Artificial Intelligence
Subject Code: 20A05502T

UNIT IV - Natural Language for Communication

Perception

- Image Formation,

- Early Image Processing Operations,

- Object Recognition by appearance,

- Reconstructing the 3D World,

- Object Recognition from Structural information,

- Using Vision.

- **Perception** provides agents with information about the world they inhabit by interpreting the response of **sensors**.
- A sensor measures some aspect of the environment in a form that can be used as input by an agent program.
- A variety of sensory modalities are available to artificial agents.
- Those they share with humans include vision, hearing, and touch.
- Modalities that are not available to the unaided human include radio, infrared, GPS, and wireless signals.
- Some robots do **active sensing**, meaning they send out a signal, such as radar or ultrasound, and sense the reflection of this signal off of the environment.

- a model-based decision-theoretic agent in a partially observable environment has a **sensor model**

- a probability distribution $\mathbf{P}(E \mid S)$ over the evidence that its sensors provide, given a state of the world.

- Bayes' rule can then be used to update the estimation of the state.

- For vision, the sensor model can be broken into two components:
- 1. An **object model**
- describes the objects that inhabit the visual world—people, buildings, trees, cars, etc.
- The object model could include a precise 3D geometric model taken from a computer-aided design (CAD) system, or it could be vague constraints, such as the fact that human eyes are usually 5 to 7 cm apart.
- 2. A **rendering model**
- describes the physical, geometric, and statistical processes that produce the stimulus from the world. Rendering models are quite accurate, but they are ambiguous.
- For example, a white object under low light may appear as the same color as a black object under intense light.
- A small nearby object may look the same as a large distant object. Without additional evidence, we cannot tell if the image that fills the frame is a toy Godzilla or a real monster.

- Ambiguity can be managed with prior knowledge

- A decision-theoretic agent is not the only architecture that can make use of vision sensors.

- Flies and many other flying animals make used of a closed-loop control architecture to land on an object.

- The visual system extracts an estimate of the distance to the object, and the control system adjusts the wing muscles accordingly, allowing very fast changes of direction, with no need for a detailed model of the object.

- Compared to the data from other sensors
- visual observations are extraordinarily rich, both in the detail they can reveal and in the sheer amount of data they produce.

- We can characterize three broad approaches to the problem.
- The **feature extraction** approach, as exhibited by *Drosophila*, emphasizes simple computations applied directly to the sensor observations.
- In the **recognition** approach an agent draws distinctions among the objects it encounters based on visual and other information.
- Recognition could mean labeling each image with a yes or no as to whether it contains food that we should forage, or contains Grandma's face.
- In the **reconstruction** approach an agent builds a geometric model of the world from an image or a set of images.

IMAGE FORMATION

- Imaging distorts the appearance of objects.

- For example, a picture taken looking down a long straight set of railway tracks will suggest that the rails converge and meet.

- another example, if you hold your hand in front of your eye, you can block out the moon, which is not smaller than your hand.

- As you move your hand back and forth or tilt it, your hand will seem to shrink and grow *in the image*, but it is not doing so in reality

- Models of these effects are essential for both recognition and reconstruction.
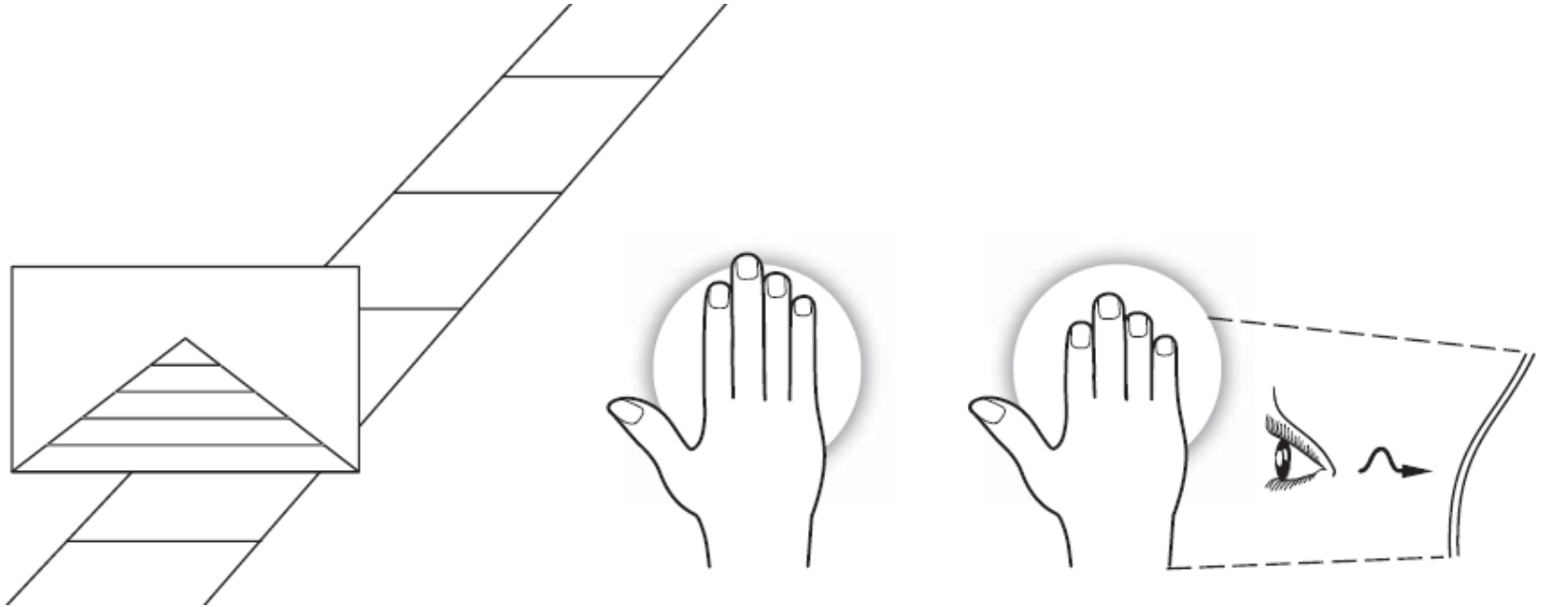
- Imaging distorts geometry.
- Parallel lines appear to meet in the distance, as in the image of the railway tracks on the left.
- In the center, a small hand blocks out most of a large moon.
- On the right is a foreshortening effect: the hand is tilted away from the eye, making it appear shorter than in the center figure.

Images without lenses: The pin-hole camera

- Image sensors gather light scattered from objects in a **scene** and create a two-dimensional **image**.

Thank You