The University Of Sheffield.

EEE116 Multimedia Systems

Topic 7 – Visual Content – II

- Image and video representation
  - Data rates and available bandwidth
  - Redundant and Irrelevant data
- Image compression
- Video compression

- International standards for images and video

- Tutorial Questions : Q13 - Q16
- The SAQ for topics 6 and 7 in MOLE

Dr. Charith Abhayaratne
c.abhayaratne@sheffield.ac.uk

How to find the bit rate?

W pixels/line
H Lines/frame

H x W pixels/frame (This is the pixels in Y channel)

S Chrominance sub sampling factor
S=3 for 4:4:4 S=2 for 4:2:2 S=1.5 for 4:2:0

N bits/pixel (bpp)

S x N x H x W bits/frame

F frames/sec

S x N x H x W x F bits/sec

Derive the bit rate for 4:2:2 format 4CIF video!

Y channel resolution of 4CIF video = 576 x 720

Considering N= 8 bits per pixel per colour channel

Cb and Cr resolutions using 4:2:2 sampling S= 2

Memory per frame = 576 x 704 x 2 x 8 = 6,488,064 bits

Memory to store 90 minutes of video (at 50 frames per sec)
6,488,064 x 50 =324,403,200 bits/s
For 90 minutes 324,403,200 x 60 x 90 = 204 G Bytes

How many DVDs are required to record this programme?
(A single layer DVD can store only about 4.5 Gbytes per disk)

The University Of Sheffield.

Another Example

Derive bit rate for video transmission over mobile networks using QCIF 4:2:0 format

Mobile phones display resolution
180 x 144

Typical Frame rate is 6.25 fps

Bits/frame = 180 x144 x 1.5 x 8 ~ 304 K bits

At 6.25 frames/s ~ 1.9 Mbits/s

What is the available mobile phone network bandwidth?

What can we do to further reduce the data rate?

Why is it necessary to compress data?

Data rate requirements:

1. The data rate for 4CIF resolution 4:2:2 video =

2. The data rate for transmitting QCIF 4:2:0 and 6.25 fps video over a mobile communication link =

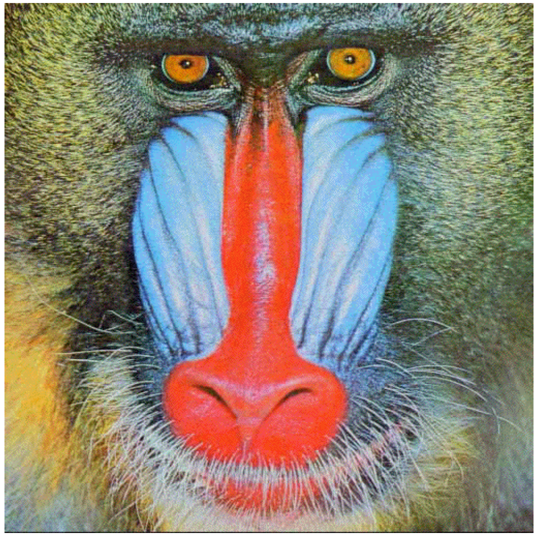
What is available:

Capacity of a DVD -	~4.5 Gbytes
Standard modem -	~20K bits/sec
Broadband modem -	~1M bits/sec
Mobile phone -	~128K bits/sec

The available bandwidth and storage space is limited.

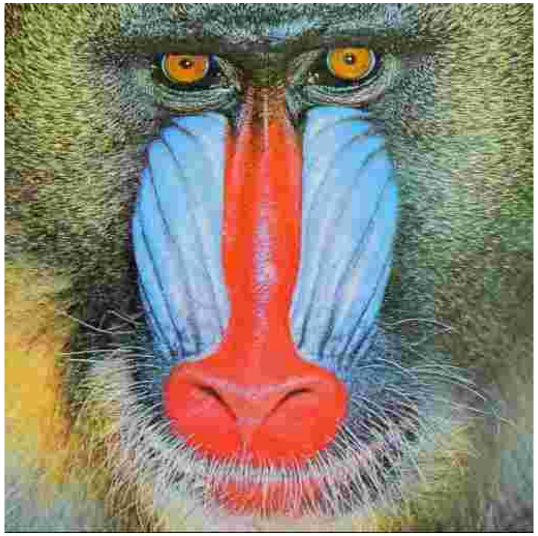
Solution is data compression

Image Compression Examples



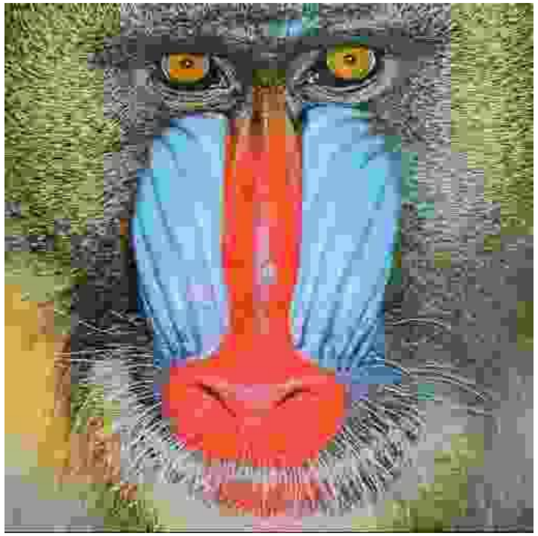
512 x 512 pixels (8 bit colour)
262144 bytes

10:1 Compression
26,200 bytes



20:1 Compression
13,100 bytes

40:1 Compression
6,600 bytes



Why is data compression possible?

Usual digital representations are redundant. For compression remove data redundancies. Data redundancy can be of several forms.

1. Coding Redundancy

We know images use N-bit (usually N=8 for monochrome or for each colour channel) codewords to represent images/video pixels per colour channel. This is a fixed-length code representation.

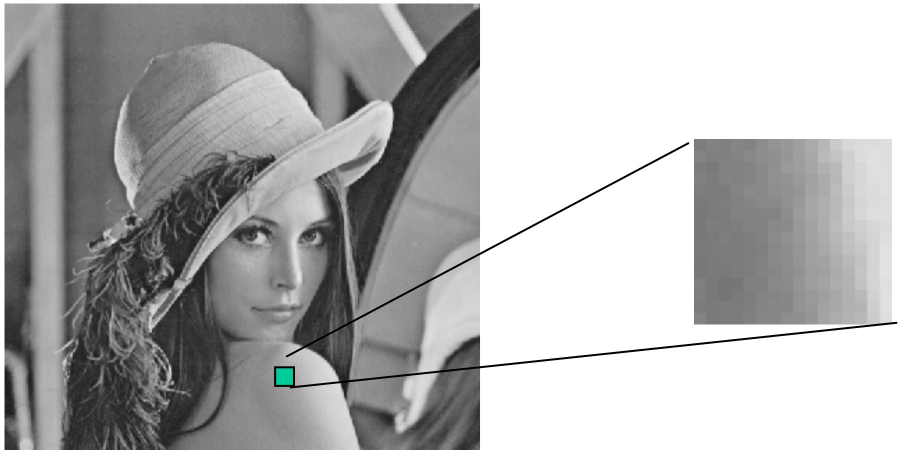
What are the disadvantages of using fixed length codes?

What is the solution?

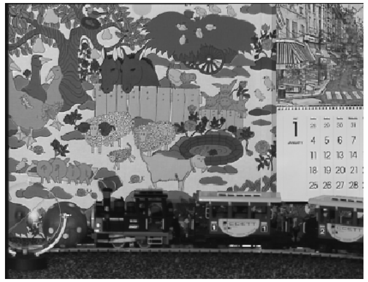
2. Inter-sample (inter-pixel) redundancy

The spatial and temporal (for video) correlations with neighbouring pixels. Usually, the neighbouring pixels have similar gray-levels.

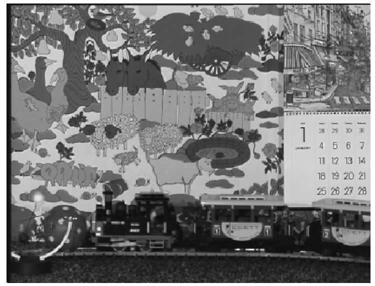
Spatial redundancy



Temporal redundancy



Frame 1



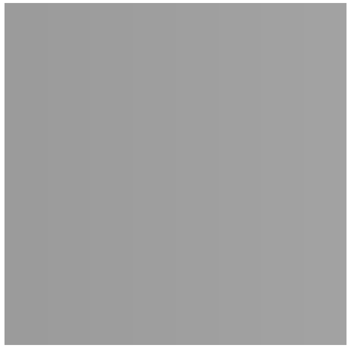
Frame 2

If data is correlated the inter-sample redundancy is high. By decorrelating data inter-sample redundancy can be removed.

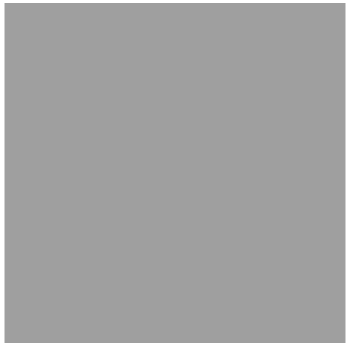
3. Psycho-visual redundancy

Consider these 8x8 pixel blocks. Visually they look the same.

A



B



The actual image matrices are:

$$A=\begin{pmatrix}156 & 157 & 158 & 159 & 160 & 161 & 162 & 163\\ 156 & 157 & 158 & 159 & 160 & 161 & 162 & 163\\ 156 & 157 & 158 & 159 & 160 & 161 & 162 & 163\\ 156 & 157 & 158 & 159 & 160 & 161 & 162 & 163\\ 156 & 157 & 158 & 159 & 160 & 161 & 162 & 163\\ 156 & 157 & 158 & 159 & 160 & 161 & 162 & 163\\ 156 & 157 & 158 & 159 & 160 & 161 & 162 & 163\\ 156 & 157 & 158 & 159 & 160 & 161 & 162 & 163\end{pmatrix}$$

$$B=\begin{pmatrix}160 & 160 & 160 & 160 & 160 & 160 & 160 & 160\\ 160 & 160 & 160 & 160 & 160 & 160 & 160 & 160\\ 160 & 160 & 160 & 160 & 160 & 160 & 160 & 160\\ 160 & 160 & 160 & 160 & 160 & 160 & 160 & 160\\ 160 & 160 & 160 & 160 & 160 & 160 & 160 & 160\\ 160 & 160 & 160 & 160 & 160 & 160 & 160 & 160\\ 160 & 160 & 160 & 160 & 160 & 160 & 160 & 160\\ 160 & 160 & 160 & 160 & 160 & 160 & 160 & 160\end{pmatrix}$$

Which block has the higher entropy?

How can we obtain B from A?

Image/video Redundancy

Inter-pixel Redundancy:

The spatial and temporal (for video) correlations with neighbouring pixels. Usually, the neighbouring pixels have similar gray-levels.

Psychovisual Redundancy:

The eye does not response with equal sensitivity to all visual information. Certain information has less relative importance than other information in normal visual processing. This information is said to be psychovisually redundant.

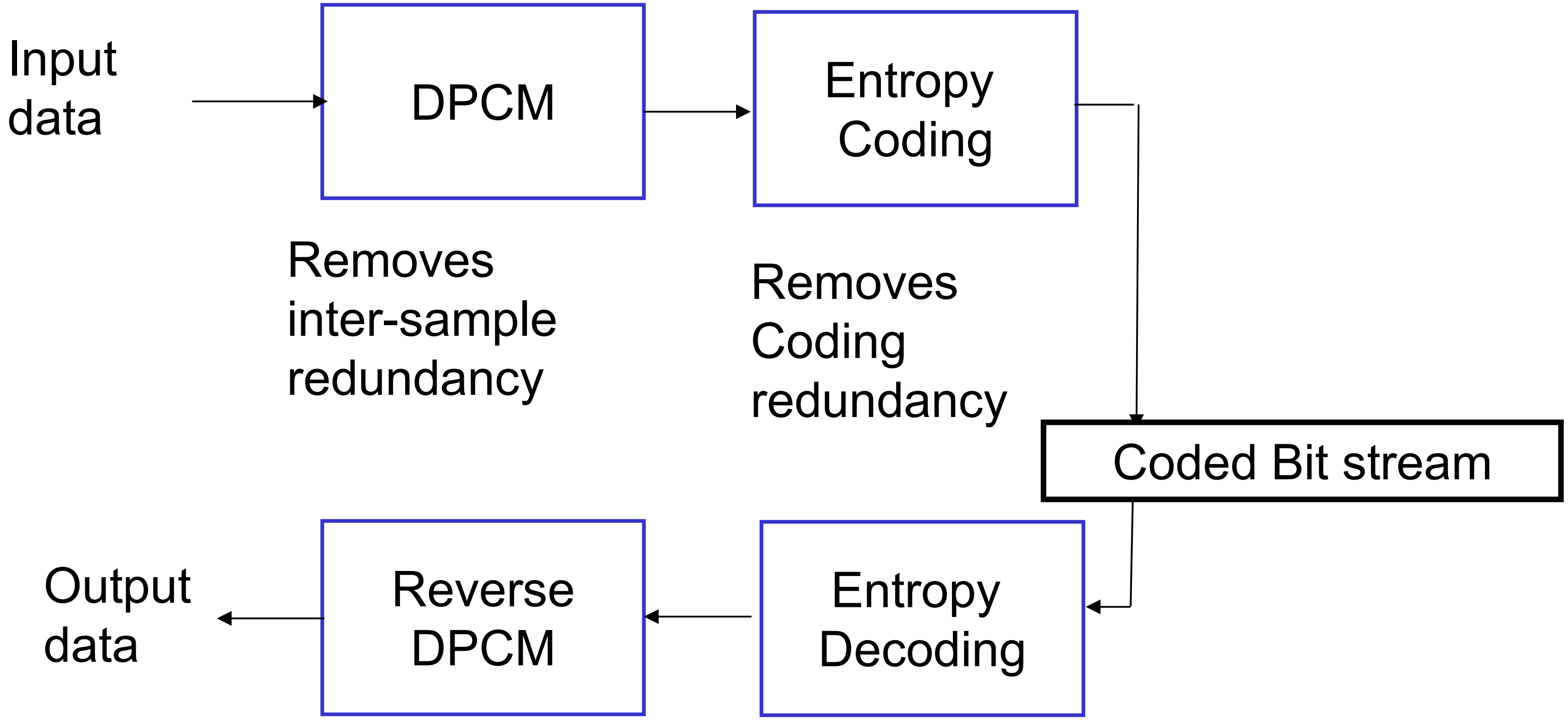
Coding Redundancy (R):

Based on the probability of occurrence for a symbol, and measured by **$R=1-E_C$** , where E_C is the coding efficiency for binary codes (L5)

Spectral Redundancy:

The correlation among different spectral bands. For example, the redundancy in RGB bands. (Refer to slide EEE116.6.57)

Data Compression Model (from Topic 4)



This results in a lossless compression scheme.

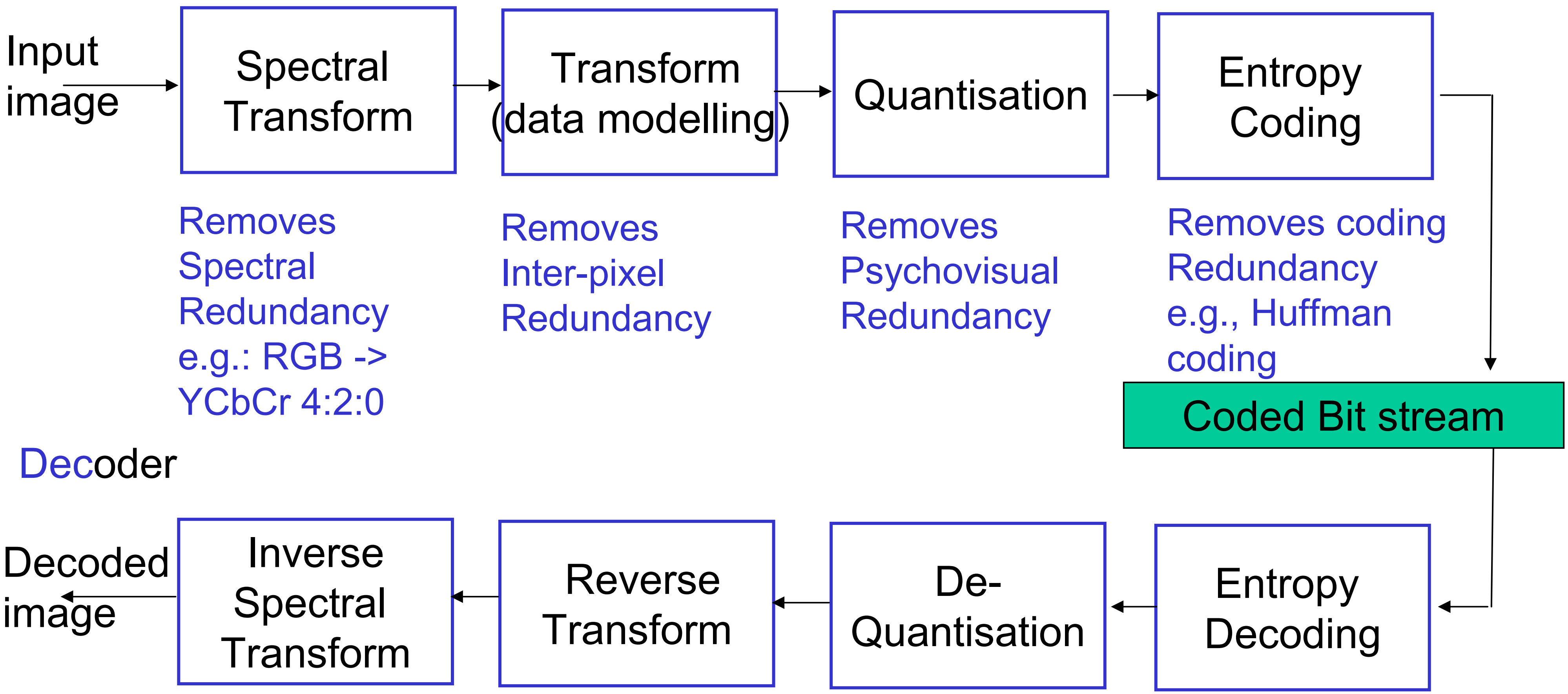
Can we do even better?

Consider each types of data and model the limitations of human perception.

The redundancy in data due to the limitations of the HVS is called psycho-visual redundancy

Image Compression Model

Coder



Coder-Decoder pair is called Codec

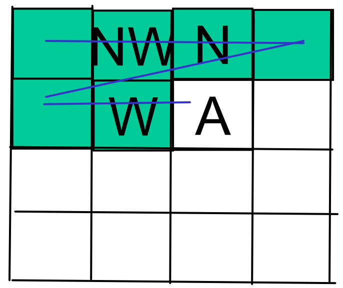
Removing spatial redundancy

We need to decorrelate the image data.

-One approach is to develop a model and use this model to predict that data and replace data values with the prediction error
- e.g.,
  - Linear Pedictive coding (for speech)
  - Differential Pulse Coded Modulation (DPCM)

Similarly for images 2D DPCM can be used.

An example of a 2D template
A_p=A-(aW+bN+cNW)
a+b+c=1;



But there are disadvantages in using DPCM

Therefore, for image coding there are more efficient methods:
-The Discrete Cosine Transform (DCT)
-The Wavelet Transform
(in Level 4 - EEE421 and EEE414)

Removing psycho-visual redundancy

We can reduce the bit resolution by further quantising the transformed image values.

Quantisation of a transformed pixel x using a quantisation factor Q is

$$\boxed{x_q = \text{Round}\{ x/Q\}}$$

Fewer bits need to represent x_q

Rounding operation discards information. Usually the appropriate Q values are determined by the HVS limits.

De-quantisation $\boxed{x_r = x_q Q}$

Note that x_r and x are not the same. This is where image coding process adds errors to pixel values.

Quantisation error is $x - x_r$

Example: Compute the value of Q for the example in slide EEE116.7.9

Lossless vs. Lossy coding

Can reduce the file size (or data rate) as most data representations are not optimally efficient.
Two basic forms of compression – **lossless** and **lossy**

- Possible to reconstruct exactly the original data.
- No information is discarded (does not include quantisation process).
- Useful for Scientific imaging, where the exact pixel values are required to do further image analysis.
- E.g., JPEG-LS
- Only low compression gains. (e.g., 1.5:1 or 2:1.
- The compressed file sizes are closer to entropy of the image.

- Possible only to reconstruct an approximation of the original data.
- Information is discarded.
- The limits of the HVS are exploited.
- High compression ratios.
- 10:1 visually lossless. (The rate at which just noticeable difference –JND- between the original and the compressed image is seen)
- e.g. JPEG

JPEG Compression

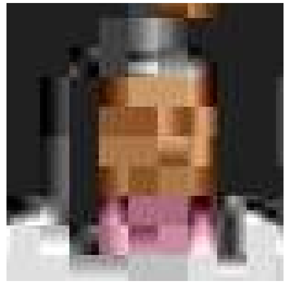
20:1

40:1

60:1

80:1

100:1



120:1

The latest standard JPEG2000 is much better
File formats: .jp2 and .j2k

Can we do better with video?

176 x 112 pixels at 8 bits/colour at 10 frames/s

Raw data rate = 4,730 k bit/s

Playing = 14.4 k bit/s

Compressed = 330:1 (0.3%)

Introduction to video compression

Can consider video as a simple sequence of digitized pictures

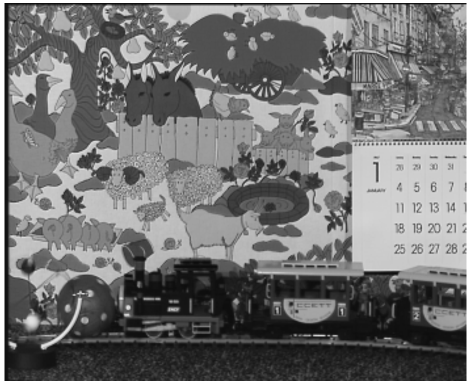
So just transmit a sequence of separately compressed JPEG images
This approach is known as Motion JPEG.
This is called Intra only coding (I frames).

Compression is limited as reasonable compression ratio per JPEG is ~20:1

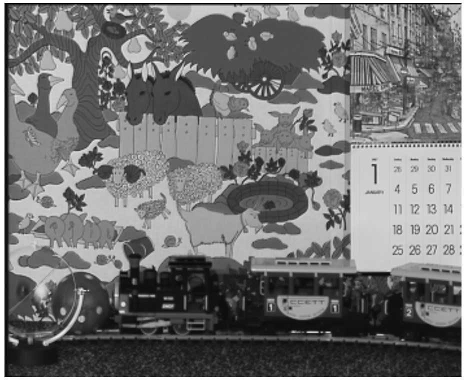
Need to use the temporal redundancy between frames (pictures) – often little change between two consecutive frames.

Video Motion

- Can transmit frame (picture) differences.



F1 7.61 bpp



F2 7.61 bpp

F2-F1 6.48 bpp

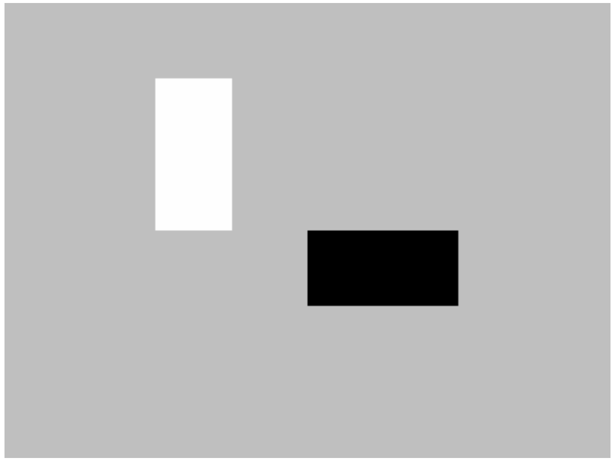
- Changes in consecutive pictures are due to
  - Consistent motion of large regions of image.
    - Horizontal, vertical and rotational
  - Camera motion
  - Zooming in and out
  - Scene and shot changes
  - Entry or exit of objects

The University Of Sheffield.

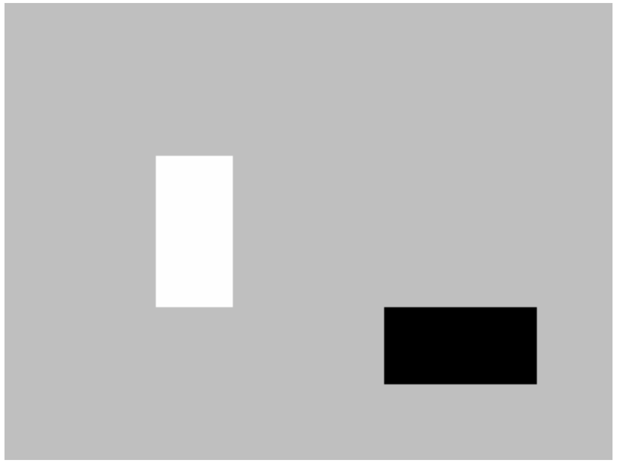
Video Motion

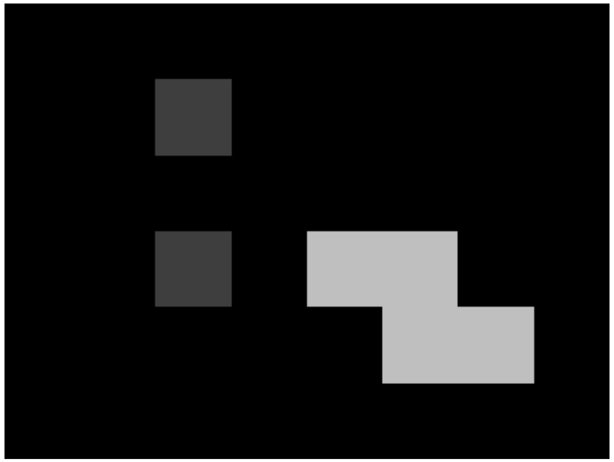
Consider two frames: A. --- Current frame (the frame to be predicted).
B. --- A previously encoded frame (the reference frame).

B



A

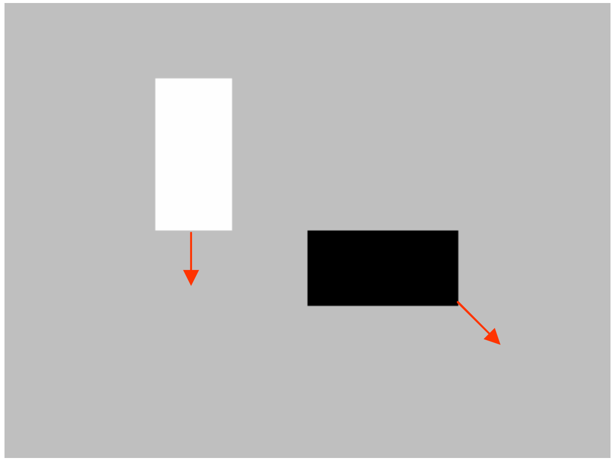




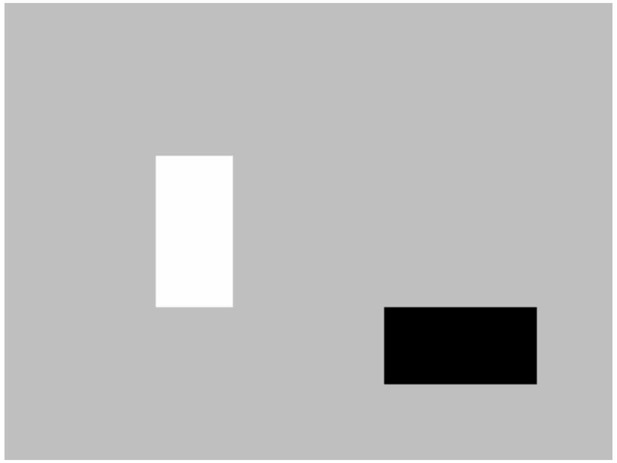
Frame difference
C= A-B (no change in background)
C is the prediction error.
Now we can write
A= C+B;
Since B is the previously coded frame. We only need to encode C to transmit Frame A to the receiver.

Video Motion

B



A



Now if we know how the displacements of the objects, we can do a better prediction.
C= A – f(B, displacement values),

f is called Motion compensated prediction.
The process to find displacement values is called Motion Estimation.

C is the prediction error.
A= C– f(B, displacement values),

Since B is the previously coded frame. We only need to encode C and displacement values to transmit Frame A to the receiver (C is zero if displacement values are accurate).

Motion compensated prediction (MCP)

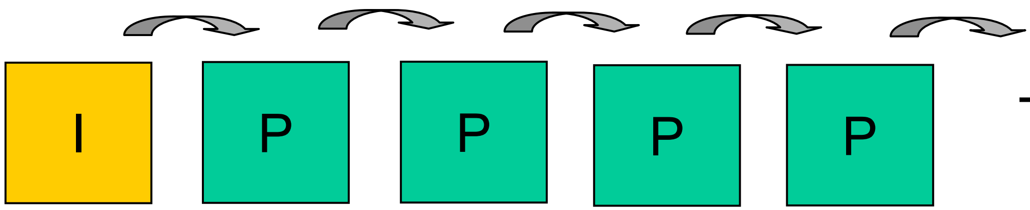
For complex scenes, it is difficult to estimate motion for each object.

Instead we partition each frame into smaller non-overlapping blocks, estimate displacement for each of the block. In this case it is difficult to accurately estimate displacement for all blocks. Hence prediction error (C) is not exactly zero.

C=F2- f(F1, displacements)

New entropy is 5.01 bpp

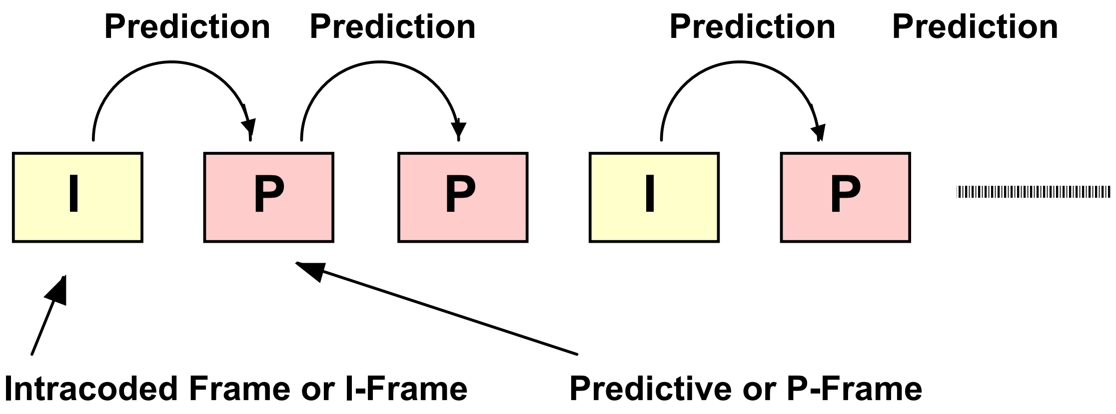
Intra frame (I) = A frame encoded as a still image.
Predicted frame (P) = A frame encoded using MCP. -- known as P frames



This is Motion compensated DPCM.

Is this a good strategy?
Why?

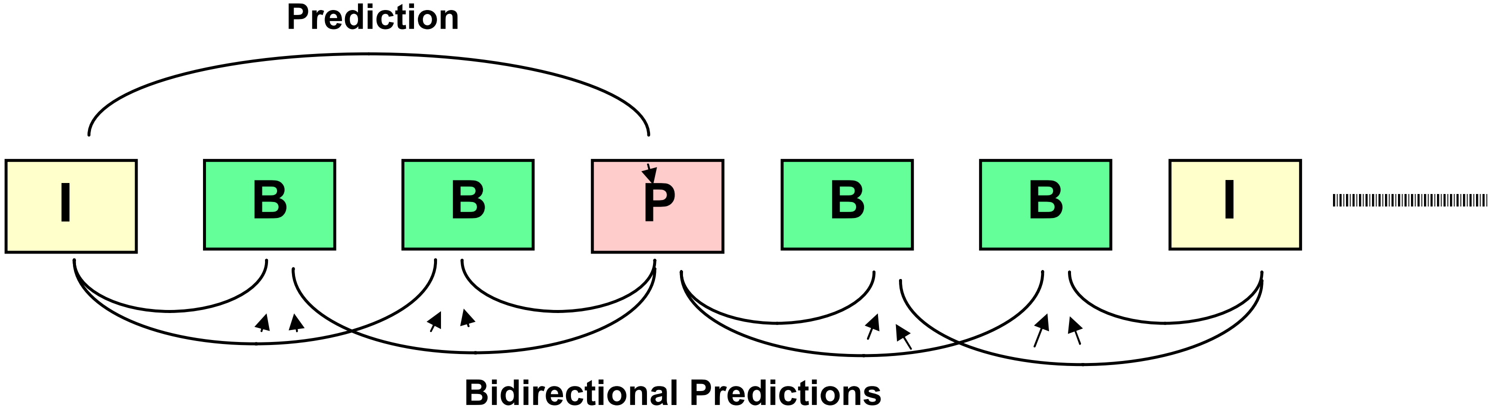
A better Strategy



- **Intra Frames** - encoded without reference to any other frame. Treated as separate picture and use JPEG type algorithm (normally) on separate Y, C_b and C_r images.
- Level of compression is fairly low. Must be the first frame of clip (and probably the first frame of a new scene).
- Need to resend a new I-frame occasionally as cannot work just with frame differences if an error occurs. An I-frame is coded, typically, at every 4-12 frames.
- Encoding of a **P-frame** depends on content of preceding I-frame or preceding P-frame. Use combination of motion estimation (displacement values) and motion compensated prediction - hence much higher compression.
- At the same time P- frames are highly computationally complex due to motion estimation.

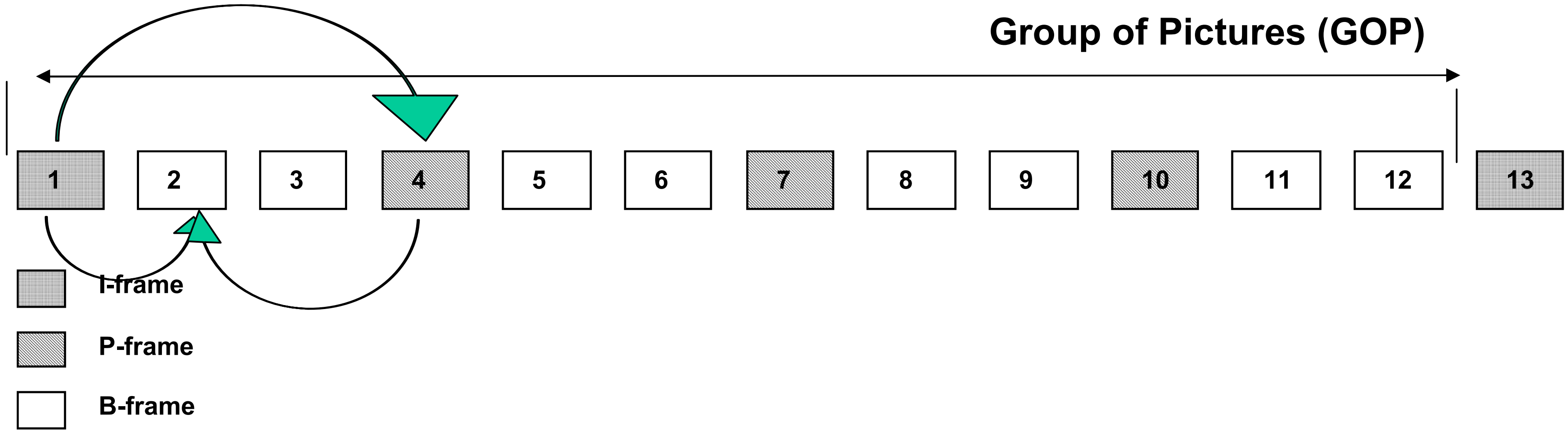
B-Frames

- Simple P-frames are fine for videoconferencing (head and shoulder shots - little movement). How about for general movies.
- Use **B-frames** - predict from past and future frames. Gives better estimate (Thus extremely high compression). Compressed heavily because they are not used as reference frames for further predictions.
- but cannot have fully real-time operation.



- B-frames have to be decoded using succeeding P- or I-frames.
- Increases coding delay. High complexity due to two motion estimations per frame.
- Need to reorder the incoming coded frames and require buffering to permit reordering for final uncompressed video stream.

Frame Reordering



A GOP is group of pictures, that contain only one I-Frame.
GOP size = (number of frames between two I frames) + 1

Display order = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, …….. Etc.

What is the coding / decoding order? What is the buffer size?

Summary

I	B	B	P	B	B	P	B	B	I	B	B	P
1	2	3	4	5	6	7	8	9	10	11	12	13

Display order of first 13 frames are shown above.

A GOP is a group of pictures that contains only one I-Frame.

What is the GOP size?

What is the coding / decoding order?

How many frames should be kept in the buffer at any given time?

Summary

I P B frames

How are they generated/coded?

Are they used as reference frames for generating other frames?

How much compression can be tolerated?

Summary

Four different frame organisations:

1. Intra only: I I I I I I I I I I I I I I ….
2. Motion compensated DPCM: I P P P P P P P P P …
3. Same as (2) but with occasional Intra frames: I P P P P P I P P P P …
4. Same as (3) but with B frames: I B B P B B I B B P B …

Which arrangement can get the highest compression?

Lowest compression?

Has the highest computational complexity?

Lowest computational complexity?

Most susceptible for propagation of errors?

Least susceptible for propagation of errors?

With the lowest coding/decoding delay?

Video compression architecture (encoder of an MPEG-2 codec)

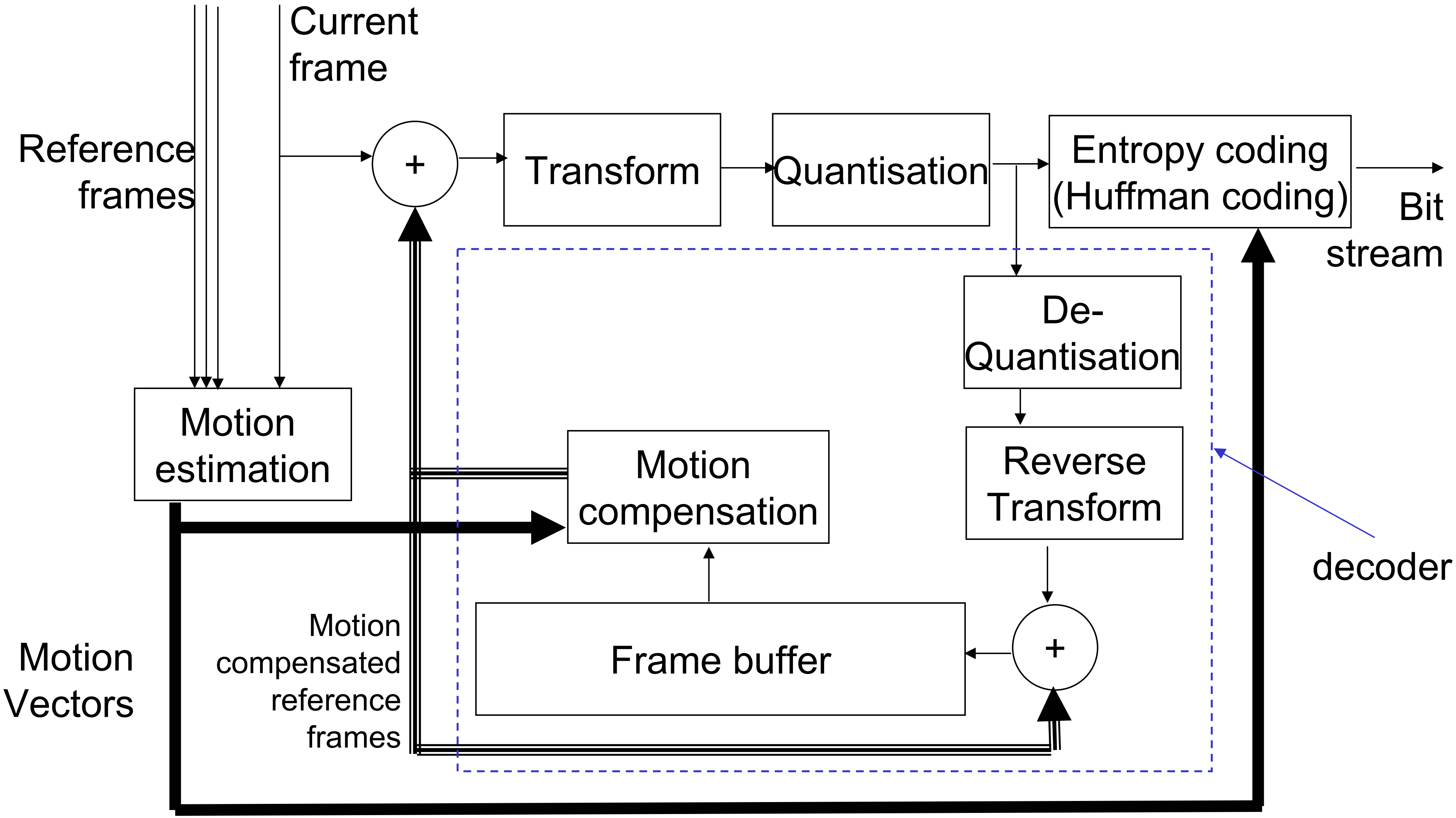


Image / Video Codec Performance Evaluation

The performance of codecs is usually evaluated using the rate-distortion (R-D) measurements.

1. Rate: This is a measure of the amount of compression. The most commonly used metric is the compression ratio (CR) .

There are many way to measure the compression ratio. A generic way to define it is as a ratio of "Original parameters" and "new parameters" for the compressed image/video.

$$CR = \frac{\text{Parameter value of Uncompressed data}}{\text{the value of the same parameter of Compressed data}}$$

Some parameters: data rates, total bits (file size)

data rates = bits per pixel (bpp) for images.
bits per second (bps) for video.

Examples

1. Using the bit rate: A CIF resolution 4:2:0 @ 25 fps 8 bits/colour sequence is to be transmitted over a link that has a capacity of 5000 Kbps. What is the compression ratio?

2. Using the total number of bits: A 1-hour long programme of the above sequence is to be recorded in a 1 Gbyte DVD. What is the compression ratio?

3. Using the bit rate: An 8 bpp image has to be compressed to 0.5 bpp. What is the compression ratio?

Rate-distortion measures

2. Distortion: In R-D analysis, distortion metrics measure the difference between the original image and the compressed image.

The simplest method is to compare two images/video pixel by pixel and compute the mean square error (mse). For an original image I_1 with WxH pixels and the decoded image I_2 the MSE is:

$$\boxed{MSE(I_1, I_2) = \underbrace{\frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H}}_{\text{Mean}} \underbrace{\left(I_2(i,j) - I_1(i,j)\right)^2}_{\text{(squared (error))}}}$$

As usual we represent this as a signal power to noise power ratio.

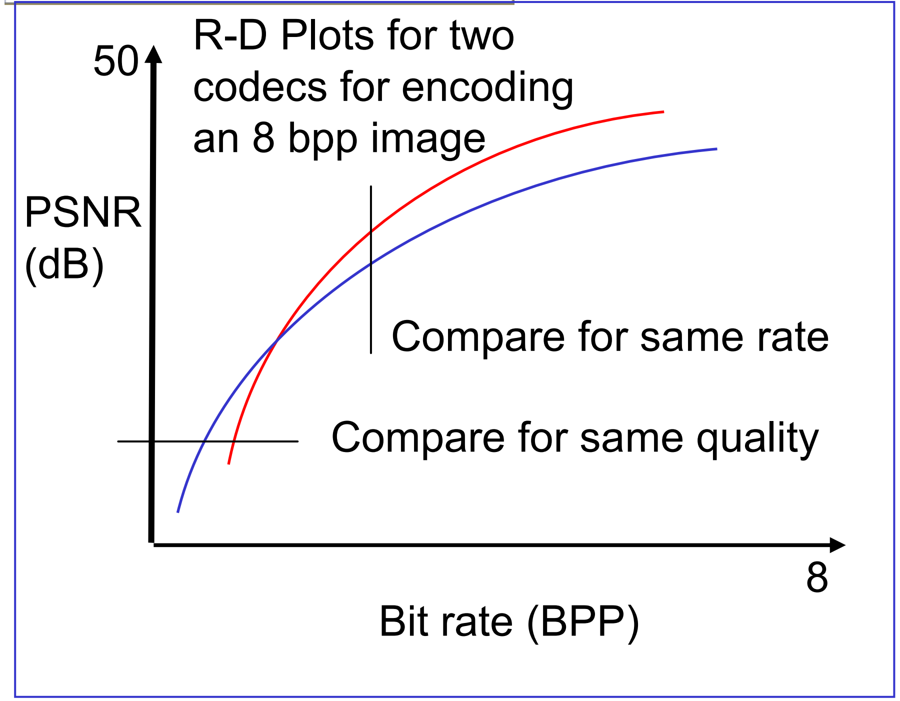
For visual applications, we define Peak Signal to Noise Ratio (PSNR).

$$\boxed{PSNR = 10\log_{10}\left(\frac{(\text{Peak signal value})^2}{MSE}\right)}$$

For an N-bit image the peak signal value is 2^N-1.

e.g. For an 8 bit image the peak signal value is $\boxed{}$

Rate-distortion measures



In R-D plots:
X- axis is rate (CR or bit rate)
Y- axis is distortion measure. (PSNR or MSE)

A good codec should provide
- Higher PSNR and lower bit rate.
- _______ MSE and _______ CR

What is the MSE and PSNR for a lossless coded image?

Is MSE, hence PSNR, a good metric for measuring image quality?

Image Compression Standards

	JPEG	JPEG2000	JPEG-LS
Usage	Lossy coding, Low compression ratios. Up to 20:1	Lossy coding, Very high compression ratios. More than 20:1	Lossless coding. Moderate compression ratios. 1.2:1 to 2:1.
Technol-ogy	Discrete Cosine Transforms (DCT), quantisation, Huffman coding.	Wavelets, Embedded quantisation, Arithmetic coding.	Advanced DPCM, Advanced entropy coding.
Function-alities.	Limited scalability	Fully scalable	No scalability.
Target Applicati-ons	Digital Cameras, Internet imaging,	Very low bit rate image applications. (mobile phones, internet)	Scientific imaging (e.g., medical imaging, remote sensing (satellite images, space images))

JPEG vs. JPEG2000 comparison
For Lena 512x512, 8 bpp



JPEG	0.1 bpp 21.9 dB	0.2 bpp 28.9 dB	1 bpp 37.8 dB



JPEG2000	0.05 bpp 27.9 dB	0.2 bpp 35.6 dB	1 bpp 41.9 dB

Video Compression Standards

MPEG-1

Technology: GOP Structure with I-P-B frames
Motion estimation – compensation
Macroblock size 16x16
JPEG like coding (DCT, quantization, Huffman coding)

Usage: Up to 1.5 Mbit/sec compression of CIF resolution non-interlaced video.

Target Applications: video storage and replay on PCs, CD-ROMs (VCD), video file transfer over the Internet (.mpg files).

Video Compression Standards

MPEG-2

Technology: An extension of MPEG-1. There are new tools for providing (limited) scalability.

Different profiles and levels within each profile have been defined.

profiles: simple, main, scalable, high, 4:2:2, multiview

levels: low, main (SD) , high-1440, high (HD)

e.g., main profile @ main level is used for DVD and DVB

Usage: Designed for between 1.5 and 15 Mbit/sec for both interlaced and non-interlaced.

Target Applications: Digital video broadcasting (DVB), Digital Television set top boxes, DVD, professional applications such as storage in studios.

Video Compression Standards

MPEG-4

Technology: Functionality for objects based processingas opposed to frames based in MPEG-1 and MPEG-2.
Additional functionalities for improving scalability and robustness (error resilience). Several profiles.

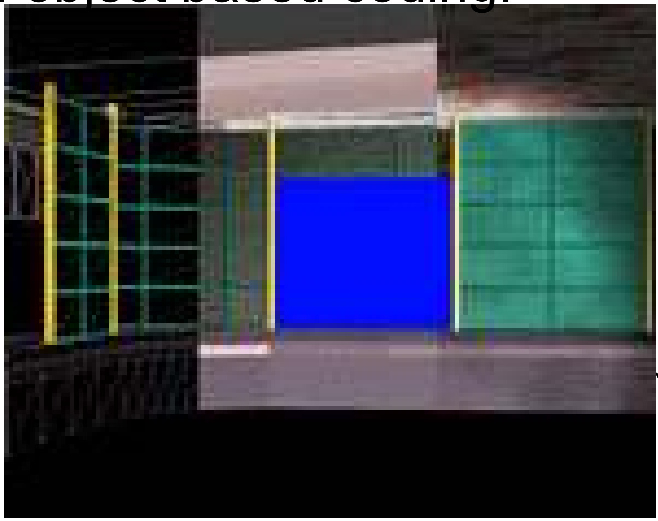
Usage: ranging from extremely low rates and resolutions as required by mobile video transmission up to high rates, resolutions and fidelity as applicable in the field of professional production.

Target Applications: Internet and mobile video (Simple Profile), video on demand (Advanced Simple Profile) and studio applications (Studio Profile)

Other MPEG standards (MPEG-7 and MPEG-21) do not focus on compression.

Video Compression Standards

Usage of object based coding:



Background

Object1

Foreground 1

Object 2

Object 3

Composite

Video Compression Standards

ITU-T Standards – Mainly target video transmission over telecommunication networks

H.261 : similar to MPEG-1. Main difference is no B frames.
H.262 : identical to MPEG-2.
H.263 : Extension of above two. Focus on low nit rates suitable for video conferencing.
H.264 : Further extensions to H.263. Very good performance. Jointly developed by ITU-T VCEG and ISO MPEG. Therefore, this is a joint standard. Known as MPEG-4 Part 10 Advanced Video Coding. This is the state-of-the art video compression scheme.

Storage requirement for 90 min DVD quality video