

# LORD: LOw-complexity, Rate-controlled, Distributed video coding system

Rami Cohen and David Malah  
 Department of Electrical Engineering  
 Technion - Israel Institute of Technology  
 Technion City, Haifa 32000, Israel  
 Email: rc@tx.technion.ac.il, malah@ee.technion.ac.il

**Abstract**—Distributed Video Coding (DVC) is an emerging coding scheme that employs principles of source coding with side information (SI) at the decoder. In this paper, we present a new DVC encoder, named LORD (LOw-complexity, Rate-controlled, Distributed video coding system). No feedback channel is used in our encoder, and an adaptive noise model that varies both spatially and temporally is employed. The SI is created using motion extrapolation, resulting in a low delay encoding process. We extend LORD for encoding videos acquired by Bayer sensors in endoscopy, in which only partial information of the colors in each pixel is known. This special video format has not been addressed yet in the DVC framework. We show that, using LORD, a significant improvement in performance is achieved over a standard intra-coding method with a similar complexity, on a set of examined videos.

**Index Terms**—Bayer format, distributed video coding, endoscopy videos, motion extrapolation, wyner-ziv coding

## 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 spatio-temporal prediction and block-wise 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 paradigm, known as Distributed Video Coding (DVC), has emerged in the last decade. This

paradigm employs principles of 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).

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.

In the case of the first two, the employed noise model is obtained by offline training, in spite of the spatially and temporally varying statistics of any typical video sequence. In the latter one, the noise is estimated online, but the estimation process requires a delay of one frame. In addition, the last two systems use a feedback channel (between the encoder and the decoder), which limits their usefulness for real time applications and also leads to complex rate control schemes.

In this paper, we propose a new feedback-less DVC system. We use a Laplace distribution model, whose parameter varies on-line, spatially and temporally, adapted to the changes in the video scene. We also employ an efficient rate control scheme for meeting channel rate constraints. The SI is obtained using motion extrapolation, avoiding the delay of at least one frame needed by interpolation.

In addition, we show how to modify our codec for compressing endoscopy videos, which are recorded using an endoscope in medical operations. These videos are usually acquired using Bayer sensors [6], in which only one color component is known for each pixel. Recently, endoscopy videos are transmitted using a wireless channel, where low complexity and low delay encoders are crucial.

This paper is organized as follows. The proposed video codec is presented in Section II. Adaptation of this codec for encoding endoscopy videos is given in III, and performance evaluation is given in Section IV. The paper is concluded in Section V.

## II. PROPOSED CODEC

The aim in the design process of the proposed codec, which we have named LORD (LOw-complexity, Rate-controlled, Distributed video coding system), is to address several issues that limit the performance and the practical use of current DVC systems.

First, the noise statistics in LORD is updated on-line, at the decoder, using both spatial and temporal information. Second, the rate control is performed at the encoder, in a highly accurate way, without being affected by the decoder. Moreover, no feedback channel is used in our implementation, and minimal delay is incurred by the process of creating the SI, which is created at the decoder.

During the design process, we focused on the general requirement of low-complexity. Accordingly, the components of LORD involve simple and straight-forward calculations only, enabling its implementation in resource limited environments. A block diagram of the LORD codec is given in Figure 1, and its components are described below.

#### A. Encoder

The main new features in this encoder, compared with existing ones, are the rate-distortion optimization (RDO) and rate control (RC) components. We work with group of pictures (GOP), which are each composed of key (intra) and WZ frames. The available budget of bits is distributed among the frames inside each GOP using the RDO module, where the RC component ensures that the transmission rate satisfies the rate constraint of the transmission channel.

We will concentrate on the case of a GOP that consists of two frames: a key frame and a WZ frame (in this order). It has been shown that using this size of GOP usually provides the best results, where in cases in which an improvement is achieved using a larger GOP, it is minor [7]. The main components of LORD's encoder are depicted in Figure 1a, and are described below.

1) *Block Classification*: In this stage, the coding mode of each block is decided. This decision is made according to the *residual energy*  $E_d$  of the difference between each block and its co-located one in the previous frame. Following a similar principle used in PRISM, there are three possible coding modes:

- **SKIP** If the energy  $E_d$  is smaller than a predefined threshold,  $SKIP_{TH}$ , the current block is not encoded, and the decoder simply copies its co-located block from the previous frame.
- **INTRA** If the energy  $E_d$  is higher than a predefined threshold,  $INTRA_{TH}$ , the block is encoded using conventional intra-coding. The intra-coding method used in LORD is JPEG, which offers a low computational complexity and good performance. In addition, human vision system considerations are also taken into account in the quantization method of JPEG.
- **COSET** If the energy  $E_d$  is between  $SKIP_{TH}$  and  $INTRA_{TH}$ , the block is encoded using DVC principles. In order to ensure a reasonable quality of the decoded block, only the first 15 AC coefficients (*WZ coefficients*) are encoded in this mode, where the DC and the remaining AC coefficients are encoded using JPEG. The encoder quantizes uniformly each of the WZ coefficients to symmetric  $2^m$  levels ( $m = 3$  in our implementation). The maximal absolute differences  $V_k$  between each of

the current 15 AC coefficients ( $k = 2, 3, \dots, 16$ ) and the coefficients of the co-located blocks in the previous frame are transmitted losslessly (using a Huffman code) to the decoder, as well as the quantization indices of the coefficients.

The thresholds  $SKIP_{TH}$  and  $INTRA_{TH}$  were chosen according to an offline training phase. After the classification stage, the INTRA and COSET blocks undergo an  $8 \times 8$  discrete cosine transform (DCT).

2) *Rate-Distortion Optimization*: Assuming that due to channel constraints, we have only  $B$  available bits (after taking into account the bits assigned to the WZ coefficients) for encoding the frames in the GOP, we would like to distribute them in an optimal way inside the GOP. For this purpose, a rate distortion model is used, which relates the distortion to the number of bits allocated for encoding the frames.

We use a relatively simple and well-known RD model, with the necessary adaptations needed for our case. A detailed description of this model is given in [8]. The main parts of this model are as follows.

Assume that we have  $P$  random variables (possibly with different probability distributions):  $X_1, X_2, \dots, X_P$ , with zero mean and variances  $\sigma_i^2$  ( $i = 1, 2, \dots, P$ ). The distortion (measured as the mean squared-error, MSE) incurred when uniformly quantizing  $X_i$  using  $b_i$  bits (i.e.,  $2^{b_i}$  quantization levels) can be modelled as:

$$D_i(b_i) = h_i \sigma_i^2 2^{-2b_i} \quad (1)$$

where the constants  $h_i$  are determined by the probability density function (PDF) of  $X_i$  [8].

In our case, we refer to the DCT coefficients as being emitted by the random variables above. We treat the DCT bands in each frame of the GOP separately, since different coding modes are used in each frame, and the content of the frames varies.

Now, assume that there are  $M_{key}$  blocks in the key frame,  $M_{coset}$  COSET blocks and  $M_{intra}$  INTRA blocks in the WZ frame (SKIP blocks are not taken into account since they do not consume bits). The overall distortion, taking into account the number of blocks and their coding modes, is:

$$D = \underbrace{\sum_{i=1}^{P/2} m_i h_i \sigma_i^2 2^{-2b_i}}_{\text{Key frame distortion}} + \underbrace{\sum_{i=P/2+1}^P m_i h_i \sigma_i^2 2^{-2b_i}}_{\text{WZ frame distortion}} \quad (2)$$

where  $m_i$  denotes the number of intra-coded coefficients in the  $i^{\text{th}}$  band ( $i = 1, 2, \dots, P$ , where  $P = 128$  for a GOP of size 2).

Writing the total distortion more compactly, we formulate an optimization problem:

$$\min_{b_i} D = \sum_{i=1}^P m_i h_i \sigma_i^2 2^{-2b_i}, \quad \text{s.t.} \quad \sum_{i=1}^P b_i \leq B \quad (3)$$

where  $b_i$  denotes the number of bits assigned to the  $i^{\text{th}}$  DCT band. Based on [8], we set  $h_i = h_G = \frac{\sqrt{3}\pi}{2}$  for the DC

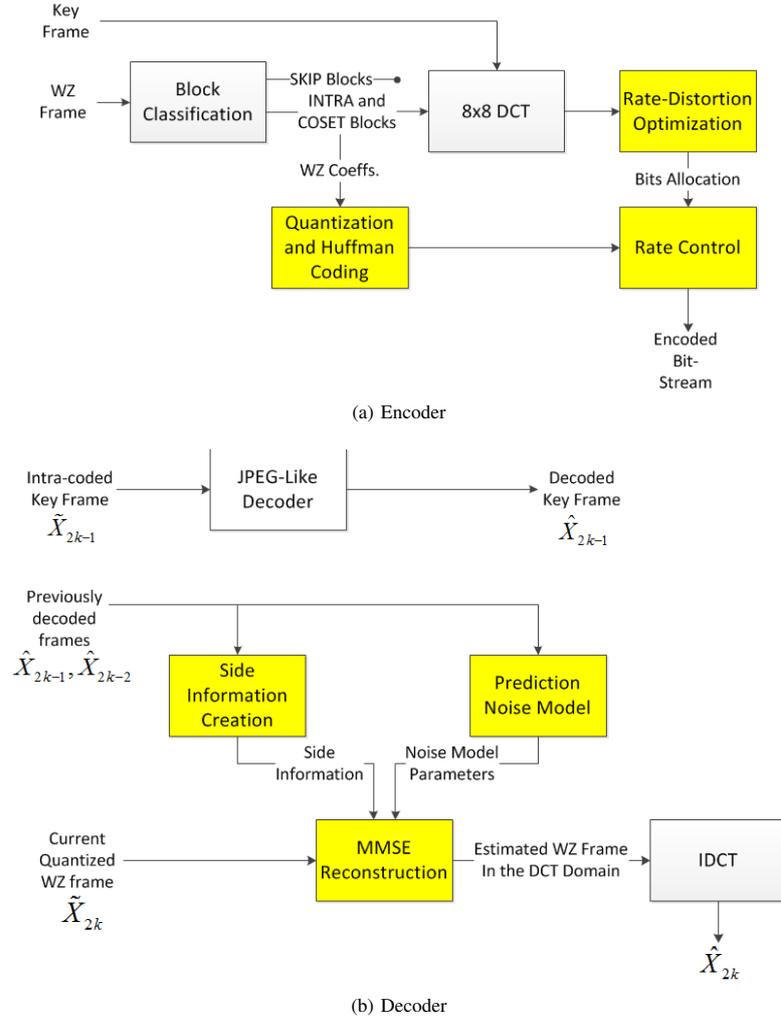


Figure 1: LORD codec (highlighted: our main contributions)

band, which is assumed to have a Gaussian distribution, and  $h_i = h_L = \frac{\theta}{2}$  for the AC bands, which are assumed to have a Laplace distribution. The empirical maximum-likelihood (ML) estimator of the variance of each band is calculated according to its distribution. For a DC band, the ML estimator is:

$$\sigma_G^2 = \frac{1}{N_G} \sum_{j=1}^{N_G} x_j^2 \quad (4)$$

where  $x_j$  denotes a realization (coefficient) that is associated with the DC band, and  $N_G$  is the total number of realizations in this band. The ML estimator for an AC band is:

$$\sigma_L^2 = 2 \left( \frac{1}{N_L} \sum_{j=1}^{N_L} |x_j| \right)^2 \quad (5)$$

where  $x_j$  denotes a realization that is associated with the AC band, and  $N_L$  is the total number of realizations in this band.

The solution of the optimization problem (3), obtained using the method of Lagrange multipliers, is:

$$b_i = \bar{b} + \frac{1}{2} \log_2 \frac{\sigma_i^2}{\sigma^2} + \frac{1}{2} \log_2 \frac{h_i}{H} + \frac{1}{2} \log_2 \frac{m_i}{M} \quad (6)$$

where  $\bar{b} = \frac{B}{P}$  and:

$$\sigma^2 = \left( \prod_{i=1}^P \sigma_i^2 \right)^{1/P}, H = \left( \prod_{i=1}^P h_i \right)^{1/P}, M = \left( \prod_{i=1}^P m_i \right)^{1/P} \quad (7)$$

Finally, the total number of bits that are allocated to each frame is:

$$B_{key} = \sum_{i=1}^{P/2} b_i, B_{wz} = \sum_{i=P/2+1}^P b_i \quad (8)$$

3) *Rate Control*: Given the number of bits allocated to each frame in the GOP, as determined in (8), our task now is to employ a rate control (RC) scheme that ensures that each frame uses its assigned share of bits.

Since the rate is affected essentially by the coefficients that are encoded using JPEG (that is, the intra-coded blocks and the intra part of the COSET blocks), the rate can be controlled by changing the quantization matrices of JPEG throughout each frame, in order to achieve the desired bitrate.

The parameter that controls the quality in JPEG is the *scale factor parameter*  $q$ . This parameter multiplies the base quantization matrix of JPEG, where each DCT band is quantized uniformly using a different quantization step  $step_i(q)$  ( $i$  denotes the index of the DCT band).

He and Mitra [9] have shown that there is a linear relationship between the coding bit rate  $R$  of an image and the fraction of zeros among its quantized transform coefficients, denoted by  $\rho$  ( $0 \leq \rho \leq 1$ ):

$$R(\rho) = \theta(1 - \rho) \quad (9)$$

where  $\theta$  (the negative of the slope) is a constant related to the image content that depends mainly on the amount of texture and edges in the image. In the framework of JPEG, we first determine  $\rho$  using  $q$ :

$$\rho(q) = \frac{1}{N} \sum_i \sum_{j: |x_{i,j}| \leq step_i(q)} 1 \quad (10)$$

where  $i = 1, 2, \dots, 64$  for INTRA blocks or  $i = 2, \dots, 16$  for COSET blocks,  $j$  denotes the indices of the realizations of these bands, and  $N$  is the total number of the DCT coefficients that are not SKIP encoded, calculated separately for each frame in the GOP. Once  $\rho$  is known, we can simply get the associated rate using (9). However, we still have to estimate the parameter  $\theta$ . For this purpose, we use an adaptive estimation for  $\theta$  (adopted from [9]) that maintains the relation (9) for each frame. An example for achieving a desired rate using this RC scheme can be seen in Figure 2.

According to our experiments, using the proposed RC scheme, the average deviation of the rate from the desired one is less than 1%. It should be noted that we can deal in a simple manner with cases in which the resulting rate using this RC scheme is higher or smaller than required. In such cases, an additional pass over the blocks can be done, in which bits are removed or added according to a predefined criterion.

## B. Decoder

The components of LORD's decoder are depicted in Figure 1b. Its main components are described next.

1) *Side Information Creation*: The side information, which serves as a prediction of the WZ frame, is created at the decoder using *motion extrapolation* (MX). That is, a prediction  $Y_{2k}$  of the current input Wyner-Ziv frame,  $X_{2k}$ , is created at the decoder, using two previously decoded frames,  $\hat{X}_{2k-1}$  and  $\hat{X}_{2k-2}$ . This is done by an iterative process, as follows:

- 1) **Initialization** An offset, which is a two-dimensional vector  $(o_x, o_y)$  that determines the 2D offset of  $\hat{X}_{2k-1}$

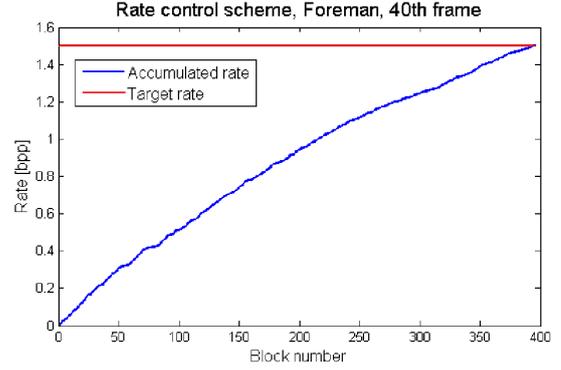


Figure 2: RC scheme, applied to Foreman 40th frame

from its upper-left corner is set to  $(0, 0)$ , and an iteration parameter  $i$  is set to 1.

- 2) **Motion estimation** Motion vectors are estimated for  $8 \times 8$  blocks, between  $\hat{X}_{2k-2}$  and  $\hat{X}_{2k-1}$ , where the latter is offsetted by  $(o_x, o_y)$ . The MVs are obtained using motion estimation with quarter-pel (qpel) precision (achieved using, twice, the standard H.264 interpolation filter).
- 3) **Smoothing** The motion field is smoothed by replacing each MV by the median of this MV and its 4 nearest MVs. It should be noted that since we deal with vectors, the median vector is selected as the vector with the minimal sum of (Euclidean) distances to the others.
- 4) **Projection** The pixels from  $\hat{X}_{2k-1}$  are projected to a frame  $Z_i$  using the motion vectors (at qpel precision) obtained before, as depicted in Figure 3. If there are several sources for the same pixel in  $Z_i$ , their average value is used.
- 5) **Update** Set  $(o_x, o_y) := (o_x + 2, o_y + 2)$  and  $i := i + 1$ .
- 6) **Loop** Stages 2-5 are repeated until the percentage of non-covered pixels in the disjoint union of  $Z_i$  falls below a predefined threshold (e.g., 1%), or until a maximal number of iterations ( $i$ ) is reached.

The extrapolated frame is defined as the average of the  $Z_i$ 's, where the averaging is performed pixel-wise, according to the total number of occurrences of an extrapolated value of each pixel in the  $Z_i$ 's. If there are pixels with no value ("holes") in the extrapolated frame, a spatial interpolation (average of 4 nearest neighbours) is used.

2) *Prediction Noise Model*: Once the side information is created at the decoder, it is used in order to decode the COSET-coded blocks. As mentioned earlier, only the first 15 AC coefficients of these blocks are encoded, by sending the indices of their quantization interval.

In order to determine the most probable location of each such coefficient within its known coarse quantization interval, that is, to "correct" the "noisy" (quantized) COSET coefficients, a Laplacian noise model between the WZ frame and its prediction is assumed. The PSNR gain using this model,

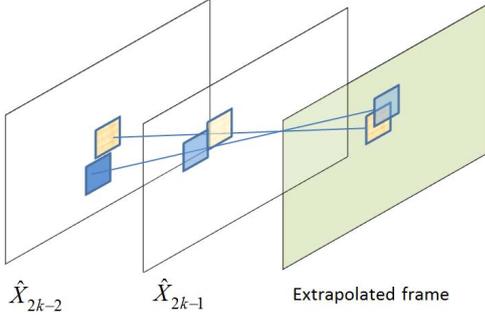


Figure 3: Extrapolation process

compared with simply using the center of each quantization interval as the decoded value, is 1.1-1.7dB, for the videos examined in our experiments.

The parameters of this noise model are obtained as follows. First, we calculate the displaced block differences between  $\hat{X}_{2k-2}$  and  $\hat{X}_{2k-1}$ , using the MVs obtained in the ME process. Only blocks that are further extrapolated to COSET blocks in the extrapolated frame are taken into account. Then, the histogram of the DCT-transformed differences is used for estimating the  $\alpha$  parameter of Laplace distribution of the noise ( $\mathcal{N}$ ) between the WZ frame ( $X$ ) and its MX-predicted frame ( $Y$ ):

$$f_{X|Y}(x) = f_{\mathcal{N}}(x - y) = \frac{\alpha}{2} e^{-\alpha|x-y|} \quad (11)$$

where  $\alpha$  is calculated using the ML estimator, for each COSET-coded DCT band, and for each WZ frame. Now, given that the boundaries of the quantization interval (to be denoted as  $z_i$  and  $z_{i+1}$ ) of the current DCT coefficient  $x$  are known, we get an MMSE estimate of this coefficient, using both these boundaries and the side information  $y$ :

$$\hat{x} = \mathbb{E}[x | x \in [z_i, z_{i+1}), y] = \frac{\int_{z_i}^{z_{i+1}} x f_{X|Y}(x) dx}{\int_{z_i}^{z_{i+1}} f_{X|Y}(x) dx} \quad (12)$$

The integrals in (12) can be carried out analytically, resulting in [10]:

$$\hat{x} = \begin{cases} z_i + \frac{1}{\alpha} + \frac{\Delta}{1 - e^{\alpha\Delta}} & \text{if } y < z_i \\ y + \frac{(\gamma + \frac{1}{\alpha}) e^{-\alpha\gamma} - (\delta + \frac{1}{\alpha}) e^{-\alpha\delta}}{2 - (e^{-\alpha\gamma} + e^{-\alpha\delta})} & \text{if } y \in [z_i, z_{i+1}) \\ z_{i+1} - \frac{1}{\alpha} - \frac{\Delta}{1 - e^{\alpha\Delta}} & \text{if } y \geq z_{i+1} \end{cases} \quad (13)$$

where  $\Delta = \Delta(V) = \frac{V}{2^{m-1}}$  is the quantization interval length, which depends on the maximal DCT difference  $V$  between co-located blocks (for simplicity we omit the band index  $k$  from  $V$ ),  $\gamma \triangleq y - z_i$  and  $\delta \triangleq z_{i+1} - y$ .

Once the DCT coefficients of the WZ frame are recon-

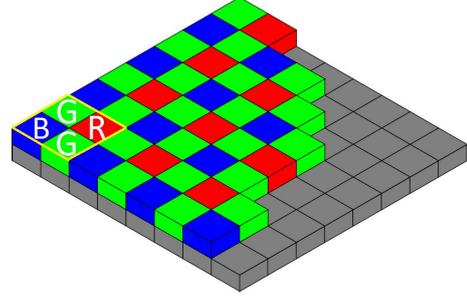


Figure 4: Bayer format

structed, using (13) and the INTRA/SKIP blocks, they undergo inverse DCT (IDCT), as defined in the JPEG standard, resulting in a reconstructed GOP in the pixel domain.

### III. ADAPTATION OF LORD TO ENDOSCOPY VIDEOS

Endoscopy (meaning "looking inside") refers to looking inside the body for medical reasons using an endoscope, an instrument used to examine the interior of a hollow organ or cavity of the body. The video recorded during the endoscopy process is obtained using a camera that is attached to the endoscope.

There is a shift recently to transmission of endoscopy videos over a wireless channel (this enables the physician to work in a more sterile environment). In this case, the transmission rate of video content during the endoscopy process is limited due to limited power resources. This fits well into the principles that are behind DVC. That is, there is a need for a low-complexity encoder, while allowing high-complexity decoder, which is essentially a computer connected to the endoscope.

Endoscopy videos are usually recorded using Bayer format, in which only one color component (R, G or B) is known for each pixel, as depicted in Figure 4. The full color image is obtained after a process called *demaosaicing* [11]. In the case of such videos, we encode each color component separately. This way, the spatial correlation of each color channel is exploited. We also employ a modified rate distortion model, which takes into account the different weights of the colors in the total distortion.

Usually, the quality measure in the case of compressing Bayer videos is calculated as the PSNR between the  $Y$  (luminance) components of the resulting color images after demosaicing (between the original image and the decompressed one). The relation between the RGB components and  $Y$  is as follows (according to ITU-R BT.601 standard):

$$Y = w_R R + w_G G + w_B B \quad (14)$$

where the weights of the color components for determining  $Y$  are  $w_R = 0.299$ ,  $w_G = 0.587$  and  $w_B = 0.114$ .

Taking these weights into account, and assuming that the RGB components are (approximately) independent, the expectation of the squared error between the  $Y$  components is

approximated as:

$$\varepsilon_Y^2 \approx w_R^2 \varepsilon_R^2 + w_G^2 \varepsilon_G^2 + w_B^2 \varepsilon_B^2 \quad (15)$$

where  $\varepsilon_X^2 = \mathbb{E} \left[ \left( X - \hat{X} \right)^2 \right]$  ( $X = R, G, B$ ) is the squared error between each of the original and the reconstructed color components.

Now, we formulate a similar optimization problem to (3), where we express the total distortion as the sum of the distortions from the RGB components, weighted according to (15). The solution of this optimization problem is a simple extension of (6), with an additional term that is the geometric mean of the weights of the colors (where only intra-coded coefficients are taken into account) [12]. Finally, the available bits are distributed among the RGB component, which are later encoded as standard videos, as described in Section II.

#### IV. PERFORMANCE EVALUATION

We begin by presenting compression results obtained using LORD on the sequences Foreman, Football and Coastguard (luminance only). The resolution used is  $176 \times 144$  (QCIF), and the frame rate is 15Hz. The first 100 frames of each sequence were encoded. The results are shown in Figure 5.

The PSNR gain, compared to Motion JPEG, is about 2dB for Foreman, and up to 1dB for Football, where the most noticeable improvement is obtained for Coastguard, in which a gain of up to 4dB in PSNR is achieved. This can be explained by the linear motion of the objects in Coastguard, which is well suited for the linear movement assumption in the motion extrapolation process.

As expected, the performance of H.264 IPPP and H.264 INTRA, which are significantly more complex than LORD (where in the case of H.264 IPPP there is also a delay of 3 frames in the encoding process), provide better results than LORD.

In addition, the performance of LORD was evaluated for Bayer endoscopy videos. Two videos were used, simulating two typical endoscopy operations. The first one is a simulation of a surgery performed on a chicken, and the second one is a simulation of capsule endoscopy of the gastrointestinal tract (in which the patient swallows a capsule that contains a tiny camera).

The results for endoscopy videos are shown in Figure 6. A significant PSNR gain over Motion JPEG is seen, where in the case of the simulation on the chicken the gain is up to 5dB. LORD also outperforms H.264 INTRA for both endoscopy videos. The good performance of LORD in this case can be attributed to the linear motion of most of the objects in endoscopy videos, resulting in a side information that is close to the WZ frame. In addition, the performance of LORD are closer to H.264 IPPP than in the case of the tested standard videos.

#### V. CONCLUSION

In this paper we focused on the design process of a distributed video coding system, which we named LORD. We

proposed an online varying noise model, which adapts itself to the content of the video.

We applied rate control at the encoder, rather than the rate control used in most of the existing DVC systems, which relies heavily on probability models that depend on the decoder, thus inexact in nature. The proposed rate control scheme is not affected by the decoder and is found to be highly accurate.

In addition, no feedback channel, which is used in most of the current DVC systems, is needed in our codec. The use of this channel incurs unknown delay in the encoding/decoding process, and is unsuitable for real time tasks or even for some offline tasks such as storage. The SI creation process does not require any additional delay, due to the use of motion extrapolation, rather than methods that use motion interpolation.

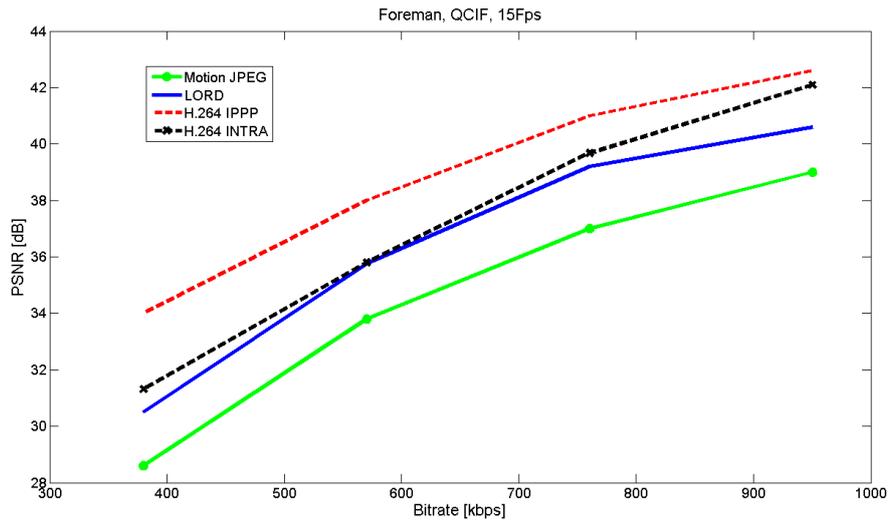
We have also shown how to adapt LORD to the task of compressing endoscopy videos, which are recorded using a Bayer sensor. Our experimental results demonstrated the better performance of LORD compared to Motion JPEG, and even to H.264 INTRA in the case of endoscopy videos.

#### Acknowledgement

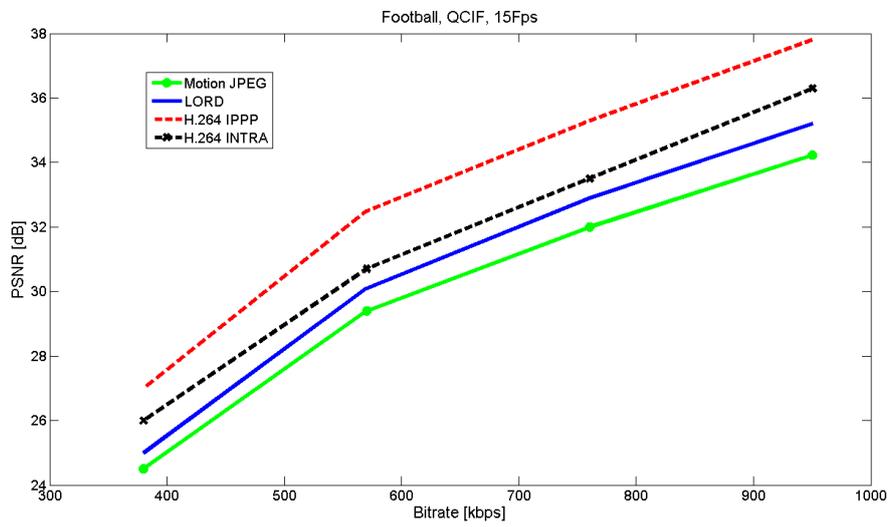
The authors would like to thank D. Adler, S. Wolf and N. Skarbnik from Gyrus ACMI Inc. (Israel branch) for providing samples of simulated endoscopy videos. The support of the staff of the Signal and Image Processing Lab (SIPL) is greatly appreciated.

#### REFERENCES

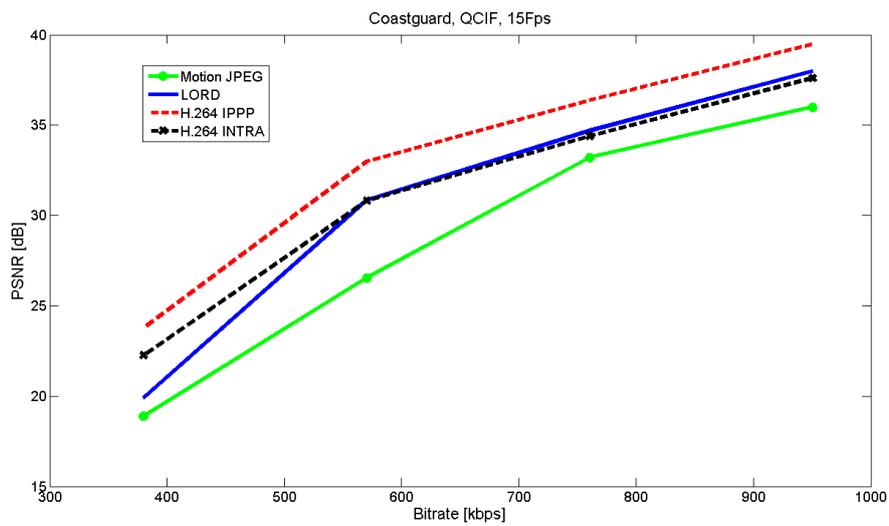
- [1] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973.
- [2] A. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 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] A. Davies and P. Fennessy. *Digital imaging for photographers (Fourth ed.)*. Focal press, 2001.
- [7] P. L. Dragotti and M. Gastpar. *Distributed Source Coding: Theory, Algorithms and Applications*. Academic Press, 2009.
- [8] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer academic publishers, 1992.
- [9] Z. He and S. K. Mitra. A linear source model and a unified rate control algorithm for DCT video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(11):970–982, 2002.
- [10] D. Kubasov, J. Nayak, and C. Guillemot. Optimal reconstruction in Wyner-Ziv video coding with multiple side information. In *IEEE 9th Workshop on Multimedia Signal Processing*, pages 183–186, Oct. 2007.
- [11] R. A. Maschal et al. Review of Bayer pattern CFA demosaicing with new quality assessment algorithms. In *SPIE Conference Series*, volume 7662, Apr. 2010.
- [12] R. Cohen. Feedback-less distributed video coding and its application in compressing endoscopy videos. M.Sc. thesis, Technion, Israel Institute of Technology, July 2012, [http://sipl.technion.ac.il/siglib/FP/Rami\\_Cohen\\_MSc\\_thesis.pdf](http://sipl.technion.ac.il/siglib/FP/Rami_Cohen_MSc_thesis.pdf).



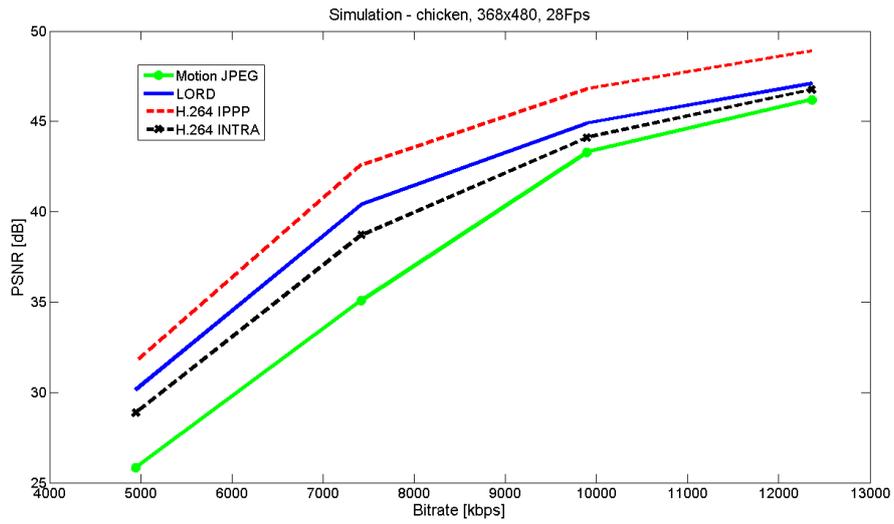
(a) Foreman



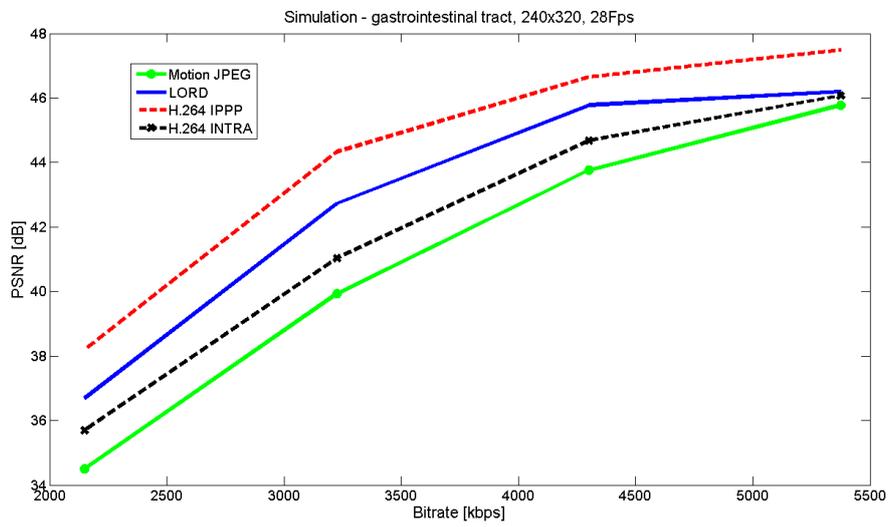
(b) Football



(c) Coastguard



(a) Simulation - chicken



(b) Simulation - gastrointestinal tract

Figure 6: LORD: Compression results, endoscopy videos