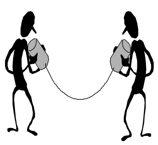
IEEE INFOCOM 2001

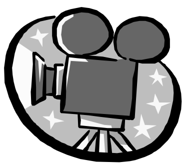
Tutorial T5

Video Over IP

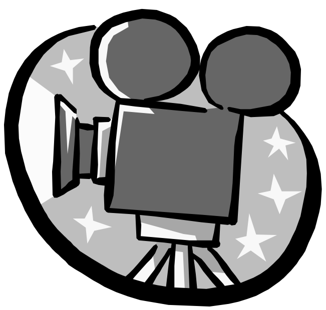
Magda El-Zarki (University of California at Irvine)

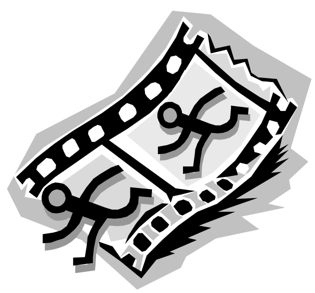
Monday, 23 April, 2001 - Morning

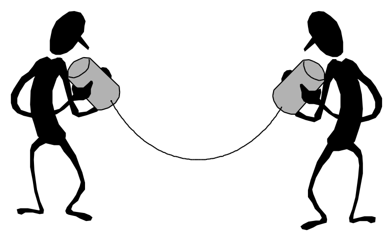




MPEG-4 over IP - Part 1



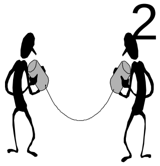


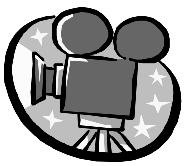


Magda El Zarki
Dept. of ICS
UC, Irvine
elzarki@uci.edu

Outline of Tutorial

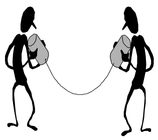
1. Part 1:
   1. Overview of Video Compression
   2. The MPEG suite
   3. Video Quality
2. Part 2:
   1. MPEG-4
3. Part 3:
   1. MPEG-4 Delivery over IP
   2. Conclusions

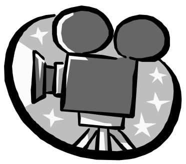




1. Video Compression: Goal

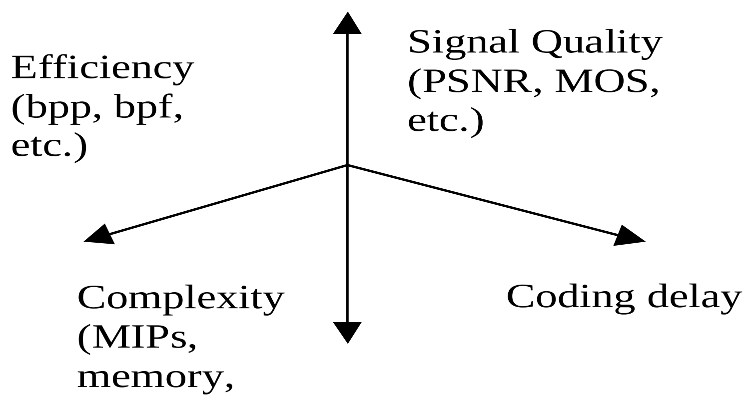
- Goal of video compression is to minimize the bit rate in the digital representation of the video signal while:
  - Maintaining required levels of signal quality
  - Minimizing the complexity of the codec
  - Containing the delay

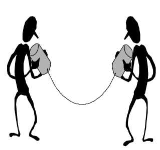


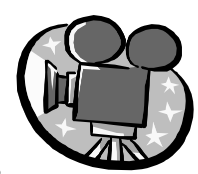


1. Video Compression: Tradeoffs

- The choice of a compression method involves a tradeoff along the following 4 dimensions:

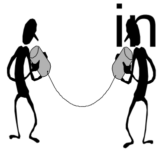


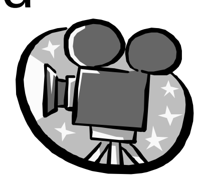




1. Video Compression: Why?

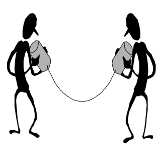
- Video signals are amenable to compression due to the following factors:
  - Spatial correlation: correlation among neighboring pixels
  - Spectral correlation: color images
  - Temporal correlation: correlation among pixels in different frames
- There is considerable irrelevant (from a perceptual viewpoint) information contained in video data.

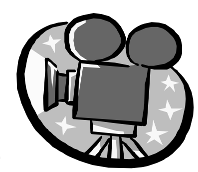




1. Video Compression: Lossless Coding

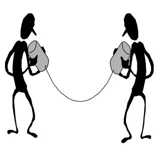
- Lossless coding is a reversible process - perfect recovery of data -> before and after are identical in value. Used regardless of media's specific characteristics. Low compression ratios.
  – Example: Entropy Coding
    - data taken as a simple digital sequence
    - decompression process regenerates data completely
    - e.g. run-length coding (RLC), Huffman coding, Arithmetic coding

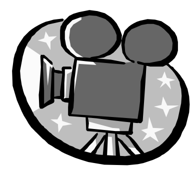




1.Video Compression: Lossy Coding

- Lossy coding is an irreversible process - recovered data is degraded -> the reconstructed video is numerically not identical to the original. Takes into account the semantics of the data. Quality is dependent on the compression method and the compression ratio.
  - Example: Source Coding
    - degree of compression depends on data content.
    - E.g. content prediction technique - DPCM, delta modulation



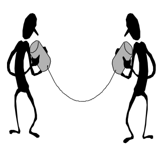


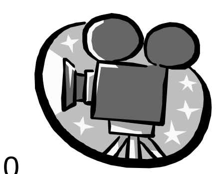
1. Video Compression: Hybrid Coding

- Used by most multimedia systems
  - combines entropy with source encoding
  - E.g. JPEG, H.263, MPEG-1, MPEG-2, MPEG-4

1. Video Compression: Degree/Ratio

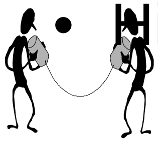
- The degree of compression achieved depends on the quality of the input data:
  - Acquisition noise (timing jitters, poor A/D, etc.)
  - Sampling (affects spatial and temporal correlation)
  - Bits per pixel

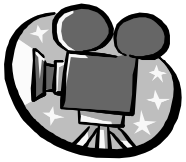




1. Video Compression: Design Choices

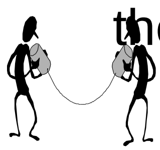
- Lossless or lossy or both
- Compression ratio
- Variability in compression ratio (fixed or variable)
- Resilience to transmission errors
- Complexity tradeoffs in codec
- Nature of degradations
- Hierarchical representation

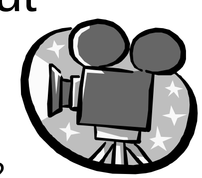




1. Video Compression: Spatial - Block Coding

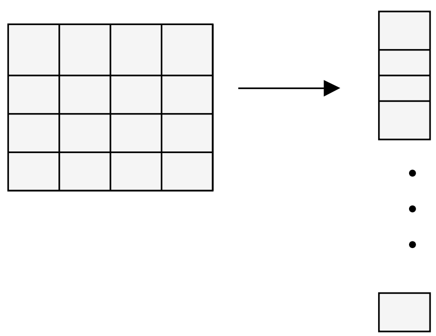
- Image is decomposed into blocks of 8x8 pixels
- Pixel values range between 0 - 255 (8 bits per pixel). These values are shifted to -128 - 127 (centered around zero)
- Discrete Cosine Transform (DCT) maps the spatial data to the frequency domain (64 coefficients). The (0,0) coeff. represents the DC value
- The DCT coefficients are then quantized (Q= 1 - 255) to reduce their spread (essentially zeroing out the higher frequencies)

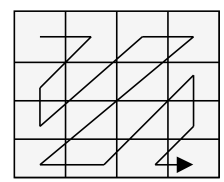




1. Video Compression: Block -> Vector

- A zigzag pattern is then used to scan the block to create a 64 element vector in preparation for entropy coding.

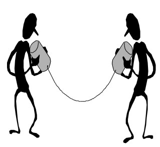


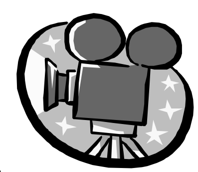


Zig zag scan

1. Video Compression: Run Length Coding

- Run Length Coding (RLC) replaces the sequence of same consecutive bytes with the number of occurrences.
- Variation - Zero Suppression: Used to encode long binary bit strings containing mostly zeros.
  - Each k-bit symbol tells how many 0's occurred between consecutive 1's.
  - e.g. 0000000 - max. runs of zeros to be encoded is 7.
    - xxx (3 bit symbol)
  - e.g. 000100000001101 (using 3 bit symbol)
    - 011 111 000 001 (3-7-0-1 zeros between 1s)



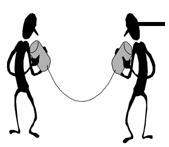


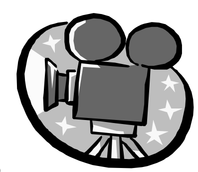
1. Video Compression: Temporal - Macro Blocks

- 16x16 pixels (2x2 blocks) are used as the basic unit for motion prediction - matches based on macro-blocks
- Search window can be any size - larger window better motion estimation but higher computation cost (longer delays)

1. Video Compression - Standards

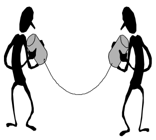
- Broadcast (high bit rate):
  - MPEG-1
  - MPEG-2
- Video Conferencing (low bit rate):
  - H.261
  - H.263
- Interactive (full range of bit rates):
  - MPEG-4

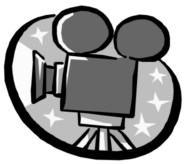




1. Video Compression: H.261

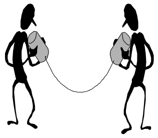
- Designed for video phone and video conference over ISDN
- Bit rate: n x 64kbps, n∈[1,30]
- QCIF (172x144), CIF (352x288)
- Coding Scheme
  - DCT based compression to reduce spatial redundancy (similar to JPEG)
  - Block based motion compensation to reduce temporal redundancy

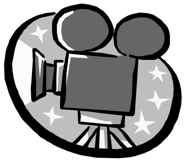




1. Video Compression: H.263

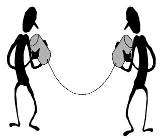
- Designed for low bitrate video applications
- Bit rate: 10 ~ 384kbps
- SQCIF (128x96) ~ 16CIF (1408x1152)
- Coding similar to H.261 but more efficient

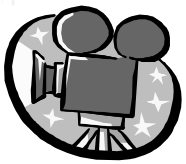




1. Video Compression: H.261 vs. H.263

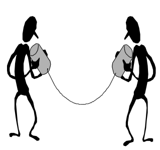
Picture format	Luminance pixels	Luminance lines	H.261 support	H.263 support	Uncompressed bitrate (Mbit/s)			
					10 frames/s		30 frames/s	
					Grey	Colour	Grey	Colour
SQCIF	128	96		Yes	1.0	1.5	3.0	4.4
QCIF	176	144	Yes	Yes	2.0	3.0	6.1	9.1
CIF	352	288	Optional	Optional	8.1	12.2	24.3	36.5
4CIF	704	576		Optional	32.4	48.7	97.3	146.0
16CIF	1408	1152		Optional	129.8	194.6	389.3	583.9

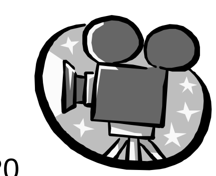




1. Video Compression: MPEG-1

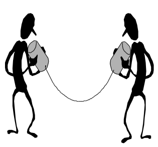
- Designed for storage/retrieval of VHS quality video on CD-ROM
- Bit rate: ~1.5Mbps
- Similar Coding scheme to H.261 with:
  - Random access support
  - Fast forward/backward support

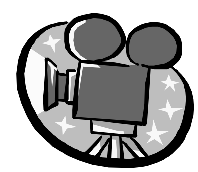




1. Video Compression: MPEG-1 vs H.261

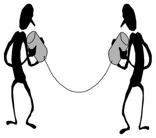
MPEG	H.261
CIF and higher resolution	QCIF & CIF
Variable image aspect ratio defined in header	Fixed 4:3
Use GOP	No GOP
I,P and B macro-blocks	No B macro-blocks
Typically 1.1Mbps	Typically 384kbps
No restrictions on skipped picture	Only 1,2, or 3 skipped pictures allowed
Sub pixel motion vector	Pixel accurate motion vector
Motion vector range ±15 pixels	Motion vector range ±7 pixels
coding delay is not critical	End-to-end delay is very critical

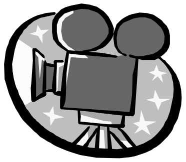




1. Video Compression: MPEG-2

- Designed for broadcast quality video storage and transport
- HDTV support
- Bit rate: 2Mbps or higher (CBR/VBR)
- Two system bit streams: Packet Stream & Transport Stream
- Used for:
  - DVD
  - DirecTV
  - Digital CATV



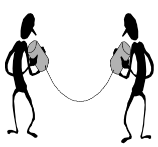


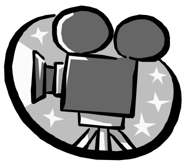
1. Video Compression: Deficiencies of existing standards

- Designed for specific usage
  - H.263 cannot be stored (no random access)
  - MPEG-1 & MPEG-2: not optimized for IP transport
- No universal file format for both local storage and network streaming
- Output cannot be reused efficiently after composition - encoded once, no versatility

1. Video Compression: Requirements for New Standard

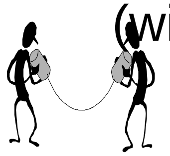
- Efficient coding scheme
  - Code once, use and reuse everywhere
  - optimized for both local access and network streaming
- Works well in both error prone and error free environment
  - Scalable for different bandwidth usage
  - Video format can be changed on the fly
  - Transparent to underlying transport network
- Support efficient interactivity over network

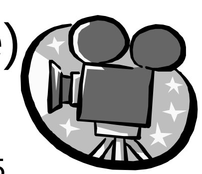




1. Video Compression: The solution: MPEG-4

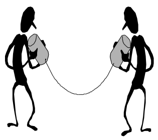
- Internet in the future
  - Not only text and graphics, but also audio and video
- Fast and versatile interactivity
  - Zoom in, zoom out (remote monitoring)
  - Fast forward and fast backward (video on demand)
  - Change viewing point (online shopping, sports)
  - Trigger a series of events (distance learning)
  - On the fly composition
  - Virtual environments
- Support both low bandwidth connections (wireless/mobile) and high bit rates (fixed/wireline)

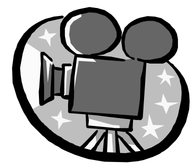




1. Video Compression: What is MPEG-4

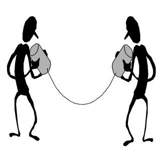
"A coded, streamable representation of audio-visual objects and their associated time-variant data along with a description of how they are combined."

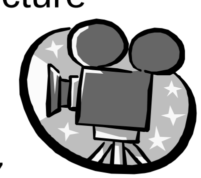




2. MPEG: Overview

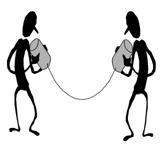
- MPEG exploits temporal (i.e frame-to-frame) redundancy present in all video sequences.
- Two Categories: intra-frame and inter-frame encoding
  - Intra: DCT based compression for the reduction of spatial redundancy
  - Inter: Block-based motion compensation for exploiting temporal redundancy
    - Causal (predictive coding) - current picture is modeled as transformation of picture at some previous time
    - Non-causal (interpolative coding) - uses past and future picture reference

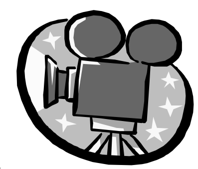




2. MPEG: Motion Representation

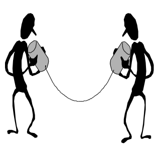
- Predictive and interpolative coding
  – Good compression but requires storage
  – Often makes sense for parts of an image and not the whole image.
- Each image is divided into areas called macro-blocks (motion compensation units)
  – Choice of macro-block size is a tradeoff between gain from motion compensation and cost of motion estimation.
  – In MPEG, each macro-block is partitioned into 16x16 pixels for luminance and 8x8 for each of the chrominance components.

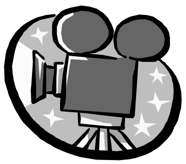




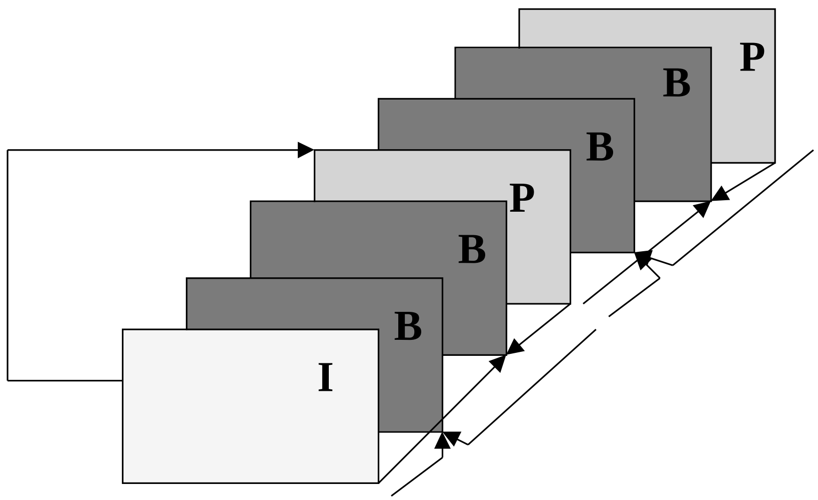
2. MPEG: Video Processing

- MPEG stream includes 3 types of image coding:
  - I-frames - Intra-coded frames - access points for random access, yields moderate compression
  - P-frames - Predictive-coded frames - encoded with reference to a previous I or P frame.
  - B-frames - Bi-directionally predictive coded frames - encoded using previous/next I and P frame, maximum compression
- Motivation for types of frames
  - Fast random access
  - More efficient coding scheme
    - temporal redundancies of both previous and subsequent pictures must be exploited





2. MPEG: Stream Components



2. MPEG: Video Decoding

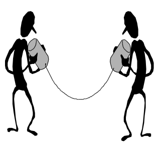
- Using B-frames, the order of images in a MPEG-coded stream differs from the actual decoding order.

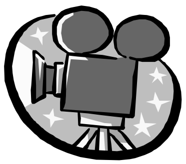
Display Order

Type of frame:	B	B	I	B	B	P	B	B	P
Frame Number:	0	1	2	3	4	5	6	7	8

Decoding Order

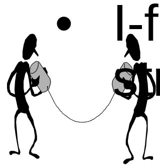
Type of frame:	I	B	B	P	B	B	P	B	B
Frame Number:	2	0	1	5	3	4	8	6	7

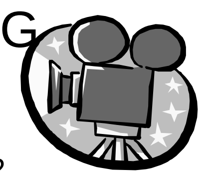




2. MPEG: I-Frames

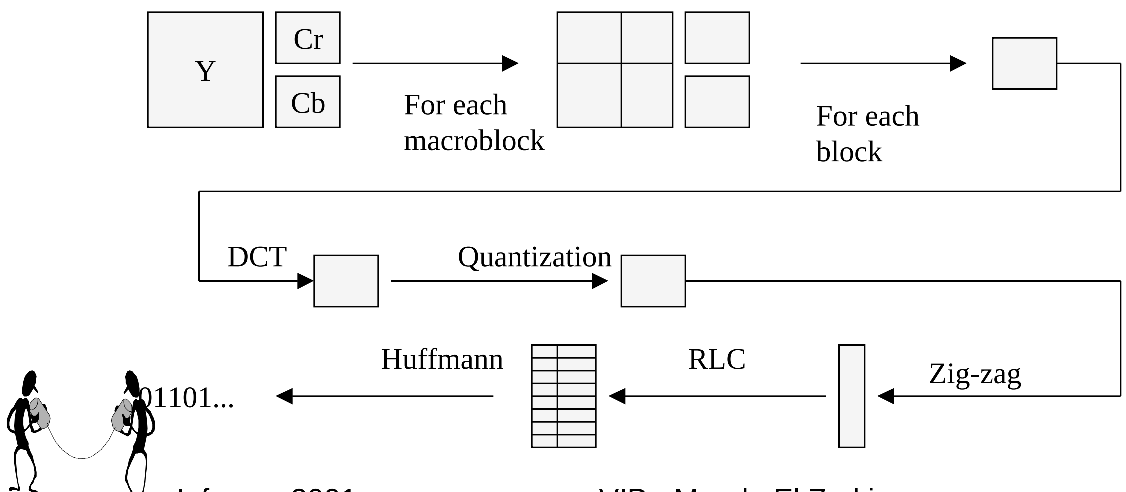
- I-frames (intra-coded images) are self-contained without any references to other images -> can be treated as still images.
- Need efficient compression scheme
  - Compression must be executed in real-time -> compression of individual frames must occur within a requested time interval.
- As only spatial compression is used -> low compression ratio when compared to other MPEG frames
- I-frames are points for random-access in an MPEG stream.





2. MPEG: Compression for I-frames

- Use 8x8 blocks defined within a macro-block
  - Perform DCT on these blocks
  - Quantization is done by a constant value for all DCT coefficients.



2. MPEG: P-frames

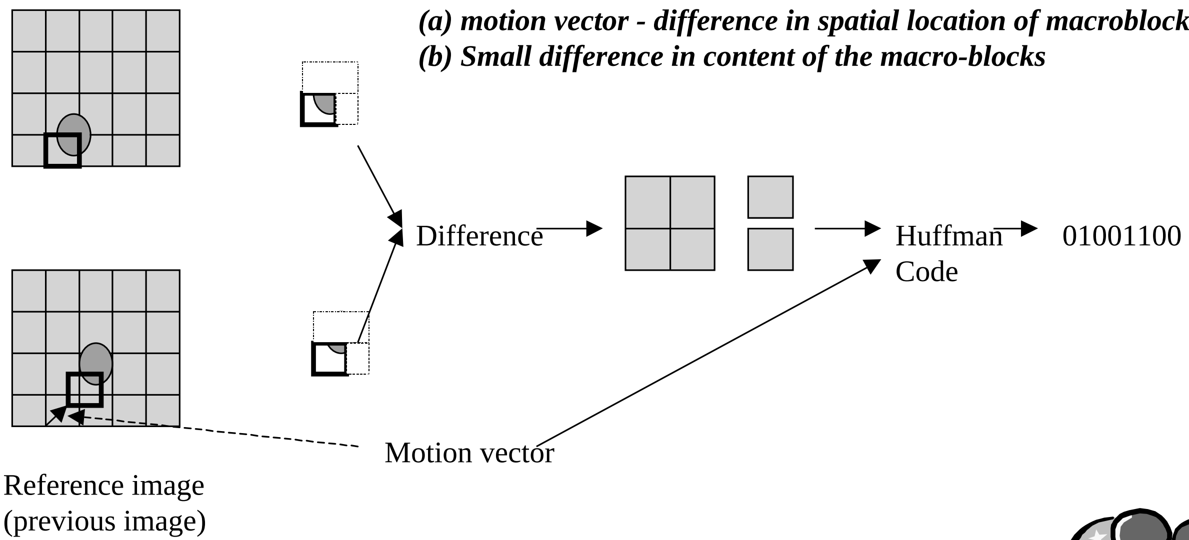
- P-frames (Predictive Coded Frames)
  - require information of the previous I-frame and/or previous P-frames for encoding and decoding.
- Motion estimation method at encoder:
  - Methods that are computation intensive often give better results. Tradeoff b/w cost and video quality.
  - MPEG does not specify a motion estimation method, it specifies the coding of the result.

2. MPEG: Compression for P-frames

Encode:

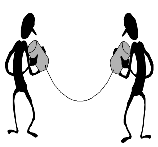
(a) motion vector - difference in spatial location of macroblocks

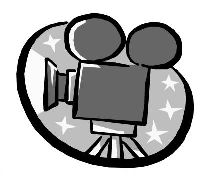
(b) Small difference in content of the macro-blocks



2. MPEG: Motion Computation for P-frames

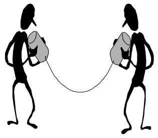
- Look for match window within search window
  - Match window corresponds to macro-block
  - Search window corresponds to arbitrary window size
  - Larger search window, better motion estimation

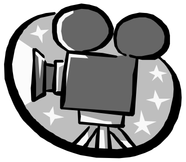




2. MPEG: Matching Methods

- Matching methods
  - SSD (Sum of squared differences)
    - SSD = $\sum_i (x_i - y_i)^2$:
      - subtract pixel by pixel
      - square the residuals
      - sum them
      - find minimal SSD correlation among matching windows.
  - SAD (Sum of absolute differences)
    - SAD = $\sum_i |x_i - y_i|$
      - correlation is absolute value of residuals
      - Can deal with outliers - In SSD, one bad point gives large difference which skews the decision of picking correct match windows.





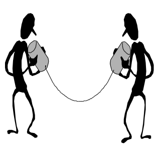
2. MPEG: Finding minimal SSD

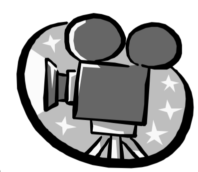
7 9 8 5 4 6 9 8 2	versus	8 7 9 7 5 4 7 5 4	→	SSD = (7-8)^2 + (9-7)^2 + (8-9)^2 + (5-7)^2 + (4-5)^2 + (6-4)^2 + (9-7)^2 + (8-5)^2 + (2-4)^2 1+4+1+4+1+4+4+9+4 = 32
7 9 8 5 4 6 9 8 2	versus	8 7 10 6 5 4 10 7 1	→	SSD = 18

Min SSD = 18

Take match windows

7 9 8 5 4 6 9 8 2	versus	8 7 10 6 5 4 10 7 1





2. MPEG: Comparing SAD and SSD

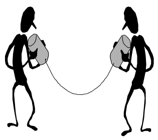
SSD:

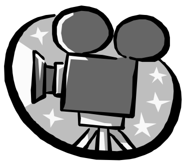
7	9	8	versus	8	7	10	→ SSD = 18
5	4	6		6	5	4	
9	8	2		10	7	1	

7	9	8	versus	8	7	10	→ SSD = 40,017
5	4	6		6	5	4	
9	8	2		10	7	202	

SAD:

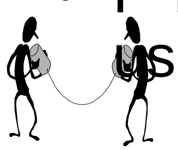
7	9	8	versus	8	7	10	→ SAD = 211
5	4	6		6	5	4	
9	8	2		10	7	202	

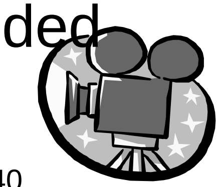




2. MPEG: Video Processing of P-frames

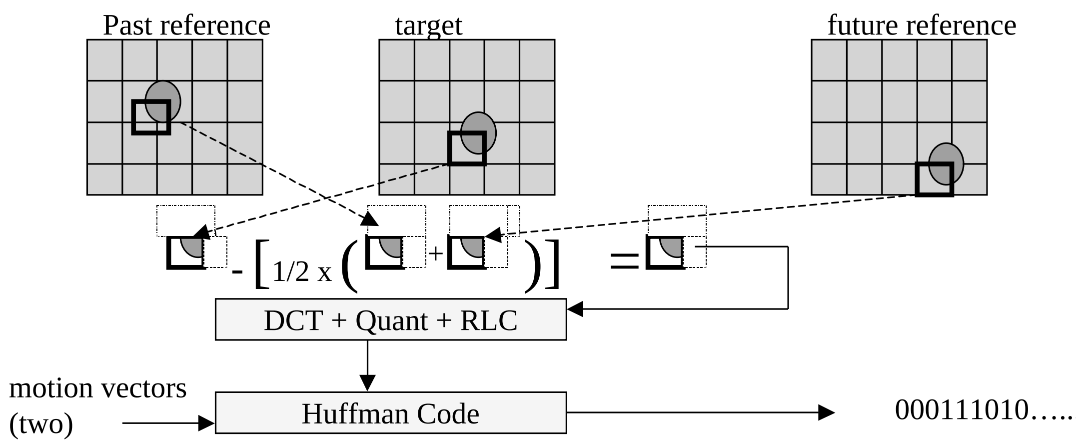
- Apply 2D DCT to macro-blocks not reduced by motion compensation.
- Motion vector of adjacent macro blocks often differs slightly, hence apply DPCM.
  - Maximum size of motion vector not defined in standard.
- P-frames consist of I-frame macro blocks and predictive macro-blocks.
- P-frames are quantized and entropy encoded using run length coding.

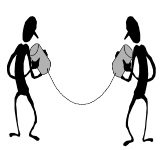


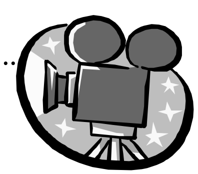


2. MPEG: B frames

- B-frames (Bidirectionally predictive coded frames)
  - requires information of previous and following I and/or P frame

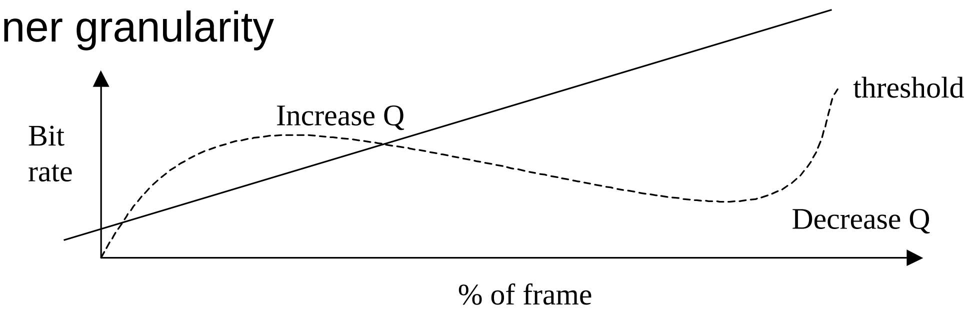






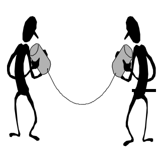
2. MPEG: Variable Quantization

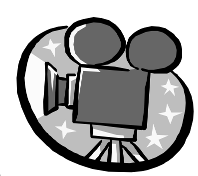
- AC coefficients of B and P-frames are usually large
- I-frames have smaller values
- MPEG quantization is adjusted as follows:
  - If data rate increases over threshold, then quantization enlarges the step size
  - If data rate decreases, then quantization is performed with finer granularity



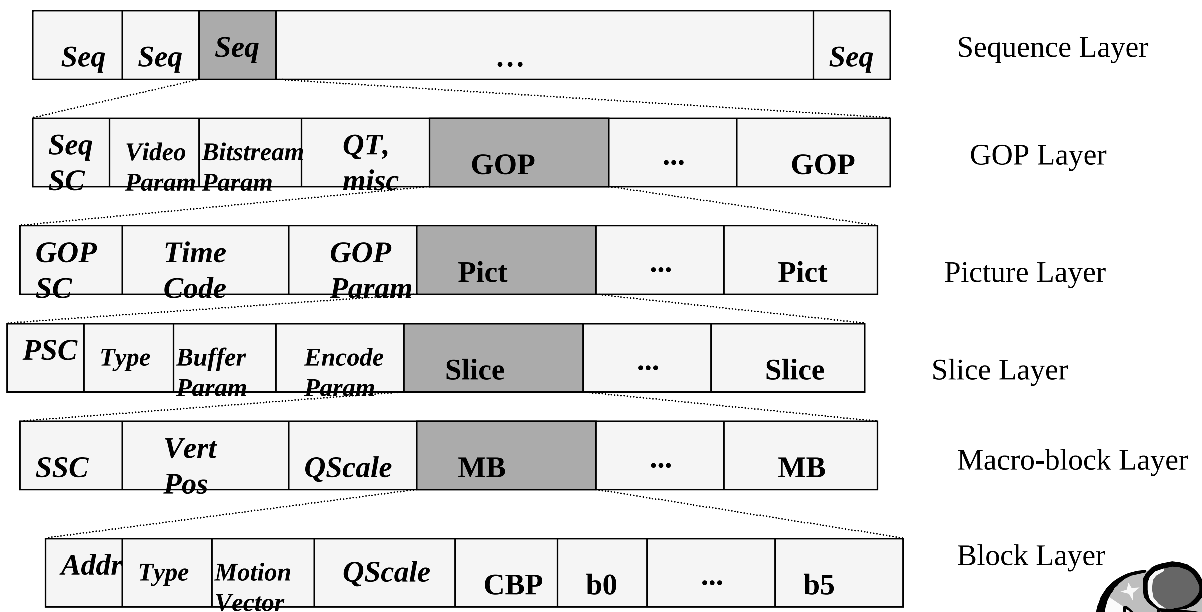
2. MPEG: System Data Stream

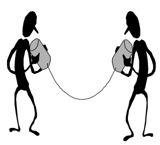
- Video Stream is interleaved with audio.
- Video Stream consists of 6 layers
  - Sequence layer
  - Group of pictures layer
    - Video Param - width, height, aspect ratio, picture rate
    - Bitstream Param - bitrate, bufsize
    - QT - intracoded blocks, intercoded blocks
  - Picture layer
    - Time code - hours, minutes, seconds
  - Slice layer
    - Type - I, P, B
    - Buffer Param - decoder's bufsize
    - Encode Param - indicates info about motion vectors
  - Macro-block layer
    - Vertical Position - what line does this slice start on?
    - Qscale - how is the quantization table scaled in this slice?
  - Block layer

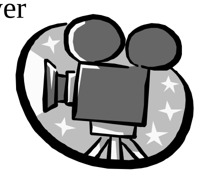




2. MPEG: Video Data Stream





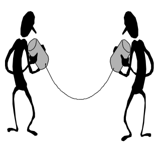


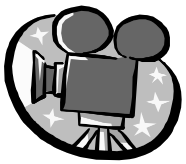
3. Quality

- What is video quality?
  - Generally judged using PSNR
    - Easy to compute

    BUT
    - Not a good estimate of quality
  - New objective quality measurements
    - Hard to compute

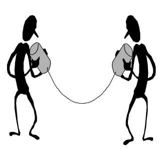
    BUT
    - More accurate

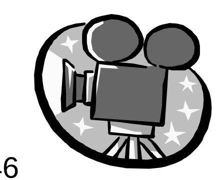




3. Quality: Assessment Techniques

- Traditional Objective Assessment - Peak Signal to Noise Ratio (PSNR)
- Subjective Assessment – DSCQS (Double Stimulus Continuous Quality Scale)
- Perceptual Objective Assessment -
  - Human visual perception based
  - Capturing image imperfections





3. Quality: Peak Signal to Noise Ratio

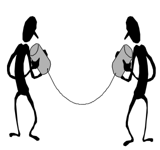
- For a video sequence of K frames of NxM dimension with 8 bit depth:

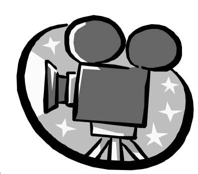
$$RMSE = \sqrt{\frac{1}{N\,M\,K}\sum_{k=1}^{K}\sum_{x=1}^{N}\sum_{y=1}^{M}\left[f(x,y,k)-f'(x,y,k)\right]^2}$$

$$P\ldots = 20\log_{10}\frac{255}{\ldots}$$

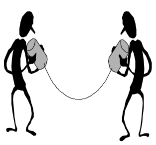
3. Quality: Advantages of PSNR

- Very easy to compute
- Well understood by most researchers
- Results are close to DSCQS but they do not translate accurately to human perception

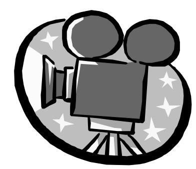




3. Quality: Disadvantages of PSNR

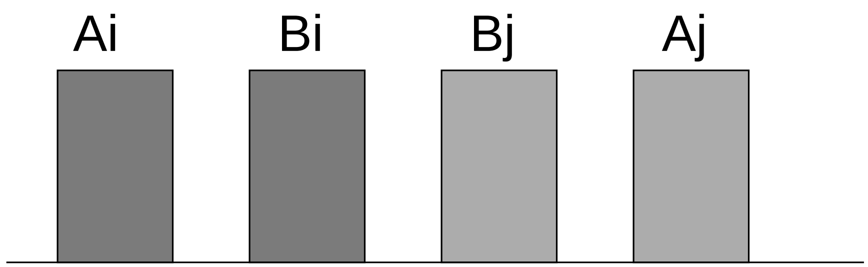


Some reconstructed images with different errors
have the same PSNR values



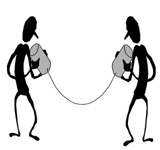
3. Quality: Subjective Assessment: DSCQS

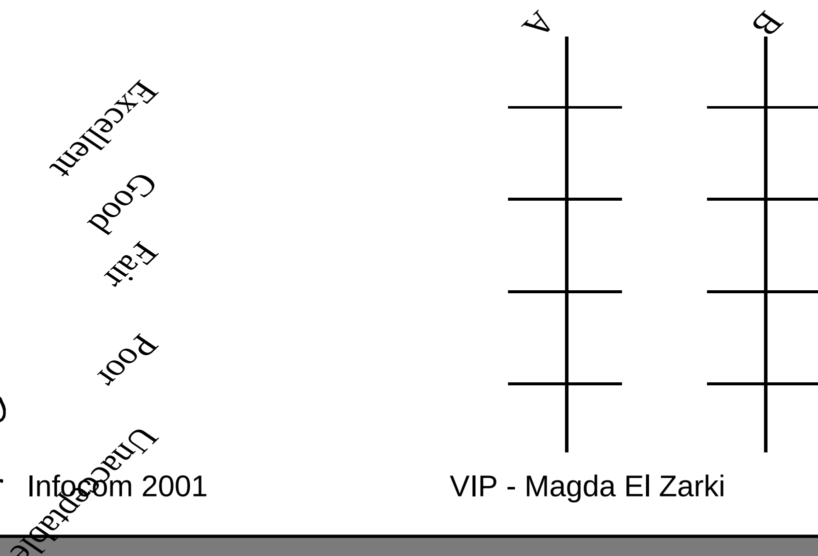
- Source (A) and Processed (B) video clips are presented in pairs
- The video presentation sequences are randomized

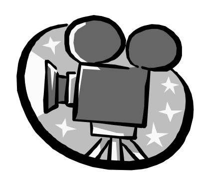


3. Quality:DSCQS scoring

- Viewers grade each clip's quality
- Data is gathered in pairs

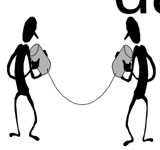


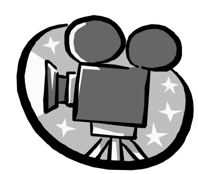




3. Quality: Issues with DSCQS

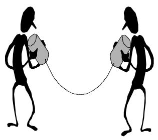
- Until now the most reliable quality measurement method
- Requires special viewing room and equipment
- Needs a large group of people
- Large amount of post processing on data

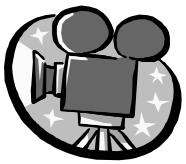




3. Quality: Objective Assessment (OA)

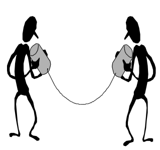
- Establish a good quality assessment model
- The model takes as inputs the source and the processed video clips.
- Compare the model output to DSCQS test score
- If the result is consistent with DSCQS measurement, the model can substitute DSCQS

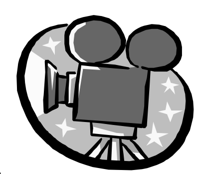




3. Quality: OA Requirements

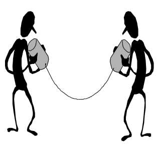
- Ability to predict subjective quality with low error
- Predictions agree monotonically with the relative magnitudes of subjective quality ratings
- Prediction is robust with respect to a variety of video impairments

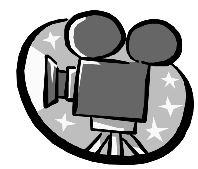




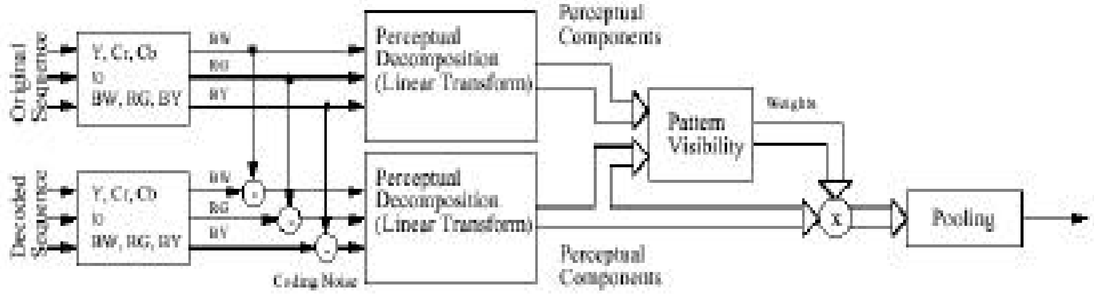
3. Design: OA Models - 2 approaches

1. Establish a model that simulates the human visual stimulation
2. Find the relationship between measurable parameters and perceptual distortion (blurring, tiling, noise)



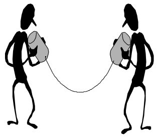


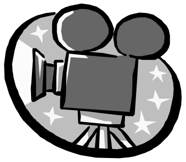
3. Quality: EPFL model



3. Quality: Issues related to EPFL Model

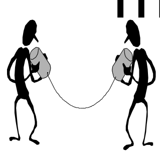
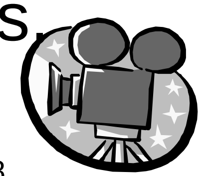
- Advantages:
  - Considers both luminance and chrominance
  - Very high correlation with DSCQS for some video sequences
- Disadvantages:
  - Not capable of in-service evaluation
  - Not consistent over all video bit rate ranges
  - Computationally complex





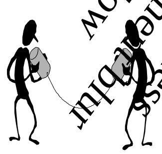
3. Quality: ITS Model

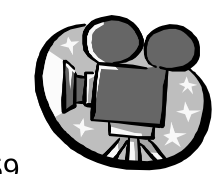
- Institute for Telecommunication Studies (ITS) were the first group to propose an objective measure several years ago.
- They have since refined (or fine tuned) the model to capture more of the image imperfections.
- They map image imperfections onto measurable mathematical parameters.

3. Quality: Perceptual Impairment Factor Vs. A Measurable Parameter

Image blur	Step response rise time
Edge busyness	Step response jitter width
False contouring	Sp-p to minimum quantizing
Granular noise	Equivalent analog signal SNR expressed as Sp-p/ Nrms
Dirty Window	Maximum noise amplitude
Movement blur	Rise time of a moving edge
Jerkiness	Field or frame difference in terms of moving edge position

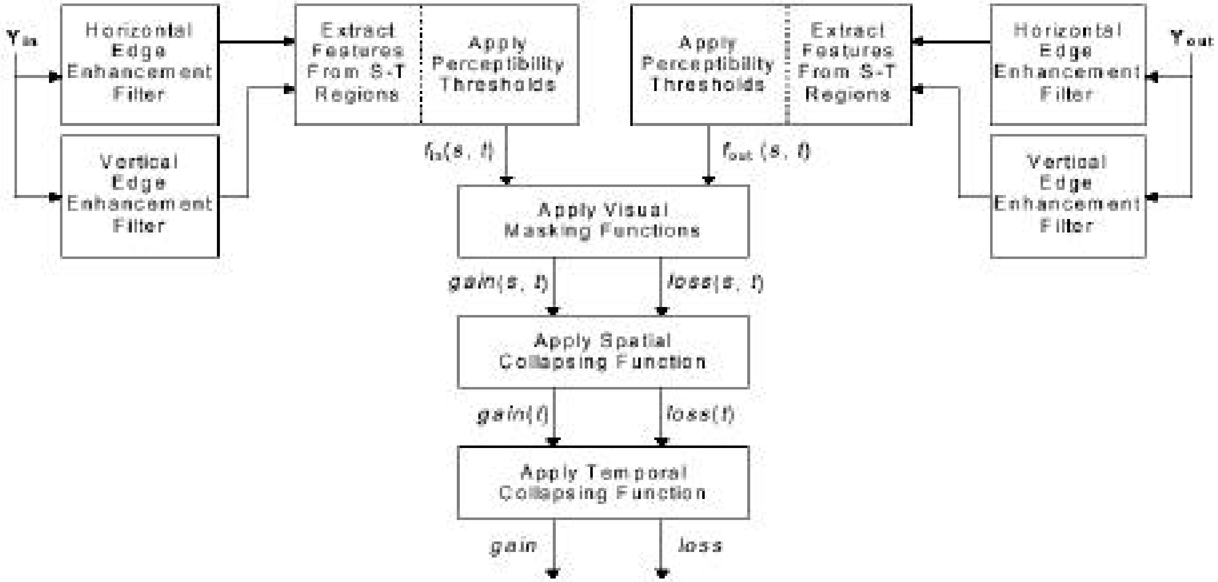




3. Quality: Perceptual Impairment Factor

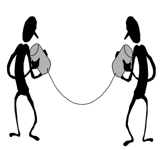
a) Original, b) Blur, c) Tiling, d) Noise

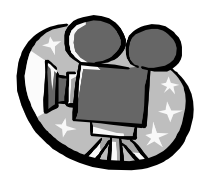
3. Quality: ITS Algorithm



3. Quality: Advantages of ITS Model

- Works well for a wide range of bit rates
- Produces results that are consistent with subjective tests (DSCQS)
- Computationally efficient
- Bandwidth efficient (384:1)
- In service quality monitoring



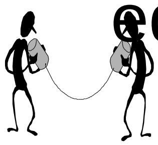


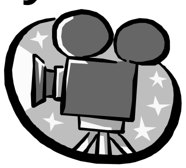
3. Quality: Disadvantages of ITS Model

- Based on no visual model (vs. EPFL)
- Only considers luminance value

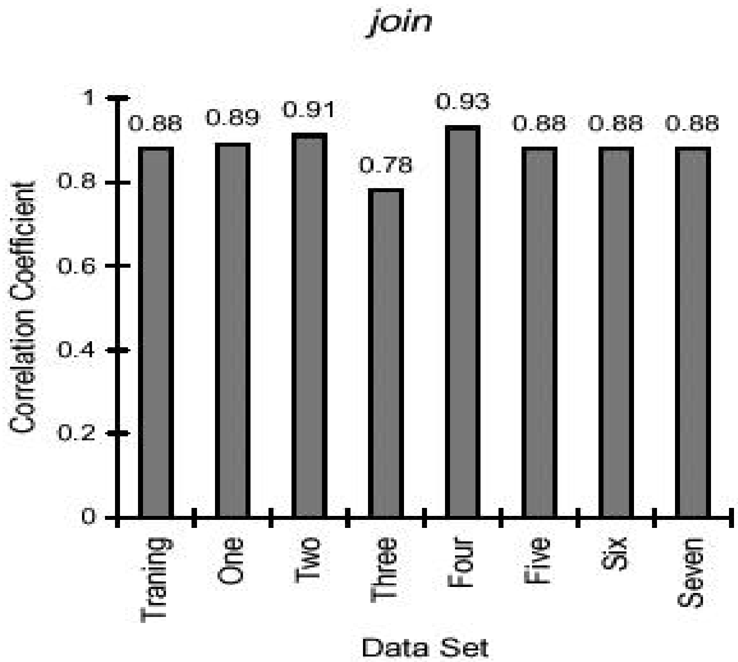
3. Quality: Video Quality Experts Group (VQEG)

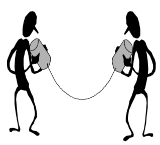
- Several models have been under evaluation
- Tested video bit rate from 768 kbps to 50Mbps (4:2:0 - 4:2:2 MPEG-2)
- Both NTSC and PAL signals tested
- Viewing Distance limit to 6:1
- Carefully calibrated and aligned display equipment

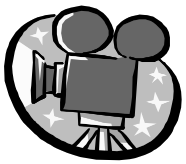




3. Quality: DSCQS vs ITS







3. Quality: Conclusions

- All models have strengths and weaknesses, not one can substitute DSCQS
- Some display fairly consistent behavior for different video resources