Modeling and Rate Control in Reversed Complexity Video Coding Systems

Yevgeny Priziment

Department of Electrical Engineering
Technion - Israeli Institute of Technology
Haifa, Israel

M.Sc. Research supervised by: Prof. David Malah

Outline

Distributed Video Coding - Motivation

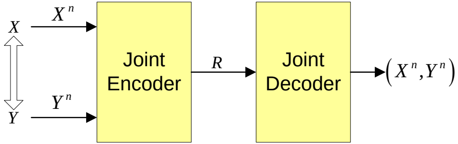
Standard Video Coders - MPEG, H.264

- Based on Motion Estimation and Transform Coding
- Complex encoder - due to ME
- Downlink oriented

New Video Applications - Wireless/Cellular Video, Surveillance

- Uplink oriented
- Low cost
- Limited power Low complexity encoder
- Limited computational resources
- Limited bandwidth $\rightarrow$ coding efficiency

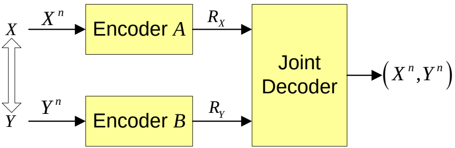
Distributed Coding of Correlated Sources

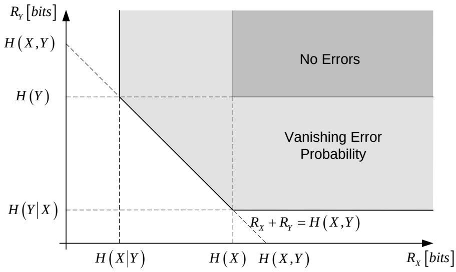


- X and Y correlated sources
- (x_i, y_i) i.i.d. $\sim p_{XY}(x,y)$; $x \in \mathcal{X}, y \in \mathcal{Y}$
- Joint encoding and joint decoding:

$$R \geq H(X,Y)$$

Distributed Coding of Correlated Sources





Naive Approach

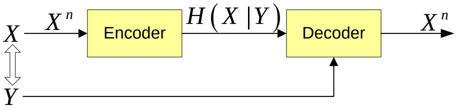
$$R_X \geq H(X)$$
$$R_Y \geq H(Y)$$
$$R_T = R_X + R_Y > H(X,Y)$$

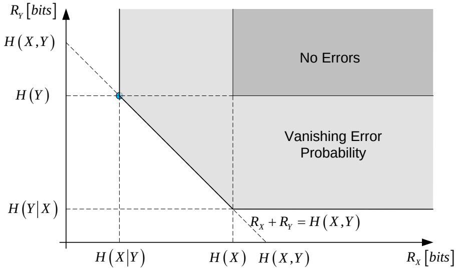
Slepian & Wolf 73

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

- SW proof is *non* constructive - based on random binning, joint typicality, asymptotic arguments ($n \to \infty$)

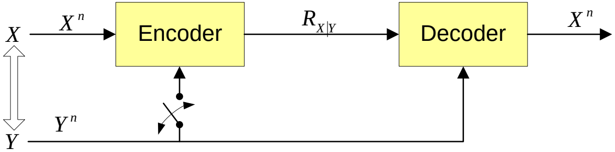
Source Coding with Side Information at the Decoder





- Y is coded at rate $R_Y \geq H(Y)$ - decoded without any information on X
- X is coded at rate $R_X \geq H(X|Y)$ - decoded using Y
- Distributed Source Coding problem is simplified into a problem of source coding with Side Information at the decoder

Practical Slepian Wolf Coding



- Y is a 'noisy' version of X at the output of a virtual channel
- Use channel coding techniques to correct the noisy Y into X

Example-Part I: SI known at the Encoder and Decoder

Setting X^n and Y^n binary words, $n = 3$ and $d_H(X^n, Y^n) \leq 1$

Encoding send the index of $U^n = X^n \bigoplus Y^n$, $i(U^n) \in \{0, 1, 2, 3\}$
$(U^n = \{000, 001, 010, 100\})$
$\textbf{ENC}(X^n = 100 | Y^n = 101) = i(100 \bigoplus 101) = 1$
$R_{X|Y} = 2/3$ bps

Decoding recover U^n and add to Y^n
$\textbf{DEC}(i | Y^n = 101) = U^n \bigoplus Y^n = 100$

Practical Slepian Wolf Coding (cont.)

Example-Part II: SI known only at the Decoder

Setting As in Part I

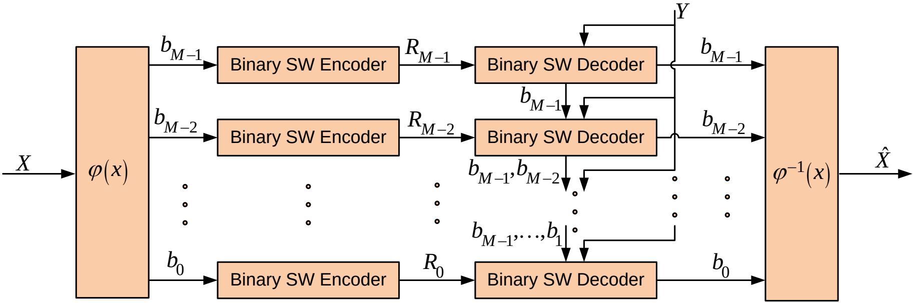
Codebook Partition all words $\{0,1\}^3$ into cosets:
$C_{00} = \{000, 111\}$, $C_{01} = \{001, 110\}$,
$C_{10} = \{010, 101\}$, $C_{11} = \{100, 011\}$

Encoding send the index of coset containing X^n
$\textbf{ENC}(X^n = 100) = 11$ $\quad R^{SW}_{X|Y} = 2/3$ bps

Decoding find a word closest to Y^n in the coset C_{11}
$\textbf{DEC}(C_{11}, Y^n = 101) = 100$

- C_{00} is a simple repetition code of rate $R_C = 1 - H(X|Y) = 1/3$ bps (C_{01}, C_{10}, C_{11} retain the distance properties of C_{00})
- In general case use capacity approaching, LDPC or Turbo codes

Slepian Wolf Coding of M-ary Sources

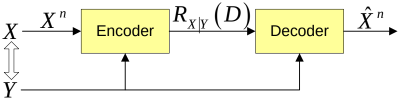


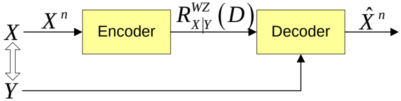
- $|\mathcal{X}| = 2^M$
- Coding rate at bitplane m:

$$R_m = H(b_{M-1}, \ldots, b_m | Y) - H(b_{M-1}, \ldots, b_{m+1}, | Y)$$
$$= H(b_m | b_{M-1}, \ldots, b_{m+1}, Y)$$

- Using Entropy chain rule:
$\sum_{m=0}^{M-1} H(b_m | b_0, \ldots, b_{m-1}, Y) = H(b_0, \ldots, b_{M-1} | Y) = H(X|Y)$

Wyner-Ziv Coding





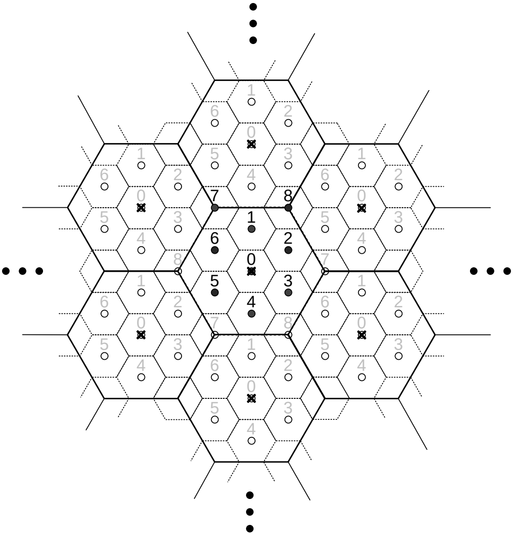
- RD function: $R^{WZ}_{X|Y}(D) \geq R_{X|Y}(D)$
- Equality holds if:
  - $X = Y + N$
  - N Gaussian and independent of Y
  - Y arbitrarily distributed
  - MSE distortion metric
- [Zamir 98] Rate loss: $R^{WZ}_{X|Y}(D) - R_{X|Y}(D) \leq 0.5$ bits/sample

Practical Wyner-Ziv Coding

Nested Lattice Quantization [Zamir *et al* 2002]

- Coarse quantizer $Q_c(\cdot)$ is nested in a fine quantizer $Q_f(\cdot)$- all bin centroids of the coarse quantizer coincide with a (regular) subset of centroids of the fine quantizer
- Attains the WZ RD bound for Gaussian sources as quantizers' dimension growth to ∞

Practical Wyner-Ziv Coding



Practical Wyner-Ziv Coding

Nested Lattice Quantization [Zamir *et al* 2002]

- Coarse quantizer $Q_c(\cdot)$ is nested in a fine quantizer $Q_f(\cdot)$- all bin centroids of the coarse quantizer coincide with a (regular) subset of centroids of the fine quantizer
- Attains the WZ RD bound for Gaussian sources as quantizers' dimension growth to ∞

Finite Dimensional Lattice Quantizers [Xiong 2006]

- Correlation between quantizer's output and the SI, Y, remains
- Use Slepian Wolf coding for further compression
- At high rates ($D < \sigma^2_{X|Y}$) nested scalar quantizer is 1.53 dB away from the WZ RD bound

Practical Wyner-Ziv Coding (cont.)

Example: Nested Scalar Quantization

Setting Nesting Ratio $\Delta_c/\Delta_f = 4$

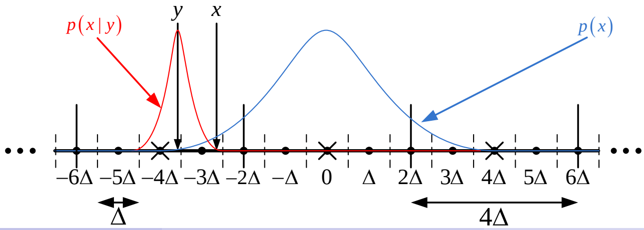
Quantization $x_{Q_f} = Q_f(x) = -3\Delta$
Send to decoder: $s = Q_c(x) - x_{Q_f} = -\Delta$

De-Quantization Recover $x_{Q_{fine}}$ using side information y:

$$x_{Q_f} = Q_c(y) - s = -4\Delta + \Delta = 3\Delta$$

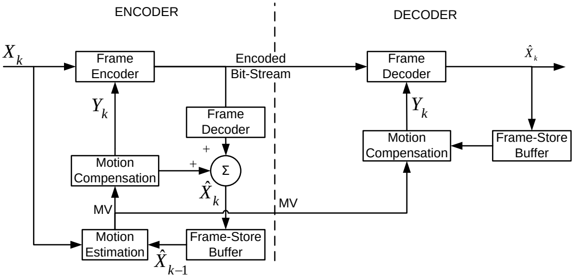
Reconstruct x as a centroid of the bin indexed by x_{Q_f}:

$$\hat{x} = E_{p(x|y)}\left[x|x_{Q_f}, y\right]$$

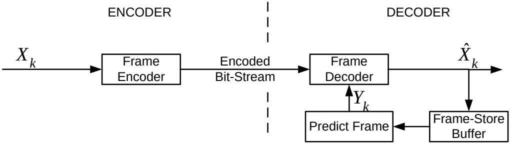


Video Coding

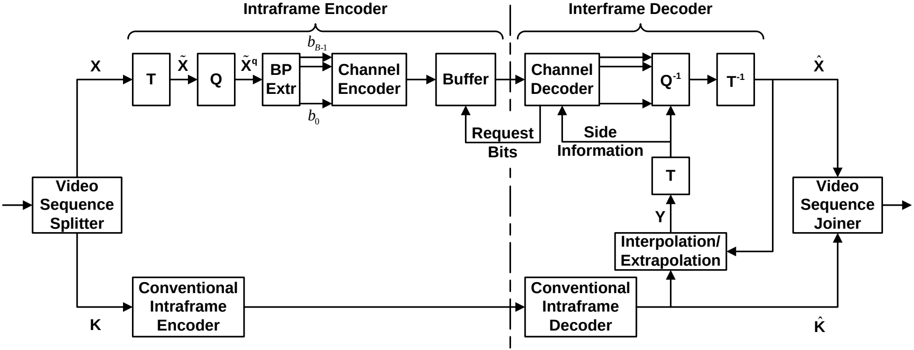
Standard System



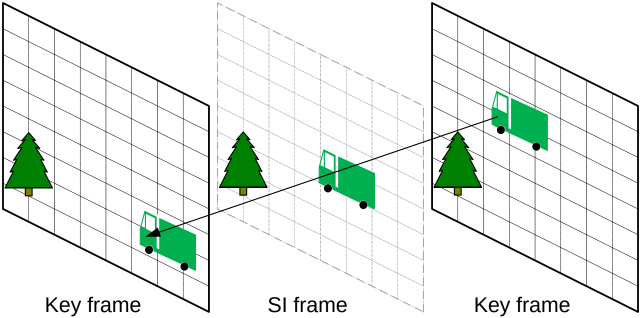
Distributed Coding System



Distributed Video Coding - Detailed Diagram



Side Information Generation



SI Generation Methods

- Motion estimation between decoded Key and/or WZ frames
  - Block matching
  - Optical flow
  - Parametric methods - affine transform
- Motion compensated extrapolation/interpolation

Joint Distribution Modeling

- Joint distribution is needed for SW decoding and de-quantization
- In DVC the joint distribution of the Source and Side Information is not known

Off-line modeling

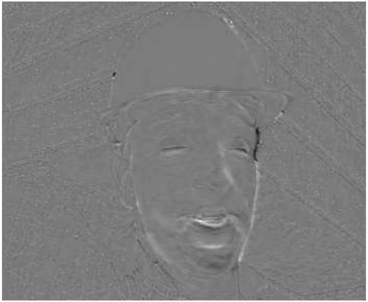
- Assume $Z = Y - X$, $Z \sim Laplace(\mu = 0, \sigma)$
- Learn the typical σ based on a set of test sequences
- Fails to capture temporal variation

On-line modeling

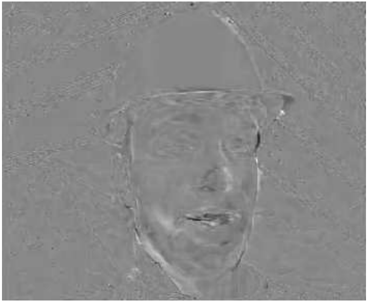
- X available only at the encoder, Y available only at the decoder
- Approximate the noise image Z using decoded Key or WZ frames
  $\tilde{Z}(x, y) = \frac{1}{2}(X_b(x + dx_b, y + dy_p) - X_f(x + dx_f, y + dy_f))$
- Approximation accuracy depends on ME algorithm, GOP size, reference frames quality

Joint Distribution Modeling (cont.)

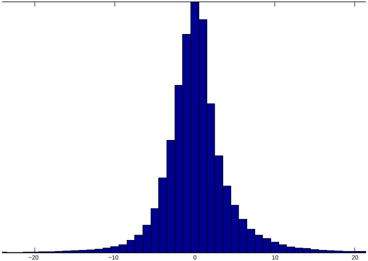
Z



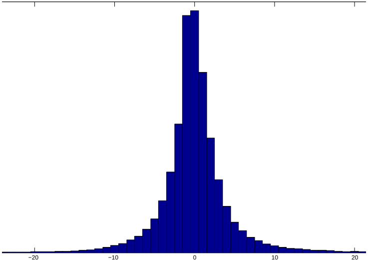
$\tilde{Z}$



Z



$\tilde{Z}$



Virtual Channel Modeling

DCT Virtual Channel - Assumptions

- Noise is additive ($Z = Y - X$), white and independent of Y
- Noise distribution in each DCT coefficient band is characterized by different set of parameters

Generalized Gamma Distribution (GΓD)

$$f_Z(z; a, v, p) = \frac{pa^{-pv}}{2\Gamma(v)} |z|^{pv-1} e^{-(|z|/a)^p}, \quad z \in \mathbb{R}, \quad a, v, p > 0$$

	Parameters	# Parameters
Generalized Gaussian	vp=1	2
Gamma	p=1, v=0.5	1
Laplace	p=1, v=1	1
Gaussian	p=2, v=0.5	1

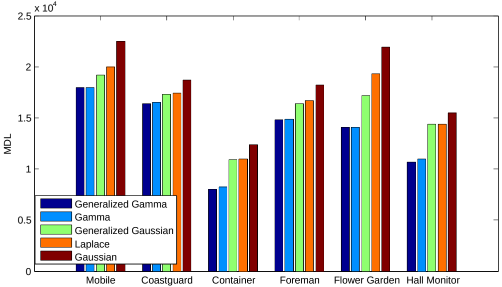
Comparing the Models

Goodness of Fit Metrics

$$AIC(\hat{\theta}_{ML}) = -2\log\left[f\left(\underline{x};\hat{\theta}_{ML}\right)\right] + 2k$$

$$MDL(\hat{\theta}_{ML}) = -\log\left[f\left(\underline{x};\hat{\theta}_{ML}\right)\right] + \frac{k}{2}\log(n)$$

where $\hat{\theta}_{ML}$ is the Maximum Likelihood estimate of model parameters and k is the number of parameters



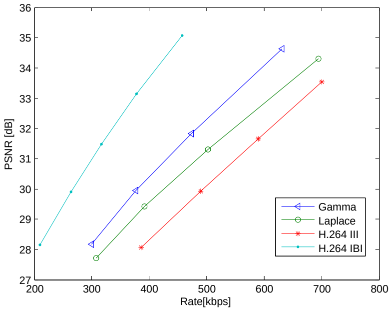
Comparing the Models (cont.)

- We integrate the Gamma distribution into DVC system
  - MDL and AIC metrics for Gamma are very close to GΓD
  - The double sided Gamma distribution has a single parameter:

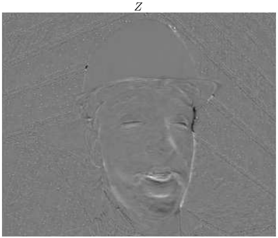
$$f(z) = \frac{1}{2\sqrt{\pi a |z|}} e^{\frac{-|z|}{a}}, \qquad a > 0$$

  - Closed form ML estimator for a:

$$\hat{a}_{ML} = \frac{2}{n} \sum_{i=1}^{n} |x_i|$$



Spatially Adaptive Modeling of the Virtual Channel



- Stationary models fail to capture the spatially varying joint distribution
- The estimation should be performed in the pixel domain
- Spatially adjacent virtual channel pixels are correlated
- The family of multivariate double sided Gamma distribution is not closed with respect to linear transformation
- Use MultiVariate Laplace (MVL) distributions to model $n_b \times n_b$ virtual channel blocks:
  $f(z) = 2(2\pi)^{-d/2} |\Sigma|^{-1/2} \left(z'\Sigma^{-1}z/2\right)^{v/2} K_v\left(\sqrt{2z'\Sigma^{-1}z}\right),$
  $z \in \mathbb{R}^d, d = n_b^2, v = (2-d)/2$ and $K_v(\cdot)$ - modified Bessel function

MVL Distribution

MVL Characteristic Function

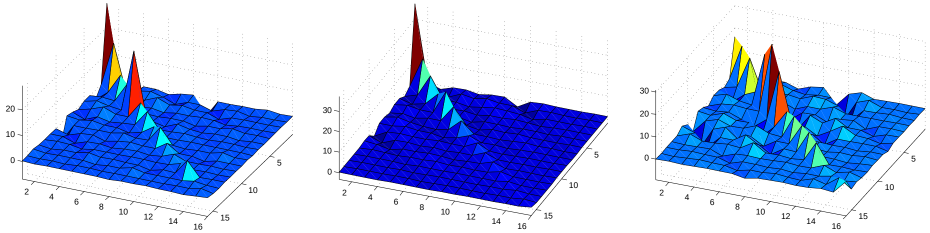
$$\Phi(t) = \frac{1}{1+\frac{1}{2}t'\Sigma t}, \quad t \in \mathbb{R}^d$$

MVL Properties

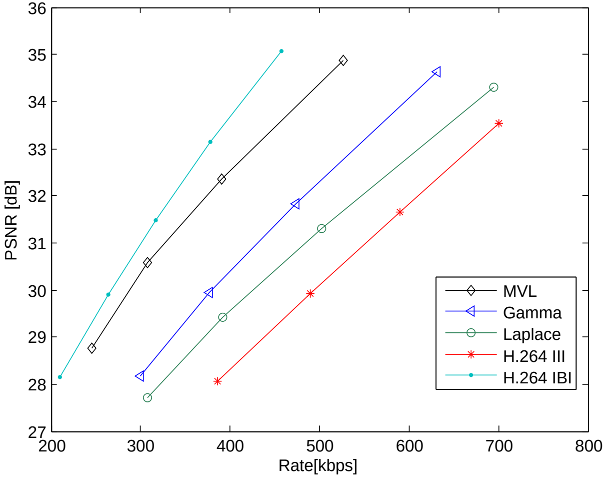
- If $Z = z_1, \ldots, z_d$ is MVL distribution then 1D marginal distributions are Laplace distributions $z_i \sim Laplace(\mu = 0, \sigma_{ii})$
- If $Z \sim MVL(\Sigma)$ then $W = AZ \sim MVL(A'\Sigma A)$:
  $\Phi_W(t) = E[e^{iWt}] = E[e^{i(AZ)'t}] = E[e^{iZ'A't}] = \Phi_Z(A't)$

MVL Distribution Parameters Estimation

- Pixel domain block-wise autocovariance matrix estimation using samples from $n_w \times n_w$ window:
  $Cov(u,v) = \frac{1}{(n_w-u)(n_w-v)} \sum_{|k-i|=u,|l-j|=v} x_w(k,l)x_w(i,j)$
- Transform Σ for each block to DCT domain
- Typically, cross-band elements of Σ are relatively small thus 1D marginals can be used without significant loss
- Results in univariate Laplace distribution with different (spatially dependent) parameter for each DCT coefficient



Spatially Adaptive Modeling - Simulation Results



Encoder-Side Rate Control

Feedback Channel

- Incurs delay $\rightarrow$ Unsuitable for real-time applications
- Not available in some apps. (e.g. storage)

Feedback Suppression

- [Morbee 08], [Brites 07] Use average of adjacent frames as encoder-side low-cost SI
- Rate estimation is based on the quantized data, R=H(Q(X)|Y)
- Quantization and rate evaluation have to be repeated until $R < R_max$
- Proposed approach: Encoder–side rate control based on a rate distortion model

Rate Distortion Model

High Rate Quantization [Rebollo *et al* 2006]

- If quantization step Δ is small enough - $p(x|y)$ is $\sim$uniform within quantization bin then:

$$D(\Delta) = \frac{\Delta^2}{12} \text{ and } R(\Delta) = h(X|Y) - \log \Delta$$

- If $Z \sim Laplace(\mu, \sigma_{X|Y})$ then:

$$R(D) = \frac{1}{2} \log \frac{e^2 \sigma^2_{X|Y}}{6D} \text{ or } D(R) = \frac{1}{6} e^2 \sigma^2_{X|Y} 2^{-2R}$$

General Case - Laplacian Sources [Sheinin 06]

- $X = Y + Z$, $Y \sim Laplace(\mu_Y, \sigma^2_Y)$ and $Z \sim Laplace(\mu_{X|Y}, \sigma^2_{X|Y})$ i.i.d., Z independent of Y
- Infinite Uniform Scalar Quantizer - $IUSQ(\Delta)$
- The RD model is given in integral form expressions

Infinite Scalar Quantizer

- Define bin probability p_i and bin centroid $\hat{x}_i$ as follows:

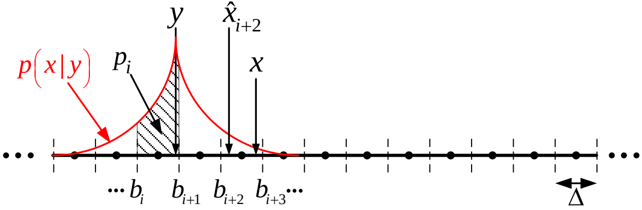
$$p_i(\Delta) = \int_{b_i}^{b_{i+1}} p(x|y)dx \qquad \hat{x}_i = \frac{1}{p_i} \int_{b_i}^{b_{i+1}} xp(x|y)dx$$

- [Sheinin 2006] In IUSQ(Δ) for given y:

$$r(\Delta|y) = -\sum_{i=-\infty}^{\infty} p_i \log p_i$$
$$d(\Delta|y) = \sum_{i=-\infty}^{\infty} \int_{b_i}^{b_{i+1}} p(x|y)(x - \hat{x}_i)^2 dx$$

- The total rate and distortion can be obtained by averaging over Y:

$$R(\Delta) = \int p(y)r(\Delta|y)dy$$
$$D(\Delta) = \int p(y)d(\Delta|y)dy$$

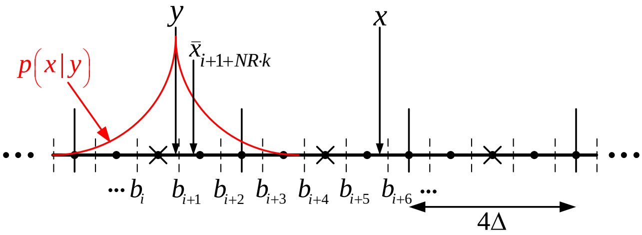


SWC-NSQ for Laplace Virtual Channel

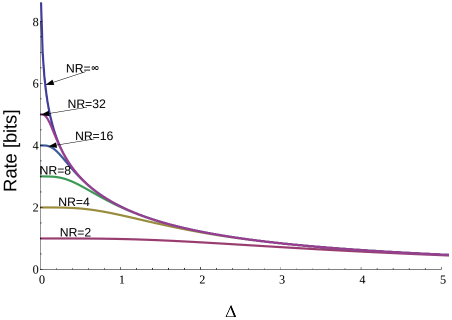
- We generalize the nesting approach for, given y and nesting ratio NR:

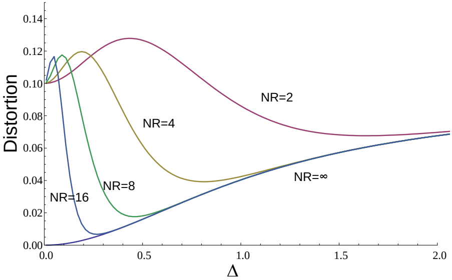
$$r(\Delta|y) = -\sum_{j=0}^{NR-1}\sum_{k=-\infty}^{\infty} p_{j+NR\cdot k}\log\left(\sum_{k=-\infty}^{\infty} p_{j+NR\cdot k}\right)$$
$$d(\Delta|y) = \sum_{i=-\infty}^{\infty}\int_{b_i}^{b_{i+1}} p(x|y)(x-\bar{x}_i)^2 dx$$

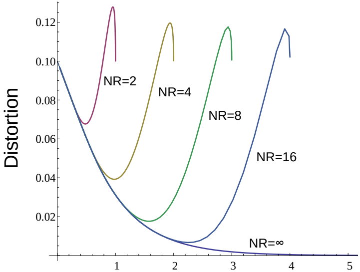
  where $\bar{x}_i$ is the centroid of the bin indexed by $Q_c(y) - Q_f(x)$, if $Q_c(x) \neq Q_c(y)$ then x will be recovered in a wrong bin



SWC-NSQ for Laplace Virtual Channel (cont.)



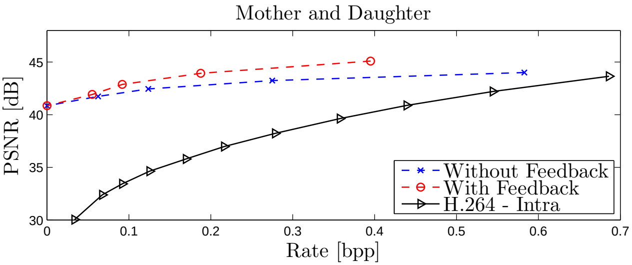




Applications of RD Model in DVC

DVC Encoder–Side Rate Control

- Feedback suppression - evaluate RD for the whole frame
- Use frame difference to estimate 'noise' statistics (applicable only to low motion sequences)

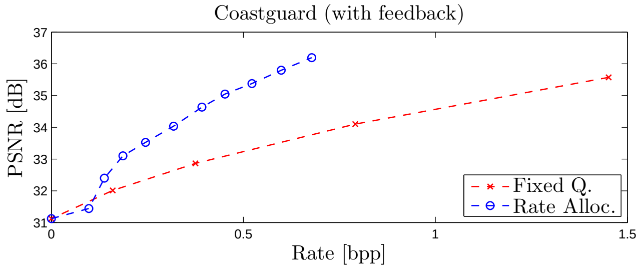


Rate Allocation

- Split WZ frames into disjoint slices, evaluate RD for each slice
- Applicable to systems with and without feedback

$$\min_{(q_0,\ldots,q_{S-1})} \sum_{s=0}^{S-1} D_s, \qquad \text{s.t.} \sum_{s=0}^{S-1} R_s(D_s) \le R_{max}$$

$$q_i \in \{\Delta_0, \ldots, \Delta_{m-1}\}$$



Aerial Video Coding

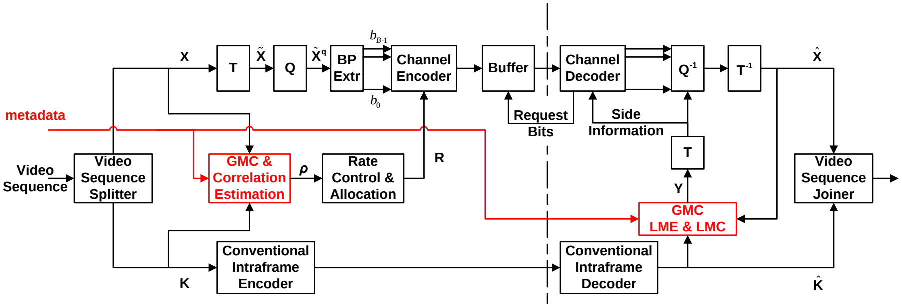
Civil and Military Applications

- Mapping
- Resources monitoring
- Intelligence gathering
- Remote real-time operation of UAV

Aerial Video Codec Design

- Design should comply with space, weight, and power constraints dictated by the desire for vehicles with longer endurance
- Aerial Video is accompanied by metadata describing cameras pose and platform's motion
- Metadata enables to compensate for the Global Motion e.g., using affine transform

Aerial DVC System



GMC – Global Motion Compensation

LMC – Local Motion Compensation

LME – Local Motion Estimation

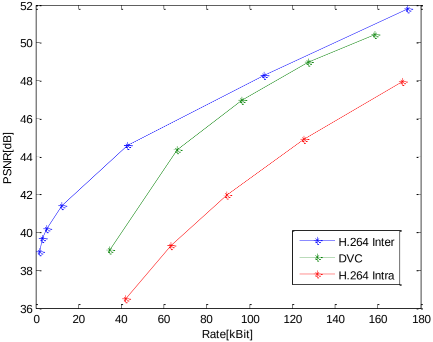
BP Extr. – Bit-Planes Extraction

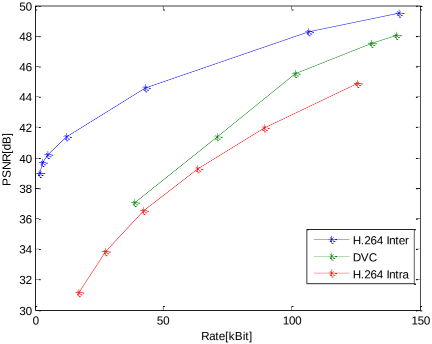
Global-Local Motion Compensated Interpolation

Global-Local Segmentation

- Segmentation is based on a difference between two Global Motion Compensated reference frames GMC(x_b) and GMC(x_f)
- Segmentation is performed in two steps, first at block level and then refined at pixel level
- Morphological operators are used to remove isolated block/pixels
- Temporal tracking after foreground segments is used to identify moving background misclassified as foreground

Aerial DVC - Simulation Results





Summary

- Stationary and spatially adaptive models of the virtual channel were considered
- It was shown that the Gamma model outperforms the widely adopted Laplace model
- A method for spatially adaptive modeling was presented. Pixel domain estimation of model parameters and their transformation to the DCT domain were shown. Performance gains relatively to the stationary case were demonstrated.
- Methods for encoder-side model-based rate control and allocation were developed.

Future Work

- Using motion field reliability information will result in a more accurate modeling of the virtual channel
- Generalizing the spatially adaptive modeling into temporally-spatially adaptive modeling will provide a more reliable estimates of model parameters