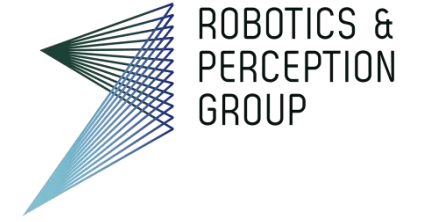




**University of
Zurich** ^{UZH}



Vision Algorithms for Mobile Robotics

Lecture 01
Introduction

Davide Scaramuzza

<https://rpg.ifi.uzh.ch>



In diesem Hause lehrten und forschten

1909–1911	Albert Einstein Nobelpreis für Physik 1921
1911–1912	Peter Debye Nobelpreis für Chemie 1936
1912–1914	Max von Laue Nobelpreis für Physik 1914
1913–1951	Walter Rudolf Hess Nobelpreis für Medizin und Physiologie 1949
1921–1927	Erwin Schrödinger Nobelpreis für Physik 1933

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 since 2012
- Dep. of Informatics (UZH) and Neuroinformatics (UZH & ETH)
- Director of the Master's program in Artificial Intelligence at the Dep. of Informatics (UZH)
- Adjunct Professor of the ETH Master in Robotics, Systems and Control and Associate faculty of the ETH AI Center



Previously

- Distinguished Visiting Scientist at NASA Jet Propulsion Laboratory

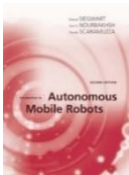
Education

- Master in Electronics Engineering at the University of Perugia, Italy, 2004
- PhD in Robotics and Computer Vision at ETH Zurich, Switzerland, 2008
- Post-doc at the University of Pennsylvania, USA
- Visiting professor at Stanford University, 2019



Book

- "Autonomous Mobile Robots, 2nd Edition" MIT Press, 2011

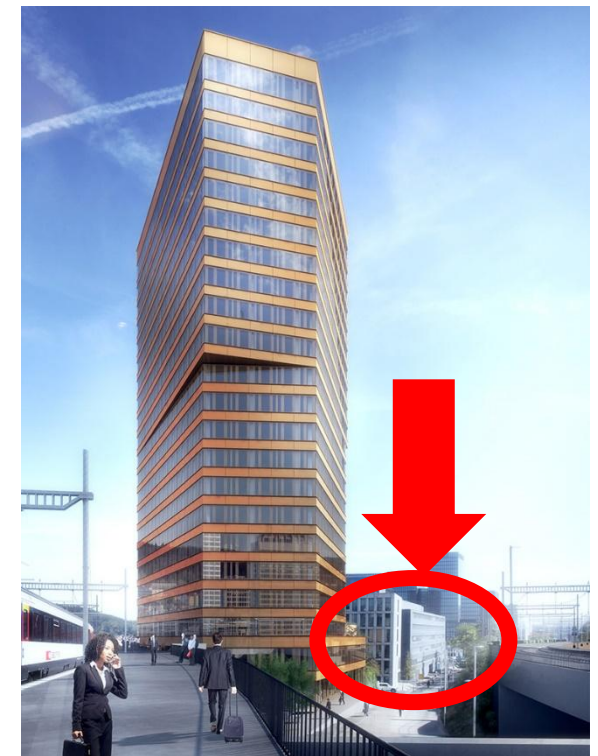
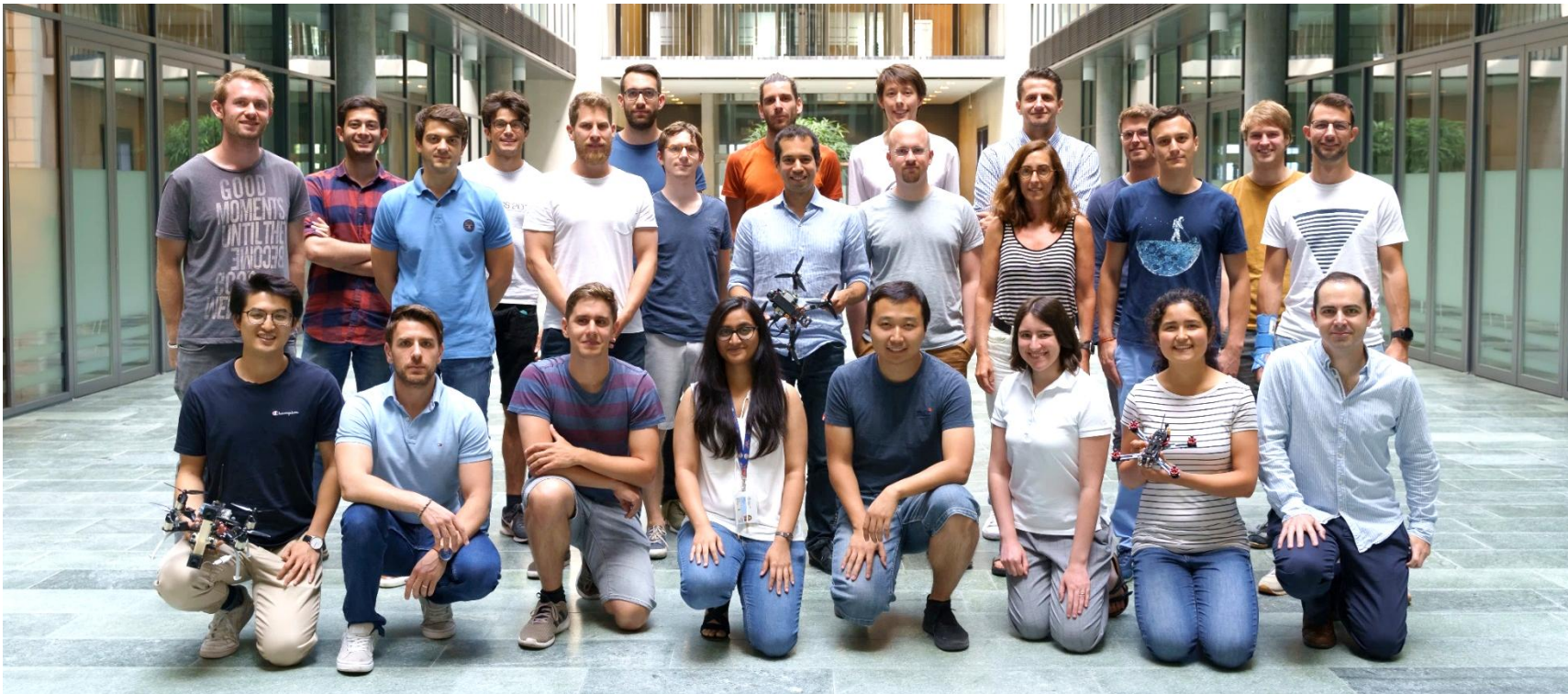


Hobbies

- Running, piano, magic tricks

My lab: the Robotics and Perception Group

- **Address:** Andreasstrasse 15, 2nd floor, next to **Zurich Oerlikon** train station
- **Webpage:** <https://rpg.ifi.uzh.ch/>
- **Lab Open Door Day:** December 11, at 2pm



Our Research: Autonomous, Vision-based Navigation

Real-time, **perception, learning, and control** for **autonomous agile robotics**
using both **standard cameras** or **neuromorphic event-based cameras**



Spatial AI: Visual Simultaneous Localization and Mapping



Agile Robotics



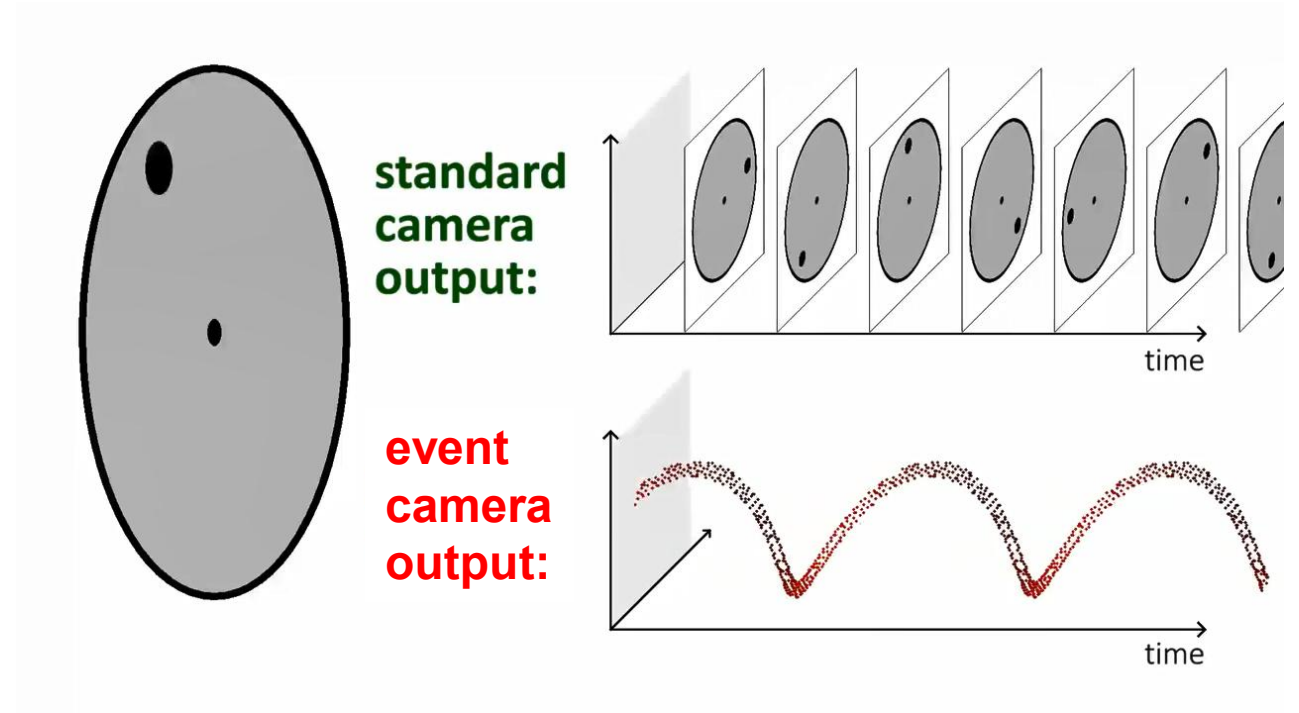
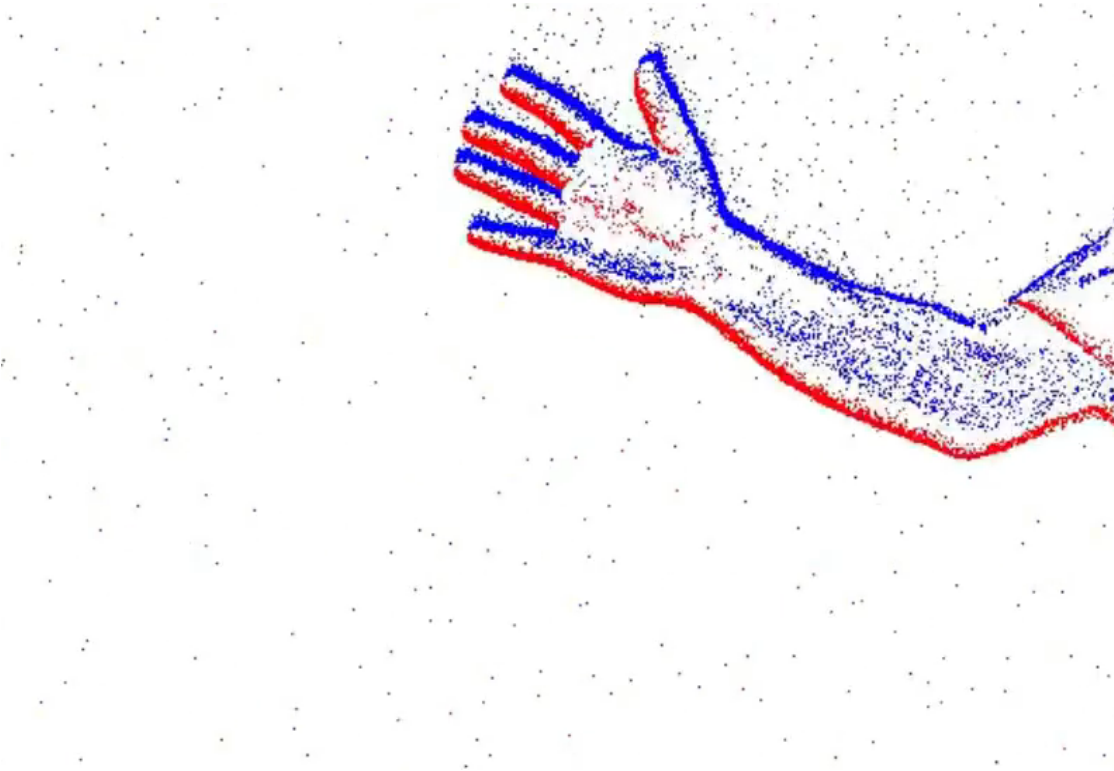
Kaufmann, Bauersfeld, Loquercio, Mueller, Koltun, Scaramuzza,
Champion-Level Drone Racing using Deep Reinforcement Learning, Nature, 2023

Reinforcement Learning



Song, Romero, Mueller, Koltun, Scaramuzza, *Reaching the Limit in Autonomous Racing: Optimal Control versus Reinforcement Learning*, **Science Robotics**, 2023

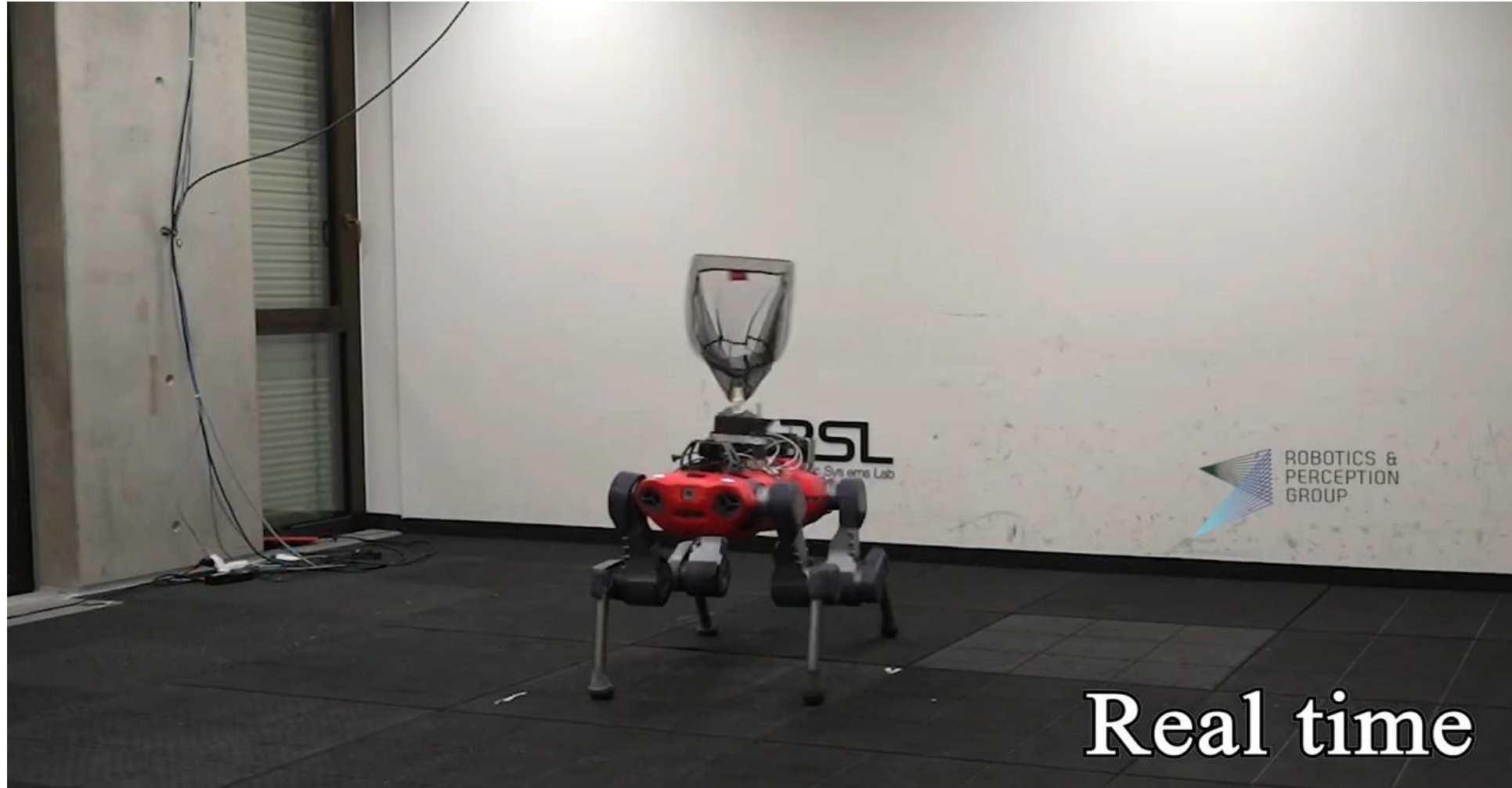
Event Cameras



[1] Lichtsteiner, Posch, Delbruck, *A 128x128 120 dB 15 μ s Latency Asynchronous Temporal Contrast Vision Sensor*, IEEE Journal of Solid-State Circuits, 2008

[2] Gallego et al., *Event-based Vision: A Survey*, T-PAMI, 2020

Low-Latency Perception and Control

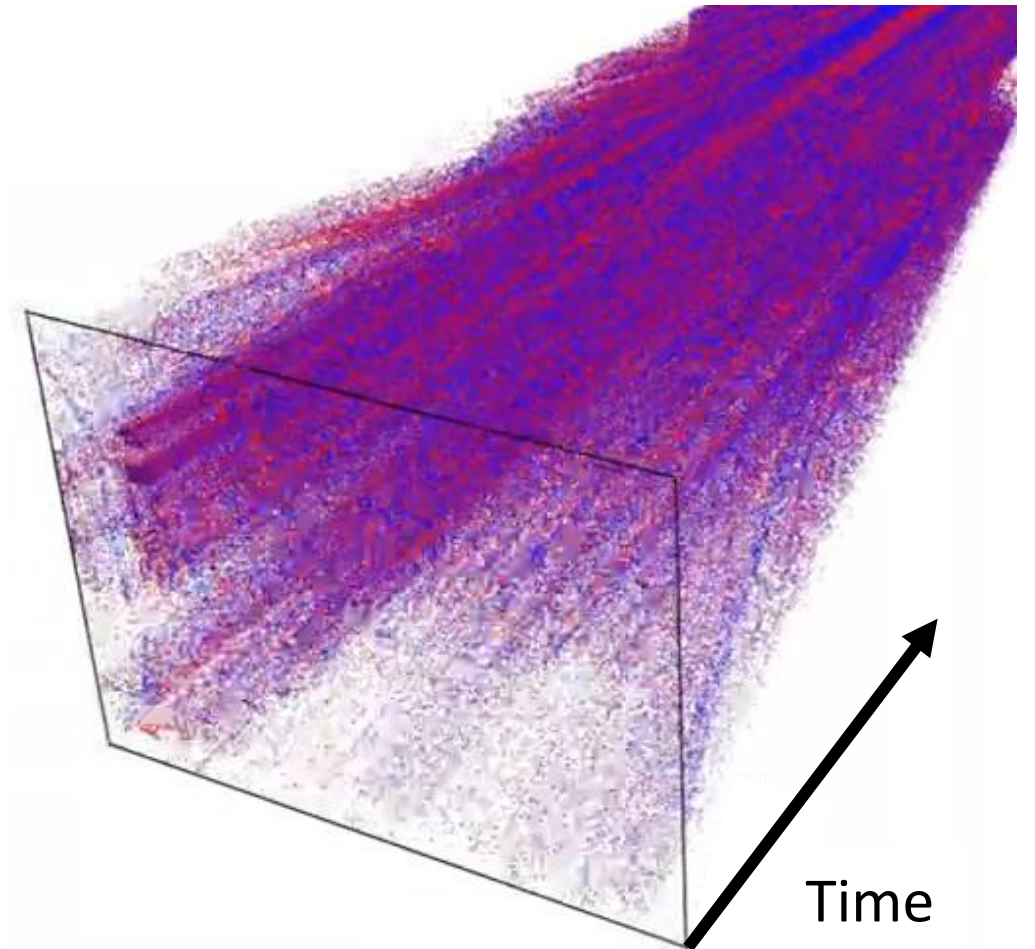


Low-Latency Perception for Autonomous Driving

Standard camera



Event camera

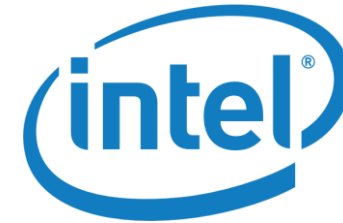


Low-Latency Perception for Autonomous Driving

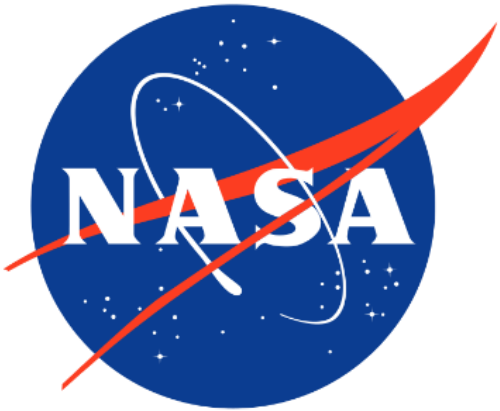


Gehrig, Scaramuzza, *Low Latency Automotive Vision with Event Cameras*, **Nature**, 2024

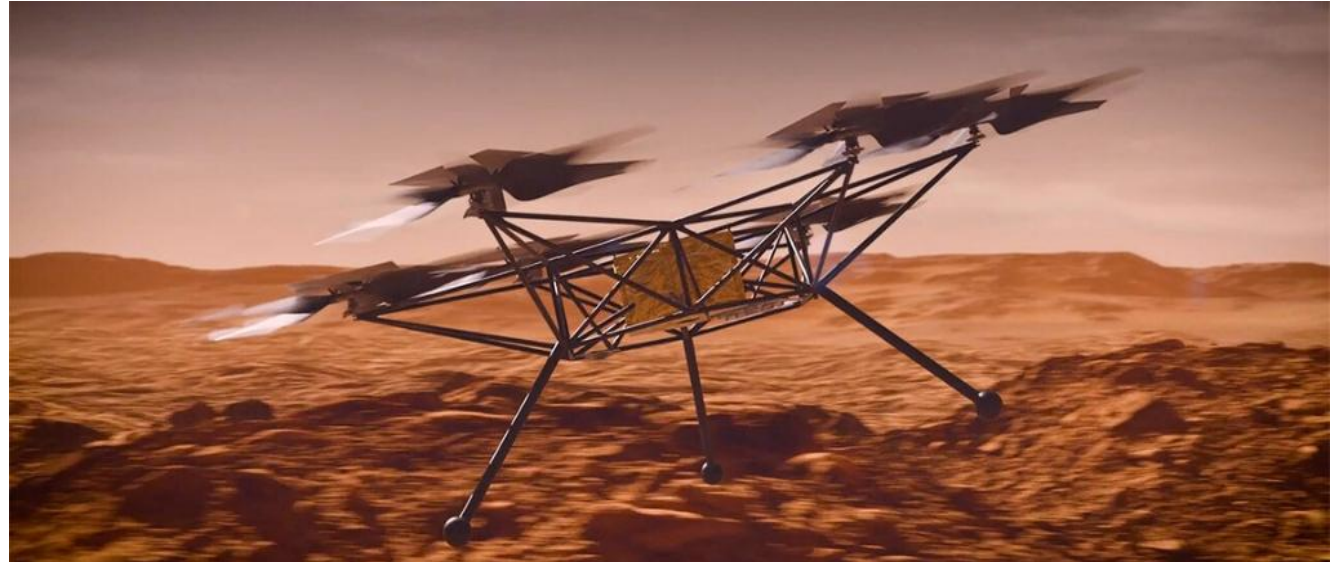
Tech Transfer and Spin-offs



Collaboration with NASA/JPL for Chopper mission



Jet Propulsion Laboratory
California Institute of Technology



1. Alberico, Delaune, Cioffi, Scaramuzza, Structure-Invariant Range-Visual-Inertial Odometry, IROS'24, [PDF](#), [Video](#)
2. Polizzi, Hewitt, Hidalgo-Carrió, Delaune, Scaramuzza, Data-Efficient Collaborative Decentralized Thermal-Inertial Odometry, RAL'22, [PDF](#), [Video](#)
3. Mahlkecht, Gehrig, Nash, Rockenbauer, Morrell, Delaune, Scaramuzza, Exploring Event Camera-based Odometry for Planetary Robots, RAL'22, [PDF](#), [Video](#)
4. More in [this Swissinfo article](#)

SUIND Autonomous Crop Spraying and Forest monitoring



Fotokite A tethered drone for first response

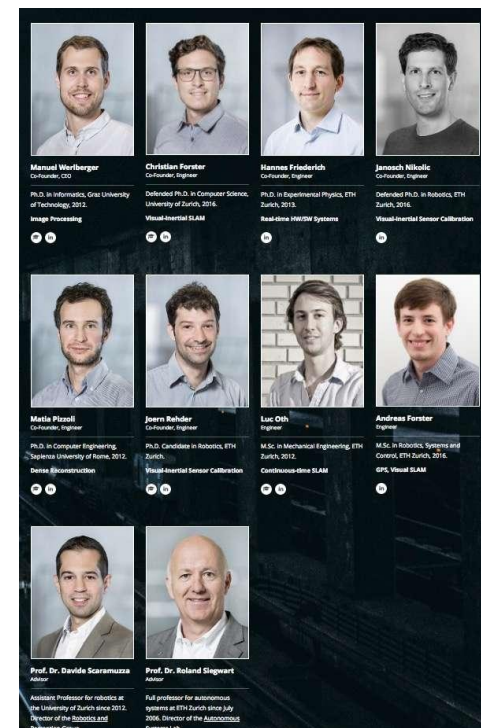
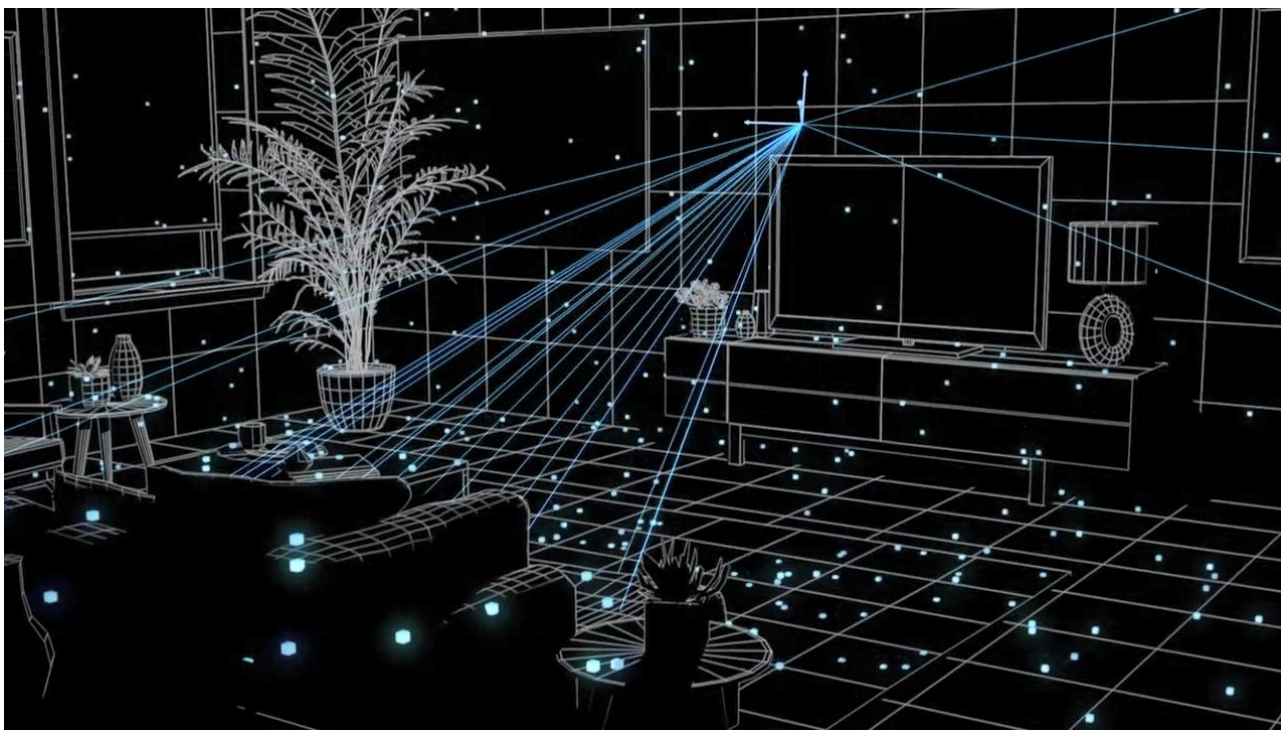
The drone receives electrical power over-tether from the ground so that it can fly “forever”





“Zurich-Eye” – Today: Meta Zurich

- **Vision-based Localization and Mapping** systems for mobile robots
- Born in Sep. 2015, became **Facebook-Oculus Zurich** in Sep. 2016. Today, over **200 employees**.
- In 2018, Zurich-Eye launched **Oculus Quest** (25 million units sold so far)



“From the lab to the living room”: The story behind Facebook’s Oculus Insight technology from Zurich-Eye to Oculus Quest:

<https://tech.fb.com/the-story-behind-oculus-insight-technology/>

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

- **Topics:** machine learning, computer vision, event cameras, foundation models, reinforcement learning, control, planning, system integration
- **Highlights:** many of our Master students have published their thesis in international conferences, won prestigious awards (e.g., ETH Medal, Fritz-Kutter Award, conference paper awards), were admitted for a PhD at top robotics labs (UZH, ETH, UPenn, CMU, MIT, etc.) or landed jobs at top-tier companies (NASA/JPL, CERN, SONY, Meta, NVIDIA, Microsoft, Skydio, etc.)

**University of Zurich**

Department of Informatics - Institute of Neuroinformatics - Robotics and Perception Group

News

People

Research

Publications

Software/Datasets

Open Positions

Student Projects

Teaching

Media Coverage

Awards

Gallery

Contact

Student Projects

How to apply

To easily browse and send your CV, your life and the transcript by email to all the contacts indicated below the project description. Do not apply on SROP! Since Prof. Davide Scaramuzza is affiliated with ETH, there is no organizational overhead for ETH students. Custom projects are occasionally available. If you would like to do a project with us but could not find an advertised project that suits you, please contact Prof. Davide Scaramuzza directly to ask for a **tailored project** (please at ifi.uzh.ch).

Upon successful completion of a project in our lab, students may also have the opportunity to get an **internship at one of our numerous industrial and academic partners worldwide** (e.g., NASA/JPL, University of Pennsylvania, UCLA, MIT, Stanford, ...).

Electrical Flow-Based Graph Embeddings for Event-based Vision and other downstream tasks - Available



Description: Besides RPG, this project will be co-supervised by Simon Meierhans (from Alg. & Opt. group at ETH) and prof. Siddhartha Mishra. This project explores a novel approach to graph embeddings using electrical flow computations. By leveraging the efficiency of solving systems of linear equations and some properties of electrical flows, we aim to develop a new method for creating low-dimensional representations of graphs. These embeddings have the potential to capture unique structural and dynamic properties of networks. The project will investigate how these electrical flow-based embeddings can be utilized in various downstream tasks such as node classification, link prediction, graph classification and event-based vision tasks.

Goal: The primary goal of this project is to design, implement, and evaluate a graph embedding technique based on electrical flow computations. The student will develop algorithms to compute these embeddings efficiently, compare them with existing graph embedding methods, and apply them to real-world network datasets. The project will also explore the effectiveness of these embeddings in downstream machine learning tasks. Applicants should have a strong background in graph theory, linear algebra, and machine learning, as well as proficiency in Python and ideally experience with graph processing libraries like NetworkX or graph-tool.

Contact Details: Nikita Zubic (zubic@ifi.uzh.ch), Simon Meierhans (simon.meierhans@inf.ethz.ch)

Thesis Type: Master Thesis

[See project on SROP](#)

What can Large Language Models offer to Event-based Vision? - Available



Description: Event-based vision algorithms process visual changes in an asynchronous manner akin to how biological visual systems function, while large language models (LLMs) specialize in parsing and generating human-like text. This project aims to explore the intersection of Large Language Models (LLMs) and Event-based Vision, leveraging the unique capabilities of each domain to create a symbiotic framework. By marrying the strengths of both technologies, the initiative aims to develop a novel, more robust paradigm that excels in challenging conditions.

Goal: The primary objective is to devise methodologies that synergize the capabilities of LLMs with Event-Based Vision systems. We intend to address identified shortcomings in existing paradigms by leveraging the inferential strengths of LLMs. Rigorous evaluations will be conducted to validate the efficacy of the integrated system under various challenging conditions.

Contact Details: Nikita Zubic (zubic@ifi.uzh.ch), Nico Messikommer (nmes@ifi.uzh.ch)

Thesis Type: Semester Project / Master Thesis

[See project on SROP](#)

Hybrid Spiking-Deep Neural Network System for Efficient Event-Based Vision Processing - Available



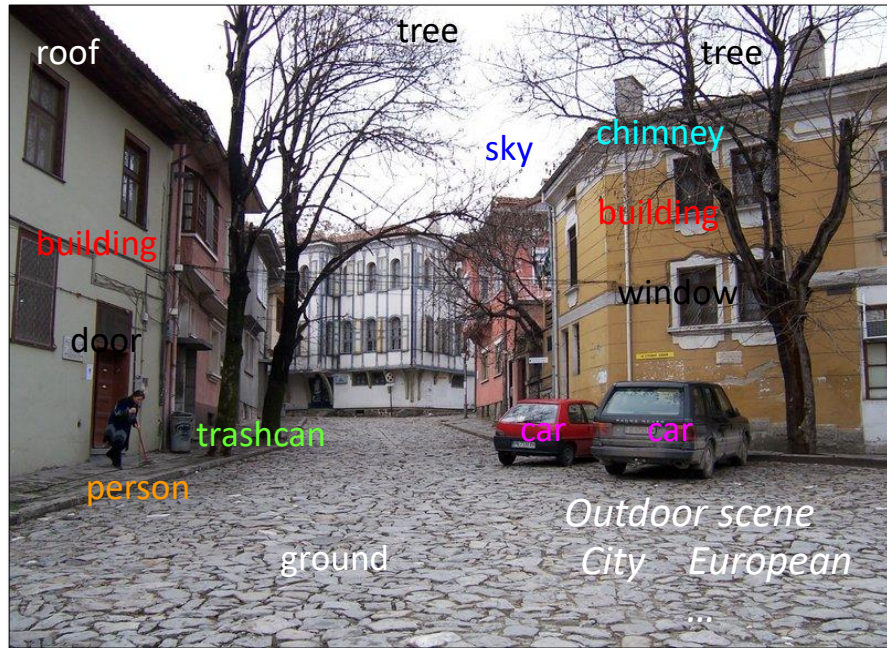
Description: Event cameras are innovative sensors that capture changes in a scene dynamically, unlike standard cameras that capture images at fixed intervals. They detect pixel-level brightness changes, providing high temporal resolution and low latency. This results in efficient data processing and reduced power consumption, typically just 1 mW. Spiking Neural Networks (SNNs) process information as discrete events or spikes, mimicking the brain's neural activity. They differ from standard Neural Networks (NNs) that process information continuously. SNNs are highly efficient in power consumption and well-suited for event-driven data from event cameras. In collaboration with SynSense, this project aims to integrate the rapid processing capabilities of SNNs with the advanced analytic powers of deep neural networks. By distilling higher-level features from raw event data, we aim to significantly reduce the volume of events needing further processing by traditional NNs, improving data quality and transmission efficiency. System will be tested on computer vision tasks like object detection and tracking, gesture recognition, and high-speed motion estimation.

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

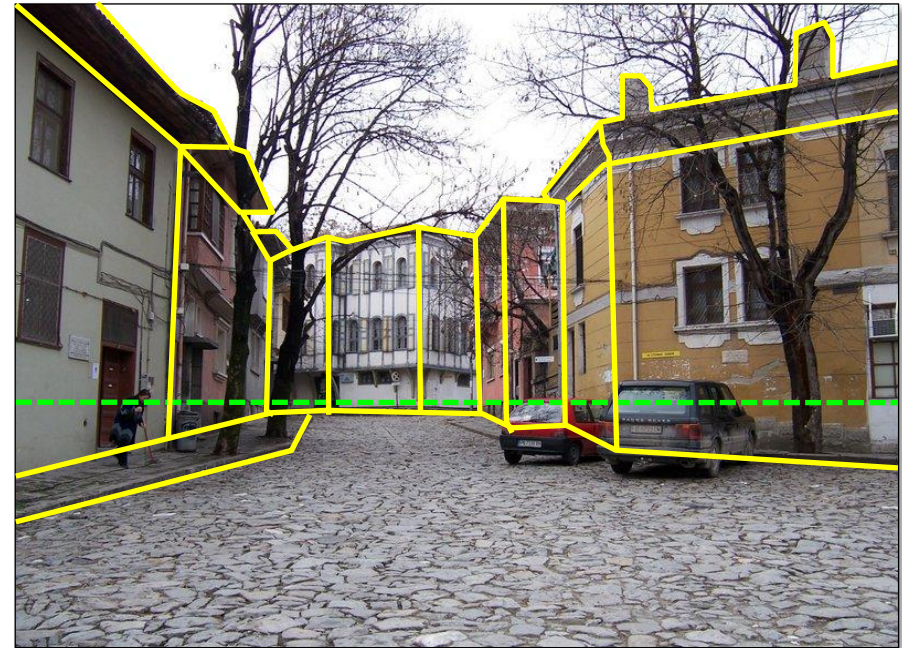
What is computer vision?

Automatic extraction of “**meaningful**” information from images and videos



Semantic information

(“Image Analysis and Computer Vision” course)



Geometric information

(this course)

Vision Demo?

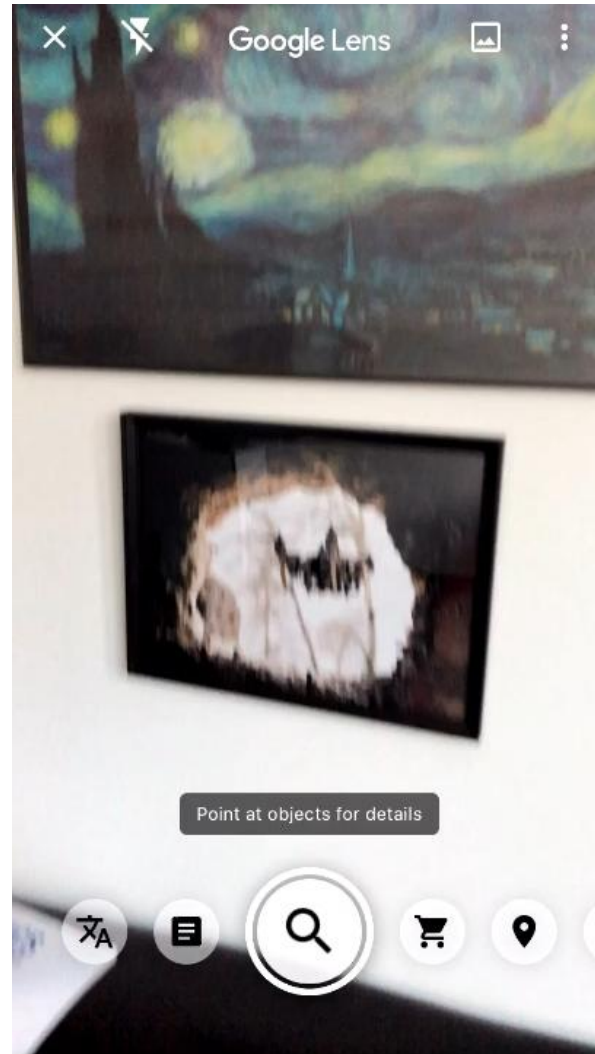


Terminator 2



Are we there? Almost!

Google App



Today's Outline

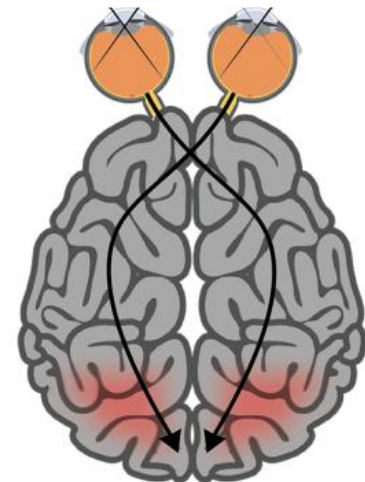
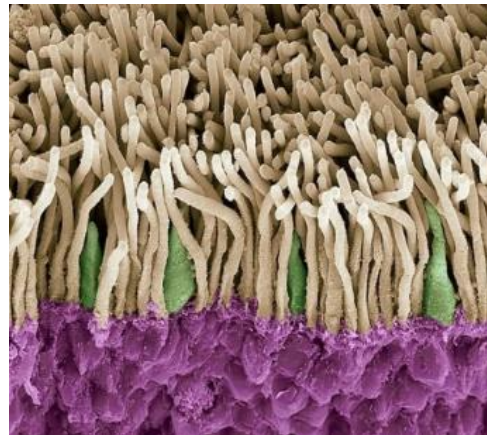
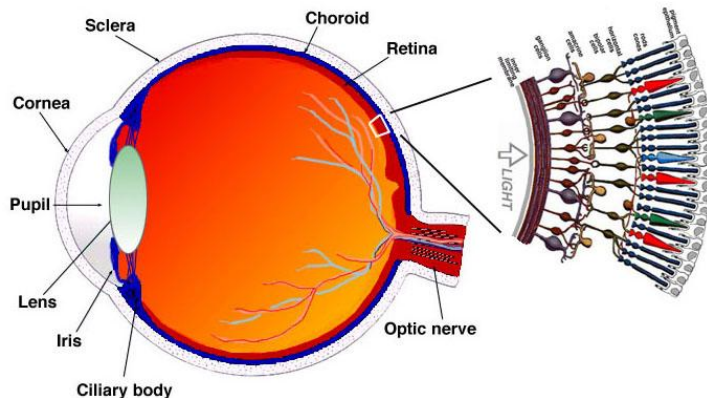
- About me and my research lab
- What is Computer Vision?
- Why study computer vision?
- Example of Vision Applications
- Organization of the course
- Start: Visual Odometry overview

Why study computer vision?

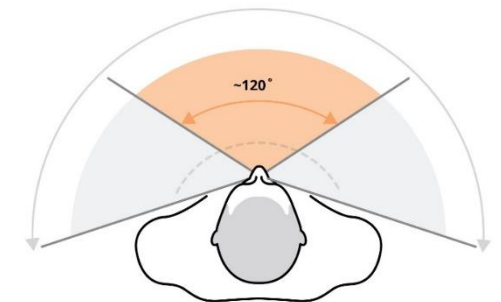
- **Relieve** humans of boring, easy tasks
- **Enhance** human abilities: human-computer interaction, visualization, augmented reality (AR)
- Perception for autonomous **robots**
- **Organize** and give access to visual **content**
- Lots of computer-vision **companies and jobs in Switzerland** (Zurich & Lausanne):
 - Meta (Zurich): AR/VR, Instagram
 - Huawei (Zurich): automotive, autonomous cars, event cameras, computational photography
 - Verity (Zurich): SLAM engineer
 - Perspective Robotics (Zurich): Computer vision engineer
 - Fixposition (Zurich): Sensor fusion engineer
 - 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)
 - VIZRT (Zurich): sport broadcasting, 3D replay
 - Pix4D (Lausanne): 3D reconstruction from drones
 - More on [glassdoor.ch](https://www.glassdoor.ch)

Vision in humans

- **Vision** is our most powerful sense. **Half of the primate cerebral cortex** is devoted to visual processing
- The retina is $\sim 1,000 \text{ mm}^2$. Contains **130 million photoreceptors** (120 mil. rods for low light vision and 10 mil. cones for color sampling) covering a **visual field of 220×135 degrees**
- Provides enormous amount of information: **data-rate of $\sim 3 \text{ GBytes/s}$**
- 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 small depression called **fovea**



visual cortex

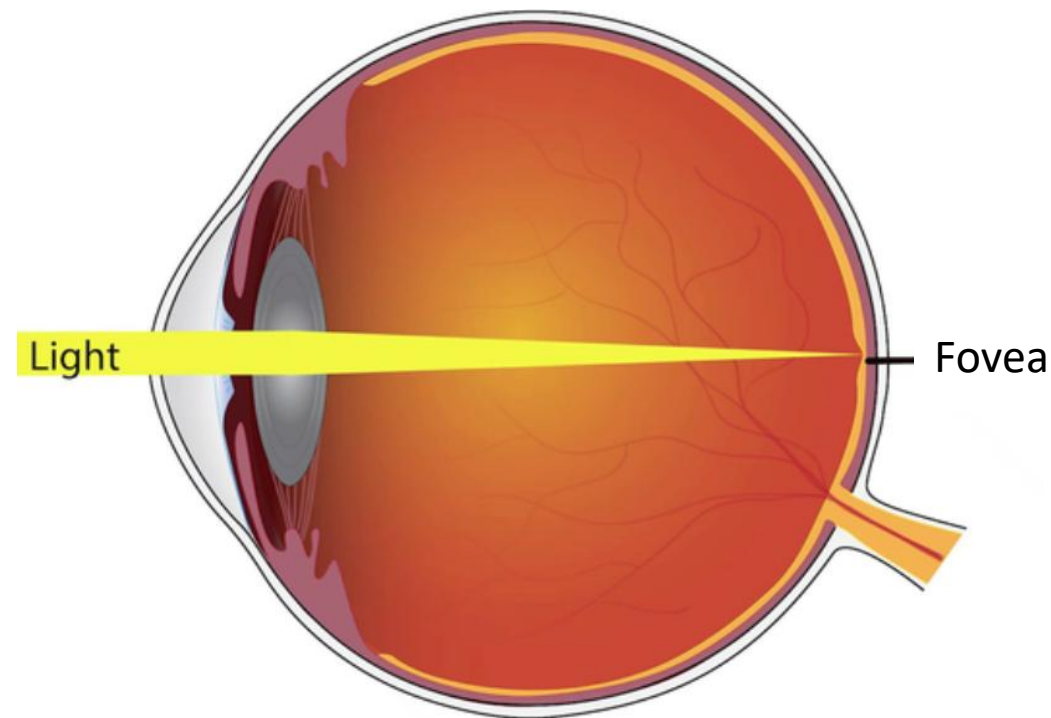


Vision in humans: how we see

- 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)**



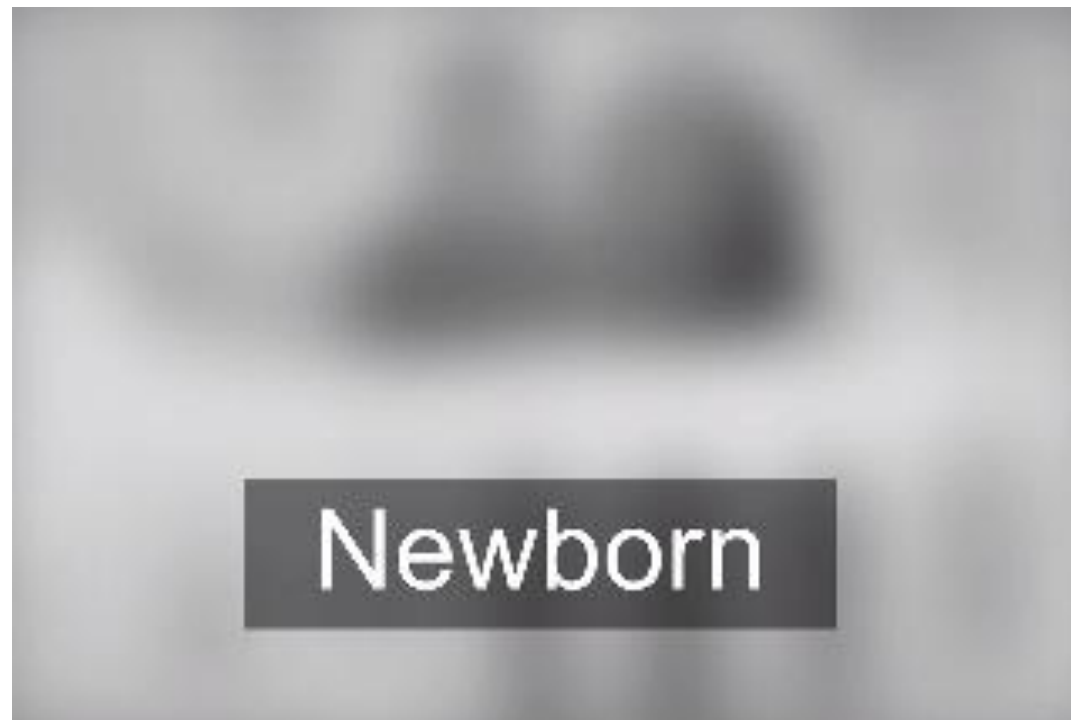
How we actually see. This principle is used in [foveated rendering](#)



If you are interested to study human perception, check out the UZH course "[Computational Vision](#)" (next semester)

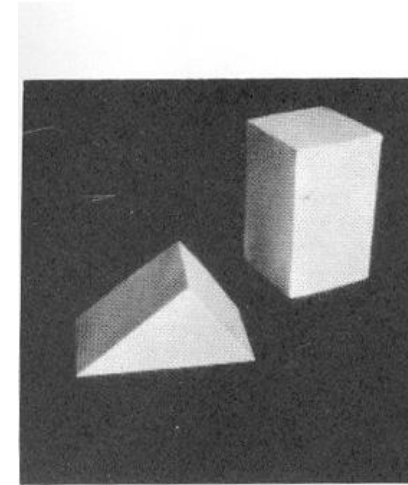
What a newborn sees every month in the first year

“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.”

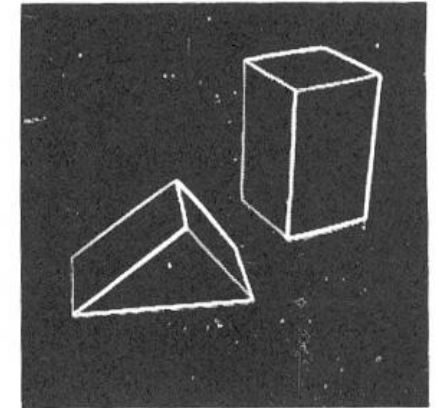


Origins of computer vision

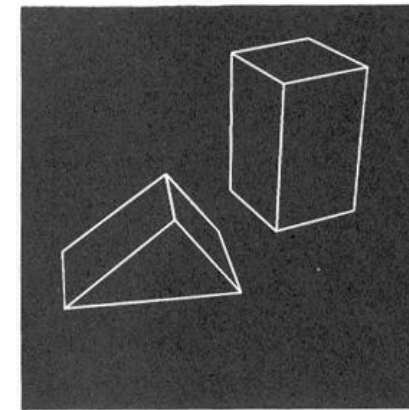
- **1963** - [L. G. Roberts](#) publishes his PhD thesis on *[Machine Perception of Three Dimensional Solids](#)*, thesis, MIT Department of Electrical Engineering
- He is the **inventor of ARPANET, the current Internet**



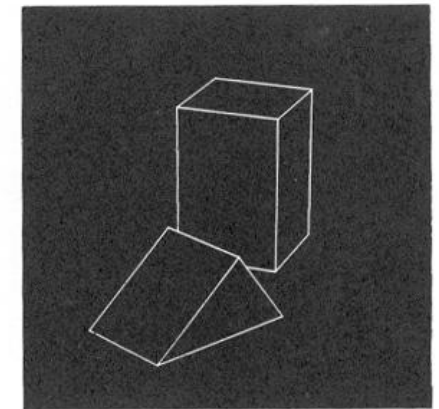
(a) Original picture.



(b) Differentiated picture.



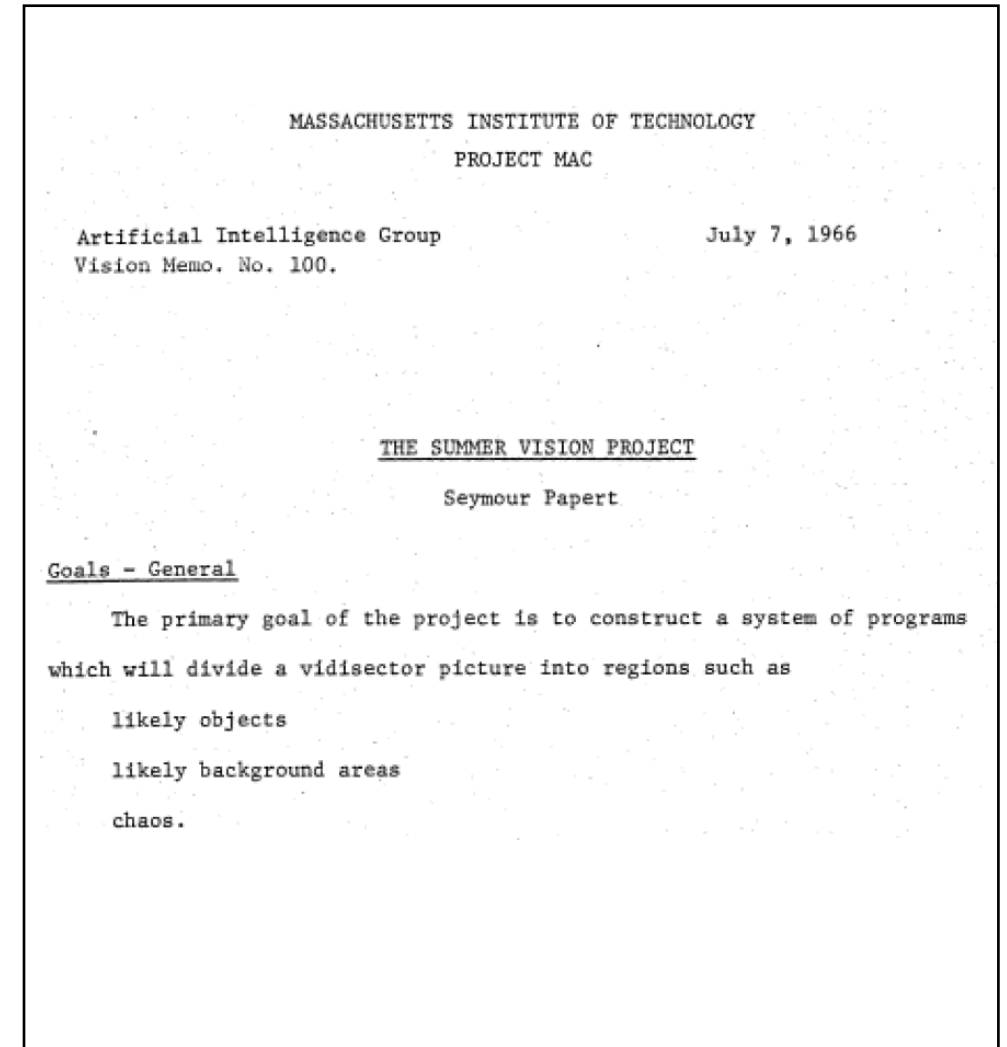
(c) Line drawing.



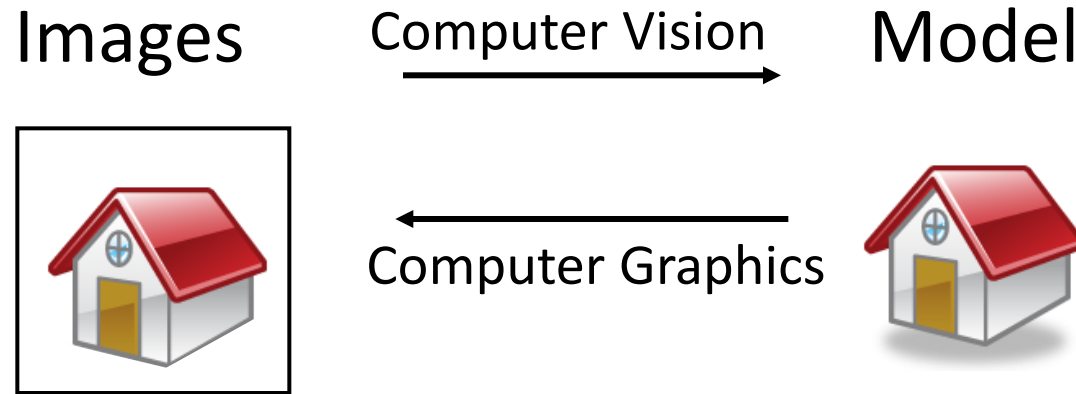
(d) Rotated view.

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](#)



Computer Vision vs Computer Graphics

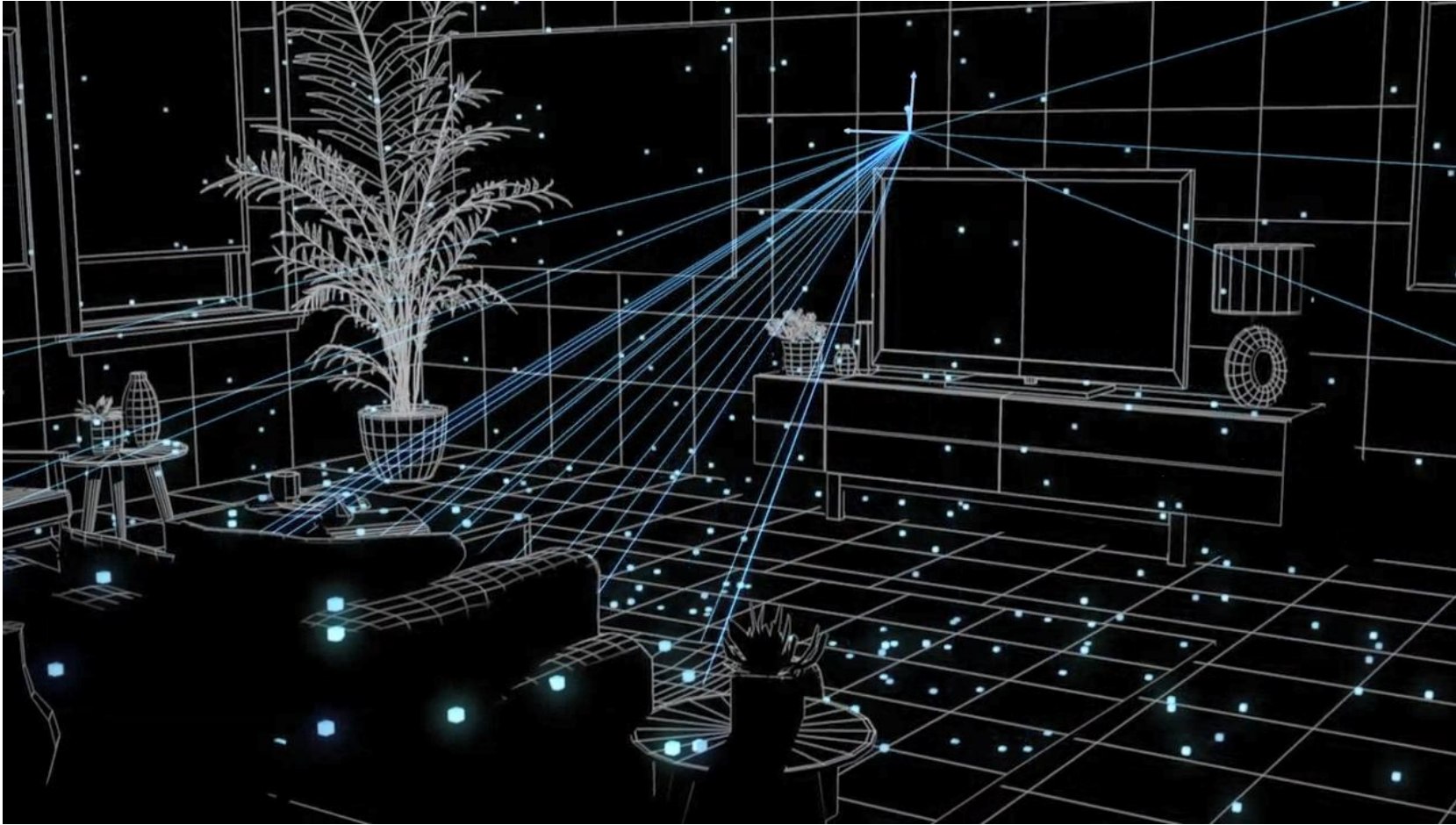


Inverse problems: analysis and synthesis.

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

VR/AR



Oculus Quest uses four cameras to track the pose of the head and the controllers

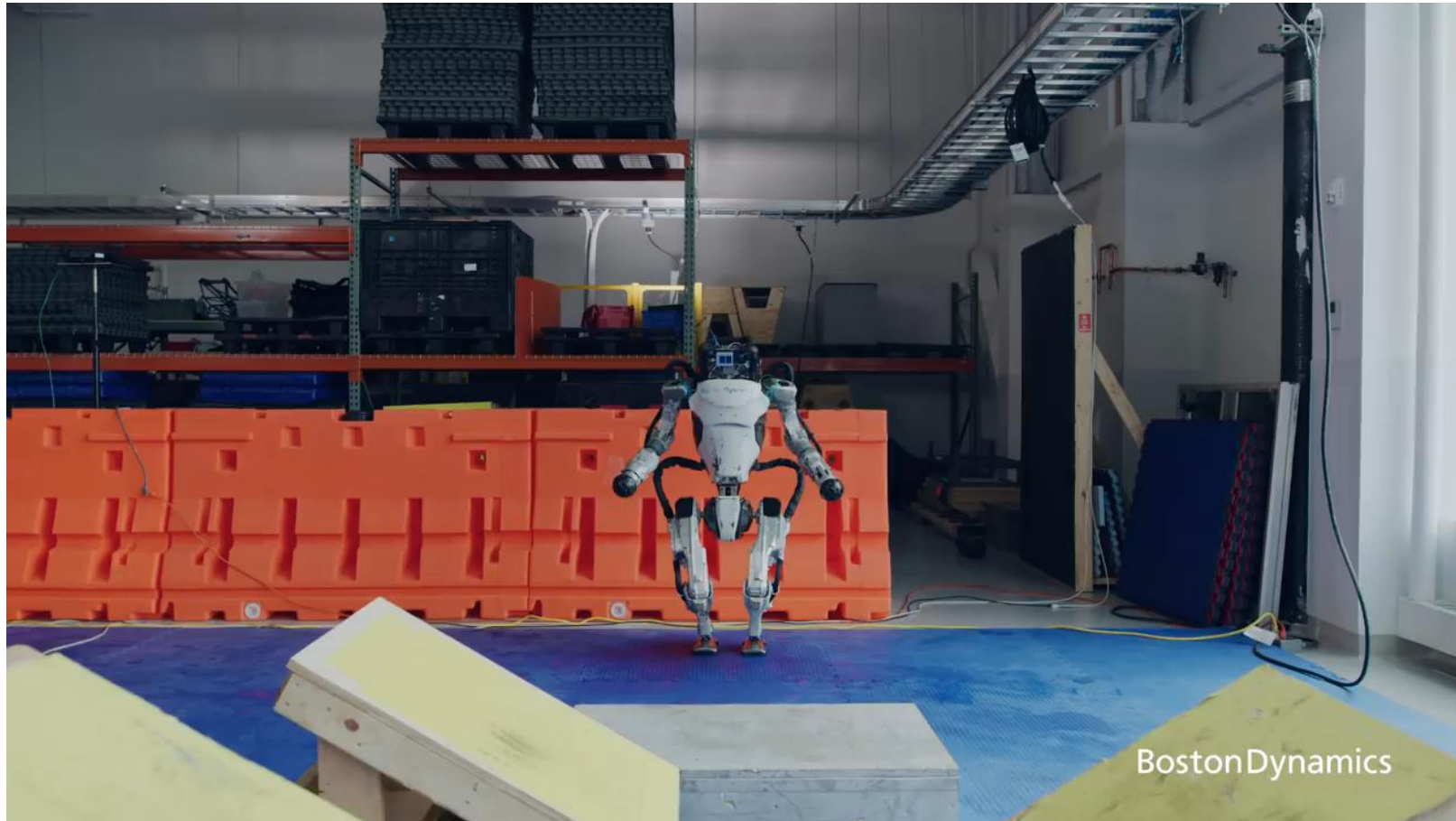
Advanced Driving Assistance Systems (ADAS)



[Mobileye](#): Vision system used at **BMW, GM, Volvo** models. Bought by **Intel** in **2017** for **15 billion USD**, it is used in **200 million cars** worldwide

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

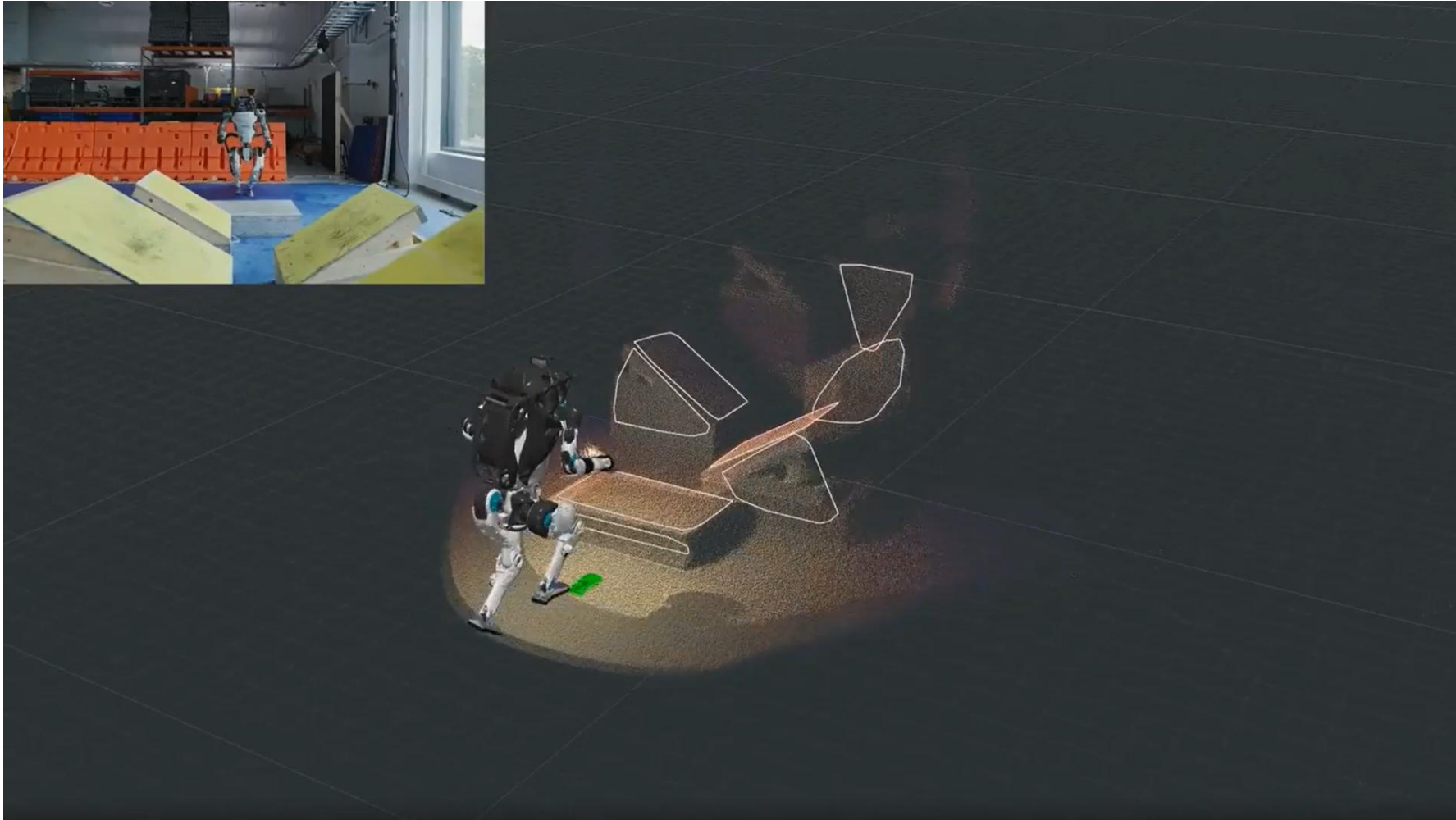
Boston Dynamics ATLAS Robot



<https://youtu.be/tF4DML7FIWk>

Watch Boston Dynamics keynote at the MIT Robotics Today Seminar Series: <https://youtu.be/EGABAx52GKI>

Boston Dynamics ATLAS Robot



<https://blog.bostondynamics.com/flipping-the-script-with-atlas>

Roomba Vacuum Cleaner

- Introduced in 2002 by iRobot; more than 50 million Roombas sold so far
- Fully autonomous, uses camera to recognize places



Until 2014, the Roomba robots navigated by zigzagging randomly across the room. Average cleaning time: 20 minutes



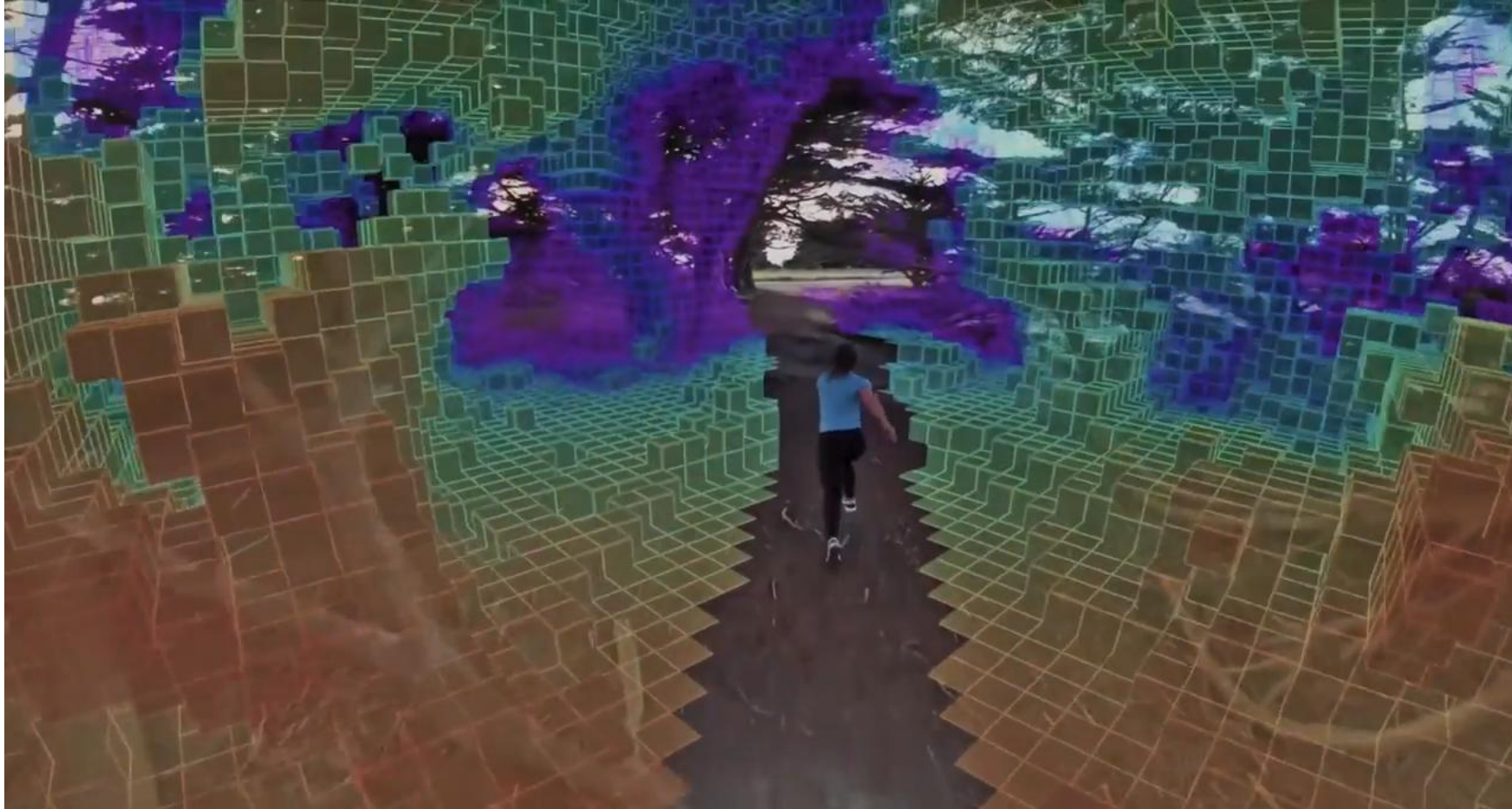
In 2015, the [Roomba 980](#) was introduced, which navigated by following a pre-defined path optimized thanks to visual SLAM - Average cleaning time: 5 minutes

Skydio and DJI Drones



The Skydio-2 drone features 7 cameras for obstacle avoidance, visual odometry, and person following
Watch Skydio keynote at the MIT Robotics Today Seminar Series: <https://youtu.be/ncZmnfIRIWE>

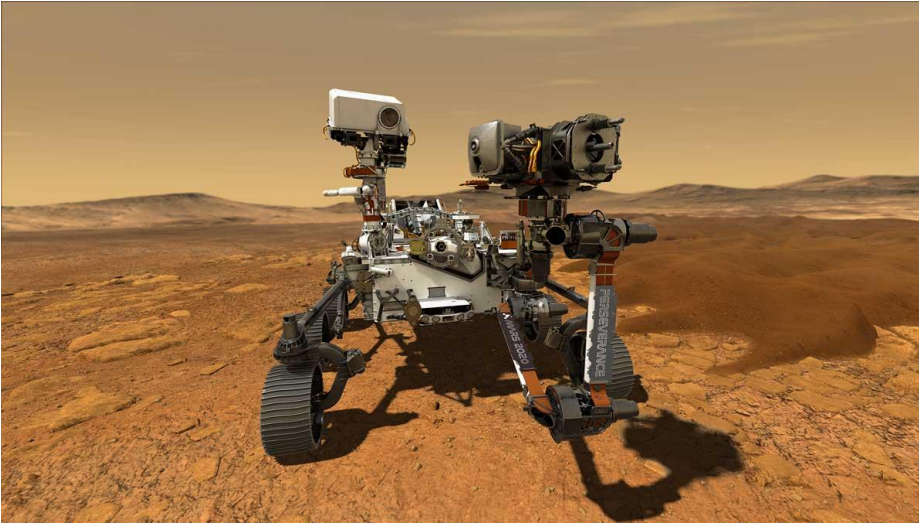
Skydio and DJI Drones



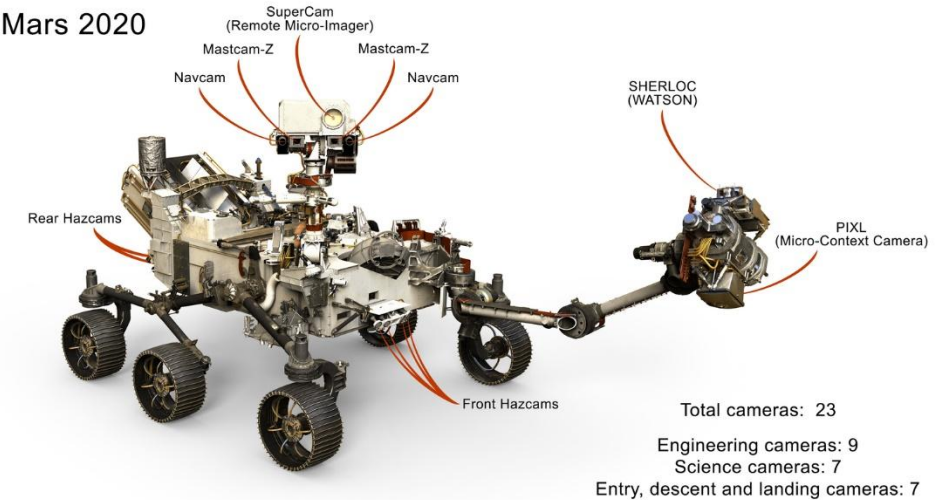
The Skydio-2 drone features 7 cameras for obstacle avoidance, visual odometry, and person following

Watch Skydio keynote at the MIT Robotics Today Seminar Series: <https://youtu.be/ncZmnfIRIWE>

NASA Mars Rovers



Mars 2020



[NASA'S Perseverance Rover](#) landed in 2021 features [23 cameras](#) used for:

- Autonomous landing on Mars (Terrain Relative Navigation, next slide)
- Autonomous navigation and positioning of the robotic arm
- Science: 3D photos and videos, analysis of chemical components of the soil

For more info, watch the RSS'21 keynote by Larry Matthies, head of the Computer Vision Group at NASA/JPL:

<https://youtu.be/Ncl6fJOzBsU>

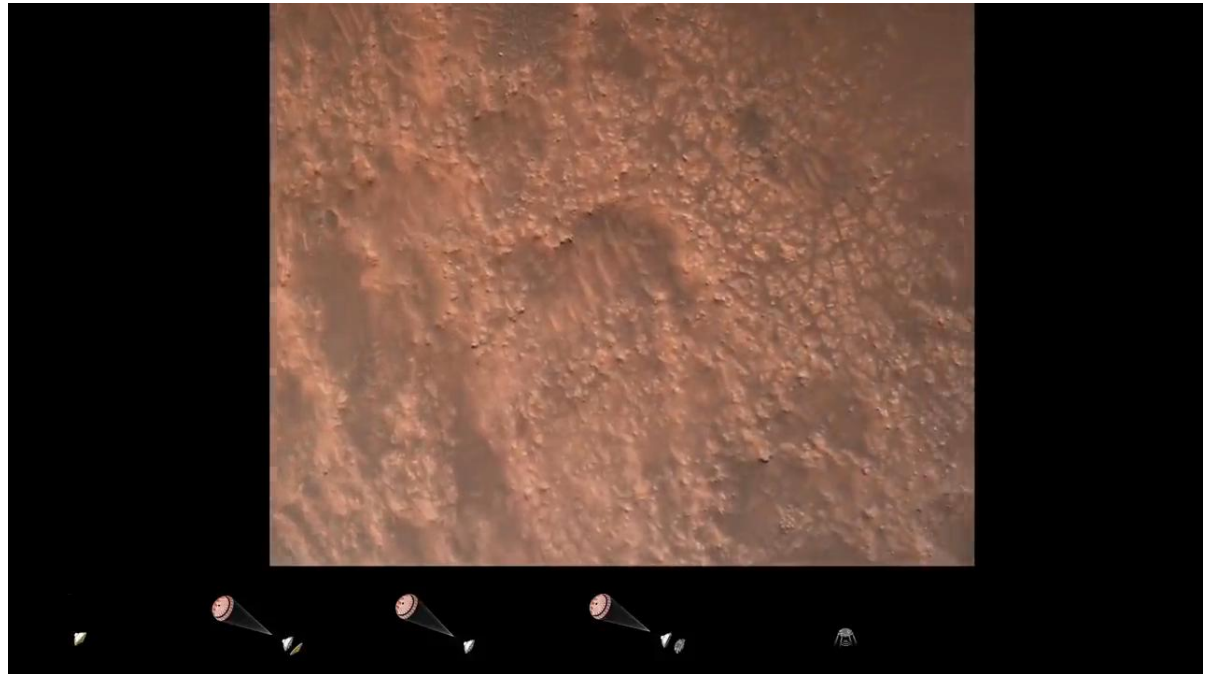
Perseverance Descent via Vision-based Terrain Relative Navigation

Landing accuracy: 40 meters.

"If we didn't have TRN, the probability of landing safely at Jezero Crater was about 80 to 85%. With TRN, the probability increased to 99%." [\[Link\]](#)



[Link 1](#), [Link 2](#)



Real footage recorded by Perseverance during descent <https://youtu.be/4czjS9h4Fpg>

Vision-based Flight on Mars



The NASA Ingenuity helicopter performs the first autonomous vision-based flight on April 19, 2021, on Mars.

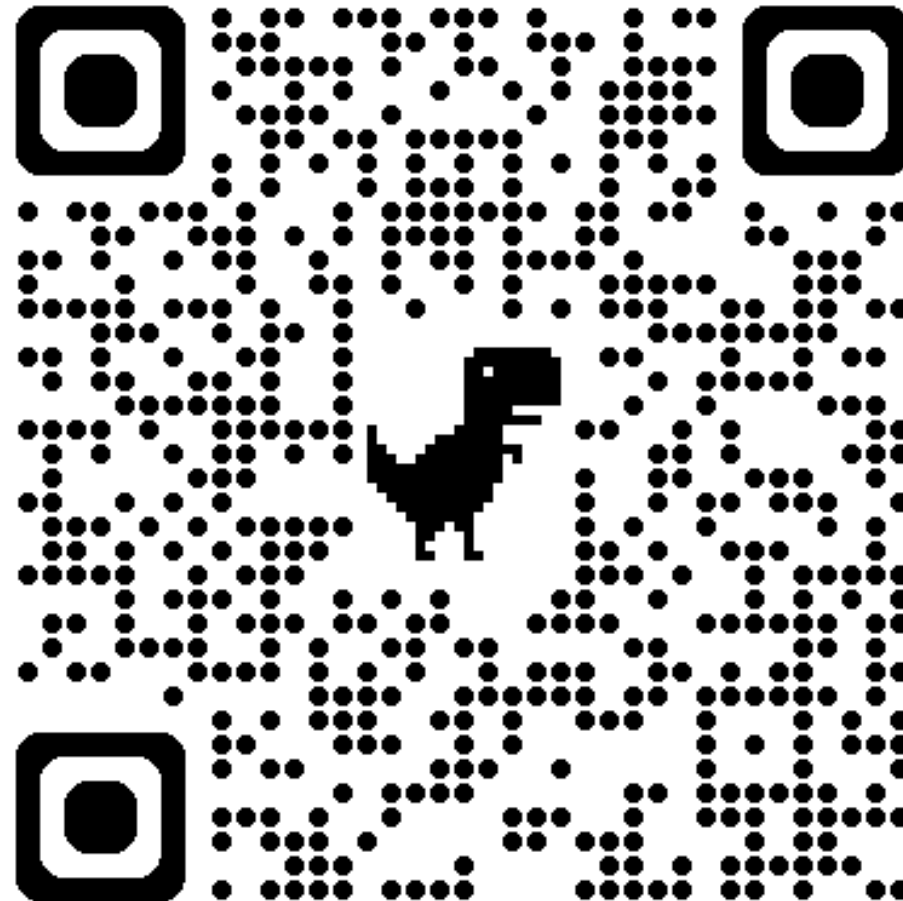
<https://youtu.be/p1KolyCqICI?t=2502>

<https://mars.nasa.gov/technology/helicopter/#>

On December 4, we will have a lecture by [Jeff Delaune](#), from NASA/JPL,
who developed the visual navigation of Ingenuity

Before the Break

Please fill this quick survey: <https://tinyurl.com/VAMR25>



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 16:00 to 18:00**
both in person or via ZOOM possible (**please announce yourself by email**)



- **Teaching Assistants:** Ismail Geles and Roberto Pellerito



Ismail Geles
geles (at) ifi (dot) uzh (dot) ch



Roberto Pellerito
rpellerito (at) ifi (dot) uzh (dot) ch

<https://rpg.ifi.uzh.ch/people.html>

Lectures and Exercises

Lectures:

- **08:00 to 09:45** every week. **After class, I usually stay for 10 more minutes** for questions
- **Breaks:** always but can vary between 5-15 min
- Room: SOC-F-106, Rämistrasse 69, 8001 Zurich

Exercises:

- **12:15 to 13:45** every week: starting today with a tutorial on camera notation
- Room: same as above

Course & Exam Registration and Cancellation

- Registration and exam cancellation deadline for the course is **October 7, 23:59 hrs**
- NB: at UZH, when you register for a course, you are also automatically registered for the exam. If you want to cancel the exam, you must unbook the course by October 7. Afterward, it will no longer be possible to cancel the exam. No show at the exam will be graded as 1.0. If you cannot take the exam because you fell ill, you must submit a petition with medical certificate no later than five business days after the examination date. If you are a student with a disability, you must request assistance in due time. Further info [here](#)

Tentative Course Schedule

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

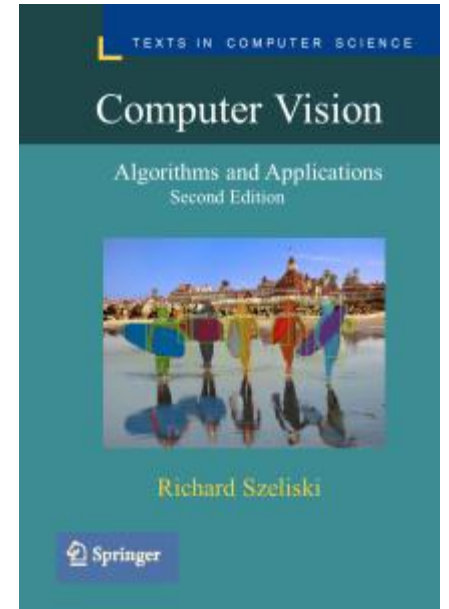
18.09.2025	Lecture 01 - Introduction to Computer Vision and Visual Odometry Exercise: Camera Notation Tutorial	Scaramuzza Ismail, Roberto
25.09.2025	Lecture 02 - Image Formation: perspective projection and camera models Exercise 01- Augmented reality wireframe cube	Scaramuzza Ismail, Roberto
02.10.2025	Lecture 03 - Camera Calibration Exercise 02 - PnP problem	Scaramuzza Ismail, Roberto
09.10.2025	Lecture 03 continued Lecture 04 - Filtering & Edge detection Exercise session replaced by continuation of Lecture 4	Scaramuzza
16.10.2025	Lecture 05 - Point Feature Detectors, Part 1 Exercise 03 - Harris detector + descriptor + matching	Scaramuzza Ismail, Roberto
23.10.2025	Lecture 06 - Point Feature Detectors, Part 2 Exercise 04 - SIFT detector + descriptor + matching	Leonard Ismail, Roberto
30.10.2025	Lecture 07 - Multiple-view Geometry 1 Exercise 05 - Stereo vision: rectification, epipolar matching, disparity, triangulation	Leonard Ismail, Roberto
06.11.2025	Lecture 08 - Multiple-view Geometry 2 Exercise 06 - Eight-Point Algorithm	Scaramuzza Ismail, Roberto
13.11.2025	Lecture 09 - Multiple-view Geometry 3 Exercise 07 - P3P algorithm and RANSAC	Scaramuzza Ismail, Roberto
20.11.2025	Lecture 10 - Multiple-view Geometry 4 Continuation of Lecture 10 + Exercise 08 on Bundle Adjustment	Scaramuzza Ismail, Roberto
27.11.2025	Lecture 11 - Optical Flow and KLT Tracking Exercise 09 - Lucas-Kanade tracker	Scaramuzza Ismail, Roberto
04.12.2025	1st hour: seminar by Dr. Delaune from NASA-JPL: "Vision-Based Navigation for Planetary Exploration" Lecture 12a (2nd hour) - Place Recognition Lecture 12b (3rd and 4th hour, replaces exercise) - Deep Learning Tutorial Optional Exercise on Place Recognition	Scaramuzza Ismail, Roberto
11.12.2025	Lecture 13 - Visual inertial fusion + lab visit after the exercise Exercise 10 – Visual Inertial fusion	Scaramuzza Ismail, Roberto
18.12.2025	Lecture 14 - Event-based vision Exercise 11: Contrast Maximization for Event Cameras	Scaramuzza Ismail, Roberto

Study Material

- **Schedule, lecture slides, exercise download, mini projects, course info** on the official course website: <https://rpg.ifi.uzh.ch/teaching.html>
- **Video Recordings** of lectures and exercises will be uploaded to **OLAT**:
<https://lms.uzh.ch/auth/RepositoryEntry/17765990822/CourseNode/112478900904563>
- Post any **questions** related to lectures or exercises in the **OLAT Forum**

Reference Textbooks

- **Computer Vision: Algorithms and Applications, 2nd Edition**, by Richard Szeliski.
Freely downloadable from the author webpage: <https://szeliski.org/Book/>
- **Chapter 4 of Autonomous Mobile Robots**, 2nd edition, 2011, by R. Siegwart, I.R. Nourbakhsh, D. Scaramuzza. [PDF](#)
- **SLAM Handbook**: <http://asrl.utias.utoronto.ca/~tdb/slam/>
- **Additional readings** (i.e., **optional and not requested at the exam**) for interested students will be provided along with the slides and linked directly from the course website
- Further readings:
 - *Robotics, Vision and Control: Fundamental Algorithms*, 2nd edition, by Peter Corke
 - *An Invitation to 3D Vision*: Y. Ma, S. Soatto, J. Kosecka, S.S. Sastry
 - *Multiple view Geometry*: R. Hartley and A. Zisserman



Prerequisites

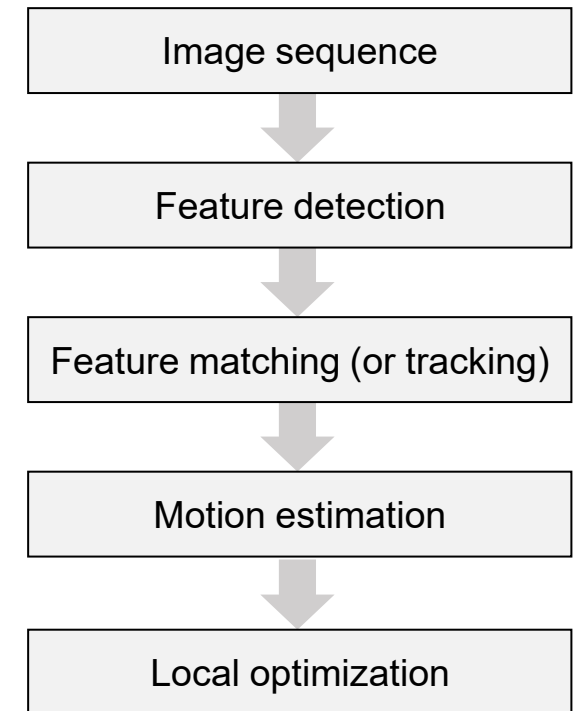
- **Linear algebra**
- **Matrix calculus:** matrix multiplication, inversion, singular value decomposition
 - Check out this [Linear Algebra Primer](#) from Stanford University
 - Check out this [Immersive Linear Algebra](#) interactive tool by Ström, Åström, and Akenine-Möller
 - Check out this [tutorial](#) on camera pose notation by Paul Furgale
- **No prior knowledge of computer vision and image processing is required**

Learning Objectives

- **High-level goal:** learn to implement the visual-inertial odometry algorithms used in current mobile robots (drones, cars, planetary robots), AR/VR products (Meta Quest, Microsoft HoloLens, Magic Leap), and Google Visual Positioning Service (e.g., Google Map Live View).
- 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,
 - feature and template tracking,
 - image retrieval,
 - event-based vision,
 - visual-inertial odometry, Simultaneous Localization And Mapping (SLAM),
 - and some basics of deep learning.

Exercises

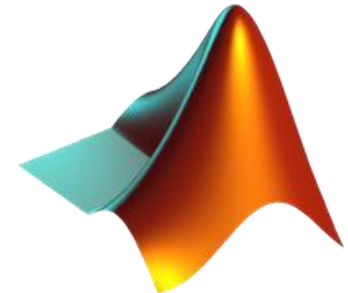
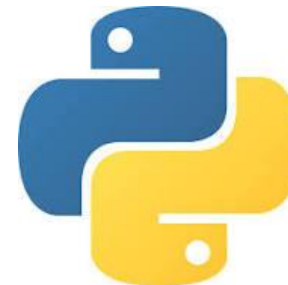
- **Learning Goal** of the exercises: **Implement a full visual odometry pipeline** (like the one running on Mars rovers).
- **Each week** you will learn how to implement a **building block** of visual odometry.
- Two exercises will be dedicated to **system integration**.
- **NB: Questions about the implementation details of each exercise can be asked at the exam.**



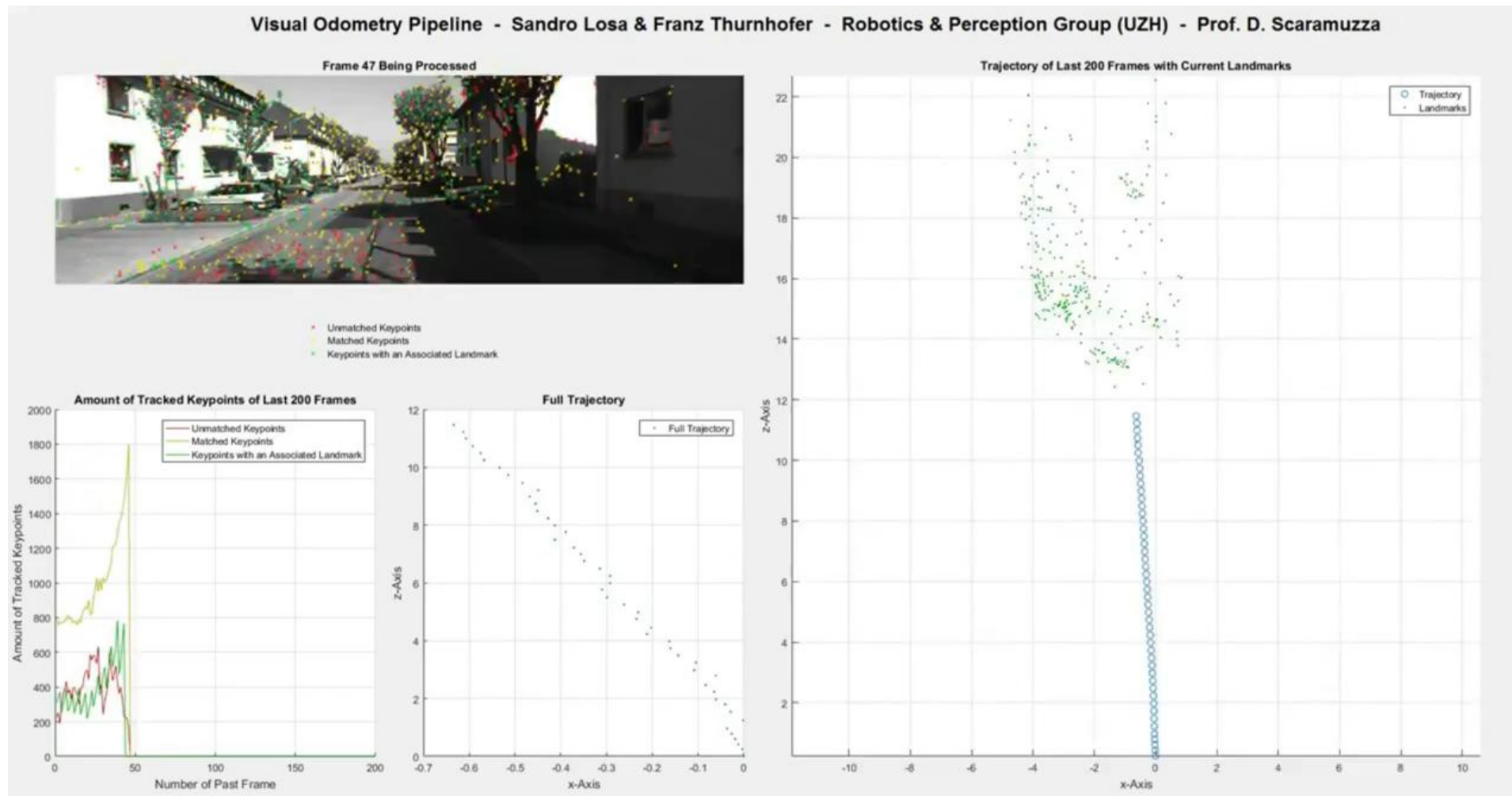
Building blocks of visual odometry
along with information flow

Exercises

- **Bring your own laptop**
- Exercises in **Python or Matlab**. You will need to have Matlab or Python already pre-installed on your machine for the exercises.
- Python can be downloaded from [here](#).
- You can download Matlab from:
 - **ETH:** Download: <https://itshop.ethz.ch/EndUser/Items/Home>
 - **UZH:** Download: <https://www.zi.uzh.ch/de/students/software-elearning/campussoft.html>
 - An introductory tutorial on Matlab can be found here: <http://rpg.ifi.uzh.ch/docs/teaching/2025/MatlabPrimer.pdf>
 - **Please install all the toolboxes included in the license.** If you don't have enough space in your PC, then install at least the Image Processing, Computer Vision, and Optimization toolboxes



Outcome of last year exercises



Grading and Exam

- The **final grade is based on a written exam** (2 hours)
 - Exam date: **January 8, 2026, from 08:00 to 10:00 on site**
 - **Closed-book exam**
 - **All the exams of the past years can be found on OLAT**
- **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** or **Python** (but C++ or 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 **05.01.2026**.
 - **Group work: minimum 2, max 4 people.**

Class Participation

- **Strong class participation is encouraged!**
- 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) ?

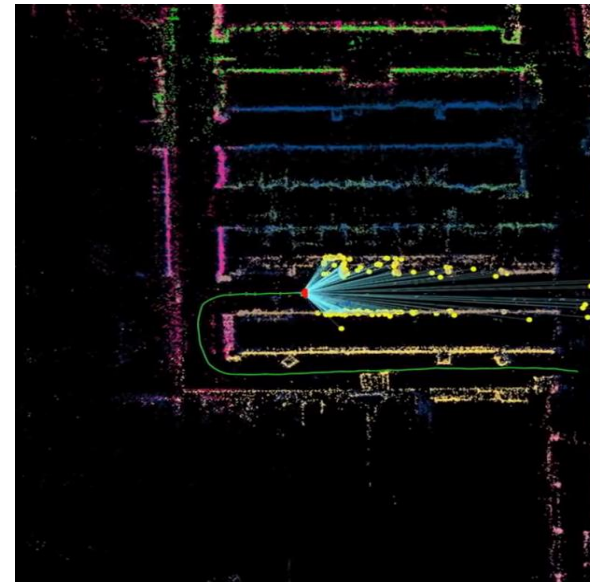
VO is the process of incrementally estimating the pose of the vehicle by examining the changes that motion induces on the images of its onboard cameras

input



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

output

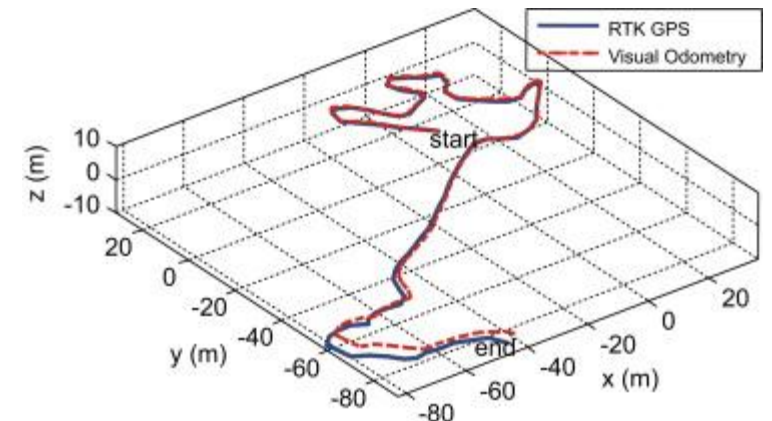
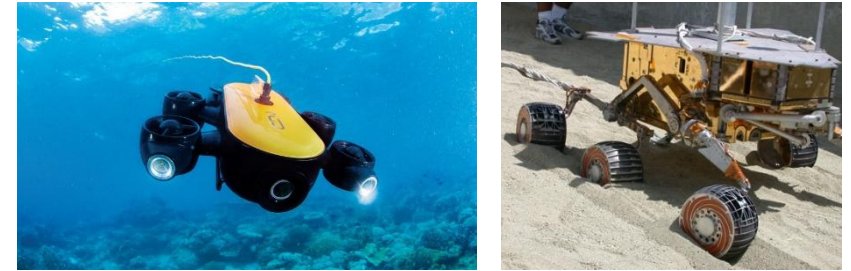


Camera trajectory (3D structure is a plus)

$$R_0, R_1, \dots, R_i$$
$$t_0, t_1, \dots, t_i$$

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?



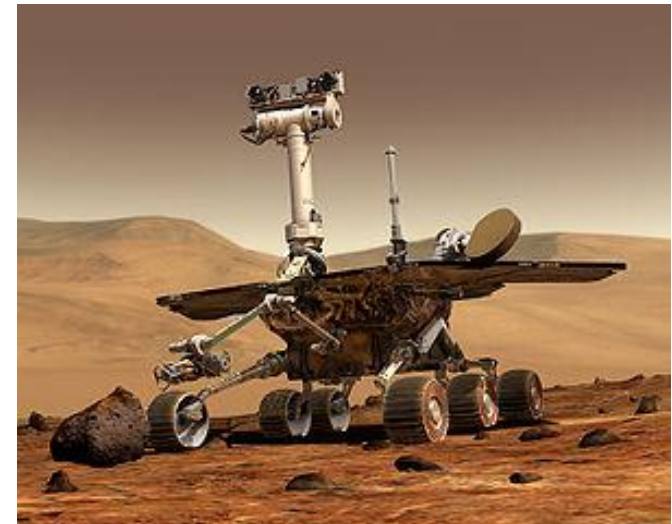
A Brief history of VO

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

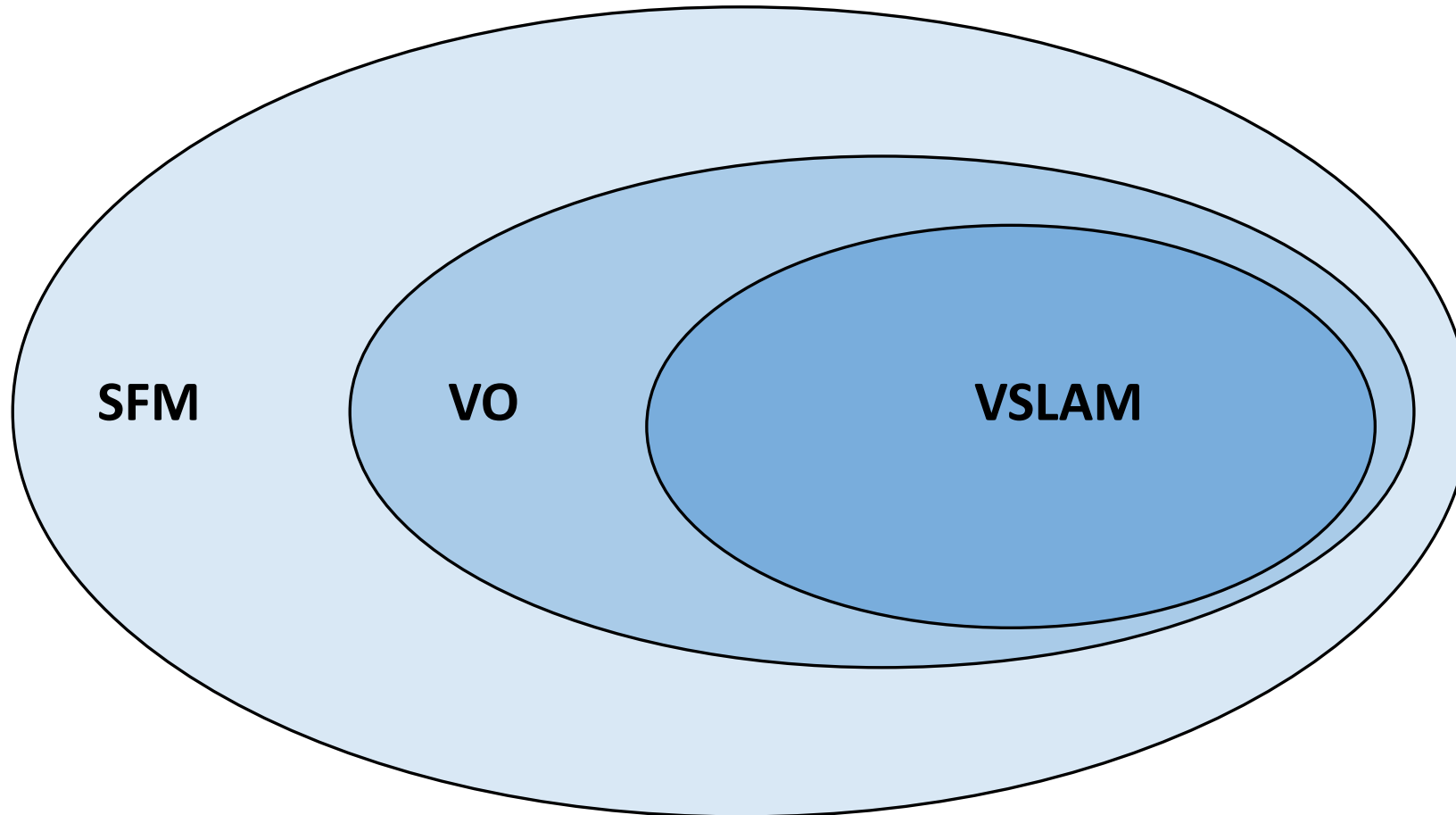


A Brief history of VO

- **1980**: First known VO real-time implementation on a robot by **Hans Moravec** 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 is used on the **Mars helicopter**

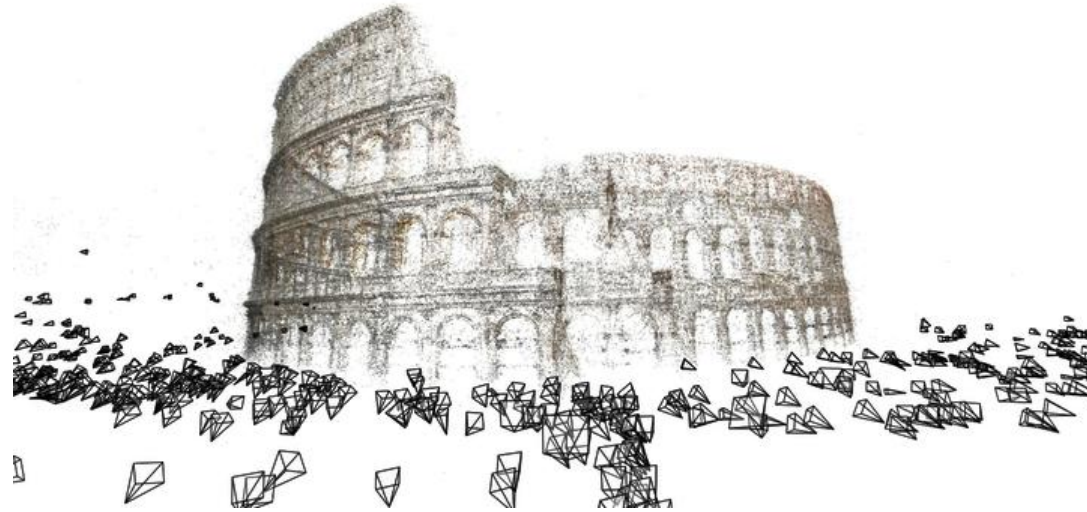


VO vs VSLAM vs SFM



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.](#)

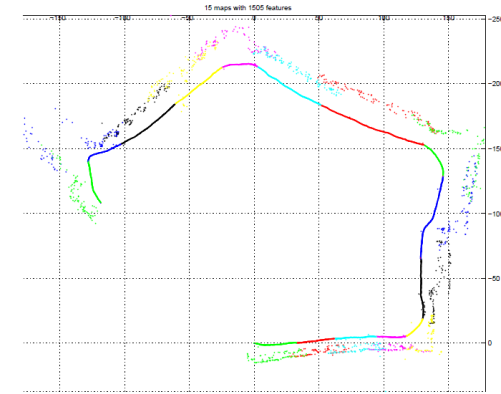
State of the art software: [COLMAP](#)

VO vs SFM

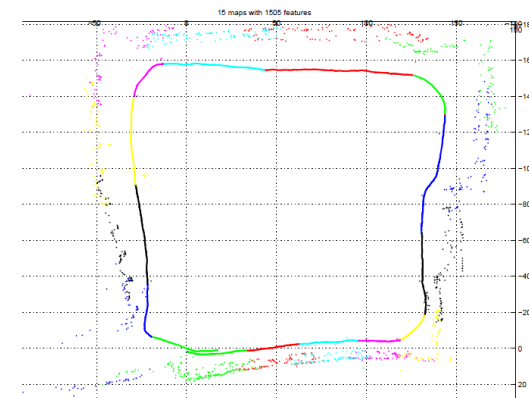
- VO is a **particular case** of SFM
- VO focuses on estimating the 6DoF motion of the camera **sequentially** (as a new frame arrives) and in **real time**
- Terminology: sometimes **SFM** is used as a **synonym** of **VO**

VO vs. Visual SLAM

- **Visual Odometry**
 - Focuses 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)
 - It's a particular case of VO
 - 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



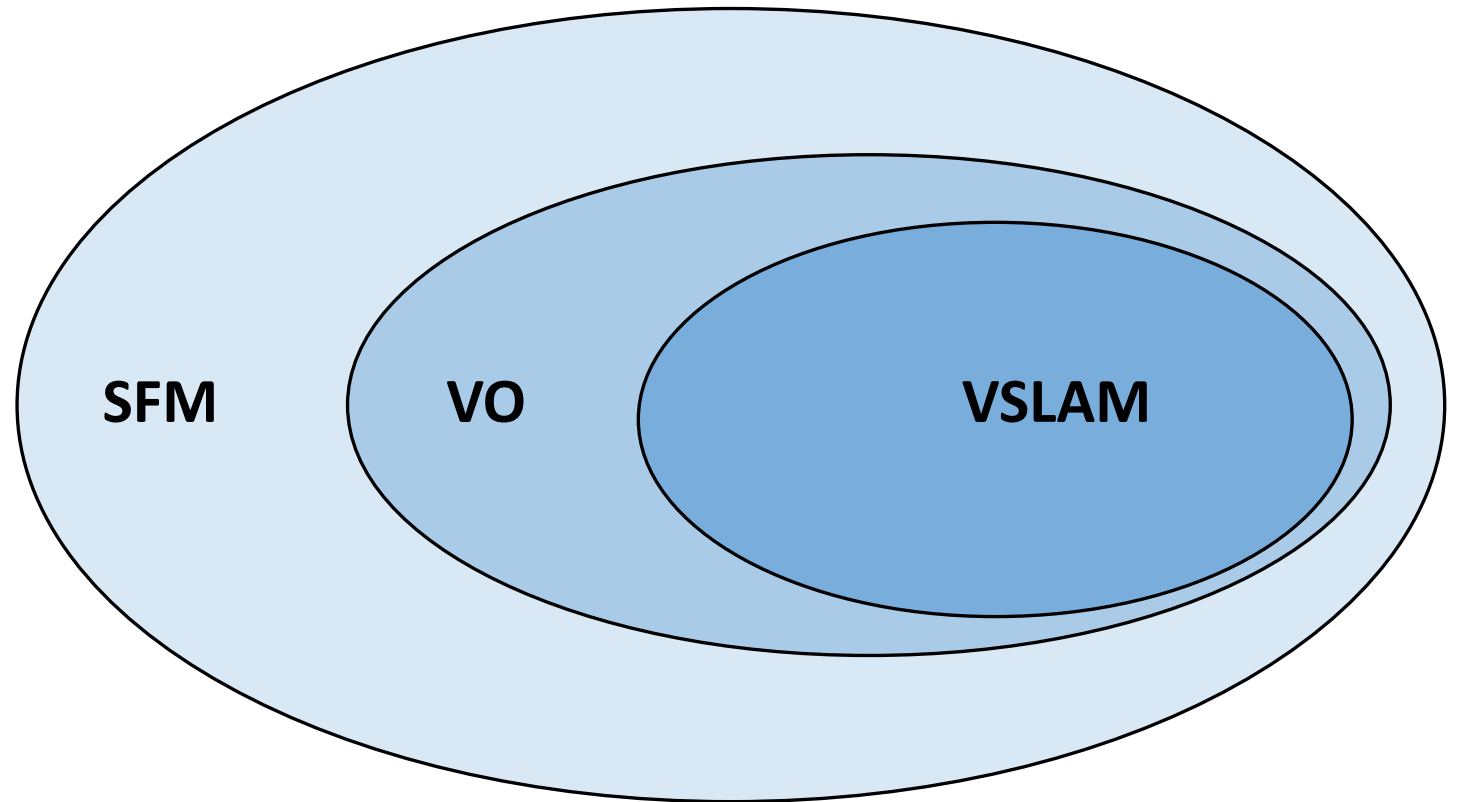
Visual SLAM

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

$$\text{VSLAM} \subseteq \text{VO} \subseteq \text{SFM}$$

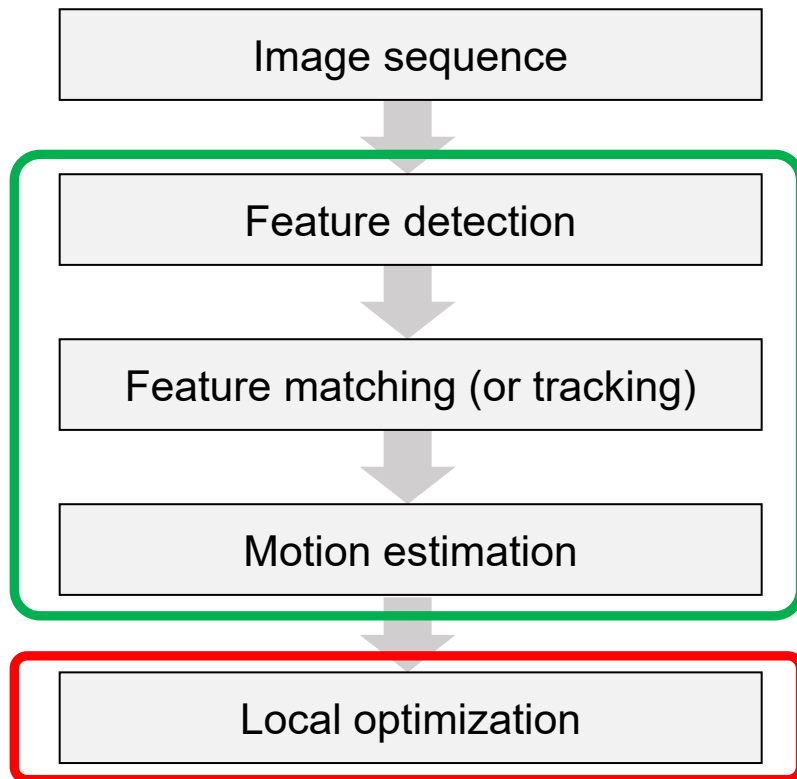
Why?

- **because every VSLAM and VO are SFM**, but not every SFM is VO or SLAM
- **because every VSLAM is a VO**, but not every VO is a SLAM
 - VSLAM applies more stringent requirements, such as loop detection and closure, than VO, making it a particular case. Every VSLAM functions as a VO, given that VSLAM, like VO, incrementally estimates poses (although Bundle Adjustment may further refine these estimations). Moreover, if VSLAM achieves global consistency, it inherently ensures local consistency as well.



VO Flow Chart

VO computes the camera path incrementally (pose after pose)

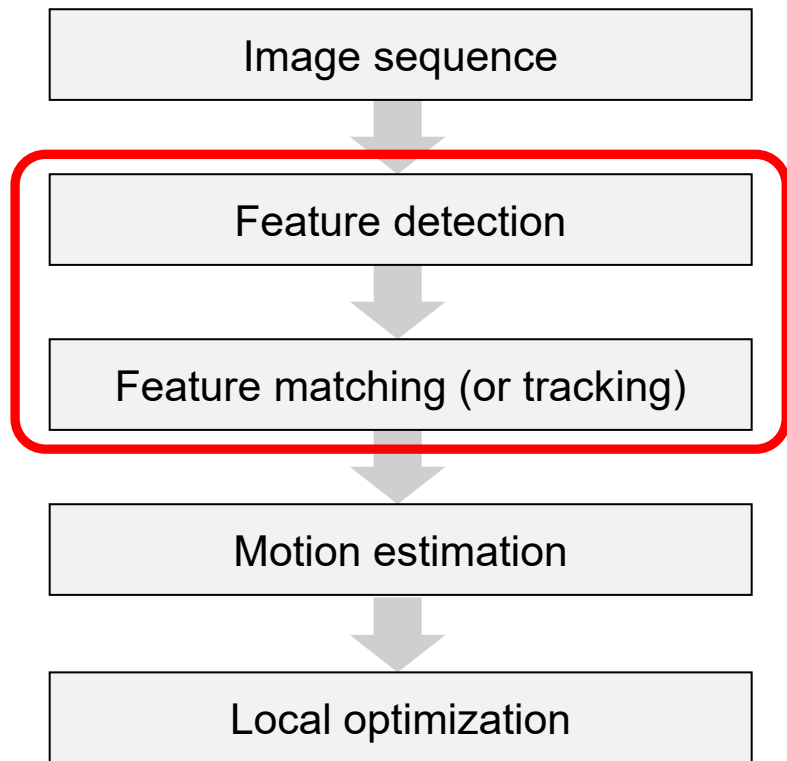


Front-end: outputs the *relative pose* between the *last two frames*

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

VO Flow Chart

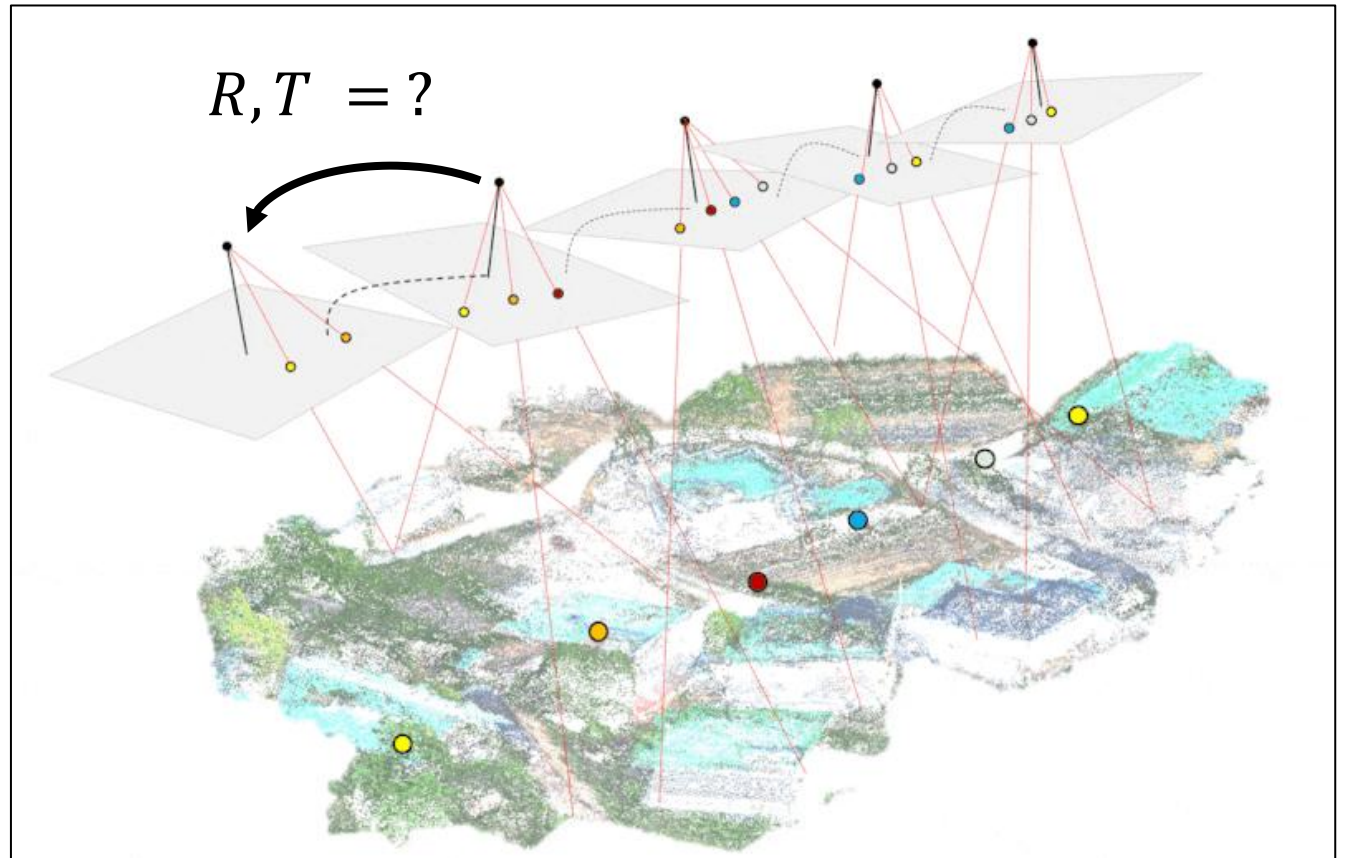
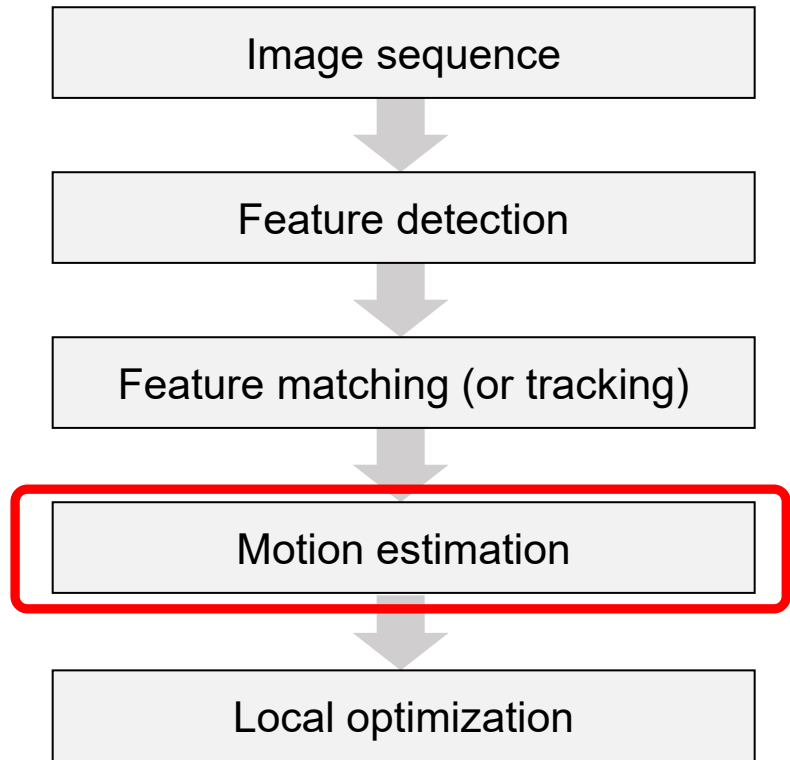
VO computes the camera path incrementally (pose after pose)



Features tracked over multiple recent frames overlaid on the last frame

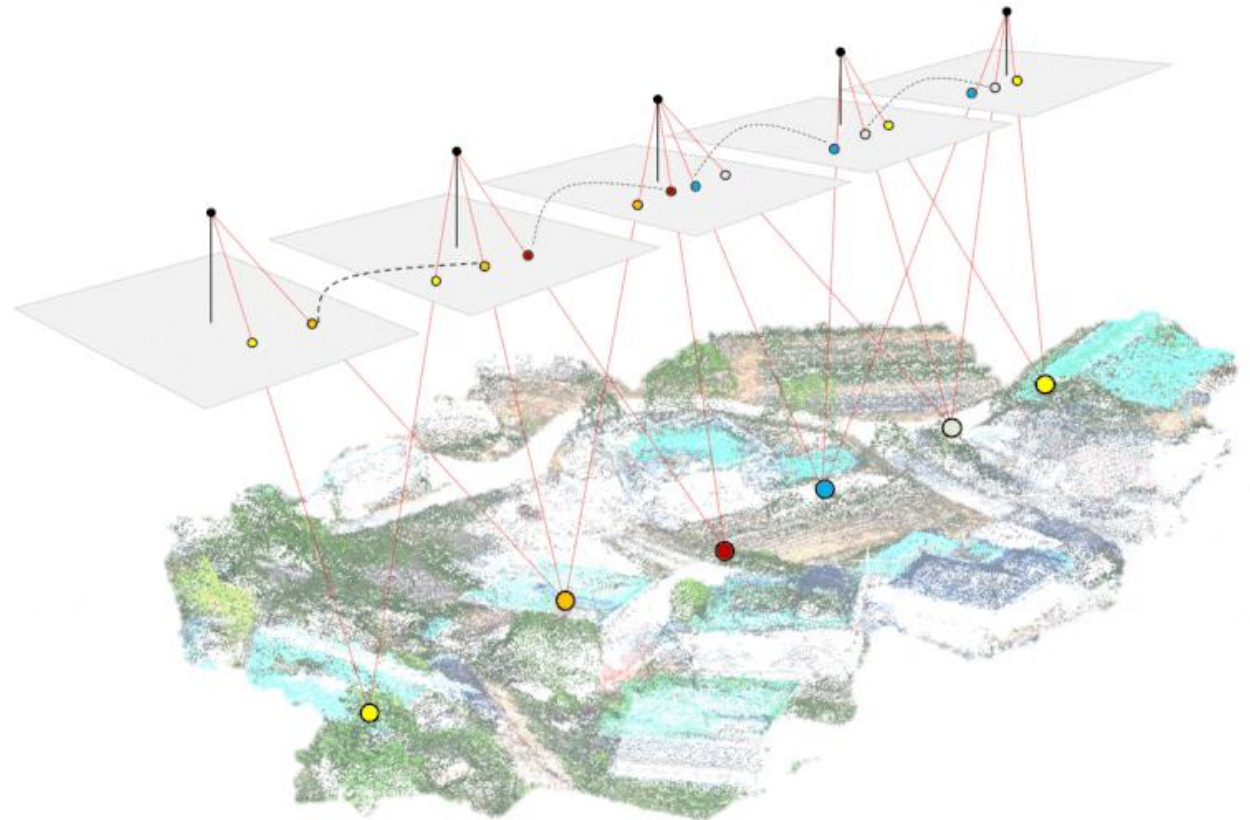
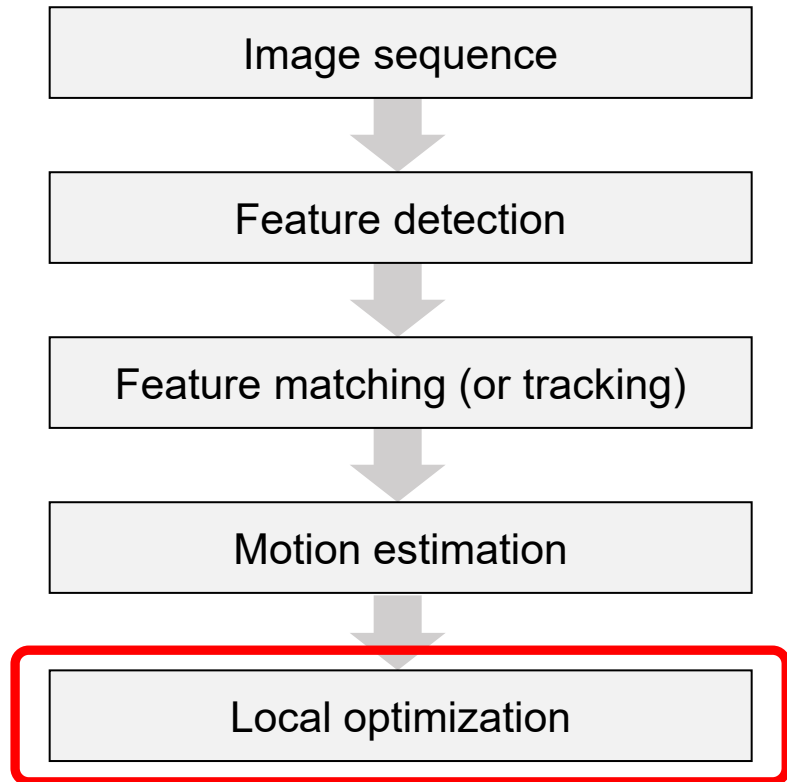
VO Flow Chart

VO computes the camera path incrementally (pose after pose)



VO Flow Chart

VO computes the camera path incrementally (pose after pose)



Course Topics

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

Course Topics

- Principles of image formation
 - Perspective projection
 - Camera calibration

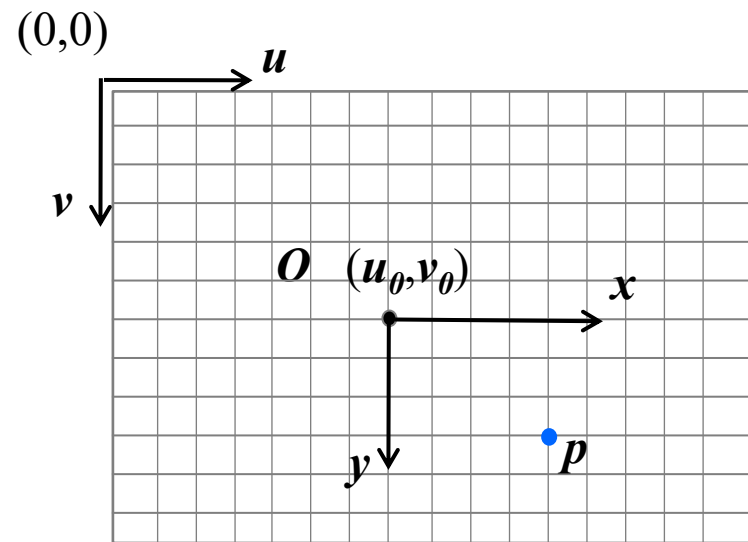


Image plane

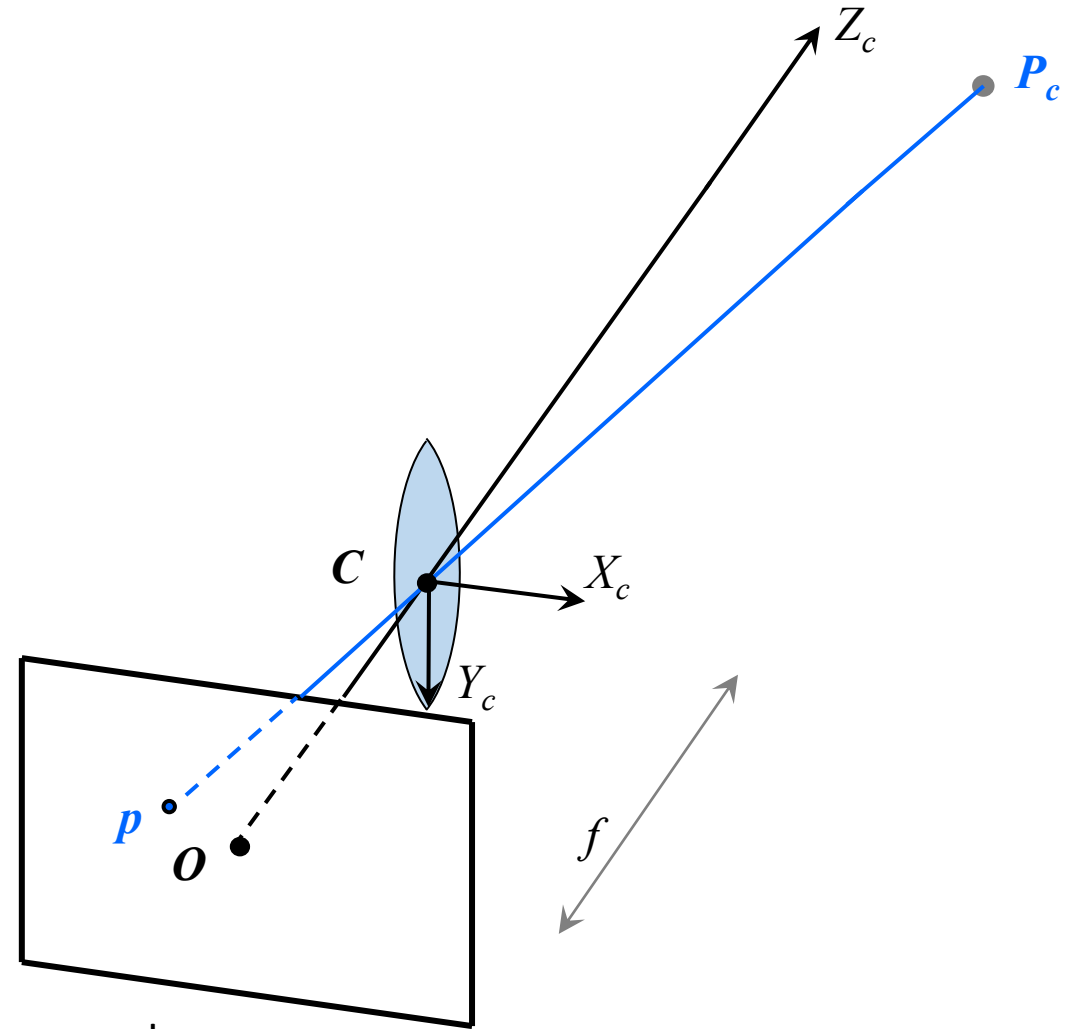
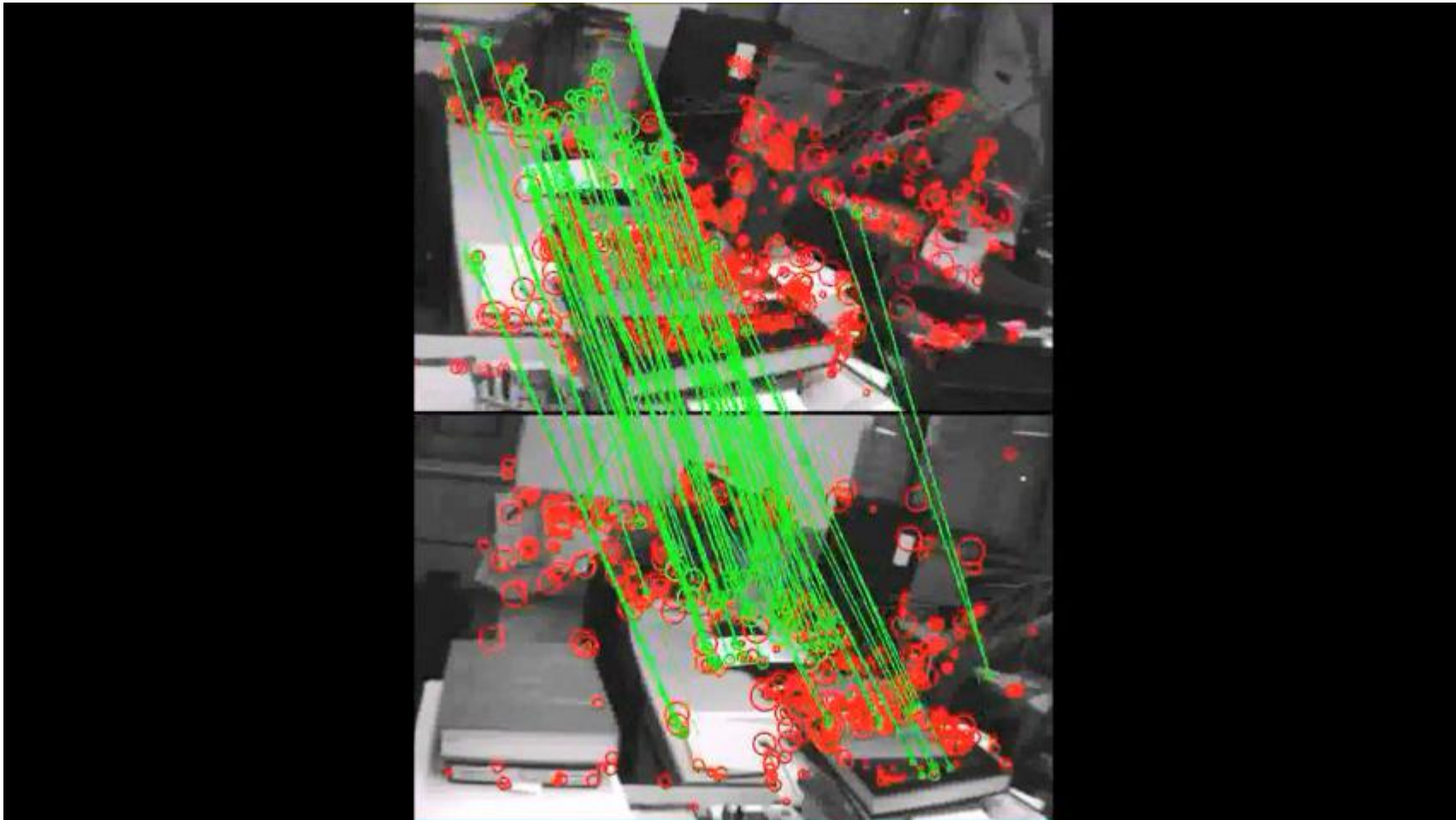


Image plane

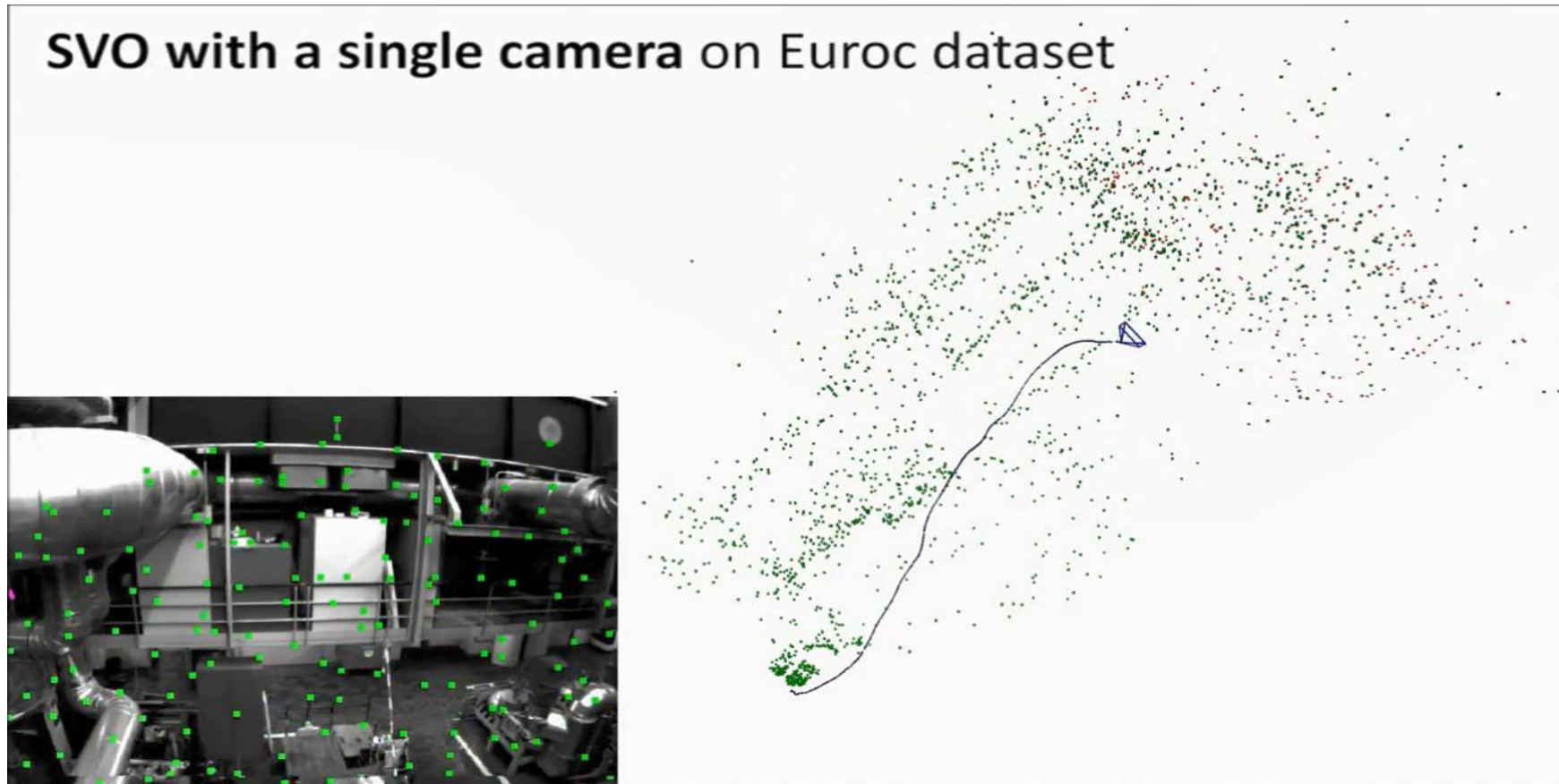
Course Topics

- Feature detection and matching



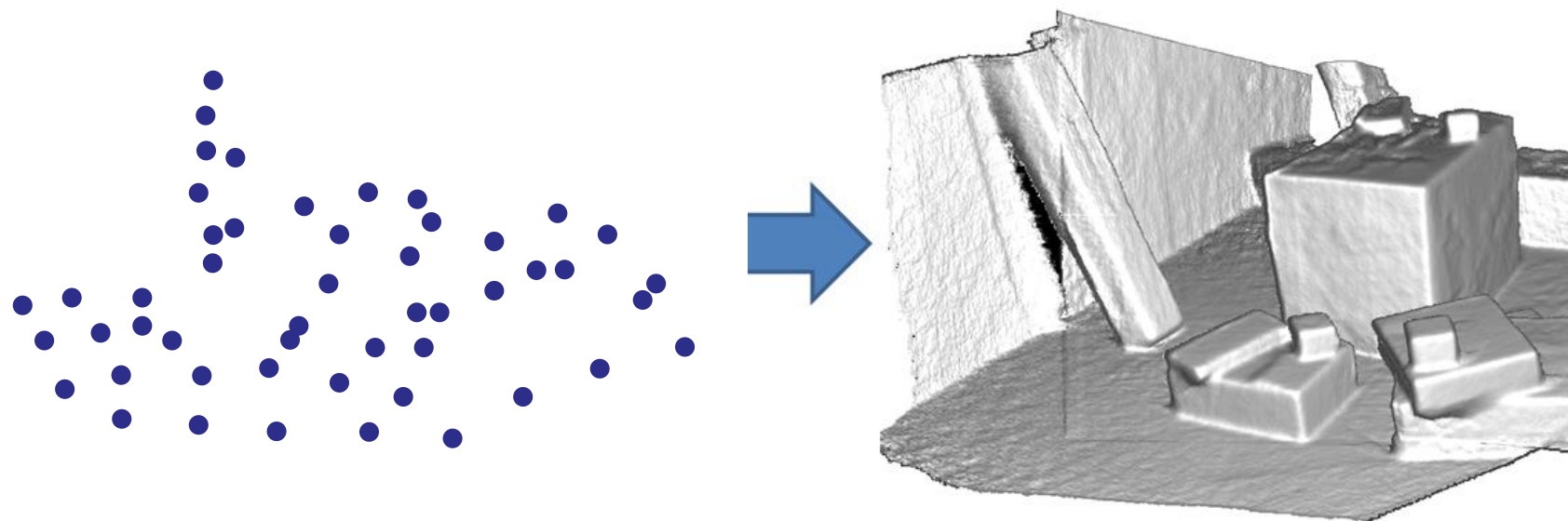
Course Topics

- Multi-view geometry and sparse 3D reconstruction



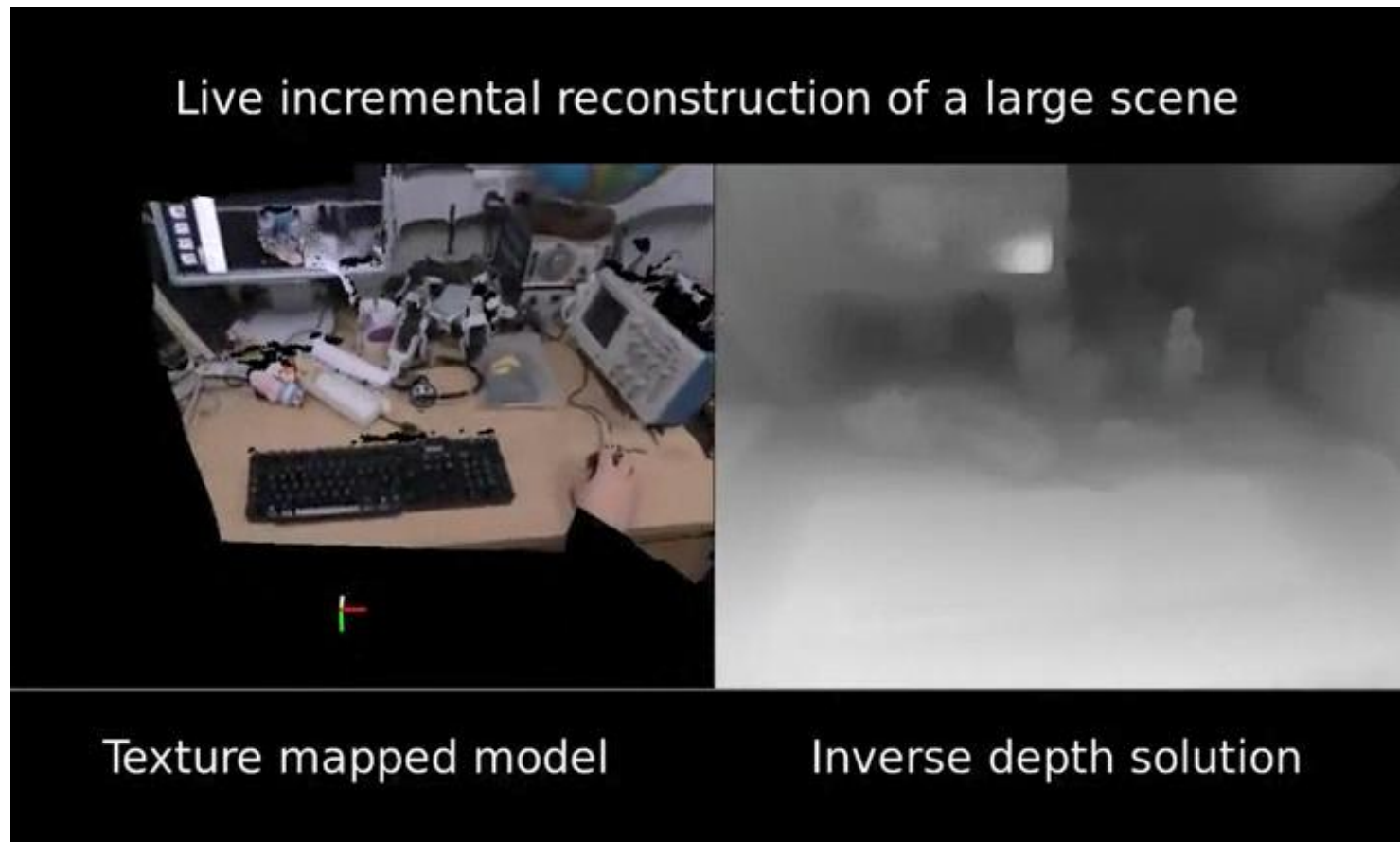
Course Topics

- Dense 3D reconstruction



Course Topics

- Dense 3D reconstruction

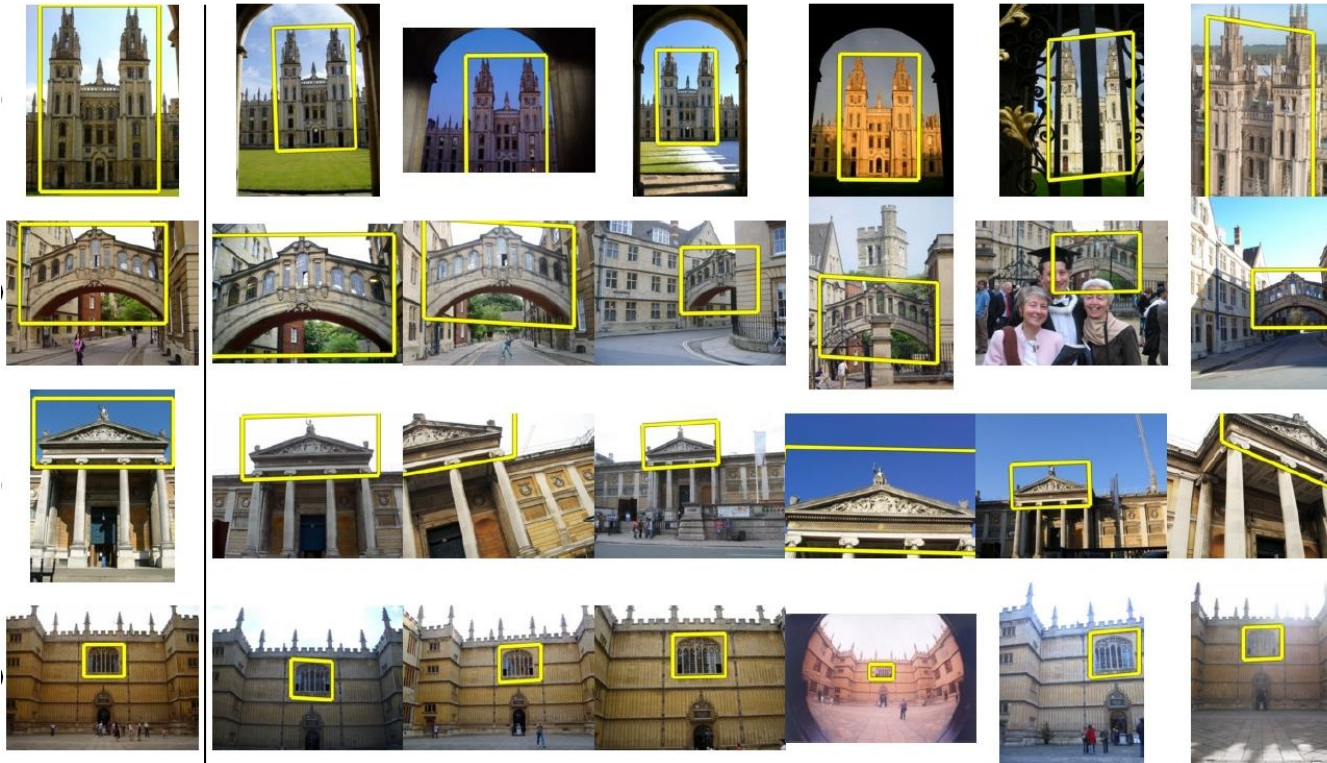


Course Topics

- Place recognition and deep learning

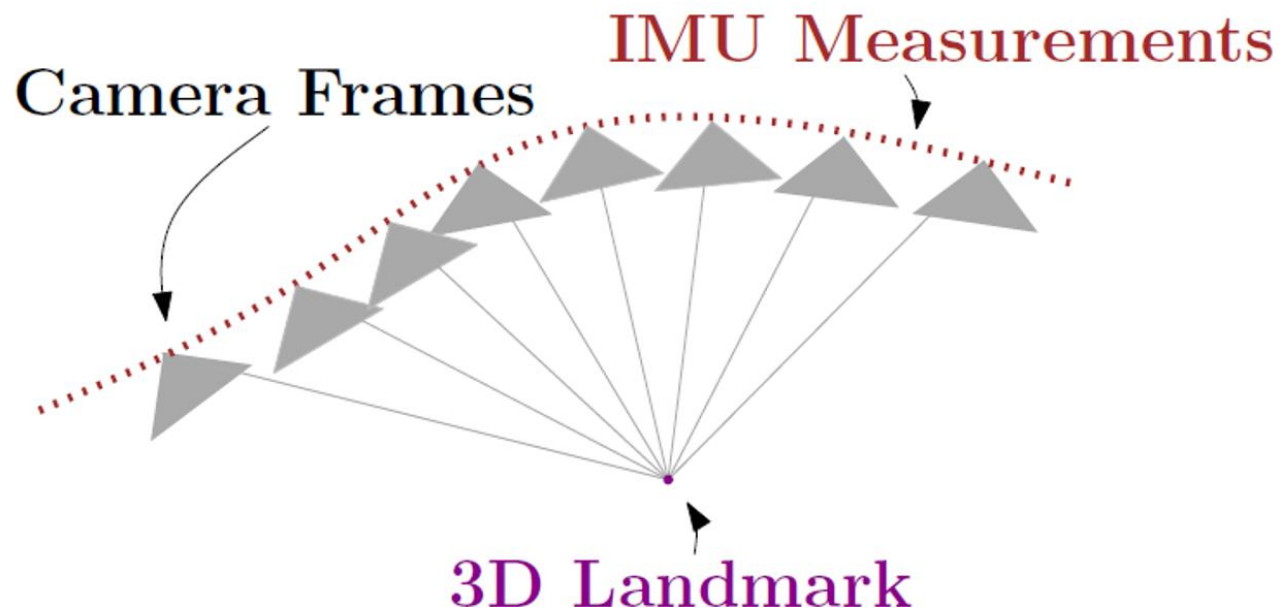
**Query
image**

Most similar places from a database of millions of images



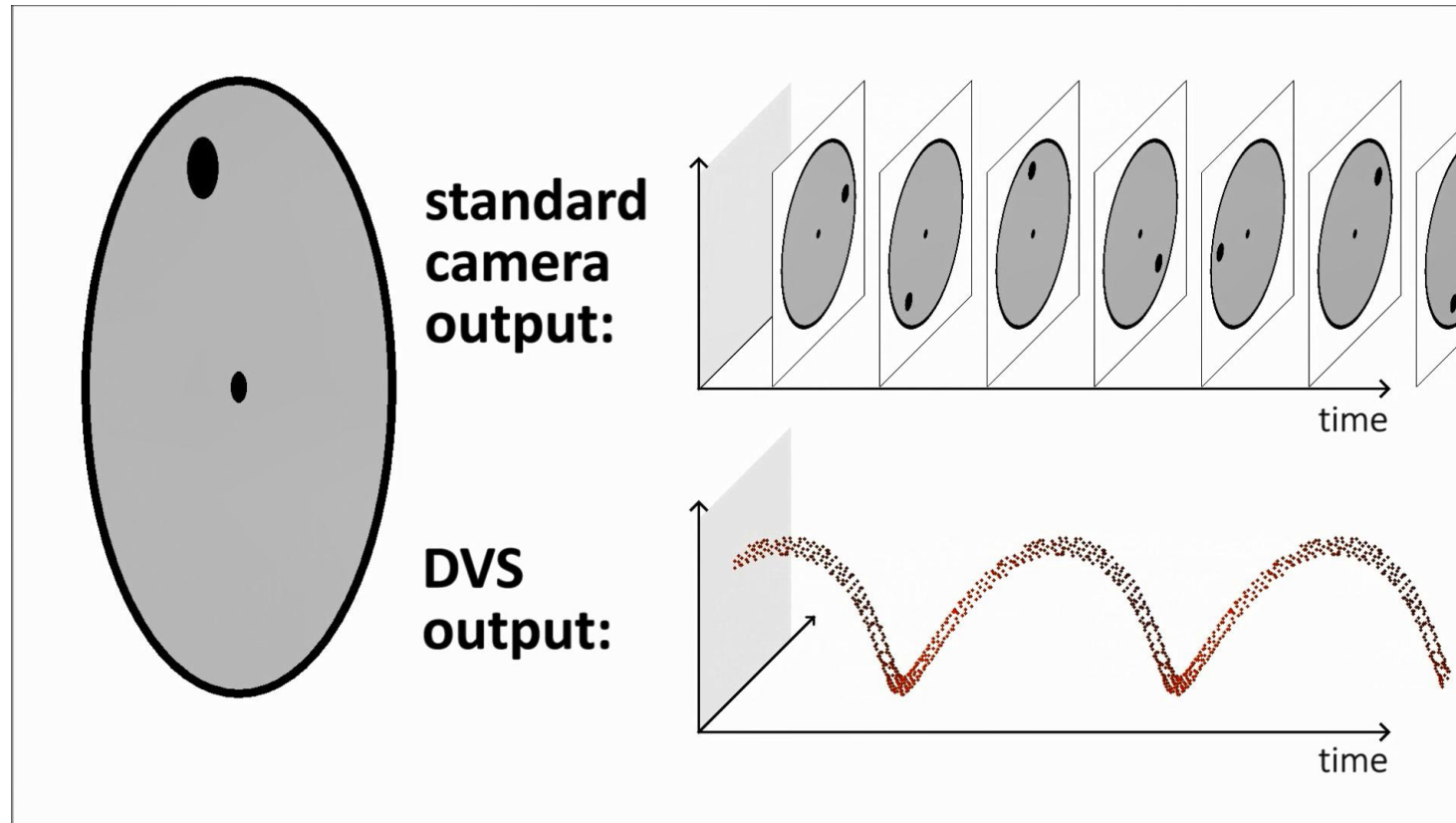
Course Topics

- Visual-inertial fusion

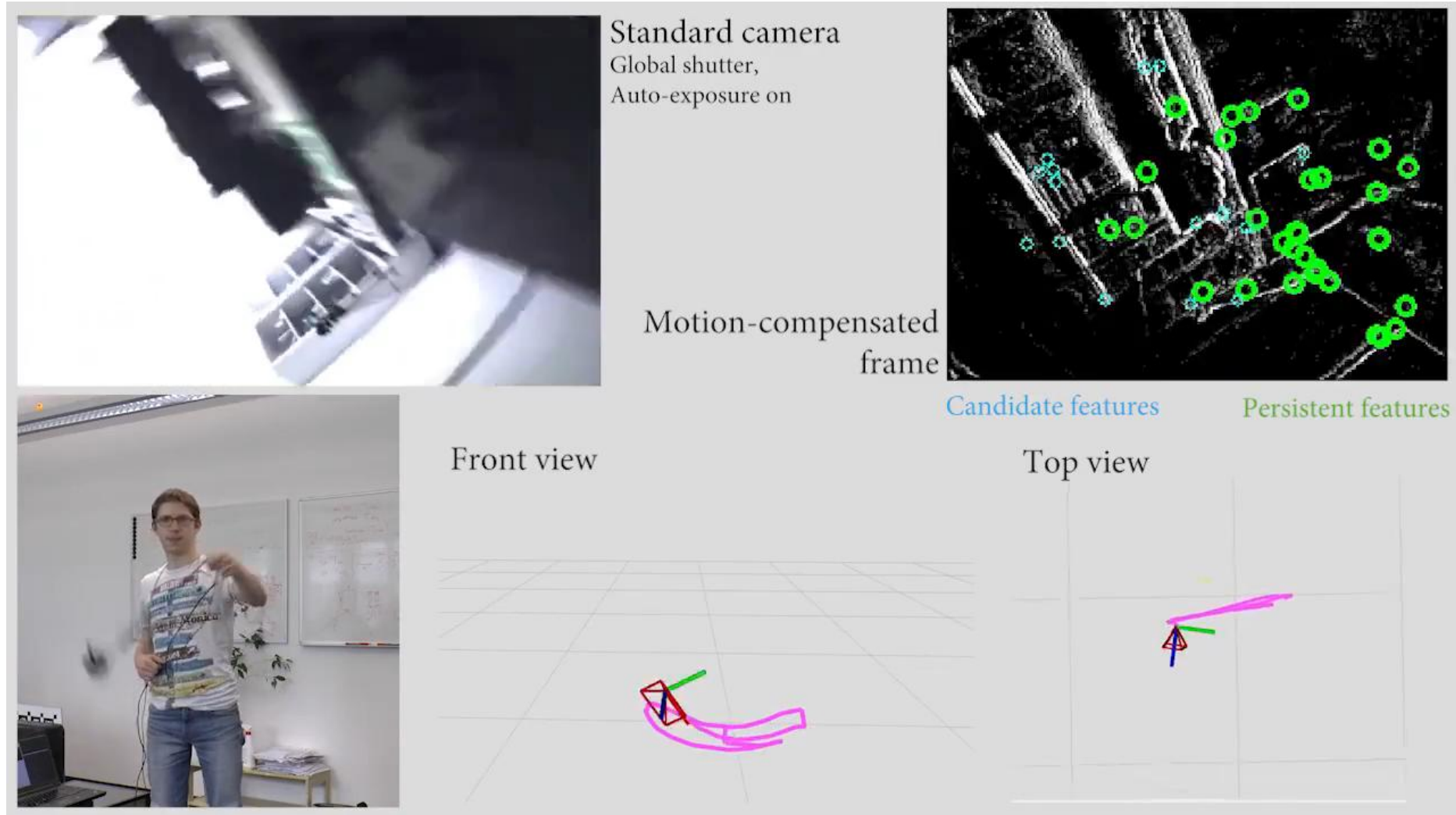


Course Topics

- Event cameras



Application of event cameras: high-speed VO



Reading

- 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](#)
- **Chapter 1 of SLAM Handbook:** [PDF](#)

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?