LECTURER: TAI LE QUY

ARTIFICIAL INTELLIGENCE

TOPIC OUTLINE

History of AI 1

Modern AI Systems 2

Reinforcement Learning 3

Natural Language Processing – Part 1 4

Natural Language Processing – Part 2 5

Computer Vision 6

UNIT 6

COMPUTER VISION

STUDY GOALS

On completion of this unit, you will be able to …

… define computer vision.

… explain how to represent images as pixels.

… distinguish between detection, description, and matching of features.

… correct distortion with calibration methods.

EXPLAIN SIMPLY

1. What are the typical computer vision tasks?

2. How are filters used in computer vision?

3. How does camera calibration work?

HISTORICAL DEVELOPMENTS

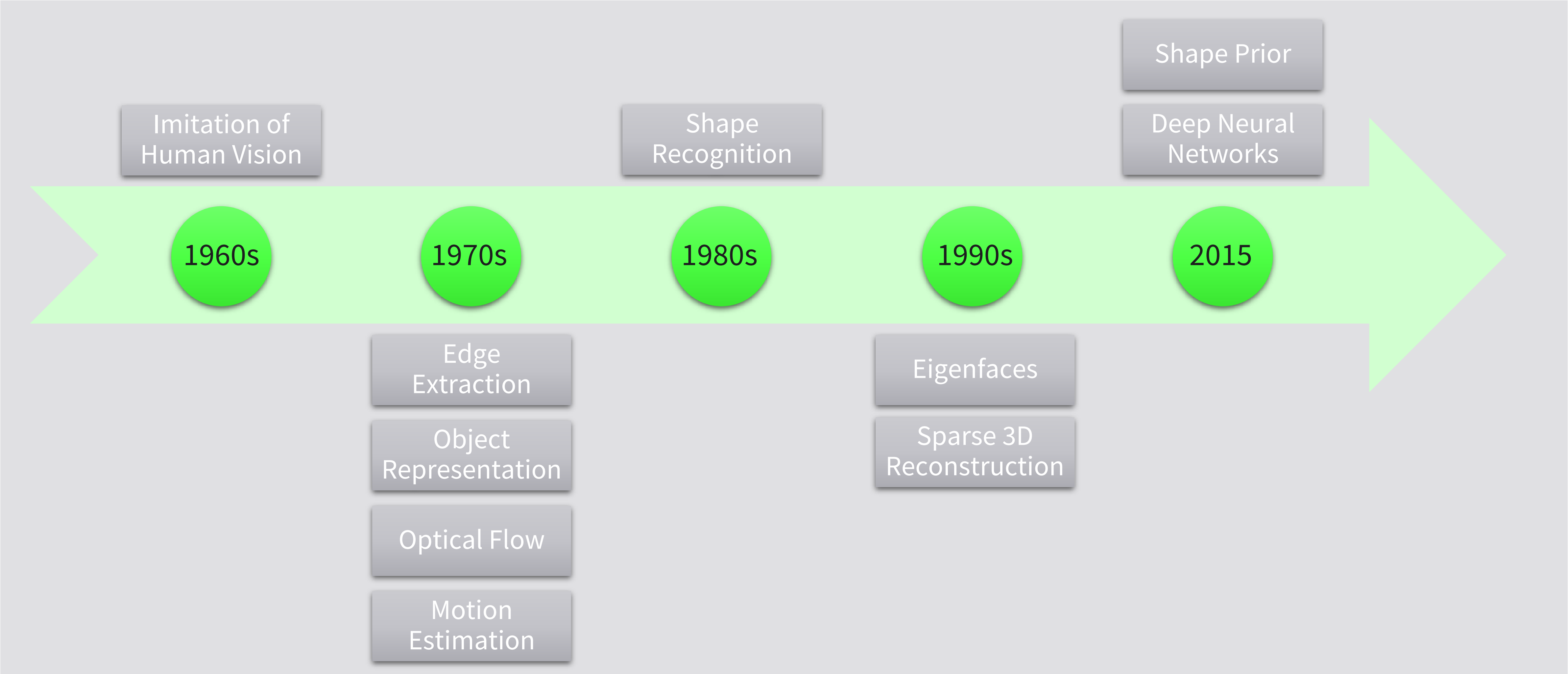


Image Source: Custom Depiction

TYPICAL COMPUTER VISION TASKS

Computer vision tasks

Recognition	Motion analysis	Image restoration	Geometry reconstruction
Detect and classify different objects or persons in images	Detect and track motion of objects, persons or the camera	Filter and remove noise from images and videos	Estimation of a 3D model of a real-world object or scene

Image Source:

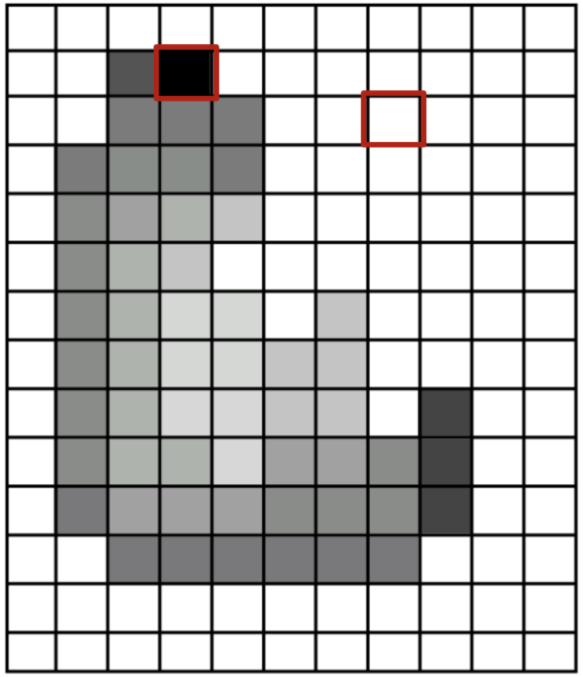
RECOGNITION TASKS

- Detection of objects, persons, poses, images
- Estimation of different classes of objects contained in images
- Person identification: fingerprint, face, handwriting, etc.
- Estimation the orientation, position of objects relative to the camera
- Optical character recognition (OCR)

CHALLENGES IN COMPUTER VISION

- The illumination of an object is very important
- Differentiating similar objects can also be difficult in recognition tasks
- The size and aspect ratios of objects in images or videos
- Algorithms must be able to deal with rotation of an object
- The location of objects can vary

- PIXEL: pix (pictures) + el (element)



=

255	255	255	255	255	255	255	255	255	255	255
255	255	20	0	255	255	255	255	255	255	255
255	255	75	75	255	255	255	255	255	255	255
255	75	95	95	75	255	255	255	255	255	255
255	96	127	145	175	255	255	255	255	255	255
255	127	145	175	175	175	255	255	255	255	255
255	127	145	200	200	175	175	95	255	255	255
255	127	145	200	200	175	175	95	47	255	255
255	127	145	145	175	127	127	95	47	255	255
255	74	127	127	127	95	95	95	47	255	255
255	255	74	74	74	74	74	74	255	255	255
255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255

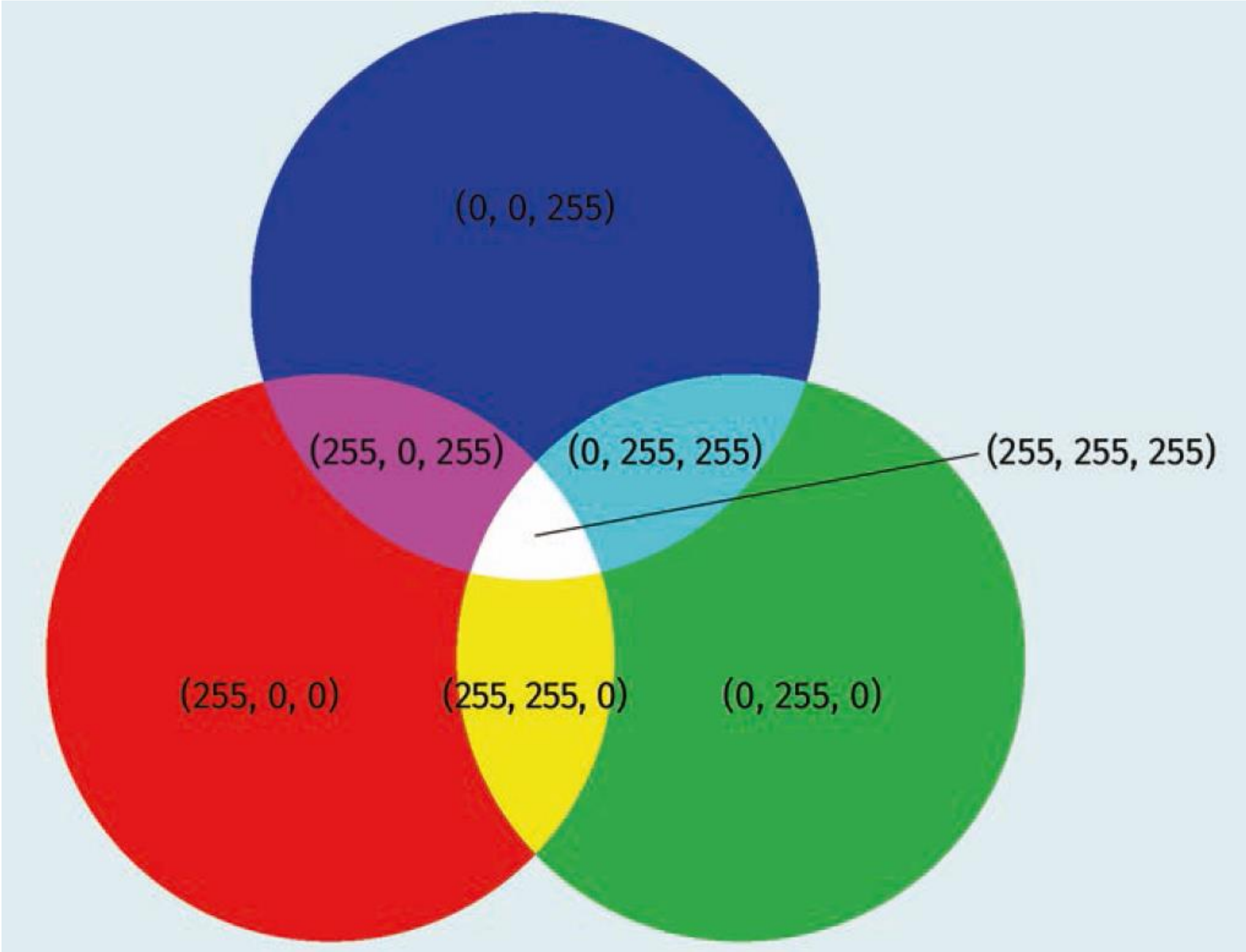
0 = black; 255 = white

Source of image: https://edtech.engineering.utoronto.ca/files/2d-image-digital-representation

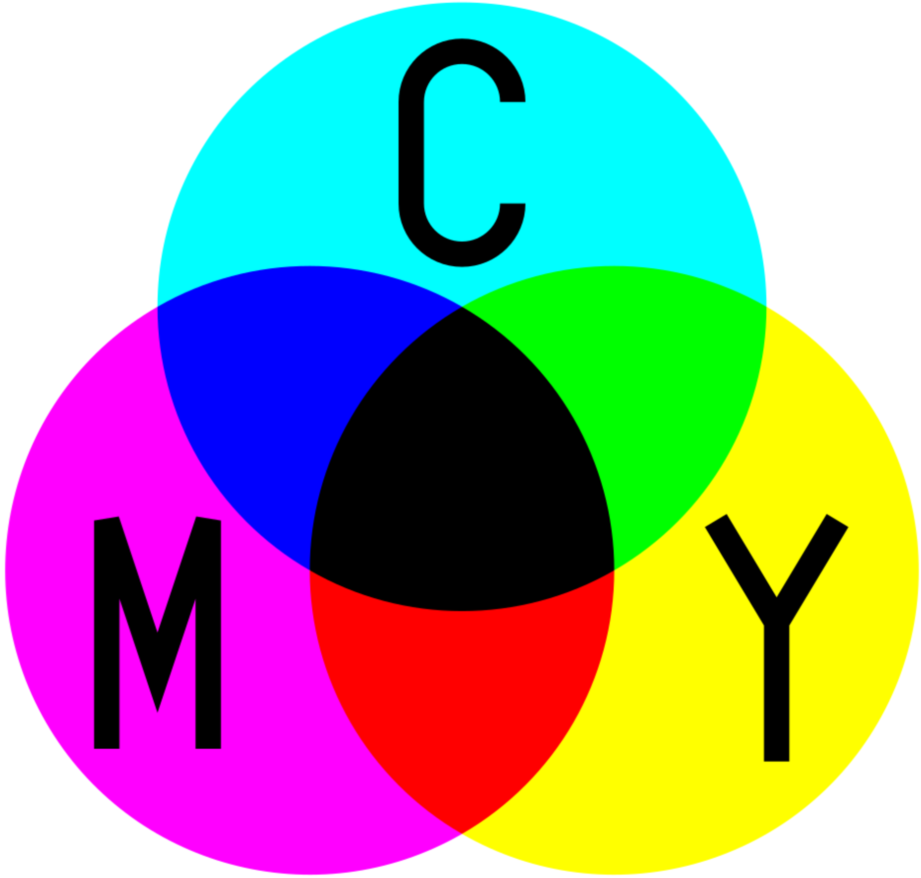
Color Representations in Images		
Name	Color representation	Color depth
Monochrome	1 bit	2 colors
	8 bit	$2^8 = 256$ grayscale intensity levels or colors
Real color	15 bit	$2^{15} = 32.768$ colors
High color	16 bit	$2^{16} = 65.536$ colors
True color	24 bit	$2^{24} = 16.777.216$colors
Deep color	30 – 48 bit	$2^{30} - 2^{48}$colors

COLOR REPRESENTATION

RGB

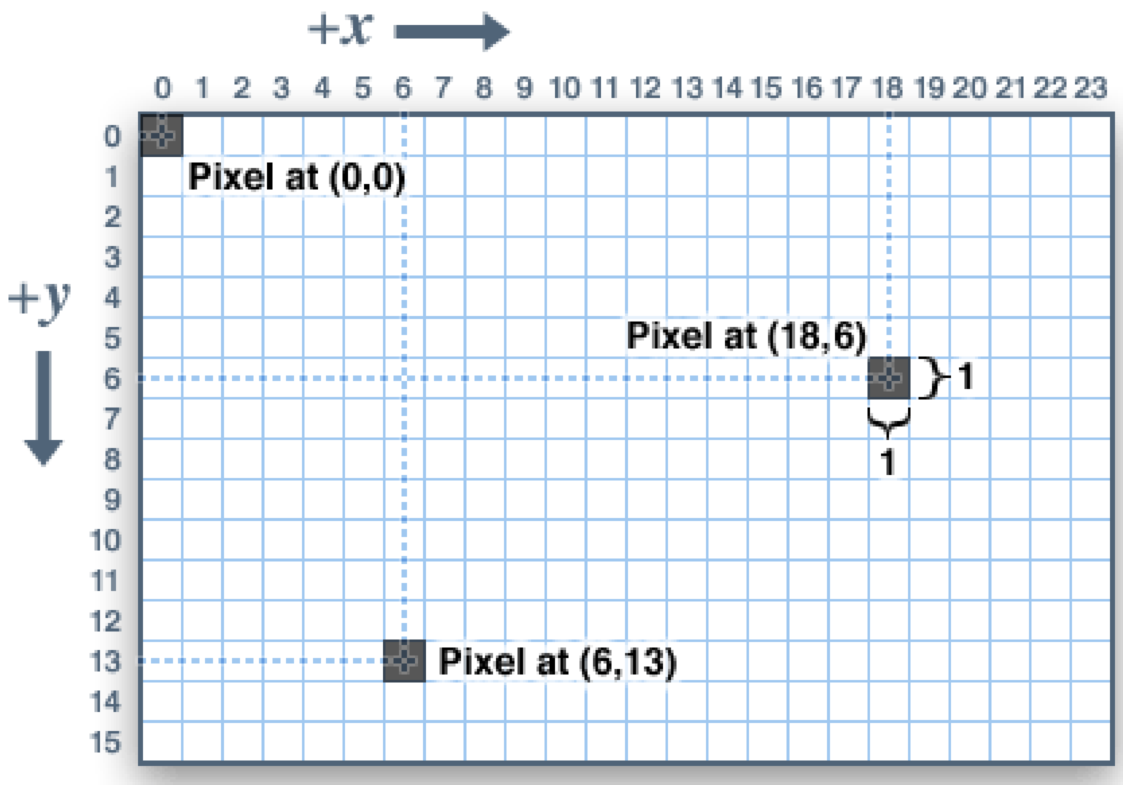


CMYK



Source of iamges: coursebook, wikipedia

- A function that can map a two-dimensional coordinate (x,y) to a specific color value
- Example:
  - f(x, y) for an 8-bit grey scale image
  - f(42, 100) = 0, i.e., black pixel



Source of image: https://learn.adafruit.com/assets/1264

Filter:

Function which receives an image as an input, applies modifications, and returns the filtered image as an output.

Example: 2D image convolution

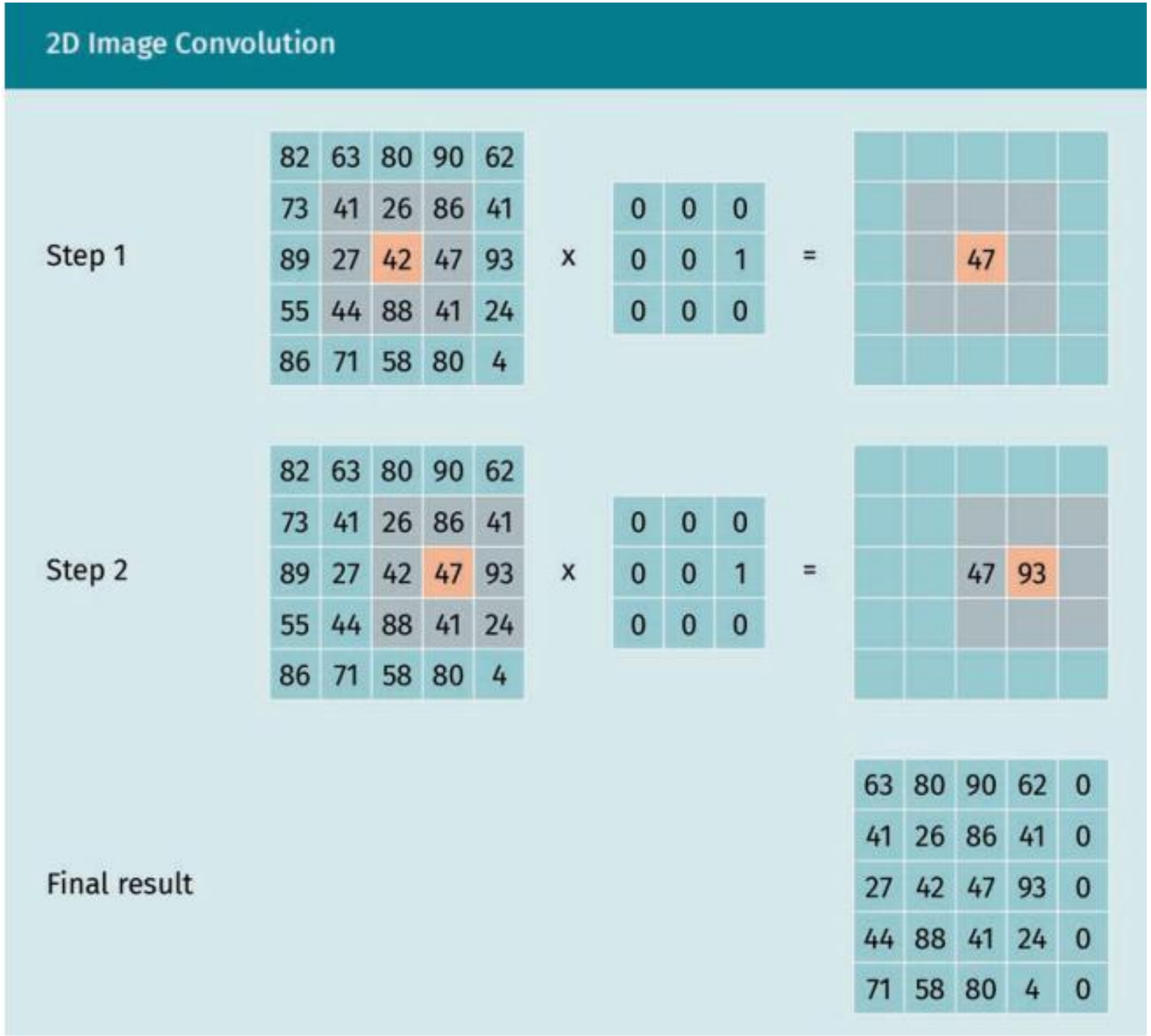
Convolution of an image I with a kernel k with a size of n and a center coordinate a:

$$\mathrm{I} \cdot (x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{I}(x - i + a, y - j + a) k(i, j)$$

$\mathrm{I} \cdot (x, y)$: value of the resulting image $\mathrm{I} \cdot$ at position (x, y), I is the original image.

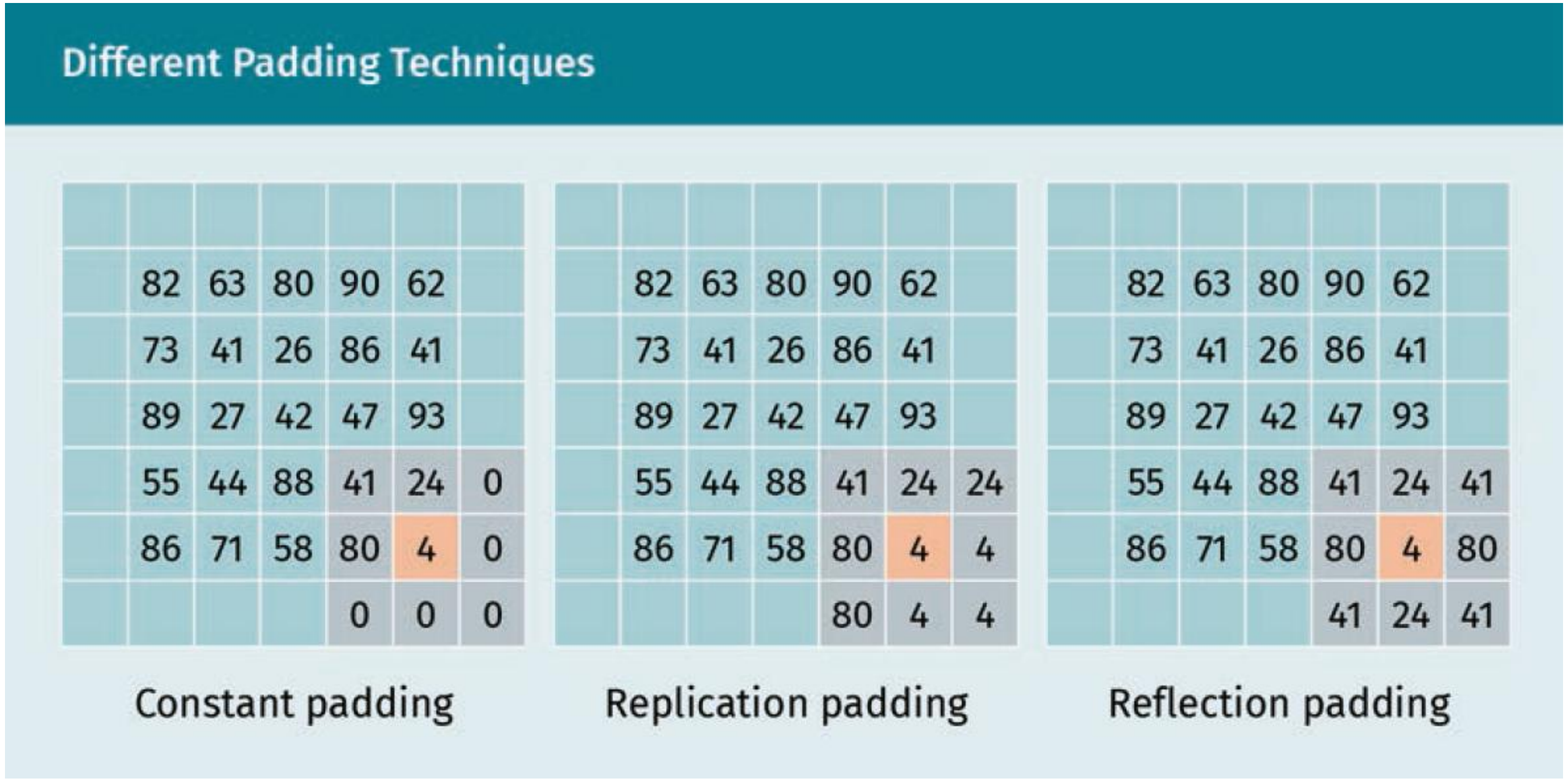
Example:

- n = 3
- a = 2



Source of image: Course book

PADDING TECHNIQUES



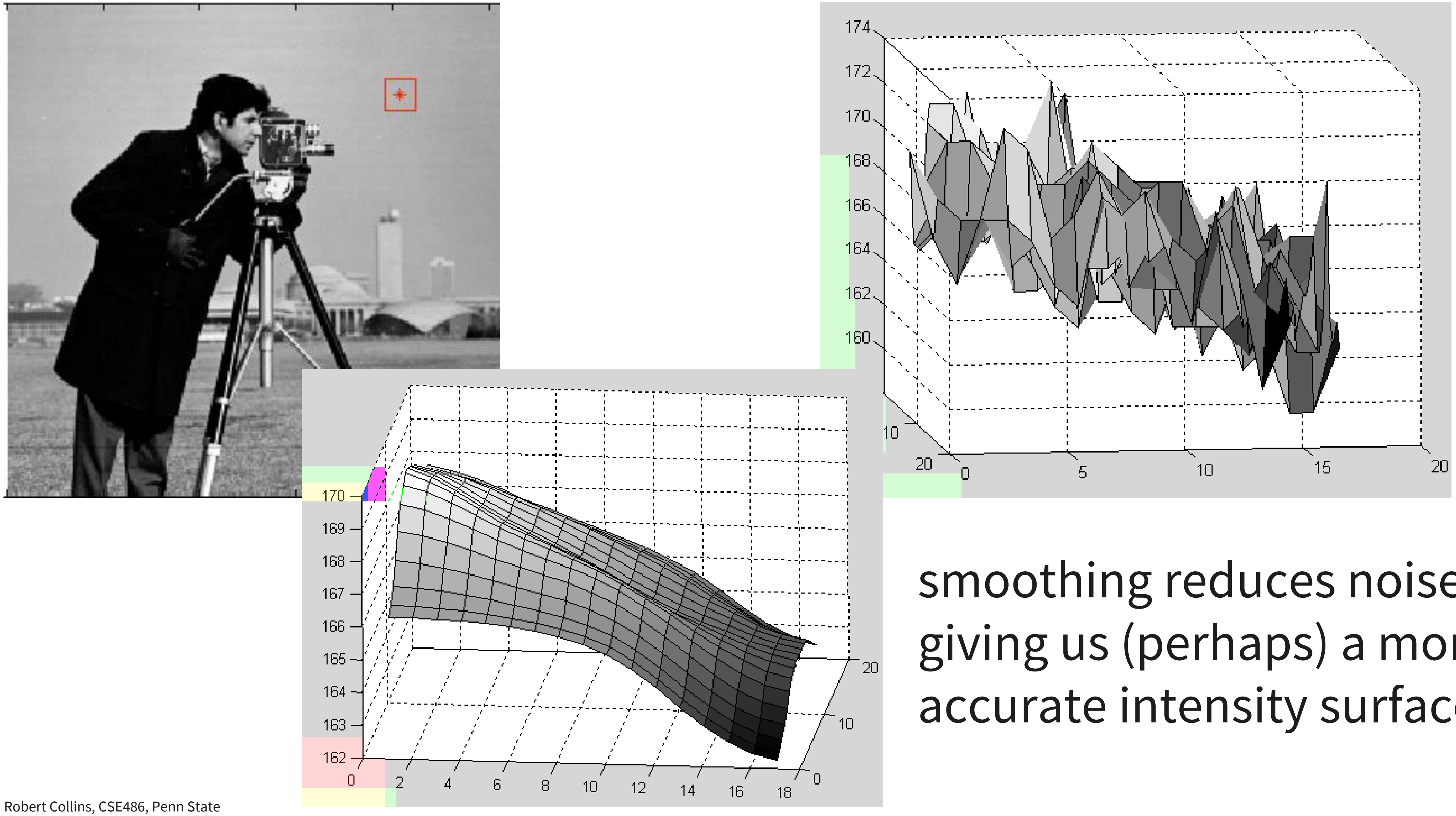
From Numerical Recipes in C:

The premise of data smoothing is that one is measuring a variable that is both slowly varying and also corrupted by random noise. Then it can sometimes be useful to replace each data point by some kind of local average of surrounding data points. Since nearby points measure very nearly the same underlying value, averaging can reduce the level of noise without (much) biasing the value obtained.

(William H. Press et al., 1992)

Source of the text: https://www.cec.uchile.cl/cinetica/pcordero/MC_libros/NumericalRecipesinC.pdf

SMOOTHING REDUCES NOISE



smoothing reduces noise, giving us (perhaps) a more accurate intensity surface.

Source: Robert Collins, CSE486, Penn State

AVERAGING / BOX FILTER

- Mask with positive entries that sum to 1.
- Replaces each pixel with an average of its neighborhood.
- Since all weights are equal, it is called a BOX filter.
- Variance of noise in the average is smaller than variance of the pixel noise

Box filter

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

Source: Robert Collins, CSE486, Penn State

Original

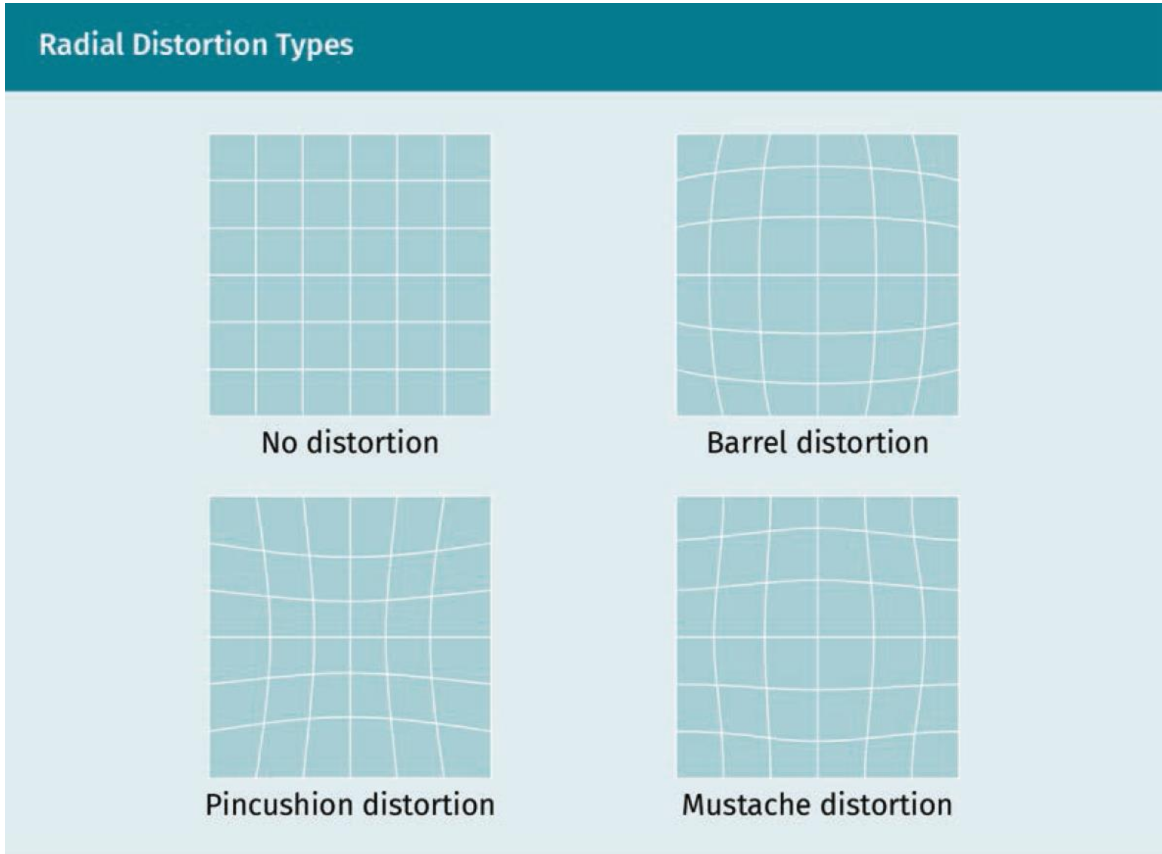
Convolved with 11x11 box filter

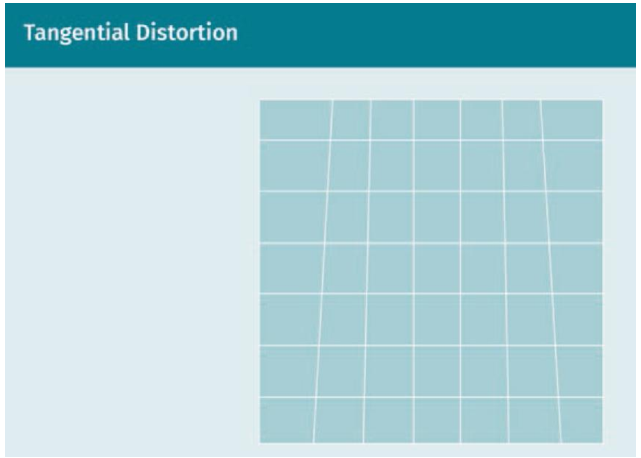
Drawback: smoothing reduces fine image detail

Source: Robert Collins, CSE486, Penn State

DISTORTION

- Assumption: linear projection of a scene
- 2 kinds of distortions:
  - Radical distortion: lines bend towards the edge of the camera lens
  - Tangential distortion: caused if the camera lens are not properly aligned
- Address distortion
  - Mathematical model: Brown Conrady model (Brown, 1966)



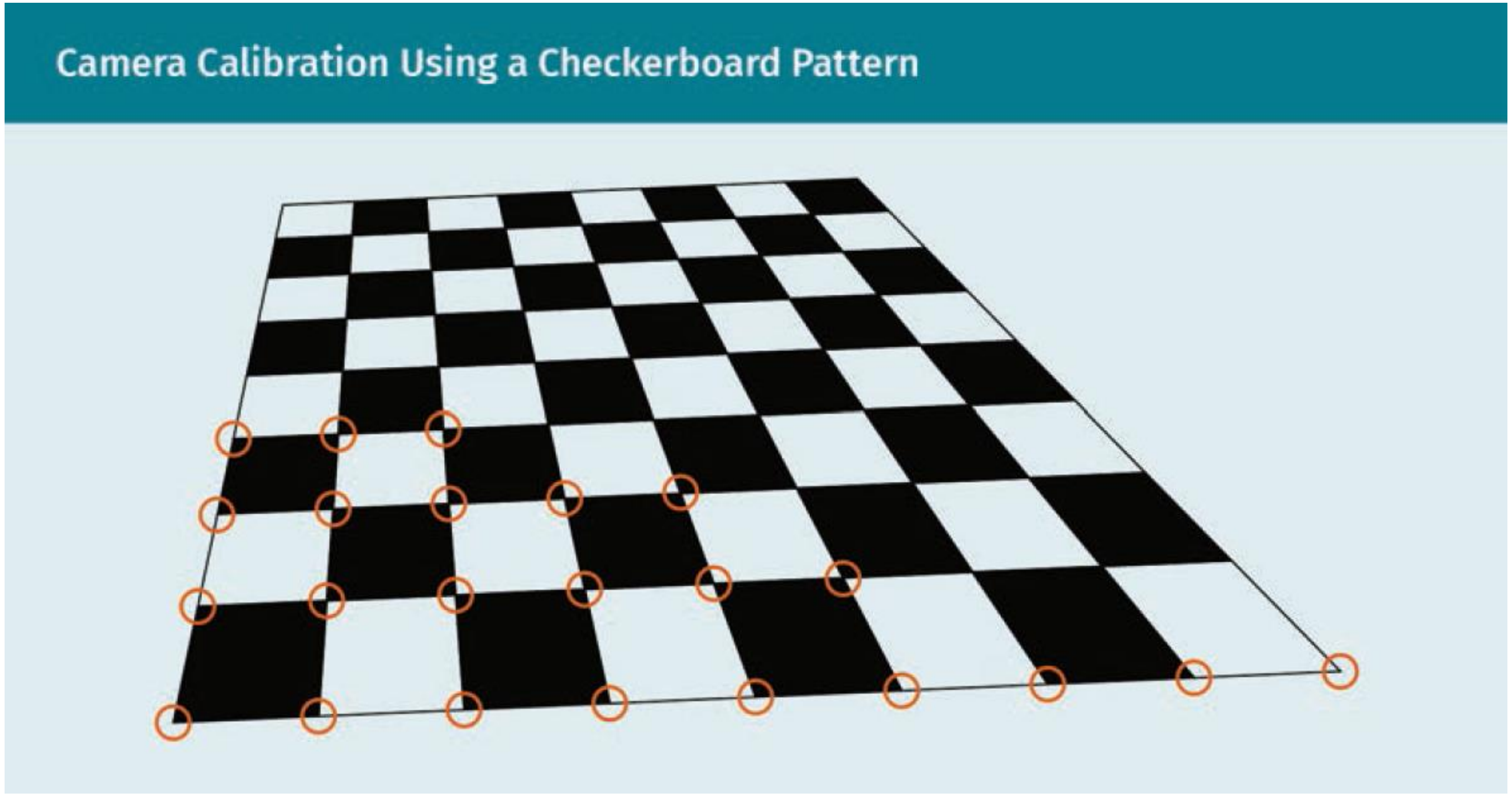


Source of images: course book

- Estimates the extrinsic and intrinsic parameters of a camera
  - Extrinsic characteristics: orientation, rotation, position of the camera
  - Intrinsic characteristics: optical center, focal length, etc.
- Reconstruct a 3D model of the underlying scenario from the real world

Image Source: Custom Depiction

- Use two or more images as an input, and the size of the object
- For example: checkerboard



Source of image: course book

- The calibration process
  - Select at least two sample images, such as checkerboard pattern
  - Identify distinctive points of each images, e.g., corners of the individual squares
  - Localization of the corners of the squares, we know:
    - The size of the checkerboard
    - The 2D coordinates of the corners in the camera's taken image

→ Calculate the camera matrix and the distortion coefficients, then apply the Brown-Conrady model.

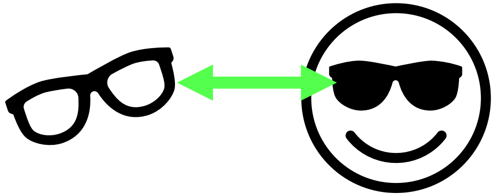
Feature Detection

- Blobs (blue)
- Edges (red)
- Corners (yellow)

Feature Description

- Common algorithms:
  - BRIEF
  - ORB
  - SIFT
  - SURF

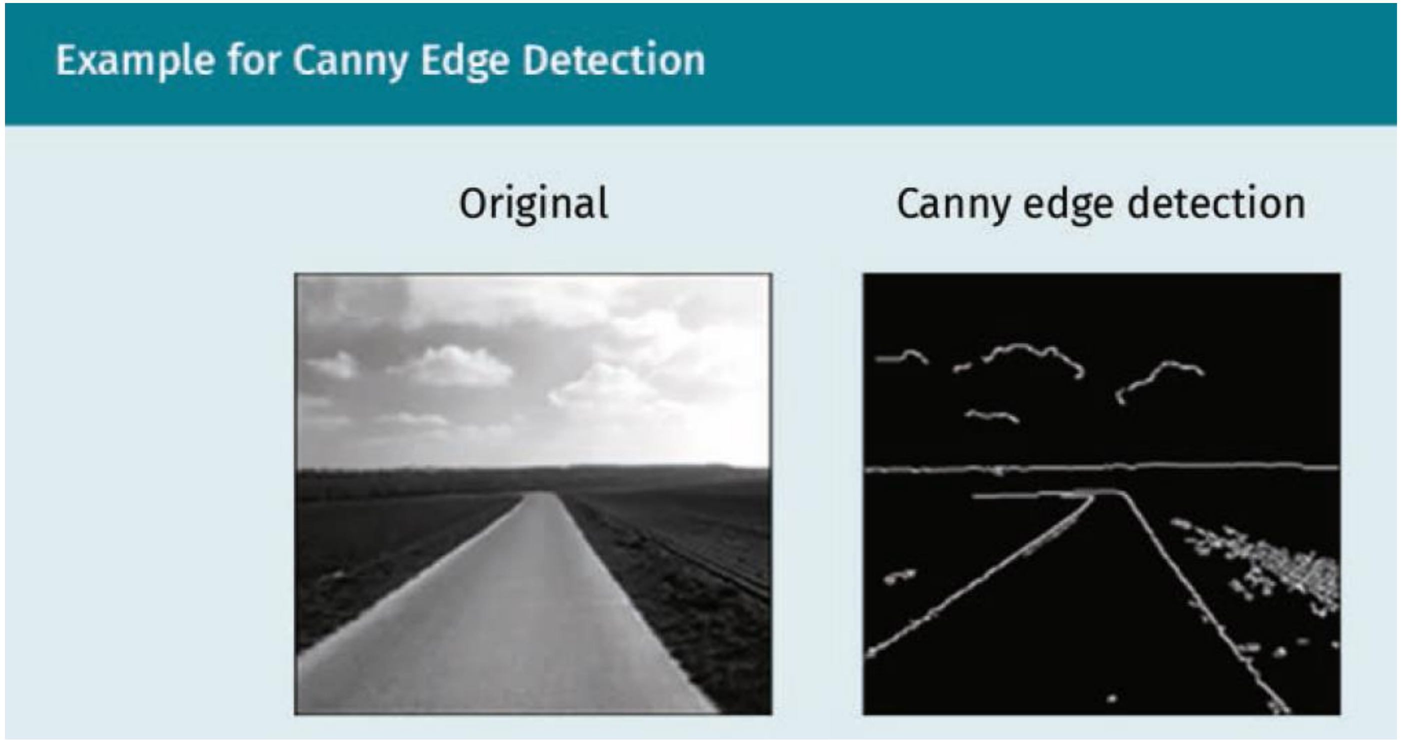
Feature Matching



- Identification of similar features in different images

Image Source: Custom Depiction

- Edge detection
  - Canny edge detection: analyzing the change between pixel values, works only on single color images (grey scaled images)

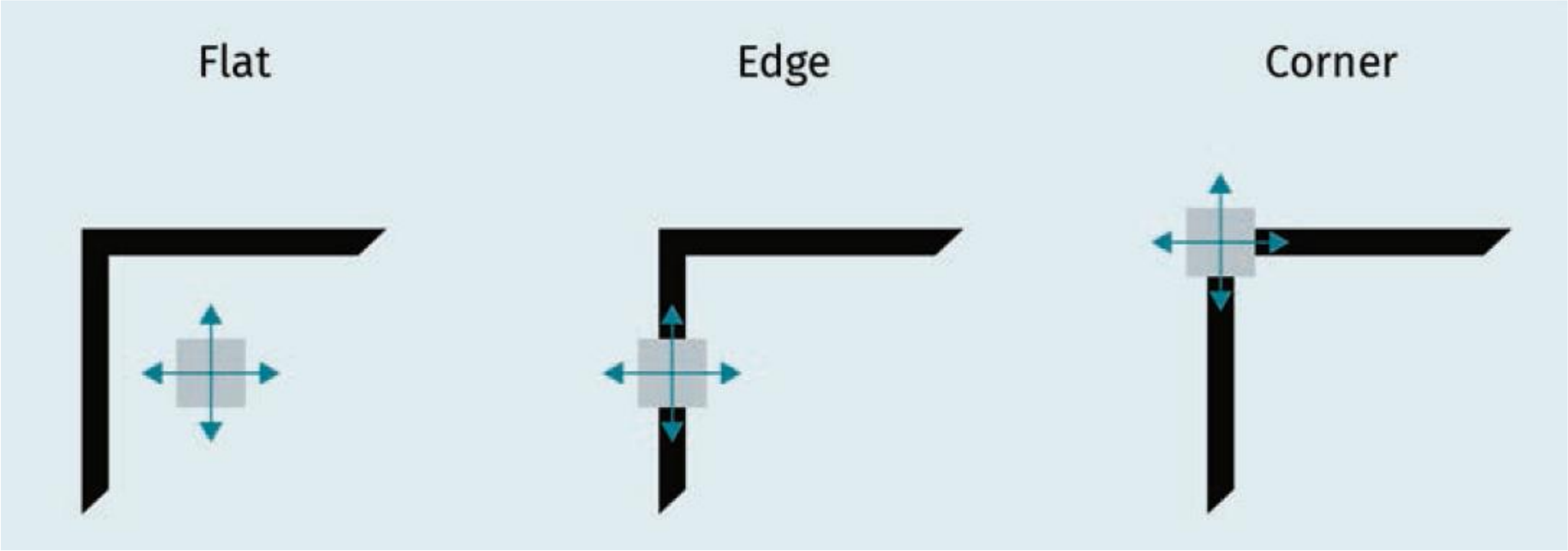


Source of image: course book

- Edge detection
  - Sobel filter: using two special kernel matrices (for each of the axes) to transfer the original image into a gradient image
    - High frequencies in the gradient image indicate areas with the high changes in pixel

Source of image: course book

- Corner detection
  - Haris method: analyzing the change of the pixel in a sliding window moving in different direction



Source of image: course book

- Binary Robust Independent Elementary Features (BRIEF)
  - Comparing the intensity of a pair of pixels
  - p(x): pixel intensity at position x

$$\tau(\mathbf{p};\mathbf{x},\mathbf{y}) := \begin{cases} 1 & \text{if } p(x) < p(y) \\ 0 & \text{otherwise} \end{cases}$$

- Oriented FAST and Rotated BRIEF (ORB)
  - Deal with the limitation of BRIEF when features are rotated more than 35^0

IMPORTANT CHARACTERISTICS OF FEATURE DETECTION ALGORITHMS

- Robustness
- Accuracy
- Repeatability
- Generality
- Efficiency
- Quantity

SEMANTIC SEGMENTATION

- Parts of an images that belong to the same object class are put into the same cluster
- Performed on the pixel-level

Original image

Segmentation map

Segmentation overlay

Background Chair Coffee table

REVIEW STUDY GOALS

On completion of this unit, you will be able to …

… define computer vision.

… explain how to represent images as pixels.

… distinguish between detection, description, and matching of features.

… correct distortion with calibration methods.

SESSION 6

TRANSFER TASK

TRANSFER TASK

1.

a) Given the following image matrix *I* and the kernel matrix *k*, compute the values of the resulting image after the convolution. What does the kernel matrix do?

$$I = \begin{bmatrix} 82 & 63 & 80 & 90 & 62 \\ 73 & 41 & 26 & 86 & 41 \\ 89 & 27 & 42 & 47 & 93 \\ 55 & 44 & 88 & 41 & 24 \\ 86 & 71 & 58 & 80 & 4 \end{bmatrix} \qquad k = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

TRANSFER TASK

1.

b) What type of filter would the kernel matrix k2 apply to an image?

c) How would the filtered image change if you applied kernel matrix k3 instead?

$$k2 = \frac{1}{9}\begin{bmatrix}1 & 1 & 1\\1 & 1 & 1\\1 & 1 & 1\end{bmatrix} \qquad k3 = \frac{1}{25}\begin{bmatrix}1 & 1 & 1 & 1 & 1\\1 & 1 & 1 & 1 & 1\\1 & 1 & 1 & 1 & 1\\1 & 1 & 1 & 1 & 1\\1 & 1 & 1 & 1 & 1\end{bmatrix}$$

2. For camera calibration, it is important to identify the external and internal parameters.
Which extrinsic and intrinsic parameters do you know?

3. Think about possible use cases for semantic image segmentation. Where could it be used?

TRANSFER TASK
PRESENTATION OF THE RESULTS

Please present your results.
The results will be discussed in plenary.

1. What are the typical tasks in computer vision?

2. What is the purpose of camera calibration?

3. What are the most commonly used types of features in computer vision?

4. What is the purpose of semantic segmentation?

How did you like the course?

HOW DID YOU
LIKE THE COURSE?