MPEG-4 and H.261/263 Video Compression

Lecture 11

MPEG-4 - SCOPE and Features

- MPEG-4 is an expansion of other video compression algorithms
- MPEG-4 addresses the needs of authors, service providers and end users
- Provides a set of interfaces and methods

MPEG-4 Functions

Content Based Interactivity

Content-Based Multimedia Data Access Tools	Content-based retrieval of information from on-line libraries and travel information databases
Content-Based Manipulation and Bitstream Editing	· Interactive home shopping · Home movie production and editing · Insertion of sign language interpreter or subtitles; · Digital effects (e.g. fade-ins);

MPEG-4 Functions

Content Based Interactivity

Hybrid Natural and Synthetic Data Coding	· Animation and synthetic sound can be composited with natural audio and video in a game. · A viewer can translate or remove a graphic overlay to view the video beneath it; · Graphics and sound can be ‘rendered’ from different points of observation;

MPEG-4 Functions

Content Based Interactivity

Improved Temporal Random Access	· Audio-visual data can be randomly accessed from a remote terminal over limited capacity media. · A ‘fast forward’ can be performed on a single AV object in the sequence.

MPEG-4 Functions

Compression

Improved Coding Efficiency	· Efficient transmission of audio-visual data on low-bandwidth channels. · Efficient storage of audio-visual data on limited capacity media, e.g. magnetic disks.
Coding of Multiple Concurrent Data Streams	· Multimedia entertainment, e.g. virtual reality games, 3D movies; · Training and flight simulations; · Multimedia presentations and education;

MPEG-4 Functions

- Universal Access

Robustness in Error-Prone Environments	· Transmitting from a database over a wireless network; · Communicating with a mobile terminal. · Gathering audio-visual data from a remote location.
Content-Based Scalability	· User or automated selection of decoded quality of objects in the scene; · Database browsing at different content levels, scales, resolutions, and qualities;

MPEG-4 Functions

- Synchronization - the ability to synchronize the audio, video, and other content data for presentation;
- Auxiliary data capability - the ability to allocate a channel for binary data bitstreams;
- Virtual channel allocation flexibility - the ability to dynamically reallocate video, audio, or data channels;

MPEG-4 Functions

- Low (end-to-end and/or decoder) delay mode - the ability for the system, audio, and video codecs to operate with low delay;
- User controls - the ability to support user controls for interactive operations;
- Transmission media interworking - the ability to operate in various media;

MPEG-4 Functions

- Interworking with other audio-visual systems - the ability to interact with various terminal types;
- Security - the ability to provide encryption, authentication, and key management;
- Multipoint capability - the ability to have multiple sources and/or destinations;

MPEG-4 Functions

- Content - the ability to code various types of visual scenes and audio content (high and intermediate quality audio, wideband, narrowband, intelligible, and synthetic speech, and synthetic audio);
- Format - the ability to code various audio and video formats;
- Quality - assessment of the decoded audio or video adequacy;

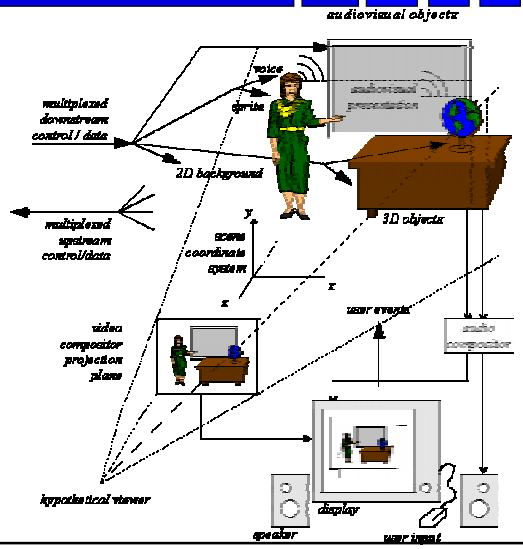
MPEG-4 Functions

- Bitrate modes - low (9.6 - 64 kbit/s), intermediate (64 - 384 kbit/s), and high (384 - 1024 kbit/s);
- Low complexity modes - the cost, with regard to the hardware, firmware, and software, required to implement the system.

MPEG-4 Standardized Scene Description

- Place media object anywhere
- Apply Transforms
- Group media objects
- Apply stream data to media objects
- Change, interactively, the users viewing and listening points

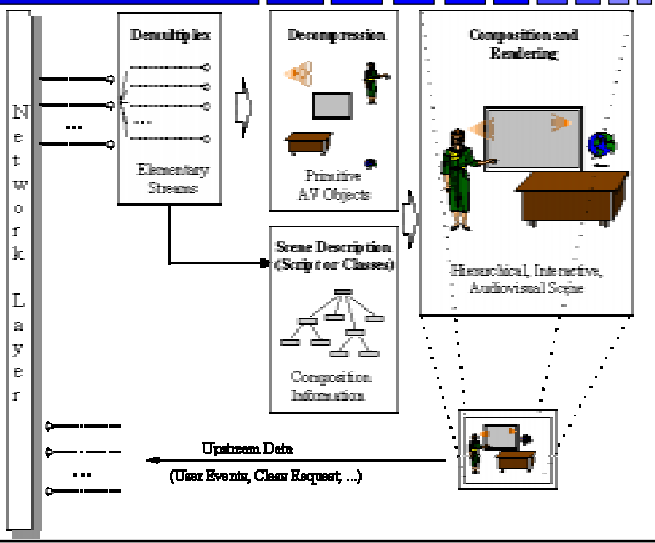
MPEG-4 Standardized Scene Description



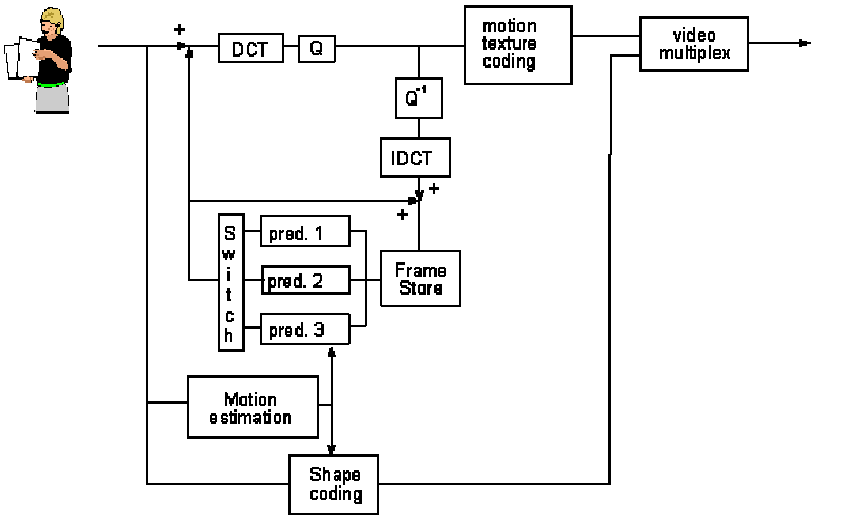
MPEG-4 Terminal (Receiver Side)

- MPEG-4 describes how to multiplex different streams of audio-video data into an complete presentation.
- Elementary Streams are parse and passed to decoders
- AV objects are created and rendered
- User inputs are sent back to network process

MPEG-4 Terminal (Receiver Side)



MPEG-4 Video Coder



MPEG-4 Video Coder

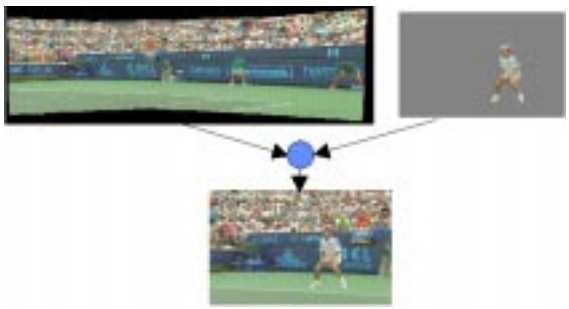
- Shape Coding for arbitrarily shaped Video Objects (VO)
- Motion Compensation as well as DCT based texture coding (8x8 DCT)
- Method allows the use of special object based motion prediction tools for each object in scene.
- Microsoft Net Show is the same verification model as MPEG-4

MPEG-4 Sprite Coding

- Analyze the video stream to find the static background
- Create a still image of the background
- Code the moving objects against the background

MPEG-4 Sprite Coding

- 8 global motion parameters describing camera motion are coded for each sequence
  - Represent an affine transform of the sprite from the first frame



MPEG-4 Status

- Committee Draft of the standard - released in 1998.
- Committee Draft:
  - 14496-1 Systems
  - 14496-2 Video
  - 14496-3 Audio
  - 14496-4 Conformance Testing
  - 14496-5 Reference Software
  - 14496-6 Delivery Multimedia Integration Framework

MPEG-4 Version 1 Work Plan

Part	Title	WD	CD	FCD	DIS	IS
1	Systems		97/11	98/3	98/10	99/02
2	Visual		97/11	98/3	98/10	99/02
3	Audio		97/11	98/3	98/10	99/02
4	Conformance Testing	97/10	98/12	99/07	99/12	00/02
5	Reference Software		97/11	98/3	99/03	00/05
6	Delivery Multimedia Integration Framework (DMIF)	97/07	97/11	98/3	98/10	99/02

Video Conferencing Coding Standards

- Video Conferencing standards tie together several technologies:
  - Video Compression
  - Audio Compression
  - Packet Communications Protocol
- Video Conferencing must adjust to different data rates, and multiple (conflicting) requirements.

Packet Communications Protocol

- Integrated Services Digital Networks (ISDN)
  - Basic Rate ISDN video coding at 112 kbits/s and speech coding at 16 kbits/s
  - H0 ISDN video coding at 320 kbits/s and audio at 64 kbits/s
- Coding of motion video at ISDN rates require conditional replenishment, and high data compression rates (below 1 bit per moving-area pel).

Basic Compression Techniques

- Code blocks of pixels using a transform (DCT)
- Variable Length Code of coefficients
- Additional efficiency can be achieved by using motion compensation.
  - subtract prediction from the moving-area pel values prior to transformation.
  - Adaptive quantization

H.261 Video Compression

- CCITT standard for videoconferencing
- Compresses video to 64Kbits to 2.048 Mbits (standard document says 1.92 Mbits
- Subsampling of Components
  - CIF resolution Y=352x288
    - » CIF -> Common Intermediate Format
  - QCIF resolution Y=176x144
- Motion Compensated Differential DCT Algorithm
- Group of Blocks, Macroblocks, Blocks

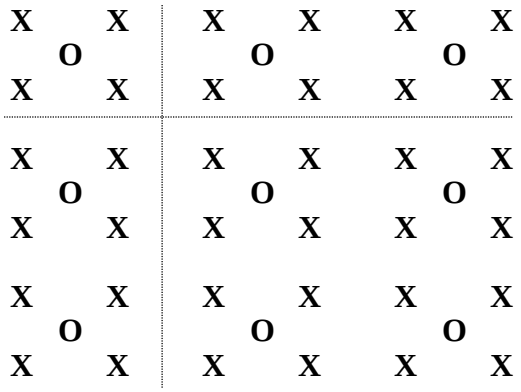
Video Formats

	CIF	QCIF
Luminance Pels Per Line	352	176
Chrominance Pels per line	176	88
Luminance Lines per field	288	144
Chrominance Lines per field	144	72
Fields Per Second	29.97	29.97

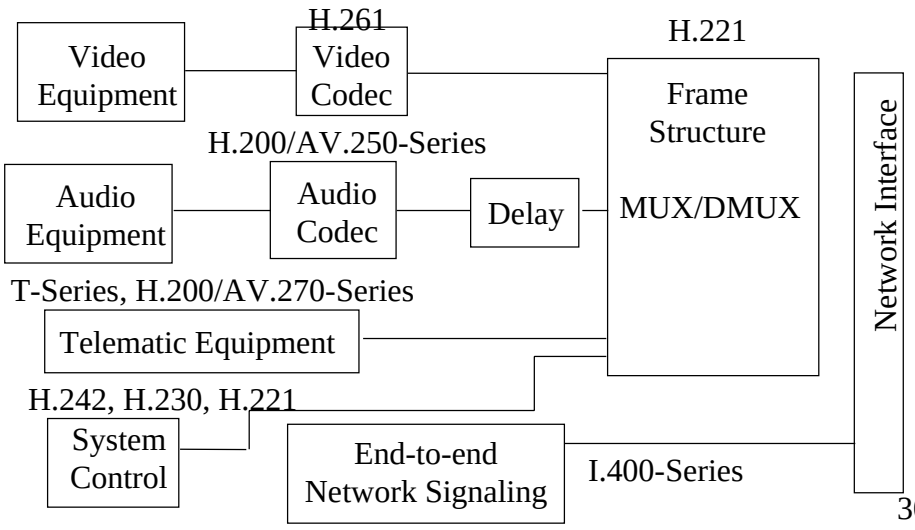
Video Formats

	CIF	QCIF
Interlace	1:1	1:1
Color Components	Y Cb Cr	Y Cb Cr
Luminance Range	16-235	16-235
Chrominance Range	16-240	16-240
Zero Color Difference Level	128	128
Bits per Pel	8	8

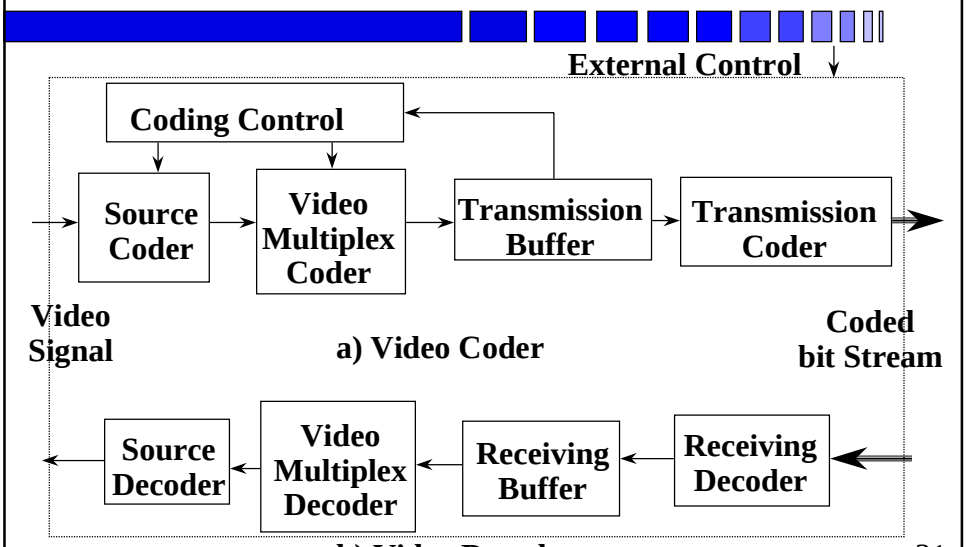
Positioning Of Luminance And Chrominance Samples



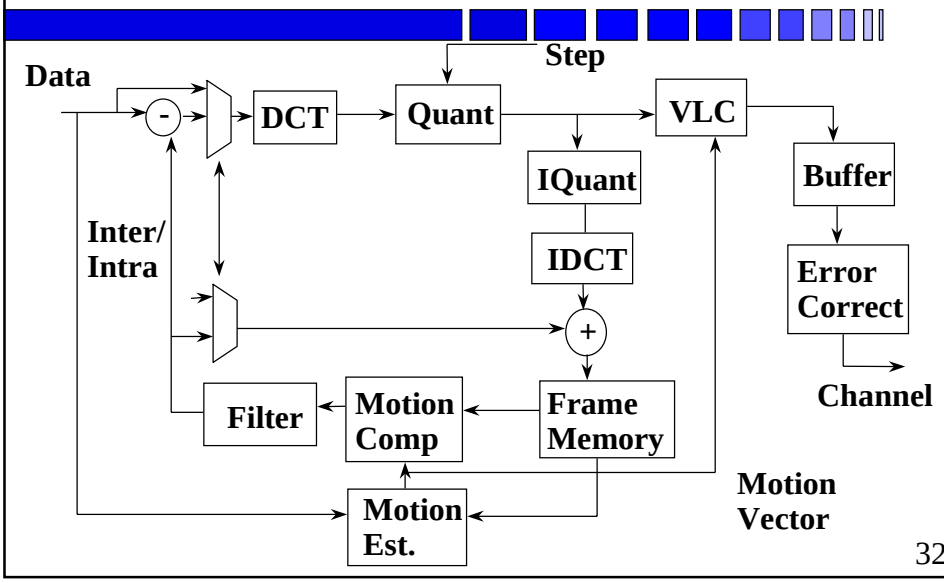
Generic Visual Telephone System



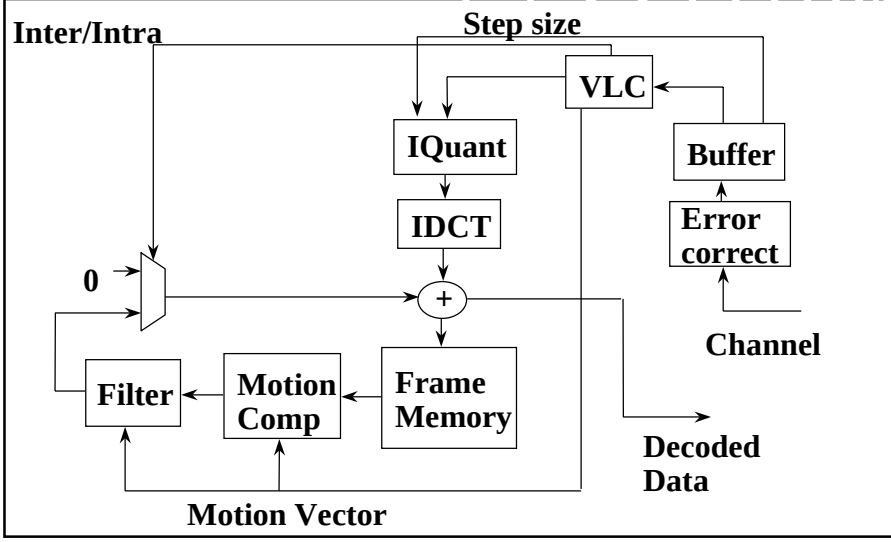
H.261 Video Codec Outline



H.261 Encoder



H.261 Decoder



Common Intermediate Format (CIF)

- CIF represent half the active lines of a 625/25 television signal and the picture rate of a 525/30 NTSC signal

Color Component	Image size (pixels x lines)
Y	352 x 288
Cb	176 x 144
Cr	176 x 144

H.261 DCT

- Coding path similar to JPEG and MPEG
  - Video frame first translated into CIF frame and stored in frame memory
  - Noise filtering and other signal processing done at this stage (sensor dependent)
  - DCT operates on 8x8 picture blocks
  - Four lumance (Y) blocks, and one Cb and one Cr color difference blocks combined to form a macroblock

H.261 Frame Prediction

- Similar to MPEG-1
  - Exception: only I-pictures and P-pictures are used.
- Differential coding allows the DCT coder to operate on either input macroblocks (INTRA mode) or differential macroblocks between the current frame and the prior frame (INTER mode).
  - Not all macroblocks need to be coded and transmitted.
    - » Low bit rate: up to three full frames can be skipped.

Output VLC

- Following DCT, output coefficients are quantized, coded using Variable Length Coder (Huffman) and stored in an output buffer
- Monitoring the status of the buffer, the rate can be controlled dynamically.
  - Consequences of a full buffer is skipping frames (freeze frame).
- An error detection and correction code can be used (optional) to add parity bits to the transmission.

H.261 Motion Estimation/Compensation and Loop Filter

- Optional in standard
- Loop filter minimizes prediction error by smoothing the pixels in the previous frame.
  - Separable 2-D filter that operates on 8x8 pixel blocks
  - 1-D filter is a three tap FIR
    - » At block edges, filter coefficients are 0,1, and 0
    - » Otherwise: 1/4, 1/2 and 1/4

Implementation Issues

- Error on DCT is controlled
- At least one intraframe coded macroblock (base frame) for every 132 interframe coded macroblock
  - For better coding efficiency, perform intraframe coding on a few macroblocks in every picture using a rotational scheme.

H.261 Motion Compensation

- Motion Compensation is optional in the encoder
- Decoder will accept one vector per macroblock
- Both horizontal and vertical components of these motion vectors have integer values not exceeding ±15
  - The vector is used for all four luminance blocks in the macroblock

Motion Estimation and Frame Rate

- Maximum range in standard is [-15, 15]
- H.261 can operated at various bit rates (64kbits/s to 1,984 kbits/s)
- At high frame rates, temporal distance between frame is smaller, can afford to have smaller search region
- Low frame rates, the situation is reversed, need larger search region for better quality

Application Derived Search Region

- Videoconferencing deals with limited movement scenes (head and shoulders)
- Smaller, diamond-shaped search regions instead of rectangular shaped searched regions work with no noticeable loss in video quality.
- Graph on next chart shows quality versus frame rate

Search Region Trades

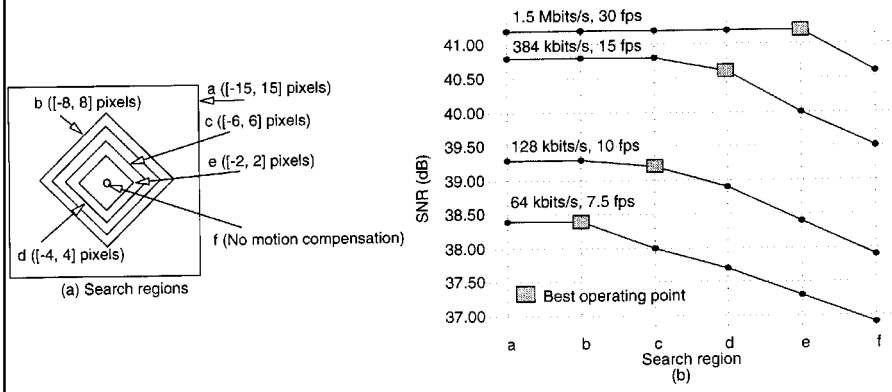
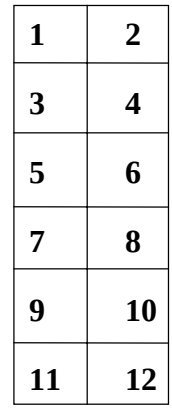


Figure 7.4 Search regions and video quality for typical H.261 encoding. (a) Diamond-shaped search regions; (b) video quality for the corresponding search regions.

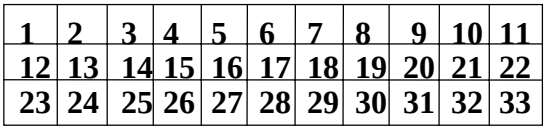
H.261 Motion Compensation

- The motion vector for both color difference blocks is derived by halving the component values of the macroblock vector and truncating the magnitude parts towards zero to yield integer components
  - Positive value of the horizontal or vertical component of the motion vector signifies that the prediction is formed from pels in the previous picture which are spatially to the right or below the pels being predicted

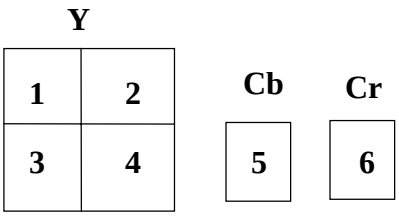
H.261 Video Compression



Group of Blocks



Macroblock Arrangement



Block Arrangement

Compressed Data Stream

- Picture is the top layer
- Each picture is divided into groups of blocks (GOB's)
  - GOB is either 1/12 of a CIF or 1/3 of a QCIF
- GOB is divided into 33 macroblocks
- Each macroblock is six 8x8 blocks (four Y, one Cb, and one Cr)

H.261 Data Structure

- Picture Layer
  - Picture Start Code (PSC) - 20 bit pattern
  - Temporal Reference (TR) - 5 bit input frame number
  - Type Information (PTYPE) - CIF or QCIF selection
  - Spare bits to be defined in later versions

H.261 Data Structure

- GOB Layer
  - Group of Blocks Start Code (GBSC) - 16 bit pattern
  - Group Number (GC) - 4 bit GOB address
  - Quantizer information (GQUANT) - Initial quantizer step size normalized to the range 1 to 31.
    - » At the start QUANT=GQUANT
  - Spare bits to be defined in later versions

H.261 Data Structure

- Macroblock (MB) layer
  - Macroblock address (MBA)
    - » Location of this MB relative to the previously encoded MB inside the GOB.
  - Type information (MTYPE) - 10 types in total
  - Quantizer (MQUANT)
    - » normalized quantizer step size to be used until the next MQUANT or GQUANT. (Range 1 to 31)

H.261 Data Structure

- Macroblock (MB) layer
  - Motion Vector Data (MVD)
    - » differential displacement vector
  - Coded Block Pattern (CBP)
    - » Indicates which blocks in the MB are coded.
    - » Blocks not coded contain zero coefficients.

H.261 Data Structure

- Block Layer
  - Lowest layer is the block layer, consisting of
    - » quantized transform coefficients (TCOEFF),
    - » End of block (EOB) symbol
  - All coded blocks have the EOB symbol.

Types of Coded MB

- Intra - Original Pels are transform Coded
- Inter - Frame difference pels (zero-motion vectors) are coded.
  - Skipped MBs are considered inter by default.
- Inter_MC - displaced (nonzero-motion vectors)
- Inter_MC with filter - displaced blocks are filtered by loop filter.
  - Used for very low bit rates.

H.261 Huffman Code Example

Prediction	MQUANT	MVD	CBP	TCOEFF	VLC
Intra				x	0001
Intra	x			x	0000 001
Inter			x	x	1
Inter	x		x	x	0000 1
Inter+MC		x			0000 0000 1
Inter+MC		x	x	x	0000 0001
Inter+MC	x	x	x	x	0000 0000 01
Inter+MC+FIL		x			001
Inter+MC+FIL		x	x	x	01
Inter+MC+FIL	x	x	x	x	0000 01

Note 1 - "x" means that the item is present in the macroblock
Note 2 - It is possible to apply the filter in a non-motion compensated macroblock by declaring it as MC+FIL but with a zero vector.

H.261 Compression

- Channel Buffer control for constant rate
- Channel Acquisition
- Channel Errors and Error Concealment
- Flexibility in algorithm for originality
- Pre and Post Processing

H.261 Multipoint Facilities

- Freeze Picture request
  - Causes the decoder to freeze its displayed picture until a freeze picture release signal is received or a timeout period of at least six seconds has expired
    - » The transmission of this signal is via external means
- Fast Update Request
  - Causes the encoder to encode its next picture in INTRA mode with coding parameters such as to avoid buffer overflow

Differences Between MPEG and H.261

MPEG	H.261
Uses CIF, SIF, or higher spatial resolutions	Uses QCIF or CIF spatial resolution
Variable image aspect ratio (defined in the header)	Fixed 4:3 aspect ratio
Uses groups of pictures	No notion of GOPs
I, P, and B macroblocks	No B macroblocks
Typical bit rates are around 1.1 Mbits/s	Typical bit rates are around 384 kbits/s. Max. bit rate is 2 Mbits/s.

Differences Between MPEG and H.261

MPEG	H.261
No restrictions on skipped pictures	Only 1, 2, or 3 skipped pictures allowed
Sub-pixel accurate motion vectors	Pixel accurate motion vectors
Typical motion vector range is +/- 15 pixels.	Typical motion vector range is +/- 7 pixels.
The end-to-end coding delay is not critical	Used mostly in interactive appications. End-to-end delay is very critical

Available Videoconferencing Products

Vendor	Name	Codec Speed	Max Frame	Comp. Alg.
BT. North America	Videocodec VC2200 Videocodec VC2100	56 and 112 kbps 56 kbps to 2048 kbps	30 per sec	H.261
GPT Video Systems	System 261 Twin Chan. System 261 Universal	56 and 112 kbps 56 kbps to 2048 kbps	30 per sec	H.261
Compres. Labs.	Rembrandt II/VP	56 kbps to 2048 kbps	30 per sec	H.261, CTX CTX Plus

Available Videoconferencing Products

Vendor	Name	Codec Speed	Max Frame	Comp. Alg.
NEC America	Visual Link 5000 M20 Visual Link 5000 M15	56 and 384kbps 56 kbps to 2048 kbps	30 per sec	H.261, NEC proprietary
PictureTel Corp	System 4000	56 kbps to 768 kbps	10 per sec mono	H.261, SG3 SG2/HVq
Video Telecon	CS350	56 kbps to 768 kbps	15 per sec	H.261, Blue Chip

Low Data Rate Products

Product	Data Rate	Compression Algorithm
AT&T Videophone 2500	16.8/ 19.2 kbps	MC DCT 10 frames/s (max)
British Telecom/Marconi Relate 2000 Videophone	9.6/ 14.4 kbps	H.261 like 7.5 (3.75) frames/s
COMTECH Labs. STU-3 Secure Videophone	9.6 kbps	MC DCT QCIF resolution
Sharevision	14.4 kbps	MC DCT

H.263 - Video Coding for Very Low Bit Rate

Picture Format	number of pixels for luminance (dx)	number of lines for luminance (dy)	number of pixels for chrominance (dx/2)	number of lines for chrominance (dy/2)
sub-QCIF	128	96	64	48
QCIF	176	144	88	72
CIF	352	288	176	144
4CIF	704	576	352	288
16CIF	1408	1152	704	576

Differences Between H.261 and H.263

- Include various video formats such as sub-QCIF, 4CIF and 16CIF
- Advanced prediction Mode: half-pixel motion estimation, median-based VM prediction, 4MV's per macroblock, and overlapped block MC
- Unrestriced MV mode: when MV points outside the picture area, use edge pixels.

Differences Between H.261 and H.263

- A syntax-based arithmetic coding (SAC) mode is possible to change the given VLC tables
- PB-frames mode (forward and bi-directional prediction): similar to those in MPEG
- Weighted quantizer matrix for B-blocks

Differences Between H.261 and H.263

- No loop filter; no macroblock addressing (due to 1/2 pixel ME)
- 1-bit coded or not-coded macroblock information in MB layer (separate coded block patterns for luminance (CBPY) and chrominance (MCBPC) components and for intra/inter mode)

Differences Between H.261 and H.263

- 2-bit differential quantizer information in MB layer and 5-bit quantizer information in picture layer and in GOB layer
- 3D VLC (last-run-level) for coding the transform coefficients
- VLC for B-blocks