Image Processing

Principles and Applications

Tinku Acharya
Avisere, Inc.
Tucson, Arizona
and
Department of Electrical Engineering
Arizona State University
Tempe, Arizona

Ajoy K. Ray
Avisere, Inc.
Tucson, Arizona
and
Electronics and Electrical Communication Engineering Department
Indian Institute of Technology
Kharagpur, India

WILEY-
INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Image Processing

This Page Intentionally Left Blank

Image Processing

Principles and Applications

Tinku Acharya
Avisere, Inc.
Tucson, Arizona
and
Department of Electrical Engineering
Arizona State University
Tempe, Arizona

Ajoy K. Ray
Avisere, Inc.
Tucson, Arizona
and
Electronics and Electrical Communication Engineering Department
Indian Institute of Technology
Kharagpur, India

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Acharya, Tinku.
Image processing : principles and applications / Tinku Acharya, Ajoy K. Ray.
p. cm.
"A Wiley-Interscience Publication."
Includes bibliographical references and index.
ISBN-13 978-0-471-71998-4 (cloth : alk. paper)
ISBN-10 0-471-71998-6 (cloth : alk. paper)
1. Image processing. I. Ray, Ajoy K., 1954- II. Title.
TA1637.A3 2005
621.36'7—dc22 2005005170

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

In memory of my father, Prohlad C. Acharya

—Tinku

In memories of my mother, father, and uncle

—Ajoy

This Page Intentionally Left Blank

Contents

Preface

There is a growing demand of image processing in diverse application areas, such as multimedia computing, secured image data communication, biomedical imaging, biometrics, remote sensing, texture understanding, pattern recognition, content-based image retrieval, compression, and so on. As a result, it has become extremely important to provide a fresh look at the contents of an introductory book on image processing. We attempted to introduce some of these recent developments, while retaining the classical ones.

The first chapter introduces the fundamentals of the image processing techniques, and also provides a window to the overall organization of the book. The second chapter deals with the principles of digital image formation and representation. The third chapter has been devoted to color and color imagery. In addition to the principles behind the perception of color and color space transforation, we have introduced the concept of color interpolation or demosaicing, which is today an integrated part of any color imaging device. We have described various image transformation techniques in Chapter 4. Wavelet transformation has become very popular in recent times for its many salient features. Chapter 5 has been devoted to wavelet transformation.

The importance of understanding the nature of noise prevalent in various types of images cannot be overemphasized. The issues of image enhancement and restoration including noise modeling and filtering have been detailed in Chapter 6. Image segmentation is an important task in image processing and pattern recognition. Various segmentation schemes have been elaborated in Chapter 7. Once an image is appropriately segmented, the next important task involves classification and recognition of the objects in the image. Various pattern classification and object recognition techniques have been presented in Chapter 8. Texture and shape play very important roles in image understanding. A number of texture and shape analysis techniques have been detailed in Chapter 9.

In sharp contrast with the classical crisp image analysis, fuzzy set theoretic approaches provide elegant methodologies for many image processing tasks. Chapter 10 deals with a number of fuzzy set theoretic approaches. We introduce content-based image retrieval and image mining in Chapter 11. Biomedical images like x-Ray, ultrasonography, and CT-Scan images provide sufficient information for medical diagnostics in biomedical engineering. We devote Chapter 12 to biomedical image analysis and interpretation. In this chapter, we also describe some of the biometric algorithms, particularly face recognition, signature verification, etc. In Chapter 13, we present techniques for remotely sensed images and their applications. In Chapter 14, we describe principles and applications of dynamic scene analysis, moving-object detection, and tracking. Image compression plays an important role for image storage and transmission. We devote Chapter 15 to fundamentals of image compression. We describe the JPEG standard for image compression in Chapter 16. In Chapters 17 and 18, we describe the new JPEG2000 standard.

The audience of this book will be undergraduate and graduate students in universities all over the world, as well as the teachers, scientists, engineers and professionals in R&D and research labs, for their ready reference.

We sincerely thank Mr. Chittabrata Mazumdar who was instrumental to bring us together to collaborate in this project. We are indebted to him for his continuous support and encouragement in our endeavors.

We thank our Editor, Val Moliere, and her staff at Wiley, for their assistance in this project. We thank all our colleagues in Avisere and Indian Institute of Technology, Kharagpur, particularly Mr. Roger Undhagen, Dr. Andrew Griffis, Prof. G. S. Sanyal, Prof. N. B. Chakrabarti, and Prof. Arun Majumdar for their continuous support and encouragement. We specially thank Odala Nagaraju, Shyama P. Choudhury, Brojeswar Bhowmick, Ananda Datta, Pawan Baheti, Milind Mushrif, Vinu Thomas, Arindam Samanta, Abhik Das, Abha Jain, Arnab Chakraborti, Sangram Ganguly, Tamalika Chaira, Anindya Moitra, Kaushik Mallick and others who have directly or indirectly helped us in the preparation of this manuscript in different ways. We thank anonymous reviewers of this book for their constructive suggestions.

Finally, we are indebted to our families for their active support throughout this project. Especially, Mrs. Baishali Acharya and Mrs. Supriya Ray stood strongly behind us in all possible ways. We would like to express our sincere appreciation to our children, Arita and Arani, and Aniruddha and Ananya, who were always excited about this work and made us proud.

Tinku Acharya
Ajoy K. Ray

1

Introduction

1.1 FUNDAMENTALS OF IMAGE PROCESSING

We are in the midst of a visually enchanting world, which manifests itself with a variety of forms and shapes, colors and textures, motion and tranquility. The human perception has the capability to acquire, integrate, and interpret all this abundant visual information around us. It is challenging to impart such capabilities to a machine in order to interpret the visual information embedded in still images, graphics, and video or moving images in our sensory world. It is thus important to understand the techniques of storage, processing, transmission, recognition, and finally interpretation of such visual scenes. In this book we attempt to provide glimpses of the diverse areas of visual information analysis techniques.

The first step towards designing an image analysis system is digital image acquisition using sensors in optical or thermal wavelengths. A two-dimensional image that is recorded by these sensors is the mapping of the three-dimensional visual world. The captured two dimensional signals are sampled and quantized to yield digital images.

Sometimes we receive noisy images that are degraded by some degrading mechanism. One common source of image degradation is the optical lens system in a digital camera that acquires the visual information. If the camera is not appropriately focused then we get blurred images. Here the blurring mechanism is the defocused camera. Very often one may come across images of outdoor scenes that were procured in a foggy environment. Thus any outdoor scene captured on a foggy winter morning could invariably result into a blurred image. In this case the degradation is due to the fog and mist in the atmosphere, and this type of degradation is known as atmospheric degradation. In some other cases there may be a relative motion between the object and the camera. Thus if the camera is given an impulsive displacement during the image capturing interval while the object is static, the resulting image will invariably be blurred and noisy. In some of the above cases, we need appropriate techniques of refining the images so that the resultant images are of better visual quality, free from aberrations and noises. Image enhancement, filtering, and restoration have been some of the important applications of image processing since the early days of the field [1]–[4].

Segmentation is the process that subdivides an image into a number of uniformly homogeneous regions. Each homogeneous region is a constituent part or object in the entire scene. In other words, segmentation of an image is defined by a set of regions that are connected and nonoverlapping, so that each pixel in a segment in the image acquires a unique region label that indicates the region it belongs to. Segmentation is one of the most important elements in automated image analysis, mainly because at this step the objects or other entities of interest are extracted from an image for subsequent processing, such as description and recognition. For example, in case of an aerial image containing the ocean and land, the problem is to segment the image initially into two parts–land segment and water body or ocean segment. Thereafter the objects on the land part of the scene need to be appropriately segmented and subsequently classified.

After extracting each segment, the next task is to extract a set of meaningful features such as texture, color, and shape. These are important measurable entities which give measures of various properties of image segments. Some of the texture properties are coarseness, smoothness, regularity, etc., while the common shape descriptors are length, breadth, aspect ratio, area, location, perimeter, compactness, etc. Each segmented region in a scene may be characterized by a set of such features.

Finally based on the set of these extracted features, each segmented object is classified to one of a set of meaningful classes. In a digital image of ocean, these classes may be ships or small boats or even naval vessels and a large class of water body. The problems of scene segmentation and object classification are two integrated areas of studies in machine vision. Expert systems, semantic networks, and neural network-based systems have been found to perform such higher-level vision tasks quite efficiently.

Another aspect of image processing involves compression and coding of the visual information. With growing demand of various imaging applications, storage requirements of digital imagery are growing explosively. Compact representation of image data and their storage and transmission through communication bandwidth is a crucial and active area of development today. Interestingly enough, image data generally contain a significant amount of superfluous and redundant information in their canonical representation. Image compression techniques helps to reduce the redundancies in raw image data in order to reduce the storage and communication bandwidth.

1.2 APPLICATIONS OF IMAGE PROCESSING

There are a large number of applications of image processing in diverse spectrum of human activities–from remotely sensed scene interpretation to biomedical image interpretation. In this section we provide only a cursory glance in some of these applications.

1.2.1 Automatic Visual Inspection System

Automated visual inspection systems are essential to improve the productivity and the quality of the product in manufacturing and allied industries [5]. We briefly present few visual inspection systems here.

- **Automatic inspection of incandescent lamp filaments:** An interesting application of automatic visual inspection involves inspection of the bulb manufacturing process. Often the filament of the bulbs get fused after short duration due to erroneous geometry of the filament, e.g., nonuniformity in the pitch of the wiring in the lamp. Manual inspection is not efficient to detect such aberrations.

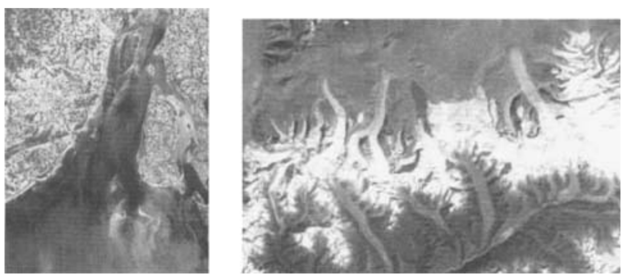
  In an automated vision-based inspection system, a binary image slice of the filament is generated, from which the silhouette of the filament is produced. This silhouette is analyzed to identify the non-uniformities in the pitch of the filament geometry inside the bulb. Such a system has been designed and installed by the General Electric Corporation.

- **Faulty component identification:** Automated visual inspection may also be used to identify faulty components in an electronic or electromechanical systems. The faulty components usually generate more thermal energy. The infra-red (IR) images can be generated from the distribution of thermal energies in the assembly. By analyzing these IR images, we can identify the faulty components in the assembly.

- **Automatic surface inspection systems:** Detection of flaws on the surfaces is important requirement in many metal industries. For example, in the hot or cold rolling mills in a steel plant, it is required to detect any aberration on the rolled metal surface. This can be accomplished by using image processing techniques like edge detection, texture identification, fractal analysis, and so on.

1.2.2 Remotely Sensed Scene Interpretation

Information regarding the natural resources, such as agricultural, hydrological, mineral, forest, geological resources, etc., can be extracted based on remotely sensed image analysis. For remotely sensed scene analysis, images of the earth's surface are captured by sensors in remote sensing satellites or by a multi-spectral scanner housed in an aircraft and then transmitted to the Earth Station for further processing [6, 7]. We show examples of two remotely sensed images in Figure 1.1 whose color version has been presented in the color figure pages. Figure 1.1(a) shows the delta of river Ganges in India. The light blue segment represents the sediments in the delta region of the river, the deep blue segment represents the water body, and the deep red regions are mangrove swamps of the adjacent islands. Figure 1.1(b) is the glacier flow in Bhutan Himalayas. The white region shows the stagnated ice with lower basal velocity.



(a) (b)

Fig. 1.1 Example of a remotely sensed image of (a) delta of river Ganges, (b) Glacier flow in Bhutan Himalayas. *Courtesy: NASA/GSFC/METI/ERSDAC/JAROS, and U.S./Japan ASTER Science Team.*

Techniques of interpreting the regions and objects in satellite images are used in city planning, resource mobilization, flood control, agricultural production monitoring, etc.

1.2.3 Biomedical Imaging Techniques

Various types of imaging devices like X-ray, computer aided tomographic (CT) images, ultrasound, etc., are used extensively for the purpose of medical diagnosis [8]–[10]. Examples of biomedical images captured by different image formation modalities such as CT-scan, X-ray, and MRI are shown in Figure 1.2.

(i) localizing the objects of interest, i.e. different organs

(ii) taking the measurements of the extracted objects, e.g. tumors in the image

(a) (b) (c)

Fig. 1.2 Examples of (a) CT-scan image of brain, (b) X-ray image of wrist, (c) MRI image of brain.

(iii) interpreting the objects for diagnosis.

Some of the biomedical imaging applications are presented below.

(A) *Lung disease identification:* In chest X-rays, the structures containing air appear as dark, while the solid tissues appear lighter. Bones are more radio opaque than soft tissue. The anatomical structures clearly visible on a normal chest X-ray film are the ribs, the thoracic spine, the heart, and the diaphragm separating the chest cavity from the abdominal cavity. These regions in the chest radiographs are examined for abnormality by analyzing the corresponding segments.

(B) *Heart disease identification:* Quantitative measurements such as heart size and shape are important diagnostic features to classify heart diseases. Image analysis techniques may be employed to radiographic images for improved diagnosis of heart diseases.

(C) *Digital mammograms:* Digital mammograms are very useful in detecting features (such as micro-calcification) in order to diagnose breast tumor. Image processing techniques such as contrast enhancement, segmentation, feature extraction, shape analysis, etc. are used to analyze mammograms. The regularity of the shape of the tumor determines whether the tumor is benign or malignant.

1.2.4 Defense surveillance

Application of image processing techniques in defense surveillance is an important area of study. There is a continuous need for monitoring the land and oceans using aerial surveillance techniques.

Suppose we are interested in locating the types and formation of Naval vessels in an aerial image of ocean surface. The primary task here is to segment different objects in the water body part of the image. After extracting the segments, the parameters like area, location, perimeter, compactness, shape, length, breadth, and aspect ratio are found, to classify each of the segmented objects. These objects may range from small boats to massive naval ships. Using the above features it is possible to recognize and localize these objects. To describe all possible formations of the vessels, it is required that we should be able to identify the distribution of these objects in the eight possible directions, namely, north, south, east, west, northeast, northwest, southeast and southwest. From the spatial distribution of these objects it is possible to interpret the entire oceanic scene, which is important for ocean surveillance.

1.2.5 Content-Based Image Retrieval

Retrieval of a query image from a large image archive is an important application in image processing. The advent of large multimedia collection and digital libraries has led to an important requirement for development of search tools for indexing and retrieving information from them. A number of good search engines are available today for retrieving the text in machine readable form, but there are not many fast tools to retrieve intensity and color images. The traditional approaches to searching and indexing images are slow and expensive. Thus there is urgent need for development of algorithms for retrieving the image using the embedded content in them.

The features of a digital image (such as shape, texture, color, topology of the objects, etc.) can be used as index keys for search and retrieval of pictorial information from large image database. Retrieval of images based on such image contents is popularly called the content-based image retrieval [11, 12].

1.2.6 Moving-Object Tracking

Tracking of moving objects, for measuring motion parameters and obtaining a visual record of the moving object, is an important area of application in image processing [13, 14]. In general there are two different approaches to object tracking:

1. Recognition-based tracking

2. Motion-based tracking.

A system for tracking fast targets (e.g., a military aircraft, missile, etc.) is developed based on motion-based predictive techniques such as Kalman filtering, extended Kalman filtering, particle filtering, etc. In automated image processing based object tracking systems, the target objects entering the sensor field of view are acquired automatically without human intervention. In recognition-based tracking, the object pattern is recognized in successive image-frames and tracking is carried-out using its positional information.

1.2.7 Image and Video Compression

Image and video compression is an active application area in image processing [12, 15]. Development of compression technologies for image and video continues to play an important role for success of multimedia communication and applications. Although the cost of storage has decreased significantly over the last two decades, the requirement of image and video data storage is also growing exponentially. A digitized 36 cm $\times$ 44 cm radiograph scanned at 70 μm requires approximately 45 Megabytes of storage. Similarly, the storage requirement of high-definition television of resolution 1280 $\times$ 720 at 60 frames per second is more than 1250 Megabits per second. Direct transmission of these video images without any compression through today's communication channels in real-time is a difficult proposition. Interestingly, both the still and video images have significant amount of visually redundant information in their canonical representation. The redundancy lies in the fact that the neighboring pixels in a smooth homogeneous region of a natural image have very little variation in their values which are not noticeable by a human observer. Similarly, the consecutive frames in a slow moving video sequence are quite similar and have redundancy embedded in them temporally. Image and video compression techniques essentially reduce such visual redundancies in data representation in order to represent the image frames with significantly smaller number of bits and hence reduces the requirements for storage and effective communication bandwidth.

1.3 HUMAN VISUAL PERCEPTION

Electromagnetic radiation in the optical band generated from our visual environment enters the visual system through eyes and are incident upon the sensitive cells of the retina. The activities start in the retina, where the signals from neighboring receivers are compared and a coded message dispatched on the optic nerves to the cortex, behind our ears. An excellent account of human visual perception may be found in [16]. The spatial characteristics of our visual system have been proposed as a nonlinear model in [17, 18].

Although the eyes can detect tranquility and static images, they are essentially motion detectors. The eyes are capable of identification of static objects and can establish spatial relationships among the various objects and regions in a static scene. Their basic functioning depends on comparison of stimuli from neighboring cells, which results in interpretation of motion. When observing a static scene, the eyes perform small repetitive motions called *saccades* that move edges past receptors. The perceptual recognition and interpretation aspects of our vision, however, take place in our brain. The objects and different regions in a scene are recognized in our brain from the edges or boundaries that encapsulate the objects or the regions inside the scene. The maximum information about the object is embedded along these edges or boundaries. The process of recognition is a result of learning that takes place in our neural organization. The orientation of lines and the directions of movements are also used in the process of object recognition.

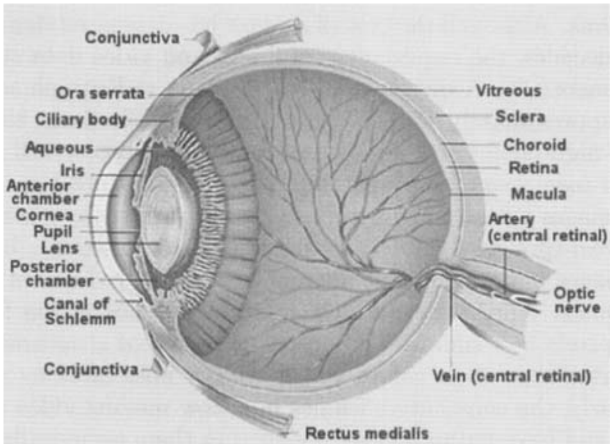


Fig. 1.3 Structure of human eye.

1.3.1 Human Eyes

The structure of an eye is shown in Figure 1.3. The transportation of the visual signal from the retina of the eye to the brain takes place through approximately one and a half million neurons via optic nerves. The retina contains a large number of photo-receptors, compactly located in a more or less regular, hexagonal array. The retinal array contains three types of color sensors, known as *cones* in the central part of the retina named as fovea centralis. The cones are distributed in such a way that they are densely populated near the central part of the retina and the density reduces near the peripheral part of the fovea. There are three different types of cones, namely red, green and blue cones which are responsible for color vision. The three distinct classes of cones contain different photosensitive pigments. The three pigments have maximum absorptions at about 430 nm (violet), 530 nm (blue-green) and 560 nm (yellow-green).

Another type of small receptors fill in the space between the cones. These receptors are called *rods* which are responsible for gray vision. These receptors are more in number than the cones.

Rods are sensitive to very low-levels of illumination and are responsible for our ability to see in dim light (scotopic vision). The cone or photopic system, on the other hand, operates at high illumination levels when lots of photons are available, and maximizes resolution at the cost of reduced sensitivity.

1.3.2 Neural Aspects of the Visual Sense

The optic nerve in our visual system enters the eyeball and connects with rods and cones located at the back of the eye.

The neurons contain dendrites (inputs), and a long axon with an arborization at the end (outputs). The neurons communicate through synapses. The transmission of signals is associated with the diffusion of the chemicals across the interface and the receiving neurons are either stimulated or inhibited by these chemicals, diffusing across the interface. The optic nerves begin as bundles of axons from the ganglion cells on one side of the retina. The rods and cones, on the other side, are connected to the ganglion cells by bipolar cells, and there are also horizontal nerve cells making lateral connections.

The signals from neighboring receptors in the retina are grouped by the horizontal cells to form a receptive field of opposing responses in the center and the periphery, so that a uniform illumination of the field results in no net stimulus. In case of nonuniform illumination, a difference in illumination at the center and the periphery creates stimulations. Some receptive fields use color differences, such as red-green or yellow-blue, so the differencing of stimuli applies to color as well as to brightness. There is further grouping of receptive field responses in the lateral geniculate bodies and the visual cortex for directional edge detection and eye dominance. This is low-level processing preceding the high-level interpretation whose mechanisms are unclear. Nevertheless, it demonstrates the important role of differencing in the senses, which lies at the root of contrast phenomena. If the retina is illuminated evenly in brightness and color, very little nerve activity occurs.

There are 6 to 7 million cones, and 110 to 130 million rods in a normal human retina. Transmission of the optical signals from rods and cones takes place through the fibers in the optic nerves. The optic nerves cross at the optic chiasma, where all signals from the right sides of the two retinas are sent to the right half of the brain, and all signals from the left, to the left half of the brain. Each half of the brain gets half a picture. This ensures that loss of an eye does not disable the visual system. The optical nerves end at the lateral geniculate bodies, halfway back through the brain, and the signals are distributed to the visual cortex from there. The visual cortex still has the topology of the retina, and is merely the first stage in perception, where information is made available. Visual regions in two cerebral hemispheres are connected in the corpus callosum, which unites the halves of the visual field.

1.4 COMPONENTS OF AN IMAGE PROCESSING SYSTEM

There are several components of an image processing system. The first major component of an image processing system is a camera that captures the images of a three-dimensional object.

1.4.1 Digital Camera

The sensors which are used in most of the cameras are either charge coupled device (CCD) or CMOS sensors. The CCD camera comprises a very large number of very small photo diodes, called photosites. The electric charges which are accumulated at each cell in the image are transported and are recorded after appropriate analog to digital conversion.

In CMOS sensors, on the other hand, a number of transistors are used for amplification of the signal at each pixel location. The resultant signal at each pixel location is read individually. Since several transistors are used the light sensitivity is lower. This is because of the fact that some of the photons are incident on these transistors (used for signal amplification), located adjacent to the photo-sensors. The current state-of-the-art CMOS sensors are more noisy compared to the CCD sensors. However, they consume low power and they are less expensive.

In case of bright sunlight the aperture, located behind the camera lens, need not be large since we do not require much light, while on cloudy days when we need more light to create an image the aperture should be enlarged. This is identical to the functioning of our eyes. The shutter speed gives a measure of the amount of time during which the light passes through the aperture. The shutter opens and closes for a time duration which depends on the requirement of light. The focal length of a digital camera is the distance between the focal plane of the lens and the surface of the sensor array. Focal length is the critical information in selecting the amount of required magnification which is desired from the camera.

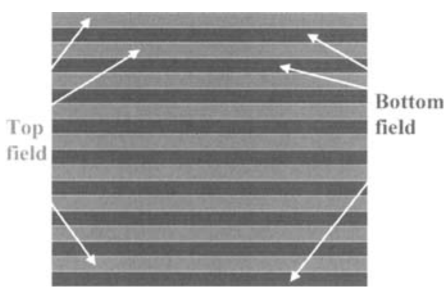


Fig. 1.4 Top and bottom fields in interlace scan.

In an interlaced video camera, each image frame is divided in two fields. Each field contains either the even (top field) or odd (bottom field) horizontal video lines. These two fields are assembled by the video display device. The mode of assembling the top and bottom fields in an interlace camera is shown in Fig. 1.4. In progressive scan cameras on the other hand, the entire frame is output as a single frame. When a moving scene is imaged, such as in robotic vision, it is captured using strobe pulse to illuminate the object in the scene. In such cases of imaging applications, progressive scan cameras are preferable.

Interlaced cameras are not used in such applications because the illumination time may be shorter than the frame time and only one field will be illuminated and captured if interlaced scanning is used.

A digital camera can capture images in various resolutions, e.g., 320 × 240, or 352 × 288, or 640 × 480 pixels on the low to medium resolution range to 1216 × 912 or 1600 × 1200 pixels on the high resolution size. The cameras that we normally use can produce about 16 million colors, i.e., at each pixel we can have one of 16 million colors.

The spatial resolution of an image refers to the image size in pixels, which corresponds to the size of the CCD array in the camera. The process of zooming an image involves performing interpolation between pixels to produce a zoomed or expanded form of the image. Zooming does not increase the information content in addition to what the imaging system provides. The resolution, however, may be decreased by subsampling which may be useful when system bandwidth is limited. Sensor resolution depends on the smallest feature size of the objects in a scene that we need our imaging system to distinguish, which is a measure of the object resolution. For example in an OCR system, the minimum object detail that needs to be discerned is the minimum width of line segments that constitute the pattern. In case of a line drawing, the minimum feature size may be chosen as two pixels wide. The sensor resolution of a camera is the number of rows and columns of the CCD array, while the field of view FOV is the area of the scene that the camera can capture. The FOV is chosen as the horizontal dimension of the inspection region that includes all the objects of interest. The sensor resolution of the camera = 2FOV/object resolution. The sensor resolution or sensor size is thus inversely proportional to the object resolution. The resolution of quantization refers to the number of quantization levels used in analog to digital (A/D) conversions. Higher resolution in this sense implies improved capability of analyzing low-contrast images.

Line scan cameras use a sensor that has just a row of CCD elements. An image may be captured by either moving the camera or by moving the image being captured by the camera. The number of elements in a line scan camera ranges from 32 to 8096. Even a single detector moved in a scanning pattern over an area can also be used to produce a video signal. A number of features, such as shutter control, focus control, exposure time control along with various triggering features are supported in cameras.

1.4.1.1 Capturing colors in a digital camera There are several ways in which a digital camera can capture colors. In one approach, one uses red, green, and blue filters and spins them in front of each single sensor sequentially one after another and records three separate images in three colors at a very fast rate. Thus the camera captures all the three color components at each pixel location. While using this strategy an automatic assumption is that during the process of spinning the three filters, the colors in the image must not change (i.e., they must remain stationary). This may not be a very practical solution.

A practical solution is based on the concept of color interpolation or demosaicing, which is a more economical way to record the three primary colors of an image. In this method, we permanently place only one type of filter over each individual photo-site. Usually the sensor placements are carried out in accordance to a pattern. The most popular pattern is called the Bayer's pattern [19], where each pixel is indicated by only one color–red, blue, or green pixel. It is possible to make very accurate guesses about the missing color component in each pixel location by a method called *color interpolation* or *demosaicing* [20, 21]. We cover different methods of color interpolation in Chapter 3.

In high-quality cameras, however, three different sensors with the three filters are used and light is directed to the different sensors by using a beam splitter. Each sensor responds only to small wavelength band of color. Thus the camera captures each of the three colors at each pixel location. These cameras will have more weight and they are costly.

1.5 ORGANIZATION OF THE BOOK

In this chapter, we introduced some fundamental concepts and a brief introduction to digital image processing. We have also presented few interesting applications of image processing in this chapter.

Chapter 2 deals with the principles of image formation and their digital representation in order to process the images by a digital computer. In this chapter, we also review the concepts of sampling and quantization, as well as the various image representation and formatting techniques.

In Chapter 3, we present the basics of color imagery, the color spaces and their transformation techniques. In this chapter, we also present a novel concept of color interpolation to reconstruct full color imagery from sub-sampled colors prevalent in low-cost digital camera type image processing devices.

Chapter 4 has been devoted to discuss various image transformation techniques and their underlying theory. Some of the popular image transformation techniques such as Discrete Fourier Transform, Discrete Cosine Transform, Karhaunen-Loeve Transform, Singular Value decomposition, Walsh-Hadamard transform and their salient properties are discussed here.

Wavelet transformation has become very popular in image processing applications in recent times for its many salient features. Chapter 5 has been devoted to wavelet transformation. We discuss both the convolution and lifting based algorithms for implementation of the DWT.

The importance of understanding the nature of noise and imprecision prevalent in various types of images cannot be overemphasized. This issue has been detailed in Chapter 6. We present a number of algorithms for enhancement, restoration, and filtering of images in this chapter.

Image segmentation is possibly one of the most important tasks in image processing. Various edge detection schemes have been elaborated in Chapter 7. Region based segmentation strategies such as thresholding, region growing, and clustering strategies have been discussed in this chapter.

Once an image is appropriately segmented, the next important task involves classification and recognition of the objects in the image. The various supervised and unsupervised pattern classification and object recognition techniques have been presented in Chapter 8. Several neural network architectures namely multilayer perceptron, Kohonen's Self Organizing feature map, and counterpropagation networks have been discussed in this chapter.

Texture and shape of objects play a very important role in image understanding. A number of different texture representation and analysis techniques have been detailed in Chapter 9. In this chapter, we have also discussed various shape discrimination strategies with examples.

In sharp contrast with the classical crisp image analysis techniques, fuzzy set theoretic approaches provide elegant methodologies which yield better results in many image processing tasks. We describe a number of image processing algorithms based on fuzzy set theoretic approaches in Chapter 10.

In today's world dealing with Internet, the application on content based image retrieval became important because of image search and other multimedia applications. We introduce the concepts of *content-based image retrieval* and *image mining*in Chapter 11.

Biomedical images like x-Ray, ultrasonography, and CT-scan images provide sufficient information for medical diagnostics in biomedical engineering. We devote Chapter 12 to biomedical image analysis and interpretation. In this chapter, we also describe two important applications of biometric recognition, viz., face recognition and signature verification.

Remote sensing is one of the most important applications in image processing. We discuss various satellite based remotely sensed image processing applications in Chapter 13.

In Chapter 14, we describe principles and applications of dynamic scene analysis, moving-object detection, and tracking. We also included recent developments such as condensation algorithm and particle filtering for object tracking.

Image compression plays an important role for image storage and transmission. We devote Chapter 15 to describe the fundamentals of image compression and principles behind it. There are many image compression techniques in the literature. However, adhering to image compression standards is important for interoperability and exchange of image data in today's networked world. The international standard organization, defined the algorithms and formats for image compression towards this goal. We describe the JPEG standard for image compression in Chapter 16.

In this era of internet and multimedia communication, it is necessary to incorporate new features and functionalities in image compression standards in order to serve diverse application requirements in the market place.

JPEG2000 is the new image compression standard to achieve this goal. In Chapters 17 and 18, we elaborate on the JPEG2000 standard, its applications and implementation issues.

1.6 HOW IS THIS BOOK DIFFERENT?

With the growth of diverse applications, it became a necessity to provide a fresh look at the contents of an introductory image processing book. In our knowledge there is no other book that covers the following aspects in detail.

- We present a set of *advanced topics*, in this book, retaining the classical ones.
- We cover several applications such as *biomedical* and *biometric* image processing, *Content based image retrieval*, *remote sensing*, *dynamic scene analysis*, *pattern recognition*, *shape* and *texture* analysis, etc.
- We include new concepts in *color interpolation* to produce full color from sub-sampled Bayer pattern color prevalent in today's digital camera and other imaging devices [21].
- The concepts of Discrete Wavelet Transform and its efficient implementation by *lifting* approach have been presented in great detail.
- In this era of internet and multimedia communication, there is necessity to incorporate many new features and functionalities in *image compression standards* to serve diverse application. *JPEG2000* is the new image compression standard to achieve this goal [15]. We devote two chapters on the *JPEG2000 standard* in great detail.
- We present the concepts and techniques of *Content based image retrieval* and *image mining* [11].
- The principles of *moving-object detection* and *tracking*, including recent developments such as *condensation algorithm* and *particle filtering* for object tracking [14] have been discussed in this book.
- Applications of *dental* and *mammogram* image analysis in biomedical image processing [9, 10] have been presented here.
- Both the *soft* and *hard* computing approaches have been dealt in greater length with respect to the major image processing tasks [11].
- The *fuzzy set theoretic* approaches are rich to solve many image processing tasks, but not much discussions are present in the classical image processing books [22, 23].

- We present the direction and development of current *research* in certain areas of image processing.
- We have provided *extensive bibliography* in the unified framework of this book.

1.7 SUMMARY

In this chapter, we have introduced the concepts, underlying principles, and applications of image processing. We have visited the role of eyes as the most important visual sensor in the human and animal world. The components constituting a computer vision system are presented briefly here. The organization of book and how this book is different from other image processing books currently in the market have also been discussed.

REFERENCES

1. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Second Edition, Volume 1, Academic Press, 1982.
2. W. K. Pratt, *Digital Image Processing*, Second Edition, Wiley, New York, 1991.
3. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA, 1992.
4. R. N. Bracewell, *Two-Dimensional Imaging*, Prentice Hall, Englewood Cliffs, NJ, 1995.
5. D. T. Pham and R. Alcock, *Smart Inspection Systems: Technicuqes and Applications of Intelligent Vision*, Academic Press, Oxford, 2003.
6. T. M. Lissesand and R. W. Kiefer, *Remote Sensing and Image Interpretation*, 4th Edition, John Wiley and Sons, 1999.
7. J. R. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective*, Prentice Hall, 2000.
8. P. Suetens, *Fundamentals of Medical Imaging* Cambridge University Press, 2002.
9. P. F. Van Der stelt and Qwil G.M.Geraets, "Computer aided interpretation and quantication of angular periodontal Bone defects on dental radiographs", *IEEE Transactions on Biomedical engineering*, 38(4), April 1998, 334–338.

10. M. A. Kupinski and M. Giger, "Automated Seeded Lesion Segmentation on Digital Mammograms," IEEE Trans. Med. Imag., Vol. 17, 1998, 510–517.

11. S. Mitra and T. Acharya, *Data Mining: Multimedia, Soft Computing, and Bioinformatics*, Wiley, Hoboken, NJ, 2003.

12. A. K. Ray and T. Acharya. *Information Technology: Principles and Applications.* Prentice Hall of India, New Delhi, India, 2004.

13. D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. on Automation and Control*, Vol. AC-24, December 1979, 84–90.

14. R. Cucchiara, C. Grana, G. Neri, M. Piccardi, and A. Prati, "The Sakbot System for Moving Object Detection and Tracking," *Video-Based Surveillance Systems-Computer Vision and Distributed Processing*, 2001, 145–157.

15. T. Acharya and P. S. Tsai. *JPEG2000 Standard for Image Compression: Concepts, Algorithms, and VLSI Architectures*, Wiley, Hoboken, NJ, 2004.

16. G. Wyszecki and W. S. Stiles, *Color Science*, Second Edition, McGraw-Hill, NY, 1982.

17. T. G. Stockham, Jr., "Image Processing in the context of a Visual Model," *Proceedings of IEEE*, 60(7), July 1972, 828–842.

18. C. F. Hall, and E. L. Hall, "A Nonlinear Model for the Spatial Characteristics of the Human Visual System," *IEEE Trans. Systems, Man, and Cybernetics*, SMC-7(3), March 1977, 161–170.

19. B. E. Bayer, "Color Imaging Array," *US Patent 3,971,065*, Eastman Kodak Company, 1976.

20. T. Sakamoto, C. Nakanishi, and T. Hase, "Software Pixel Interpolation for Digital Still Cameras Suitable for A 32-bit MCU," *IEEE Transactions on Consumer Electronics*, 44(4), November 1998, 1342–1352.

21. P. Tsai, T. Acharya, and A. K. Ray, "Adaptive Fuzzy Color Interpolation," *Journal of Electronic Imaging*, 11(3), July 2002, 293–305.

22. L. A. Zadeh, "Fuzzy Sets," *Information and Control*, 8, 1965, 338–353.

23. C. V. Jawahar and A. K. Ray, "Fuzzy Statistics of Digital Images," *IEEE Signal Processing Letter*, 3, 1996, 225–227.

2

Image Formation and Representation

2.1 INTRODUCTION

There are three basic components of image formation, i.e., the illumination, the reflectance models of surfaces which are imaged, and the process of image formation at the retina of human eyes or at the sensor plane of the camera. Once the images are formed (which is a two-dimensional analog signal), the next process involves sampling and digitization of the analog image. The digital images so formed after all these processes need to be represented in appropriate format so that they may be processed and manipulated by a digital computer for various applications. In this chapter, we discuss the principles of image formation and the various representation schemes.

2.2 IMAGE FORMATION

Understanding of physics of illumination is the first step of understanding of image formation. We start our discussion with the physics of illumination.

2.2.1 Illumination

Illumination is a fundamental component in the image formation process, which generates sense in our visual organ. Light produces the psychological sensation when it impinges on our eyes and excites our visual sense. The strength of this sensation, which is the sensation of brightness, can be quantified by averaging the responses of many human observers. The average response, i.e., the psychovisual sensation is determined at different spectral wavelengths. The peak spectral sensitivity of a human observer happens at 555 nm wavelength. If this sensitivity is normalized to one, then the sensitivity drops down to 0.0004 at the two ends of the optical spectrum (i.e., at 400 nm and 735 nm).

It may be noted here that equal amounts of luminous flux produce equal brightness, which is proportional to the logarithm of the luminous flux. Fechner's Law defines the brightness by the relation

$$B = k \log(\frac{F}{F_0}),$$

where F_0 is a reference luminous flux, measured in lumane (lm). The above relation shows that doubling the luminous flux does not double the apparent brightness.

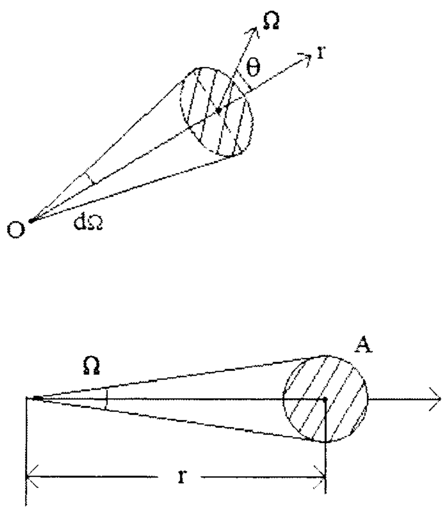


Fig. 2.1 Differential solid angle formation.

Let us consider a point source which emits luminous flux along radial lines. This point source of illumination may be anisotropic. A finite amount of radiation is emitted from the anisotropic point source in a finite cone. This cone has its vertex at the point source O, and its base of area dA at a distance r from the point source O, the normal to dA making an angle θ with the radius. Then, this cone is measured by the differential solid angle

$$d\Omega = \frac{dA \cos\theta}{r^2}$$

measured in steradians as shown in Figure 2.1. It is positive or negative as the normal to dA points outwards or inwards.

It is clear that the total solid angle surrounding a point is 4π. The luminous intensity I of a point source is the ratio $\frac{dF}{d\Omega}$, where F is the luminance flux. The luminous intensity is in general a function of direction and it is measured in candela (*cd*). If 1 lm (lumane) is emitted per steradian, the intensity is 1 *cd*. An isotropic point source of intensity I candela emits $4\pi I$ lumane.

The luminous flux incident on area dA from a source of intensity I is

$$dF = I\frac{dA\cos\theta}{r^2},$$

as shown in Figure 2.1. This follows directly from the definition of I as luminous flux per unit solid angle and the definition of solid angle. If the source is an extended one, then this must be integrated over the source area. The luminous flux per unit area falling on a surface is called the illumination E of the surface, and is measured in lm/m^2 (lumane per square meter), which is called a lux. For a point source,

$$E = \frac{dF}{dA} = I\frac{\cos\theta}{r^2}.$$

When a surface is illuminated, the response to incident light differs quite significantly, depending upon the nature of the surface. There are different types of surfaces with different characteristics. Some surfaces may be perfectly absorbing (e.g., black absorbing surfaces), which absorb the entire incident luminous flux and do not reflect any light. Other surfaces reflect the light incident on them.

2.2.2 Reflectance Models

Depending on the nature of reflection we group them in three categories– *Lambertian*, *Specular*, and *Hybrid* surfaces.

- **Lambertian Reflectance:** The *Lambertian* surfaces are those surfaces from which light is reflected in all directions. The nature of such reflectance is a diffused one. The reflection from the wall painted with flat paints, papers, fabrics, ground surfaces are some of the examples of *Lambertian reflection*. The illuminated region of the surface emits the entire incident light in all directions covering solid angle 2π radians. The Lambertian surface appears equally bright from all directions (i.e., equal projected areas radiate equal amounts of luminous flux). Many real surfaces approach to be nearly Lambertian. The reflectance map of the Lambertian surface may be modelled as

$$I_L = E_0 A\cos\theta,$$

where E_0 is the strength of the incident light source, A is the surface area of the Lambertian patch and θ is the angle of incidence. Such a model applies better when the angle of incidence as well as the angle of reflection are both small.

- **Specular Reflectance:** A specularly reflecting surface, such as that of a metal or mirror reflects the light according to the laws of reflection (i.e., the angle of reflection is equal to the angle of incidence). The reflectance from such a surface is known as *specular* reflection.

- **Hybrid Reflectance Model:** There exists another type of reflection, which are mostly found in display devices. These are known as *Hazes.* In the real world most of the surfaces we come across are neither Lambertian nor specular. They possess the combination of both the properties and are termed as *hybrid* surfaces. For example, the cathode-ray oscilloscopes may be considered as having considerable specular reflection and very low to moderate Lambertian reflection. The specular components of reflection from these surfaces may be reduced by using antireflection coatings on these surfaces. The reflectance models from such surfaces may be described as

$$I = wI_S + (1 - w)I_L,$$

where w is the weight of the specular component of the hybrid surface, and I_S and I_L are the specular and Lambertian intensities of the hybrid surface.

The problem of sun glint and glare assumes importance while working with optical imagery of water, snow, or even roads. This problem increases as the sun angle increases. This is due to the specular reflection of light from the object surface. In scenes containing water body the glint increases at high sun angle. At high sun angle much of the sunlight reaches the bottom of the water body and as a result the bottom of the water body gets illuminated and the potential for glint increases. The effect of glint depends on the sun angle, and also on the focal length and the field of view of the imaging devices. The glare is much more common over water, which has a much higher natural reflectance than vegetation. This can be seen on the waters, where the glare appears grayish-silver.

2.2.3 Point Spread Function

The basis of image formation can be explained by the *point spread function* (PSF). The PSF indicates how a point source of light results in a spread image in the spatial dimension.

Let us assume that we want to find the image of a single point at (x, y). If the imaging system is perfectly focused and without any stochastic disturbance, then all the photons from the point source will strike the detector focal plane at the same point and will produce a point image. However, the resultant image of this point source will interestingly not be a point, or a perfect copy of the point; but a blurred version of it. Usually the intensity at the center will be maximum and it will progressively reduce away from the center, resulting in a Gaussian distribution function. The blurring results from several factors – the blurring may be due to inappropriate focusing, imperfection of the lens, scatter of photons or the interaction of photons with the detector array. The resultant image is described in terms of its point spread function (PSF), as defined below

$$I_{res}(x, y) = I_{id}(x, y) \otimes P(x, y)$$

where $\otimes$ is the convolution operation, and I_{res} is the resultant image when the input image I_{id} is convolved with the point spread function $P(x, y)$ at location $(x,\ y)$. The width of the PSF decides the nature of the resultant image.

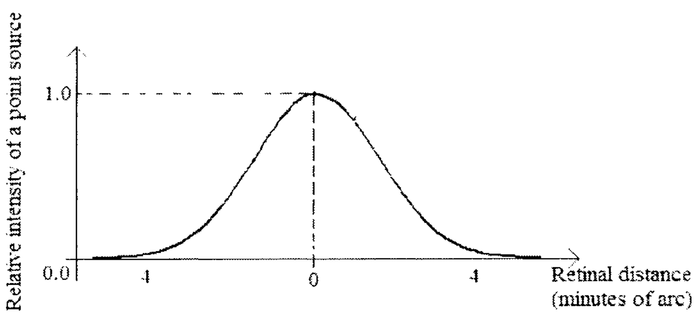


Fig. 2.2 Example of point spread function.

Thus if we know the point spread function, it is possible to restore the image by deconvolution. We know that the convolution in the time domain is analogous to the multiplication in the frequency domain. In the Fourier Transform domain

$$F\left(I_{res}(x, y)\right) = F\left(I_{id}(x, y)\right) \cdot F\left(P(x, y)\right)$$

or

$$F\left(I_{id}(x, y)\right) = \frac{F\left(I_{res}(x, y)\right)}{F\left(P(x, y)\right)},$$

where $F\left(f(x, y)\right)$ represents the Fourier transform of the two-dimensional image function $f(x, y)$. Thus given the Fourier transform of the resultant image along with the Fourier transform of the point spread function, we can reconstruct the original point object by taking the inverse transform of $F\left(I_{id}(x, y)\right)$.

Figure 2.2 shows the PSF of a typical imaging device. The width within which the PSF drops to half on both the sides of the center point is known as *full width half maximum* (FWHM). If now there are two points which are separated by a distance of FWHM or more, then the two points can be distinguished in the image. Otherwise the points will be indistinguishable in the image plane. This is shown in Figure 2.3. The PSF is not necessarily symmetrical and it may have different spreads in different directions.

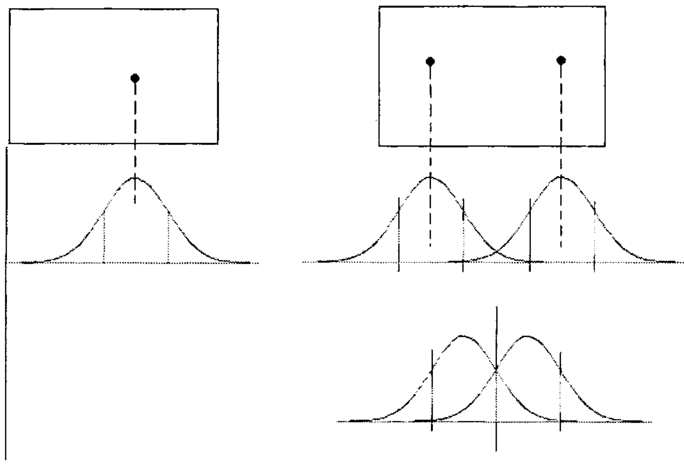


Fig. 2.3 Indistinguishability of point sources.

Often it is difficult to produce a perfect point source to measure the point spread function. In this case a line or edge is often used instead, giving the line spread function (LSF), or edge response function (ERF). The line spread function is a simple extension of the concept of PSF. As in case of a PSF, profiles can be generated orthogonally through the line image, and as in case of PSF, FWHM is used for defining the resolution.

2.3 SAMPLING AND QUANTIZATION

Understanding the process of *sampling and quantization* is one of the key areas in image processing. A comprehensive and detailed description of the theory may be found in [1, 2]. The phenomenal research of Shannon on the diverse aspects of communications in a noisy environment has led to the understanding of the process of sampling continuous signals [3]. The theories of image sampling and quantization have been investigated from two viewpoints. The two-dimensional images may be viewed as deterministic systems, where a continuous-valued image, representing the intensity or luminance at each point of the image, is sampled by an array of *Dirac-Delta* functions of infinite size. The results of sampling and reconstruction of such a deterministic image field may be found in [4]. In an alternative view images have been considered as samples of two-dimensional random processes. In this approach an image is viewed as a two-dimensional stationary random process with a certain mean and autocorrelation function. The practical images may always be viewed as the ideal image with additive noise, which is modelled as a random field. Sampling of such a two-dimensional random field model of images has been discussed in [5].

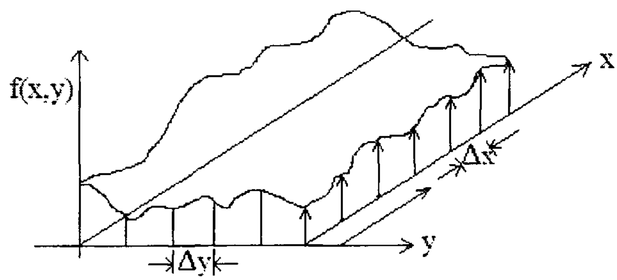


Fig. 2.4 Two-dimensional sampling array.

Let $f(x,y)$ be a continuous-valued intensity image and let $s(x,y)$ be a two-dimensional sampling function of the form

$$s(x,y) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(x - j\Delta x, y - k\Delta y.)$$

The two-dimensional sampling function is an infinite array of dirac delta functions as shown in Figure 2.4. The sampling function, also known as a comb function, is arranged in a regular grid of spacing Δx and Δy along X- and Y axes respectively. The sampled image may be represented as

$$\begin{aligned} f_s(x,y) &= f(x,y)s(x,y) \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(j\Delta x, k\Delta y) \cdot \delta(x - j\Delta x, y - k\Delta y) \end{aligned}$$

The sampled image $f_s(x,y)$ is an array of image intensity values at the sample points $(j\Delta x, k\Delta y)$ in a regular two-dimensional grid. Images may be sampled using rectangular and hexagonal lattice structures as shown in Figure 2.5. One of the important questions is how small Δx and Δy should be, so that we will be able to reconstruct the original image from the sampled image. The answer to this question lies in the Nyquist theorem, which states that a time varying signal should be sampled at a frequency which is at least twice of the maximum frequency component present in the signal. Comprehensive discussions may be found in [1, 2, 4, 6].

2.3.1 Image Sampling

A static image is a two-dimensional spatially varying signal. The sampling period, according to Nyquist criterion, should be smaller than or at the most

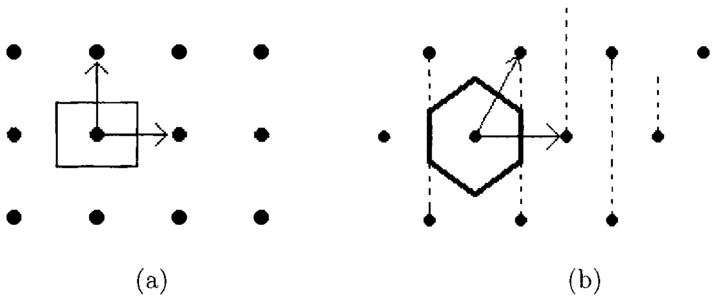


Fig. 2.5 (a) Rectangular and (b) hexagonal lattice structure of the sampling grid.

equal to half of the period of the finest detail present within an image. This implies that the sampling frequency along x axis $w_{xs} \geq 2w_x^L$ and along y axis $w_{ys} \geq 2w_y^L$, where w_x^L and w_y^L are the limiting factors of sampling along x and y directions. Since we have chosen sampling of Δx along X-axis and Δy along Y-axis, $\Delta x \leq \frac{\pi}{w_x^L}$ and $\Delta y \leq \frac{\pi}{w_y^L}$. The values of Δx and Δy should be chosen in such a way that the image is sampled at Nyquist frequency. If Δx and Δy values are smaller, the image is called oversampled, while if we choose large values of Δx and Δy the image will be undersampled. If the image is oversampled or exactly sampled, it is possible to reconstruct the bandlimited image. If the image is undersampled, then there will be spectral overlapping, which results in *aliasing effect*. We have shown images sampled at different spatial resolutions in Figure 2.6 to demonstrate that the aliasing effect increases as the sampling resolution decreases.

Fig. 2.6 Images sampled at 256 × 256, 128 × 128 , 64 × 64, 32 × 32, and 16 × 16 rectangular sampling grids.