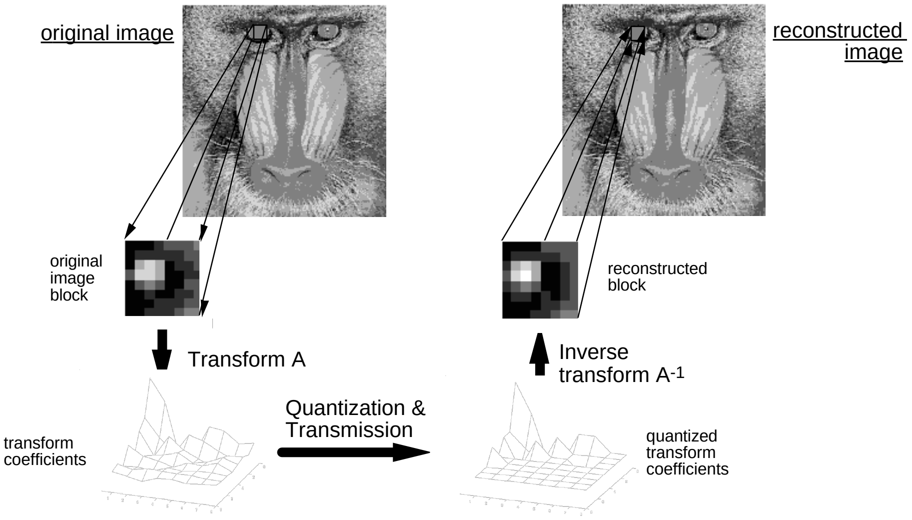
Transform coding - topics

- 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



original image

reconstructed image

original image block

reconstructed block

Transform A

Inverse transform A^{-1}

Quantization & Transmission

transform coefficients

quantized transform coefficients

Properties of orthonormal transforms

- Forward transform

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

NxN transform coefficients, arranged as a vector

Transform matrix of size N^2xN^2

Input signal block of size NxN, arranged as a 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“.
- 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 NxN can be expressed by

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

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

NxN transform coefficients

Orthonormal transform matrix of size NxN

NxN block of input signal

Kronecker product

- The inverse transform is

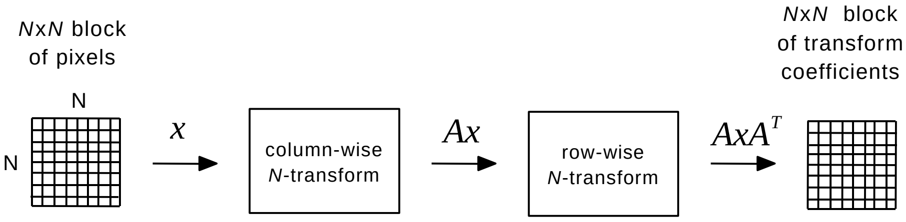
$$\boxed{x = A^T y A}$$

- 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

$\rightarrow$ Reduction of the complexity from $O(N^4)$ 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)



*N*x*N* block of pixels

N

N

x

column-wise *N*-transform

Ax

row-wise *N*-transform

AxA^T

*N*x*N* block of transform coefficients

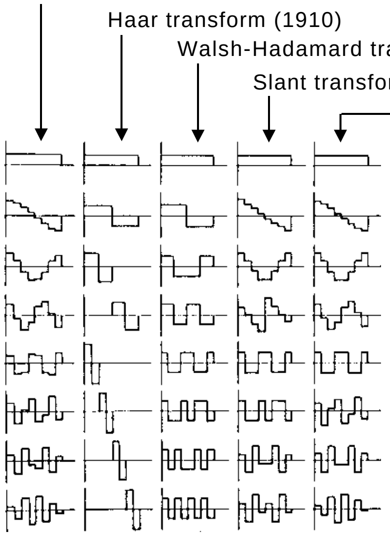
Criteria for the selection of a particular transform

- 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



Karhunen Loève transform (1948/1960)

Haar transform (1910)

Walsh-Hadamard transform(1923)

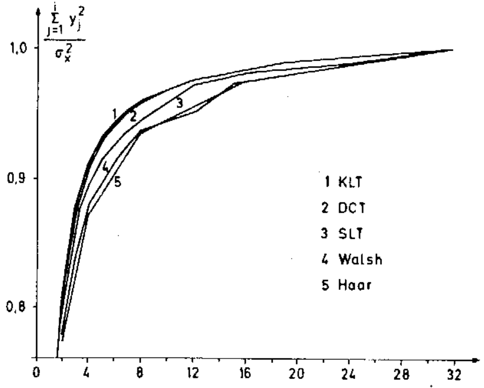
Slant transform (Enomoto, Shibata, 1971)

Discrete CosineTransform (DCT)
(Ahmet, Natarajan, Rao, 1974)

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

Comparison of various transforms, II

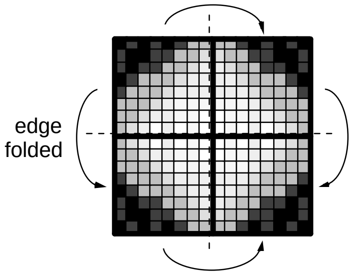
- Energy concentration measured for typical natural images, block size 1x32 *[Lohscheller]:*

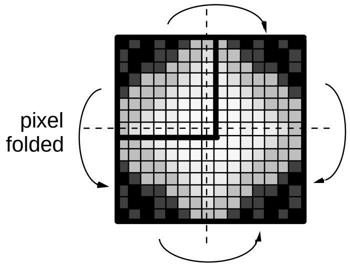


- KLT is optimum
- DCT performs only slightly worse than KLT

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 blocksize MxM is defined by transform matrix A containing elements

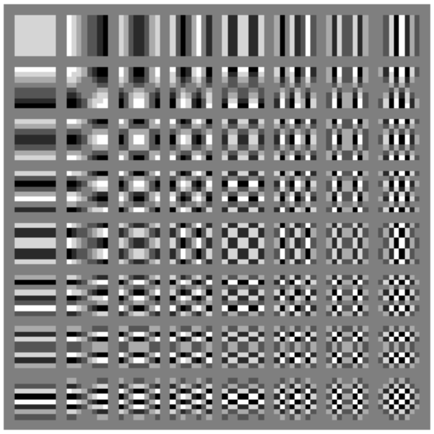
$$a_{ik} = a_i \cos\frac{p(2k+1)i}{2M}$$

for $i,k = 0,\dots,M-1$

with $a_0 = \sqrt{\frac{1}{M}}$

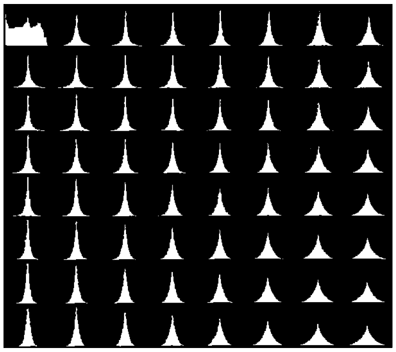
$$a_i = \sqrt{\frac{2}{M}} \quad \forall i \neq 0$$

- 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

<u>Problem:</u> divide bit-rate R among MxM transform coefficients i such that resulting distortion D is minimized.

<u>Assumptions</u>

$$\boxed{R = \sum_i R_i} \qquad \boxed{D = \sum_i D_i}$$

Total rate

Rate for coefficient *i*

Total distortion

Distortion contributed by coefficient *i*

lead to "Pareto condition"

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

Bit allocation for transform coefficients II

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

variance of transform coefficient *i*

$$R_i = \max\left\{0, \frac{1}{2}\log_2\left(\frac{s_i^2}{q}\right)\right\} \text{ bit}$$

$$D_i = \min\left\{s_i^2, q\right\}$$

Threshold parameter, same for all *i*

- 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

$$D = \sum_i w_i D_i$$

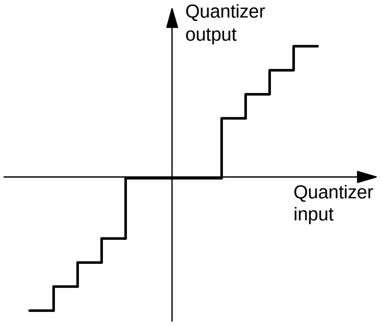
$$R_i = \max\left\{0, \frac{1}{2}\log_2\left(\frac{w_i s_i^2}{q}\right)\right\} \text{ bit}$$

$$D_i = \min\left\{s_i^2, \frac{q}{w_i}\right\}$$

- Often implemented by scaling coefficients by $(w_i)^{-1/2}$ prior to quantization (“weighting matrix”)

Threshold coding, I

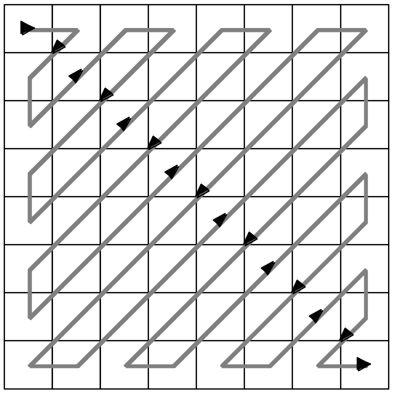
- Transform coefficients that fall below a threshold are discarded.
- Implementation by uniform quantizer with threshold characteristic:



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

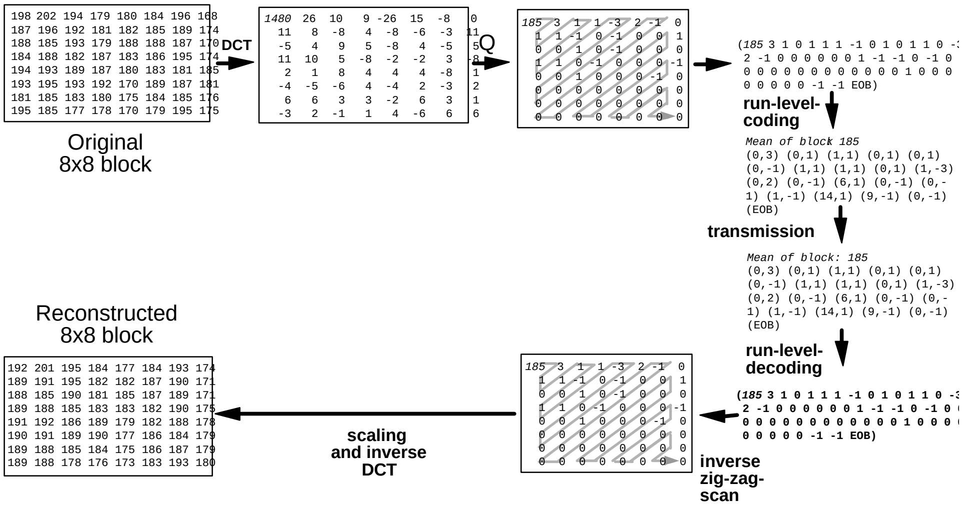
Threshold coding, II

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



ordering of the transform coefficients by zig-zag-scan

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	10	9	-26	15	-8	0
11	8	-8	4	-8	-6	-3	11
-5	4	9	5	-8	4	-5	5
11	10	5	-8	-2	-2	3	-8
2	1	8	4	4	4	-8	1
-4	-5	-6	4	-4	2	-3	2
6	6	3	3	-2	6	3	1
-3	2	-1	1	4	-6	6	6

Q

185	3	1	1	-3	2	-1	0
1	1	-1	0	-1	0	0	1
0	0	1	0	-1	0	0	0
1	1	0	-1	0	0	0	-1
0	0	1	0	0	0	-1	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(*185* 3 1 0 1 1 1 -1 0 1 0 1 1 0 -
2 -1 0 0 0 0 0 0 1 -1 -1 0 -1 0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 -1 -1 EOB)

run-level-coding

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) (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) (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

(*185* 3 1 0 1 1 1 -1 0 1 0 1 1 0 -
2 -1 0 0 0 0 0 0 1 -1 -1 0 -1 0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 -1 -1 EOB)

185	3	1	1	-3	2	-1	0
1	1	-1	0	-1	0	0	1
0	0	1	0	-1	0	0	0
1	1	0	-1	0	0	0	-1
0	0	1	0	0	0	-1	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

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

Detail in a block vs. DCT coefficients transmitted

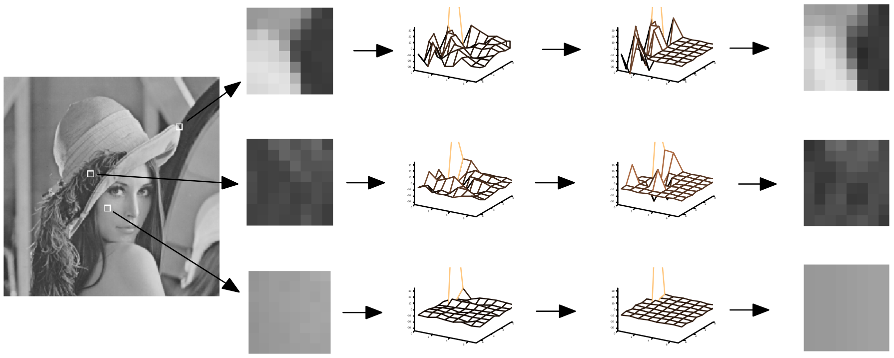


image block

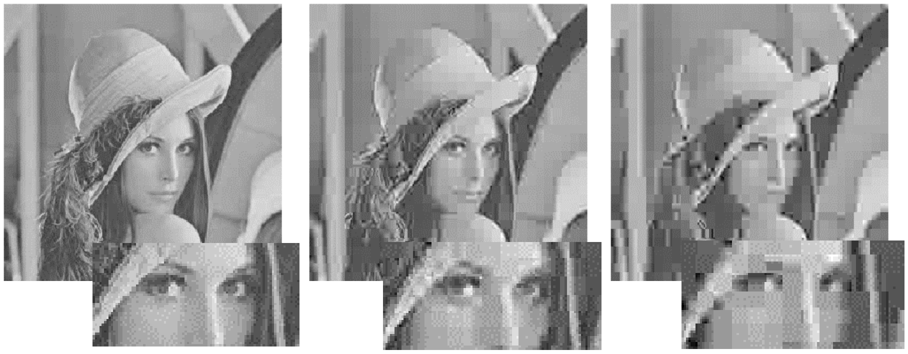
DCT coefficients of block

quantized DCT coefficients of block

block reconstructed from quantized coefficients

Typical DCT coding artifacts

DCT coding with increasingly coarse quantization, block size 8x8

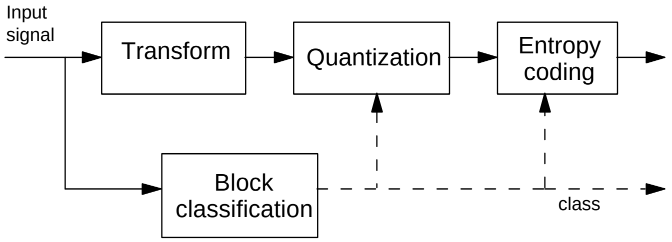


quantizer stepsize for AC coefficients: 25

quantizer stepsize for AC coefficients: 100

quantizer stepsize for AC coefficients: 200

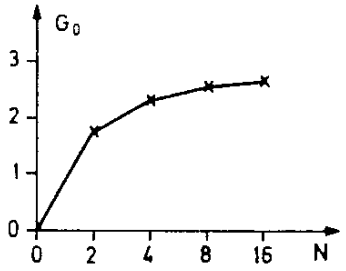
Adaptive transform coding



- 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*x*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):

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

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

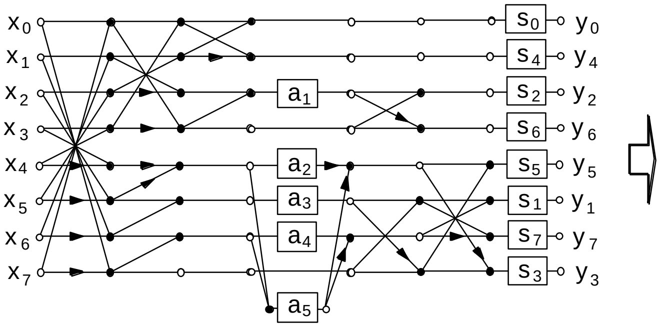
$$M_1 = \begin{pmatrix} 1 &&&&&&& \\ & 1 &&&&& 0 & \\ && 1 &&&&& \\ &&& 1 &&&& \\ &&&& 1 &&& 1 \\ &&&&& 1 & 1 & \\ & 0 &&&& 1 & -1 & \\ &&&& -1 &&& 1 \end{pmatrix} \quad M_2 = \begin{pmatrix} 1 &&&&&&& \\ & 1 &&&&& 0 & \\ && 1 & 1 &&&& \\ && -1 & 1 &&&& \\ &&&& 1 &&& \\ &&&&& 1 && 1 \\ & 0 &&&&& 1 &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

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

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

Fast DCT algorithm II

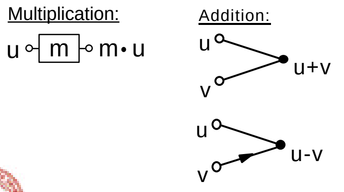
- Signal flow graph for fast (scaled) 8-DCT according to Arai, Agui, Nakajima:



scaling

only 5 + 8 multiplications

(direct matrix multiplication: 64 multiplications)



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

$$s_0 = \frac{1}{2\sqrt{2}}$$
$$s_k = \frac{1}{4C_k} \qquad k = 1, \ldots, 7$$
$$C_k = \cos\left(\frac{p}{16}k\right)$$

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