Slides 1 - 6 = Images
Slides 7 - 9 = Video
Slides 10 - 11 = Audio
Slides 12 - 13 = Delivery

Questions
- An encoder when encoding a macroblock in a P picture can decide to encode it as an INTER mode or INTRA mode macroblock. Explain why the MPEG standard allows such a choice and suggest a method that an encoder could use for deciding which macroblock mode to use.
    - if intra
        - you effectively just have texture your coding directly - doing dct on values
    - if inter
        - you have texture coding also, but a different texture, and you will have motion vector
    - in mpeg 1 in notes series of steps - can be enhanced by looking at variation & sum of absolute differences

Hi Alan,
I will not expect that you will know "low level" details such as the conversion between colour spaces. However, you may be expected to know that such a conversion takes place (for example it can be part of JPEG compression).

It is dangerous to assume that certain topics will not be examined as they may contribute indirectly to an answer. However, chroma keying material as covered in the image and video part of the course will not feature.

Regards,
John.

<u>Notes</u>
Quantisation
- The compressing of image, audio or video files by removing data that does not affect the overall quality of the file. ie. lossy.

DCT
- The discrete cosine transform (DCT) represents an image as a sum of sinusoids of varying magnitudes and frequencies. The dct2 function computes the two-dimensional discrete cosine transform (DCT) of an image. The DCT has the property that, for a typical image, most of the visually significant information about the image is concentrated in just a

few coefficients of the DCT. For this reason, the DCT is often used in image compression applications. For example, the DCT is at the heart of the international standard lossy image compression algorithm known as JPEG. For most images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT.

- Compression is achieved since the lower right values represent higher frequencies, and are often small - small enough to be neglected with little visible distortion.

Macroblock

- This is an image compression component and technique based on discrete cosine transform used on still images and video frames. Macroblocks are usually composed of two or more blocks of pixels. The size of a block depends on the codec and is usually a multiple of 4.

Colour Model

- The purpose of a color model is to facilitate the specification of colors in some standard generally accepted way. In essence, a color model is a specification of a 3-D coordinate system and a subspace within that system where each color is represented by a single point.
- Each industry that uses color employs the most suitable color model. For example, the RGB color model is used in computer graphics, YUV or YCbCr are used in video systems and so on.

MPEG-4

- Method of defining compression of audio and visual (AV) digital data.

Video Encoding

- In video editing and production video encoding is the process of preparing the video for output, where the digital video is encoded to meet proper formats and specifications for recording and playback through the use of video encoder software. Also called video conversion.

A packet is considered lost if;
- Packet never arrives
- Arrives later than its scheduled playout time

Automatic Repeat reQuest (ARQ) may not be viable (to combat packet loss) as it is unlikely that the retransmitted packet will arrive on time.

Low-pass filter: frequencies above a certain threshold are discarded.

Entropy Coding e.g. Huffman, run-length coding (lossless)
Transform coding e.g. DCT

Use a motion vector to identify where parts of the image in the target frame were located in the reference frame.

PSNR (Pulse Noise Reduction)

Visual acuity (accuracy is distinguishing closely spaced lines) is much greater for grey (intensity only) than for colour

The human is more sensitive to luminance (brightness) than chrominance (colour).

Chroma sub-sampling
- ● Practice of encoding images by implementing less resolution for chroma information than for luminance information.

JPEG Compression Workflow
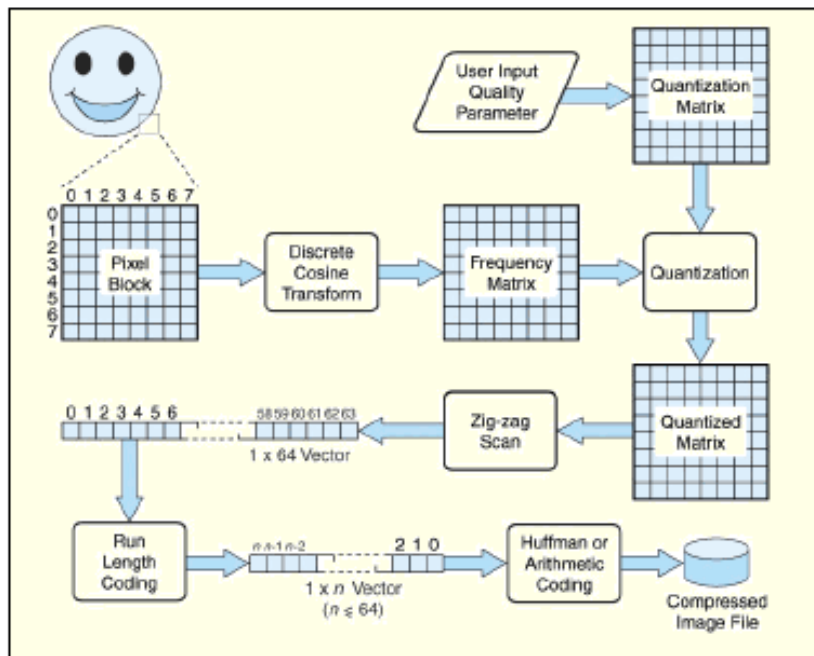
Example Transform
- ● If we have a block of 4 sections. If the pixels are 8 bits each then the block would use 32 bits.
- ● Using the transform we could assign 4 bits each for the difference values and 8 bits for the base pixel.
- ● This would reduce the data to 8 + (3x4) = 20 bits for the 2x2 block.

The first value in the list represents the DC Coefficient. It represents the average value of the block.

Digital compositing is the digitally manipulated integration of at least two source images to produce a new image.

Alpha blending is blending foreground and background

Keying is separating foreground and background
- ● Techniques: luma, chroma and difference keying

Rig removal is removing unwanted elements

Rig removal tasks
- ● Apply a mask to mask out the rig
- ● Replace pixels in masked area by corresponding pixels in clean plate background

The Matting Problem
- ● It is difficult to separate a foreground from the background when there are many colours e.g. hair where it is difficult to decide what belongs to the background or the foreground. Easier to do in controlled environment e.g. green screening.

Solution to the Matting Problem
- ● Vlahos Assumption
- ● Assume that all foreground colours have the magnitude of their blue component related to the magnitude of the green by some ratio k2.

Triangulation Matting (Smith & Blinn)
- ● Instead of reducing the number of unknowns, increase the number of equations.
- ● Capture an image of an object against two known backgrounds of different colour.
- ● Results in six equations and four unknowns - three equations for each unknown.

Problems with Matting
- ● Images do not look realistic
- ● Lack of refracted light
- ● Lack of reflected light
- ● Composite objects do not affect the image

Quantisation Tables
- ● Farid has identified the quantisation tables for over 200 cameras and software - on average each table is common to 1.4 cameras / software.
- ● Can (roughly) identify if an image was edited and re-compressed - contains a table different to the alleged source camera. However Photoshop signatures are unique.

JPEG encoding / decoding
- ● The encoding of a JPEG image involves DCT, quantisation, and entropy encoding
- ● The decoding of a JPEG image involves the reverse - entropy decoding, de-quantisation, and inverse DCT

Double Quantisation
- ● Is a point-wise operation described by a two-parameter family of functions.

- Effect of Double Quantisation
  - When the step size decreases some bins in the histogram are empty
  - When the step size increases some bins contain more samples than their neighbouring bins

Block Artefacts
- The block-based DCT technique used in JPEG partitions an image into blocks
- A block artefact grid (BAG) is the grid embedded in an image where block artefacts appear.

Difference Coding Techniques
- Difference Coding (conditional replenishment)
  - Each frame of a sequence is compared with its predecessor and only pixels that have changed are updated. Very often very little changes.
- Temporal Redundancy
  - Difference between corresponding pixels of two video frames

Macroblock Assumptions
- The effectiveness of macroblock based motion compensation depends on several assumptions;
- Objects move in a plane that is parallel to the camera plane
- The effects of zoom and object rotation are not considered
- The level of lighting is constant throughout the image and does not change over time

Motion Compensation Compression
- The steps in compressing digital video using Motion Compensation (MC);
- 1. Search for motion vectors (Motion Estimation)
- 2. MC-bases Prediction
- 3. Calculating the prediction error (the residuals)

Motion Compensation Stages
- The current frame to be compressed is divided into uniform non-overlapping macroblocks
- Each macroblock (target) in the current frame is compared to the content in other frames to find one that is similar
  - The location of similar pixels in the other frame can be different from the location of the target macroblock
  - The relative difference in locations is known as the motion vector
- The motion vector detailing the position of the target macroblock's match is encoded instead of the target macroblock itself
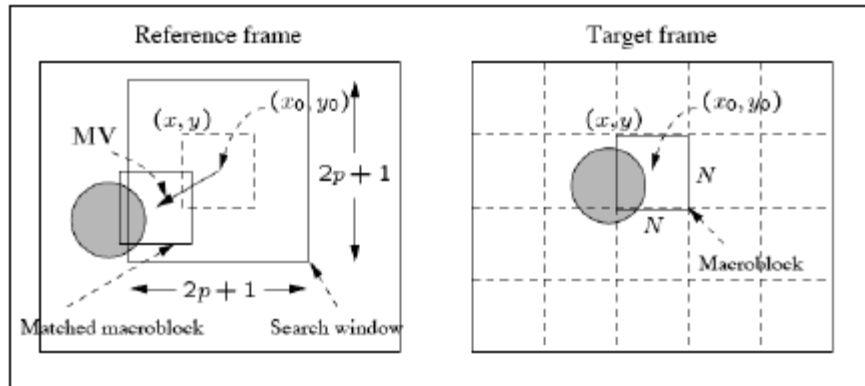
Forward Prediction
- To compensate inter-frame differences due to motion: Find the best matching part of the previous frame.

Bi-directional Prediction
- Use parts from previous or future frames.

- Look for best match.
- Possibly use values from both frames.

Motion Estimation

Search Window
- Skip candidates that are unlikely to give a good match

Avoid Searching
- If the difference between the target macroblock and the macroblock at the same position in the reference frame is below some threshold

Exhaustive Search
- Search every possibility over a specified range for the best matching macroblock
- Computationally intensive
- Requires the matching criteria to be evaluated many times

Measuring the Difference
- The difference between two macroblocks can be measured using various techniques
- A simple measure is the Mean Absolute Difference (MAD)

Residuals
- Even using motion compensation there are usually differences between the best candidate and the original. We cannot expect to find exact matching macroblocks so instead look for close matches.
- The differences between the macroblock being coded and its best match are known as residuals.
- SAD - Sum of absolute differences
- MSE - Mean squared error

Key Ideas in MPEG Video Encoding
- Work on each macroblock independently
- Motion compensation done at the macroblock (16x16 pixels) level
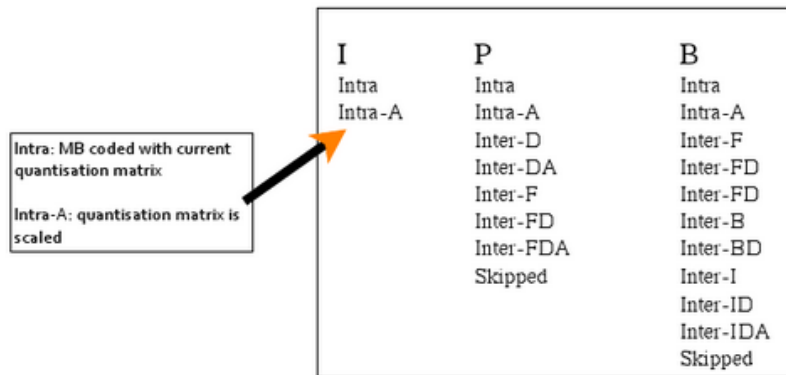- Image coding is performed at the block (8x8 pixels) level

Open & closed GOP
- A closed GOP can be decoded without reference to another GOP
- An open GOP uses an anchor frame from another GOP
- Open GOPs generally provide slightly better compression than do closed GOPs of the same structure and size
  - They allow an extra B-frame in the GOP pattern
  - A closed GOP contains one more P-frame than an open GOP of the same length
  - P-frames generally require more bits that B-frames so the open GOP achieves slightly better compression

MPEG-1 Features
- MPEG-1 allows motion vectors to be of sub-pixel precision
- Bilinear interpolation generates the needed values at half-pixel locations

Macroblock Types

I	P	B
Intra	Intra	Intra
Intra-A	Intra-A	Intra-A
	Inter-D	Inter-F
	Inter-DA	Inter-FD
	Inter-F	Inter-FD
	Inter-FD	Inter-B
	Inter-FDA	Inter-BD
	Skipped	Inter-I
		Inter-ID
		Inter-IDA
		Skipped

Intra: MB coded with current quantisation matrix

Intra-A: quantisation matrix is scaled

Macroblock Types

▸ F: forward MC

▸ B: bi-directional MC

▸ D: the DCT of prediction error is coded

▸ A: adaptive quantisation

▸ FD: transmit motion vector and DCTs of prediction error

▸ FDA: a motion vector, DCTs of error, and adaptive quantisation

▸ I: Interpolated prediction with motion prediction

Rotoscoping
- The process of manually tracing shapes through a captured image sequence
- Use to;

- ○ Create mattes to place an actor into a different scene
- ○ Replace a real prop with a CGI element
- ○ Apply image filters selectively over parts of a video frame
- ○ Create 2D animation from captured video
- Modern uses: effects painting, matte creation, motion tracking.
- However, difficult to define area (line) directly around object in some situation (think of why green screen needed)

Wave Properties of Sound & Behaviours
- Reflection (bouncing)
- Refraction (change of angle when entering a medium with different density)
- Diffraction (bending around an obstacle)

Wave Properties

▶ Wavelength is the distance between one part of a wave and the same part of the next wave

▶ Frequency is the number of waves moving past a point in one second

▶ Amplitude is the strength or power of a wave signal
  - Higher amplitudes are interpreted as a higher volume
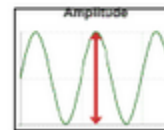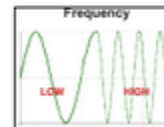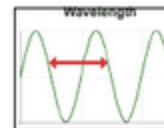
- Pitch is a measure of how high or low a sound is
- Volume is the amount of sound energy reaching your ears

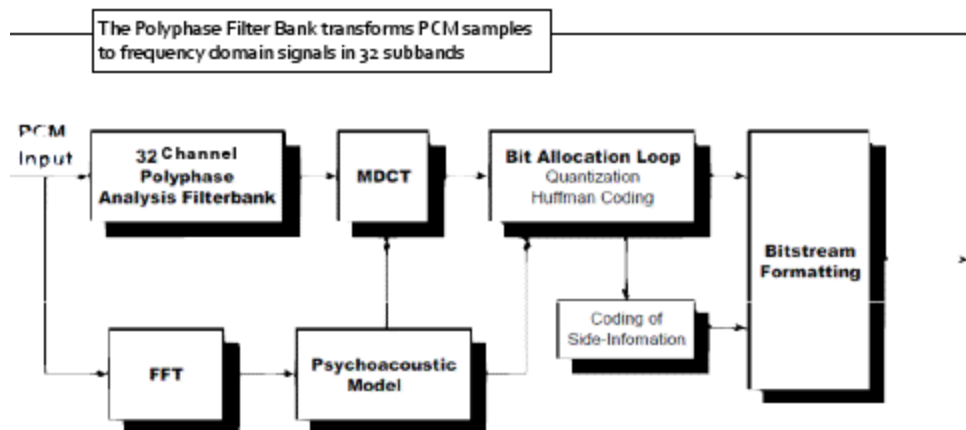| Threshold of hearing	0
Rustle of leaves	10
Very quiet room	20
Average room	40
Conversation	60
Busy street	70
Loud radio	80
Train through station	90
Riveter	100
Threshold of discomfort	120
Threshold of pain	140
Damage to ear drum	160

- Bit Depth describes the accuracy of the audio data, increasing the number of bits increases the quality of the audio
- Sampling Rate: The more samples taken per second the higher the accuracy
- A side effect of digitisation is to introduce noise (quantisation noise)

- The human ear can hear from about 20Hz (a very deep rumble) to 20kHz

Design Limitations of MP3
- Several limitations that cannot be overcome by using a better encoder
- Bitrate is limited to a maximum of 320 kbits/s
- Time resolution can be too low for highly transient signals
- Joint stereo is done on a frame-to-frame basis

MP3 Encoder Architecture

The Polyphase Filter Bank transforms PCM samples to frequency domain signals in 32 subbands

PCM Input → 32 Channel Polyphase Analysis Filterbank → MDCT → Bit Allocation Loop (Quantization, Huffman Coding) → Bitstream Formatting

FFT → Psychoacoustic Model

Coding of Side-Information

Types of Distribution
- Files are downloaded and will not display until all content is downloaded
    - Progressive download enables a fast start but playback pauses if content is not available. Content is seen as a single resource.
    - The size of each chunk is sent with the chunk, client knows when it has finished receiving that chunk. Transfer is terminated by a final chunk of length zero.
- Streaming
    - Content is seen as a sequence of segments.

Streaming Overview
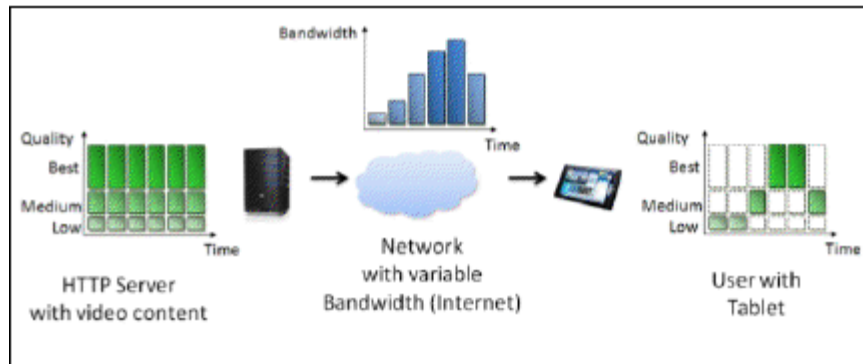
Streaming Systems
- ● HTTP LIve Streaming (HLS) - adaptive streaming protocol
- ● Example System
    - ○ ffmpeg can encapsulate content in an HLS compatible format
    - ○ Takes an input file and generates a series of equal-length files
    - ○ Generate an index file (playlist)

Challenges of Media Delivery
- ● High bandwidth requirements
- ● Non-uniform frame sizes
- ● Variable bit rate audio and video encoding

Handling Variability
- ● Constant bitrate encoding of each stream
- ● Multiplexing of variable rate streams
- ● Selective discard of packets / frames in the stream

Routers
- ● A router schedules packets
- ● Discard policy if packet arrives to full queue
    - ○ Tail drop: drop arriving packet
    - ○ Priority: drop / remove on priority basis
    - ○ Random: drop / remove randomly
- ● Scheduling policy: transmit highest priority queued packet

Packet Loss
- ● packet loss has an even greater effect than packet jitter

Flexible Macroblock Ordering (FMO)
- ● Allows mapping of macroblocks to slice groups
- ● A slice group may contain several slices
- ● Spatially co-located image areas can be interleaved in different slices to increase the probability of concealing lost macroblocks
- ● Can map a region of interest into a separate slice group

Data Partitioning results in "graceful degradation"

Error Concealment Objectives
- ● The decoder generates a representation of lost data which should match as closely as possible the lost information within a manageable level of complexity

Estimating the Motion Vector
- ● Use the;
  - ○ Motion vector of the corresponding macroblock in the previous decoded frame
  - ○ Average of the motion vectors from adjacent macroblocks
  - ○ Median of the motion vectors from adjacent macroblocks

Other
- ● AC Coefficient Encoding - L3
- ● Video Interleaving - L13
- ● Video packetisation modes - L13
- ● Hybrid Concealment - L13