In The Name of Allah

Digital Media Laboratory
Advanced Information & Communication Technology Center
Sharif University of Technology

Video
(Fundamentals, Compression Techniques & Standards)

Hamid R. Rabiee
Mostafa Salehi, Fatemeh Dabiran, Hoda Ayatollahi

Spring 2011
Outlines

- Frame Types
- Color
- Video Compression Techniques
  - H.261
  - H.263
  - MPEG Family
- Video Coding Standards
Digital Video

≤ Video; a multi-dimensional signal

≤ Video is a sequence of 2D images called frames. Digital video is digitized version of a 3D function \( f(x,y,t) \).
Frame Types

- **I-frames** are coded without reference to other frames. Serve as reference pictures for predictive-coded frames.

- **P-frames** are coded using motion compensated prediction from a past I-frame or P-frame.

- **B-frames** are bidirectionally predictive-coded. Highest degree of compression, but require both past and future reference pictures for motion compensation.

- **D-frames** are DC-coded. Of the DCT coefficients only the DC coefficients are present. Used in interactive applications like VoD for rewind and fast-forward operations.
I-frames Compression

I-frame

For each macroblock

For each $8 \times 8$ block

DCT
Quantization
Entropy coding

1010010...
Motion Compensation

- Exploits temporal redundancy in video frames
- Prediction:
  - Assumes that "locally" current frame can be modeled as a translation of a previous frame
  - Displacement need not be the same everywhere in the frame; encode motion information properly for accurate reconstruction
- Bi-directional Interpolation:
  - High degree of compression
  - Areas just uncovered are not predictable from the past, but can be predicted from the future
  - Effect of noise and errors can be reduced by averaging between past and future references
  - Frequency of B-frames: increasing the frequency of B-frames
    - Improves compression efficiency, but ...
    - Decreases the correlation between the B-frame and the references, as well as between the references
      ‡ reasonable to space references by 1/10th of a second
Motion Compensation

Previous Frame

Current Frame

Previous Frame

Current Frame

only transmit changing blocks
Motion Compensation

Reference frame

Target frame

Matched macroblock
Search window

Macroblock
P-frame Compression

Target frame

Reference frame

Current macroblock

Best match

Difference macroblock

For each $8 \times 8$ block

DCT

Quantization

Entropy coding

Motion vector

$01100010 \ldots$
Why do we need B-frames?

- Bi-directional prediction works better than only using previous frames when occlusion occurs.

For this example, the prediction from next frame is used and the prediction from previous frame is not considered.
Frame 1

Previous frame doesn’t contain this information

Frame 2

Frame 3

Next frame does. Can we code from this?

I-frame
(Intra-frame, for error resilience and movie joining)

B-frame
(Bi-directional predicted frame)

from P-frame

from I-frame

P-frame
(Predicted frame from previous I or P frames)
B-frames Compression

Previous reference frame  Target frame  Future reference frame

Difference macroblock
\[ Y \]

For each 8 × 8 block

- DCT
- Quantization
- Entropy coding

Motion vectors

0011101...
B-frame Advantage

- B-frames increase compression.
- Typically use twice as many B frames as I+P frames.

<table>
<thead>
<tr>
<th>Type</th>
<th>Size</th>
<th>Compression</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>18KB</td>
<td>7:1</td>
</tr>
<tr>
<td>P</td>
<td>6KB</td>
<td>20:1</td>
</tr>
<tr>
<td>B</td>
<td>2.5KB</td>
<td>50:1</td>
</tr>
<tr>
<td>Average</td>
<td>4.8KB</td>
<td>27:1</td>
</tr>
</tbody>
</table>

Typical MPEG-1 values.
Really depends on video content.
B-frame Disadvantages

- Computational complexity.
  - More motion search, need to decide whether or not to average.

- Increase in memory bandwidth.
  - Extra picture buffer needed.
  - Need to store frames and encode or playback out of order.

- Delay
  - Adds several frames delay at encoder waiting for need later frame.
  - Adds several frames delay at decoder holding decoded I/P frame, while decoding and playing prior B-frames that depend on it.
Color

- Human Eye has receptors for brightness (in low light), and separate receptors for red, green, and blue.
- Can make any color we can see by mixing red, green and blue light in different intensities
Color TV

≤ Original TV standards were black and white.
    ≤ AM: Amplitude of signal determines brightness.

≤ How to add color without changing TV transmitters, and in such a way that it’s compatible with existing B&W TVs?

≤ Add a high frequency subcarrier in band within B&W TV signal.
    ≤ Not noticeable on B&W TV - would show as high frequency pattern, but human eye can’t really see this well.

≤ Modulate the phase of the sub carrier to indicate the color.
    ≤ Problem: how to calibrate the absolute phase.
    ≤ Get this wrong, and the colors display incorrectly.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Phase (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burst</td>
<td>0</td>
</tr>
<tr>
<td>Yellow</td>
<td>15</td>
</tr>
<tr>
<td>Red</td>
<td>75</td>
</tr>
<tr>
<td>Magenta</td>
<td>135</td>
</tr>
<tr>
<td>Blue</td>
<td>195</td>
</tr>
<tr>
<td>Cyan</td>
<td>255</td>
</tr>
<tr>
<td>Green</td>
<td>315</td>
</tr>
</tbody>
</table>
NTSC
National Television Standard Committee

- Introduced in 1953 (in US)
- Used in US, Canada, Japan
- 30 frames per second (Actually 29.97)
  - Interlaced (even/odd field lines), so 60 fields per second.
  - Same as 60Hz AC power in these countries
- 525 lines
  - Picture only on 480 of these => 640x480 monitors
  - Rest are the vertical rescan.
- Aspect ratio is 4:3
- Needs color calibration
- Uses a color burst signal at start of each line, but needs TV to be adjusted relative to this. “NTSC = Never Twice Same Color”
PAL
Phase Alternating Line

- Introduced in 1967 (by Walter Bruch in Germany)
- PAL-I(UK), PAL-B/G (much of Europe), PAL-M (Brazil) ...
  - Differ mainly in audio subcarrier frequency.
- 25 frames per second
  - Interlaced (even/odd field lines), so 50 fields per second.
  - Same as 50Hz AC power in these countries
- 625 lines
  - Only 576 lines used for picture.
  - Rest are vertical retrace, but often carry teletext information.
- Color phase is reversed on every alternate line.
  - Originally human eye would average to derive correct color.
  - Now TV sets auto calibrate to derive correct color.
SÉCAM
Séquentiel Couleur Avec Mémoire

- Introduced in 1967 (in France)
  - "System Essentially Contrary to American Method"
- Used in France, Russia, Eastern Europe...
- 625 lines, 25 fps interlaced, like PAL
- Uses FM modulation of subcarrier.
  - Red-Luminance difference on one line
  - Blue-Luminance difference on next line
  - Uses a video line store to recombine the two signals
  - Vertical color resolution is halved relative to NTSC and PAL.
- Human eye is not sensitive to lack of spatial color information.
Colorsace Representations

≤ RGB (Red, Green, Blue)
   ≤ Basic analog components (from camera/to TV tube)

≤ YPbPr (Y, B-Y, R-Y)
   ≤ Color space derived from RGB used in component video. Y= Luminance, B = Blue, R = Red

≤ YUV
   ≤ Similar to YPbPr but scaled to be carried on a composite carrier.
   ≤ YCbCr Digital representation of YPbPr colorspace (8 bit, two's complement)
Color System in Video

\[ YUV \text{ was used in PAL (an analog video standard) and also for digital video.} \]

\[ Y \text{ is the luminance component (brightness)} \]

\[ Y = 0.299 \, R + 0.587 \, G + 0.144 \, B \]

\[ U \text{ and } V \text{ are color components} \]

\[ U = B - Y \]

\[ V = R - Y \]
RGB vs. YUV
YUV Formats

≤ YUV 4:4:4
   ≤ 8 bits per Y,U,V channel (no chroma down sampling)

≤ YUV 4:2:2
   ≤ 4 Y pixels sample for every 2 U and 2V
   ≤ 2:1 horizontal down sampling, no vertical down sampling

≤ YUV 4:2:0
   ≤ 2:1 horizontal down sampling
   ≤ 2:1 vertical down sampling

≤ YUV 4:1:1
   ≤ 4 Y pixels sample for every 1 U and 1V
   ≤ 4:1 horizontal down sampling, no vertical down sampling
YIQ is the color standard in NTSC.

\[
\begin{bmatrix}
Y' \\
I \\
Q
\end{bmatrix} =
\begin{bmatrix}
0.299 & 0.587 & 0.144 \\
0.595879 & -0.274133 & -0.321746 \\
0.211205 & -0.523083 & 0.311878
\end{bmatrix}
\begin{bmatrix}
R' \\
G' \\
B'
\end{bmatrix}
\]

\[
\begin{bmatrix}
Y' \\
C_b \\
C_r
\end{bmatrix} =
\begin{bmatrix}
65.481 & 128.553 & 24.966 \\
-37.797 & -74.203 & 112 \\
112 & -93.786 & -18.214
\end{bmatrix}
\begin{bmatrix}
R' \\
G' \\
B'
\end{bmatrix}
+ \begin{bmatrix}
16 \\
128 \\
128
\end{bmatrix}
\]
Digital Video Formats

- Common Intermediate Format (CIF):
  - This format was defined by CCITT (TSS) for H.261 coding standard (teleconferencing and videophone).
  - Several size formats:
    - SQCIF: 88x72 pixels.
    - QCIF: 176x144 pixels.
    - CIF: 352x288 pixels.
    - 4CIF: 704x576 pixels.
  - Non-interlaced (progressive), and chrominace sub-sampling using 4:2:0.
  - Frame rates up to 25 frames/sec
Digital Video Formats

- Source Input Format (SIF):
  - Utilized in MPEG as a compromise with Rec. 601.
  - Two size formats (similar to CIF):
    - QSIF: 180x120 or 176x144 pixels at 30 or 25 fps
    - SIF: 360x240 or 352x288 pixels at 30 or 25 fps
  - Non-interlaced (progressive), and chrominance sub-sampling using 4:2:0.

- High Definition Television (HDTV):
  - 1080x720 pixels.
  - 1920x1080 pixels.
Uncompressed Video Data Rate

≤ Examples (CCIR 601)

≤ PAL signal: 864x625, YUV 4:2:2 20 bits/pixel, 25fps. 270Mb/s
≤ PAL signal: 864x625, YUV 4:2:2 16 bits/pixel, 25fps. 216Mb/s
≤ PAL video: 720x576, YUV 4:2:2 16 bits/pixel, 25fps. 166Mb/s (~1GByte/min)

≤ Firewire: 400Mb/s (800Mb/s)

≤ USB 2.0: 480Mb/s
VIDEO COMPRESSION REVIEW
Need for Compression

- Large data rate and storage capacity requirement

<table>
<thead>
<tr>
<th>Satellite imagery</th>
<th>180x180 km² 30 m² resolution</th>
<th>600 MB/image</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTSC video</td>
<td>30 frames/s, 640x480 pixels, 3 bytes/pixel</td>
<td>30 Mbytes/s</td>
</tr>
</tbody>
</table>

- Compression algorithms exploit:
  - Spatial redundancy (i.e., correlation between neighboring pixels)
  - Spectral redundancy (i.e., correlation between different frequency spectrum)
  - Temporal redundancy (i.e., correlation between successive frames)
Requirement for Compression Algorithms

**Objectives**

- Minimize the complexity of the encoding and decoding process
- Ensure a good quality of decoded images
- Achieve high compression ratios

**Other general requirements**

- Independence of specific size and frame rate
- Support various data rates
Classification of Compression Algorithms

≤ Lossless compression
  ≤ Reconstructed image is mathematically equivalent to the original image (i.e., reconstruction is perfect)
  ≤ Drawback: achieves only a modest level of compression (about a factor of 5)

≤ Lossy compression
  ≤ Reconstructed image demonstrates degradation in the quality of the image
    † the techniques are irreversible
  ≤ Advantage: achieves very high degree of compression (compression ratios up to 200)
  ≤ Objective: maximize the degree of compression while maintaining the quality of the image to be virtually lossless
Compression Techniques: Fundamentals

≤ Entropy encoding

≤ Ignores semantics of input data and compresses media streams by regarding them as sequences of bits

≤ Examples: run-length encoding, Huffman encoding, ...

≤ Source encoding

≤ Optimizes the compression ratio by considering media specific characteristics

≤ Examples:

≤ Predictive coding: e.g., DPCM

≤ Layered coding: e.g., bit-plane coding, sub-sampling

≤ Transform coding: e.g., DCT, FFT, Wavelet, ...

≤ Most compression algorithms employ a hybrid of the above techniques
Entropy Coding

≤ Run-length encoding

\[ \alpha \alpha \alpha \alpha \rightarrow 4\alpha \]

≤ Huffman encoding

≤ Employ variable length codes

≤ Assign fewer bits to encode more frequently occurring values

≤ exploit the statistical distribution of the values within an data sequence

≤ Share codebook between encoder and decoder
Source Coding: Predictive Coding

≤ Basic technique

≤ Predict the value at a pixel by using the values of the neighboring pixels; and
≤ Encode the difference between the actual value and the predicted value

\[ e_m = x_m - \hat{x}_m \]

≤ Predictor

≤ Dimension of the predictor
≤ Order of the predictor: number of pixels used

≤ Example of a third order predictor:

\[ \hat{x}_m = 0.75a - 0.5b + 0.75c \]

≤ Huffman encoding of differential images
Source Coding: Bit-plane Encoding

- An $N \times N$ image with $k$ bits per pixel can be viewed as $k N \times N$ bit planes
- Encode each bit plane separately
- Advantages:
  - Permits progressive transmission of encoded images (most significant bit plane first - since it generally contains more information)
- Encoding should be carried out such that separate encoding yields better performance than jointly encoding the bit planes
  - Gray codes are better suited as compared to binary encoding
Source Coding: Transform Coding

- Subdivide an individual $N \times N$ image into several $n \times n$ blocks
- Each $n \times n$ block undergoes a reversible transformation
- Basic approach:
  - De-correlate the original block; radiant energy is redistributed amongst only a small number of transform coefficients
  - Discard many of the low energy coefficients (through quantization)
- General requirements:
  - Image independence
  - Should be computationally efficient
VIDEO CODING STANDARDS
Video Coding Standards

≤ H.261
≤ H.263
≤ H.263+
≤ MPEG Family
≤ MPEG-1
≤ MPEG-2
≤ MPEG-4
≤ MPEG-7
≤ H.264

ITU-T Standards
H.261
H.263
H.263+
H.263++

Joint ITU-T / MPEG Standards
H.262/MPEG-2
H.26L

MPEG Standards
MPEG-1
MPEG-4

Timeline:
H.261

- H.261 is an ITU video compression standard finalized in 1990.
- The basic scheme of H.261 has been retained in the newer video standards.
- H.261 supports bit rates at p*64 kbps (p=1..30).

<table>
<thead>
<tr>
<th>Video format</th>
<th>Luminance resolution</th>
<th>Chrominance resolution</th>
<th>Bitrate (Mbps) (if 30 fps and uncompressed)</th>
<th>H.261 support</th>
</tr>
</thead>
<tbody>
<tr>
<td>QCIF</td>
<td>176 × 144</td>
<td>88 × 72</td>
<td>9.1</td>
<td>Required</td>
</tr>
<tr>
<td>CIF</td>
<td>352 × 288</td>
<td>176 × 144</td>
<td>36.5</td>
<td>Optional</td>
</tr>
</tbody>
</table>
H.261

\[ \leq \] In H.261, motion vectors are in the range \([-15,15] \times [-15,15]\).

\[ \leq \] H.261 uses a constant step-size for different DCT coefficients.

\[ \leq \] For DC coefficients

\[ QDCT = \text{round}\left(\frac{DCT}{\text{step\_size}}\right) = \text{round}\left(\frac{DCT}{8}\right) \]

\[ \leq \] For AC coefficients

\[ QDCT = \left\lfloor \frac{DCT}{\text{step\_size}} \right\rfloor = \left\lfloor \frac{DCT}{2 \times \text{scale}} \right\rfloor \]

Where scale = 1 .. 31
Group of macroBlocks (GOB)

To reduce the error propagation problem, H.261 makes sure that a “group” of Macro-Blocks can be decoded independently.

<table>
<thead>
<tr>
<th>GOB 0</th>
<th>GOB 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOB 2</td>
<td>GOB 3</td>
</tr>
<tr>
<td>GOB 4</td>
<td>GOB 5</td>
</tr>
<tr>
<td>GOB 6</td>
<td>GOB 7</td>
</tr>
<tr>
<td>GOB 8</td>
<td>GOB 9</td>
</tr>
<tr>
<td>GOB 10</td>
<td>GOB 11</td>
</tr>
</tbody>
</table>

CIF

<table>
<thead>
<tr>
<th>GOB 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOB 1</td>
</tr>
<tr>
<td>GOB 2</td>
</tr>
</tbody>
</table>

QCIF
**H.261 Bit Stream Syntax**

![Diagram of H.261 Bit Stream Syntax](image)

<table>
<thead>
<tr>
<th>Layer</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture layer</td>
<td>PSC, TR, PType, GOB, GOB, ... GOB</td>
</tr>
<tr>
<td>GOB layer</td>
<td>GBSC, GN, GQuant, MB, ... MB</td>
</tr>
<tr>
<td>Macroblock layer</td>
<td>Address, Type, MQuant, MVD, CBP, b0, b1, ... b5</td>
</tr>
<tr>
<td>Block layer</td>
<td>DC, (Run, Level), ... (Run, Level), EOB</td>
</tr>
</tbody>
</table>

**Definitions**

- **PSC**: Picture Start Code
- **PType**: Picture Type
- **GBSC**: GOB Start Code
- **GQuant**: GOB Quantizer
- **MQuant**: MB Quantizer
- **CBP**: Coded Block Pattern
- **TR**: Temporal Reference
- **GOB**: Group of Blocks
- **GN**: Group Number
- **MB**: Macroblock
- **MVD**: Motion Vector Data
- **EOB**: End of Block

Digital Media Lab - Sharif University of Technology
H.263

- H.263 is an improved video coding standard for video conferencing through PSTN (public switching telecommunication network).
- Apart from QCIF and CIF, it supports SubQCIF, 4CIF and 16CIF.
- H.263 has a different GOB scheme.
The motion compensation in this standard is a bit different from the MPEG method!

The motion compensation in the core H.263 is based on one motion vector per macroblock of 16 × 16 pixels, with half pixel precision.

Motion vector prediction

\[ \text{pred}_x = \text{median} (\text{MV}_{1x}, \text{MV}_{2x}, \text{MV}_{3x}) \]
\[ \text{pred}_y = \text{median} (\text{MV}_{1y}, \text{MV}_{2y}, \text{MV}_{3y}) \]

\[ \text{MVD}_x = \text{MV}_x - \text{pred}_x \]
\[ \text{MVD}_y = \text{MV}_y - \text{pred}_y \]

Motion vector prediction for the border macroblocks

(a)

\[ (0, 0) \rightarrow \text{MV} \]

\[ \text{MV} \rightarrow \text{MV}_1 \text{ MV}_1 \]

\[ \text{MV} \rightarrow \text{MV}_1 \text{ MV}_1 \rightarrow \text{MV} \rightarrow \text{MV} \rightarrow \text{Border} \]
H.263+

≤ H.263 Ver. 2 (H.263+), ITU-T

≤ Additional negotiable options for H.263.
≤ New features include: deblocking filter, scalability, slicing for network packetization and local decode, square pixel support, arbitrary frame size, chromakey transparency, etc…
≤ Arbitrary frame size, pixel aspect ratio (including square), and picture clock frequency
≤ Advanced INTRA frame coding
≤ Loop de-blocking filter
≤ Slice structures
≤ Supplemental enhancement information
≤ Improved PB-frames
MPEG

- Moving Picture Experts Group

- JPEG does not exploit temporal (i.e., frame-to-frame) redundancy present in all video sequences

- MPEG exploits temporal redundancy

- MPEG requirements:
  - Random access
  - Fast searches - both forward and reverse
  - Reverse playback
  - Audio-video synchronization
  - Robustness to errors
  - Low encoding/decoding delay
  - Editability
MPEG Family

≤ MPEG-1
  ≤ Similar to H.263 CIF in quality

≤ MPEG-2
  ≤ Higher quality: DVD, Digital TV, HDTV

≤ MPEG-4/H.264
  ≤ More modern codec.
  ≤ Aimed at lower bitrates.
  ≤ Works well for HDTV too.
MPEG-1 Video

- MPEG-1 was approved by ISO and IEC in 1991 for “Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbps”.

- MPEG-1 standard is composed of
  - System
  - Video
  - Audio
  - Conformance
  - And Software
MPEG-1 Standard - An Overview

≤ Two categories: intra-frame and inter-frame encoding
≤ Contrasting requirements: delicate balance between intra- and inter-frame encoding
≤ Need for high compression ‡ only intra-frame encoding is not sufficient
≤ Need for random access ‡ best satisfied by intra-frame encoding
≤ Overview of the MPEG algorithm:
  ≤ DCT-based compression for the reduction of spatial redundancy (similar to JPEG)
  ≤ Block-based motion compensation for exploiting the temporal redundancy
    ≤ Motion compensation using both causal (predictive coding) and non causal (interpolative coding) predictors
Exploiting Temporal Redundancy
Three types of frames in MPEG-1

- **I-frames:**
  - Intra-coded frames, provide access points for random access – yield moderate compression

- **P-frames:**
  - Predicted frames are encoded with reference to a previous I or P frame

- **B-frames:**
  - Bi-directional frames encoded using the previous and the next I/P frame
  - Achieves maximum compression
Example

≤ The figure illustrates the relationship between these three types of picture. Since B-pictures use I and P-pictures as predictions, they have to be coded later. This requires reordering the incoming picture order, which is carried out at the preprocessor.
Motion Representation

- 16 x 16 blocks used as motion-compensation units (referred to as macro blocks)
  - Macro block size selected based on the tradeoff between the gain due to motion compensation and the cost of coding motion information

- Types of macro blocks:
  - Intra, forward-predicted, backward-predicted, average

- Two types of information are maintained:
  - Motion vector:
    - The difference between the spatial locations of the macro blocks
    - One motion vector for forward/backward predicted blocks, and two vectors for average blocks
    - Adjacent motion vectors typically differ only slightly - encode them using differential encoding techniques (e.g., DPCM)
  - Difference between the macro block being encoded and its predictor block(s) - encode the difference using DCT-based transform coding techniques
Motion Estimation

- Block-matching techniques employed for motion estimation
- Motion vector obtained by minimizing the mismatch between the block being encoded and its predictor
- Exhaustive search for such a block yield good results – but the complexity can be prohibitive
- Tradeoff between the quality of the motion vector versus the complexity of the motion estimation process is left to the implementer
Difference of MPEG-1 with H.261

≤ Picture formats (SIF vs. CIF)
≤ GOB structure

Slices in MPEG-1
Difference of MPEG-1 with H.261 (cont)

MPEG-1 uses different quantization tables for I and P or B frames.

<table>
<thead>
<tr>
<th>Intra-coding quantization table</th>
<th>Inter-coding quantization table</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 16 19 22 26 27 29 34</td>
<td>16 16 16 16 16 16 16 16</td>
</tr>
<tr>
<td>16 16 22 24 27 29 34 37</td>
<td>16 16 16 16 16 16 16 16</td>
</tr>
<tr>
<td>19 22 26 27 29 34 34 38</td>
<td>16 16 16 16 16 16 16 16</td>
</tr>
<tr>
<td>22 22 26 27 29 34 37 40</td>
<td>16 16 16 16 16 16 16 16</td>
</tr>
<tr>
<td>22 26 27 29 32 35 40 48</td>
<td>16 16 16 16 16 16 16 16</td>
</tr>
<tr>
<td>26 27 29 32 35 40 48 58</td>
<td>16 16 16 16 16 16 16 16</td>
</tr>
<tr>
<td>26 27 29 34 38 46 56 69</td>
<td>16 16 16 16 16 16 16 16</td>
</tr>
<tr>
<td>27 29 35 38 46 56 69 83</td>
<td>16 16 16 16 16 16 16 16</td>
</tr>
</tbody>
</table>

(the prediction error is like noise and their DCT coefficients are quite “flat”. We can use a uniform quantization table.)

\[
\begin{align*}
\text{Intra mode:} & \quad Q_{DCT}[i, j] &= \text{round} \left( \frac{8 \times DCT[i, j]}{\text{step}_\text{size}[i, j]} \right) \\
\text{Inter mode:} & \quad Q_{DCT}[i, j] &= \text{round} \left( \frac{8 \times DCT[i, j]}{Q_1[i, j] \times \text{scale}} \right)
\end{align*}
\]
Difference of MPEG-1 with H.261 (cont)

≤ Sub pixel motion estimation in MPEG-1.

≤ Motion range up to 512 pixels.

≤ MPEG adds another layer called "Group Of Pictures" (GOP) to allow random video access.
MPEG-1 Video Stream

Video sequence

- Sequence header
- GOP
- GOP
- GOP
- ... Sequence end code

Group of picture layer

- GOP header
- Picture
- Picture
- Picture
- ... Picture

Picture layer

- Picture header
- Slice
- Slice
- Slice
- ... Slice

Slice layer

- Slice header
- Macroblock
- Macroblock
- ... Macroblock

Macroblock layer

- Macroblock header
- Block 0
- Block 1
- Block 2
- Block 3
- Block 4
- Block 5

Block layer

- Differential DC coefficient
- VLC run
- VLC run
- ... end_of_block

(if intra macroblock)
MPEG-2

MPEG-2 profiles and levels:

Profiles and Levels in MPEG-2

<table>
<thead>
<tr>
<th>Level</th>
<th>Simple profile</th>
<th>Main profile</th>
<th>SNR scalable profile</th>
<th>Spatially scalable profile</th>
<th>High profile</th>
<th>4:2:2 profile</th>
<th>Multiview profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>High 1440</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Main</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Low</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

**TABLE 11.6:** Four levels in the main profile of MPEG-2.

<table>
<thead>
<tr>
<th>Level</th>
<th>Maximum resolution</th>
<th>Maximum fps</th>
<th>Maximum pixels/sec</th>
<th>Maximum coded data rate (Mbps)</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1,920 × 1,152</td>
<td>60</td>
<td>62.7 × 10^6</td>
<td>80</td>
<td>Film production</td>
</tr>
<tr>
<td>High 1440</td>
<td>1,440 × 1,152</td>
<td>60</td>
<td>47.0 × 10^6</td>
<td>60</td>
<td>Consumer HDTV</td>
</tr>
<tr>
<td>Main</td>
<td>720 × 576</td>
<td>30</td>
<td>10.4 × 10^6</td>
<td>15</td>
<td>Studio TV</td>
</tr>
<tr>
<td>Low</td>
<td>352 × 288</td>
<td>30</td>
<td>3.0 × 10^6</td>
<td>4</td>
<td>Consumer tape equivalent</td>
</tr>
</tbody>
</table>
Scalable Layered Coding

≤ Need for Hierarchical Coding/Scalable Compression

≤ Facilitate access to images at different quality levels or resolutions

≤ Progressive transmission:
  ≤ Transmit image information in stages; at each stage, the reconstructed image is progressively improved
  ≤ Motivated by the need for transmitting images over low bandwidth channels
  ≤ Permits progressive transmission to be stopped either if an intermediate version is of satisfactory quality or the image is found to be of no interest
  ≤ Examples: multimedia databases, tele-browsing, etc.

≤ Multi-use environments:
  ≤ Support a number of display devices with differing resolutions
  ≤ Optimizes utilization of storage server and network resources
  ≤ Example: video-in-a-window archical Coding/Scalable Compression
Scalability

≤ SNR scalability
≤ Base layer uses rough quantization, while enhancement layers encode the residue errors.

≤ Spatial scalability
≤ Base layer encodes a small resolution video; enhancement layers encode the difference of bigger resolution video with the “un-sampled” lower resolution one.

≤ Temporal scalability
≤ Base layer down-samples the video in time; enhancement layers include the rest of the frames.

≤ Hybrid scalability
Scalability Example

Spatial Scalability

SNR Scalability
MPEG-2 vs. MPEG-1

- Sequence layer:
  - progressive vs. interlaced
  - More aspect ratios (e.g. 16x9)
  - Syntax can now signal frames sizes up to 16383x16383
  - Pictures must be a multiple of 16 pixels

- MPEG-2 can use a modified zig-zag for run-length encoding of the coefficients:
MPEG-2 vs. MPEG-1

- Picture Layer:
  - All MPEG-2 motion vectors are always half-pixel accuracy
    - MPEG-1 can opt out, and do one-pixel accuracy.
  - DC coefficient can be coded as 8, 9, 10, or 11 bits.
    - MPEG-1 always uses 8 bits.
  - Optional non-linear macroblock quantization, giving a more dynamic step size range:
    - 0.5 to 56 vs. 1 to 32 in MPEG-1.
    - Good for high-rate high-quality video.
Interlacing

- Although MPEG-2 only codes full frames (both fields), it supports both field prediction and frame prediction for interlaced sources.
  - The current uncompressed frame has two fields.
  - Can do the motion search independently for each field.
  - Half the lines use one motion vector and half use the other to produce the reference block.
MPEG-4

- ISO/IEC designation 'ISO/IEC 14496': 1999
- Aimed at low bitrate (10Kb/s)
- Can scale very high (1Gb/s)
- Based around the concept of the composition of basic video objects into a scene.
MPEG-4

- Initial goal of MPEG-4
  - Very low bit rate coding of audio visual data.
- MPEG-4 (at the end)
  - Officially up to 10 Mbits/sec.
  - Improved encoding efficiency.
  - Content-based interactivity.
  - Content-based and temporal random access.
  - Integration of both natural and synthetic objects.
  - Temporal, spatial, quality and object-based scalability.
  - Improved error resilience.
Audio-Video Object

MPEG-4 Standard

- Defines the scheme of encoding audio and video objects
  - Encoding of shaped video objects.
  - Sprite encoding.
  - Encoding of synthesized 2D and 3D objects.
- Defines the scheme of decoding media objects.
- Defines the composition and synchronization scheme.
- Defines how media objects interact with users.
MPEG-7 Standard

- (2001) MPEG7, ISO “Content Representation for Info Search”
  - Specify a standardized description of various types of multimedia information. This description shall be associated with the content itself, to allow fast and efficient searching for material that is of a user’s interest.

- Mpeg-7
  - Independent of the coding format of the media & physical location of the media
  - Facilitates searching for media content with ease

- Mpeg-7 supports
  - text-based queries
  - complex content-based queries.
Structure of the standard

- MPEG-7 standardizes a representation of meta-data
- Media content has metadata
- Metadata provides context for data
- Normative representations and semantics of metadata is common in MPEG-7
MPEG-7 Applications

- Storage and retrieval of audiovisual databases (*image, film, radio archives*)
- Broadcast media selection (*radio, TV programs*)
- Surveillance (*traffic control, surface transportation, production chains*)
- E-commerce and Tele-shopping (*searching for clothes / patterns*)
- Remote sensing (*cartography, ecology, natural resources management*)
- Entertainment (*searching for a game, for a karaoke*)
- Cultural services (*museums, art galleries*)
- Journalism (*searching for events, persons*)
- Personalized news service on Internet (*push media filtering*)
- Intelligent multimedia presentations
- Educational applications
- Bio-medical applications
Why do we need MPEG-7?

Support for Advanced Query

Need

- Fast & Accurate Access
- Personalized Content Production and Consumption
- Content Management Automation

Visual Audio Sketch
H.264 (MPEG-4, Part 10)

≤ MPEG-4, Part 10 is also known as H.264.

≤ Advanced video coding standard, finalized in 2003.
H.264 vs. MPEG-2

- Multi-picture motion compensation.
  - Can use up to 32 different frames to predict a single frame.
  - B-frames in MPEG-2 only code from two.

- Variable block-size motion compensation
  - From 4x4 to 16x16 pixels.
  - Allows precise segmentation of edges of moving regions.

- Quarter-pixel precision for motion compensation.

- Weighted prediction (can scale or offset predicted block)
  - Useful in fade-to-black or cross-fade between scenes.

- Spatial prediction from the edges of neighboring blocks for "intra" coding.

- Choice of several more advanced context-aware variable length coding schemes (instead of Huffman).
H.264 Performance

- Typically half the data rate of MPEG-2.
- HDTV:
  - MPEG-2: 1920x1080 typically 12-20 Mbps
  - H.264: 1920x1080 content at 7-8 Mbps
H.264 Usage

- Pretty new, but expanding use.
  - Included in MacOS 10 (Tiger) for iChat video conferencing.
  - Used by Video iPod.
  - Adopted by 3GPP for Mobile Video.
  - Mandatory in both the HD-DVD and Blu-ray specifications for High Definition DVD.
Video Standards Applications

≤ H.261, ITU-T
   ≤ Designed to work at multiples of 64 kb/s (px64).

≤ MPEG-1, ISO “Storage & Retrieval of Audio & Video”
   ≤ Main application is CD-ROM based video (~1.5 Mb/s).

≤ MPEG-2, ISO “Digital Television”
   ≤ Main application is video broadcast (DirecTV, DVD, HDTV).
   ≤ Typically operates at data rates of 2-3 Mb/s and above.

≤ H.263, ITU-T
   ≤ Evolution of all of the above.
   ≤ Targeted low bit rate video <64 kb/s. Works well at high rates, too.
Video Standards Applications

≤ H.263 Ver. 2 (H.263+), ITU-T
   ≤ Additional negotiable options for H.263.

≤ MPEG-4, ISO “Multimedia Applications”
   ≤ Support for multi-layered, non-rectangular video display

≤ MPEG7, ISO “Content Representation for Info Search”
Next Session

NGN
Any Question

Thank you!

Winter 2011