Introduction to Computer Vision for Robotics

AE640A Autonomous Navigation

5th March, 2019

Lecture Outline

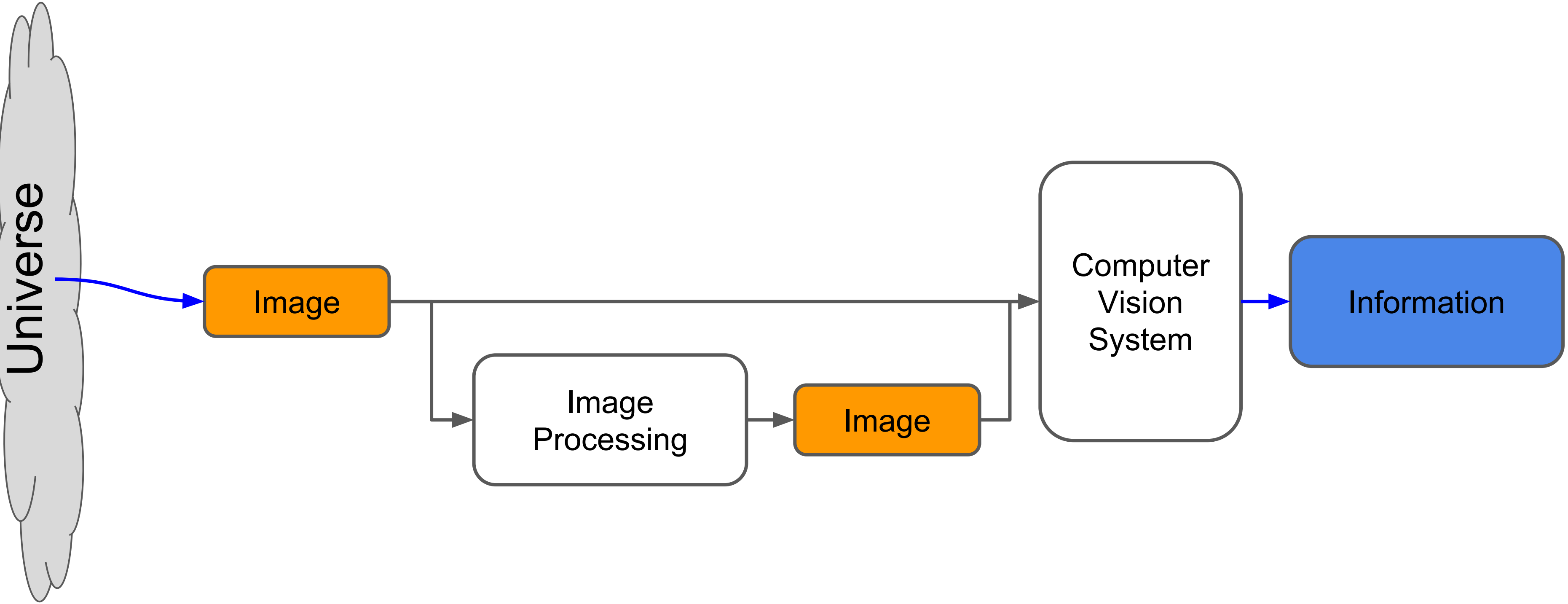
- Introduction
  - What is CV?
  - Overview of the field
  - A look at history
  - Hard Problem?
- Human Vision System & the Machine
  - The human vision system
  - Fooling humans
  - The computer vision system
- Images as matrices.
  - How cameras work to produce these matrices
  - Meaning of Intensity, Color etc
  - Shoutout to Image Processing

Lecture Outline

- Camera Model
  - Pinhole Camera Model
  - Intrinsic Camera Matrix
  - Camera Calibration

Introduction

What is Computer Vision?



What is Computer Vision?

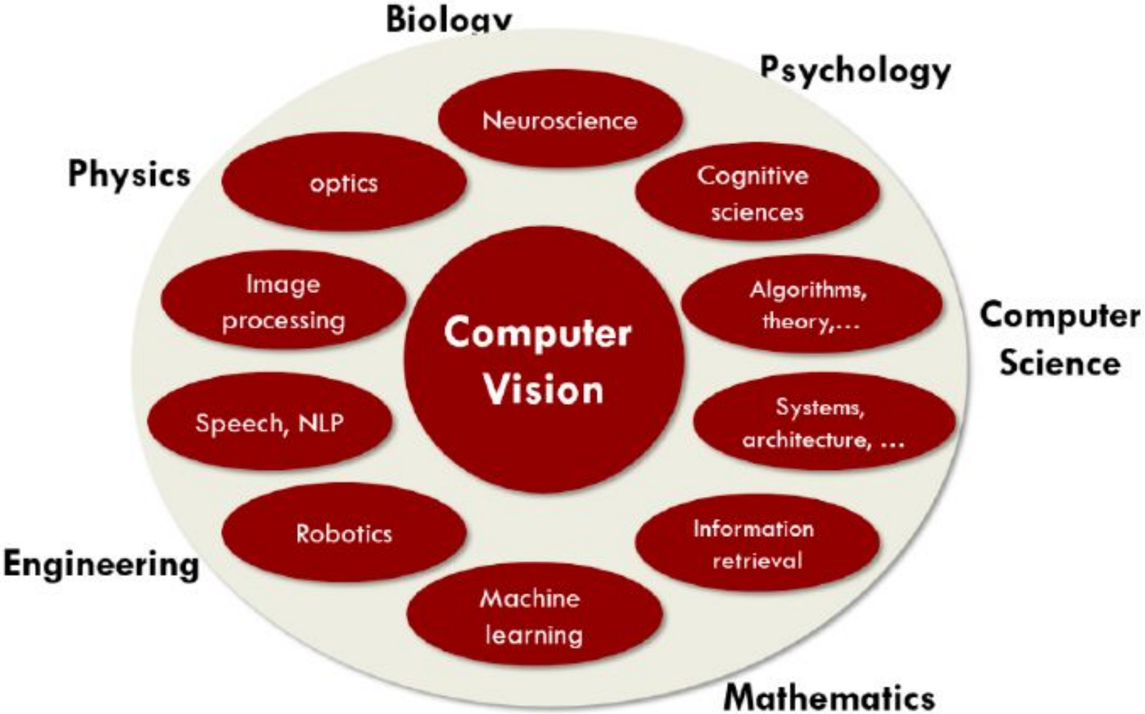


Image Credits: CS131, Fall ‘18, Stanford

What is Computer Vision?

- Computer Vision is deals with extracting information regarding the 3D world we live in using a single or a bunch of images.
- Computer Vision like most other fields today, is at the junction of numerous disciplines from Biology to Computer Science and has applications only limited by our imagination.

Overview of the field

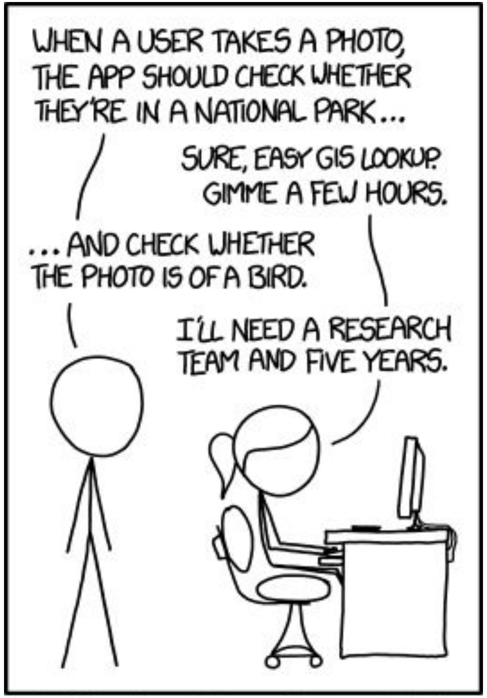


Image Credits: XKCD, 1425, 2014

Overview of the field

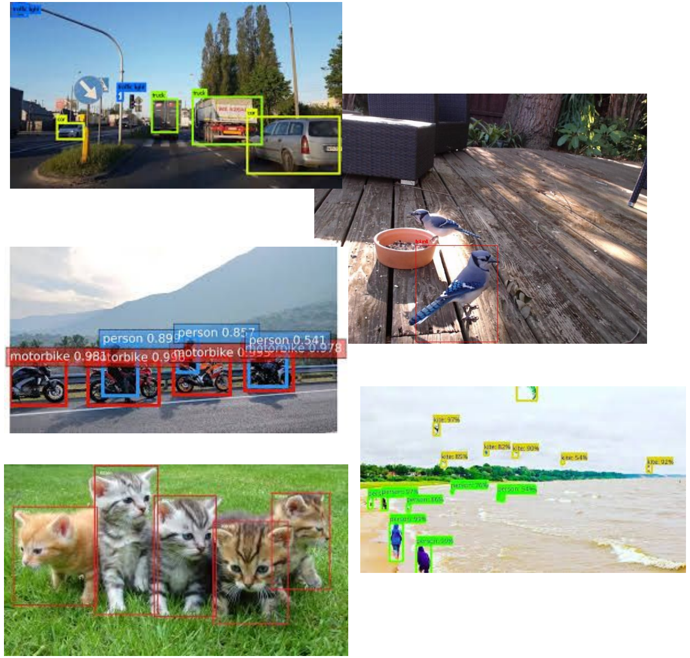


Image Credits: https://tinyurl.com/y53by9pr

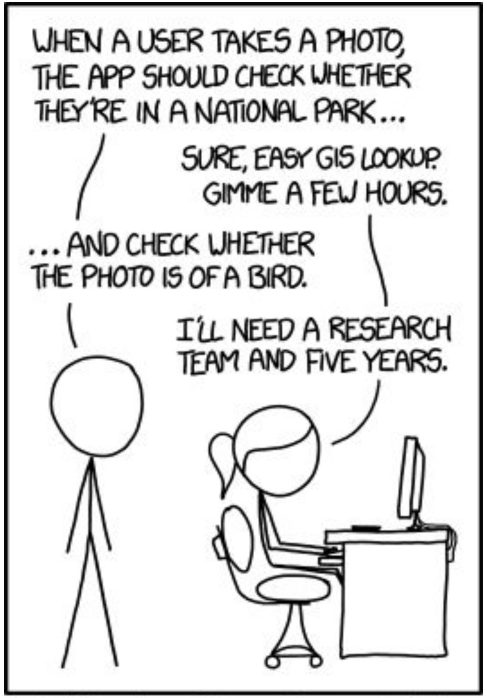
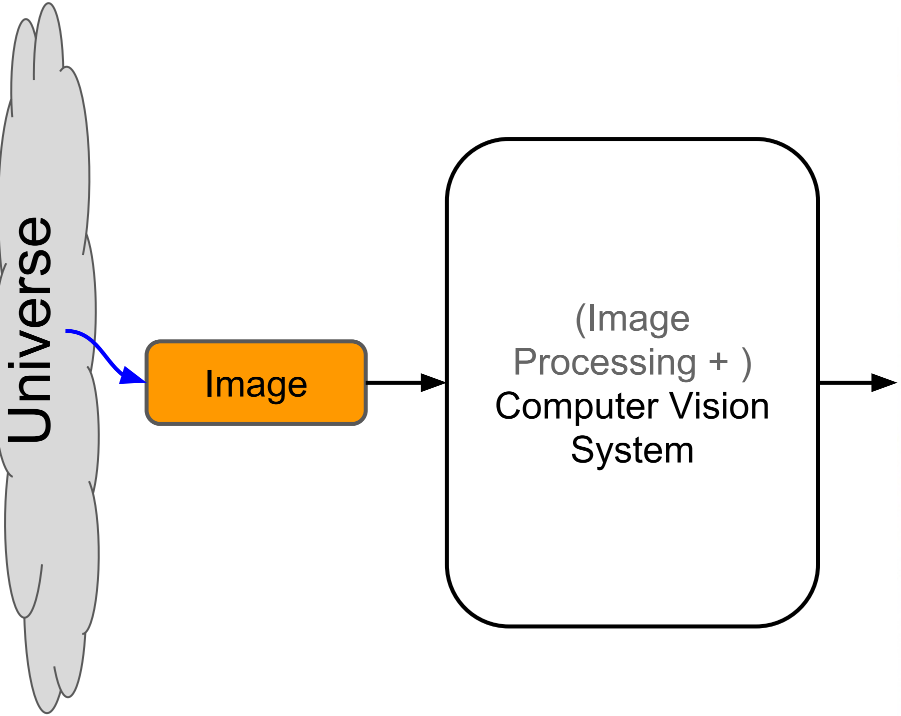


Image Credits: XKCD, 1425, **2014**

Image Credits: https://tinyurl.com/y53by9pr

Overview of the field



What kind of Information?

Overview of the field

Image Credits: Karpathy, CVPR'15

What kind of Information?

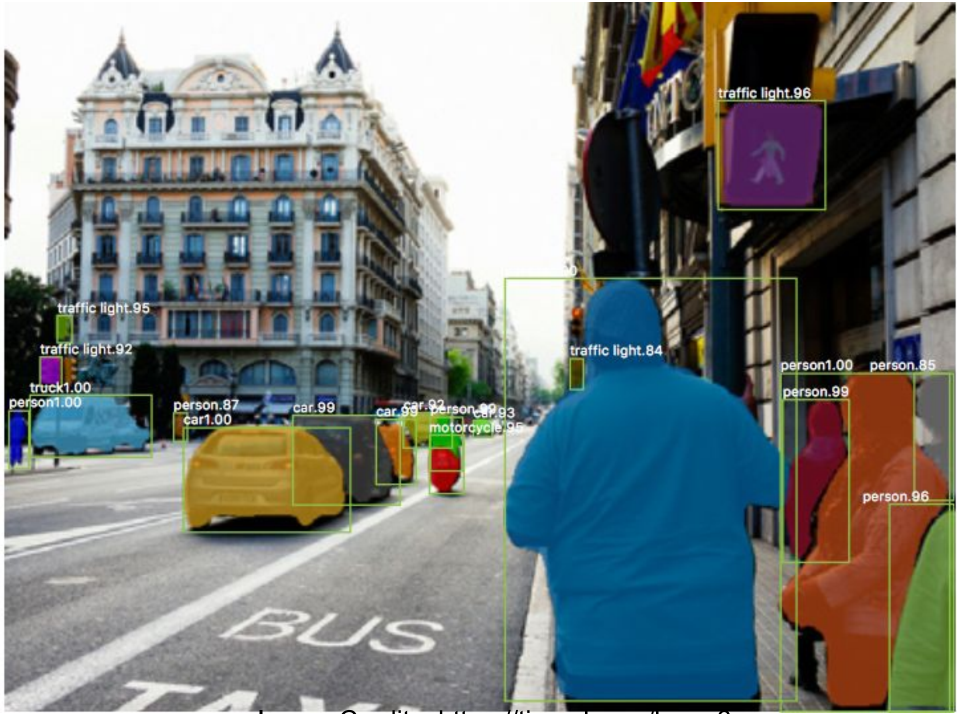


Image Credits: https://tinyurl.com/lxuex6o

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection

Recognition: Cat?

Image: https://tinyurl.com/yanp2o5e

Overview of the field

Primary themes in Computer Vision are:

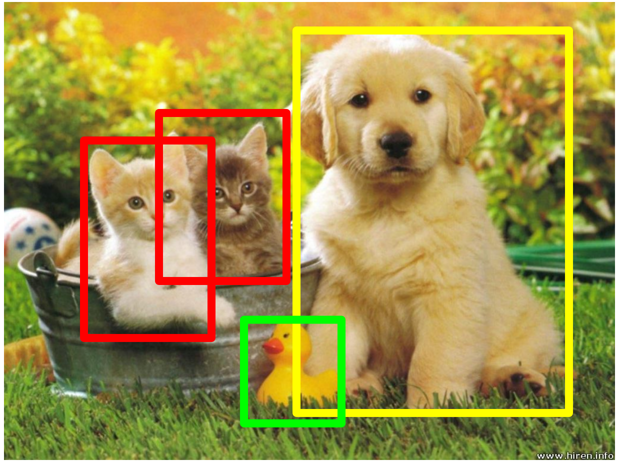
1. Object Detection

Recognition: Cat?
Localization: Where is the cat?
Image: https://tinyurl.com/yanp2o5e

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection



Object Detection: Which Objects are here and where?
Image: https://tinyurl.com/y4ly96rd

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection
2. Segmentation

Segmentation: Which pixels belong to which object?
Credits: Own Work

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements

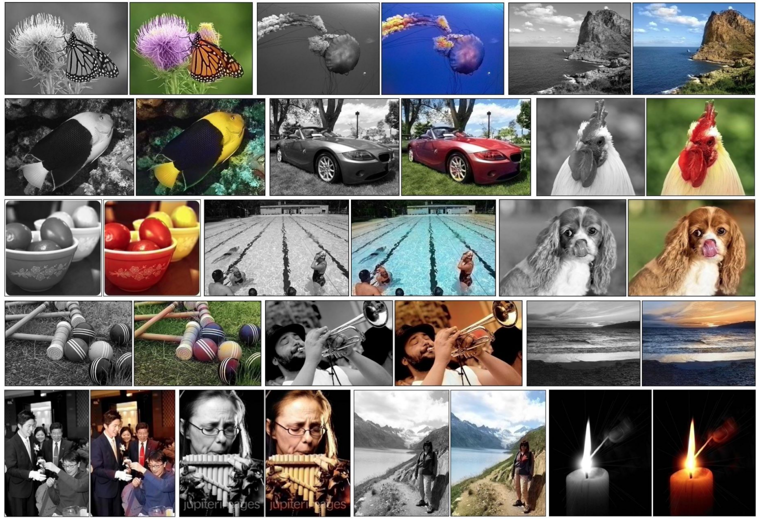


Image Colorization: From Grayscale to Colored Images
Credits: Richard Zhang, CVPR 2016

Overview of the field

Primary themes in Computer Vision are:

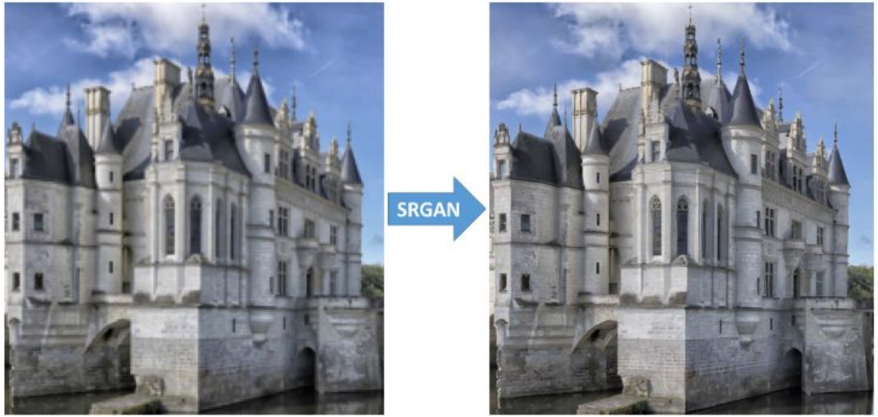
1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements

Image Enhancement: Real Time Image Enhancement
Credits: Michael Gharbi, ACM Graphics 2017

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements



Super Resolution: Upsampling Images while preserving quality
Credits: https://github.com/tensorlayer/srgan

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements
4. Image to Text

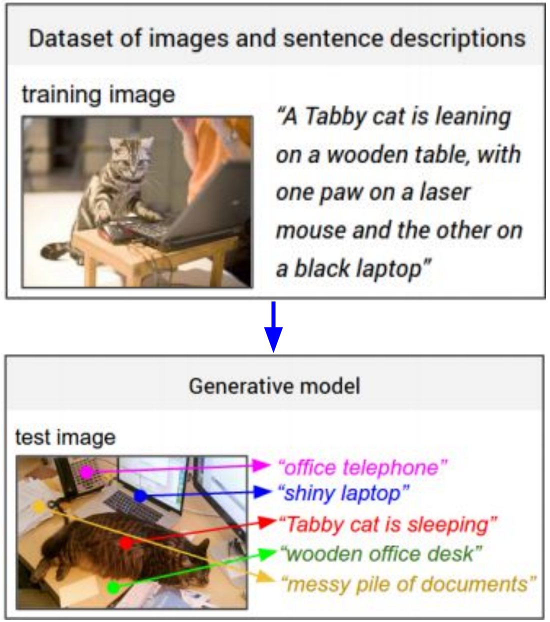


Image Description: Automatic semantic description for images
Credits: Karpathy, CVPR 2015

Overview of the field

Primary themes in Computer Vision are:

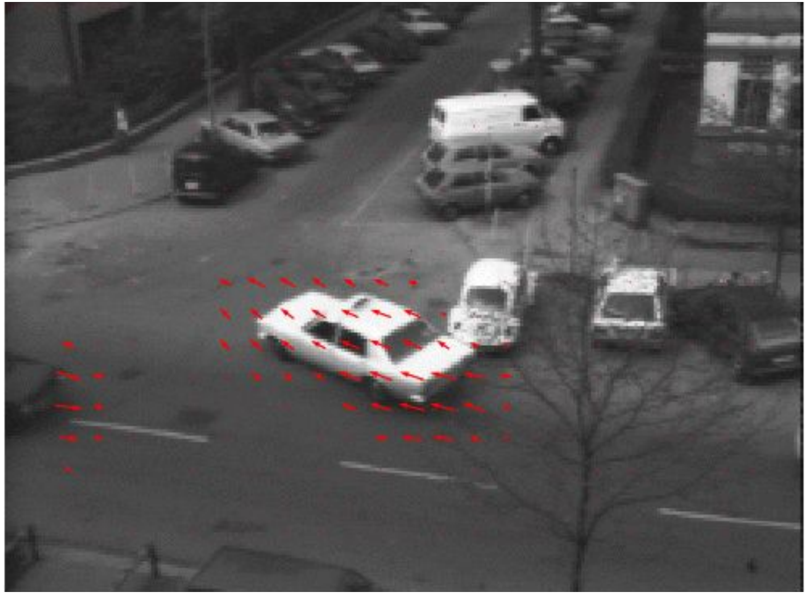
1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements
4. Image to Text
5. Image Generation

Image Generation: A style based generator architecture for GANs
Credits: Tero Karras, arXiv 2018

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements
4. Image to Text
5. Image Generation
6. Motion Estimation



Optical Flow: Lucas Kanade method for motion estimation
Credits: https://tinyurl.com/y5rloh3g

Overview of the field

Primary themes in Computer Vision are:

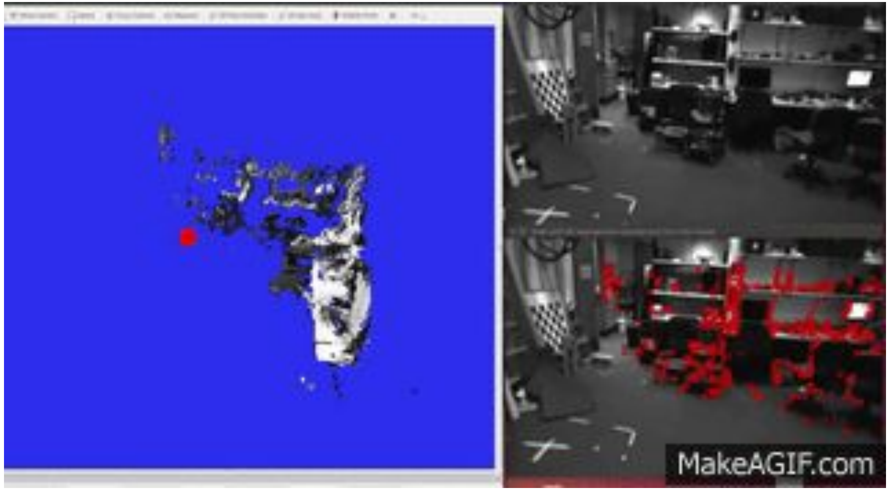
1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements
4. Image to Text
5. Image Generation
6. Motion Estimation
7. 3D reconstruction from Images

3D Reconstruction: REMODE,
Real Time Reconstruction
Credits: Matia Pizzoli, ICRA 2014

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements
4. Image to Text
5. Image Generation
6. Motion Estimation
7. 3D reconstruction from Images
8. Visual SLAM

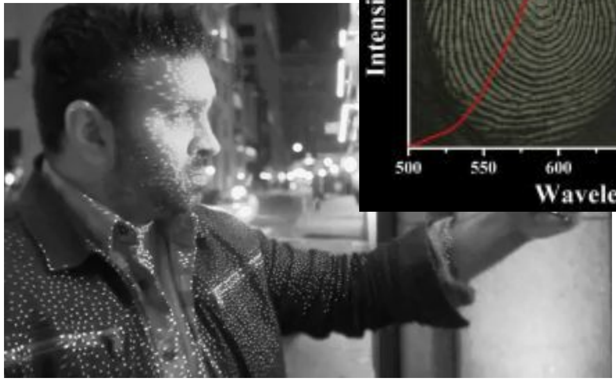


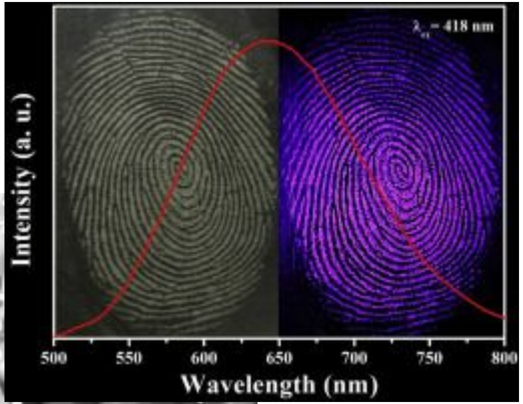
3D Reconstruction: REMODE,
Real Time Reconstruction
Credits: Matia Pizzoli, ICRA 2014

Overview of the field

Primary themes in Computer Vision are:

1. Object Detection
2. Segmentation
3. Image Modifications/Enhancements
4. Image to Text
5. Image Generation
6. Motion Estimation
7. 3D reconstruction from Images
8. Visual SLAM
9. Biometrics and more ...





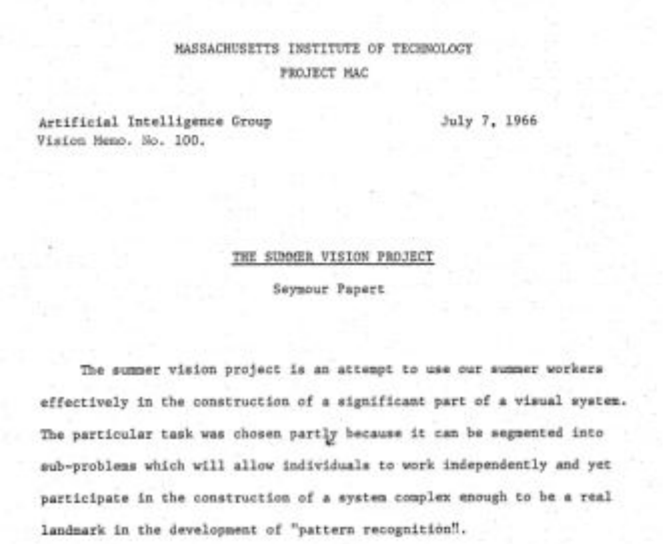
Biometrics : Fingerprint Detection, Apple Face ID

Credits:https://tinyurl.com/y2a7wybz, TheVerge Youtube

A look at history

- **Robert Nathan** started writing computer programs for enhancing images from NASA's spacecraft's at Jet Propulsion Lab, NASA.
- The Summer Vision Project: Project at MIT to **solve a significant part of visual system.** Primary Objective was to divide the image into object, background and chaos regions, **over the course of a summer.**

Credits: EE604, nasa.gov



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

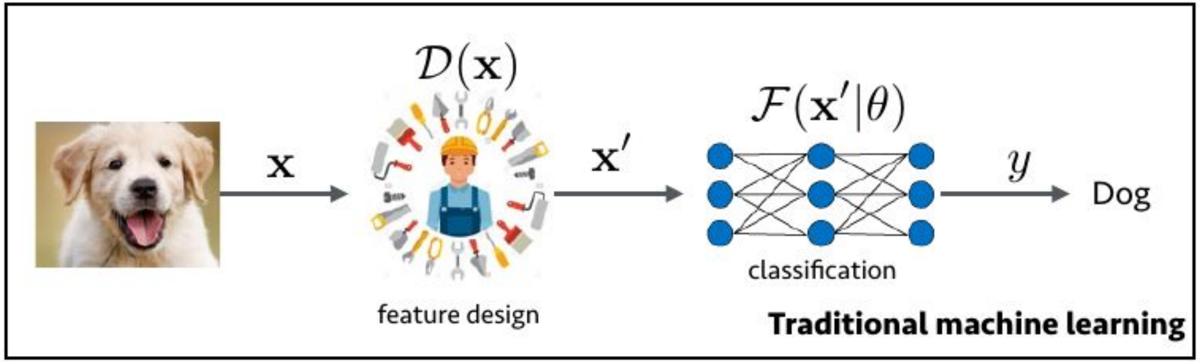
THE SUMMER VISION PROJECT

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".

Credits: https://tinyurl.com/y6bpo4nk

A look at history



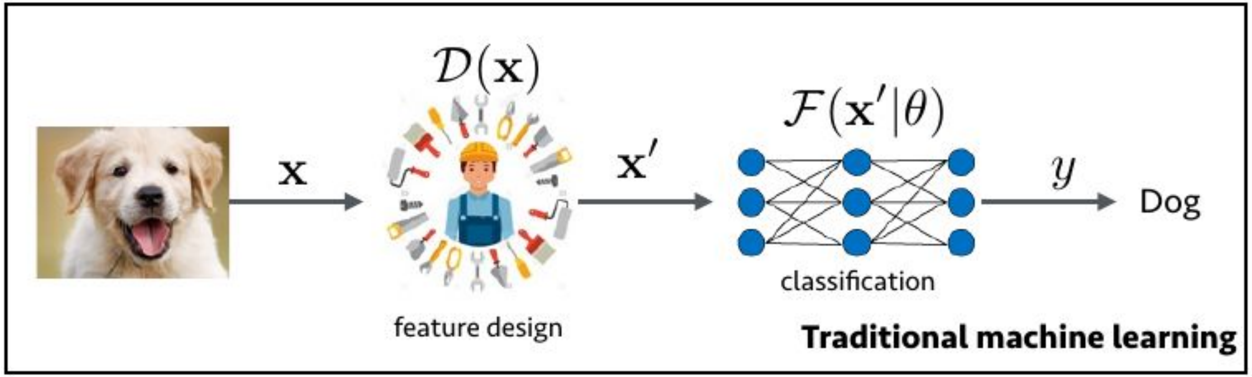
$\mathcal{D}(\mathbf{x})$

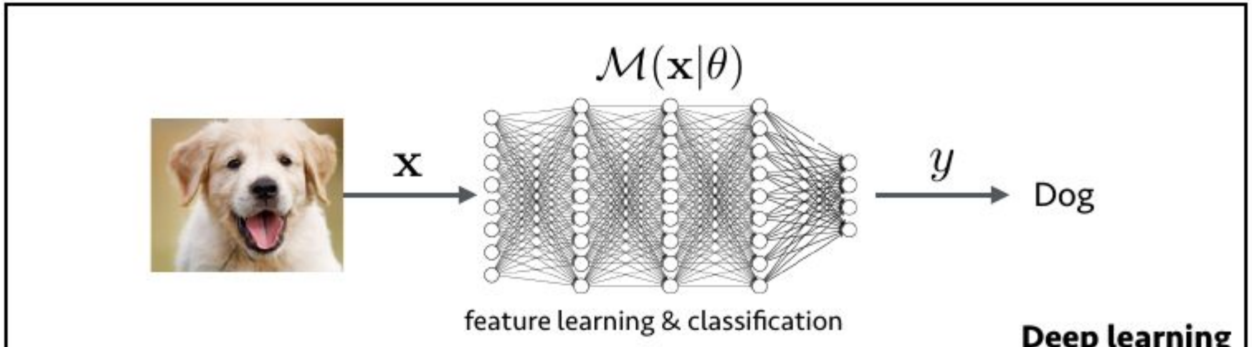
SIFT	SURF	TEXTON
STIP	HOG	GIST

Many more ...

Credits: Prof. Tanaya Guha, EE698K

A look at history





Credits: Prof. Tanaya Guha, EE698K

Hard Problem?

- Why are we still working on roughly the same problem as the “summer vision project”?
- Why is it that creating 3D models of chairs is easier than identifying them?

Hard Problem?

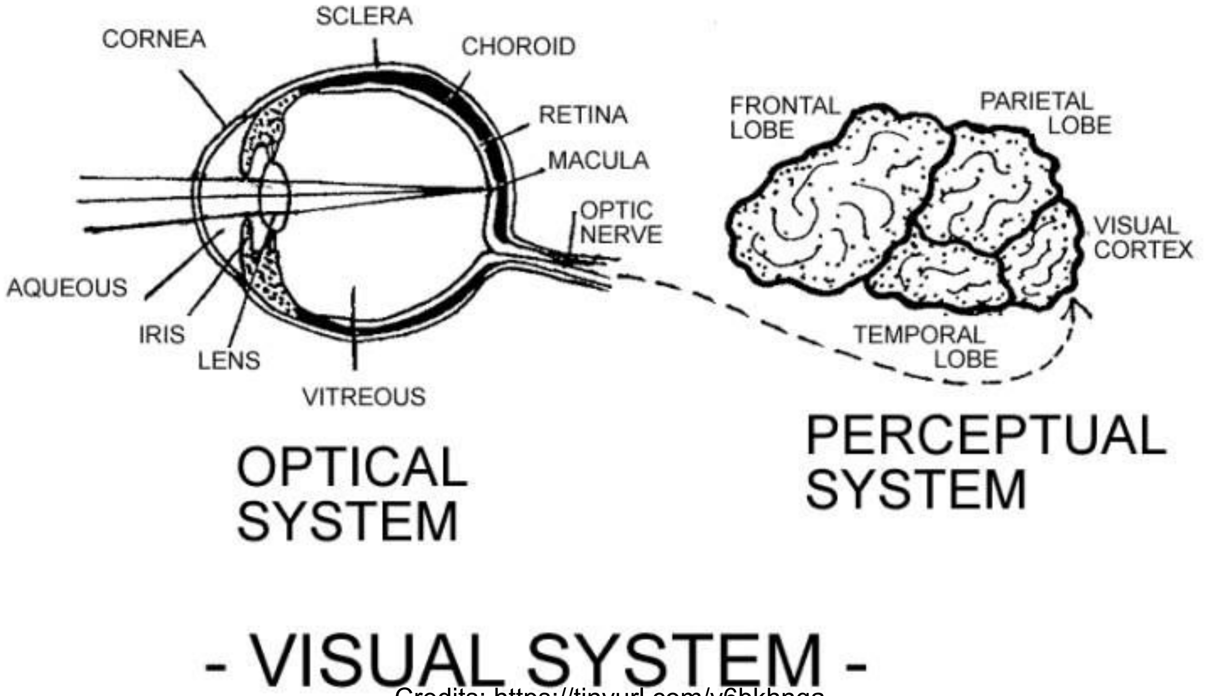
- Why are we still working on roughly the same problem as the “summer vision project”?
- Why is it that creating 3D models of chairs is easier than identifying them?

➔ There is a large between some ~1920x1080x3 numbers and the high-level abstract meaning we associate with them.

➔ Images are **2D** representation of information from **3D** world.

Human Vision System & Computer Vision System

The human vision system



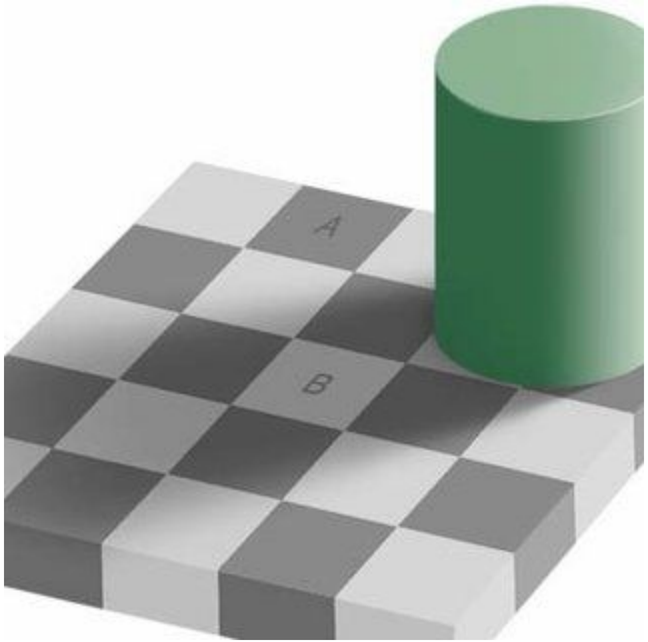
Credits: https://tinyurl.com/y6bkhnqa

The human vision system

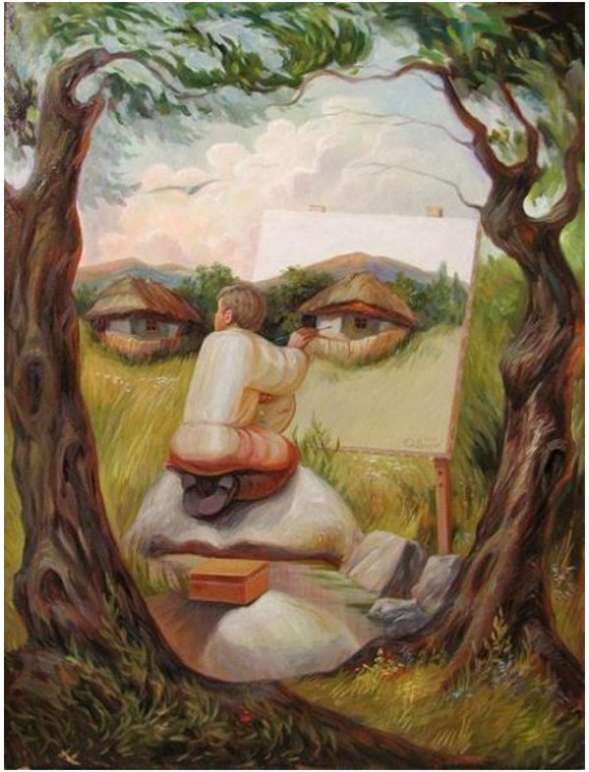
The human vision system

Credits: Ulas Bagci, UCF

Fooling humans



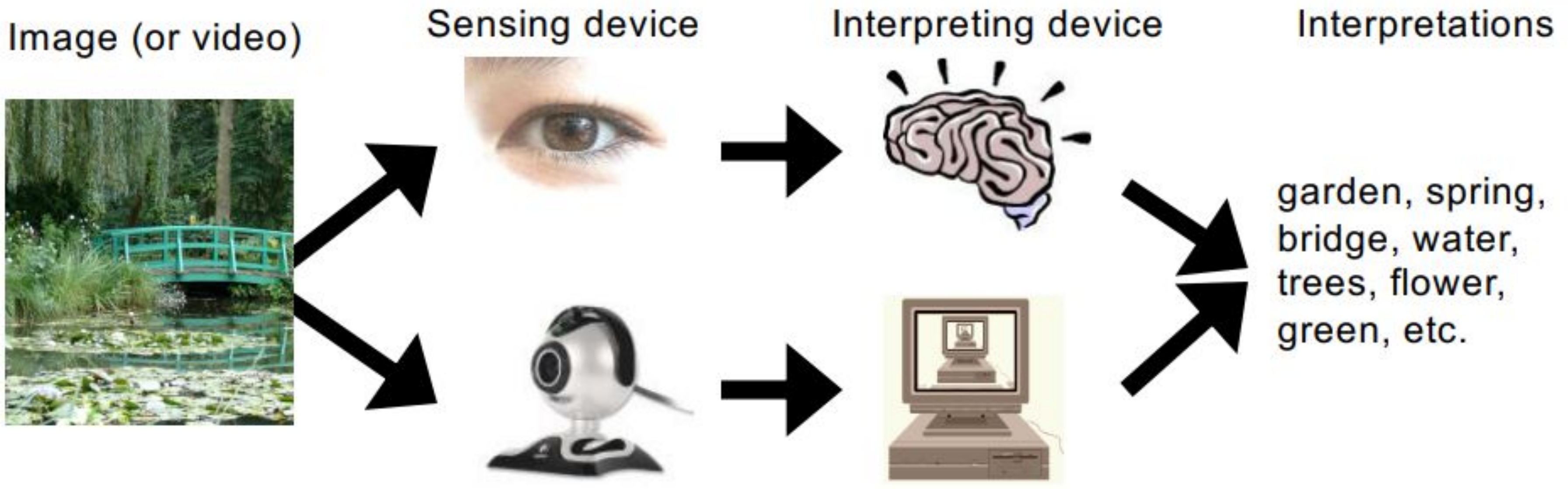
Credits: https://tinyurl.com/y49rp7sd



Credits: Oleg Shuplyak, Pinterest

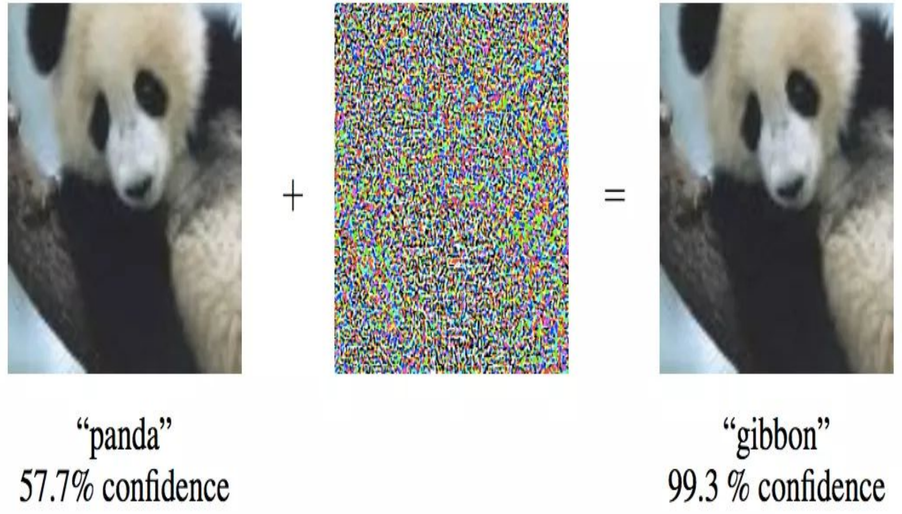
Credits: Wikipedia, Spinning Dancer

The ~~(human) (computer)~~ vision system

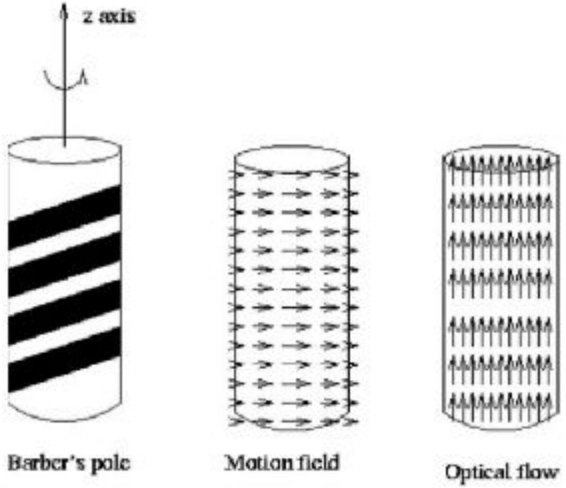


Credits: CS131, Stanford

Fooling Computers



Credits: https://tinyurl.com/l5pwp6t



Credits: Wikipedia, Barber Pole Illusion

Images as Matrices

Camera Models

Camera Models

Credits: https://tinyurl.com/y6qen2vb

Camera Model~~s~~

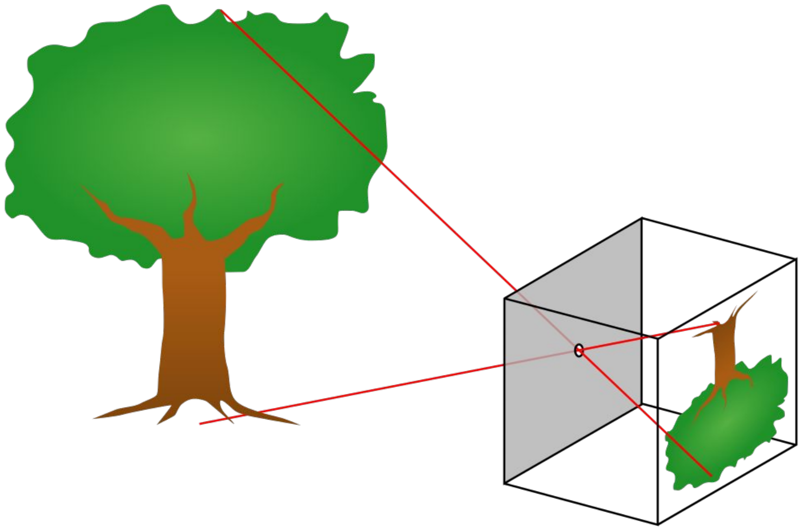
Not this one but models as in modelling a phenomena

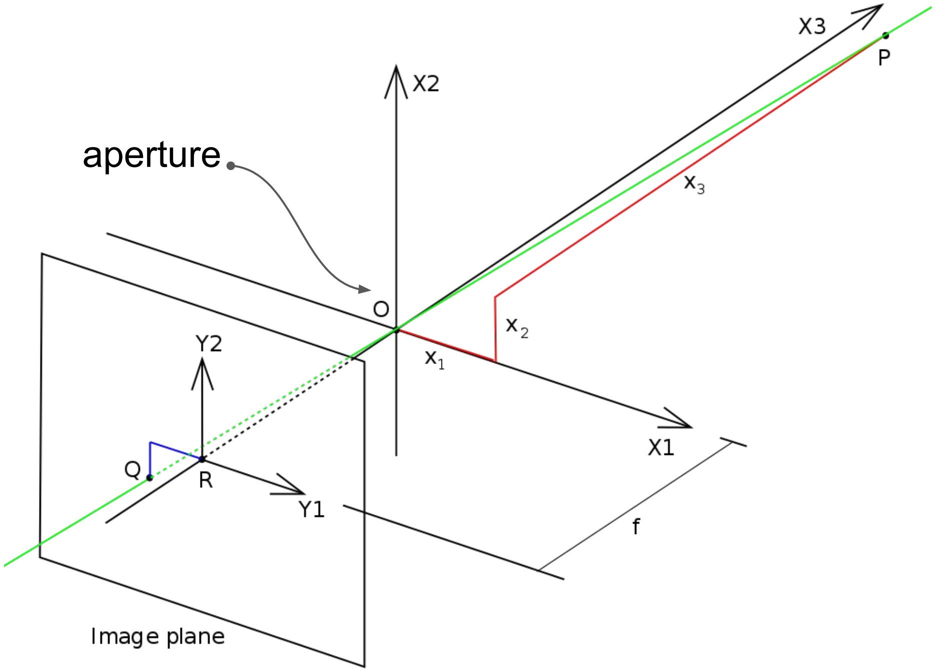
Credits: https://tinyurl.com/y6qen2vb

Camera Models

- Like so many things in engineering, we create a simple “model” of a camera to which is easy to understand and can approximate the actual functioning of a camera to a good degree.
- There are different models:
    - Pinhole camera model
    - Lens model
    - ...

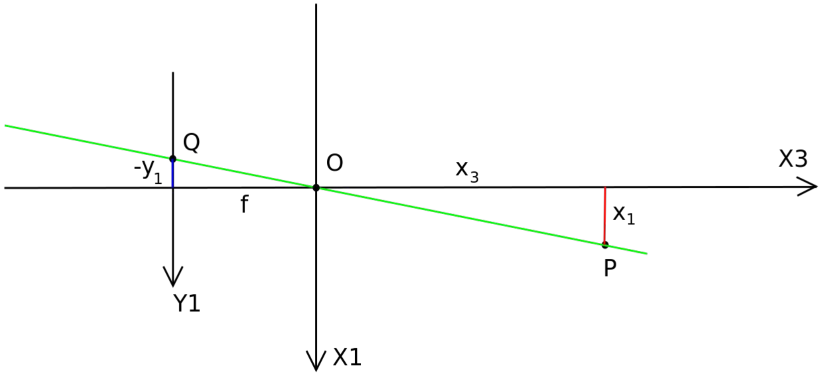
Pinhole camera model

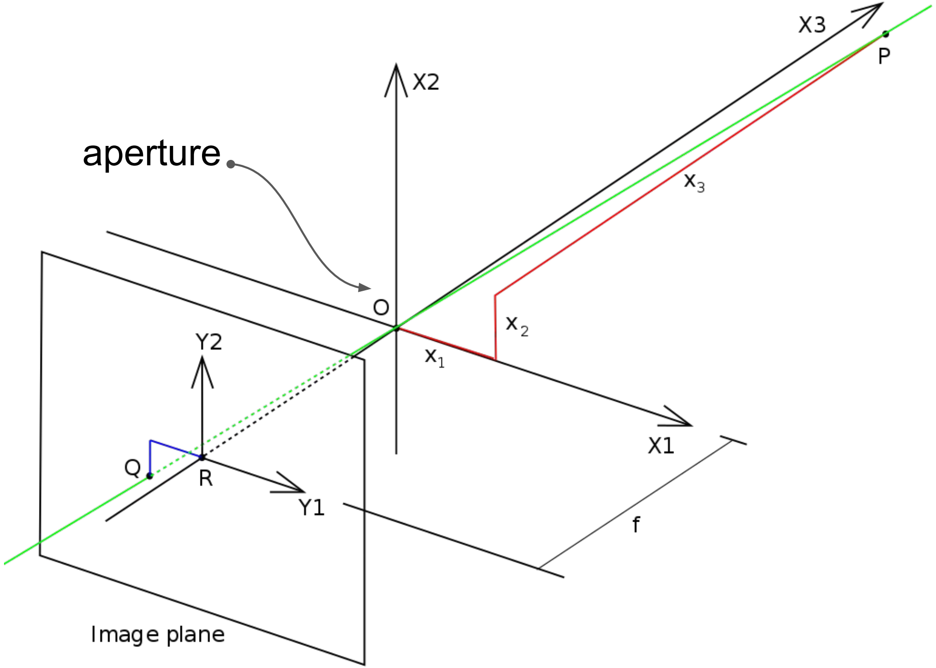




Credits: Wikipedia, Pinhole Camera Model

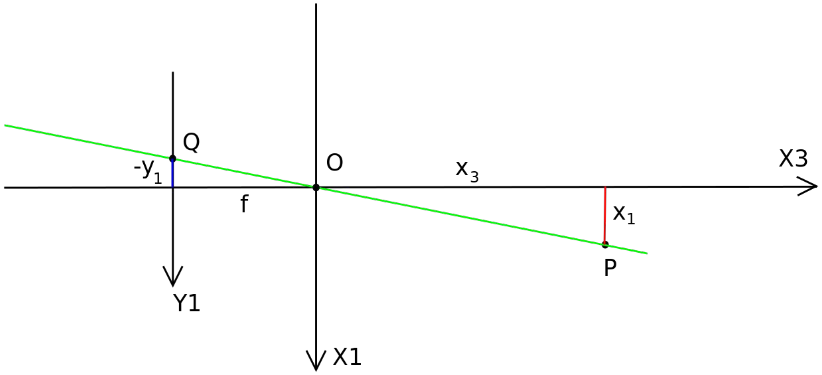
Pinhole camera model



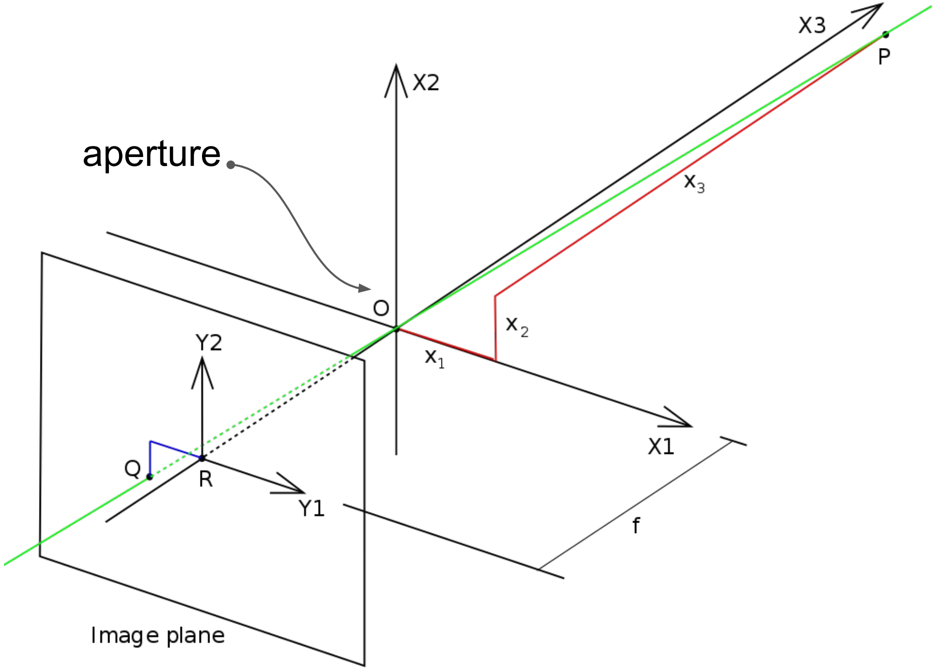


Credits: Wikipedia, Pinhole Camera Model

Pinhole camera model



$$y_1 = -\frac{f\,x_1}{x_3}$$



Credits: Wikipedia, Pinhole Camera Model

Pinhole camera model

$x' = -xf/z$ where $x'_i = y_i$ and $z = x_3$

$x' = fx/z + c$ where c is an offset in pixels

$$p' = \begin{bmatrix} fx/z + c_x \\ fy/z + c_y \end{bmatrix} \qquad p = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Can we make this into a matrix multiplication of the form **p' = Mp**?

Intrinsic camera matrix

- We'll introduce a new convention, *homogenous coordinates*.
- We write points just the way we did before, but add an extra row:
  - The extra row is a *scale factor* for the whole vector.

$$p = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{becomes} \quad p = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Credits: Edwin Olson, University of Michigan

Intrinsic camera matrix

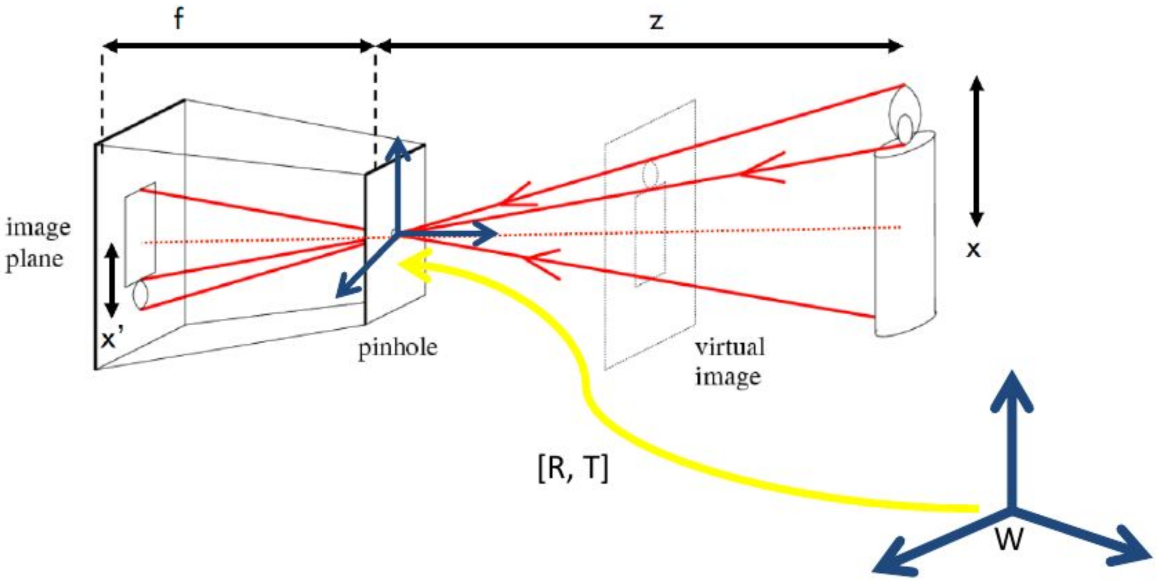
- Eureka!

$$p' = \begin{bmatrix} fx/z + c_x \\ fy/z + c_y \\ 1 \end{bmatrix} = \begin{bmatrix} fx + c_x z \\ fy + c_y z \\ z \end{bmatrix} = \begin{bmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

2d homogeneous coordinates — 2d homogeneous coordinates — perspective transform — 3d homogeneous coordinates

Credits: Edwin Olson, University of Michigan

Intrinsic camera matrix



Credits: Edwin Olson, University of Michigan

Intrinsic camera matrix

- The product of two rigid-body transformations is ***always*** another rigid-body transformation!
- So no matter how the object has been translated or rotated, we can describe its position with a single 4x4 matrix, which has the structure:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} R_{00} & R_{01} & R_{02} & T_x \\ R_{10} & R_{11} & R_{12} & T_y \\ R_{20} & R_{21} & R_{22} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Credits: Edwin Olson, University of Michigan

Intrinsic camera matrix

$$\begin{bmatrix} x' \\ y' \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{00} & R_{01} & R_{02} & T_x \\ R_{10} & R_{11} & R_{12} & T_y \\ R_{20} & R_{21} & R_{22} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

2d homogeneous pixel (camera) coordinates

Focal length and focal center of camera

Rigidly move every object in the world to simulate the camera's true position

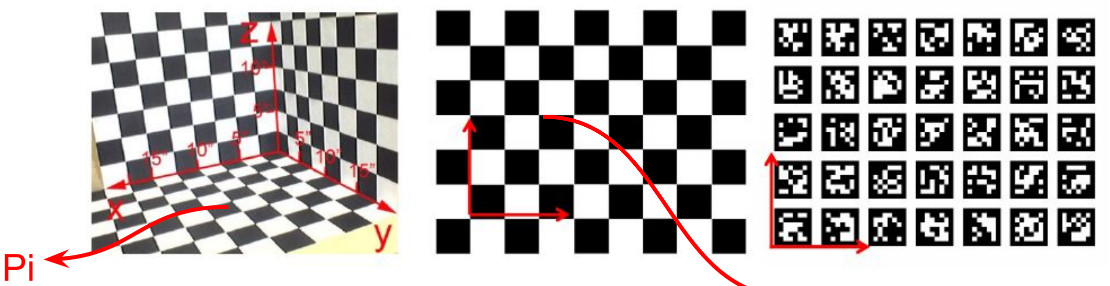
3d homogenous (world) coordinates

"Intrinsics"

"Extrinsics"

Credits: Edwin Olson, University of Michigan

Camera calibration



Calibration Rig Image

$$M = \arg\min_{M} \sum \|P_{ci} - MP_i\|^2$$

$$P_{ci} \Leftrightarrow P_i$$

Credits: Gaurav Pandey, Ford