Transform coding - topics

- Principle of block-wise transform coding
- Properties of orthonormal transforms
- Discrete cosine transform (DCT)
- Bit allocation for transform coefficients
- Threshold coding
- Typical coding artifacts
- Fast implementation of the DCT

Principle of block-wise transform coding

![Diagram of transform coding process]

Original image
Transform A
Reconstructed image

Transform coefficients
Quantization & Transmission

Quantized transform coefficients
Inverse transform A⁻¹
Reconstructed block
Properties of orthonormal transforms

- **Forward transform**
  \[ y = Ax \]
  - \( N \times N \) transform coefficients, arranged as a vector
  - Transform matrix of size \( N^2 \times N^2 \)
  - Input signal block of size \( N \times N \), arranged as a vector

- **Inverse transform**
  \[ x = A^{-1}y = A^T y \]

- Linearity: \( x \) is represented as linear combination of "basis functions".
- Parseval's Theorem holds: transform is a rotation of the signal vector around the origin of an \( N^2 \)-dimensional vector space.

Separable orthonormal transforms, I

- An orthonormal transform is separable, if the transform of a signal block of size \( N \times N \) can be expressed by
  \[ y = AxA^T \]
  - \( N \times N \) transform coefficients
  - Orthogonal transform matrix of size \( N \times N \)
  - \( N \times N \) block of input signal
  - Kronecker product

- The inverse transform is
  \[ x = A^T y A \]

- Great practical importance: The transform requires 2 matrix multiplications of size \( N \times N \) instead one multiplication of a vector of size \( 1 \times N^2 \) with a matrix of size \( N^2 \times N^2 \)

  \[ \text{Reduction of the complexity from } O(N^4) \text{ to } O(N^3) \]
Separable orthonormal transforms, II

- 2-d transform realized by 2 one-dimensional transforms (along rows and columns of the signal block)

Criteria for the selection of a particular transform

- Decorrelation, energy concentration (e.g., KLT, DCT, . . .)
- Visually pleasant basis functions (e.g., pseudo-random-noise, m-sequences, lapped transforms)
- Low complexity of computation
Karhunen Loève Transform (KLT)

- Karhunen Loève Transform (KLT) yields decorrelated transform coefficients.
- Basis functions are eigenvectors of the covariance matrix of the input signal.
- KLT achieves optimum energy concentration.
- Disadvantages:
  - KLT dependent on signal statistics
  - KLT not separable for image blocks
  - Transform matrix cannot be factored into sparse matrices.

Comparison of various transforms, I

<table>
<thead>
<tr>
<th>Karhunen Loève transform (1948/1960)</th>
<th>Haar transform (1910)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walsh-Hadamard transform (1923)</td>
<td>Slant transform (Enomoto, Shibata, 1971)</td>
</tr>
<tr>
<td>Discrete Cosine Transform (DCT)</td>
<td></td>
</tr>
<tr>
<td>(Ahmet, Natarajan, Rao, 1974)</td>
<td></td>
</tr>
</tbody>
</table>

Comparison of 1-d basis functions for block size $N=8$
Comparison of various transforms, II

- Energy concentration measured for typical natural images, block size 1x32 [Lohscheller]:

  ![Graph showing energy concentration comparison]

- KLT is optimum
- DCT performs only slightly worse than KLT

Discrete cosine transform and discrete Fourier transform

- Transform coding of images using the Discrete Fourier Transform (DFT):
  - For stationary image statistics, the energy concentration properties of the DFT converge against those of the KLT for large block sizes.
  - Problem of blockwise DFT coding: blocking effects due to circular topology of the DFT and Gibbs phenomena.
  - Remedy: reflect image at block boundaries, DFT of larger symmetric block -> “DCT”
DCT

- Type II-DCT of block size \( M \times M \) is defined by transform matrix \( A \) containing elements

\[
a_{ik} = \alpha_i \cos \left( \frac{\pi (2k + 1)i}{2M} \right)
\]

for \( i, k = 0, \ldots, M - 1 \)

with \( \alpha_0 = \sqrt{\frac{1}{M}} \)

\( \alpha_i = \sqrt{\frac{2}{M}} \quad \forall i \neq 0 \)

2D basis functions of the DCT:

Amplitude distribution of the DCT coefficients

- Histograms for 8x8 DCT coefficient amplitudes measured for natural images (from Mauersberger):

- DC coefficient is typically uniformly distributed.

- For the other coefficients, the distribution resembles a Laplacian pdf.
Bit allocation for transform coefficients I

Problem: divide bit-rate $R$ among $M \times M$ transform coefficients $i$ such that resulting distortion $D$ is minimized.

Assumptions

\[ R = \sum_i R_i \quad \text{Total rate} \]

\[ D = \sum_i D_i \quad \text{Total distortion} \]

Rate for coefficient $i$  
Distortion contributed by coefficient $i$

lead to "Pareto condition"

\[ \frac{\partial D_i}{\partial R_i} = \frac{\partial D_j}{\partial R_j} \quad \text{for all } i, j \]

Bit allocation for transform coefficients II

- Additional assumptions “Gaussian r.v.” and “mse distortion” yield the optimum rate for each transform coefficient $i$:

\[ R_i = \max \left\{ 0, \frac{1}{2} \log_2 \left( \frac{\sigma_i^2}{\theta} \right) \right\} \text{ bit} \]

\[ D_i = \min \left\{ \sigma_i^2, \theta \right\} \]

- In practice, with variable length coding, one often uses “distortion allocation” instead of bit allocation
Bit allocation for transform coefficients III

- Extension to weighted m.s.e. distortion measure
  \[ D = \sum_i w_i D_i \]

  \[ R_i = \max \left\{ 0, \frac{1}{2} \log_2 \left( \frac{w_i \sigma_i^2}{\theta} \right) \right\} \text{ bit} \]

  \[ D_i = \min \left\{ \sigma_i^2, \frac{\theta}{w_i} \right\} \]

- Often implemented by scaling coefficients by \((w_i)^{\nu_2}\) prior to quantization (“weighting matrix”)

---

Threshold coding, I

- Transform coefficients that fall below a threshold are discarded.
- Implementation by uniform quantizer with threshold characteristic:

  ![Quantizer diagram](image)

  - Positions of non-zero transform coefficients are transmitted in addition to their amplitude values.
Threshold coding, II

Efficient encoding of the position of non-zero transform coefficients: zig-zag-scan + run-level-coding

order of the transform coefficients by zig-zag-scan

Threshold coding, III

DCT run-level-coding

Original 8x8 block

Reconstructed 8x8 block

scaling and inverse DCT

Mean of block: 185

(0,3) (0,1) (1,1) (0,1) (0,1)
(0,-1) (1,1) (1,1) (0,1) (1,-3)
(0,2) (0,-1) (6,1) (0,-1) (0,-1)
(EOB)

Mean of block: 185

(0,3) (0,1) (1,1) (0,1) (0,1)
(0,-1) (1,1) (1,1) (0,1) (1,-3)
(0,2) (0,-1) (6,1) (0,-1) (0,-1)
(EOB)

Mean of block: 185

(0,3) (0,1) (1,1) (0,1) (0,1)
(0,-1) (1,1) (1,1) (0,1) (1,-3)
(0,2) (0,-1) (6,1) (0,-1) (0,-1)
(EOB)

Mean of block: 185

(0,3) (0,1) (1,1) (0,1) (0,1)
(0,-1) (1,1) (1,1) (0,1) (1,-3)
(0,2) (0,-1) (6,1) (0,-1) (0,-1)
(EOB)
Detail in a block vs. DCT coefficients transmitted

Typical DCT coding artifacts

DCT coding with increasingly coarse quantization, block size 8x8
Adaptive transform coding

- Quantization and entropy coding optimized separately for each class.
- Typical classes:
  - Blocks without detail
  - Horizontal structures
  - Vertical structures
  - Diagonals
  - Textures without preferred orientation

Influence of DCT block size

- Efficiency as a function of blocksize $N \times N$, measured for 8 bit quantization in the original domain and equivalent quantization in the transform domain.

$$G_0 = \frac{\text{memoryless entropy of original signal}}{\text{mean entropy of transform coefficients}}$$

- Block size 8x8 is a good compromise.
Fast DCT algorithm I

- DCT matrix factored into sparse matrices (Arai, Agui, and Nakajima; 1988):

\[ y = Ax = SPM_1M_2M_3M_4M_5M_6x \]

\[
\begin{pmatrix}
S_0 & S_1 & S_2 & 0 \\
S_1 & S_2 & 0 & S_3 \\
0 & S_3 & S_4 & S_5 \\
S_5 & 0 & S_6 & S_7
\end{pmatrix}
\]

\[
P = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix}
\]

\[
M_1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ -1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}
\]

\[
M_2 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix}
\]

\[
M_3 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ -1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix}
\]

\[
M_4 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix}
\]

\[
M_5 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ -1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix}
\]

Fast DCT algorithm II

- Signal flow graph for fast (scaled) 8-DCT according to Arai, Agui, Nakajima:

Multiplication:
\[ u = m \Rightarrow m \cdot u \]
Addition:
\[ u \rightarrow u + v \]
\[ u \leftarrow u - v \]

\[ a_1 = C_4 \]
\[ a_2 = C_4 - C_6 \]
\[ a_3 = C_4 \]
\[ a_4 = C_4 + C_2 \]
\[ a_5 = C_6 \]
\[ s_0 = \frac{1}{2\sqrt{2}} \]
\[ s_k = \frac{1}{4C_k} \]
\[ C_k = \cos\left(\frac{\pi}{16}k\right) \]
Transform coding: summary

- Orthonormal transform: rotation of coordinate system in signal space
- Purpose of transform: decorrelation, energy concentration
- KLT is optimum, but signal dependent and, hence, without a fast algorithm
- DCT shows reduced blocking artifacts compared to DFT
- Bit allocation proportional to logarithm of variance
- Threshold coding + zig-zag-scan + 8x8 block size is widely used today (e.g. JPEG, MPEG, ITU-T H.263)
- Fast algorithm for scaled 8-DCT: 5 multiplications, 29 additions