The Art of IMAGE PROCESSING with Java

KENNY A. HUNT

The Art of
Image Processing
with Java

The Art of Image Processing with Java

Kenny A. Hunt

A K Peters, Ltd.
Natick, Massachusetts

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

Version Date: 20130905

International Standard Book Number-13: 978-1-4398-6559-0 (eBook - PDF)

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

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

Contents

1 Introduction

1.1 What Is Digital Image Processing?

We must begin our journey by taking issue with the philosophical adage that "a picture is worth a thousand words." It is my belief that a picture cannot begin to convey the depth of human experience and wisdom embedded in the words of Shakespeare, Dostoevsky, Dante, or Moses. A picture cannot convey with due precision the mathematical underpinnings of the discoveries of Galileo or Pascal nor can a picture give expression to the philosophy of Augustine, Plato, or Edwards. Nonetheless, while pictures do not carry the precision of written language, they do contain a wealth of information and have been used throughout the centuries as an important and useful means of communication. An *image* is a picture representing visual information. A *digital image* is an image that can be stored in digital form.

Prior to the advent of computation, images were rendered on papyrus, paper, film, or canvas using ink or paint or photosensitive chemicals. These non-digital images are prone to fading and hence suffer loss of image quality due to exposure to light or temperature extremes. Also, since non-digital images are fixed in some physical medium it is not possible to precisely copy a non-digital image. Throughout the annals of art history, forgers have attempted to copy paintings of well-known masters but usually fail due to their inability to precisely duplicate either a style or an original work. Han van Meegeren is one of the best known art forgers of the 20th century. His technique so closely mimicked the style and colors of the art masters that he was able to deceive even the most expert art critics of his time. His most famous forgery, *The Disciples at Emmaus*, was created in 1936 and was purportedly created by the well-known Dutch artist Johannes Vermeer. His work was finally exposed as fraudulent, however, at least in part by a chemical analysis of the paint, which showed traces of a plastic compound that was not manufactured until the 20th century!

Digital images, however, are pictures that are stored in digital form and that are viewable on some computing system. Since digital images are stored as binary data, the digital image never fades or degrades over time and the only way to destroy a digital image is to delete or corrupt the file itself. In addition, a digital image can be transmitted across the globe in seconds and can be efficiently and precisely copied without any loss of quality.

Digital image processing is a field of study that seeks to analyze, process, or enhance a digital image to achieve some desired outcome. More formally, digital image processing can be defined as the study of techniques for transforming a digital image into another (improved) digital image or for analyzing a digital image to obtain specific information about the image.

From the cradle to the grave we are accustomed to viewing life through *digital* images. A parent's first portrait of their child is often taken before they are even born through the use of sophisticated ultrasound imaging technology. As the child grows, the parents capture developmental milestones using palm-sized digital video cameras. Portraits are sent over email to relatives and friends and short video clips are posted on the family's website. When the child breaks an arm playing soccer, the emergency-room physician orders an x-ray image and transmits it over the Internet to a specialist hundreds of miles away for immediate advice. During his lifetime the child will watch television images that have been digitally transmitted to the dish on top of his house, view weather satellite images on the Internet to determine whether or not to travel, and see images of war where smart bombs find their target by "seeing" the enemy.

Computer graphics is a closely related field but has a different goal than image processing. While the primary goal of computer graphics is the efficient generation of digital images, the input to a graphics system is generally a geometric model specifying the shape, texture, and color of all objects in the virtual scene. Image processing, by contrast, begins with a digital image as input and generates, typically, a digital image as output.

Computer vision, or machine vision, is another increasingly important relative of image processing where an input image is analyzed in order to determine its content. The primary goal of computer vision systems is the inverse of

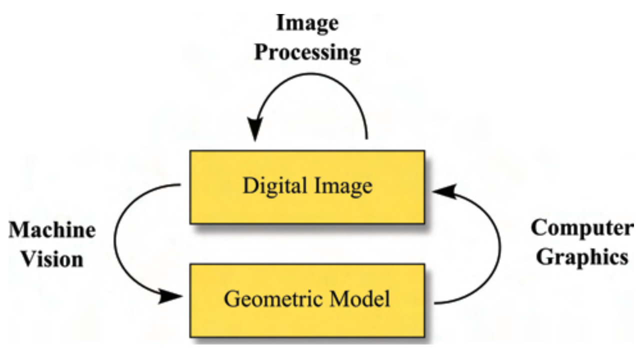


Figure 1.1. Disciplines related to image processing.

Figure 1.2. Image processing pipeline.

computer graphics: to analyze a digital image and infer meaningful information about the scene depicted by the image. Figure 1.1 illustrates the roles and relationships between each of these three disciplines where boxes represent a type of data while the connecting arrows show the typical input and output for a field of study.

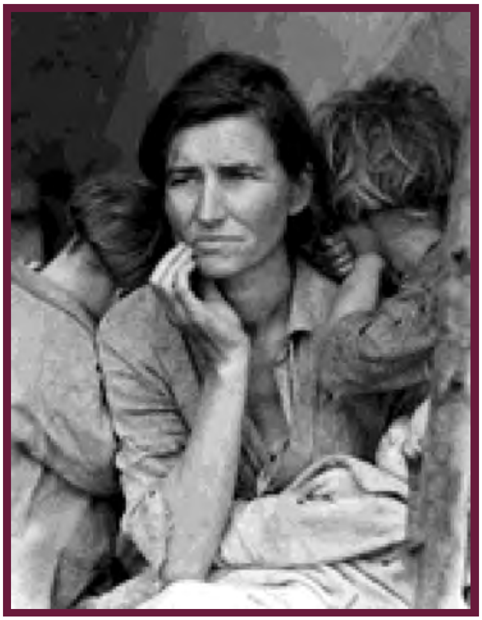
A complete digital image processing system is able to service every aspect of digital image handling. Figure 1.2 shows the five typical stages in an image processing pipeline: image acquisition, image processing, image archival, image transmission, and image display. Image acquisition is the process by which digital images are obtained or generated. Image processing is the stage where a digital image is enhanced or analyzed. Image archival is concerned with how digital images are represented in memory. Image transmission is likewise concerned with data representation but places added emphasis on the robust reconstruction of potentially corrupted data due to transmission noise. Image display deals with the visual display of digital image data whether on a computer monitor, television screen, or printed page.

A visual example of the pipeline stages is given in Figure 1.3. During the image acquisition stage, an approximation of a continuous tone or analog scene is recorded. Since the captured image is an approximation, it includes some error which is introduced through sampling and quantization. During archival, a further degradation of quality may occur as the concern to conserve memory and hence conserve transmission bandwidth competes with the desire to maintain a high quality image. When the image is displayed, in this case through printing in black and white, image quality may be compromised if the output display is unable to reproduce the image with sufficient resolution or depth of color.

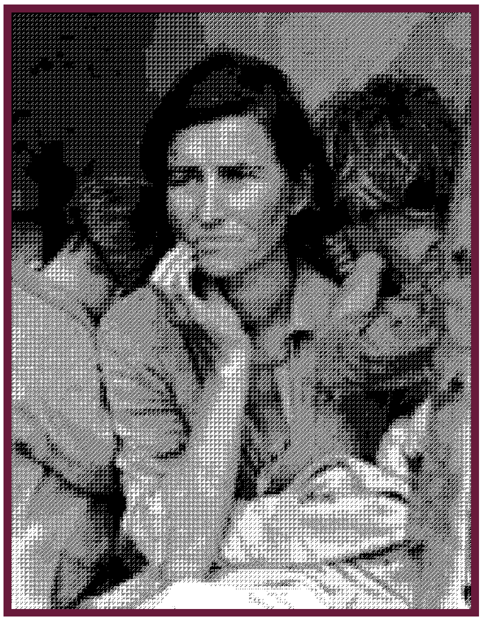
Construction of a complete image processing system requires specialized knowledge of how hardware architecture, the physical properties of light, the workings of the human visual system, and the structure of computational techniques affects each stage in the pipeline. Table 1.1 summarizes the most important topics of study as they correspond to each of the five primary stages in an image processing system. Of course a deep understanding of each of the listed areas of study is required to construct an efficient and effective processing module within any stage of the pipeline. Nevertheless, each stage of the processing pipeline raises unique concerns regarding memory requirements, computational efficiency, and image quality. A thorough understanding of the affects of each stage on image processing is required in order to achieve the best possible balance among memory, computation time, and image quality.

(a) Scene.

(b) Acquired.



(c) Archived.



(d) Displayed.

Figure 1.3. Effects of image processing stages on a processed image.

Processing Stage	Topic of Study
acquisition	physical properties of light human perception mathematical models of color
processing	software architecture data representation algorithm design
archival	compression techniques data representation
transmission	data representation transmission protocols
display	digital halftoning color models human perception

Table 1.1. Topics of study in image processing.

These five stages serve as a general outline for the remainder of this text. The image processing topics associated with each stage of the processing pipeline will be discussed with an emphasis on the processing stage which lies at the heart of image processing. By contrast, little coverage will be allocated to transmission issues in particular.

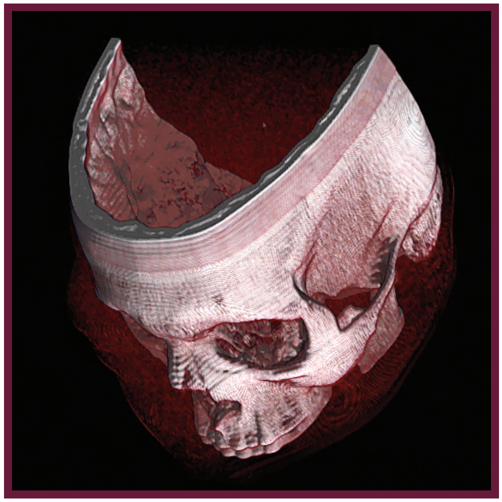
1.2 Why Digital Image Processing?

Digital images are used across an exceptionally wide spectrum of modern life. Ranging from digital cameras and cell phones to medical scans and web technology, digital image processing plays a central role in modern culture. This section provides examples of practical applications of image processing techniques. A general overview of these applications suffices to illustrate the importance, power, and pervasiveness of image processing techniques.

1.2.1 Medicine

Digital imaging is beginning to supplant film within the medical field. Computed tomography (CT) is a noninvasive imaging technique used to diagnose various ailments such as cancers, trauma, and musculoskeletal disorders. Magnetic resonance imaging (MRI) is a similarly noninvasive method for imaging the internal structure and function of the body. MRI scans are more amenable to diagnosing neurological and cardiovascular function than CT scans due to their greater contrast among soft tissue volumes. Figure 1.4 gives an example of both MRI and CT images where the MRI highlights contrast in the internal soft-tissue organs of a human pelvis while the CT image captures the internal skeletal structure of a human skull.

(a) MRI.



(b) CT.

Figure 1.4. Medical images.

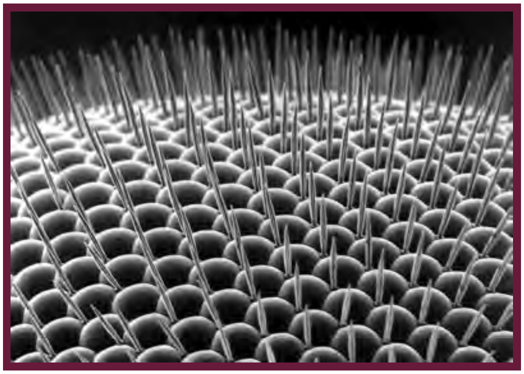
Since errors in the acquisition, processing, archival, or display of medical images could lead to serious health risks for patients, rigorous standards have been developed to ensure that digital images for medical use are properly archived and displayed. The Digital Imaging and Communications in Medicine (DICOM) is one such standard and has become the de facto standard for image processing in the health professions.

1.2.2 Biology

Biology is a natural science that studies living organisms and how they interact with the environment. Biological research covers a vast array of specialized subdisciplines such as botany, zoology, cell biology, microbiology, and biochemistry. Each of these disciplines relies to some degree on sophisticated computing systems to acquire and analyze large amounts of image-based data. These measurements ultimately provide information required for tasks such as deciphering complex cellular processes and identifying the structure and behavior of DNA.

Since image-based measurement is becoming increasingly vital to biological research, biologists must have basic knowledge in image processing to correctly interpret and process their results. Part (a) of Figure 1.5 shows a scanning electron microscope (SEM) image of a rust mite where the length of the mite is on the order of 60 μm. Part (b) shows the structure of the eye of a fruit fly where each spherical sensor is on the order of 10 μm in diameter.

(a) A rust mite (*Aceria anthocoptes*).



(b) The eye of a fruit fly (*Drosophilidae*).

Figure 1.5. Images in biology.

1.2.3 Biometrics

The security of national, corporate, and individual assets has become a topic of great importance in the modern global economy as terrorists, con men, and white-collar criminals pose an ongoing threat to society. When a person boards an airplane, enters credit card information over the Internet, or attempts to access medical records for a hospital patient, it is desirable to verify that the person actually is who they claim to be. The field of biometrics seeks to verify the identity of individuals by measuring and analyzing biological characteristics such as fingerprints, voice patterns, gait, facial appearance, or retinal scans. In most of these techniques, with the exception of voice recognition, the biological traits are obtained by the analysis of a digital image.

Biometrics has been used for decades in law enforcement to identify criminals from fingerprint images. Highly trained experts have traditionally performed fingerprint identification manually by comparing fingerprints of criminal suspects with fingerprints obtained from a crime scene. Systems are now commonly used to match fingerprints against large databases of suspects or known criminals. Specialized hardware is used to first acquire a digital image of an individual's fingerprint. Software is then used to analyze the image and compare it with a large database of known fingerprint images. Since the process is automated, it is possible to quickly search a very large database and quickly obtain accurate verification.

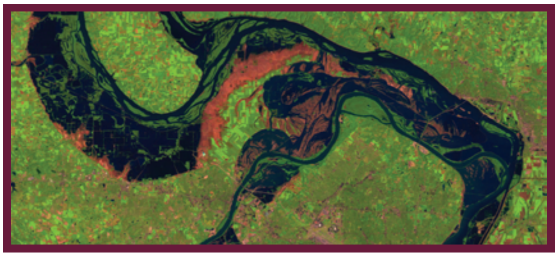
The use of palm scans is proving increasingly effective in the field of biometrics. A palm scanner is used to acquire an image of the blood flow through the veins of the hand in a completely noninvasive and contact-free fashion. Since the veins form a complex three-dimensional structure within a person's palm, individuals can be identified with extremely high accuracy, and forgery is extremely difficult.

1.2.4 Environmental Science

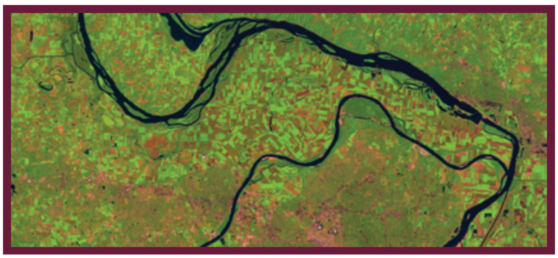
All life depends upon a healthy environment and the environmental sciences seek to understand the forces that affect our natural world. Environmental science is a broad and interdisciplinary field that includes the study of weather patterns (meteorology), oceans (oceanography), pollution as it affects life (ecology), and the study of the earth itself (the geosciences).

Data acquisition and analysis plays a key role in each of these fields since monitoring oceans, forests, farms, rivers, and even cities is critical to proper stewardship. Computer and imaging systems play an increasingly active and central role in these tasks. Satellite imaging is used to monitor and assess all types of environmental phenomena, including the effects of wildfires, hurricanes, drought, and volcanic eruptions. Motion-sensitive cameras have been installed in remote regions to monitor wildlife population densities. In recent years, these systems have discovered many new species and have even taken photographs of animals long believed extinct.

Figure 1.6 shows two enhanced satellite images of St. Louis, Missouri. The image in Figure 1.6(a) was taken during the great flood of 1993 while the image in Figure 1.6(b) was taken the following year. Environmental scientists tracked and measured the extent of the flood and the effect of the flood on terrain, vegetation, and city structures through sophisticated imaging systems and software.



(a) Satellite image in 1993.



(b) Satellite image in 1994.

Figure 1.6. Satellite images of the St. Louis flood. (Image courtesy of NASA/Goddard Space Flight Center Scientific Visualization Studio.)

1.2.5 Robotics

The field of robotics has made astounding progress in recent years. Robots now appear on the shelves of commercial toy stores, in industrial manufacturing lines, and in search and rescue missions. At the heart of most intelligent robots is a set of image processing routines that is able to process images gathered by the robot's "eyes" and determine how the robots should respond to their visually perceived environment. A team of robotics experts from the University of Southern Florida was brought in to assist in the search and rescue mission during the days after the World Trade Center collapse. These robots were specifically designed to navigate through dangerous situations looking for signs of life.

1.2.6 Professional Sports

Most professional sports leagues are developing computer systems to improve either the sports telecast or to assist umpires and referees throughout the game. The US Tennis Association, for example, uses specialized image processing systems to assist in making line calls. Officials were having increased difficulty with making correct calls as skilled tennis players can now generate 150 mile-per-hour serves and 100 mile-per-hour backhands.

Major League Baseball has also installed complex image processing systems to record the trajectory of each pitch made during a baseball game. Two cameras track the motion of the ball and are able to triangulate the position to within 1/2 inch accuracy over the entire trajectory of the pitch. A third camera is used to monitor the batter and determine the strike zone by computing the batter's knee-to-chest position. While the system is not used during game play it is used to augment television broadcasts. High-performance image processing algorithms superimpose the pitch trajectory and strike zone on instant replays. This gives sports fans an objective way to decide if the pitch was a ball or a strike. Major League Baseball does use the system to rate the performance of plate umpires in calling balls and strikes. At the conclusion of each game, the plate umpire is given a CD-ROM containing the trajectories of every pitch along with a comparison between the computer and umpire calls made.

Other sports have successfully used image-processing techniques for both decision-making and aesthetic purposes. Most major networks airing National Football League games superimpose yellow "first down" markers onto the playing field. These yellow stripes are obviously not actually on the field, but are applied using real-time image processing techniques. With the decreasing cost of computational power, it is to be expected that image processing will become more prevalent in all areas of professional sports.

1.2.7 Astronomy

Astronomers have long used digital images to study deep space over much of the electromagnetic spectrum: the Compton Gamma Ray Observatory captures digital images primarily in the gamma ray spectrum; the Chandra X-Ray Observatory and the Space Infrared Telescope Facility (also known as the Spitzer Space Telescope) provide coverage of the x-ray and infrared portions of the spectrum, respectively. The most well known telescope covering the visible portion of the spectrum is the Hubble Space Telescope, which was launched in 1990. The Hubble Telescope orbits the earth with a reflector-style optics system and a mirror of 2.4 meters in diameter. The focal length is 57.6 meters and it is able to take infrared images as well as images in the visible spectrum. Of course the images are digital since they must be transmitted to ground stations for viewing and analysis. The Hubble has produced some of the most remarkable images ever taken of created order.

Figure 1.7 is an image of the Antennae galaxies. These two galaxies are located in the constellation Corvus and are in the process of collapsing into a

Figure 1.7. Hubble Space Telescope image of the Antennae galaxies. (Image courtesy of NASA, ESA, and the Hubble Heritage Team.)

single galaxy. These galaxies are approximately 45 million light years away, and scientists predict that within 400 million years the two galaxies will have merged to form a single elliptical galaxy.

1.2.8 Conclusion

Ours is an increasingly visual culture and digital imaging is pervasive across nearly all professions, disciplines, and academic fields of study. The study of digital image processing will provide a foundation for understanding how best to acquire digital images, the nature of information contained within a digital image, and how to best archive and display images for specific purposes or applications.

Artwork

Figure 1.3. *"Migrant Mother"* by Dorothea Lange (1895–1965). Dorothea Lange was born in Hoboken, New Jersey in 1895 and devoted herself to portrait photography at a young age. After apprenticing with a photographer in New York City, she moved to San Francisco and worked predominantly with the upper class. After about 13 years she developed the desire to see things from a different point of view and Lange began shooting among San Francisco's unemployed and documenting the increasing labor unrest. She was eventually hired by the Farm Security Administration (FSA) as a photographer and photojournalist. She is best known for her work with the FSA, which put a human face on the tragedy of the Great Depression and profoundly influenced the field of photojournalism in subsequent years. She died on October 11, 1965. Her most famous portrait is entitled "Migrant Mother," which is shown in Figure 1.3. The image is available from the United States Library of Congress's Prints and Photographs Division using the digital ID fsa.8b29516.

Optics and Human Vision 2

This chapter gives an overview of the physical properties of light, optics and the human visual system. The chapter provides important background for understanding how image processing mirrors the biological mechanisms of human perception and how the properties of human perception can be leveraged for computational advantage.

2.1 Light

In a physical sense light is composed of particles known as photons that act like waves. The two fundamental properties of light are the amplitude of the wave and its wavelength. Amplitude is a measure of the strength of the light wave where higher amplitude light waves are perceived as brighter. Wavelength measures, as the term itself indicates, the length of the wave and is typically given in meters. Light waves are also commonly characterized by their frequency, which is inversely proportional to the wavelength. The relationship between the wavelength λ and frequency f of a light wave traveling through a vacuum is

$$\lambda = \frac{c}{f},$$

where c is the speed of light, which is equal to 299,792,458 m/s. Since wavelength is given in meters, frequency is therefore given in units of $\mathrm{seconds}^{-1}$, also known as hertz. Frequency measures the number of oscillations that a wave generates over a duration of one second of time. In terms of human perception, lower frequency (longer wavelength) light is perceived as the so-called *warmer* colors (red, yellow) while higher frequency (shorter wavelength) light is perceived as the *cooler* colors (violet, blue). While wavelength corresponds to color perception, amplitude corresponds to brightness, since brightness is proportional to the average energy over some time period. Figure 2.1 depicts a sinusoidal wave showing the relationship of amplitude and wavelength.

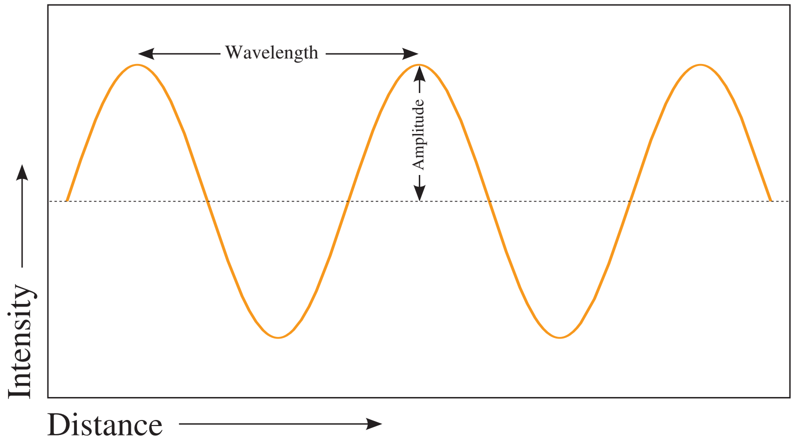


Figure 2.1. Diagram of a light wave.

The human eye can only perceive energy waves that have wavelengths within a very narrow band. Visible light is a relatively small portion of the larger electromagnetic spectrum which is composed of energy waves outside of our ability to perceive. The electromagnetic spectrum includes wavelengths as short as approximately 10^{-34} meters and as long as the size of the universe itself, although these are theoretical limits that have never been empirically verified. Electromagnetic radiation with wavelengths too small for the human eye to see are known, in decreasing order, as ultraviolet (UV), x-rays, and gamma rays. Electromagnetic radiation with wavelengths too large for the human eye to see are, in increasing order, infrared (IR), microwaves, and radio/TV.

Some living creatures are able to detect light outside of the range of wavelengths perceptible by humans. Bees, for example, are able to see UV radiation and pit viper snakes can see into the IR region using sensitive biological sensors in the pits of their heads. For human perception, however, the visible portion of the spectrum includes light with wavelengths ranging between about 400 and 700 nanometers. Light that has a wavelength of 650 nm appears red, while light that has a wavelength of 550 nm appears green; and light with a wavelength of about 475 nm appears blue. Figure 2.2 depicts the electromagnetic spectrum, highlighting the relatively narrow band of energy which is visible to the human eye. This figure depicts the portion of the spectrum spanning gamma rays, which have wavelengths on the order of 10^{-14} meters in length, to radio and television signals, which have wavelengths on the order of 10^4 meters.

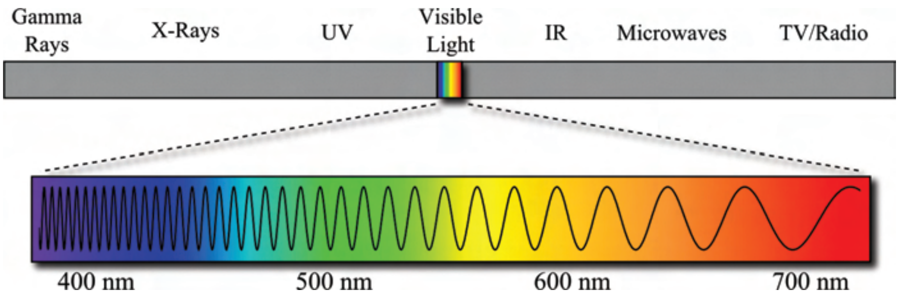


Figure 2.2. Electromagnetic spectrum.

2.2 Camera Optics

A lens is a transparent device, usually glass, that allows light to pass through while causing it to either converge or diverge. Assume that a camera is focused upon a target object. If the distance from the target to the lens is S_1 and the distance from the lens to the film is S_2 then these values are related by the thin lens equation shown in Equation (2.1), where f is the focal length:

$$\frac{1}{S_1} + \frac{1}{S_2} = \frac{1}{f}. \tag{2.1}$$

The focal length is a measure of how strongly a lens converges light and is defined as the distance from the lens to the point at which parallel rays passing through the lens converge. According to the thin lens equation, if we assume that an object is positioned a fixed distance S_1 from the lens, then a lens having a short focal length must produce correspondingly smaller images than a lens with a larger focal length. This follows since the focal length is directly proportional to the size of the image. Figure 2.3 illustrates light originating from a real-world target object, and passing through the lens and onto the imaging plane as a real image.

The magnification factor m of a lens is another measure of how strongly a lens converges light and is given by Equation (2.2). This formulation may initially seem inverted since digital cameras are often characterized by *optical zoom*, which is often misunderstood to be the magnification factor. The optical system of a camera is composed of multiple lenses where the positions of the lenses can vary and thereby adjust the focal length of a camera within some narrow range. The amount by which the focal length can be changed, and hence the resulting image size, is described as the *optical zoom*. The magnification factor

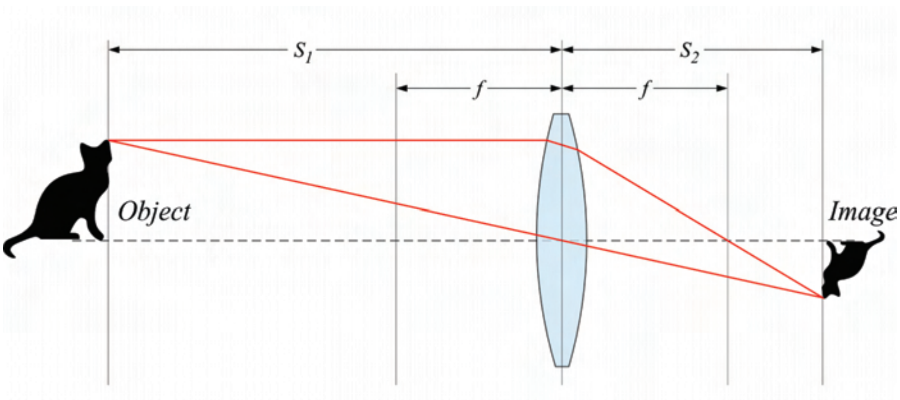


Figure 2.3. Optical properties of a convex lens.

of a single thin lens is typically less than 1 and should be understood to mean that an object of size S_1 appears to be of size S_2 when projected onto the focal plane:

$$m = \frac{\text{image size}}{\text{object size}} = \frac{S_2}{S_1}. \tag{2.2}$$

The aperture of a camera is the opening that determines the amount of light that is focused onto the imaging plane. The aperture of a camera is typically controlled by an adjustable diaphragm, which is allowed to expand or contract to either increase or decrease the amount of light entering the system. Of course the aperture can be no larger than the lens itself and is generally set smaller than the lens by the diaphragm. In photography the lens aperture is usually specified as an *f-number*, which is the ratio of focal length to the diameter of the aperture and is denoted as *f*/#. A lens typically has preset aperture openings named *f-stops*, which allow a predetermined amount of light into the camera. These f-stops are conventionally set such that each successive f-stop either halves or doubles the total amount of light passing through the lens.

A circle of diameter d has an area of $\pi \cdot (\frac{d}{2})^2$, which implies that a halving of the area requires decreasing the diameter by a factor of $\sqrt{2}$. For this reason, the conventional f-stop scale is a geometric sequence involving powers of $\sqrt{2}$ as in the sequence $\{1, 1.4, 2, 2.8, 4, 5.6, \ldots\}$. Figure 2.4 illustrates how the diaphragm of the lens controls the amount of light passing through the optical system and shows the significance of the f-number. The open area of the lens labeled as $f/1.4$ is twice the area of the opening labeled as $f/2$, which is itself twice the area of the opening labeled as $f/2.8$.

Figure 2.4. F stops.

2.3 Human Visual System

2.3.1 The Human Eye

The human visual system is an exceptionally sophisticated biological imaging system composed of many interrelated biomechanical and biochemical subcomponents. The eye is, of course, the primary component in this system and is itself composed of various parts. The eye is spherical in shape, having a diameter of approximately 20 mm. Figure 2.5 shows a cross section of the human eye and labels each of the primary structural components.

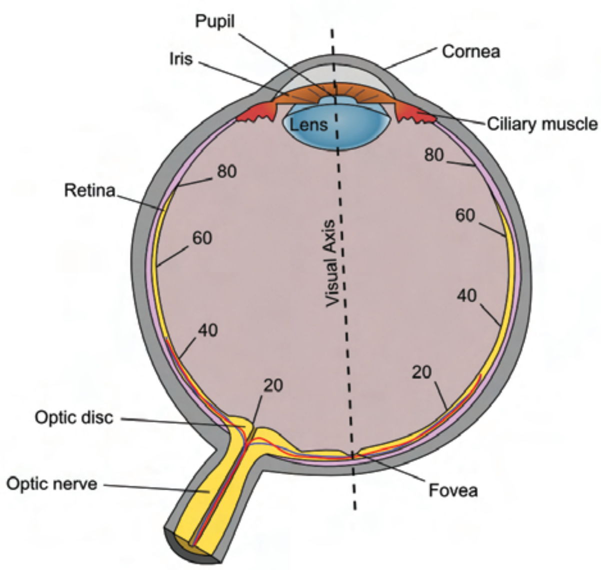


Figure 2.5. Primary structural elements of the human eye.

The cornea is a tough, transparent, dome-shaped membrane covering the front of the eye. The cornea is a protective covering over the eye and also functions to roughly focus incoming light. The iris, a diaphragm that expands and contracts to control the amount of light entering the eye, lies immediately behind the cornea. The central hole of the iris is the pupil, which varies in size from about 2 to 8 mm.

The lens is a convex and highly malleable disk positioned immediately behind the iris. While the cornea is responsible for most of the overall refraction of the eye, the curvature of the cornea is fixed, which requires the lens to fine-tune the focus of the eye. Ciliary muscles are used to adjust the thickness of the lens which adjusts both the curvature and the degree of refraction of the lens. When focusing at a distance, the lens flattens and therefore diminishes the degree of refraction while the lens thickens to focus on nearby objects.

The rear portion of the eye, known as the retina, is analogous to film in a conventional camera. The retina covers approximately 72% of the interior surface of the eye and is covered with two types of photoreceptor cells known as rods and cones. These biological sensors are able to convert light energy into electrical impulses, which are then transmitted to the brain through the optic nerve and interpreted as colors and images. There are approximately 6 million cones and 120 million rods in the human eye [Oyster 99]. The visual axis of the eye can be defined as a line extending between the center of the lens and the center of the retina, known as the fovea. The angle between any point on the retinal surface and the visual axis is given as the perimetric angle. Figure 2.5 shows the visual axis and the perimetric angles of various points on the retinal surface, spanning approximately -80° to $+80^\circ$.

Rods and cones are types of sensors that serve different purposes. Rods are able to function in relatively low light environments while cones function in relatively bright light environments. A single rod cell is able to detect a single photon of light and is about 100 times more sensitive than a cone cell. Scotopic vision, also known as night vision, is produced exclusively through rod cells since cone cells are not responsive to low light levels. Photopic vision is the normal vision of the eye under well-lit environments and is produced through the cone cells.

The rods and cones are also distributed throughout the retina in different ways. The 120 million rods of the human eye are concentrated around the outer edges of the retina and are therefore used in peripheral vision, while the 6 million cones in the retina are concentrated near the center of the retina, becoming less dense near the outer ring. This implies that in dark conditions the eye can more easily see objects by focusing slightly off-center of the object of interest, while in well-lit conditions the eye can more easily see by focusing directly on the target. Figure 2.6 plots the distribution of the rods and cones as a function of perimetric angle. Whereas the cones are clustered near the fovea (at a perimetric angle of 0°), there are essentially no rods. As the perimetric angle moves away from

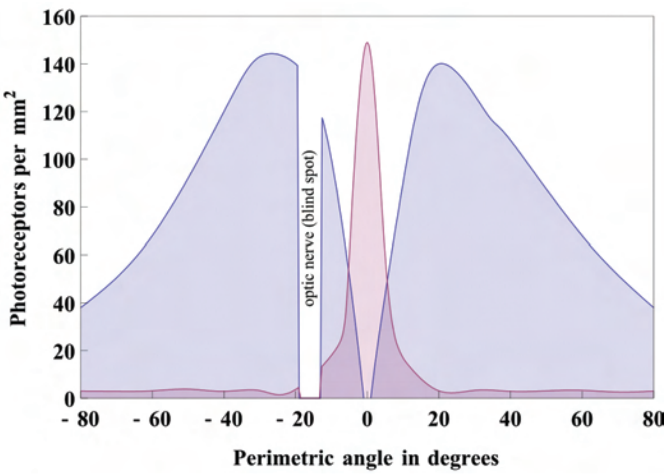


Figure 2.6. Photoreceptor density in the human eye as a function of perimetric angle.

the visual axis, however, the density of cones decreases sharply while the rod density increases sharply, peaking at approximately $\pm 35°$.

The optic nerve is responsible for carrying signals from the photoreceptors of the retina to the brain for analysis and interpretation. The optic nerve is composed of 1.2 million synaptic fibers and introduces a blind spot in the human eye since there are no photoreceptors in that area of the retina. Figure 2.6 indicates that the blind spot is at a perimetric angle of about $15°$ on the nasal periphery (the hemisphere of the eye oriented toward the nose).

Perhaps the most important distinction to make between the rods and cones is the role each has in discerning color. The rods, which are operative in scotopic vision, have no ability to discern color and hence see only in shades of gray. The primary reason that objects look washed out in dim light is that only the rods are active and can only detect differences in light intensity, not differences in color. Cones, by contrast, are able to discern differences in both intensity and color. The ability to see color derives from the fact that there are three separate types of cones. The three cone types respond differently to different light wavelengths such that one cone is sensitive to red, another to green, and another to blue wavelengths of light. These three types of cone are known as **L**, **M**, and **S** since they respond to **l**ong, **m**edium and **s**hort wavelengths of light, respectively. The **L** type cone has a peak sensitivity at wavelengths near 564–580 nm (red) while the **M** type peaks at 534–545 nm (green) and the **S** type peaks at 420–440 nm (blue). Color perception is thus gained by combining the sensory information of the three cone types. Roughly 65% of all cones are **L** type while 30% are **M** and 5% are **S** [Roorda et al. 01]. Table 2.1 is derived from [Kandel et al. 00] and summarizes the differences between the rods and cones.

Rods	Cones
Used for night vision	Used for day vision
Loss causes night blindness	Loss causes legal blindness
Low spatial resolution with higher noise	High spatial resolution with lower noise
Not present in the fovea	Concentrated in the fovea
Slower response to light	Quicker response to light
One type of photosensitive pigment	Three types of photosensitive pigment
Emphasis on detecting motion	Emphasis on detecting fine details

Table 2.1. Comparison between rods and cones.

2.3.2 Log Compression

Experimental results show that the relationship between the *perceived* amount of light and the *actual* amount of light in a scene are generally related logarithmically. The human visual system perceives brightness as the logarithm of the actual light intensity and interprets the image accordingly. Consider, for example, a bright light source that is approximately six times brighter than another. The eye will perceive the brighter light as being approximately twice the brightness of the darker. Log compression helps to explain how the human visual system is able to sense light across an exceptionally broad range of intensity levels.

2.3.3 Brightness Adaptation

The human visual system has the remarkable capacity of seeing light over a tremendously large range of intensities. The difference between the least amount of light needed to see (the scotopic threshold) and the maximum amount of light that we can see (the glare limit) is on the order of 10^{10}. The key to understanding how it is possible to perceive images across such a vast range of light levels is to understand that we cannot perceive light at both the low and upper limits *simultaneously*. At any particular instant we can discriminate among a very small portion of our full dynamic range. The sensitivity of the eye changes, or adapts, to the average brightness of the region on which it is focused. The eye calibrates itself to the average intensity of the region on which the eye is focused; it is then able to discriminate among a few dozen light levels which are centered about the calibration point. The range of intensities that the eye can see at one point in time is known as the instantaneous range of vision, while the calibration or adjustment to ambient lighting conditions is known as brightness adaptation.

Figure 2.7 shows that perceived brightness is a nonlinear function of the actual light intensity. The eye can perceive light across the full adaptive range but at any point in time can see only within the much narrower instantaneous range. The adaptive range is on the order of 10^{10} units and hence the graph vertical axis is shown using a log-compressed scale.

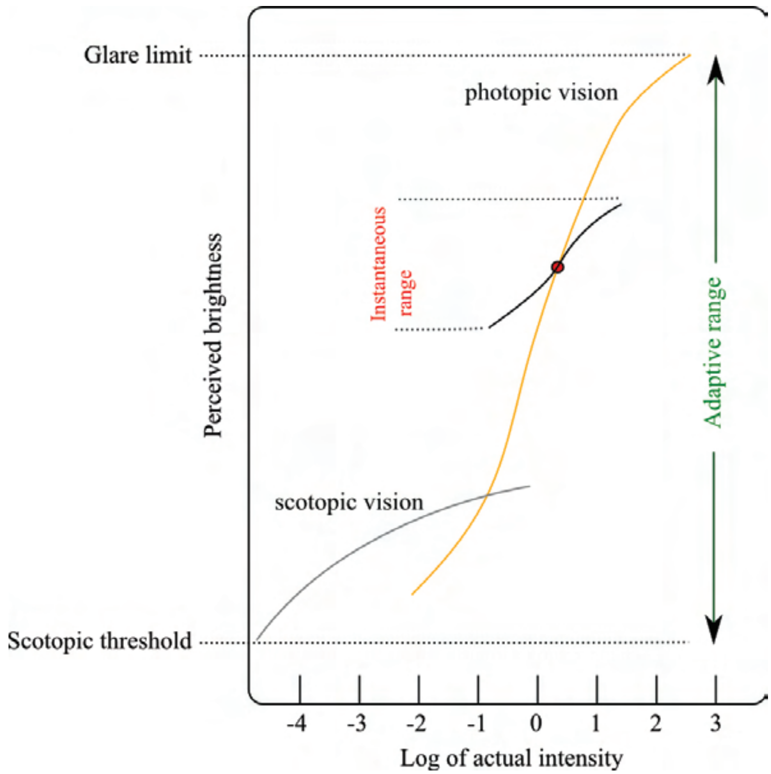


Figure 2.7. Brightness adaptation.

Consider taking a walk around noon on a cloudless and bright summer day. While the light is bright you are nonetheless able to discern with clarity your surroundings and make out details of the flowers and faces that you encounter. As you return home, however, and step through the doorway into a dark hallway entrance, the image is dull and dark. For a few moments the eye perceives almost pure darkness but as your visual system adapts to the overall brightness of the interior room, the sensitivity increases and more details become clear and visible. After some time it may even appear as if the interior room is as well lit as the sunny outdoors. While brightness adaptation allows us to see in extreme lighting conditions it has the disadvantage that we are unable to accurately estimate the actual intensity of light in our environment.

High dynamic range imaging (HDR) is an image processing technique that gives a displayed image the appearance of greater contrast than is actually displayed. HDR works by taking advantage of local brightness adaptation to fool the eye into thinking that it is seeing a much larger range of values than it really is.

2.3.4 Mach Banding

When viewing any scene the eye rapidly scans across the field of view while coming to momentary rest at each point of particular interest. At each of these points the eye adapts to the average brightness of the local region immediately surrounding the point of interest. This phenomena is known as *local* brightness adaptation. Mach banding is a visual effect that results, in part, from local brightness adaptation.

Figure 2.8 gives a popular illustration of Mach banding. While the figure of part (a) contains 10 bands of solid intensity the eye perceives each band as being either brighter or darker near the band edges. These changes in brightness are due to the eye scanning across the edge boundary, thus causing local adaptation to occur; this results in a changing perception of light intensity in that region when compared to the centers of the bands. The graph of part (b) is an approximation to perceived brightness as the visual system sharpens edges by increasing the contrast at object boundaries.

(a) Mach banding. (b) Row profile of perceived brightness.

Figure 2.8. Mach banding effect.

2.3.5 Simultaneous Contrast

Simultaneous contrast refers to the way in which two adjacent intensities (or colors) affect each other. The image of Figure 2.9 is a common way of illustrating that the perceived intensity of a region is dependent upon the contrast of the region with its local background. In this figure, the four inner squares are of

Figure 2.9. Simultaneous contrast.

identical intensity but are contextualized by the four surrounding squares, and thus the perceived intensity of the inner squares varies from bright on the left to dark on the right. This phenomena can be intuitively explained by noting that a blank sheet of paper may appear white when placed on a desktop but may appear black when used to shield the eyes against the sun [Gonzalez and Woods 92].

2.4 Exercises

1. A professional baseball league intends to save money and increase accuracy by replacing all umpires by automated computer vision systems. Miller Park will participate in a test run of the system and is purchasing a set of digital cameras that will serve as the front-end to the system. Each camera must be able to view a baseball at a resolution of at least 75 pixels per inch (75 digital pixels per inch of actual baseball) from a distance of at most 90 feet. Specify the minimum magnification and focal length required if the camera uses a CCD chip having 2048×2048 square pixels in an active sensor area of 3.55×3.55 mm. Clearly state any assumptions that you make concerning your design.

2. Section 2.3.2 states that perception of light intensity is logarithmically related to the actual light intensity. Empirical studies have shown that a logarithmic base of approximately 2.5 should be used to model this logarithmic compression when viewing a point light source. Using this model, what is the perceived intensity of a point light source emitting 120 units of light intensity?

3. A graphic artist wants to construct an image similar to that shown in Figure 2.8(a) but which contains no Mach banding. Discuss how such an image could be constructed and provide a convincing argument that your image minimizes the Mach banding effect.

3 Digital Images

3.1 Introduction

A digital image is visual information that is represented in digital form. This chapter describes common mathematical models of color and also gives various techniques for representing visual information in digital form.

3.2 Color

As described earlier, the human visual system perceives color through three types of biological photo-sensors known as cones. Each cone is attuned to one of three wavelengths that correspond roughly to red, green, or blue light. The L-cone, or long wavelength cone, responds to red light, the M-cone, or medium wavelength cone, responds to green light, and the S-cone, or short wavelength cone, responds to blue light. The individual responses of all cones within a small region of the retina are then combined to form the perception of a single color at a single point within the field of view. The design of this biological system suggests that color is a three-dimensional entity.

A color model is an abstract mathematical system for representing colors. Since color is a three dimensional entity, a color model defines three primary colors (corresponding to three dimensions or axes) from which all possible colors are derived by mixing together various amounts of these primaries. Color models are typically limited in the range of colors they can produce and hence represent only a portion of the visible spectrum. The range of colors covered by a color model is referred to as either the gamut or the color space.

Color models can be classified as either an additive or subtractive. Additive color models assume that light is used to generate colors for display. In an additive color model, the color black represents a complete lack of the primary colors while the color white corresponds to maximal and equal amounts of each of the primaries. Additive color models assume that the individual primaries sum together into a single color. Human perception is an additive system since the perception of black is caused by the complete absence of light while white is perceived when large and equal amounts of red, green, and blue are present. Human perception is often described as a *tristimulus color space*, since it is based on an additive color model composed of three primaries. Computer monitors and LCD projectors are light-emitting devices that are naturally modeled using an additive color model.

Subtractive color models assume that pigment will be used to create colors such that a complete absence of any pigment corresponds to the color white while combining the three primaries in maximal and equal amounts yields the color black. Subtractive color models tend to be the more intuitive of these two ways of defining color since people are generally accustomed to this mode; since they tend to use pencil, ink, or paint pigments to create images. Subtractive color models assume that when white light is projected onto pigment, the pigment will absorb power from certain wavelengths and reflect power at other wavelengths. This absorption is described as *subtraction* since certain wavelengths are subtracted from any light that is projected onto the pigment. In terms of electronic systems, images rendered with ink-based printers are most naturally described using a subtractive color model.

Numerous color models are in common use, each having qualities that are well suited for a certain class of applications. These include the RGB, CMY, CMYK, HSB, and NTSC (or YIQ) models. Since matching the characteristics of a particular color model to the requirements of a particular application is an important task in the design of an image processing system, it is important to understand how color models are designed along with their properties. The following subsections discuss the most commonly used color models for image processing applications and describe their basic properties in greater detail.

3.2.1 RGB Color Model

The RGB color model is the most common way of representing color in image processing systems. The RGB model is additive and uses red, green, and blue as the primary colors or primary axes such that any color can be obtained by combining different amounts of these three primaries. By way of example, consider a flashlight that has a slider allowing one to choose the strength of light emitted. In setting the slider to zero, the flashlight is turned completely off and generates no light; in setting the slider to one, the flashlight generates as much light as it is capable of generating. Now consider three such flashlights: the first emits purely red light, the second emits purely green light, and the third emits purely blue light. If all three flashlights are aimed at the same spot on a white wall any color can be projected onto the wall by adjusting the slider values on the three lights

(a) RGB (additive). (b) CMY (subtractive).

Figure 3.1. Additive and subtractive models.

in different ways. If all sliders are set to zero, black is projected onto the wall. If all sliders are set to one, white is projected onto the wall, and if all sliders are set to 0.5, then gray is projected.

Figure 3.1 illustrates the RGB color model with three lights projecting onto a white wall in an otherwise black room. The location at which they overlap indicates the sum of all three lights which is seen as white. The secondary colors are produced by the overlap of only two of the three primaries. Figure 3.1 also shows the CMY model which is described in the following subsection.

The system of three flashlights is a precise analog of the RGB color model. A color within the RGB color space is defined by three numeric values (a *tuple*) that specify the amount of red, green, and blue that comprise the specified color. A color specified using the RGB model is said to lie within the RGB color space. In this text we adopt the convention that an RGB color specification is specified in normalized coordinates such that the value 0 indicates that none of the specified primary is present and the value 1 indicates that the maximum amount of the specified primary is present. For example, the color red is specified in RGB space as the tuple $\langle 1, 0, 0 \rangle$, the color cyan is specified as $\langle 0, 1, 1 \rangle$, and middle gray is specified as $\langle .5, .5, .5 \rangle$.

The RGB color space can be visualized as the unit cube shown in Figure 3.2. All colors representable in RGB space lie within the volume defined by the three primary axes corresponding to pure red, green, and blue. The origin lies at $\langle 0, 0, 0 \rangle$ (black) while the opposite corner lies at $\langle 1, 1, 1 \rangle$ and corresponds to the color white. Each point within the RGB color space corresponds to a unique color as illustrated in Figure 3.2(b). Note that the line connecting pure black and pure white within the RGB color space is known as the gray scale, as shown in Figure 3.2(c) since any point lying on that line is a shade of gray.