Low Bit Rate H.263 + Video Coding: Efficiency, Scalability and Error Resilience

Faouzi Kossentini

Signal Processing and Multimedia Group
Department of Electrical and Computer Engineering
University of British Columbia

http://spmg.ece.ubc.ca/

Outline

- Introduction: Low bit rate video coding
- The H.263/H.263+ standards and their optional modes
- Efficiency: Performance and complexity of individual modes and combinations of modes
- Efficiency: Real-time software-only encoding
- Efficiency: Rate-distortion optimized video coding
- Scalability: Description & Characteristics
- Scalability: Rate-distortion optimized framework
- Error resilience: Synchronization & error concealment
- Error resilience: Multiple description video coding
- Conclusions: H.263/H.263+, MPEG-4 and research directions

Low Bit Rate Video coding

Why?: Increasing demand for video conferencing and telephony applications, limited bandwidth in PSTN and wireless networks

Video coding algorithms:

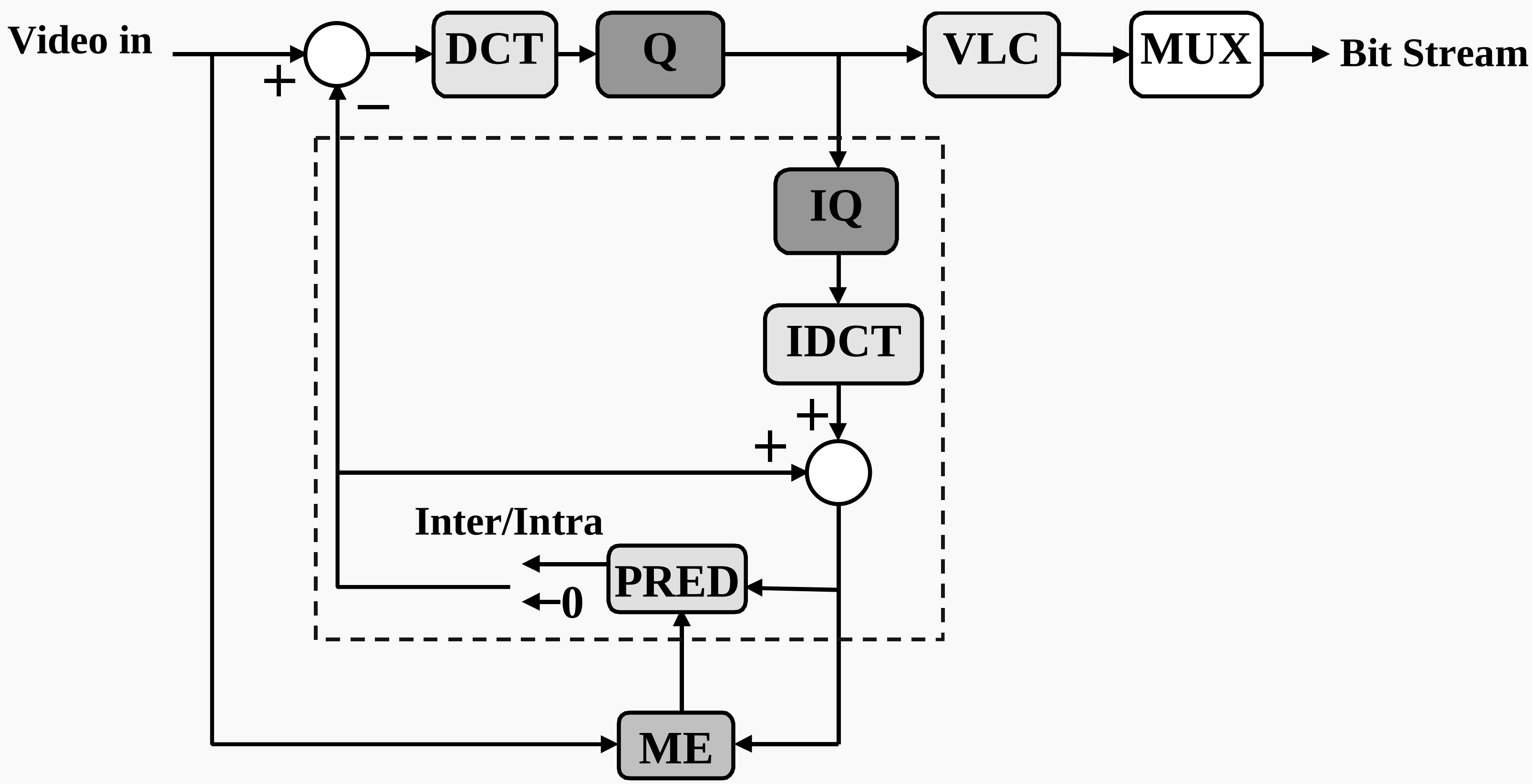
- Waveform based coding: MC+DCT/wavelets, 3D subband, etc.
- Object- and model-based coding : shape coding, wireframes, etc.

Video coding standards:

- ITU-T H.261(1990), H.262 (1994), H.263 (1995), H.263+ (1998)
- ISO/IEC MPEG1 (1992), MPEG2 (1994), MPEG4 (1999)

H.263 version 2 (H.263+): Higher coding efficiency, more flexibility, scalability support, error resilience support

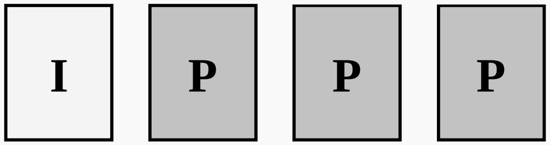
The H.263 Standard



The H.263 Standard

Structure

- Inter picture prediction to reduce temporal redundancies
- DCT coding to reduce spatial redundancies in difference frames



Major enhancements to H.261

- More standardized input picture formats
- Half pixel motion compensation
- Optimized VLC tables
- Better motion vector prediction
- Four (optional) negotiable modes

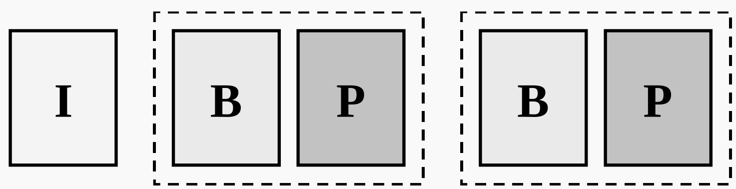
The H.263 Standard: Optional Modes

Unrestricted Motion Vector (UMV) mode, Annex D: **Motion vectors (MVs) allowed to point outside the picture, MV range extended to ±31.5**

Syntax-based Arithmetic Coding (SAC) mode, Annex E : **SAC used instead of VLCs**

Advanced Prediction (AP) mode, Annex F: **Four MVs per macroblock, over-lapped motion compensation, MVs allowed to point outside the picture area**

PB-frames (PB) mode, Annex G: **A Frame with a P-picture and a B-picture used as a unit**



The H.263+ Standard: Optional Modes

Unrestricted Motion Vector (UMV) mode, Annex D: MV range extended to ±256 depending on the picture size, reversible VLCs used for MVs

Advanced Intra Coding (AIC) mode, Annex I: Inter block prediction from neighboring intra coded blocks, modified quantization, optimized VLCs

Deblocking Filter (DF) mode, Annex J: Deblocking filter inside the coding loop, four motion vectors per macroblock, MVs outside picture boundaries

Slice Structured (SS) mode, Annex K: Slices used instead of GOBs

Supplemental Enhancement Information (SEI) mode, Annex L: Supplemental information included in the stream to offer display capabilities

The H.263+ Standard: Optional Modes

Improved PB-frames (IBP) mode, Annex M: Forward, backward and bi-directional prediction supported, delta vector not transmitted

Reference Picture Selection (RPS) mode, Annex N: Reference picture selected for prediction to suppress temporal error propagation

Temporal, SNR, and Spatial Scalability mode, Annex O: Syntax to support temporal, SNR, and spatial scalability

Reference Picture Resampling (RPR) mode, Annex P: Warping of the reference picture prior to its use for prediction

The H.263+ Standard: Optional Modes

Reduced Resolution Update (RRU) mode, Annex Q: Encoder allowed to send update information for a picture encoded at a lower resolution, while still maintaining a higher resolution for the reference picture

Independently Segmented Decoding (ISD) mode, Annex R: Dependencies across the segment boundaries not allowed

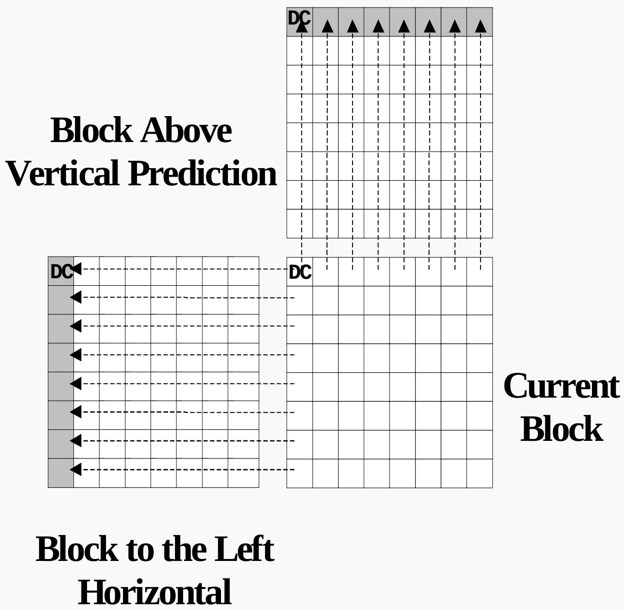
Alternative Inter VLC (AIV) mode, Annex S: The intra VLC table designed for encoding quantized intra DCT coefficients in the AIC mode used for inter coding

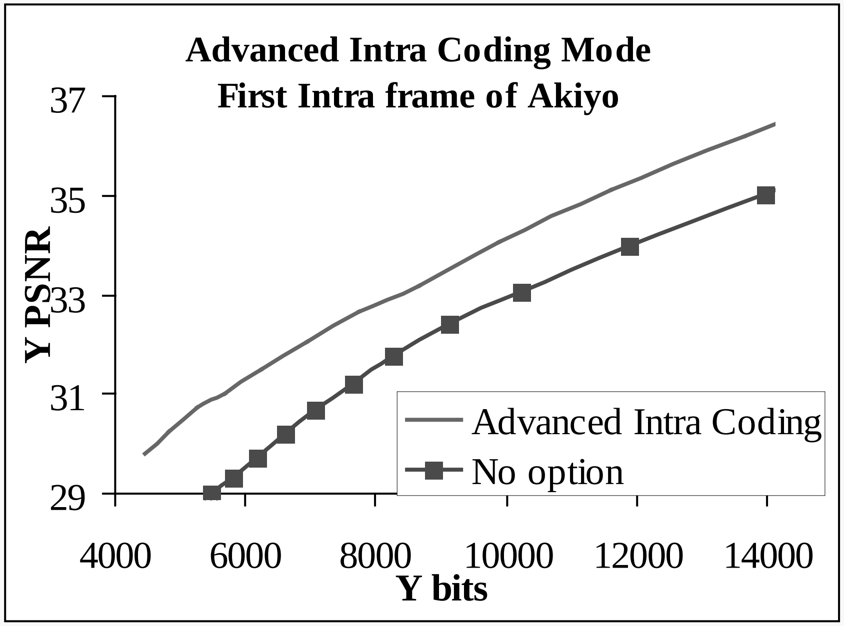
Modified Quantization (MQ) mode, Annex T: Quantizer allowed to change at the macroblock layer, finer chrominance quantization employed, range of representable quantized DCT coefficients extended to [-127, +127]

Efficiency: Performance & Complexity of Individual Modes

Advanced Intra Coding (AIC) mode, Annex I: Inter block prediction from neighboring intra coded blocks, modified quantization, optimized VLCs

Performance & Complexity: Compression efficiency for intra macroblocks increased, encoding time increased by ~5%, required memory slightly increased

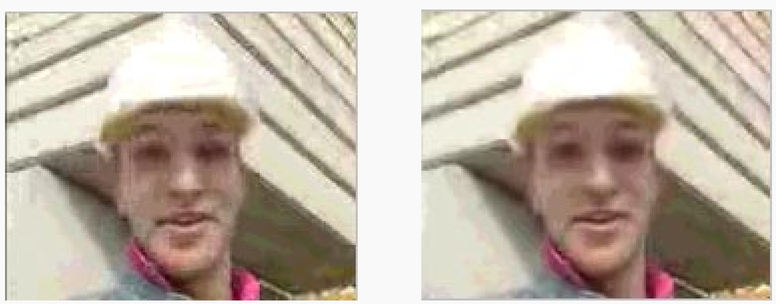




Efficiency: Performance & Complexity of Individual Modes

Deblocking Filter (DF) mode, Annex J: Deblocking filter inside the coding loop, four motion vectors per macroblock, MVs outside picture boundaries

Performance: Subjective quality improved (blocking & mosquito artifacts removed), 5-10% additional encoding time



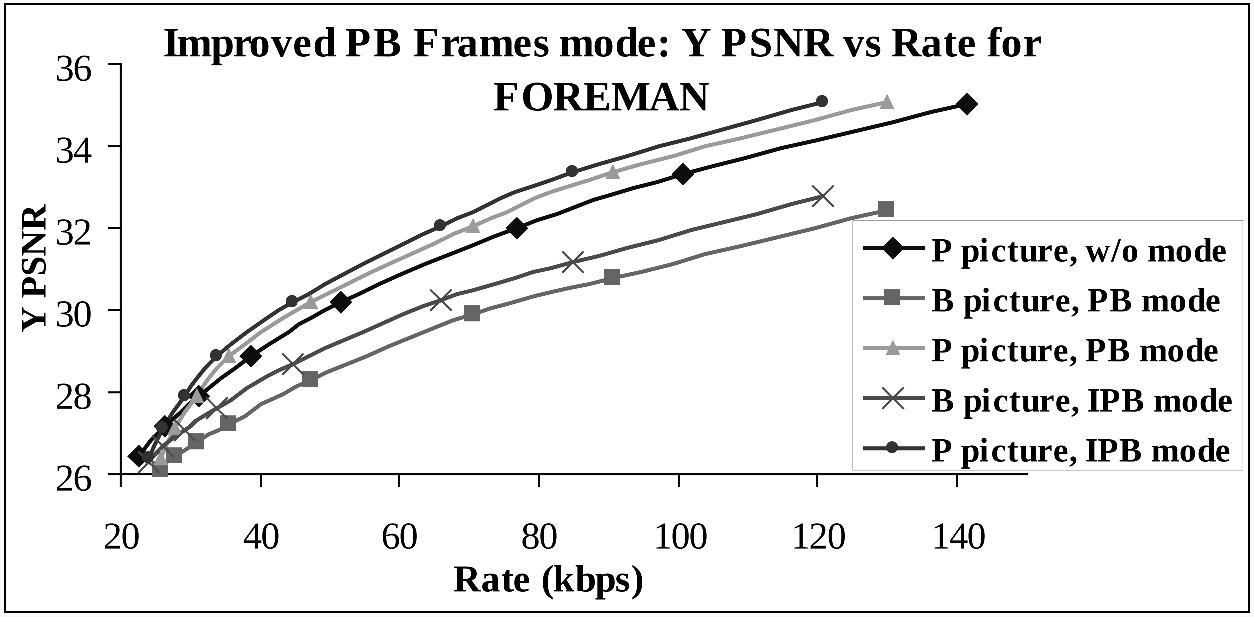
(a) (b)

Without (a) and with (b) DF mode set on

Efficiency: Performance & Complexity of Individual Modes

Improved PB-frames (IBP) mode, Annex M: Forward, backward and bi-directional prediction supported, delta vector not transmitted

Performance: Picture rate doubled, complexity increased, more memory required, one frame delay incurred



Efficiency: Performance & Complexity of Individual Modes

Alternative Inter VLC (AIV) mode, Annex S: The intra VLC table designed for encoding quantized intra DCT coefficients in the AIC mode used for inter coding

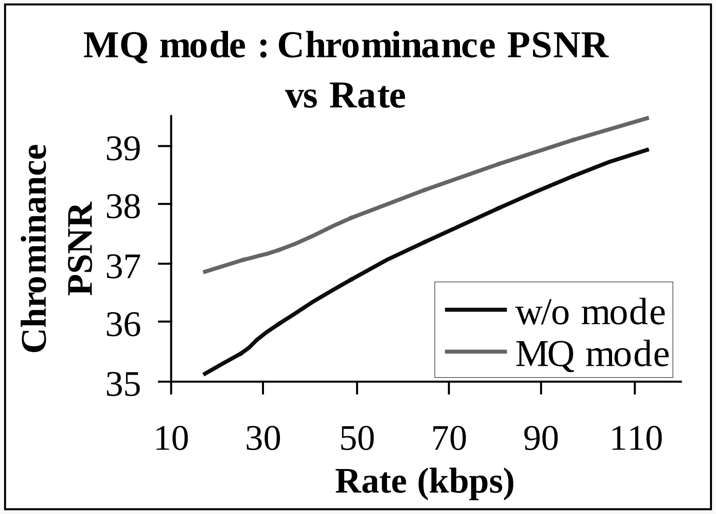
Performance: Useful at high bit rates, bit savings of as much as 10% achieved, less than 2% additional encoding/decoding time required

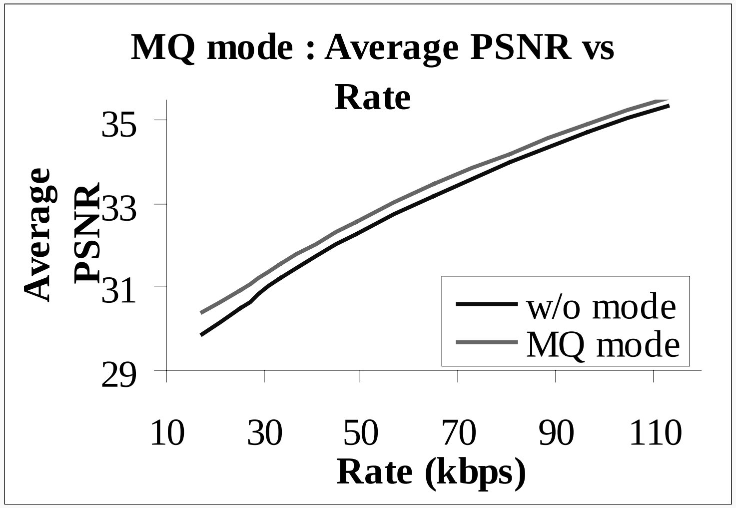
Sequence	Quantizer Step Size	Bit savings using AIV mode
AKIYO	4	5%
	12	0%
FOREMAN	4	7%
	8	4%
	16	1%

Efficiency: Performance & Complexity of Individual Modes

Modified Quantization (MQ) mode, Annex T: Quantizer allowed to change at the macroblock layer, finer chrominance quantization employed, range of representable quantized DCT coefficients extended to [-127, +127]

Performance: Chrominance PSNR increased substantially at low bit rates, more flexible rate control supported, very little complexity and computation time added to the coder





Efficiency: Performance & Complexity of Combinations of Modes

H.261 PSNR = 31.91

H.263 PSNR = 32.44

H.263+ PSNR = 33.29

H.263+ modes: AIC, MQ, UMV, DF, AP

Efficiency: Real-time Software-only Encoding

- **Required for real-time video telephony and conferencing: encoding at a minimum of 8-10 fps for QCIF resolution**

	FOREMAN (fps)	***AKIYO (fps)***
H.263+ Baseline coding (Full ME)	1.8	2.4
H.263+ Baseline coding (Fast ME)	4.3	5.8
H.263+: MQ, DF, AIC, AIV, IPB, UMV, AP modes turned on (Fast ME)	3	3.8

Efficiency: Real-time Software-only Encoding

- Algorithmic approach: Developing efficient platform-independent video coding algorithms
- Hardware dependent approach: Mapping SIMD oriented coding components onto Intel's MMX architecture

Function	Computational Load
Reference IDCT	25%
Integer Pixel ME	12%
Fast DCT	11%
Half Pixel ME	10%
Quantization	9%

Efficiency: Real-time Software-only Encoding

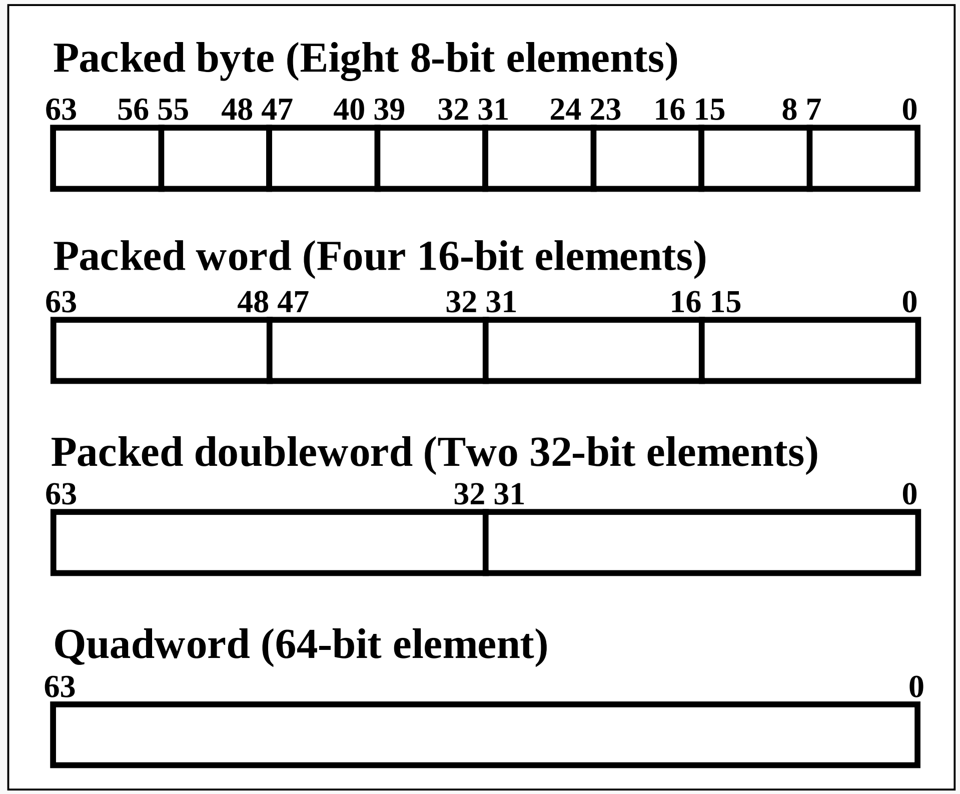
Efficient Video Coding Algorithms:

- IDCT for blocks with all or most of the coefficients equal to zero avoided or efficiently computed (respectively)
- SAD for most macroblocks computed efficiently via prediction
- DCT & Quantization for most blocks avoided via prediction
- Linear approximation methods used for half-pixel ME
- Quantization: look-up tables employed

Efficiency: Real-time Software-only Encoding

Intel's MMX Architecture:

- **Features :**
- **SIMD structure**
- **Four new data types**
- **57 new instructions**
- **Saturation arithmetic**

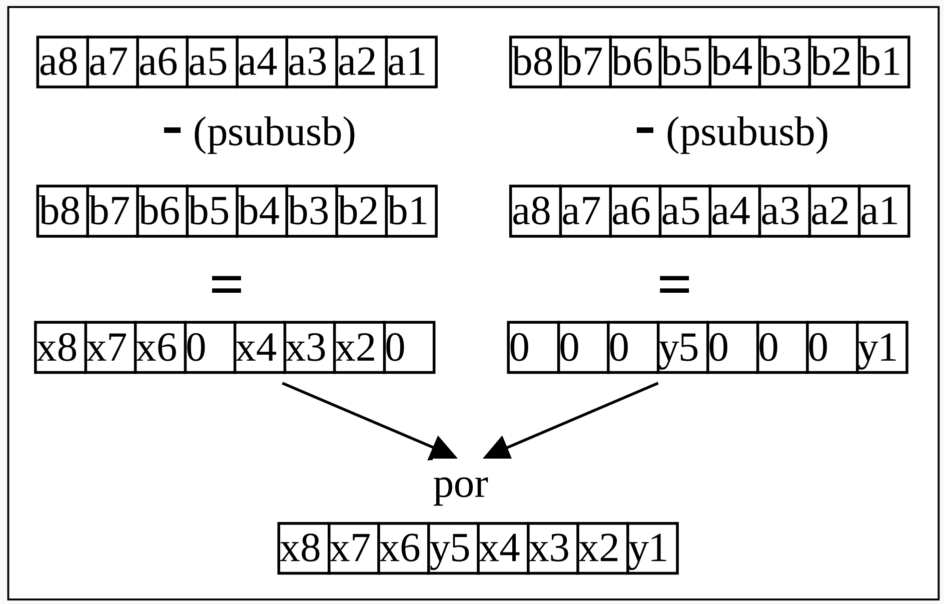


Efficiency: Real-time Software-only Encoding

MMX Mapping of SAD Computations:

- ➢SAD computation function implemented in MMX for 16x16 blocks (baseline coding) and 8x8 blocks (optional modes)

- ➢Saturation arithmetic is used to perform absolute difference operations



- ➢Result: 4-5 times faster SAD computations

Efficiency: Real-time Software-only Encoding

DCT MMX Implementation:

- ➢Floating point arithmetic to fixed point arithmetic
- ➢Four rows processed at a time
- ➢Results: 4 times faster using fixed point arithmetic, 3 times faster using MMX as compared to the integer "C" implementation

	Foreman 48 Kb/sec	Akiyo 8 Kb/sec	Foreman 64 Kb/sec	Akiyo 28 Kb/sec
PSNR with floating point DCT	30.47	32.63	31.56	37.92
Variation in PSNR with integer or MMX DCT	-0.01	+0.02	-0.01	0.00

Other MMX Implementations: Interleaved transfers for half pixel ME (7x), interp. (3x), motion comp. (2x), block data transfers (2x)

Overall Performance:

- 14-17 fps baseline H.263+ encoding with fast ME
- 8-10 fps with all of the H.263+ optional modes
- More than a 100% increase in speed using the efficient algorithms and more than a 25% additional increase in speed using MMX

Function	Computational Load
DCT	9%
IDCT	9%
Quantization	8%
Load ME Area	7%
Scan	6%
Interpolation	5.5%
Fast ME	5%
Reconstruct Full Image	5%
Half Pixel ME	4%
Motion Compensation	4%

	encoded frame rate before optimization		encoded frame rate after optimization	
	FOREMAN	*AKIYO*	*FOREMAN*	*AKIYO*
H.263+ Baseline coding (Full ME)	2.0	2.4	4.1	4.4
H.263+ Baseline coding (Fast ME)	6.4	7.5	14.3	17.0

Efficiency: Rate-Distortion Optimized Video Coding

Motivation:

- Best possible tradeoffs between quality (distortion) and bit rate obtained via Rate-Distortion (RD) optimization, through the use of the Lagrangian cost measure

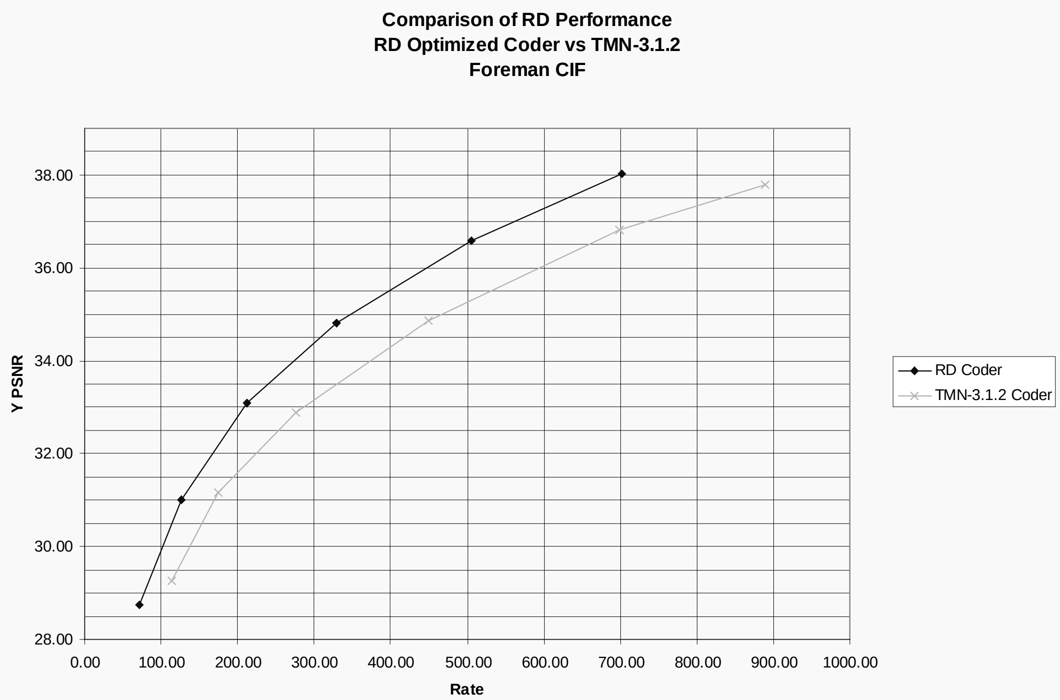
Solution:

- Motion vector selected that yields the best rate-quality tradeoff
- Coding mode (Intra, Inter16, Inter8, Skipped) selected that yields the best rate-quality tradeoff based on the already determined motion vectors

Efficiency: Rate-Distortion Optimized Video Coding

Result:

- 10 - 30% savings in bit rate obtained using the RD optimized coder for most sequences over our reference software



Scalability: Description & Characteristics

Description:

- Multi-representation coding at different quality and resolution levels allowed
- Three general types (SNR, spatial, temporal) of scalability supported

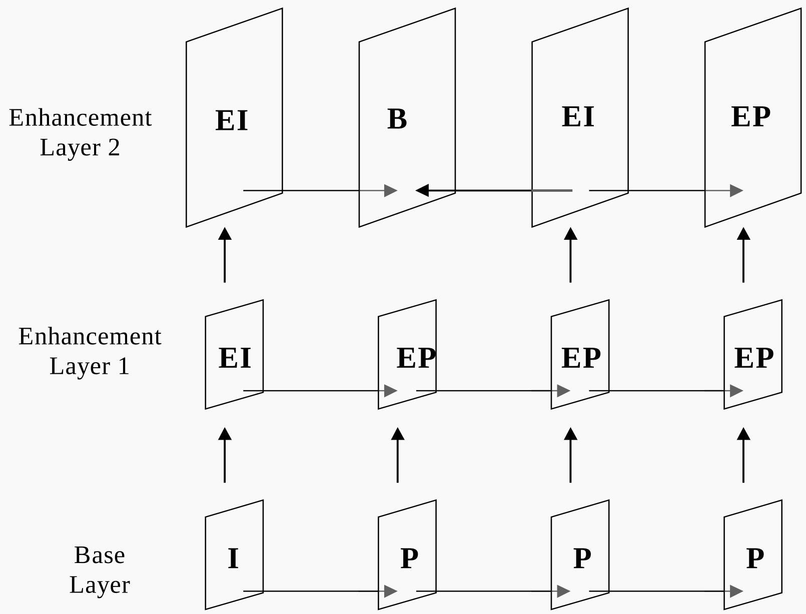
Characteristics:

- Each layer built incrementally on its previous layer, increasing the decoded picture quality, resolution, or both
- Only the first (base) layer independently decoded
- Information from temporally previous pictures in the same layer or temporally simultaneous pictures in a lower layer used

Scalability: Description & Characteristics

Example:

- Enhancement layer 1: Example of SNR scalability
- Enhancement layer 2: Example of spatial and temporal scalability



Scalability: RD Optimized Framework

Motivation:

- Compression efficiency of a higher layer dependent on that of all preceding (lower) layers
- Higher compression efficiency and more flexible joint rate-distortion control achieved via an RD optimized framework

Current research direction:

- Employ (1) multiple rate constraints (explicit for each layer) OR (2) a single rate constraint with explicit distortion constraints for each layer.
- Apply RD optimization to temporal, spatial and SNR scalable coding algorithms

Error Resilience: Synchronization & Error Concealment

Use of Synchronization Markers:

- Different synchronization markers (between GOBs, slices, etc.) used for higher error resilience

Error concealment:

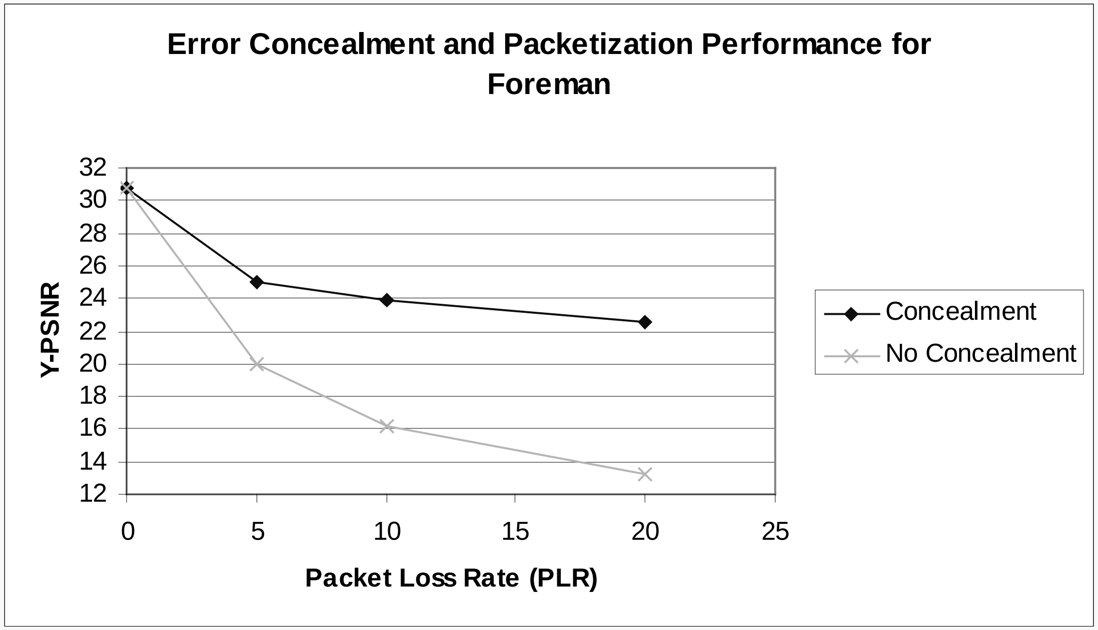
- Missing motion vectors for a macroblock received in error recovered from neighboring macroblocks (via spatial methods)
- Texture information from the previous frame is compensated using the recovered motion vectors

Error Resilience: Synchronization & Error Concealment

Experimental Results:

➢**Packetization: one packet consists of a complete GOB.**

➢**Using synchronization markers and error concealment yields PSNR gains of as much as 9 dB.**



Error Resilience: Multiple Description Video Coding

Use of Reference Picture Selection Mode: Use of different pictures (or picture segments) for prediction allowed by H.263+, hence multiple description video coding

Multiple Description:

- Different descriptions of the source sent on different channels, may be available at the decoder
- Redundancy among descriptions minimized
- Useful even if only one description is available

Temporal Reference for Prediction:

- Different temporal reference pictures employed for prediction of different pictures
- Error propagation confined to those pictures dependent on the missing reference

Conclusions

H.263+ versus H.263:

- **H.263+ backward compatible with the H.263 (same baseline)**
- **Efficient ways provided by H.263+ of trading additional complexity for more compression efficiency, scalability, resilience and flexibility**

H.263/H.263+ versus MPEG-4:

- **H.263 output decodable by MPEG-4, H.263+/MPEG-4 sharing tools**
- **Although somewhat available via chroma keying in H.263+, object-based functionality better provided by MPEG-4**
- **H.263/H.263+ public-domain implementation:** http://spmg.ece.ubc.ca/

Future directions:

- **Still higher coding efficiency and better scalability/resilience possible**
- **More object-based tools and capabilities, better (efficient) shape coding**