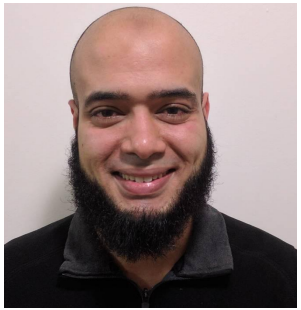
Machine Learning

Intro to Computer Vision Problems

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)



Video

- **See full arabic lecture on my youtube**

Problems of interest during the course

- There **might** be some tasks that requires you know:
- Image classification
- Video classification
- Object Detection
- 2D/3D body pose estimation

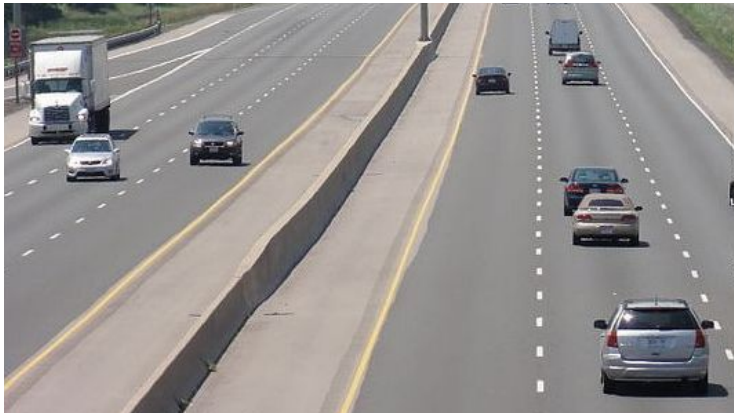
Image

- Grayscale image: 2D matrix (height x width)
  - Array position image[row, col] is a pixel
  - Each pixel represents intensity information in **range** 0 (for black) up to 255 (for white)
  - Binary Image: has only 2 values for black and white (e.g. 0 and 255)
- RGB image: 3D matrix (height x width x 3 channels)
  - Access: image[row, col, channel]
  - Other color spaces: HSL and HSV, CMYK, CIELAB. Conversions.
- Video = Sequence of frames (images)

Image: RGB vs Gray

Src: Article

Image: RGB vs Binary Image



Black for Background

White for Foreground

Src: Article

The goal of computer vision

- To bridge the gap between pixels and “meaning”

What we see

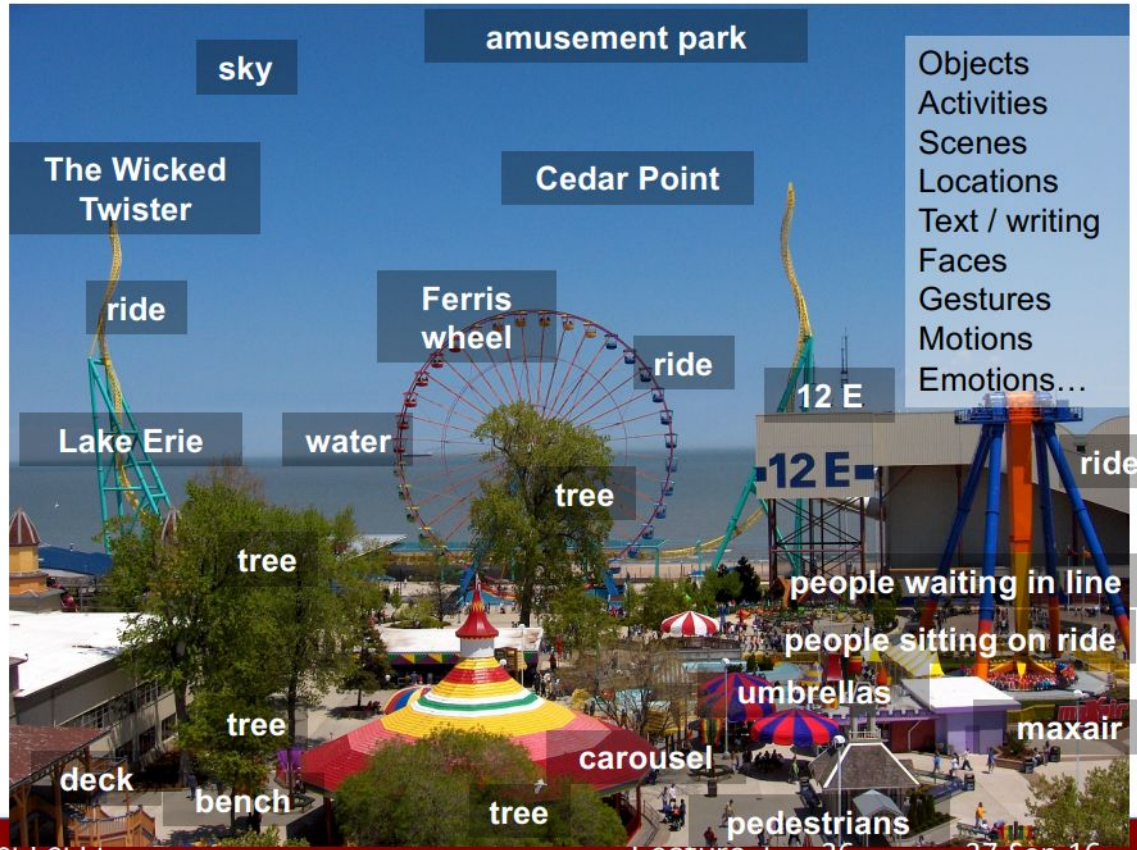
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Source: S. Narasimhan

Src: Stanford CS 131

Vision as a source of semantic information



Slide credit: Kristen Grauman

2D and 3D Computer vision

- Both are important. Both receives input of 2D Images
- 2D models understands images based on given 2D positions
- 3D models make use of multi view / depth
  - E.g. Building Depth for the 2D view or Building 3D model/coordinates
  - In some problems require Camera parameters or several views of same scene
- 2D real-life scenarios/research seems more
  - Nature of several apps just understand given image/video
  - All these uploads on the web don’t provide camera parameters
- RGB-**D** images (D for depth channel)
  - Depth of distance between image plane and corresponding object in RGB image
  - Now more RGB-D Smartphones and Tablets (Useful for apps such as AR/VR)

RGB-D Example

Src: List of RGBD datasets

2D Computer vision problems

- Images
  - **Image Classification**
  - **Object Detection**
  - Semantic Segmentation and Instance Segmentation
  - Edge Detection
  - Human Pose Estimation
  - More
- Videos
  - Action Recognition and Action Localization
  - Object Tracking
  - Group Activity Recognition Problem
  - More

3D Computer vision problems

- Stereo Vision
- 3D reconstruction
- Structure-from-Motion and SLAM
- Depth Estimation
- Pose Estimation
- Panorama Stitching
- Optical Flow

2D Vision - Image Problems

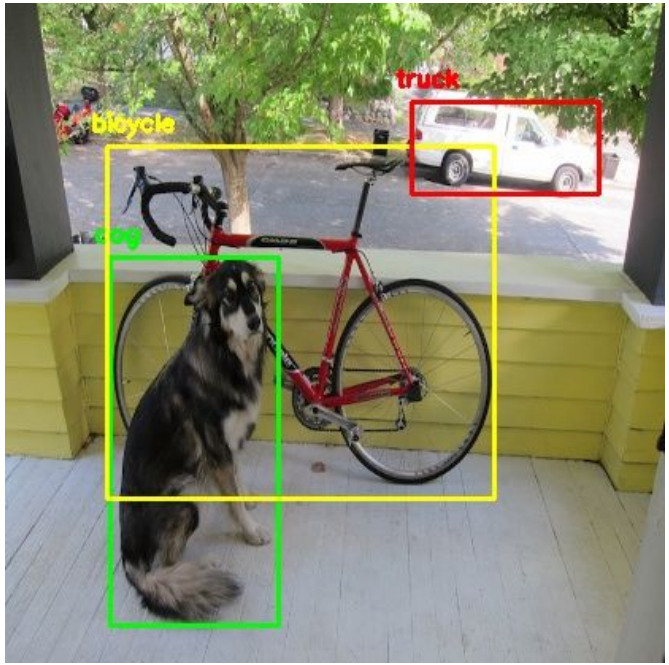
Problem: Image Classification

- Let say we have 1000 classes of interest
  - E.g. Cat, Dog, Chair, Car, BMW Car, Bird, ...
- Given an image: Identify its major class (e.g. Image for Leopard)

Src: CMU course

Problem: Object Detection

- Now harder problems.
- Let say we have objects of interest
  - e.g. Cat, Chair, Cow, Bus, …
- Given an image, return:
  - rectangles for their positions
- Aka Object **Localization**
  - Sometimes localization query has specific number of items. E.g. retrieve 3 cars
- Object Proposals



Src: Tutorial

Problem: Semantic Segmentation

- Given an image, for each pixel decides its class

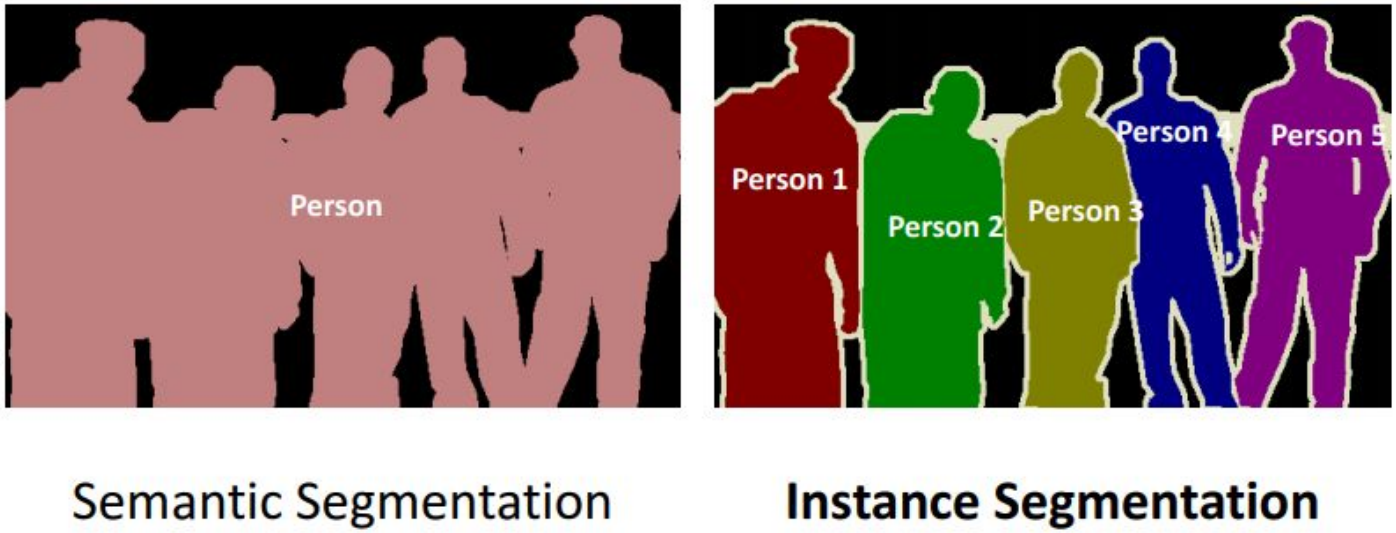
predict

Person
Bicycle
Background

Src: Tutorial

Problem: Instance Segmentation / Panoptic

- Same as previous, but identify the instance of each category



Src: Tutorial

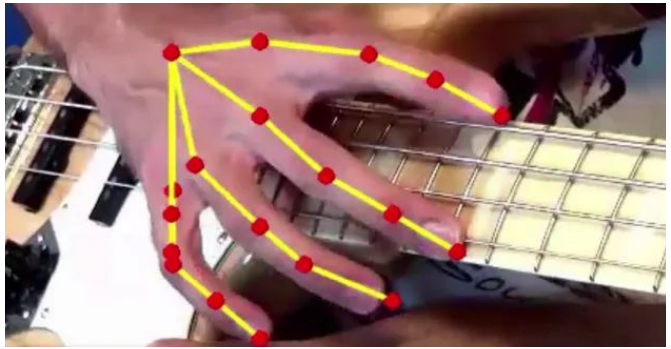
Problem: Edge Detection

- No object of interests. Identify the boundaries/borders

Src: Article

Problem: Human Pose Estimation

- Given an image of people, for each person identify his body joints (specific e.g. wrist/shoulder)
- Similar task: Hand pose estimation
  - Find 21 joints of hand (e.g. use for sign language)



Src: learnopencv.com

Face Recognition & Identification

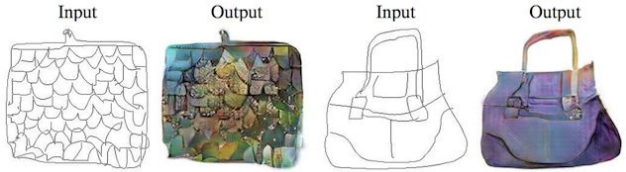
- Recognition: Find a face
- Identification: Who is this face?
- Authentication: Is this face for mostafa?

Src / Src

Crowd counting

Src / Src

GANs



Src: machinelearningmastery.com

Problem: Image Captioning

- Given an image ⇒ generating textual description
  - CV and NLP intersection problem

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

Src: [towardsdatascience.com]

Problem: Visual Question Answering

- Given an image and question: Answer it (CV/NLP)

Who is wearing glasses?

man

woman

Where is the child sitting?

fridge

arms

Is the umbrella upside down?

yes

no

How many children are in the bed?

2

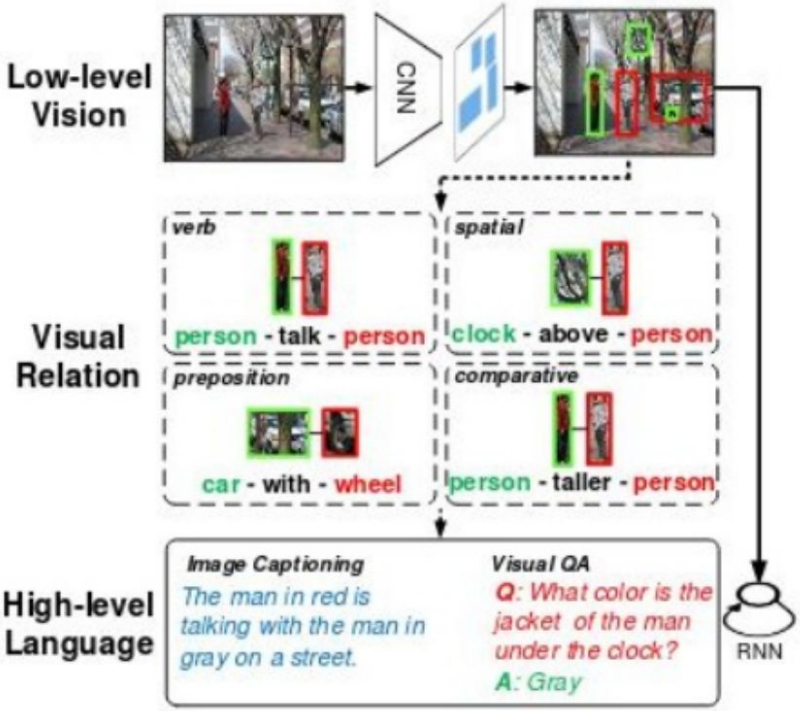
1

Src: [mc.ai](mc.ai)

Visual Relation Detection

- Modeling and understanding the relationships between objects in a scene (i.e. “person ride bike”).
- Better generalization for other tasks such as image captioning or VQA.
- Visual relations are *subject-predicate-object* triplets, which we can model jointly or separately.

Src



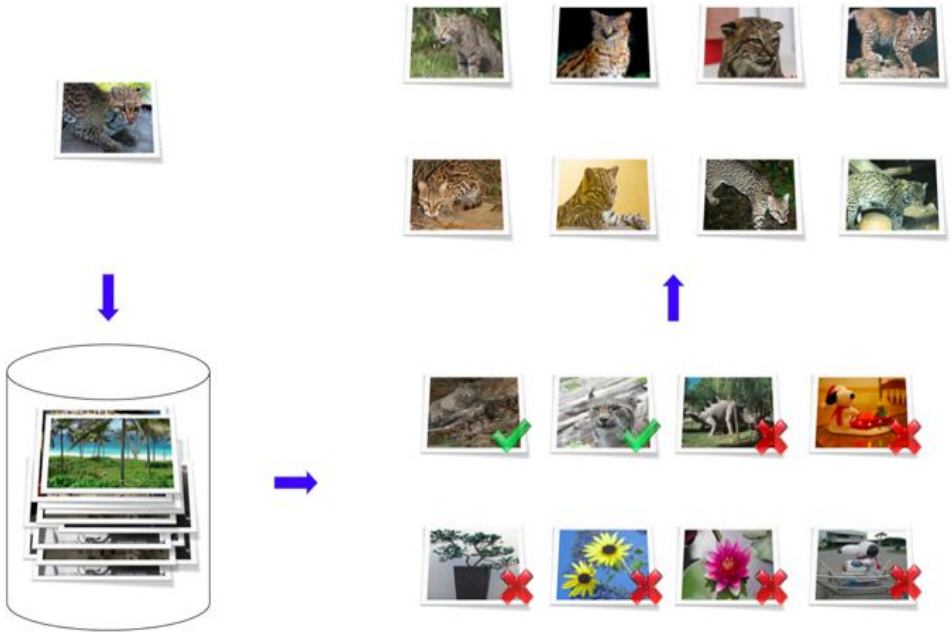
Problem: Image inpainting

- Given an image and bounding box of object
  - Remove the object and replace with background
  - Useful in apps such as Photoshop, Films making, removing someone from your photos

Src: paper

Problem: Content-based Image Retrieval

- Assume **dataset** of images
- Query: Image to find similar ones in the database
- Output: **Rank** all dataset images according to their **similarity** with query

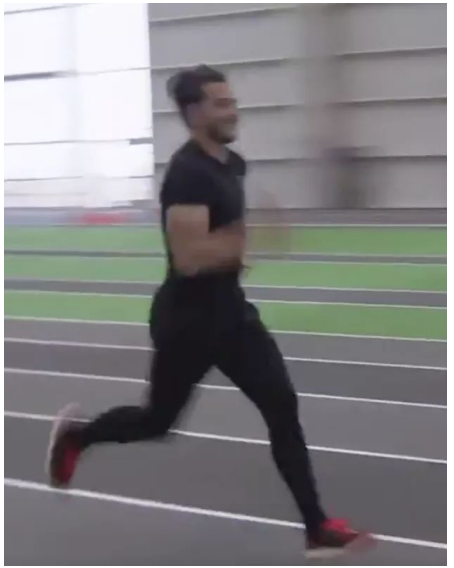


Src: Article

2D Vision - Video Problems

Problem: Action Recognition

- The video version of Image Classification
- Action: Sequence of Simple steps (Running)



Src: Site

Problem: Action Localization

- The video version of Object Localization
- We find a tublet (aka trajectory = set of consecutive bbox)
- Find **action** of each tublet

Problem: Object Tracking

- We track objects based on their appearance
- We don’t label the actions
- If a human: might do several kind of actions: walk, run, jump

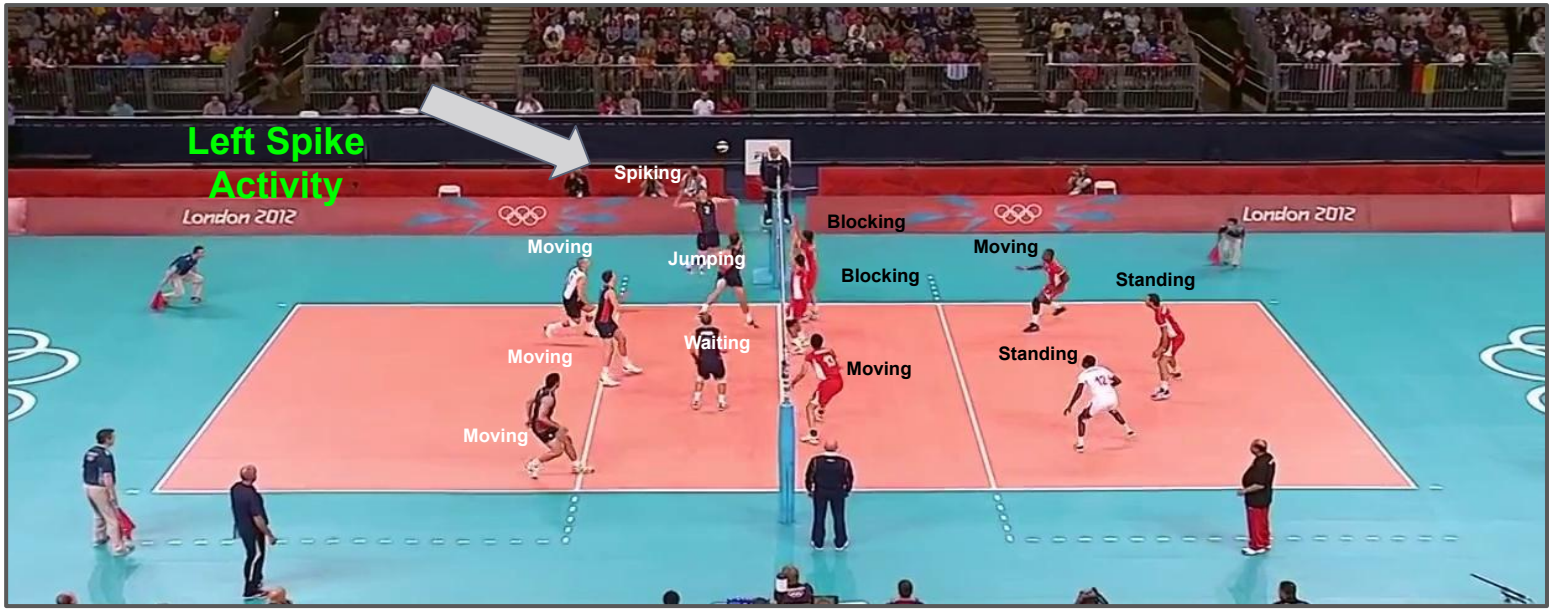
Problem: Group Activity Recognition Problem

Group Activity = Major Action = Walking

Collective activity, Choi et al, ICCV Workshop 09

Problem: Group Activity Recognition Problem

Group Activity = Key Action(s) = Left Spike



Volleyball Dataset, Ibrahim et al, CVPR 16

Person Re-identification

When someone disappears and come back, we wanna still link with the old person

Video Prediction

What will happen in the next 10 frames?

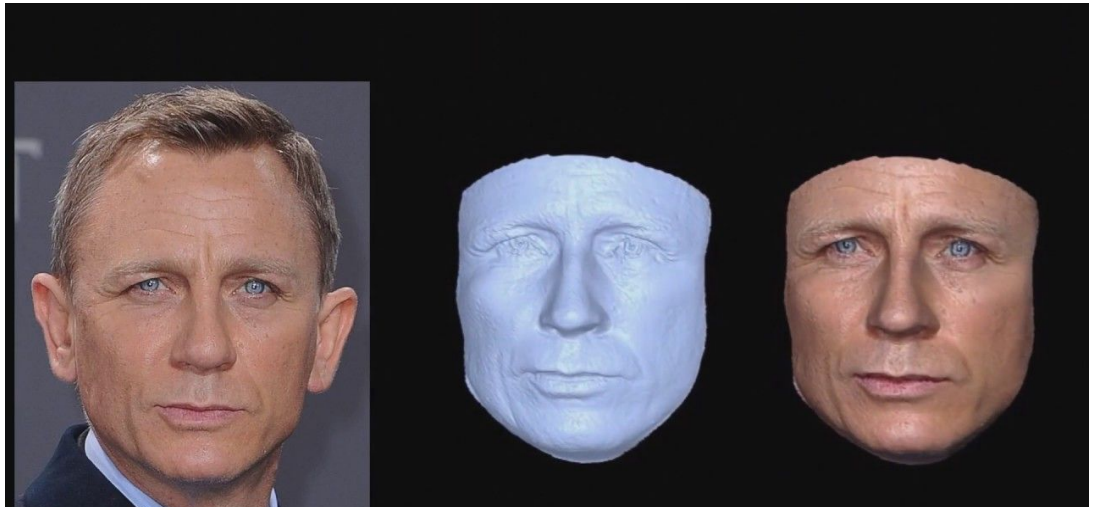
Some ML Perspectives

- Supervised Learning
  - Weekly, Semi, Self Supervised learning
- Zero & Few-shots learning / Closed vs Open Set
- Multi-tasking
- Knowledge Transfer / Domain adaptation / Meta Learning
- Knowledge Distillation
- GNN, Active Learning, Attention mechanisms

3D Vision

What have we lost when projecting:

3D world scene to 2D image?



Src: Site

3D vision

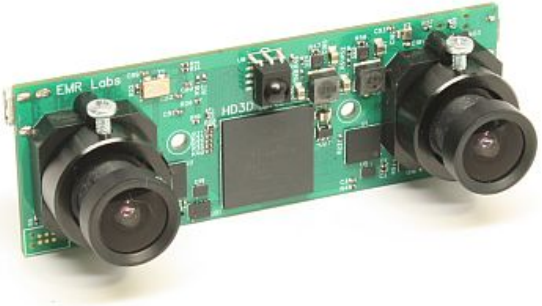
- Most of classical algorithms are explained in non-ML context
  - Involves camera model and camera matrices (intrinsic/extrinsic)
  - Involves single camera, two cameras, or more than two cameras
  - A lot of linear algebra and optimizations!
- In deep learning context, some problems are
  - solved by ML training (e.g. optical flow) or
  - networks involve some 3D information (e.g. hand pose estimation, gaze estimation, ...)

3D point clouds

- Representation for 3D objects

Src: Site

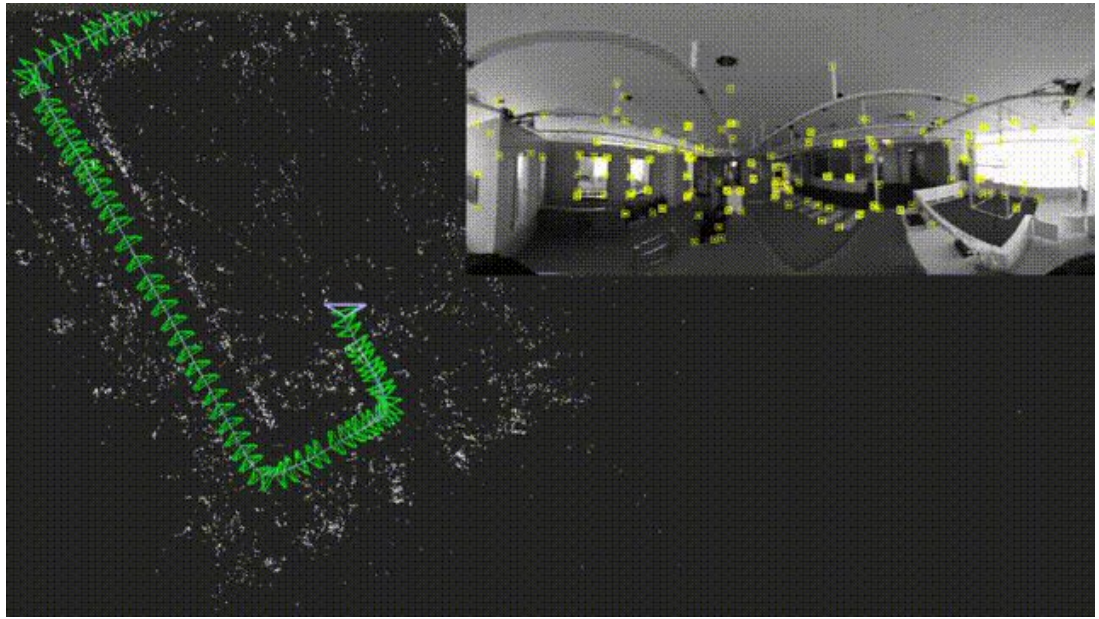
Stereo Camera



Src: Site Site

Stereo Vision

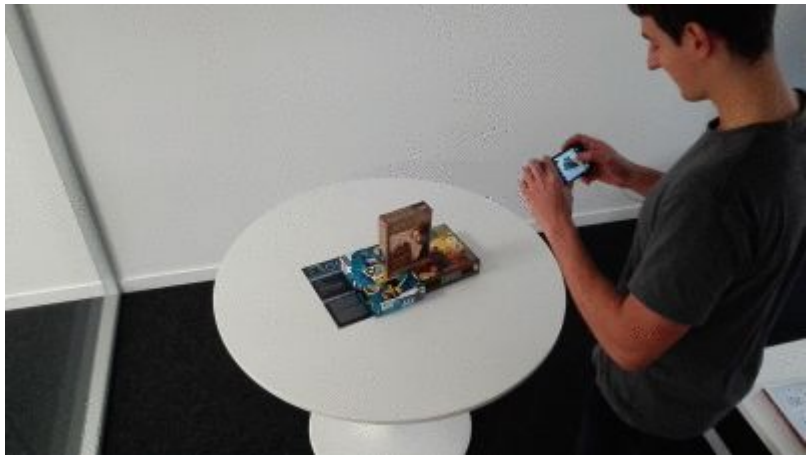
- important in fields such as **robotics**, to extract information about the **relative position of 3D objects**
- Right side: **VSLAM**



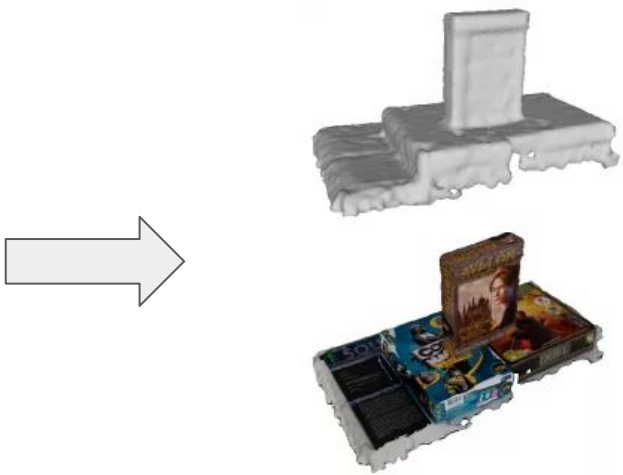
Src: Site

3D reconstruction

- Given set of images of object (e.g. building), construct its 3D object

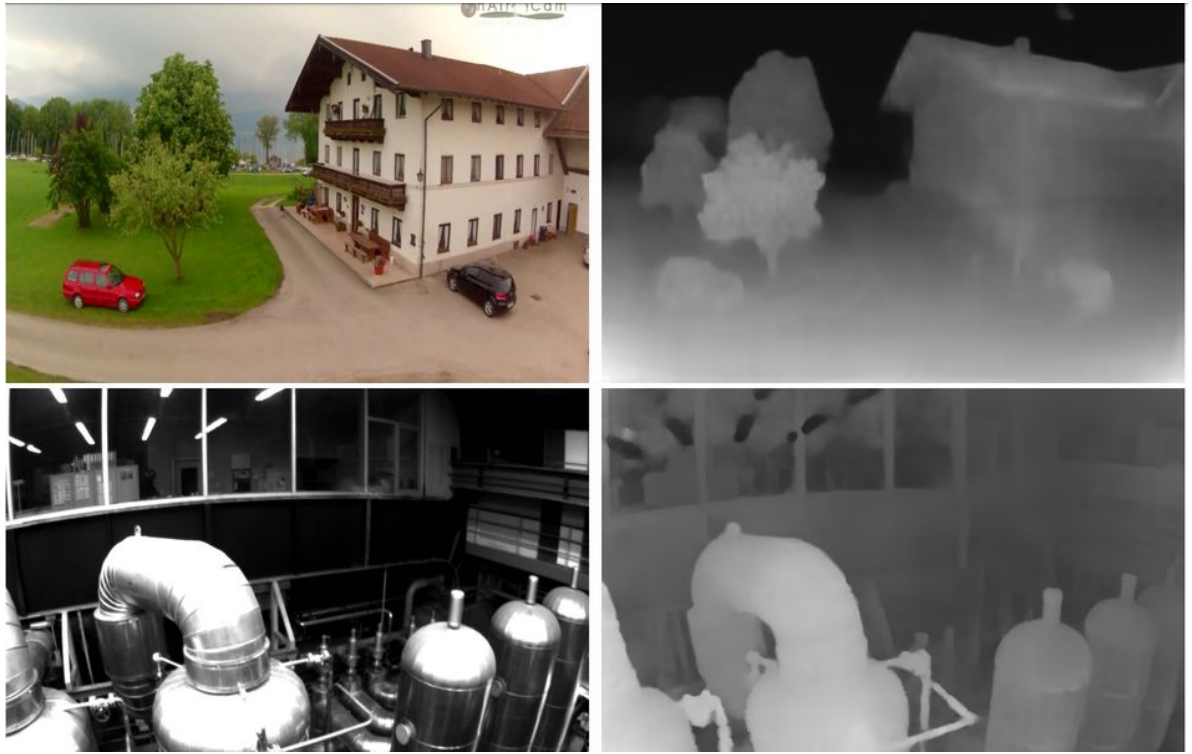


Src: Site



Depth Estimation

- Depth ~= Distance



Src: Site

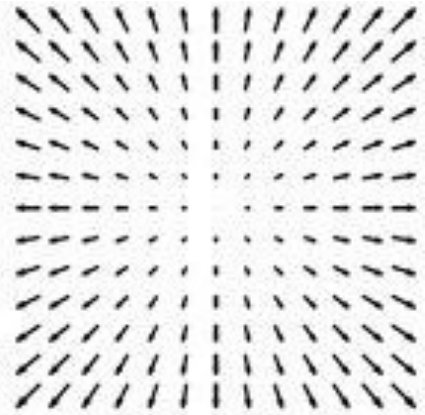
Panorama stitching

- Take **several photos** of a wide view and merge then nicely to one **big** photo

Src: Site

Optical Flow

- Given **2 consecutive frames**, find **displacement vector** showing the **movement** of points from first frame to second
- Can be casted as learning problem (E.g. FlowNet)



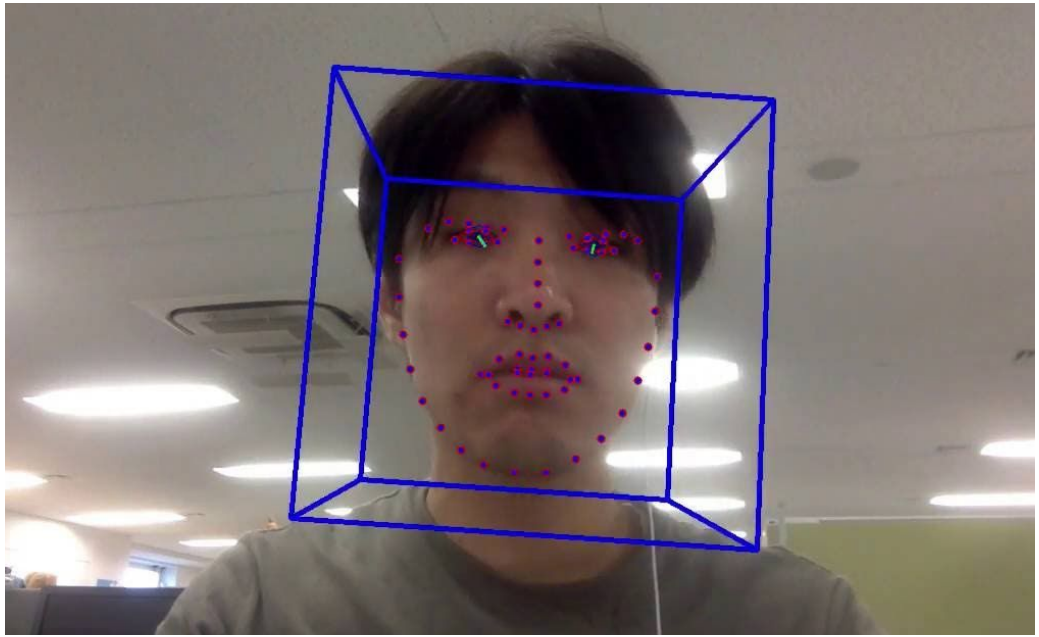
Src: Site

3D Body Pose Estimation

Img src

3D Head Pose Estimation

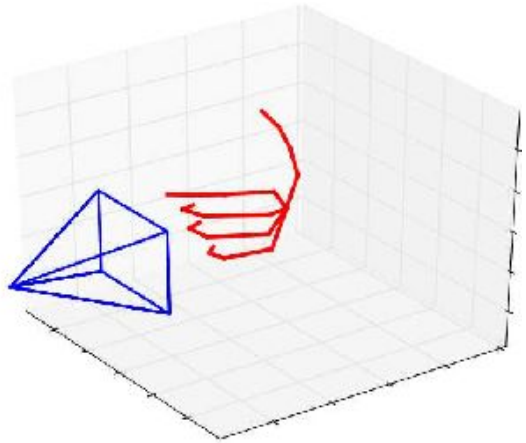
- Finding the translation and rotation of the head



Src: Site

3D Hand Pose Estimation

- For each joint, its 3D position



Src: Paper

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."