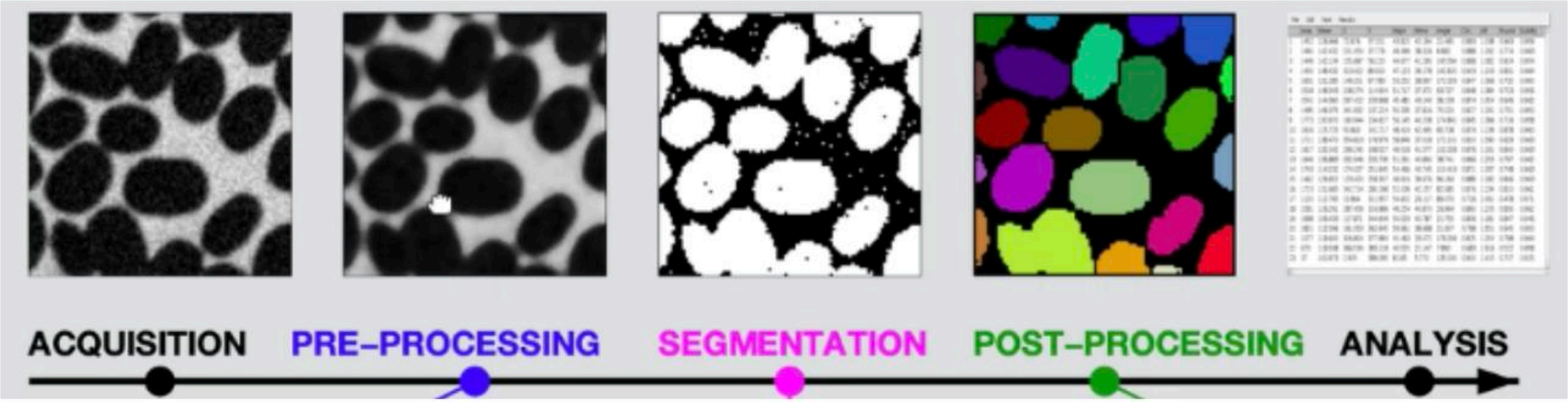
Announcement

1/23/2025 Thursday

- Zoom class

- Digital Image Fundamentals
  -

Roadmap: image processing → computer vision



Low level

Mid level

High level

Image acquisition

Image enhancement

Image compression

Image segmentation

Object recognition

Scene understanding

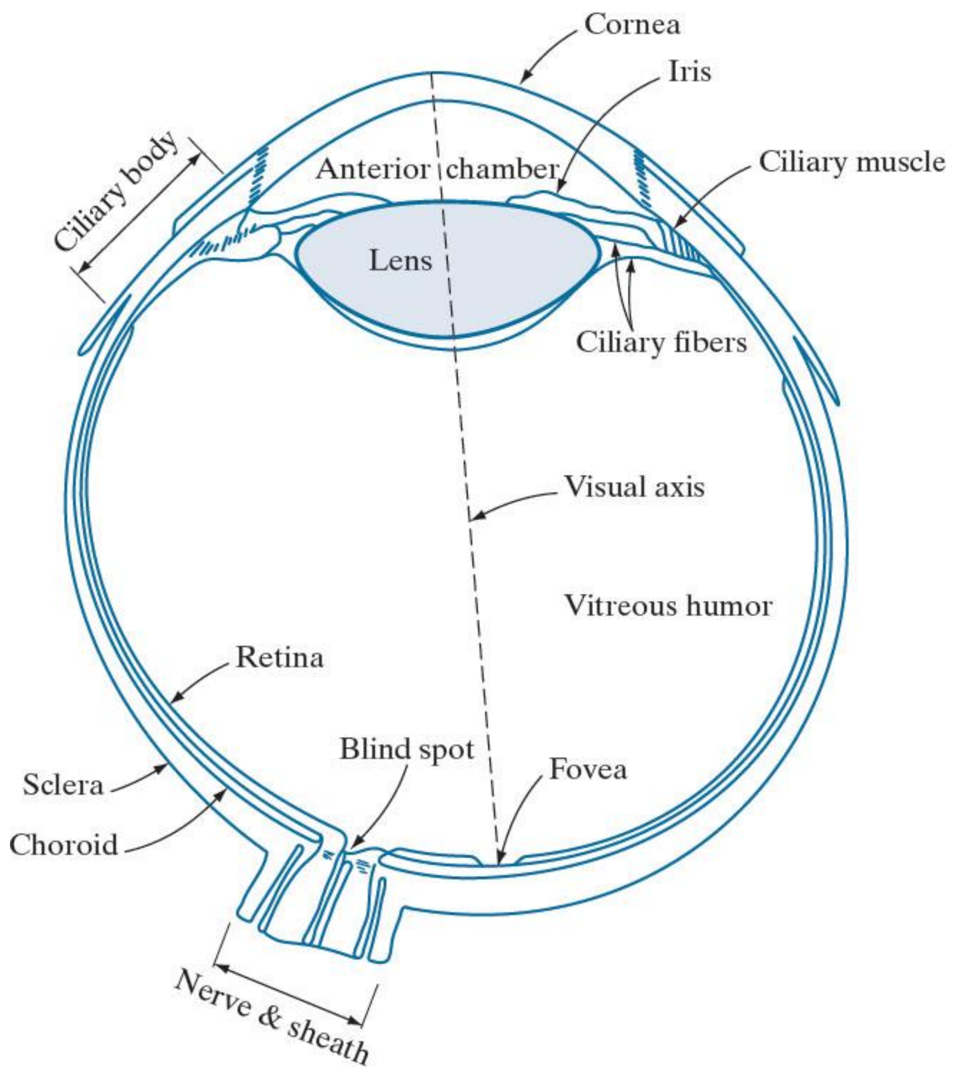
Semantics

Image processing

Image analysis
(Computer vision,
Pattern recognition, etc.)

Elements of Human Visual Perception

- Human visual perception plays a key role in selecting a technique
- Lens and Cornea: focusing on the objects
- Two receptors in the retina: Cones and rods
  - Cones located in fovea and are highly sensitive to color
  - Rods give a general overall picture of view, are insensitive to color and are sensitive to low level of illumination



Distribution of Rods and Cones in the Retina

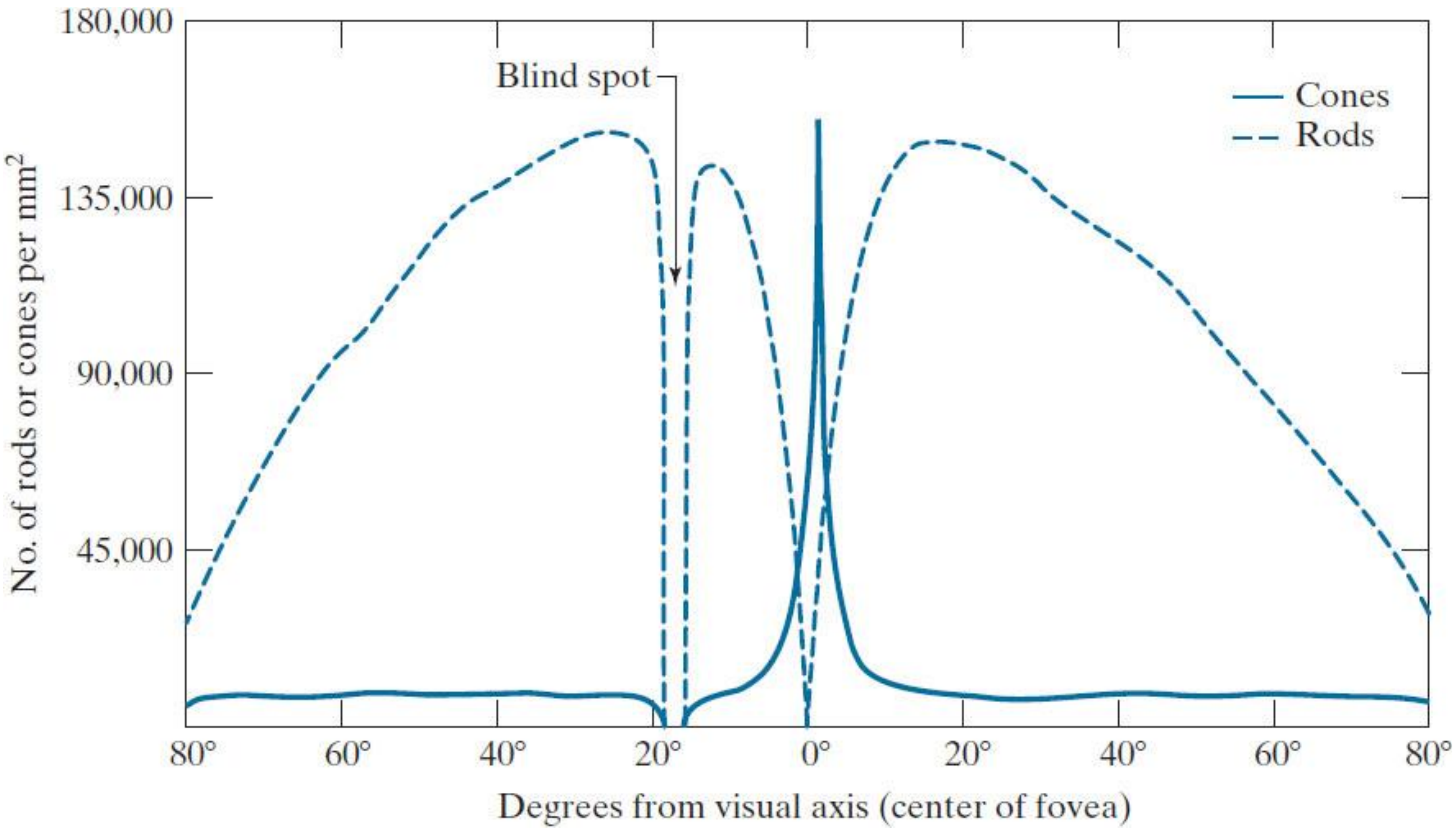
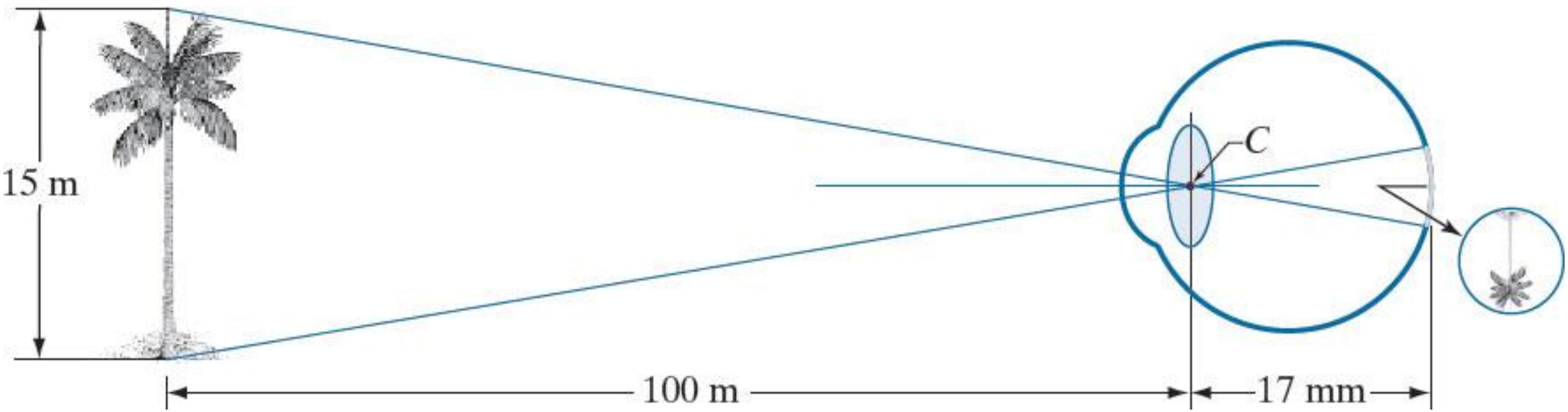


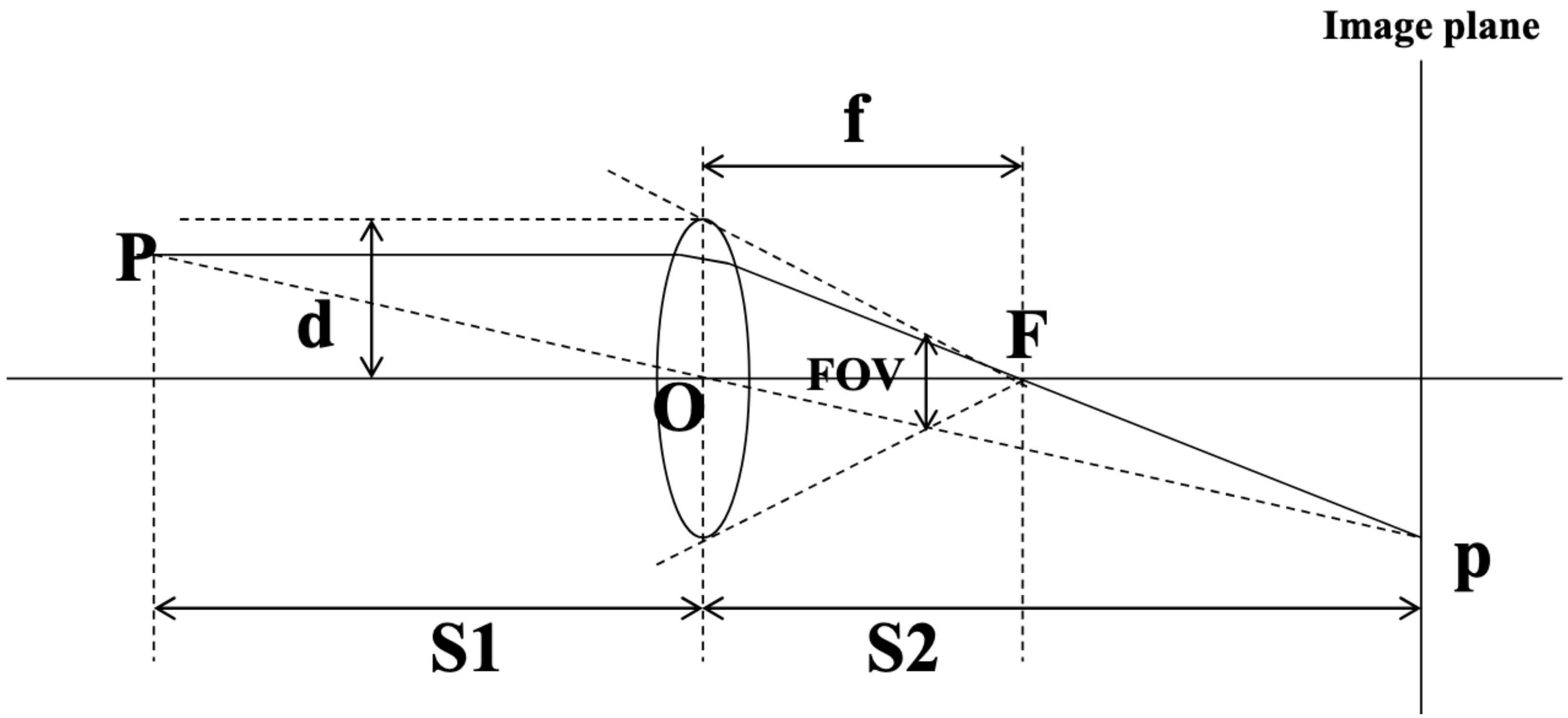
Image Formation in the Eye

- Image is upside down in the retina/imaging plane!



Adjust focus length
- Camera
- Human eye

Lens Parameters

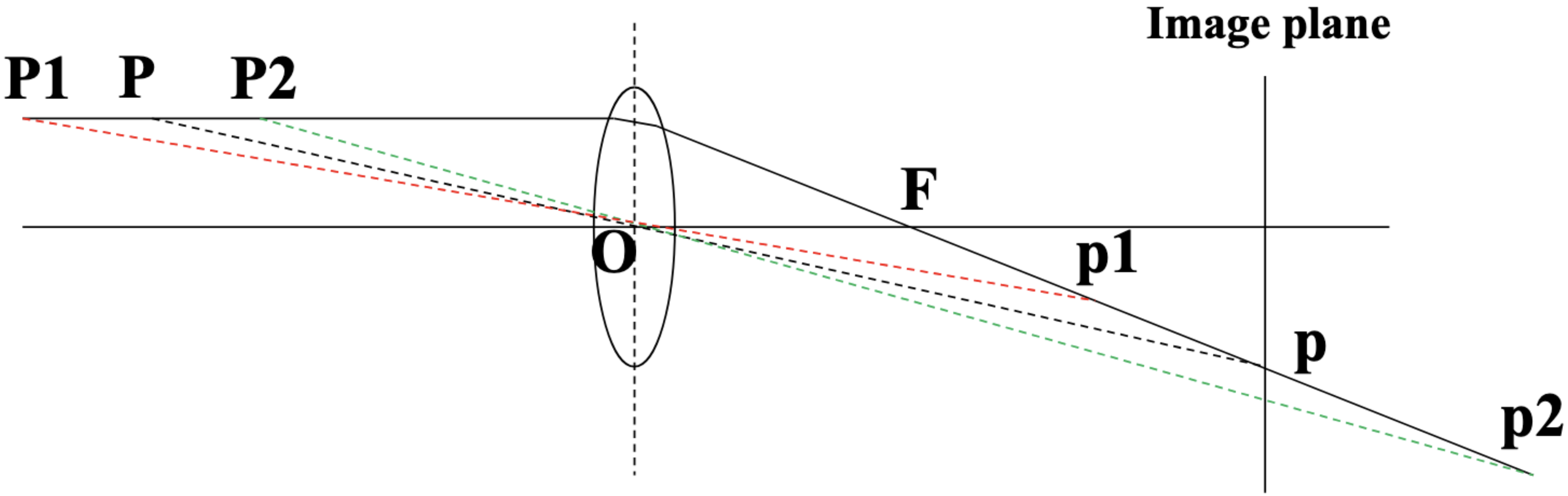


- Thin lens theory: $\frac{1}{S_1} + \frac{1}{S_2} = \frac{1}{f}$
- Field of view (FOV): $\omega = 2\arctan\frac{d}{f}$

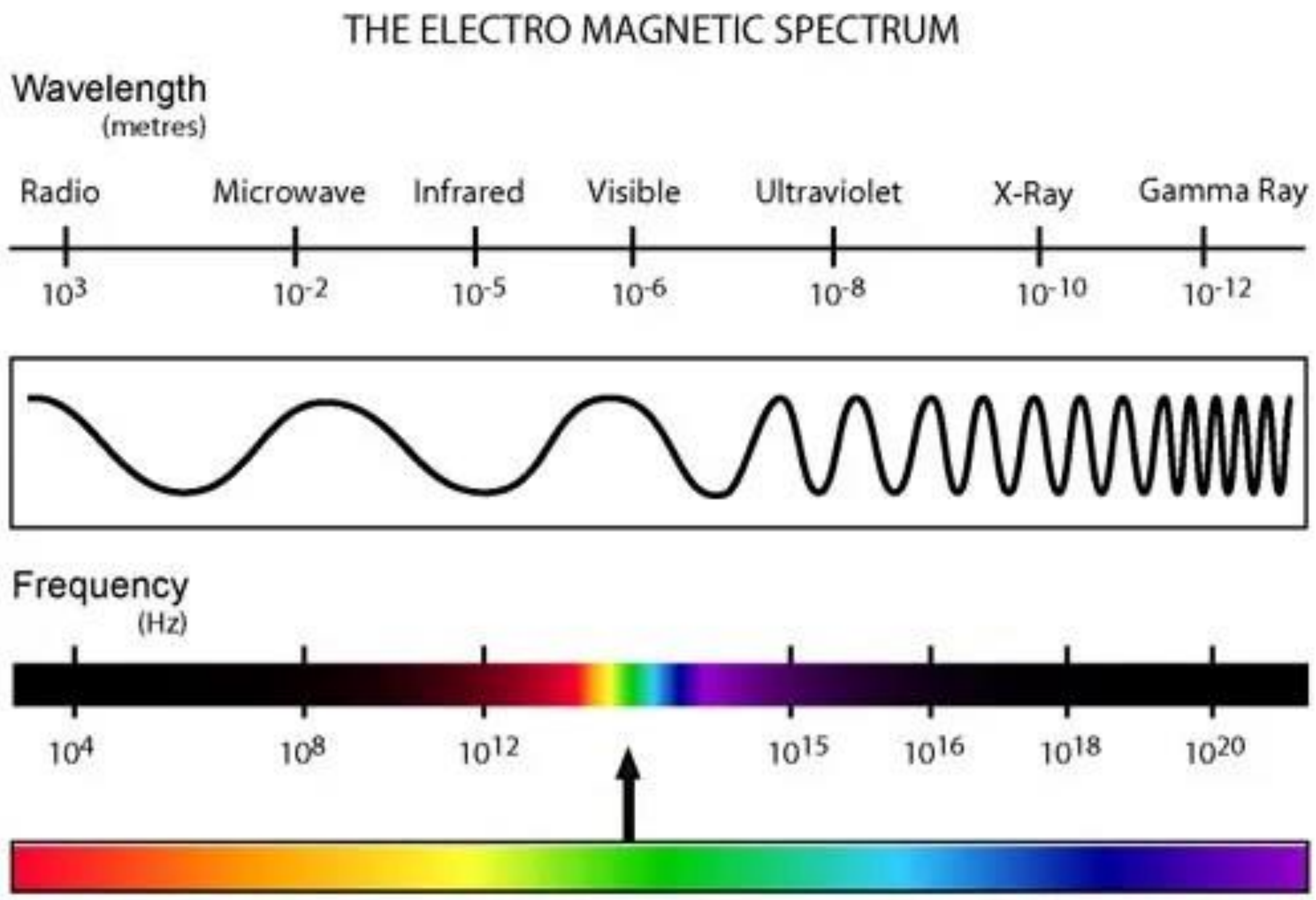
- Increasing the distance from the object to the lens will reduce the size of image
- Large focus length will give a small FOV

Depth of Field & Out of Focus

- DOF is inversely proportional to the focus length
- DOF is proportional to S1



Light and EM Spectrum

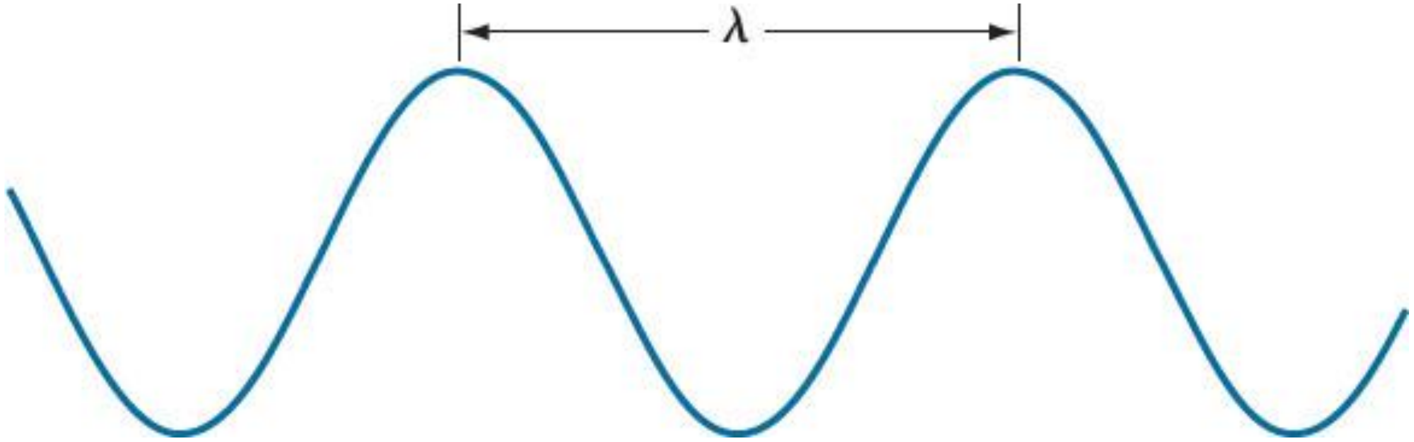


380 to 750 nm

380 to 750 nm

https://colourware.org/2013/11/20/is-there-such-a-thing-as-visible-light/

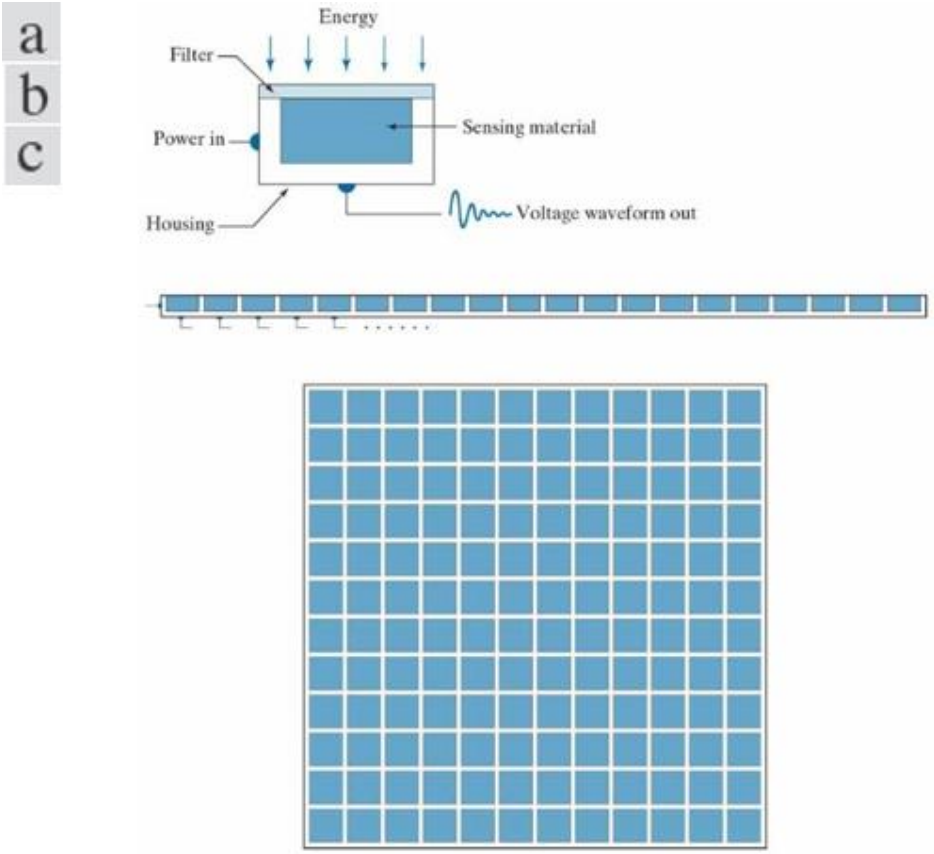
Relation Among Wavelength, Frequency and Energy



- Wavelength λ, frequency ν, energy E
- $\lambda = \frac{c}{\nu}$ $c = 2.988 \times 10^{8}\ m/s$ is the speed of light
- $E = h\nu$ h is the Planck's constant, $6.626068 \times 10^{-34} m^2 kg/s$

Image Sensing and Acquisition

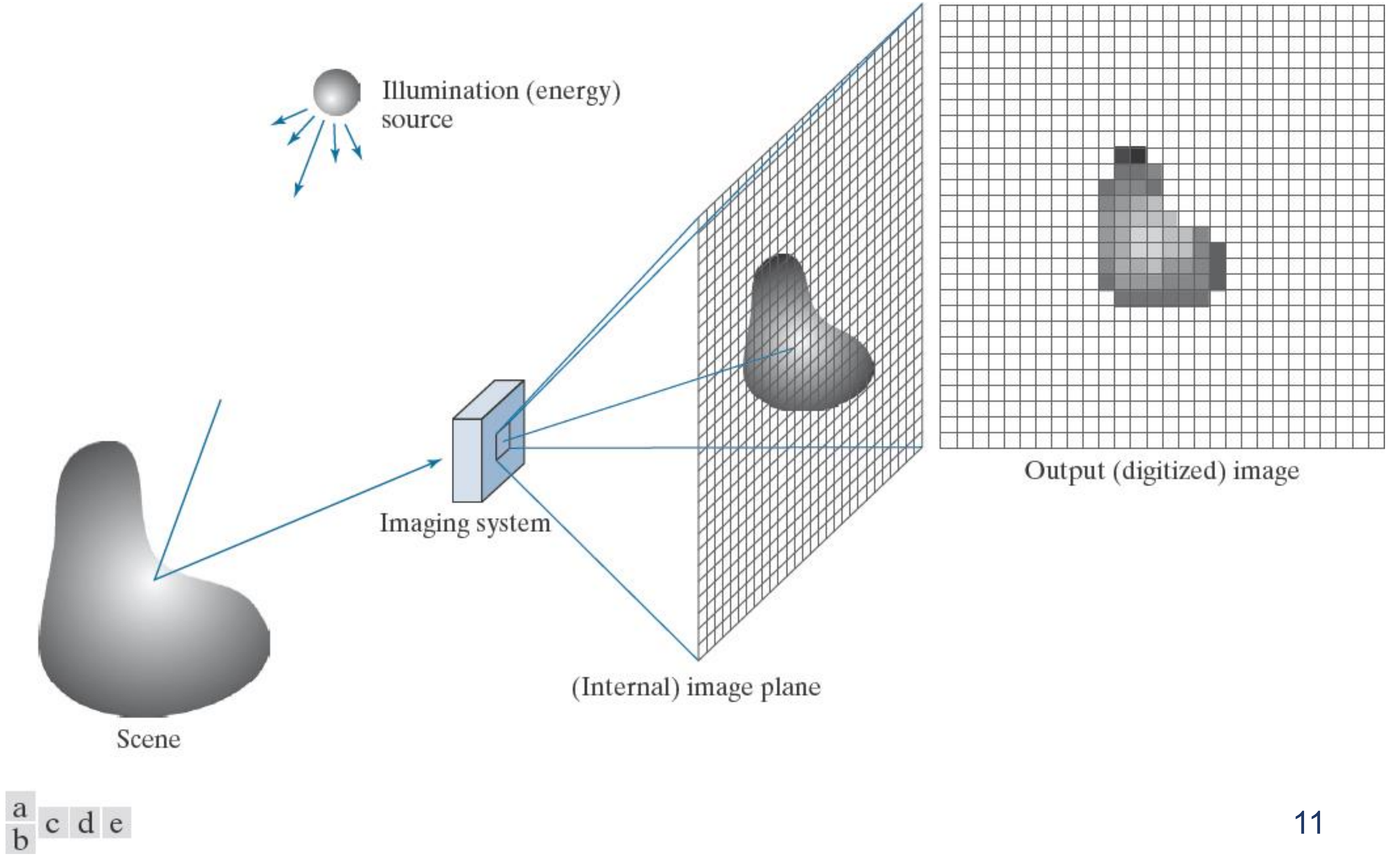
- Illumination energy → digital images
- Incoming energy is transformed into a voltage
- Digitizing the response



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

Digital Image Acquisition

- Illumination (energy) source.
- Imaging system
- Projection of the scene onto the image plane
- Digitized image



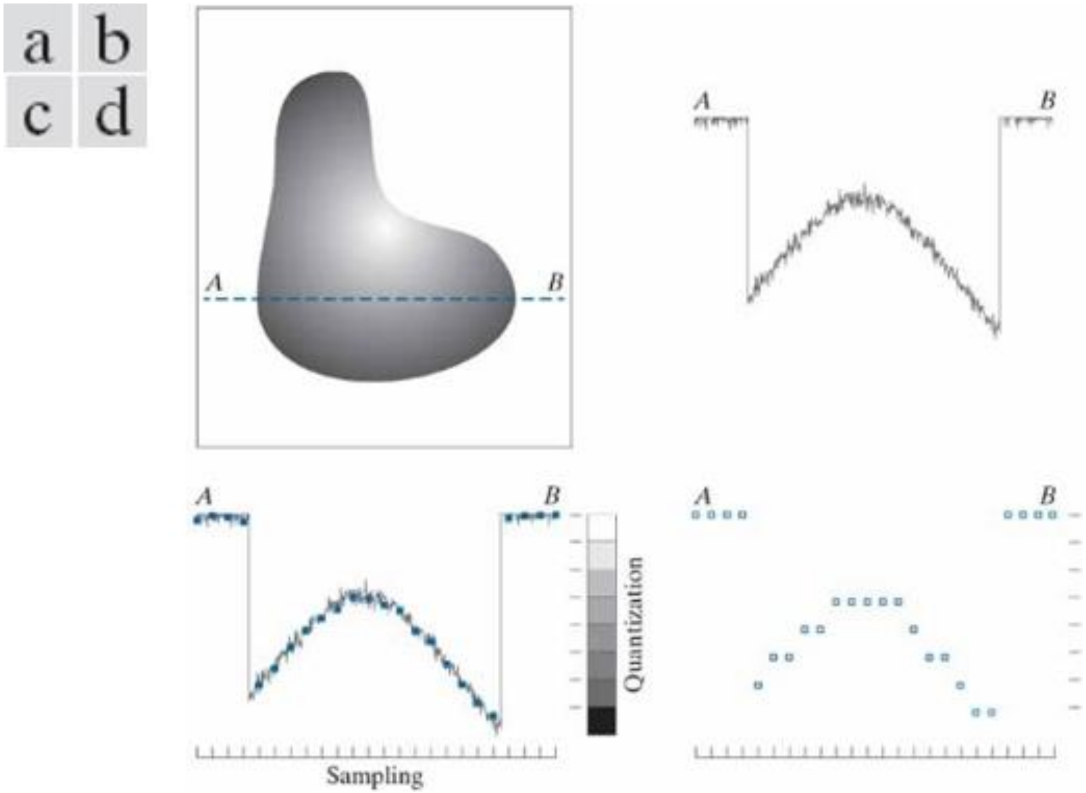
A (2D) Image

- An image = a 2D function $f(x, y)$ where
  - x and y are spatial coordinates
  - $f(x, y)$ is the intensity or gray level
- A digital image:
  - x, y , and $f(x, y)$ are all finite
  - For example, $x \in \{1,2, \dots, M\}, y \in \{1,2, \dots, N\}, f(x, y) \in \{0,1,2, \dots, 255\}$

- Digital image processing → processing digital images by means of a digital computer

- Each element (x, y) in a digital image is called a pixel (picture element)

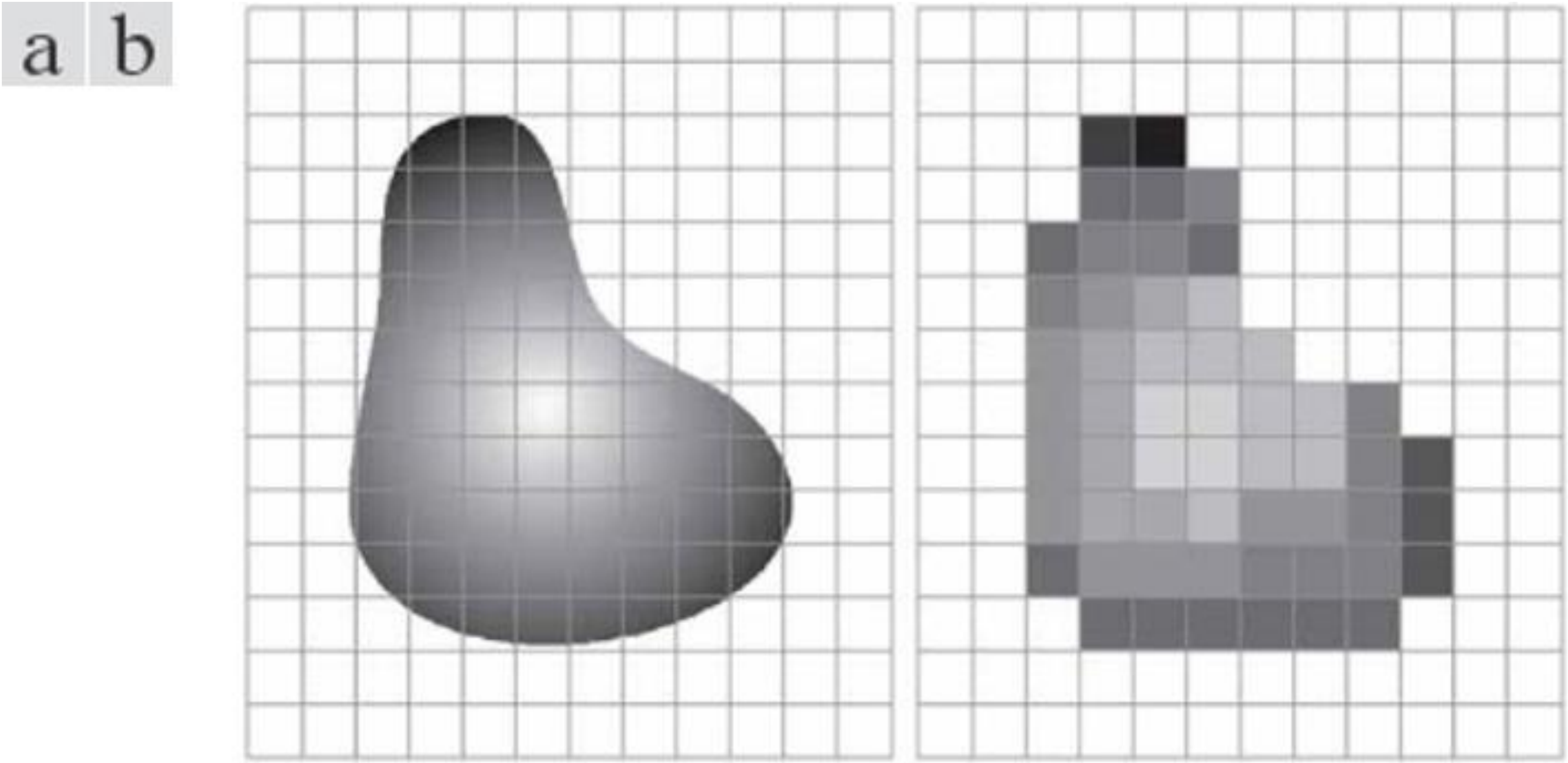
Image Sampling and Quantization



(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.

- Sampling: Digitizing the coordinate values (usually determined by sensors)
- Quantization: Digitizing the amplitude values

Image Sampling and Quantization in a Sensor Array



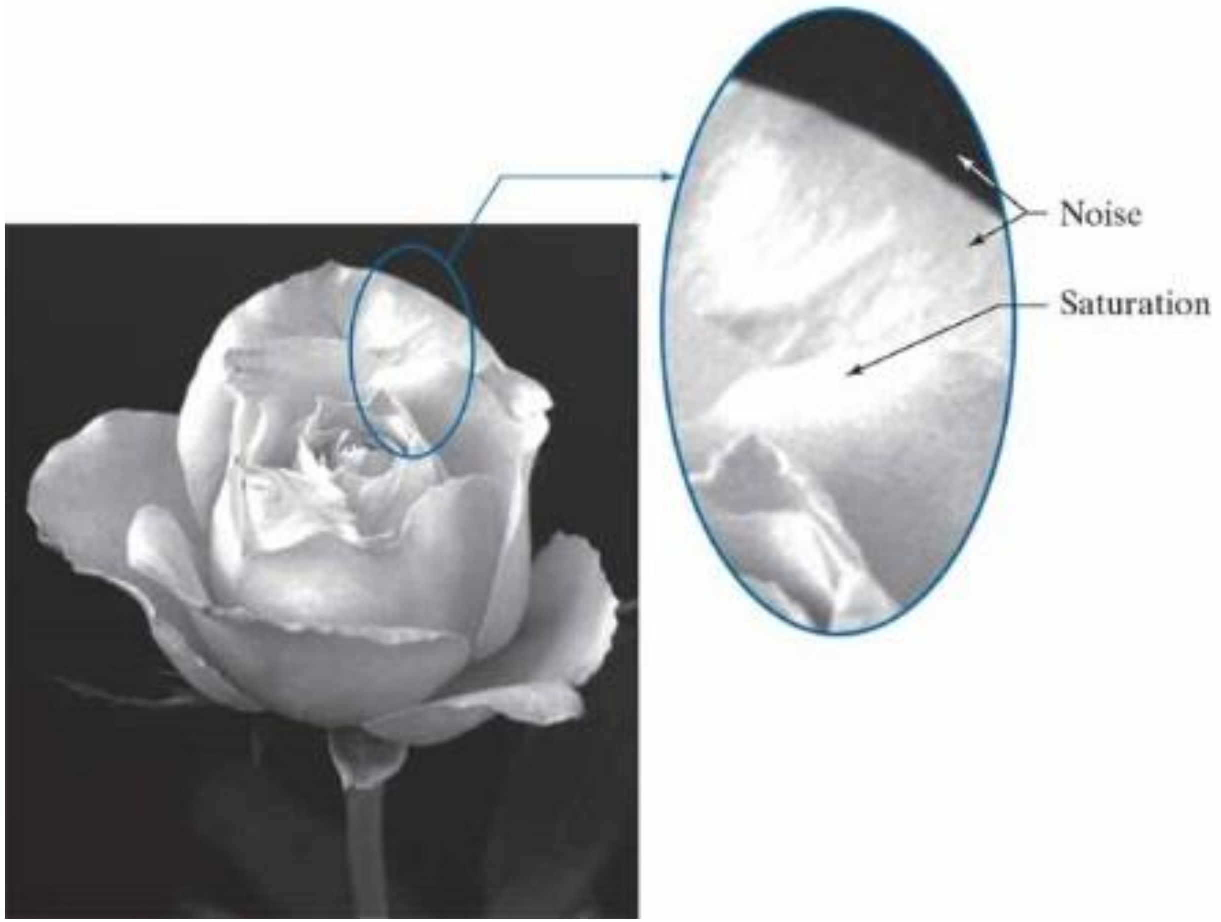
(a) Continuous image projected onto a sensor array.
(b) Result of image sampling and quantization.

Dynamic Range

- $L_{min} < f(x, y) < L_{max}$
  →

$0 \leq f(x, y) \leq L - 1$ and $L = 2^k$ (in practice)

- Dynamic range/contrast ratio
  - The ratio of the maximum detectable intensity level (saturation) to the minimum detectable intensity level (noise)
  - $\frac{I_{max}}{I_{min}}$



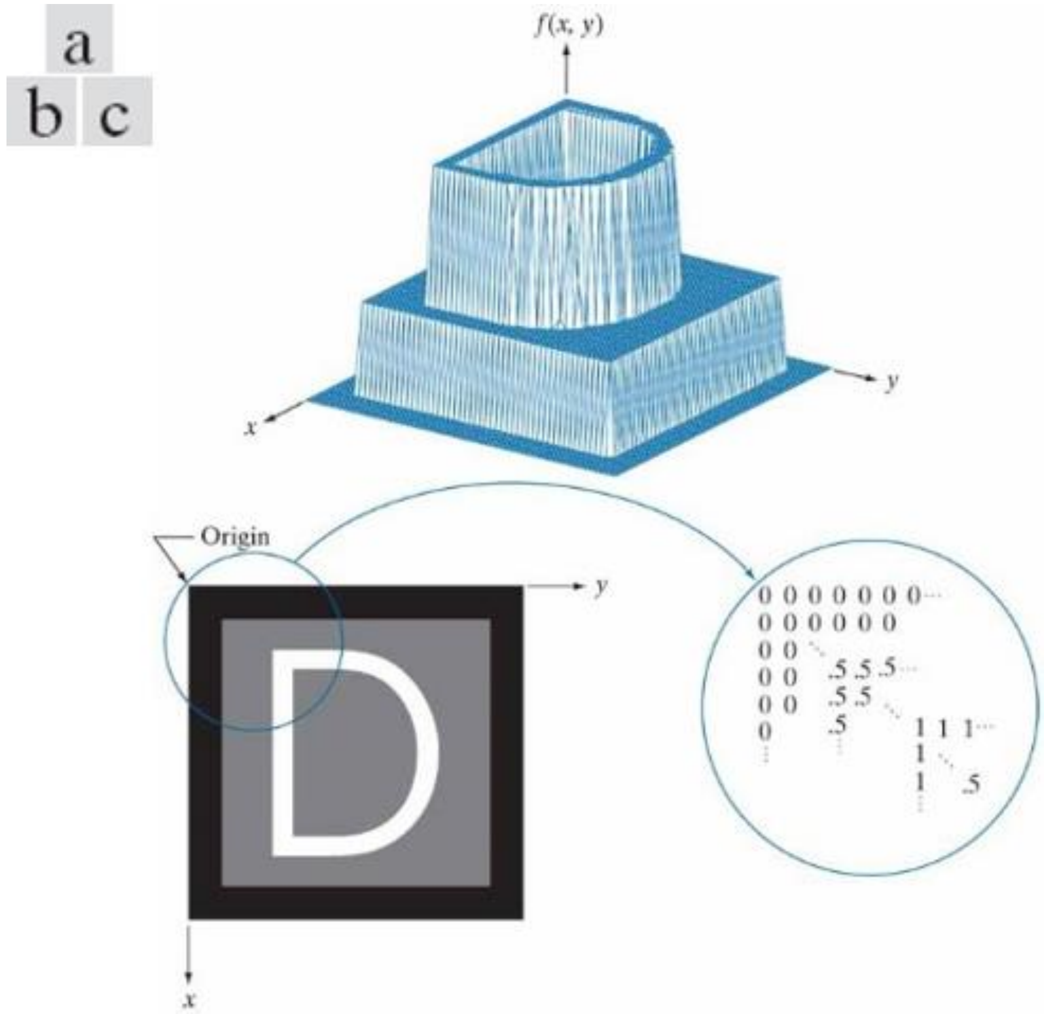
Representing Digital Images

- (a) Image plotted as a surface.
- (b) Image displayed as a visual intensity array. → Suitable for visualization
- (c) Image shown as a 2-D numerical array. → processing and algorithm development

Number of bits storing the image

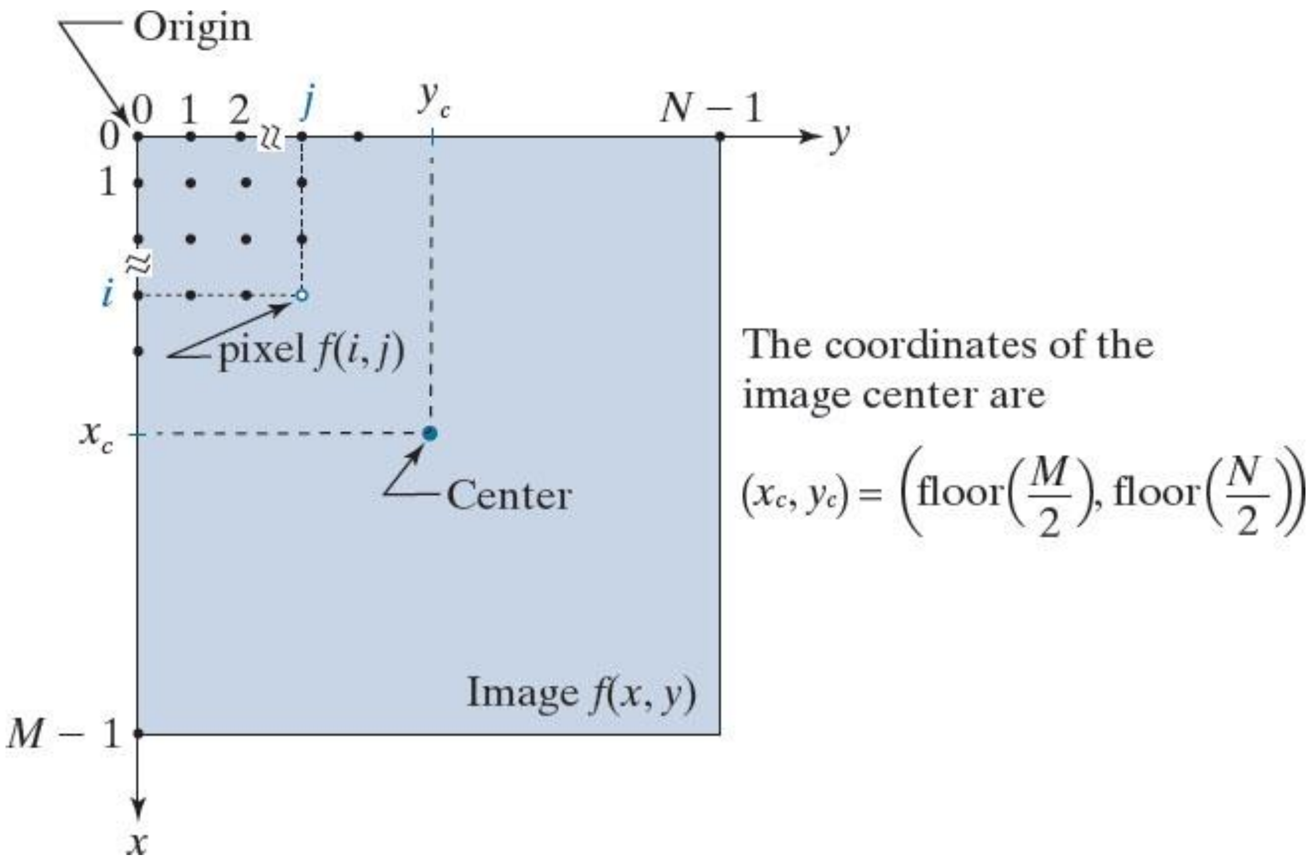
↑

$$b = M \times N \times k$$



Representing Digital Images

- Coordinate convention used to represent digital images. Because coordinate values are integers, there is a one-to-one correspondence between **x** and **y** and the rows (**r**) and columns (**c**) of a matrix.

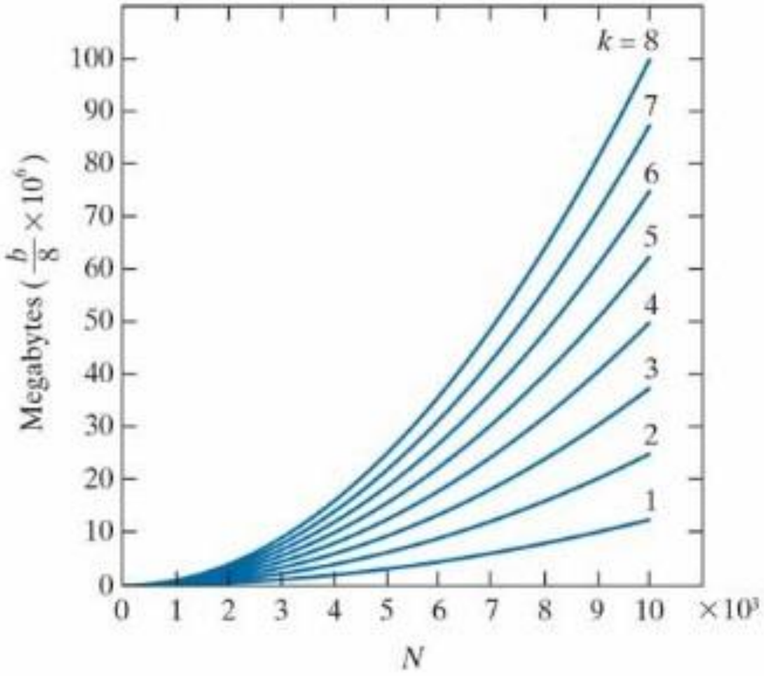


Store an Image

- Number of megabytes required to store images for various values of **N** and **k**.

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



Spatial Resolution

- Spatial resolution: smallest discernible details
- # of line pairs per unit distance
- # of dots (pixels) per unit distance
  - Printing and publishing
  - In US, dots per inch (dpi)

- Large image size itself does not mean high spatial resolution!
- Scene/object size in the image matters

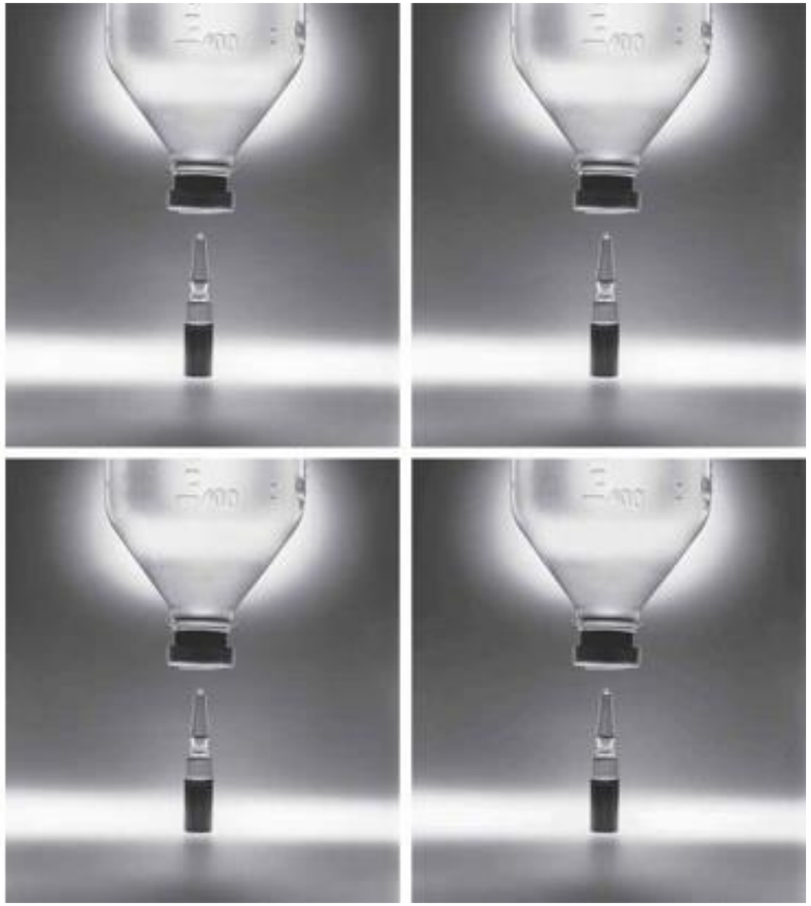
930 dpi 300 dpi

150 dpi 72 dpi

Intensity Resolution

- Keep the image size constant (2022×1800)

256 level 128 level

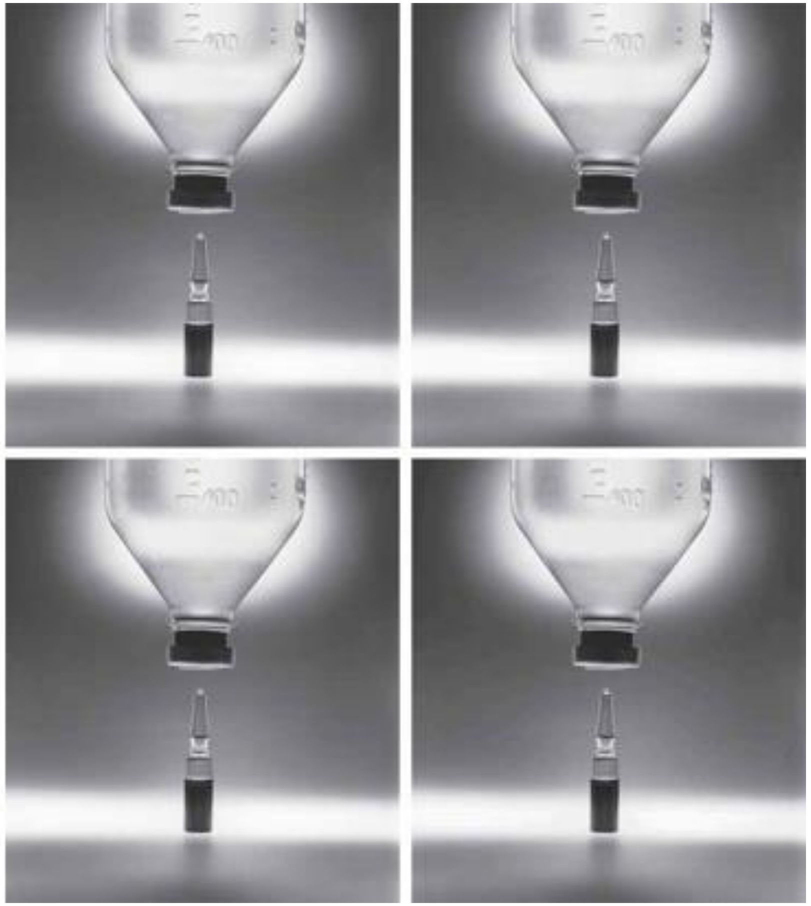


64 level 32 level

Intensity Resolution

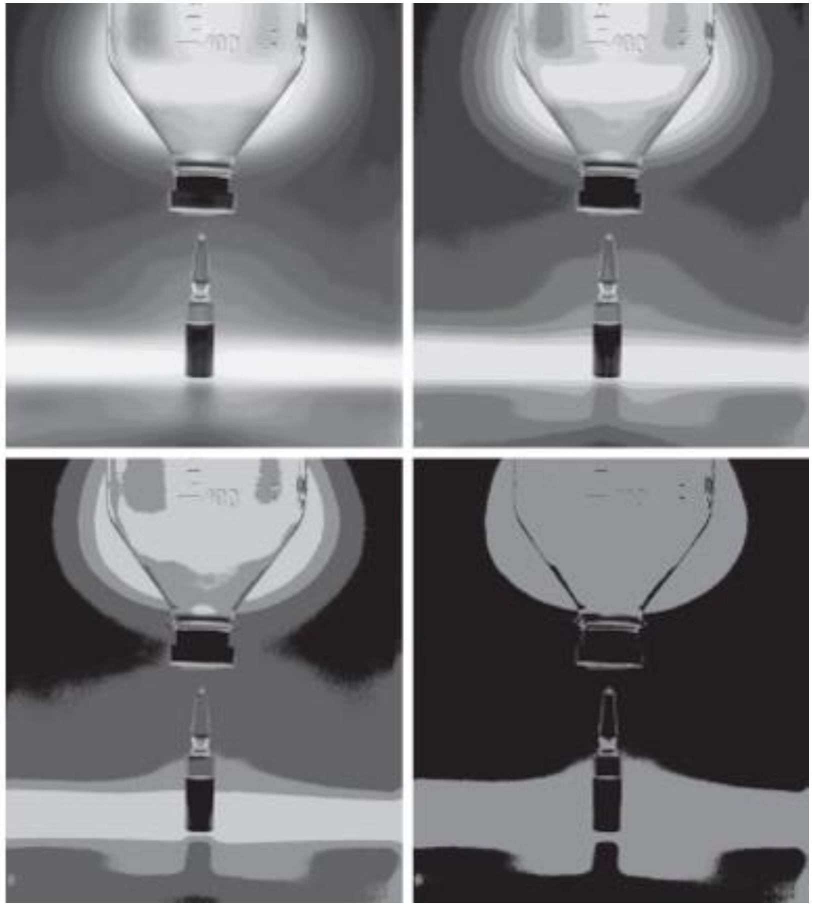
- Keep the image size constant (2022×1800)

256 level 128 level



64 level 32 level

16 level 8 level



4 level 2 level

Some Basic Relationships between Pixels

Neighbors of a pixel

$N_4(p)$

	(x-1,y)	
(x,y-1)	(x,y)	(x,y+1)
	(x+1,y)	

$N_D(p)$

(x-1,y-1)		(x-1,y+1)
	(x,y)	
(x+1,y-1)		(x+1,y+1)

$N_8(p)$

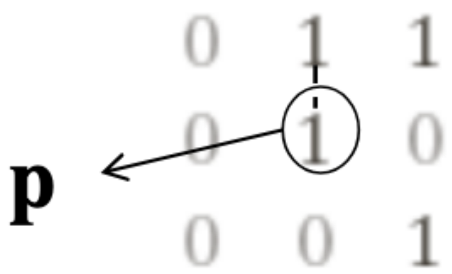
(x-1,y-1)	(x-1,y)	(x-1,y+1)
(x,y-1)	(x,y)	(x,y+1)
(x+1,y-1)	(x+1,y)	(x+1,y+1)

Adjacency

- Adjacency is the relationship between two pixels p and q

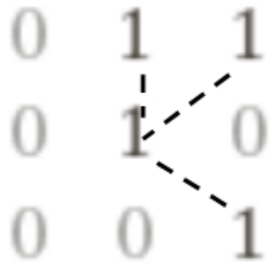
- Three types of adjacency

4-adjacency



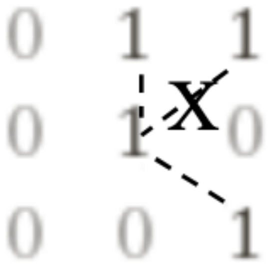
$q \in N_4(p)$

8-adjacency



$q \in N_8(p)$

m-adjacency



$q \in N_D(p)$ **and** $N_4(q) \cap N_4(p) = \emptyset$

$or\ q \in N_4(p)$

Connectivity

- Path from p to q: a sequence of distinct and adjacent pixels with coordinates

$$\text{Starting point p} \leftarrow \underbrace{(x_0, y_0), (x_1, y_1)}_{\text{adjacent}}, \dots, (x_n, y_n) \rightarrow \text{ending point q}$$

- p and q are connected: if there is a path from p to q in S
- Closed path: if the starting point is the same as the ending point
- Connected component: all the pixels in S connected to p
- Connected set: S has only one connected component

Regions

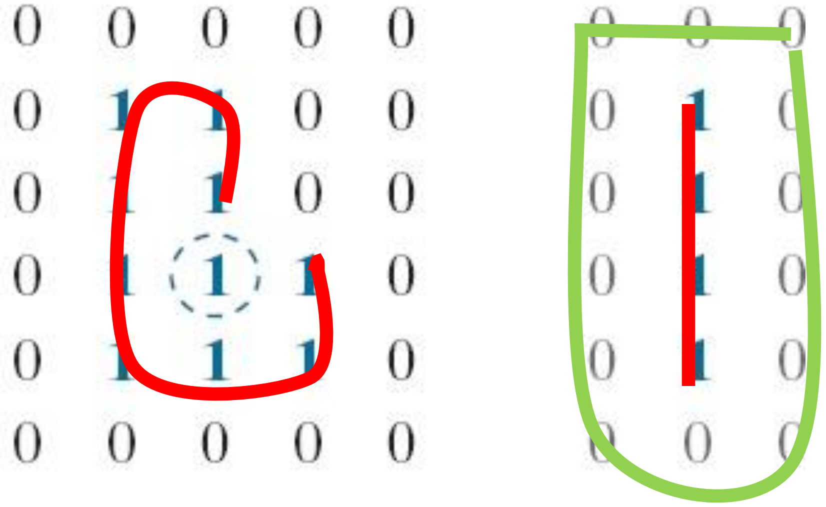
- R is a region if R is a connected set

- R_i and R_j are adjacent if is a connected set

$$\begin{array}{l} \left.\begin{array}{ccc}1&1&1\\1&0&1\\0&1&0\end{array}\right\} R_i \\ \left.\begin{array}{ccc}0&0&1\\1&1&1\\1&1&1\end{array}\right\} R_j \end{array}$$

Boundaries

- Inner boundary (boundary) -- the set of pixels each of which has at least one background neighbor
- Outer boundary – the boundary pixels in the background



Distance Measures

- For pixels p, q, and z, with coordinates (x,y), (s,t) and (v,w), D is a distance function or metric if

$$(a)\ D(p,q) \geq 0 \quad D(p,q) = 0 \ iff \ p = q$$

$$(b)\ D(p,q) = D(q,p), \quad and$$

$$(c)\ D(p,z) \leq D(p,q) + D(q,z)$$

Distance Measures

- Euclidean distance

$$D_e(p,q) = \sqrt{(x-s)^2 + (y-t)^2}$$

- City-block (D4) distance

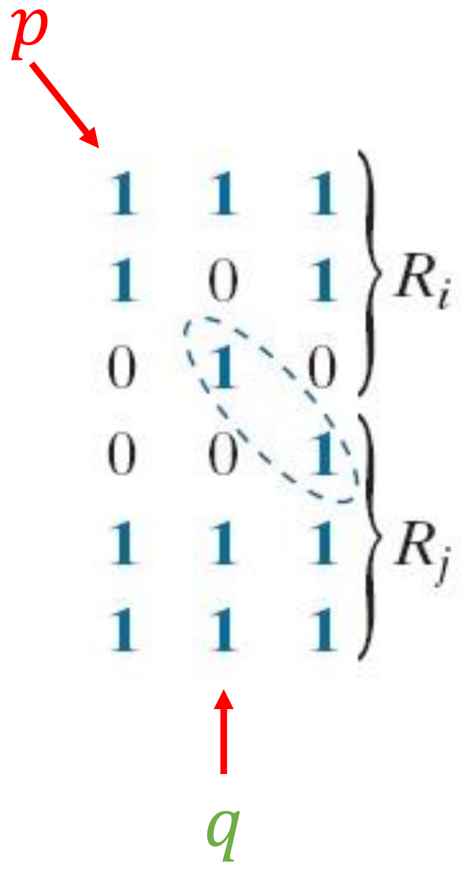
$$D_4(p,q) = |x-s| + |y-t|$$

- Chessboard (D8) distance (Chebyshev distance)

$$D_8(p,q) = \max(|x-s|, |y-t|)$$

Distance: Sample Problem

- D4 distance
- D8 distance
- Euclidean distance



Mathematic Tools

- Array/Matrix operations
- Linear/nonlinear operations
  - Linearity $H\left[a_i f_i(x, y) + a_j f_j(x, y)\right] = a_i H\left[f_i(x, y)\right] + a_j H\left[f_j(x, y)\right]$
- Arithmetic Operations – single pixel operations
  - Image averaging, image subtraction, image multiplication
- Set and logic operations
- Spatial operations
  - Single pixel operations and neighborhood operations
- Image transformation
- Probabilistic methods

Array versus Matrix operations

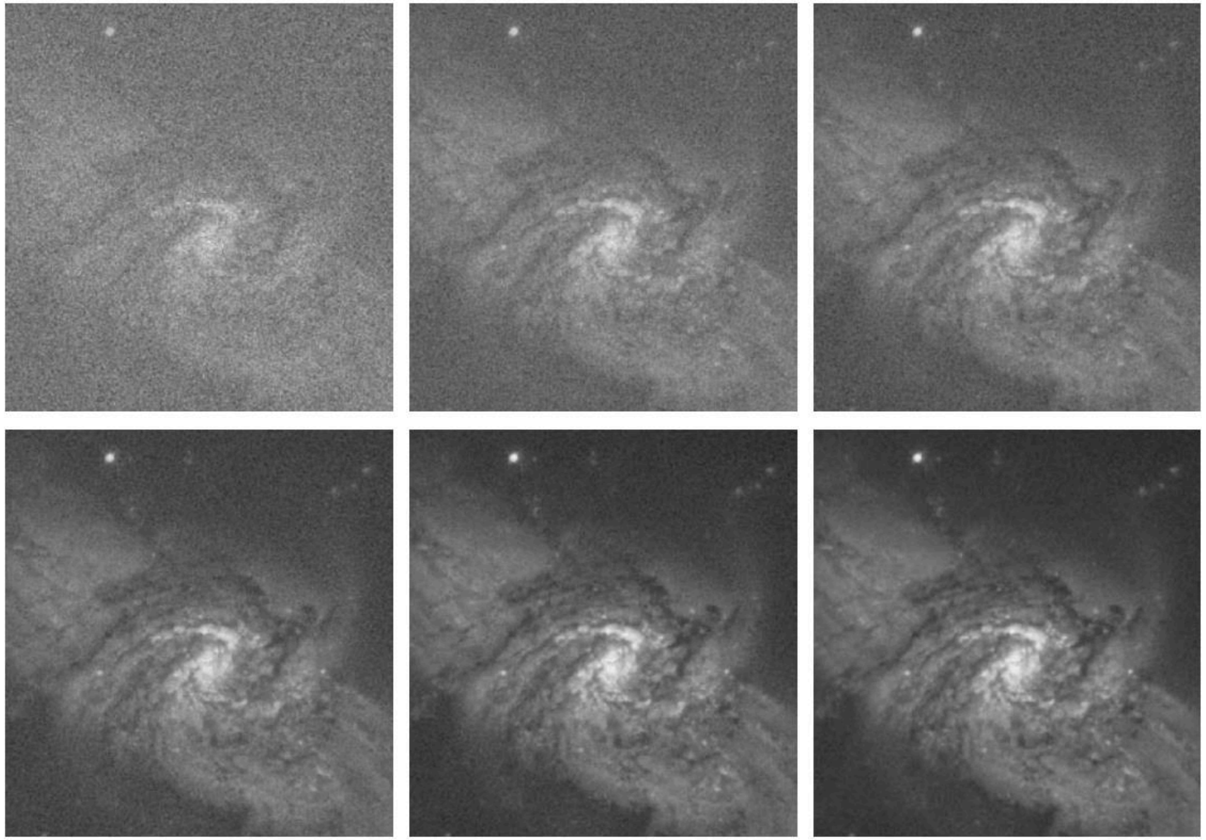
- Array Multiplications

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$$

- Matrix Multiplications

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

Image Averaging – Noise Reduction



a b c
d e f

FIGURE 2.29 (a) Image of Galaxy Pair NGC 3314 corrupted by additive Gaussian noise. (b)-(f) Result of averaging 5, 10, 20, 50, and 1,00 noisy images, respectively. All images are of size 566×598 pixels, and all were scaled so that their intensities would span the full [0, 255] intensity scale. (Original image courtesy of NASA.)

$$g(x, y) = f(x, y) + \eta(x, y)$$

$$\downarrow$$

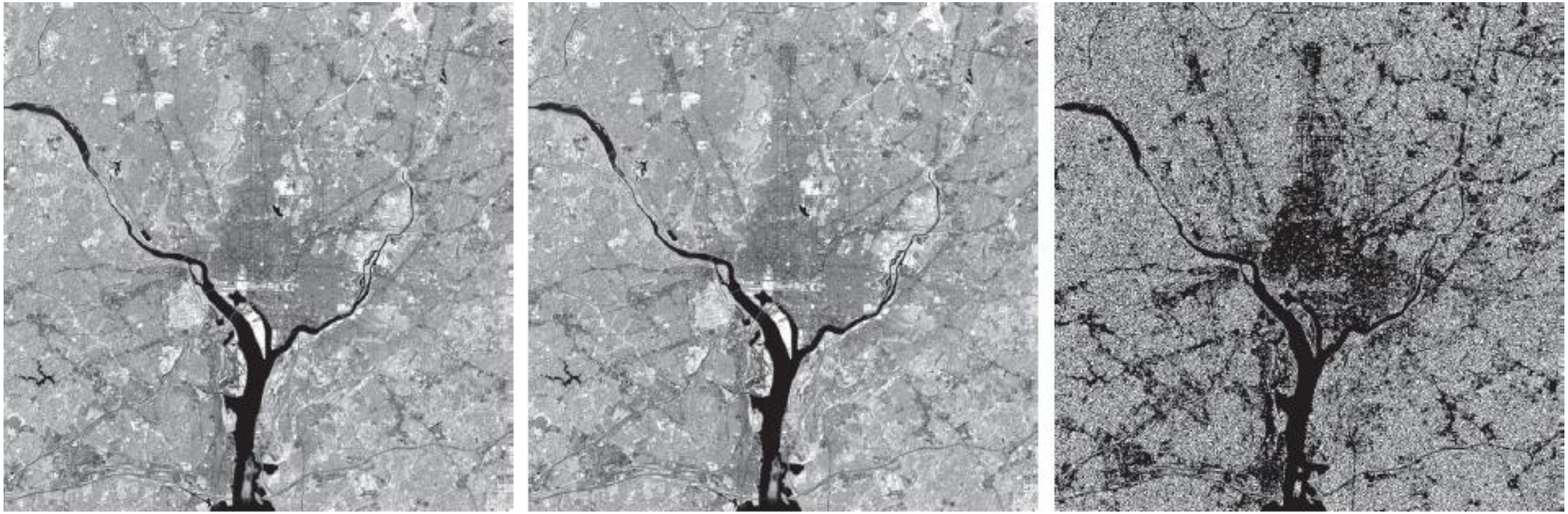
$$\bar{g}(x, y) = \frac{1}{K} \sum_{i=1}^{K} g_i(x, y)$$

$$E\{\bar{g}(x, y)\} = f(x, y)$$

$$\sigma^2_{\bar{g}(x,y)} = \frac{1}{K} \sigma^2_{\eta(x,y)}$$

- Assumption: the noise is uncorrelated in image and has zero mean

Image Subtraction – Enhance Difference



a b c

(a) Infrared image of the Washington, D.C. area.
(b) Image resulting from setting to zero the least significant bit of every pixel in (a).
(c) Difference of the two images, scaled to the range [0, 255] for clarity.

The images used in averaging & subtraction must be registered!

Image Multiplication (Division)

- Patch/Region-of-Interest

ROI mask for isolating teeth with fillings
(white corresponds to 1 and black corresponds to 0)

Product of the original image and the ROI masks

$$g(x, y) = f(x, y)h(x, y)$$

Image Multiplication (Division)

- Shading correction

Estimated shading pattern.

Product of the original image by the reciprocal of the shading pattern

$$g(x,y) = f(x,y)/h(x,y)$$

Notes on Arithmetic Operations

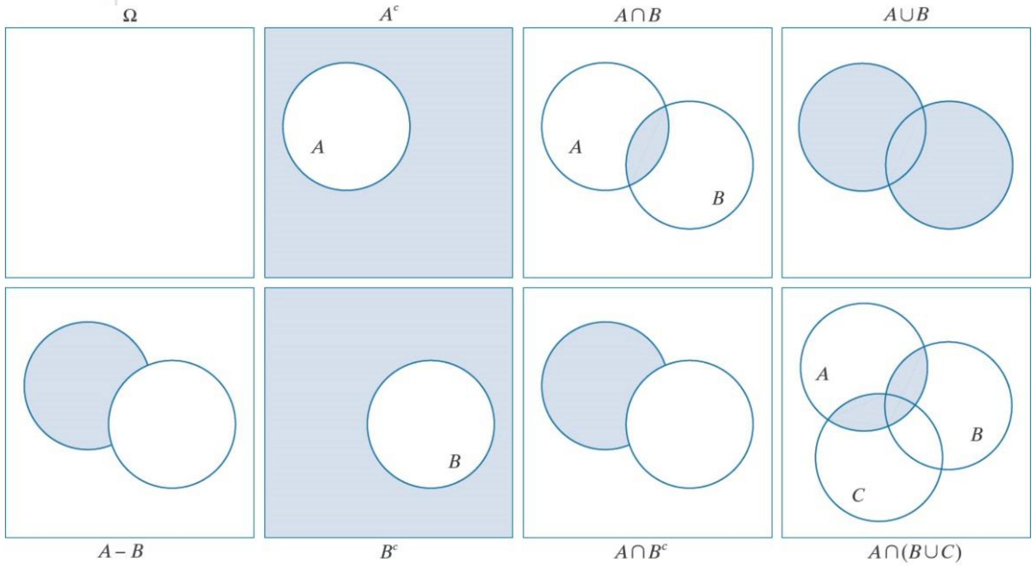
- The images used in averaging & subtraction must be registered!

- The output images should be normalized to the range of [0,255]

$$f_m = f - \min(f)$$
$$f_s = K[f_m / \max(f_m)]$$

Basic Set and Logical Operations

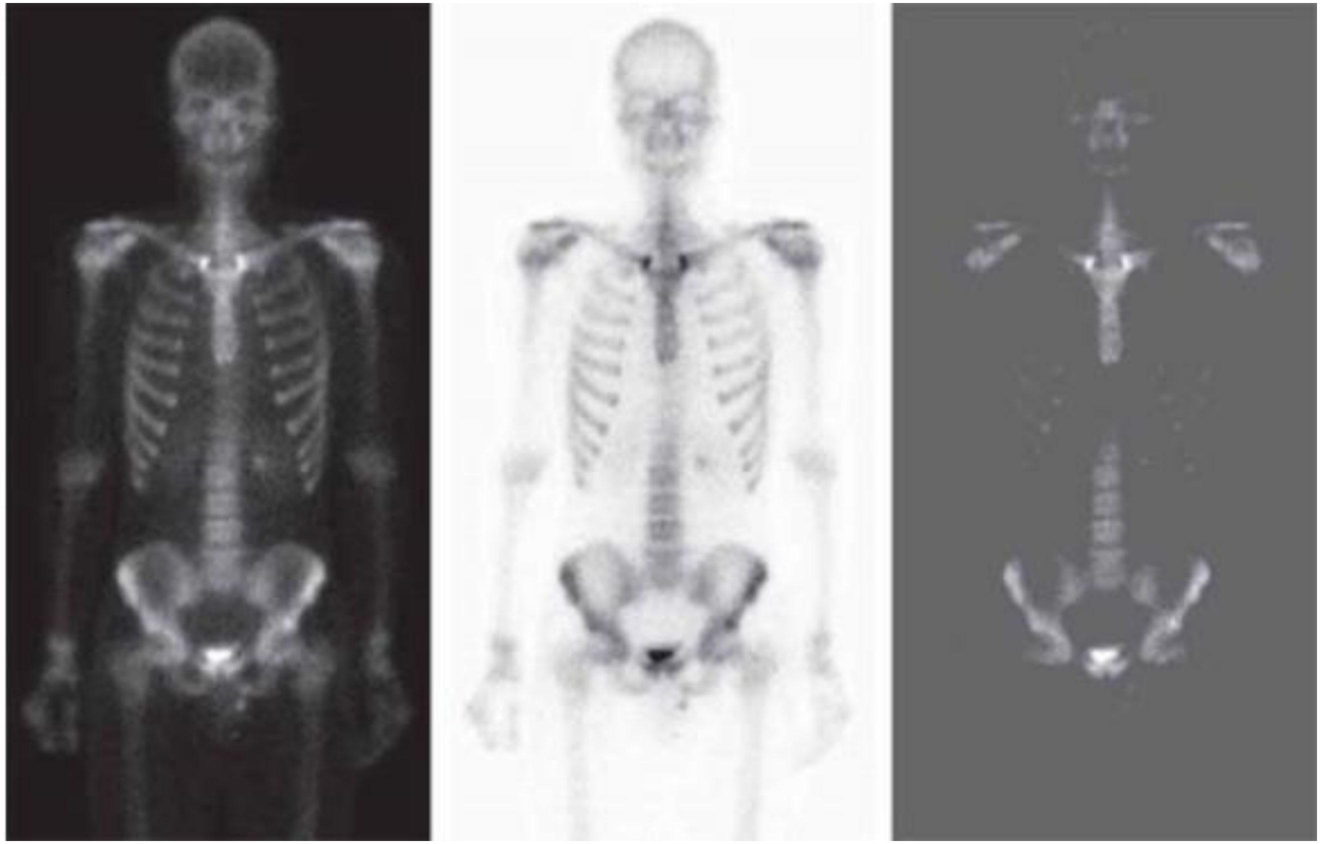
Description	Expressions
Operations between the sample space and null sets	$\Omega^c = \varnothing;\ \varnothing^c = \Omega;\ \Omega \cup \varnothing = \Omega;\ \Omega \cap \varnothing = \varnothing$
Union and intersection with the null and sample space sets	$A \cup \varnothing = A;\ A \cap \varnothing = \varnothing;\ A \cup \Omega = \Omega;\ A \cap \Omega = A$
Union and intersection of a set with itself	$A \cup A = A;\ A \cap A = A$
Union and intersection of a set with its complement	$A \cup A^c = \Omega;\ A \cap A^c = \varnothing$
Commutative laws	$A \cup B = B \cup A$ $A \cap B = B \cap A$
Associative laws	$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$
Distributive laws	$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
DeMorgan's laws	$(A \cup B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \cup B^c$



Venn diagrams corresponding to some of the set operations

Set Operations Based on Intensities

$$A^c = \{(x, y, K - z) | (x, y, z) \in A\}$$



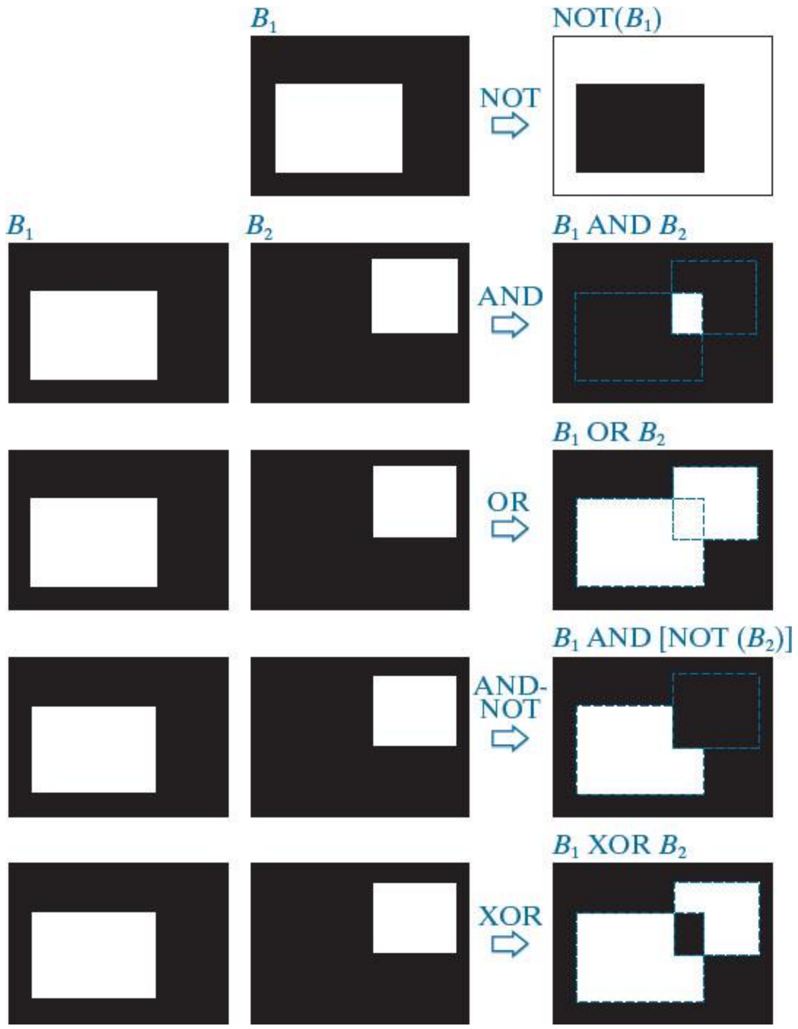
(a) Original image.
(b) Image negative obtained using grayscale set complementation.
(c) The union of image (a) and a constant image.

$$A \cup B = \left\{\left(x, y, \max(z_a, z_b)\right) \middle| (x, y, z_a) \in A, (x, y, z_b) \in B\right\}$$

Logic Operations for Binary Image

- Logic operations will be used a lot in morphological image processing

a	b	a AND b	a OR b	NOT(a)
0	0	0	0	1
0	1	0	1	1
1	0	0	1	0
1	1	1	1	0



Spatial Operations

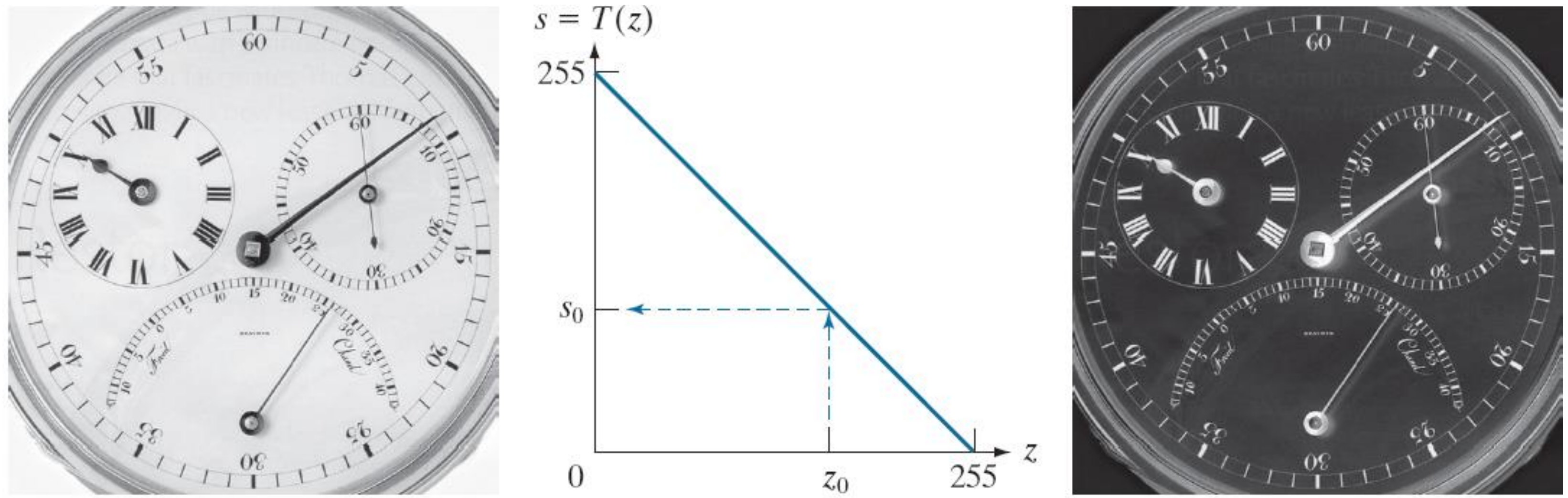
- Perform directly on the pixels of the given image
  - Intensity transformation – change the intensity
  - Single pixel operations $s=T(z)$
  - Neighborhood operations

- Geometric spatial transformations – change the coordinates

Single pixel operations

- Determined by
  - Transformation function T
  - Input intensity value

- Not depend on other pixels and position



a b c

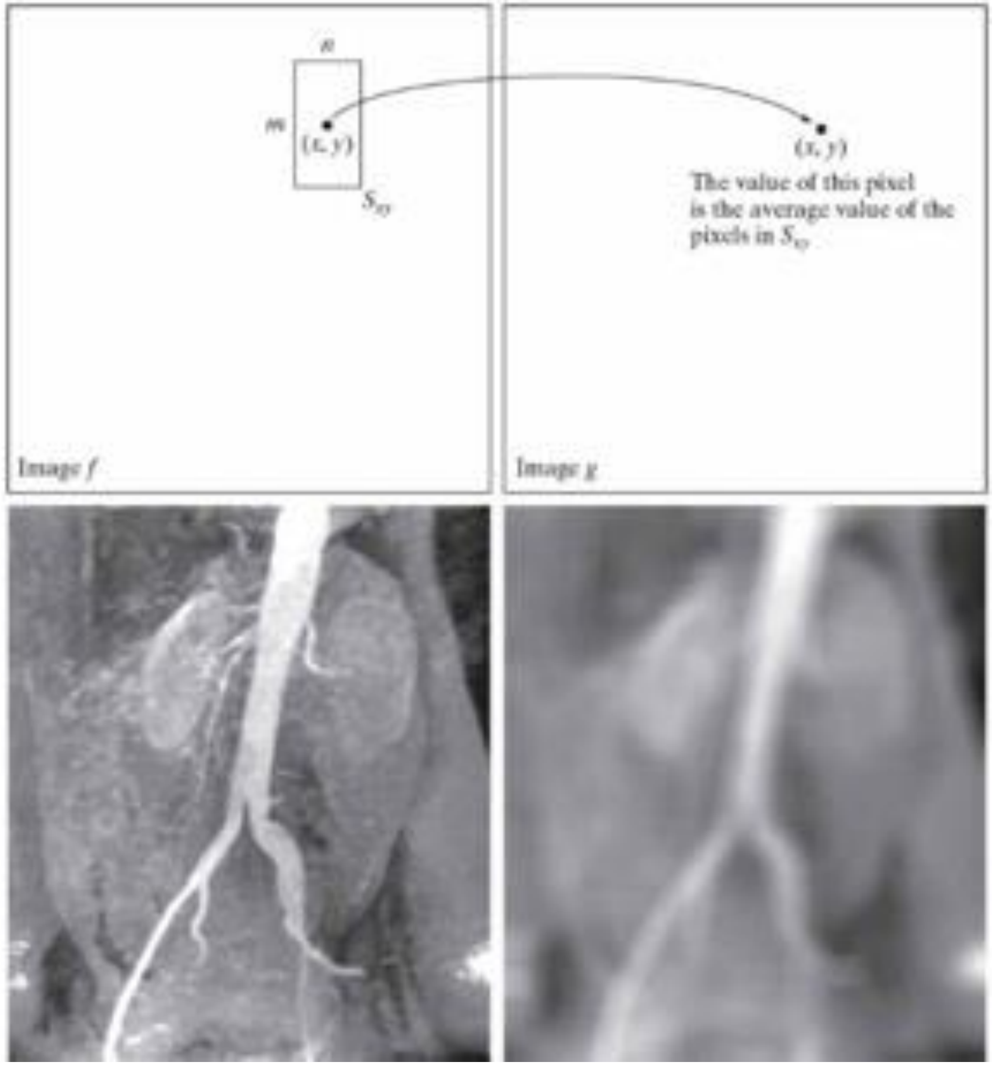
Negative of (a)

Neighborhood Operations

- Image smoothing

$$g(x, y) = \frac{1}{mn} \sum_{(r,c) \in S_{xy}} f(r, c)$$

- Other examples:
  - Interpolation
  - Image filtering



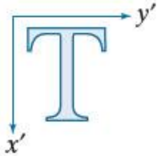
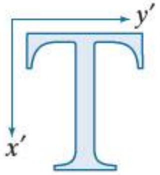
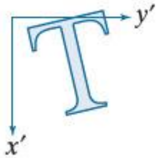
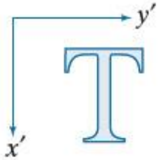
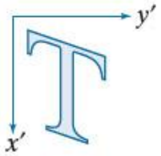
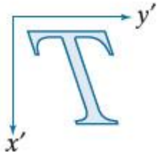
Geometric Spatial Transformations – Rubber Sheet Transformation

- $(x, y) = T\{(v, w)\}$
- Affine transformation

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{T} \begin{bmatrix} v \\ w \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & 1 \end{bmatrix} \begin{bmatrix} v \\ w \\ 1 \end{bmatrix}$$

- Inverse mapping

$$\begin{bmatrix} v \\ w \\ 1 \end{bmatrix} = \mathbf{T}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Transformation Name	Affine Matrix, A	Coordinate Equations	Example
Identity	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x$ $y' = y$	
Scaling/Reflection (For reflection, set one scaling factor to −1 and the other to 0)	$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = c_x x$ $y' = c_y y$	
Rotation (about the origin)	$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x\cos\theta - y\sin\theta$ $y' = x\sin\theta + y\cos\theta$	
Translation	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x + t_x$ $y' = y + t_y$	
Shear (vertical)	$\begin{bmatrix} 1 & s_v & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x + s_v y$ $y' = y$	
Shear (horizontal)	$\begin{bmatrix} 1 & 0 & 0 \\ s_h & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x$ $y' = s_h x + y$	

Geometric Spatial Transformations

- Note: a neighborhood operation, i.e., interpolation, is required following geometric transformation

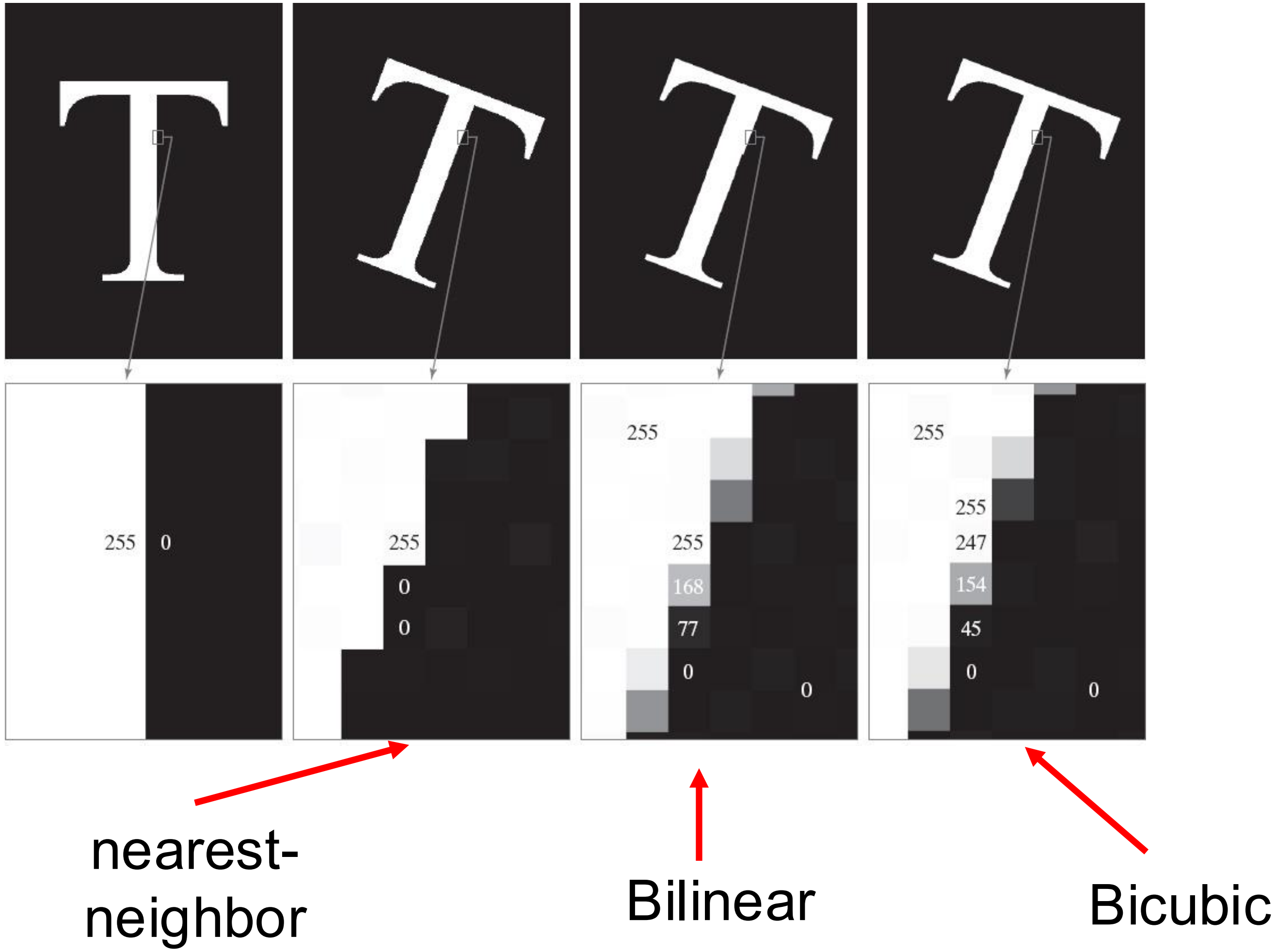


Image Registration

- Compensate the geometric change in:
  - view angle
  - distance
  - orientation
  - sensor resolution
  - object motion

- Four major steps:
  - Feature detection
  - Feature matching
  - Transformation model
  - Resampling

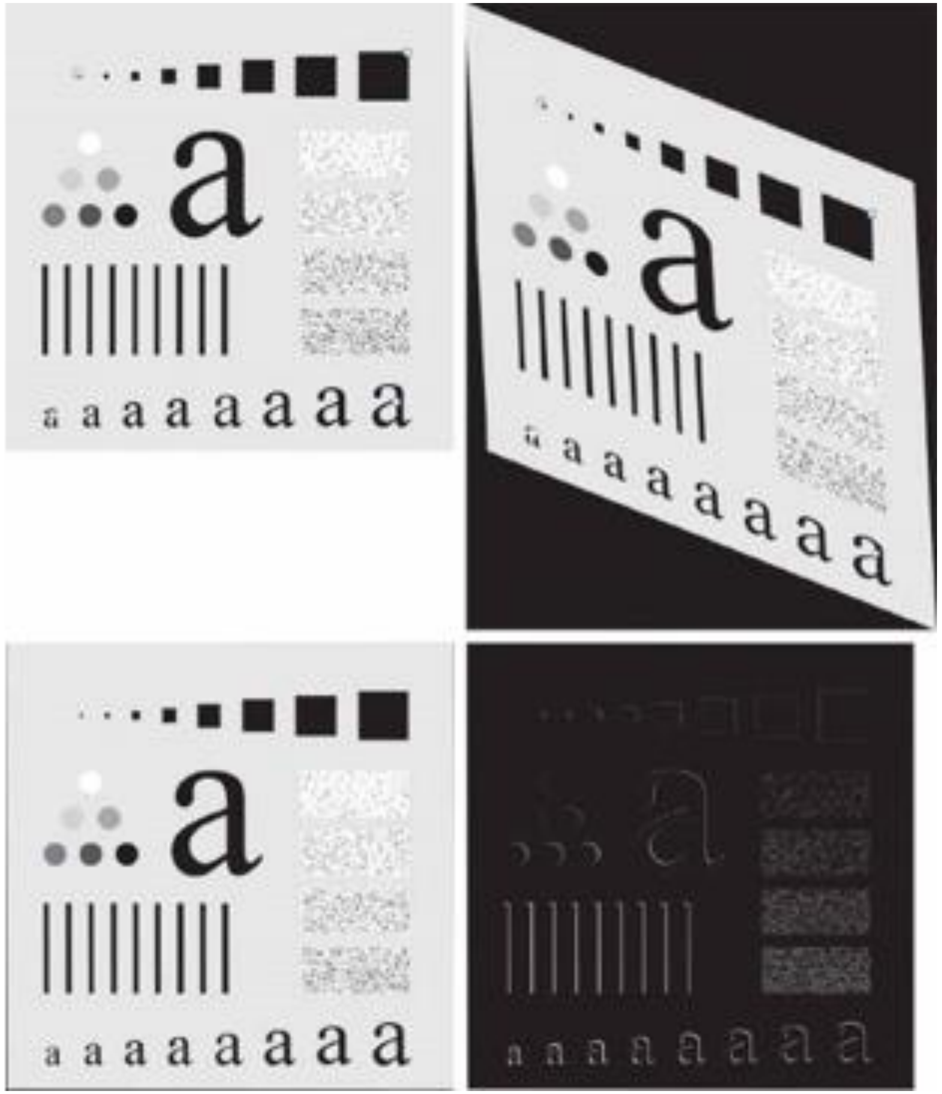
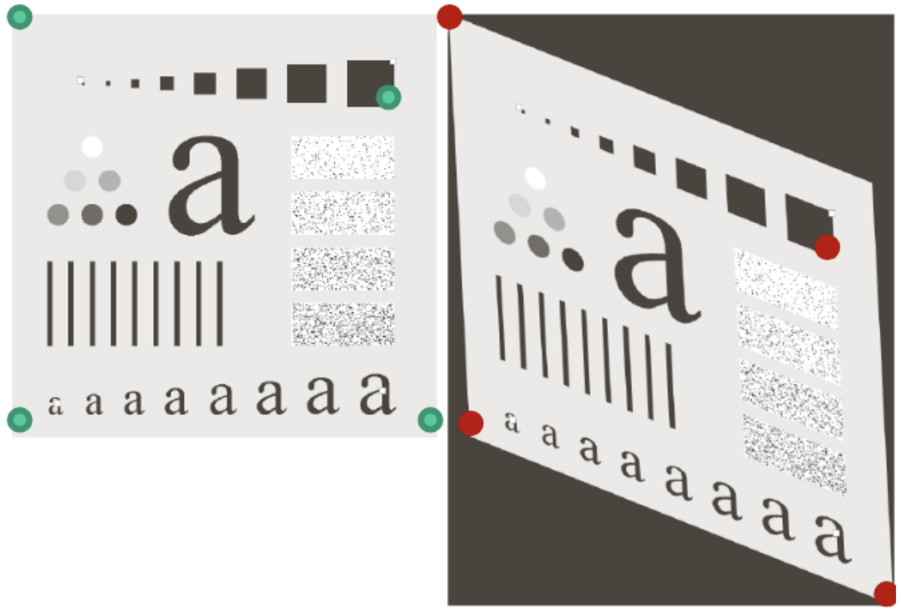


Image Registration

- Coordinates in the moving image (v, w)
- Coordinates in the template image (x, y)



$$x = c_1 v + c_2 w + c_3 vw + c_4$$

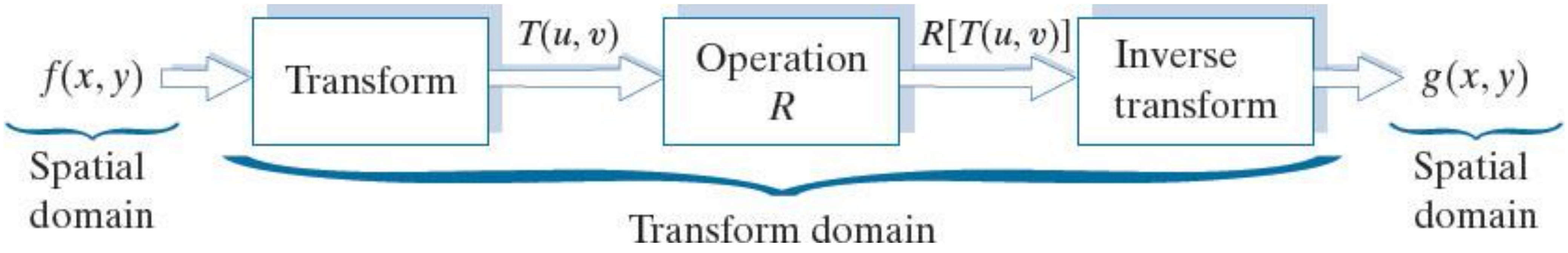
$$y = c_5 v + c_6 w + c_7 vw + c_8$$

- Known: coordinates of the points (x, y) and (v, w)
- Unknown: $\boldsymbol{c1}$ to $\boldsymbol{c8}$

4 tie points → 8 equations

Spatial-Frequency Domain Transformation

- General approach for working in the linear transform domain.



Forward transformation kernel

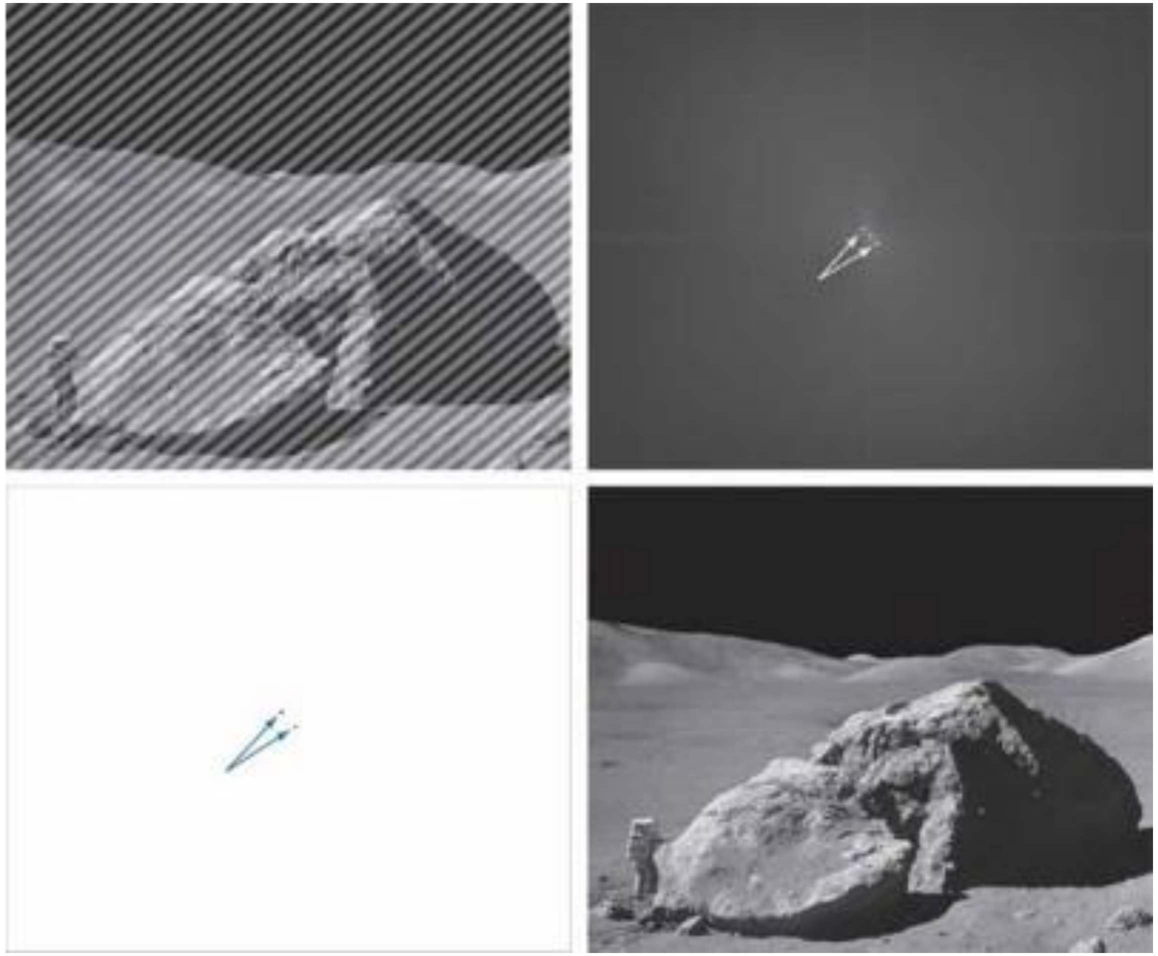
$$T(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\boxed{r(x,y,u,v)}, \qquad \begin{array}{l} u = 0,1,\ldots, M-1 \\ v = 0,1,\ldots, N-1 \end{array}$$

$$f(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} T(u,v)\boxed{s(x,y,u,v)} \qquad \begin{array}{l} x = 0,1,\ldots, M-1 \\ y = 0,1,\ldots, N-1 \end{array}$$

Inverse transformation kernel

Fourier Transforms and Filtering

(a) Image corrupted by sinusoidal interference.



(b) Magnitude of the Fourier transform showing the bursts of energy caused by the interference (the bursts were enlarged for display purposes

(c) Mask used to eliminate the energy bursts

(d) Result of computing the inverse of the modified Fourier transform.

Probability Methods

z_k is the kth intensity value
n_k is the number of pixels having the intensity value z_k

Probability of an intensity value

$$p(z_k) = \frac{n_k}{MN}, \quad \sum_{k=1}^{L-1} p(z_k) = 1$$

Probability Methods

$$m = \sum_{k=1}^{L-1} z_k p(z_k), \quad \sigma^2 = \sum_{k=1}^{L-1} (z_k - m)^2 p(z_k)$$

What do they mean?

$$\mu_n(z) = \sum_{k=1}^{L-1} (z_k - m)^n p(z_k)$$

n^{th} moment of z

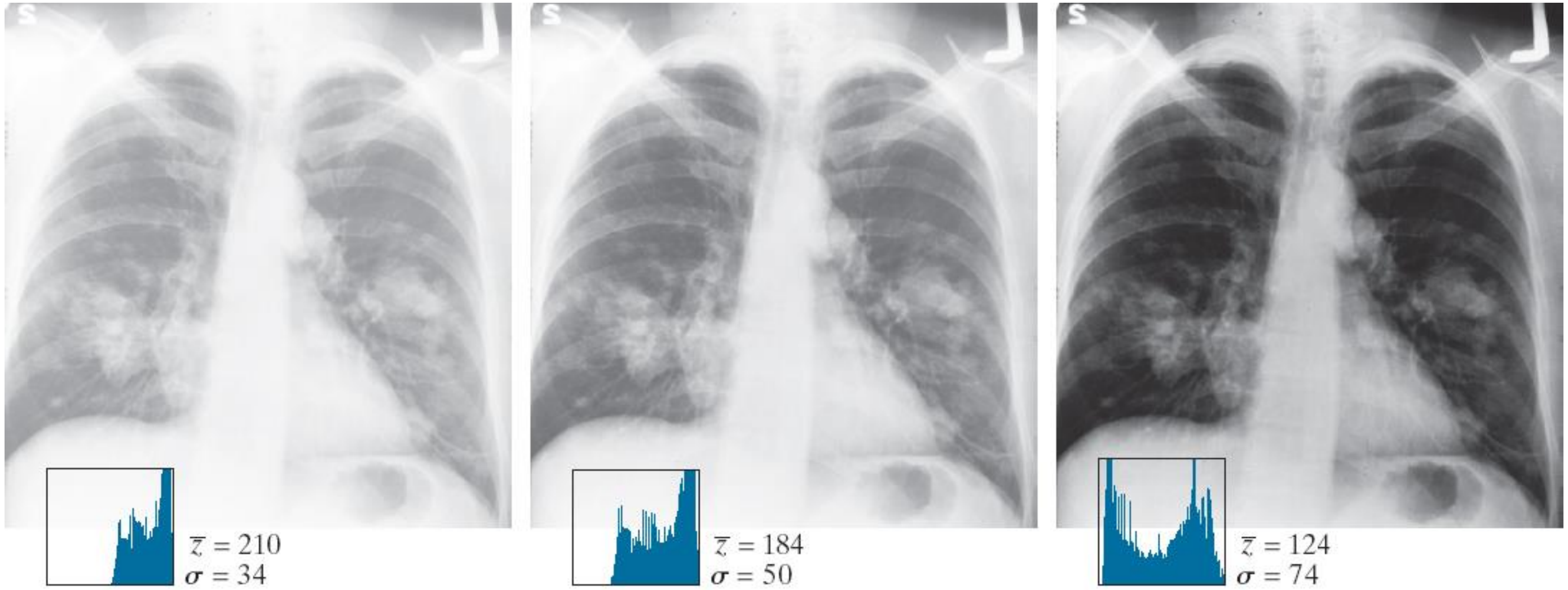
Probability Methods

$$m = \sum_{k=1}^{L-1} z_k p(z_k), \quad \sigma^2 = \sum_{k=1}^{L-1} (z_k - m)^2 p(z_k)$$

What do they mean?

$$\mu_n(z) = \sum_{k=1}^{L-1} (z_k - m)^n p(z_k)$$

n^{th} moment of z



a b c

Stochastic Image-Sequence Processing

- Using probability and random-process tools
- Each pixel is a random event → each image frame is a random event, related to time
- Probability plays a central role in modern image processing and computer vision