STREAMING REAL-TIME AUDIO AND VIDEO DATA ON THE INTERNET WITH ERROR CONCEALMENT AND RECONSTRUCTION

Benjamin W. Wah

Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
http://manip.crhc.uiuc.edu

Contributors: D. Lin and X. Su

June 19, 2000

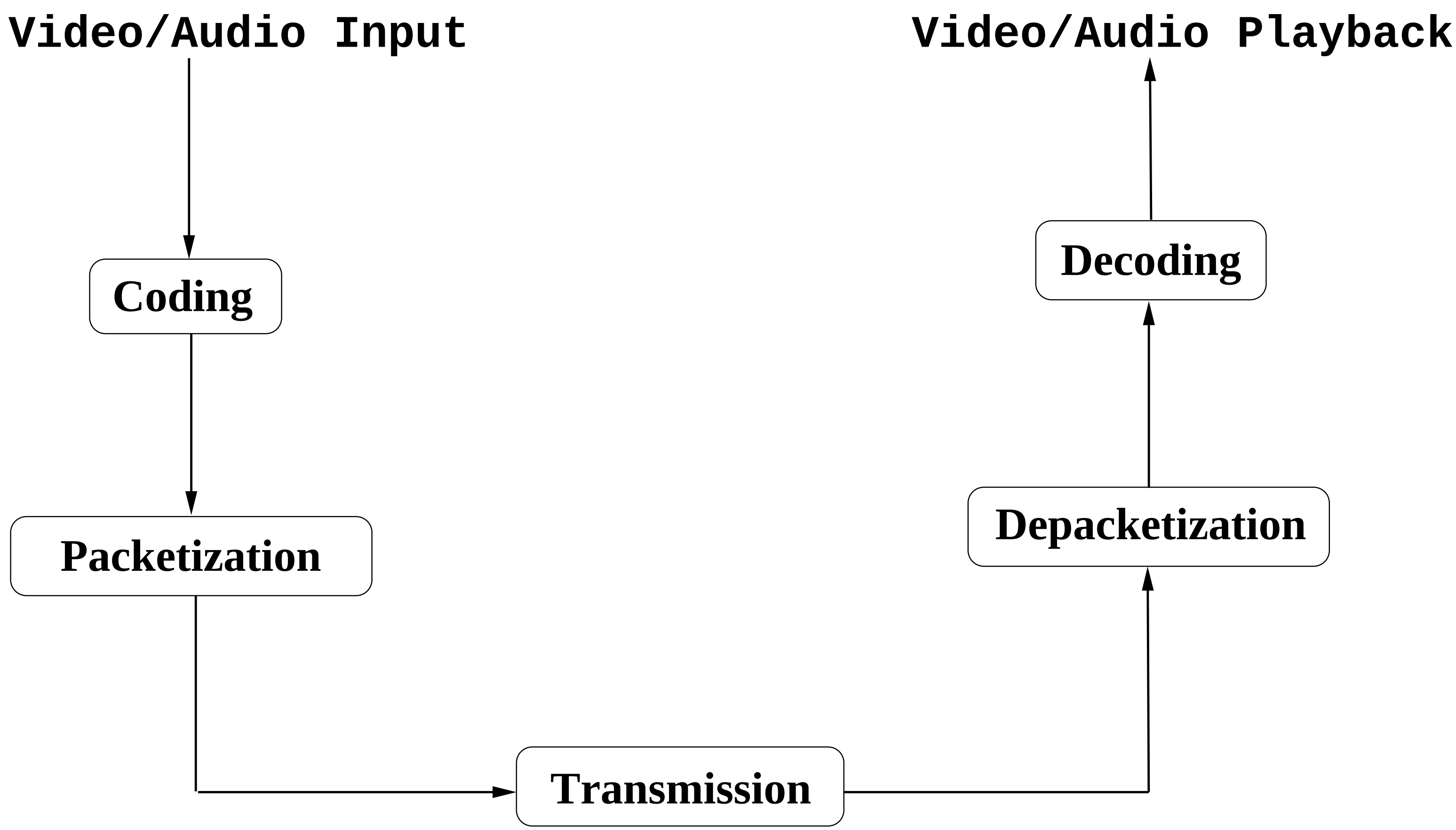
Outline

- Motivations: dealing with information loss
  - IP-based multimedia applications
  - Internet loss behavior
  - Performance of commercial players under loss
- Error concealment problem
- Sender-receiver IP-based multi-description coding
  - *Coder-independent*: transformation-based
  - *Coder-dependent*: optimized reconstruction-based DCT and CELP
- Experimental demonstration

IP-Based Multimedia Applications

- Voice over IP
- Internet telephony
- Video and audio conferencing
- Real-time video and audio on demand
- Distance learning

Real-Time Coding and Decoding



Assumptions

- Real-time:
  - Interactive communications
  - End-to-end delay less than 300 msec (soft constraint)
- Transmissions over the Internet and wireless channels:
  - Limited and variable bandwidth
  - High and bursty packet losses
  - Delay jitters
- Built on top of UDP/IP, no QoS support from underlying network

Performance Measures

- Subjective quality:
  - Video: no freeze on losses, no flickering, no visible lost blocks
  - Audio (for waveform codecs): Mean Opinion Score (MOS), 5 classes (excellent, good, fair, poor, bad)
- Objective quality:

$$PSNR = 10\log_{10}\frac{255^2}{\sum_i (s_i - \hat{s}_i)^2} \quad \text{(video)}$$

$$SNR = 10\log_{10}\frac{\sum_i s_i^2}{\sum_i (s_i - \hat{s}_i)^2} \quad \text{(audio using waveform codecs)}$$

Itakura-Saito likelihood ratio – (LPC-based low-bit rate speech codecs)

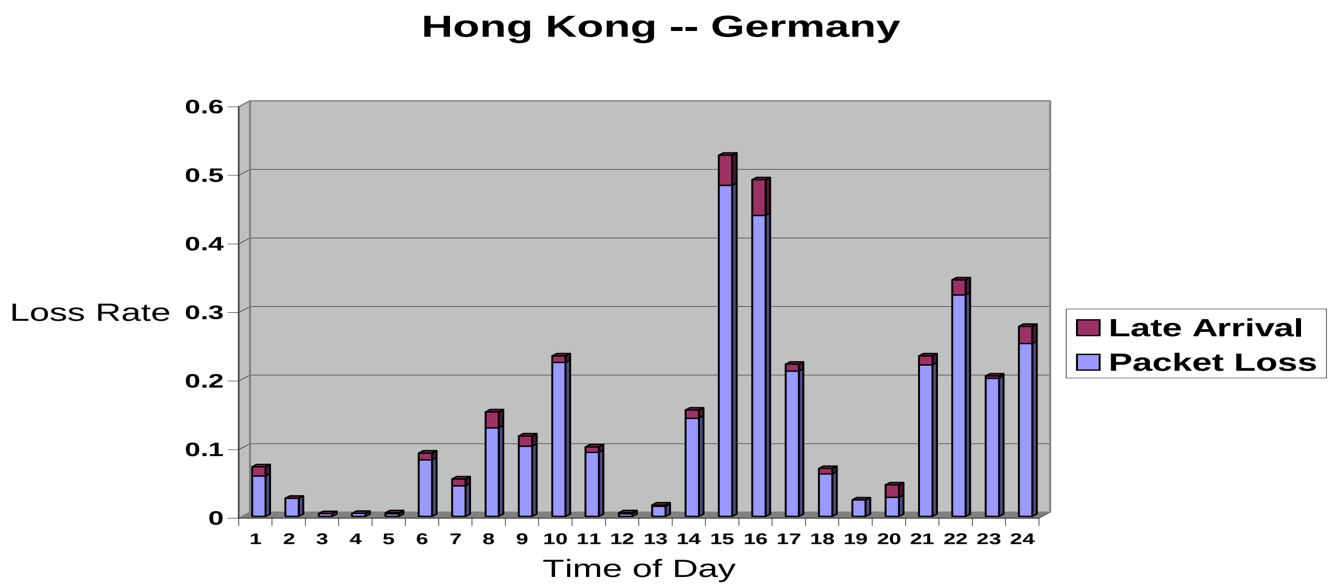
Cepstral distance – (LPC-based low-bit rate speech codecs)

Three Major Challenges in Real-Time Internet Transmissions

- Limited bandwidth:
  - CIF (352x288) H.263 video at 30 f/s needs 80 Kbps-1.5 Mbps
  - Internet sustained bandwidth ranges from 30 - 800 Kbps
- Strict timing constraints:
  - Real-time streaming requires playback at prescribed time instances
  - Jitters make packets arrive too late for playback
- High playback quality:
  - High-quality real-time streaming needs robust delivery mechanisms
  - Poor quality due to information loss and limited bandwidth

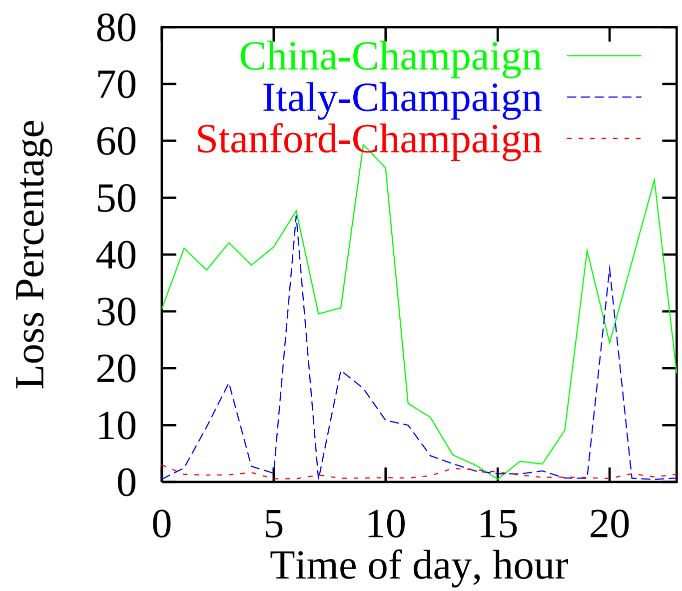
Sources of Information Loss

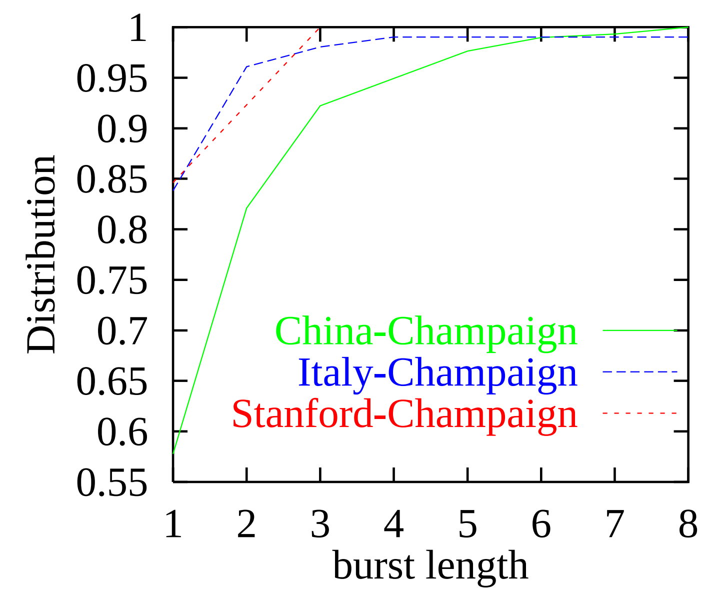
- Compression loss: lossy quantization
- Bitstream loss
  - Due to network – over 50% (network and jitter losses)
  - Due to scaling – dropped by sender in limited bandwidth (video)



- Propagation loss due to dependencies (video)

Statistics on Burst Lengths

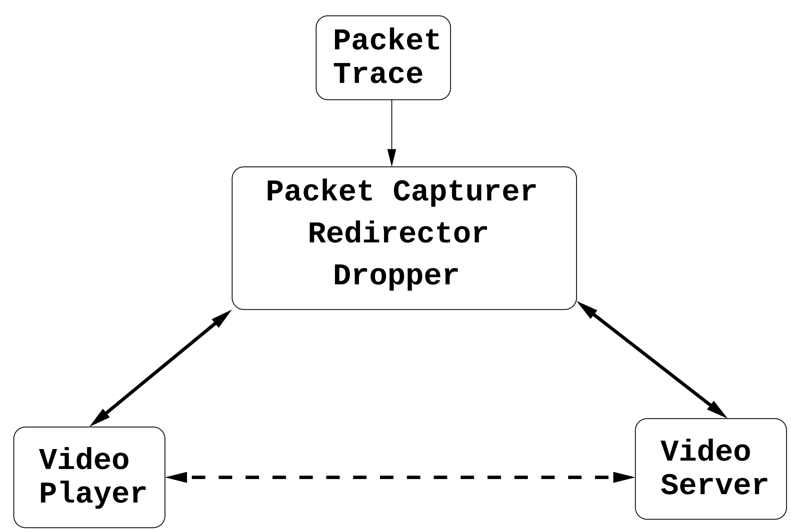




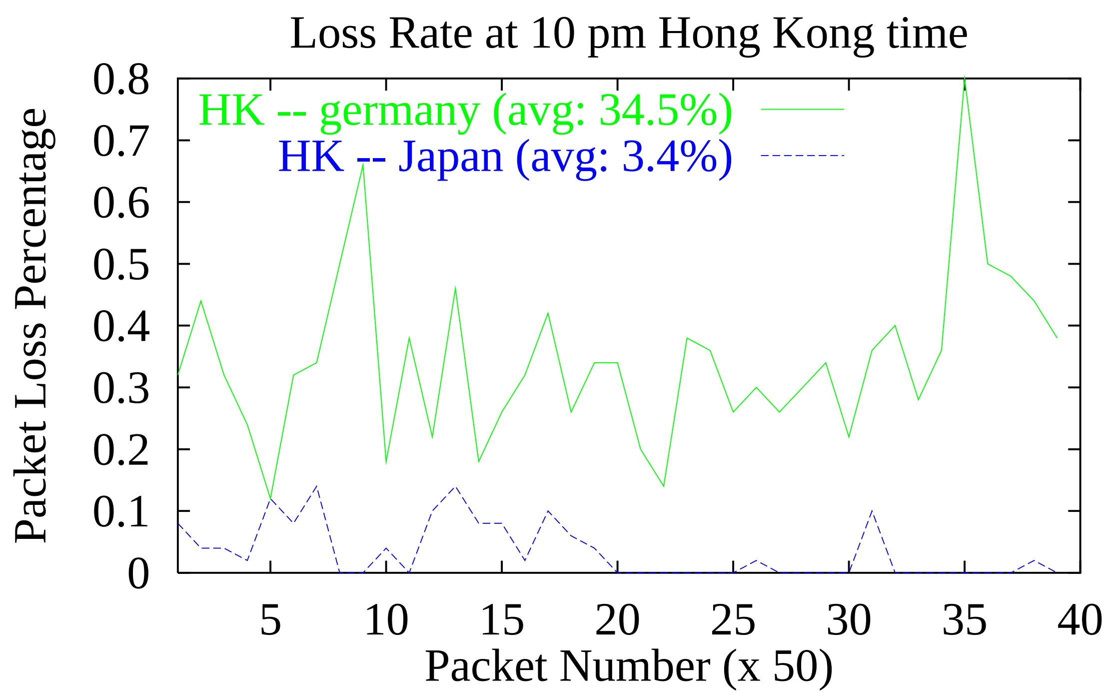
- Probability of burst length > 1 is 2% for connections within US
- Probability of burst length > 3 is 5% for transcontinental connections

Evaluation of Commercial Video Streaming Systems

- Leading Internet video streaming players – RealPlayer, MediaPlayer
  - Proprietary codecs
  - Unknown error concealment schemes
  - Initial buffering delay of several seconds
- Trace-based evaluations to compare quality under same traffic conditions and bit rate



Playback Quality of Commercial Players under Loss



- Test sequence: boxing at 320×224, 5 fps and 80 Kbps
- Observations: like a slide show
  - Video freezes on packet losses, 5-15 seconds to recover
  - Visible corrupted blocks

Error-Concealment Schemes: Previous Work

- Redundant transmissions
  - Forward error correction (FEC)
  - Joint source-channel coding (JSC)
    - Achieve trade-off between data and protection, assuming bit error probability p
  - Reversible variable length coding (RVLC): MPEG-IV
    - Bitstream can be decoded backwards in case of loss
    - Unknown channel model
  - Sending extracted information in adjacent packets
  - Exploiting time constraints of applications to do retransmissions
    - Extra bandwidth

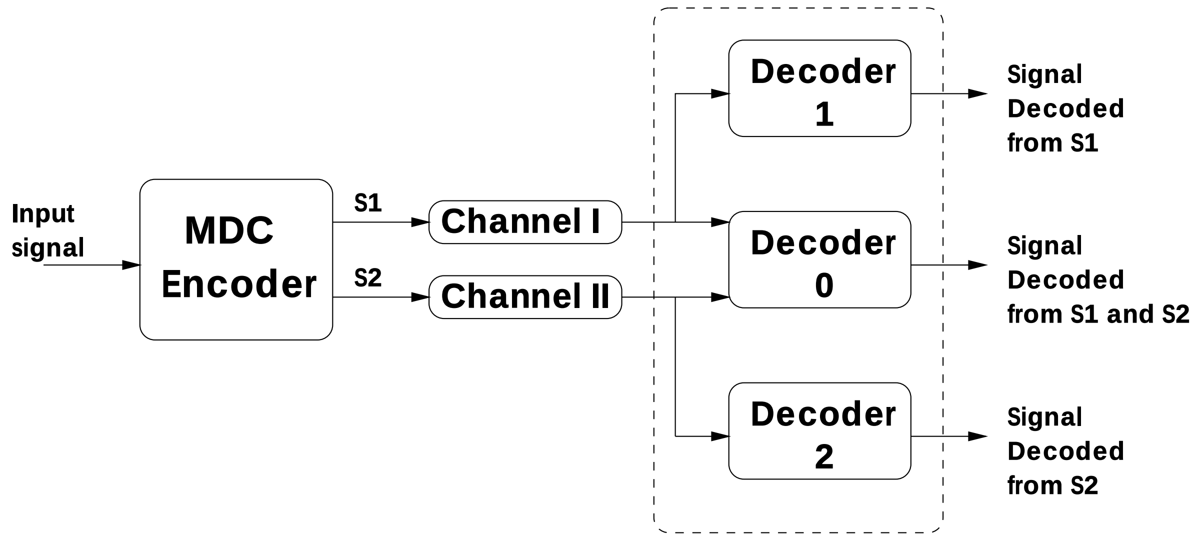
Error-Concealment Schemes: Previous Work (cont'd)

- Nonredundant transmissions
  - Replaying last packet received
  - Padding lost packets by silence or white noise
  - Waveform substitution
    - Failure under long bursty losses
  - Exploiting source-data properties (edge orientations & geometric str.)
    - Computational expensive
  - Motion JPEG (MJPEG)
    - No motion compensation to avoid error propagation
  - Independent Segment Decoding (ISD): MPEG-IV
    - Motion estimation not exceeding segment boundary

Robust Coding Algorithms: Previous Work

- Layered coding schemes: base layer + enhancement layers,
  - Different priority of frames or parameters of frames:
    - Voiced, unvoiced, silence, etc.
    - I-frames and P-frames
  - MSBs and LSB interleaving
  - Issues
    - No prioritized delivery in the Internet
    - Difficulty in reconstruction when base layer is lost
- Multi-description coding (MDC) schemes
  - Partitioning data into equally important streams (descriptions)
  - Uncorrelated losses to different descriptions

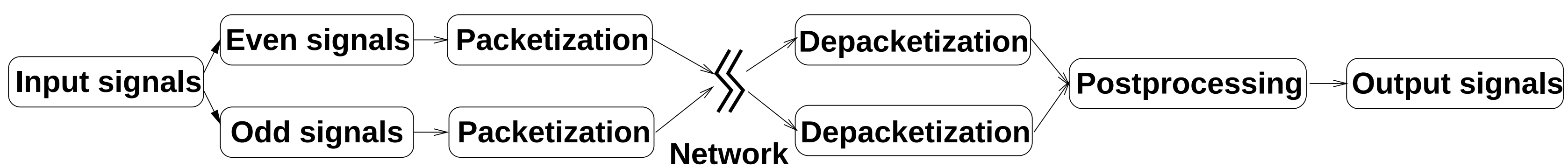
Multi-Description Coding



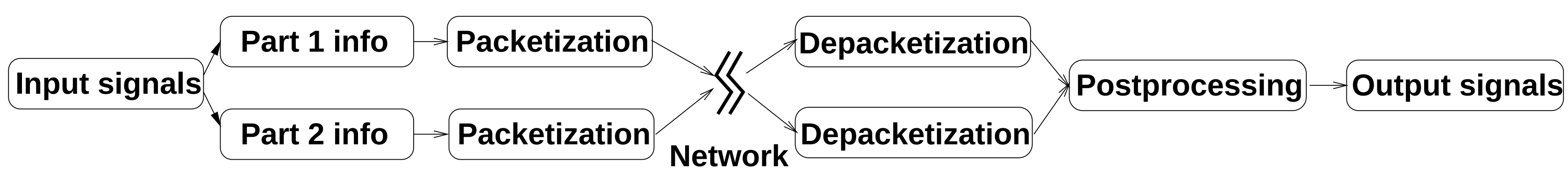
- Mediocre coding efficiency with only one description
- Higher bit rate
- Encoder and decoder should be designed together to achieve high performance

Interleaving: A Simple MDC

- Sample-based interleaving: degree 2 to 4 to cope with bursty losses



- Interleaving of meta-information



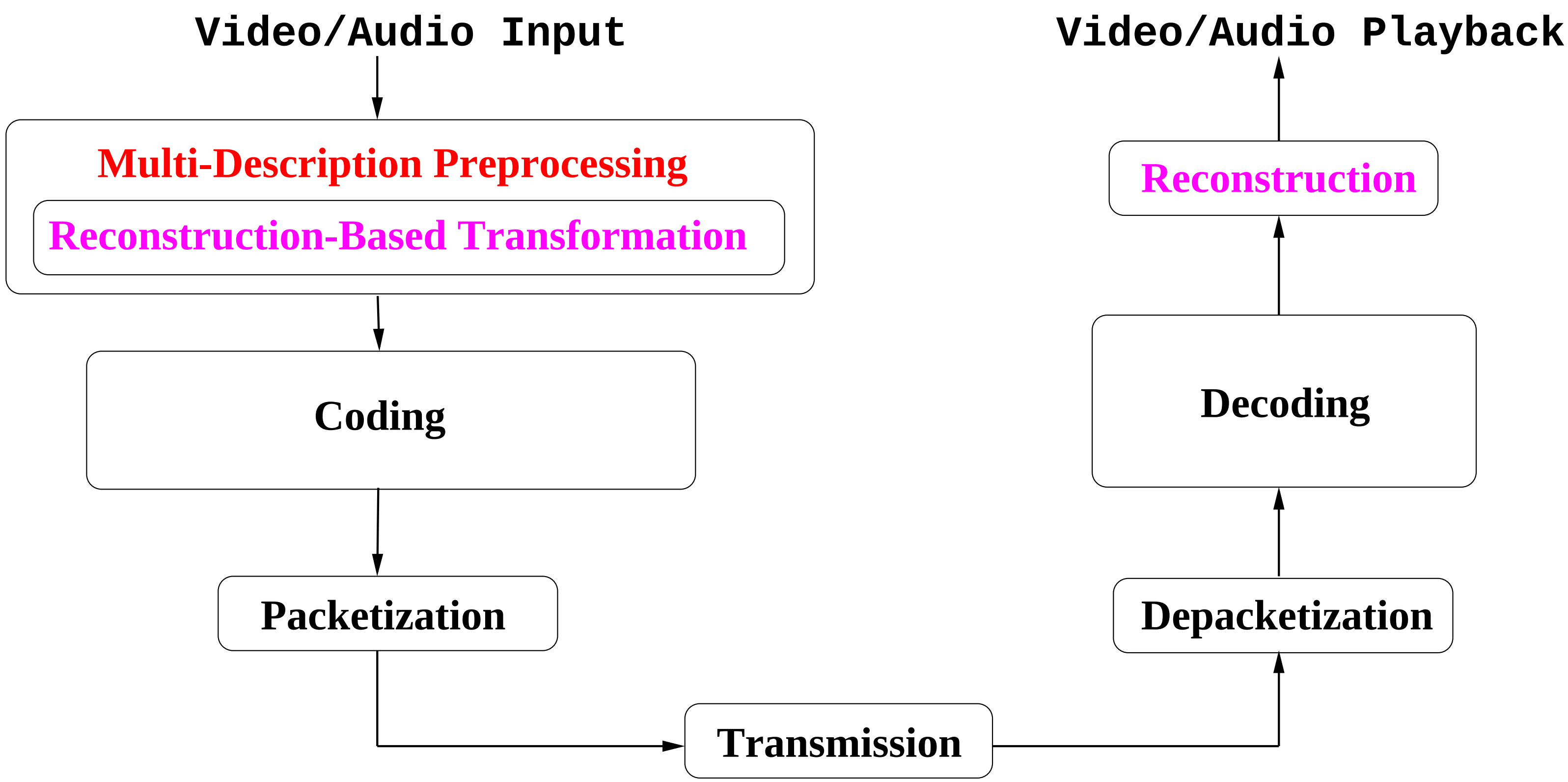
Error Concealment Problem

- Design, analyze and evaluate robust end-to-end multi-description coding and error-concealment schemes in order to allow robust real-time audio and video streaming over unreliable IP networks
- Focus on coding schemes in this talk

Outline

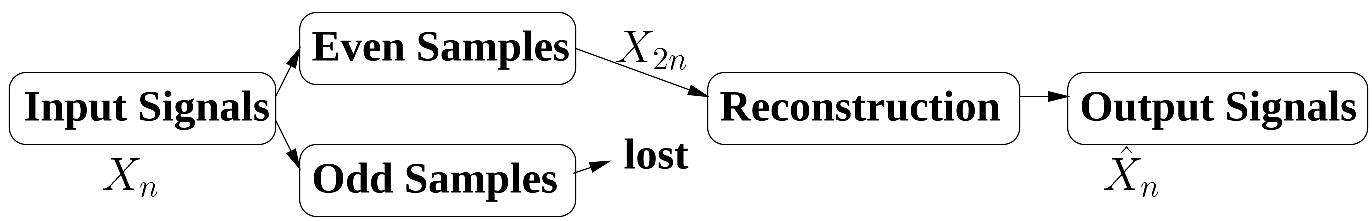
- Motivations: dealing with information loss
  - IP-based multimedia applications
  - Internet loss behavior
  - How commercial players perform under loss
- Error concealment problem
- Sender-receiver IP-based multi-description coding
  - *Coder-independent*: transformation-based
  - *Coder-dependent*: optimized reconstruction-based DCT and CELP
- Experimental demonstration

Coder-Independent Error Concealment and Reconstruction



Multi-Description Preprocessing and Reconstruction

- Multi-description preprocessing using 2-way interleaving:



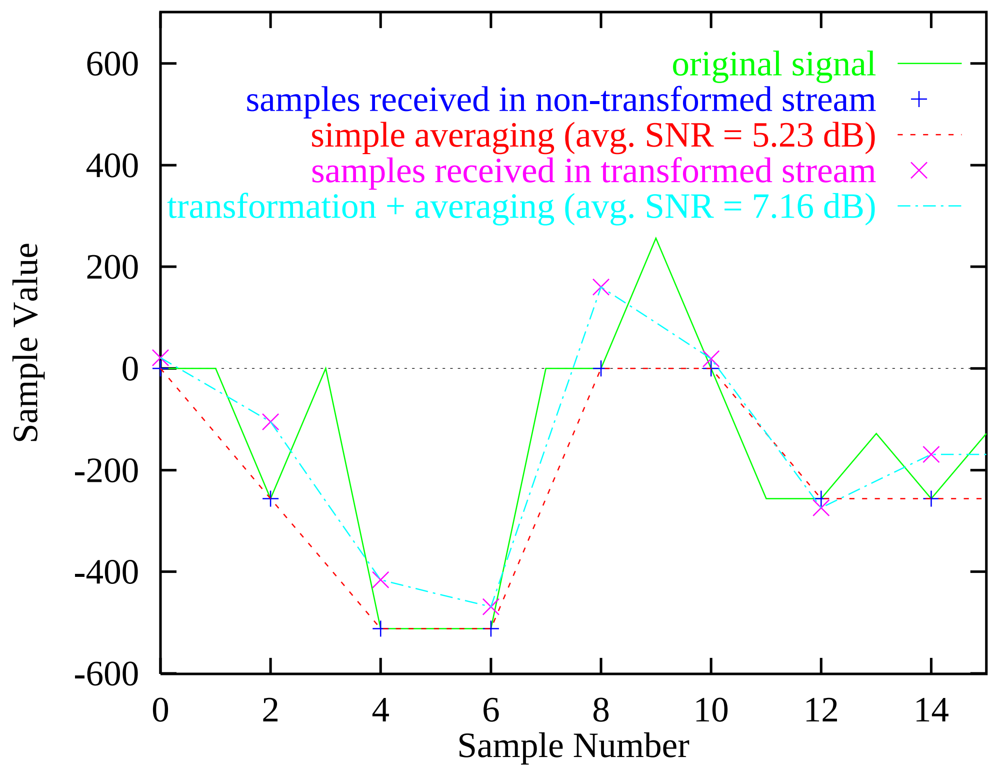
- Simple reconstruction of missing samples by linear interpolation

$$\hat{x}_i = \begin{cases} x_i & i \text{ even} \\ \frac{x_{i-1}+x_{i+1}}{2} & i \text{ odd and } i \neq 2N-1 \\ \frac{x_{2N-2}}{2} & i = 2N-1 \end{cases}$$

- Issue: does not work well when adjacent samples change rapidly

Reconstruction-Based Transformation

- Sender transforms input signals according to reconstruction method used at the receiver to enable better reconstruction quality if some information is lost during transmission (form of anti-aliasing filters)



Optimization of Transformation T

$$\vec{x} = x_0, \cdots, x_{2N-1} \xrightarrow[\text{Original Stream}]{} \boxed{\begin{array}{c}\text{Linear Transformation}\\ \vec{y} = T \times \vec{x}\end{array}} \xrightarrow[\text{Transformed Stream}]{} \vec{y} = y_0, \cdots, y_{2N-1}$$

- Suppose only $\vec{y}_{even} = y_0, y_2, \cdots, y_{2N-2}$ received;
  Reconstructed stream is $y_0, \frac{y_0+y_2}{2}, y_2, \cdots, y_{2N-2}, \frac{y_{2N-2}}{2}$
- Reconstruction Error

$$\mathcal{E}_r = \sum_{n=0}^{N-1} (x_{2n} - y_{2n})^2 + \sum_{n=0}^{N-2} \left(x_{2n+1} - \frac{y_{2n} + y_{2n+2}}{2} \right)^2 + \left(x_{2N-1} - \frac{y_{2N-2}}{2} \right)^2$$

- Minimize $\mathcal{E}_r$ to get the transformation T_{even} for $\vec{y}_{even}$:

$$\frac{\partial \mathcal{E}_r}{\partial y_i} = 0, \quad i = 0, 2, \cdots, 2N-2.$$

- Compute the transformation T_{odd} for $\vec{y}_{odd}$ similarly

Linear Transformation T

- Combine transformations T_{even} and T_{odd} to get signal-independent T

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & \frac{1}{5} & 0 & & & & & & & \\ 0 & 1 & 0 & \frac{1}{6} & & & & & & & \\ \frac{1}{6} & 0 & 1 & 0 & \frac{1}{6} & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & \frac{1}{6} & 0 & 1 & 0 & \frac{1}{6} \\ & & & & & & & \frac{1}{6} & 0 & 1 & 0 \\ & & & & & & & & \frac{1}{5} & 0 & 1 \end{pmatrix}^{-1} \times \begin{pmatrix} \frac{4}{5} & \frac{2}{5} & & & & & & & & \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & & & & & & & \\ 0 & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & \\ & & & & & & & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ & & & & & & & & \frac{2}{5} & \frac{4}{5} \end{pmatrix}$$

- Perfect reconstruction if both packets are received and no precision loss in transmission and coding

Performance of Coder-Independent Error Concealment

- Transformation size: $L =$ image width (video), $L = 64$ (audio)
- Compression – H.263 (video), G.723 (audio)
- Transformation consistently improves over the original scheme
  - One of the streams is lost under 2-way interleaving

Video Sequence	PSNR(dB), loss					
	without compression			with compression		
	org	tran	gain	org	tran	gain
Missa	39.44	41.11	**1.67**	36.20	36.38	**0.18**
football	36.05	37.29	**1.24**	29.55	29.88	**0.33**

Audio Files	SNR(dB), loss					
	without compression			with compression		
	org	tran	gain	org	tran	gain
Audio file (1)	11.54	12.66	**1.12**	7.74	8.94	**1.2**
Audio file (2)	12.72	14.19	**1.47**	12.54	13.77	**1.23**

Pixel-Based versus Block-Based Interleaving

- Block-based: Put adjacent blocks (basic coding unit) into different interleaved streams
- Pixel-based: Put adjacent pixels into different interleaved streams

Coding Efficiency

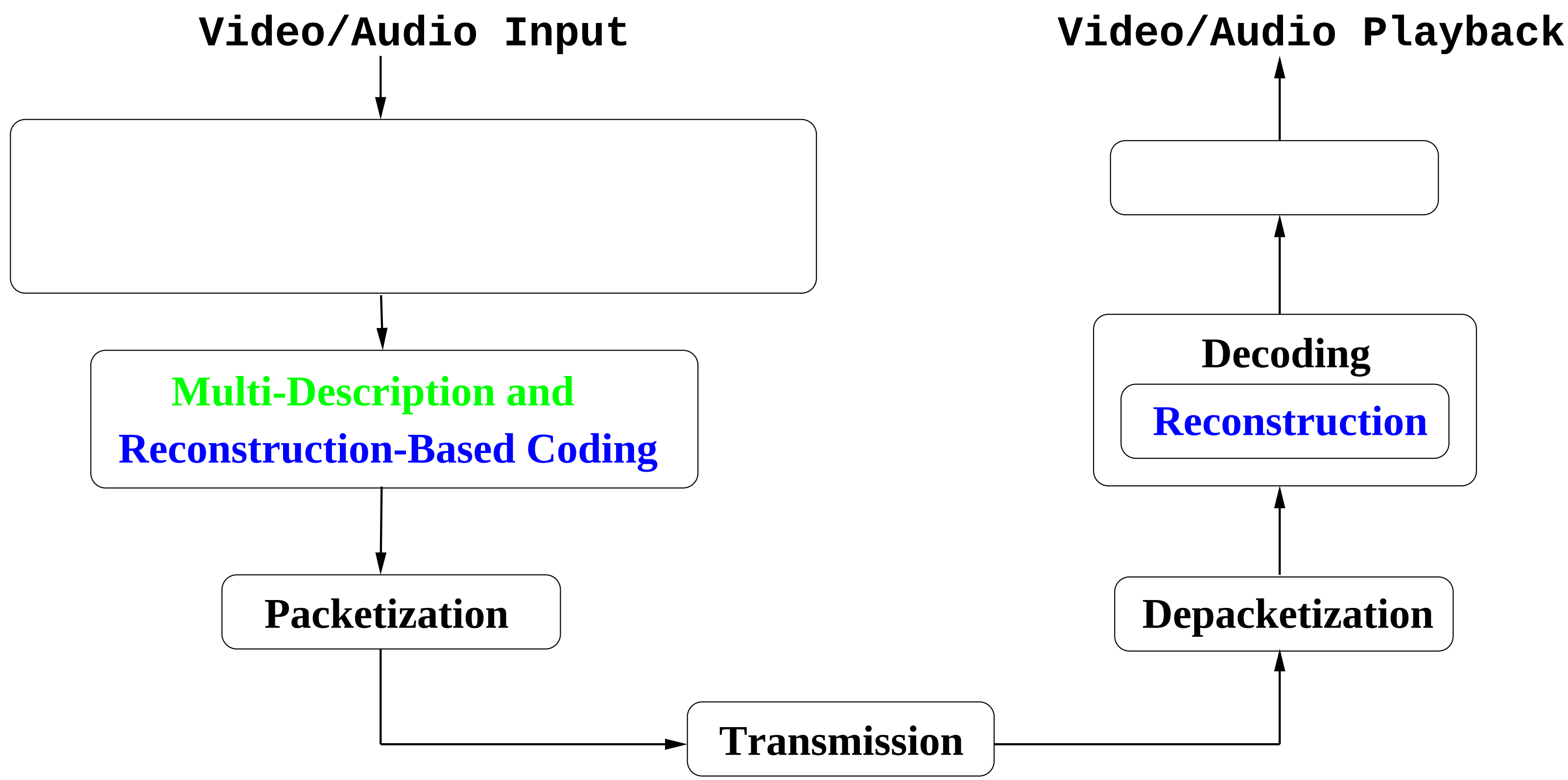
Video Sequence	Pixel-Based Interleaving		Block-Based Interleaving		Comparisons	
	BW (kbps)	PSNR	BW (kbps)	PSNR	BW (kbps)	PSNR
Missa	110 (30f/s)	36.74	190 (30f/s)	36.60	+80	-0.14
football	220 (5f/s)	30.15	270 (5f/s)	30.55	+50	+0.4

Reconstruction Quality

Video Sequence	Pixel-Based	Block-Based			Comparisons
	Int.	Int.	Directional	Max-smooth	
Missa	36.20	33.03	32.27	33.20	-3
football	29.55	26.06	24.90	26.13	-3.42

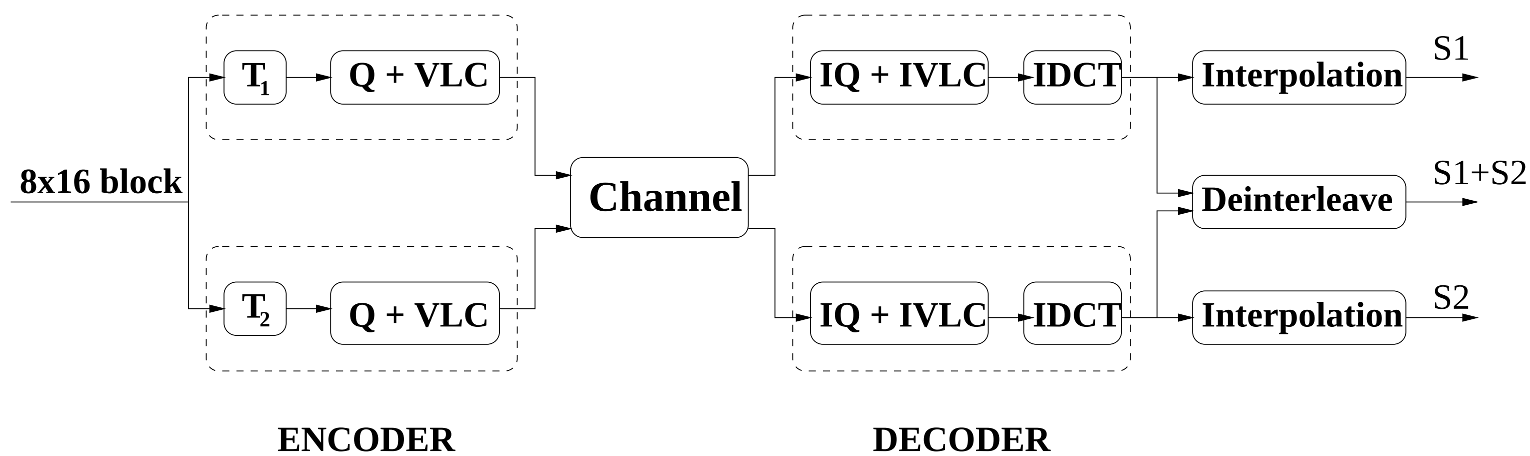
- Summary: Pixel-based interleaving/reconstruction is preferable.

Coder-Dependent Multi-Description Processing and Coding

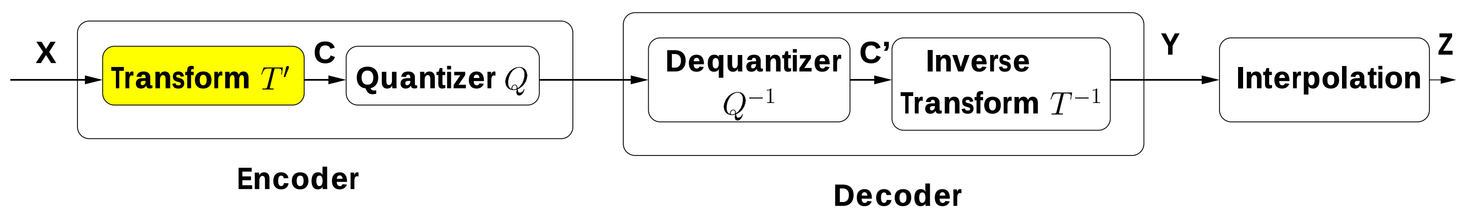


Coder-Dependent MDC Video Processing

- Use block as a basic unit
- T_1 and T_2 are of dimension 128×64



Derivation of Transformations T_1 and T_2



- Objective: assume T^{-1} is fixed, find quantized transform coefficients $C'_{n\times n}$ in order to minimize reconstruction error $\mathcal{E}_r$, after inverse DCT transform T^{-1} and average interpolation.
- Feature:
  - Standard-compliant decoder

Optimized Reconstruction-Based DCT

- Intra-coded blocks:

$$\min \mathcal{E}_r(C) = \left\| \sum_{i=1}^{n} \sum_{j=1}^{n} C(i,j) b_i e_j^T - X \right\|^2$$

$$\Longrightarrow \vec{C}_{64\times 1} = \vec{T'}_{64\times 128} \; \vec{X}_{128\times 1}$$

- Inter-coded blocks:

$$\min \mathcal{E}_r(C) = \left\| \sum_{i=1}^{n} \sum_{j=1}^{n} C(i,j) b_i e_j^T - (X - P) \right\|^2$$

$$\Longrightarrow \vec{C}_{64\times 1} = \vec{T'}_{64\times 128} \; (\vec{X} - \vec{P})_{128\times 1}$$

where P denotes its interpolated reference block

One Description (out of 2) Consistently Lost

Video	Odd received			Even received		
Sequence	DCT	ORB-DCT	Gain	DCT	ORB-DCT	Gain
	No quantization effects					
Missa	39.44	41.31	**1.87**	39.51	41.45	**1.94**
Football	36.05	37.48	**1.43**	36.01	37.47	**1.46**
	With quantization effects					
Missa	36.20	36.61	**0.41**	36.14	36.59	**0.45**
Football	29.43	29.82	**0.39**	29.40	29.83	**0.43**

Reconstruction Quality with 4 Descriptions

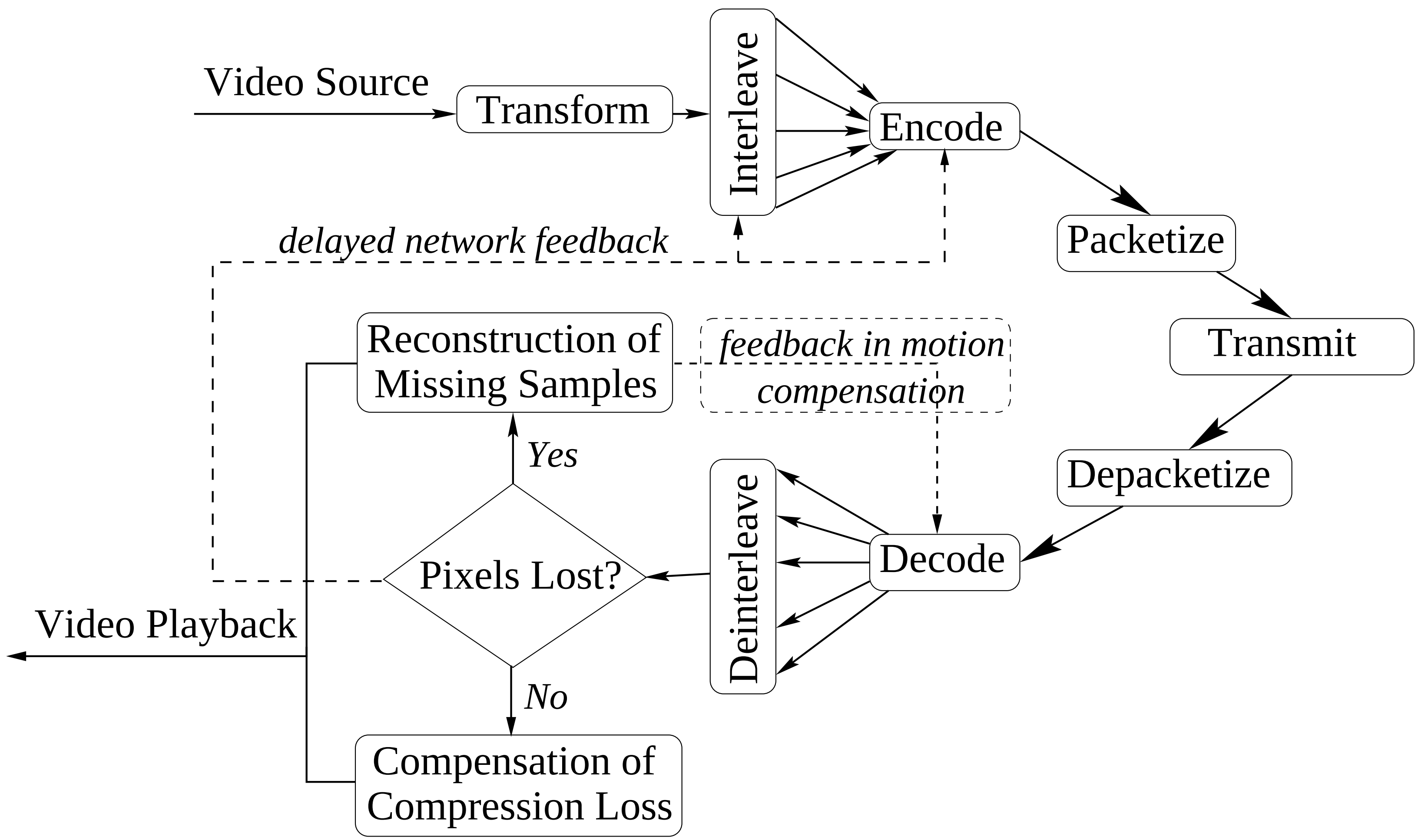
Video Sequence	Quant. Effects	Case I			Case II		
		DCT	ORBDCT	gain	DCT	ORBDCT	gain
Missa	No	35.84	37.27	**1.43**	39.35	39.88	**0.53**
football		34.92	35.97	**1.05**	35.72	36.15	**0.43**
Missa	Yes	33.58	33.93	**0.35**	34.01	34.23	**0.22**
Football		24.32	24.68	**0.36**	27.76	27.96	**0.20**
		Case III			Case IV		
Missa	No	39.38	41.25	**1.87**	42.82	42.74	**-0.08**
Football		35.99	37.35	**1.36**	40.15	40.00	**-0.15**
Missa	Yes	34.47	34.89	**0.42**	35.07	35.16	**0.09**
Football		28.43	28.83	**0.40**	29.24	29.37	**0.13**

I: three out of the four interleaved descriptions were lost;
II: two descriptions, both from the same horizontal group, were lost;
III: two descriptions, each from a different horizontal group, were lost;
IV: one out of the four interleaved descriptions was lost.

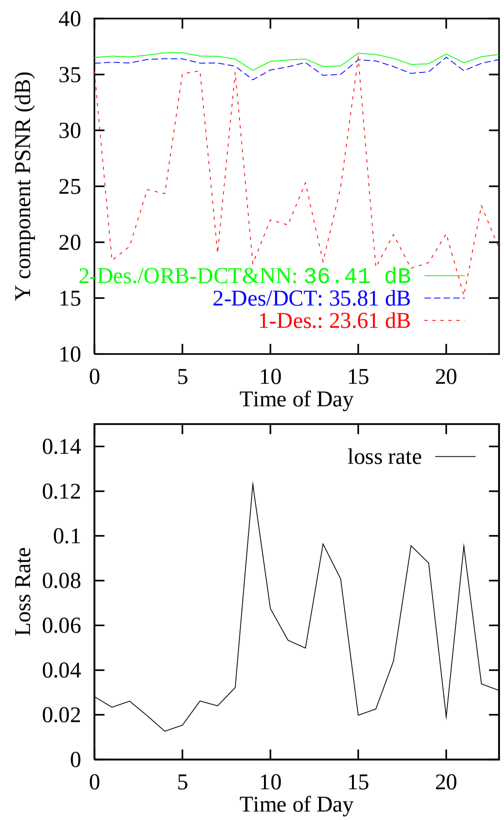
All Descriptions Correctly Received

Video Sequence	Interleaving Degree	DCT	DCT & NN	ORB -DCT	ORB-DCT & NN	Gain in PSNR (dB)
Missa	2	36.74	37.06	36.70	37.05	**0.31**
Football		30.16	30.69	30.09	30.67	**0.51**
Missa	4	35.53	36.09	35.43	36.02	**0.49**
Football		29.73	30.35	29.72	30.31	**0.58**

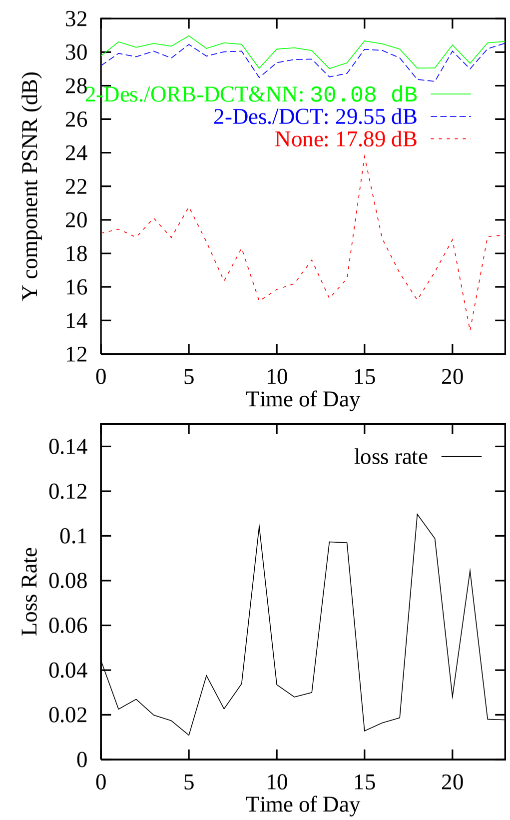
Video Streaming Prototype



Reconstruction Quality for the Champaign-Berkeley Connection

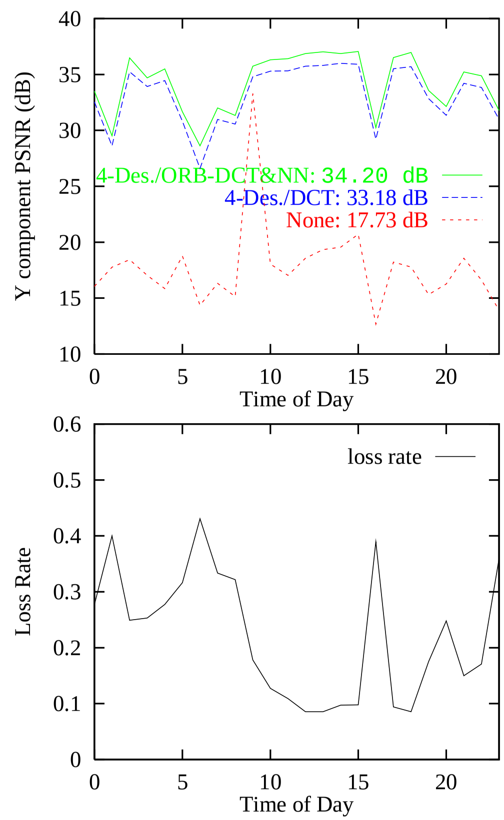


a) *missa* sequence

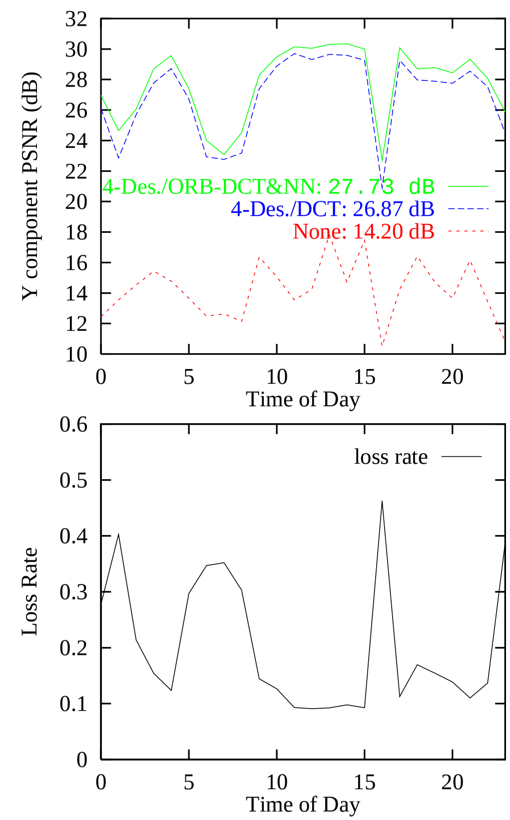


b) *football* sequence

Reconstruction Quality for the Champaign-China Connection



a) *missa* sequence



b) *football* sequence

Conclusions

- Essential in integrating error concealment and reconstruction in coding
- Future work
  - Error concealment in motion estimation
  - Block-dependent coding
  - Bandwidth-restricted concealment schemes
  - Stereo voice transmissions