CMPE 185 Autonomous Mobile Robots

Perception: Computer Vision and Image Processing

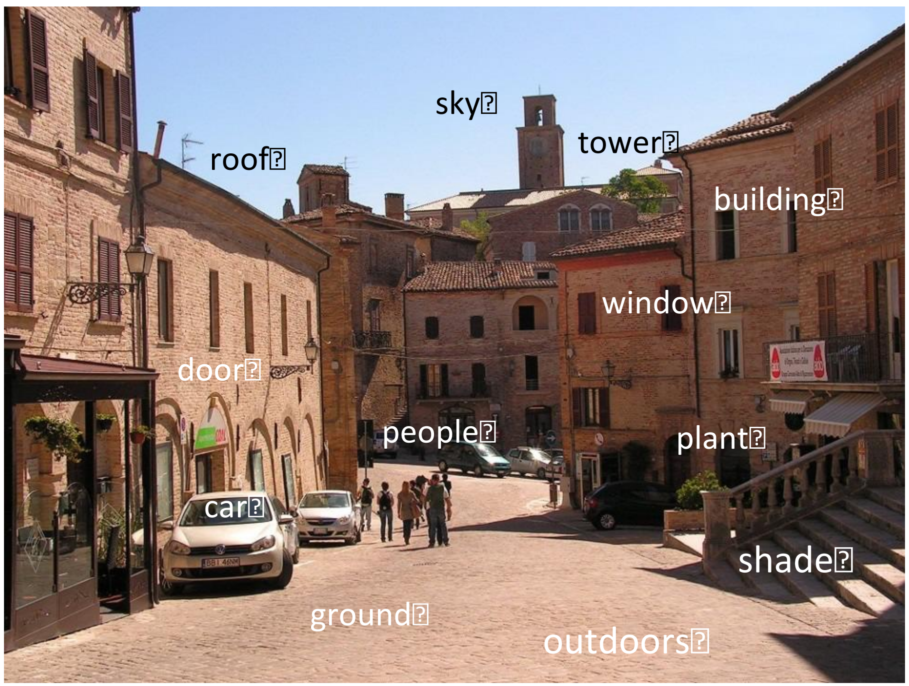
Dr. Wencen Wu
Computer Engineering Department
San Jose State University

Human Visual Capabilities

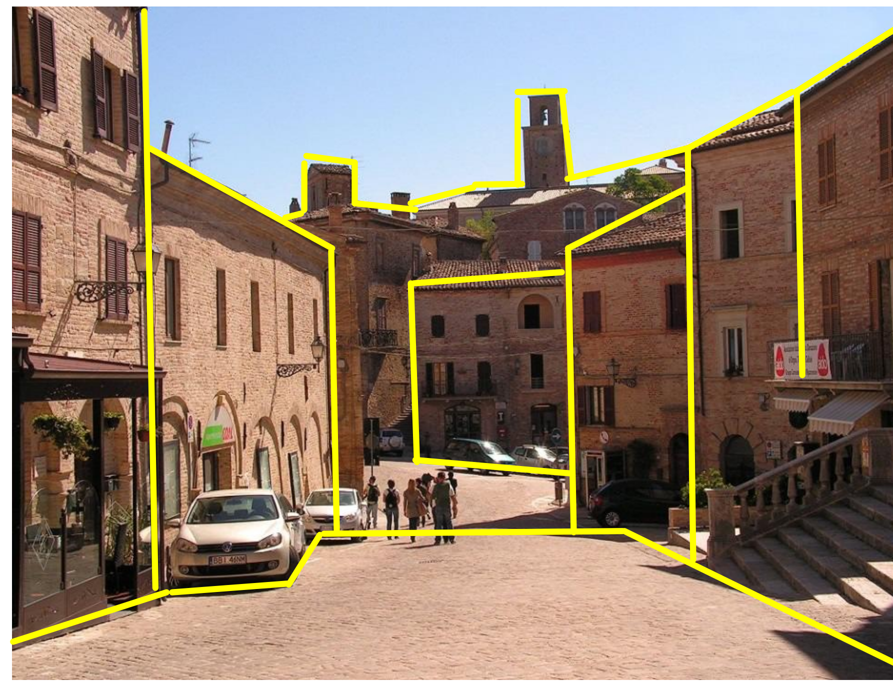
- Our visual system is very sophisticated
- Humans can interpret images successfully under a wide range of conditions – even in the presence of very limited cues

Computer Vision – What is it?

- Automatic extraction of "meaningful" information from images and videos
  - varies depending on the application



Semantic information



Geometric information

Computer Vision for Robotics

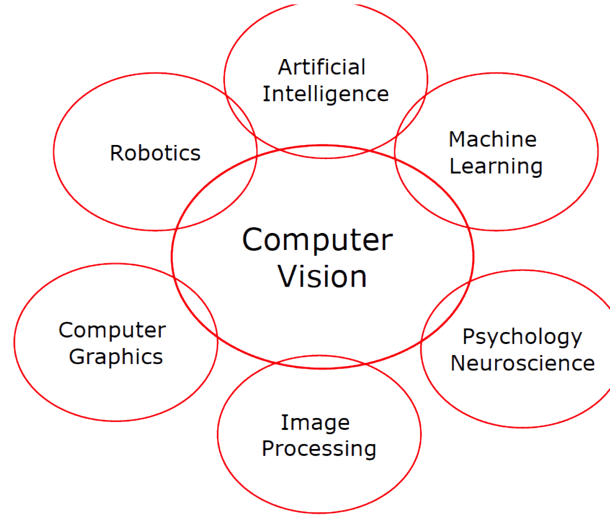
- Enormous descriptability of images → a lot of data to process (human vision involves 60 billion neurons!)
- Vision provides humans with a great deal of useful cues to explore the power of vision towards intelligent robots

Cameras:

- Vision is increasingly popular as a sensing modality:
  - descriptive
  - compactness, compatibility, ...
  - low cost
  - HW advances necessary to support the processing of images

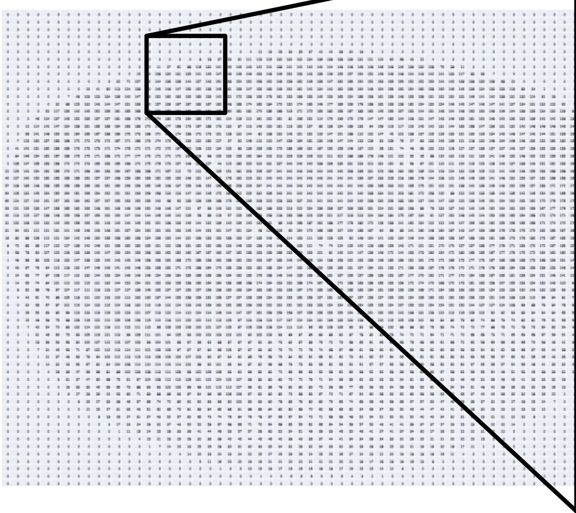
Computer Vision – Applications

- 3D reconstruction and modeling
- Recognition
- Motion capture
- Augmented reality:
- Video games and tele-operation
- Robot navigation and automotive
- Medical imaging



Computer Vision – Why is it hard?

- Achieving human-level visual perception is probably “AI-complete”



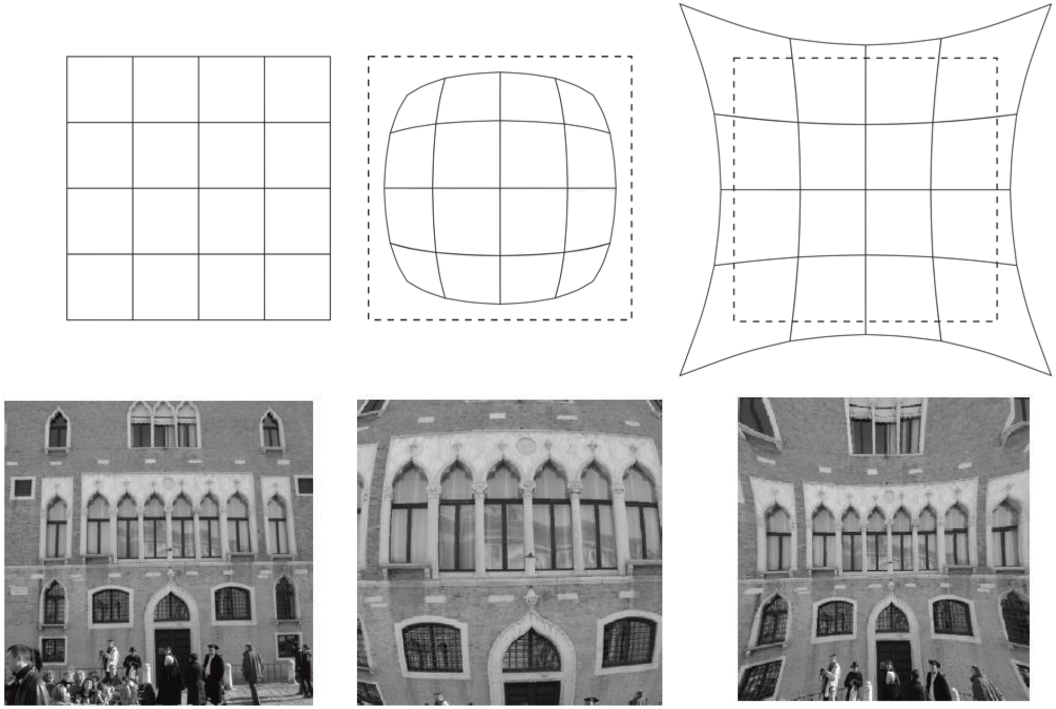
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	8	28
0	0	0	0	0	0	11	37	61	88	116	132
0	0	0	0	15	64	108	130	131	135	141	145
0	3	32	71	107	132	144	139	139	144	137	143
41	90	113	124	138	140	145	148	147	155	152	139
123	134	136	140	147	149	152	160	160	155	163	155
143	144	147	151	156	160	157	159	167	167	160	167
152	156	161	165	166	169	170	170	164	169	173	164
157	157	161	168	176	175	174	180	173	164	165	171
165	166	164	163	166	172	179	177	168	173	167	168
167	168	169	175	173	168	171	177	174	168	172	173

What a computer sees

What we see

Image Distortion

- The camera image is a projection of the three dimensional space into a two dimensional space,
- The projection process is affected by the characteristics of each camera



No distortion　　Barrel distortion　　Pincushion

Camera Calibration

- Camera calibration: calculating the camera's unique parameters
- Camera calibration is necessary if you are measuring distance from images acquired with a stereo camera or processing images for object detection
  - Need to know the information of the camera: lens characteristics, the gap between the lens and the image sensor, and the twisted angle of the image sensor, etc.

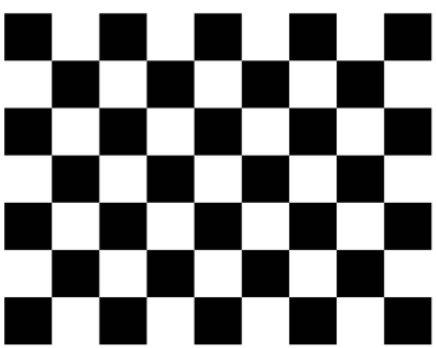


Figure 8-8 Chessboard for calibration (8 x 6)

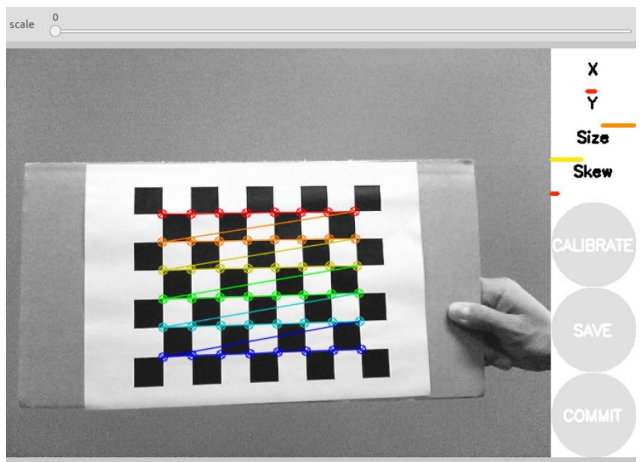


Figure 8-9 Calibration GUI initial state

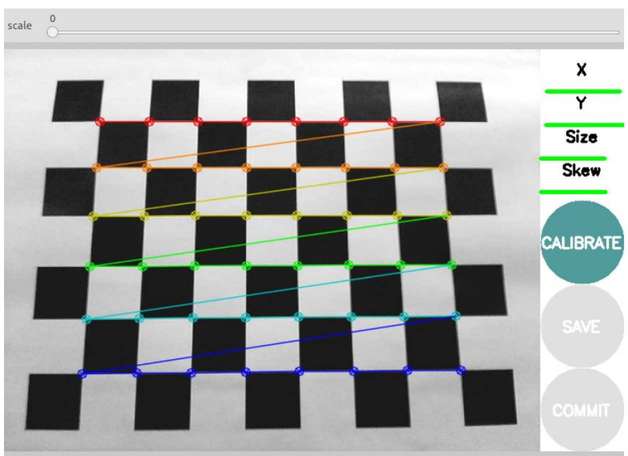


Figure 8-10 Calibration process using the calibration GUI

How do we measure distances with cameras?

- From a single image: we can only deduct the **ray** along which each image-point lies
- Stereo vision
  - using 2 cameras with known relative position T and orientation R, recover the 3D scene information

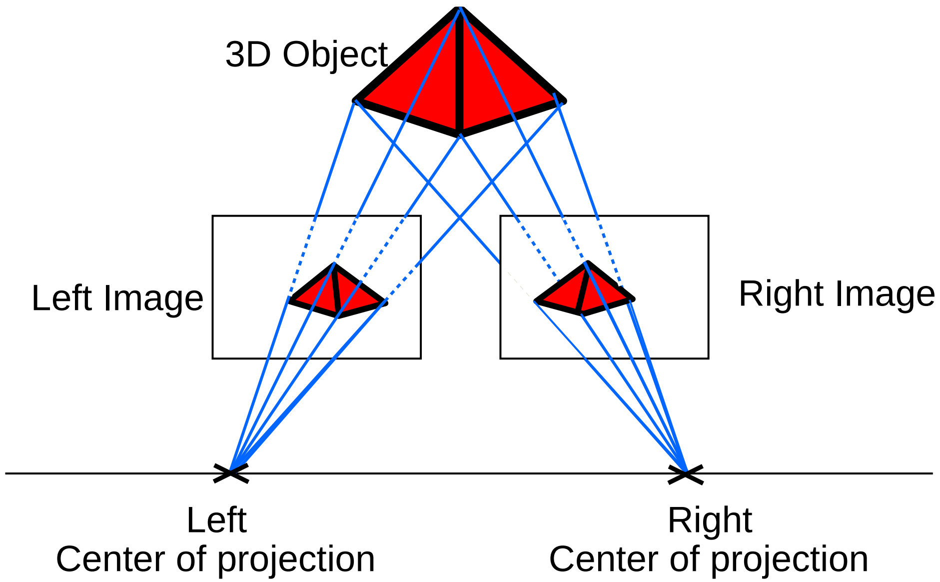
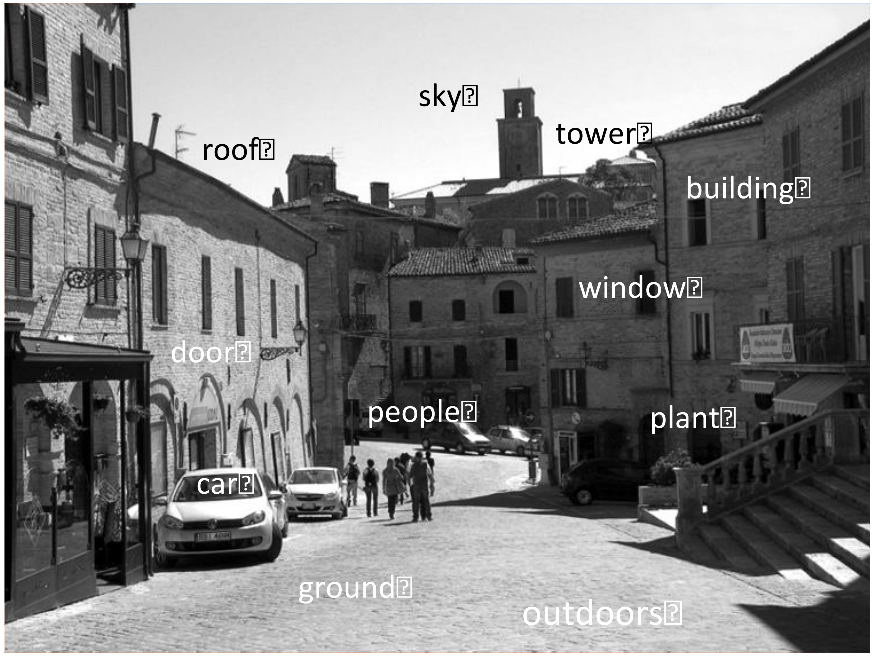


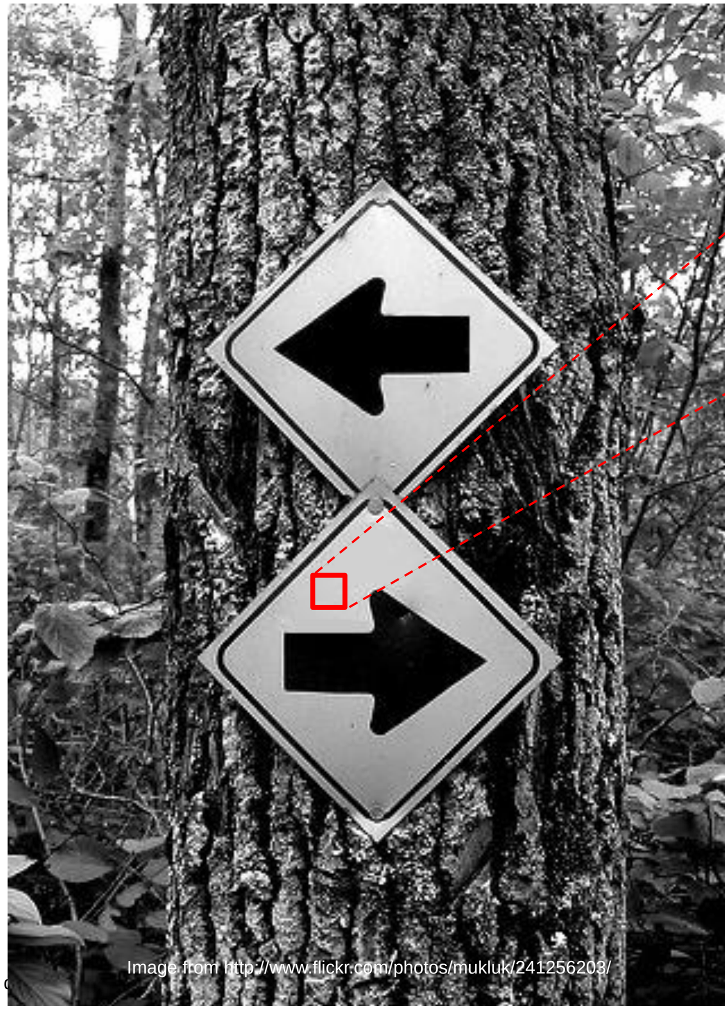
Image Processing

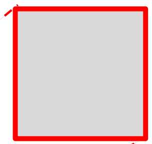
Image Intensities & Data Reduction

- Image capture a lot of information
- Monochrome image → matrix of intensity values
- Typical sizes:
  - 320 * 240 (QVGA)
  - 640 * 480 (VGA)
  - 1280 * 720 (HD)
- Intensity sampled to 256 grey levels – 8bits

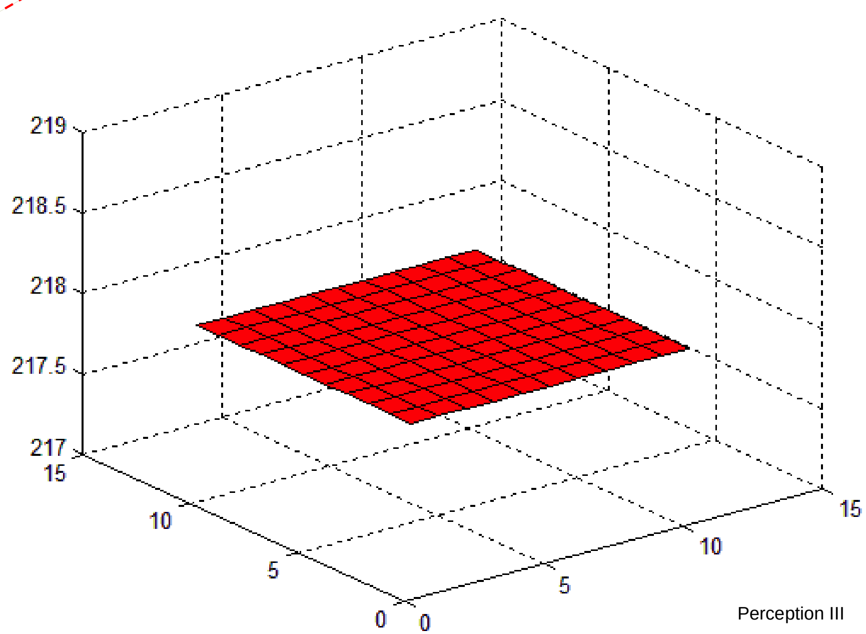


What is Useful, What is Redundant?



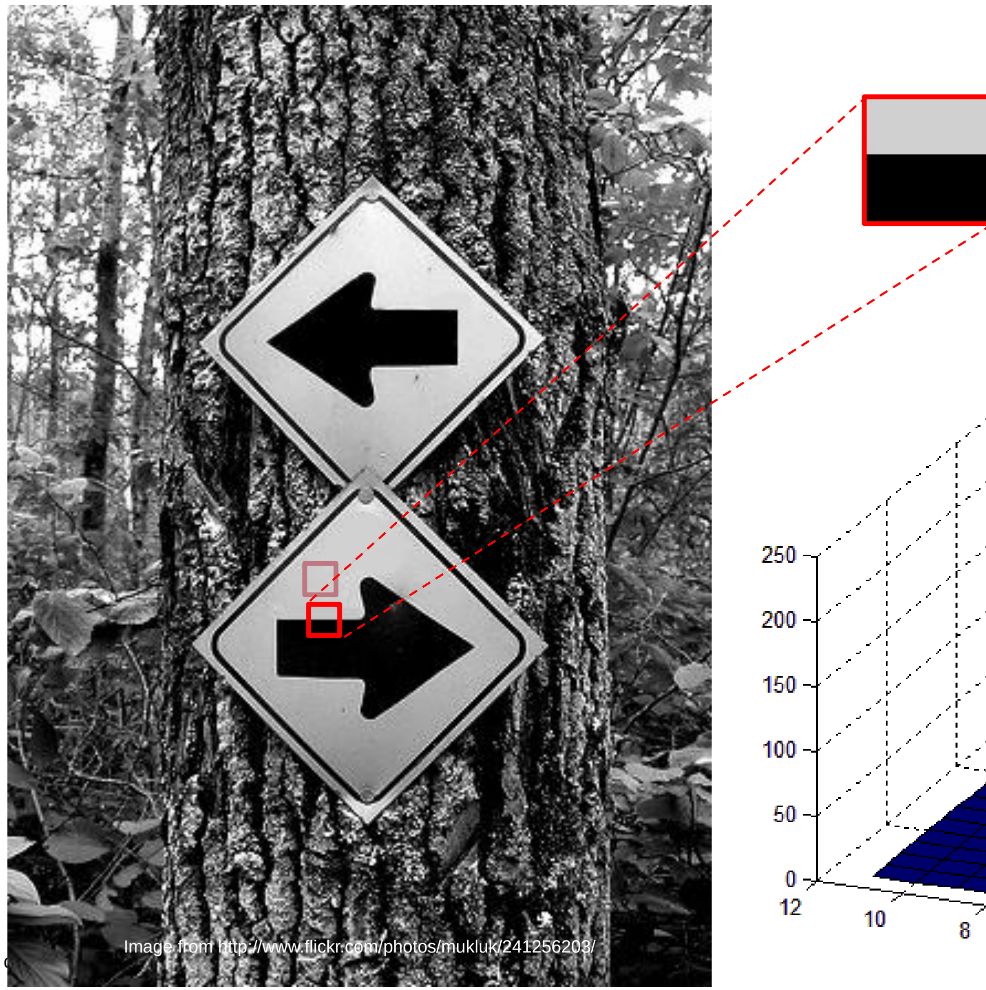


218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218
218	218	218	218	218	218	218	218	218	218	218

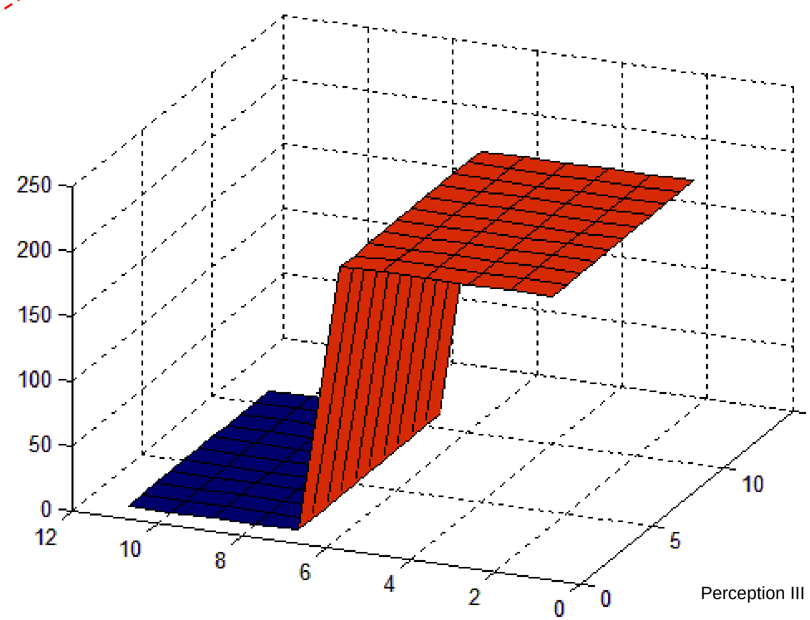


Perception III |

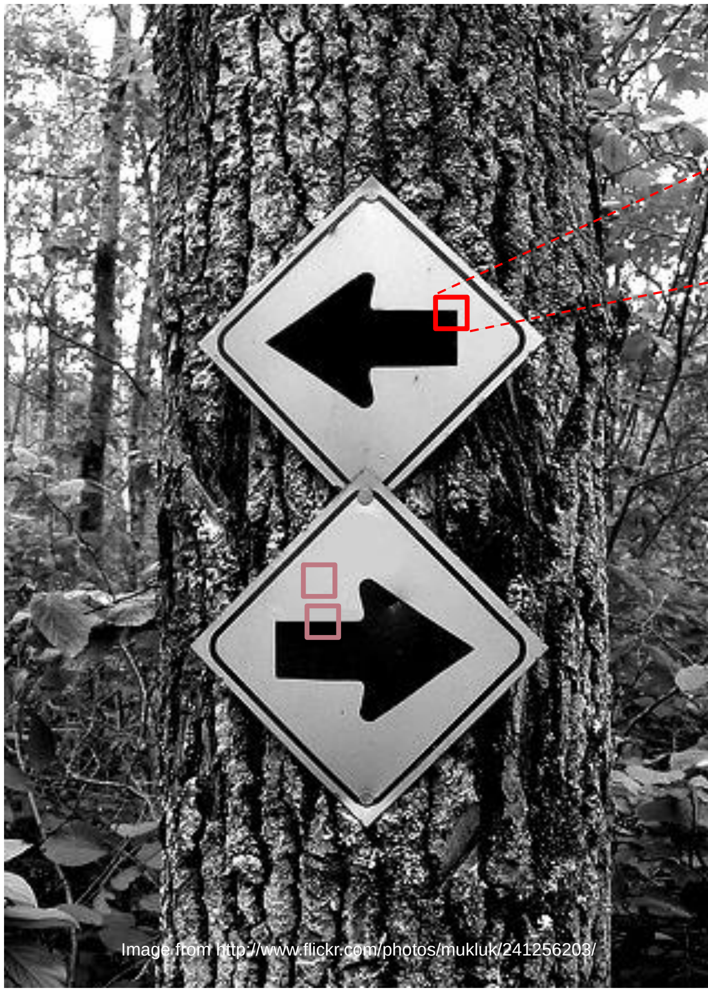
What is Useful, What is Redundant?

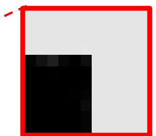


208	208	208	208	208	208	208	208	208	208	208
208	208	208	208	208	208	208	208	208	208	208
208	208	208	208	208	208	208	208	208	208	208
208	208	208	208	208	208	208	208	208	208	208
208	208	208	208	208	208	208	208	208	208	208
208	207	208	208	208	208	208	208	208	208	208
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0

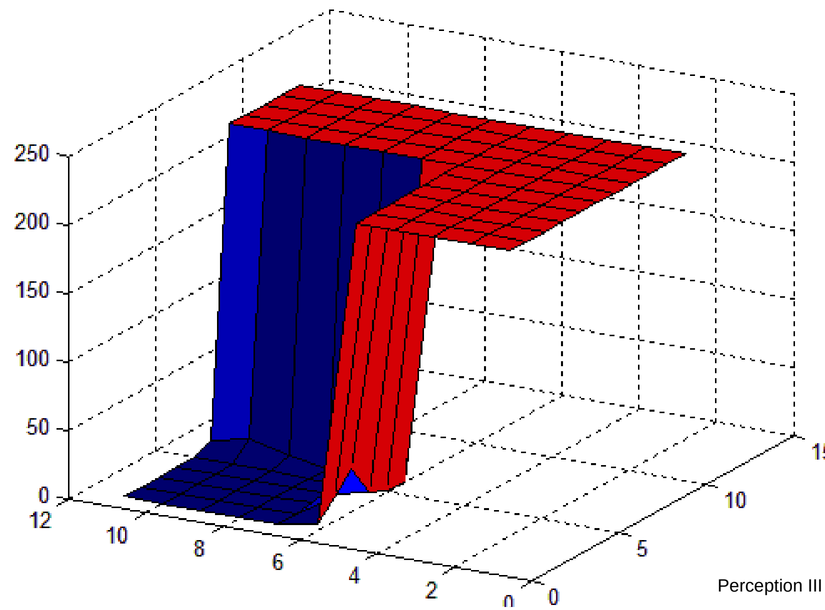


What is Useful, What is Redundant?





229	229	229	229	229	229	229	229	229	229	229
229	229	229	229	229	229	229	229	229	229	229
229	229	229	229	229	229	229	229	229	229	229
229	229	229	229	229	229	229	229	229	229	229
229	229	229	229	229	230	229	229	229	229	229
5	17	31	7	1	0	229	229	229	229	229
0	0	1	0	0	0	229	229	229	229	229
0	0	0	0	0	0	229	229	229	229	229
0	0	0	0	1	4	229	229	229	229	229
0	0	0	0	0	11	229	229	229	229	229
0	0	0	0	0	5	229	229	229	229	229



Perception III

Image Filtering

- **Filtering:** accept / reject certain components
- Example: a low-pass filter allows low frequencies a blurring (smoothing) effect on an image – used to reduce image noise
- Smoothing can be achieved not only with **frequency filters**, but also with **spatial filters**.

Low-pass filtering:
retains low-frequency components (smoothing)

High-pass filtering:
retains high-frequency components (edge detection)

Image Filtering – Spatial Filters

- S_{xy}: neighborhood of pixels around the point *(x,y)* in an image
- Spatial filtering operates on S_{xy} to generate a new value for the corresponding pixel at output image *J*

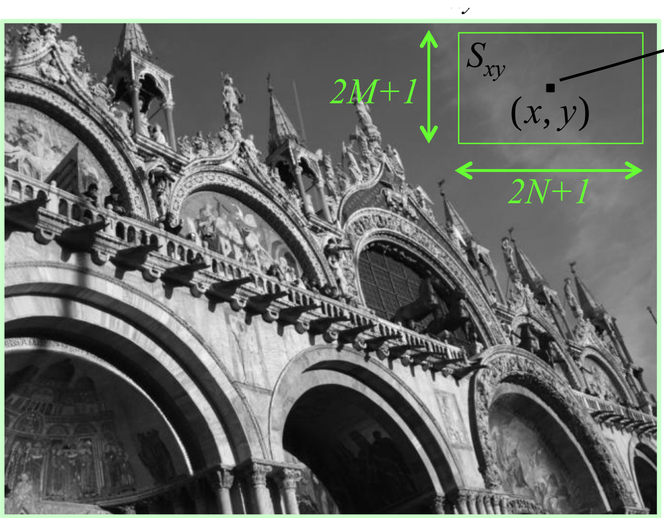
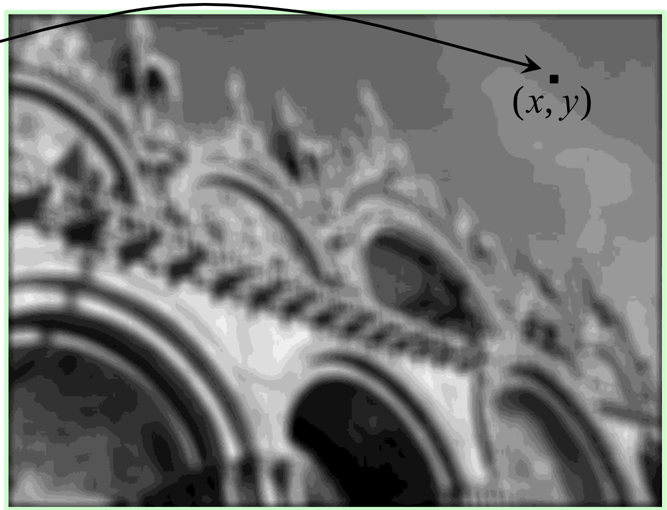


Image I



Filtered Image $J = F(I)$

For example, an averaging filter is: $J(x,y) = \dfrac{\sum_{u,v \in S_{xy}} I(u,v)}{(2M+1)(2N+1)}$

Image Filtering – Linear, shift-invariant filters

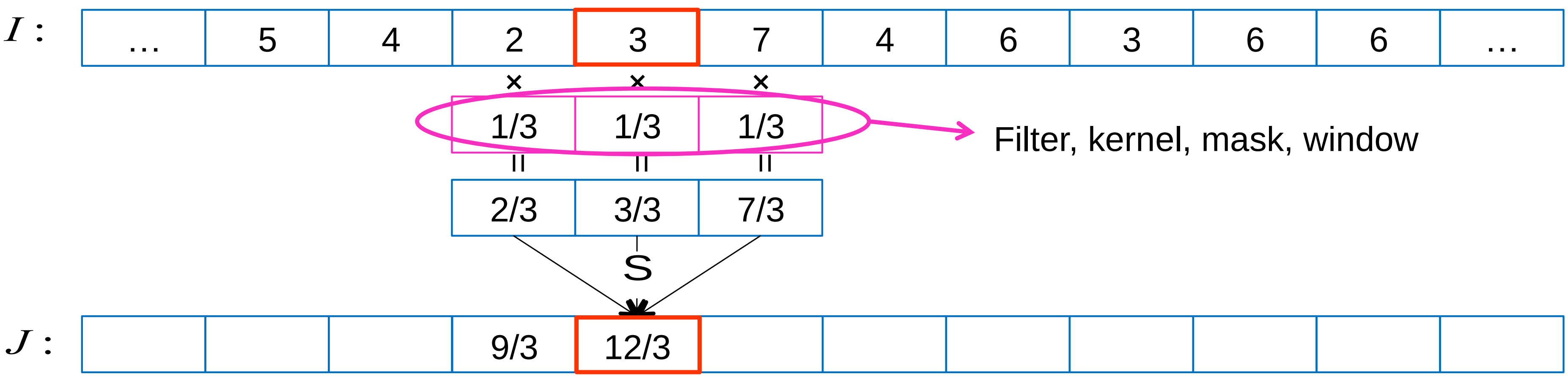
- Linear: every pixel is replaced by a linear combination of its neighbors

- Shift-invariant: the same operation is performed on every point on the image

- Why filter?
  - noise reduction, image enhancement, feature extraction, …

Image Filtering – Correlation

- An averaging filter in 1D



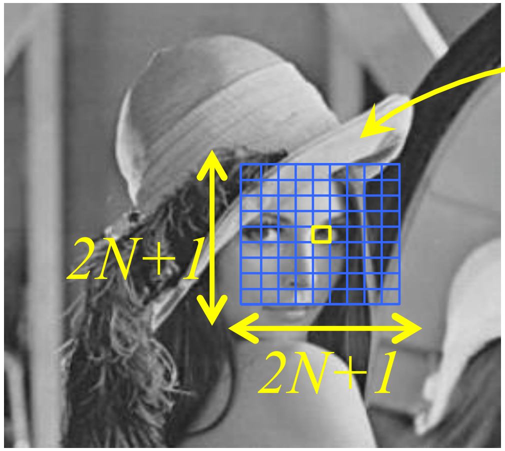
- Formally, Correlation is $J(x) = F \cdot I(x) = \sum_{i \in [-N,N]} F(i) I(x+i)$

- In this smoothing example $F(i) = \begin{cases} 1/3, i \in [-1,1] \\ 0, i \notin [-1,1] \end{cases}$

Image Filtering – Correlation in 2D

- Example: constant averaging filter

$$F = \begin{bmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix}$$



- If $size(F) = (2N + 1)^2$, i.e., a square filter
  - # of multiplications per pixel = $(2N + 1)^2$
  - # of additions per pixel = $(2N + 1)^2 - 1$

Image Filtering – Matching Using Correlation

- Find locations in an image that are similar to a template
- Filter = template

3	8	3

→ test it against all image locations

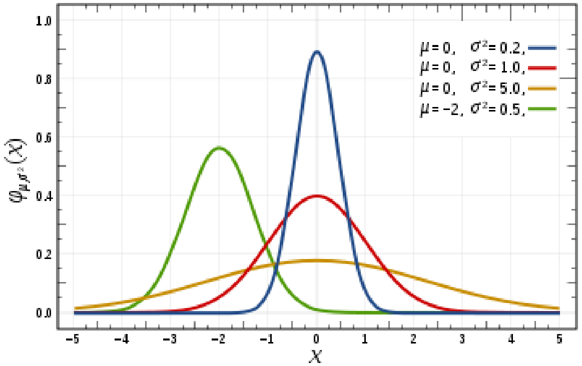
I :	3	2	4	1	3	8	4	0	3	8	7	7
J :	26	37	21	50	54	1	50	65	59	16	42	17

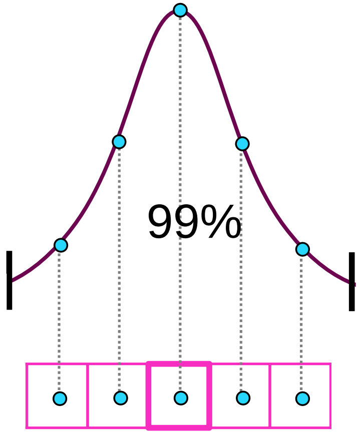
- Similarity measure: Sum of Squared Differences (SSD) minimizes

$$\sum_{i=-N}^{N} \left(F(i) - I(x+i)\right)^2$$

Image Filtering – Gaussian Filter

- Common practice for image smoothing: use a Gaussian





$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-u)^2}{2\sigma^2}}$$

$\mu = 0$

σ: controls the amount of smoothing

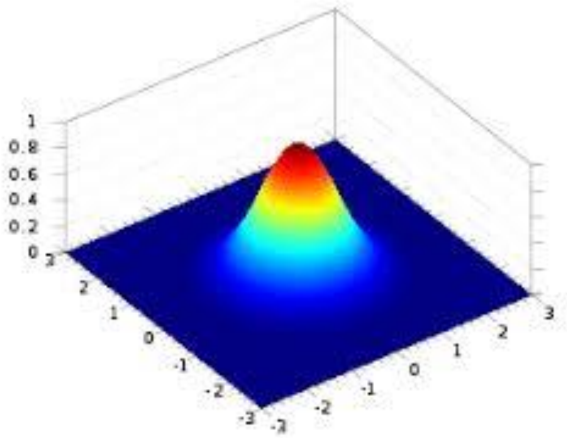
Normalize filter so that values always add up to 1

- Near-by pixels have a bigger influence on the averaged value rather than more distant ones

Image Filtering – 2D Gaussian Smoothing

- A general, 2D Gaussian

$$G(x,y) = \frac{1}{2\pi|S|^{1/2}} e^{-\frac{1}{2}\binom{x}{y} S^{-1}(x,y)}$$



- We usually want to smooth by the same amount in both *x* and *y* directions

$$S = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

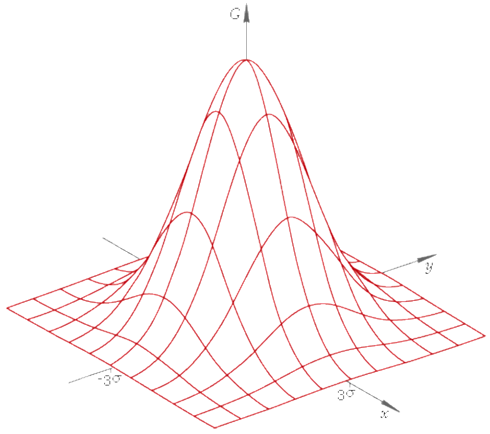
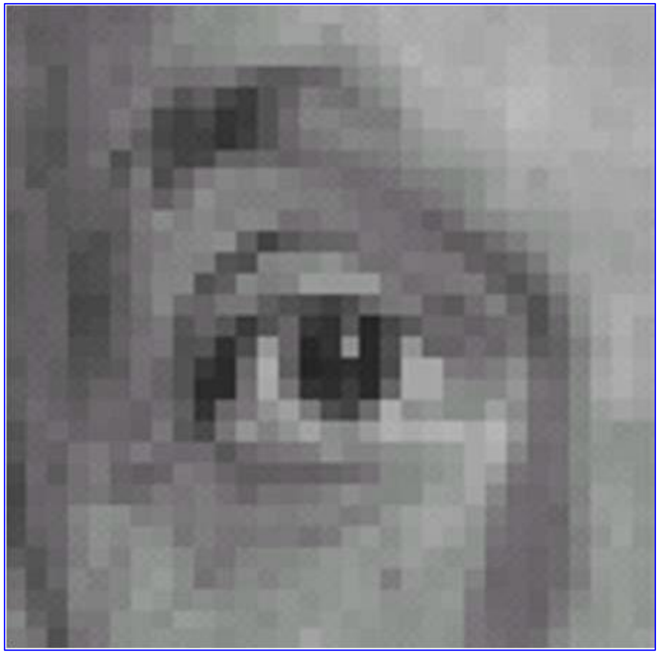


Image Filtering – Examples


original image

$*$

0	0	0
0	1	0
0	0	0

=

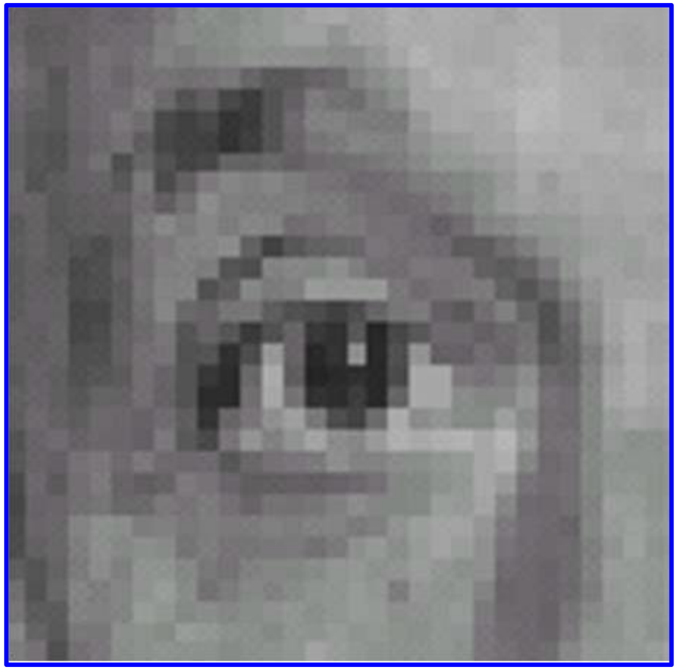
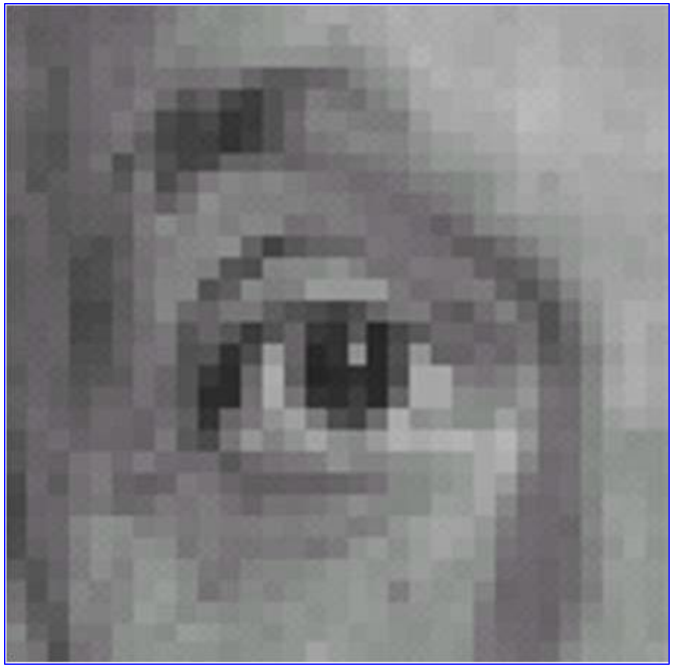


Image Filtering – Examples

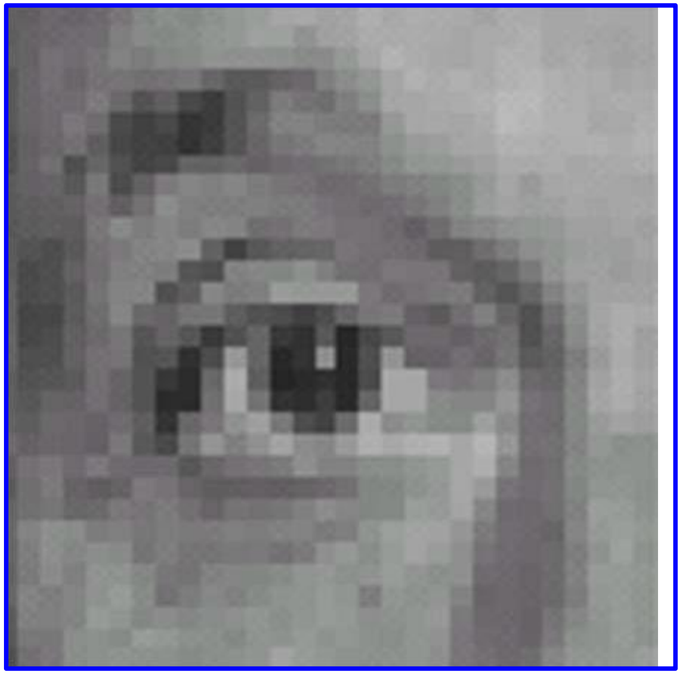


original image

*

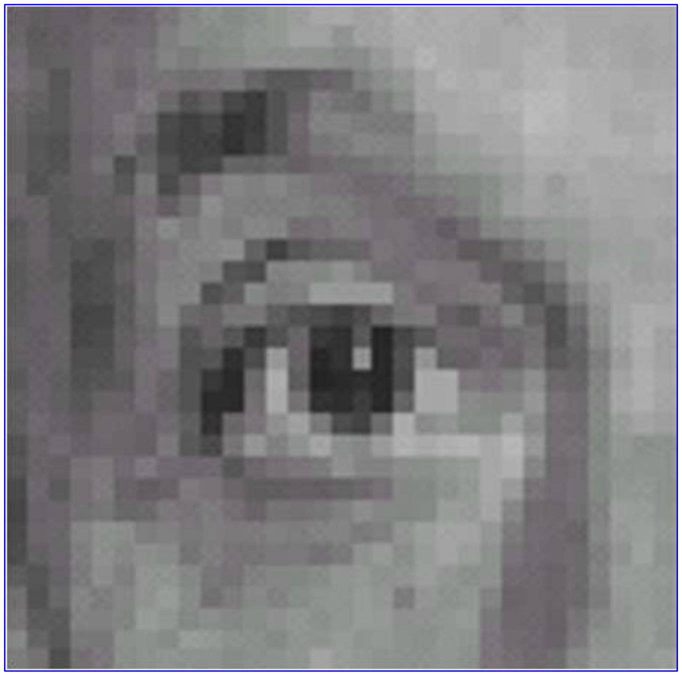
0	0	0
0	0	1
0	0	0

=



filtered (shifted left by 1 pixel)

Image Filtering – Examples

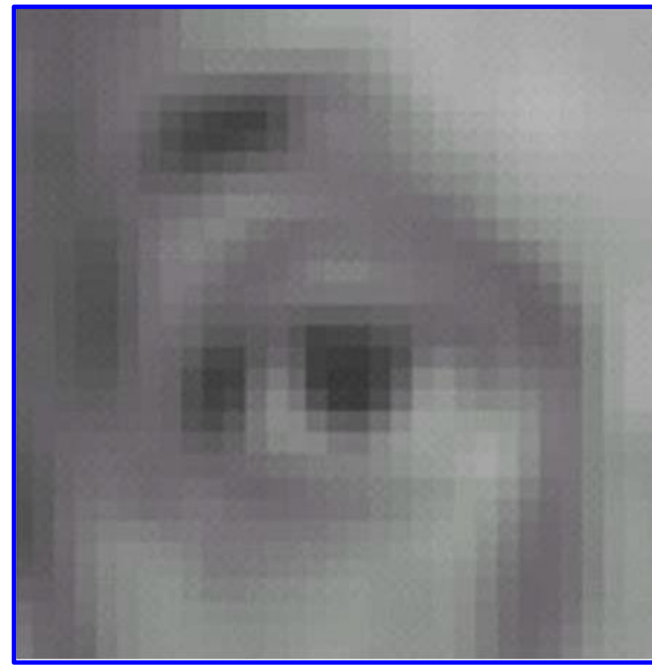


original image

*

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

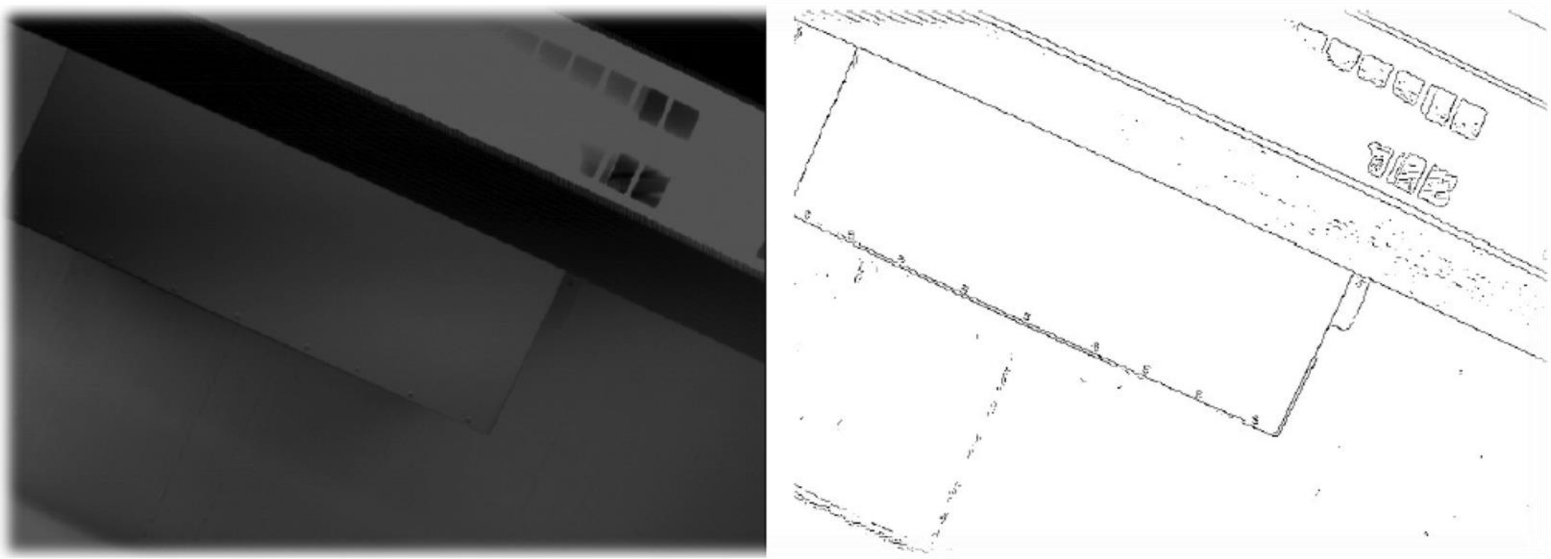
=



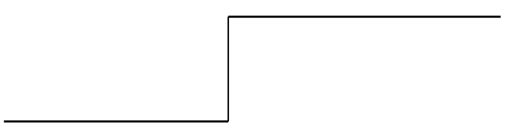
filtered (blurred with a box filter)

Image Filtering – Edge Detection

- Ultimate goal of edge detection: an idealized line drawing
- Edge contours in the image correspond to important scene contours



- Edges correspond to sharp changes of intensity



- How to detect an edge?
  - Big intensity change → magnitude of derivative is large

Image Filtering – Edge Detection

- Examples of edge detection filters

  - Prewitt:

$$F_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \qquad F_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

  - Sobel:

$$F_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad F_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

  - Roberts:

$$F_x = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad F_y = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Image Filtering – Edge Detection

I : original image (Lena)

- Lidar-camera filtering:
  - https://www.mathworks.com/help/lidar/ug/lidar-camera-calibration.html

- Camera calibration using AprilTag markers
  - https://www.mathworks.com/help/vision/ug/camera-calibration-using-apriltag-markers.html

- Thank you!