INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY

HYDERABAD

# CSE578:
# Computer Vision

Professor : Anoop Namboodiri
Notes By : Sai Manaswini Reddy I

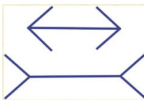
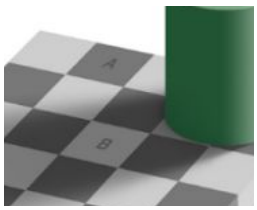
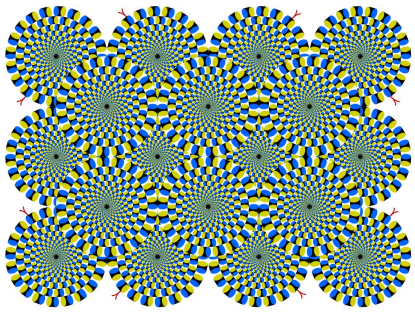
2021

## What is Computer Vision?

- Understanding of visual inputs (images/videos) by computers.
- Making sense out of them. Describing them.
- Does computer vision mimic the human vision?
    - Certainly in many of its goals
    - Why? Human vision is among the best!
    - Sophisticated and efficient but not understood well
- Should computers process visual inputs like humans?
    - Not necessarily!
    - Human visual system need not limit computer vision
    - We draw inspiration from it as often as is convenient
- Goal: Extract all possible information about a visual scene by computer processing
    - What? When? Where? Who? How? Why? How many?
- Over 50% of the brain is devoted to vision for humans.

    – Must be important to us!

## Shortcomings of Human Vision :

| | |
|---|---|
|  | |
|  | Line in the below one seems to be lengthy, in fact this statement is filthy |
|  | A seems to be darker which in fact is not true. |
|  | It is not the picture with motion, it is our brain that is causing this. |

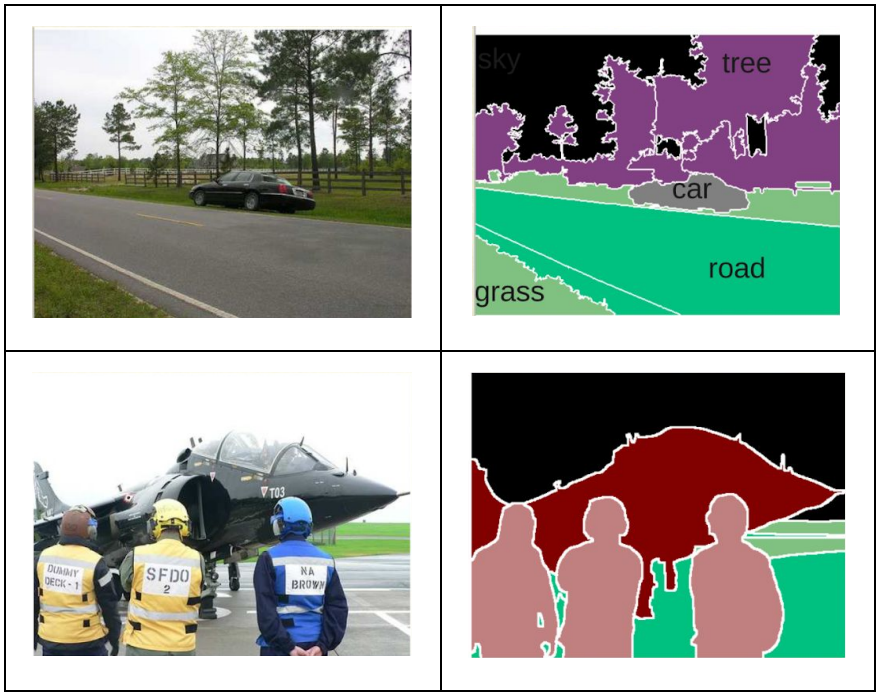
## Three “Urges” on seeing a Picture* :

1. Segmentation : To group proximate and similar parts of the image into meaningful “regions”.
2. Recognition : To connect to memory to recollect previously seen “objects” (brain automatically recognises)
3. reconstruction : To measure quantitative aspects such as number and sizes of objects, distances to/between them, etc.

## Urge to Group :

We don’t see individual pixels (like the computer does!).

- We see groups of pixels together. This is grouping. We label the pictures after grouping.
- What is the basis for “correct” grouping?
    - Group similar pixels together as objects.
    - Group semantically meaningful pixels together as objects.




## Urge to Touch Memory :

Recognizing objects from (visual) input is fundamental to human cognition of the world. We involuntarily recognise or try to recognise.

## Urge to Measure :

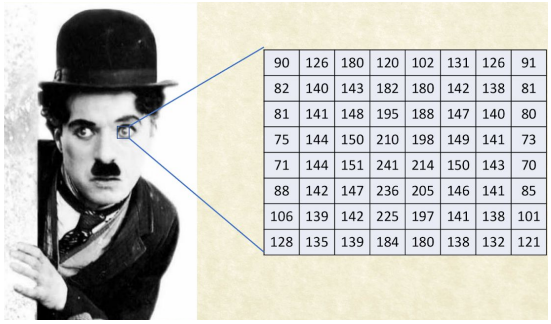
We measure number, distance, colour, etc, from images. Sometimes correctly too!

## Why is it difficult for computers to do these ?

We recognise the person in this image as soon as we look at the image involuntarily.

Computer will just look at it like an array of numbers. We can not recognise the object from the array of numbers.

Extra effort is needed to figure out to process this array of numbers and recognise it as an eye.

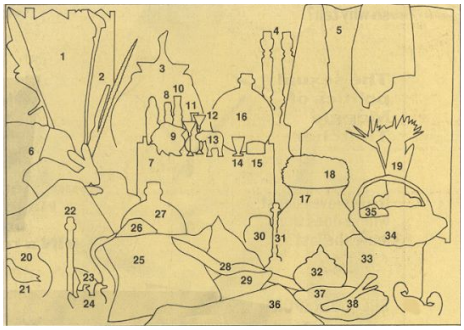


| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 90 | 126 | 180 | 120 | 102 | 131 | 126 | 91 |
| 82 | 140 | 143 | 182 | 180 | 142 | 138 | 81 |
| 81 | 141 | 148 | 195 | 188 | 147 | 140 | 80 |
| 75 | 144 | 150 | 210 | 198 | 149 | 141 | 73 |
| 71 | 144 | 151 | 241 | 214 | 150 | 143 | 70 |
| 88 | 142 | 147 | 236 | 205 | 146 | 141 | 85 |
| 106 | 139 | 142 | 225 | 197 | 141 | 138 | 101 |
| 128 | 135 | 139 | 184 | 180 | 138 | 132 | 121 |

Even for human vision it's very trivial to recognise all the objects.

Example :

Scene interpretation ; segmentation and labelling.



We can not recognise all these until and unless we are expertees in this.

## Examples :

Chair :

An object with a platform to sit, a backrest and four legs. But all the below images contradict.

There is a large amount of variation (intra class variation)

In our brain we have a property noted called sittability!

We can understand the world, but we do not understand how we understand. So it is difficult to teach computers about how we understand.

## Applications of Computer Vision :

- Medical :
    - CT scan
    - Computer assisted surgery
    - Segmentation
- Space imaging
- Automated inspection
    - PCB Inspection
- Biometrics
    - Iris
    - Finger print
    - Face
- Broad casting
    - Field understanding : virtual line
    - Chroma keying : replacing backgrounds
    - Ball Tracking : Hawk eye
    - Player Tracking
- 3D Shape and Motion Recovery
    - Structure light scanner, laser range finder
    - Multi-camera stereo, structure recovery
    - Reverse Engineering
    - Virtualized/Augmented reality
- Surveillance
- Automated Assembly
- Mail Sorting
- Face detection (photography)
- Robot Navigation
- Content-Based Image Retrieval
- Entertainment
    - Gaming : Natural Gaming
- Automotive Safety

Why Automated Vision?

1. High reliability
2. High reliability
3. More objective evaluation
4. Lower cost
5. Higher speed
6. Ability to operate in hazardous environments

Problem :

Finding out what is happening in the other person's brain. What we predict may be completely different from what is gonna happen.