Distributed video coding that needs to incorporate within itself the merits of interframe coding scheme and intraframe coding scheme. Such an architecture promises higher compression efficiency, robustness to wireless channel loss and at the same time distribution of complexity between encoder and decoder.

Distributed Video coding concept is based on the information theoretic bounds established in 1970s by Slepian- Wolf [2] for distributed lossless coding and by Wyner-Ziv [3] for lossy coding with decoder side information. Distributed video coding specifically works on the principle of Wyner- Ziv coding considering a distortion measure. In such a coding system the encoder encodes each video frame separately with respect to the correlation statistics between itself and the side information. The decoder decodes the frames jointly using the side information available only at the decoder. This video paradigm is as opposed to the conventional coding system where the side information is available both at the encoder and decoder.

In distributed video coding environment side information Y is treated as the noisy version of the source X i.e. $X = Y + N$. Statistically dependent side information Y, is available only at the decoder and let X be a source that is to be transmitted using least average number of bits. The encoder must therefore encode X in the absence of Y, where as the decoder jointly decodes X using Y. In this context X is compressed to syndrome S of a channel code [5]. These syndromes identify the coset to which X belongs to. The receiver on receiving the syndrome S identifies the code word from the corresponding coset that is close to the side information Y. In this paper, we present an approach used for syndrome coding of video based on the principle of distributed source coding and compare it with H.263+ Intra coding [6], H.264 Intra coding [7], H.263+ Inter coding [6]. We also compare the complexity and performance of this work with [8] which presents a syndrome coding approach based on LDPC codes.

Manuscript received May 19, 2012; revised June 13, 2012.
The authors are with the National Institute of Technology, Surathkal, Karnataka, India (email: p.aparnadinesh@gmail.com; sumam@ieee.org)

II. Proposed Scheme: Syndrome Coding with Multilevel Coset Coding and Golay Codes

A. Encoder

The encoder block diagram of the current implementation is as shown in the Fig. 1.

The first frame is coded as intra frame. Each of the consecutive frames is intercoded. An intra frame is introduced after an interval of *i (15 or 30)* frames.

Intracoding **:** Each 8x8 block of the frame is transformed using Discrete Cosine Transform and then these coefficients are zig zag scanned. These coefficients are then quantized and entropy coded using Huffman and run length coding.

Intercoding **:** Block DCT of 8x8 is applied to each block in the frame that is to be intercoded. The transformed coefficients are zigzag scanned so that they are arranged in the order of their importance. These transformed coefficients are then formed into coefficient bands such that each coefficient at the same spatial location within a block is combined together in one coefficient band. Coefficient band *coeffBand0* corresponds to all DC coefficients and hence is very significant. Around ¼th of the coefficients in a block are chosen for intercoding and hence number of bands is limited to 16 in a block of size 8x8. The remaining 3/4th of the coefficients are less important and hence can be quantized and entropy coded like intra blocks. These coefficients are insignificant and hence contribute less to the compression performance. Each of the coefficient bands is assigned different number of bits for quantization. More number of bits is assigned to higher bands and less number of bits to lower bands. Bit allocation is also based on offline training done on various video data sets. Proper decoding of the syndrome coded bits requires that the band step size is greater than the correlation noise. Hence different sets of bit allocation are pre-defined. One of these data sets is chosen based on the input data. The bit allocation data set is chosen based on the average of the correlation noise between each band and the corresponding coefficient band of the side information, such that the above criteria are fulfilled. Each of these coefficient bands are uniformly quantized with reference to the bits allocated for each band.

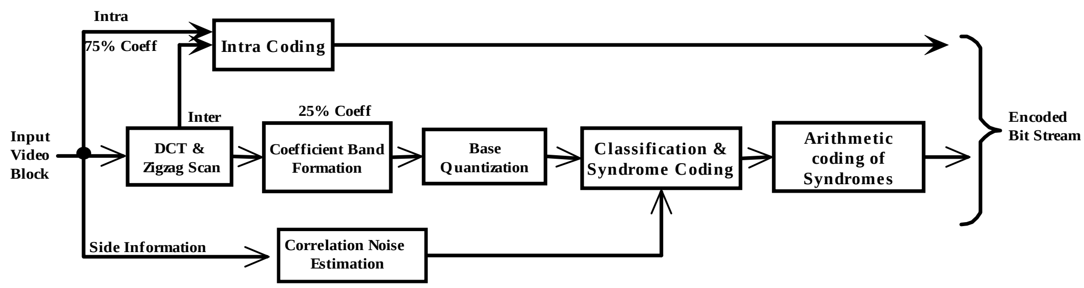

Fig. 1. Video Encoder

1) Classification and Syndrome Generation:

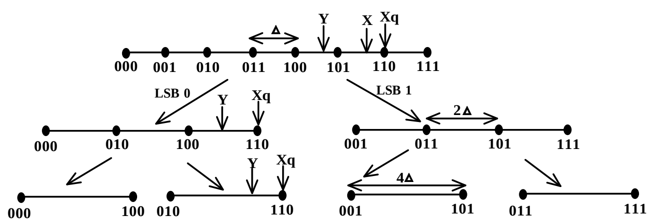

Fig. 2. The quantization bin [5]

Syndrome generation is based on the principle of multilevel coset code proposed in [5]. Syndrome for each quantized coefficient is generated based on the correlation noise between the source and the side information. In this case, syndrome corresponds to the bits that cannot be inferred from the best side information. Fig. 2 shows the quantization bin, where Xi represents the source, Xqi quantized value of the source, Yi represents side information, Δ represents the stepsize of the corresponding band. Ni represents the correlation noise, given by Ni =Xqi− Yi. In the presence of side information Yi the number of least significant bits that needs to be communicated to the encoder is given by

$$l_i = 2 + [\log_2(\frac{|(N_i)|}{\Delta})]; N_i > \Delta \quad (1)$$

$$l_i = 0; else \quad (2)$$

The syndrome bits to be communicated to the decoder can be obtained by

$$S_i = X_{qi} \,\&\, (2^{li} - 1) \quad (3)$$

where & represents bitwise AND operation.

The number of least significant bits l_i, for each coefficient should also be sent along with these bits. Hence for each coefficient, the value of l_i and l_i number of least significant bits are mapped to a unique symbol

$$S_{ti} = 2^{l_i} + S_i \quad (4)$$

where + denotes bitwise OR operation.

Transmitting syndrome bits is equivalent to dividing the quantization lattice into sub lattices as shown in the Fig. 2 up to the level specified by l_i.

Thus the number of least significant bits that needs to be communicated to decoder is dependent on the correlation noise between the source *X* and the side information *Y*. Bit planes marked in gray in Fig.3 are transmitted to the decoder and the bit planes in white can be inferred from the side information. More the correlation noise, more number of bits is to be transmitted to the decoder. This corresponds to

Wyner-Ziv coding where in the least significant bits that cannot be obtained from the side information is transmitted to the decoder. The unique symbol S_{ti} obtained from l_i and S_i is then arithmetic coded and transmitted to the decoder.

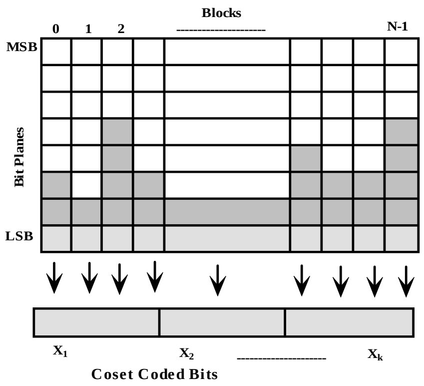

Fig. 3. Bitplanes of a single coefficient band

2) Coset Channel Coding Using Golay Code:

The bit rate can be further reduced by coding few bitplanes with respect to a channel code. If high complexity of the decoder is acceptable coset channel coding can be considered. These bits in grey as shown in Fig.3 cannot be generated by side information and hence are transmitted to the decoder. However this may incur higher bitrate. In order to further reduce the bitrate some of the least significant bitplanes are further compressed based on the principle of distributed source coding [5]. Distributed video coding in [8] uses LDPC codes for coset channel coding. This gives only a marginal gain in the bitrate with reasonable quality, but at the cost of very high complexity at the decoder. It also needs a very long block length n to give a good performance. In this method (23,12,7) Golay code is used where the block length n is 23 and the message bit length k is 12 and the hamming distance d_{min} between the code words is 7. In this method the parity check matrix H of a (23,12,7) Golay code is used to generate the syndrome bits S_i from the input bits X_i. One or two least significant bitplanes of the syndrome bitplanes can be considered for coset channel coding which are formed into a block of n bits each as shown in Fig.3. Each of the n bits X_i data block is transformed into $(n - k)$ syndrome bits by using the parity check matrix H according to $S_i = HX_i$. Thus n data bits are compressed to $(n - k)$ syndrome bits giving a compression rate of $(n/n - k)$.

3) Side Information

If a less complex encoder as well as decoder is desirable we consider the co located block of the reference frame as side information. Instead we can generate additional CRC bits on the quantized codewords so that motion search for the side information can be incorporated at the decoder.

B Decoder

The block diagram of the decoder is shown in the Fig.4. The frames that are intracoded are passed through an entropy decoder, then dequantized and Inverse transformed to get back the intra coded frame.

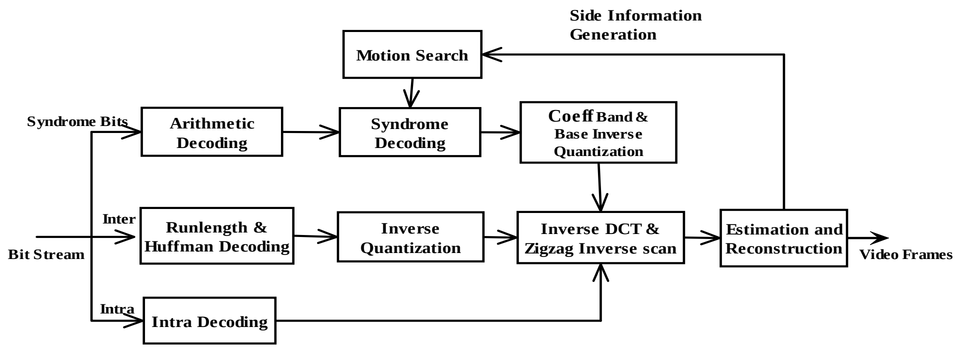

Fig. 4. Video Decoder

1) Syndrome Decoding and Coset Channel Decoding:

The coefficients that are syndrome coded are first passed through the arithmetic decoder to decode the unique symbol S_{ti}. From this symbol S_{ti} the number of least significant bits l_i and the syndrome bits S_i are obtained. At the decoder only the syndrome bits S_i and the side information bits Y_i are available. As long as the hamming distance between X_i and Y_i is less than d_{min} of the Golay code, the data X_i can be recovered. The syndrome bits S_i indicates the coset to which X_i belongs to. The coset leader of that coset is chosen as say A_i. We then find $X_i^{'}$ a codeword corresponding to the Golay code closest to $Y_i^{'}$. The required data bits X_i is then obtained from $X_i^{'}$ as $X_i = X_i^{'} \oplus A_i$. These decoded bits X_i are combined with rest of the multilevel coded bits to reconstruct back the current frame. We then find $Y_i^{'}$ such that $Y_i^{'} = Y_i \oplus A_i$, where $\oplus$ is bitwise EXOR operation. Side information generation unit provide the best side information Y_i for the current frame. Based on the side information Y_i the rest of the MSBs are retrieved and combined together with the LSBs to form the current frame coefficient band for X. The coefficient bands are uniform dequantized based on the bit allocation set chosen. Rest of the coefficient that is intracoded are further combined with syndrome decoded coefficients and then block IDCT is applied to get back the original frame.

2) Side Information Generation:

In case of complex decoder, side information is generated by performing the motion search and quantizing the candidate block and computing the CRC. If the CRC matches, that block is considered as side information.

By using a *(n, k)* linear channel code, the encoding rate achieved is $(n - k)/n$. In the work [8] the syndrome coding of the inter coefficients was done using 3/4 or 1/2 rate LDPC code. It is observed that for a ½ rate LDPC code, the correlation of the sources with the side information should be very high, which other wise would result in high distortion. On the other hand the use of 3/4 rate LDPC encoder results in less distortion, but the compression achieved is quite low. The other issues with the LDPC code are the parity check matrix H, which needs to be generated in real-time or stored at the encoder. This would increase the complexity of the encoder as storing increases the resource requirement and power consumption. Also LDPC codes require long block lengths and high decoding complexity. As the block length of the data to be coset channel coded is small, LDPC code seems to be unsuitable for the current implementation. Hence in this work Golay codes have been considered for coset channel coding. The encoder of this method is simple satisfying the main objective of this work. Also quality of the reconstructed sequence is good with Golay codes for a 1/2 rate encoder.

III. Simulation Results

In this work video codec is designed for a single camera scenario which is an application to wireless network of video camera equipped with cell phones. Encoder allows the storage of one previous frame. Objective performance evaluation of the system is done by comparing the bit rate and the Peak Signal to Noise Ratio (PSNR) between the original and the reconstructed video sequence. Four test video QCIF sequences with a resolution of 176x144 are considered for evaluating the rate distortion. These video files are considered based on their motion content. It is seen that *container.qcif* has least motion content when compared to all the other sequences where as *football.qcif* has highest motion content, *foreman.qcif* and *news.qcif* have moderate motion content. The Luma PSNR is computed for various bit rates as shown in the Table I. The performance of the Distributed video coding discussed in this paper are compared with the H.263+ Intra [6], H.264/AVC Intra [7], syndrome coding with LDPC codes [8] and H.263 + Inter coding standards [6] at frame rate of 30fps. The H.264/AVC Intra coder is the H.264/AVC coder (JM 16.0 reference software) in the intra mode of operation without exploiting temporal redundancies, but with a very efficient spatial redundancy reduction technique. For a specific condition it is observed from Table I, that H.263+ Inter gives good rate distortion performance for all the files except *football* and *foreman*. This is because *football* video sequence has high motion content (less correlation) which the predictive coder cannot efficiently code. From the simulation results we can see that the DVC implementation perform considerably well for video sequences with higher motion content i.e foreman and football. For video sequences with less motion content (i.e more correlation between adjacent frames) H.263+ Inter coder performs better than the DVC implementation. However the results of DVC implementation discussed in this paper are consistently better than H.263+ Intracoder and H.264 Intra coder. Also an improvement of at least 3dB is seen in the current implementation when compared to the syndrome coding with LDPC codes [8].

TABLE I: Comparison of Rate-Distortion Performance of All the Qcif Files with The Resolution of 176x144 for A Frame Rate of 30fps

File	Bitrate Kbps	Luma PSNR (dB) for different Methods				
		H.263+Intra	H.264 Intra	Syndrome coding with LDPC codes	DVC Method	H.263+Inter
container	220	28.72	29.47	29.03	34.47	37.83
	280	30.16	30.94	30.41	35.92	39.06
News	250	28.44	29.93	30.04	32.29	36.89
	320	29.06	30.68	30.95	33.91	39.47
Foreman	290	29.56	30.98	31.16	33.14	34.65
	360	30.65	32.37	32.84	35.30	35.51
Football	380	26.07	26.22	28.09	29.52	25.24
	480	26.81	27.33	28.76	30.93	25.60

IV. Conclusion

In this method we have introduced low complexity distributed video coder using Golay codes. From the simulation results presented it is shown that the rate-distortion performance of a distributed video coder is better than intraframe coders, with a same complexity level. Also we observe that performance of this coder is very close to that of interframe coders. However by proper modeling of correlation noise we can further improve the performance. The main aim of this work is to reduce encoder complexity making it pertinent to uplink friendly architecture which seems to be satisfied. It is also observed that the current implementation operates well in high quality (PSNR of order of 30dB) regime. The extension to lower bit rates without any compromise in the quality so that it is comparable with the conventional codecs can be further considered without increasing the complexity.

References

[1] Bernd Girod, Anne Margot Aaron, Shantanu Rane and David Rebollo-Monedero,"Distributed Video Coding ", in *Proc. IEEE*, vol. 93, Issue 1, pp 71 - 83 , Jan 2005.

[2] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources ", *IEEE Trans. on Information 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. on Information Theory*, vol. IT-22, no. 1, pp. 1-10, Jan. 1976.
[4] A.D Wyner, "Recent Results in the Shannon Theory ", *IEEE Trans. on Information Theory*, vol. 20, no. 1, pp. 2-10, Jan. 1974.
[5] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction", *Proc. IEEE Data Compression Conf.* , Snowbird, UT, pp. 158 -167, Mar. 1999.
[6] G.Cote, B.E rol, M.Gallant, and F.Kosssentini, "H.263+:Video coding at low bitrates ", *IEEE Trans. on circuits and systems for video technology*, vol. 8, no. 7, pp 849-866, November 1998.
[7] T. Wiegand, G.J.Sullivan, G.Bjontegaard and A.Luthra, "Overview of the H. 264/AVC video coding standard ", *IEEE Trans on circuits and systems for video technology*, vol. 13, no. 7, pp 560-576, 2003.
[8] Aparna P, Sivaprakash Reddy, and fSumam David, "Distributed video coding using LDPC codes for wireless video ", *Journal on Wireless Sensor Networks*, Scientific Research Publishing, Inc. USA, pp 334-339, Nov 2009,
[9] Rohit Puri, Abhik Majumdar and Kannan Ramachandran, "PRISM: A video Coding Paradigm with Motion Estimation at the Decoder ", *IEEE Trans On Image Processing*, vol. 16, no. 10, October 2007.
[10] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles", *Proc. Allerton Conf. on Communication, Control and Computing*, Allerton, IL, Oct. 2002.

Aparna P. did her BE in Electronics and Communication Engineering at NMAMIT, Nitte, Mangalore University, MTech in Digital Electronics and Advanced Communication at National Institute of Technology Karnataka, Surathkal, India and Ph.D from National Institute of Technology Karnataka, Surathkal, India. Currently, she is working as Assistant Professor in the Electronics and Communication Engineering Department, National Institute of Technology Karnataka, Surathkal, India. Her areas of interest are image processing, video compression, signal processing and embedded systems.

Sumam David completed her BTech in Electronics and Communication Engineering from College of Engineering, Thiruvananthapuram, MTech in Communication Systems and High Frequency Techniques and PhD in Electrical Engineering at Indian Institute of Technology, Madras, in 1985, 1986 and 1992, respectively. She is currently Professor in the Electronics and Communication Engineering Department, National Institute of Technology Karnataka, Surathkal, India. Her research interests are in multimedia signal compression and VLSI architectures for signal processing. She is a Senior Member of IEEE.