Computer Vision

Dr. Syed Faisal Bukhari

Associate Professor

Department of Data Science

Faculty of Computing and Information Technology

University of the Punjab

Textbook

Multiple View Geometry in Computer Vision, Hartley, R., and Zisserman

Richard Szeliski, **Computer Vision: Algorithms and Applications**, 1st edition, 2010

Reference books

Readings for these lecture notes:

- Hartley, R., and Zisserman, A. **Multiple View Geometry in Computer Vision**, Cambridge University Press, 2004, Chapters 1-3.

- Forsyth, D., and Ponce, J. **Computer Vision: A Modern Approach**, Prentice-Hall, 2003, Chapter 2.

These notes contain material c Hartley and Zisserman (2004) and Forsyth and Ponce (2003).

References

These notes are based

- Dr. Matthew N. Dailey's course: AT70.20: Machine Vision for Robotics and HCI
- Prof. Fei Fei Li's CS131 class at Stanford
- Dr. Sohaib Ahmad Khan CS436 / CS5310 Computer Vision Fundamentals at LUMS
- Dr. Nazar Khan, PUCIT
- https://www.sciencedirect.com/topics/engineering/stereovision
- http://serendip.brynmawr.edu/bb/kinser/Structure1.html
- https://www.tutorialspoint.com/dip/computer_vision_and_graphics.htm

Grading breakup

I. Midterm = 35 points
II. Final term = 40 points
III. Quizzes = 6 points (A total of 6 quizzes)
IV. Group project = 15 points
    a. Pitch your project idea = 2 points
    b. Research paper presentation relevant to your project = 3 points
    c. Project prototype and its presentation = 5 points
    d. Research paper in IEEE conference template = 5 points
V. OpenCV based on Python presentation = 2.5 points
VI. Matlab presentation = 2.5 points

Some top tier conferences of computer vision

I. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition **(CVPR).**
II. Proceedings of the European Conference on Computer Vision **(ECCV).**
III. Proceedings of the Asian Conference on Computer Vision **(ACCV).**
IV. Proceedings of the International Conference on Robotics and Automation **(ICRA).**
V. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems **(IROS).**

Some well known Journals

I. International Journal of Computer Vision **(IJCV).**
II. IEEE Transactions on Pattern Analysis and Machine Intelligence **(PAMI).**
III. Image and Vision Computing.
IV. Pattern Recognition.
V. Computer Vision and Image Understanding.
VI. IEEE Transactions on Robotics.
VII. Journal of Mathematical Imaging and Vision

Vision

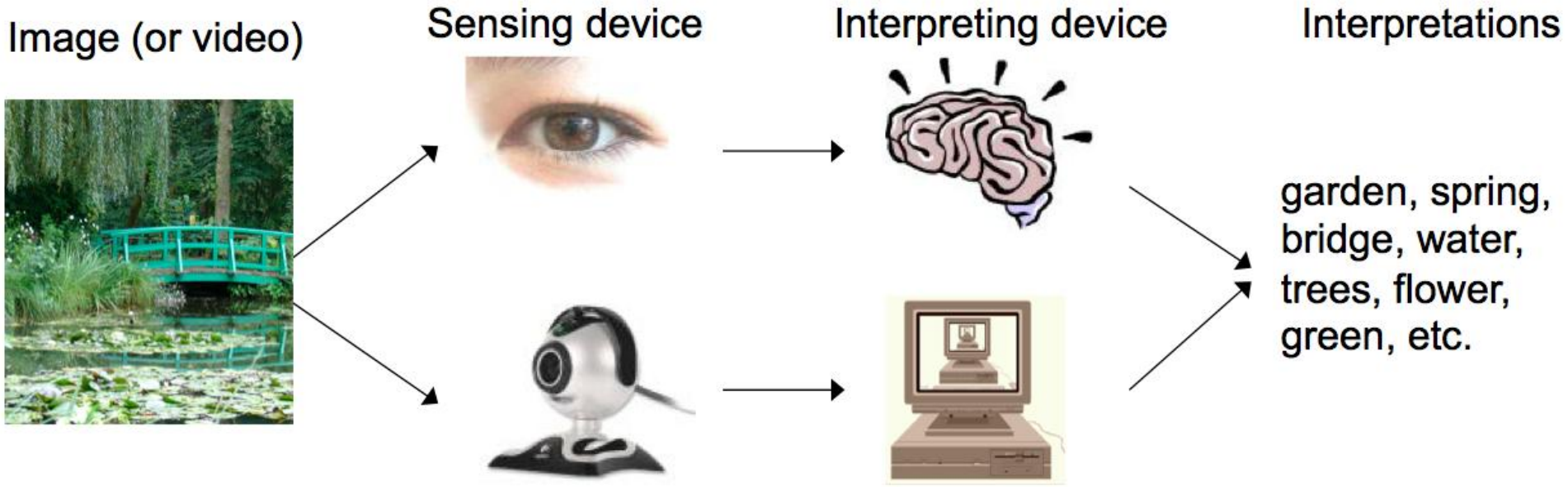
- Sight is our primary sensation

- We perceive about **80%** of all impressions by **means of sight**

- **30%** of neurons in brain's cortex are dedicated to vision, compared to **8%** for **touch**, **2%** for **hearing**

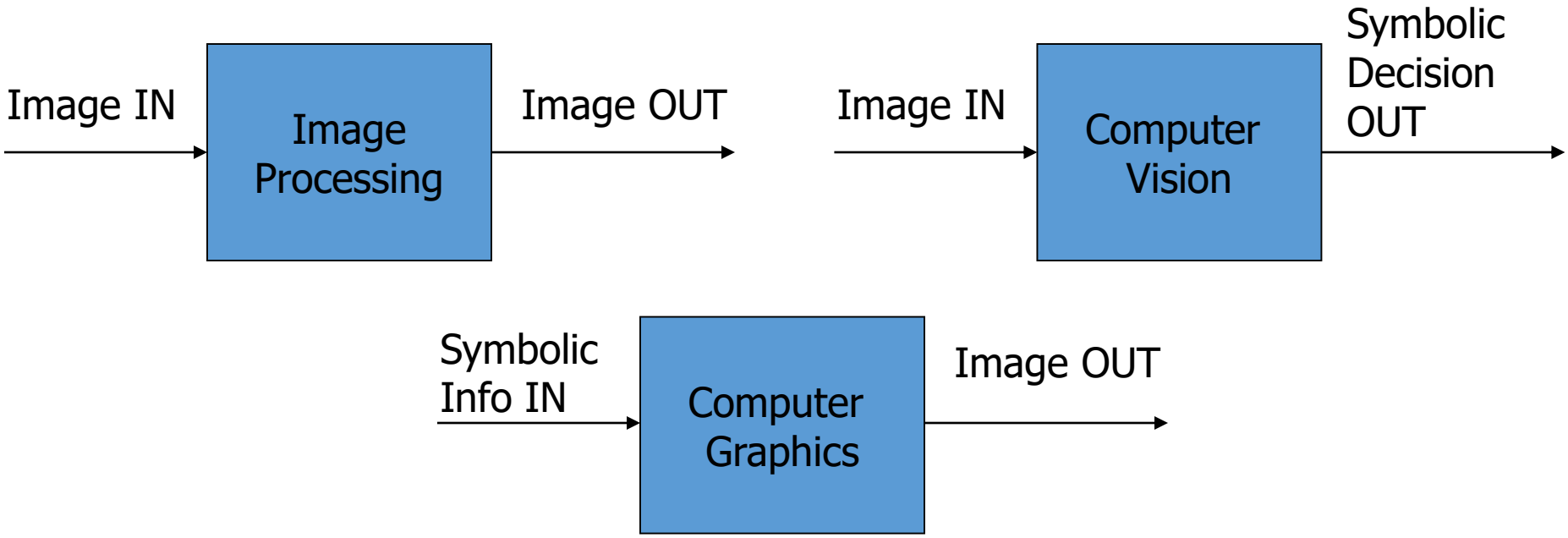
- **Note:** The **cerebrum or cortex** is the **largest part** of the human brain, associated with higher brain function such as **thought** and **action**

What is (Computer) Vision?



What is Computer Vision?

- "The goal of Computer Vision is to make **useful decisions** about real **physical objects and scenes** based on **sensed images**"



What is Symbolic Decision?

- Symbolic Decision in Computer Vision refers to **extracting meaningful information from images** and representing them in **a structured, symbolic form.**
- Converts **raw images** into **high-level representations**
- Helps AI recognize, classify, and make decisions
- Enables pattern recognition and feature extraction

Examples of Symbolic Decisions

1. Object Recognition: Identifies and labels objects (e.g., detecting a car and tagging it as **"Car"**)

2. Facial Recognition: Detecting human faces and associating them with identities

3. Autonomous Vehicles: Recognizing a stop sign and applying brakes

Examples of Symbolic Decisions

4. Medical Imaging: Detecting anomalies in X-rays and categorizing them as diseases

5. Scene Understanding: Identifying relationships between objects in an image

Process of Symbolic Decision

1. Image Processing: Enhancing image quality and extracting features

2. Feature Extraction: Identifying edges, textures, and key points

3. Classification & Recognition: Using AI to recognize patterns

4. Decision-Making: Converting recognized objects into symbolic outputs (e.g., **'Stop Sign Detected' → 'Apply Brakes'**)

Conclusion

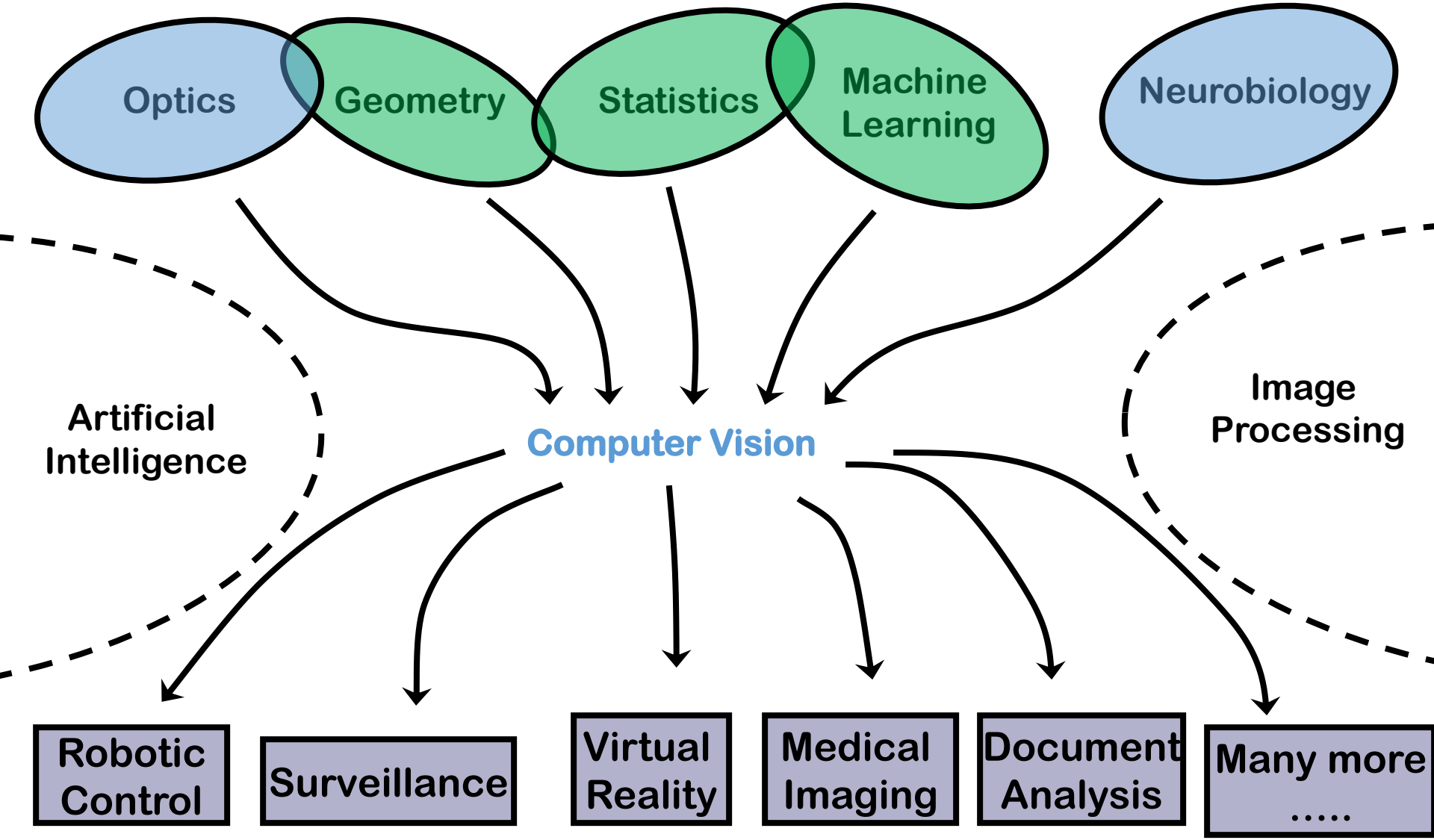
Symbolic decisions bridge the gap between image perception and actionable insights, allowing AI-powered systems to make meaningful decisions based on visual input.

Computer Graphics

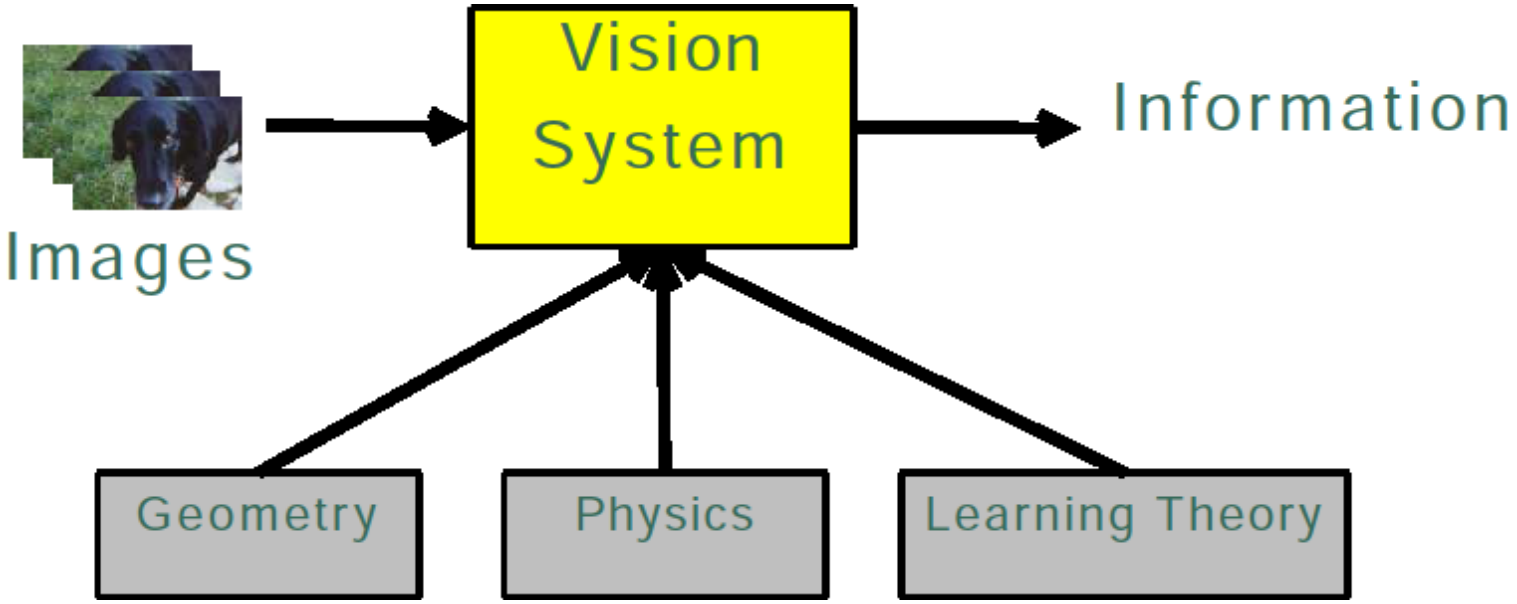
- Computer graphics are graphics created using computers and the representation of image data by a computer specifically with help from **specialized graphic hardware and software**.

- Formally we can say that Computer graphics is **creation, manipulation and storage of geometric objects** (modeling) and their **images** (Rendering).

What is computer vision?



Introduction: Vision systems



The kind of information we want is application specific:

3D models

Object categories

Object poses

Camera poses

Reconstruction of 3D Structure

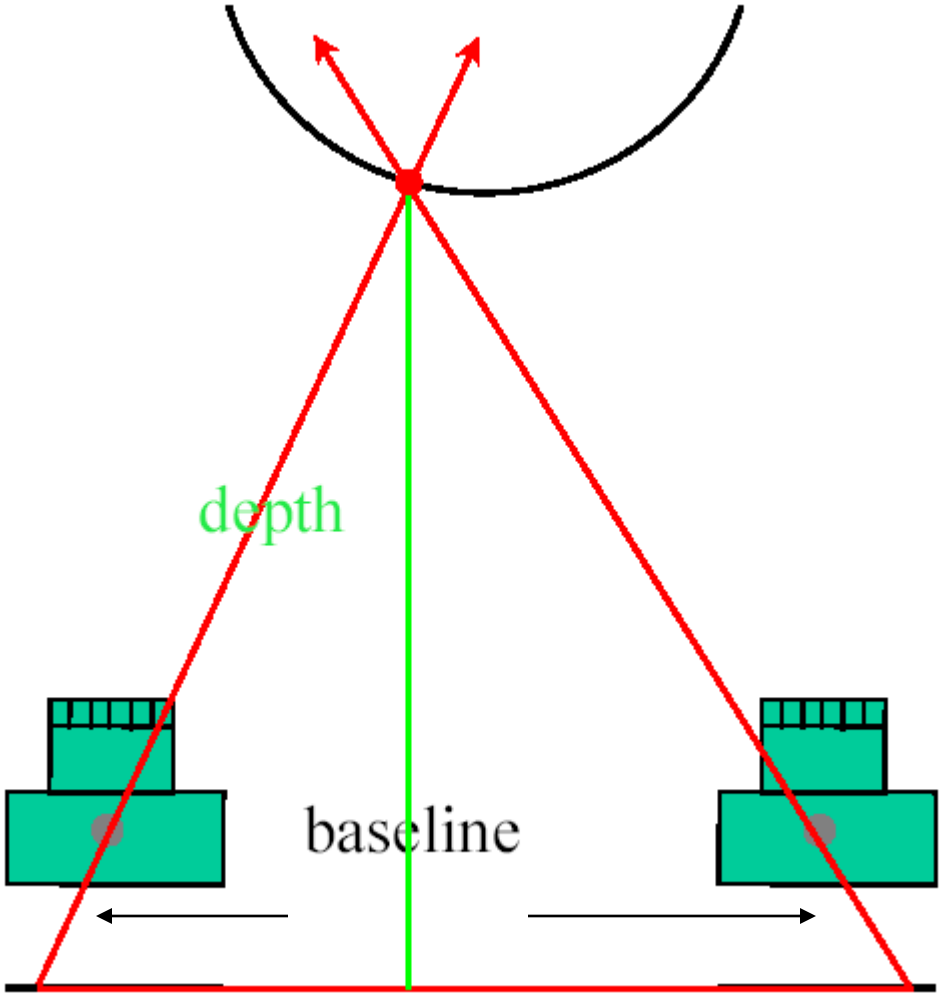
- An Image is a 2-Dimensional projection of 3D World
- 3D can be reconstructed from
  - Two images
    - Stereo Problem
  - Video with moving camera
    - Structure from Motion Problem
  - Some understanding about what is being viewed
    - Geometrical inference
    - Shape from shading or texture

Stereovision

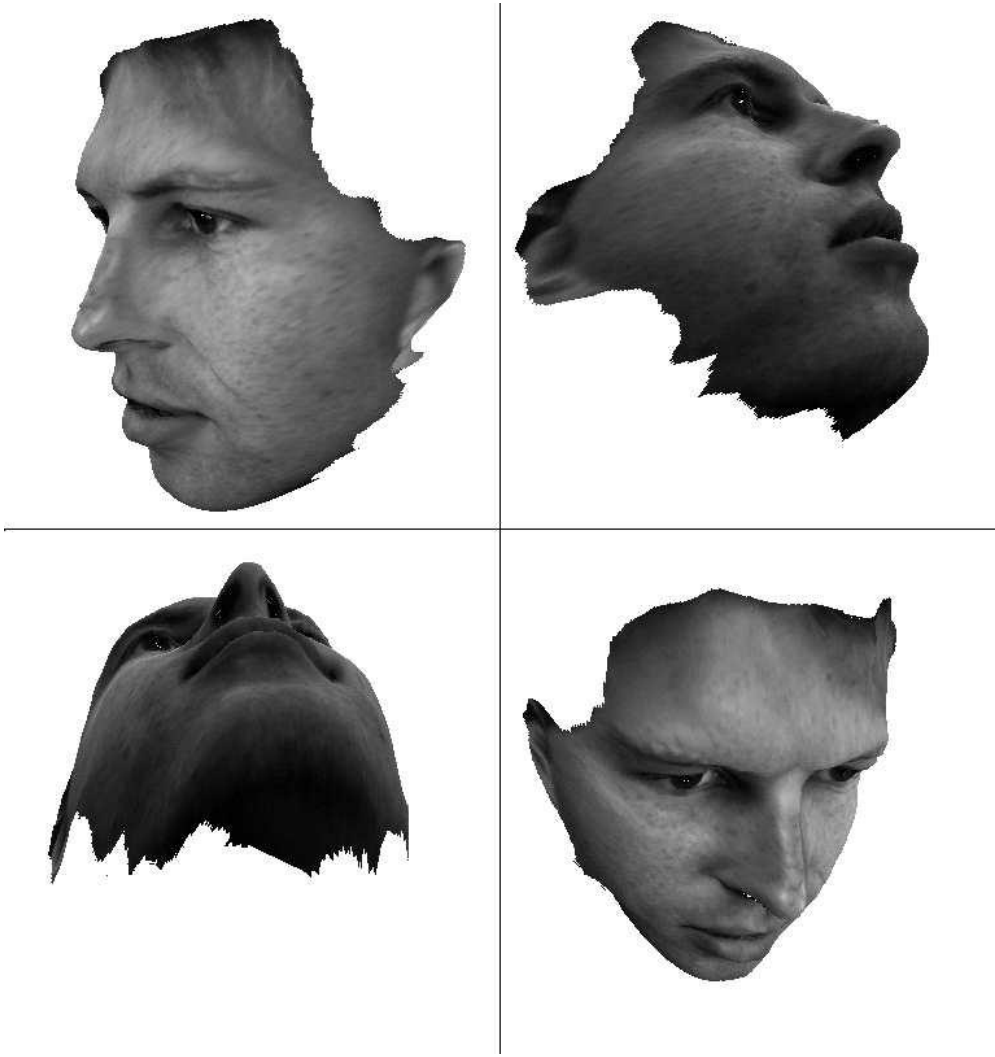
Stereovision techniques use two cameras to see the same object.

- The two cameras are separated by a baseline, the distance for which is assumed to be known accurately.
- The two cameras simultaneously capture two images.
- The two images are analyzed to note the differences between the images. Essentially, one needs to accurately identify the same pixel in both images, known as the problem of correspondence between the two cameras.

Stereo



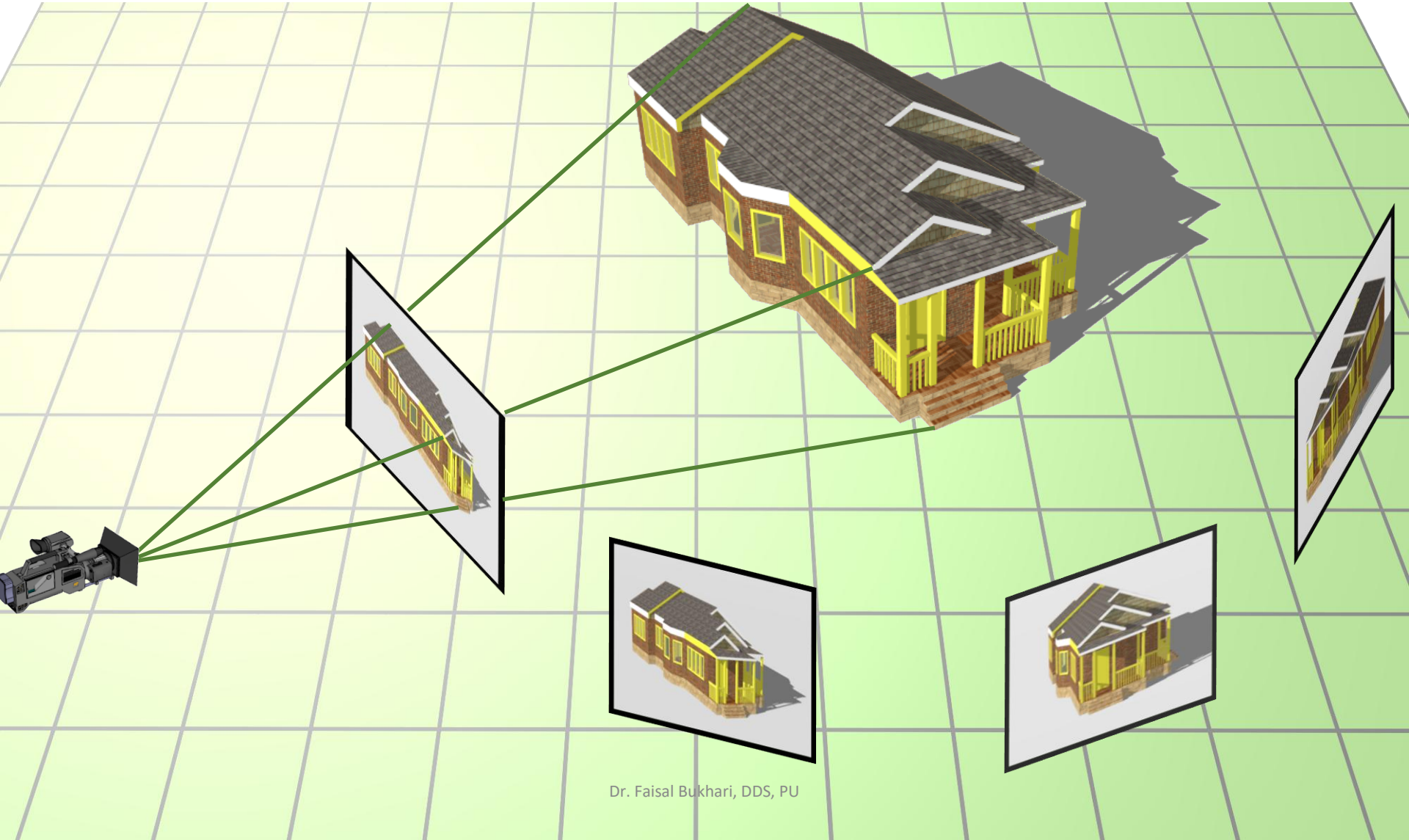
Stereo image pair.



3-D reconstructions

L. Alvarez, R. Deriche, J. S´anchez, J. Weickert (2002).

Rigid Structure from Motion



A vision system

"A vision system is something that takes images and output information of some kind".

- The information that we want could be a ton of different possible things. We might want to know what is a **3D structure of a given scene**.

- **Object categories**

We might want to know what is the **contents of the images that we are seeing**.

- Is it a dog?
- Is it a cat?
- Is it a person?
- Is it my friend?
- Is it my relative?
- Etc.

A vision system

- We might also figure out **something about parts of the objects in the scene to build some models of the objects in the scene based on their parts**. A lot of different things we might want to output from an image.
- **3D structure** and **object categories** might be slightly apparent types of things we might want to output, but then there are other things, interesting in certain situations.

- For example, **camera poses.**

Introduction to Camera Position

- What is camera position?
- Why do we need to know the camera position?

Mathematical Representation

- Camera coordinate system

- Pinhole camera model (including projection)

Applications of Camera Position

- 3D Reconstruction
- Augmented Reality (AR) & Virtual Reality (VR)
- Robotics & Autonomous Navigation
- Computer Graphics & Rendering

Techniques for Determining Camera Position

Structure-from-Motion (SfM)

Simultaneous Localization and Mapping (SLAM)

Camera Calibration & Extrinsic Parameters

What is Structure from Motion (SfM)?

A technique for reconstructing **3D structures** from **multiple 2D images**.

Estimates both camera motion and 3D scene structure simultaneously.

Commonly used in photogrammetry, AR/VR, and robotics.

What is SLAM?

SLAM enables a system to localize itself while **mapping an unknown environment**.

Used in robotics, self-driving cars, drones, and AR/VR systems.

Works by detecting environmental features and tracking motion.

What is Camera Calibration?

- **Camera calibration** determines **the camera's internal and external parameters**.

- Helps correct lens distortion and improve accuracy in 3D reconstruction.

- Essential for computer vision applications like AR, robotics, and SLAM.

Real-World Examples

- Object tracking

- Drone-based photogrammetry

- Self-driving cars

Camera poses [1]

- Suppose each image in a sequence is **an image of the same scene**. It means for every image in a sequence.

- What we want to know, what was the **position of the camera** at each point in time.

If each image in a sequence captures the same scene, our goal is to determine the camera's position at each moment in time throughout the sequence.

Camera poses [2]

- Why do we want to know the position of the camera?
- **Actually, two things:**
  - To know the **position of the object** in the real world, the **things we are looking at** (position of **object** in the real world).
  - But also, often time you might want to know the position of the thing that is **seeing the object in the real-world** (position of the **camera** in the real world).

Why Determine the Camera's Position?

- To accurately determine the position of objects in the real world.
- To track the motion of the camera itself as it moves through space.
- Essential for applications like robotics, augmented reality, and 3D reconstruction.

Camera poses [3]

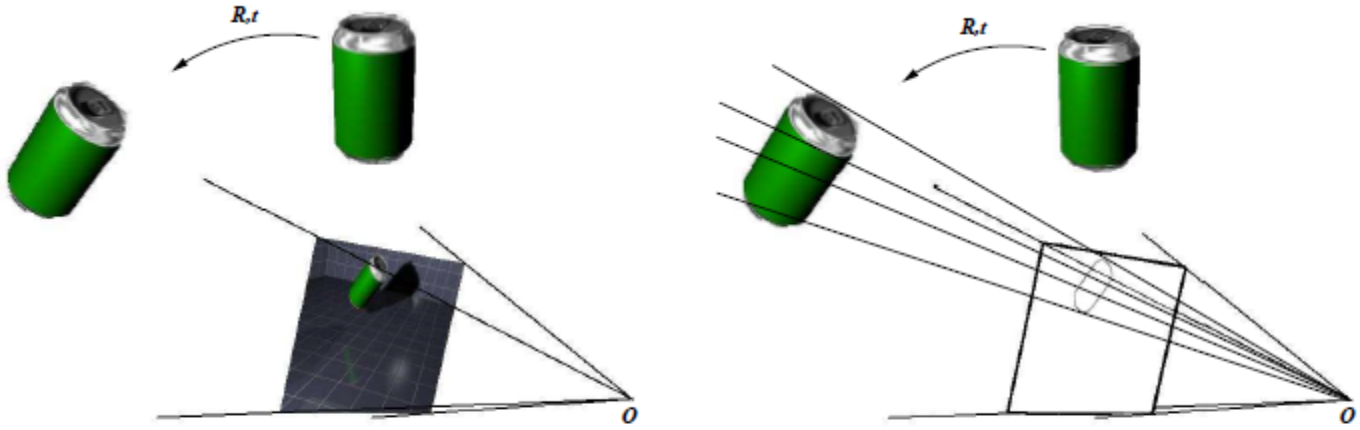
- The **camera poses** is related to **localization**, which means we have a **camera moving around** in a world.

- Maybe this **camera** is **mounted on a robot** and we want to know where is the **robot at each point in time.**

- Maybe we are working in indoor, and we don't have **GPS** and so on.

- It turns out **camera poses** can be **extracted from a sequence of images** if we add a little bit of extra information.

Pose Estimation



(a) 2D-3D Pose Estimation Scenario

(b) Silhouette-based *explicit* 2D-3D Pose Estimation

Figure 1. 2D-3D Pose Estimation

Author: Nazar Khan (2007)

Find the rotation ***R*** and translation ***t*** that aligns a **3D object** with its **2D image**.

Object poses

- **For example**, if we **track people** for our specific application, we might be interested in whether the person is walking this way and facing this way and what the person is looking at.

If our application involves tracking people, we may be particularly interested in determining their direction of movement, the orientation they are facing, and what they are looking at.

What is Camera Pose?

- Defines the position and orientation of a camera in 3D space.
- Essential for 3D reconstruction, AR, robotics, and computer vision.
- Includes translation (position) and rotation (orientation).

Components of Camera Pose

1. **Translation (T):** Camera's position in 3D space (x, y, z). It represents position using a **3 × 1 vector**

2. **Rotation (R):**Camera's orientation, defined by rotation matrices or Euler angles. It represents orientation using a **3 × 3 matrix**

3. **Transformation Matrix:** Combines rotation & translation into a **4 × 4 matrix**.

$$P = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \quad or \quad \begin{bmatrix} R_{3\times3} & T_{3\times1} \\ \mathbf{0}^{T} & 1 \end{bmatrix}_{4\times4}$$

Applications of Camera Poses

Augmented Reality (AR): Aligns virtual objects with real world.

3D Reconstruction: Builds 3D models from multiple images.

Robotics & Drones: Helps in navigation and object recognition.

Autonomous Vehicles: Used in SLAM (Simultaneous Localization and Mapping).

Challenges in Camera Pose Estimation

Motion Blur: Reduces accuracy in moving environments.

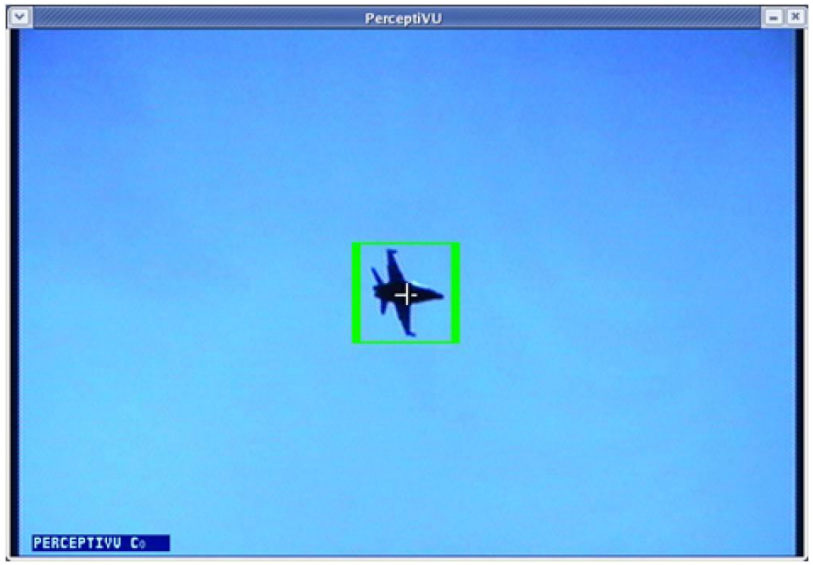
Occlusions: Objects blocking the scene affect results.

Feature Matching Errors: Poor feature detection can lead to incorrect pose estimation.

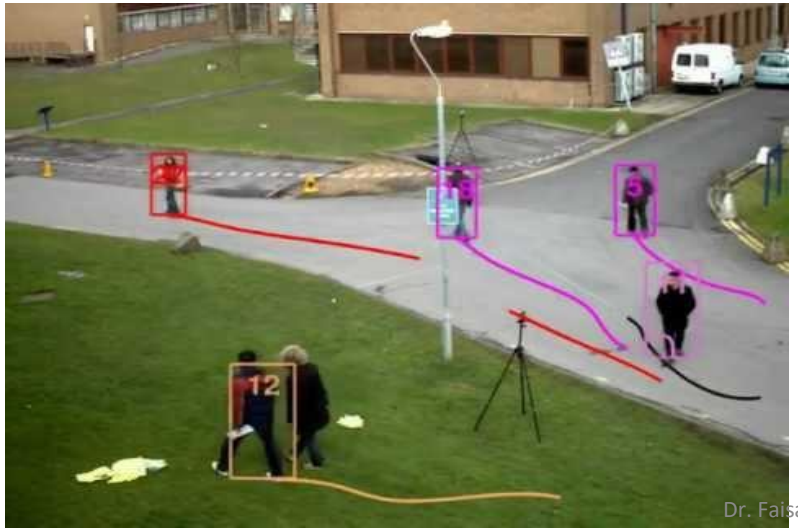
Conclusion

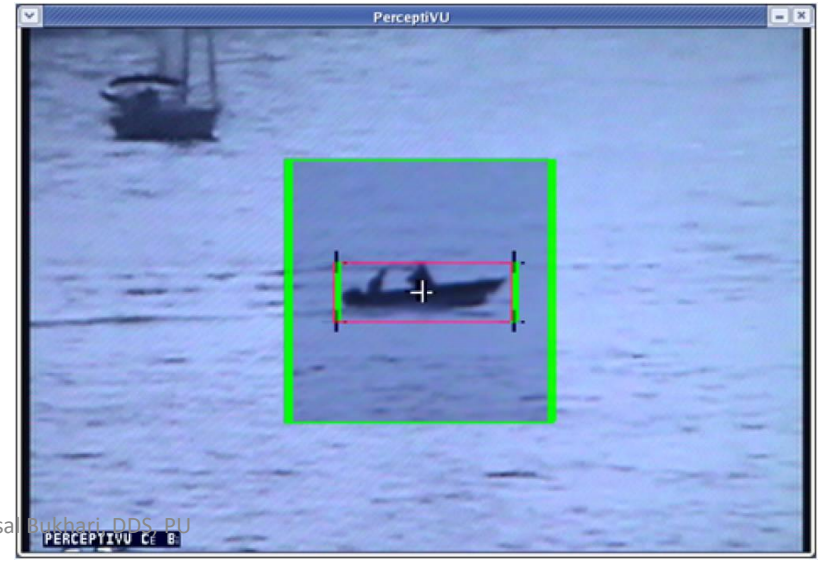
- Camera poses define how cameras perceive 3D space.
- They are crucial in applications like AR, 3D reconstruction, and robotics.
- Accurate pose estimation enables better visual understanding and decision-making.

Tracking


PerceptiVU
PERCEPTIVU Co


PerceptiVU
PERCEPTIVU Hs B


12


PerceptiVU
PERCEPTIVU Ce B

A vision system

- The **output of a vision system** depends on a specific application that we have in mind. Still, all vision systems have this in common taking **image as input and outputting information as output**. How do we do that?

- There are variety of different things in order to create that information. We can't just analyze the **pixel information directly, but** we should have **some source of intelligence**.

- There is another significant source of information that we are going to use to perform some of these **tasks is machine learning.**

A vision system

- We will incorporate:
  - **Geometry**
  - **Physics**
  - **Machine learning**

in our vision system in order to **transduce images** into **information**. That's the goal of this course.

The Complexity of Perception

The goal of Computer Vision

- To **bridge the gap** between **‘pixels’** and **‘meaning’**

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 the computer gets as input**

Introduction: Applications

- Important applications include:
  - Mobile robot navigation
  - Industrial inspection and control
  - Military intelligence
  - Security
  - Human-computer interaction
  - Image retrieval from digital libraries
  - Medical image analysis
  - 3D model capture for visualization and animation

Example: Applications

- Robotics
- Medicine
- Security
- Transportation
- Industrial Automation

Robotics Application

- Localization-determine robot location automatically
- Navigation
- Obstacles avoidance
- Assembly (peg-in-hole, welding, painting)
- Manipulation (e.g. PUMA robot manipulator)
- Human Robot Interaction (HRI): Intelligent robotics to interact with and serve people

Medicine Application

- Classification and detection (e.g. lesion or cells classification and tumor detection)
- 2D/3D segmentation
- 3D human organ reconstruction (MRI or ultrasound)
- Vision-guided robotics surgery

Industrial Automation Application

- Industrial inspection (defect detection)
- Assembly
- Barcode and package label reading
- Object sorting
- Document understanding (e.g. OCR)

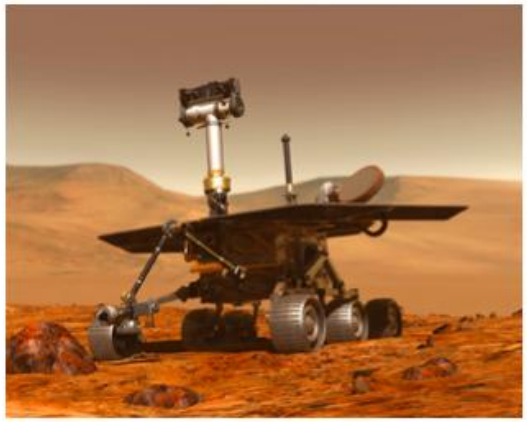
Security Application

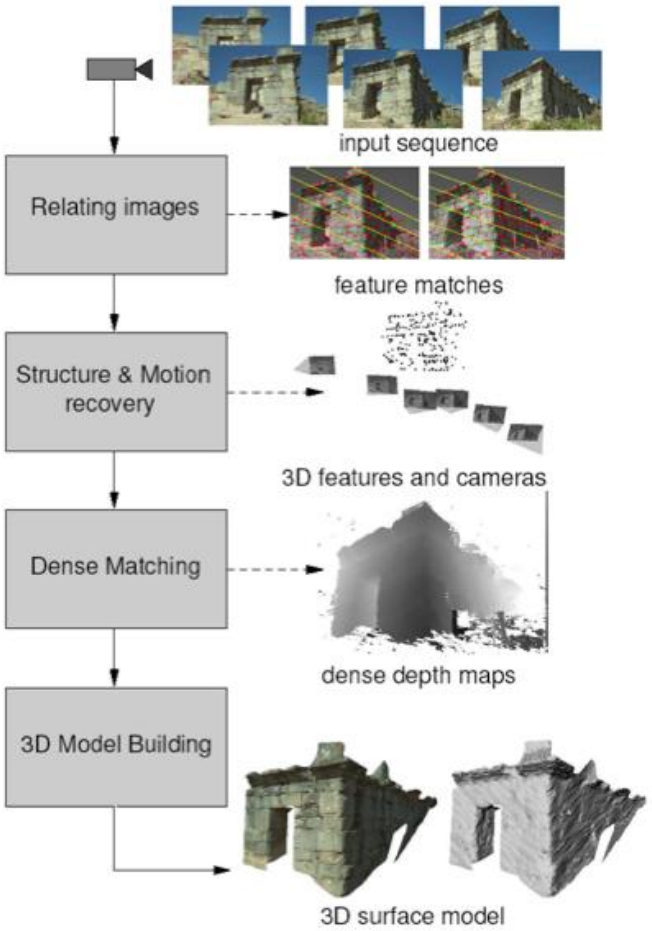
- Biometrics (iris, finger print, face recognition)
- Surveillance-detecting certain suspicious activities or behaviors

Transportation Application

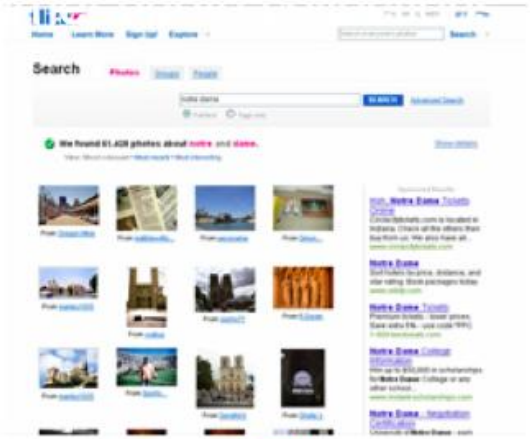
- Autonomous vehicle
- Safety, e.g., driver vigilance monitoring

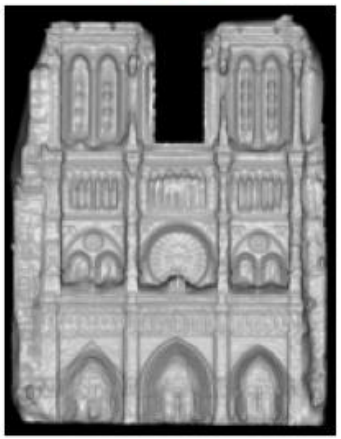
Vision as a measurement device





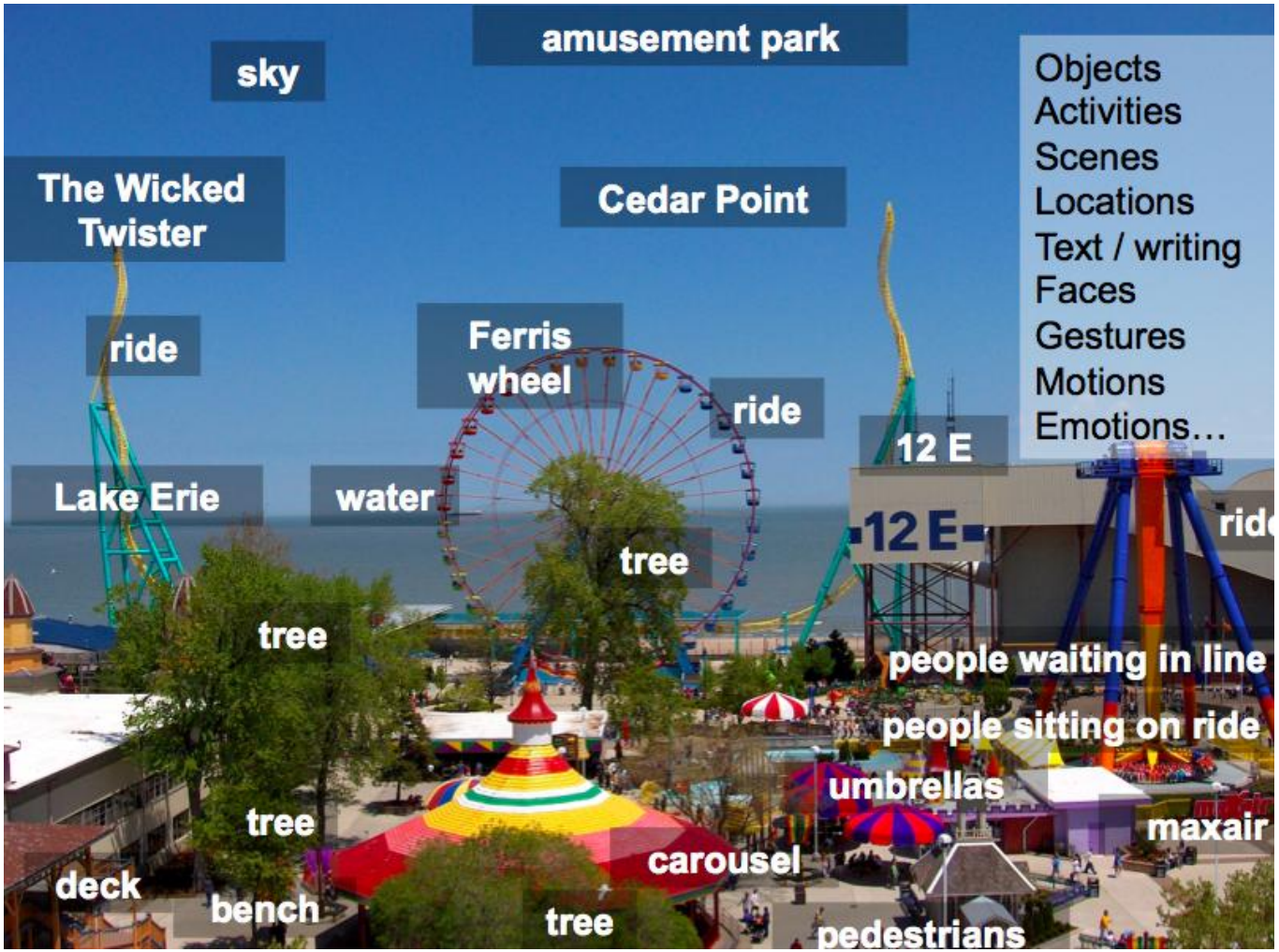
Pollefeys et al.

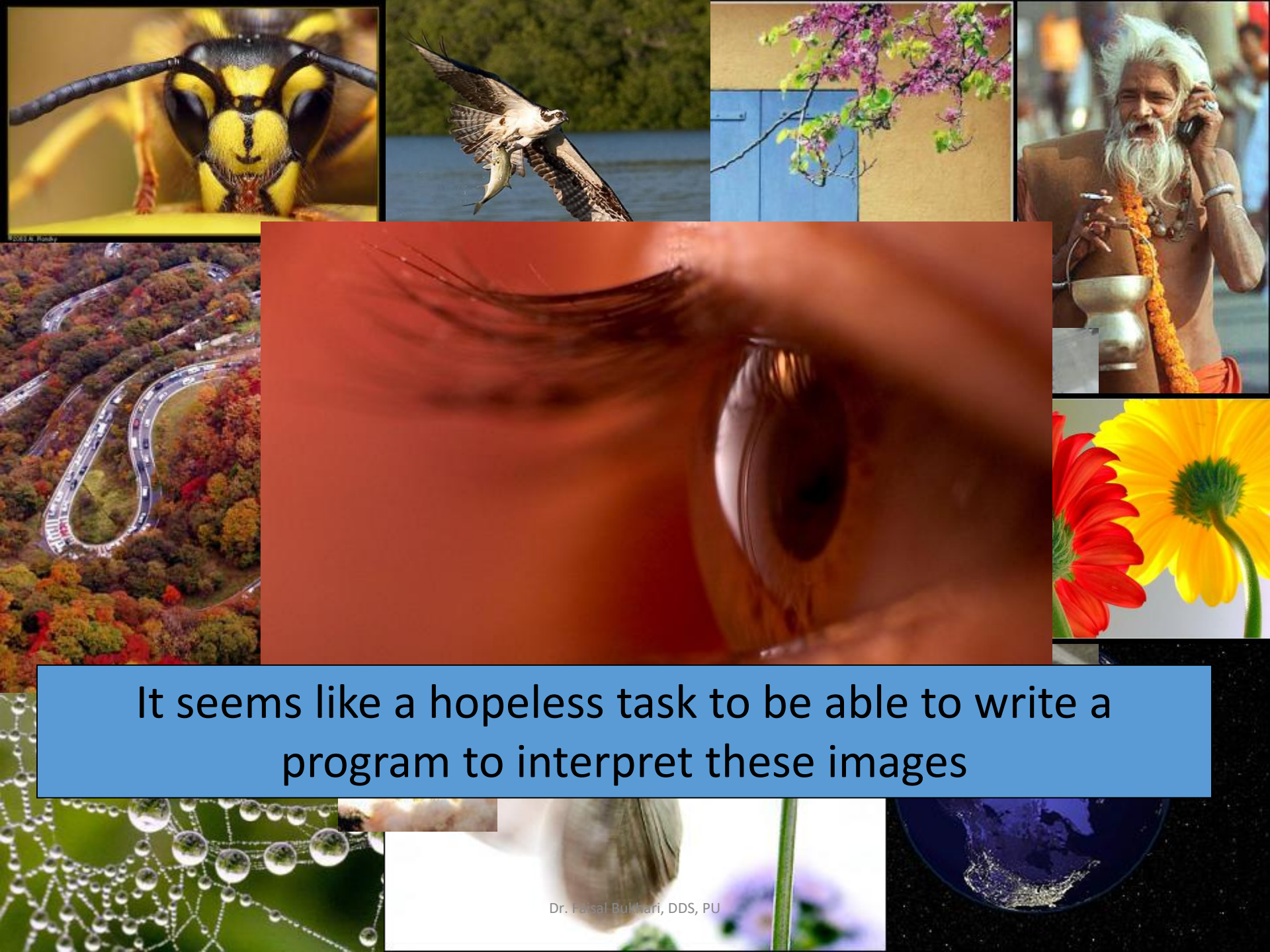




Goesele et al.

Vision as a source of semantic information



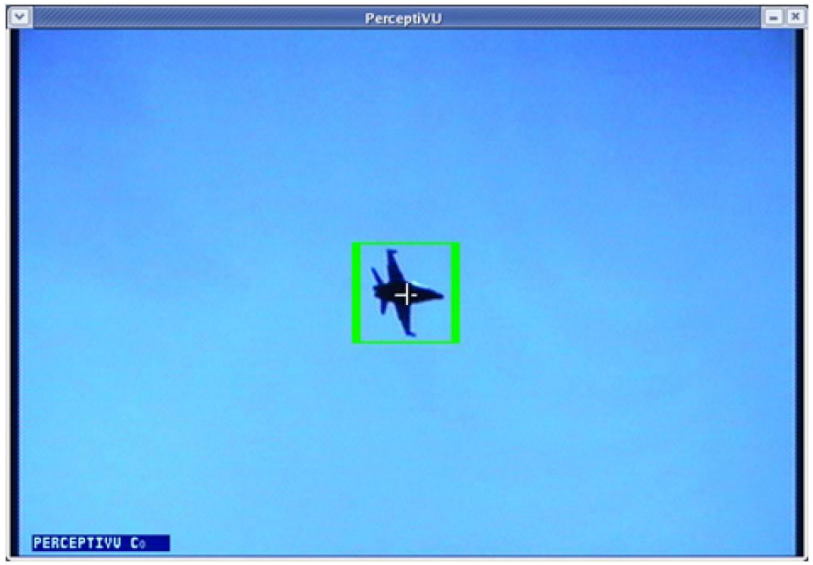


It seems like a hopeless task to be able to write a program to interpret these images

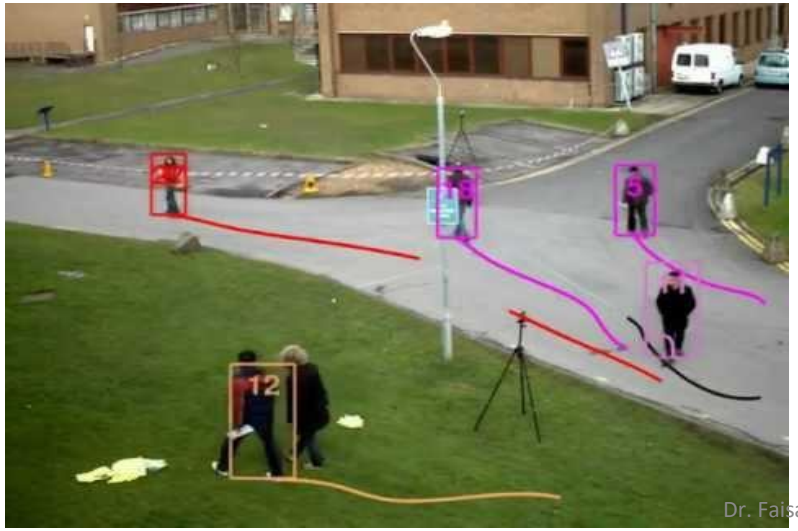
3D Reconstruction from a Single View

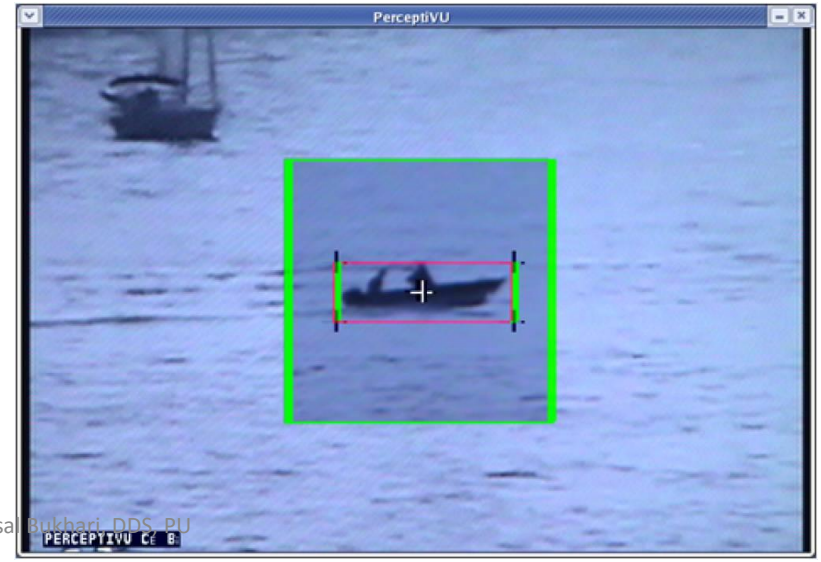
Saint Jerome in his study (1630) Joseph R.Ritman Collection
by Henry V Steinwick (1560-1649)

Tracking


PerceptiVU
PERCEPTIVU Co


PerceptiVU
PERCEPTIVU Hs B


12


PerceptiVU
PERCEPTIVU Ce B

Action Recognition