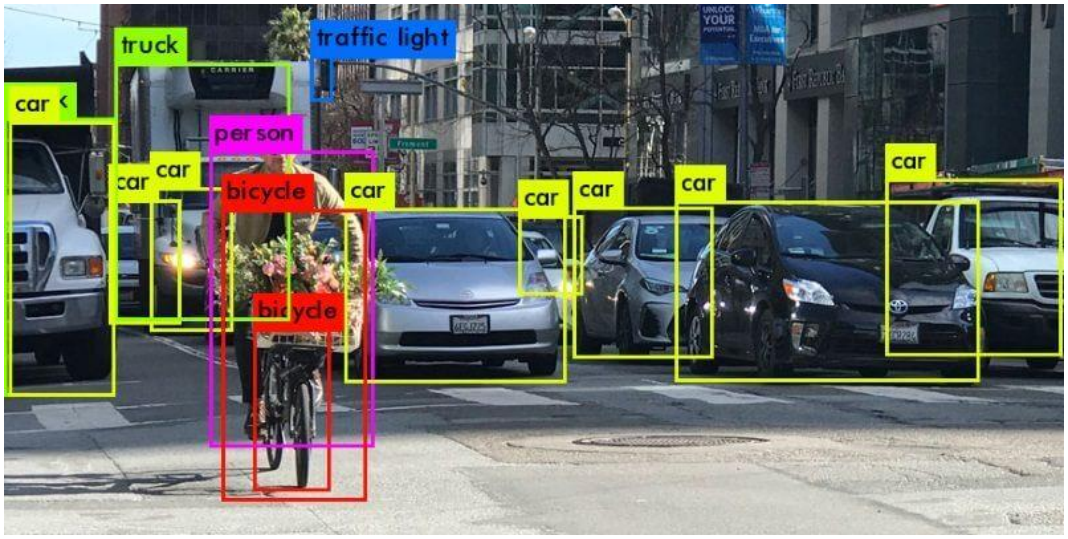
ARTIFICIAL INTELLIGENCE

Lecture Notes

Unit-IV

Natural Language Communication & Perception



Prepared by

Dr.V V N BHASKAR

HOD of Mechanical Engineering

ADITYA COLLEGE OF ENGINEERING

MADANAPALLE

Unit – IV

NATURAL LANGUAGE COMMUNICATION & PERCEPTION

JNTUA SYLLABUS

Unit – IV: **Natural Language for Communication**: Phrase structure grammars, Syntactic Analysis, Augmented Grammars and semantic Interpretation, Machine Translation, Speech Recognition

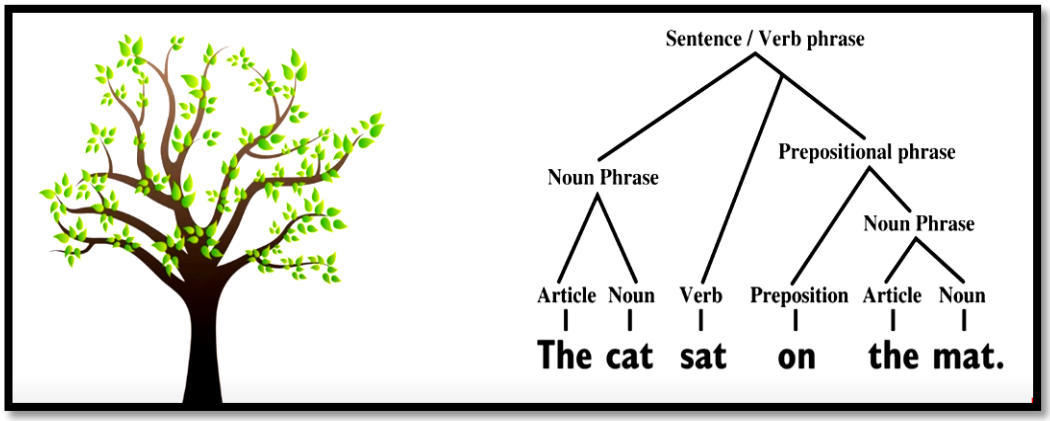
Perception: Image Formation, Early Image Processing Operations, Object Recognition by appearance, Reconstructing the 3D World, Object Recognition from Structural information, Using Vision.

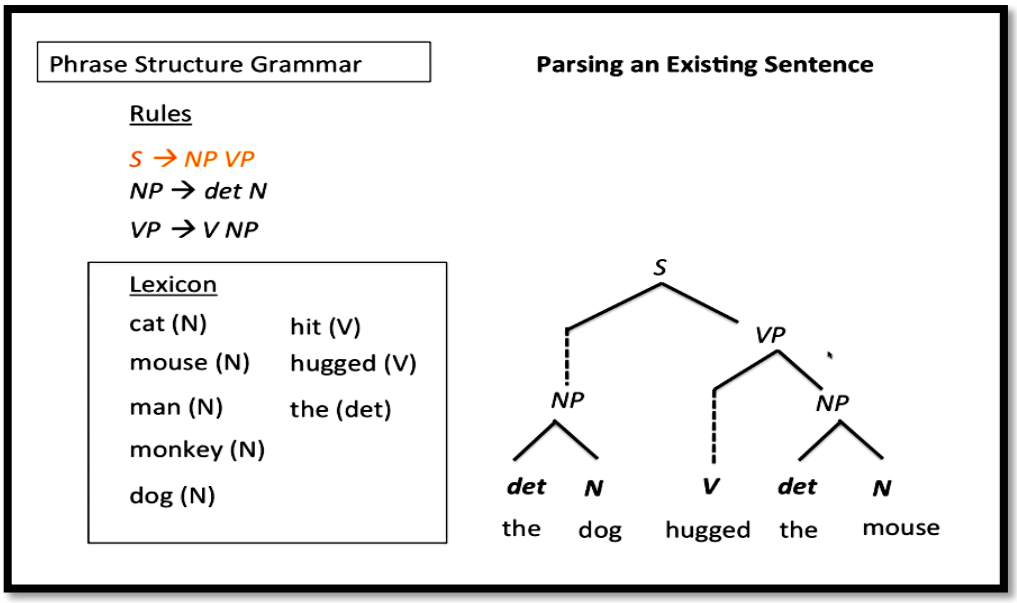
CONTENTS

Unit –IV
NATURAL LANGUAGE FOR COMMUNICATION

4.1 PHRASE STRUCTURE GRAMMAR [PSG]

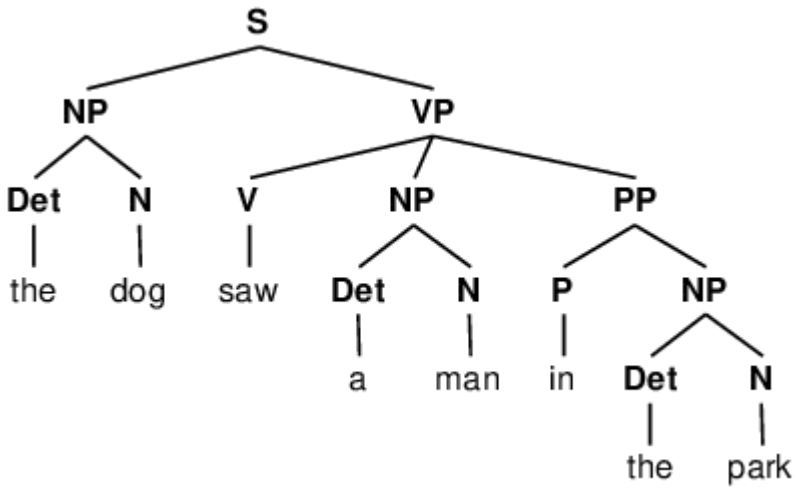
- A grammar is a set of rules that generates valid sentences in a language as a set of allowable strings of words.
- Phrase structure grammar is a type of generative grammar in which constituent structures are represented by phrase structure rules.
- A phrase is a short or long group of words that does not convey a complete thought.
- A sentence is also a group of words, but it conveys a complete thought
- The N-gram language models are based on sequence of words.
- The big issue for these models is huge vocabulary of, say, 10^5 words; there are 10^{15} trigram probabilities to estimate.
- A corpus of even a trillion words will not be able to supply reliable estimates for all of them.
- This problem can be addressed with generalization.
- For example, "black dog" is more frequent than "dog black". and similar observations like adjectives tend to come before nouns in English.
- **Probabilistic Context Free Grammar (PCFG):**
  - One of the popular language models for phrase structure
  - PCFG is a probabilistic version of CFG where each production has a probability.
  - PCFG rule: VP → V [0.70]
    | VP NP [0.30]
    Where V – Verb VP – Verb Phrase NP – Noun Phrase
  - This rule says that, with probability 0.70 a verb phrase consists solely of a verb, & with probability 0.30 it is a VP followed by an NP.
- **Parse Tree or Syntax tree**:
  - It is a tree representation of syntactic structure of sentences or strings.



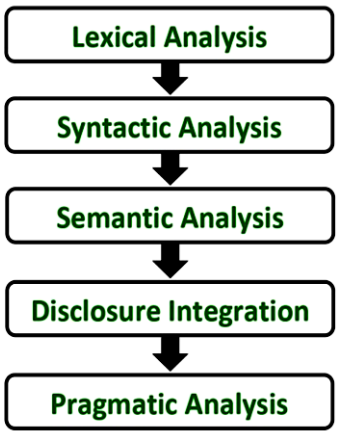


4.2 SYNTACTIC ANALYSIS (PARSING)

- In Syntax analysis, the sentence is checked whether it is well-formed or not.
- **Parsing** – Creates a parse tree of a sentence.
- The purpose of this phase is to draw exact meaning from the text.
- This is used to check grammar, word arrangements, and shows the relationship among the words.
- This is used to assess how Natural Language aligns with grammatical rules.
- Syntax refers to arrangement of words in a sentence such that they are grammatically correct.
  - Example 1: **"The school goes to boy"**
    In the real world, the above sentence does not make any sense, so this sentence is rejected by the Syntactic analyzer
  - Example2: **"Innocent peacefully cats sleep little"** is rejected.
    "Innocent little cats sleep peacefully" is accepted.



4.3 AUGMENTED GRAMMARS AND SEMANTIC INTERPRETATION



1. **Lexical Analysis/Morphological Processing.**
   This phase scans the source code as a stream of characters and converts it into meaningful lexemes. It divides the whole text into paragraphs, sentences, and words.
   For example, the sentence "He goes to college." is divided into ['He', 'goes', 'to', 'college', '.']. There are five tokens in the sentence.
2. **Syntactic Analysis**
   Syntactic Analysis is used to check grammar, word arrangements, and shows the relationship among the words.
   Example: "**Delhi goes to the Poonam"**
   In the real world, Delhi goes to the Poonam, does not make any sense, so this sentence is rejected by the Syntactic analyzer.
3. **Semantic Analysis**
   Semantic analysis is the extension of Syntactic analysis. It helps in checking the meaningfulness in text. In this phase, the sentence is checked for the literal meaning of each words and phrases. For example, the sentences like **"I ate hot ice cream"**, **"dry water"** will get rejected by the semantic analyzer because it doesn't make sense.
   Issues in Semantics:
   How to understand the meaning, specially that words are ambiguous and polysemous (may have multiple meanings)
   Example: Buy this table? Serve that table? Sort the table?
4. **Disclosure Integration**
   The meaning of the individual sentence may depend on the previous sentences. It may influence the meaning of the sentences that follow it.
   It involves building the relationship and meaningfulness between two consecutive sentences or phrases.
   E.g. **"Rahul wanted it."** The word "it" depends on the previous sentence.

5. **Pragmatic Analysis**

   Pragmatics is the study of speaker meaning. It incorporates the current situation into account. It helps to discover the intended effect by applying a set of rules that characterize cooperative dialogues.

   For Example: **"Open the door"** is interpreted as a request instead of an order.

4.3.1 Ambiguity

A system with a large grammar and lexicon might find thousands of interpretations for perfectly ordinary sentence. This is difficult because of Ambiguity and Uncertainty in the language.

Following are the three types of ambiguity.

1. Lexical Ambiguity

- Lexical Ambiguity exists in the presence of two or more possible meanings within a single word.
- **Example:**
  - Manya is looking for a **match**. (Ambiguity: looking for a partner or a cricket match?
  - The fisherman went to the **bank**. (Ambiguity: Regular bank of River bank?)
  - The **tank** was full of water. (Ambiguity: regular tank or military tank?)

2. Syntactic Ambiguity (Structure)

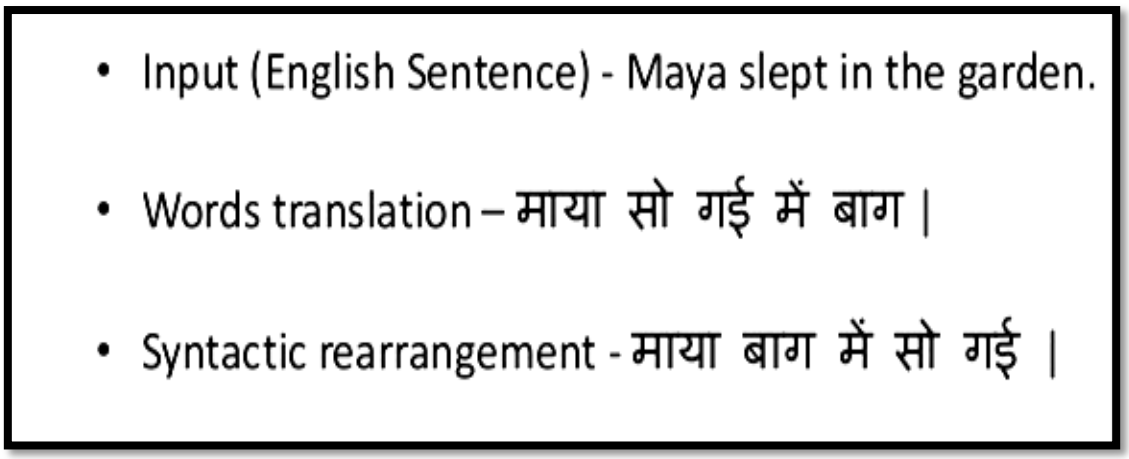
- Syntactic Ambiguity exists in the presence of two or more possible meanings within the sentence.
- **Example:**
  - The chicken is ready to eat.(Ambiguity: Chicken is ready to take his own food? or chicken is ready for us to eat)
  - I saw the man with the binoculars. (Ambiguity: Who has the binocular- me or him?
  - The car hit the pole while it was moving.(Ambiguity: which is moving? Either car or pole?)

3. Referential Ambiguity

- Referential Ambiguity exists when you are referring to something using the pronoun.
- **Example:**
  - Lasya went to Sunita. She said, "I am hungry."(Ambiguity: Who is hungry? Either Lasya or Sunita)
  - The boy told his father the theft. He was very upset (Who is upset here? Boy or father or thief)
  - She is a big YouTube star.(Ambiguity: big means famous or fat?

4.4 MACHINE TRANSLATION (MT)

- Machine translation is an automatic translation from one natural language (the source) to another (the target).
- The input source can be text or speech.



- There are four types of machine translation.
    1. Statistical Machine Translation
    2. Rule-based Machine Translation
    3. Hybrid Machine Translation
    4. Neural Machine Translation

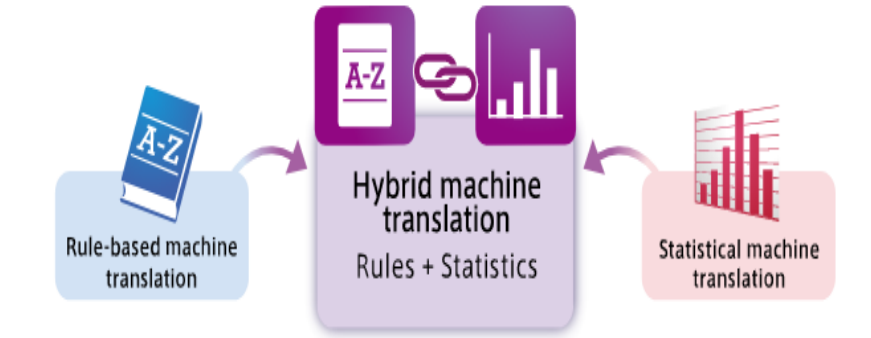
1. Statistical Machine Translation

- This works by referring to statistical models that are based on the analysis of large volumes of bilingual text.
- It aims to determine the correspondence between a word from the source language and a word from the target language.
- Its greatest drawback is that translations can often be incorrect.
- High-quality translations may not be possible in few cases.

2. Rule based Machine Translation

- This translates on the basis of grammatical rules.
- This is an example of Dictionary based machine translation.
- It conducts a grammatical analysis of the source language and the target language to generate the translated sentence.
- However, this requires extensive proofreading, and its efficiency is achieved after a long period of time.

3. Hybrid Machine Translation

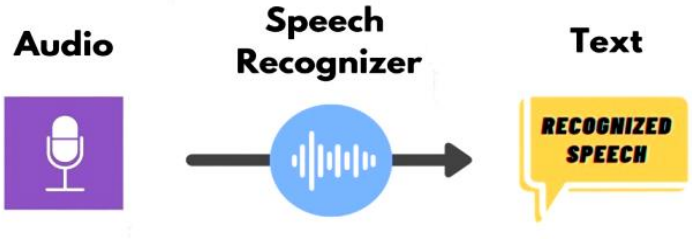


- This is a blend of statistical and rule based machine translation.
- This is more effective in terms of quality.
- However, it requires extensive editing. Human translators will be required.

4. Neural Machine Translation

- This depends on neural network models (based on the human brain) to develop statistical models for the purpose of translation.
- The primary benefit is that it provides a single system that can be trained to decode the source and target text.

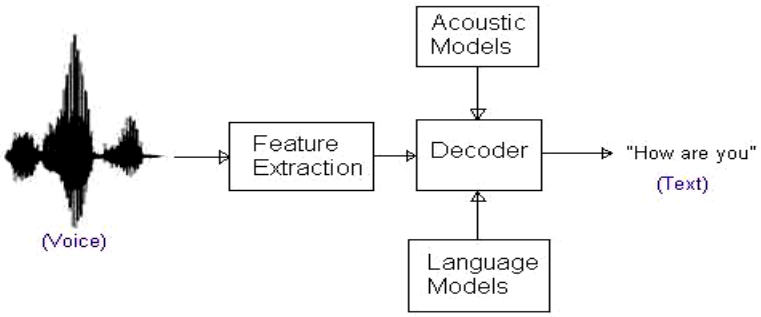
4.5 SPEECH RECOGNITION /VOICE RECOGNITION



- Speech recognition converts an acoustic signal to a string of words.
- It converts spoken words to machine readable input.
- Speech recognition is a technology which enables a machine to understand the spoken language and translate into a machine-readable format.
- It is a way to talk with a computer, and on the basis of that command, a computer can perform a specific task.
- It includes **Speech to text, Text to speech.**
- Speech is an attractive option when hands-free operation is necessary, as when operating machinery.
- Speech recognition is difficult because the sounds made by a speaker are ambiguous and noisy.
- Words like “to,” “too,” and “two” that sound the same but differ in meaning.

How Speech Recognition Works?

- Automatic Speech Recognition pipeline consists of the following blocks.



Speech Recognition System

1. **Feature extraction:** Audio signal preprocessing using normalization, spectrogram.

2. **Acoustic Model:** It predicts the probability distributions over vocabulary characters per each time step.
3. **Language model:** This is used to add context and correct mistakes in the acoustic model. It tries to determine what was spoken by combining both what the acoustic model thinks it heard with what a likely next word is.

- The most likely sequence can be computed with the help of **Navie Bayes' rule**:
- **Word Sequence** **Acoustic Model** **Language Model**

$$\underset{word_{1:t}}{argmax}\ P(word_{1:t} \mid sound_{1:t}) = \underset{word_{1:t}}{argmax}\ P(sound_{1:t} \mid word_{1:t}) P(word_{1:t})$$

  - Where $P(sound_{1:t} \mid word_{1:t})$ is the **acoustic model**. It describes the sounds of words—that **"ceiling"** begins with a soft "c" and sounds the same as **"sealing."**
  - $P(word_{1:t})$ is known as the **language model**. It specifies the prior probability of each sound—for example, that **"ceiling fan"** is about 500 times more likely as a word sequence than **"sealing fan."**
  - This approach was named as **noisy channel model.**
  - Argmax is an operation that finds the argument that gives the maximum value from a target function.
  - Argmax is most commonly used in machine learning for finding the class with the largest predicted probability.
- Once the acoustic and language models were defined, then prediction of sequence of words is very easy.
- Most speech recognition systems use a language model that makes the Markov assumption—that the current state $\boldsymbol{Word_t}$ depends only on a fixed number n of previous states—and represent $\boldsymbol{Word_t}$
- This is known as **Hidden Markov Model (HMM)**
- **HMM - Hidden Markov Model is a probabilistic graphical model that allows predicting a sequence of unknown (hidden) variables from a set of observed variables.**

4.6 TERMINOLOGIES/KEY TERMS RELATED TO NLP

1. Phonology

It is a field that studies how sound is organized systematically.

2. Morphology

It is a field that studies how meaningful words are constructed from primitive units.

3. Syntax

Syntax refers to the set of rules which defines how the words can be organized to form the sentences and phrases.

4. Semantics

Semantics is a branch that focuses on the meaning of words and sentences. It deals with how to combine words so that it converts into meaningful phrases and sentences.

5. Pragmatics

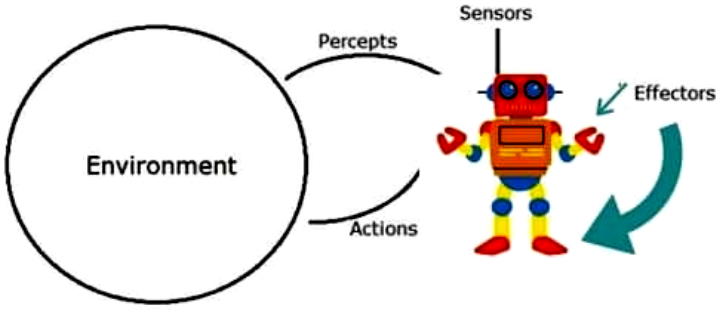
It is the branch that deals with how the sentences should be interpreted according to different situations.

4.7 CHALLENGES OF NLP

- NLP is one of the most difficult problems in Computer Science.
- Understanding what the humans are saying is a very complex and difficult task for machines.
- Humans are capable of using a number of languages and a wide range of words.
- It is difficult for computers to understand so many words and other rules for delivering a meaningful message.
- Whenever a paragraph from any language is used, number of ambiguities arises in the implementation of NLP by machines.
- In such situations, it becomes difficult for NLP to decode the exact meaning of what a person is trying to say.

PERCEPTION

4.8 MACHINE PERCEPTION



Agent and Environment

- Perception is the process of acquiring, interpreting, selecting and organizing sensory information.
- Perception provides the information about the world to the agents.
- Machine perception is the ability to use input from sensors such as cameras visible spectrum or infrared, microphones, wireless signals, active sonar, radar etc., to analyze the real aspects of the World.
- **Types of Perception**

  Hearing Perception – Speech Recognition

  Vision Perception –Image or Object Recognition, Facial Recognition Computer Vision, Robot Vision
- **VISION Perception**

  In this Vision, the sensor model can be broken into two components:
  - Object model
  - Rendering model

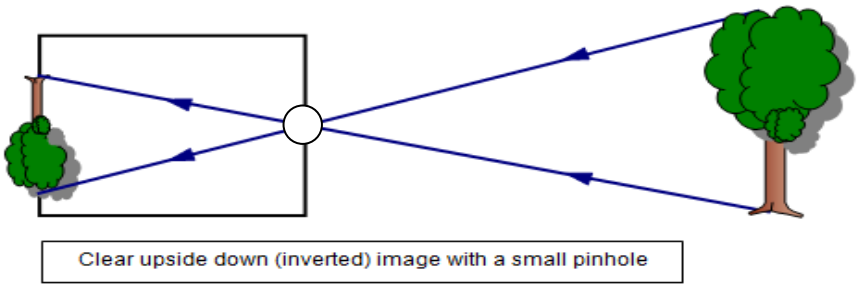
  - **Object Model -**
  - It describes the objects that reside in the visual world such as people, buildings, trees, cars, etc.
  - The object model could include a precise 3D geometric model taken from a computer-aided design (CAD) system.

  - **Rendering Model -**
  - It describes the physical, geometric, and statistical processes from the world.
  - Rendering models are quite accurate, but they are ambiguous.
  - For example, a white object under low light may appear as a black object.
  - A small nearby object may look as a large object.

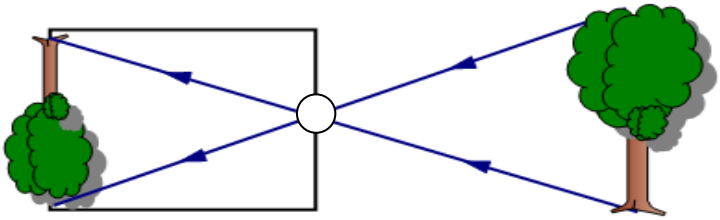
4.9 IMAGE FORMATION

- Image formation is a physical process that captures object in the scene through lens and creates a 2-D image.
- Imaging may misrepresent the appearance of objects.
  For example, if you hold your hand in front of your eye, you can block out the moon, which is not smaller than your hand.
- **Computer vision** aims at giving computers the ability to understand the environment as we do.
- It focuses on looking at the world through multiple images or videos and reconstructing properties like the shape of objects, intensity, color distributions, etc..
- Image formation also depends on discrete color and intensity values. It needs to know the lighting of the environment, camera optics, sensor properties, etc.,
- The following aspects are required to understand the Image Formation.
  1. ***Images without lenses: The pinhole camera***
  2. ***Lens System***
  3. ***Scaled orthographic projection***
  4. ***Lighting and Shading***
  5. ***Color***

4.9.1 Images without lenses: The pinhole camera



Clear upside down (inverted) image with a small pinhole

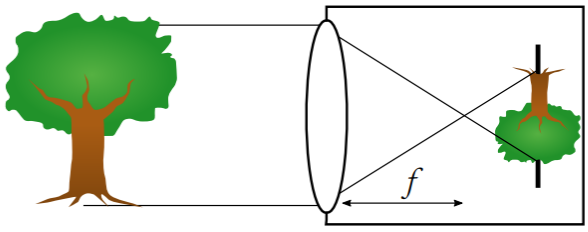


Camera closer to the object – a larger image

- The pinhole camera Image sensors gather light that are scattered from objects in a scene and create a 2D image.
- Through a pinhole camera, we can view stationary objects which help to form a focused image.
- It consists of a pinhole opening, O, at the front of a box and an image plane at the back of the box.
- The main drawback of the pinhole camera is that, a small pinhole is needed to keep the image in focus.
    - If the pinhole is smaller, the image will be dark, due to less number of photons entering the box.
    - If the pin hole is bigger, then image will be a blurred image due to the appearance of photons at multiple locations.

4.9.2 Lens System

- Modern cameras use a lens system to gather sufficient light while keeping the image in focus.
- A large opening is covered with a lens that focuses light from nearby object locations down to nearby locations in the image plane.
- However, lens systems can focus light only from points that lie within a range.
- Objects outside this range will be out of focus in the image.



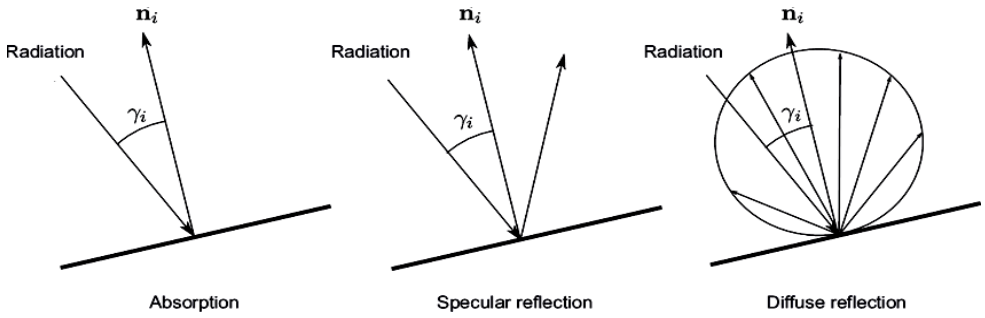
camera with lens to focus image

4.9.3 Scaled orthographic projection

- Position effects aren't always well-defined.
- For example, spots on a distant leopard may look small because the leopard is far away.
- Scaled orthographic projection is an approximation that is valid only for those parts of the scene with not much internal depth variation.
- Scaled orthographic projection can be a good model for the features on the front of a distant building.

4.9.4 Lighting and Shading

- Images cannot exist without light. Light sources can be a point or an area light source.
- When the light hits a surface, three major reactions might occur.
    1. **Absorption** - Some light is absorbed.
    2. **Specular reflection** – Some light is reflected speularly.
       E.g., mirror, polished metals like Aluminum or foil, wet floor
    3. **Diffuse reflection**- Some light is reflected diffusively.
       E.g., cloth, paints, wooden surfaces



4.9.5 Color

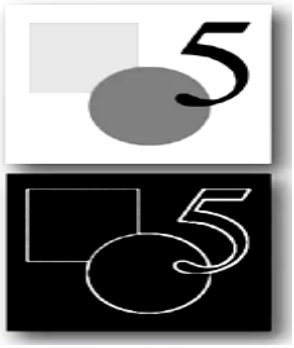
- Light arriving at the eye has different amounts of energy at different wavelengths.
- From a viewpoint of color, visible light is only a small portion of a large electromagnetic spectrum.
- Two factors are noticed when a colored light arrives at a sensor:
    - Colour of the light
    - Colour of the surface
- In a camera, not every sensor captures all the three components (RGB-Rred, Green and Blue) of light.
- Inspired by human visual preceptors, a colour grid may contain 50% Green, 25 % Red, and 25% Blue sensors.
- Quite accurate color constancy algorithms are now available.
  Example: “auto white balance” feature in camera.

4.10 EARLY IMAGE-PROCESSING OPERATIONS

- Image processing is manipulating an image in order to enhance it or extract information from it.
- There are two methods of Image Processing
  - **Analog Image Processing** – This is used processing physical photographs, printouts and other hard copies of images.
  - **Digital Image Processing** – This is used for manipulating digital images with the help of computer algorithms.
- The main image processing operations are
  1. ***Edge detection***
  2. ***Texture analysis***
  3. ***Optical flow***
  4. ***Segmentation of images***

4.10.1 Edge detection

- Edge detection is an image processing technique for finding the boundaries of objects within images.
- Edges are straight lines or curves in the image plane across which there is a "significant" change in image brightness.
- Edge detection is used in image analysis.
- It locates areas with strong intensity contrasts.
- The main goal of edge detection is to construct the ideal outline of an image.

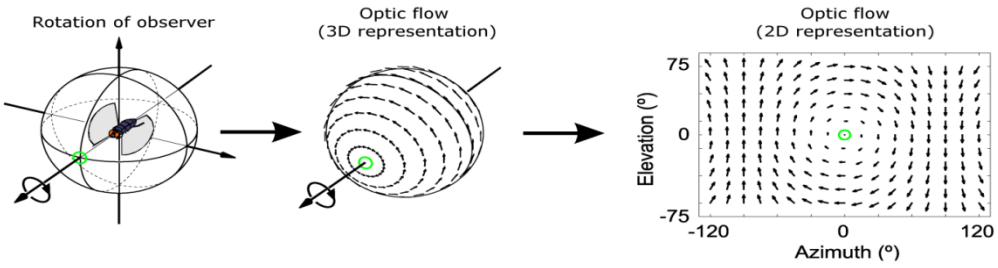


4.10.2 Texture

- Texture is the visual feel of a surface.
- In computational vision, texture refers to a spatially repeating pattern on a surface that can be sensed visually.
- Image texture gives the information about the spatial arrangement of colour or intensities in an image.
- **Examples:** Pattern of windows on a building, stitches on a sweater, spots on a leopard, blades of grass on a lawn, pebbles on a beach, and people in a stadium.
- Texture is used to estimate distances and for segmenting objects.

4.10.3 Optical flow

- When an object in the video is moving, the resulting apparent motion in the image is called **optical flow**.
- Optical flow describes the direction and speed of motion of features in the image.
- The optical flow of a video of a race car would be measured in pixels per second, not miles per hour.
- The optical flow encodes useful information about scene structure.
- For example, in a video of scenery taken from a moving train, far-away objects have slower apparent motion than close objects.
- The rate of apparent motion can tell us something about distance.



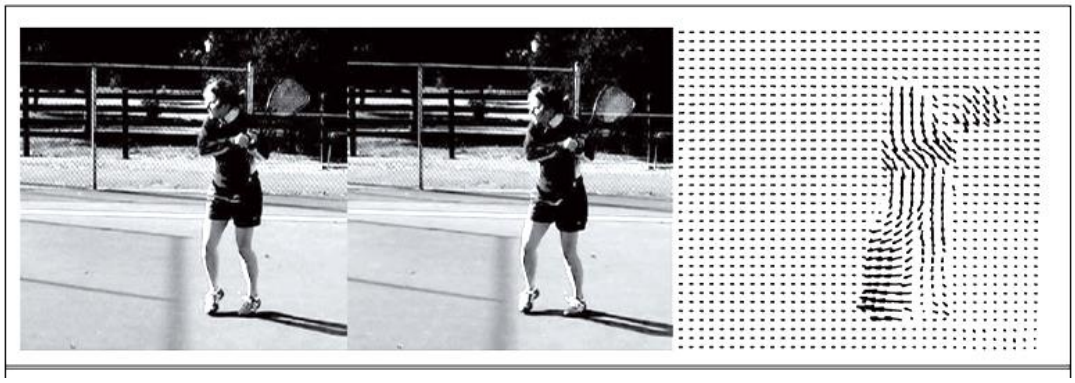
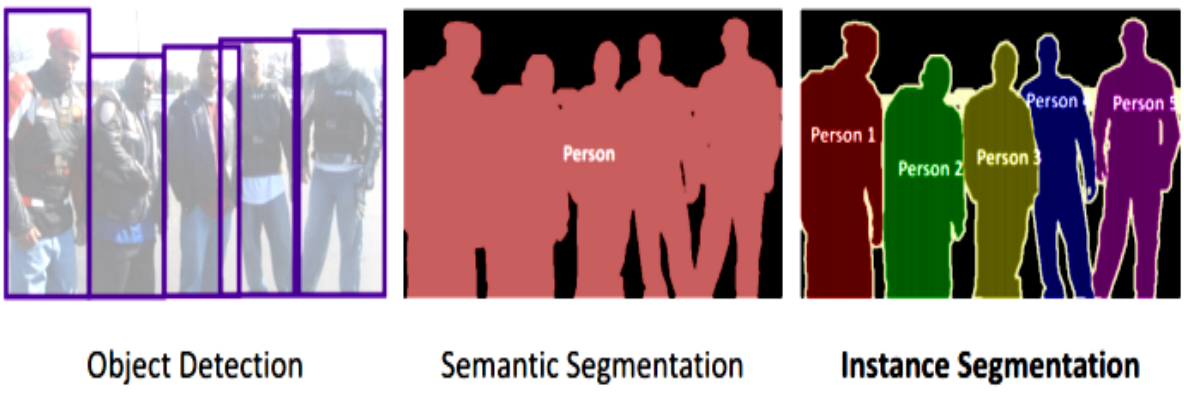


Fig: Two frames of video sequence. Right side. The movement of the tennis racket and the front leg is captured by the directions of the arrows.

4.10.4 Segmentation of Images



- Segmentation is the process of breaking an image into regions of similar pixels.
- Each image pixel can be associated with certain visual properties, such as brightness, color, and texture.
- Image segmentation focuses on separating an image into different parts according to their features and properties.
- The primary goal is to simplify the image for easier analysis.
- In image segmentation, an image will be divided into various parts that have similar attributes.
- The parts which are divided in the image are called Image Objects.

4.11 OBJECT RECOGNITION BY APPEARANCE

- Appearance means what an object tends to look like.
- For example, football is rather round in shape.
- It is important to know every class of images with a classifier.
- A house can have different size, color, and shape and can look different from different angles.
- A dancer looks different in each pose, or when the stage lights change colors.
- Training data is quite easily obtained. Testing each class of images with a learned classifier is important.

4.11.1 Complex appearance and pattern elements

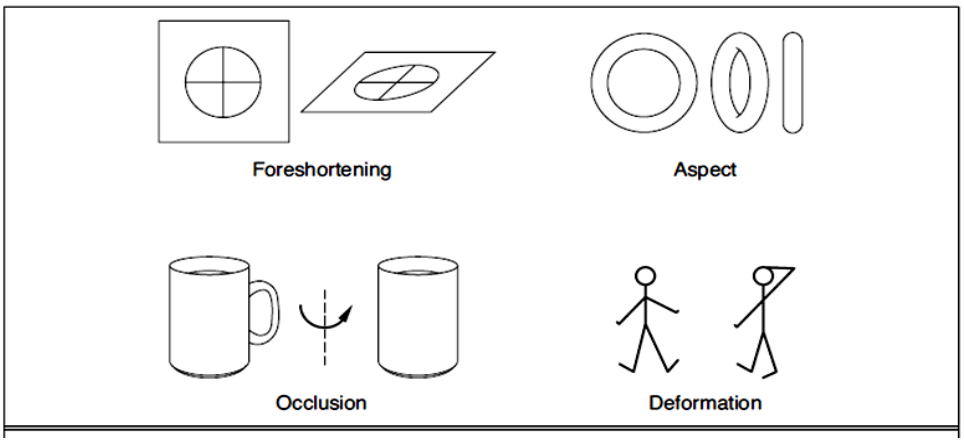
- Many objects produce much more complex patterns than faces do.
- This is because several effects can move features in the object.
- For example, a picture of a car is likely to show some of headlights, doors, wheels, windows, though they may be in somewhat different arrangements in different pictures.
- This suggests modeling objects with pattern elements (collections of parts).
- An object recognizer is required, which can identify various features of an object.
- Some of the Effects which may change the features of an object include:

  a. **Foreshortening** - It causes a pattern viewed at a slant to be significantly distorted.

  b. **Aspect** – It causes objects to look different when seen from different directions.

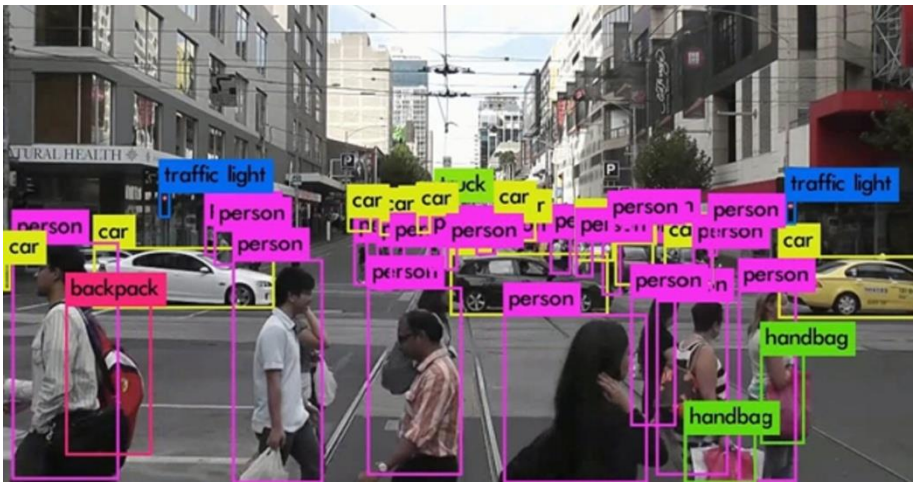
  c. **Occlusion(Closure)** – Some of the parts are hidden from some viewing directions.

d. **Deformation** -Internal degrees of freedom of the object change its appearance. For example, people can move their arms and legs around, generating a very wide range of different body configurations.

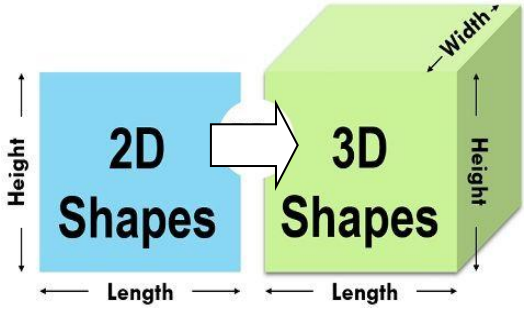


4.11.2 Pedestrian detection with HOG (Histogram Of Gradient orientations) features

- *HOG is a feature description used in computer vision and image processing for the purpose of object detection.*
- In self driving cars application, detecting pedestrians is an important aspect.
- Pedestrians wear many different kinds of clothing and appear in many different configurations, but, at relatively low resolution.
- The most usual cases are frontal views of a walk. In these cases, we see either a “lollipop” shape” or a “scissor” shape.
- We expect to see some evidence of arms and legs, and the curve around the shoulders and head also tends to visible and quite distinctive.
- This means that, with a careful feature construction, we can build a useful moving-window pedestrian detector.



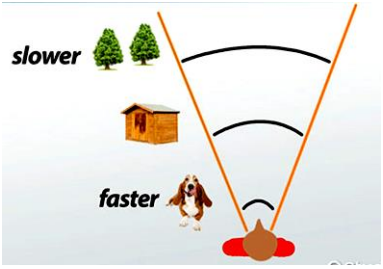
4.12 RECONSTRUCTING THE 3D WORLD



- In computer vision and computer graphics, 3D reconstruction is the process of capturing the shape and appearance of real objects.
- Using 3D reconstruction one can determine any object's 3D profile, as well as the 3D coordinates of any point on the profile.
- Some of the applications include Computer Aided Geometric Design (CAGD), Computer Graphics, Computer Animation, Computer Vision, Medical Imaging, Computational Science, Virtual Reality, Digital Media etc.,
- In 3D reconstruction, the input data can be acquired by several means, such as image cameras or scanner devices.
- 3D Computer Vision algorithms require considerable computational and imaging resources.
- It also requires suitable hardware, software, processing speed, accuracy and practicality of the results.
- Following are the main visual cues/ effects in 3D World.
    1. ***Motion Parallax***
    2. ***Binocular stereopsis.***
    3. ***Multiple views***
    4. ***Texture***
    5. ***Shading Contour***

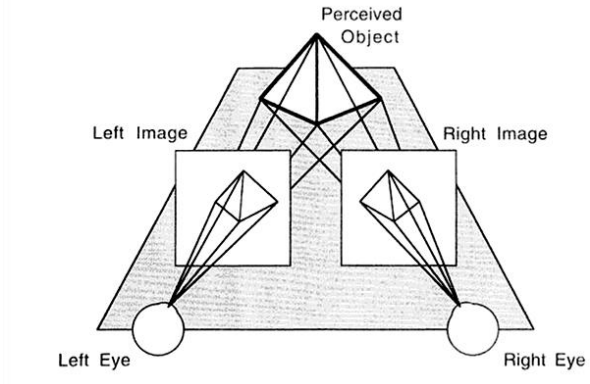
4.12.1 Motion Parallax

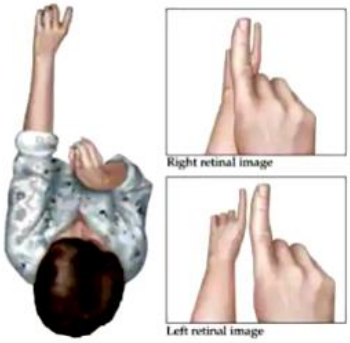
- Motion parallax refers to the fact that objects moving at a constant speed across the frame will appear to move a greater amount if they are closer to an observer (or camera) than they would if they were at a greater distance.
- Example 1: While riding in a car, objects that are close seem to move faster, but objects that are far-away appear to move much more slowly.
  Example 2: A video of scenery taken from a moving train, far-away objects have slower apparent motion than close objects.



4.12.2 Binocular stereopsis

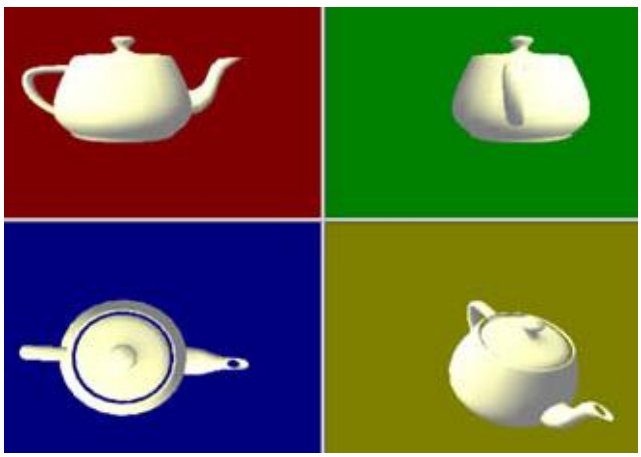
- It is the ability of both eyes to see the same object as one image and to create a perception of depth.
- It is a measure of binocular visual function, i.e. how well both eyes work together.
- Human eyes are separated horizontally by approximately 6.3cm.
- Existence of different retinal images leads to binocular disparity.

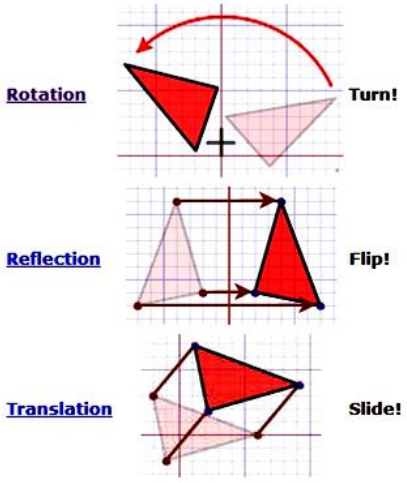




4.12.3 Multiple Views

- Multiple visualizations allow the user to explore large amounts of complex information more easily and rapidly.
- Following are the main issues in analyzing multiple views.
  - The correspondence problem- identifying features in the different images that are projections of the same feature in the 3D world.
  - The relative orientation problem - determining the transformation (rotation, reflection and translation) between the coordinate systems fixed to the different cameras.
  - The depth estimation problem- determining the depths of various points in the world.



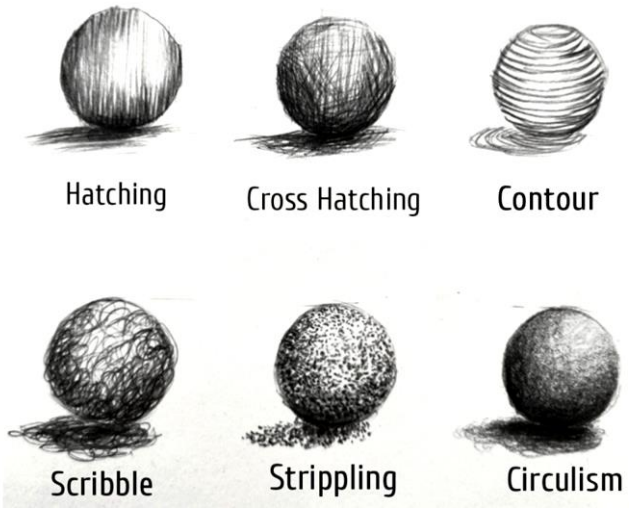


4.12.4. Texture

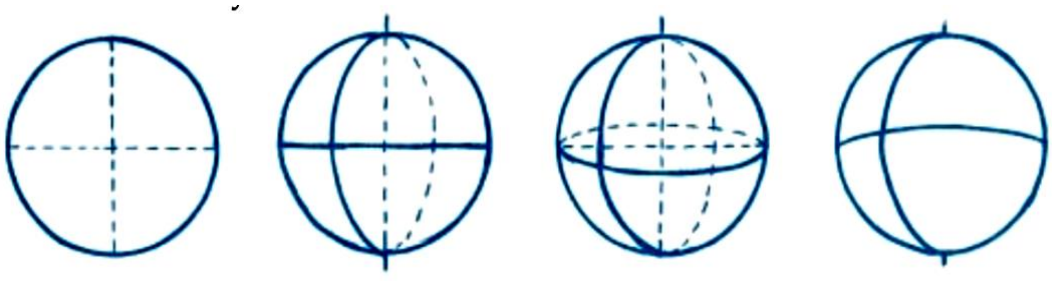
- Texture is the visual feel of a surface.
- In computational vision, texture refers to a spatially repeating pattern on a surface that can be sensed visually.
- Image texture gives the information about the spatial arrangement of colour or intensities in an image.
- Examples: Pattern of windows on a building, stitches on a sweater, spots on a leopard, blades of grass on a lawn, pebbles on a beach, and people in a stadium.
- Texture is used to estimate distances and for segmenting objects.

4.12.5 Shading

- Shading is the use of light and dark values to give the illusion of form and depth.
- Shading is a very slight variation, typically in colour.
- From different portions of a surface in a scene, receiving variation in the intensity of light called shading.
- Some of the shading techniques include hatching, cross-hatching, contour, scribble, circulism(smooth surface) and stripling(dots).



4.12.6 Contour

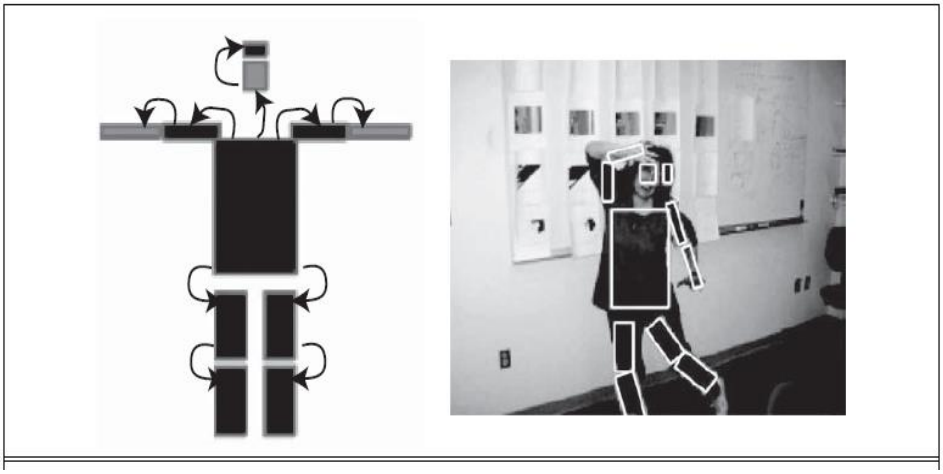


- Contour is a curve joining all the continuous points (along the boundary), having same color or intensity.
- Contour is a boundary around something that has well defined edges.
- The contour is a useful tool for shape analysis and object recognition and detection.
- These contours represent an abrupt change of the luminance, color, or texture in the image and characteristic features in the 3D information.
- There are infinitely many 3D interpretations of a 2D contour-drawing.

4.13 OBJECT RECOGNITION FROM STRUCTURAL INFORMATION

- Putting a box around pedestrians in an image may be enough to avoid accidents.
- Human body parts are very small in images and vary in color, texture among individuals.
- It is difficult to detect them using moving window method.
- Some images are very small to a size of two to three pixels wide.
- The configuration can be described by a model called **deformable template**.
- The simplest deformable template model of a person connects lower arms to upper arms, upper arms to torso (trunk of the human body), and so on.

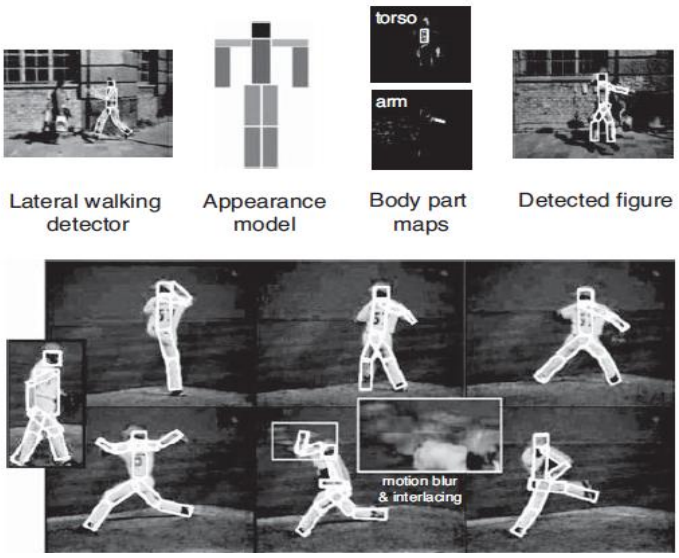
4.13.1 The geometry of bodies: finding arms and legs



- A tree of different segments is used to model the geometry of body.
- Segments are rectangular in shape.

- "Cardboard people" models are used to assume the position and pose (orientation) of human body parts and segments in the images.
- A rectangle in an image will look like its segment.
- It describes how well an image rectangle matches a body segment.
- This model is usually known as **pictorial structure model.**

4.13.2 Coherent appearance: tracking people in video



- There are several ways to develop a good **appearance model**.
- The video is a collection of large stack of pictures.
- Tracking people in video is an important practical problem.
- People in the videos tend to move fast and produce large accelerations.
- The people in the video can be tracked by detecting their body segments.
- The effective methods state the fact that, from frame to frame the appearances change.
- Concept of tracking people will be highly useful in surveillance systems.
- In some cases, the segment detector may generate false positives because the people are appearing against a near fixed background.
- This problem can be solved by applying a detector to a fixed body configuration for all the frames.
- This detector can identify a real person in a video with low false positive rate.

4.14 USING VISION

- **Computer vision** is a field of Artificial Intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs.
- It takes action or makes recommendations based on the visual inputs.
- If AI enables computers to think, computer vision enables them to see, observe and understand.
- Examples: Image classification, Object recognition, Object detection, Object tracking, Image retrieval.
- Vision technology is widely used in the following applications.
    1. ***Words and Pictures***
    2. ***Reconstruction from many views***
    3. ***Using vision for controlling movement (Automated Cars)***

- **1.Words and Pictures**
    - ✓ Many websites offer collections of images for viewing.
    - ✓ If the user enters a text query, such as "bike race." some of the images will appear which contains text near the image.
    - ✓ For these, **image retrieval** can be like text retrieval.
    - ✓ A picture of a dog playing in the garden might be tagged with words like "dog" and "garden".

- **2. Reconstruction from many views**
    - ✓ Reconstruction from multiple images is the creation of 3D models from a set of images.
    - ✓ A computer vision based reconstruction method allows the use of rich image resources from the internet.
    - ✓ **Example: Google Earth Studio** -The 3D scene reconstruction applications such as Google Earth allow people to take flight over entire metropolitan areas in a virtually real 3D world, explore 3D tours of buildings, cities and famous landmarks, as well as take a virtual walk around natural and cultural landmarks without having to be physically there.

- **3. Using vision for controlling movement (Self driving cars)**
  - ✓ The tasks in an automated vehicle include the following:
    - **i. Lateral control**—ensures that the vehicle remains within its lane or changes lanes smoothly when required.
      For lateral control edge-detection algorithms can be used to find edges corresponding to the lane-marker segments.
    - **ii. Longitudinal control**—maintains a safe distance with the vehicle in front.
      For longitudinal control, binocular stereopsis or optical flow can be used to know distances to the vehicles in front.
      Using these techniques, vision controlled cars can now drive reliably at highway speeds.
    - **ii. Obstacle avoidance**—monitors vehicles in neighboring lanes and be prepared if one of them decides to change lanes.
      For obstacle avoidance, two stereoscopic camera systems, one looking front and one looking back can be fixed for tracking the vehicles.

  - ✓ Using vision, appropriate steering, acceleration, and braking actions can be accomplished in autonomous cars.
  - ✓ The above driving example makes few points very clear:
    - For a specific task, there is no need to recover all the information from an image.
    - There is no need to recover the exact shape of every vehicle, solve for shape-from-texture on the grass surface adjacent to the road.
    - Instead, a vision system should compute just what is needed to achieve the task.

*****

"Computer Vision and Machine Learning have really started to take off, but for most people, the whole idea of what is computer seeing when it's looking at an image is relatively obscure"

- **Mike Krieger,** a Brazilian-American entrepreneur & software engineer who co-founded **Instagram** along with Kevin Systrom, and served as its CTO.

Question Bank

UNIT – IV: NATURAL LANGUAGE COMMUNICATION & PERCEPTION

2 Marks Questions

1. What is phase structure grammar?
2. What is Probabilistic Context Free Grammar(PCFG)?
3. Define parsing.
4. What is lexical analysis?
5. What is disclosure integration?
6. Compare syntactic analysis and semantic analysis.
7. Give the list of ambiguity in NLP.
8. Define machine translation.
9. Give the differences between statistical MT and rule based MT.
10. What is neural machine translation?
11. What are the two modes of speech recognition?
12. How speech recognition works?
13. Explain Hidden Markov's Model(HMM).
14. What are the challenges of NLP?
15. What is perception and machine perception?
16. Give differences between analog and digital image processing.
17. What is motion parallax?
18. Explain Binocular stereopsis.
19. Explain texture and shading.
20. What is contour in 3D World?

Essay Questions

1. With the help of a suitable example explain about phrase structure grammar and parse tree.
2. What is augmented grammars and semantic interpretation .Explain various phases.
3. Explain various types of machine translation methods.
4. With the help of acoustic and language models explain speech recognition.
5. Explain various aspects of image formation.
6. Explain various image processing operations.
7. Describe the concept of object recognition by appearance.
8. Briefly explain various phases of reconstructing the 3D World.
9. How objects are recognized from structural information. Explain.
10. Explain the concept of using vision and various applications.