Quiz

On digital video.

Previous Quiz

On OSI Model.

Number	Question/answer
1	List seven layers of OSI model: physical, data link, network, transport, session, presentation, application.
2	OSI stands for Open Systems Interconnection
3	What defines the use of CAT 5e as the cable to use? The answer is which OSI layer. This is the physical layer.
4	Media Access Control (MAC) layer is located where in the OSI? The data link layer.
5	Transmission Control Protocol is located where in the OSI? Transport layer.
6	Which layer deals with just bits? Physical.
7	Packets appear at which layer? Network (and transport).
8	What part of the data stream identifies a particular layer? Packet header.
9	What is the OSI model related to in the video world? The elementary stream.
10	The data stream is made up of packets.

Next Quiz

On MPEG compression, on Tuesday.

MPEG Compression

DCT

Pixel samples presented to DCT conversion in 8x8 blocks. Why work on an 8x8 block instead of just take the first 64 pixels in order in which they appear in the serial data stream? There's less likelihood of change in a square block than there would be across an entire line of pixels.

The spectral selection option encodes the lower-frequency DCT coefficients first (to obtain an image quickly), followed by the higher-frequency ones (to add more detail). The successive approximation option encodes the more significant bits of the DCT coefficients first, followed by the less significant bits.

When the camera acquires an image at the CCD blocks, it is either interlaced or progressive at that time.

We'd have to research what actually takes the pixel data and groups them into the 8x8 blocks. How much buffering is done for frames? There is a frame store block in the block diagram of the MPEG encoder that does the DCT. Some video is stored, likely that I frames are stored so that motion vectors on P and B frames can be calculated based on the previous I or P frames.

I Frame

Intra pictures. Are intra coded; coded and reconstructed without reference to other frames. Called anchor frames; they provide access points to where decoding can begin. Reduction of spatial (within the frame) redundancy, not temporal (between frames) redundancy. Still reduces bandwidth significantly even just with spatial reduction.

8x8 DCT. Arbitrary weighting matrix for coefficients. Differential coding of DC-coefficients (the upper left corner of the matrixes). Uniform quantization. Zig-zag-scan, run-level coding. Entropy coding. Unfortunately, not quite JPEG.

Encode Y, R-Y, and B-Y simultaneously. No I or Q. Colors are encoded within macroblocks since color is scanned half as often (4:2:2 or 4:2:0).

One DC coefficient in upper left corner of matrix; all remaining coefficients are AC coefficients. DC is the reference level, and the AC coefficients are referenced back to the DC coefficient. The DC coefficient is the average picture level (or the equivalent in color).

Difference between MPEG and motion JPEG? Motion JPEG has only I frames. Where do we use motion JPEG in broadcast? The cameras (newer ones). Our BTS video server stores in motion JPEG, so video can be edited. What rate is motion JPEG is at (4:2:2)?

P Frames

Predictive pictures. Can use the previous I or P frame for motion compensation. May be used as a reference for further prediction. Each block can be either predicted or intra-coded.

In the receiver:

- the P frames are forward predicted from the last I frame or P frame
- it is impossible to reconstruct P frames without the data of another frame (I or P)
- by reducing spatial and temporal redundancy, increased compression over I pictures

Co-sited refers to sampling Cr and Cb at the same time as sampling Y for a given pixel. The next pixel just gets its Y sampled.

For 4:2:0. Diagram 1. On one row, chrominance is sampled every other pixel. On the next row, chrominance is not sampled at all.

The order gets reworked so that we get Cb, Y, Cr, Y, Cb, Y, Cr, Y, so that Y appears in every other data point.

Coding of P pictures

Motion compensated prediction from an encoded I picture or P picture. Half-pixel accuracy of motion compensation, by linear interpolation (or is this bilinear interpolation as stated on the slide?). One displacement vector per macroblock. Differential coding of displacement vectors. Coding of prediction error with 8x8-DCT, uniform threshold quantization, zig-zag-scan as in I pictures.

B Frames

Highest degree of compression. Bidirectionally predictive pictures. Each block can be forward, backward, or bidirectionally predicted or intracoded from the last/next I frame or P frame. Two other frames necessary to reconstruct them. P frames and B frames are referred to as **inter coded** frames.

Coding of B pictures. Motion-compensated prediction from two consecutive P or I pictures. Either:

only forward prediction (1 vector per macroblock)

only backward prediction (1 vector per macroblock)

average of forward and backward prediction aka interpolation (2 vectors per macroblock)

Half-pixel accuracy of motion compensation, bilinear interpolation. Coding of prediction error with 8x8 DCT uniform quantization, etc.

Backward Prediction from a Future Frame?

Encoder must reorder the pictures. Any picture the B frame references must be sent first. Introduces a reordering delay related to the number of consecutive B pictures.

Display order: I1 B2 B3 B4 P5 B6 B7 B8 P9

Bitstream order: I1 P5 B2 B3 B4 P9 B6 B7 B8

This allows the decoder to be made more cheaply since it doesn't have to buffer as many frames in the bitstream order as it would have to in the display order. The complexity and expense is put into a much smaller number of decoders.

Frames and Blocks

A P or B frame may have individual blocks (or maybe even all the blocks in the entire frame) that are I encoded. Maybe the change from the corresponding block in the previous reference frame is too great to capture efficiently in prediction, so go to an intra coded block.

Group of Pictures

The sets of frames. As an example, a frame sequence I B B P B B B P gets transferred as I P B B B P B B B. Frame numbers are in the frame header. Ascending frame number sequence.

Discrete Cosine Transform

Of the image compression techniques available, transform coding is the preferred method. Images tend to compact their energy in the frequency domain, making compression in this domain much more effective. Transform coding is simple the compression of the images in the frequency domain. Transform coefficients are used to maximize compression.

The DCT is an example of transform coding. It relocates the highest energies to the upper left corner of the image. The lesser energy or information is relocated into other areas.

The DCT is fast. It can be quickly calculated and is best for images with smooth edges like photos with human subjects.

The inverse DCT (IDCT) can be used to retrieve the image from its transform representation.

The two dimensional DCT is just a one-dimensional DCT applied twice, once in x direction and once in y direction. Applying DCT to 8x8 matrix yields an 8x8 matrix of weighted values corresponding to how much of the DCT function is present in the original image. **For most 8x8 images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT. The lower right values represent higher frequencies, and are often small**.

Almost all values are equal to zero.

Quantization

DCT'd matrix has larger values in upper left corner area. The other stuff is probably not needed in order to show a good picture. Object of quantization is to get a lot of zeroes in the quantized DCT'd matrix. The nonzero coefficients appear in the upper left corner area. This makes the matrix more amenable to compression via run-length encoding. The tradeoff is the irreversible loss of information as the result of the quantization process.

The DCT always tends to compute zeros. This effect is assisted by quantization which zeros small values.

Value in upper left corner is called the **DC value**. Abbreviation for direct current, refers to AC theory where an AC can have a DC component. In DCT the DC value determines the average brightness of the block.

Main Encoding Path

Encoder decodes the quantizer output so it can feed the decoded information back to the subtractor.

The Need for a Local Decoder

Motion compensated predictor requires signal on which to base its prediction. Signal delayed internally in predictor; important to see that the predictor is in a loop which must be replicated in the remote decoder. The predictor must base its prediction on a signal that is available in the remote decoder. This is why we have to decode in the encoder. An adder takes the decoded prediction error and adds it back in to a suitable delayed version of the prediction to produce a locally decoded signal.

Motion Vectors

I frame shows a triangle. P frame shows same triangle at another position. Prediction means to supply a motion vector which declares how to move the triangle on the I frame to obtain the triangle on the P frame. This motion vector is part of the MPEG stream and it is divided positive or negative horizontal or vertical. A positive value indicates motion rightward or downward. A negative value indicates motion left or upward. The parts of the motion vector are in a range of -64 to +63. The referred area can be up to 64x64 pixels away.

Both luminance and chrominance are moved by the motion vector.

A diagram in the slide shows the red rectangle is shifted and rotated by 5 degrees to the right. A simle displacement of the red rectangle will cause a prediction error. The MPEG stream must contain a matrix for compensating this prediction error.

The reconstruction of inter coded (P & B) frames proceeds in two steps: application of the motion vector to the referred frame, adding the prediction error compensation to the result.

Prediction error compensation requires fewer bytes than a whole block. DCT compression is then applied to the prediction error, further decreasing its size.

Frame divided into macro blocks of 16x16 pixels, and each macro block has its own motion vector.

If contradictory motion occurs. If prediction error is too big the coder can decide to intra-code a macro block. Similarly each macro block in B frames can be either:

- forward predicted
- backward predicted
- forward and backward predicted
- intra-coded

Motion Estimation and Prediction Selection

Skip this slide.