Multimedia System

Media Compression (Images)

Dr. Reham Reda Mostafa

Mansoura University
Faculty of Computers and Information
Dept. of Information System

dr.reham2206@yahoo.com

4. December 2019

Outlines

- Overview
- Redundancy and Relevancy of image data
- Classes of image compression techniques
- Lossless Image Coding
  - Image Coding Based on Run Length
  - Dictionary-Based Image Coding (GIF, PNG)
- Lossy Image Coding
  - TRANSFORM IMAGE CODING
    - DCT Image Coding and the JPEG Standard
    - JPEG Bit Stream
    - Drawbacks of JPEG
  - WAVELET BASED CODING (JPEG 2000)
  - JPEG 2000 Versus JPEG

Overview

- Visual media types such as images, video, and graphics are now ubiquitous in today's digital world.
- One important effect of the change is an ever-increasing demand for the access, transfer, storage, and interpretation of increasing amounts of image information, quickly and conveniently.
- Image compression technology is central to meeting this demand, which is manifested in many different applications.
- This lecture is devoted to explaining the state of the art in image compression and the technical issues involved in compression of images.
  - Section 1 explains the redundancy in image information and how that can be exploited by compression algorithms.
  - Section 2 describes generic classes of image compression techniques.
  - A few of these algorithms are selected for detailed explanations because of their popular use and their adoption into the ISO/ITU standards, such as JPEG and JPEG 2000.

Overview

- Following figure provides disk memory requirements and transmission times across different bandwidths for a variety of consumer- and industry-used image formats.
- From figure, it should be clear that compression techniques that produce compression ratios from 20:1 to 30:1 and even more are necessary to make image storage and distribution feasible for consumer-level and industry-grade applications.

Multimedia image data	Grayscale image	Color image	HDTV video frame	Medical image	Super High Definition (SHD) image
Size/duration	512 × 512	512 × 512	1280 × 720	2048 × 1680	2048 × 2048
Bits/pixel or bits/sample	8 bpp	24 bpp	12 bpp	12 bpp	24 bpp
Uncompressed size (B for bytes)	262 KB	786 KB	1.3 MB	5.16 MB	12.58 MB
Transmission bandwidth (b for bits)	2.1 Mb/image	6.29 Mb/ image	8.85 Mb/ frame	41.3 Mb/ image	100 Mb/ image
Transmission time (56 K modem)	42 seconds	110 seconds	158 seconds	12 min.	29 min.
Transmission time (780 Kb DSL)	3 seconds	7.9 seconds	11.3 seconds	51.4 seconds	2 min.

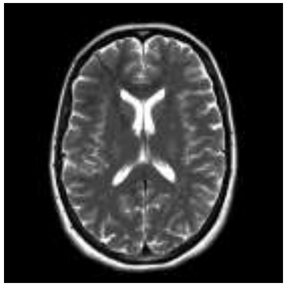
Figure 7-1 Examples showing storage space, transmission bandwidth, and transmission time required for uncompressed images

Outlines

- Overview
- Redundancy and Relevancy of image data
- Classes of image compression techniques
- Lossless Image Coding
  - Image Coding Based on Run Length
  - Dictionary-Based Image Coding (GIF, PNG)
- Lossy Image Coding
  - TRANSFORM IMAGE CODING
    - DCT Image Coding and the JPEG Standard
    - JPEG Bit Stream
    - Drawbacks of JPEG
  - WAVELET BASED CODING (JPEG 2000)
  - JPEG 2000 Versus JPEG

Redundancy and Relevancy of image data

- Image compression techniques can be purely lossless or lossy, but good image compression techniques often work as hybrid schemes, making efficient use of lossy and lossless algorithms to compress image data.
- These schemes **aim to get** compression by typically analyzing the image data according to two important aspects:
  - ***Irrelevancy reduction***—In many cases, information associated with some pixels might be irrelevant and can, therefore, be removed. These irrelevancies can be further categorized
    - *Visual irrelevancy:* Information can be deemed visually irrelevant when the image sample density exceeds the limits of visual acuity given the display/viewing conditions.
    - *Application-specific irrelevancy* occurs when an image or image region might not be required for the application.



Redundancy and Relevancy of image data

- ***Redundancy reduction***—The image signal, like any other signal, has statistical redundancy because pixel values are not random but highly correlated, either in local areas or globally. The repeated use of pixels can be statistically analyzed by entropy-coding techniques to reduce the amount of bits required to represent groups of pixels.
  - The redundancy can be further classified as follows:
    - ***Spatial redundancy***—The pixels' intensities in local regions are very similar.
    - ***Spectral redundancy***—When the data is mapped to the frequency domain, a few frequencies dominate over the others.
    - There is also another kind of redundancy—***temporal redundancy*—which** occurs when a sequence of images, such as video, is captured.

Redundancy and Relevancy of image data

- The relevancy deals with the removal of image data using a lossy scheme due to the data's lack of relevancy either in perception or importance for the required application.
  - This is achieved by applying quantization in the spatial and/or frequency domains.
  - The lossy quantization effects often introduce a distortion, which must be minimized, so as to go either unnoticed or be perceptually acceptable by the human visual system (HVS).
- The redundancy is reduced by lossless or entropy-coding techniques, which use statistical analysis to reduce the average number of bits used per pixel.
  - Although there are standards such as GIF, TIFF, and PNG that use lossless coding techniques only, most commercially used standards that require higher compression use both lossy and lossless techniques in combination.

Redundancy and Relevancy of image data

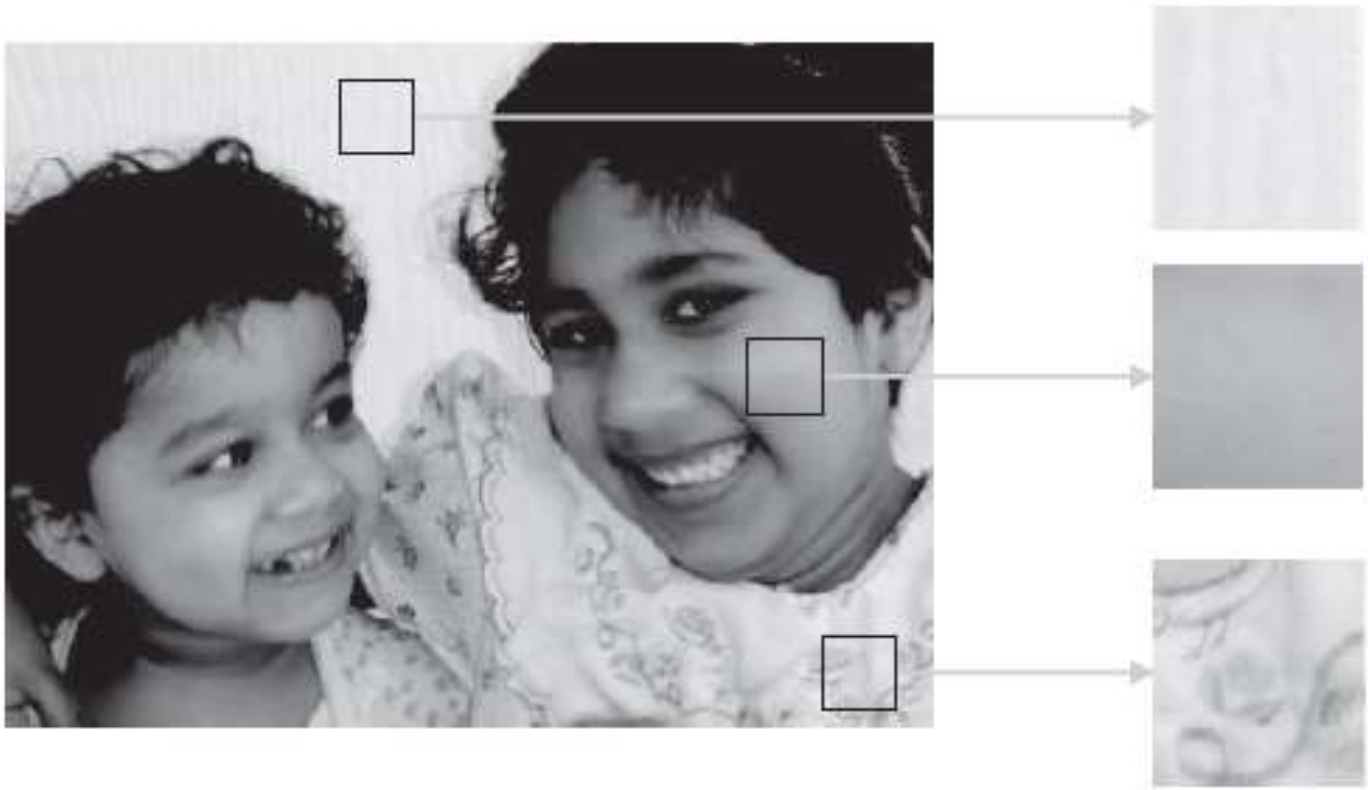


Figure 7-2 Local pixel correlation. Images are not a random collection of pixels, but exhibit a similar structure in local neighborhoods. Three magnified local areas are shown on the right.

Outlines

- Overview
- Redundancy and Relevancy of image data
- Classes of image compression techniques
- Lossless Image Coding
  - Image Coding Based on Run Length
  - Dictionary-Based Image Coding (GIF, PNG)
- Lossy Image Coding
  - TRANSFORM IMAGE CODING
    - DCT Image Coding and the JPEG Standard
    - JPEG Bit Stream
    - Drawbacks of JPEG
  - WAVELET BASED CODING (JPEG 2000)
  - JPEG 2000 Versus JPEG

Classes of image compression techniques

- Image compression techniques can be generically categorized as either *lossless* or *lossy*.
- In lossless compression schemes, the reconstructed image, after compression, is numerically identical to the original image.
  - The classes of lossless techniques mainly exploit spatial and statistical redundancies.
    - The entropy-coding techniques (huffman and arithmetic coding) are sufficient to fully exploit the statistical redundancy. In practice, however, these techniques by themselves do not achieve sufficient compression.
    - The run-length coding are sufficient to fully exploit the spatial redundancy.
- The lossy schemes normally operate by quantization of image data either directly in the pixel domain or in the frequency domain obtained via different transformations.

Classes of image compression techniques (cont'd)

An overall taxonomy is shown in following figure.

- Also shown in bold are popular standards based on the respective algorithms.
- Several selected techniques among these are discussed in more detail in the following few sections.

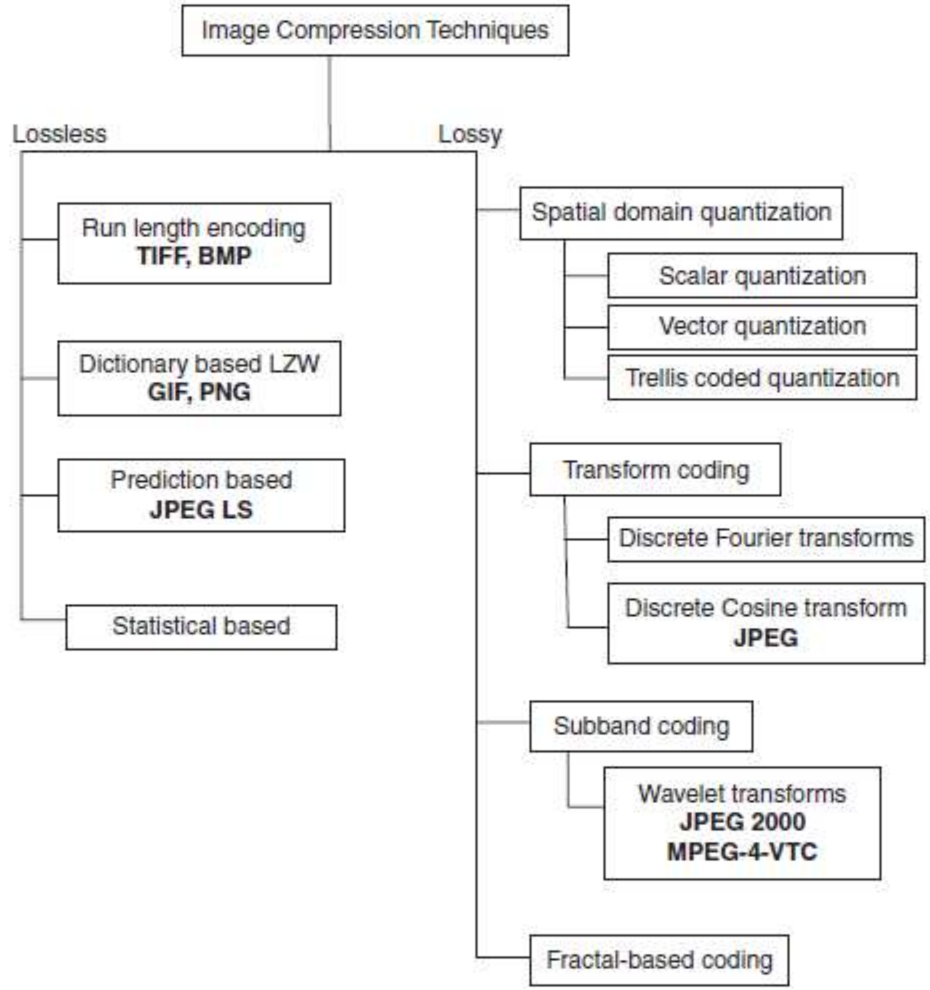


Figure 7-3 Taxonomy of image compression techniques. The left column shows lossless categories and the right column shows lossy categories. Commonly used standards based on the techniques are shown in bold.

Outlines

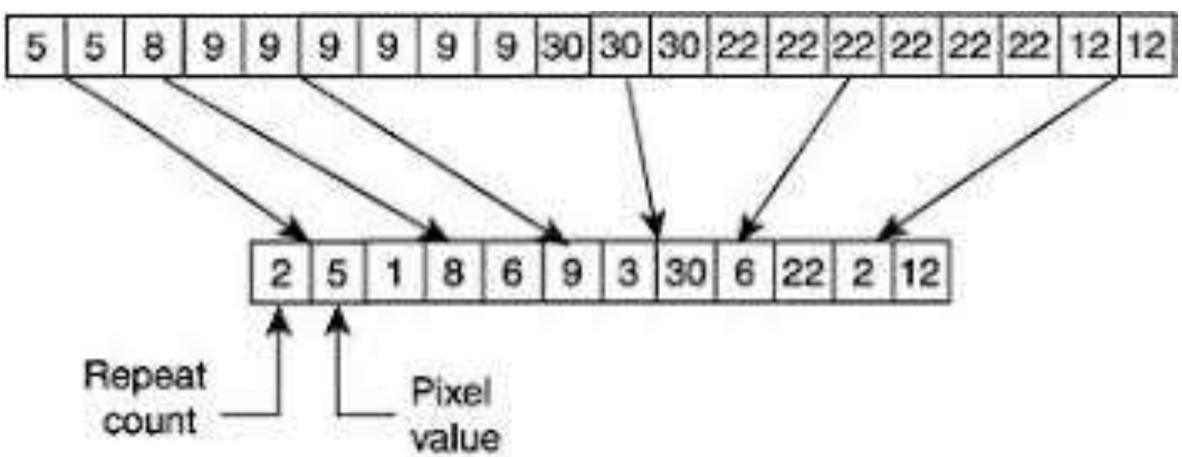
- Overview
- Redundancy and Relevancy of image data
- Classes of image compression techniques
- Lossless Image Coding
  - Image Coding Based on Run Length
  - Dictionary-Based Image Coding (GIF, PNG)
- Lossy Image Coding
  - TRANSFORM IMAGE CODING
    - DCT Image Coding and the JPEG Standard
    - JPEG Bit Stream
    - Drawbacks of JPEG
  - WAVELET BASED CODING (JPEG 2000)
  - JPEG 2000 Versus JPEG

Lossless Image Coding

- The first compression schemes were designed to compress digital images for storage purposes, and were based on lossless coding techniques.
- Compared with lossy techniques, purely lossless compression can only achieve a modest amount of compression, so their use is somewhat limited in practice today.
- However, certain applications do need to store images in lossless form—for example, the film industry renders mages and the original image sequences are always stored in lossless form to preserve image quality.
- Lossless image compression schemes correspond to the BMP, TIFF, GIF, and PICT formats, among others.
- Example:
  - Run-length coding
  - Dictionary-based coding
  - Entropy-based coding (huffman coding, arithmetic coding)

Image Coding Based on Run Length (BMP, TIFF)

- Run length encoding (RLE) is probably one of the earliest lossless coding schemes that was applied to images, and was incorporated in the earlier formats such as BMP, TIFF, and RLA on PCs and PICT on Macintosh machines.
- As described in the previous lecture, this method simply replaces a run of the same pixel by the pixel value or index along with the frequency.



Dictionary-Based Image Coding (GIF, PNG)

- Dictionary-based coding algorithms, such as LZW, work by substituting shorter codes for longer patterns of pixel values occurring in an image.
- A dictionary of code words and indexes is built depending on the pixel patterns occurring in the image. This technique is used by both GIF and PNG standards.

Outlines

- Overview
- Redundancy and Relevancy of image data
- Classes of image compression techniques
- Lossless Image Coding
  - Image Coding Based on Run Length
  - Dictionary-Based Image Coding (GIF, PNG)
- Lossy Image Coding
  - TRANSFORM IMAGE CODING
    - DCT Image Coding and the JPEG Standard
    - JPEG Bit Stream
    - Drawbacks of JPEG
  - WAVELET BASED CODING (JPEG 2000)
  - JPEG 2000 Versus JPEG

Loosy image compression

- Unlike the error-free compression, lossy encoding is based on the concept of compromising the accuracy of the reconstructed image in exchange for increased compression.
- The lossy compression method produces distortion which is irreversible. On the other hand, very high compression ratios ranging between 10:1 to 50:1 can be achieved with visually indistinguishable from the original. The error-free methods rarely give results more than 3:1.
- Transform coding is the most popular lossy image compression method which operates directly on the pixels of an image.
- The method uses a reversible transform (i.e. Fourier, Cosine transform) to map the image into a set of transform coefficients which are then quantized and coded.
- The **goal** of the transformation is to decorrelate the pixels of a given image block such the most of the information is packed into smallest number of transform coefficients.

TRANSFORM IMAGE CODING

- Transform coding techniques work by performing a mathematical transformation on the input signal that results in a different signal.
- The transformation must be invertible so that the original signal can be recovered.
- The motivation for the use of such techniques is that the transformation changes the signal representation to a domain, which can result in reducing the signal entropy and, hence, the number of bits required to compress the signal.
- Transform coding techniques by themselves are not lossy; however, they frequently employ quantization after a transform, which results in loss of data.

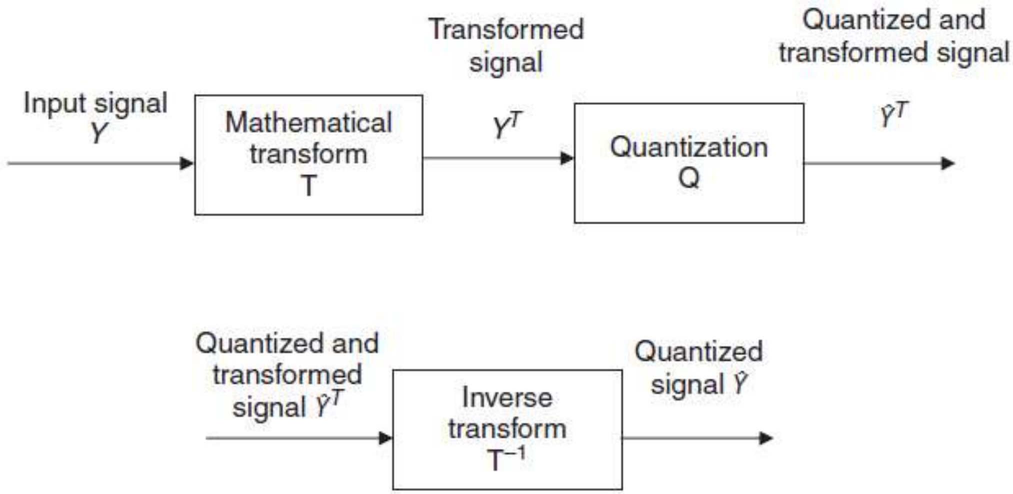
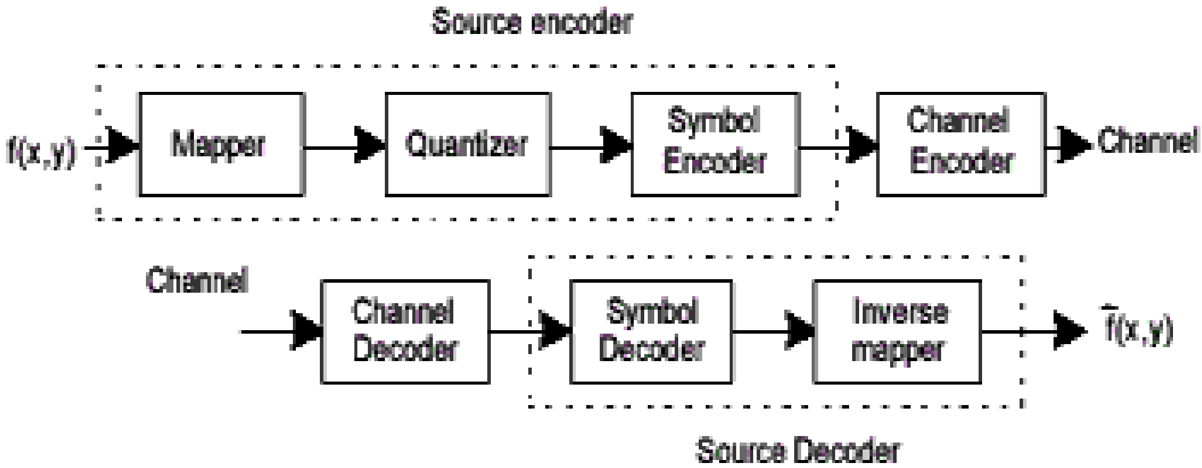


Figure 6-14 Transform coding. The top figure shows a block diagram at the encoder. The input signal Y is transformed to Y^T and then quantized to $\hat{Y}^T$. The bottom figure shows the decoder where $\hat{Y}^T$ is inverse transformed to produce the signal $\hat{Y}$.

TRANSFORM IMAGE CODING

- Image Compression Model



– **Mapper**: transforms the input data into a (usually non visual) format designed to reduce spatial redundancies in the input image.
– **Quantizer**: This process reduces the accuracy and hence irrelevant information of a given image. This process is irreversible and therefore lossy.
– **Symbol Encoder**: This is the source encoding process where fixed or variable-length code is used to represent mapped and quantized data sets.

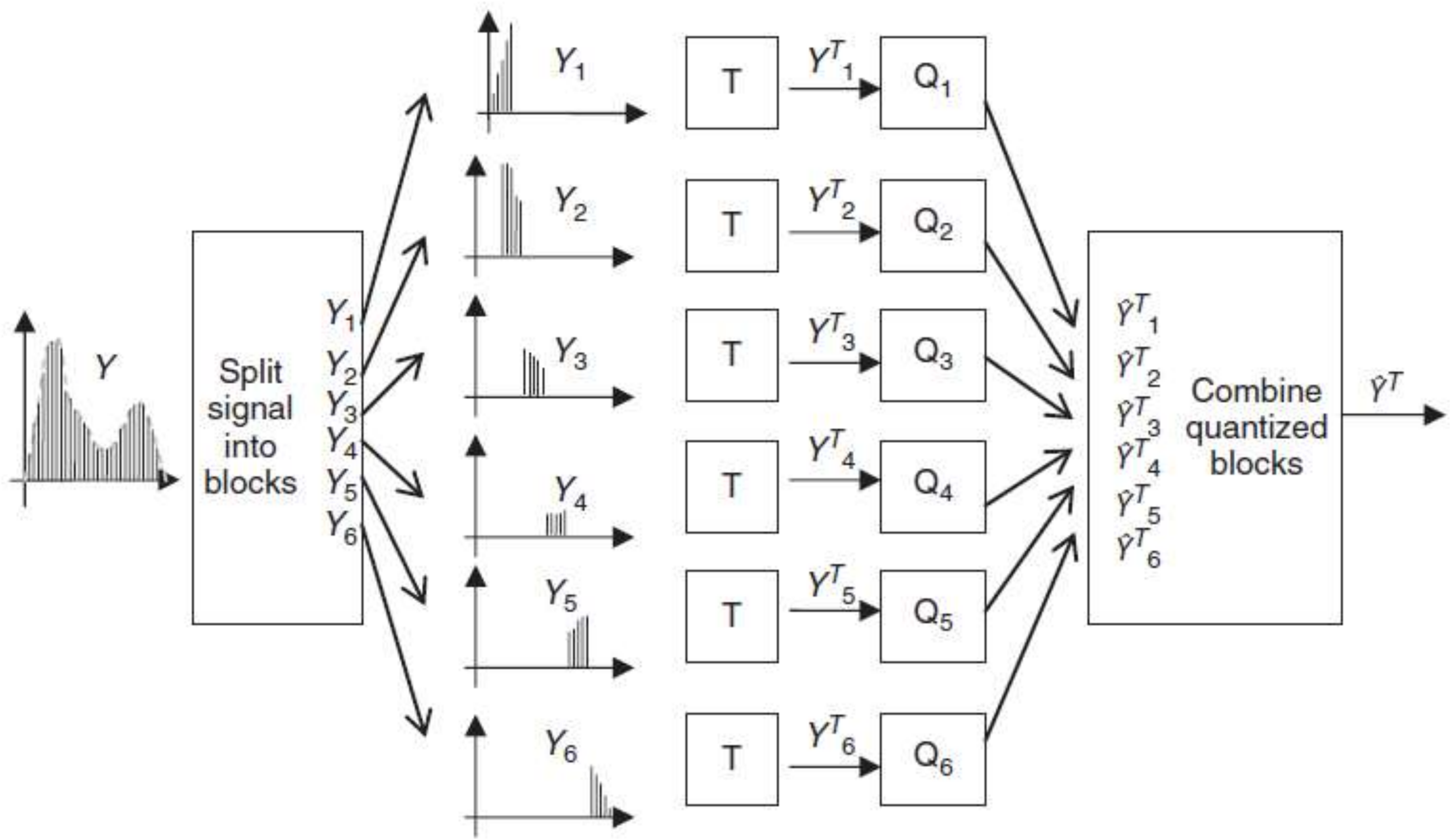
TRANSFORM IMAGE CODING (cont'd)

- Transform techniques are very well suited to both 1D signal (sound) and 2D signal (images) coding, thus a large number of such approaches have been proposed.
- These can be grouped as follows:
  - *Frequency transforms* — Discrete Fourier transforms, Hadamard transforms, Discrete Cosine transforms
  - *Wavelet transforms*—While similar to frequency transforms, these transforms work more efficiently because the input is transformed to a multiresolution frequency representation.
- Compression is obtained by quantizing the coefficients of the basis frequencies, thus reducing the number of bits used by all the coefficients.
- Quantization causes a loss and data cannot be recovered.

Block-transform based coding

- Computationally, it is not feasible to compute transformations for the entire media signal, for example an entire image or a few minutes of audio.
- Most standard techniques use frequency transforms by breaking the signal into blocks and transforming the blocks individually to keep the computation manageable.
- This block-based transformation technique is illustrated in following figure.
  - During encoding, the input signal Y is broken into different blocks $Y_1, Y_2, \dots, Y_n$. Each of the blocks is individually transformed and then quantized.
  - The decoder receives the quantized blocks, which are then inverse transformed to get the reconstructed signal.

Block-transform based coding (cont'd)



Block-transform based coding (cont'd)

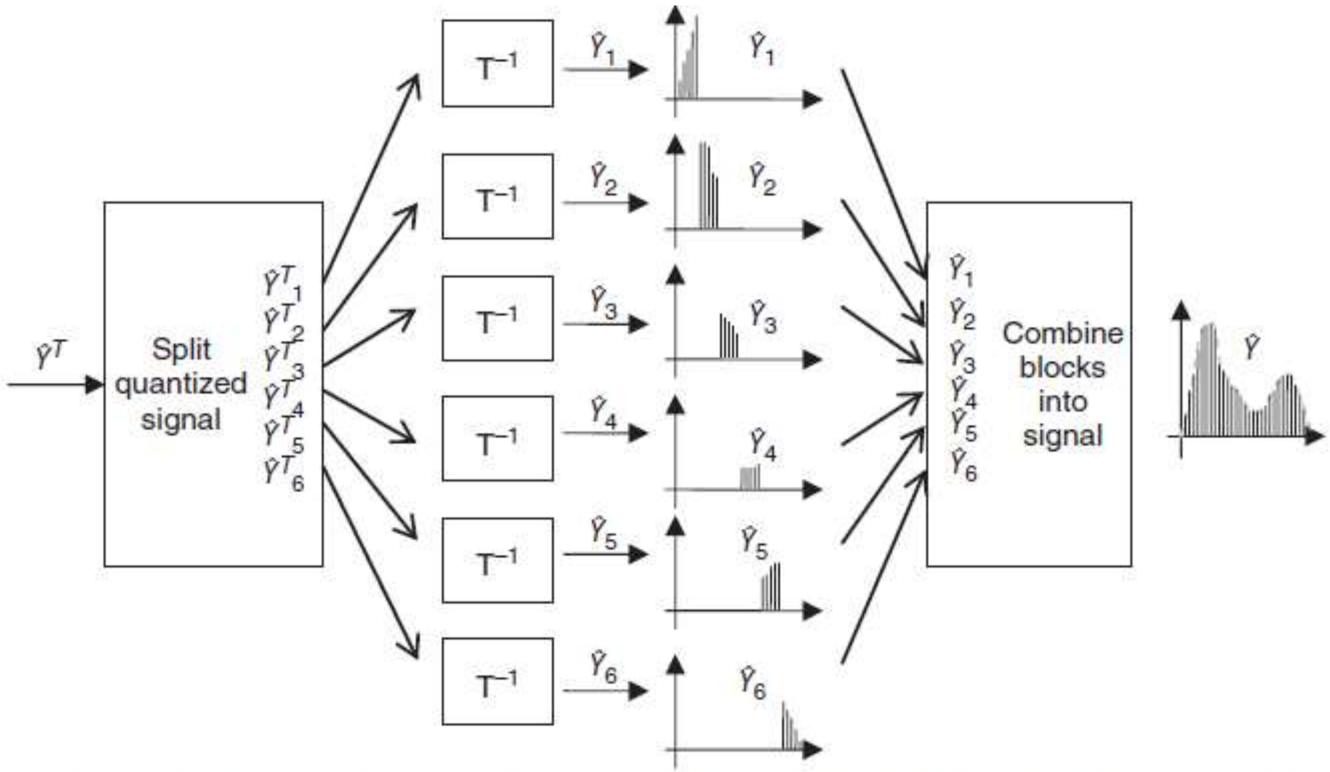


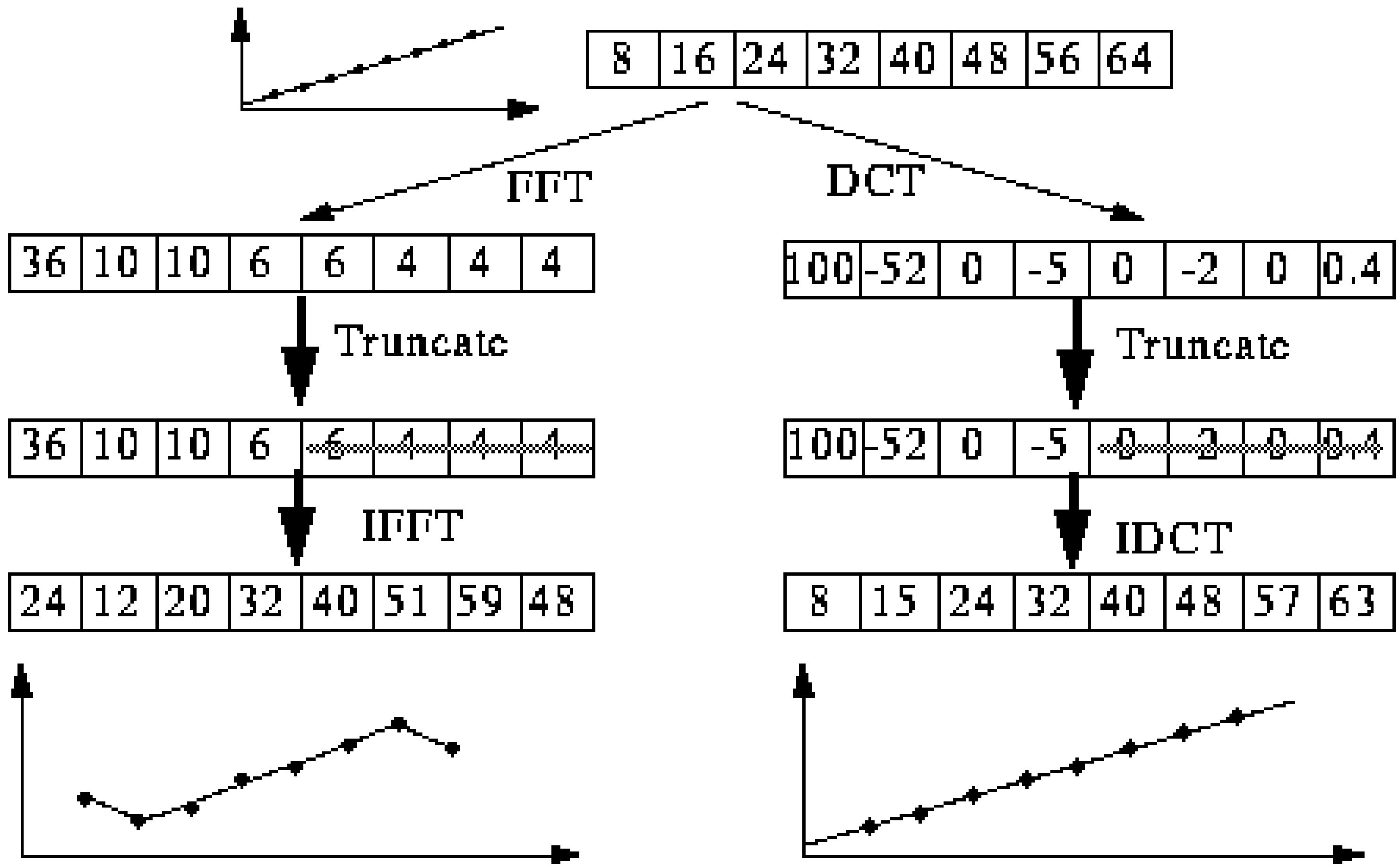
Figure 6-15 Block-based transform coding. The top figure shows the encoding process. The input signal Y is split into blocks Y_1, Y_2 . . . Y_6, each of which are transformed using the same transformation T into Y^T_1, Y^T_2. . . . Y^T_6. Each of these are in turn quantized separately by quantizers Q_1, Q_2 . . . Q_6 and combined to produce the resultant $\hat{Y}^T$. The bottom figure shows the decoder where $\hat{Y}^T$ is split into its constituent blocks and inverse transformed to produce the signal blocks $\hat{Y}_1$, $\hat{Y}_2$ . . . $\hat{Y}_3$, which are then combined to produce the decoded signal $\hat{Y}$.

DCT Image Coding and the JPEG Standard

- The JPEG standard is based on a transform image coding technique that utilizes the Discrete Cosine transform (DCT).
- The DCT was chosen by the JPEG community because it has an information packing ability that is superior to other transforms, and provides a good compromise or balance between information packing ability and computational complexity

DCT Image Coding and the JPEG Standard

- Why DCT not FFT? -- DCT is like FFT, but can approximate linear signals well with few coefficients.



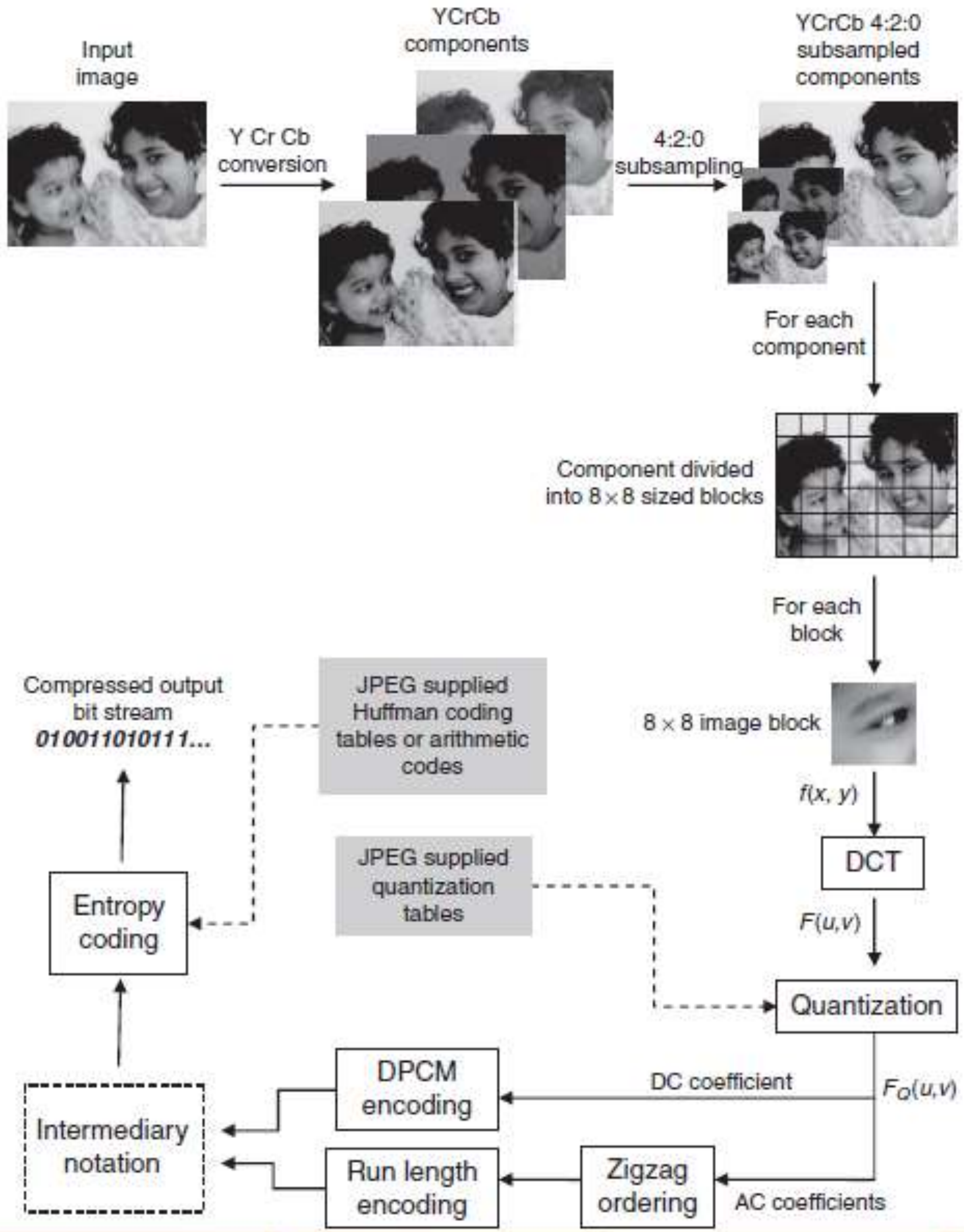
DCT Image Coding and the JPEG Standard

- The entire JPEG encoder pipeline is depicted in following Figure.
  - JPEG compression start by transforming the image representation into a YCrCb component representation. The YCrCb color space is very similar to the YUV color space.
  - Each of the Y, Cr, and Cb components is treated independently by breaking it into 8×8 blocks.
  - In the baseline mode, the blocks are scanned in scan line order from left to right and top to bottom, each one undergoing the DCT computation, quantization of frequency coefficients, and, finally, entropy coding.

Mansoura University

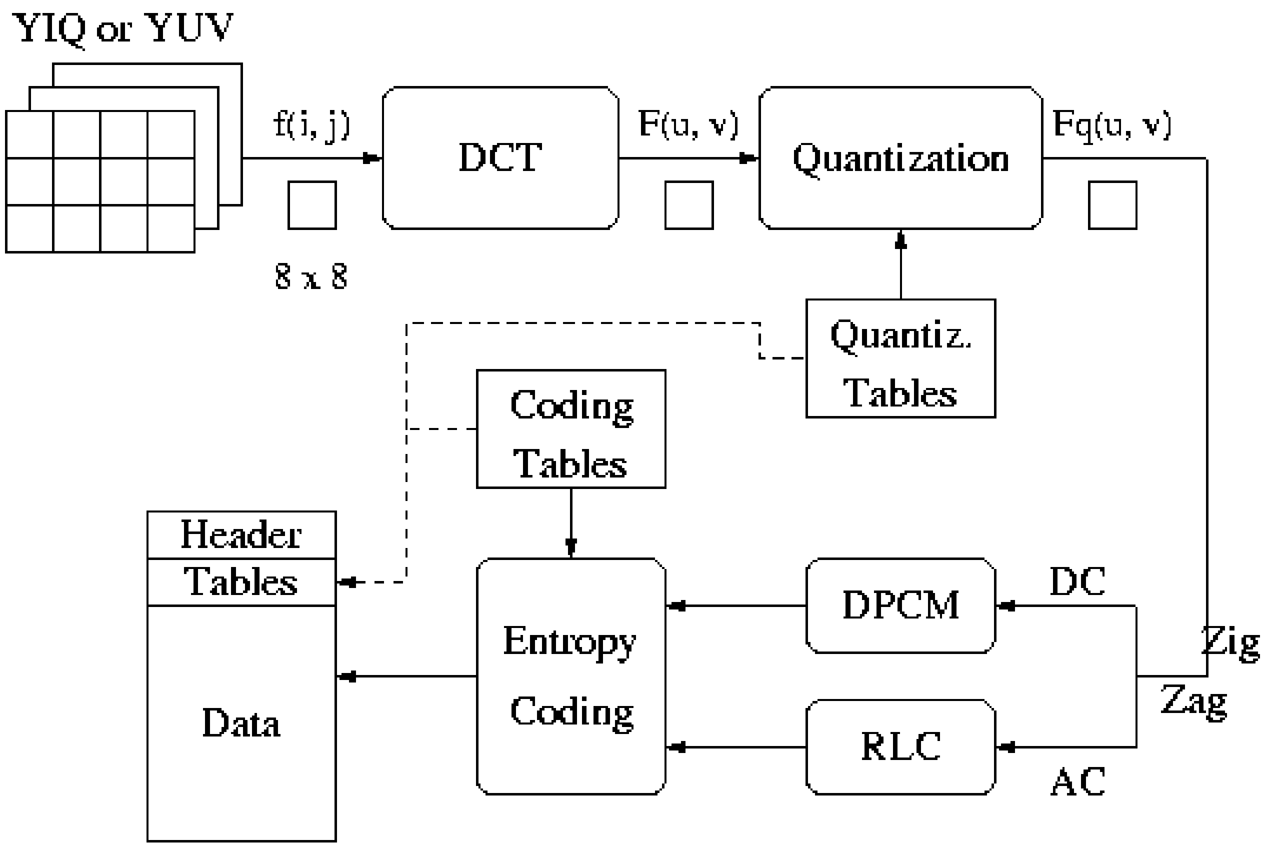
JPEG Standard (cont'd)

The JPEG compression pipeline



JPEG Standard (cont'd)

The JPEG compression pipeline



DCT Image Coding and the JPEG Standard (cont'd)

- The salient features of each step are summarized in the following list:
  - Any image can be taken as input, but is always first converted to the YCrCb format to decouple the image chrominance from the luminance.
  - The YCrCb representation undergoes a 4:2:0 subsampling, where the chrominance channels Cr and Cb are subsampled to one-fourth the original size.
    - The subsampling is based on psychovisual experiments, which suggest that the human visual system is more sensitive to luminance, or intensity, than to color.
  - Next, each channel (Y, Cr, and Cb) is processed independently. Each channel is divided into 8×8 blocks. The JPEG community choose this block size after much experimentation on natural, continuous tone images. On the average, the 8×8 size seems to be the most optimal area for spatial and spectral correlation that the DCT quantization can exploit.
    - If the image width (or height) is not a multiple of 8×8 , the boundary blocks get padded with zeros to attain the required size. The blocks are processed independently.

DCT Image Coding and the JPEG Standard (cont'd)

- The salient features of each step are summarized in the following list:
  - Each 8×8 block (for all the channels) undergoes a DCT transformation, which takes the image samples $f(x, y)$ and computes frequency coefficients $F(u, v)$.
    - The DCT computation of a sample 8×8 block is shown in following Figure. The image $f(x, y)$ and its numerical intensity values are shown on the upper left. The middle section shows the computed DCT transform $F(u, v)$.
    - Of these, the first coefficient, that is the coefficient in the location given by index (0,0), is normally the highest, and is called the DC coefficient.
    - This special status for the DC coefficient is deliberate because most of the energy in natural photographs is concentrated among the lowest frequencies.
    - The remaining coefficients are called AC coefficients.

DCT Image Coding and the JPEG Standard (cont'd)

Figure 7-7: DCT 8 × 8 block example. The arrays show the values of $f(x, y)$, $F(u, v)$, $F_Q(u, v)$, and the decoded $F_Q(u, v)$ and $ff(x, y)$,. The frequency coefficient values are also shown depicted in 3D. The coefficient in the first location given by index (0,0) is normally the highest, whereas higher-frequency coefficients are almost zero.

178	187	183	175	178	177	150	183
191	174	171	182	176	171	170	188
199	153	128	177	171	167	173	183
195	178	158	167	167	165	166	177
190	186	158	155	159	164	158	178
194	184	137	148	157	158	150	173
200	194	148	151	161	155	148	167
200	195	172	159	159	152	156	154

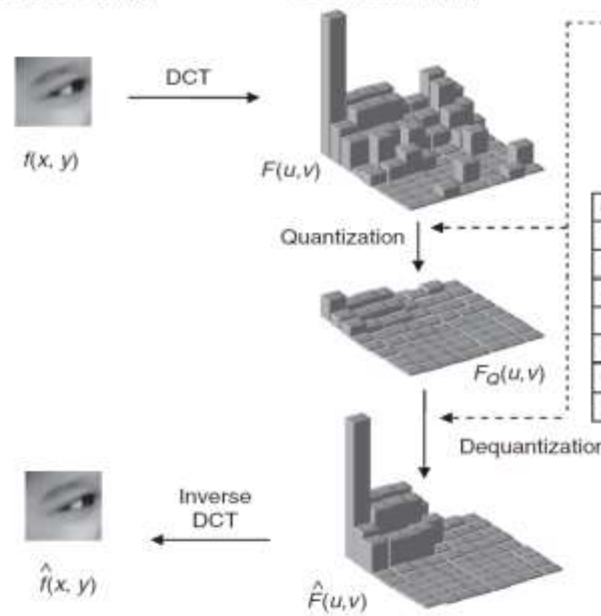
Pixel values $f(x, y)$

1359	46	61	26	38	–21	–5	–18
31	–35	–25	–11	13	10	12	–3
13	20	–17	–14	–11	–7	6	5
–5	5	2	–8	–11	–26	8	–4
10	15	–10	–16	–21	–7	8	7
–6	1	0	7	5	–7	–1	–3
–13	–8	1	10	8	4	–3	–4
–5	–5	–2	5	5	0	0	–3

DCT values $F(u,v)$

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

Quantization table

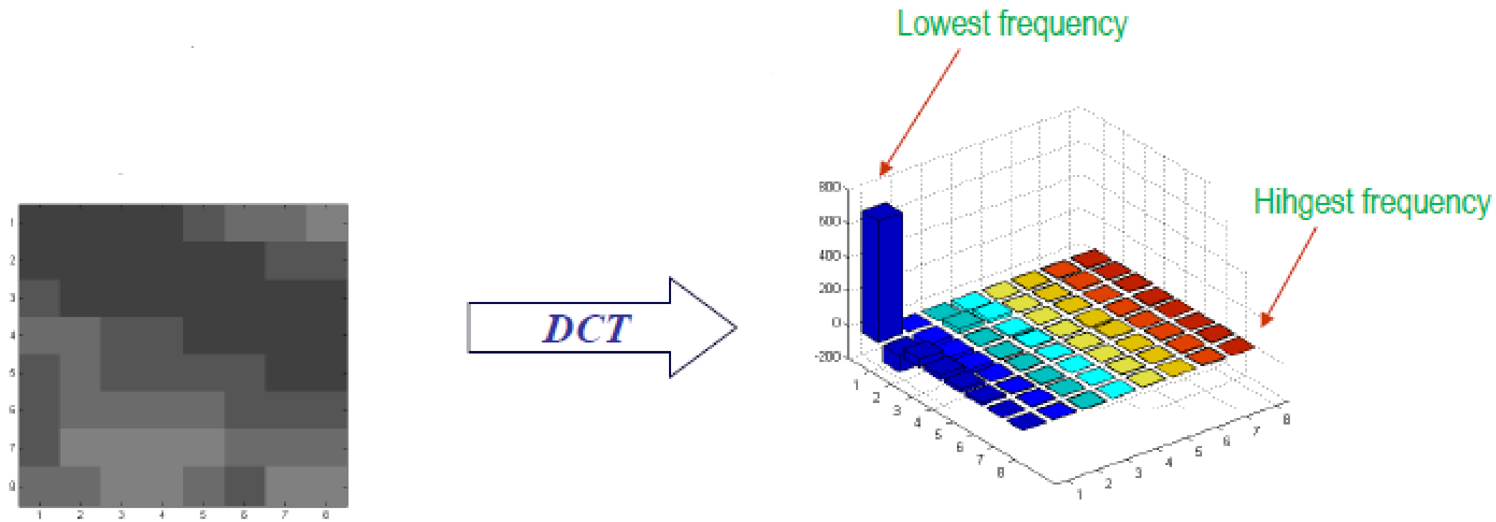


85	4	6	2	2	–1	0	0
3	–3	–2	–1	1	0	0	0
1	2	–1	–1	0	0	0	0
0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

192	185	178	152	193	182	155	190
181	172	162	154	187	164	159	192
183	168	150	158	181	167	182	190
198	177	150	155	177	169	161	182
202	180	148	148	171	167	159	175
193	176	145	141	162	162	156	170
195	184	155	145	159	156	150	164
209	200	170	154	160	153	144	156

1360	44	60	32	48	–40	0	0
36	–36	–28	–19	26	0	0	0
14	26	–16	–24	0	0	0	0
0	0	0	0	0	0	0	0
18	22	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	0	0	0	0	0

DCT Image Coding and the JPEG Standard (cont'd)



69	71	74	76	89	106	111	122
59	70	61	61	68	76	88	94
82	70	77	67	65	63	57	70
97	99	87	83	72	72	68	63
91	105	90	95	85	84	79	75
92	110	101	106	100	94	87	93
89	113	115	124	113	105	100	110
104	110	124	125	107	95	117	116

717.6	0.2	0.4	-19.8	-2.1	-6.2	-5.7	-7.6
-99.0	-35.8	27.4	19.4	-2.6	-3.8	9.0	2.7
51.8	-60.8	3.9	-11.8	1.9	4.1	1.0	6.4
30.0	-25.1	-6.7	6.2	-4.4	-10.7	-4.2	-8.0
22.6	2.7	4.9	3.4	-3.6	8.7	-2.7	0.9
15.6	4.9	-7.0	1.1	2.3	-2.2	6.6	-1.7
0.0	5.9	2.3	0.5	5.8	3.1	8.0	4.8
-0.7	-2.3	-5.2	-1.0	3.6	-0.5	5.1	-0.1

DCT Image Coding and the JPEG Standard (cont'd)

- The salient features of each step are summarized in the following list:
  - Next, the DCT coefficients $F(u, v)$ are quantized using a quantization table supplied by JPEG. Each number at position (u, v) gives the quantization interval size for the corresponding $F(u, v)$ value. The quantization table values might appear random, but, in fact, they are based on experimental evaluations with human subjects, which have shown that low frequencies are dominant in images, and the human visual system is more sensitive to loss in the low-frequency range. Correspondingly, the numbers in the low-frequency area (upper-left corner of the table) are smaller and increase as you move toward the high-frequency coefficients in the other three corners.
  - Using this table, $F_Q(u, v)$ is computed as

$$F_Q(u, v) = \left\lfloor \frac{F(u, v)}{Q(u, v)} \right\rceil$$

where $Q(u, v)$ *is the value in the quantization table*

DCT Image Coding and the JPEG Standard (cont'd)

717.6	0.2	0.4	-19.8	-2.1	-6.2	-5.7	-7.6
-99.0	-35.8	27.4	19.4	-2.6	-3.8	9.0	2.7
51.8	-60.8	3.9	-11.8	1.9	4.1	1.0	6.4
30.0	-25.1	-6.7	6.2	-4.4	-10.7	-4.2	-8.0
22.6	2.7	4.9	3.4	-3.6	8.7	-2.7	0.9
15.6	4.9	-7.0	1.1	2.3	-2.2	6.6	-1.7
0.0	5.9	2.3	0.5	5.8	3.1	8.0	4.8
-0.7	-2.3	-5.2	-1.0	3.6	-0.5	5.1	-0.1

÷

Quantization Matrix

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

=

45	0	0	-1	0	0	0	0
-8	-3	2	1	0	0	0	0
4	-5	0	0	0	0	0	0
2	-1	0	0	0	0	0	0
1	0	0	0	0	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

Quantized 8x8 block

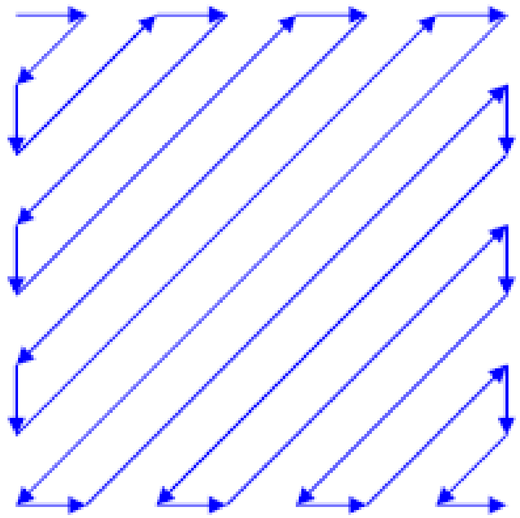
DCT Image Coding and the JPEG Standard (cont'd)

- The salient features of each step are summarized in the following list:
  - After quantization, almost all the high-frequency $F_Q(u,v)$ are zero, while a few low frequency values remain.
    - For evaluation purposes, we also show the decoder process in Figure, which takes $F_Q(u,v)$ and dequantizes the value to produce $\hat{F}(u,v)$. The loss of data in the frequency coefficients should be pretty obvious when you compare $F(u,v)$ and $\hat{F}(u,v)$.
  - The quantized coefficients $F_Q(u,v)$ are then encoded into an intermediary pattern. Here, DC $F_Q(0,0)$, which normally corresponds to the highest energy (lowest-frequency coefficients), is treated differently when compared with the other higher-frequency coefficients, called AC coefficients.
    - The DC coefficients of the blocks are encoded using differential pulse code modulation.
    - The AC coefficients are first scanned in a zigzag order, as shown in Figure. The quantization using the JPEG quantization table produces a lot more zeros in the higher-frequency area. Consequently, the zigzag ordering produces a longer run of zeros towards the end of the scan.
    - This produces a lower entropy of scanned AC coefficients, which are run length encoded.

DCT Image Coding and the JPEG Standard (cont'd)

45	0	0	-1	0	0	0	0
-8	-3	2	1	0	0	0	0
4	-5	0	0	0	0	0	0
2	-1	0	0	0	0	0	0
1	0	0	0	0	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

Input



Zigzag scan procedure

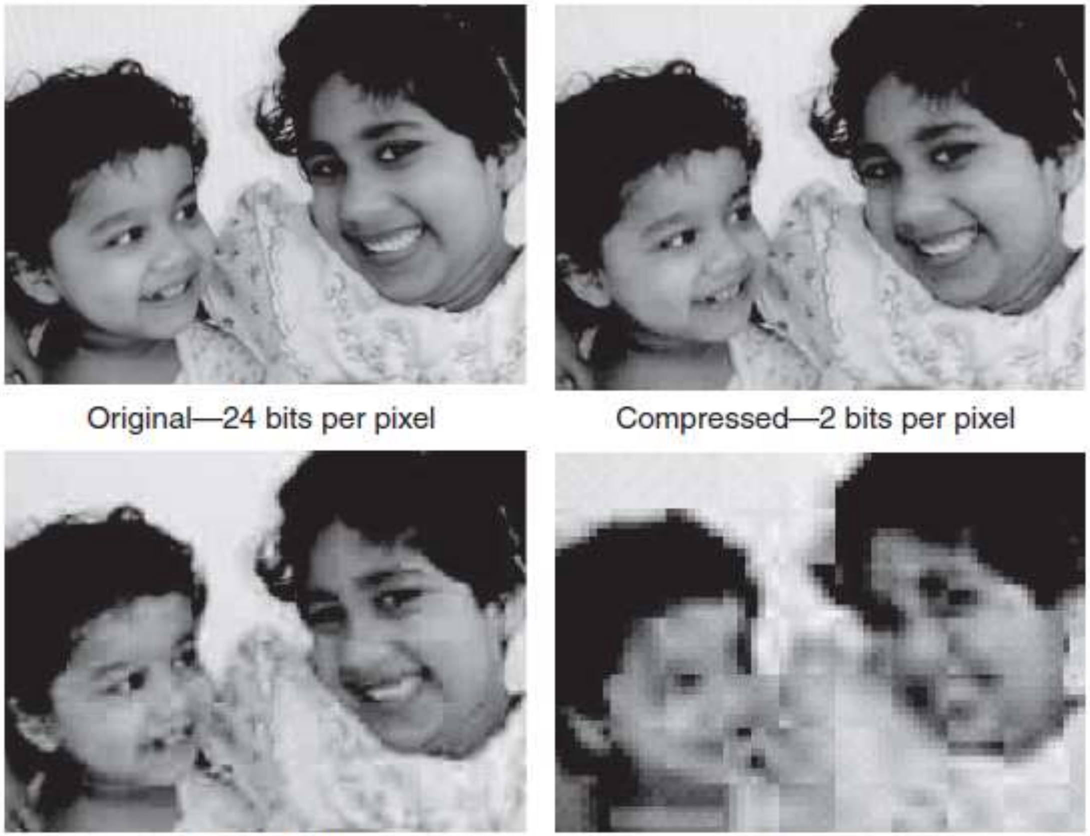
Result = 45,0,-8,4,-3,0,-1,2,-5,2,1,-1,0,1,0,0,0,0,0,0,1,EOB

EOB symbol denotes the end-of-block condition

DCT Image Coding and the JPEG Standard (cont'd)

- The salient features of each step are summarized in the following list:
  - Both the DPCM codes of DC coefficients and the run length coded AC coefficients produce an intermediary representation.
  - The intermediary representations are then entropy coded using codes supplied by the JPEG organization.
    - The first symbol in the intermediary representation for both DC and AC coefficients is encoded using Huffman coding, where the Huffman codes supplied by JPEG are based on statistical analysis of the frequency of these intermediary notations.
- The results of the JPEG algorithm at different compression rates is shown if following figure. You can see that as the compression ratios go down to more than 50:1, the distortion gets unacceptable.

DCT Image Coding and the JPEG Standard (cont'd)



Original—24 bits per pixel
Compressed—2 bits per pixel
Compressed—0.5 bits per pixel
Compressed—0.15 bits per pixel

Figure 7-10 JPEG compression results. The upper-left image shows the original at 24 bits per pixel. The remaining three images show the reconstructed outputs after JPEG compression at different compression ratios. The blocky artifacts increase at lower bit rates.

JPEG Bit Stream

- In JPEG algorithm, the image is divided into blocks and each block is encoded in a scan line manner. The resultant compressed bit representations of all the scanned blocks are packed and organized to form a bit stream as per the JPEG specification standard.
- Figure 7-11 shows a hierarchical view of this standardized representation.
  - The first level shows a header and tail to represent the start and end of the image.
  - The image here is termed as a *frame*. Each frame is organized as a combination of *scans* or components of the frame.
  - Each scan is further organized as a group of *segments*.
  - A segment is a series of 8×8 blocks.

JPEG Bit Stream

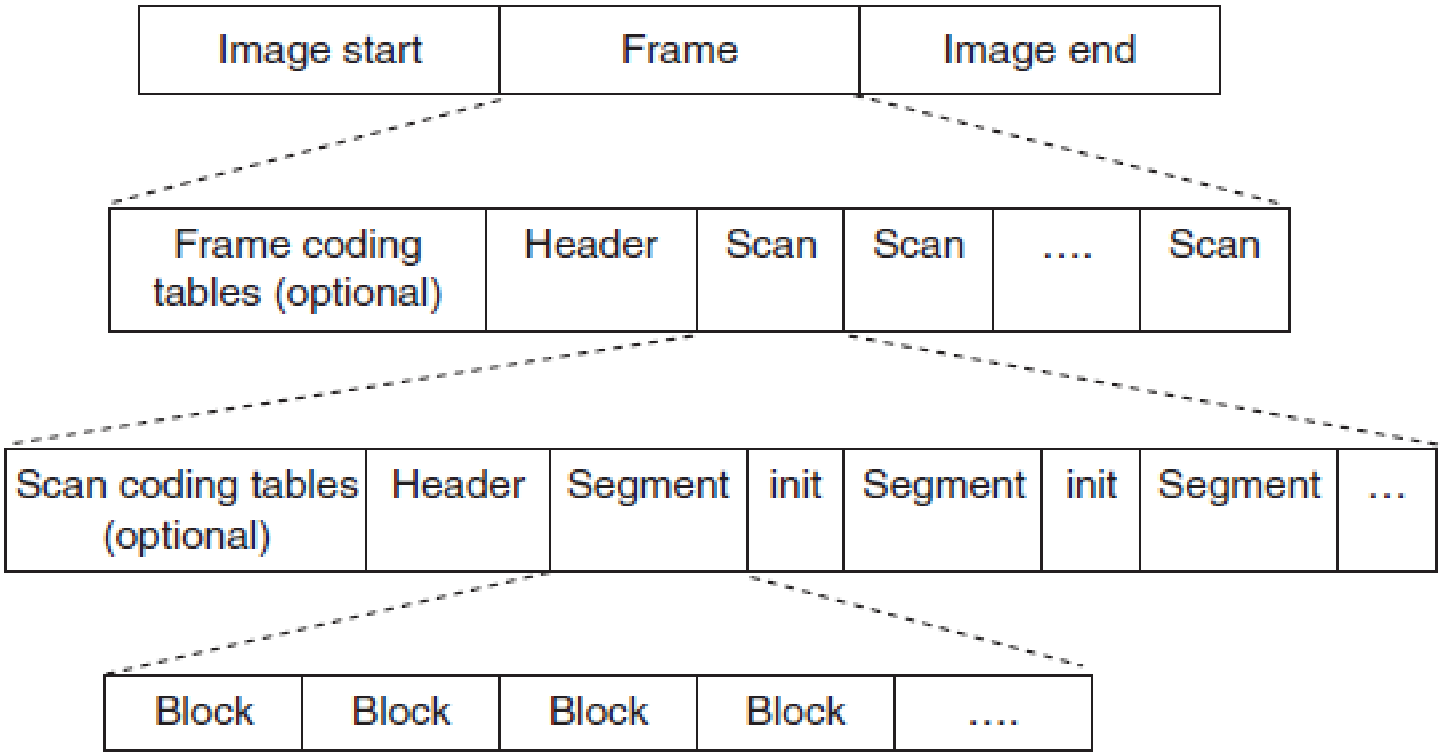


Figure 7-11 Hierarchical representation of the JPEG bit stream.

JPEG Compression Modes

- **JPEG defines four modes of operation**
  1. **Sequential Lossless Mode:** Compress the image in a single scan and the decoded image is an exact replica of the original image.
  2. **Sequential DCT-Based Mode:** Compress the image in a single scan using DCT-based lossy compression technique. As a result, the decoded image is not an exact replica, but an approximation of the original image.
  3. **Progressive DCT-based Mode:** Compress the image in multiple scans and also decompress the image in multiple scans with each successive scan producing a better-quality image.
  4. **Hierarchal Mode:** compress the image at multiple resolutions, so that the lower resolution of the image can be accessed first without decompressing the whole resolution of the image, for display on different devices.
- The last three DCT-based modes (2, 3, and 4) are lossy compression because precision limitation to compute DCT and the quantization process introduce distortion in the reconstructed image.
- The lossless mode uses predictive method and does not have quantization process.

JPEG Compression Modes (cont’d)

Sequential

Progressive

Drawbacks of JPEG

- ***Poor low bit-rate compression***—JPEG offers excellent rate-distortion performance in the mid and high bit rates, but at low bit rates, the perceived distortion becomes unacceptable.
- ***Lossy and lossless compression***—There is currently no standard that can provide superior lossless and lossy compression in a single coded stream.
- ***Single compression architecture***—The current JPEG standard has about 44 modes. Many of these are application specific and not used by the majority of decoders.
- ***Transmission in noisy environments***—JPEG was created before wireless communications became an everyday reality; therefore, it does not acceptably handle such an error-prone channel.
- ***Computer-generated images and documents***—JPEG was optimized for natural images and does not perform well on computer-generated images and document imagery. This is because JPEG is well suited to continuous tone imagery but not constant tone or slow changing tone imagery.

WAVELET BASED CODING (JPEG 2000)

- Consequently, the ISO community set forth a new image coding standard for different types of still images (bilevel, grayscale, color, multicomponent), with different characteristics (natural, scientific, remote sensing, text rendered graphics, compound, etc.), allowing different imaging models (client/server, real-time transmission, image library archival, limited buffer and bandwidth resources, etc.) preferably within a unified and integrated system.
- This is the JPEG 2000 standard, which is based on the wavelet transform, as explained in the next section.

Outlines

- Overview
- Redundancy and Relevancy of image data
- Classes of image compression techniques
- Lossless Image Coding
  - Image Coding Based on Run Length
  - Dictionary-Based Image Coding (GIF, PNG)
- Lossy Image Coding
  - TRANSFORM IMAGE CODING
    - DCT Image Coding and the JPEG Standard
    - JPEG Bit Stream
    - Drawbacks of JPEG
  - WAVELET BASED CODING (JPEG 2000)
  - JPEG 2000 Versus JPEG

WAVELET BASED CODING (JPEG 2000)

- The JPEG 2000 compression pipeline makes use of the Discrete Wavelet transform (DWT) to compress images. The data flow of JPEG 2000 is similar to transform coding flow.

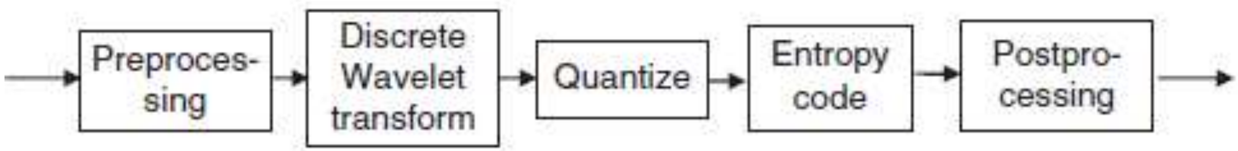


Figure 7-12 The JPEG 2000 pipeline

- Central to the whole encoding process in JPEG 2000 is the Discrete Wavelet transform, which is known to work better than the Discrete Cosine transform in the way it distributes energy among the frequency coefficients.
- In JPEG, the DCT works on individual 8×8 blocks, which results in “blockiness” at higher compression ratios.
- In contrast, the DWT in JPEG 2000 converts the whole image into a series of wavelets, which can be stored more efficiently than 8×8 pixel blocks.
- We now discuss each of the subprocesses in the pipeline.

The Preprocessing Step

- The preprocessing stage is responsible for tiling, conversion into the YCrCb formats, and level offsetting.
- The preprocessing steps, all of which are not compulsory, are performed prior to applying the Discrete Wavelet transform to the image.
  - The *tiling process* partitions the image into rectangular, but equal-sized and nonoverlapping blocks. Each tile is then independently processed for DWT analysis, quantization, entropy coding, and so on. The tiling process is purely optional and is done only to reduce memory requirements, as needed to deal with very large images. Additionally, because each tile is encoded independently, they can also be decoded independently. The tiling process is shown in Figure.

Figure 7-13 Sample tiling on an image. The tiles are square with nonoverlapping areas. The tiles at the boundary might not contain all the image data. In such cases, the out-of-range area is padded with zeros.

The Preprocessing Step (cont'd)

- The preprocessing stage is responsible for tiling, conversion into the YCrCb formats, and level offsetting.
- The preprocessing steps, all of which are not compulsory, are performed prior to applying the Discrete Wavelet transform to the image.
  - The YCrCb conversion process is similar to the JPEG pipeline. It is mainly done to take advantage of the human visual system's diminished tolerance to chrominance when compared with luminance. Each of the Y, Cr, and Cb channels are independently processed afterward, with different tolerances used for Y, Cr and Cb.
  - The level offsetting process refers to shifting the DC levels. Before the DWT can be applied to the image (or tiles), the image (or tiles) is DC level shifted by subtracting a constant value from each pixel value. This is done primarily for the DWT to operate properly, which requires that the dynamic range of the pixel values be centered about zero.

The Discrete Wavelet Transform

- JPEG 2000 uses the Discrete Wavelet transform (DWT) to decompose the incoming image signal (or an individual tile) into its high and low subbands at every stage.
- One such stage is shown in Figure 7-14, where the input signal is passed through a lowpass filter and a high-pass filter. This results in a low-band output and a high-band output.
- The low-pass output has the same number of samples as the input image, but the image signal does not contain any high frequency.

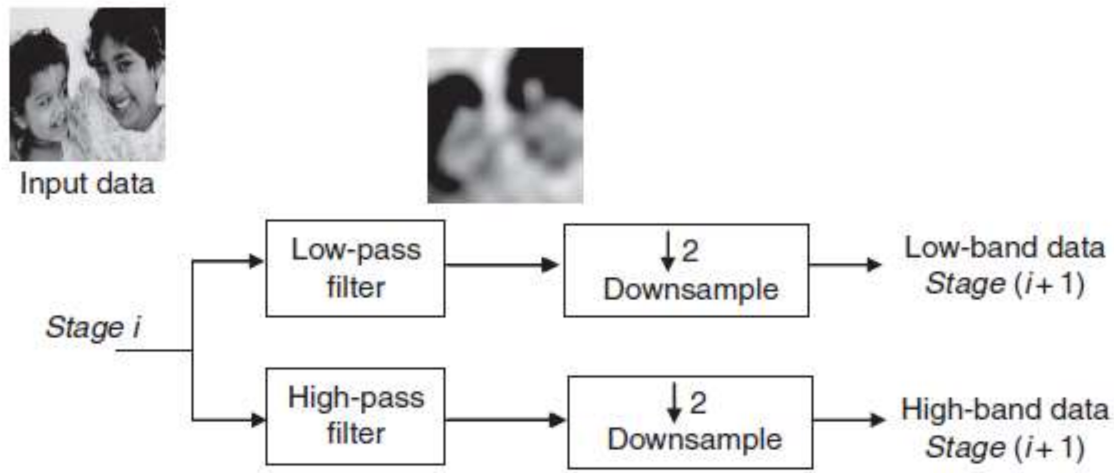


Figure 7-14 The basic Discrete Wavelet structure. Any input data at stage i is passed through a high- and low-frequency filter to produce two outputs. Each contains the same data samples as the input but has either the low or high frequencies removed. The resulting filter outputs are then subsampled to reduce the number of samples by a factor of 2. The combined samples of the outputs are now equal to the number of samples at the input.

The Discrete Wavelet Transform (cont'd)

- JPEG 2000 uses the Discrete Wavelet transform (DWT) to decompose the incoming image signal (or an individual tile) into its high and low subbands at every stage.
- Similarly, the high-pass output also contains the same number of samples as the input, but without any low-frequency content. Thus, the number of samples at each stage gets doubled and, therefore, the output of each filter is downsampled by a factor of 2 to keep the total number of samples at each stage's output the same as the input.

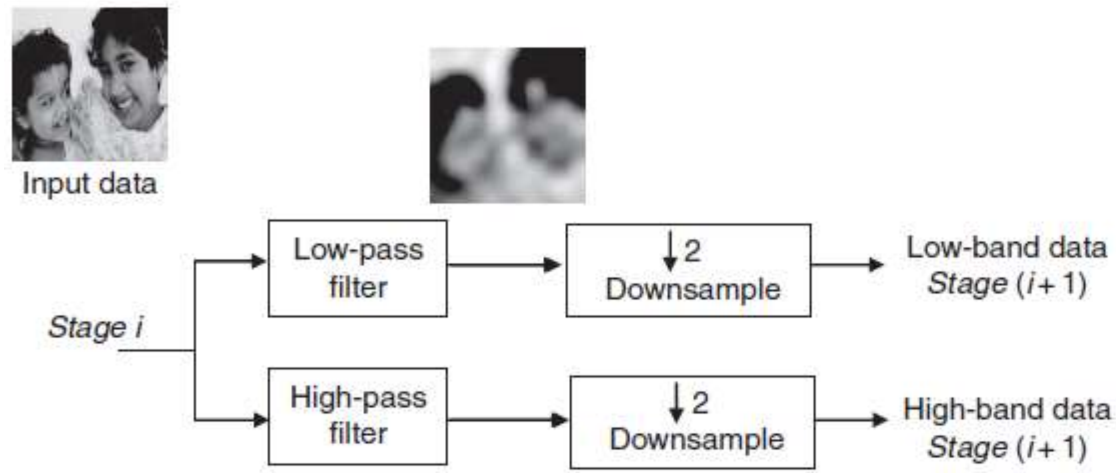


Figure 7-14 The basic Discrete Wavelet structure. Any input data at stage i is passed through a high- and low-frequency filter to produce two outputs. Each contains the same data samples as the input but has either the low or high frequencies removed. The resulting filter outputs are then subsampled to reduce the number of samples by a factor of 2. The combined samples of the outputs are now equal to the number of samples at the input.

The Discrete Wavelet Transform (cont'd)

- Although Figure 7-14 shows the general process on two-dimensional data, all the wavelet transforms used in JPEG 2000 are applied in one dimension only.
- A two-dimensional transform is achieved by applying one-dimensional transforms along two orthogonal dimensions.
- This is better depicted in Figure 7-15, where first a row 1D wavelet transform is applied and the output subsampled by 2, and then the output of the first 1D transform stage undergoes a column 1D wavelet transform to create four bands as shown.
- All the channels are independently processed, but we show the output on the luminance channel only.

The Discrete Wavelet Transform (cont'd)

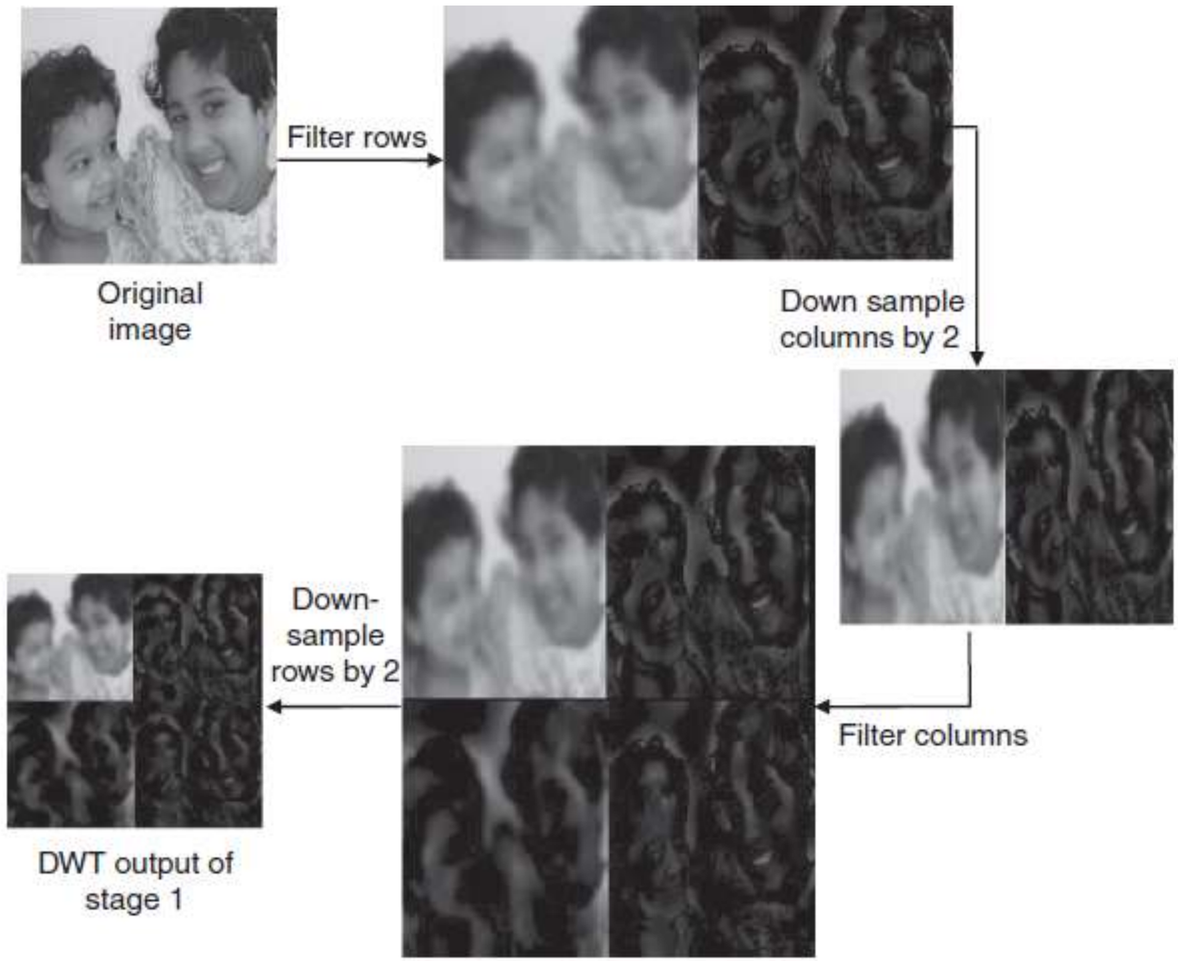


Figure 7-15 DWT process for the Y component of the sample image used.

The Discrete Wavelet Transform (cont'd)

- In JPEG 2000, every original input image signal goes through multiple stages of the DWT. The number of stages performed depends on the implementation. Two stages are demonstrated in Figure 7-16; although all the Y, Cr, and Cb are independently processed, here we show the output together. The first stages output has four quadrants:
  - *LL*—Low subbands of the filtering in both dimensions, rows and columns
  - *HL*—High subbands after row filtering and low subbands after column filtering
  - *LH*—Low subbands after row filtering and high subbands after column filtering
  - *HH*—High subbands after both row and column filtering
- The second stage repeats the same process with the LL subband output of the previous stage. The higher subbands (HL, LH, and HH) hardly contain any significant samples and, therefore, only the LL subband is further transformed.
- Although JPEG 2000 supports from 0 to 32 stages, usually 4 to 8 stages are used for natural images.

The Discrete Wavelet Transform (cont’d)

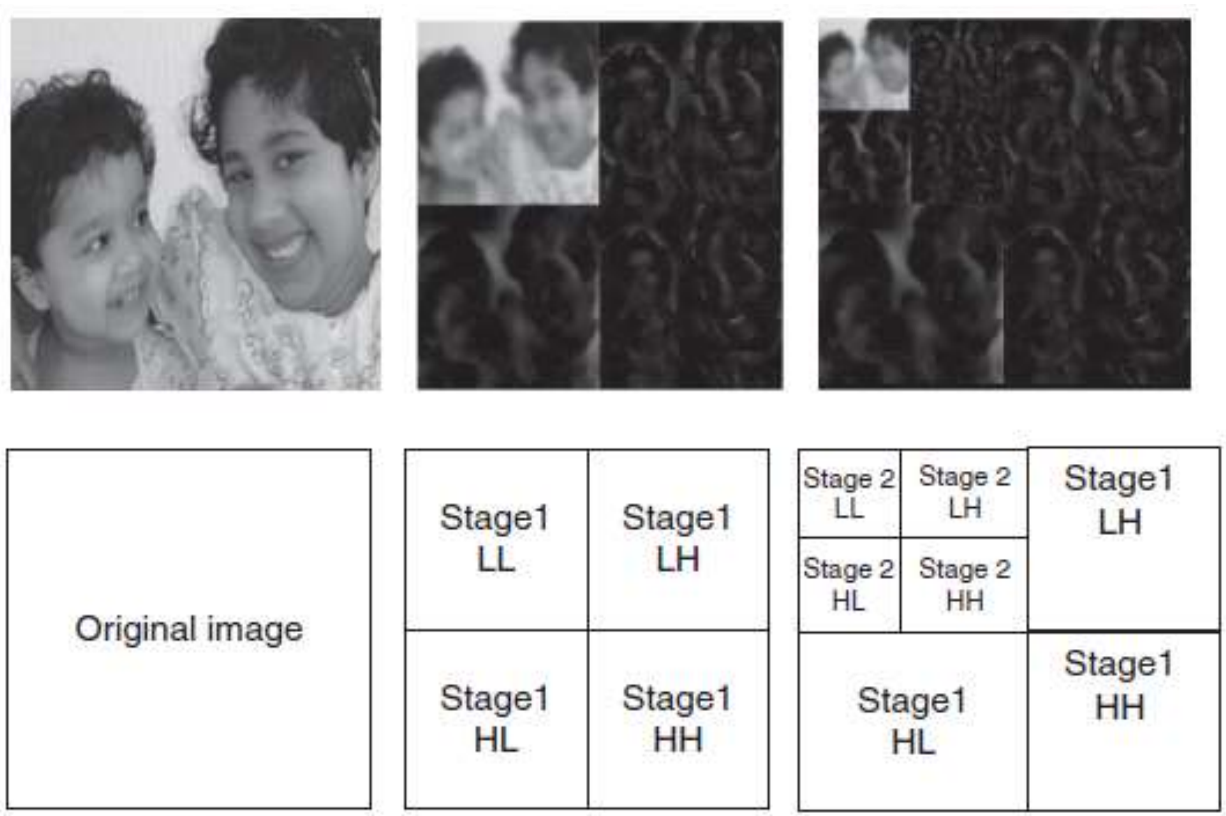


Figure 7-16 Original input and the output of discrete wavelet processing at the first two levels. The top row shows the imaged outputs. The bottom row shows what each block at a level contains.

The Discrete Wavelet Transform (cont’d)

- Each component (or the tiles in each component) is decomposed using the DWT into a series of decomposition levels. Each level contains a number of subbands. These subbands contain frequency coefficients that describe the horizontal and vertical frequency characteristics of the original input.
- These coefficients are quantized to reduce precision, which causes a loss.
- The quantized coefficients are next reorganized into packets where each packet contains the quantized coefficients at a particular resolution. This results in the first packets containing low-quality approximations of the image, with each successive packet containing frequency coefficient information that improves the quality of the decoded output.
- The packets are organized into blocks and entropy coding is performed on each block independently.

Outlines

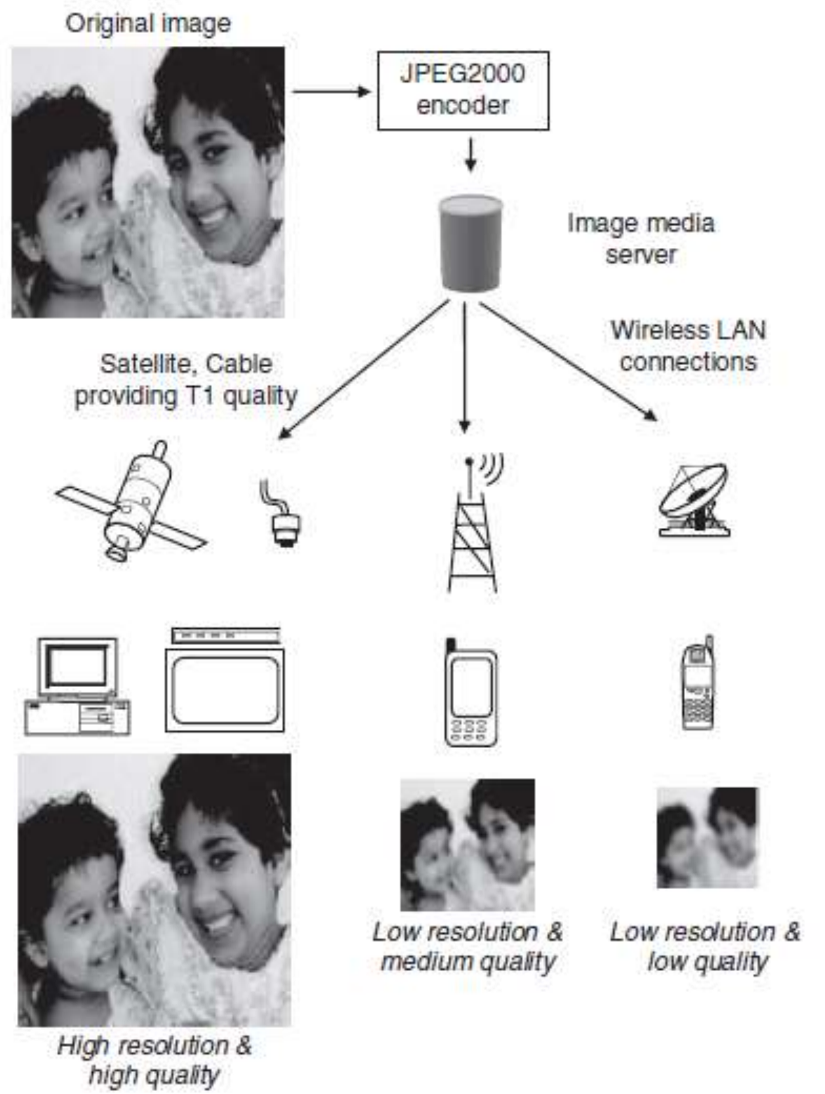
- Overview
- Redundancy and Relevancy of image data
- Classes of image compression techniques
- Lossless Image Coding
  - Image Coding Based on Run Length
  - Dictionary-Based Image Coding (GIF, PNG)
- Lossy Image Coding
  - TRANSFORM IMAGE CODING
    - DCT Image Coding and the JPEG Standard
    - JPEG Bit Stream
    - Drawbacks of JPEG
  - WAVELET BASED CODING (JPEG 2000)
  - JPEG 2000 Versus JPEG

JPEG 2000 Versus JPEG

- The DWT used in JPEG 2000 is more state of the art and provides a better compression ratio when compared with the DCT used in JPEG. Besides that, the organization of the compressed bit stream in JPEG 2000, which included bit plane coding, packetization, and organization of packets into blocks, provides far more powerful features that had not been present in other standards prior to it. Examples of these features include the following:
  - *Encode once—platform-dependent decoding*—In JPEG 2000, the encoder decides the maximum resolution and image quality to be produced—from the highly compressed to completely lossless. Because of the hierarchical organization of each band's coefficients, any image quality can be decompressed from the resulting bit stream. Additionally, the organization of the bit stream also makes it possible to perform random access by decompressing a desired part of the image or even a specific image component. Unlike JPEG, it is not necessary to decode and decompress the entire bit stream to display section(s) of an image or the image itself at various resolutions. This is powerful for practical scenarios, as shown in Figure 7-17. The figure depicts how you can encode an image once with JPEG 2000 and decode it at various frequency resolutions and spatial resolutions depending on the bandwidth availability or the display platform.

JPEG 2000 Versus JPEG (cont’d)

Figure 7-17 A practical application of JPEG2000. The image is encoded once using the DWT, but various end terminals can receive different resolutions from the same bit stream.



JPEG 2000 Versus JPEG (cont'd)

- *Working with compressed images*—Normally, imaging operations such as simple geometrical transformations (cropping, flipping, rotating, and so on) and filtering (using the frequency domain) are performed on the pixel domain representation of the image. If the image is compressed, it is decompressed, and recompressed after the required operation. However, with JPEG 2000, such operations can be applied to the compressed representation of the image.
- *Region-of-interest encoding*—Region-of-interest (ROI) coding pertains to coding a specific region with higher quality compared with the rest of the image. An example of this is shown in Figure 7-18. This process can be predefined, or can change dynamically. Predefined ROIs are normally set at encoding time, whereas dynamic ROIs are used in applications where specific regions of the image are chosen for decoding at finer resolutions compared with other parts of the image. JPEG 2000 elegantly allows for ROI control by specifying a parameter known as an ROI mask. The ROI mask informs the encoder (if static selection) or the decoder (if dynamic selection) about the range of wavelet coefficients that contribute to a region's reconstruction. These coefficients can be decoded first, before all the background is decoded.

JPEG 2000 Versus JPEG (cont'd)

Figure 7-18 Region of interest (ROI). Using JPEG 2000 ROI coding, specific regions can be encoded or decoded with higher resolutions compared with the rest of the image.

Readings

- Readings
  - Multimedia Systems: Algorithms, Standards, and Industry Practices, Parag Havaldar and Gérard Medioni
    - Chapter 7

Thank You