Computer Vision (CSE3010)

Dr. Susant Kumar Panigrahi
Assistant Professor
School of Electrical & Electronics Engineering
VIT BHOPAL

Module-1 Syllabus

Digital Image Formation And Low Level Processing:

- Overview and State-of-the-art, Fundamentals of Image Formation, Transformation: Orthogonal, Euclidean, Affine, Projective, Fourier Transform,
- Convolution and Filtering, Image Enhancement, Restoration, Histogram Processing.

Module-2 Syllabus

Depth Estimation And Multi-Camera Views:

Depth Estimation and Multi-Camera Views: Perspective, Binocular Stereopsis: Camera and Epipolar Geometry; Homography, Rectification, DLT, RANSAC, 3-D reconstruction framework; Auto-calibration. apparel.

Module-3 Syllabus

Feature Extraction And Image Segmentation:

- **Feature Extraction**: Edges - Canny, LOG, DOG; Line detectors (Hough Transform), Corners - Harris and Hessian Affine, Orientation Histogram, SIFT, SURF, HOG, GLOH, Scale-Space Analysis- Image Pyramids and Gaussian derivative filters, Gabor Filters and DWT.
- **Image Segmentation:** Region Growing, Edge Based approaches to segmentation, Graph-Cut, Mean-Shift, MRFs, Texture Segmentation; Object detection.

Module-4 Syllabus

Pattern Analysis And Motion Analysis:

- **Pattern Analysis**: Clustering: K-Means, K-Medoids, Mixture of Gaussians, Classification: Discriminant Function, Supervised, Un-supervised, Semi-supervised; Classifiers: Bayes, KNN, ANN models;
- **Dimensionality Reduction**: PCA, LDA, ICA; Non-parametric methods. Motion Analysis: Background Subtraction and Modelling, Optical Flow, KLT, Spatio-Temporal Analysis, Dynamic Stereo; Motion parameter estimation.

Module-5 Syllabus

Shape From X:

Light at Surfaces; Phong Model; Reflectance Map;

Albedo estimation; Photometric Stereo; Use of Surface Smoothness

Constraint; Shape from Texture, color, motion and edges.

Guest Lecture on Contemporary Topics

Text Books

1. Richard Szeliski, Computer Vision: Algorithms and Applications, Springer-Verlag London Limited 2011.
2. Computer Vision: A Modern Approach, D. A. Forsyth, J. Ponce, Pearson Education, 2003.

Reference Book(s):

1. R.C. Gonzalez and R.E. Woods, Digital Image Processing, Addison- Wesley, 1992.
2. Richard Hartley and Andrew Zisserman, Multiple View Geometry in Computer Vision, Second Edition, Cambridge University Press, March 2004.
3. K. Fukunaga; Introduction to Statistical Pattern Recognition, Second Edition, Academic Press, Morgan Kaufmann, 1990.

Required Tools/Software/IDLE:

1. Python/jupyter-notebook/google-colab
2. OpenCV
3. MATLAB

Indicative List of Experiments:

1. Implement image preprocessing and Edge
2. Implement camera calibration methods
3. Implement Projection
4. Determine depth map from Stereo pair
5. Construct 3D model from Stereo pair
6. Implement Segmentation methods
7. Construct 3D model from defocus image
8. Construct 3D model from Images
9. Implement optical flow method
10. Implement object detection and tracking from video
11. Face detection and Recognition
12. Object detection from dynamic Background for Surveillance
13. Content based video retrieval
14. Construct 3D model from single image

Computer Vision
Unit – 01
Low-Level Vision (Digital Image Formation And Low Level Processing)

VISION: HUMANS ARE VISUAL CREATURES

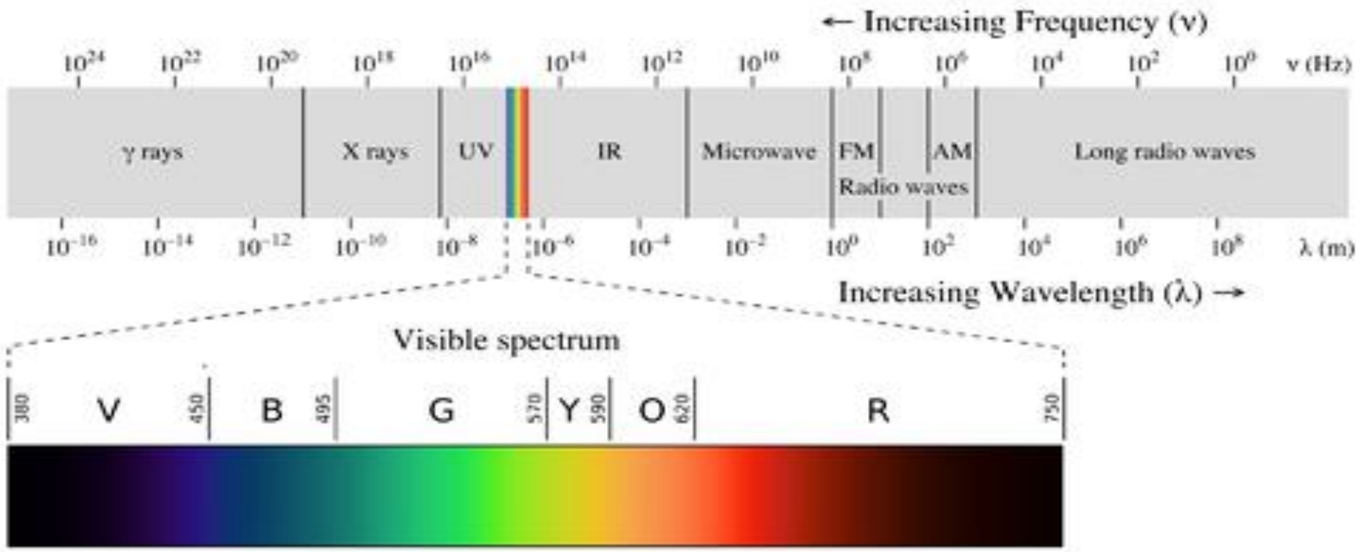
- *Half of the human brain is directly or indirectly devoted to processing visual information.*
  - The eye's retina, which contains 150 million light-sensitive rod and cone cells
- *The brain can identify images seen for as little as 13 milliseconds.*
  - Helps the brain as it decides where to focus the eyes
  - Deciding where to move the eyes can take 100 to 140 ms, so very high-speed understanding must occur before that.
- *At least 65 % of people are "visual learners"*
- *Humans have a remarkable ability to remember pictures.*
  - More than 2000 pictures with at least 90 % accuracy.
- *What our eyes see can influence what we hear, which is called the "McGurk Effect".*

Why we want Machines to Emulate Human Vision?

- ✓ To engage machines to perform our mundane works.
- ✓ Human visions are more qualitative rather quantitative.
- ✓ Human vision is limited.

Human Perception VS Machine Vision

Limited vs Entire EM spectrum

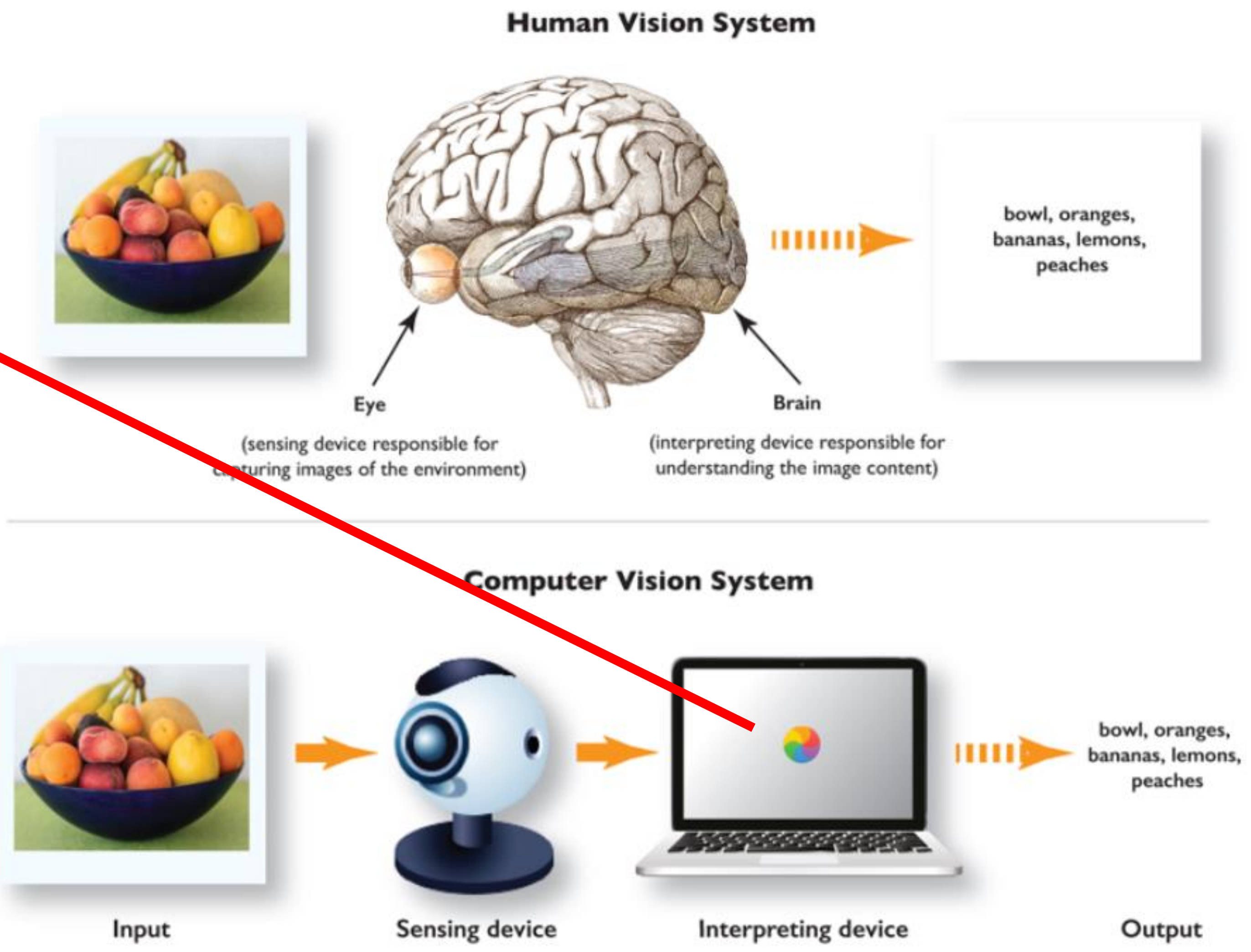


http://www.kollewin.com/blog/electromagnetic-spectrum/

HUMAN VS COMPUTER VISION

Interpreting Device or Vision Software: To come up with some symbolic description of the scene. [Bowl, Orange, Lemon, Banana and Peach].

More sophisticated algorithms may also illustrate "How fresh the fruits are what color/shape/size the bowl is"



Computer Vision*: where humans teach computers to see and interpret the world around them.

*a subfield of Deep Learning and Artificial Intelligence

But, what really is Computer Vision?

VISION is.....

... Automating human visual processes.

-- David Mahar

Computer vision that emulates human vision.

..... An information processing task.

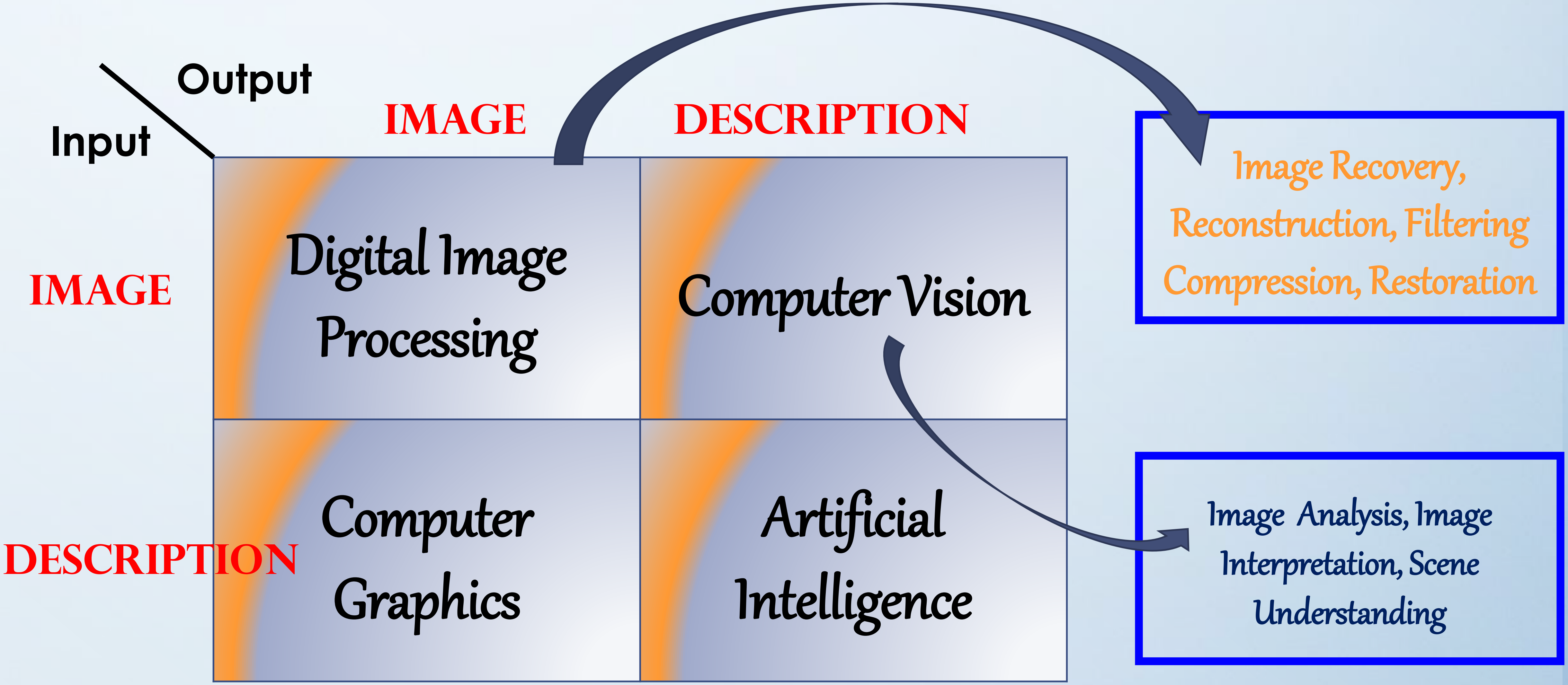
..... Inverting image formation.

--- Berthold Horn

..... Inverse of computer graphics.

...... REALY USEFUL.

Computer Vision & Other Related Fields…



Why Computer Vision Difficult?

- Inverse Problem

Perceives the "Story" behind the picture/video

Loss of dimensionality; while capturing image using sensing devices.

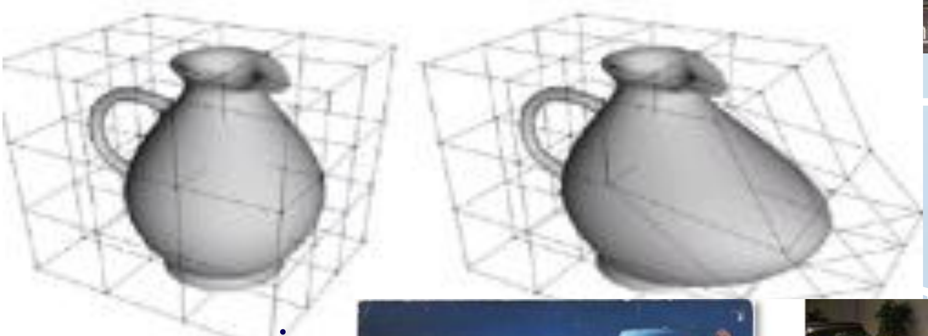
Seeks to recover some unknowns given insufficient information to fully specify the problem

Challenges:

1. **Viewpoint Variation:** Input image may align in different directions that leads the computer vision system to predict inaccurate results.

2. **Scale Variation:** Images captured closer to camera looks bigger and vice versa. Variation in size or scale affects the decision taking capabilities of the system.

3. **Deformation:** System may learn from perfect image and depicts a particular perception about the shape, size and other features. In real-world shape may change that leads to inaccuracy when shape of the object is deformed.



4. **Inter-class variation:** Objects of same class may come in different shape, size, colour and texture; but the algorithm need to identify them as single class.

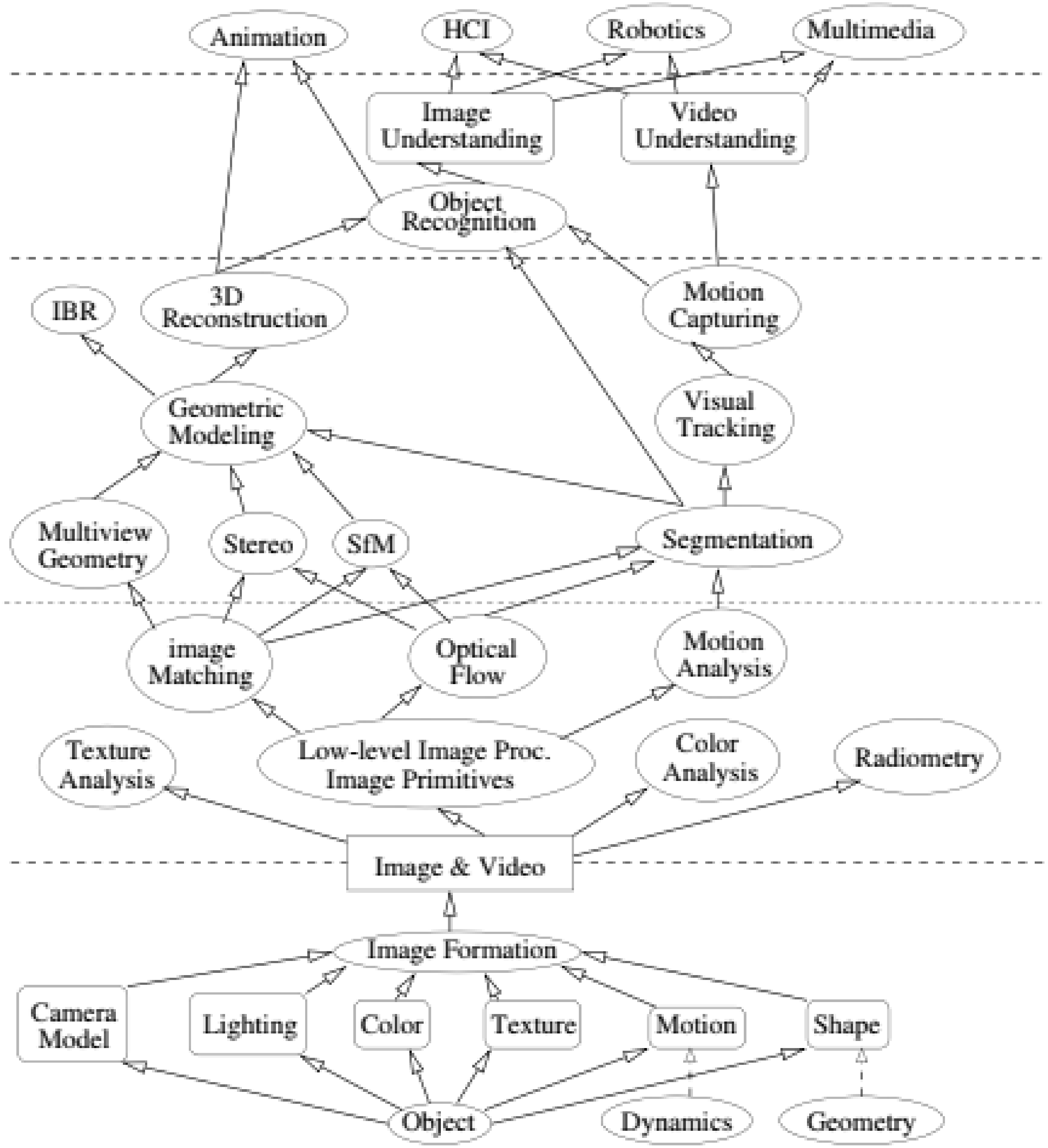
5. **Scale Variation:** Incomplete information due to occlusion results in inaccurate interpretation.

6. **Illumination Variation:** Same image captured under different illumination levels.

Computer Vision System: How Does CV Works?

Computer Vision System:

1) Image Formation: Cameras to obtain visual data,
2) Low-level image processing:
3) Low level vision
4) Middle-level Vision
5) High-level vision: Image Understanding
6) Decision making



Vision deals with Images...!

A picture worth more than thousand words.

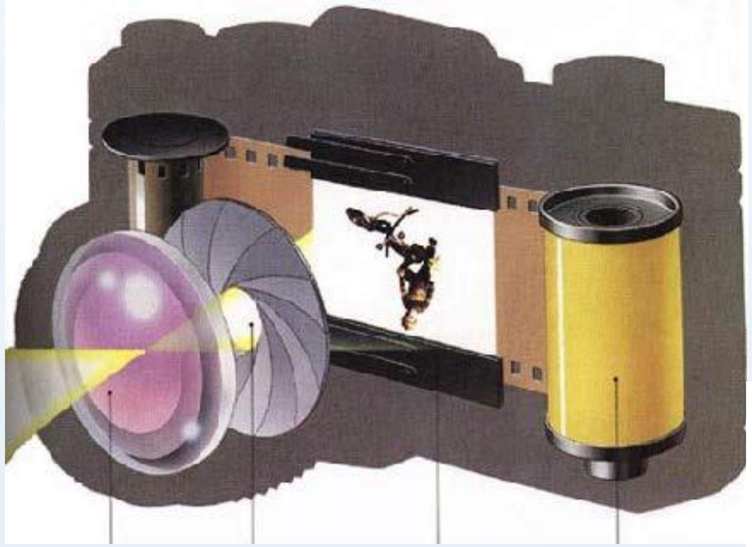
0	18	238	255	248	134	21	76	248	255	255	162	0	0	0	0	0	0	0	0	0	0	0	0
0	87	255	255	255	255	255	255	254	253	255	87	0	0	0	0	0	0	0	0	0	0	0	0
0	75	255	255	249	253	255	254	253	255	255	132	50	41	0	0	0	0	79	143	27	0	0	0
0	39	255	255	255	245	200	186	187	192	222	255	255	255	51	0	43	202	255	255	75	0	0	0
0	67	255	255	239	162	154	174	179	172	155	197	227	236	164	166	255	255	255	255	39	0	0	0
0	146	255	255	173	159	206	196	194	200	184	155	241	253	255	255	240	217	250	255	44	0	0	0
54	255	255	242	167	193	192	193	192	205	172	150	247	255	248	186	163	166	163	255	178	0	0	0
167	255	250	255	189	157	196	207	199	167	141	186	219	248	185	156	195	204	171	165	255	207	92	33
193	255	251	255	245	179	153	138	144	190	241	255	227	175	169	202	193	192	193	168	239	255	255	174
154	255	247	207	214	233	198	190	235	255	255	255	219	160	178	197	193	197	177	168	252	255	236	30
54	255	225	207	245	255	255	254	210	181	177	231	220	146	190	198	196	199	145	203	255	255	73	0
0	66	215	255	252	255	245	165	154	187	179	156	230	154	125	192	184	151	192	253	255	235	12	0
0	0	82	255	255	253	179	169	205	192	205	156	197	253	174	159	163	220	254	253	255	255	75	0
0	0	39	255	255	224	161	198	192	192	204	154	192	255	255	214	197	255	251	249	254	255	193	0
0	0	104	255	255	206	159	203	192	197	190	152	235	255	240	195	241	255	251	255	255	251	118	0
0	74	255	255	255	227	142	189	202	186	145	206	255	255	206	222	255	251	255	190	36	0	0	0
95	255	255	249	251	254	200	153	167	171	215	254	252	255	183	234	255	255	170	0	0	0	0	0
103	255	252	253	255	255	255	233	221	245	255	250	250	255	193	164	254	150	0	0	0	0	0	0
0	206	255	255	172	175	255	255	255	253	249	249	253	255	255	62	0	0	0	0	0	0	0	0
0	26	144	70	0	0	106	255	255	250	250	255	255	240	172	18	0	0	0	0	0	0	0	0
0	0	0	0	0	0	8	230	255	250	255	239	77	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	198	255	255	242	37	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	67	231	204	36	0	0	0	0	0	0	0	0	0	0	0	0	0

Scope and Challenges in CV

- ✓ Vision is a hard problem.
- ✓ Vision is multi-disciplinary.
- ✓ Considerable progress has been made.
- ✓ Many successful real-world applications.
- ✓ Computer vision as part of AI

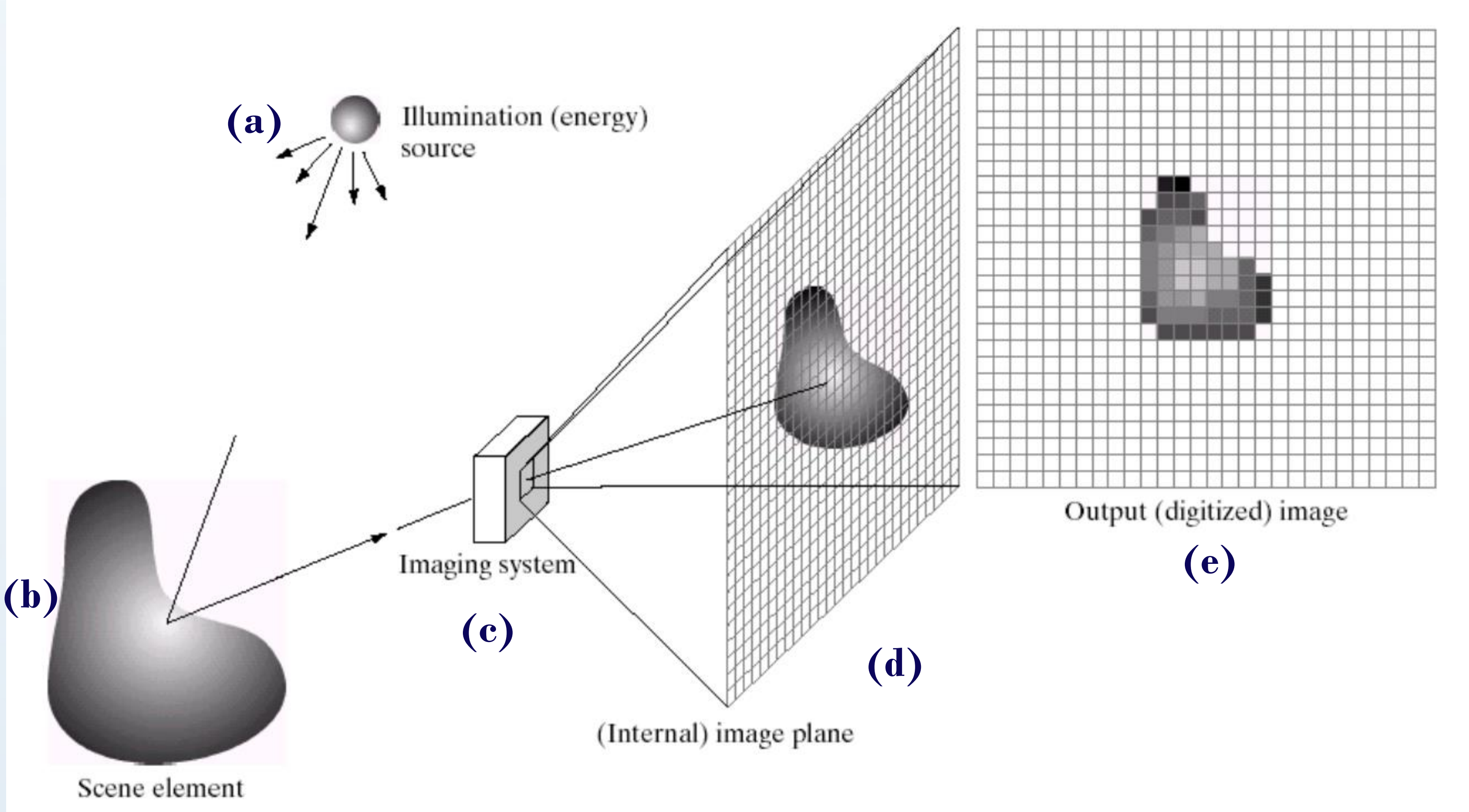
IMAGE ACQUISITION: IMAGING SYSTEM

Camera + Scanner → Digital Camera: Get images into computer



https://cse.sc.edu/~tongy/csce763/csce763.html

IMAGE ACQUISITION: IMAGING SYSTEM



An example of digital image acquisition process. (a) Energy (illumination) Source. (b) An element of a scene. (c) Imaging System. (d) Projection of the scene on to the imaging plane. (e) Digitized Image

Formulation:

$$f(x, y) = i(x, y) \times r(x, y)$$

Where:

$i(x, y)$ = Illumination (Energy) source.

$$0 < i(x, y) < \infty$$

$r(x, y)$ = Reflectance component of the scene.

$$0 \leq r(x, y) \leq 1$$

DIGITAL IMAGE

- Reflectance of some scenes
  - 0.01 for black velvet
  - 0.65 for stainless steel
  - 0.80 for flat-white wall paint
  - 0.90 for silver-plated metal
  - 0.93 for snow

Formulation:

$$f(x, y) = i(x, y) \times r(x, y)$$

Where:

$i(x, y)$ = Illumination (Energy) source.

$$0 < i(x, y) < \infty$$

$r(x, y)$ = Reflectance component of the scene.

$$0 \leq r(x, y) \leq 1$$

DIGITAL IMAGE

- Image: Two dimensional function $f(x, y)$

  Where x *and* y spatial(plane) coordinates

- Intensity or gray level: The amplitude of f at any pair of coordinates (x, y)

- Image as function

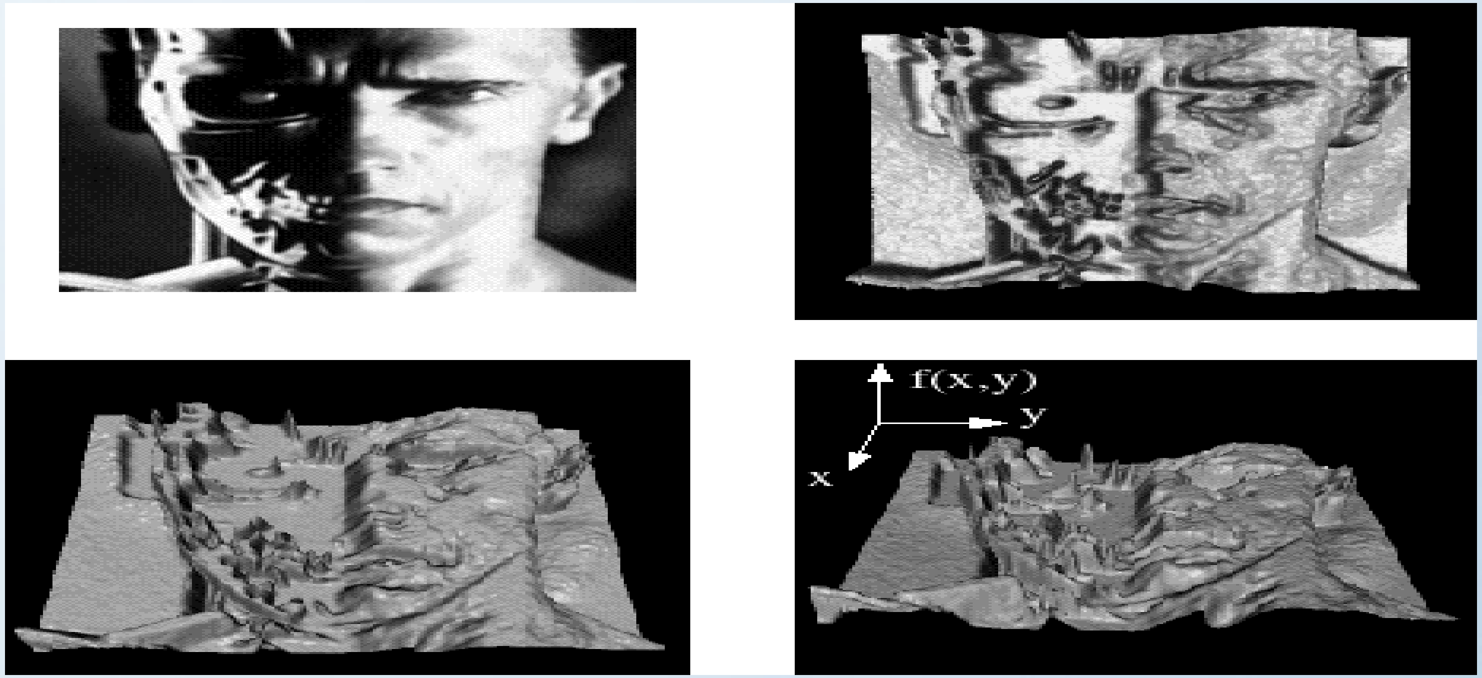
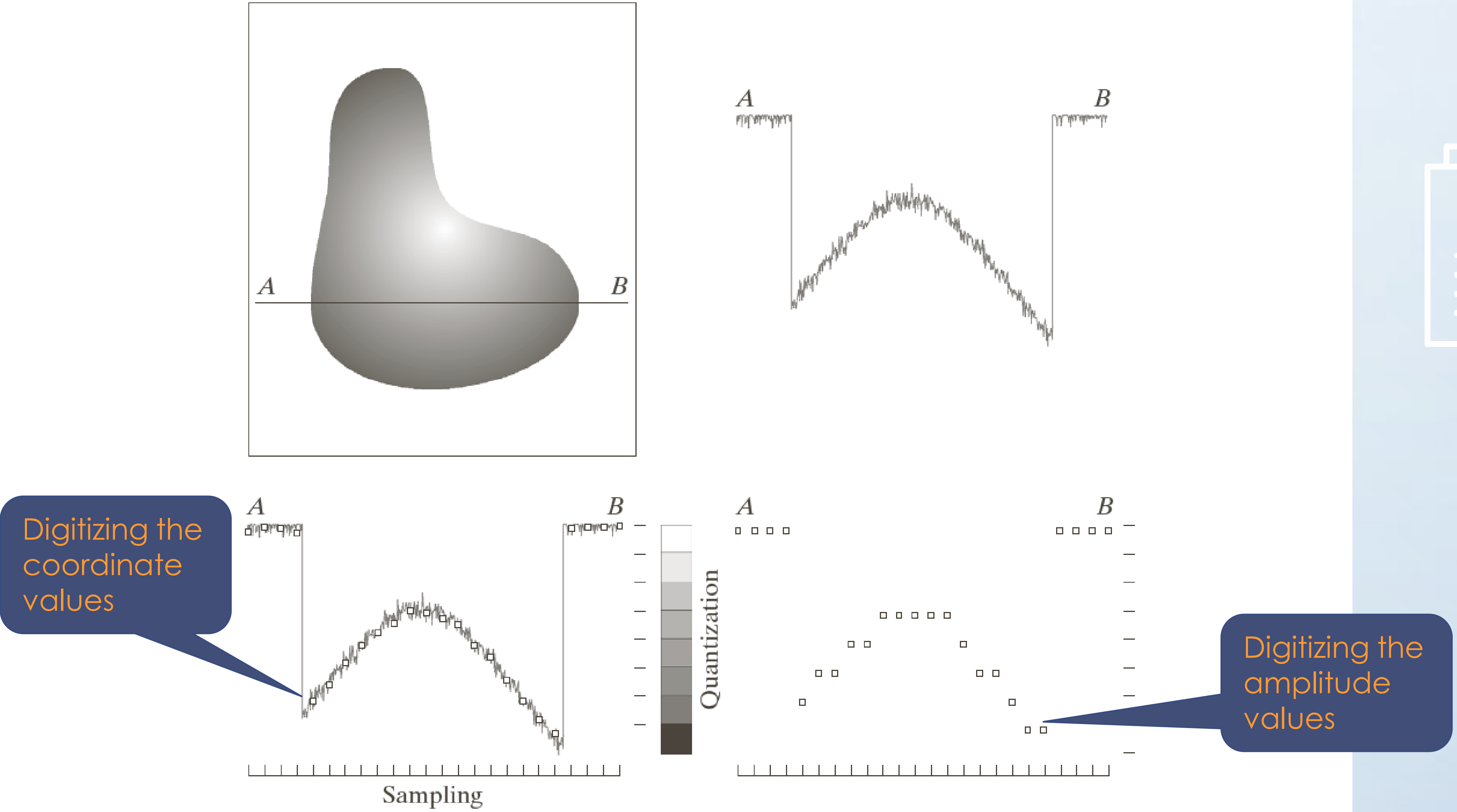
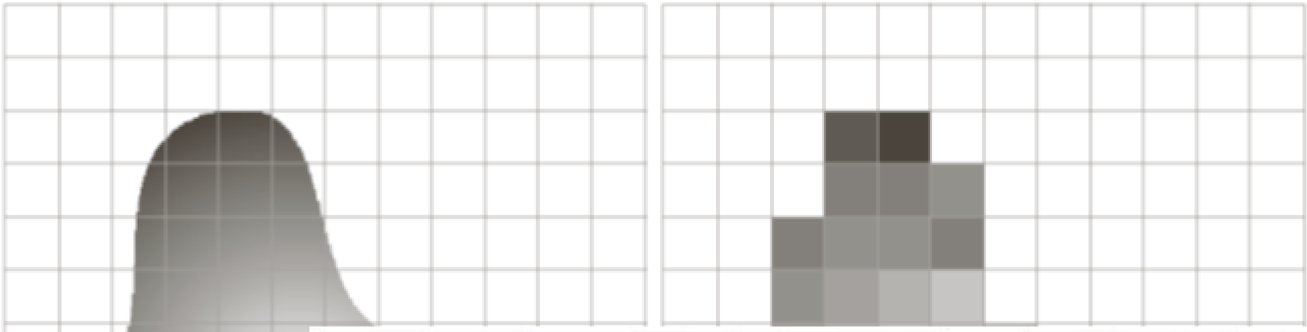


IMAGE SAMPLING & QUANTIZATION



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

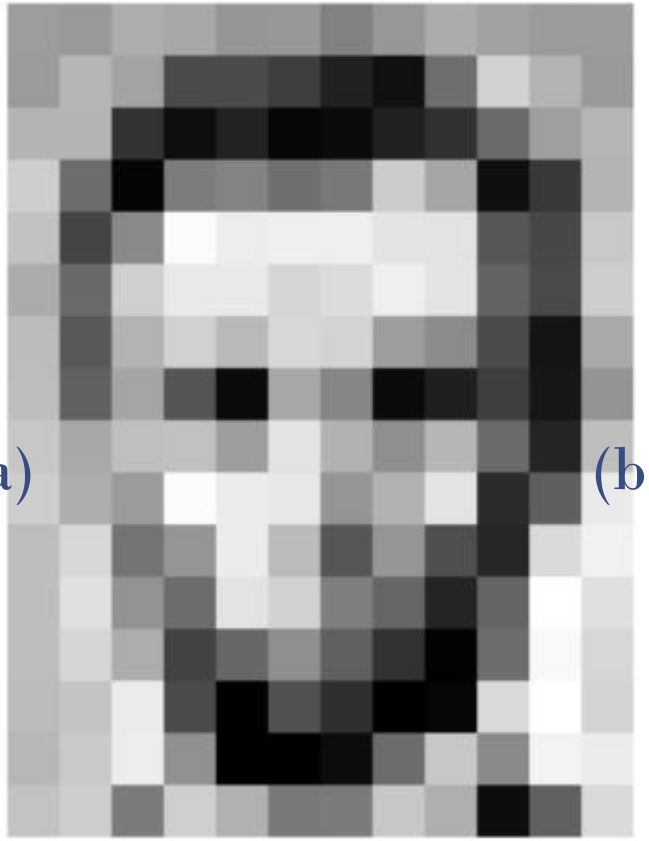
IMAGE SAMPLING & QUANTIZATION

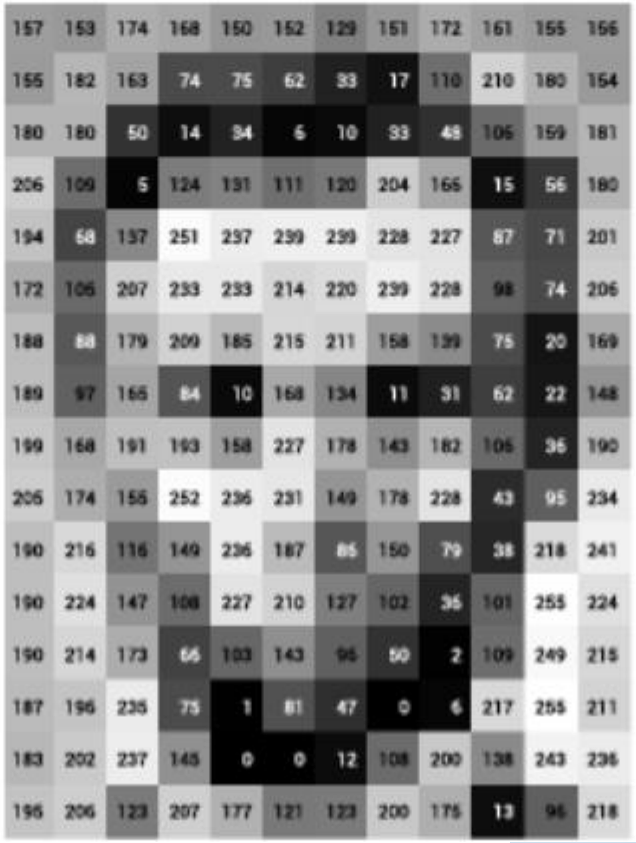


(a)

(b)

(a) Continuous image





157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

DIGITAL IMAGE REPRESENTATION

A monochrome image and the convention used to represent rows (x) and columns (y).

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

2*D* Sampling:

- Represented in a matrix with M rows and N columns.
- Equally spaced samples in two dimension.

[Digitizing the coordinate values are called ***sampling***]

Quantization:

- Digitizing the amplitude values are called Quantization.
- The amplitude values depend on the “data type” 0r “class” by which it is represented.
- The following are few popular representations:
  - ✓ *uint8*
  - ✓ *uint16*
  - ✓ *unit32*
  - ✓ *double*
  - ✓ *single*

DIGITAL IMAGE REPRESENTATION

- Discrete intensity interval $[0, L-1], L = 2^k$; L = Dynamic Range
- The number b of bits required to store a $M \times N$ digitized image

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

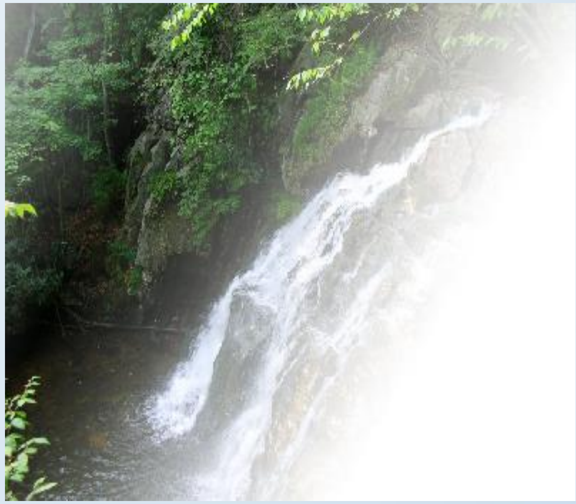
Number of storage bits for various values of N and k.

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

TYPES DIGITAL IMAGE

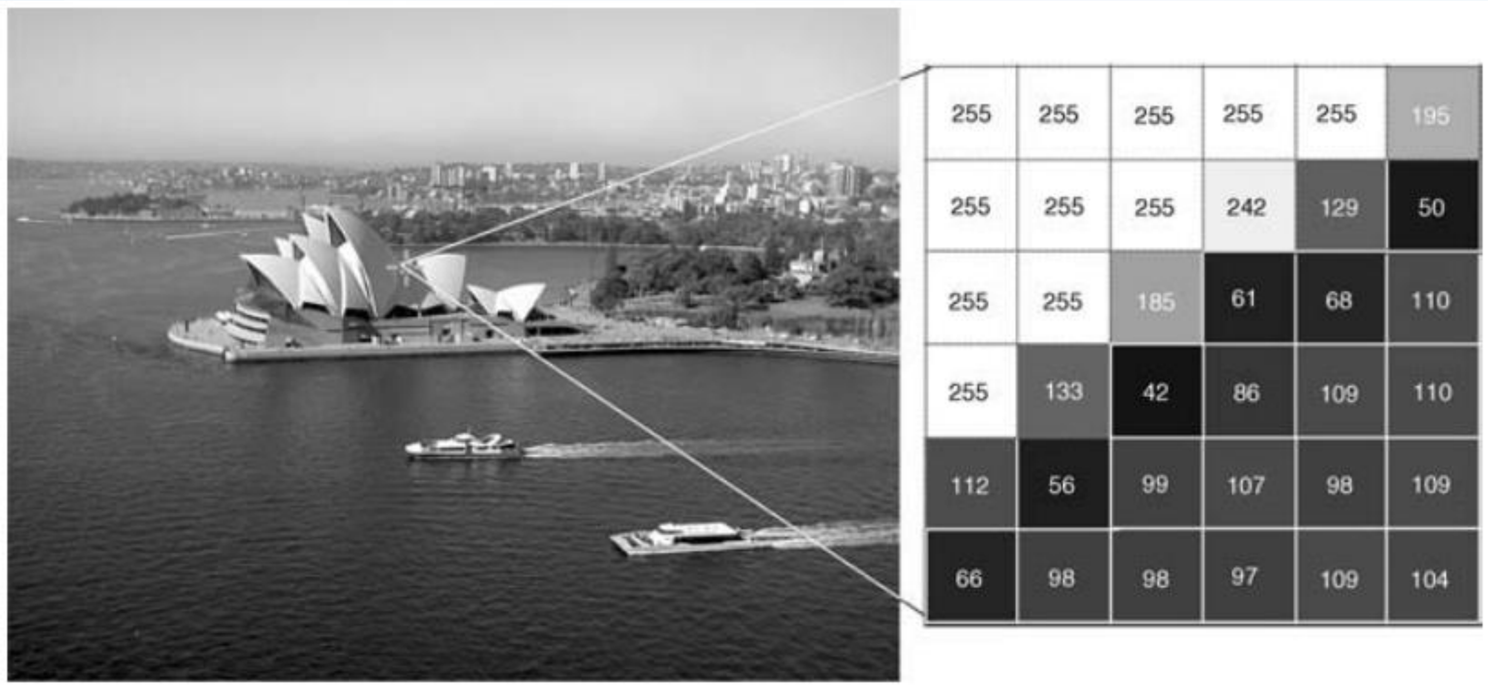
- Common image formats include:
  - 1 sample per point (B&W or Grayscale)
  - 3 samples per point (Red, Green, and Blue)
  - 4 samples per point (Red, Green, Blue, and “Alpha”, a.k.a. Opacity)





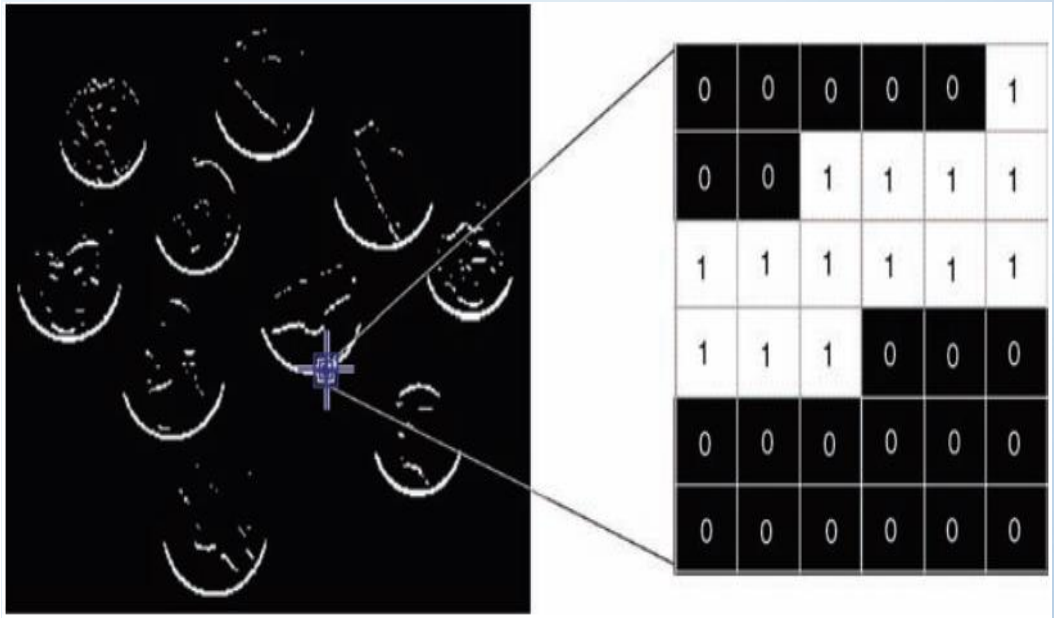
- For most of this course we will focus on grey-scale and colour images

TYPES OF DIGITAL IMAGES



255	255	255	255	255	195
255	255	255	242	129	50
255	255	185	61	68	110
255	133	42	86	109	110
112	56	99	107	98	109
66	98	98	97	109	104

A grayscale (monochrome) image and the pixel values in a 6 × 6 neighborhood.



0	0	0	0	0	1
0	0	1	1	1	1
1	1	1	1	1	1
1	1	1	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

A binary image and the pixel values in a 6 × 6 neighborhood.

MAX-MIN Pixel Intensity:

Gray-scale Image:

$$min = 0$$
$$max = 2^b - 1$$

b = Number of bits required to represent the pixel intensity values.

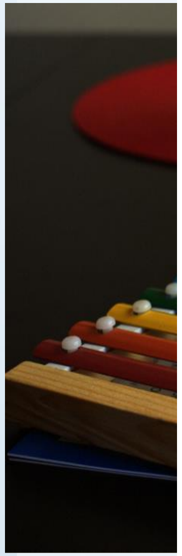
MAX-MIN Pixel Intensity:

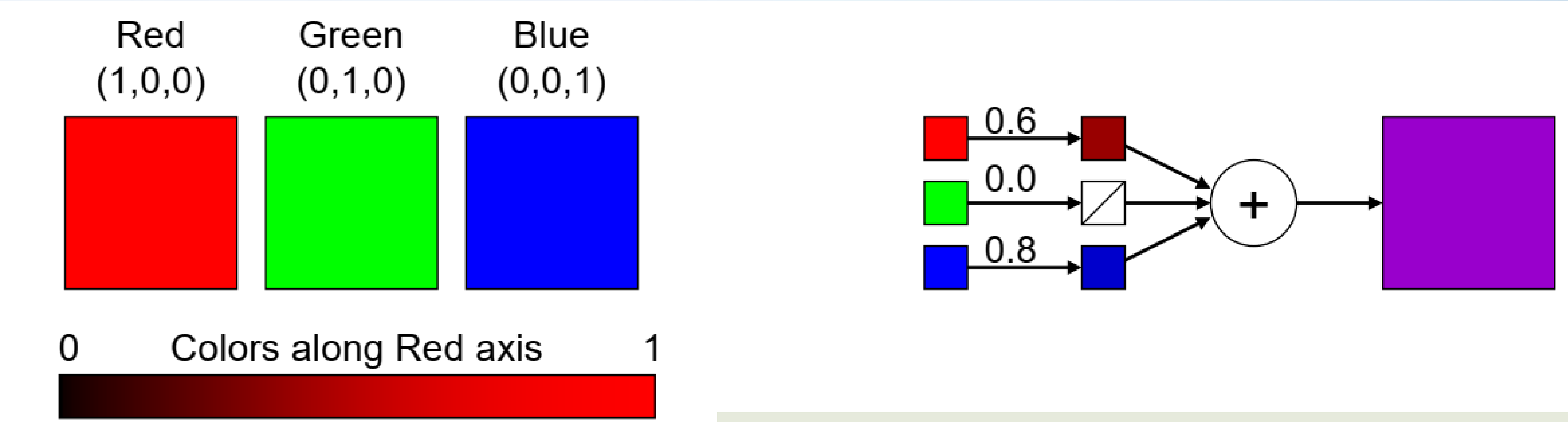
Binary Image:

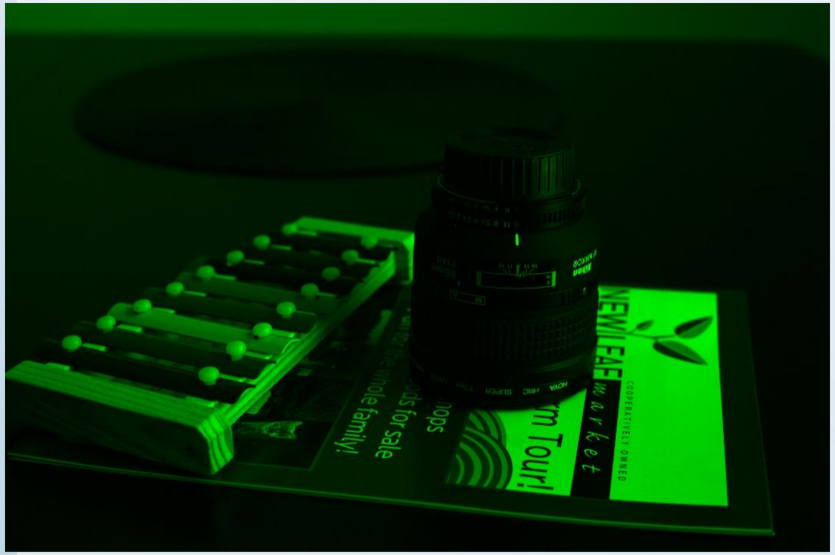
$$min = 0$$
$$max = 1$$

[Only two intensity values can be represented]

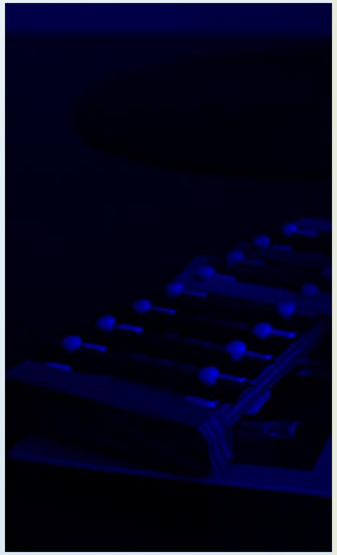
COLOR IMAGE



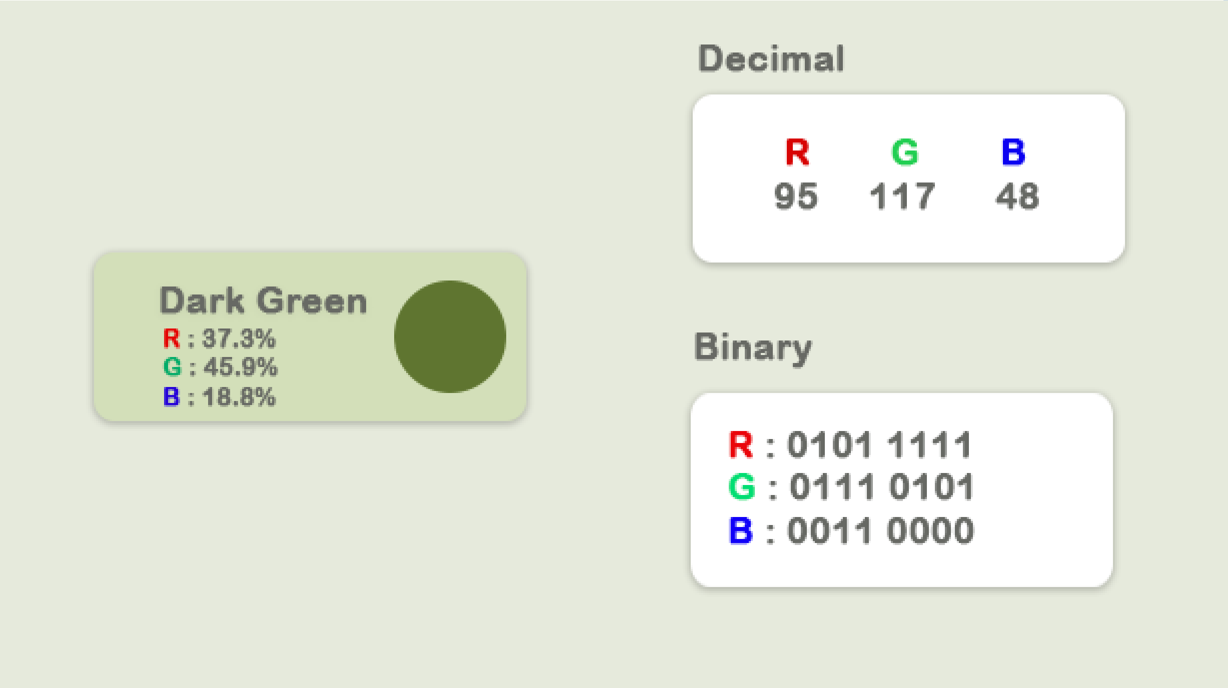


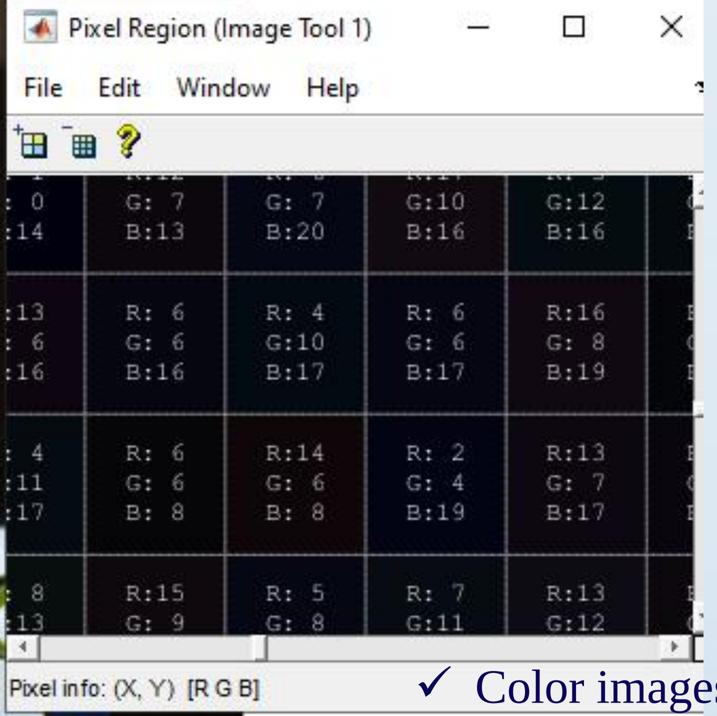


(c) Green-Channel of the color image



(d) Blue-Chan





A color/RGB image and the pixel values in a 3 × 3 neighborhood.

- ✓ Color images can be represented using three 2D arrays of same size, one for each color channel: **red (R), green (G), and blue (B) .**
- ✓ Each array element contains an 8-bit value, indicating the amount of red, green, or blue at that point in a [0, 255] scale .
- ✓ The combination of the three 8-bit values into a 24-bit number allows $\mathbf{2^{24}}$ (16,777,216, usually referred to as 16 million or **16 M**) color combinations.

Image Transformation

- An **image processing** operation typically defines a new image g in terms of an existing image f.

- We can transform either the range of f.

$$\boldsymbol{g(x,y) = t(f(x,y))}$$

- Or the domain of f:

$$g(x,y) = f\left(t_x(x,y), t_y(x,y)\right)$$

- What kinds of operations can each perform?