

Digital Image Processing

Second Edition

Rafael C. Gonzalez

University of Tennessee

Richard E. Woods

MedData Interactive



Prentice Hall
Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data

Gonzalez, Rafael C.

Digital Image Processing / Richard E. Woods
p. cm.

Includes bibliographical references

ISBN 0-201-18075-8

1. Digital Imaging. 2. Digital Techniques. I. Title.

TA1632.G66 2001

621.3—dc21

2001035846

CIP

Vice-President and Editorial Director, ECS: *Marcia J. Horton*

Publisher: *Tom Robbins*

Associate Editor: *Alice Dworkin*

Editorial Assistant: *Jody McDonnell*

Vice President and Director of Production and Manufacturing, ESM: *David W. Riccardi*

Executive Managing Editor: *Vince O'Brien*

Managing Editor: *David A. George*

Production Editor: *Rose Kernan*

Composition: *Prepare, Inc.*

Director of Creative Services: *Paul Belfanti*

Creative Director: *Carole Anson*

Art Director and Cover Designer: *Heather Scott*

Art Editor: *Greg Dulles*

Manufacturing Manager: *Trudy Pisciotti*

Manufacturing Buyer: *Lisa McDowell*

Senior Marketing Manager: *Jennie Burger*



© 2002 by Prentice-Hall, Inc.
Upper Saddle River, New Jersey 07458

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN: 0-201-18075-8

Pearson Education Ltd., *London*

Pearson Education Australia Pty., Limited, *Sydney*

Pearson Education Singapore, Pte. Ltd.

Pearson Education North Asia Ltd., *Hong Kong*

Pearson Education Canada, Ltd., *Toronto*

Pearson Education de Mexico, S.A. de C.V.

Pearson Education—Japan, *Tokyo*

Pearson Education Malaysia, Pte. Ltd.

Pearson Education, *Upper Saddle River, New Jersey*

Preface

When something can be read without effort,
great effort has gone into its writing.
Enrique Jardiel Poncela

This edition is the most comprehensive revision of *Digital Image Processing* since the book first appeared in 1977. As the 1977 and 1987 editions by Gonzalez and Wintz, and the 1992 edition by Gonzalez and Woods, the present edition was prepared with students and instructors in mind. Thus, the principal objectives of the book continue to be to provide an introduction to basic concepts and methodologies for digital image processing, and to develop a foundation that can be used as the basis for further study and research in this field. To achieve these objectives, we again focused on material that we believe is fundamental and has a scope of application that is not limited to the solution of specialized problems. The mathematical complexity of the book remains at a level well within the grasp of college seniors and first-year graduate students who have introductory preparation in mathematical analysis, vectors, matrices, probability, statistics, and rudimentary computer programming.

The present edition was influenced significantly by a recent market survey conducted by Prentice Hall. The major findings of this survey were:

1. A need for more motivation in the introductory chapter regarding the spectrum of applications of digital image processing.
2. A simplification and shortening of material in the early chapters in order to “get to the subject matter” as quickly as possible.
3. A more intuitive presentation in some areas, such as image transforms and image restoration.
4. Individual chapter coverage of color image processing, wavelets, and image morphology.
5. An increase in the breadth of problems at the end of each chapter.

The reorganization that resulted in this edition is our attempt at providing a reasonable degree of balance between rigor in the presentation, the findings of the market survey, and suggestions made by students, readers, and colleagues since the last edition of the book. The major changes made in the book are as follows.

Chapter 1 was rewritten completely. The main focus of the current treatment is on examples of areas that use digital image processing. While far from exhaustive, the examples shown will leave little doubt in the reader’s mind regarding the breadth of application of digital image processing methodologies. Chapter 2 is totally new also. The focus of the presentation in this chapter is on how digital images are generated, and on the closely related concepts of

sampling, aliasing, Moiré patterns, and image zooming and shrinking. The new material and the manner in which these two chapters were reorganized address directly the first two findings in the market survey mentioned above.

Chapters 3 through 6 in the current edition cover the same concepts as Chapters 3 through 5 in the previous edition, but the scope is expanded and the presentation is totally different. In the previous edition, Chapter 3 was devoted exclusively to image transforms. One of the major changes in the book is that image transforms are now introduced when they are needed. This allowed us to begin discussion of image processing techniques much earlier than before, further addressing the second finding of the market survey. Chapters 3 and 4 in the current edition deal with image enhancement, as opposed to a single chapter (Chapter 4) in the previous edition. The new organization of this material does not imply that image enhancement is more important than other areas. Rather, we used it as an avenue to introduce spatial methods for image processing (Chapter 3), as well as the Fourier transform, the frequency domain, and image filtering (Chapter 4). Our purpose for introducing these concepts in the context of image enhancement (a subject particularly appealing to beginners) was to increase the level of intuitiveness in the presentation, thus addressing partially the third major finding in the marketing survey. This organization also gives instructors flexibility in the amount of frequency-domain material they wish to cover.

Chapter 5 also was rewritten completely in a more intuitive manner. The coverage of this topic in earlier editions of the book was based on matrix theory. Although unified and elegant, this type of presentation is difficult to follow, particularly by undergraduates. The new presentation covers essentially the same ground, but the discussion does not rely on matrix theory and is much easier to understand, due in part to numerous new examples. The price paid for this newly gained simplicity is the loss of a unified approach, in the sense that in the earlier treatment a number of restoration results could be derived from one basic formulation. On balance, however, we believe that readers (especially beginners) will find the new treatment much more appealing and easier to follow. Also, as indicated below, the old material is stored in the book Web site for easy access by individuals preferring to follow a matrix-theory formulation.

Chapter 6 dealing with color image processing is new. Interest in this area has increased significantly in the past few years as a result of growth in the use of digital images for Internet applications. Our treatment of this topic represents a significant expansion of the material from previous editions. Similarly Chapter 7, dealing with wavelets, is new. In addition to a number of signal processing applications, interest in this area is motivated by the need for more sophisticated methods for image compression, a topic that in turn is motivated by a increase in the number of images transmitted over the Internet or stored in Web servers. Chapter 8 dealing with image compression was updated to include new compression methods and standards, but its fundamental structure remains the same as in the previous edition. Several image transforms, previously covered in Chapter 3 and whose principal use is compression, were moved to this chapter.

Chapter 9, dealing with image morphology, is new. It is based on a significant expansion of the material previously included as a section in the chapter on image representation and description. Chapter 10, dealing with image segmentation, has the same basic structure as before, but numerous new examples were included and a new section on segmentation by morphological watersheds was added. Chapter 11, dealing with image representation and description, was shortened slightly by the removal of the material now included in Chapter 9. New examples were added and the Hotelling transform (description by principal components), previously included in Chapter 3, was moved to this chapter. Chapter 12 dealing with object recognition was shortened by the removal of topics dealing with knowledge-based image analysis, a topic now covered in considerable detail in a number of books which we reference in Chapters 1 and 12. Experience since the last edition of *Digital Image Processing* indicates that the new, shortened coverage of object recognition is a logical place at which to conclude the book.

Although the book is totally self-contained, we have established a companion web site (see inside front cover) designed to provide support to users of the book. For students following a formal course of study or individuals embarked on a program of self study, the site contains a number of tutorial reviews on background material such as probability, statistics, vectors, and matrices, prepared at a basic level and written using the same notation as in the book. Detailed solutions to many of the exercises in the book also are provided. For instruction, the site contains suggested teaching outlines, classroom presentation materials, laboratory experiments, and various image databases (including most images from the book). In addition, part of the material removed from the previous edition is stored in the Web site for easy download and classroom use, at the discretion of the instructor. A downloadable instructor's manual containing sample curricula, solutions to sample laboratory experiments, and solutions to all problems in the book is available to instructors who have adopted the book for classroom use.

This edition of *Digital Image Processing* is a reflection of the significant progress that has been made in this field in just the past decade. As is usual in a project such as this, progress continues after work on the manuscript stops. One of the reasons earlier versions of this book have been so well accepted throughout the world is their emphasis on fundamental concepts, an approach that, among other things, attempts to provide a measure of constancy in a rapidly-evolving body of knowledge. We have tried to observe that same principle in preparing this edition of the book.

R.C.G.
R.E.W.

Contents

Preface xv

Acknowledgements xviii

About the Authors xix

1 *Introduction* 15

- 1.1 **What Is Digital Image Processing?** 15
- 1.2 **The Origins of Digital Image Processing** 17
- 1.3 **Examples of Fields that Use Digital Image Processing** 21
 - 1.3.1 Gamma-Ray Imaging 22
 - 1.3.2 X-ray Imaging 23
 - 1.3.3 Imaging in the Ultraviolet Band 25
 - 1.3.4 Imaging in the Visible and Infrared Bands 26
 - 1.3.5 Imaging in the Microwave Band 32
 - 1.3.6 Imaging in the Radio Band 34
 - 1.3.7 Examples in which Other Imaging Modalities Are Used 34
- 1.4 **Fundamental Steps in Digital Image Processing** 39
- 1.5 **Components of an Image Processing System** 42
 - Summary 44
 - References and Further Reading 45

2 *Digital Image Fundamentals* 34

- 2.1 **Elements of Visual Perception** 34
 - 2.1.1 Structure of the Human Eye 35
 - 2.1.2 **Image Formation in the Eye** 37
 - 2.1.3 Brightness Adaptation and Discrimination 38
- 2.2 **Light and the Electromagnetic Spectrum** 42
- 2.3 **Image Sensing and Acquisition** 45
 - 2.3.1 Image Acquisition Using a Single Sensor 47
 - 2.3.2 Image Acquisition Using Sensor Strips 48
 - 2.3.3 Image Acquisition Using Sensor Arrays 49
 - 2.3.4 A Simple Image Formation Model 50
- 2.4 **Image Sampling and Quantization** 52
 - 2.4.1 Basic Concepts in Sampling and Quantization 52
 - 2.4.2 Representing Digital Images 54
 - 2.4.3 Spatial and Gray-Level Resolution 57
 - 2.4.4 Aliasing and Moiré Patterns 62
 - 2.4.5 Zooming and Shrinking Digital Images 64

2.5	Some Basic Relationships Between Pixels	66
2.5.1	Neighbors of a Pixel	66
2.5.2	Adjacency, Connectivity, Regions, and Boundaries	66
2.5.3	Distance Measures	68
2.5.4	Image Operations on a Pixel Basis	69
2.6	Linear and Nonlinear Operations	70
	Summary	70
	References and Further Reading	70
	Problems	71

3 *Image Enhancement in the Spatial Domain* 75

3.1	Background	76
3.2	Some Basic Gray Level Transformations	78
3.2.1	Image Negatives	78
3.2.2	Log Transformations	79
3.2.3	Power-Law Transformations	80
3.2.4	Piecewise-Linear Transformation Functions	85
3.3	Histogram Processing	88
3.3.1	Histogram Equalization	91
3.3.2	Histogram Matching (Specification)	94
3.3.3	Local Enhancement	103
3.3.4	Use of Histogram Statistics for Image Enhancement	103
3.4	Enhancement Using Arithmetic/Logic Operations	108
3.4.1	Image Subtraction	110
3.4.2	Image Averaging	112
3.5	Basics of Spatial Filtering	116
3.6	Smoothing Spatial Filters	119
3.6.1	Smoothing Linear Filters	119
3.6.2	Order-Statistics Filters	123
3.7	Sharpening Spatial Filters	125
3.7.1	Foundation	125
3.7.2	Use of Second Derivatives for Enhancement—The Laplacian	128
3.7.3	Use of First Derivatives for Enhancement—The Gradient	134
3.8	Combining Spatial Enhancement Methods	137
	Summary	141
	References and Further Reading	142
	Problems	142

4 *Image Enhancement in the Frequency Domain* 147

4.1	Background	148
------------	-------------------	------------

4.2	Introduction to the Fourier Transform and the Frequency Domain	149
4.2.1	The One-Dimensional Fourier Transform and its Inverse	150
4.2.2	The Two-Dimensional DFT and Its Inverse	154
4.2.3	Filtering in the Frequency Domain	156
4.2.4	Correspondence between Filtering in the Spatial and Frequency Domains	161
4.3	Smoothing Frequency-Domain Filters	167
4.3.1	Ideal Lowpass Filters	167
4.3.2	Butterworth Lowpass Filters	173
4.3.3	Gaussian Lowpass Filters	175
4.3.4	Additional Examples of Lowpass Filtering	178
4.4	Sharpening Frequency Domain Filters	180
4.4.1	Ideal Highpass Filters	182
4.4.2	Butterworth Highpass Filters	183
4.4.3	Gaussian Highpass Filters	184
4.4.4	The Laplacian in the Frequency Domain	185
4.4.5	Unsharp Masking, High-Boost Filtering, and High-Frequency Emphasis Filtering	187
4.5	Homomorphic Filtering	191
4.6	Implementation	194
4.6.1	Some Additional Properties of the 2-D Fourier Transform	194
4.6.2	Computing the Inverse Fourier Transform Using a Forward Transform Algorithm	198
4.6.3	More on Periodicity: the Need for Padding	199
4.6.4	The Convolution and Correlation Theorems	205
4.6.5	Summary of Properties of the 2-D Fourier Transform	208
4.6.6	The Fast Fourier Transform	208
4.6.7	Some Comments on Filter Design	213
Summary 214		
References 214		
Problems 215		

5 *Image Restoration* 220

5.1	A Model of the Image Degradation/Restoration Process	221
5.2	Noise Models	222
5.2.1	Spatial and Frequency Properties of Noise	222
5.2.2	Some Important Noise Probability Density Functions	222
5.2.3	Periodic Noise	227
5.2.4	Estimation of Noise Parameters	227
5.3	Restoration in the Presence of Noise Only–Spatial Filtering	230
5.3.1	Mean Filters	231
5.3.2	Order-Statistics Filters	233
5.3.3	Adaptive Filters	237

5.4	Periodic Noise Reduction by Frequency Domain Filtering	243
5.4.1	Bandreject Filters	244
5.4.2	Bandpass Filters	245
5.4.3	Notch Filters	246
5.4.4	Optimum Notch Filtering	248
5.5	Linear, Position-Invariant Degradations	254
5.6	Estimating the Degradation Function	256
5.6.1	Estimation by Image Observation	256
5.6.2	Estimation by Experimentation	257
5.6.3	Estimation by Modeling	258
5.7	Inverse Filtering	261
5.8	Minimum Mean Square Error (Wiener) Filtering	262
5.9	Constrained Least Squares Filtering	266
5.10	Geometric Mean Filter	270
5.11	Geometric Transformations	270
5.11.1	Spatial Transformations	271
5.11.2	Gray-Level Interpolation	272
	Summary	276
	References and Further Reading	277
	Problems	278

6 *Color Image Processing* 282

6.1	Color Fundamentals	283
6.2	Color Models	289
6.2.1	The RGB Color Model	290
6.2.2	The CMY and CMYK Color Models	294
6.2.3	The HSI Color Model	295
6.3	Pseudocolor Image Processing	302
6.3.1	Intensity Slicing	303
6.3.2	Gray Level to Color Transformations	308
6.4	Basics of Full-Color Image Processing	313
6.5	Color Transformations	315
6.5.1	Formulation	315
6.5.2	Color Complements	318
6.5.3	Color Slicing	320
6.5.4	Tone and Color Corrections	322
6.5.5	Histogram Processing	326
6.6	Smoothing and Sharpening	327
6.6.1	Color Image Smoothing	328
6.6.2	Color Image Sharpening	330
6.7	Color Segmentation	331
6.7.1	Segmentation in HSI Color Space	331
6.7.2	Segmentation in RGB Vector Space	333
6.7.3	Color Edge Detection	335

6.8	Noise in Color Images	339
6.9	Color Image Compression	342
	Summary	343
	References and Further Reading	344
	Problems	344

7 Wavelets and Multiresolution Processing 349

7.1	Background	350
7.1.1	Image Pyramids	351
7.1.2	Subband Coding	354
7.1.3	The Haar Transform	360
7.2	Multiresolution Expansions	363
7.2.1	Series Expansions	364
7.2.2	Scaling Functions	365
7.2.3	Wavelet Functions	369
7.3	Wavelet Transforms in One Dimension	372
7.3.1	The Wavelet Series Expansions	372
7.3.2	The Discrete Wavelet Transform	375
7.3.3	The Continuous Wavelet Transform	376
7.4	The Fast Wavelet Transform	379
7.5	Wavelet Transforms in Two Dimensions	386
7.6	Wavelet Packets	394
	Summary	402
	References and Further Reading	404
	Problems	404

8 Image Compression 409

8.1	Fundamentals	411
8.1.1	Coding Redundancy	412
8.1.2	Interpixel Redundancy	414
8.1.3	Psychovisual Redundancy	417
8.1.4	Fidelity Criteria	419
8.2	Image Compression Models	421
8.2.1	The Source Encoder and Decoder	421
8.2.2	The Channel Encoder and Decoder	423
8.3	Elements of Information Theory	424
8.3.1	Measuring Information	424
8.3.2	The Information Channel	425
8.3.3	Fundamental Coding Theorems	430
8.3.4	Using Information Theory	437
8.4	Error-Free Compression	440
8.4.1	Variable-Length Coding	440

8.4.2	LZW Coding	446
8.4.3	Bit-Plane Coding	448
8.4.4	Lossless Predictive Coding	456
8.5	Lossy Compression	459
8.5.1	Lossy Predictive Coding	459
8.5.2	Transform Coding	467
8.5.3	Wavelet Coding	486
8.6	Image Compression Standards	492
8.6.1	Binary Image Compression Standards	493
8.6.2	Continuous Tone Still Image Compression Standards	498
8.6.3	Video Compression Standards	510
	Summary	513
	References and Further Reading	513
	Problems	514

9

Morphological Image Processing 519

9.1	Preliminaries	520
9.1.1	Some Basic Concepts from Set Theory	520
9.1.2	Logic Operations Involving Binary Images	522
9.2	Dilation and Erosion	523
9.2.1	Dilation	523
9.2.2	Erosion	525
9.3	Opening and Closing	528
9.4	The Hit-or-Miss Transformation	532
9.5	Some Basic Morphological Algorithms	534
9.5.1	Boundary Extraction	534
9.5.2	Region Filling	535
9.5.3	Extraction of Connected Components	536
9.5.4	Convex Hull	539
9.5.5	Thinning	541
9.5.6	Thickening	541
9.5.7	Skeletons	543
9.5.8	Pruning	545
9.5.9	Summary of Morphological Operations on Binary Images	547
9.6	Extensions to Gray-Scale Images	550
9.6.1	Dilation	550
9.6.2	Erosion	552
9.6.3	Opening and Closing	554
9.6.4	Some Applications of Gray-Scale Morphology	556
	Summary	560
	References and Further Reading	560
	Problems	560

10 *Image Segmentation* 567

10.1 Detection of Discontinuities 568

10.1.1 Point Detection 569

10.1.2 Line Detection 570

10.1.3 Edge Detection 572

10.2 Edge Linking and Boundary Detection 585

10.2.1 Local Processing 585

10.2.2 Global Processing via the Hough Transform 587

10.2.3 Global Processing via Graph-Theoretic Techniques 591

10.3 Thresholding 595

10.3.1 Foundation 595

10.3.2 The Role of Illumination 596

10.3.3 Basic Global Thresholding 598

10.3.4 Basic Adaptive Thresholding 600

10.3.5 Optimal Global and Adaptive Thresholding 602

10.3.6 Use of Boundary Characteristics for Histogram Improvement
and Local Thresholding 608

10.3.7 Thresholds Based on Several Variables 611

10.4 Region-Based Segmentation 612

10.4.1 Basic Formulation 612

10.4.2 Region Growing 613

10.4.3 Region Splitting and Merging 615

10.5 Segmentation by Morphological Watersheds 617

10.5.1 Basic Concepts 617

10.5.2 Dam Construction 620

10.5.3 Watershed Segmentation Algorithm 622

10.5.4 The Use of Markers 624

10.6 The Use of Motion in Segmentation 626

10.6.1 Spatial Techniques 626

10.6.2 Frequency Domain Techniques 630

Summary 634

References and Further Reading 634

Problems 636

11 *Representation and Description* 643

11.1 Representation 644

11.1.1 Chain Codes 644

11.1.2 Polygonal Approximations 646

11.1.3 Signatures 648

11.1.4 Boundary Segments 649

11.1.5 Skeletons 650

11.2 Boundary Descriptors	653
11.2.1 Some Simple Descriptors	653
11.2.2 Shape Numbers	654
11.2.3 Fourier Descriptors	655
11.2.4 Statistical Moments	659
11.3 Regional Descriptors	660
11.3.1 Some Simple Descriptors	661
11.3.2 Topological Descriptors	661
11.3.3 Texture	665
11.3.4 Moments of Two-Dimensional Functions	672
11.4 Use of Principal Components for Description	675
11.5 Relational Descriptors	683
Summary	687
References and Further Reading	687
Problems	689

12 *Object Recognition* 693

12.1 Patterns and Pattern Classes	693
12.2 Recognition Based on Decision-Theoretic Methods	698
12.2.1 Matching	698
12.2.2 Optimum Statistical Classifiers	704
12.2.3 Neural Networks	712
12.3 Structural Methods	732
12.3.1 Matching Shape Numbers	732
12.3.2 String Matching	734
12.3.3 Syntactic Recognition of Strings	735
12.3.4 Syntactic Recognition of Trees	740
Summary	750
References and Further Reading	750
Problems	750
<i>Bibliography</i>	755
<i>Index</i>	779

1

Introduction

One picture is worth more than ten thousand words.

Anonymous

Preview

Interest in digital image processing methods stems from two principal application areas: improvement of pictorial information for human interpretation; and processing of image data for storage, transmission, and representation for autonomous machine perception. This chapter has several objectives: (1) to define the scope of the field that we call image processing; (2) to give a historical perspective of the origins of this field; (3) to give an idea of the state of the art in image processing by examining some of the principal areas in which it is applied; (4) to discuss briefly the principal approaches used in digital image processing; (5) to give an overview of the components contained in a typical, general-purpose image processing system; and (6) to provide direction to the books and other literature where image processing work normally is reported.

1.1 What Is Digital Image Processing?

An image may be defined as a two-dimensional function, $f(x, y)$, where x and y are *spatial* (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the *intensity* or *gray level* of the image at that point. When x , y , and the amplitude values of f are all finite, discrete quantities, we call the image a *digital image*. The field of *digital image processing* refers to processing digital images by means of a digital computer. Note that a digital image is composed of a finite number of elements, each of which has a particular location and

value. These elements are referred to as *picture elements*, *image elements*, *pels*, and *pixels*. *Pixel* is the term most widely used to denote the elements of a digital image. We consider these definitions in more formal terms in Chapter 2.

Vision is the most advanced of our senses, so it is not surprising that images play the single most important role in human perception. However, unlike humans, who are limited to the visual band of the electromagnetic (EM) spectrum, imaging machines cover almost the entire EM spectrum, ranging from gamma to radio waves. They can operate on images generated by sources that humans are not accustomed to associating with images. These include ultrasound, electron microscopy, and computer-generated images. Thus, digital image processing encompasses a wide and varied field of applications.

There is no general agreement among authors regarding where image processing stops and other related areas, such as image analysis and computer vision, start. Sometimes a distinction is made by defining image processing as a discipline in which both the input and output of a process are images. We believe this to be a limiting and somewhat artificial boundary. For example, under this definition, even the trivial task of computing the average intensity of an image (which yields a single number) would not be considered an image processing operation. On the other hand, there are fields such as computer vision whose ultimate goal is to use computers to emulate human vision, including learning and being able to make inferences and take actions based on visual inputs. This area itself is a branch of artificial intelligence (AI) whose objective is to emulate human intelligence. The field of AI is in its earliest stages of infancy in terms of development, with progress having been much slower than originally anticipated. The area of image analysis (also called image understanding) is in between image processing and computer vision.

There are no clear-cut boundaries in the continuum from image processing at one end to computer vision at the other. However, one useful paradigm is to consider three types of computerized processes in this continuum: low-, mid-, and high-level processes. Low-level processes involve primitive operations such as image preprocessing to reduce noise, contrast enhancement, and image sharpening. A low-level process is characterized by the fact that both its inputs and outputs are images. Mid-level processing on images involves tasks such as segmentation (partitioning an image into regions or objects), description of those objects to reduce them to a form suitable for computer processing, and classification (recognition) of individual objects. A mid-level process is characterized by the fact that its inputs generally are images, but its outputs are attributes extracted from those images (e.g., edges, contours, and the identity of individual objects). Finally, higher-level processing involves “making sense” of an ensemble of recognized objects, as in image analysis, and, at the far end of the continuum, performing the cognitive functions normally associated with vision.

Based on the preceding comments, we see that a logical place of overlap between image processing and image analysis is the area of recognition of individual regions or objects in an image. Thus, what we call in this book *digital image processing* encompasses processes whose inputs and outputs are images

and, in addition, encompasses processes that extract attributes from images, up to and including the recognition of individual objects. As a simple illustration to clarify these concepts, consider the area of automated analysis of text. The processes of acquiring an image of the area containing the text, preprocessing that image, extracting (segmenting) the individual characters, describing the characters in a form suitable for computer processing, and recognizing those individual characters are in the scope of what we call digital image processing in this book. Making sense of the content of the page may be viewed as being in the domain of image analysis and even computer vision, depending on the level of complexity implied by the statement “making sense.” As will become evident shortly, digital image processing, as we have defined it, is used successfully in a broad range of areas of exceptional social and economic value. The concepts developed in the following chapters are the foundation for the methods used in those application areas.

1.2 The Origins of Digital Image Processing

One of the first applications of digital images was in the newspaper industry, when pictures were first sent by submarine cable between London and New York. Introduction of the Bartlane cable picture transmission system in the early 1920s reduced the time required to transport a picture across the Atlantic from more than a week to less than three hours. Specialized printing equipment coded pictures for cable transmission and then reconstructed them at the receiving end. Figure 1.1 was transmitted in this way and reproduced on a telegraph printer fitted with typefaces simulating a halftone pattern.

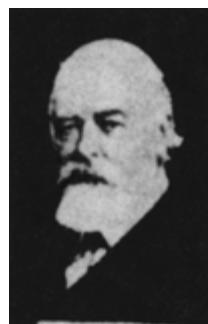
Some of the initial problems in improving the visual quality of these early digital pictures were related to the selection of printing procedures and the distribution of intensity levels. The printing method used to obtain Fig. 1.1 was abandoned toward the end of 1921 in favor of a technique based on photographic reproduction made from tapes perforated at the telegraph receiving terminal. Figure 1.2 shows an image obtained using this method. The improvements over Fig. 1.1 are evident, both in tonal quality and in resolution.



FIGURE 1.1 A digital picture produced in 1921 from a coded tape by a telegraph printer with special type faces. (McFarlane.[†])

[†]References in the Bibliography at the end of the book are listed in alphabetical order by authors' last names.

FIGURE 1.2 A digital picture made in 1922 from a tape punched after the signals had crossed the Atlantic twice. Some errors are visible. (McFarlane.)



The early Bartlane systems were capable of coding images in five distinct levels of gray. This capability was increased to 15 levels in 1929. Figure 1.3 is typical of the type of images that could be obtained using the 15-tone equipment. During this period, introduction of a system for developing a film plate via light beams that were modulated by the coded picture tape improved the reproduction process considerably.

Although the examples just cited involve digital images, they are not considered digital image processing results in the context of our definition because computers were not involved in their creation. Thus, the history of digital image processing is intimately tied to the development of the digital computer. In fact, digital images require so much storage and computational power that progress in the field of digital image processing has been dependent on the development of digital computers and of supporting technologies that include data storage, display, and transmission.

The idea of a computer goes back to the invention of the abacus in Asia Minor, more than 5000 years ago. More recently, there were developments in the past two centuries that are the foundation of what we call a computer today. However, the basis for what we call a *modern* digital computer dates back to only the 1940s with the introduction by John von Neumann of two key concepts: (1) a memory to hold a stored program and data, and (2) conditional branching. These two ideas are the foundation of a central processing unit (CPU), which is at the heart of computers today. Starting with von Neumann, there were

FIGURE 1.3 Unretouched cable picture of Generals Pershing and Foch, transmitted in 1929 from London to New York by 15-tone equipment. (McFarlane.)



a series of key advances that led to computers powerful enough to be used for digital image processing. Briefly, these advances may be summarized as follows: (1) the invention of the transistor by Bell Laboratories in 1948; (2) the development in the 1950s and 1960s of the high-level programming languages COBOL (Common Business-Oriented Language) and FORTRAN (Formula Translator); (3) the invention of the integrated circuit (IC) at Texas Instruments in 1958; (4) the development of operating systems in the early 1960s; (5) the development of the microprocessor (a single chip consisting of the central processing unit, memory, and input and output controls) by Intel in the early 1970s; (6) introduction by IBM of the personal computer in 1981; and (7) progressive miniaturization of components, starting with large scale integration (LSI) in the late 1970s, then very large scale integration (VLSI) in the 1980s, to the present use of ultra large scale integration (ULSI). Concurrent with these advances were developments in the areas of mass storage and display systems, both of which are fundamental requirements for digital image processing.

The first computers powerful enough to carry out meaningful image processing tasks appeared in the early 1960s. The birth of what we call digital image processing today can be traced to the availability of those machines and the onset of the space program during that period. It took the combination of those two developments to bring into focus the potential of digital image processing concepts. Work on using computer techniques for improving images from a space probe began at the Jet Propulsion Laboratory (Pasadena, California) in 1964 when pictures of the moon transmitted by *Ranger 7* were processed by a computer to correct various types of image distortion inherent in the on-board television camera. Figure 1.4 shows the first image of the moon taken by *Ranger 7* on July 31, 1964 at 9:09 A.M. Eastern Daylight Time (EDT), about 17 minutes before impacting the lunar surface (the markers, called *reseau* marks, are used for geometric corrections, as discussed in Chapter 5). This also is the first image of the moon taken by a U.S. spacecraft. The imaging lessons learned with *Ranger 7* served as the basis for improved methods used to enhance and restore images from the Surveyor missions to the moon, the Mariner series of flyby missions to Mars, the Apollo manned flights to the moon, and others.



FIGURE 1.4 The first picture of the moon by a U.S. spacecraft. *Ranger 7* took this image on July 31, 1964 at 9:09 A.M. EDT, about 17 minutes before impacting the lunar surface. (Courtesy of NASA.)

In parallel with space applications, digital image processing techniques began in the late 1960s and early 1970s to be used in medical imaging, remote Earth resources observations, and astronomy. The invention in the early 1970s of computerized axial tomography (CAT), also called computerized tomography (CT) for short, is one of the most important events in the application of image processing in medical diagnosis. Computerized axial tomography is a process in which a ring of detectors encircles an object (or patient) and an X-ray source, concentric with the detector ring, rotates about the object. The X-rays pass through the object and are collected at the opposite end by the corresponding detectors in the ring. As the source rotates, this procedure is repeated. Tomography consists of algorithms that use the sensed data to construct an image that represents a "slice" through the object. Motion of the object in a direction perpendicular to the ring of detectors produces a set of such slices, which constitute a three-dimensional (3-D) rendition of the inside of the object. Tomography was invented independently by Sir Godfrey N. Hounsfield and Professor Allan M. Cormack, who shared the 1979 Nobel Prize in Medicine for their invention. It is interesting to note that X-rays were discovered in 1895 by Wilhelm Conrad Roentgen, for which he received the 1901 Nobel Prize for Physics. These two inventions, nearly 100 years apart, led to some of the most active application areas of image processing today.

From the 1960s until the present, the field of image processing has grown vigorously. In addition to applications in medicine and the space program, digital image processing techniques now are used in a broad range of applications. Computer procedures are used to enhance the contrast or code the intensity levels into color for easier interpretation of X-rays and other images used in industry, medicine, and the biological sciences. Geographers use the same or similar techniques to study pollution patterns from aerial and satellite imagery. Image enhancement and restoration procedures are used to process degraded images of unrecoverable objects or experimental results too expensive to duplicate. In archeology, image processing methods have successfully restored blurred pictures that were the only available records of rare artifacts lost or damaged after being photographed. In physics and related fields, computer techniques routinely enhance images of experiments in areas such as high-energy plasmas and electron microscopy. Similarly successful applications of image processing concepts can be found in astronomy, biology, nuclear medicine, law enforcement, defense, and industrial applications.

These examples illustrate processing results intended for human interpretation. The second major area of application of digital image processing techniques mentioned at the beginning of this chapter is in solving problems dealing with machine perception. In this case, interest focuses on procedures for extracting from an image information in a form suitable for computer processing. Often, this information bears little resemblance to visual features that humans use in interpreting the content of an image. Examples of the type of information used in machine perception are statistical moments, Fourier transform coefficients, and multidimensional distance measures. Typical problems in machine perception that routinely utilize image processing techniques are automatic character recognition, industrial machine vision for product assembly and inspection, military recognizance, automatic processing of fingerprints, screening of X-rays and blood samples, and machine processing of aerial and satellite imagery for weather

prediction and environmental assessment. The continuing decline in the ratio of computer price to performance and the expansion of networking and communication bandwidth via the World Wide Web and the Internet have created unprecedented opportunities for continued growth of digital image processing. Some of these application areas are illustrated in the following section.

1.3 Examples of Fields that Use Digital Image Processing

Today, there is almost no area of technical endeavor that is not impacted in some way by digital image processing. We can cover only a few of these applications in the context and space of the current discussion. However, limited as it is, the material presented in this section will leave no doubt in the reader's mind regarding the breadth and importance of digital image processing. We show in this section numerous areas of application, each of which routinely utilizes the digital image processing techniques developed in the following chapters. Many of the images shown in this section are used later in one or more of the examples given in the book. All images shown are digital.

The areas of application of digital image processing are so varied that some form of organization is desirable in attempting to capture the breadth of this field. One of the simplest ways to develop a basic understanding of the extent of image processing applications is to categorize images according to their source (e.g., visual, X-ray, and so on). The principal energy source for images in use today is the electromagnetic energy spectrum. Other important sources of energy include acoustic, ultrasonic, and electronic (in the form of electron beams used in electron microscopy). Synthetic images, used for modeling and visualization, are generated by computer. In this section we discuss briefly how images are generated in these various categories and the areas in which they are applied. Methods for converting images into digital form are discussed in the next chapter.

Images based on radiation from the EM spectrum are the most familiar, especially images in the X-ray and visual bands of the spectrum. Electromagnetic waves can be conceptualized as propagating sinusoidal waves of varying wavelengths, or they can be thought of as a stream of massless particles, each traveling in a wavelike pattern and moving at the speed of light. Each massless particle contains a certain amount (or bundle) of energy. Each bundle of energy is called a *photon*. If spectral bands are grouped according to energy per photon, we obtain the spectrum shown in Fig. 1.5, ranging from gamma rays (highest energy) at one end to radio waves (lowest energy) at the other. The bands are shown shaded to convey the fact that bands of the EM spectrum are not distinct but rather transition smoothly from one to the other.

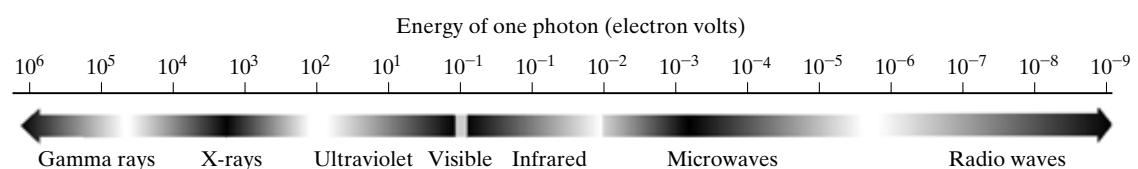


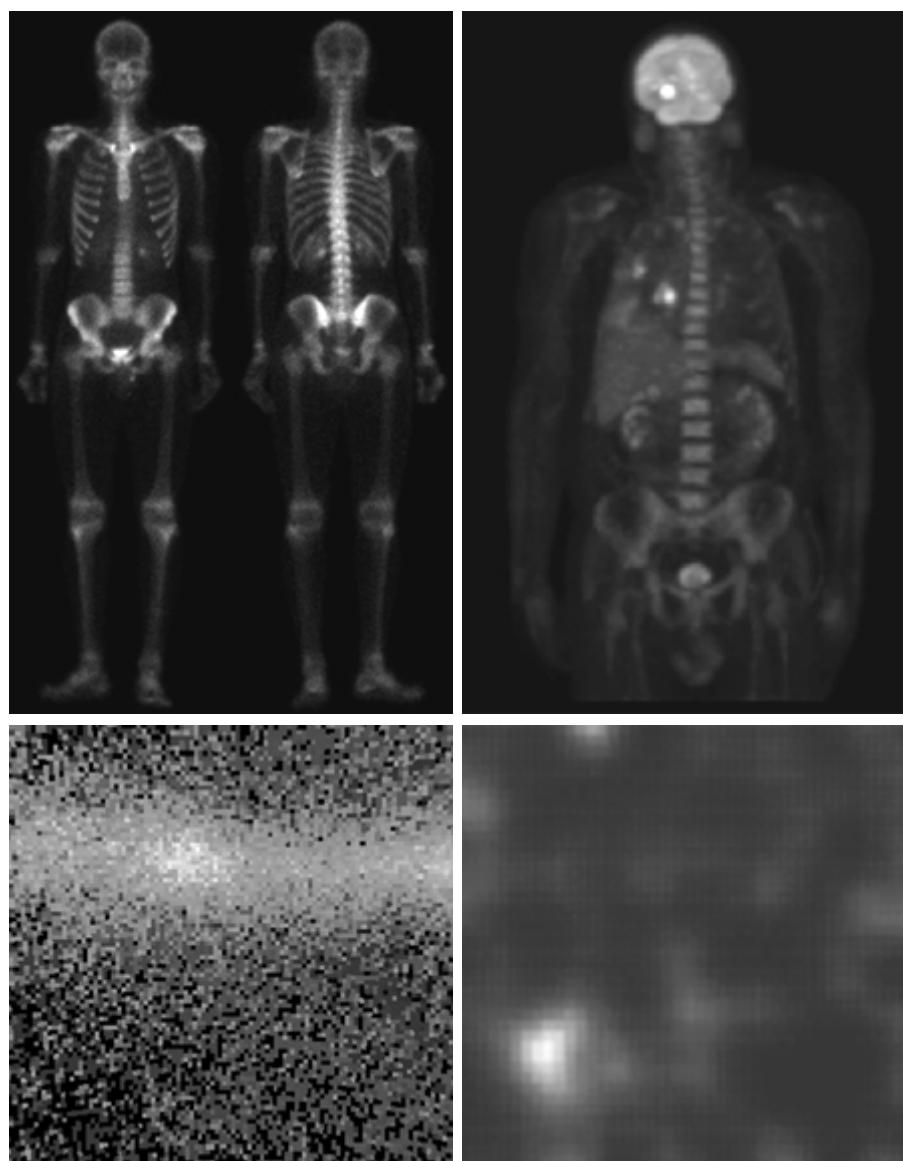
FIGURE 1.5 The electromagnetic spectrum arranged according to energy per photon.

1.3.1 Gamma-Ray Imaging

Major uses of imaging based on gamma rays include nuclear medicine and astronomical observations. In nuclear medicine, the approach is to inject a patient with a radioactive isotope that emits gamma rays as it decays. Images are produced from the emissions collected by gamma ray detectors. Figure 1.6(a) shows an image of a complete bone scan obtained by using gamma-ray imaging. Images of this sort are used to locate sites of bone pathology, such as infections or tumors. Figure 1.6(b) shows another major modality of nuclear imaging called positron emission tomography (PET). The principle is the same

a b
c d

FIGURE 1.6
Examples of gamma-ray imaging. (a) Bone scan. (b) PET image. (c) Cygnus Loop. (d) Gamma radiation (bright spot) from a reactor valve.
(Images courtesy of (a) G.E. Medical Systems, (b) Dr. Michael E. Casey, CTI PET Systems, (c) NASA, (d) Professors Zhong He and David K. Wehe, University of Michigan.)



as with X-ray tomography, mentioned briefly in Section 1.2. However, instead of using an external source of X-ray energy, the patient is given a radioactive isotope that emits positrons as it decays. When a positron meets an electron, both are annihilated and two gamma rays are given off. These are detected and a tomographic image is created using the basic principles of tomography. The image shown in Fig. 1.6(b) is one sample of a sequence that constitutes a 3-D rendition of the patient. This image shows a tumor in the brain and one in the lung, easily visible as small white masses.

A star in the constellation of Cygnus exploded about 15,000 years ago, generating a superheated stationary gas cloud (known as the Cygnus Loop) that glows in a spectacular array of colors. Figure 1.6(c) shows the Cygnus Loop imaged in the gamma-ray band. Unlike the two examples shown in Figs. 1.6(a) and (b), this image was obtained using the natural radiation of the object being imaged. Finally, Fig. 1.6(d) shows an image of gamma radiation from a valve in a nuclear reactor. An area of strong radiation is seen in the lower, left side of the image.

1.3.2 X-ray Imaging

X-rays are among the oldest sources of EM radiation used for imaging. The best known use of X-rays is medical diagnostics, but they also are used extensively in industry and other areas, like astronomy. X-rays for medical and industrial imaging are generated using an X-ray tube, which is a vacuum tube with a cathode and anode. The cathode is heated, causing free electrons to be released. These electrons flow at high speed to the positively charged anode. When the electrons strike a nucleus, energy is released in the form of X-ray radiation. The energy (penetrating power) of the X-rays is controlled by a voltage applied across the anode, and the number of X-rays is controlled by a current applied to the filament in the cathode. Figure 1.7(a) shows a familiar chest X-ray generated simply by placing the patient between an X-ray source and a film sensitive to X-ray energy. The intensity of the X-rays is modified by absorption as they pass through the patient, and the resulting energy falling on the film develops it, much in the same way that light develops photographic film. In digital radiography, digital images are obtained by one of two methods: (1) by digitizing X-ray films; or (2) by having the X-rays that pass through the patient fall directly onto devices (such as a phosphor screen) that convert X-rays to light. The light signal in turn is captured by a light-sensitive digitizing system. We discuss digitization in detail in Chapter 2.

Angiography is another major application in an area called contrast-enhancement radiography. This procedure is used to obtain images (called *angiograms*) of blood vessels. A catheter (a small, flexible, hollow tube) is inserted, for example, into an artery or vein in the groin. The catheter is threaded into the blood vessel and guided to the area to be studied. When the catheter reaches the site under investigation, an X-ray contrast medium is injected through the catheter. This enhances contrast of the blood vessels and enables the radiologist to see any irregularities or blockages. Figure 1.7(b) shows an example of an aortic angiogram. The catheter can be seen being inserted into the large blood vessel on the lower left of the picture. Note the high contrast of the

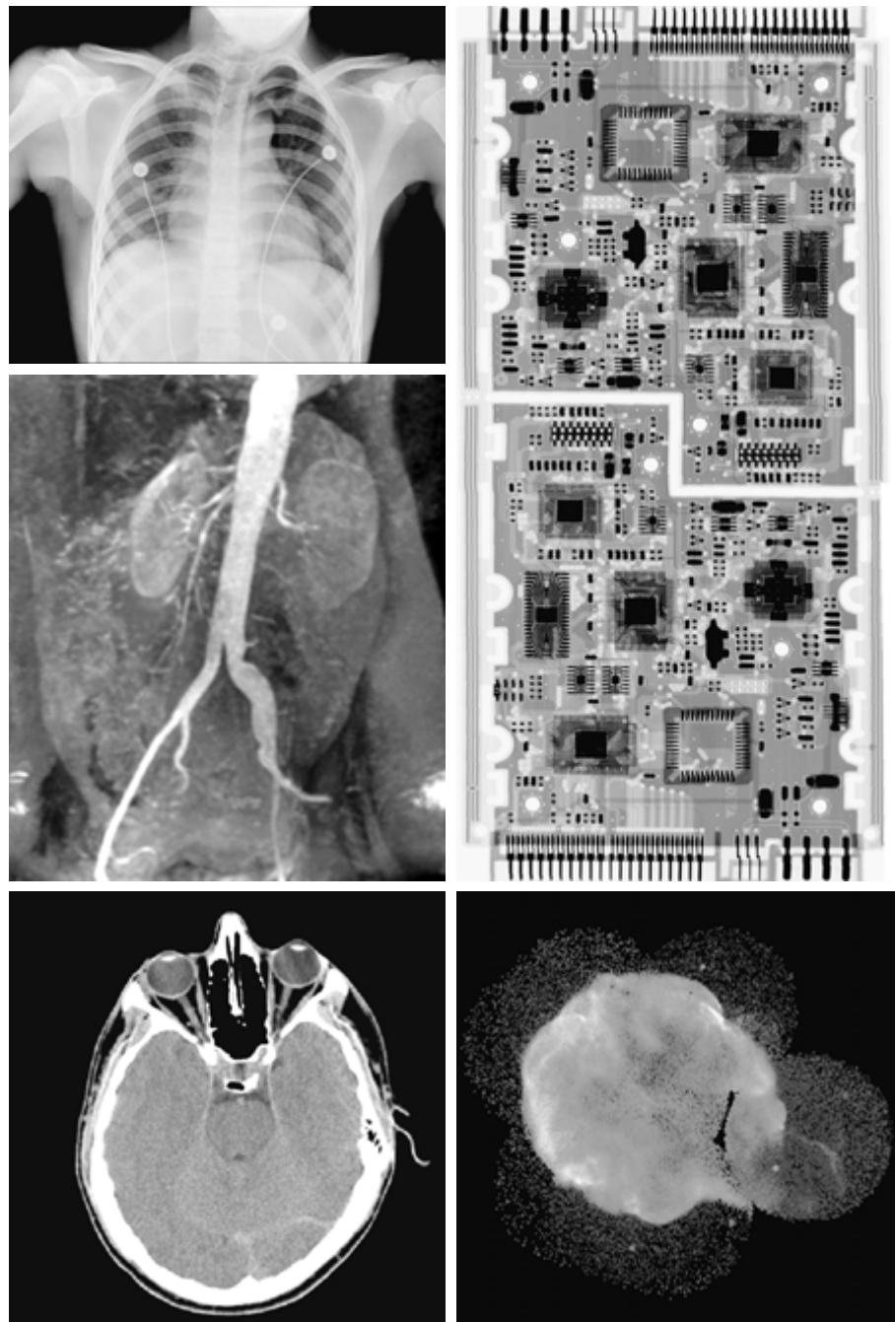


FIGURE 1.7 Examples of X-ray imaging. (a) Chest X-ray. (b) Aortic angiogram. (c) Head CT. (d) Circuit boards. (e) Cygnus Loop. (Images courtesy of (a) and (c) Dr. David R. Pickens, Dept. of Radiology & Radiological Sciences, Vanderbilt University Medical Center, (b) Dr. Thomas R. Gest, Division of Anatomical Sciences, University of Michigan Medical School, (d) Mr. Joseph E. Pascente, Lixi, Inc., and (e) NASA.)

large vessel as the contrast medium flows up in the direction of the kidneys, which are also visible in the image. As discussed in Chapter 3, angiography is a major area of digital image processing, where image subtraction is used to enhance further the blood vessels being studied.

Perhaps the best known of all uses of X-rays in medical imaging is computerized axial tomography. Due to their resolution and 3-D capabilities, CAT scans revolutionized medicine from the moment they first became available in the early 1970s. As noted in Section 1.2, each CAT image is a “slice” taken perpendicularly through the patient. Numerous slices are generated as the patient is moved in a longitudinal direction. The ensemble of such images constitutes a 3-D rendition of the inside of the patient, with the longitudinal resolution being proportional to the number of slice images taken. Figure 1.7(c) shows a typical head CAT slice image.

Techniques similar to the ones just discussed, but generally involving higher-energy X-rays, are applicable in industrial processes. Figure 1.7(d) shows an X-ray image of an electronic circuit board. Such images, representative of literally hundreds of industrial applications of X-rays, are used to examine circuit boards for flaws in manufacturing, such as missing components or broken traces. Industrial CAT scans are useful when the parts can be penetrated by X-rays, such as in plastic assemblies, and even large bodies, like solid-propellant rocket motors. Figure 1.7(e) shows an example of X-ray imaging in astronomy. This image is the Cygnus Loop of Fig. 1.6(c), but imaged this time in the X-ray band.

1.3.3 Imaging in the Ultraviolet Band

Applications of ultraviolet “light” are varied. They include lithography, industrial inspection, microscopy, lasers, biological imaging, and astronomical observations. We illustrate imaging in this band with examples from microscopy and astronomy.

Ultraviolet light is used in fluorescence microscopy, one of the fastest growing areas of microscopy. Fluorescence is a phenomenon discovered in the middle of the nineteenth century, when it was first observed that the mineral fluorspar fluoresces when ultraviolet light is directed upon it. The ultraviolet light itself is not visible, but when a photon of ultraviolet radiation collides with an electron in an atom of a fluorescent material, it elevates the electron to a higher energy level. Subsequently, the excited electron relaxes to a lower level and emits light in the form of a lower-energy photon in the visible (red) light region. The basic task of the fluorescence microscope is to use an excitation light to irradiate a prepared specimen and then to separate the much weaker radiating fluorescent light from the brighter excitation light. Thus, only the emission light reaches the eye or other detector. The resulting fluorescing areas shine against a dark background with sufficient contrast to permit detection. The darker the background of the nonfluorescing material, the more efficient the instrument.

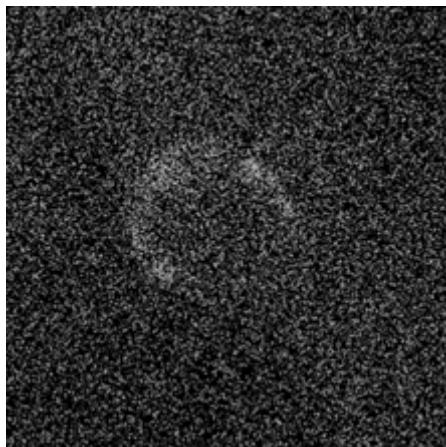
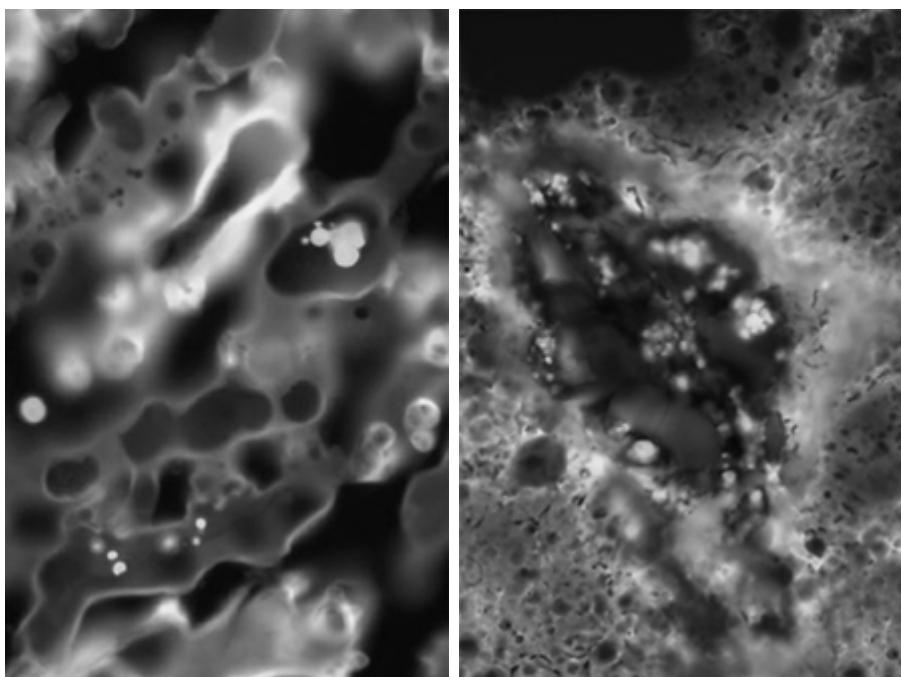
Fluorescence microscopy is an excellent method for studying materials that can be made to fluoresce, either in their natural form (primary fluorescence) or when treated with chemicals capable of fluorescing (secondary fluorescence). Figures 1.8(a) and (b) show results typical of the capability of fluorescence

a b
c

FIGURE 1.8

Examples of ultraviolet imaging.

- (a) Normal corn.
 - (b) Smut corn.
 - (c) Cygnus Loop.
- (Images courtesy of (a) and (b) Dr. Michael W. Davidson, Florida State University, (c) NASA.)



microscopy. Figure 1.8(a) shows a fluorescence microscope image of normal corn, and Fig. 1.8(b) shows corn infected by “smut,” a disease of cereals, corn, grasses, onions, and sorghum that can be caused by any of more than 700 species of parasitic fungi. Corn smut is particularly harmful because corn is one of the principal food sources in the world. As another illustration, Fig. 1.8(c) shows the Cygnus Loop imaged in the high-energy region of the ultraviolet band.

1.3.4 Imaging in the Visible and Infrared Bands

Considering that the visual band of the electromagnetic spectrum is the most familiar in all our activities, it is not surprising that imaging in this band outweighs by far all the others in terms of scope of application. The infrared band

often is used in conjunction with visual imaging, so we have grouped the visible and infrared bands in this section for the purpose of illustration. We consider in the following discussion applications in light microscopy, astronomy, remote sensing, industry, and law enforcement.

Figure 1.9 shows several examples of images obtained with a light microscope. The examples range from pharmaceuticals and microinspection to materials characterization. Even in just microscopy, the application areas are too numerous to detail here. It is not difficult to conceptualize the types of processes one might apply to these images, ranging from enhancement to measurements.

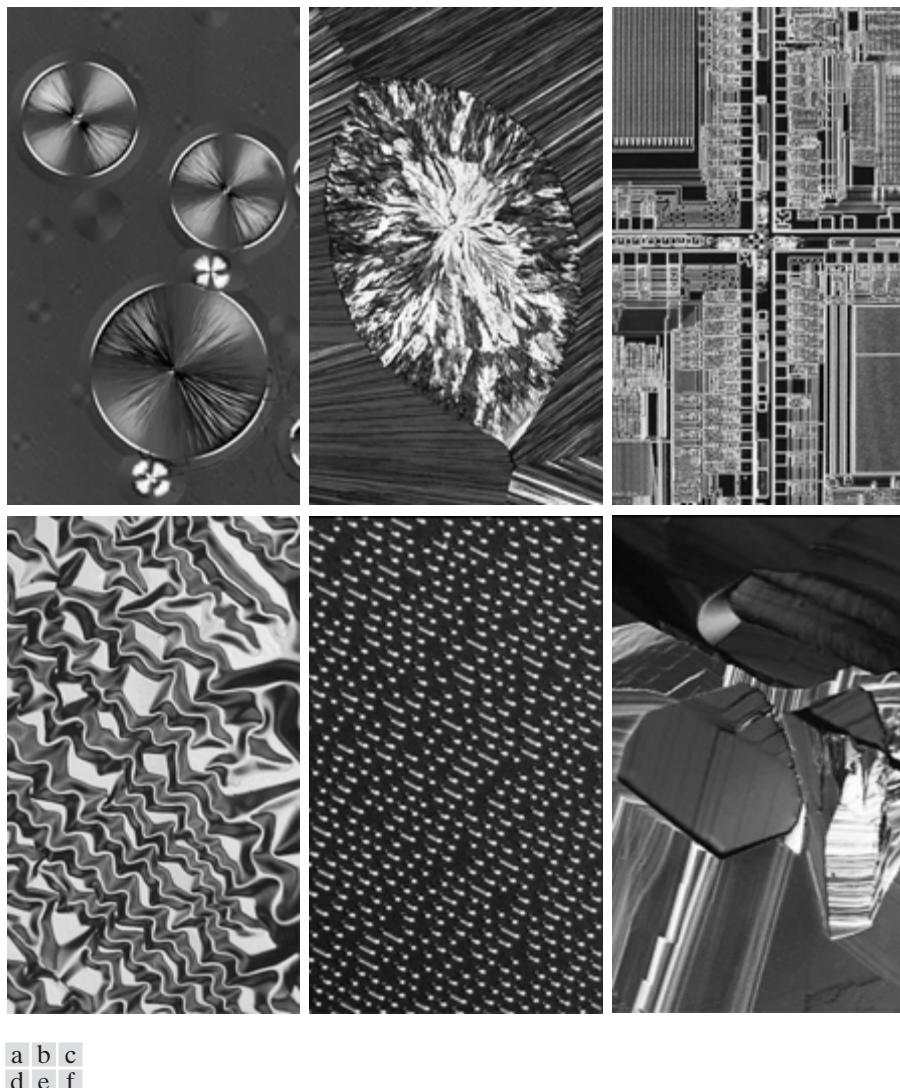


FIGURE 1.9 Examples of light microscopy images. (a) Taxol (anticancer agent), magnified 250×. (b) Cholesterol—40×. (c) Microprocessor—60×. (d) Nickel oxide thin film—600×. (e) Surface of audio CD—1750×. (f) Organic superconductor—450×. (Images courtesy of Dr. Michael W. Davidson, Florida State University.)

TABLE 1.1
Thematic bands
in NASA's
LANDSAT
satellite.

Band No.	Name	Wavelength (μm)	Characteristics and Uses
1	Visible blue	0.45–0.52	Maximum water penetration
2	Visible green	0.52–0.60	Good for measuring plant vigor
3	Visible red	0.63–0.69	Vegetation discrimination
4	Near infrared	0.76–0.90	Biomass and shoreline mapping
5	Middle infrared	1.55–1.75	Moisture content of soil and vegetation
6	Thermal infrared	10.4–12.5	Soil moisture; thermal mapping
7	Middle infrared	2.08–2.35	Mineral mapping

Another major area of visual processing is remote sensing, which usually includes several bands in the visual and infrared regions of the spectrum. Table 1.1 shows the so-called *thematic bands* in NASA's LANDSAT satellite. The primary function of LANDSAT is to obtain and transmit images of the Earth from space, for purposes of monitoring environmental conditions on the planet. The bands are expressed in terms of wavelength, with $1 \mu\text{m}$ being equal to 10^{-6} m (we discuss the wavelength regions of the electromagnetic spectrum in more detail in Chapter 2). Note the characteristics and uses of each band.

In order to develop a basic appreciation for the power of this type of *multispectral* imaging, consider Fig. 1.10, which shows one image for each of the spec-

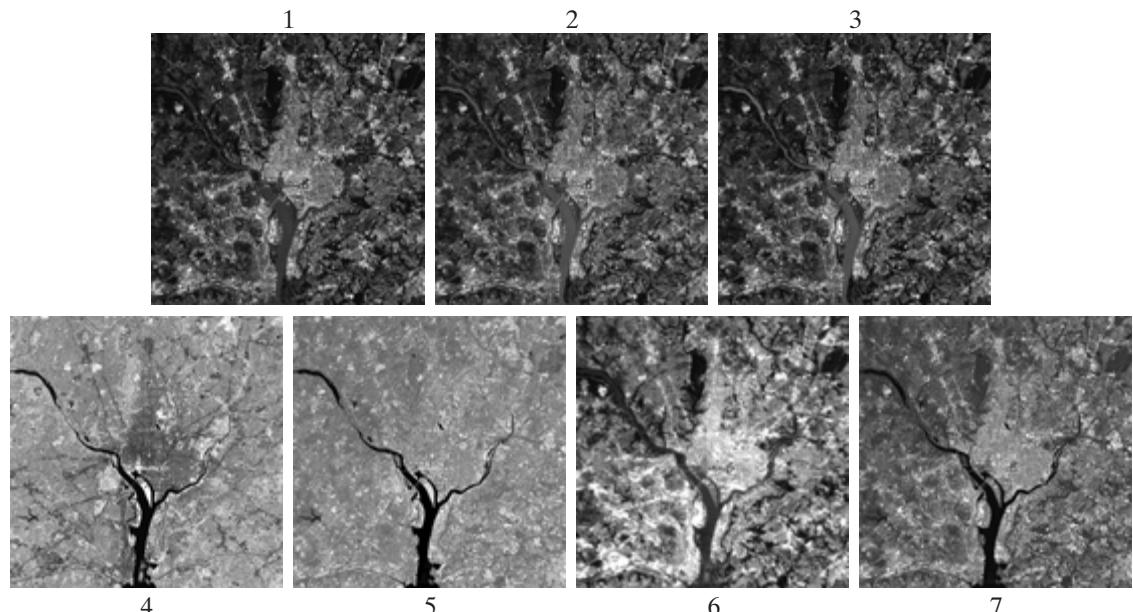


FIGURE 1.10 LANDSAT satellite images of the Washington, D.C. area. The numbers refer to the thematic bands in Table 1.1. (Images courtesy of NASA.)

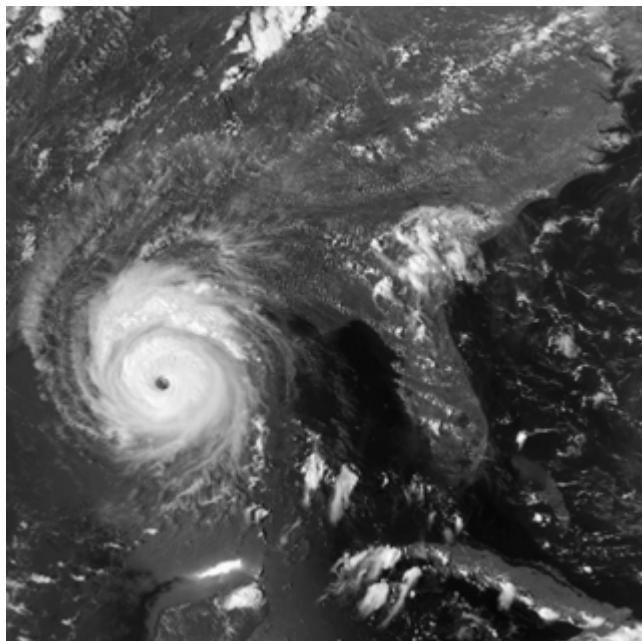


FIGURE 1.11
Multispectral
image of
Hurricane
Andrew taken by
NOAA GEOS
(Geostationary
Environmental
Operational
Satellite) sensors.
(Courtesy of
NOAA.)

tral bands in Table 1.1. The area imaged is Washington D.C., which includes features such as buildings, roads, vegetation, and a major river (the Potomac) going through the city. Images of population centers are used routinely (over time) to assess population growth and shift patterns, pollution, and other factors harmful to the environment. The differences between visual and infrared image features are quite noticeable in these images. Observe, for example, how well defined the river is from its surroundings in Bands 4 and 5.

Weather observation and prediction also are major applications of multispectral imaging from satellites. For example, Fig. 1.11 is an image of a hurricane taken by a National Oceanographic and Atmospheric Administration (NOAA) satellite using sensors in the visible and infrared bands. The eye of the hurricane is clearly visible in this image.

Figures 1.12 and 1.13 show an application of infrared imaging. These images are part of the *Nighttime Lights of the World* data set, which provides a global inventory of human settlements. The images were generated by the infrared imaging system mounted on a NOAA DMSP (Defense Meteorological Satellite Program) satellite. The infrared imaging system operates in the band 10.0 to 13.4 μm , and has the unique capability to observe faint sources of visible-near infrared emissions present on the Earth's surface, including cities, towns, villages, gas flares, and fires. Even without formal training in image processing, it is not difficult to imagine writing a computer program that would use these images to estimate the percent of total electrical energy used by various regions of the world.

FIGURE 1.12

Infrared satellite images of the Americas. The small gray map is provided for reference.
(Courtesy of NOAA.)



A major area of imaging in the visual spectrum is in automated visual inspection of manufactured goods. Figure 1.14 shows some examples. Figure 1.14(a) is a controller board for a CD-ROM drive. A typical image processing task with products like this is to inspect them for missing parts (the black square on the top, right quadrant of the image is an example of a missing component). Figure 1.14(b) is an imaged pill container. The objective here is to have a machine look for missing pills. Figure 1.14(c) shows an application in which image processing is used to look for bottles that are not filled up to an acceptable level. Figure 1.14(d) shows



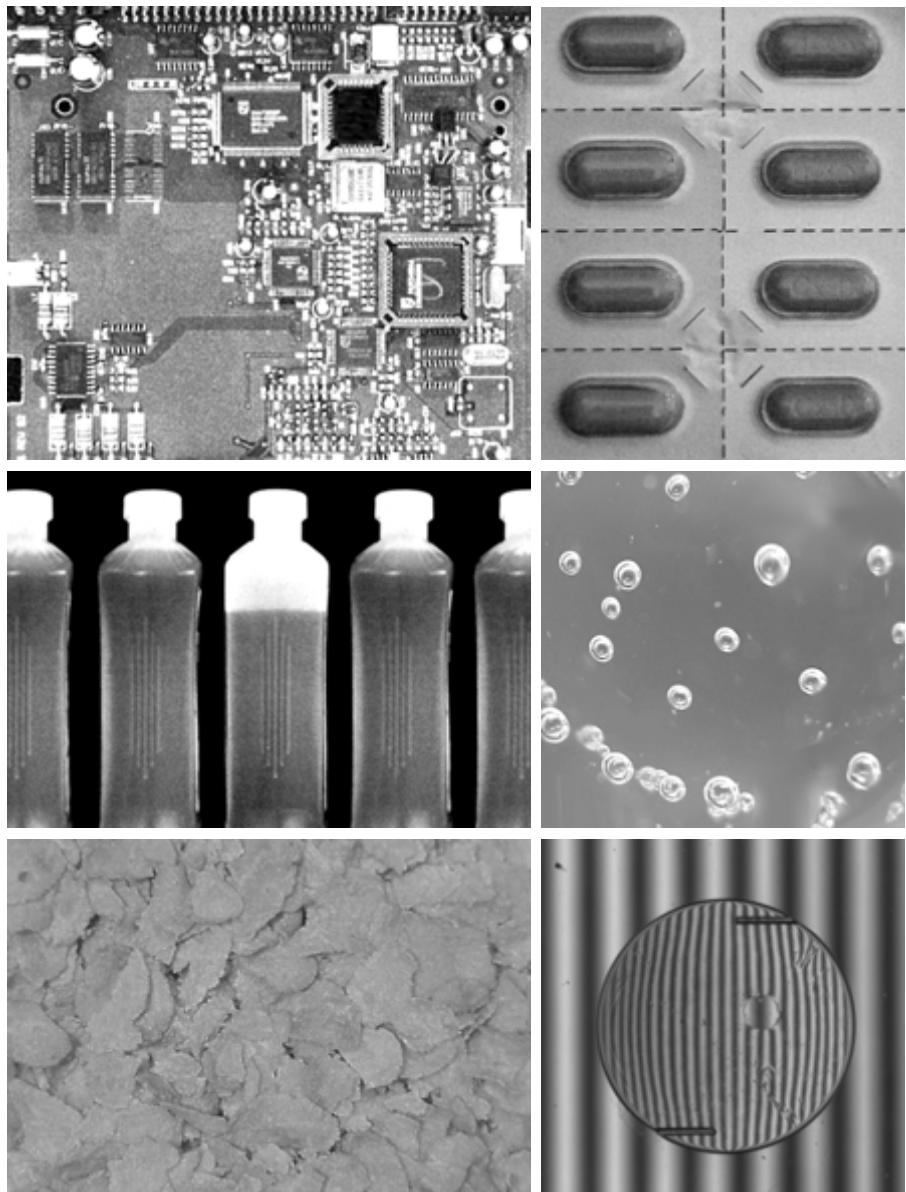
FIGURE 1.13
Infrared satellite images of the remaining populated part of the world. The small gray map is provided for reference.
(Courtesy of NOAA.)

a clear-plastic part with an unacceptable number of air pockets in it. Detecting anomalies like these is a major theme of industrial inspection that includes other products such as wood and cloth. Figure 1.14(e) shows a batch of cereal during inspection for color and the presence of anomalies such as burned flakes. Finally, Fig. 1.14(f) shows an image of an intraocular implant (replacement lens for the human eye). A “structured light” illumination technique was used to highlight for easier detection flat lens deformations toward the center of the lens. The markings at 1 o’clock and 5 o’clock are tweezers damage. Most of the other small speckle detail is debris. The objective in this type of inspection is to find damaged or incorrectly manufactured implants automatically, prior to packaging.

As a final illustration of image processing in the visual spectrum, consider Fig. 1.15. Figure 1.15(a) shows a thumb print. Images of fingerprints are routinely processed by computer, either to enhance them or to find features that aid in the automated search of a database for potential matches. Figure 1.15(b) shows an image of paper currency. Applications of digital image processing in this area include automated counting and, in law enforcement, the reading of the serial number for the purpose of tracking and identifying bills. The two vehicle images shown in Figs. 1.15 (c) and (d) are examples of automated license plate reading.

a
b
c
d
e
f

FIGURE 1.14
 Some examples of manufactured goods often checked using digital image processing. (a) A circuit board controller.
 (b) Packaged pills.
 (c) Bottles.
 (d) Bubbles in clear-plastic product.
 (e) Cereal.
 (f) Image of intraocular implant.
 (Fig. (f) courtesy of Mr. Pete Sites, Perceptics Corporation.)



The light rectangles indicate the area in which the imaging system detected the plate. The black rectangles show the results of automated reading of the plate content by the system. License plate and other applications of character recognition are used extensively for traffic monitoring and surveillance.

1.3.5 Imaging in the Microwave Band

The dominant application of imaging in the microwave band is radar. The unique feature of imaging radar is its ability to collect data over virtually any region at any time, regardless of weather or ambient lighting conditions. Some radar



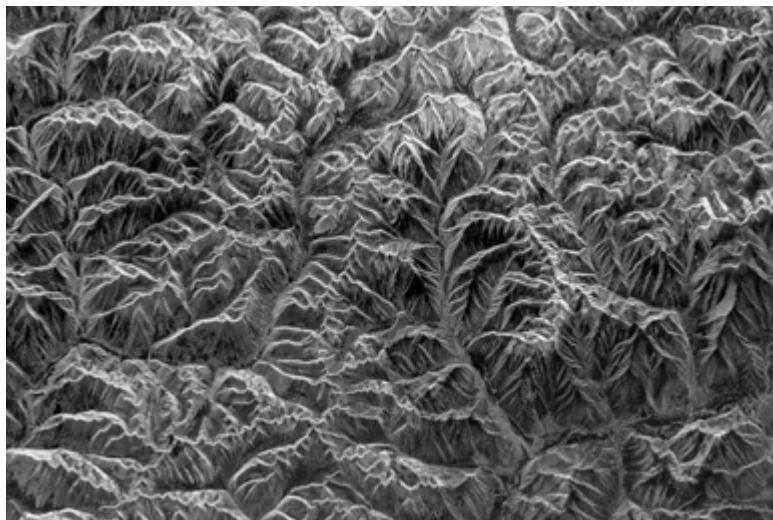
a b
c
d

FIGURE 1.15
Some additional examples of imaging in the visual spectrum.
(a) Thumb print.
(b) Paper currency.
(c) and (d). Automated license plate reading. (Figure (a) courtesy of the National Institute of Standards and Technology. Figures (c) and (d) courtesy of Dr. Juan Herrera, Perceptics Corporation.)

waves can penetrate clouds, and under certain conditions can also see through vegetation, ice, and extremely dry sand. In many cases, radar is the only way to explore inaccessible regions of the Earth's surface. An imaging radar works like a flash camera in that it provides its own illumination (microwave pulses) to illuminate an area on the ground and take a snapshot image. Instead of a camera lens, a radar uses an antenna and digital computer processing to record its images. In a radar image, one can see only the microwave energy that was reflected back toward the radar antenna.

Figure 1.16 shows a spaceborne radar image covering a rugged mountainous area of southeast Tibet, about 90 km east of the city of Lhasa. In the lower right corner is a wide valley of the Lhasa River, which is populated by Tibetan farmers and yak herders and includes the village of Menba. Mountains in this area reach about 5800 m (19,000 ft) above sea level, while the valley floors lie about 4300 m (14,000 ft) above sea level. Note the clarity and detail of the image, unencumbered by clouds or other atmospheric conditions that normally interfere with images in the visual band.

FIGURE 1.16
Spaceborne radar
image of
mountains in
southeast Tibet.
(Courtesy of
NASA.)



1.3.6 Imaging in the Radio Band

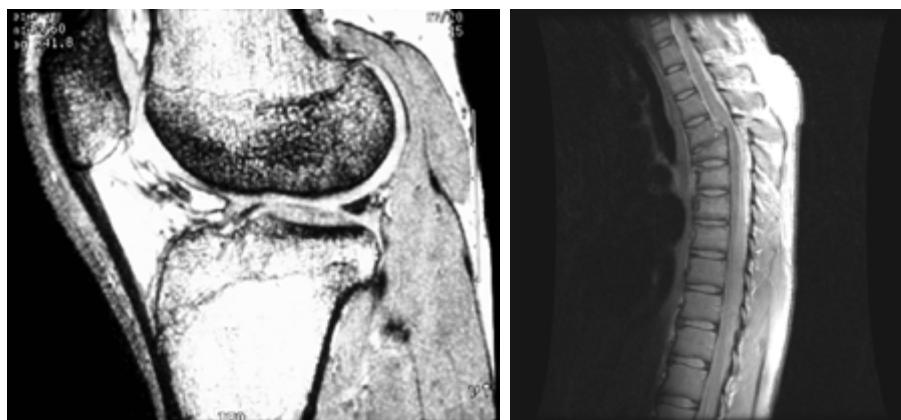
As in the case of imaging at the other end of the spectrum (gamma rays), the major applications of imaging in the radio band are in medicine and astronomy. In medicine radio waves are used in magnetic resonance imaging (MRI). This technique places a patient in a powerful magnet and passes radio waves through his or her body in short pulses. Each pulse causes a responding pulse of radio waves to be emitted by the patient's tissues. The location from which these signals originate and their strength are determined by a computer, which produces a two-dimensional picture of a section of the patient. MRI can produce pictures in any plane. Figure 1.17 shows MRI images of a human knee and spine.

The last image to the right in Fig. 1.18 shows an image of the Crab Pulsar in the radio band. Also shown for an interesting comparison are images of the same region but taken in most of the bands discussed earlier. Note that each image gives a totally different “view” of the Pulsar.

1.3.7 Examples in which Other Imaging Modalities Are Used

Although imaging in the electromagnetic spectrum is dominant by far, there are a number of other imaging modalities that also are important. Specifically, we discuss in this section acoustic imaging, electron microscopy, and synthetic (computer-generated) imaging.

Imaging using “sound” finds application in geological exploration, industry, and medicine. Geological applications use sound in the low end of the sound spectrum (hundreds of Hertz) while imaging in other areas use ultrasound (millions of Hertz). The most important commercial applications of image processing in geology are in mineral and oil exploration. For image acquisition over land, one of the main approaches is to use a large truck and a large flat steel plate. The plate is pressed on the ground by the truck, and the truck is vibrated through a fre-



a b

FIGURE 1.17 MRI images of a human (a) knee, and (b) spine. (Image (a) courtesy of Dr. Thomas R. Gest, Division of Anatomical Sciences, University of Michigan Medical School, and (b) Dr. David R. Pickens, Department of Radiology and Radiological Sciences, Vanderbilt University Medical Center.)

quency spectrum up to 100 Hz. The strength and speed of the returning sound waves are determined by the composition of the earth below the surface. These are analyzed by computer, and images are generated from the resulting analysis.

For marine acquisition, the energy source consists usually of two air guns towed behind a ship. Returning sound waves are detected by hydrophones placed in cables that are either towed behind the ship, laid on the bottom of the ocean, or hung from buoys (vertical cables). The two air guns are alternately pressurized to ~ 2000 psi and then set off. The constant motion of the ship provides a transversal direction of motion that, together with the returning sound waves, is used to generate a 3-D map of the composition of the Earth below the bottom of the ocean.

Figure 1.19 shows a cross-sectional image of a well-known 3-D model against which the performance of seismic imaging algorithms is tested. The arrow points to a hydrocarbon (oil and/or gas) trap. This target is brighter than the surrounding layers because of the change in density in the target region is larger.

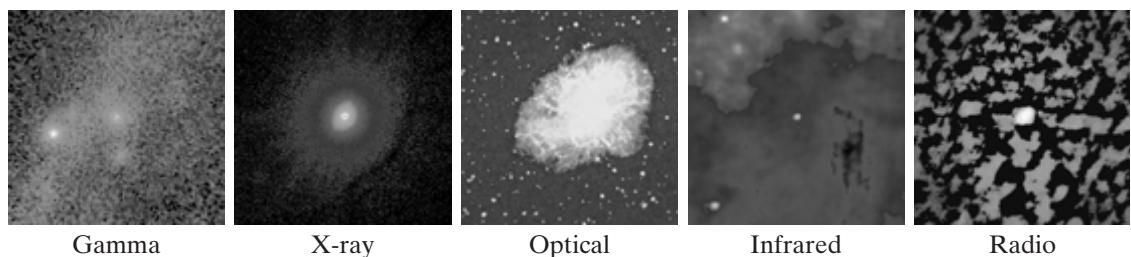
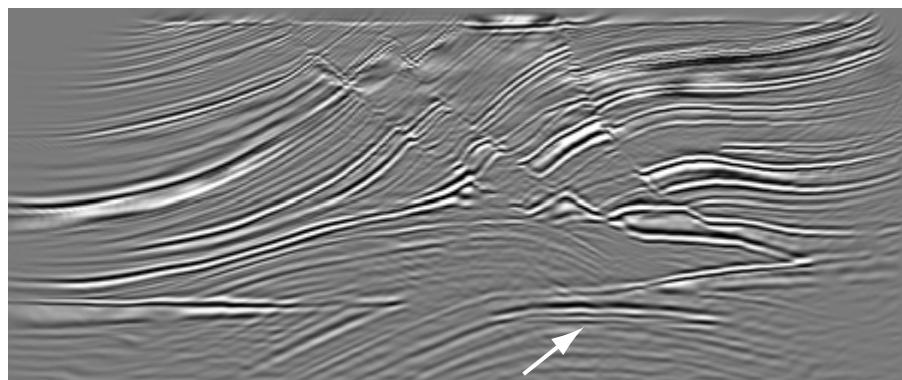


FIGURE 1.18 Images of the Crab Pulsar (in the center of images) covering the electromagnetic spectrum. (Courtesy of NASA.)

FIGURE 1.19

Cross-sectional image of a seismic model. The arrow points to a hydrocarbon (oil and/or gas) trap. (Courtesy of Dr. Curtis Ober, Sandia National Laboratories.)



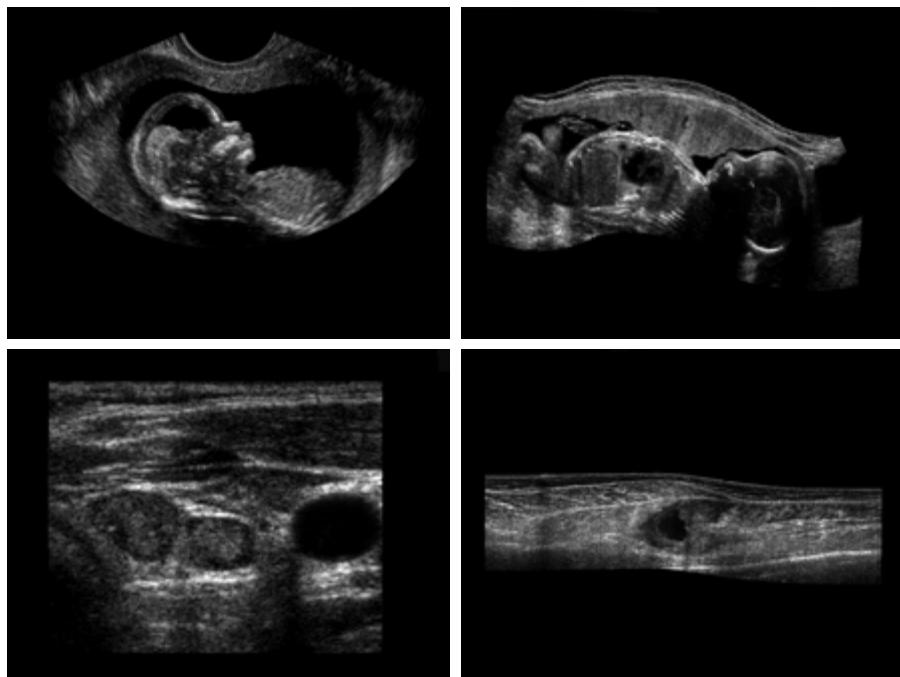
Seismic interpreters look for these “bright spots” to find oil and gas. The layers above also are bright, but their brightness does not vary as strongly across the layers. Many seismic reconstruction algorithms have difficulty imaging this target because of the faults above it.

Although ultrasound imaging is used routinely in manufacturing, the best known applications of this technique are in medicine, especially in obstetrics, where unborn babies are imaged to determine the health of their development. A byproduct of this examination is determining the sex of the baby. Ultrasound images are generated using the following basic procedure:

1. The ultrasound system (a computer, ultrasound probe consisting of a source and receiver, and a display) transmits high-frequency (1 to 5 MHz) sound pulses into the body.
2. The sound waves travel into the body and hit a boundary between tissues (e.g., between fluid and soft tissue, soft tissue and bone). Some of the sound waves are reflected back to the probe, while some travel on further until they reach another boundary and get reflected.
3. The reflected waves are picked up by the probe and relayed to the computer.
4. The machine calculates the distance from the probe to the tissue or organ boundaries using the speed of sound in tissue (1540 m/s) and the time of the each echo’s return.
5. The system displays the distances and intensities of the echoes on the screen, forming a two-dimensional image.

In a typical ultrasound image, millions of pulses and echoes are sent and received each second. The probe can be moved along the surface of the body and angled to obtain various views. Figure 1.20 shows several examples.

We continue the discussion on imaging modalities with some examples of electron microscopy. Electron microscopes function as their optical counterparts, except that they use a focused beam of electrons instead of light to image a specimen. The operation of electron microscopes involves the following basic steps: A stream of electrons is produced by an electron source and accelerated toward the specimen using a positive electrical potential. This stream is con-



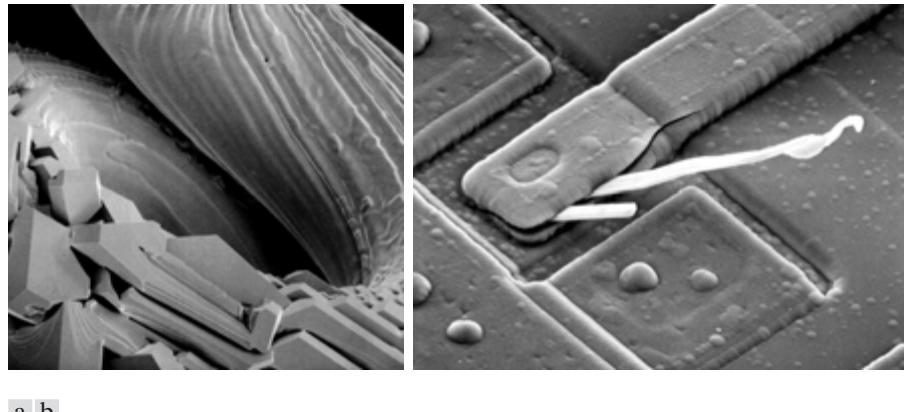
a b
c d

FIGURE 1.20
Examples of ultrasound imaging. (a) Baby. (2) Another view of baby. (c) Thyroids. (d) Muscle layers showing lesion. (Courtesy of Siemens Medical Systems, Inc., Ultrasound Group.)

fined and focused using metal apertures and magnetic lenses into a thin, focused, monochromatic beam. This beam is focused onto the sample using a magnetic lens. Interactions occur inside the irradiated sample, affecting the electron beam. These interactions and effects are detected and transformed into an image, much in the same way that light is reflected from, or absorbed by, objects in a scene. These basic steps are carried out in all electron microscopes, regardless of type.

A *transmission electron microscope* (TEM) works much like a slide projector. A projector shines (transmits) a beam of light through the slide; as the light passes through the slide, it is affected by the contents of the slide. This transmitted beam is then projected onto the viewing screen, forming an enlarged image of the slide. TEMs work the same way, except that they shine a beam of electrons through a specimen (analogous to the slide). The fraction of the beam transmitted through the specimen is projected onto a phosphor screen. The interaction of the electrons with the phosphor produces light and, therefore, a viewable image. A *scanning electron microscope* (SEM), on the other hand, actually scans the electron beam and records the interaction of beam and sample at each location. This produces one dot on a phosphor screen. A complete image is formed by a raster scan of the beam through the sample, much like a TV camera. The electrons interact with a phosphor screen and produce light. SEMs are suitable for “bulky” samples, while TEMs require very thin samples.

Electron microscopes are capable of very high magnification. While light microscopy is limited to magnifications on the order $1000\times$, electron microscopes



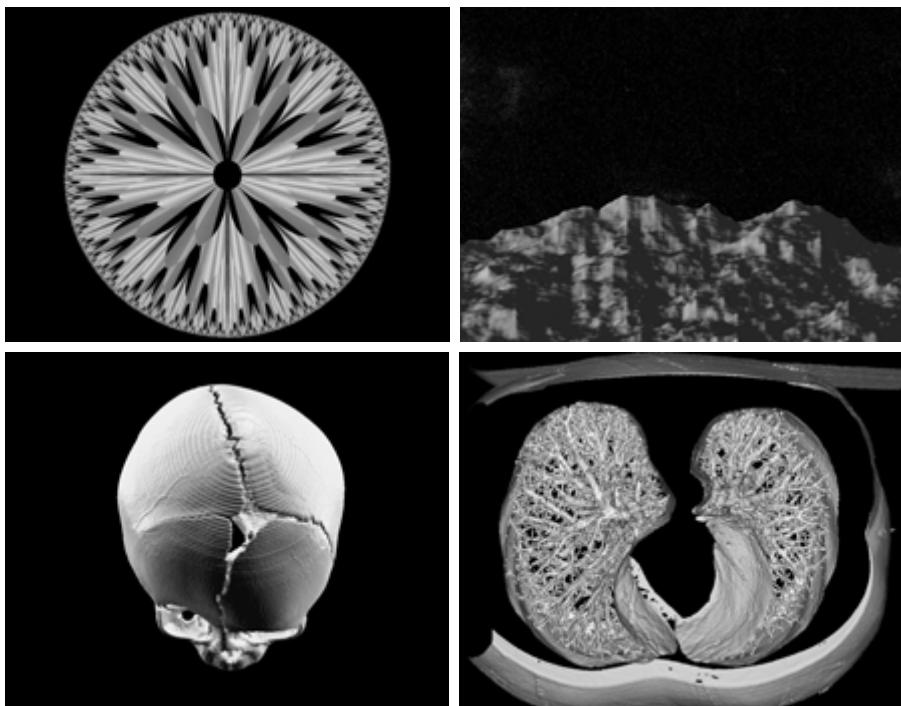
a b

FIGURE 1.21 (a) 250 \times SEM image of a tungsten filament following thermal failure. (b) 2500 \times SEM image of damaged integrated circuit. The white fibers are oxides resulting from thermal destruction. (Figure (a) courtesy of Mr. Michael Shaffer, Department of Geological Sciences, University of Oregon, Eugene; (b) courtesy of Dr. J. M. Hudak, McMaster University, Hamilton, Ontario, Canada.)

can achieve magnification of 10,000 \times or more. Figure 1.21 shows two SEM images of specimen failures due to thermal overload.

We conclude the discussion of imaging modalities by looking briefly at images that are not obtained from physical objects. Instead, they are generated by computer. *Fractals* are striking examples of computer-generated images [Lu (1997)]. Basically, a fractal is nothing more than an iterative reproduction of a basic pattern according to some mathematical rules. For instance, *tiling* is one of the simplest ways to generate a fractal image. A square can be subdivided into four square subregions, each of which can be further subdivided into four smaller square regions, and so on. Depending on the complexity of the rules for filling each subsquare, some beautiful tile images can be generated using this method. Of course, the geometry can be arbitrary. For instance, the fractal image could be grown radially out of a center point. Figure 1.22(a) shows a fractal grown in this way. The reader will recognize this image as the theme image used in the beginning page of each chapter in this book, selected because of its artistic simplicity and abstract analogy to a human eye. Figure 1.22(b) shows another fractal (a “moonscape”) that provides an interesting analogy to the images of space used as illustrations in some of the preceding sections.

Fractal images tend toward artistic, mathematical formulations of “growth” of subimage elements according to some rules. They are useful sometimes as random textures. A more structured approach to image generation by computer lies in 3-D modeling. This is an area that provides an important intersection between image processing and computer graphics and is the basis for many 3-D visualization systems (e.g., flight simulators). Figures 1.22(c) and (d) show examples of computer-generated images. Since the original object is created in 3-D, images can be generated in any perspective from plane projections of the 3-D volume. Images of this type can be used for medical training and for a host of other applications, such as criminal forensics and special effects.



a b
c d

FIGURE 1.22
(a) and (b) Fractal images. (c) and (d) Images generated from 3-D computer models of the objects shown. (Figures (a) and (b) courtesy of Ms. Melissa D. Binde, Swarthmore College, (c) and (d) courtesy of NASA.)

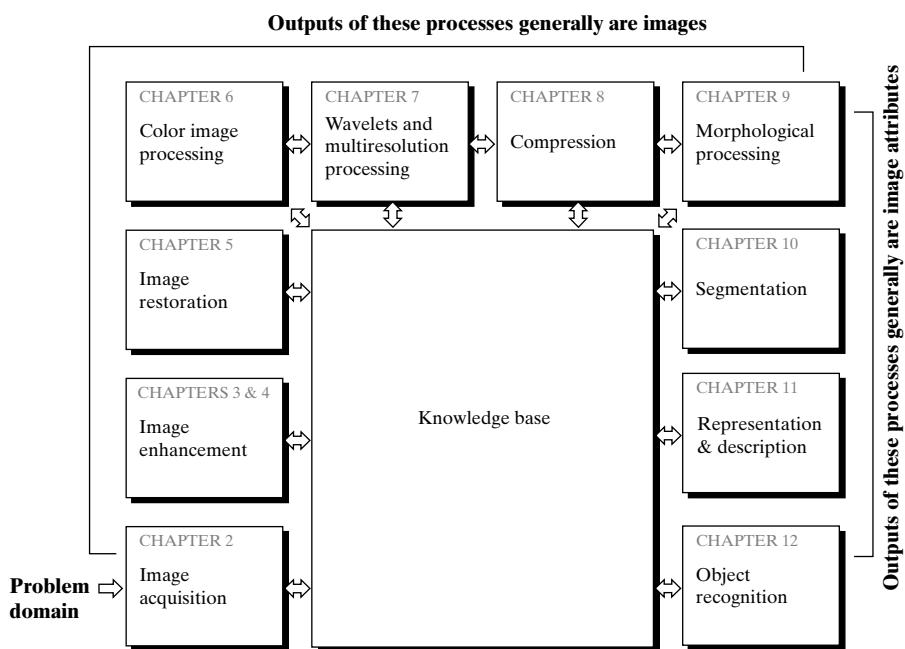
1.4 Fundamental Steps in Digital Image Processing

It is helpful to divide the material covered in the following chapters into the two broad categories defined in Section 1.1: methods whose input and output are images, and methods whose inputs may be images, but whose outputs are attributes extracted from those images. This organization is summarized in Fig. 1.23. The diagram does not imply that every process is applied to an image. Rather, the intention is to convey an idea of all the methodologies that can be applied to images for different purposes and possibly with different objectives. The discussion in this section may be viewed as a brief overview of the material in the remainder of the book.

Image acquisition is the first process shown in Fig. 1.23. The discussion in Section 1.3 gave some hints regarding the origin of digital images. This topic is considered in much more detail in Chapter 2, where we also introduce a number of basic digital image concepts that are used throughout the book. Note that acquisition could be as simple as being given an image that is already in digital form. Generally, the image acquisition stage involves preprocessing, such as scaling.

Image enhancement is among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interest in an image. A familiar example of enhancement is when we increase the contrast of an image because “it looks better.” It is important to keep in mind that

FIGURE 1.23
Fundamental steps in digital image processing.



enhancement is a very subjective area of image processing. Two chapters are devoted to enhancement, not because it is more important than the other topics covered in the book but because we use enhancement as an avenue to introduce the reader to techniques that are used in other chapters as well. Thus, rather than having a chapter dedicated to mathematical preliminaries, we introduce a number of needed mathematical concepts by showing how they apply to enhancement. This approach allows the reader to gain familiarity with these concepts in the context of image processing. A good example of this is the Fourier transform, which is introduced in Chapter 4 but is used also in several of the other chapters.

Image restoration is an area that also deals with improving the appearance of an image. However, unlike enhancement, which is subjective, image restoration is objective, in the sense that restoration techniques tend to be based on mathematical or probabilistic models of image degradation. Enhancement, on the other hand, is based on human subjective preferences regarding what constitutes a “good” enhancement result.

Color image processing is an area that has been gaining in importance because of the significant increase in the use of digital images over the Internet. Chapter 5 covers a number of fundamental concepts in color models and basic color processing in a digital domain. Color is used also in later chapters as the basis for extracting features of interest in an image.

Wavelets are the foundation for representing images in various degrees of resolution. In particular, this material is used in this book for image data compression and for pyramidal representation, in which images are subdivided successively into smaller regions.

Compression, as the name implies, deals with techniques for reducing the storage required to save an image, or the bandwidth required to transmit it. Although storage technology has improved significantly over the past decade, the same cannot be said for transmission capacity. This is true particularly in uses of the Internet, which are characterized by significant pictorial content. Image compression is familiar (perhaps inadvertently) to most users of computers in the form of image file extensions, such as the jpg file extension used in the JPEG (Joint Photographic Experts Group) image compression standard.

Morphological processing deals with tools for extracting image components that are useful in the representation and description of shape. The material in this chapter begins a transition from processes that output images to processes that output image attributes, as indicated in Section 1.1.

Segmentation procedures partition an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually. On the other hand, weak or erratic segmentation algorithms almost always guarantee eventual failure. In general, the more accurate the segmentation, the more likely recognition is to succeed.

Representation and description almost always follow the output of a segmentation stage, which usually is raw pixel data, constituting either the boundary of a region (i.e., the set of pixels separating one image region from another) or all the points in the region itself. In either case, converting the data to a form suitable for computer processing is necessary. The first decision that must be made is whether the data should be represented as a boundary or as a complete region. Boundary representation is appropriate when the focus is on external shape characteristics, such as corners and inflections. Regional representation is appropriate when the focus is on internal properties, such as texture or skeletal shape. In some applications, these representations complement each other. Choosing a representation is only part of the solution for transforming raw data into a form suitable for subsequent computer processing. A method must also be specified for describing the data so that features of interest are highlighted. *Description*, also called *feature selection*, deals with extracting attributes that result in some quantitative information of interest or are basic for differentiating one class of objects from another.

Recognition is the process that assigns a label (e.g., “vehicle”) to an object based on its descriptors. As detailed in Section 1.1, we conclude our coverage of digital image processing with the development of methods for recognition of individual objects.

So far we have said nothing about the need for prior knowledge or about the interaction between the *knowledge base* and the processing modules in Fig. 1.23. Knowledge about a problem domain is coded into an image processing system in the form of a knowledge database. This knowledge may be as simple as detailing regions of an image where the information of interest is known to be located, thus limiting the search that has to be conducted in seeking that information. The knowledge base also can be quite complex, such as an inter-related list of all major possible defects in a materials inspection problem or an

image database containing high-resolution satellite images of a region in connection with change-detection applications. In addition to guiding the operation of each processing module, the knowledge base also controls the interaction between modules. This distinction is made in Fig. 1.23 by the use of double-headed arrows between the processing modules and the knowledge base, as opposed to single-headed arrows linking the processing modules.

Although we do not discuss image display explicitly at this point, it is important to keep in mind that viewing the results of image processing can take place at the output of any stage in Fig. 1.23. We also note that not all image processing applications require the complexity of interactions implied by Fig. 1.23. In fact, not even all those modules are needed in some cases. For example, image enhancement for human visual interpretation seldom requires use of any of the other stages in Fig. 1.23. In general, however, as the complexity of an image processing task increases, so does the number of processes required to solve the problem.

1.5 Components of an Image Processing System

As recently as the mid-1980s, numerous models of image processing systems being sold throughout the world were rather substantial peripheral devices that attached to equally substantial host computers. Late in the 1980s and early in the 1990s, the market shifted to image processing hardware in the form of single boards designed to be compatible with industry standard buses and to fit into engineering workstation cabinets and personal computers. In addition to lowering costs, this market shift also served as a catalyst for a significant number of new companies whose specialty is the development of software written specifically for image processing.

Although large-scale image processing systems still are being sold for massive imaging applications, such as processing of satellite images, the trend continues toward miniaturizing and blending of general-purpose small computers with specialized image processing hardware. Figure 1.24 shows the basic components comprising a typical *general-purpose* system used for digital image processing. The function of each component is discussed in the following paragraphs, starting with image sensing.

With reference to *sensing*, two elements are required to acquire digital images. The first is a physical device that is sensitive to the energy radiated by the object we wish to image. The second, called a *digitizer*, is a device for converting the output of the physical sensing device into digital form. For instance, in a digital video camera, the sensors produce an electrical output proportional to light intensity. The digitizer converts these outputs to digital data. These topics are covered in some detail in Chapter 2.

Specialized image processing hardware usually consists of the digitizer just mentioned, plus hardware that performs other primitive operations, such as an arithmetic logic unit (ALU), which performs arithmetic and logical operations in parallel on entire images. One example of how an ALU is used is in averaging images as quickly as they are digitized, for the purpose of noise reduction. This type of hardware sometimes is called a *front-end subsystem*, and its most

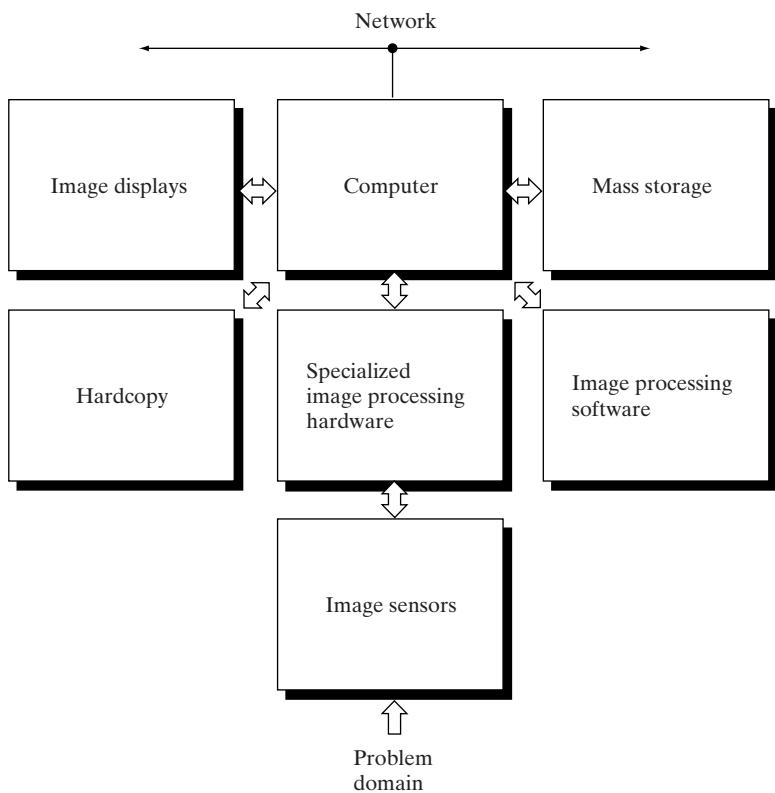


FIGURE 1.24
Components of a general-purpose image processing system.

distinguishing characteristic is speed. In other words, this unit performs functions that require fast data throughputs (e.g., digitizing and averaging video images at 30 frames/s) that the typical main computer cannot handle.

The *computer* in an image processing system is a general-purpose computer and can range from a PC to a supercomputer. In dedicated applications, sometimes specially designed computers are used to achieve a required level of performance, but our interest here is on general-purpose image processing systems. In these systems, almost any well-equipped PC-type machine is suitable for off-line image processing tasks.

Software for image processing consists of specialized modules that perform specific tasks. A well-designed package also includes the capability for the user to write code that, as a minimum, utilizes the specialized modules. More sophisticated software packages allow the integration of those modules and general-purpose software commands from at least one computer language.

Mass storage capability is a must in image processing applications. An image of size 1024×1024 pixels, in which the intensity of each pixel is an 8-bit quantity, requires one megabyte of storage space if the image is not compressed. When dealing with thousands, or even millions, of images, providing adequate storage in an image processing system can be a challenge. Digital storage for

image processing applications falls into three principal categories: (1) short-term storage for use during processing, (2) on-line storage for relatively fast recall, and (3) archival storage, characterized by infrequent access. Storage is measured in bytes (eight bits), Kbytes (one thousand bytes), Mbytes (one million bytes), Gbytes (meaning giga, or one billion, bytes), and Tbytes (meaning tera, or one trillion, bytes).

One method of providing short-term storage is computer memory. Another is by specialized boards, called *frame buffers*, that store one or more images and can be accessed rapidly, usually at video rates (e.g., at 30 complete images per second). The latter method allows virtually instantaneous image *zoom*, as well as *scroll* (vertical shifts) and *pan* (horizontal shifts). Frame buffers usually are housed in the specialized image processing hardware unit shown in Fig. 1.24. On-line storage generally takes the form of magnetic disks or optical-media storage. The key factor characterizing on-line storage is frequent access to the stored data. Finally, archival storage is characterized by massive storage requirements but infrequent need for access. Magnetic tapes and optical disks housed in “jukeboxes” are the usual media for archival applications.

Image displays in use today are mainly color (preferably flat screen) TV monitors. Monitors are driven by the outputs of image and graphics display cards that are an integral part of the computer system. Seldom are there requirements for image display applications that cannot be met by display cards available commercially as part of the computer system. In some cases, it is necessary to have stereo displays, and these are implemented in the form of headgear containing two small displays embedded in goggles worn by the user.

Hardcopy devices for recording images include laser printers, film cameras, heat-sensitive devices, inkjet units, and digital units, such as optical and CD-ROM disks. Film provides the highest possible resolution, but paper is the obvious medium of choice for written material. For presentations, images are displayed on film transparencies or in a digital medium if image projection equipment is used. The latter approach is gaining acceptance as the standard for image presentations.

Networking is almost a default function in any computer system in use today. Because of the large amount of data inherent in image processing applications, the key consideration in image transmission is bandwidth. In dedicated networks, this typically is not a problem, but communications with remote sites via the Internet are not always as efficient. Fortunately, this situation is improving quickly as a result of optical fiber and other broadband technologies.

Summary

The main purpose of the material presented in this chapter is to provide a sense of perspective about the origins of digital image processing and, more important, about current and future areas of application of this technology. Although the coverage of these topics in this chapter was necessarily incomplete due to space limitations, it should have left the reader with a clear impression of the breadth and practical scope of digital image processing. As we proceed in the following chapters with the development of image processing theory and applications, numerous examples are provided to keep a clear focus

on the utility and promise of these techniques. Upon concluding the study of the final chapter, the reader of this book will have arrived at a level of understanding that is the foundation for most of the work currently underway in this field.

References and Further Reading

References at the end of later chapters address specific topics discussed in those chapters, and are keyed to the Bibliography at the end of the book. However, in this chapter we follow a different format in order to summarize in one place a body of journals that publish material on image processing and related topics. We also provide a list of books from which the reader can readily develop a historical and current perspective of activities in this field. Thus, the reference material cited in this chapter is intended as a general-purpose, easily accessible guide to the published literature on image processing.

Major refereed journals that publish articles on image processing and related topics include: *IEEE Transactions on Image Processing*; *IEEE Transactions on Pattern Analysis and Machine Intelligence*; *Computer Vision, Graphics, and Image Processing* (prior to 1991); *Computer Vision and Image Understanding*; *IEEE Transactions on Systems, Man and Cybernetics*; *Artificial Intelligence*; *Pattern Recognition*; *Pattern Recognition Letters*; *Journal of the Optical Society of America* (prior to 1984); *Journal of the Optical Society of America—A: Optics, Image Science and Vision*; *Optical Engineering*; *Applied Optics—Information Processing*; *IEEE Transactions on Medical Imaging*; *Journal of Electronic Imaging*; *IEEE Transactions on Information Theory*; *IEEE Transactions on Communications*; *IEEE Transactions on Acoustics, Speech and Signal Processing*; *Proceedings of the IEEE*; and issues of the *IEEE Transactions on Computers* prior to 1980. Publications of the International Society for Optical Engineering (SPIE) also are of interest.

The following books, listed in reverse chronological order (with the number of books being biased toward more recent publications), contain material that complements our treatment of digital image processing. These books represent an easily accessible overview of the area for the past 30 years and were selected to provide a variety of treatments. They range from textbooks, which cover foundation material; to handbooks, which give an overview of techniques; and finally to edited books, which contain material representative of current research in the field.

- Duda, R. O., Hart, P. E., and Stork, D. G. [2001]. *Pattern Classification*, 2nd ed., John Wiley & Sons, NY.
- Ritter, G. X. and Wilson, J. N. [2001]. *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, Boca Raton, FL.
- Shapiro, L. G. and Stockman, G. C. [2001]. *Computer Vision*, Prentice Hall, Upper Saddle River, NJ.
- Dougherty, E. R. (ed.) [2000]. *Random Processes for Image and Signal Processing*, IEEE Press, NY.
- Etienne, E. K. and Nachtegael, M. (eds.). [2000]. *Fuzzy Techniques in Image Processing*, Springer-Verlag, NY.
- Goutsias, J, Vincent, L., and Bloomberg, D. S. (eds.). [2000]. *Mathematical Morphology and Its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Boston, MA.
- Mallot, A. H. [2000]. *Computational Vision*, The MIT Press, Cambridge, MA.
- Marchand-Maillet, S. and Sharaiha, Y. M. [2000]. *Binary Digital Image Processing: A Discrete Approach*, Academic Press, NY.

- Mitra, S. K. and Sicuranza, G. L. (eds.) [2000]. *Nonlinear Image Processing*, Academic Press, NY.
- Edelman, S. [1999]. *Representation and Recognition in Vision*, The MIT Press, Cambridge, MA.
- Lillesand, T. M. and Kiefer, R. W. [1999]. *Remote Sensing and Image Interpretation*, John Wiley & Sons, NY.
- Mather, P. M. [1999]. *Computer Processing of Remotely Sensed Images: An Introduction*, John Wiley & Sons, NY.
- Petrou, M. and Bosdogianni, P. [1999]. *Image Processing: The Fundamentals*, John Wiley & Sons, UK.
- Russ, J. C. [1999]. *The Image Processing Handbook*, 3rd ed., CRC Press, Boca Raton, FL.
- Smirnov, A. [1999]. *Processing of Multidimensional Signals*, Springer-Verlag, NY.
- Sonka, M., Hlavac, V., and Boyle, R. [1999]. *Image Processing, Analysis, and Computer Vision*, PWS Publishing, NY.
- Umbaugh, S. E. [1998]. *Computer Vision and Image Processing: A Practical Approach Using CVIPtools*, Prentice Hall, Upper Saddle River, NJ.
- Haskell, B. G. and Netravali, A. N. [1997]. *Digital Pictures: Representation, Compression, and Standards*, Perseus Publishing, NY.
- Jahne, B. [1997]. *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*, Springer-Verlag, NY.
- Castleman, K. R. [1996]. *Digital Image Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.
- Geladi, P. and Grahn, H. [1996]. *Multivariate Image Analysis*, John Wiley & Sons, NY.
- Bracewell, R. N. [1995]. *Two-Dimensional Imaging*, Prentice Hall, Upper Saddle River, NJ.
- Sid-Ahmed, M. A. [1995]. *Image Processing: Theory, Algorithms, and Architectures*, McGraw-Hill, NY.
- Jain, R., Rangachar, K., and Schunk, B. [1995]. *Computer Vision*, McGraw-Hill, NY.
- Mitiche, A. [1994]. *Computational Analysis of Visual Motion*, Perseus Publishing, NY.
- Baxes, G. A. [1994]. *Digital Image Processing: Principles and Applications*, John Wiley & Sons, NY.
- Gonzalez, R. C. and Woods, R. E. [1992]. *Digital Image Processing*, Addison-Wesley, Reading, MA.
- Haralick, R. M. and Shapiro, L. G. [1992]. *Computer and Robot Vision*, vols. 1 & 2, Addison-Wesley, Reading, MA.
- Pratt, W. K. [1991] *Digital Image Processing*, 2nd ed., Wiley-Interscience, NY.
- Lim, J. S. [1990]. *Two-Dimensional Signal and Image Processing*, Prentice Hall, Upper Saddle River, NJ.
- Jain, A. K. [1989]. *Fundamentals of Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ.
- Schalkoff, R. J. [1989]. *Digital Image Processing and Computer Vision*, John Wiley & Sons, NY.
- Giardina, C. R. and Dougherty, E. R. [1988]. *Morphological Methods in Image and Signal Processing*, Prentice Hall, Upper Saddle River, NJ.

- Levine, M. D. [1985]. *Vision in Man and Machine*, McGraw-Hill, NY.
- Serra, J. [1982]. *Image Analysis and Mathematical Morphology*, Academic Press, NY.
- Ballard, D. H. and Brown, C. M. [1982]. *Computer Vision*, Prentice Hall, Upper Saddle River, NJ.
- Fu, K. S. [1982]. *Syntactic Pattern Recognition and Applications*, Prentice Hall, Upper Saddle River, NJ.
- Nevatia, R. [1982]. *Machine Perception*, Prentice Hall, Upper Saddle River, NJ.
- Pavlidis, T. [1982]. *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, MD.
- Rosenfeld, R. and Kak, A. C. [1982]. *Digital Picture Processing*, 2nd ed., vols. 1 & 2, Academic Press, NY.
- Hall, E. L. [1979]. *Computer Image Processing and Recognition*, Academic Press, NY.
- Gonzalez, R. C. and Thomason, M. G. [1978]. *Syntactic Pattern Recognition: An Introduction*, Addison-Wesley, Reading, MA.
- Andrews, H. C. and Hunt, B. R. [1977]. *Digital Image Restoration*, Prentice Hall, Upper Saddle River, NJ.
- Pavlidis, T. [1977]. *Structural Pattern Recognition*, Springer-Verlag, NY, 1977.
- Tou, J. T. and Gonzalez, R. C. [1974]. *Pattern Recognition Principles*, Addison-Wesley, Reading, MA, 1974.
- Andrews, H. C. [1970]. *Computer Techniques in Image Processing*, Academic Press, NY.

2

Digital Image Fundamentals



Those who wish to succeed must ask the right preliminary questions.

Aristotle

Preview

The purpose of this chapter is to introduce several concepts related to digital images and some of the notation used throughout the book. Section 2.1 briefly summarizes the mechanics of the human visual system, including image formation in the eye and its capabilities for brightness adaptation and discrimination. Section 2.2 discusses light, other components of the electromagnetic spectrum, and their imaging characteristics. Section 2.3 discusses imaging sensors and how they are used to generate digital images. Section 2.4 introduces the concepts of uniform image sampling and gray-level quantization. Additional topics discussed in that section include digital image representation, the effects of varying the number of samples and gray levels in an image, some important phenomena associated with sampling, and techniques for image zooming and shrinking. Section 2.5 deals with some basic relationships between pixels that are used throughout the book. Finally, Section 2.6 defines the conditions for linear operations. As noted in that section, linear operators play a central role in the development of image processing techniques.

2.1 Elements of Visual Perception

Although the digital image processing field is built on a foundation of mathematical and probabilistic formulations, human intuition and analysis play a central role in the choice of one technique versus another, and this choice often is

made based on subjective, visual judgments. Hence, developing a basic understanding of human visual perception as a first step in our journey through this book is appropriate. Given the complexity and breadth of this topic, we can only aspire to cover the most rudimentary aspects of human vision. In particular, our interest lies in the mechanics and parameters related to how images are formed in the eye. We are interested in learning the physical limitations of human vision in terms of factors that also are used in our work with digital images. Thus, factors such as how human and electronic imaging compare in terms of resolution and ability to adapt to changes in illumination are not only interesting, they also are important from a practical point of view.

2.1.1 Structure of the Human Eye

Figure 2.1 shows a simplified horizontal cross section of the human eye. The eye is nearly a sphere, with an average diameter of approximately 20 mm. Three membranes enclose the eye: the *cornea* and *sclera* outer cover; the *choroid*; and the *retina*. The cornea is a tough, transparent tissue that covers the anterior

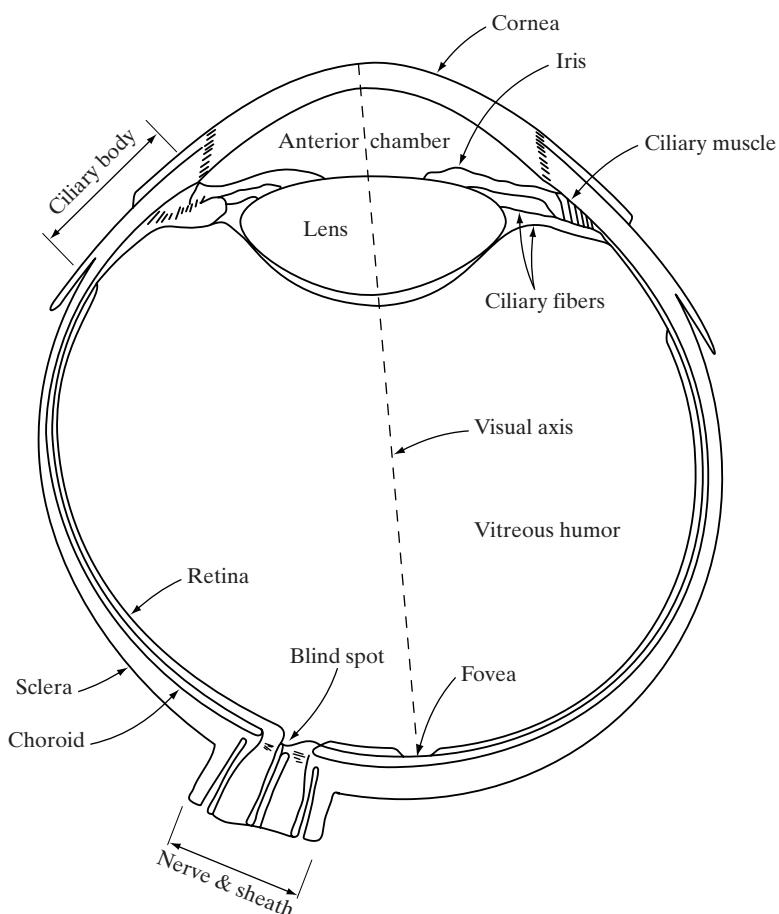


FIGURE 2.1
Simplified diagram of a cross section of the human eye.

surface of the eye. Continuous with the cornea, the sclera is an opaque membrane that encloses the remainder of the optic globe.

The choroid lies directly below the sclera. This membrane contains a network of blood vessels that serve as the major source of nutrition to the eye. Even superficial injury to the choroid, often not deemed serious, can lead to severe eye damage as a result of inflammation that restricts blood flow. The choroid coat is heavily pigmented and hence helps to reduce the amount of extraneous light entering the eye and the backscatter within the optical globe. At its anterior extreme, the choroid is divided into the *ciliary body* and the *iris diaphragm*. The latter contracts or expands to control the amount of light that enters the eye. The central opening of the iris (the *pupil*) varies in diameter from approximately 2 to 8 mm. The front of the iris contains the visible pigment of the eye, whereas the back contains a black pigment.

The *lens* is made up of concentric layers of fibrous cells and is suspended by fibers that attach to the ciliary body. It contains 60 to 70% water, about 6% fat, and more protein than any other tissue in the eye. The lens is colored by a slightly yellow pigmentation that increases with age. In extreme cases, excessive clouding of the lens, caused by the affliction commonly referred to as *cataracts*, can lead to poor color discrimination and loss of clear vision. The lens absorbs approximately 8% of the visible light spectrum, with relatively higher absorption at shorter wavelengths. Both infrared and ultraviolet light are absorbed appreciably by proteins within the lens structure and, in excessive amounts, can damage the eye.

The innermost membrane of the eye is the retina, which lines the inside of the wall's entire posterior portion. When the eye is properly focused, light from an object outside the eye is imaged on the retina. Pattern vision is afforded by the distribution of discrete light receptors over the surface of the retina. There are two classes of receptors: *cones* and *rods*. The cones in each eye number between 6 and 7 million. They are located primarily in the central portion of the retina, called the *fovea*, and are highly sensitive to color. Humans can resolve fine details with these cones largely because each one is connected to its own nerve end. Muscles controlling the eye rotate the eyeball until the image of an object of interest falls on the fovea. Cone vision is called *photopic* or bright-light vision.

The number of rods is much larger: Some 75 to 150 million are distributed over the retinal surface. The larger area of distribution and the fact that several rods are connected to a single nerve end reduce the amount of detail discernible by these receptors. Rods serve to give a general, overall picture of the field of view. They are not involved in color vision and are sensitive to low levels of illumination. For example, objects that appear brightly colored in daylight when seen by moonlight appear as colorless forms because only the rods are stimulated. This phenomenon is known as *scotopic* or dim-light vision.

Figure 2.2 shows the density of rods and cones for a cross section of the right eye passing through the region of emergence of the optic nerve from the eye. The absence of receptors in this area results in the so-called *blind spot* (see Fig. 2.1). Except for this region, the distribution of receptors is radially symmetric about the fovea. Receptor density is measured in degrees from the fovea (that is, in degrees off axis, as measured by the angle formed by the visual axis and a line passing through the center of the lens and intersecting the retina).

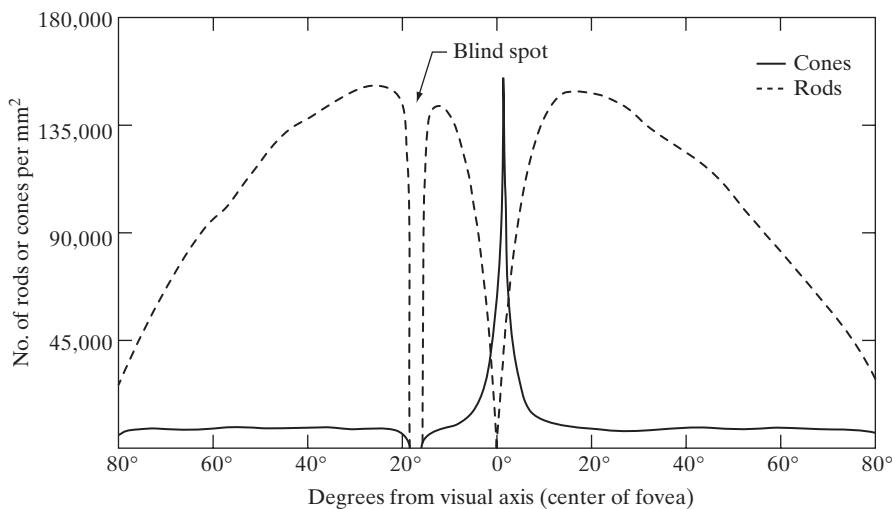


FIGURE 2.2
Distribution of rods and cones in the retina.

Note in Fig. 2.2 that cones are most dense in the center of the retina (in the center area of the fovea). Note also that rods increase in density from the center out to approximately 20° off axis and then decrease in density out to the extreme periphery of the retina.

The fovea itself is a circular indentation in the retina of about 1.5 mm in diameter. However, in terms of future discussions, talking about square or rectangular arrays of sensing elements is more useful. Thus, by taking some liberty in interpretation, we can view the fovea as a square sensor array of size $1.5\text{ mm} \times 1.5\text{ mm}$. The density of cones in that area of the retina is approximately 150,000 elements per mm^2 . Based on these approximations, the number of cones in the region of highest acuity in the eye is about 337,000 elements. Just in terms of raw resolving power, a charge-coupled device (CCD) imaging chip of medium resolution can have this number of elements in a receptor array no larger than $5\text{ mm} \times 5\text{ mm}$. While the ability of humans to integrate intelligence and experience with vision makes this type of comparison dangerous. Keep in mind for future discussions that the basic ability of the eye to resolve detail is certainly within the realm of current electronic imaging sensors.

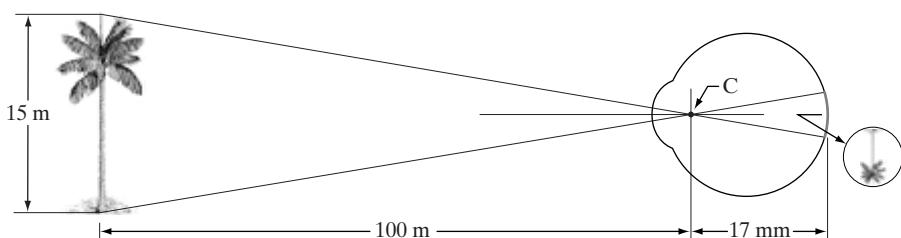
2.1.2 Image Formation in the Eye

The principal difference between the lens of the eye and an ordinary optical lens is that the former is flexible. As illustrated in Fig. 2.1, the radius of curvature of the anterior surface of the lens is greater than the radius of its posterior surface. The shape of the lens is controlled by tension in the fibers of the ciliary body. To focus on distant objects, the controlling muscles cause the lens to be relatively flattened. Similarly, these muscles allow the lens to become thicker in order to focus on objects near the eye.

The distance between the center of the lens and the retina (called the *focal length*) varies from approximately 17 mm to about 14 mm, as the refractive power of the lens increases from its minimum to its maximum. When the eye

FIGURE 2.3

Graphical representation of the eye looking at a palm tree. Point C is the optical center of the lens.



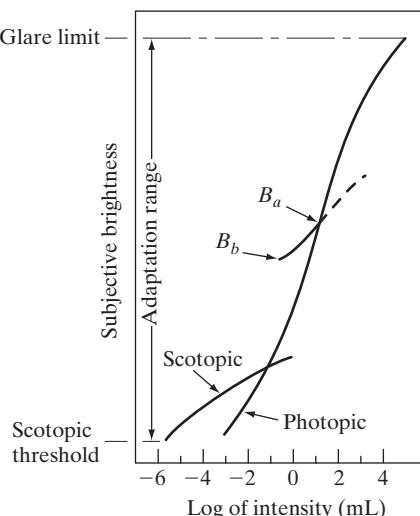
focuses on an object farther away than about 3 m, the lens exhibits its lowest refractive power. When the eye focuses on a nearby object, the lens is most strongly refractive. This information makes it easy to calculate the size of the retinal image of any object. In Fig. 2.3, for example, the observer is looking at a tree 15 m high at a distance of 100 m. If h is the height in mm of that object in the retinal image, the geometry of Fig. 2.3 yields $15/100 = h/17$ or $h = 2.55$ mm. As indicated in Section 2.1.1, the retinal image is reflected primarily in the area of the fovea. Perception then takes place by the relative excitation of light receptors, which transform radiant energy into electrical impulses that are ultimately decoded by the brain.

2.1.3 Brightness Adaptation and Discrimination

Because digital images are displayed as a discrete set of intensities, the eye's ability to discriminate between different intensity levels is an important consideration in presenting image-processing results. The range of light intensity levels to which the human visual system can adapt is enormous—on the order of 10^{10} —from the scotopic threshold to the glare limit. Experimental evidence indicates that *subjective brightness* (intensity as *perceived* by the human visual system) is a logarithmic function of the light intensity incident on the eye. Figure 2.4, a plot of light intensity versus subjective brightness, illustrates this char-

FIGURE 2.4

Range of subjective brightness sensations showing a particular adaptation level.



acteristic. The long solid curve represents the range of intensities to which the visual system can adapt. In photopic vision alone, the range is about 10^6 . The transition from scotopic to photopic vision is gradual over the approximate range from 0.001 to 0.1 millilambert (-3 to -1 mL in the log scale), as the double branches of the adaptation curve in this range show.

The essential point in interpreting the impressive dynamic range depicted in Fig. 2.4 is that the visual system cannot operate over such a range *simultaneously*. Rather, it accomplishes this large variation by changes in its overall sensitivity, a phenomenon known as *brightness adaptation*. The total range of distinct intensity levels it can discriminate simultaneously is rather small when compared with the total adaptation range. For any given set of conditions, the current sensitivity level of the visual system is called the *brightness adaptation level*, which may correspond, for example, to brightness B_a in Fig. 2.4. The short intersecting curve represents the range of subjective brightness that the eye can perceive when adapted to this level. This range is rather restricted, having a level B_b at and below which all stimuli are perceived as indistinguishable blacks. The upper (dashed) portion of the curve is not actually restricted but, if extended too far, loses its meaning because much higher intensities would simply raise the adaptation level higher than B_a .

The ability of the eye to discriminate between *changes* in light intensity at any specific adaptation level is also of considerable interest. A classic experiment used to determine the capability of the human visual system for brightness discrimination consists of having a subject look at a flat, uniformly illuminated area large enough to occupy the entire field of view. This area typically is a diffuser, such as opaque glass, that is illuminated from behind by a light source whose intensity, I , can be varied. To this field is added an increment of illumination, ΔI , in the form of a short-duration flash that appears as a circle in the center of the uniformly illuminated field, as Fig. 2.5 shows.

If ΔI is not bright enough, the subject says “no,” indicating no perceivable change. As ΔI gets stronger, the subject may give a positive response of “yes,” indicating a perceived change. Finally, when ΔI is strong enough, the subject will give a response of “yes” all the time. The quantity $\Delta I_c/I$, where ΔI_c is the increment of illumination discriminable 50% of the time with background illumination I , is called the *Weber ratio*. A small value of $\Delta I_c/I$, means that a small percentage change in intensity is discriminable. This represents “good” brightness discrimination. Conversely, a large value of $\Delta I_c/I$, means that a large percentage change in intensity is required. This represents “poor” brightness discrimination.

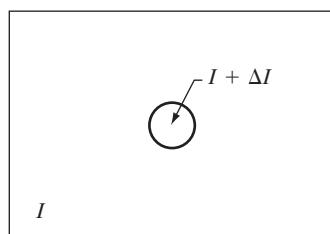
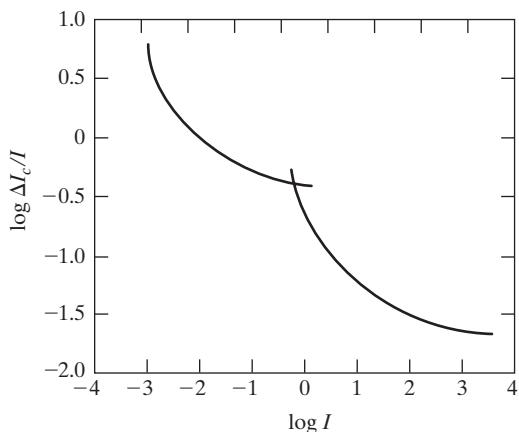


FIGURE 2.5 Basic experimental setup used to characterize brightness discrimination.

FIGURE 2.6
Typical Weber ratio as a function of intensity.



A plot of $\log \Delta I_c/I$, as a function of $\log I$ has the general shape shown in Fig. 2.6. This curve shows that brightness discrimination is poor (the Weber ratio is large) at low levels of illumination, and it improves significantly (the Weber ratio decreases) as background illumination increases. The two branches in the curve reflect the fact that at low levels of illumination vision is carried out by activity of the rods, whereas at high levels (showing better discrimination) vision is the function of cones.

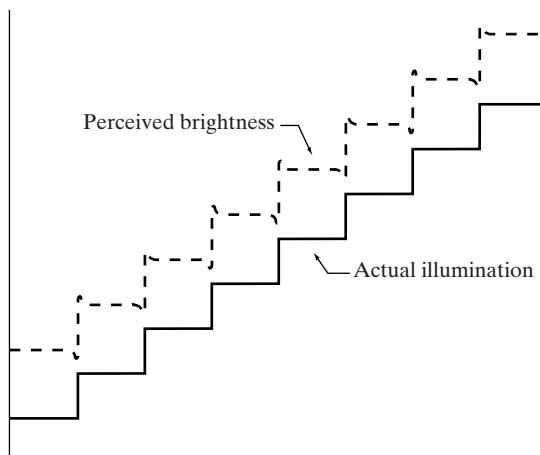
If the background illumination is held constant and the intensity of the other source, instead of flashing, is now allowed to vary incrementally from never being perceived to always being perceived, the typical observer can discern a total of one to two dozen different intensity changes. Roughly, this result is related to the number of different intensities a person can see at any one point in a monochrome image. This result does not mean that an image can be represented by such a small number of intensity values because, as the eye roams about the image, the average background changes, thus allowing a *different* set of incremental changes to be detected at each new adaptation level. The net consequence is that the eye is capable of a much broader range of *overall* intensity discrimination. In fact, we show in Section 2.4.3 that the eye is capable of detecting objectionable contouring effects in monochrome images whose overall intensity is represented by fewer than approximately two dozen levels.

Two phenomena clearly demonstrate that perceived brightness is not a simple function of intensity. The first is based on the fact that the visual system tends to undershoot or overshoot around the boundary of regions of different intensities. Figure 2.7(a) shows a striking example of this phenomenon. Although the intensity of the stripes is constant, we actually perceive a brightness pattern that is strongly scalloped, especially near the boundaries [Fig. 2.7(b)]. These seemingly scalloped bands are called *Mach bands* after Ernst Mach, who first described the phenomenon in 1865.

The second phenomenon, called *simultaneous contrast*, is related to the fact that a region's perceived brightness does not depend simply on its intensity, as Fig. 2.8 demonstrates. All the center squares have exactly the same intensity.

a
b

FIGURE 2.7
 (a) An example showing that perceived brightness is not a simple function of intensity. The relative vertical positions between the two profiles in (b) have no special significance; they were chosen for clarity.



However, they appear to the eye to become darker as the background gets lighter. A more familiar example is a piece of paper that seems white when lying on a desk, but can appear totally black when used to shield the eyes while looking directly at a bright sky.

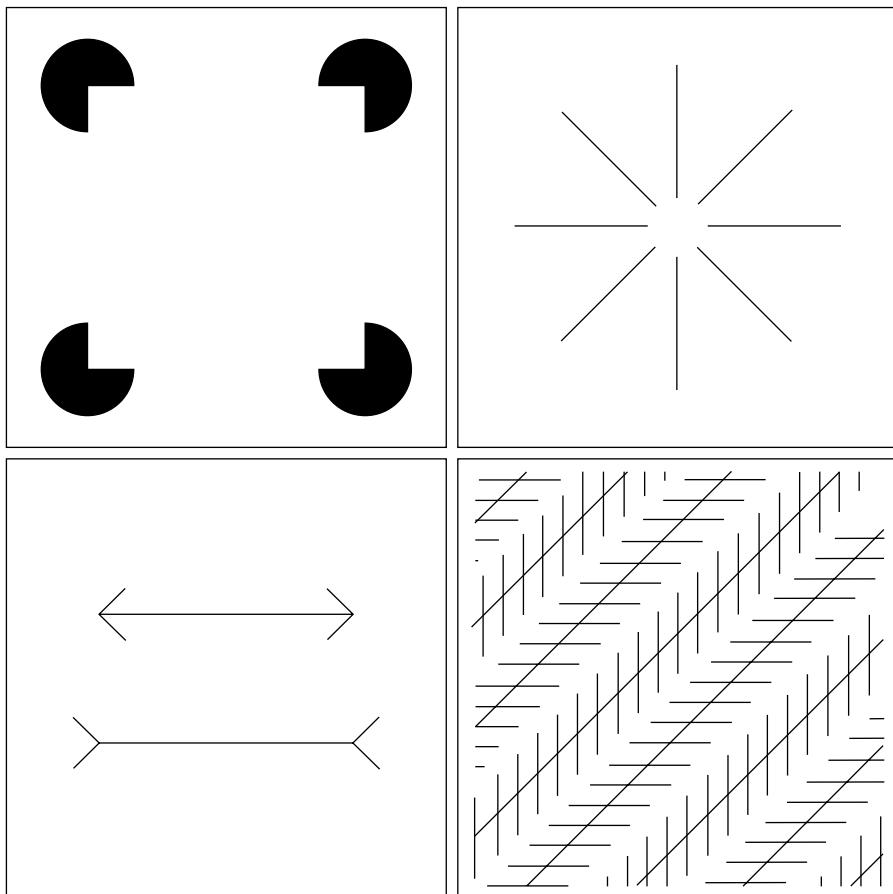


a b c

FIGURE 2.8 Examples of simultaneous contrast. All the inner squares have the same intensity, but they appear progressively darker as the background becomes lighter.

a
b
c
d

FIGURE 2.9 Some well-known optical illusions.



Other examples of human perception phenomena are optical illusions, in which the eye fills in nonexistent information or wrongly perceives geometrical properties of objects. Some examples are shown in Fig. 2.9. In Fig. 2.9(a), the outline of a square is seen clearly, in spite of the fact that no lines defining such a figure are part of the image. The same effect, this time with a circle, can be seen in Fig. 2.9(b); note how just a few lines are sufficient to give the illusion of a complete circle. The two horizontal line segments in Fig. 2.9(c) are of the same length, but one appears shorter than the other. Finally, all lines in Fig. 2.9(d) that are oriented at 45° are equidistant and parallel. Yet the crosshatching creates the illusion that those lines are far from being parallel. Optical illusions are a characteristic of the human visual system that is not fully understood.

2.2 Light and the Electromagnetic Spectrum

The electromagnetic spectrum was introduced in Section 1.3. We now consider this topic in more detail. In 1666, Sir Isaac Newton discovered that when a beam of sunlight is passed through a glass prism, the emerging beam of light is not

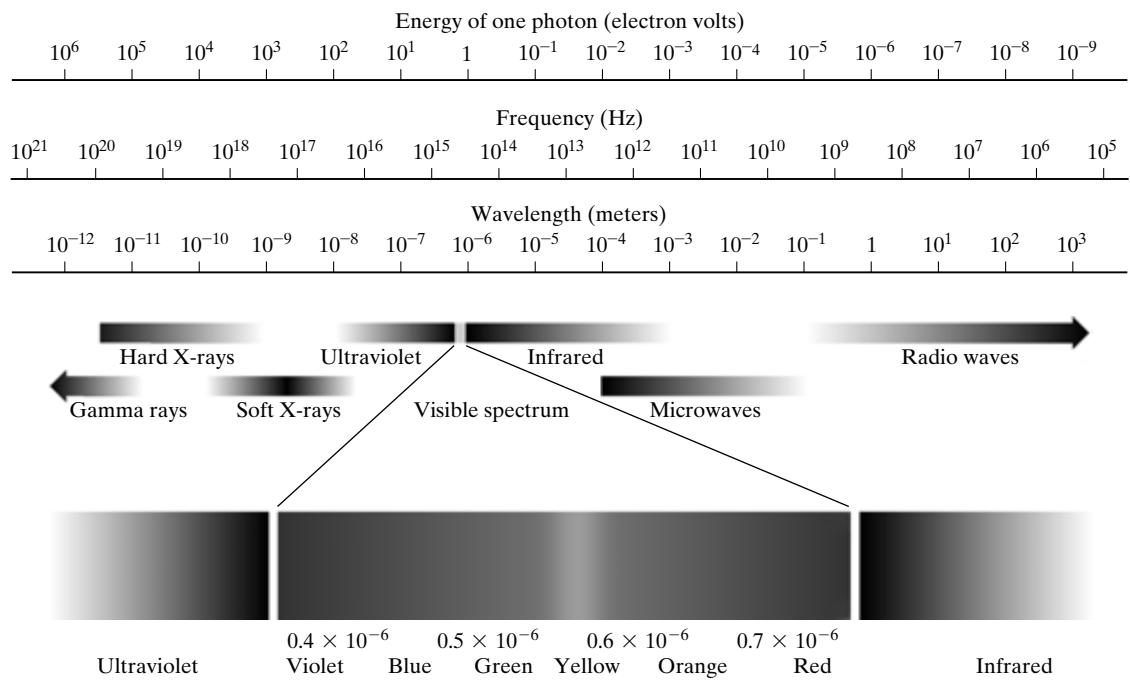


FIGURE 2.10 The electromagnetic spectrum. The visible spectrum is shown zoomed to facilitate explanation, but note that the visible spectrum is a rather narrow portion of the EM spectrum.

white but consists instead of a continuous spectrum of colors ranging from violet at one end to red at the other. As shown in Fig. 2.10, the range of colors we perceive in visible light represents a very small portion of the electromagnetic spectrum. On one end of the spectrum are radio waves with wavelengths billions of times longer than those of visible light. On the other end of the spectrum are gamma rays with wavelengths millions of times smaller than those of visible light. The electromagnetic spectrum can be expressed in terms of wavelength, frequency, or energy. Wavelength (λ) and frequency (ν) are related by the expression

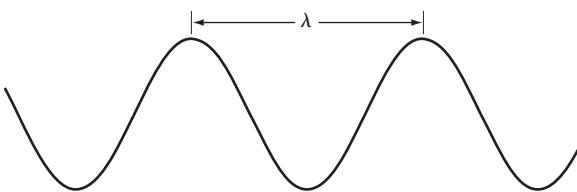
$$\lambda = \frac{c}{\nu} \quad (2.2-1)$$

where c is the speed of light (2.998×10^8 m/s). The energy of the various components of the electromagnetic spectrum is given by the expression

$$E = h\nu \quad (2.2-2)$$

where h is Planck's constant. The units of wavelength are meters, with the terms *microns* (denoted μm and equal to 10^{-6} m) and *nanometers* (10^{-9} m) being used just as frequently. Frequency is measured in Hertz (Hz), with one Hertz being equal to one cycle of a sinusoidal wave per second. A commonly used unit of energy is the electron-volt.

FIGURE 2.11
Graphical representation of one wavelength.



Electromagnetic waves can be visualized as propagating sinusoidal waves with wavelength λ (Fig. 2.11), or they can be thought of as a stream of massless particles, each traveling in a wavelike pattern and moving at the speed of light. Each massless particle contains a certain amount (or bundle) of energy. Each bundle of energy is called a *photon*. We see from Eq. (2.2-2) that energy is proportional to frequency, so the higher-frequency (shorter wavelength) electromagnetic phenomena carry more energy per photon. Thus, radio waves have photons with low energies, microwaves have more energy than radio waves, infrared still more, then visible, ultraviolet, X-rays, and finally gamma rays, the most energetic of all. This is the reason that gamma rays are so dangerous to living organisms.

Light is a particular type of electromagnetic radiation that can be seen and sensed by the human eye. The visible (color) spectrum is shown expanded in Fig. 2.10 for the purpose of discussion (we consider color in much more detail in Chapter 6). The visible band of the electromagnetic spectrum spans the range from approximately $0.43\text{ }\mu\text{m}$ (violet) to about $0.79\text{ }\mu\text{m}$ (red). For convenience, the color spectrum is divided into six broad regions: violet, blue, green, yellow, orange, and red. No color (or other component of the electromagnetic spectrum) ends abruptly, but rather each range blends smoothly into the next, as shown in Fig. 2.10.

The colors that humans perceive in an object are determined by the nature of the light *reflected* from the object. A body that reflects light and is relatively balanced in all visible wavelengths appears white to the observer. However, a body that favors reflectance in a limited range of the visible spectrum exhibits some shades of color. For example, green objects reflect light with wavelengths primarily in the 500 to 570 nm range while absorbing most of the energy at other wavelengths.

Light that is void of color is called *achromatic* or *monochromatic* light. The only attribute of such light is its *intensity*, or amount. The term *gray level* generally is used to describe monochromatic intensity because it ranges from black, to grays, and finally to white. Chromatic light spans the electromagnetic energy spectrum from approximately 0.43 to $0.79\text{ }\mu\text{m}$, as noted previously. Three basic quantities are used to describe the quality of a chromatic light source: radiance; luminance; and brightness. *Radiance* is the total amount of energy that flows from the light source, and it is usually measured in watts (W). *Luminance*, measured in lumens (lm), gives a measure of the amount of energy an observer *perceives* from a light source. For example, light emitted from a source operating in the far infrared region of the spectrum could have significant energy (radiance), but an observer would hardly perceive it; its luminance would be almost zero. Finally, as discussed in Section 2.1, *brightness* is a subjective descriptor of light perception that is practically impossible to measure. It embod-

ies the achromatic notion of intensity and is one of the key factors in describing color sensation.

Continuing with the discussion of Fig. 2.10, we note that at the short-wavelength end of the electromagnetic spectrum, we have gamma rays and hard X-rays. As discussed in Section 1.3.1, gamma radiation is important for medical and astronomical imaging, and for imaging radiation in nuclear environments. Hard (high-energy) X-rays are used in industrial applications. Chest X-rays are in the high end (shorter wavelength) of the soft X-rays region and dental X-rays are in the lower energy end of that band. The soft X-ray band transitions into the far ultraviolet light region, which in turn blends with the visible spectrum at longer wavelengths. Moving still higher in wavelength, we encounter the infrared band, which radiates heat, a fact that makes it useful in imaging applications that rely on “heat signatures.” The part of the infrared band close to the visible spectrum is called the *near-infrared* region. The opposite end of this band is called the *far-infrared* region. This latter region blends with the microwave band. This band is well known as the source of energy in microwave ovens, but it has many other uses, including communication and radar. Finally, the radio wave band encompasses television as well as AM and FM radio. In the higher energies, radio signals emanating from certain stellar bodies are useful in astronomical observations. Examples of images in most of the bands just discussed are given in Section 1.3.

In principle, if a sensor can be developed that is capable of detecting energy radiated by a band of the electromagnetic spectrum, we can image events of interest in that band. It is important to note, however, that the wavelength of an electromagnetic wave required to “see” an object must be of the same size as or smaller than the object. For example, a water molecule has a diameter on the order of 10^{-10} m. Thus, to study molecules, we would need a source capable of emitting in the far ultraviolet or soft X-ray region. This limitation, along with the physical properties of the sensor material, establishes the fundamental limits on the capability of imaging sensors, such as visible, infrared, and other sensors in use today.

Although imaging is based predominantly on energy radiated by electromagnetic waves, this is not the only method for image generation. For example, as discussed in Section 1.3.7, sound reflected from objects can be used to form ultrasonic images. Other major sources of digital images are electron beams for electron microscopy and synthetic images used in graphics and visualization.

2.3 Image Sensing and Acquisition

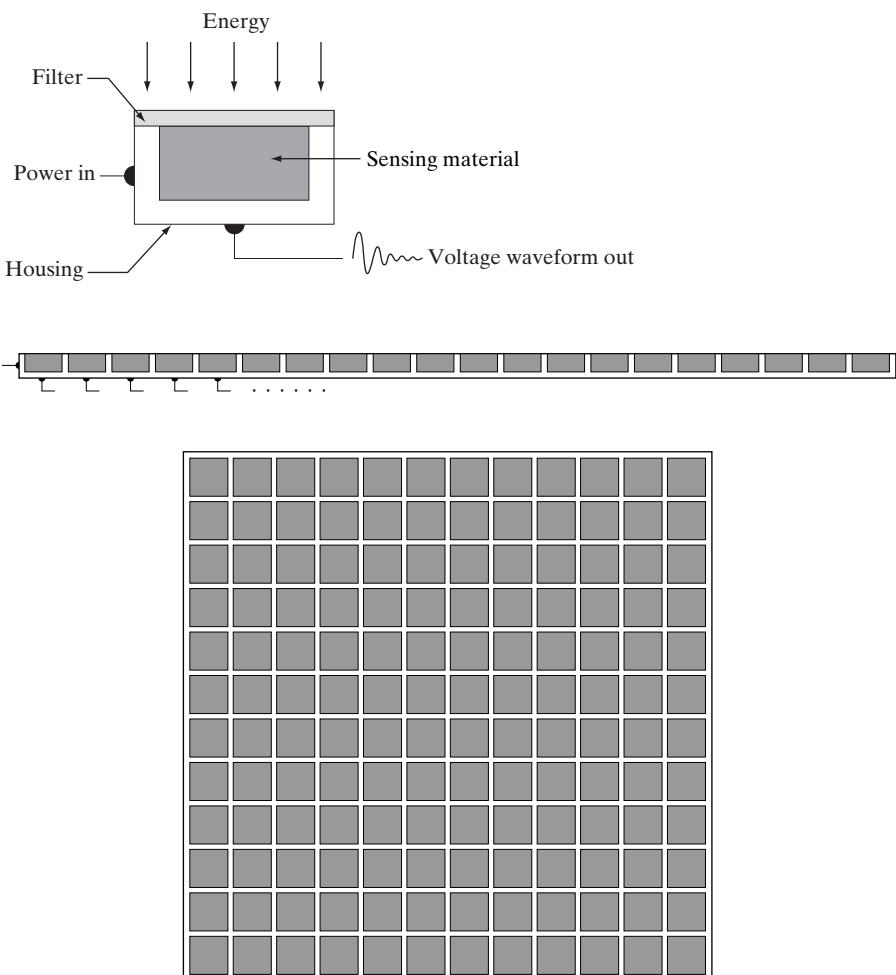
The types of images in which we are interested are generated by the combination of an “illumination” source and the reflection or absorption of energy from that source by the elements of the “scene” being imaged. We enclose *illumination* and *scene* in quotes to emphasize the fact that they are considerably more general than the familiar situation in which a visible light source illuminates a common everyday 3-D (three-dimensional) scene. For example, the illumination may originate from a source of electromagnetic energy such as radar, infrared,

or X-ray energy. But, as noted earlier, it could originate from less traditional sources, such as ultrasound or even a computer-generated illumination pattern. Similarly, the scene elements could be familiar objects, but they can just as easily be molecules, buried rock formations, or a human brain. We could even image a source, such as acquiring images of the sun. Depending on the nature of the source, illumination energy is reflected from, or transmitted through, objects. An example in the first category is light reflected from a planar surface. An example in the second category is when X-rays pass through a patient's body for the purpose of generating a diagnostic X-ray film. In some applications, the reflected or transmitted energy is focused onto a photoconverter (e.g., a phosphor screen), which converts the energy into visible light. Electron microscopy and some applications of gamma imaging use this approach.

Figure 2.12 shows the three principal sensor arrangements used to transform illumination energy into digital images. The idea is simple: Incoming energy is

a
b
c

FIGURE 2.12
(a) Single imaging sensor.
(b) Line sensor.
(c) Array sensor.



transformed into a voltage by the combination of input electrical power and sensor material that is responsive to the particular type of energy being detected. The output voltage waveform is the response of the sensor(s), and a digital quantity is obtained from each sensor by digitizing its response. In this section, we look at the principal modalities for image sensing and generation. Image digitizing is discussed in Section 2.4.

2.3.1 Image Acquisition Using a Single Sensor

Figure 2.12(a) shows the components of a single sensor. Perhaps the most familiar sensor of this type is the photodiode, which is constructed of silicon materials and whose output voltage waveform is proportional to light. The use of a filter in front of a sensor improves selectivity. For example, a green (pass) filter in front of a light sensor favors light in the green band of the color spectrum. As a consequence, the sensor output will be stronger for green light than for other components in the visible spectrum.

In order to generate a 2-D image using a single sensor, there has to be relative displacements in both the x - and y -directions between the sensor and the area to be imaged. Figure 2.13 shows an arrangement used in high-precision scanning, where a film negative is mounted onto a drum whose mechanical rotation provides displacement in one dimension. The single sensor is mounted on a lead screw that provides motion in the perpendicular direction. Since mechanical motion can be controlled with high precision, this method is an inexpensive (but slow) way to obtain high-resolution images. Other similar mechanical arrangements use a flat bed, with the sensor moving in two linear directions. These types of mechanical digitizers sometimes are referred to as *microdensitometers*.

Another example of imaging with a single sensor places a laser source coincident with the sensor. Moving mirrors are used to control the outgoing beam in a scanning pattern and to direct the reflected laser signal onto the sensor. This arrangement also can be used to acquire images using strip and array sensors, which are discussed in the following two sections.

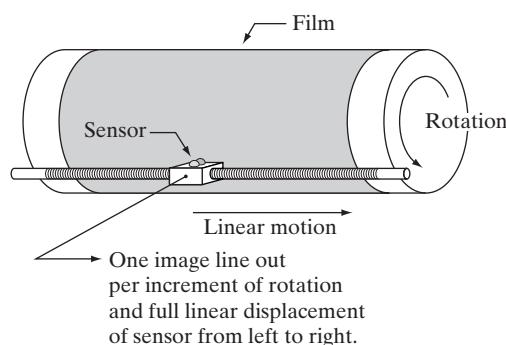
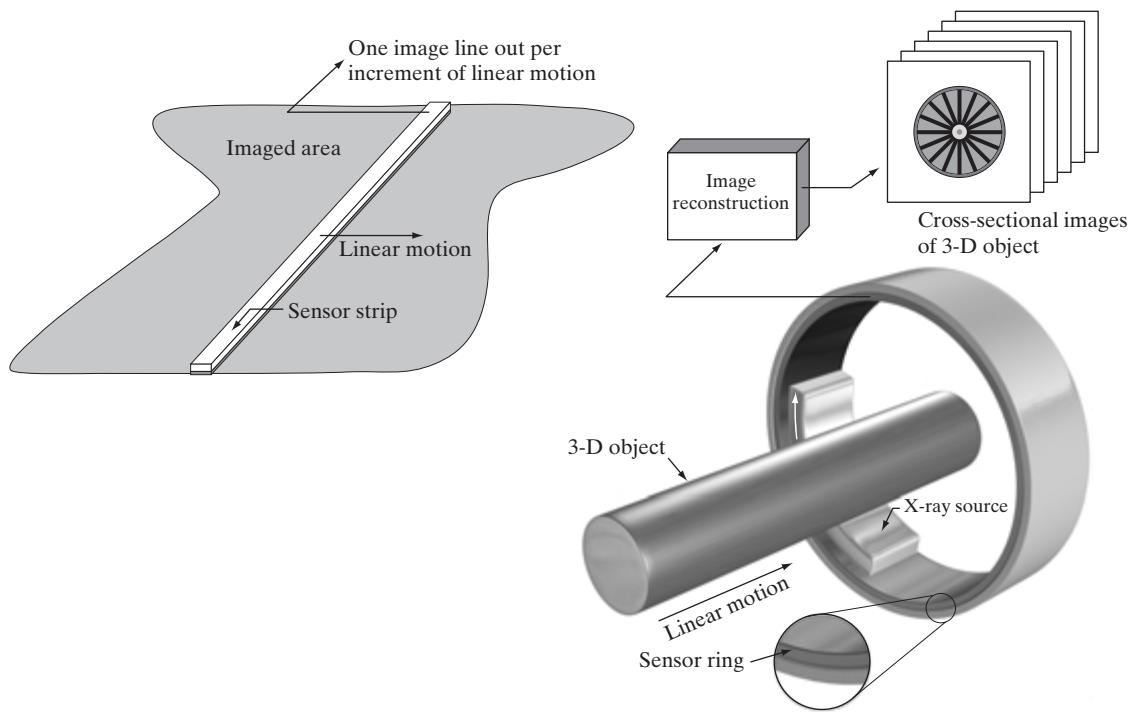


FIGURE 2.13 Combining a single sensor with motion to generate a 2-D image.

2.3.2 Image Acquisition Using Sensor Strips

A geometry that is used much more frequently than single sensors consists of an in-line arrangement of sensors in the form of a sensor strip, as Fig. 2.12(b) shows. The strip provides imaging elements in one direction. Motion perpendicular to the strip provides imaging in the other direction, as shown in Fig. 2.14(a). This is the type of arrangement used in most flat bed scanners. Sensing devices with 4000 or more in-line sensors are possible. In-line sensors are used routinely in airborne imaging applications, in which the imaging system is mounted on an aircraft that flies at a constant altitude and speed over the geographical area to be imaged. One-dimensional imaging sensor strips that respond to various bands of the electromagnetic spectrum are mounted perpendicular to the direction of flight. The imaging strip gives one line of an image at a time, and the motion of the strip completes the other dimension of a two-dimensional image. Lenses or other focusing schemes are used to project the area to be scanned onto the sensors.

Sensor strips mounted in a ring configuration are used in medical and industrial imaging to obtain cross-sectional (“slice”) images of 3-D objects, as Fig. 2.14(b) shows. A rotating X-ray source provides illumination and the por-



a b

FIGURE 2.14 (a) Image acquisition using a linear sensor strip. (b) Image acquisition using a circular sensor strip.

tion of the sensors opposite the source collect the X-ray energy that pass through the object (the sensors obviously have to be sensitive to X-ray energy). This is the basis for medical and industrial computerized axial tomography (CAT) imaging as indicated in Sections 1.2 and 1.3.2. It is important to note that the output of the sensors must be processed by reconstruction algorithms whose objective is to transform the sensed data into meaningful cross-sectional images. In other words, images are not obtained directly from the sensors by motion alone; they require extensive processing. A 3-D digital volume consisting of stacked images is generated as the object is moved in a direction perpendicular to the sensor ring. Other modalities of imaging based on the CAT principle include magnetic resonance imaging (MRI) and positron emission tomography (PET). The illumination sources, sensors, and types of images are different, but conceptually they are very similar to the basic imaging approach shown in Fig. 2.14(b).

2.3.3 Image Acquisition Using Sensor Arrays

Figure 2.12(c) shows individual sensors arranged in the form of a 2-D array. Numerous electromagnetic and some ultrasonic sensing devices frequently are arranged in an array format. This is also the predominant arrangement found in digital cameras. A typical sensor for these cameras is a CCD array, which can be manufactured with a broad range of sensing properties and can be packaged in rugged arrays of 4000×4000 elements or more. CCD sensors are used widely in digital cameras and other light sensing instruments. The response of each sensor is proportional to the integral of the light energy projected onto the surface of the sensor, a property that is used in astronomical and other applications requiring low noise images. Noise reduction is achieved by letting the sensor integrate the input light signal over minutes or even hours (we discuss noise reduction by integration in Chapter 3). Since the sensor array shown in Fig. 2.15(c) is two dimensional, its key advantage is that a complete image can be obtained by focusing the energy pattern onto the surface of the array. Motion obviously is not necessary, as is the case with the sensor arrangements discussed in the preceding two sections.

The principal manner in which array sensors are used is shown in Fig. 2.15. This figure shows the energy from an illumination source being reflected from a scene element, but, as mentioned at the beginning of this section, the energy also could be transmitted through the scene elements. The first function performed by the imaging system shown in Fig. 2.15(c) is to collect the incoming energy and focus it onto an image plane. If the illumination is light, the front end of the imaging system is a lens, which projects the viewed scene onto the lens focal plane, as Fig. 2.15(d) shows. The sensor array, which is coincident with the focal plane, produces outputs proportional to the integral of the light received at each sensor. Digital and analog circuitry sweep these outputs and convert them to a video signal, which is then digitized by another section of the imaging system. The output is a digital image, as shown diagrammatically in Fig. 2.15(e). Conversion of an image into digital form is the topic of Section 2.4.

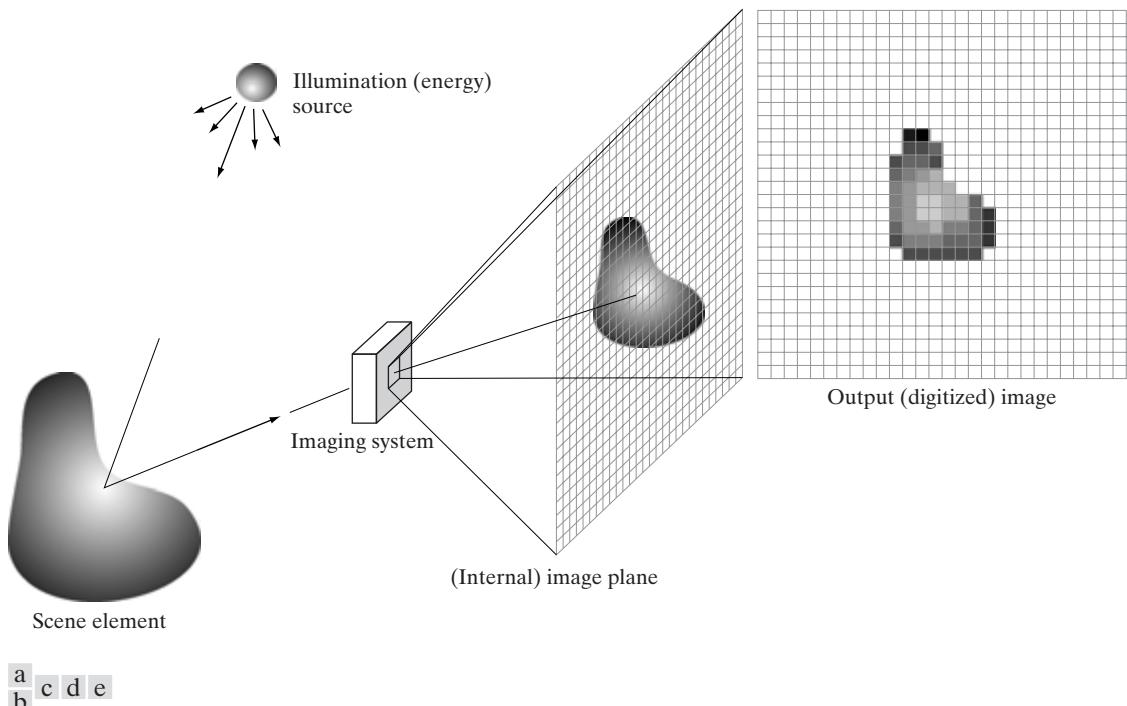


FIGURE 2.15 An example of the digital image acquisition process. (a) Energy (“illumination”) source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

2.3.4 A Simple Image Formation Model

As introduced in Section 1.1, we shall denote images by two-dimensional functions of the form $f(x, y)$. The value or amplitude of f at spatial coordinates (x, y) is a positive scalar quantity whose physical meaning is determined by the source of the image. Most of the images in which we are interested in this book are monochromatic images, whose values are said to span the gray scale, as discussed in Section 2.2. When an image is generated from a physical process, its values are proportional to energy radiated by a physical source (e.g., electromagnetic waves). As a consequence, $f(x, y)$ must be nonzero and finite; that is,

$$0 < f(x, y) < \infty. \quad (2.3-1)$$

The function $f(x, y)$ may be characterized by two components: (1) the amount of source illumination incident on the scene being viewed, and (2) the amount of illumination reflected by the objects in the scene. Appropriately, these are called the *illumination* and *reflectance* components and are denoted by $i(x, y)$ and $r(x, y)$, respectively. The two functions combine as a product to form $f(x, y)$:

$$f(x, y) = i(x, y)r(x, y) \quad (2.3-2)$$

where

$$0 < i(x, y) < \infty \quad (2.3-3)$$

and

$$0 < r(x, y) < 1. \quad (2.3-4)$$

Equation (2.3-4) indicates that reflectance is bounded by 0 (total absorption) and 1 (total reflectance). The nature of $i(x, y)$ is determined by the illumination source, and $r(x, y)$ is determined by the characteristics of the imaged objects. It is noted that these expressions also are applicable to images formed via transmission of the illumination through a medium, such as a chest X-ray. In this case, we would deal with a *transmissivity* instead of a *reflectivity* function, but the limits would be the same as in Eq. (2.3-4), and the image function formed would be modeled as the product in Eq. (2.3-2).

■ The values given in Eqs. (2.3-3) and (2.3-4) are theoretical bounds. The following *average* numerical figures illustrate some typical ranges of $i(x, y)$ for visible light. On a clear day, the sun may produce in excess of 90,000 lm/m² of illumination on the surface of the Earth. This figure decreases to less than 10,000 lm/m² on a cloudy day. On a clear evening, a full moon yields about 0.1 lm/m² of illumination. The typical illumination level in a commercial office is about 1000 lm/m². Similarly, the following are some typical values of $r(x, y)$: 0.01 for black velvet, 0.65 for stainless steel, 0.80 for flat-white wall paint, 0.90 for silver-plated metal, and 0.93 for snow. ■

EXAMPLE 2.1:
Some typical
values of
illumination and
reflectance.

As noted in Section 2.2, we call the intensity of a monochrome image at any coordinates (x_0, y_0) the *gray level* (ℓ) of the image at that point. That is,

$$\ell = f(x_0, y_0) \quad (2.3-5)$$

From Eqs. (2.3-2) through (2.3-4), it is evident that ℓ lies in the range

$$L_{\min} \leq \ell \leq L_{\max} \quad (2.3-6)$$

In theory, the only requirement on L_{\min} is that it be positive, and on L_{\max} that it be finite. In practice, $L_{\min} = i_{\min}r_{\min}$ and $L_{\max} = i_{\max}r_{\max}$. Using the preceding average office illumination and range of reflectance values as guidelines, we may expect $L_{\min} \approx 10$ and $L_{\max} \approx 1000$ to be typical limits for indoor values in the absence of additional illumination.

The interval $[L_{\min}, L_{\max}]$ is called the *gray scale*. Common practice is to shift this interval numerically to the interval $[0, L - 1]$, where $\ell = 0$ is considered black and $\ell = L - 1$ is considered white on the gray scale. All intermediate values are shades of gray varying from black to white.

2.4 Image Sampling and Quantization

From the discussion in the preceding section, we see that there are numerous ways to acquire images, but our objective in all is the same: to generate digital images from sensed data. The output of most sensors is a continuous voltage waveform whose amplitude and spatial behavior are related to the physical phenomenon being sensed. To create a digital image, we need to convert the continuous sensed data into digital form. This involves two processes: *sampling* and *quantization*.

2.4.1 Basic Concepts in Sampling and Quantization

The basic idea behind sampling and quantization is illustrated in Fig. 2.16. Figure 2.16(a) shows a continuous image, $f(x, y)$, that we want to convert to digital form. An image may be continuous with respect to the x - and y -coordinates, and also in amplitude. To convert it to digital form, we have to sample the function in both coordinates and in amplitude. **Digitizing the coordinate values is called *sampling*. Digitizing the amplitude values is called *quantization*.**

The one-dimensional function shown in Fig. 2.16(b) is a plot of amplitude (gray level) values of the continuous image along the line segment AB in Fig. 2.16(a). The random variations are due to image noise. To sample this function, we take equally spaced samples along line AB , as shown in Fig. 2.16(c). The location of each sample is given by a vertical tick mark in the bottom part of the figure. The samples are shown as small white squares superimposed on the function. The set of these discrete locations gives the sampled function. However, the values of the samples still span (vertically) a continuous range of gray-level values. In order to form a digital function, the gray-level values also must be converted (*quantized*) into discrete quantities. The right side of Fig. 2.16(c) shows the gray-level scale divided into eight discrete levels, ranging from black to white. The vertical tick marks indicate the specific value assigned to each of the eight gray levels. The continuous gray levels are quantized simply by assigning one of the eight discrete gray levels to each sample. The assignment is made depending on the vertical proximity of a sample to a vertical tick mark. The digital samples resulting from both sampling and quantization are shown in Fig. 2.16(d). Starting at the top of the image and carrying out this procedure line by line produces a two-dimensional digital image.

Sampling in the manner just described assumes that we have a continuous image in both coordinate directions as well as in amplitude. In practice, the method of sampling is determined by the sensor arrangement used to generate the image. When an image is generated by a single sensing element combined with mechanical motion, as in Fig. 2.13, the output of the sensor is quantized in the manner described above. However, sampling is accomplished by selecting the number of individual mechanical increments at which we activate the sensor to collect data. Mechanical motion can be made very exact so, in principle, there is almost no limit as to how fine we can sample an image. However, practical limits are established by imperfections in the optics used to focus on the

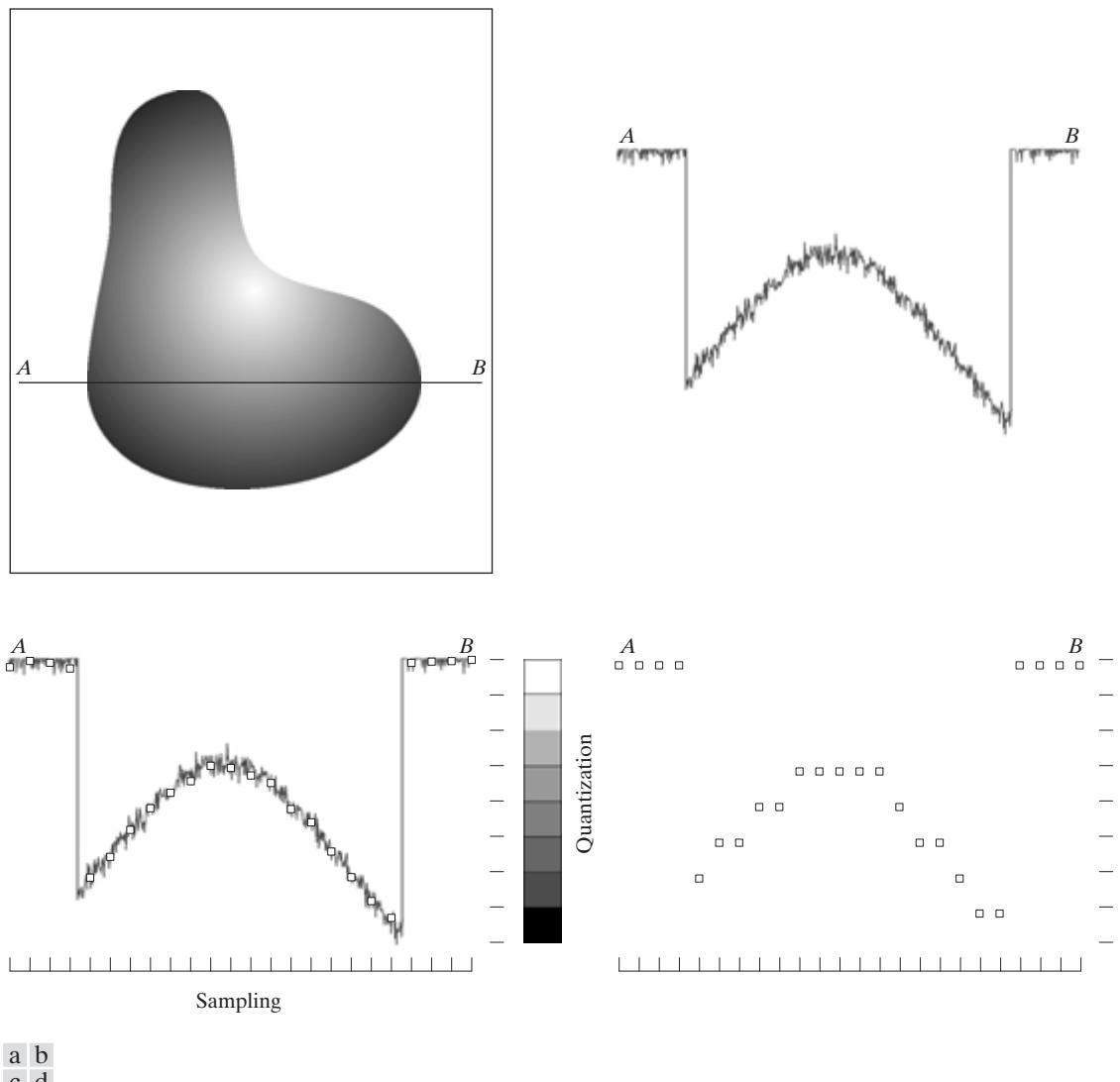


FIGURE 2.16 Generating a digital image. (a) Continuous image. (b) A scan line from A to B in the continuous image, used to illustrate the concepts of sampling and quantization. (c) Sampling and quantization. (d) Digital scan line.

sensor an illumination spot that is inconsistent with the fine resolution achievable with mechanical displacements.

When a sensing strip is used for image acquisition, the number of sensors in the strip establishes the sampling limitations in one image direction. Mechanical motion in the other direction can be controlled more accurately, but it makes little sense to try to achieve sampling density in one direction that exceeds the

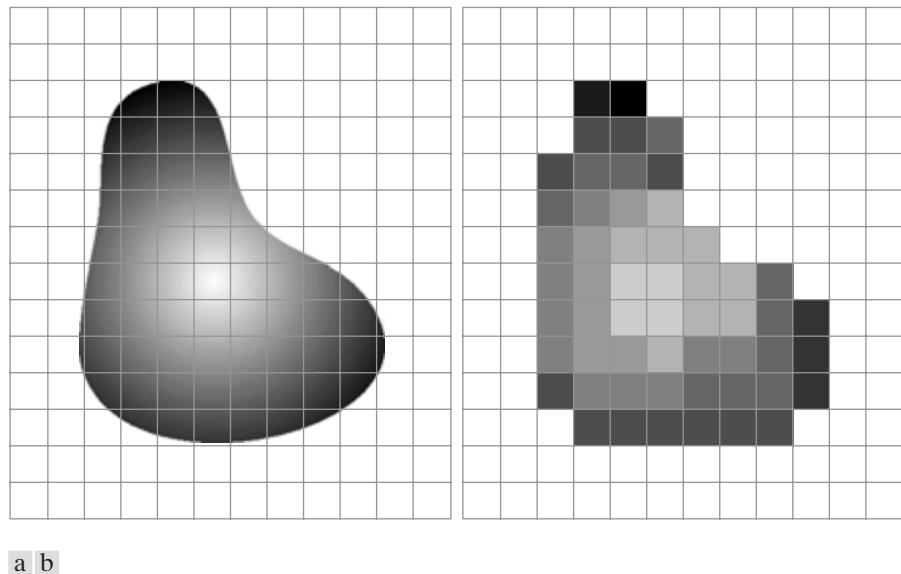


FIGURE 2.17 (a) Continuous image projected onto a sensor array. (b) Result of image sampling and quantization.

sampling limits established by the number of sensors in the other. Quantization of the sensor outputs completes the process of generating a digital image.

When a sensing array is used for image acquisition, there is no motion and the number of sensors in the array establishes the limits of sampling in both directions. Quantization of the sensor outputs is as before. Figure 2.17 illustrates this concept. Figure 2.17(a) shows a continuous image projected onto the plane of an array sensor. Figure 2.17(b) shows the image after sampling and quantization. Clearly, the quality of a digital image is determined to a large degree by the number of samples and discrete gray levels used in sampling and quantization. However, as shown in Section 2.4.3, image content is an important consideration in choosing these parameters.

2.4.2 Representing Digital Images

The result of sampling and quantization is a matrix of real numbers. We will use two principal ways in this book to represent digital images. Assume that an image $f(x, y)$ is sampled so that the resulting digital image has M rows and N columns. The values of the coordinates (x, y) now become *discrete* quantities. For notational clarity and convenience, we shall use integer values for these discrete coordinates. Thus, the values of the coordinates at the origin are $(x, y) = (0, 0)$. The next coordinate values along the first row of the image are represented as $(x, y) = (0, 1)$. It is important to keep in mind that the notation $(0, 1)$ is used to signify the second sample along the first row. It does *not* mean that these are the actual values of physical coordinates when the image was sampled. Figure 2.18 shows the coordinate convention used throughout this book.

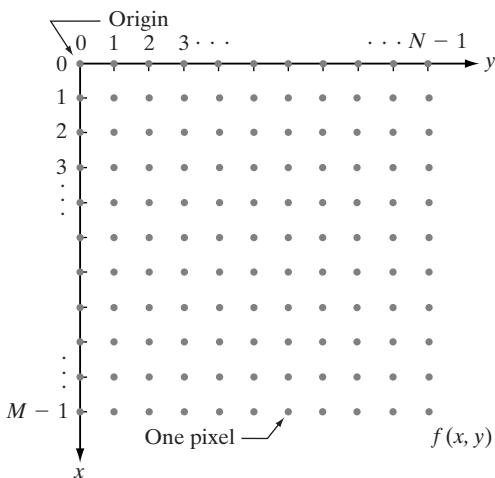


FIGURE 2.18
Coordinate convention used in this book to represent digital images.

The notation introduced in the preceding paragraph allows us to write the complete $M \times N$ digital image in the following compact matrix form:

$$f(x, y) = \begin{bmatrix} f(0, 0) & f(0, 1) & \cdots & f(0, N - 1) \\ f(1, 0) & f(1, 1) & \cdots & f(1, N - 1) \\ \vdots & \vdots & & \vdots \\ f(M - 1, 0) & f(M - 1, 1) & \cdots & f(M - 1, N - 1) \end{bmatrix}. \quad (2.4-1)$$

The right side of this equation is by definition a digital image. Each element of this matrix array is called an *image element*, *picture element*, *pixel*, or *pel*. The terms *image* and *pixel* will be used throughout the rest of our discussions to denote a digital image and its elements.

In some discussions, it is advantageous to use a more traditional matrix notation to denote a digital image and its elements:

$$\mathbf{A} = \begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,N-1} \\ \vdots & \vdots & & \vdots \\ a_{M-1,0} & a_{M-1,1} & \cdots & a_{M-1,N-1} \end{bmatrix}. \quad (2.4-2)$$

Clearly, $a_{ij} = f(x = i, y = j) = f(i, j)$, so Eqs. (2.4-1) and (2.4-2) are identical matrices.

Expressing sampling and quantization in more formal mathematical terms can be useful at times. Let Z and R denote the set of real integers and the set of real numbers, respectively. The sampling process may be viewed as partitioning the xy plane into a grid, with the coordinates of the center of each grid being a pair of elements from the Cartesian product Z^2 , which is the set of all ordered pairs of elements (z_i, z_j) , with z_i and z_j being integers from Z . Hence, $f(x, y)$ is a digital image if (x, y) are integers from Z^2 and f is a function that assigns a gray-level value (that is, a real number from the set of real numbers, R) to each distinct pair of coordinates (x, y) . This functional assignment

obviously is the quantization process described earlier. If the gray levels also are integers (as usually is the case in this and subsequent chapters), Z replaces R , and a digital image then becomes a 2-D function whose coordinates and amplitude values are integers.

This digitization process requires decisions about values for M , N , and for the number, L , of discrete gray levels allowed for each pixel. There are no requirements on M and N , other than that they have to be positive integers. However, due to processing, storage, and sampling hardware considerations, the number of gray levels typically is an integer power of 2:

$$L = 2^k. \quad (2.4-3)$$

We assume that the discrete levels are equally spaced and that they are integers in the interval $[0, L - 1]$. Sometimes the range of values spanned by the gray scale is called the *dynamic range* of an image, and we refer to images whose gray levels span a significant portion of the gray scale as having a high dynamic range. When an appreciable number of pixels exhibit this property, the image will have high contrast. Conversely, an image with low dynamic range tends to have a dull, washed out gray look. This is discussed in much more detail in Section 3.3.

The number, b , of bits required to store a digitized image is

$$b = M \times N \times k. \quad (2.4-4)$$

When $M = N$, this equation becomes

$$b = N^2k. \quad (2.4-5)$$

Table 2.1 shows the number of bits required to store square images with various values of N and k . The number of gray levels corresponding to each value of k is shown in parentheses. When an image can have 2^k gray levels, it is common practice to refer to the image as a “ k -bit image.” For example, an image with 256 possible gray-level values is called an 8-bit image. Note that storage requirements for 8-bit images of size 1024×1024 and higher are not insignificant.

TABLE 2.1

Number of storage bits for various values of N and k .

N/k	1 ($L = 2$)	2 ($L = 4$)	3 ($L = 8$)	4 ($L = 16$)	5 ($L = 32$)	6 ($L = 64$)	7 ($L = 128$)	8 ($L = 256$)
32	1,024	2,048	3,072	4,096	5,120	6,144	7,168	8,192
64	4,096	8,192	12,288	16,384	20,480	24,576	28,672	32,768
128	16,384	32,768	49,152	65,536	81,920	98,304	114,688	131,072
256	65,536	131,072	196,608	262,144	327,680	393,216	458,752	524,288
512	262,144	524,288	786,432	1,048,576	1,310,720	1,572,864	1,835,008	2,097,152
1024	1,048,576	2,097,152	3,145,728	4,194,304	5,242,880	6,291,456	7,340,032	8,388,608
2048	4,194,304	8,388,608	12,582,912	16,777,216	20,971,520	25,165,824	29,369,128	33,554,432
4096	16,777,216	33,554,432	50,331,648	67,108,864	83,886,080	100,663,296	117,440,512	134,217,728
8192	67,108,864	134,217,728	201,326,592	268,435,456	335,544,320	402,653,184	469,762,048	536,870,912

2.4.3 Spatial and Gray-Level Resolution

Sampling is the principal factor determining the *spatial resolution* of an image. Basically, **spatial resolution** is the smallest discernible detail in an image. Suppose that we construct a chart with vertical lines of width W , with the space between the lines also having width W . A *line pair* consists of one such line and its adjacent space. Thus, the width of a line pair is $2W$, and there are $1/2W$ line pairs per unit distance. A widely used definition of resolution is simply the smallest number of discernible line pairs per unit distance; for example, 100 line pairs per millimeter.

Gray-level resolution similarly refers to the smallest discernible change in gray level, but, as noted in Section 2.1.3, measuring discernible changes in gray level is a highly subjective process. We have considerable discretion regarding the number of samples used to generate a digital image, but this is not true for the number of gray levels. Due to hardware considerations, **the number of gray levels is usually an integer power of 2**, as mentioned in the previous section. The most common number is 8 bits, with 16 bits being used in some applications where enhancement of specific gray-level ranges is necessary. Sometimes we find systems that can digitize the gray levels of an image with 10 or 12 bits of accuracy, but these are the exception rather than the rule.

When an actual measure of physical resolution relating pixels and the level of detail they resolve in the original scene are not necessary, it is not uncommon to refer to an L -level digital image of size $M \times N$ as having a spatial resolution of $M \times N$ pixels and a gray-level resolution of L levels. We will use this terminology from time to time in subsequent discussions, making a reference to actual resolvable detail only when necessary for clarity.

■ Figure 2.19 shows an image of size 1024×1024 pixels whose gray levels are represented by 8 bits. The other images shown in Fig. 2.19 are the results of

EXAMPLE 2.2:
Typical effects of
varying the
number of
samples in a
digital image.

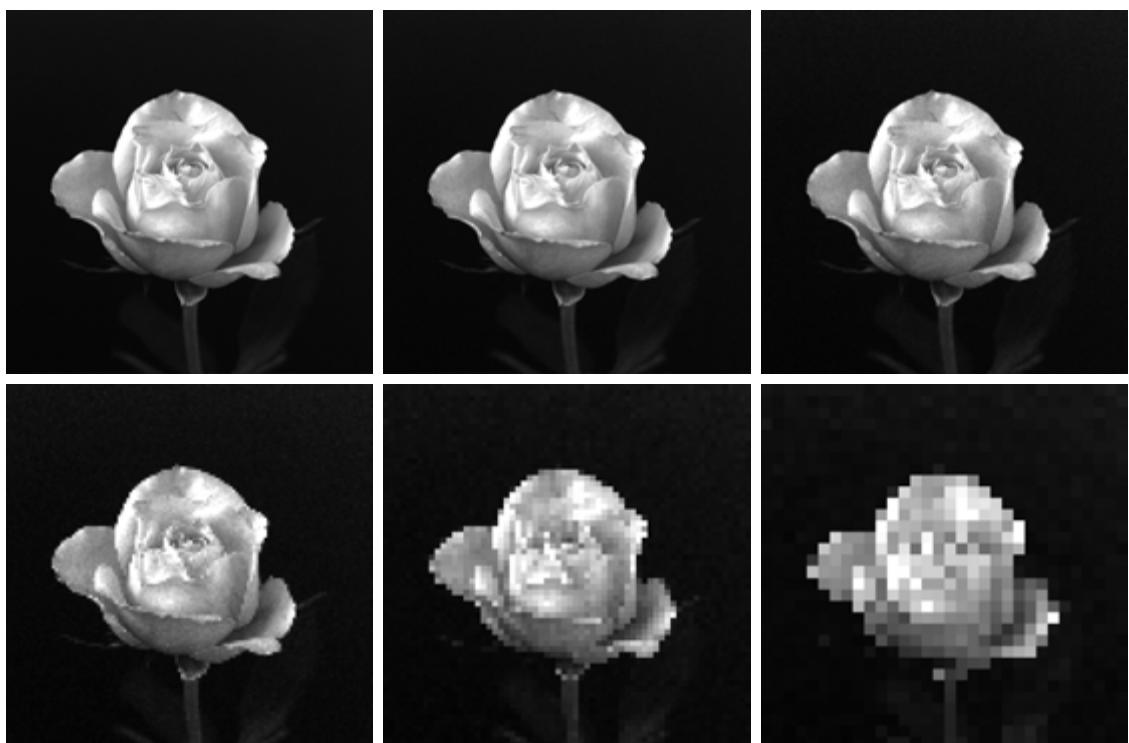


FIGURE 2.19 A 1024×1024 , 8-bit image subsampled down to size 32×32 pixels. The number of allowable gray levels was kept at 256.

subsampling the 1024×1024 image. The subsampling was accomplished by deleting the appropriate number of rows and columns from the original image. For example, the 512×512 image was obtained by deleting every other row and column from the 1024×1024 image. The 256×256 image was generated by deleting every other row and column in the 512×512 image, and so on. The number of allowed gray levels was kept at 256.

These images show the dimensional proportions between various sampling densities, but their size differences make it difficult to see the effects resulting from a reduction in the number of samples. The simplest way to compare these effects is to bring all the subsampled images up to size 1024×1024 by row and column pixel replication. The results are shown in Figs. 2.20(b) through (f). Figure 2.20(a) is the same 1024×1024 , 256-level image shown in Fig. 2.19; it is repeated to facilitate comparisons.

Compare Fig. 2.20(a) with the 512×512 image in Fig. 2.20(b) and note that it is virtually impossible to tell these two images apart. The level of detail lost is simply too fine to be seen on the printed page at the scale in which these im-



a	b	c
d	e	f

FIGURE 2.20 (a) 1024×1024 , 8-bit image. (b) 512×512 image resampled into 1024×1024 pixels by row and column duplication. (c) through (f) 256×256 , 128×128 , 64×64 , and 32×32 images resampled into 1024×1024 pixels.

ages are shown. Next, the 256×256 image in Fig. 2.20(c) shows a very slight fine checkerboard pattern in the borders between flower petals and the black background. A slightly more pronounced graininess throughout the image also is beginning to appear. These effects are much more visible in the 128×128 image in Fig. 2.20(d), and they become pronounced in the 64×64 and 32×32 images in Figs. 2.20(e) and (f), respectively.

■ In this example, we keep the number of samples constant and reduce the number of gray levels from 256 to 2, in integer powers of 2. Figure 2.21(a) is a 452×374 CAT projection image, displayed with $k = 8$ (256 gray levels). Images such as this are obtained by fixing the X-ray source in one position, thus producing a 2-D image

EXAMPLE 2.3:
Typical effects of
varying the
number of gray
levels in a digital
image.

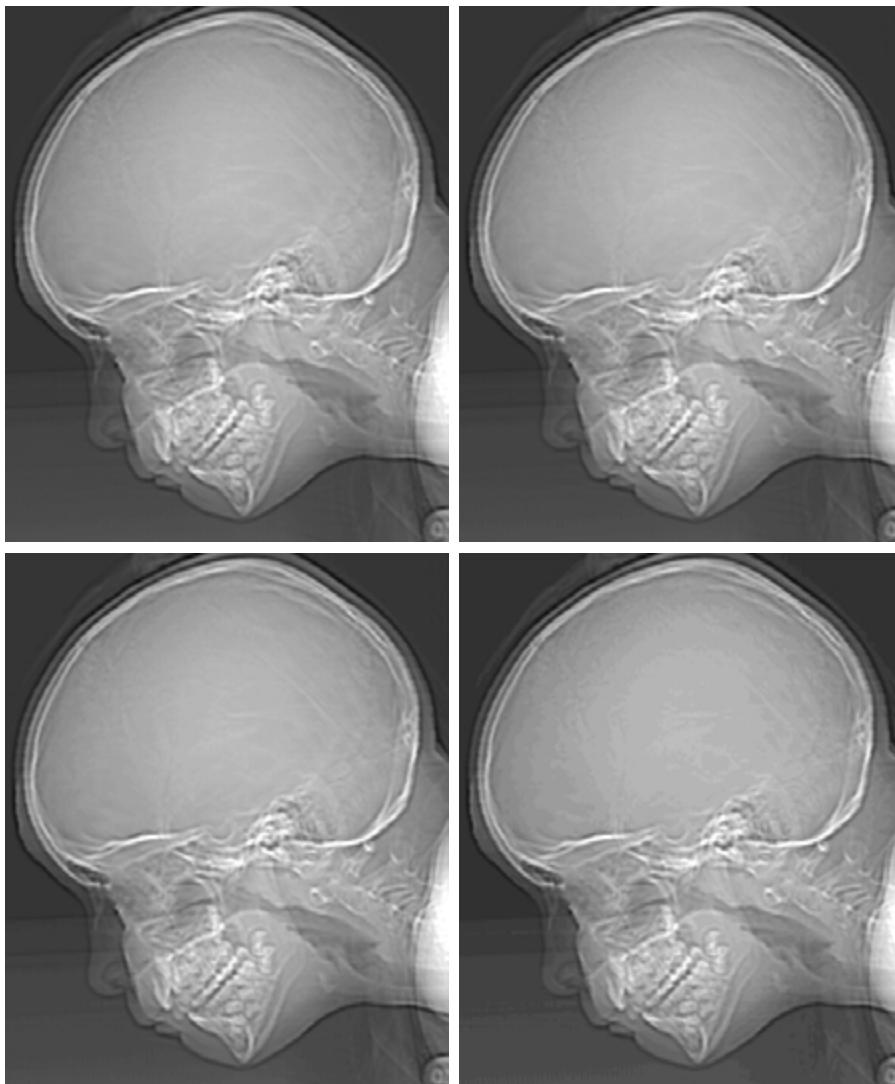


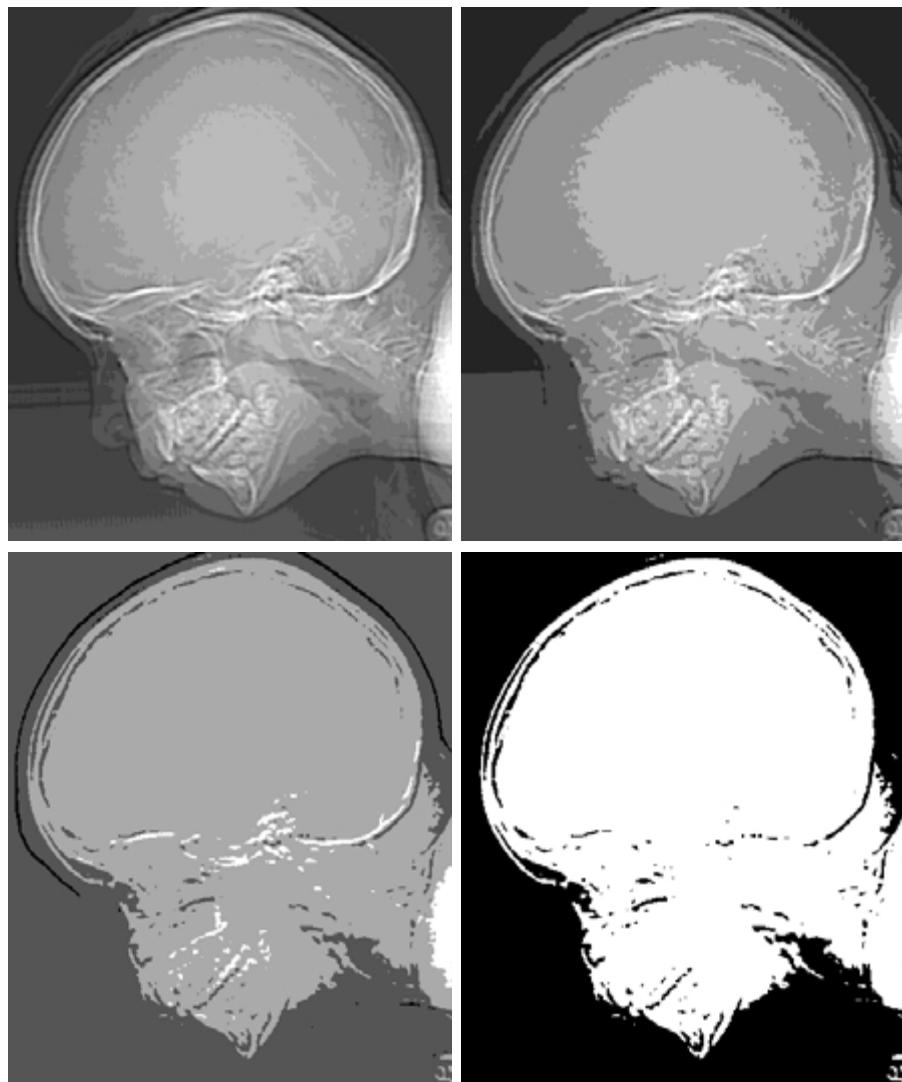
FIGURE 2.21
(a) 452×374 ,
256-level image.
(b)–(d) Image
displayed in 128,
64, and 32 gray
levels, while
keeping the
spatial resolution
constant.

in any desired direction. Projection images are used as guides to set up the parameters for a CAT scanner, including tilt, number of slices, and range.

Figures 2.21(b) through (h) were obtained by reducing the number of bits from $k = 7$ to $k = 1$ while keeping the spatial resolution constant at 452×374 pixels. The 256-, 128-, and 64-level images are visually identical for all practical purposes. The 32-level image shown in Fig. 2.21(d), however, has an almost imperceptible set of very fine ridgelike structures in areas of smooth gray levels (particularly in the skull). This effect, caused by the use of an insufficient number of gray levels in smooth areas of a digital image, is called *false contouring*, so called because the ridges resemble topographic contours in a map. False contouring generally is quite visible in images displayed using 16 or less uniformly spaced gray levels, as the images in Figs. 2.21(e) through (h) show.

e f
g h

FIGURE 2.21
(Continued)
(e)–(g) Image displayed in 16, 8, 4, and 2 gray levels. (Original courtesy of Dr. David R. Pickens, Department of Radiology & Radiological Sciences, Vanderbilt University Medical Center.)



As a very rough rule of thumb, and assuming powers of 2 for convenience, images of size 256×256 pixels and 64 gray levels are about the smallest images that can be expected to be reasonably free of objectionable sampling checkerboards and false contouring. ■

The results in Examples 2.2 and 2.3 illustrate the effects produced on image quality by varying N and k independently. However, these results only partially answer the question of how varying N and k affect images because we have not considered yet any relationships that might exist between these two parameters. An early study by Huang [1965] attempted to quantify experimentally the effects on image quality produced by varying N and k simultaneously. The experiment consisted of a set of subjective tests. Images similar to those shown in Fig. 2.22 were used. The woman's face is representative of an image with relatively little detail; the picture of the cameraman contains an intermediate amount of detail; and the crowd picture contains, by comparison, a large amount of detail.

Sets of these three types of images were generated by varying N and k , and observers were then asked to rank them according to their subjective quality. Results were summarized in the form of so-called *isopreference curves* in the Nk -plane (Fig. 2.23 shows average isopreference curves representative of curves corresponding to the images shown in Fig. 2.22). Each point in the Nk -plane represents an image having values of N and k equal to the coordinates of that point. Points lying on an isopreference curve correspond to images of equal subjective quality. It was found in the course of the experiments that the isopreference curves tended to shift right and upward, but their shapes in each of the three image categories were similar to those shown in Fig. 2.23. This is not unexpected, since a shift up and right in the curves simply means larger values for N and k , which implies better picture quality.

The key point of interest in the context of the present discussion is that isopreference curves tend to become more vertical as the detail in the image increases. This result suggests that for images with a large amount of detail only

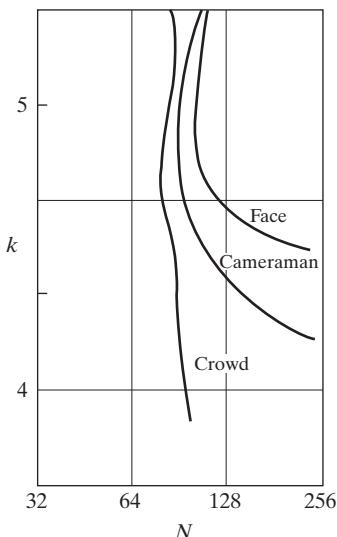


a b c

FIGURE 2.22 (a) Image with a low level of detail. (b) Image with a medium level of detail. (c) Image with a relatively large amount of detail. (Image (b) courtesy of the Massachusetts Institute of Technology.)

FIGURE 2.23

Representative isopreference curves for the three types of images in Fig. 2.22.



a few gray levels may be needed. For example, the isopreference curve in Fig. 2.23 corresponding to the crowd is nearly vertical. This indicates that, for a fixed value of N , the perceived quality for this type of image is nearly independent of the number of gray levels used (for the range of gray levels shown in Fig. 2.23). It is also of interest to note that perceived quality in the other two image categories remained the same in some intervals in which the spatial resolution was increased, but the number of gray levels actually decreased. The most likely reason for this result is that a decrease in k tends to increase the apparent contrast of an image, a visual effect that humans often perceive as improved quality in an image.

2.4.4 Aliasing and Moiré Patterns

As discussed in more detail in Chapter 4, functions whose area under the curve is finite can be represented in terms of sines and cosines of various frequencies. The sine/cosine component with the highest frequency determines the highest “frequency content” of the function. Suppose that this highest frequency is finite and that the function is of unlimited duration (these functions are called *band-limited functions*). Then, the Shannon sampling theorem [Bracewell (1995)] tells us that, if the function is sampled at a rate equal to or greater than twice its highest frequency, it is possible to recover completely the original function from its samples. If the function is *undersampled*, then a phenomenon called *aliasing* corrupts the sampled image. The corruption is in the form of additional frequency components being introduced into the sampled function. These are called *aliased frequencies*. Note that the *sampling rate* in images is the number of samples taken (in both spatial directions) per unit distance.

As it turns out, except for a special case discussed in the following paragraph, it is impossible to satisfy the sampling theorem in practice. We can only work with sampled data that are finite in duration. We can model the process of convert-

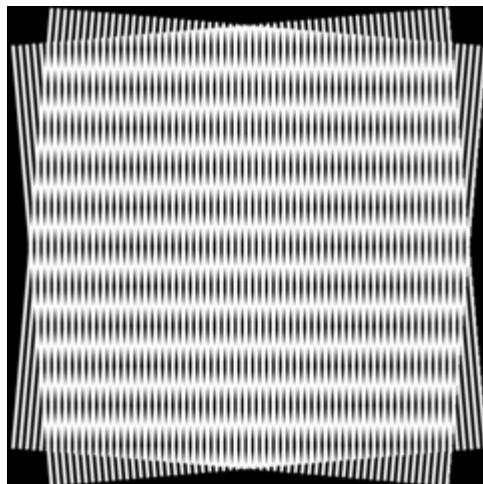


FIGURE 2.24 Illustration of the Moiré pattern effect.

ing a function of unlimited duration into a function of finite duration simply by multiplying the unlimited function by a “gating function” that is valued 1 for some interval and 0 elsewhere. Unfortunately, this function itself has frequency components that extend to infinity. Thus, the very act of limiting the duration of a band-limited function causes it to cease being band limited, which causes it to violate the key condition of the sampling theorem. The principal approach for reducing the aliasing effects on an image is to reduce its high-frequency components by blurring the image (we discuss blurring in detail in Chapter 4) *prior* to sampling. However, aliasing is always present in a sampled image. The effect of aliased frequencies can be seen under the right conditions in the form of so-called *Moiré patterns*[†], as discussed next.

There is one special case of significant importance in which a function of infinite duration can be sampled over a finite interval without violating the sampling theorem. When a function is periodic, it may be sampled at a rate equal to or exceeding twice its highest frequency, and it is possible to recover the function from its samples *provided* that the sampling captures *exactly* an integer number of periods of the function. This special case allows us to illustrate vividly the Moiré effect. Figure 2.24 shows two identical periodic patterns of equally spaced vertical bars, rotated in opposite directions and then superimposed on each other by multiplying the two images. A Moiré pattern, caused by a break-up of the periodicity, is seen in Fig. 2.24 as a 2-D sinusoidal (aliased) waveform (which looks like a corrugated tin roof) running in a vertical direction. A similar pattern can appear when images are digitized (e.g., scanned) from a printed page, which consists of periodic ink dots.

[†]The word *Moiré* appears to have originated with weavers and comes from the word *mohair*, a cloth made from Angora goat hairs.

2.4.5 Zooming and Shrinking Digital Images

We conclude the treatment of sampling and quantization with a brief discussion on how to zoom and shrink a digital image. This topic is related to image sampling and quantization because zooming may be viewed as oversampling, while shrinking may be viewed as undersampling. The key difference between these two operations and sampling and quantizing an original continuous image is that zooming and shrinking are applied to a *digital* image.

Zooming requires two steps: the creation of new pixel locations, and the assignment of gray levels to those new locations. Let us start with a simple example. Suppose that we have an image of size 500×500 pixels and we want to enlarge it 1.5 times to 750×750 pixels. Conceptually, one of the easiest ways to visualize zooming is laying an imaginary 750×750 grid over the original image. Obviously, the spacing in the grid would be less than one pixel because we are fitting it over a smaller image. In order to perform gray-level assignment for any point in the overlay, we look for the closest pixel in the original image and assign its gray level to the new pixel in the grid. When we are done with all points in the overlay grid, we simply expand it to the original specified size to obtain the zoomed image. This method of gray-level assignment is called *nearest neighbor interpolation*. (Pixel neighborhoods are discussed in the next section.)

Pixel replication, the method used to generate Figs. 2.20(b) through (f), is a special case of nearest neighbor interpolation. Pixel replication is applicable when we want to increase the size of an image an integer number of times. For instance, to double the size of an image, we can duplicate each column. This doubles the image size in the horizontal direction. Then, we duplicate each row of the enlarged image to double the size in the vertical direction. The same procedure is used to enlarge the image by any integer number of times (triple, quadruple, and so on). Duplication is just done the required number of times to achieve the desired size. The gray-level assignment of each pixel is predetermined by the fact that new locations are exact duplicates of old locations.

Although nearest neighbor interpolation is fast, it has the undesirable feature that it produces a checkerboard effect that is particularly objectionable at high factors of magnification. Figures 2.20(e) and (f) are good examples of this. A slightly more sophisticated way of accomplishing gray-level assignments is *bilinear interpolation* using the four nearest neighbors of a point. Let (x', y') denote the coordinates of a point in the zoomed image (think of it as a point on the grid described previously), and let $v(x', y')$ denote the gray level assigned to it. For bilinear interpolation, the assigned gray level is given by

$$v(x', y') = ax' + by' + cx'y' + d \quad (2.4-6)$$

where the four coefficients are determined from the four equations in four unknowns that can be written using the four nearest neighbors of point (x', y') .

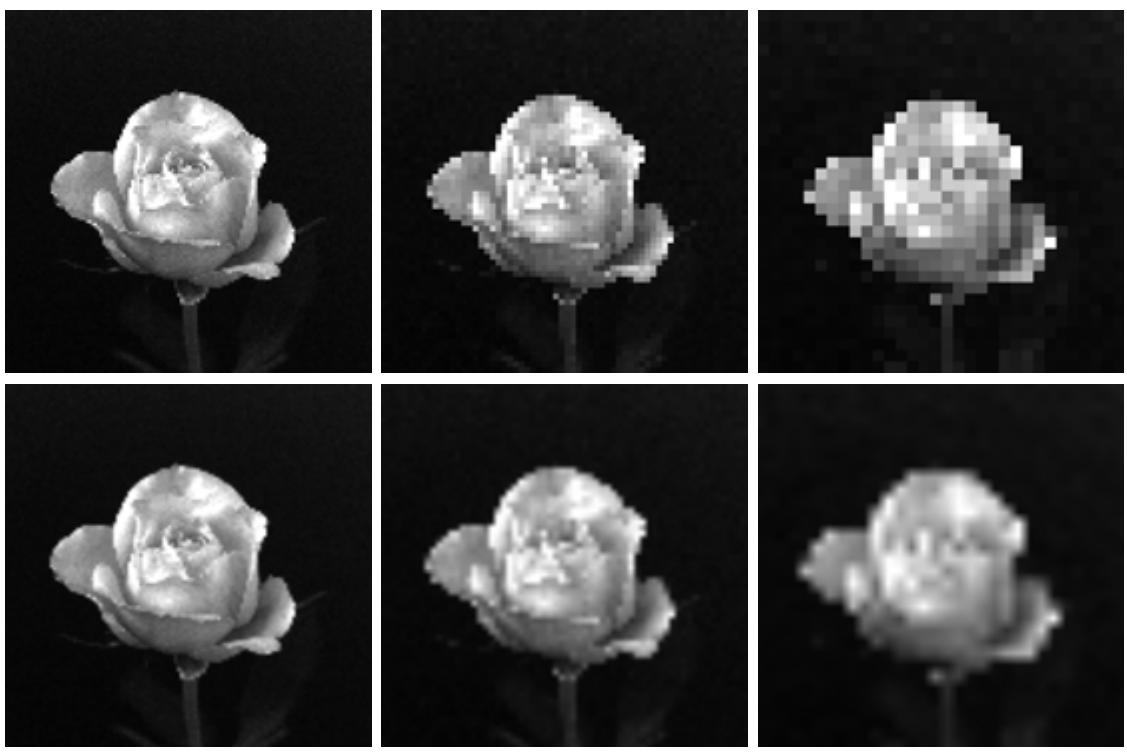
Image shrinking is done in a similar manner as just described for zooming. The equivalent process of pixel replication is row-column deletion. For example, to shrink an image by one-half, we delete every other row and column. We can use the zooming grid analogy to visualize the concept of shrinking by a noninteger factor, except

that we now *expand* the grid to fit over the original image, do gray-level nearest neighbor or bilinear interpolation, and then shrink the grid back to its original specified size. To reduce possible aliasing effects, it is a good idea to blur an image slightly before shrinking it. Blurring of digital images is discussed in Chapters 3 and 4.

It is possible to use more neighbors for interpolation. Using more neighbors implies fitting the points with a more complex surface, which generally gives smoother results. This is an exceptionally important consideration in image generation for 3-D graphics [Watt (1993)] and in medical image processing [Lehmann et al. (1999)], but the extra computational burden seldom is justifiable for general-purpose digital image zooming and shrinking, where bilinear interpolation generally is the method of choice.

■ Figures 2.20(d) through (f) are shown again in the top row of Fig. 2.25. As noted earlier, these images were zoomed from 128×128 , 64×64 , and 32×32 to 1024×1024 pixels using nearest neighbor interpolation. The equivalent results using bilinear interpolation are shown in the second row of Fig. 2.25. The improvements in overall appearance are clear, especially in the 128×128 and

EXAMPLE 2.4:
Image zooming
using bilinear
interpolation.



a	b	c
d	e	f

FIGURE 2.25 Top row: images zoomed from 128×128 , 64×64 , and 32×32 pixels to 1024×1024 pixels, using nearest neighbor gray-level interpolation. Bottom row: same sequence, but using bilinear interpolation.

64×64 cases. The 32×32 to 1024×1024 image is blurry, but keep in mind that this image was zoomed by a factor of 32. In spite of this, the result of bilinear interpolation shown in Fig. 2.25(f) is a reasonably good rendition of the original image shape, something that is lost in Fig. 2.25(c). ■

2.5 Some Basic Relationships Between Pixels

In this section, we consider several important relationships between pixels in a digital image. As mentioned before, an image is denoted by $f(x, y)$. When referring in this section to a particular pixel, we use lowercase letters, such as p and q .

2.5.1 Neighbors of a Pixel

A pixel p at coordinates (x, y) has four *horizontal* and *vertical* neighbors whose coordinates are given by

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)$$

This set of pixels, called the *4-neighbors* of p , is denoted by $N_4(p)$. Each pixel is a unit distance from (x, y) , and some of the neighbors of p lie outside the digital image if (x, y) is on the border of the image.

The four *diagonal* neighbors of p have coordinates

$$(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)$$

and are denoted by $N_D(p)$. These points, together with the 4-neighbors, are called the *8-neighbors* of p , denoted by $N_8(p)$. As before, some of the points in $N_D(p)$ and $N_8(p)$ fall outside the image if (x, y) is on the border of the image.

2.5.2 Adjacency, Connectivity, Regions, and Boundaries

Connectivity between pixels is a fundamental concept that simplifies the definition of numerous digital image concepts, such as regions and boundaries. To establish if two pixels are connected, it must be determined if they are neighbors and if their gray levels satisfy a specified criterion of similarity (say, if their gray levels are equal). For instance, in a binary image with values 0 and 1, two pixels may be 4-neighbors, but they are said to be connected only if they have the same value.

Let V be the set of gray-level values used to define adjacency. In a binary image, $V = \{1\}$ if we are referring to adjacency of pixels with value 1. In a gray-scale image, the idea is the same, but set V typically contains more elements. For example, in the adjacency of pixels with a range of possible gray-level values 0 to 255, set V could be any subset of these 256 values. We consider three types of adjacency:

- (a) *4-adjacency*. Two pixels p and q with values from V are 4-adjacent if q is in the set $N_4(p)$.
- (b) *8-adjacency*. Two pixels p and q with values from V are 8-adjacent if q is in the set $N_8(p)$.

- (c) *m-adjacency* (mixed adjacency). Two pixels p and q with values from V are *m-adjacent* if
- (i) q is in $N_4(p)$, or
 - (ii) q is in $N_D(p)$ and the set $N_4(p) \cap N_4(q)$ has no pixels whose values are from V .

Mixed adjacency is a modification of 8-adjacency. It is introduced to eliminate the ambiguities that often arise when 8-adjacency is used. For example, consider the pixel arrangement shown in Fig. 2.26(a) for $V = \{1\}$. The three pixels at the top of Fig. 2.26(b) show multiple (ambiguous) 8-adjacency, as indicated by the dashed lines. This ambiguity is removed by using *m*-adjacency, as shown in Fig. 2.26(c). Two image subsets S_1 and S_2 are adjacent if some pixel in S_1 is adjacent to some pixel in S_2 . It is understood here and in the following definitions that *adjacent* means 4-, 8-, or *m*-adjacent.

A (*digital*) *path* (or *curve*) from pixel p with coordinates (x, y) to pixel q with coordinates (s, t) is a sequence of distinct pixels with coordinates

$$(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$$

where $(x_0, y_0) = (x, y)$, $(x_n, y_n) = (s, t)$, and pixels (x_i, y_i) and (x_{i-1}, y_{i-1}) are adjacent for $1 \leq i \leq n$. In this case, n is the *length* of the path. If $(x_0, y_0) = (x_n, y_n)$, the path is a *closed path*. We can define 4-, 8-, or *m*-paths depending on the type of adjacency specified. For example, the paths shown in Fig. 2.26(b) between the northeast and southeast points are 8-paths, and the path in Fig. 2.26(c) is an *m*-path. Note the absence of ambiguity in the *m*-path.

Let S represent a subset of pixels in an image. Two pixels p and q are said to be *connected* in S if there exists a path between them consisting entirely of pixels in S . For any pixel p in S , the *set* of pixels that are connected to it in S is called a *connected component* of S . If it only has one connected component, then set S is called a *connected set*.

Let R be a subset of pixels in an image. We call R a *region* of the image if R is a connected set. The *boundary* (also called *border* or *contour*) of a region R is the set of pixels in the region that have one or more neighbors that are not in R . If R happens to be an entire image (which we recall is a rectangular set of pixels), then its boundary is defined as the set of pixels in the first and last rows and columns of the image. This extra definition is required because an image has no neighbors beyond its border. Normally, when we refer to a region, we are

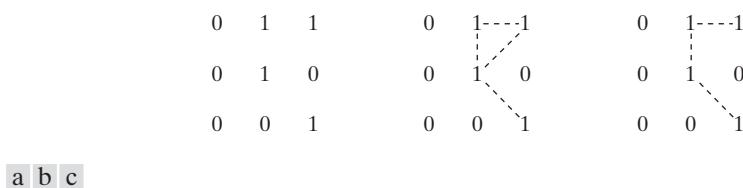


FIGURE 2.26 (a) Arrangement of pixels; (b) pixels that are 8-adjacent (shown dashed) to the center pixel; (c) *m*-adjacency.

referring to a subset of an image, and any pixels in the boundary of the region that happen to coincide with the border of the image are included implicitly as part of the region boundary.

The concept of an *edge* is found frequently in discussions dealing with regions and boundaries. There is a key difference between these concepts, however. The boundary of a finite region forms a closed path (Problem 2.14) and is thus a “global” concept. As discussed in detail in Chapter 10, edges are formed from pixels with derivative values that exceed a preset threshold. Thus, the idea of an edge is a “local” concept that is based on a measure of gray-level discontinuity at a point. It is possible to link edge points into edge segments, and sometimes these segments are linked in such a way that correspond to boundaries, but this is not always the case. The one exception in which edges and boundaries correspond is in binary images. Depending on the type of connectivity and edge operators used (we discuss these in Chapter 10), the edge extracted from a binary region will be the same as the region boundary. This is intuitive. Conceptually, until we arrive at Chapter 10, it is helpful to think of edges as intensity discontinuities and boundaries as closed paths.

2.5.3 Distance Measures

For pixels p , q , and z , with coordinates (x, y) , (s, t) , and (v, w) , respectively, D is a *distance function* or *metric* if

- (a) $D(p, q) \geq 0$ ($D(p, q) = 0$ iff $p = q$),
- (b) $D(p, q) = D(q, p)$, and
- (c) $D(p, z) \leq D(p, q) + D(q, z)$.

The *Euclidean distance* between p and q is defined as

$$D_e(p, q) = [(x - s)^2 + (y - t)^2]^{\frac{1}{2}}. \quad (2.5-1)$$

For this distance measure, the pixels having a distance less than or equal to some value r from (x, y) are the points contained in a disk of radius r centered at (x, y) .

The D_4 distance (also called *city-block distance*) between p and q is defined as

$$D_4(p, q) = |x - s| + |y - t|. \quad (2.5-2)$$

In this case, the pixels having a D_4 distance from (x, y) less than or equal to some value r form a diamond centered at (x, y) . For example, the pixels with D_4 distance ≤ 2 from (x, y) (the center point) form the following contours of constant distance:

$$\begin{matrix} & & & 2 \\ & & 2 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 \\ & 2 & 1 & 2 \\ & & 2 \end{matrix}$$

The pixels with $D_4 = 1$ are the 4-neighbors of (x, y) .

The D_8 distance (also called *chessboard distance*) between p and q is defined as

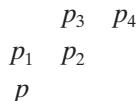
$$D_8(p, q) = \max(|x - s|, |y - t|). \quad (2.5-3)$$

In this case, the pixels with D_8 distance from (x, y) less than or equal to some value r form a square centered at (x, y) . For example, the pixels with D_8 distance ≤ 2 from (x, y) (the center point) form the following contours of constant distance:

2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

The pixels with $D_8 = 1$ are the 8-neighbors of (x, y) .

Note that the D_4 and D_8 distances between p and q are independent of any paths that might exist between the points because these distances involve only the coordinates of the points. If we elect to consider m -adjacency, however, the D_m distance between two points is defined as the shortest m -path between the points. In this case, the distance between two pixels will depend on the values of the pixels along the path, as well as the values of their neighbors. For instance, consider the following arrangement of pixels and assume that p, p_2 , and p_4 have value 1 and that p_1 and p_3 can have a value of 0 or 1:



Suppose that we consider adjacency of pixels valued 1 (i.e., $V = \{1\}$). If p_1 and p_3 are 0, the length of the shortest m -path (the D_m distance) between p and p_4 is 2. If p_1 is 1, then p_2 and p will no longer be m -adjacent (see the definition of m -adjacency) and the length of the shortest m -path becomes 3 (the path goes through the points $pp_1p_2p_4$). Similar comments apply if p_3 is 1 (and p_1 is 0); in this case, the length of the shortest m -path also is 3. Finally, if both p_1 and p_3 are 1 the length of the shortest m -path between p and p_4 is 4. In this case, the path goes through the sequence of points $pp_1p_2p_3p_4$.

2.5.4 Image Operations on a Pixel Basis

Numerous references are made in the following chapters to operations between images, such as dividing one image by another. In Eq. (2.4-2), images were represented in the form of matrices. As we know, matrix division is not defined. However, when we refer to an operation like “dividing one image by another,” we mean specifically that the division is carried out between *corresponding* pixels in the two images. Thus, for example, if f and g are images, the first element of the image formed by “dividing” f by g is simply the first pixel in f divided by the first pixel in g ; of course, the assumption is that none of the pixels in g have value 0. Other arithmetic and logic operations are similarly defined between corresponding pixels in the images involved.

2.6 Linear and Nonlinear Operations

Let H be an operator whose input and output are images. H is said to be a *linear* operator if, for any two images f and g and any two scalars a and b ,

$$H(af + bg) = aH(f) + bH(g). \quad (2.6-1)$$

In other words, the result of applying a linear operator to the sum of two images (that have been multiplied by the constants shown) is identical to applying the operator to the images individually, multiplying the results by the appropriate constants, and then adding those results. For example, an operator whose function is to compute the sum of K images is a linear operator. An operator that computes the absolute value of the difference of two images is not. An operator that fails the test of Eq. (2.6-1) is by definition *nonlinear*.

Linear operations are exceptionally important in image processing because they are based on a significant body of well-understood theoretical and practical results. Although nonlinear operations sometimes offer better performance, they are not always predictable, and for the most part are not well understood theoretically.

Summary

The material in this chapter is primarily background information for subsequent discussions. Our treatment of the human visual system, although brief, provides a basic idea of the capabilities of the eye in perceiving pictorial information. The discussion of light and the electromagnetic spectrum is fundamental in understanding the origin of the many images we use in this book. Similarly, the image model developed in Section 2.3.4 is used in the Chapter 4 as the basis for an image enhancement technique called *homomorphic filtering*, and again in Chapter 10 to explain the effect of illumination on the shape of image histograms.

The sampling ideas introduced in Section 2.4 are the foundation for many of the digitizing phenomena likely to be encountered in practice. These ideas can be expanded further once a basic understanding of frequency content is mastered. A detailed discussion of the frequency domain is given in Chapter 4. The concepts of sampling and aliasing effects also are of importance in the context of image acquisition.

The concepts introduced in Section 2.5 are the basic building blocks for processing techniques based on pixel neighborhoods. As shown in the following chapter and in Chapter 5, neighborhood processing methods are at the core of many image enhancement and restoration procedures. When applicable, neighborhood processing is favored in commercial applications of image processing due to their operational speed and simplicity of implementation in hardware and/or firmware. Finally, the concept of a linear operator and the theoretical and conceptual power associated with it will be used extensively in the following three chapters.

References and Further Reading

Additional reading for the material in Section 2.1 regarding the structure of the human eye may be found in Atchison and Smith [2000], and Oyster [1999]. For additional reading on visual perception, see Regan [2000] and Gordon [1997]. The book by Hubel [1988] and the now classic book by Cornsweet [1970] also are of interest. Born and Wolf [1999]

is a basic reference that discusses light in terms of electromagnetic theory. Electromagnetic energy propagation is covered in some detail by Felsen and Marcuvitz [1994].

The area of image sensing is quite broad and very fast moving. An excellent source of information on optical and other imaging sensors is the International Society for Optical Engineering (SPIE). The following are representative publications by the SPIE in this area: Blouke et al. [2001], Hoover and Doty [1996], and Freeman [1987].

The image model presented in Section 2.3.4 is from Oppenheim, Schafer, and Stockham [1968]. A reference for the illumination and reflectance values used in that section is the *IES Lighting Handbook* [2000]. For additional reading on image sampling and some of its effects, such as aliasing, see Bracewell [1995]. The early experiments mentioned in Section 2.4.3 on perceived image quality as a function of sampling and quantization were reported by Huang [1965]. The issue of reducing the number of samples and gray levels in an image while minimizing the ensuing degradation is still of current interest, as exemplified by Papamarkos and Atsalakis [2000]. For further reading on image shrinking and zooming, see Sid-Ahmed [1995], Unser et al. [1995], Umbaugh [1998], and Lehmann et al. [1999]. For further reading on the topics covered in Section 2.5, see Rosenfeld and Kak [1982], Marchand-Maillet and Sharaiha [2000], and Ritter and Wilson [2001]. Additional reading on linear systems in the context of image processing may be found in Castleman [1996].

Problems

- ★ 2.1** Using the background information provided in Section 2.1, and thinking purely in geometric terms, estimate the diameter of the smallest printed dot that the eye can discern if the page on which the dot is printed is 0.2 m away from the eyes. Assume for simplicity that the visual system ceases to detect the dot when the image of the dot on the fovea becomes smaller than the diameter of one receptor (cone) in that area of the retina. Assume further that the fovea can be modeled as a square array of dimensions 1.5 mm \times 1.5 mm, and that the cones and spaces between the cones are distributed uniformly throughout this array.
- 2.2** When you enter a dark theater on a bright day, it takes an appreciable interval of time before you can see well enough to find an empty seat. Which of the visual processes explained in Section 2.1 is at play in this situation?
- ★ 2.3** Although it is not shown in Fig. 2.10, alternating current certainly is part of the electromagnetic spectrum. Commercial alternating current in the United States has a frequency of 60 Hz. What is the wavelength in kilometers of this component of the spectrum?
- 2.4** You are hired to design the front end of an imaging system for studying the boundary shapes of cells, bacteria, viruses, and protein. The front end consists, in this case, of the illumination source(s) and corresponding imaging camera(s). The diameters of circles required to enclose individual specimens in each of these categories are 50, 1, 0.1, and 0.01 μm , respectively.
- (a) Can you solve the imaging aspects of this problem with a single sensor and camera? If your answer is yes, specify the illumination wavelength band and the type of camera needed. Identify the camera as being a color camera, far-infrared camera, or whatever appropriate name corresponds to the illumination source.
- (b) If your answer in (a) is no, what type of illumination sources and corresponding imaging sensors would you recommend? Specify the light sources



See inside front cover

Detailed solutions to the problems marked with a star can be found in the book web site. The site also contains suggested projects based on the material in this chapter.

and cameras as requested in part (a). Use the *minimum* number of illumination sources and cameras needed to solve the problem.

- 2.5** A CCD camera chip of dimensions 7×7 mm, and having 1024×1024 elements, is focused on a square, flat area, located 0.5 m away. How many line pairs per mm will this camera be able to resolve? The camera is equipped with a 35-mm lens. (*Hint:* Model the imaging process as in Fig. 2.3, with the focal length of the camera lens substituting for the focal length of the eye.)
- ★2.6** An automobile manufacturer is automating the placement of certain components on the bumpers of a limited-edition line of sports cars. The components are color coordinated, so the robots need to know the color of each car in order to select the appropriate bumper component. Models come in only four colors: blue, green, red, and white. You are hired to propose a solution based on imaging. How would you solve the problem of automatically determining the color of each car, keeping in mind that *cost* is the most important consideration in your choice of components?
- 2.7** Suppose that a flat area with center at (x_0, y_0) is illuminated by a light source with intensity distribution

$$i(x, y) = Ke^{-[(x-x_0)^2 + (y-y_0)^2]}.$$

Assume for simplicity that the reflectance of the area is constant and equal to 1.0, and let $K = 255$. If the resulting image is digitized with k bits of intensity resolution, and the eye can detect an abrupt change of eight shades of intensity between adjacent pixels, what value of k will cause visible false contouring?

- 2.8** Sketch the image in Problem 2.7 for $k = 2$.
- ★2.9** A common measure of transmission for digital data is the *baud rate*, defined as the number of bits transmitted per second. Generally, transmission is accomplished in packets consisting of a start bit, a byte (8 bits) of information, and a stop bit. Using these facts, answer the following:
- (a) How many minutes would it take to transmit a 1024×1024 image with 256 gray levels using a 56K baud modem?
- (b) What would the time be at 750K baud, a representative speed of a phone DSL (digital subscriber line) connection?
- 2.10** High-definition television (HDTV) generates images with a resolution of 1125 horizontal TV lines interlaced (where every other line is painted on the tube face in each of two fields, each field being 1/60th of a second in duration). The width-to-height aspect ratio of the images is 16:9. The fact that the horizontal lines are distinct fixes the vertical resolution of the images. A company has designed an image capture system that generates digital images from HDTV images. The resolution of each TV (horizontal) line in their system is in proportion to vertical resolution, with the proportion being the width-to-height ratio of the images. Each pixel in the color image has 24 bits of intensity resolution, 8 pixels each for a red, a green, and a blue image. These three “primary” images form a color image. How many bits would it take to store a 2-hour HDTV program?
- ★2.11** Consider the two image subsets, S_1 and S_2 , shown in the following figure. For $V = \{1\}$, determine whether these two subsets are (a) 4-adjacent, (b) 8-adjacent, or (c) m -adjacent.

	S_1					S_2					
0	0	0	0	0	0	0	0	1	1	0	
1	0	0	1	0	0	0	1	0	0	1	
1	0	0	1	0	1	1	0	0	0	0	
0	0	1	1	1	0	0	0	0	0	0	
0	0	1	1	1	0	0	0	1	1	1	

- ★ 2.12 Develop an algorithm for converting a one-pixel-thick 8-path to a 4-path.
- 2.13 Develop an algorithm for converting a one-pixel-thick m -path to a 4-path.
- 2.14 Show that the boundary of the region, as defined in Section 2.5.2, is a closed path.
- ★ 2.15 Consider the image segment shown.
- (a) Let $V = \{0, 1\}$ and compute the lengths of the shortest 4-, 8-, and m -path between p and q . If a particular path does not exist between these two points, explain why.
- (b) Repeat for $V = \{1, 2\}$.

3	1	2	1	(q)
2	2	0	2	
1	2	1	1	
(p)	1	0	1	2

- ★ 2.16 (a) Give the condition(s) under which the D_4 distance between two points p and q is equal to the shortest 4-path between these points.
- (b) Is this path unique?
- 2.17 Repeat Problem 2.16 for the D_8 distance.
- ★ 2.18 In the following chapter, we will deal with operators whose function is to compute the sum of pixel values in a small subimage area, S . Show that these are linear operators.
- 2.19 The median, ζ , of a set of numbers is such that half the values in the set are below ζ and the other half are above it. For example, the median of the set of values $\{2, 3, 8, 20, 21, 25, 31\}$ is 20. Show that an operator that computes the median of a subimage area, S , is nonlinear.
- 2.20 A plant produces a line of translucent miniature polymer squares. Stringent quality requirements dictate 100% visual inspection, and the plant manager finds the use of human inspectors increasingly expensive. Inspection is semiautomated. At each inspection station, a robotic mechanism places each polymer square over a light located under an optical system that produces a magnified image of the square. The image completely fills a viewing screen measuring 80×80 mm. Defects appear as dark circular blobs, and the inspector's job is to look at the screen and reject any sample that has one or more such dark blobs with a diameter of 0.8 mm or larger, as measured on the scale of the screen. The manager believes that, if she can find a way to automate the process completely, she will increase profits by 50%. She also believes that success in this project will aid her climb up the corporate ladder. After much investigation, the manager decides that the way to solve the problem is to view each inspection screen with a CCD TV camera and feed the output of the

camera into an image processing system capable of detecting the blobs, measuring their diameter, and activating the accept/reject buttons previously operated by an inspector. She is able to find a system that can do the job, as long as the smallest defect occupies an area of at least 2×2 pixels in the digital image. The manager hires you to help her specify the camera and lens system, but requires that you use off-the-shelf components. For the lenses, assume that this constraint means any integer multiple of 25 mm or 35 mm, up to 200 mm. For the cameras, it means resolutions of 512×512 , 1024×1024 , or 2048×2048 pixels. The *individual* imaging elements in these cameras are squares measuring $8 \times 8 \mu\text{m}$, and the spaces between imaging elements are $2 \mu\text{m}$. For this application, the cameras cost much more than the lenses, so the problem should be solved with the lowest-resolution camera possible, based on the choice of lenses. As a consultant, you are to provide a written recommendation, showing in reasonable detail the analysis that led to your conclusion. Use the same imaging geometry suggested in Problem 2.5.

3

Image Enhancement in the Spatial Domain



It makes all the difference whether one sees darkness through the light or brightness through the shadows.

David Lindsay

Preview

The principal objective of enhancement is to process an image so that the result is more suitable than the original image for a specific application. The word *specific* is important, because it establishes at the outset that the techniques discussed in this chapter are very much problem oriented. Thus, for example, a method that is quite useful for enhancing X-ray images may not necessarily be the best approach for enhancing pictures of Mars transmitted by a space probe. Regardless of the method used, however, image enhancement is one of the most interesting and visually appealing areas of image processing.

Image enhancement approaches fall into two broad categories: spatial domain methods and frequency domain methods. The term *spatial domain* refers to the image plane itself, and approaches in this category are based on direct manipulation of pixels in an image. Frequency domain processing techniques are based on modifying the Fourier transform of an image. Spatial methods are covered in this chapter, and frequency domain enhancement is discussed in Chapter 4. Enhancement techniques based on various combinations of methods from these two categories are not unusual. We note also that many of the fundamental techniques introduced in this chapter in the context of enhancement are used in subsequent chapters for a variety of other image processing applications.

There is no general theory of image enhancement. When an image is processed for visual interpretation, the viewer is the ultimate judge of how well

a particular method works. Visual evaluation of image quality is a highly subjective process, thus making the definition of a “good image” an elusive standard by which to compare algorithm performance. When the problem is one of processing images for machine perception, the evaluation task is somewhat easier. For example, in dealing with a character recognition application, and leaving aside other issues such as computational requirements, the best image processing method would be the one yielding the best machine recognition results. However, even in situations when a clear-cut criterion of performance can be imposed on the problem, a certain amount of trial and error usually is required before a particular image enhancement approach is selected.

3.1 Background

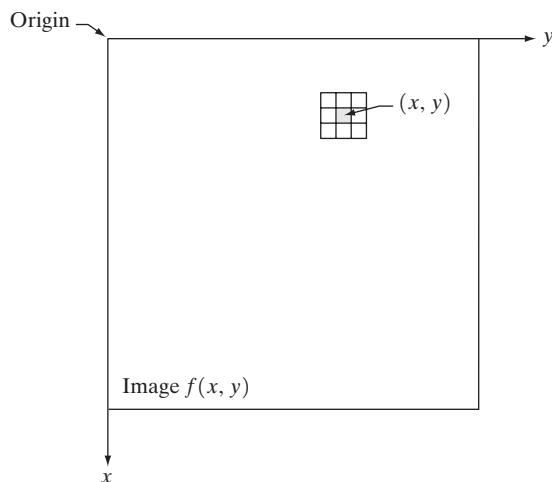
As indicated previously, the term *spatial domain* refers to the aggregate of pixels composing an image. Spatial domain methods are procedures that operate directly on these pixels. Spatial domain processes will be denoted by the expression

$$g(x, y) = T[f(x, y)] \quad (3.1-1)$$

where $f(x, y)$ is the input image, $g(x, y)$ is the processed image, and T is an operator on f , defined over some neighborhood of (x, y) . In addition, T can operate on a set of input images, such as performing the pixel-by-pixel sum of K images for noise reduction, as discussed in Section 3.4.2.

The principal approach in defining a neighborhood about a point (x, y) is to use a square or rectangular subimage area centered at (x, y) , as Fig. 3.1 shows. The center of the subimage is moved from pixel to pixel starting, say, at the top left corner. The operator T is applied at each location (x, y) to yield the output, g , at that location. The process utilizes only the pixels in the area of the image spanned by the neighborhood. Although other neighborhood shapes, such as ap-

FIGURE 3.1 A 3×3 neighborhood about a point (x, y) in an image.



proximations to a circle, sometimes are used, square and rectangular arrays are by far the most predominant because of their ease of implementation.

The simplest form of T is when the neighborhood is of size 1×1 (that is, a single pixel). In this case, g depends only on the value of f at (x, y) , and T becomes a *gray-level* (also called an *intensity* or *mapping*) *transformation function* of the form

$$s = T(r) \quad (3.1-2)$$

where, for simplicity in notation, r and s are variables denoting, respectively, the gray level of $f(x, y)$ and $g(x, y)$ at any point (x, y) . For example, if $T(r)$ has the form shown in Fig. 3.2(a), the effect of this transformation would be to produce an image of higher contrast than the original by darkening the levels below m and brightening the levels above m in the original image. In this technique, known as *contrast stretching*, the values of r below m are compressed by the transformation function into a narrow range of s , toward black. The opposite effect takes place for values of r above m . In the limiting case shown in Fig. 3.2(b), $T(r)$ produces a two-level (binary) image. A mapping of this form is called a *thresholding* function. Some fairly simple, yet powerful, processing approaches can be formulated with gray-level transformations. Because enhancement at any point in an image depends only on the gray level at that point, techniques in this category often are referred to as *point processing*.

Larger neighborhoods allow considerably more flexibility. The general approach is to use a function of the values of f in a predefined neighborhood of (x, y) to determine the value of g at (x, y) . One of the principal approaches in this formulation is based on the use of so-called *masks* (also referred to as *filters*, *kernels*, *templates*, or *windows*). Basically, a mask is a small (say, 3×3) 2-D array, such as the one shown in Fig. 3.1, in which the values of the mask coefficients determine the nature of the process, such as image sharpening. Enhancement techniques based on this type of approach often are referred to as *mask processing* or *filtering*. These concepts are discussed in Section 3.5.

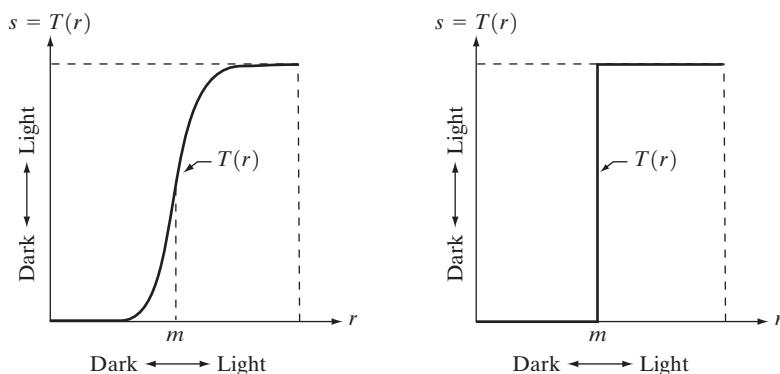


FIGURE 3.2 Gray-level transformation functions for contrast enhancement.

3.2 Some Basic Gray Level Transformations

We begin the study of image enhancement techniques by discussing gray-level transformation functions. These are among the simplest of all image enhancement techniques. The values of pixels, before and after processing, will be denoted by r and s , respectively. As indicated in the previous section, these values are related by an expression of the form $s = T(r)$, where T is a transformation that maps a pixel value r into a pixel value s . Since we are dealing with digital quantities, values of the transformation function typically are stored in a one-dimensional array and the mappings from r to s are implemented via table lookups. For an 8-bit environment, a lookup table containing the values of T will have 256 entries.

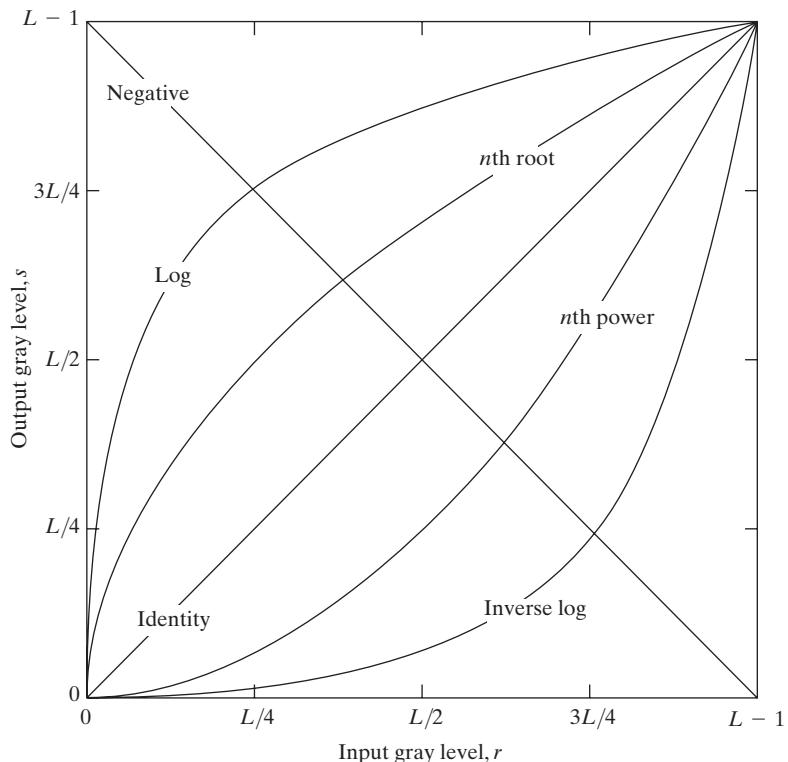
As an introduction to gray-level transformations, consider Fig. 3.3, which shows three basic types of functions used frequently for image enhancement: linear (negative and identity transformations), logarithmic (log and inverse-log transformations), and power-law (n th power and n th root transformations). The identity function is the trivial case in which output intensities are identical to input intensities. It is included in the graph only for completeness.

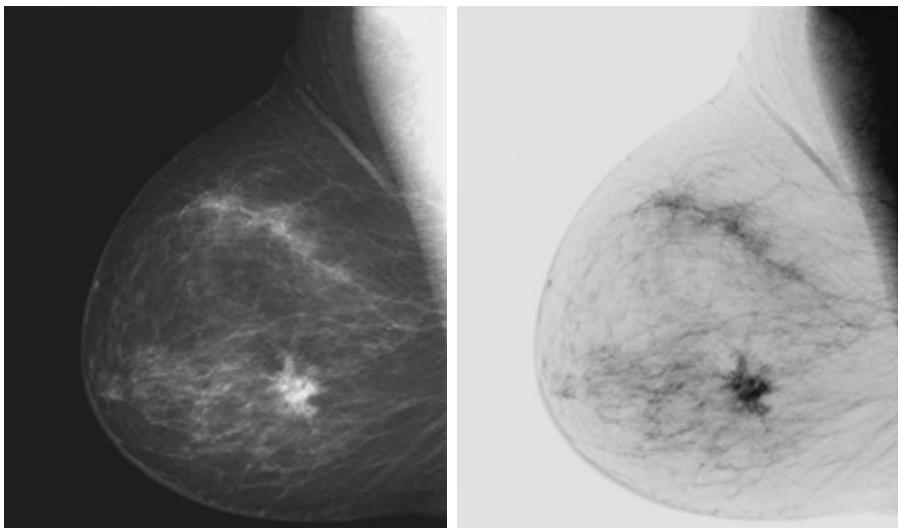
3.2.1 Image Negatives

The negative of an image with gray levels in the range $[0, L - 1]$ is obtained by using the **negative transformation** shown in Fig. 3.3, which is given by the expression

$$s = L - 1 - r. \quad (3.2-1)$$

FIGURE 3.3 Some basic gray-level transformation functions used for image enhancement.





a b

FIGURE 3.4
 (a) Original digital mammogram.
 (b) Negative image obtained using the negative transformation in Eq. (3.2-1).
 (Courtesy of G.E. Medical Systems.)

Reversing the intensity levels of an image in this manner produces the equivalent of a photographic negative. This type of processing is particularly suited for enhancing white or gray detail embedded in dark regions of an image, especially when the black areas are dominant in size. An example is shown in Fig. 3.4. The original image is a digital mammogram showing a small lesion. In spite of the fact that the visual content is the same in both images, note how much easier it is to analyze the breast tissue in the negative image in this particular case.

3.2.2 Log Transformations

The general form of the log transformation shown in Fig. 3.3 is

$$s = c \log(1 + r) \quad (3.2-2)$$

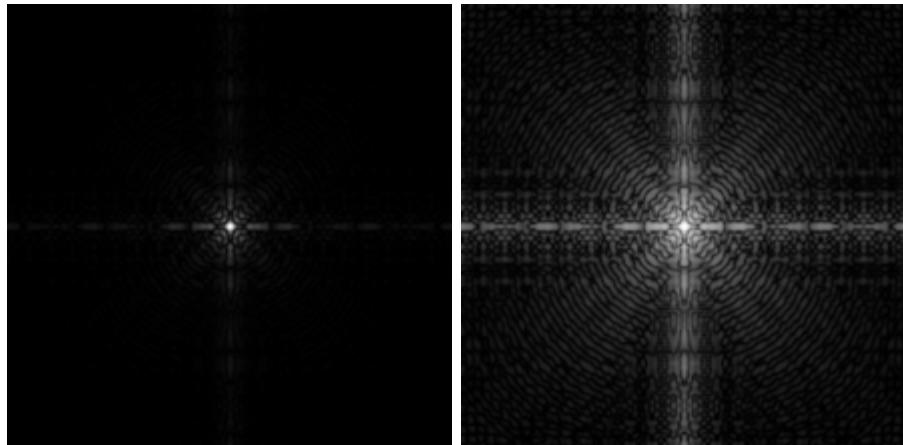
where c is a constant, and it is assumed that $r \geq 0$. The shape of the log curve in Fig. 3.3 shows that this transformation maps a narrow range of low gray-level values in the input image into a wider range of output levels. The opposite is true of higher values of input levels. We would use a transformation of this type to expand the values of dark pixels in an image while compressing the higher-level values. The opposite is true of the inverse log transformation.

Any curve having the general shape of the log functions shown in Fig. 3.3 would accomplish this spreading/compressing of gray levels in an image. In fact, the power-law transformations discussed in the next section are much more versatile for this purpose than the log transformation. However, the log function has the important characteristic that it compresses the dynamic range of images with large variations in pixel values. A classic illustration of an application in which pixel values have a large dynamic range is the Fourier spectrum, which will be discussed in Chapter 4. At the moment, we are concerned only with the image characteristics of spectra. It is not unusual to encounter spectrum values

a b

FIGURE 3.5

- (a) Fourier spectrum.
 (b) Result of applying the log transformation given in Eq. (3.2-2) with $c = 1$.



that range from 0 to 10^6 or higher. While processing numbers such as these presents no problems for a computer, image display systems generally will not be able to reproduce faithfully such a wide range of intensity values. The net effect is that a significant degree of detail will be lost in the display of a typical Fourier spectrum.

As an illustration of log transformations, Fig. 3.5(a) shows a Fourier spectrum with values in the range 0 to 1.5×10^6 . When these values are scaled linearly for display in an 8-bit system, the brightest pixels will dominate the display, at the expense of lower (and just as important) values of the spectrum. The effect of this dominance is illustrated vividly by the relatively small area of the image in Fig. 3.5(a) that is not perceived as black. If, instead of displaying the values in this manner, we first apply Eq. (3.2-2) (with $c = 1$ in this case) to the spectrum values, then the range of values of the result become 0 to 6.2, a more manageable number. Figure 3.5(b) shows the result of scaling this new range linearly and displaying the spectrum in the same 8-bit display. The wealth of detail visible in this image as compared to a straight display of the spectrum is evident from these pictures. Most of the Fourier spectra seen in image processing publications have been scaled in just this manner.

3.2.3 Power-Law Transformations

Power-law transformations have the basic form

$$s = cr^\gamma \quad (3.2-3)$$

where c and γ are positive constants. Sometimes Eq. (3.2-3) is written as $s = c(r + \varepsilon)^\gamma$ to account for an offset (that is, a measurable output when the input is zero). However, offsets typically are an issue of display calibration and as a result they are normally ignored in Eq. (3.2-3). Plots of s versus r for various values of γ are shown in Fig. 3.6. As in the case of the log transformation, power-law curves with fractional values of γ map a narrow range of dark input values into a wider range of output values, with the opposite being true for high-

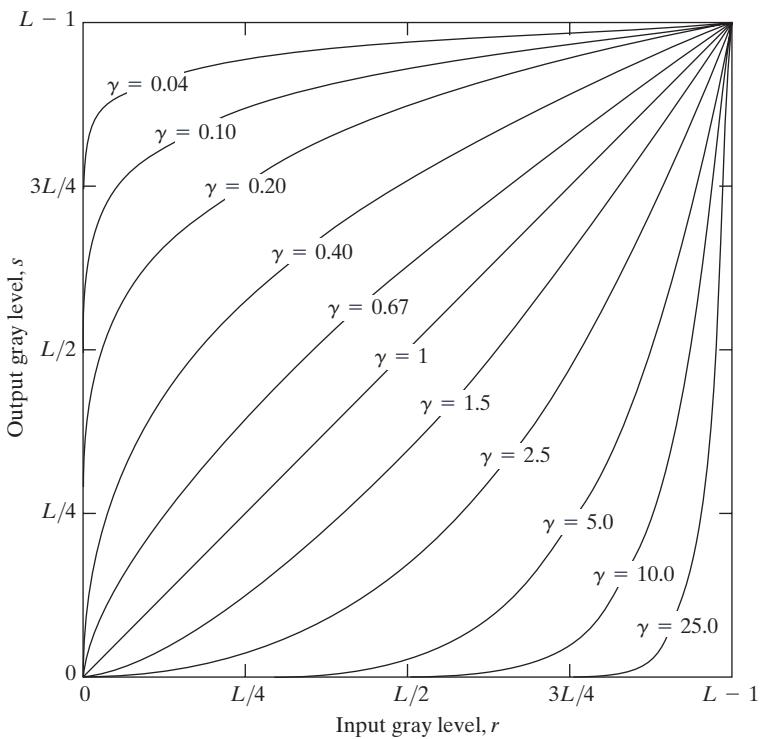


FIGURE 3.6 Plots of the equation $s = cr^\gamma$ for various values of γ ($c = 1$ in all cases).

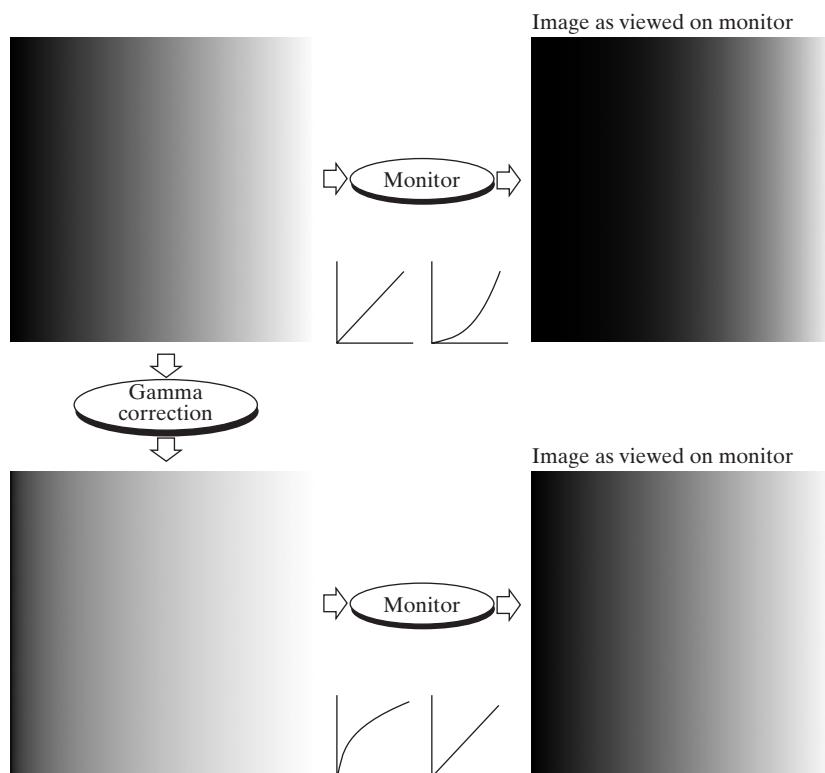
er values of input levels. Unlike the log function, however, we notice here a family of possible transformation curves obtained simply by varying γ . As expected, we see in Fig. 3.6 that curves generated with values of $\gamma > 1$ have exactly the opposite effect as those generated with values of $\gamma < 1$. Finally, we note that Eq. (3.2-3) reduces to the identity transformation when $c = \gamma = 1$.

A variety of devices used for image capture, printing, and display respond according to a power law. By convention, the exponent in the power-law equation is referred to as *gamma* [hence our use of this symbol in Eq. (3.2-3)]. The process used to correct this power-law response phenomena is called *gamma correction*. For example, cathode ray tube (CRT) devices have an intensity-to-voltage response that is a power function, with exponents varying from approximately 1.8 to 2.5. With reference to the curve for $\gamma = 2.5$ in Fig. 3.6, we see that such display systems would tend to produce images that are darker than intended. This effect is illustrated in Fig. 3.7. Figure 3.7(a) shows a simple gray-scale linear wedge input into a CRT monitor. As expected, the output of the monitor appears darker than the input, as shown in Fig. 3.7(b). Gamma correction in this case is straightforward. All we need to do is preprocess the input image before inputting it into the monitor by performing the transformation $s = r^{1/2.5} = r^{0.4}$. The result is shown in Fig. 3.7(c). When input into the same monitor, this gamma-corrected input produces an output that is close in appearance to the original image, as shown in Fig. 3.7(d). A similar analysis would

a
b
c
d

FIGURE 3.7

- (a) Linear-wedge gray-scale image.
- (b) Response of monitor to linear wedge.
- (c) Gamma-corrected wedge.
- (d) Output of monitor.



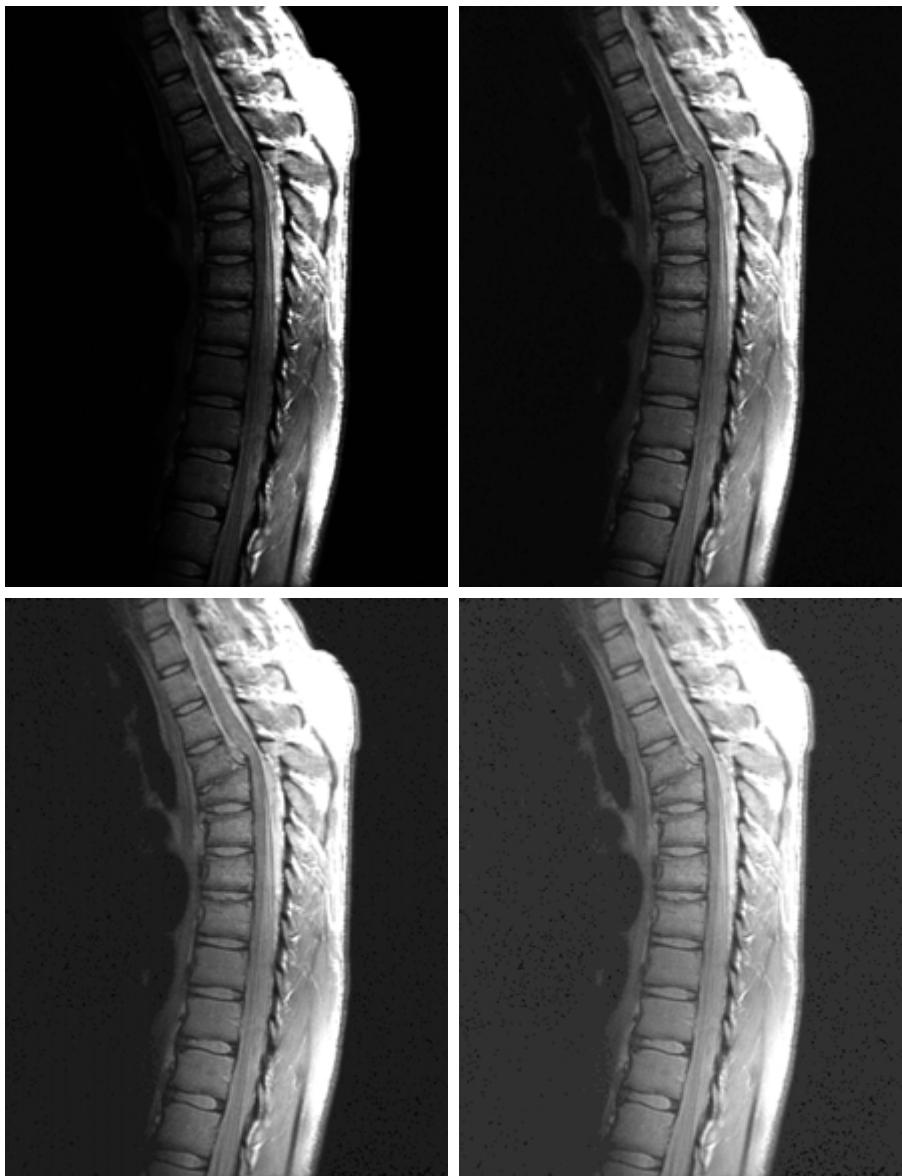
apply to other imaging devices such as scanners and printers. The only difference would be the device-dependent value of gamma (Poynton [1996]).

Gamma correction is important if displaying an image accurately on a computer screen is of concern. Images that are not corrected properly can look either bleached out, or, what is more likely, too dark. Trying to reproduce colors accurately also requires some knowledge of gamma correction because varying the value of gamma correction changes not only the brightness, but also the ratios of red to green to blue. Gamma correction has become increasingly important in the past few years, as use of digital images for commercial purposes over the Internet has increased. It is not unusual that images created for a popular Web site will be viewed by millions of people, the majority of whom will have different monitors and/or monitor settings. Some computer systems even have partial gamma correction built in. Also, current image standards do not contain the value of gamma with which an image was created, thus complicating the issue further. Given these constraints, a reasonable approach when storing images in a Web site is to preprocess the images with a gamma that represents an “average” of the types of monitors and computer systems that one expects in the open market at any given point in time.

EXAMPLE 3.1:

Contrast enhancement using power-law transformations.

- In addition to gamma correction, power-law transformations are useful for general-purpose contrast manipulation. Figure 3.8(a) shows a magnetic resonance (MR) image of an upper thoracic human spine with a fracture dislocation



a b
c d

FIGURE 3.8
 (a) Magnetic resonance (MR) image of a fractured human spine.
 (b)–(d) Results of applying the transformation in Eq. (3.2-3) with $c = 1$ and $\gamma = 0.6, 0.4$, and 0.3 , respectively. (Original image for this example courtesy of Dr. David R. Pickens, Department of Radiology and Radiological Sciences, Vanderbilt University Medical Center.)

and spinal cord impingement. The fracture is visible near the vertical center of the spine, approximately one-fourth of the way down from the top of the picture. Since the given image is predominantly dark, an expansion of gray levels are desirable. This can be accomplished with a power-law transformation with a fractional exponent. The other images shown in the Figure were obtained by processing Fig. 3.8(a) with the power-law transformation function of Eq. (3.2-3). The values of gamma corresponding to images (b) through (d) are 0.6, 0.4, and 0.3, respectively (the value of c was 1 in all cases). We note that, as gamma decreased from 0.6 to 0.4, more detail became visible. A further decrease of gamma

to 0.3 enhanced a little more detail in the background, but began to reduce contrast to the point where the image started to have a very slight “washed-out” look, especially in the background. By comparing all results, we see that the best enhancement in terms of contrast and discernable detail was obtained with $\gamma = 0.4$. A value of $\gamma = 0.3$ is an approximate limit below which contrast in this particular image would be reduced to an unacceptable level. ■

EXAMPLE 3.2:

Another illustration of power-law transformations.

■ Figure 3.9(a) shows the opposite problem of Fig. 3.8(a). The image to be enhanced now has a washed-out appearance, indicating that a compression of gray levels is desirable. This can be accomplished with Eq. (3.2-3) using values of γ greater than 1. The results of processing Fig. 3.9(a) with $\gamma = 3.0, 4.0$, and 5.0 are shown in Figs. 3.9(b) through (d). Suitable results were obtained with gamma values of 3.0 and 4.0, the latter having a slightly more appealing appearance because it has higher contrast. The result obtained with $\gamma = 5.0$ has areas that are too dark, in which some detail is lost. The dark region to the left of the main road in the upper left quadrant is an example of such an area. ■

a b
c d

FIGURE 3.9

(a) Aerial image.
(b)–(d) Results of applying the transformation in Eq. (3.2-3) with $c = 1$ and $\gamma = 3.0, 4.0$, and 5.0 , respectively.
(Original image for this example courtesy of NASA.)



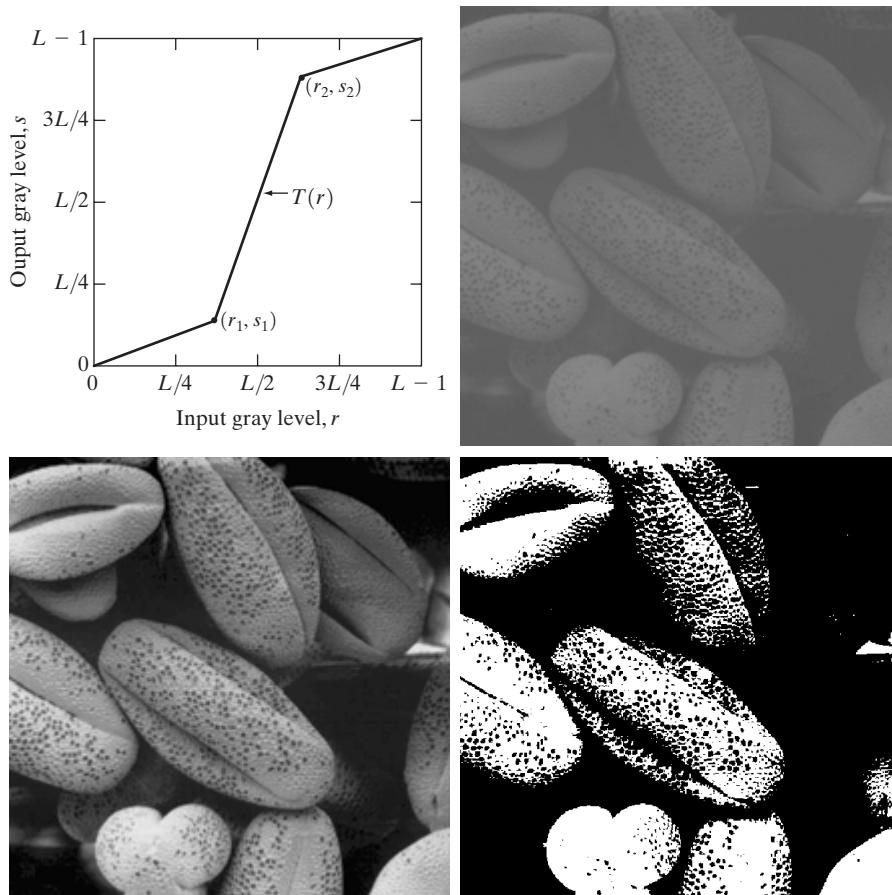
3.2.4 Piecewise-Linear Transformation Functions

A complementary approach to the methods discussed in the previous three sections is to use piecewise linear functions. The principal advantage of piecewise linear functions over the types of functions we have discussed thus far is that the form of piecewise functions can be arbitrarily complex. In fact, as we will see shortly, a practical implementation of some important transformations can be formulated only as piecewise functions. The principal disadvantage of piecewise functions is that their specification requires considerably more user input.

Contrast stretching

One of the simplest piecewise linear functions is a contrast-stretching transformation. Low-contrast images can result from poor illumination, lack of dynamic range in the imaging sensor, or even wrong setting of a lens aperture during image acquisition. The idea behind contrast stretching is to increase the dynamic range of the gray levels in the image being processed.

Figure 3.10(a) shows a typical transformation used for contrast stretching. The locations of points (r_1, s_1) and (r_2, s_2) control the shape of the transformation



a b
c d

FIGURE 3.10
Contrast stretching.
(a) Form of transformation function. (b) A low-contrast image. (c) Result of contrast stretching. (d) Result of thresholding. (Original image courtesy of Dr. Roger Heady, Research School of Biological Sciences, Australian National University, Canberra, Australia.)

function. If $r_1 = s_1$ and $r_2 = s_2$, the transformation is a linear function that produces no changes in gray levels. If $r_1 = r_2$, $s_1 = 0$ and $s_2 = L - 1$, the transformation becomes a *thresholding function* that creates a binary image, as illustrated in Fig. 3.2(b). Intermediate values of (r_1, s_1) and (r_2, s_2) produce various degrees of spread in the gray levels of the output image, thus affecting its contrast. In general, $r_1 \leq r_2$ and $s_1 \leq s_2$ is assumed so that the function is single valued and monotonically increasing. This condition preserves the order of gray levels, thus preventing the creation of intensity artifacts in the processed image.

Figure 3.10(b) shows an 8-bit image with low contrast. Fig. 3.10(c) shows the result of contrast stretching, obtained by setting $(r_1, s_1) = (r_{\min}, 0)$ and $(r_2, s_2) = (r_{\max}, L - 1)$ where r_{\min} and r_{\max} denote the minimum and maximum gray levels in the image, respectively. Thus, the transformation function stretched the levels linearly from their original range to the full range $[0, L - 1]$. Finally, Fig. 3.10(d) shows the result of using the thresholding function defined previously, with $r_1 = r_2 = m$, the mean gray level in the image. The original image on which these results are based is a scanning electron microscope image of pollen, magnified approximately 700 times.

Gray-level slicing

Highlighting a specific range of gray levels in an image often is desired. Applications include enhancing features such as masses of water in satellite imagery and enhancing flaws in X-ray images. There are several ways of doing level slicing, but most of them are variations of two basic themes. One approach is to display a high value for all gray levels in the range of interest and a low value for all other gray levels. This transformation, shown in Fig. 3.11(a), produces a binary image. The second approach, based on the transformation shown in Fig. 3.11(b), brightens the desired range of gray levels but preserves the background and gray-level tonalities in the image. Figure 3.11(c) shows a gray-scale image, and Fig. 3.11(d) shows the result of using the transformation in Fig. 3.11(a). Variations of the two transformations shown in Fig. 3.11 are easy to formulate.

Bit-plane slicing

Instead of highlighting gray-level ranges, highlighting the contribution made to total image appearance by specific bits might be desired. Suppose that each pixel in an image is represented by 8 bits. Imagine that the image is composed of eight 1-bit planes, ranging from bit-plane 0 for the least significant bit to bit-plane 7 for the most significant bit. In terms of 8-bit bytes, plane 0 contains all the lowest order bits in the bytes comprising the pixels in the image and plane 7 contains all the high-order bits. Figure 3.12 illustrates these ideas, and Fig. 3.13 shows the various bit planes for the image shown in Fig. 3.13. Note that the higher-order bits (especially the top four) contain the majority of the visually significant data. The other bit planes contribute to more subtle details in the image. Separating a digital image into its bit planes is useful for analyzing the relative importance played by each bit of the image, a process that aids in determining the adequacy of the number of bits used to quantize each pixel. Also, this type of decomposition is useful for image compression, as discussed in Chapter 8.

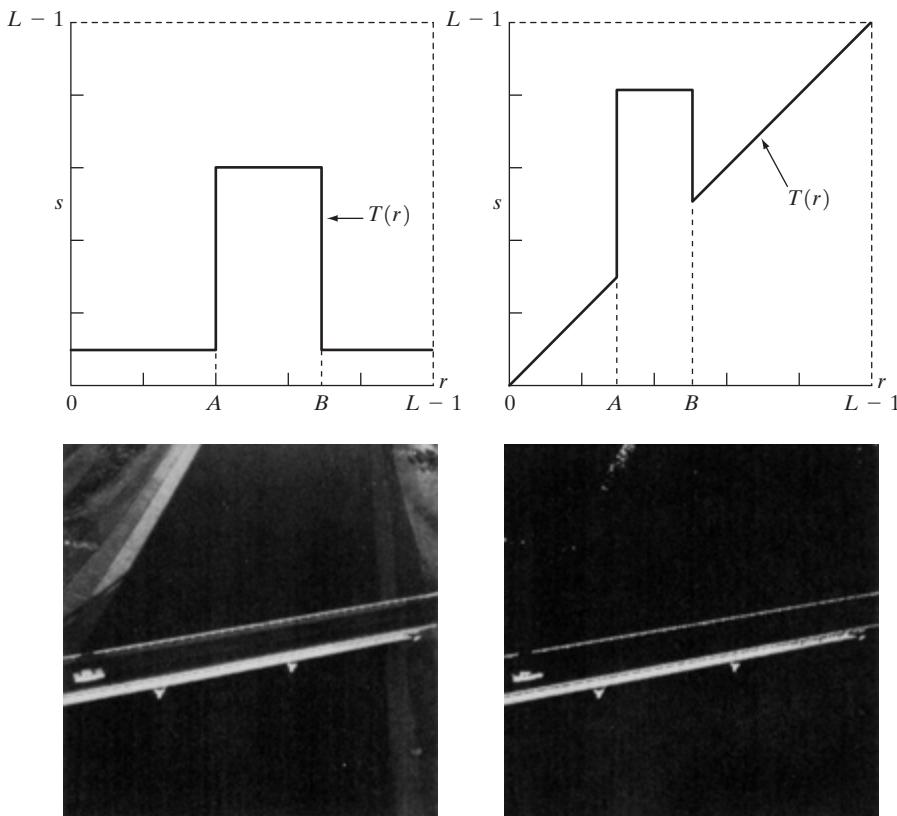


FIGURE 3.11
 (a) This transformation highlights range $[A, B]$ of gray levels and reduces all others to a constant level.
 (b) This transformation highlights range $[A, B]$ but preserves all other levels.
 (c) An image.
 (d) Result of using the transformation in (a).

In terms of bit-plane extraction for an 8-bit image, it is not difficult to show that the (binary) image for bit-plane 7 can be obtained by processing the input image with a thresholding gray-level transformation function that (1) maps all levels in the image between 0 and 127 to one level (for example, 0); and (2) maps all levels between 129 and 255 to another (for example, 255). The binary image for bit-plane 7 in Fig. 3.14 was obtained in just this manner. It is left as an exercise (Problem 3.3) to obtain the gray-level transformation functions that would yield the other bit planes.

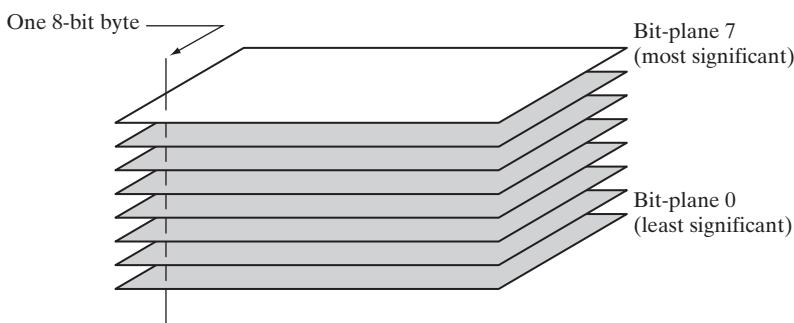


FIGURE 3.12
 Bit-plane representation of an 8-bit image.

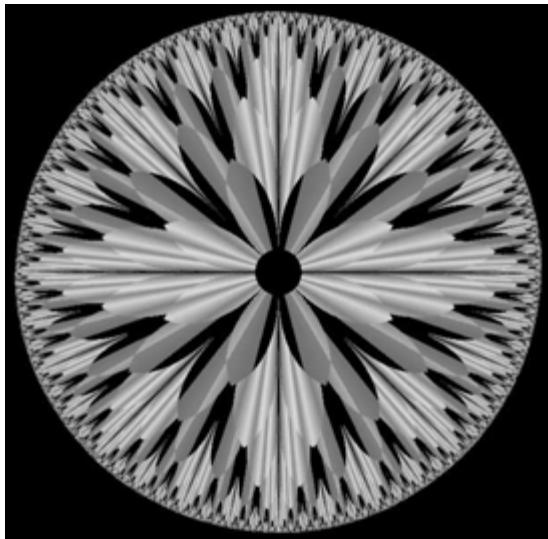


FIGURE 3.13 An 8-bit fractal image. (A fractal is an image generated from mathematical expressions). (Courtesy of Ms. Melissa D. Binde, Swarthmore College, Swarthmore, PA.)



See inside front cover
Consult the book web site
for a review of basic probability theory.

3.3 Histogram Processing

The histogram of a digital image with gray levels in the range $[0, L - 1]$ is a discrete function $h(r_k) = n_k$, where r_k is the k th gray level and n_k is the number of pixels in the image having gray level r_k . It is common practice to normalize a histogram by dividing each of its values by the total number of pixels in the image, denoted by n . Thus, a normalized histogram is given by $p(r_k) = n_k/n$, for $k = 0, 1, \dots, L - 1$. Loosely speaking, $p(r_k)$ gives an estimate of the probability of occurrence of gray level r_k . Note that the sum of all components of a normalized histogram is equal to 1.

Histograms are the basis for numerous spatial domain processing techniques. Histogram manipulation can be used effectively for image enhancement, as shown in this section. In addition to providing useful image statistics, we shall see in subsequent chapters that the information inherent in histograms also is quite useful in other image processing applications, such as image compression and segmentation. Histograms are simple to calculate in software and also lend themselves to economic hardware implementations, thus making them a popular tool for real-time image processing.

As an introduction to the role of histogram processing in image enhancement, consider Fig. 3.15, which is the pollen image of Fig. 3.10 shown in four basic gray-level characteristics: dark, light, low contrast, and high contrast. The right side of the figure shows the histograms corresponding to these images. The horizontal axis of each histogram plot corresponds to gray level values, r_k . The vertical axis corresponds to values of $h(r_k) = n_k$ or $p(r_k) = n_k/n$ if the values are normalized. Thus, as indicated previously, these histogram plots are simply plots of $h(r_k) = n_k$ versus r_k or $p(r_k) = n_k/n$ versus r_k .

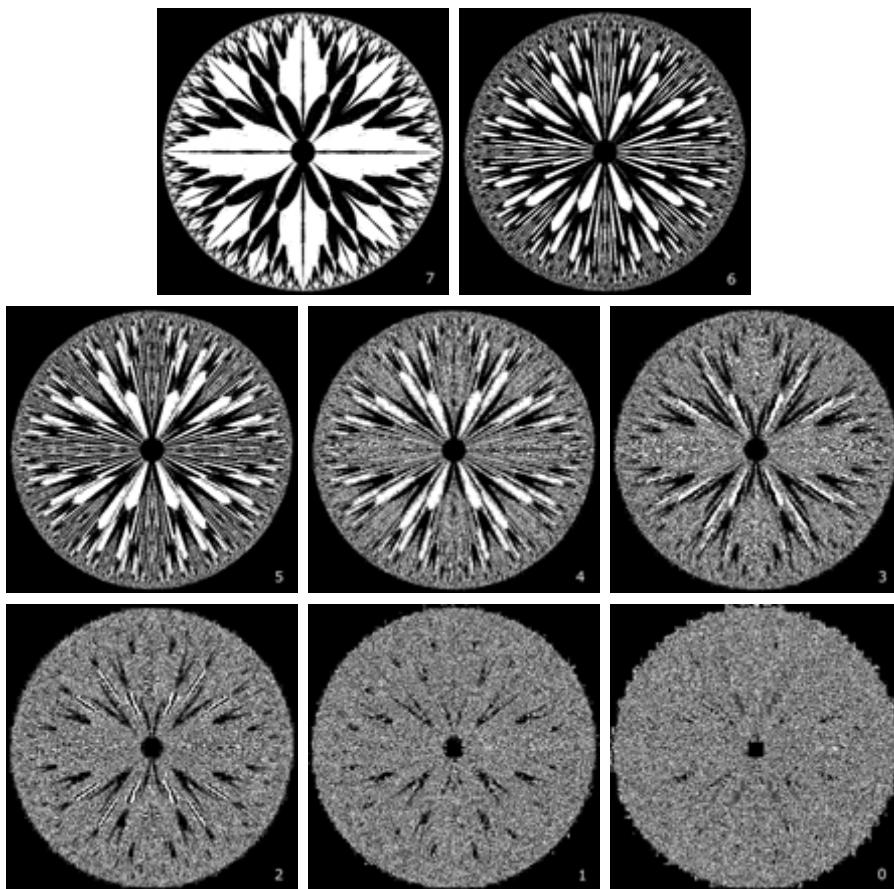


FIGURE 3.14 The eight bit planes of the image in Fig. 3.13. The number at the bottom, right of each image identifies the bit plane.

We note in the dark image that the components of the histogram are concentrated on the low (dark) side of the gray scale. Similarly, the components of the histogram of the bright image are biased toward the high side of the gray scale. An image with low contrast has a histogram that will be narrow and will be centered toward the middle of the gray scale. For a monochrome image this implies a dull, washed-out gray look. Finally, we see that the components of the histogram in the high-contrast image cover a broad range of the gray scale and, further, that the distribution of pixels is not too far from uniform, with very few vertical lines being much higher than the others. Intuitively, it is reasonable to conclude that an image whose pixels tend to occupy the entire range of possible gray levels and, in addition, tend to be distributed uniformly, will have an appearance of high contrast and will exhibit a large variety of gray tones. The net effect will be an image that shows a great deal of gray-level detail and has high dynamic range. It will be shown shortly that it is possible to develop a transformation function that can automatically achieve this effect, based only on information available in the histogram of the input image.

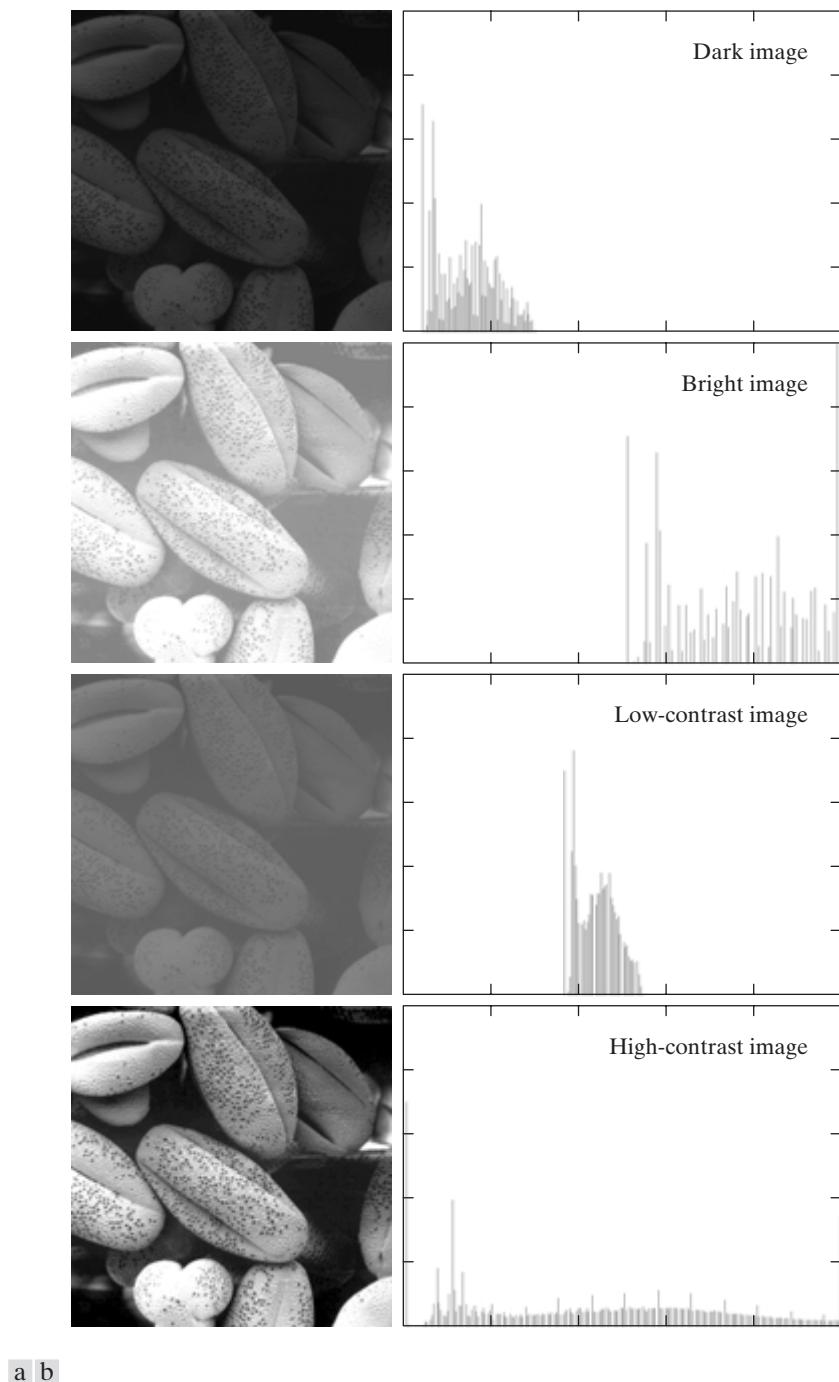


FIGURE 3.15 Four basic image types: dark, light, low contrast, high contrast, and their corresponding histograms. (Original image courtesy of Dr. Roger Heady, Research School of Biological Sciences, Australian National University, Canberra, Australia.)

3.3.1 Histogram Equalization

Consider for a moment continuous functions, and let the variable r represent the gray levels of the image to be enhanced. In the initial part of our discussion we assume that r has been normalized to the interval $[0, 1]$, with $r = 0$ representing black and $r = 1$ representing white. Later, we consider a discrete formulation and allow pixel values to be in the interval $[0, L - 1]$.

For any r satisfying the aforementioned conditions, we focus attention on transformations of the form

$$s = T(r) \quad 0 \leq r \leq 1 \quad (3.3-1)$$

that produce a level s for every pixel value r in the original image. For reasons that will become obvious shortly, we assume that the transformation function $T(r)$ satisfies the following conditions:

- (a) $T(r)$ is single-valued and monotonically increasing in the interval $0 \leq r \leq 1$; and
- (b) $0 \leq T(r) \leq 1$ for $0 \leq r \leq 1$.

The requirement in (a) that $T(r)$ be single valued is needed to guarantee that the inverse transformation will exist, and the monotonicity condition preserves the increasing order from black to white in the output image. A transformation function that is not monotonically increasing could result in at least a section of the intensity range being inverted, thus producing some inverted gray levels in the output image. While this may be a desirable effect in some cases, that is not what we are after in the present discussion. Finally, condition (b) guarantees that the output gray levels will be in the same range as the input levels. Figure 3.16 gives an example of a transformation function that satisfies these two conditions. The inverse transformation from s back to r is denoted

$$r = T^{-1}(s) \quad 0 \leq s \leq 1. \quad (3.3-2)$$

It can be shown by example (Problem 3.8) that even if $T(r)$ satisfies conditions (a) and (b), it is possible that the corresponding inverse $T^{-1}(s)$ may fail to be single valued.

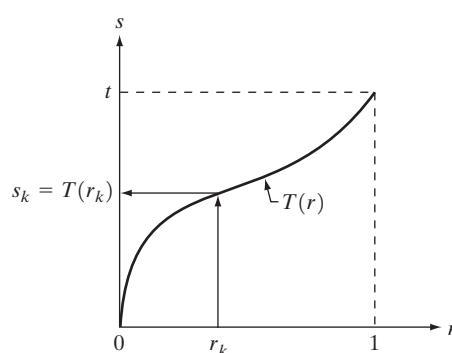


FIGURE 3.16 A gray-level transformation function that is both single valued and monotonically increasing.

The gray levels in an image may be viewed as random variables in the interval $[0, 1]$. One of the most fundamental descriptors of a random variable is its probability density function (PDF). Let $p_r(r)$ and $p_s(s)$ denote the probability density functions of random variables r and s , respectively, where the subscripts on p are used to denote that p_r and p_s are different functions. A basic result from an elementary probability theory is that, if $p_r(r)$ and $T(r)$ are known and $T^{-1}(s)$ satisfies condition (a), then the probability density function $p_s(s)$ of the transformed variable s can be obtained using a rather simple formula:

$$p_s(s) = p_r(r) \left| \frac{dr}{ds} \right|. \quad (3.3-3)$$

Thus, the probability density function of the transformed variable, s , is determined by the gray-level PDF of the input image and by the chosen transformation function.

A transformation function of particular importance in image processing has the form

$$s = T(r) = \int_0^r p_r(w) dw \quad (3.3-4)$$

where w is a dummy variable of integration. The right side of Eq. (3.3-4) is recognized as the cumulative distribution function (CDF) of random variable r . Since probability density functions are always positive, and recalling that the integral of a function is the area under the function, it follows that this transformation function is single valued and monotonically increasing, and, therefore, satisfies condition (a). Similarly, the integral of a probability density function for variables in the range $[0, 1]$ also is in the range $[0, 1]$, so condition (b) is satisfied as well.

Given transformation function $T(r)$, we find $p_s(s)$ by applying Eq. (3.3-3). We know from basic calculus (Leibniz's rule) that the derivative of a definite integral with respect to its upper limit is simply the integrand evaluated at that limit. In other words,

$$\begin{aligned} \frac{ds}{dr} &= \frac{dT(r)}{dr} \\ &= \frac{d}{dr} \left[\int_0^r p_r(w) dw \right] \\ &= p_r(r). \end{aligned} \quad (3.3-5)$$

Substituting this result for dr/ds into Eq. (3.3-3), and keeping in mind that all probability values are positive, yields

$$\begin{aligned} p_s(s) &= p_r(r) \left| \frac{dr}{ds} \right| \\ &= p_r(r) \left| \frac{1}{p_r(r)} \right| \\ &= 1 \quad 0 \leq s \leq 1. \end{aligned} \quad (3.3-6)$$

Because $p_s(s)$ is a probability density function, it follows that it must be zero outside the interval $[0, 1]$ in this case because its integral over all values of s must equal 1. We recognize the form of $p_s(s)$ given in Eq. (3.3-6) as a *uniform* probability density function. Simply stated, we have demonstrated that performing the transformation function given in Eq. (3.3-4) yields a random variable s characterized by a uniform probability density function. It is important to note from Eq. (3.3-4) that $T(r)$ depends on $p_r(r)$, but, as indicated by Eq. (3.3-6), the resulting $p_s(s)$ always is uniform, *independent* of the form of $p_r(r)$.

For discrete values we deal with probabilities and summations instead of probability density functions and integrals. The probability of occurrence of gray level r_k in an image is approximated by

$$p_r(r_k) = \frac{n_k}{n} \quad k = 0, 1, 2, \dots, L - 1 \quad (3.3-7)$$

where, as noted at the beginning of this section, n is the total number of pixels in the image, n_k is the number of pixels that have gray level r_k , and L is the total number of possible gray levels in the image. The discrete version of the transformation function given in Eq. (3.3-4) is

$$\begin{aligned} s_k &= T(r_k) = \sum_{j=0}^k p_r(r_j) \\ &= \sum_{j=0}^k \frac{n_j}{n} \quad k = 0, 1, 2, \dots, L - 1. \end{aligned} \quad (3.3-8)$$

Thus, a processed (output) image is obtained by mapping each pixel with level r_k in the input image into a corresponding pixel with level s_k in the output image via Eq. (3.3-8). As indicated earlier, a plot of $p_r(r_k)$ versus r_k is called a *histogram*. The transformation (mapping) given in Eq. (3.3-8) is called *histogram equalization* or *histogram linearization*. It is not difficult to show (Problem 3.9) that the transformation in Eq. (3.3-8) satisfies conditions (a) and (b) stated previously in this section.

Unlike its continuous counterpart, it cannot be proved in general that this discrete transformation will produce the discrete equivalent of a uniform probability density function, which would be a uniform histogram. However, as will be seen shortly, use of Eq. (3.3-8) does have the general tendency of spreading the histogram of the input image so that the levels of the histogram-equalized image will span a fuller range of the gray scale.

We discussed earlier in this section the many advantages of having gray-level values that cover the entire gray scale. In addition to producing gray levels that have this tendency, the method just derived has the additional advantage that it is fully “automatic.” In other words, given an image, the process of histogram equalization consists simply of implementing Eq. (3.3-8), which is based on information that can be extracted directly from the given image, without the need for further parameter specifications. We note also the simplicity of the computations that would be required to implement the technique.

The inverse transformation from s back to r is denoted by

$$r_k = T^{-1}(s_k) \quad k = 0, 1, 2, \dots, L - 1 \quad (3.3-9)$$

It can be shown (Problem 3.9) that the inverse transformation in Eq. (3.3-9) satisfies conditions (a) and (b) stated previously in this section only if none of the levels, r_k , $k = 0, 1, 2, \dots, L - 1$, are missing from the input image. Although the inverse transformation is not used in histogram equalization, it plays a central role in the histogram-matching scheme developed in the next section. We also discuss in that section details of how to implement histogram processing techniques.

EXAMPLE 3.3:
Histogram
equalization.

■ Figure 3.17(a) shows the four images from Fig. 3.15, and Fig. 3.17(b) shows the result of performing histogram equalization on each of these images. The first three results (top to bottom) show significant improvement. As expected, histogram equalization did not produce a significant visual difference in the fourth image because the histogram of this image already spans the full spectrum of the gray scale. The transformation functions used to generate the images in Fig. 3.17(b) are shown in Fig. 3.18. These functions were generated from the histograms of the original images [see Fig. 3.15(b)] using Eq. (3.3-8). Note that transformation (4) has a basic linear shape, again indicating that the gray levels in the fourth input image are nearly uniformly distributed. As was just noted, we would expect histogram equalization in this case to have negligible effect on the appearance of the image.

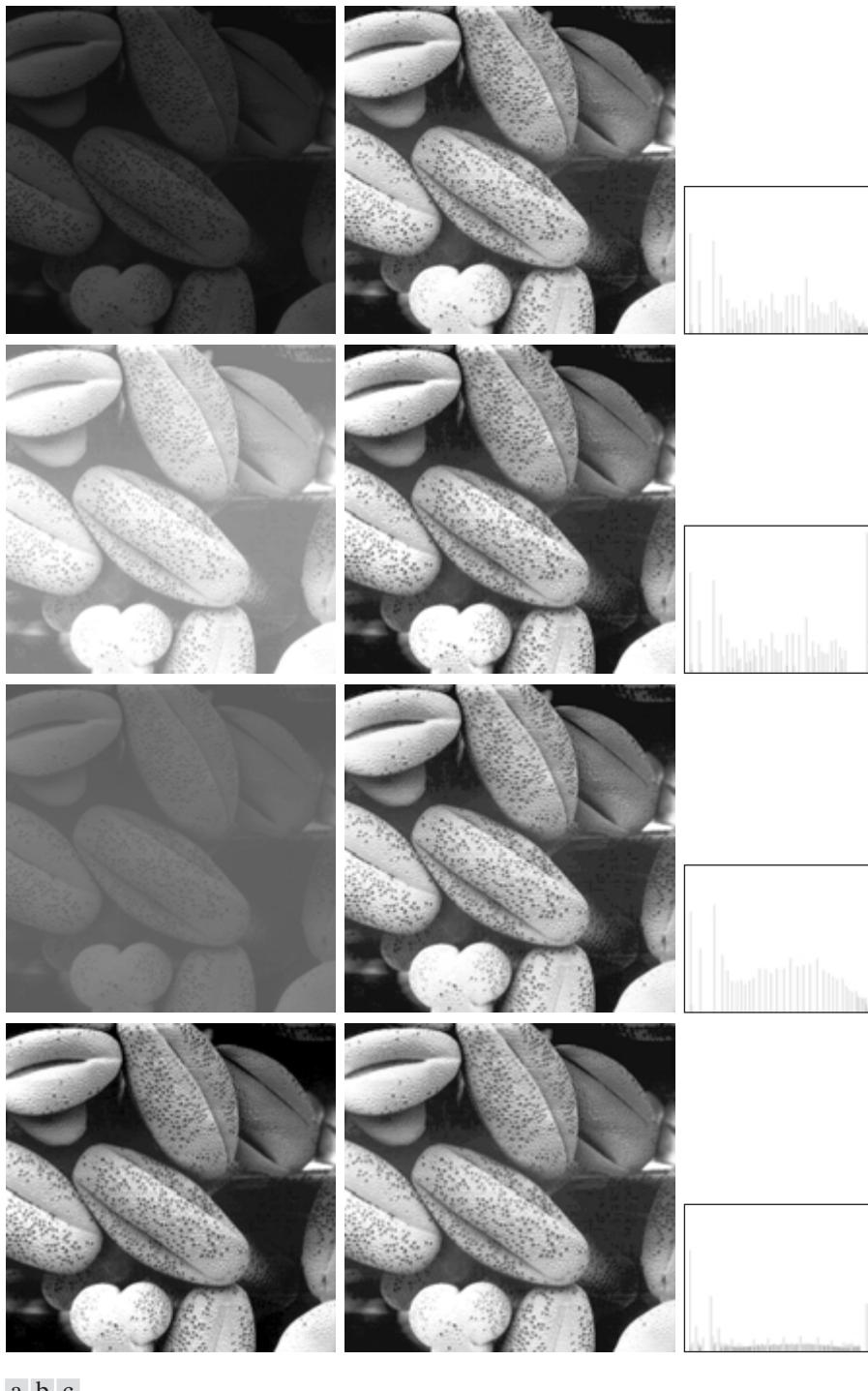
The histograms of the equalized images are shown in Fig. 3.17(c). It is of interest to note that, while all these histograms are different, the histogram-equalized images themselves are visually very similar. This is not unexpected because the difference between the images in the left column is simply one of contrast, not of content. In other words, since the images have the same content, the increase in contrast resulting from histogram equalization was enough to render any gray-level differences in the resulting images visually indistinguishable. Given the significant contrast differences of the images in the left column, this example illustrates the power of histogram equalization as an adaptive enhancement tool. ■

3.3.2 Histogram Matching (Specification)

As indicated in the preceding discussion, histogram equalization automatically determines a transformation function that seeks to produce an output image that has a uniform histogram. When automatic enhancement is desired, this is a good approach because the results from this technique are predictable and the method is simple to implement. We show in this section that there are applications in which attempting to base enhancement on a uniform histogram is not the best approach. In particular, it is useful sometimes to be able to specify the shape of the histogram that we wish the processed image to have. **The method used to generate a processed image that has a specified histogram is called histogram matching or histogram specification.**

Development of the method

Let us return for a moment to continuous gray levels r and z (considered continuous random variables), and let $p_r(r)$ and $p_z(z)$ denote their corresponding continuous probability density functions. In this notation, r and z denote

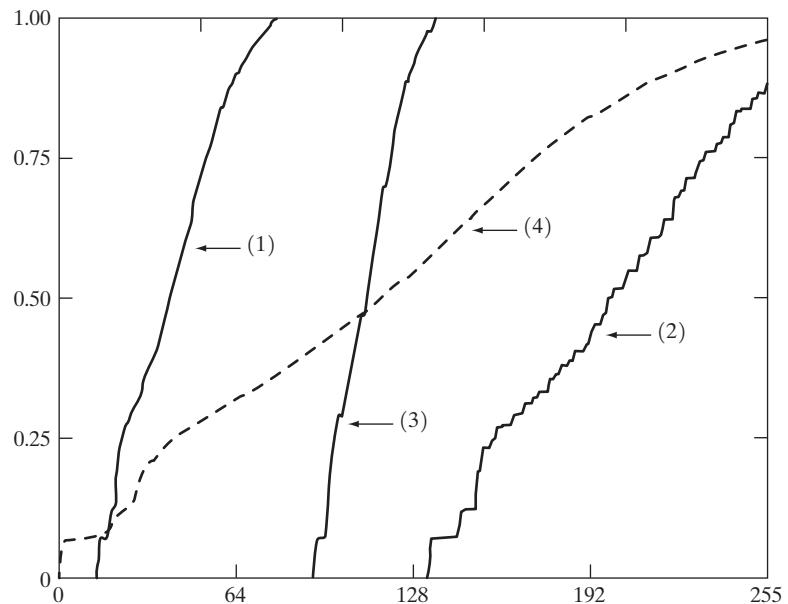


a b c

FIGURE 3.17 (a) Images from Fig. 3.15. (b) Results of histogram equalization. (c) Corresponding histograms.

FIGURE 3.18

Transformation functions (1) through (4) were obtained from the histograms of the images in Fig. 3.17(a), using Eq. (3.3-8).



the gray levels of the input and output (processed) images, respectively. We can estimate $p_r(r)$ from the given input image, while $p_z(z)$ is the *specified* probability density function that we wish the output image to have.

Let s be a random variable with the property

$$s = T(r) = \int_0^r p_r(w) dw \quad (3.3-10)$$

where w is a dummy variable of integration. We recognize this expression as the continuous version of histogram equalization given in Eq. (3.3-4). Suppose next that we define a random variable z with the property

$$G(z) = \int_0^z p_z(t) dt = s \quad (3.3-11)$$

where t is a dummy variable of integration. It then follows from these two equations that $G(z) = T(r)$ and, therefore, that z must satisfy the condition

$$z = G^{-1}(s) = G^{-1}[T(r)]. \quad (3.3-12)$$

The transformation $T(r)$ can be obtained from Eq. (3.3-10) once $p_r(r)$ has been estimated from the input image. Similarly, the transformation function $G(z)$ can be obtained using Eq. (3.3-11) because $p_z(z)$ is given.

Assuming that G^{-1} exists and that it satisfies conditions (a) and (b) in the previous section, Eqs. (3.3-10) through (3.3-12) show that an image with a specified probability density function can be obtained from an input image by using the following procedure: (1) Obtain the transformation function $T(r)$ using Eq. (3.3-10). (2) Use Eq. (3.3-11) to obtain the transformation function $G(z)$. (3) Obtain the inverse transformation function G^{-1} . (4) Obtain the output image

by applying Eq. (3.3-12) to all the pixels in the input image. The result of this procedure will be an image whose gray levels, z , have the specified probability density function $p_z(z)$.

Although the procedure just described is straightforward in principle, it is seldom possible in practice to obtain analytical expressions for $T(r)$ and for G^{-1} . Fortunately, this problem is simplified considerably in the case of discrete values. The price we pay is the same as in histogram equalization, where only an approximation to the desired histogram is achievable. In spite of this, however, some very useful results can be obtained even with crude approximations.

The discrete formulation of Eq. (3.3-10) is given by Eq. (3.3-8), which we repeat here for convenience:

$$\begin{aligned} s_k &= T(r_k) = \sum_{j=0}^k p_r(r_j) \\ &= \sum_{j=0}^k \frac{n_j}{n} \quad k = 0, 1, 2, \dots, L - 1 \end{aligned} \quad (3.3-13)$$

where n is the total number of pixels in the image, n_j is the number of pixels with gray level r_j , and L is the number of discrete gray levels. Similarly, the discrete formulation of Eq. (3.3-11) is obtained from the given histogram $p_z(z_i)$, $i = 0, 1, 2, \dots, L - 1$, and has the form

$$v_k = G(z_k) = \sum_{i=0}^k p_z(z_i) = s_k \quad k = 0, 1, 2, \dots, L - 1. \quad (3.3-14)$$

As in the continuous case, we are seeking values of z that satisfy this equation. The variable v_k was added here for clarity in the discussion that follows. Finally, the discrete version of Eq. (3.3-12) is given by

$$z_k = G^{-1}[T(r_k)] \quad k = 0, 1, 2, \dots, L - 1 \quad (3.3-15)$$

or, from Eq. (3.3-13),

$$z_k = G^{-1}(s_k) \quad k = 0, 1, 2, \dots, L - 1. \quad (3.3-16)$$

Equations (3.3-13) through (3.3-16) are the foundation for implementing histogram matching for digital images. Equation (3.3-13) is a mapping from the levels in the original image into corresponding levels s_k based on the histogram of the original image, which we compute from the pixels in the image. Equation (3.3-14) computes a transformation function G from the given histogram $p_z(z)$. Finally, Eq. (3.3-15) or its equivalent, Eq. (3.3-16), gives us (an approximation of) the desired levels of the image with that histogram. The first two equations can be implemented easily because all the quantities are known. Implementation of Eq. (3.3-16) is straightforward, but requires additional explanation.

Implementation

We start by noting the following: (1) Each set of gray levels $\{r_j\}$, $\{s_j\}$, and $\{z_j\}$, $j = 0, 1, 2, \dots, L - 1$, is a one-dimensional array of dimension $L \times 1$. (2) All mappings from r to s and from s to z are simple table lookups between a given

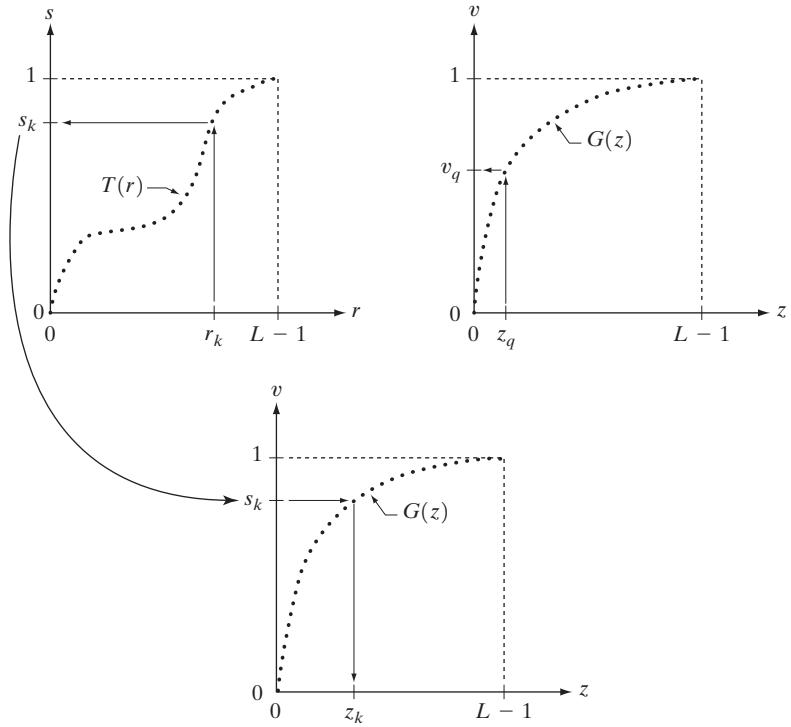
pixel value and these arrays. (3) Each of the elements of these arrays, for example, s_k , contains two important pieces of information: The subscript k denotes the location of the element in the array, and s denotes the value at that location. (4) We need to be concerned only with integer pixel values. For example, in the case of an 8-bit image, $L = 256$ and the elements of each of the arrays just mentioned are integers between 0 and 255. This implies that we now work with gray level values in the interval $[0, L - 1]$ instead of the normalized interval $[0, 1]$ that we used before to simplify the development of histogram processing techniques.

In order to see how histogram matching actually can be implemented, consider Fig. 3.19(a), ignoring for a moment the connection shown between this figure and Fig. 3.19(c). Figure 3.19(a) shows a hypothetical discrete transformation function $s = T(r)$ obtained from a given image. The first gray level in the image, r_1 , maps to s_1 ; the second gray level, r_2 , maps to s_2 ; the k th level r_k maps to s_k ; and so on (the important point here is the *ordered* correspondence between these values). Each value s_j in the array is precomputed using Eq. (3.3-13), so the process of mapping simply uses the actual value of a pixel as an index in an array to determine the corresponding value of s . This process is particularly easy because we are dealing with integers. For example, the s mapping for an 8-bit pixel with value 127 would be found in the 128th position in array $\{s_j\}$ (recall that we start at 0) out of the possible 256 positions. If we stopped here and mapped the value of each pixel of an input image by the

a b
c

FIGURE 3.19

- (a) Graphical interpretation of mapping from r_k to s_k via $T(r)$.
- (b) Mapping of z_q to its corresponding value v_q via $G(z)$.
- (c) Inverse mapping from s_k to its corresponding value of z_k .



method just described, the output would be a histogram-equalized image, according to Eq. (3.3-8).

In order to implement histogram matching we have to go one step further. Figure 3.19(b) is a hypothetical transformation function G obtained from a given histogram $p_z(z)$ by using Eq. (3.3-14). For any z_q , this transformation function yields a corresponding value v_q . This mapping is shown by the arrows in Fig. 3.19(b). Conversely, given any value v_q , we would find the corresponding value z_q from G^{-1} . In terms of the figure, all this means graphically is that we would reverse the direction of the arrows to map v_q into its corresponding z_q . However, we know from the definition in Eq. (3.3-14) that $v = s$ for corresponding subscripts, so we can use exactly this process to find the z_k corresponding to any value s_k that we computed previously from the equation $s_k = T(r_k)$. This idea is shown in Fig. 3.19(c).

Since we really do not have the z 's (recall that finding these values is precisely the objective of histogram matching), we must resort to some sort of iterative scheme to find z from s . The fact that we are dealing with integers makes this a particularly simple process. Basically, because $v_k = s_k$, we have from Eq. (3.3-14) that the z 's for which we are looking must satisfy the equation $G(z_k) = s_k$, or $(G(z_k) - s_k) = 0$. Thus, all we have to do to find the value of z_k corresponding to s_k is to iterate on values of z such that this equation is satisfied for $k = 0, 1, 2, \dots, L - 1$. This is the same thing as Eq. (3.3-16), except that we do not have to find the inverse of G because we are going to iterate on z . Since we are dealing with integers, the closest we can get to satisfying the equation $(G(z_k) - s_k) = 0$ is to let $z_k = \hat{z}$ for each value of k , where \hat{z} is the *smallest* integer in the interval $[0, L - 1]$ such that

$$(G(\hat{z}) - s_k) \geq 0 \quad k = 0, 1, 2, \dots, L - 1. \quad (3.3-17)$$

Given a value s_k , all this means conceptually in terms of Fig. 3.19(c) is that we would start with $\hat{z} = 0$ and increase it in integer steps until Eq. (3.3-17) is satisfied, at which point we let $z_k = \hat{z}$. Repeating this process for all values of k would yield all the required mappings from s to z , which constitutes the implementation of Eq. (3.3-16). In practice, we would not have to start with $\hat{z} = 0$ each time because the values of s_k are known to increase monotonically. Thus, for $k = k + 1$, we would start with $\hat{z} = z_k$ and increment in integer values from there.

The procedure we have just developed for histogram matching may be summarized as follows:

1. Obtain the histogram of the given image.
2. Use Eq. (3.3-13) to precompute a mapped level s_k for each level r_k .
3. Obtain the transformation function G from the given $p_z(z)$ using Eq. (3.3-14).
4. Precompute z_k for each value of s_k using the iterative scheme defined in connection with Eq. (3.3-17).
5. For each pixel in the original image, if the value of that pixel is r_k , map this value to its corresponding level s_k ; then map level s_k into the final level z_k . Use the precomputed values from Steps (2) and (4) for these mappings.

Note that Step (5) implements two mappings for each pixel in the image being processed. The first mapping is nothing more than histogram equalization. If the histogram-equalized image is not required, it obviously would be beneficial to combine both transformations into one in order to save an intermediate step.

Finally, we note that, even in the discrete case, we need to be concerned about G^{-1} satisfying conditions (a) and (b) of the previous section. It is not difficult to show (Problem 3.9) that the only way to guarantee that G^{-1} be single valued and monotonic is to require that G be strictly monotonic (i.e., always increasing), which means simply that none of the values of the specified histogram $p_z(z_i)$ in Eq. (3.3-14) can be zero.

EXAMPLE 3.4:
Comparison
between
histogram
equalization and
histogram
matching.

■ Figure 3.20(a) shows an image of the Mars moon, Phobos, taken by NASA's *Mars Global Surveyor*. Figure 3.20(b) shows the histogram of Fig. 3.20(a). The image is dominated by large, dark areas, resulting in a histogram characterized by a large concentration of pixels in the dark end of the gray scale. At first glance, one might conclude that histogram equalization would be a good approach to enhance this image, so that details in the dark areas become more visible. It is demonstrated in the following discussion that this is not so.

Figure 3.21(a) shows the histogram equalization transformation [Eq. (3.3-8) or (3.3-13)] obtained from the histogram shown in Fig. 3.20(b). The most relevant characteristic of this transformation function is how fast it rises from gray level 0 to a level near 190. This is caused by the large concentration of pixels in the input histogram having levels very near 0. When this transformation is applied to the levels of the input image to obtain a histogram-equalized result, the net effect is to map a very narrow interval of dark pixels into the upper end of the gray scale of the output image. Because numerous pixels in the input image have levels precisely in this interval, we would expect the result to be an

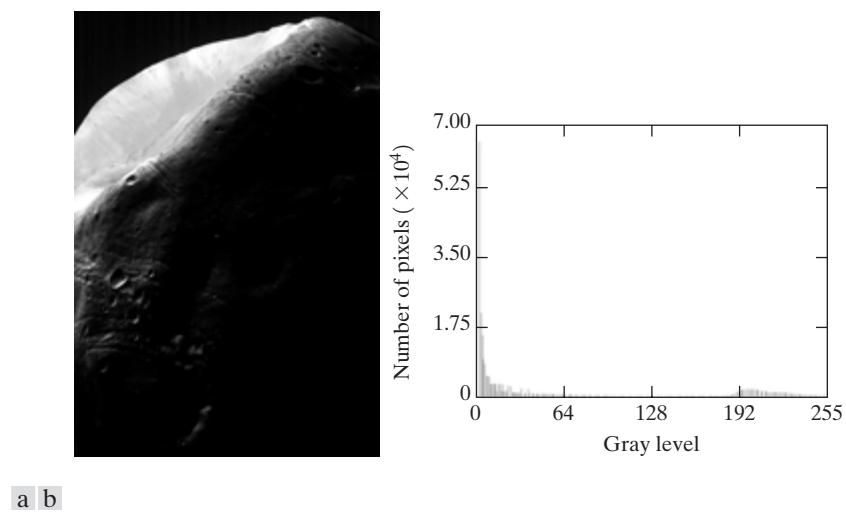


FIGURE 3.20 (a) Image of the Mars moon Photos taken by NASA's *Mars Global Surveyor*. (b) Histogram. (Original image courtesy of NASA.)

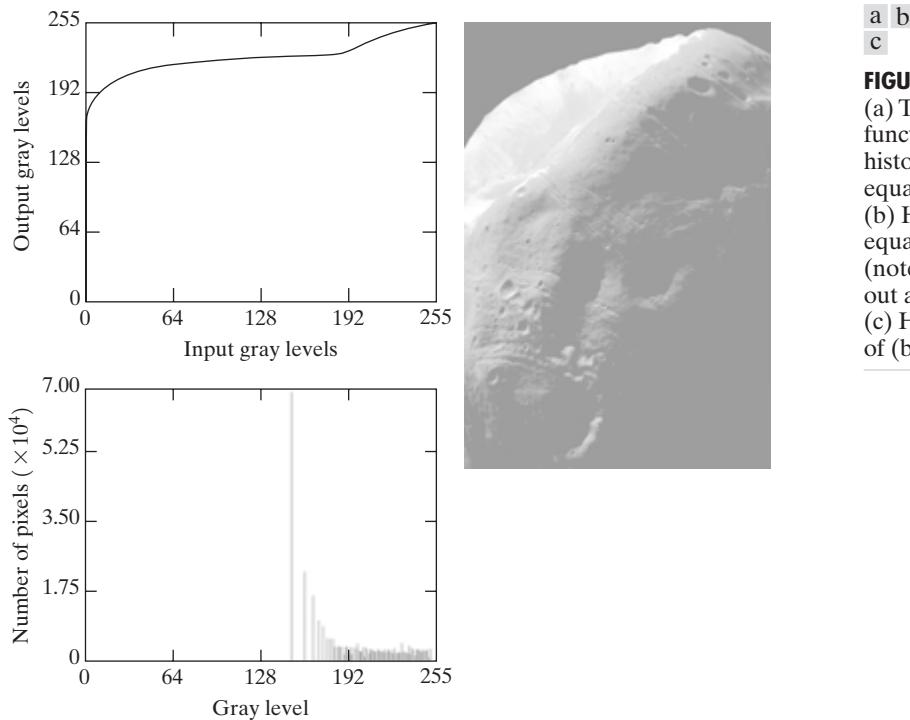


FIGURE 3.21
 (a) Transformation function for histogram equalization.
 (b) Histogram-equalized image (note the washed-out appearance).
 (c) Histogram of (b).

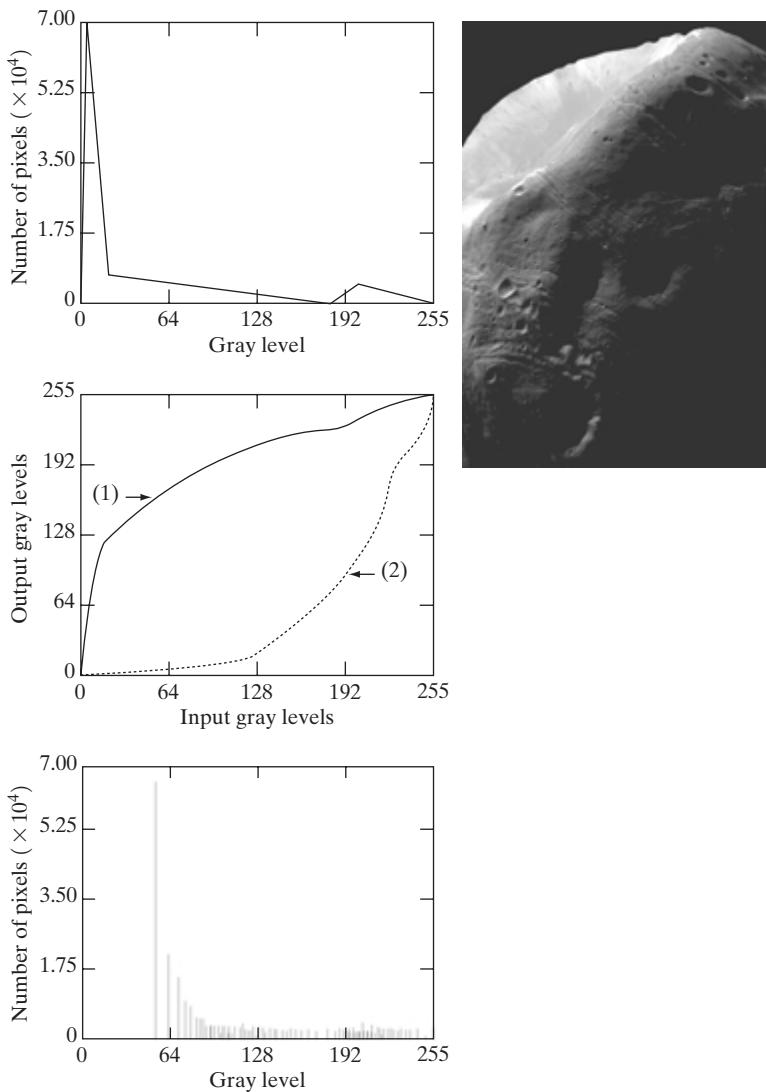
image with a light, washed-out appearance. As shown in Fig. 3.21(b), this is indeed the case. The histogram of this image is shown in Fig. 3.21(c). Note how all the gray levels are biased toward the upper one-half of the gray scale.

Since the problem with the transformation function in Fig. 3.21(a) was caused by a large concentration of pixels in the original image with levels near 0, a reasonable approach is to modify the histogram of that image so that it does not have this property. Figure 3.22(a) shows a *manually specified* function that preserves the general shape of the original histogram, but has a smoother transition of levels in the dark region of the gray scale. Sampling this function into 256 equally spaced discrete values produced the desired specified histogram. The transformation function $G(z)$ obtained from this histogram using Eq. (3.3-14) is labeled transformation (1) in Fig. 3.22(b). Similarly, the inverse transformation $G^{-1}(s)$ from Eq. (3.3-16) [obtained using the iterative technique discussed in connection with Eq. (3.3-17)] is labeled transformation (2) in Fig. 3.22(b). The enhanced image in Fig. 3.22(c) was obtained by applying transformation (2) to the pixels of the histogram-equalized image in Fig. 3.21(b). The improvement of the histogram-specified image over the result obtained by histogram equalization is evident by comparing these two images. It is of interest to note that a rather modest change in the original histogram was all that was required to obtain a significant improvement in enhancement. The histogram of Fig. 3.22(d) is shown in Fig. 3.22(d). The most distinguishing feature of this histogram is how its low end has shifted right toward the lighter region of the gray scale, as desired. ■

a c
b d

FIGURE 3.22

- (a) Specified histogram.
 - (b) Curve (1) is from Eq. (3.3-14), using the histogram in (a); curve (2) was obtained using the iterative procedure in Eq. (3.3-17).
 - (c) Enhanced image using mappings from curve (2).
 - (d) Histogram of (c).
-



Although it probably is obvious by now, we emphasize before leaving this section that histogram specification is, for the most part, a trial-and-error process. One can use guidelines learned from the problem at hand, just as we did in the preceding example. At times, there may be cases in which it is possible to formulate what an “average” histogram should look like and use that as the specified histogram. In cases such as these, histogram specification becomes a straightforward process. In general, however, there are no rules for specifying histograms, and one must resort to analysis on a case-by-case basis for any given enhancement task.

3.3.3 Local Enhancement

The histogram processing methods discussed in the previous two sections are *global*, in the sense that pixels are modified by a transformation function based on the gray-level content of an entire image. Although this global approach is suitable for overall enhancement, there are cases in which it is necessary to enhance details over small areas in an image. The number of pixels in these areas may have negligible influence on the computation of a global transformation whose shape does not necessarily guarantee the desired local enhancement. The solution is to devise transformation functions based on the gray-level distribution—or other properties—in the neighborhood of every pixel in the image. Although processing methods based on neighborhoods are the topic of Section 3.5, we discuss local histogram processing here for the sake of clarity and continuity. The reader will have no difficulty in following the discussion.

The histogram processing techniques previously described are easily adaptable to local enhancement. The procedure is to define a square or rectangular neighborhood and move the center of this area from pixel to pixel. At each location, the histogram of the points in the neighborhood is computed and either a histogram equalization or histogram specification transformation function is obtained. This function is finally used to map the gray level of the pixel centered in the neighborhood. The center of the neighborhood region is then moved to an adjacent pixel location and the procedure is repeated. Since only one new row or column of the neighborhood changes during a pixel-to-pixel translation of the region, updating the histogram obtained in the previous location with the new data introduced at each motion step is possible (Problem 3.11). This approach has obvious advantages over repeatedly computing the histogram over all pixels in the neighborhood region each time the region is moved one pixel location. Another approach used some times to reduce computation is to utilize nonoverlapping regions, but this method usually produces an undesirable checkerboard effect.

Figure 3.23(a) shows an image that has been slightly blurred to reduce its noise content (see Section 3.6.1 regarding blurring). Figure 3.23(b) shows the result of global histogram equalization. As is often the case when this technique is applied to smooth, noisy areas, Fig. 3.23(b) shows considerable enhancement of the noise, with a slight increase in contrast. Note that no new structural details were brought out by this method. However, local histogram equalization using a 7×7 neighborhood revealed the presence of small squares inside the larger dark squares. The small squares were too close in gray level to the larger ones, and their sizes were too small to influence global histogram equalization significantly. Note also the finer noise texture in Fig. 3.23(c), a result of local processing using relatively small neighborhoods. ■

EXAMPLE 3.5:
Enhancement
using local
histograms.

3.3.4 Use of Histogram Statistics for Image Enhancement

Instead of using the image histogram directly for enhancement, we can use instead some statistical parameters obtainable directly from the histogram. Let r denote a discrete random variable representing discrete gray-levels in the range

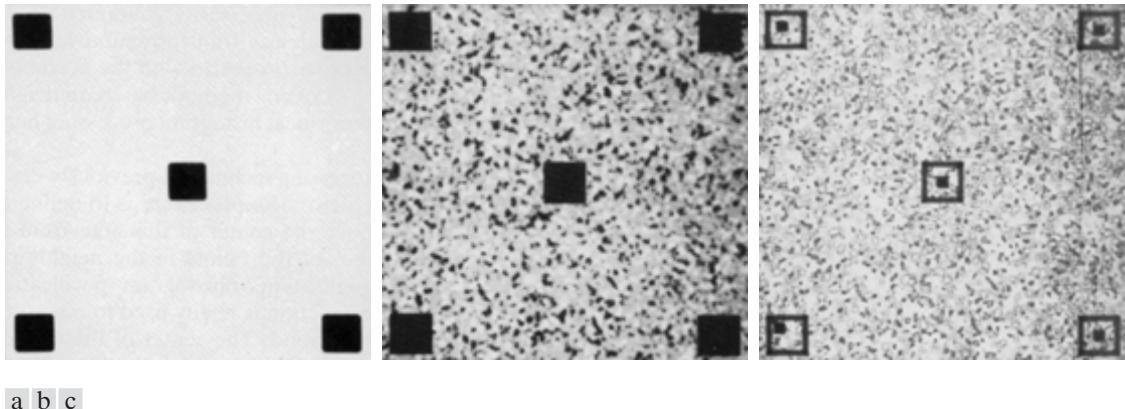


FIGURE 3.23 (a) Original image. (b) Result of global histogram equalization. (c) Result of local histogram equalization using a 7×7 neighborhood about each pixel.

$[0, L - 1]$, and let $p(r_i)$ denote the normalized histogram component corresponding to the i th value of r . As indicated previously in this section, we may view $p(r_i)$ as an estimate of the probability of occurrence of gray level r_i . The n th moment of r about its mean is defined as

$$\mu_n(r) = \sum_{i=0}^{L-1} (r_i - m)^n p(r_i) \quad (3.3-18)$$

where m is the mean value of r (its average gray level):

$$m = \sum_{i=0}^{L-1} r_i p(r_i). \quad (3.3-19)$$

It follows from Eqs. (3.3-18) and (3.3-19) that $\mu_0 = 1$ and $\mu_1 = 0$. The second moment is given by

$$\mu_2(r) = \sum_{i=0}^{L-1} (r_i - m)^2 p(r_i). \quad (3.3-20)$$

We recognize this expression as the variance of r , which is denoted conventionally by $\sigma^2(r)$. The standard deviation is defined simply as the square root of the variance. We will revisit moments in Chapter 11 in connection with image description. In terms of enhancement, however, we are interested primarily in the mean, which is a measure of average gray level in an image, and the variance (or standard deviation), which is a measure of average contrast.

We consider two uses of the mean and variance for enhancement purposes. The *global* mean and variance are measured over an entire image and are useful primarily for gross adjustments of overall intensity and contrast. A much more powerful use of these two measures is in local enhancement, where the *local* mean and variance are used as the basis for making changes that depend on image characteristics in a predefined region about each pixel in the image.

Let (x, y) be the coordinates of a pixel in an image, and let S_{xy} denote a neighborhood (subimage) of specified size, centered at (x, y) . From Eq. (3.3-19) the mean value $m_{S_{xy}}$ of the pixels in S_{xy} can be computed using the expression

$$m_{S_{xy}} = \sum_{(s,t) \in S_{xy}} r_{s,t} p(r_{s,t}) \quad (3.3-21)$$

where $r_{s,t}$ is the gray level at coordinates (s, t) in the neighborhood, and $p(r_{s,t})$ is the neighborhood normalized histogram component corresponding to that value of gray level. Similarly, from Eq. (3.3-20), the gray-level variance of the pixels in region S_{xy} is given by

$$\sigma_{S_{xy}}^2 = \sum_{(s,t) \in S_{xy}} [r_{s,t} - m_{S_{xy}}]^2 p(r_{s,t}). \quad (3.3-22)$$

The local mean is a measure of average gray level in neighborhood S_{xy} , and the variance (or standard deviation) is a measure of contrast in that neighborhood.

An important aspect of image processing using the local mean and variance is the flexibility they afford in developing simple, yet powerful enhancement techniques based on statistical measures that have a close, predictable correspondence with image appearance. We illustrate these characteristics by means of an example.

■ Figure 3.24 shows an SEM (scanning electron microscope) image of a tungsten filament wrapped around a support. The filament in the center of the image and its support are quite clear and easy to study. There is another filament structure on the right side of the image, but it is much darker and its size and other features are not as easily discernable. Local enhancement by contrast manipulation is an ideal approach to try on problems such as this, where part of the image is acceptable, but other parts may contain hidden features of interest.

In this particular case, the problem is to enhance dark areas while leaving the light area as unchanged as possible since it does not require enhancement. We can use the concepts presented in this section to formulate an enhancement method that can tell the difference between dark and light and, at the same time, is capable of enhancing only the dark areas. A measure of whether an area is relatively light or dark at a point (x, y) is to compare the local average gray level $m_{S_{xy}}$ to the average image gray level, called the global mean and denoted M_G . This latter quantity is obtained by letting S encompass the entire image. Thus, we have the first element of our enhancement scheme: We will consider the pixel at a point (x, y) as a candidate for processing if $m_{S_{xy}} \leq k_0 M_G$, where k_0 is a positive constant with value less than 1.0. Since we are interested in enhancing areas that have low contrast, we also need a measure to determine whether the contrast of an area makes it a candidate for enhancement. Thus, we will consider the pixel at a point (x, y) as a candidate for enhancement if $\sigma_{S_{xy}} \leq k_2 D_G$, where D_G is the global standard deviation and k_2 is a positive constant. The value of this constant will be greater than 1.0 if we are interested in enhancing light areas and less than 1.0 for dark areas. Finally, we need to restrict

EXAMPLE 3.6:
Enhancement
based on local
statistics.

the lowest values of contrast we are willing to accept, otherwise the procedure would attempt to enhance even constant areas, whose standard deviation is zero. Thus, we also set a lower limit on the local standard deviation by requiring that $k_1 D_G \leq \sigma_{S_{xy}}$, with $k < k_2$. A pixel at (x, y) that meets all the conditions for local enhancement is processed simply by multiplying it by a specified constant, E , to increase (or decrease) the value of its gray level relative to the rest of the image. The values of pixels that do not meet the enhancement conditions are left unchanged.

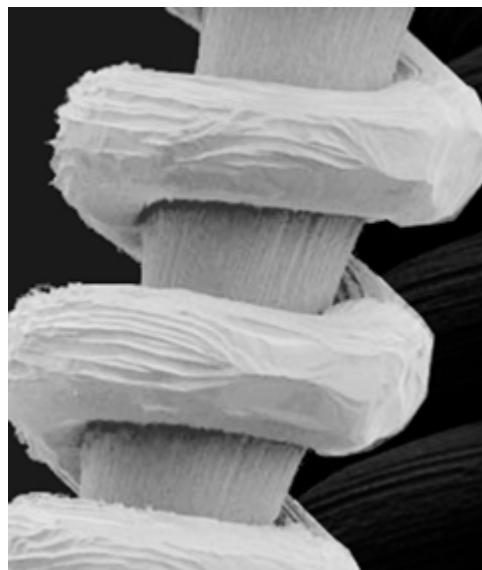
A summary of the enhancement method is as follows. Let $f(x, y)$ represent the value of an image pixel at any image coordinates (x, y) , and let $g(x, y)$ represent the corresponding enhanced pixel at those coordinates. Then

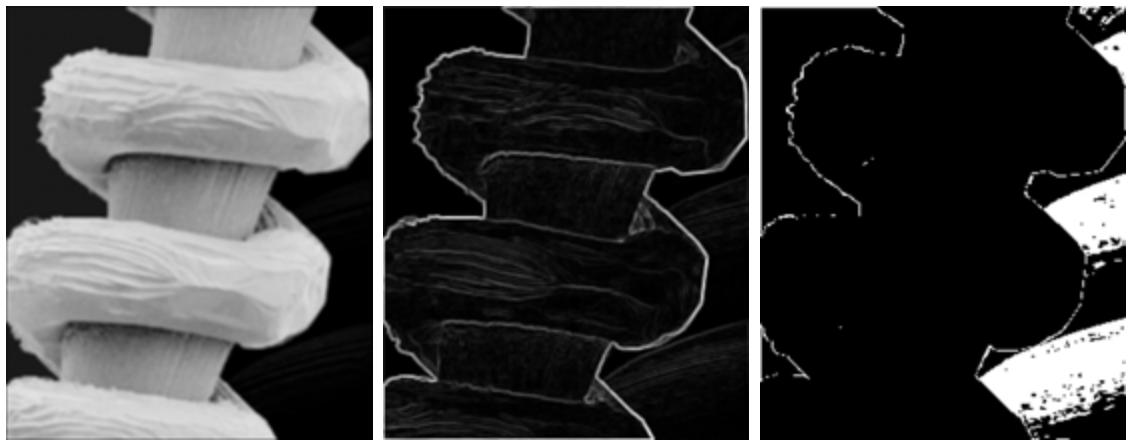
$$g(x, y) = \begin{cases} E \cdot f(x, y) & \text{if } m_{S_{xy}} \leq k_0 M_G \text{ AND } k_1 D_G \leq \sigma_{S_{xy}} \leq k_2 D_G \\ f(x, y) & \text{otherwise} \end{cases}$$

where, as indicated previously, E, k_0, k_1 , and k_2 are specified parameters; M_G is the global mean of the input image; and D_G is its global standard deviation.

Normally, making a successful selection of parameters requires a bit of experimentation to gain familiarity with a given image or class of images. In this case, the following values were selected: $E = 4.0$, $k_0 = 0.4$, $k_1 = 0.02$, and $k_2 = 0.4$. The relatively low value of 4.0 for E was chosen so that, when it was multiplied by the levels in the areas being enhanced (which are dark), the result would still tend toward the dark end of the scale, and thus preserve the general visual balance of the image. The value of k_0 was chosen as somewhat less than half the global mean since it is obvious by looking at the image that the areas that require enhancement definitely are dark enough to be below half the global mean. A similar analysis led to the choice of values for k_1 and k_2 . Choosing these constants is not a difficult task in general, but their choice

FIGURE 3.24 SEM image of a tungsten filament and support, magnified approximately 130×. (Original image courtesy of Mr. Michael Shaffer, Department of Geological Sciences, University of Oregon, Eugene).





a b c

FIGURE 3.25 (a) Image formed from all local means obtained from Fig. 3.24 using Eq. (3.3-21). (b) Image formed from all local standard deviations obtained from Fig. 3.24 using Eq. (3.3-22). (c) Image formed from all multiplication constants used to produce the enhanced image shown in Fig. 3.26.

definitely must be guided by a logical analysis of the enhancement problem at hand. Finally, the choice of size for the local area should be as small as possible in order to preserve detail and keep the computational burden as low as possible. We chose a small (3×3) local region.

Figure 3.25(a) shows the values of $m_{S_{xy}}$ for all values of (x, y) . Since the value of $m_{S_{xy}}$ for each (x, y) is the average of the neighboring pixels in a 3×3 area centered at (x, y) , we expect the result to be similar to the original image, but



FIGURE 3.26
Enhanced SEM image. Compare with Fig. 3.24. Note in particular the enhanced area on the right, bottom side of the image.

slightly blurred. This indeed is the case in Fig. 3.25(a). Figure 3.25(b) shows an image formed using all the values of $\sigma_{S_{xy}}$. Similarly, we can construct an image out of the values that multiply $f(x, y)$ at each coordinate pair (x, y) to form $g(x, y)$. Since the values are either 1 or E , the image is binary, as shown in Fig. 3.25(c). The dark areas correspond to 1 and the light areas to E . Thus, any light point in Fig. 3.25(c) signifies a coordinate pair (x, y) at which the enhancement procedure multiplied $f(x, y)$ by E to produce an enhanced pixel. The dark points represent coordinates at which the procedure did not modify the pixel values.

The enhanced image obtained with the method just described is shown in Fig. 3.26. In comparing this image with the original in Fig. 3.24, we note the obvious detail that has been brought out on the right side of the enhanced image. It is worthwhile to point out that the unenhanced portions of the image (the light areas) were left intact for the most part. We do note the appearance of some small bright dots in the shadow areas where the coil meets the support stem, and around some of the borders between the filament and the background. These are undesirable artifacts created by the enhancement technique. In other words, the points appearing as light dots met the criteria for enhancement and their values were amplified by factor E . Introduction of artifacts is a definite drawback of a method such as the one just described because of the nonlinear way in which they process an image. The key point here, however, is that the image was enhanced in a most satisfactory way as far as bringing out the desired detail. ■

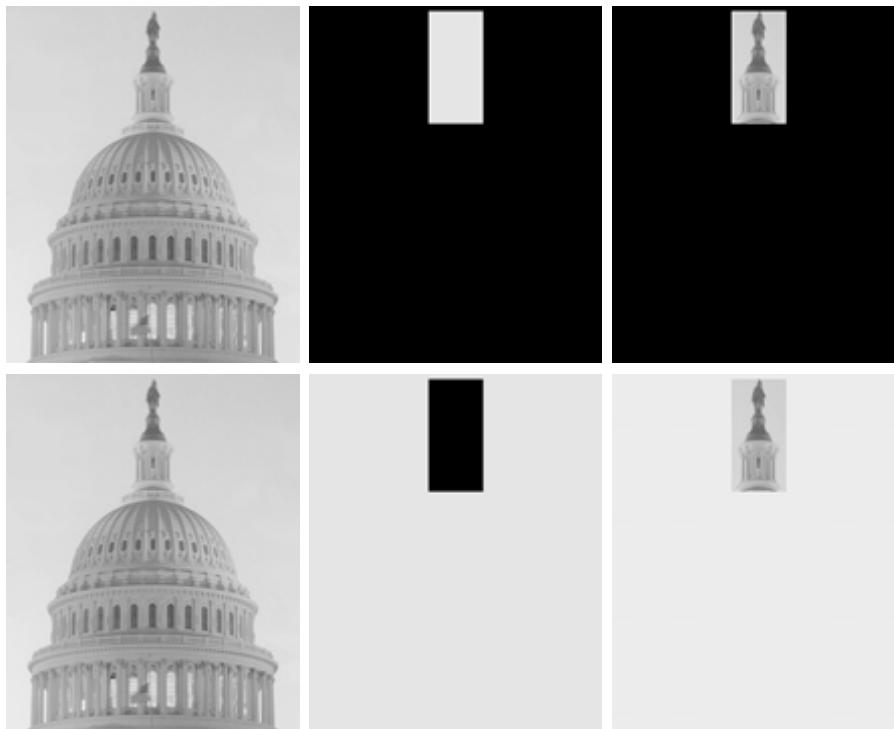
It is not difficult to imagine the numerous ways in which the example just given could be adapted or extended to other situations in which local enhancement is applicable.

3.4 Enhancement Using Arithmetic/Logic Operations

Arithmetic/logic operations involving images are performed on a pixel-by-pixel basis between two or more images (this excludes the logic operation NOT, which is performed on a single image). As an example, subtraction of two images results in a new image whose pixel at coordinates (x, y) is the difference between the pixels in that same location in the two images being subtracted. Depending on the hardware and/or software being used, the actual mechanics of implementing arithmetic/logic operations can be done sequentially, one pixel at a time, or in parallel, where all operations are performed simultaneously.

Logic operations similarly operate on a pixel-by-pixel basis[†]. We need only be concerned with the ability to implement the AND, OR, and NOT logic operators because these three operators are *functionally complete*. In other words, any other logic operator can be implemented by using only these three basic functions. When dealing with logic operations on gray-scale images, pixel values are processed as strings of binary numbers. For example, performing the NOT operation on a black, 8-bit pixel (a string of eight 0's) produces a white pixel

[†] Recall that, for two binary variables a and b : $a \text{AND} b$ yields 1 only when both a and b are 1; otherwise the result is 0. Similarly, $a \text{OR} b$ is 0 when both variables are 0; otherwise the result is 1. Finally, if a is 1, $\text{NOT}(a)$ is 0, and vice versa.



a	b	c
d	e	f

FIGURE 3.27
 (a) Original image. (b) AND image mask.
 (c) Result of the AND operation on images (a) and (b). (d) Original image. (e) OR image mask.
 (f) Result of operation OR on images (d) and (e).

(a string of eight 1's). Intermediate values are processed the same way, changing all 1's to 0's and vice versa. Thus, the NOT logic operator performs the same function as the negative transformation of Eq. (3.2-1). The AND and OR operations are used for *masking*; that is, for selecting subimages in an image, as illustrated in Fig. 3.27. In the AND and OR image masks, light represents a binary 1 and dark represents a binary 0. Masking sometimes is referred to as *region of interest* (ROI) processing. In terms of enhancement, masking is used primarily to isolate an area for processing. This is done to highlight that area and differentiate it from the rest of the image. Logic operations also are used frequently in conjunction with morphological operations, as discussed in Chapter 9.

Of the four arithmetic operations, subtraction and addition (in that order) are the most useful for image enhancement. We consider division of two images simply as multiplication of one image by the reciprocal of the other. Aside from the obvious operation of multiplying an image by a constant to increase its average gray level, image multiplication finds use in enhancement primarily as a masking operation that is more general than the logical masks discussed in the previous paragraph. In other words, multiplication of one image by another can be used to implement gray-level, rather than binary, masks. We give an example in Section 3.8 of how such a masking operation can be a useful tool. In the remainder of this section, we develop and illustrate methods based on subtraction and addition for image enhancement. Other uses of image multiplication are discussed in Chapter 5, in the context of image restoration.

3.4.1 Image Subtraction

The difference between two images $f(x, y)$ and $h(x, y)$, expressed as

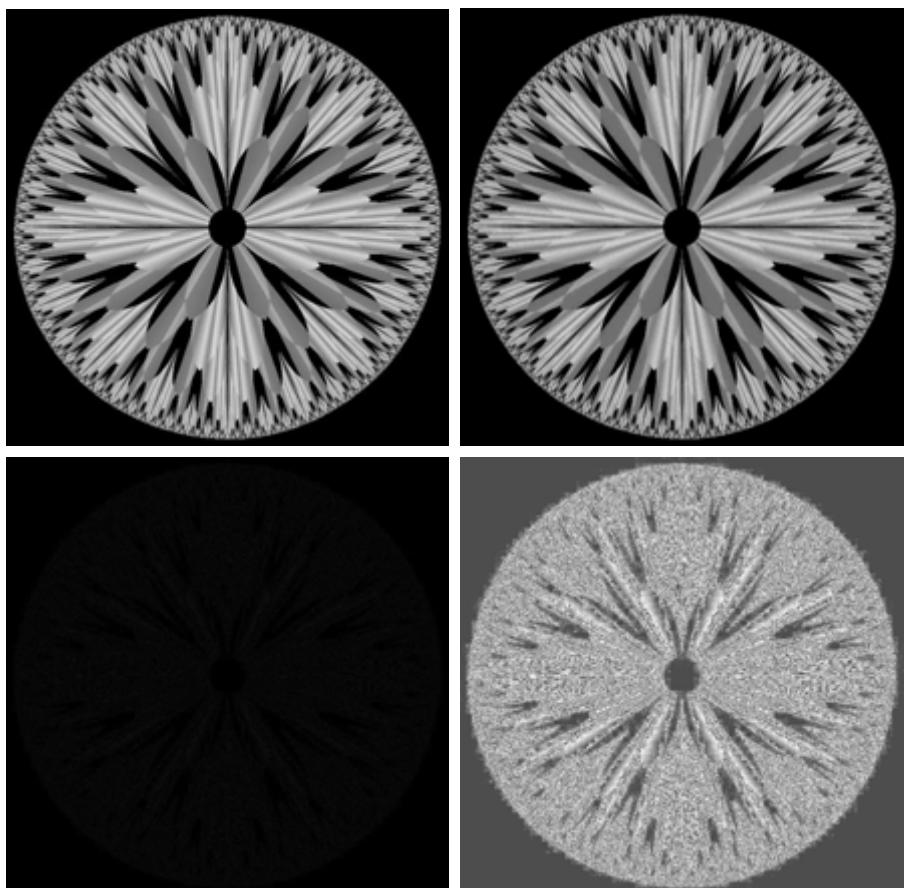
$$g(x, y) = f(x, y) - h(x, y), \quad (3.4-1)$$

is obtained by computing the difference between all pairs of corresponding pixels from f and h . The key usefulness of subtraction is the enhancement of *differences* between images. We illustrate this concept by returning briefly to the discussion in Section 3.2.4, where we showed that the higher-order bit planes of an image carry a significant amount of visually relevant detail, while the lower planes contribute more to fine (often imperceptible) detail. Figure 3.28(a) shows the fractal image used earlier to illustrate the concept of bit planes. Figure 3.28(b) shows the result of discarding (setting to zero) the four least significant bit planes of the original image. The images are nearly identical visually, with the exception of a very slight drop in overall contrast due to less variability of the gray-level values in the image of Fig. 3.28(b). The pixel-by-pixel difference between these two images is shown in Fig. 3.28(c). The differences in pixel values are so small that the difference image appears nearly black when displayed on an 8-bit

a	b
c	d

FIGURE 3.28

- (a) Original fractal image.
- (b) Result of setting the four lower-order bit planes to zero.
- (c) Difference between (a) and (b).
- (d) Histogram-equalized difference image. (Original image courtesy of Ms. Melissa D. Binde, Swarthmore College, Swarthmore, PA).

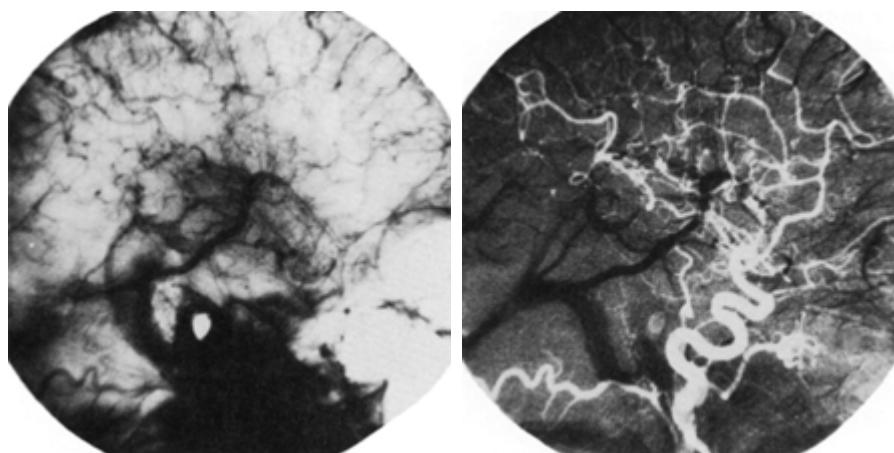


display. In order to bring out more detail, we can perform a contrast stretching transformation, such as those discussed in Sections 3.2 or 3.3. We chose histogram equalization, but an appropriate power-law transformation would have done the job also. The result is shown in Fig. 3.28(d). This is a very useful image for evaluating the effect of setting to zero the lower-order planes.

■ One of the most commercially successful and beneficial uses of image subtraction is in the area of medical imaging called *mask mode radiography*. In this case $h(x, y)$, the *mask*, is an X-ray image of a region of a patient's body captured by an intensified TV camera (instead of traditional X-ray film) located opposite an X-ray source. The procedure consists of injecting a contrast medium into the patient's bloodstream, taking a series of images of the same anatomical region as $h(x, y)$, and subtracting this mask from the series of incoming images after injection of the contrast medium. The net effect of subtracting the mask from each sample in the incoming stream of TV images is that the areas that are different between $f(x, y)$ and $h(x, y)$ appear in the output image as enhanced detail. Because images can be captured at TV rates, this procedure in essence gives a movie showing how the contrast medium propagates through the various arteries in the area being observed.

Figure 3.29(a) shows an X-ray image of the top of a patient's head prior to injection of an iodine medium into the bloodstream. The camera yielding this image was positioned above the patient's head, looking down. As a reference point, the bright spot in the lower one-third of the image is the core of the spinal column. Figure 3.29(b) shows the difference between the mask (Fig. 3.29a) and an image taken some time after the medium was introduced into the bloodstream. The bright arterial paths carrying the medium are unmistakably enhanced in Fig. 3.29(b). These arteries appear quite bright because they are not subtracted out (that is, they are not part of the mask image). The overall background is much darker than that in Fig. 3.29(a) because differences between areas of little change yield low values, which in turn appear as dark shades of gray in the difference image. Note, for instance, that the spinal cord, which is bright in Fig. 3.29(a), appears quite dark in Fig. 3.29(b) as a result of subtraction. ■

EXAMPLE 3.7:
Use of image subtraction in mask mode radiography.



a b

FIGURE 3.29
Enhancement by image subtraction.
(a) Mask image.
(b) An image (taken after injection of a contrast medium into the bloodstream) with mask subtracted out.

A few comments on implementation are in order before we leave this section. In practice, most images are displayed using 8 bits (even 24-bit color images consist of three separate 8-bit channels). Thus, we expect image values not to be outside the range from 0 to 255. The values in a difference image can range from a minimum of -255 to a maximum of 255, so some sort of scaling is required to display the results. There are two principal ways to scale a difference image. One method is to add 255 to every pixel and then divide by 2. It is not guaranteed that the values will cover the entire 8-bit range from 0 to 255, but all pixel values definitely will be within this range. This method is fast and simple to implement, but it has the limitations that the full range of the display may not be utilized and, potentially more serious, the truncation inherent in the division by 2 will generally cause loss in accuracy.

If more accuracy and full coverage of the 8-bit range are desired, then we can resort to another approach. First, the value of the minimum difference is obtained and its negative added to all the pixels in the difference image (this will create a modified difference image whose minimum value is 0). Then, all the pixels in the image are scaled to the interval [0, 255] by multiplying each pixel by the quantity 255/Max, where Max is the maximum pixel value in the modified difference image. It is evident that this approach is considerably more complex and difficult to implement.

Before leaving this section we note also that change detection via image subtraction finds another major application in the area of segmentation, which is the topic of Chapter 10. Basically, segmentation techniques attempt to subdivide an image into regions based on a specified criterion. Image subtraction for segmentation is used when the criterion is “changes.” For instance, in tracking (segmenting) moving vehicles in a sequence of images, subtraction is used to remove all stationary components in an image. What is left should be the moving elements in the image, plus noise.

3.4.2 Image Averaging

Consider a noisy image $g(x, y)$ formed by the addition of noise $\eta(x, y)$ to an original image $f(x, y)$; that is,

$$g(x, y) = f(x, y) + \eta(x, y) \quad (3.4-2)$$

where the assumption is that at every pair of coordinates (x, y) the noise is uncorrelated[†] and has zero average value. The objective of the following procedure is to reduce the noise content by adding a set of noisy images, $\{g_i(x, y)\}$.

If the noise satisfies the constraints just stated, it can be shown (Problem 3.15) that if an image $\bar{g}(x, y)$ is formed by averaging K different noisy images,

$$\bar{g}(x, y) = \frac{1}{K} \sum_{i=1}^K g_i(x, y) \quad (3.4-3)$$

[†]Recall that the variance of a random variable x with mean m is defined as $E[(x - m)^2]$, where $E\{\cdot\}$ is the expected value of the argument. The covariance of two random variables x_i and x_j is defined as $E[(x_i - m_i)(x_j - m_j)]$. If the variables are *uncorrelated*, their covariance is 0.

then it follows that

$$E\{\bar{g}(x, y)\} = f(x, y) \quad (3.4-4)$$

and

$$\sigma_{\bar{g}(x, y)}^2 = \frac{1}{K} \sigma_{\eta(x, y)}^2 \quad (3.4-5)$$

where $E\{\bar{g}(x, y)\}$ is the expected value of \bar{g} , and $\sigma_{\bar{g}(x, y)}^2$ and $\sigma_{\eta(x, y)}^2$ are the variances of \bar{g} and η , all at coordinates (x, y) . The standard deviation at any point in the average image is

$$\sigma_{\bar{g}(x, y)} = \frac{1}{\sqrt{K}} \sigma_{\eta(x, y)}. \quad (3.4-6)$$

As K increases, Eqs. (3.4-5) and (3.4-6) indicate that the variability (noise) of the pixel values at each location (x, y) decreases. Because $E\{\bar{g}(x, y)\} = f(x, y)$, this means that $\bar{g}(x, y)$ approaches $f(x, y)$ as the number of noisy images used in the averaging process increases. In practice, the images $g_i(x, y)$ must be registered (aligned) in order to avoid the introduction of blurring and other artifacts in the output image.

■ An important application of image averaging is in the field of astronomy, where imaging with very low light levels is routine, causing sensor noise frequently to render single images virtually useless for analysis. Figure 3.30(a) shows an image of a galaxy pair called NGC 3314, taken by NASA's Hubble Space Telescope with a wide field planetary camera. NGC 3314 lies about 140 million light-years from Earth, in the direction of the southern-hemisphere constellation Hydra. The bright stars forming a pinwheel shape near the center of the front galaxy have formed recently from interstellar gas and dust. Figure 3.30(b) shows the same image, but corrupted by uncorrelated Gaussian noise with zero mean and a standard deviation of 64 gray levels. This image is useless for all practical purposes. Figures 3.30(c) through (f) show the results of averaging 8, 16, 64, and 128 images, respectively. We see that the result obtained with $K = 128$ is reasonably close to the original in visual appearance.

We can get a better appreciation from Fig. 3.31 for how reduction in the visual appearance of noise takes place as a function of increasing K . This figure shows the difference images between the original [Fig. 3.30(a)] and each of the averaged images in Figs. 3.30(c) through (f). The histograms corresponding to the difference images are also shown in the figure. As usual, the vertical scale in the histograms represents number of pixels and is in the range $[0, 2.6 \times 10^4]$. The horizontal scale represents gray level and is in the range $[0, 255]$. Notice in the histograms that the mean and standard deviation of the difference images decrease as K increases. This is as expected because, according to Eqs. (3.4-3) and (3.4-4), the average image should approach the original as K increases. We can also see the effect of a decreasing mean in the difference images on the left column of Fig. 3.31, which become darker as the K increases.

EXAMPLE 3.8:
Noise reduction
by image
averaging.

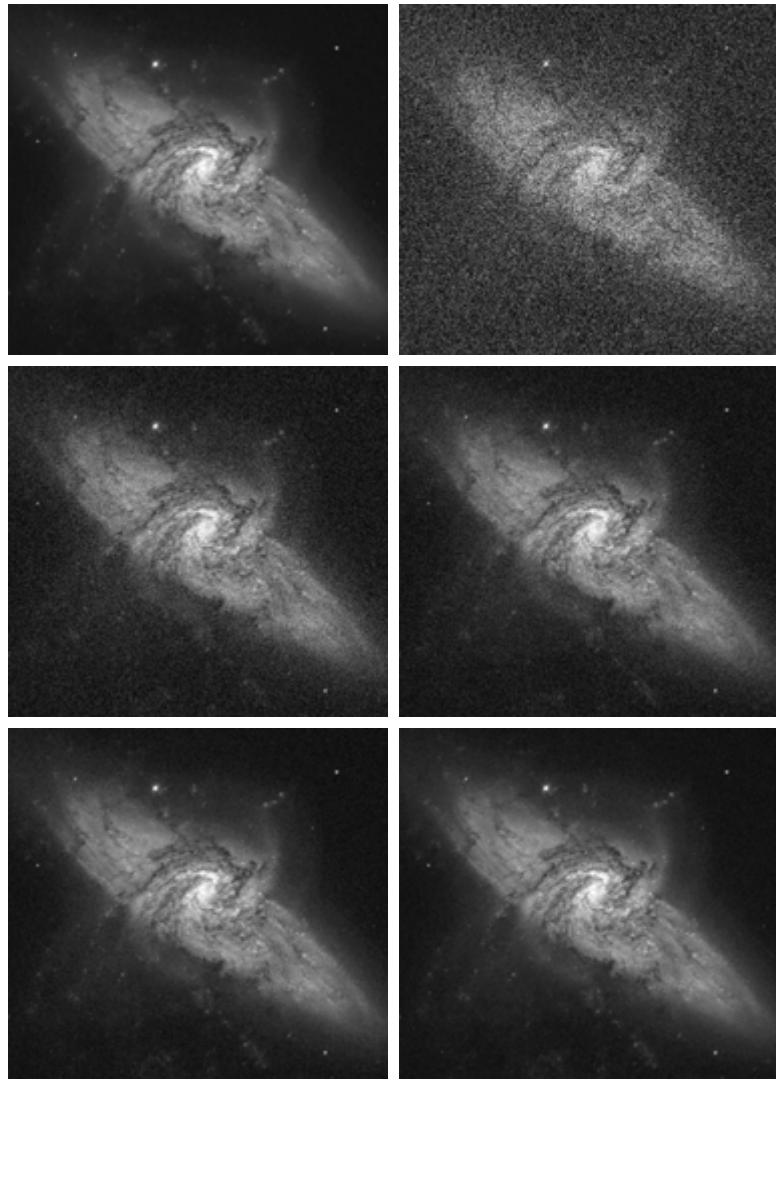


FIGURE 3.30 (a) Image of Galaxy Pair NGC 3314. (b) Image corrupted by additive Gaussian noise with zero mean and a standard deviation of 64 gray levels. (c)–(f) Results of averaging $K = 8, 16, 64$, and 128 noisy images. (Original image courtesy of NASA.)

Addition is the discrete formulation of continuous integration. In astronomical observations, a process equivalent to the method just described is to use the integrating capabilities of CCD or similar sensors for noise reduction by observing the same scene over long periods of time. The net effect, however, is analogous to the procedure just discussed. Cooling the sensor further reduces its noise level. ■

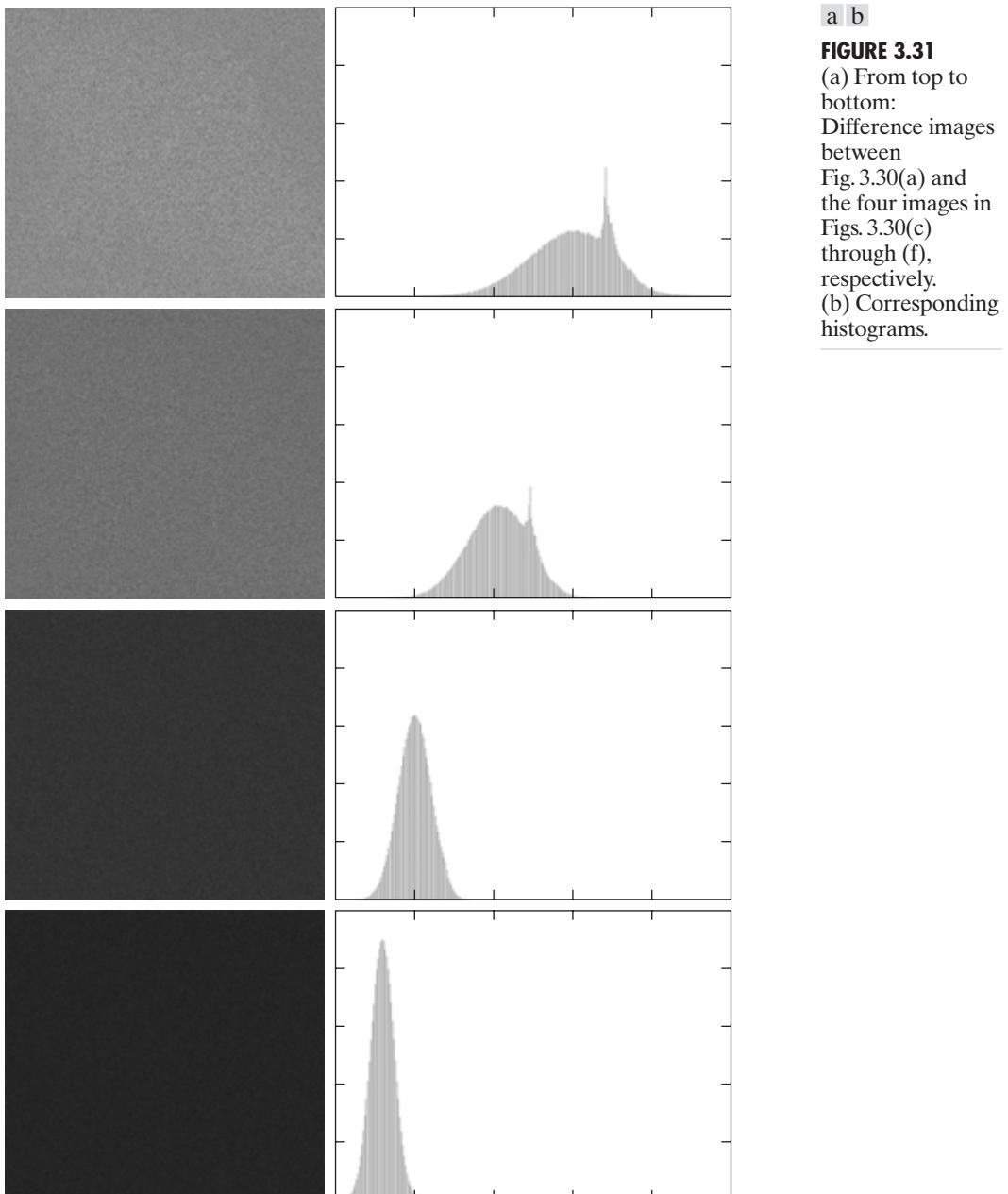


FIGURE 3.31
 (a) From top to bottom:
 Difference images between
 Fig. 3.30(a) and the four images in
 Figs. 3.30(c) through (f),
 respectively.
 (b) Corresponding histograms.

As in the case of image subtraction, adding two or more 8-bit images requires special care when it comes to displaying the result on an 8-bit display. The values in the sum of K , 8-bit images can range from 0 to $255 \times K$. Scaling back to 8 bits in this case consists simply of dividing the result by K . Naturally, some accuracy will be lost in the process, but this is unavoidable if the display has to be limited to 8 bits.

It is possible in some implementations of image averaging to have negative values when noise is added to an image. In fact, in the example just given, this was precisely the case because Gaussian random variables with zero mean and nonzero variance have negative as well as positive values. The images in the example were scaled using the second scaling method discussed at the end of the previous section. That is, the minimum value in a given average image was obtained and its negative was added to the image. Then all the pixels in the modified image were scaled to the range [0, 255] by multiplying each pixel in the modified image by the quantity $255/\text{Max}$, where Max was the maximum pixel value in that image.

3.5 Basics of Spatial Filtering

As mentioned in Section 3.1, some neighborhood operations work with the values of the image pixels in the neighborhood *and* the corresponding values of a subimage that has the same dimensions as the neighborhood. The subimage is called a *filter, mask, kernel, template, or window*, with the first three terms being the most prevalent terminology. The values in a filter subimage are referred to as *coefficients, rather than pixels*.

The concept of filtering has its roots in the use of the Fourier transform for signal processing in the so-called *frequency domain*. This topic is discussed in more detail in Chapter 4. In the present chapter, we are interested in filtering operations that are performed directly on the pixels of an image. We use the term *spatial filtering* to differentiate this type of process from the more traditional frequency domain filtering.

The mechanics of spatial filtering are illustrated in Fig. 3.32. The process consists simply of moving the filter mask from point to point in an image. At each point (x, y) , the response of the filter at that point is calculated using a predefined relationship. For *linear* spatial filtering (see Section 2.6 regarding linearity), the response is given by a sum of products of the filter coefficients and the corresponding image pixels in the area spanned by the filter mask. For the 3×3 mask shown in Fig. 3.32, the result (or response), R , of linear filtering with the filter mask at a point (x, y) in the image is

$$R = w(-1, -1)f(x - 1, y - 1) + w(-1, 0)f(x - 1, y) + \dots \\ + w(0, 0)f(x, y) + \dots + w(1, 0)f(x + 1, y) + w(1, 1)f(x + 1, y + 1),$$

which we see is the sum of products of the mask coefficients with the corresponding pixels directly under the mask. Note in particular that the coefficient $w(0, 0)$ coincides with image value $f(x, y)$, indicating that the mask is centered at (x, y) when the computation of the sum of products takes place. For a mask of size $m \times n$, we assume that $m = 2a + 1$ and $n = 2b + 1$, where a and b are nonnegative integers. All this says is that our focus in the following discussion will be on masks of *odd sizes*, with the smallest meaningful size being 3×3 (we exclude from our discussion the trivial case of a 1×1 mask).

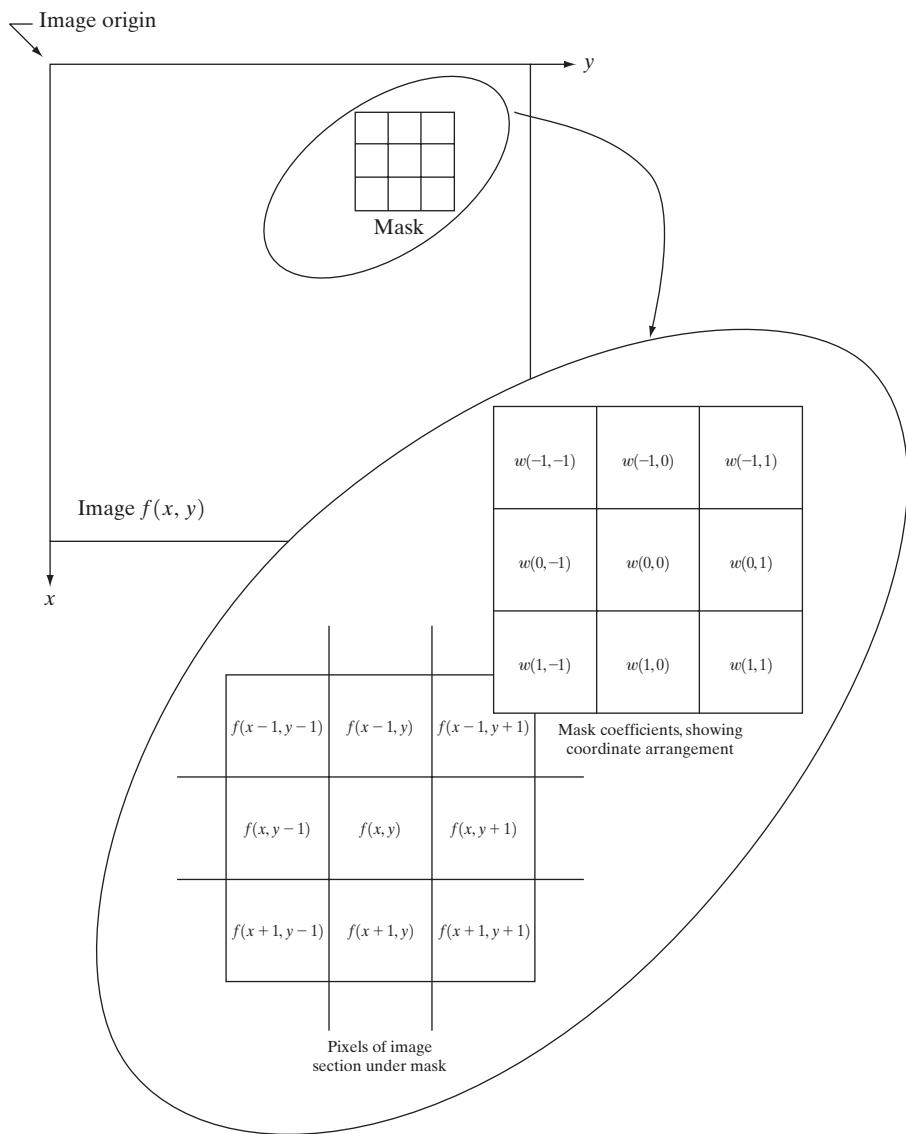


FIGURE 3.32 The mechanics of spatial filtering. The magnified drawing shows a 3×3 mask and the image section directly under it; the image section is shown displaced out from under the mask for ease of readability.

In general, linear filtering of an image f of size $M \times N$ with a filter mask of size $m \times n$ is given by the expression:

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t)f(x + s, y + t) \quad (3.5-1)$$

where, from the previous paragraph, $a = (m - 1)/2$ and $b = (n - 1)/2$. To generate a complete filtered image this equation must be applied for $x = 0, 1, 2, \dots, M - 1$ and $y = 0, 1, 2, \dots, N - 1$. In this way, we are assured that the

mask processes all pixels in the image. It is easily verified when $m = n = 3$ that this expression reduces to the example given in the previous paragraph.

As discussed in Chapter 4, the process of linear filtering given in Eq. (3.5-1) is similar to a frequency domain concept called convolution. For this reason, linear spatial filtering often is referred to as “convolving a mask with an image.” Similarly, filter masks are sometimes called convolution masks. The term convolution kernel also is in common use.

When interest lies on the response, R , of an $m \times n$ mask at any point (x, y) , and not on the mechanics of implementing mask convolution, it is common practice to simplify the notation by using the following expression:

$$\begin{aligned} R &= w_1 z_1 + w_2 z_2 + \dots + w_{mn} z_{mn} \\ &= \sum_{i=1}^{mn} w_i z_i \end{aligned} \quad (3.5-2)$$

where the w 's are mask coefficients, the z 's are the values of the image gray levels corresponding to those coefficients, and mn is the total number of coefficients in the mask. For the 3×3 general mask shown in Fig. 3.33 the response at any point (x, y) in the image is given by

$$\begin{aligned} R &= w_1 z_1 + w_2 z_2 + \dots + w_9 z_9 \\ &= \sum_{i=1}^9 w_i z_i. \end{aligned} \quad (3.5-3)$$

We make special mention of this simple formula because it is seen frequently in the published literature on image processing.

Nonlinear spatial filters also operate on neighborhoods, and the mechanics of sliding a mask past an image are the same as was just outlined. In general, however, the filtering operation is based conditionally on the values of the pixels in the neighborhood under consideration, and they do not explicitly use coefficients in the sum-of-products manner described in Eqs. (3.5-1) and (3.5-2). As shown in Section 3.6.2, for example, noise reduction can be achieved effectively with a nonlinear filter whose basic function is to compute the median gray-level value in the neighborhood in which the filter is located. Computation of the median is a nonlinear operation, as is computation of the variance, which we used in Section 3.3.4.

FIGURE 3.33

Another representation of a general 3×3 spatial filter mask.

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

An important consideration in implementing neighborhood operations for spatial filtering is the issue of what happens when the center of the filter approaches the border of the image. Consider for simplicity a square mask of size $n \times n$. At least one edge of such a mask will coincide with the border of the image when the center of the mask is at a distance of $(n - 1)/2$ pixels away from the border of the image. If the center of the mask moves any closer to the border, one or more rows or columns of the mask will be located outside the image plane. There are several ways to handle this situation. The simplest is to limit the excursions of the center of the mask to be at a distance no less than $(n - 1)/2$ pixels from the border. The resulting filtered image will be smaller than the original, but all the pixels in the filtered image will have been processed with the full mask. If the result is required to be the same size as the original, then the approach typically employed is to filter all pixels only with the section of the mask that is fully contained in the image. With this approach, there will be bands of pixels near the border that will have been processed with a partial filter mask. Other approaches include “padding” the image by adding rows and columns of 0’s (or other constant gray level), or padding by replicating rows or columns. The padding is then stripped off at the end of the process. This keeps the size of the filtered image the same as the original, but the values of the padding will have an effect near the edges that becomes more prevalent as the size of the mask increases. The only way to obtain a perfectly filtered result is to accept a somewhat smaller filtered image by limiting the excursions of the center of the filter mask to a distance no less than $(n - 1)/2$ pixels from the border of the original image.

3.6 Smoothing Spatial Filters

Smoothing filters are used for blurring and for noise reduction. Blurring is used in preprocessing steps, such as removal of small details from an image prior to (large) object extraction, and bridging of small gaps in lines or curves. Noise reduction can be accomplished by blurring with a linear filter and also by non-linear filtering.

3.6.1 Smoothing Linear Filters

The output (response) of a smoothing, linear spatial filter is simply the average of the pixels contained in the neighborhood of the filter mask. These filters sometimes are called *averaging filters*. For reasons explained in Chapter 4, they also are referred to as *lowpass filters*.

The idea behind smoothing filters is straightforward. By replacing the value of every pixel in an image by the average of the gray levels in the neighborhood defined by the filter mask, this process results in an image with reduced “sharp” transitions in gray levels. Because random noise typically consists of sharp transitions in gray levels, the most obvious application of smoothing is noise reduction. However, edges (which almost always are desirable features of an image) also are characterized by sharp transitions in gray levels, so averaging filters have the undesirable side effect that they blur edges. Another application of this type of process includes the smoothing of false contours that result

a b

FIGURE 3.34 Two 3×3 smoothing (averaging) filter masks. The constant multiplier in front of each mask is equal to the sum of the values of its coefficients, as is required to compute an average.

$\frac{1}{9} \times$ <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </table>	1	1	1	1	1	1	1	1	1	$\frac{1}{16} \times$ <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>1</td><td>2</td><td>1</td></tr> <tr><td>2</td><td>4</td><td>2</td></tr> <tr><td>1</td><td>2</td><td>1</td></tr> </table>	1	2	1	2	4	2	1	2	1
1	1	1																	
1	1	1																	
1	1	1																	
1	2	1																	
2	4	2																	
1	2	1																	

from using an insufficient number of gray levels, as discussed in Section 2.4.3. A major use of averaging filters is in the reduction of “irrelevant” detail in an image. By “irrelevant” we mean pixel regions that are small with respect to the size of the filter mask. This latter application is illustrated later in this section.

Figure 3.34 shows two 3×3 smoothing filters. Use of the first filter yields the standard average of the pixels under the mask. This can best be seen by substituting the coefficients of the mask into Eq. (3.5-3):

$$R = \frac{1}{9} \sum_{i=1}^9 z_i,$$

which is the average of the gray levels of the pixels in the 3×3 neighborhood defined by the mask. Note that, instead of being $1/9$, the coefficients of the filter are all 1's. The idea here is that it is computationally more efficient to have coefficients valued 1. At the end of the filtering process the entire image is divided by 9. An $m \times n$ mask would have a normalizing constant equal to $1/mn$. A spatial averaging filter in which all coefficients are equal is sometimes called a *box filter*.

The second mask shown in Fig. 3.34 is a little more interesting. This mask yields a so-called *weighted average*, terminology used to indicate that pixels are multiplied by different coefficients, thus giving more importance (weight) to some pixels at the expense of others. In the mask shown in Fig. 3.34(b) the pixel at the center of the mask is multiplied by a higher value than any other, thus giving this pixel more importance in the calculation of the average. The other pixels are inversely weighted as a function of their distance from the center of the mask. The diagonal terms are further away from the center than the orthogonal neighbors (by a factor of $\sqrt{2}$) and, thus, are weighed less than these immediate neighbors of the center pixel. The basic strategy behind weighing the center point the highest and then reducing the value of the coefficients as a function of increasing distance from the origin is simply an attempt to reduce blurring in the smoothing process. We could have picked other weights to accomplish the same general objective. However, the sum of all the coefficients in the mask of Fig. 3.34(b) is equal to 16, an attractive feature for computer implementation because it has an integer power of 2. In practice, it is difficult in general to see differences between images smoothed by using either of the masks in Fig. 3.34, or similar arrangements, because the area these masks span at any one location in an image is so small.

With reference to Eq. (3.5-1), the general implementation for filtering an $M \times N$ image with a weighted averaging filter of size $m \times n$ (m and n odd) is given by the expression

$$g(x, y) = \frac{\sum_{s=-a}^a \sum_{t=-b}^b w(s, t)f(x + s, y + t)}{\sum_{s=-a}^a \sum_{t=-b}^b w(s, t)} \quad (3.6-1)$$

The parameters in this equation are as defined in Eq. (3.5-1). As before, it is understood that the complete filtered image is obtained by applying Eq. (3.6-1) for $x = 0, 1, 2, \dots, M - 1$ and $y = 0, 1, 2, \dots, N - 1$. The denominator in Eq. (3.6-1) is simply the sum of the mask coefficients and, therefore, it is a constant that needs to be computed only once. Typically, this scale factor is applied to all the pixels of the output image after the filtering process is completed.

■ The effects of smoothing as a function of filter size are illustrated in Fig. 3.35, which shows an original image and the corresponding smoothed results obtained using square averaging filters of sizes $n = 3, 5, 9, 15$, and 35 pixels, respectively. The principal features of these results are as follows: For $n = 3$, we note a general slight blurring throughout the entire image but, as expected, details that are of approximately the same size as the filter mask are affected considerably more. For example, the 3×3 and 5×5 squares, the small letter “a,” and the fine grain noise show significant blurring when compared to the rest of the image. A positive result is that the noise is less pronounced. Note that the jagged borders of the characters and gray circles have been pleasingly smoothed.

The result for $n = 5$ is somewhat similar, with a slight further increase in blurring. For $n = 9$ we see considerably more blurring, and the 20% black circle is not nearly as distinct from the background as in the previous three images, illustrating the blending effect that blurring has on objects whose gray level content is close to that of its neighboring pixels. Note the significant further smoothing of the noisy rectangles. The results for $n = 15$ and 35 are extreme with respect to the sizes of the objects in the image. This type of excessive blurring is generally used to eliminate small objects from an image. For instance, the three small squares, two of the circles, and most of the noisy rectangle areas have been blended into the background of the image in Fig. 3.35(f). Note also in this figure the pronounced black border. This is a result of padding the border of the original image with 0’s (black) and then trimming off the padded area. Some of the black was blended into all filtered images, but became truly objectionable for the images smoothed with the larger filters. ■

EXAMPLE 3.9:
Image smoothing
with masks of
various sizes.

As mentioned earlier, an important application of spatial averaging is to blur an image for the purpose getting a gross representation of objects of interest, such that the intensity of smaller objects blends with the background and larger objects become “bloblike” and easy to detect. The size of the mask establishes the relative size of the objects that will be blended with the background. As an illustration, consider Fig. 3.36(a), which is an image from the Hubble telescope in orbit around the Earth. Figure 3.36(b) shows the result of applying a

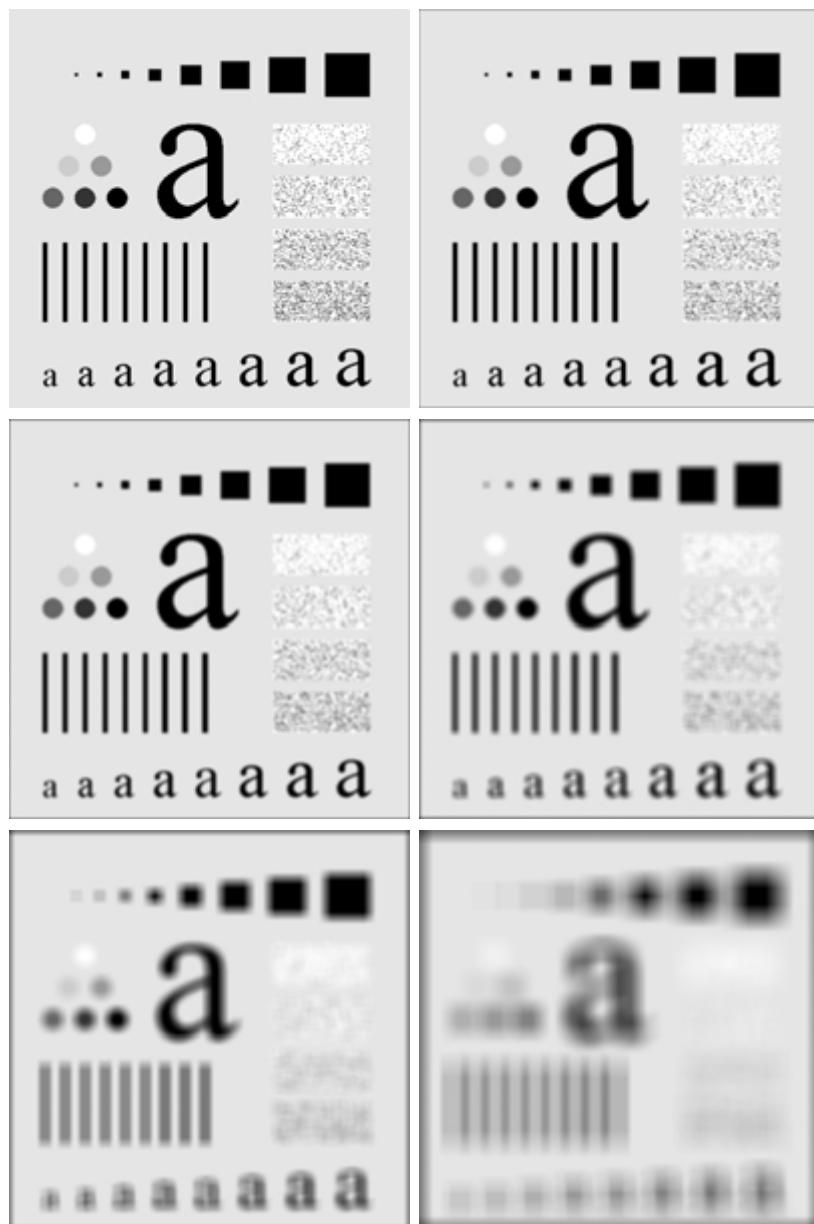
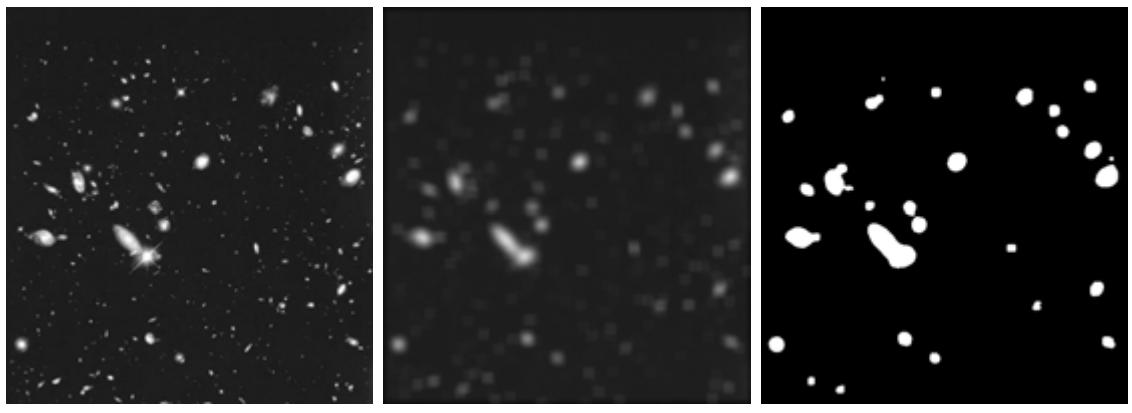


FIGURE 3.35 (a) Original image, of size 500×500 pixels. (b)–(f) Results of smoothing with square averaging filter masks of sizes $n = 3, 5, 9, 15$, and 35 , respectively. The black squares at the top are of sizes $3, 5, 9, 15, 25, 35, 45$, and 55 pixels; their borders are 25 pixels apart. The letters at the bottom range in size from 10 to 24 points, in increments of 2 points; the large letter at the top is 60 points. The vertical bars are 5 pixels wide and 100 pixels high; their separation is 20 pixels. The diameter of the circles is 25 pixels, and their borders are 15 pixels apart; their gray levels range from 0% to 100% black in increments of 20%. The background of the image is 10% black. The noisy rectangles are of size 50×120 pixels.



a b c

FIGURE 3.36 (a) Image from the Hubble Space Telescope. (b) Image processed by a 15×15 averaging mask. (c) Result of thresholding (b). (Original image courtesy of NASA.)

15×15 averaging mask to this image. We see that a number of objects have either blended with the background or their intensity has diminished considerably. It is typical to follow an operation like this with thresholding to eliminate objects based on their intensity. The result of using the thresholding function of Fig. 3.2(b) with a threshold value equal to 25% of the highest intensity in the blurred image is shown in Fig. 3.36(c). Comparing this result with the original image, we see that it is a reasonable representation of what we would consider to be the largest, brightest objects in that image.

3.6.2 Order-Statistics Filters

Order-statistics filters are nonlinear spatial filters whose response is based on ordering (ranking) the pixels contained in the image area encompassed by the filter, and then replacing the value of the center pixel with the value determined by the ranking result. The best-known example in this category is the *median filter*, which, as its name implies, replaces the value of a pixel by the median of the gray levels in the neighborhood of that pixel (the original value of the pixel is included in the computation of the median). Median filters are quite popular because, for certain types of random noise, they provide excellent noise-reduction capabilities, with considerably less blurring than linear smoothing filters of similar size. Median filters are particularly effective in the presence of *impulse noise*, also called *salt-and-pepper noise* because of its appearance as white and black dots superimposed on an image.

The median, ξ , of a set of values is such that half the values in the set are less than or equal to ξ , and half are greater than or equal to ξ . In order to perform median filtering at a point in an image, we first sort the values of the pixel in question and its neighbors, determine their median, and assign this value to that pixel. For example, in a 3×3 neighborhood the median is the 5th largest value, in a 5×5 neighborhood the 13th largest value, and so on. When several values

in a neighborhood are the same, all equal values are grouped. For example, suppose that a 3×3 neighborhood has values (10, 20, 20, 20, 15, 20, 20, 25, 100). These values are sorted as (10, 15, 20, 20, 20, 20, 20, 25, 100), which results in a median of 20. Thus, the principal function of median filters is to force points with distinct gray levels to be more like their neighbors. In fact, isolated clusters of pixels that are light or dark with respect to their neighbors, and whose area is less than $n^2/2$ (one-half the filter area), are eliminated by an $n \times n$ median filter. In this case “eliminated” means forced to the median intensity of the neighbors. Larger clusters are affected considerably less.

Although the median filter is by far the most useful order-statistics filter in image processing, it is by no means the only one. The median represents the 50th percentile of a ranked set of numbers, but the reader will recall from basic statistics that ranking lends itself to many other possibilities. For example, using the 100th percentile results in the so-called *max filter*, which is useful in finding the brightest points in an image. The response of a 3×3 max filter is given by $R = \max\{z_k | k = 1, 2, \dots, 9\}$. The 0th percentile filter is the *min filter*, used for the opposite purpose. Median, max, and mean filters are considered in more detail in Chapter 5.

EXAMPLE 3.10:
Use of median filtering for noise reduction.

Figure 3.37(a) shows an X-ray image of a circuit board heavily corrupted by salt-and-pepper noise. To illustrate the point about the superiority of median filtering over average filtering in situations such as this, we show in Fig. 3.37(b) the result of processing the noisy image with a 3×3 neighborhood averaging mask, and in Fig. 3.37(c) the result of using a 3×3 median filter. The image processed with the averaging filter has less visible noise, but the price paid is significant blurring. The superiority in all respects of median over average filtering in this case is quite evident. In general, median filtering is much better suited than averaging for the removal of additive salt-and-pepper noise. ■

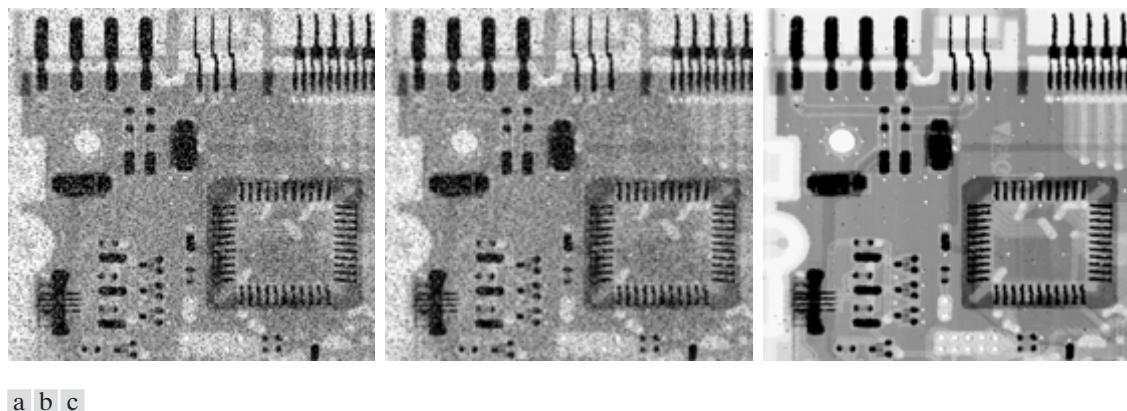


FIGURE 3.37 (a) X-ray image of circuit board corrupted by salt-and-pepper noise. (b) Noise reduction with a 3×3 averaging mask. (c) Noise reduction with a 3×3 median filter. (Original image courtesy of Mr. Joseph E. Pascente, Lixi, Inc.)

3.7 Sharpening Spatial Filters

The principal objective of sharpening is to highlight fine detail in an image or to enhance detail that has been blurred, either in error or as a natural effect of a particular method of image acquisition. Uses of image sharpening vary and include applications ranging from electronic printing and medical imaging to industrial inspection and autonomous guidance in military systems.

In the last section, we saw that image blurring could be accomplished in the spatial domain by pixel averaging in a neighborhood. Since averaging is analogous to integration, it is logical to conclude that sharpening could be accomplished by spatial differentiation. This, in fact, is the case, and the discussion in this section deals with various ways of defining and implementing operators for sharpening by digital differentiation. Fundamentally, the strength of the response of a derivative operator is proportional to the degree of discontinuity of the image at the point at which the operator is applied. Thus, image differentiation enhances edges and other discontinuities (such as noise) and deemphasizes areas with slowly varying gray-level values.

3.7.1 Foundation

In the two sections that follow, we consider in some detail sharpening filters that are based on first- and second-order derivatives, respectively. Before proceeding with that discussion, however, we stop to look at some of the fundamental properties of these derivatives in a digital context. To simplify the explanation, we focus attention on one-dimensional derivatives. In particular, we are interested in the behavior of these derivatives in areas of constant gray level (flat segments), at the onset and end of discontinuities (step and ramp discontinuities), and along gray-level ramps. These types of discontinuities can be used to model noise points, lines, and edges in an image. The behavior of derivatives during transitions into and out of these image features also is of interest.

The derivatives of a digital function are defined in terms of differences. There are various ways to define these differences. However, we require that any definition we use for a first derivative (1) must be zero in flat segments (areas of constant gray-level values); (2) must be nonzero at the onset of a gray-level step or ramp; and (3) must be nonzero along ramps. Similarly, any definition of a second derivative (1) must be zero in flat areas; (2) must be nonzero at the onset and end of a gray-level step or ramp; and (3) must be zero along ramps of constant slope. Since we are dealing with digital quantities whose values are finite, the maximum possible gray-level change also is finite, and the shortest distance over which that change can occur is between adjacent pixels.

A basic definition of the first-order derivative of a one-dimensional function $f(x)$ is the difference

$$\frac{\partial f}{\partial x} = f(x + 1) - f(x).$$

We used a partial derivative here in order to keep the notation the same as when we consider an image function of two variables, $f(x, y)$, at which time we

will be dealing with partial derivatives along the two spatial axes. Use of a partial derivative in the present discussion does not affect in any way the nature of what we are trying to accomplish.

Similarly, we define a second-order derivative as the difference

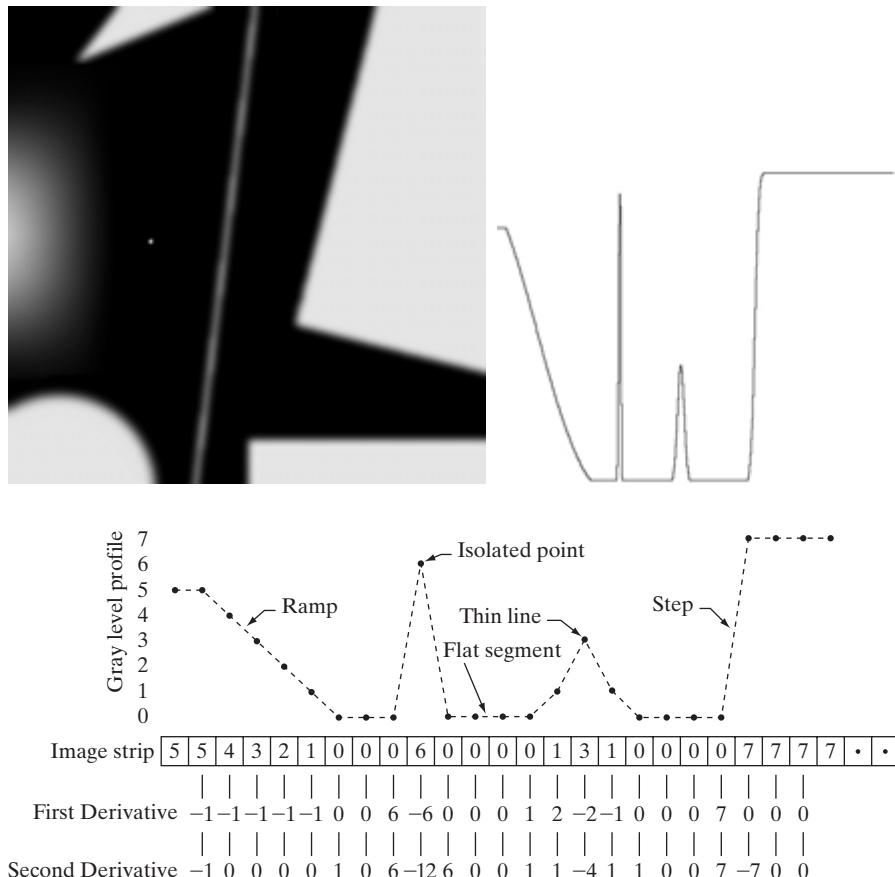
$$\frac{\partial^2 f}{\partial x^2} = f(x + 1) + f(x - 1) - 2f(x).$$

It is easily verified that these two definitions satisfy the conditions stated previously regarding derivatives of the first and second order. To see this, and also to highlight the fundamental similarities and differences between first- and second-order derivatives in the context of image processing, consider the example shown in Fig. 3.38.

Figure 3.38(a) shows a simple image that contains various solid objects, a line, and a single noise point. Figure 3.38(b) shows a horizontal gray-level profile (scan line) of the image along the center and including the noise point. This profile is the one-dimensional function we will use for illustrations regarding this figure. Figure 3.38(c) shows a simplification of the profile, with just enough num-

a
b
c

FIGURE 3.38
 (a) A simple image. (b) 1-D horizontal gray-level profile along the center of the image and including the isolated noise point.
 (c) Simplified profile (the points are joined by dashed lines to simplify interpretation).



bers to make it possible for us to analyze how the first- and second-order derivatives behave as they encounter a noise point, a line, and then the edge of an object. In our simplified diagram the transition in the ramp spans four pixels, the noise point is a single pixel, the line is three pixels thick, and the transition into the gray-level step takes place between adjacent pixels. The number of gray levels was simplified to only eight levels.

Let us consider the properties of the first and second derivatives as we traverse the profile from left to right. First, we note that the first-order derivative is nonzero along the entire ramp, while the second-order derivative is nonzero only at the onset and end of the ramp. Because edges in an image resemble this type of transition, we conclude that first-order derivatives produce “thick” edges and second-order derivatives, much finer ones. Next we encounter the isolated noise point. Here, the response at and around the point is much stronger for the second- than for the first-order derivative. Of course, this is not unexpected. A second-order derivative is much more aggressive than a first-order derivative in enhancing sharp changes. Thus, we can expect a second-order derivative to enhance fine detail (including noise) much more than a first-order derivative. The thin line is a fine detail, and we see essentially the same difference between the two derivatives. If the maximum gray level of the line had been the same as the isolated point, the response of the second derivative would have been stronger for the latter. Finally, in this case, the response of the two derivatives is the same at the gray-level step (in most cases when the transition into a step is not from zero, the second derivative will be weaker). We also note that the second derivative has a transition from positive back to negative. In an image, this shows as a thin double line. This “double-edge” effect is an issue that will be important in Chapter 10, where we use derivatives for edge detection. It is of interest also to note that if the gray level of the thin line had been the same as the step, the response of the second derivative would have been stronger for the line than for the step.

In summary, comparing the response between first- and second-order derivatives, we arrive at the following conclusions. (1) First-order derivatives generally produce thicker edges in an image. (2) Second-order derivatives have a stronger response to fine detail, such as thin lines and isolated points. (3) First-order derivatives generally have a stronger response to a gray-level step. (4) Second-order derivatives produce a double response at step changes in gray level. We also note of second-order derivatives that, for similar changes in gray-level values in an image, their response is stronger to a line than to a step, and to a point than to a line.

In most applications, the second derivative is better suited than the first derivative for image enhancement because of the ability of the former to enhance fine detail. For this, and for reasons of simpler implementation and extensions, we will focus attention initially on uses of the second derivative for enhancement. First-order derivatives are discussed in Section 3.7.3. Although the principle of use of first derivatives in image processing is for edge extraction, they do have important uses in image enhancement. In fact, we show in Section 3.8 that they can be used in conjunction with the second derivative to obtain some impressive enhancement results.

3.7.2 Use of Second Derivatives for Enhancement—The Laplacian

In this section we consider in some detail the use of two-dimensional, second-order derivatives for image enhancement. The approach basically consists of defining a discrete formulation of the second-order derivative and then constructing a filter mask based on that formulation. We are interested in *isotropic* filters, whose response is independent of the direction of the discontinuities in the image to which the filter is applied. In other words, isotropic filters are *rotation invariant*, in the sense that rotating the image and then applying the filter gives the same result as applying the filter to the image first and then rotating the result.

Development of the method

It can be shown (Rosenfeld and Kak [1982]) that the simplest isotropic derivative operator is the *Laplacian*, which, for a function (image) $f(x, y)$ of two variables, is defined as

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \quad (3.7-1)$$

Because derivatives of any order are linear operations, the Laplacian is a linear operator.

In order to be useful for digital image processing, this equation needs to be expressed in discrete form. There are several ways to define a digital Laplacian using neighborhoods. Whatever the definition, however, it has to satisfy the properties of a second derivative outlined in Section 3.7.1. The definition of the digital second derivative given in that section is one of the most used. Taking into account that we now have two variables, we use the following notation for the partial second-order derivative in the x -direction:

$$\frac{\partial^2 f}{\partial x^2} = f(x + 1, y) + f(x - 1, y) - 2f(x, y) \quad (3.7-2)$$

and, similarly in the y -direction, as

$$\frac{\partial^2 f}{\partial y^2} = f(x, y + 1) + f(x, y - 1) - 2f(x, y) \quad (3.7-3)$$

The digital implementation of the two-dimensional Laplacian in Eq. (3.7-1) is obtained by summing these two components:

$$\begin{aligned} \nabla^2 f = & [f(x + 1, y) + f(x - 1, y) + f(x, y + 1) + f(x, y - 1)] \\ & - 4f(x, y). \end{aligned} \quad (3.7-4)$$

This equation can be implemented using the mask shown in Fig. 3.39(a), which gives an isotropic result for rotations in increments of 90° . The mechanics of implementation are given in Eq. (3.5-1) and are illustrated in Section 3.6.1 for the linear smoothing filters. We simply are using different coefficients here.

The diagonal directions can be incorporated in the definition of the digital Laplacian by adding two more terms to Eq. (3.7-4), one for each of the two diagonal directions. The form of each new term is the same as either Eq. (3.7-2)

0	1	0	1	1	1
1	-4	1	1	-8	1
0	1	0	1	1	1
0	-1	0	-1	-1	-1
-1	4	-1	-1	8	-1
0	-1	0	-1	-1	-1

a b
c d

FIGURE 3.39

(a) Filter mask used to implement the digital Laplacian, as defined in Eq. (3.7-4).

(b) Mask used to implement an extension of this equation that includes the diagonal neighbors.

(c) and (d) Two other implementations of the Laplacian.

or (3.7-3), but the coordinates are along the diagonals. Since each diagonal term also contains a $-2f(x, y)$ term, the total subtracted from the difference terms now would be $-8f(x, y)$. The mask used to implement this new definition is shown in Fig. 3.39(b). This mask yields isotropic results for increments of 45° . The other two masks shown in Fig. 3.39 also are used frequently in practice. They are based on a definition of the Laplacian that is the negative of the one we used here. As such, they yield equivalent results, but the difference in sign must be kept in mind when combining (by addition or subtraction) a Laplacian-filtered image with another image.

Because the Laplacian is a derivative operator, its use highlights gray-level discontinuities in an image and deemphasizes regions with slowly varying gray levels. This will tend to produce images that have grayish edge lines and other discontinuities, all superimposed on a dark, featureless background. Background features can be “recovered” while still preserving the sharpening effect of the Laplacian operation simply by adding the original and Laplacian images. As noted in the previous paragraph, it is important to keep in mind which definition of the Laplacian is used. If the definition used has a negative center coefficient, then we *subtract*, rather than add, the Laplacian image to obtain a sharpened result. Thus, the basic way in which we use the Laplacian for image enhancement is as follows:

$$g(x, y) = \begin{cases} f(x, y) - \nabla^2 f(x, y) & \text{if the center coefficient of the Laplacian mask is negative} \\ f(x, y) + \nabla^2 f(x, y) & \text{if the center coefficient of the Laplacian mask is positive.} \end{cases} \quad (3.7-5)$$

Use of this equation is illustrated next.

EXAMPLE 3.11:
Imaging
sharpening with
the Laplacian.

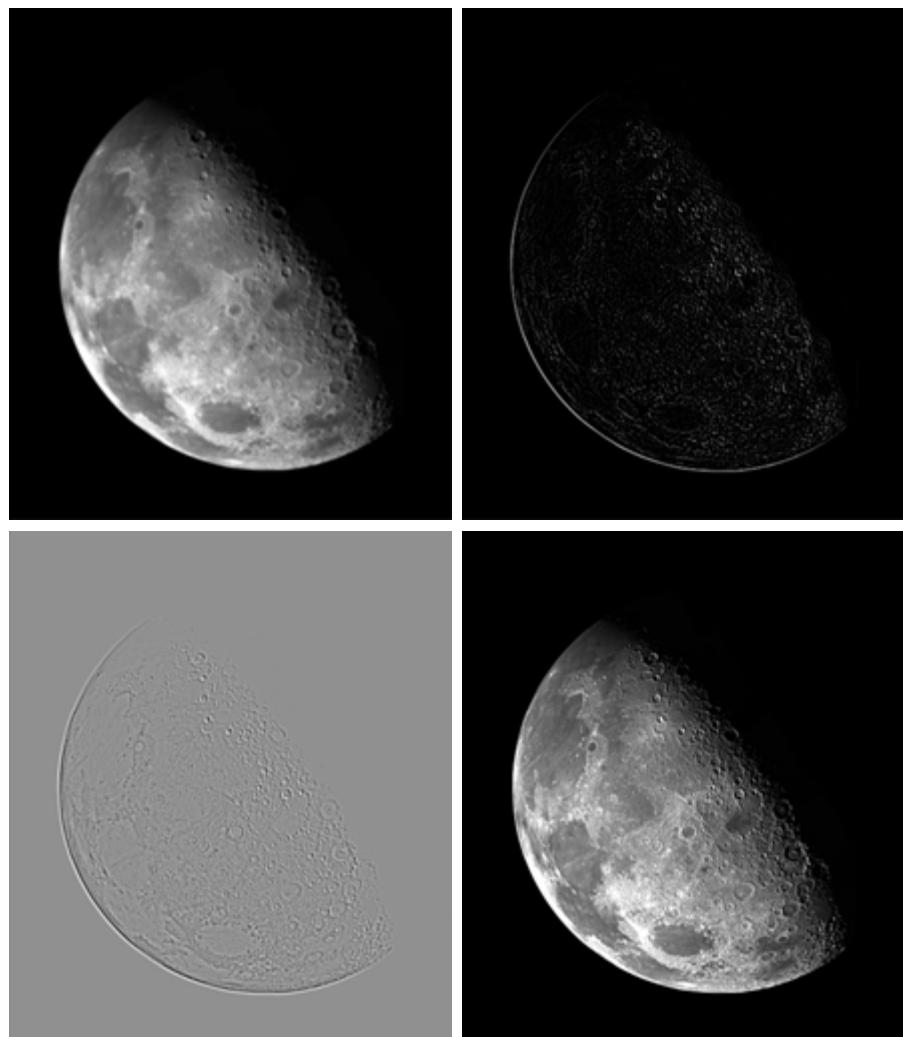
Figure 3.40(a) shows an image of the North Pole of the moon. Figure 3.40(b) shows the result of filtering this image with the Laplacian mask in Fig. 3.39(b). Since the Laplacian image contains both positive and negative values, a typical way to scale it is to use the approach discussed at the end of Section 3.4.1. Sometimes one encounters the absolute value being used for this purpose, but this really is not correct because it produces double lines of nearly equal magnitude, which can be confusing.

The image shown in Fig. 3.40(c) was scaled in the manner just described for display purposes. Note that the dominant features of the image are edges and sharp gray-level discontinuities of various gray-level values. The background, previously near black, is now gray due to the scaling. This grayish appearance is typical of Laplacian images that have been scaled properly. Finally, Fig. 3.40(d)

a b
c d

FIGURE 3.40

- (a) Image of the North Pole of the moon.
- (b) Laplacian-filtered image.
- (c) Laplacian image scaled for display purposes.
- (d) Image enhanced by using Eq. (3.7-5). (Original image courtesy of NASA.)



shows the result obtained using Eq. (3.7-5). The detail in this image is unmistakably clearer and sharper than in the original image. Adding the image to the Laplacian restored the overall gray level variations in the image, with the Laplacian increasing the contrast at the locations of gray-level discontinuities. The net result is an image in which small details were enhanced and the background tonality was perfectly preserved. Results like these have made Laplacian-based enhancement a fundamental tool used frequently for sharpening digital images. ■

Simplifications

In the previous example, we implemented Eq. (3.7-5) by first computing the Laplacian-filtered image and then subtracting it from the original image. This was done for instructional purposes to illustrate each step in the procedure. In practice, Eq. (3.7-5) is usually implemented with one pass of a single mask. The coefficients of the single mask are easily obtained by substituting Eq. (3.7-4) for $\nabla^2 f(x, y)$ in the first line of Eq. (3.7-5):

$$\begin{aligned} g(x, y) &= f(x, y) - [f(x + 1, y) + f(x - 1, y) \\ &\quad + f(x, y + 1) + f(x, y - 1)] + 4f(x, y) \\ &= 5f(x, y) - [f(x + 1, y) + f(x - 1, y) \\ &\quad + f(x, y + 1) + f(x, y - 1)]. \end{aligned} \quad (3.7-6)$$

This equation can be implemented using the mask shown in Fig. 3.41(a). The mask shown in Fig. 3.41(b) would be used if the diagonal neighbors also were included in the calculation of the Laplacian. Identical masks would have resulted if we had substituted the negative of Eq. (3.7-4) into the second line of Eq. (3.7-5).

■ The results obtainable with the mask containing the diagonal terms usually are a little sharper than those obtained with the more basic mask of Fig. 3.41(a). This property is illustrated by the Laplacian-filtered images shown in Figs. 3.41(d) and (e), which were obtained by using the masks in Figs. 3.41(a) and (b), respectively. By comparing the filtered images with the original image shown in Fig. 3.41(c), we note that both masks produced effective enhancement, but the result using the mask in Fig. 3.41(b) is visibly sharper. Figure 3.41(c) is a scanning electron microscope (SEM) image of a tungsten filament following thermal failure; the magnification is approximately 250×.) ■

EXAMPLE 3.12:
Image
enhancement
using a composite
Laplacian mask.

Because the Laplacian is a linear operator, we could have arrived at the same composite masks in Figs. 3.41(a) and (b) by noting that Eq. (3.7-5) is the difference between (sum of) two linear processes. That is, $f(x, y)$ may be viewed as itself processed with a mask that has a unit coefficient in the center and zeros elsewhere. The second term in the equation is the same image processed with one of the Laplacian masks of Fig. 3.39. Due to linearity, the result obtained in Eq. (3.7-5) with the unit-center mask and one of those Laplacian masks would be the same as the result obtained with a single mask formed by subtracting (adding) the Laplacian mask from (to) the unity-center mask.

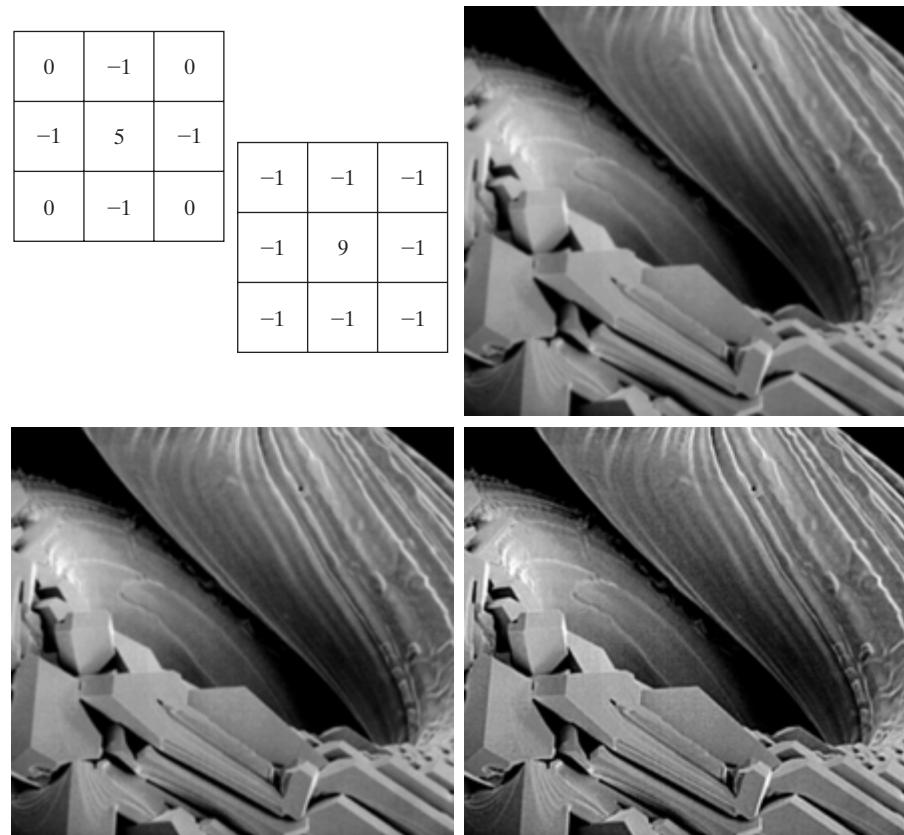


FIGURE 3.41 (a) Composite Laplacian mask. (b) A second composite mask. (c) Scanning electron microscope image. (d) and (e) Results of filtering with the masks in (a) and (b), respectively. Note how much sharper (e) is than (d). (Original image courtesy of Mr. Michael Shaffer, Department of Geological Sciences, University of Oregon, Eugene.)

Unsharp masking and high-boost filtering

A process used for many years in the publishing industry to sharpen images consists of subtracting a blurred version of an image from the image itself. This process, called *unsharp masking*, is expressed as

$$f_s(x, y) = f(x, y) - \bar{f}(x, y) \quad (3.7-7)$$

where $f_s(x, y)$ denotes the sharpened image obtained by unsharp masking, and $\bar{f}(x, y)$ is a blurred version of $f(x, y)$. The origin of unsharp masking is in dark-room photography, where it consists of clamping together a blurred negative to a corresponding positive film and then developing this combination to produce a sharper image.

A slight further generalization of unsharp masking is called *high-boost filtering*. A high-boost filtered image, f_{hb} , is defined at any point (x, y) as

$$f_{hb}(x, y) = Af(x, y) - \bar{f}(x, y) \quad (3.7-8)$$

0	-1	0	-1	-1	-1
-1	$A + 4$	-1	-1	$A + 8$	-1
0	-1	0	-1	-1	-1

a b

FIGURE 3.42 The high-boost filtering technique can be implemented with either one of these masks, with $A \geq 1$.

where $A \geq 1$ and, as before, \bar{f} is a blurred version of f . This equation may be written as

$$f_{\text{hb}}(x, y) = (A - 1)f(x, y) + f(x, y) - \bar{f}(x, y). \quad (3.7-9)$$

By using Eq. (3.7-7), we obtain

$$f_{\text{hb}}(x, y) = (A - 1)f(x, y) + f_s(x, y) \quad (3.7-10)$$

as the expression for computing a high-boost-filtered image.

Equation (3.7-10) is applicable in general and does not state explicitly how the sharp image is obtained. If we elect to use the Laplacian, then we know that $f_s(x, y)$ can be obtained using Eq. (3.7-5). In this case, Eq. (3.7-10) becomes

$$f_{\text{hb}} = \begin{cases} Af(x, y) - \nabla^2 f(x, y) & \text{if the center coefficient of the} \\ & \text{Laplacian mask is negative} \\ Af(x, y) + \nabla^2 f(x, y) & \text{if the center coefficient of the} \\ & \text{Laplacian mask is positive.} \end{cases} \quad (3.7-11)$$

High-boost filtering can be implemented with one pass using either of the two masks shown in Fig. 3.42. Note that, when $A = 1$, high-boost filtering becomes “standard” Laplacian sharpening. As the value of A increases past 1, the contribution of the sharpening process becomes less and less important. Eventually, if A is large enough, the high-boost image will be approximately equal to the original image multiplied by a constant.

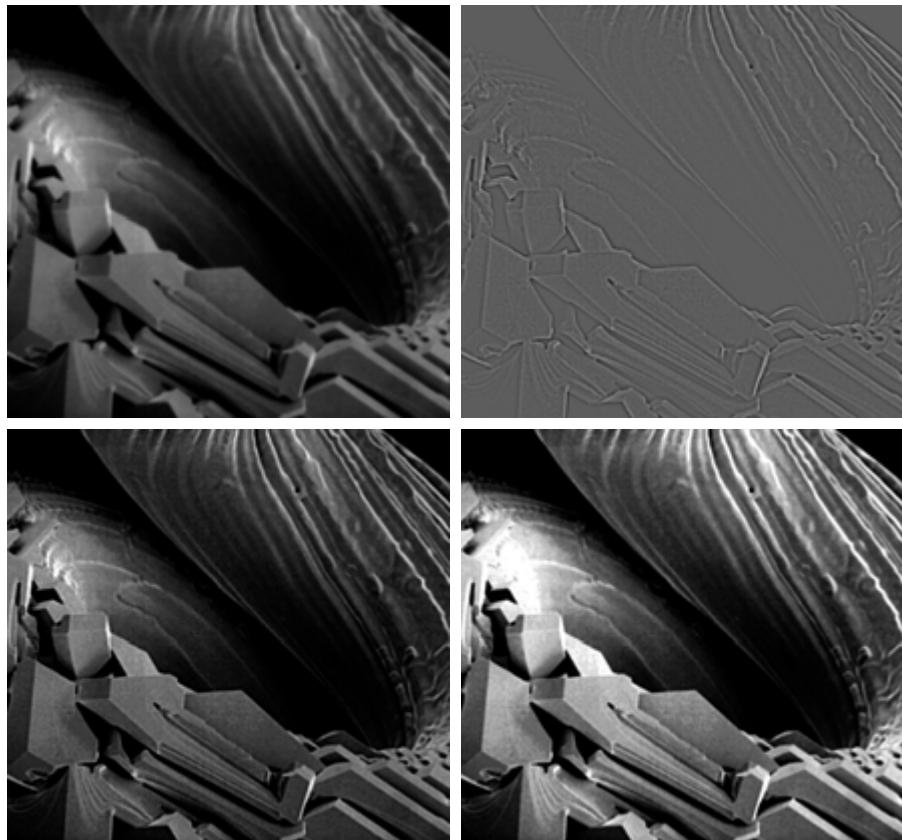
■ One of the principal applications of boost filtering is when the input image is darker than desired. By varying the boost coefficient, it generally is possible to obtain an overall increase in average gray level of the image, thus helping to brighten the final result. Figure 3.43 shows such an application. Part (a) of this figure is a darker version of the image in Fig. 3.41(c). Figure 3.43(b) shows the Laplacian computed using the mask in Fig. 3.42(b), with $A = 0$. Figure 3.43(c) was obtained using the mask in Fig. 3.42(b) with $A = 1$. As expected, the image has been sharpened, but it is still as dark as the original. Finally, Fig. 3.43(d) shows the result of using $A = 1.7$. This is a much more acceptable result, in which the average gray level has increased, thus making the image lighter and more natural. ■

EXAMPLE 3.13:
Image enhancement with a high-boost filter.

a	b
c	d

FIGURE 3.43

- (a) Same as Fig. 3.41(c), but darker.
 (a) Laplacian of (a) computed with the mask in Fig. 3.42(b) using $A = 0$.
 (c) Laplacian enhanced image using the mask in Fig. 3.42(b) with $A = 1$. (d) Same as (c), but using $A = 1.7$.



3.7.3 Use of First Derivatives for Enhancement—The Gradient

First derivatives in image processing are implemented using the magnitude of the gradient. For a function $f(x, y)$, the gradient of f at coordinates (x, y) is defined as the two-dimensional column vector

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}. \quad (3.7-12)$$

The magnitude of this vector is given by

$$\begin{aligned} \nabla f &= \text{mag}(\nabla f) \\ &= [G_x^2 + G_y^2]^{1/2} \\ &= \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{1/2}. \end{aligned} \quad (3.7-13)$$

The components of the gradient vector itself are linear operators, but the magnitude of this vector obviously is not because of the squaring and square root

operations. On the other hand, the partial derivatives in Eq. (3.7-12) are not rotation invariant (isotropic), but the magnitude of the gradient vector is. Although it is not strictly correct, the magnitude of the gradient vector often is referred to as the *gradient*. In keeping with tradition, we will use this term in the following discussions, explicitly referring to the vector or its magnitude only in cases where confusion is likely.

The computational burden of implementing Eq. (3.7-13) over an entire image is not trivial, and it is common practice to approximate the magnitude of the gradient by using absolute values instead of squares and square roots:

$$\nabla f \approx |G_x| + |G_y|. \quad (3.7-14)$$

This equation is simpler to compute and it still preserves relative changes in gray levels, but the isotropic feature property is lost in general. However, as in the case of the Laplacian, the isotropic properties of the digital gradient defined in the following paragraph are preserved only for a limited number of rotational increments that depend on the masks used to approximate the derivatives. As it turns out, the most popular masks used to approximate the gradient give the same result only for vertical and horizontal edges and thus the isotropic properties of the gradient are preserved only for multiples of 90°. These results are independent of whether Eq. (3.7-13) or (3.7-14) is used, so nothing of significance is lost in using the simpler of the two equations.

As in the case of the Laplacian, we now define digital approximations to the preceding equations, and from there formulate the appropriate filter masks. In order to simplify the discussion that follows, we will use the notation in Fig. 3.44(a) to denote image points in a 3×3 region. For example, the center point, z_5 , denotes $f(x, y)$, z_1 denotes $f(x - 1, y - 1)$, and so on. As indicated in Section 3.7.1, the simplest approximations to a first-order derivative that satisfy the conditions stated in that section are $G_x = (z_8 - z_5)$ and $G_y = (z_6 - z_5)$. Two other definitions proposed by Roberts [1965] in the early development of digital image processing use cross differences:

$$G_x = (z_9 - z_5) \quad \text{and} \quad G_y = (z_8 - z_6). \quad (3.7-15)$$

If we elect to use Eq. (3.7-13), then we compute the gradient as

$$\nabla f = [(z_9 - z_5)^2 + (z_8 - z_6)^2]^{1/2} \quad (3.7-16)$$

If we use absolute values, then substituting the quantities in Eq. (3.7-15) into Eq. (3.7-14) gives us the following approximation to the gradient:

$$\nabla f \approx |z_9 - z_5| + |z_8 - z_6|. \quad (3.7-17)$$

This equation can be implemented with the two masks shown in Figs. 3.44(b) and (c). These masks are referred to as the *Roberts cross-gradient operators*.

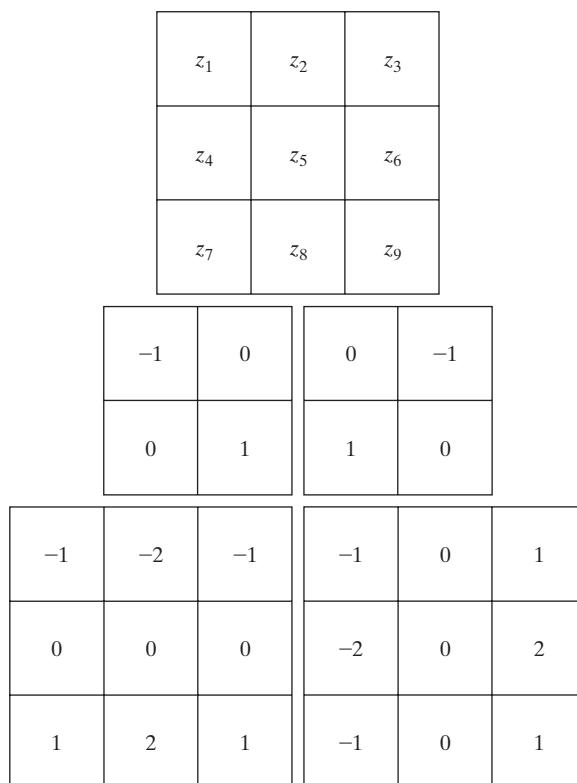
Masks of even size are awkward to implement. The smallest filter mask in which we are interested is of size 3×3 . An approximation using absolute values, still at point z_5 , but using a 3×3 mask, is

$$\begin{aligned} \nabla f \approx & |(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)| \\ & + |(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)|. \end{aligned} \quad (3.7-18)$$

a
b
c
d
e

FIGURE 3.44

A 3×3 region of an image (the z 's are gray-level values) and masks used to compute the gradient at point labeled z_5 . All masks coefficients sum to zero, as expected of a derivative operator.



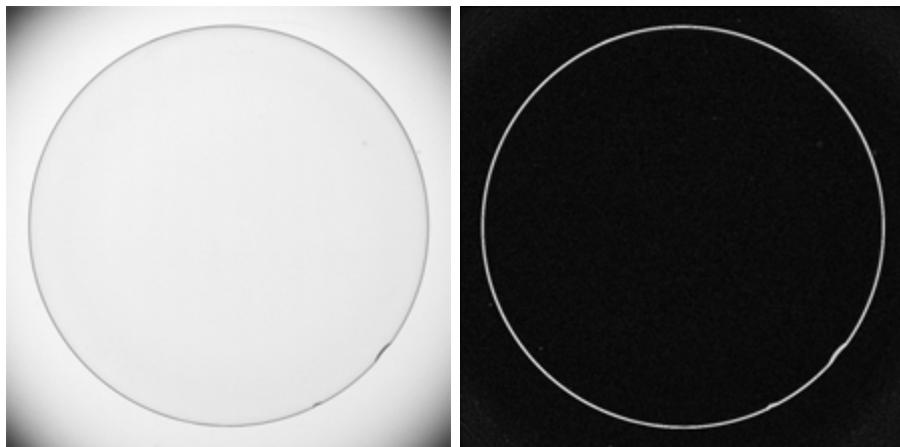
The difference between the third and first rows of the 3×3 image region approximates the derivative in the x -direction, and the difference between the third and first columns approximates the derivative in the y -direction. The masks shown in Figs. 3.44(d) and (e), called the *Sobel operators*, can be used to implement Eq. (3.7-18) via the mechanics given in Eq. (3.5-1). The idea behind using a weight value of 2 is to achieve some smoothing by giving more importance to the center point (we discuss this in more detail in Chapter 10). Note that the coefficients in all the masks shown in Fig. 3.44 sum to 0, indicating that they would give a response of 0 in an area of constant gray level, as expected of a derivative operator.

EXAMPLE 3.14:

Use of the gradient for edge enhancement.

■ The gradient is used frequently in industrial inspection, either to aid humans in the detection of defects or, what is more common, as a preprocessing step in automated inspection. We will have more to say about this in Chapters 10 and 11. However, it will be instructive at this point to consider a simple example to show how the gradient can be used to enhance defects and eliminate slowly changing background features. In this particular example, the enhancement is used as a preprocessing step for automated inspection, rather than for human analysis.

Figure 3.45(a) shows an optical image of a contact lens, illuminated by a lighting arrangement designed to highlight imperfections, such as the two edge



a b

FIGURE 3.45
Optical image of contact lens (note defects on the boundary at 4 and 5 o'clock).
(b) Sobel gradient.
(Original image courtesy of Mr. Pete Sites, Perceptics Corporation.)

defects in the lens boundary seen at 4 and 5 o'clock. Figure 3.45(b) shows the gradient obtained using Eq. (3.7-14) with the two Sobel masks in Figs. 3.44(d) and (e). The edge defects also are quite visible in this image, but with the added advantage that constant or slowly varying shades of gray have been eliminated, thus simplifying considerably the computational task required for automated inspection. Note also that the gradient process highlighted small specs that are not readily visible in the gray-scale image (specs like these can be foreign matter, air pockets in a supporting solution, or minuscule imperfections in the lens). The ability to enhance small discontinuities in an otherwise flat gray field is another important feature of the gradient. ■

3.8 Combining Spatial Enhancement Methods

With a few exceptions, like combining blurring with thresholding in Section 3.6.1, we have focused attention thus far on individual enhancement approaches. Frequently, a given enhancement task will require application of several complementary enhancement techniques in order to achieve an acceptable result. In this section we illustrate by means of an example how to combine several of the approaches developed in this chapter to address a difficult enhancement task.

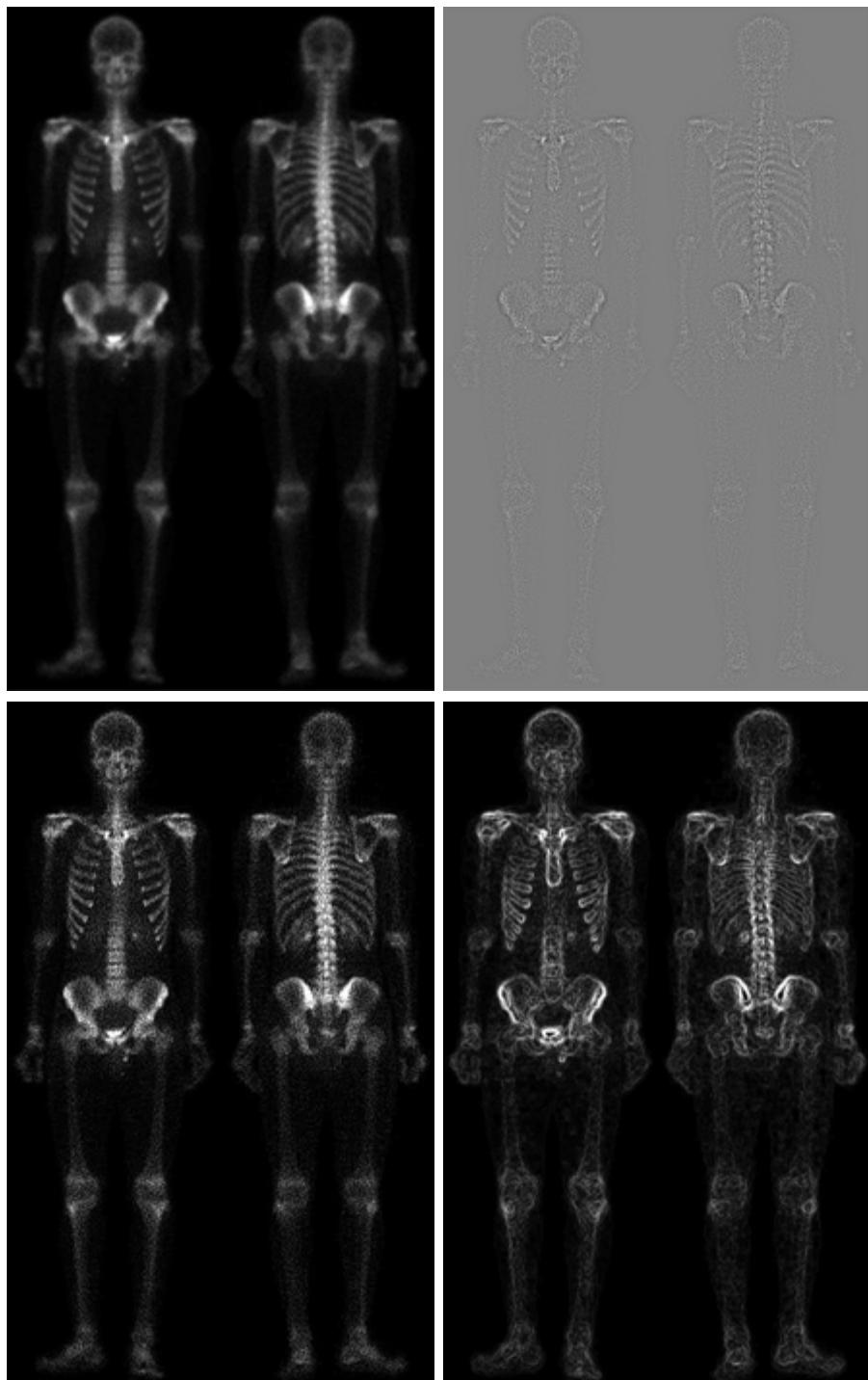
The image shown in Fig. 3.46(a) is a nuclear whole body bone scan, used to detect diseases such as bone infection and tumors. Our objective is to enhance this image by sharpening it and by bringing out more of the skeletal detail. The narrow dynamic range of the gray levels and high noise content make this image difficult to enhance. The strategy we will follow is to utilize the Laplacian to highlight fine detail, and the gradient to enhance prominent edges. For reasons that will be explained shortly, a smoothed version of the gradient image will be used to mask the Laplacian image (see Section 3.4 regarding masking). Finally, we will attempt to increase the dynamic range of the gray levels by using a gray-level transformation.

Figure 3.46 (b) shows the Laplacian of the original image, obtained using the mask in Fig. 3.39(d). This image was scaled (for display only) using the same technique as in Fig. 3.40. We can obtain a sharpened image at this point

a b
c d

FIGURE 3.46

- (a) Image of whole body bone scan.
(b) Laplacian of (a). (c) Sharpened image obtained by adding (a) and (b). (d) Sobel of (a).



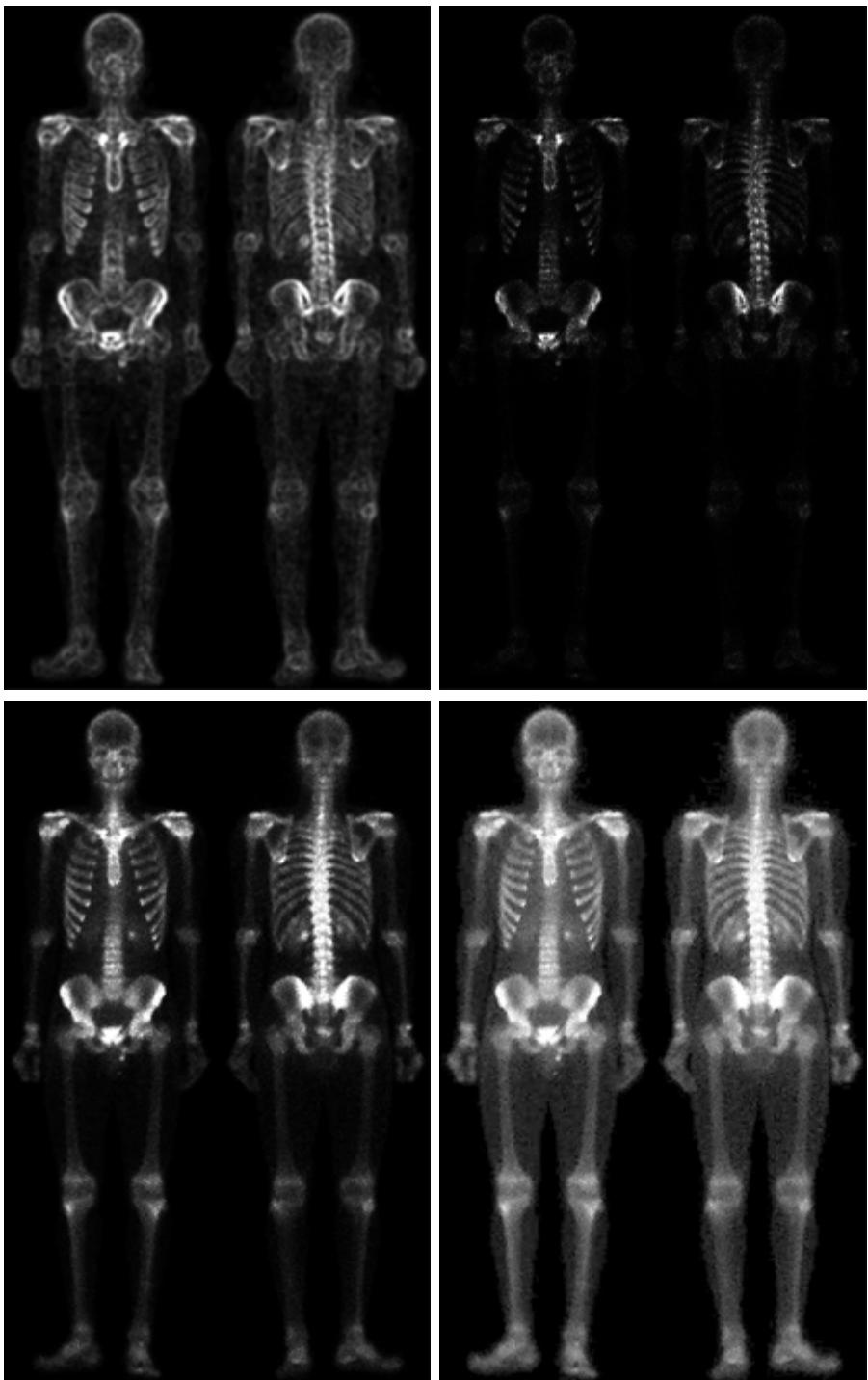


FIGURE 3.46
(Continued)
(e) Sobel image smoothed with a 5×5 averaging filter.
(f) Mask image formed by the product of (c) and (e).
(g) Sharpened image obtained by the sum of (a) and (f).
(h) Final result obtained by applying a power-law transformation to (g). Compare (g) and (h) with (a).
(Original image courtesy of G.E. Medical Systems.)

simply by adding Figs. 3.46(a) and (b), which are an implementation of the second line in Eq. (3.7-5) (we used a mask with a positive center coefficient). Just by looking at the noise level in (b), we would expect a rather noisy sharpened image if we added Figs. 3.46(a) and (b), a fact that is confirmed by the result shown in Fig. 3.46(c). One way that comes immediately to mind to reduce the noise is to use a median filter. However, median filtering is a non-linear process capable of removing image features. This is unacceptable in medical image processing.

An alternate approach is to use a mask formed from a smoothed version of the gradient of the original image. The motivation behind this is straightforward and is based on the properties of first- and second-order derivatives explained in Section 3.7.1. The Laplacian, being a second-order derivative operator, has the definite advantage that it is superior in enhancing fine detail. However, this causes it to produce noisier results than the gradient. This noise is most objectionable in smooth areas, where it tends to be more visible. The gradient has a stronger response in areas of significant gray-level transitions (gray-level ramps and steps) than does the Laplacian. The response of the gradient to noise and fine detail is lower than the Laplacian's and can be lowered further by smoothing the gradient with an averaging filter. The idea, then, is to smooth the gradient and multiply it by the Laplacian image. In this context, we may view the smoothed gradient as a mask image. The product will preserve details in the strong areas while reducing noise in the relatively flat areas. This process can be viewed roughly as combining the best features of the Laplacian and the gradient. The result is added to the original to obtain a final sharpened image, and could even be used in boost filtering.

Figure 3.46(d) shows the Sobel gradient of the original image, computed using Eq. (3.7-14). Components G_x and G_y were obtained using the masks in Figs. 3.44(d) and (e), respectively. As expected from our discussion in Section 3.7.1, edges are much more dominant in this image than in the Laplacian image. The smoothed gradient image shown in Fig. 3.46(e) was obtained by using an averaging filter of size 5×5 . The two gradient images were scaled for display in the same manner as the two Laplacian images. Because the smallest possible value of a gradient image is 0, the background is black in the scaled gradient images, rather than gray as in the scaled Laplacian. The fact that Figs. 3.46(d) and (e) are much brighter than Fig. 3.46(b) is again evidence that the gradient of an image with significant edge content has values that are higher in general than in a Laplacian image.

The product of the Laplacian and smoothed-gradient image is shown in Fig. 3.46(f). Note the dominance of the strong edges and the relative lack of visible noise, which is the key objective behind masking the Laplacian with a smoothed gradient image. Adding the product image to the original resulted in the sharpened image shown in Fig. 3.46(g). The significant increase in sharpness of detail in this image over the original is evident in most parts of the image, including the ribs, spinal chord, pelvis, and skull. This type of improvement would not have been possible by using the Laplacian or gradient alone.

The sharpening procedure just discussed does not affect in an appreciable way the dynamic range of the gray levels in an image. Thus, the final step in our

enhancement task is to increase the dynamic range of the sharpened image. As we discussed in some detail in Sections 3.2 and 3.3, there are a number of gray-level transformation functions that can accomplish this objective. We do know from the results in Section 3.3.2 that histogram equalization is not likely to work well on images that have dark gray-level distributions like our images have here. Histogram specification could be a solution, but the dark characteristics of the images with which we are dealing lend themselves much better to a power-law transformation. Since we wish to spread the gray levels, the value of γ in Eq. (3.2-3) has to be less than 1. After a few trials with this equation we arrived at the result shown in Fig. 3.46(h), obtained with $\gamma = 0.5$ and $c = 1$. Comparing this image with Fig. 3.46(g), we see that significant new detail is visible in Fig. 3.46(h). The areas around the wrists, hands, ankles, and feet are good examples of this. The skeletal bone structure also is much more pronounced, including the arm and leg bones. Note also the faint definition of the outline of the body, and of body tissue. Bringing out detail of this nature by expanding the dynamic range of the gray levels also enhanced noise, but Fig. 3.46(h) represents a significant visual improvement over the original image.

The approach just discussed is representative of the types of processes that can be linked in order to achieve results that are not possible with a single technique. The way in which the results are used depends on the application. The final user of the type of images shown in this section is likely to be a radiologist. For a number of reasons that are beyond the scope of our discussion, physicians are unlikely to rely on enhanced results to arrive at a diagnosis. However, enhanced images are quite useful in highlighting details that can serve as clues for further analysis in the original image or sequence of images. In other areas, the enhanced result may indeed be the final product. Examples are found in the printing industry, in image-based product inspection, in forensics, in microscopy, in surveillance, and in a host of other areas where the principal objective of enhancement is to obtain an image with a higher content of visual detail.

Summary

The material presented in this chapter is representative of spatial domain techniques commonly used in practice for image enhancement. This area of image processing is a dynamic field, and new techniques and applications are reported routinely in professional literature and in new product announcements. For this reason, the topics included in this chapter were selected for their value as fundamental material that would serve as a foundation for understanding the state of the art in enhancement techniques, as well as for further study in this field. In addition to enhancement, this chapter served the purpose of introducing a number of concepts, such as filtering with spatial masks, that will be used in numerous occasions throughout the remainder of the book. In the following chapter, we deal with enhancement from a complementary viewpoint in the frequency domain. Between these two chapters, the reader will have developed a solid foundation for the terminology and some of the most fundamental tools used in image processing. The fact that these tools were introduced in the context of image enhancement is likely to aid in the understanding of how they operate on digital images.

References and Further Reading

The material in Section 3.1 is from Gonzalez [1986]. Additional reading for the material in Section 3.2 may be found in Schowengerdt [1983], Poyton [1996], and Russ [1999]. See also the paper by Tsuji et al. [1998] regarding the optimization of image displays. Early references on histogram processing are Hummel [1974], Gonzalez and Fitts [1977], and Woods and Gonzalez [1981]. Stark [2000] gives some interesting generalizations of histogram equalization for adaptive contrast enhancement. Other approaches for contrast enhancement are exemplified by Centeno and Haertel [1997] and Cheng and Xu [2000]. For enhancement based on an ideal image model, see Highnam and Brady [1997]. For extensions of the local histogram equalization method, see Caselles et al. [1999], and Zhu et al. [1999]. See Narendra and Fitch [1981] on the use and implementation of local statistics for image enhancement. Kim et al. [1997] present an interesting approach combining the gradient with local statistics for image enhancement.

Image subtraction (Section 3.4.1) is a generic image processing tool widely used for change detection. As noted in that section, one of the principal applications of digital image subtraction is in mask mode radiography, where patient motion is a problem because motion smears the results. The problem of motion during image subtraction has received significant attention over the years, as exemplified in the survey article by Meijering et al. [1999]. The method of noise reduction by image averaging (Section 3.4.2) was first proposed by Kohler and Howell [1963]. See Peebles [1993] regarding the expected value of the mean and variance of a sum of random variables.

For additional reading on linear spatial filters and their implementation, see Ummbaugh [1998], Jain [1989], and Rosenfeld and Kak [1982]. Rank-order filters are discussed in these references as well. Wilburn [1998] discusses generalizations of rank-order filters. The book by Pitas and Venetsanopoulos [1990] also deals with median and other nonlinear spatial filters. A special issue of *IEEE Transactions in Image Processing* [1996] is dedicated to the topic of nonlinear image processing. The material on high-boost filtering is from Schowengerdt [1983]. We will encounter again many of the spatial filters introduced in this chapter in discussions dealing with image restoration (Chapter 5) and edge detection (Chapter 10).

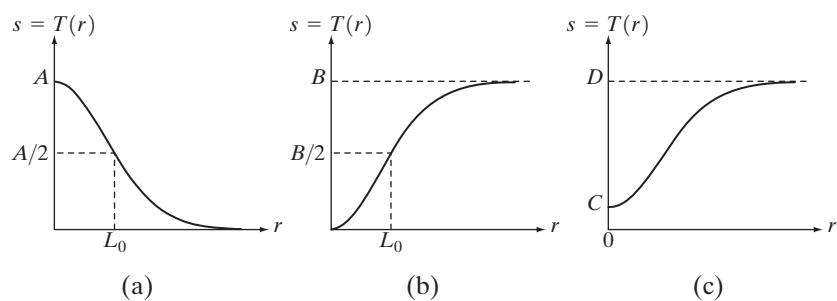
Problems



See inside front cover

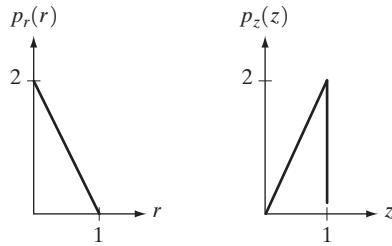
Detailed solutions to the problems marked with a star can be found in the book web site. The site also contains suggested projects based on the material in this chapter.

3.1 Exponentials of the form $e^{-\alpha r^2}$, with α a positive constant, are useful for constructing smooth gray-level transformation functions. Start with this basic function and construct transformation functions having the general shapes shown in the following figures. The constants shown are *input* parameters, and your proposed transformations must include them in their specification. (For simplicity in your answers, L_0 is not a required parameter in the third curve.)



- 3.2 ★ (a)** Give a continuous function for implementing the contrast stretching transformation shown in Fig. 3.2(a). In addition to m , your function must include a parameter, E , for controlling the slope of the function as it transitions from low to high gray-level values. Your function should be normalized so that its minimum and maximum values are 0 and 1, respectively.
- (b)** Sketch a family of transformations as a function of parameter E , for a fixed value $m = L/2$, where L is the number of gray levels in the image.
- (c)** What is the smallest value of s that will make your function *effectively* perform as the function in Fig. 3.2(b)? In other words, your function does not have to be identical to Fig. 3.2(b). It just has to yield the same result of producing a binary image. Assume that you are working with 8-bit images, and let $m = 128$. Also, let C be the smallest positive number representable in the computer you are using.
- 3.3** Propose a set of gray-level-slicing transformations capable of producing all the individual bit planes of an 8-bit monochrome image. (For example, a transformation function with the property $T(r) = 0$ for r in the range $[0, 127]$, and $T(r) = 255$ for r in the range $[128, 255]$ produces an image of the 7th bit plane in an 8-bit image.)
- 3.4 ★ (a)** What effect would setting to zero the lower-order bit planes have on the histogram of an image in general?
- (b)** What would be the effect on the histogram if we set to zero the higher-order bit planes instead?
- ★3.5** Explain why the discrete histogram equalization technique does not, in general, yield a flat histogram.
- 3.6** Suppose that a digital image is subjected to histogram equalization. Show that a second pass of histogram equalization will produce exactly the same result as the first pass.
- 3.7** In some applications it is useful to model the histogram of input images as Gaussian probability density functions of the form
- $$p_r(r) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-m)^2}{2\sigma^2}}$$
- where m and σ are the mean and standard deviation of the Gaussian PDF. The approach is to let m and σ be measures of average gray level and contrast of a given image. What is the transformation function you would use for histogram equalization?
- ★3.8** Assuming continuous values, show by example that it is possible to have a case in which the transformation function given in Eq. (3.3-4) satisfies Conditions (a) and (b) in Section 3.3.1, but its inverse may fail to be single valued.
- 3.9 (a)** Show that the discrete transformation function given in Eq. (3.3-8) for histogram equalization satisfies conditions (a) and (b) in Section 3.3.1.
- (b)** Show by example that this does not hold in general for the inverse discrete transformation function given in Eq. (3.3-9).
- ★ (c)** Show that the inverse discrete transformation in Eq. (3.3-9) satisfies Conditions (a) and (b) in Section 3.3.1 if none of the gray levels $r_k, k = 0, 1, \dots, L - 1$, are missing.

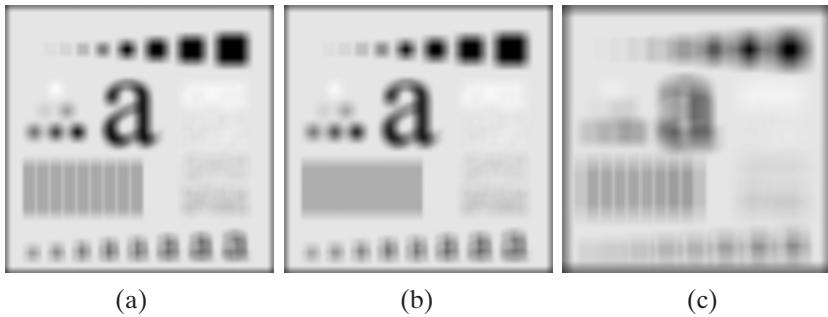
- 3.10** An image has the gray level PDF $p_r(r)$ shown in the following diagram. It is desired to transform the gray levels of this image so that they will have the specified $p_z(z)$ shown. Assume continuous quantities and find the transformation (in terms of r and z) that will accomplish this.



- ★3.11** Propose a method for updating the local histogram for use in the local enhancement technique discussed in Section 3.3.3.
- 3.12** Two images, $f(x, y)$ and $g(x, y)$, have histograms h_f and h_g . Give the conditions under which you can determine the histograms of
- (a) $f(x, y) + g(x, y)$
 - (b) $f(x, y) - g(x, y)$
 - (c) $f(x, y) \times g(x, y)$
 - (d) $f(x, y) \div g(x, y)$
- in terms of h_f and h_g . Explain how to obtain the histogram in each case.
- 3.13** Consider two 8-bit images whose gray levels span the full range from 0 to 255.
- (a) Discuss the limiting effect of repeatedly subtracting image (b) from image (a).
 - (b) Would reversing the order of the images yield a different result?
- ★3.14** Image subtraction is used often in industrial applications for detecting missing components in product assembly. The approach is to store a “golden” image that corresponds to a correct assembly; this image is then subtracted from incoming images of the same product. Ideally, the differences would be zero if the new products are assembled correctly. Difference images for products with missing components would be nonzero in the area where they differ from the golden image. What conditions do you think have to be met in practice for this method to work?
- 3.15** Prove the validity of Eqs. (3.4-4) and (3.4-5).
- 3.16** In an industrial application, X-ray imaging is to be used to inspect the inside of certain composite castings. The objective is to look for voids in the castings, which typically appear as small blobs in the image. However, due to properties of the casting material and X-ray energy used, high noise content often makes inspection difficult, so the decision is made to use image averaging to reduce the noise and thus improve visible contrast. In computing the average, it is important to keep the number of images as small as possible to reduce the time the parts have to remain stationary during imaging. After numerous experiments, it is concluded that decreasing the noise variance by a factor of 10 is sufficient. If the imaging device can produce 30 frames/s, how long would the castings have to remain stationary during imaging to achieve the desired decrease in variance? Assume that the noise is uncorrelated and has zero mean.

- 3.17** The implementation of linear spatial filters requires moving the center of a mask throughout an image and, at each location, computing the sum of products of the mask coefficients with the corresponding pixels at that location (see Section 3.5). In the case of lowpass filtering, all coefficients are 1, allowing use of a so-called *box-filter* or *moving-average* algorithm, which consists of updating only the part of the computation that changes from one location to the next.
- ★ (a) Formulate such an algorithm for an $n \times n$ filter, showing the nature of the computations involved and the scanning sequence used for moving the mask around the image.
- (b) The ratio of the number of computations performed by a brute-force implementation to the number of computations performed by the box-filter algorithm is called the *computational advantage*. Obtain the computational advantage in this case and plot it as a function of n for $n > 1$. The $1/n^2$ scaling factor is common to both approaches, so you need not consider it in obtaining the computational advantage. Assume that the image has an outer border of zeros that is thick enough to allow you to ignore border effects in your analysis.
- 3.18** Discuss the limiting effect of repeatedly applying a 3×3 lowpass spatial filter to a digital image. You may ignore border effects.
- 3.19** ★ (a) It was stated in Section 3.6.2 that isolated clusters of dark or light (with respect to the background) pixels whose area is less than one-half the area of a median filter are eliminated (forced to the median value of the neighbors) by the filter. Assume a filter of size $n \times n$, with n odd, and explain why this is so.
- (b) Consider an image having various sets of pixel clusters. Assume that all points in a cluster are lighter or darker than the background (but not both simultaneously in the same cluster), and that the area of each cluster is less than or equal to $n^2/2$. In terms of n , under what condition would one or more of these clusters cease to be isolated in the sense described in part (a)?
- ★ **3.20** (a) Develop a procedure for computing the median of an $n \times n$ neighborhood.
- (b) Propose a technique for updating the median as the center of the neighborhood is moved from pixel to pixel.
- 3.21** (a) In a character recognition application, text pages are reduced to binary form using a thresholding transformation function of the form shown in Fig. 3.2(b). This is followed by a procedure that thins the characters until they become strings of binary 1's on a background of 0's. Due to noise, the binarization and thinning processes result in broken strings of characters with gaps ranging from 1 to 3 pixels. One way to "repair" the gaps is to run an averaging mask over the binary image to blur it, and thus create bridges of nonzero pixels between gaps. Give the (odd) size of the smallest averaging mask capable of performing this task.
- (b) After bridging the gaps, it is desired to threshold the image in order to convert it back to binary form. For your answer in (a), what is the minimum value of the threshold required to accomplish this, without causing the segments to break up again?
- ★ **3.22** The three images shown were blurred using square averaging masks of sizes $n = 23, 25$, and 45 , respectively. The vertical bars on the left lower part of (a) and (c) are blurred, but a clear separation exists between them. However, the bars

have merged in image (b), in spite of the fact that the mask that produced this image is significantly smaller than the mask that produced image (c). Explain this.



- 3.23** Consider an application such as the one shown in Fig. 3.36, in which it is desired to eliminate objects smaller than those enclosed in a square of size $q \times q$ pixels. Suppose that we want to reduce the average gray level of those objects to one-tenth of their original average gray level. In this way, those objects will be closer to the gray level of the background and they can then be eliminated by thresholding. Give the (odd) size of the smallest averaging mask that will accomplish the desired reduction in average gray level in only one pass of the mask over the image.
- 3.24** In a given application an averaging mask is applied to input images to reduce noise, and then a Laplacian mask is applied to enhance small details. Would the result be the same if the order of these operations were reversed?
- ★ 3.25** Show that the Laplacian operation defined in Eq. (3.7-1) is isotropic (invariant to rotation). You will need the following equations relating coordinates after axis rotation by an angle θ :

$$\begin{aligned} x &= x' \cos \theta - y' \sin \theta \\ y &= x' \sin \theta + y' \cos \theta \end{aligned}$$

where (x, y) are the unrotated and (x', y') are the rotated coordinates.

- 3.26** Give a 3×3 mask for performing unsharp masking in a single pass through an image.
- ★ 3.27** Show that subtracting the Laplacian from an image is proportional to unsharp masking. Use the definition for the Laplacian given in Eq. (3.7-4).
- 3.28** **(a)** Show that the magnitude of the gradient given in Eq. (3.7-13) is an isotropic operation. (See Problem 3.25.)
- (b)** Show that the isotropic property is lost in general if the gradient is computed using Eq. (3.7-14).
- 3.29** A CCD TV camera is used to perform a long-term study by observing the same area 24 hours a day, for 30 days. Digital images are captured and transmitted to a central location every 5 minutes. The illumination of the scene changes from natural daylight to artificial lighting. At no time is the scene without illumination, so it is always possible to obtain an image. Because the range of illumination is such that it is always in the linear operating range of the camera, it is decided not to employ any compensating mechanisms on the camera itself. Rather, it is decided to use digital techniques to postprocess, and thus normalize, the images to the equivalent of constant illumination. Propose a method to do this. You are at liberty to use any method you wish, but state clearly all the assumptions you made in arriving at your design.

Digital Image Processing

Second Edition

Instructor's Manual

Rafael C. Gonzalez
Richard E. Woods

Digital Image Processing

Second Edition

Instructor's Manual

(Evaluation copy—Contains only representative, partial problem solutions)

Rafael C. Gonzalez
Richard E. Woods

Prentice Hall

Upper Saddle River, NJ 07458

www.prenhall.com/gonzalezwoods

or

www.imageprocessingbook.com

Revision history

10 9 8 7 6 5 4 3 2 1

Copyright ©1992-2002 by Rafael C. Gonzalez and Richard E. Woods

Preface

This manual contains detailed solutions to all problems in *Digital Image Processing*, 2nd Edition. We also include a suggested set of guidelines for using the book, and discuss the use of computer projects designed to promote a deeper understanding of the subject matter. The notation used throughout this manual corresponds to the notation used in the text.

The decision of what material to cover in a course rests with the instructor, and it depends on the purpose of the course and the background of the students. We have found that the course outlines suggested here can be covered comfortably in the time frames indicated when the course is being taught in an electrical engineering or computer science curriculum. In each case, no prior exposure to image processing is assumed. We give suggested guidelines for one-semester courses at the senior and first-year graduate levels. It is possible to cover most of the book in a two-semester graduate sequence.

The book was completely revised in this edition, with the purpose not only of updating the material, but just as important, making the book a better teaching aid. To this end, the instructor will find the new organization to be much more flexible and better illustrated. Although the book is self contained, we recommend use of the companion web site, where the student will find detailed solutions to the problems marked with a star in the text, review material, suggested projects, and images from the book. One of the principal reasons for creating the web site was to free the instructor from having to prepare materials and handouts beyond what is required to teach from the book.

Computer projects such as those described in the web site are an important part of a course on image processing. These projects give the student hands-on experience with algorithm implementation and reinforce the material covered in the classroom. The projects suggested at the web site can be implemented on almost any reasonably-equipped multi-user or personal computer having a hard copy output device.

1 Introduction

The purpose of this chapter is to present suggested guidelines for teaching material from this book at the senior and first-year graduate level. We also discuss use of the book web site. Although the book is totally self-contained, the web site offers, among other things, complementary review material and computer projects that can be assigned in conjunction with classroom work. Detailed solutions to all problems in the book also are included in the remaining chapters of this manual.

Teaching Features of the Book

Undergraduate programs that offer digital image processing typically limit coverage to one semester. Graduate programs vary, and can include one or two semesters of the material. In the following discussion we give general guidelines for a one-semester senior course, a one-semester graduate course, and a full-year course of study covering two semesters. We assume a 15-week program per semester with three lectures per week. In order to provide flexibility for exams and review sessions, the guidelines discussed in the following sections are based on forty, 50-minute lectures per semester. The background assumed on the part of the student is senior-level preparation in mathematical analysis, matrix theory, probability, and computer programming.

The suggested teaching guidelines are presented in terms of general objectives, and not as time schedules. There is so much variety in the way image processing material is taught that it makes little sense to attempt a breakdown of the material by class period. In particular, the organization of the present edition of the book is such that it makes it much easier than before to adopt significantly different teaching strategies, depending on course objectives and student background. For example, it is possible with the new organization to offer a course that emphasizes spatial techniques and covers little or no transform material. This is not something we recommend, but it is an option that often is attractive in programs that place little emphasis on the signal processing aspects of the field and prefer to focus more on the implementation of spatial techniques.

2 Chapter 1 Introduction

The companion web site

www.prenhall.com/gonzalezwoods

or

www.imageprocessingbook.com

is a valuable teaching aid, in the sense that it includes material that previously was covered in class. In particular, the review material on probability, matrices, vectors, and linear systems, was prepared using the same notation as in the book, and is focused on areas that are directly relevant to discussions in the text. This allows the instructor to assign the material as independent reading, and spend no more than one total lecture period reviewing those subjects. Another major feature is the set of solutions to problems marked with a star in the book. These solutions are quite detailed, and were prepared with the idea of using them as teaching support. The on-line availability of projects and digital images frees the instructor from having to prepare experiments, data, and handouts for students. The fact that most of the images in the book are available for downloading further enhances the value of the web site as a teaching resource.

One Semester Senior Course

A basic strategy in teaching a senior course is to focus on aspects of image processing in which both the inputs and outputs of those processes are images. In the scope of a senior course, this usually means the material contained in Chapters 1 through 6. Depending on instructor preferences, wavelets (Chapter 7) usually are beyond the scope of coverage in a typical senior curriculum). However, we recommend covering at least some material on image compression (Chapter 8) as outlined below.

We have found in more than two decades of teaching this material to seniors in electrical engineering, computer science, and other technical disciplines, that one of the keys to success is to spend at least one lecture on motivation and the equivalent of one lecture on review of background material, as the need arises. The motivational material is provided in the numerous application areas discussed in Chapter 1. This chapter was totally rewritten with this objective in mind. Some of this material can be covered in class and the rest assigned as independent reading. Background review should cover probability theory (of one random variable) before histogram processing (Section 3.3). A brief review of vectors and matrices may be required later, depending on the material covered. The review material included in the book web site was designed for just this purpose.

Chapter 2 should be covered in its entirety. Some of the material (such as parts of Sections 2.1 and 2.3) can be assigned as independent reading, but a detailed explanation of Sections 2.4 through 2.6 is time well spent.

Chapter 3 serves two principal purposes. It covers image enhancement (a topic of significant appeal to the beginning student) and it introduces a host of basic spatial processing tools used throughout the book. For a senior course, we recommend coverage of Sections 3.2.1 through 3.2.2; Section 3.3.1; Section 3.4; Section 3.5; Section 3.6; Section 3.7.1, 3.7.2 (through Example 3.11), and 3.7.3. Section 3.8 can be assigned as independent reading, depending on time.

Chapter 4 also discusses enhancement, but from a frequency-domain point of view. The instructor has significant flexibility here. As mentioned earlier, it is possible to skip the chapter altogether, but this will typically preclude meaningful coverage of other areas based on the Fourier transform (such as filtering and restoration). The key in covering the frequency domain is to get to the convolution theorem and thus develop a tie between the frequency and spatial domains. All this material is presented in very readable form in Section 4.2. “Light” coverage of frequency-domain concepts can be based on discussing all the material through this section and then selecting a few simple filtering examples (say, low- and highpass filtering using Butterworth filters, as discussed in Sections 4.3.2 and 4.4.2). At the discretion of the instructor, additional material can include full coverage of Sections 4.3 and 4.4. It is seldom possible to go beyond this point in a senior course.

Chapter 5 can be covered as a continuation of Chapter 4. Section 5.1 makes this an easy approach. Then, it is possible give the student a “flavor” of what restoration is (and still keep the discussion brief) by covering only Gaussian and impulse noise in Section 5.2.1, and a couple of spatial filters in Section 5.3. This latter section is a frequent source of confusion to the student who, based on discussions earlier in the chapter, is expecting to see a more objective approach. It is worthwhile to emphasize at this point that spatial enhancement and restoration are the same thing when it comes to noise reduction by spatial filtering. A good way to keep it brief and conclude coverage of restoration is to jump at this point to inverse filtering (which follows directly from the model in Section 5.1) and show the problems with this approach. Then, with a brief explanation regarding the fact that much of restoration centers around the instabilities inherent in inverse filtering, it is possible to introduce the “interactive” form of the Wiener filter in Eq. (5.8-3) and conclude the chapter with Examples 5.12 and 5.13.

Chapter 6 on color image processing is a new feature of the book. Coverage of this

4 Chapter 1 Introduction

chapter also can be brief at the senior level by focusing on enough material to give the student a foundation on the physics of color (Section 6.1), two basic color models (RGB and CMY/CMYK), and then concluding with a brief coverage of pseudocolor processing (Section 6.3).

We typically conclude a senior course by covering some of the basic aspects of image compression (Chapter 8). Interest on this topic has increased significantly as a result of the heavy use of images and graphics over the Internet, and students usually are easily motivated by the topic. Minimum coverage of this material includes Sections 8.1.1 and 8.1.2, Section 8.2, and Section 8.4.1. In this limited scope, it is worthwhile spending one-half of a lecture period filling in any gaps that may arise by skipping earlier parts of the chapter.

One Semester Graduate Course (No Background in DIP)

The main difference between a senior and a first-year graduate course in which neither group has formal background in image processing is mostly in the scope of material covered, in the sense that we simply go faster in a graduate course, and feel much freer in assigning independent reading. In addition to the material discussed in the previous section, we add the following material in a graduate course.

Coverage of histogram matching (Section 3.3.2) is added. Sections 4.3, 4.4, and 4.5 are covered in full. Section 4.6 is touched upon briefly regarding the fact that implementation of discrete Fourier transform techniques requires non-intuitive concepts such as function padding. The separability of the Fourier transform should be covered, and mention of the advantages of the FFT should be made. In Chapter 5 we add Sections 5.5 through 5.8. In Chapter 6 we add the HSI model (Section 6.3.2), Section 6.4, and Section 6.6. A nice introduction to wavelets (Chapter 7) can be achieved by a combination of classroom discussions and independent reading. The minimum number of sections in that chapter are 7.1, 7.2, 7.3, and 7.5, with appropriate (but brief) mention of the existence of fast wavelet transforms. Finally, in Chapter 8 we add coverage of Sections 8.3, 8.4.2, 8.5.1 (through Example 8.16), Section 8.5.2 (through Example 8.20) and Section 8.5.3.

If additional time is available, a natural topic to cover next is morphological image processing (Chapter 9). The material in this chapter begins a transition from methods whose inputs and outputs are images to methods in which the inputs are images, but the outputs are attributes about those images, in the sense defined in Section 1.1. We

recommend coverage of Sections 9.1 through 9.4, and some of the algorithms in Section 9.5.

One Semester Graduate Course (with Background in DIP)

Some programs have an undergraduate course in image processing as a prerequisite to a graduate course on the subject. In this case, it is possible to cover material from the first eleven chapters of the book. Using the undergraduate guidelines described above, we add the following material to form a teaching outline for a one semester graduate course that has that undergraduate material as prerequisite. Given that students have the appropriate background on the subject, independent reading assignments can be used to control the schedule.

Coverage of histogram matching (Section 3.3.2) is added. Sections 4.3, 4.4, 4.5, and 4.6 are added. This strengthens the student's background in frequency-domain concepts. A more extensive coverage of Chapter 5 is possible by adding sections 5.2.3, 5.3.3, 5.4.3, 5.5, 5.6, and 5.8. In Chapter 6 we add full-color image processing (Sections 6.4 through 6.7). Chapters 7 and 8 are covered as in the previous section. As noted in the previous section, Chapter 9 begins a transition from methods whose inputs and outputs are images to methods in which the inputs are images, but the outputs are attributes about those images. As a minimum, we recommend coverage of binary morphology: Sections 9.1 through 9.4, and some of the algorithms in Section 9.5. Mention should be made about possible extensions to gray-scale images, but coverage of this material may not be possible, depending on the schedule. In Chapter 10, we recommend Sections 10.1, 10.2.1 and 10.2.2, 10.3.1 through 10.3.4, 10.4, and 10.5. In Chapter 11 we typically cover Sections 11.1 through 11.4.

Two Semester Graduate Course (No Background in DIP)

A full-year graduate course consists of the material covered in the one semester undergraduate course, the material outlined in the previous section, and Sections 12.1, 12.2, 12.3.1, and 12.3.2.

Projects

One of the most interesting aspects of a course in digital image processing is the pictorial

nature of the subject. It has been our experience that students truly enjoy and benefit from judicious use of computer projects to complement the material covered in class. Since computer projects are in addition to course work and homework assignments, we try to keep the formal project reporting as brief as possible. In order to facilitate grading, we try to achieve uniformity in the way project reports are prepared. A useful report format is as follows:

Page 1: Cover page.

- Project title
- Project number
- Course number
- Student's name
- Date due
- Date handed in
- Abstract (not to exceed 1/2 page)

Page 2: One to two pages (max) of technical discussion.

Page 3 (or 4): Discussion of results. One to two pages (max).

Results: Image results (printed typically on a laser or inkjet printer). All images must contain a number and title referred to in the discussion of results.

Appendix: Program listings, focused on any original code prepared by the student. For brevity, functions and routines provided to the student are referred to by name, but the code is not included.

Layout: The entire report must be on a standard sheet size (e.g., 8.5×11 inches), stapled with three or more staples on the left margin to form a booklet, or bound using clear plastic standard binding products.

Project resources available in the book web site include a sample project, a list of suggested projects from which the instructor can select, book and other images, and MATLAB functions. Instructors who do not wish to use MATLAB will find additional software suggestions in the Support/Software section of the web site.

2 Problem Solutions

Problem 2.1

The diameter, x , of the retinal image corresponding to the dot is obtained from similar triangles, as shown in Fig. P2.1. That is,

$$\frac{(d/2)}{0.2} = \frac{(x/2)}{0.014}$$

which gives $x = 0.07d$. From the discussion in Section 2.1.1, and taking some liberties of interpretation, we can think of the fovea as a square sensor array having on the order of 337,000 elements, which translates into an array of size 580×580 elements. Assuming equal spacing between elements, this gives 580 elements and 579 spaces on a line 1.5 mm long. The size of each element and each space is then $s = [(1.5\text{mm})/1, 159] = 1.3 \times 10^{-6} \text{ m}$. If the size (on the fovea) of the imaged dot is less than the size of a single resolution element, we assume that the dot will be invisible to the eye. In other words, the eye will not detect a dot if its diameter, d , is such that $0.07(d) < 1.3 \times 10^{-6} \text{ m}$, or $d < 18.6 \times 10^{-6} \text{ m}$.

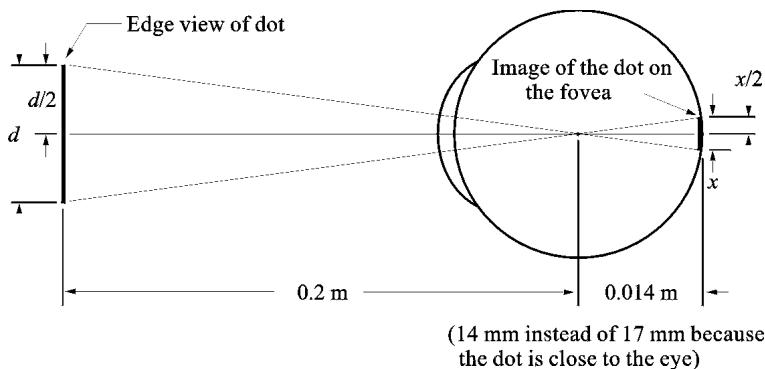


Figure P2.1

Problem 2.3

$$\lambda = c/v = 2.998 \times 10^8 (\text{m/s}) / 60(\text{1/s}) = 4.99 \times 10^6 \text{ m} = 5000 \text{ Km.}$$

Problem 2.6

One possible solution is to equip a monochrome camera with a mechanical device that sequentially places a red, a green, and a blue pass filter in front of the lens. The strongest camera response determines the color. If all three responses are approximately equal, the object is white. A faster system would utilize three different cameras, each equipped with an individual filter. The analysis would be then based on polling the response of each camera. This system would be a little more expensive, but it would be faster and more reliable. Note that both solutions assume that the field of view of the camera(s) is such that it is completely filled by a uniform color [i.e., the camera(s) is(are) focused on a part of the vehicle where only its color is seen. Otherwise further analysis would be required to isolate the region of uniform color, which is all that is of interest in solving this problem].

Problem 2.9

- (a) The total amount of data (including the start and stop bit) in an 8-bit, 1024×1024 image, is $(1024)^2 \times [8 + 2]$ bits. The total time required to transmit this image over a At 56K baud link is $(1024)^2 \times [8 + 2]/56000 = 187.25$ sec or about 3.1 min. (b) At 750K this time goes down to about 14 sec.

Problem 2.11

Let p and q be as shown in Fig. P2.11. Then, (a) S_1 and S_2 are not 4-connected because q is not in the set $N_4(p)$; (b) S_1 and S_2 are 8-connected because q is in the set $N_8(p)$; (c) S_1 and S_2 are m -connected because (i) q is in $N_D(p)$, and (ii) the set $N_4(p) \cap N_4(q)$ is empty.

	S_1					S_2					
0	0	0	0	0	0	0	0	1	1	0	0
1	0	0	1	0	0	0	1	0	0	1	1
1	0	0	1	0	0	1	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0	1	1	1	1

Figure P2.11

Problem 2.12

The solution to this problem consists of defining all possible neighborhood shapes to go from a diagonal segment to a corresponding 4-connected segment, as shown in Fig. P2.12. The algorithm then simply looks for the appropriate match every time a diagonal segment is encountered in the boundary.

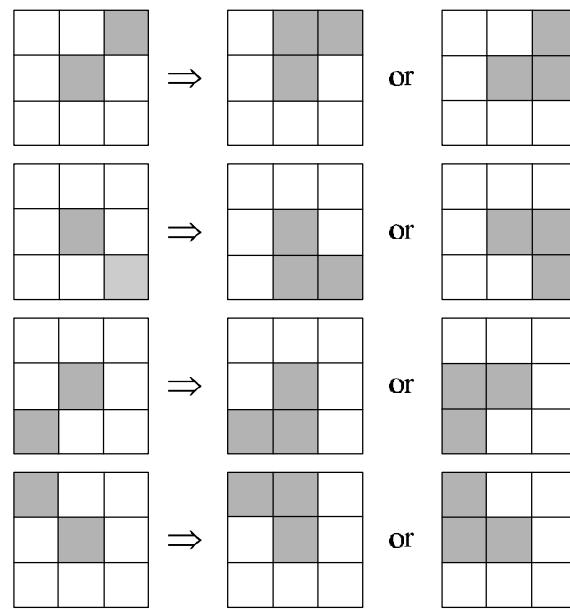


Figure P2.12

Problem 2.15

- (a) When $V = \{0, 1\}$, 4-path does not exist between p and q because it is impossible to

get from p to q by traveling along points that are both 4-adjacent and also have values from V . Figure P2.15(a) shows this condition; it is not possible to get to q . The shortest 8-path is shown in Fig. P2.15(b); its length is 4. In this case the length of shortest m - and 8-paths is the same. Both of these shortest paths are unique in this case. (b) One possibility for the shortest 4-path when $V = \{1, 2\}$ is shown in Fig. P2.15(c); its length is 6. It is easily verified that another 4-path of the same length exists between p and q . One possibility for the shortest 8-path (it is not unique) is shown in Fig. P2.15(d); its length is 4. The length of a shortest m -path similarly is 4.

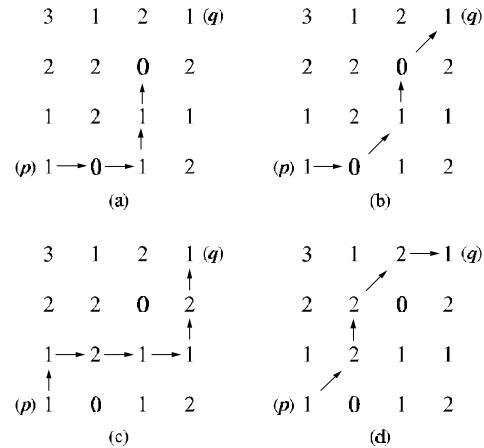


Figure P2.15

Problem 2.16

(a) A shortest 4-path between a point p with coordinates (x, y) and a point q with coordinates (s, t) is shown in Fig. P2.16, where the assumption is that all points along the path are from V . The length of the segments of the path are $|x - s|$ and $|y - t|$, respectively. The total path length is $|x - s| + |y - t|$, which we recognize as the definition of the D_4 distance, as given in Eq. (2.5-16). (Recall that this distance is independent of any paths that may exist between the points.) The D_4 distance obviously is equal to the length of the shortest 4-path when the length of the path is $|x - s| + |y - t|$. This occurs whenever we can get from p to q by following a path whose elements (1) are from V , and (2) are arranged in such a way that we can traverse the path from p to q by making turns in at most two directions (e.g., right and up). (b) The path may or may not be unique, depending on V and the values of the points along the way.

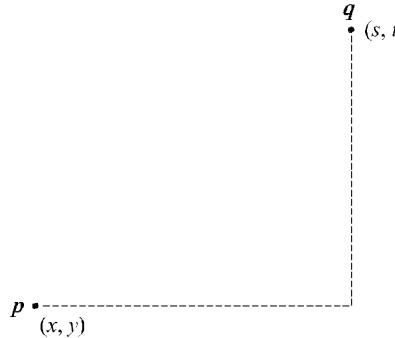


Figure P2.16

Problem 2.18

With reference to Eq. (2.6-1), let H denote the neighborhood sum operator, let S_1 and S_2 denote two different small subimage areas of the same size, and let $S_1 + S_2$ denote the corresponding pixel-by-pixel sum of the elements in S_1 and S_2 , as explained in Section 2.5.4. Note that the size of the neighborhood (i.e., number of pixels) is not changed by this pixel-by-pixel sum. The operator H computes the sum of pixel values in a given neighborhood. Then, $H(aS_1 + bS_2)$ means: (1) multiplying the pixels in each of the subimage areas by the constants shown, (2) adding the pixel-by-pixel values from S_1 and S_2 (which produces a single subimage area), and (3) computing the sum of the values of all the pixels in that single subimage area. Let ap_1 and bp_2 denote two arbitrary (but corresponding) pixels from $aS_1 + bS_2$. Then we can write

$$\begin{aligned}
 H(aS_1 + bS_2) &= \sum_{p_1 \in S_1 \text{ and } p_2 \in S_2} ap_1 + bp_2 \\
 &= \sum_{p_1 \in S_1} ap_1 + \sum_{p_2 \in S_2} bp_2 \\
 &= a \sum_{p_1 \in S_1} p_1 + b \sum_{p_2 \in S_2} p_2 \\
 &= aH(S_1) + bH(S_2)
 \end{aligned}$$

which, according to Eq. (2.6-1), indicates that H is a linear operator.

3 Problem Solutions

Problem 3.2

(a)

$$s = T(r) = \frac{1}{1 + (m/r)^E}.$$

Problem 3.4

- (a) The number of pixels having different gray level values would decrease, thus causing the number of components in the histogram to decrease. Since the number of pixels would not change, this would cause the height some of the remaining histogram peaks to increase in general. Typically, less variability in gray level values will reduce contrast.

Problem 3.5

All that histogram equalization does is remap histogram components on the intensity scale. To obtain a uniform (flat) histogram would require in general that pixel intensities be actually redistributed so that there are L groups of n/L pixels with the same intensity, where L is the number of allowed discrete intensity levels and n is the total number of pixels in the input image. The histogram equalization method has no provisions for this type of (artificial) redistribution process.

Problem 3.8

We are interested in just one example in order to satisfy the statement of the problem. Consider the probability density function shown in Fig. P3.8(a). A plot of the transformation $T(r)$ in Eq. (3.3-4) using this particular density function is shown in Fig. P3.8(b). Because $p_r(r)$ is a probability density function we know from the discussion

in Section 3.3.1 that the transformation $T(r)$ satisfies conditions (a) and (b) stated in that section. However, we see from Fig. P3.8(b) that the inverse transformation from s back to r is not single valued, as there are an infinite number of possible mappings from $s = 1/2$ back to r . It is important to note that the reason the inverse transformation function turned out not to be single valued is the gap in $p_r(r)$ in the interval $[1/4, 3/4]$.

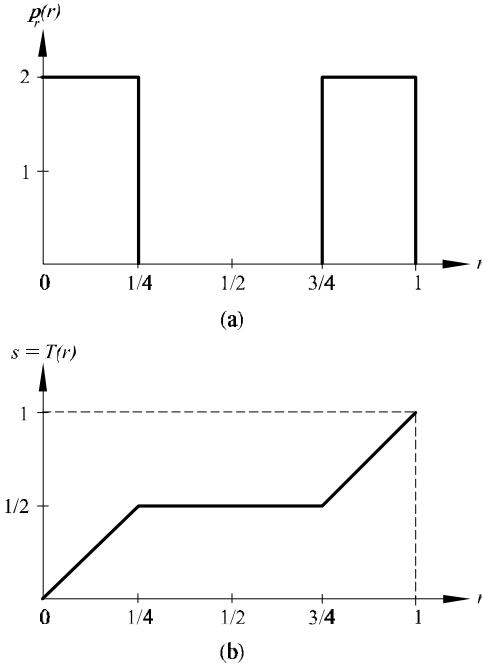


Figure P3.8.

Problem 3.9

- (c) If none of the gray levels r_k , $k = 1, 2, \dots, L - 1$, are 0, then $T(r_k)$ will be strictly monotonic. This implies that the inverse transformation will be of finite slope and this will be single-valued.

Problem 3.11

The value of the histogram component corresponding to the k th intensity level in a neighborhood is

$$p_r(r_k) = \frac{n_k}{n}$$

for $k = 1, 2, \dots, K - 1$, where n_k is the number of pixels having gray level value r_k , n is the total number of pixels in the neighborhood, and K is the total number of possible gray levels. Suppose that the neighborhood is moved one pixel to the right. This deletes the leftmost column and introduces a new column on the right. The updated histogram then becomes

$$p'_r(r_k) = \frac{1}{n} [n_k - n_{L_k} + n_{R_k}]$$

for $k = 0, 1, \dots, K - 1$, where n_{L_k} is the number of occurrences of level r_k on the left column and n_{R_k} is the similar quantity on the right column. The preceding equation can be written also as

$$p'_r(r_k) = p_r(r_k) + \frac{1}{n} [n_{R_k} - n_{L_k}]$$

for $k = 0, 1, \dots, K - 1$. The same concept applies to other modes of neighborhood motion:

$$p'_r(r_k) = p_r(r_k) + \frac{1}{n} [b_k - a_k]$$

for $k = 0, 1, \dots, K - 1$, where a_k is the number of pixels with value r_k in the neighborhood area deleted by the move, and b_k is the corresponding number introduced by the move.

$$\sigma_g^2 = \sigma_f^2 + \frac{1}{K^2} [\sigma_{\eta_1}^2 + \sigma_{\eta_2}^2 + \dots + \sigma_{\eta_K}^2]$$

The first term on the right side is 0 because the elements of f are constants. The various $\sigma_{\eta_i}^2$ are simply samples of the noise, which has variance σ_η^2 . Thus, $\sigma_{\eta_i}^2 = \sigma_\eta^2$ and we have

$$\sigma_g^2 = \frac{K}{K^2} \sigma_\eta^2 = \frac{1}{K} \sigma_\eta^2$$

which proves the validity of Eq. (3.4-5).

Problem 3.14

Let $g(x, y)$ denote the golden image, and let $f(x, y)$ denote any input image acquired during routine operation of the system. Change detection via subtraction is based on computing the simple difference $d(x, y) = g(x, y) - f(x, y)$. The resulting image $d(x, y)$ can be used in two fundamental ways for change detection. One way is to use a pixel-by-pixel analysis. In this case we say that $f(x, y)$ is "close enough" to the golden image if all the pixels in $d(x, y)$ fall within a specified threshold band $[T_{min}, T_{max}]$ where T_{min} is negative and T_{max} is positive. Usually, the same value of threshold is

used for both negative and positive differences, in which case we have a band $[-T, T]$ in which all pixels of $d(x, y)$ must fall in order for $f(x, y)$ to be declared acceptable. The second major approach is simply to sum all the pixels in $|d(x, y)|$ and compare the sum against a threshold S . Note that the absolute value needs to be used to avoid errors cancelling out. This is a much cruder test, so we will concentrate on the first approach.

There are three fundamental factors that need tight control for difference-based inspection to work: (1) proper registration, (2) controlled illumination, and (3) noise levels that are low enough so that difference values are not affected appreciably by variations due to noise. The first condition basically addresses the requirement that comparisons be made between corresponding pixels. Two images can be identical, but if they are displaced with respect to each other, comparing the differences between them makes no sense. Often, special markings are manufactured into the product for mechanical or image-based alignment

Controlled illumination (note that “illumination” is not limited to visible light) obviously is important because changes in illumination can affect dramatically the values in a difference image. One approach often used in conjunction with illumination control is intensity scaling based on actual conditions. For example, the products could have one or more small patches of a tightly controlled color, and the intensity (and perhaps even color) of each pixels in the entire image would be modified based on the actual versus expected intensity and/or color of the patches in the image being processed.

Finally, the noise content of a difference image needs to be low enough so that it does not materially affect comparisons between the golden and input images. Good signal strength goes a long way toward reducing the effects of noise. Another (sometimes complementary) approach is to implement image processing techniques (e.g., image averaging) to reduce noise.

Obviously there are a number of variations of the basic theme just described. For example, additional intelligence in the form of tests that are more sophisticated than pixel-by-pixel threshold comparisons can be implemented. A technique often used in this regard is to subdivide the golden image into different regions and perform different (usually more than one) tests in each of the regions, based on expected region content.

Problem 3.17

- (a) Consider a 3×3 mask first. Since all the coefficients are 1 (we are ignoring the 1/9

scale factor), the net effect of the lowpass filter operation is to add all the gray levels of pixels under the mask. Initially, it takes 8 additions to produce the response of the mask. However, when the mask moves one pixel location to the right, it picks up only one new column. The new response can be computed as

$$R_{\text{new}} = R_{\text{old}} - C_1 + C_3$$

where C_1 is the sum of pixels under the first column of the mask before it was moved, and C_3 is the similar sum in the column it picked up after it moved. This is the basic box-filter or moving-average equation. For a 3×3 mask it takes 2 additions to get C_3 (C_1 was already computed). To this we add one subtraction and one addition to get R_{new} . Thus, a total of 4 arithmetic operations are needed to update the response after one move. This is a recursive procedure for moving from left to right along one row of the image. When we get to the end of a row, we move down one pixel (the nature of the computation is the same) and continue the scan in the opposite direction.

For a mask of size $n \times n$, $(n - 1)$ additions are needed to obtain C_3 , plus the single subtraction and addition needed to obtain R_{new} , which gives a total of $(n + 1)$ arithmetic operations after each move. A brute-force implementation would require $n^2 - 1$ additions after each move.

Problem 3.19

- (a) There are n^2 points in an $n \times n$ median filter mask. Since n is odd, the median value, ζ , is such that there are $(n^2 - 1)/2$ points with values less than or equal to ζ and the same number with values greater than or equal to ζ . However, since the area A (number of points) in the cluster is less than one half n^2 , and A and n are integers, it follows that A is always less than or equal to $(n^2 - 1)/2$. Thus, even in the extreme case when all cluster points are encompassed by the filter mask, there are not enough points in the cluster for any of them to be equal to the value of the median (remember, we are assuming that all cluster points are lighter or darker than the background points). Therefore, if the center point in the mask is a cluster point, it will be set to the median value, which is a background shade, and thus it will be “eliminated” from the cluster. This conclusion obviously applies to the less extreme case when the number of cluster points encompassed by the mask is less than the maximum size of the cluster.

Problem 3.20

(a) Numerically sort the n^2 values. The median is

$$\zeta = [(n^2 + 1)/2]\text{-th largest value.}$$

(b) Once the values have been sorted one time, we simply delete the values in the trailing edge of the neighborhood and insert the values in the leading edge in the appropriate locations in the sorted array.

Problem 3.22

From Fig. 3.35, the vertical bars are 5 pixels wide, 100 pixels high, and their separation is 20 pixels. The phenomenon in question is related to the horizontal separation between bars, so we can simplify the problem by considering a single scan line through the bars in the image. The key to answering this question lies in the fact that the distance (in pixels) between the onset of one bar and the onset of the next one (say, to its right) is 25 pixels. Consider the scan line shown in Fig. P3.22. Also shown is a cross section of a 25×25 mask. The response of the mask is the average of the pixels that it encompasses. We note that when the mask moves one pixel to the right, it loses one pixel of the vertical bar on the left, but it picks up an identical one on the right, so the response doesn't change. In fact, the number of pixels belonging to the vertical bars and contained within the mask does not change, regardless of where the mask is located (as long as it is contained within the bars, and not near the edges of the set of bars). The fact that the number of bar pixels under the mask does not change is due to the peculiar separation between bars and the width of the lines in relation to the 25-pixel width of the mask. This constant response is the reason no white gaps are seen in the image shown in the problem statement. Note that this constant response does not happen with the 23×23 or the 45×45 masks because they are not "synchronized" with the width of the bars and their separation.

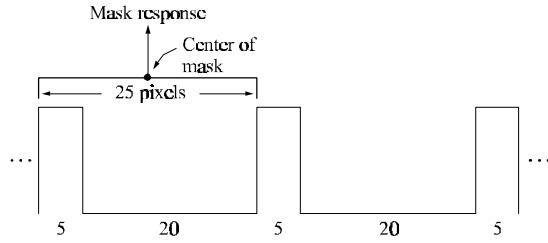


Figure P3.22

Problem 3.25

The Laplacian operator is defined as

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

for the unrotated coordinates and as

$$\nabla^2 f = \frac{\partial^2 f}{\partial x'^2} + \frac{\partial^2 f}{\partial y'^2}.$$

for rotated coordinates. It is given that

$$x = x' \cos \theta - y' \sin \theta \quad \text{and} \quad y = x' \sin \theta + y' \cos \theta$$

where θ is the angle of rotation. We want to show that the right sides of the first two equations are equal. We start with

$$\begin{aligned} \frac{\partial f}{\partial x'} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial x'} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial x'} \\ &= \frac{\partial f}{\partial x} \cos \theta + \frac{\partial f}{\partial y} \sin \theta \end{aligned}$$

Taking the partial derivative of this expression again with respect to x' yields

$$\frac{\partial^2 f}{\partial x'^2} = \frac{\partial^2 f}{\partial x^2} \cos^2 \theta + \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) \sin \theta \cos \theta + \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) \cos \theta \sin \theta + \frac{\partial^2 f}{\partial y^2} \sin^2 \theta$$

Next, we compute

$$\begin{aligned} \frac{\partial f}{\partial y'} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial y'} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial y'} \\ &= -\frac{\partial f}{\partial x} \sin \theta + \frac{\partial f}{\partial y} \cos \theta \end{aligned}$$

Taking the derivative of this expression again with respect to y' gives

$$\frac{\partial^2 f}{\partial y'^2} = \frac{\partial^2 f}{\partial x^2} \sin^2 \theta - \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) \cos \theta \sin \theta - \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) \sin \theta \cos \theta + \frac{\partial^2 f}{\partial y^2} \cos^2 \theta$$

Adding the two expressions for the second derivatives yields

$$\frac{\partial^2 f}{\partial x'^2} + \frac{\partial^2 f}{\partial y'^2} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

which proves that the Laplacian operator is independent of rotation.

Problem 3.27

Consider the following equation:

$$\begin{aligned}
f(x, y) - \nabla^2 f(x, y) &= f(x, y) - [f(x+1, y) + f(x-1, y) + f(x, y+1) \\
&\quad + f(x, y-1) - 4f(x, y)] \\
&= 6f(x, y) - [f(x+1, y) + f(x-1, y) + f(x, y+1) \\
&\quad + f(x, y-1) + f(x, y)] \\
&= 5 \{1.2f(x, y) - \\
&\quad \frac{1}{5} [f(x+1, y) + f(x-1, y) + f(x, y+1) \\
&\quad + f(x, y-1) + f(x, y)]\} \\
&= 5 [1.2f(x, y) - \bar{f}(x, y)]
\end{aligned}$$

where $\bar{f}(x, y)$ denotes the average of $f(x, y)$ in a predefined neighborhood that is centered at (x, y) and includes the center pixel and its four immediate neighbors. Treating the constants in the last line of the above equation as proportionality factors, we may write

$$f(x, y) - \nabla^2 f(x, y) \sim f(x, y) - \bar{f}(x, y).$$

The right side of this equation is recognized as the definition of unsharp masking given in Eq. (3.7-7). Thus, it has been demonstrated that subtracting the Laplacian from an image is proportional to unsharp masking.