Ch. 4: Video Compression
Multimedia Systems

Prof. Ben Lee (modified by Prof. Nguyen)
Oregon State University
School of Electrical Engineering and
Computer Science

Outline

- Introduction
- MPEG Overview
- MPEG Encoding
  - Motion Compensation
  - Frame Encoding
- MPEG Video Bit Stream
- MPEG Video Standards
- H.264/AVC
Chapter 4: Video Compression

Introduction

- Used by
  - Digital set-top boxes
  - HDTV decoders
  - DVD players
  - Video conferencing
  - Internet video
- ISO Motion Picture Experts Group (MPEG): MPEG-x series:
  - MPEG-1, MPEG-2, MPEG-4, MPEG-7, and MPEG-21.
- ITU-T Video Coding Experts Group (VCEG): H.26x series:
  - Generic coding of moving pictures/video for conferencing, streaming, etc: H.261, H.262 (MPEG-2), H.263, H.264

Outline

- Introduction
- MPEG Overview
- MPEG Encoding
  - Motion Compensation
  - Frame Encoding
- MPEG Video Bit Stream
- MPEG Video Standards
- H.264/AVC
MPEG Overview

- A moving picture is simply a succession of still pictures or frames.
- Although significant compression can be achieved by simply processing a video signal as a sequence of images, using JPEG...
- We can do a lot better by exploiting the fact that considerable redundancy exists in all video sequences:
  - Many of the pixels change very little or not at all, or the change simply involves the movement of a pattern of pixels from one location to another.
- Additional compression in the order of a factor of 2-4.

Basic Idea

- Spatial compression (JPEG)
- Temporal compression (Motion compensation)
  - Improves compression ratio by 2 - 4 times
  - Motion Estimation and Compensation
Frame Types

- Frame Types:
  - Intrapictures (I)
  - Unidirectional predicted pictures (P)
  - Bidirectional predicted pictures (B)
- Allows
  - Random access
  - Fast forward/reverse searches

Frame Reconstruction

- I-frame is a complete image (JPEG).
- P-frames provide series of updates to most recent I-frame.
  - Predicted from past I- or P-frames using Motion Compensation.
Frame Reconstruction

- B-frames interpolate between frames represented by I's and P's.
- Predicted from either past and/or future I or P frames using motion estimation.
- Never used as reference pictures.

\[ \begin{array}{cccccccc}
I_1 & I_1+P_1 & I_1+P_1+P_2 & \cdots & \cdots \\
B_1 & B_2 & B_3 & \cdots & \cdots \\
B_4 & B_5 & B_6 & \cdots & \cdots \\
B_7 & B_8 & B_9 & \cdots & \cdots \\
\end{array} \]

Interpolations

Group of Pictures (GOP)

- What is the best combination of I-, P-, and B-frames?
  - Random access and fast reverse/forward searches.
    - Need more I-frames.
  - Computational complexity vs. the number of P- and B-frames.
    - More P- and B-frames, more computation, but better compression.
    - B-frames provide the best compression ratio, followed by P-frames.
  - Should there be lots of B-frames?
    - Having too many results in less correlation between B-frames and its reference frames.
Typical Ordering of Frames

- 1 second
- 28 pictures between 1 picture every 15th frame
- Reference (P) pictures
- (1/2 second at 30 Hz)

Display Order: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30

Macroblocks

- Works in units of 16×16 pixel macroblocks.
- 4:2:0 subsampling
Chapter 4: Video Compression

Outline
- Introduction
- MPEG Overview
- MPEG Encoding
  - Motion Compensation
  - Frame Encoding
- MPEG Video Bit Stream
- MPEG Video Standards
- H.264/AVC

MPEG Encoding

- DCT
- Motion Compensation
- IDCT
- Quantizer (Q)
- regulator
- Controls Encoding Rate
- VLC
- Buffer
- Output
- Motion vectors
- Current Frame
- Reference Frame
**I-frame Encoding**

- Same as JPEG encoding.
- Also called *intra-frame coding*.

---

**P-frame Encoding**

- Motion Compensation
Motion Compensation

- Exploit temporal redundancy of video.
- Detect the displacement of corresponding pixels or regions (i.e., macroblocks) in these frames and measure their differences.
- Motion compensation
  - Motion vector search (Motion estimation)
  - Derivation of prediction error
- Also called inter-frame prediction.

Motion Estimation

- For every macro-block, find the region in the previous frame that most closely matches the macro-block.

Types of Motion Estimation
- P-frames use Forward Prediction
- B-frames use Bidirectional or Motion Compensated Interpolation
**Best-Match Criterion**

- Block matching is done only on the Luminance component of frames.
- The best matching block is determined by minimizing a cost function, **Mean Absolute Difference** (MAD):

\[
MAD(i, j) = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} |S_{c}(x+i, y+j) - S_{r}(x+m+i, y+n+j)|
\]

where \(S_c\) and \(S_r\) are pixel values of \(M \times N\) block in the current and reference frames, respectively, and \(-p \leq i, j \leq p\).

**Search Algorithms**

- Searching over the entire previous frame is too expensive.
- Limit the search to a given area, centered around the macro-block and larger than the macro-block.
- **How large is large enough?**
  - MPEG-1 supports up to a range of \([-1024, 1023]\)!
Full Search

- Check every possible match within some spatial region.
- Must perform \((2p+1)^2 = 169\) absolute difference operations per macro-block.
- Many heuristic algorithms:
  - 3-step search
  - 2-D Logarithmic Search
  - Orthogonal Search
  - and others...
- Based on Principle of Locality, which suggests that very good matches, if they exist, are likely to be found in the neighborhood of other good matches.

3-Step Search

- Tests 8 points around the center. Step size initially set to \(p/2\).
- Reduces step size by 1 at each step.
2D Logarithmic Search

- Initial step size is $2^{\lfloor \log p \rfloor - 1}$. For $p = 6$, step size is 2.
- Step size is halved if the best match is at the center or the current minimum point reaches the search window boundary.
- When step size is 1, examine all blocks around the center.

```
+6-6
+6
-6
```

M(0,0)

Step 1
Step 2
Step 3
Step 4
Step 5

2D Logarithmic Search (cont.)

- Many variations of this algorithm (e.g., how step size is reduced).
- Below is an example with initial step size = 4.

```
+6-6
+6
-6
```

M(0,0)

Step 1
Step 2
Step 3
Step 4

Chapter 4: Video Compression
Orthogonal Search Algorithm

- Hybrid of Three Step Search and 2D Logarithmic
- Initial step size is $\lceil (p+1)/2 \rceil$, and step size halved after each horizontal + vertical search.

Motion Vector Encoding

- Motion vector = $(x, y)$ offset to best match
  - Sign is always relative to the current picture.
- Number of vectors needed
  - 1 for forward or backward prediction.
  - 2 for bi-directional (interpolated) prediction.
- Differentially encoded relative to motion vectors of previous adjacent block.
- Entropy code the result.
- Whenever a motion vector is determined for luminance pixels, corresponding chrominance pixels are used to form chrominance prediction.
Deviation of Prediction Error

- But wait! Just having a motion vector is not enough.
- What if the object is shifted and rotated? It will cause prediction error.
- Difference between pixels in predicted and referenced blocks are encoded.

Non-MC Coded Macroblock

- Sometimes, a good match cannot be found => prediction error exceeds predefined threshold.
- Macroblock itself is encoded => called non-motion-compensated macroblock.
- Treated just like a macroblock for I-frame plus motion vector.
### Motion Compensation Example

- **Previous frame**
- **Current frame**
- **Current frame with displacement vectors**
- **Motion-compensated Prediction error**

### B-frame Encoding

- Some macroblocks need information from past and future (I or P) frames.
- Example: The darken macroblock in the Current frame does not have a good match from the Past frame, but will find a good match in Future frame.
### B-frame Encoding (cont.)

Current Frame

\[ F_p(x', y') = \frac{1}{2} \left( F_{past}(x, y) + F_{future}(x, y) + MB_{reference} \left( MB_{past} + MB_{future} \right) \right) \]

\[ MB_{reference} = \left( MB_{past} + MB_{future} \right) / 2 \]

Future Frame

DCT + Q + RLE

Huffman Code

Past Frame

### Half-Pixel Precision

- Reduces prediction error for MC.
  - For half-pixel precision, MPEG-1 supports a range of \([-512, 511.5]\)

\[
\begin{align*}
  a &= A \\
  b &= (A + B + 1)/2 \\
  c &= (A + C + 1)/2 \\
  d &= (A + B + C + D + 2)/4
\end{align*}
\]
Quantization

- Uses different Quantization table for intra- and inter-coding.

<table>
<thead>
<tr>
<th>Intra-coding</th>
<th>Inter-coding</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>

Quantization

- Quantization for DCT coefficients in intra-mode:

\[ DCT_{Q_1}(i,j) = \text{round} \left( \frac{8 \times DCT(i,j)}{Q_1(i,j) \times \text{scale}} \right) \]

- Quantization for DCT coefficients in inter-mode:

\[ DCT_{Q_2}(i,j) = \left\lfloor \frac{8 \times DCT(i,j)}{Q_2(i,j) \times \text{scale}} \right\rfloor \]

- \textit{scale} (Q-scale) is an inter in the range of \([1, 31]\)
  - Q-scale can be changed on macroblock basis to provide bit-rate control.
MPEG Decoding

Input
Buffer → VLD → 𝑄⁻¹ → IDCT → Output

Frame Memory → Motion Compensation

Motion vectors

GOP Sequence

- Frames are not transmitted in order, must be buffered and re-ordered at the decoder

Display order
I B B P B B P

Transmit order
I P B B P B B
Chapter 4: Video Compression

Outline

- Introduction
- MPEG Overview
- MPEG Encoding
  - Motion Compensation
  - Frame Encoding
- MPEG Video Bit Stream
- MPEG Video Standards
- H.264/AVC

MPEG Video Bitstream
Sequence Information

- Bitrate
- Buffer size
- Constrained parameter flag

- Width/height
- Aspect ratio
- Frame rate

Group of Pictures Information

- Describes structure of GOP
  - GOP closed (no IBP or IBBP)
  - Broken link

- Bit field with SMPTE (Society of Motion Picture and Television Engineers) time code (Hours, minutes, seconds, frame).
Closed GOP & Broken Link

- **Broken link**
  - Set to '0' during encoding.
  - Set to '1' to indicate that the first consecutive B-Pictures (if any) immediately following the first coded I-frame following the GOP header may not be correctly decoded because the reference frame which is used for prediction is not available (because of the action of editing).
  - So that a decoder may avoid displaying these B-Pictures

- **Closed GOP**
  - Set to '1' to indicate that these B-pictures have been encoded using only backward prediction.
  - If the previous pictures have been removed by editing, broken_link may be set to '1'. However if the closed_gop bit is set to '1', then the editor may choose not to set the broken_link bit as these B-Pictures can be correctly decoded.

---

Picture Information

- **picture_start_code**
  - Indicates how full decoder's buffer should be before starting decode

- **0x00000100**
  - Indicates whether half pixel motion vectors are used
  - Indicates whether I, P, or B-frame
Chapter 4: Video Compression

Slice Information

- Each slice may contain variable number of MBs.
- Provides additional flexibility in bit-rate control.
- Slice concept is important for error recovery.
  - If part of a slice is lost, the decoder will skip the rest of the slice and start decoding from the beginning of the next slice.

Slice start code

- How is the quantization table scaled in this slice?
- One slice of macroblocks
- What line does this slice start on

---

Chapter 4 Video Compression

43

44
Macroblock Information

- Indicates
  - Intra or non-intra
  - Q-scale is specified or not
  - Motion Vector exists or not
  - CBP exists or not.
- How is the Q-table scaled in this MB.
- Encoded as 1-31 (5 bits).
- (6 bit) Bitmap indicating which blocks are coded.
  - None for I-frame
  - 1 for P-frame
  - 2 for B-frame
  - Address of MB

Address Increment

- Each macroblock has an address.
  - MB_WIDTH = width of luminance / 16
  - MB_ROW = row # of upper left pixel / 16
  - MB_COL = col. # of upper left pixel / 16
  - MB_ADDR = MB_ROW * MB_WIDTH + MB_COL
- Decoder maintains PREV_MBADDR.
  - Set to -1 at beginning of picture.
  - Set to \((SLICE\_ROW*MB\_WIDTH-1)\) at slice header.
- MB address increment added to PREV_MBADDR provides current macroblock address.
  - PREV_MBADDR set to current macroblock address.
I-frame Macroblock

- First Non-zero AC Coeff. (variable bit length)
- Last Non-zero AC Coeff. (variable bit length)
- DC Size (2-7 bits)
- DC Bits (0-8 bits)
- EOB (2 bits)
- Luminance Blocks
  - U Block
  - V Block
- Q Scale (5 bits)
- Macroblock Type (1 or 2 bits)
- Macroblock Address Increment (1-bit)

P-frame Macroblock

- CBP indicates which blocks have enough prediction error to warrant coding.
- CBP = 0000002 indicates no blocks needed coding.
- May or may not exist depending on 6-bit Coded Block Pattern (Huffman coded to 3-9 bits).
- Differentially encoded relative to motion vectors of previous adjacent macroblock.
**B-frame Macroblock**

May or may not exist depending on 6-bit Coded Block Pattern (Huffman coded to 3-9 bits).

- **Luminance Blocks**
- **Block Pattern** (3-9 bits)
- **Backward Motion Vector** (variable)
- **Forward Motion Vector** (variable)
- **Q Scale** (5 bits)
- **Macroblock Type** (1-6 bits)
- **Macroblock Address Increment** (variable)

**Skipped Macroblocks**

- No motion vector and no appreciable prediction error to encode (it might be all zeroed after quantization) => *skip macroblock*.
- Decoder detects this by looking at the address increment.
- First and last blocks of a slice must not be skipped.
- For B-frame macroblock, not allowed to skip following an I-frame macroblock.
MPEG-1 Compression Ratio

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

Outline

- Introduction
- MPEG Overview
- MPEG Encoding
  - Motion Compensation
  - Frame Encoding
- MPEG Video Bit Stream
- MPEG Video Standards
- H.264/AVC
MPEG Video Standards

- **MPEG-1**
  - Became a standard in 1991
  - 1.2 - 1.5 Mbps
  - 352 × 240 at 30 fps (NTSC) or 352 × 288 pixels at 25 fps (PAL)
  - “VHS quality”

- **MPEG-2**
  - 3 - 6 Mbps (352 × 240, 30 fps)
  - Up to 1920 × 1152 pixels, 30 fps (60-80 Mbps)
  - Interlaced video
  - “TV and HDTV quality”

- **MPEG-4**
  - Coding of Audio-Visual Objects
  - One generic toolbox for many different kinds of application
  - Interactivity
  - Low bitrate

- **MPEG-7**
  - Multimedia content description interface

MPEG-2

- **Targeted for Digital TV**
- **Supports higher bit rates, e.g., 4-80 Mbps.**
  - 1920x1152, 60 fps, HDTV (US), 80 Mbps
  - 1440x1152, 60 fps, HDTV (Europe), 60 Mbps
  - 720x576, 30 fps, CCIR601 video (NTSC/PAL), 15 Mbps
  - 352x288, 30 fps, SIF video (VHS quality), 4 Mbps, backward compatible with MPEG-1
MPEG-2

- Supports *interlaced* and non-interlaced (progressive) frames
- 4:2:0 as well as 4:2:2 and 4:4:4 subsampling
- Non-linear quantization
- Provides *base layer* (offers basic video quality, i.e., MPEG-1 quality) and *enhanced layer*.
  - Suitable for networks with slow bit rate or VBR.
  - Scalable coding

Interlacing
MPEG-4

- To address the requirements of a new generation of highly interactive multimedia applications.
  - Content based
  - Random access (in time & objects)
- **Object-based coding instead of frame-base coding** to allow
  - Efficient coding
  - Interactivity
  - Scalability of audio and video content
  - Support for natural and synthetic audio and video
- To become standard for streaming AV media on the internet and via wireless networks.
- An audiovisual scene is coded representation of audiovisual objects related in space and time.
- **MPEG-4 applications**
  - Interactive TV
  - Streaming media on the web
  - Multimedia titles
  - Network video games
  - Mobile multimedia

Outline

- Introduction
- MPEG Overview
- MPEG Encoding
  - Motion Compensation
  - Frame Encoding
- MPEG Video Bit Stream
- MPEG Video Standards
- H.264/AVC
H.264/AVC

- Joint Video Team (JVT)
  - ITU-T Video Coding Experts Group (VCEG): H.26x series
  - ISO Motion Picture Experts Group (MPEG): MPEG-x series
  - H.264 (JVC) and MPEG4 Part 10 AVC (Advanced Video Coding) => H.264/AVC

- Key new features:
  - Enhanced motion-prediction capability
  - Use of a small block-size exact –match transform
  - Adaptive in-loop deblocking filter
  - Enhanced entropy coding methods

- Average bit rate reduction of 50% compared to any other standard and significant improvement in perceptual quality.

H.264/AVC

- Addresses a full range of video applications:
  - Low bit-rate wireless applications
  - Standard Definition/High Definition broadcast TV
  - Video streaming over internet
  - HD DVD content
  - Digital cinema application

- Represents the single largest improvement in coding efficiency and quality since the introduction of MPEG-2.
- H.264 is expected to displace MPEG-2/4 in many existing applications.
- Much higher complexity than previous standards (5/6 times more complex than MPEG-2)
Timeline of Video Development

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint ITU-T/MPEG Standards</td>
<td>H.262/MPEG-2</td>
<td>H.264/MPEG-4 AVC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPEG Standards</td>
<td>MPEG-1</td>
<td>MPEG-4 (Version 1)</td>
<td>MPEG-4 (Version 2)</td>
<td></td>
</tr>
</tbody>
</table>

Video Coding Layer (VLC)

[Diagram showing the process of video coding layer]
Design Feature Highlights

- **Improved Prediction Methods**
  - Variable block-size motion compensation with small block sizes.
  - Quarter-sample-accurate motion compensation.
  - Motion vectors over picture boundaries.
  - Multiple reference picture motion compensation.
  - Decoupling of referencing order from display order.
  - Decoupling of picture representation methods from picture referencing capability.
  - Weighted prediction.
  - Improved “skipped” and “direct” motion inference.
  - Directional spatial prediction for intra coding.
  - In-the-loop deblocking filtering.

- **Improved Coding Efficiency**
  - Small block-size transform
  - Exact-match inverse transform
  - Short word-length transform
  - Hierarchical block transform
  - Arithmetic entropy coding
  - Context-adaptive entropy coding

- **Features for robustness to data errors/losses**
  - Parameter set structure
  - NAL unit syntax structure
  - Flexible slice size
  - Flexible macroblock ordering (FMO)
  - Arbitrary slice ordering (ASO)
  - Redundant pictures
  - Data Partitioning
  - SPS/PSI synchronization/switching pictures

Variable Block-Size MC with Small Block Sizes

- Allows very precise segmentation of moving regions.
- Partitioned in 2 stages
  - 1st stage, determine first 4 modes: $16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8$
  - Once a mode is chosen, further partition into smaller blocks, e.g., $8 \times 8$ block: $8 \times 4, 4 \times 8, 4 \times 4$
  - At most 16 motion vectors may be transmitted for a $16 \times 16$ macroblock
  - Large computational complexity to determine the modes.
Variable Block-Size MC with Small Block Sizes

Various block sizes and shapes

Quarter-Sample-Accurate MC

\[ b = (b_1 + 16) >> 5 \]
\[ h = (h_1 + 16) >> 5 \]
clipped to 0~255

\[ j = (j_1 + 512) >> 10 \]
clipped to 0~255

\[ a = (G + b + 1) >> 1 \]

\[ e = (b + h + 1) >> 1 \]
Multiple Reference Picture MC

- The encoder can select for motion compensation among a larger number of pictures previously decoded and stored.
- Useful when dealing with motion that is periodic or in presence of camera switching between 2 scenes.

Directional Spatial Prediction for Intra-coding

- Intra-frame coding, such as for I-frame, generates high bit-rate.
- Intra prediction predicts the texture in current block using the pixel samples from neighboring blocks.
- Intra prediction for 4×4 and 16×16 blocks are supported.
Chapter 4: Video Compression

Directional Spatial Prediction for Intra-coding (4x4)

Mode 7 is selected

In-the-loop Deblocking Filtering

- Visible block boundaries at low bit-rates, called blocking artifacts.
- Applied when a large difference between samples near a block edge is detected.
In-the-loop Deblocking Filtering

Need to distinguish between true edges in the image and those created by quantization of the DCT coefficients.