# Distributed Video Coding with Unsupervised Learning of Motion Compensation

***Project Proposal for EE398B Image Communication II***

Young Min Kim, Stephanie Kwan, and Karen Zhu

{jinhae, skwan1, kyzhu}@stanford.edu

## I. Introduction

In the conventional approach of video coding, the statistical dependencies of frames are exploited at the encoder side. This approach makes the encoder much more complicated then the decoder, and calls for huge computational power at the encoder side. While this design achieves high rate-distortion performance, it is not suitable for applications such as surveillance camera and sensor networks where memory and computation are scarce. A new paradigm of video coding, namely Distributed Video Coding [1], which shift the complex computation from the encoder to the decoder side, is recently introduced. This new paradigm is based on channel coding theory that exploits the statistical dependencies between the frames at the decoder side. It encodes the frames independently, but decodes them jointly with some side information. Different approaches based on Slepian-Wolf coder for lossless coding, and Wyner-Ziv coder for lossy coding have been proposed.

## II. Background

### *A. Slepian-Wolf Theorem for Lossless Distributed Coding*

Distributed coding refers to a coding scheme where two dependent random sequences are separately encoded but jointly decoded. Suppose we have two statistically dependent i.i.d. sequences X and Y. When these two sequences are encoded and decoded separately, we can achieve rates of $R_X \geq H(X)$, $R_Y \geq H(Y)$. However, if these two sequences are encoded separately, but decoded jointly, the Slepian-Wolf Theorem [2] shows that we can achieve lower rates. In particular, the theorem states that

$$R_X + R_Y \geq H(X,Y)$$
$$R_X \geq H(X|Y) \;,\; R_Y \geq H(Y|X)$$

Consider a special case of distributed coding where Y is some side information available only at the decoder but not at the encoder side. If X is a sequence on the encoder side that is statistically dependent on Y, then, from the Slepian-Wolf Theorem, X can be encoded at a rate of $R_X \geq H(X|Y)$, even though the encoder does not have any information on Y [1].

### *B. Wyner-Ziv Coding*

While Slepian and Wolf established the lower bound for lossless distributed coding, Wyner and Ziv extended this idea of separately encoding and jointly decoding to lossy compression with side information at the decoder side. Suppose again, Y is some i.i.d. side information at the decoder side, and X is an i.i.d. random sequence at the encoder side that is statistically dependent on Y, and X' is the reconstruction of X at the decoder side. Denote distortion D = E[d(X,X')], $R^{WZ}_{X|Y}(D)$ is the achievable bit rate lower bound for distortion D using Wyner-Ziv coding, and $R_{X|Y}(D)$ is the rate required if side information is available at the encoder side. Wyner and Ziv showed that in general, there is a rate loss of $R^{WZ}_{X|Y}(D)$ - $R_{X|Y}(D) \geq 0$ when the side information is not available at the encoder. However, when the source is Gaussian and memoryless. When distortion D is calculated using mean-squared error distortion, then $R^{WZ}_{X|Y}(D)$ - $R_{X|Y}(D)$ =0 [1].

## III. Previous work

### *A. Wyner-Ziv Video Codec with Intra-frame Encoding and Inter-frame Decoding*

In [3] and [4], a lossy distributed video coding scheme is suggested using the Wyner-Ziv coder and a reconstruction/interpolation algorithm depicted as in the following diagram.

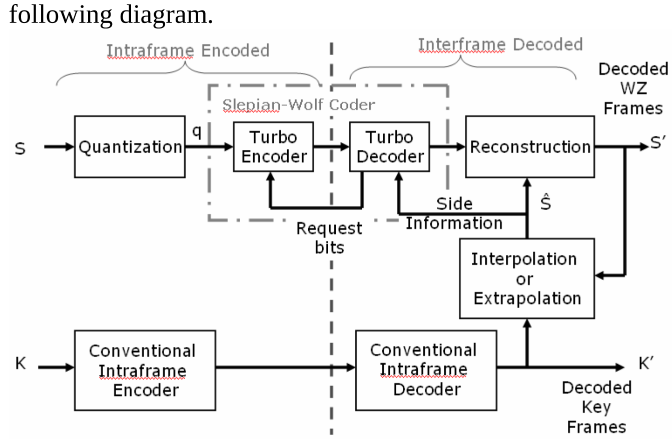


Fig. 1. Wyner-Ziv video codec with intraframe encoding and interframe decoding. [4]

At the encoder side, a subset of frames, regularly spaced in sequence, are intra-coded in the conventional way to serve as key frames in order to provide side information. The rest of the frames are turbo coded, and only the parity bits are sent across the network. At the decoder side, the parity bits of the current frame are decoded together with the interpolated key frames.

### *B. Distributed Compression of Stereo Images*

[5] and [6], on the other hand, applied the above approach to a stereo image system, but using a LDPC based Slepian-Wolf coder for lossless compression. A machine learning algorithm based on expectation maximization (EM) is also introduced to do unsupervised learning for disparity estimation. The system is shown in the following diagram.

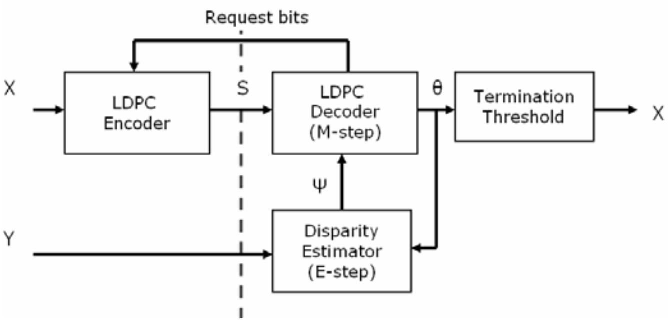


Fig. 2. Distributed compression of stereo images with unsupervised learning of disparity D via EM. [5]

## IV. Proposed work

We are going to implement a lossless pixel-domain distributed video coder with frame to frame motion field estimation achieved by EM algorithm as depicted below.

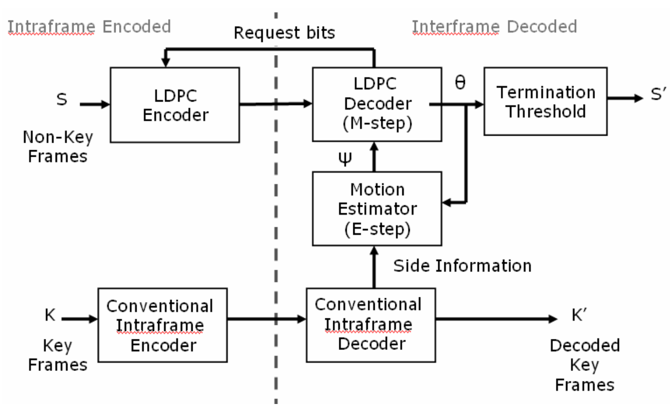


Fig.3 Proposed Distributed Video Coder with Unsupervised Learning of Motion

We will evaluate the performance of this specific distributed video coding scheme in comparison with the following schemes:

1) Slepian-Wolf bound.
2) Distributed coding with no motion estimation.
3) Distributed coding using a motion oracle.
4) Wyner-Ziv video codec with intraframe encoding and interframe decoding.

Our approach will be similar to the Wyner-Ziv coding scheme described in [3] and [4], but we will adopt a lossless scheme without the quantization step and use a LDPC coder instead of turbo coder to generate syndromes. We will also implement EM algorithm introduced in [5] and [6] to perform frame to frame motion field estimation. In addition, we will look into the Wyner-Ziv residual coder [7], and possibly apply the concept to the current proposed coder. Finally, we will explore possibilities of the following extensions:

1) Extending motion compensation from one-dimensional space to two-dimensional space.
2) Extend and optimize searching range for motion compensation.
3) In the EM algorithm, try to shorten the learning time by starting with the motion vector found for previous frame.

## References


[1] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, “Distributed Video Coding”, *Proceedings of the IEEE,* vol. 93, no. 1, pp. 71-83, January 2005. *Invited Paper.*

[2] D. Slepian, J. K. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inform. Theory 19 (4) (1973) 471–480.

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

[4] A. Aaron, E. Setton and B. Girod, "Towards practical Wyner-Ziv coding of video", Proc. IEEE International Conference on Image Processing, ICIP-2003, Barcelona, Spain, Sept. 2003.

[5] D. Varodayan, A. Mavlankar, M. Flierl and B. Girod, "Distributed coding of random dot stereograms with unsupervised learning of disparity," *Proc. IEEE International Workshop on Multimedia Signal Processing, MMSP 2006,* Victoria, Canada, October 2006.

[6] D. Varodayan, A. Mavlankar, M. Flierl and B. Girod, "Distributed grayscale stereo image coding with unsupervised learning of disparity," *Proc. IEEE Data Compression Conference, DCC 2007,* Snowbird, UT, March 2007.

[7] A. Aaron, D. Varodayan and B. Girod, "Wyner-Ziv residual coding of video," Proc. Picture Coding Symposium, PCS-2006, Beijing, China, April 2006.