IMAGE PROCESSING AND ANALYSIS WITH VISION SYSTEMS

INTRODUCTION

There is a very large body of work associated with vision systems, image processing, and pattern recognition that addresses many different hardware- and software- related topics This information has been accumulated since the 1950s, and with the added interest in the subject from different sectors of the industry and economy, it is growing rapidly. The enormous number of papers published every year indicates that there must be many useful techniques constantly appearing in the literature. At the same time, it also means that a lot of these techniques may be unsuitable for other applications. In this chapter, we will study and discuss some fundamental techniques for image processing and image analysis, with a few examples of routines developed for certain purposes. The chapter does not profess to be a complete survey of all possible vision routines, but only an introduction.. It is recommended that the interested reader continue studying the subject through other references.

The next few sections present some fundamental definitions of terms and basic concepts that we will use throughout the chapter.

IMAGE PROCESSING VERSUS IMAGE ANALYSIS

Image processing relates to the preparation of an image for later analysis and use. Images captured by a camera or a similar technique (e.g., by a scanner) are not necessarily in a form that can be used by image analysis routines. Some may need improvement to reduce noise, others may need to be simplified, and still others may need to be enhanced, altered, segmented, filtered, etc. Image processing is the collection of routines and techniques that improve, simplify, enhance, or otherwise alter an image.

What Is an Image ?

Image analysis is the collection of processes in which a captured image that is prepared by image processing is analyzed in order to extract information about the image and to identify objects or facts about the object or its environment.

TWO- AND THREE-DIMENSIONAL IMAGES

Although all real scenes are three dimensional, images can either be two or three dimensional. Two-dimensional images are used when the depth of the scene or its features need not be determined. As an example, consider defining the surrounding contour or the silhouette of an object. In that case, it will not be necessary to determine the depth of any point on the object. Another example is the use of a vision system for inspection of an integrated circuit board. Here too, there is no need to know the depth relationship between different parts, and since all parts are fixed to a flat plane, no information about the surface is necessary. Thus, a two-dimensional image analysis and inspection will suffice.

Three-dimensional image processing deals with operations that require motion detection, depth measurement, remote sensing, relative positioning, and navigation. CAD/CAM-related operations also require three-dimensional image processing, as do many inspection and object recognition tasks. Other techniques, such as computed tomography (CT) scan, are also three dimensional. In computed tomography, either X-rays or ultrasonic pulses are used to get images of one slice (Section) of the object at a time, and later, all of the images are put together to create' a three-dimensional image of the internal characteristics of the object.

All three-dimensional vision systems share the problem of coping with many- to-one mappings of scenes to images. To extract information from these scenes, image-processing techniques are combined with artificial intelligence techniques. When the system is working in environments with known characteristics (e.g., controlled lighting), it functions with high accuracy and speed. On the contrary, when the environment is unknown or noisy and uncontrolled (e.g., in underwater operations), the systems are not very accurate and require additional processing of the information. Thus, they operate at low speeds. In addition, a three-dimensional coordinate system has to be dealt with.

WHAT IS AN IMAGE ?

An image is a representation of a real scene, either in black and white or in color, and either in print form or in a digital form. Printed images may have been reproduced either by multiple colors and gray scales (as in color print or half-tone print) or by a single ink source. For example, in order to reproduce a photograph with real half tones, one has to use multiple gray inks, which, when combined, produce an image that is somewhat realistic. However, in most print applications, only one color of ink is available (such as black ink on white paper in a newspaper or copier). In that case, all gray levels must be produced by changing the ratio of black versus white areas (the size of the black dot). Imagine that a picture to be printed is divided into small sections.

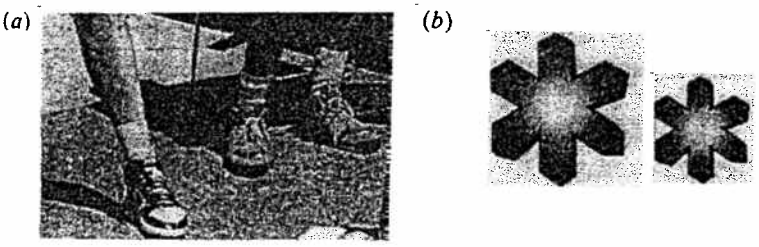


Figure 8.1 Examples of gray intensity creation in printed images. In print, only one color of ink is used, while the ratio of the black to the white area of the pixel is changed to create different gray levels.

In each section, if the ink portion of the section is smaller compared to the white-blank area, the section will look lighter gray. (See examples in Figure 8.1.) If the black ink area is larger compared to the white area, it will look darker gray. By changing the size of the printed dot, many gray levels may be produced, and collectively, a gray-scale picture may be printed.

Unlike printed images, television and digital images are divided into small sections called picture cells, or pixels (in three-dimensional images, they are called volume cells or voxels), where the size of all pixels are the same, while the **intensity of light in each pixel** is varied to create the gray images. Since we deal with digital images, we will always refer to pixels of the same size with varying intensities.

ACQUISITION OF IMAGES

There are two types of vision cameras: analog and digital. Analog cameras are not very common any more, but are still around; they used to be standard at television stations. Digital cameras are much more common and are mostly similar to each other. A video camera is a digital camera with an added videotape recording section. Otherwise, the mechanism of image acquisition is the same as in other cameras that do not record an image. Whether the captured image is analog or digital, in vision systems the image is eventually digitized. In a digital form, all data are binary and are stored in a computer file or memory chip.

The following short discussion is about **analog** and digital cameras and how their images are captured. Although analog cameras are not common anymore, **since the television sets available today are still mostly analog,** understanding the way the camera works will help in understanding how the television set works. Thus, both analog and digital cameras are examined here.

8.5.1 Vidicon Camera

A vidicon camera is an analog camera that transforms an image into an analog electrical signal. The signal, a variable voltage (or current) versus time, can be stored, digitized, broadcast, or reconstructed into an image. Figure 8.2 shows a simple schematic of a vidicon camera. With the use of a lens, the scene is projected onto a screen made up of two layers: a transparent metallic film and a **photoconductive mosaic** that is sensitive to light. The mosaic reacts to the varying intensity of light by varying its resistance. As a result, as the image is projected onto it, the magnitude of the **resistance at each location varies** with the intensity of the light. An electron gun generates and sends a continuous cathode beam (a stream of electrons with a negative

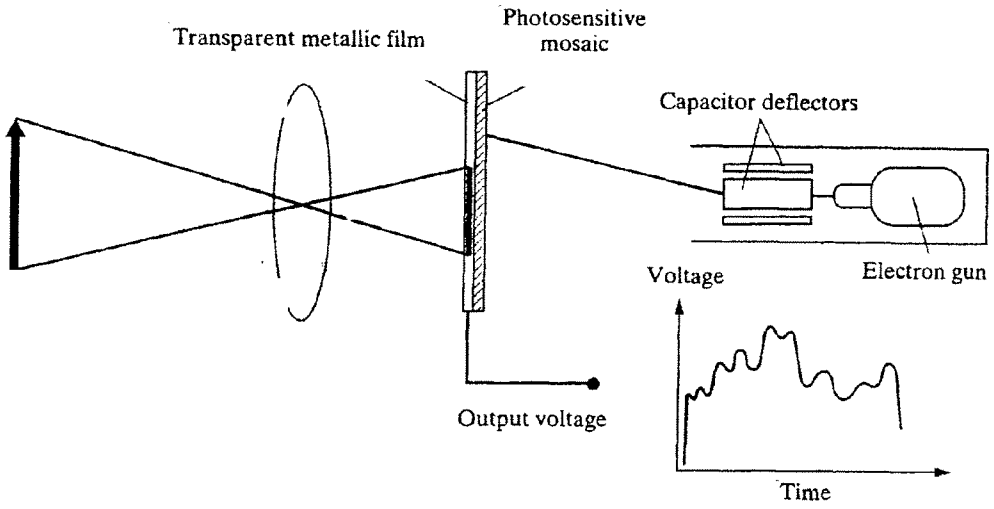


Figure 8.2 Schematic of a vidicon camera.

Voltage Electron gun Time

Acquisition Photosensitive Transparent metallic film mosaic Capacitor deflectors

Figure 8.2 Schematic of a vidicon camera.

---

charge) through two pairs of capacitors (deflectors) that are perpendicular to each other. Depending on the charge on each pair of capacitors, the electron beam is deflected up or down, and left or right, and is projected onto the photoconductive mosaic. At each instant, as the beam of electrons hits the mosaic, the charge is conducted to the metallic film and can be measured at the output port. The voltage measured at the output is V = IR, where I is the current (of the beam of electrons), and R is the resistance of the mosaic at the point of interest.

Now suppose that we routinely change the charges in the two capacitors and thus deflect the beam both sideways and up and down, so as to cause it to scan the mosaic (a process called a raster scan). As the beam scans the image, at each instant the **output** is proportional to the resistance of the mosaic or **proportional to the intensity of the light** on the mosaic. By reading the output voltage continuously, an analog representation of the image can be obtained.

To create moving images in televisions, the image is scanned and reconstructed **30 times a second**. Since human eyes possess a temporary hysteresis effect of about 1/10 second, images changing at 30 times a second are perceived as continuous and thus moving. The image is divided into two 240-line sub-images, interlaced onto each other. Thus, a television image is composed of **480 image lines**, **changing 30 times a second**. In order to return the beam to the top of the mosiac, another **45 lines are used,** creating a total of **525 lines**. In most other countries, 625 lines are the standard. Figure 8.3 depicts a raster scan in a vidicon camera.

If the signal is to be broadcast, it is usually frequency modulated (FM); that is, the frequency of the carrier signal is a function of the amplitude of the signal. The signal is broadcast and is received by a receiver, where it is de-modulated back to the original signal, creating a variable voltage with respect to time. To re-create the image — for example, in a television set — this voltage must be converted back to an image. To do this, the voltage is fed into a cathode-ray tube (CRT) with an electron gun and similar deflecting capacitors, as in a vidicon camera. The intensity of the electron beam in the television is now proportional to the voltage of the signal, and is scanned similar to the way a camera does. In the television set, however, the beam Output voltage

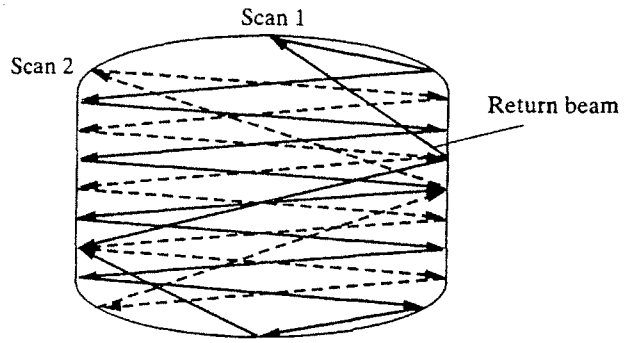


Figure 8.3 A raster scan depiction of a vidicon camera.

Scan 2

Scan 1

Figure 8.3 A raster scan depiction of a vidicon camera.

Return beam is projected onto a phosphorous-based material on the screen, which glows proportionally to the intensity of the beam, thus re-creating the image.

For color images, the projected image is decomposed into the **three colors of red, green, and blue (ROB).** The exact same process is repeated for the three images, and three simultaneous signals are produced and broadcast. In the television set, three electron guns regenerate three simultaneous images in RGB on the screen, except that the screen has three sets of small dots (pixels) that react by glowing in ROB colors and are repeated over the entire screen. All color images in any system are divided into ROB images and are dealt with as three separate images.

If the signal is not to be broadcast, it either is recorded for later use, is digitized (as discussed later), or is fed into a monitor for direct viewing.

8.5.2 Digital Camera

A digital camera is based on solid-state technology. As with other cameras, a set of lenses is used to project the area of interest onto the image area of the camera. The main part of the camera is a solid-state silicon wafer image area that has hundreds of thousands of extremely small photosensitive areas called **photo sites printed on it**. Each small area of the wafer is a pixel. As the image is projected onto the image area, at each pixel location of the wafer a charge is developed that is proportional to the intensity of light at that location. (Thus, a digital camera is also called a charge coupled device, or CCD camera, and a charge integrated device, or CID camera). The collection of charges, if read sequentially, would be a representation of the image pixels. (See Figure 8.4).

The wafer may have as many as 520,000 pixels in an area with dimensions of a fraction of an inch (/16 x ¼). Obviously, it is impossible to have direct wire connections to all of these pixels to measure the charge in each one. To read such an enormous number of pixels, 30 times a second the charges are moved to optically isolated shift registers next to each photo site, are moved down to an output line, and then are read [1,2]. The result is that every thirtieth of a second the charges in all pixel locations are read sequentially and stored or recorded. The output is a discrete representation of the image — a voltage sampled in time — as shown in Figure 8.5(a). Figure 8.5(b) is the CCD element of a VHS camera.

Similar to CCD cameras for visible lights, long-wavelength infrared cameras yield a television like image of the infrared emissions of a scene [3].

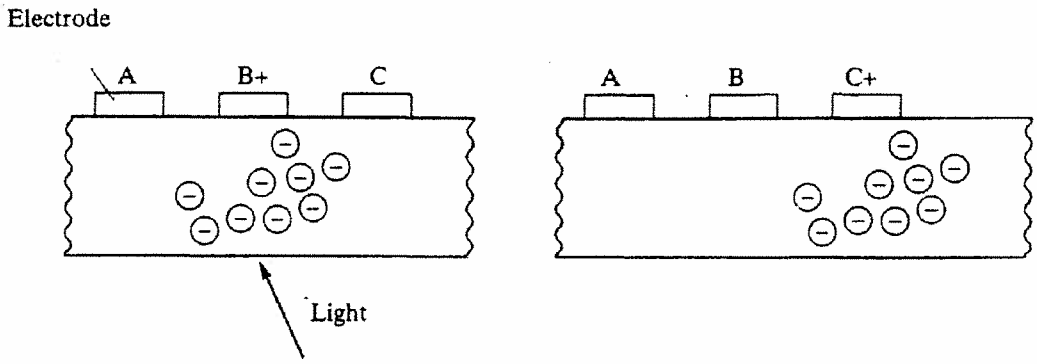
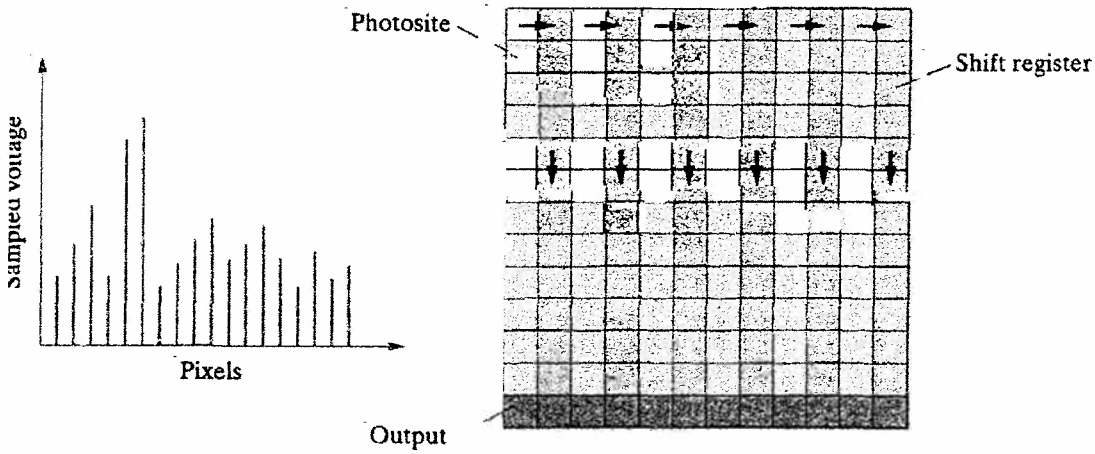
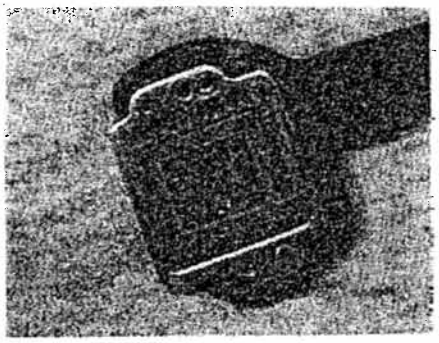


Figure 8.4 Image acquisition with a digital camera involves the development, at each pixel location, of a charge proportional to the light at the pixel. The image is then read by moving the charges to optically isolated shift registers and reading them at a known rate.



(*a*)



(*b*)

Figure 8.5 (a) Image data collection model. (b) The CCD element of a VHS camera

8.6 DIGITAL IMAGES

The sampled voltages from the aforementioned process are first digitized through an analog-to-digital converter (ADC) and then either stored in the computer storage unit in an image format such as TIFF, JPG, Bitmap, etc., or displayed on a monitor. Since it is digitized, the stored information is a collection of U's and l's that represent the intensity of light at each pixel; a digitized image is nothing more than a computer file that contains the collection of these 0's and l's, sequentially stored to **represent the intensity of light at each pixel.** The files can be accessed and read by a program. can be duplicated and manipulated, or can be rewritten in a different form. Vision routines generally access this information, perform some function on the data, and either display the result or store the manipulated result in a new file.

An image that has different gray levels at each pixel location is called a gray image. The gray values are digitized by a digitizer, yielding strings of 0's and l's that are subsequently displayed or stored. **A color image is obtained by superimposing three images of red, green, and blue hues, each with a varying intensity and each equivalent to a gray image (but in a colored state).** Thus, when the image is digitized, it will similarly have strings of U's and l's for each hue. **A binary image is an image such that each pixel is either fully light or fully dark — a 0 or a 1.** To achieve a binary image, in most cases a gray image is converted by using the histogram of the image and a cut-off value called a threshold. A histogram determines the distribution of the different gray levels. One can pick a value that best determines a cutoff level with least distortion and use that value as a threshold to assign 0's (or "off") to **all pixels whose gray levels are below the threshold value and to assign l's (or "on") to all pixels whose gray values are above the threshold**. Changing the threshold will change the binary image. The advantage of a binary image is that it requires far less memory and can be processed much faster than gray or colored images.

8.7 FREQUENCY DOMAIN VS. SPATIAL DOMAIN

Many processes that are used in image processing and analysis are based on the frequency domain or the spatial domain. In frequency-domain processing, the frequency spectrum of the image is used to alter, analyze, or process the image. In this case, the individual pixels and their contents are not used. Instead, **a frequency representation of the whole image** is used for the process. **In spatial-domain** processing, the process is applied to the **individual pixels** of the image. As a result, each pixel is affected directly by the process. However, the two techniques are equally important and powerful and are used for different purposes. Note that although spatial- and frequency-domain techniques are used differently, they are related. For example, suppose that a spatial filter is used to reduce noise in an image. As a result, noise level in the image will be reduced, but at the same time, the frequency spectrum of the image will also be affected, due to the reduction in noise.

The next several sections discuss some fundamental issues about frequency and spatial domains. The discussion, although general, will help us throughout the entire chapter.

8.8 FOURIER TRANSFORM AND FREQUENCY CONTENT OF A SIGNAL

As you may remember from your mathematics or other courses, any periodic signal may be decomposed into a number of sines and cosines of different amplitudes and frequencies as follows:

$$f(t) \quad \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\omega t + \sum_{n=1}^{\infty} b_n \sin n\omega t.$$

If you add these sines and cosines together again, you will have reconstructed the original signal. Equation (1) is called a Fourier series, and the collection of different frequencies present in the equation is called the frequency spectrum or frequency content of the signal. Of course, although the signal is in the amplitude—time domain, the frequency spectrum is in the amplitude—frequency domain. To understand this concept better, let's look at an example.

Consider a signal in the form of a simple sine function like f(t) = sin (t). Since this signal consists of only one frequency with a constant amplitude, if we were to plot the signal in the frequency domain, it would be represented by a single line at the given frequency, as shown in Figure 8.6. Obviously, if we plot the function represented by the arrow in Figure 8.6(b) with the given frequency and amplitude, we will have reconstructed the same sine function. The plots in Figure 8.7 are similar and represent

$$f(t) = \Sigma_{n=1,3...15}(1/n)\ \sin(nt)$$

The frequencies are also plotted in the frequency—amplitude domain. Clearly, when the number of frequencies contained in f(t) increases, the summation becomes closer to a square function.

Theoretically, to reconstruct a square wave from sine functions, an infinite number of sines must be added together. Since a square wave function represents a sharp change, this means that rapid changes (such as an impulse, a pulse, a square wave, or anything else similar to them or modeled by them) have a large number of frequencies. The sharper the change, the higher is the number of frequencies needed to reproduce it. Thus, any video (or other) signal that contains sharp changes (such as noise, high contrasts, or an impulse or step function) or that has

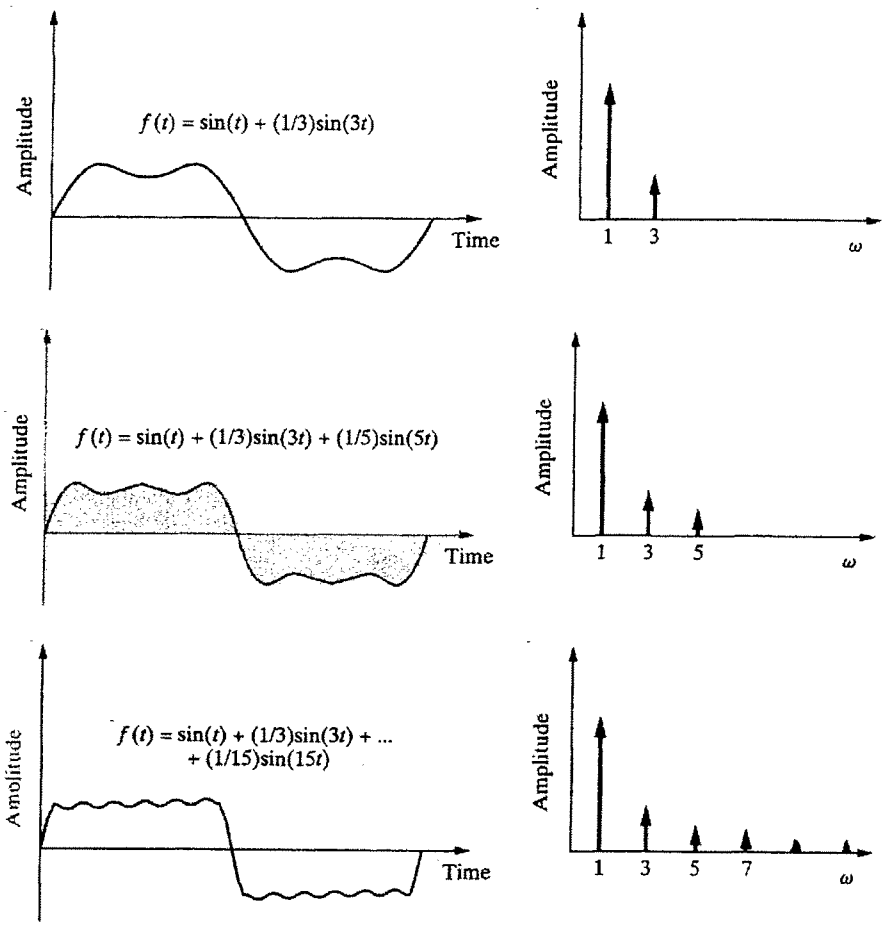


Figure 8.7 Sine functions in the time and frequency domain for a successive set of frequencies. As the number of frequencies increases, the resulting signal becomes closer to a square function.

detailed information (high-resolution signals with fast, varying changes) will have a larger number of frequencies in its frequency spectrum.

A similar analysis can be applied to **non-repeating signals** as well. (**The equation used is a Fourier transform or, sometimes, a fast Fourier Transform, or FFT**) Although we will not discuss the details of the Fourier transform in this book, suffice it to say that an **approximate frequency spectrum of any signal can be found.** Although, theoretically, there will be infinite frequencies in the spectrum, generally, some of the major frequencies within the spectrum will have **larger amplitudes**. **These major frequencies, or harmonics,** are used in identifying and labeling a signal, including recognizing voices, shapes, objects, etc.

8.9 FREQUENCY CONTENT OF AN IMAGE: NOISE, EDGES

Consider sequentially plotting the gray values of the pixels of an image (on the y axis) against (a) time or (b) pixel location (on the x-axis) as the image is scanned. (See Section 8.5.) The result will be a discrete time plot of varying amplitudes showing the intensity of light at each pixel, as indicated in Figure 8.8. Let's say that we are on the ninth row and are looking at pixels 129—144. The intensity of pixel **136** is very different from the intensities of the pixels around it and may be considered to be **noise**. (Generally, noise is information that does not belong to the surrounding environment.) The intensities of pixels **134 and 141** are also different from the neighboring pixels and may indicate a **transition** between the object and the background; thus, these pixels can be construed as representing the **edges** of the object.

Although we are discussing a discrete (digitized) signal, it may be transformed into a large number of sines and cosines with different amplitudes and frequencies that, if added together, will reconstruct the signal. As discussed earlier, slowly changing signals (such as small changes between succeeding pixel gray values) will require few sines and cosines in order to be reconstructed, and thus have low frequency content. On the other hand, quickly varying signals (such as large differences between pixel gray levels) will require many more frequencies to be reconstructed and thus have high frequency content. Both noises and edges are instances in which one pixel value is very different from the neighboring ones. Thus, **noises and edges create the larger frequencies of a typical frequency spectrum**, whereas slowly varying gray level sets of pixels, representing the object, contribute to the lower frequencies of the spectrum.
However, if a high-frequency signal is passed through a low-pass filter — a filter that allows lower frequencies to go through without much attenuation in amplitude, but that severely attenuates the amplitudes of the higher frequencies in the

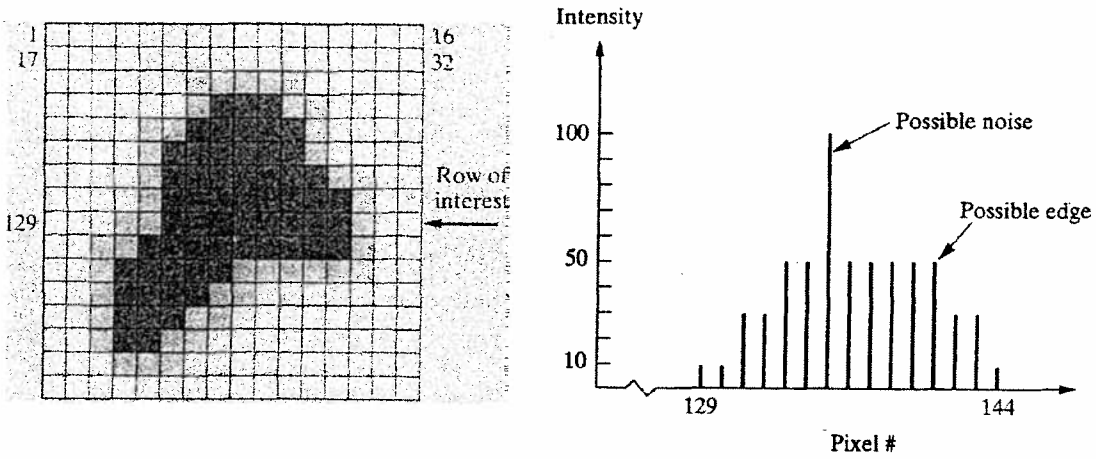


Figure 8.8 Noise and edge information in an intensity diagram of an image. The pixels with intensities that are much different from the intensities of neighboring pixels can be considered to be edges or noise.

signal — the filter will reduce the influence of all high frequencies, including the noises and edges. This means that, although a **low-pass filter** will reduce noises, it will also reduce the clarity of an image by **attenuating the edges**, thus softening the image throughout. A **high-pass filter**, on the other hand, will increase the apparent effect of higher frequencies by **severely attenuating the low-frequency** amplitudes. In such cases, noises and edges will be left alone, but **slowly changing areas will disappear** from the image.

To see how the Fourier transform can be applied in this case, let's look at the data of Figure 8.8 once again. The grayness level of the pixels of **row 9** is repeated in Figure 8.9(a). A simple first-approximation Fourier transform of the gray values [4] was performed for the first four harmonic frequencies, and then the signal was reconstructed, as shown in Figure 8.9(b). Comparing the two graphs reveals that a digital, discrete signal can be reconstructed, even if its accuracy is dependent on the number of sines and cosines, as well as the method of integration, etc.

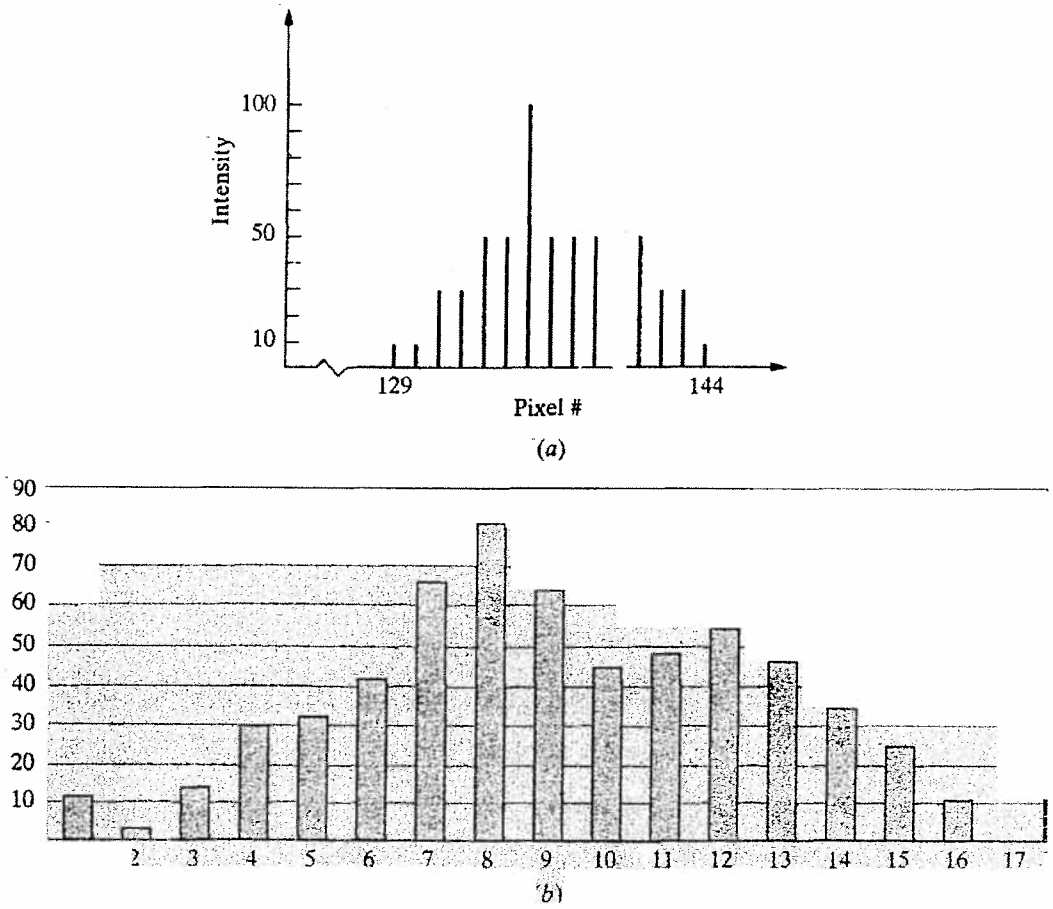


Figure 8.9 (a) Signal. (b) Discrete signal reconstructed from the Fourier transform of the signal in (a), using only four of the first frequencies in the spectrum.

Figure 8.9 (a) Signal. (b) Discrete signal reconstructed from the Fourier transform of the signal in (a), using only four of the first frequencies in the spectrum.