CS231A

Computer Vision:

From 3D Perception to 3D Reconstruction and Beyond

Class Time and Location

M-W; 1:30—2:50PM
Gates B1

CS231A

Instructors

Silvio Savarese

- ssilvio@stanford.edu
- Office: Gates Building, room: tbd
- Office hour: tbd and by appointment

Jeannette Bohg

- bohg@stanford.edu
- Office: Gates Building, room: 244
- Office hour: Tuesday 9am (not this week)

Teaching Assistants

Aditya Dutt

Ishikaa Lunawat

Adam Sun

Lecture 1

Introduction

- An introduction to computer vision
- Course overview

AI is a propelling force of today's technology

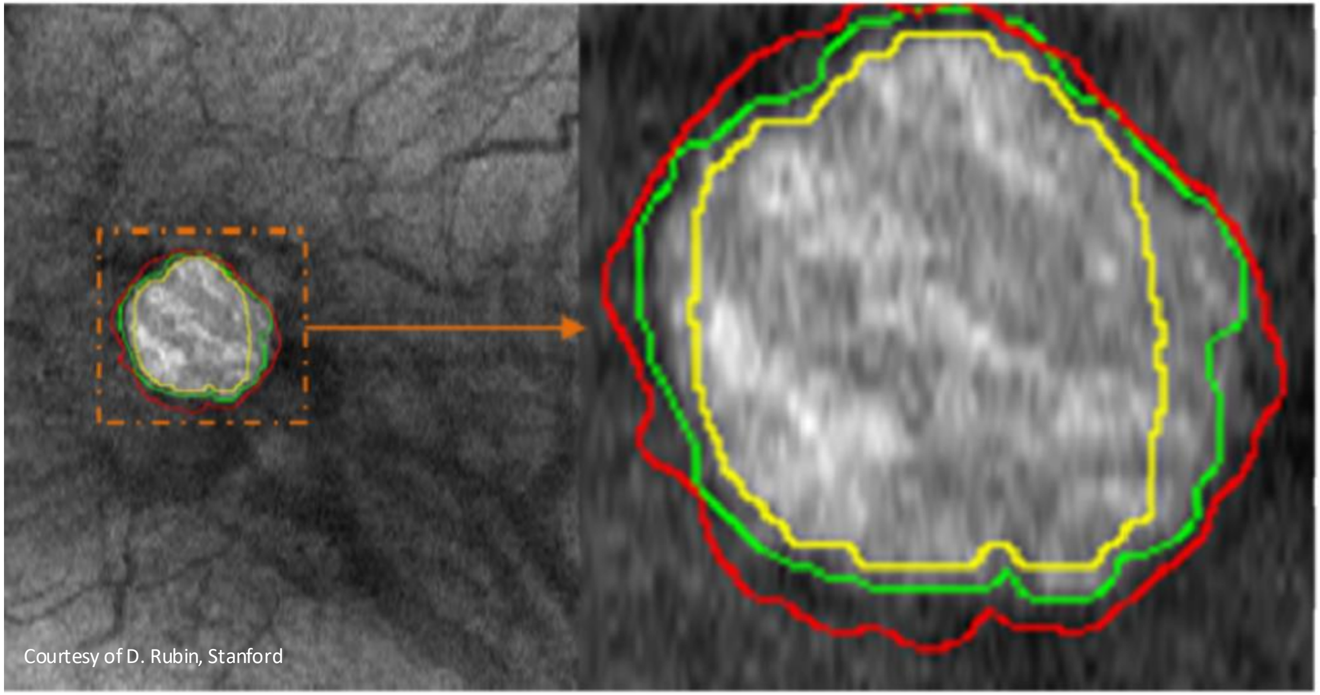
OpenAI's DALL-E

Zipline

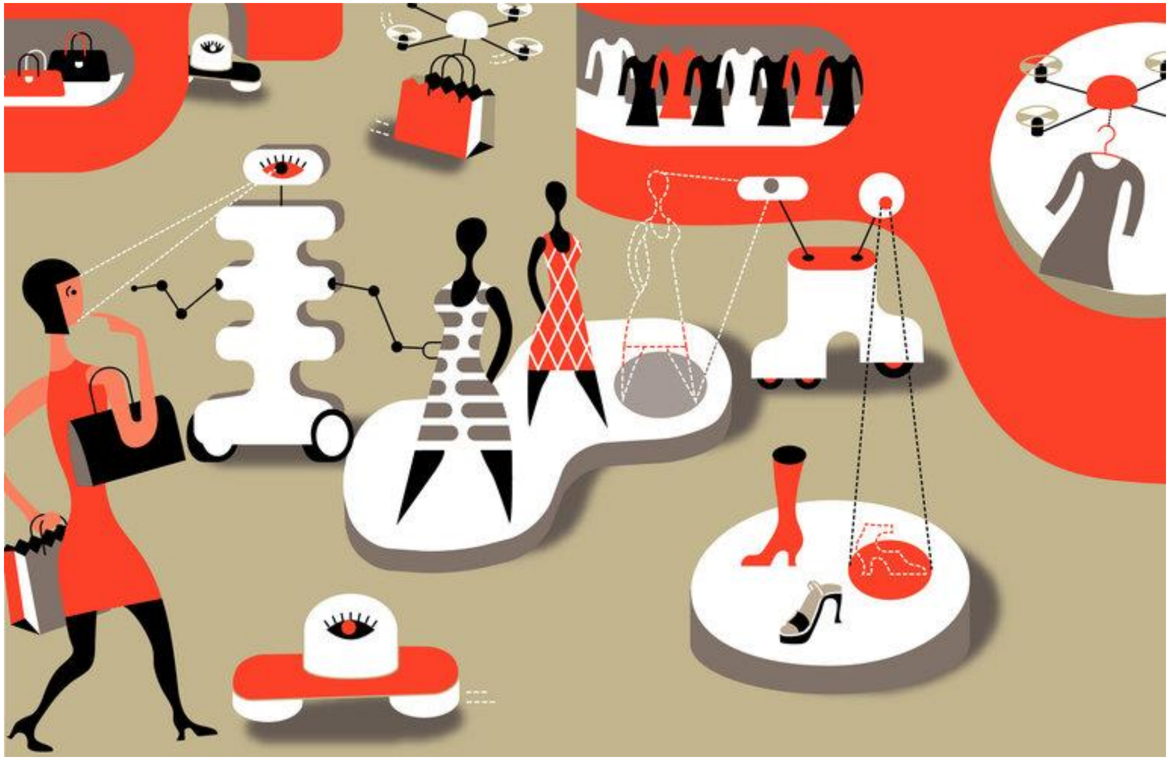

Waymo Self-Driving

Smart Agriculture

Health care



Retail



From Imagining the Retail Store of the Future - The New York Times, April 12, 2017

Manufacturing

Robots working in a German Bakery

Robots Welding

Robot in a foundry

Warehouse Logistics and Transportation

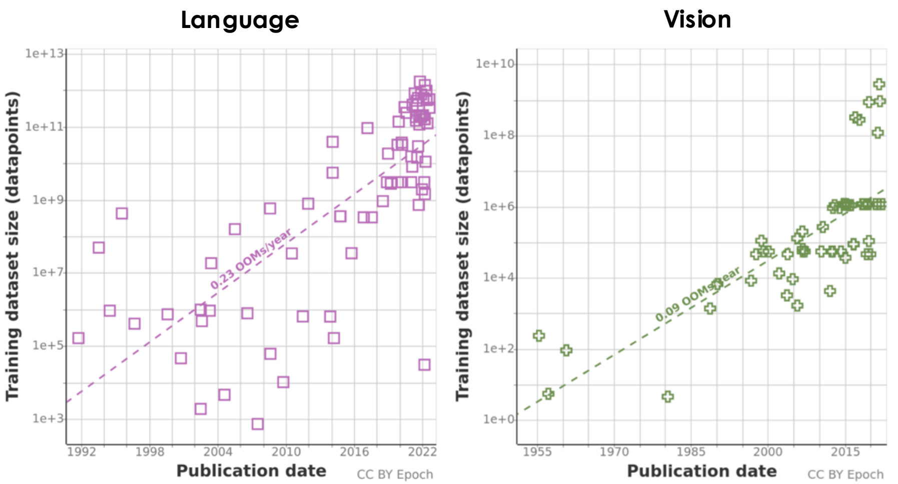
Courtesy of Dexterity.ai

Construction Management

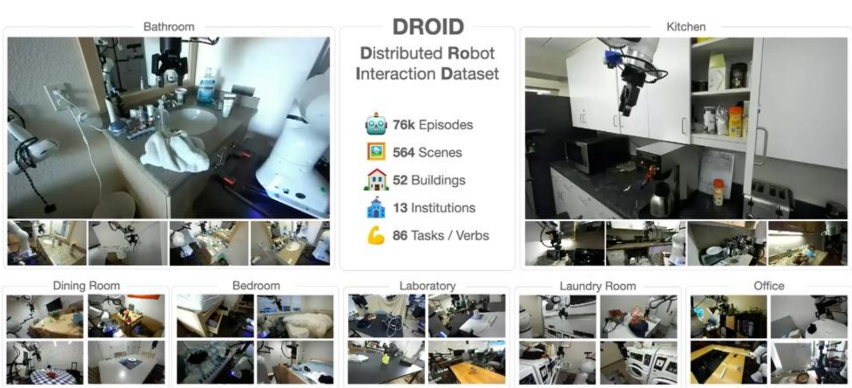
Why is this acceleration happening now?

Enabling factors

- Large Amounts of Data



https://epochai.org/blog/trends-in-training-dataset-sizes



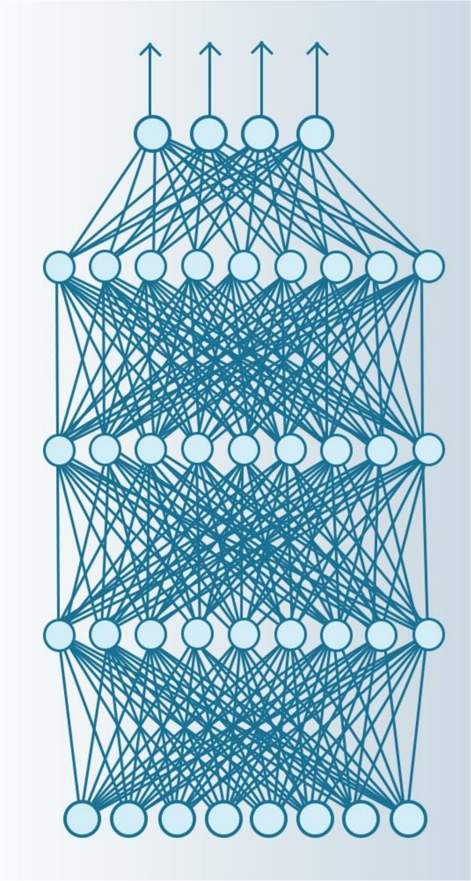
ImageNet, 2009 ShapeNet, 2015

Enabling factors

- Large Amounts of Data
- Faster hardware

Enabling factors

- Large Amounts of Data
- Faster hardware
- Improved algorithms
  - Neural Networks
  - Representation learning
  - Imitation Learning
  - Reinforcement Learning



Computer vision

→

Sensing device

→

Computational device

→

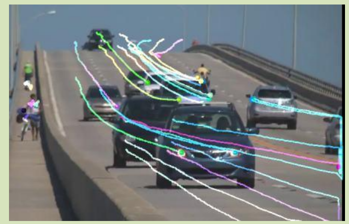
- **Information extraction**
- **Interpretation**

1. **Information extraction:** features, 3D structure, motion flows, etc…
2. **Interpretation:** recognize objects, scenes, actions, events in either single or multiple frames

Major areas in Computer Vision

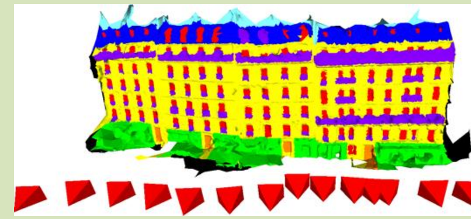
Space / Geometry

- Object shape recovery
- Depth estimation
- 3D scene reconstruction



Time / Dynamics

- Object tracking
- Dynamic scene understanding
- Motion Understanding



Semantics

- Object detection and pose estimation
- Semantic Scene understanding

Major areas in Computer Vision

Space / Geometry

- Object shape recovery
- Depth estimation
- 3D scene reconstruction

Time / Dynamics

- Object tracking
- Dynamic scene understanding
- Motion Understanding

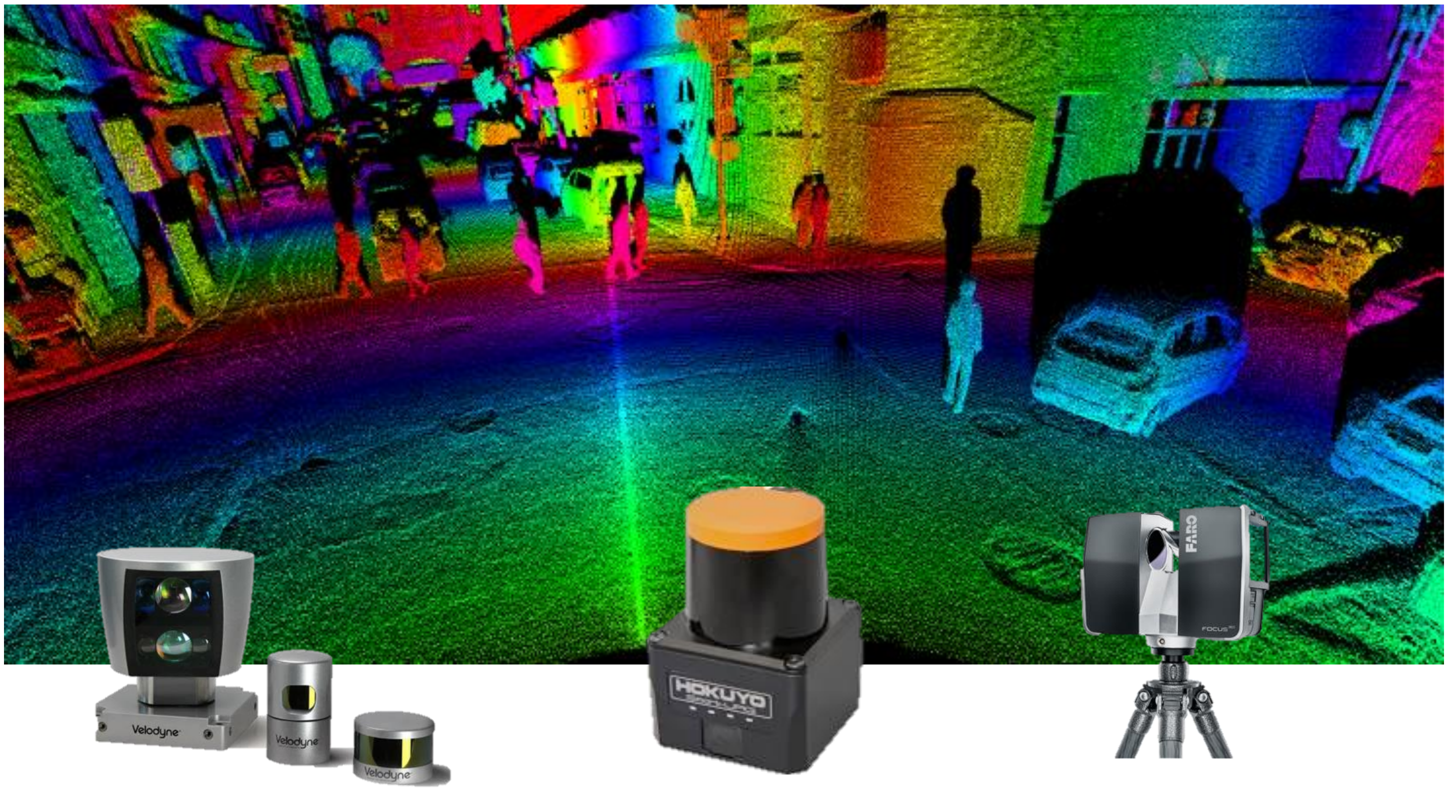
Semantics

- Object detection and pose estimation
- Semantic Scene understanding

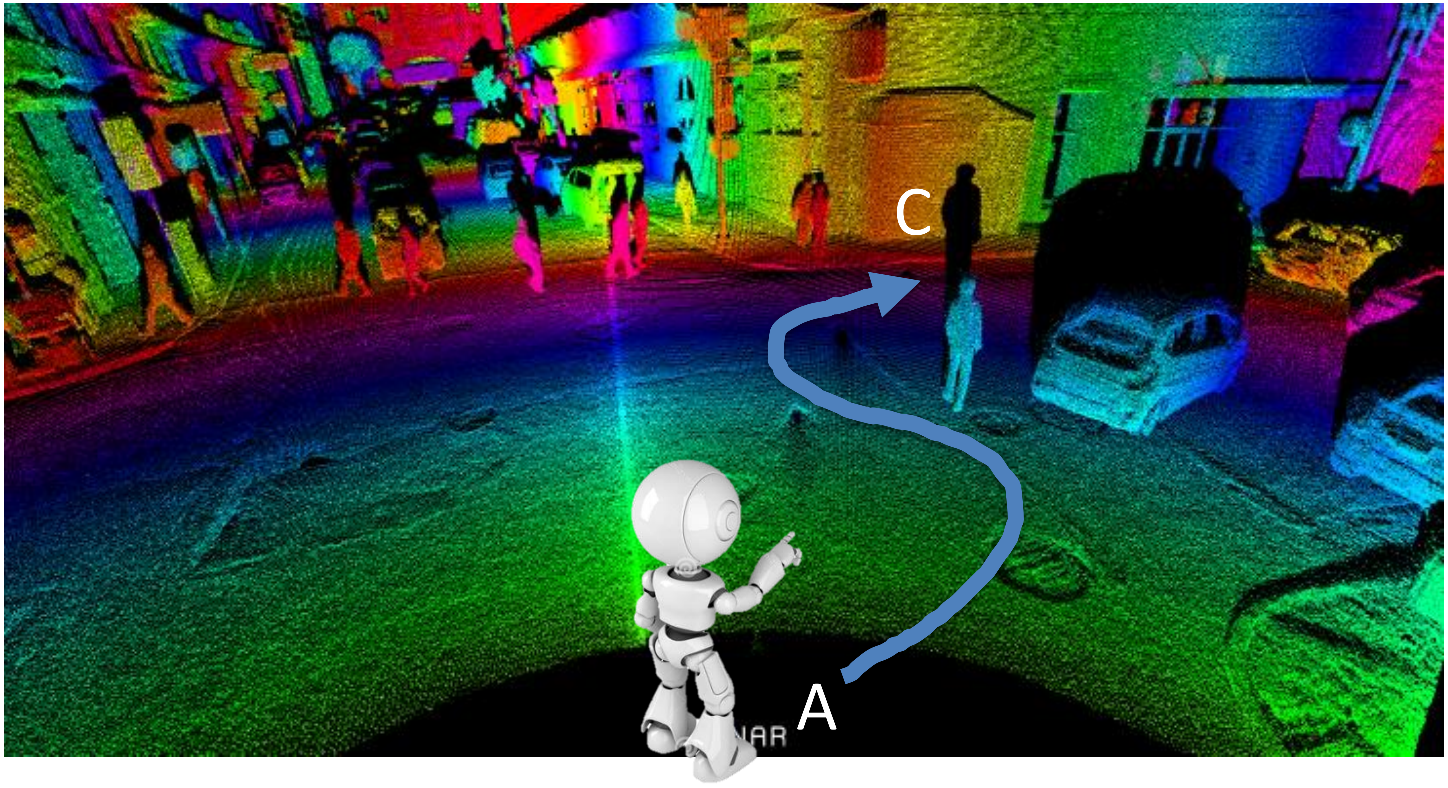
Recovering 3D models of the environments

Armeni et al. 2016

Recovering 3D models of the environments



This is critical for autonomous driving or navigation!



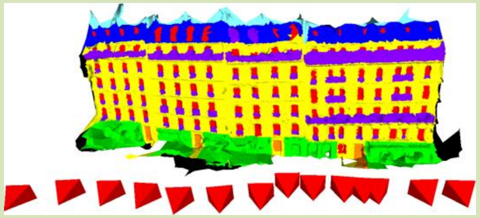
Major areas in Computer Vision

Space / Geometry

- Object shape recovery
- Depth estimation
- 3D scene reconstruction

Time / Dynamics

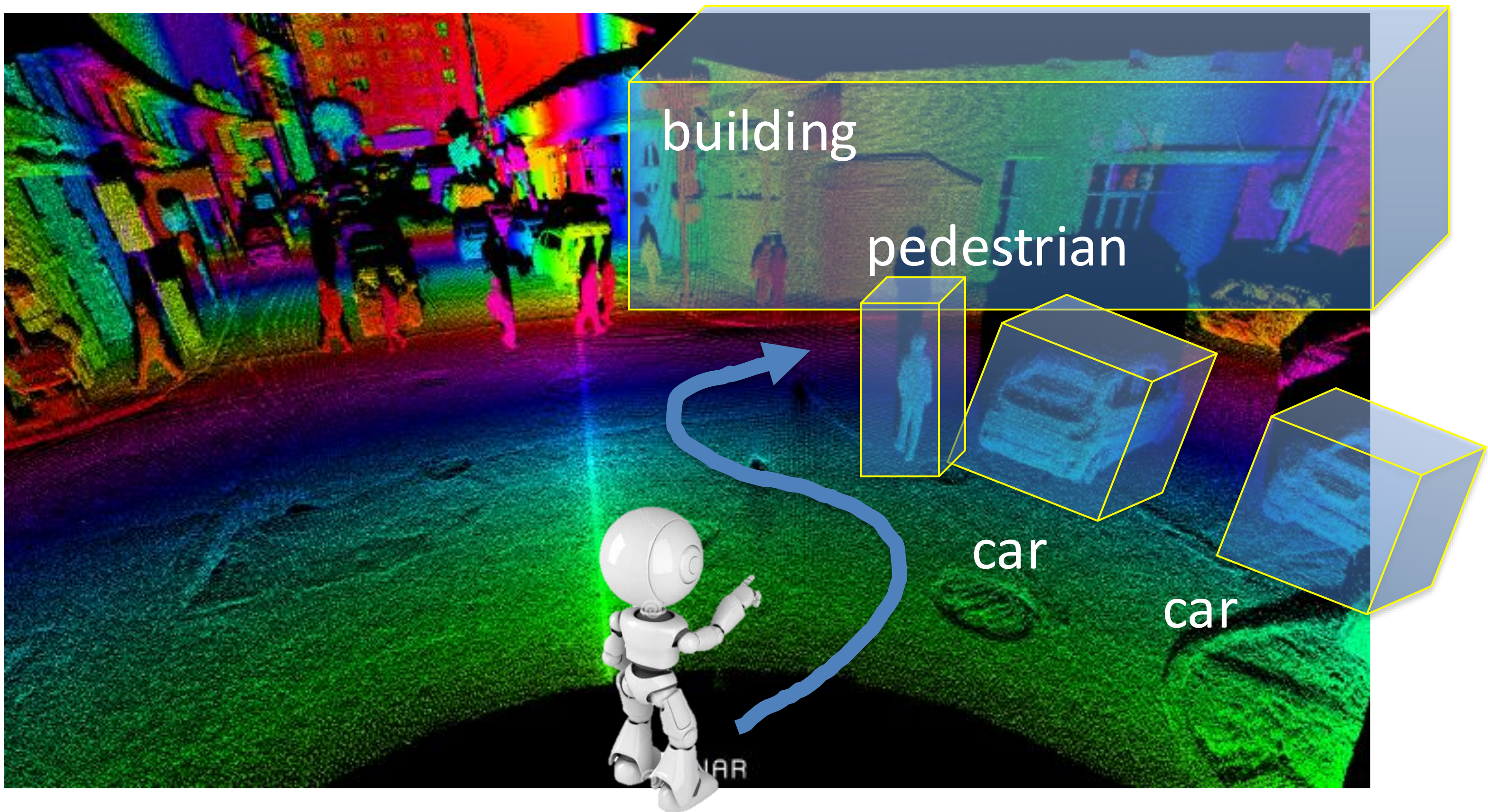
- Object tracking
- Dynamic scene understanding
- Motion Understanding



Semantics

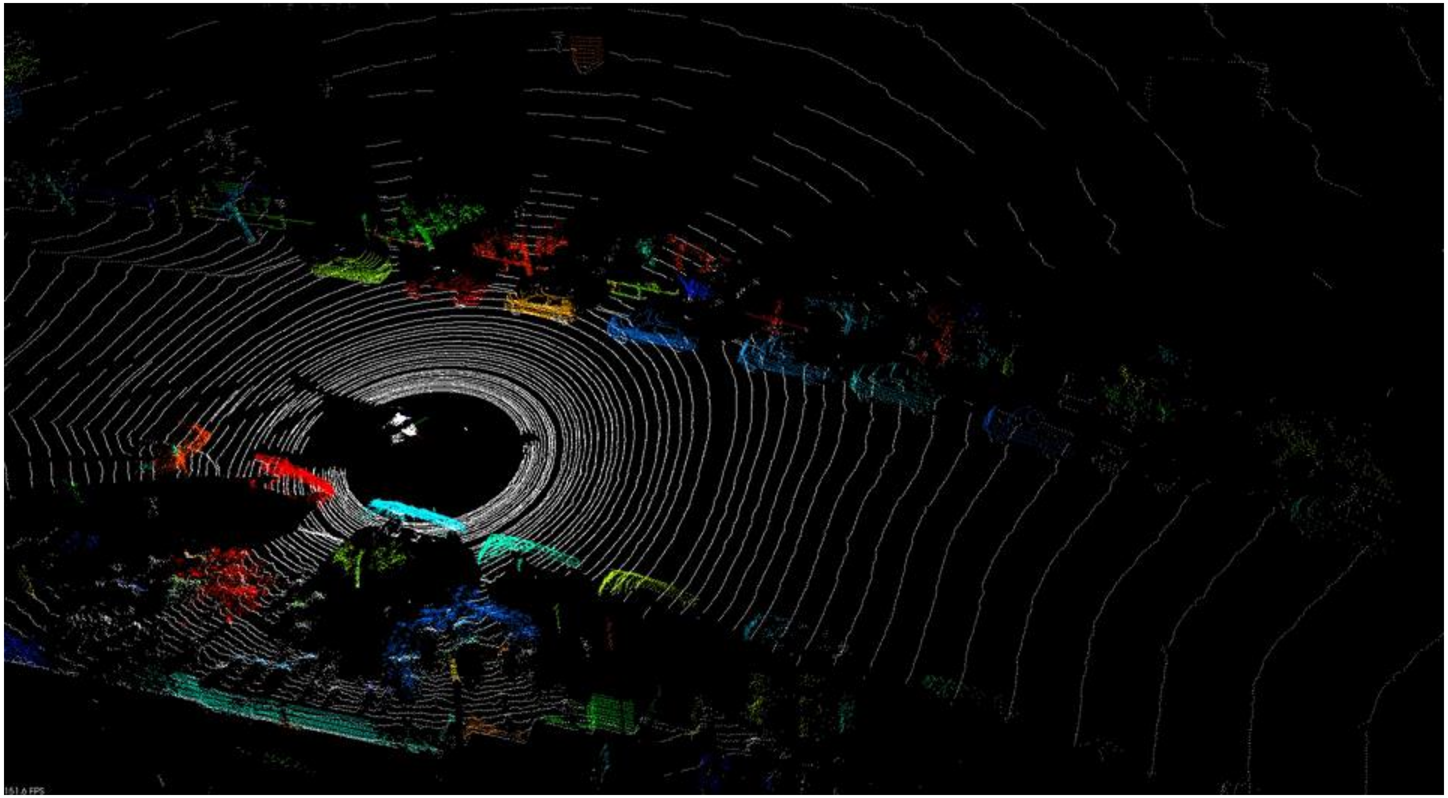
- Object detection and pose estimation
- Semantic Scene understanding

Detecting and tracking objects in the environments

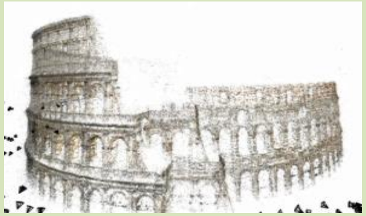


3D Scene Parsing

Held, Thrun, Savarese, 2016-206



Major areas in Computer Vision

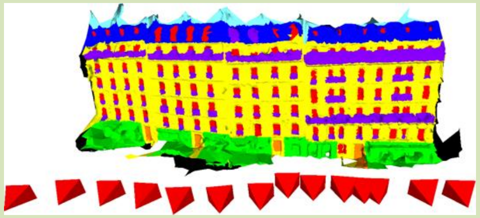


Space / Geometry

- Object shape recovery
- Depth estimation
- 3D scene reconstruction

Time / Dynamics

- Object tracking
- Dynamic scene understanding
- Motion Understanding



Semantics

- Object detection and pose estimation
- Semantic Scene understanding

CS231a = Focus on 3D w/ little Semantics

CS231n = Focus on 2D w/ a lot of Semantics

CS 231A course overview

1. Space/Geometry

   Estimating spatial properties of objects and scene from images through geometrical methods

1. Time/Dynamics + Learning

CS 231A course overview

1. Space/Geometry

   Estimating spatial properties of objects and scene from images through geometrical methods

2. Time/Dynamics + Learning

   Estimating semantic and dynamic properties of scene elements from images through learning methods

CS 231A course overview

1. Space/Geometry

   Estimating spatial properties of objects and scene from images through geometrical methods

2. Time/Dynamics + Learning

   Estimating semantic and dynamic properties of scene elements from images through learning methods

Camera systems

Establish a mapping from 3D to 2D

How to calibrate a camera

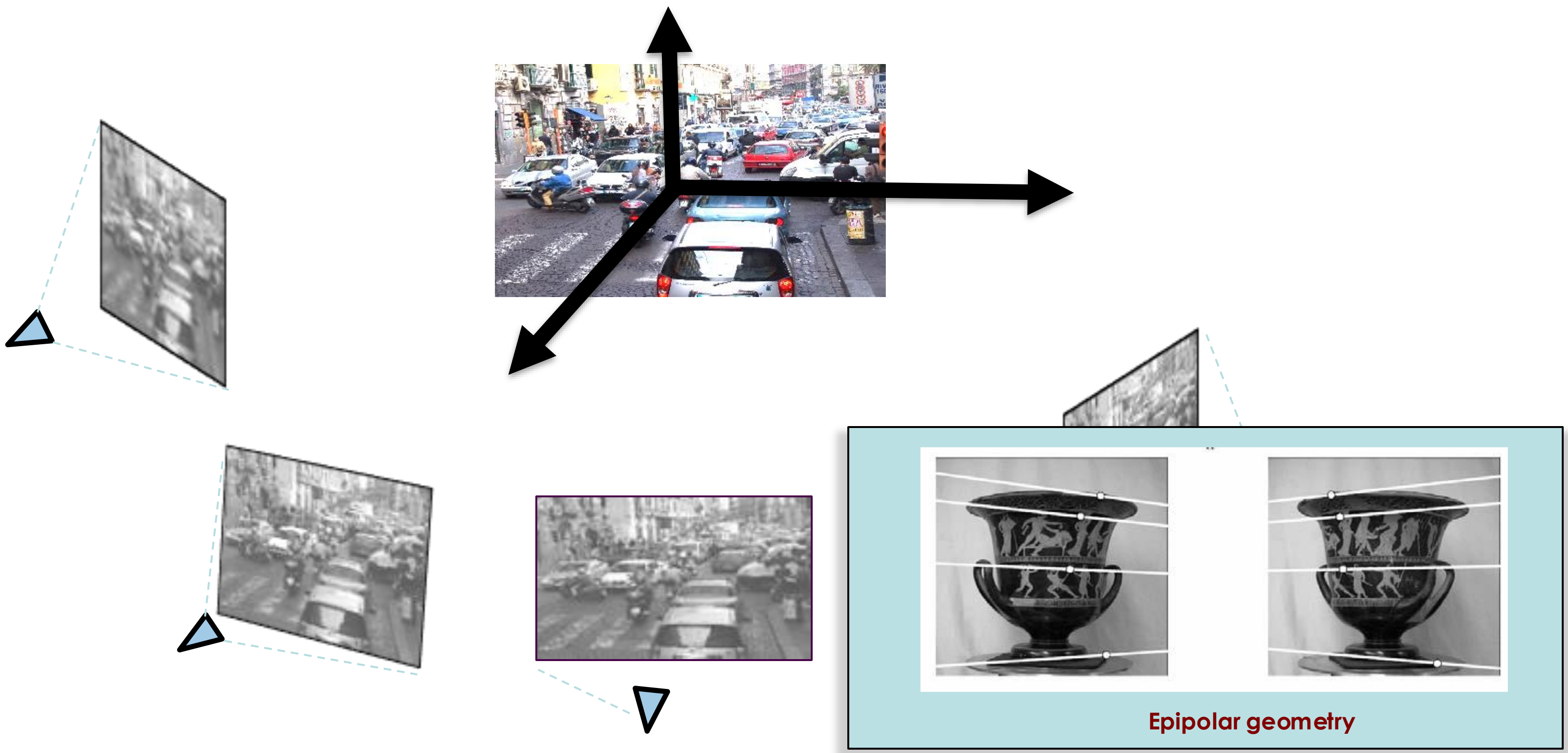
Estimate camera parameters such pose or focal length

Single view metrology

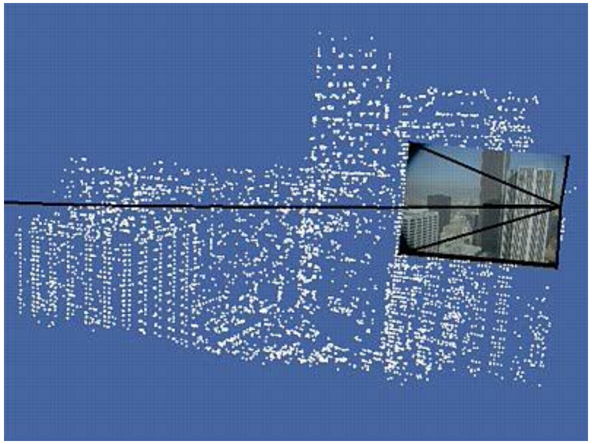
Estimate 3D properties of the world from a single image

Multiple view geometry

Estimate 3D properties of the world from multiple views



Structure from motion

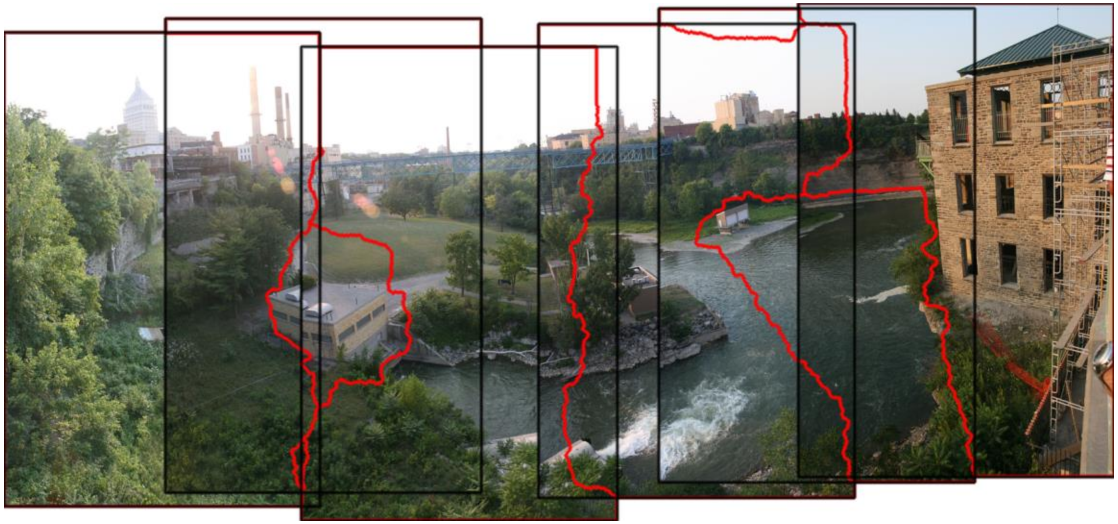


Courtesy of Oxford **Visual Geometry Group**

SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM. Keetha. CVPR'24.

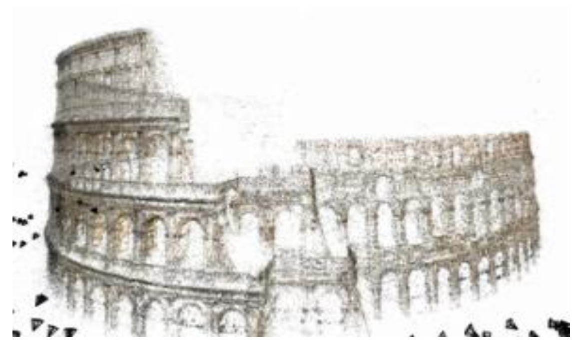
Panoramic Photography

Alcatraz Island, shown in a panorama created by image stitching. Source: Wikipedia.



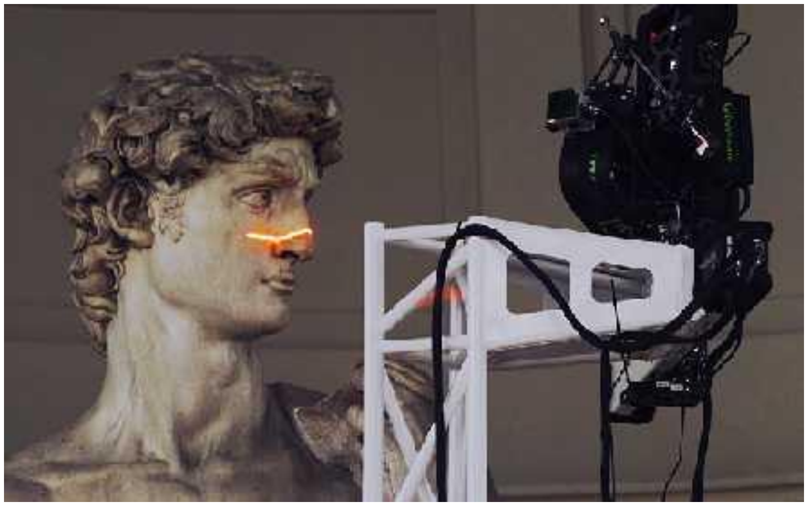
This sample image shows geometrical registration and stitching lines in panorama creation.

3D Modeling of landmarks



Microsoft Photosynth

Accurate 3D Object Prototyping



Scanning Michelangelo's *"The David"*

- The Digital Michelangelo Project
  - http://graphics.stanford.edu/projects/mich/
- 2 BILLION polygons, accuracy to .29mm

Augmented Reality

Mirriad
Advertising for the Skip Generation

AUGMENT

Meta Aria Glasses

Apple Vision Pro

CS 231A course overview

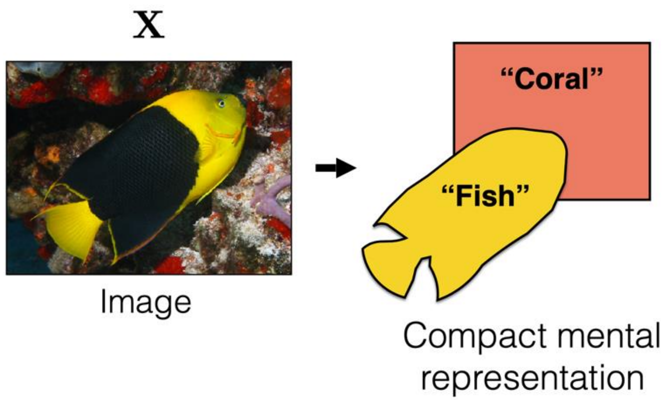
1. Space/Geometry

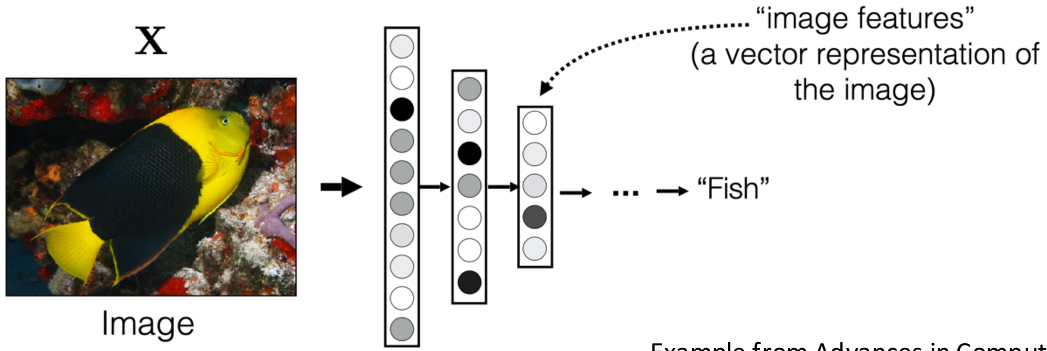
   Estimating spatial properties of objects and scene from images through geometrical methods

2. Time/Dynamics + Learning

   Estimating semantic and dynamic properties of scene elements from images through learning methods

Representations and Representation Learning

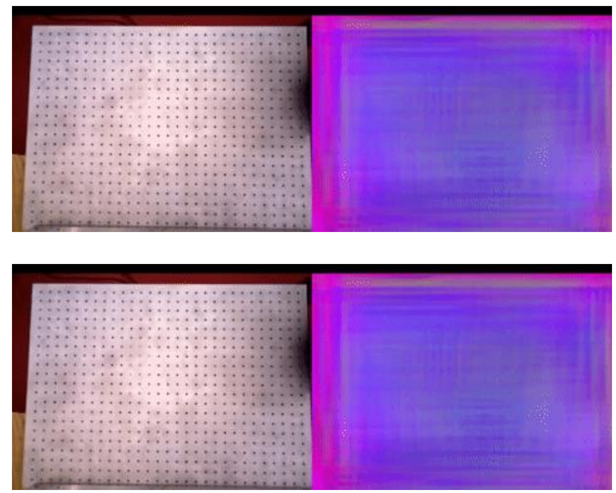




Example from Advances in Computer Vision – MIT – 6.869/6.819

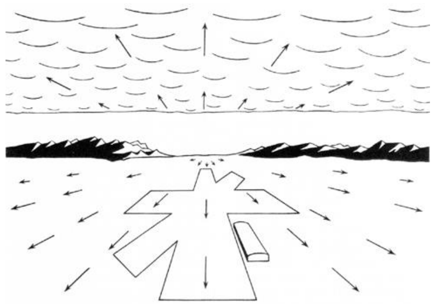
Monocular Depth Estimation and Feature Tracking

Monocular Depth Estimation



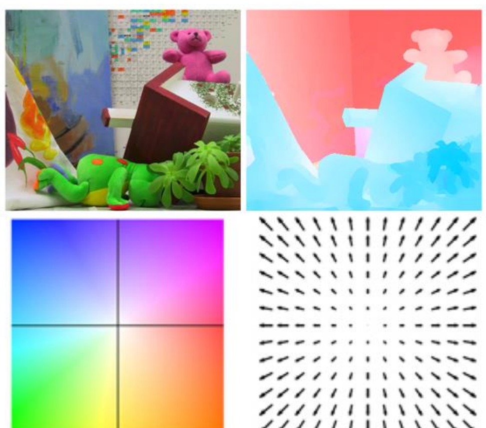
Feature Tracking

Optical and Scene Flow

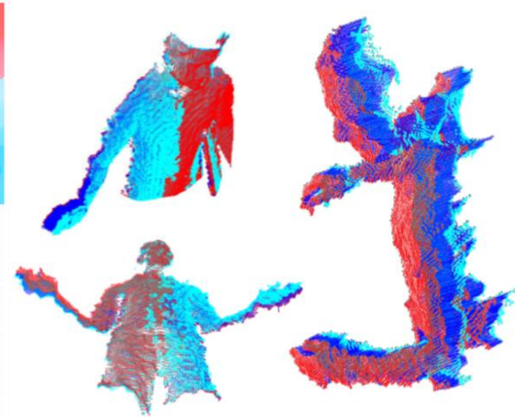


J. J. Gibson, The Ecological Approach to Visual Perception

Lucas-Kanade Feature Tracking over multiple frames. Picture adopted from OpenCV Webpage.

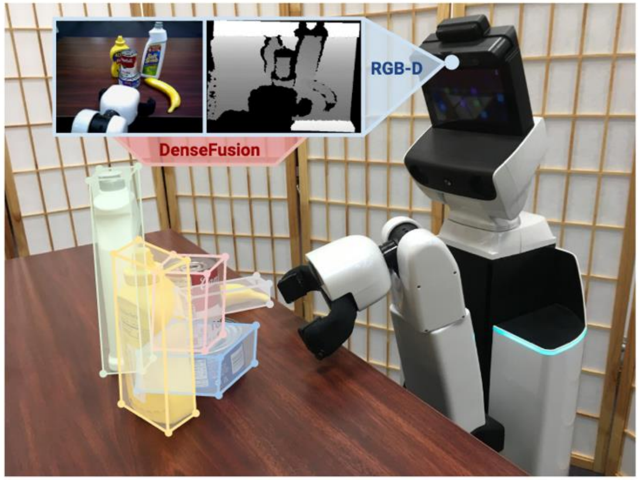


A Database and Evaluation Methodology for Optical Flow. Baker et al. IJCV. 2011



A Primal-Dual Framework for Real-Time Dense RGB-D Scene Flow. Jaimez at al. ICRA, 2015.

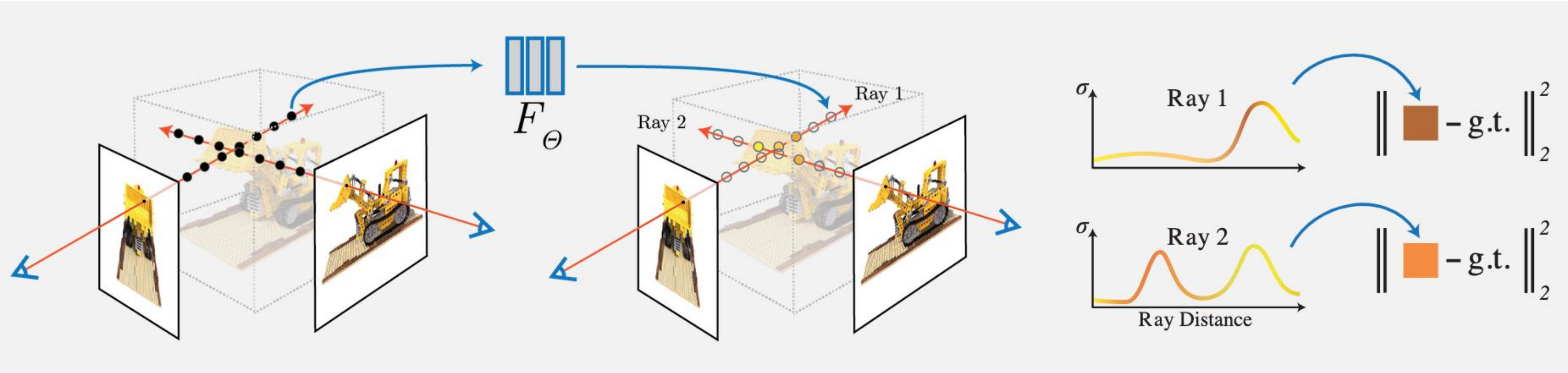
Optimal Estimation for Object Tracking



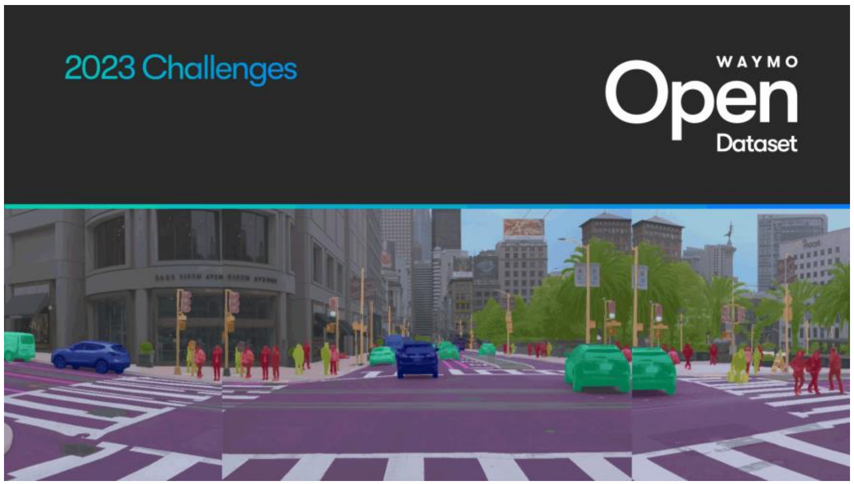
Wang et al. “Dense Fusion: 6D Object Pose Estimation by Iterative Dense Fusion”, CVPR 2019

Manuel Wühtrich et al. “Probabilistic Object Tracking using a Depth Camera”, IROS 2013

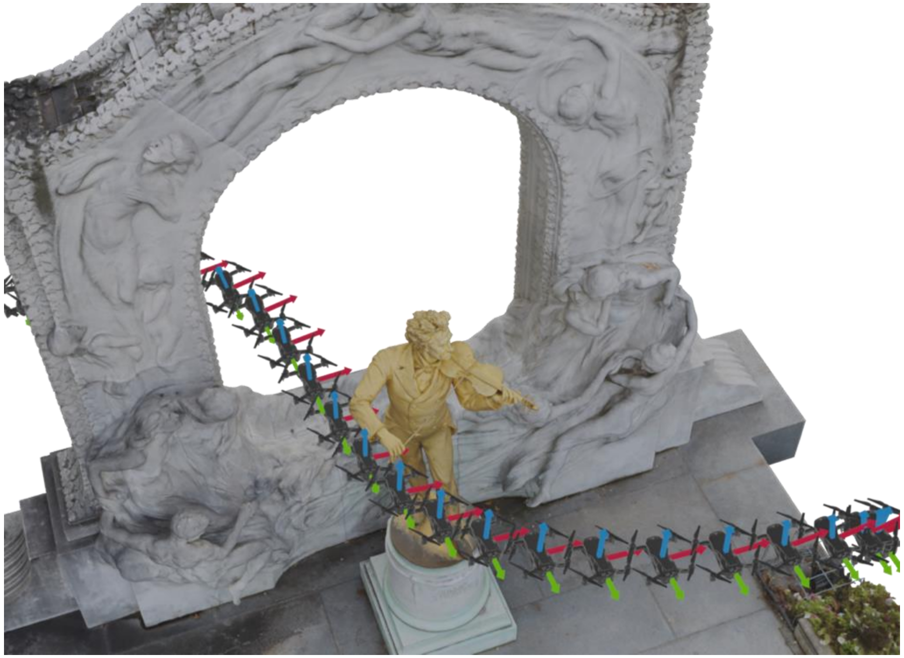
Neural Fields for View Synthesis



Autonomous navigation and safety



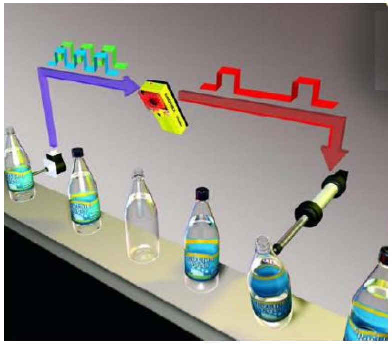
Driving Research Forward: The Waymo Open Dataset Updates and 2023 Challenges. Drago Anguelov. 2023.



Navigation in a Neural Radiance World using a Monocular camera only. Adamkiewicz, Chen et al. 2021

Personal robotics

More Applications



Factory inspection

Assistive technologies

Surveillance

Exploration and remote operations

Syllabus

Lecture	Topic
1	Introduction
2	Camera models
3	Camera calibration
4	Single view metrology
5	Epipolar geometry
6	Stereo Systems
7	Structure from Motion
8	Active and Volumetric Stereo
9	Fitting and Matching
10	Representations and Representation Learning
11	**Midterm**
12	Depth Estimation, Feature Tracking
13	Optical and Scene Flow
14	Optimal Estimation
15	Applications of Optimal Estimation
16	Neural Radiance Fields
17	Gaussian Splatting
18	Guest Lecture

April: lectures 1–9; May: lectures 10–16; June: lectures 17–18

Geometry: lectures 1–9; Dynamics & Learning: lectures 10–18

Proposal due (at lecture 8)

Milestone due (at lecture 14)

Final Project (after lecture 18)

Prerequisites

- This course requires knowledge of linear algebra, probability, statistics, machine learning and computer vision, as well as decent programming skills (CS106a,b).
- It is strongly recommended that you have at least taken either CS221 or CS229 or CS131A or have equivalent knowledge.
- We will leverage concepts from low-level image processing (CS131A) (e.g., linear filters, edge detectors, corner detectors, etc…) and machine learning (CS229) (e.g., SVM, basic Bayesian inference, clustering, neural networks, etc…) which we won’t cover in this class.
- We will provide links to background material related to CS131A and CS229 (or discuss during TA sessions) so students can refresh or study those topics if needed.

Text books

Required:

- [FP] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach* (2nd Edition). Prentice Hall, 2011.
- [HZ] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Academic Press, 2002.

Recommended:

- R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2011.
- D. Hoiem and S. Savarese. *Representations and Techniques for 3D Object Recognition and Scene Interpretation*, Synthesis lecture on Artificial Intelligence and Machine Learning. Morgan Claypool Publishers, 2011
- Learning OpenCV, by Gary Bradski & Adrian Kaehler, O'Reilly Media, 2008.
- [PB] Probabilistic Robotics, by Thrun, Fox and Burgard, MIT Press, 2005. (PDF of relevant chapter will be provided)

Course assignments

- 1 warm up problem set (HW-0) released 8pm today
- 4 problem sets
- 1 mid-term exam
- 1 project

- Look up class schedule for release and due dates.
- Problems will be released through the webpage and must be submitted through Gradescope (Use code **ZYWE5J**).

Midterm Exam

- The exam will be in person and during class time on 05/05/2025.
- CGOE students will do the exam through CGOE
- You will be updated with more details, e.g. material to be covered, review sessions etc., as we approach the midterm.

Course Projects

- Replicate an interesting paper
- Comparing different methods to a test bed
- A new approach to an existing problem
- Original research

- Write a 10-page paper summarizing your results
- Release the final code
- Give a final poster presentation
- CGOE students submit presentation videos instead.

- We will introduce project ideas in 1-2 weeks
- Important dates: look up class schedule

Course Projects

- Form your team:
  - 1-3 people
  - The larger is the team, the more work we expect from the team
  - Be nice to your partner: do you plan to drop the course?
- Evaluation
  - Quality of the project (including writing)
  - Final project poster presentation

Grading policy

- Homeworks: 29%
  - 1% for HW0
  - 7% for HW1, HW2, HW3, HW4 (each)
- Mid term exam: 20%
- Course project: 46%
  - Project proposal 4%
  - Project Milestone 10%
  - Final Project Report 22%
  - Final Poster Presentation 10%
    - For the project presentation, CGOE students send videos instead.
- Class participation: 5%
  - Questions, answers, remarks on Ed

Grading policy (HWs)

- 25% will be deducted per day late.
- Four 24-hours, one-time late submission “bonuses” are available; that is, you can use this bonus to submit your HW late after at most 24 hours. This is one time deal: After you use all your bonuses, you must adhere to the standard late submission policy.
- Max 2 bonuses can be used per assignment.
- Exceptions are made in case of medical emergencies or other exceptional and unforeseeable circumstances

Grading policy (project)

- If 1 day late, 25% off the grade for the project
- If 2 days late, 50% off the grade for the project
- Zero credits if more than 2 days
- No "late submission bonus" is allowed when submitting your progress report or project report

CS231
Introduction to Computer Vision

Next lecture: Camera systems

Instructor: Silvio Savarese