CHAPTER 9 Multimedia Signal Processing (Data Compression)

Learning Outcomes

- Understand the meaning of multimedia and different components of multimedia

    - ❖ Understand what is multimedia, the different type of multimedia and multimedia signal processing.

- Understand Data compression and its process

    - ❖ Define compression and its significance.
    - ❖ Explain the process of data compression.
    - ❖ Define some compression terms.

- Understand the two major classifications of data compression and the various common data compression algorithm applied to text, image and audio.

    - ❖ Classification of Data Compression – Lossless and Lossy.
    - ❖ Know the different formats and data compression standards for image and video.
    - ❖ Apply Run-length algorithm.
    - ❖ Apply Arithmetic Algorithm.
    - ❖ Apply Lempel-Ziv-Welch (LZW) Algorithm.

- Understand the various terminologies, JPEG, MPEG, Discrete Cosine Transform in multimedia signal processing.

    - ❖ Able to understand the basics of JPEG and MPEG/MPEG 4.
    - ❖ Apply Discrete Cosine Transform (DCT) in image compression.
    - ❖ Briefly explain next generation MPEG 7.

CHAPTER 9 Multimedia Signal Processing (Data Compression)

9.1 The meaning of multimedia and different components of multimedia

The proliferation of consumer desktop personal computers started in the 1980s with accelerated growth at an insane pace where the internet came in the picture in 1990s. This resulted in a number of innovative devices, IPod, digital cameras and tablets just to name a few. Coupled with that the demand on on-line services in the service, banking and gaming industries.

This huge growth of multimedia technology over the past decades drives the demand for digital information dramatically. Consumers desire higher resolution of cameras, screen resolution, huge storage in devices or on cloud platform. In other applications such as tele-medical and satellite images, geographical data cause a huge constraint on the network traffic.

There are many definitions of **multimedia**. Multi means many and media, channel of communication forms. Hence, **multimedia** is a combination of text, audio, images and videos inclusive of animations for the consumers. When a viewer of a multimedia presentation is allowed to control what elements are delivered and when, it is *interactive* multimedia.

Multimedia signal processing the techniques to manage the whole range of multimedia types such text, graphics, speech, audio, image, video; standards for multimedia coding.

Multimedia signal processing which deals the technology that are needed to support multimedia applications, such as:

- Data Compression
- Data and file format
- Multimedia input, output and storage

Audio, image or video is almost always compressed to save disk storage and network bandwidth. Data compression is becoming even more important in online games and content based image retrieval system that require lesser latency time and high resolutions. In some video games, 4K is the standard video resolution replacing 1080p video, this means a fourfold increase of pixels on screen. As the contents for both events are high data sizes, the storage, transmission and interaction post challenging requirement. The receivers could be a low end hand phone where the data storage and random access memory may not be very large.

Data or information needs to be captured, stored and transmitted from the source which could be a hand phone to the receiver whether a server or another hand phone. There is also a need for information like images to be retrieved.

Fraunhofer-Institute IOSB develops a software module for content based image retrieval which the demonstration software can be obtained at https://www.iosb.fraunhofer.de/servlet/is/28046/. Given an input image, the aim of the software is to find those images within an image database which show the same object or scene as the input image. With such enormous amount of images to be stored and retrieved, the compression technology is critical to keep the size of the images small and also fast retrieval.

Content based image retrieval in images from the server. This is particular important e-commerce and also tele-presence. With on-line gaming, online teleconferencing, Virtual, Augmented and Mixed reality in the Covid 19 situation, the demand on the bandwidth the also the need to compress the data so that it will take lesser time to be transmitted over the network.

There is a need to understand this new technology, multimedia signal processing especially the data compression techniques is one of the ways to address all these challenges. Different compression techniques are more suitable for different media.

The storage size will depend on the type of media. For example
, plain text will take up 20-30 kbits per page, CD-quality music 1.5Mbits/s. An email without images is about 20kbits. A 1024x1024 still image is about 1Mbits and high definition television 1 Gbits/s. MP3 audio is roughly 1 megabyte per minute.

The size of uncompressed DVD Movie for 1 hour with 720 x 576 pixels resolution can easily reaches 100 Gbytes. This put an enormous constraint on the storage, transmission and computation time.

Entropy also plays a part in data compression as it is also a measure of how much the data can be compressed.

Compression is a technology or coding that fundamentally reduces the size of video, data, image and voice files so that lesser space will be required to be store or send through the networks. . Hence, it dramatically reduces the amount of network capacity (bandwidth) needed to transmit high definition TV, music, movies, computer games, without keep increasing the size of the data storage especially for small hand held devices where the data storage is still limited in gigabytes instead of petabytes at the data centers.

Compression thus increases throughput, without changing the actual bandwidth or capacity of the medium. At the receiver, the compressed files are decompressed and in the exact transmitted form or at a slightly lower quality depending on the applications.

For text, it is should not be altered especially for critical numerical information or crucial production information. However, for other applications such as images (not medical images such as MRI or X-ray), video or voice in varying degrees of lower resolution or quality with acceptable or hardly noticeable alteration.

Companies using a large number of IoT (internet of Thing) devices need to make sure they can communicate quickly, without incurring high cost. For example, one smart meter company is able to cut down on the amount of data they consumed from 400 MB per day per device with data compression technique, to a 10:1 reduction in data.

For the mobile Internet, the operators need to collect vehicles information such as the location, speed, height, license plates, vehicle state, driving conditions, drivers, passenger positions, and other pieces of text information. Different data formats, sizes, and types through data exchange, through the vehicle terminal wireless communication module for uploading. a compression/decompression mode can ensure the integrity of data transmission, can reduce data traffic, and can reduce communication costs.

Compression is one of the enabling technologies that supports the multimedia revolution. Without successful techniques and standards for efficiently compressing multimedia in general (e.g., text, speech, audio, images, videos) there will be difficulties in sending large text, sounds, images and videos files across the networks, storage, retrieved and performance simulation at the clouds through the electronic gadgets such as hand phones, smart screens, laptops and other IoT devices.

Multimedia issues without good compression techniques will affect the storage, retrieval, analysis and creation of the content. With good compression technology support, it allows better applications such as video on demand, video broadcasting (pay TV, videoconferencing, multimedia databases and internet TV.

9.2 Data Compression and its process

9.2.1 Purpose of Data Compression

1) Reduce storage size
2) Save transmission bandwidth
3) Save cost by reducing the transmission time

Although the cost of storage and transmission bandwidth for digital data have dropped dramatically, the demand for increasing their capacity in many applications has been growing rapidly ever since. There are cases in which extra storage or extra bandwidth is difficult to achieve, if not impossible.

Data compression as a means may make much more efficient use of existing resources with less cost. For an image size of 500 kilobytes, 100 000 images will be easily occupied 50 Gbytes. A colour image of 1600 x1200 resolution will take up 5.76 Mbytes.

Data compression techniques is motivated mainly by the need to improve efficiency of information processing.

Image data compression eliminates redundancy and irrelevancy in the data:

1) Temporal - Pixels in two video frames that have the same values in the same location.
2) Spatial - correlation between neighboring pixels or data items.
3) Spectral - correlation between color or luminescence components. This uses the frequency domain to exploit relationships between frequencies of change in data.
4) Psycho-visual - exploit perceptual properties of the human visual system.

9.2.2 Process of Data Compression

Data compression in its simplest sense is to reduce the size of the data by removing unused or redundant data. The compression method depends intrinsically on the type of data to be compressed: an image will not be compressed in the same way as an audio or text file.

Compression can be defined by the compression factor, that is, the number of bits in the compressed image divided by the number of bits in the original image.

The compression ratio, which is often used, is the inverse of the compression factor; it is usually expressed as a percentage.

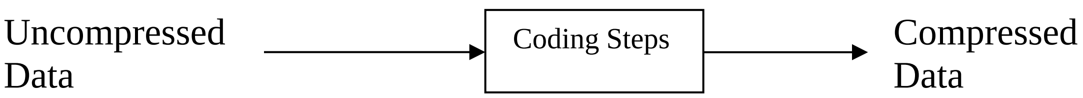
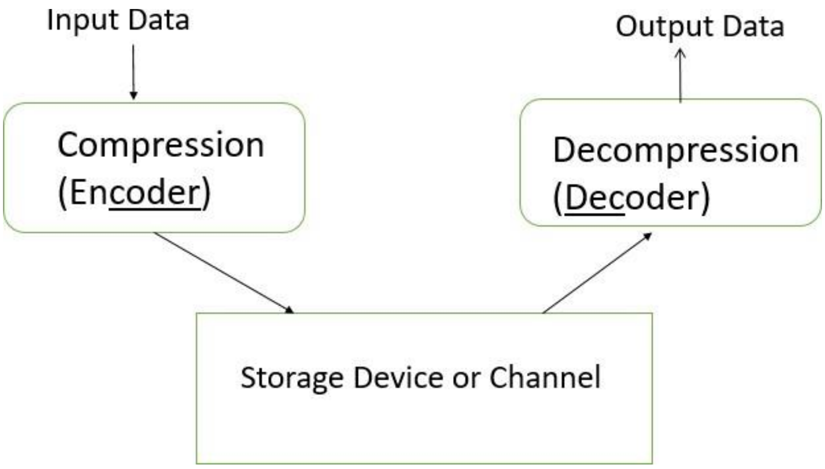


Figure 9.1: Data Compression Encoding Process



Codec: Encoder + Decoder

Figure 9.2 : Data Compression Scheme

Compression ratio, C_r or **CR**= n1/n2

Where

n1 = the size of data before compression

n2 = the size of data after compression

Relative Data Redundancy, $R_d = 1 - 1/C_r$

Space saving, Ss = $1 - 1/C_r$

Compression gain, $C_g = 10 \log(C_r)$ dB

Example 9.1

If n1=n2, find C_r and R_d.

Solution

$C_r = 1$ and $R_d = 0$

Example 9.2

If n1:n2 = 10:1, calculate R_d and S_s.

Solution

R_d=0.9 → 90% of data redundant

S_s = 1-0.9 =0.1

Example 9.3

An image file (pixels 512 x 512) with 262,144 bytes is compressed into a file with 32,768 bytes.

Find the compression ratio, C_r , Space saving, S_s and Compression gain, C_g

Solution

Compression ratio, C_r = 262144/16,384= 8

Space saving, S_s = 1 -1/8 = 0.875 or 87.5%

Compression gain, G_g = 10 log(8) = 9.031 dB

The mean-square error (MSE) and the peak signal-to-noise ratio (PSNR) are used to compare image compression quality. MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the lower the error.

For the original data or image, I_1 and I_2, reconstructed data or image.

PSNR, Peak signal-to-noise ratio = 20 log (max|Xi|/MSE)

MSE, Mean Square Error = $\left[\frac{1}{mX\ n} \sum_{n,m} (I_1(m,n) - I_2(m,n)) \right]$

When the original and reconstructed data are exactly identical, MSE will be zero and PSNR will be infinity. For good quality reconstructed image, the value of MSE will be low and PSNR value will be high.

Example 9.4

255	250
10	15

Original Image Pixels

249	249
9	14

Reconstructed Image Pixels

MSE = $\left[\frac{1}{4\ x\ 4} (1^2+1^2+1^2+1^2)\right] = 0.25$

PSNR = 10 log(255^2/0.25) = 54.151 dB

Dctdemo.m is an educational demo Matlab program that demonstrate Discrete cosine transform (DCT) image compression. It allows user to select the number of coefficients and it shows original and reconstructed image and an error image, plus the MSE value.

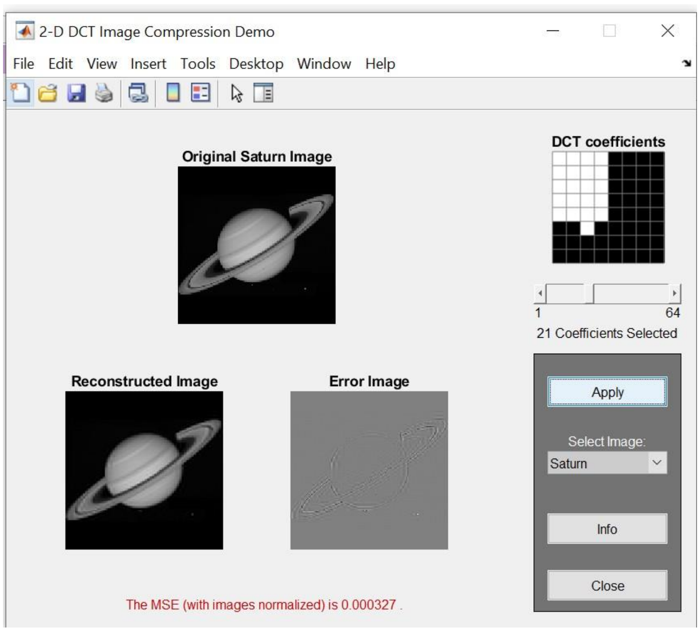


Figure 9.3: DCTDEMO.m

9.3 Classification of Data Compression, Common Techniques and Various Standards

9.3.1 Two major classifications of Data Compression

Data Compression can be divided into 2 major categories:

1) **Lossless Compression** (Reversible Compressions)

2) **Lossy Compression** (Irreversible Compression)

In lossless compression, the decompressed data is the “same” as the original data.

For lossless compression, the original data can be reconstructed back while lossy compression some of the data will be lost. In short, lossless compression algorithms retain all of the original data of the file, preserving its original quality precisely.

- Lossless compression does not involve the process of quantization, but makes use of encoding to provide compression. Almost every lossy compression system contains a lossless compression system.

The lossy compression compresses the original data and replaces them with an “approximation”. Multimedia data (audio, video) can tolerate a certain level of degradation without the sensory organs (eye, tympanum, etc) distinguishing any significant degradation.

Nearly all video clips available to consumers use lossy compression methods, due to the large file sizes. When downloading video clips, quality is largely determined by the file size of Gbytes.

Lossy compression especially for image processing, involves at least three steps; image transformation, quantization, and encoding. Transformation is a lossless step in which the image is transformed from gray scale values in the spatial domain to coefficients in frequency domain.

Some of these transforms are Discrete Cosine Transform (DCT) and the Discrete Wavelet Transform (DWT) which are commonly used for image compression.

No loss of information occurs in the transformation step. Quantization is the step in which the data integrity is lost. It attempts to minimize information loss by preferentially preserving the most important coefficients where less important coefficients are roughly approximated to zero and ignored.

A few of the popular coding methods or transforms for both lossless and lossy compression are given.

Lossless Compression Techniques (Information Preserved, Low Cr)

a. Run Length
b. Arithmetic
c. Lempel-Ziv-Welch, LZW

Lossy Compression Techniques (Some Information lost, High Cr)

a. JPEG, Joint Photographic Experts Group
b. MPEG, Moving Picture Experts Group

9.3.2 Lossless Compression Methods

Run-length Coding, RLC

It is the simplest method of compression. Run-length encoding or algorithm, performs lossless data compression and applies to palette-based iconic images. However, it does not work so well on continuous-tone images such as photographs. However, JPEG uses RLC for the coefficients that remains after transforming and quantizing the image blocks. It replace consecutive repeating occurrences of a symbol by 1 occurrence of the symbol itself, then followed by the number of occurrences.

Run-length Algorithm/Encoding is frequently applied to images (or pixels in a scan line). It is a small compression component used in JPEG compression. In this instance, sequences of image elements X_1, X_2, …, X_n are mapped to pairs (c_1, l_1), (c_1, l_2), …, (c_n, l_n) where c_i represent image intensity or colour and l_i the length of the ith run of pixels.

The idea of Run-length coding is to replace consecutively repeated symbols in a source with a code pair which consists of either the repeating symbol or the number of its occurrences, or sequence of non-repeating symbols.

Example 9.5

String ABBBBBBBCC can be represented by (A,1),(B,7), (C,2).

All the symbols are represented by an 8-bit ASCII codeword.

For example,

Original Data: DDDDDDDDDDDDDDDDOOOOOOOOOOOGGGGGGGGGG

Compressed Data:

(D,16), (O,11), (G,10)

Original Data:

0000000111111111100000001111 1

Compressed Data:

(7,0), (1,10), (0,7),(1,5)

Arithmetic Algorithm/Coding

Arithmetic coding is a form of entropy **encoding** used in lossless and lossy data compression. Arithmetic perform encoding the whole message or signal into a single number rather than separating it into the component symbols and changing each with the code.

In arithmetic coding a source ensemble is represented by an interval between 0 and 1 on the real number line, according to the probabilities of the occurrence of the intensities. Each symbol of the ensemble narrows this interval. As the interval becomes smaller, the number of bits needed to specify it grows. Arithmetic coding assumes an explicit probabilistic model of the source.

It is a defined-word scheme which uses the probabilities of the source messages to successively narrow the interval used to represent the ensemble. A high probability message narrows the interval less than a low probability message, so that high probability messages contribute fewer bits to the coded ensemble.

The method begins with an unordered list of source messages and their probabilities. The number line is partitioned into subintervals based on cumulative probabilities.

Example 9.6

The source of information A generates the symbols {a, b, c, d and e} with the corresponding probabilities {0.4, 0.3, 0.2 and 0.1}. The message to be encoded is "aadbc".

Step 1

Source Symbol	**P_i**	**Sub-interval**
a	0.4	[0.0;0.4)
b	0.3	[0.4;0.7)
c	0.2	[0.7;0.9)
d	0.1	[0.9;1.0)

Step 2

The first symbol to be encoded is **a**.

Source Symbol	**New ‘a’ interval**
→**a**	[0.0;0.16)
b	[0.16;0.28)
c	[0.28;0.36)
d	[0.36;0.4)

Step 3

The second symbol to be encoded is **a**.

Source Symbol	**New ‘a’ interval**
→**a**	[0.0;0.064)
b	[0.064;0.112)
c	[0.112;0.144)
d	[0.144;0.16)

Step 4

The third symbol to be encoded is **d**.

Source Symbol	**New ‘d’ interval**
a	[0.144;0.1504)
b	[0.1504;0.1552)
c	[0.1552;0.1584)
→**d**	[0.1584;0.16)

Step 5

The fourth symbol to be encoded is **b**.

Source Symbol	**New 'b' interval**
a	[0.144;0.1504)
→**b**	[0.1504;0.1552)
c	[0.1552;0.1584)
d	[0.1584;0.16)

Step 6

The final step for the fourth symbol to be encoded is **c**.

Source Symbol	**New 'c' interval**
a	[0.1504;0.15232)
b	[0.15232;0.15376)
→**c**	[0.15376;0.15472)
d	[0.15472;0.1552)

The number of symbols encoded will be stated in the protocol of the image format, so any number, such as 0.1543 within [0.15376, 0.15472) will be acceptable. The binary representation of this number is **0.001001111** which is 9 bits.

Lempel-Ziv-Welch (LZW) Algorithm

LZW compression or technique uses a table-based lookup algorithm invented by Abraham Lempel, Jacob Ziv, and Terry Welch. Two commonly-used file formats in which LZV **compression** is used are the GIF image format served from Web sites and the TIFF image format.

The LZW algorithm is a very common compression technique.

- LZ encoding is an example of a category of algorithms called dictionary-based encoding. The idea is to create a dictionary (table) of strings used during the communication session. The compression algorithm extracts the smallest substring that cannot be found in the dictionary from the remaining non-compressed string.

- Original methods due to Ziv and Lempel in 1977 and 1978. Terry Welch improved the scheme in 1984 (called LZW compression).

- Applications: Unix Compress, gzip, GIF

Example 9.7

Given the input symbols is abababa
Solution

Input	Output	Index	String
a	(0,a)	1	a
b	(0,b)	2	b
ab	(1,b)	3	ab
aba	(3,a)	4	aba

Compressed string (codewords) = (0,a),(0,b),(1,b),(3,a)
Total number of bits = (1+8)+(1+8)+(2+8)+(2+8)= 38 bits
Hence there is a saving of the number of bits to be transmitted.

9.3.3 Lossy Compression

Lossy compression aims to obtain the best possible ***fidelity*** for a given bit-rate or minimizing the bit-rate to achieve a given fidelity measure. Video and audio compression techniques are most suited to this form of compression.

If an image is compressed it clearly needs to be uncompressed (decoded) before it can viewed/listened to. Some processing of data may be possible in encoded form however.

Lossless compression frequently involves some form of ***entropy encoding*** and are based in information theoretic techniques.
Lossy compression use source encoding techniques that may involve transform encoding, differential encoding or vector quantization.

Video and Audio files are very large. Unless we develop and maintain very high bandwidth networks (Gigabytes per second or more) we have to compress the data.

Multimedia consists of several media sources, such as video, audio, graphics, animation, text integrated in a meaningful way to convey some information.

Media can be broadly divided into continuous, temporal that changes with time such as audio or discrete, which is time independent such as texts, still images or graphics.

File formats are agreed standard rules or protocols. The followings are some the international standards

Two of the most important organizations that are responsible for the Multimedia compression standards:

(i) ITU International Telecommunication Union

(ii) ISO International Organization for Standardization

Different data compression standards for Continuous-Tone Still Images

JPEG (Joint Photographic Experts Group) is an ISO/IEC group of experts that develops and maintains standards for a suite of compression algorithms for computer image files.

Name	Description
JPEG	Joint Photographic Experts Group, JPEG standard for images of photographic quality based on lossy baseline coding systems. It uses quantized discrete cosine transforms (DCT) on image block, Huffman, and run length coding and is one of the most popular methods for compressing images on the Internet.
JPEG-LS	A lossless to near-lossless standard for continuous-tone images.
JPEG-2000	Arithmetic coding and quantized discrete wavelet transform (DWT) are used. The compression can be lossy or loseless.

Table 9.1 Different data compression standards for Continuous-Tone Still Images

The Moving Picture Experts Group (MPEG) is the family of standards and file formats which are used in digital video.

MPEG was developed by the working group formed by IEC and ISO, which is also known as moving picture experts group. Because of the various sophisticated compression techniques used, MPEGs, when compared to most audio and video formats, are smaller in size and more or less of same quality.

Name	Description
MPEG-1	Audio/video standards designed for digital storage media (such as an MP3 file).
MPEG-2	Standards for digital television and DVD video. Higher data rates for high-quality video
MPEG-4	Multimedia standards for the computers, mobile devices, and the web
MPEG-7	Standards for the description and search of multimedia content

Table 9.2 Internationally sanctioned video compression standards.

The basic MATLAB® data structure is the *array*, an ordered set of real or complex elements. An array is naturally suited to the representation of *images*, real-valued, ordered sets of color or intensity data. (An array is suited for complex-valued images.)

In the MATLAB workspace, most images are represented as two-dimensional arrays (matrices), in which each element of the matrix corresponds to a single pixel in the displayed image. For example, an image composed of 200 rows and 300 columns of different coloured dots stored as a 200-by-300 matrix. Some images, such as RGB, require a three-dimensional array, where the first plane in the third dimension represents the red pixel intensities, the second plane represents the green pixel intensities, and the third plane represents the blue pixel intensities.

These are some of the popular image used in hand phone, internet and computer.

Name	Description
Bitmap (.bmp)	Bitmap Image File is a format developed by Microsoft for Windows. There is no compression or information loss with BMP files which allow images to have very high quality, but also very large file sizes. Due to BMP being a proprietary format, it is generally recommended to use TIFF files. **Compression:** None
TIFF (.tif, .tiff)	Tagged Image File Format are lossless images files meaning that they do not need to compress or lose any image quality or information (although there are options for compression), allowing for very high-quality images but also larger file sizes. **Compression:** Lossless - no compression. Very high-quality images.
JPEG (.jpg, jpeg)	JPG (short for Joint Photographic Experts Group, and pronounced jay-peg) is a file format best used for photo images which must be very small files, for example, for web sites or for email. JPG uses lossy compression (lossy meaning "with losses to quality")

GIF (.gif)	GIF or Graphics Interchange Format files are widely used for web graphics, because they are limited to only 256 colors, can allow for transparency, and can be animated. GIF files are typically small is size and are very portable. **Compression:** Lossless - compression without loss of quality

Table 9.3 Popular image standards and file formats

In Matlab, imwrite command allows the user to compress the image files through the setting of qualify factor, Q.

Quality	Size	Compression Ratio
Raw Lena_Color.TIFF	769KB	1:1
7_zip TIFF	620KB	1.2:1
Q=100	225KB	3.5:1
Q=75	37KB	17:1
Q=50	24KB	27:1
Q=10	16KB	72:1

Table 9.4 Quality factor vs Compression Ratio

Original Image	Quality, 10	Quality, 50	Quality, 75	Quality, 100
				
769 kB	10 kB	23.7 kB	36.9 kB	224 kB
CR= 1:1	76.9	32.45	20.84	3.43

Table 9.5 Photo's Quality vs Qualify Factor

JPEG (Joint Photographic Experts Groups)

JPEG (Joint Photographic Experts Group) is an international compression standard for continuous-tone still image, both grayscale and color. This standard is designed to support a wide variety of applications for continuous-tone images. Because of the distinct requirement for each of the applications, the JPEG standard have two basic compression methods.

It uses the lossy compression technique for either full-color or gray-scale images.

In JPEG, Forward Discreet Cosine Transform (FDCT) transforms the 64 values in 8x8 pixel block in a way that the relative relationships between pixels are kept but the redundancies are revealed.

In order to understand JPEG compression better, it helps to understand how computers represent photographic images. An image file can be thought of as a grid of individual blocks called pixels. Each pixel has its own color value, and the larger the image, the more pixels. The more pixels, the larger the resulting file will be.

The DCT-based method is specified for lossy compression, and the predictive method is specified for lossless compression. A simple lossy technique called baseline, which is a DCT-based methods, has been widely used today and is sufficient for a large number of applications.

Together with the Graphic Interchange Format (GIF) and Portable Network Graphics (PNG) file formats, the JPEG is one of the image file formats supported on the World Wide Web, usually with the file suffix of ".jpg".

Main Concept

Change the picture into a linear (vector) sets of numbers that reveals the redundancies. The redundancies is then removed by one of lossless compression methods.

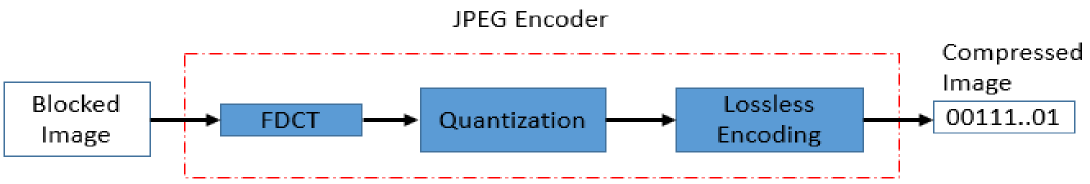


Figure 9.4 JPEG Encoder

4 Important building blocks of JPEG

1. **Image/Block Preparation** - Partitioning of input images into 8x8 sub-block of pixels.

2. **Forward Discrete Cosine Transform, FDCT** - Perform FDCT which lays the foundation for achieving data compression by concentrating most of the signal in the lower spatial frequencies. The coefficient with zero frequency in both dimensions is called the "DC coefficient: and the remaining 63 coefficients are called the "AC coefficients".

3. **Quantization**– quantized using uniform quantization tables based upon psych visual experiments.

4. **Entropy Encoding** - Zigzag – the scan is based upon that most of the high-frequency coefficients are zero. Encoding – uses Run length coding to generate intermediate sequence, and then Huffman coded for transmission or storage.

JPEG compression uses the 2D-DCT transform for both grayscale and color images in the YIQ color space used by the NTSC (NTSC is an abbreviation for **National Television Standards Committee**, named for the group that originally developed the black & white and subsequently color television system that is used in the United States, Japan and many other countries) color TV system. The Y component represents the luminance information, and is the only component used by black-and-white television receivers. I and Q represent the chrominance information.

The **RGB color model** in which red, green, and blue light are added together in various ways to reproduce a broad array of colors. The name of the model comes from the initials of the three additive primary colors, red, green, and blue. An **RGB image**, sometimes referred to as a true color **image**, is stored in MATLAB as an m-by-n-by-3 data array that defines red, green, and blue color components for each individual pixel. There is additional step of [R,G,B] to [Y,Cb,Cr] conversion after the division of image to 8x8 blocks.

JPEG uses different quality factors to normalize DCT coefficients for the luminance(Y) channel and the chrominance (IQ) channels.

<u>Discrete Cosine Transform, DCT</u>

- DCT is the “heart” for JPEG (Joint Photo-graphic Experts Group) for still-image compression and MPEG (Motion Picture Experts Group) for video compression with compression ratio more than 15.

- The principle of the DCT is to transform the original image pixels to an identical number of DCT coefficients. Frequency domains can be obtained through the transformation from one (time or spatial) domain to the other (frequency) via Fourier Transform (or Discrete Fourier Transform in the early chapter) or through Discrete Cosine Transform.

- Large magnitudes or values at high frequency components mean the data is changing rapidly on a short distance scale.

- The DCT coefficients have a non-uniform distribution of direct-current (DC) terms representing the average values, and alternate-current (AC) terms representing fluctuations.

- The coefficients with zero frequency is called the "DC coefficient" and the remaining 63 coefficients are called the "AC coefficients.

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

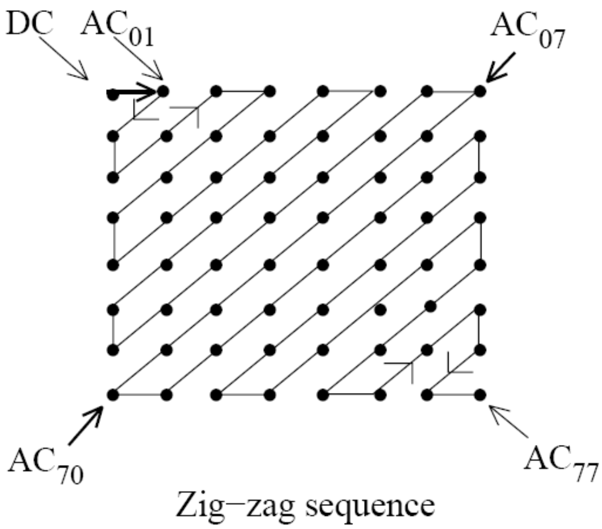


Fig 9.5 Zig-Zag Sequence

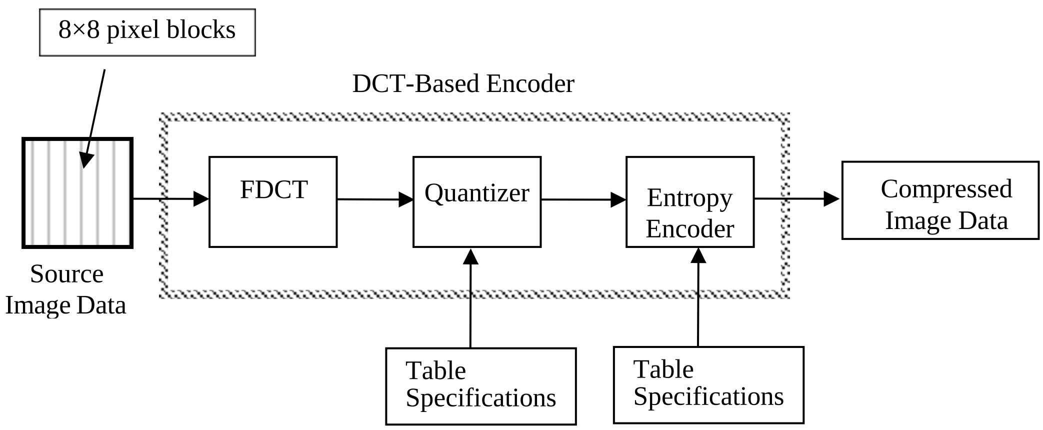


Fig. 9.6 DCT Encoder-Based Processing Step

The DCT transforms the spatial signal to frequency domain coefficients.

- No complex number operations (Discrete Fourier Transform)
- DCT decomposes a block of data
  - DC coefficient ->average of the data samples
  - AC coefficients, each corresponding to the frequency components.

<u>One-Dimensional (1D) DCT, 1D-DCT</u>

DCT(Forward Transform) for N data samples

$$X_{DCT}(k) = \sqrt{\frac{2}{N}}C(k)\sum_{n=0}^{N-1} x(n)\cos\left[\frac{(2n+1)k\pi}{2N}\right] \qquad k = 0,1,\cdots,N-1 \tag{9.1}$$

- IDCT(Inverse transform)

$$x(n) = \sqrt{\frac{2}{N}}\sum_{k=0}^{N-1} C(k)X_{DCT}(k)\cos\left[\frac{(2n+1)k\pi}{2N}\right] \qquad n = 0,1,\cdots,N-1$$

$$C(k) = \begin{cases} \frac{\sqrt{2}}{2} & k = 0 \\ 1 & otherwise \end{cases} \tag{9.2}$$

<u>Example 9.8</u>

Samples: x(n)={10,8,10,12} where N=4

$$X_{DCT}(0) = \sqrt{\frac{2}{4}} \times \frac{\sqrt{2}}{2}\sum_{n=0}^{4-1} x(n)\cos(0) = \frac{1}{2}(10+8+10+12)$$

Hence DCT coefficients:

$X_{DCT}(0) = 20$, $X_{DCT}(1) = -1.8478$, $X_{DCT}(2) = 2$, $X_{DCT}(3) = 0.7654$

Using Matlab,

```
>> dct([10 8 10 12])

ans =
  20.0000   -1.8478    2.0000    0.7654
```

If the whole array is a fixed number and only the DC coefficient will have a value but the AC coefficients will be zero.

The DCT coefficients can be further quantized and encode them into binary data. 1D-DCTcan be used to audio coding. A modified DCT is used in MPEG-1 MP3 audio coding.

Similarly, 2D-DCT can be used for JPEG image compression.

Two-Dimensional (2D) DCT, 2D-DCT

$$F(u,v) = \frac{2C(u)C(v)}{\sqrt{MN}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p(i,j) \cos\left(\frac{(2i+1)u\pi}{2M}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \tag{9.3}$$

Inverse DCT

$$p(i,j) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \frac{2C(u)C(v)}{\sqrt{MN}} F(u,v) \cos\left(\frac{(2i+1)u\pi}{2M}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \tag{9.4}$$

$$C(m) = \begin{cases} \frac{\sqrt{2}}{2} & if\ m = 0 \\ 1 & otherwise \end{cases}$$

p(I,j) = pixel level at the location (i,j)
F(u,v) =DCT coefficient at the frequency indices (u,v)

The DCT transformation creates amples, *p(i,j)* from *F(u,v)*, DCT coefficients

The DC value gives the average value of the pixels. The AC values gives the changes.

Lack of changes in neighboring pixels creates 0s.

The DCT transformation is reversible.

Mathematical formula for DCT transformation – complex and required practice.

<u>Quantization</u>

After the *F* table is created, the values are quantized to reduce the number of bits needed for encoding.

Divide the number by a constant and then drop the fraction.

The quantizing phase is not reversible.

Some information will be lost.

<u>Compression</u>

After **quantization**, the values are read from the table, and redundant 0s are removed.

The reason is that if the picture does not have fine changes, the bottom right corner of the *F* table is all 0s.

Example 9.9

Find the 2D-DCT coefficients for the following image:

p(i,j) = $\begin{vmatrix} 100 & -50 \\ 100 & 10 \end{vmatrix}$

DC coefficient, F(0,0)= $\frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1} p(i,j)$=$\frac{1}{2}(100-50+100+10)$

Hence, F(0,0)= 80

Using Matlab,

```
>> F=dct2([100 -50; 100 10])

F=
   80.0000   120.0000
  -30.0000   30.0000
```

2-D JPEG Grayscale Image Compression

a. DCT coefficients have a big DC component but small AC components values.

b. These coefficients are further normalized (quantized) with a quality factor Q where the DC coefficient is further reduced slightly but the AC coefficients much more leaving a few AC coefficients.

c. Encode and transmit only nonzero DCT coefficients and omit transmitting zeros, hence DATA COMPRESSION!

Video compression—MPEG

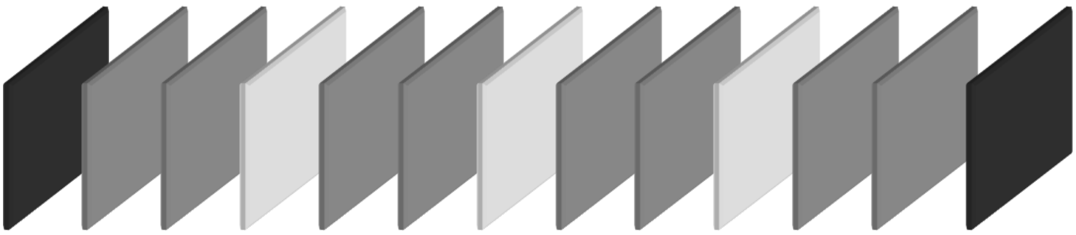
- In a motion sequence, individual frames of pictures are grouped together (called a *group of pictures or frames*, or *GOP*). MPEG method first divides frames into three categories: I-frames, P-frames, B-frames.

- An I-frame (Intra-coded picture or key frame), a complete image, like a JPG or BMP image file. One would expect the compression ratio will be the least for I-frame than the other 2 types of frame, namely B and P frames.

- **P** (Predicted Picture) and **B** (Bi-directionally predicted pictures) **frames** hold only part of the image information (the part that changes **between frames**), so they need less space **in the** output file than an I-**frame**.

- **Intraframe** compression, when applied to image sequences, reduces only the spatial redundancies present in an image sequence.

- **Interframe** compression employs temporal predictions and thus aims to reduce temporal as well as spatial redundancies, increasing the efficiency of data compression.



I_1 B_1 B_2 P_1 B_3 B_4 P_2 B_5 B_6 P_3 B_7 B_8 I_2

Fig. 9.7 I, P and B Frames

Basic idea

1) Spatial compression of each frame is done with JPEG.
2) Temporal compression removes the redundant frames.
3) Each video is a rapid sequence of a set of frames. Each frame is a spatial combination of pixels, or a picture.
4) Compressing video = spatially compressing each frame
   +
   Temporally compressing a set of frames.

MPEG is a popular file format for audio and video.

For broadcasting applications with limited resources, MPEG is one of the best options as it allows streaming high quality video on platforms like the Internet. The file format is supported by most electronic products and can be played using software like Windows Media Player, Cyberlink PowerDVD, Apple Quick Time Player, etc.

8 important MPEG Features

1) Compared to most formats, are smaller in size.
2) Sophisticated compression techniques.
3) Compared to most formats, superior audio and video quality.
4) Major standards are: MPEG1, MPEG2, MPEG3, MPEG4, and MPEG7.
5) Many simple, cheap decoders.
6) Supported by all popular browsers.
7) Both a non-commercial and cross-bros
8) High image resolution and multi-channel sound technique.

The MPEG standards are an evolving set of standards for video and audio compression and for multimedia delivery developed by the Moving Picture Experts Group (MPEG).

The MPEG compression is achieved using **5 different techniques**:

1) The use of a frequency-based transform called Discrete Cosine Transform (DCT).
2) Quantization, a technique for losing selective information (sometimes known as lossy compression) that can be acceptably lost from visual information.
3) Huffman coding, a technique of lossless compression that uses code tables based on statistics about the encoded data.
4) Motion compensated predictive coding, in which the differences in what has changed between an image and its preceding image are calculated and only the differences are encoded.
5) Bi-directional prediction, in which some images are predicted from the pictures immediately preceding and following the image.

The first three techniques are also used in JPEG file compression.

MPEG-1 was designed for coding progressive video at a transmission rate of about 1.5 million bits per second. It was designed specifically for Video-CD and CD-i media. MPEG-1 audio layer-3 (MP3) has also evolved from early MPEG work. One of MPEG Audio Lossy Compression is MPEG-1.

MPEG-1

The MPEG algorithm exploits perceptual limitations of the human auditory system to determine which part of an audio signal is acoustically irrelevant and can be removed for compression.

MPEG Audio Compression – layer 3

1) Exploit areas where the human ear is less sensitive to sound to achieve compression.
2) MPEG audio compresses by removing acoustically irrelevant parts of audio signals
3) Dividing the audio signal up into a set of frequency sub bands.
4) Uses a bank of DCT-based or FFT-based (or even FIR/IIR) analyzer filters.
5) Perform quantization process.
6) Uses Huffman coding on quantized samples.

MPEG-2 was designed for coding interlaced images at transmission rates above 4 million bits per second. MPEG-2 is used for digital TV broadcast and DVD. An MPEG-2 player can handle MPEG-1 data as well.

MPEG 2 Video Compression

1) Pictures are made up of pixels.

2) Each 8x8 array of pixels is known as a block.

3) A 2x2 array of blocks is termed a macroblock.

4) Compression is achieved through certain techniques of prediction.

5) Pictures called I pictures are encoded without prediction. Pictures termed P pictures may be encoded with prediction from previous pictures. B pictures may be encoded using prediction from both previous and subsequent pictures.

MPEG-1 and -2 define techniques for compressing digital video by factors varying from 25:1 to 50:1.

A proposed MPEG-3 standard, intended for High Definition TV (HDTV), was merged with the MPEG-2 standard when it became apparent that the MPEG-2 standard met the HDTV requirements.

MPEG-4 is a much more ambitious standard and addresses speech and video synthesis, fractal geometry, computer visualization, and an artificial intelligence (AI) approach to reconstructing images. MPEG-4 addresses a standard way for authors to create and define the media objects in a multimedia presentation, how these can be synchronized and related to each other in transmission, and how users are to be able to interact with the media objects.

MPEG-7 is the later addition in the family of MPEG standards and added by September 2000. There will be more MPEG's standards to meet the multimedia evolution. MPEG-7 will be a standardized description of various types of multimedia information. This description will be associated with the content itself, to allow fast and efficient searching for material that is of interest to the user. MPEG-7 is formally called "Multimedia Content Description Interface".

The increasing availability of potentially interesting audio/video material makes its search more difficult. This challenging situation led to the need of a solution to the problem of quickly and efficiently searching for various types of multimedia material interesting to the user. MPEG-7 wants to answer to this need, providing this solution.

The people taking part in defining MPEG-7 represent broadcasters, equipment manufacturers, digital content creators and managers, transmission providers, publishers and intellectual property rights managers, as well as university researchers.

MPEG-7 will not replace MPEG-1, MPEG-2 or MPEG-4. It is intended to provide complementary functionality to these other MPEG standards, representing information about the content, not the content itself ("the bits about the bits"). This functionality is the standardization of multimedia content descriptions. MPEG-7 can be used independently of the other MPEG standards - the description might even be attached to an analogue movie. The representation that is defined within MPEG-4, i.e. the representation of audio-visual data in terms of objects, is however very well suited to what will be built on the MPEG-7 standard. This representation is basic to the process of categorization. In addition, MPEG-7 descriptions could be used to improve the functionality of previous MPEG standards.

There are many applications and application domains which will benefit from the MPEG-7 standard. A few application examples are:

1) Digital libraries (image catalogue, musical dictionary)

2) Multimedia directory services (e.g. yellow pages)

3) Broadcast media selection (radio channel, TV channel)

4) Multimedia editing (personalized electronic news service, media authoring)

The full list of latest MPEG standards can be obtained from this link,

https://mpeg.chiariglione.org/docs/full-list-mpeg-standards