Multimedia Communications

Transform Coding

McMaster University

Transform coding

- Transform coding: source output is transformed into components that are coded according to their characteristics
- If a sequence of inputs is transformed into another sequence in which most of the information is contained in only a few elements, we can encode and transmit those elements resulting in data compression
- Issues:
  - transform
  - quantization and encoding of the transformed coefficients

McMaster University

Transform coding

- Transform coding consists of three steps:

1. Data sequence x[n] is divided into blocks of size N and each block is mapped into a transform sequence y[n]
2. Quantization of transformed sequence which depends of three factors:
    1. desired average bit rate
    2. statistics of various elements of transformed sequence
    3. effects of distortion in the transform coefficients on the reconstructed sequence
3. Quantized values need to be encoded using some binary encoding technique (e.g., run length coding, Huffman coding)

Linear Transforms

- Linear transform: $$y[k] = \sum_{n=0}^{N-1} x[n] a_k[n], \quad k = 0, \ldots, N-1$$
  $$\underline{y} = \underline{A}\underline{x}$$
  $$A = [a_k[n]]$$
  $$\underline{x} = \underline{A}^{-1}\underline{y}$$
- Desired properties of a transform:
  - invertible
  - energy preserving $$\sum_{n=0}^{N-1} x^2[n] = \sum_{k=0}^{N-1} |y[k]|^2$$
  - decorrelating
  - energy compacting: information contained in only a few elements

McMaster University

2-D Transforms

$$y[k_1,k_2]=\sum_{n=0}^{N-1} x[n_1,n_2]a_{k_1,k_2}[n_1,n_2], \quad k_1,k_2=0,N-1$$

- Most useful 2-D transforms are separable: the 2-D transform of a 2-D signal can be obtained by applying the corresponding 1-D transform to the rows and columns.

$$a_{k_1,k_2}[n_1,n_2]=a_{k_1}[n_1]a_{k_2}[n_2]$$

$$\underline{y}=\underline{A}\underline{x}\underline{A}^{T}$$

$$x=A^{-1}y(A^{-1})^{T}$$

where $\underline{x}$ is the matrix holding the samples of the 2-D signal x.

McMaster University

Orthogonal transforms

- Orthogonal transforms: inverse of the transformation matrix is its transpose because the rows of the transform matrix form an orthogonal basis set
- Orthogonal transforms are energy preserving:

$$\sum_{n=1}^{N} y^2[n] = y^T y = (Ax)^T Ax = x^T A^T Ax = x^T x = \sum_{n=1}^{N} x^2[n]$$

Performance measures

- To choose among several transforms, we need some performance measures.
- Let's assume the autocorrelation matrix of the input is W and the transform is V.
- Decorrelation efficiency:

$$\eta_c = 1 - \frac{\sum_{j \neq k} |v_{jk}|}{\sum_{j \neq k} |w_{jk}|}$$

- Energy packing efficiency:

$$\eta_E = \frac{\sum_{j=1}^{M} |v_{jj}|}{\sum_{j=1}^{N} |v_{jj}|} = \frac{\sum_{j=1}^{M} |\sigma_j^2|}{\sum_{j=1}^{N} |\sigma_j^2|}$$

- Coding gain:

$$G = \frac{\frac{1}{N}\sum_{k=0}^{N-1} \sigma_k^{\ 2}}{\left[\prod_{k=0}^{N-1} \sigma_k^{\ 2}\right]^{1/N}} \qquad \sigma_k^{\ 2} = \mathrm{var}(y[k])$$

McMaster University

Karhunen-Loève Transform

- Karhunen-Loève transform (Hotelling transform or method of principle component analysis) decorelates the input sequence
- Sample-to-sample correlation of the transformed sequence is zero

$$\underline{A} = \left[\underline{\phi}_0, \underline{\phi}_1, \ldots, \underline{\phi}_{N-1}\right]^T$$

$$\text{where } \underline{\mathrm{R}}\underline{\phi}_k = \lambda_k \underline{\phi}_k$$

$\underline{\mathrm{R}}$ = the autocorrelation matrix of x

$\underline{\phi}_k$ = eigenvectors of R

McMaster University

Karhunen-Loève Transform

- Advantages:
  - the transform samples are completely uncorrelated
  - KLT maximizes transform coding gain

$$G = \frac{\frac{1}{N}\sum_{k=0}^{N-1}\sigma_k^{\ 2}}{\left[\prod_{k=0}^{N-1}\sigma_k^{\ 2}\right]^{1/N}} \qquad \sigma_k^{\ 2} = \mathrm{var}(y[k])$$

- Disadvantages:
  - data dependent: most be computed for each data set and sent to receiver (significant overhead)
  - computationally intensive

The Discrete Cosine Transform

$$y[k] = \alpha[k] \sum_{n=0}^{N-1} x[n] \cos\left[\frac{\pi(2n+1)k}{2N}\right] \quad 0 \le k \le N-1$$

where $\alpha[0] = \sqrt{\frac{1}{N}} \quad \alpha[k] = \sqrt{\frac{2}{N}} \quad 1 \le k \le N-1$

- It can be computed from the DFT.
- Data independent
- Fast transform exists, $O(N\log_2 N)$
- Approximates well the KLT in terms of energy compaction especially Markov sources with high correlation coefficient

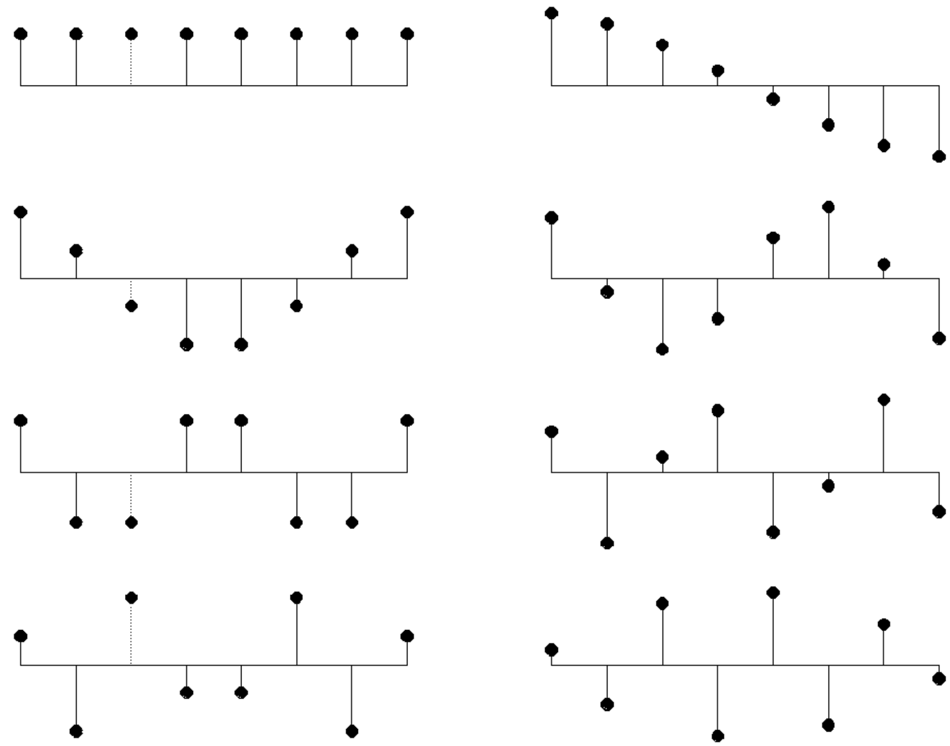
$$\rho = \frac{E[x_n x_{n+1}]}{E[x_n^2]}$$

- Used in JPEG, MPEG, MJPEG, etc.

McMaster University

The DCT Basis Vectors (N=8)

- The basis functions of the DCT are real sinusoids.
- Similar Fourier-type frequency-domain interpretation holds.



McMaster University

Discrete Walsh-Hadamard Transform

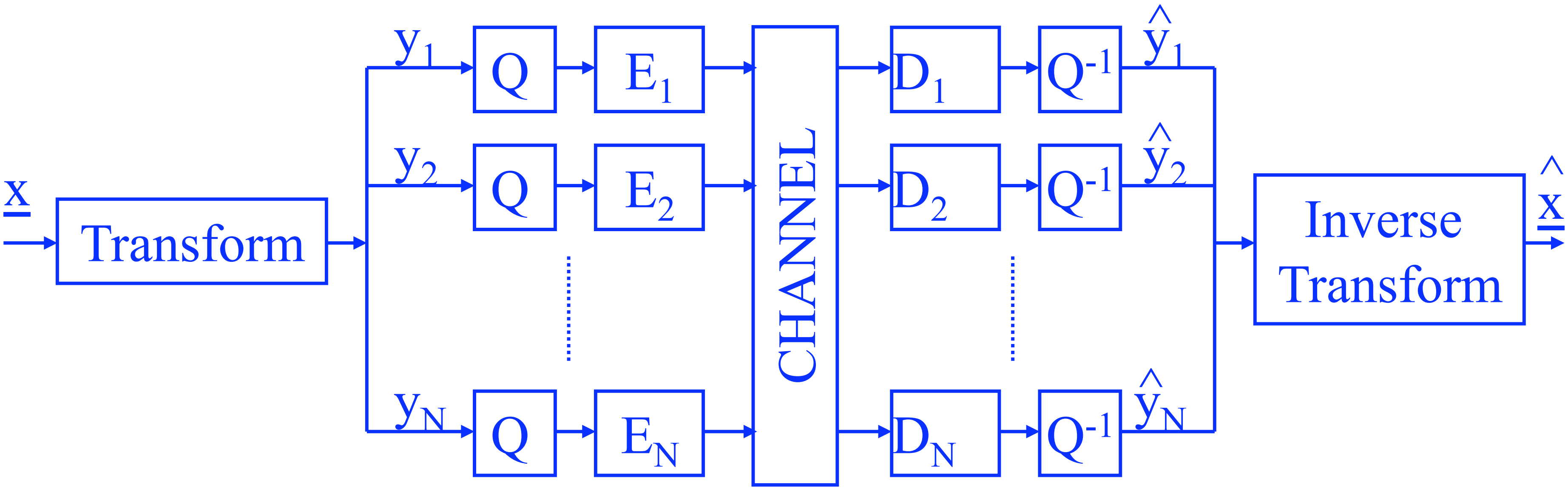
- Hadamard matrix of order N: $HH^T=NI$
- Hadamard matrices with dimensions of power of two can be constructed in the following manner:

$$H_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad H_{n+1} = \frac{1}{\sqrt{2}}\begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}$$

- The Walsh-Hadamard transform: order the rows of Hadamard matrix in increasing order of sequency (number of zero crossings).
- Advantage: very efficient multiplier-less implementation.
- Disadvantage: less energy compacting.

McMaster University

Transform Coding



Bit allocation

- Bit allocation: which transform coefficients are kept and what is the precision that is used to represent them.
- Retaining the coefficients:
  1. Maximum variance (zonal coding)
  2. Maximum magnitude (threshold coding)

Zonal Coding

- Uses the information theory concept of viewing information as uncertainty.
- Transform coefficients of maximum variance carry the most picture information and should be retained.
- How to calculate the variance:
  1. From the (N/n)(N/n) transformed subimage arrays
  2. An assumed image model
- Coefficients of maximum variance are located around the origin of the transform
- The retained coefficients must be quantized and coded.

Threshold coding

- Threshold coding is adaptive: the location of the transform coefficients retained for each subimage vary from one subimage to another one.
- For each subimage, the transform coefficients of largest magnitude make the most significant contribution to reconstructed subimage quality.
- What is the threshold and how it is obtained?
  1. A single global threshold for all subimages
  2. A single threshold for each subimage
  3. The threshold can be varied as a function of the location of each coefficient within the subimage.

JPEG

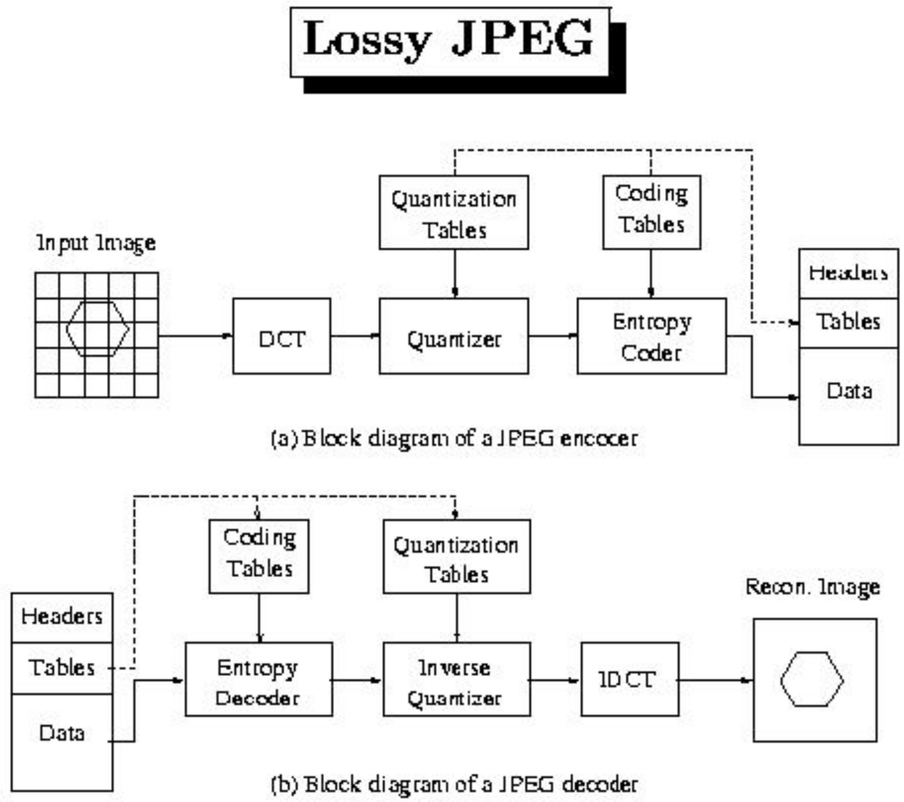
- JPEG: Joint Photographic Experts Group
- Scope: development of international standard for the compression and decompression and encoding of digital continuous tone still pictures.
- Benefit: To make use of continuous tone digital images more economical during both storage and transmission
- Applicable fields:
  - Databases, electronic mail and photoediting
  - Medical imaging and scientific imaging

McMaster University

JPEG

- Requirements:
  - Generic still image compression
  - Modest to low software/hardware complexity
  - Sequential, progressive and layered coding
  - Solution:
    - Differential and Huffman coding (lossless)
    - DCT, qunatization run-length and Huffman/ arithmetic coding (lossy)
- Features:
  - Psychovisual-based quantization
  - Sequential, progressive and hierarchical modes
  - Interleaving between color components

JPEG


Lossy JPEG
Input Image
DCT
Quantization Tables
Quantizer
Coding Tables
Entropy Coder
Headers
Tables
Data
(a) Block diagram of a JPEG encoder
Headers
Tables
Data
Coding Tables
Entropy Decoder
Quantization Tables
Inverse Quantizer
IDCT
Recon. Image
(b) Block diagram of a JPEG decoder

JPEG

- Input: 1-4 color components (8 bits/pixel)
- DCT: the most computationally demanding
- Quantization: uniform, midtread, quantization
- Quantizers step sizes are arranged in a table called Q-table

$$\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 81 & 104 & 113 & 92 \\ 24 & 35 & 55 & 64 & 103 & 121 & 120 & 101 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

- The Q-table can be changed by scaling the prototype with a quality factor.

JPEG

- DC coefficients: first coded differentially (difference between neighboring labels are coded)
- The number of values that the difference can get is large
- It is difficult to manage a Huffman code
- The possible values of the difference are partitioned into categories
- Size of these categories grow as power of two
- Category number is Huffman coded
- Elements within each category are specified by tacking on extra bits to the end of the Huffman code for that category

JPEG

Range	DC Difference Category	AC Category
0	0	N/A
−1, 1	1	1
−3, −2, 2, 3	2	2
−7, ..., −4, 4, ..., 7	3	3
−15, ..., −8, 8, ..., 15	4	4
−31, ..., −16, 16, ..., 31	5	5
−63, ..., −32, 32, ..., 63	6	6
−127, ..., −64, 64, ..., 127	7	7
−255, ..., −128, 128, ..., 255	8	8
−511, ..., −256, 256, ..., 511	9	9
−1023, ..., −512, 512, ..., 1023	A	A
−2047, ..., −1024, 1024, ..., 2047	B	B
−4095, ..., −2048, 2048, ..., 4095	C	C
−8191, ..., −4096, 4096, ..., 8191	D	D
−16383, ..., −8192, 8192, ..., 16383	E	E
−32767, ..., −16384, 16384, ..., 32767	F	N/A

JPEG

Category	Base Code	Length	Category	Base Code	Length
0	010	3	6	1110	10
1	011	4	7	11110	12
2	100	5	8	111110	14
3	00	5	9	1111110	16
4	101	7	A	11111110	18
5	110	8	B	111111110	20

JPEG

Run/ Category	Base Code	Length	Run/ Category	Base Code	Length
0/0	**1010 (= EOB)**	**4**			
0/1	00	3	8/1	11111010	9
0/2	01	4	8/2	111111111000000	17
0/3	100	6	8/3	1111111110110111	19
0/4	1011	8	8/4	1111111110111000	20
0/5	11010	10	8/5	1111111110111001	21
0/6	111000	12	8/6	1111111110111010	22
0/7	1111000	14	8/7	1111111110111011	23
0/8	1111110110	18	8/8	1111111110111100	24
0/9	1111111110000010	25	8/9	1111111110111101	25
0/A	1111111110000011	26	8/A	1111111110111110	26
1/1	1100	5	9/1	111111000	10
1/2	111001	8	9/2	1111111110111111	18
1/3	1111001	10	9/3	1111111111000000	19
1/4	111110110	13	9/4	1111111111000001	20
1/5	11111110110	16	9/5	1111111111000010	21
1/6	1111111110000100	22	9/6	1111111111000011	22
1/7	1111111110000101	23	9/7	1111111111000100	23
1/8	1111111110000110	24	9/8	1111111111000101	24
1/9	1111111110000111	25	9/9	1111111111000110	25
1/A	1111111110001000	26	9/A	1111111111000111	26
2/1	11011	6	A/1	111111001	10
2/2	11111000	10	A/2	1111111111001000	18
2/3	1111110111	13	A/3	1111111111001001	19
2/4	1111111110001001	20	A/4	1111111111001010	20
2/5	1111111110001010	21	A/5	1111111111001011	21
2/6	1111111110001011	22	A/6	1111111111001100	22
2/7	1111111110001100	23	A/7	1111111111001101	23

JPEG

- As the categories are of different size, we need differing number of bits to identify the value in each category
- For AC coefficients the category number (C) and the number of zero-valued labels Z since the last non-zero label form a pointer to the Huffman code
- Extra bits are added to the Huffman code to determine the value
- If a particular coefficient is the last nonzero value along the zigzag scan, the code for it is followed by EOB.

McMaster University