Chapter 4 Transform Coding

Outline

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

Transform Coding (TC)

- TC: another efficient coding scheme based on utilization of inter-pixel correlation.
  - A core technique recommended by JPEG
  - Efficient in terms of coding prediction error in motion compensated predictive coding, adopted by all international video coding standards.
- Transformation decides which format of input source is quantized and encoded
  - In DPCM: the difference signal (smaller variance)
  - In TC: the transformed version of a signal (**less correlated**)

Hotelling Transform

- Consider an *N*-dimensional vector $\vec{z}_s$
- The ensemble of such vectors can be modeled by a random vector

$$\vec{z} = (z_1, z_2, \cdots, z_N)^T$$

Where each component z_i is a random variable

- The mean vector of the population

$$m_{\vec{z}} = E[\vec{z}] = (m_1, m_2, \cdots, m_N)^T, m_i = E[z_i]$$

- The covariance matrix:

$$C_{\vec{z}} = E[(\vec{z} - m_{\vec{z}})(\vec{z} - m_{\vec{z}})^T]$$

with element $C_{i,j} = Cov(z_i, z_j)$

$C_{i,i}$: variance of z_i

  - $C_{\vec{z}}$: real and symmetric

Hotelling Transform

- According to theory of linear algebra, always possible to find a set of *N* orthonormal eigenvectors of the matrix $C_{\vec{z}}$ to convert it into a full ranked diagonal matrix
  - *N* orthonoromal eigenvectors $\vec{e}_i$
  - With corresponding eigenvalues: λ_i
  - Form an orthogonal matrix

$$\phi = (\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_N)^T$$

- Hotelling transform (or eigenvector transform):

$$\vec{y} = \phi(\vec{z} - m_{\vec{z}})$$

Hotelling Transform

- Two features
  - $m_{\vec{y}} = 0$
  - $$C_{\vec{y}} = \phi C_{\vec{z}} \phi^T = \begin{bmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & & \lambda_N \end{bmatrix}$$
    - Correlation previously existing between different components has been removed in the transformed domain
- Inverse Hotelling transform

$$\vec{z} = \phi^{-1}\vec{y} + m_{\vec{z}} = \phi^T \vec{y} + m_{\vec{z}}$$

- Can be viewed as discrete version of the Karhunen-Loeve transform (KLT)

Statistical Interpretation

- The elements in the main diagonal of $C_{\vec{y}}$
  - Eigenvalues of $C_{\vec{y}}$.
  - Variances of the components of vector $\vec{y}$, denoted by

    $\sigma^2_{y,1}, \sigma^2_{y,2}, \cdots, \sigma^2_{y,N}$
- Arrange eigenvalues (variances) in non-increasing order:

  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$
- Choose an integer *L* where *L*<*N*
- Using the corresponding *L* eigenvectors $\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_L$, to form a new matrix $\bar{\phi}$ of dimension *L*x*N*
  - transform changes to $\vec{y} = \bar{\phi}(\vec{z} - m_{\vec{z}})$
  - Inverse transform changes to $\vec{z}' = \bar{\phi}^T \vec{y} + m_{\vec{z}}$

Statistical Interpretation

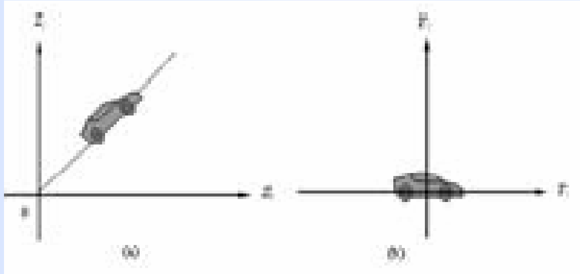
- It can be shown that *MSE* between $\vec{z}$ and $\vec{z}'$ is given by

$$MSE_r = \sum_{i=L+1}^{N} \sigma_{y,i}^2$$

- Quantizing and encoding only *L* components of vector $\vec{y}$ lead to higher coding efficiency – basic idea behind transform coding
- The linear unitary transform provides two functions.
  - Decorrelate input data
  - Some transform coefficients more significant than others
    - Some can be discarded
    - Some can be coarsely quantized
    - Some can be finely quantized

Geometrical Interpretation

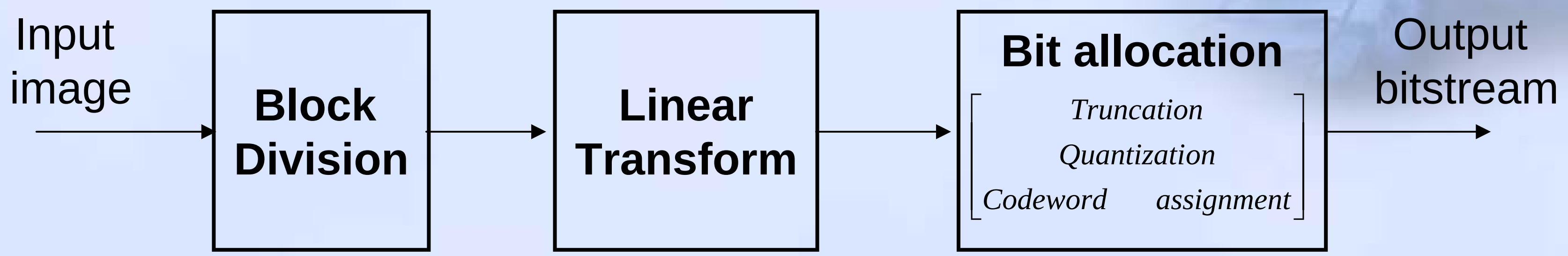
- A binary image of a car in Fig. 4.1 (a)
  - Each pixel in the shaded object region: a 2-D vector with its two components being coordinates z_1 and z_2 respectively.
  - The Hotelling transform
  - Note the two features



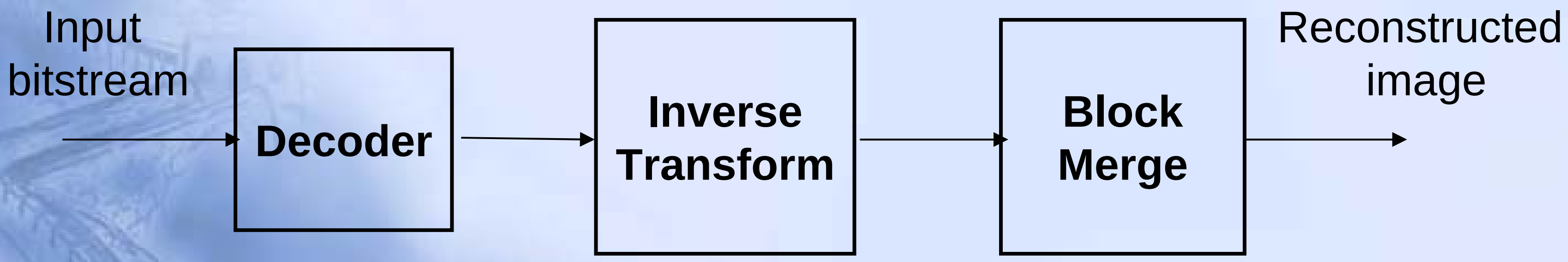
4.1(a) 4.1(b)

Fig. 4.1 (b): after transform, object is aligned with principal axes

Procedures of Transform Coding



(a) Transmitter



(b) Receiver

Linear Transforms

- A digital image: a 2-D array g(x, y)
  - (x,y) coordinates of a pixel in 2-D array
  - g(): gray level value of the pixel
- T(u,v): 2-D transform of g(x,y)
  - (u,v): coordinates in transformed domain.
- 2-D forward and inverse transforms:

$$T(u,v) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} g(x,y) f(x,y,u,v)$$

$$g(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} T(u,v) i(x,y,u,v)$$

  - f(x,y,u,v) and i(x,y,u,v): forward and inverse ***transformation Kernels***

Separability

- A transformation kernel is ***separable*** if

$$f(x,y,u,v) = f_1(x,u)\,f_2(y,v)$$
$$i(x,y,u,v) = i_1(x,u)\,i_2(y,v)$$

- A 2-D separable transform can be decomposed into two 1-D transforms

$$T_1(x,v) = \sum_{y=0}^{N-1} g(x,y)\,f_2(y,v)$$

$$T(u,v) = \sum_{x=0}^{N-1} T_1(x,v)\,f_1(x,u)$$

- A transformation kernel is ***symmetric*** if it is separable and

$$f_1(y,v) = f_2(y,v)$$

Matrix Form

- If a transformation kernel is symmetric (hence separable), then 2-D image transform can be expressed in matrix form.
  - Denote an image matrix by *G:* $G = \{g_{i,j}\} = \{g(i-1, j-1)\}$
  - Denote **forward transform matrix** by *F* (NxN)

    $$F = \{f_{i,j}\} = \{f_1(i-1, j-1)\}$$

  - Denote inverse transform matrix by *I:* $I = \{i_{j,k}\} = \{i_1(j-1, k-1)\}$

    $$I = F^{-1}$$

  - Then $$T = F^T G F$$ $$G = I^T T I$$
  - DFT (complex quantities) $$T = F^{*T} G F$$ $$G = I^{*T} T I$$ $$I = F^{-1} = F^{*T}$$
    - "*": complex conjugation
    - Unitary transform

Orthogonality

- Definition: a transform is orthogonal if

$$F^T = F^{-1}$$

- An orthogonal matrix (transform) is a special case of a unitary matrix (transform):
  - only real quantities are involved.
- All 2-D image transforms to be presented:
  - separable, symmetric, and unitary.

Basis Image Interpretation

- Consider 2-D inverse transform

$$g(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}T(u,v)i(x,y,u,v)$$

- Consider an *N*x*N* **basis image** for a specific (u,v)

$$I_{u,v}=\{i(x,y,u,v),0\le x,y\le N-1\}$$

  - That is:

$$I_{u,v}=\begin{bmatrix} i(0,0,u,v) & i(0,1,u,v) & \cdots & \cdots & i(0,N-1,u,v) \\ i(1,0,u,v) & i(1,1,u,v) & \cdots & \cdots & i(1,N-1,u,v) \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ i(N-1,0,u,v) & i(N-1,1,u,v) & \cdots & \cdots & i(N-1,N-1,u,v) \end{bmatrix}$$

Basis Image Interpretation

- The inverse transform in a collective form

$$G = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} T(u,v) I_{u,v}$$

- Interpretation:
  - A series expansion of the original image *G* into a set of N^2 basis images
  - Transform coefficients $T(u,v), 0 \le u,v \le N-1$, become the coefficients of the expansion
  - Coefficient (weight) T(u,v): **a correlation measure between image *G* and basis image** [Wintz'72]

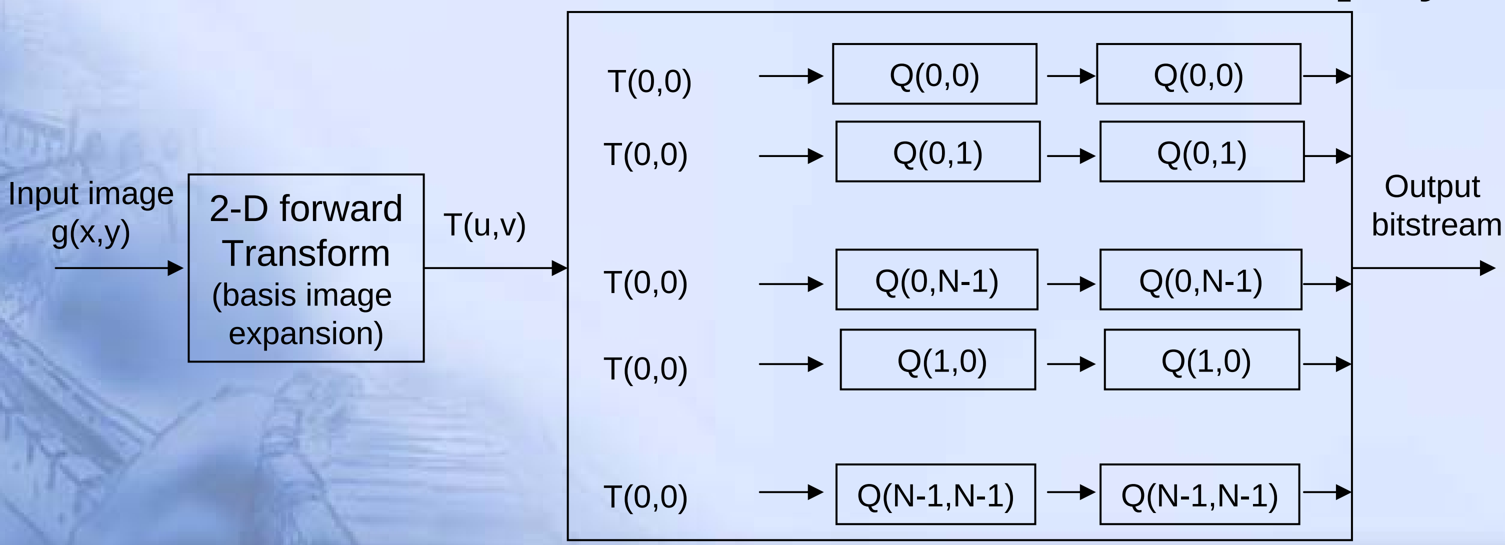
Basis Image Interpretation

- Comments on basis images:
  - Have nothing to do with the input image
  - Completely defined by transform itself
    - Different transforms ⇔ different sets of basis images
- Motivations:
  - With proper transform, transform coefficients are more *independent* than the gray scales of original input image.
  - Optimum linear transform: uncorrelated coefficients
  - The coefficient variance varies widely
    - Insignificant coefficients ignored
    - Significant coefficients are allocated more bits
  - Coding efficiency is thus enhanced.

Basis Image Interpretation

- TC coding can be considered a special case of subband coding, although TC was devised much earlier
- An alternative way to define basis images
  - $I_{u,v} = \vec{b}_u \vec{b}_v^T$

    the outer product of the basis vectors, where $\vec{b}_u$ is the *u*th column vector of the inverse transform matrix [Jayant'84]



Subimage Size Selection

- The larger the size, the more decorrelation TC can achieve
- Correlation between image pixels, however, becomes insignificant when the distance becomes large
- Problems with large size:
  - A large block size can NOT adapt to local statistics well
  - A transmission error in TC affects the whole subimage
    - A possibly severe effect of transmission error for large size
  - In video coding, TC is typically used together with motion compensated (MC) coding. Large block size not used in MC.
- Subimage sizes (N) of 4,8,16 are used most often

Transforms of Particular Interest

- Discrete Fourier Transform (DFT)
- Discrete Walsh Transform (DWT)
  - DWT transformation kernels [Walsh'23]:

$$f(x,y,u,v)=\frac{1}{n}\prod_{i=0}^{n-1}[(-1)^{p_i(x)p_{n-1-i}(u)}(-1)^{p_i(y)p_{n-1-i}(v)}]$$

Where $n=\log_2 N$

P_i(arg): *i*th bit in NBC (natural binary representation) of the arg

The *0*th bit: the least significant bit

The (*n-1*)th bit: the most significant bit

  - For instance, consider N=16 (n=4)

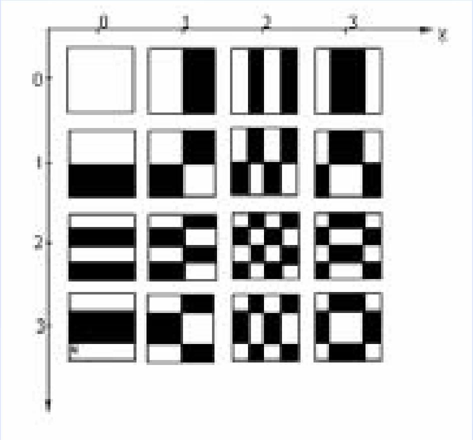
The NBC of 8 is1000, $p_0(8)=p_1(8)=p_2(8)=0,\ p_3(8)=1$

Discrete Walsh Transforms (DWT)

- DWT transform matrix for N=4:

$$F = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

- The DWT implementation is simple
  - *F* is simple
  - $i(x, y, u, v) = f(x, y, u, v)$



A set of 16 basis images for N=4

Discrete Hadamard Transform

- Closely related to DWT
  - DHT transformation kernels [Hadamard'1893]:

$$f(x,y,u,v) = \frac{1}{n}\prod_{i=0}^{n}[(-1)^{p_i(x)p_i(u)}(-1)^{p_i(y)p_i(v)}]$$

$$i(x,y,u,v) = f(x,y,u,v)$$

Where *n*, *i*, and P_i(arg) are the same as in DWT

- Walsh-Hadamard transform (DWHT) is frequently used to represent either of the two transforms

Discrete Cosine Transform (DCT)

- Background
  - Most commonly used transform for image and video coding
  - Established by Ahmed, Natarajan and Rao [Ahmed'74]
- The basis vectors of 1-D DCT: a good approximation to the eigenvectors of the class of Toeplitz matrices defined as (where $0 < \rho < 1$)

$$\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix}$$

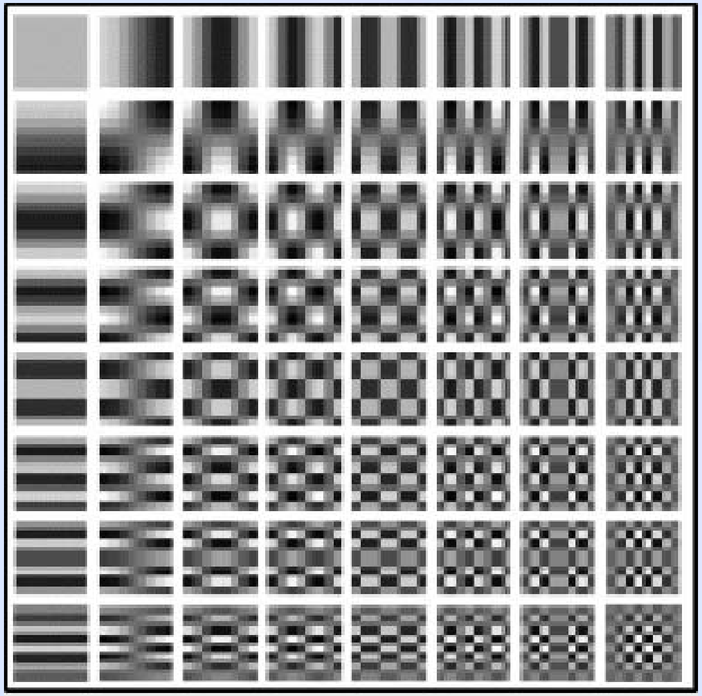
DCT Transformation Kernel

$$f(x,y,u,v) = C(u)C(v)\cos\left(\frac{(2x+1)u\pi}{2N}\right)\cos\left(\frac{(2y+1)v\pi}{2N}\right)$$

Where

$$C(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for} \quad u = 0 \\ \sqrt{\frac{2}{N}} & \text{for} \quad u = 1, 2, \cdots, N-1 \end{cases}$$

$C(v)$ is defined in the same way



64 basis images for N=8

Relationship with DFT

- The DCT of an *N*-point sequence $g_N(n)$ is obtained via the following three steps
  - Form a *2N*-point sequence, $g_{2N}(n)$, then form a periodic sequence $\tilde{g}_{2N}(n)$
  - Find Fourier Series (FS) coefficients of $\tilde{g}_{2N}(n)$
  - Only keep the *N* coefficients with indexes $0,1,\cdots,N-1$ which are the DCT coefficients of the given *N*-point sequence $g_N(n)$

Relationship with DFT

- Observations:
  - $\tilde{g}_N(n)$ not smooth
    - End-head discontinuities cause high frequency distribution in DFT
  - $\tilde{g}_{2N}(n)$ does not have this discontinuity
- DCT has better energy compaction in low frequency than DFT
  - In terms of energy compaction, DCT is the best among DFT, DWT, DHT and discrete Harr transform
  - KLT: image dependent, not practical

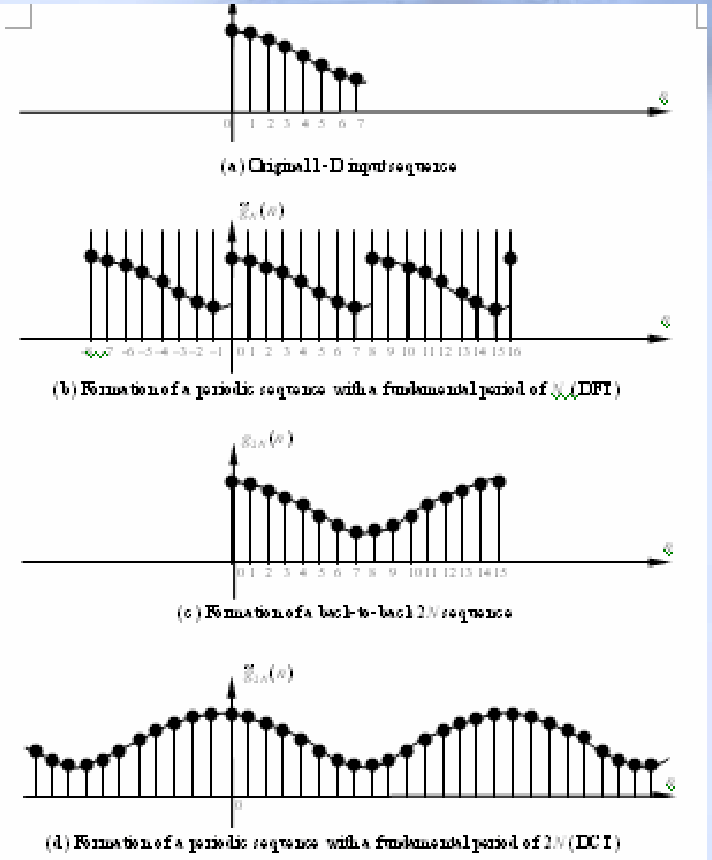


Figure 4.5 An example to illustrate the differences and similarities between DFT and DCT

Relationship with DFT

- DCT can be implemented using the FFT
  - DCT of an *N*-point sequence can be obtained from DFT of a *2N*-point sequence,
  - *Even symmetry* of $\tilde{g}_{2N}(n)$ makes computation required for DCT of an *N*-point sequence equal to that required for DFT of an *N*-point sequence
- As a result, DCT is the most popular image transform used in image and video coding

Performance Comparison

- Energy compaction
  - Mean square approximation error

$$MSE_r = E\left[\left|(\vec{z} - \vec{z}')\right|^2\right] = \sum_{i=L+1}^{N} \sigma_i^2$$

Where $\vec{z}'$ denotes the reconstructed vector

- DWT, DFT, DCT can be implemented using FFT

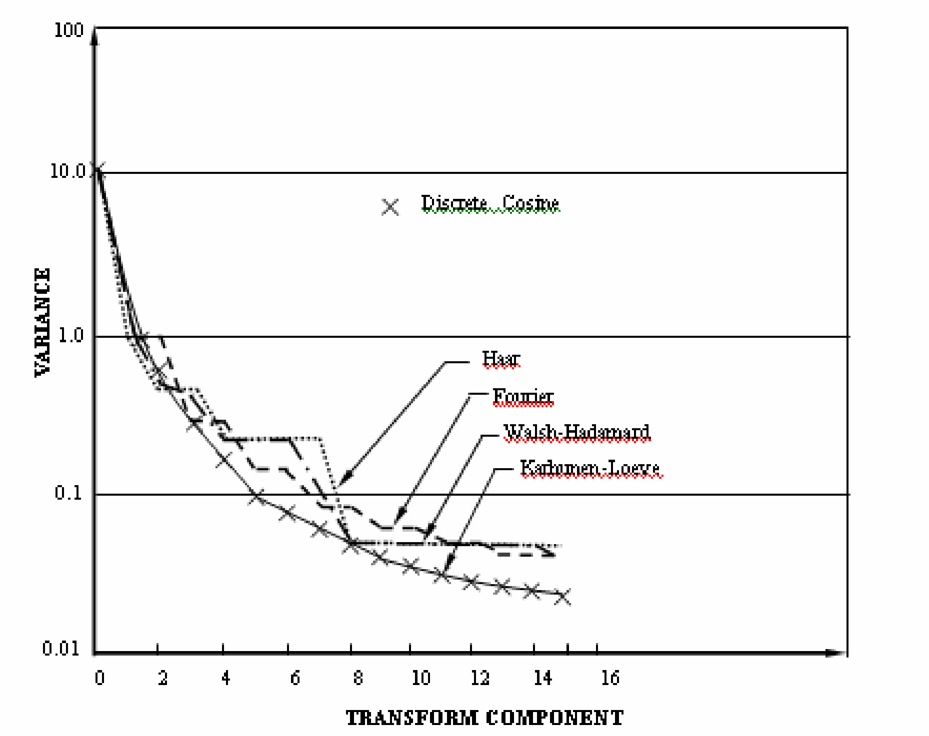


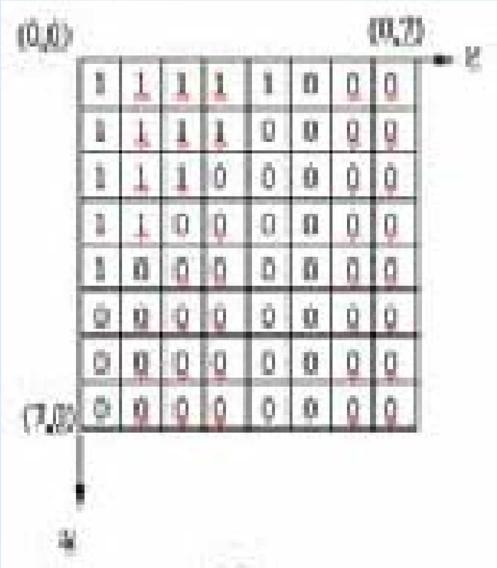
Figure 4.6 Transform coefficient variances when N=16, ρ=0.95 [ahmed 1974]

Bit Allocation

- Bit allocation: truncation, quantization, and codeword assignment
- Three types of error in TC
  - Truncation error: majority of coefficients are truncated to zero
  - Quantization error
  - Transmission error
- Two different ways in truncation
  - Zonal coding
  - Threshold coding

Zonal Coding

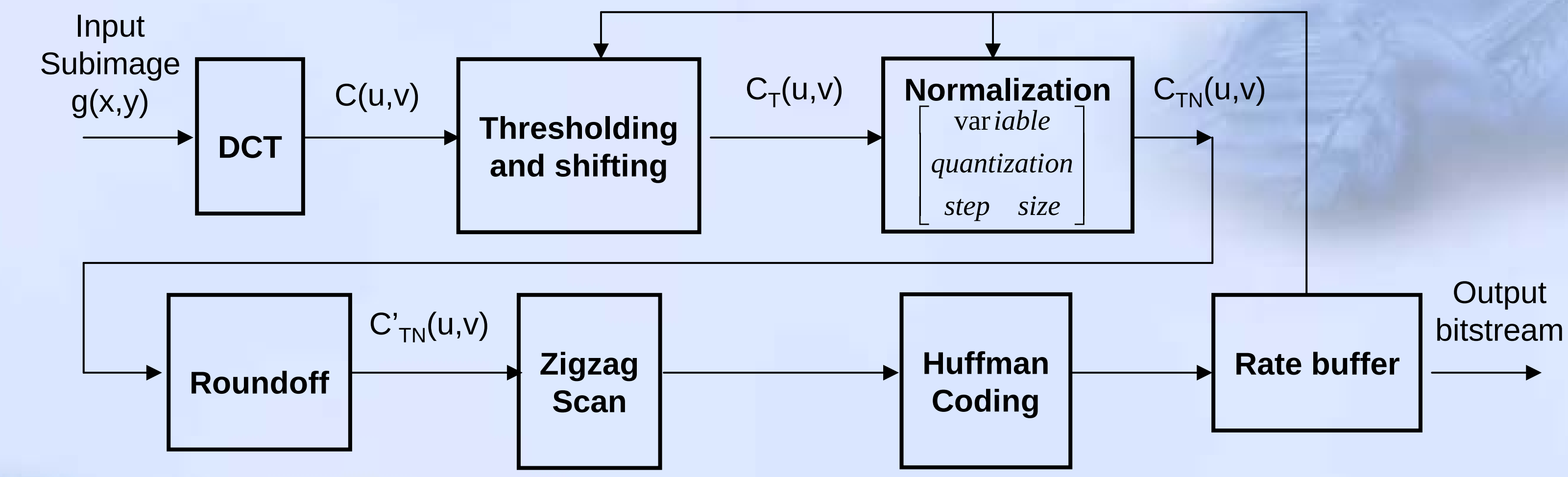
- No overhead side information
- Coding efficiency, however, may not be high
  - Some coefficients outside the zone might be large
  - Some coefficients inside the zone may be small
- For further improvement, adaptive scheme has to be used.



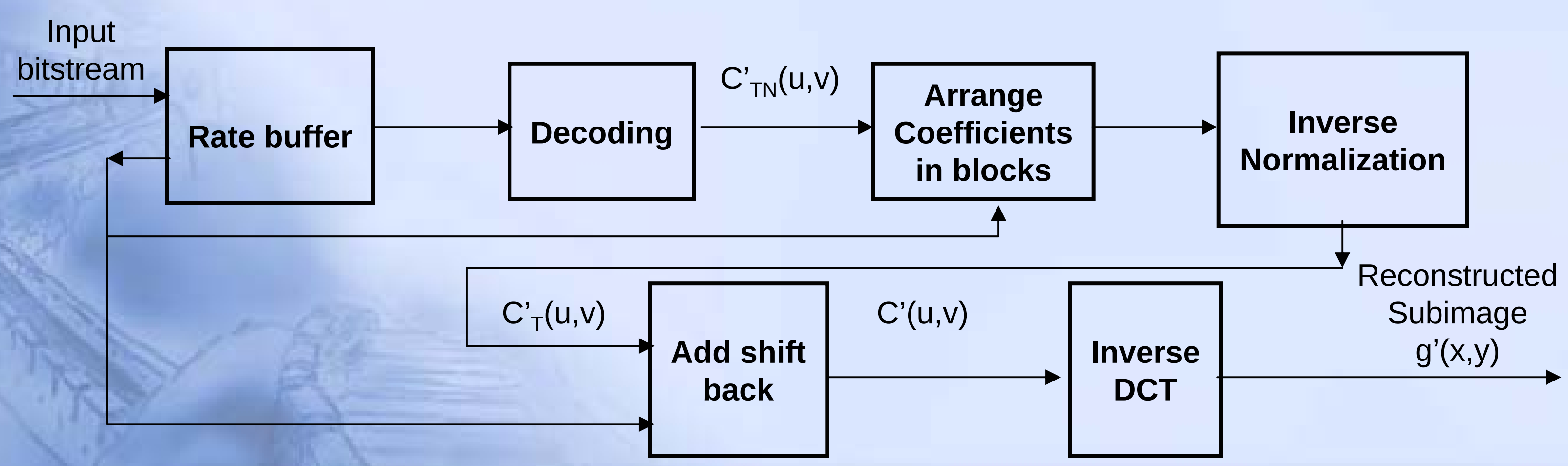
Threshold Coding

- No predefined zone
- Each transform coefficient compared with a threshold
  - Smaller than threshold, set to zero
  - Larger than threshold, retained for quantization/encoding
  - Threshold determined after evaluation of all coefficients
- Address of retained coefficients has to be sent to receiver as side information
  - Usually a two-pass adaptive technique

Threshold Coding



(a) Transmitter



(b) Receiver

Threshold Coding

- Chen and Pratt: an efficient adaptive coding scheme [Chen'84]
  - A ***one-pass*** fast adaptive scheme
  - With several effective techniques, it achieved excellent results in TC
  - Demonstrated satisfactory quality of reconstructed frames at 0.4 bpp for coding color images
    - $\Rightarrow$ real-time color television transmission over a 1.5 Mbps channel
  - Adopted by JPEG

Thresholding and Shifting

- Formula

$$C_T(u,v) = \begin{cases} C(u,v) - T & \text{if} \quad C(u,v) > T \\ 0 & \text{otherwise} \end{cases}$$

  - T: threshold
- More than 60% of DCT coefficients normally fall below a threshold as low as 5

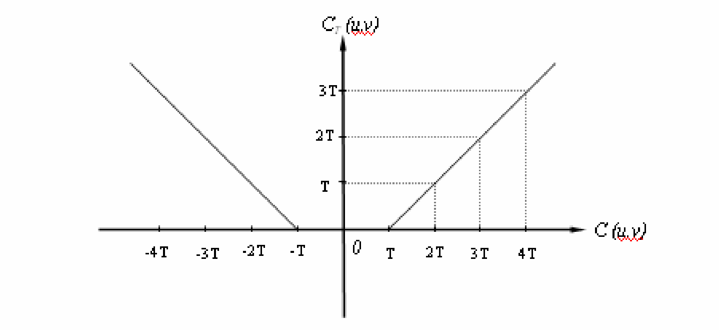


Figure 4.9 Input-output characteristic of thresholding and shifting

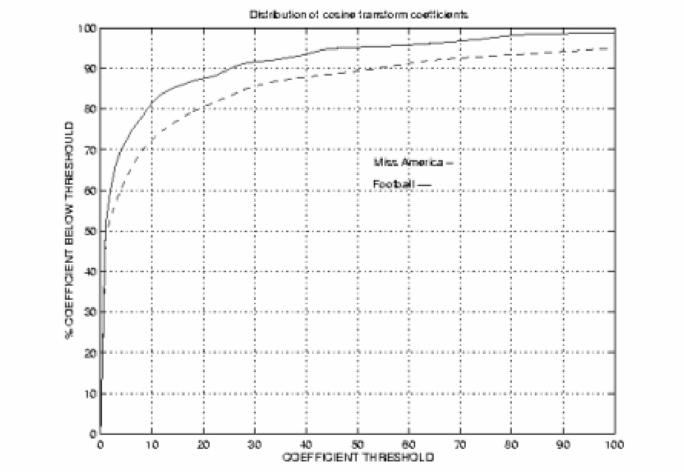


Figure 4.10 Distribution of DCT coefficients

Percentage vs threshold

Normalization and Roundoff

- Normalization is implemented as follows:

$$C_{TN}(u,v) = \frac{C_T(u,v)}{\Gamma_{u,v}}$$

  - $\Gamma_{u,v}$: **normalization factor** controlled by the rate buffer
- Then roundoff process converts floating point to nearest integer
  - Uniform midtread quantizer with unit quantization step
  - Normalization is a scaling process
    - Makes resultant uniform midtread quantizer **adapt to dynamic range** of the coefficients.
  - By adjusting $\Gamma_{u,v}$, variable bit rate and MSE can be achieved.
    - Take into account image statistics and HVS characteristics.

Normalization and Roundoff

- HVS more sensitive to luminance component than to chrominance components
- HVS more sensitive to low frequency components
- Quantization table in JPEG
  - a matrix consisting of all normalization factors

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	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

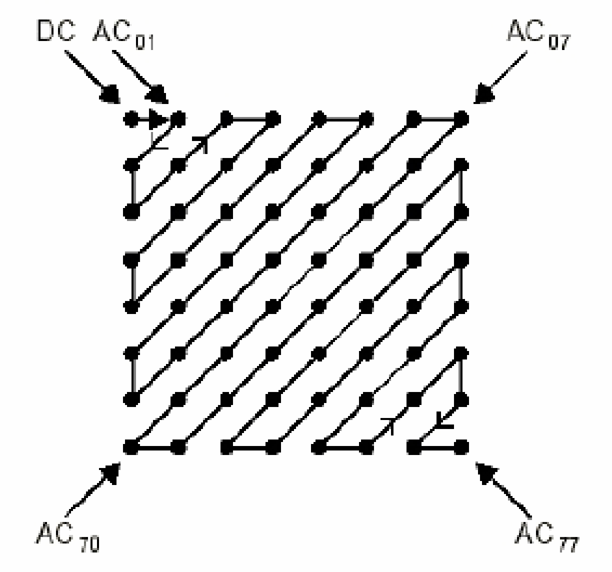
Luminance quantization table

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Chrominance quantization table

Zigzag Scan

- Most quantized coefficients are zero
- Run-length code (RLC)
  - Very efficient to encode address information of nonzero coefficients
  - Run-length of zero coefficients: # of consecutive zeros in zigzag scan
  - Zigzag scanning minimizes use of RLC in the block



Huffman Coding

- Statistical studies of magnitude of nonzero DCT coefficients and run-length of zeros were conducted in [Chen'84]
  - Domination of coefficients with small amplitude and short run-lengths was found
- This justifies the applications of Huffman coding to magnitude of nonzero coefficients and run-lengths of zeros
- **Special codewords**
  - End of block (EOB)
    - Indicate the termination of coding a block
    - save bits
  - *Run-length prefix*:
    - differentiate run-length codewards from amplitude codewords.

Rate Buffer Feedback and Equalization

- A rate buffer accepts a ***variable***-rate data input from the encoding process and provides a ***fixed***-rate data output to the channel
- The status of the rate buffer is monitored and fed back to control the threshold and the normalization factor
  - A one-pass adaptation is achieved.

Some Issues

- Effect of transmission error
  - In TC, each pixel in reconstructed image relies on all transform coefficients of the subimage where the pixel is located
  - A bit reversal transmission error will be spread
    - Subimage size should be limited
  - The effect on the reconstructed image varies
    - An error in DC or low frequency AC coefficients may be objectionable
    - An error in high frequency AC coefficients less noticeable

Reconstruction Error Sources

- Three sources for reconstruction error
  - Truncation, quantization, transmission
- Block artifacts
  - TC are usually conducted blockwise
  - Block boundary artifacts can be annoying to HVS, especially at low bit rate
- To alleviate blocking artifacts
  - Block overlapping
  - Post-filtering reconstructed image along block boundary
  - Advanced transform

Reconstruction Error Sources

- Block overlapping
  - Each pixel in overlapped regions takes an average of all its reconstructed pixel values from multiple blocks
  - Extra bits are used for those pixels in overlapped regions
    - Overlapped region is usually only one pixel wide
- Post-filtering
  - *Low pass* filtering block boundaries
  - No bit overhead, better result
  - Adopted by international coding standards.

Comparison: DPCM vs TC

- Both utilize inter-pixel correlation, and are efficient
- In terms of computational complexity, DPCM simpler
- In terms of memory requirement and processing delay, DPCM superior
- Design of DPCM system, and its performance, is sensitive to image-to-image variation.
  - Optimum DPCM design is matched to statistics of certain image
  - TC less sensitive to variation in image statistics
- In general, optimum DPCM with a third or higher order predictor performs better than TC for bit rate about 2-3 bits per pixel
  - When bit rate is lower, TC is preferred

Hybrid Coding

- Hybrid transform/waveform coding
  - In waveform coding, the waveform of a signal is coded, e.g., DPCM
  - Hybrid coding combines TC and DPCM coding, i.e., TC applied first rowwise followed by DPCM coding columnwise, or vice versa.
  - Both used in JPEG
- Two techniques complement each other: has TC's small sensitivity to variable statistics and DPCM's simplicity
- A successful hybrid coding scheme in interframe coding
  - Use motion compensated predictive coding
  - Prediction error is transform coded
  - Efficient and adopted by international video coding standards
- HW #2: Ex. 4-8