Image and Video Processing

Dr. Badri Narayan Subudhi
Department of Electrical Engineering
Indian Institute of Technology Jammu
subudhi.badri@iitjammu.ac.in

Module no.	Topic	No. of hours
1.	**Fundamentals of Image Processing:** Introduction, Image sampling, Quantization, Resolution, Image file formats, Elements of image processing system, Applications of Digital image processing.	3
2.	**Image Enhancement:** Spatial domain methods: Histogram processing, Fundamentals of Spatial filtering, Smoothingspatial filters, Sharpening spatial filters. Frequency domain methods: Basics of filtering in frequency domain, image smoothing, image sharpening, Selective filtering.	6
3.	**Image Restoration:** Introduction to Image restoration, Image degradation, Types of image blur, Classification of image restoration techniques, Image restoration model, Linear and Nonlinear image restoration techniques.	5
4.	**Image Segmentation:** Introduction to image segmentation, Point, Line and Edge Detection, Region based segmentation., Classification of segmentation techniques, Region approach to image segmentation, clustering techniques, Image segmentation based on thresholding, Edge based segmentation	6
5.	**Image Compression:** Introduction, Need for image compression, Redundancy in images, Classification of redundancy in images, image compression scheme, Classification of image compression schemes, Image compression standard, JPEG Standards.	5
6.	**Basic Steps of Video Processing:** Analog Video, Digital Video. Time-Varying Image Formation models: Three-Dimensional Motion Models, Geometric Image Formation, Photometric Image Formation, Sampling of Videosignals, Filtering operations	6
7.	**2-D Motion Estimation:** Optical flow, General Methodologies, Pixel Based Motion Estimation, Block Matching Algorithm, Mesh based Motion Estimation, Global Motion Estimation, Region based Motion Estimation.	6
8.	**Visual scene analysis:** Basics of background modeling and foreground detection connected component labeling, shot boundary detection	5
	Total Lecture hours	**42**

Human Perception

Humans have developed highly sophisticated skills for sensing their environment and taking actions according to what they observe, e.g.,

– Recognizing a face.

– Understanding spoken words.

– Reading handwriting.

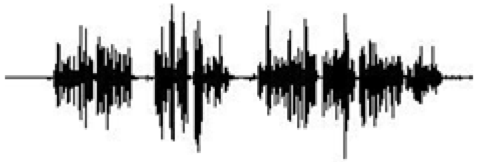
– Distinguishing fresh food from its smell.

Machine Perception

- Build a machine that can recognize patterns:
  - Speech recognition
  - Fingerprint identification
  - OCR (Optical Character Recognition)
  - DNA sequence identification

Intelligence Mechanism

- Speech

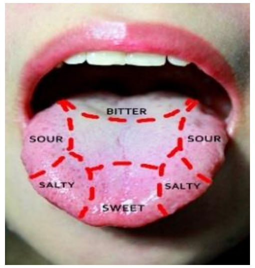


- Vision

- Touch

- Smell

- Taste



When Human Fails?

- A human can perfor

However,

  - It is quite time consu
  - It may be Erroneous

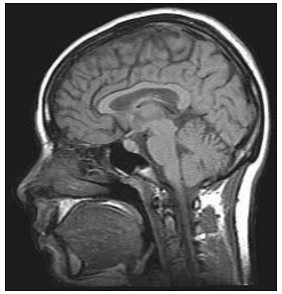
What is Computer Vision?

- **Computer vision** is the science and technology of machines that see.

- Concerned with the theory for building artificial systems that obtain information from images.

- The image data can take many forms, such as a video sequence, depth images, views from multiple cameras, or multi-dimensional data from a medical scanner





Computer Vision

Make computers understand images and videos.

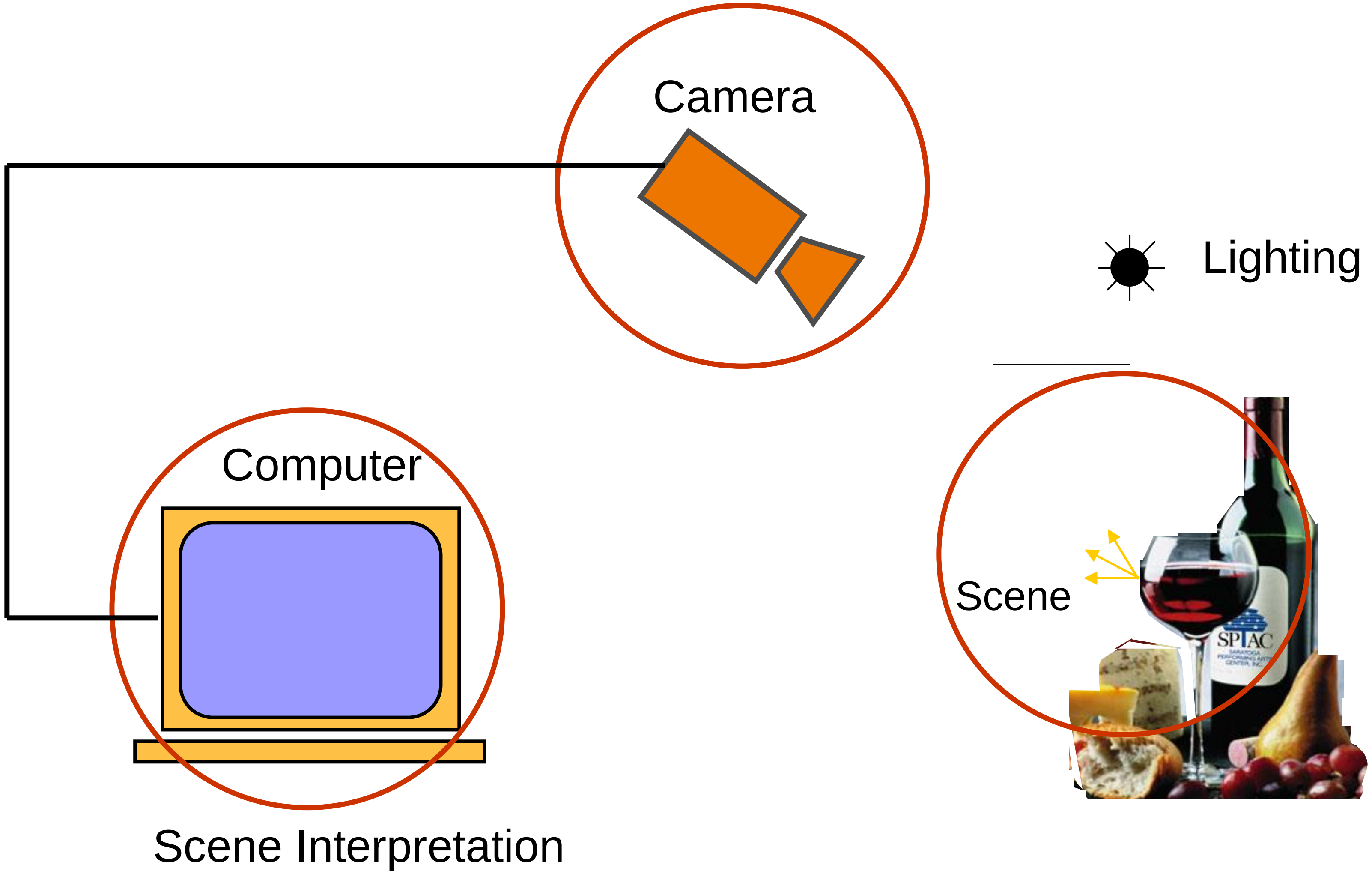
What kind of scene?

Where are the cars?

How far is the building?

…

Components of a computer vision system



Srinivasa Narasimhan's slide

Computer vision vs human vision

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What we see

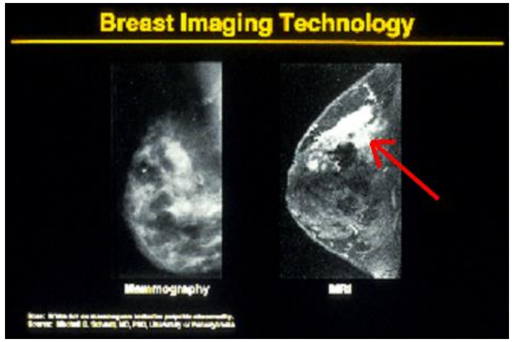
What a computer sees

Vision is really hard

- Vision is an amazing feat of natural intelligence
  - Visual cortex occupies about 50% of a brain
  - More human brain devoted to vision than anything else

Why computer vision matters

Safety

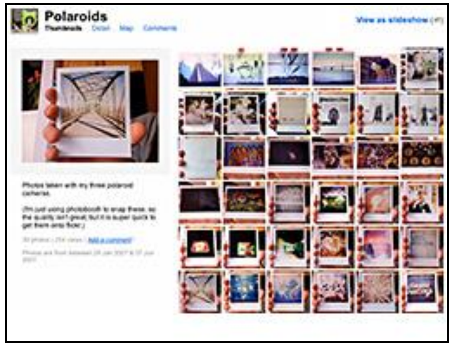


Health

Security

Comfort

Fun

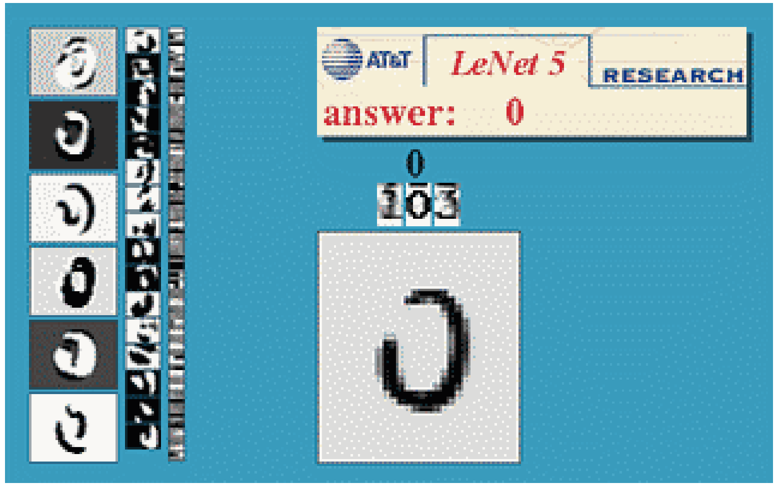


Access

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



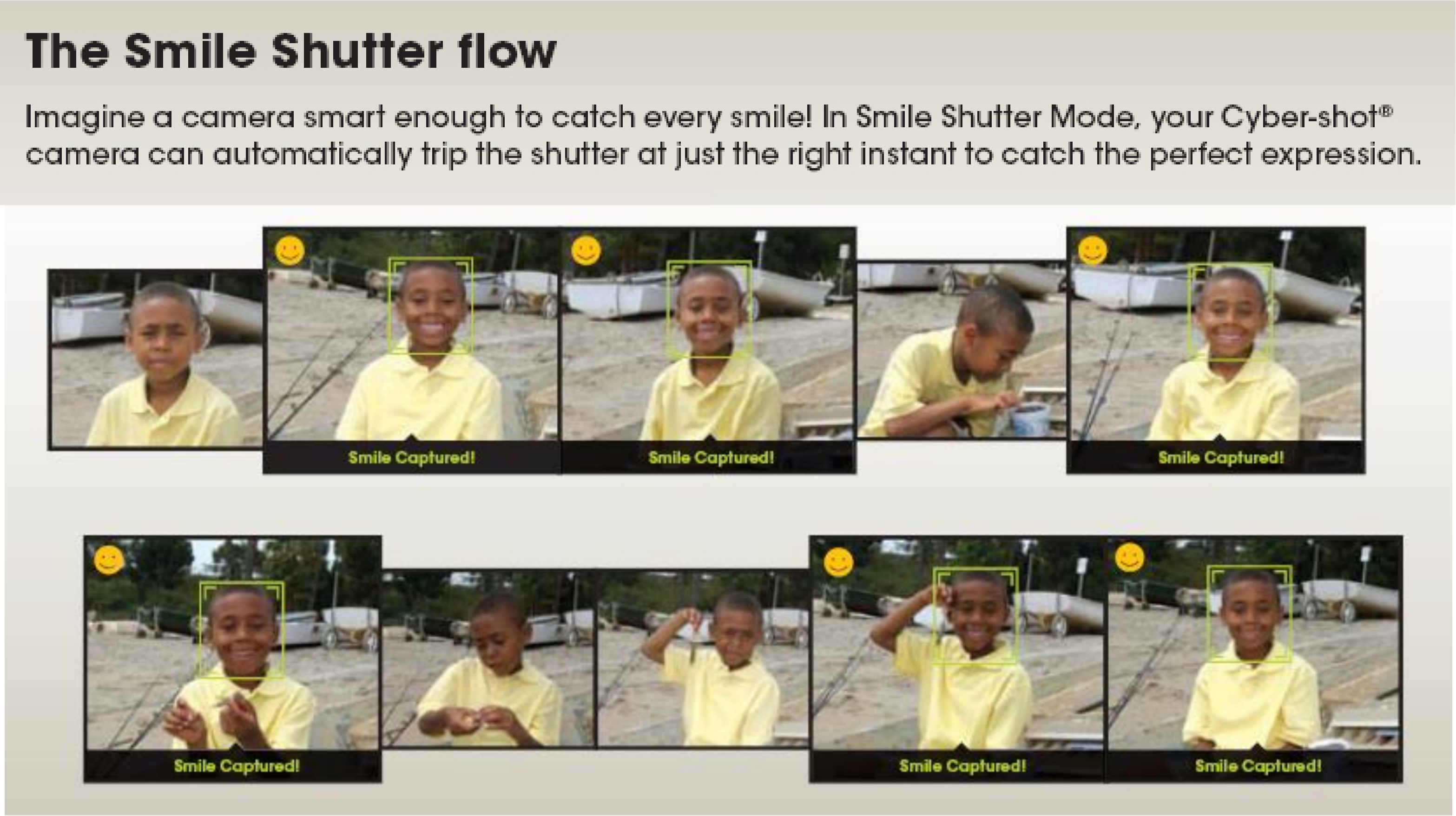
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Smile detection



[Sony Cyber-shot® T70 Digital Still Camera](#)

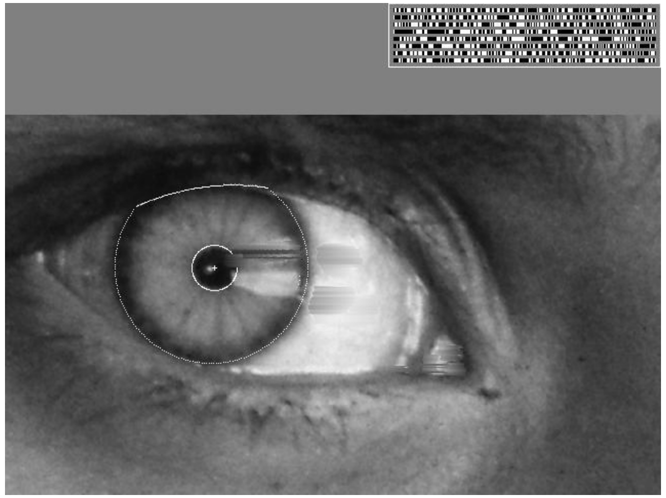
Object recognition (in supermarkets)

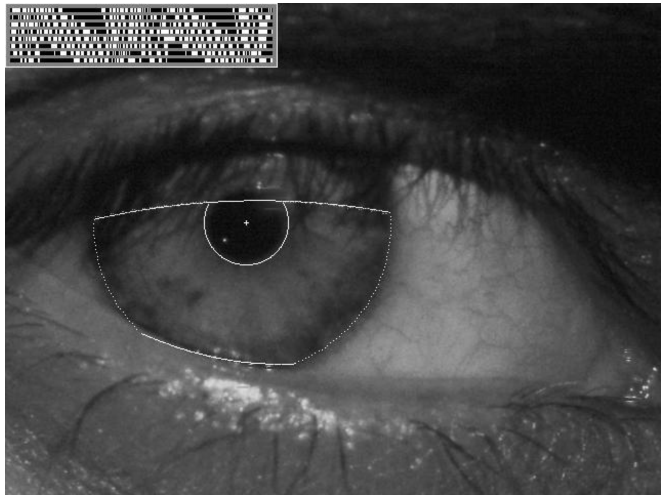
[LaneHawk by EvolutionRobotics]

"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… "

Vision-based biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story
wikipedia





Login without a password...

Fingerprint scanners on
many new laptops,
other devices

Face recognition systems now
beginning to appear more widely
http://www.sensiblevision.com/

Object recognition (in mobile phones)

[Point & Find](), [Nokia]()
[Google Goggles]()

Special effects: shape capture

The Matrix movies, ESC Entertainment

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

http://www.sportvision.com/video.html

Smart cars

Slide content courtesy of Amnon Shashua

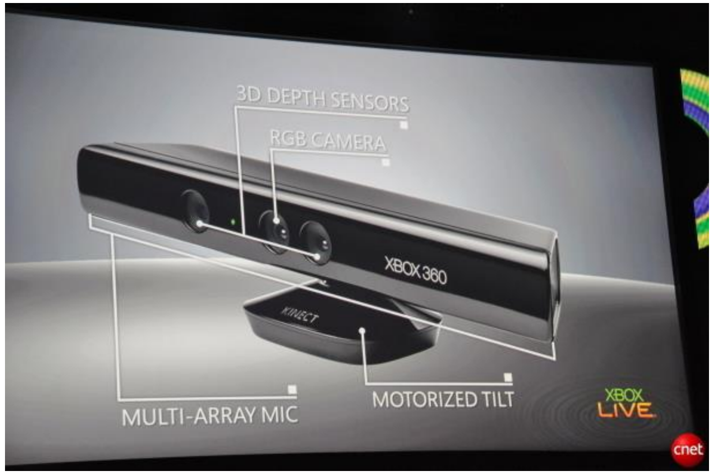
- Mobileye [wiki article]
  - Vision systems currently in high-end BMW, GM, Volvo models
  - By 2010: 70% of car manufacturers.

Google cars

http://www.nytimes.com/2010/10/10/science/10google.html?ref=artificialintelligence

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY
- 3D tracking, reconstruction, and interaction: http://research.microsoft.com/en-us/projects/surfacerecon/default.aspx



Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

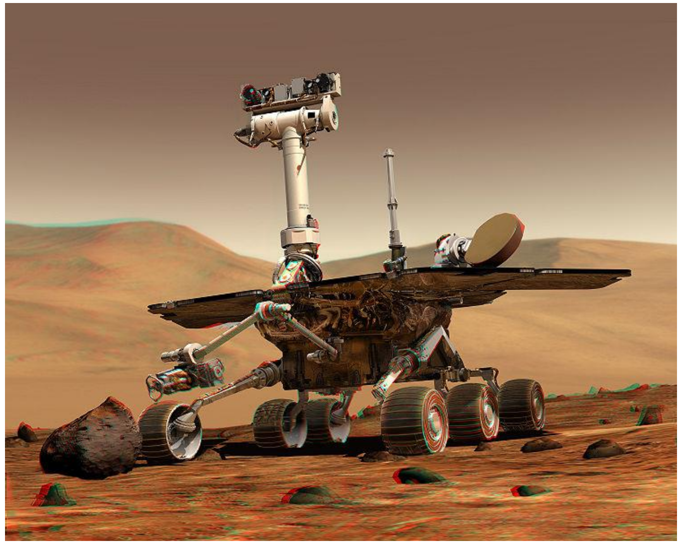
Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “Computer Vision on Mars” by Matthies et al.

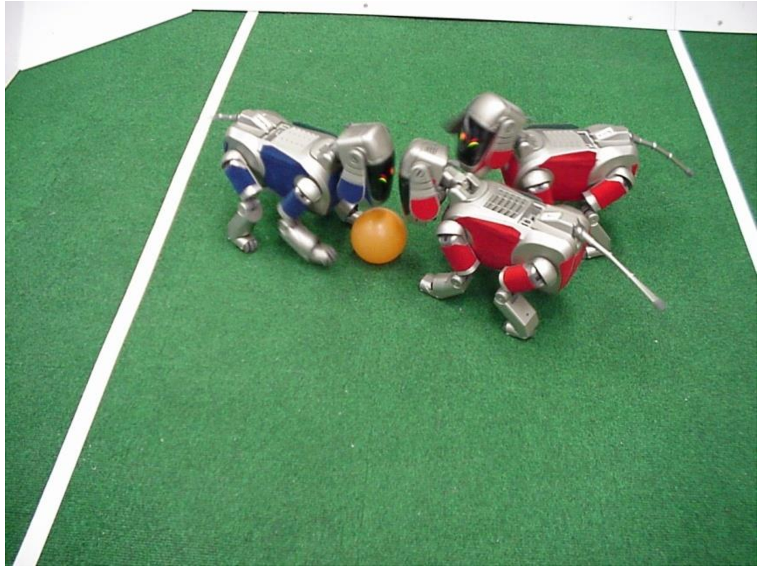
Industrial robots

Vision-guided robots position nut runners on wheels

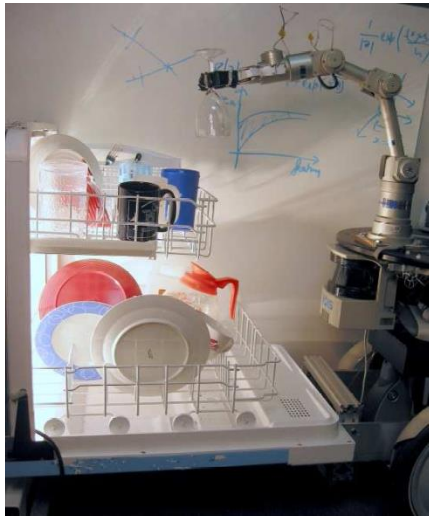
Mobile robots



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

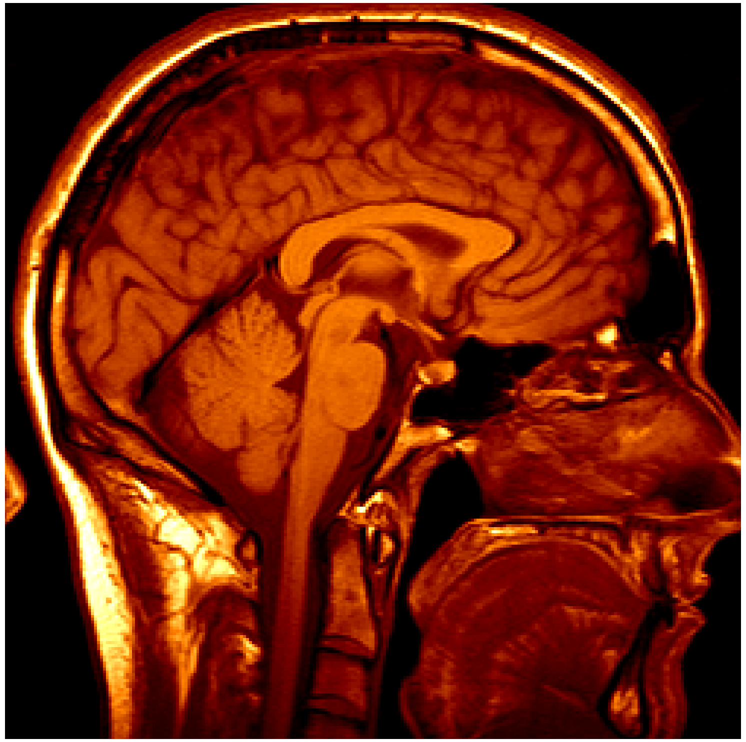


http://www.robocup.org/



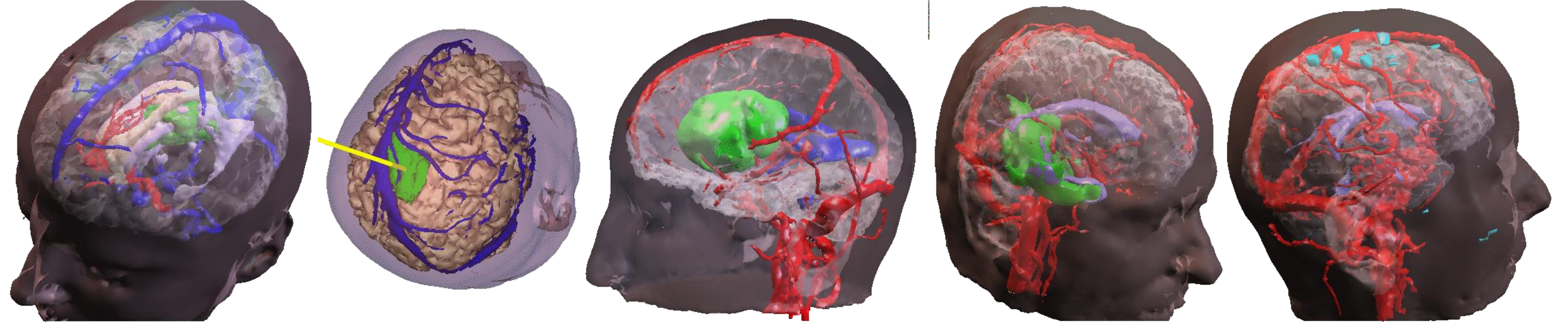
Saxena et al. 2008
STAIR at Stanford

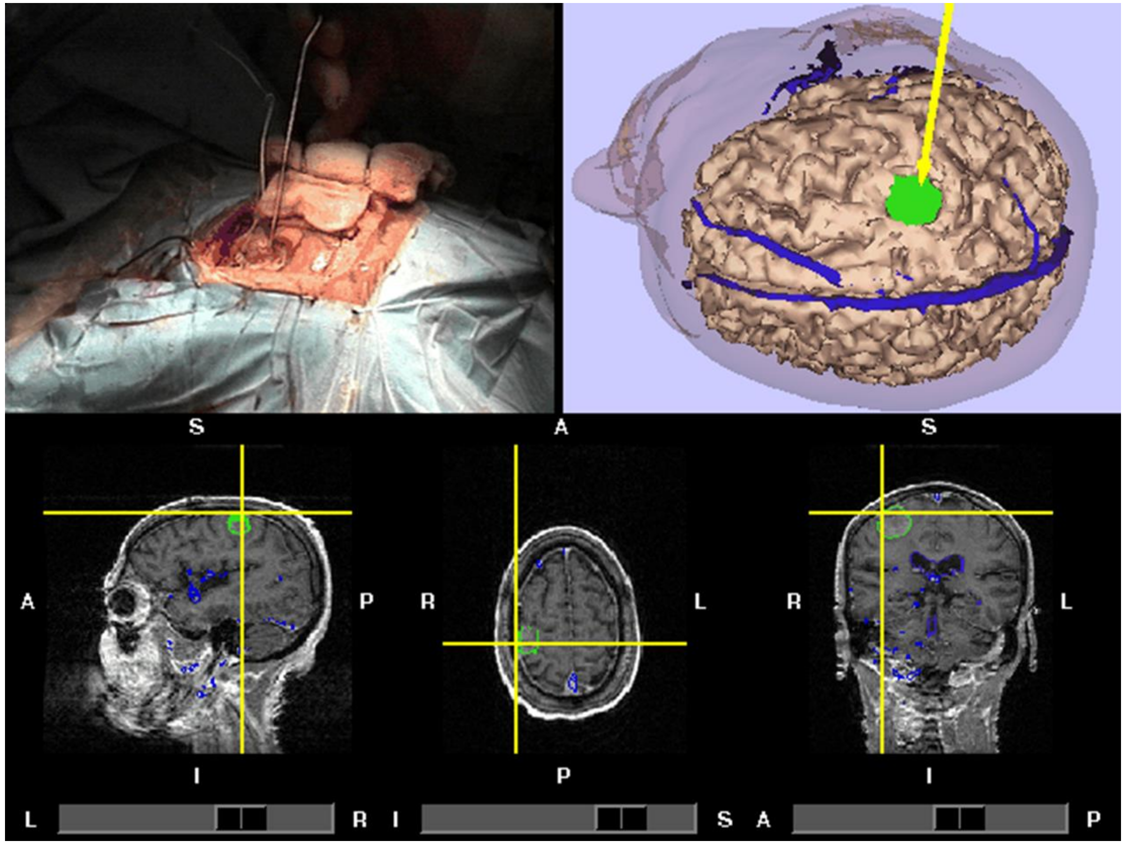
Medical imaging



3D imaging
MRI, CT

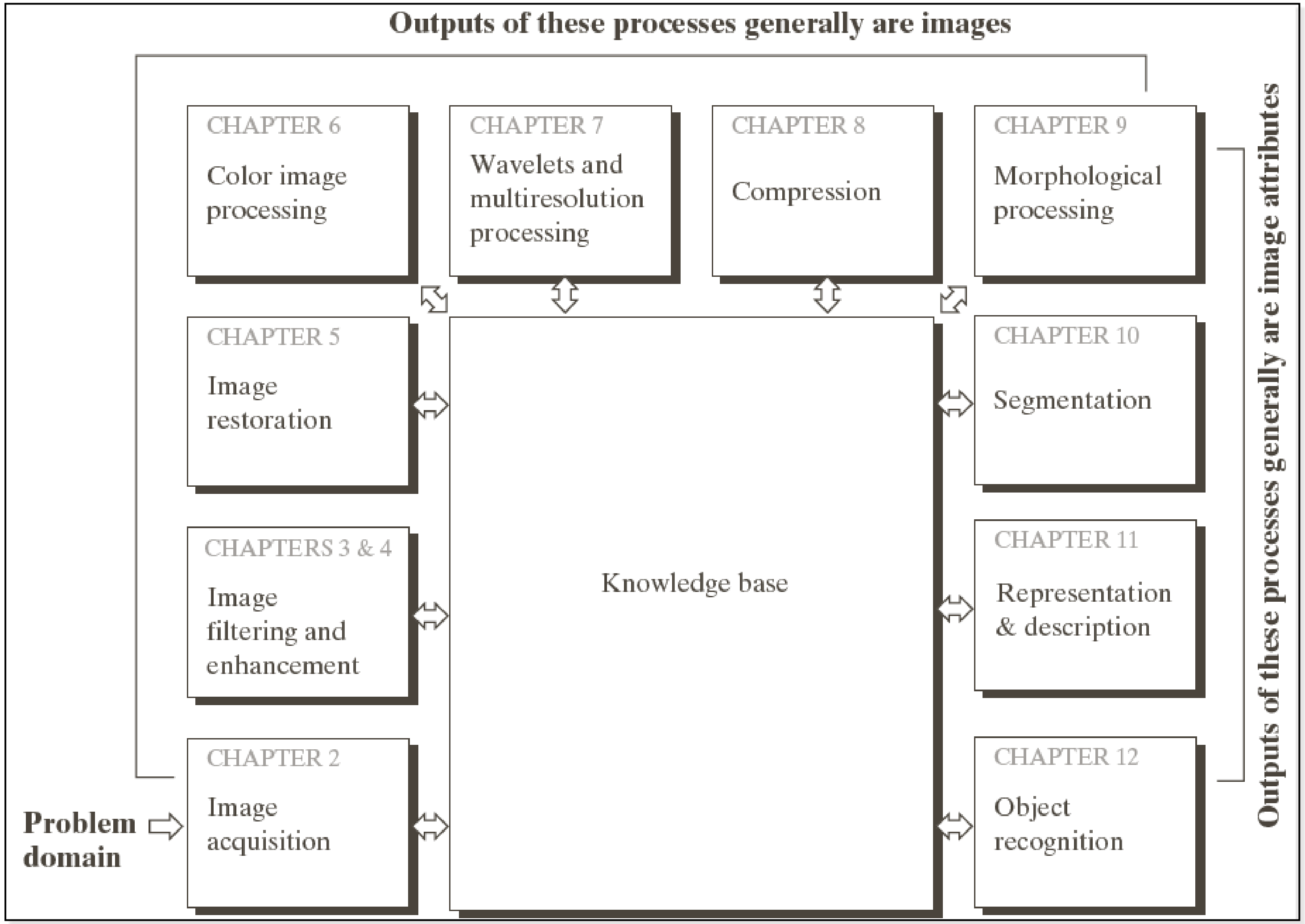
Image guided surgery
[Grimson et al., MIT]




S
A
S
A
P
R
L
R
L
I
P
I
L
R
I
S
A
P

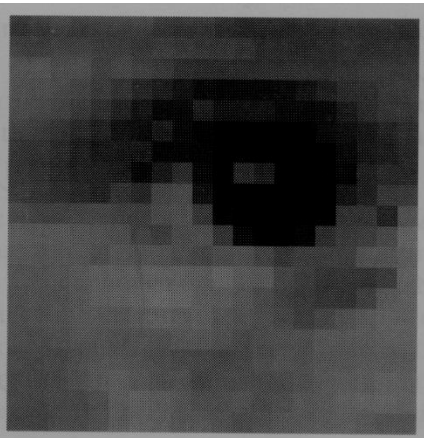
Visual effects

BASIC IMAGE PROCESSING BLOCKS



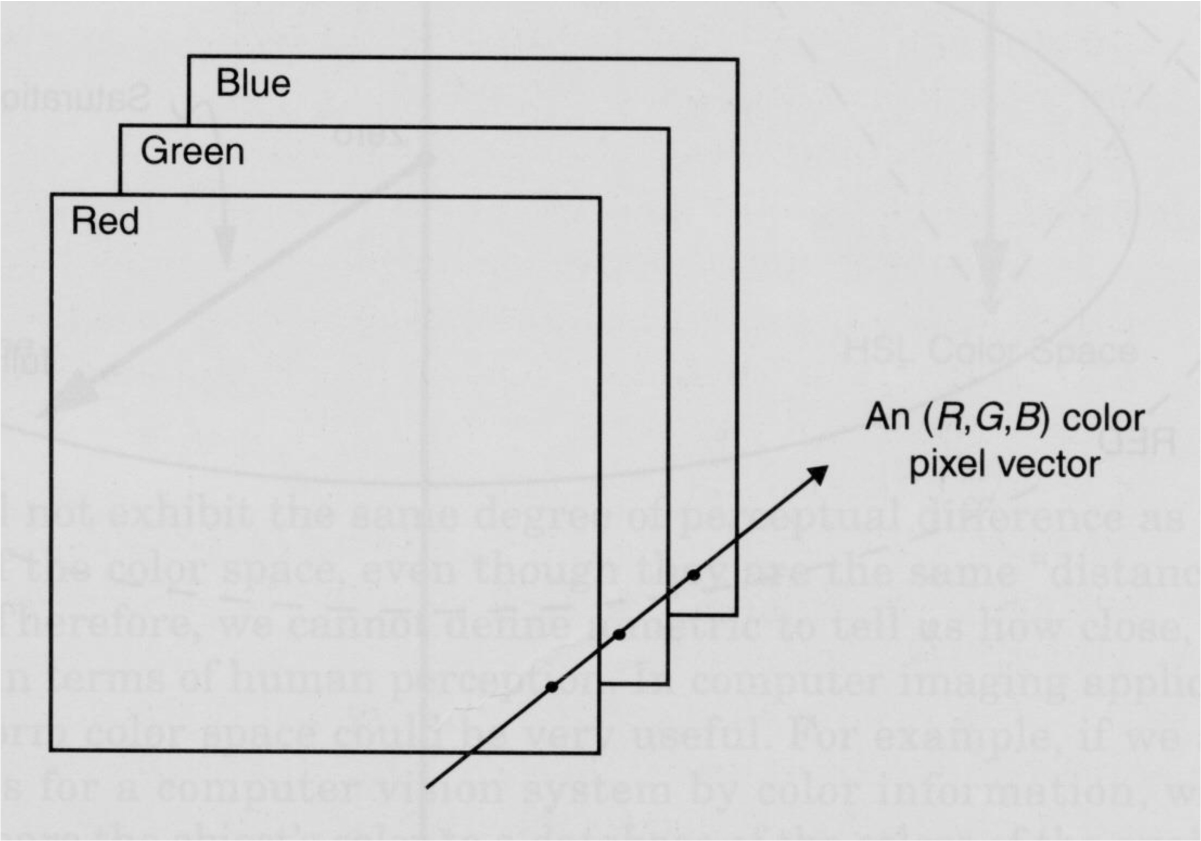
Fundamental steps in digital image processing. The chapter(s) indicated in the boxes is where the material described in the box is discussed.

How are images represented in the computer?



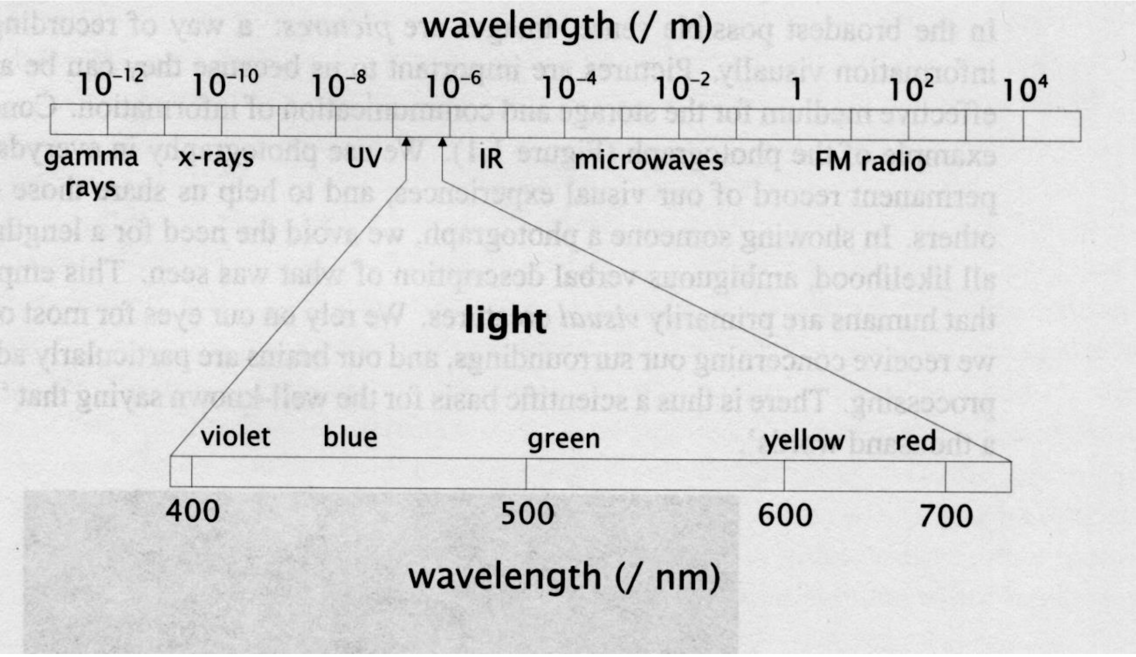
117	125	133	127	130	130	133	121	116	115	100	91	93	94	99	103	112	105	109	106
134	133	138	138	132	134	130	133	128	123	121	113	106	102	99	106	113	109	109	113
146	147	138	140	125	134	124	115	102	96	93	94	99	96	99	100	103	110	109	110
144	141	136	130	120	108	88	74	53	37	31	37	35	39	53	79	93	100	109	116
139	136	129	119	102	85	58	31	41	77	51	53	53	33	37	41	69	94	105	108
132	127	117	102	87	57	49	77	42	28	17	15	13	13	17	41	53	69	88	100
124	120	108	94	72	74	72	31	35	31	15	13	15	11	15	13	46	75	83	96
125	115	102	93	88	82	42	79	113	41	19	100	82	11	11	17	31	91	99	100
124	116	109	99	91	113	99	140	144	57	20	20	15	11	15	17	63	87	119	124
136	133	133	135	138	133	132	144	150	120	24	17	15	15	17	20	115	113	88	150
158	157	157	154	149	145	133	127	146	150	116	35	20	19	28	105	124	128	141	171
155	154	156	155	146	155	154	154	147	139	148	150	138	120	128	129	130	151	156	165
150	151	154	162	166	167	169	174	172	167	177	166	164	140	134	120	121	120	127	172
145	149	151	157	165	169	173	179	176	166	166	157	145	136	129	124	120	136	163	168
144	148	153	160	159	158	165	172	165	169	157	151	149	141	130	140	151	162	169	167
144	141	147	155	154	149	156	151	157	157	151	144	147	147	149	159	158	159	166	165
139	140	140	150	153	151	150	146	140	139	138	140	145	151	149	156	156	162	162	161
136	134	138	146	156	164	153	146	145	136	139	139	140	141	149	157	159	161	169	166
136	133	136	135	144	159	168	159	151	142	141	145	139	146	153	156	164	167	172	168
133	129	140	142	146	159	167	165	154	151	146	141	147	154	156	160	161	157	153	154

Color images



What is light?

- The visible portion of the **electromagnetic** (EM) spectrum.
- It occurs between wavelengths of approximately 400 and 700 nanometers.



Short wavelengths

- Different wavelengths of radiation have different properties.
- The **x-ray** region of the spectrum, it carries sufficient energy to penetrate a significant volume or material.

Long wavelengths

- Copious quantities of **infrared** (IR) radiation are emitted from warm objects (e.g., locate people in total darkness).

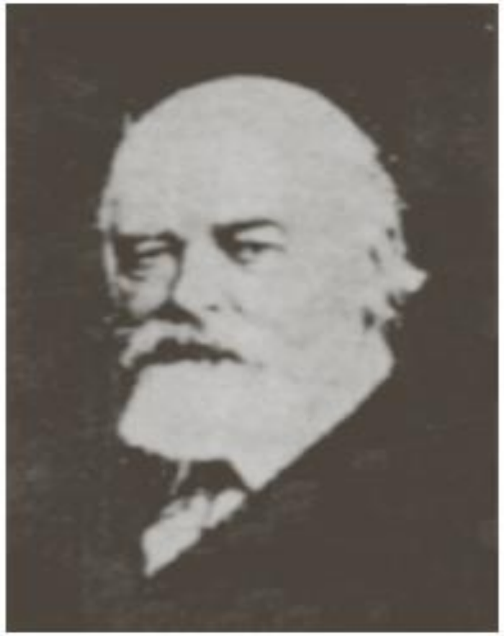
Long wavelengths (cont'd)

- "**Synthetic aperture radar**" (SAR) imaging techniques use an artificially generated source of microwaves to probe a scene.
- SAR is unaffected by weather conditions and clouds (e.g., has provided us images of the surface of Venus).

Sonic images

- Produced by the reflection of sound waves off an object.
- High sound frequencies are used to improve resolution.

A digital picture produced in 1921 from a coded tape by a telegraph printer with special type faces. (McFarlane.[†])



A digital picture made in 1922 from a tape punched after the signals had crossed the Atlantic twice. (McFarlane.)

Unretouched cable picture of Generals Pershing and Foch, transmitted in 1929 from London to New York by 15-tone equipment. (McFarlane.)

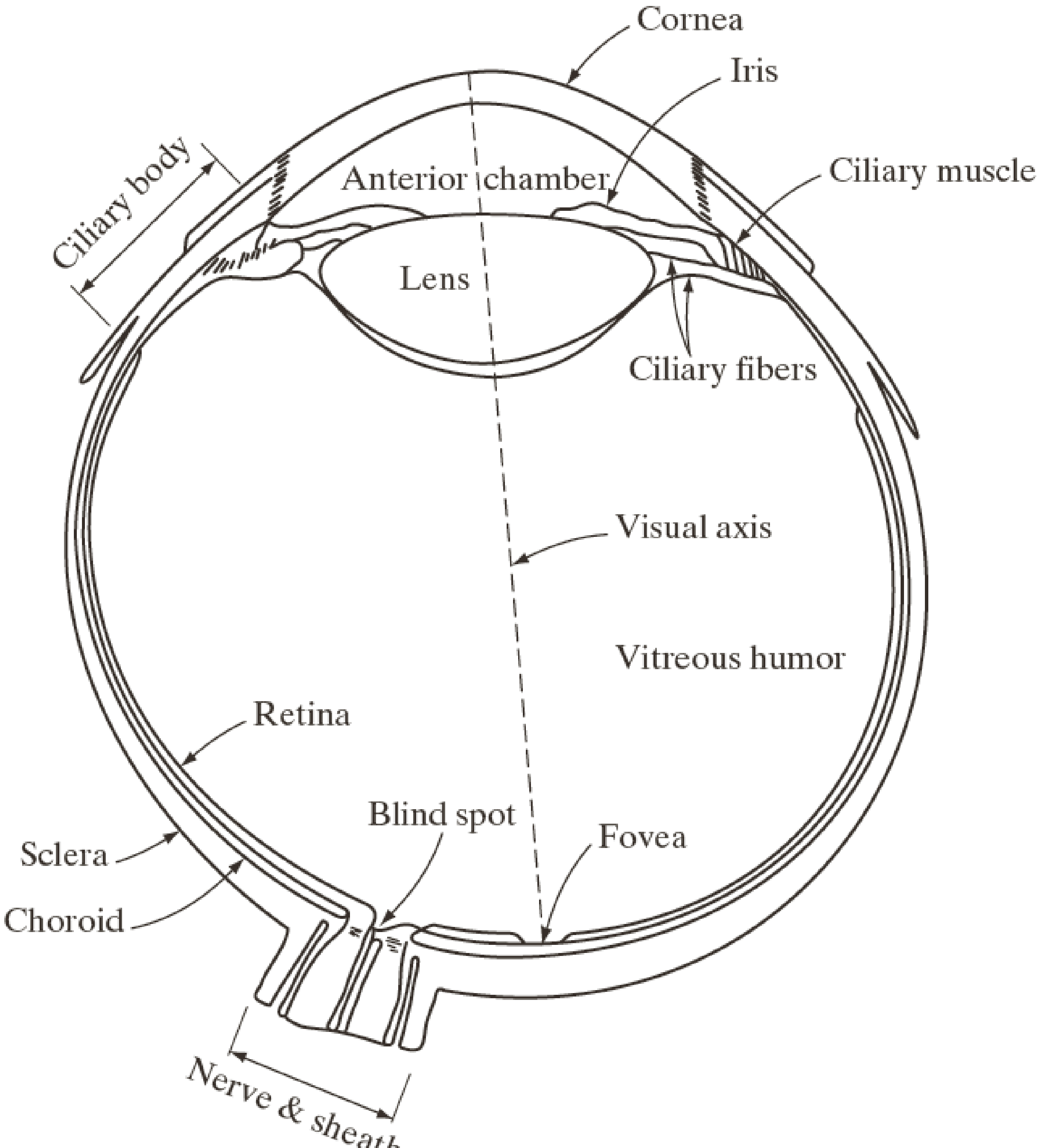
The first picture of the moon by a U.S. spacecraft. *Ranger 7* took this image on July 31, 1964 at 9 : 09 A.M. EDT, about 17 minutes before impacting the lunar surface. (Courtesy of NASA.)

Image formation

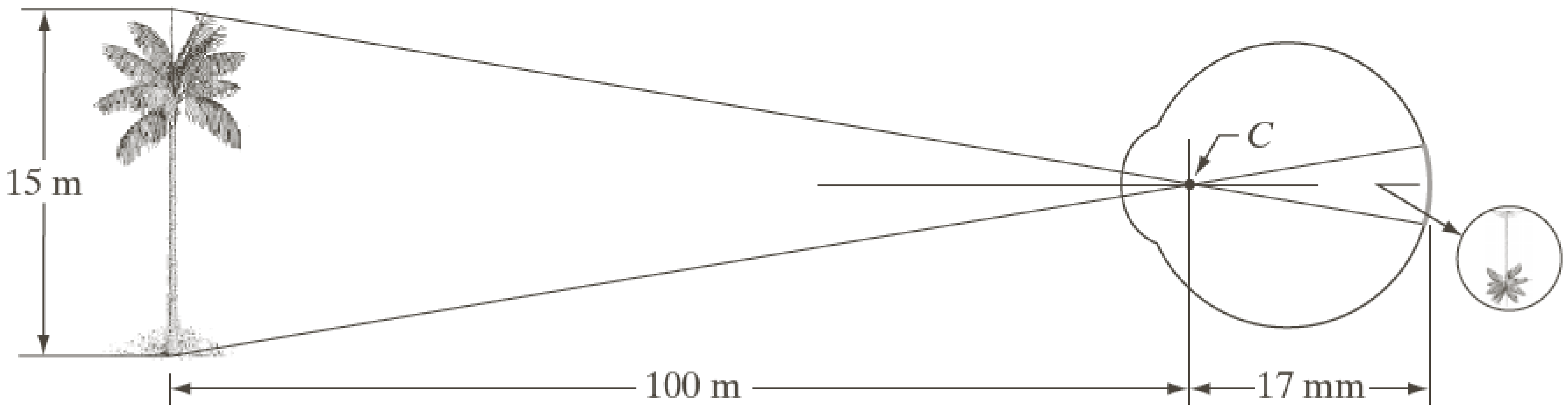
- There are two parts to the image formation process:

  - The **geometry of image formation**, which determines where in the image plane the projection of a point in the scene will be located.

  - The **physics of light**, which determines the brightness of a point in the image plane as a function of illumination and surface properties.



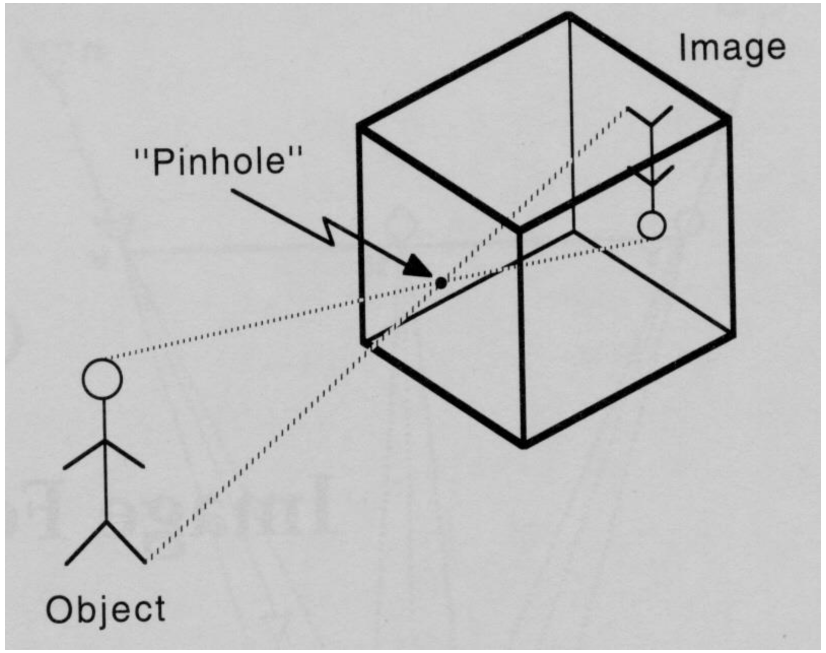
Cornea
Iris
Ciliary body
Anterior chamber
Ciliary muscle
Lens
Ciliary fibers
Visual axis
Vitreous humor
Retina
Blind spot
Fovea
Sclera
Choroid
Nerve & sheath



Graphical representation of the eye looking at a palm tree. Point C is the optical center of the lens.

Pinhole camera

- This is the simplest device to form an image of a 3D scene on a 2D surface.
- Straight rays of light pass through a "pinhole" and form an inverted image of the object on the image plane.

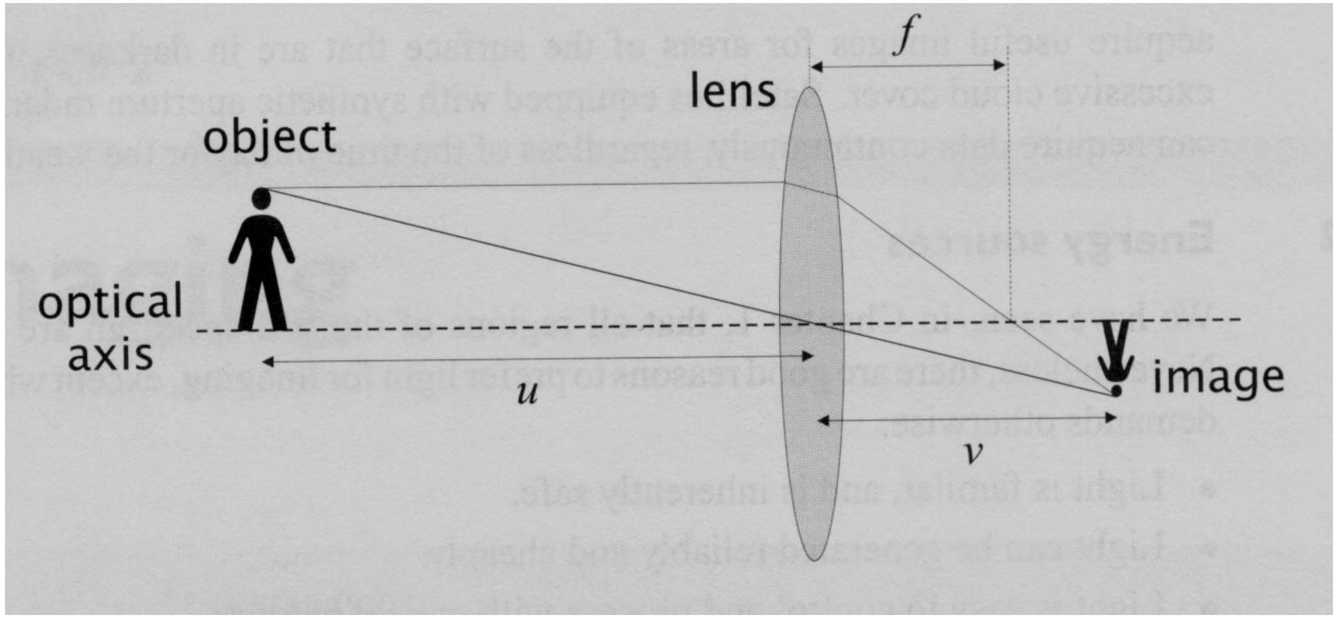


$$x = \frac{fX}{Z}$$

$$y = \frac{fY}{Z}$$

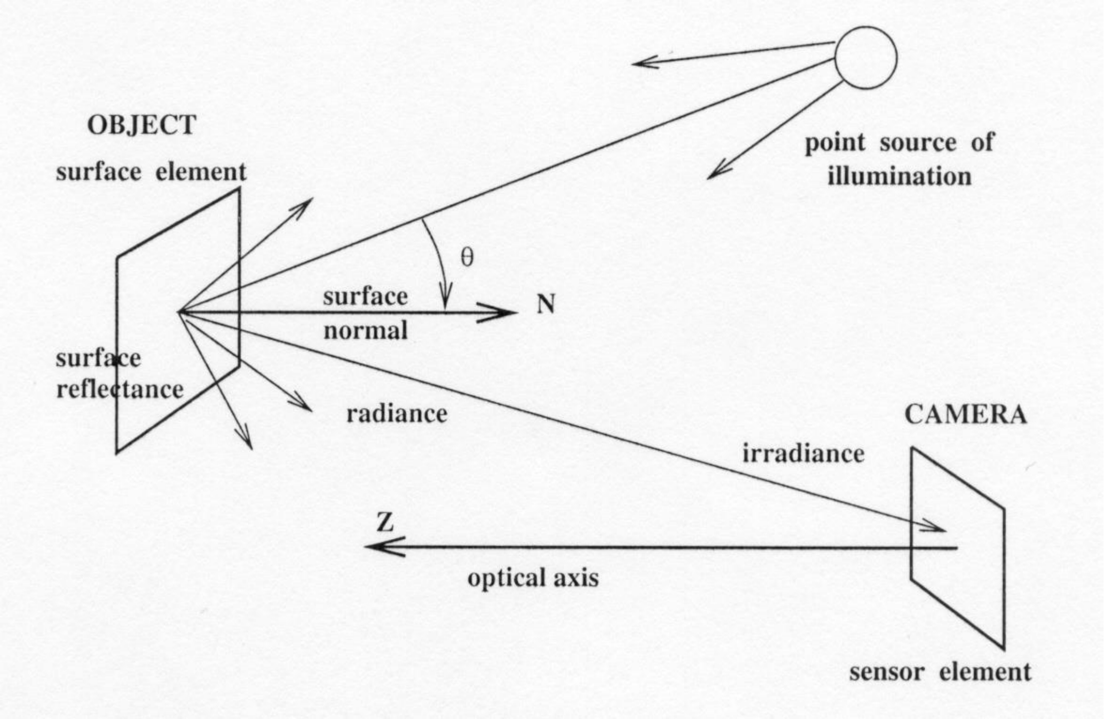
Camera optics

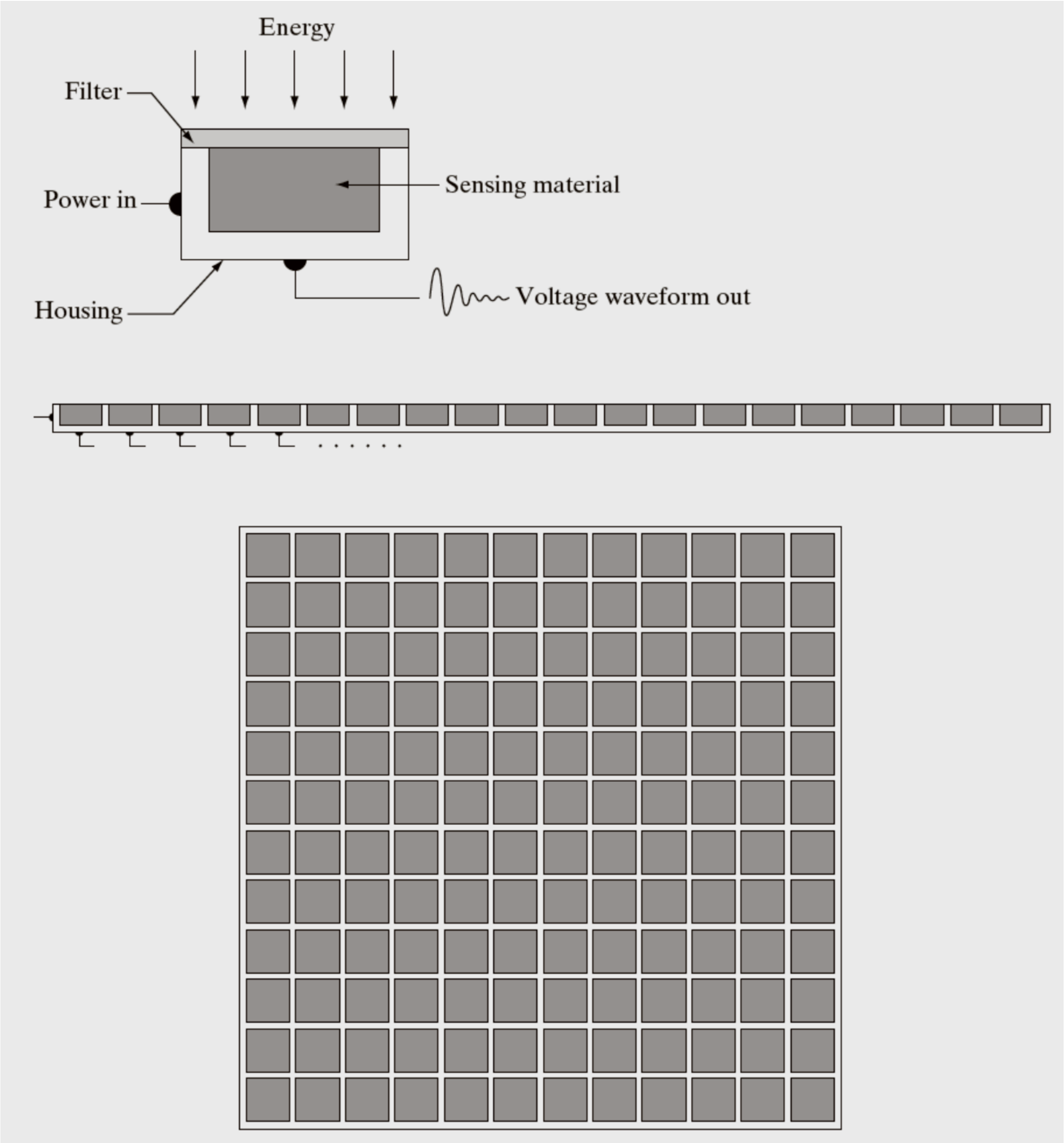
- In practice, the aperture must be larger to admit more light.
- Lenses are placed to in the aperture to <u>**focus**</u> the bundle of rays from each scene point onto the corresponding point in the image plane

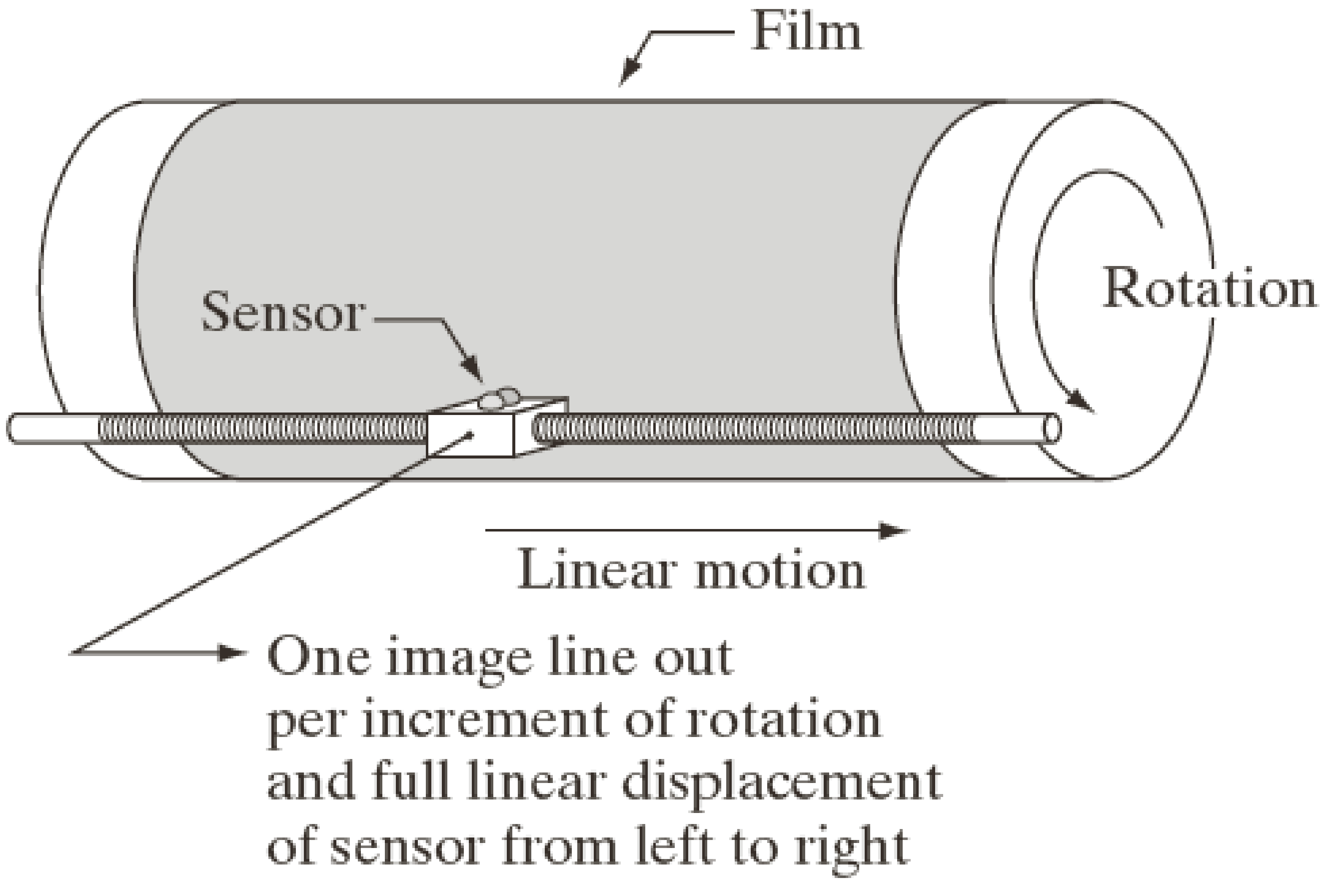


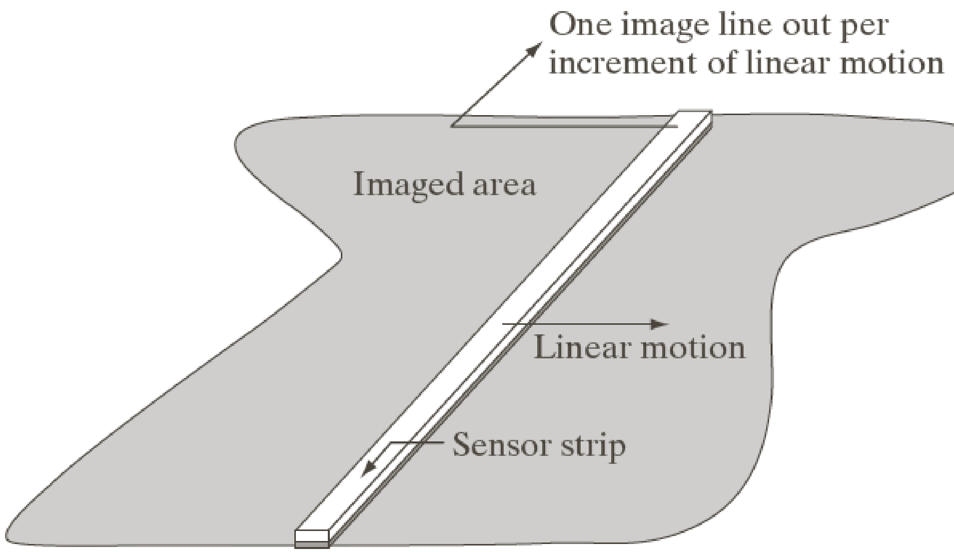
A Simple model of image formation

- The scene is illuminated by a single source.
- The scene reflects radiation towards the camera.
- The camera senses it via chemicals on film.




Energy
Filter
Power in
Sensing material
Housing
Voltage waveform out


Film
Sensor
Rotation
Linear motion
One image line out
per increment of rotation
and full linear displacement
of sensor from left to right


One image line out per
increment of linear motion
Imaged area
Linear motion
Sensor strip

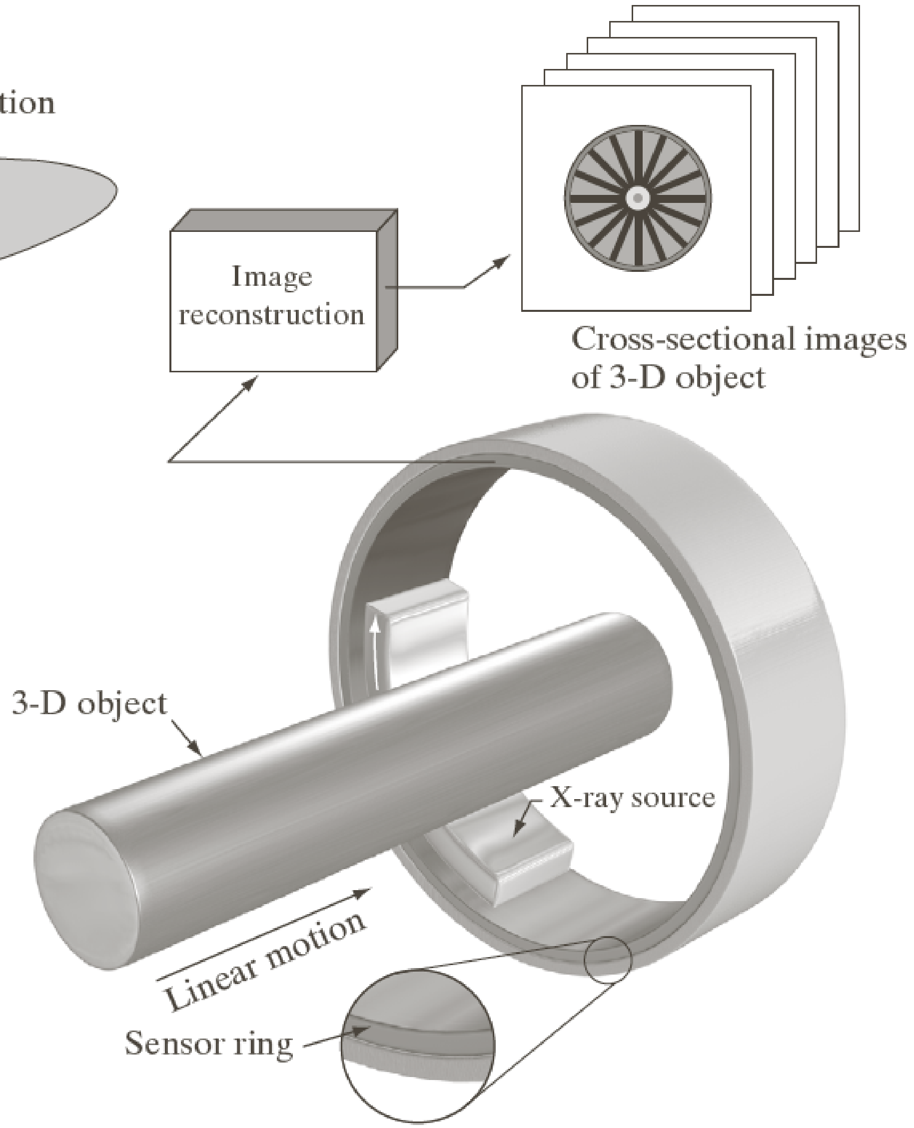
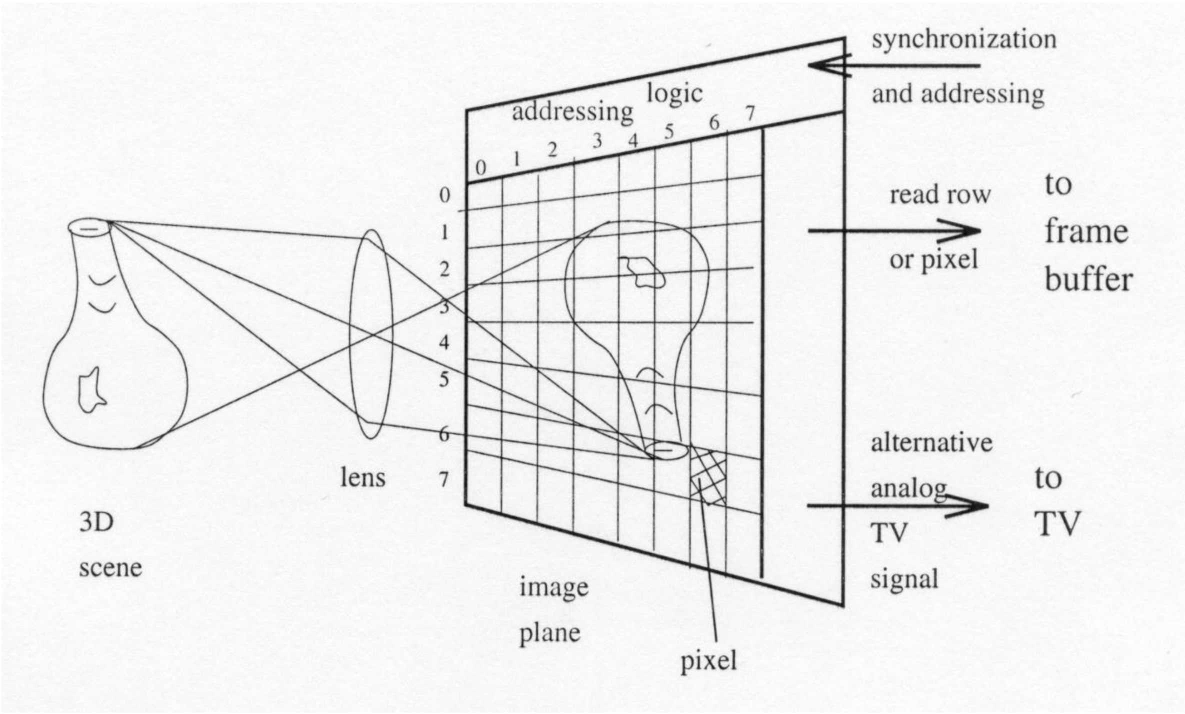


Image
reconstruction
Cross-sectional images
of 3-D object
3-D object
X-ray source
Linear motion
Sensor ring

CCD (Charged-Coupled Device) cameras

- Tiny <u>solid state cells</u> convert light energy into electrical charge.
- The image plane acts as a digital memory that can be read row by row by a computer.



Frame grabber

- Usually, a CCD camera plugs into a computer board (**frame grabber)**.
- The frame grabber digitizes the signal and stores it in its memory (**frame buffer**).

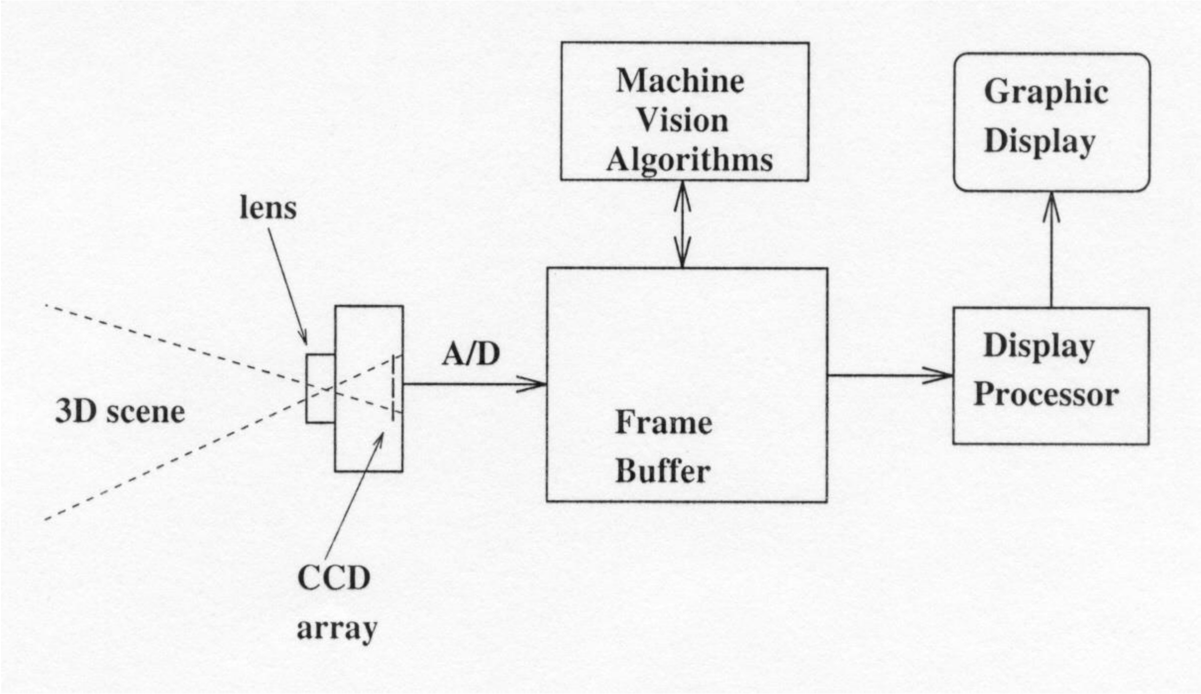
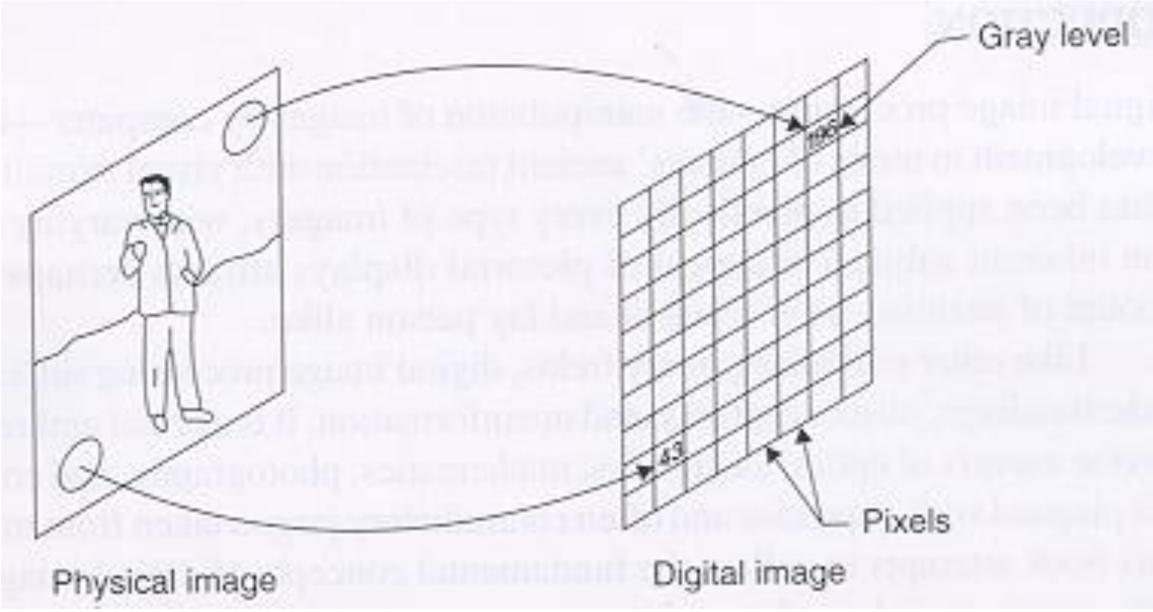
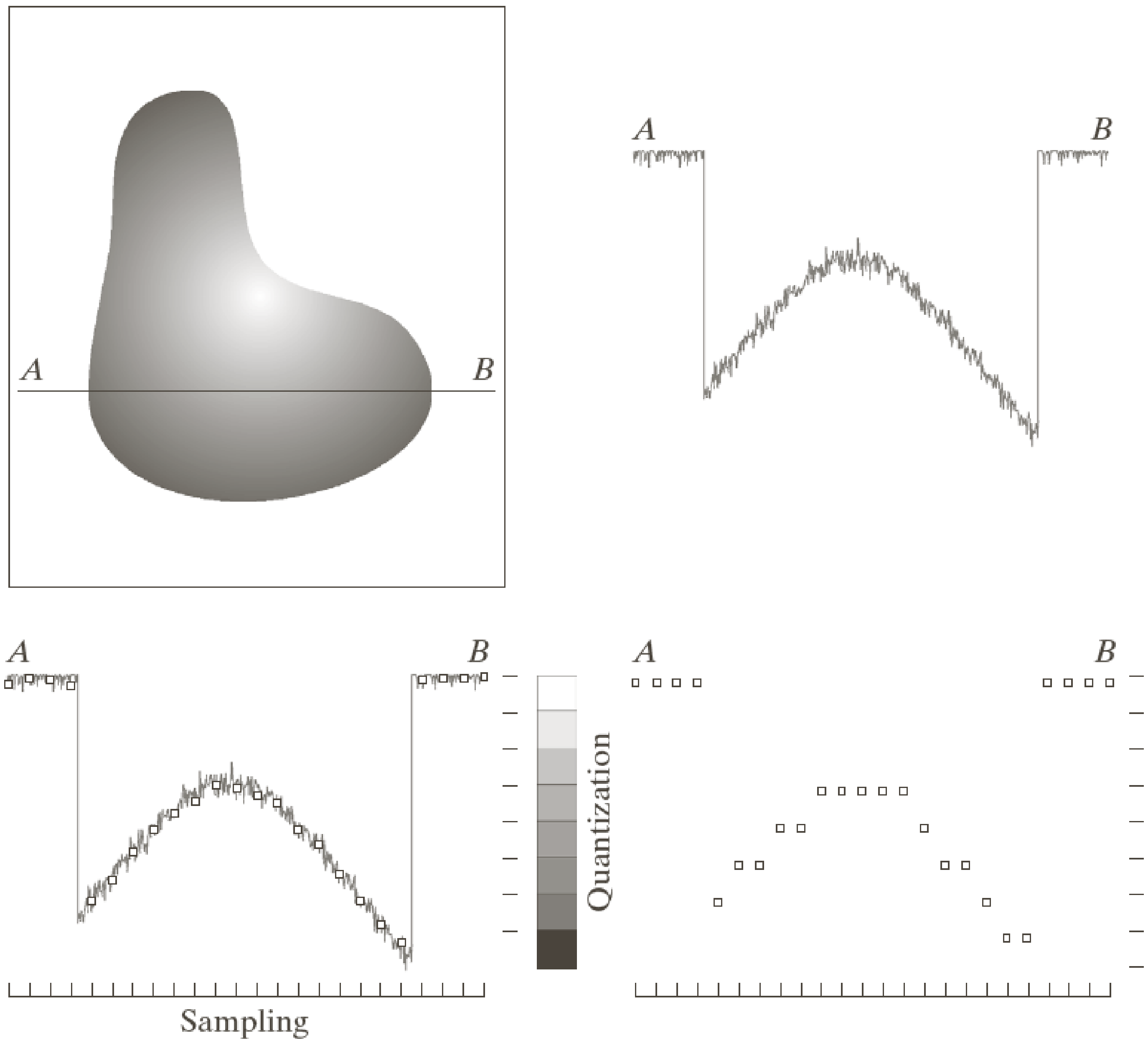
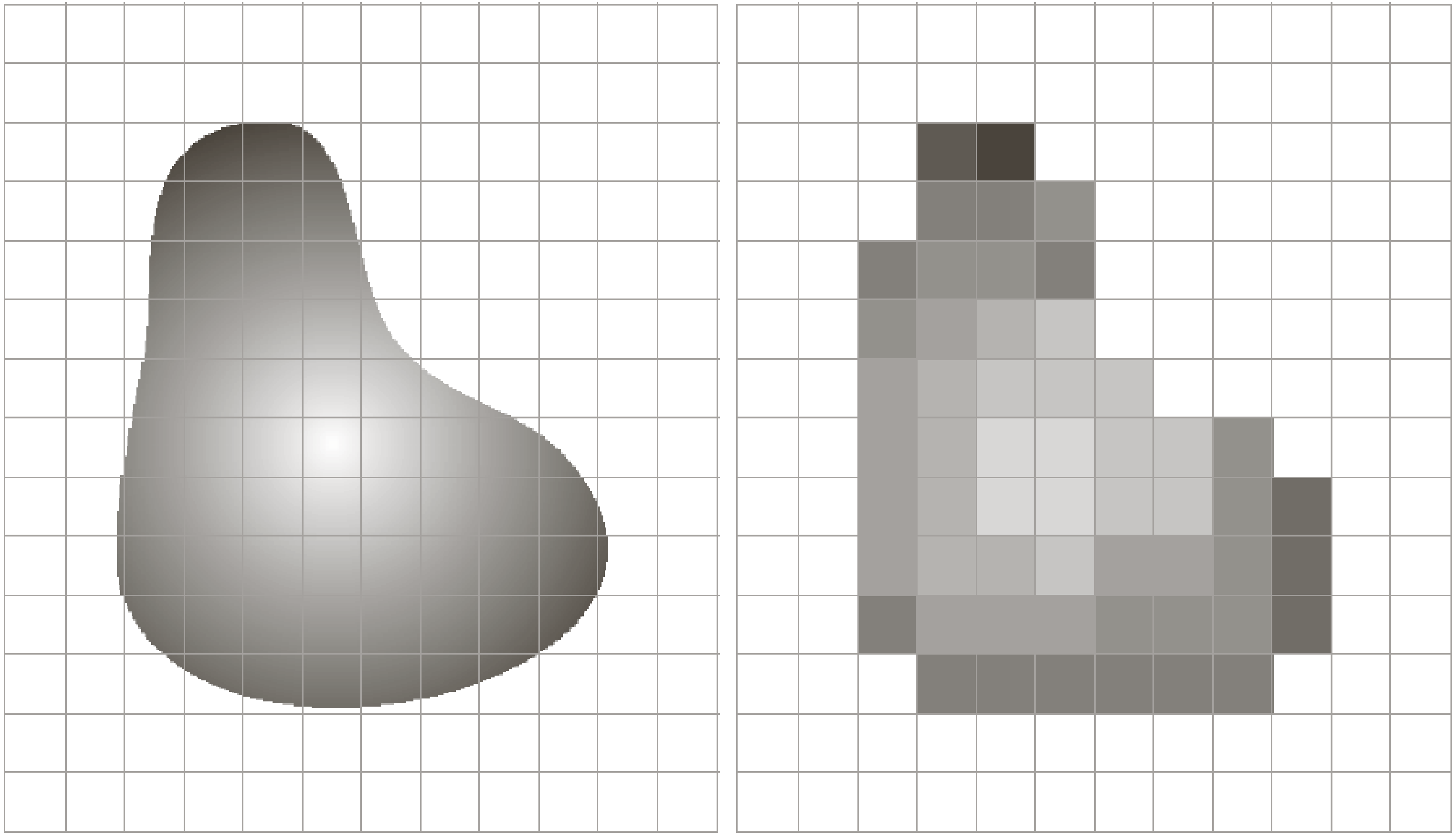


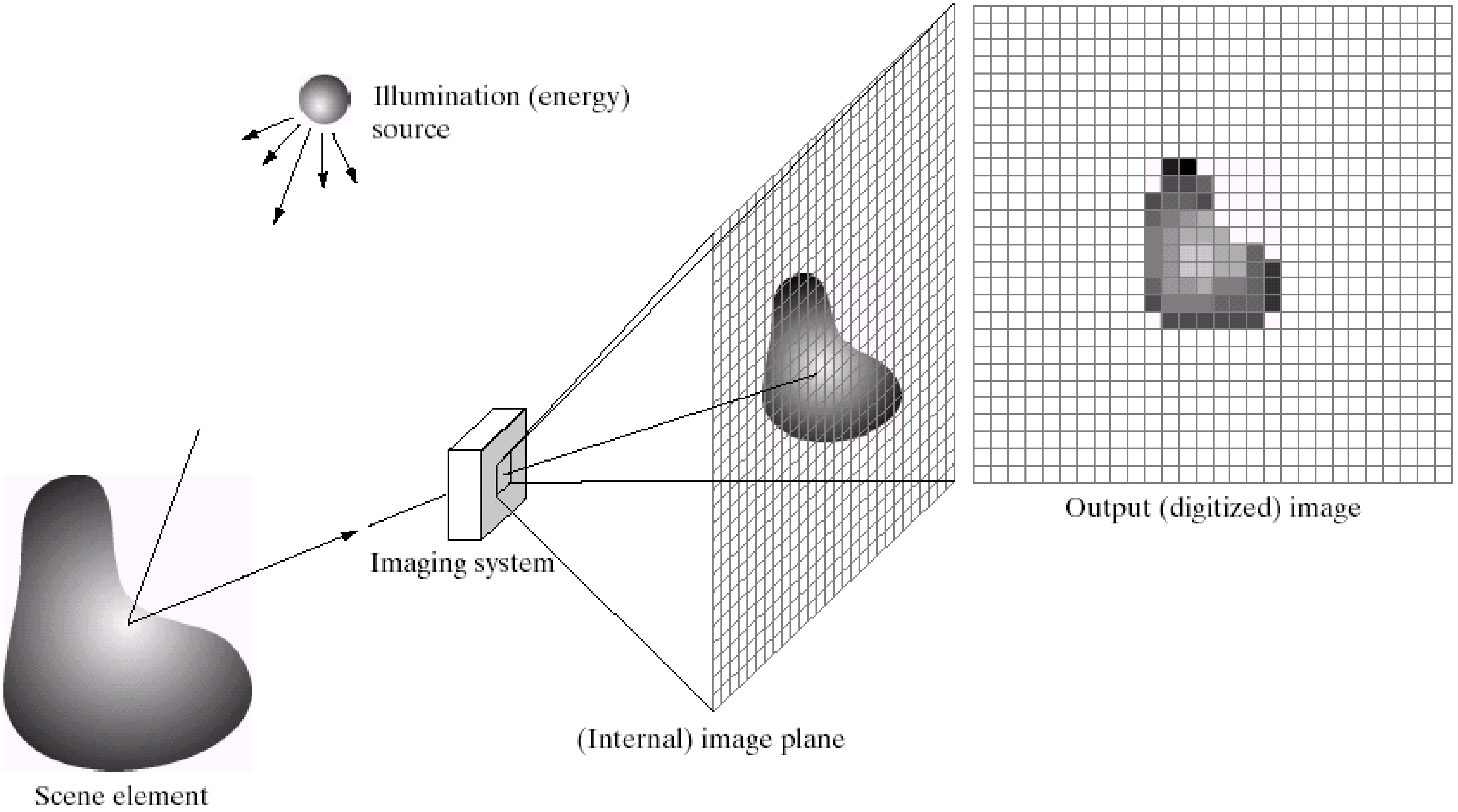
Image digitization



- **<u>Sampling</u>** means measuring the value of an image at a finite number of points.
- **<u>Quantization</u>** is the representation of the measured value at the sampled point by an integer.


A
B
A
B
A
B
Quantization
Sampling
A
B





Illumination (energy)
source
Imaging system
Output (digitized) image
(Internal) image plane
Scene element

Digital image

- An image is represented by a rectangular array of integers.
- An integer represents the brightness or darkness of the image at that point.
- N: # of rows, M: # of columns, Q: # of gray levels
  - N = 2^n , M = 2^m , Q = 2^q (q is the # of bits/pixel)
  - Storage requirements: NxMxq (e.g., N=M=1024, q=8, 1MB)

$$
\begin{matrix}
f(0,0) & f(0,1) & \dots & f(0,M-1) \\
f(1,0) & f(1,1) & \dots & f(1,M-1) \\
\dots & \dots & \dots & \dots \\
f(N-1,0) & f(N-1,1) & \dots & f(N-1,M-1)
\end{matrix}
$$

Image sampling (example)

Typical effects of reducing spatial resolution. Images shown at: (a) 1250 dpi, (b) 300 dpi, (c) 150 dpi, and (d) 72 dpi. The thin black borders were added for clarity. They are not part of the data.

Resolution: Image sampling (example)

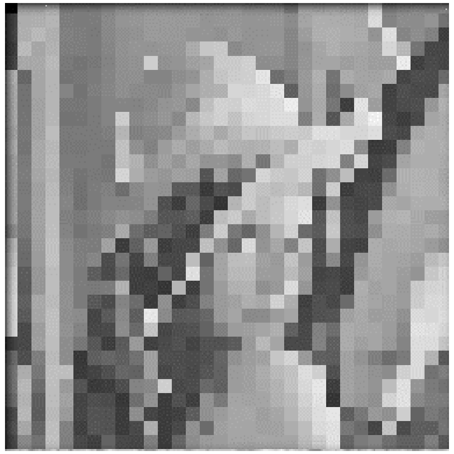
Typical effects of reducing spatial resolution. Images shown at: (a) 1250 dpi, (b) 300 dpi, (c) 150 dpi, and (d) 72 dpi. The thin black borders were added for clarity. They are not part of the data.

original image

sampled by a factor of 2

sampled by a factor of 4

sampled by a factor of 8



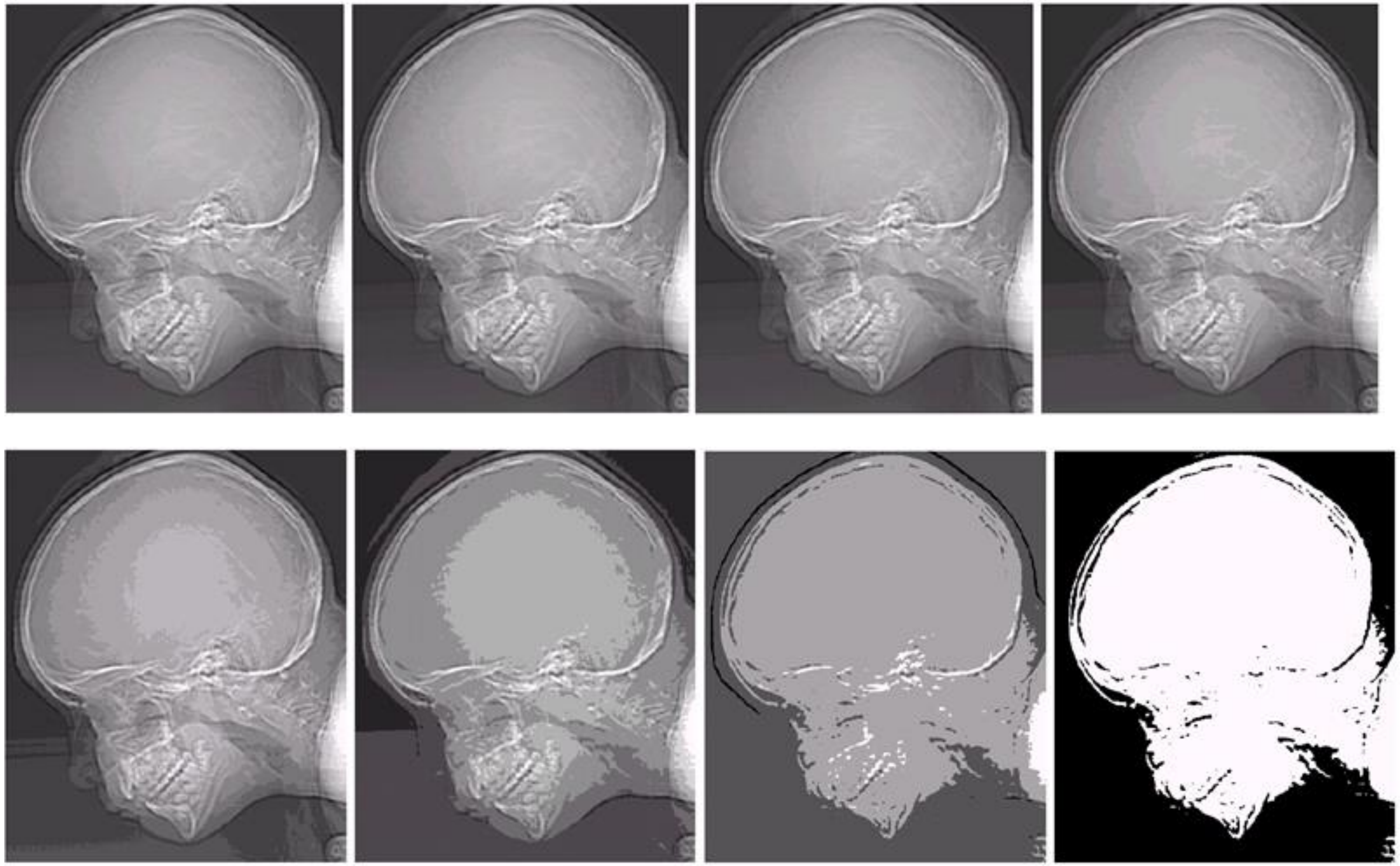
Gray-level resolution: Image quantization

- 256 gray levels (8bits/pixel) 32 gray levels (5 bits/pixel) 16 gray levels (4 bits/pixel)

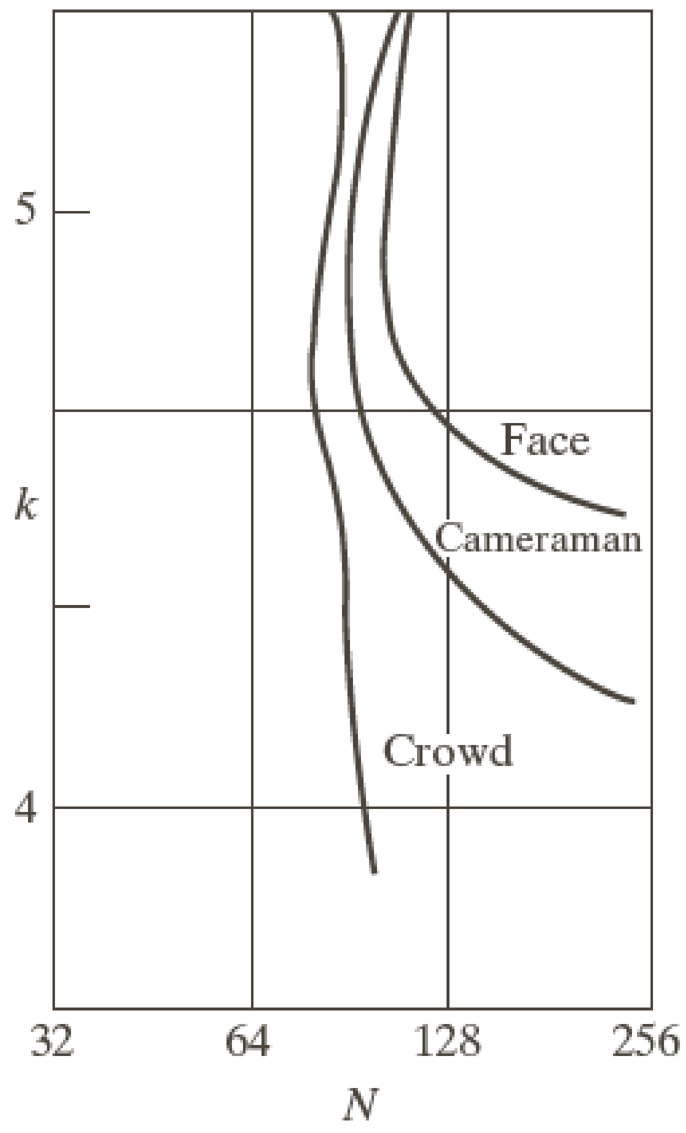


- 8 gray levels (3 bits/pixel) 4 gray levels (2 bits/pixel) 2 gray levels (1 bit/pixel)





(a) Image with a low level of detail. (b) Image with a medium level of detail. (c) Image with a relatively large amount of detail. (Image (b) courtesy of the Massachusetts Institute of Technology.)

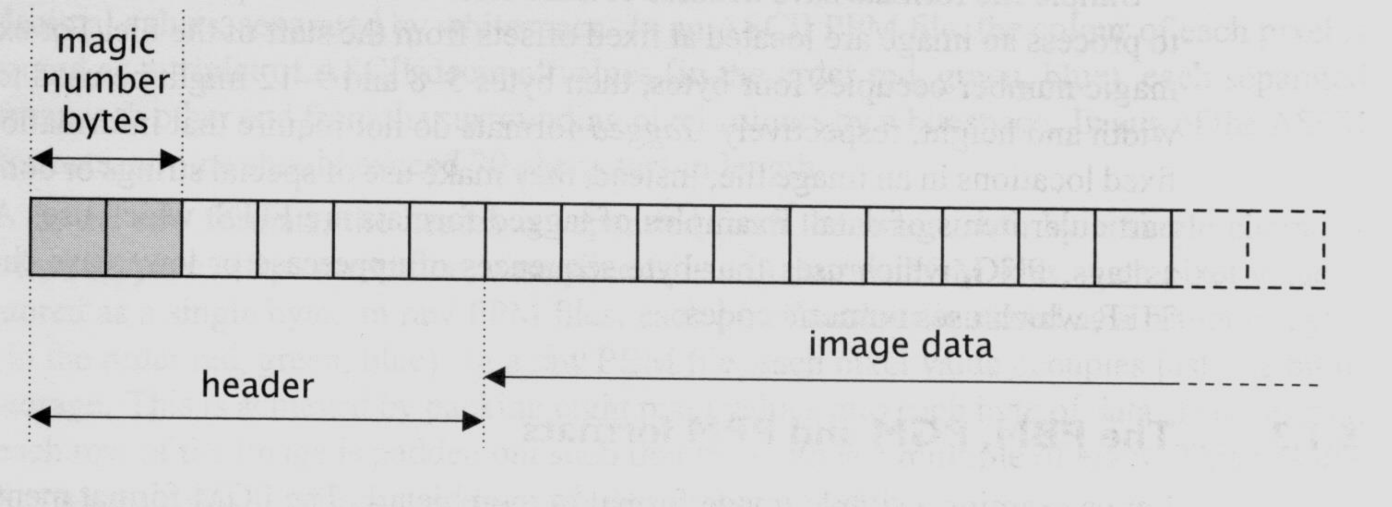


5
4
k
32
64
128
256
N
Face
Cameraman
Crowd

(a) Image reduced to 72 dpi and zoomed back to its original size (3692 × 2812 pixels) using nearest neighbor interpolation. This figure is the same as Fig. 2.20(d). (b) Image shrunk and zoomed using bilinear interpolation. (c) Same as (b) but using bicubic interpolation. (d)–(f) Same sequence, but shrinking down to 150 dpi instead of 72 dpi [Fig. 2.24(d) is the same as Fig. 2.20(c)]. Compare Figs. 2.24(e) and (f), especially the latter, with the original image in Fig. 2.20(a).

Image file formats

- Many image formats adhere to the simple model shown below (line by line, no breaks between lines).
- The header contains at least the width and height of the image.
- Most headers begin with a **<u>signature</u>** or “magic number” - a short sequence of bytes for identifying the file format.



Common image file formats

- GIF (Graphic Interchange Format) -
- PNG (Portable Network Graphics)
- JPEG (Joint Photographic Experts Group)
- TIFF (Tagged Image File Format)
- PGM (Portable Gray Map)
- FITS (Flexible Image Transport System)

PGM format

- A popular format for grayscale images (8 bits/pixel)
- Closely-related formats are:
  - PBM (Portable Bitmap), for binary images (1 bit/pixel)
  - PPM (Portable Pixelmap), for color images (24 bits/pixel)

```
P2
# a simple PGM image
7 7 255
  120 120 120 120 120 120 120
  120 120 120  33 120 120 120
  120 120 120  33 120 120 120
  120  33  33  33  33  33 120
  120 120 120  33 120 120 120
  120 120 120  33 120 120 120
  120 120 120 120 120 120 120
```

```
P5
# a simple PGM image
7 7 255
xxxxxxxxxx!xxxxxx!xxxx!!!!!xxxx!xxxxxx!xxxxxxxxxx
```

ASCII or binary (raw) storage

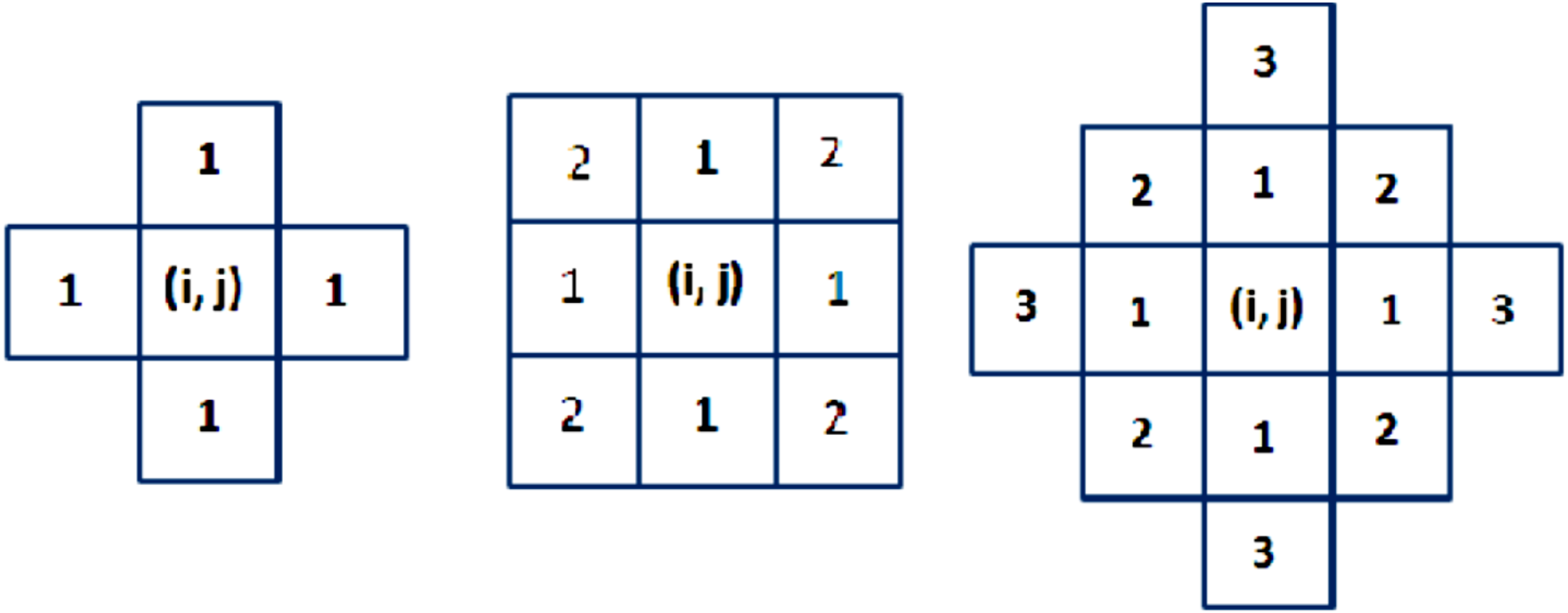
Signatures of the various PBM, PGM and PPM image formats.

Signature	Image type	Storage type
P1	binary	ASCII
P2	greyscale	ASCII
P3	RGB	ASCII
P4	binary	raw bytes
P5	greyscale	raw bytes
P6	RGB	raw bytes

ASCII vs Raw format

- ASCII format has the following advantages:
  - Pixel values can be examined or modified very easily using a standard text editor.
  - Files in raw format cannot be modified in this way since they contain many unprintable characters.
- Raw format has the following advantages:
  - It is much more compact compared to the ASCII format.
  - Pixel values are coded using only a single character !

Neighborhood



Distance Measures

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

(a) $D(p, q) \geq (D(p, q) = 0$ iff $p=q)$,
(b) $D(p, q) = D(q, p)$, *and*
(c) $D(p, z) \leq D(p, q) + D(q, z)$.

Euclidean Distance

- The *Euclidean distance* between *p* and *q* is defined as

$$D_e(p, q) = \left[(x - s)^2 + (y - t)^2\right]^{\frac{1}{2}}$$

		2		
	√2	1	√2	
2	1	0	1	2
	√2	1	√2	
		2		

City-block distance

- The *D_4 distance* (also called *city-block distance)* between *p* and *q* is defined as

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

$$\begin{array}{ccccc} & & 2 & & \\ & 2 & 1 & 2 & \\ 2 & 1 & 0 & 1 & 2 \\ & 2 & 1 & 2 & \\ & & 2 & & \end{array}$$

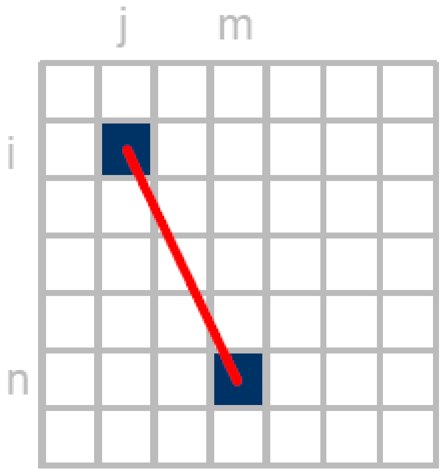
Chessboard distance

- The D_8 distance (also called *chessboard distance)* between *p* and *q* is defined as

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

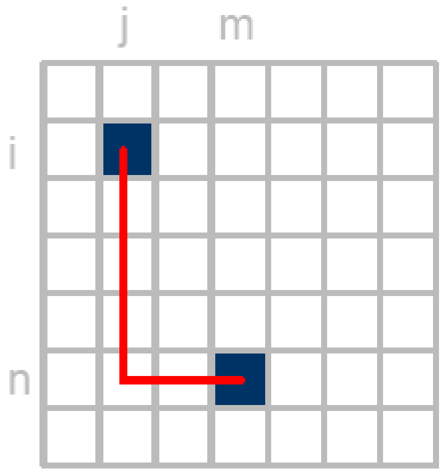
$$\begin{matrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{matrix}$$

Distance measures



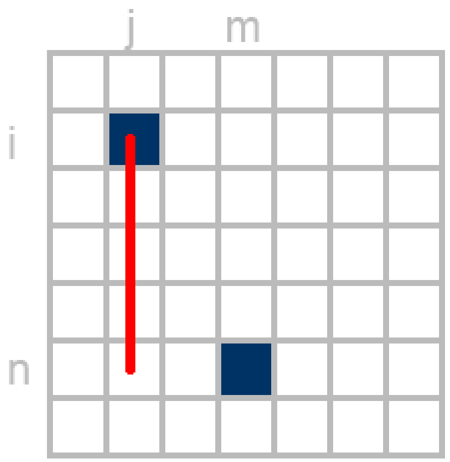
Euclidean Distance

$$= \sqrt{(i-n)^2 + (j-m)^2}$$



City Block Distance

$$= |i-n| + |j-m|$$



Chessboard Distance

$$= \max[\ |i-n|, |j-m|\ ]$$

Adjacency and Connectivity

- Two pixels are said to be connected if they are neighbors and satisfy some specific criteria of similarity (e.g., same gray value).
- Let V: a set of intensity values used to define adjacency and connectivity.
- In a binary image, $V = \{1\}$, if we are referring to adjacency of pixels with value 1.
- In a gray-scale image, the idea is the same, but V typically contains more elements, for example, $V = \{180, 181, 182, \ldots, 200\}$
- If the possible intensity values 0 – 255, V set can be any subset of these 256 values.

Types of Adjacency

1. **4-adjacency:** Two pixels p and q with values from V are 4-adjacent if q is in the set $N_4(p)$.
2. **8-adjacency:** Two pixels p and q with values from V are 8-adjacent if q is in the set $N_8(p)$.
3. **m-adjacency =(mixed)**

Types of Adjacency

- **m-adjacency:**

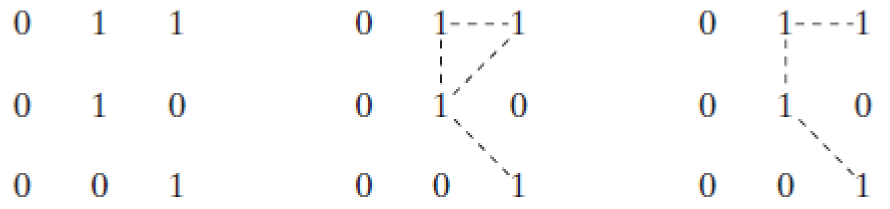
  Two pixels p and q with values from V are m-adjacent if :

  - q is in $N_4(p)$ or
  - q is in $N_D(p)$ and the set $N_4(p) \cap N_4(q)$ has no pixel whose values are from V (no intersection)

- **Important Note:** the type of adjacency used must be specified

Types of Adjacency

- Mixed adjacency is a modification of 8-adjacency. It is introduced to eliminate the ambiguities that often arise when 8-adjacency is used.
- For example:



a b c

FIGURE 2.26 (a) Arrangement of pixels; (b) pixels that are 8-adjacent (shown dashed) to the center pixel; (c) *m*-adjacency.

Types of Adjacency

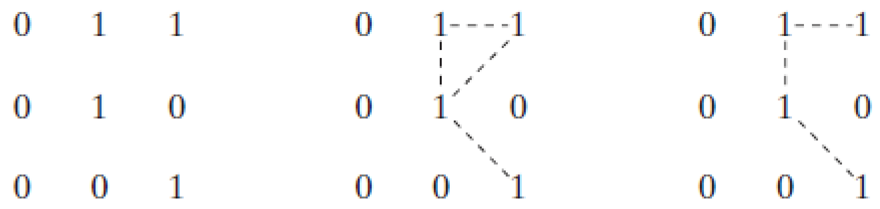
- In this example, we can note that to connect between two pixels (finding a path between two pixels):
  - In 8-adjacency way, you can find multiple paths between two pixels
  - While, in m-adjacency, you can find only one path between two pixels
- So, m-adjacency has eliminated the multiple path connection that has been generated by the 8-adjacency.
- Two subsets *S1* and *S2* are adjacent, if some pixel in *S1* is adjacent to some pixel in *S2*. Adjacent means, either 4-, 8- or m-adjacency.

A Digital Path

- A digital path (or curve) from pixel p with coordinate (x,y) to pixel q with coordinate (s,t) is a sequence of distinct pixels with coordinates (x_0,y_0), (x_1,y_1), …, (x_n, y_n) where $(x_0,y_0) = (x,y)$ and $(x_n, y_n) = (s,t)$ and pixels (x_i, y_i) and (x_{i-1}, y_{i-1}) are adjacent for $1 \le i \le n$
- n is the length of the path
- If $(x_0,y_0) = (x_n, y_n)$, the path is closed.
- We can specify 4-, 8- or m-paths depending on the type of adjacency specified.

A Digital Path

- Return to the previous example:



a b c

FIGURE 2.26 (a) Arrangement of pixels; (b) pixels that are 8-adjacent (shown dashed) to the center pixel; (c) *m*-adjacency.

In figure (b) the paths between the top right and bottom right pixels are 8-paths. And the path between the same 2 pixels in figure (c) is m-path

Connectivity

- Let *S* represent a subset of pixels in an image, two pixels *p* and *q* are said to be connected in *S* if there exists a path between them consisting entirely of pixels in *S*.
- For any pixel *p* in *S*, the set of pixels that are connected to it in *S* is called a *connected component* of *S*. If it only has one connected component, then set *S* is called a *connected set*.

Region and Boundary

- **Region**

  Let *R* be a subset of pixels in an image, we call *R* a region of the image if *R* is a connected set.

- **Boundary**

  The *boundary* (also called *border* or *contour*) of a region *R* is the set of pixels in the region that have one or more neighbors that are not in *R*.

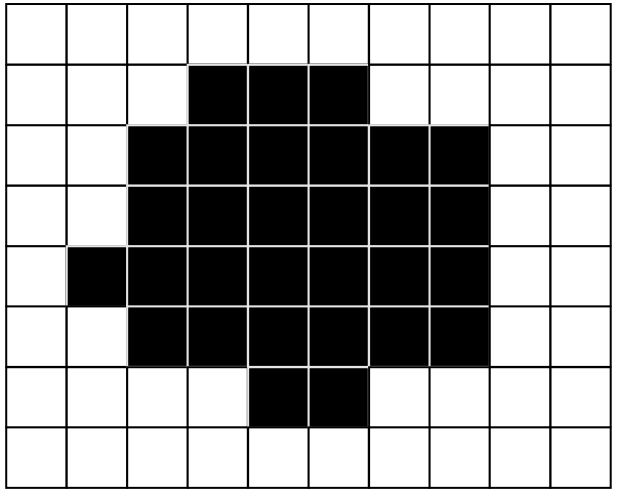
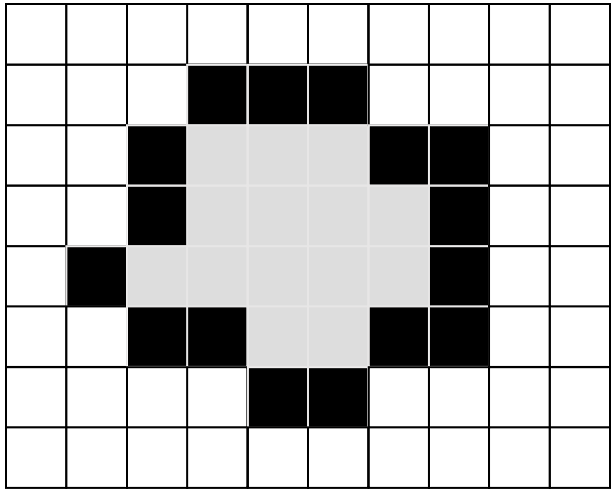
Region and Boundary

If *R* happens to be an entire image, then its boundary is defined as the set of pixels in the first and last rows and columns in the image.

This extra definition is required because an image has no neighbors beyond its borders

Regions

- R is a region if R is a connected set.
- The pixel in the boundary (contour) has at least one 4-adjacent neighbor whose value is 255.

	P_3	P_4
P_1	P_2	
P		

1. $P=P_2=P_4=1$ and $P_1=P_3=0$
2. $P=P_2=P_4=1$ and $P_1=P_3=1$
3. $P=P_2=P_4=P_1=1$ and $P_3=0$
4. $P=P_2=P_4=P_3=1$ and $P_1=0$