2. Review of Video Coding Standards

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

Outline

- JPEG, H.261, MPEG-1, MPEG-2, H.262, H.263, H.263+, MPEG-4, MPEG-7, H.264
- Trends in video coding standards
- Is video compression dead?
- Trends of video compression technologies

Video Coding Standards

- Every standard was developed with an initial driving-application which may be different from the current driving applications
- The requirements from the driving-applications determined how the standard was developed
- Current driving applications determine how the standards will evolve
- The application ultimately determines if the standard will be successful
- ITU-T vs. ISO

Video Compression Standards Evolution

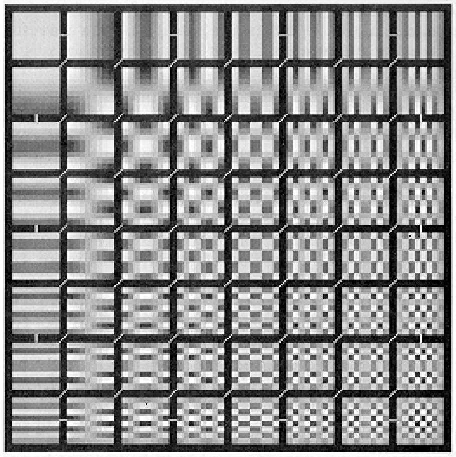
- JPEG (mid 80's)
- H.261 (mid 80's, ~2x improvement compared to JPEG)
- MPEG-1 (late 80's)
- MPEG-2/H.262 (late 80's to early 90's, ~2x improvement compared to H.261)
- H.263 (early to mid 90's)
- MPEG-4 ASP (mid to late 90's, ~1.5x improvement compared to MPEG-2)
- MPEG-4 AVC/H.264 (00's, ~3x improvement compared to MPEG-2)

JPEG

– Driving application: image compression for storage
– Technology: Integration of existing techniques
Discrete Cosine Transform (DCT) and Entropy coding

Discrete Cosine Transform (DCT)

8x8 DCT
Basis Functions



Entropy Coding

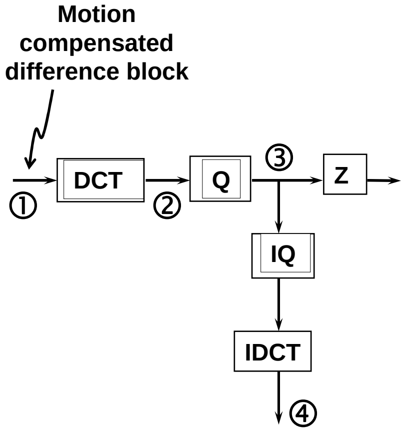
Run-length Coding:

Input sequence:
0,0,-3,5,1,0,-2,0,0,0,0,2,
-4,3,-2,0,0,0,1,0,0,-2

Run-length sequence:
(2,-3)(0,5)(0,1)(1,-2)(4,2)
(0,-4)(0,3)(0,-2)(3,1)(2,-2)

Symbol s_i	p_i	l_i	**Codeword**
0	0.28200	2	11
1	0.27860	2	10
2	0.14190	3	011
3	0.13890	3	010
4	0.05140	4	0011
5	0.05130	4	0010
6	0.01530	5	00011
7	0.01530	5	00010
8	0.00720	6	000011
9	0.00680	6	000010
10	0.00380	7	0000011
11	0.00320	7	0000010
12	0.00190	7	0000001
13	0.00130	8	00000001
14	0.00070	9	000000001
15	0.00040	9	000000000

JPEG: DCT + Quantization + Entropy Coding



①

DCT

②

994	-34	20	38	22	0	5	-1
-49	17	8	-17	-4	-1	4	-2
149	35	-37	-15	1	3	-1	-4
58	8	-32	-3	-9	-11	2	-4
-27	5	26	6	9	7	8	0
9	41	53	8	0	1	-1	2
0	2	-1	-6	-2	-3	0	1
3	-14	-12	2	-1	-7	-5	-1

QUANTIZATION
(Step size = 30)

③

33	-1	1	1	1	0	0	0
-2	1	0	-1	0	0	0	0
5	1	-1	0	0	0	0	0
2	0	-1	0	0	0	0	0
-1	0	1	0	0	0	0	0
0	1	2	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

IQ, IDCT

④

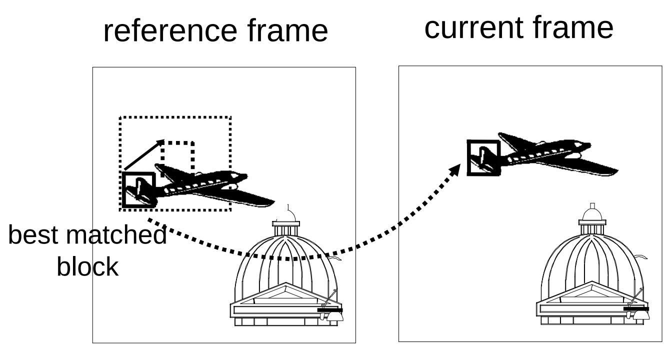
Mean Square Error = 38.75

H.261

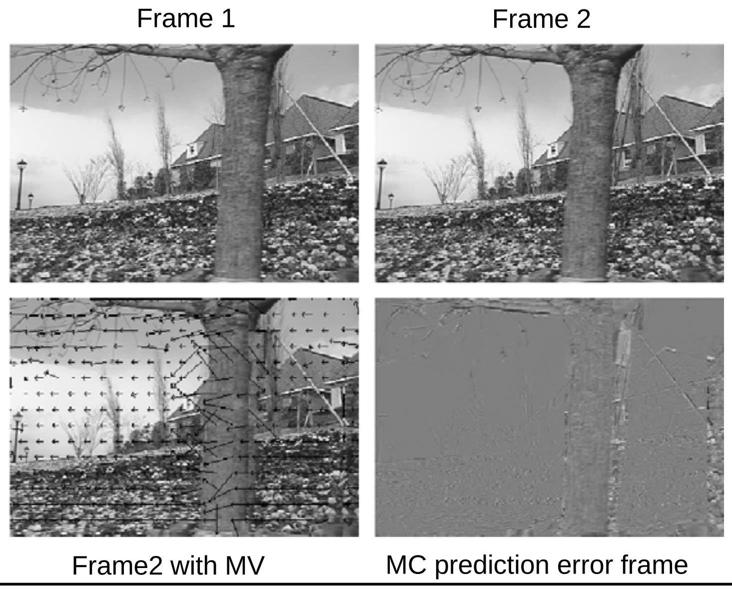
– Initial driving application: ISDN videoconferencing
– New technology used:
  Motion compensated predictive coding

H.261

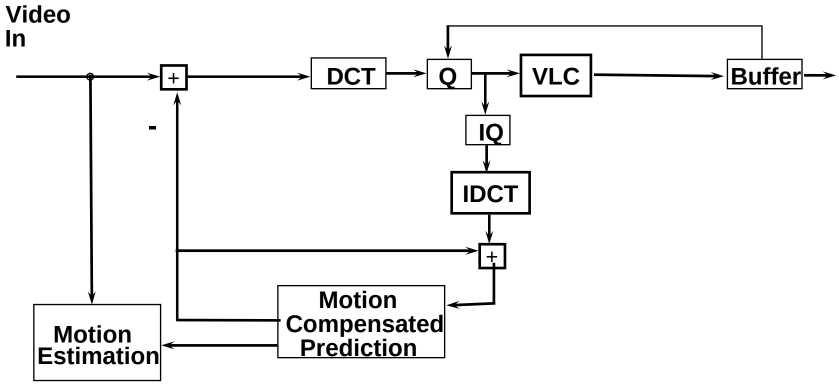
JPEG + Motion Compensated Predictive Coding (MCPC)



Motion Compensated Predictive Coding



H.261 Video Codec



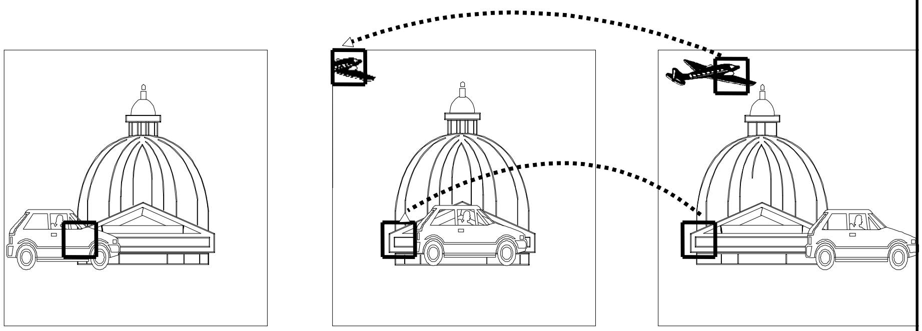
VLC: Variable-Length Coding

MPEG-1

– Initial driving application: Video CD
– New concepts:
Bi-directional motion estimation with half-pixel accuracy

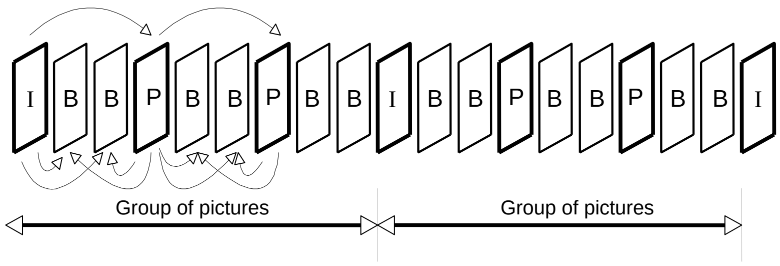
MPEG-1

H.261 + Half-pixel bi-directional MCPC



Frame N-1 Frame N Frame N+1

Group of Picture



I = **Intra-Picture** Coding, allow random access, for reference
P = **Predictive** coding, causal prediction only, can be referenced
B = **Bi-directional** coding, noncausal prediction, never referenced

MPEG-2

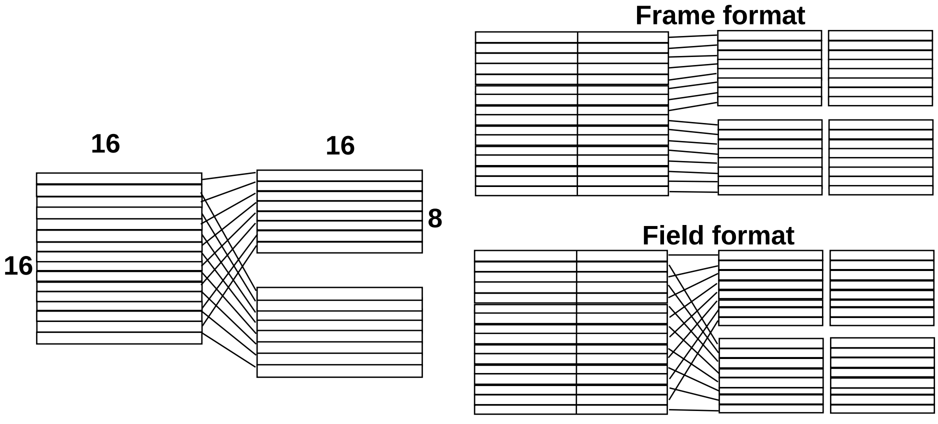
– Initial driving application: Video-On-Demand over HFC (Hybrid Fiber Coaxial) networks
– Main new concepts:
  * Frame-field adaptive DCT and Motion Estimation
  * Profiles and Levels
  * Scalable coding
- The most successful video standard for high quality applications (DVD, Digital TV, HDTV, ...)
- Not suitable for low-bit-rate applications due to the overheads

Profiles and Levels

Level	Profile				
	Simple	Main	SNR Scalable	Spatially Scalable	High
High 1920x1152 (60 frames/s)		62.7 Ms/s 80 Mbit/s			100 Mbit/s for 3 layers
High-1440 1440x1152 (60 frames/s)		47 Ms/s 60 Mbit/s		47 Ms/s 60 Mbit/s for 3 layers	80 Mbit/s for 3 layers
Main 720x576 (30 frames/s)	10.4 Ms/s 15 Mbit/s	10.4 Ms/s 15 Mbit/s	10.4 Ms/s 15 Mbit/s for 2 layers		20 Mbit/s for 3 layers
Low 352x288 (30 frames/s)		3.04 Ms/s 4 Mbit/s	3.04 Ms/s 4 Mbit/s for 2 layers		

* numbers in the table are maximum allowed

Frame/Field Adaptive Coding



H.263

- Initial driving application: Videophone / Wireless video
- New technologies: 4 optional modes
- Suitable for low-bit-rate applications due to the optimized overheads
- Current driving applications: Internet / Wireless Video

Four Negotiable Options

- Annex D: Unrestricted Motion Vector

  Motion vectors can point outside the picture, maximum range for motion vectors: -31,5 to 31.5 instead of -16 to 15.5 (useful in case of higher resolution format and movement along the edge of the picture and camera movement)

- Annex F: Advanced Prediction Mode

  8x8 motion vectors, overlapped block motion compensation (OBMC), motion vector can point outside the picture

- Annex E: Syntax-based Arithmetic coding

  About 5% decrease in bit-rate

- Annex G: PB-frame mode

H.263+ / H.263 Version 2

- Annex D: New Unrestricted Motion Vector Mode (use RVLC, motion vector range up to +/- 256)
- Annex I: Advanced Intra Coding Mode
- Annex J: Deblocking Filter Mode
- Annex L: Supplemental enhancement information specification
- Annex M: Improved PB-Frame Mode
- Annex O: Temporal, Spatial, and SNR Scalability Mode
- Annex P: Reference Picture Resampling Mode
- Annex Q: Reduced Resolution Update Mode
- Annex S: Alternative Inter VLC Mode
- Annex T: Modified Quantization Mode

Error resilience:
- Annex K: Slice Structured Mode
- Annex R: Independent Segment Decoding Mode
- Annex N: Reference Picture Selection Mode

MPEG-4

– Initial driving application:
  1. Next generation video coding for very low bit-rate video
  2. Object based manipulation
  - New concepts: Object-based coding

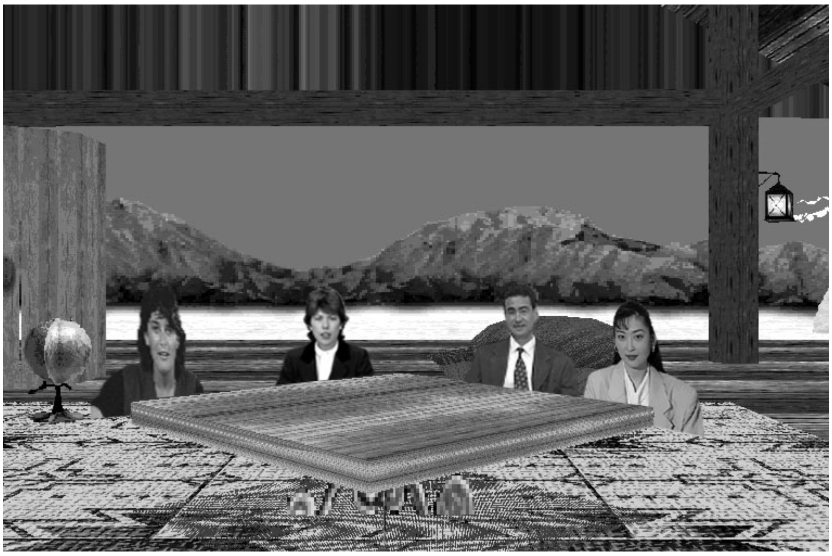
– Current driving application:

MPEG-4 attempts to become *THE* standard for streaming AV media on the Internet and via wireless networks

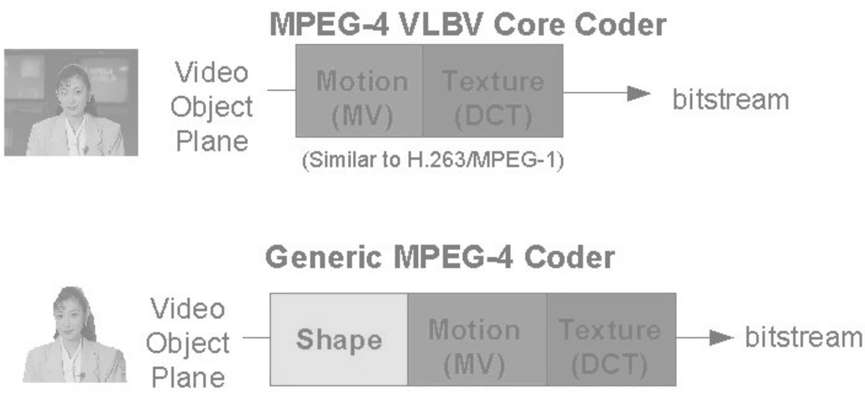
- Error-resilience features, Optimized for low-bit-rate applications, Fine granular scalability coding
- MPEG-4 File Format defines first standard for streaming AV media on the Internet

MPEG-4 Video Object Plane (VOP)

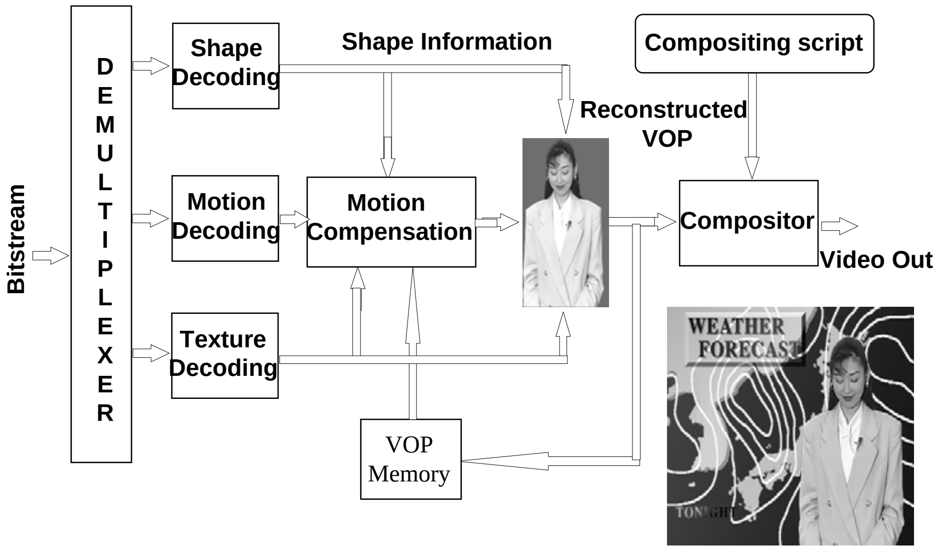
Synthetic and Natural Hybrid Coding (SNHC)



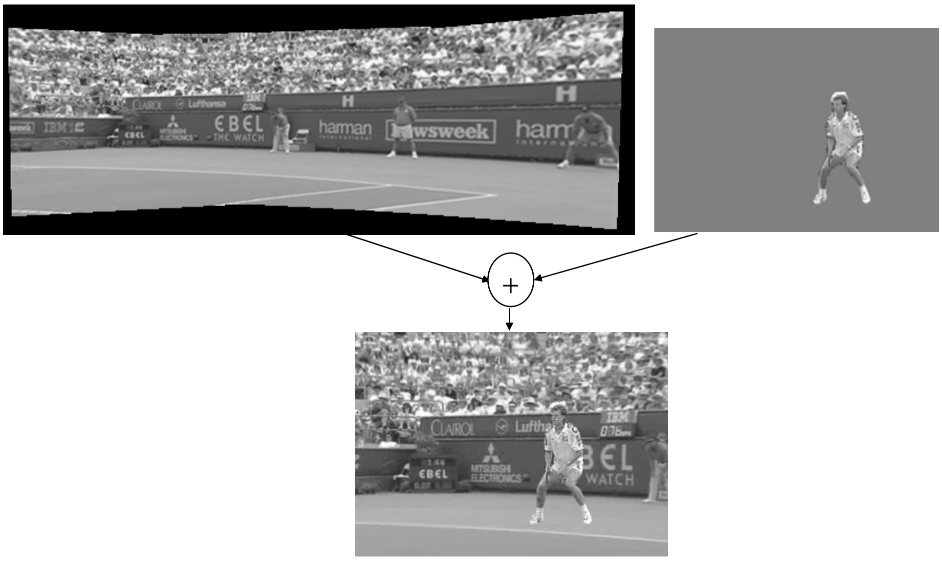
MPEG-4 Texture and Shape Coding



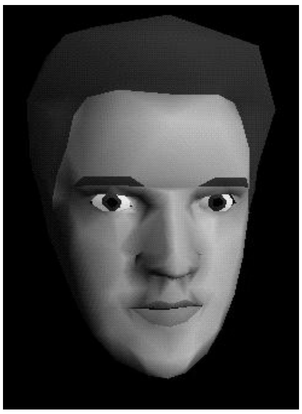
MPEG-4 VOP Decoder

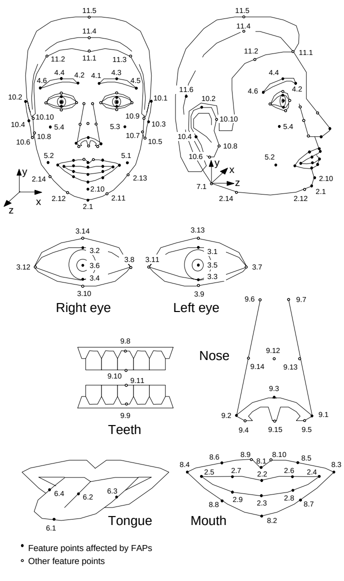


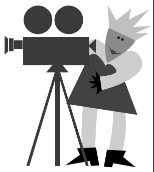
Sprite Coding



MPEG-4 Visual Face Animation

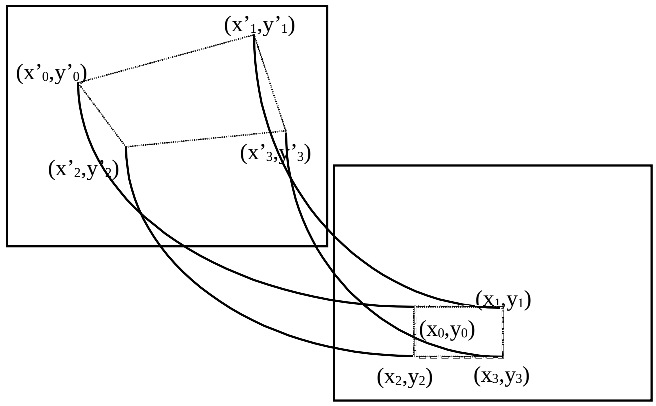




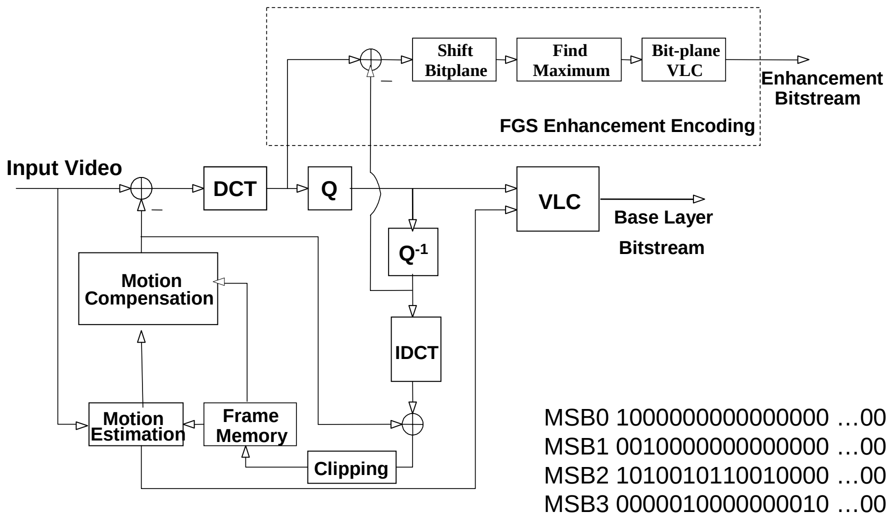


MPEG-4 ASP (Advanced Simple Profile)

Previous techniques + ¼-pixel ME + Global ME

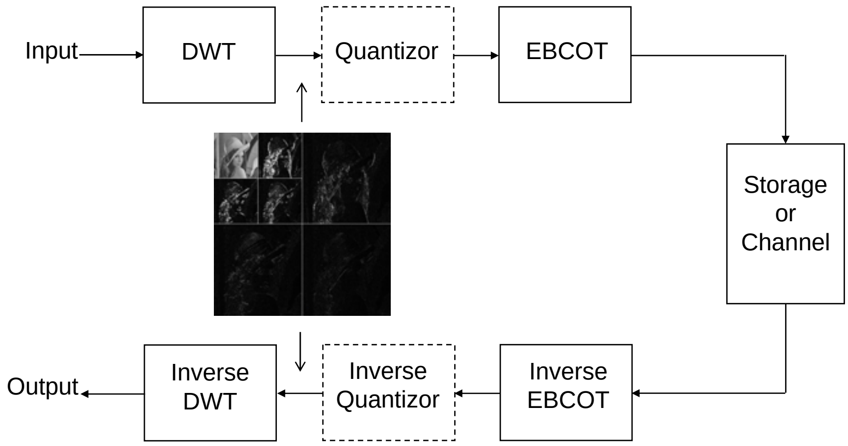


MPEG-4 FGS (Fine Granularity Scalability)



JPEG-2000 vs. JPEG (1/2)

JPEG-2000: Discrete Wavelet Transform (DWT) + Embedded Block Coding with Optimum Truncation (EBCOT)



JPEG-2000 vs. JPEG (1/3)

0.125bpp, JPEG

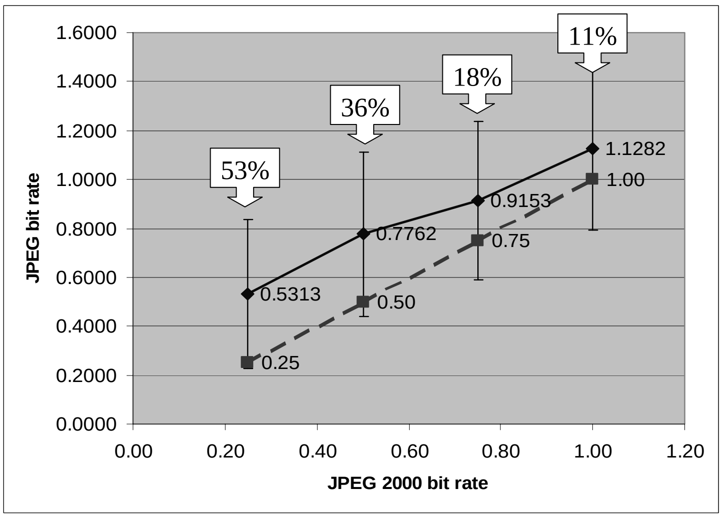
0.125bpp, JPEG-2000

JPEG-2000 vs. JPEG (2/3)

0.25bpp, JPEG

0.25bpp, JPEG-2000

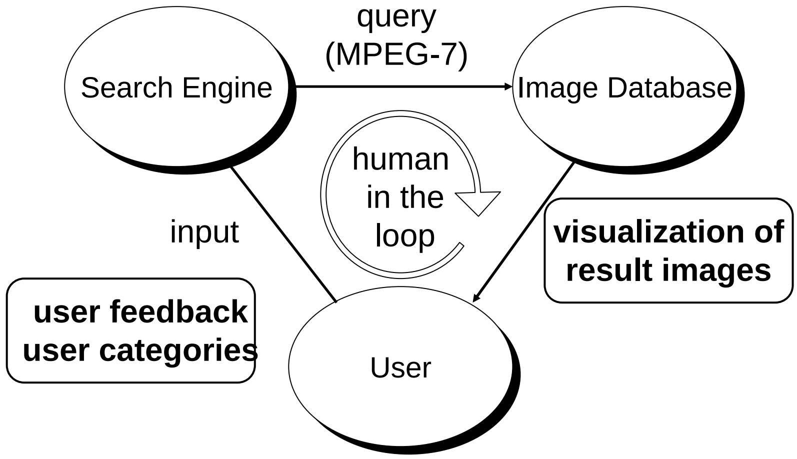
JPEG-2000 vs. JPEG (3/3)



6 observers
6 color images
300 dpi

MPEG-7: Multimedia Content Description Interface

To define standardized solutions that allow users or agents to ***Search, Browse, Filter*** audiovisual content.



MPEG-7: Description of Multimedia Information

How can “Non-Text-Information” be described in a useful way?

Text

Speech- and Audio-Descriptors

- *Pitch*
- *Melody*
- *Frequency*
- *Sex of speaker*

Visual Clues

- ***Texture***
- ***Color***
- ***Form and Shape (2D/3D)***
- ***Motion Activity***
- ***Higher Order Object-Models***
- ***Persons, Gestures***

What MPEG-7 Standardize?

Description Definition Language (DDL)

Flexible description of description schemes or descriptors

- *XML Schema*

Description Schemes

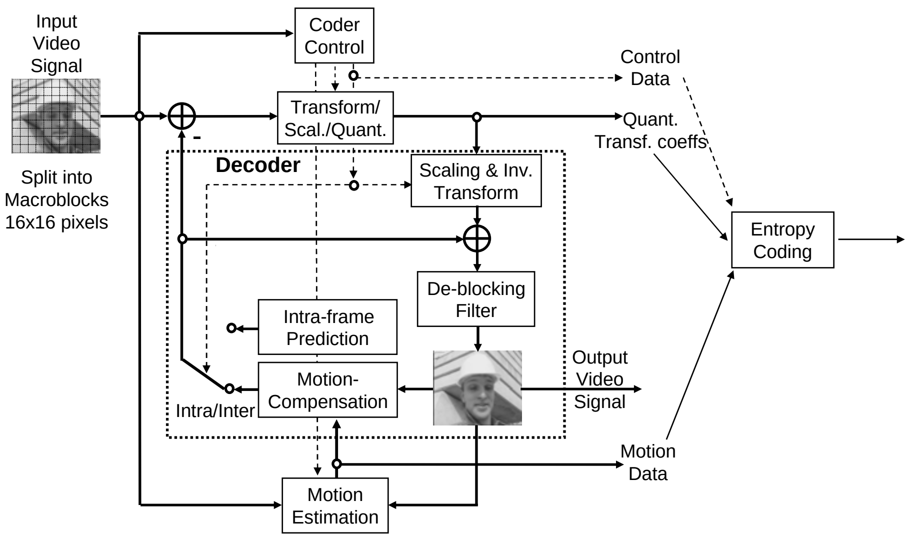
Relations between descriptors

- *shape + motion*
- *shape + frame no*

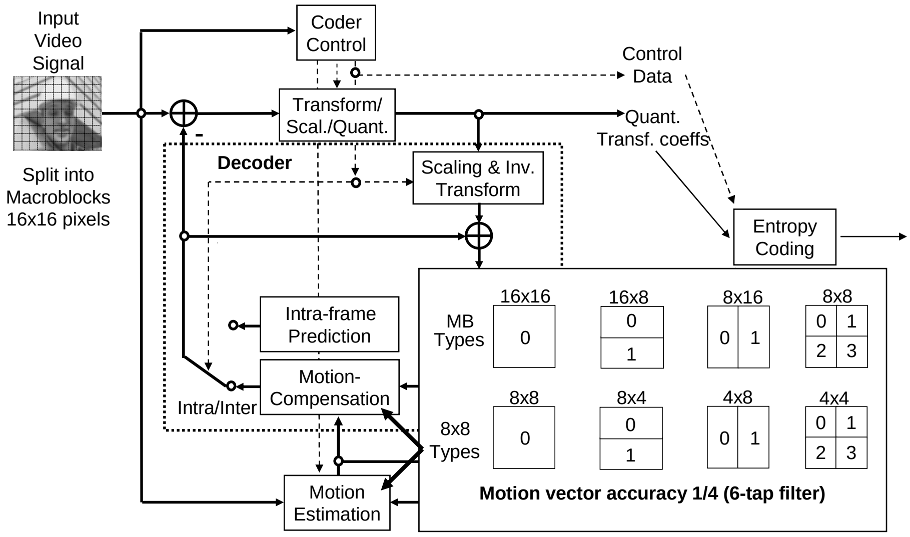
Descriptors

- *Text,*
- *Color, Texture*
- *Shape, Motion*
- *Pitch, Melody*

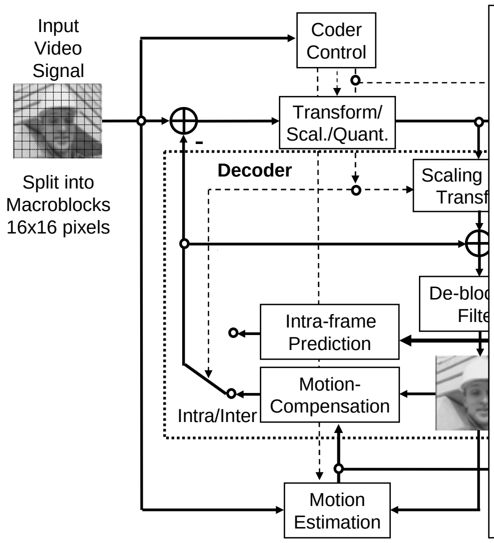
MPEG-4 AVC/H.264: Encoder Architecture



MPEG-4 AVC/H.264: Motion Compensation

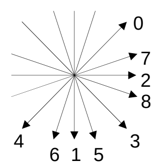


MPEG-4 AVC/H.264: Intra Prediction



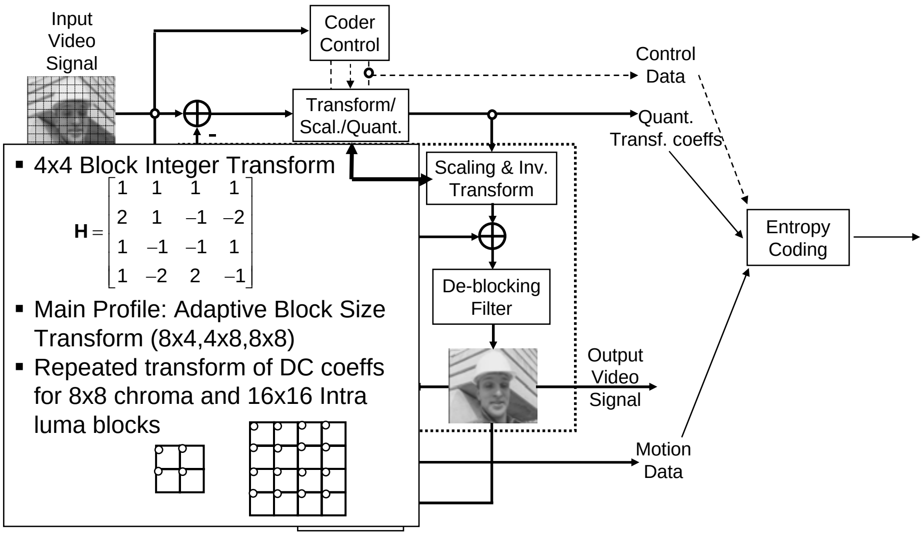
- Directional spatial prediction (9 types for luma, 1 chroma)

```
Q A B C D E F G H
I a b c d
J e f g h
K i j k l
L m n o p
M
N
O
P
```



- e.g., Mode 3: diagonal down/right prediction a, f, k, p are predicted by (A + 2Q + I + 2) >> 2

MPEG-4 AVC/H.264: Transform Coding

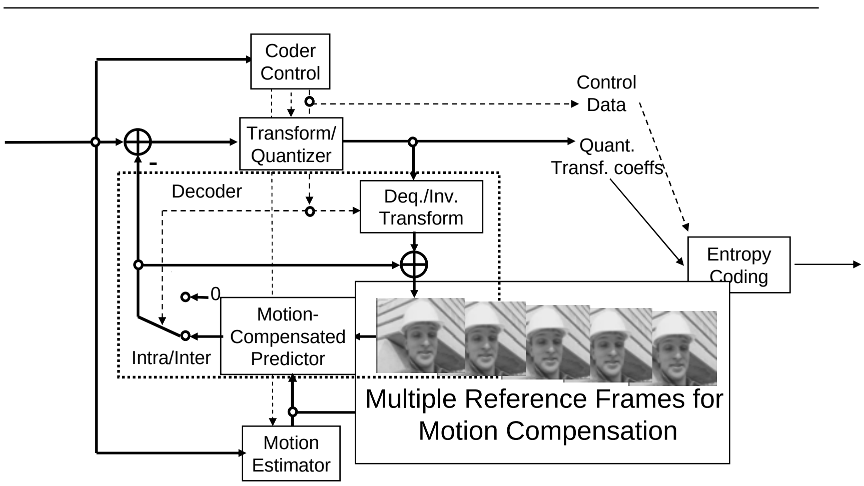


- 4x4 Block Integer Transform

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

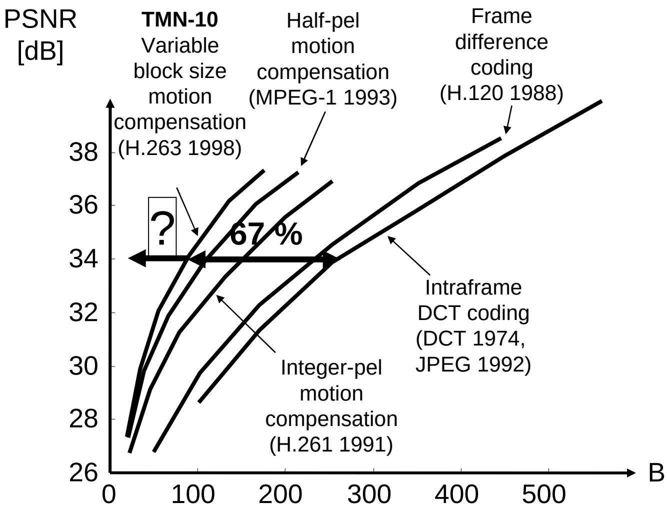
- Main Profile: Adaptive Block Size Transform (8x4,4x8,8x8)
- Repeated transform of DC coeffs for 8x8 chroma and 16x16 Intra luma blocks

MPEG-4 AVC/H.264: Multiple Reference Frames



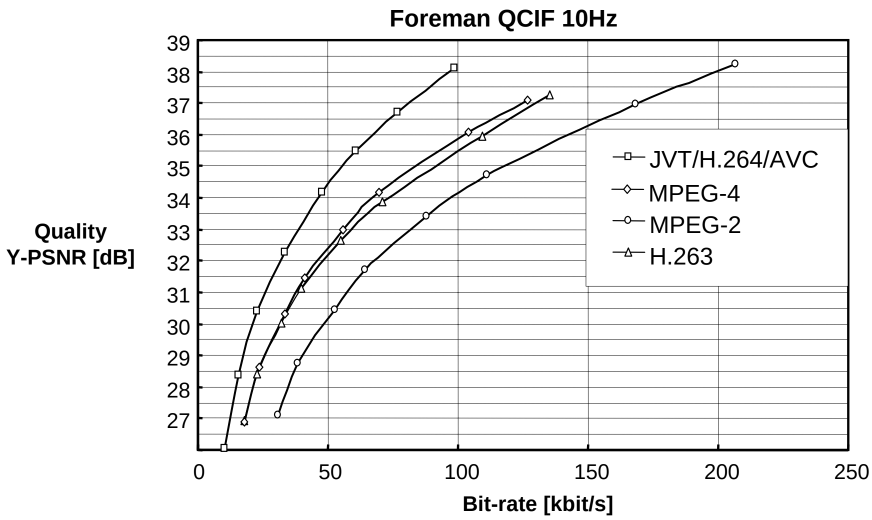
Multiple Reference Frames for Motion Compensation

Coding Efficiency Comparison (1/3)

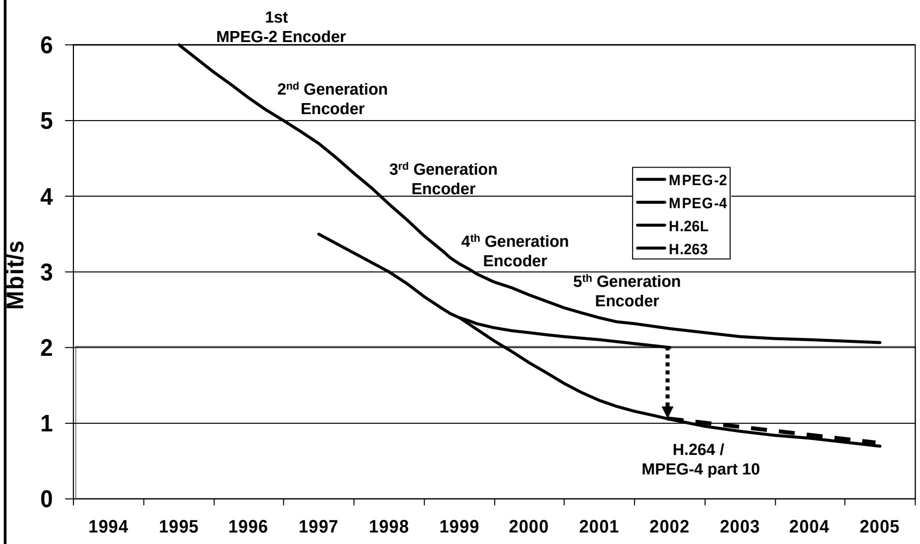


Foreman
10 Hz, QCIF
100 frames encoded

Coding Efficiency Comparison (2/3)



Coding Efficiency Comparison (3/3)



Source: Modulus Video

How were the Improvements Obtained

- It mainly comes from incremental improvements:
  - Better prediction
  - More computation
  - More memory
- No fundamental changes in the basic algorithm (DCT + MCPC)

Is Compression Dead? (1/2)

NO!!!

- Contrary to many people’s thought that the compression has saturated, we actually have steady improvements over the last decade.
- Bandwidth is never sufficient: millions people may use streaming video, wireless video, games, HDTV, digital TV, 3-D TV, virtual reality, enhanced reality (needs multiple views), …
- Every time you increase the bandwidth, people will always figure out innovative ways to fill up the bandwidth.

Is Compression Dead? (2/2)

- New applications need new compression algorithms for optimal performance
- Always need better quality
- As long as there are needs, people will eventually figure out better ways to do it
- Hardware and technology improvements will relax the constraints

What Aspects Can We Explore to Get Better Compression?

- Better prediction
- More memory
- More computation
- More content-aware
- More adaptive processing
- More delay
- Utilize human perception characteristics
- …

Performance evaluation by PSNR?

What Else Besides Compression? (1/2)

- Networking (wired and wireless):

  Scalable coding, Multiple description coding, Transcoding, Statistical multiplexing, Error-resilience coding, Error-concealment, Pre/post processing, Dynamic rate-control, Content-aware delivery, Joint source channel coding, Peer-to-peer streaming, Wireless games, …

What Else Besides Compression? (2/2)

- Asynchronous applications may see more early success: needs authoring, retrieval, and content filtering tools
- Security: advanced video surveillance, face recognition, …
- Content protection: IP Protection, encryption, watermarking, …
- Implementation: for low-cost low-power high-performance devices, media processors