III.3 Scalable Video Coding

Prof. Chia-Wen Lin
Department of CS
National Chung Cheng University
886-5-272-0411 ext. 33120
cwlin@cs.ccu.edu.tw

Outline

- Introduction
- Pyramidal coding
- Scalability in the standard codecs
- Layered coding with wavelets
- MPEG-4 FGS
- MPEG-21 SVC
- Summary

Introduction

- Layered video coding (scalable coding): a concept that enables video layers to interwork
- The codec generates two bit-streams
  - Base layer: most vital video information
  - Enhancement layer: residual information to enhance the quality of the base layer image
- three general layered coding techniques:
  - Pyramidal coding
  - Scalability in the standard video codecs (MPEG-2, H.263+, MPEG-4)
  - Wavelet-based coding (MPEG-4 I-frame, JPEG 2000)

Pyramidal Coding

- Pyramid: a data structure that provides successively condensed information of an image
- Coding schemes based on the pyramid structure are called *pyramidal coding*
  - the *apex* picture: the top of the pyramid, which gives the minimum acceptable picture resolution
  - other levels reconstruct images of higher quality by including additional information
  - lower levels toward to the bottom of the pyramid are of less significant importance

Pyramidal Coding (Cont.)

- can be used to reconstruct images of varying quality, depending on the network resources
- Two methods of pyramidal image coding:
  - Laplacian pyramid (Burt and Adelson 1983)
  - DCT pyramid

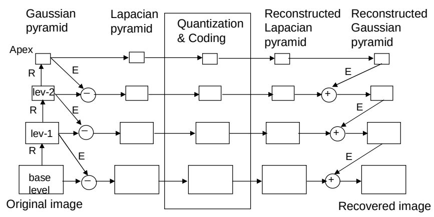
Laplacian Pyramid

- LPC (Laplacian pyramidal coding) includes two distinct types of pyramid:
  - The Gaussian pyramid
  - The Laplacian pyramid
- Coding efficiency is limited

$$G_{k+1} = reduce(G_k)$$

$$G_k = L_k + expand(G_{k+1})$$

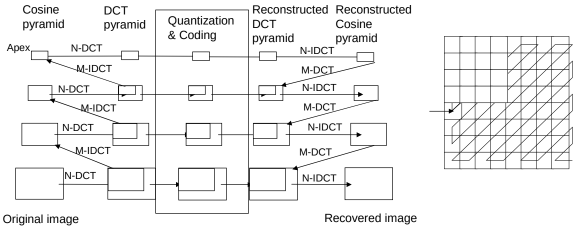
$$L_k = G_k - \text{expand}\{\text{reduce}(G_k)\}$$



E: *expand* operator R: *reduction* operator

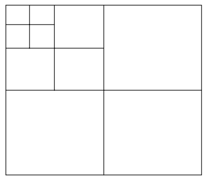
DCT Pyramid

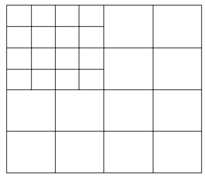
- N->M decimation:
  - retains only the *MxM* low-freq. components out of *NxN* DCT coef.
  - then performs *MxM* IDCT to obtain the decimated image
  - forms the cosine pyramid
- The remaining high-freq. coef. are quantized and coded (DCT pyramid)



Wavelet Decomposition vs. DCT Pyramid

- The Laplacian coding increase the coding are
  - 1+1/4+1/16+1/64+…=4/3
- The DCT pyramid and the wavelet transform do not increase the coding area


(a) wavelet


(b) DCT pyramid

Subband Decomposition

- The DCT pyramid implicitly embodies subband decomposition
- The effective bandwidth of these bands decreases from level to level
- Quantization and coding of each band of the pyramid can be adapted to reflect the sensitivity of the HVS
  - Coarser quantization for the higher frequency bands
  - Finer quantization for the lower frequency bands

Performance of DCT Pyramidal Coding with Data Loss from Levels


(a) 1


(b) 1+2


(c) 1+2+3


(d) All levels except apex

PSNR Performance of DCT Pyramidal Coding: Parrot

Layers received	Bit /picture [kbits]	Bit/pixel	Discard rate [%]	Quality [dB]
Apex = 5	8.1	0.02	92	21.75
4+5	28.4	0.07	72	26.48
3+4+5	56.8	0.14	44	31.06
2+3+4+5	77	0.19	24	34.78
All	101.4	0.25	0	39.2

Scalability in Standard Codecs

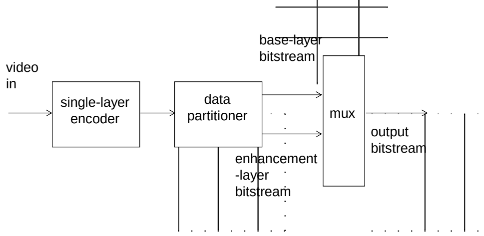
- The basic scalability tools offered are:
  - Data partitioning
  - SNR scalability
  - Spatial scalability
  - Temporal scalability
  - Hybrid scalability

Data Partitioning

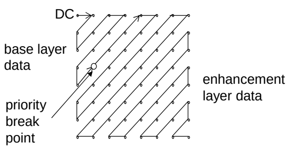
- not a true scalable coding
- a means of dividing the bitstream of a single-layer nonscalable DCT-based codec into two parts(layers):
  - The first layer
    - comprises the critical parts of the bitstream (e.g., headers, motion vectors, lower order DCT coefficients)
  - The second layer
    - is made of less critical data (e.g., higher DCT coefficients)
- Data from the second layer cannot be used unless the decoded base layer data are available

Data Partitioning (Cont.)

- At the encoder, during the quantization and zigzag scanning of each 8 x 8 DCT coefficient, the scanning is broken at the *priority break point* (PBP)



Data Partitioning (Cont.)

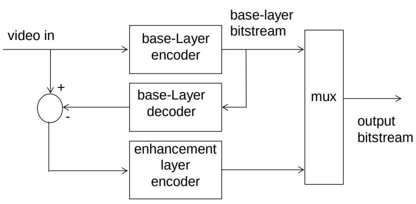


- May cause *picture drift*
- I-pictures can clean up the drift, but cause higher bit rates
- One of the limitations in data partitioning is the need for a high allocated bit rate to the base layer to avoid “blockiness”
- The simplest kind of scalability, has no extra complexity over the nonscalable encoder

SNR Scalability

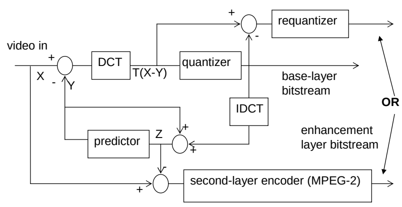
- Generates two video layers
  - Base layer and enhancement layer
  - With same spatio-temporal resolution, but different video qualities
  - The enhancement layer enhances the *signal-to-noise ratio* (SNR) of the base layer, so called SNR scalability

Two-Layer SNR Scalable Coder



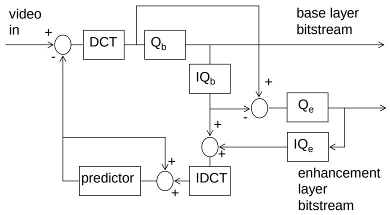
- First, the input video is coded at a low bit rate (lower image quality), to generate the base layer bitstream
- With a higher precision, to generate the enhancement layer
- May use an identical or different encoder at two layers
- The encoder is much more complex than the data-partitioning, it requires at least two nonscalable encoders

DCT-Based Base-Layer Encoder



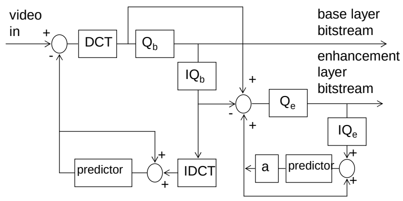
- Input pixels block X and motion-compensated predictions *Y*
- After transform coding: $T(X\text{-}Y)$
- After quantization (a quantization distortion Q): $T(X\text{-}Y)\text{-}Q$
- After the inverse DCT: $T^{-1}[T(X\text{-}Y)\text{-}Q]$ -> $T^{-1}T(X\text{-}Y)\text{-}\ T^{-1}(Q)$
  -> $X\text{-}Y\text{-}\ T^{-1}Q$
- When this error is added to the Y : $Z = Y+X\text{-}Y - T^{-1}(Q) = X - T^{-1}(Q)$
- Thus the signal coded by the second-layer encoder is
  $X - Z = X\text{-}X+T^{-1}(Q) = T^{-1}(Q)$

Two-Layer SNR Scalable Encoder with a Drift-Free Enhancement Layer



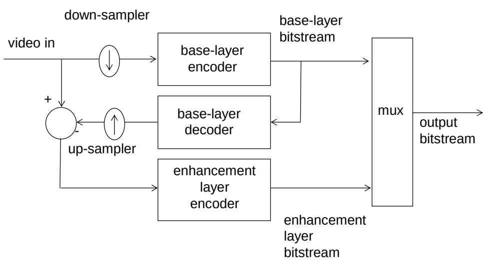
- Q_b and Q_e are the base and enhancement layer quantization step sizes, which $Q_e < Q_b$
- Compared to data partitioning, this requires only a second quantizer, an inverse quantizer, and two adders, and the complexity is not extremely great
- May cause picture drift

A Two-Layer SNR Scalable Encoder without Drift

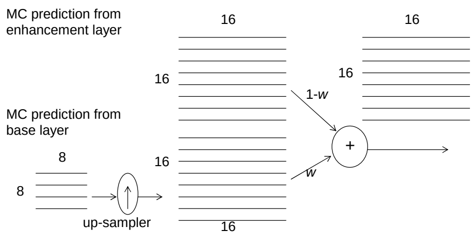


- One way to prevent picture drift is not to feed back the enhancement data into the base-layer prediction
- Intra-coding is used in the enhancement layer, which results in a very high bit rate
- To reduce bit rate, needs another encoder, then becomes much more complex
- To reduce the complexity, a leaky prediction is used
- Optimal tradeoff between bit-rate and picture drift can be achieved for a = 0.9 ~ 0.95

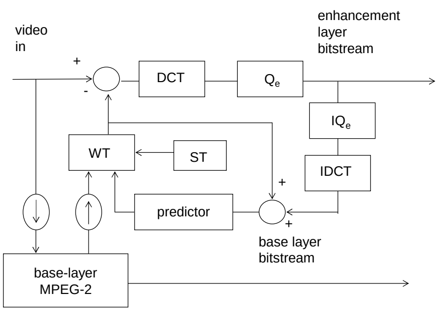
Spatial Scalability



Principle of Spatio-Temporal Prediction in The Spatial Scalable Encoder



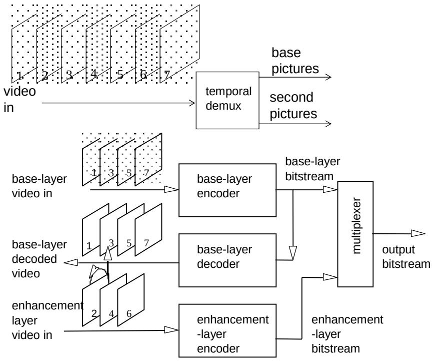
Details of Spatial Scalable Encoder



Comments on Spatial Scalable Encoder

- Comparing to data-partitioning and SNR scalable coders
  - Base-layer picture is almost free from blockiness
  - Some of very high frequency information is still missing
  - Base-layer picture can be used alone without picture drift
  - Higher price and more complexity

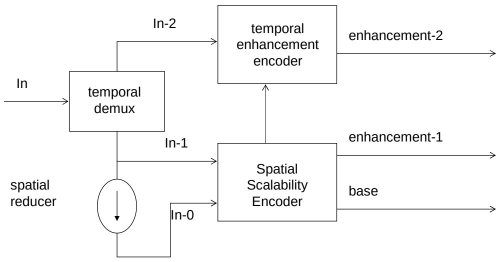
Block Diagram of a Two-Layer Temporal Scalable Encoder



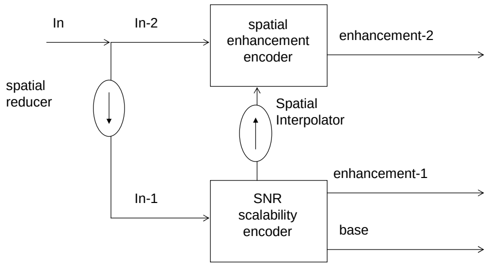
Temporal Scalability

- In fact the B pictures in MPEG-1 and MPEG-2 provide a very simple temporal scalability
  - Base layer: I and P pictures
  - Enhancement layer: B picture
- The encoder needs not be more complex than a single-layer encoder
- Free from picture drift

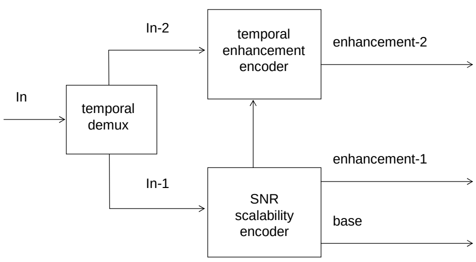
Hybrid Scalability



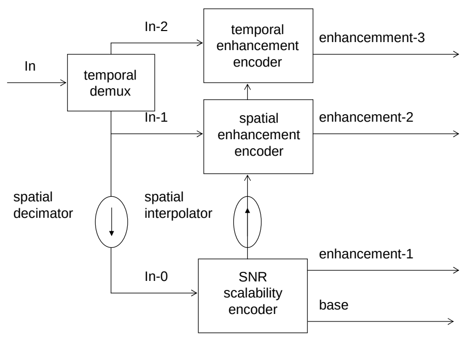
Hybrid Scalability: SNR and Spatial



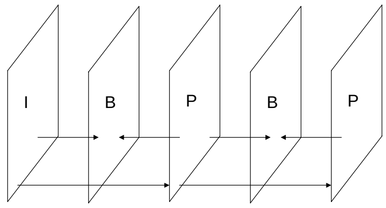
Hybrid Scalability: SNR and Temporal



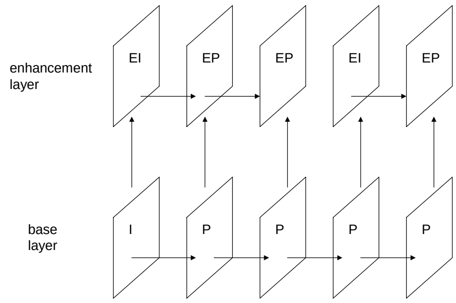
Hybrid Scalability: Spatial and Temporal



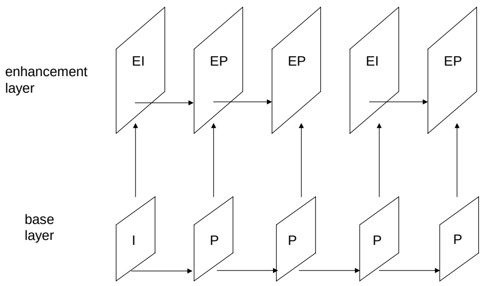
H.263+ Temporal Scalability



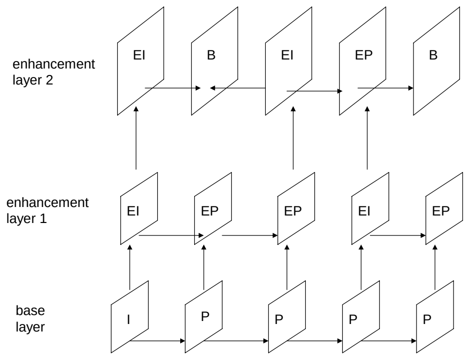
H.263+ SNR Scalability



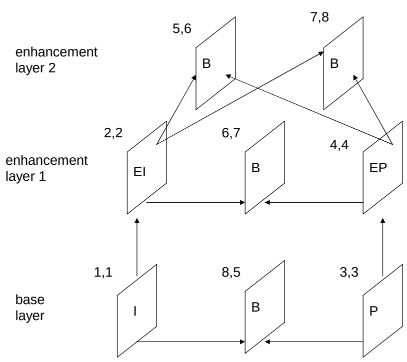
H.263+ Spatial Scalability



H.263+ Multilayer Scalability



Transmission Order of Pictures



Applications of Scalability

- Data partitioning (simplest):
  - Video over low-loss networks (e.g., ATM with congestion control)
- SNR scalability:
  - Transmission of video at different qualities
    - multiquality video, video on demand, broadcasting of TV and enhanced TV
  - Video over networks with high error or packet loss rates
    - the Internet,
    - heavily congested ATM networks

Applications of Spatial Scalability

- Spatial scalability (most complex):
  - Interworking between two different standard video codecs or heterogeneous data networks
  - Simulcasting of drift-free, good-quality video at two spatial resolutions, such as standard TV and HDTV
  - Distribution of video over computer networks
  - Video browsing
  - Reception of good quality low spatial resolution pictures over mobile networks
  - Similar to other scalable coders, transmission of error resilient video over packet networks.

Applications of Temporal Scalability

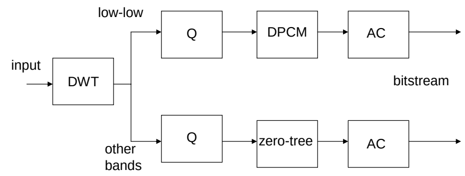
- Temporal scalability (moderately complex):
  - Migration to progressive HDTV from the current interlaced broadcast TV.
  - Internetworking between lower bit rate mobile and higher bit rate fixed networks.
  - Video over LANs, Internet and ATM for computer work stations.
  - Video over packet (Internet/ATM) networks for loss resilience.

Layered Coding with Wavelets

- One of the advantages of wavelet over DCT-based codecs is the absence of blocking artifacts
- With wavelet transforms, one can generate several layers having various spatial and quality resolutions
- The number of data layers can be much higher than what with the DCT-based codecs
- Better delivery of images over networks

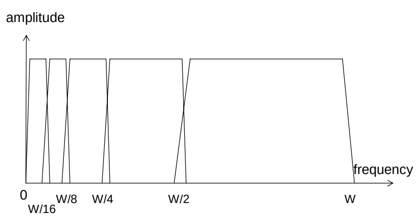
Wavelet-Based Still Image Coder

- The coding principle is based on the discrete wavelet transform, which is a subclass of subband coding
- The lowest subband is coded with a differential pulse code modulation (DPCM)
- Higher bands with the zero-tree coding technique

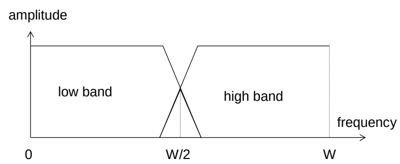


Discrete Wavelet Transform

- The basic principle is the partitioning of the signal spectrum into several frequency bands

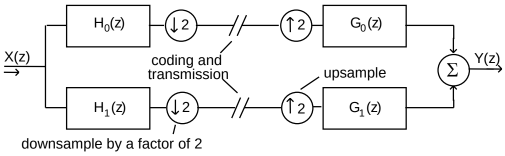


Two-band Analysis Filter



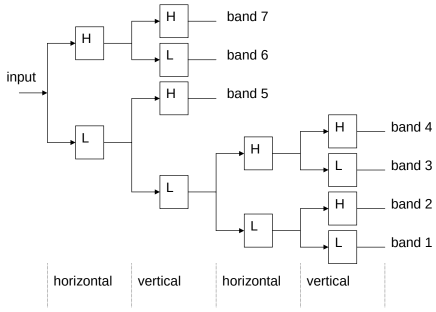
- To eliminate aliasing distortion, the synthesis and analysis filters must have certain relationships

Two-band Wavelet Transform Codec

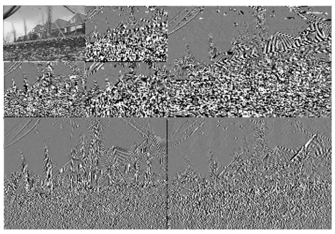


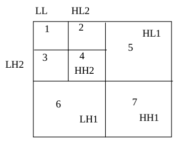
- H0(z) : the z-transform of the low-pass analysis filter
- H1(z) : the z-transform of high-pass analysis filter
- G0(z) and G1(z) are the corresponding synthesis filters
- The downsampling factor is 2, so as the upsampling

Tree-Structured Multiband Wavelet Transform



Tree-Structured Multiband Wavelet Transform

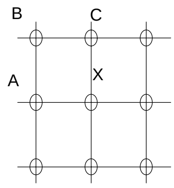




Daubechies (9,3) filter bank
Low-pass = {0.033, -0.066, -0.177, 0.420, 0.994, 0.420. -0.177, -0.066, 0.033}
High-pass = {-0.354, 0.707, -0.354}

Prediction for Coding The Lowest Band Coefficients

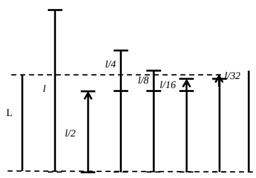
- wprd = wC, if |wA-wB| < |wA-wC|, otherwise wprd = wA



Zero-Tree Coding of Higher Bands

- The higher order wavelet coefficients are coded with the embedded zero-tree wavelet (EZW)
- The method
  - based on the concept of quantization by *successive approximation*,
  - exploits the similarities of the bands of the same orientation.

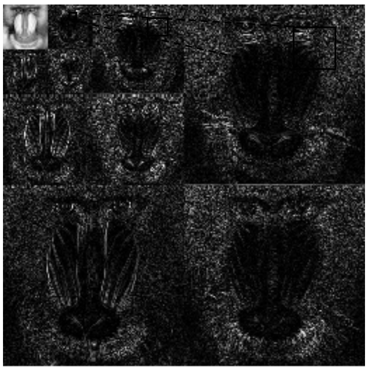
Principle of Successive Approximation



- As shown above, the quantized length, can be expressed as

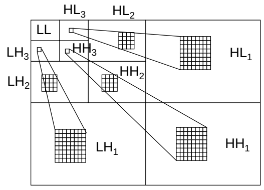
$$\hat{L} = 0 \times l + 1 \times \frac{l}{2} + 0 \times \frac{l}{4} + 0 \times \frac{l}{8} + 1 \times \frac{l}{16} \cdot + 1 \times \frac{l}{32} \cdots\cdots = \frac{l}{2} + \frac{l}{16} + \frac{l}{32}$$

Similarities among Image Subbands

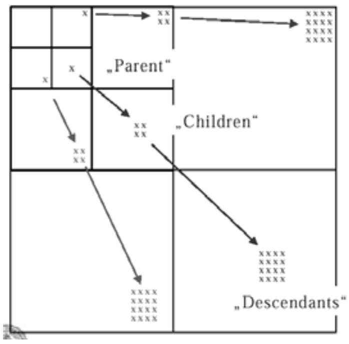


Zero-tree Coding: Quadtree Representation of Higher Bands

- Subimages of lower bands have quarter dimensions of their higher bands
- A quad-tree representation of the bands of the same orientation for a 10-band splitting is shown below (three-stage wavelet transform)
- If a coefficient in LH_3 is zero, it's more likely that its children in higher bands of LH_2 and LH_1 will also be zero => "*zero tree*"



Embedded Zerotree Wavelet (EZW) Algorithm



- Idea: Conditional coding of all descendants (incl. children)
- Coefficient magnitude > threshold: significant coefficients
- Four cases
  - ZTR: zero-tree, coefficient and all descendants are **not** significant
  - IZ: coefficient is not significant, but some descendants are significant
  - POS: positive significant
  - NEG: negative significant

Embedded Zerotree Wavelet (EZW) Algorithm

- For the highest bands, ZTR and IZ symbols are merged into one symbol Z
- Successive approximation quantization and encoding
  - Initial "dominant" pass
    - Set initial threshold T, determine significant coefficients
    - Arithmetic coding of symbols ZTR, IZ, POS, NEG
  - Subordinate pass
    - Refine magnitude of coefficients **found significant so far** by one bit (subdivide magnitude bin by two)
    - Arithmetic coding of sequence of zeros and ones.
  - Repeat dominant pass
    - Set previously found significant coefficients to zero
    - Decrease threshold by factor of 2, determine new significant coefficients
    - Arithmetic coding of symbols ZTR, IZ, POS, NEG
  - Repeat subordinate and dominate passes, until bit budget is exhausted.

Embedded Zerotree Wavelet (EZW) Algorithm

- Decoding: bitstream can be truncated to yield a coarser approximation: "embedded" representation
- Further details: *J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3445-3462, December 1993.*

Summary

- Layered coding is a means of facilitating unequal protection of image/video information at various important levels
- Three general layered coding schemes are discussed
  - Pyramidal coding:
    - only has a historical importance
    - DCT pyramid has proven to be very efficient in image condensation
  - Layered coding based on standard DCT-Based codec
    - only three methods of scalability have been recognized (spatial, SNR, and temporal)
    - supported in H.263+ and MPEG-2
  - Wavelet transform
    - has been adopted in JPEG-2000 and MPEG-4
    - generates more layers than DCT-based codec => very attractive in video networking

MPEG-4 Fine Granularity Scalability

Prof. Chia-Wen Lin
Department of CS
National Chung Cheng University
886-5-272-0411 ext. 33120
cwlin@cs.ccu.edu.tw

MPEG-4 Fine Granularity Scalability

- ***Internet applications***
- ***broadcast applications over packet networks***

- Low complexity
- Supports both unicast & multicasting capabilities
- Supports various layers of SNR enhancements
- Covers a “range” of bitrates instead of a few discrete bitrates
- Base-layer compatible to MPEG-4
- Error robustness

Challenges for Internet Video

- Challenges
  – No QOS guarantees (bandwidth, delay, packet loss)
  – Bandwidth differences of heterogeneous networks
  – Bandwidth variation with time
- Conventional video coding techniques
  – Optimizing perceived quality at a given bitrate

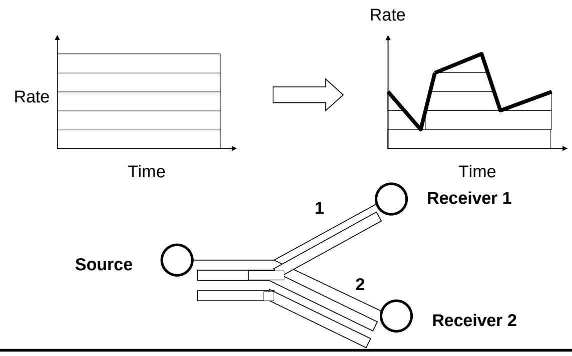
Bandwidth Variation

- “Broadband” Internet access has wider variation:
  – Cable modem: from < 100 to > 1000 Kbit/sec
  – DSL: from < 600 to > 6000 Kbit/sec

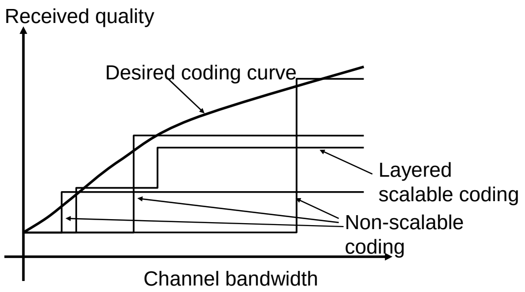
Bandwidth Variation - Solutions

- Multiple Bit-streams pre-encoded with different bit-rates
  – Large storage
  – Complexity in bit-streams management and switching
- Real-time Transcoder
  – High complexity in the streaming sever
- **Scalable video**
  – **Degree of scalability**

Why Scalable Video – Rate Control



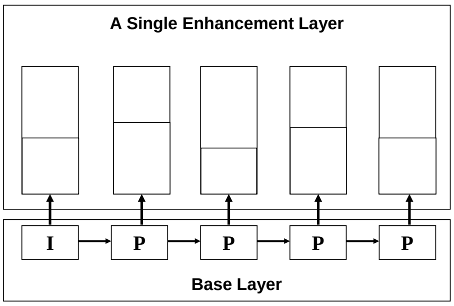
Performance Comparison of Various Coding Schemes



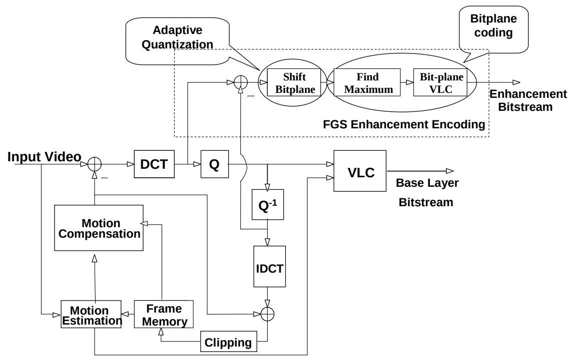
FGS Standard

- July '99 **FGS Working Draft issued**
- Oct '99 **MPEG adopted FGS as "MPEG-4-Version4"**
  First proposals on FGS profiles presented
- Dec '99 **Large number of proposals in support of FGS**
- Mar '00 FGS was issued as MPEG-4-V4 (PDAM)

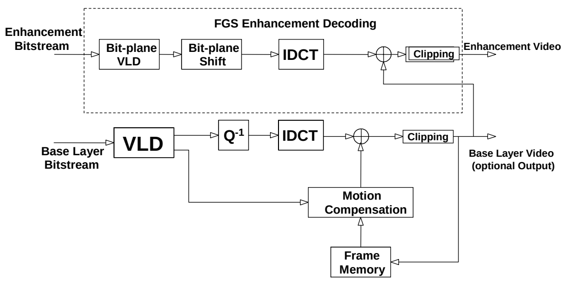
Rate Adaptation with MPEG-4 FGS



FGS Encoder



FGS Decoder



FGS Bit-plane Coding

- Bitplane coding considers each DCT coefficients as *a binary number of several bits* instead of a decimal integer
- Coding the DCT coefficients from MSB plane to LSB plane
- Find maximum number of bitplanes

FGS Bit-plane Coding

- Symbols
  - (RUN , EOP)
    - VLC
    - Escape, FLC
  - All-Zero
- Different VLC tables for different bitplanes
  - 4 VLC tables (MSB, MSB-1, MSB-2, others)
- At the first 2 bitplanes, All-Zero in macroblock layer (special pattern)

FGS Bit-plane Coding Example

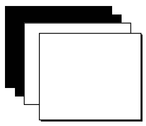
❖ The absolute residue values after zigzag ordering are given as follows:

10,0,6,0,3,0,…,0,0

$(10)_{10}=(1010)_2$

1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,…0,0 (MSB)
0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,…0,0 (MSB-1)
1,0,1,0,0,1,0,1,1,0,0,1,0,0,0,0,…0,0 (MSB-2)
0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,…0,0 (MSB-3)

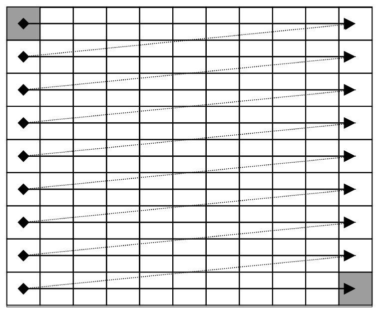
FGS Bit-plane Coding Example



- Converting the four bit-planes into (RUN,EOP)symbols:

(0,1)	(MSB)
(2,1)	(MSB-1)
(0,0),(1,0),(2,0)(1,0),(0,0),(2,1)	(MSB-2)
(5,0),(8,1)	(MSB-3)

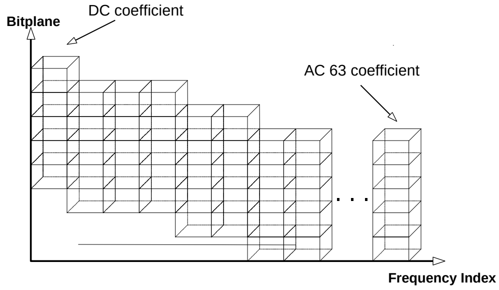
Macroblock Scan Order in FGS



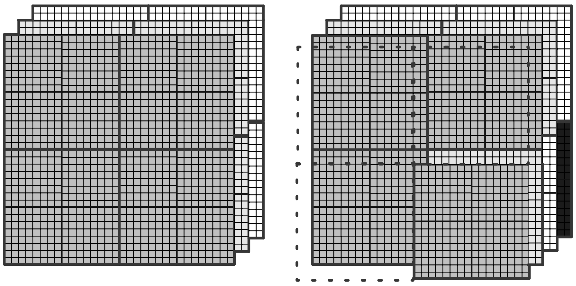
Selective Enhancement Tools in FGS

- Improving perceived visual-quality
- Base layer
  - Quantization matrix for different coefficients
  - Quantization factor varies on macroblocks
- Enhancement layer
  - Bitplane shifting
    - Frequency weighting (FW)
    - Selective enhancement (SE)

Frequency Weighting



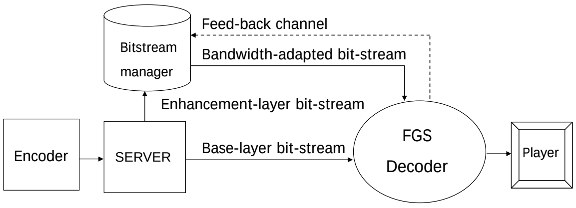
Selective Enhancement



Selective Enhancement on Face Area

without SE on face   with SE on face

FGS-based Bandwidth Adaptation

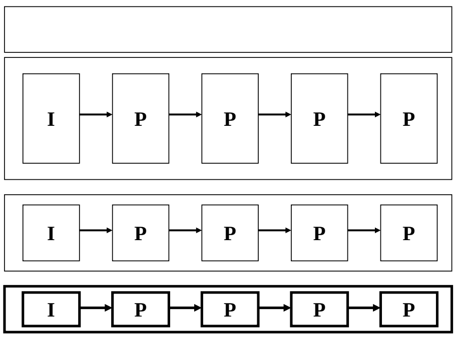


Packet-Loss Resilience

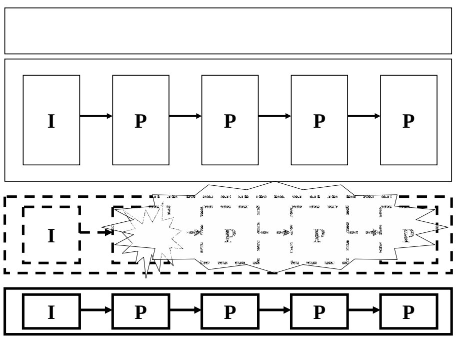
- Retransmission- and FEC-based recovery methods have been shown to be viable for real-time applications
- HOWEVER, there are NO guarantees for *on-time* delivery

⇒ Video has to be packet-loss resilient

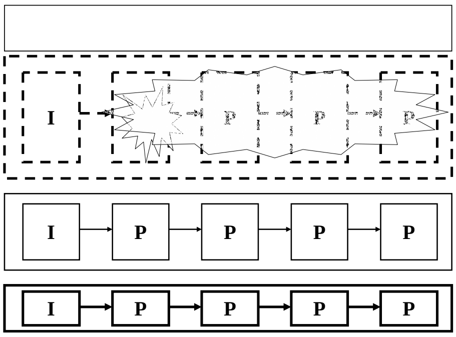
Packet-Loss Resilience – Non-Scalable Option



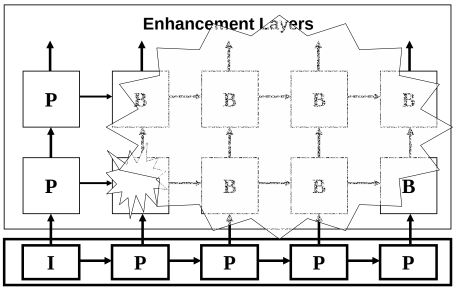
Packet-Loss Resilience – Non-Scalable Option



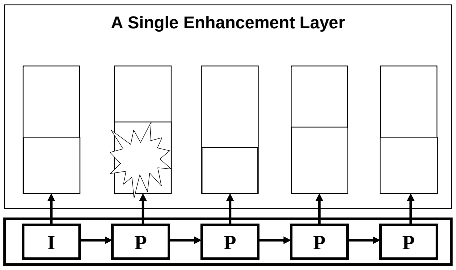
Packet-Loss Resilience – Non-Scalable Option



Packet-Loss Resilience – Multilayer Scalable Option



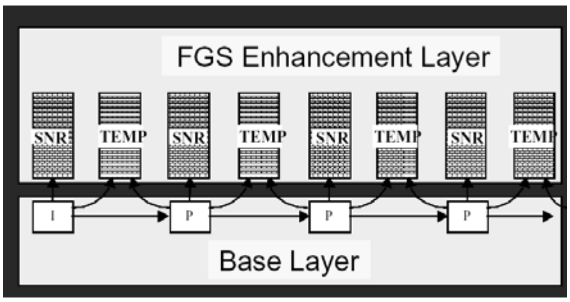
Packet-Loss Resilience - FGS



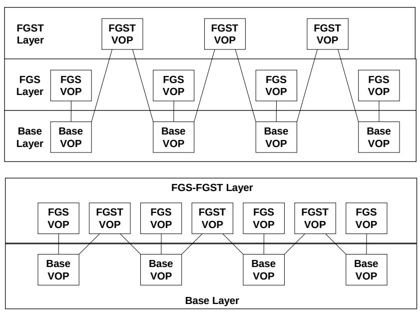
Packet-Loss Resilience

MPEG-4 Single-Layer FGS

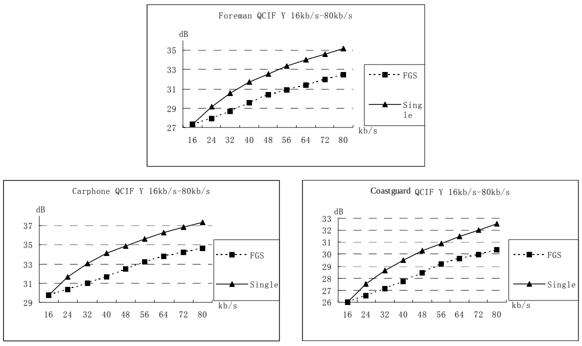
Hybrid SNR/Temporal Scalability



Hybrid SNR/Temporal Scalability



Performance Comparison of FGS Coder and Single-Layer Coder



1.5~2.5 dB degradation was observed

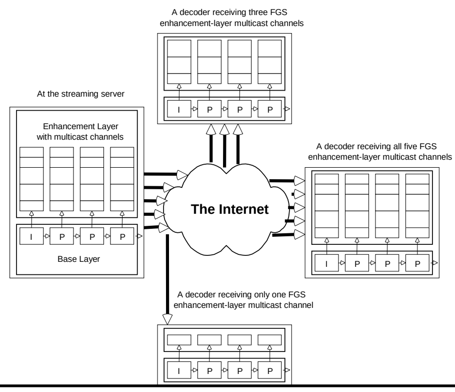
Other Proposals for Improving FGS

- Progressive FGS (PFGS)
  - Proposed by Microsoft Research Asia
- Adaptive Motion-Compensated FGS (AMC-FGS)
  - Proposed by Philips Research
- Reliable FGS (RFGS)
  - Proposed by NCTU
  - Uses leaky prediction

FGS Multicasting

- MPEG-4 FGS method
  - consists of layered video coding that supports SNR, temporal, and hybrid temporal-SNR scalabilities
  - simplicity and flexibility in supporting multicast streaming applications
  - Base layer and one or more enhancement layer send into different multicasting group individually
- Problems in multimedia transmission
  - Network heterogeneous
    - Different codec, resource, network conditions, user requirements, etc.
  - Feedback implosion

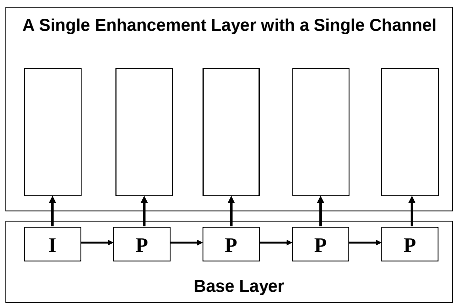
FGS Multicasting



Current FGS Multicasting Approaches

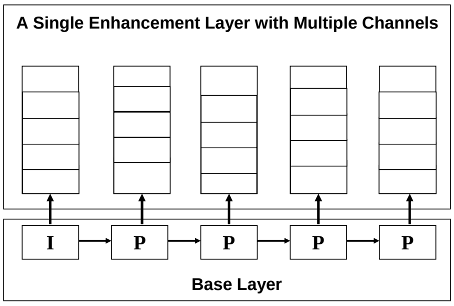
- Receiver-driven multicast
  - Multicasting the base-layer over one MC group
  - Multicasting the enhancement-layer over one or more MC groups
  - Total flexibility in creating “customized MC channels”
- Sender Adaptive & Receiver-driven multicast
  - Better layer arrangement and resource allocation

FGS Multicast Channels

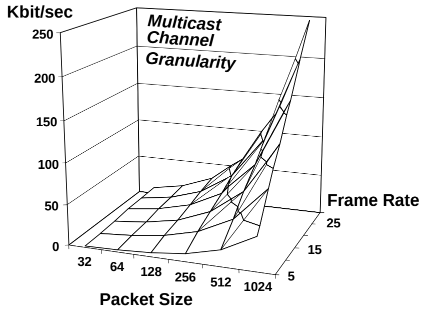


Page 15

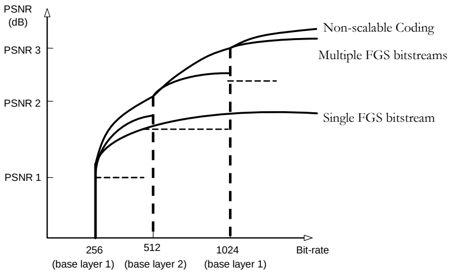
FGS Multicast Channels



Multicasting with FGS



Combination of Scalable Coding & Bistream Switching

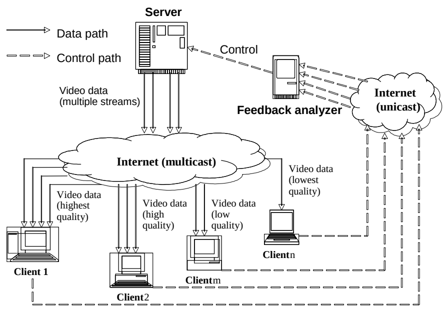


The coding efficiency of MPEG-4 FGS with "large" bit-rate range is usually not good

⇒FGS + multiple-bitstream switching

(integrated with receiver-driven multicasting)

Receiver-Driven Multicasting



Summary

- MPEG-4 FGS solves the bandwidth-variation problem over the Internet
  - A single enhancement-layer stream
- Totally flexible, efficient, and simple solution
  - For both unicast and multicast
- Packet loss resilient
- Open standard

MPEG-21 Scalable Video Coding

Prof. Chia-Wen Lin
Department of CS
National Chung Cheng University
886-5-272-0411 ext. 33120
cwlin@cs.ccu.edu.tw