Feedback-less Distributed Video Coding and its Application in Compressing Endoscopy Videos

Rami Cohen

Signal and Image Processing Lab
Department of Electrical Engineering
Technion - Israel Institute of Technology

26 June, 2012

M.Sc. Research
Supervised by Prof. David Malah
Distributed Video Coding (DVC)

- Why DVC?
- Theoretical Background
- Standard Video Encoders
- DVC Systems - Overview
Outline

1. Distributed Video Coding (DVC)
   - Why DVC?
   - Theoretical Background
   - Standard Video Encoders
   - DVC Systems - Overview

2. LORD: LOw-complexity, Rate-controlled, Distributed video coding system
   - Motivation
   - Encoder
   - Decoder
   - Adaptation to Endoscopy Video Compression
Outline

1. Distributed Video Coding (DVC)
   - Why DVC?
   - Theoretical Background
   - Standard Video Encoders
   - DVC Systems - Overview

2. LORD: LOw-complexity, Rate-controlled, Distributed video coding system
   - Motivation
   - Encoder
   - Decoder
   - Adaptation to Endoscopy Video Compression

3. Experimental Results
Outline

1. Distributed Video Coding (DVC)
   - Why DVC?
   - Theoretical Background
   - Standard Video Encoders
   - DVC Systems - Overview

2. LORD: LOw-complexity, Rate-controlled, Distributed video coding system
   - Motivation
   - Encoder
   - Decoder
   - Adaptation to Endoscopy Video Compression

3. Experimental Results

4. Conclusion
Why DVC?

- There are cases in which standard (complex) encoders are impractical.
- DVC paradigm offers low complexity encoders with good performance.

Limited-complexity video encoders: Examples
What is the minimal rate required for encoding two correlated sources $X$ and $Y$ separately when they are jointly decoded?

Slepian-Wolf Theorem (1973)

- Surprisingly, given that: $R_X \geq H(X|Y), R_Y \geq H(Y|X)$, Slepian & Wolf have shown that:

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

- $Y$ is referred to as side information
Wyner-Ziv Theorem (1976)

- When a distortion $D$ is allowed, Wyner and Ziv have shown that:

$$R_{X|Y}^{WZ}(D) \geq R_{X|Y}(D)$$

- Special case in which equality holds:
  1. $X = Y + N$ where $N$ is Gaussian and independent of $Y$
  2. MSE distortion metric

- The rate loss is bounded [Zamir 98]:

$$R_{X|Y}^{WZ}(D) - R_{X|Y}(D) \leq 0.5\text{bits/sample}$$
Hybrid video encoders

- The side information (SI) in modern (hybrid) video encoders such as MPEG-2 and H.264 is created by:
  1. Temporal prediction (constituting up to 70% of the encoder’s complexity)
  2. Spatial prediction (known as H.264 INTRA)
- The SI is available both at the encoder and the decoder
- *Master-Slave*: Complex encoder, simple decoder
- Impractical in power or resources limited encoders
Main Parts

- Usually, the input is separated into key (intra-coded) frames and Wyner–Ziv (WZ) frames (Group of Pictures (GOP) of size 2)
- Side information creation: prediction ($Y$) of the frame to be encoded ($X$), created at the decoder
- Noise prediction model: estimating $X$ from $Y$
  - Probability distribution models for $N = X - Y$

![Diagram of DVC System](image)
SI Creation Example: Interpolation/Extrapolation
Motivation for developing LORD

Adaptation to the Video statistics
- On-line estimation of the parameters of the noise model

Rate control
- Not affected by the decoder
- Suitable for channels with constant rate constraint

Low delay
- No feedback-channel

Medical application
- Adaptation to the compression of endoscopy videos
Scheme

Key Frame → Block Classification → SKIP Blocks → INTRA and COSET Blocks → 8x8 DCT → Rate-Distortion Optimization

WZ Frame → WZ Coeffs. → Quantization and Huffman Coding → Rate Control

Bits Allocation → Rate Control

Encoded Bit-Stream
Encoder

Block Classification

![Diagram showing block classification with different colors indicating SKIP, COSET, and INTRA operations.]

Rami Cohen (Technion)
**COSET mode**

- The differences between the first 15 AC coeffs. are quantized uniformly to symmetric $2^m$ levels.
- The quantization indices are sent using Huffman code.
- The remaining coeffs. are INTRA coded.

**Diagram:**

- $m=3$
- $2^{m-1}$ bins
- Residual
- $V_x$ for $p_4, p_3, p_2, p_1$
- $-V_x/2^{m-1}$ for $p_4, p_3, p_2, p_1$

**Table:**

<table>
<thead>
<tr>
<th>Index</th>
<th>Codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01</td>
</tr>
<tr>
<td>2</td>
<td>00</td>
</tr>
<tr>
<td>3</td>
<td>101</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>1101</td>
</tr>
<tr>
<td>6</td>
<td>1100</td>
</tr>
<tr>
<td>7</td>
<td>1111</td>
</tr>
<tr>
<td>8</td>
<td>1110</td>
</tr>
</tbody>
</table>
Rate-Distortion Optimization

- Assuming $P$ DCT bands in the GOP:

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

- The solution:

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

$$
\bar{b} = \frac{B}{P}, \quad \sigma^2 = \left( \prod_{i=1}^{P} \sigma_i^2 \right)^{1/P}, \quad H = \left( \prod_{i=1}^{P} h_i \right)^{1/P}, \quad M = \left( \prod_{i=1}^{P} m_i \right)^{1/P}
$$
Rate Control

- Rate control (RC) algorithm is used for enforcing the optimal bit distribution.
- The following relationship between the coding bit rate $R$ and the fraction $\rho$ of zeros among the quantized intra-coded coefficients is employed [He & Mitra, 2002]:

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

- $\theta$ is a constant related to the image content.
- The number of zeros is controlled by the parameter $q$ used in JPEG.
Encoder

RC algorithm: Example

- Low complexity, one-pass algorithm
SI Creation: 1. Motion Estimation

- Qpel (quarter-pixel) full search motion estimation is performed between two already decoded frames, $\hat{X}_{2k-2}$ and $\hat{X}_{2k-1}$
- Qpel precision is obtained in $\hat{X}_{2k-2}$ using H.264 interpolation filter:

$$h = \left[ \begin{array}{ccccccccc} 1 & 0 & -5 & 0 & 20 & 32 & 20 & 0 & -5 & 0 & 1 \end{array} \right] / 32$$
1. Assuming linear motion, the pixels from $\hat{X}_{2k-1}$ are projected to the next (extrapolated) frame, which is used as side information.

2. The average value of the predictors is used in case of multiple ones.

3. Spatial interpolation is used in case of "holes".
Our motion extrapolation algorithm outperforms standard (ipel based) ones by 3-4dB on average.
The noise model is used for ”improving” the basic estimate provided by the SI

- Estimated \textit{on-line}, for each WZ frame

The noise ($N$) between $\hat{X}_{2k-2}$ and $\hat{X}_{2k-1}$ serves as an estimate of the noise between the SI ($Y$) and the WZ frame ($X$)

$N$ is assumed to be Laplace-distributed:

$$f_{X|Y}(x) = f_N(x - y) = \frac{\alpha}{2} e^{-\alpha|x-y|}$$

$\alpha$ is calculated for each band, using the ML estimator
Noise Prediction Model (Cont.)

![Histograms for Football - AC1, Football - AC2, Football - AC3, Foreman - AC1, Foreman - AC2, Foreman - AC3 with RMS Error values.](image)
MMSE Reconstruction

- The boundaries of the quantization interval (denoted \( z_i \) and \( z_{i+1} \)) of the COSET coeffs. are provided by encoder.
- An MMSE estimate of the source \( X \), using both these boundaries and the side information, is obtained:

\[
\hat{x} = \mathbb{E} [x | x \in [z_i, z_{i+1}) , y] = \frac{\int_{z_i}^{z_{i+1}} x f_{X|y}(x) \, dx}{\int_{z_i}^{z_{i+1}} f_{X|y}(x) \, dx}
\]
MMSE Reconstruction (Cont.)

- The last integrals can be carried out analytically, resulting in:

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

- \(\Delta, \gamma, \delta\) are defined according to:

\[
\begin{align*}
\Delta &= \text{Parameter Here} \\
\gamma &= \text{Parameter Here} \\
\delta &= \text{Parameter Here}
\end{align*}
\]
Endoscopy videos

Endoscopy

- Endoscopy refers to looking inside the body for medical reasons using an endoscope
- An endoscope is consisted of a long, thin, flexible tube that has a light source and an attached camera
- Recently, a shift towards transmission of endoscopy videos over a wireless channel - limited power resources
Bayer Filter

- Bayer color filter array (CFA) is composed of filter blocks of size 2x2, which are 50% green, 25% red and 25% blue.
- Conforms with the strong sensitivity of the human vision system (HVS) to green light.
- Each physical pixel has an optical filter placed over it, allowing penetration of only particular color of light (red, green or blue).
- Almost universal on consumer digital cameras, used in endoscopes.
Endoscopy Videos

Bayer Filter

- Bayer demosaicing is the process of translating a Bayer image into a full color (RGB) image

From video provided by Gyrus ACMI, Inc.
Endoscopy Videos

RGB Separation

- Each color component is compressed separately
- Exploiting the correlation between pixels of the same color

Raw Bayer

R component

G component

B component
Endoscopy Videos

Rate-Distortion Optimization

- The PSNR calculation for Bayer format is taken into account for RDO
The relation between $Y$ and $RGB$ components is (ITU-R BT.601):

$$Y = w_R R + w_G G + w_B B$$

where: $w_R = 0.299$, $w_G = 0.587$, $w_B = 0.114$

Assuming that the correlation between different color components is relatively small, we get:

$$\mathbb{E} \left[ (Y - \hat{Y})^2 \right] = \mathbb{E} \left[ (w_R (R - \hat{R}) + w_G (G - \hat{G}) + w_B (B - \hat{B}))^2 \right]$$

$$\approx w_R^2 \cdot \mathbb{E} \left[ (R - \hat{R})^2 \right] + w_G^2 \cdot \mathbb{E} \left[ (G - \hat{G})^2 \right] + w_B^2 \cdot \mathbb{E} \left[ (B - \hat{B})^2 \right]$$
Endoscopy Videos

Rate-Distortion Optimization

- The distortion is calculated separately for each color component.
- Each distortion is weighted according to $w_C^2$ ($C = R, G, B$):

$$D = w_R^2 \cdot \sum_{\text{DCT bands}} m_R h_i \sigma_i^2 2^{-2b_i} + w_G^2 \cdot \sum_{\text{DCT bands}} m_G h_i \sigma_i^2 2^{-2b_i} + w_B^2 \cdot \sum_{\text{DCT bands}} m_B h_i \sigma_i^2 2^{-2b_i}$$

- Considering the available bits $B$, we get an optimization problem.
- The solution is a simple extension of the previous one.
Experimental Results

Standard Videos

PSNR Results

- **Foreman, QCIF, 15Fps**
  - INTRA
  - LORD
  - H.264 IPPP
  - H.264 INTRA

- **Coastguard, QCIF, 15Fps**
  - INTRA
  - LORD
  - H.264 IPPP
  - H.264 INTRA
Endoscopy Videos

PSNR Results

Simulation - chicken, 368x480, 28Fps

Simulation - gastrointestinal tract, 240x320, 28Fps
New DVC codec was developed
On-line estimation of the parameters of the noise model
Rate-distortion model and rate control algorithm are used, at the encoder
No feedback channel is used
Adaptation to endoscopy videos (Bayer format)
Improvement over standard intra coding, for both standard videos and endoscopy videos
Thank You