01076566 Multimedia Systems

Chapter 7: Media Compression: Images

Pakorn Watanachaturaporn
pakorn.wa@KMITL.ac.th

Bachelor Program in Computer Engineering (B.Eng.)
Faculty of Engineering

King Mongkut's Institute of Technology Ladkrabang

พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

Outline

- Redundancy and Relevancy of Image Data
- Classes of Image Compression Techniques
- Lossless Image Coding
- Transform Image Coding
- Wavelet Based Coding (JPEG 2000)
- Transmission Issues in Compressed Image

pakorn.wa@kmitl.ac.th

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

Redundancy and Relevancy of Image Data

pakorn.wa@kmitl.ac.th

- Image compression techniques can be purely lossless or lossy
- Good image compression techniques often work as hybrid schemes, making efficient use of lossy and lossless algorithms to compress image data.

pakorn.wa@kmitl.ac.th

- Get compression by typically analyzing the image data according to two important aspects:
  - Irrelevancy reduction
    - Visual irrelevancy
    - Application-specific irrelevancy
  - Redundancy reduction
    - The image 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.

pakorn.wa@kmitl.ac.th

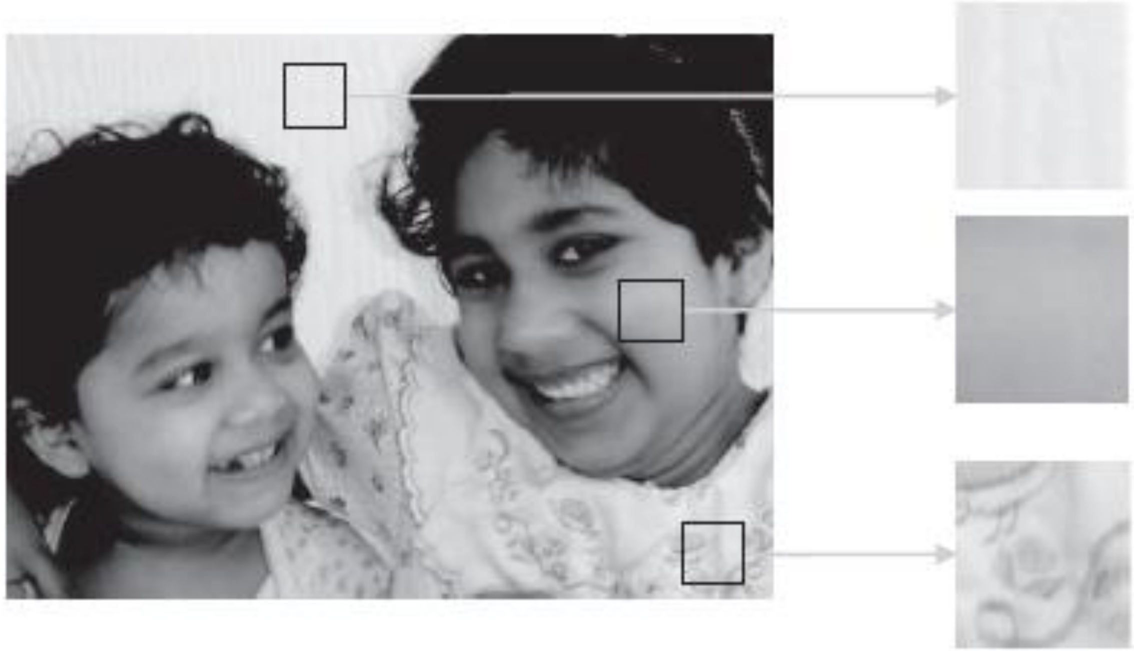


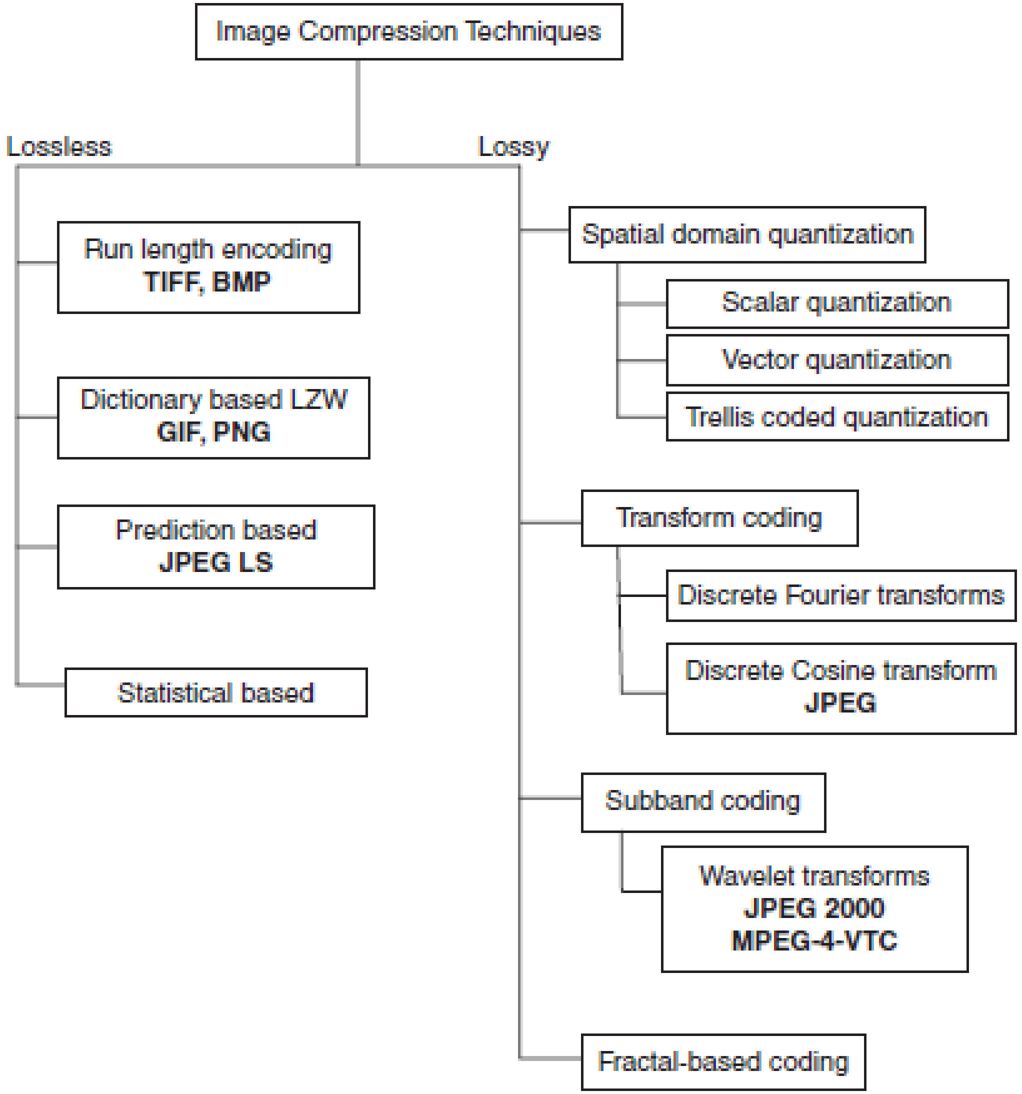
Figure 7-2 Local pixel correlation. Images are not a random collection of pixels, but exhibit a similar structure in local neighborhoods. Three magnified local areas are shown on the right.

pakorn.wa@kmitl.ac.th

- 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.
- Another kind of redundancy— ***temporal redundancy*** —which occurs when a sequence of images, such as video, is captured.

Classes of Image Compression Techniques

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th



Lossless Image Coding (BMP, TIFF, RLA, PICT)

pakorn.wa@kmitl.ac.th

- **Image Coding Based on Run Length**
  - One of the earliest lossless coding schemes that was applied to images
  - This method simply replaces a run of the same pixel by the pixel value or index along with the frequency.
  - Although better schemes than BMP and RLA exist, this is still used sometimes for compatibility of formats between various imaging software.
  - TIFF (Tagged Image File Format) was designed to be extensible by supplying the compression scheme used in the header, so TIFF can be adjusted to make use of RLE or any other supported compression algorithm.

Dictionary-Based Image Coding (GIF, PNG)

KMITL
พระจอมเกล้าลาดกระบัง

pakorn.wa@kmitl.ac.th

- 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.
- Used by both GIF and PNG standards.
- Instead of using the LZW algorithm on the initial pixels in the image, it is performed on an indexed set of colors.

- When compressing using GIF, you normally supply a palette size, for example, 8 bits, which corresponds to a maximal number of 256 colors.
- The best 256 colors are then chosen as the palette colors and the image pixel values are replaced by indexes that best represent the original pixel values from the palette.
- The LZW algorithm is then applied to the indexes.
- The GIF format was initially developed by the UNISYS Corporation and was one of first compression formats that made image transfer on the Internet possible.

Prediction-Based Coding

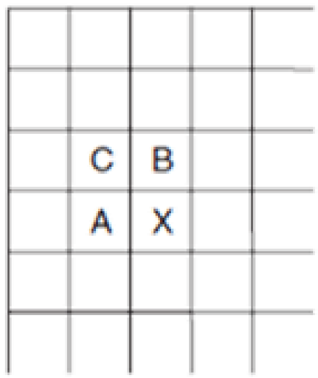
KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- Standard prediction techniques such as DPCM or its variants modified to 2D can be used to predict pixel values in the neighborhood.
- The errors between the actual values and their predictions can then be entropy coded to achieve compression.
- The lossless mode of JPEG works by using a predictive scheme that exploits redundancy in two dimensions.

pakorn.wa@kmitl.ac.th

- Given a pixel value X, a prediction mechanism based on the previously used neighborhood pixels is used to predict the value of X.
- The difference $D(i,j)$ between the original and predicted value is computed.
- Repeating this process for all the pixels results in an *error image*, which has lower entropy than the original image and can be effectively entropy coded.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

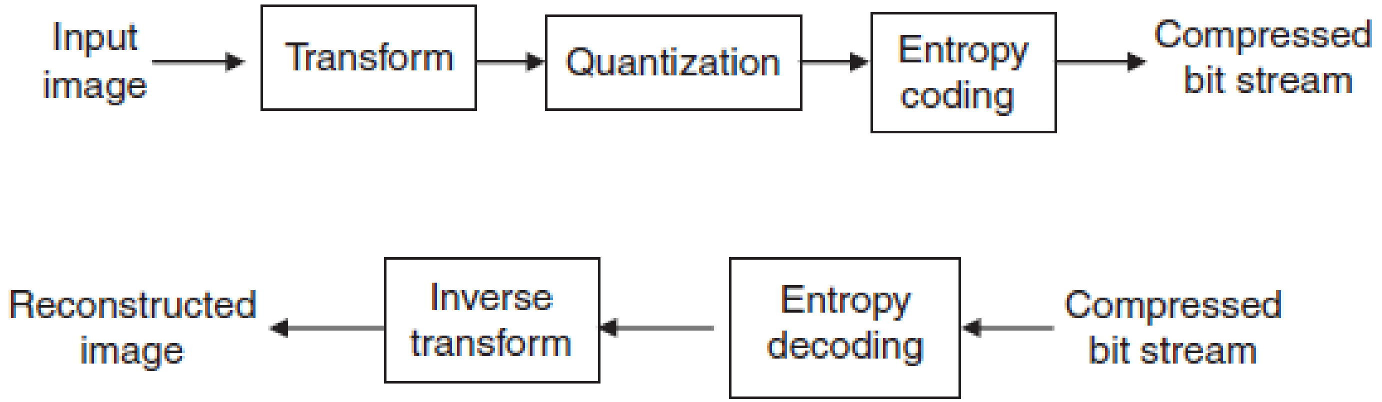


Prediction index	Prediction
0	No prediction
1	A
2	B
3	C
4	A + B − C
5	A + ((B − C)/2)
6	B + ((A − C)/2)
7	(A + B)/2

- The process is illustrated in Figure 7-4, where a pixel with value *X* is shown, and the predicted value obtained from the neighbors *A*, *B*, and *C* depends on a prediction rule, which might be one-dimensional or two-dimensional.

Transform Image Coding

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th



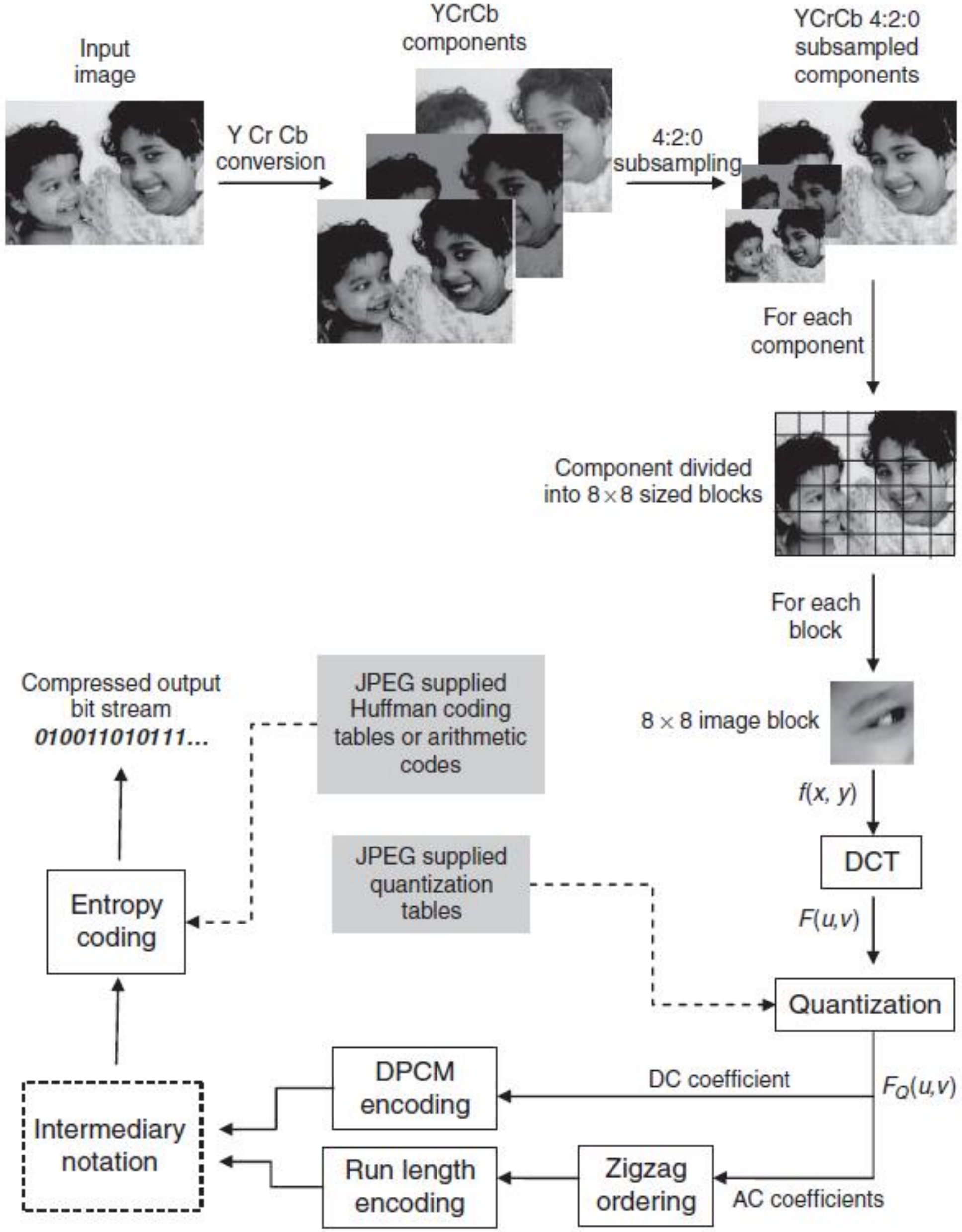
- The encoder proceeds in three steps: transform computation, quantization of the transform coefficients, and entropy coding of the quantized values.
- The decoder only has two steps: the entropy decoding step that regenerates the quantized transform coefficients and the inverse transform step to reconstruct the image.
- The quantization step present in the encoder is the main cause of the loss.

DCT Image Coding and the JPEG Standard

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- 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 of its good frequency domain energy distribution for natural images, as well as its ease of adoption for efficient hardware-based computations.

พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th



pakorn.wa@kmitl.ac.th

- 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.
- 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.

pakorn.wa@kmitl.ac.th

- Each 8 × 8 block (for all the channels) undergoes a DCT transformation,
  - which takes the image samples $f(x,y)$ and computes frequency coefficients $F(u,v)$.
  - Although the values are real numbers, the table shows the DCT values rounded to the nearest integer for compactness in depiction.
  - 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.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

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

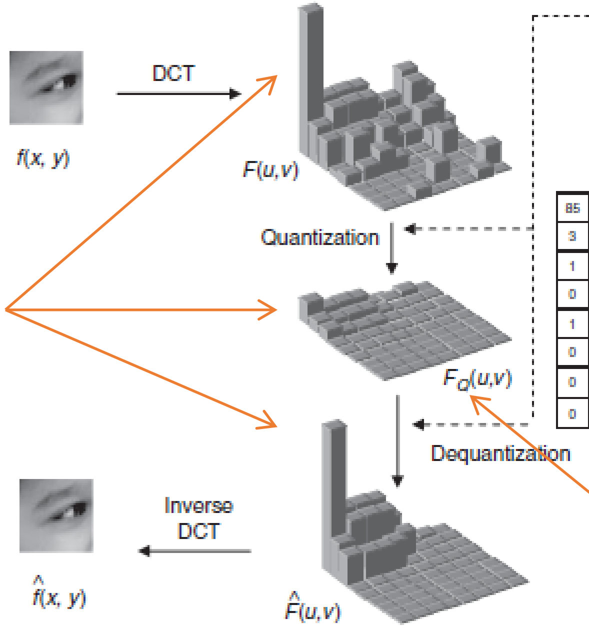
Pixel values $f(x, y)$

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



DC coefficient

Almost all the high-frequency are zero.

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

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

192	185	178	152	193	162	155	190
181	172	162	154	187	164	159	192
183	168	150	158	181	167	162	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

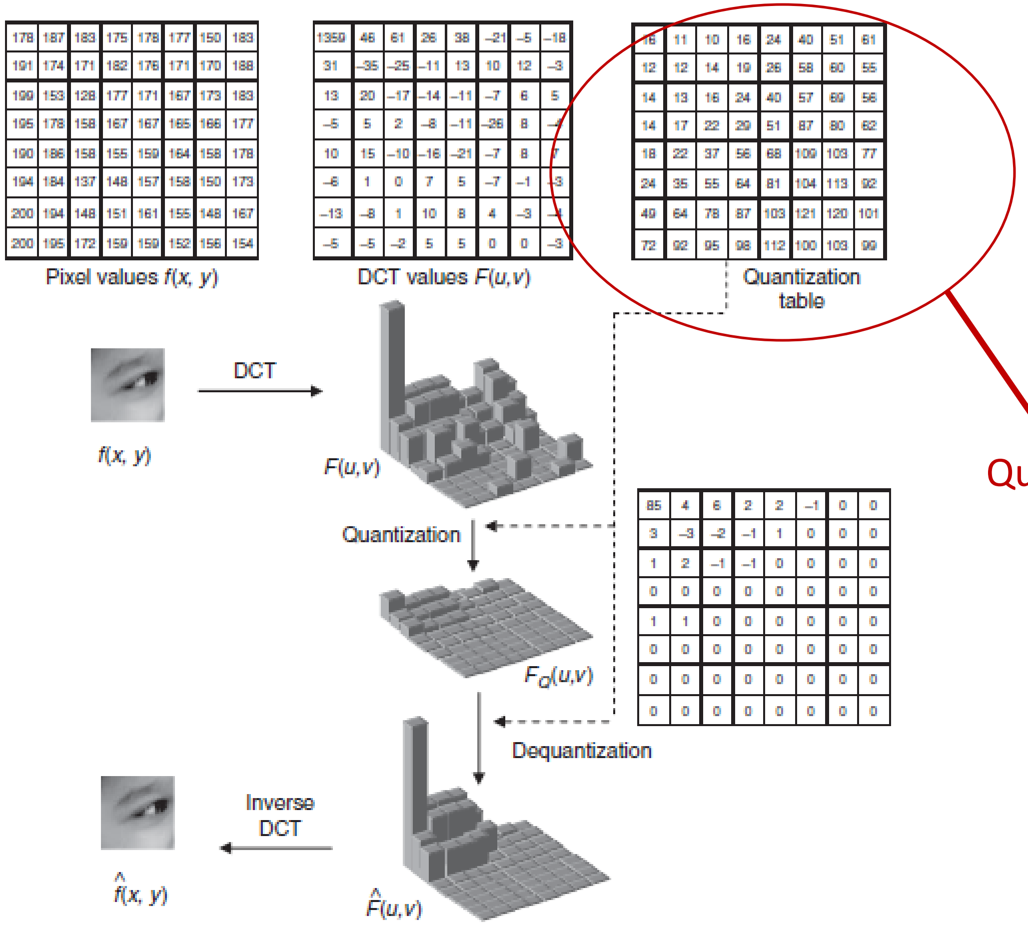
1360	44	60	32	48	−40	0	0
36	−36	−28	−19	26	0	0	0
14	26	−16	−24	0	0	0	0
0	0	0	0	0	0	0	0
18	22	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

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 f(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.

pakorn.wa@kmitl.ac.th

- 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.

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



Pixel values $f(x, y)$

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

DCT values $F(u,v)$

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

Quantization table

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

$F_Q(u,v)$

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

$\hat{f}(x, y)$

192	185	178	152	193	162	155	190
181	172	162	154	187	164	159	192
183	168	150	158	181	167	162	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

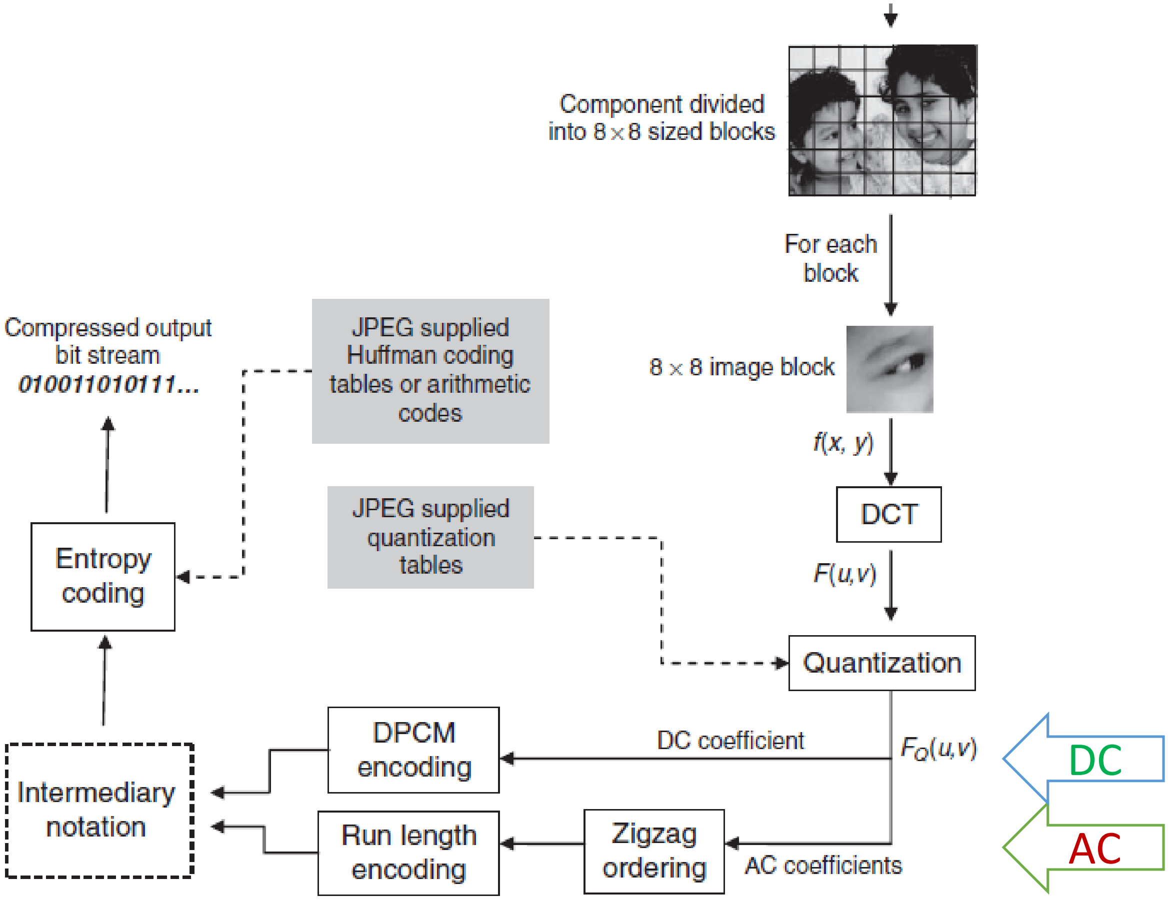
$\hat{F}(u,v)$

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

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 f(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.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

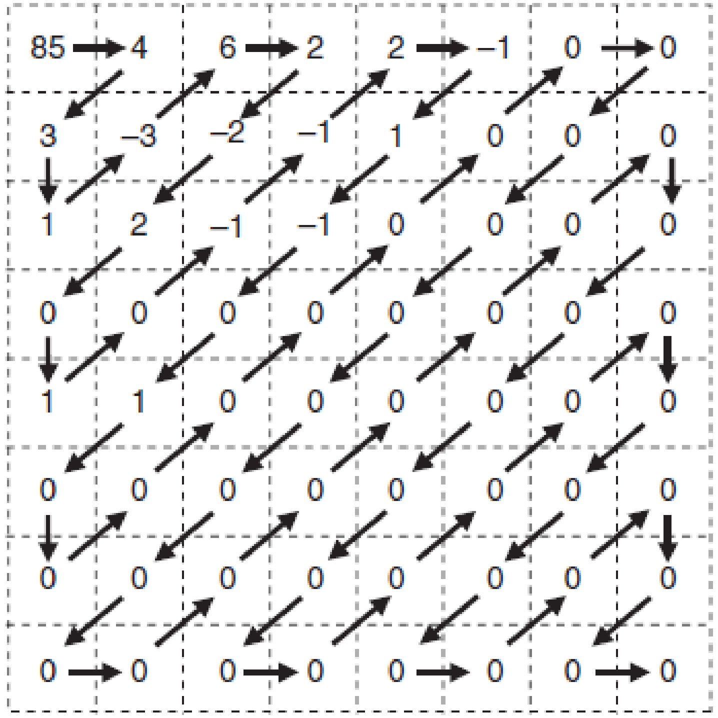
pakorn.wa@kmitl.ac.th



pakorn.wa@kmitl.ac.th

- The quantized coefficients $F_Q(u, v)$ are then encoded into an intermediary
- pattern.
  - The DC coefficients of the blocks are encoded using differential pulse code modulation.
  - The AC coefficients are first scanned in a zigzag order. Consequently, the zigzag ordering produces a longer run of zeros towards the end of the scan because the high-frequency coefficients appear at the tail end.
  - Both the DPCM codes of DC coefficients and the run length coded AC coefficients produce an intermediary representation.

pakorn.wa@kmitl.ac.th



DC coefficient = 85

AC coefficient stream

4 3 1 -3 6 2 -2 2 0 1 0 -1 -1 2
-1 1 -1 0 1 0 0 0 0 0 0 0 0 0 ...

A lot more zeros ...

pakorn.wa@kmitl.ac.th

- The next step is to entropy code the run of DC and AC coefficients.
  - Prior to entropy coding, these coefficients are converted into intermediary representations.
- The intermediary representations are then entropy coded using codes supplied by the JPEG organization.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

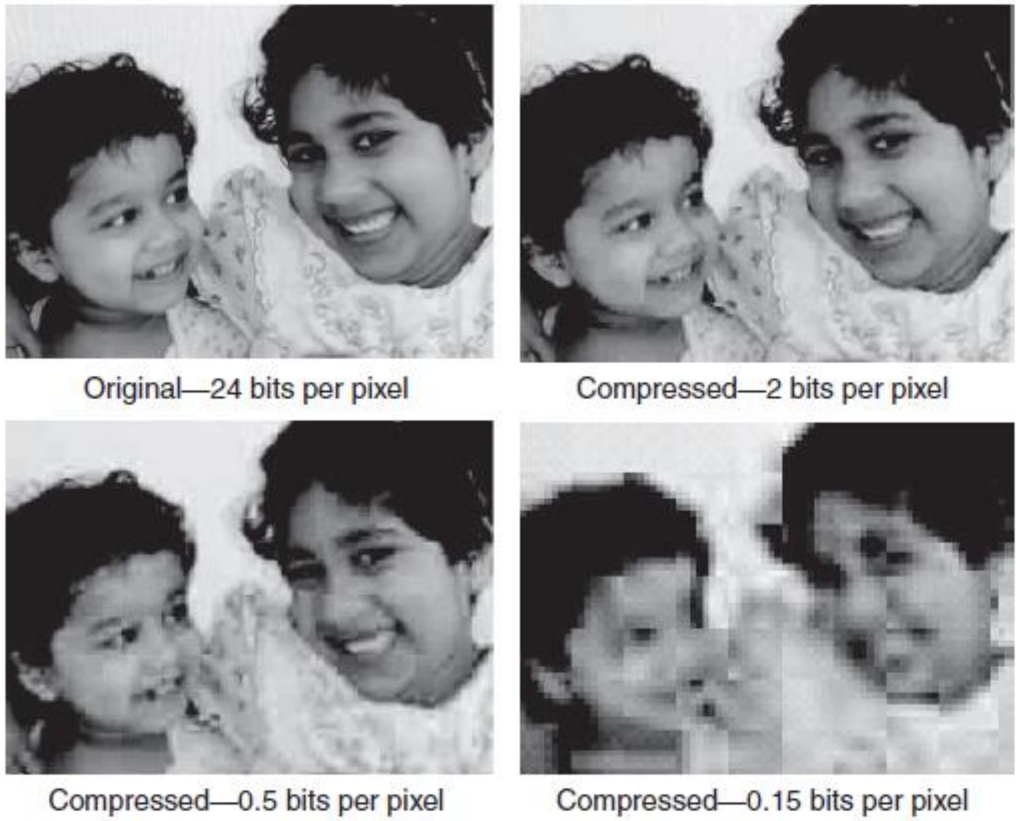


Figure 7-10 JPEG compression results. The upper-left image shows the original at 24 bits per pixel. The remaining three images show the reconstructed outputs after JPEG compression at different compression ratios. The blocky artifacts increase at lower bit rates.

pakorn.wa@kmitl.ac.th

JPEG Bit Stream

- The JPEG algorithm where the image is divided into blocks and each block is encoded in a scan line manner.
- The resultant compressed bit representation was also derived for a sample block.
- The compressed bit representations of all the scanned blocks are packed and organized to form a

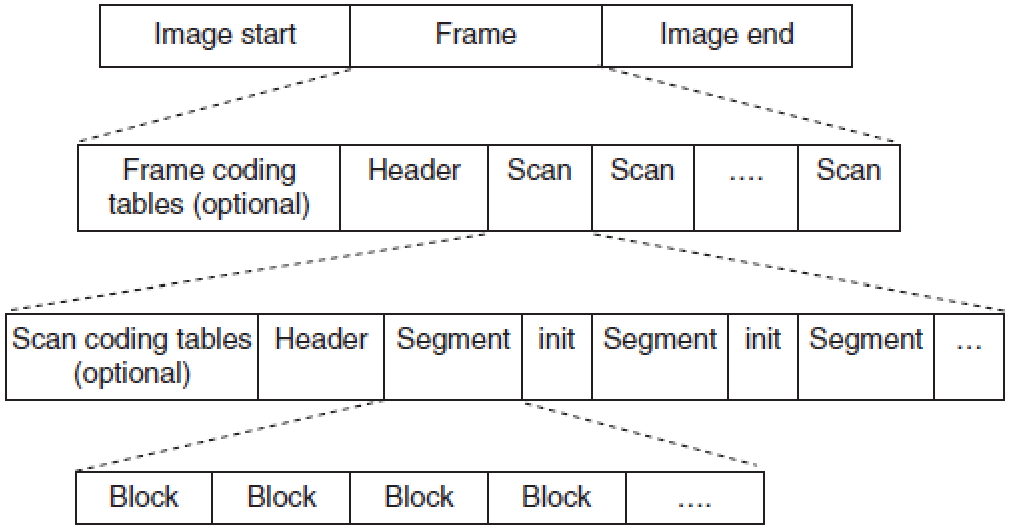


Figure 7-11 Hierarchical representation of the JPEG bit stream.

Drawbacks of JPEG

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- *Poor low bit-rate compression*
  - 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.
- *Random access of the bit stream*
  - JPEG does not allow random access because each of the 8×8 blocks is interdependent.
- *Large image handling*
  - JPEG does not allow for the compression of images larger than 64K by 64K without tiling.
- *Single compression architecture*
  - The current JPEG standard has about 44 modes.

pakorn.wa@kmitl.ac.th

- *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 (JPEG2000)

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

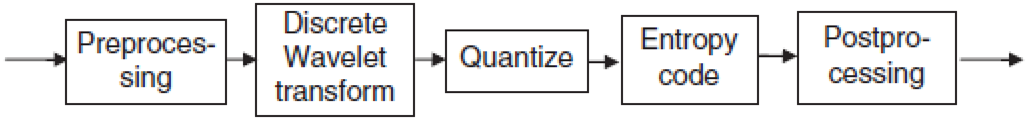


Figure 7-12 The JPEG 2000 pipeline

- The JPEG 2000 compression pipeline makes use of the Discrete Wavelet transform (DWT) to compress images
- DWT work better than DCT in the way it distributes energy among the frequency coefficients.
- Instead of working on individual 8 × 8 blocks, 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

The Preprocessing Step

pakorn.wa@kmitl.ac.th

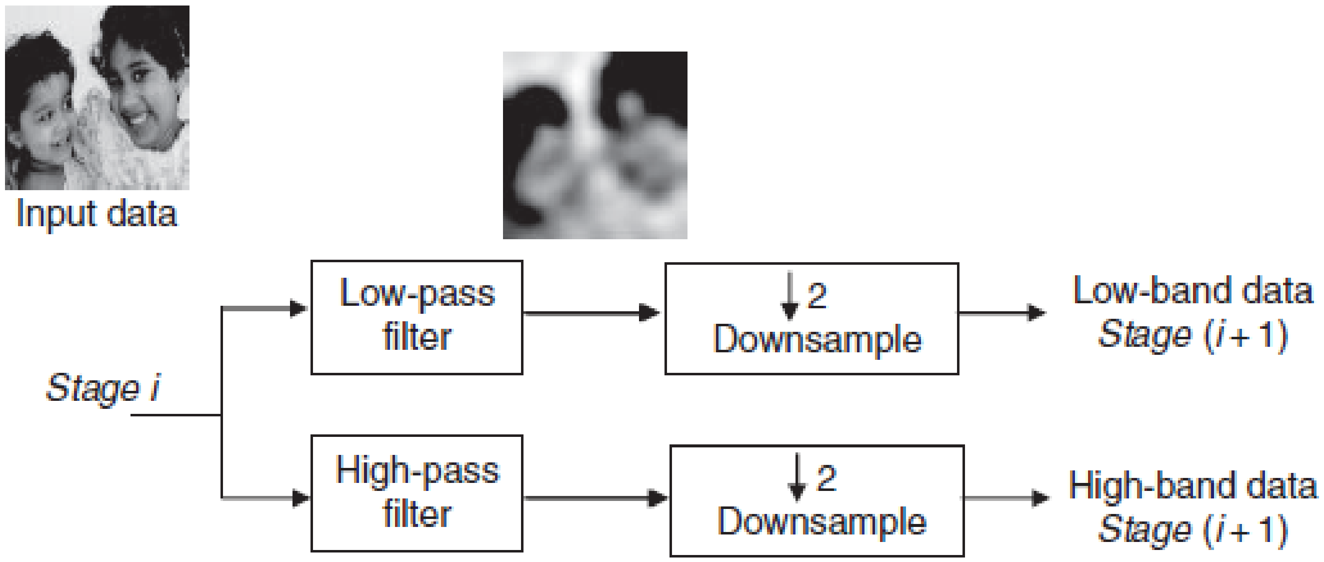
- The *tiling process*
  - Partitions the image into rectangular, but equal-sized and non-overlapping blocks.
  - Each tile is then independently processed for DWT analysis, quantization, entropy coding, and so on.
  - The tiling process is purely optional and is done only to reduce memory requirements, as needed to deal with very large images.

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 YCrCb conversion process is similar to the JPEG pipeline. 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.

The Discrete Wavelet Transform

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th



พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

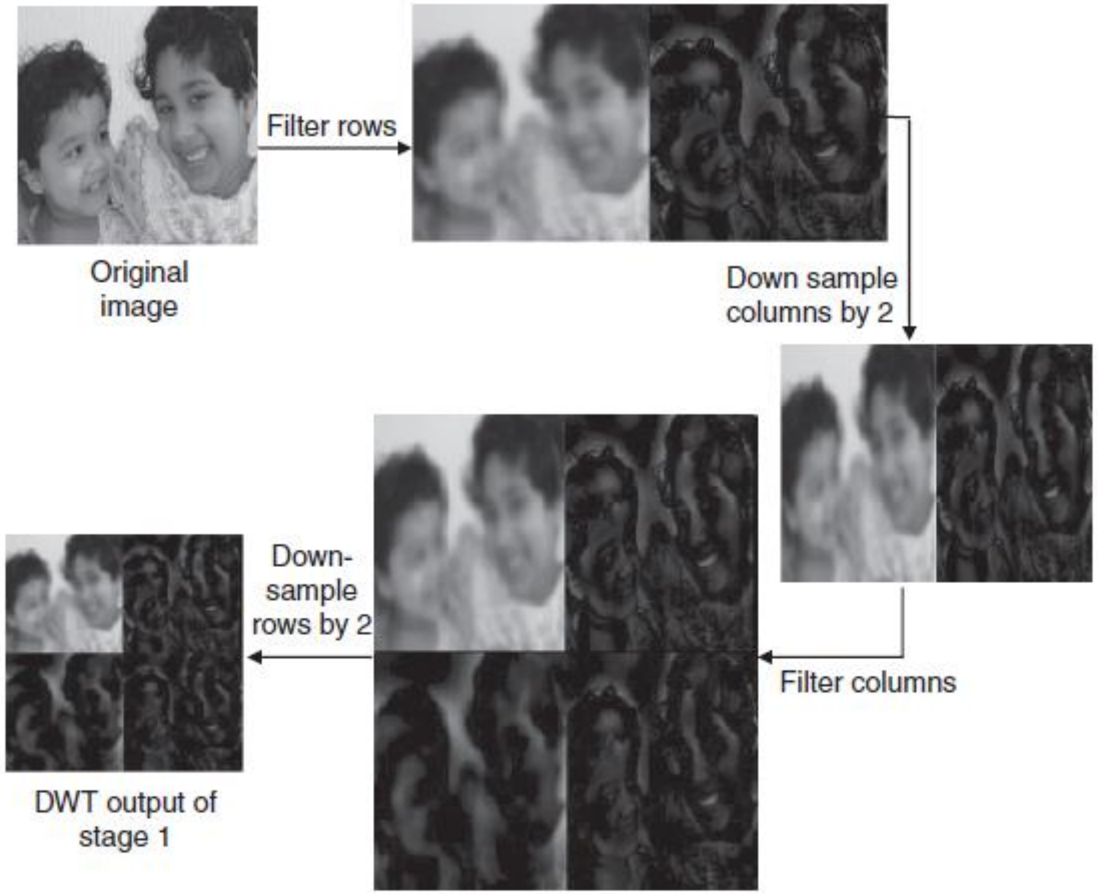


Figure 7-15 DWT process for the Y component of the sample image used.

pakorn.wa@kmitl.ac.th

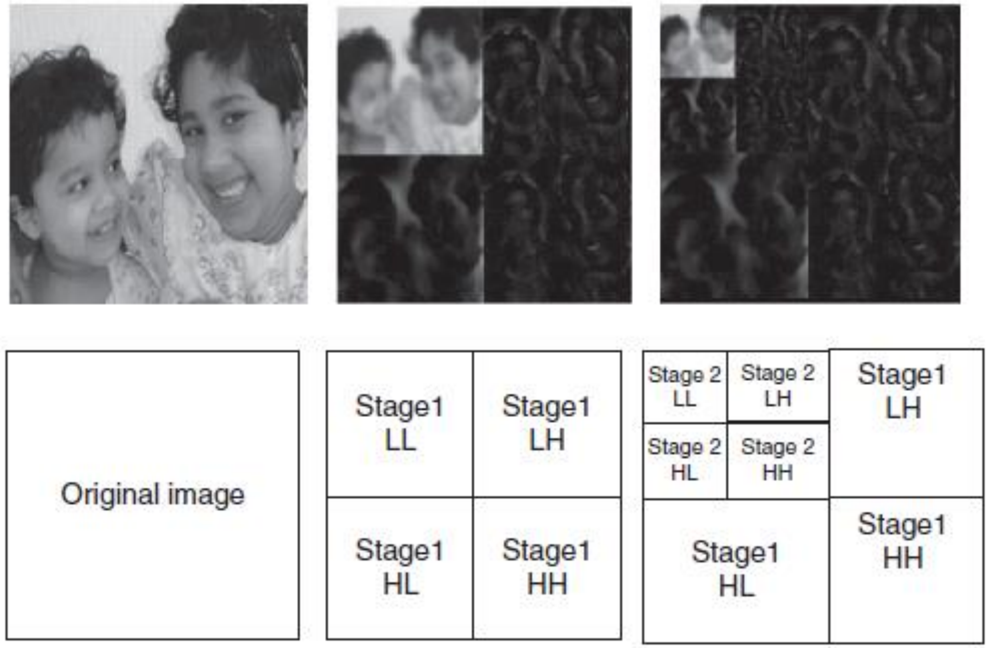


Figure 7-16 Original input and the output of discrete wavelet processing at the first two levels. The top row shows the imaged outputs. The bottom row shows what each block at a level contains.

- *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

JPEG 2000 Versus JPEG

พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- Encode once – platform-dependent decoding
- Working with compressed images – Simple geometrical transformation (cropping, flipping, rotating, etc.) can be applied to the compressed representation of the image.
- Region-of-interest encoding

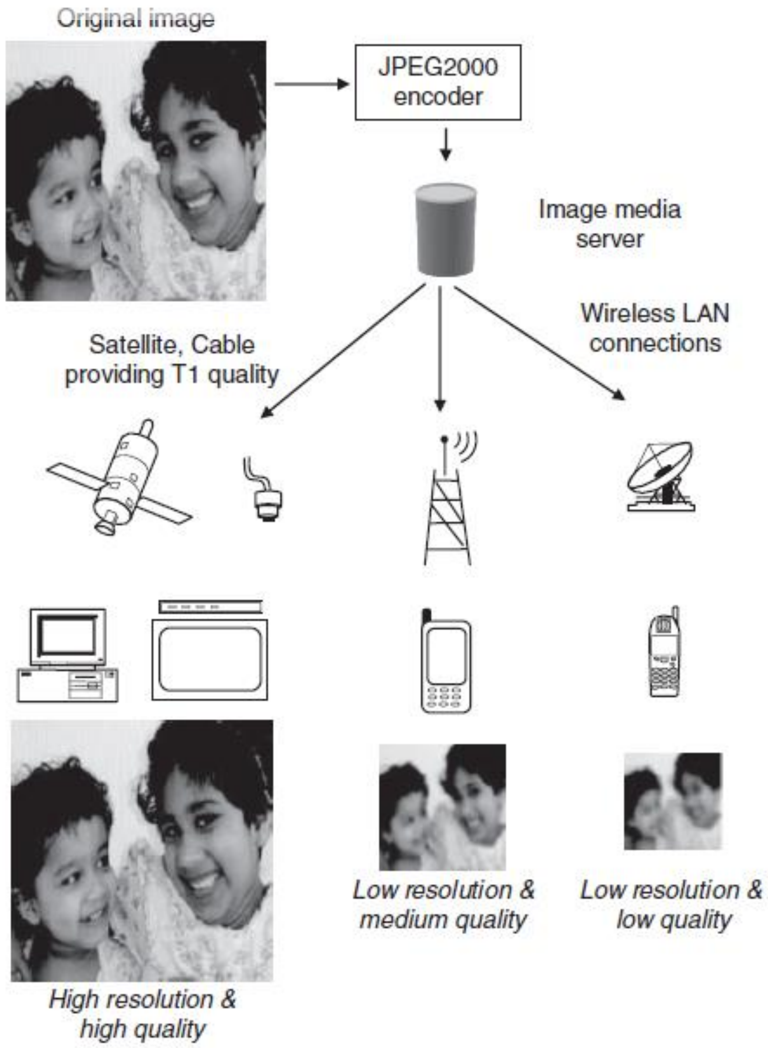


Figure 7-18 Region of interest (ROI). Using JPEG 2000 ROI coding, specific regions can be encoded or decoded with higher resolutions compared with the rest of the image.

Progressive Transmission Using DCTs in JPEG

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

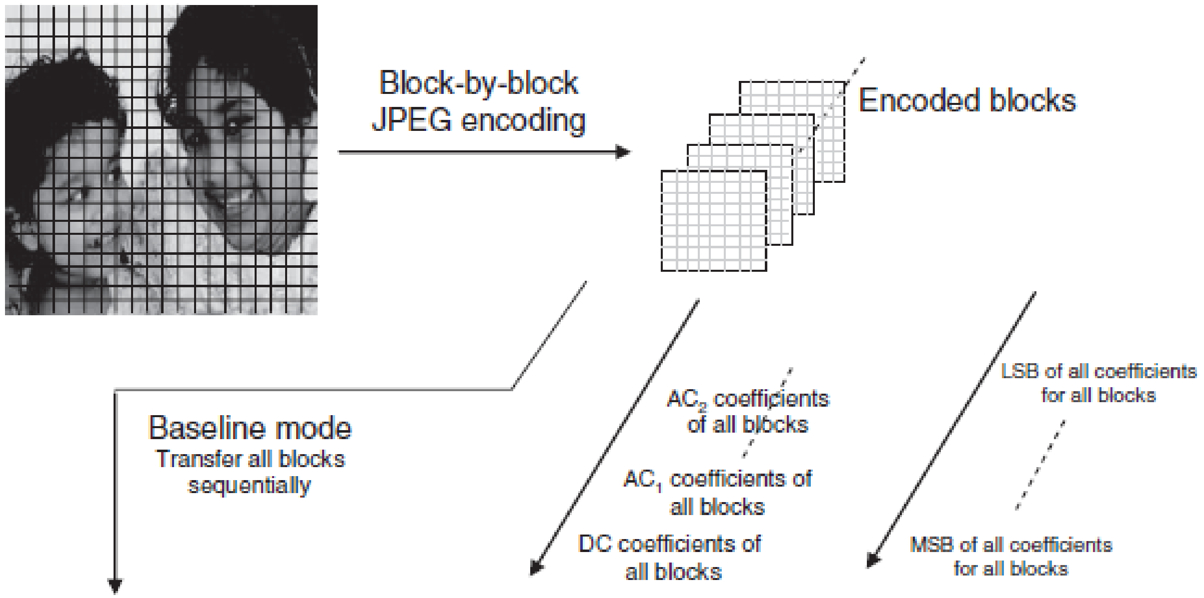
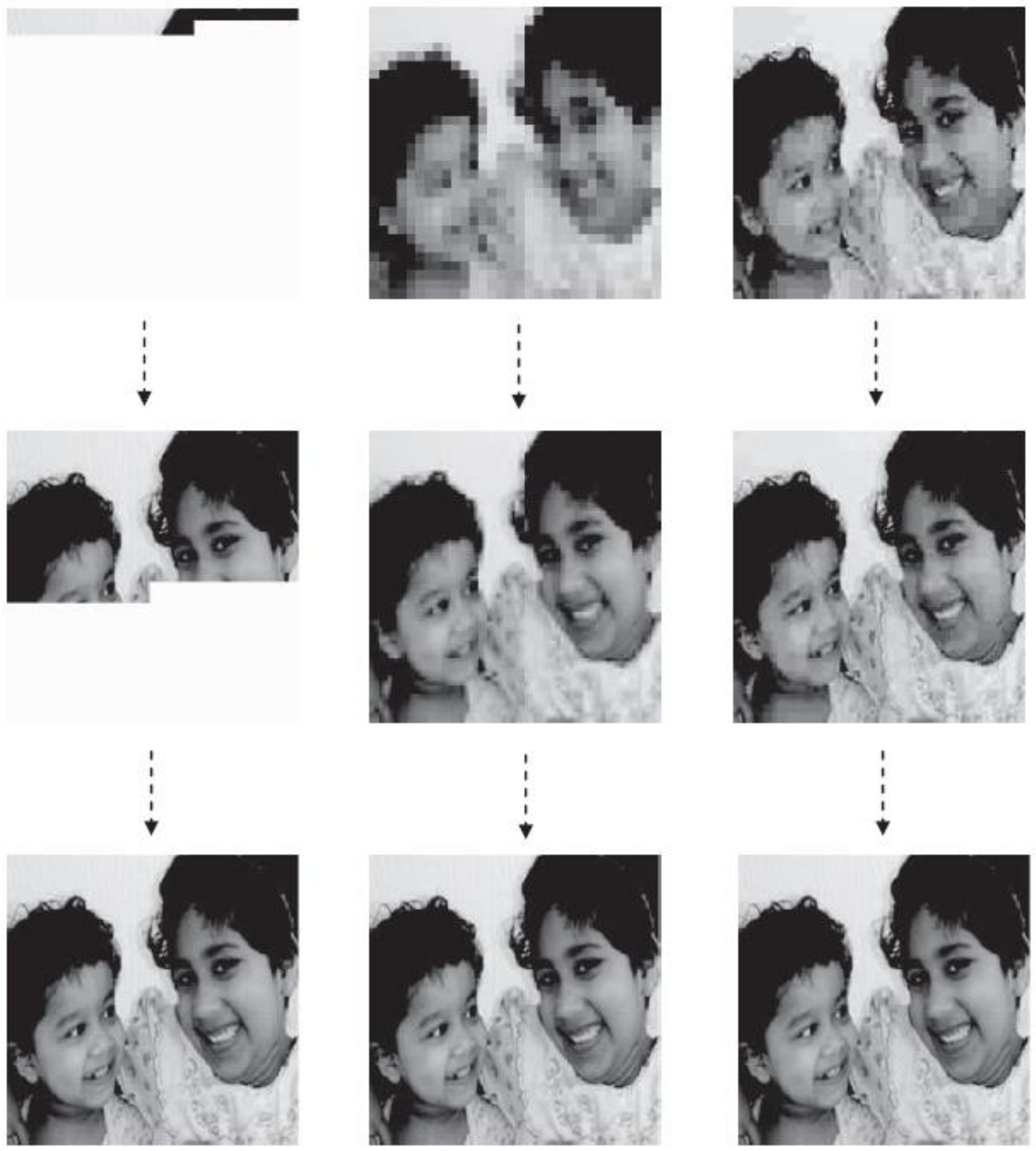


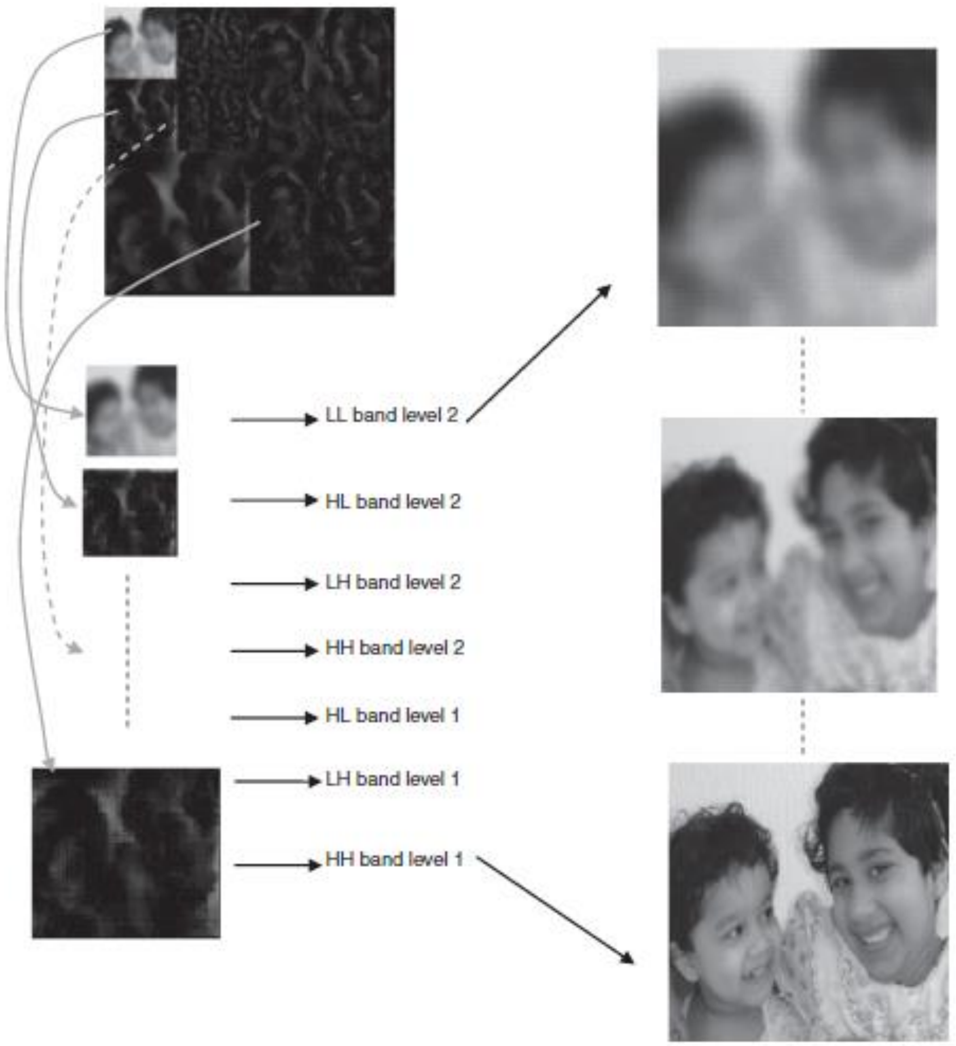
Figure 7-21 Progressive transmission in JPEG.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th



Progressive Transmission Using Wavelets in JPEG 2000

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th



KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

Q & A