Turn in Papers on MPEG Encoding and Decoding

Discussed what we learned.

Video Compression

First part is a review.

What is compression? Employs redundancy in data.

Temporal: in one-dimensional data, one-dimensional signals, audio, etc.

Spatial: correlation between neighboring pixels or data items

Spectral: correlation between color or luminescence components. This uses the frequency domain to exploit relationships between frequency of change in data.

Psycho-visual: exploit perceptual properties of the human visual system.

Categorizing

Compression can be categorized in two broad ways.

Lossless Compression

Where data is compressed and uncompressed without loss of information. Referred to as **bit-preserving** or **reversible compression** systems.

Lossless compression frequently involves some form of entropy encoding and are based on information theoretic techniques.

Lossy Compression

To obtain the best possible *fidelity* for a given bit-rate. Video and audio compression techniques are most suited to this form of compression.

Lossy compression uses source encoding techniques that may involve transform encoding, differential encoding, or vector quantization.

MPEG

A few tokens amounting to only 100 bits can represent an entire block of 64 samples transparently (as in not being able to tell the difference) which otherwise normally consumes 64 * 8 or 512 bits.

Early Video Compression in Analog Domain

Composite video.

Bandwidth limiting to 4.2 MHz. This is lossy compression.

Vestigial sideband filtered down – this removes redundant information, so it is lossless.

R-Y and B-Y compression.

Rounding Considerations

At some point, a result must be rounded to some lower precision (16 bits to 8 bits or 32 bits to 16 bits) in order to realize a cost effective hardware implementation.

There are several rounding techniques: truncation, conventional rounding, error feedback rounding.

Truncation

Drops any fractional data during each rounding operation. As a result, after only a few operations, a significant error may be introduced. This may result in contours being visible in areas of solid colors.

Conventional Rounding

Uses the fractional data bits to determine whether to round up or down. If the fractional data is 0.5 or greater, round up, else round down.

Error Feedback Rounding

Follows the principle of never throwing away anything. Stores the residue of a truncation and adds it to the next video sample. Substitutes less visible noise-like quantization errors in place of contouring effects caused by simple truncation. People expect this kind of noise and it is more pleasing. Diagram 1.

Dithering

A form of noise or erroneous signal or data which is deliberately added to data for the purpose of minimizing quantization error. Routinely used in processing of both digital audio and digital video data.

Dynamic Rounding

Dithers the LSB according to the weighting of the discarded fractional bits. Original data word divided into two parts, one representing resolution of final output word, one dealing with remaining fractional data. Fractional data compared to output of random number generator equal in resolution to the fractional data. Output of comparator is a 1-bit random pattern weighted by the value of the fractional data, and serves as a carry-in to the adder. In all instances, only one LSB of the output word is changed, in a random fashion. Diagram 2.

JPEG

Major steps include:

- DCT (Discrete Cosine Transformation)
- Quantization

- Zigzag scan
- DPCM (differential pulse code modulation) on DC component encodes the difference from previous 8x8 blocks
- RLE (run-length encoding) on AC components
- Entropy coding

This shows how MPEG took a lot of what was incorporated into JPEG.

Signal analysis. Each event is then assigned a code.

Quantization. Commonly occurring events are assigned few bits.

Variable length encoding. Rare events will have codes with more bits.

Four Methods For Compression

Discrete Cosine Transform (DCT)

Vector Quantization (VQ)

Fractal Compression

Discrete Wavelet Transform (DWT)

DCT

A lossy compression algorithm that samples an image at regular intervals. Analyzes the frequency components present in the sample. Discards those frequencies which do not affect the image as the human eye perceives it. Actually is the quantization that is lossy that is used as part of this.

MPEG-1

Designed for up to 1.5 Mbit/sec. Standard for compression of moving pictures and audio. Based on CD-ROM video applications, and popular standard for video in Internet, transmitted as .mpg files. In addition, level 3 of MPEG-1 is the most popular standard for digital compression of audio, known as MP3.

Compare with motion JPEG, that creates a bunch of I frames. Takes up a lot more bandwidth, but better for recording, and great for editing.

MPEG-2

Designed for between 1.5 and 15 Mbits/sec. Standard on which DTV set top boxes and DVD compression based. Based on MPEG-1 but designed for the compression and transmission of digital broadcast TV. Most significant enhancement from MPEG-1 is its ability to efficiently compress interlaced video. MPEG-2 scales well to HDTV resolution and bit rates, obviating the need for an MPEG-3 (which was originally designed for high definition TV).

MPEG-4

Standard for multimedia and Web compression. Based on object-based compression, similar in nature to VRML. Individual objects within scene tracked separately and compressed together to create an MPEG-4 file. Results in very efficient compression that is very scalable, from low bit rates to very high. Also allows developers to control objects independently in a scene, and therefore introduce interactivity.

MPEG-7

This standard, currently under development, is also called the Multimedia Content Description Interface. The hope is the standard will provide a framework for multimedia content that will include information on content manipulation, filtering, and personalization, as well as the integrity and security of the content. Contrary to previous MPEG standards, which described actual content, MPEG-7 will represent information about the content.

This isn't a compression standard; relies on MPEG-2 for the compression standard.

MPEG-21

Also called the Multimedia Framework. Attempts to describe the elements needed to build an infrastructure for the delivery and consumption of multimedia content, and how they will relate to each other.

DV

DV is a high-resolution digital video format used with video cameras and camcorders. The standard uses DCT to compress the pixel data and is a form of lossy compression. The resulting video stream is transferred from the recording device via FireWire (**IEEE 1394** – important to know and have in black book), a high-speed serial bus capable of transferring data up to 50 MB/sec. FireWire is Apple's trademark name, ILink is Sony's trademark name.

Codec

A video codec is software that can compress a video source (encoding) as well as play compressed video (decompress). Coder/decoder or compress/decompress.

How does MPEG Achieve Compression?

Spatial correlation: transform coding with 8x8 DCT. Transforms spatial domain to frequency domain.

MPEG-2 Levels and Profiles

Profiles determine the set of compression tools, compromise between compression rate and decoder cost.

Simple profile: simplify the encoder and decoder, higher bit-rate, no bidirectional prediction (B pictures)

Main profile: best compromise between rate and cost, use all three image types (I, P, and B)

SNR scalable profile: allow the quantization accuracy of a basic quality picture to be enhanced

Spatially scalable profile: allow the spatial resolution to be enhanced

High profile: for HDTV broadcast applications

Don't memorize levels and profiles, just know we use main at main (for SDTV), main at high (for 720), and high at high (for 1080).

There is an MPEG-2 4:2:2 profile at high (1920 x 1088, 300 Mb/sec) and main (720 x 608, 50 Mb/sec) levels.

GIGO (garbage in, garbage out)

Video Production Formats

Shows a block diagram. Diagram 3, but Steve printed out the slide and handed it out to everyone. Outputs are serial digital composite (143.8181 and 177 Mb/sec) (D2) and serial digital component (270 Mb/sec) (601).

Typical Video Data Rates

10-bit rec 601 270 Mbps

8-bit Rec 601 216 Mbps

8-bit rec 601 (active only) 167 Mbps

Digital Betacam ~90 Mbps

MPEG-2 4:2:2P@ML 15-50 Mbps

MPEG-2 MP@ML 1.5 – 15 Mbps

MPEG-1 constrain. Param. 0.5 – 1.8 Mbps

H.261 videoconferencing 64 kbps – 1.5 Mbps

H.263 videoconferencing

MPEG-2 Video Compression

Pre-processing: clean up pictures and prepare video samples Remove noise, which looks like signal since it is always changing, and thus will participate in compression and reduce compression ratios.

Temporal compression (interframe): compresses the data from multiple frames

Spatial compression (intraframe): compresses the data within one frame, similar to JPEG

Rate control: provides constant bit rate for constant or nearly constant quality

Important Points about MPEG

Only specifies bitstream syntax and decoding. Encoding algorithms not defined 0 they are open to

invention and generally proprietary, and future improvements are compatible with all decoders.

Asymmetric Compression

Encoder is very complex. Encoder contains a decoder. Decoder definition emphasizes low complexity and cost.

Pre-Processing of Video

Required

Decode from composite to component. Why? To pick up the 4:2:0 in the component domain, where we will get the Y, R-Y, B-Y.

Produce correct picture size.

Reduce 10-bit samples (601) to 8-bit samples (MPEG-2).

Convert to 4:2:0 sampling (entertainment quality).

Optional

Noise reduction.

Other picture clean up (pre-filter).

Redundancies

Spatial

Redundant information in the horizontal and vertical picture dimensions. Data similar or repeats itself in picture areas which are close to one another.

Temporal

Redundant data over given time. Data similar or repeats itself from moment to moment, even if its location in the picture area changes.

DCT Function

8x8 pixel blocks converted from spatial domain to spatial frequency domain. Transformed blocks numerically represented as 8x8 DCT coefficients. DCT coefficients are more suitable for bit rate reduction techniques. **The transform process does not result in bit rate reduction**.

DCT Example

Have a gray scale of vertical stripes. Sample values in columns: 223, 191, 159, 129, 98, 72, 39, 16. DCT coefficients all zeros except for top left row: 43.8, -40, 0, -4.1, 0, -1.1.

Quantization

Divides each DCT coefficient by a frequency-dependent value and truncates the results to an integer. Many of resulting integers are zero or small values. Quantization coefficients can be tailored to complement limitations of the human visual system. Quantization causes information to be irretrievably lost. Reconstructed pixels usually differ in value from the original.

Processing Sequences

Zigzag or Classic (nominally for frames).

Alternate (nominally for fields).

Entropy Coding

Run length coding uses a special code for repeating values (like 13 zeros, 5 ones, 4 twos). Variable length coding uses shorter code words for more probable symbols (like Morse code).

Example table.

Symbol	*Probability*	*Code Word*
A	0.5	0
B	0.25	10
C	0.125	110
D	0.0625	1110
E	0.03125	11110
F	0.03125	11111

Intraframe Coding

Shows a block diagram for encoding.

Convert 4:2:2 to 8-bit 4:2:0 – information lost, data reduced

DCT – no loss, no data reduced

Quantize – data reduced, information lost

Entropy coding – data reduced (no loss)

Buffer – feeds back to quantization step, accepts quantizing data from quantizer as input

Quantizing reduces the number of bits for each coefficient, gives preferenceto certain coefficients, reduction can differ for each coefficient.

Temporal Reductions

Frame to frame redundancies. New location, same data. New data uncovered. Sailboat moves, uncovering buildings behind the sailboat.

Motion Estimation

Use macroblock of 16x16 pixels. Frame N + 1 compared to Frame N calculates a motion vector to specify where the macroblock in frameN now appears in frame N + 1.

Interframe Coding (forward prediction)

Shows block diagram.

Group of Pictures

Always starts with an I picture. Suggested standard of15 pictures within a GOP.