1. IMAGE SENSING AND ACQUISITION

Most of the images in which we are interested are generated by the combination of an "illumination" source and the reflection or absorption of energy from that source by the elements of the "scene" being imaged. We enclose *illumination* and *scene* in quotes to emphasize the fact that they are considerably more general than the familiar situation in which a visible light source illuminates a familiar 3-D scene. For example, the illumination may originate from a source of electromagnetic energy, such as a radar, infrared, or X-ray system. But, as noted earlier, it could originate from less traditional sources, such as ultrasound or even a computer-generated illumination pattern. Similarly, the scene elements could be familiar objects, but they can just as easily be molecules, buried rock formations, or a human brain. Depending on the nature of the source, illumination energy is reflected from, or transmitted through, objects. An example in the first category is light reflected from a planar surface. An example in the second category is when X-rays pass through a patient's body for the purpose of generating a diagnostic X-ray image. In some applications, the reflected or transmitted energy is focused onto a photo converter (e.g., a phosphor screen) that converts the energy into visible light. Electron microscopy and some applications of gamma imaging use this approach.

Figure 1 shows the three principal sensor arrangements used to transform incident energy into digital images. The idea is simple: Incoming energy is transformed into a voltage by a combination of the input electrical power and sensor material that is responsive to the type of energy being detected. The output voltage waveform is the response of the sensor, and a digital quantity is obtained by digitizing that response. In this section, we look at the principal modalities for image sensing and generation.

IMAGE ACQUISITION USING A SINGLE SENSING ELEMENT

Figure 1(a) shows the components of a single sensing element. A familiar sensor of this type is the photodiode, which is constructed of silicon materials and whose output is a voltage proportional to light intensity. Using a filter in front of a sensor improves its selectivity. For example, an optical green-transmission filter favors light in the green band of the color spectrum. As a consequence, the sensor output would be stronger for green light than for other visible light components.

In order to generate a 2-D image using a single sensing element, there has to be relative displacements in both the x- and y-directions between the sensor and the area to be imaged. Figure 2 shows an arrangement used in high-precision scanning, where a film negative is mounted onto a drum whose mechanical rotation provides displacement in one dimension. The sensor is mounted on a lead screw that provides motion in the perpendicular direction. A light source is contained inside the drum. As the light passes through the film, its intensity is modified by the film density before it is captured by the sensor. This "modulation" of the light intensity causes corresponding variations in the sensor voltage, which are ultimately converted to image intensity levels by digitization.

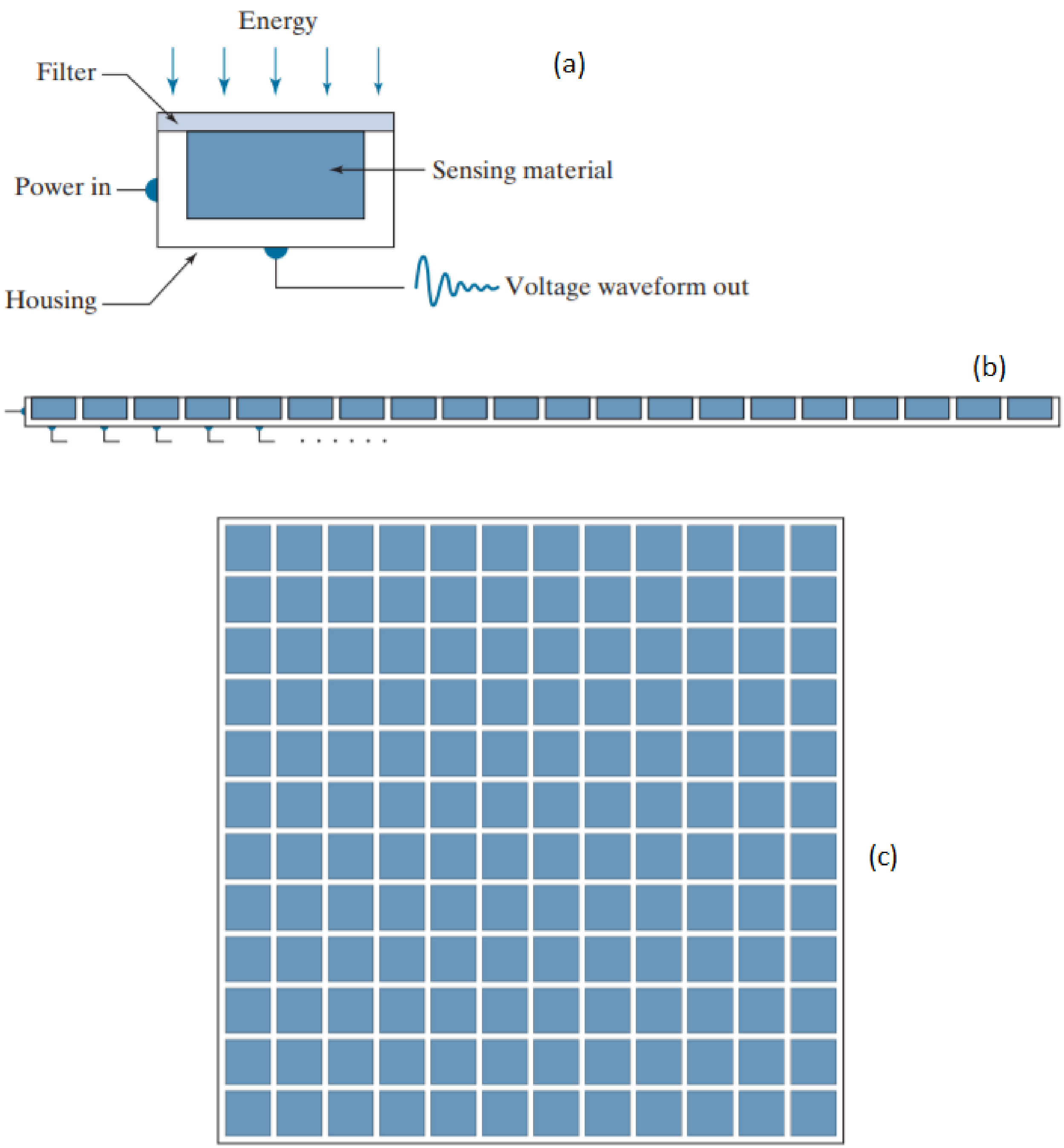

Figure 1: (a) Single sensing element. (b) Line sensor. (c) Array sensor.

This method is an inexpensive way to obtain high-resolution images because mechanical motion can be controlled with high precision. The main disadvantages of this method are that it is slow and not readily portable. Other similar mechanical arrangements use a flat imaging bed, with the sensor moving in two linear directions. These types of mechanical digitizers sometimes are referred to as *transmission microdensitometers*. Systems in which light is reflected from the medium, instead of passing through it, are called *reflection microdensitometers*. Another example of imaging with a single sensing element places a laser source coincident with the sensor. Moving mirrors are used to control the outgoing beam in a scanning pattern and to direct the reflected laser signal onto the sensor.

IMAGE ACQUISITION USING SENSOR STRIPS

A geometry used more frequently than single sensors is an in-line sensor strip, as in Fig. 1(b). The strip provides imaging elements in one direction. Motion perpendicular to the strip provides imaging in the other direction, as shown in Fig. 1(a). This arrangement is used in most flat bed scanners. Sensing devices with 4000 or more in-line sensors are possible. In-line sensors are used routinely in airborne imaging applications, in which the imaging system is mounted on an aircraft that flies at a constant altitude and speed over the geographical area to be imaged. One dimensional imaging sensor strips that respond to various bands of the electromagnetic spectrum are mounted perpendicular to the direction of flight. An imaging strip gives one line of an image at a time, and the motion of the strip relative to the scene completes the other dimension of a 2-D image. Lenses or other focusing schemes are used to project the area to be scanned onto the sensors.

Sensor strips in a ring configuration are used in medical and industrial imaging to obtain cross-sectional ("slice") images of 3-D objects, as Fig. 3(b) shows. A rotating X-ray source provides illumination, and X-ray sensitive sensors opposite the source collect the energy that passes through the object. This is the basis for medical and industrial computerized axial tomography (CAT) imaging. The output of the sensors is processed by reconstruction algorithms whose objective is to transform the sensed data into meaningful cross sectional images. In other words, images are not obtained directly from the sensors by motion alone; they also require extensive computer processing. A 3-D digital volume consisting of stacked images is generated as the object is moved in a direction perpendicular to the sensor ring. Other modalities of imaging based on the CAT principle include magnetic resonance imaging (MRI) and positron emission tomography (PET). The illumination sources, sensors, and types of images are different, but conceptually their applications are very similar to the basic imaging approach shown in Fig. 3(b).

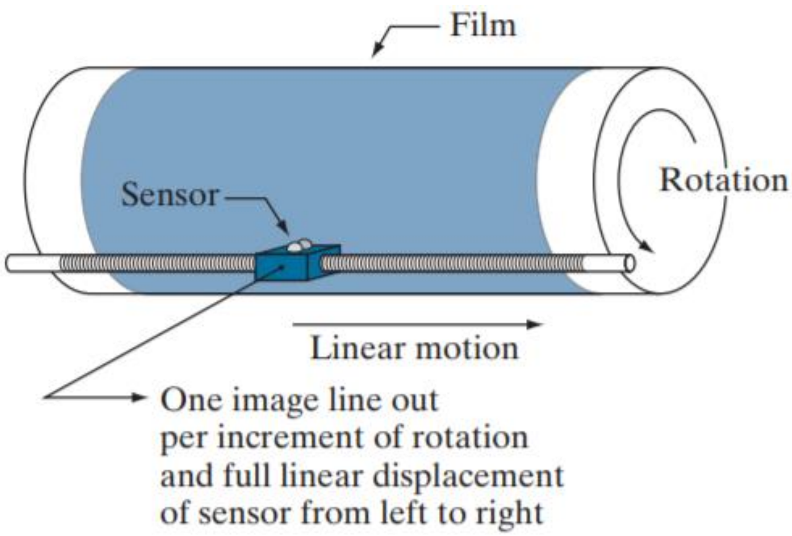

Figure 2: Combining a single sensing element with mechanical motion to generate a 2-D image.

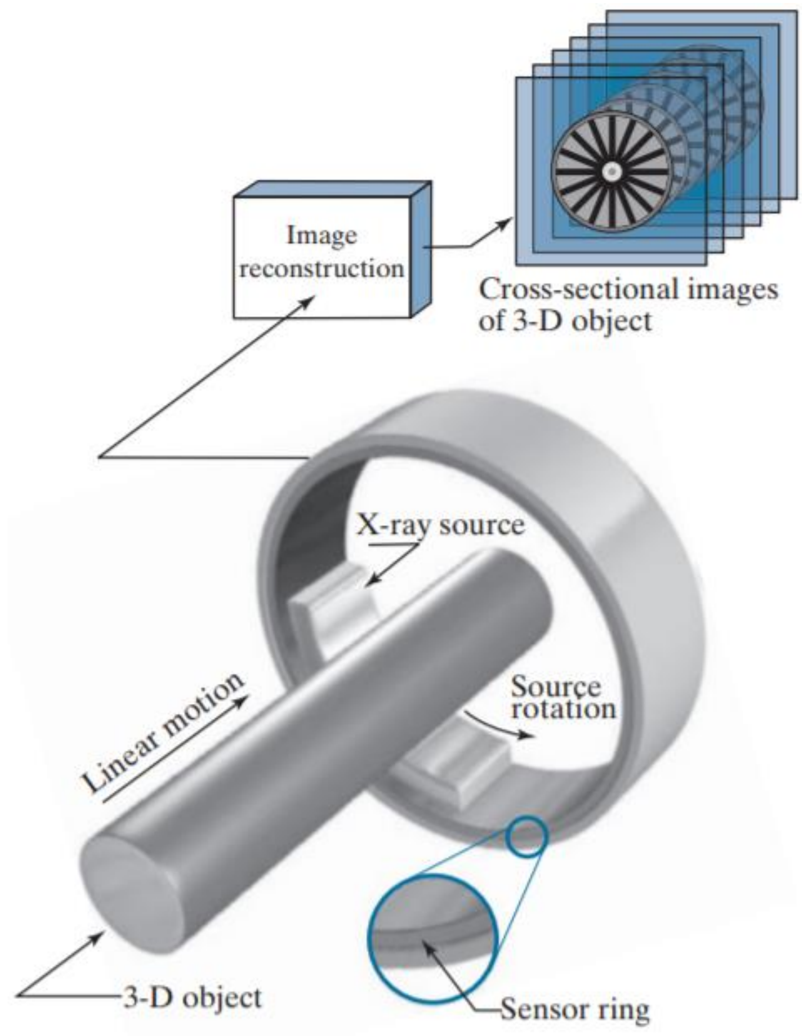

Figure 3: (a) Image acquisition using a linear sensor strip. (b) Image acquisition using a circular sensor strip.

IMAGE ACQUISITION USING SENSOR ARRAYS

Figure 1(c) shows individual sensing elements arranged in the form of a 2-D array. Electromagnetic and ultrasonic sensing devices frequently are arranged in this manner. This is also the predominant arrangement found in digital cameras. A typical sensor for these cameras is a CCD (charge-coupled device) array, which can be manufactured with a broad range of sensing properties and can be packaged in rugged arrays of 4000×4000 elements or more. CCD sensors are used widely in digital cameras and other light-sensing instruments. The response of each sensor is proportional to the integral of the light energy projected onto the surface of the sensor, a property that is used in astronomical and other applications requiring low noise images. Noise reduction is achieved by letting the sensor integrate the input light signal over minutes or even hours. Because the sensor array in Fig. 1(c) is two dimensional, its key advantage is that a complete image can be obtained by focusing the energy pattern onto the surface of the array. Motion obviously is not necessary, as is the case with the sensor arrangements discussed in the preceding two sections. Figure 4 shows the principal manner in which array sensors are used. This figure shows the energy from an illumination source being reflected from a scene (as mentioned at the beginning of this section, the energy also could be transmitted through the scene). The first function performed by the imaging system in Fig. 4(c) is to collect the incoming energy and focus it onto an image plane. If the illumination is light, the front end of the imaging system is an optical lens that projects the viewed scene onto the focal plane of the lens, as Fig. 4(d) shows. The sensor array, which is coincident with the focal plane, produces outputs proportional to the integral of the light received at each sensor. Digital and analog circuitry sweep these outputs and convert them to an analog signal, which is then digitized by another section of the imaging system. The output is a digital image, as shown diagrammatically in Fig. 4(e).

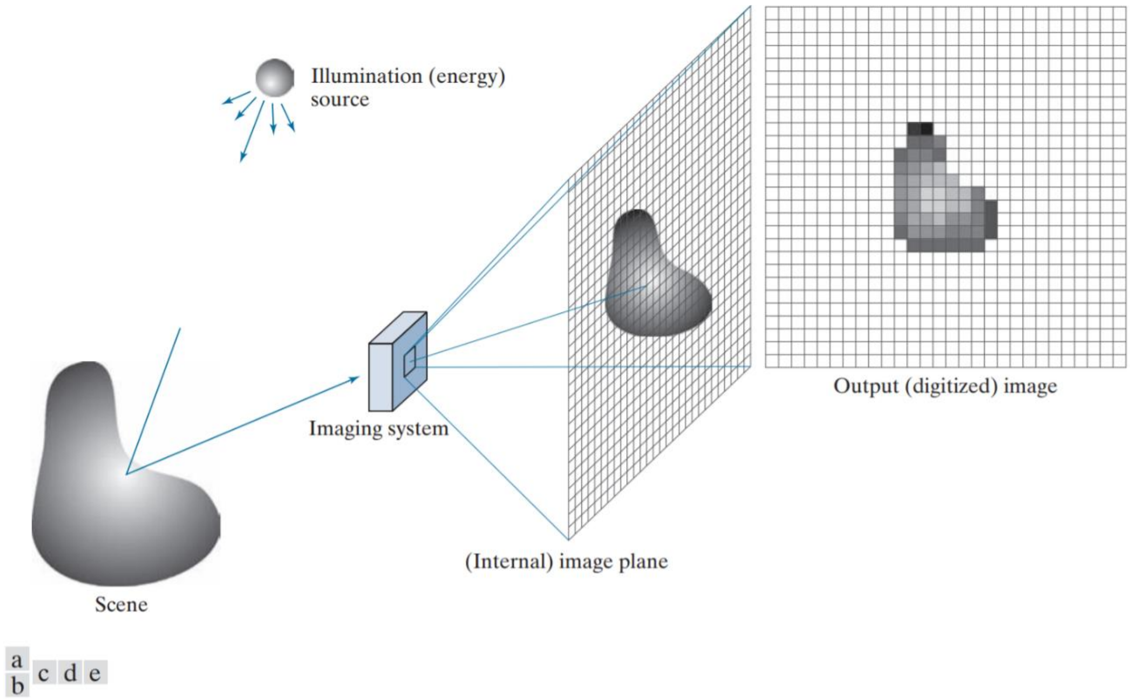

Figure 4: An example of digital image acquisition. (a) Illumination (energy) source. (b) A scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

2. A SIMPLE IMAGE FORMATION MODEL

We denote images by two-dimensional functions of the form $f(x, y)$. The value of f at spatial coordinates (x, y) is a scalar quantity whose physical meaning is determined by the source of the image, and whose values are proportional to energy radiated by a physical source (e.g., electromagnetic waves). As a consequence, $f(x, y)$ must be nonnegative† and finite; that is,

$$0 \leq f(x, y) < \infty$$

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

$$f(x, y) = i(x, y)r(x, y) \tag{1}$$

where

$$0 \leq i(x, y) < \infty \tag{2}$$

and

$$0 \leq r(x\,,y\,) \leq 1 \tag{3}$$

Thus, reflectance is bounded by 0 (total absorption) and 1 (total reflectance). The nature of $i(x\,,y\,)$ is determined by the illumination source, and $r(x\,,y\,)$ is determined by the characteristics of the imaged objects. These expressions are applicable also to images formed via transmission of the illumination through a medium, such as a chest X-ray. In this case, we would deal with a transmissivity instead of a reflectivity function, but the limits would be the same as in Eq. (3), and the image function formed would be modeled as the product in Eq. (1).

3. IMAGE SAMPLING AND QUANTIZATION

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

BASIC CONCEPTS IN SAMPLING AND QUANTIZATION

Figure 1(a) shows a continuous image f that we want to convert to digital form. An image may be continuous with respect to the x- and y-coordinates, and also in amplitude. To digitize it, we have to sample the function in both coordinates and also in amplitude. Digitizing the coordinate values is called sampling. Digitizing the amplitude values is called quantization.

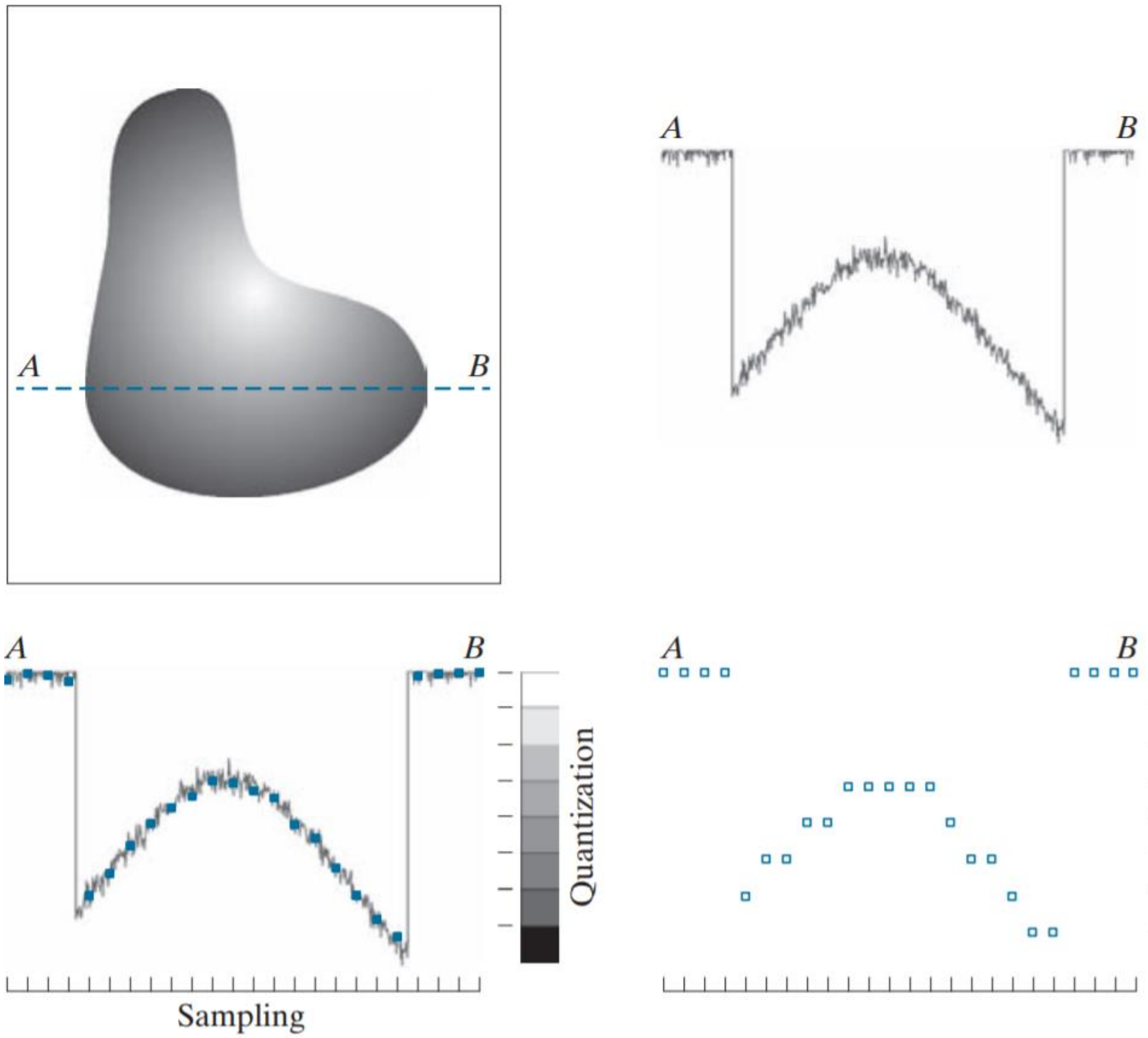

Figure 5: (a) Continuous image. (b) A scan line showing intensity variations along line AB in the continuous image. (c) Sampling and quantization. (d) Digital scan line. (The black border in (a) is included for clarity. It is not part of the image).

The one-dimensional function in Fig. 1(b) is a plot of amplitude (intensity level) values of the continuous image along the line segment AB in Fig. 1(a). The random variations are due to image noise. To sample this function, we take equally spaced samples along line AB, as shown in Fig. 1(c). The samples are shown as small dark squares superimposed on the function, and their (discrete) spatial locations are indicated by corresponding tick marks in the bottom of the figure. The set of dark squares constitute the $sampled$ function. However, the $values$ of the samples still span (vertically) a continuous range of intensity values. In order to form a digital function, the intensity values also must be converted ($quantized$) into $discrete$ quantities. The vertical gray bar in Fig. 1(c) depicts the intensity scale divided into eight discrete intervals, ranging from black to white. The vertical tick marks indicate the specific value assigned to each of the eight intensity intervals. The continuous intensity levels are quantized by assigning one of the eight values to each sample, depending on the vertical proximity of a sample to a vertical tick mark. The digital samples resulting from both sampling and quantization are shown as white squares in Fig. 1(d). Starting at the top of the continuous image and carrying out this procedure downward, line by line, produces a two-dimensional digital image. It is implied in Fig. 1 that, in addition to the number of discrete levels used, the accuracy achieved in quantization is highly dependent on the noise content of the sampled signal.

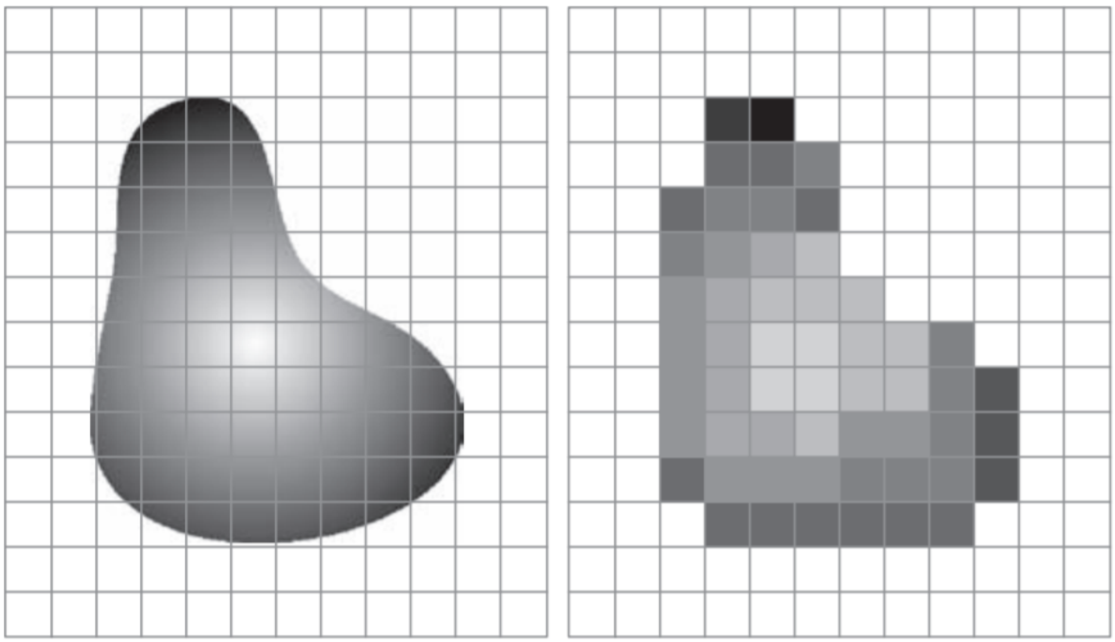

Figure 6: (a) Continuous image projected onto a sensor array. (b) Result of image sampling and quantization.

In practice, the method of sampling is determined by the sensor arrangement used to generate the image. When an image is generated by a single sensing element combined with mechanical motion, as in Fig. 2.13, the output of the sensor is quantized in the manner described above. However, spatial sampling is accomplished by selecting the number of individual mechanical increments at which we activate the sensor to collect data. Mechanical motion can be very exact so, in principle, there is almost no limit on how fine we can sample an image using this approach. In practice, limits on sampling accuracy are determined by other factors, such as the quality of the optical components used in the system.

When a sensing strip is used for image acquisition, the number of sensors in the strip establishes the samples in the resulting image in one direction, and mechanical motion establishes the number of samples in the other. Quantization of the sensor outputs completes the process of generating a digital image.

When a sensing array is used for image acquisition, no motion is required. The number of sensors in the array establishes the limits of sampling in both directions. Quantization of the sensor outputs is as explained above. Figure 2 illustrates this concept. Figure 2(a) shows a continuous image projected onto the plane of a 2-D sensor. Figure 2(b) shows the image after sampling and quantization. The quality of a digital image is determined to a large degree by the number of samples and discrete intensity levels used in sampling and quantization. However, as we will show later in this section, image content also plays a role in the choice of these parameters.

4. REPRESENTING DIGITAL IMAGES

Let $f(s, t)$ represent a continuous image function of two continuous variables, s and t. We convert this function into a *digital image* by sampling and quantization, as explained in the previous section. Suppose that we sample the continuous image into a digital image, $f(x, y)$, containing M rows and N columns, where (x, y) are discrete coordinates. For notational clarity and convenience, we use integer values for these discrete coordinates: $x = 0, 1, 2, \dots M - 1$ and $y = 0, 1, 2, \dots N - 1$. Thus, for example, the value of the digital image at the origin is $f(0, 0)$, and its value at the next coordinates along the first row is $f(0, 1)$. Here, the notation $(0, 1)$ is used to denote the second sample along the first row. It does not mean that these are the values of the physical coordinates when the image was sampled. In general, the value of a digital image at any coordinates (x, y) is denoted $f(x, y)$, where x and y are integers. When we need to refer to specific coordinates (i, j), we use the notation $f(i, j)$, where the arguments are integers. The section of the real plane spanned by the coordinates of an image is called the *spatial domain*, with x and y being referred to as *spatial variables* or *spatial coordinates*.

Figure 7 shows three ways of representing $f(x, y)$. Figure 7(a) is a plot of the function, with two axes determining spatial location and the third axis being the values of f as a function of x and y. This representation is useful when working with grayscale sets whose elements are expressed as triplets of the form (x, y, z), where x and y are spatial coordinates and z is the value of f at coordinates (x, y).

The representation in Fig. 7(b) is more common, and it shows $f(x, y)$ as it would appear on a computer display or photograph. Here, the intensity of each point in the display is proportional to the value of f at that point. In this figure, there are only three equally spaced intensity values. If the intensity is normalized to the interval $[0, 1]$, then each point in the image has the value 0, 0.5, or 1. A monitor or printer converts these three values to black, gray, or white, respectively, as in Fig. 7(b). This type of representation includes color images, and allows us to view results at a glance.

As Fig. 7(c) shows, the third representation is an array (matrix) composed of the numerical values of $f(x, y)$. This is the representation used for computer processing. In equation form, we write the representation of an $M \times N$ numerical array as

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & \begin{matrix} f(0,N-1) \\ f(1,N-1) \end{matrix} \\ f(1,0) & f(1,1) & \dots & \vdots \\ \vdots & \vdots & & \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix}$$

The right side of this equation is a digital image represented as an array of real numbers. Each element of this array is called an image element, picture element, pixel, or pel.

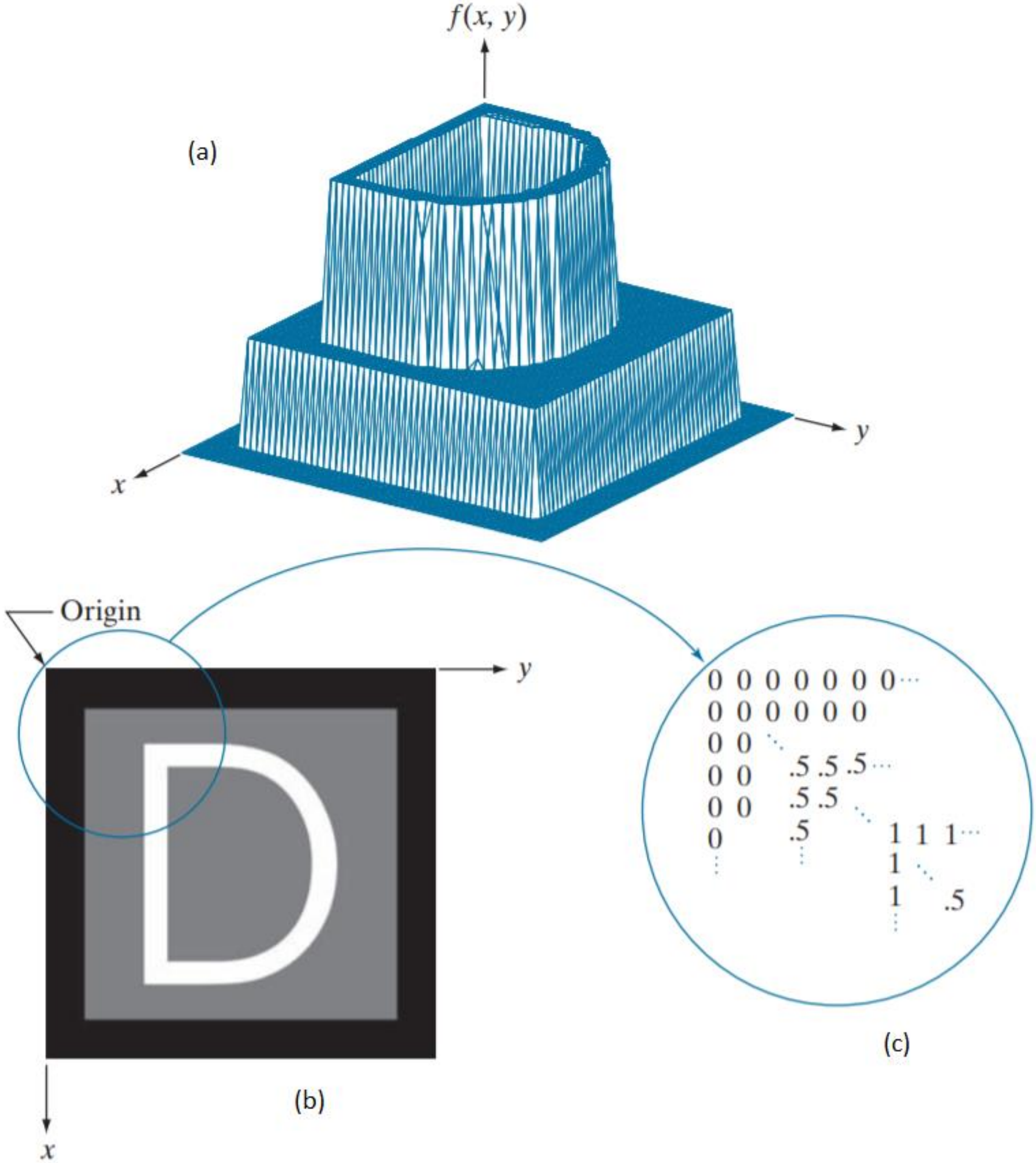

Figure 7: (a) Image plotted as a surface. (b) Image displayed as a visual intensity array. (c) Image shown as a 2-D numerical array. (The numbers 0, .5, and 1 represent black, gray, and white, respectively.)

The center of an $M \times N$ digital image with origin at $(0, 0)$ and range to $(M-1, N-1)$ is obtained by dividing M and N by 2 and rounding down to the nearest integer. This operation sometimes is denoted using the floor operator, $\lfloor\cdot\rfloor$. This holds true for M and N even or odd. For example, the center of an image of size 1023×1024 is at (511, 512). Some programming languages (e.g., MATLAB) start indexing at 1 instead of at 0. The center of an image in that case is found at

$$(x_c, y_c) = (floor\left(\frac{M}{2}\right) + 1, floor\left(\frac{N}{2}\right) + 1).$$

References and further reading:

Digital Image Processing, 4th edition, Gonzalez, Rafael and Woods, Richard, 2018