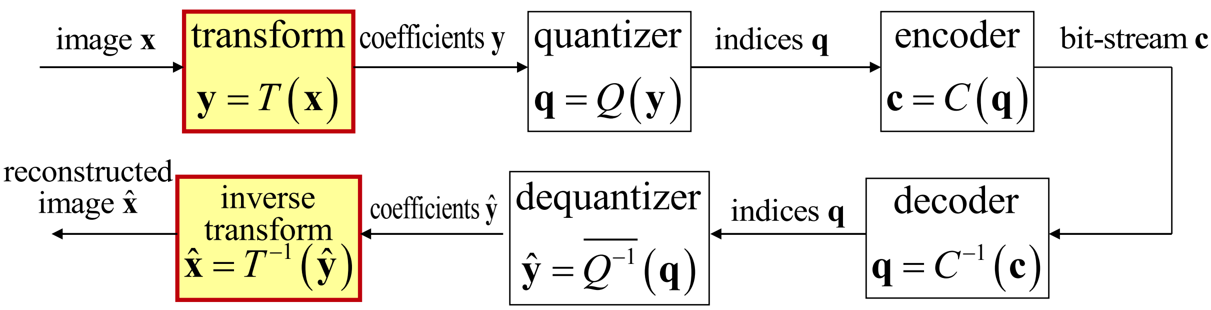
Transform coding - topics

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

Typical structured codec



- Transform $T(\boldsymbol{x})$ usually invertible
- Quantization $Q(\mathbf{y})$ not invertible, introduces distortion
- Combination of encoder $C(\mathbf{q})$ and decoder $C^{-1}(\mathbf{c})$ lossless

Properties of orthonormal transforms

- Forward transform

$$\boxed{\mathbf{y} = \mathbf{A}\mathbf{x}}$$

NxN transform coefficients, arranged as a column vector

Transform matrix of size N^2xN^2

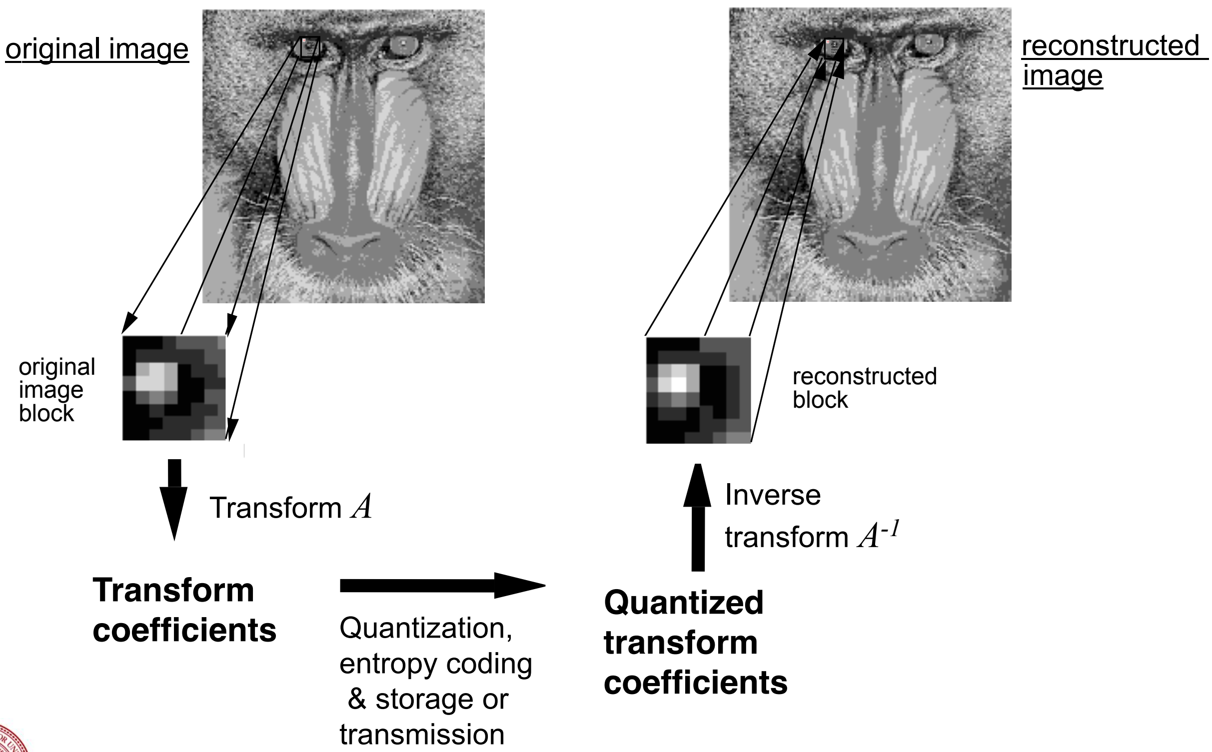
Image block of size *NxN*, arranged as a column vector

- Inverse transform

$$\boxed{\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} = \mathbf{A}^{T}\mathbf{y}}$$

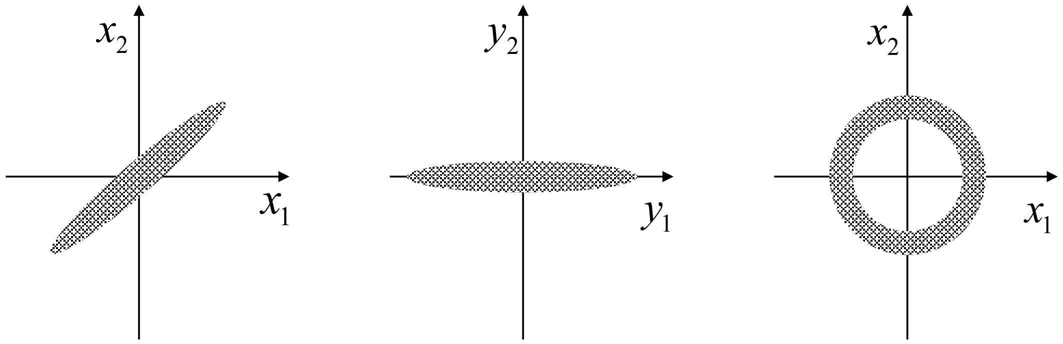
- Linearity: $\mathbf{x}$ is represented as linear combination of "basis functions" (i.e., columns of $\mathbf{A}^T$)

Block-wise transform coding



2-d orthonormal transform

$$\mathbf{A} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$



Strongly correlated samples, equal energies	After transform: uncorrelated samples, most of the energy in first coefficient	Despite statistical dependence, orthonormal transform won't help.

Energy conservation

- For any orthonormal transform $\mathbf{y} = \mathbf{A}\mathbf{x}$

$$\boxed{\|\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = \|\mathbf{x}\|^2}$$

- Interpretation
  - Vector length („energies") conserved
  - Orthonormal transform is a rotation of the coordinate system around the origin (plus possible sign flips)

Coding gain of orthonormal transform

- Assume distortion rate functions for image samples

$$\boxed{d(R) \cong \varepsilon^2 \sigma_X^2 2^{-2R}}$$

. . . and for encoding transform coefficients

$$\boxed{d^{XFORM}(R) = \frac{1}{N}\sum_{n=0}^{N-1} d_n(R_n) \cong \frac{1}{N}\sum_{n=0}^{N-1} \varepsilon^2 \sigma_{Y_n}^2 2^{-2R_n}; \qquad R = \frac{1}{N}\sum_{n=0}^{N-1} R_n}$$

- Transform coding gain

$$\boxed{G_T = \frac{d(R)}{d^{XFORM}(R)}}$$

Unequal variances of transform coefficients

- Total energy conserved, but unevenly distributed among coefficients.
- Covariance matrix

$$\mathbf{R}_{\mathbf{yy}} = E\left[(\mathbf{y}-\mu_{\mathbf{Y}})(\mathbf{y}-\mu_{\mathbf{Y}})^{\mathbf{T}}\right]$$
$$= E\left[\mathbf{A}(\mathbf{x}-\mu_{\mathbf{X}})(\mathbf{x}-\mu_{\mathbf{X}})^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\right] = \mathbf{A}\mathbf{R}_{\mathbf{xx}}\mathbf{A}^{\mathbf{T}}$$

- Variances of the coefficients y_i are diagonal elements of $\mathbf{R}_{\mathbf{yy}}$

$$\sigma_{Y_i}^2 = \left[\mathbf{R}_{\mathbf{yy}}\right]_{i,i} = \left[\mathbf{A}\mathbf{R}_{\mathbf{xx}}\mathbf{A}^{\mathbf{T}}\right]_{i,i}$$

Coding gain of orthonormal transform (cont.)

- Optimum distortion and rate per coefficient

$$\boxed{d_n(R_n) = d^{XFORM}(R) \quad \text{for all } n} \qquad \boxed{R_n = \frac{1}{2}\log_2 \frac{\varepsilon^2 \sigma_{Y_n}^2}{d^{XFORM}} \quad \text{for all } n}$$

- Transform coding gain

$$\boxed{G_T = \frac{d(R)}{d^{XFORM}(R)} = \frac{\sigma_X^2}{\sqrt[N]{\prod_{n=0}^{N-1}\sigma_{Y_n}^2}} = \frac{\frac{1}{N}\sum_{n=0}^{N-1}\sigma_{Y_n}^2}{\sqrt[N]{\prod_{n=0}^{N-1}\sigma_{Y_n}^2}}}$$

Coding gain of orthonormal transform (cont.)

- Find optimum bit allocation using Lagrangian formulation

$$\boxed{J = d^{XFORM}(R) + \lambda R = \frac{1}{N}\sum_{n=0}^{N-1}\varepsilon^2\sigma_{Y_n}^2 2^{-2R_n} + \lambda\frac{1}{N}\sum_{n=0}^{N-1}R_n \xrightarrow{R_0, R_1, \ldots, R_{N-1}} \min.}$$

- Solution by setting $\frac{\partial J}{\partial R_n} = 0 \quad \text{for all } n$

Distortion of individual coefficient

$$\boxed{\frac{\partial d_i}{\partial R_i} = \frac{\partial d_j}{\partial R_j} \quad \text{for all } i, j}$$

"Pareto condition"

Vilfredo Pareto
Economist
1848-1923

Karhunen Loève Transform (KLT)

- Karhunen Loève Transform (KLT): basis functions are eigenvectors of the covariance matrix R_{XX} of the input signal.
- KLT yields decorrelated transform coefficients (covariance matrix R_{YY} is diagonal).
- KLT achieves optimum energy concentration.
- KLT maximizes coding gain G_T

"Reverse water filling"

- With additional constraints $R_n \geq 0$ for all n and $\varepsilon = 1$ use Karush-Kuhn-Tucker conditions

$$\boxed{\frac{\partial J}{\partial R_n} \begin{cases} = 0, & \text{if } d_n < \sigma_{Y_n}^2 \\ \geq 0, & \text{if } d_n = \sigma_{Y_n}^2 \end{cases}}$$

- Optimum distortion and rate allocation

$$\boxed{d_n(R_n) = \begin{cases} \theta, & \text{if } \sigma_{Y_n}^2 > \theta \\ \sigma_{Y_n}^2, & \text{if } \sigma_{Y_n}^2 \leq \theta \end{cases}} \qquad \boxed{R_n = \frac{1}{2}\log_2 \frac{\sigma_{Y_n}^2}{d_n} \quad \text{for all } n}$$

where θ is chosen to yield $\sum_n d_n(R_n) = d^{XFORM}$

Disadvantages of KLT

- KLT dependent on signal statistics
- KLT not separable for image blocks
- Transform matrix cannot be factored into sparse matrices

$\boxed{\Rightarrow \text{ Find structured transforms that perform close to KLT}}$

KLT maximizes coding gain

- Determinant of any orthonormal transform $\det(\mathbf{A}) = \pm 1$
- Determinant of covariance matrix for any orthonormal transform

$$\det(\mathbf{R_{YY}}) = \det(\mathbf{A})\det(\mathbf{R_{XX}})\det(\mathbf{A^T}) = \det(\mathbf{R_{XX}})$$

- Determinant of (diagonal) covariance matrix after KLT

$$\det(\mathbf{R_{YY}}) = \prod_{n=0}^{N-1} \sigma_{Y_n}^2$$

- Hadamard inequality: determinant of any symmetric, positive semi-definite matrix is less than or equal to the product of its diagonal elements

$$\boxed{\prod_{n=0}^{N-1} \sigma_{Y_n}^2 (\mathbf{KLT}) = \det(\mathbf{R_{YY}}) \le \prod_{n=0}^{N-1} \sigma_{Y_n}^2 (\mathbf{A})}$$

Separable transforms, I

- A transform is separable, if the transform of a signal block of size NxN can be expressed by

$$\boxed{y = AxA^T}$$

NxN transform coefficients

Orthonormal transform matrix of size NxN

NxN block of input signal

- The inverse transform is

$$\boxed{x = A^T yA}$$

Note: $\mathbf{A} = A \otimes A$

Transform matrix for vectors $\mathbf{y} = \mathbf{A}\mathbf{x}$

Kronecker product

- Great practical importance: The transform requires 2 matrix multiplications of size NxN instead one multiplication of a vector of size 1xN^2 with a matrix of size N^2xN^2

$\longrightarrow$ Reduction of the complexity from $O(N^4)$ to $O(N^3)$

Various orthonormal transforms

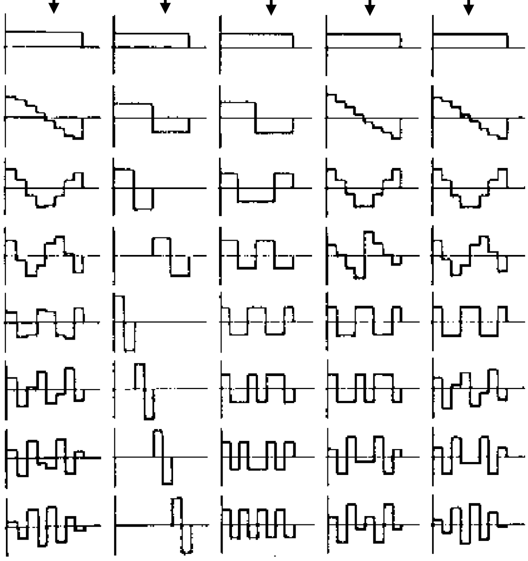
Karhunen Loève transform *[1948/1960]*

Haar transform *[1910]*

Walsh-Hadamard transform *[1923]*

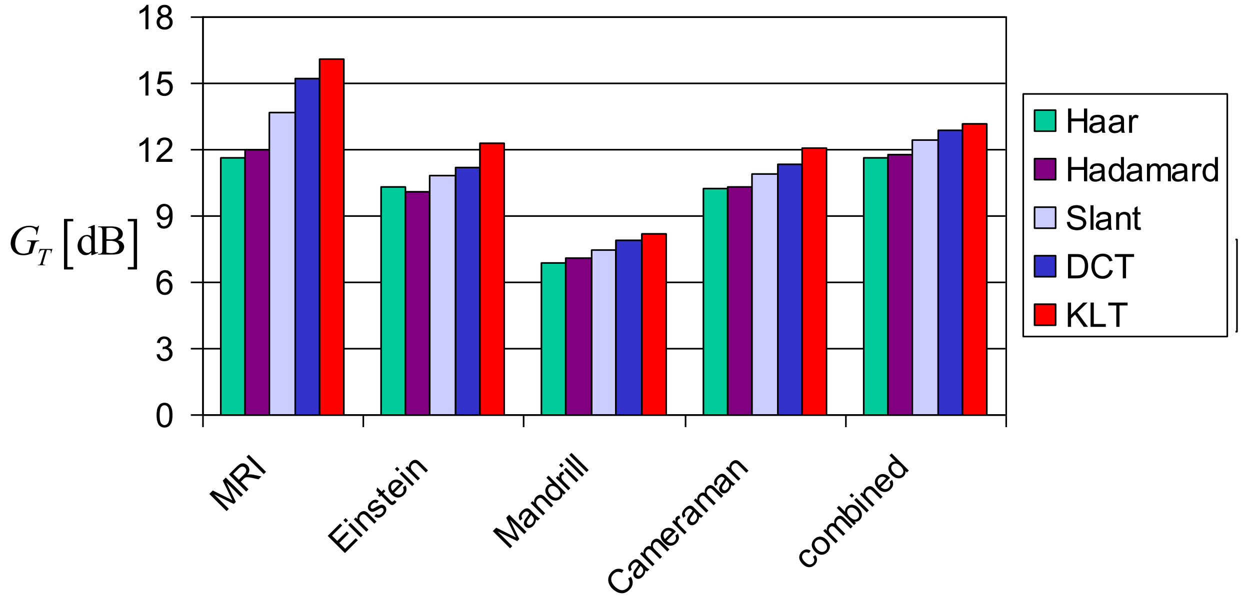
Slant transform *[Enomoto, Shibata, 1971]*

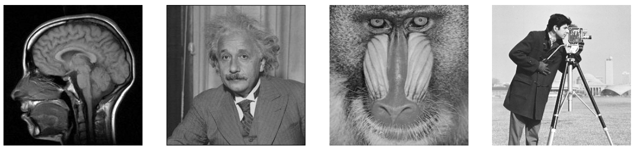
Discrete CosineTransform (DCT) *[Ahmet, Natarajan, Rao, 1974]*



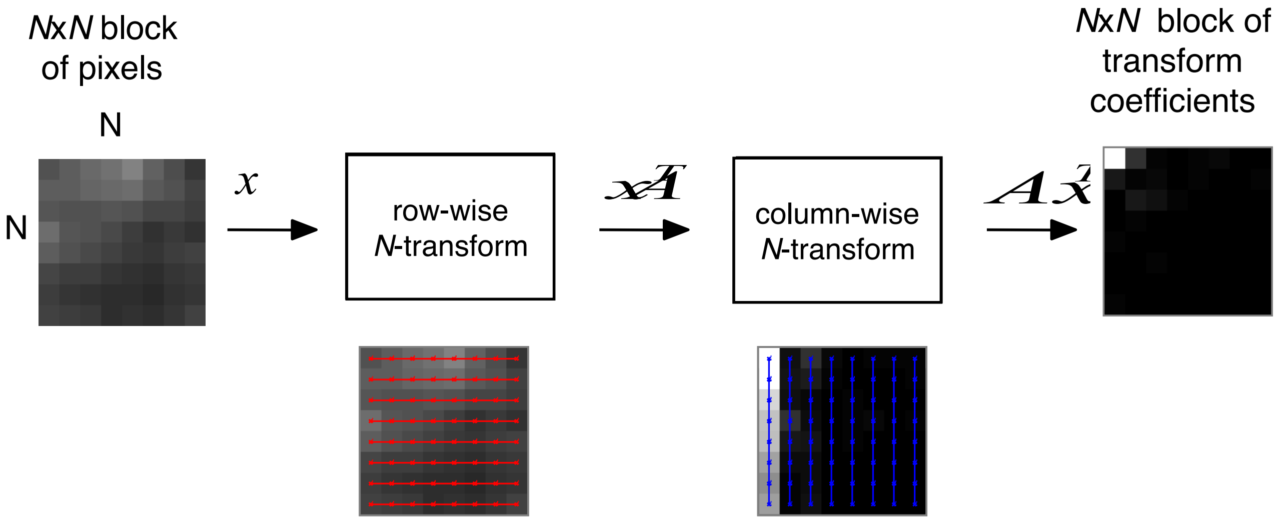
Comparison of 1-d basis functions for block size $N=8$

Coding gain with 8x8 transforms





Separable transforms, II



DCT

- Type II-DCT of blocksize NxN is defined by transform matrix A containing elements

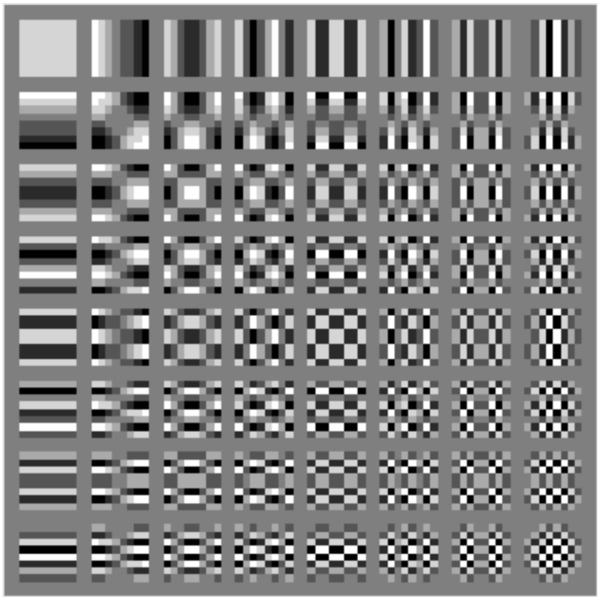
$$a_{ik} = \alpha_i \cos \frac{\pi(2k+1)i}{2N}$$

$$\text{for } i,k = 0,\ldots,N-1$$

$$\text{with } \alpha_0 = \sqrt{\frac{1}{N}}$$

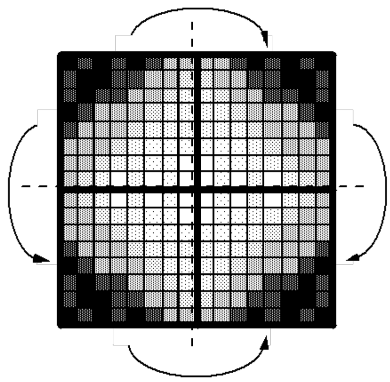
$$\alpha_i = \sqrt{\frac{2}{N}} \quad \forall i \neq 0$$

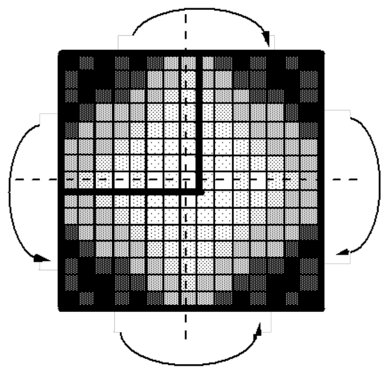
- 2D DCT basis functions:



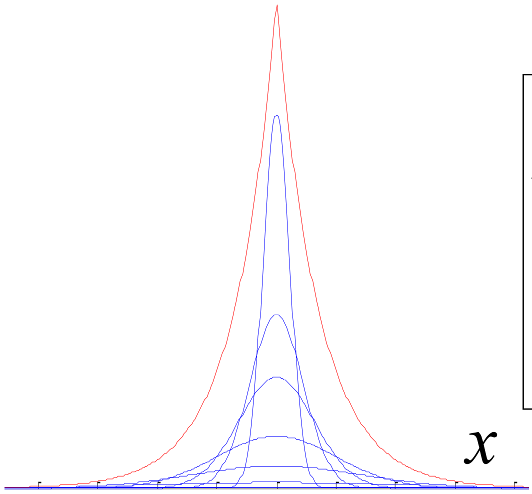
Discrete Cosine Transform and Discrete Fourier Transform

- 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“





Infinite Gaussian mixture modeling



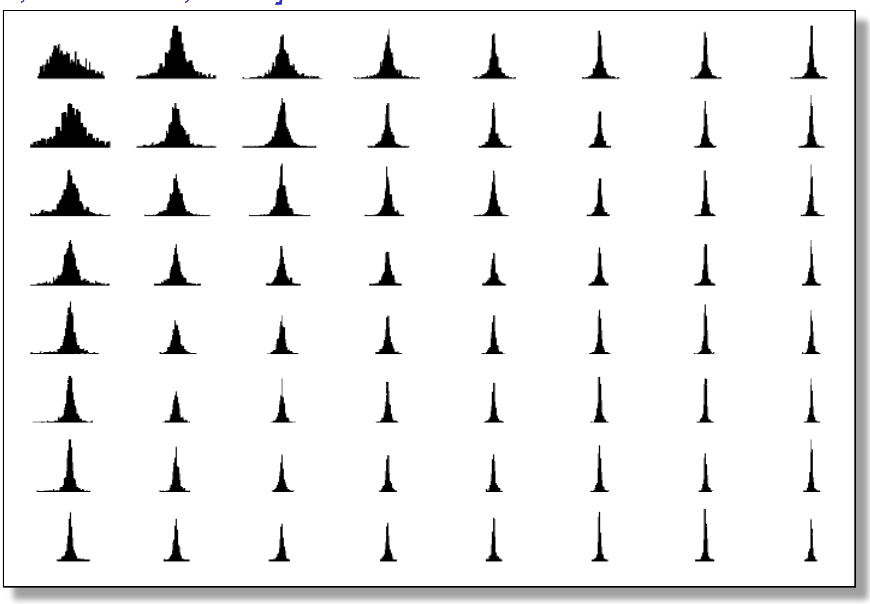
$$p_{Y_n}(y) = \int_0^\infty \frac{1}{\sqrt{2\pi v}} \cdot e^{-y^2/2v} \, \frac{1}{\sigma^2} e^{-v/\sigma_{y_n}^2} dv$$

$$= \sqrt{\frac{1}{2\sigma_{y_n}^2}} \cdot e^{-\sqrt{2} \cdot |y| / \sigma_{y_n}}$$

- For a given block variance, coefficient pdfs are Gaussian
- Gaussian mixture w/ exponential variance distribution yields a Laplacian
- Gaussian mixture w/ half-Gaussian variance distribution yields pdf very close to Laplacian *[Lam, Goodman, 2000]*
- Elegant explanation of Laplacian pdfs of DCT coefficients

Amplitude distribution of the DCT coefficients

- Histograms for 8x8 DCT coefficient amplitudes measured for test image *[Lam, Goodman, 2000]*

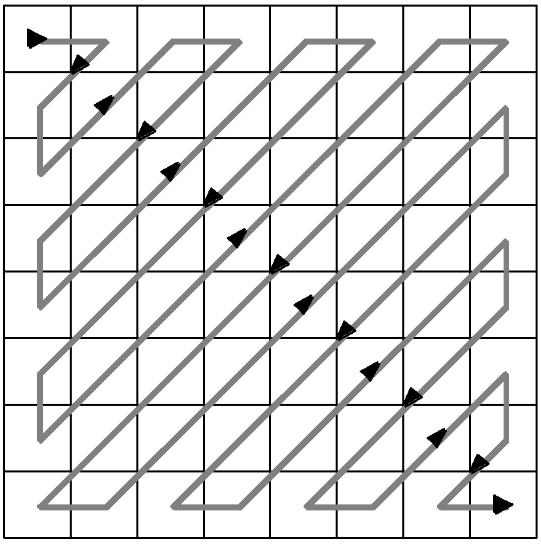


Test image
Bridge

- AC coefficients: Laplacian PDF
- DC coefficient distribution similar to the original image

Threshold coding, II

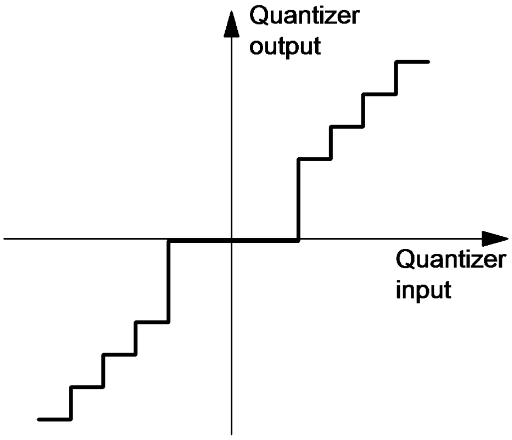
- Efficient encoding of the position of non-zero transform coefficients: zig-zag-scan + run-level-coding



ordering of the transform coefficients by zig-zag-scan

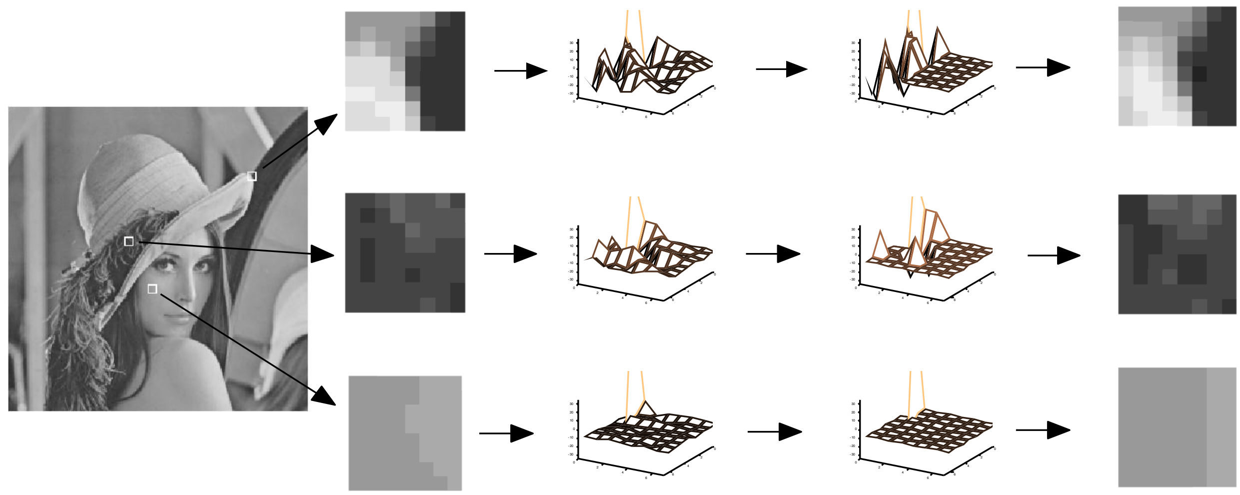
Threshold coding, I

- Uniform deadzone quantizer: transform coefficients that fall below a threshold are discarded.



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

Detail in a block vs. DCT coefficients



Threshold coding, III

198	202	194	179	180	184	196	168
187	196	192	181	182	185	189	174
188	185	193	179	188	188	187	170
184	188	182	187	183	186	195	174
194	193	189	187	180	183	181	185
193	195	193	192	170	189	187	181
181	185	183	180	175	184	185	176
195	185	177	178	170	179	195	175

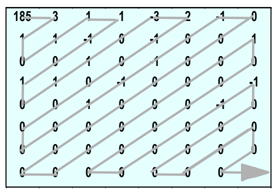
Original 8x8 block

DCT

1480	26.0	9.5	8.9	-26.4	15.1	-8.1	0.3
11.0	8.3	-8.2	3.8	-8.4	-6.0	-2.8	10.6
-5.5	4.5	9.0	5.3	-8.0	4.0	-5.1	4.9
10.7	9.8	4.9	-8.3	-2.1	-1.9	2.8	-8.1
1.6	1.4	8.2	4.3	3.4	4.1	-7.9	1.0
-4.5	-5.0	-6.4	4.1	-4.4	1.8	-3.2	2.1
5.9	5.8	2.4	2.8	-2.0	5.9	3.2	1.1
-3.0	2.5	-1.0	0.7	4.1	-6.1	6.0	5.7

Transformed 8x8 block

Q



Zig-zag scan

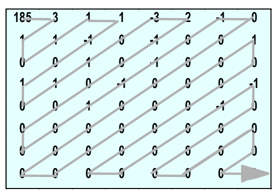
Run-level coding

Mean of Block: 185
(0,3) (0,1) (1,1) (0,1) (0,1) (0,1) (0,-1) (1,1)
(1,1) (0,1) (1,-3) (0,2) (0,-1) (6,1) (0,-1) (0,-1)
(1,-1) (14,1) (9,-1) (0,-1) EOB

Transmission

Mean of Block: 185
(0,3) (0,1) (1,1) (0,1) (0,1) (0,1) (0,-1) (1,1)
(1,1) (0,1) (1,-3) (0,2) (0,-1) (6,1) (0,-1) (0,-1)
(1,-1) (14,1) (9,-1) (0,-1) EOB

Run-level decoding



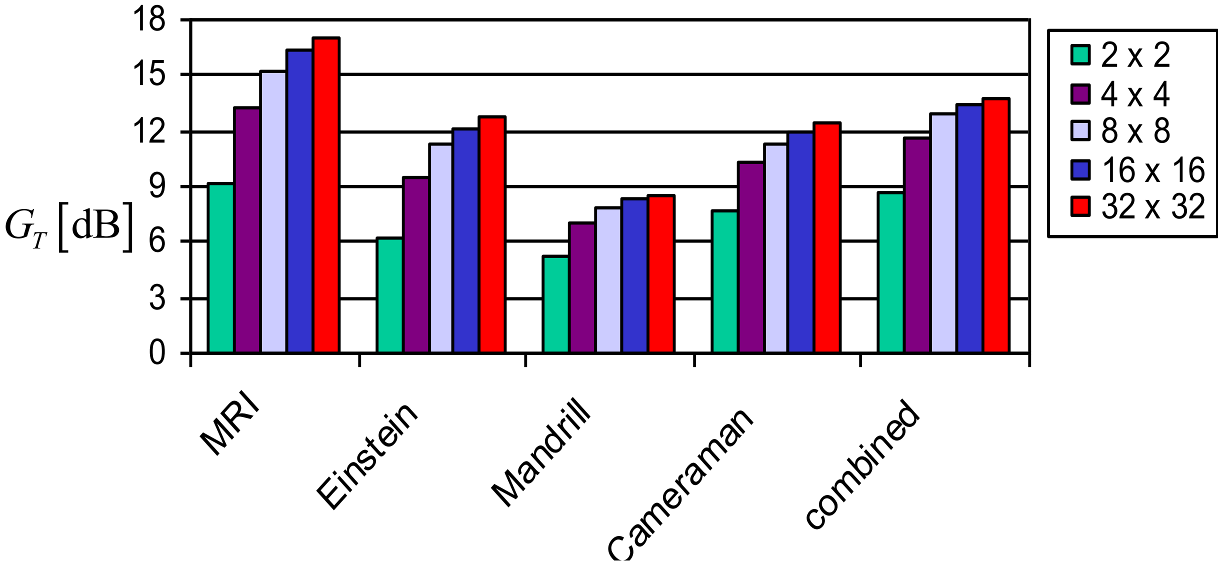
Inverse zig-zag scan

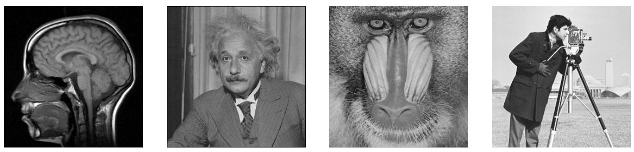
Scaling and inverse DCT

Reconstructed 8x8 block

192	201	195	184	177	184	193	174
189	191	195	182	182	187	190	171
188	185	190	181	185	187	189	171
189	188	185	183	183	182	190	175
191	192	186	189	179	182	188	178
190	191	189	190	177	186	184	179
189	188	185	184	175	186	187	179
189	188	178	176	173	183	193	180

Influence of DCT block size





Typical DCT coding artifacts

DCT coding with increasingly coarse quantization, block size 8x8

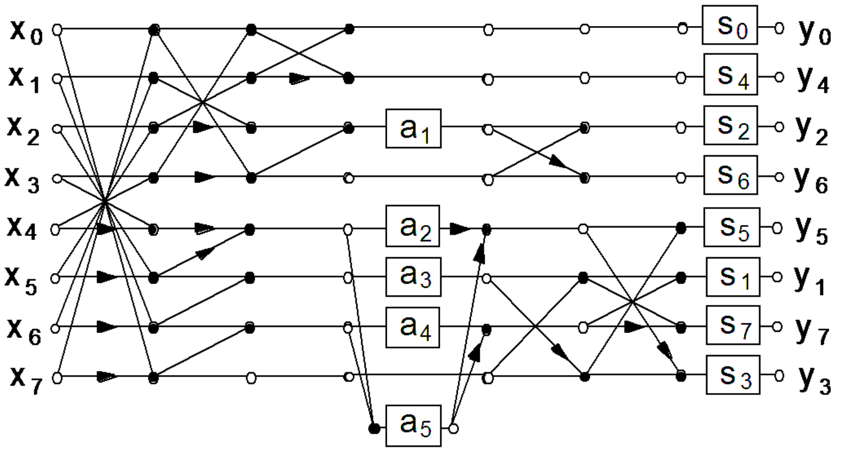
quantizer stepsize for AC coefficients: 25

quantizer stepsize for AC coefficients: 100

quantizer stepsize for AC coefficients: 200

Fast DCT algorithm II

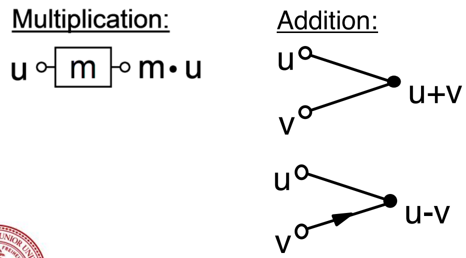
- Signal flow graph for fast (scaled) 8-DCT *[Arai, Agui, Nakajima, 1988]*



scaling

only 5 + 8 multiplications

(direct matrix multiplication: 64 multiplications)



$$a_1 = C_4$$
$$a_2 = C_2 - C_6$$
$$a_3 = C_4$$
$$a_4 = C_6 + C_2$$
$$a_5 = C_6$$

$$s_0 = \frac{1}{2\sqrt{2}}$$

$$s_k = \frac{1}{4C_k} \qquad k = 1,\ldots,7$$

$$C_k = \cos\left(\frac{\pi}{16}k\right)$$

Fast DCT algorithm I

- DCT matrix factored into sparse matrices
  [Arai, Agui, and Nakajima; 1988]

$$\begin{aligned} y &= Ax \\ &= SPM_1M_2M_3M_4M_5M_6x \end{aligned}$$

$$S = \begin{pmatrix} S_0 & & & & & & & \\ & S_1 & & & & & 0 & \\ & & S_2 & & & & & \\ & & & S_3 & & & & \\ & & & & S_4 & & & \\ & & & & & S_5 & & \\ & 0 & & & & & S_6 & \\ & & & & & & & S_7 \end{pmatrix} \qquad P = \begin{pmatrix} 1 & & & & & & & \\ & & & & & 1 & & \\ & & 1 & & & & & \\ & & & & & & & 1 \\ & 1 & & & & & & \\ & & & & 1 & & & \\ & & & 1 & & & & \\ & & & & & & 1 & \end{pmatrix}$$

$$M_1 = \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & 0 & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ & & & & 1 & & & 1 \\ & & & & & 1 & 1 & \\ & 0 & & & & 1 & -1 & \\ & & & & -1 & & & 1 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & 0 & \\ & & 1 & 1 & & & & \\ & & -1 & 1 & & & & \\ & & & & 1 & & & \\ & & & & & 1 & & 1 \\ & 0 & & & & & 1 & \\ & & & & & -1 & & 1 \end{pmatrix}$$

$$M_3 = \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & 0 & \\ & & C_4 & & & & & \\ & & & 1 & & & & \\ & & & & -C_2 & & & \\ & & & & & C_4 & -C_6 & \\ & 0 & & & & -C_6 & C_2 & \\ & & & & & & & 1 \end{pmatrix} \qquad M_4 = \begin{pmatrix} 1 & 1 & & & & & & \\ 1 & -1 & & & & & 0 & \\ & & 1 & 1 & & & & \\ & & & 1 & & & & \\ & & & & 1 & & & \\ & & & & & 1 & & \\ & 0 & & & & & 1 & \\ & & & & & & & 1 \end{pmatrix}$$

$$M_5 = \begin{pmatrix} 1 & & & 1 & & & & \\ & 1 & 1 & & & & 0 & \\ & 1 & -1 & & & & & \\ 1 & & & -1 & & & & \\ & & & & -1 & -1 & & \\ & & & & & 1 & 1 & \\ & 0 & & & & & 1 & 1 \\ & & & & & & & 1 \end{pmatrix} \qquad M_6 = \begin{pmatrix} 1 & & & & 0 & & & 1 \\ & 1 & & & & & 1 & \\ & & 1 & & & 1 & & \\ 0 & & & 1 & 1 & & & \\ & & & 1 & -1 & & & 0 \\ & & 1 & & & -1 & & \\ & 1 & & & & & -1 & \\ 1 & & & & 0 & & & -1 \end{pmatrix}$$

Reading

- Wiegand, Schwarz, Chapter 7
- Marcellin, Taubman, sections 4.1, 4.3
- V. K. Goyal, “Theoretical foundations of transform coding,” IEEE Signal Processing Magazine, vol. 18, no. 5, pp. 9-21, Sept. 2001
- W.-H. Chen, W. Pratt, “Scene Adaptive Coder,” IEEE Transactions on Communications, vol. 32, no. 3, pp. 225-232, March 1984.
- E. Y. Lam, J. W. Goodman, “A Mathematical Analysis of the DCT Coefficient Distributions for Images,” IEEE Transactions on Image Processing, vol. 9, no. 10, pp. 1661-1666, October 2000.

Transform coding: summary

- Orthonormal transform: rotation of coordinate system in signal space
- Purpose of transform: decorrelation, energy concentration
- Bit allocation proportional to logarithm of variance, equal distortion
- KLT is optimum, but signal dependent and, hence, without a fast algorithm
- DCT shows reduced blocking artifacts compared to DFT
- 8x8 block size, uniform quantization, zig-zag-scan + run-level coding is widely used today (e.g. JPEG, MPEG, ITU-T H.261, H.263)
- Fast algorithm for scaled 8-DCT: 5 multiplications, 29 additions