Video Compression - An Introduction

Introduction to an Image

A colour image is represented in the RGB component system.
(Image = R-component + B-component + G-component). This system also contains information about the brightness (Luminance) as well as the colour (Chrominance).

Red = luminance (Y) + red colour difference (Cr)
Blue = Y + Cb
Green = Y + Cg
Because Cr + Cb + Cg = constant, we need only to encode and store two chrominance components, usually Cr and Cb.

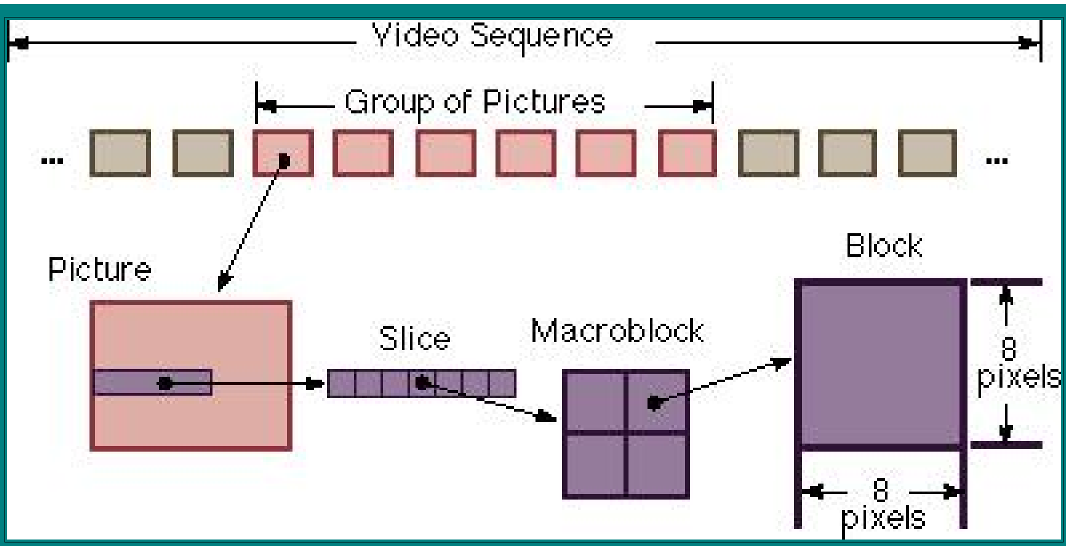
The Human Visual System (HVS) is less sensitive to colour information than it is to brightness information. The chrominance signals can therefore be represented with a lower resolution than the luminance. In short, the luminance component (Y) has a high sampling rate while the color difference components (Cb and Cr) have low sampling rate.

The need for compression

The digital video transmission (for example CCIR 601 format) requires about 216 MBps whereas an analog TV signal requires 6.-7 MHz. Hence the digital video information must be compressed (encoded) prior to transmission and decompressed (decoded) before displaying at the receiver.

Digital video data contains lot of redundancy, namely, spatial and the temporal redundancy. This is a motivation for compressing the video data.

The MPEG Data Hierarchy



Video Sequence = Seq. Header + n(GOP's) + End of Seq.

GOP = Group of Pictures (frames) = header + series of pictures (frames)

Picture (Frame) = A primary coding unit of a video sequence. It consists of 3 matrices Y, Cb and Cr. An RGB system can be easily converted to the Y,Cb,Cr system by a linear transformation.

Slice = Group of Macroblocks

Macroblock = 2 x 2 matrix of Blocks

Block = 8 x 8 pixel set

In simple terms the ascending order of hierarchy is [Block --> Macroblock --> Slice --> Frame]

Types of Frames (as per the MPEG standard)

- Intra Frames (I-frames) : These frames are coded using the data in the frame itself.
- Predicted Frames (P-frames) : These frames are coded using the data from the nearest previous I-frame or P-frame.
- Bidirectional Frames (B-frames) : These frames are coded using the data from the past as well as future I/P frames.

Video Stream Composition

The encoder decides the frequency and location of the I-frames. The MPEG encoder re-orders the frames in the video stream to present the frames to the decoder in the most efficient sequence. That means the display order of frames in most cases will be different from the video stream order.

P-frame and B-frame coding using Motion Compensation

Motion compensation is done at macroblock level. The temporal redundancy is eliminated by using motion vectors. Instead of sending the entire frame, only the motion vectors and the content difference between the macroblocks (error) are sent in a compressed form. This is the coding technique used for P-frames and B-frames to eliminate the temporal redundancy.

I-frame coding (Transform Coding)

The I-frames usually have a lot of spatial redundancy. For example the background of an image may be of one single color.
A 8 x 8 block of pixels is transformed from the spatial domain to the frequency domain using the DCT (Discrete Cosine Transform)

Discrete Cosine Transform Coding

The coding consists of the following steps:

a) *Separation of the image (I-frame) into blocks*
Typically a block consists of 8 x 8 (or 16 x 16) pixels.
(8x8 is optimum for trade-off between compression efficiency and computational complexity)
Note : larger block size leads to more efficient coding but requires more computational power

b) *Applying the DCT*
DCT converts a block of pixels into a block of transform coefficients. The coefficients represent the spatial frequency
components which make up the original block. Each coefficient can be thought of as a weight which is applied to an
appropriate basis function. Click here to see the basis functions for a DCT.

c) *Quantization*
For a typical block, most of the DCT coefficients will be almost zero. The larger DCT coefficients will be clustered
around the DC value (the top-left basis function) i.e.: they will be low spatial frequency coefficients
The DCT coefficients are quantized so that the near-zero coefficients are set to zero and the remaining coefficients are
represented with a reduced precision - this can be achieved by dividing each coefficient by a positive integer which
results in loss of information as well as compression

d) *Encoding*

Further compression is given by exploiting the statistical redundancies within the quantized DCT coefficient data
non-zero coefficients are encoded using a variable length coding scheme and zero coefficients are encoded using
run-length encoding (i.e.: transmit a number representing the current run of zeroes).

- *Some links:*
  - A detailed overview of Video Compression
  - Java Applet for DCT