Look at MPEG Encoded Still Images

If quantizer step size for AC coefficients is 25, picture is probably acceptable with some blocking. Coefficient of 100 gets blocks larger and noticeable. Coefficient of 200 is not really acceptable.

Why does this happen? This step size is the divisor in the matrix used for dividing the DCT matrix to create the quantized matrix. Higher coefficients zero out more information, which provides more compression. DC coefficient in the quantized matrix in the upper left reflects average picture level. AC coefficients are differences from the average picture level.

Look at MPEG Encoded Images of Moving Scene (Sports)

Show concept of motion vector tracking a portion of the image that moves from one set of blocks to another.

MPEG Encoding of Movement using Triangle Example

The I frame shows a triangle. A P frame shows the triangle at another position. Prediction means to supply a *motion vector* which declares how to move the triangle on I frame to obtain the triangle in P frame.

This *motion vector* is part of the MPEG stream and it is divided positive or negative horizontal and vertical. A positive value indicates motion right or downwards. A negative value indicates motion left or upwards.

The parts of the motion vector are in a range of -64 to +63. The referred area can be up to 64x64 pixels away.

MPEG Encoding of Movement using Rotated Rectangle

Another figure shows a red rectangle shifted and rotated by 5 degrees to the right. A simple displacement of the red rectangle will cause a *prediction error*. The MPEG stream must contain a matrix for compensating this prediction error.

The I frame is in a way of thinking the background; the I frame contains the compressed background. The P frame contains motion vectors that are based on the information contained in the I frame being referenced by the P frame.

MPEG 2 doesn't deal with objects being moved; MPEG 4 does deal with objects.

Reconstruction (Decoding) of the Rotated Rectangle Frames

The reconstruction (decoding) of inter coded frames (P and B) proceeds in two steps.

1. Application of the motion vector to the referred frame.
2. Adding the prediction error compensation to the result.

Prediction error compensation requires fewer bytes than a whole frame. DCT compression is then

applied to the prediction error, further decreasing its size.

The different meanings of the two + signs.

- The first means adding the motion vector to the x-, y- coordinates of each pixel.
- The second means adding an error value to the appropriate pixel.

Contradictory Motion

What if some parts move to the left and others move to the right? The motion vector isn't valid for the entire frame. The frame is divided into macro blocks of 16x16 pixels. Every macro block has its own motion vector. This does not avoid contradictory motion, but it minimizes its probability. What does this mean?

The motion vector is calculated just on luminance. The luminance vector is then applied to the chrominance.

If Contradictory Motion Occurs

One of the greatest misunderstandings of the MPEG compression technique is to assume that all macro blocks of P frames are predicted. If the prediction error is too big the coder can decide to intra-code a macro block. Similarly, the macro blocks in B frames can be forward predicted or backward predicted or forward **and** backward predicted or intra-coded. These frames are still P or B frames even though they might be completely intra-frame encoded. Won't stop the GOP and start a new one; will maintain the standard of 15 frame sequence even though this P or B frame is encoded like it was an I frame.

Consequences of Encoding

We encode to save bandwidth, but if use it up, picture starts to fall apart. Cable is very susceptible to this. Cable tries to maximize everything; they get paid money for every station they carry. Even a moving mouth can exceed the bandwidth limit.

One previous test question has been: if broadcasting Super Bowl, and boss decided wants three different camera angles broadcast simultaneously, can you do it? Three camera angles on three simultaneous channels of SDTV. Might not be able to do it if all three cameras pan over complex scenes at the same time.

(Skips prediction selection block notes).

Results of motion estimation and prediction selection processes need to be transmitted to the decoder as part of the MPEG bit stream. Like the quantized DCT coefficients, motion vectors and prediction modes are also variable-length and run-length encoded.

Break

Video Elementary Stream

This is important stuff.

The output of an MPEG-2 video encoder is known as an Elementary Stream (ES). If such a stream is held on a server, it contains all the information necessary to produce a decoded video signal.

Macro Blocks

The lowest-level entity in the stream is a coded block of DCT coefficients. Each block is terminated by an end-of-block code and the four luminance and two chrominance blocks in the macroblock are simply concatenated.

The coded blocks are preceded by a macroblock header which contains all the control information belonging to the macroblock: spatial address, motion vectors, prediction modes, field/frame DCT mode, quantizer step size. The result is a coded macroblock.

The above describes the macroblock portion of an ES.

Coded macroblock is a macroblock header, coded blocks for Y, Y, Y, Y, U, V, and the end-of-block code.

In the macro, they identify field or frame (the field/frame DCT mode flag). This tells us which zigzag pattern we are using. Important to know. Maybe this flag name should refer to the zigzag pattern rather than field or frame. We should check this. Apparently there are numerous (over 30) DCT modes, and this may refer to that instead. Where does the zigzag pattern get specified?

Slices and Pictures

A slice is a set of consecutive macroblocks preceded by a slice header.

The slice is the smallest entity in an MPEG ES on which synchronization can be attained. This is important. This has to do with the decoder getting lost, and being able to find where it is in the encoded MPEG 2 bit stream.

Each slice has a slice header and concatenated coded macroblocks.

The slice header contains a unique start code which cannot be duplicated elsewhere in the bit stream.

There is usually one slice for each row of macroblocks, but in error-prone channels or in other applications where quick resynchronization is very important, it is possible to have several slices per row, even one slice per macroblock.

Slices are then grouped together into coded pictures. Encoded pictures have a picture header followed by concatenated slices.

What are error-prone channels?

GOP Header

Picture header again has a unique start code and contains picture-specific control information, for example whether the picture is a field or a frame, its picture coding type (I, P, or B), and the quantizer weighting matrices.

There is then the option of grouping pictures together and preceding them with a GOP header to make a group of pictures. Usually 15 pictures in a GOP; ATSC recommends 15 pictures in a GOP.

The GOP header contains a time code and some flags relating to editing. Why have flags for editing? What one thing? There is one I frame in the GOP.

Whitaker's book says that GOPs are optional in both MPEG-2 and ATSC digital video, in the chapter on ATSC digital television. We don't know why Whitaker said this. The GOP header flag relating to editing got moved someplace else (we will cover that later); maybe this is why GOP is optional.

Sequence

Finally, GOPs or pictures are grouped into a coded sequence. Sequence header, GOPs or pictures.

The sequence header can be repeated within the sequence. It contains basic information needed before decoding can begin, such as picture size, frame rate.

Because of its importance, the sequence header is repeated, maybe about twice a second.