Basic Transform coding *Tutorial 2*

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

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

- The distribution of 2-D DCT Coefficients

Ref: H. Wu

$$\begin{bmatrix} 1 & 3 & 0 & -2 & 0 & 0 & -2 & 0 \\ 3 & 5 & -2 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & -2 & -2 & 0 & 0 \\ -2 & -2 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
JPEG DCT-Based Encoding

Coded DCT Coefficients (DC)
- DC coefficient is coded differently as (size, amplitude). There are 12 size categories

Coded DCT Coefficients (AC)
- AC coefficients are re-arranged to a sequence of (run, level) pairs through a zig-zag 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) RRRRSSS=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) RRRRSSS=00000000 represents end-of-block (EOB)
**Inter-frame Encoder**

**Tutorial 2**

![Diagram of inter-frame encoder and decoder process](image)

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

**Tutorial 2**

- **Block base search**

![Diagram of block based motion estimation](image)

- **Current Frame**
- **Search Window**
- **Position of Current Block**
- **Motion Vector**
- **W = search range**
- **16x16 -- Macroblock**

![Motion Compensated Frame](image)

**Motion Compensated MB**

**16x16 -- Macroblock**
4.1 Introduction to Video Coders – Motion Compensated Coder

- Motion Compensated Coder is a lossless coder
- Central to the operation of the coder is the frame store
- It contains one or more previously transmitted frames.
- A Macroblock (MB) can be transmitted directly to the decoder is called “intra” mode transmission
- A different block between the current block and a corresponding block in a transmitted frame in the frame store is called “inter” mode transmission
- This is a hybrid MC/DPCM

4.1 Introduction to Video Coders – Motion Compensated Coder

This is a lossy coder since a quantization scheme is introduced
- The key part for the coder is the quantizer to reduce the residue values after the motion compensation for inter-mode.
- This is a hybrid MC/DPCM with quantization coder
4.1 Introduction to Video Coders – Motion Compensated Coder with Quantization

- The pixel data for a motion compensated MB is DCT transformed before quantization.
- The 2-D DCT coef. of each 8x8 block is then re-arranged by zig-zag into a 1-D arrangement.
- A quantization scheme is applied to differently quantize each DCT coef. according to its frequency spectrum.
- This will achieve a better compression result based on perceptual weight to the HVS.
  - The HVS is less sensitive to high spatial frequencies than lower frequencies.
  - The low DCT coef. the finely quantization.
- This is a hybrid MC/DPCM/DCT coder.

4.1 Introduction to Video Coders – Motion Compensated DCT/Quantization Decoder

- The procedure followed by the decoder does the inverse of the process performed at encoder.
- Delta of Frame N and N+1
- Motion compensation Prediction
- Frame N
- Frame N+1
- NICTA
4.2 Digital Video Coding (DVC) Structure

- All the DVC standards are based on the Hybrid MC/DPCM/DCT video coding structure.

- Since the constant rate constrain for most of current video codec applications, the quantization scheme should be considered to achieve the maximum of rate/distortion (R-D) ratio.

- Standards defined the decoder process while provide verification model for industry to develop their encoder. Therefore, there are many challenges about how to develop advanced algorithms to realize encoder at low complexity, low power and high performance (e.g: R-D ratio).

4.2 Digital Video Coding (DVC) Structure – Hybrid MC/DPCM/DCT

- Encoder

4.2 Digital Video Coding (DVC) Structure – Video Block Data Structure

- 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.
4.3 Digital Video Coding (DVC) Standards -- Overview

- MPEG: Motion Pictures Experts Group
- MPEG-1 (ISO/IEC 11172, Nov 92).
  - Audio and video storage media such as CD-ROM 1x CD-ROM: 150 KB/s = 1.2 Mbps).
  - Targeted at 1 to 1.5 Mbps (~1.2 Mbps for video, and ~250 kbps for audio).
  - Digital TV: SDTV, HDTV, DVD, etc.
  - Wider range of bitrates: 4 to 80 Mbps (optimised for 4 Mbps).
  - Supports interlaced video and scalable coding.

4.3 Digital Video Coding (DVC) Standards-- ITU-T H.261

- Target on a very specific area -- videophone and video conferencing.
  - Originally targeted for m x 384 kbit/s (m=1,...,5), changed to p x 64 kbit/s (p=1,...,30) (ISDN rates) in 1988. Also called "p x 64".
  - 40 kbit/s to 2 Mbit/s.
- Required for low bit rates and low delays.
- It is part of an entire suite of standards which takes care of other aspects:
  - H.221 – Multiplexing, H.320 – Control & Indication
  - H.242 – Call setup, signaling, H.320 Terminal specification
- Fixed video formats: CIF and QCIF (YCbCr, 4:2:0) at ~ 30, 15, 10 and 7.5 frame/sec.
- A typical hybrid MC/DPCM/DCT coding structure is applied with addition of a loop filter after motion compensation. This is a low-pass filter with taps [1/4,1/2,1/4].
- Ref:

4.3 Digital Video Coding (DVC) Standards-- ITU-T H.261

- Macroblocks & mode selection

<table>
<thead>
<tr>
<th>Coded</th>
<th>Non-coded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q Same</td>
<td>Q changed</td>
</tr>
<tr>
<td>Intra</td>
<td>Inter</td>
</tr>
<tr>
<td>0001</td>
<td>0000 001</td>
</tr>
<tr>
<td>01</td>
<td>0000 0001</td>
</tr>
<tr>
<td>00</td>
<td>0000 0000 01</td>
</tr>
</tbody>
</table>

- MTYPE Coding type applied to each macroblock can vary (MTYPE):
  - Non-coded: Skipped, MC and MC+FIL
  - In intra-frame: with/without Q scale.
  - In inter-frame: with/without Q scale, and either MC or MC+FIL.
  - MTYPE is coded using Huffman VLC.
4.3 Digital Video Coding (DVC)
Standards– ITU-T H.261

- Key points for H.261 codec
  - Zigzag scan of DCT coefficients.
    - DC value is differentially coded (same as JPEG).
    - AC coefficients run-length coded with run and level jointly coded using truncated Huffman coding.
  - Rate control involves
    - Quantisation step size selection (at MB or GOB or Picture level)
    - MTYPE Coding type (mode) selection
  - The range of Motion vectors (MV) is restricted to \([-15,15]\).
    - Differentially coded using the MVs of the neighboring block.
    - Motion vector differences are Huffman VLC coded.
  - Intra-refreshment
    - To avoid error propagation: every MB should be intra-coded at least once in every 132 transmitted picture frames.

4.4 Digital Video Coding (DVC)
Standards– MPEG-1 (ISO/IEC 11172)

- Coded representation of moving pictures and associated audio stored on digital storage media

Basic Requirements:
- Generic video coding at 1 to 1.5 Mbps (~VHS and 1.2 Mbps for video, and ~250 kbps for audio)
- Fast forward/reverse: seek and play in FF/FR using access points.
- Random access to a frame in limited time: frequent access points
- System supporting audio-visual synchronized play and access

Typical features and parameters:
- Bi-directional in temporal processing (I,P,B frame)
- Larger motion compensation range with half pixel MC (no loop filters)
- Quantization table
- 4:2:0 format and SIF (~CIF) resolution 352x240@30 or 352x288@25
4.4 Digital Video Coding (DVC) Standards – MPEG-1 (ISO/IEC 11172)

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

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

The decoder needs to reorder using a delay, before displaying pictures.

Ref: H. Wu

Half-pel refinement of motion vectors using simple linear interpolation.

- Half-pel causes filtering of prediction image: loop-filter not required.

- The use of bi-directional prediction in MPEG-1 can also lead to sub-pixel accuracy motion compensation and top of half pixel MC to give better-than-half-pixel accuracy.

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

- Quantization weighting matrix (QWM)
  - Diff quantizers for diff. DCT Coef.
  - Default QWM
    \[ C_{ij} = 16 \times \frac{Q_{ij}}{(Q \times W_{ij})} \]

**Recommended MPEG quantization matrix**

\[
\begin{pmatrix}
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
\end{pmatrix}
\]

4.5 Digital Video Coding (DVC) Standards—MPEG-2 (ISO/IEC 13818)

- ISO/IEC 13818-2 (or ITU-T H.262) -- An extension of MPEG-1 standard
- Broadcast TV, cable/satellite TV, HDTV, DVD, etc
  - To achieve PAL quality at rates between 4 and 9 Mbits/s -- SDTV
  - To cover HDTV around 20 Mbits/s
  - To cope with fully interlaced content such as ITU-R 601 resolution (i.e. 720x576@50 Hz interlaced field rate).
- Extremely flexible to support adaptive field/frame processing (ME and DCT) and higher chroma resolution (4:2:2 and 4:4:4)
- Downloadable quantization matrix
- Scalable video for multi-quality video applications

4.5 Digital Video Coding (DVC) Standards—MPEG-2 (ISO/IEC 13818)

- Profile and level
  - DVC standard subset to satisfy the functional requirements of different applications
  - Each profile “levels” are defined to restrict the values of various parameters to build flexible encoder/decoder
- Support for interlaced content (selected on frame-by-frame basis):
  - Frame-pictures: frame encoded as a single picture (as in MPEG1).
  - Field-pictures: two fields encoded as two pictures (can use prediction).
- Motion estimation from either alternate field, or previous frame
- For Intra_coding
  - Alternative Intra VLC table, Zig-zag scan order….
4.6 Digital Video Coding (DVC) Standards– MPEG-2 Profile/Levels

- Profiles and levels provide a means of defining subsets of the standard syntax and thereby the decoder capabilities.
- A profile defines a subset of constraints upon the allowed values of parameters within the full syntax.
- Conformance tests will be carried out against defined profiles at defined levels.
- Three profiles have been defined at this stage:
  - Main Profile
  - Next Profile
  - Simple Profile

### MPEG-2 Profile/Levels

<table>
<thead>
<tr>
<th>Level/Profile</th>
<th>Simple 4:2:0</th>
<th>Main 4:2:0</th>
<th>Nextg 4:2:2 Scalable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pixels/line</td>
<td>1920</td>
<td>1920</td>
<td>1920</td>
</tr>
<tr>
<td>Lines/frame</td>
<td>1152</td>
<td>1152</td>
<td>1152</td>
</tr>
<tr>
<td>Frame/s</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Pixels/s</td>
<td>62.7 million</td>
<td>62.7 million</td>
<td>62.7 million</td>
</tr>
<tr>
<td><strong>High-1440</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pixels/line</td>
<td>1440</td>
<td>1440</td>
<td>1440</td>
</tr>
<tr>
<td>Lines/frame</td>
<td>1152</td>
<td>1152</td>
<td>1152</td>
</tr>
<tr>
<td>Frame/s</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Pixels/s</td>
<td>44.9 million</td>
<td>44.9 million</td>
<td>44.9 million</td>
</tr>
<tr>
<td><strong>Main</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pixels/line</td>
<td>720</td>
<td>720</td>
<td>720</td>
</tr>
<tr>
<td>Lines/frame</td>
<td>576</td>
<td>576</td>
<td>576</td>
</tr>
<tr>
<td>Frame/s</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Pixels/s</td>
<td>10.4 million</td>
<td>10.4 million</td>
<td>10.4 million</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pixels/line</td>
<td>352</td>
<td>352</td>
<td>Not defined</td>
</tr>
<tr>
<td>Lines/frame</td>
<td>288</td>
<td>288</td>
<td></td>
</tr>
<tr>
<td>Frame/s</td>
<td>30</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Pixels/s</td>
<td>2.53 million</td>
<td>2.53 million</td>
<td>Not defined</td>
</tr>
</tbody>
</table>

4.7 Digital Video Coding (DVC) Standards– MPEG-2 Scalability

- Scalable video coding means the ability to achieve more than one video resolution or quality simultaneously.
4.7 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 predicator 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.

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

- **Spatiotemporal weighted Prediction**

**Example Weight Table**

<table>
<thead>
<tr>
<th>Layer 1</th>
<th>Layer 2</th>
<th>Layer 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**SNR Scalability**

- It provides different quality reconstructions of the same spatial and temporal resolution at different layers
- The BL encoder is the same as a single layer encoder
- The EL bitstream is derived:
  - Calculating Delta of DCT = before quantization – after de-quantizing
  - Re-quantizing this Delta with a finer quantizer
- It provides high coding efficiency with small overhead compared to single layer service
- It may be subject to drift problem in various cases.

**SNR Scalability (mode 1 encoder)**

Drift will be introduced to the enhanced layer. This is because that the diff. refinement coeff. do not feed back into the lower MC Pred' loop at the encoder whereas they do so at the decoder. If only the base layer works (error or packet loss in enhanced layer), no drift is expected.
4.7 Digital Video Coding (DVC) Standards – MPEG-2 Scalability

- SNR Scalability (mode 2 encode)

Drift will not be to the enhanced layer. This is because that the diff. refinement coef. do feed back into the lower MC Pred' loop at the encoder. However, the drift will be introduced once the error or packet loss is occurred in the enhanced layer.

4.7 Digital Video Coding (DVC) Standards – MPEG-2 Scalability

- SNR Scalability (mode 2 decoder)

Data partitioning

- 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 coef.
  - The HL contains the rest info. of the bitstream
  - The syntax elements in BL are indicated by propery 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.
4.7 Digital Video Coding (DVC) Standards– MPEG-2 Scalability

- Data partitioning – bitstream example (PBP = 64)

<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 1st 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 2nd 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>