Basic Transform coding **Tutorial 2**

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

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

The distribution of 2-D DCT Coefficients

Ref: H. Wu
JPEG DCT-Based Encoding

Routing 2

Coding of DCT Coefficients (DC)

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

<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</td>
<td>2</td>
<td>011</td>
<td>3+2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0110</td>
<td>3+3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>101</td>
<td>3+4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>110</td>
<td>3+5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1110</td>
<td>4+6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>11110</td>
<td>5+7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>111110</td>
<td>6+8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>1111110</td>
<td>7+9</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>11111110</td>
<td>8+10</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>111111110</td>
<td>9+11</td>
</tr>
</tbody>
</table>

Final code: 01100

Coding of DCT Coefficients (AC)

Routing 2

- AC coefficients are re-arranged to a sequence of (run, level) pairs through a zigzag scanning process.

- Level is further divided into (size categories, amplitude).

- Run and size are then combined and coded as a single event (2D VLC).

- An 8-bit code ‘RRRRSSSS’ is used to represent the nonzero coefficients:
  - The SSSS is defined as size categories from 1 to 11.
  - The RRRR is defined as run-length of zeros in the zig-zag scan or number of zeros before a nonzero coefficient.
  - The composite value of RRRRSSSS is then Huffman coded.

Ex: 1) RRRRSSSS=11110000 represents 15 run ‘0’ coef. and followed by a ‘0’ coef.

2) Multiple symbols used for run-length of ‘0’ coef. exceeds 15.

3) RRRRSSSS=000000000 represents end-of-block (EOB).

Zig-Zag scan
Inter-frame Encoder

Tutorial 2

Frame x(n)
Error image e(n)
Dequantised error image e'(n)
Reconstructed frame x'(n-1)
Reconstructed frame x'(n)

Encoder
Transmission or Storage Media
Decoder
Reconstructed frame 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 Estimation

Tutorial 2

- Block base search

Reconstructed Frame

Motion Vector

W=Search Range

16

16

Search Window
Position of Current Block

Current Frame

Motion Compensated Frame

Reconstructed Frame

Motion Vector

W=Search Range

16

16

Search Window
Position of Current Block

Motion Compensated MB
16x16 -- Macroblock
Digital Video Coding (DVC) Structure

Hybrid MC/DPCM/DCT Tutorial 2

Codec = encoder/decoder

4.1 Digital Video Coding (DVC) Standards—MPEG-2 Scalability

Spatial Scalability
- A spatially scalable coder operates by filtering and decimating a video sequence to a smaller size prior to coding.
- An up-sampled version of this coded base layer representation is then available as a predictor for the enhanced layer.
- As prediction is performed in the spatial domain, the coding at the base layer can take any other standards including (MPEG-1 or H.261).
- This is an important feature to address compatibility in layered codec.
4.1 Digital Video Coding (DVC) Standards– MPEG-2 Scalability

- Spatial Scalability Types
  - Progress to progress
  - Progress to interlaced
  - Interlaced to progress
  - Interlaced to interlaced

- Spatially interpolated fractional layer
- Spatially interpolated fractional layer
- Spatially interpolated fractional layer
- Spatially interpolated fractional layer

2 layer spatially scalable coder

Spatiotemporal weighted Prediction

- Data partitioning permits a video bitstream to be divided into two separate bitstreams
  - The BL contains the more info. including address and control info. as well as lower order DCT coeffients.
  - The HL contains the rest info. of the bitstream
  - The syntax elements in BL are indicated by properly breakpoint (PBP)
  - Some syntax elements in BL are redundant in HL to facilitate error recovery
  - It has the advantage to introduce almost no additional overhead
  - The disadvantage of this scheme: considerable drift occurs if only the BL is available to a decoder.
4.1 Digital Video Coding (DVC) Standards– MPEG-2 Scalability

- **Data partitioning**

<table>
<thead>
<tr>
<th>Priority Break Point</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>All data at sequence, GOP, Pic and slice layers</td>
</tr>
<tr>
<td>66</td>
<td>PBP=65 plus MB data to MB type</td>
</tr>
<tr>
<td>67</td>
<td>PBP=66 plus data to MB motion Vectors</td>
</tr>
<tr>
<td>0</td>
<td>PBP=67 plus MB data from CBP to DC (or 1\textsuperscript{st} non-zero) Coeff.</td>
</tr>
<tr>
<td>1</td>
<td>PBP=0 plus to first coeff. Following DC to first non-zero coeff after the first coeff. in the scan order</td>
</tr>
<tr>
<td>2</td>
<td>PBP=0 plus up to first non-zero coeff after the 2\textsuperscript{nd} coeff in the scan order</td>
</tr>
<tr>
<td>j</td>
<td>PBP=0 plus to first non-zero coeff after the jth coeff in the scan order</td>
</tr>
</tbody>
</table>

4.2 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.
4.2 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


4.2 MPEG-4 Visual Standard

- Video Object Planes (VOP’s)

Original
Binary Segmentation Mask

The binary segmentation Mask is to extract the back/foreground layers

Ref: MPEG-4 AKIYO testing video sequence

COMP9519 Multimedia Systems – Lecture 4 – Slide 26 – J Zhang

4.2 MPEG-4 Visual Standard

- Decomposition into VOP’s

Background Layer VOP
Foreground Layer VOP

The overlapping VOP’s bring the opportunity to do the manipulation of Scene content


4.2 MPEG-4 Visual Standard

- Video Object Plane” layered coding

Arbitrary VOP
Rectangular VOP

MPEG-4 VOP-coder

Shape Motion (MV) Texture DCT

Similar to H.263

Motion (MV) Texture DCT

Similar to H.263

Ref: Thomas Sikora

4.2 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

4.2 MPEG-4 Visual Standard

- Coding of a “Video Object Plane”

Ref: Thomas Sikora

4.2 MPEG-4 Visual Standard

- Background Padding for Motion Compensation

Ref: Thomas Sikora

Previous Frame

Current Frame

Padded background

4.2 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
4.3 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)

4.3 Video Codec Structure of H.264

- 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)

4.3 H.264: Motion Compensation Accuracy

- Motion Estimator
- Motion Comp. Prediction
- Deblocking Filter
- Intra_Frame_Transform
- Deq./Inv. Transform
- Quant. Transf. coeffs
- Entropy Coding
- Bitstream Output

- Mode 1: 0 0 0 0
- Mode 2: 1 0 0 0
- Mode 3: 0 1 0 0
- Mode 4: 0 0 1 0
- Mode 5: 0 0 0 1
- Mode 6: 1 0 1 0
- Mode 7: 0 1 1 0
- Mode 8: 0 0 1 1

- 1/4 (QCIF) or 1/8 (CIF) pel
4.3 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