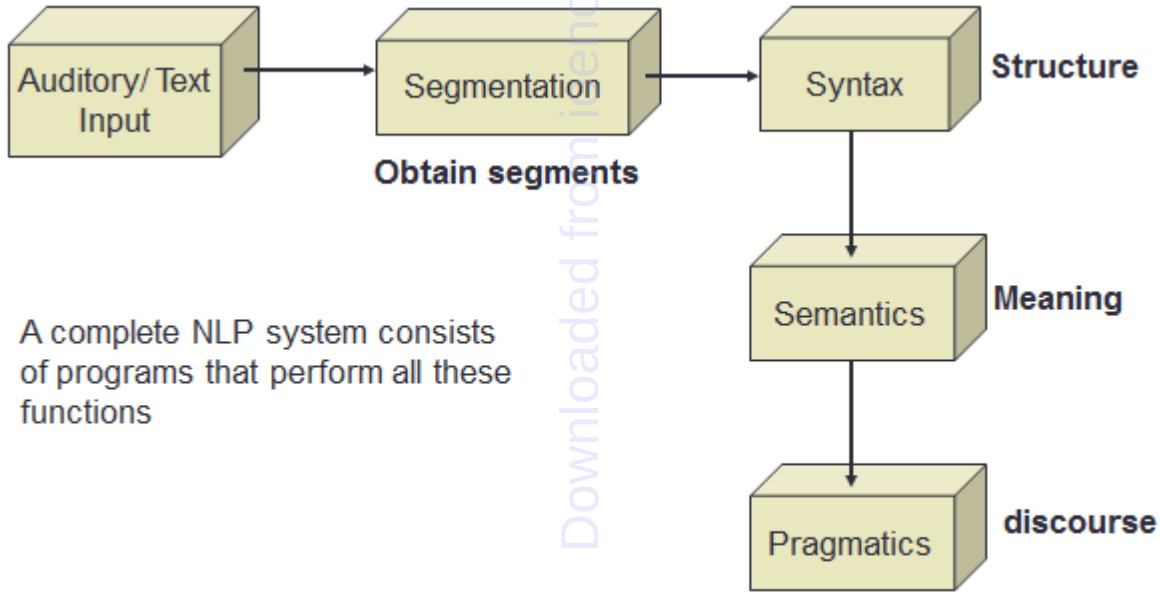
# Natural Language Processing

- Language is means of Communication for humans. By studying language, we come to understand more about the world.
- If we can succeed at building computational mode of language, we will have a powerful tool for communicating about the world.
- We look at how we can exploit knowledge about the world, in combination with linguistic facts, to build computational natural language systems.
- Natural Language Processing (NLP) is the process of computer analysis of input provided in a human language (natural language), and conversion of this input into a useful form of representation.
- NLP is one of field of AI that processes or analyzes written or spoken language.
- NLP involve processing of speech, grammar and meaning.
- NLP is composed of two part: NLU (Natural Language Understanding) and NLG (Natural Language generation).
- Understanding any language requires detailed knowledge of that language.

## NLP Steps/ Processes






**Input/Source**

- The input of a NLP system can be written text or speech.
- Quality of input decides the possible errors in language processing that is high quality input leads to correct language understanding.

**Segmentation**

- The text inputs are divided in segments (Chunks) and the segments are analyzed. Each such chunk is called frames.

**Syntactic Analysis**

- Syntactic analysis takes an input sentence and produces a representation of its grammatical structure.
- A grammar describes the valid parts of speech of a language and how to combine them into phrases.
- The grammar of English is nearly context free.

**Grammar**: A computer grammar specifies which sentences are in a language and their parse trees. A parse tree is a hierarchical structure that shows how the grammar applies to the input. Each level of the tree corresponds to the application of one grammar rule.

**Syntax**

1. sentence ↔ noun_phrase verb_phrase
2. noun_phrase ↔ noun
3. noun_phrase ↔ article noun
4. verb_phrase ↔ verb
5. verb_phrase ↔ verb noun_phrase
6. article ↔ a
7. article ↔ the
8. noun ↔ man
9. noun ↔ dog
10. verb ↔ likes
11. verb ↔ bites

**Parse tree**

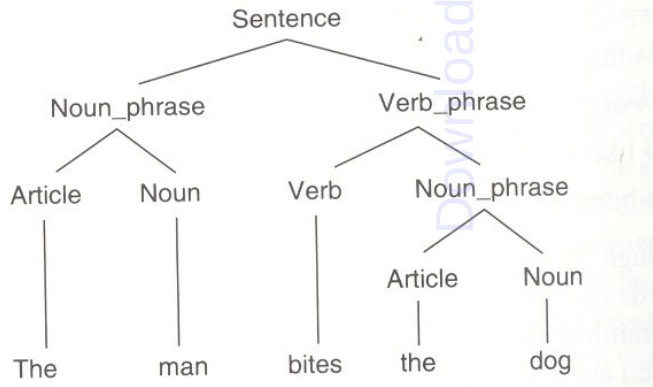



**Figure 10.3** Parse tree for the sentence "The man bites the dog."

**Semantic Analysis**

- Semantic analysis is a process of converting the syntactic representations into a meaning representation.
- This involves the following tasks:

- Word sense determination
- Sentence level analysis

**Word sense**: Words have different meanings in different contexts.
Example:*Mary had a bat in her office.*
bat = `a baseball thing'
bat = `a flying mammal'
**Sentence Level Meaning**
Once the words are understood, the sentence must be assigned some meaning
Examples:

- She saw her duck.
- I saw a man with a telescope.

Non-examples: Colorless green ideas sleep furiously - >This would be rejected semantically as *colorless green* would make no sense

**Pragmatic Analysis**

- Pragmatics comprises aspects of meaning that depend upon the context or upon facts about real world.
- These aspects include:

- Pronouns and referring expressions
- Logical inferences, that can be drawn from the meanings of a set of propositions.
- Discourse structure: the meaning of a collection of sentences taken together.

Examples

- **Jack fell. Jill brought him a band-aid.**
  Jack got hurt and Jill wanted to help.

- **We got seven letters today**
  We got only seven letters (and not eight).

- **Understanding toothpaste directions.**
  *For best results, squeeze tube from the bottom and flatten as you go up.*
  - The tube has an opening at only one end.
  - The tube can be squeezed,
  - The instructee wants to use the paste.
  - The tube can be emptied only through its mouth.
- Lexicon or dictionary
- Morphological Analysis system

**NLP AIDS**

- To understand or analyze language some aids or tools can be used in NLP

**i. Lexicon or Dictionary**

- A lexicon defines the words of a language that a system knows about.
- This includes common words and words that are specific to the domain of the application.
- Entries include meanings for each word and its syntactic and morphological behavior.

### ii. Morphological Analysis System

- Morphological analysis is the process of recognizing the suffixes and prefixes that have been attached to a word.
- We do this by having a table of affixes and trying to match the input as: prefixes+root+suffixes.
    - For example: adjective + ly -> adverb
    - We may not get a unique result.
    - "-s, -es" can be either a plural noun or a verb
    - "-d, -ed" can be either a past tense or a perfect participle
- **Morphological Information**

- Transform part of speech
    - *green, greenness (adjective, noun)*
    - *walk, walker (verb, noun)*
- Change features of nouns
    - *boat, boats (singular, plural)*
- Bill slept , Bill's bed
    - (subjective case, possessive case)
- Change features of verbs
    - Aspect
        - *I walk. I am walking.* (present, progressive)
    - Tense
        - *I walked. I will walk. I had been walking. (past, future, past progressive)*
    - Number and person
        - *I walk. They walk. (first person singular, third person plural)*

- **Problems**

- doubled consonants (sit, sitting)
- changing y to i (fly, flies)
- abbreviations (CAN0
- naming word (Cook)
- wrong spelled word (Hare instead hair)
- unrecognized words:
    - real words not in the dictionary
    - names, etc that cannot be listed (names, account nos., library call nos)

### iii. Syntactic analysis system

- Takes an i/p sentence and produces a representation of its grammatic structure.
- Grammar (mostly context free grammar) and parser are used.

### iv. Semantic Analysis System

- Convert the syntactic representation into a meaning representation.
- It involves:
    - Word sense determination
    - Sentence level analysis

- Knowledge Representation

**v. Pragmatic Analysis System**

- Study how knowledge about the world and language conventions interacts with the literal meaning.
- Depends upon context and facts about real world.
- These aspects includes:
  - Pronouns and referring expressions.
  - Logical interfaces that can be drawn from meaning of a set of propositions.
  - Inference about what the producer of a sentence is trying to do.

**Natural Language Analysis Techniques**

- Two commonly used techniques used in natural language processing are

**i. Keyword analysis or Pattern matching**

- Accept the given sentence as input.
- Segment the sentence.
- Identify keywords in each segment.
- If keyword is present
  - If only one keyword is present , give suitable reply as of keyword
  - If more the on keywords are present, prioritize them and give suitable reply as of keyword
- If no keyword present in the segment , give a random reply.

**ii. Syntactic driven parsing technique**

- Words can fit together higher level units such as phrases, clauses and sentences.
- Interpretations of larger groups of words are built up out of the interpretation of their syntactic constituent words or phrases.
- Interpretation i/p done as a whole.
- Obtained by application of grammar that determines what sentences are legal in the language that is being parsed.



## NLP Problems

1. The same expression means different things in different context.
   - Where's the water? ( Chemistry lab? Must be pure)
   - Where's the water? ( Thirsty? Must be drinking water)
   - Where's the water? ( Leaky roof? It can be dirty)
2. No natural language program can be complete because of new words, expression, and meaning can be generated quite freely.
   - I'll fax it to you
3. There are lots of ways to say the same thing.
   - Ram was born on October 11.
   - Ram's birthday is October 11.
4. Sentence and phrases might have hidden meanings
   - "Out of sight, out of mind"-> " invisible idiot"
   - "The spirit was willing but the flesh was weak" - > " the vodka was good, but the meat was bad"

5. Problem due to syntax and semantics
6. Problem due to extensive use of pronouns. (semantic issue)
    - Eg. Ravi went to the supermarket. He found his favorite brand of coffee in rack. He paid for it and left.
    - It denotes??
7. Use of grammatically incorrect sentence
    - He rice eats. (syntax issue)
8. Use of conjunctions to avoid repetition of phrases cause problem in NLP
    - Eg. Ram and Hari went to restaurant. While Ram had a cup of coffee, Hari had tea.
    - Hari had a cup of tea.

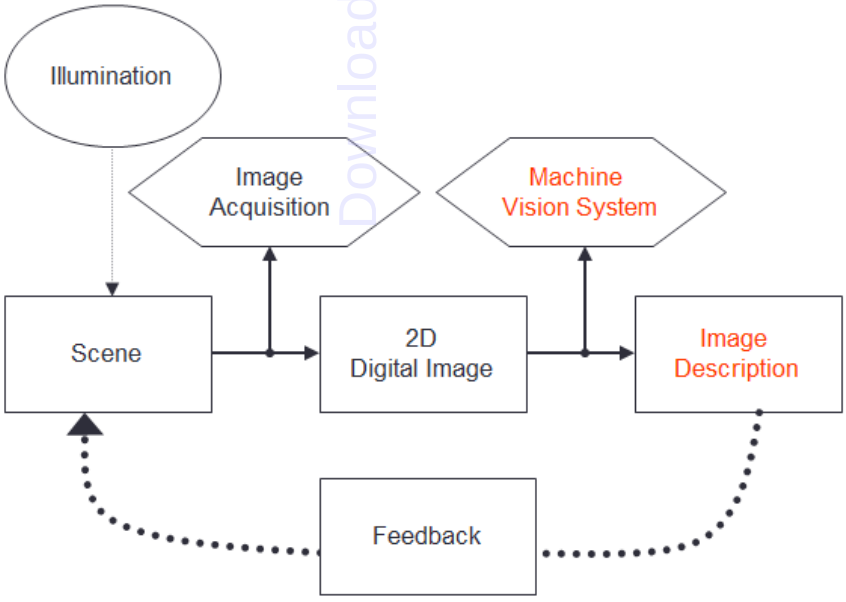


## Introduction to machine Vision

The goal of Machine Vision is to create a model of the real world from images

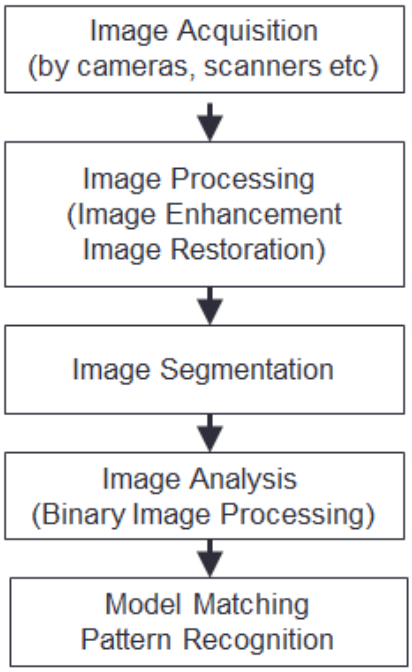
- A machine vision system recovers useful information about a scene from its two dimensional projections
- The world is three dimensional
- Two dimensional digitized image

Knowledge about the objects (regions) in a scene and projection geometry is required.The information which is recovered differs depending on the application such as Satellite, medical images etc.



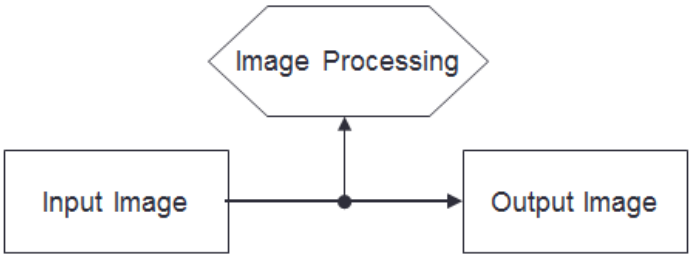

The goal of a machine vision system is to compute a meaningful description of the scene (e.g., object)

## Machine Vision Stages




- Analog to digital conversion
- Remove noise/patterns, improve contrast
- Find regions (objects) in the image
- Take measurements of objects/relationships
- Match the above description with similar description of known objects (models)

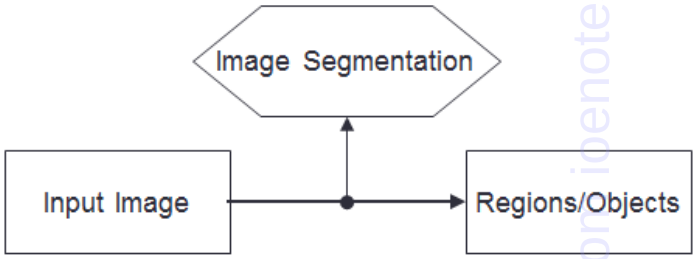
## Image Processing




Various procedures such as image enhancement (filtering, edge detection, surface detection, computation of depth) Image restoration (remove point/pattern degradation: there exist a mathematical expression of the type of degradation like e.g. Added multiplicative noise, sin/cos pattern degradation etc) takes place.



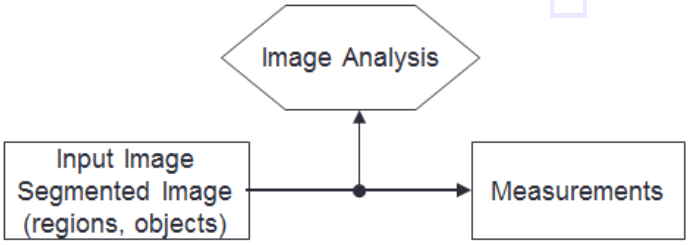
## Image Segmentation




Classify pixels into groups (regions/objects of interest) sharing common characteristics.

- Intensity/Color, texture, motion etc.

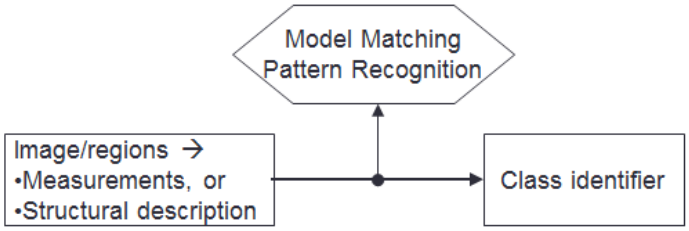
## Image Analysis




Take useful measurements from pixels, regions, spatial relationships, motion etc.

- Grey scale / color intensity values;
- Size, distance;
- Velocity;

## Pattern Recognition




Classify an image (region) into one of a number of known classes

- Statistical pattern recognition (the measurements form vectors which are classified into classes);
- Structural pattern recognition (decompose the image into primitive structures).



## Digital Image Representation

Image: 2D array of gray level or color values

- Pixel: array element;
- Pixel value: arithmetic value of gray level or color intensity.

Gray level image: $f = f(x,y)$

- 3D image $f=f(x,y,z)$

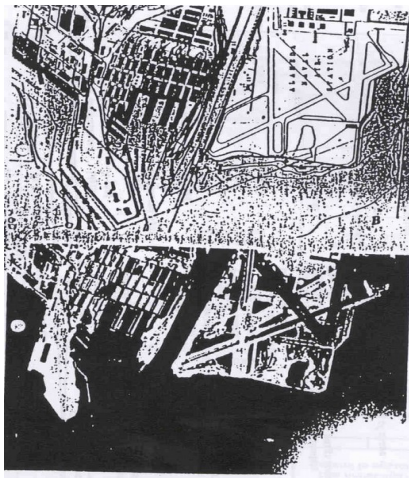
Color image (multi-spectral)

$f = \{R_{red}(x,y), G_{green}(x,y), B_{blue}(x,y)\}$

## Machine Vision Applications

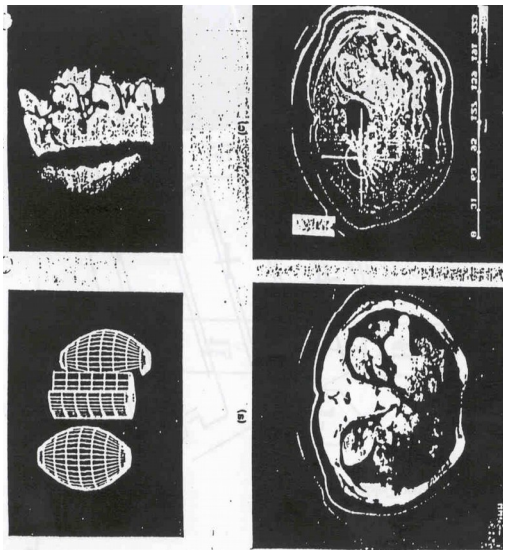
- Robotics
- Medicine
- Remote Sensing
- Meteorology
- Quality inspection

**Remote Sensing**



- Take images from high altitudes (from aircrafts, satellites).
- Find ships in the aerial image of the dock.
- Find if new ships have arrived.
- What kind of ships?

**Medical Applications**



- Assist a physician to reach a diagnosis.
- Construct 2D, 3D anatomy models of the human body.
- CG geometric models.
- Analyze the image to extract useful features.