EE 5178 : Modern Computer Vision

Kaushik Mitra
Dept. of Electrical Engineering
IIT Madras

The goal of computer vision

- To extract “meaning” from pixels

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

What kind of information can be extracted from an image?

- Semantic information

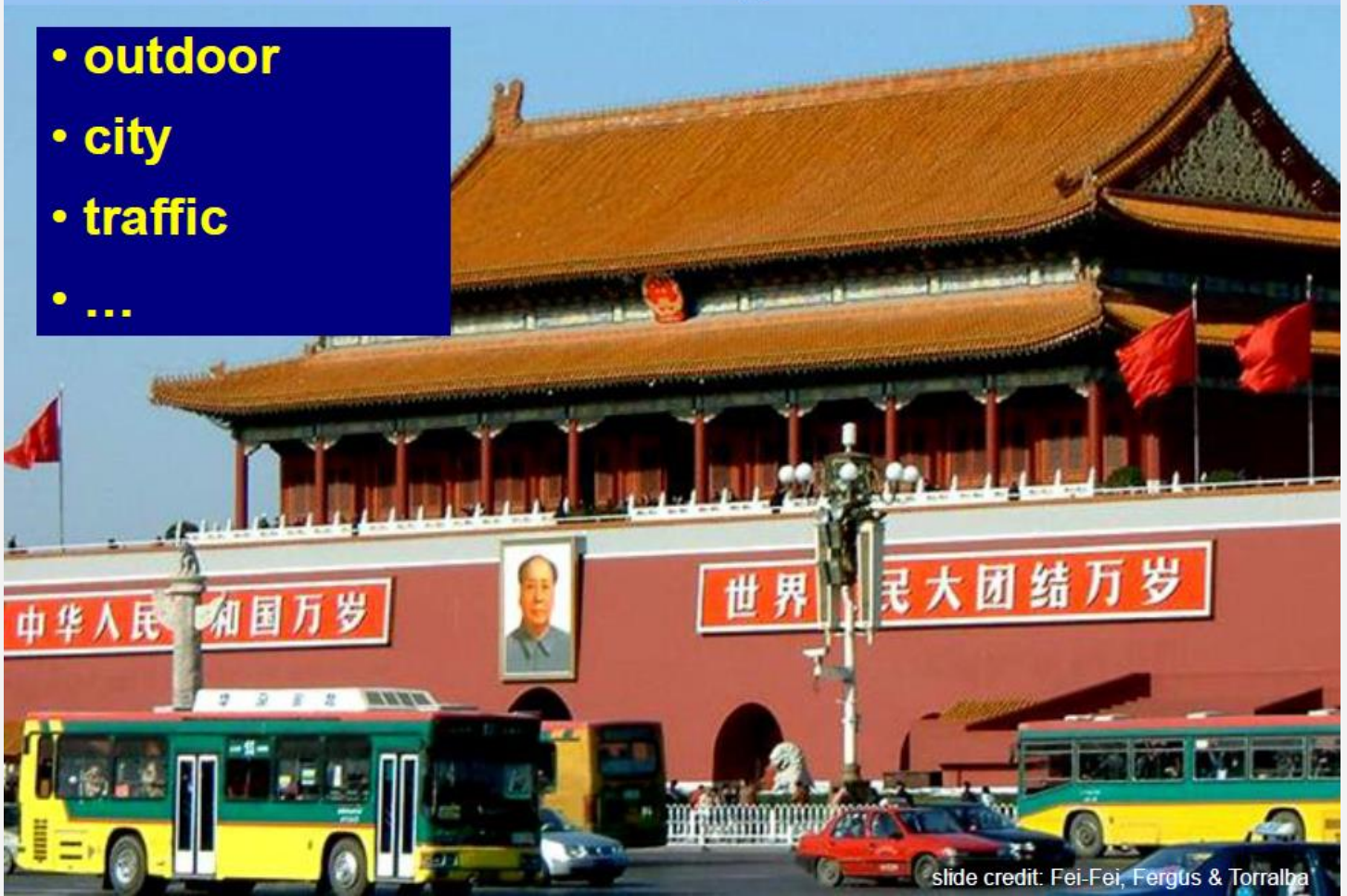
- Metric information

Vision as a source of semantic information

Object categorization

Scene and context categorization

- outdoor
- city
- traffic
- …



slide credit: Fei-Fei, Fergus & Torralba

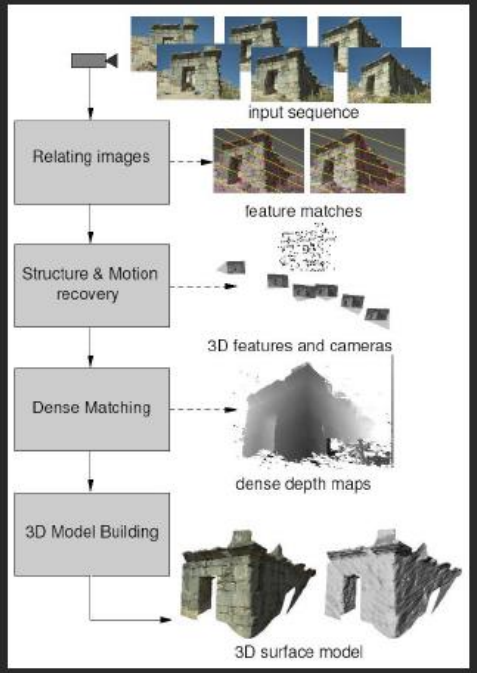
Vision as measurement device

Real-time stereo

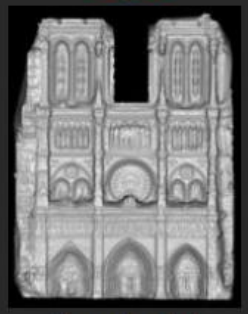
NASA Mars Rover

Pollefeys et al.

Structure from motion



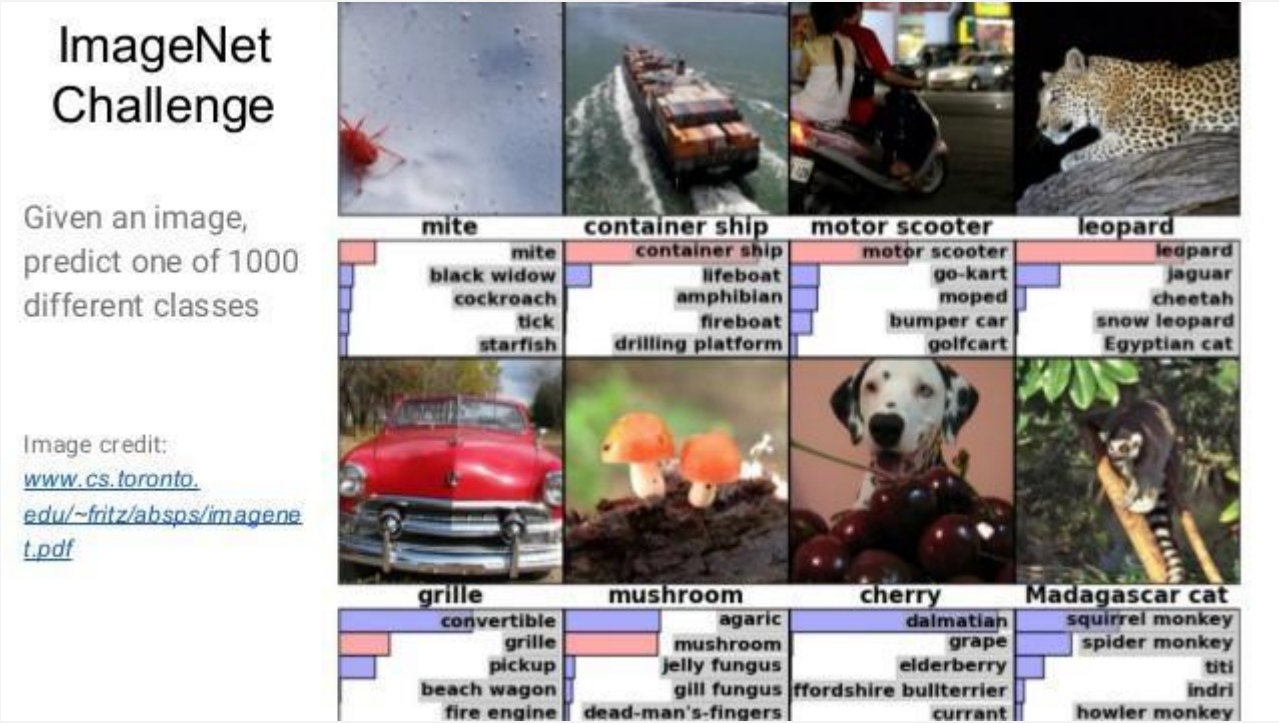
Reconstruction from Internet photo collections



Goesele et al.

Semantic Computer Vision tasks

Object Recognition



Object/face Detection (Recognition + Localization)

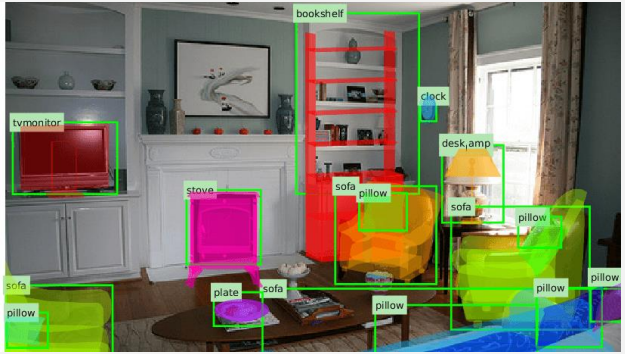
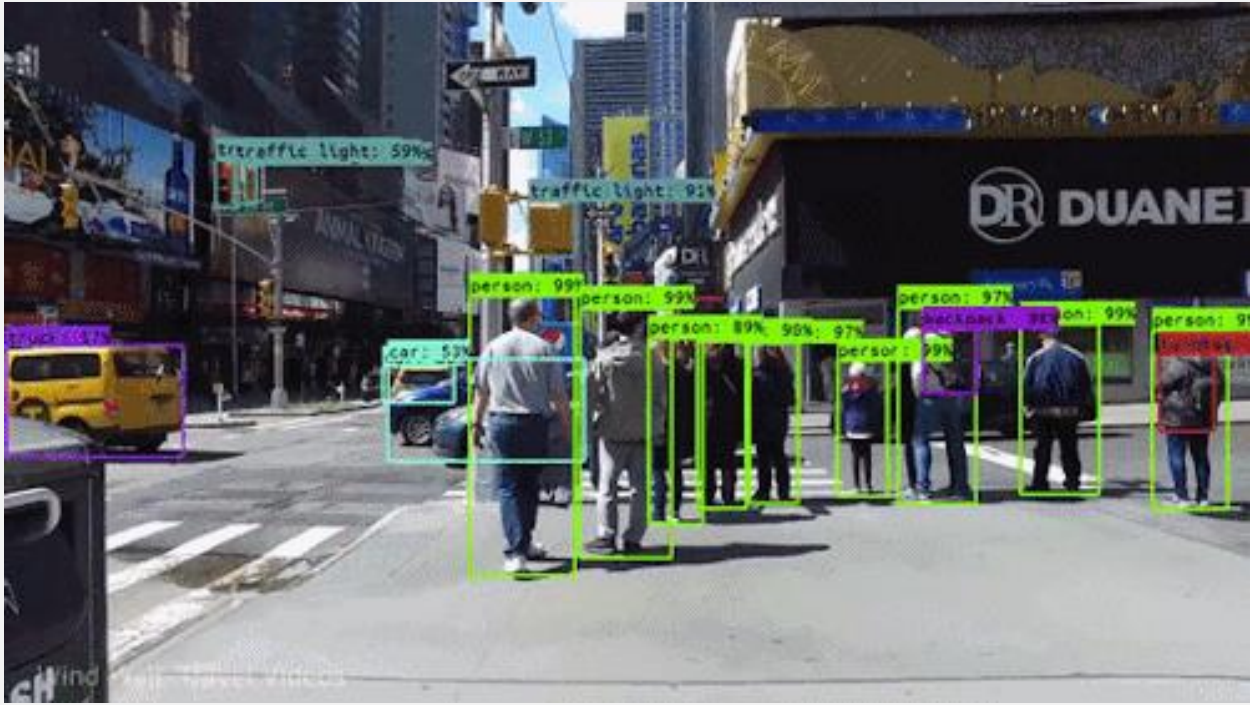


Image Source



You Only Look Once: Unified, Real-Time Object Detection

Face detection + Localization + Classification

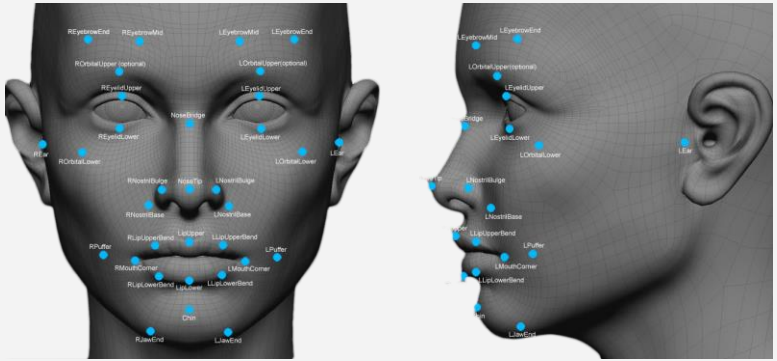


Image Source

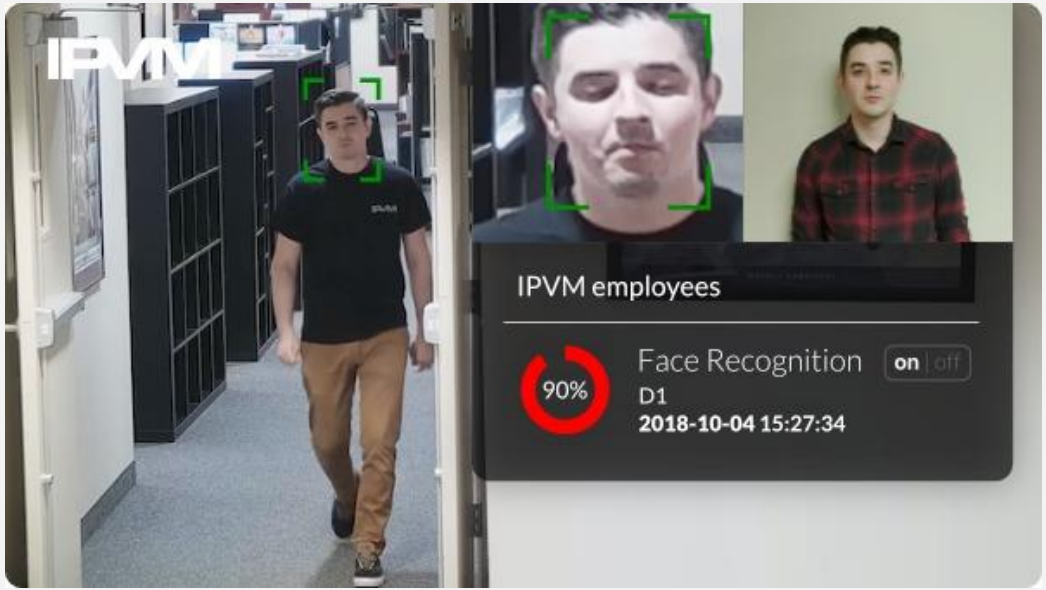


Image Source

Number Plate Reader

License plate readers

http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

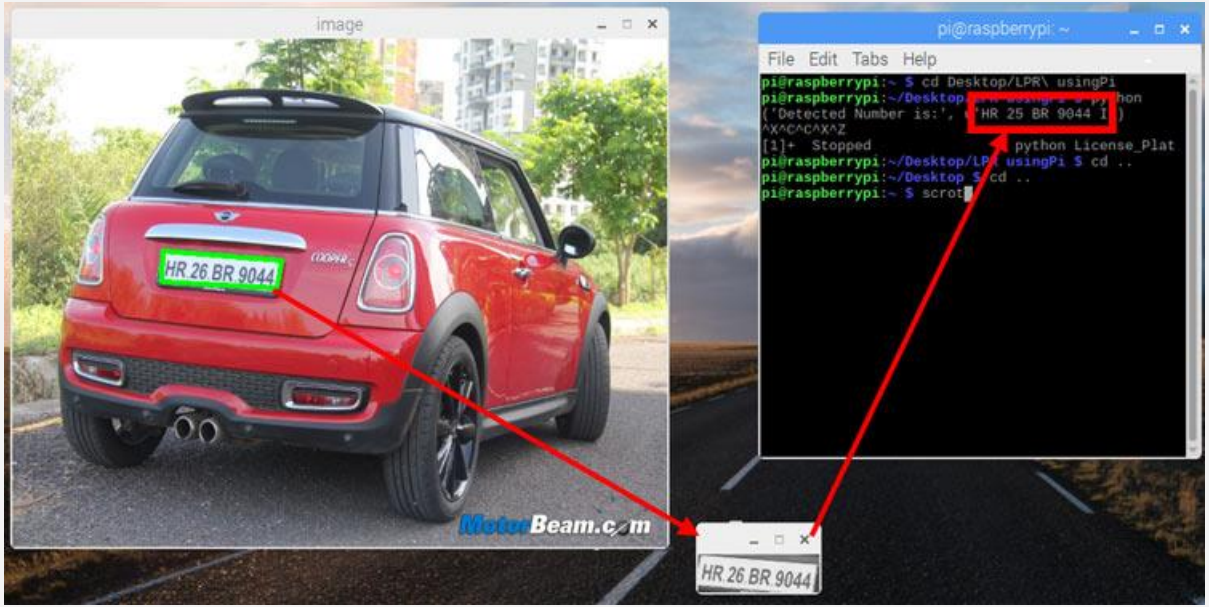
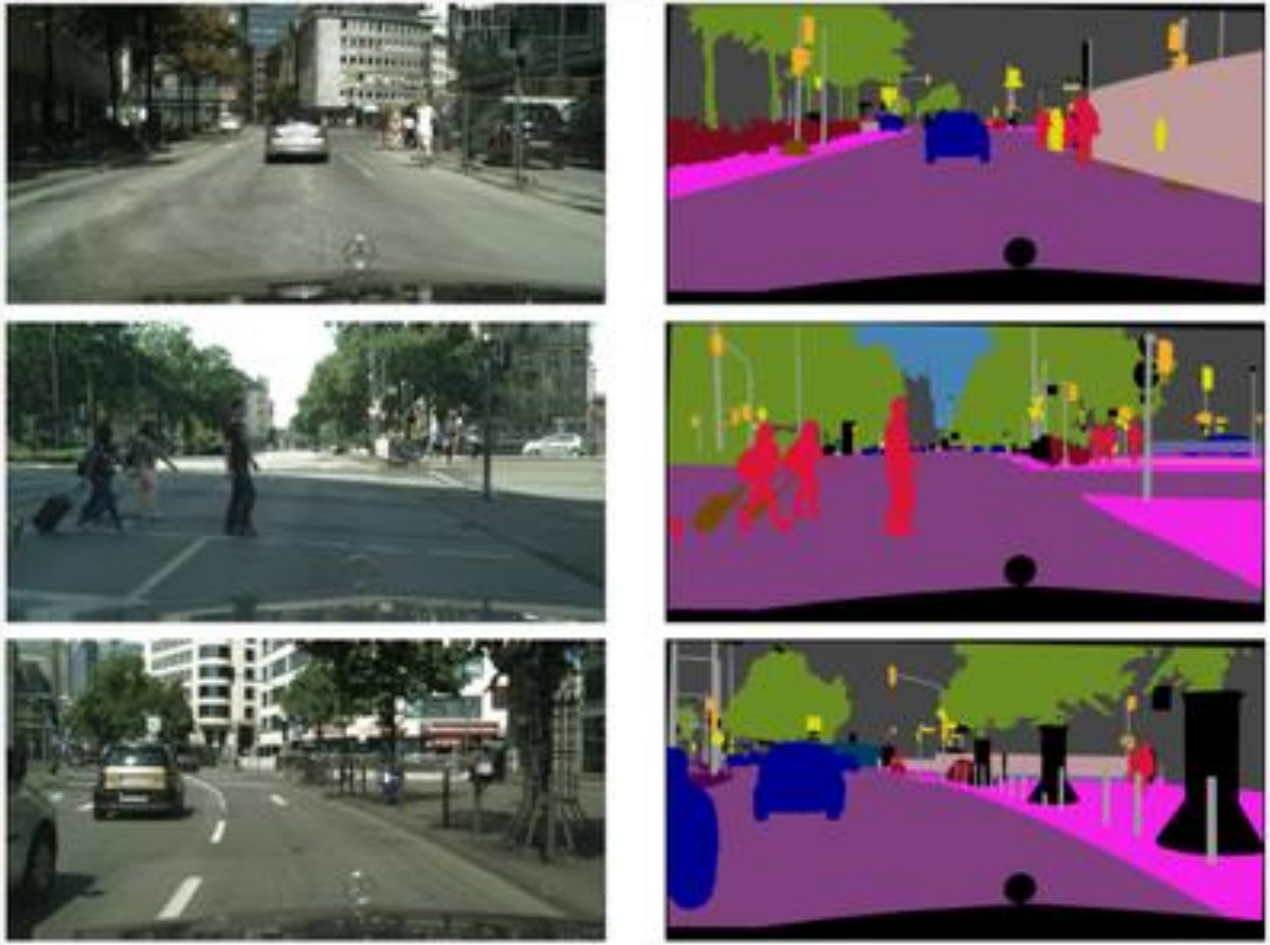
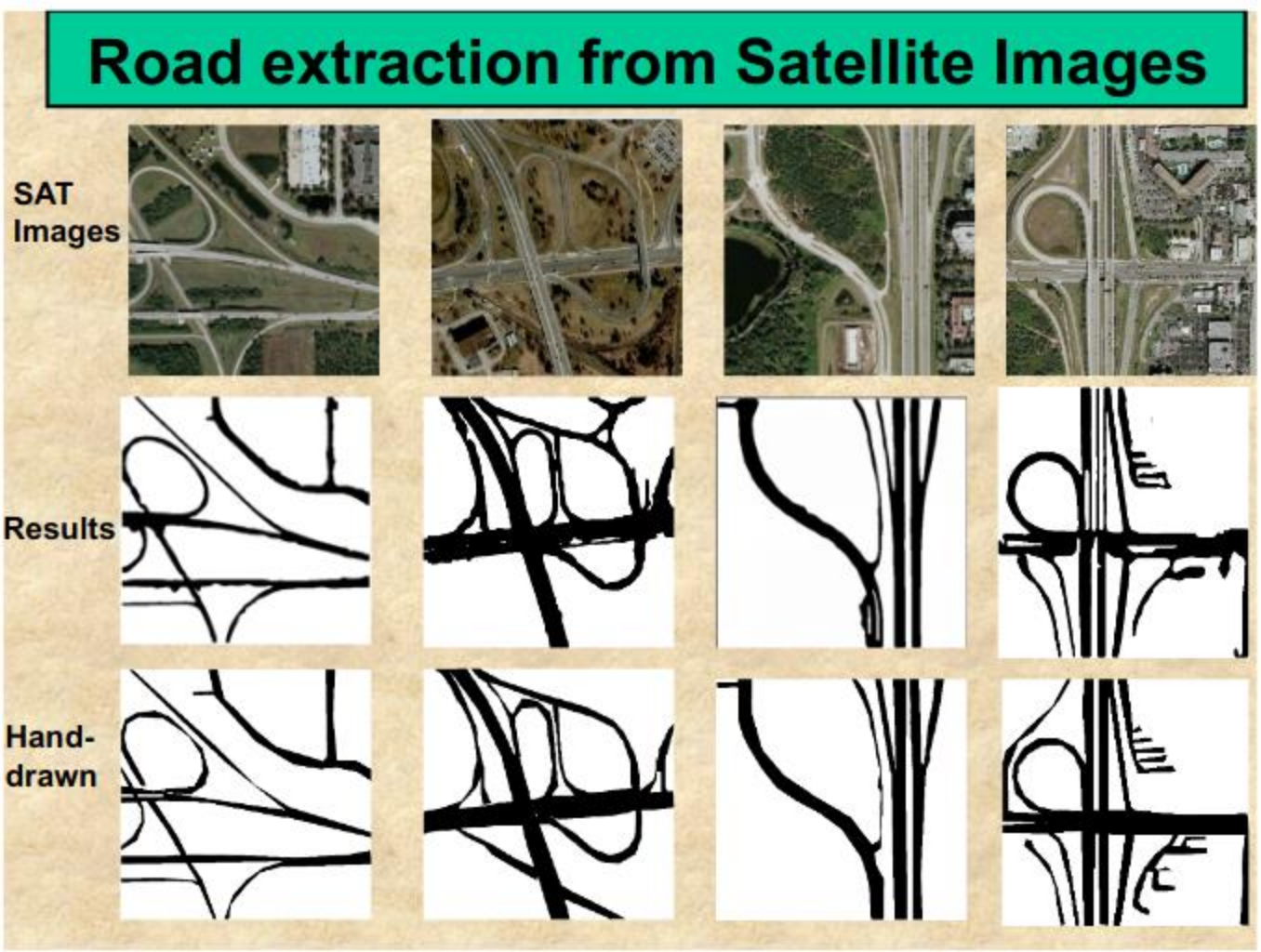


Image Segmentation



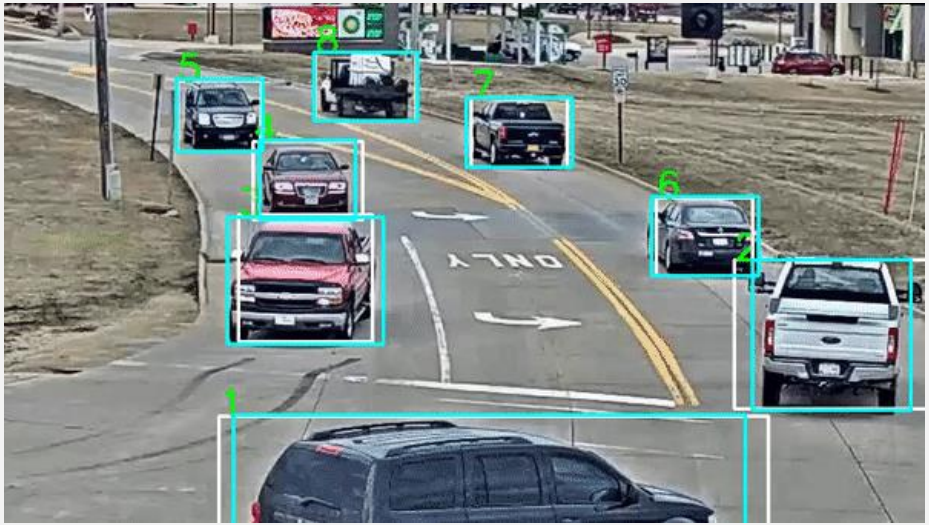
Semantic Image Segmentation



Courtesy: Prof. Sukhendu Das

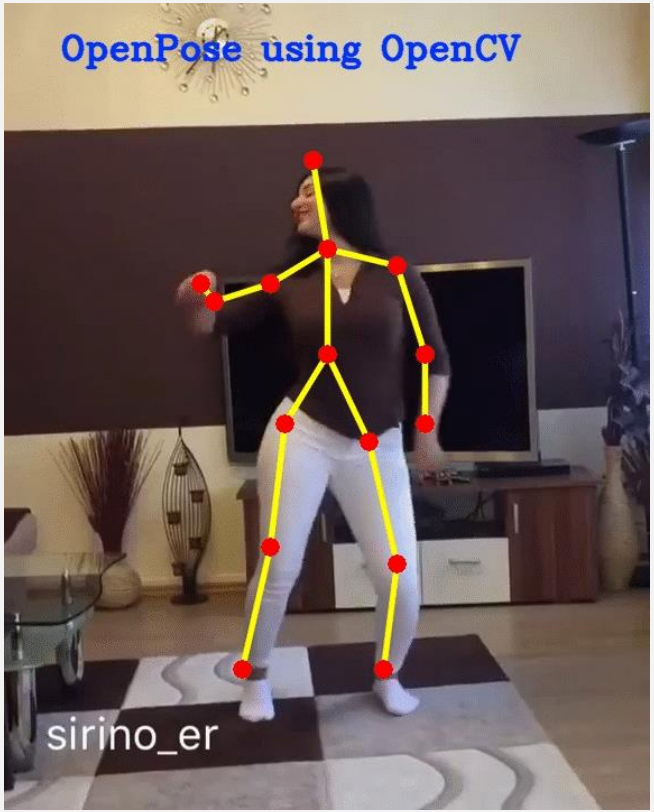
Object Tracking

Image Source



DeepSORT: Deep Learning to track custom objects in a video

Pose estimation



Activity Recognition

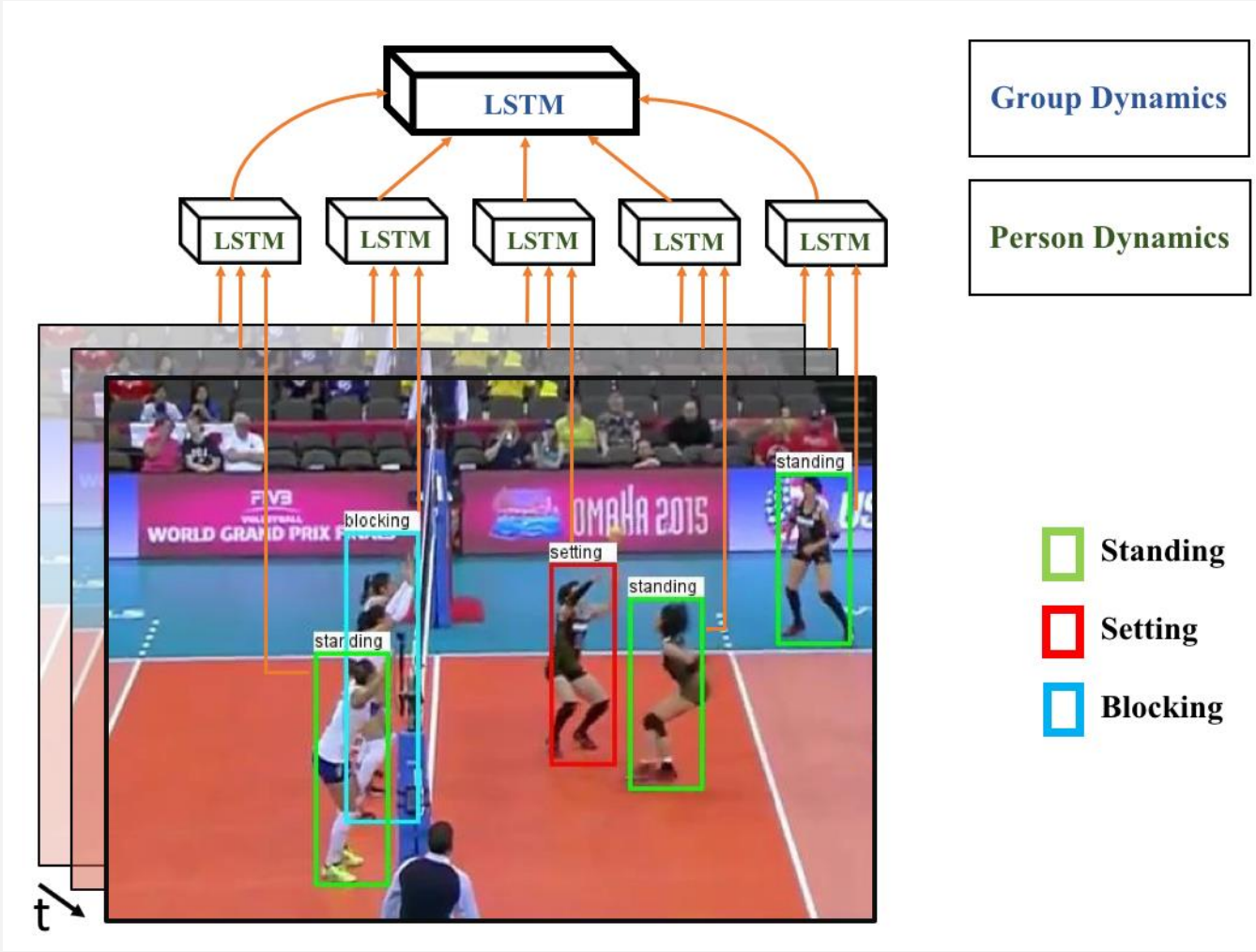
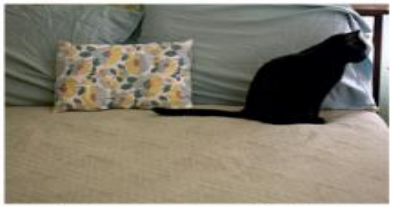


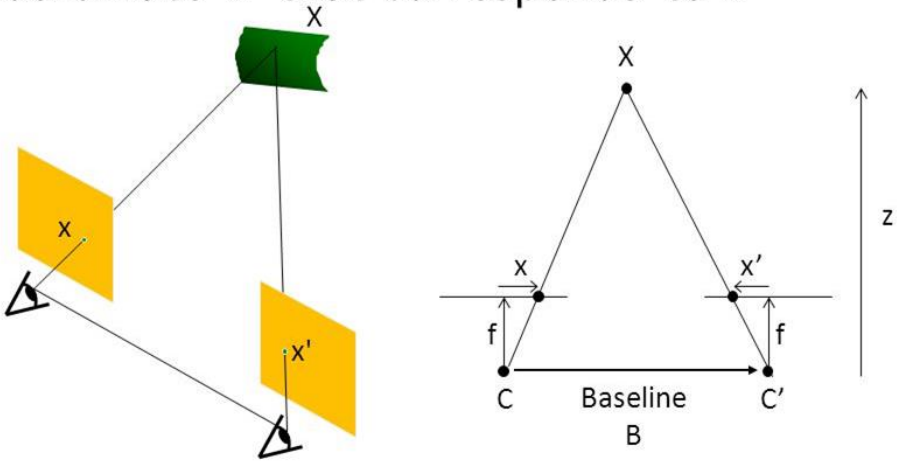
Image Captioning

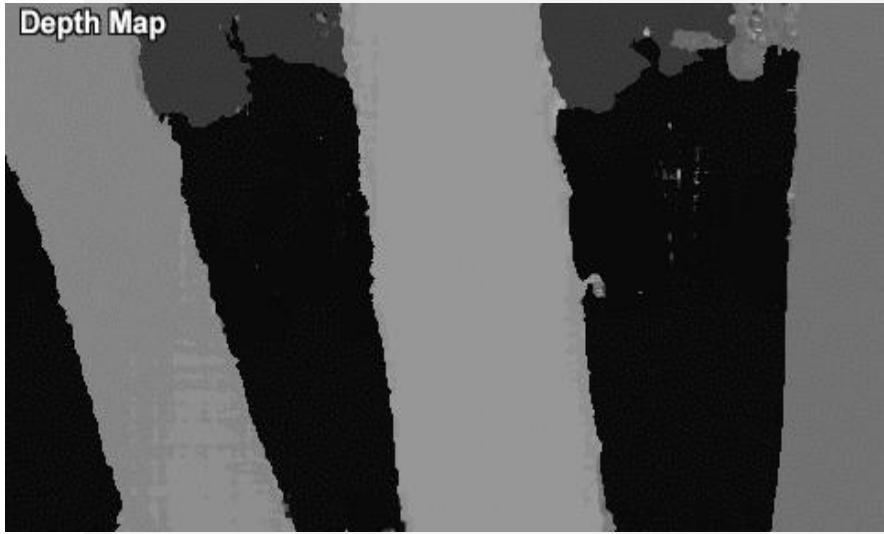
Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
 A person riding a motorcycle on a dirt road.	 Two dogs play in the grass.	 A skateboarder does a trick on a ramp.	 A dog is jumping to catch a frisbee.
 A group of young people playing a game of frisbee.	 Two hockey players are fighting over the puck.	 A little girl in a pink hat is blowing bubbles.	 A refrigerator filled with lots of food and drinks.
 A herd of elephants walking across a dry grass field.	 A close up of a cat laying on a couch.	 A red motorcycle parked on the side of the road.	 A yellow school bus parked in a parking lot.

Geometric Computer Vision : Vision for measurement

Depth from stereo

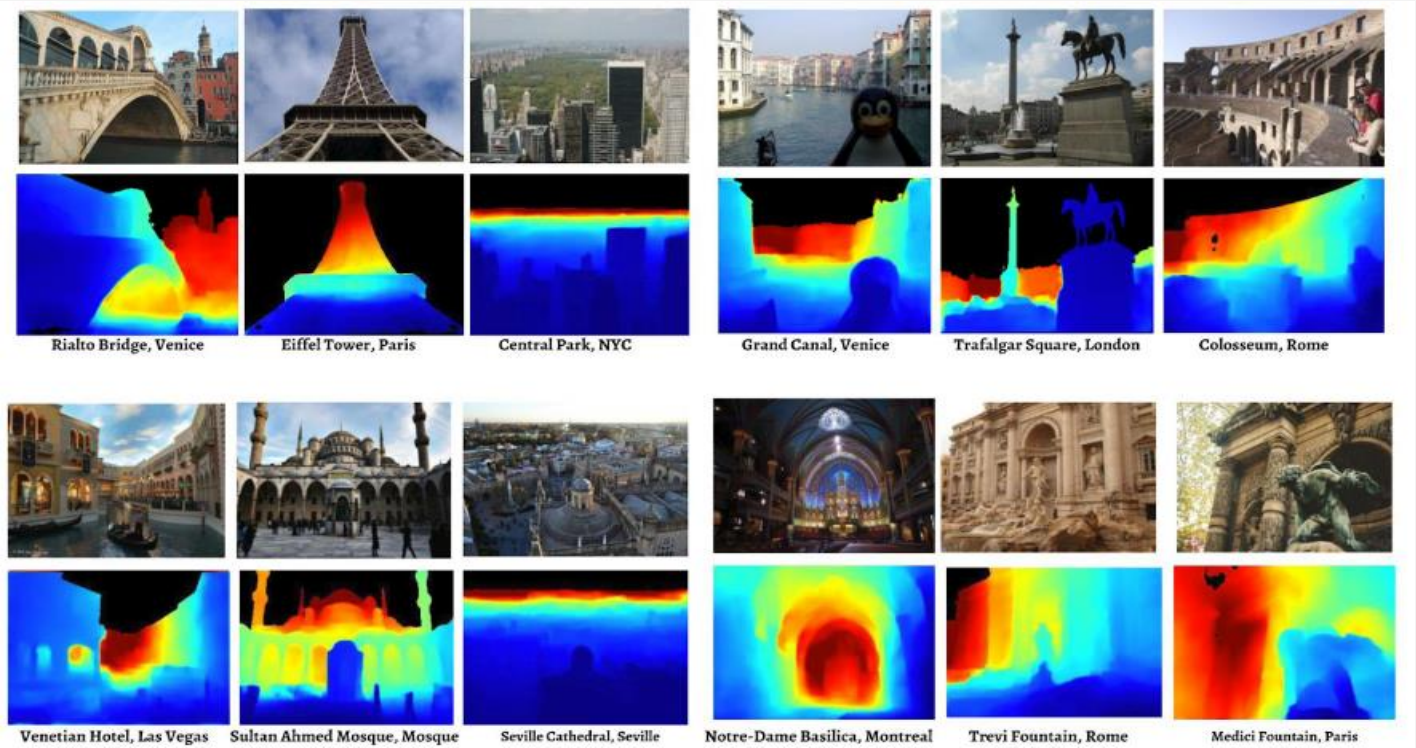
- Goal: recover depth by finding image coordinate x′ that corresponds to x





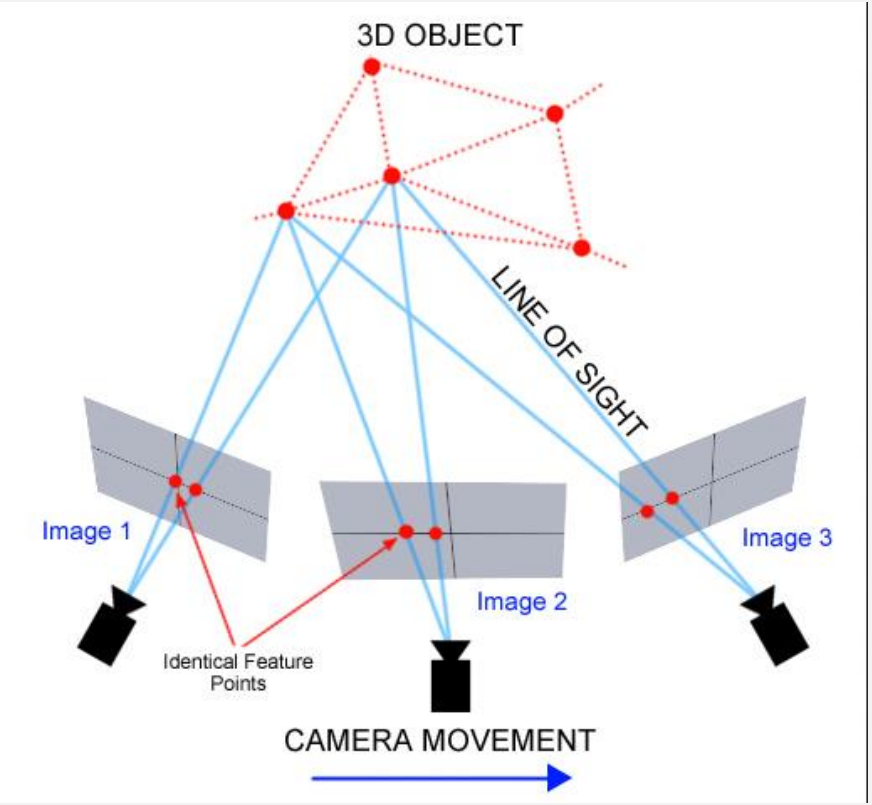
Source: Simon Jacobs

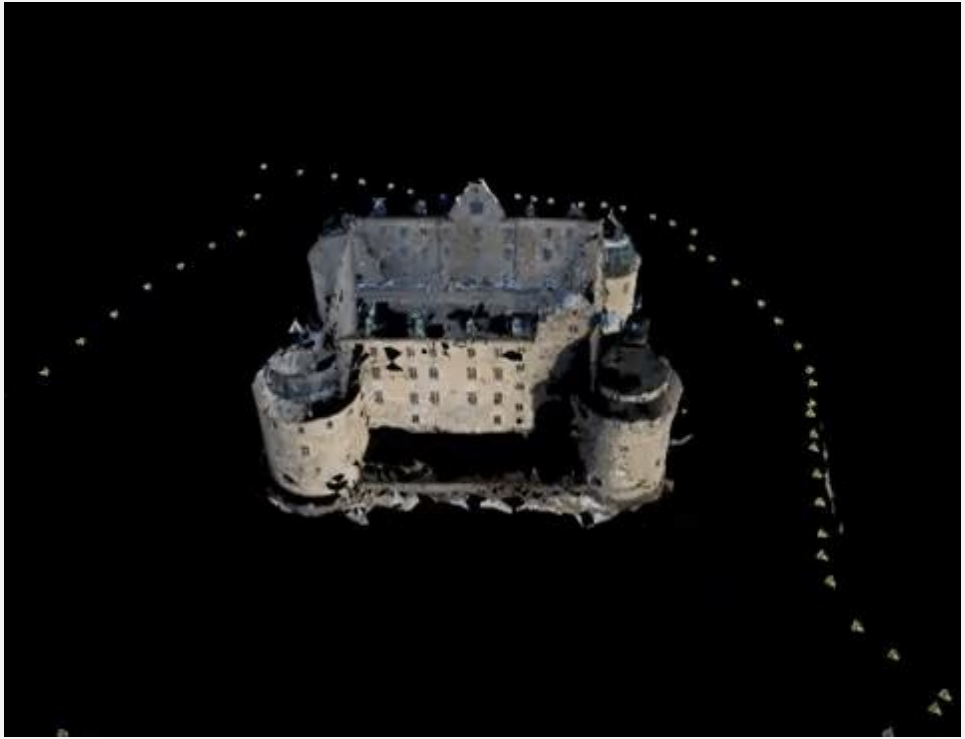
Depth from a single image



Courtesy Noah Snavely

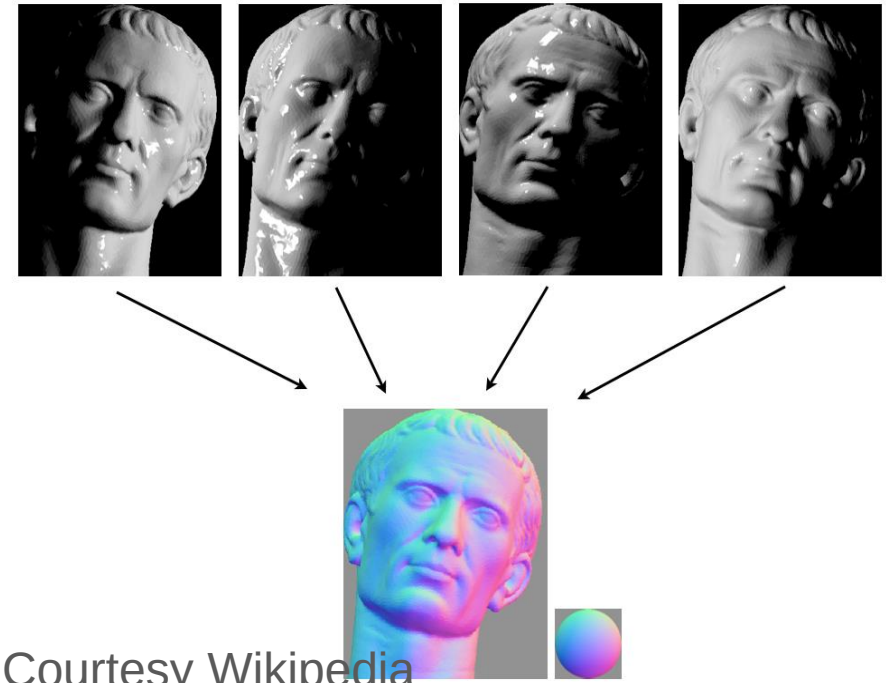
Structure From Motion



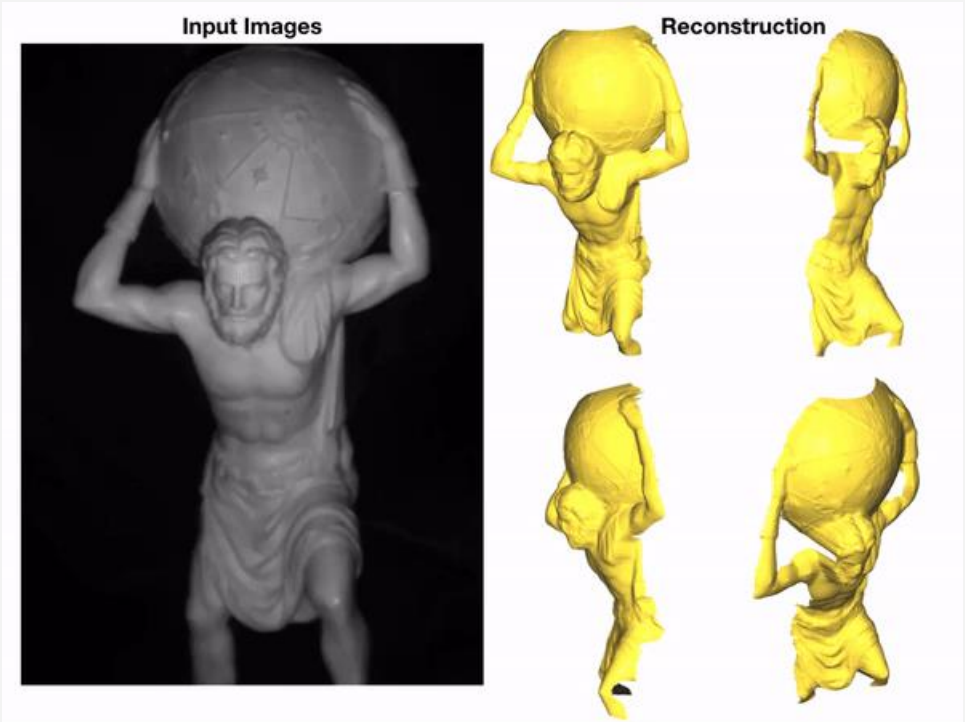


Source: Dr. Calle Olsson

Photometric stereo



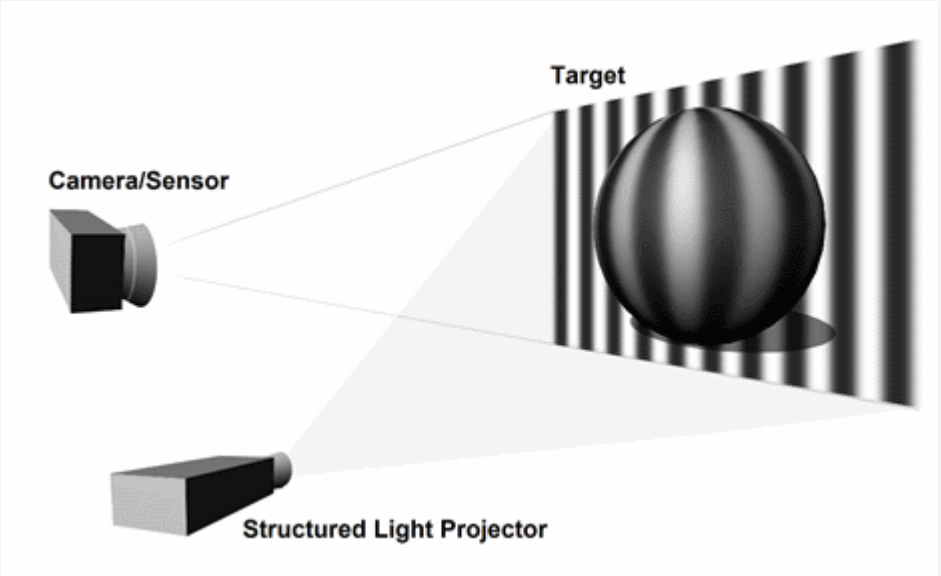
Courtesy Wikipedia

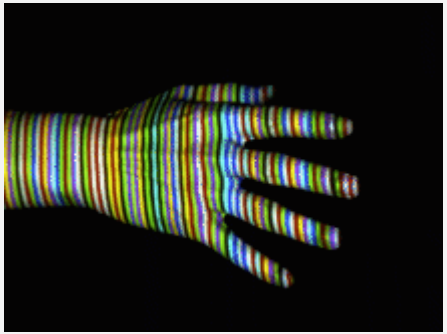


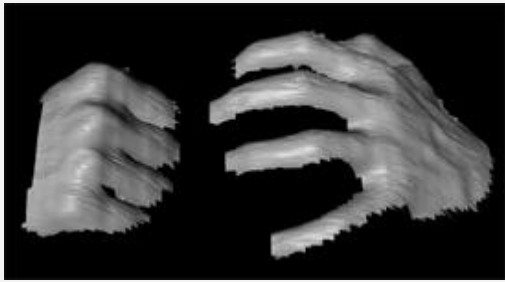
" Near-Light Photometric Stereo using Circularly Placed Point Light Sources "

Structured light

Like Stereo, but replace one camera with Projector







“Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming”

You just saw many examples of current systems.

But what challenges do they face ?

Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Challenges: illumination

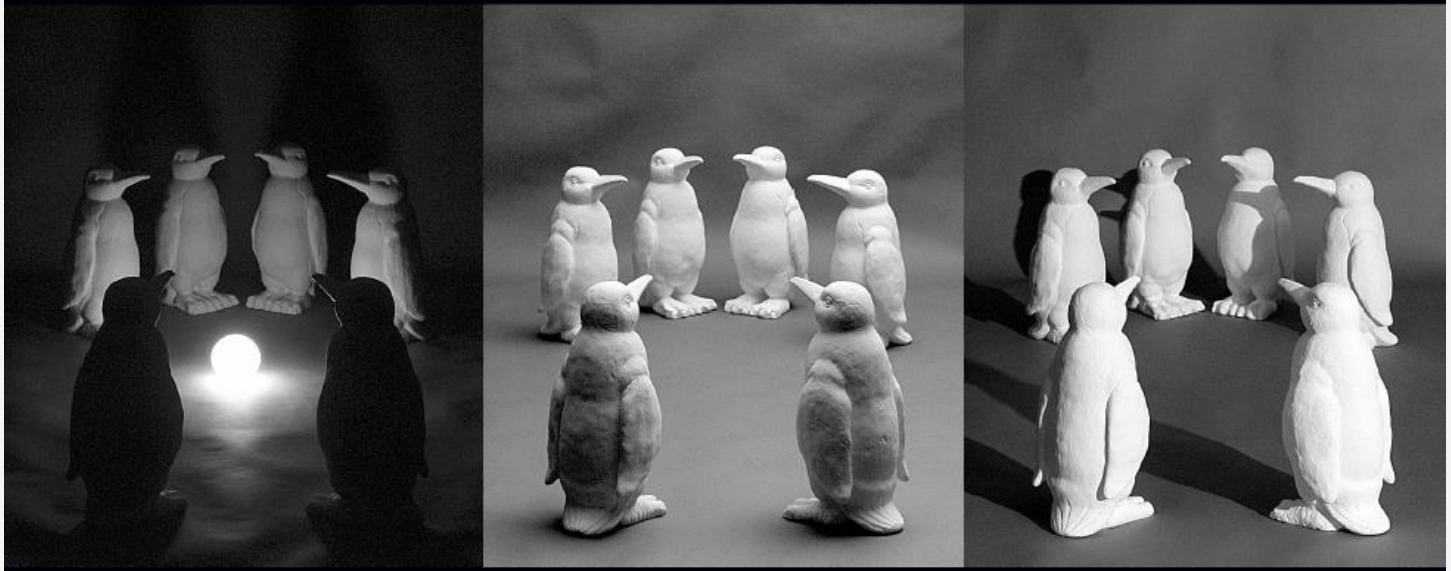
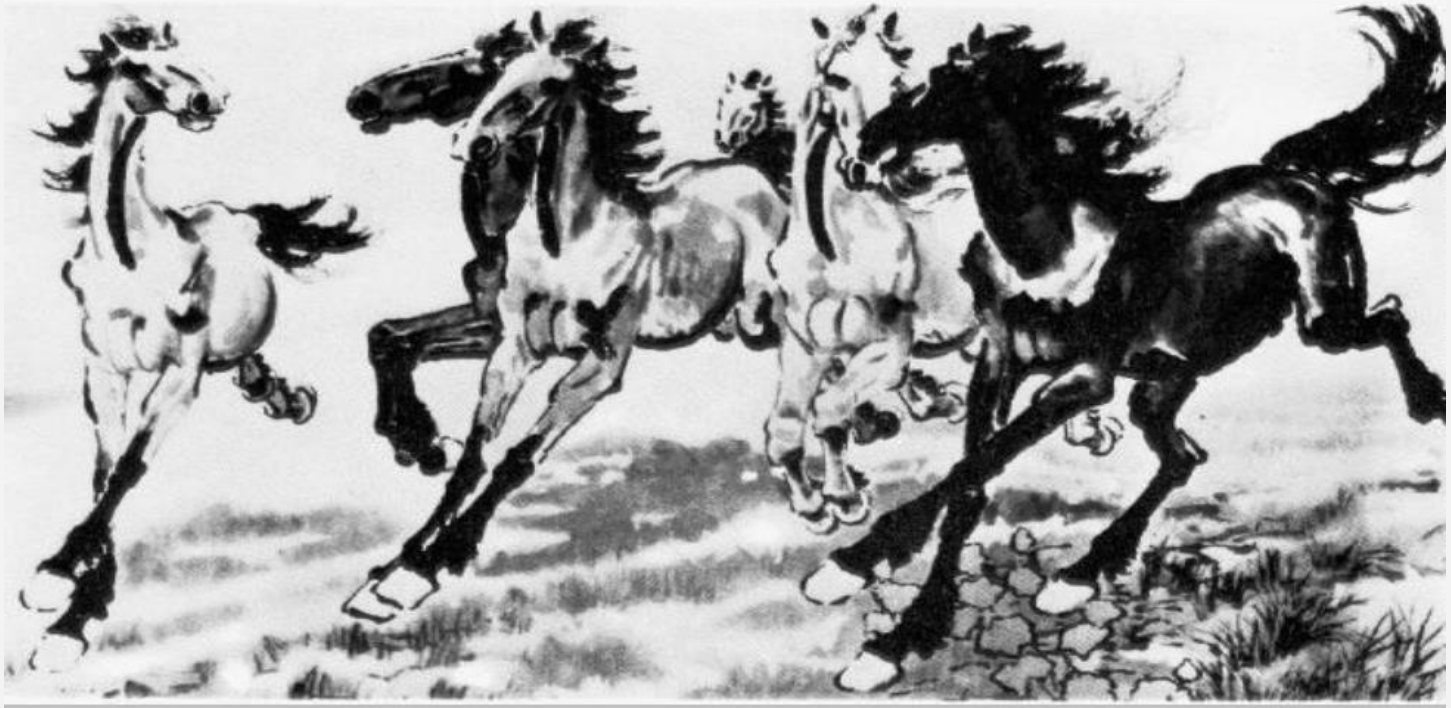


image credit: J. Koenderink

Challenges: scale

slide credit: Fei-Fei, Fergus & Torralba

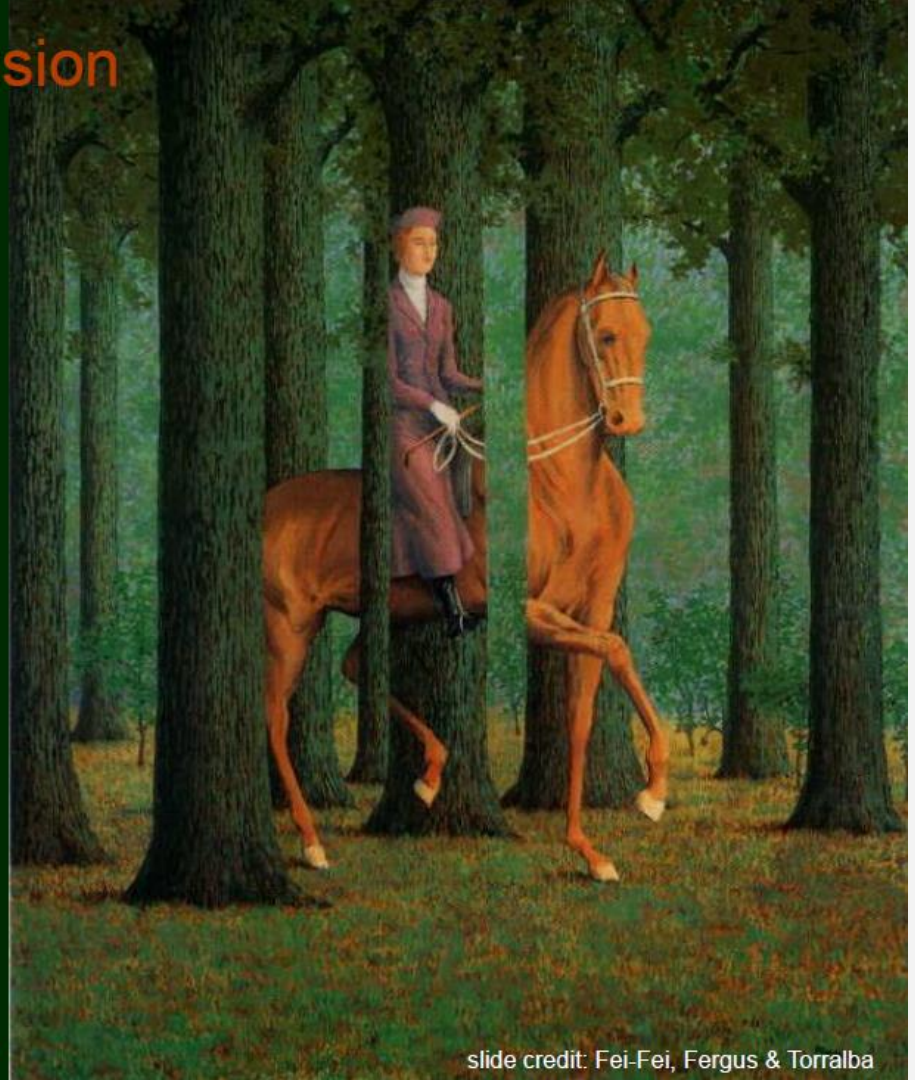
Challenges: deformation



Xu, Beihong 1943

slide credit: Fei-Fei, Fergus & Torralba

Challenges: occlusion



Magritte, 1957

slide credit: Fei-Fei, Fergus & Torralba

Challenges: background clutter

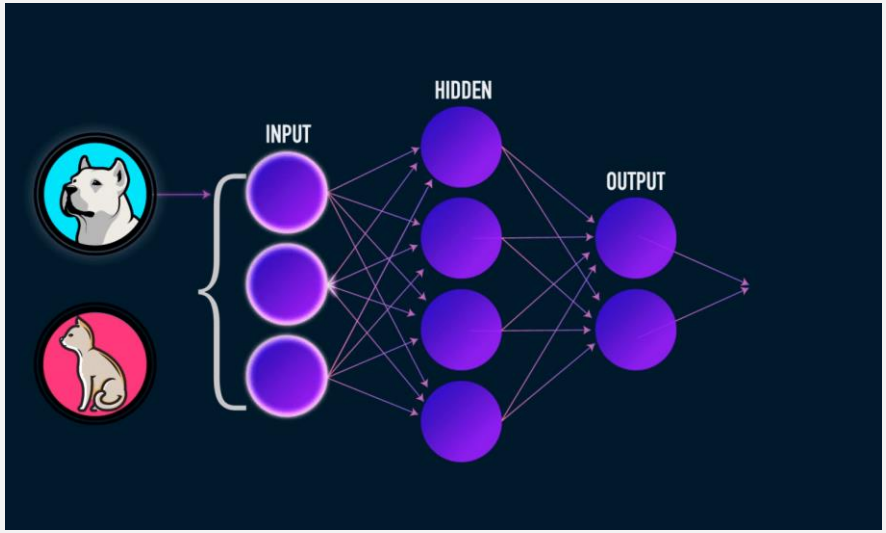
Challenges: Motion

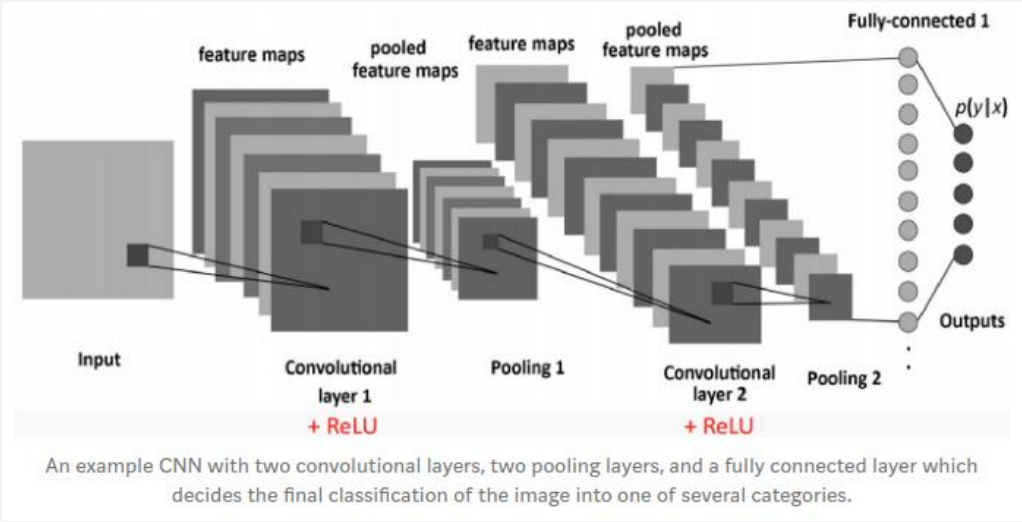
Challenges: object intra-class variation

slide credit: Fei-Fei, Fergus & Torralba

Course Content

Quick review of Deep Learning

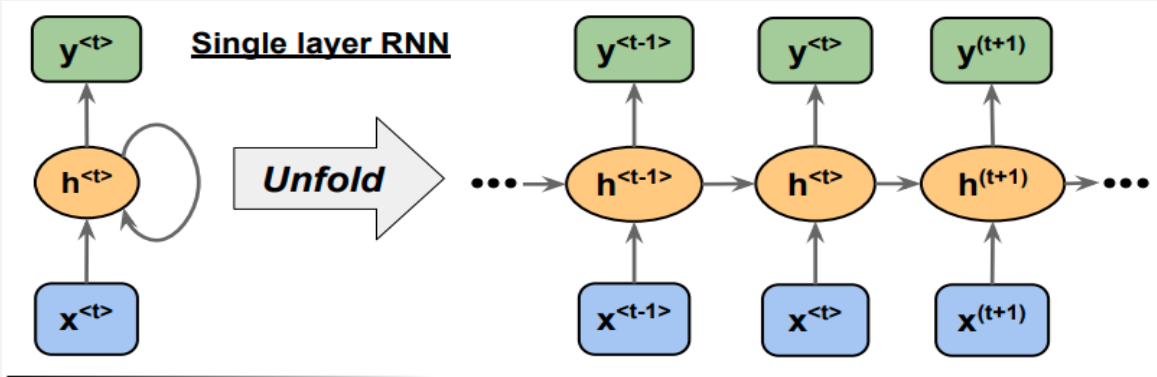




An example CNN with two convolutional layers, two pooling layers, and a fully connected layer which decides the final classification of the image into one of several categories.

Basic Neural Network (MLP)

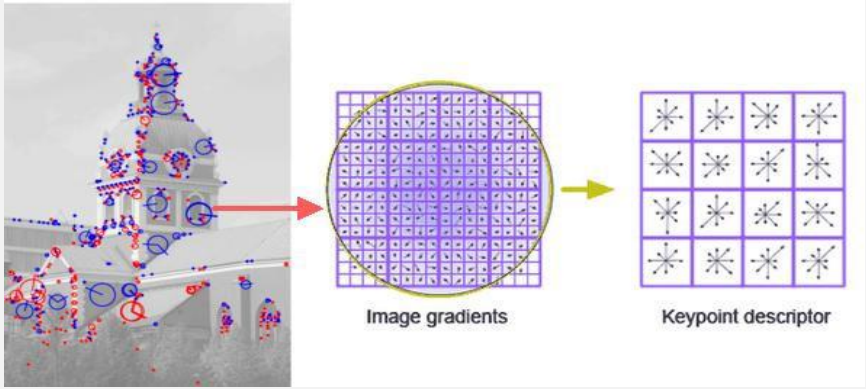
Convolutional Neural Network (CNN)



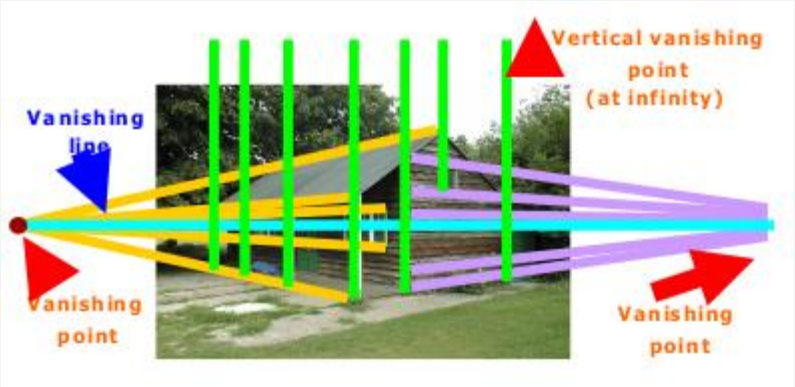
Recurrent Neural Network (RNN)

Low level vision

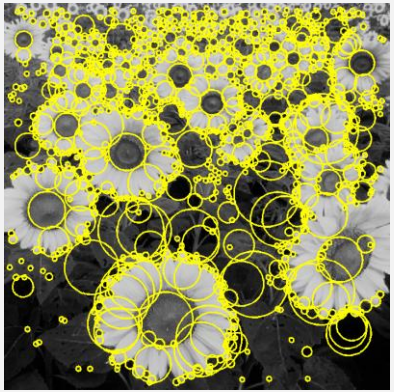
Edge Extraction



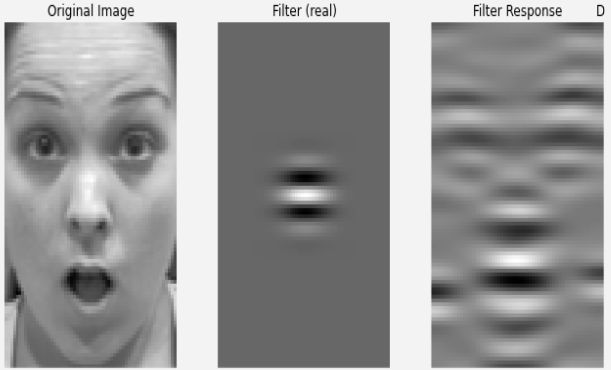
Feature Extraction



Line Detection

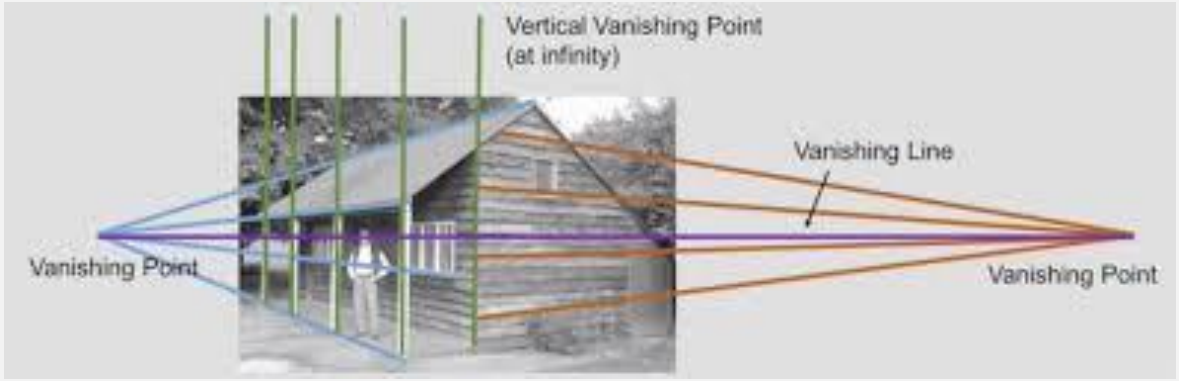


Blob detection

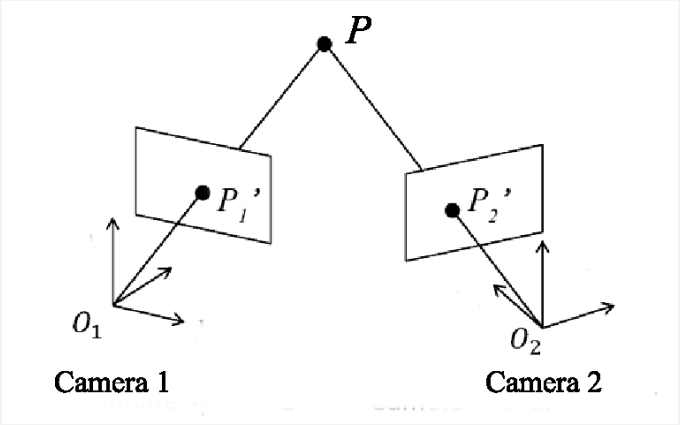


Filtering

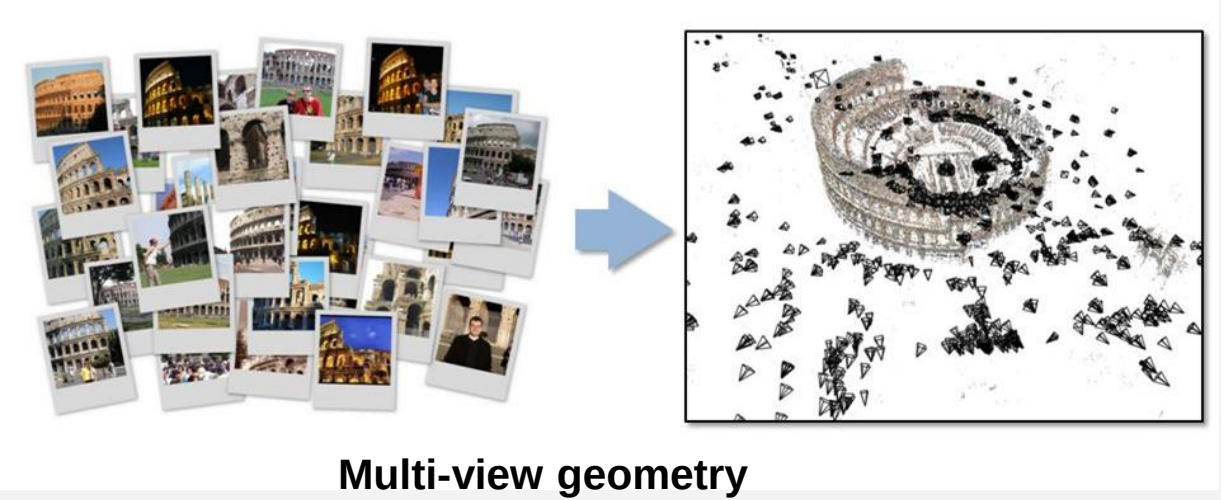
Geometry



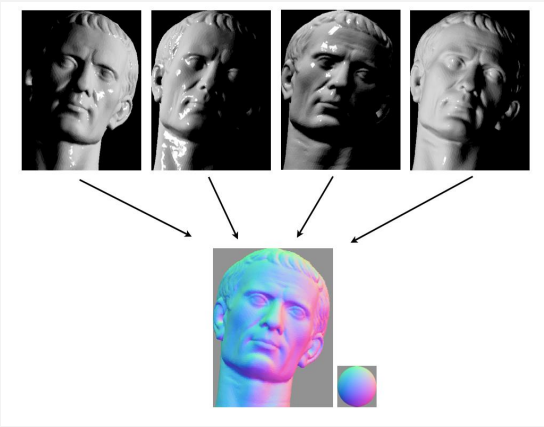
Single view geometry



Two-view geometry

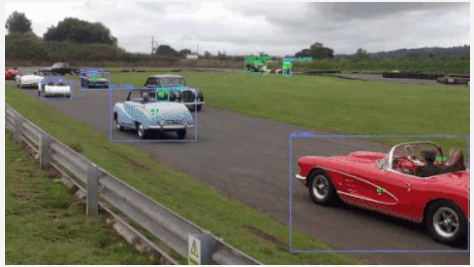


Multi-view geometry

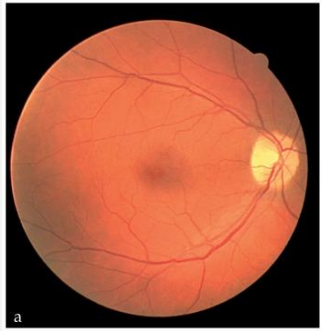


Photometric stereo

Mid-level Vision



Tracking



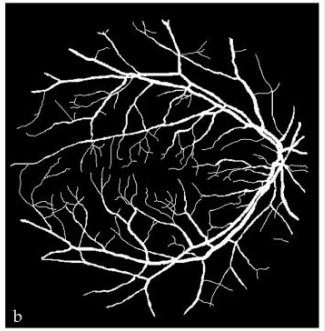
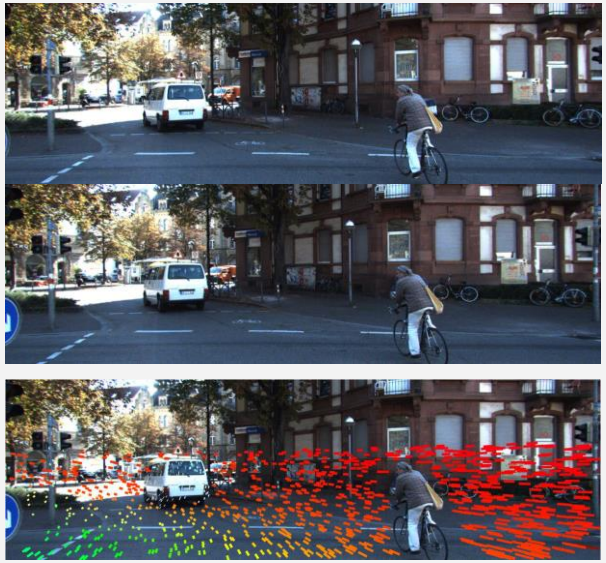


Image segmentation



Optical Flow

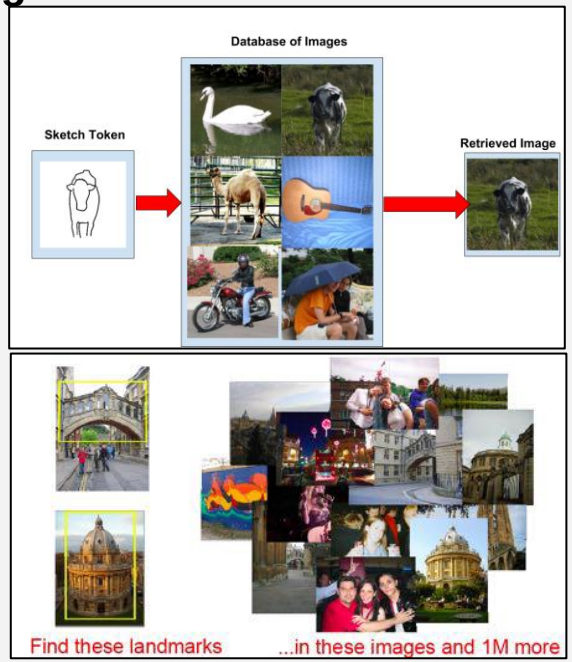
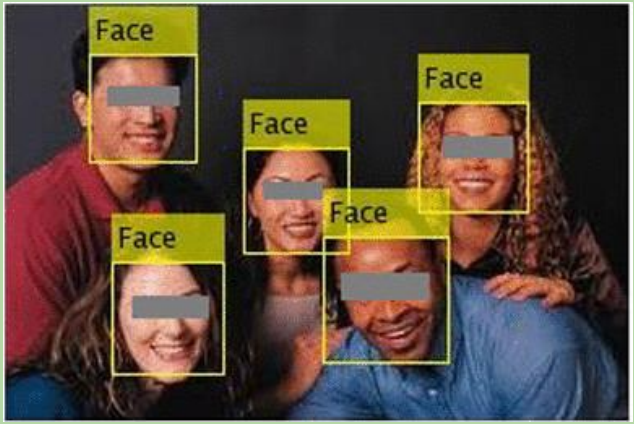


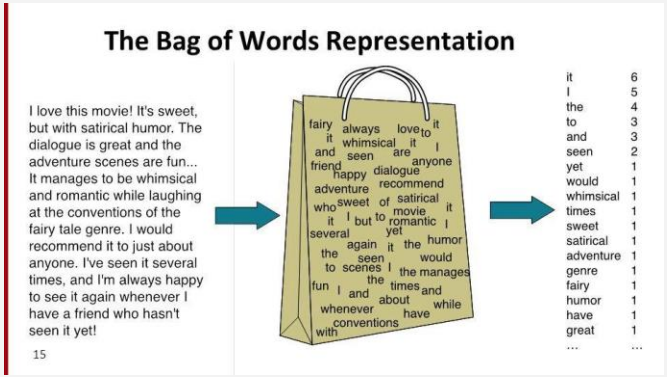
Image Retrieval

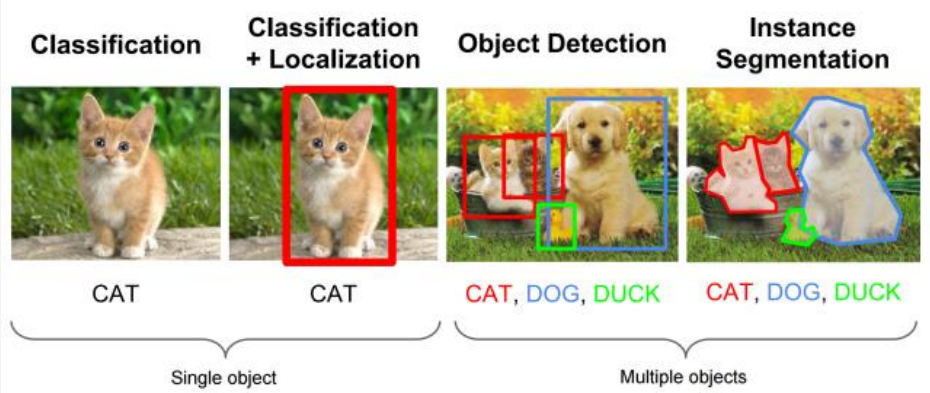
High level Vision



Viola Jones Face Detector

Image Captioning





Tentative grading policy

- Regular (MCQ based) mini-quizzes (20%) – most likely 4 miniquiz
- Programming assignments (25%)
  - Around 3-4 assignments
- Kaggle Competition or Term Project (25%)
- End semester exam (30%)

Reading Material

1. Szeliski, Computer Vision: Algorithms and Applications, Springer, 2010

2. Hartley and Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004

3. Forsyth and Ponce, Computer Vision: A Modern Approach, Prentice Hall, 2002

Course Contents

● **Quick review of Deep Learning:** Multilayer perceptron (MLP), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN)

● **Low level vision:** Edge, line and corner detections; Image filtering; Features

● **Geometry:** Single-view geometry; Stereo geometry; Multi-view geometry; Photometric stereo

● **Mid-level vision:** Optical flow, Image segmentation; Tracking; CB Image Retrieval

● **High-level vision:** Viola-Jones detector; Bag of words model; Deformable parts model; Object recognition and detection; Image captioning