Overview: Image/Video Coding Techniques

Dr. Jian Zhang

NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2007
Pixel Representation

- Y,U,V Colour Space
  - Colour can be represented by Red, Green and Blue components (RGB).
  - Transform to YUV or YCbCr with less correlated representation.

\[
Y = 0.299R + 0.587G + 0.114B \\
U_t = \frac{B - Y}{2.03} \\
V_t = \frac{R - Y}{1.14}
\]

Note:
The Y component represents luminance and the two chrominance components (U,V) contain considerably less information than the luminance component. For this reason, chrominance is often subsampled such as YUV 420. The 420 format results in a \( \frac{3}{4} \) data reduction in each U and V component.
Basic Concepts

- Chrominance sub-sampling and formats
  - Y, Cb & Cr (YUV) Formats: 4:4:4 -> 4:2:2 -> 4:2:0 or 4:1:1 format

```
4:4:4

4:2:2

4:2:0
```

Formats:
Basic Transform coding

- 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 $Y = C U C^T$.
  - The $N \times N$ discrete cosine transform matrix $C = c(k,n)$ is defined as:

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

- The distribution of 2-D DCT Coefficients

Ref: H. Wu
**Simple Inter-frame Encoder**

**Encoder**
- Frame \( x(n) \)
- Error image \( e(n) \)
- Dequantised error image \( e'(n) \)
- Reconstructed frame \( \hat{x}(n-1) \)

**Transmission or Storage Media**

**Decoder**
- Reconstruction frame \( x'(n) \)
- Reconstructed frame \( \hat{x}(n) \)

- Step 1: Calculate the difference between the current and previous frames.
- Step 2: Quantise and encode the difference image.
- Step 3: Add the dequantised (residual) image to the previous frame to reconstruct the current frame of image.
Block Based Motion Est.

- Block base search

Reference Frame

![Diagram of motion estimation](attachment:image.png)

- Motion Vector
- Search Window: 16x16
- Position of Current Block: 16x16 -- Macroblock
- Current Frame: grid of squares

\[ w = \text{search range} \]

16

16

16
Block Based Motion Estimation

- Block base search

Reconstructed Frame

Current Frame

W = Search Range

Current Block
16x16 -- Macroblock
Block Based Motion Estimation

- Block base search

Reconstructed Frame

Motion Vector

16

W=Search Range

W

Search Window

Position of Current Block

Motion Compensated Frame

W

16

16

16

Motion Compensated MB

16x16 -- Macroblock
Motion Compensated DCT based Codec

Codec = encoder/decoder
The Key Operation in Video Codec – Motion Estimation

- Inter-frame coding (Zero MV) scheme achieves a low computation complexity while possibly results in a high error residuals due to no other search positions that are considered to find an optimal (low error residuals) position.

- Motion compensated coding scheme achieves a low error residuals while results in a high computationally complex due to an exhaustive search strategy within a defined search window.

- Several sub-optimum fast search techniques have been developed. However, the quality-cost trade-off is usually worthwhile.
Digital Video Coding (DVC) Standards—MPEG-1/2

- Intra coded picture (I-Picture):
  - Coded on their own (all MBs are intra) and served 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...

\[ N = \text{number of pictures in GOP} \]
\[ M = \text{prediction distance} \]
\[ (M-1 \text{ in-between B-pics}) \]

Example: \( N=12, M=3 \)

NICTA
Digital Video Coding (DVC) Standards—MPEG-1 (ISO/IEC 11172)

Forward prediction: Predict where the pixels in a current frame were in a past frame.
Backward prediction: Predict where the pixels in a current frame will go in a future frame.
Prediction for a macroblock may be backward, forward, or an average of both.
Advantages
Main Advantage:
  • High coding efficiency (gain/cost is significant)
  • No uncovered background problem
Main disadvantage: long delay and more memory to store two anchor frames

Ref: H. Wu
Digital Video Coding (DVC) Standards– MPEG-2 Scalability

- Spatial Scalability Types
  - Progress to progress
  - Progress to interlaced
  - Interlaced to progress
  - Interlaced to interlaced
Digital Video Coding (DVC) Standards—MPEG-2 Scalability

2 layer spatially scalable coder

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

- Spatiotemporal weighted Prediction
Comparison Figures

Uncompressed ITU-R Rec. 601:
- 720 x 576 pixels at 25 frames per second (4:2:2);
- 8-bit per component, leading to
  Total bit rate  = 25 frames/s x (720+360+360) x 576 component pixels/frame x 8 bits per component pixel
  = 166 Mbps,
  and a 90 minute movie requires over 100 GBytes.
- 10-bit per component, leading to
  Total bit rate  = 25 frames/s x (720+360+360) component pixels/frame x 10 bits per component pixel
  = 207 Mbps,
  a 90 minute movie requires over 135 GBytes.

Digital Storage Capacities:
- CD/VCD: 640~700 MB, typically at 1.2Mbps, (142:1 compression required for storage and 138:1 compression required for bandwidth);
- DVD: 4.6 GB, at 4.5 to 6 Mbps (21:1 compression required for storage, 28:1 for bandwidth);
- PC Hard Disk: 40 to 80 GB with PCI bus bandwidth of ~132 MBps

Ref: H. Wu
Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- Concealment Techniques
  - Spatial concealment:

  ![Diagram](Simple interpolation with above/below slices)

  MB before/above the lost MB

  Lost MB

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

- Concealment Techniques
  - Spatial concealment:

  ![Error damaged picture](image1)
  ![Spatial concealed picture](image2)

Error damaged picture
Spatial concealed picture
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

![Diagram showing previous and current frames with a lost macroblock (MB)]
Digital Video Coding (DVC) Standards–MPEG-2 Error Resilience

- The impact of errors within decoded picture

Error damaged picture

Temporal concealed picture
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
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
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
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
MPEG-4 Visual Standard

- Access and manipulation of arbitrarily shaped images

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

Ref: Thomas Sikora
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
MPEG-4 Visual Standard

- Decomposition into VOP’s

Ref: Thomas Sikora

Background Layer VOP

Foreground Layer VOP

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

- Video Object Plane” layered coding

Ref: Thomas Sikora

MPEG-4 VOP-coder

Arbitrary VOP

Rectangular VOP

Similar to H.263

Shape

Motion (MV)

Texture DCT

Motion (MV)

Texture DCT

Similar to H.263
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
Consultation time

- A special consultation time will be arranged before the Exam. Please check your email. The special consultation time will be sent from comp9519@cse.unsw.edu.au

- Good Luck for your final Exam.