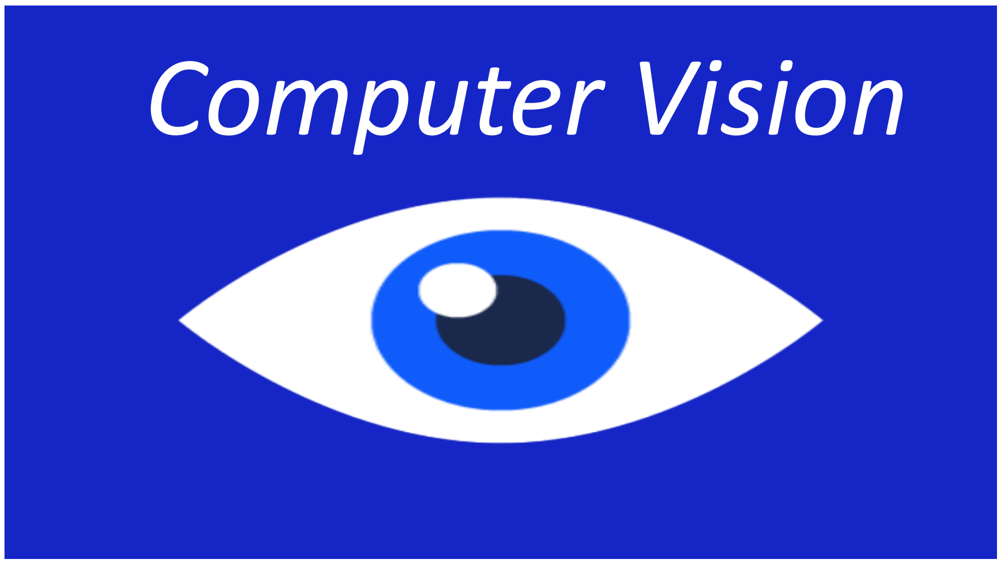
Welcome Class
to the course



Course Instructor
Dr. Harkeerat Kaur

The History of Vision

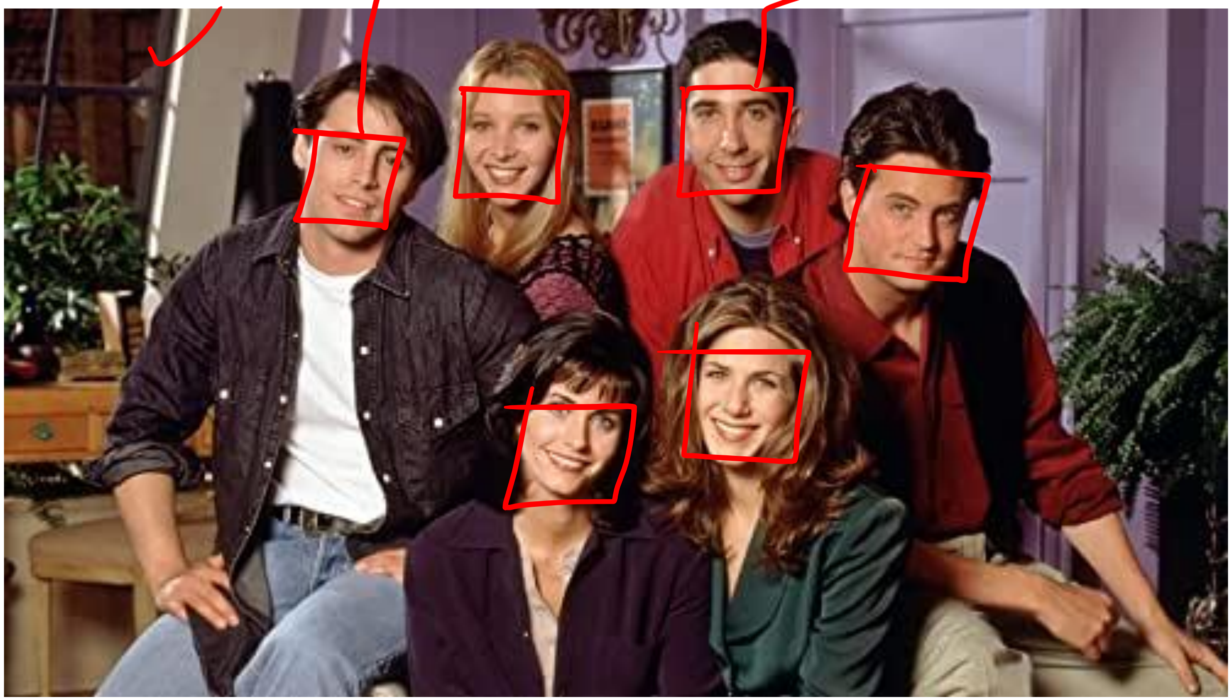
- https://www.youtube.com/watch?v=2W5hOJaFjxU

What is computer vision

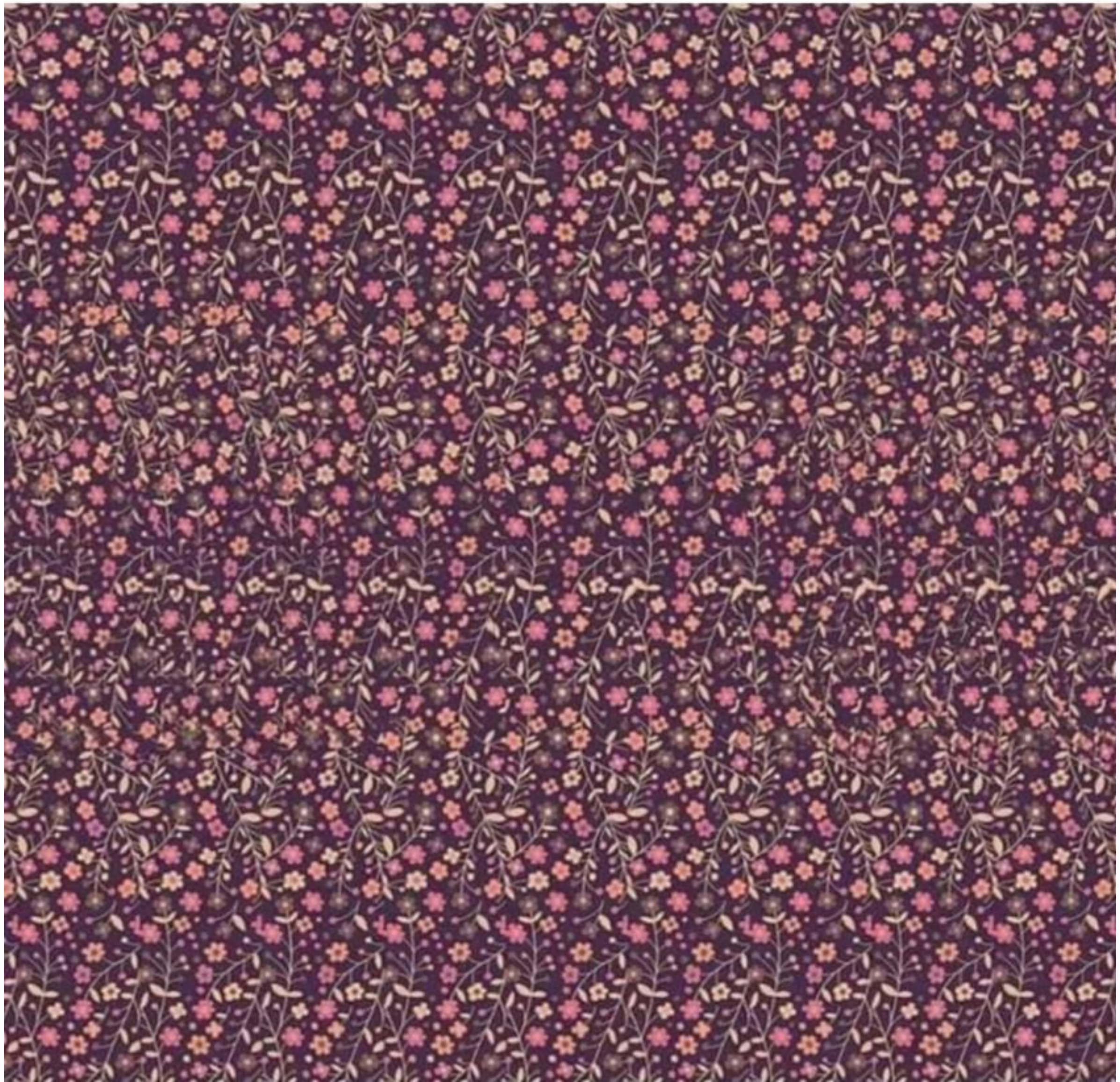
Figure 1.1 The human visual system has no problem interpreting the subtle variations in translucency and shading in this photograph and correctly segmenting the object from its background.

What is computer vision

face detection & recog.
males & females

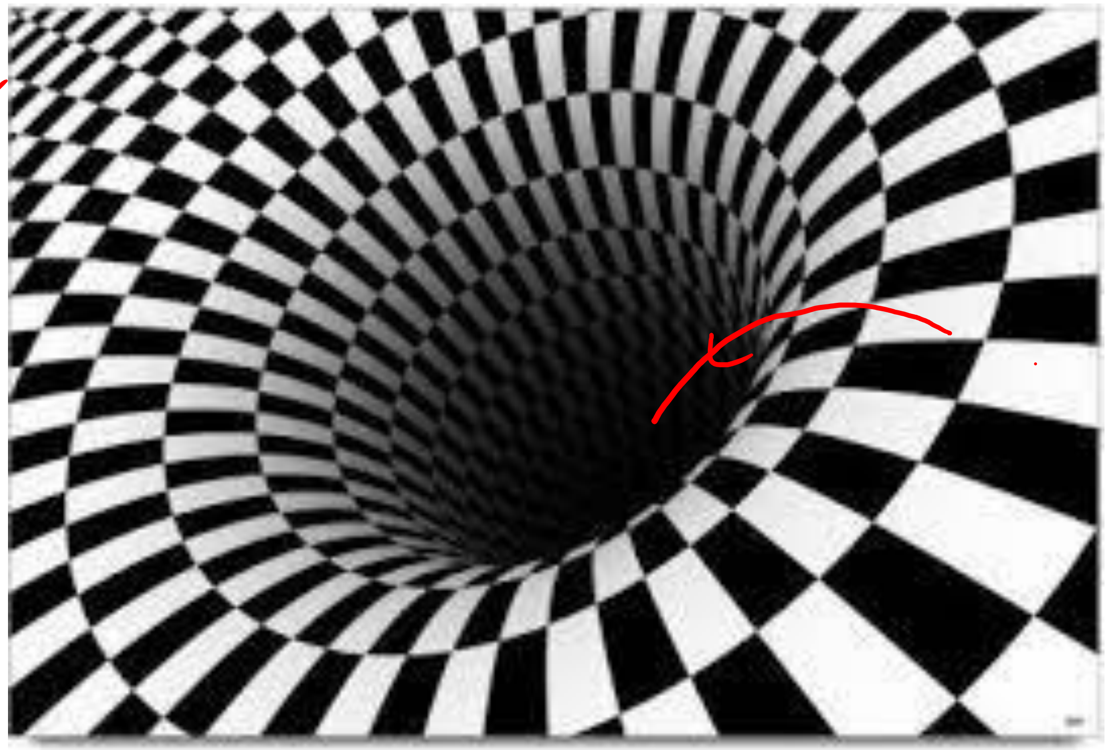


Put your nose to the center of the screen.
Focus on the center of the screen for 10 seconds.
Then pull away from the screen, slowly.



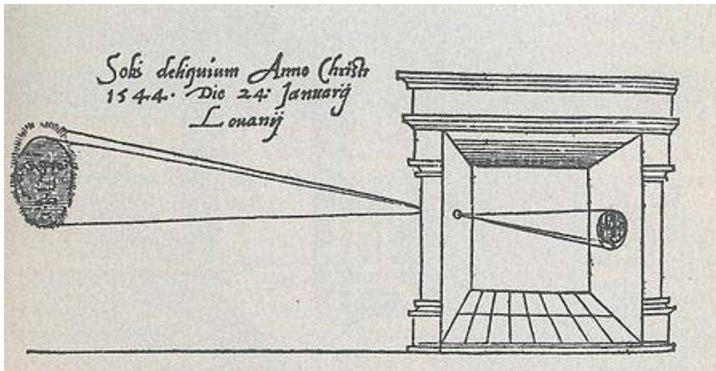
What is computer vision

2D → effect of vision 3D



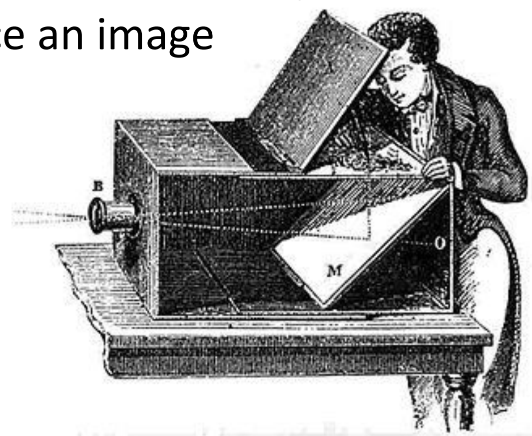
Picture is worth a thousand words

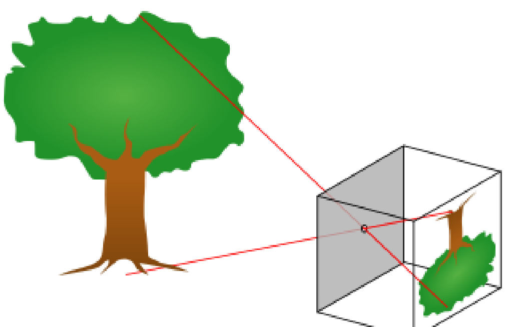
History of computer vision



First published picture of a camera obscura in Gemma Frisius' 1545 book *De Radio Astronomica et Geometrica*

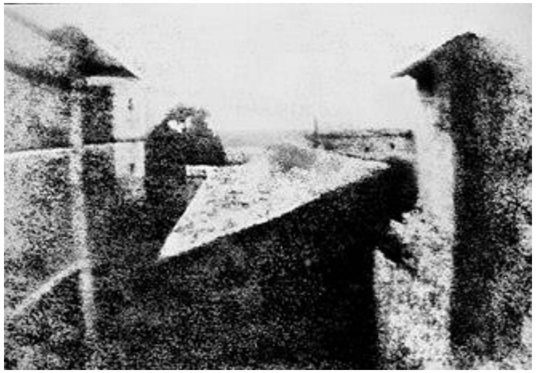
An artist using an 18th-century camera obscura to trace an image





Pinhole camera. Light enters a dark box through a small hole and creates an inverted image on the wall opposite the hole.[1]

History of computer vision



[View from the Window at Le Gras](#) (1825), the earliest surviving photograph

Late 19th-century studio camera

History of computer vision

Around 1990s……

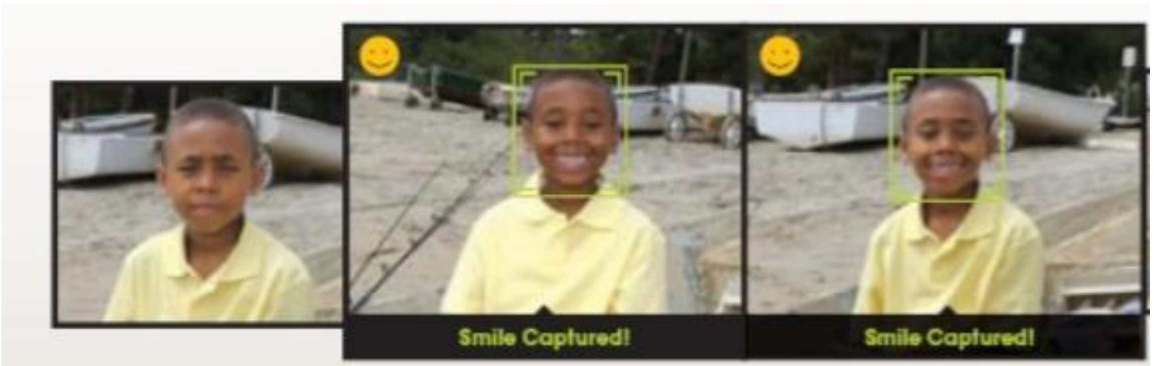
Evolution of computer vision

Voila Jone Face detector 2001

Many new digital cameras now detect faces

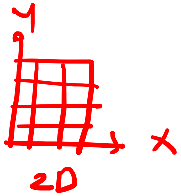
– Canon, Sony, Fuji, …

Evolution of computer vision



Images???

2D → matrix.
only discrete values.



1 Binary Image

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

→ pixel.
value = 1/0
✓ intensity
✓ brightness value.

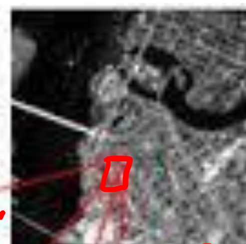
discrete

→ [0, 1, 1, ... 1, 0, 00...]

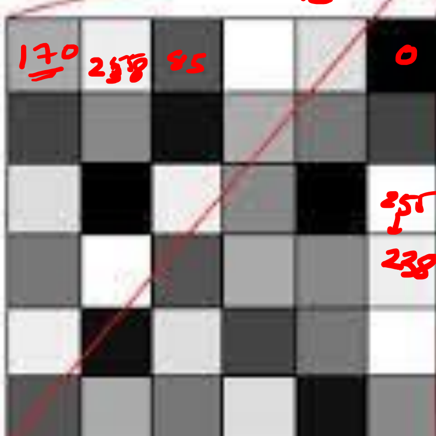
x y 2D ✓ 2D

2

0 → black
gray
255 → white



← grayscale values
[0, 255].

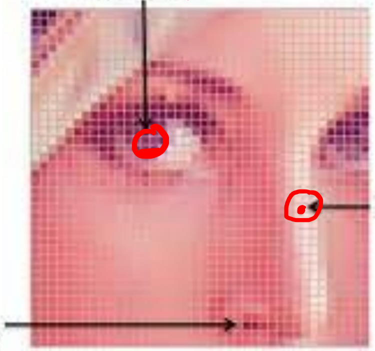


170

170	238	85	255	221	0
68	136	17	170	119	68
221	0	238	136	0	255
119	255	85	170	136	238
238	17	221	68	119	255
85	170	119	221	17	136

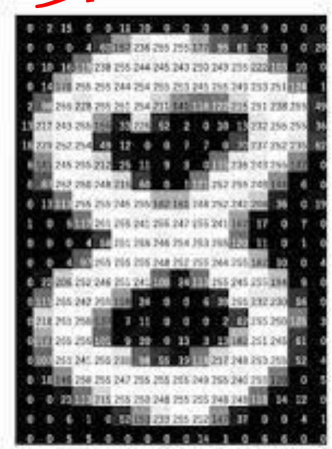
⇒ [170, 23...]

3



(0,0,0)
(255,255,255)
R G B
(R, G, B)
[0 - 255) *Red

4 Colorimage [R, G, B]
grayscale



The goal of computer vision

- Design algorithms to bridge the gap between pixels and "meaning"

2D - array

What we see

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 a computer sees

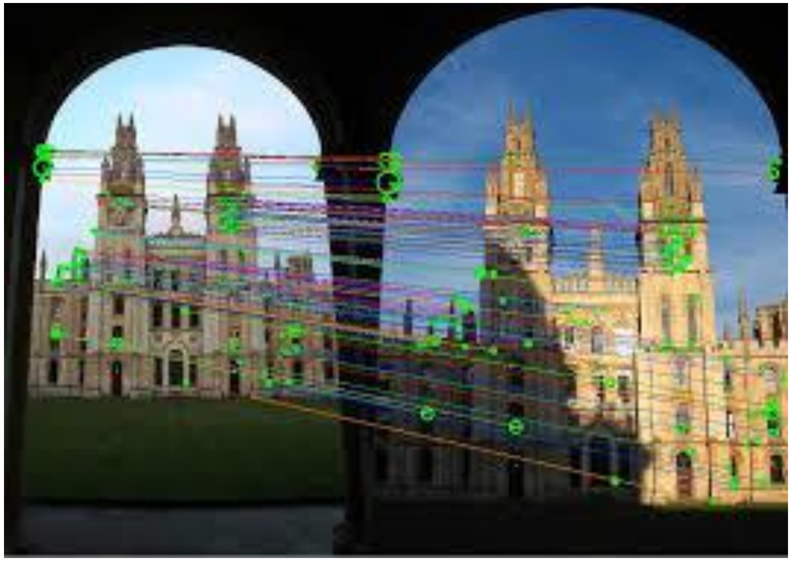
How do we interpret from images

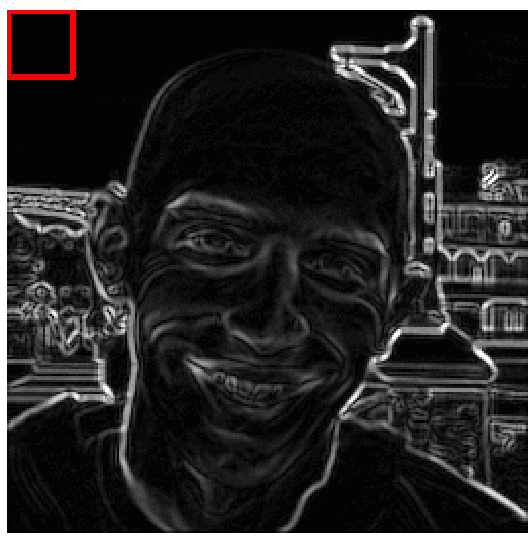
- Derive meaning full information from intensity values

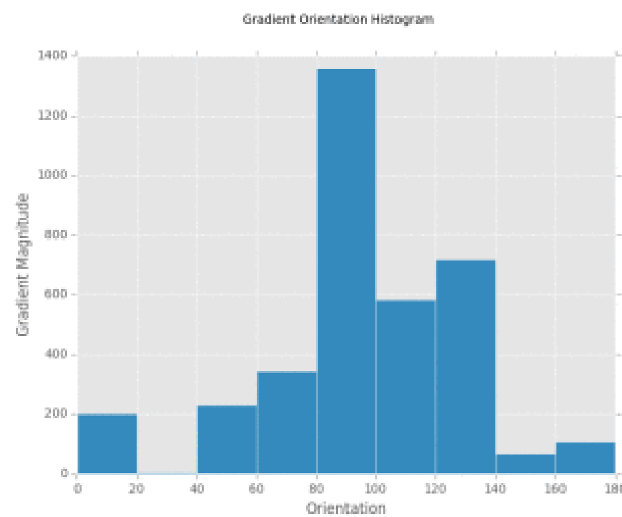
Segmentation.

Learning Motion from Images

Feature Extraction

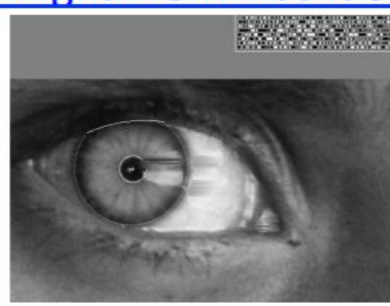


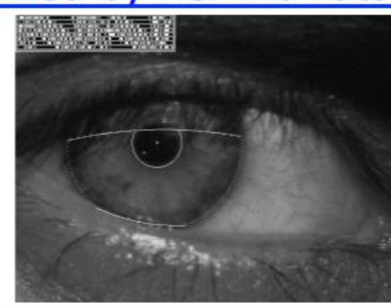




Biometrics

[How the Afghan Girl was Identified by Her Iris Patterns](#)





Source: S. Seitz

Convolution Neural networks

CNN$_L$*+RHN* : a black and white cat looking at itself in a mirror
CNN$_L$*+RNN* : a black and white cat sitting in front of a mirror
GRU : a black and white cat standing next to a mirror
LSTM : a black and white cat sitting in a bathroom sink
RNN : a cat sitting on the floor in a bathroom

- *there is a black tuxedo cat looking in the mirror*
- *two cats sitting on top of a wooden floor*
- *a cat looking at itself in the mirror next to a tripod*
- *a cat and a tripod sitting in front of a mirror*
- *a close up of a cat in a mirror*

CNN$_L$ *+RHN* : a man standing next to a child on a snow covered slope
CNN$_L$ *+RNN* : a man and a woman standing on a snow covered slope
GRU : a man and a child standing on a snow covered slope
LSTM : a man and a child are standing in the snow
RNN : a man and a woman are skiing on the snow

- *a woman and child in ski gear next to a lodge*
- *a man and a child are smiling while standing on skiis*
- *a young man poses with a little kid in the snow*
- *an adult and a small child dressed for skiing*
- *a man and a little girl in skis stand in front of a mountain lodge*

Joining the dots…….

The contents

Course Contents:	Topics	No. of Hours
1	Introduction and Digital Image Fundamentals	3
2	Image Enhancement and Restoration	5
3	Image Enhancement in the Frequency domain	5
4	Image Segmentation, Representation and Description	3
5	Harris Detector, Sift, Point Matching, Ransac, Local Binary Pattern	3
6	Projective geometry for Computer Vision:	3

The contents

Course Contents:	Topics	No. of Hours
7	Multiple views geometry: skew	3
8	Simultaneous Localization and Mapping	3
9	Optical Flow & Motion Analysis	3
10	3D reconstruction with a calibrated camera	3
11	Object Tracking	3
12	Object Recognition	3
13	Advanced Topics	2

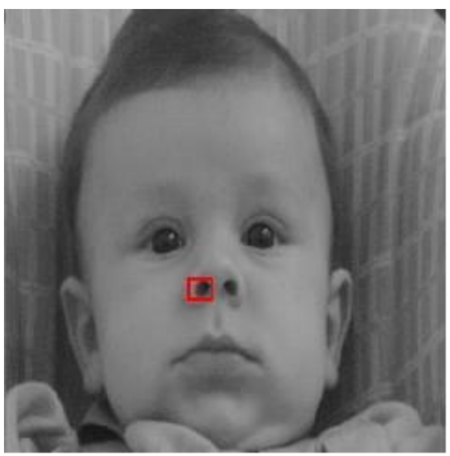
Text Book:	1. Forsyth & Ponce, Computer Vision: A Modern Approach, Pearson, 2002, ISBN 0130851981 2. Gonzalez R. C., Woods R. E., Digital Image Processing, Pearson Education (2007).

20
CT
(ms | 25%) (end sem 50%)

Image Processing Fundamentals

Digital Image Representation

- A digital image is an image f(x,y) that has been digitized both in spatial coordinates and brightness
- The value of f at any point (x,y) is proportional to the brightness (or gray level) of the image at that point



Pixel values in highlighted region

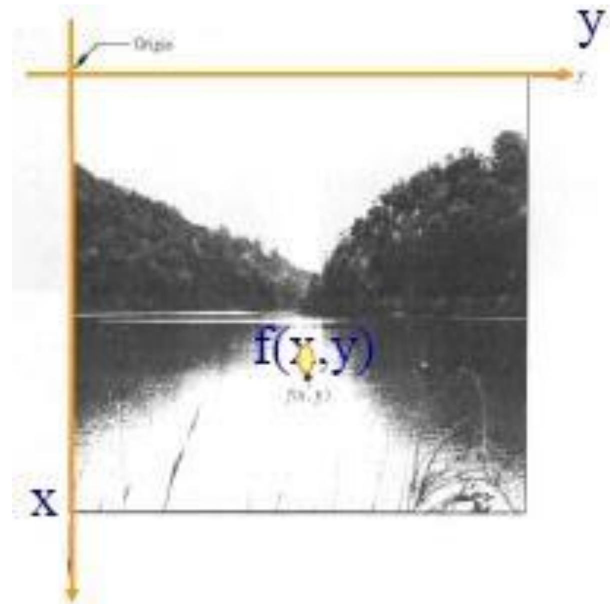
99	71	61	51	49	40	35	53	86	99
93	74	53	56	48	46	48	72	85	102
101	69	57	53	54	52	64	82	88	101
107	82	64	63	59	60	81	90	93	100
114	93	76	69	72	85	94	99	95	99
117	108	94	92	97	101	100	108	105	99
116	114	109	106	105	108	108	102	107	110
115	113	109	114	111	111	113	108	111	115
110	113	111	109	106	108	110	115	120	122
103	107	106	108	109	114	120	124	124	132

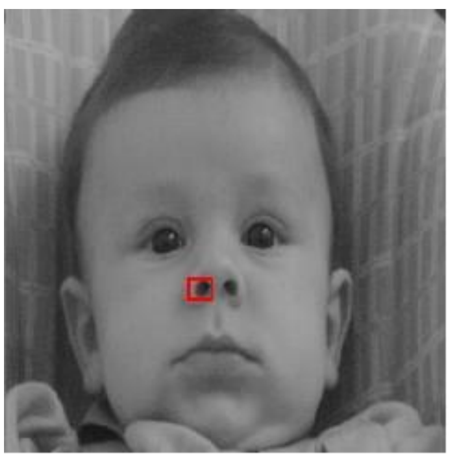
CAMERA ⇒ DIGITIZER ⇒ A set of number in 2D grid

Samples the analog data and digitizes it.

Digital Image Representation

A digital image is a representation of a two dimensional signal as a finite set of digital values, called picture elements or pixels



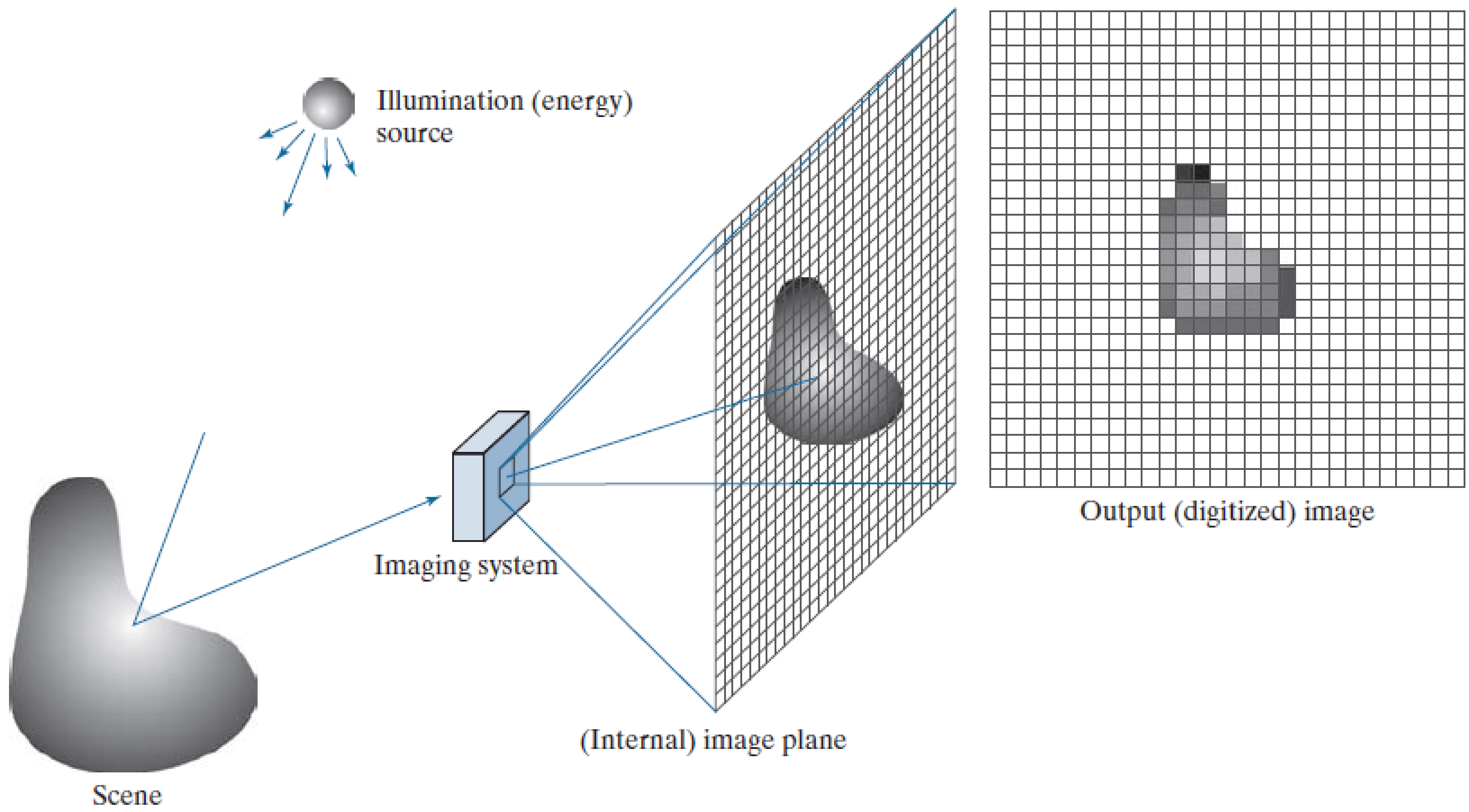


Pixel values in highlighted region

99	71	61	51	49	40	35	53	86	99
93	74	53	56	48	46	48	72	85	102
101	69	57	53	54	52	64	82	88	101
107	82	64	63	59	60	81	90	93	100
114	93	76	69	72	85	94	99	95	99
117	108	94	92	97	101	100	108	105	99
116	114	109	106	105	108	108	102	107	110
115	113	109	114	111	111	113	108	111	115
110	113	111	109	106	108	110	115	120	122
103	107	106	108	109	114	120	124	124	132

CAMERA ⇒ DIGITIZER ⇒ A set of number in 2D grid

Samples the analog data and digitizes it.


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

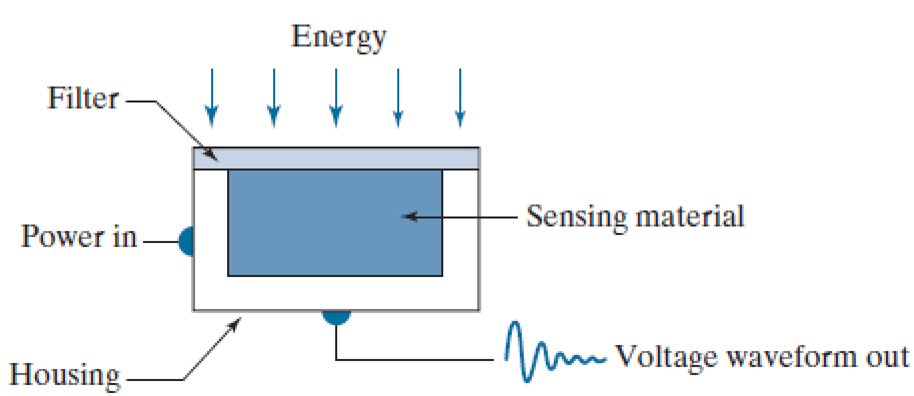

Energy
Filter
Sensing material
Power in
Housing
Voltage waveform out

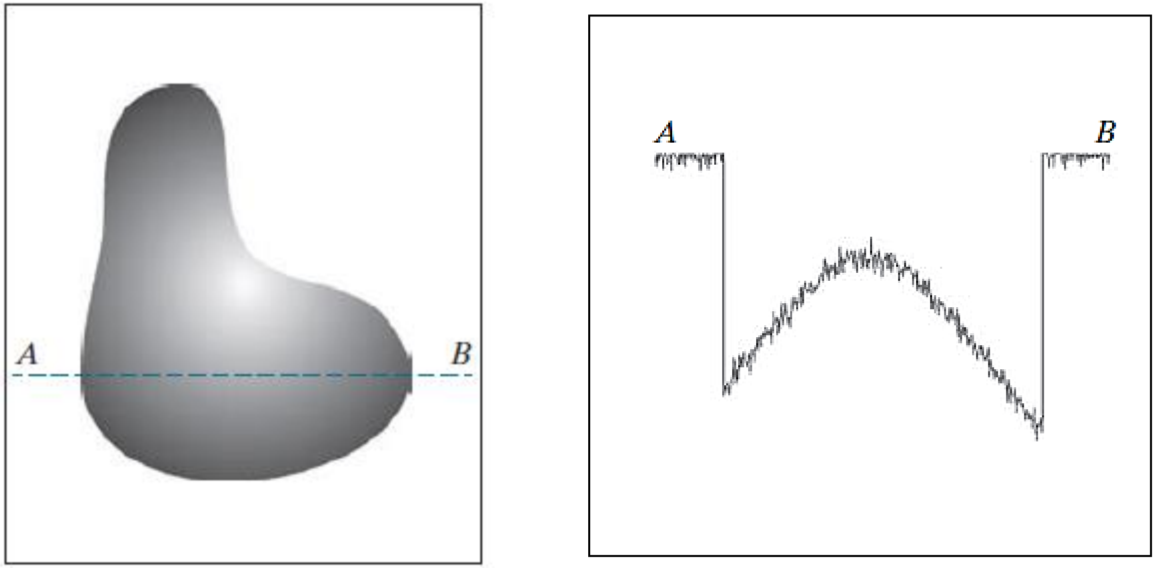
Image Sampling and Quantization

- The output of most sensors is a continuous voltage waveform whose amplitude and spatial behavior are related to the physical phenomenon being sensed.
- To create a digital image, we need to convert the continuous sensed data into a digital format:
  - Sampling: digitizing the coordinate values
  - Quantization: digitizing the amplitude values

    Usually thought of as the number of bits per sample of the signal (e.g., 1 bit/pixel b/w images, 8 bit/pixel grayscale images

    16-bit audio, 24-bit color images etc.)

Image Sampling and Quantization



a b
c d

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 a quantizat (c) Sampl

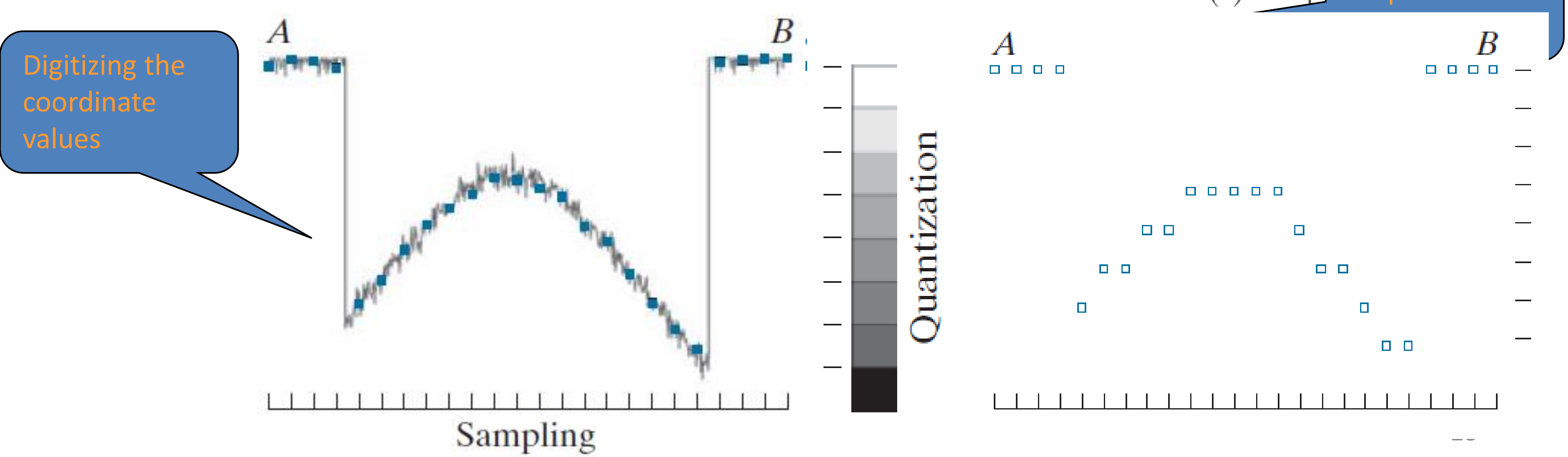
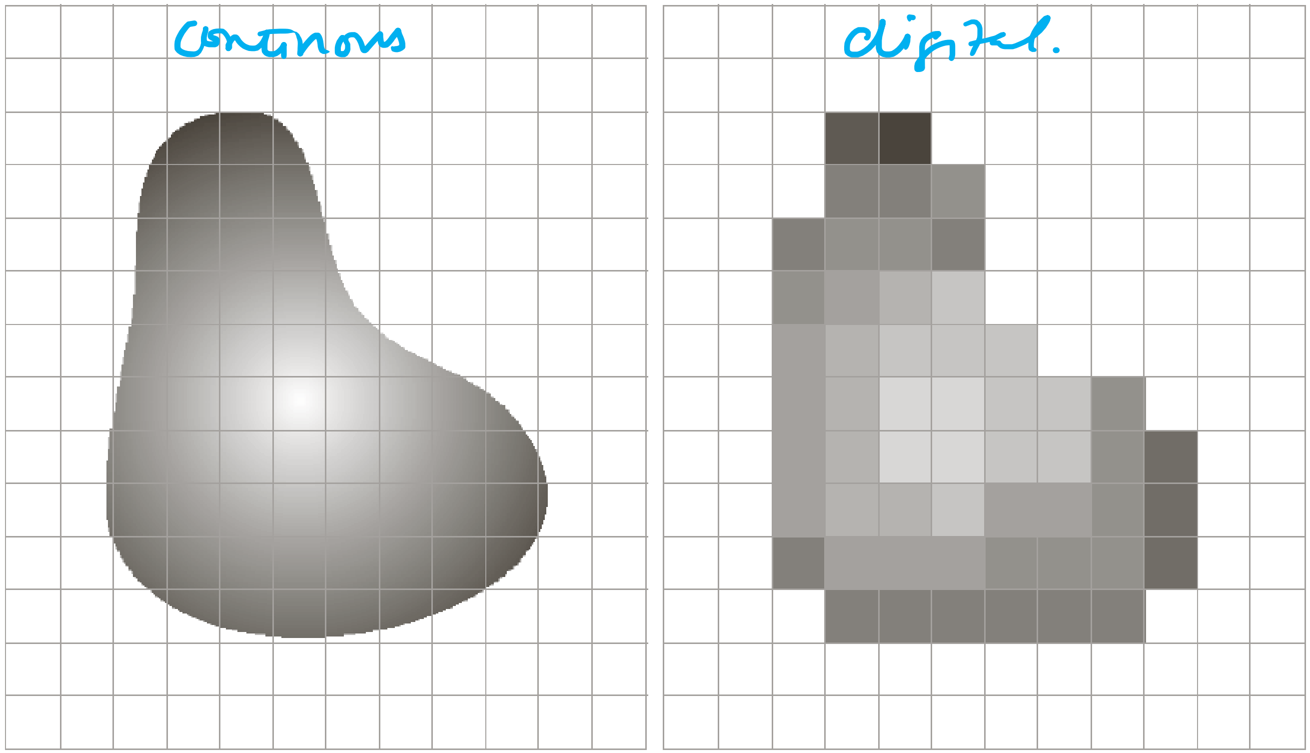


Image Sampling and Quantization



a b

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

The uniform quantizer

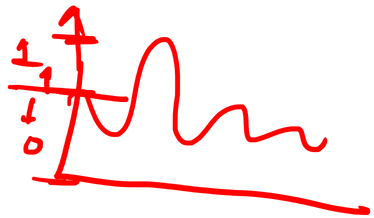
- The input brightness range $\mathrm{u} \in [Lmin, Lmax]$
- No. of bits of the quantizer = B
- Reconstruction levels =L= 2^b
- The expressions of the decision levels:
- T_1=Lmin, T_{L+1}=Lmax
- t_k-t_{k-1}=q = $(Lmax - Lmin)/\mathrm{L}$
- Computation of the quantization error:

  for a given image of size M×N pixels,
  - U – non-quantized, and U' – quantized
  - estimate the MSE: $err = \frac{1}{MN}\sum_0^{M-1}\sum_o^{N-1}(u(m,n) - u'(m,n))^2$

Results

- B=1 => L=2

$2^1=2$



Non-quantized image

$u(m,n)$

Quantized image

2 bits per pixel

$u'(m,n)$

Quantization error; MSE=36.2

$$\sum_{i,j=0}^{m,n} \sum \left(u(m,n)-u'(m,n)\right)^2$$

Some Results

bits per pixel

$(m,n) \rightarrow$ pixel coord

- B=2 => L=4 (2^B)

(0,0)

Non-quantized image

(u)

$u(m,n)$

(0,0)

Quantized image

(u')

$u'(m,n)$

Quantization error; MSE=15

$$\frac{1}{(m \times n)} \sum \sum \left(u'(m,n) - u(m,n)\right)^2$$

Some Basics

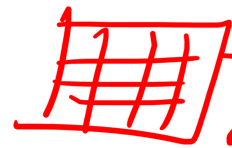
- Dynamic Range:

- Image Contrast:

- Contrast Ratio:
- appreciable number of pixels in image have a high dynamic range-> high contrast.
- Image low dynamic range -> dull, low contrast

Some Basics

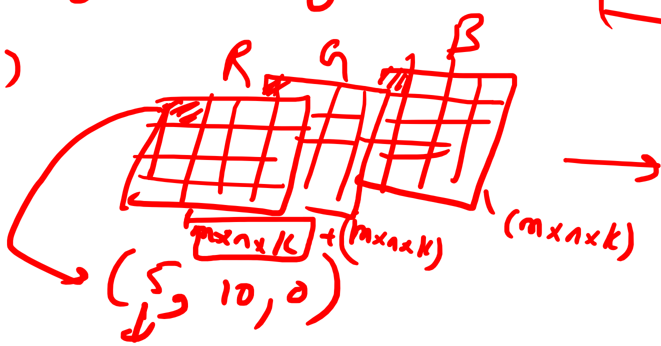
- No. of bits required to store an image ×3
- Size: Mx N
- Intensity levels= 2^K?



→ K-bits per pixel

Size of the image = $\boxed{m \times n \times K \times 3}$ - bits.

(m,n)



Spatial and Intensity Resolution

- Spatial resolution

- Intensity resolution

Image → Interpolation → Tech — NN
B-linear
B-cubic

Spatial and Intensity Resolution

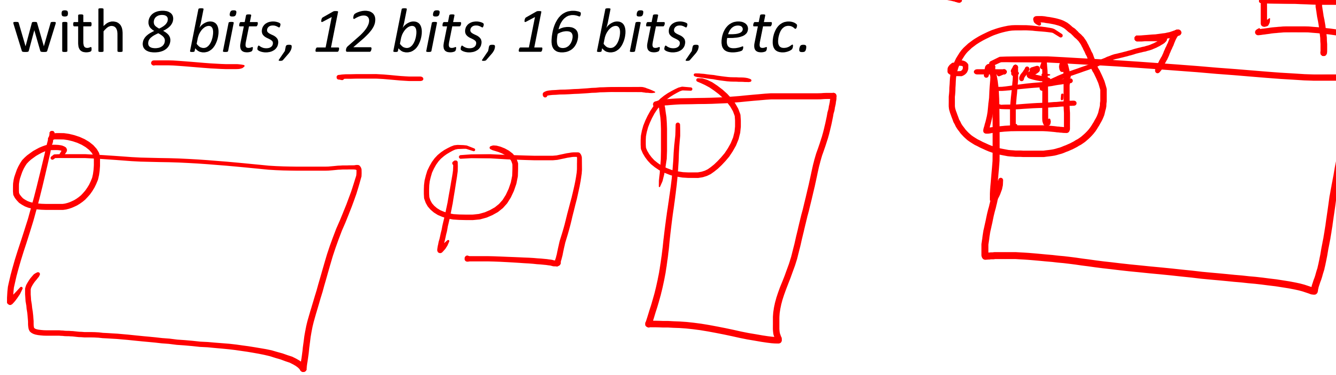
sampling

- Spatial resolution = sampling → per unit inch.
  - A measure of the smallest discernible detail in an image
  - stated with *line pairs per unit distance, dots (pixels) per unit distance, dots per inch (dpi)*

1024x1024

= quantization.

- Intensity resolution
  - The smallest discernible change in intensity level
  - stated with *8 bits, 12 bits, 16 bits, etc.*



Spatial and Intensity Resolution

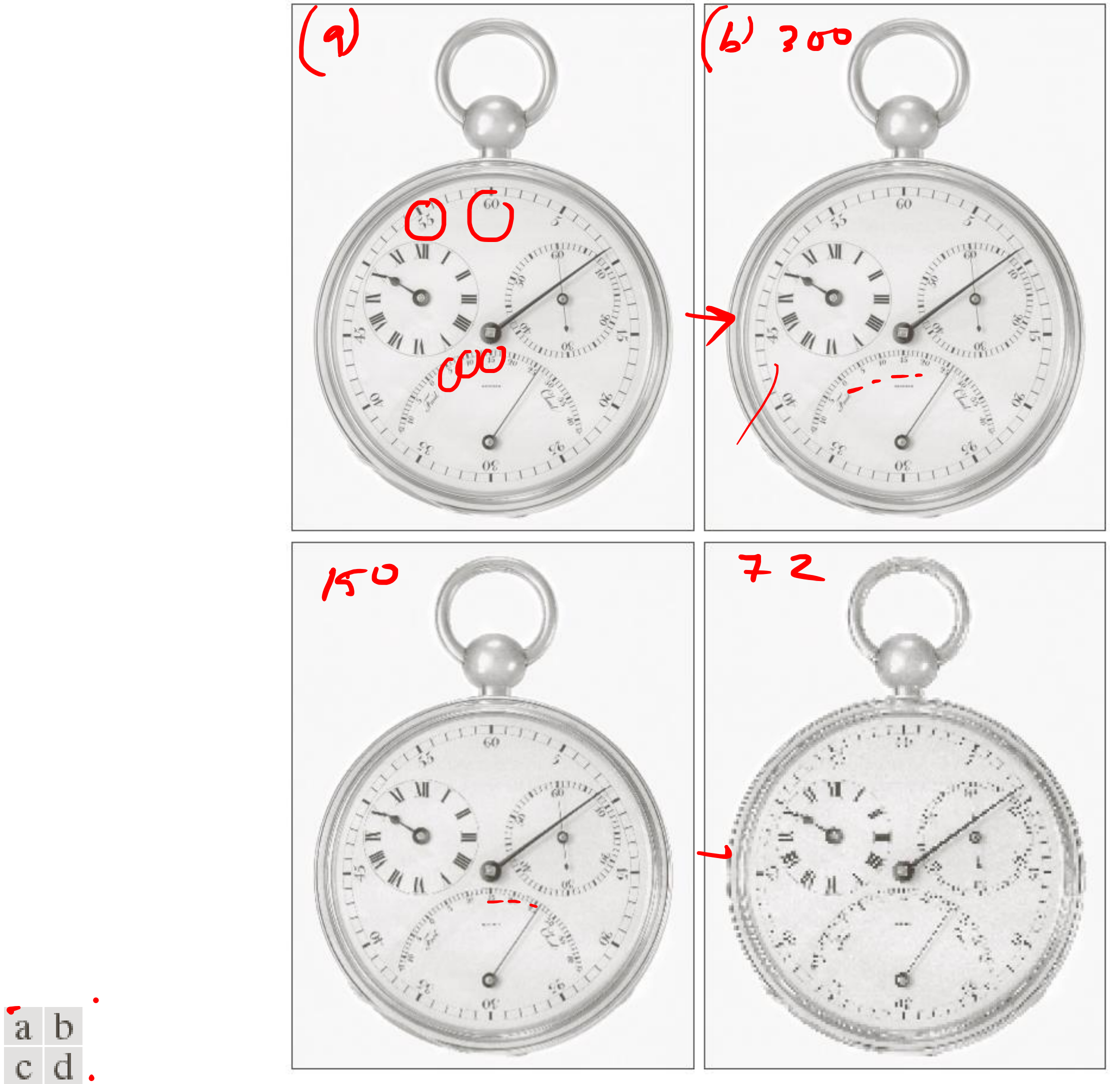
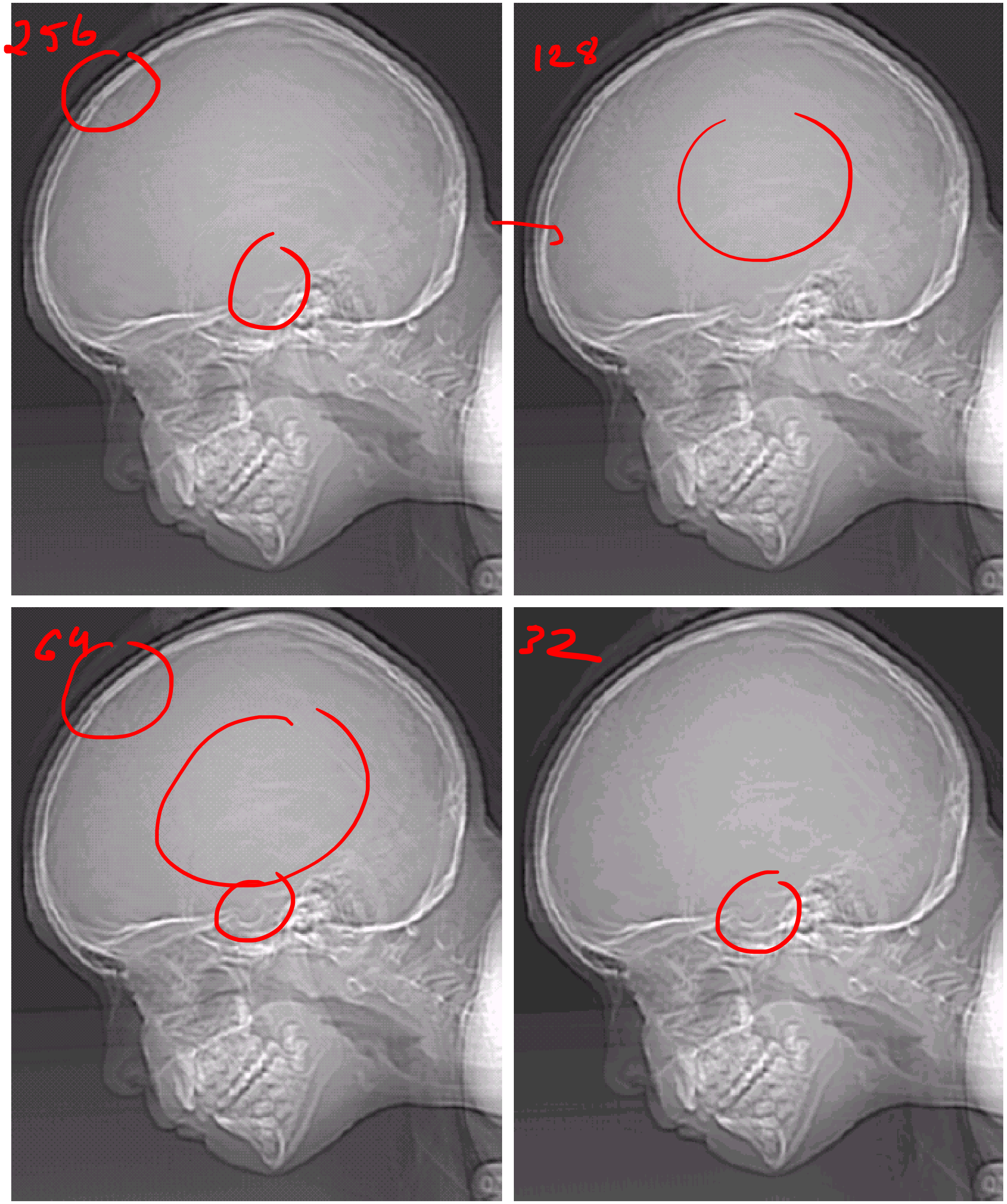


FIGURE 2.20 Typical effects of reducing spatial resolution. Images shown at: (a) 1250 dpi, (b) 300 dpi, (c) 150 dpi, and (d) 72 dpi. The thin black borders were added for clarity. They are not part of the data.

Spatial and Intensity Resolution



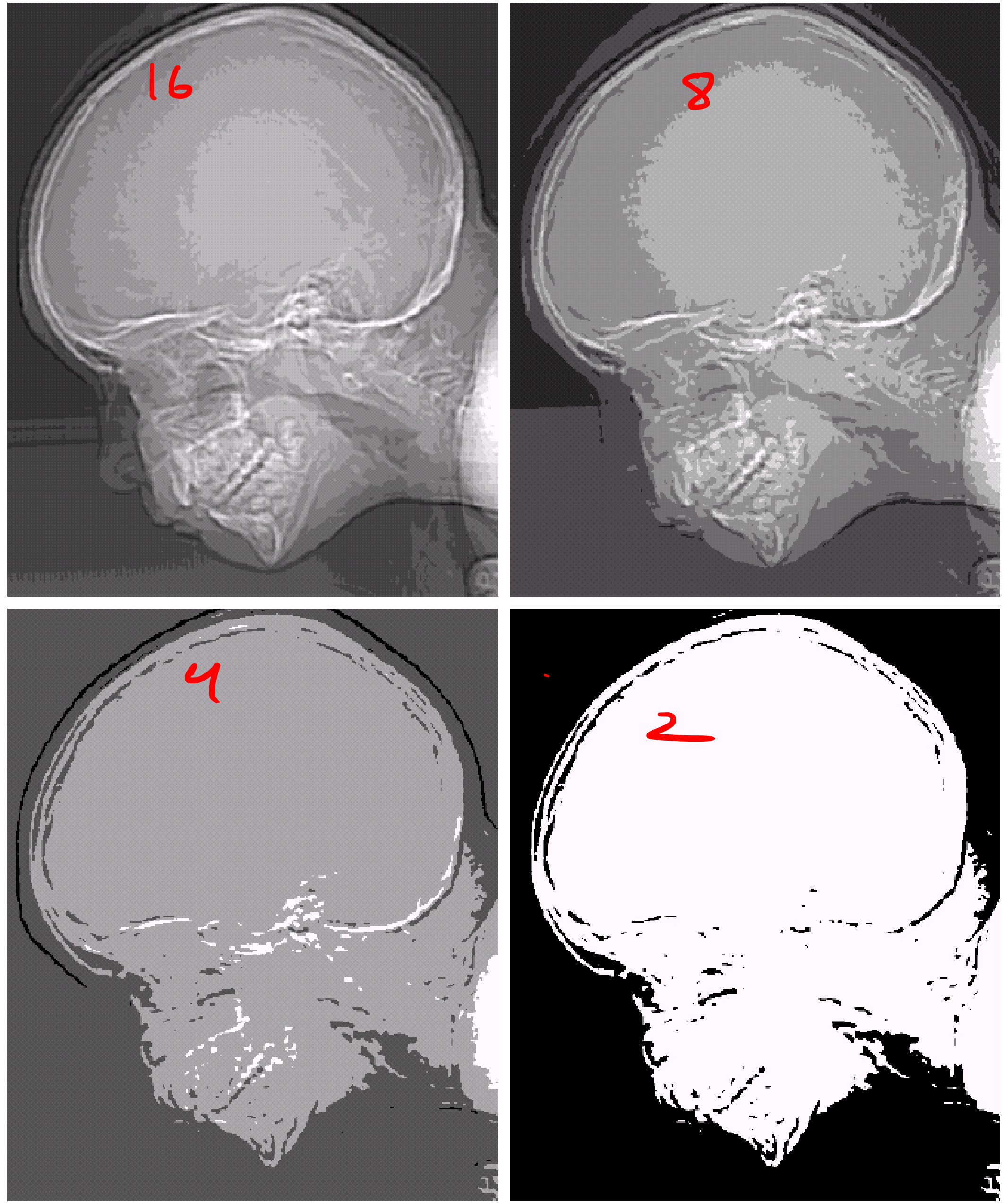
a b
c d

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.

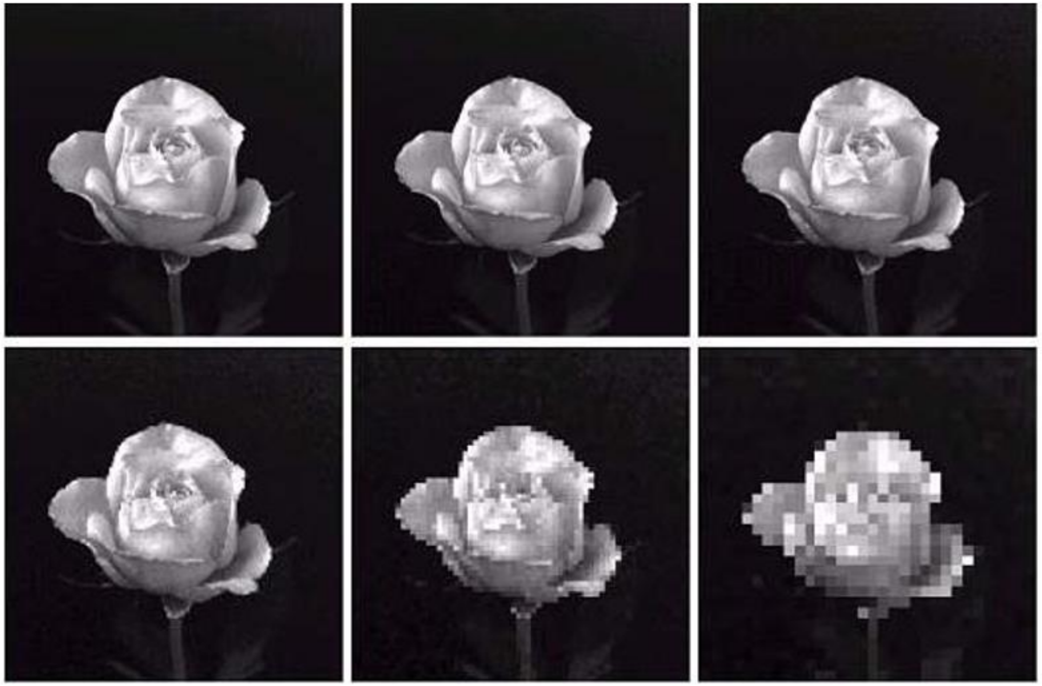
Spatial and Intensity Resolution

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.)



Checkerboard Effect-Spatial Resolution



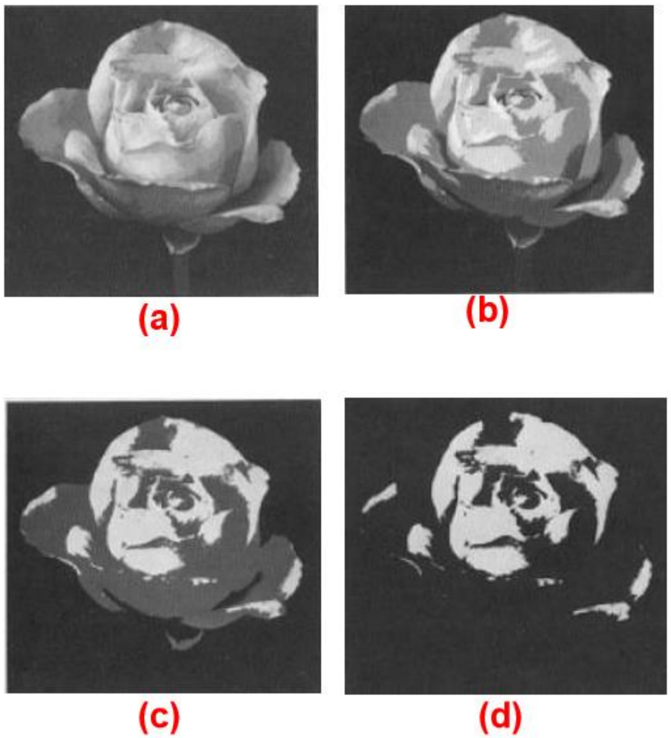
a	b	c
d	e	f

(a) 1024x1024
(b) 512x512
(c) 256x256
(d) 128x128
(e) 64x64
(f) 32x32

All images are of 8 bits

- If the resolution is decreased too much, the checkerboard effect can occur

False Contouring –Intensity Resolution



(a) Gray level = 16
(b) Gray level = 8
(c) Gray level = 4
(d) Gray level = 2

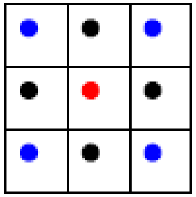
- If the gray scale is not enough, the smooth area will be effected
- False contouring can occur on the smooth area which has fine gray scales

Basic Relationship b/w Pixels

- Neighbors of a pixel
- Adjacency, and its types
- What is meant by Connectivity
- How to define Path, Boundary
- Region
- Labeling of connected components
- Distance Measures

Basic Relationship b/w Pixels

- Neighbors of a Pixel: A pixel p at coordinate (x,y) has
  - $N_4(p)$: 4 neighbors of p
    - (x-1,y), (x+1,y), (x,y-1), (x,y+1)
  - $N_D(p)$: 4 diagonal neighbors of p
    - (x+1,y+1), (x+1,y-1), (x-1,y+1), (x-1,y-1)
  - $N_8(p)$: 8 neighbors of p
    - a combination of $N_4(p)$ and $N_D(p)$



	x	
x	p	x
	x	

x		x
	p	
x		x

x	x	x
x	p	x
x	x	x

Connectivity

- Useful for define object boundaries
- Image components or regions

If f(x,y)>Th
$\Rightarrow$ f(x,y) is object or fore ground
Else
f(x,y) is Background

- Two pixels are connected if they are adjacent in some sense.
- Adjacent
  - A pixel p is adjacent to a pixel q if they are connected

Connectivity

- Two pixels are connected iff
  - They are neighbors $N_{4,}$ $N_{8,}$ N_D
  - They have similar intensity values

- Let V be the set of gray-level values used to defined connectivity
  - 4-connectivity: 2 pixels p and q with values from V are 4-connected if q is in the set $N_4(p)$
  - 8-connectivity: 2 pixels p and q with values from V are 8-connected if q is in the set $N_8(p)$

Connectivity

- m-connectivity (mixed connectivity)
  - 2 pixels p and q having value belonging to set V{ } are m-connected if
    - q is in the set $N_4(p)$ or
    - q is in the set $N_D(p)$ and the set $N_4(p) \cap N_4(q)$ has no pixels whose values are from V
    - **(the set of pixels that are 4-neighbors of both p and q whose values are from V)**

- Example:

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

Pixel arrangement for V = {1}

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

Pixels that are 8-adjacent to the center pixel (shown dashed)

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

m-adjacency

m-connectivity eliminates the multiple path connections that arise in 8-connectivity

Path

- Path
  - A path from pixel p with coordinates (x,y) to pixel q with coordinates (s,t) is a sequence of distinct pixels with coordinates $(x_0,y_0), (x_1,y_1),\ldots(x_n,y_n)$

    where $(x_0,y_0)=(x,y)$, $(x_n,y_n)=(s,t)$ and pixels (x_i,y_i), (x_{i-1},y_{i-1}) are adjacent for $1 \leq i \leq n$
  - n is the length of the path
  - We can define 4-,8-, or m-paths depending on type of adjacency specified

Adjacency

- Two pixels p,q are ajacent iff
  - 4 connected
  - 8 connected
  - M connected
- Two image area subsets S1 and S2 are adjacent if some pixel in S1 is adjacent to some pixel S2

Examples: Adjacency and Path

V = {1, 2}

$0_{1,1}$	$1_{1,2}$	$1_{1,3}$
$0_{2,1}$	$2_{2,2}$	$0_{2,3}$
$0_{3,1}$	$0_{3,2}$	$1_{3,3}$

0	1	1
0	2	0
0	0	1

8-adjacent

0	1	1
0	2	0
0	0	1

m-adjacent

Examples: Adjacency and Path

V = {1, 2}

$$\begin{matrix} 0_{1,1} & 1_{1,2} & 1_{1,3} \\ 0_{2,1} & 2_{2,2} & 0_{2,3} \\ 0_{3,1} & 0_{3,2} & 1_{3,3} \end{matrix} \qquad \begin{matrix} 0 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{matrix} \qquad \begin{matrix} 0 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{matrix}$$

8-adjacent

m-adjacent

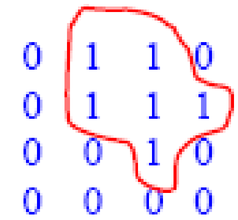
The 8-path from (1,3) to (3,3):
(i) (1,3), (1,2), (2,2), (3,3)
(ii) (1,3), (2,2), (3,3)

The m-path from (1,3) to (3,3):
(1,3), (1,2), (2,2), (3,3)

m-connectivity eliminates the multiple path connections that arise in 8-connectivity

Basic Relationship b/w Pixels

- Connected Components

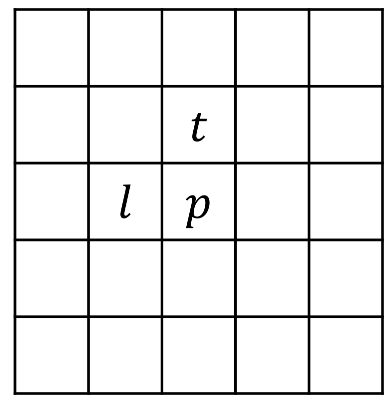


  - S: a subset of pixels in an image
  - *Two pixels p and q are said to be connected in S* if there exists a path between them consisting entirely of pixels in S
  - For any pixel p in S, the set of pixels that are connected to it in S is called a connected component of S
- Region

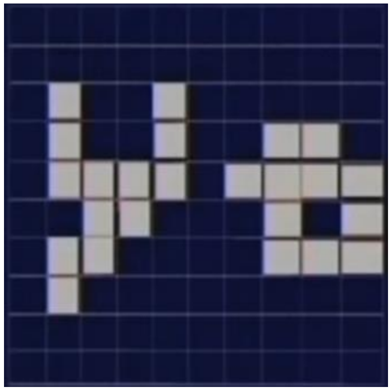
  - R: a subset of pixels in an image
  - R is called a region if every pixel in R is connected to any other pixel in R
  - Boundary (border or contour) of a region: set of pixels in the region that have one or more neighbors that are not in R

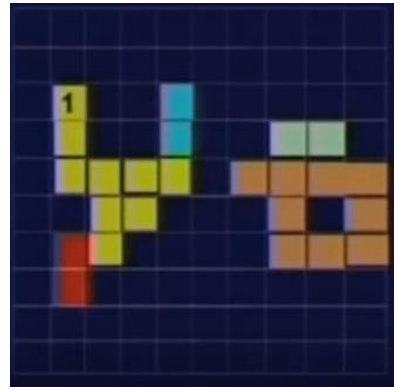
Labeling of Connected Components

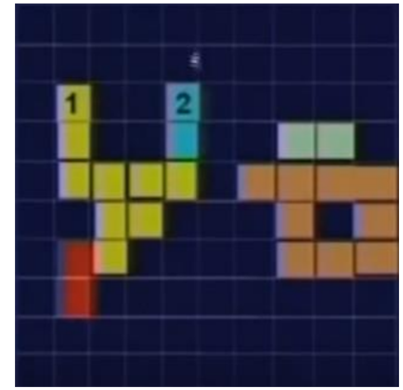
- Scan the image from left to right, assume 4 -connectivity
- Let p denote the pixel at any step in the scanning process
- Let t denote the upper neighbor of p
- Let l denote the left-hand neighbors of p, respectively
- When we get to p, points t and l have already been encountered and labeled if they were 1’s
- If the value of p = 0, move on
- If the value of p = 1, examine l and t
  - If they are both 0, assign a new label to p
  - If only one of them is 1, assign it’s label to p
  - If they are both 1
    - If they have same label, assign that label to p
    - If not, assign one of the labels to p and make a note that the two labels are equivalent (r and t are connected through p)
- At the end of the scan, all points with value 1 have been labeled



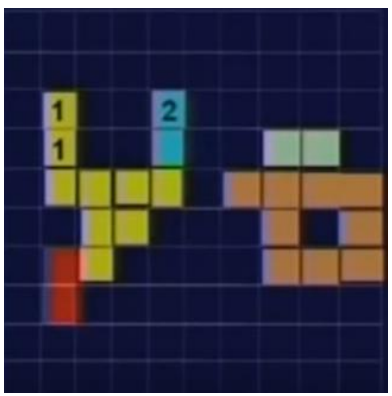
Example: Labeling of Connected Components



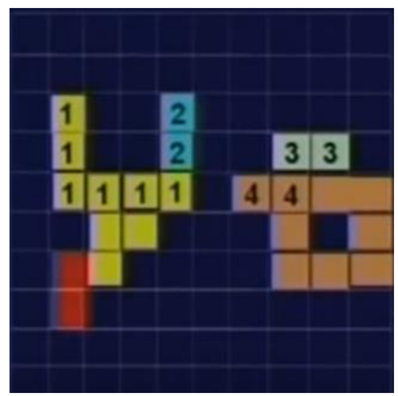




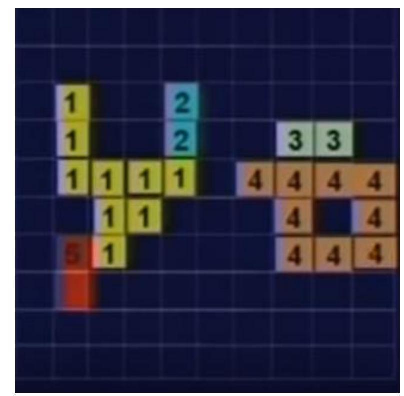
{1}{2}



{1,2} , {3}



{1,2} {3,4}



{1,2,5} {3,4}

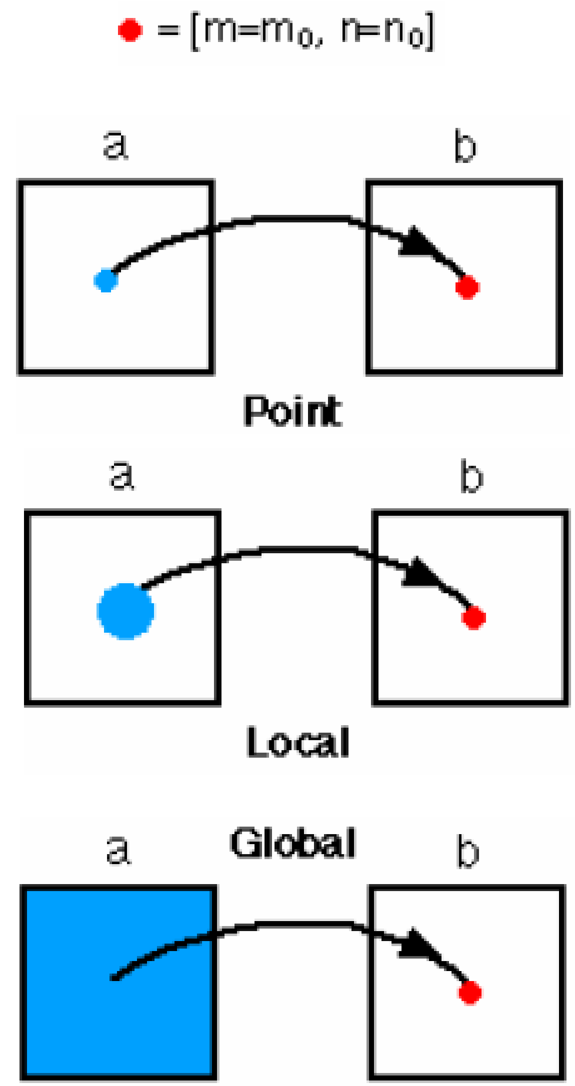
After completing the scan, do the second round and introduce a unique label to each equivalent class

Labeling of Connected Components

- What shall we do with 8-connected components?
- Do the same way but examine also the upper diagonal neighbors of p
  - If p = 0, move on
  - If p is 1
    - If all four neighbors are 0, assign a new label to p
    - If only one of the neighbors is 1, assign its label to p
    - If two or more neighbors are 1, assign one of the label to p and make a note of equivalent classes
    - After completing the scan, do the second round and introduce a unique label to each equivalent class

Types of Operations

- Operations that can be applied to digital images to transform an input image a[m,n] in to an output image b[m,n]
- Point
  - The output value at a specific coordinate is dependent only on the input value at that same coordinate
- Local
  - The output value at a specific coordinate is dependent on the input values in the neighborhood of that same coordinate
- Global
  - The output value at a specific coordinate is dependent on all the input values in the input image



Types of Operations

- Point Operation
  - Contrast Stretching
  - Binarization
    - Reduction of gray scale image to binary
- Local Operations
  - Convolution, Filtering
- Global Operations
  - Image Profiling
    - Histogram of ON-valued pixels summed across rows or columns of a binary image
  - Frequency Analysis
    - Images are often characterized not by particular objects but by the global configuration of all image objects and resulting patterns or texture

Arithmetic Operations on Images

- Used extensively, Carried out pixel by pixel
- Let I_1 and I_2 are two images of samedimension
  - Addition: used in image average to reduce noise
    - $I(x,y) = \min(I_1(x,y) + I_2(x,y), I_{max})$
  - Subtraction: basic tool in medical imaging
    - $I(x,y) = \max(I_1(x,y) - I_2(x,y), I_{min})$
  - Possibility of overflow and underflow are taken care by clipping
  - Clipping can be avoided by rescaling the range of intensity values encountered in both input images
  - Modified addition:
    - $I(x,y) = (I_1(x,y) + I_2(x,y))*255/(I_1+I_2)max$
  - Modified Subtraction :
    - $I(x,y) = (I_1(x,y)-I_2(x,y))*255/(I_1-I_2)min$

An Example of Image Multiplication and Division for shading correction

a b c

FIGURE 2.34 (a) Digital dental X-ray image. (b) ROI mask for isolating teeth with fillings (white corresponds to 1 and black corresponds to 0). (c) Product of (a) and (b).

Logical Operations on Images

- Carried out pixel by pixel
- Logical operations apply only to binary images, while arithmetic operations apply to multivalued pixels
- Used for tasks such as masking, feature detection, and shape analysis

- An Example

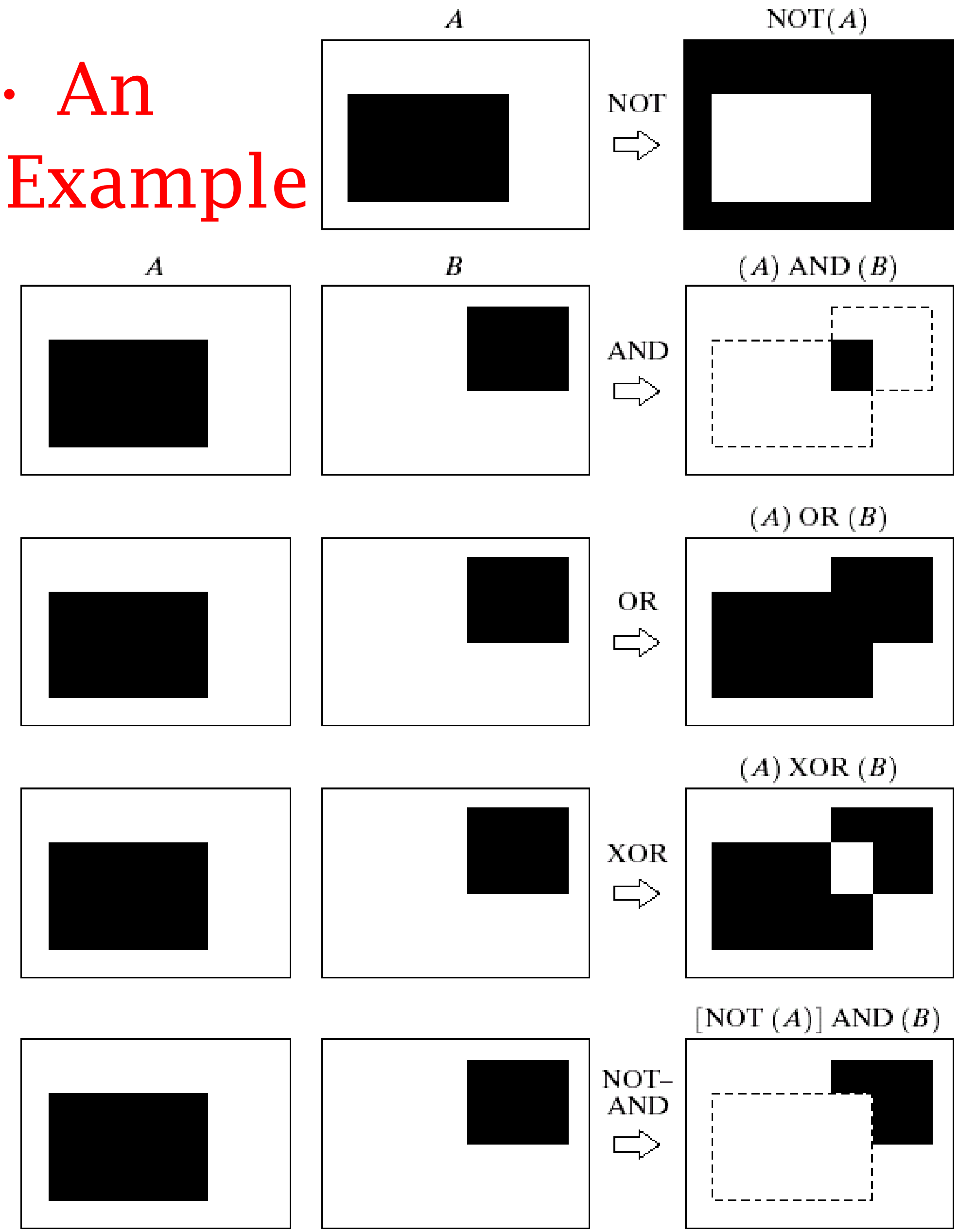
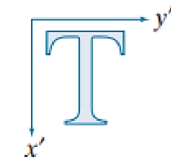
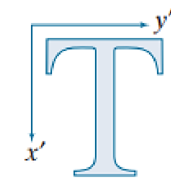
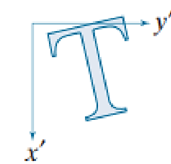
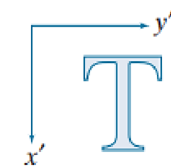
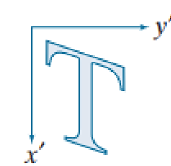
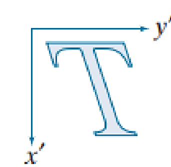


Image Geometry

- Basic Transformations: expressed in 3D Cartesian coordinate system (x,y,z)
  - Translation
  - Scaling
  - Rotation
  - Concatenation and Inverse Transformation

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)	$\begin{bmatrix} 1 & 0 & 0 \\ s_h & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x$ $y' = s_h x + y$	

Distance Measure

- The distance between two pixels ($p(x_1,y_1)$ and $q(x_2,y_2)$) in a given image can be given by three different types of measures
- D_e (Euclidian) distance
  - $D_e(p,q) = ((x_1-x_2)^2 + (y_1-y_2)^2)^{1/2}$
- D_4 distance
  - Also called city-blocking distance
  - $D_4(p,q) = |(x_1-x_2)| + |(y_1-y_2)|$
- D_8 distance
  - Also called chessboard distance
  - $D_8(p,q) = \max(|(x_1-x_2)|, |(y_1-y_2)|)$

Distance Metric

- The distance function is called metric if the following properties are satisfied:
  - D(p,q) is well-defined and finite for all p and q.
  - $D(p,q) \geq 0$; if $p=q$, then $D(p,q)=0$.
  - The distance $D(p,q) = D(q,p)$.
  - $D(p,q)+D(q,t) = D(p,t)$. (triangular inequality)