Homework

Two-page handout. Contains a diagram and explanation of MPEG-1 encoding and decoding.

Write a 3-page paper on how MPEG encoding and decoding works.

The handout leaves a lot of questions unanswered. Do some research to answer these questions and incorporate this into your paper.

Due Tuesday.

Find out what *pel* really means. It is associated with MPEG.

Europe and High Definition Television

By the way, Europe finally going to go high definition over the air, but it will be pay-per-view only.

MPEG Encoding (continued)

What goes into the DCT circuitry? It is the luminances or chrominances in an 8 by 8 pixel block.

What does the DCT do? It changes numbers in the amplitude domain to numbers in the frequency domain. Why do we do this? It transforms the data into a more compressible form. We end up with levels within the matrix.

The upper left value is the average picture level within the 8x8 block. It is the DC coefficient. It is the reference value for all other values within the 8x8 block.

Typically won't see much of a change of APL within an 8x8 block. Our systems don't provide the bandwidth to support extreme changes. Just using 8x8 blocks is a form of compression.

The eye pattern in SDI is rounded because we don't have the bandwidth to support the higher harmonics necessary for clean square waves.

Push just about all data, not just frames, through the DCT.

Quiz Review

Which of the three pictures is used for random access points within the video stream?

I. Why do we use it for cutting? It is a complete frame, not dependent on any other frame for its data.

Which of the three pictures is used at scene cuts?

I. Why? Same reason as above. It is a complete frame. MPEG is a delivery codec, not an editing codec.

Which of the three pictures is forward predicted only?

P. A P frame can reference either an I frame or another P frame, but it will reference an I frame if the I frame is more recent.

List the three pictures in order from least to most compression.

I, P, B.

By averaging two pictures in the backward and forward methods, we can realize an improvement in what?

Compression efficiency and perceived picture quality. That wasn't what he was looking for, but he gave people credit for that.

Noise. Improving perceived picture quality is just another way of referring to noise reduction. The way this could happen is that if one value of the two being averaged is garbage because it represents noise, then the average of the two values (good value and garbage value) is much closer to the good value than if we just used the garbage value. It's much less likely that both of the two values being averaged will both be garbage. This is the same reason why scientific research studies take multiple readings for a given set of conditions, to minimize the effect of a single incorrect reading.

Noise is the enemy of compression, whether spatial or temporal.

GOP stands for

Group of pictures. Sometimes called gawp other times called gee oh pee.

A GOP must have at least one ___ picture?

I. The I picture is the direct or indirect reference for all the others.

Which picture(s) concentrates on the removal of spatial redundancies?

I.

Which picture(s) concentrates on the removal of temporal redundancies?

P and B.

What is the formula associated with 601?

(13.5 MHz for Y + 6.75 MHz for R-Y + 6.75 MHz for B-Y) * 10 bits per sample = 270 Mbits/second.

More on MPEG Encoding

Main Encoding Path

The heart of the MPEG encoder is actually the **subtractor**.

Here, the motion compensated prediction for a picture subtracted from the input picture, which has been presented in the correct order for encoding according to the desired GOP structure.

The subtractor is disabled (or the prediction set to zero) for I-frames. The output of the subtractor, *which is the prediction error* (or the picture in the case of I-frames), is passed to the DCT, whose output coefficients are quantized.

The difference information represents the motion between frames.

Need periodic I frames in order to ensure that encoding errors don't build up too much, and so that people tuning in can get in sync.

The output table of the quantizer is zigzag scanned, run-length and variable-length coded. The variable rate bitstream is buffered and the occupancy of the buffer is monitored by the rate control block which controls the quantizer.

Block Diagram of an MPEG Encoder

There is a decoder within the encoder, that inverse quantizer and inverse DCT. The output of this gets fed to the subtractor.

The decoder provides the data that is available at the receiver/monitor, since the receiver/monitor does not have the original data to work with. This may make the subtractor more accurate since it is working with the same data that the receiver/monitor will have access to. Quantization is a lossy process. DCT is a lossy process also (right?).

The need for a local decoder

The motion compensated predictor requires a signal on which to base its prediction. This signal will be delayed internally in the predictor. It is important to see that the predictor is in a loop *which must be replicated in the remote decoder*. The predictor must therefore base its prediction on a signal which is also available in the remote decoder. The locally decoded signal in the encoder consists of blocks to undo the encoding stages of quantizer and DCT. This is followed by an adder which takes the decodd prediction error and adds it back in (missed some of this) .

A P-frame is encoded relative to the past reference frame. A reference frame is a P or I frame. The past reference frame is the closest preceding reference frame.

Each macroblock in a P-frame can be encoded either as an I macroblock or a P macroblock. An I macroblock is encoded just like a macroblock in an I frame. A P macroblock is encoded as a 16x16 area of the past reference frame, plus an error term.

To specify the 16x16 area of the reference frame, a motion vector is included. A motion vector (0, 0) means that the 16x16 area is in the same position as the macroblock we are encoding. Other motion vectors are relative to that position. Motion vectors may include half-pixel values, in which case pixels are averaged. The error term is encoded using the DCT, quantization, and run-length encoding.

A macroblock may also be skipped, which is equivalent to a (0, 0) vector and an all-zero error term. The search for good motion vector (the one that gives small error term and good compression) is the heart of any MPEG-1 video encoder and it is the primary reason why encoders are slow.

A B-frame is encoded relative to the past reference frame, the future reference frame, or both frames. The future reference frame is the closest following reference frame (I or P). The encoding for B-frames is similar to P-frames, except that motion vectors may refer to areas in the future reference frames. For macroblocks that use both past and future reference frames, the two 16x16 areas are averaged.

P picture represents anything that moves from the preceding P or I picture.

What's a motion vector?

A bi-dimensional pointer that tells the decoder how much left/right and up/down, from the position of the macroblock, is located the prediction macroblock in the reference field or frame.

Motion vectors have a **half-pel resolution**. Pel is *picture element*; pel and pixel mean the samething. An interpolation process is necessary to get the prediction. It must be noticed that the same motion vector is applied both to luminance and, after being scaled, to chrominance.

What does it mean “after being scaled”? The chrominance is subsampled (4:2:0)

Sports is the most difficult subject for MPEG because of so much motion.

What's motion estimation?

The process, performed by the coder, that should find the motion vector pointing to the best prediction macroblock in a reference frame or field.

The *absolute error* or the *mean squared error* is applied to the macroblock (ora part of it), such technique is called **block matching** and is the most used in video coding.

In general every possible prediction in a given range is evaluated, so we speak about **full search**.

Unfortunately the computation's complexity is proportional to the **search area** and can be quite heavy, and on the other hand the search area has to be wide enough to include every movement.

It must be noticed that the capability to perform a good motion estimation is a key point for the quality of a coder.

B frames are calculated based on the 16x16 macroblocks. The 16x16 is sampled in the P frame and is moved in the B frame. P frame encodes motion of the I frame in 8x8. The B frame assesses the movement of 16x16 macroblocks and encodes the *movement* seen into the B frame. B frames can be just motion vectors. The major advantage to B frames are the motion vector encoding capabilities.

The prediction menu for frame pictures

Frame-based prediction

A single motion vector for the whole macroblock. It's used when movements between two fields are insignificant. It's the only possible choice for progressive images.

Field-based prediction

Two different motion vectors. One for the samples belonging to the first field and one for those belonging to the second field. It's used when movements between the two fields are important.

Quantization

Quantization techniques generally compress by compressing a range of values to a single quantum value. By reducing the number of discrete symbols in a given stream, the stream becomes more compressible. The quantization table is already set up. Transmitted periodically. Is a matrix used in matrix division. Divide the DCT matrix by the quantization matrix. This is another form of compression. The coefficients in the lower right of the quantization table look larger than the ones on the upper left, so it tends to zero out values in the lower right that aren't fairly large. Not sent as ancillary data, the quantization table is sent as part of the data stream, periodically.

Therefore the MPEG compression could be regarded as lossless, except that the transformed values are quantized.

The DCT supplies values up to 2047 (for the DC values).

In MPEG there is a quantization matrix which defines a different quantization value for every transform value depending on its position. These values are divided by a certain value supplied by the quantization table.

The decoder multiplies the result by the same value. Of course the **result differs from the original value**. But because of some properties of the human eye the errors are not visible.

Entropy Coding

Chaos – disorder. It is the loss of information in a transmitted signal or message. After quantization, nonzero coefficients are further encoded using an entropy coder such as a Huffman coder. In entropy coding schemes, the more frequent values are represented with shorter codes and the less frequent values with longer codes.

First, take a look at the data that needs to be compressed and create a table that lists how many times each piece of unique data occurs. Now assign a very small code word to the piece of data that occurs most frequently. Continue until all unique pieces of data are assigned unique code words. Data occurs most frequently gets a small code word, less frequent gets a larger code word, resulting in space savings.

Run-length Encoding

Zero coefficients can be efficiently encoded using **run-length encoding**. Instead of transmitting all the zero values one by one, run length coding simply transmits the total number of the current run of zeros. The result is a compressed representation of the original image.

Run Length Coding

First the DC quantized coefficient receives its own Huffman code – up to 11 bits. The linear stream of quantized frequency coefficients is converted into a series of [run, amplitude] pairs. [Run] indicates the number of zero coefficients, and [Amplitude] the nonzero coefficient that ends the run.