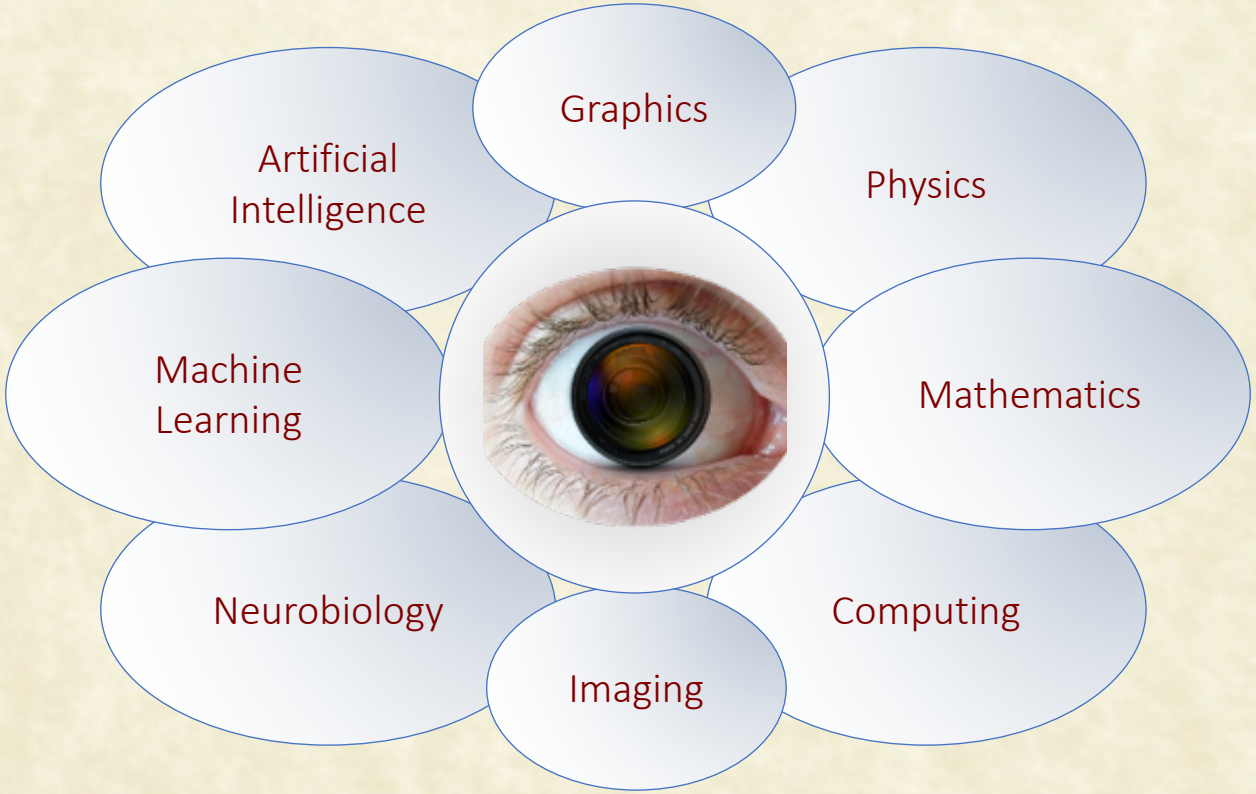
CSE 578: Computer Vision

Spring 2021: Introduction



Anoop M. Namboodiri

Biometrics and Secure ID Lab, CVIT,

IIIT Hyderabad

Course Outline, Topics

1

Computer Vision

M1: Geometry	M2: Image Grouping	M3: Recognition
Pinhole Camera Model	Segmentation as Labelling	Feature Detection, Descriptors
2 Proj. Geometry, Camera Matrix	3 Graphcut, Binary Segmentation	4 Face Detection
Camera Calibration	MRF for Segmentation	Pedestrian Detection (HoG,SVM)
2-View Geometry, Homography	Multi-label MRFs	Bag of Words, SURF, Others
Fundamental Matrix	Deep Learning for Segmentation	Deformable Parts Model
Stereo Corr., Depth Estimation		Indexing and Retrieval
	Image-to-Image Networks	CNNs for Detection and Recog.
5 SFM and Bundle Adjustment		
Image Rectification		Vision and Language
Computational Imaging		

What about Deep Learning?

- DL has become the primary driving force behind most recent success in CV. However, this is the first course on Computer Vision.
- Computer vision has a strong mathematical and conceptual basis developed over 4 decades
  - Geometry
  - Optimization
  - Visual object representations
  - Optics, Lighting, Appearance models
- You need to know the basics to build on it

Pre-Requisites for the Course

- Linear algebra and a good mathematical outlook
  - Vectors, matrices, eigenvalues, singular values
  - 2D/3D geometry
  - See course page for a more detailed list of topics
- Image/Signal processing
  - Filtering, edge detection, segmentation
  - Transforms, analysis
- Pattern Analysis, Algorithms, Programming
  - Features, classifiers
  - Training, testing, validation
  - Python/C++, OpenCV

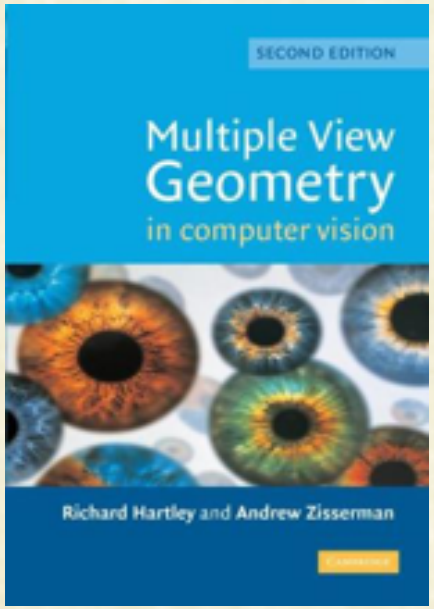
Brush up these topics if you aren't certain. A reading list of online material will be prepared for the preliminaries

Reference Books

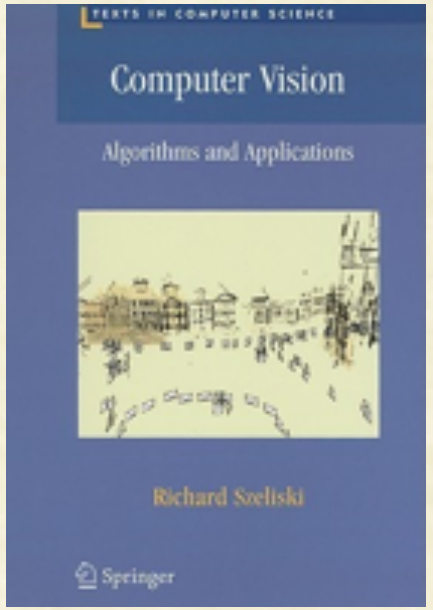
No single textbook

Forsyth & Ponce
Indian Edition

Hartley & Zisserman
Indian Edition



Rick Szeliski
Free online version



… and read several research papers

Administrivia

- Grade Distribution
  - In-class Quizzes: (~25%)
  - Tests/Quizzes: (~25%)
  - Homeworks/Assignments: 5 Assignments (~25%)
  - Project: In groups of 4 (~25%)
- This is an advanced elective that you opted for
  - We expect you to work hard to learn well.
  - Class participation lifts the level of the class
  - We don’t want credit-seekers or resume-padders here
- Inverted Classes
  - Most classes will be conducted in the inverted class model
  - Content videos will be uploaded the day before.
  - Classes will start with a short quiz on the video followed by discussions and problem solving

Class Etiquettes

- Be in the class before 2pm
  - Quiz will be available only for a fixed time starting at 2pm
- Reduce noise in the class
  - Switch off your cameras and microphones
  - Put your hand up if you have a question
  - You may also type your questions in the chat
- Watch the content video (multiple times if required) well in advance and be clear about the topic or with questions.
- If you have a doubt, ask. Others are also likely to have the same doubt.

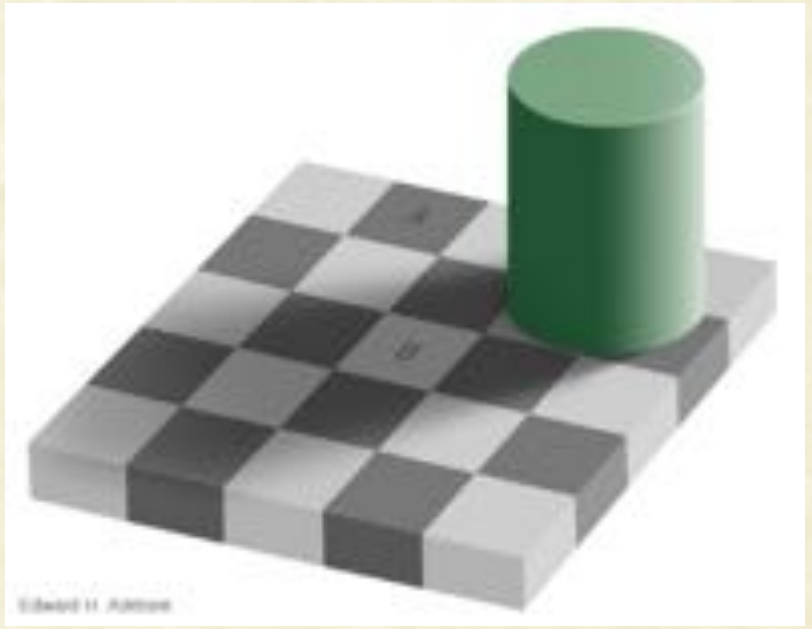
What is Computer Vision?

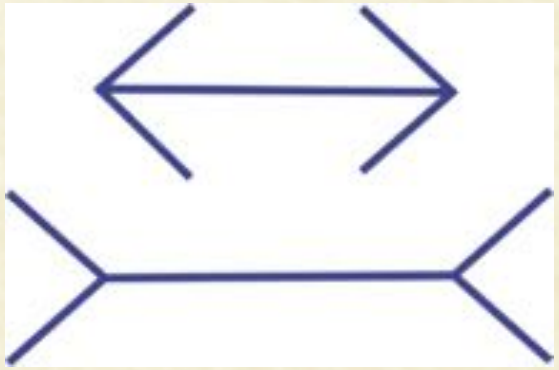
- Understanding of visual inputs (images/videos) by computers.
- Making sense out of them. Describing them.
- Does computer vision mimic the human vision?
  - Certainly in many of its goals
  - Why? Human vision is among the best!
  - Sophisticated and efficient but not understood well
- Should computers process visual inputs like humans?

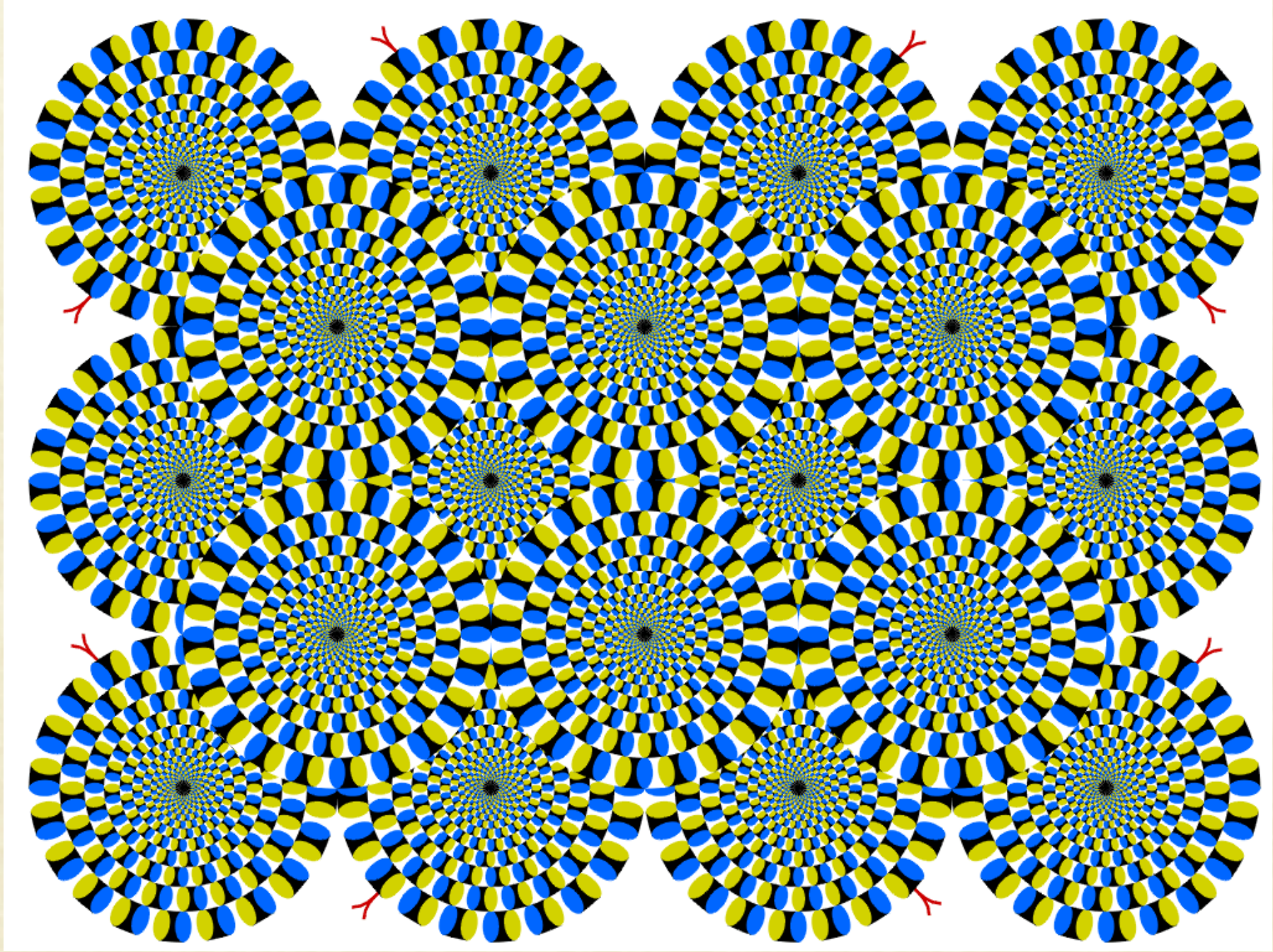
  Not necessarily!
  - Human visual system need not limit computer vision
  - We draw inspiration from it as often as is convenient

Human perception has its shortcomings…

Sinha and Poggio (Image: Ron Rensick)







Copyright A.Kitaoka 2003

Three “Urges” on seeing a Picture*

1. **To group** proximate and similar parts of the image into meaningful “regions”.

   Called segmentation in computer vision.

2. **To connect to memory** to recollect previously seen “objects”.

   Called recognition in computer vision.

3. **To measure** quantitative aspects such as number and sizes of objects, distances to/between them, etc.

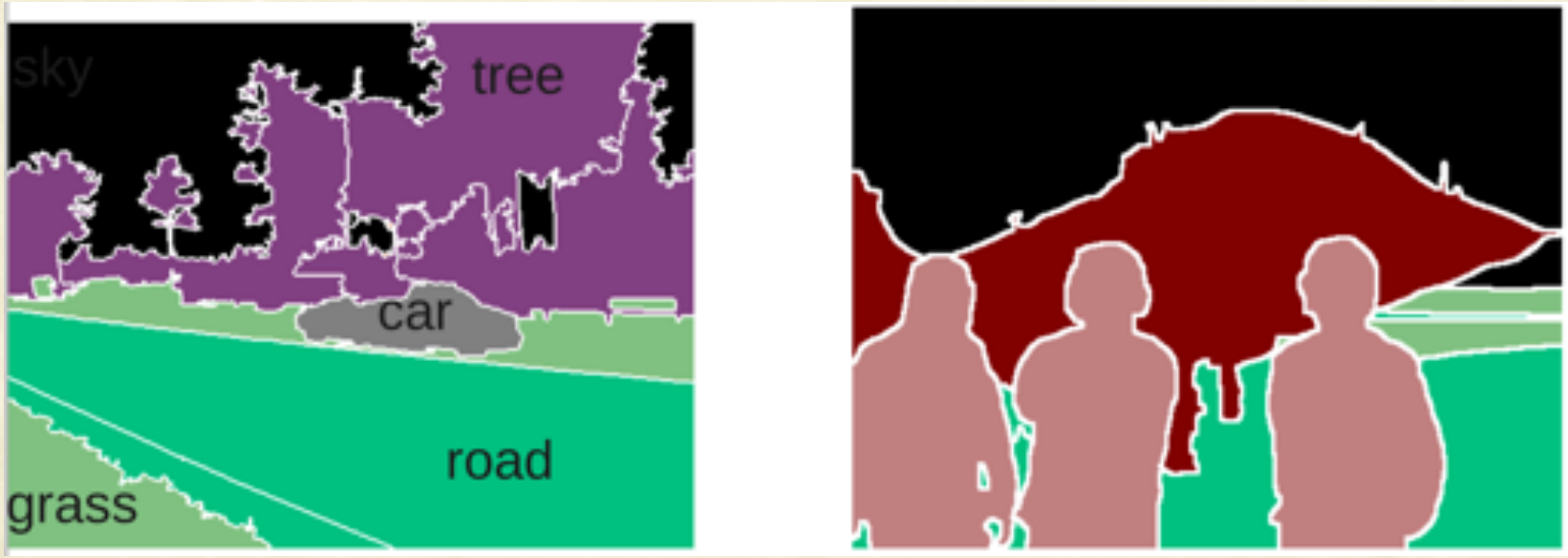
   Called reconstruction in computer vision.

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

Urge to Group

- We don’t see individual pixels (like the computer does!).
- We see groups of pixels together.
- What is the basis for “correct” grouping?

Urge to Group



- Group similar pixels together as objects.
- Group semantically meaningful pixels together as objects.
- Is appearance similarity the same as semantic similarity?

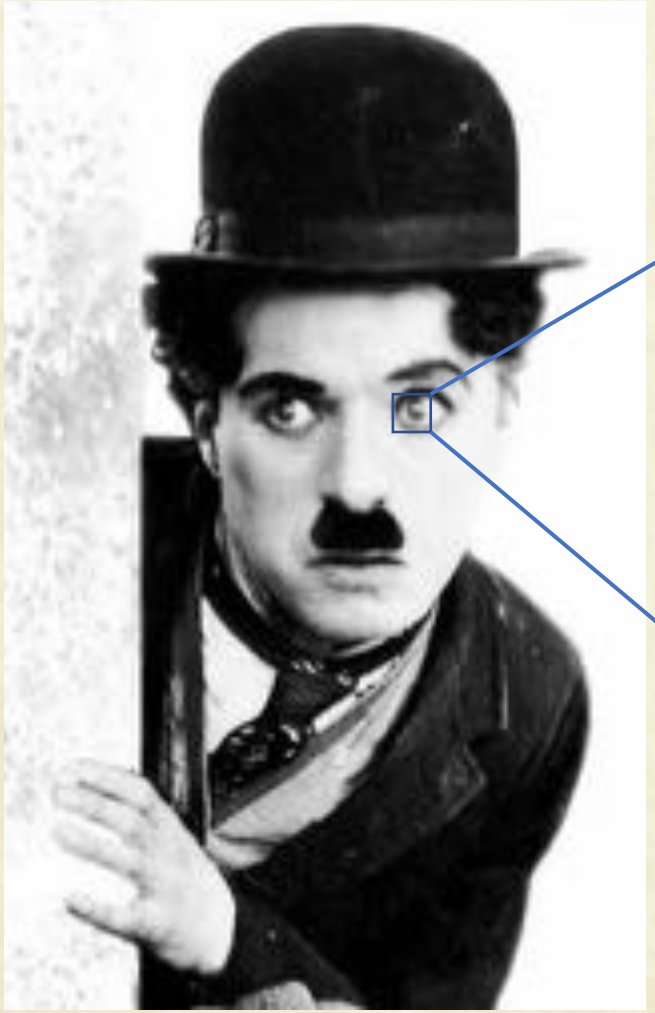
Urge to Touch Memory

Recognizing objects from (visual) input is fundamental to human cognition of the world.

Urge to Measure

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

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

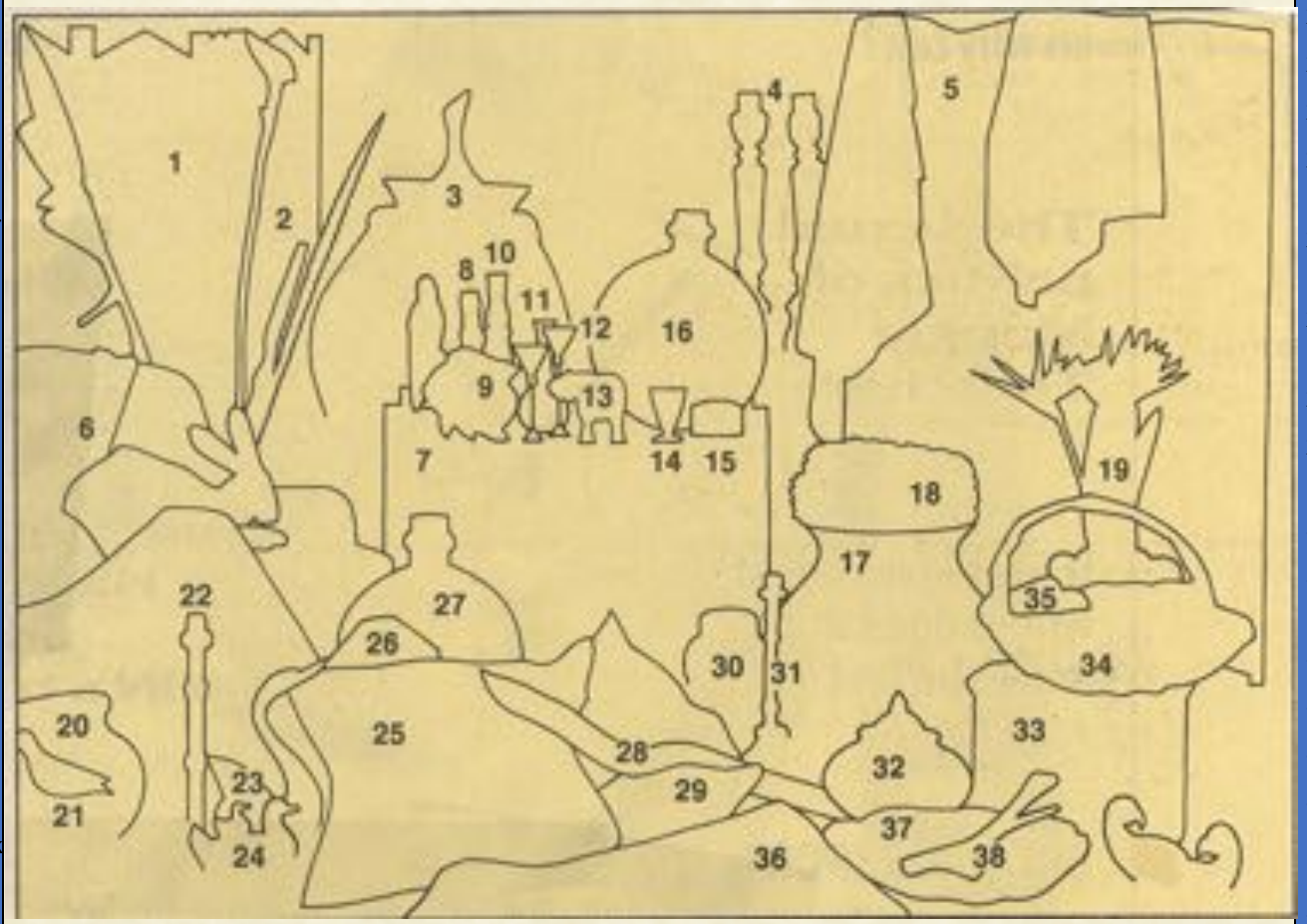
15.Indian brass filigree box

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!

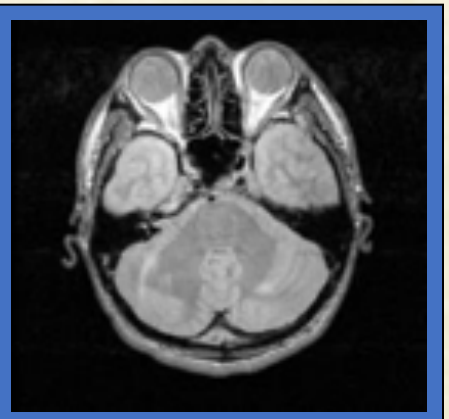
- Why is it difficult?

Chairs and Chairs

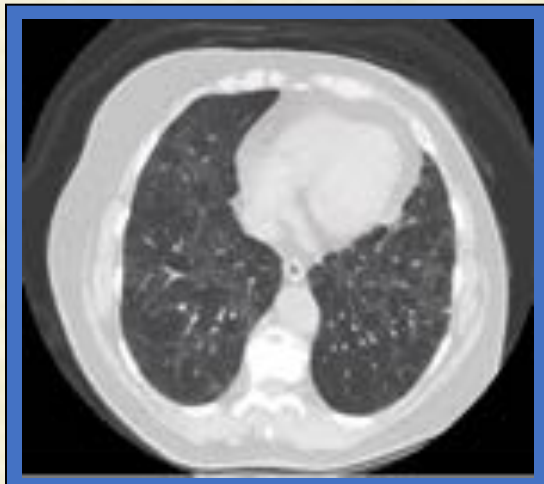
- Which are chairs?
- Large intra-class variations
- How do we describe a chair?
- Basic property: Sittability!
- We infer a lot from pictures. Can we instruct a computer to do the same?
- Do we understand how we infer?

Applications: Medical

Computer Assisted Surgery

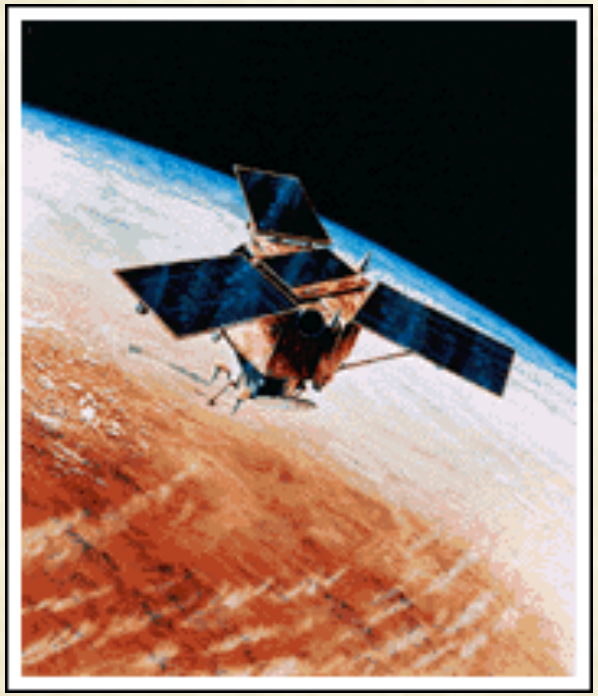


CT Scan

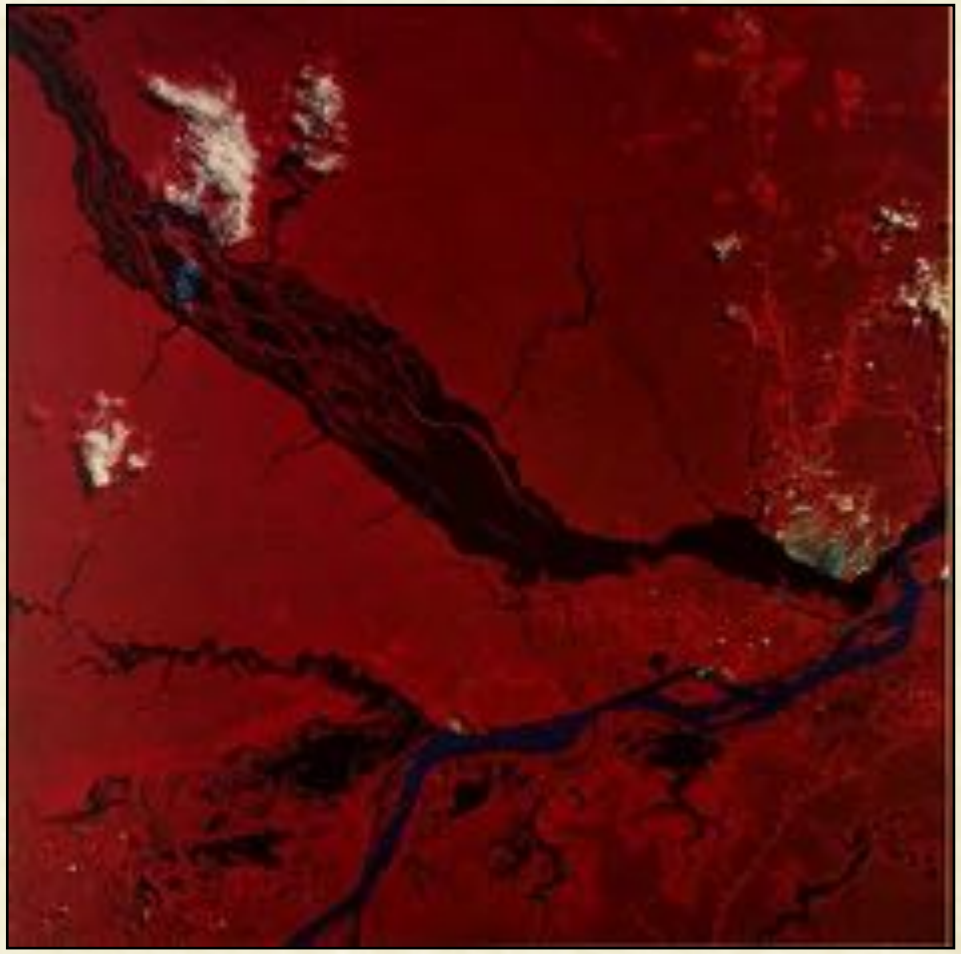


Segmentation

Applications: Space Imaging



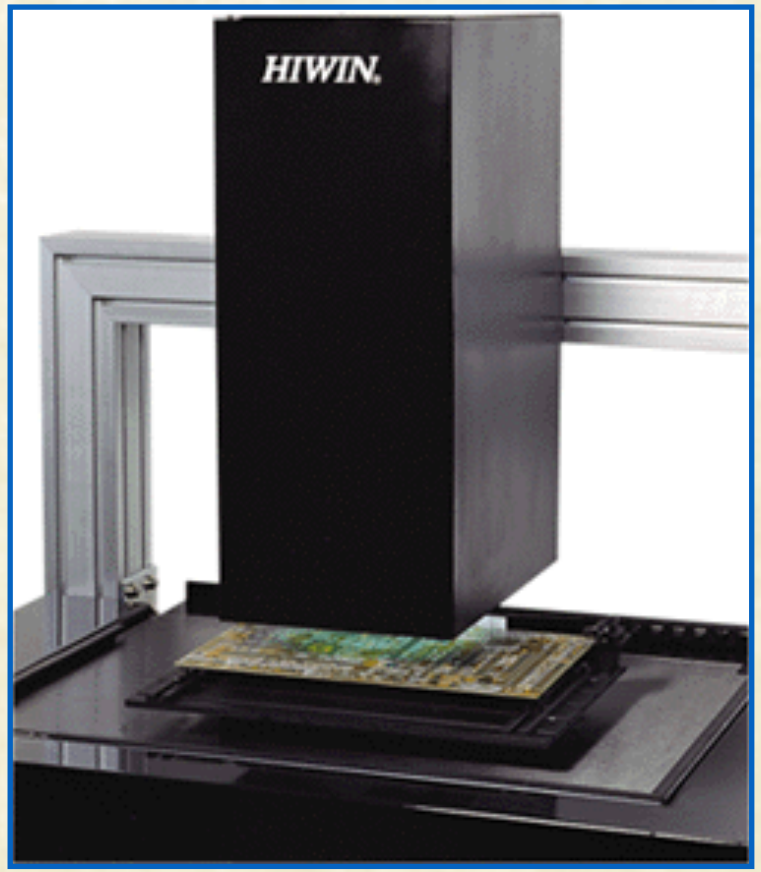
Ikonos



Rio Negro (black) meets Amazon (blue)

Applications: Automated Inspection

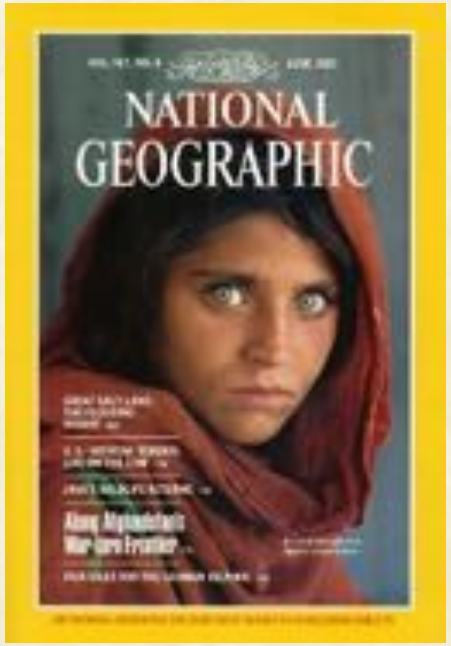
Manual PCB Inspection



Automated PCB Inspection

Applications: Biometrics

Travel

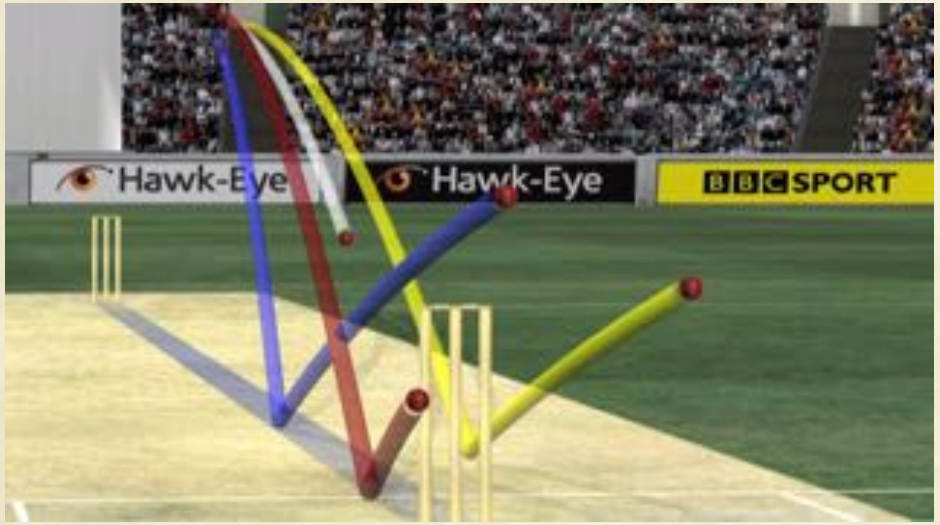


Computer Access

Disney Land

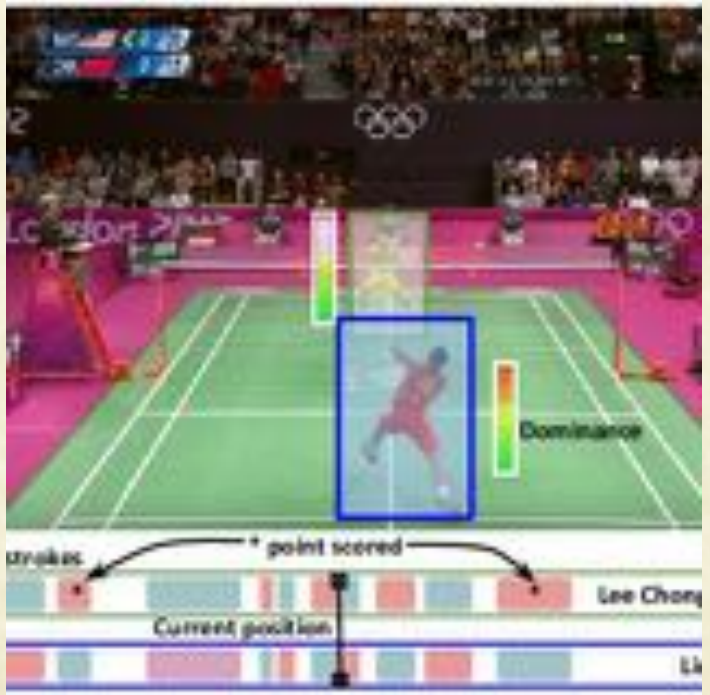
Applications: Broadcasting

Field Understanding: Virtual Line



Ball Tracking: Hawk Eye

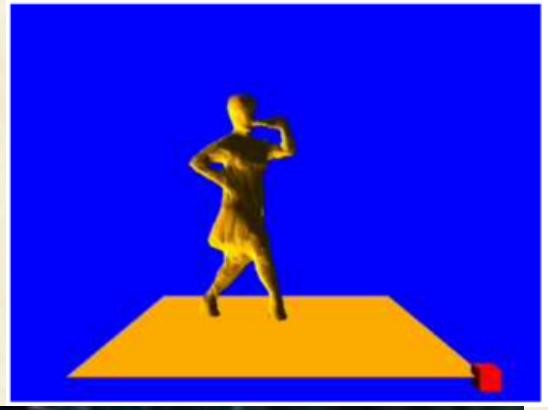
Chroma Keying: Replacing Backgrounds



Player Tracking: CVIT, IIITH

3D Shape and Motion Recovery

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



Applications: Others

- Surveillance
- Automated Assembly
- Mail Sorting
- Face detection (photography)
- Robot Navigation
- Content-Based Image Retrieval
- Entertainment
- 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

General purpose machine vision system do not exist.

Recent: Structure from Motion

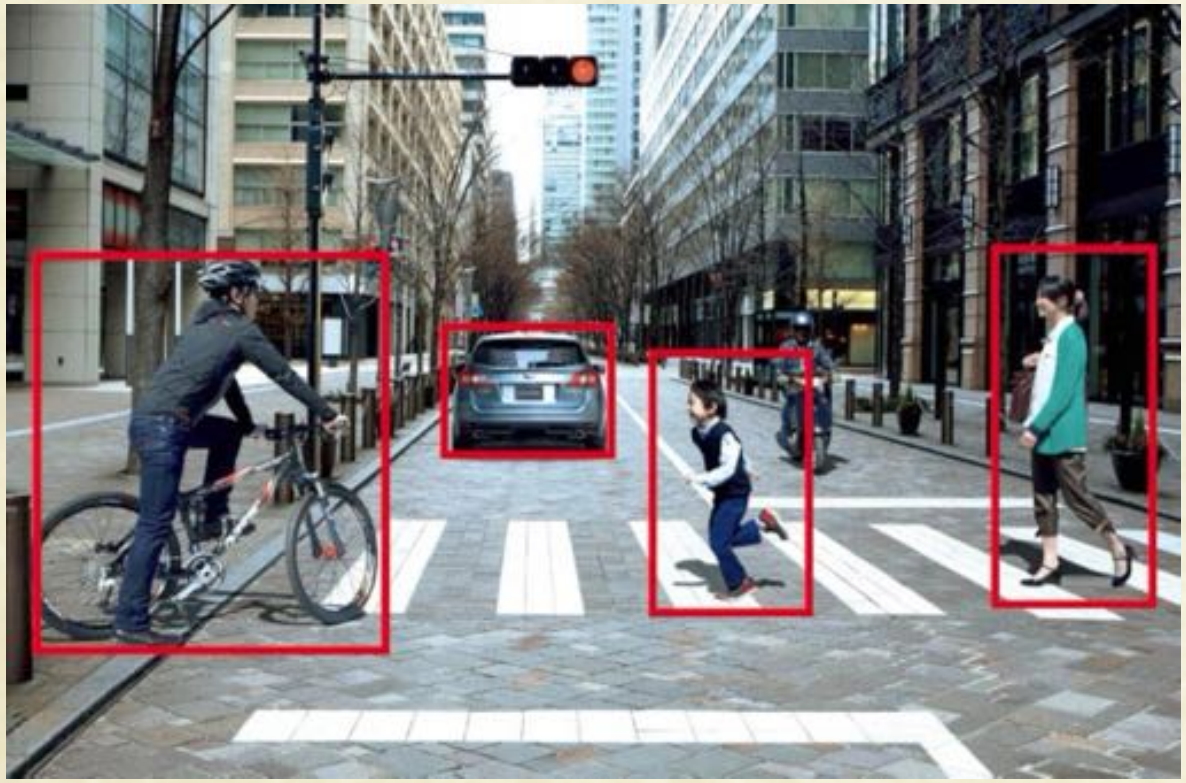
- Approximate 3D structure from an unstructured collection of images! [PhotoTourism, SIGGRAPH2006]
- PhotoSynth
- Autodesk 123D: Your pictures to model
- And many more to follow soon

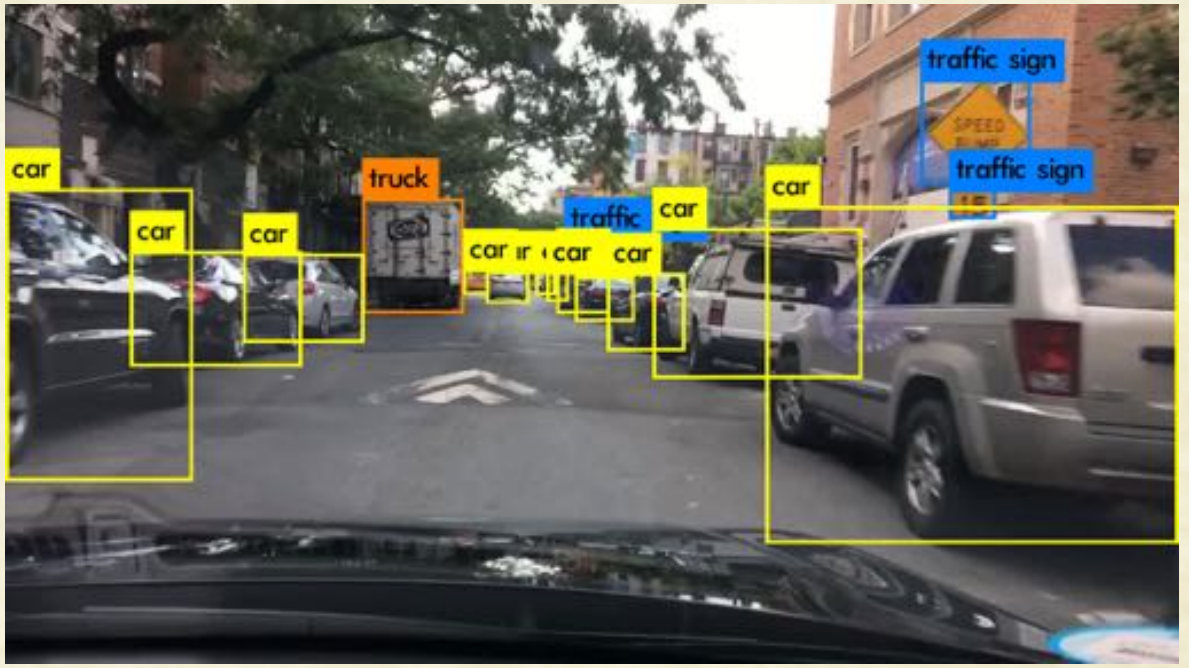
Recent: Natural Gaming

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.

Recent: Automotive Safety





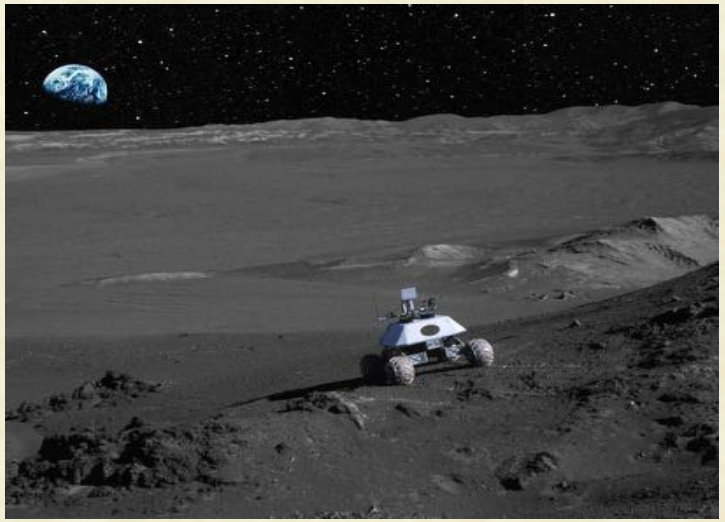
Can help avoid accidents greatly!

The Real Problem

Develop something similar for Indian roads!

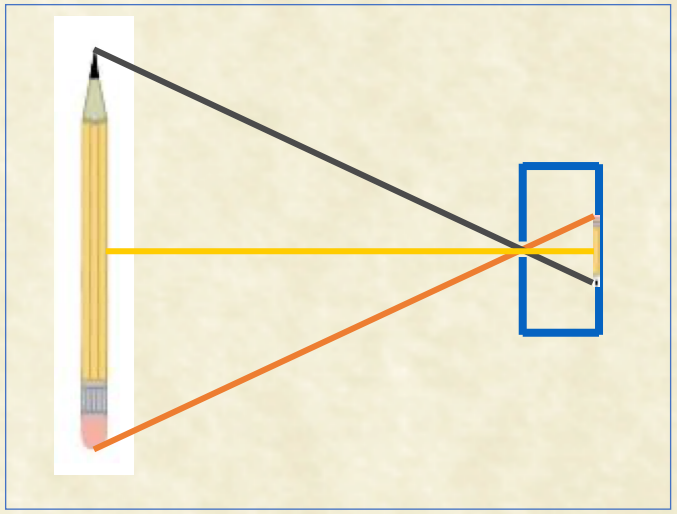
What More is Possible?

- Much much more .....
- The journey has just begun for computer vision.
- Large amount of data, high computing power, machine learning algorithms continue to transform computer vision.
- Big things are yet to come.

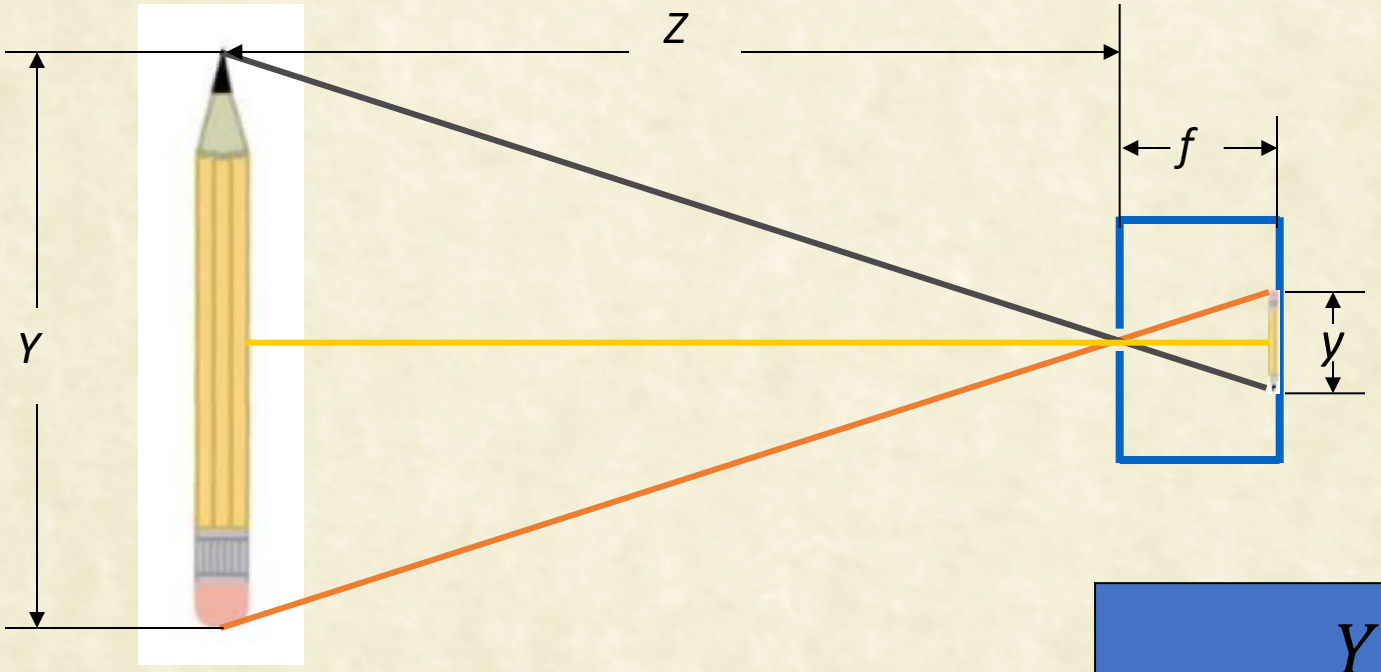


Questions?

M1 Geometry: Imaging and Camera Model

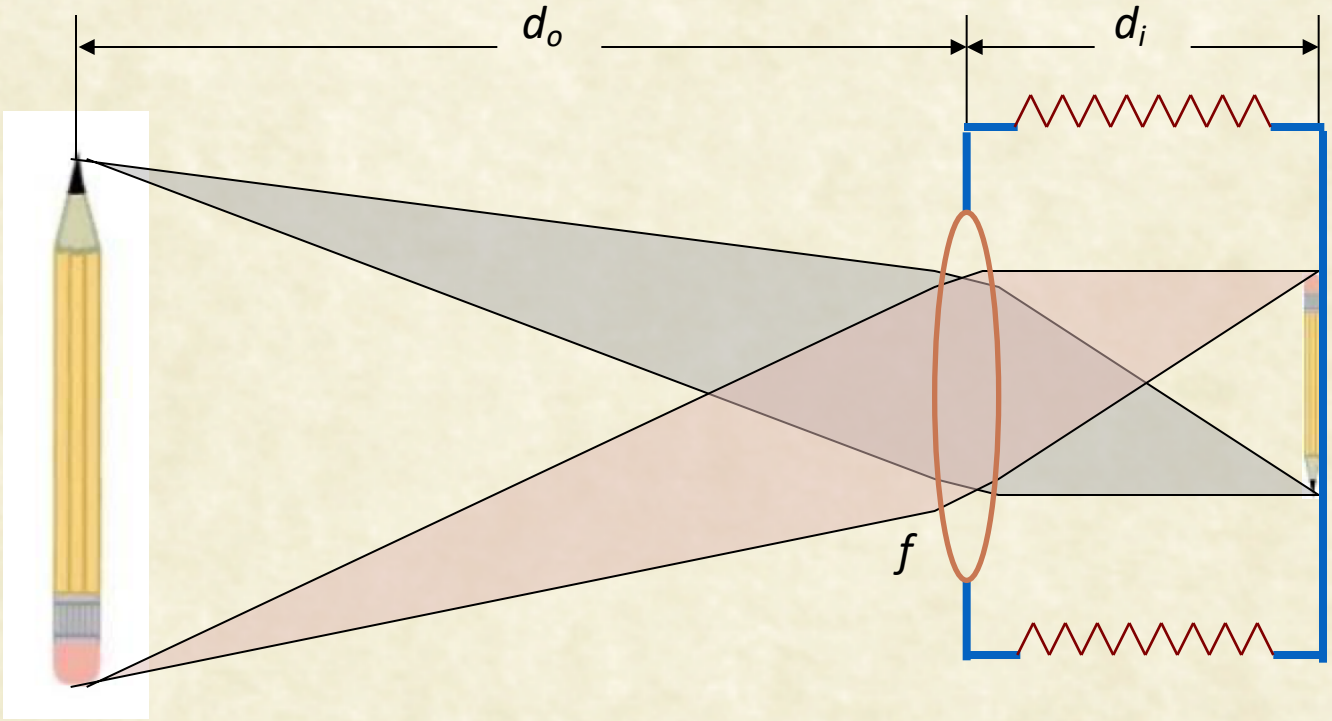


The Pinhole Camera



$$y = f\frac{Y}{Z}$$

Camera with Lens

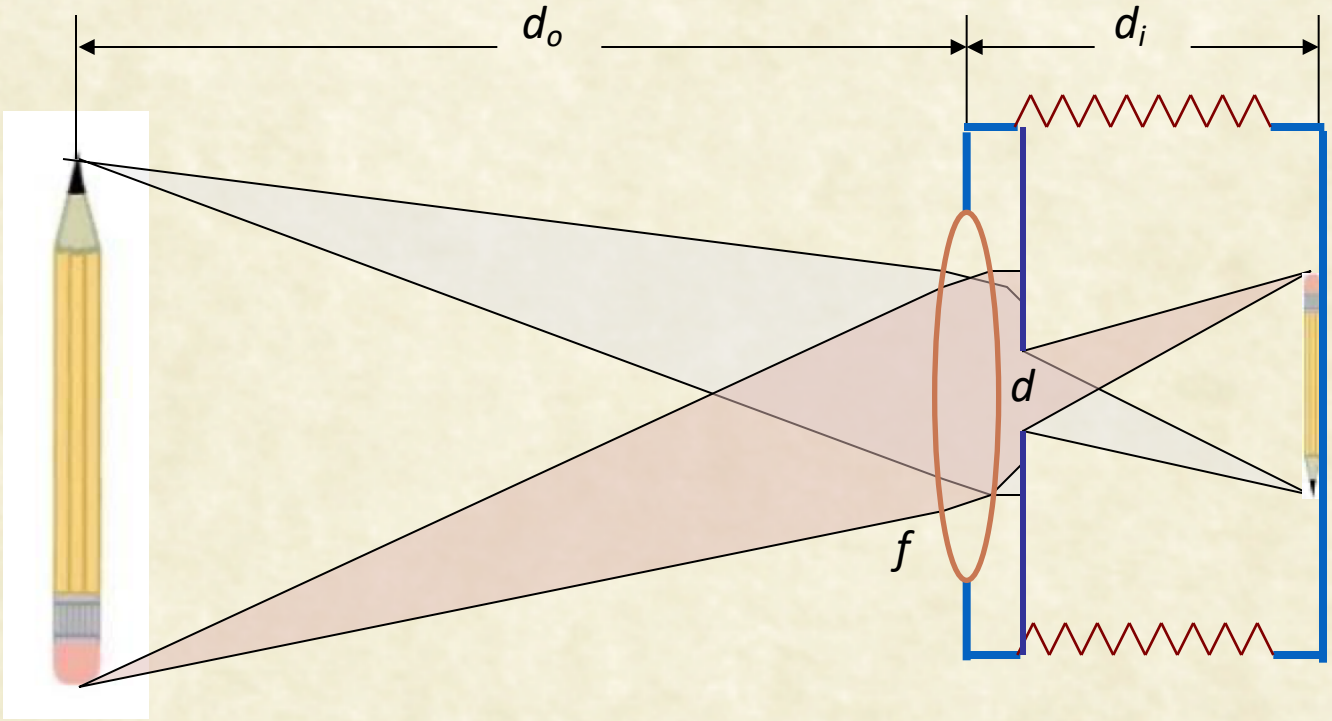


Thin lens equation: $\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$

$$d_i = f\frac{d_o}{(d_o - f)}$$

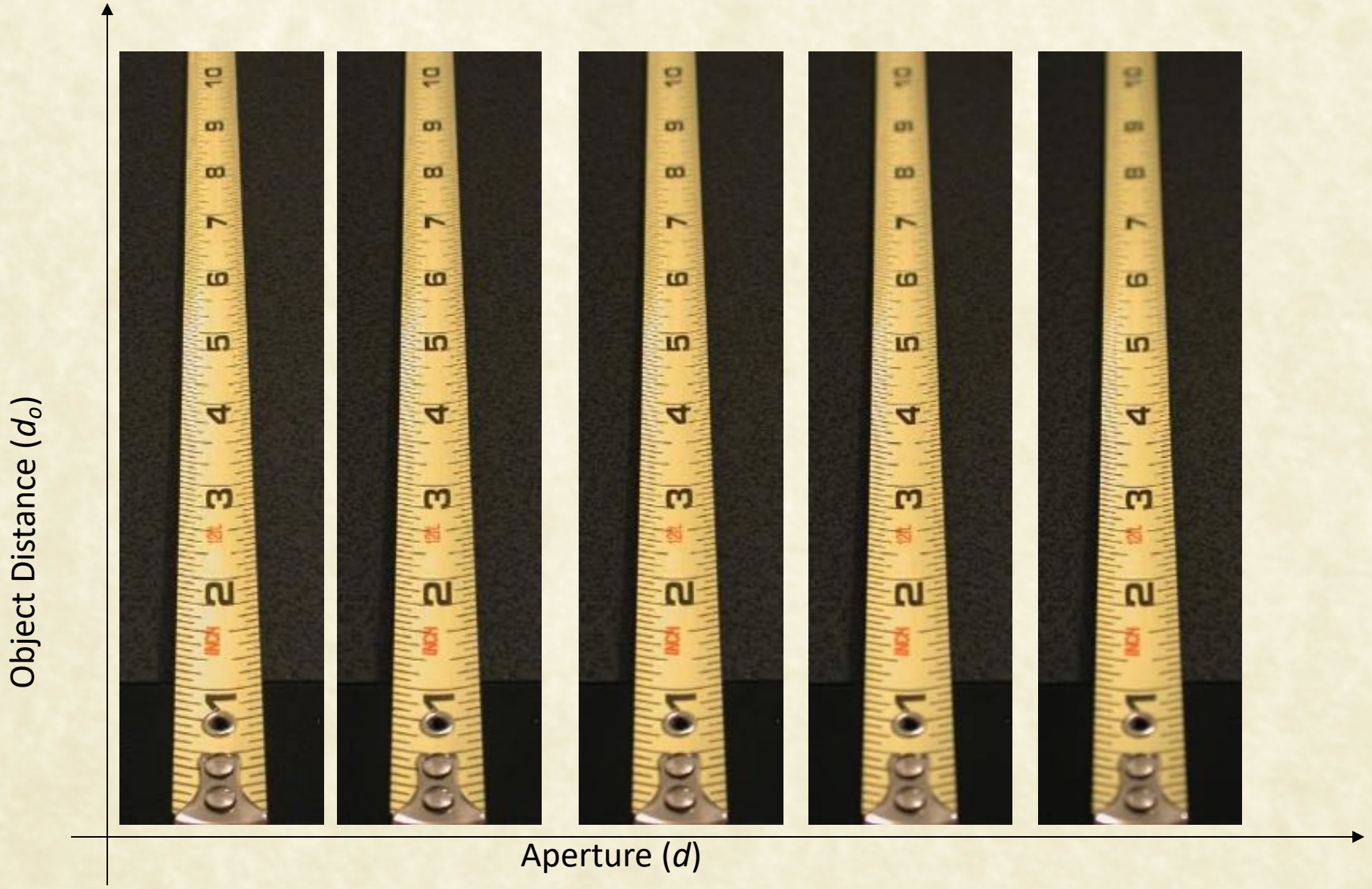
Focus and DOF

Aperture

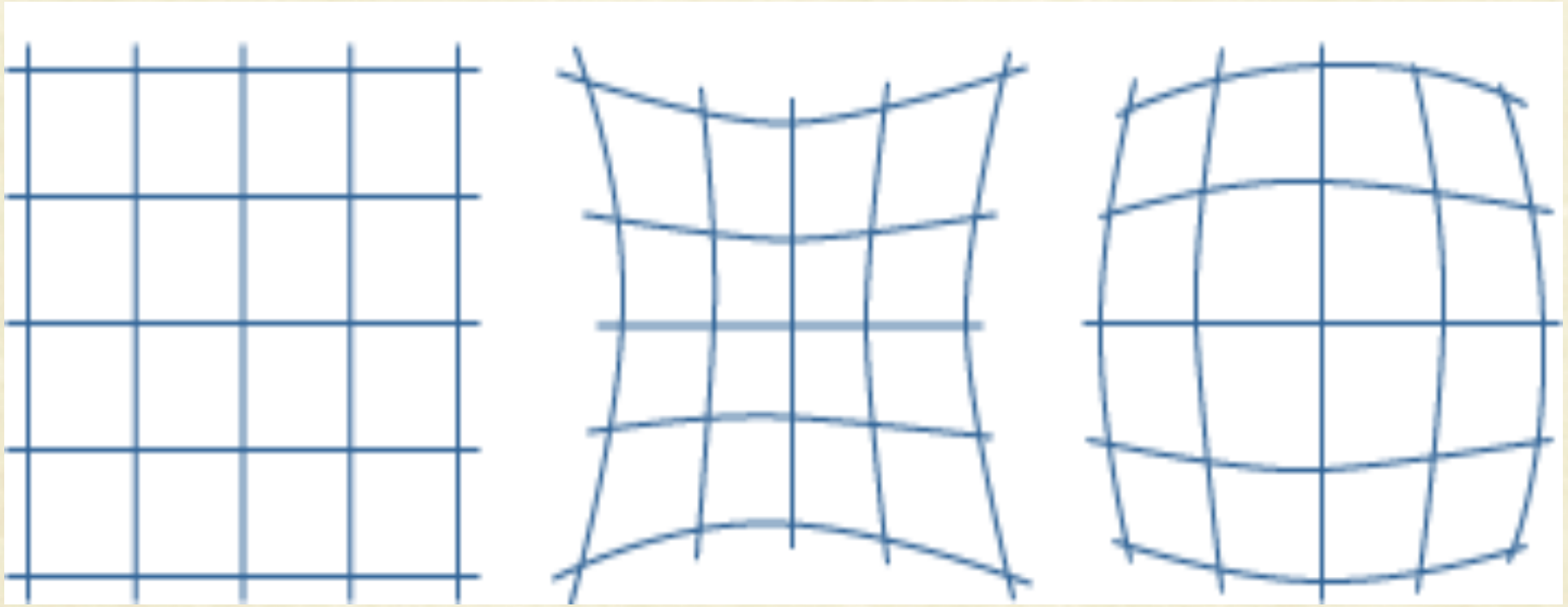


Focal Ratio = f / d

Aperture vs. DOF



Geometric Distortions



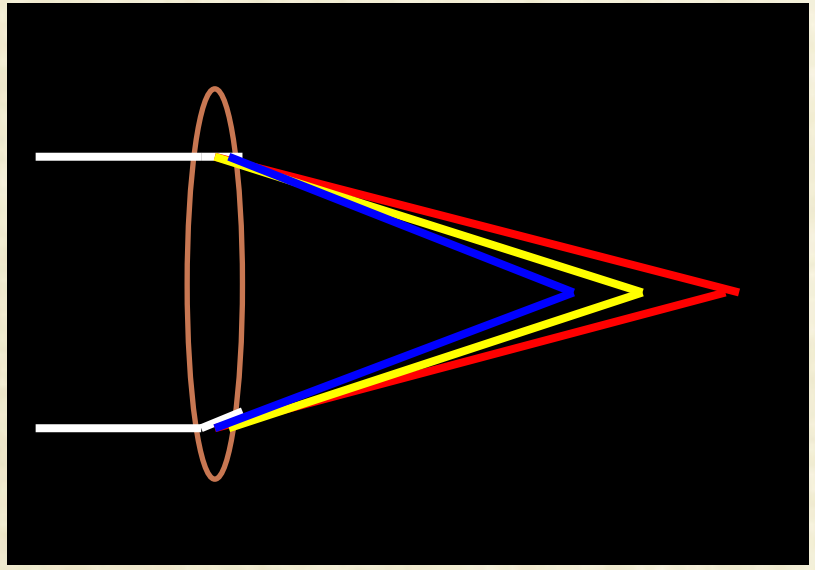
original pincushion barrel

Geometric Distortions

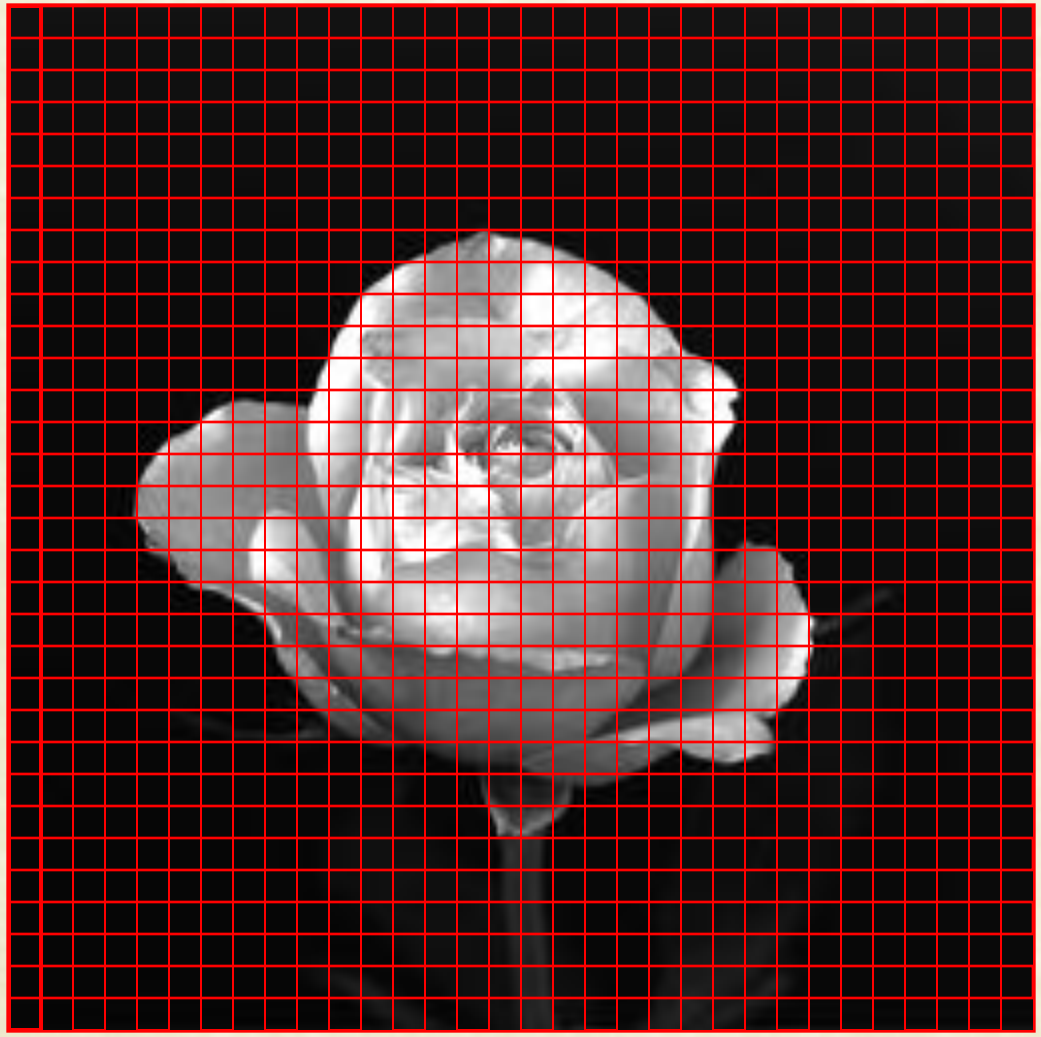
Lens Flare

Chromatic Aberration

Normal lenses diffract different wavelengths to different degree



Sampling an Image: Resolution



Resolution

- The number of samples in an image (number of sensor elements) is referred to as its resolution
- The resolution is typically represented as the product of number of samples in the horizontal and vertical directions in the image. e.g.: 32x32, 256x256, 640x480

Common Resolutions:

NTSC:	648 x 486
Typical Webcam:	1280 x 720
High-end SLR:	11,648 × 8,736 *
Hubbles Telescope:	1,600 x 1,600

Fujifilm GFX100

Camera Model: Objectives

- Mathematically model what a camera does
  - Also understand what the model means
- Getting the model for a real-world camera
  - Estimation from real world measurements
- Special imaging configurations with simpler properties
  - Simpler relationships
- General theory on fitting linear models under noisy observations
  - Techniques that work across problems

What does a Camera do?

- Form an image on the 2D image plane of the 3D world visible to it.
- Image is *behind* the lens; the scene is in front.
- 3D world is **projected** down to a 2D plane.
- Significant loss of information as one dimension is dropped.
- Mathematical depiction of this projection ...

Questions?