Digital Image Fundamentals

Prof. George Wolberg
Dept. of Computer Science
City College of New York

Objectives

- In this lecture we discuss:
  - Image formation
  - Image acquisition
  - Sampling and quantization
  - Spatial and graylevel resolution

Pinhole Camera

- To form a recognizable image, the light rays must be constrained.
  - One way to do this is to construct an empty, opaque box that is so tight that no light can enter the box.
  - Then, a small hole the size of a pin is pierced into one side of the box, which allows light to enter the box only through the hole.

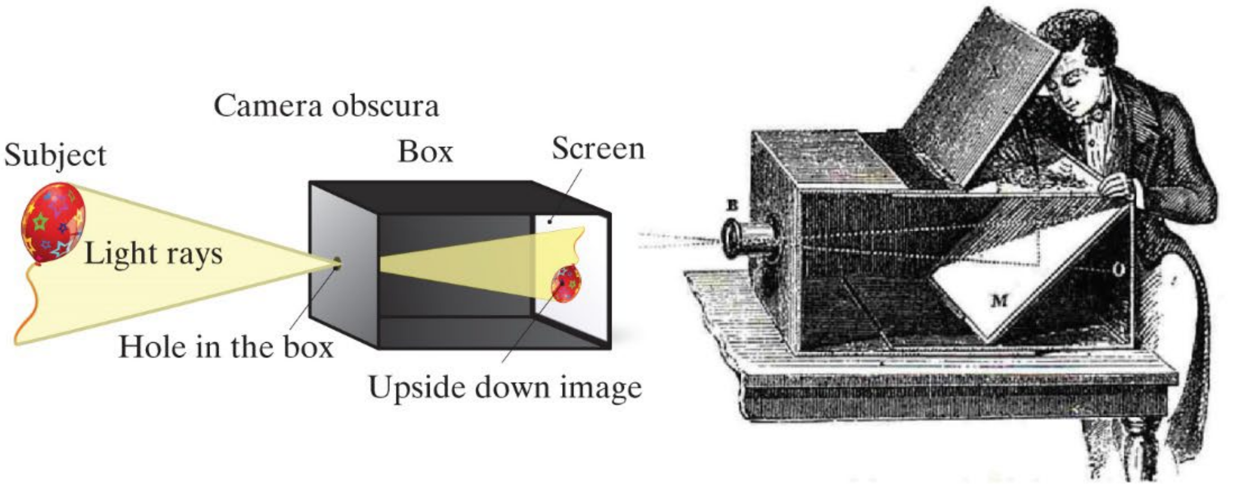


Figure 2.22 In a pinhole camera, light rays pass through the tiny aperture and form an upside-down image on the opposite wall. A camera obscura was an early form of pinhole camera in which light rays pass through the small aperture, reflect off the mirror, and form an image on the top horizontal surface near the rear of the enclosed box.

Pinhole Camera Properties

- **Focal point:** the pinhole through which all rays of light pass.
- **Image plane:** the sensor surface on which the image is formed.
- **Optical axis:** the line through the focal point perpendicular to the image plane.
- **Focal length:** the distance from the focal point to the image plane along this line.

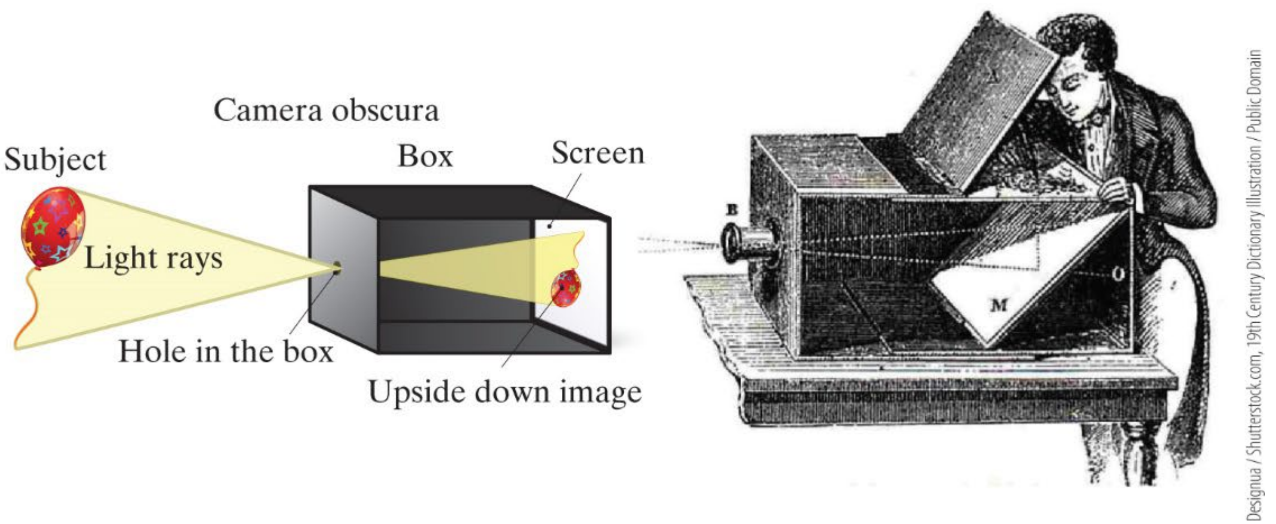
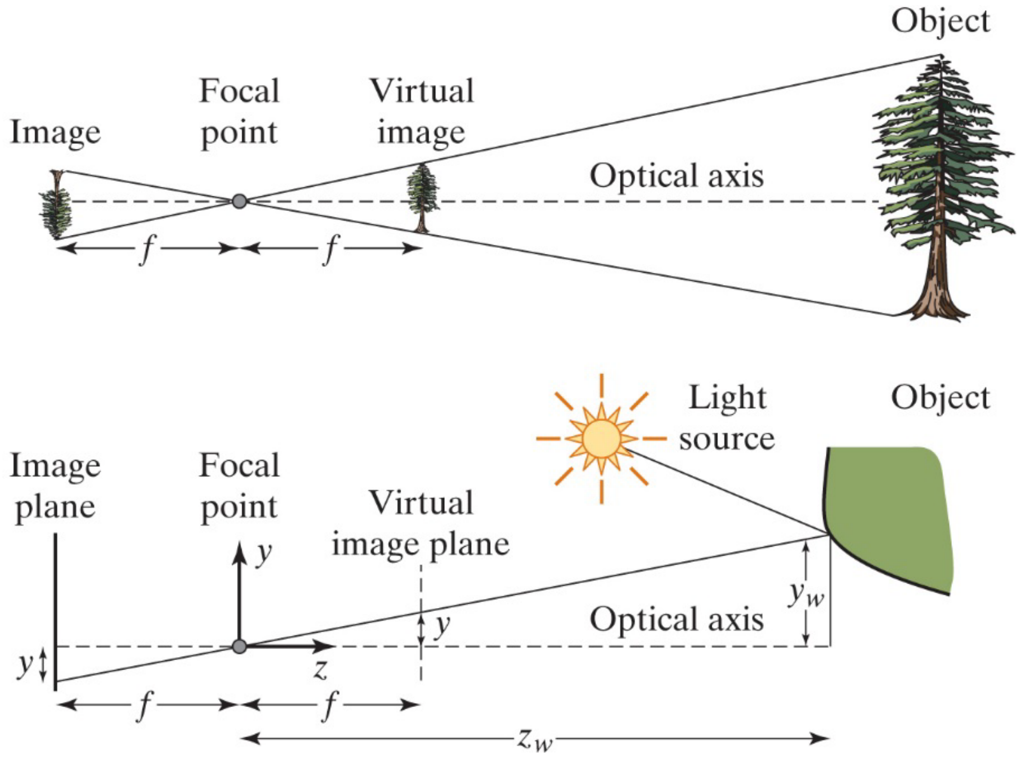


Figure 2.22 In a pinhole camera, light rays pass through the tiny aperture and form an upside-down image on the opposite wall. A camera obscura was an early form of pinhole camera in which light rays pass through the small aperture, reflect off the mirror, and form an image on the top horizontal surface near the rear of the enclosed box.

Perspective Projection

Light rays from the source reflect off the surface of an object in the scene, travel through the focal point (also called the center of projection), then land on the image plane, which may be film, the retina, or CCD.

Figure 2.23 Perspective projection caused by a pinhole camera, showing the focal point (pinhole), image plane, focal length, and optical axis. The light rays emitted by the light source reflect off the surface in the world and pass through the aperture to form an upside-down image on the image plane. This is mathematically equivalent to producing a rightside-up image on the virtual image plane in front of the focal point.

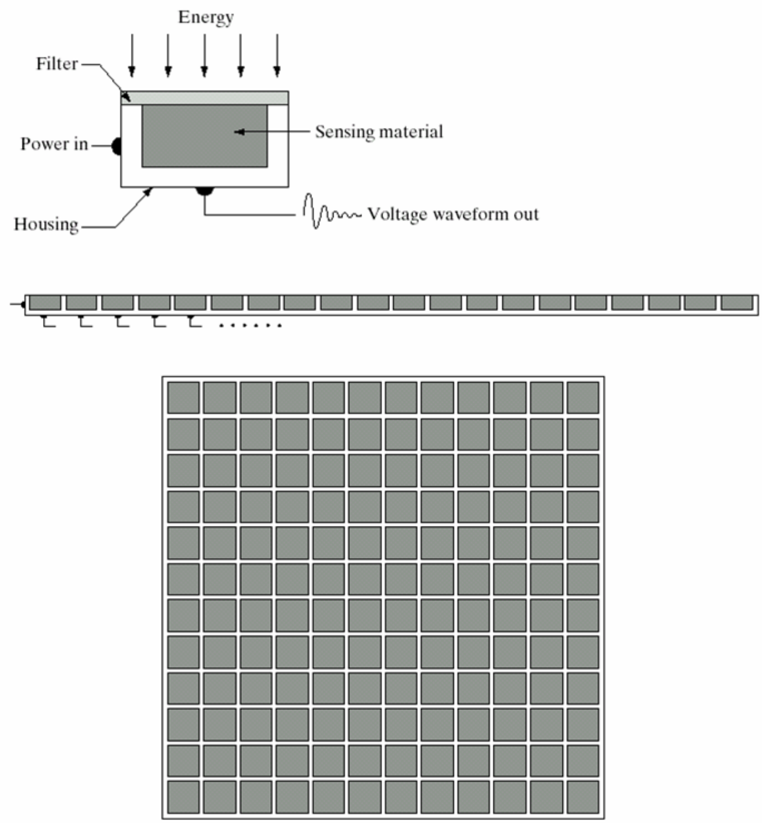


Analog vs Digital Images

- Up to now, the images acquired by the camera formed analog (continuous) images such as photographs recorded on film or landing directly on our retina from the 3D world.
- We shall be interested in converting these images into finite precision samples to form a digital image.
- A digital image captured by a camera is usually a digitized version of some two-dimensional sensory input.

Sensor Arrangements

- Three principal sensor arrangements:
  - Single, line, and array



a
b
c

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

Single Sensor

- Photodiode: constructed of silicon materials whose output voltage waveform is proportional to light.
- To generate a 2D image using a single sensor, there must be relative displacements in the horizontal and vertical directions between the sensor and the area to be imaged.
- Microdensitometers: mechanical digitizers that use a flat bed with the single sensor moving in two linear directions.

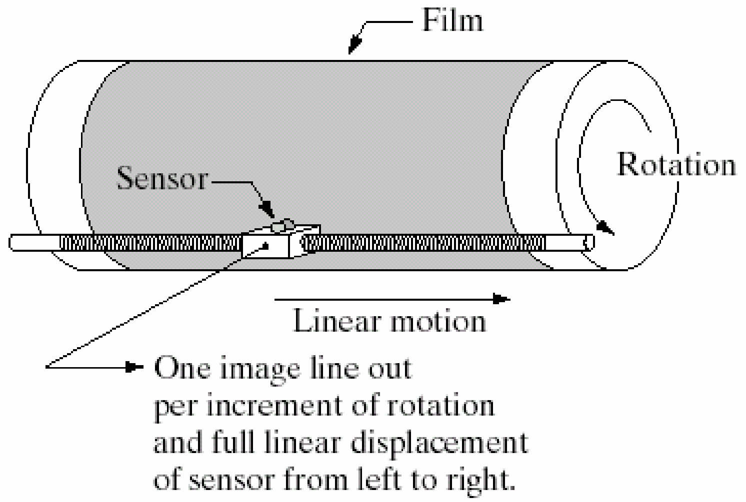


FIGURE 2.13 Combining a single sensor with motion to generate a 2-D image.

Sensor Strips

- In-line arrangement of sensors in the form of a sensor strip.
- The strip provides imaging elements in one direction.
- Motion perpendicular to strip images in the other direction.
- Used in flat bed scanners, with 4000 or more in-line sensors.

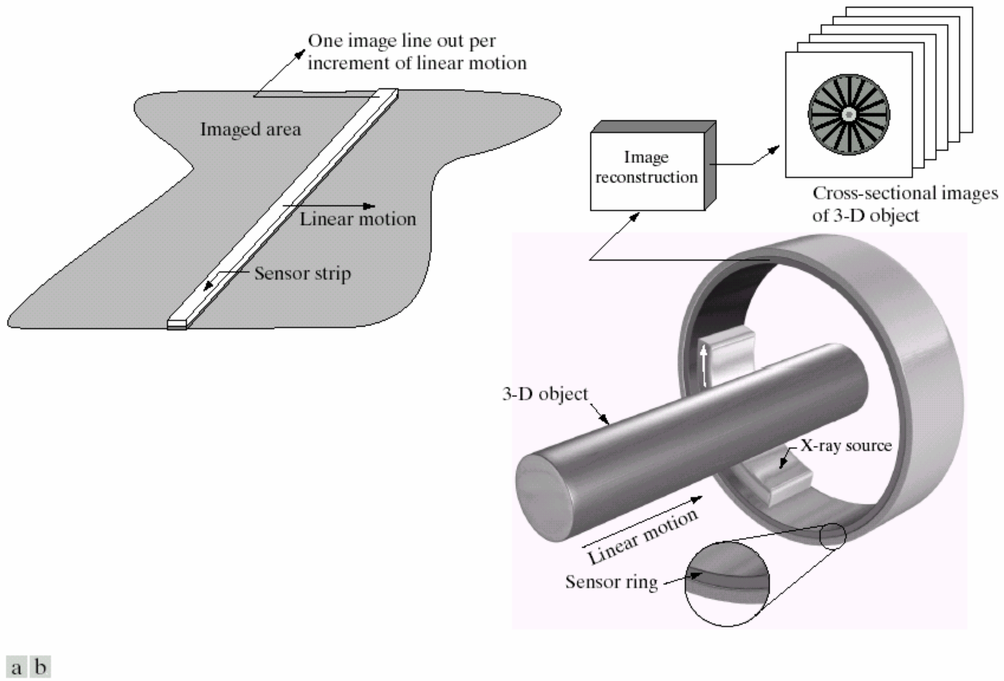


FIGURE 2.14 (a) Image acquisition using a linear sensor strip. (b) Image acquisition using a circular sensor strip.

Sensor Arrays

- Individual sensors are arranged in the form of a 2D array.
- Used in digital cameras and camcorders.
- Entire image formed at once; no motion necessary.

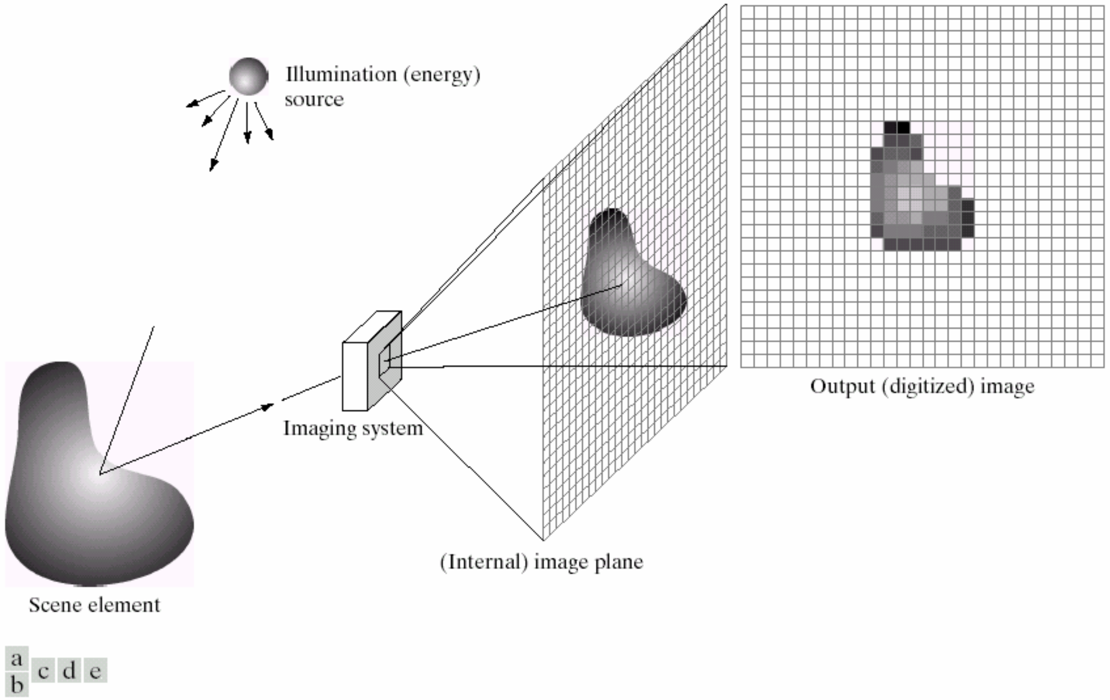


FIGURE 2.15 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

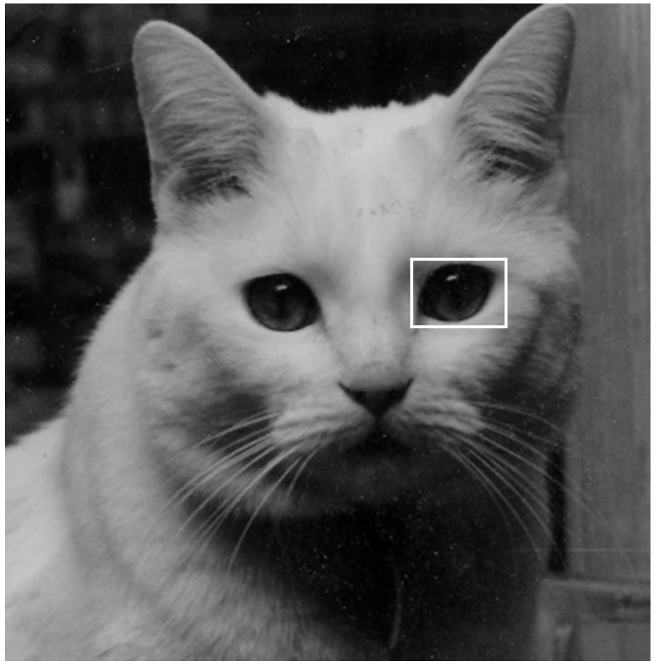
Signals

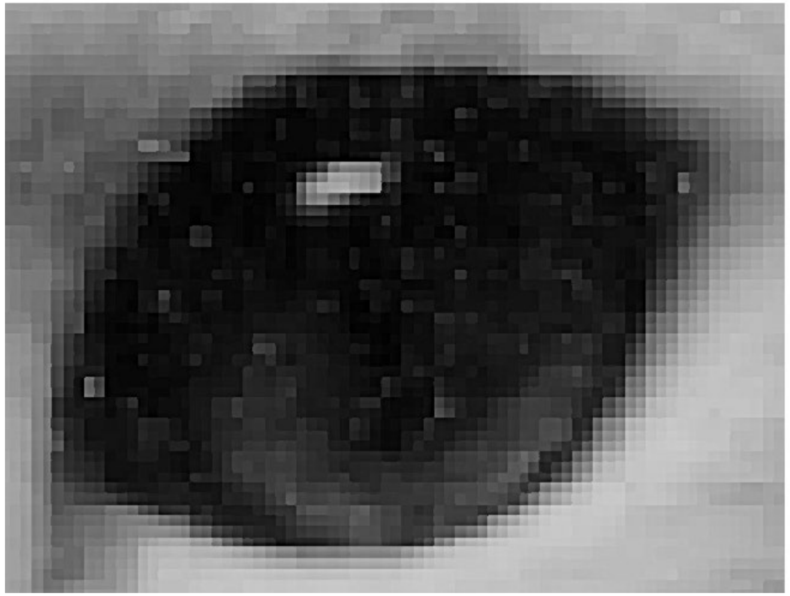
- A signal is a function that conveys information
  - 1D signal: f(x) waveform
  - 2D signal: f(x,y) image
  - 3D signal: f(x,y,z) volumetric data
    or f(x,y,t) animation (spatiotemporal volume)
  - 4D signal: f(x,y,z,t) snapshots of volumetric data over time

- The dimension of the signal is equal to its number of indices.
- In this course, we focus on 2D images: f(x,y)
- Efficient implementation often calls for 1D row or column processing. That is, process the rows independently and then process the columns of the resulting image.

Digital Image

- A digital image is produced as an array (the *raster*) of picture elements (*pixels* or *pels*) in the *frame buffer.*





Accessing Image Data (1)

- **Digital image:** a discrete two-dimensional array of values, like a matrix.
  - *Width:* refers to the number of columns in the image.
  - *Height:* refers to the number of rows in the image.
- **Aspect ratio:** width divided by height.
- **Pixel:** Each element of the array.
- An image is stored in memory as a 1D array in two ways:
  - *Column major order:* column 1 is stored, then column 2, and so on until the final column.
  - *Row major order:* row 1 is stored, then row 2, and so on until the final row.

Accessing Image Data (2)

Figure 1.5: Top: Image as a 2D array, showing the 1D index of each pixel. Bottom: Internal representation of image as a 1D array using row major order.

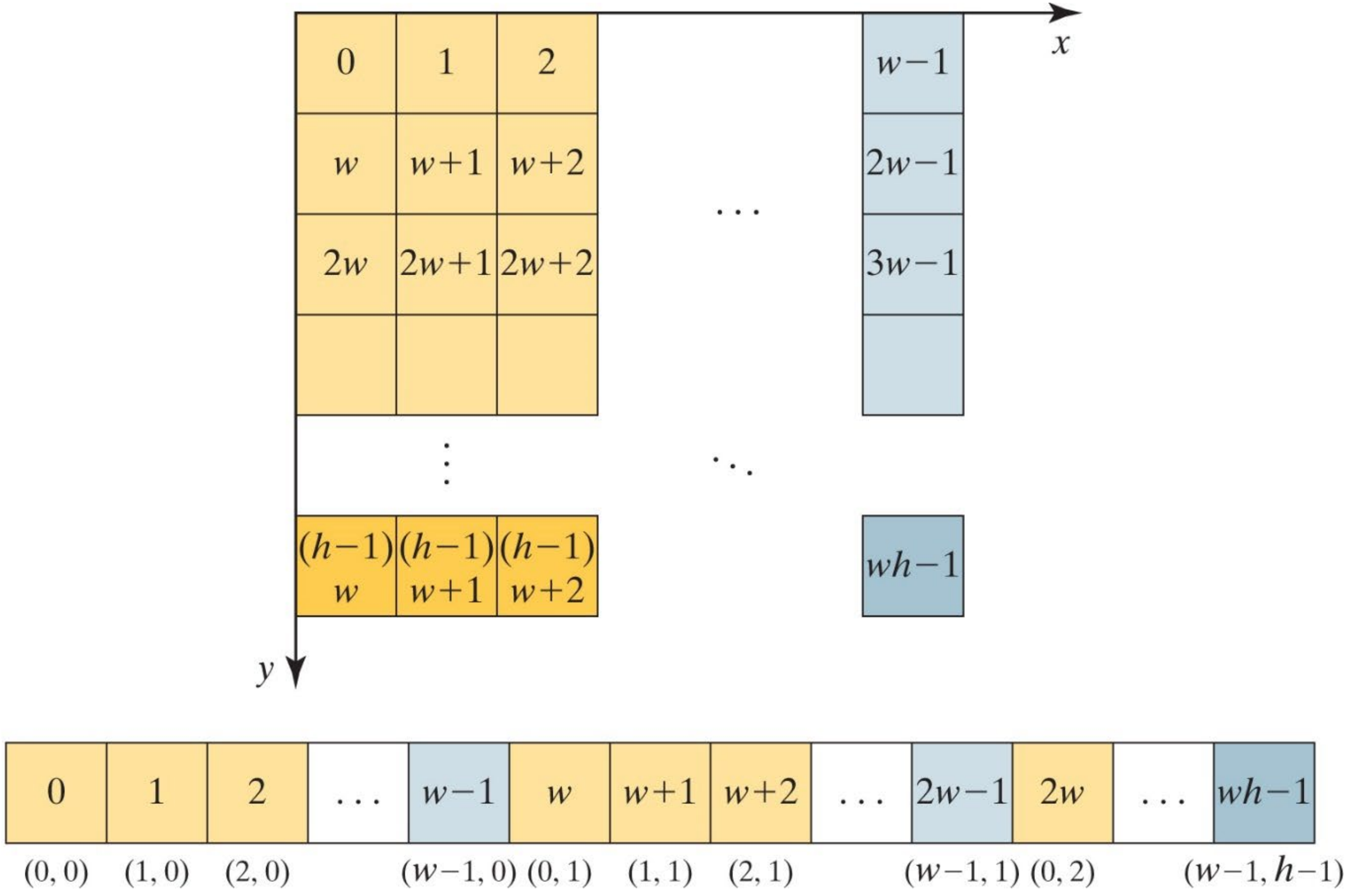


Image Types (1)

- **Grayscale image**: the value of each pixel is a scalar indicating the amount of light captured.
  - These values are quantized into a finite number of discrete levels called **gray levels**.

- In an 8-bit grayscale image, a pixel whose value is 0 represents black, whereas a pixel whose value is 255 represents white. In general, the highest intensity value for a pixel having *b* bits is 2^b - 1.

- **RGB color image**: the pixel values are triples containing the amount of light captured in the three color channels: red, green, and blue.
  - **Interleaved:** all three values for one pixel are stored before the three values of the next pixel
  - **Planar:** the red, green, and blue channels are stored as separate one-byte-per-pixel images
  - **Alpha value or opacity:** used for blending multiple images
  - The total number of colors in a color image is 2^b x 2^b x 2^b = 2^{3b}. For $b = 8$, this yields 2^{24} (16M colors)

Image Types (2)

- **Binary image:** The logical values can be stored using one bit per pixel, (0 for off or 1 for on), or they can be stored using one byte per pixel, where their values are usually 0 (hexadecimal 00) or 255 (hexadecimal FF).

- **Real-valued image**, or **floating-point image:** each pixel contains a real number.
  - The number is stored in the computer as an IEEE single- or double-precision floating point number

- **Integer-valued image:** the value of each pixel is an integer.

Image Types (3)

- **Channels:** For example, an RGB color image is an 8-bit image with three channels.
  - **Complex-valued image:** arises from computing the Fourier transform of an image.
    - It contains two floating-point values for each pixel, one for the real component and one for the imaginary component.

	grayscale	RGB color	binary	integer-valued	real-valued	complex-valued
channels	1	3	1	1	1	2
bit depth	8	24	1	32/64	32/64	64/128
value range	$\{0, \ldots, 255\}$	$\{0, \ldots, 255\}^3$	$\{0, 1\}$	$\mathbb{Z}$	$\mathbb{R}$	$\mathbb{R}^2$

TABLE 1.3: Common image types, shown with the number of channels, the most commonly encountered bit depth (number of bits per pixel), and the set of possible values. In the final three columns this set is conceptual only, since the integers $\mathbb{Z}$ and real numbers $\mathbb{R}$ are infinite sets.

Conceptualizing Images (1)

- A digital image is stored in the computer as a discrete array of values, which can be visualized either by:
  - considering the raw pixel values themselves arranged in a 2D lattice
  - equivalently as a height map, or 3D surface plot

- $I(x, y)$: evaluate the function at the position (x, y).

Conceptualizing Images (2)

- Set of pixels
  - The grayscale image: Example: $I = \begin{bmatrix} 3 & 8 & 0 \\ 2 & 9 & 4 \end{bmatrix}$
  - The binary image: Example: $I = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$
  - The following matrix **A** is an $m \times n$ matrix whose $(i, j)^{th}$ entry is given by a_{ij}

$$\mathbf{A}_{\{m \times n\}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Conceptualizing Images (3)

Figure 1.6: Different ways to visualize an image: as a picture, as a height map, as an array of values, as a function, as a set, as a graph, and as a vector. The 5 × 4 array is a small portion of the image; the set contains the coordinates of all pixels in the array whose value is greater than 80; and the weights of the edges in the graph are the absolute differences between values in the array.

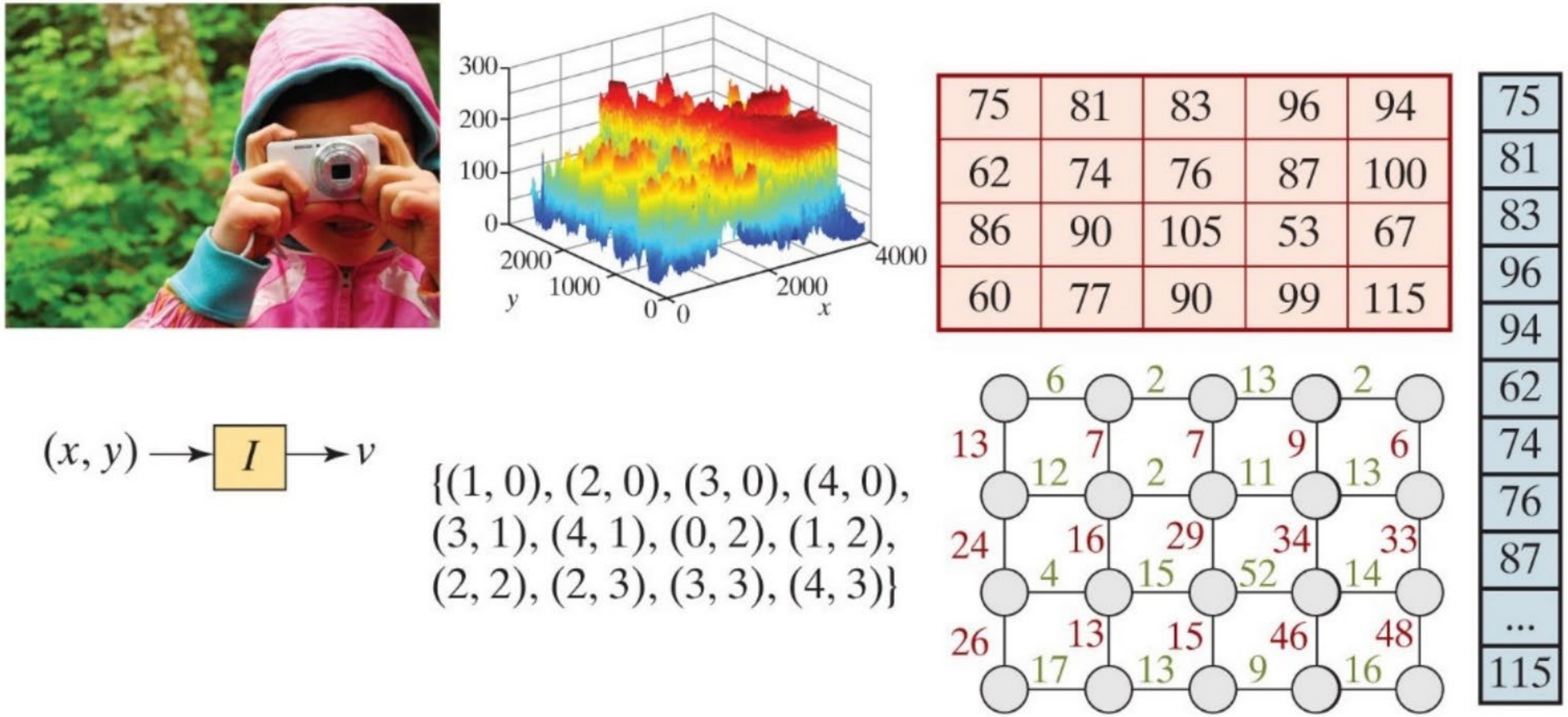


Image Classification (1)

- Images can be classified by whether they are defined over all points in the spatial domain and whether their image values have finite or infinite precision.
- If the position variables (x,y) are continuous, then the function is defined over all points in the spatial domain.
- If (x,y) is discrete, then the function can be sampled at only a finite set of points, i.e., the set of integers.
- The value that the function returns can also be classified by its precision, independently of x and y.

Image Classification (2)

- Quantization refers to the mapping of real numbers onto a finite set: a many-to-one mapping.
- Akin to casting from double precision to an integer.

Space	Image Values	Classification
continuous	continuous	analog (continuous) image
continuous	discrete	intensity quantization
discrete	continuous	spatial quantization
discrete	discrete	digital (discrete) image

Image Formation

- The values of an acquired image are always positive. There are no negative intensities: $0 < f(x,y) < \infty$
- Continuous function $f(x,y) = i(x,y)\, r(x,y)$, where

  $0 < i(x,y) < \infty$ is the illumination

  $0 < r(x,y) < 1$ is the reflectance of the object
- $r(x,y) = 0$ is total absorption and $r(x,y) = 1$ is total reflectance.
- Replace $r(x,y)$ with transmissivity term $t(x,y)$ for chest X-ray.

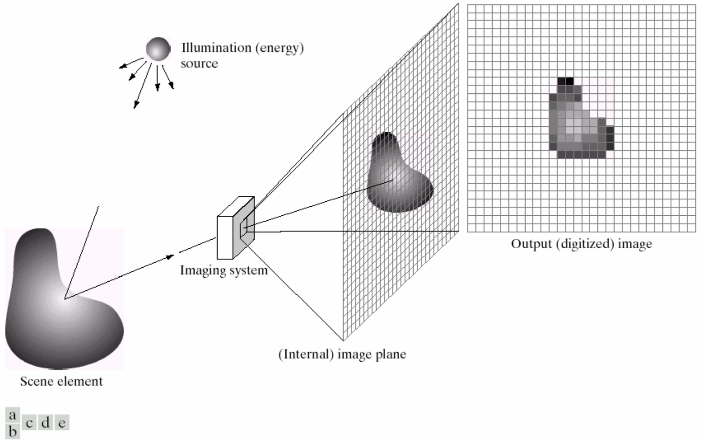
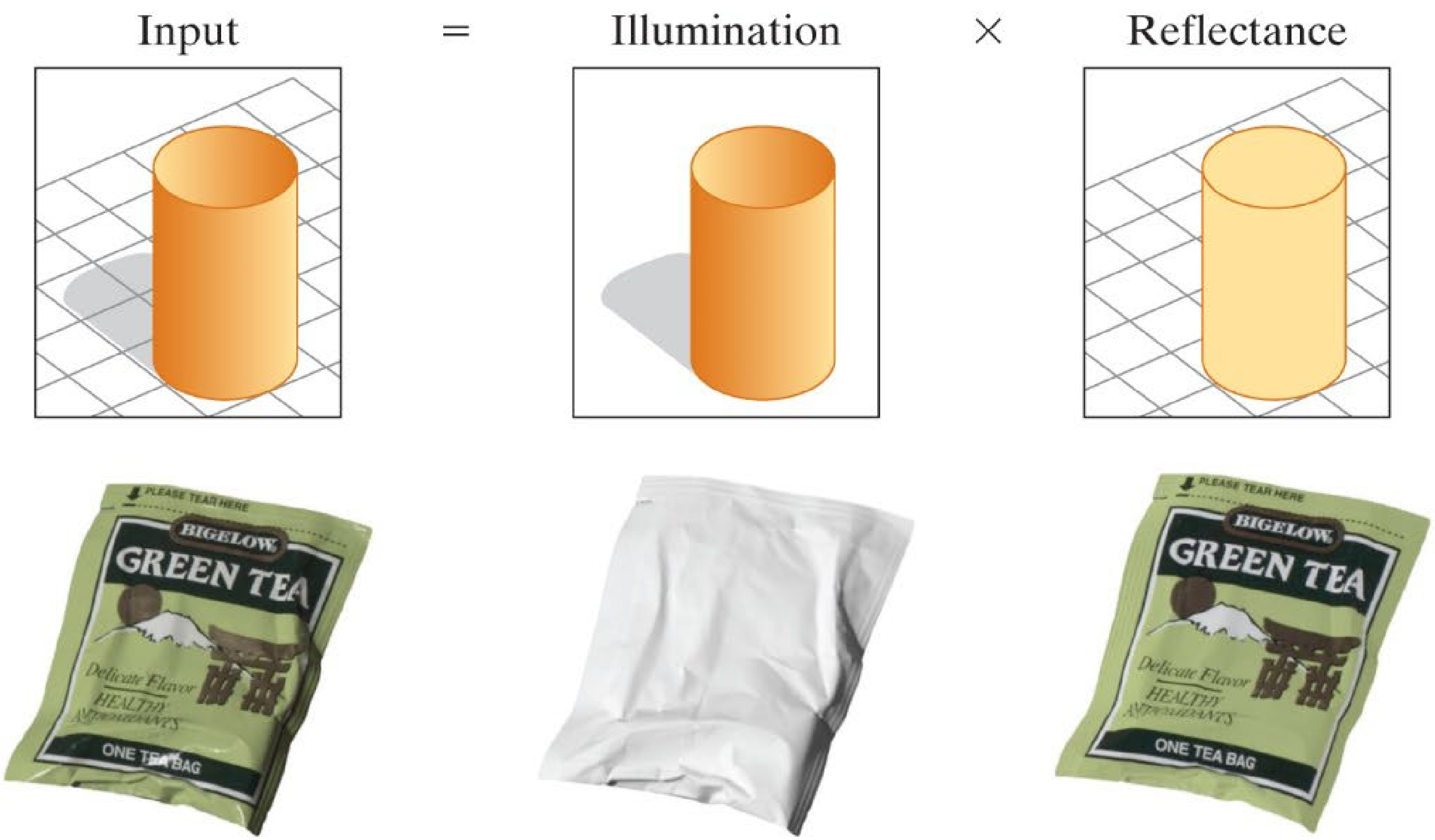


FIGURE 2.15 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

Simplified Imaging Model

Figure 2.28 Intrinsic images are a mid-level description of scenes determined by decomposing an image into constituent components, such as an illumination image and a reflectance image. Based on Y. Weiss, "Deriving intrinsic images from image sequences," *Proceedings of the International Conference on Computer Vision*, pages 68-75, July 2001.



Typical Values of Illumination and Reflectance

- The following i(x,y) illumination values are typical (in lumens/m^2):
  - Illumination of sun on Earth on a clear day: 90,000
  - Illumination of sun on Earth on a cloudy day: 10,000
  - Illumination of moon on Earth on a clear night: 0.1
  - Illumination level in a commercial office: 1000
  - Illumination level of video projectors: 1000-1500
- The following r(x,y) reflectance values are typical:
  - Black velvet: 0.01
  - Stainless steel: 0.65
  - Flat-white wall paint: 0.80
  - Silver-plated metal: 0.90
  - Snow: 0.93

Graylevels

- The intensity of a monochrome image at any coordinate *(x,y)* is called *graylevel L*, where Lmin ≤ *L* ≤ Lmax
- The office illumination example indicates that we may expect

  Lmin ≈ .01 * 1000 = 10 (virtual black)

  Lmax ≈ 1 * 1000 = 1000 (white)
- Interval [Lmin, Lmax] is called the *grayscale.*
- In practice, the interval is shifted to the [0, 255] range so that intensity can be represented in one byte (unsigned char).
- 0 is black, 255 is white, and all intermediate values are different shades of gray varying from black to white.

Generating a Digital Image

- Sample and quantize continuous input image.

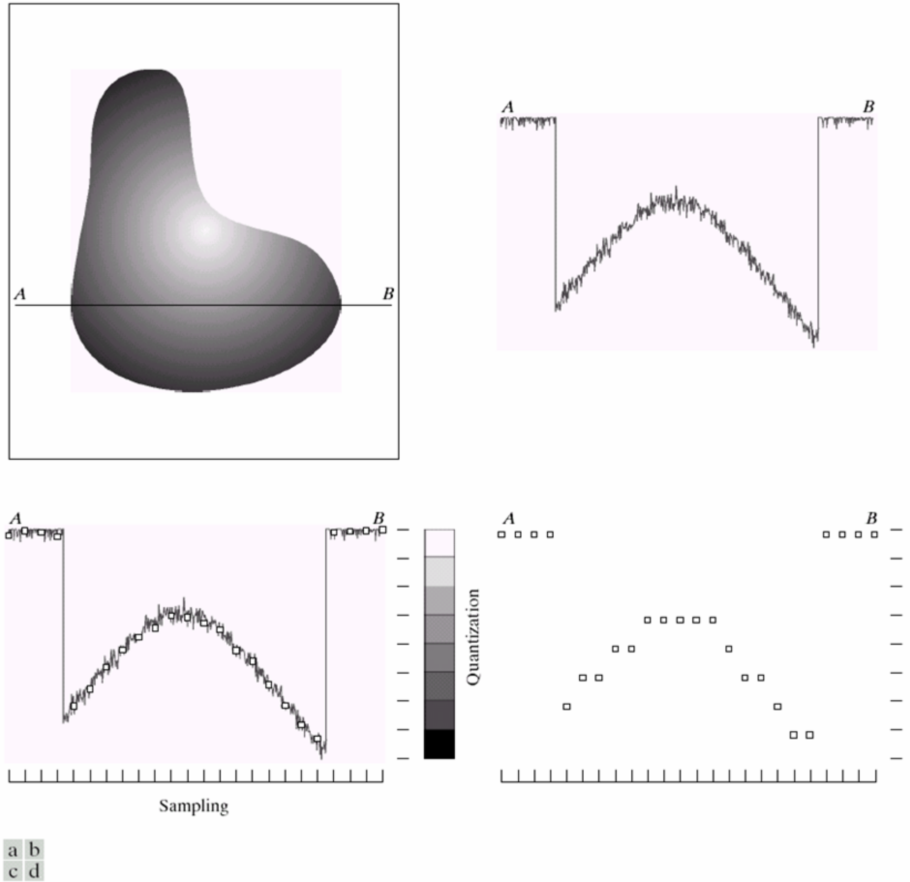


FIGURE 2.16 Generating a digital image. (a) Continuous image. (b) A scan line from A to B in the continuous image, used to illustrate the concepts of sampling and quantization. (c) Sampling and quantization. (d) Digital scan line.

Image Sampling and Quantization

- Sampling: digitize (discretize) spatial coordinate (x,y)
- Quantization: digitize intensity level *L*

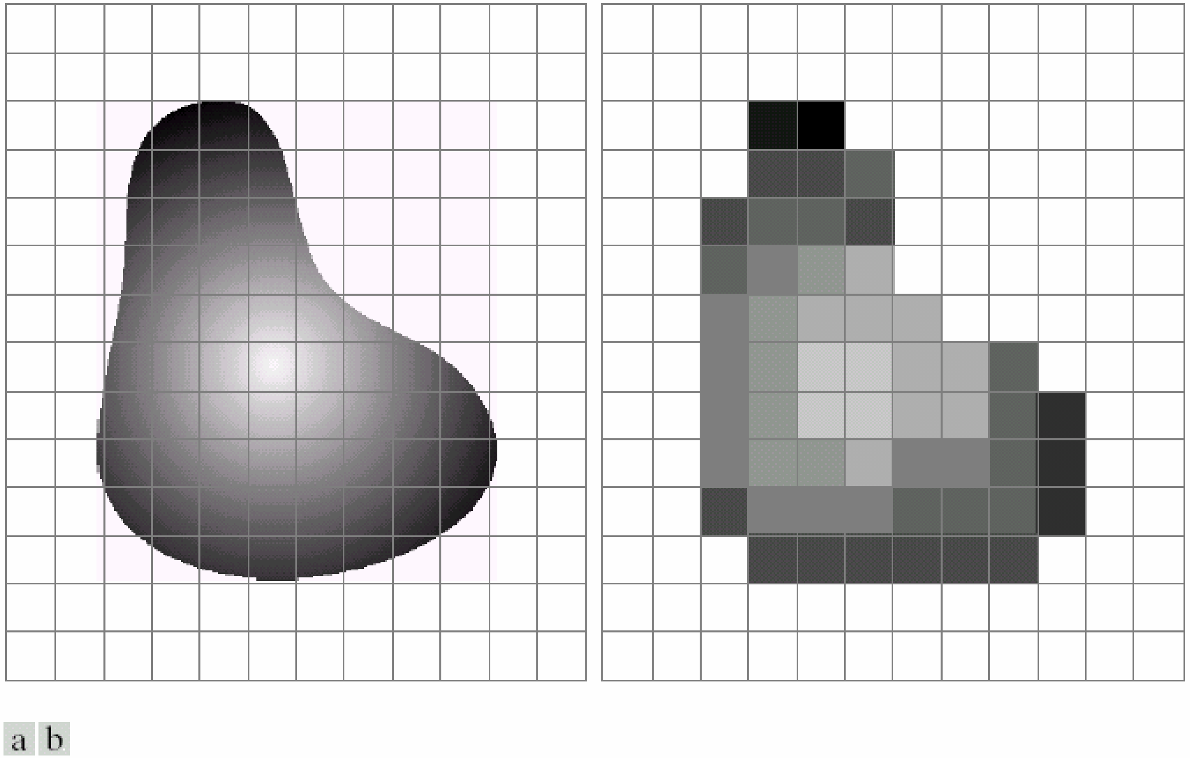


FIGURE 2.17 (a) Continuos image projected onto a sensor array. (b) Result of image sampling and quantization.

Effects of Varying Sampling Rate (1)

- Subsampling was performed by dropping rows and columns.
- The number of gray levels was kept constant at 256.

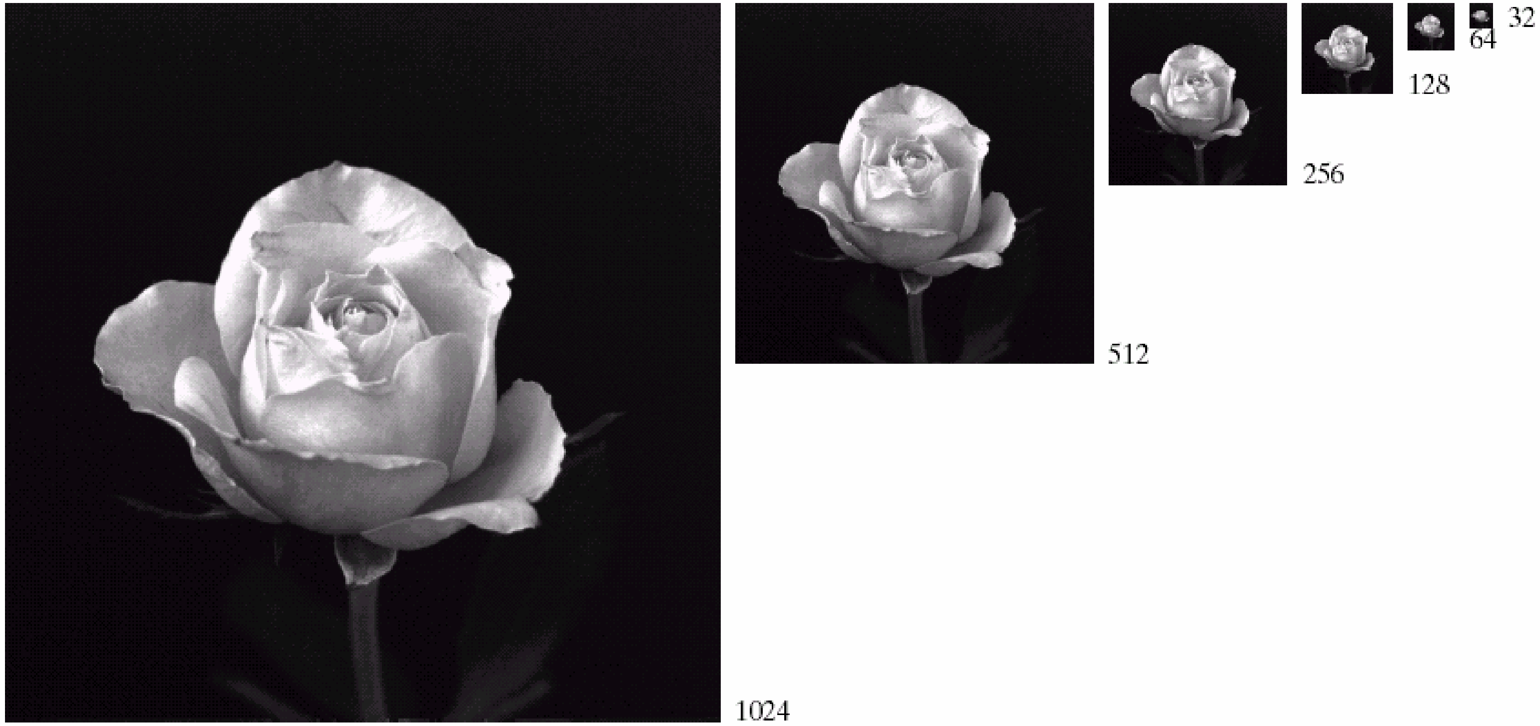
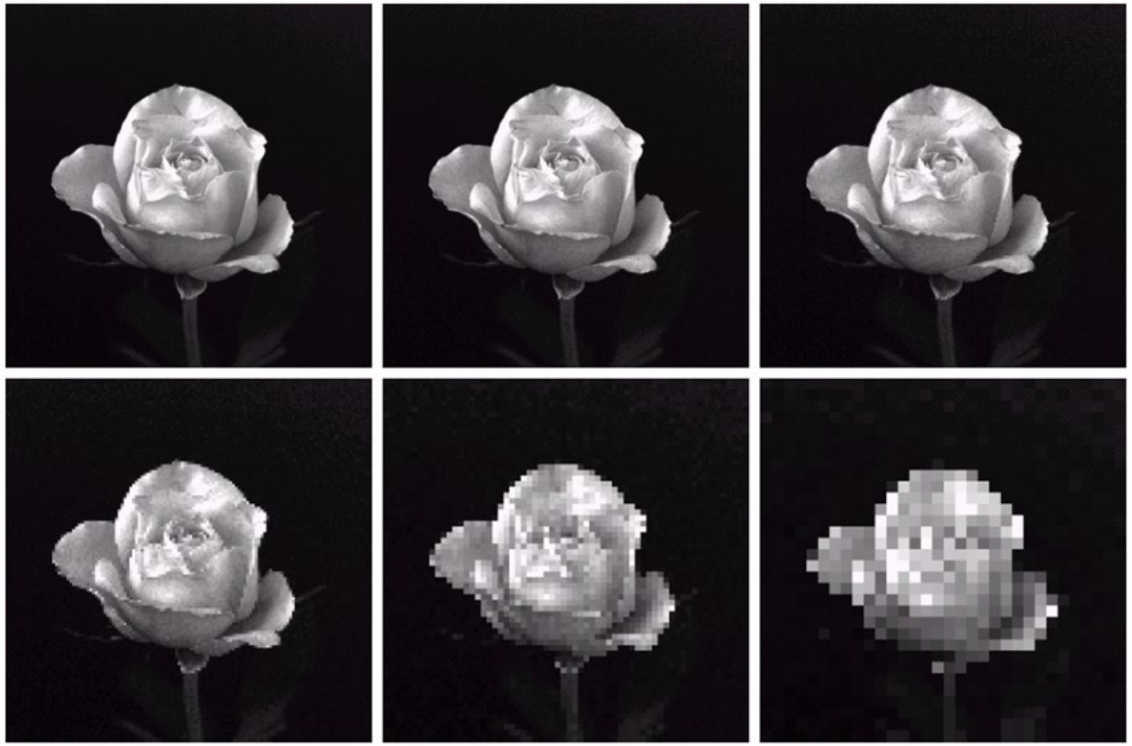


FIGURE 2.19 A 1024 × 1024, 8-bit image subsampled down to size 32 × 32 pixels. The number of allowable gray levels was kept at 256.

Effects of Varying Sampling Rate (2)

- Size differences make it difficult to see effects of subsampling.



a b c
d e f

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

Spatial Resolution

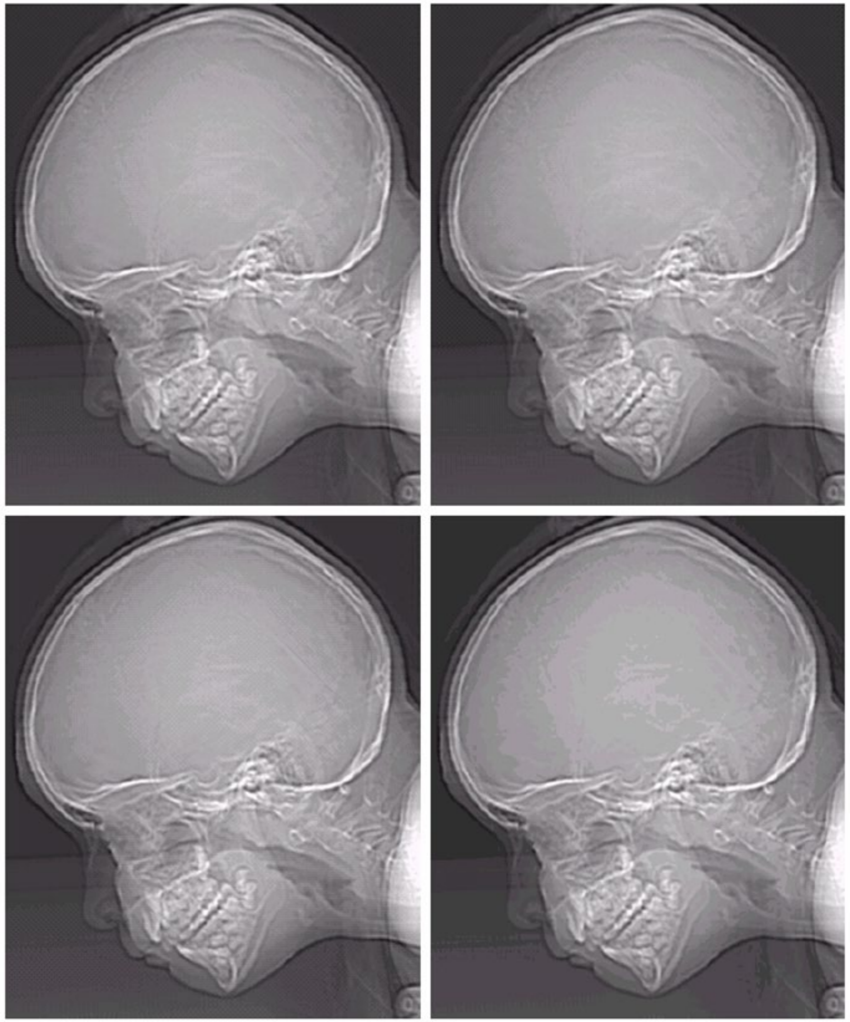
- Defined as the smallest discernable detail in an image.
- Widely used definition: smallest number of discernable line pairs per unit distance (100 line pairs/millimeter).
- A line pair consists of one line and its adjacent space.
- When an actual measure of physical resolution is not necessary, it is common to refer to an MxN image as having spatial resolution of MxN pixels.

Graylevel Resolution

- Defined as the smallest discernable change in graylevel.
- Highly subjective process.
- The number of graylevels is usually a power of two:
  - k bits of resolution yields 2^k graylevels.
  - When k=8, there are 256 graylevels ← most typical case
- Black-and-white television uses k=6, or 64 graylevels.

Effects of Varying Graylevels (1)

- Number of graylevels reduced by dropping bits from k=8 to k=1
- Spatial resolution remains constant.

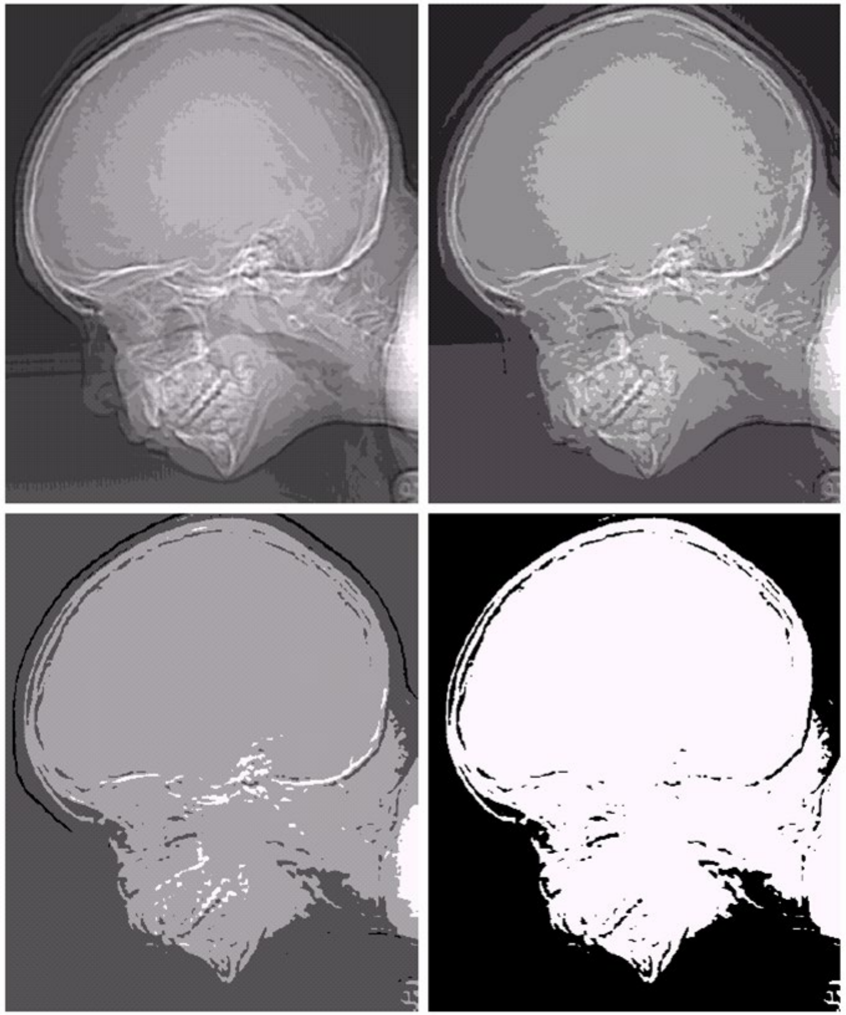


a b
c d

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

Effects of Varying Graylevels (2)

- Notice false contouring in coarsely quantized images.
- Appear as fine ridgelike structures in areas of smooth gray levels.



e f
g h

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

RGB Images

Figure 3.24 Top: Original RGB image and three separate color channels. Middle: RGB image obtained by swapping the color channels, and same three color channels of original image shown as grayscale images. Bottom: RGB image obtained by swapping the color channels in the reverse order, and three different grayscale transformations of the original image. See text for details.

RGB-Grayscale Transformations

- Gray = .30 R + .59 G + .11 B
- Note that the human visual system is not equally sensitive to all frequencies.
- Naïve approximations include:
  - Gray = (R + G + B) / 3
  - Gray = (R + 2G + B) / 4 (efficiently implemented using bitwise shifts)
- To convert from gray to RGB, simply replicate the gray value to each channel:

  R = gray

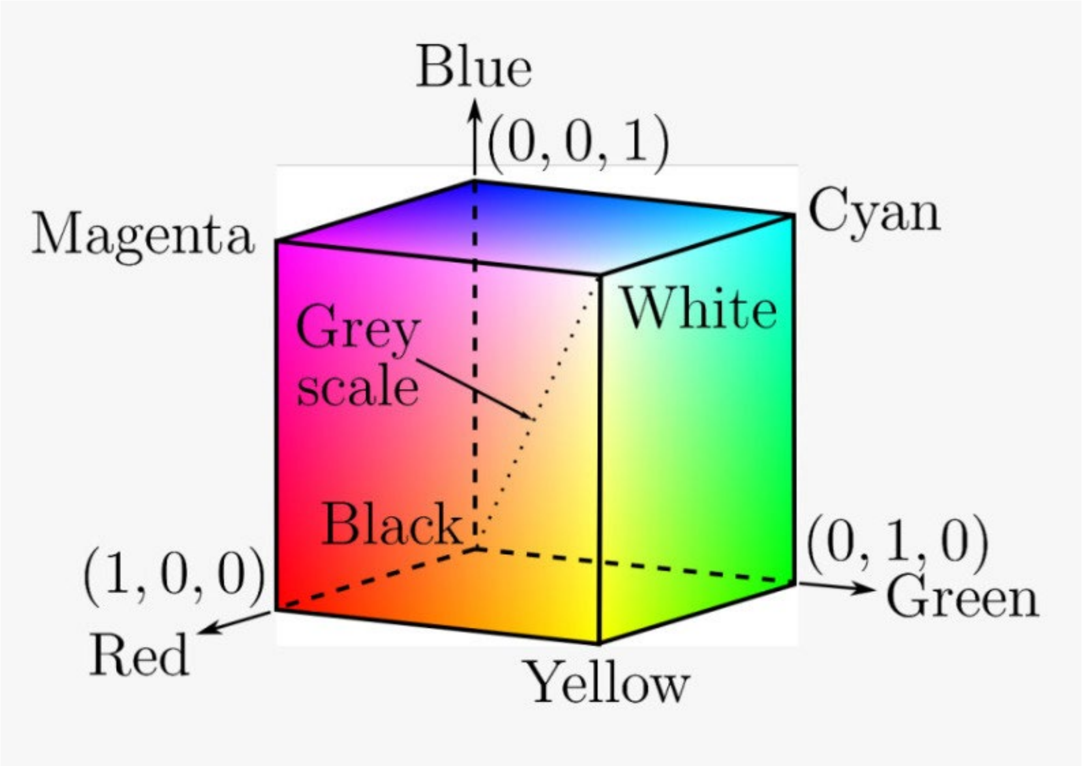
  G = gray

  B = gray

  There are many RGB values that transform to a given gray value.
- Colorization is a field in which we attempt to meaningfully convert B/W movies to color movies.

RGB Color Cube

- Consider a cube defined by the R-, G-, and, B-axes, each quantized to 8-bits.
- There are 2^8 x 2^8 x 2^8 = 2^{24} = 16M unique colors.
- All 256 colors along the diagonal (R = G = B) are gray values.



Storage Requirements

- Consider an NxN image having k bits per pixel.
- Color (RGB) images require three times the storage (assuming no compression).

TABLE 2.1
Number of storage bits for various values of N and k.

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

Large Space of Images

- Any image can be downsampled and represented in a few bits/pixel for use on small coarse displays.
- How many unique images can be displayed on an NxN k-bit display?
  - 2^k possible values at each pixel
  - N^2 pixels
  - Total: $(2^k)^{N^2}$
- This total is huge even for k=1 and N=8:

  18,446,744,073,709,551,616 $\leftarrow 2^{64}$
- It would take 19,498,080,578 years to view this if it were laid out on video at 30 frames/sec.