Understanding Digital Image Processing

VIPIN TYAGI

CRC Press
Taylor & Francis Group
A SCIENCE PUBLISHERS BOOK

CPD CERTIFIED
The CPD Certification Service

Understanding Digital Image Processing

Vipin Tyagi

Department of Computer Science and Engineering
Jaypee University of Engineering and Technology
Raghogarh, Guna (MP), India

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A SCIENCE PUBLISHERS BOOK

Cover photo credits: Prahlad Bubbar collection

MATLAB® and Simulink® are trademarks of The MathWorks, Inc. and are used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® and Simulink® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® and Simulink® software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20180730

International Standard Book Number-13: 978-1-138-56684-2 (Hardback)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface

Digital Images are an integral part of our digital life. Digital images are used in every aspect of or daily life. Digital image processing concerns with techniques to perform processing on an image, to get an enhanced image or to extract some useful information from it to make some decisions based on it. Digital image processing techniques are growing at a very fast speed. This book,*Understanding Digital Image Processing* aims on providing digital image processing concepts in a simple language with the help of examples. The major objective of this text book is to introduce the subject to the readers and motivate for further research in the area.

To explain the concepts, MATLAB® functions are used throughout the book. MATLAB® Version 9.3 (R2017b), Image Acquisition Toolbox Version 5.3 (R2017b), Image Processing Toolbox, Version 10.1 (R2017b) are used to create the book material. My thanks to MathWorks for providing support in preparation of this book.

The functions provided by Image Processing Toolbox™ are given in Appendix A. The details of these functions are available at on the website https://in.mathworks.com/help/images/index.html

For MATLAB® product information, please contact:

The MathWorks, Inc.

3 Apple Hill Drive

Natick, MA, 01760-2098 USA

Tel: 508-647-7000

Fax: 508-647-7001

E-mail: info@mathworks.com

Web: mathworks.com

How to buy: www.mathworks.com/store

'C' language is a very popular programming language. Various functions written in 'C' language are given in Appendix B. My sincere thanks to Mr. Dwayne Phillips for permitting to use the code written by him. Related sub-functions and their descriptions are available at http://homepages.inf.ed.ac.uk/rbf/BOOKS/PHILLIPS/.

A glossary of common image processing terms is provided in Appendix C.

A bibliography of the work in the area of image processing is also given at end of the book. We are thankful to all authors whose work has been used in preparation of the manu-script. We have tried to include all such work in bibliography, but if we have skipped some work by error, we will include in next version of the book, Color version of some figures have been provided at the end of the book. These figures also have corresponding black & white versions in the text.

The target audience spans the range from the undergraduate with less exposure to subject to research students interested in learning digital image processing. Many texts are available in the area of digital image processing. In this book, objective is to explain the concepts in a very simple and understandable manner. Hope this book will succeed in its aim.

This work would not have been possible without the help and mentoring from many,in particular, my teacher Prof. Vinod K. Agarwal, Meerut. Special thanks to my dear scholars Mr. K. B. Meena, Dr. Ashwani Kumat and Dr. Deepshikha Tiwari, for their help and support in preparation of the manuscript.

The research work of several researchers contributed to a substantial part of some sections of the book. I thankfully acknowledge their contributions.

It has been a pleasure working with Taylor and Francis Publishers in the development of the book. Thanks to Mr. Vijay Primlani for his kind and timely support in publishing the book and for handling the publication.

Contents

Preface iii

1. Introduction to Digital Image Processing 1
2. Digital Image Representation 13
3. Mathematical Tools for Image Processing 23
4. Image Enhancement in Spatial Domain 36
5. Image Processing in Frequency Domain 57
6. Image Denoising 76
7. Image Segmentation 85
8. Mathematical Morphology 103
9. Image Understanding 118
10. Image Compression 127
11. Image Retrieval 147
12. Digital Image Forgery 155

Appendix A 164

Appendix B 190

Appendix C 333

Bibliography 349

Index 361

Color Plate Section 365

1

Introduction to Digital Image Processing

1.1 Introduction

In today's digital life, digital images are everywhere around us. An image is a visual representation of an object, a person, or a scene. A digital image is a two-dimensional function f(x, y) that is a projection of a 3-dimesional scene into a 2-dimensional projection plane, where x, y represents the location of the picture element or pixel and contains the intensity value. When values of x, y and intensity are discrete, then the image is said to be a digital image. Mathematically, a digital image is a matrix representation of a two-dimensional image using a finite number of points cell elements, usually referred to as pixels (picture elements, or pels). Each pixel is represented by numerical values: for grayscale images, a single value representing the intensity of the pixel (usually in a [0, 255] range) is enough; for color images, three values (representing the amount of red (R), green (G), and blue (B)) are stored. If an image has only two intensities, then the image is known as a binary image. Figure 1.1 shows a color image and its red, green and blue components. The color image is a combination of these three images. Figure 1.1e shows the 8-bit grayscale image corresponding to color image shown in Fig. 1.1a. Figure 1.1 also shows the matrix representation of a small part of these images.

MATLAB® supports the following image types:

1. **Grayscale:** A grayscale image, having M×N pixels is represented as a matrix of double datatype of M×N size in MATLAB. Element values denote the pixel grayscale intensities in the range [0,1] with 1 = white and 0 = black.
2. **True-color RGB:** A true-color red-green-blue (RGB) image is represented as a three-dimensional M×N×3 double matrix in MATLAB. Each pixel has

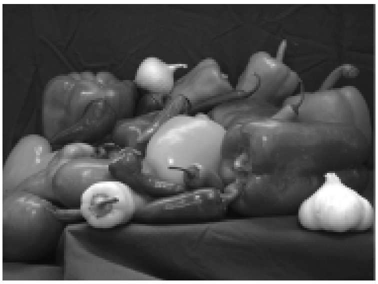
(a)

(b)

(c)

(d)

(e)



(f)

103	105	102	99	97	97	96	95
139	144	144	143	142	133	125	111
174	181	184	189	193	193	187	179
174	179	185	187	189	193	194	192
183	185	184	187	187	187	188	196
184	182	180	184	183	185	186	191
184	182	182	182	181	181	180	184
194	184	182	180	178	177	175	179
199	189	184	178	175	173	172	171
190	186	185	178	174	169	168	167

(g)

76	82	82	83	83	85	85	88
83	83	82	80	81	82	83	84
80	80	78	73	75	80	80	75
78	80	76	76	78	78	75	73
74	79	74	74	77	79	80	80
73	78	77	74	75	78	78	75
76	72	75	75	73	73	75	75
73	74	74	70	71	74	76	78
75	76	78	79	79	75	76	80
78	78	80	83	82	80	78	78
81	78	82	81	81	84	82	83

(h)

67	71	72	72	73	74	70	66
70	72	73	73	71	73	72	68
72	72	71	71	71	68	70	70
74	73	66	66	69	68	71	71
70	67	66	69	71	70	70	68
69	64	69	71	68	68	72	70
67	67	71	70	63	66	70	72
71	69	68	68	66	69	68	71
68	69	67	67	73	69	62	70
63	65	62	61	68	68	61	66

(i)

51	54	53	52	53	54	53	51
52	53	55	54	52	54	54	51
52	53	53	52	51	52	53	53
55	53	51	52	53	52	53	53
53	50	49	50	52	51	52	50
52	48	49	49	49	51	52	51
51	50	52	51	48	51	52	53
52	52	52	52	52	53	52	54
50	53	52	51	52	53	50	52
48	48	49	48	51	51	48	49

Fig. 1.1 contd. ...

red, green and blue components. The color components of a pixel (m,n) are denoted as (m,n,1) = red, (m,n,2) = green, (m,n,3) = blue.

3. **Indexed:** In MATLAB, Indexed (paletted) images are represented with an index matrix of size M×N and a colormap matrix of size K×3. The colormap matrix holds all colors used in the image and the index matrix represents the pixels by referring to colors in the colormap.
4. **Binary:** In MATLAB, a binary image having two values, 1 (white) or 0 (black), is represented by an M×N logical matrix.
5. **uint8:** In MATLAB, this type uses less memory and some operations compute faster than with double types.

In image processing operations, most of the operations are performed in grayscale images. For color image processing applications, a color image can be decomposed into Red, Green and Blue components and each component is processed independently as a grayscale image. For processing, an indexed image is converted to grayscale image or RGB color image for most operations.

MATLAB commands imread and imwrite are used for reading and writing image files as:

I = imread(filename);
imwrite(I, filename) writes image data I to the file specified by filename, inferring the file format from the extension. The bit depth of the output image depends on the data type of I and the file format. For most formats:

- If I is of data type uint8, then imwrite outputs 8-bit values.
- If I is of data type uint16 and the output file format supports 16-bit data (JPEG, PNG and TIFF), then imwrite outputs 16-bit values. If the output file format does not support 16-bit data, then imwrite returns an error.
- If I is a grayscale or RGB color image of double or single data type, then imwrite assumes that the dynamic range is [0,1] and automatically scales the data by 255 before writing it to the file as 8-bit values. If the data in I is single, convert I to double before writing to a GIF or TIFF file.
- If I is of logical data type, then imwrite assumes that the data is a binary image and writes it to the file with a bit depth of 1, if the format allows it. BMP, PNG, or TIFF formats accept binary images as input arrays.

...Fig. 1.1 contd.

Fig. 1.1. (a) A color image; (b) Red Component of color image (a); (c) Green Component of color image (a); (d) Blue Component of color image (a); e) Color image (a) converted into 8-bit grayscale image; (f) Matrix representation of upper left corner of image (b); (g) Matrix representation of upper left corner of image (c); (h) Matrix representation of upper left corner of image (d); (i) Matrix representation of upper left corner of image (e).

Original image reprinted with permission of The MathWorks, Inc.

(For color images of Fig. 1.1(a), (b), (c), (d) see Color Figures Section at the end of the book)

In literature, the following three levels of image processing operations are defined:

- Low-Level image processing: Primitive operations on images (e.g., contrast enhancement, noise reduction, etc.) are under this category, where both the input and the output are images.
- Mid-Level image processing: In this category, operations involving extraction of attributes (e.g., edges, contours, regions, etc.), from images are included.
- High-Level image processing: This category involves complex image processing operations related to analysis and interpretation of the contents of a scene for some decision making.

Image processing involves many disciplines, mainly computer science, mathematics, psychology and physics. Other areas, such as artificial intelligence, pattern recognition, machine learning, and human vision, are also involved in image processing.

1.2 Typical Image Processing Operations

Image processing involves a number of techniques and algorithms. The most representative image processing operations are:

- **Binarization:** Many image processing tasks can be performed by converting a color image or a grayscale image into binary in order to simplify and speed up processing. Conversion of a color or grayscale image to a binary image having only two levels of gray (black and white) is known as binarization.
- **Smoothing:** A technique that is used to blur or smoothen the details of objects in an image.
- **Sharpening:** Image processing techniques, by which the edges and fine details of objects in an image are enhanced for human viewing, are known as sharpening techniques.
- **Noise Removal and De-blurring:** Before processing, the amount of noise in images is reduced using noise removal filters. Image removal technique can sometimes be used, depending on the type of noise or blur in the image.
- **Edge Extraction:** To find various objects before analyzing image contents, edge extraction is performed.
- **Segmentation:** The process of dividing an image into various parts is known as segmentation. For object recognition and classification segmentation is a pre-processing step.

1.3 History of Digital Image Processing

Earlier digital image processing was mainly used in the newspaper industry to make images look better or for simply converting black and white images into color images. In the 1920s, digital images were transmitted electronically between London and New York. Initial Bartlane cable picture systems were capable of coding an image using five gray levels; this was later enhanced to 15 gray levels in 1929. Actual digital image processing started after the invention of digital computers and related technologies, including image storage, display and transmission. In the 1960s, powerful computers gave birth to meaningful digital image processing. Satellite Images of the moon, taken by Ranger 7 U.S. spacecraft, were processed at the Jet Propulsion laboratory at California. At the same time, use of digital image processing began in various activities relating to astronomy, medical image processing, remote sensing, etc. From 1960 onwards, the use of digital image processing techniques has grown tremendously. These techniques are now used in almost every part of our life. They have applications in the fields of defence, astronomy, medicine, law, etc.

Some common examples of digital image processing are fingerprint recognition, processing of satellite images, weather prediction, character recognition, face recognition, product inspection and assembly.

1.4 Human Visual System

The human visual system consists of two parts: eye and brain. The human eye acts as a receptor of images by capturing light and converting it into signals. These signals are then transmitted to the brain for further analysis. Eyes and brain work in combination to form a picture.

The human eye is analogous to a camera. The structure of human eye is shown in Fig. 1.2:

Various parts of the human eye are identified:

- Primary Lens: Cornea and aqueous humour, used to focus incoming light signal.
- Iris: The iris dynamically controls the amount of light entering the eye, so that the eye can function in various lighting conditions, from dim light to very bright light. The portion of the lens not covered by the iris is called the pupil.
- Zonula: This is a muscle that controls the shape and positioning (forward and backward) of the eye's lens.

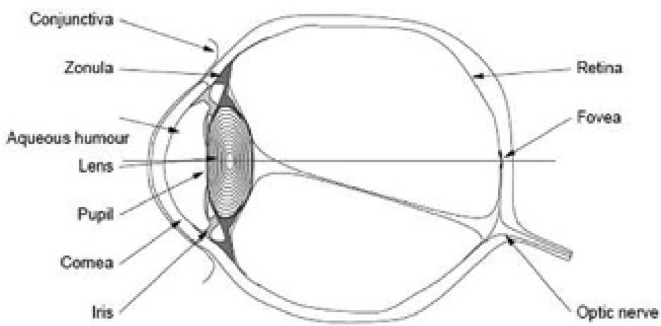


Fig. 1.2. Human eye.

- Retina: provides a photo-sensitive screen at the back of the eye; it converts the light hitting the retina into nerve signals.
- Fovea: A small central region of the retina that contains a large number of photo-sensitive cells and provides very good resolution.
- Optic nerve: These nerves carry the signals generated by the retina to the brain.

Human vision has a “blind spot” in the area of the retina where the optic nerves are attached. This blind spot does not have any photosensitive cells.

Light sensitive cells in the brain are of two types: rods and cones. There are about 120 million rod cells and 6 million cone cells. Rod cells provide visual capability at very low light levels and are very sensitive. Cone cells provide our daytime visual facilities and perform best at normal light levels.

Rods and cones are not uniformly distributed across the retina. The cones are concentrated in the center, while the rods are away from the center (Fig. 1.3). A yellow spot (macula), of size 2.5 to 3 mm in diameter, is found in the middle of the retina. Cones are very tightly packed together and the blood vessels within the fovea and other cells are pulled aside in order to expose them directly to the light.

In dim light, such as moonlight, rods in our eyes are activated and the fovea effectively acts as a second blindspot. To see small objects at night, one must shift the vision slightly to one side, around 4 to 12 degrees, so that the light falls on some rods.

Although there are about 120 million rods and 6 million cone cells in the retina, there are less than a million optic nerve fibres which connect them to the brain. This means that there cannot be a single one-to-one connection between the photoreceptors and the nerve fibres. The number of receptors connecting to each fibre is location dependent.

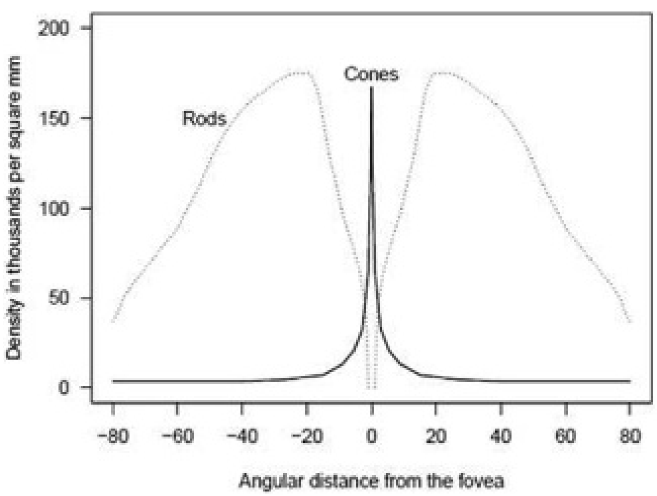


Fig. 1.3. Rods and Cones cells distribution in the retina.

In the fovea, cones have a one-to-one interaction, while in the outer part of the retina, about 600 rods are connected to each nerve fibre.

1.5 Classification of Digital Images

With regard to the manner in which they are stored, digital images can be classified into two categories: (1) Raster or Bitmap image, (2) Vector image.

A bitmap or raster image is a rectangular array of sampled values or pixels. These images have a fixed number of pixels. In the zooming of a raster image, mathematical interpolation is applied. The quality of a zoomed image degrades after a particular value of zooming factor, as shown in Fig. 1.4.

The resolution of a bitmap image is determined by the sensing device. BMP, GIF, PNG, TIFF and JPEG common image formats are bitmap or raster image formats.

On the other hand, vector images are stored in the form of mathematical lines and curves. Information like length, color, thickness, etc., is stored in the form of a vector. These images can be displayed in any size, any resolution and on any output

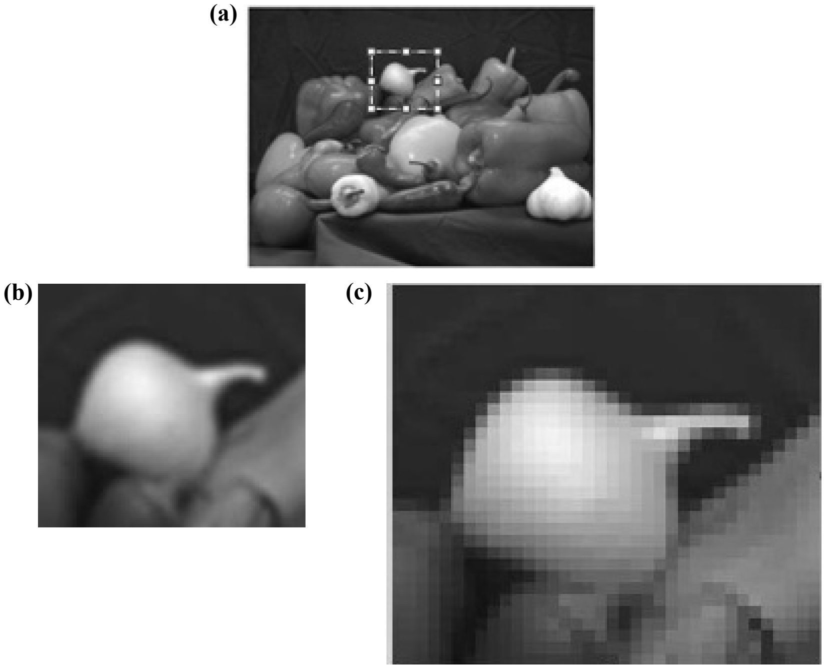


Fig. 1.4. Zooming of a Bitmap image (a) Original image (b) Part of image zoomed 4 times (c) Same part of the image zoomed 8 times.
Original image reprinted with permission of The MathWorks, Inc.
(For color image of Fig. 1.4(a), see Color Figures Section at the end of the book)

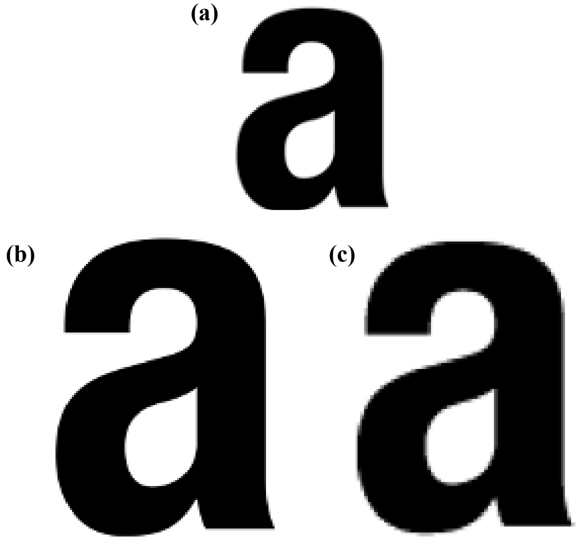


Fig. 1.5. (a) An image (b) Image in (a) zoomed in vector format (c) Image in (a) zoomed in Raster format.

device. Vector images are suitable in illustrations, line art, font, etc. The difference in zooming of a vector image and raster image can be observed in Fig. 1.5. The degradation in the quality due to zooming is clearly visible on the boundaries of the character stored in raster format.

1.6 Digital Image File Types

There are a number of digital image file types available these days. The most commonly used image file types are: JPEG, GIF, TIFF, PNG and BMP. Image file types are based on the compression technique used for reducing the size of the image file. Images in various file types may differ in color, if color has been used. An image in its simplest form may contain only two intensities, i.e., black and white, and needs only 1 bit to represent intensity at each pixel.

- **TIFF** (Tagged Image File Format): This format is a very flexible and may be based on a lossy or lossless compression technique. The details of the compression technique are stored in the image itself. Generally, TIFF files use a lossless image storage format and, hence, are quite large in size.
- **Portable Network Graphics** (PNG): This format is a lossless storage format and uses patterns in the image to compress the image. The compression in PNG files is exactly reversible, meaning the uncompressed image is identical to the original image.
- **Graphical Interchange Format** (GIF): This format creates a table of upto 256 colors from 16 million colors. If the image to be compressed has less

than 256 colors, then the GIF image has exactly the same color. But if the number of colors is greater than 256, then the GIF approximates the colors in the image using a table of the 256 colors that are available.

- **Joint Picture Experts Group** (JPG or JPEG): This format is an optimized format for photographs and continuous tone images that contain a large number of colors. JPEG files can achieve high compression ratios while maintaining clear image quality.
- **RAW:** This is a lossless image format available on some digital cameras. These files are very small but the format is manufacturer dependent, therefore, the manufacturer's software is required in order to view the images.
- **Bitmapped Image** (BMP) is an uncompressed proprietary format invented by Microsoft.

Some other common image file formats are PSD, PSP, etc., which are proprietary formats used by graphics programs. Photoshop's files have the PSD extension, while Paintshop Pro files use the PSP extension.

MATLAB supports a number of image file formats, e.g., Windows Bitmap, Windows Cursor resources, Flexible Image Transport, Graphics Interchange Format, Windows Icon resources, JPEG, Portable Bitmap, Windows Paintbrush, Tagged Image File Format.

The details can be obtained using MATLAB function imformats. The imformats function can be used if the file format exists in the registry as:

formatStruct = imformats ('bmp'), then the output will look like:
format Struct = struct with fields:
  ext: {'bmp'}
  isa: @isbmp
  info: @imbmpinfo
  read: @readbmp
  write: @writebmp
  alpha: 0
  description: 'Windows Bitmap'

1.7 Components of an Image Processing System

A complete digital image processing system comprises of various elements, such as image acquisition, storage, image processing, displays, etc. (Fig. 1.6).

Sensing devices are used to capture the image. The sensing device senses the energy radiated by the object and converts it into digital form. For example, a digital camera senses the light intensity and converts into the digital image form. Image processing elements are used to perform various operations on a digital image. It requires a combination of hardware and software. Storage is a very important part of an image processing system. The size of an image or video file is very large. For instance, an 8-bit image having 1024 x 1024 pixels requires 1 megabyte of storage space. Therefore, mass storage devices are required in image processing systems.

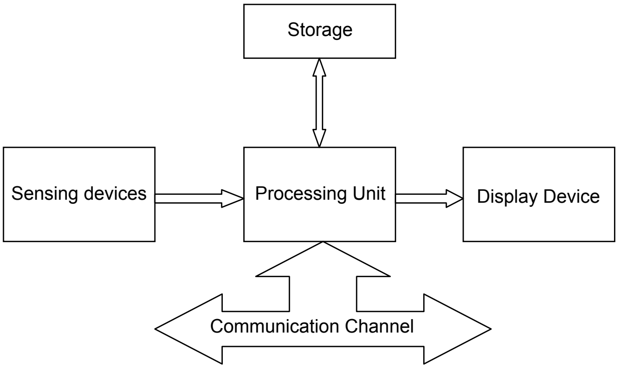


Fig. 1.6. A digital image processing system.

Display devices are required to display the images. These can be a computer monitor, mobile screen, projector for display or hardcopy devices, such as printers. A communication channel is also essential for sending and receiving images.

1.8 Applications of Digital Image Processing

Digital image processing techniques are now used in a number of applications; some common applications are given below.

In medicine: Several medical tools use image processing for various purposes, such as image enhancement, image compression, object recognition, etc. X-radiation (X-rays), computed tomography scan (CT scan), positron-emission tomography (PET), Single-photon emission computed tomography (SPECT), nuclear magnetic resonance (NMR) spectroscopy and Ultra-Sonography are some popular pieces of medical equipment based on image processing.

In agriculture: Image processing plays a vital role in the field of agriculture. Various paramount tasks such as weed detection, food grading, harvest control and fruit picking are done automatically with the help of image processing. Irrigated land mapping, determination of vegetation indices, canopy measurement, etc., are possible with good accuracy through the use of imaging techniques in various spectrums, such as hyper spectral imaging, infrared, etc.

In weather forecasting: Image processing also plays a crucial role in weather forecasting, such as prediction of rainfall, hailstorms, flooding. Meteorological radars are widely used to detect rainfall clouds and, based on this information, systems predict immediate rainfall intensity.

In photography and film: Retouched and spliced photos are extensively used in newspapers and magazines for the purpose of picture quality enhancement. In movies, many complex scenes are created with image and video editing tools which are based on image and video processing operations. Image processing-based methods are used to predict the success of upcoming movies. For a global media and entertainment company, Latent View extracted over 6000 movie posters from IMDB along with their metadata (genre, cast, production, ratings, etc.), in order to predict the movies' success using image analytics. The colour schemes and objects in the movie posters were analyzed using Machine Learning (ML) algorithms and image processing techniques.

In entertainment and social media: Face detection and recognition are widely used in social networking sites where, as soon as a user uploads the photograph, the system automatically identifies and gives suggestion to tag the person by name.

In security: Biometric verification systems provide a high level of authenticity and confidentiality. Biometric verification techniques are used for recognition of humans based on their behaviours or characteristics. To create alerts for particularly undesirable behaviour, video surveillance systems are being employed in order to analyze peoples' movements and activities. Several banks and other departments are using these image processing-based video surveillance systems in order to detect undesired activities.

In banking and finance: The use of image processing-based techniques is rapidly increasing in the field of financial services and banking. 'Remote deposit capture' is a banking facility that allows customers to deposit checks electronically using mobile devices or scanners. The data from the check image is extracted and used in place of a physical check. Face detection is also being used in the bank customer authentication process. Some banks use 'facial-biometric' to protect sensitive information. Signature verification and recognition also plays a significant role in authenticating the signature of the customers. However, a robust system used to verify handwritten signatures is still in need of development. This process has many challenges because handwritten signatures are imprecise in nature, as their corners are not always sharp, lines are not perfectly straight, and curves are not necessarily smooth.

In marketing and advertisement: Some companies are using image-sharing through social media in order to track the influence of the company's latest products/advertisement. The tourist department uses images to advertise tourist destinations.

In defence: Image processing, along with artificial intelligence, is contributing to defence based on two fundamental needs of the military: one is autonomous operation and the other is the use of outputs from a diverse range of sophisticated sensors for predicting danger/threats. In the Iran-Iraq war, remote sensing technologies were employed for the reconnaissance of enemy territory. Satellite images are analyzed in order to detect, locate and destroy weapons and defence systems used by enemy forces.

In industrial automation: An unprecedented use of image processing has been seen in industrial automation. The 'Automation of assembly lines' system detects the position and orientation of the components. Bolting robots are used to detect the moving bolts. Automation of inspection of surface imperfection is possible due to image processing. The main objectives are to determine object quality and detect any abnormality in the products. Many industries also use classification of products by shape automation.

In forensics: Tampered documents are widely used in criminal and civil cases, such as contested wills, financial paper work and professional business documentation. Documents like passports and driving licenses are frequently tampered with in order to be used illegally as identification proof. Forensic departments have to identify the authenticity of such suspicious documents. Identifying document forgery becomes increasingly challenging due to the availability of advanced document-editing tools. The forger uses the latest technology to perfect his art. Computer scan documents are copied from one document to another to make them genuine. Forgery is not only confined to documents, it is also gaining popularity in images. Imagery has a remarkable role in various areas, such as forensic investigation, criminal investigation, surveillance systems, intelligence systems, sports, legal services, medical imaging, insurance claims and journalism. Almost a decade ago, Iran was accused of doctoring an image from its missile tests; the image was released on the official website, Iran's Revolutionary Guard, which claimed that four missiles were heading skyward simultaneously. Almost all the newspaper and news magazine published this photo including, The Los Angeles Times, The Chicago Tribune and BBC News. Later on, it was revealed that only three missiles were launched successfully, one missile failed. The image was doctored in order to exaggerate Iran's military capabilities.

In underwater image restoration and enhancement: Underwater images are often not clear. These images have various problems, such as noise, low contrast, blurring, non-uniform lighting, etc. In order to restore visual clarity, image enhancement techniques are utilized.

Summary

- Digital images are very important in today's digital life.
- An image may be a grayscale image or a color image. A grayscale image, having only two intensities, is termed as a binary image.
- Digital images can be classified as Raster image or Vector image, with regard to the manner in which the image is stored.
- Some common image file types which are used in our daily life are JPEG, GIF, TIFF, PNG and BMP.
- The main components of an image processing system are (1) Sensing device, (2) Image processing elements, (3) Storage device and (4) Display device.

2

Digital Image Representation

2.1 Digital Image

A digital image is made up of $M \times N$ pixels, and each of these pixels is represented by k bits. A pixel represented by k bits can have 2^k different shades of gray in a grayscale image. These pixel values are generally integers, having values varying from 0 (black pixel) to $2^k–1$ (white pixel). The number of pixels in an image defines resolution and determines image quality.

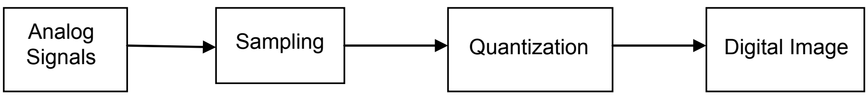


Fig. 2.1. Digital image formation process.

2.2 Sampling and Quantization

When an image is sensed and converted into digital form, the image is discretized in two ways: sampling and quantization. The concepts of sampling and quantization can be understood by following example in Fig. 2.2, in which a one-dimensional signal is quantized on a uniform grid. The signal is sampled at ten positions (x = 0,..., 9), and each sampled value is then quantized to one of seven levels (y = 0,..., 6). It can be observed that the samples collectively describe the gross shape of the original signal but that smaller variations and structures may not be represented, i.e., information may have been lost mathematically. The digital image is an approximation of a real image and the quality depends on the number of samples taken and the quantization.

Sampling is the process of converting coordinate values in digital form and Quantization is the process of digitizing amplitude values. The quality of an image and display device is measured in terms of resolution that depends on the sampling

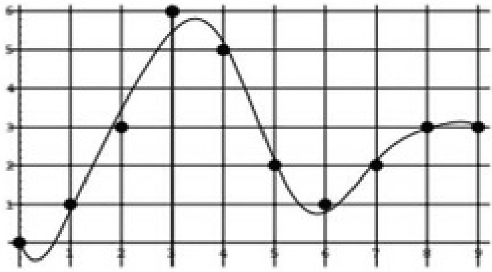
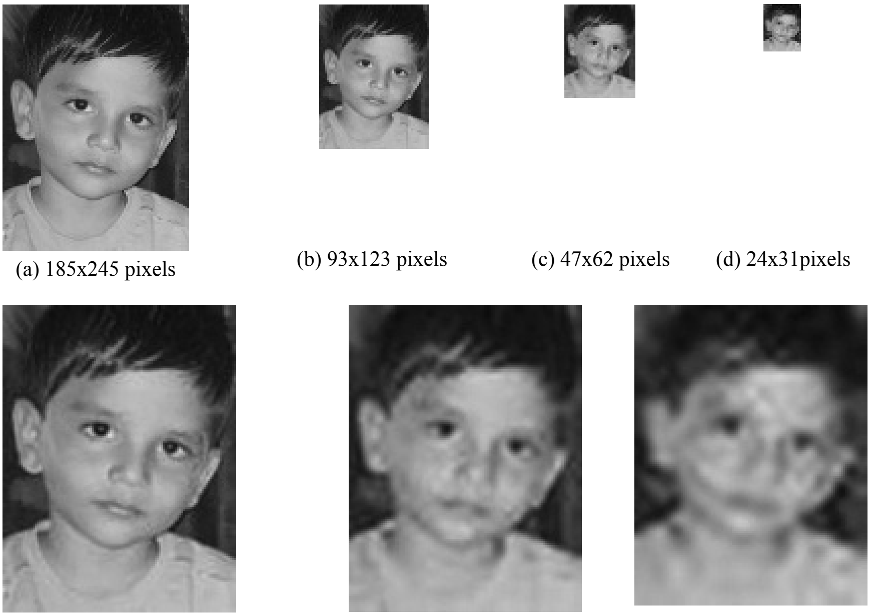


Fig. 2.2. A 1-dimensional signal quantized on a uniform grid.

and quantization process. The number of samples taken in digitization is known as image resolution and the quantization decides the accuracy of pixel intensities. The number of intensity values or pixels in an image is known as spatial resolution (Fig. 2.3) and the number of various graylevel values in an image is known as graylevel resolution (Fig. 2.4). These two factors also determine the space required to store an image, as can be observed from Table 2.1. If in an image, there is not a sufficient number of graylevels, then the smooth parts in the image exhibit 'false contouring'. This effect is clearly visible in Fig. 2.4.



(a) 185x245 pixels (b) 93x123 pixels (c) 47x62 pixels (d) 24x31pixels

(e) Resized from image (b) (f) Resized from image (c) (g) Resized from image (d)

Fig. 2.3. The effect of spatial resolution on the image quality.

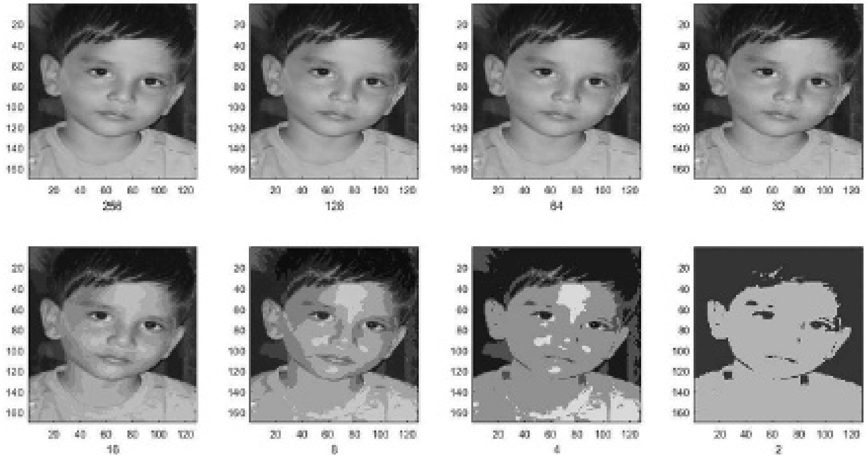


Fig. 2.4. The effect of gray level resolution on the image quality.

Table 2.1. Storage Requirement (in bytes) for an image with different parameter values.

	Bits per pixel				
Resolution	**1**	**2**	**4**	**6**	**8**
32 × 32	128	256	512	768	1,024
64 × 64	512	1,024	2,048	3,072	4,096
128 × 128	2,048	4,096	8,192	12,288	16,384
256 × 256	8,192	16,384	32,768	49,152	65,536
512 × 512	32,768	65,536	131,072	196,608	262,144
1024 × 1024	131,072	262,144	524,288	786,432	1,048,576

2.3 Color Models

Visible light is composed of various frequencies, approximately between 400 and 700 nm in the electromagnetic energy spectrum. Newton discovered that if a white sunlight beam is passed through a glass prism, the beam is divided into a continuous spectrum of colors ranging from violet to red (Fig. 2.5). The colors in this color spectrum are VIBGYOR (Violet, Indigo, Blue, Green, Yellow, Orange, Red). In this spectrum, the red color has the longest wavelength, and the violet color has the shortest wavelength. Wavelengths of various colors are given in Table 2.2.

The colors are observed by the human eye according to the wavelength of the light reflected by the object. For example, a yellow object reflects light having wavelength 577–597 nm, and absorbs most of the energy at other wavelengths. Similarly, an object appears white if it reflects relatively balanced light that is in all visible wavelengths.

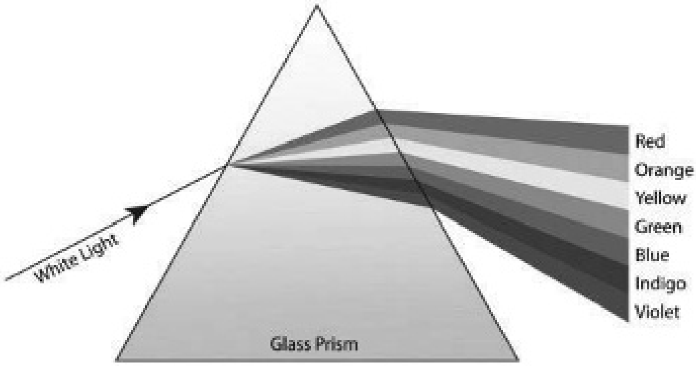


Fig. 2.5. Color spectrum after passing through a Prism.
(For color image of this Figure see Color Figures Section at the end of the book)

Table 2.2. Wavelength of various colors.

Color	Wavelength (nm)
Violet	380–435
Indigo	~ 446
Blue	435–500
Green	500–565
Yellow	565–590
Orange	590–625
Red	625–740

According to color theory, color formation is based on two processes: (i) Additive color formation and (ii) Subtractive color formation.

2.3.1 RGB Color Model

Additive colors are formed by combining spectral light in varying combinations. For example, computer screen produces color pixels by generating red, green, and blue electron guns at phosphors on the monitor screen. Additive color formation begins with black and ends with white and, as more color is added, the result becomes lighter and tends towards white (Fig. 2.6). Primary additive colors are Red, Green and Blue (R, G, B). The CIE (Commission Internationale de l'Eclairage - International Commission on Illumination) has standardized the wavelength values to the primary colors, which are as follows:

Red (R) = 700.0 nm
Green (G) = 546.1 nm
Blue (B) = 435.8 nm

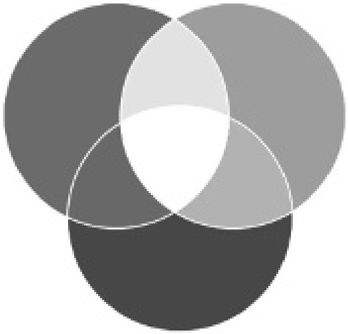


Fig. 2.6. Additive color system.
(For color image of this Figure see Color Figures Section at the end of the book)

In the RGB color model, the primary colors are represented as:

Red = (1,0,0), Green = (0,1,0), Blue = (0,0,1).
and the secondary colors of RGB are represented as
Cyan = (0,1,1), Magenta = (1,0,1), Yellow = (1,1,0).

The RGB model can be represented in the form of a color cube (Fig. 2.7). In this cube, black is at the origin (0,0,0), and white is at the corner, where $R = G = B = 1$. The grayscale values from black to white are represented along the line joining these two points. Thus a grayscale value can be represented as (x,x,x) starting from black = (0,0,0) to white = (1,1,1).

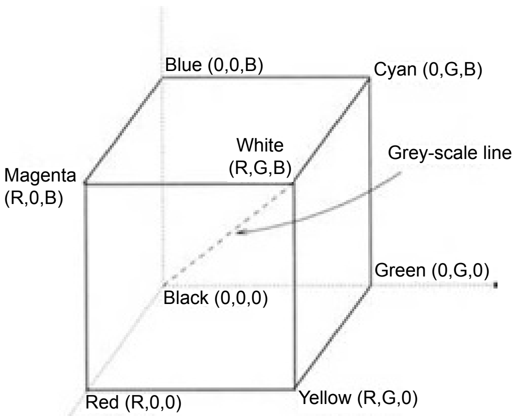


Fig. 2.7. RGB color cube.

2.3.2 CMY Color Model

In subtractive color formation, the color is generated by light reflected from its surroundings and the color does not emit any light of its own. In the subtractive colour system, black color is produced by a combination of all the colors (Fig. 2.8). For example, in printing, black color is generated by mixing all colors,

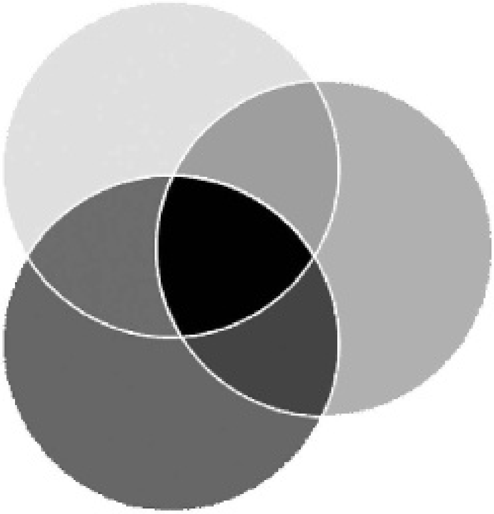


Fig. 2.8. Subtractive color system.
(For color image of this Figure see Color Figures Section at the end of the book)

and, hence, subtractive color method is used. In subtractive color mixing, white means no color and black is a combination of all colors. As more colors are added, the result becomes darker and tends towards black. Cyan, magenta and yellow (CMY) colors correspond to primary colors in subtractive color system.

The primary colors are added to generate the secondary colors yellow (red + green), cyan (green + blue), magenta (red + blue). Several other color models, such as YUV, YIQ and HSI also exist and can be used as per the requirement.

The RGB to CMY conversion can be performed using:

$$\begin{aligned} C&=1-R \\ M&=1-G \\ Y&=1-B \end{aligned} \tag{2.1}$$

In the CMY model, a pixel of color cyan, for example, reflects all other RGB colors but red. A pixel with the color of magenta, on the other hand, reflects all other RGB colors but green. Now, if cyan and magenta colors are mixed, blue will be generated, rather than white as in the additive color system. In the CMY model, the black color generated by combining cyan, magenta and yellow is not very impressive, therefore a new color, black, is added in CMY color model, generating CMYK model. In publishing industry this CMYK model is referred as four-color printing.

There is another popular color model known as the YUV model. In the YUV color model, Y is luminance and U and V represent chrominance or color information. The YUV color model is based on color information along with brightness information. The components of YUV are:

$$\begin{aligned} Y &= 0.3\ R + 0.6\ G + 0.1\ B \\ U &= B\text{-}Y \qquad (2.2) \\ V &= R\text{-}Y \end{aligned}$$

The luminance component can be considered as a grayscale version of the RGB image. There are certain advantages of YUV compared to RGB, which are:

- The brightness information is separated from the color information.
- The correlations between the color components are reduced.
- Most of the information is in the *Y* component, while the information content in the *U* and *V* is less.

The latter two properties are very useful in various applications, such as image compression. This is because the correlations between the different color components are reduced, thus, each of the components can be compressed separately. Besides that, more bits can be allocated to the *Y* component than to *U* and *V*. The YUV color system is adopted in the JPEG image compression standard.

2.3.3 YIQ Color Model

The YIQ color model takes advantage of the human visual system which is more sensitive to luminance variations in comparison to variance in hue or saturation. In the YIQ color system, Y is the luminance component, and I and Q represent chrominance U and V similar to YUV color systems. The RGB to YIQ conversion can be done using

$$\begin{aligned} Y &= 0.299\ R + 0.587\ G + 0.114\ B \\ I &= 0.596\ R – 0.275\ G – 0.321\ B \qquad (2.3) \\ Q &= 0.212\ R – 0.523\ G + 0.311\ B \end{aligned}$$

The YIQ color model can also be described in terms of YUV:

$$\begin{aligned} Y &= 0.3\ R + 0.6\ G + 0.1\ B \\ I &= 0.74\ V – 0.27\ U \qquad (2.4) \\ Q &= 0.48\ V + 0.41\ U \end{aligned}$$

2.3.4 HSI Color Model

The HSI model is based on hue (H), saturation (S), and intensity (I). Hue is an attribute associated with the dominant wavelength in a mixture of light waves, i.e., the dominant color as perceived by a human. Saturation refers to the relative purity of the amount of white light mixed with the hue. Intensity corresponds to the luminance component (*Y*) of the YUV and YIQ models. The advantages of HSI are:

- The intensity is separated from the color information (similar to YUV and YIQ models).
- The hue and saturation components are intimately related to the way in which human beings perceive color.

whereas RGB can be described by a three-dimensional cube, the HSI model is represented as a color triangle, as shown in Fig. 2.9.

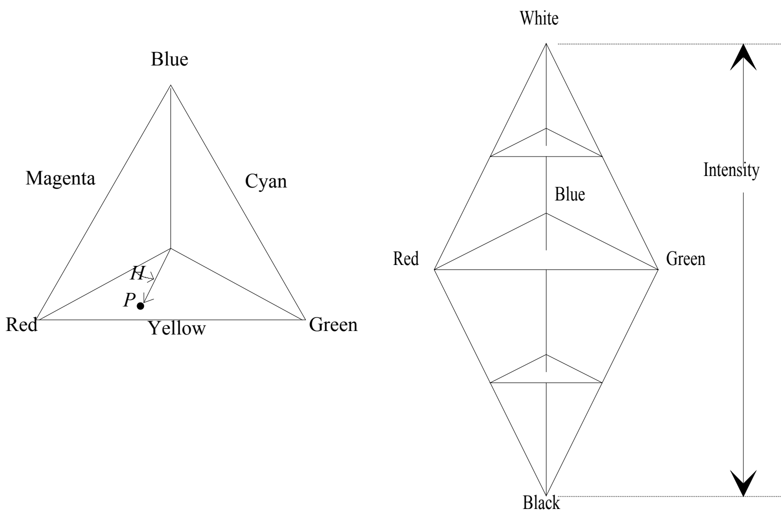


Fig. 2.9. HSI color triangle.

All colors lie inside the triangle whose vertices are defined by the three basic colors, red, green and blue. If a vector is drawn from the central point of the triangle to the color point *P*, then hue (*H*) is the angle of the vector with respect to the red axis. For example, 0º indicates red color, 60º yellow, 120º green, and so on. Saturation (*S*) is the degree to which the color is undiluted by white and is proportional to the distance to the center of the triangle.

The RGB to HSI conversion can be performed as follows:

$$H = \frac{1}{360^{o}} \cdot \cos^{-1}\left[\frac{\frac{1}{2}\cdot\left[(R-G)+(R-B)\right]}{\sqrt{(R-G)^{2}+(R-B)(G-B)}}\right], \text{ if } B \le G$$

$$H = 1 - \frac{1}{360^{o}} \cdot \cos^{-1}\left[\frac{\frac{1}{2}\cdot\left[(R-G)+(R-B)\right]}{\sqrt{(R-G)^{2}+(R-B)(G-B)}}\right], \text{ otherwise}$$

$$S=1-\frac{3}{(R+G+B)}\cdot\min(R,G,B)$$
$$I=\frac{1}{3}\cdot(R+G+B) \tag{2.5}$$

2.4 Basic Relationships between Pixels

The neighboring relationship between pixels is very important in an image processing task. A pixel p(x,y) has four neighbors (two vertical and two horizontal neighbors) specified by [(x+1, y), (x–1, y), (x, y+1), (x, y–1)].

This set of pixels is known as the 4-neighbors of P, and is denoted by N_4(P). All these neighbors are at a unit distance from P.

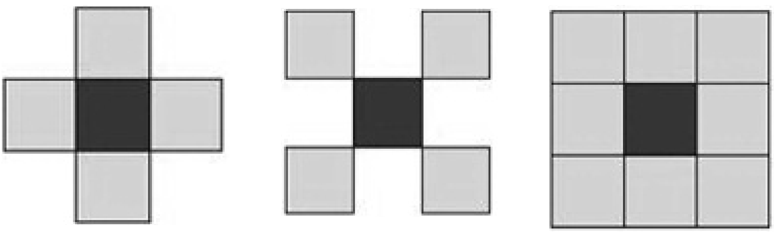
A pixel p(x,y) has four diagonal neighbors specified by, [(x+1, y+1), (x+1, y–1), (x–1, y+1), (x–1, y–1)].

This set of pixels is denoted by N_D(P). All these neighbors are at a distance of 1.414 in Euclidean space from P.

The points N_D(P) and N_4(P) are together known as 8-neighbors of the point P, denoted by N_8(P). In 8-neighbors, all the neighbors of a pixel p(x,y) are:

[(x+1, y), (x–1, y), (x, y+1), (x, y–1), (x+1, y+1), (x+1, y–1), (x–1, y+1), (x–1, y–1)].

Figure 2.10 shows 4-neigbours, diagonal neighbors and 8-neigbours of a pixel. For a pixel on the boundary, some neighboring pixels of a pixel may not exist in the image.



(a) **(b)** **(c)**

Fig. 2.10. (a) 4-neighbors of a pixel (b) Diagonal neighbors of a pixel (c) 8-neighbors of a pixel.

2.5 Adjacency

Adjacency of two pixels is defined in the terms of neighbors of a pixel. Two pixels p and q are said to be 4-adjacent if these pixels are 4-neighbors to each other, i.e., if these pixels have the same value in a binary image and q is 4-neighbor of p. Similarly, two pixels p and q are 8-adjacent, if these pixels are 8-neighbors to each other, i.e., if these pixels have the same value in a binary image and q is 8-neighbor of p. This concept can also be extended to grayscale images. In grayscale images, the value of intensity at pixels may not be same but they must belong to a defined set of intensities.

Two pixels, p and q, are called m-adjacent (mixed adjacent) if these two pixels are either 4-adjacent or diagonal adjacent but not both. m-adjacency is an improvement over 8-adjacency as sometimes 8-adjacency may be ambiguous.

2.6 Digital Path

A (digital) path (or curve) from a pixel p (x_0, y_0) to pixel $q(x_n, y_n)$ is a sequence of distinct pixels with coordinates $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$, where (x_i, y_i) and (x_{i-1}, y_{i-1}) are adjacent for $1 \le i \le n$. If $(x_0, y_0) = (x_n, y_n)$, the path is known as a closed path and n is the length of the path. 4-, 8-, and m-paths can be defined based on the type of adjacency used.

2.7 Connected Set

Let S represents a subset of pixels in an image. Two pixels $p(x_0, y_0)$ and q (x_n, y_n) are said to be connected in S if there exists a path $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$, where $\forall i, 0 \le i \le n, (x_i, y_i) \in S$. This set S is called a region of the image if the set S is a connected component and two regions are known to be adjacent to each other if their union forms a connected component, otherwise regions are said to be disjoint.

The boundary of the region S is the set of pixels in the region that have one or more neighbors that are not in S.

Summary

- An image is sensed and converted into digital form by the process of sampling and quantization.
- Sampling determines the spatial resolution of a digital image.
- Quantization determines the number of gray levels in the digital image.
- The quality of an image and file size is based on sampling and quantization.
- A color model is a well-defined system for generating a spectrum of colors.
- The two basic color models types are: Additive color model and Subtractive color model.
- Some common color models are RGB, CMY, CMYK and HSI.
- 4-Neigbours of a pixel are the pixels which are on the left, right, top and bottom of the pixel.
- 8-Neighbors of a pixel are the pixels which are on the left, right, top, bottom of the pixel and also diagonal neighbors of the pixel.

3

Mathematical Tools for Image Processing

3.1 Introduction

In various image processing activities, mathematical operations are very important. An image itself is represented in the form of a numerical matrix, having intensity or color information at each cell. In image processing operations, a large number of mathematical operations are used, e.g., interpolation in zooming and set operations in mathematical morphological operations. Various mathematical concepts, useful in implementing a number of image processing operations, are explained below.

3.2 Distance Function

Digital images follow certain metric and topological properties.

A distance function or metric D on a set S is a bivariate operator, i.e., it takes two operands, say $x \in S$ and $y \in S$, in order to map the set of non-negative real numbers, $[0, \infty)$.

$$D : S \times S \to [0,\infty)$$

A distance measure D is said to be a valid distance metric if, for all x, y, $z \in S$, the following conditions are satisfied:

$$\begin{aligned} &D(x,y) \geq 0 \text{ ; } D(x, y) = 0 \text{ if } x = y && \text{(non-negative or separation axiom)} \\ &D(x, y) = D(y, x) && \text{(symmetry)} \\ &D(x, z) \leq D(x, y) + D(y, z) && \text{(triangle inequality or subadditivity)} \end{aligned} \tag{3.1}$$

The distance between two pixels may be defined in various ways. A very simple way to define the distance is Euclidean distance, used in simple geometry.

Euclidean distance, also known as Pythagoras distance, is a simple straight-line distance defined as:

$$D((x_1,y_1), (x_2, y_2)) = \sqrt{((x_1 - x_2)^2 - (y_1 - y_2)^2)} \tag{3.2}$$

Apart from Euclidean, another popular distance metric is Manhattan distance, also known as 'taxi-cab' distance or 'city block' distance, defined as:

$$D_4 ((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2| \tag{3.3}$$

City block distance D_4 is the minimum number of elementary steps required to move from a starting point to the end point in a grid, withonly horizontal and vertical moves being permitted. If diagonal moves are also permitted then another distance D_8, called Chessboard distance, is obtained. Chessboard distance, or Chebyshev distance, is defined as:

$$D_8((x_1, y_1), (x_2, y_2)) = \max\{|x_1 - x_2|, |y_1 - y_2|\} \tag{3.4}$$

This distance computes the distance between two pixels by taking the maximum of their differences along any coordinate dimension. In two-dimensional space, this distance is equivalent to the minimum number of moves a king needs to make in order to travel between any two squares on the chessboard, hence the name Chessboard Distance.

Two pixels are called 4-neighbors if the D_4 distance between these two pixels is 1. This means that in 4-adjacency neighborhood only left, right, top and down adjacent pixels are considered while in 8-adjacency, the diagonal neighbors are also included. Two pixels are called 8-neighbors if the D_8 distance between these two pixels is 1.

If there exists a path between two pixels in the set of pixels in the image, then these pixels are known as contiguous pixels.

3.3 Convexity Property

If a line connecting any two points within a region lies within the region, then the region is defined as convex, otherwise it is non-convex.

3.4 Topological Properties

Topology or rubber sheet distortions is the study of properties of any object that remains unaffected by any deformation as long as there is no joining or tearing of the figure. For example, if any object has three holes in it, then even after rotating the object or stretching it, the number of holes will remain the same. Topological properties do not depend on the concept of distance. Any object that has a non-regular shape can be represented using its topological components.

3.5 Interpolation

Interpolation is the process used to find any missing value in a set of given values. For example, the value of a variable y is known at x = 1, 2, 4, 5, so the process used to find the value of y at x = 3 is known as interpolation. This is very useful in a number of image operations where there is a need to find the unknown value at any pixel, based on the known pixel values, i.e., where resampling is required, e.g., zooming of an image. A number of interpolation techniques are available in the literature. Traditional interpolating techniques are:

3.5.1 Nearest Neighbor Interpolation

This is a very simple interpolation technique, in which the nearest pixel intensity is replicated in order to find the unknown intensity value at any pixel location.

3.5.2 Bilinear Interpolation

In the bilinear interpolation technique, the weighted average of intensities of 2 x 2 neighboring pixels is used to interpolate an unknown intensity value.

In bilinear interpolation, the value of a pixel $x_{i,j}$ at the location (i,j) can be calculated as:

$$f(x) = a + \Delta i(b-a) + \Delta j(c-a) + \Delta i \Delta j(a-b-c+d) \tag{3.5}$$

where a, b, c, d are the nearest known pixel values to x; Δi and Δj define the relative distance from a to x (varying from 0 to 1).

3.5.3 Bicubic Interpolation

Bicubic interpolation uses the weighted average of the values of intensities of 4 x 4 neighboring pixels for finding the unknown intensity value at a pixel.

3.5.4 Cubic Spline Interpolation

Given a function *f* defined on [a, b] and a set of nodes $a = x_0 < x_1 < \cdots < x_n = b$, a cubic spline interpolant S (Fig. 3.1) for *f* is a function that satisfies the following conditions:

a) S(x) is a cubic polynomial, denoted $S_j(x)$, on the subinterval $[x_j, x_{j+1}]$, for each $j = 0, 1, \ldots, n-1$;

b) $S_j(x_j) = f(x_j)$ and $S_j(x_{j+1}) = f(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;

c) $S_{j+1}(x_{j+1}) = S_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;

d) $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;

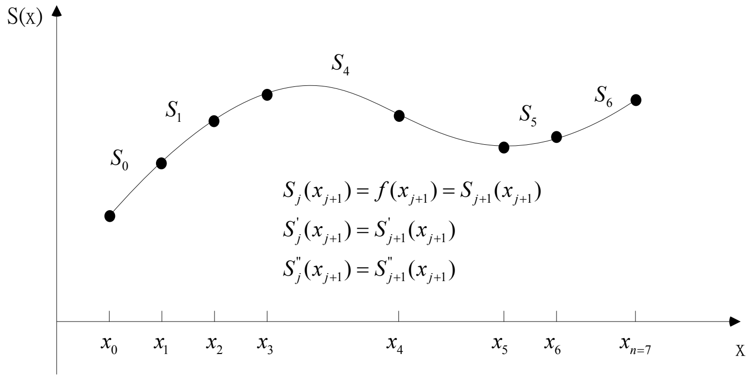


Fig. 3.1. Cubic Spline Interpolation.

e) $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;

f) One of the following sets of boundary conditions is satisfied:

i) $S''(x_0) = S''(x_n) = 0$ (free or natural boundary);

ii) $S'(x_0) = f'(x_0)$ and $S'(x_n) = f'(x_n)$ (clamped boundary).

To construct the cubic spline interpolant for a given function f, the conditions in the definition are applied to the cubic polynomials

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3$$

for each $j = 0, 1, \ldots, n-1$. (3.6)

3.6 Circularly Symmetric Signals

An arbitrary 2D signal $a(x, y)$ can be expressed in the polar coordinate system as $a(r, \theta)$. When the 2D signal exhibits a circular symmetry,

$$a(x, y) = a(r, \theta) = a(r) \tag{3.7}$$

where $r^2 = x^2 + y^2$ and $\tan \theta = y/x$. A number of physical systems exhibit circular symmetry (e.g., lenses), it is useful to be able to compute an appropriate Fourier representation.

3.7 Statistics

- Probability distribution function of the brightness
- Probability density function of the brightness
- Average
- Standard deviation
- Coefficient-of-variation
- Signal-to-Noise (SNR) ratio

In image processing it is quite common to use simple statistical descriptions of images and sub-images. The notion of a statistics is intimately connected to the concept of a probability distribution, generally the distribution of signal amplitudes. For a given region, which could conceivably be an entire image, the probability distribution function of the brightness in that region and probability density function of the brightness in that region can be defined.

3.7.1 Probability Distribution Function

The probability distribution function $P(a)$ is the probability that a brightness chosen from the region is less than or equal to a given brightness value a. As a increases from $-\infty$ to $+\infty$, $P(a)$ increases from 0 to 1. $P(a)$ is monotonic, non-decreasing in a and thus $dP/da >= 0$.

3.7.2 Probability Density Function

The probability that a brightness in a region falls between a and $a + \Delta a$, given the probability distribution function $P(a)$, can be expressed as $p(a)\Delta a$ where $p(a)$ is the probability density function.

$$p(a)\Delta a = \left(\frac{dp(a)}{da}\right)\Delta a \tag{3.8}$$

Because of monotonic, non-decreasing character of $p(a)$ we have $P(a) \geq 0$ and $\int_{-\propto}^{\propto} p(a)da = 1$. For an image with quantized (integer) brightness amplitudes, the interpretation of Δa is the width of a brightness interval. The brightness probability density function is frequently estimated by counting the number of times that each brightness occurs in the region in order to generate a histogram, $h[a]$. The histogram can then be normalized so that the total area under the histogram is 1. Or the $p(a)$ for region is the normalized count of the number of pixels, N, in a region that have quantized brightness a:

$$p[a] = \frac{1}{N} h[a] \qquad with \qquad N = \sum_{\alpha} h[a] \tag{3.9}$$

Both the distribution function and the histogram as measured from a region are statistical descriptions of that region. It must be emphasized that both $P(a)$ and $p(a)$ should be viewed as estimates of true distributions when they are computed from a specific region.

3.7.3 Average

The average brightness of a region is defined as the sample mean of the pixel brightness within that region. The average, m_a, of the brightness over the N pixels within a region is given by:

$$m_\alpha = \frac{1}{N} \sum_{m,n \varepsilon R} a[m,n] \tag{3.10}$$

3.7.4 Standard Deviation

The unbiased estimate of the standard deviation, S_a, of the brightness within a region R with N pixels is called the sample standard deviation and is given by:

$$\begin{aligned} S_a &= \sqrt{\frac{1}{N-1} \sum_{m,n \varepsilon R} (a[m,n] - m_a)^2} \\ &= \sqrt{\frac{\sum_{m,n \varepsilon R} a^2[m,n] - N m_a^2}{N-1}} \end{aligned} \tag{3.11}$$

Using the histogram formulation gives:

$$s_a = \sqrt{\frac{\left(\sum_a a^2 - h[a] - N \cdot m_a^2\right)}{N-1}}$$

The standard deviation, S_a, is an estimate σ_a of the underlying brightness probability distribution.

3.7.5 Coefficient-of-variation

The dimensionless coefficient-of-variation is defined as:

$$CV = \frac{s_a}{m_a} \times 100\% \tag{3.12}$$

3.7.6 Percentiles

The percentile, p%, of an unquantized brightness distribution is defined as the value of the brightness such that:

$$P(a) = p\%$$

or equivalently

$$\int_{-\infty}^{a} p(\alpha) d\alpha = p\%$$

Three special cases are frequently used in digital image processing.

- 0% of the minimum value in the region
- 50% of the median value in the region
- 100% of the maximum value in the region.

3.7.7 Mode

The mode of the distribution is the most frequent brightness value. There is no guarantee that a mode exists or that it is unique.

3.7.8 Signal-to-Noise Ratio

The noise is characterized by its standard deviation, S_n. The characterization of the signal can differ. If the signal is known to lie between two boundaries, $a_{\min} <= a <= a_{\max}$, then the signal-to-noise ratio (SNR) is defined as:

Bounded signal $SNR = 20\ log_{10} \left(\frac{S_a}{S_n}\right)$ dB (3.13)

3.8 Transforms

An image can be represented using pixel values (spatial domain) or using frequency spectrum (frequency domain). In frequency domain, an image can be expressed as a linear combination of some basis functions of some linear integral transform, e.g., Fourier transform, cosine transform, wavelet transform.

3.8.1 Two dimensional signals (images)

As a one-dimensional signal can be represented by an orthonormal set of basis vectors, an image can also be expanded in terms of a discrete set of basis arrays, called basis images, through a two dimensional (image) transform.

For an $N \times N$ image $f(x, y)$ the forward and inverse transforms are:

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

$$f(x,y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} I(x,y,u,v) g(u,v) \quad (3.14)$$

where, again, $T(u, v, x, y)$ and $I(x, y, u, v)$ are called the forward and inverse transformation kernels, respectively.

The forward kernel is said to be separable if

$$T(u, v, x, y) = T_1(u, x)\ T_2(v, y) \quad (3.15)$$

It is said to be symmetric if T_1 is functionally equal to T_2 such that

$$T(u,v,x,y) = T_1(u,x) T_1(v,y) \quad (3.16)$$

These properties are valid for the inverse kernel.

If the kernel $T(u, v, x, y)$ of an image transform is separable and symmetric, then the transform $g(u,v) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} T(u,v,x,y) f(x,y) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} T_1(u,x) T_1(v,y) f(x,y)$ can be written in matrix form:

$$g = T_1 \cdot f \cdot T_1^T \tag{3.17}$$

where f is the original image of size $N \times N$, and T_1 is an $N \times N$ transformation matrix with elements $t_{ij} = T_1(i, j)$. If, in addition, T_1 is a unitary matrix then the transform is called separable unitary and the original image is recovered through the relationship

$$f = T_1^{*T} \cdot g \cdot T_1^* \tag{3.18}$$

3.8.2 Discrete Fourier Transform

• Continuous space and continuous frequency

The Fourier transform is extended to a function $f(x, y)$ of two variables. If $f(x, y)$ is continuous and integrable and $F(u, v)$ is integrable, the following Fourier transform pair exists:

$$F(u,v) = \int_{-\infty}^{\infty}\int f(x,y) e^{-j2\pi(ux+vy)} dxdy \tag{3.19}$$

$$f(x,y) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int F(u,v) e^{j2\pi(ux+vy)} dudv \tag{3.20}$$

In general, $F(u, v)$ is a complex-valued function of two real frequency variables u, v and hence, it can be written as:

$$F(u,v) = R(u,v) + jI(u,v) \tag{3.21}$$

The amplitude spectrum, phase spectrum and power spectrum, respectively, are defined as follows:

$$|F(u,v)| = \sqrt{R^2(u,v) + I^2(u,v)}$$

$$\phi(u,v) = \tan^{-1}\left[\frac{I(u,v)}{R(u,v)}\right]$$

$$P(u,v) = |F(u,v)|^2 = R^2(u,v) + I^2(u,v) \tag{3.22}$$

• Discrete space and continuous frequency

For the case of a discrete sequence $f(x, y)$ of infinite duration, we can define the 2-D discrete space Fourier transform pair as follows:

$$F(u,v) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} f(x,y) e^{-j(xu+vy)} \tag{3.23}$$

$$f(x,y) = \frac{1}{(2\pi)^2} \int_{u=-\pi}^{\pi} \int_{v=-\pi}^{\pi} F(u,v) e^{j(xu+vy)} du dv \tag{3.24}$$

$F(u, v)$ is, again, a complex-valued function of two real frequency variables u, v and it is periodic with a period of $2\pi \times 2\pi$, that is $F(u,v) = F(u+2\pi, v) = F(u, v+2\pi)$.

The Fourier transform of $f(x, y)$ is said to converge uniformly when $F(u, v)$ is finite and

$$\lim_{N_1 \to \infty} \lim_{N_2 \to \infty} \sum_{x=-N_1}^{N_1} \sum_{y=-N_2}^{N_2} f(x,y) e^{-j(xu+vy)} = F(u,v) \text{ for all } u, v. \tag{3.25}$$

when the Fourier transform of $f(x, y)$ converges uniformly, $F(u, v)$ is an analytic function and is infinitely differentiable with respect to u and v.

- **Discrete space and discrete frequency: The two-dimensional Discrete Fourier Transform (2-D DFT)**

If $f(x, y)$ is an $M \times N$ array, such as the one obtained by sampling a continuous function of two dimensions at dimensions M and N on a rectangular grid, then its two dimensional Discrete Fourier transform (DFT) is the array given by

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi(ux/M + vy/N)} \tag{3.26}$$

where u = 0, 1, 2, ...M–1 and v = 0, 1 ,2 .... N–1

and the inverse DFT (IDFT) is

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{j2\pi(ux/M + vy/N)} \tag{3.27}$$

where x = 0, 1,2, ... M–1 and y = 0, 1, 2, ... N–1

When images are sampled in a square array, i.e., $M = N$, then

$$F(u,v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi(ux+vy)/N} \tag{3.28}$$

$$f(x,y) = \frac{1}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) e^{j2\pi(ux+vy)/N} \tag{3.29}$$

Two-dimensional Discrete Fourier Transform is separable, symmetric and unitary.

3.9 Wavelet Transform

The decomposition of an image g into wavelet coefficients through a Discrete Wavelet Transform (DWT) W can be expressed as: