Lecture 5: Video Compression Standards (Part2)

Dr. Jian Zhang
NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2007
jzhang@cse.unsw.edu.au
Level 4, 223 Anzac PRD
2780 Ext
Tutorial 3 :
Introduction to Histogram

- Image Histogram
4.2 Digital Video Coding (DVC) Structure – Hybrid MC/DPCM/DCT -- *Tutorial*

Diff. Image = \text{abs} \ [\text{diff. image}]
Block Based Motion Estimation

Tutorial 2

- Block base search

Reconstructed Frame

Current Frame

Search Window

Position of Current Block

Motion Vector

W = Search Range

16x16 -- Macroblock
Block Based Motion Estimation

**Tutorial 2**

- Block base search

Reconstructed Frame

Motion Compensated Frame

**Motion Vector**

W = Search Range

Search Window

Position of Current Block

**Motion Compensated MB**

16x16 -- Macroblock
Information Entropy — *Tutorial*

Information Entropy: The Entropy is defined as the average information content per symbol of the source. The Entropy, $H$, *can be expressed as follows*:

$$H = - \sum_{i=1}^{m} p_i \log_2 p_i \text{ bits}$$

Look at a source that contains $m$ possible symbols: $\{s_i, i=1,2..m\}$. The occurrence probabilities: $\{P_i, i=1,2..m\}$, So The info. content of a symbol $s_i$;

$$I_i = i( s ) = - \log_2 p_i \text{ bits}$$

The entropy of an information source is a function of occurrence probabilities. The entropy reaches the Max. when all symbols in the set are equally probable.

$$H = - \sum_{i=1}^{m} p_i \log_2 p_i \text{ bits}$$
Histogram & Entropy--**Tutorial**

Calculate: $P_1, P_2, \ldots, P_m$.

$$
H = - \sum_{i=1}^{m} p_i \log_2 p_i
$$

Entropy = 7.63 bits/pixel
Key operations in image/video coding -- Tutorial

- A 2-dimensional transform can be extended according to: $[C] = [C'][T] = [T][F][T]^T$

$$[C'] = 1/2 \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 5 & 6 & 8 & 10 \\ 6 & 6 & 5 & 7 \\ 4 & 5 & 3 & 6 \\ 8 & 7 & 5 & 5 \end{bmatrix}$$

$$[C] = 1/2 \begin{bmatrix} 11.5 & 12.0 & 10.5 & 14.0 \\ -0.5 & 0.0 & 2.5 & 3.0 \\ 1.5 & 1.0 & 2.5 & 1.0 \\ -2.5 & -1.0 & 0.5 & 2.0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 11.5 & 12.0 & 10.5 & 14.0 \\ -0.5 & 0.0 & 2.5 & 3.0 \\ 1.5 & 1.0 & 2.5 & 1.0 \\ -2.5 & -1.0 & 0.5 & 2.0 \end{bmatrix} \begin{bmatrix} 24.0 & -0.5 & 1.5 & -2.0 \\ 2.5 & -3.0 & 0.0 & -0.5 \\ 3.0 & -0.5 & -0.5 & 1.0 \\ -0.5 & -3.0 & 0.0 & -1.5 \end{bmatrix}$$

93% of energy now in one term that is in position (0,0)
Key operations in image/video coding -- Tutorial

- Discrete Cosine Transform
  - For a 2-D input block $U$, the transform coefficients can be found as $Y = C U C^T$
  - The inverse transform can be found as $U = C^T Y C$
  - The $N \times N$ discrete cosine transform matrix $C = c(k,n)$ is defined as:

$$
c(k,n) = \begin{cases} 
1 \sqrt{N} & \text{for } k = 0 \text{ and } 0 \leq n \leq N - 1, \\
\sqrt{2} N \cos \frac{\pi(2n+1)k}{2N} & \text{for } 1 \leq k \leq N - 1 \text{ and } 0 \leq n \leq N - 1.
\end{cases}
$$

- [1] Do you get any hint to do DCT transform for a 8x8 block?
Key operations in image/video coding
-- Tutorial

- Quantization weighting matrix (QWM)
  - Diff quantizers for diff. DCT Coef.
  - Default Qant. QWM

\[
\hat{C}_{ij} = 16 \times C_{ij} \div (Q \times W_{ij})
\]

Recommended MPEG quantization matrix

\[
\begin{bmatrix}
8 & 16 & 19 & 22 & 26 & 27 & 29 & 34 \\
16 & 16 & 22 & 24 & 27 & 29 & 34 & 37 \\
19 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\
22 & 22 & 26 & 27 & 29 & 34 & 37 & 40 \\
22 & 26 & 27 & 29 & 32 & 35 & 40 & 48 \\
26 & 27 & 29 & 32 & 35 & 40 & 48 & 58 \\
26 & 27 & 29 & 34 & 38 & 46 & 56 & 69 \\
27 & 29 & 35 & 38 & 46 & 56 & 69 & 83
\end{bmatrix}
\]

[2] Quantization results of the DCT coefficients, you can choose an appropriate quantization step-size Q \{4,32\}. W-Matrix. You need to tell the Q value and the QWM (left or right?)
Key operations in image/video coding --

**Tutorial**

- The distribution of 2-D DCT Coefficients after the quantization

Ref: H. Wu

- [3] Run-length (Zig-zag) results

Zig-Zag scan

\[
\begin{bmatrix}
68 & 3 & 5 & -2 & 0 & 0 & -2 & 0 \\
-10 & 0 & -4 & 3 & 0 & 0 & 0 & 0 \\
9 & 3 & 0 & 0 & -2 & 0 & 0 \\
3 & 2 & 0 & 3 & 0 & 2 & -2 & 0 \\
0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\
0 & 2 & -2 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]
Key operations in image/video coding --

Tutorial

- DC coefficient is coded differentially as (size, amplitude). There are 12 categories of size

\[ \Delta D_{C_i} = D_{C_i} - D_{C_{i-1}} \]

<table>
<thead>
<tr>
<th>Coeff</th>
<th>Size</th>
<th>[Code]</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>00</td>
<td>2+0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>010</td>
<td>3+1</td>
</tr>
<tr>
<td>2..3</td>
<td>2</td>
<td>011</td>
<td>3+2</td>
</tr>
<tr>
<td>4..7</td>
<td>3</td>
<td>100</td>
<td>3+3</td>
</tr>
<tr>
<td>8..15</td>
<td>4</td>
<td>101</td>
<td>3+4</td>
</tr>
<tr>
<td>16..31</td>
<td>5</td>
<td>110</td>
<td>3+5</td>
</tr>
<tr>
<td>32..63</td>
<td>6</td>
<td>1110</td>
<td>4+6</td>
</tr>
<tr>
<td>64..127</td>
<td>7</td>
<td>11110</td>
<td>5+7</td>
</tr>
<tr>
<td>128..255</td>
<td>8</td>
<td>111110</td>
<td>6+8</td>
</tr>
<tr>
<td>256..511</td>
<td>9</td>
<td>1111110</td>
<td>7+9</td>
</tr>
<tr>
<td>512..1023</td>
<td>10</td>
<td>11111110</td>
<td>8+10</td>
</tr>
<tr>
<td>1024..2047</td>
<td>11</td>
<td>111111110</td>
<td>9+11</td>
</tr>
</tbody>
</table>

Final code: 01100

Note: We have block₀; block₁; block₂; block₃.
The predicated DC value for block₀ is 0
The predicated DC value for block₁ is the DC value of block₀ ……
Key operations in image/video coding -- Tutorial

Zig-Zag scan

[4] VLC of the Zig-zag scan run-length pairs for the four 8x8 sub-image blocks
Digital Video Coding (DVC) Structure – Video Block Data Structure—Tutorial

- Typical MC/DPCM/DCT video coding architectures use the 4:2:0 (YCbCr) format as block data structure.
- The architecture consists of the following layered data structure for each picture:
  - Picture layer
  - Slice or group of blocks (GOB) layer
  - Macroblock (MB) layer: 16x61 pixels 4Y,Cb & Cr
  - Block size layer: 8x8 pixels.
- MC is applied to the MBs.

Ref: H. Wu
Digital Video Coding (DVC) — Tutorial Standards— MPEG-1 (ISO/IEC 11172)

- Intra coded picture (I-Picture):
  - Coded on their own (all MBs are intra) and server as random access.

- Predicted picture (P-Picture):
  - Coded with reference (MC predictions) to the previous anchor I or P picture.

- Bidirectionally predicted picture (B-Picture):
  - Coded with reference to the previous and/or future anchor I or P pictures (forward or backward MC prediction and/or linear interpolation).

Group of Pictures (GOP): start with I or B, end with I or P.

-2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12

Coding order: ...0,3,1,2,6,4,5,9,7,8...

N = number of pictures in GOP
M = prediction distance
(M−1 in-between B-pics)

Example: N=12, M=3
Digital Video Coding (DVC) — Tutorial Standards – MPEG-2 Scalability

2 layer spatially scalable coder

Spatiotemporal weighted Prediction in Spa-Scal. + ‘Pred’
5.1 Digital Video Coding (DVC) Standards– MPEG-2 Error Resilience

- Most digital storage media and communication channel are not error free.
- Channel coding is beyond the scope of the MPEG specification
- The compression scheme needs to robust to residual errors
- The error resilience schemes can be applied in both sides Encoder/decoder
  - The slice structure of MPEG allows resync. after an error to limit the bit errors within a certain area
  - Decoders may also attempt to conceal the effect of errors
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Five major parts form the prevision of error resilience
  - Error Detection:
    - It is desirable to detect quickly that an error has occurred and immediately take steps to prevent catastrophic degradation in video quality.
    - Error detection can be based on a syntactic or a semantic basis such as illegal VLC, more 64 DCT coef. etc.
  - Resynchronization:
    - When synchronization is lost with the arriving bitstream, the decoder will normally go looking for a unique resync. codeword
    - The resync. word is followed by sufficient data to allow the decoder to continue the decoding process from that point
  - Data recovery:
    - It explores to find the boundary of corrupted bitstream and then determine if the regain for decoding is utilized
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- **Temporal/Special Localization:**
  - It prevents the error propagation by inserting regular intra-coded picture, slides or macroblocks due to any transmission error.
  - Two type of error propagations:
    - temporal error propagation – errors are propagated into next frames
    - spatial error propagation – errors are propagated within a video picture

- **Concealment:**
  - The impact of any errors can be concealed using correctly decoded information from the current or previous pictures

- **Conclusion:**
  - While all of the previous parts are important, the use of effective error concealment will lead to the greatest improvement in subjective quality of decoded for error corrupted bitstreams
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- The impact of errors within decoded picture

Original picture

Error damaged picture
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

● Concealment Techniques
  ● Spatial concealment:

  Simple interpolation with above/below slices

  ● This technique is best suited to little spatial activity but is far less successful in areas where there is significant spatial details.
5.1 Digital Video Coding (DVC) Standards—MPEG-2 Error Resilience

- **Concealment Techniques**
  - Spatial concealment:
    - Error damaged picture
    - Spatial concealed picture
5.1 Digital Video Coding (DVC) Standards—MPEG-2 Error Resilience

- Concealment Techniques
  - Simple temporal concealment

- This technique is effective in the relatively stationary area but much less effective in the fast moving background
5.1 Digital Video Coding (DVC) Standards—MPEG-2 Error Resilience

- The impact of errors within decoded picture

Error damaged picture

Temporal concealed picture
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Concealment Techniques
  - Motion compensated concealment
    - It combines both temporal replacement and motion estimation
    - High correlation among nearby MVs in a picture
    - The assumption is the linear changes on MVs from the below to above MBs.
    - It has been proven that this scheme can improve significantly error concealment in moving areas of the picture.
5.1 Digital Video Coding (DVC) Standards—MPEG-2 Error Resilience

- Concealment Techniques
  - Motion compensated concealment
    - Encoder has option to transmit the concealment MVs with Intra-coded MB if they are above/below a corrupted MB
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- **Temporal Localization**
  - Cyclic intra-coded pictures

- Extra intra-coded I-pictures can be sent to replace B or P-pictures. Error propagation can be reduced at the cost of extra overhead in the bitstream
- This scheme will reduce the coding efficiency
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Temporal Localization
  - Cyclic intra-coded slices

- Extra intra-coded I-slices can be used to periodically refresh the frame from the top to the bottom over a number of frames
- The disadvantage is that the partial updating of the screen in a frame period will produce a noticeable “windscreen wiper” effect
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Temporal Localization
  - Cyclic intra-coded macroblock (MB) refreshment
    - Extra intra-coded MBs can be inserted to refresh the frame periodically over a fixed number of frames
    - This scheme has been widely used in current MPEG-4 codec
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Spatial Localization
  - Small Slices Mode (slice mode in MPEG-4 or H.263)
    - The 11MBs or 22MBs per slice can reduce the damage to decoded picture after a corrupted MB is detected.
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Spatial Localization
  - Small Slices Mode (slice mode in MPEG-4 or H.263)
    - Some significant factors impact on the number of bits required to code a slice in a picture (coding mode, picture activities etc.)

![Diagram showing spatial localization and small slices mode in digital video coding standards.](image)
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Spatial Localization
  - Fixed Bits Payload (Adaptive Slice Mode)
    - The disadvantage of the small slice mode does not take into account of the packing structure to accommodate one slice per packet
    - The fixed bits payload – adaptive slice mode addresses the weak point caused by small slice

![Diagram showing coding structure with start slices and packet cells]
5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- The Combination of Error Resilience Schemes
  - Spatial interpolation
  - Simple macroblock temporal replacement
  - Small slice mode
  - Adaptive slice mode
  - Small slice mode + Motion compensated concealment
  - Advanced error concealment

5.1 Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Layered Coding for error concealment
  - Base layer – important data/information
  - High layer – less important

- It applies to Spatial Scalability, SNR Scalability and Data Partitioning Scalability & Frequency Scalability
5.1 Demo for Error Resilience
5.2 Introduction to MPEG-4 Standard

- New Functionalities & Associated Application for MPEG-4

**Content-based Multimedia**
- Database searching
- Video Games
- 3-D movies

**Improved Coding Efficiency**
- Video telephone
- Video streaming
- Video authoring

**Content-based Interactivity**
- Mobility access
- Remote Monitoring
- Wireless LAN

**Universal Access**
- Video telephone
- Video streaming
- Video authoring

- Video Games
- 3-D movies
- Content-based Multimedia Database searching
- Mobility access
- Remote Monitoring
- Wireless LAN

- Improved Coding Efficiency
- Video telephone
- Video streaming
- Video authoring

- Content-based Interactivity
- Mobility access
- Remote Monitoring
- Wireless LAN

- Universal Access
- Video telephone
- Video streaming
- Video authoring
5.2 Introduction to MPEG-4 Standard

1. Systems
   - Defines the architecture, multiplex and syntax of the MPEG-4 standard.
   - Syntax component for configuring decoders with multiple coding tools
   - Defining the compositing operation permitted by decoders for user interaction

2. Video
   - Define the video coding algorithms and tools (normative parts)
   - specifies the coding and animation of synthetic and natural hybrid video (SNHC)

3. Audio

4. Conformance Testing

5. Technical Report – Software codec specification

6. DSM-CC Multimedia integration framework
5.3 MPEG-4 visual standard

- Video Coding and Communication
  - MPEG-4 standard: video part -- content based video coding scheme
    - To enable all these content-based functionalities, MPEG-4 relies on a revolutionary, content based representation of audiovisual objects.
    - As opposed to classical rectangular video (eg: MPEG1/2), MPEG-4 treats a scene as a composition of several objects that are separately encoded and decoded
  - The scalability at the object or content level enables to distribute the available bit-rate among the objects in the scene
    - Visually, more important objects are allocated more bits.
    - Encoded once and automatically played out at different rates with acceptable quality for the communication environment and bandwidth at hand.
5.3 MPEG-4 Visual Standard

Major Functionalities in MPEG-4 Visual Standard

- **Compression:**
  - Improved Coding Efficiency
  - Coding of Multiple Concurrent Data Streams

- **Content-Based Interactivity:**
  - Content-based Manipulation and Bitstream Editing
  - Hybrid Natural & Synthetic Data Coding
  - Improved Temp. Random Access

- **Universal Access:**
  - Content-based Scalability in Textures, Image and Video
    - Complexity scalability for en/de
    - Spatial scalability
    - Temporal scalability
    - Quality scalability
    - Fine grain scalability
  - Shape & Alpha Channel Coding
  - Robustness in Error-Prone Environments
5.3 MPEG-4 Visual Standard

- Access and manipulation of arbitrarily shaped images

Ref: Thomas Sikora

Object Based MPEG-4 Video Verification Model
1. In MPEG-4, scenes are composed of different objects to enable content-based functionalities.
2. Flexible coding of video objects
3. Coding of a “Video Object Plane” (VOP) Layer
5.3 MPEG-4 Visual Standard

- Video Object Planes (VOP’s)

The binary segmentation Mask is to extract the back/fore-ground layers

Ref: MPEG-4 AKIYO testing video sequence

Ref: Thomas Sikora
5.3 MPEG-4 Visual Standard

- Decomposition into VOP’s

Ref: Thomas Sikora

The overlapping VOP’s bring the opportunity to do the manipulation of Scene content.
5.3 MPEG-4 Visual Standard

- Video Object Plane” layered coding

MPEG-4 VOP-coder

Arbitrary VOP

Shape

Motion (MV)

Texture DCT

Rectangular VOP

Motion (MV)

Texture DCT

Ref: Thomas Sikora

Similar to H.263

Similar to H.263
5.3 MPEG-4 Visual Standard

- DCT-Based Approach for Coding VOP’s

Ref: Thomas Sikora

Block diagram of the basic MPEG-4 hybrid DPCM/transform codec structure
5.3 MPEG-4 Visual Standard

- Coding of a “Video Object Plane”

Ref: Thomas Sikora
5.3 MPEG-4 Visual Standard

- Background Padding for Motion Compensation

Ref: Thomas Sikora

Padded background

Previous Frame

Current Frame
5.3 MPEG-4 Visual Standard
One Typical Example -- Sprite Coding

1. A non-changing background only has to be transmitted once
2. Only foreground objects transmitted and re-inserted at the decoder
3. Object are much smaller than full video
5.3 MPEG-4 Visual Standard

- Scalability of Video Content – Spatial Scalability
5.3 MPEG-4 Visual Standard

- Scalability of Video Content – Temporal Scalability
5.3 MPEG-4 Visual Standard

- The MPEG-4 Video Verification Model (VM)
  - Coding of arbitrarily shaped images (VOP’s)
  - Coding of shape and Transparency information for each VOP
  - I,P and B frames (VOP’s)
  - Block-based hybrid MC/DCT coding approach based on ITU H.263 and MPEG techniques
  - Temporal and Spatial Scalability of VOP’s
5.4 Introduction to H.264 Video Coding Standard

- It started from the ITU-T H.26L Project (Long term)
- It aims to improve the coding efficiency up to 50% compared to MPEG-4 video coding standard
- In Dec. 2001, MPEG and ITU-T experts set up joint video team (JVT) to focus on this new standard.
- The final version of the standard has been approved by ITU-T 2003. H.264 video coding standard or MPEG-4 Part 10.
- The new technical approaches:
  - An Adaptive deblocking loop filter to remove the artifacts
  - Multiple frame for ME/MC
  - Predication in Intra mode
  - Integer transform
  - Optimized rate control strategy (my opinion)
5.4.1 Video Codec Structure of H.264
5.4.2 Video Codec Structure of H.264 (H.26L TML-8 Design Part 1 of 4)

- Hybrid of DPCM/MC/Trans coding as in Prior standards. Common elements include:
  - 16x16 macroblocks
  - Conventional sampling of chrominance and association of luminance and chrominance data
  - Block motion displacement
  - Motion vectors over picture boundaries
  - Variable block-size motion
  - Block transforms (not DCT, wavelets or fractals)
  - Scalar quantization (weighted)
5.4.3 H.264: Motion Compensation Accuracy

- Coder Control
- Transform/Quantizer
- Deq./Inv. Transform
- Deblocking Filter
- Intra_Frame Prediction
- Motion Comp. Predication
- Motion Estimator
- Entropy Coding
- MB of Input Image Signal
- Control Data
- Quant. Transf. coeffs
- Bitstream Output

- Intra/Inter
- Mode 1
- Mode 2
- Mode 3
- Mode 4
- Mode 5
- Mode 6
- Mode 7

1/4 (QCIF) or 1/8 (CIF) pel
5.4.4 H.264: Multiple Reference Frames

Multiple Reference Frames for Motion Compensation
5.4.4 H.264: Multiple Reference Frames

- Motion Compensation:
  - Multiple reference pictures (per H.263++ Annex U)
  - B picture prediction weighting
  - New “SP” transition pictures for sequence switching
  - Various block sizes and shapes for motion compensation (7 segmentations of the macroblock: 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4)
  - 1/4 sample (sort of per MPEG-4) and 1/8 sample accuracy motion