Tech Topic #2

January 25, 2010

Today's Objectives

- Video coding techniques
  - We'll leave audio…
  - Most of what happens in video happens in audio
- We are not (yet) covering
  - Error (network errors) recovery techniques

Video Basics

- Represented by 3 8-bit values
  - RGB: Red, Green, Blue
  - YUV: Luma (Y) and two chrominance values (UV) Cr and Cb
- HDTV as an example
  - 1920 x 1080 x 24 bpp x 30 fps = 1.5 Gbps (60 Hz interlaced)
- Other smaller formats
  - NTSC: 352x240 @ 30fps
  - PAL: Source Input Format (SIF): 352x288 @ 25 fps
  - Quarter Common Intermediate Format (QCIF): 176x144
  - Sub QCIF: 128x96
  - 4CIF: 704x576
  - 16CIF: 1408x1152

Video Basics

- Three mechanisms used as part of encoding/compression scheme
  - Spatial: similarities around a given location of a frame
  - Temporal: similarities around a given location across time
  - Lossy: eliminate details not visible to the naked eye

Spatial Redundancy

- Take advantage of the fact that most video has similar values in nearby positions
- One option is to use differential encoding
  - Assume that the next value is the same and encode only the difference
- Compress the resulting stream of bits
  - More later

Temporal Redundancy

- Check for similarities between video frames
- If no other frames exist (single image) or no temporal redundancy is used, the frame is intra-encoded (sometimes called an I-frame)

Temporal Redundancy

- Predicted images (P-frames) are based on other I-or P- frames
  - Encoder does an expanding ring search to find image components (motion compensation)
  - How far from original location to look corresponds to how much processing is necessary and how much compression is had (key reason encoder is more complex)

- Bi-directionally predicted images (B-frames) are based on combination of forward and backward prediction
  - If imagine component is in Location A now and Location C in the future. Half way between now and then, it should be in Location B (interpolation).
  - Encode the “error”: the difference between predicted Location B and where it actually is

Temporal Redundancy

- Thresholds are used for P and B frames
  - If there are enough differences (e.g., a scene change) such that an P or B frame would not result in any less data, encode frame as an I frame
- Typically use a pattern of I, P, and B frames
  - Ex: I B B P B B P B B … I B B P B B P B B … repeat
  - Could encode all I frames (essentially motion JPEG)
- For real-time video, typically no B frames
  - B frames depend on future frames, can't encode and send until the future I frame is generated (so adds delay)
- For compressed and stored video, different I/P/B patterns can be tried

Transform Coding

- Transform coding is used to convert spatial image pixel values to transform coefficient values
  - No information is lost, the number of coefficients produced is equal to the number of pixels transformed
- The result is that most of the energy in the image will be contained in a few large transform coefficients
  - Generally, only a few coefficients will contain most of the energy in a block
  - Smaller coefficients can be coarsely quantized or deleted without doing visible damage to the reproduced image

Transform Coding

- Many types of transforms have been tried for picture coding
  - Fourier, Karhonen-Loeve, Walsh-Hadamard, lapped orthogonal, discrete cosine, and wavelets

- Goal is to have the most concentration of energy and least number of artifacts

Quantization

- The level of quantization provides clearest tradeoff between quality and level of compression
  - More quantization means more compression which means less bandwidth but more artifacts
- Quantization can be adjusted dynamically
  - Constant Bit Rate (CBR): same amount of bandwidth no matter the amount of energy/action in a picture
  - Variable Bit Rate (VBR): bandwidth requirements vary based on complexity and motion in video
- Use of quantization is the source of noise/error in a compressed stream (different than network data loss)

Error/Noise

- Coding error: the difference between the source picture and the reproduced picture
- Coding error is measured as the root-mean-square between the two values
  - A common metric for evaluating the performance of an encoding system

Huffman/Run-Length Coding

- An example block of 8x8 DCT samples:

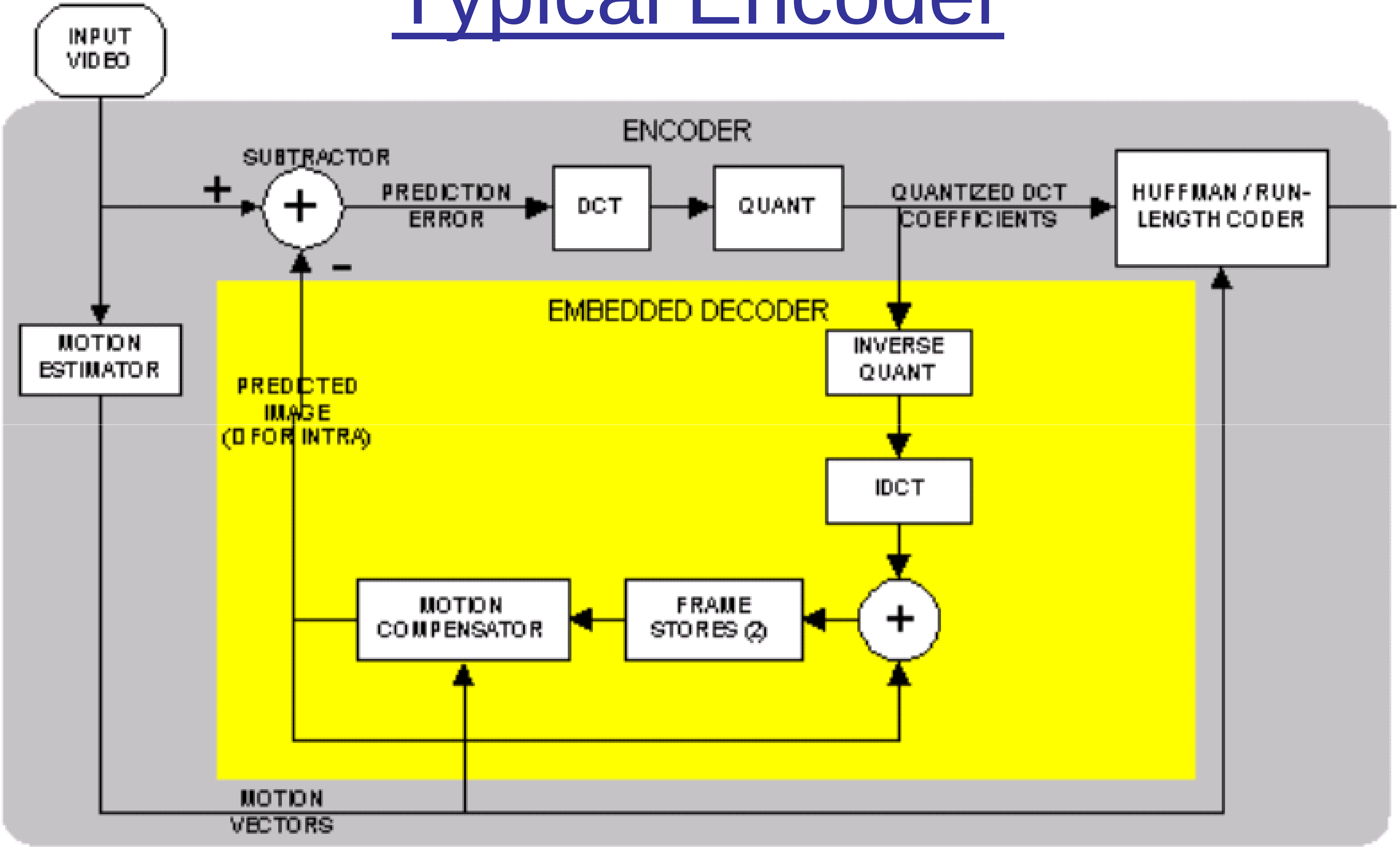
12	34	0	54	0	0	0	0
87	0	0	12	0	0	0	0
16	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

- Zig-zag scanning is used to create single sequence
  - 12 34 87 16 0 0 54 0 0 0 0 0 0 12 0 0 3 0 0 0 .....

Huffman/Run-Length Coding

- Zig-zag scanning is used to create single sequence:
  - 12 34 87 16 0 0 54 0 0 0 0 0 0 12 0 0 3 0 0 0 .....
- Could apply quantization:
  - 12 36 88 16 0 0 56 0 0 0 0 0 0 12 0 0 0 0 0 0 .....
- Or eliminate elements
  - http://www.john-wiseman.com/technical/MPEG_tutorial.htm
  - See Figures 7-14
- Then use combination of Huffman and Run-Length
  - Huffman: most common sequence conversion
  - Run-Length: runs of single value

Typical Encoder



Standards

- MPEG-1 (1993)
  - Designed up to 1.5 Mbps
  - Standard CD-ROM, NTSC video quality
- MPEG-2 (1995)
  - Designed for between 1.5 and 15 Mbps
  - Standard for DVD, HDTV
- MPEG-4 (1999)
  - Object-based compression
- MPEG-7 (2002)
  - Provides framework for adding descriptive information about video contents: uses XML to store meta-data
- MPEG-21(2001)
  - Adds digital rights/permissions/restrictions

Other Standards

- MPEG came out of ISO

- Also CCITT (which became ITU)
  - Early on, principally designed encoding for low-bit rate video conferencing (Ex: H-261, H-263, H-264, etc.)
  - Typically use the same components (e.g., temporal, spatial, etc.) in the encoding scheme

- Typically there are separate standards for audio and video and then for a combination of the two

MPEG-1

- Consists of 5 parts:
  - Systems (storage and synchronization of video, audio, and other data together)
  - Video (compressed video content)
  - Audio (compressed audio content)
    - 3 different layers, the third is most commonly used (MP3)
  - Conformance testing (testing the correctness of implementations of the standard)
  - Reference software (example software showing how to encode and decode according to the standard)

MPEG-2 Video Compression

- Pictures is made up of pixels (RGB values)
- Block is 8x8 array of pixels
  - Uses YUV and sub-sampling
  - Eye is most sensitive to changes in luminance, and less sensitive to variations in chrominance

- a:b:c (Luma:Cr:Cb)
- 4:2:0 (X=Luma, O=Cr and Cb) [Reduction from 12 to 6]

```
X X X X X X X X
O   O   O   O
X X X X X X X X

X X X X X X X X
O   O   O   O
X X X X X X X X
```

MPEG-2 Video Compression

- Pictures is made up of pixels (RGB values)
- Block is 8x8 array of pixels
  - Uses YUV and sub-sampling

- a:b:c (Luma:Cr:Cb)
- 4:2:2 (X=Luma,   O=Cr and Cb,   Z=Y, Cr and Cb)

Z X Z X Z X Z X

Z X Z X Z X Z X

Z X Z X Z X Z X

Z X Z X Z X Z X

MPEG-2 Video Compression

- Pictures are made up of pixels (RGB values)
- Block is 8x8 array of pixels
  - Uses YUV and sub-sampling
  - Basis for DCT coding
- Macroblock is 16x16 array of blocks
- Slice is typically a row of macroblocks

MPEG-2 Video Compression

- Pictures are either “frame picture” or “field picture”
  - “frame picture”: complete frame
  - “field picture”: half of interlaced frame
- Has a header
  - Picture type (I, P, B)
  - Temporal reference information
  - Motion vector search range
  - Optional user data
- Group of Pictures (GOP)
  - Series of I, P, and B frames

MPEG-2 Video Compression

- Sequence is a group of GOP
  - Has header including:
    - picture size
    - aspect ratio
    - frame rate and bit rate
    - optional quantizer matrices
    - required decoder buffer size
    - chroma pixel structure

- Need, for example, when changing channels

MPEG Transmission

- Not designed for transmission over networks where losses can occur
  - What happens when part of an I frame or P frame is lost
  - What happens when header information is lost
- There are steps that can be taken to improve the resiliency to network losses, but there are better formats
  - Basic idea is to have fewer critical elements (header information) sent only once
  - Increase redundancy means less susceptible to loss, but at the tradeoff of sending more information

H.264 Scalable Video Coding

- Definitely a dense paper
- Required some very close reading and definitely some background in the area
- Basic idea is how to do scalable video coding

Scalable Video Coding

- Create set of dependent layers such that each additional layer adds detail and clarity
- Very useful for creating and sending a video stream to lots of people simultaneously but who have different end-to-end throughput levels
  - Useful when combined with multicast
  - Each layer is sent as different group
  - Join group transmitting base layer and then join additional layers as capacity allows

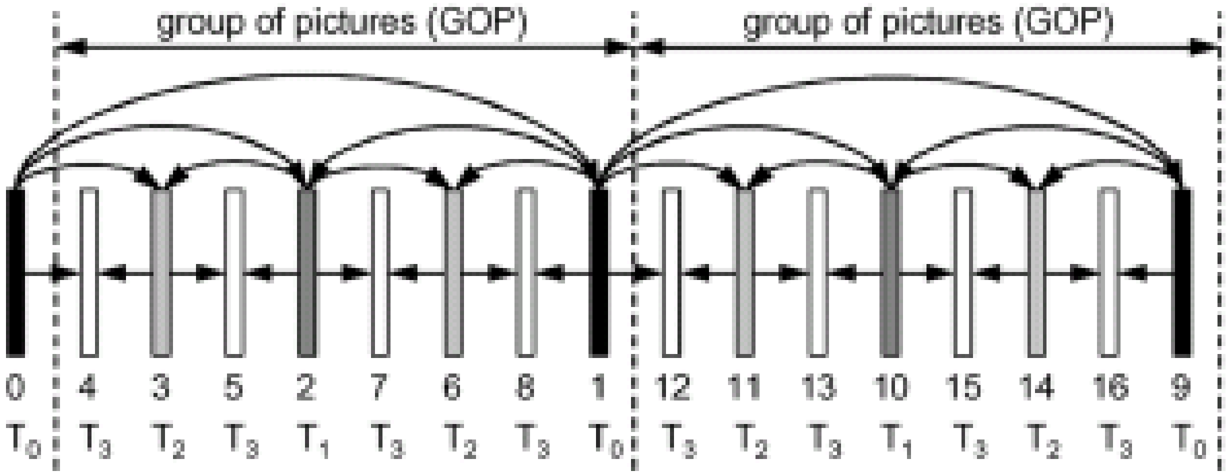
- Alternative is to send separately encoded streams
  - Same stream encoded at different levels
  - Compare between complexity, bandwidth required, performance, and overhead

H.264 Scalable Video Coding

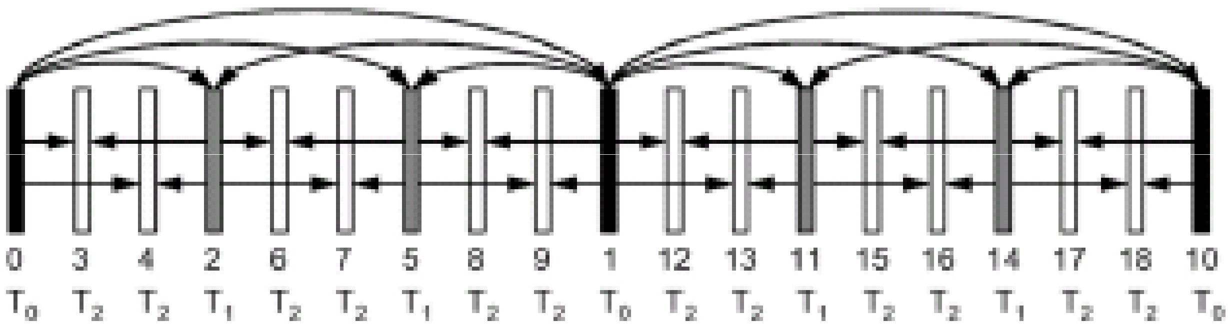
- Definitely a dense paper
- Required some very close reading and definitely some background in the area
- Basic idea is how to do scalable video coding
  - Additional B frames can be added
  - Additional B frames are dependent on the higher level B frames directly on either side of the current level

H.264 Scalable Video Coding

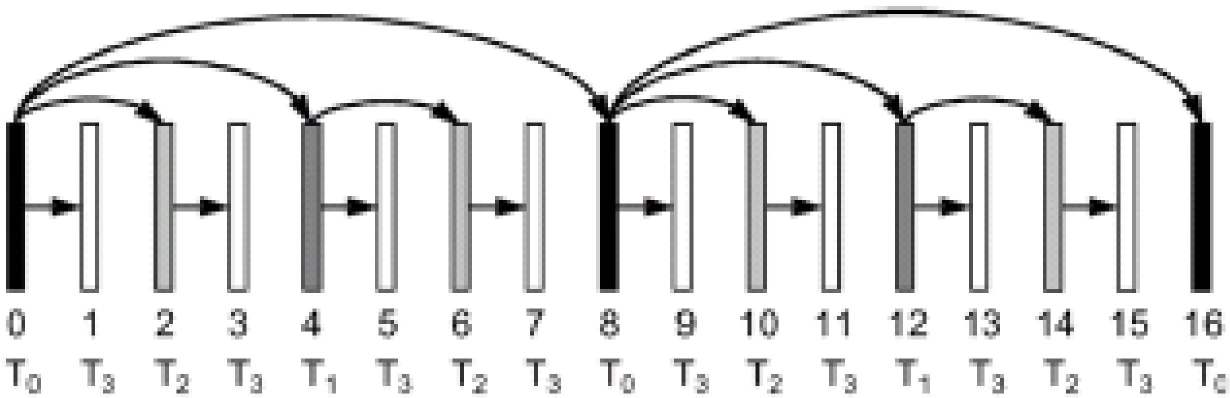
- Definitely a dense paper
- Required some very close reading and definitely some background in the area
- Basic idea is how to do scalable video coding
  - Additional B frames can be added
  - Additional B frames are dependent on the higher level B frames directly on either side of the current level


group of pictures (GOP)
group of pictures (GOP)
0 4 3 5 2 7 6 8 1 12 11 13 10 15 14 16 9
T_0 T_3 T_2 T_3 T_1 T_3 T_2 T_3 T_0 T_3 T_2 T_3 T_1 T_3 T_2 T_3 T_0

(a)


0 3 4 2 6 7 5 8 9 1 12 13 11 15 16 14 17 18 10
T_0 T_2 T_2 T_1 T_2 T_2 T_1 T_2 T_2 T_0 T_2 T_2 T_1 T_2 T_2 T_1 T_2 T_2 T_0

(b)


0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
T_0 T_3 T_2 T_3 T_1 T_3 T_2 T_3 T_0 T_3 T_2 T_3 T_1 T_3 T_2 T_3 T_0

(c)

H.264 Scalable Video Coding

- Lots of experiments
  - Effects for different GOP configurations
  - Bandwidth savings and overhead
  - PSNR values for different combinations of layers
  - Additional tradeoffs that can be made
  - Different types of video