Lecture 3: Introduction to Audio, Video & Image Coding Techniques (II) – Coding Techniques

DR. JIAN ZHANG
Conjoint Associate Professor
NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2006
jzhang@cse.unsw.edu.au

3. Introduction to audio, video & image coding techniques (II)
- 3.1 Transform Coding
- 3.2 Still Image Coding Standard (JPEG)
- 3.3 Subband Coding
- 3.4 Temporal Redundancy and Prediction for Video Coding
- 3.5 Motion Estimation & Compensation Algorithms
- 3.6 Tutorial 2

3.1.1 Basic Transform coding
- Block Based Transform Coding
  - Transform coding is performed by taking an image and breaking it down into sub-image (block) of size nxn. The transform is then applied to each sub-image (block) and the resulting transform coefficients are quantized and entropy coded.

\[
F = \begin{bmatrix}
5 & 6 & 8 & 10 \\
6 & 6 & 5 & 7 \\
4 & 5 & 3 & 6 \\
8 & 7 & 5 & 5
\end{bmatrix}
\]

- A 2-Dimensional transform can be carried out in a separable way (i.e. first down the columns and then along the rows).
3.1.1 Basic Transform coding

- The 1-Dimensional transform is calculated according to 
  \[
  [C'] = [T][F];
  \]
- Where \([T]\) is the transform matrix

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

4 Coefficients contain 94% of total energy

- A 2-dimensional transform can be extended according to:

\[
[C] = [C'][T] = [T][F][T]^T
\]

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

93% of energy now in one term that is in position (0,0)

3.1.1 Basic Transform coding

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

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

8x8 2-D DCT

Original 8x8 pixel image

8x8 2-D DCT coefficients

\[
\begin{bmatrix}
545.87 & 22.53 & 50.40 & -21.20 & -0.62 & -4.91 & -22.37 & -12.56 \\
-76.02 & 8.84 & -40.14 & 30.71 & -4.25 & 2.03 & 15.01 & 8.62 \\
83.64 & 32.46 & 5.76 & -2.03 & -2.42 & -34.87 & 20.62 & -2.62 \\
32.91 & 25.27 & -10.25 & 34.64 & -4.21 & 29.62 & -39.55 & -11.59 \\
12.27 & 20.77 & -31.24 & 36.16 & -2.38 & -17.96 & 2.94 & -0.60 \\
11.49 & 5.52 & 11.37 & -12.96 & -0.92 & 14.98 & -3.26 & 9.73 \\
7.64 & 1.15 & -5.81 & 9.25 & -3.84 & -15.97 & 2.03 & -1.61
\end{bmatrix}
\]
3.1.1 Basic Transform coding

- The distribution of 2-D DCT Coefficients

### Ref: H. Wu

$$
\begin{bmatrix}
68 & 3 & -5 & 0 & 0 & -2 & 0 \\
-10 & 0 & 4 & -3 & 0 & 0 & 0 \\
9 & 3 & 0 & 0 & -2 & 0 & 0 \\
3 & 2 & 0 & 3 & 0 & -2 & -2 \\
0 & 0 & -2 & 2 & 0 & 0 & 0 \\
0 & -2 & 2 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

### Zig-Zag scan

3.1.2 Optimal Transform

- An Optimal transform should achieve:
  - Completely de-correlate the data
  - Maximize the amount of energy packed into the lowest order coefficients

**Given a 1-D general data column vector** $X$ **with the mean vector**, the covariance matrix $(2D)$ is defined as:

$$
COV(X) = E[(X - \bar{X})(X - \bar{X})^T]
$$

Where $E(\bullet)$ is the expectation operator. Given an orthogonal transform represented by the matrix $[T]$ and the corresponding transform coeff. Vector by $[Y] = [T][X]$, we have:

$$
COV(Y) = E((T(X - \bar{X}))(X - \bar{X})^T) = TCOV(X)T^T
$$

3.1.3 Transform Efficiency

- De-correlation and energy packing effects (Eg. DCT)

  **The transform domain covariance matrix** $COV(Y_{dct})$ **is generated by using the 1-D 4-point discrete cosine transform (DCT)**

$$
COV(Y_{dct}) = T_{dct}COV(X)T_{dct}^T
$$

**From the previous slides, there are two distinct effects by the DCT**

- All off-diagonal elements of the transform domain covariance matrix are significantly reduced, indicating that the transform coefficients are effectively de-correlated as a result of the transform;

- The majority of the data energy (as represented by the variance) has been transferred to a few low-order coefficients after the transform.

**The most desirable transform domain covariance matrix will have a diagonal form (with values corresponding to the eigenvalues of the COV matrix) which achieves 100% decorrelation efficiency for transform coefficients**

$$
[T][COV(X)][T]^T = \Lambda
$$
3.1.2 Optimal Transform

- The optimal Transform exists – eg. Karhunen-Loeve Transform (KLT), but has some limitations
  - The KLT has the maximum energy packing capability with completely de-correlation for signal in transform domain
  - KLT involves the estimation of the COV, its diagonalisation and the contraction of the basis vectors
  - The basis vectors (transform matrix) are not fixed, and must be generated for images with different correlation characteristics.
  - The transform remains of more theoretical than practical interest as far as image coding is concerned.
  - Due to the lack of a fast implementation, it is not used practically.

3.1.4 Discrete Cosine Transform

- A transform is calculated by finding the correlation between the input sub-image and a series of basis vectors
- The transform has separable property (2D transform = 1D transform along rows and then down the columns)
- The DCT has the advantage for video coding of good redundancy reduction with performance close to the KLT [K.R.Rao]

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

3.2 Still Image Coding Standard (JPEG)

- Joint Photographic Experts Group.
- Formed in 1986 by ISO and ITU-T (CCITT) and Became an international standard (IS) in 1991.
- General purpose, applicable to almost all continuous-tone still image application
- Lossy (DCT based) vs Lossless (2D DPCM)
- 8/12 bits sample precision up to 65535 lines and 65535 pixels per line
- Huffman vs arithmetic coding
- Sequential, progressive & hierarchical modes
- Compression ratio of 10:1 to 50:1.
- Colour-space independent: up to 255 colour components, components can be sub-sampled and interleaved, YUV is better than RGB
3.2 Still Image Coding Standard (JPEG)

- 3.2.1 JPEG DCT-Based Encoding
- 3.2.2 Coding of DCT Coeff.
- 3.2.3 JPEG DCT-based coding

3.2.1 JPEG DCT-Based Encoding

- Recommended JPEG quantisation matrix \( N_{jk} \) =

For luminance:

\[
\begin{bmatrix}
16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\
12 & 12 & 14 & 19 & 26 & 58 & 66 & 74 \\
14 & 13 & 16 & 24 & 40 & 57 & 69 & 86 \\
14 & 17 & 29 & 51 & 87 & 80 & 62 & 87 \\
18 & 22 & 37 & 68 & 109 & 103 & 77 & 99 \\
24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\
49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\
72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \\
\end{bmatrix}
\]

For chrominance:

\[
\begin{bmatrix}
17 & 18 & 24 & 47 & 99 & 99 & 99 & 99 \\
12 & 18 & 26 & 66 & 99 & 99 & 99 & 99 \\
14 & 24 & 26 & 99 & 99 & 99 & 99 & 99 \\
\end{bmatrix}
\]

- Quantisation of coefficients \( Y_{jk} \) with a quality factor \( Q_s \)

\[
\hat{Y}_{jk} = Q(Y_{jk}) = \text{round}\left(\frac{Y_{jk}}{N_{jk} \times Q_s}\right)
\]

- High-frequency coefficients and colour components can be quantised more.

3.2.2 Coding of DCT Coefficients (DC)

- DC coefficient is coded differentially as (size, amplitude). There are 12 categories of size.
3.2.2 Coding of DCT Coefficients (AC)

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 zigzag scan or number of zeros before a nonzero coefficient.
    - The composite value of RRRRSSSS is then Huffman coded.
- Examples:
  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=00000000 represents end-of-block (EOB).

1. Code the amplitude (sign and values) of the nonzero AC coefficient with VLC that is similar to the DC code table.
2. The Huffman tables for AC coef. are defined in Annex K of the JPEG standard.
3. Support for arithmetic coding is included (but is patented).
4. Arithmetic coding can give a 8.8%-17.4% improvement in compression over Huffman coding (mean of entropy coding) with fixed table.
3.2.3 JPEG DCT-Based decoding

- **Lossless JPEG**
  - Pixel based prediction
  - Prediction error entropy coded
  - Compression of around 2:1

- **Progressive JPEG**
  - Gradual build-up of an Image
    - Spectral selection
      - A subset of coefficients is transmitted during each pass for all blocks. One usually starts with the low frequency coef. and moves towards the high freq.
    - Successive approximation
      - The most selected coef. are coded and transmitted during the first pass. In each subsequent pass, one extra bit is coded and transmitted to increase the precision of the coef.

3.3 Subband Coding

- The fundamental concept behind Subband Coding is to split up the frequency band of a signal and then to code each subband using a coder and bit rate accurately matched to the statistics of the band
- This makes each subband at the lower sampling rate
- At the receiver, the subband are resampled, fed through interpolation filters and added to reconstruct the image
- With an appropriate choice of filters, perfect reconstructions can be achieved

3.3.1 Analysis/Synthesis Stages

- **Analysis**
  - A signal is first filtered to create a set of signals, each of which contains a limited range of frequencies. These signals are called subbands.
  - Since each subband has a reduced bandwidth compared to the original fullband signal, they may be downsampled. That is, a reduced number of samples may be taken of the signal without causing aliasing
- **Synthesis-- Reconstruction is achieved by:**
  - Upsampling the decoded subbands.
  - Applying appropriate filters to reverse the subbanding process.
  - Adding the reconstructed subbands together.
3.3.1 Analysis/Synthesis Stage

- 2D dimensional decomposition structure

- The formation of subbands does not create any compression in itself.
  - The same total number of samples is required to represent the subbands as is required to represent the original signal.
  - The subbands can be encoded efficiently:
    - The significance of the different spatial frequencies are not uniform, and this fact may be exploited by different bit allocations to the various subbands
    - The subbands are then encoded using one or more coders. Different bit rates or even different coding techniques may be used for each subband.

3.4 Temporal Redundancy and Prediction for Video Coding

- In video coding, it is necessary to explore the compression in the spatial domain (spatial redundancy in intra-picture) and in time domain (temporal redundancy in inter-picture)

- A video sequence is a series of images with limited motion between adjacent images in most of the time
3.4.1 Temporal Statistical Redundancy

- Temporal redundancy between pixels of adjacent frames of video sequence
- The pixel differences at the same spatial location between consecutive frames are typically small.

3.4.2 Inter-frame Encoder

- Conditional Replacement
  - Inter-frame coders produce differential signals that are effectively zero in the non-changing parts of the picture and non-zero only in the moving areas
  - It is needed to transmit differential signal values only for the moving areas of the picture
  - We call this technique – Conditional frame replacement or Inter-frame DPCM

3.5 Motion Estimation & Compensation Algorithms

- The football has moved a limited distance but its shape remains almost constant between two adjacent frames.
- To reduce these temporal redundancies, several motion compensation methods can be applied
3.5 Motion Estimation & Compensation Algorithms

- 3.5.1 Block based Motion Estimation
- 3.5.2 Full search algorithm
- 3.5.3 Motion Compensation
- 3.5.4 Fast Motion Estimation Algorithm
- 3.5.5 Motion Estimation to sub-pixel accuracy
- 3.5.6 Introduction to Video Coders

3.5.1 Block Based Motion Est.

- A common approach in predictive encoder is to attempt to compensate for motion which occurs between frames using block matching.
- Using this approach, the reconstruction of the previous frame is searched for the best match to the current frame on a block by block basis. The location of this matching block (called the motion vector) and the prediction error (residual) is then transmitted.

- Each picture is divided into block based subimage for motion estimation rather than using each pixel as a unit.
- To reduce computational and storage requirements, a limited search area is defined to the position around the current block.
- The range of possible displacement in the X and Y directions is +/- 16.
  - 5 bits to specify the Hor/Ver displacement.
- The total motion vector overhead is 10 bits/pixel (more than the entropy of the orig. image).
- 16x16 Macroblock is defined for ME.
3.5.2 Full Search Algorithm

- Step 1
- Step 2
- Step 3
- Step 33
- Search Window Size $W = \pm 16$
- Step 34
- Step 66
- Step 1089
- Step 1057

$W = \pm 16$

Total search steps = $33 \times 33 = 1089$
Total pixel operations = $1089 \times 256 = 278784$
For CCIR Rec.601 video (704 pixels x 576 pixels)
Pixel operation/frame = 441.6 million/frame
Pixel operation/second = 11.0 billion/second
(25 frame/second)

Conclusion:
Motion Estimation is Highly Computationally Intensive !!!!!!

3.5.3 Motion Compensation

“Previous frame (No. 1)”

“Current original frame (No. 2)”
3.5.3 Motion Compensation

“Motion compensated prediction (No. 2)”

3.5.3 Motion Compensation

“Motion compensated prediction error (No. x)”

3.5.4 Fast Motion Estimation Techniques

- Full search motion estimation is computationally complex.
- Several sub-optimum fast search techniques have been developed.
- Many work on the assumption that block matching will improve monotonically as the search moves closer to the optimum point.
- Since they do not examine all of the candidate blocks, the choice of matching block might not be as good as that chosen by a full search. However, the quality-cost trade-off is usually worthwhile.

3.5.4 Fast Motion Estimation Techniques

**Two-dimensional logarithmic search (TDL)**

- This and the following techniques are quadrant monotonic searches.
- Quadrant monotonic assumes that the value of the distortion function increases as the distance from the point of minimum distortion increases.
- Special case of the principle of locality: not only locality to optimal block, but also distance from optimal block.
- TDL is a multi-stage search which successively reduces the search area during each stage until the search area is trivially small.
- Search positions: $2 + 7 \log w$. Example: for $w=16$, 30 positions are searched, compared with 1089 for the full search.
- The processing of a search step relies on the previous search steps; therefore, the search steps cannot be performed in parallel.
3.5.4 Fast Motion Estimation Techniques

**Two-dimensional logarithmic search (TDL)**

Step 1: The block at the centre of the search area and four blocks at distance s from the centre on the X and Y axes are searched for a best match.

Step 2: If the position of best match is the centre, halve the step size (s/2). Else, if the best match is in one of the four outer positions, then it becomes the new centre point ([cx, cy]) for the next stage.

Step 3: If the step size s is 1, then all nine blocks around the centre are examined, and the best match chosen for the target block. Otherwise, blocks at positions ([cx, cy], [cx+s, cy], [cx-s, cy], [cx, cy+s], and [cx, cy-s]) are searched, and the algorithm goes to stage 2.

**Note:** The points [0, +4], [+4, +4], [+6, +4] are the minima at each stage, and finally [+7, +4] is chosen as the matching block.

For step size s and search window size w, the step update algorithm is given by:

\[ s_0 = 2^{\log_2 w - 1}, \quad s_n = s_{n-1} / 2 \]

---

3.5.4 Fast Motion Estimation Techniques

**“Three Step Search (TSS)”**

- Very similar to TDL search, and developed around the same time.
- The three step search tests eight points around the centre instead of four, with the position of minimum distortion becoming the new centre.
- After each stage the step size is reduced.
- For step size s and search window size w, the step update algorithm is given by:

\[ s_0 = 2^{\log_2 w - 1}, \quad s_n = 2^{\log_2 s_{n-1} - 1}. \]

---

3.5.4 Fast Motion Estimation Techniques

**Hierarchical and sub-pixel block matching**

**Algorithm**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Maximum number of search points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full search</td>
<td>$(2w+1)^2$</td>
</tr>
<tr>
<td>Ghanbari</td>
<td>$2+7\log_2 w$</td>
</tr>
<tr>
<td>PHS</td>
<td>$1+4\log_2 w$</td>
</tr>
<tr>
<td>SR</td>
<td>$3+2w$</td>
</tr>
<tr>
<td>KR</td>
<td>$1+6\log_2 w$</td>
</tr>
<tr>
<td>w</td>
<td>7 15 31</td>
</tr>
<tr>
<td></td>
<td>225 961 3969</td>
</tr>
<tr>
<td></td>
<td>22 30 37 13 17 21</td>
</tr>
<tr>
<td></td>
<td>17 33 65</td>
</tr>
<tr>
<td></td>
<td>18 25 31</td>
</tr>
</tbody>
</table>

---

3.5.4 Fast Motion Estimation Techniques


Project – Fast motion estimation

- Diamond Search algorithm
  - A two step gradient-based diamond search algorithm was proposed. It is based on the observation, that 53% (in larger motion case) – 99% (in small motion case) of the motion vectors are enclosed in a circular area with a radius of 2 pixels around the zero motion vector.

- Large diamond search pattern (LDSP) is used for the gradient-based coarse search. When the centred search position of the LDSP shows the minimum SDA, the small diamond search pattern (SDSP) is chosen for fine search.

3.5.4 Fast Motion Estimation Techniques

- Issues with fast motion estimation algorithms
  - Full search motion estimation is computationally complex.
  - Fast search algorithms only work based on the assumption that block matching will improve monotonically as the search moves closer to the optimum point, which may not be the case in general.
  - All block matching based motion estimation algorithms assume pixel motion within the block is uniform, which will not work if the block involves object deformation or varying motion speeds.
  - Overlapped block matching algorithm and hexagonal matching algorithms have been introduced to address these two issues.

3.5.5 Motion Estimation to Sub-pixel Accuracy

- After a single pixel accuracy of motion estimation, it can be extended to sub-pixel accuracy by performing bi-linear interpolation between pixels in the search area.
- This can lead to a worthwhile improvement in the performance of motion compensated prediction.