Vision Algorithms for Mobile Robotics

Lecture 01
Introduction

Davide Scaramuzza
http://rpg.ifi.uzh.ch

Today's Outline

- About me and my research lab
- What is Computer Vision?
- Why study Computer Vision?
- Example of vision applications
- Organization of the course
- Start: Visual Odometry overview

Who am I?

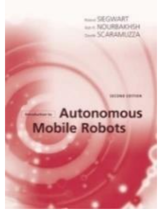
Current position

- Professor of Robotics & computer vision
  Dep. of Informatics (UZH) and Neuroinformatics (UZH & ETH)

Education

- PhD from ETH Zurich with Roland Siegwart
- Post-doc at the University of Pennsylvania with Vijay Kumar & Kostas Daniilidis

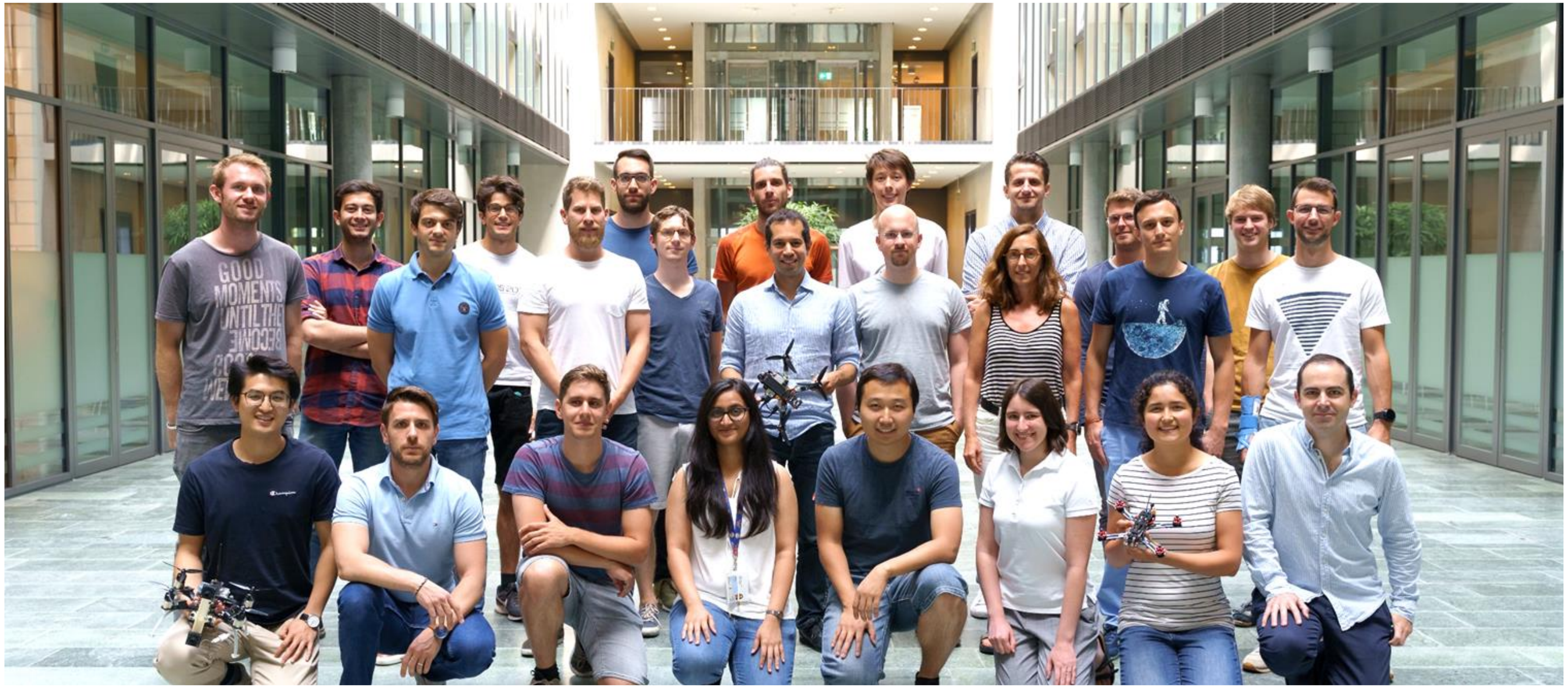
Highlights



- Book “Autonomous Mobile Robots,” 2011, MIT Press, co-authored with Roland Siegwart
- Coordinator of the European project sFly on visual navigation of micro drones, which introduced the PX4 autopilot and visual navigation of drones

My lab: the Robotics and Perception Group

- **Address**: Andreasstrasse 15, 2nd floor, next to Zurich Oerlikon train station
- **Webpage**: http://rpg.ifi.uzh.ch



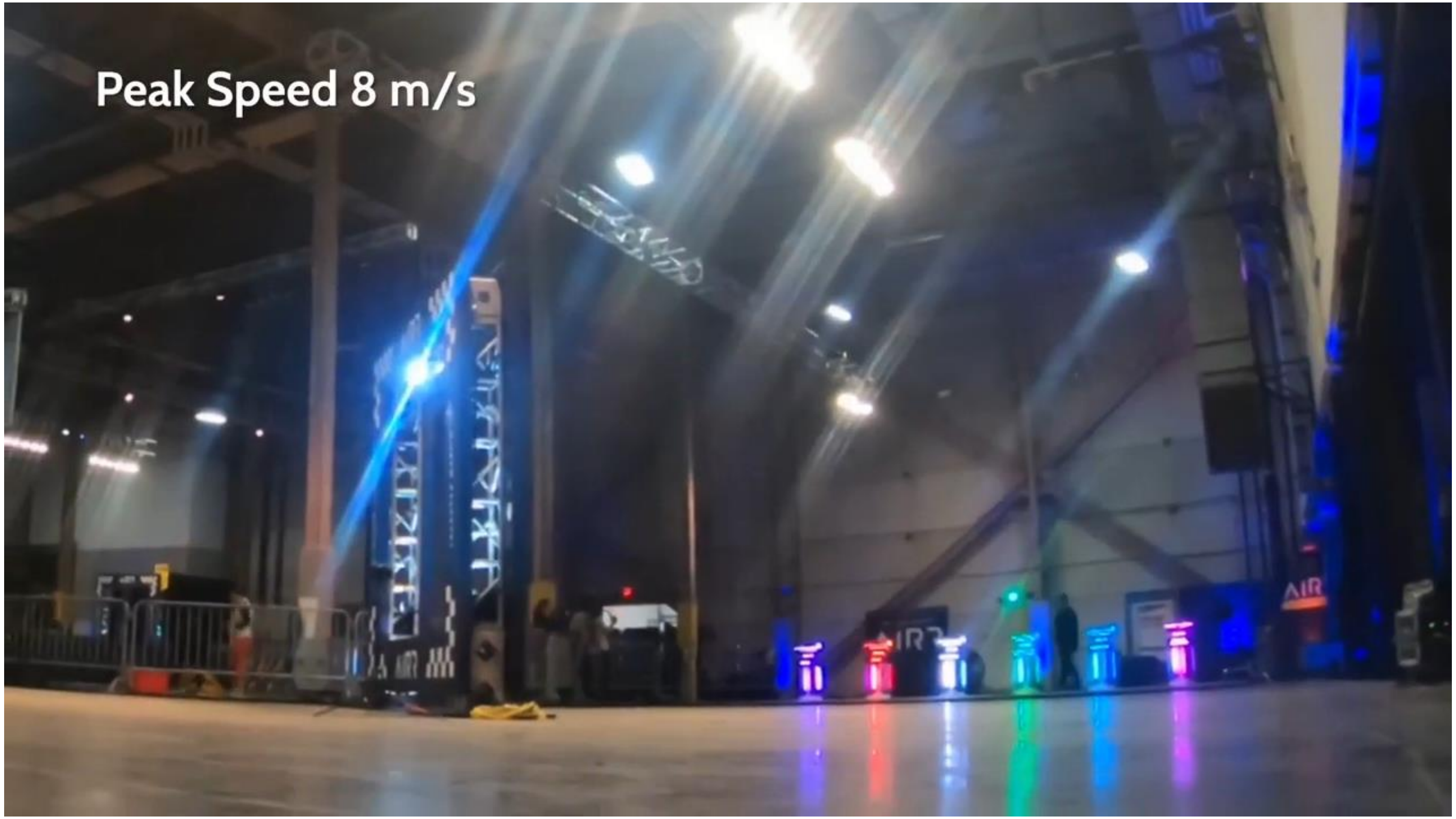
ROBOTICS &
PERCEPTION
GROUP

Research Topics

Real-time, Onboard Computer Vision and Control for Autonomous, Agile Drone Flight:

- Machine Learning
- Computer Vision
- Motion Planning & Control

Autonomous Drone Racing



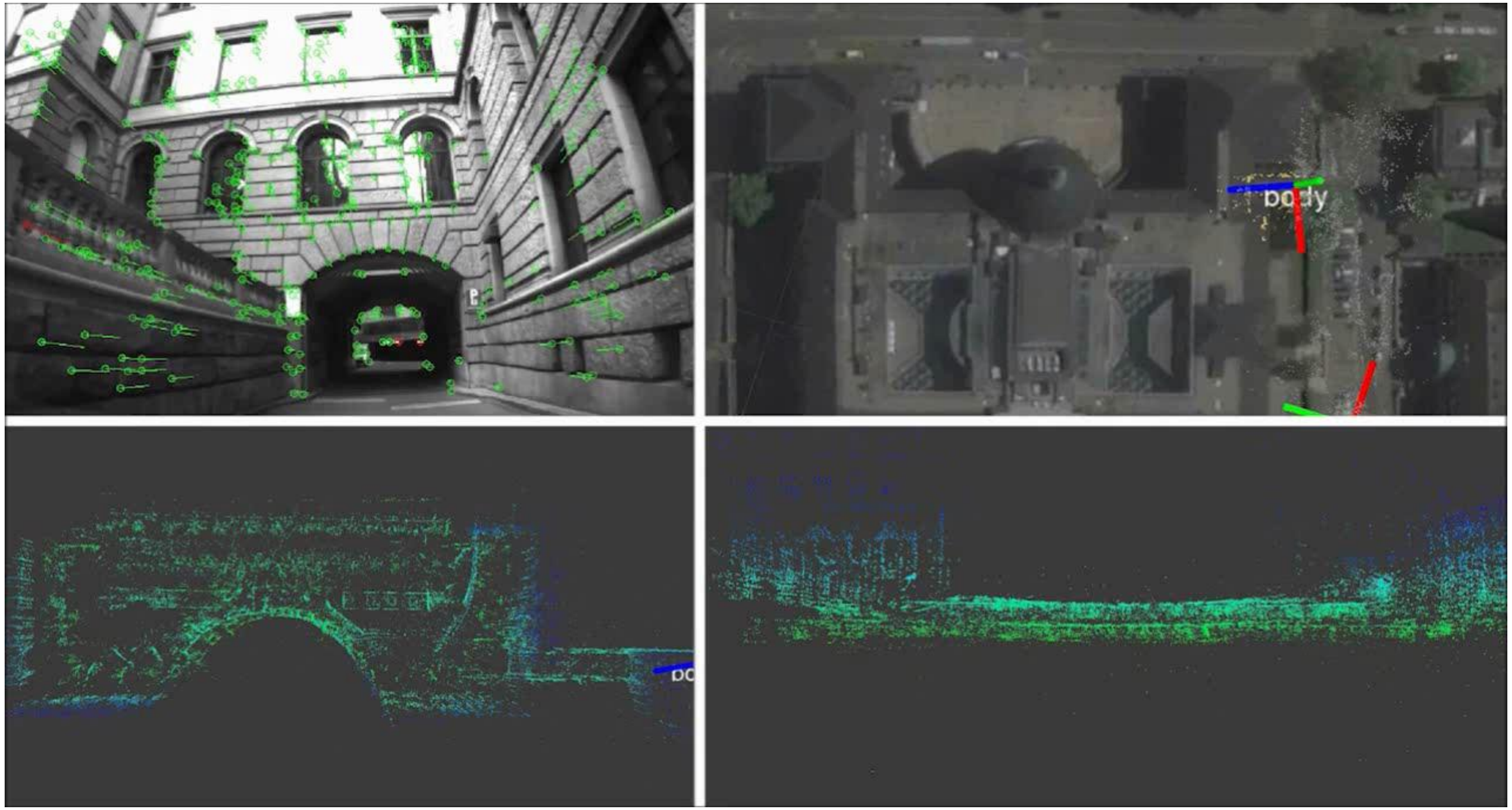
P. Foehn et al., **AlphaPilot: Autonomous Drone Racing,** RSS 2020, **Best System Paper Award**. [PDF](PDF) Video

Drone Acrobatics

Kaufmann et al., ***Deep Drone Acrobatics*, **RSS 2020, **Best Paper Award finalist.** PDF. Video.

Startup: "Zurich-Eye" – Today: Facebook-Oculus Zurich

- **Vision-based Localization and Mapping** systems for mobile robots
- Born in Sep. 2015, became **Facebook-Oculus Zurich** in Sep. 2016. Today, **100 employees**.



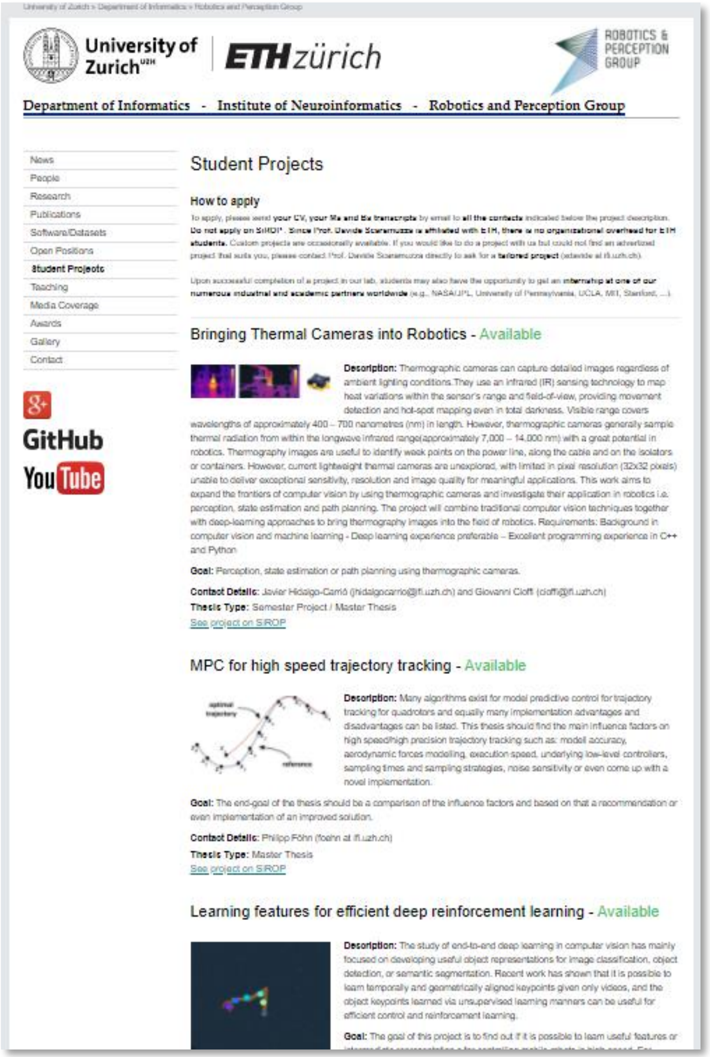
Startup: “Zurich-Eye” – Today: Facebook-Oculus Zurich

- **Vision-based Localization and Mapping** systems for mobile robots
- Born in Sep. 2015, became **Facebook-Oculus Zurich** in Sep. 2016. Today, **100 employees**.
- In 2018, Zurich-Eye launched **Oculus Quest**: https://youtu.be/xwW-1mbemGc

We will have a lecture by Christian Forster, from Oculus Zurich, end of November.

Student Projects: http://rpg.ifi.uzh.ch/student_projects.php

- **Topics**: machine learning, computer vision, control, planning, robot integration
- **Highlights**: many of our Master students have published their thesis in international conferences and won prestigious awards (e.g., ETH Medal, Fritz-Kutter Award, conference paper awards)



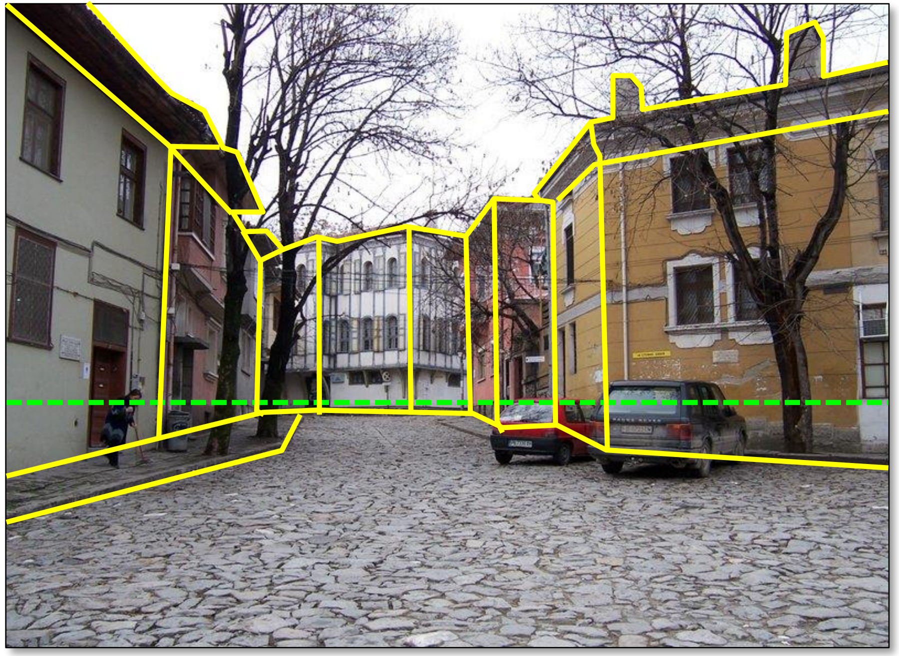
Today’s Outline

- About me and my research lab
- What is Computer Vision?
- Why study computer vision?
- Example of Vision Applications
- Organization of the course
- Start: Visual Odometry overview

What is computer vision?

Automatic extraction of "meaningful" information from images and videos

Semantic information



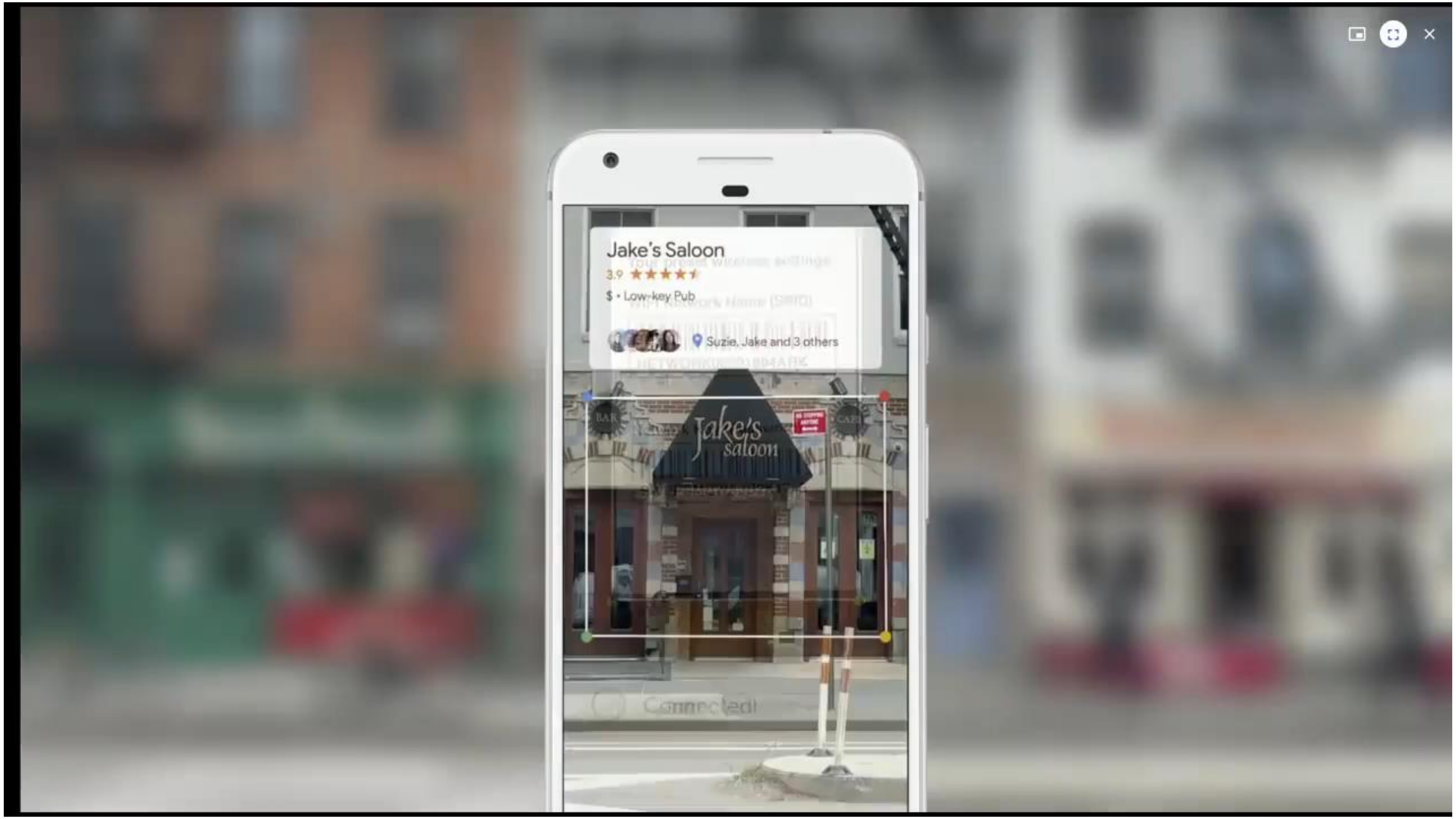
Geometric information
(this course)

Vision Demo?

Terminator 2

We are almost there!

Google Lens



Jake's Saloon
3.9
$ • Low-key Pub
Suzie, Jake and 3 others
Jake's saloon

Today's Outline

- About me and my research lab
- What is Computer Vision?
- Why study computer vision?
- Example of Vision Applications
- Organization of the course
- Start: Visual Odometry overview

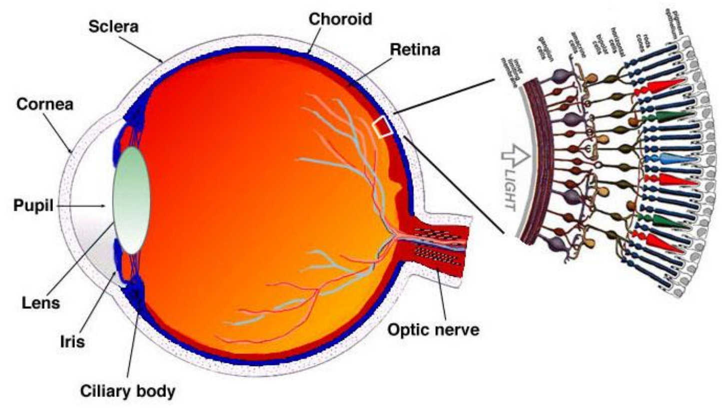
Why study computer vision?

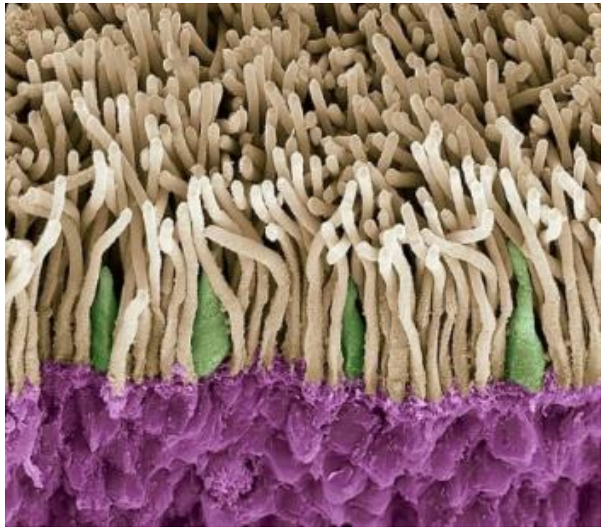
- **Relieve** humans of boring, easy tasks
- **Enhance** human abilities: human-computer interaction, visualization, augmented reality (AR)
- Perception for autonomous **robots**
- **Organize** and give access to visual **content**
- Lots of computer-vision **companies and jobs in Switzerland** (Zurich & Lausanne):

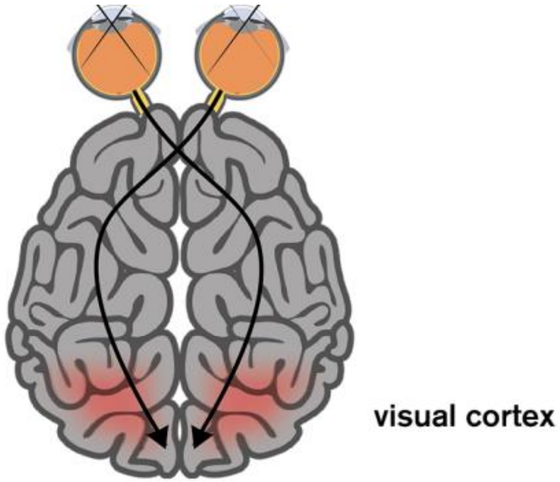
  - Facebook-Oculus (Zurich): AR/VR
  - Huawei (Zurich): event cameras, computational photography
  - Magic-Leap (Zurich): AR/VR
  - Microsoft Research (Zurich): Robotics and Hololens AR
  - Google (Zurich): Brain, Positioning Services, Street View, YouTube
  - Apple (Zurich): Autonomous Driving, face tracking
  - NVIDIA (Zurich): simulation, autonomous driving
  - Logitech (Zurich, Lausanne)
  - Disney-Research (Zurich)
  - Pix4D (Lausanne): 3D reconstruction from drones
  - VIZRT (Zurich): sport broadcasting, 3D replay
  - More on glassdoor.ch

Vision in humans

- **Vision** is our most powerful sense. **Half of primate cerebral cortex** is devoted to visual processing
- The retina is ~1,000 mm^2. Contains **130 million photoreceptors** (120 mil. rods for low light vision and 10 mil. cones for color sampling) covering a **visual field** of $\mathbf{220 \times 135}$ **degrees**
- Provides enormous amount of information: **data-rate of ~3GBytes/s**
- To match the eye resolution, we would need a **500 Megapixel** camera. But in practice the acuity of an eye is **8 Megapixels** within a **18-degree field of view** (5.5 mm diameter) around a region called **fovea**



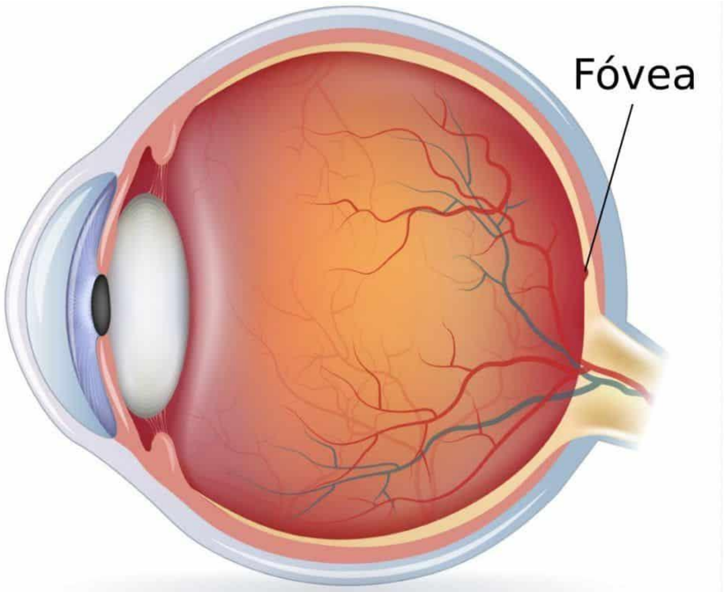




Vision in humans: how we see

- The **area we see in focus** and in **full color** represents the part of the visual field that is covered by the **fovea**
- The **fovea** is 0.35 mm in diameter, covers a visual field of **1-2 degrees**, has **high density of cone cells**
- Within the rest of the **peripheral visual field**, the image we perceive becomes more **blurry** (**rod cells**)

How we actually see. This principle is used in [foveated rendering]



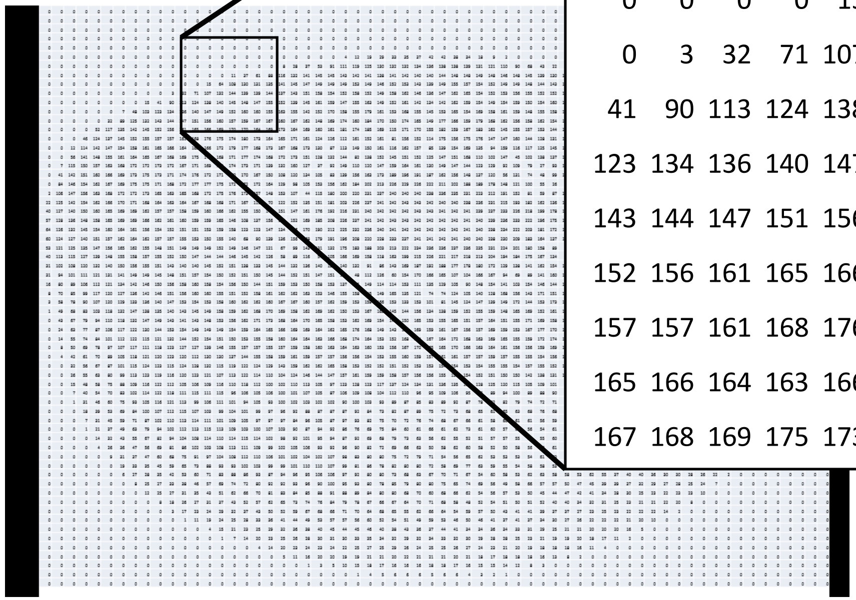
What a newborn sees every month in the first year

"Your baby sees things best from 15 to 30 cm away. This is the perfect distance for gazing up into the eyes of mom or dad. Any farther than that, and the newborn sees mostly blurry shapes because they're nearsighted. At birth, a newborn's eyesight is between 20/200 and 20/400."

http://uk.businessinsider.com/what-a-baby-can-see-every-month-for-the-first-year-of-its-life-2017-1?r=US&IR=T

Why is vision hard?

How do we go from an array of number to recognizing a fruit?



0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	8	28
0	0	0	0	0	0	11	37	61	88	116	132
0	0	0	0	15	64	108	130	131	135	141	145
0	3	32	71	107	132	144	139	139	144	137	143
41	90	113	124	138	140	145	148	147	155	152	139
123	134	136	140	147	149	152	160	160	155	163	155
143	144	147	151	156	160	157	159	167	167	160	167
152	156	161	165	166	169	170	170	164	169	173	164
157	157	161	168	176	175	174	180	173	164	165	171
165	166	164	163	166	172	179	177	168	173	167	168
167	168	169	175	173	168	171	177	174	168	172	173

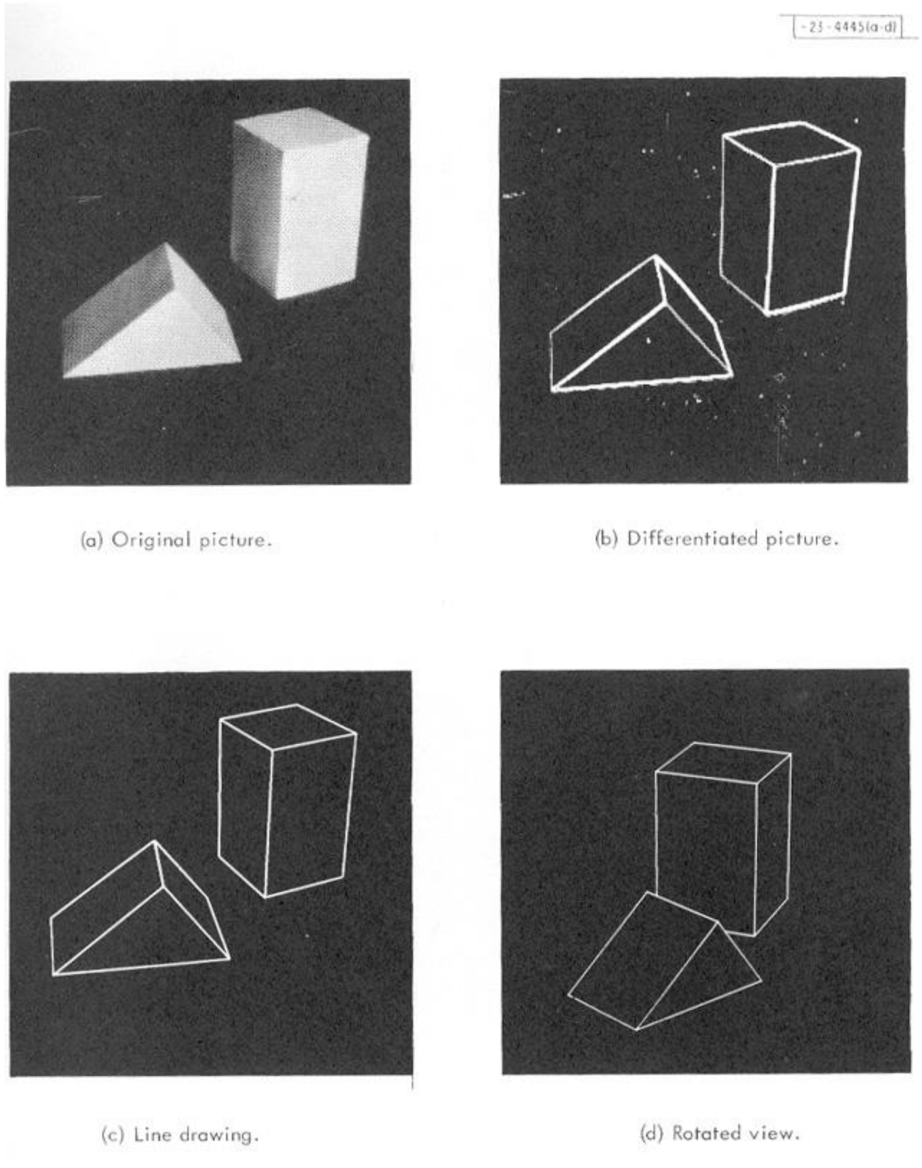
What a computer sees

What we see

Origins of computer vision

- **1963** - L. G. Roberts publishes his PhD thesis on *Machine Perception of Three Dimensional Solids*, thesis, MIT Department of Electrical Engineering

- He is the **inventor of ARPANET, the current Internet**



(a) Original picture.

(b) Differentiated picture.

(c) Line drawing.

(d) Rotated view.

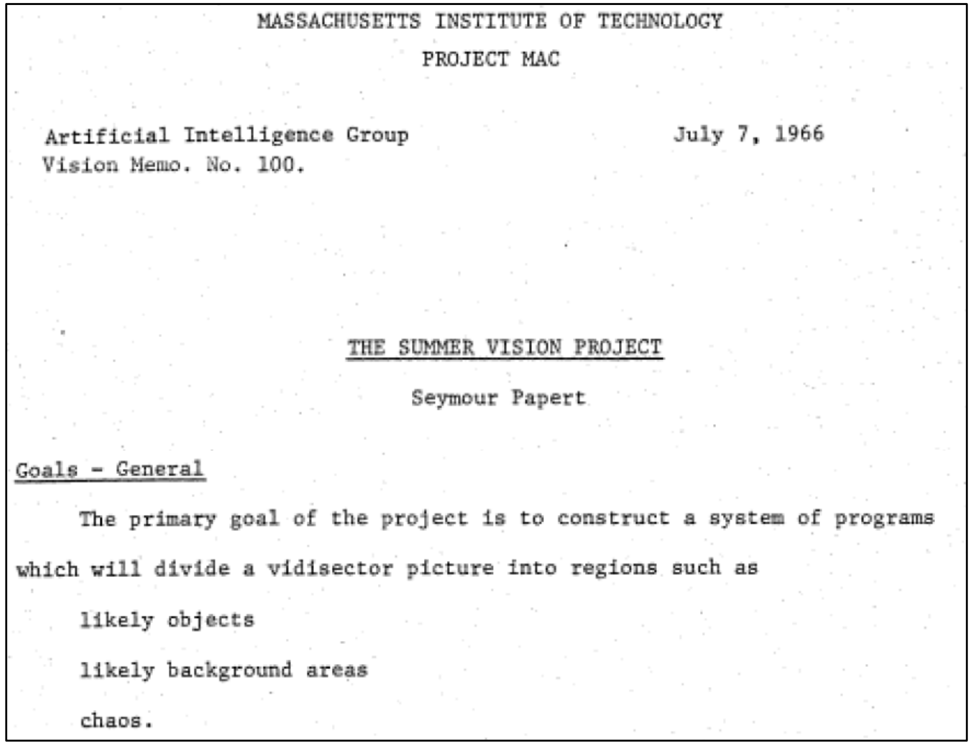
Origins of computer vision

- **1963** - L. G. Roberts publishes his PhD thesis on *Machine Perception of Three Dimensional Solids*, thesis, MIT Department of Electrical Engineering

- He is the **inventor of ARPANET, the current Internet**

- **1966** – Seymour Papert, MIT, publishes the Summer Vision Project asking students to design an algorithm to segment an image into objects and background… within summer!

- **1969** – David Marr starts developing a framework for processing visual information



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

Goals - General

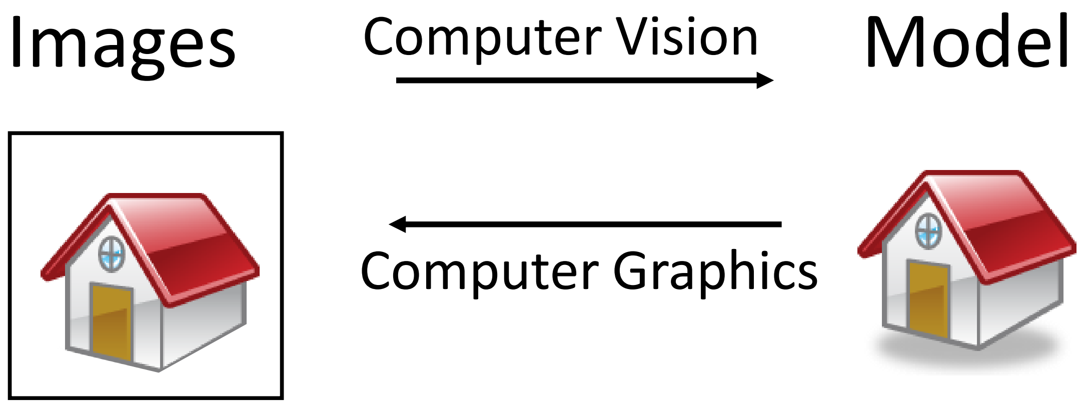
The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

likely objects

likely background areas

chaos.

Computer Vision vs Computer Graphics



Inverse problems: analysis and synthesis.

Today's Outline

- About me and my research lab
- What is Computer Vision?
- Why study computer vision?
- Example of Vision Applications
- Organization of the course
- Start: Visual Odometry overview

High-fidelity 3D Dense Reconstruction

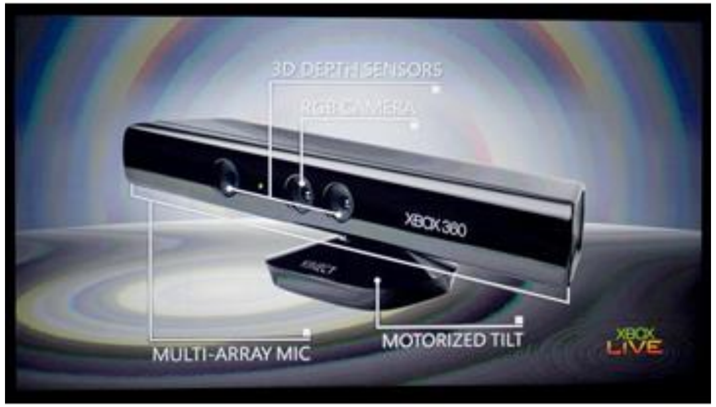
Pix4D: EPFL startup – Now a company

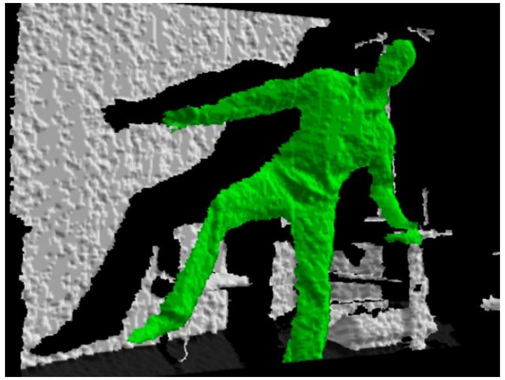
Advanced Driving Assistance Systems (ADAS)

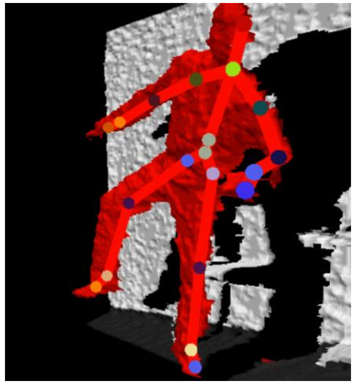
Mobileye: Vision system used at **Tesla, BMW, GM, Volvo** models. Bought by **Intel in 2017 for 15 billion USD**!

- **Pedestrian & car** collision warning
- **Lane departure** warning
- **Safety distance** monitoring and warning

Video gaming

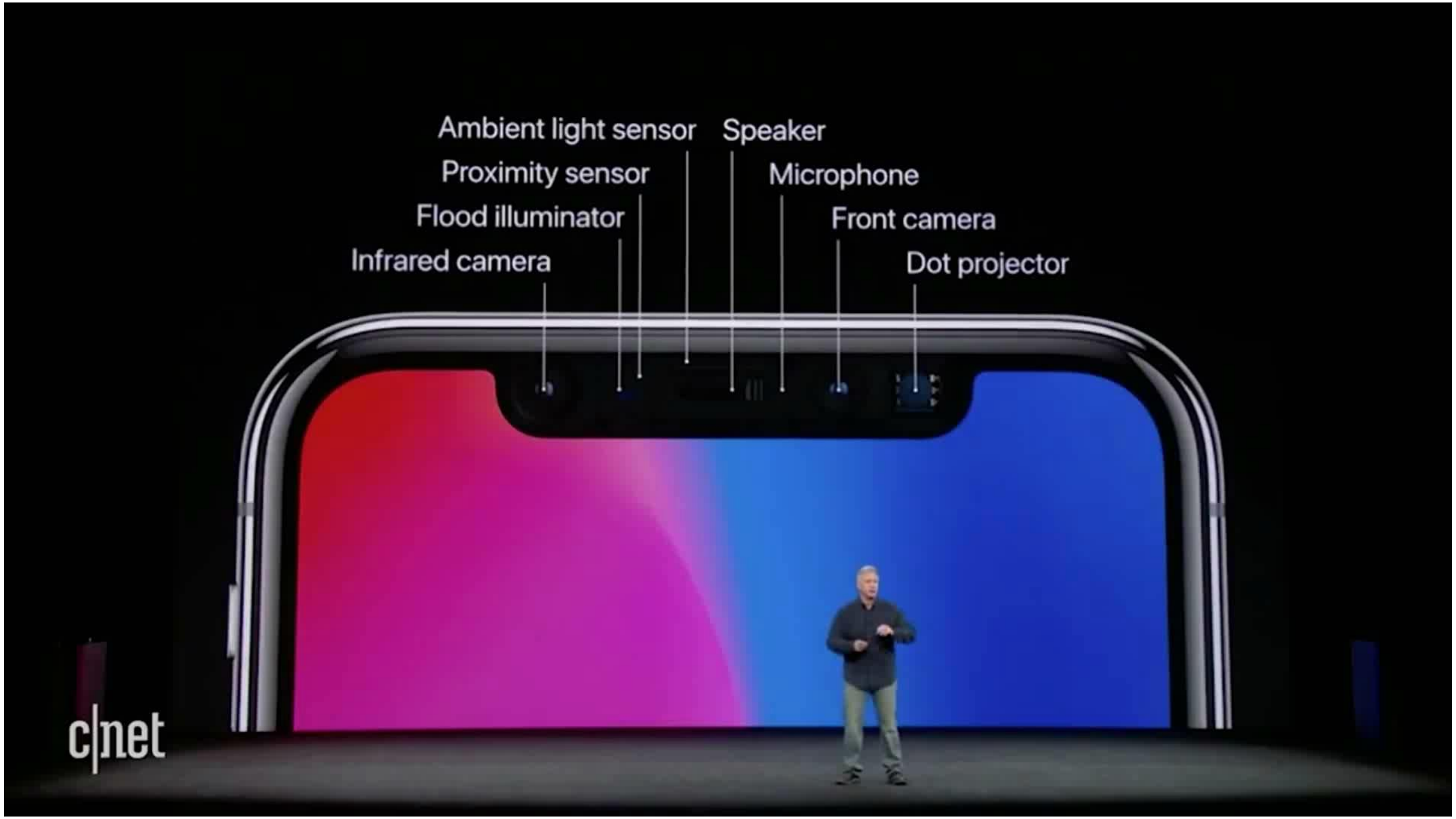






Microsoft Xbox Kinect sensor

Smartphones



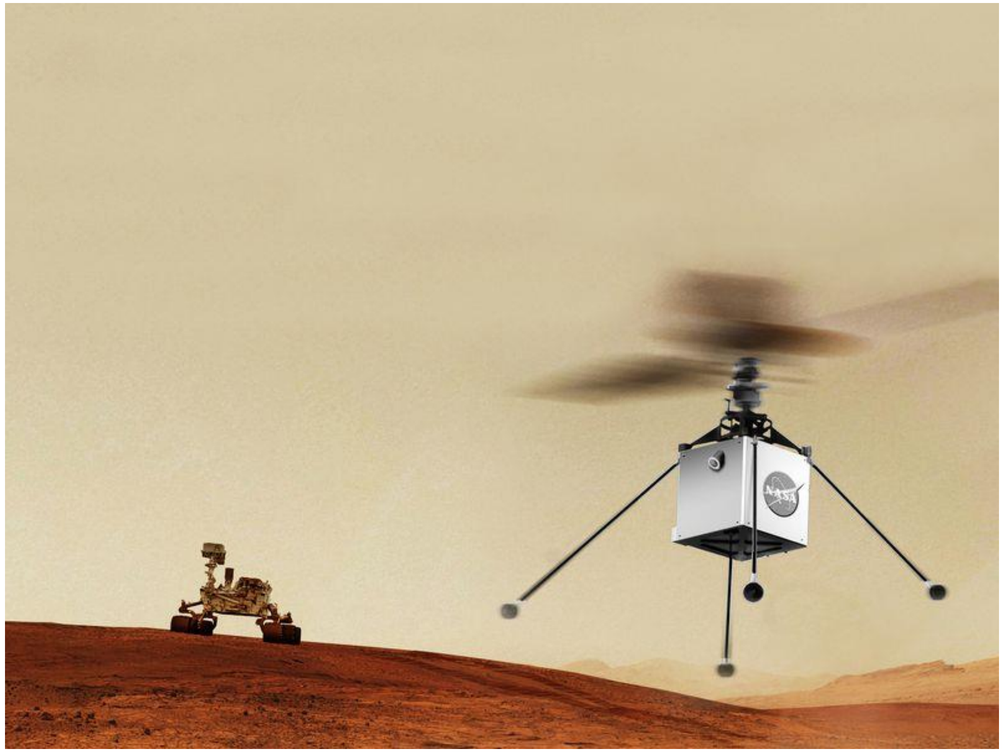
iPhone X presentation

Space applications

NASA'S Mars Exploration Curiosity Rover launched in 2011 featured 17 cameras. 19 cameras are on the Perseverance rover launched in 2020. Computer vision used for:

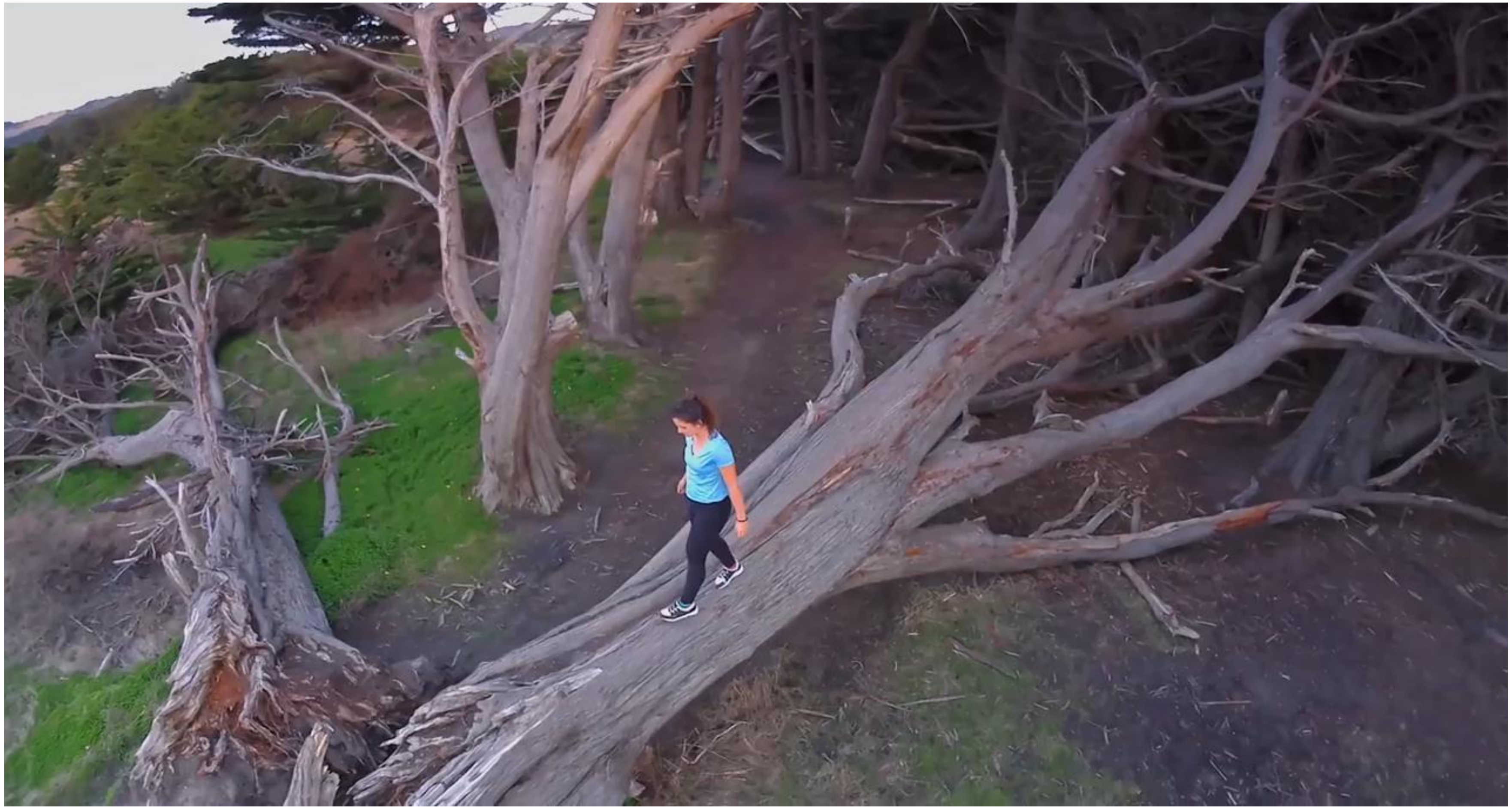
- Panorama stitching
- 3D terrain modeling
- Obstacle detection
- Visual odometry
- For more, read “Computer Vision on Mars” by Matthies et al.

Space applications



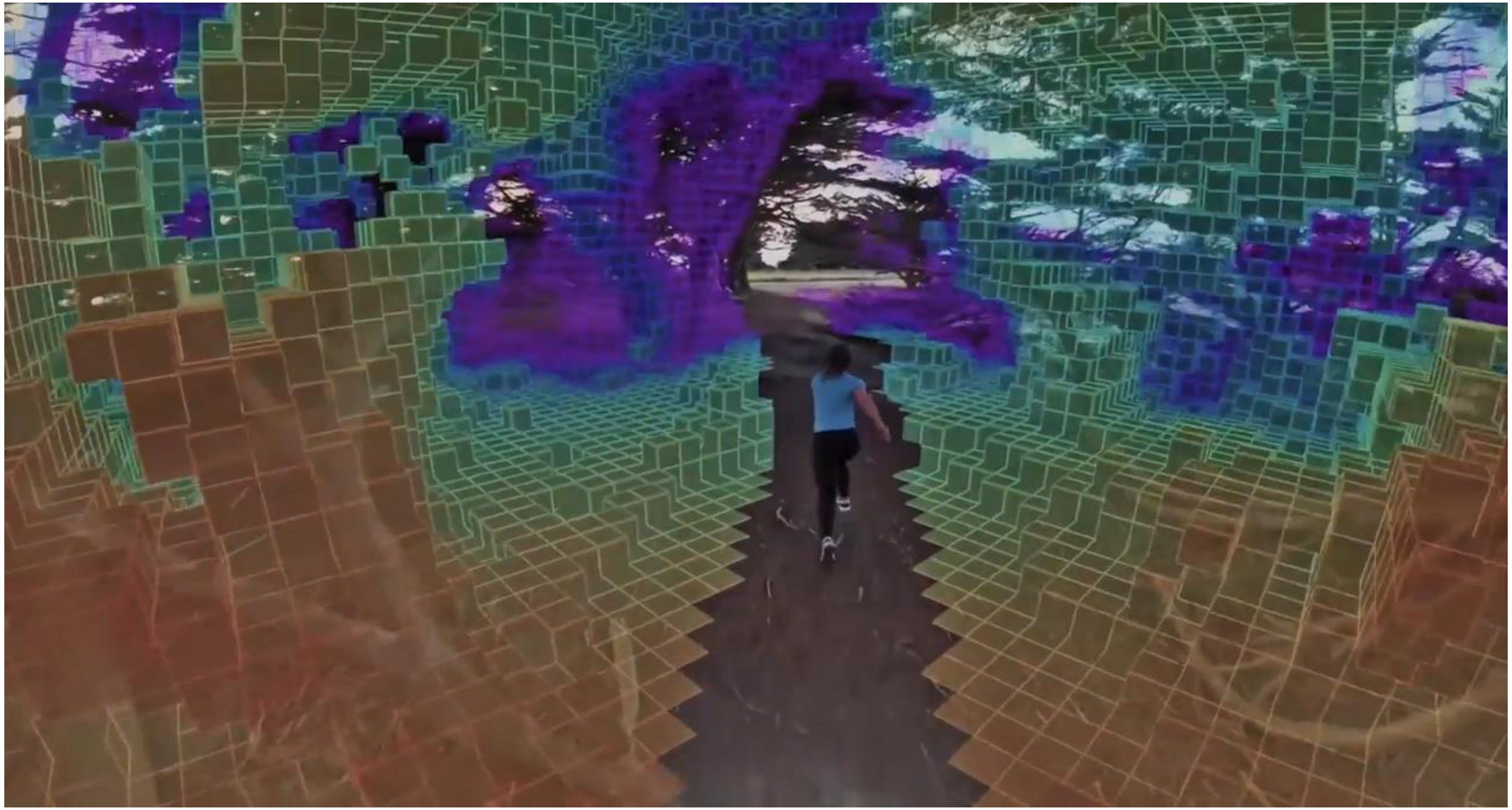
The NASA Mars Helicopter supposed to land on Mars on February 18, 2020.
It will use onboard visual inertial odometry.
https://en.wikipedia.org/wiki/Mars_Helicopter_Ingenuity

Drones



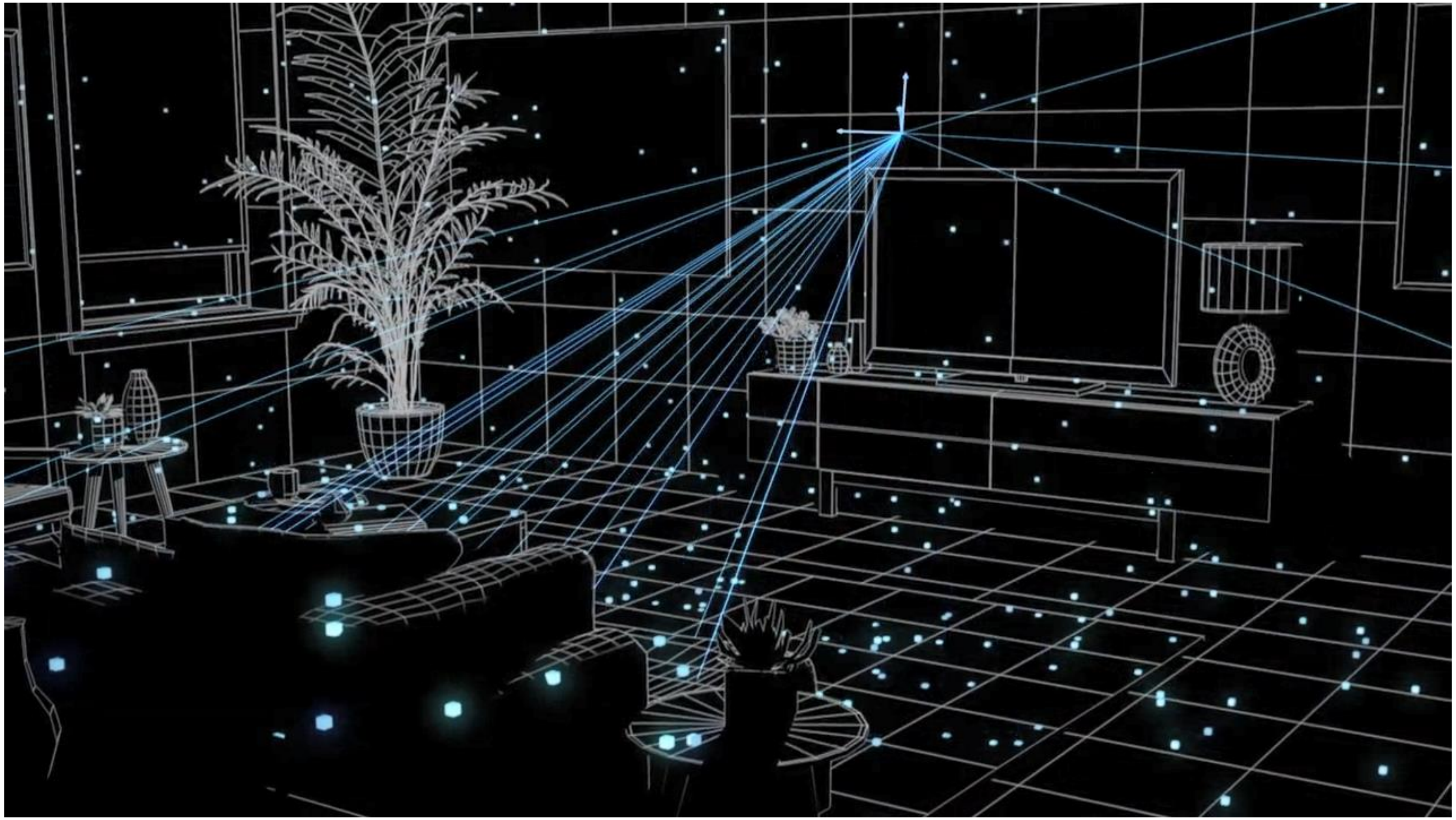
The Skydio R1 drone features 13 cameras for obstacle avoidance, visual odometry, and person following

Drones



The Skydio R1 drone features 13 cameras for obstacle avoidance, visual odometry, and person following

VR/AR



Oculus Quest features four cameras for head-pose tracking

Today’s Outline

- About me and my research lab
- What is Computer Vision?
- Why study computer vision?
- Example of Vision Applications
- Organization of the course
- Start: Visual Odometry overview

Instructors

Lecturer: Davide Scaramuzza

- **Contact**: sdavide (at) ifi (dot) uzh (dot) ch
- **Office hours**: every Thursday from 16:30 to 18:30
  both in person or via ZOOM possible
  (please announce yourself by email)

- **Exercise instructors and Teaching Assistants**: Manasi Muglikar and Daniel Gehrig

Manasi Muglikar
muglikar (at) ifi (dot) uzh (dot) ch

Daniel Gehrig
dgehrig (at) ifi (dot) uzh (dot) ch

http://rpg.ifi.uzh.ch/people.html

Organization of this Course

Lectures:

- **10:15 to 12:00** every week. **After class, I can stay for 10 more minutes** for questions
- Room: ETH HG D 1.2, Rämistrasse 101, 8092 Zurich.

Exercises:

- 14:15 to 16:00: Starting from Sep. 24 (Lecture 02). Then roughly every week.
- Room: same as above

Official course website: http://rpg.ifi.uzh.ch/teaching.html

- Check it out for the course **schedule updates**, **lecture slides**, and **updates**

Learning Objectives

- **High-level goal:** learn to implement visual odometry algorithms used in current mobile robots (drones, cars, Mars rovers), and Virtual-reality (VR) and Augmented reality (AR) products: e.g., Google Visual Positioning Service, Oculus Quest, Microsoft HoloLens, Magic Leap.

- You will also learn **to implement the fundamental computer vision algorithms** used in mobile robotics, in particular:
  - image formation,
  - filtering,
  - feature extraction,
  - multiple view geometry,
  - dense reconstruction,
  - tracking,
  - image retrieval,
  - event-based vision,
  - visual-inertial odometry, Simultaneous Localization And Mapping (SLAM),
  - and some basics of deep learning.

Course Schedule

- For updates, slides, and additional material: http://rpg.ifi.uzh.ch/teaching.html

17.09.2020	Lecture 01 - Introduction to Computer Vision and Visual Odometry No exercise today
24.09.2020	Lecture 02 - Image Formation 1: perspective projection and camera models Exercise 01- Augmented reality wireframe cube
01.10.2020	Lecture 03 - Image Formation 2: camera calibration algorithms Exercise 02 - PnP problem
08.10.2020	Lecture 04 - Filtering & Edge detection No Exercise today
15.10.2020	Lecture 05 - Point Feature Detectors, Part 1 Exercise 03 - Harris detector + descriptor + matching
22.10.2020	Lecture 06 - Point Feature Detectors, Part 2 Exercise 04 - SIFT detector + descriptor + matching
29.10.2020	Lecture 07 - Multiple-view geometry Exercise 05 - Stereo vision: rectification, epipolar matching, disparity, triangulation
05.11.2020	Lecture 08 - Multiple-view geometry 2 Exercise 06 - Eight-Point Algorithm
12.11.2020	Lecture 09 - Multiple-view geometry 3 Exercise 07 - P3P algorithm and RANSAC
19.11.2020	Lecture 10 - Multiple-view geometry 3 continued Exercise session: Intermediate VO Integration
26.11.2020	Lecture 11 - Optical Flow and Tracking (Lucas-Kanade) Exercise 08 - Lucas-Kanade tracker
03.12.2020	Lecture 12a (1st hour) - Dense 3D Reconstruction and Place recognition Exercise session replaced by Lecture 12c: Deep Learning Tutorial
10.12.2020	Lecture 13 - Visual inertial fusion Exercise 09 - Bundle Adjustment
17.12.2020	Lecture 14 - Event based vision After the lecture, if the Covid-19 rules allow, we will visit Scaramuzza's lab.

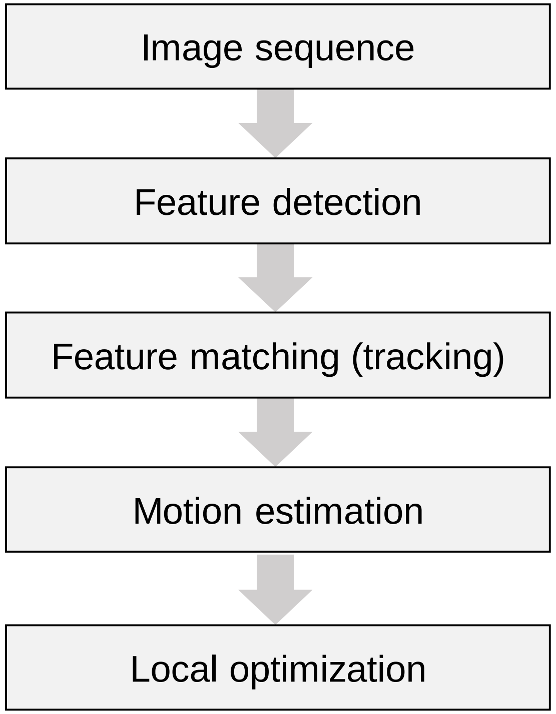
Exercises

- Almost every week starting from next week (check out course schedule)
- **Participation in the exercise sessions is mandatory. Questions about the implementation details might be asked at the exam.**
- **Bring your own laptop**
- Each exercise will consist of coding a building block of a visual odometry pipeline. There will be two exercises dedicated to integrating these blocks together.
- Have **Matlab** pre-installed!
  - **ETH:** Download: https://idesnx.ethz.ch/
  - **UZH:** Download: https://www.zi.uzh.ch/de/students/software-elearning/softwareinstructions/Matlab.html
  - An introductory tutorial on Matlab can be found here: http://rpg.ifi.uzh.ch/docs/teaching/2020/MatlabPrimer.pdf
  - **Please install all the toolboxes included in the license.**

Exercises

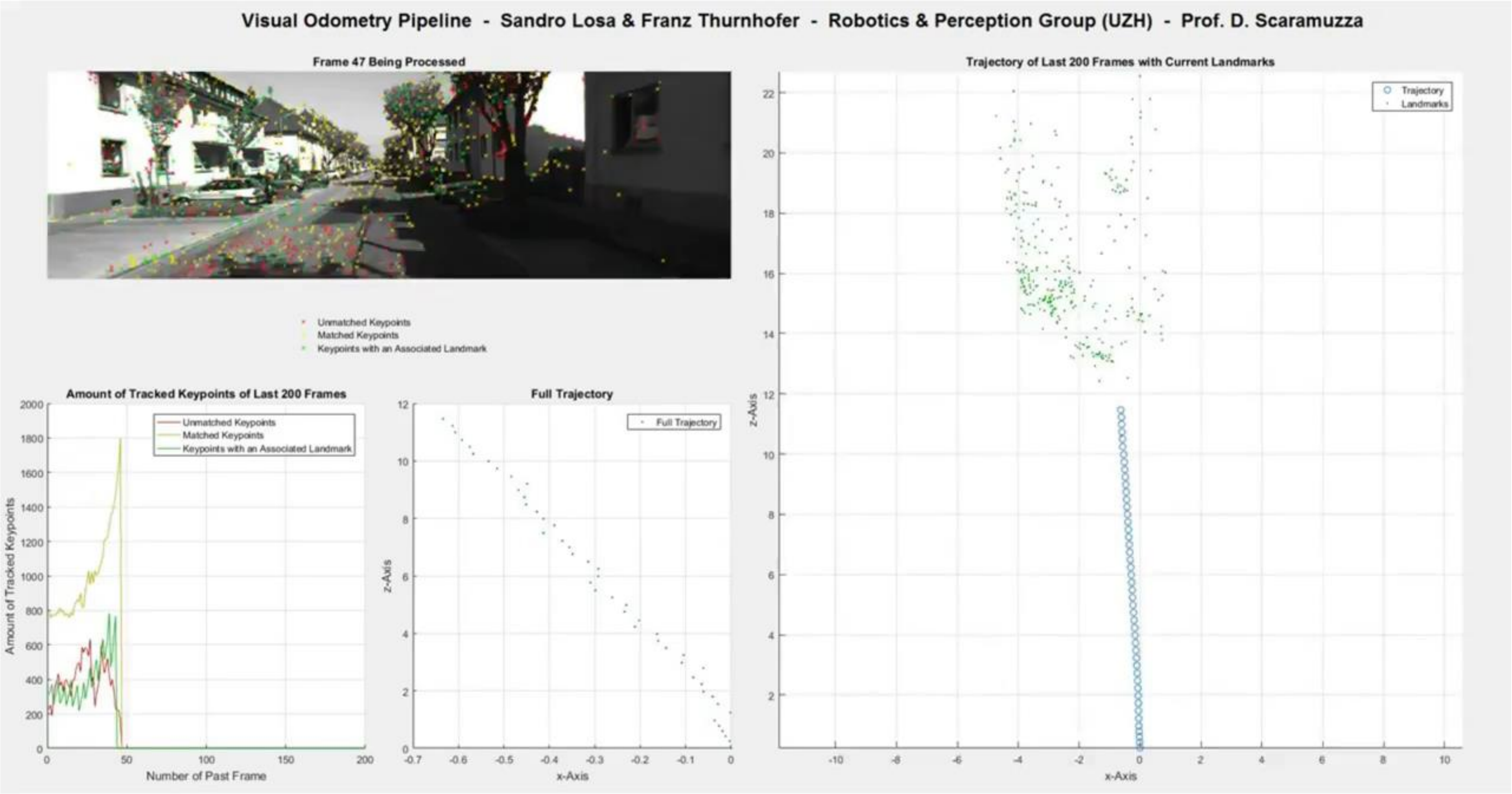
- **Learning Goal** of the exercises: **Implement a full visual odometry pipeline** (similar to that running on Mars rovers but actually much better ☺).

- **Each week** you will learn how to implement a **building block** of visual odometry.



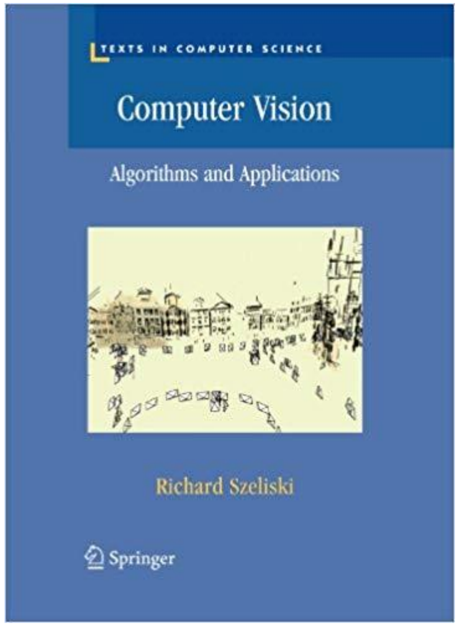
Building blocks of visual odometry along with information flow

Outcome of last year exercises



Reference Textbooks

- **Computer Vision: Algorithms and Applications**, by Richard Szeliski, 2009. PDF freely downloadable from the author webpage: http://szeliski.org/Book/
- **Chapter 4 of Autonomous Mobile Robots,** by R. Siegwart, I.R. Nourbakhsh, D. Scaramuzza. PDF
- Additional readings for curious students will be provided along with the slides and linked directly from the course website
- Alternative books:
  - *Robotics, Vision and Control: Fundamental Algorithms*, by Peter Corke 2011.
  - *An Invitation to 3D Vision*: Y. Ma, S. Soatto, J. Kosecka, S.S. Sastry
  - *Multiple view Geometry*: R. Hartley and A. Zisserman



Prerequisites

- **Linear algebra**
- **Matrix calculus**: matrix multiplication, inversion, singular value decomposition
- **No prior knowledge** of **computer vision** and **image processing** is required

Grading and Exam

- The **final grade is based on an oral exam** (30 minutes). Example exam questions can be found here.
  - Exam dates:
    - **UZH**: January 21-22, 2021
    - **ETH**: January 18 to February 5, 2021

- **Optional mini project:**
  - you have the **option** (i.e., not mandatory) to do a **mini project**, which consists of implementing a working visual odometry algorithm in **Matlab** (C++ or Python are also accepted)
  - If the algorithm runs smoothly, producing a reasonable result, you will be rewarded with an **up to 0.5 grade increase on the final grade**. However, notice that the mini project can be very time consuming!
  - The **deadline** to hand in the mini project is 10.01.2021.
  - **Group work (up to 4) possible.**

- **Strong class participation can offset negative performance** at the oral exam.

Class Participation

- Class participation includes
  - **ask** and **answer questions**
  - being able to **articulate key points from last lecture**

Today’s Outline

- About me and my research lab
- What is Computer Vision?
- Why study computer vision?
- Example of Vision Applications
- Organization of the course
- Start: Visual Odometry overview

What is Visual Odometry (VO) ?

VO is the process of incrementally estimating the pose of the vehicle by examining the changes that motion induces on the images of its onboard cameras

input

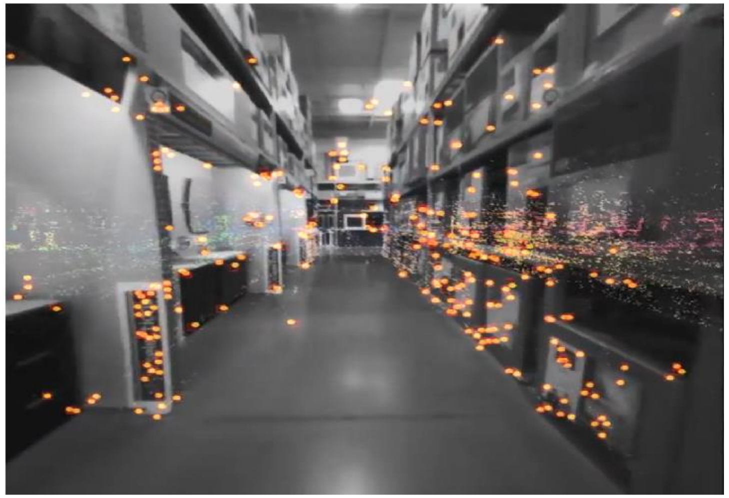
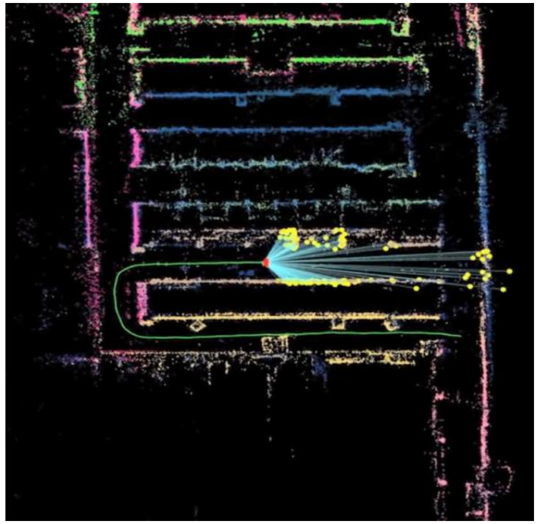


Image sequence (or video stream)
from one or more cameras attached to a moving vehicle

output



$$R_0, R_1, \dots, R_i$$

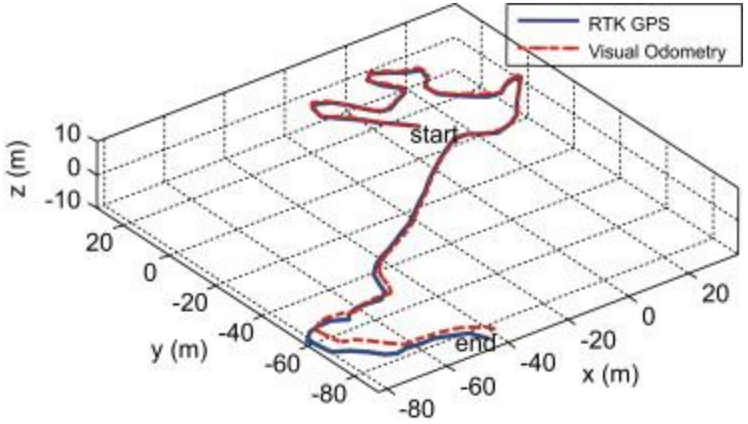
$$t_0, t_1, \dots, t_i$$

Camera trajectory (3D structure is a plus)

Why VO?

- VO is crucial for **flying**, **walking**, and **underwater** robots
- Contrary to wheel odometry, VO is **not affected by wheel slippage** (e.g., on sand or wet floor)
- Very accurate:
  relative position error is 0.1% – 2% of the travelled distance
- VO can be used as a complement to
  - wheel encoders (wheel odometry)
  - GPS (when GPS is degraded)
  - Inertial Measurement Units (IMUs)
  - laser odometry



Assumptions

- **Sufficient illumination** in the environment
- **Dominance of static scene** over moving objects
- **Enough texture** to allow apparent motion to be extracted
- Sufficient **scene overlap** between consecutive frames

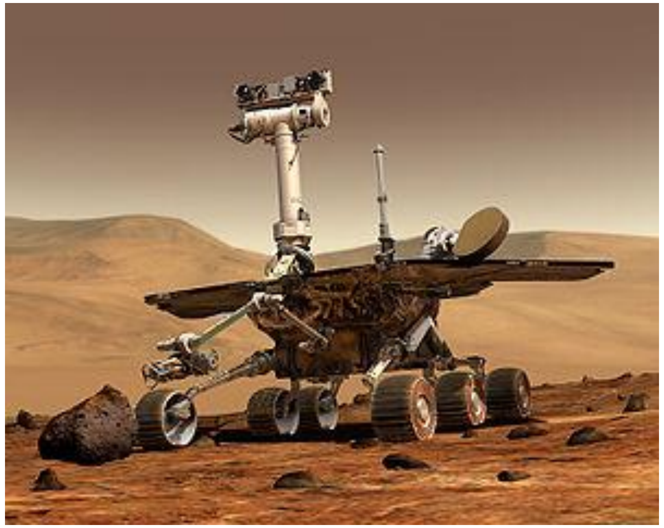
Is any of these scenes good for VO? Why?

A Brief history of VO

- **1980**: First known VO real-time implementation on a robot by **Hans Moraveck** PhD thesis **(NASA/JPL)** for Mars rovers using one sliding camera (*sliding stereo*).

A Brief history of VO

- **1980**: First known VO real-time implementation on a robot by **Hans Moraveck** PhD thesis **(NASA/JPL)** for Mars rovers using one sliding camera (*sliding stereo*).
- **1980 to 2000**: The VO research was dominated by **NASA/JPL** in preparation of the **2004 mission to Mars**
- **2004**: VO was used on a robot on another planet: Mars rovers Spirit and Opportunity (see seminal paper from NASA/JPL, 2007)
- **2004**. VO was revived in the academic environment by **David Nister**'s «Visual Odometry» paper. The term VO became popular.
- **2015-today**: VO becomes a **fundamental tool of several products**: VR/AR, drones, smartphones
- **2021**. VO will be used on the **Mars helicopter**



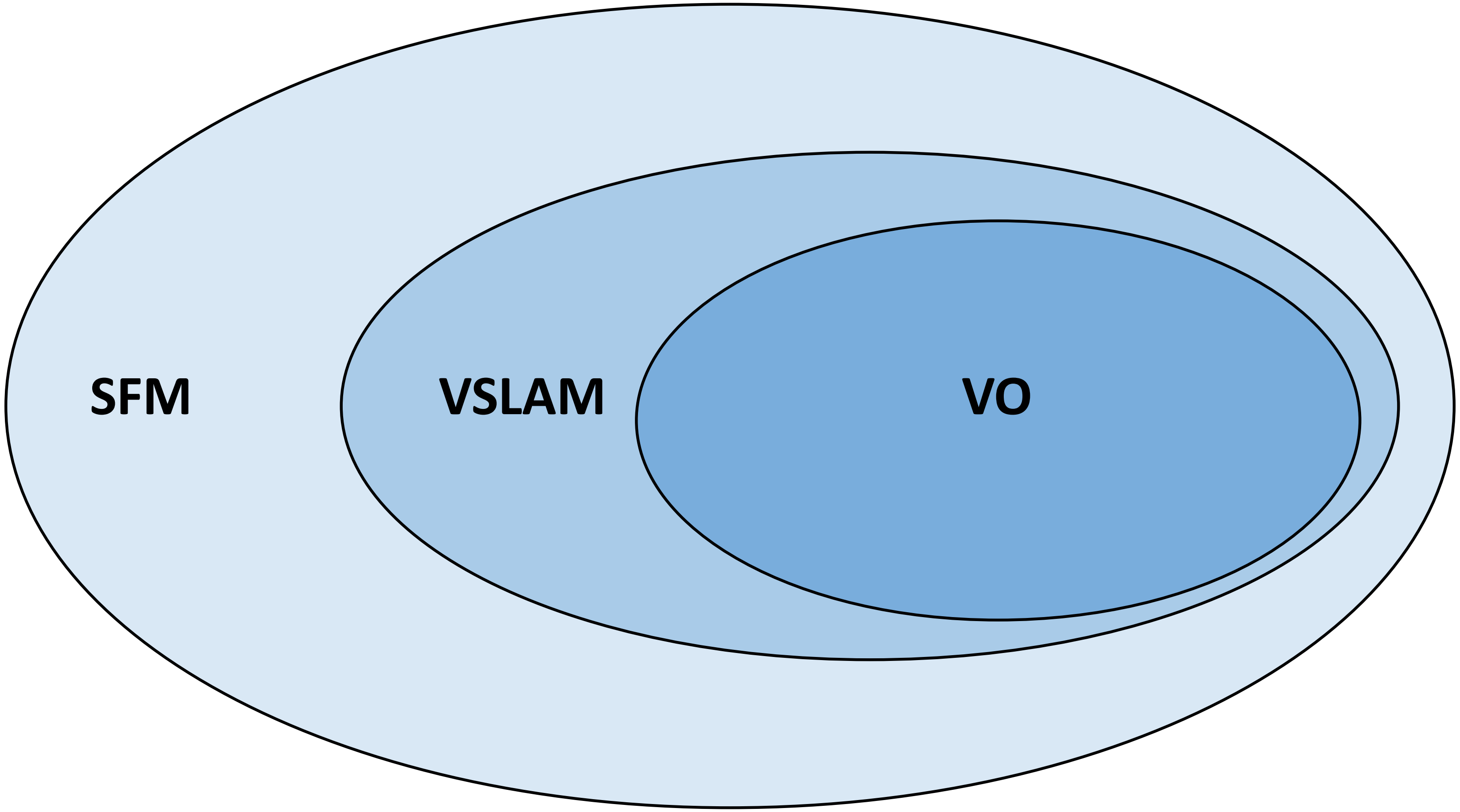
More about history and tutorials

- Scaramuzza, D., Fraundorfer, F., **Visual Odometry: Part I - The First 30 Years and Fundamentals**, *IEEE Robotics and Automation Magazine*, Volume 18, issue 4, 2011. PDF

- Fraundorfer, F., Scaramuzza, D., **Visual Odometry: Part II - Matching, Robustness, and Applications**, *IEEE Robotics and Automation Magazine*, Volume 19, issue 1, 2012. PDF

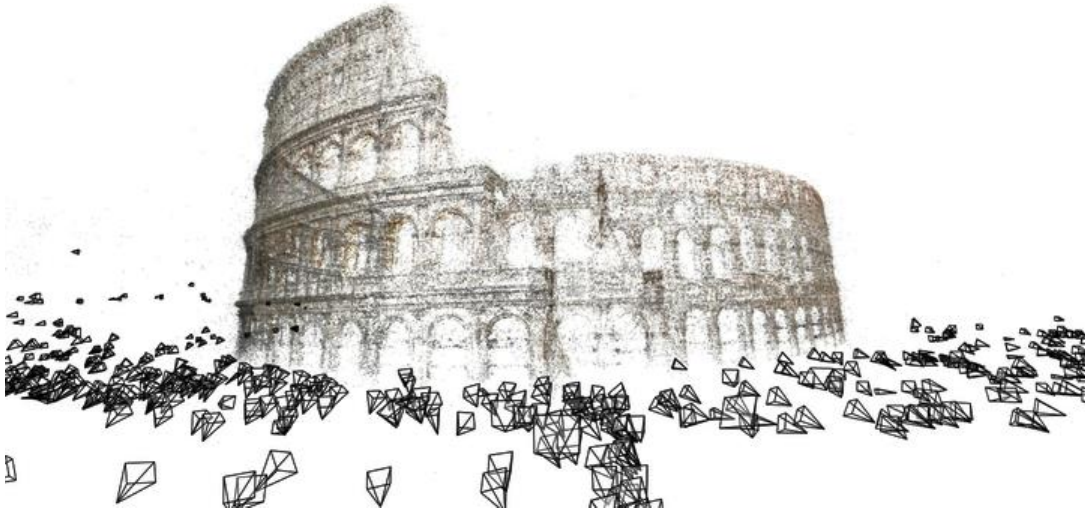
- C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I.D. Reid, J.J. Leonard, **Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age**, *IEEE Transactions on Robotics*, Vol. 32, Issue 6, 2016. PDF

VO vs VSLAM vs SFM



Structure from Motion (SFM)

- SFM is more general than VO and tackles the problem of 3D reconstruction and 6DOF pose estimation from **unordered image sets**



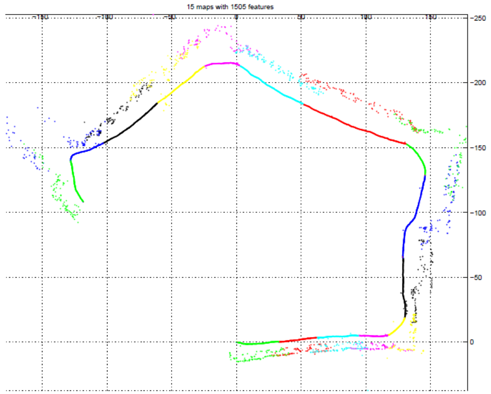
Reconstruction from 3 million images from Flickr.com on a cluster of 250 computers, 24 hours of computation
Paper: "Building Rome in a Day", ICCV'09.
State of the art software: COLMAP

VO vs SFM

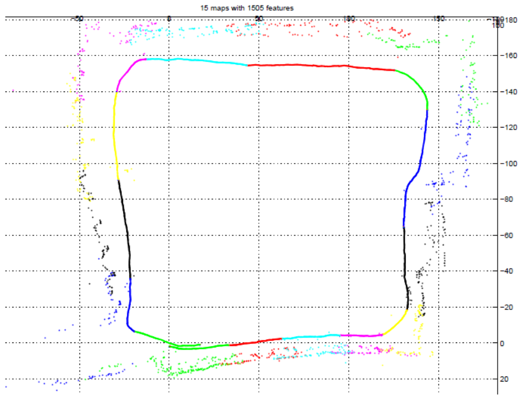
- VO is a **particular case** of SFM
- VO focuses on estimating the 6DoF motion of the camera **sequentially** (as a new frame arrives) and in **real time**
- Terminology: sometimes **SFM** is used as a **synonym** of **VO**

VO vs. Visual SLAM

- **Visual Odometry**
  - Focus on incremental estimation
  - **Guarantees local consistency** (i.e., estimated trajectory is locally correct, but not globally, i.e. from the start to the end)
- **Visual SLAM** (Simultaneous Localization And Mapping)
  - **SLAM = visual odometry + loop detection & closure**
  - **Guarantees global consistency** (the estimated trajectory is globally correct, i.e. from the start to the end)



Visual odometry

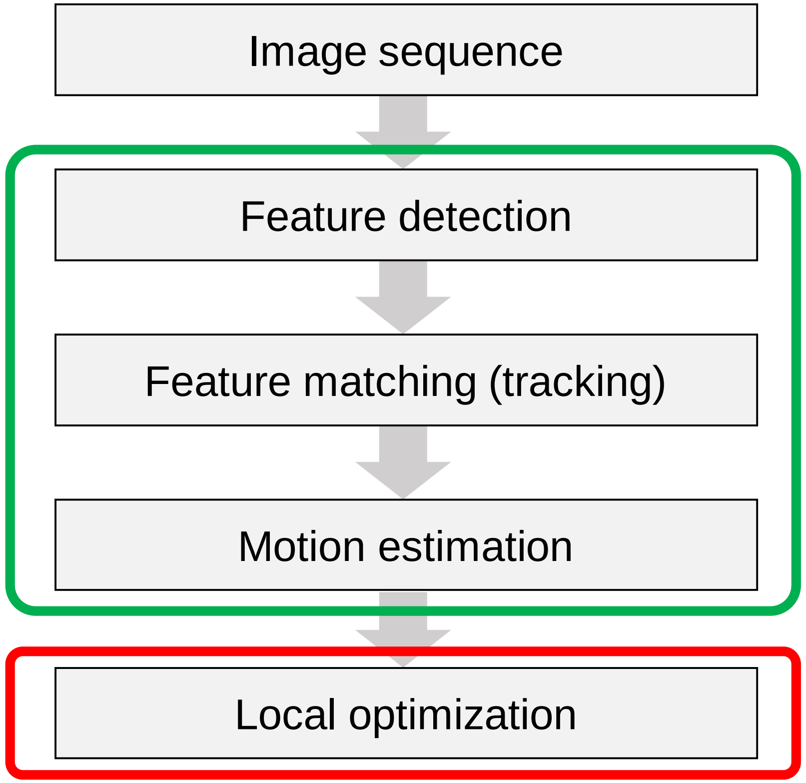


Visual SLAM

Image courtesy of [Clemente et al., RSS’07]

VO Flow Chart

- VO computes the camera path incrementally (pose after pose)

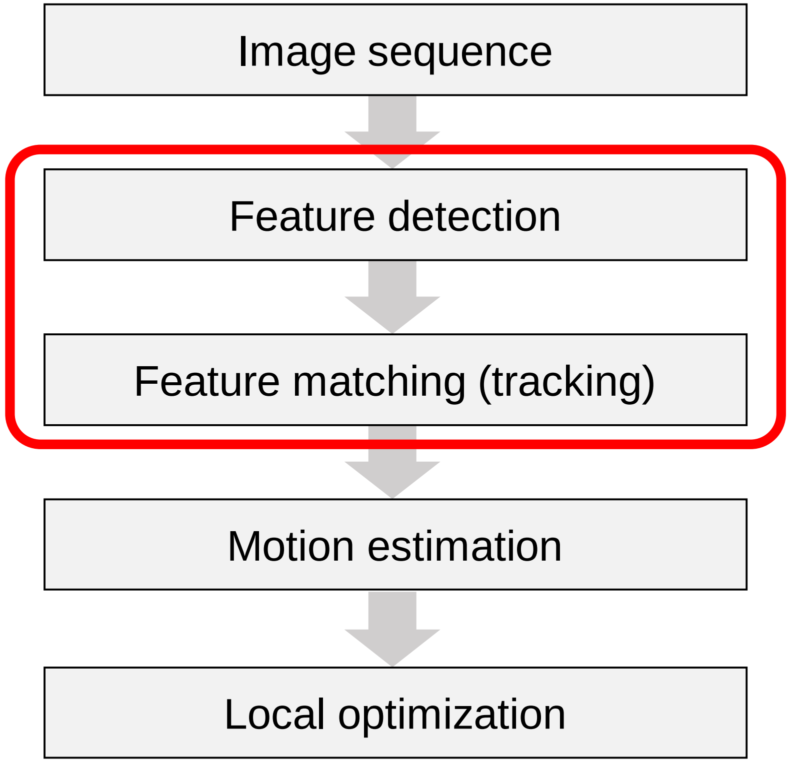


Front-end: outputs the *relative pose* between the *last two frames*

Back-end: “*adjusts*” the relative poses among *multiple recent frames*

VO Flow Chart

- VO computes the camera path incrementally (pose after pose)



Features tracked over multiple recent frames overlaid on the last frame

VO Flow Chart

- VO computes the camera path incrementally (pose after pose)

Image sequence

↓

Feature detection

↓

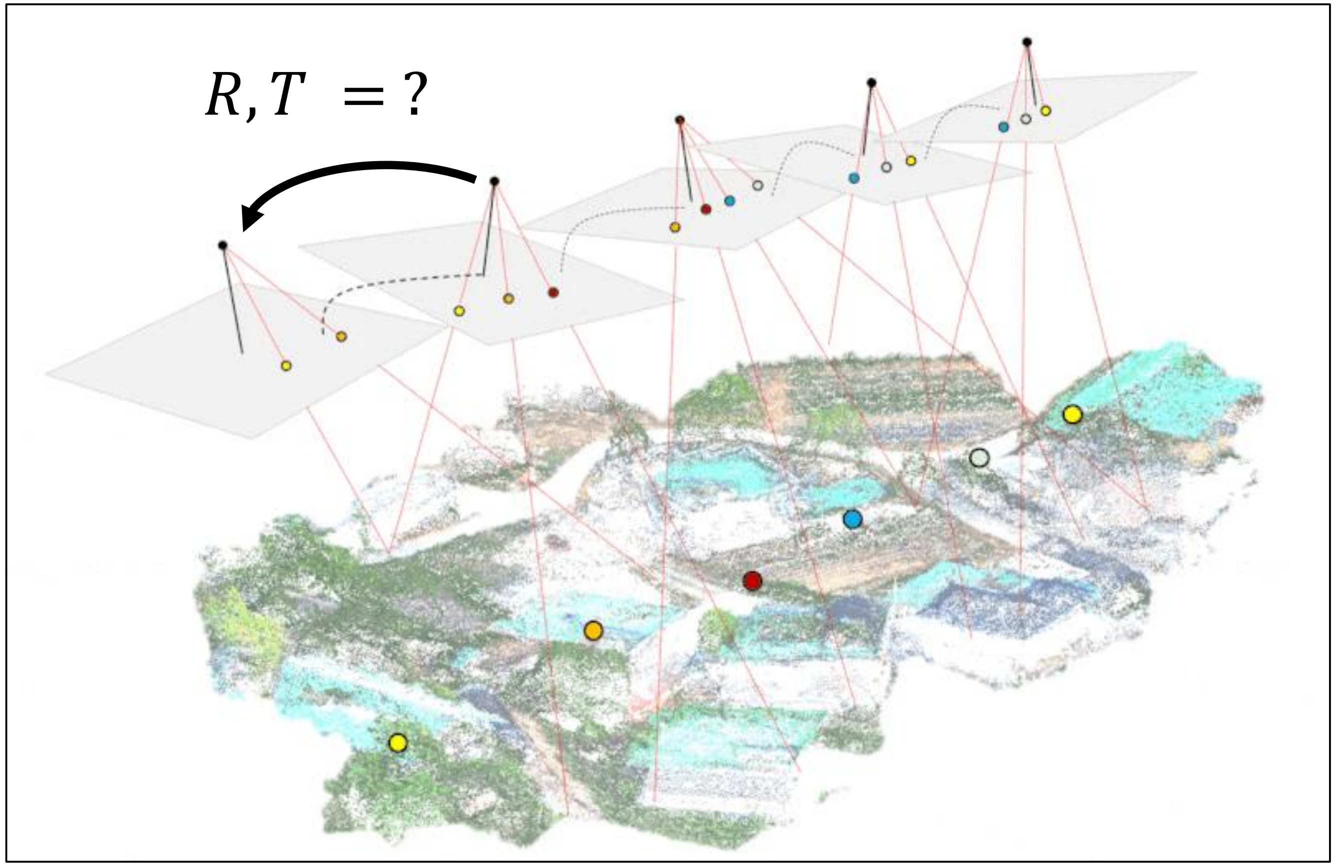
Feature matching (tracking)

↓

Motion estimation

↓

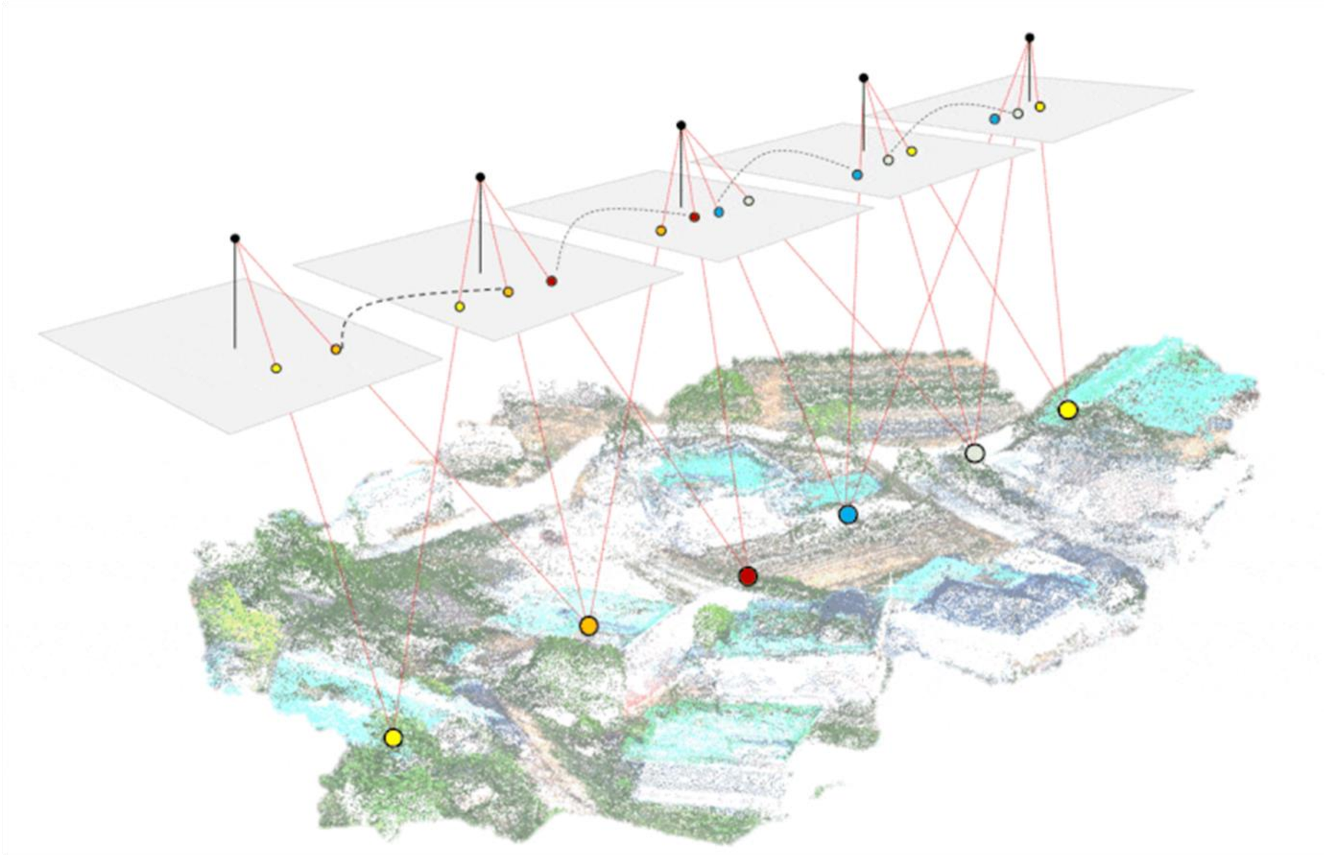
Local optimization



VO Flow Chart

- VO computes the camera path incrementally (pose after pose)

Image sequence
↓
Feature detection
↓
Feature matching (tracking)
↓
Motion estimation
↓
Local optimization

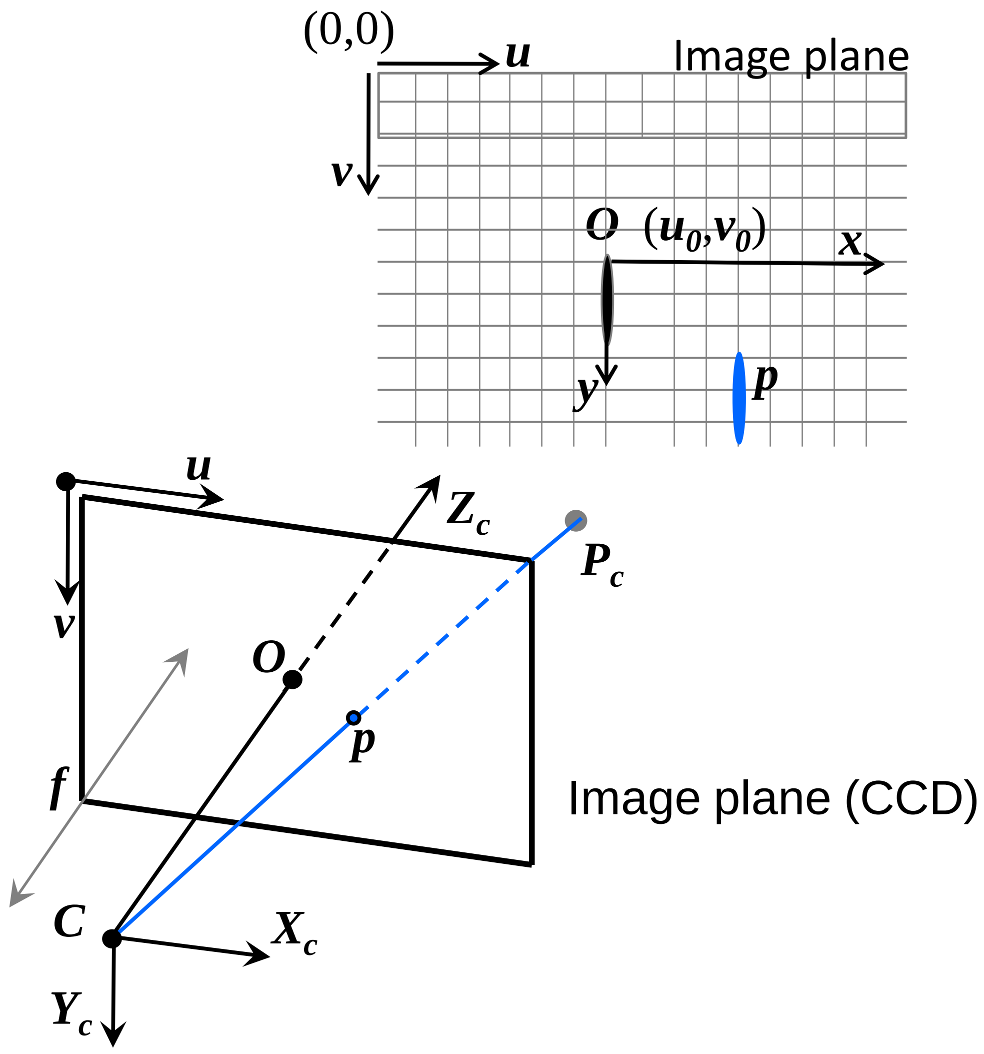


Course Topics

- Principles of image formation
- Image Filtering
- Feature detection and matching
- Multi-view geometry
- Dense reconstruction
- Visual place recognition
- Visual inertial fusion
- Event-based Vision

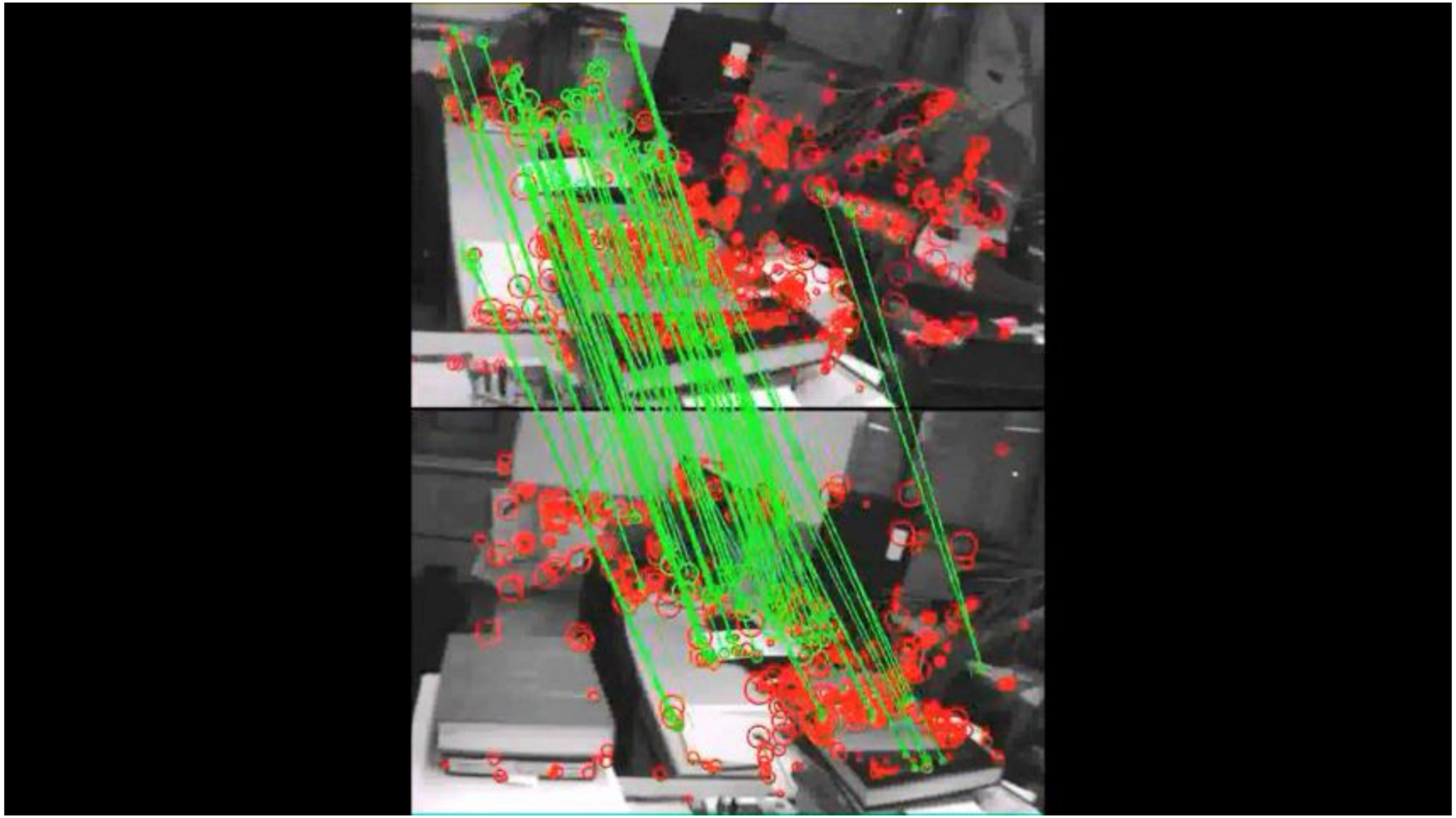
Course Topics

- Principles of image formation
  - Perspective projection
  - Camera calibration



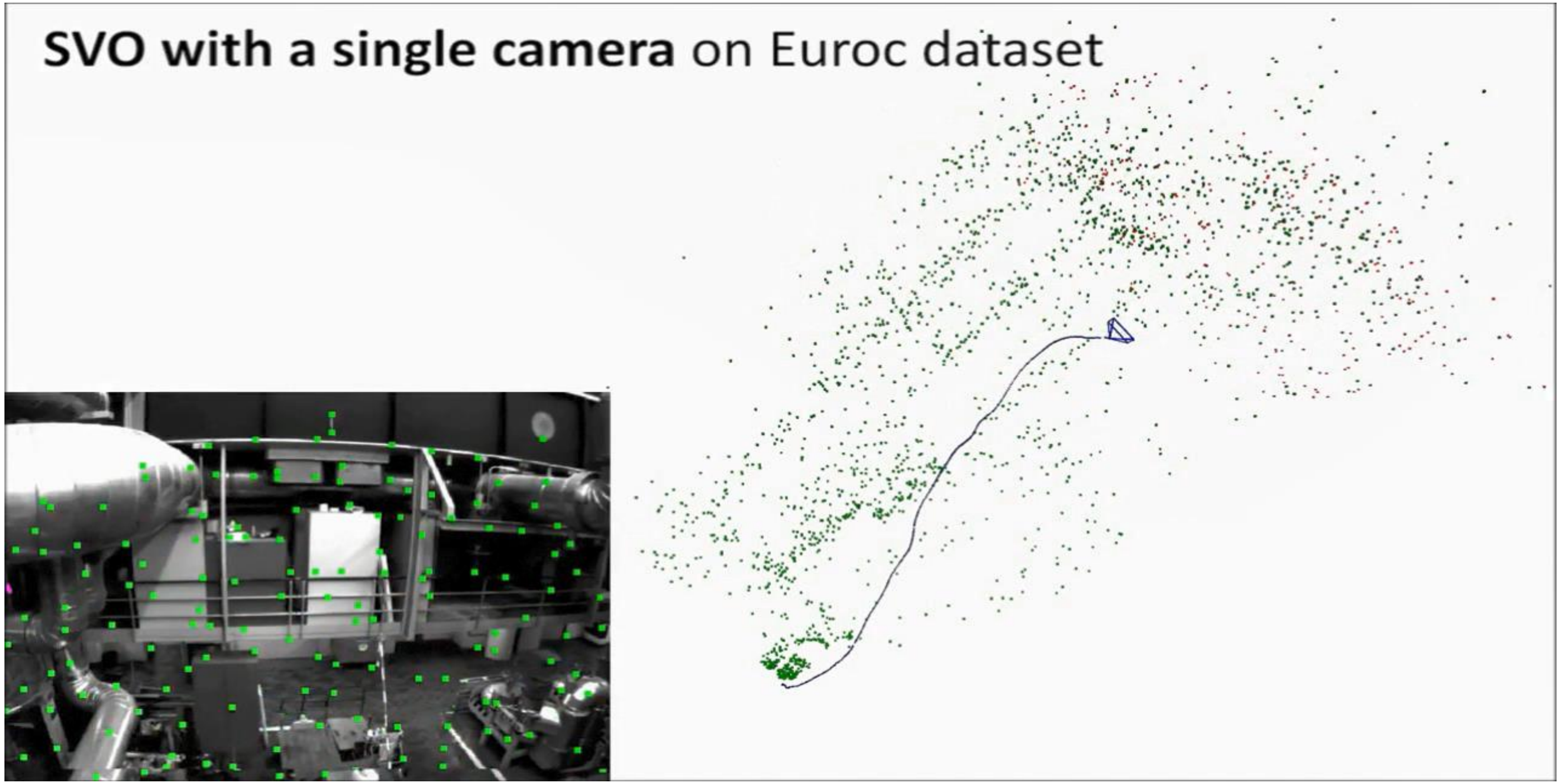
Course Topics

- Feature detection and matching



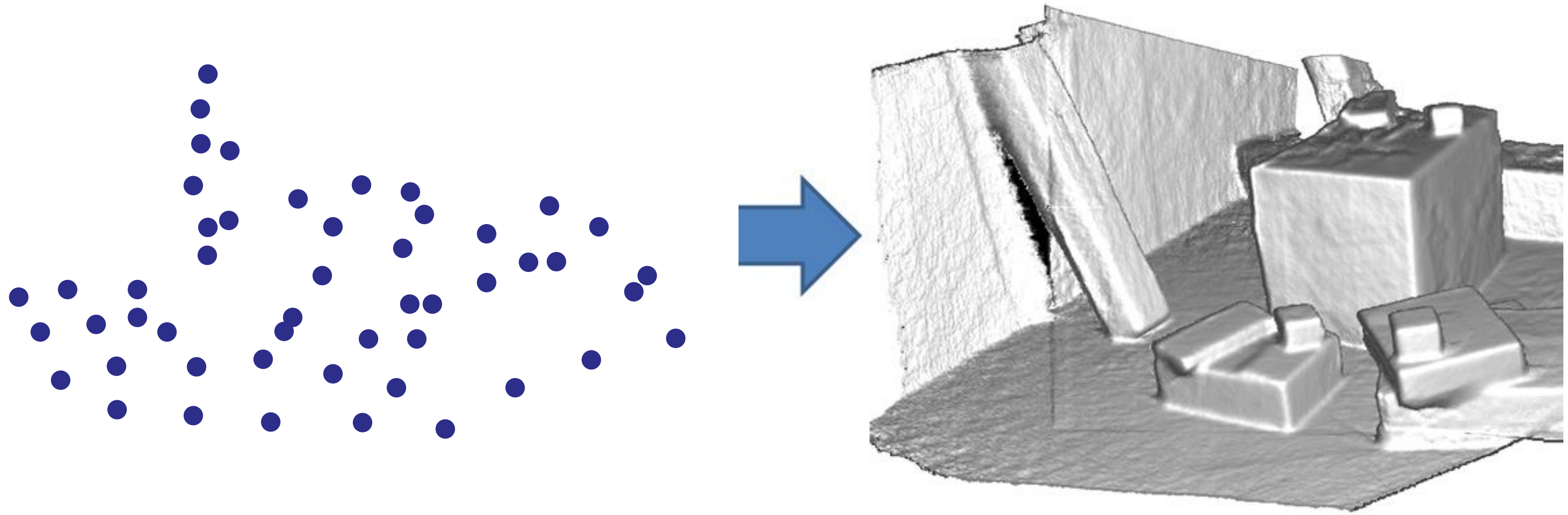
Course Topics

- Multi-view geometry and sparse 3D reconstruction



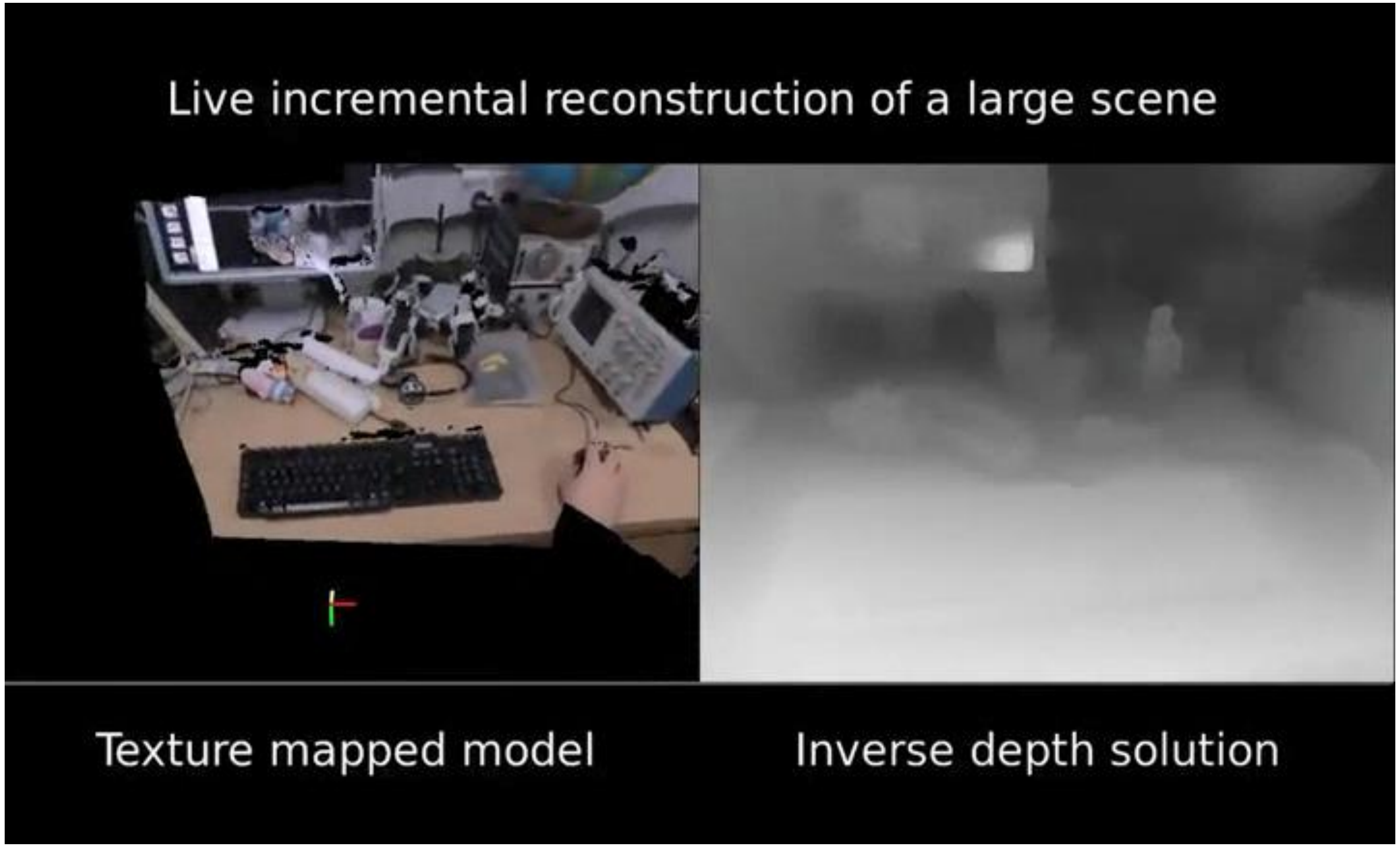
Course Topics

- Dense 3D reconstruction



Course Topics

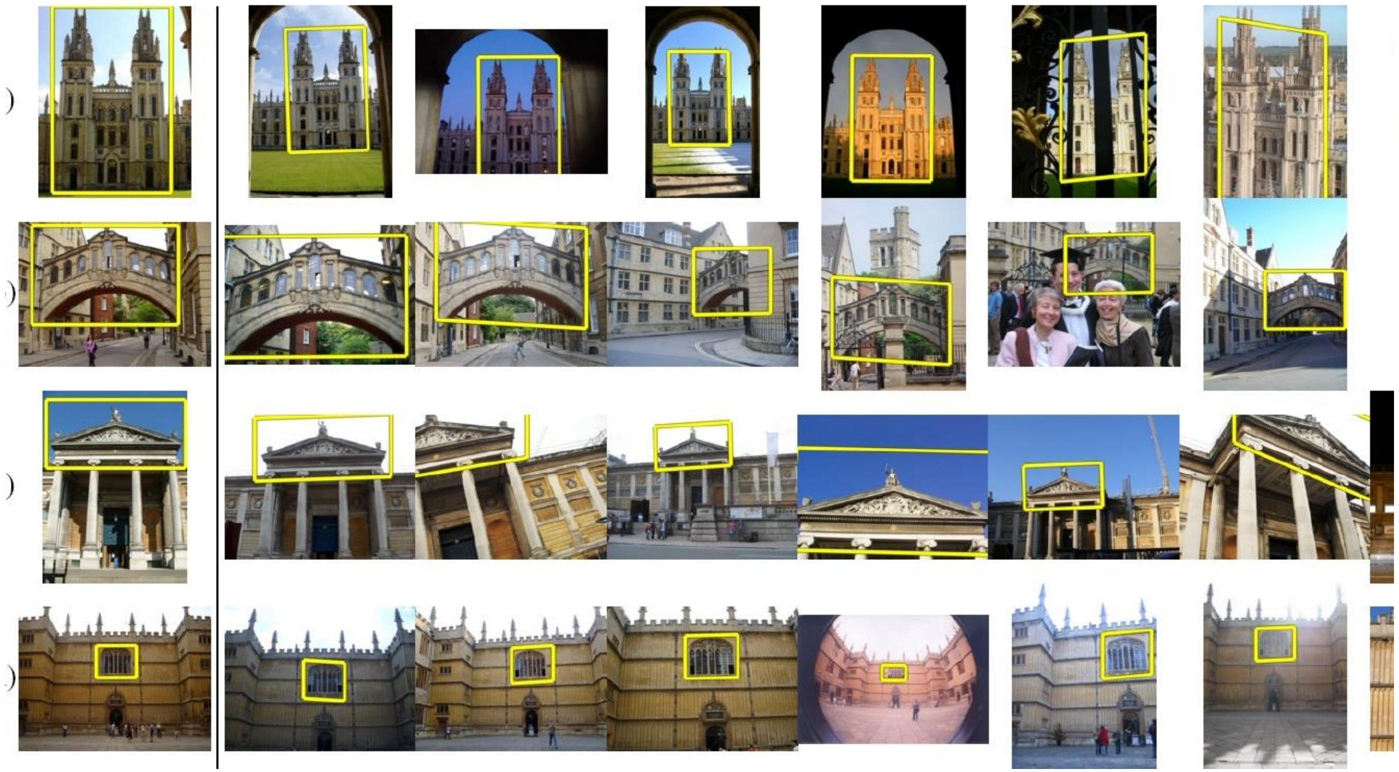
- Dense 3D reconstruction



Course Topics

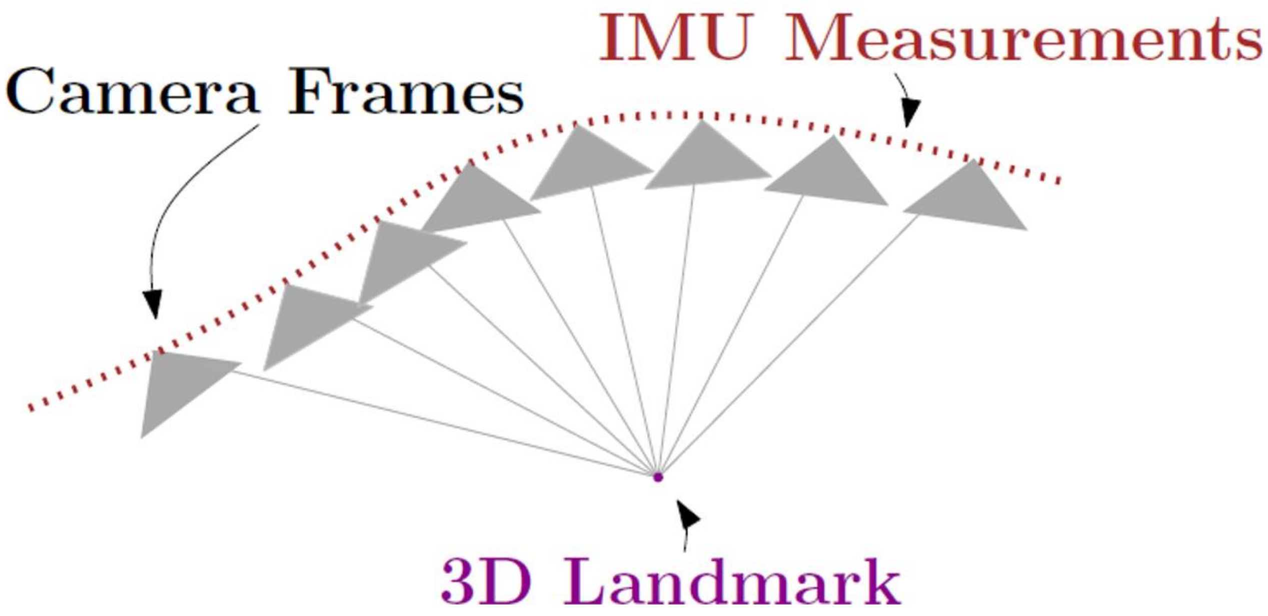
- Place recognition

Query image Most similar places from a database of millions of images



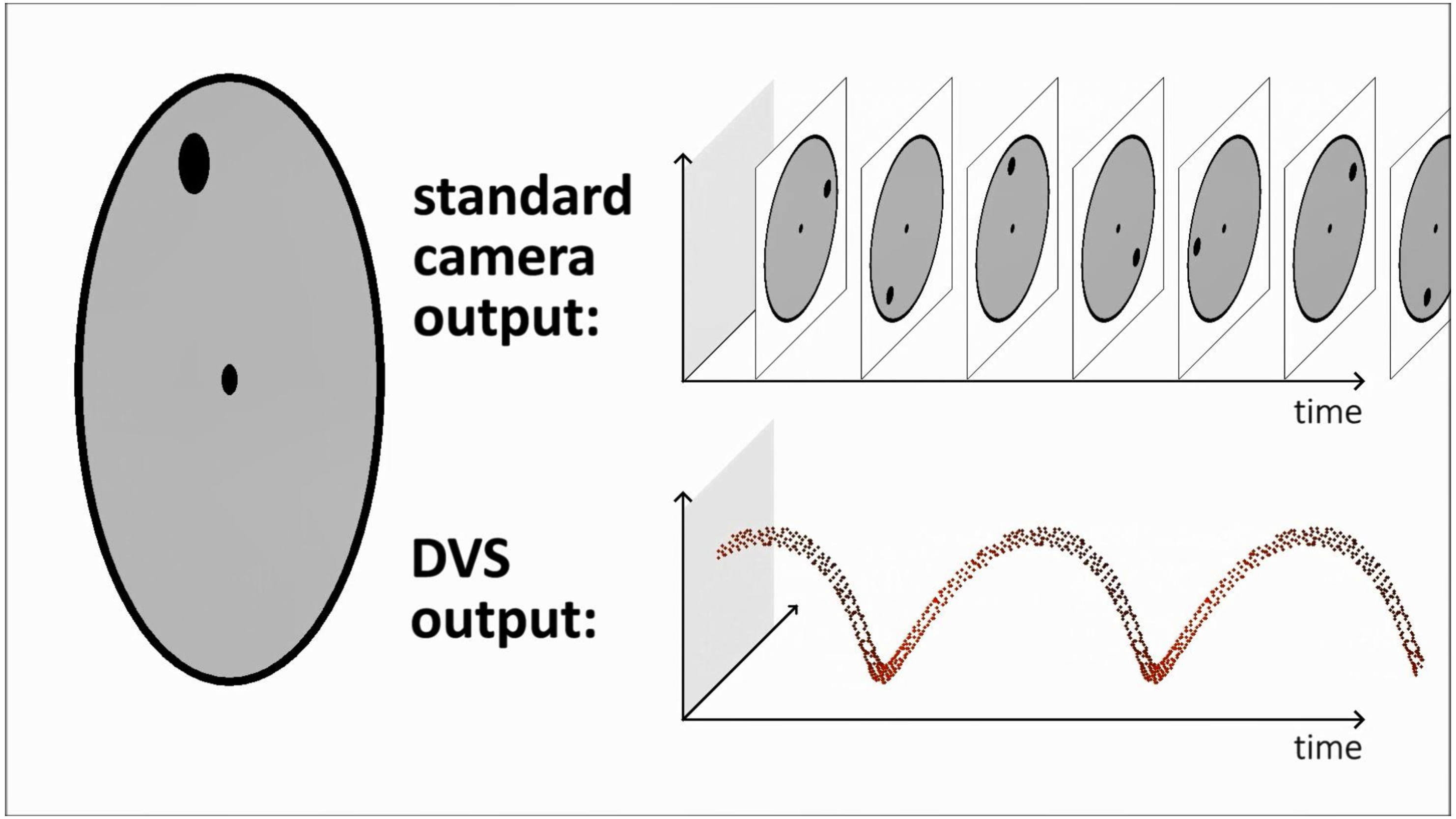
Course Topics

- Visual-inertial fusion

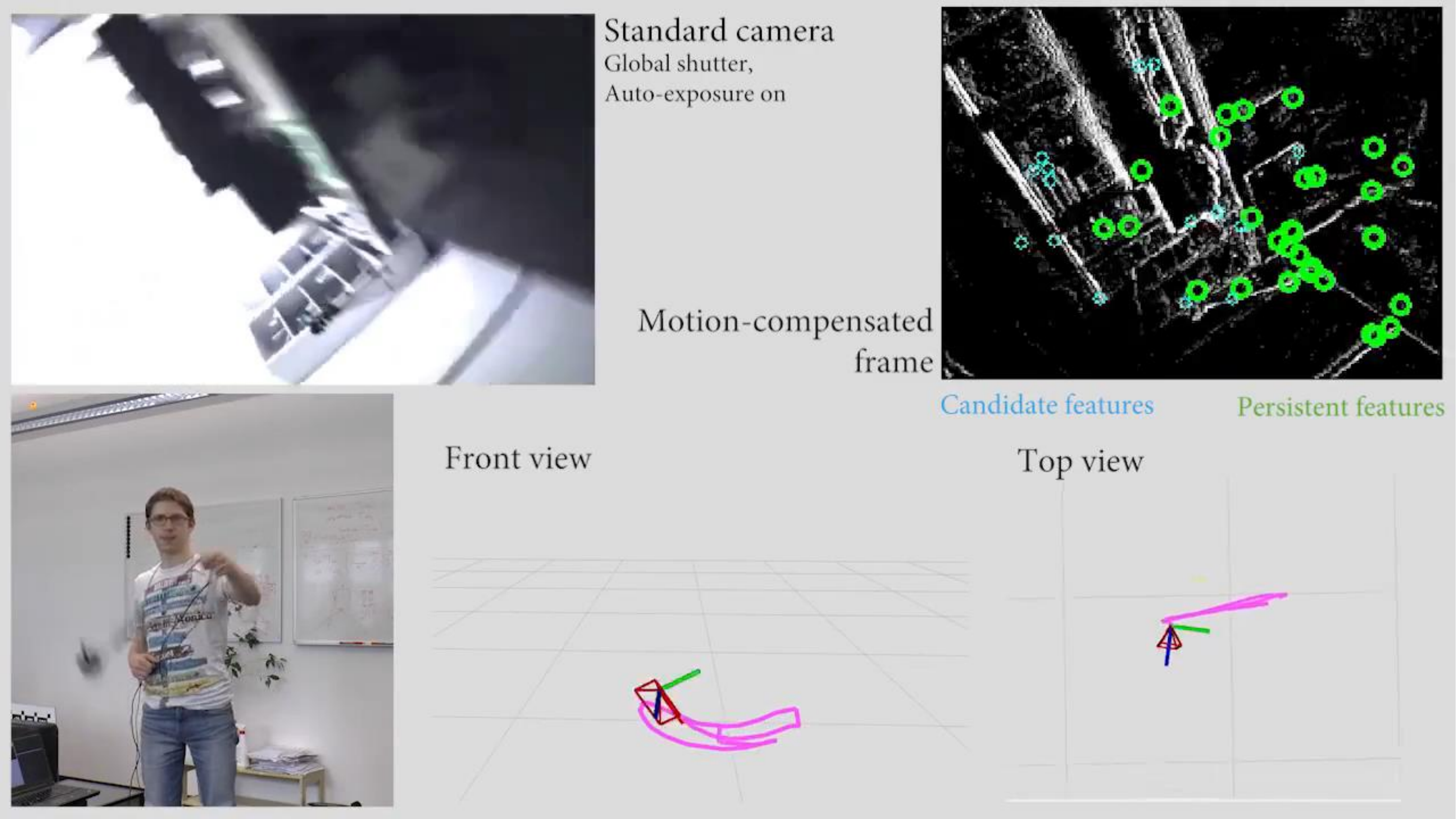


Course Topics

- Event cameras



Application of event cameras: high-speed VO



Rosinol et al., *Ultimate SLAM?* **IEEE RAL'18 best Paper Award Honorable Mention** [PDF](). [Video](). [IEEE Spectrum]().

Understanding Check

Are you able to:

- Provide a definition of Visual Odometry?
- Explain the most important differences between VO, VSLAM, and SFM?
- What assumptions does VO rely on?
- Illustrate the flow chart of VO?