Indira Gandhi National Open University
School of Sciences

MMTE-003
PATTERN RECOGNITION AND IMAGE PROCESSING

Volume-I
IMAGE PROCESSING

Unitised Course Outline

Volume I Image Processing

Block 1 Digital Images

Unit 1: Introduction to Digital Image

Unit 2: Image Transforms I

Unit 3: Image Transforms II

Unit 4: Colour Image Processing

Block 2 Image Improvement I

Unit 5: Image Enhancement in Spatial Domain

Unit 6: Histogram Modelling

Unit 7: Neighbourhood Operations

Block 3 Image Improvement II

Unit 8: Image Enhancement in Frequency Domain
Unit 9: Image Restoration I
Unit 10: Image Restoration II

Volume II Pattern Recognition and Scilab

Block 4 Pattern Recognition

Unit 11: Image Segmentation

Unit 12: Unsupervised Learning

Unit 13: Pattern Classification

Unit 14: Supervised Learning

Block 5 Scilab Manual

Unit 15: Introduction to Scilab

Unit 16: Introduction to SIVP Toolbox

Session 1: Basic Commands

Session 2: Plotting in Scilab

Session 3: SIVP Basic Commands

Session 4: Image Tansforms

Session 5: Edge Detection

Session 6: Histogram

Session 7: DFFT and IFFT

Session 8: Noise

Session 9: Colour Transformation

Session 10: Image Segmentation

COURSE INTRODUCTION

Welcome to the course "Pattern Recognition and Image Processing". Vision is the most powerful of the five human senses. Visual information, conveyed in the form of images gives better impact than textual information. The fields of digital image processing have grown tremendously over the past 30 years. Digital image processing is concerned primarily with the extraction of useful information from images. In this process, it deals with 1) image representation, 2) enhacing the quality of an image, 3) restoration of the original image from its degraded version, and 4) compression of the large amount of data in the images for efficient archiving and retrieval.

In this course, the image processing algorithms are simulated using SCILAB. This course contains five blocks, out of which one block is lab manual.

Block 1 gives an overview of the digital image-processing system. The different elements of image-processing systems like image acquisition, image sampling, quantization, image sensors, image scanners and image storage devices are covered in this chapter. We highlight the different types of image file formats in practice. It deals with two-dimensional signals and systems. Different types of two-dimensional signals, and properties of two-dimensional signals. The need for image transforms, different types of image transforms and their properties are explained in this block. The concepts related to colour image processing includes human perception of colour, colours models, colour image quantisation, colour image filtering, pseudo-colouring.

In Block 2, we discuss the various techniques used to enhance the image. We focus image enhancement in spatial domain. We also highlight histogram modeling for image. The aim of this block is to highlight the importance of image enhancement and to explain various algorithms for doing it. In Block 3, we discuss spatial domain operations on image along with the image enhancement in frequency domain. We also explain low pass and high pass filters with their transfer functions. We also discuss image degradation models and image restoration techniques.

Block 4 is the last theory block. In this block we begin with image segmentation. In which you will see how various objects present in any image are segmented with the help of edge detection, line detection, boundary detection, etc. we begin pattern recognition here by discussing clustering approaches. We continue recognition to various techniques such as Bayesian classifiers, minimum distance classifiers, etc.

As you know this course has a practical component. A lab manual consists of two units and ten sessions of practical. Practicals are to be done in scilab environment.

Now, a word about our notation. Each unit is divided into sections, which may be divided into sub-sections. These sections/sub-sections are numbered sequentially, as are the exercises and important equations, within a unit. Regarding the exercises in each unit, these have been placed at many places in a unit. They are meant for you to assess if you have understood the concepts and processes that have been discussed till that point. Therefore, you must attempt to solve all the exercises in the unit as you go along, before you look up the solution.

Since the material in the different units is heavily interlinked, there will be cross referencing. For this we will be using the notation Sec. x.y to mean Section y of Unit x. Further, the end of an example is shown by the sign *** and the end of a proof of, theorem/proposition/corollary is shown by the mark ■.

Another compulsory component of this course is its assignment, which covers the whole course. Your academic counselor at the study centre will evaluate it, and return it to you with detailed comments. Thus, the assignment is meant to be a teaching as well as an assessment aid. Further, you will not be allowed to take the exam of this course till you submit your assignment response. So please submit it well in time.

The course material that we have sent you is self-sufficient. If you have a problem in understanding any portion, please ask your academic counselor for help. In addition to these exercises, you may like to also look up some other books in the library of your study centre, and try to solve some exercises from these books too.

Some useful books and websites are the following:

1) "Digital Image Processing", S. Jayaraman, S. Esakkirajan, T. Veerakumar, Mc Graw Hill.
2) "Digital Image Processing", Gonzalez, Woods, Pearson.
3) "Fundamentals of Digital Image Processing", Jain, PHI.
4) nptel.ac.in/courses/117/117105079.

Best wishes for an enjoyable time "Pattern Recognition and Image Processing."

The Course Team

Indira Gandhi National Open
University School of Sciences

MMTE-003
Pattern Recognition and Image Processing

Block

1

Digital Images

Curriculum Design Committee

Dr. B.D. Acharya
Dept. of Science & Technology, New Delhi

Prof. Adimurthi
School of Mathematics
TIFR, Bangalore

Prof. Archana Aggarwal
CESP, School of Social Sciences
JNU, New Delhi

Prof. R.B. Bapat
Indian Statistical Institute
New Delhi

Prof. M.C. Bhandari
Dept. of Mathematics
IIT, Kanpur

Prof. R. Bhatia
Indian Statistical Institute, New Delhi

Prof. A.D. Dharmadhikari
Dept. of Statistics
University of Pune

Prof. O.P. Gupta
Dept. of Financial Studies
University of Delhi

Prof. S.D. Joshi
Dept. of Electrical Engineering
IIT, Delhi

Dr. R.K. Khanna
Scientific Analysis Group, DRDO, Delhi

Prof. Susheel Kumar
Dept. of Management Studies
IIT, Delhi

Prof. Veni Madhavan
Scientific Analysis Group
DRDO, Delhi

Prof. J.C. Misra
Dept. of Mathematics
IIT, Kharagpur

Prof. C. Musili
Dept. of Mathematics and Statistics
University of Hyderabad

Prof. Sankar Pal
ISI, Kolkata

Prof. A.P. Singh
PG Dept. of Mathematics
University of Jammu

Faculty Members
School of Sciences, IGNOU
Prof. Poornima Mital
Dr. Atul Razdan
Prof. Parvin Sinclair
Prof. Sujatha Varma
Dr. S. Venkataraman

Course Design Committee

Dr. A.K Sinha
DRDO, Delhi

Prof. Vipula Singh
RNSIT, Bangalore

Prof. Muthukumar Subramanyaom
NIT, Puducherry

Faculty Members
School of Sciences, IGNOU
Prof. Sujatha Varma
Dr. S. Venkataraman
Dr. Deepika

Block Preparation Team

Prof. K.K. Biswas (Retd.) (Editor)
IIT, Delhi

Dr. Ayesha Choudhary
School of Computer and Systems Sciences, JNU

Dr. Suddhansha Sharma
School of Computer Information Sciences,
IGNOU

Dr. Deepika
School of Sciences, IGNOU

Course Coordinator: Dr. Deepika

Acknowledgements: Mr. Santosh Kumar Pal for preparing the CRC.

March, 2021

ISBN-

Further information on the Indira Gandhi National Open University courses may be obtained from the University's Office at Maidan Garhi, New Delhi-110 068.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by the Director, School of Sciences.

BLOCK 1 DIGITAL IMAGES

Digital images play an important role, both in daily-life applications such as satellite television, magnetic resonance imaging, computer tomography as well as in areas of research and technology such as geographical information systems and astronomy. An image is a 2D representation of a three-dimensional scene. A digital image is basically a numerical representation of an object. The term digital image processing refers to the manipulation of an image by means of a processor. The different elements of an image-processing system include image acquisition, image storage, image processing and display. Unit-1 begins with the basic definition of a digital image and is followed by a detailed discussion on two-dimensional sampling. This is followed by different elements of image processing systems.

Unit-2 and Unit-3 deal with different image transforms. Transform is basically a representation of the signal. Efficient representation of visual information lies at the foundation of many image-processing tasks which include image filtering, image compression and feature extraction. Efficiency of a representation refers to the ability to capture significant information of an image in a small description.

Humans use colour information to distinguish objects, materials, food, places and time of the day. Colour images surround us in print, television, computer displays, photographs and movies. Colour images in general occupy more space when compared to grayscale and binary images.

In Unit-4, we shall discuss basic concepts related to colour formation, human perception of colour information, different types of colour models and CIE chromaticity diagram.

UNIT 1

INTRODUCTION TO DIGITAL IMAGE

Structure **Page No.**

1.1 INTRODUCTION

There is a famous saying that a picture is equivalent to thousands of words, and this turns out to be true when we come to the field of image processing and pattern recognition. It is the interpretation of image which deteremines various types of outcomes viz. for an arts person it may cary a different meaning but for a science person the meaning of that image may be entirely different. Based on the utility of the images, different formats are deviced to store the different type of images, which are stored in different file formats and they contain a variety of information in the form of colour, contrast, intensity, brightness etc., but resolution of these images holds a great importance in the field of image processing, because it affects the quality of image when it is scaled up. This unit relates to the understanding of essentials of image processing and tofind major applications of image processing in our everyday life. Broadly speaking, image processing is an area that deals with manipulation of visual information called images, and one of the major objectives of image processing is to improve the quality of pictorial information for better human interpretation and to facilitate the automatic machine interpretation of images.

We shall begin our discussion by defining an image in Sec. 1.2 and will continue image acquisition and image digitization in Sec. 1.3 and Sec. 1.4 respectively. In the subsequent sections, we shall define the basic terms and concepts, which will be used throughout the course such as representation of image, resolution of image, characteristics of image and types of images. We shall end the unit by discussing various areas in which digital image processing is used.

Now, we shall list the objectives of this unit. After going through the unit, please read this list again and make sure that you have achieved the objectives.

Objectives

After going through this unit, you should be able to

- define and represent an image;
- define the various terms used in digital image processing;
- apply sampling and quantization for image digitization;
- list various classifications of images along with their descriptions;
- relate the areas where digital image processing can be applied.

Now, let us find the answer to the question “what an image is?” in the following section.

1.2 INTRODUCTION TO AN IMAGE?

When we studied optics as a subject in the field of physics, we learned that when two or more rays of light meet at a point an image is formed, and we studied various phenomenons like reflection, refraction, diffraction, dispersion, scattering etc., this means we were in the process of image generation and analysis, even much before the development of computers. But, with the advent of the computers in our day to day life this field of image processing has transformed in to an entirely new field i.e. digital image processing, where sampling and quantization techniques are applied to convert the image into its discrete form, this process of conversion is collectively known as image digitization, after digitization the image is readily available in a form suitable for further processing by digital computers. We shall discuss about the concept of sampling and quantization in the subsequent sections, here we will discuss about the concept of image only.

Many times the term ‘picture’ is used, but this term relates to the analog or raw image data and the term ‘image’ is used to refer to digital data that is suitable for the processing of images using digital computers. Infact Images are imitations of a real world objects.Image may be considered as a projection of the real world (to be more precise 2D projection of 3D world). From a photographers point of view it is a photograph (i.e. projection of real world), and from the point of view of a computer engineer an image may be a two-dimensional (2D) signal. Therefore, an image is a two-dimensional function $f(x, y)$ where for

each position (x, y) in the projection plane, the values of the function $f(x, y)$ represent the amplitude or light intensity of the image.

Images are of two types, analog and digital. Analog image can be mathematically represented as a continuous range of values representing position and intensity. A digital image is composed of picture elements called **pixels**.

You may note here that through out the course we shall use the word image which refers to digital image.

As we mentioned earlier an image is a projection of the environment under consideration, if the environment is 3D then its projection will be in 2D. If we are having an n-Dimensional space then its projection will be of n-1 dimensions. Magnetic resonance images and computerized tomography (CT) images, which are 3D images, are mathematically represented by 3D function $f(x, y, z)$, where x, y, z are spatial coordinates which may represent the amplitude or intensity or any other parameter of the image.

A digital image is defined as a 2D discrete signal that varies over the spatial coordinates x and y, and can be written mathematically as $f(x, y)$. It is also an $n \times n$ array of elements, and each element represents the sampled intensity.

In general, the image $f(x, y)$ is divided into X rows and Y columns. The coordinate ranges are $\{X = 0,1,...,m-1\}$ and $\{Y = 0,1,2,...,n-1\}$. The image can be written as a mathematical function $f(x, y)$ as given below:

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & f(0,2)..... & f(0,y-1) \\ f(1,0) & f(1,1) & f(1,2)..... & f(1,y-1) \\ ..... & ..... & ..... & ..... \\ ..... & ..... & ..... & ..... \\ f(m-1,0) & f(m-1,1) & f(m-1,2) & f(m-1,n-1) \end{bmatrix}$$

A sample digital image is shown in Fig. 1(a) and its equivalent matrix is shown in Fig.1(b).

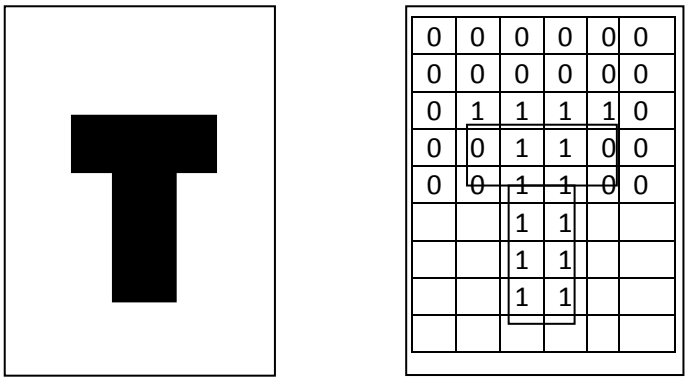


Fig. 1: (a)Sample Digital Image and (b) Matrix of the Image

In Fig. 1(b), the elements are shown as either 0 or 1, which represents the value of the function $f(x, y)$, each element is known as **pixels** i.e. 'picture element'. They represent the discrete data of any digital image, and serve as the actual building block of digital images. The value of the function $f(x, y)$ at every point indexed by a row and a column is a number and has no units, and it is known as the **grey value** of intensity of the image at that point.

You may again refer to Fig. 1(b) , the number of rows in a digital image is called **vertical resolution**. The number of columns is called **horizontal resolution**. We shall discuss about these concepts of image resolution, later in this unit. The number of rows and columns describes the **dimensions** of the image. The image size is often expressed in terms of the rectangular pixel dimensions of the array. Images can be of various sizes. Some examples of image size are 256 x 256, 512 x 512, etc. For a digital camera, the image size is defined as the total number of pixels (specified in megapixels). For example an image with resolution 2048x2048 will have $4x10^6$ pixels, that is, 4 Megapixels. Millions of pixels combine together to give a digital image, and their meaning varies with context i.e. a pixel can be considered a single sensor, photosite (physical element of the sensor array of a digital camera), element of a matrix, or display element on a monitor.

Generally, the value of the pixel is the intensity value of the image at that point, which is quantized value of the light that is captured by the sensor at the point. Typically the quantization is done in 256 levels. Thus the pixels will have values going from 0 to 255, and every pixel will need 8bits to store this information. But many times, the value of the pixel is not always the intensity value. For example, in the case of an X-ray image, the value of the pixel indicates the attenuation of the X-ray at the point. Similarly, the average Magnetic Resonance (MR) signal intensity denotes the pixel value in an MRI.

You may try the following exercises.

E1) What do you mean by the term image file format? Mention some of the frequently used image file formats.

E2) Differentiate analog image and digital image. Also, give example in support of your answer.

Now, in the following section, we shall discuss image acquisition.

1.3 IMAGE ACQUISITION

According to the fundamental concept of optics, when two or more rays of light meet at a point due to reflection or refraction or any other phenomenon, an image is formed. The information of happened phenomenon is recorded or acquired by the respective sensors as an image of that event. The process of capturing real world images and storing them into a computer is called **image acquisition**.

The question arises here, that how do we get the image in digital form. The answer is that the sensors are actually the electronic devices which taps the various types of signals, like thermal or optical or electromagnetic or any other type. Generally, the data is gathered in the analog form, which is digitized through a digitizer, and this digital signal is finally utilized by the digital computer for its processing. This leads to three types of image processing :–

- Optical Image Processing
- Analog Image Processing
- Digital Image Processing

We define these types briefly:

- **Optical Image Processing:** An optical image of 3D object is is in fact its 2D projection, which is the continuous distribution of light on a 2D surface. This 2D projection holds various types of information like 3D objects that are in focus. It also includes the study of the radiation source, and other optical processes. This optical image is available in the optical form until it is converted into analog form, leading to the area of analog image processing.

- **Analog Image Processing:** This relates to the processing of analog signals using analog circuits. Analog signals are the time varying continuous signals, often referred to as pictures. These analog signals are transformed in to digital image through the process of sampling and quantization, performed by using a digitizer and the process is known as digitization. Fig. 2 shows the steps of analog image processing.

Fig. 2: Analog Image Processing

You may note that the imaging systems that use film for recording images are also known as analog imaging systems, in medical imaging, still films are used, because these films provide better quality than digital systems.

- **Digital Image Processing:** It relates to the field of using digital circuits, systems, and software algorithms to carry out the image processing operations, which include quality enhancement of an image, counting of objects, image analysis, and many more.

Now, we know that when an interaction happens between object and rays of light (or signals) some optical or thermal or relevant phenomenon happens. The phenomenon is recorded or acquired by the respective sensors as an image of that event. This image relates to the combination of millions of pixels, each of which is carrying the information for function $f(x, y)$ or $f(x, y, z)$, which may be intensity or X-ray attenuation value, or Magnetic Resonance Intensity or any other parameter. This is gathered by using respective sensors or other

devices, this data gathering relates to the field of image acquisition, which is actually the first phase of Digital Imaging. This acquired information is stored in to the memory of the digital computer for necessary processing, the final outcome of processed image is transmitted to the concerned as an information. This means that the relevant tasks performed by any digital imaging system are **data acquisition**, **storage**, **manipulation**, and **transmission**.

Thus, a typical digital imaging system or digital imaging workstation involves components for the image acquisition, storage, processing, and transmission. Types of imaging systems and acquisition systems are shown in Fig. 3.

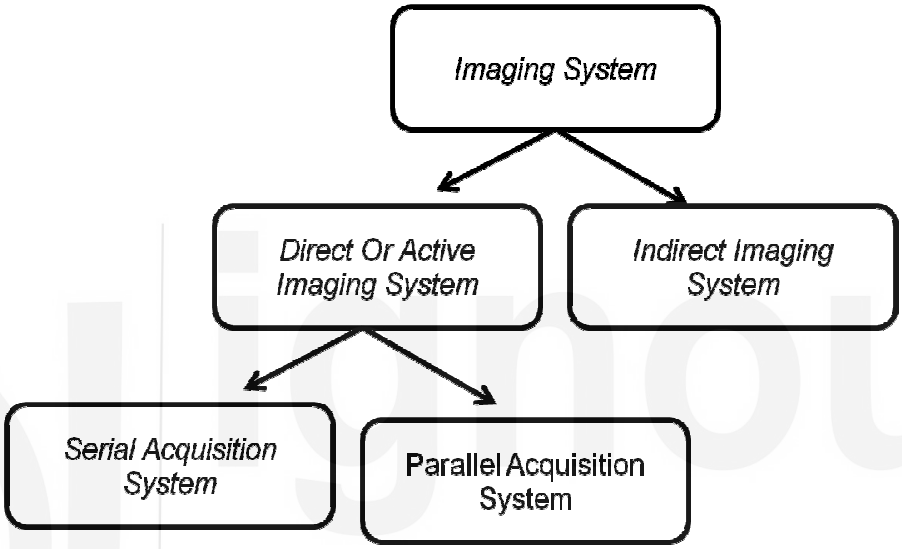


Fig. 3: Types of Imaging and Data Acquisition Systems

Let us now discuss these types briefly.

Direct or Active Imaging Systems: These are the systems where no temporary storage on film or chemical processes is required. In such systems the acquired images are in a recognizable form. We can say that in such systems the final image is a digital image. For example, human eye, digital camera, scanning confocal microscopes, etc.

The direct or active imaging systems are divided in to two sub categories i.e. serial and parallel acquisition systems. The human eye and the digital camera are example of parallel active imaging systems, where as scanning microdensitometer and scanning confocal microscopes are examples of serial acquisition imaging systems. The advantage of direct imaging systems is that the final image is a digital image.

Indirect Imaging Systems: In comparison to Direct imaging systems, the Indirect imaging systems involves data processing or reconstruction before producing an image for observation. In indirect imaging systems, the image is stored in a film, which is rendered observable by a chemical process. Therefore, there is a delay in the production of images from films.

You may now try the following exercises.

E3) Distinguish among optical image processing, analog image processing and digital image processing using examples.

E4) Differentiate between direct and indirect imaging system.

After data or image acquisition, image digitization is to be performed. We shall be using the image digitization process in image digitization.

1.4 DIGITIZATION OF IMAGES

In this section, we will try to find out the answers to three basic questions. The first question is why do we need digitization? Then we will try to find the answer to what is meant by digitization and thirdly, we will go to how to digitize an image.

We have learned from the previous sections that images are actually the imitations of real-world objects, which is generally a two-dimensional (2D) signal $f(x,y)$, where the values of the function $f(x,y)$ represent the amplitude or intensity at different points (pixels) of an image. In the processing of the analog signals or information by using digital computers, we need to convert this analog image into a discrete form using the process of digitization. This process of Digitization involves subprocess of sampling and quantization, performed through a Digitizer. The process of sampling and quantization transforms the acquired data in to the form which is suitable for further processing by digital computers.

Thus, image digitization refers to the process of sampling and quantization of the analog signals, which is required to convert the analog signal in to digital form. The digital image processing is performed using Digitizers, and this digitization process consists of two steps. One is called **sampling** and the other is called **quantization**. Sampling refers to considering the image only at a finite number of points, each image sample is called a pixel and Quantization refers to the representation of the grey level value at the sampling point using finite number of bits. For example 256 level quantization results in image storage of 8 bits per pixel.

Let us discuss the two concepts in detail.

Sampling

Consider any image. As you know that an image can be viewed as a 2 dimensional function given in the form of $f(x,y)$. Now, the image has certain length and certain height. Suppose the height of the image is H and the length of the image is L. The units of length and height would be the same. We can identify the coordinates X any Y at any space on the image. As you have seen the matrix representation of function

$f(x,y)$ in Sec. 1.2, x-axis is taken vertically and y-axis is taken horizontally, so X coordinate will vary from 0 to H and Y coordinate will vary from 0 to L. At any point (X,Y) we identify two features and write XY as product of them. Suppose these features are $r(X,Y)$ and $i(X,Y)$ represent the reflectance of the surface point and intensity of light respectively. If we consider that the maximum and minimum values of intensity are I_{max} and $I_{min,}$ then in case of continuous image $f(x,y)$ cannot attain the values either according to XY or $r(X,Y).i(x,y)$.

From the theory of real numbers you know that given any 2 points, there are infinite numbers of points. So again, when I come to this image as X varies from0 to H, there can be infinite possible values of X between 0 and H. Similarly, there can be infinite values of Y between 0 and L. So effectively, that means that if we want to represent this image in a computer, then this image has to be represented by infinite number of points and secondly when we consider the intensity value at a particular point, it varies between certain minimum I min and certain maximum Imax values, which is infinite in number.

It is clear from here that the number of points in any case would be infinite which would not be possible for computer to work on. Therefore, a way to overcome from this problem is to consider some discrete set of points, and this process of taking discrete data for any image is known as sampling.

From this discussion, we can say that since the computers can not handle the continuous data, so the continuous signal needs to be converted into digital signal and the process of this conversion is referred to as sampling.

We can visualize the process as overlaying an uniform grid on the image and sampling the image function at the center of each grid square. As we make the grid finer, we get better resolution of the image, but as we make the gridcoarser, we observe more "*pixelization*". At each pixel (or at each grid square) we usually represent the gray level value using an integer ranging from 0 for black to 255 for fully white.

To sample the signal, the signal should be frozen, as sampling cannot be applied to moving signals. If the signal is frozen for T seconds, T is called the sampling period. The **sampling period** is measured in seconds, milliseconds, or microseconds. For example, if the sampling rate is 1000Hz, it means that the signal is sampled every millisecond.

The **sampling rate** or **sampling frequency** is the reciprocal of the sampling period and is measured in samples per second or Hertz. The sampling process is the multiplication of a continuous signal $f(t)$ by a railing function $r(t)$. Railing function can be a pulse of unit amplitude at exactly the sampling time instant. We can also say that If the signal is frozen for T seconds ($T$is the sampling period) thensampled function $f(n)$ is mathematically represented as the product of continuous signal $f(t)$ by a railing function $r(t)$,

$$f(n) = f(t) \times r(t) = f(nT)$$

All these components are graphically shown in Fig. 4.

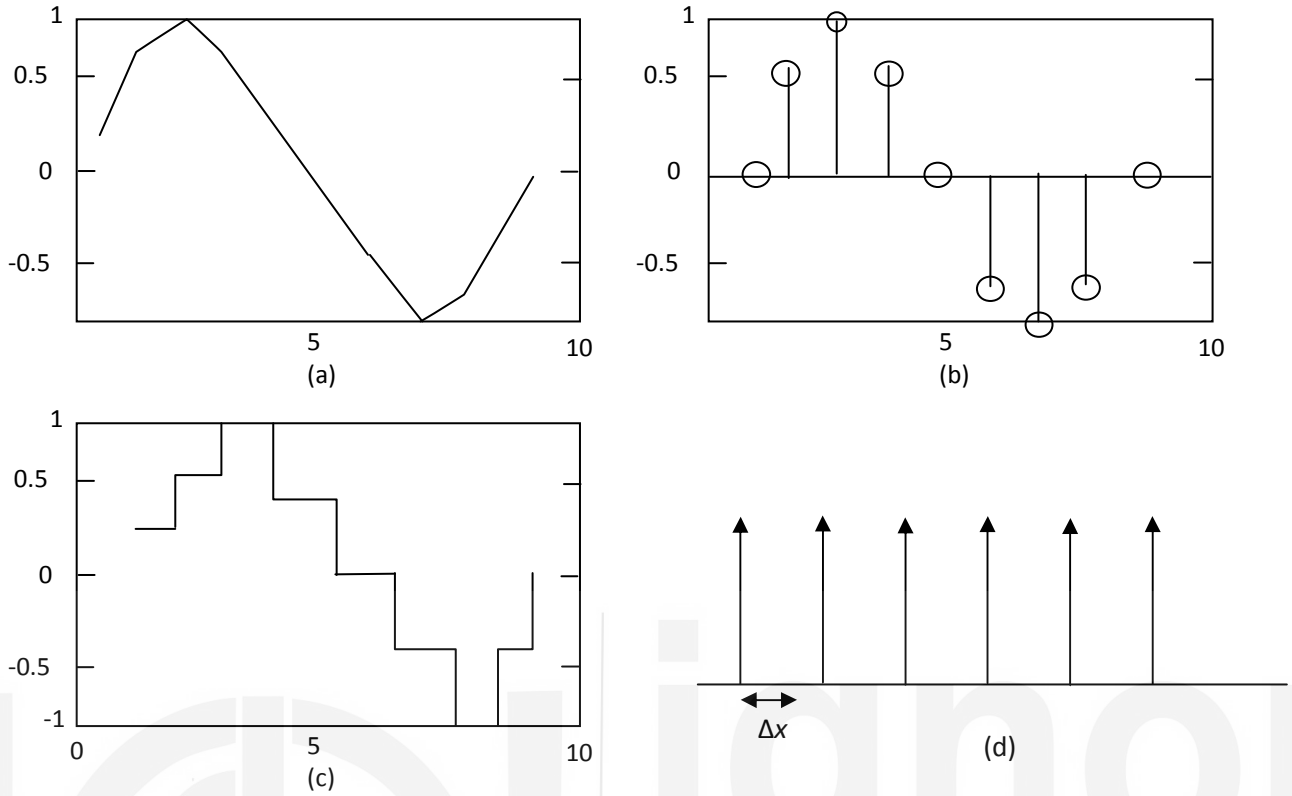


Fig. 4: **Sampling process (a) Original signal** $f(t)$ **(b) Sampled image** $f(n) = f(t) * r(t)$ **(c) Reconstructed image (d) Train of the impulse function *r*(t)**

Since, sampling is a reversible process, the reconstruction of original signal from the sampled signal is possible in both frequency and time domain. In frequency domain, the information of the base band component is required, and this base-band component over the band $\pm f_s$ (where f_s is the sampling frequency) can be extracted from the infinite spectrum of the infinite spectrum of the sampled signal, by using the low pass filters.(This statement is derived from the Shannon's Sampling Theorem).In image processing, the sampling frequency decides the distance between the samples. This distance determines the linear pixel size. Alternatively, the equivalent operation in time domain is the interpolation operation where an ideallow-pass filter is used in convolution with the interpolation function (e.g., sine function). The function $f(t)$, the railing function $r(t)$, and the sampled and reconstructed functions are all shown in Fig. 5.

This idea of one-dimensional sampling can be extended to 2D images also. The 2D railing function is known as **comb function**. It is arranged as a rectangular grid of unit impulses separated by Δx and Δy. The 2D sampling process can again be viewed as the multiplication of the railing function with the continuous function to give discrete samples. The values of Δx and Δy play an important role in image processing. Normally they are kept the same in both the horizontal and the vertical directions, so that the pixels will be a square pixels. This is called **pixel aspect ratio**. In addition, the size of the pixel is important for image quality. If the size is very large, there will be a lesser number of pixels. Hence, the details become less, which makes the image distracting and

meaningless. This is called **pixelization error** where the grey level discontinuities at the edges of the pixel become poor. Then question arises that, What should be the ideal size of the pixel? Should it be big or small? The answer is give by the Shannon-Nyquist theorem.

As per Shannon-Nyquist theorem, the sampling frequency should be greater than or equal to $2\times f_{max}$, where f_{max} is the highest frequency present in the image. Otherwise, the original signal cannot be reconstructed. In other words, the number of samples required is dictated by Shannon-Nyquist theorem, this sampling theorem can be stated in terms of distance d as $d \le \frac{1}{2f_{max}}$. Alternatively, it should be less than or equal to the smallest detail that is present in the image.

Another frequency that is helpful in image processing is the Nyquist frequency. The Nyquist frequency (f_N) is $1/2\times$ (sampling frequency). Therefore, the Nyquist frequency (f_N) should be greater than or equal to f_{max}. If the sampled image has frequencies higher than the Nyquist frequency, it results in a condition called aliasing where the high frequencies masquerade as low frequency components. This results in an image where interpretation becomes difficult. This problem is called **aliasing.**

Now, let us solve the following examples to understand this concept better.

Example 1: What should be the physical size of a 2D image of a document with dimensions ~~is~~ 2400×2400, when scanned at 300dpi. Here dpi stands for dots per inch.

Solution: The physical size of image is

$$= \frac{\text{Number of pixels in width}}{\text{Resolution}} \times \frac{\text{Number of pixels in height}}{\text{Resolution}}$$

$$= \frac{2400}{300} \times \frac{2400}{300}$$

$$= 8\,\text{inches} \times 8\,\text{inches}$$

***

Example 2: If the physical size of a medical image is 8×8 inches and the sampling resolution is 5 cycles/mm, then how many pixels per cycle are required to have a better quality image? Will an image of size 256×256 be enough?

Solution: It is given that the sampling resolution is 5 cycles/mm. Therefore, for better quality, 2 pixels per cycle are required. This is because sampling theorem states that the sampling frequency should be greater than twice the maximum signal frequency. This means that 10 pixels per mm are required. So, the pixel size is 0.1 mm.

You may note here that this double rates in both directions are called Nyquist rates.
The given image size (since $1\text{inch} = 2.54\text{cm}$) is

$= 8 \times 2.54 \times 8 \times 2.54 \text{ cm}^2$

$= 20.32 \times 20.32 \text{ cm}^2$

Therefore, the minimum number of required pixels $= 2032 \times 2032$ pixels.
So, an image of size 256×256 is not enough.

***

Now try the following exercises.

E5) The dimension of an image is 5 x 8 inches and the frequency is 500 dots per inches in each direction. Find the number of samples required to preserve the information in the image.

Now let us discuss the other step of image digitization, which is image quantization.

Quantization

Image quantization is the process of converting the sampled analog value of the function $f(x, y)$ into a discrete-valued integer. An analog signal has an infinitely larger number of distinct values. So, it is necessary to convert the continuous values into a smaller set through the process of quantization. The sampled analog image i.e. a natural image has continuously varying shades and colours and is known as a continuous tone image. This continuous tone image is required to be converted into a discrete image. The image quantizer maps a continuous value x into a discrete variable x', where discrete points of grey-tone or brightness are used.

The process of quantization involves the partitioning of input values into equally spaced intervals. The end points of the interval are called **decision boundaries**. Let the decision boundaries be given by $B = \{b_0, b_1, ..., b_m\}$, and let the input values be in the range $-X_{max}$ to $+X_{max}$. The length of the interval between successive decision boundaries i.e. the step size (Δ) is given by $\frac{2X_{max}}{m}$. The midpoint between successive decision boundaries is called **output or reconstruction level** and is given by $(R) = \{y_1, y_2, .., y_m\}$. Then quantization is in the range $\left[\text{from} -\frac{\Delta}{2} \text{to} \frac{\Delta}{2}\right]$, and the number of bits necessary to encode the output levels is given by $R = [\log_2 m]$.

Try the following exercises.

E6) What is the role of Sampling and Quantization in the process of Digitization?

E7) What do you mean by pixel aspect ratio and pixelization error?

E8) How Shannon-Nyquist theorem relates to the determination of the ideal size of the pixel?

In the following section, we shall discuss the ways how an image is represented.

1.5 REPRESENTATION OF DIGITAL IMAGE

In conventional image processing, a matrix is used to represent intensity values. Many times we need to convert this matrix to a vector and then use the normal equations or regularization based methods to represent intensity values. The normal matrix representation of image is called the Spatial Domain. There are other representation of images obtained by transforming the spatial domain image like Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) etc. Each of these transformations represents the image in the form of a mathematical formula. Further, some Partial Differential Equations (PDE) based methods use kernels to process images using 2D convolution. Sometimes we use the notion of images as point clouds and then construct a graph based on some distance measure. Then we study the Laplacians (Diffusion maps) or the Hamiltonian operators of these graphs. The Eigen values of these operators lead to the notion of heat kernel signatures and wave kernel signatures. We have these alternate views of an image based on the application, we will study some of these image representation mechanism in this section.

Going back to the basics, a digital image represents a mapping on a visual scene recorded by an optical sensor into an array of picture element intensities. Here we are discussing some of the models of image representation.

- **Pixel model:** In this model, the RGB optical sensors maps to a vector of discrete intensities which are integers. Let $p : R \times G \times B \to Z \times Z \times Z$ defined by $p(r,g,b) = (zR, zG, zB)$, where r, g, b are real valued colour intensities and zR, zG, zB are integer-valued colour intensities in the range from 0 to 255. In this case, a pixel is represented by a 1×3 column vector used to "paint" a tiny square in an image display.
- **Vector model:** In this model, the RGB optical sensor values are mapped to vectors and the displayed image conforms to a form of triangulated image space instead of a collection of pixels. In this model, a vector image is a collection of vertices and edges. The end result of vector images is the formation of a basis for applications such as Adobe Photoshop and various video games.

You can try the following exercise.

E9) Compare Pixel Model and Vector Model for the representation of Digital Image.

In the following section we will discuss about the various types of images, and their respective characteristics.

1.6 TYPES OF IMAGES

Classification of images can be performed on the basis of various criteria like attributes, colour, dimension and data types. Based on Attributes criteria the images are classified as raster and vector, whereas on the basis of colour we can classify an image as binary, greyscale, true colour and pseudo colour. Further on the basis of dimensions, we have 2D images and 3D images, and on the basis of data types the images are classified as signed integer type, unsigned integer type, float type, logical and double type. So there is no single accepted way of classifying images. The image classification is described in detail now:

- **Classification of Images on the basis of Attributes:** Based on attributes of any image, it is classified as raster and vector graphic images.
    - **Vector graphics** uses graphic primitives (point/line/circle/ellipse etc)to describe an image. Hence, the notion of resolution is practically not present in graphic.
    - **Raster images** are pixel-based, and hence their quality is dependent on the number of pixels. So, operations such as enlarging or blowing-up of a raster image often result in quality reduction.
- **Classification of images on the basis of colour:** On the basis of colour, the images can be classified into the following categories:
    - **Monochrome images** are the images where the colour component is absent, they are further classified as Grey scale and binary images.
    - **Grey scale images** : The term grey scale refers to the range of shades between white and black or vice versa, such images have many shades of grey, and eight bits ($2^8 = 256$) are enough to represent the grey scale because the human visual system can not distinguish more than 32 different grey levels, and the additional bits are necessary to cover noise margins. Most medical images such as X-rays, CT images, MRIs, and ultrasound images are grey scale images.
    - **Binary images** :The binary images are just special case of grey scale images, where the process of Thresholding is applied, they are actually Bi-level images where the pixels assumes the values of 0 or 1, i.e. only one bit is sufficient to represent the pixel value. The thresholding process involves the comparison of the threshold value i.e. the pixel value is compared with the threshold value, if the pixel value of the grey scale image is greater than the threshold value, the pixel value in the binary

image is considered as 1, Otherwise, the pixel value is 0. So far as the utility of Binary images is concerned, they are often used in representing basic shapes and line drawings. They are also used as masks. In addition, image processing operations produce binary images at intermediate stages.

- **True colour (or full colour)** images are the images where the pixel has a colour that is obtained by mixing the primary colours red, green, and blue. Each colour component is represented like a grey scale image using eight bits. Mostly, true colour images use 24 bits to represent all the colour. Hence true colour image can be considered as three-band images. The number of colours that is possible is 256^3 (i.e. 256*256*256= 1,67,77,216 colours). A display controller then uses a digital-to-analog converter (DAC) to convert the colour value to the pixel intensity of the monitor, refer to Fig.5 .

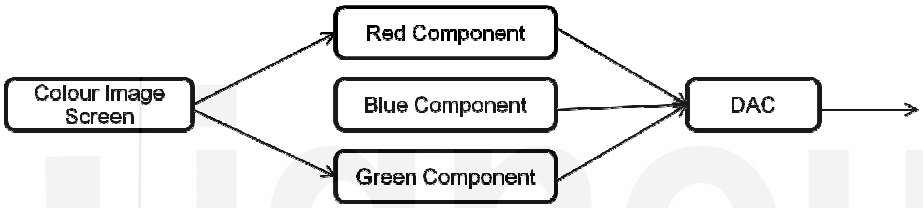


Fig. 5: True ColourImage

True colour images carries the full range of available colours in themselves. Thus, they are quite similar to the actual object and hence called true colour images. Further, the true colour images do not use any lookup table but store the pixel information with full precision.

- **Pseudocolour images** are infact false colour images, their colour component is manipulated artificially, based on the interpretation of data. Like true colour images, pseudocolour images are also used widely in images processing. We know that true colour images are called three-band images. But, in remote sensing application, multi-band images or multi-spectral images are generally used. These images, which are captured by satellites, contain many bands. A typical remote sensing image may have 3 to 11 bands in an image.This information is beyond the human perceptual range. Hence it is mostly not visible to the human observer. So colour is artificially added to these bands, so as to distinguish the bands and to increase operational convenience. These are called artificial colour or pseudocolour images. Pseudocolour images are popular in the medical domain also. For example, the doppler colour image is a pseudocolour image.

We shall discuss few examples for the better understanding of the topic.

Example3: Determine the storage space required to store 1024 x1024 pixels of a binary image?

Solution: For a binary image, one bit is sufficient for representing the pixel value. So, the number of bits required will be 1024 x 1024 x 1 =

10,48,576 bits = [(10,48,576)/8]bytes = 1,31,072 bytes = 128 KB.Since, 1 KB is 1024 bytes, the storage requirement is 128 KB.

***

Example 4: What is the storage requirement for a 1024 x 1024, 24-bit colour image?

Solution: Since colour images are three-band images (red, green, and blue components), the storage requirement is 1024*1024*3 bytes = 31,45,728 bytes. Since 1 KB is 1024 bytes, the storage requirement is 3072 KB or 3 Mega pixels.

***

We shall resume the discussion on image classification.

- **Classification of Images on the basis of Dimensions :**Images can be classified on the basis of dimensions also. Normally, digital images are a 2D rectangular array of pixels. If another dimension, of depth or any other characteristic, is considered, which may be necessary to use, then a higher-order stack of images like 3D images are produced. A good example of a 3D image is a volume image, where pixels are called voxels. By '3D image', it is meant that the dimension of the target in the imaging system (may be a scene or an object.) is three dimensional $(x, y,$ depth). In medical imaging, some of the frequently encountered 3D images are CT images, MRIs, and microscopy images. These are stored basically as 2D image slices taken across the body or the skull. Range images (often used in remote sensing application) are also 3D images as they also incorporate the depth information.

- **Classification of Images on the basis of Data Types:** Images may be classified based on their data type. Sometimes, image processing operations produce images with negative number, decimal fraction, and complex number. For example, Fourier transforms produce image involving complex numbers. To handle negative numbers, signed and unsigned integer types are used. In these data types, the first bit is used to encode whether the number is positive or negative. For example, the signed data type encodes the numbers from – 128 to 127 where one bit is used to encode the sign. In general, an $n-1$ bit signed integer can represent integers from -2^{n-1} to $2^{n-1}-1$, a total of 2^n. Unsigned integers represent all integers from 0 to 2^n-1 with n bits.

  Floating-point involves storing the data in scientific notation. For example, 1,230 can be represented as $0.123*10^4$, where 0.123 is called the significant and the power is called exponent. There are many floating-point conventions.

  The quality of such data representation is characterized by parameters such as data accuracy and precision. Data accuracy is the property of how well the pixel values of an image are able to represent the physical properties of the object that is being imaged. Data accuracy is an important parameter, as the failure to capture the actual physical properties of the image leads to the loss of vital

information that can affect the quality of the application. While accuracy refers to the correctness of measurement, precision refers to the repeatability of the measurement. In other words, repeated measurements of the physical properties of the object should give the same result. Most software use the data type 'double' to maintain precision as well as accuracy.

Now try the following exercises.

E10) Make a list of all the types of classifications of images based on various parameters.

E11) Compare true Colour Images and Monochromatic Images.

In this section, we learned various types of images. Now it is time to learn about the fundamental characteristics of images because to improve the quality of an image, it is necessary to assess its quality, and this assessment requires understanding of characteristics of images, which is discussed in the following section.

1.7 IMAGE CHARACTERISTICS

To improve the quality of an image, it is necessary to assess its quality, and this assessment requires understanding of characteristics of images. some of the essential characteristics of image are Intensity, Contrast, Brightness, Noise and Resolution, we will discuss them one by one.

- **Intensity:** The term Intensity refers to the amount of light or the numerical value of a pixel, it is the measure of energy of a wave, which is directly proportional to square of amplitude of the signal. From the point of view of image processing it is a numerical value which represents a pixel, for *example* for an 8-bit gray scale image the value of a pixel can be in the range of [0,255], where (say) numericvalue "0" represents black colour and the numeric value "255" represents white colour. Therefore, *the higher the value of a pixel(i.e. intensity) whiter the* pixel will appear, i.e. in a grey scale images, intensity is depicted by the grey level value at each pixel i.e., two pixels with grey level value 127 and 220, then it can be interpreted that pixel with value 127 is darker than pixel with value 220 .

- **Contrast:** The term Contrast of an image relates to recording of the differences in the magnitude of the intensity at the surface of an object. There are many ways in which contrast can be measured,i.e.it can be described as a product of the sensor signal contrast (this depends on the energy source and the physical properties of the object) and detector contrast (this depends on the way signal is detected, captured, and stored). But, a common

measurer of contrast (C) involves intensity of foreground (I_{object}) and background $(I_{background})$ objects, i.e.

$$C = \frac{I_{object} - I_{background}}{I_{object} + I_{background}}$$

Where, I_{object} is the average pixel intensity of the object pixels and $I_{background}$ is the average pixel intensity of the background. Another useful contrast measure using the same parameters I_{object} and $I_{background}$ is

$$C = \log_{10} \frac{I_{object}}{I_{background}}$$

We know, Contrast of an image relates to recording of the differences in the magnitude of the intensity at the surface of an object. *So, Contrast is the difference between maximum and minimum pixel intensities in an image.* Consider two images A& B having pixel intensities between 30 to 200 and 70 to 130, respectively. Then A has more contrast than B. Again contrast is also relative. Contrast can be simply explained as the difference between maximum and minimum pixel intensity in an image. For example, consider an image, whose maximum and minimum value of intensity is 100 i.e. same then contrast will be zero because, Contrast = maximum pixel intensity – minimum pixel intensity $= 100 - 100 = 0.$

- **Brightness: It** refers to the average pixel intensity of the image, Similar to contrast, the image should be reasonably bright to show all the information to the viewer. However, excessive brightness may affect the quality of the image. An image may have a higher brightness but if the intensity is not optimal, the image again will not exhibit good quality. *Brightness resolution* is defined as the number of identified distinct colours and it is also called as *colour resolution*.

  Brightness is a relative term which can be understood visually you can say the higher the intensity the brighter is the pixel. Since brightness is a relative term, so brightness can be defined as the amount of energy output by a source of light relative to the source we are comparing it to. In some cases we can easily say that the image is bright, and in some cases, it's not easy to perceive. Now, Question is How to make an image brighter? Brightness can be simply increased or decreased by simple addition or subtraction, to the pixel values of the image matrix. Consider a black image of 5 rows and 5 columns. All the entries of the image matrix are going to be zero since it is a completely black image. To make it brighter we shall perform some operation onthis mage. What we will do is, that we will simply add a value of 50 (or any value from 1 to 255) to each of the matrix value of image. After adding the image would become somewhat brighter then its previous state of brightness.

- **Noise:** *Noise* is an unwanted disturbance that causes fluctuations in the pixel value. It is a random or stochastic process and hence its true value cannot be predicted accurately. However, noise obeys all the statistic properties. So a pixel is characterized as a random

variable for statistical analysis. It is the next important characteristic, related to the quality of image, Image applications are frequently affected by the noise present in the image.

- **Resolution:** Resolution refers to the number of pixels in an image. Resolution is sometimes identified by the width and height of the image as well as the total number of pixels in the image. For example, an image that is 2048 pixels wide and 1536 pixels high (2048 x 1536) contains (multiply) 3,145,728 pixels (or 3.1 Megapixels). You could call it a 2048 x 1536 or a 3.1 Megapixel image.

You may now try the following exercises.

E12) What do you understand by the term “Brightness of Image”? How it is different from the Contrast of any Image?

So far, we have been using the word resolution and discussed it at several places. In the following section, we shall highlight resolution in detail.

1.8 IMAGE RESOLUTION

How does image resolution play out on my computer monitor? The computer screen you are looking at right now is set at a particular resolution as well. The larger the screen, the larger you likely have your screen resolution set. If you have a 17" monitor, it is likely you have it set at 800 x 600 pixels. If you have a 19" screen it is likely set at 1024 x 768. You can change the settings but these are optimum for those screen sizes.

Now, if your monitor is set to 800 x 600 and you open up an image that is 640 x 480, it will only fill up a part of your screen. If you open up an image that is 2048 x 1536 (3.1 megapixels) then you will find yourself moving the slider bar around to see all the different parts of the image. It just won't fit. Add to that the fact that the computer monitor has a finite number of pixels per inch available (like 72) so if you are going to display your image on a monitor only, you would have to reduce the quality down to 72 to save file space. If you are going to put your image on a webpage or email it to a friend then you will want to first make it a useful size. Not too big, not too small. May be 200 or 300 pixels high would be a nice size. This means that Image resolution is highly related to the size of image and its pixel density.

We learned from the last section that the term Image Resolution relates to pixel density i.e. the number of pixels in an image, which relates to clarity of image. Parameters used to determine the Resolution of any image involves the dimensions of image i.e. its width and height and the total number of pixels in the image, as well. Say, an image has following dimensions i.e. 2048 pixels wide and 1536 pixels high thus total number of pixels in the image is (2048 x 1536) i.e. 3,145,728 pixels (or 3.1 Megapixels), i.e the image has resolution of 3.1 megapixels. As the

megapixels in the pickup device in your camera increase so does the possible maximum size image you can produce. This means that a 5 megapixel camera is capable of capturing a larger image than a 3 megapixel camera.

In general the Resolution is classified as Spatial Resolution and Intensity. In simple terms, Spatial Resolution relates to the concentration of pixels in the creation of digital images and Intensity resolution of an image relates to blurredness or sharpness of an image, depending on the intensity of resolution.

- **Intensity resolution** deals with the Intensity of resolution i.e. the number of pixels per square inch, which determines the clarity or sharpness of an image. Whereas, Spatial resolution refers to the number of pixels used in making an image.

- **Spatial resolution**, in general is defined as the number of pixels per inch of an image, the term deals with the number of pixels, pixel density, and quantization levels. Thus, it depends on the sampling and quantization processes. If the spatial resolution is not satisfactory, the quality of the image would not be satisfactory. In General, if the image has more pixels, the quality of the image is supposed to be higher. but, this is not correct, because an image may have more pixels, but it may not be providing sufficient information regarding other factors such as optical resolution, the distance of an object from the camera ,the field of view etc. which plays an important role in defining image quality.

Images with a higher number of pixels per square inch are sharp and hence said to have a higher Spatial resolution i.e. Spatial resolution is the number of pixels used for the construction of a digital image. Images that have a higher spatial resolution are made of a greater number of pixels and they are much clear as compared to the images with lower spatial resolution. In radiology, spatial resolution is used to differentiate between objects located close to each other. What do we understand by ~~what~~ spatial resolution? If we have to compare two images, to see which one is more clear or which has more spatial resolution, we need to bring the two images to same size.

Since, the spatial resolution is a measure to the clarity of image, so for different devices, different measure has been made to measure it. For example Dots per inch(Dots per inch or DPI is usually used in printers), Lines per inch(Lines per inch or LPI is usually used in laser printers), Pixels per inch (Pixel per inch or PPI is measure for different devices such as tablets , Mobile phones etc.).The higher is the PPI, the higher is the quality. In order to understand that how PPI is calculated, lets calculate the PPI of a mobile phone.

First of all we will use Pythagoras theorem to calculate the diagonal resolution in pixels. It can be given as $C = \sqrt{a^2 + b^2}$, where a and b are the height and width resolutions in pixel and C is the diagonal resolution in pixels. Now we will calculate PPI
PPI = C / diagonal size in inches

We shall see the following example.

Example 5: Calculate PPI or pixel density. For the sample mobile phone with the height and width resolutions in pixel as 1080 x 1920 pixels, and the diagonal size in inches of sample mobile phone is 5.0 inches.

Solution: So, putting those values in Pythagoras theorem i.e. the equation $C = \sqrt{a^2 + b^2}$ gives the result C = 2202.90717, Given, the diagonal size in inches of sample mobile phone is 5.0 inches, therefore PPI = 2202.90717/5.0 = 440.58 = 441 (approx).

That means that the pixel density of sample mobile phone is 441 PPI.

***

Pixel resolution, the term resolution refers to the total number of count of pixels in an digital image. For example. If an image has M rows and N columns, then its resolution can be defined as M X N. If we define resolution as the total number of pixels, then pixel resolution can be defined with set of two numbers. The first number the width of the picture, or the pixels across columns, and the second number is height of the picture, or the pixels across its height (rows).We can say that the higher is the pixel resolution, the higher is the quality of the image.We can calculate mega pixels of a camera using pixel resolution.

Pixel resolution in Megapixels = {Column pixels (width) X Row pixels (height)} / 1 Million.

Thus the size of an image can also be defined by using its pixel resolution.

Size of image = pixel resolution X bpp(bits per pixel)

Example 6: Calculate pixel resolution of a camera in mega pixels, capturing an image of dimension: 2500 X 3192.

Solution: Its pixel resolution = 2500 * 3192 = 7982350 bytes. Dividing it by 1 million = 7.9 = 8 mega pixel (approximately).

***

Now, let us discuss some important terms like *Aspect ratio, Dots per Inch, Lines per Inch*, etc. which are closely associated with the concept of image resolution

- Aspect ratio *is another important concept with the pixel resolution*. Aspect ratio is the ratio between width of an image and the height of an image. It is commonly explained as two numbers separated by a colon (8:9). This ratio differs in different images, and in different screens. The common aspect ratios are: 1.33:1, 1.37:1, 1.43:1, 1.50:1, 1.56:1, 1.66:1, 1.75:1, 1.78:1, 1.85:1, 2.00:1, etc.

  Advantage of Aspect ratio is that it maintains a balance between the appearance of an image on the screen, means it maintains a ratio

between horizontal and vertical pixels. It does not let the image get distorted when aspect ratio is increased. Aspect ratio tells us many things. With the aspect ratio, you can calculate the dimensions of the image along with the size of the image.

Example 7: Given an image is an gray scale image with aspect ratio of 6:2 and pixel resolution of 480000 pixels, calculate the following:

a) Resolve pixel resolution to calculate the dimensions of image
b) Calculate the size of the image

Solution: Following are given:
Aspect ratio (i.e $c:r) = 6:2$

Pixel resolution (i.e. $r \times c) = 480000$

We know that bits per pixel for a grayscale image is 8bpp.

Here, we are required to find the number of rows(r) and the number of columns(c).

Using the Aspect ratio, we get $c:r = 6:2$ that is $c = 6r/2$

Using the pixel resolution, we get $c = 480000/r$

Comparing both relations in c and r, we get

$$\frac{6r}{2} = \frac{480000}{r} \Rightarrow r^2 = \sqrt{\frac{(480000*2}{6}} \text{ i.e. } r = 400$$

Putting $r = 400$, we get $c = 1200$;

Thus, Rows $= 400$ and Columns $= 1200$

Now we solve for the size of image.

Size of image in bits = rows * cols * bpp

Size of image in bits $= 400*1200*8 = 3840000$ bits

Size of image in bytes = Size of image in bits /8 = 480000 bytes

Size of image in kilo bytes $= 469$kb (approx).

***

- **Dots per inch(DPI):** If you have an image that is 640 x 480. How big a print can you make? Well, the true answer is you can make as big a print as you want but very quickly you will start to see "blocks" (pixelization) and the quality will drop off. To maximize the capability of your printer, you should print a picture a size that the printer can handle. Here we introduce a new term "dots per inch" (dpi) or "pixels per inch" (ppi).

The DPI is often relates to PPI (Pixels Per Inch), whereas there is a difference between these two. DPI or dots per inch is a measure of spatial resolution of printers. In case of printers, DPI means that how many dots of ink are printed per inch when an image gets printed out from the printer. Remember, it is not necessary that each Pixel per inch is printed by one dot per inch. There may be many dots per inch used for printing one pixel. The reason behind this that most of the colour printers uses CMYK model. The colours are limited. Printer has to choose from these colours to make the colour of the pixel whereas within pc, you have hundreds of thousands of colours. The higher is the dpi of the printer, the higher is the quality of the printed document or image on paper. Usually some of the laser printers have dpi of 300 and some have 600 or more.

Example 8: Determine the optimum size of the 640 X 480 image that can be printed at 200 DPI.

Solution: We have a 640 x 480 image and you want to print it at 200 dpi. 640 divided by 200 equals 3.2 and 480 divided by 200 equals 2.4 so if you print this picture at 3.2" x 2.4" you will get a print with 200 dots per inch. We recommend 200 dpi as a minimum for good quality prints.

***

Example 9 : Let's say we want to print an 8" X 10" picture at 300 dpi. What resolution must we have?

Solution:300 times 8 is 2400 and 300 times 10 is 3,000. So, we would need a 3,000 x 2,400 image i.e., 7.2 megapixels.

***

- **Lines per inch:** As dpi refers to dots per inch, liner per inch refers to lines of dots per inch. The resolution of halftone screen is measured in lines per inch. Table 2 shows some of the lines per inch capacity of the printers.

Table 2

Printer	LPI
Screen printing	45-65 lpi
Laser printer (300 dpi)	65 lpi
Laser printer (600 dpi)	85-105 lpi
Offset Press (newsprint paper)	85 lpi
Offset Press (coated paper)	85-185 lpi

Try an exercise.

E13) What do you understand by the term " Resolution of an Image"? How Intensity Resolution differs from Spatial Resolution?

In this section we learned about various terminologies related to the Image Resolution, now its time to understand, how the pixels are related to each other ? The Relation between the pixels is discussed in the subsequent section.

1.9 RELATIONS BETWEEN THE PIXELS

Image topology is the branch of image processing that deals with the study of Relation between the Pixels, it relates to the study of the fundamental properties of the image such as image neighbourhood, paths among pixels, boundary, and connected components. It characterizes the image with topological properties such as neighbourhood, connectivity, distance etc.

- Neighbourhood is fundamental to the understanding of image topology. In the simplest case, the neighbours of a given reference pixel are those pixels with which the given reference pixel shares its edges and corners.

In $N_4(p)$, the reference pixel $p(x,y)$ at the coordinate position (x,y) has two horizontal and two vertical pixels as neighbours. This is shown graphically in Fig.6.

$$\begin{pmatrix} 0 & X & 0 \\ X & p(x,y) & X \\ 0 & X & 0 \end{pmatrix}$$

Fig. 6: 4-Neighbourhood $N_4(p)$

The set of pixels $\{(x+1,y),(x-1,y),(x,y+1),(x,y-1)\}$, called the 4-neighbours of P, is denoted as $N_4(p)$. Thus 4-neighbourhood includes the four neighbours of the pixel $P(x,y)$. By duality, these are pixels that share a common edge with the given reference pixel $p(x,y)$. Further we can also consider the 4 diagonal neighbors. They are $(x-1,y-1),(x+1,y+1),(x-1,y+1)$, and $(x+1,y-1)$. The diagonal pixels for the reference pixel $p(x,y)$ are shown graphically inFig. 7.

$$\begin{pmatrix} X & 0 & X \\ 0 & p(x,y) & 0 \\ X & 0 & X \end{pmatrix}$$

Fig. 7: Diagonal Elements $N_D(p)$

The diagonal neighbours of pixel $p(x,y)$ are represented as $N_D(p)$. The 4-neighbourhood $N_4(p)$ and $N_D(p)$ are collectively called the 8-neighbourhood. This refers to all the neighbours and pixels that share a common corner with the reference pixel $p(x,y)$. These pixels are called indirect neighbours. This is represented as $N_8(p)$ and is shown graphically in Fig. 8, this set of pixels $N_8(x) = N_D(x) \cup N_4(x)$.

$$\begin{pmatrix} X & X & X \\ X & p(x,y) & X \\ X & X & X \end{pmatrix}$$

Fig. 8: 8-Neighbourhood $N_8(p)$

- Connectivity is the next important property where the relationship between two or more pixels is defined by pixel connectivity. Connectivity information is used to establish the boundaries of the objects. The pixels *p* and *q* are said to be connected if certain conditions on pixel brightness specified by the set V and spatial adjacency are satisfied. For a binary image, this set V will be $\{0,1\}$ and for grey scale image, V might be any range of grey levels.

  - **4-Connectivity:**The pixels p and q are said to be in 4-connectivity when both have the same values as specified by the set V and if q is said to be in the set $N_4(p)$. This implies any path from p to q on which every other pixel is 4-connected to the next pixel.

  - **8-Connectivity:**It is assumed that the pixels p and q share a common grey scale value. The pixels *p* and *q* are said to be in m-connectivity if q is in the set $N_8(p)$.

- Subsequently **Distance measures** are vital component, and the distance between the pixels p and q in an image can be given by distance measures such as Euclidian distance D_e, D_4 distance, and D_8 distance. Consider three pixels p, q and z. If the coordinates of the pixels are $P(x,y), Q(s,t),$ and $Z(u,w)$ as shown in Fig. 9, the distance between the pixels can be calculated.

$$\begin{matrix} 0 & 1 & 1 & 1\,(z) \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1\,(q) \\ 1 & 1 & 1 & 1 \\ (p) & & & \end{matrix}$$

Fig. 9 Sample image

  - The *Euclidean distance* between the pixels p and $q(D_e(p,q))$, with coordinates (x,y) and (s,t), respectively, can be defined as $D_e(p,q) = \sqrt{(x-s)^2 + (y-t)^2}$
  - The advantage of the Euclidean distance is its simplicity. However, since its calculation involves a square root operation, it is computationally very costly.
  - The D_4 *distance or city block distance* can be simply calculated as $D_4(p,q) = |x-s| + |y-t|$
  - The D_8 *distance or chessboard distance* can be calculated as $D_8(p,q) = \max(|x-s| + |y-t|)$

The distance function can be called metric if the following properties are satisfied:

1. $D(p,q)$ is well-defined and finite for all p and q.
2. $D(p,q) \geq 0$ if $p = q$, then $D(p,q) = 0$.
3. The distance $D(p,q) = D(q,p)$.
4. $D(p,q) + D(q,z) \geq D(p,z)$. This is called the property of triangular inequality.

Example 10: Let $V = \{0,1\}$. Compute the D_e, D_4, and D_8 distances between pixels p and q. The coordinates of *p* and *q* be (3, 0) and (2, 3), Find the distance measures, for the image given below

	0	1	2	3
0	0	1	1	1
1	1	0	0	1
2	1	1	1	**1 (q)**
3	**1**	1	1	1

(p)

Solution : The Euclidean distance is $D_e = \sqrt{(x-s)^2 + (y-t)^2}$

$$D_e = \sqrt{(3-2)^2 + (0-3)^2} = \sqrt{1+9} = \sqrt{10}$$
$$D_4 = |x-s| + |y-t| = |3-2| + |0-3| = 1+3 = 4$$
$$D_8 = \max(|x-s|, |y-t|) = \max(|3-2|, |0-3|) = \max(1,3) = 3$$

***

Some important topological aspects of an image are as follows:

1. The set of pixels that has connectivity in a binary image is said to be characterized by the connected set.
2. A digital path or curve from pixel p to another pixel q is a set of points $p_1, p_2, ..., p_n$. If the coordinates of those points are $(x_0, y_0), (x_1, y_1), ..., (x_n, y_n)$, then $p = (x_0, y_0)$ and and $q = (x_n, y_n)$. The number of pixels is called the length. If $x_0 = x_n$ and $y_0 = y_n$, then the path is called a closed path.
3. R is called a region if it is a connected component.
4. If a path between any two pixels p and q lies within the connected set S, it is called a connected component of S. If the set has only one connected component, then the set S is called a connected set. A connected set is called a region.
5. Two regions $R1$ and $R2$ are called adjacent if the union of these sets also forms a connected component. If the regions are not adjacent, it is called disjoint set. In Fig.10, two regions $R1$ and $R2$ are shown. These regions are 8-connected because the

pixels (underlined pixel '1') have 8-connectivity. If the regions are not adjacent, they are called disjoint.

0	1	0	0	1	1
1	1	0	1	1	1
1	1	1	0	1	1
Region 1			Region 2		

Fig. 10: Neighboring regions

6. The border of the image is called contour or boundary. A boundary is a set of pixels covering a region that have one or more neighbours outside the region. Typically, in a binary image, there is a foreground object and a background object. The border of the foreground object may have at least one neighbor in the background. If the border pixels are within the region itself, it is called inner boundary. This need not be closed.

7. Edges are present wherever there is an abrupt intensity change among the pixels. Edges are similar to boundaries, but may or may not be connected. If edges are disjoint, they have to be linked together by edge linking algorithms. However boundaries are global and have a closed path.

Try an exercise.

E14) How Image topology relates to image processing ?

In the following section, we shall discuss various applications where we user digital images.

1.10 APPLICATIONS THAT USE DIGITAL IMAGES

Image processing has got wide variety of application areas, and it is widely utilized by engineers, scientists, media , fine arts professionals and many more. Nowadays it is an exciting interdisciplinary field that borrows ideas freely from many fields. Fig. 11 illustrates the relationships between image processing and other related fields.

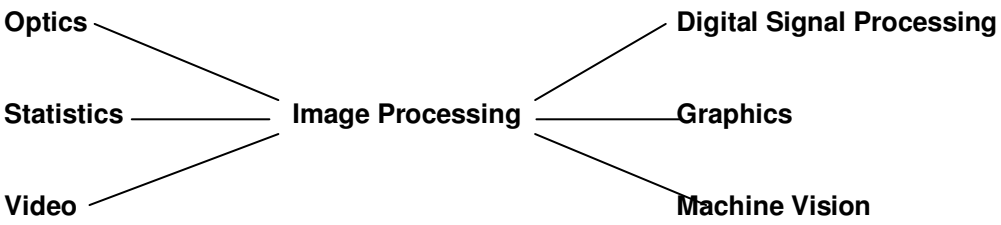


Fig. 11: Image processing and other closely related fields.

Image processing and Computer Graphics: Computer graphics and image processing are very closely related areas. Image processing deals with raster data or bitmaps, whereas computer graphics primarily deals with vector data. Raster data or bitmaps are stored in a 2D matrix form and often used to depict real images. However, vector images are composed of vectors, which represent the mathematical relationships

between the objects. Vectors are lines or primitive curves that are used to describe an image. Vector graphics are often used to represent abstract, basic line drawings.

The algorithms in computer graphics often take numerical data as input and produce an image as output. However in image processing, the input is often an image. The goal of image processing is to enhance the quality of the image to assist in its interpretation. Hence, the result of image processing is often an image or the description of an image. Thus, image processing is a logical extension of computer graphics and serves as a complementary field.

Image Processing and Machine Vision : The main goal of machine vision is to interpret the image and to extract its physical, geometric, or topological properties. Thus, the output of image processing operations can be subjected to more techniques, to produce additional information for interpretation. *Artificial vision* is a vast field, with two main subfields- machine vision and computer vision. The domain of *machine vision* includes many aspects such as lighting and camera, as part of the implementation of industrial projects, since most of the applications associated with machine vision are automated visual inspection systems. The applications involving machine vision aim to inspect a large number of products and achieve improved quality controls. *Computer vision* is more ambitious. It tries to mimic the human visual system and is often associated with scene *understanding*. Most image processing algorithms produce results that can serve as the first input for machine vision algorithms.

Image Processing and Video Processing: Image processing is about still images. In fact, analog video cameras can be used to capture still images. A video can be considered as a collection of images indexed by time. Most image processing algorithms work with video readily. Thus, video processing is an extension of image processing. In addition, images are strongly related to multimedia, as the field of multimedia broadly includes the study of audio, video images, graphics, and animation.

Image Process and Optics: Optical image processing deals with lenses, light, lighting conditions, and associated optical circuits. The study of lenses and lighting conditions has an important role in the study of image processing.

Image Processing and Statistics: Image analysis is an area that concerns the extraction and analysis of object information from the image. Imaging applications involve both simple statistics such as counting and Mensuration and complex statistics such as advanced statistical inference. So statistics play an important role in imaging applications.

You may try the following exercise.

E15) Write any three applications of digital image processing.

Now, we summarise what we have studied in the unit.

1.11 SUMMARY

We have covered the following points:

1. A digital image is defined as a 2D discrete signal that varies over the spatial coordinates x and y, and can be written mathematically as $f(x,y)$. It is also an $n \times n$ array of elements, and each element represents the sampled intensity.
2. Image acquisition can be done using various ways and the concepts required to understand the acquisition of any image.
3. The two steps of image digitization are sampling and quantization.
4. Two models for representing an image.
5. Classification of the image on the basis of various attributes of image like colour, dimension and data types, etc.
6. Various terms which will be used again and again in the course.
7. The topological aspects of image are also covered in this unit.
8. Finally, the application areas of image processing are also discussed.

1.12 SOLUTIONS/ANSWERS

E1) Image file format is an algorithm or a method used to store and display an image. Some of the frequently used file formats are JPEG, PNG, BMP, GIF, TIFF, etc.

E2) Some of the points of difference in analogue and digital images are:

i. Digital image's pixel value must be discrete whereas Analog image's pixel value must be continuous.

ii. The amplitude of digital image is finite whereas the amplitude of analog image is infinite.

iii. It is quite possible to store all the pixels of digital image whereas It is quite impossible to store all the pixels of analog image.

E3) **Analog Image Processing:** It is applied on analog signals/images and it processes only 2 D images. Examples are television, images, photographs, painting and medical images, etc.

Digital Image Processing: It is applied to digital images (a matrix of small pixels and elements). For example, colour processing, image recognition, video processing, etc.

Optical Image Processing: An optical image of 3D object is is in fact its 2D projection, which is the continuous distribution of light on a 2D surface. This 2D projection holds various types of information like 3D objects that are in focus. It also includes the study of the radiation source, and other optical processes. This

optical image is available in the optical form until it is converted into analog form, leading to the area of analog image processing.

E4) **Direct or Active Imaging Systems**: These are the systems where no temporary storage on film or chemical processes is required. In such systems the acquired images are in a recognizable form. We can say that in such systems the final image is a digital image. For example, human eye, digital camera, scanning confocal microscopes, etc.

Indirect Imaging Systems: Incomparison to Direct imaging systems, the Indirect imaging systems involves data processing or reconstruction before producing an image for observation. In indirect imaging systems, the image is stored in a film, which is rendered observable by a chemical process. Therefore, there is a delay in the production of images from films.

E5) The bandwidth = 500 dots per inch in both directions
Therefore sample size = 1000 dots per inch [using sample theorem]

Total number of samples = 5 x 1000 x 8 x 1000 = 40000000

E6) An image may be continuous with respect to the x - and y - coordinates, and also in amplitude. Converting such an image to digital form requires that the coordinates, as well as the amplitude, be digitized.

Sampling is digitizing the coordinate values and quantization is digitizing the amplitude values.

E7) The dimensions of pixel is represented by Δx and Δy, and they are kept the same in both the horizontal and vertical directions, so that the pixels will look like square pixels. This is called pixel aspect ratio. Sometimes the size of the pixel is very large and numbers of pixels and very less, which makes the image distracting and meaning less. This is known as pixelization error.

E8) As per Shannon-Nyquist theorem, the sampling frequency should be greater than or equal to $2\times f_{max}$, where f_{max} is the highest frequency present in the image. Otherwise, the original signal cannot be reconstructed. In other words, the number of samples required is dictated by Shannon-Nyquist theorem, this sampling theorem can be stated in terms of distance d as $d \le \frac{1}{2f_{max}}$.

Alternatively, it should be less than or equal to the smallest detail that is present in the image.

E9) **Pixel model:** In this model, the RGB optical sensors maps to a vector of discrete intensities which are integers. Let $p: R\times G\times B \to Z\times Z\times Z$ defined by $p(r,g,b) = (zR, zG, zB)$, where where r, g, b are real valued colour intensities and zR, zG, zB are integer-valued colour intensities in the range from 0 to 255. In this case, a pixel is represented by a 1×3 column vector used to "paint" a tiny square in an image display.

Vector model: In this model, the RGB optical sensor values are mapped to vectors and the displayed image conforms to a form of triangulated image space instead of a collection of pixels. In this model, a vector image is a collection of vertices and edges. The end result of vector images is the formation of a basis for applications such as Adobe Photoshop and various video games.

E10) The classification of an image can be done on the basis of

i) Attributes
ii) Colour
iii) Dimensions
iv) Data type.

E11) **Monochrome images**are the images where the colour component is absent, they are further classified as Grey scale and binary images.

True colour (or full colour) images are the images where the pixel has a colour that is obtained by mixing the primary colours red, green, and blue. Each colour component is represented like a grey scale image using eight bits. Mostly, true colour images use 24 bits to represent all the colour. Hence true colour image can be considered as three-band images. The number of colours that is possible is 256^3 (i.e. 256*256*256= 1,67,77,216 colours).

E12) **Brightness: It** refers to the average pixel intensity of the image, Similar to contrast, the image should be reasonably bright to show all the information to the viewer. However, excessive brightness may affect the quality of the image. An image may have a higher brightness but if the intensity is not optimal, the image again will not exhibit good quality. *Brightness resolution* is defined as the number of identified distinct colours and it is also called as *colour resolution*.

We know, Contrast of an image relates to recording of the differences in the magnitude of the intensity at the surface of an object. So, Contrast is the difference between maximum and minimum pixel intensities in an image. Consider two images A& B having pixel intensities between 30 to 200 and 70 to 130, respectively. Then A has more contrast than B. Again contrast is also relative. Contrast can be simply explained as the difference between maximum and minimum pixel intensity in an image. For example, consider an image, whose maximum and minimum value of intensity is 100 i.e. same then contrast will be zero because, Contrast = maximum pixel intensity – minimum pixel intensity $=100-100=0$.

E13) **Pixel resolution**, the term resolution refers to the total number of count of pixels in an digital image. For example. If an image has M rows and N columns, then its resolution can be defined as $M \times N$. If we define resolution as the total number of pixels, then pixel resolution can be defined with set of two numbers.

E14) You may like to write it yourself.

E15) Any three applications may be listed.

UNIT 2

IMAGE TRANSFORMS I

Structure **Page No.**

2.1 INTRODUCTION

As we discussed in the previous unit, an image is a two dimensional signal. In this unit, we shall look into the definitions of signals and their properties. These properties help in defining the relationship between images and 2D signals. Once an image is considered as a 2D signal, transformations on the signals can be applied to transform images as well. The transformations that we shall focus upon in this unit are the orthogonal transforms and unitary transforms of 2D signals. Finally, we shall discuss the properties of these transforms that are useful in image processing.

You may be familiar with some or all of these concepts from Unit-1. However, a quick look through this unit will help you in refreshing and revising your knowledge in this domain.

In Unit-1, you have gone through the definitions and properties of digital images. Therefore, you already know that when we capture an image from a camera, the image formed is a two-dimensional function that contains certain information regarding the luminance of scene objects in an RGB image or the temperature of the objects such as in a thermal image.

In Sec. 2.2, we shall discuss the definition of 1-D and 2-D signals. We then discuss the relationship between images and 2-D signals in Sec. 2.3. We discuss the orthogonal transforms and unitary transforms in Sec. 2.4 and their properties in Sec. 2.5. Finally, we summarise the discussion in Sec. 2.6 and in Sec. 2.7, we give the solutions/answers/hints to the exercises.

First, we list the objectives of this unit. After going through this unit, please read this list again and make sure that you have achieved the objectives.

Objectives

After studying this unit, you should be able to:

- define the 1-dimensional (1-D) and 2-dimensional (2-D) signals.
- relate how an image is also a 2-dimensional signal
- define and apply the unitary and orthogonal transforms of signals.
- list the properties of unitary transforms, especially useful for image processing.

Let us begin with signals in the following section.

2.2 SIGNALS

You have read about the signals in Unit-1. We begin this section by defining them. A signal is a function of one or more independent variables, where the variables represent some physical meaning. These signals carry information through variables. The dimension of a signal is defined by the number of independent variables that the function is defined on. For example, an $\mathrm{E} \subset \mathrm{G}$ signal is one-dimentional, whereas the intensity of a still image is 2-dimensional. You may recall the definition of a function as given below.

A function, f, is a mapping from the domain, D, that is, a set of values to another set of values called the range, $\mathbb{R}$. Therefore, $\mathrm{f} : \mathrm{D} \rightarrow \mathbb{R}$ such that for any element $\mathrm{x} \in \mathrm{D} \Rightarrow \mathrm{f}(\mathrm{x}) \in \mathbb{R}$.

Therefore, a way to differentiate the signals is by the dimension of their domain, which is also the number of independent variables that the function/signal is defined on.

Let us begin our discussion by defining a signal in 1-D.

Definition(1-D Signal):A 1-dimesional (1-D) signal is a function of a single variable. A one-dimensional continuous signal is represented as a function, $f(t)$, of the independent variable *t*, representing the evolution of a physical phenomenon across time '*t*'. For example, a 1-D audio signal, such as the electrical signal at the output of a microphone is a function of time.

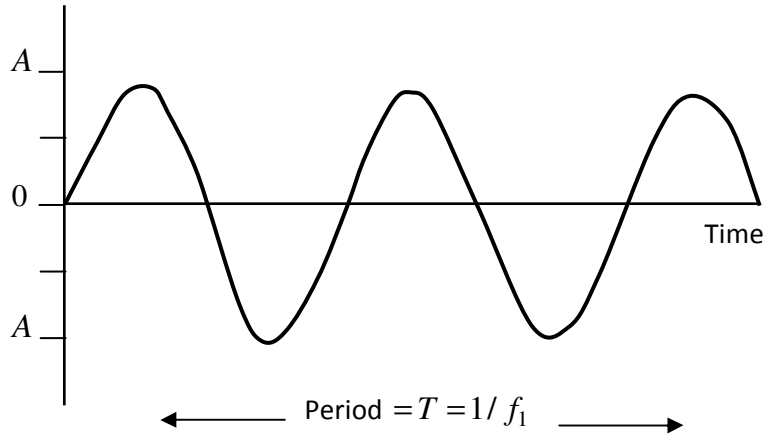


Fig. 1: 1-D Signal.

Now, we shall give the definition of 2-D signals.

Definition (2-D Signals): A 2-dimensional (2 – D) signal $f(x,y)$, is a function of two independent variables, x and y, such that (x,y) indicate a point in the 2D space.

For example, if the function f shows the intensity in a still image, then $f(x,y)$ represents the value of the intensity for the pixel at the location (x,y). Functions for 2-D signals are defined below:

i) The 2-D impulse function is defined by

$$f(x,y) = \delta(x,y) = \begin{cases} 1, & x = y = 0 \\ 0, & \text{otherwise} \end{cases}$$

ii) The line impulses are in horizontal line, vertical line or diagonal line. Accordingly,

$$f(x,y) = \begin{cases} \delta(x), & [\text{horizontal}] \\ \delta(y), & [\text{vertical}] \\ \delta(x \pm y), & [\text{diagonal}] \end{cases}$$

iii) Exponential signals is defined as $f(x,y) = n^x y^x$, where m and n are complex numbers. This function can also be written as $f(x,y) = \cos(z_1 x + z_2 y) + i\sin(z_2 x + z_2 y)$ where, $m = e^{iz_1}$ and $n = e^{iz_2}$.

Example 1: Draw the signal of $f(x,y) = \delta(x-2y)$. Identify the signal impulse type.

Solution: Using the δ-function, we get $\delta(x-2y) = \begin{cases} 1, & x-2y=0 \\ 0, & \text{otherwise} \end{cases}$.

We plot these points

x	0	1
y	0	0.5

on the graph as shown in Fig. 2.

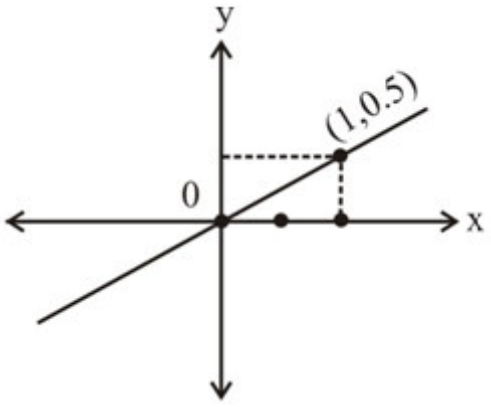


Fig. 2: Representation of $f(x,y)=\delta(x-2y)$.

The type of signal impulse is diagonal line.

***

Try an exercise.

E1) Draw the signal of $f(x,y)=\delta(x+y-2)$ and also, identify the type of the signal.

In the following section, we shall see how an image can be related with a 2-D signal.

2.3 IMAGE AS A 2-D SIGNAL

A grayscale image is a discrete 2-D signal $f(x,y)$, having two independent variables, x and y, such that $f(x,y)$ is the value of the signal at a pixel whose location in the image is given by integers x and y. The 2-D signals can be of various types such as seperable, periodic, etc.

Therefore, the 2D signal represents a digital image, where the (x,y) is the location of the pixel and $f(x,y)$ is the value at this pixel. Fig.3 showsthe discrete 2D signal.

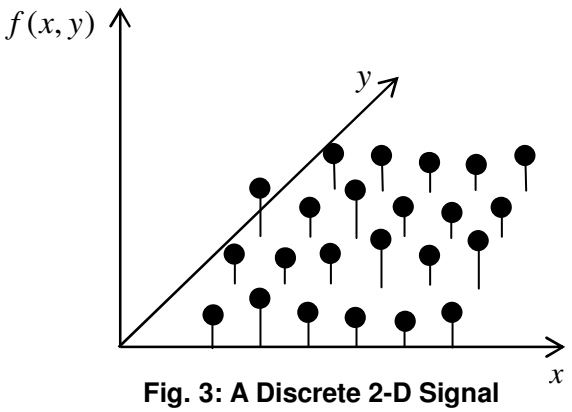


Fig. 3: A Discrete 2-D Signal

Fig.4 is another representation of a gray-scale image, where each square represents a pixel and the colour of the square gives the signal

value at that location. In case of an image, the signal value represents the intensity at that location.

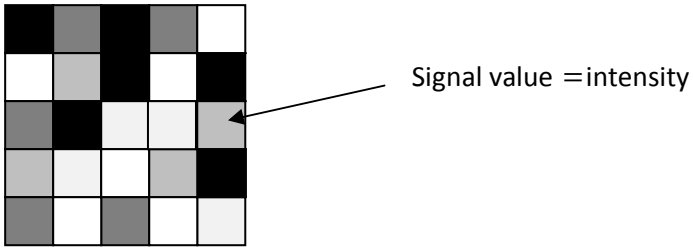
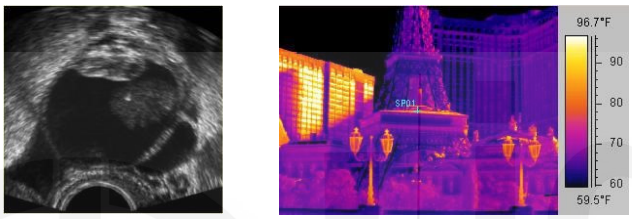


Fig.4: A Gray-Scale Image, An Example of a 2-D Signal.

In general, brightness or intensity is usually the value of the function in a image. However, in an image, the signal value can represent physical values such as temperature, pressure, depth, etc. also. We are showing some of these examples in Fig. 5.



(a) ultrasound **(b) temperature**

(c) RGB image **(d) CT scan image**

Fig.5: Examples of different types of images, depending on the physical phenomenon represented by the 2D signal.

Try an exercise.

E2) Define a 2-D image in terms of 2-D signal.

Sometimes the image in 2-D signals are difficult to process and analyse due to the complexity of the function involved in them. To make such task easier we need to transform the image. Now we discuss transformation of 2-D signals needed to process an image.

2.4 TRANSFORMATIONS OF 2 D SIGNALS

Image transforms are necessary for image analysis and image processing. Transformations are mathematical functions that allow us to convert from one domain to another. Therefore, transformation of a signal $f(x)$ will convert the signal into a new signal, say, $g(y)$, in a

different domain. However, transformation does not change the information content present in the signal/image.

Image transforms are important for computing correlation and convolutions. Therefore, image transforms find various applications. Depending on the transform used, the transformed image represents the image data in a more compact form, which helps in storage and transmission of the image easily. Some transforms also separate the noise from the image, making the information in the image clearer. Transformations such as Fourier transformation, cosine transformation which we shall study in the next unit, provide us the information about the frequency content in an image.

For our understanding, we first consider a one-dimensional discrete signal with N samples and represent the signal as $\mathrm{f(x)}, 0 \le \mathrm{x} \le \mathrm{N}-1$. Then, a transform of the signal $\mathrm{(x)}$ will convert the signal into a new signal $\mathrm{g(y)}$, such that if $\mathrm{(x)}$ has N samples, then $\mathrm{g(y)}$ also has N samples.

The generic form of transforms is:

$$\mathrm{g(u)} = \sum_{\mathrm{x}=0}^{\mathrm{N}-1} \mathrm{T(u,x)\,f(x)}, 0 \le \mathrm{u} \le \mathrm{N}-1 \qquad (1)$$

where, $T(u,x)$is called the **forward transformation kernel**.

The Eqn. (1) shows that to carry out the transformation, all values of $\mathrm{f(x)}$ are required to compute each of the N values of $g(u)$. Since we can write the N values of $\mathrm{f(x)}$ and $g(u)$in a column vector, we can write the transform in form of matrix multiplication.

Let f and g denote these column vectors. Then, the transformation equation will be of the form

$$\mathrm{g} = \mathrm{T} \cdot \mathrm{f} \qquad (2)$$

where, the matrix T is of dimension $\mathrm{N} \times \mathrm{N}$ and contains the values $\mathrm{T(y,x)}$ for different $\mathrm{y,x}$.

The Eqn. (1) can also be extended for transformation of 2-D signals.

A question that comes to our mind is why do we need to transform signals, especially images. Why we cannot do image processing in the spatial domain itself, where it is easy to see the picture? To answer these questions, we need to understand that many a time it is easier to carry out image processing tasks in another domain (or frequency domain) rather than in the spatial domain.

Now, we list the key steps for 2-D image transformations.

The key steps for image transformation from the spatial to the frequency domain are:

Step 1: Transform the image from the spatial domain to the frequency domain.

Step 2: Carry the required image processing task in the transformed domain.

Step 3: Apply inverse transform to return to the spatial domain.

These key steps are also listed in Fig. 6.

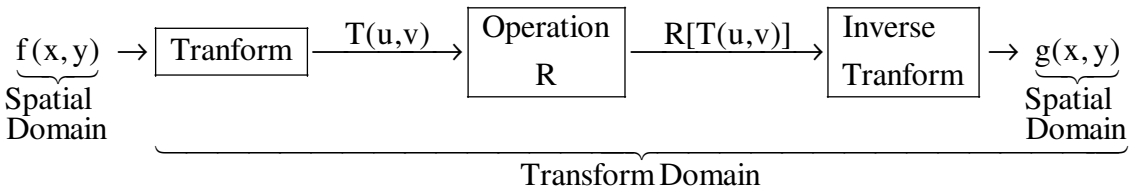


Fig. 6: Key Steps of 2-D Image Transformation.

We shall show in subsequent sections how to apply these transformations.

Try an exercise.

E3) List the steps from 2-D signal to 2-D image transformation.

In the following section, we shall discuss orthogonal transformation of 2-D signals.

2.5 ORTHOGONAL TRANSFORMATIONS OF 2-D SIGNALS

A square matrix $A = [A_1 A_2 ... A_n]$ where, A_i is the i^{th} column vector of A is real and has the property that $A^{-1} = A^T$, then it is said to be an orthogonal matrix. If, A is not real, it is said to be a unitary matrix if $A^{-1} = A^T$, that is the inverse of A is equal to its conjugate transpose.

The columns of a unitary and an orthogonal matrix are perpendicular to each other and therefore, are said to be **orthogonal**. Also, the columns of these matrices are of unit length, that is, **normalized**. Therefore, the columns of these matrices are said to be **orthonormal**. Their inner product therefore, satisfies the following:

$$(A_i, A_j) = \delta(i,j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

The n columns of these matrices can be treated as the **basis vectors** of an n-dimensional vector space.

Let us consider an $N \times N$ image $f(x,y)$. The forward and inverse transforms of $f(x,y)$ are given by

$$g(u,v)=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}T(u,v,x,y)f(x,y),\ 0\le u,\ v\in N-1. \quad (3)$$

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}I(x,y,u,v)g(u,v),\ 0\le x,y\in N-1 \quad (4)$$

where, $g(u,v)$ is the transformed image, $T(u,v,x,y)$ is called the **forward transformation kernel** and $I(x,y,u,v)$ is called the**inverse transformation kernel.** It is said to be an **orthogonal transform**, when T is an **orthogonal matrix**.

The forward transformation kernel satisfies the following properties:

i) **Orthonormality Property:**

$$\sum_{x,=0}^{N-1}\sum_{y=0}^{N-1}T(u,v,x,y)I(u',v',x,y)=\delta(u-u',v-v') \quad (5)$$

ii) Completeness Property:

$$\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}T(u,v,x,y)I\ (u,v,x',y')=\delta(x-x',y-y') \quad (6)$$

Eqns. (3) and (4) are not easy to solve. If the transform is restricted to be separable, that is $T(u,v,x,y)=A(u,x)\ B(v,y)$, then we can rewrite Eqn. (3) and Eqn. (4) as

$$g(u,v)=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}A(u,x)f(x,y)B(v,y)=AfA^{T} \quad (7)$$

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}A^{*}(u,v)g(u,v)\ B^{*}(v,y)=A^{*T}g\ A^{*} \quad (8)$$

For the rectangular image $M\times N$, we get

$$g(u,v)=A_M\ f\ A_N \quad (9)$$

$$f(x,y)=A_M^{*}\ g\ A_N^{*\,T}. \quad (10)$$

Example 2: Consider the following orthogonal matrix A and image matrix f

$$A=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix},\quad f=\begin{bmatrix}1 & 3\\ 5 & 7\end{bmatrix}$$

Apply the orthogonal transform and its inverse.

Solution: The transformed image, obtained according to the Eqn. (3) is

$$g=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 3\\ 5 & 7\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}=\frac{1}{2}\begin{bmatrix}6 & 10\\ -4 & -4\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}=\begin{bmatrix}8 & -2\\ -4 & 0\end{bmatrix}$$

We compute the outer product of the columns of A^{*T}.

$$A^*(0,0) = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$$

Similarly, we compute $A^*(0,1)$

$$A^*(0,1) = \frac{1}{2}\begin{bmatrix}1 & -1\\1 & -1\end{bmatrix} = A^{*T}(1,0) \text{ and } A^*(1,1) = \frac{1}{2}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix}$$

Now, we obtain the inverse transformation. We get

$$A^{*T}gA^* = \frac{1}{2}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}8 & -2\\-4 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}4 & -2\\12 & -2\end{bmatrix}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix} = \begin{bmatrix}1 & 3\\5 & 7\end{bmatrix}.$$

We can verify that the inverse transformation gives us the original image.

***

Try following exercises.

E4) Is the identity matrix, I, an orthonormal matrix? Can it be used to define an orthonormal transform? What will be the inverse transformation kernel in this case?

E5) For the 2×2 transform A and the image f, which are given as:

$$A = \frac{1}{2}\begin{bmatrix}\sqrt{3} & 1\\-1 & \sqrt{3}\end{bmatrix},\ f = \begin{bmatrix}2 & 3\\1 & 1\end{bmatrix}$$

Calculate the transformed image g and the basis images.

In the following section, we shall discuss unitary transforms.

2.6 UNITARY TRANSFORMS

We first discuss the unitary transforms for one dimensional signals for ease of understanding and will then move on to discussing them for 2D signals.

Unitary Transforms for 1-D signals

A one dimensional sequence $\{f(x), 0 \le x \le N-1\}$ can always be represented as a vector of dimension N, as $\mathbf{f} = [f(0)f(1)..f(N-1)]$.

Then, a transformation T of $f(x)$ can be written as:

$$\mathbf{g} = \mathbf{T} \cdot \mathbf{f} \Rightarrow g(u) = \sum_{x=0}^{N-1} T(u,x) f(x),\ 0 \le u \le N-1 \tag{11}$$

where, T(u,x) is called the **forward transformation kernel** and $g(u)$ is the transform of $f(x)$, that is $g(u)$ is the result of applying the transformation $T(u,x)$ on $f(x)$.

There exists an inverse relation that transforms $g(u)$ back to $f(x)$, which is given by $I(x,u)$ such that

$$f(x) = \sum_{u=0}^{N-1} I(x,u) g(u),\ 0 \le x \le N-1 \tag{12}$$

where, $I(x,u)$ is called the **inverse transformation kernel**.

Therefore, in the matrix form, we can write

$$\mathbf{f} = \mathbf{I} \cdot \mathbf{g} = \mathbf{T}^{-1} \cdot \mathbf{g}$$

Recall, that a matrix is said to be a unitary matrix, if the inverse of the matrix is also its conjugate transpose. That is, if

$$\mathbf{I} = \mathbf{T}^{-1} = \mathbf{T}^{*T}$$

then the matrix $\mathbf{T}$ is called **a unitary matrix** and therefore, the transformation is also known as the **unitary transformation**.

In general, the rows (or columns) of an $N \times N$ unitary matrix are orthonormal and hence, they form a complete set of basis vectors in the N- dimensional vector space. Since,

$$\mathbf{f} = \mathbf{I} \cdot \mathbf{g} = \mathbf{T}^{*T} \cdot \mathbf{g} \Rightarrow f(x) = \sum_{u=0}^{N-1} T^*(u,x) g(u) \tag{13}$$

therefore, the columns of $\mathbf{T}^{*T}$, that is, the vectors $\mathbf{T}_u^* = \left[T^*(u,0)\ T^*(u,1) \ldots T^*(u,N-1)\right]^T$ are called the **basis vectors** of T.

Now let us discuss the unitary transforms of 2-D signals.

Similar to the 1D unitary transform, which allows us to represent a 1D signal by a set of orthonormal basis vectors, there exists unitary transforms in the higher dimensions. The 2D unitary transform allows us to represent a 2D signal/ image as set of basis arrays or basis images.

Given an $N \times N$ image $f(x,y)$ the forward and inverse transforms are given by

$$g(u,v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} T(u,v,x,y) f(x,y) \tag{14}$$

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} I(x,y,u,v)g(u,v) \quad (15)$$

where, $T(u,v,x,y)$ is called the **forward transformation kernel** and $I(x,y,u,v)$ is called the **inverse transformation kernel.** It is said to be a unitary transform when T is a unitary matrix.

In general, the transformations (both orthonormal and unitary) can have the following properties:

i) **Separability:** The transformation kernel is said to be **separable** if $T(u,v,x,y)=T_1(u,x)T_2(v,y)$

ii) **Symmetry**: The transformation kernel is said to be **symmetric** if T_1 is functionally equal to T_2 such that $T(u,v,x,y)=T_1(u,x)T_1(v,y)$

The unitary transformation is an important transformation in image processing. If the forward transformation kernel $T(u,v,x,y)$ is both separable and symmetric, then the transform

$g(u,v)=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} T(u,v,x,y)f(x,y)=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} T_1(u,x)T_1(v,y)f(x,y)$ implies:

$$\mathbf{g}=\mathbf{T_1}\cdot\mathbf{f}\cdot\mathbf{T_1^T} \quad (16)$$

in the matrix form, where $\mathbf{f}$ denotes the original image of size $N\times N$, and $\mathbf{T_1}$ is an $N\times N$ transformation matrix with elements $t_{ij}=T_1(i,j)$. If, in addition, $\mathbf{T_1}$ is a unitary matrix then the original image is recovered using the following equation

$$\mathbf{f}=\mathbf{T_1}^{*T}\cdot g\cdot\mathbf{T_1}^{*} \quad (17)$$

Such a transformation is called a **separable unitary transformation**.

To understand this better, let us study the following example.

Example 3: Check whether the matrix $A=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}$ is unitary or not.

Solution: The members of A are real, therefore the condition for A to be unitary is $A^{-1}=A^{T}$. We compute A^{-1} and A^{T}.

$$A^{-1}=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix} \quad (18)$$

and $$A^{T}=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix} \quad (19)$$

from Eqn (18) and Eqn (19), it is clear that

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

Hence, A is unitary matrix.

***

Try the following exercises.

E6) Why do we need image transform?

E7) Check whether the matrix $A = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$ is unitary or not.

Now, we shall discuss some important fundamental properties of unitary transforms.

2.7 FUNDAMENTAL PROPERTIES OF UNITARY TRANSFORMS

Here we discuss the properties of unitary transforms.

i) **Energy conservation and rotation:** Let $\mathbf{T}$ be a unitary transformation, then

$$\mathbf{g} = \mathbf{T} \cdot \mathbf{f}$$

implies that

$$\|\mathbf{g}\|^2 = \|\mathbf{f}\|^2 \qquad (20)$$

This means that the signal energy is preserved after the application of a unitary transformation. This property is called **energy preservation property**. It also implies that every unitary transformation is a rotation of the vector $\mathbf{f}$ in the N - dimensional vector space.

ii) Energy compaction

A large fraction of the energy in the image is packed into relatively few transform coefficients in most unitary transforms. By the above mentioned property, since the energy is preserved, this implies that most of the transform coefficients have insignificant values and only a few of the transform coefficients that are close to the origin have significant values. This property is very useful for image compression purposes.

Try an exercise.

E8) Show that if T is a unitary transformation, such that $\mathbf{g} = \mathbf{T} \cdot \mathbf{f}$ then, $\|\mathbf{g}\|^2 = \|\mathbf{f}\|^2$.

Now, let us summarise what we have discussed so far.

2.8 SUMMARY

In this unit, we discussed the following:

1. 1-D and 2-D signals and that an image is also a 2-D signal.

2. We also discussed the orthonormal and the unitary transformation and the properties of the unitary transformation that makes it very useful for image processing, especially image compression.

2.9 SOLUTION AND ANSWER

E1) Here, $x + y - 2 = 0 \Rightarrow y = 2 - x$.

The values of x and y are

x	0	1
y	2	1

We plot these points as shown in Fig. 7.

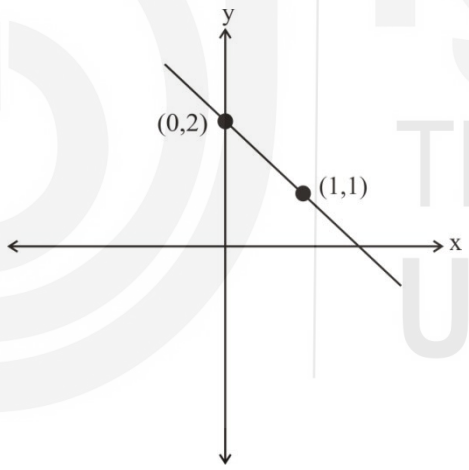


Fig. 7: Representation

E2) A grayscale image is a discrete 2-D signal $f(x, y)$, having two independent variables, x and y, such that $f(x, y)$ is the value of the signal at a pixel whose location in the image is given by x and y where, x and y are integers and the value of the signal is not defined when x and y are non-integers.

E3) First, transform the image from the spatial domain to the frequency domain. Secondly, do the image processing, and finally, apply inverse transform to return to the spatial domain.

E4) The identity matrix $A = I = [E_0, ..., E_i, ..., E_n]$, where the i^{th} column is $E_i = [0, 0, ..., 1, 0, ..., 0]$ with the i^{th} element is 1 and all other values are

0. Then, $A = I$ is an orthonormal matrix since each column is perpendicular to the others and the length of each column is unity. Therefore, it defines an orthogonal transform $Y = IX = X$. Then, the inverse transform is given by:

$$X = AY = \sum_{i=1}^{n} y_i A_i = \sum_{i=1}^{n} x_i E_i$$

Since this is an identical transformation, in this special case, the signal X and its transform, Y are identical.

E5) $$g = \frac{1}{4}\begin{bmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix} 2\sqrt{3}+1 & 3\sqrt{3}+2 \\ -2+\sqrt{3} & -3+2\sqrt{3} \end{bmatrix}\begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix} 8+4\sqrt{3} & 8 \\ 0 & 8-4\sqrt{3} \end{bmatrix}$$

$$= \begin{bmatrix} 2+\sqrt{3} & 2 \\ 0 & 2-\sqrt{3} \end{bmatrix}$$

$$f = \frac{1}{4}\begin{bmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{bmatrix}\begin{bmatrix} 2+\sqrt{3} & 2 \\ 0 & 2-\sqrt{3} \end{bmatrix}\begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix} 8 & 12 \\ 4 & 8 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$$

E6) Image transforms make image analysis and image processing easier without changing the information content present in the image.

E7) Check whether $A^{-1} = A^{T}$

E8) If $\mathbf{g} = \mathbf{T}\cdot\mathbf{f}$, then,

$$\|g\|^2 = \|Tf\|^2 = (Tf)^{*T}(Tf) = f^{*T}T^{*T}Tf = f^{*T}If = \|f\|^2$$

UNIT 3

IMAGE TRANSFORMS II

Structure **Page No.**

3.1 INTRODUCTION

As we saw in Unit 2, an image can be transformed, to show or hide information in the image. Image transformations can be done in both the spatial and the frequency domain. In the spatial domain, image transformation is carried out by changing the value of the pixels based on certain constraints. These transformations can change the brightness and clarity of the images.

In this unit, we shall focus on the transformations in the frequency domain. We recall at this point that an image is also a 2D signal and that the transformations that can be applied on a signal can also be applied on any image. The transformations in the frequency domain provide us information on the frequency content of the image. These transformations can help represent the information in the image in a more compact form, thereby making it computationally easier to store and transmit the images. The transformations may also help in separating the noise and the salient information present in the image.

In Sec. 3.2, we shall focus on very important and useful image transformations, namely the Discrete Fourier transformation (DFT). We shall continue our discussion in Sec. 3.3 with the Discrete Cosine Transformation (DCT). In Sec. 3.4, Haar transform. As we go through this unit, we shall see the unique properties of each of these transforms.

Now we shall list the objectives of this unit. After going through the unit, please read this list again and make sure that you have achieved the objectives.

Objectives

After studying this unit you should be able to:

- find the Discrete Fourier Transform (DFT)
- compute the Discrete Cosine Transform (DCT)
- find the Haar Transform
- apply the above mentioned transforms

We shall begin the unit with discrete fourier transform.

3.2 DISCRETE FOURIER TRANSFORM

The Discrete Fourier Transform (DFT) transfers an image from the spatial domain to the frequency domain. It is one of the most important transforms in image processing, which enables us to decompose an image into its sine and cosine components. The output image after applying the Fourier transformation is represented as a linear combination of a collection of sine and cosine waves of different frequencies.

Consider a 1D function, $\{f(x), 0 \le x \le N-1\}$. The general form of a transformation is

$$g(u) = \sum_{x=0}^{N-1} T(u,x) f(x);\ 0 \le u \le N-1 \tag{1}$$

where $T(u,x)$ is called the **forward kernel** of transformation and $g(u)$ is the transformed image.

If the transformation is the Discrete Fourier Transform.

Then,

$$g(u) = \sum_{x=0}^{N-1} \frac{1}{N} e^{-i2\pi \frac{x}{N}} f(x); u = 0,1,2,...N-1 \tag{2}$$

The inverse 1-D DFT will then be,

$$f(x) = \sum_{x=0}^{N-1} e^{i2\pi \frac{ux}{N}} g(u) \tag{3}$$

As can be seen the signal is written as a linear combination of an orthogonal set of basis functions. Similarly, an image can be transformed into a set of “basis images”, which can be used for representing the image.

We can extend the transform to 2-D image.

Consider an image $f(x,y)$ of size $M \times N$. The 2-D DFT of $f(x,y)$ is defined as follows:

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) e^{-i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \quad (4)$$

And the inverse 2-D DFT is given by:

$$f(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v) e^{i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \quad (5)$$

The 2-D DFT is separable, symmetric and unitary. In case of square images, $M = N$. Many a time in image processing we work with square images. Additionally if the imagesize is a power of 2, then DFT implementation becomes very easy. Computational complexity can be reduced by efficient algorithms such as FFT.

The representation of intensity as a function of frequency is called '**spectrum**'. In the Fourier domain image, each point to a particular frequency contained in the spatial domain image. The coordinates of the Fourier spectrum are the spatial frequencies. The spatial position information of an image is encoded as the difference between the coefficients of the real and imaginary parts. This difference is called the "**phase angle**". The phase information is very useful for recovering the original information. The phase information represents the edge information or boundary information of the objects present in an image. For applications such as medical image analysis, the phase information is very crucial in getting information from the image.

Note that while the image values $f(x,y)$ are going to be real, the corresponding frequency domain data is going to be complex. There will be one matrix containing real values $R(u,v)$ and the other matrix $I(u,v)$ will contain the imaginary component of the complex value. The amplitude spectrum or the magnitude for 2D DFT is given by

$$|F(u,v)| = [R^2(u,v) + I^2(u,v)]^{1/2} \quad (6)$$

where, R and I are real and imaginary parts of $F(u,v)$ and and all computations are carried out for the discrete variables $u = 0,1,2,...,M-1$ and $v = 0,1,2,...,N-1$. The spectrum tells us the relative magnitude at each frequency.

The power spectrum of the 2-D DFT is defined as

$$P(u,v) = |F(u,v)|^2 = R^2(u,v) + I^2(u,v) \quad (7)$$

and the phase spectrum of the 2-D DFT is given by

$$\text{ø}(u,v) = \tan^{-1}\frac{I(u,v)}{R(u,v)} \tag{8}$$

Note that the size of the image remains the same as the original image in spatial domain.Therefore, the magnitude, Fourier (phase) spectrum andthe power spectrum are all matrices of size $M \times N$.

Remark: We can find 2-D DFT of an image by simply computing a set of 1-D DFT is for all rows of $f(x,y)$.Thus, the 2-D DFT of an image $f(x,y)$ is

$$F(u,v) = \sum_{x=0}^{M-1} e^{\frac{-2\pi i u x}{M}} \sum_{y=0}^{N-1} f(x,y) e^{\frac{-2\pi i v y}{N}}.$$

$$= \sum_{x=0}^{M-1} F(x,v) e^{\frac{-2\pi i u x}{M}},$$

where $F(x,v) = \sum_{y=0}^{N-1} f(x,y) e^{\frac{-2\pi i v y}{N}}$.

Also, the 2-D DFT can also be found using the Eqn. (4) with the condition of separability as we used in Unit-2.

Let us discuss properties of 2-D DFT.

DFT has several useful properties which makes it an important transformation.

i) **Separability:** It is separable because a 2D transform is separable if $T(u,x,v,y) = T_1(u,x) \cdot T_2(v,y)$.

ii) **Symmetry:** It is symmetric because a 2D transform is symmetric if $T_1(u,x) = T_2(u,x)$.

iii) **Periodicity:**The 2-D DFT and the 2-D IDFT are both periodic, that is, $F(u,v) = F(u+M,v) = F(u,v+N) = F(u+M,v+N)$.

iv) **Conjugate symmetry:** $F(u,v) = F^*(-u+pM,-v+qN)$, where p and q being integers. The property of conjugate symmetry implies that $|F(u,v)| = |F(-u,-v)|$

v) If $f(x,y)$ is real and even then $F(u,v)$ is real and even.

vi) If $f(x,y)$ is real and odd then $F(u,v)$ is imaginary and odd.

vii) Let F be the DFT operator, then

$$F(f(x,y) + g(x,y) = F(f(x,y) + F(g(x,y)$$

However, $F(f(x,y) \cdot g(x,y) \neq F(f(x,y) \cdot F(g(x,y)$

viii) Translation in the spatial domain by (x_0, y_0) implies

$$f(x-x_0, y-y_0) \leftrightarrow F(u,v) e^{-j2\pi\left(\frac{ux_0}{M} + \frac{vy_0}{N}\right)}$$

While translation in the frequency domain by (u_0, v_0) implies

$$F(u-u_0, v-v_0) \leftrightarrow f(x,y)e^{j2\pi\left(\frac{xu_0}{M}+\frac{yv_0}{N}\right)}$$

ix) The average value of the signal is given by

$$\bar{f}(x,y) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)$$

If we see the value of $F(0,0) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) \Rightarrow F(0,0) = \bar{f}(x,y)$.

x) **Rotation:** Rotating $f(x,y)$ by θ rotates $F(u,v)$ by θ.

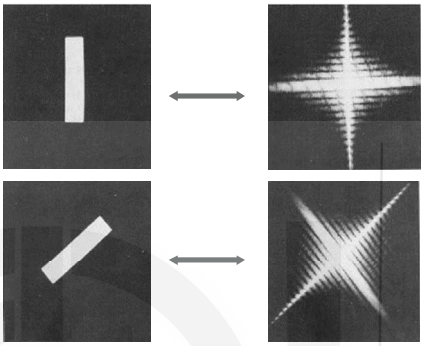


Fig. 1: Rotation in $f(x,y)$

Fig. 1 shows how the rotation in the spatial domain (on left) affects the rotation in the frequency domain (right).

After going through all the properties of DFT, let us see how do we visualize 2-D DFT. We need to translate the origin of the transformed image to the center of the image $(u,v) = (M/2, N/2)$ to be able to display the full period of the 2-D DFT. As we saw above, translating the Fourier image to the center, requires us to use the translation property of $F(u,v)$ with $u_0 = M/2$ and $v_0 = N/2$.

Then, $F\left\{f(x,y)e^{i2\pi\left(\frac{xu_0}{M}+\frac{yv_0}{N}\right)}\right\} = F(u-u_0, v-v_0)$ becomes

$$F\{f(x,y)e^{i\pi(x+y)}\} = F\{f(x,y)(-1)^{(x+y)}\} = F(u-M/2, v-N/2)$$

a) Original Image **b) Original DFT image** **c) Translated DFT image**

Fig. 2: DFT of an Original Image

We can see in Fig. 2 what changes do we see after DFT. Fig 2. (a) is the original image in the spatial domain, (b) is the 2-D DFT image, and

(c) is the translated DFT image to show the full period of the 2D DFT of image in (a).

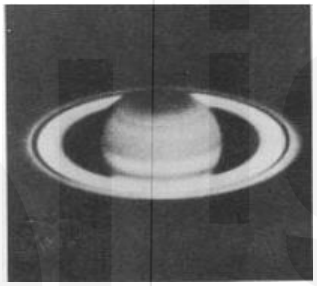
Let us see how do we visualize the range of 2-D DFT.

In general, the range of values the 2-D DFT $F(u,v)$ is very large. Therefore, when we attempt to display the values of $F(u,v)$, smaller values are not distinguishable because of quantization as can be seen in Fig. 3 (b). Therefore, to enhance the small values, weapply a logarithmic transformation given by

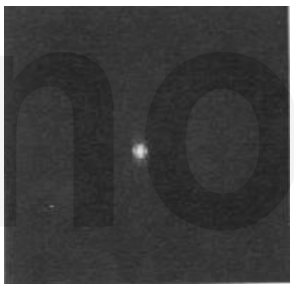
$$D(u,v) = c\log(1+|F(u,v)|)$$

Where, the parameter *c*is chosen so that the range of $D(u,v)$ is $[0,255]$.

$$c = \frac{255}{\log(1+\max\{|F(u,v)|\})}$$



a) The original image



b) The 2D DFT image

c) The 2D DFT image after log transform

Fig. 3: DFT image after log tranform

We can visualise the display of the amplitude of the 2-D DFT after logarithmic transformation in Fig. 3(b) and Fig. 3(c) respectively for the original image as shown in Fig. 3(a).

Example 1: Compute the DFT of the 1D sequence $f(x) = [1,0,-1,0]$.

Solution: Here $N = 4$. Using Eqn. (2) we get

$$g(u) = \frac{1}{4}\sum_{x=0}^{3} f(x)\cdot e^{\frac{-i,2\pi\cdot ux}{4}}; \; u = 0,1,2,3$$

$$= \frac{1}{4}\sum_{x=0}^{3} f(x)\cdot\left(e^{\frac{-2\pi i}{4}}\right)^{ux};\ u = 0,1,2,3$$

$$= \frac{1}{4}\sum_{x=0}^{3} f(x)\ (-i)^{ux};\ \ u = 0,1,2,3$$

$$= \frac{1}{4}[f(0)(-i)^{0} + f(10(-i)^{u} + f(2)(-i)^{2u} + f(3)(i)^{3u}];u = 0,.1,2,3$$

$$= \frac{1}{4}[1+0+(-1)(-i)^{2u} + 0];u = 0,1,2,3$$

$$= \frac{1}{4}[1-(i)^{2u}];u = 0,1,2,3$$

This gives $\mathbf{g} = \frac{1}{4}[0,\ 2,\ 0,\ 2]$, which is the DFT of $f(x)$.

***

Example 2: Construct a DFT matrix of order 2.

Solution: Here $N = 2.$

$$g(u) = \frac{1}{2}\sum_{x=0}^{1} f(x)e^{\frac{-i.2\pi.ux}{2}};\ u = 0,1.$$

$$= \frac{1}{2}\sum_{x=0}^{1} f(x)(-1)^{ux};\ u = 0,1$$

$$= \frac{1}{2}[f(0)(-1)^{0} + f(1)(-1)^{u}];\ \ u = 0,1$$

$$= \frac{1}{2}[f(0) + (-1)^{u} f(1);u = 0,1$$

Now, $g(0) = \frac{1}{2}[f(0) + f(1)]$

$$g(1) = \frac{1}{2}[f(0) - f(1)]$$

DFT matrix $= \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}.$

***

Example 3: Compute the 2-D DFT of the 2×2 image

$$f(x,y) = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}.$$

Solution: Let the DFT of $f(x,y)$ be $F(u,v)$, which is given in Eqn. (4).

$$F(u,v) = \sum_{x=0}^{1}\sum_{y=0}^{1} f(x,y)e^{-2\pi i\left(\frac{ux}{2}+\frac{vy}{2}\right)};\ u,v = 0,\ 1$$

$$= \sum_{x=0}^{1}\sum_{y=0}^{1} f(x,y)(-1)^{ux}(-1)^{vy};\ u,v = 0,\ 1$$

$$= [f(0,0)(-1)^0(-1)^0 + f(0,1)(-1)^0(-1)^v + f(1,0)(-1)^u(-1)^0 + f(1,1)(-1)^u(-1)^v];\ u, v = 0, 1$$

$$= [f(0,0) + (-1)^v f(0,1) + (-1)^u f(1,0) + (-1)^u(-1)^v f(1,1)];\ u, v = 0, 1, 1.$$

$$F(0,0) = f(0,0) + f(0,1) + f(1,0) + f(1,1) = 4$$
$$F(1,0) = f(0,0) + f(0,1) - f(1,0) - f(1,1) = 0$$
$$F(0,1) = f(0,0) - f(0,1) + f(1,0) - f(1,1) = 0$$
$$F(1,1) = f(0,0) - f(0,1) - f(1,0) + f(1,1) = 0$$

Thus, the 2-D DFT of the given image is $\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$.

F(0,0) is 4 which happens to be the average of all the intensity values in original image. The other values represent frequency values. But since there is no variation in values of original image, there is no frequency involved, and that is why the frequency values in DFT are zeroes.

Alternatively, the 2-D DFT can also be found using the DFT basis matrix formed by finding 1-D DFT of each row of $f(x, y)$ and then using that as kernel.

That is $F(u, v) = \text{kernel} \times f(x, y) \times (\text{kernel})^T$

We have already found the DFT matrix of order 2 in Example 2.

Therefore,

$$\text{kernel} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Hence, $F(u, v) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ [we are skipping $\frac{1}{2}$ as in 2-D we have taken $\frac{1}{MN}$ in inverse DFT]

$$= \begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$
$$= \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

Both the results are same.

***

Try the following exercises.

E1) Does the implementation of a separable and symmetric transform, such as the DFT in an image requires the sequential implementation of the corresponding one-dimensional transform row-by-row and then column-by-column? Justify your answer.

E2) Find the DFT of the sequence $f(x) = [i, 0, i, 1]$.

E3) Construct a DFT matrix of order 4. Also, check whether DFT matrix is unitary matrix or not.

E4) Find the inverse 2-D DFT of $F(u,v)$ found in Example 3.

In the following section, we shall discuss discrete cosine transform.

3.3 DISCRETE COSINE TRANSFORM

The Discrete Cosine Transform DCT is a family of unitary transformations that transforms the real values of input image to another set of real values. Unlike the DFT that is complex, the DCT is a real transform because it projects the signal onto real cosinewaves.

The 1-D DCT is given as:

$$C(u) = a(u)\sum_{x=0}^{N-1} f(x)\cos\left[\frac{(2x+1)u\pi}{2N}\right]; 0 \le u \le N-1, \qquad (9)$$

where, $a(u) = \begin{cases} \sqrt{\frac{1}{N}}; & u = 0 \\ \sqrt{\frac{2}{N}}; & u = 1,\ldots,N-1 \end{cases}$

Then, the inverse transform is given by

$$f(x) = \sum_{u=0}^{N-1} a(u)\,C(u)\cos\left[\frac{(2x+1)\,u\pi}{2N}\right], \qquad (10)$$

where $a(u)$ is the same function as used for DCT.

Let us visualize the effect of 1-D DCT through Fig. 4, where in the rows of the 8×8 transformation matrix of the DCT for a signal $f(x)$ with 8 samples are shown.

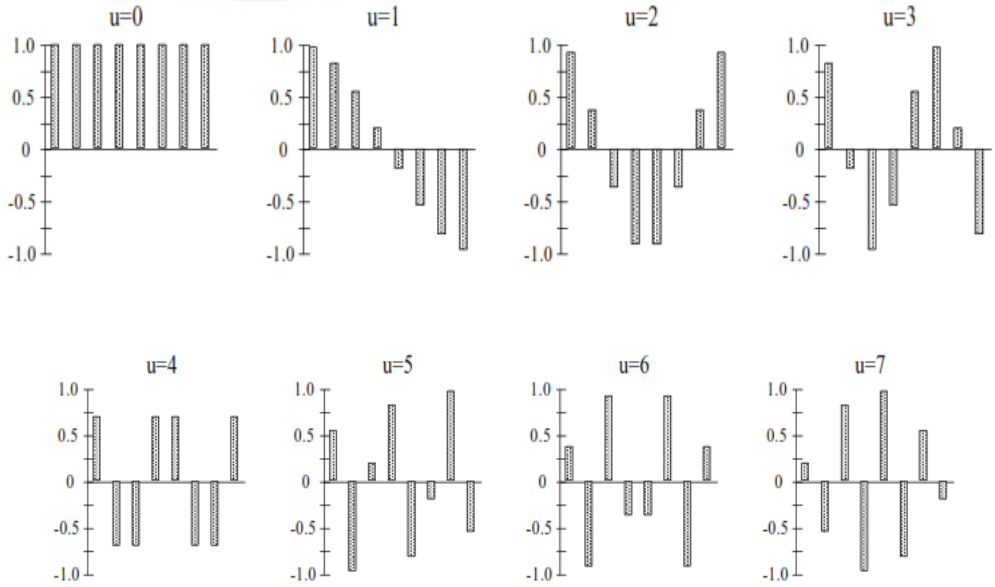


Fig. 4: 1-D DCT

The figures for values of u from 0 to 7 show the various rows of the 8×8 transformation matrix of the DCT for a 1D signal $f(x)$ with 8 samples.

Let us now expand 1-D DCT to 2-D DCT.

Consider an image $f(x,y)$ of size $M\times N$. Then, the 2-D DCT of the image is defined as:

$$C(u,v)=a(u)\,a(v)\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\cos\left[\frac{(2x+1)u\pi}{2M}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right], \quad (11)$$
$$0\le u\le M-1, 0\le v\le N-1$$

And the inverse 2-D DCT is given by

$$f(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}a(u)\,a(v)\,C(u,v)\cos\left[\frac{(2x+1)u\pi}{2M}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (12)$$
$$0\le x\le M-1, 0\le y\le N-1$$

Where $a(u)$ is same as defined earlier for 1D DCT.

DCT and DFT are very similar, however, DCT has the advantage over DFT that DCT is real while DFT is complex. Moreover. DCT has better energy compaction in comparison to the energy compaction of DFT. Energy compaction is the ability to pack the energy of the spatial sequence into as few frequency coefficients as possible. This property is exploited for image compression and is a very important property.You can see in the Fig.5, that most of the DCT image is dark, which means DCT values are concentrated only in few pixels very near the origin. This indicates that DCT has high compaction.

a) Original image – Lena **b) 2D DCT of Lena**

Fig.5: The 2D DCT in (b) of the image Lena in (a) shows the high compaction capability of DCT.

Example 4: Compute the discrete cosine transform (DCT) matrix for order 2.

Solution: Using Eqn. (9), we substitute $N=2$, and we get

$$C(u)=a(u)\sum_{x=0}^{1}f(x)\cos\frac{(2x+1)u\pi}{2\times2};\ 0\le u\le1.$$

where $a(u)=\begin{cases}\frac{1}{\sqrt{2}} & ; \quad u=0\\ \sqrt{\frac{2}{2}}=1; & u=1\end{cases}$

At $u=0$, we get

$$C(0)=\frac{1}{\sqrt{2}}\sum_{x=0}^{1} f(x)\cos\frac{(2x+1)\pi\times 0}{4}$$

$$=\frac{1}{\sqrt{2}}\sum_{x=0}^{1} f(x)\cdot 1$$

$$=\frac{1}{2}\sum_{x=0}^{1} f(x)$$

$$=\frac{1}{2}[f(0)+f(1)]$$

At $u=1$, we get

$$C(1)=1\cdot\sum_{x=0}^{1} f(x)\cdot\cos\frac{(2x+1)\cdot 1\cdot\pi}{4}$$

$$=\sum_{x=0}^{1} f(x)\cos\frac{(2x+1)\pi}{4}$$

$$=\left[f(0)\cos\frac{\pi}{4}+f(1)\cos\frac{3\pi}{4}\right]$$

$$=\frac{1}{\sqrt{2}}f(0)-\frac{1}{\sqrt{2}}f(1)$$

By collecting the coefficients, we get the required DCT. Therefore, DCT is

$$C(u)=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\end{bmatrix}$$

***

Now try the following exercises.

E5) Why is DCT important for image compression?

E6) Find the DCT of the matrix of order 4.

So far, we have discussed discrete fourier transform and discrete cosine transform.

3.4 HAAR TRANSFORM

The Haar transform is a wavelet transform. Wavelet transforms are based on small waves called wavelets which are of varying frequencies

and limited duration. These are different from the Fourier transform, where the basis functions are sinusoids. Haar transform is a transform whose basis functions are orthonormal wavelets. The Haar transform can be expressed as

$$T = HFH^T \tag{13}$$

where, F is an $N \times N$ image matrix, H is the $N \times N$ Haar transform matrix and T is the resulting $N \times N$ transform.

The Haar transform, H, contains the Haar basis functions, $h_k(t)$. They are defined on a continuous interval, $t \in [0,1]$ for $k = 0,1,...,N-1$, where $N = 2^n$. Then, H is generated by uniquely decomposing the integer k as $k = 2^p + q - 1$, where, $0 \le p \le n-1$ and when $p = 0, q = 0,1; p \ne 0$ then, $1 \le q \le 2^n$.

For example, when $N = 4$, k will take the values $k = 0,1,2,3$. For these the corresponding values of p and q have to satisfy that $k = 2^p + q - 1$.

Therefore, we compute the values of k, p and q in Table 1.

Table 1

k	0	1	2	3
p	0	0	1	1
q	0	1	1	2

Let t take the values from the set $\left\{\frac{0}{N}, \frac{1}{N}, ..., \frac{N-1}{N}\right\}$.

Then, the Haar basis functions are recursively defined as:

- For $k = 0$, the Haar function is defined as a constant

$$h_0(t) = 1/\sqrt{N} \tag{14}$$

- When $k > 0$, the Haar function is defined by

$$h_k(t) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2} & ; \quad \frac{(q-1)}{2^p} \le t < \frac{(q-0.5)}{2^p} \\ -2^{p/2} & ; \quad \frac{(q-0.5)}{2^p} \le t < \frac{q}{2^p} \\ 0 & ; \quad \text{othewise} \end{cases} \tag{15}$$

As can be seen from the definition of the Haar basis functions, for the non-zero part of the function, the amplitude and width is determined by p while its position is determined by q.

We now show how the Haar transform matrix can be computed at $n = m/N$, where $n = 0,1,...,N-1$ to form the $N \times N$ discrete Haar transform matrix through the following examples.

Example 5: For, $N = 2$, compute the discrete Haar transform of a 2×2 matrix.

Solution: Here, $N = 2$, we know that $N = 2^n$.

Substituting the value of N, we get $2 = 2^n$, which gives $n = 1$.

Since, $0 \le p \le n-1$, we get $0 \le p \le 0$.

Therefore, $p = 0$, and hence $q = 0, 1, 2$.

We determine the value of k using the relation $k = 2^p + q - 1$, we obtain

p	0	0
q	0	1
k	0	1

for $k = 0, h_0(t) = \frac{1}{\sqrt{2}}$ [using Eqn. (14)]

for $k = 1, h, (t) = \frac{1}{\sqrt{2}} \begin{cases} 1 & ; t = 0 \\ -1 & ; t = 1/2 \end{cases}$

Thus, Haar transform is $h_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

***

Example 6: For $N = 8$, the 8×8 discrete Haar transform matrix.

Solution: As you know we need to find various parameters of Haar transform. So, we find them as follows:

i) Here $N = 8$,

ii) $N = 2^n \Rightarrow n = 3$

iii) when $p = 0, q = 0,1$
$p = 1, q = 1,2$
$p = 2, q = 1,2,3,4$

iv) All the possible values of k for each set of p and q are given below:

p	0	0	1	1	2	2	2	2
q	0	1	1	2	1	2	3	4
k	0	1	2	3	4	5	6	7

v) Accordingly, $t = \left\{0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}\right\}$.

Now, we compute $h_k(t)$ for each k and t.

For $k = 0, h_0(t) = 0.$ for all t.

In general, $h_k(t) = \frac{1}{\sqrt{8}} \begin{cases} 2^{p/2} ; \frac{q-1}{2^p} \le t < \frac{q-0.5}{2^p} \\ -2^{p/2} ; \frac{q-0.5}{2^p} \le t < \frac{q}{2^p} \\ 0 \quad ; \text{otherwise} \end{cases}$. (16)

Now, let us find each $h_k(t)$ for each of the interval of t for a particular k using Eqn. (16) in the following table:

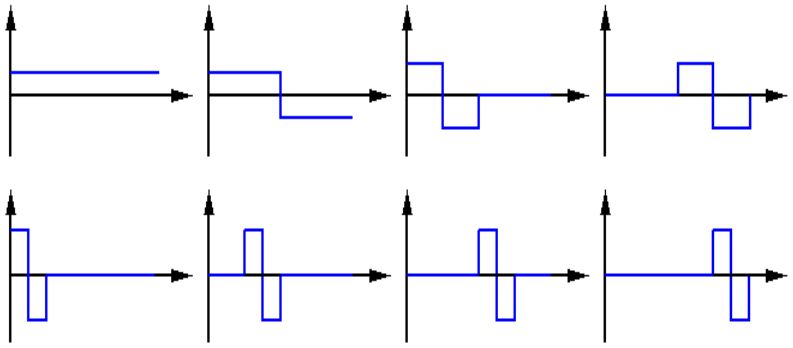
For $k = 1$

Parameters k, q, p	$h_k(t)$	Haar Transform after simplification
$k = 1,$ $q = 1,$ $p = 0$	$h_1(t) = \frac{1}{\sqrt{8}} \begin{cases} 1; \frac{0}{2^0} \le t < \frac{1-0.5}{2} \Rightarrow 1 \le t < \frac{1}{2} \\ -1; \frac{1-0.5}{2} \le t < \frac{1}{2^0} \Rightarrow \frac{1}{2} \le t < 1 \\ 0; \text{otherwise} \end{cases}$	$h_1(t) = \frac{1}{2\sqrt{2}}$; for $t = 0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}$ $h_1(t) = \frac{-1}{2\sqrt{2}}$; for $t = \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}$
$k = 2,$ $q = 1,$ $p = 1$	$h_2(t) = \frac{1}{\sqrt{8}} \begin{cases} \sqrt{2}\ ; 0 \le t < \frac{1}{4} \\ -\sqrt{2}\ ; \frac{1}{4} \le t < \frac{1}{2} \\ 0\ ; \text{otherwise} \end{cases}$	$h_2(t) = \frac{1}{2}$; for $t = 0, \frac{1}{8}$ $h_2(t) = \frac{-1}{2}$; for $t = \frac{2}{8}, \frac{3}{8}$ $h_2(t) = 0$; for $t = \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}.$
$k = 3,$ $q = 2,$ $p = 1$	$h_3(t) = \frac{1}{\sqrt{8}} \begin{cases} \sqrt{2}\ ; \frac{1}{2} \le t < \frac{3}{4} \\ -\sqrt{2}\ ; \frac{3}{4} \le t < 1 \\ 0\ ; \text{otherwise} \end{cases}$	$h_3(t) = \frac{1}{2}$; for $t = \frac{4}{8}, \frac{5}{8}$ $h_3(t) = \frac{-1}{2}$; for $t = \frac{6}{8}, \frac{7}{8}$ $h_3(t) = 0$; for $t = 0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}.$
$k = 4,$ $q = 1,$ $p = 2$	$h_4(t) = \frac{1}{\sqrt{8}} \begin{cases} 2\ ; 0 \le t < \frac{1}{8} \\ -2\ ; \frac{1}{8} \le t < \frac{1}{4} \\ 0\ ; \text{otherwise} \end{cases}$	$h_4(t) = \frac{1}{\sqrt{2}}$; for $t = 0$ $h_4(t) = \frac{-1}{\sqrt{2}}$; for $t = \frac{1}{8}$ $h_4(t) = 0$; for $t = \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}$

$k=-5,$ $q=2,$ $p=2$	$h_5(t)=\frac{1}{\sqrt{8}}\begin{cases}2\,;\,\frac{1}{4}\le t<\frac{3}{8}\\-2\,;\,\frac{3}{8}\le t<\frac{1}{2}\\0\,;\ \text{otherwise}\end{cases}$	$h_5(t)=\frac{1}{\sqrt{2}}$; for $t=\frac{2}{8}$. $h_5(t)=\frac{-1}{\sqrt{2}}$; for $t=\frac{3}{8}$ $h_5(t)=0$; for $t=0,\frac{1}{8},\frac{4}{8},\frac{5}{8},\frac{6}{8},\frac{7}{8}$
$k=6,$ $q=3,$ $p=2$	$h_6(t)=\frac{1}{\sqrt{8}}\begin{cases}2\,;\frac{1}{2}\le t<\frac{5}{8}\\-2\,;\frac{5}{8}\le t<\frac{3}{4}\\0\,;\ \text{otherwise}\end{cases}$	$h_6(t)=\frac{1}{\sqrt{2}}$; for $t=\frac{4}{8}$ $h_6(t)=\frac{-1}{\sqrt{2}}$; for $t=\frac{5}{8}$ $h_6(t)=0$; for $t=0,\frac{1}{8},\frac{2}{8},\frac{3}{8},\frac{6}{8},\frac{7}{8}$
$k=7,$ $q=4,$ $p=2$	$h_7(t)=\frac{1}{\sqrt{8}}\begin{cases}2\,;\frac{3}{4}\le t<\frac{7}{8}\\-2\,;\frac{7}{8}\le t<1\\0\,;\ \text{otherwise}\end{cases}$	$h_7(t)=\frac{1}{\sqrt{2}}$; for $t=\frac{6}{8}$ $h_7(t)=\frac{-1}{\sqrt{2}}$; for $t=\frac{7}{8}$ $h_7(t)=0$; for $t=0,\frac{1}{8},\frac{2}{8},\frac{3}{8},\frac{4}{8},\frac{5}{8}.$

The Haar transform is given in the following matrix.

$$h_k(t)=\begin{bmatrix}\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}\\ \frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{1}{2\sqrt{2}}&\frac{-1}{2\sqrt{2}}&\frac{-1}{2\sqrt{2}}&\frac{-1}{2\sqrt{2}}&\frac{-1}{2\sqrt{2}}\\ \frac{1}{2}&\frac{1}{2}&\frac{-1}{2}&\frac{-1}{2}&0&0&0&0\\ 0&0&0&0&\frac{1}{2}&\frac{1}{2}&\frac{-1}{2}&\frac{-1}{2}\\ \frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}&0&0&0&0&0&0\\ 0&0&\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}&0&0&0&0\\ 0&0&0&0&\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}&0&0\\ 0&0&0&0&0&0&\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\end{bmatrix}$$



The plot of these 8 basis functions are shown in Fig. 6.

Fig. 6: Haar Basis Functions

As can be seen by Fig.6, all non-zero Haar functions $h_k(t), k > 0$ consists of a square wave and its negative version, and the parametersp defines the magnitude and width of the shape while q specifies the position (or shift) of the shape. This gives the unique property to the Haar transform that it not only represents the signal at different scales based on the different frequencies, but also represents their locations across time.

Moreover, an important property of the Haar transform matrix is that it is real and orthogonal, that is, $H = H^*$ and $H^{-1} = H^T$. The orthogonal property of the Haar transform allows the analysis of the frequency components of the input signal. The Haar transform can also be used for analyzing the localized feature of the signal.

a) Original Image **b) Output of Haar Transform**

Fig. 7: Haar Transform

Fig. 7 (b) shows the output of Haar Transform of the image in Fig. 7 (a).

Try the following exercises.

E7) Let $X = [x[0], x[1], x[2], x[3]]^T = [1, 2, 3, 4]^T$. Then, X is a 4-point signal. Find the Haar transform coefficients and show that the signal can be expressed as a linear combination of the basis functions by the inverse transform.

E8) For $N = 4$, compute h_4, which represents the 4×4 discrete Haar transform matrix.

Now let us, summarize what we have discussed in this unit.

3.5 SUMMARY

In this unit, we discussed transformations which convert the spatial domain image to the frequency domain. We saw that these transform provide a variety of information based on the frequency content of the image.We discussed in depth three very important image transforms, namely the Discrete Fourier transformation (DFT), the Discrete Cosine Transformation (DCT) and the Haar transform. We also discussed the properties of each of these transforms, which shall help us in using them for image filtering in the frequency domain.

3.6 SOLUTIONS AND ANSWERS

E1) Consider an image $f(x,y)$ of size $M \times N$ and the generic image transform T where, x indicates row and y indicates column.

Then,

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

If, T is separable and symmetric, then we can write $g(u,v)$ as

$$g(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} T_1(u,x) T_2(v,y) f(x,y)$$

$$\Rightarrow g(u,v) = \sum_{x=0}^{M-1} T_1(u,x) \sum_{y=0}^{N-1} T_2(v,y) f(x,y)$$

Then, $\sum_{y=0}^{N-1} T_2(v,y) f(x,y)$ is the same as applying the one-dimensional transform long the x-th row of the image. By doing this for each of the M rows of the image, we obtain an intermediate image $F(x,v)$.

Then, $g(u,v) = \sum_{x=0}^{M-1} T_1(u,x) F(x,v)$

Therefore, we can see, the above sum corresponds to applying the one dimensional transform along the v-th column of the intermediate image (x,v).

Therefore, we have shown that the implementation of a separable and symmetric transform in an image requires the sequential implementation of the corresponding one-dimensional transform row-by-row and then column-by-column (or the inverse). We explain this in the Fig. 9:

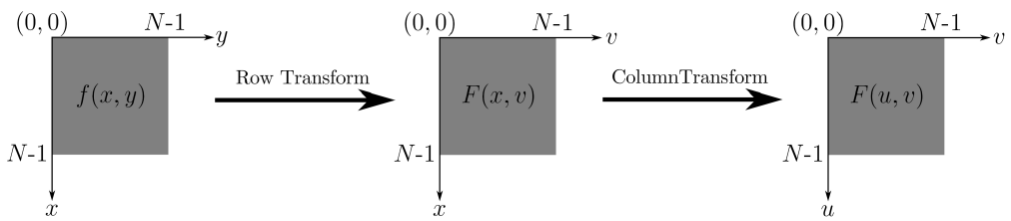


Fig. 9

E2) $$g'(u) = \frac{1}{4}\sum_{x=0}^{3} f(x) e^{\frac{-i \cdot 2\pi ux}{4}}; \; u = 0, 1, 2, 3$$

$$= \frac{1}{4}[i + i(-i)^{2u} + (-i)^{3u}]; \; u = 0, 1, 2, 3$$

$$= \frac{1}{4}[i + i(-1)^{2u} + (i)^{u}]; \; u = 0, 1, 2, 3$$

$$\mathbf{g} = \frac{1}{4}[1 + 2i, \; i, \; -1 + 2i, \; -i].$$

E3) Here $N = 4$.

$$g(u) = \frac{1}{4}\sum_{x=0}^{3} f(x) e^{\frac{-i\,2\pi \cdot ux}{4}}; \; u = 0, 1, 2, 3$$

$$g(u) = \frac{1}{4}[f(0) + (-i)^{u} f(1) + (-i)^{2u} f(2) + (-i)^{3u} f(3)]; \; u = 0, 1, 2, 3.$$

$$g(0) = \frac{1}{4}[f(0) + f(1) + f(2) + f(3)]$$

$$g(1) = \frac{1}{4}[f(0) - i f(1) + f(2) + i f(3)]$$

$$g(2) = \frac{1}{4}[f(0) - f(1) + f(2) - f(3)]$$

$$g(3) = \frac{1}{4}[f(0) + i f(1) - f(2) + i f(3)]$$

Hence, the DFT matrix is $A = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix}$.

You may check if $A \times A^{*T} = I$.

E4) Using Eqn. (5), $f(x, y) = \frac{1}{2 \times 2}\sum_{x=0}^{1}\sum_{y=0}^{1} F(u, v) . e^{2\pi i\left(\frac{ux}{2} + \frac{vy}{2}\right)}; u, v = 0, 1$

$$= \frac{1}{4}\sum_{x=0}^{1}\sum_{y=0}^{1} F(u, v) (1)^{ux} (1)^{vy}; u, v = 0, 1$$

$$= \frac{1}{4}\left[f(0,0) + (1)^{v} F(0,1) + (1)^{u} F(1,0) + (1)^{u} (1)^{v} F(1,1)\right]; u, v = 0, 1$$

which gives $f(x, y) = \frac{1}{4}\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

E5) In general, in most of the images, large part of the signal energy lies at the low frequencies which appear in the upper left corner of the DCT image. Since the higher frequencies present in the lower right of the image are small enough to be neglected, the original image can be represented in less number of coefficients, thereby achieving compression. Therefore, as DCT has good compaction property, it can represent the original image in less

number of coefficients and therefore, storage and transmission of the image is better and faster. Moreover, the original image can be recreated close to the original from the most important components of the DCT.

E6) $$C(u)=\begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.65 & 0.27 & -0.27 & -0.65 \\ 0.5 & -0.5 & -0.5 & 0.5 \\ 0.27 & -0.65 & 0.65 & -0.27 \end{bmatrix}$$

E7) $X=[x[0],x[1],x[2],x[3]]^T=[1,\ 2,\ 3,\ 4]^T$ be the 4-point signal. Then, we shall use the basis matrix, H_4 to compute the Haar transform coefficients.

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}=\begin{bmatrix} 5 \\ -2 \\ -1\sqrt{2} \\ -1\sqrt{2} \end{bmatrix}$$

The inverse transform will be:

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & \sqrt{2} & 0 \\ 1 & 1 & -\sqrt{2} & 0 \\ 1 & -1 & 0 & \sqrt{2} \\ 1 & -1 & 0 & -\sqrt{2} \end{bmatrix}\begin{bmatrix} 5 \\ -2 \\ -1\sqrt{2} \\ -1\sqrt{2} \end{bmatrix}$$

$$=\frac{1}{2}\left[5\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}-2\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}-\frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{2} \\ -\sqrt{2} \\ 0 \\ 0 \end{bmatrix}-\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 0 \\ \sqrt{2} \\ -\sqrt{2} \end{bmatrix}\right]=\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

As can be seen, X is a linear combination of the basis vectors.

E8) Here $N=4; n=2; p=0,1; t=\left\{0,\frac{1}{4},\frac{2}{4},\frac{3}{4}\right\}$. Let us write all the values of Haar transform in the following table:

k,q,p	$h_k(t)$	$h_k(t)$ after simplification
$p=0, q=0, k=0$	$h_0(t)=\frac{1}{\sqrt{4}}=\frac{1}{2}$ for all t	$h_0(t)=\frac{1}{2}$ for $t=0,\frac{1}{4},\frac{2}{4},\frac{3}{4}$.

$p=0,q=1,k=1$	$h_1(t)=\frac{1}{2}\begin{cases}1;\ 0\le t<\frac{1}{2}\\ -1;\ \frac{1}{2}\le t<1\\ 0;\text{otherwise}\end{cases}$	$h_1(t)=\frac{1}{2};t=0,\frac{1}{4}$ $h_1(t)=\frac{-1}{2};t=\frac{2}{4},\frac{3}{4}$
$p=1,q=1,k=2$	$h_2(t)=\frac{1}{2}\begin{cases}\sqrt{2};0\le t<\frac{1}{4}\\ -\sqrt{2};\frac{1}{4}\le t<\frac{1}{2}\\ 0;\text{otherwise}\end{cases}$	$h_2(t)=\frac{1}{\sqrt{2}};t=0$ $h_2(t)=\frac{-1}{\sqrt{2}};t=\frac{1}{4}$ $h_2(t)=0;t=\frac{2}{4},\frac{3}{4}.$
$p=1,q=2,k=3$	$h_3(t)=\frac{1}{2}\begin{cases}\sqrt{2};\ \frac{1}{2}\le t<\frac{3}{4}\\ -\sqrt{2};\frac{3}{4}\le t<1\\ 0;\text{otherwise}\end{cases}$	$h_3(t)=\frac{1}{\sqrt{2}};t=\frac{2}{4}$ $h_3(t)=\frac{1}{\sqrt{2}};t=\frac{3}{4}$ $h_3(t)=0;t=0,\frac{1}{4}.$

Hence,

$$h_k(t)=\begin{bmatrix}\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & \frac{-1}{2} & \frac{-1}{1}\\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0\\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\end{bmatrix}$$

UNIT 4

COLOUR IMAGE PROCESSING

Structure **Page No.**

4.1 INTRODUCTION

The purpose of this unit is to introduce the concepts related to colour image processing. As we have been working with grayscale images till now, we would like to have an in-depth discussion of how colour images are formed and the various colour models that exist.

We shall first discuss the human vision system in Sec. 4.2. A healthy vision system is capable of seeing the world in colour.

As you read further, we shall discuss the various colour models that exist and the advantages and disadvantages of each in Sec. 4.3.

Finally, we shall discuss pseudo-colour processing, also called false colour, which is the process of assigning colours to grey values based on specified conditions. We shall discuss colour processing in Sec. 4.5. We finally summarise the discussion in Sec. 4.6 and in Sec. 4.7, we give the solutions/answers/hints to the exercises.

Now we shall list the objectives of this unit. After going through the unit, please read this list again and make sure that you have achieved the objectives.

Objectives

After studying this unit, you should be able to:

- to differentiate between thousands of thousands of colours and their shades in the colour images.
- To define the different colour models and use them as per the requirements.
- To apply different pseudo colour models

A colour image is a powerful source of information. Human visual system has the ability to differentiate between hundreds of colours and their shades. Therefore, colour images contain a large amount of extra information compared to grey-scale images, that give a better understanding of the contents of the image, for example, in object detection and segmentation. If an image is captured by a full-colour sensor, then the resulting image is a full colour image.

A grayscale image can be converted into a colour image using the technique of pseudo-colour processing, where each intensity is assigned a colour.

Full colour image processing is primarily used in most applications such as visualisation and publishing. We start with discussion on human vision system in the following section.

4.2 HUMAN VISION SYSTEM

The human eye is nearly spherical in shape with a 20mm diameter on an average, as shown in Fig. 4.1. The three main parts of the eye are:

i) The cornea and sclera outer cover,

ii) the choroid and

iii) the retina.

Let us discuss them one by one briefly.

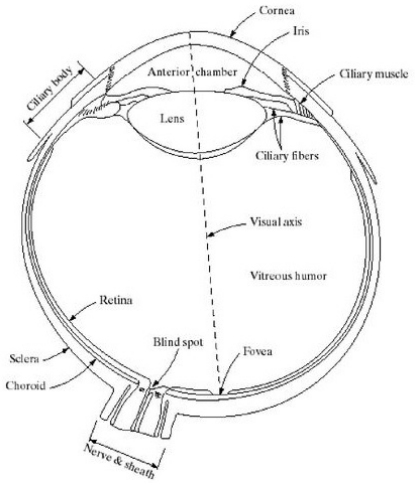


Fig.1: Structure of the human eye source

i) As you see in Fig. 1, the sclera is an opaque member that encloses theoptic globe all around, except at the anterior end, which is covered by the cornea. The cornea is a tough, transparent cover of the anterior chamber.

ii) Choroid is the layer under the sclera. The membrane choroid contains a network of blood vessels. These blood vessels form the major source of nutrition for the eyes. If the choroid is damaged and inflamed, it can restrict blood flow in the eye, resulting in serious damage of the eye cells. The role of the choroid is also to control the amount of light entering the eye as well as reduce the backscatter inside the eye.

The choroid is divided into two parts:

a. The ciliary muscles which relax and tighten to enable the lens to focus by changing its shape,
b. The iris diaphragm, that contracts and expands to control the amount of light that enters the eye.

The lens is a transparent, biconvex structure that helps to refract light into the eye such that the image is formed on the retina. The lens is flexible and ca change shape to change the focal length of the eye. THSI ensures that objects at various distances can be focussed upon and their images can be formed on the retina.

iii) The retina is the innermost membrane of the eye. It lines the wall of the complete posterior portion of the eye. The retina can be thought of as the image plane in the eye, since on properly focussing the eye on an object, light from that object is passed through the lens such that the image is formed on the retina.

The retina consists of two types of cells called rods and cones. The cones are highly sensitive to colour and are around 6-7 million in a human eye. The cones are located on the focea which is the central portion of the retina.

However, there are 75-150 million rod cells which are completely distributed all over the retina. The rod cells give the overall picture of the object in the scene, and reduce the amount of detail. Rods are also responsible for low light vision, also known as **SCOTOPIC vision**, while cones are responsible for bright light vision, also known as **PHOTOPIC vision**.

In the human eye, the distance between the retina and lens, that is the focal length, varies between 14 and 17 mm as the refractive power of the lens increases from min to max. For a nearby object, the lens is most strongly refractive. Moreover, the lens of the eye is very flexible and is flattened by controlling muscles to enable the eye to focus on distant objects.

To allow the eye to focus on objects close to the eye, the controlling muscles allow the lens to become thicker.

Here, you might be wondering how human eye adapts to different levels of brightness and how it discriminates various levels of brightness. The answer to your question is given below.

Brightness Adaptation and Discrimination: Human vision system is highly complex and can adapt to an enormous range of light intensity levels-of the order of 10^{10}. The range starts from the scotopic threshold and goes upto the glare limit. The subjective brightness, the perceived intensity by the human eye, has been experimentally found to be a logarithmic function of the light intensity that falls on the eye. Since the human eye cannot interpret this dynamic range simultaneously, brightness adaptation is carried out by the eye. The eye can discriminate only a small range of distinct intensity levels simultaneously. Brightness adaption level is the current sensitivity level of a human eye for a given set of conditions.

Now, try the following exercises.

E1) If an observer is looking at a tree that is 100m far and if h is the height of the tree in mm in the retinal image, what is h?

So, by now you know the fundamental concepts about human vision system. In the following section, we are going to highlight various colour models. You must have heard about some of them in your day-to-day life.

4.3 COLOUR FUNDAMENTALS

Every colour is defined using three quantities that are independent of each other, however, taken together they define a particular shade of a colour. These quantities are:

i) **Hue:** Hue component of a colour is defined by the dominant wavelength. The wavelength range on the electromagnetic spectrum that defines the visible colour spectrum lies between 400nm [nm is nanometre] that represents the violet colour and 700nm that represents the red colour as can be seen in Fig. 2.

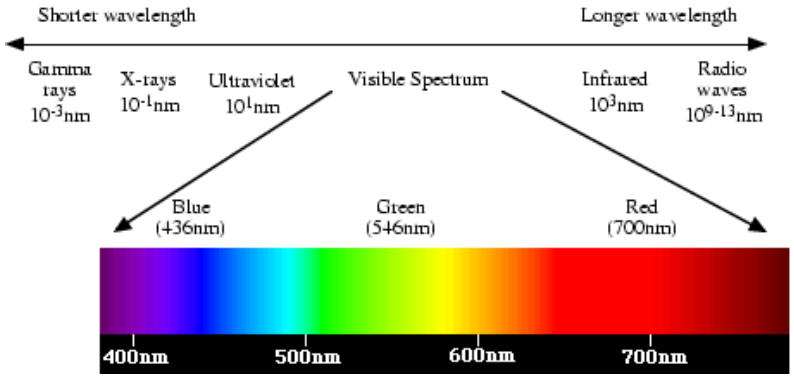


Fig. 2: Part of the electromagnetic spectrum that shows the visible spectrum
Source:

ii) **Saturation**: The excitation purity of the colour is determined by the quantity known as **saturation**.It is dependent on the amount of white light that is mixed with hue of that colour. A fully saturated colour implies that no white light is mixed with that hue.

iii) **Chromaticity**: The sum of hue and saturation constitutes the *chromaticity* of the colour. Therefore, if there is no colour, it is called achromatic light.

iv) **Intensity**: The amount of light actually present defines the *intensity*. Therefore, intensity is a physical quantity. If more light is present, the colour is more intense. Achromatic light has only intensity but no colour. Grayscale images have only intensity

v) **Luminance or Brightness**: The perception of colour is the quantity known as luminance or brightness. For example, given two colours of the same intensity, such as blue and green, it is perceived that blue is much darker than green.

vi) **Reflectance**: The ability of an object to reflect light, is the reflectance property of the object. The reflectance property determines the colour of the object, since, we see those colours that are reflected back and not the ones that are absorbed. For example, an object that reflects green, absorbs all other colours in the white light spectrum except green.

There are about 6 to 7 million cones in the human eye and they are responsible for recognising colours. Nearly, 65% of the cones recognise red, 33% are sensitive to green and about 2% to blue. Red, green and blue are known as the primary colours and nearly all other colours are seen as a combination of these primary colours. However, there is a difference between the primary colours of light and the primary colours of pigments. The primary colours of light are red, blue and green and they can be added to produce the secondary colours of light that are yellow(red plus green), magenta (red+blue) and cyan (green + blue). Moreover, the primary colours of pigments are said to be those that absorbs a primary colour of light and reflects the other two. Therefore, in tHSI case, the primary colours are cyan, magenta and yellow while the secondary colours are red, blue and green.

For standardisation, in 1931, the *Commission Internationale de l'Éclairage (CIE)* defined specific wavelengths for the three primary colours: red = 700nm, blue: 435.8 nm and green = **546.1** nm. The amounts of red, blue and green required to form a colour are known as the Tristimulus values and are denoted as X, Y and Z.

Given X, Y and Z the tristimulus coefficients which define a colour. If we define x, y and z as the relative values of the primary colours, then these values can be found by

$$x = \frac{X}{X+Y+Z},\ Y = \frac{Y}{X+Y+Z} \text{ and } z = \frac{Z}{X+Y+Z} \quad \ldots (1)$$

It is obvious that

$$x + y + z = 1. \quad \ldots (2)$$

Thus a 2-D diagram is adequate to show the coordinates x and y.

If we specify colours as a composition as x (red) and y (green). Then, given the values of x and y, the value of z (blue) can be computed as:

$$z = 1 - (x + y) \quad \text{... (3)}$$

Here, we can see only two variables are independent. Therefore, we can show these variables in 2-D coordinate system.

The point on the boundary of the chromaticity chart is fully saturated, while as a point moves farther from the boundary, more white light is added and is therefore, less saturated. The saturation is zero at the point of equal energy. A straightine joining any two points in the chromaticity diagram, determines all possible colours that can be obtained by combining the two colours at the endpoints of the segment. This can be extended to combining three colours. The three line segments joining the points pairwise form a triangle and various combinations of the colours at the vertices of this triangle give all colours inside the triangle or on the boundary of the triangle.

To understand this more clearly, we shall discuss few examples.

Example 1: Consider the coordinates of warm white $(0.45, 0.4)$ and the coordinates of deep blue $(0.15, 0.2)$. Find the percentage of the three colours red (X), green (Y) and blue (Z).

Solution: We first find the trichromatic coefficients x, y and z. At the point warm white, $x = 0.45$ and $y = 0.4$, therefore

$$\begin{aligned} z &= 1 - (x + y) \\ &= 1 - (0.45 + 0.4) \\ &= 0.15 \end{aligned}$$

Now, we shall find the tristimulus values X, Y and Z.

$$\frac{X}{X+Y+Z} = 0.45$$
$$\frac{Y}{X+Y+Z} = 0.4$$
$$\frac{Z}{X+Y+Z} = 0.15$$

Here, $X : Y : Z = 0.45 : 0.4 : 0.15$

Therefore the percentage of each colour would be as follows:

Percentage of red $(X) = 45\%$
Percentage of green $(Y) = 40\%$
Percentage of Blue $(Z) = 15\%$

At the point deep blue, $x = 0.15, y = 0.2,$ therefore $z = 0.65.$

We can find the percentage of each colour as we found in case of warm white. We get percentage of red colour as $15\%,$ percentage of green colour as 20% and the percentage of blue colour as $65\%.$

We can see the percentage is justified for each colour name.

***

Example 2: Find the relative percentage of colours warm white and deep blue which mixes the give the colour which lies on the line joining them. Use the coordinates of these points as given in Example 1.

Solution: Let the colour C lies on the line have the coordinate $(x, y).$

The distance of C from the warm white colour $= \sqrt{(x-0.45)^2 + (y-0.4)^2}$

Similarly, the distance of C from the deep blue colour
$= \sqrt{(0.15-x)^2 + (0.2-y)^2}$

The percentage of warm white in

$$C = \frac{\sqrt{(x-0.45)^2 + (y-0.4)^2} - \sqrt{(0.15-x)^2 + (0.2-y)^2}}{\sqrt{(0.45-0.15)^2 + (0.4-0.2)^2}} \times 100$$

This expression can be used to find the percentage of warm white colour at C by substituting the coordinates of the point C as per the situation. Also, the percentage of the deep blue colour would be $(100 -$ percentage of warm white colour).

***

Now try the following exercises.

E2) Derive an expression to find the percentage of each colours $C_1, C_2,$ and C_3 at the point C which lies within the triangle having vertices as $C_1, C_2,$ and $C_3.$

In the following section, we shall discuss the most commonly used colour models such as the RGB (red, green, blue), CMY (Cyan, magenta, yellow) and HSI (hue, saturation, intensity).

4.4 COLOUR MODELS

Colour models or Colour spaces or Colour systems have been introduced so as to be able to specify each colour in a generally accepted manner. There are various colour models or colour spaces. Each colour space specifies a particular colour in a standard manner, by specifying a 3-D coordinate system and a subspace that contains all possible colours in that colour model. Then, each colour in that colour space is represented as a point in that subspace, given by three coordinates $(x, y, z).$ These colour models are either oriented towards

specific hardware or image processing applications. In this section, we shall discuss three important colour models and the conversion of one colour model into other.

Before we discuss each colour model, let us discuss the principles of absorption of colours of any model by human eye.

i) The human eye has absorption characteristics of colours and recognises them as variables. Thus, the colours red (R), green (G) and blue (B) are called **primary colours** of light.

ii) Secondary colours of light are produced by adding primary colours. For example red and blue produces magenta, red and green produces yellow, green and blue produces cyan, etc.

iii) Proportion of primary and secondary colours in appropriate amount produces white light.

Now, let us discuss each model separately.

4.4.1 The RGB Model

The RGB colour is based on a cartesian coordinate system, where the colour subspace is a cube with axes representing red, green and blue. A colour in the RGB model is therefore, specified as a 3-tuple (R,G,B) where, R,G and B represent the amount of red, green and blue, respectively, present in that colour. The geometry of the RGB colour model is a cube as shown in Fig. 3. It is assumed that all colour values are normalised, that is, it is a unit cube and all values of R, G and B lie between 0 and 1. It is clear from Fig. 3, that it is a model of a cube with the three axes for Red, Green and blue. The grayscale values lies on the diagonal of the cube, which joins the black $(0,0,0)$ and white $(1,1,1)$ vertices of the cube.

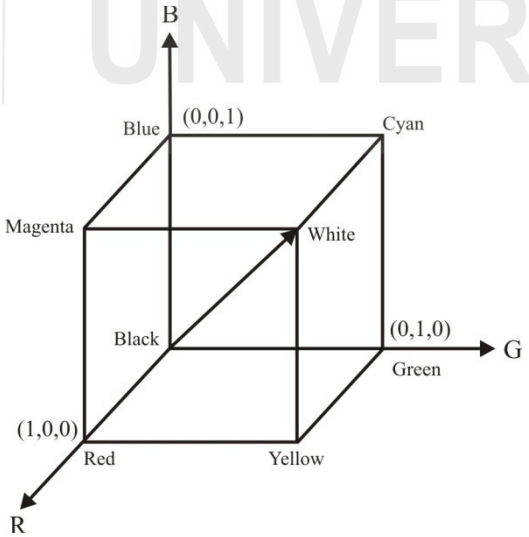


Fig. 3: The RGB colour (Image taken from [1])

A colour image in the RGB model consists of three images corresponding to each of the three colours: Red (R), Green (G) and Blue (B) colours. These three images combine to form one composite colour image on a monitor. To covert an RGB image to a grayscale

image, the intensity of the gray-pixel is given by the average of R, G and B values. The RGB colour model is mainly used for colour monitors and screens.

Now the question arises how do we find the composite colour in RGB colour model at any point. For this we follow the following steps:

Step 1: Pixel depth is the number of bits used to represent each pixel. If an image in RGB model has 8-bit image in each of its three colours, then each RGB pixel has a depth of 3 image planes $\times$8-bit per plane that is 24 bits. This gives rise to 2^{24} colour shades.

Step 2: We fix one of the three colours and let the other two colours to vary. Suppose we fix $R = 127$ and let G and B to vary. Then the colour at any point on the plane parallel to GB plane would be $(127, G, B)$, where, $G, B = 0, 1, ..., 255.$

Example 3: In a RGB image, the R and B components are at mid and the G component is at 1, then which colour would be seen by a person?

Solution: At the given point, we have

$$\frac{R}{2} + \frac{B}{2} + G = \frac{1}{2}(R + G + B) + \frac{G}{2}$$

$$= \text{mid grey} + \frac{1}{2}G.$$

$$= \text{Pure green with some grey component.}$$

***

Now, try an exercise.

E3) How many different shades of grey are there in a colour RGB system if each RGB image is an 8 bit image?

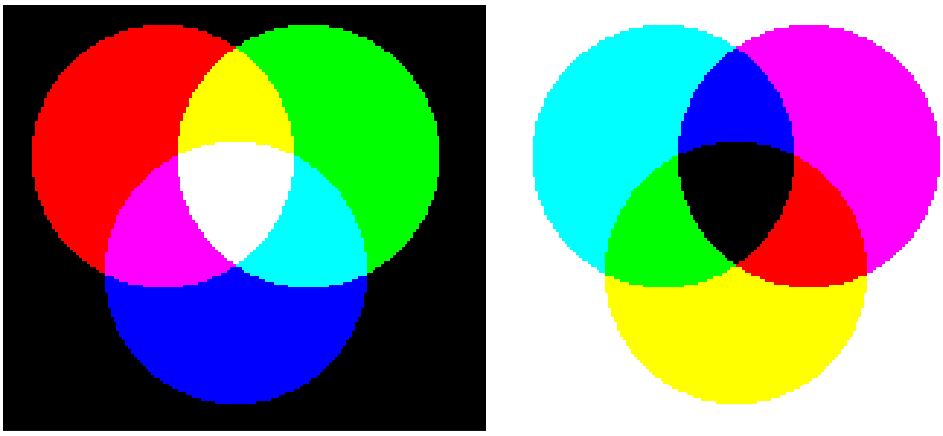
After discussing RGB model, we discuss CMY and CMYK colour models.

4.4.2 The CMY and CMYK Colour Model

Cyan (C), Magenta (M) and Yellow (Y) are the primary colours of pigments and the secondary colours of light. The CMY model is a subtractive model, implying that it subtracts a colour from the white light and reflects the rest. For example, when white light is reflected on cyan, it subtracts Red and reflects the rest. While RGB is an additive model, where something is added to black (0,0,0) to get the desired colour CMY is a **subtractive model**. The conversion between CMY and RGB model is given by the Equation below.

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \text{... (4)}$$

where, the RGB values have been normalised. THSI also gives a method to convert from RGB to CMY to enable printing hardcopy, since the CMY model is used by printers and plotters.



(a) (b)

Fig. 4: (a) RGB and (b) CMY

Fig. 4(a) shows the RGB model, in which the white colour is produced by adding the three primary colours Red, Green and Blue. Fig. 4 (b) shows the CMY model, where black is obtained as the sum of Cyan, Magenta and Yellow. The inverse relation between the RGB and CMY models are also shown by these two images.

In practice, black is obtained by combining cyan, yellow and magenta, however, this leads to a muddy looking black. For publishing, the black colour plays an important role, therefore, in the CMYK colour model, black is added as the fourth colour, where K stands for black.

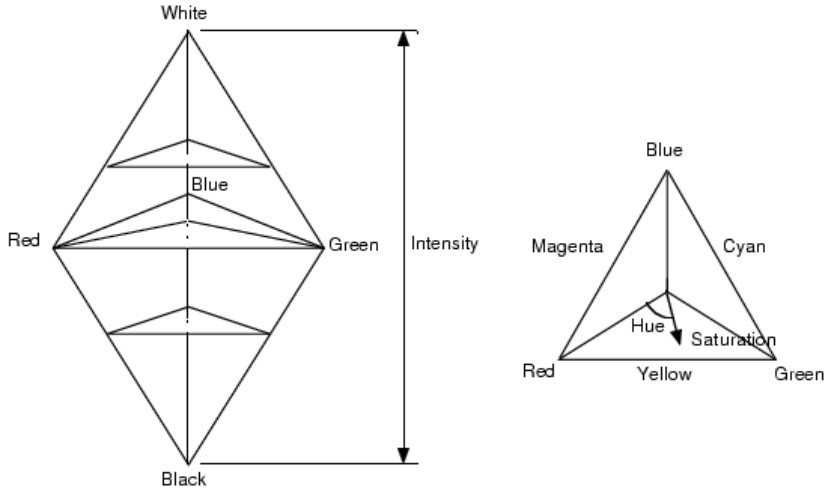
Now check by doing the following exercise, what have you understood.

E4) Why do we get green coloured paint on mixing blue and yellow coloured paints?

Now, let us discuss HSI model.

4.4.3 The HSI Model

This colour model is very close to human colour perception which uses the hue, saturation and intensity components of a colour, when we see a colour, we cannot describe it in terms of the amount of cyan, magenta and yellow that the colour contains. Therefore, the HSI colour model was introduced to enable describing a colour by its hue, saturation and intensity/ brightness. Hue describes the pure colour, saturation describes the degree of purity of that colour while intensity describes the brightness or colour, sensation. In a grayscale image, intensity defines the graylevel. Fig. 6 shows the HSI colour model and the way colours may be specified by this colour model.



(a) **(b)**

Fig. 5: (a) HSI and (b) RGB

In Fig.5, the HSI colour model is represented and its relation to RGB model is shown in the Fig. 5 (b). The HSI triangle in Fig. 5 (b) shows a slice from the HSI solid at a particular intensity as shown in Fig. 5 (a).

You may notice that in Fig. 5, the hue, saturation and intensity values required to form the HSI colour space can be computed using the RGB values.

Try the following exercises.

E6) Write the full form of HSI and define each of the components.

E7) What is colour space? Mention its classification.

Now we indicate how a RGB colour model is converted to a HSI colour model.

To convert an image in RGB format to HSI colour space, the RGB value of each pixel in the image, is converted to the corresponding HSI value in the following manner. Hue, H is given by

$$H = \begin{cases} \theta & \text{if } B \le G \\ 360 - \theta & \text{if } B > G \end{cases} \quad \text{... (5)}$$

where,

$$\theta = \cos^{-1}\left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\}$$

Saturation, S is given by

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)]$$

And, intensity, I is given by

$$I = \frac{1}{3}(R + G + B)$$

Where, the RGB values have been normalised in the range $[0,1]$ and the angle θ is measured with respect to the red axis in the HSI space.

Now we would convert HSI colour model to RGB colour space.

Given pixel values in the HSI colour space in the interval $[0,1]$, the RGB values can be computed in the same range. However, depending on the H value, the RGB values are computed in different sectors, based on the separation of the RGB colours by 120° intervals.

First, multiply the H value by 360°, to get the hue value in the interval [0°, 360°].

In the RG sector, H takes the value in the interval [0°, 120°[,

that is, $0^o \le H < 120^o$ and the RGB values are given by

$$B = 1(1 - S) \quad \text{... (8)}$$

$$R = 1\left[1 + \frac{S\cos H}{\cos(60^o - H)}\right] \quad \text{... (9)}$$

$$G = 31 - (R + B) \quad \text{... (10)}$$

In the **GB Sector,** $120^o \le H < 240^o$ then in this case, we first convert the value of H as

Then, the RGB values are computed as

$$R = I(1 - S) \quad \text{... (11)}$$

$$G = I\left[1 + \frac{S\cos H}{\cos(60^o - H)}\right] \quad \text{... (12)}$$

$$B = 31 - (R + B) \quad \text{... (13)}$$

In BR sector, when, $240^o \le H \le 360^o$, we first convert H as $H = H - 240^o$

Then, the RGB values are

$$G = I(1 - S) \quad \text{... (14)}$$

$$B = I\left[1 + \frac{S\cos H}{\cos(60^o - H)}\right] \quad \text{... (15)}$$

$$R = 3I - (R + B) \quad \text{... (16)}$$

Example 4: Consider the image with different colours as given in Fig. 6 Write the RGB colours which would appear on monochrome display. You may assume that all colours are at maximum intensity and saturation. Also show each of the colour in black and white considering them as 0 and 255 respectively.

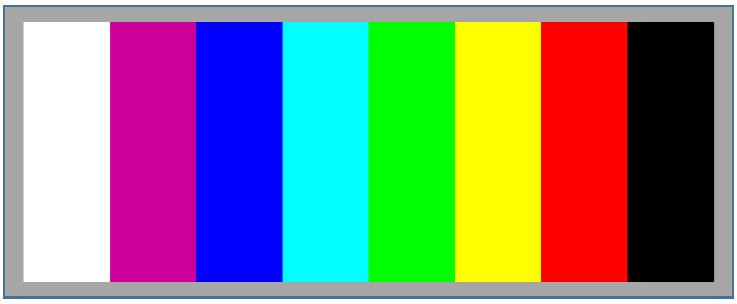


Fig. 6

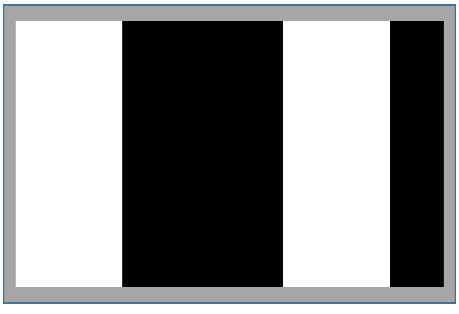
Solution: It is given here that the intensity and saturation are maximum, therefore the value of each of the components RGB would be 0 or 1.

Let us check each colour of the image one by one by starting from the left most colour.

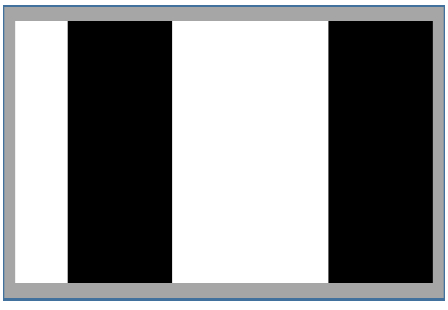
Colour	RGB combination	Intensity/Saturation			Monochrome colours		
		R	G	B	R	G	B
White	R+G+B	1	1	1	255	255	255
Magenta	R+B	1	0	1	255	0	255
Blue	B	0	0	1	0	0	255
Cyan	G+B	0	1	1	0	255	255
Green	G	0	1	0	0	255	0
Yellow	R+G	1	1	0	255	255	0
Red	R	1	0	0	255	0	0
Black	NIL	0	0	0	0	0	0

Now hence forth we shall follow the conversion that 0 represents black and 255 represents white. Also, the grey is represented by 128. You see that the table has R colour series as 255, 255, 0, 0, 255, 255, 0. Thus, it would show W, W, B, B, B, W, W, B in monochrome display, which is shown in Fig. 7 (a).

Similarly monochrome display of green colour would be shown by the series W,B,B,W,W,W,B,B and blue would be shown as W, W, W, W, B, B, B, B, B as shown in Fig. 7 (b) and Fig. 7 (c).



(a)



(b)

(c)

Fig. 7

***

Example 5: Let us sketch the HSI components of the image considered in Fig. 6 [Given in Example 4] on a monochrome display.

Solution: We transform HSI by computing values of H, S and I for each colour.

For white $R=1, G=1, B=1$

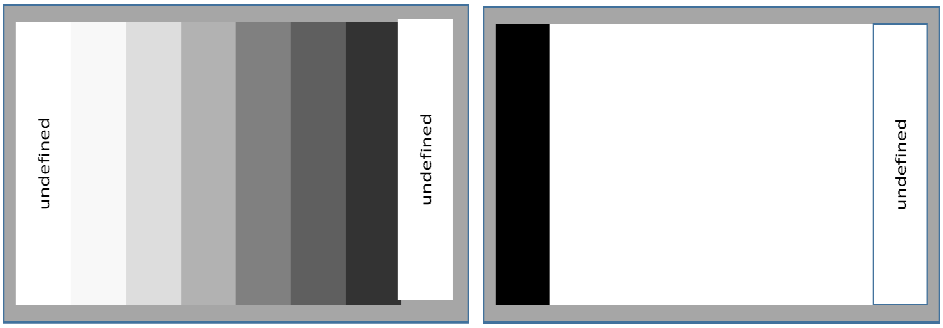
Using Eqn. (5), we can say that H does not exist as denominator is zero while computing θ.

Using Eqn. (7), we get $I=\frac{1}{3}(1+1+1)=1$ and using Eqn. (6), we get

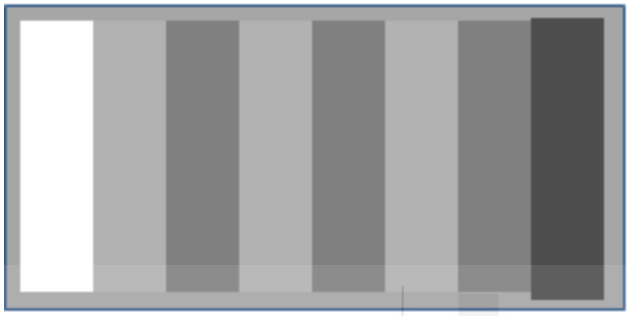
$$S=1-\frac{3}{1+1+1}[\min(1,1,1)]=0.$$

Similarly we can find H, S, I for each of the colours as shown in the following table.

Colour	R	G	B	H	S	I	Monochromatic		
							H	S	I
White	1	1	1	Cannot be computed	0	1	–	0	255
Magenta	1	0	1	$\frac{5}{6}$	1	$\frac{2}{3}$	213	255	170
Blue	0	0	1	$\frac{2}{3}$	1	$\frac{1}{3}$	170	255	85
Cyan	0	1	1	$\frac{1}{2}$	1	$\frac{2}{3}$	128	255	170
Green	0	1	0	$\frac{1}{3}$	1	$\frac{1}{3}$	85	255	85
Yellow	1	1	0	$\frac{1}{6}$	1	$\frac{2}{3}$	43	255	170
Red	1	0	0	0	1	$\frac{1}{3}$	0	255	85
Black	0	0	0	–	0	0	–	–	0



(a) (b)



(c)

Fig. 8

The output is given in Fig. 8 (a), Fig. 8 (b) and Fig. 8 (c) for each of the attribute H,S and I.

***

Now try the following exercises.

E8) Describe how the grey levels vary in RGB primary images that make up the font face of the colour cube.

E9) Transform the RGB cube by its CMY cube. Lable all the vertices. Also, interpret the colours at the edges with respect to saturation.

In the following section, we discuss pseudocolour image processing.

4.5 PSEUDOCOLOUR IMAGE PROCESSING

Pseudocolour image processing is the process of assigning colour to each pixel of a grayscale image based on specific conditions. As mentioned above, colour carries with it a large amount of information regarding the objects that we are viewing and therefore, for better visualisation, converting a grayscale image to a colour image helps in improved interpretation of the image.

Intensity slicing or density slicing is one of the simplest forms of pseudocolour image processing technique. In this technique, the image is interpreted as a 3D function and can be imagined as a set of 2D grid which are parallel to the coordinate planes and placed at each intensity value. Each plane can then be thought of as a slice of the image function in the area of intersection. For example, the plane at $f(x,y) = I_i$ slices the image function into two parts. Then, any pixel whose graylevel is on or above the plane can be coded in one colour and

whose graylevel is below the plane can be coded in another colour, thereby converting the grayscale image into a two colour image.

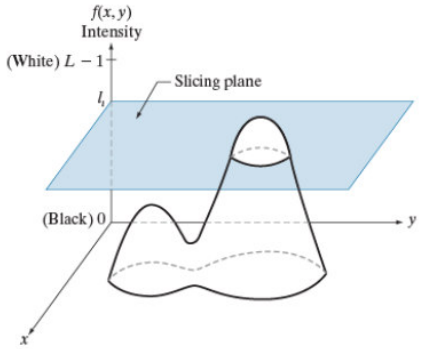


Fig. 9: Intensity slicing technique represented as slicing planes perpendicular to the intensity axis at various intensity levels.

In general, intensity slicing technique is given as follows:

Let $[0, L-1]$ be the existing grayscale values such that black is represented by l_0, that is, $f(x,y)=0$ and white is represented by l_{L-1}, that is, $f(x,y)=L-1$. Then, we define P planes such that they are perpendicular to the intensity axis at levels $l_1, l_2, ..., l_P$, where $0<P<L-1$. These P planes are parallel to the coordinate plane. The grayscale intensity levels are partitioned into $P+1$ intervals, namely, $V_1, V_2, ...V_{P+1}$. Then, assign a colour to each pixel in the following manner

$$f(x,y)=c_k \text{ if } f(x,y)\in V_k$$

where, V_k is the k^{th} intensity interval defined by planes at levels $lk-1$ and lk and c_k is the colour associated with V_k.

You may try an exercise:

E10) Define an application of intensity level slicing.

Now let us, summarise what we have discussed in this unit.

4.6 SUMMARY

In this unit, we discussed the following points:

1. The need for colour image processing. Since the human eye has the wonderful capability of seeing millions of colour, we realise that colour gives a large amount of information about the objects and scene in the images.
2. We first discussed the structure of the human eye and then the tristimulus theory that connects the perception of colour with the various colour models that exist.
3. We then discussed the main colour models or colour spaces that are mainly used in both TV and print.

4.7 SOLUTIONS/ANSWERS

E1) Since when object is far, the focal length is 17 mm for the human eye, therefore, $15/100=h/17 \Rightarrow h(17*15)/100=2.55m$

E2) THSI problem is the extention of the problem solved in Example 2. Here, we consider two possibilities.

i) When the point C at which percentage of colours C_1, C_2 and C_3 to be found is on the sides of triangle. In this case the percentage is found by considering the point on the line joining the corresponding vertices as we solved in Example 2. There would be 0% from the vertex which does not lie on the line. For example, if the point lies on the line joining C_1 and C_2, then the percentage of C_1 and C_2 can be found as given below.

Let the coordinates of C_1 be $(x_1, y_1), C_2$ be $(x_2, y_2), C_3$ be (x_3, y_3) and C be (x, y).

Percentage of C_1 in

$$C = \frac{\sqrt{(x-x_1)^2+(y-y_1)^2} - \sqrt{(x_2-x)^2+(y_2-y)^2}}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}} \times 100$$

Percentage of C_2 in $C = (100 - \text{percentage of } C_1)\%$

Percentage of C_3 in $C = 0\%$.

ii) When the point C does not lie on the sides of the triangle with vertices C_1, C_2 and C_3.

In HSI case we join the point C with any of the vertices say C_3. We follow the following steps.

iii) Join the points C and C_3 and extend the line towards the side C_1C_2. Suppose it intersects C_1C_2 at C_4.

iv) Find the percentage of C_1 and C_2 at C_4.

v) Use the concept that the ratio of C_1 and C_2 will remain same at each of the points on the line C_3C_4.

vi) Now, we can easily find the coordinates of the point C_4 by writing equation of the lines C_1C_2 and C_3C. C_4 is the point of intersection of C_1C_2 and C_3C.

vii) Finally, we can find the percentage of C_4 and C_3 for the colour C.

E3) For an 8-bit image, there are $2^8 = 256$ possible values. A colour will be grey if each of the colour in RGB is same. Therefore, there can be 256 shades of grey.

E4) You can see in Fig. 5, yellow paint is made by combining green and red while imperfections in blue leads to reflection of some amount of green from blue paint also. Therefore, when both blue and yellow are mixed, both reflect the green colour, while all other colours are absorbed. Therefore, green coloured paint results from mixing of blue and yellow paints.

E5) H stands for Hue, which represents dominant colour as observed by an observer and the corresponding wavelength is also dominant. S stands for Saturation, which is the amount of white

light mixed with a hue. I stands for intensity which reflects the brightness.

E7) A colour space allows one to represent all the colour perceived by human eye. The colour space can be broadly classified into (i) RGB, (ii) GMY and (iii) HSI colour space.

E8) Each of the components in RGB model would vary from 0 to 255. Here, we are discussing the front face. So, we fix all pixel values in the Red image as 255 and let the columns to vary from 0 to 255 in the green image and rows to vary from 255 to 0 in the blue image.

E9) The vertices would be as given below:
White $= (0,0,0)$
Cyan $= (1,0,0)$
Magenta $= (0,1,0)$
Blue $= (1,1,0)$
Green $= (1,0,1)$
Red $= (0,1,1)$
Black $= (1,1,1)$.
The edges which are free from black or white pixels are fully saturated. The saturation decreases towards the ends having black or white pixel.

E10) Detecting blockages in medical image processing is an application of intensity level slicing.

References

[1] R.C. Gonzoles and R.E. Woods, Digital Image Processing, Addision-wesley, 1992.

[2] A.K. Jain, Fundamentals of Digital Image Processing, PHI.