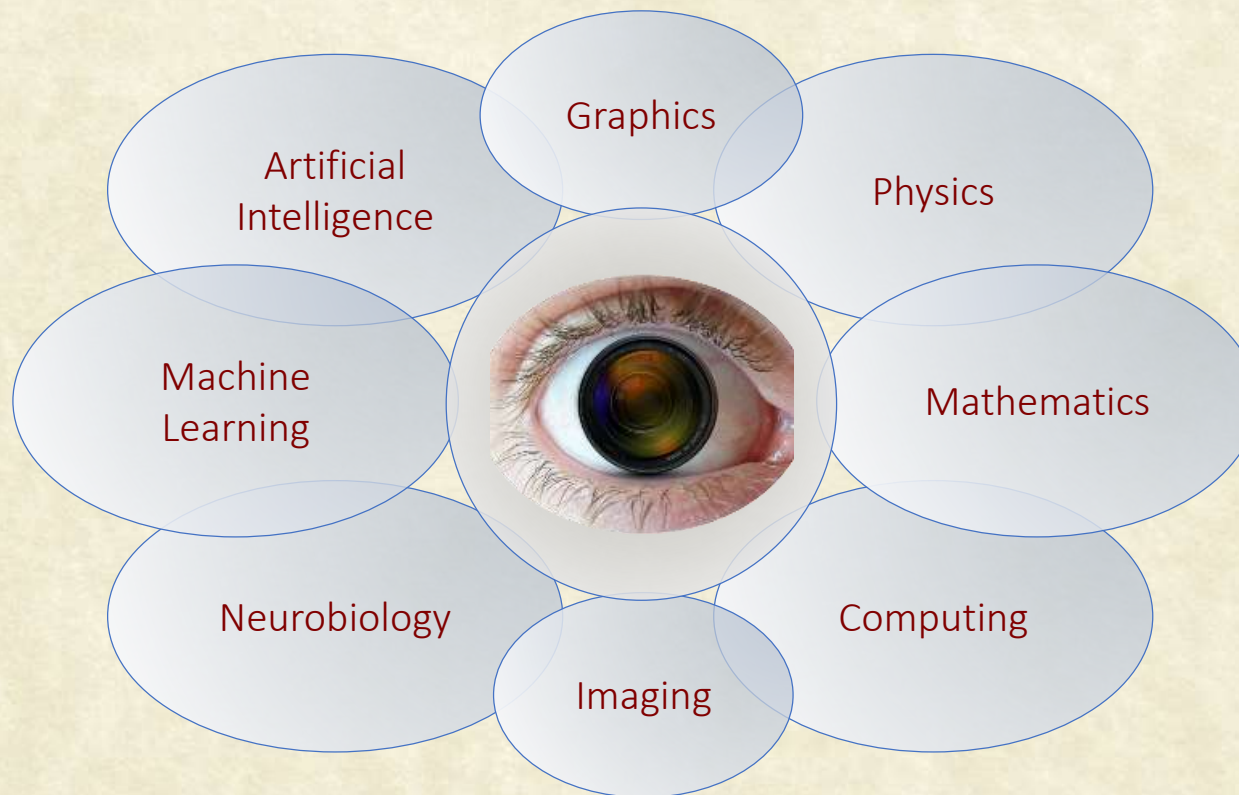




# CS7.505: Computer Vision

Spring 2025: Introduction



Makarand Tapaswi  
CVIT, IIIT Hyderabad

Ravi Kiran Sarvadevabhatla  
CVIT, IIIT Hyderabad



# What is Computer Vision?

Understanding of visual inputs (images/videos) by computers.







# What is Computer Vision?

A field trying to develop computer algorithms that “see”

- What does it mean to “see”?
  - To know what is where, by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D / material properties?
  - What actions are taking place?



# Not a summer project

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
PROJECT MAC

Artificial Intelligence Group  
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

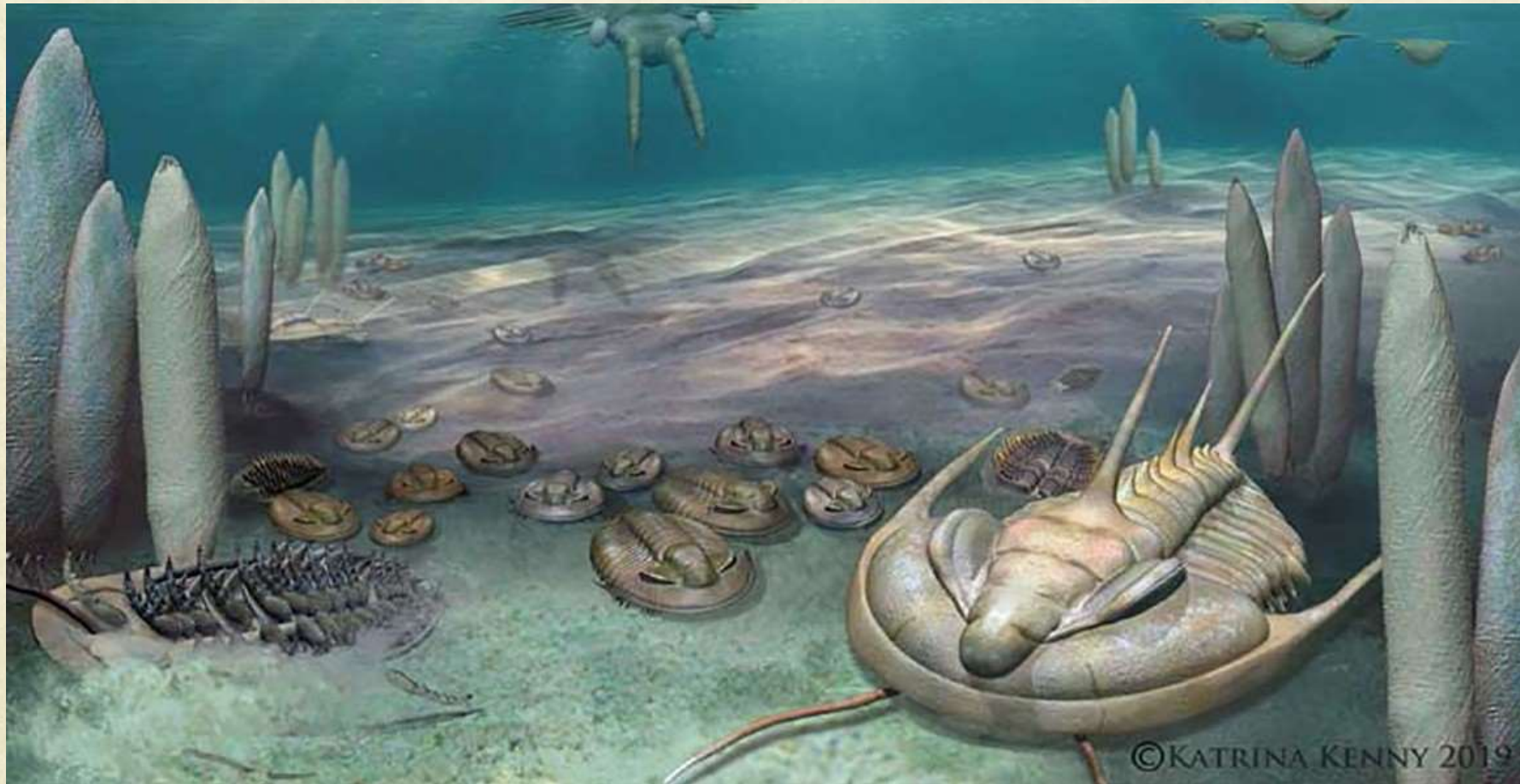
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".





# Cambrian Explosion

- Vision, from about 540 million years ago (although debated among historians 😊)





# Does computer vision mimic human vision?

- Certainly, in many of its goals
- Why? Human vision is among the best! Sophisticated and efficient but not fully understood
- Should computers process visual inputs like humans?

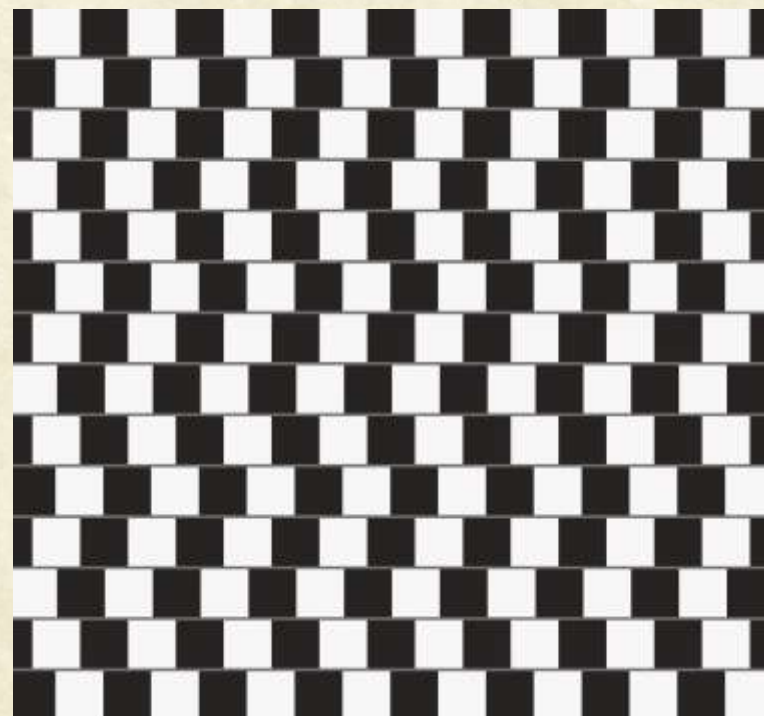
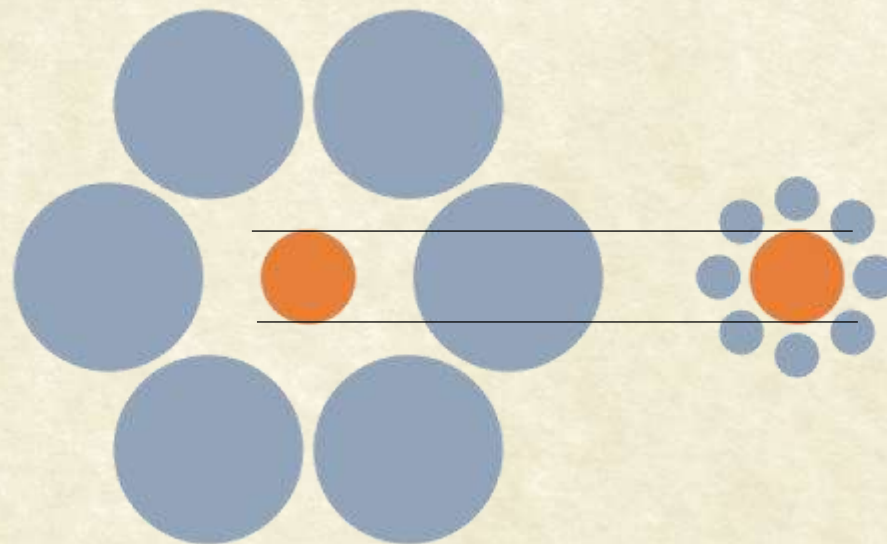
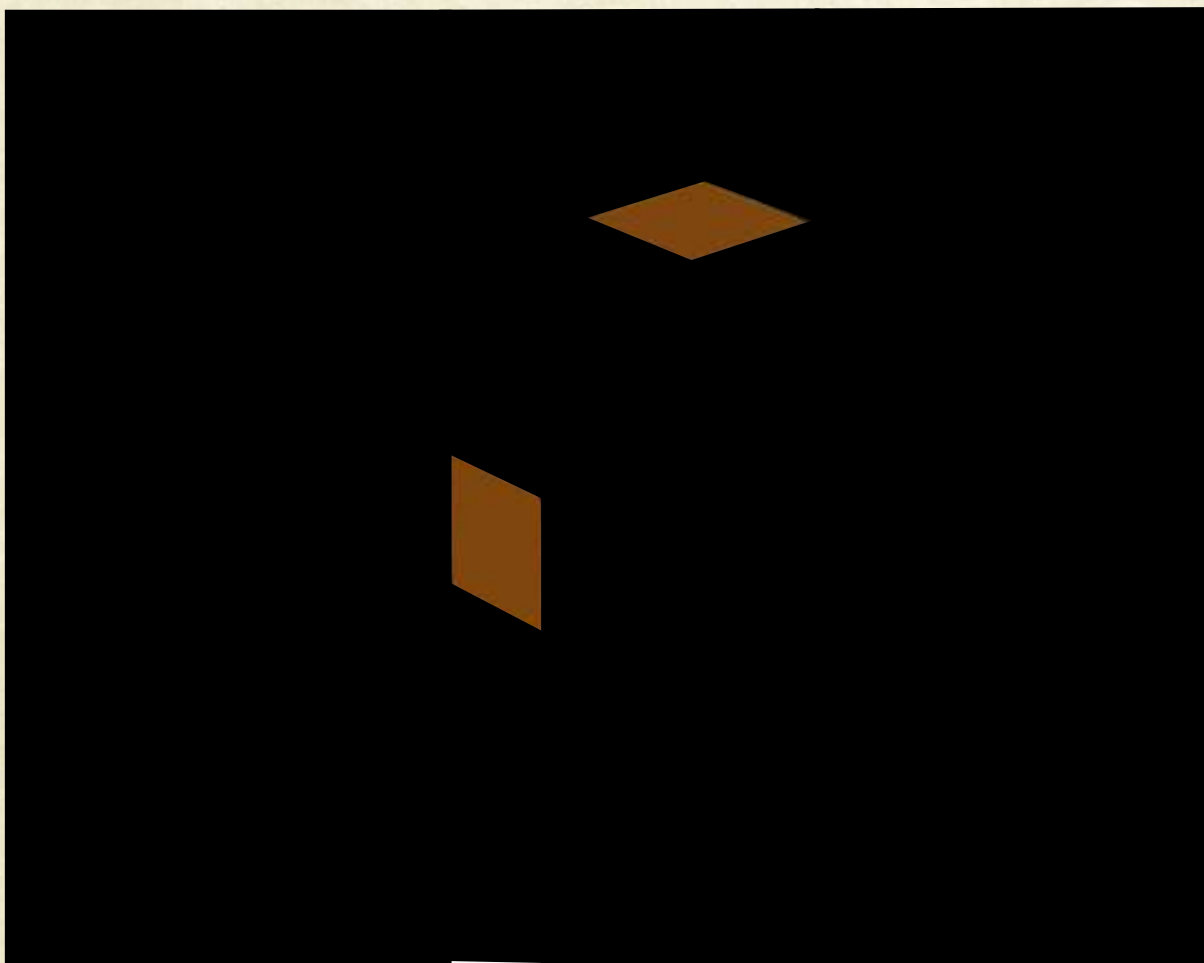
**Not necessarily!**

- Human visual system need not limit computer vision
- We draw inspiration from it as often as is convenient

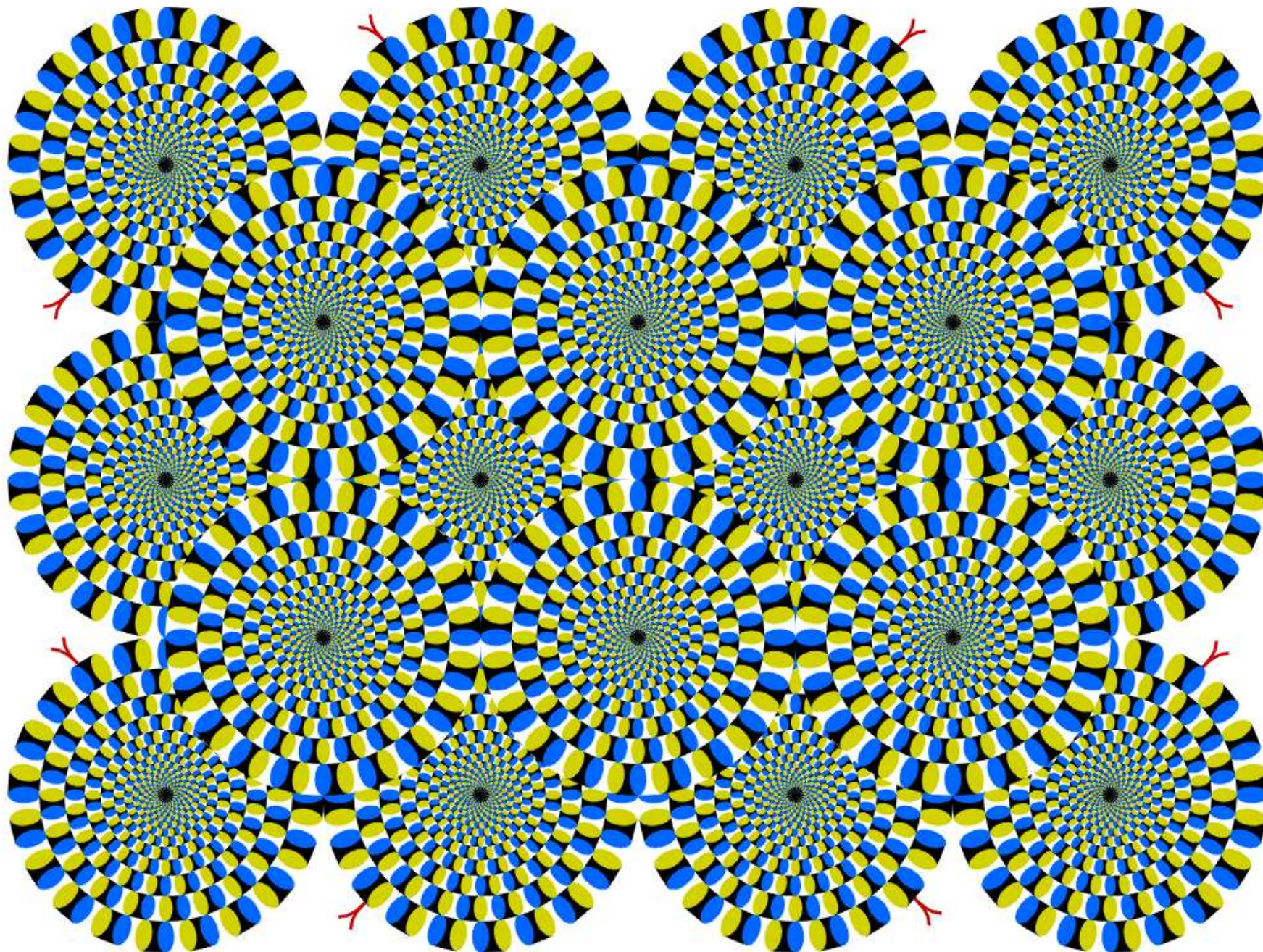




Human perception is not perfect...











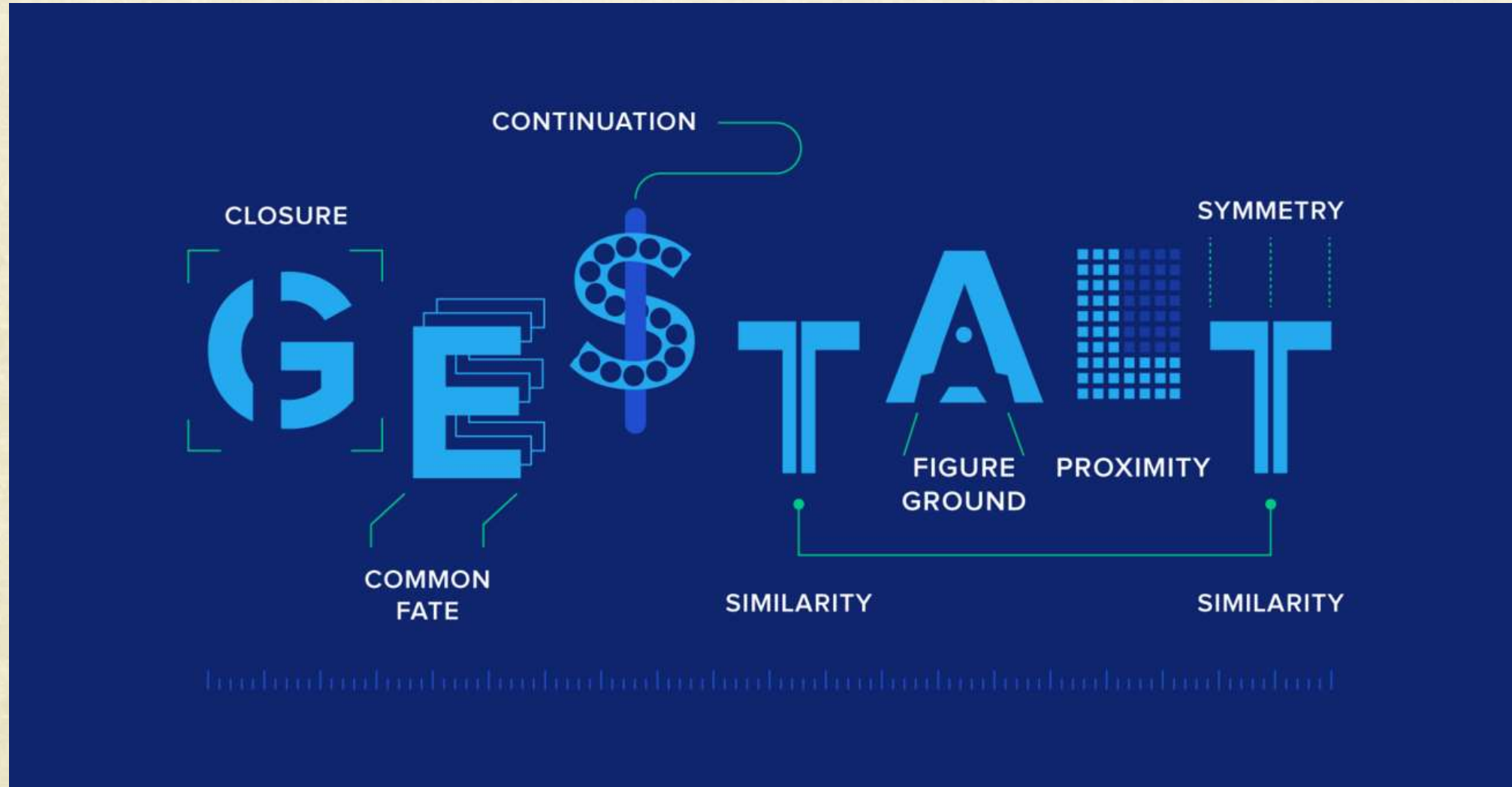
# Selective Attention



<https://www.youtube.com/watch?v=vJG698U2Mvo>



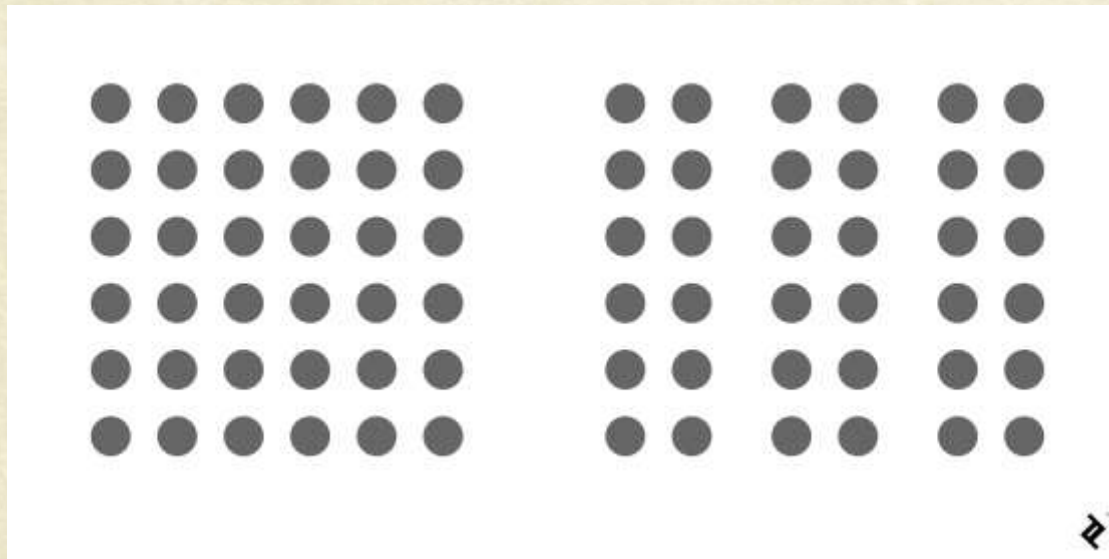
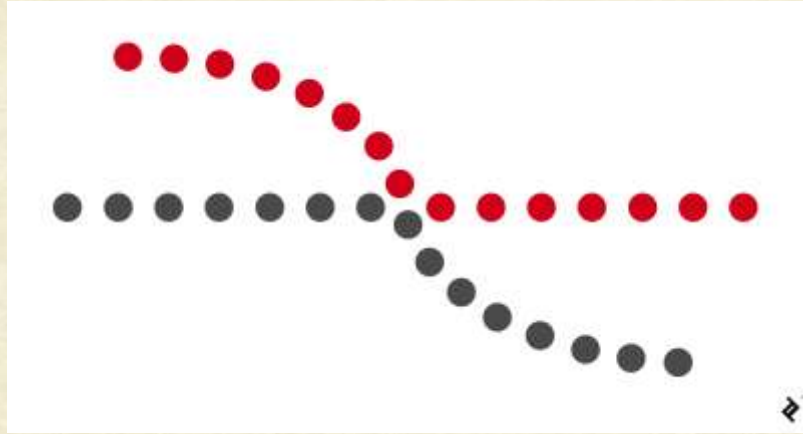
# Look for Gestalt principles







# Gestalt principles in action





# Three “Urges” on seeing a Picture\*

- Given an image, you want to do:



## Re-organization

Group proximate and similar parts into meaningful regions

## Recognition

Recollect previously seen objects from memory

## Reconstruction

Measure quantitative aspects:  
Number, Size, Distance, etc.

*\*Jitendra Malik; Mysore Park, Dec. 2011*

*The three R's of computer vision*

<https://www.sciencedirect.com/science/article/abs/pii/S0167865516000313>





# The Three Rs of Computer Vision

## Reorganization (Segm.)



Group semantically similar pixels



## Recognition

Connecting what we see to our memory



## Reconstruction

Measure/recreate a 3D model of what we see in the world



# Why is it Difficult?



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





# Scene Interpretation

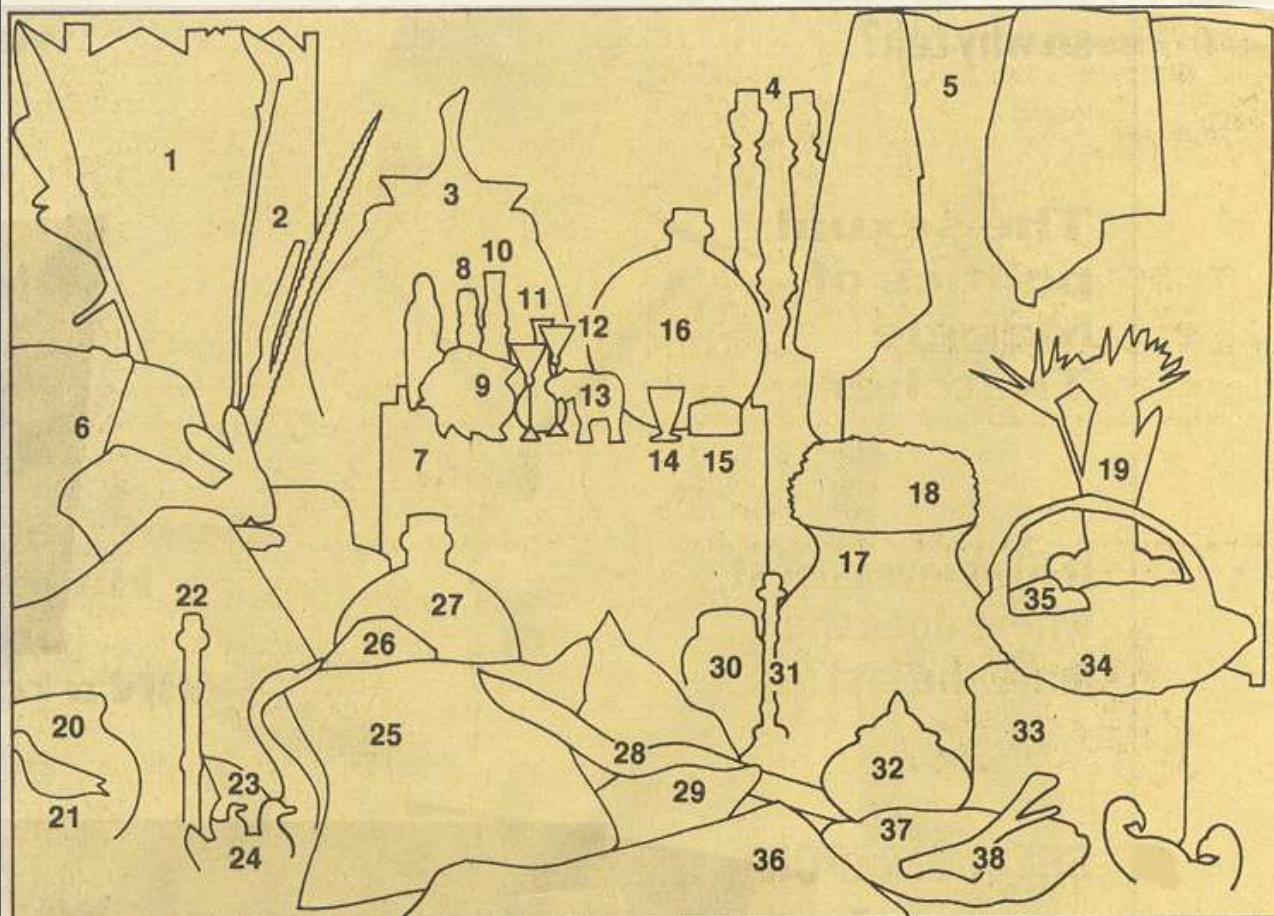






# Segmentation and Labeling

1. Hand-carved Shesham wooden screen
2. Wooden flowers
3. Wicker basket
4. Pair of hand-carved Thai candlesticks
5. Indonesian rattan screen
6. Dhurry covered armchair
7. Hand-painted chest
8. Striped wooden Indian candlestick
9. Stone terracotta Thai
10. Moroccan ceramic candlestick
11. Blue Egyptian glass decanter
12. Bronze goblet-shaped candlesticks
13. Painted wooden Indian elephant
14. Blue Egyptian glass goblets
15. Indian brass filigree box
16. Painted Indian oil bottle
17. Large African water pot
18. Philippino twig basket
19. Philippino bamboo covered urn



20. African cooking pot
21. Decoy bird
22. Painted candlestick
23. Thai wooden swan
24. Carved wooden duck
25. Embroidered mirror cushion covers
26. Green hexagonal Indian box
27. Painted Indian oil bottle
28. Joint wooden snake
29. Black embroidered cushion
30. Moroccan ceramic jar
31. Painted wooden candlestick
32. Thai pot with lid
33. Octagonal Indian box
34. Shallow twig baskets
35. Mexican paper mache fake fruit and vegs
36. Nakshe Kantha Bengali wall-hanging
37. Wooden shell bowl
38. Wooden servers





# Computer Vision

- 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!
- Why is it difficult?



# Chairs and Chairs

- Which are chairs?
- Large intra-class variations
- How do we describe a chair?
- Basic property: Sittability!
- We infer a lot from pictures.  
Can we instruct a computer to do the same?
- Do we understand how we infer?







# Applications



# Applications: Sports







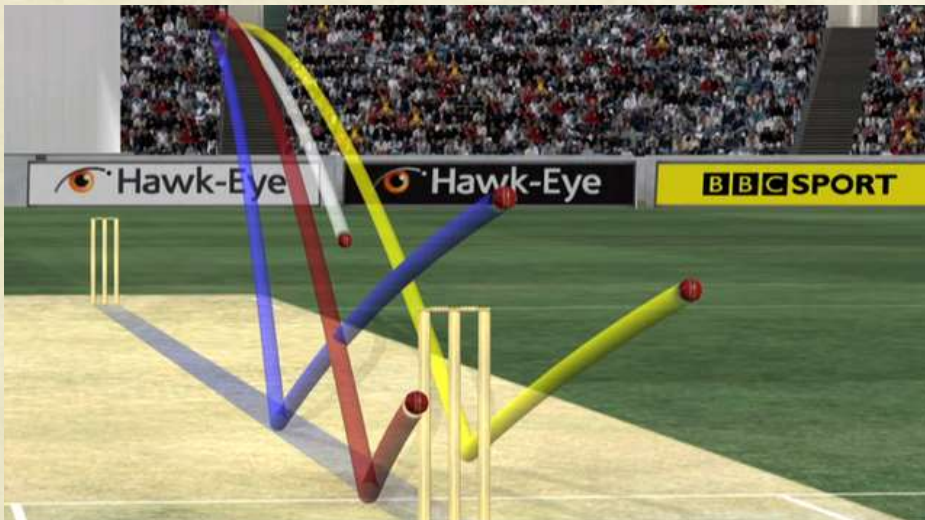
# Applications: Broadcast video / Sports!



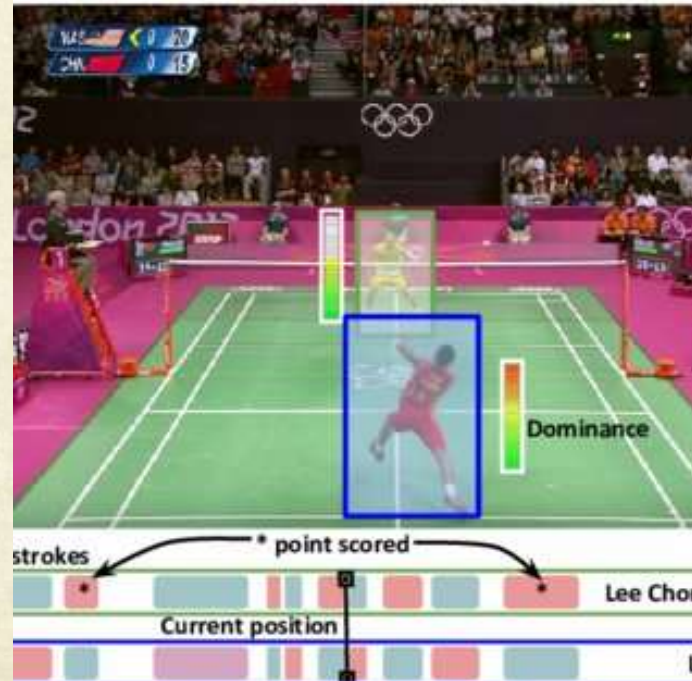
Field Understanding: Virtual Line



Chroma Keying: Replacing Backgrounds



Ball Tracking: Hawk Eye



Player Tracking: CVIT, IIITH





# Applications: Reversing your car



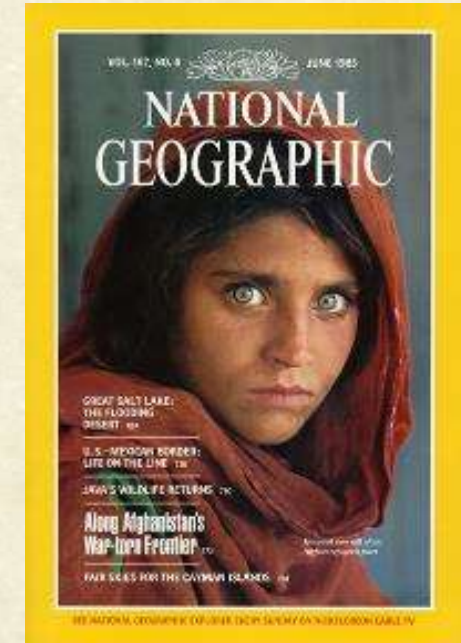




# Applications: Biometrics



Computer Access

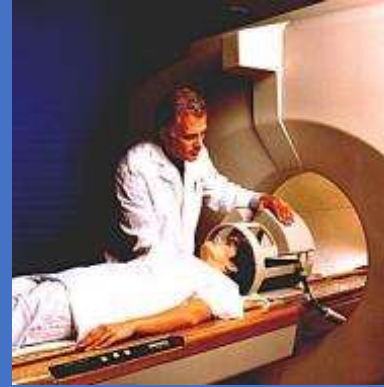




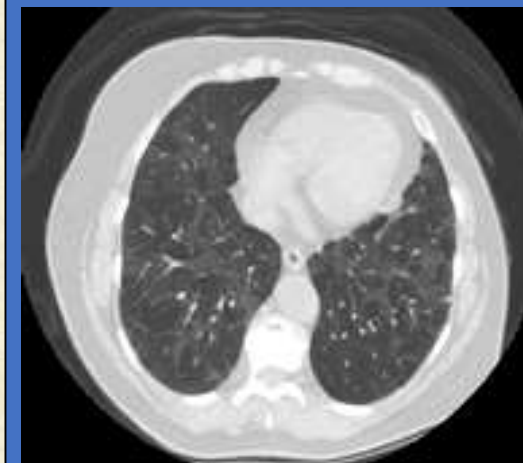
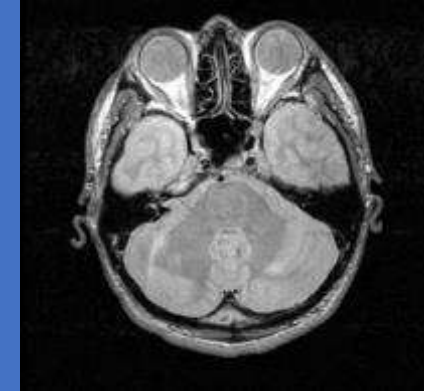
# Applications: Medical Imaging



Computer Assisted Surgery



CT Scan



Segmentation





# Applications: Automatic PCB Inspection



Manual PCB Inspection



Automated PCB Inspection



# Applications: Making Good Movies

- Structure light scanner, laser range finder
- Multi-camera stereo, structure recovery
- Reverse Engineering
- Virtualized/Augmented reality







# Applications: Reconstructing the 3D world



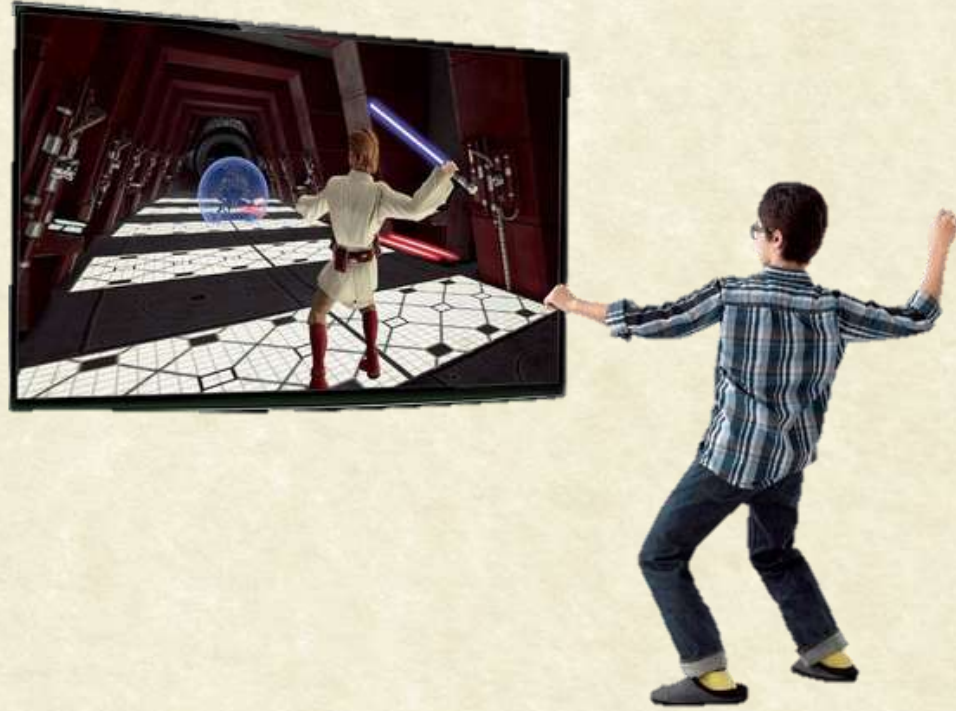
- Approximate 3D structure from an unstructured collection of images!  
[PhotoTourism, SIGGRAPH2006]
- PhotoSynth
- Autodesk 123D: Your pictures to model
- And many more to follow soon



# Applications: Gaming



Microsoft Kinect

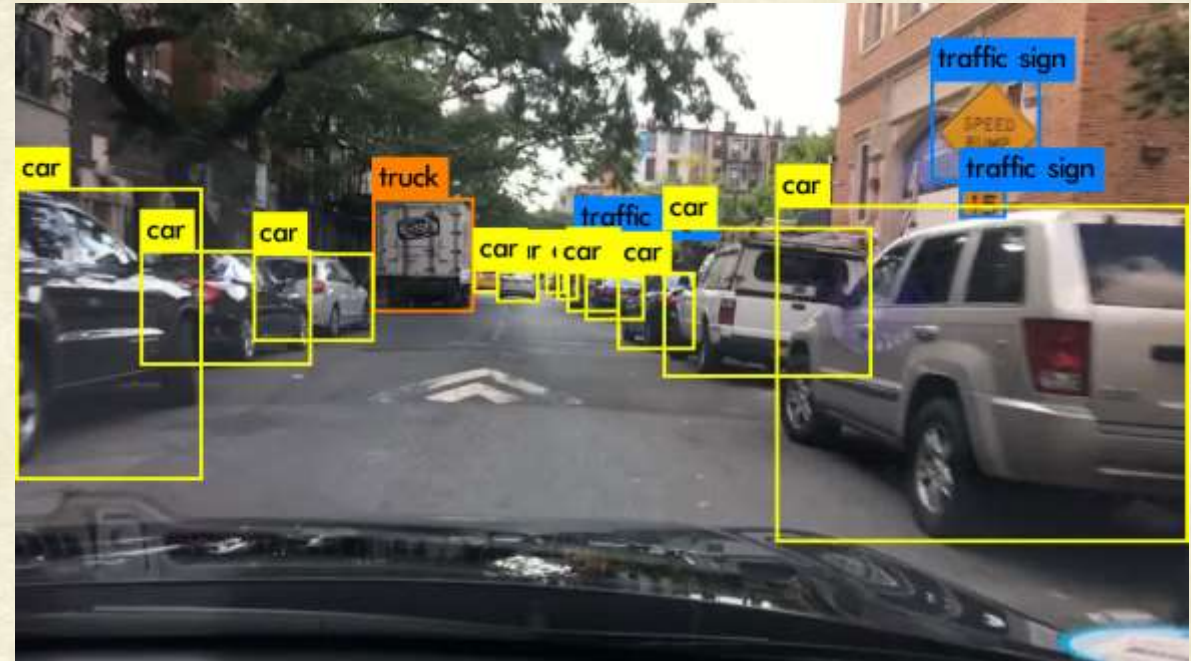
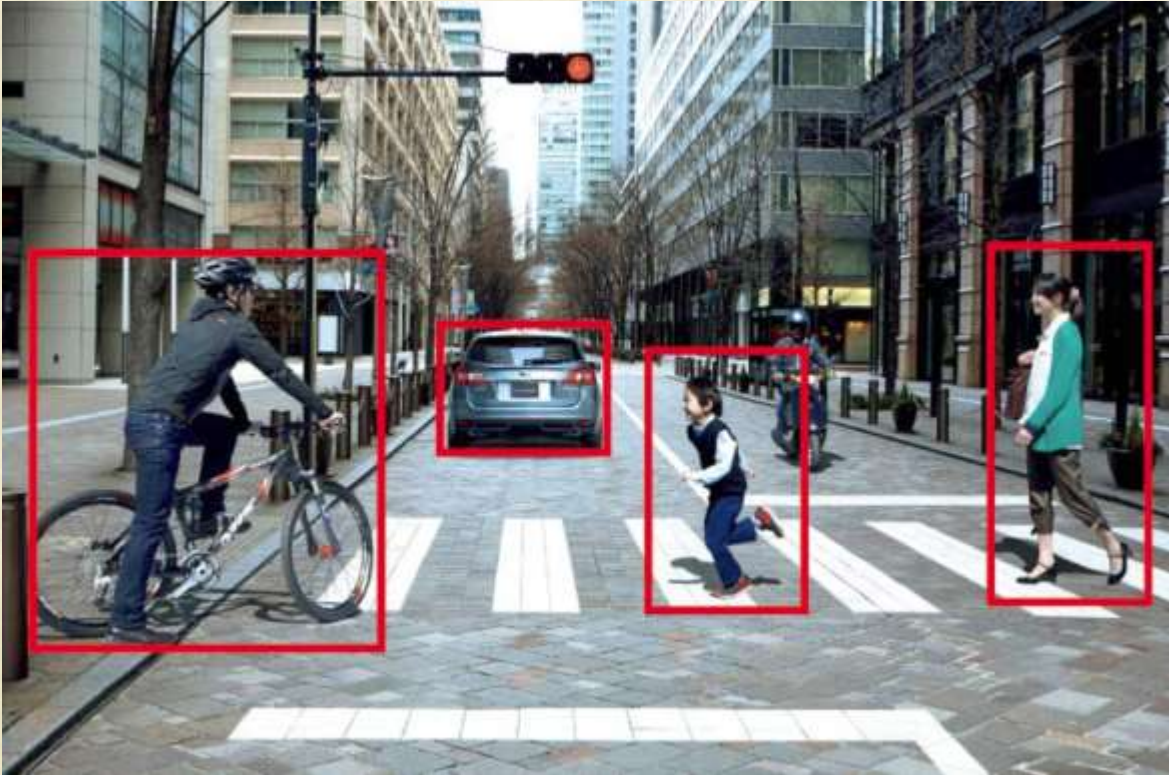


- You are the controller. Interact naturally with the game.
  - Fastest Selling Electronic Device Ever: 80 lakh units in 60 days!!
- Finding great use in Computer Vision, Robotics, etc.





# Applications: Automotive Safety



Can help avoid accidents greatly!





# Applications: Many others!

- Surveillance
- Automated Assembly
- Mail Sorting
- Face detection (photography)
- Robot Navigation
- Content-Based Image Retrieval
- Entertainment
- Generation
- And many more... with your help...







# Why Automated Vision?

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

A general-purpose machine vision system does not exist!



# Well .. but there is GPT-4o / Gemini?

Have you tried it?

When does it fail?

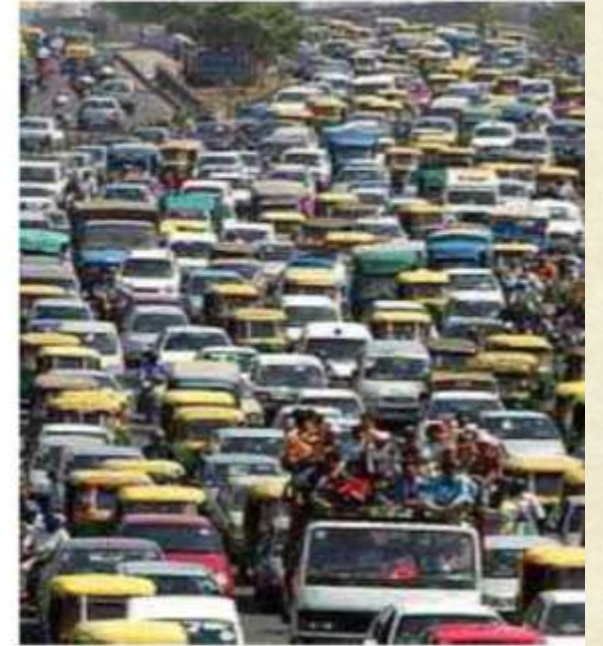
- My hypothesis (probably wrong!)
- LLMs are great, they train on ALL the world's text.
- Vision models, are not trained on EVERYTHING (yet)
- Rich Sutton's bitter lesson (?)

<http://www.incompleteideas.net/IncIdeas/BitterLesson.html>





# The Real Problem



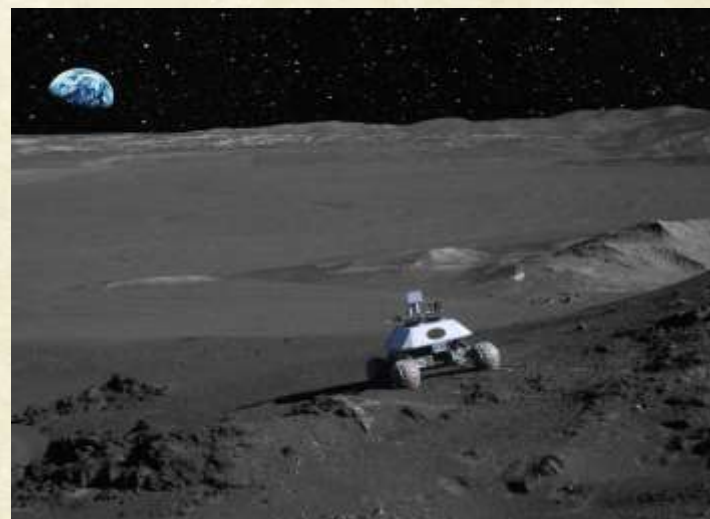
Develop something similar for Indian roads!





# What More is Possible?

- Much much more ...
- The journey is still underway for computer vision.
- Large amount of data, high computing power, machine learning algorithms continue to transform computer vision.
- Big(ger) things are yet to come.







# Logistics



# What about Deep Learning?

- DL has become the primary driving force behind most recent success in CV. We will study it in this course.
- Computer vision has a strong mathematical and conceptual basis developed over 4 decades
  - Geometry
  - Optimization
  - Visual object representations
  - Optics, Lighting, Appearance models
- We will refer to the basics when required, but not cover above in detail





# Course Outline, Topics

- Recap (DIP, SMAI)
- Datasets in Computer Vision
- Recognition with CNNs
- Detection with CNNs
- Dense Prediction
- Transformers
- Vision and Language
- Self-Supervised Learning



# Pre-Requisites for the Course

- Linear algebra and a good mathematical outlook
  - Vectors, matrices, eigenvalues, singular values
  - 2D/3D geometry
- Image/Signal processing
  - Filtering, edge detection, segmentation
  - Transforms, analysis
- Pattern Recognition
  - Features, classifiers
  - Training, validation, testing
- Programming
  - Python, NumPy, [familiarity with PyTorch is a bonus!  
[https://www.youtube.com/watch?v=Z\\_ikDlimN6A](https://www.youtube.com/watch?v=Z_ikDlimN6A)]

You may struggle if you are not comfortable with these!

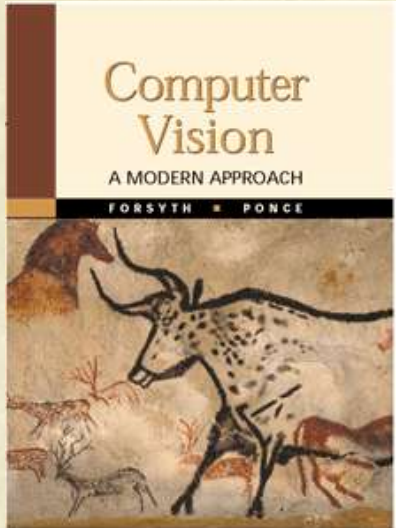




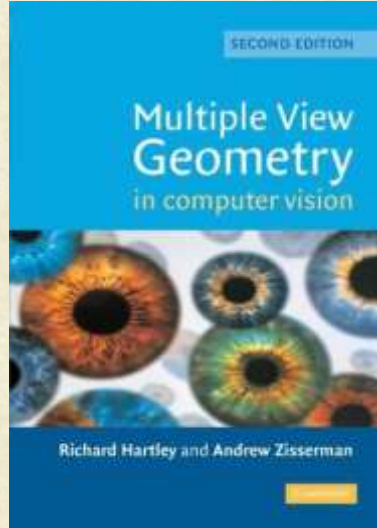
# Some popular CV books

No single textbook will be followed. Material will be shared as necessary.

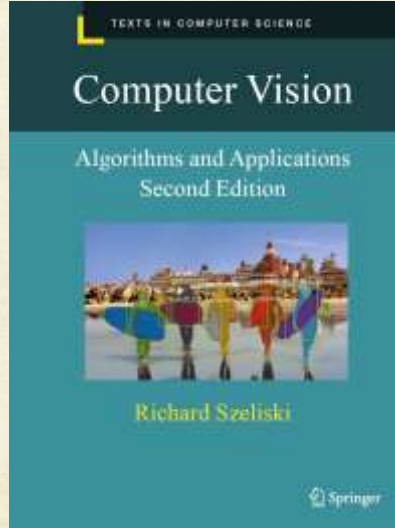
Forsyth & Ponce



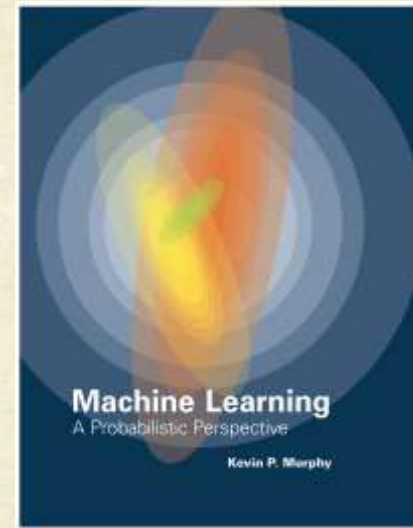
Hartley & Zisserman



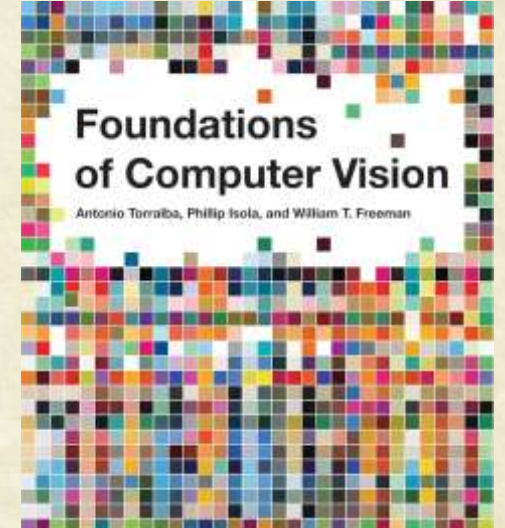
Rick Szeliski



Kevin P. Murphy



Torralba, Isola, Freeman



... and several papers and resources.



# Teaching Assistants

- Keshav Gupta
- Kyrylo Shyvam Kumar
- P Harshavardhan
- Sanika Sachin Damle
- Mohd Hozafa Khan
- Shivank Saxena





# Grading

- Grade Distribution
  - Assignments: 50%
  - Mid-term exam: 20%
  - End-term exam / Project: 30%
- Projects for a subset of students based on proposals. Pairs of 2.
- This is an **advanced elective**
  - We expect you to work hard and learn well
  - Class participation helps everyone in the class
  - We don't want credit-seekers or resume-padders



# Late Days Policy

- **Late submission bonus:** Maximum of **three days**. You must adhere to standard late submission policy below after using your bonus days. No exceptions will be made. You'll need to inform TAs before assignment deadline if you wish to use the late submission bonus.
- **Assignment Late Policy:** 50% if one day late; zero percent if more than one day late





# Grading Policy

- Grades Policy:
  - No revisiting grade components after assignment evals
  - No revisiting grade components after paper checking
  - Largest gaps in sorted order typically form the grade boundaries (A,A-,B,B- etc.).
  - Alternate gap locations may be used (subjective!)



# Exam Absence Policy

- If you want to stay for paper checking, make travel plans after May 10<sup>th</sup> 2025.
- **Exam Absence policy (applicable only for mid-term)**
  - Approval from Academic Office necessary.
  - Marks will be calibrated based on your relative class performance.
  - Does not apply for end-term exam (you'll need to re-take if you miss this)





# Class Etiquette

- Be in the class before 10:05 am
- Be curious, respectful, and in the present
- Attendance is counted
- If you have a doubt, **please ask!** Others are likely to have the same doubt
- Devices
  - Cellphones silent and away; notifications can wait
  - Laptops are a distraction; consider taking notes on paper
- Please write to us by email
  - Suggestions and questions are welcome
  - Genuine challenges with evaluations



# Plagiarism ✖

- Any kind of plagiarism is **strictly prohibited**
  - Using someone else's ideas or work and represent as your own.
  - Copying or allowing to copy on any assessments.
  - If you share ideas or receive help, give fair attribution.
- You are encouraged to discuss and learn from each other, but not during the exams :)
- You may (will anyway?!) use ChatGPT / Github copilot / etc. But you are responsible for what it generates and should be able to defend your answers in the assignment viva





Questions?