University of Zurich[UZH] | ETH zürich
Institute of Informatics – Institute of Neuroinformatics
ROBOTICS & PERCEPTION GROUP

Vision Algorithms for Mobile Robotics

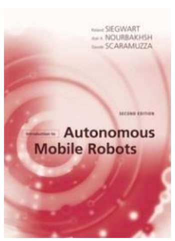
Lecture 01
Introduction

Davide Scaramuzza

Today’s Class

- About me
- What is Computer Vision?
- Example of Vision Applications
- Specifics of this course
- Introduction to Visual Odometry

Who am I?



Current positions

- ➢ Professor of Robotics, Dep. of Informatics 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

- ➢ Coordinator of the European project *sFly* on visual navigation of micro drones
  - ➢ Which introduced the PX4 autopilot and visual navigation of drones
- ➢ Book “Autonomous Mobile Robots,” 2011, MIT Press

Spinoffs & Tech Transfer

- **Zurich-Eye,** enabling machines to see, now **Facebook-Oculus Zurich**
- Former strategic advisor of **Dacuda**, now **Magic Leap Zurich**
- **Fotokite**, aerial filming made simple, incubated in my lab

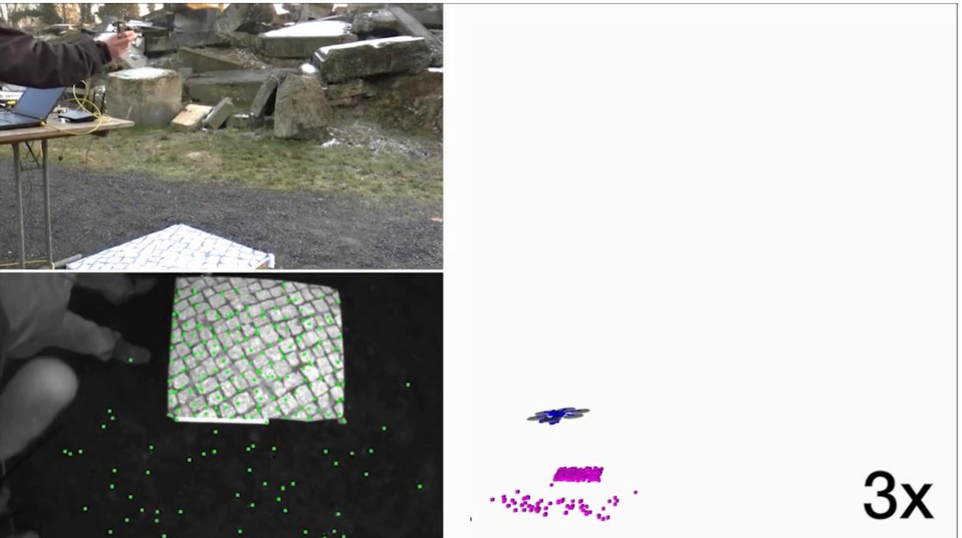
My Research Background

Computer Vision

- Visual Odometry and SLAM
- Sensor fusion
- Camera calibration

Autonomous Robot Navigation

- Self driving cars
- Micro Flying Robots



My lab

ROBOTICS &
PERCEPTION
GROUP

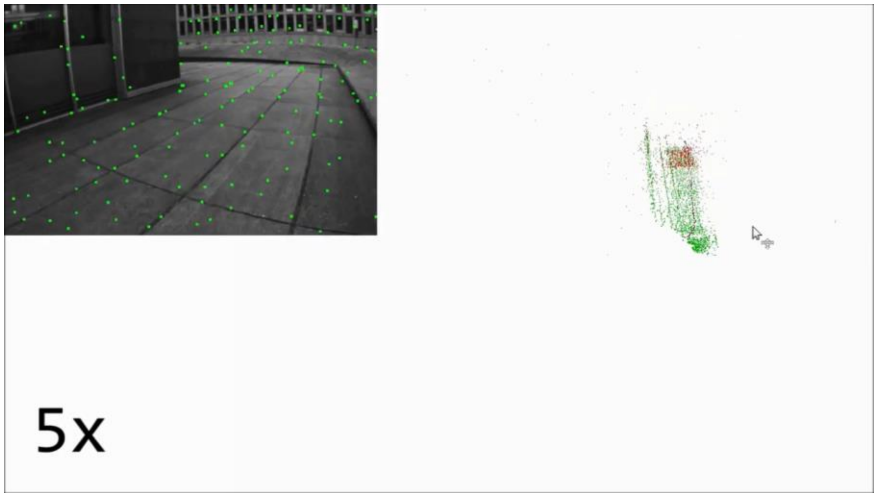
http://rpg.ifi.uzh.ch

Closed to bahnhof Oerlikon,
Andreasstrasse 15, 2nd floor

Our Research Areas

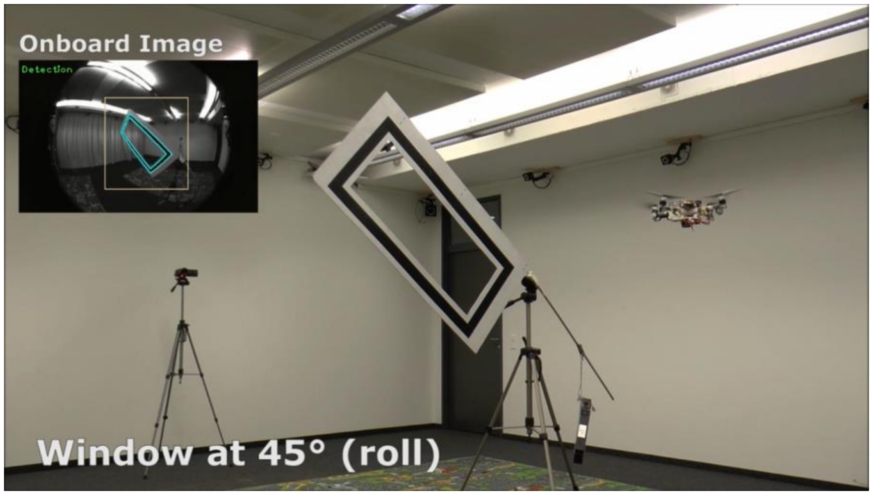
Visual-Inertial State Estimation

[IJCV'11, PAMI'13, RSS'15, TRO'16]



Vision-based Navigation of Flying Robots

[AURO'12, RAM'14, JFR'15]

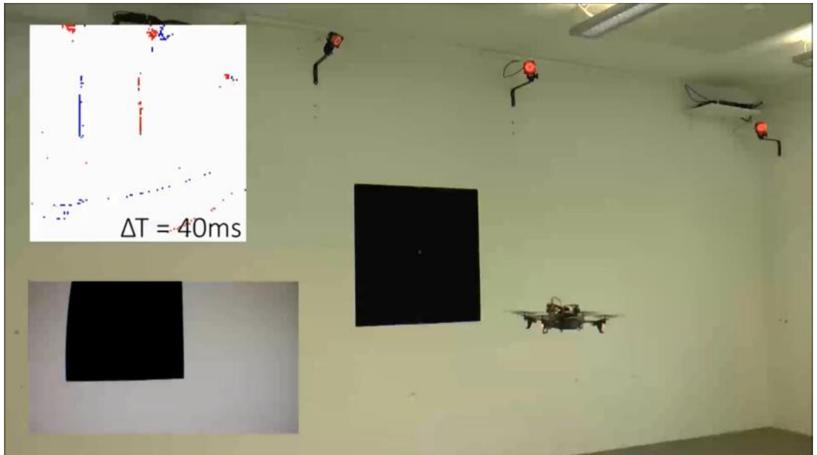


End-to-End Learning

[RAL'16-17]

Event-based Vision

[IROS'3, ICRA'14, RSS'15, PAMI'17]



Parrot: Autonomous Inspection of Bridges and Power Masts

Parrot senseFly

Albris drone

Dacuda 3D (now Magic Leap Zurich)

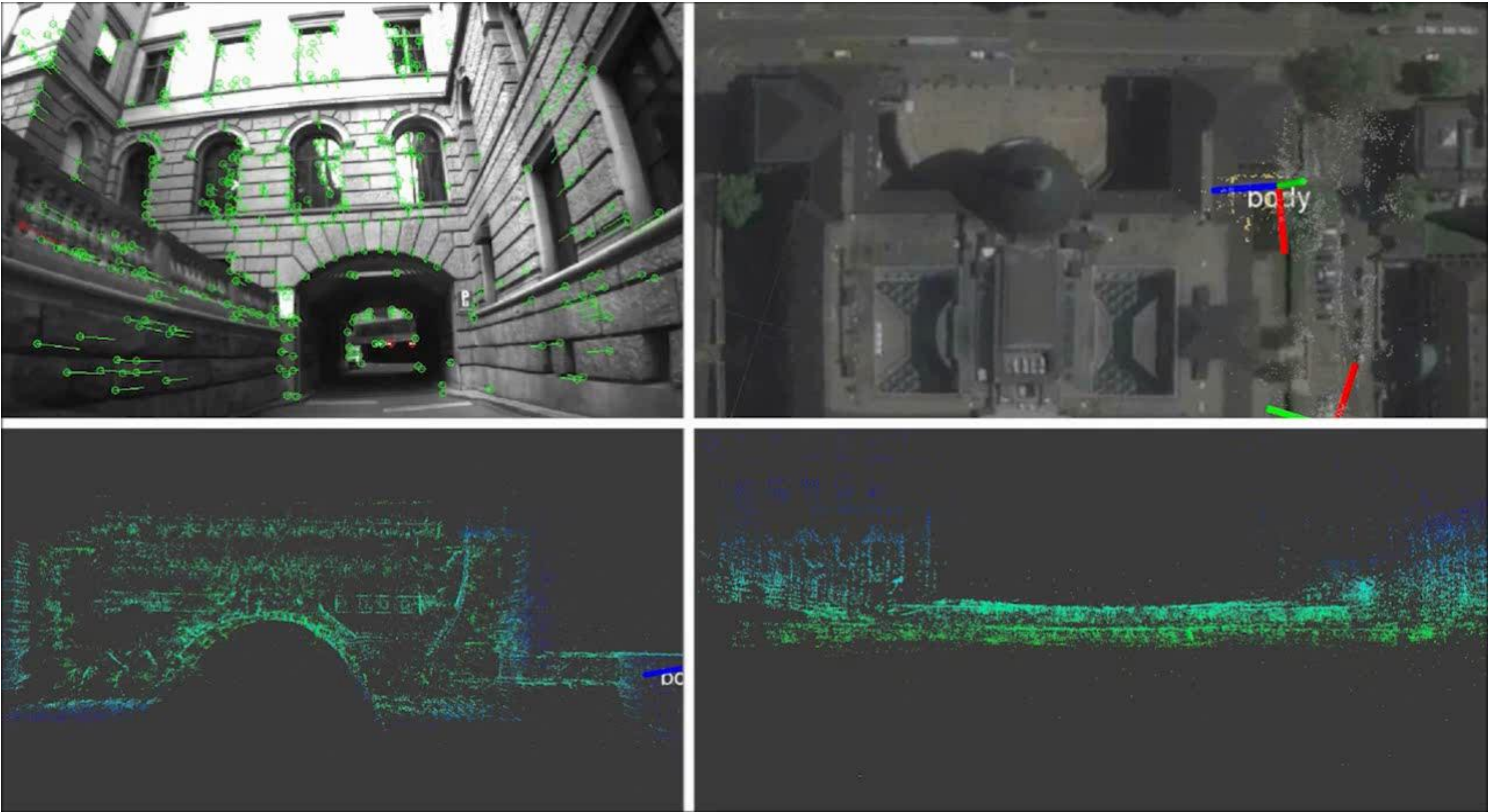
- Fully immersive VR (running on iPhone)
- Powered by SVO

Dacuda's 3D divison → magic leap

Zurich-Eye (now Oculus Zurich)

Vision-based Localization and Mapping Solutions for Mobile Robots

Launched in Sep. 2015, **became Facebook-Oculus Zurich in Sep. 2016**



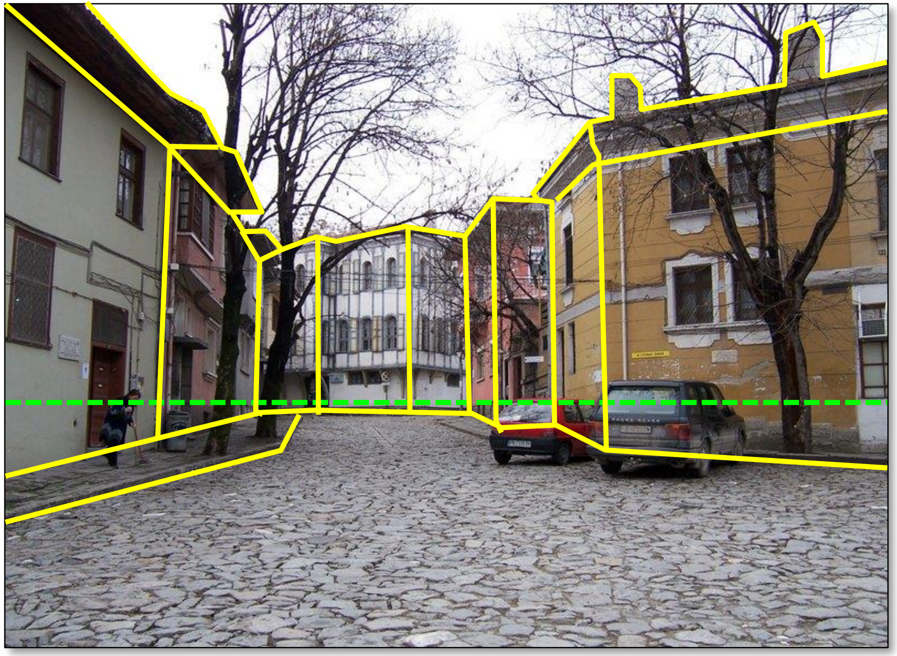
Today's Class

- What is Computer Vision?
- Example of Vision Applications
- Specifics of this course
- Overview of Visual Odometry

What is computer vision?

Automatic extraction of “meaningful” information from images and videos

Semantic information



Geometric information
(this course)

Vision Demo?

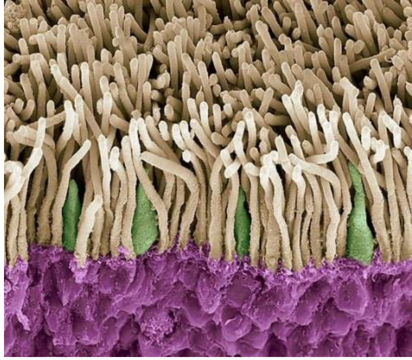
Terminator 2

We are almost there!

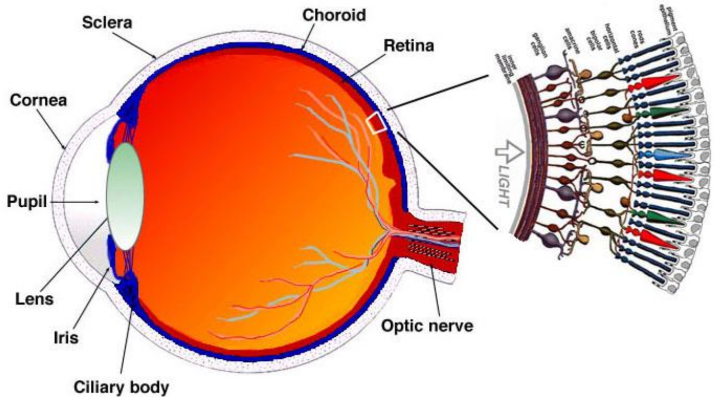
Why study computer vision?

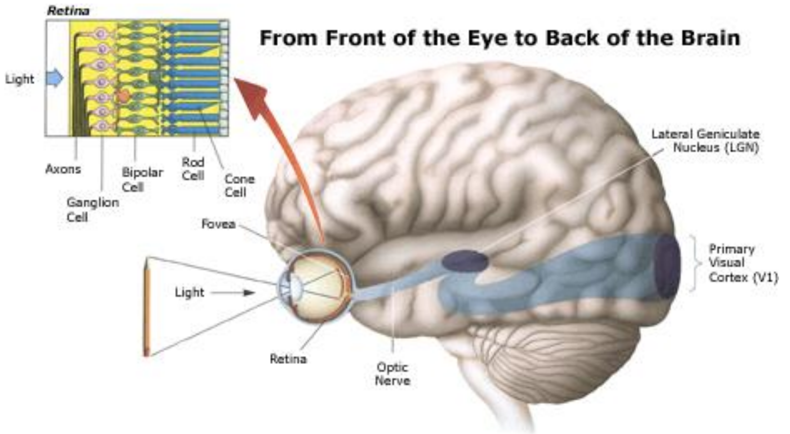
- Relieve humans of boring, easy tasks
- Enhance human abilities: human-computer interaction, visualization, augmented reality (AR)
- Perception for autonomous robots
- Organize and give access to visual content

Vision in humans



- **Vision** is our most powerful sense
- Retina is ~1000mm^2. Contains 130 million **photoreceptors** (120 mil. rods and 10 mil. cones for color sampling)
- Provides **enormous** amount of information: data-rate of **~3GBytes/s**
  - Half of primate cerebral cortex is devoted to visual processing!
- 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 **15-degree field of view** (around the fovea)!





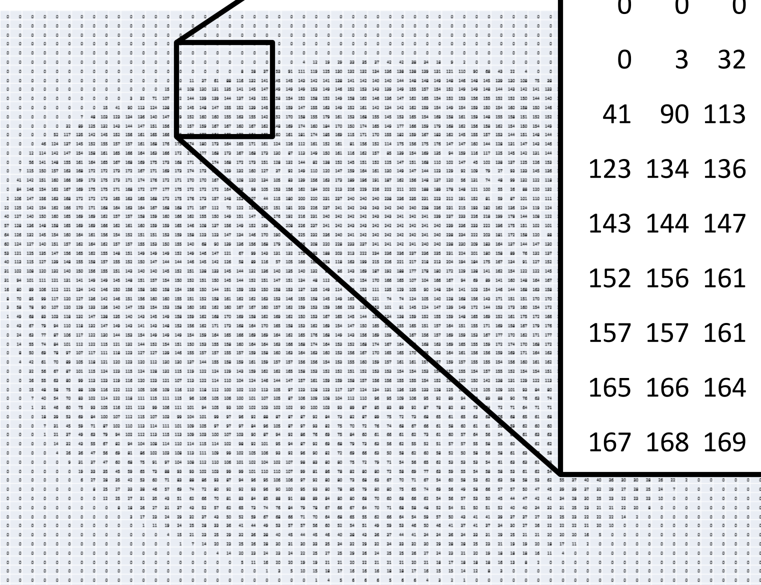
What A Baby Can See Every Month For The First Year Of Its Life

"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://www.iflscience.com/plants-and-animals/this-gif-shows-what-a-baby-can-see-every-month-for-the-first-year-of-its-life/

Why is vision hard?

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

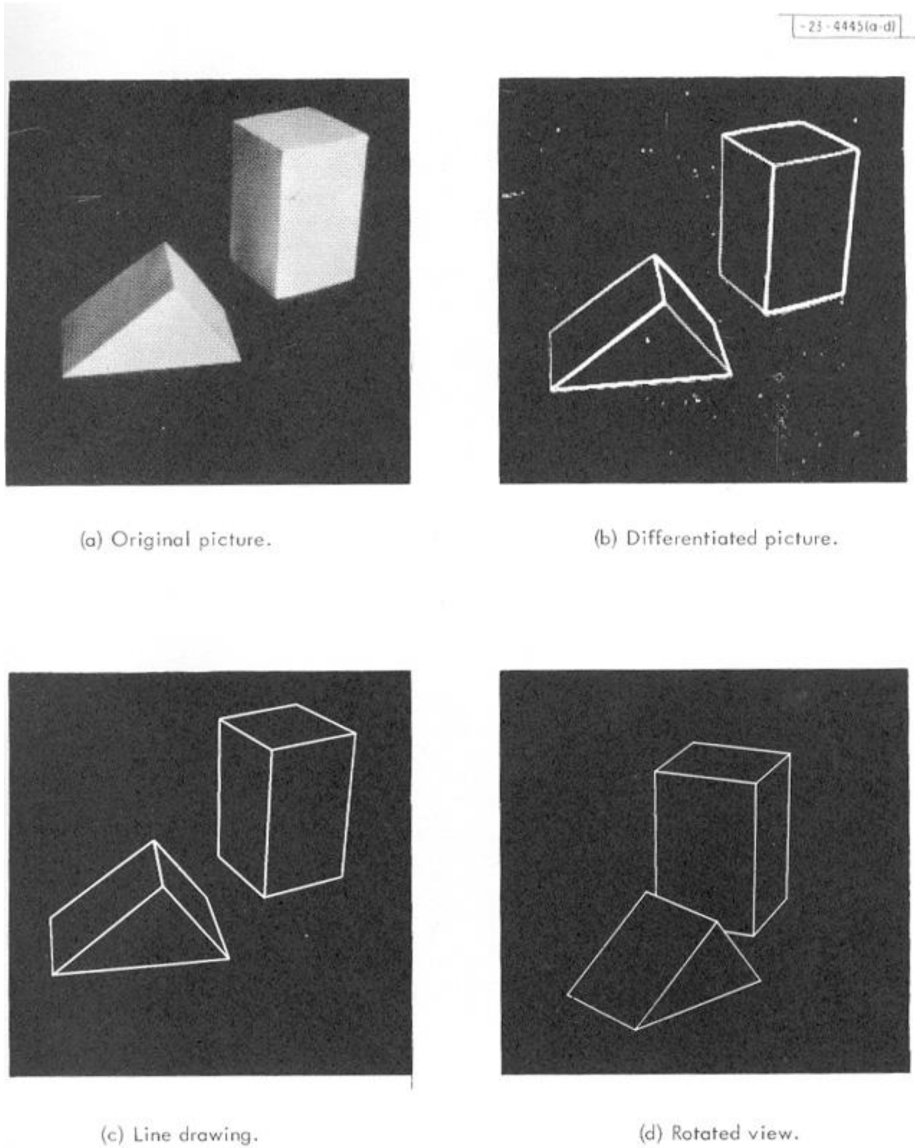


0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	8	28
0	0	0	0	0	0	11	37	61	88	116	132
0	0	0	0	15	64	108	130	131	135	141	145
0	3	32	71	107	132	144	139	139	144	137	143
41	90	113	124	138	140	145	148	147	155	152	139
123	134	136	140	147	149	152	160	160	155	163	155
143	144	147	151	156	160	157	159	167	167	160	167
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

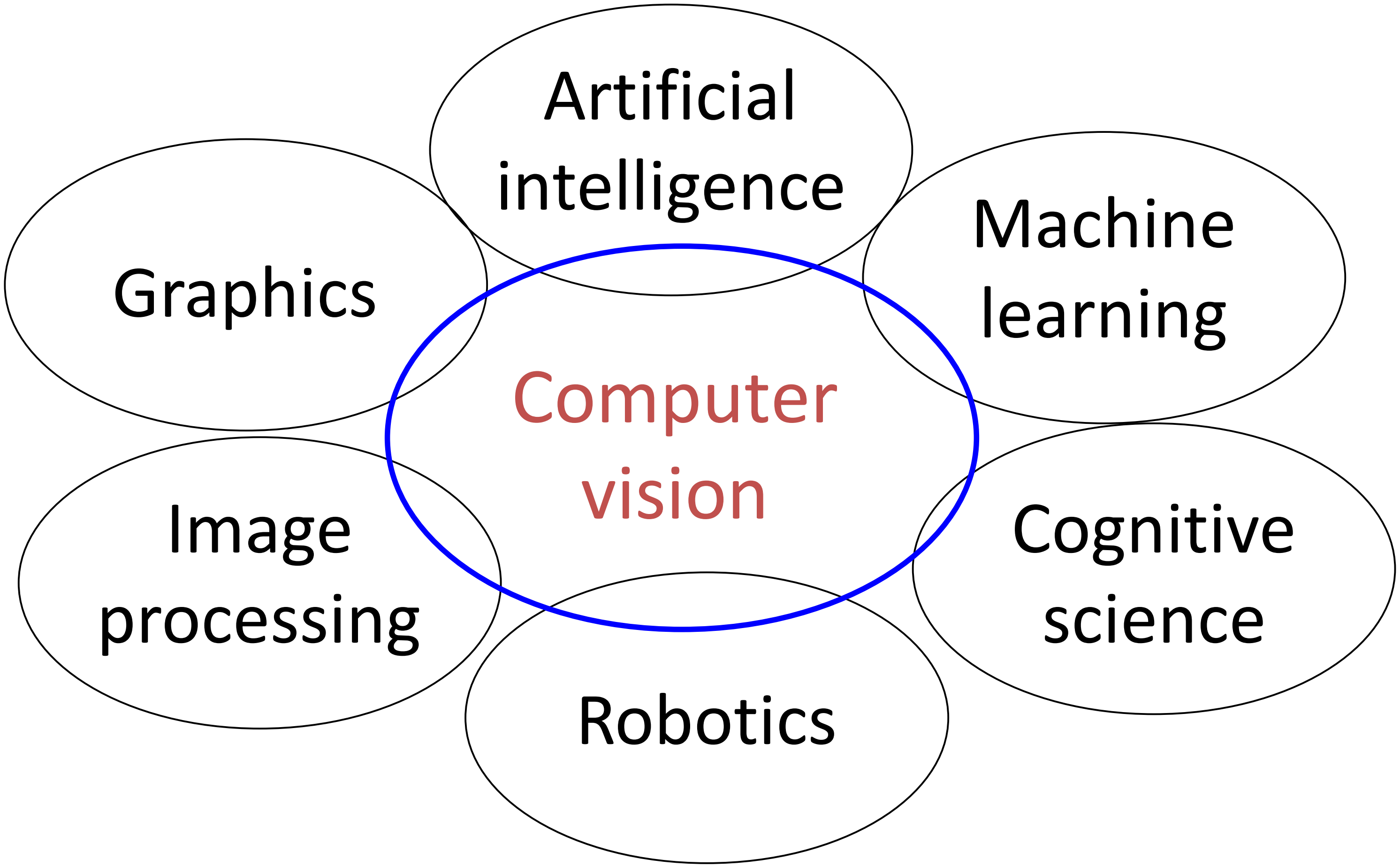
Origins of computer vision



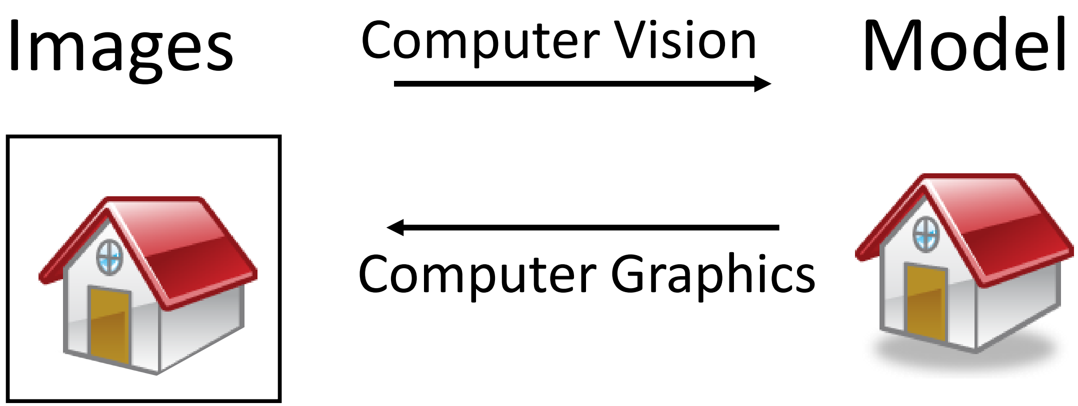
L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

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

Related disciplines



Computer Vision vs Computer Graphics



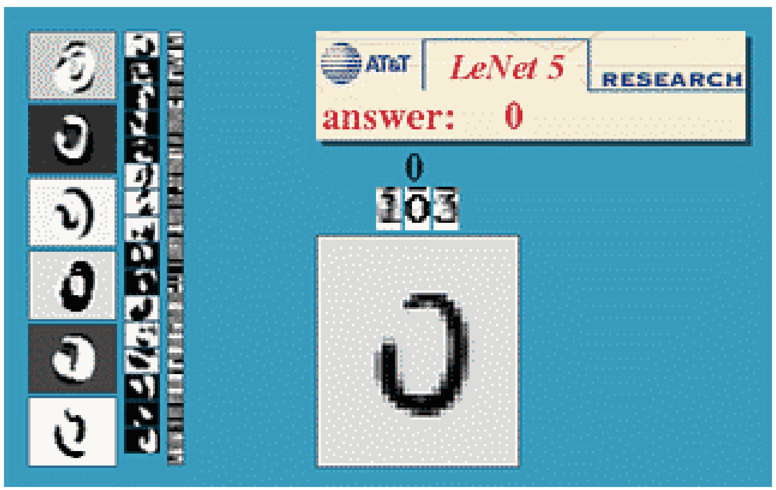
Inverse problems: analysis and synthesis.

Today's Class

- About me
- What is Computer Vision?
- Examples of Vision Applications
- Specifics of this course
- Image Formation

Optical character recognition (OCR)

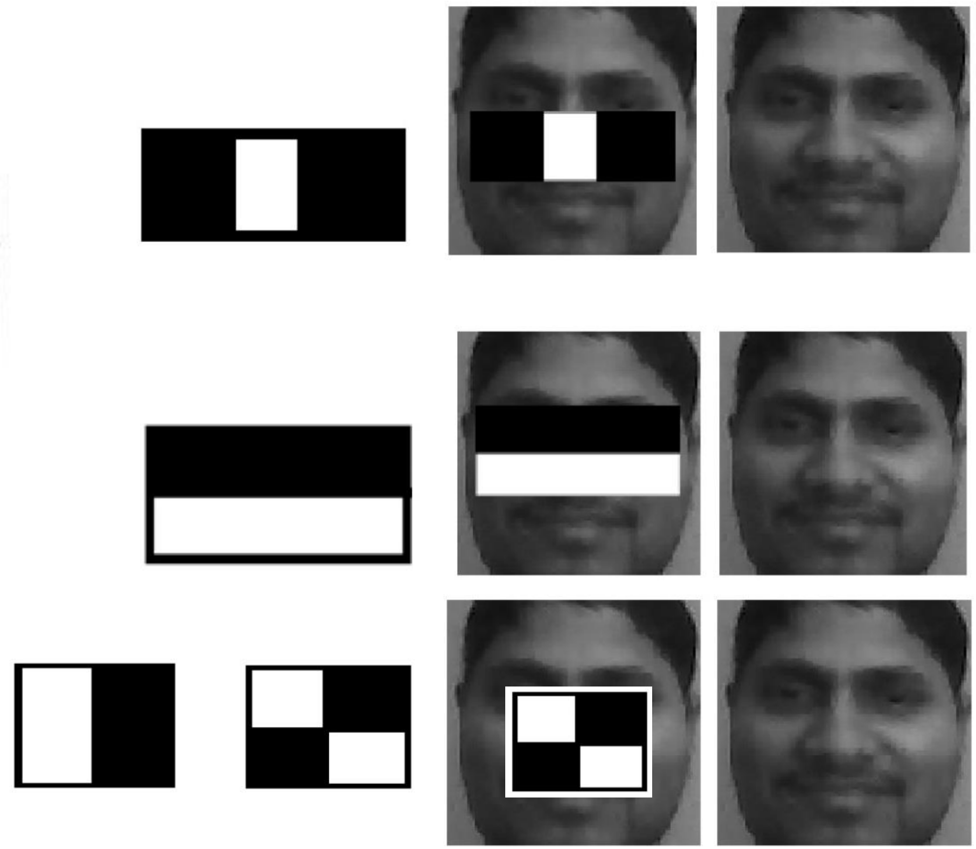
Technology to convert scanned docs to text



Digit recognition, AT&T labs, using CNN,
by Yann LeCun (1993),
now head of Deep Learning at Facebook
http://yann.lecun.com/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face detection



Now in all new digital cameras and smartphones

P. Viola, M. Jones: Robust Real-time Object Detection, Int. Journal of Computer Vision 2001
(NB. Paul Viola is now Vice President of Amazon Prime Air)

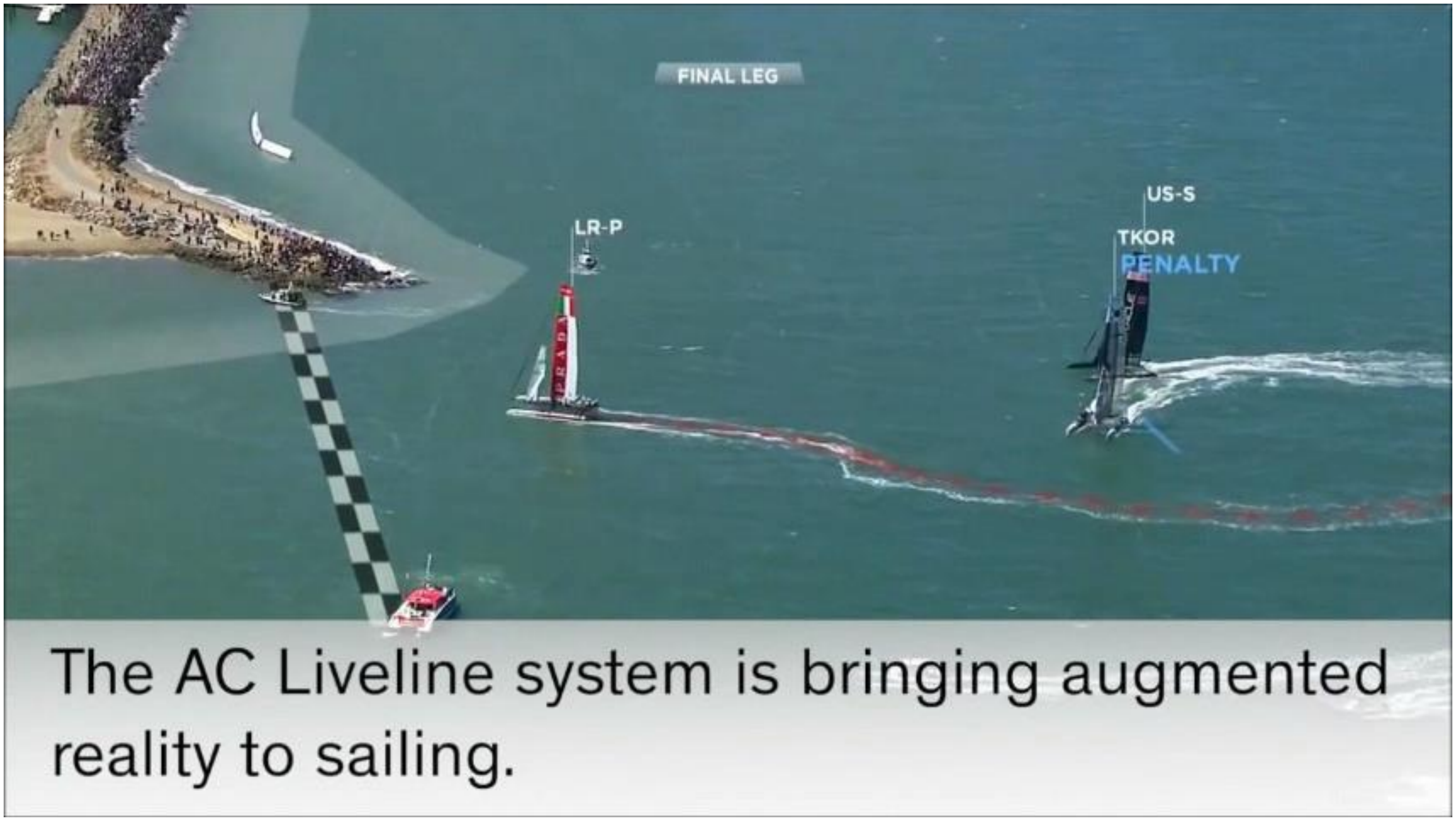
Object recognition (in mobile phones)

- This is becoming real:
  - Lincoln Microsoft Research
  - Point & Find, Nokia
  - SnapTell.com (Amazon)
  - Google Goggles

Special effects: shape and motion capture

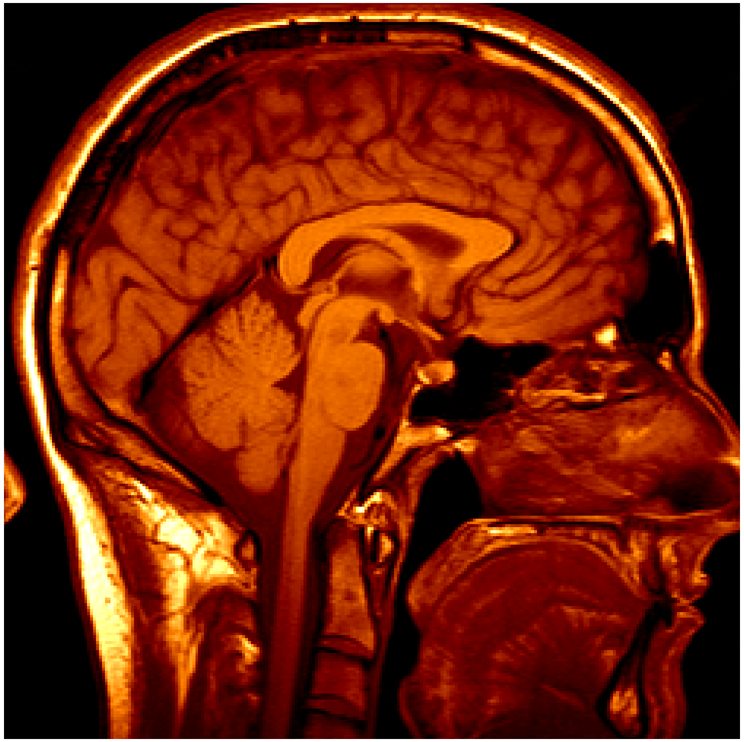
Sports

- Augmented Reality



2013 America's Cup

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

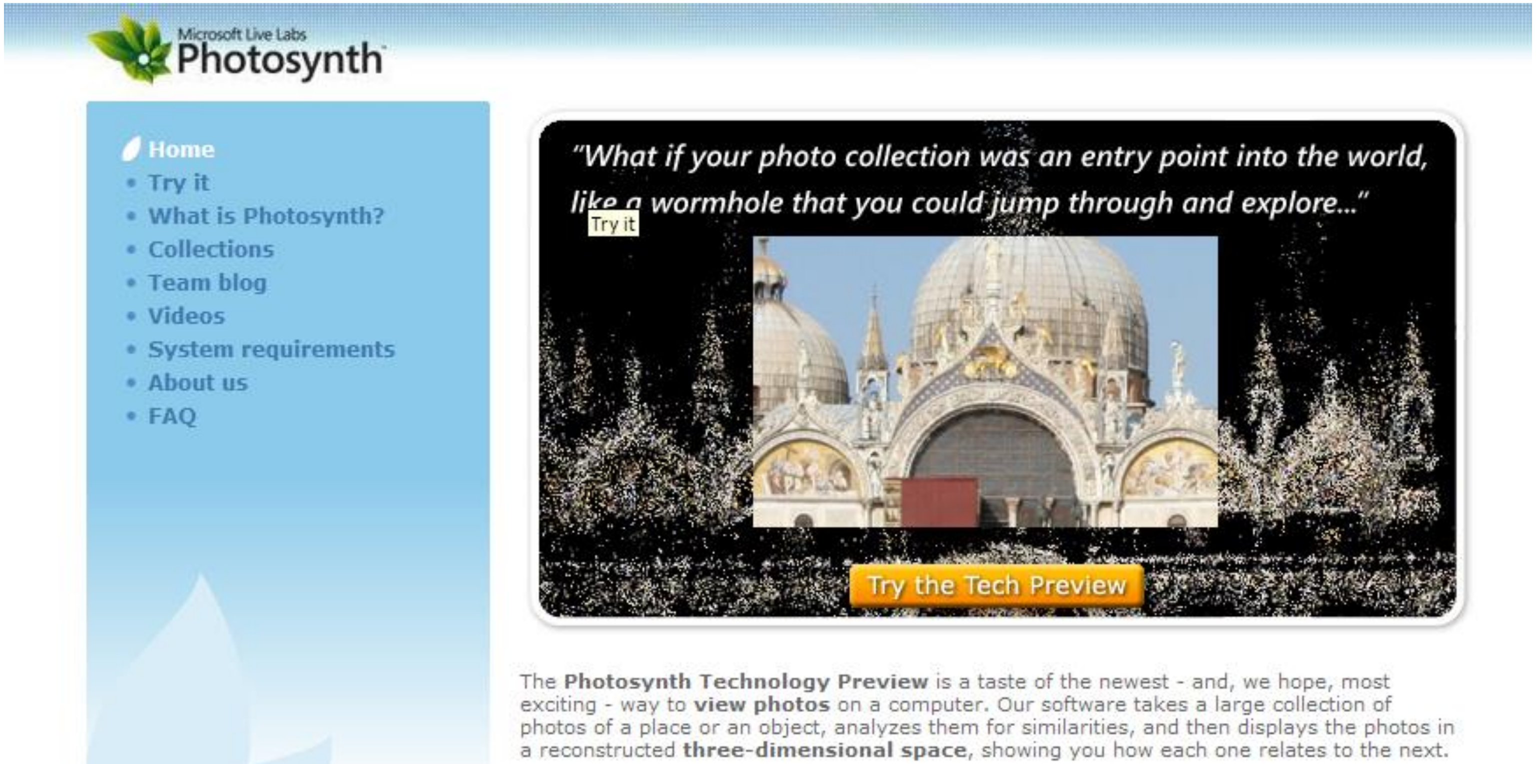
3D Reconstruction by Multi-View Stereo

[YouTube Video]

3D Reconstruction: Multi-View Stereo

YouTube Video

Microsoft Photosynth



http://labs.live.com/photosynth/

Based on Photo Tourism technology developed by Noah Snavely, Steve Seitz, and Rick Szeliski

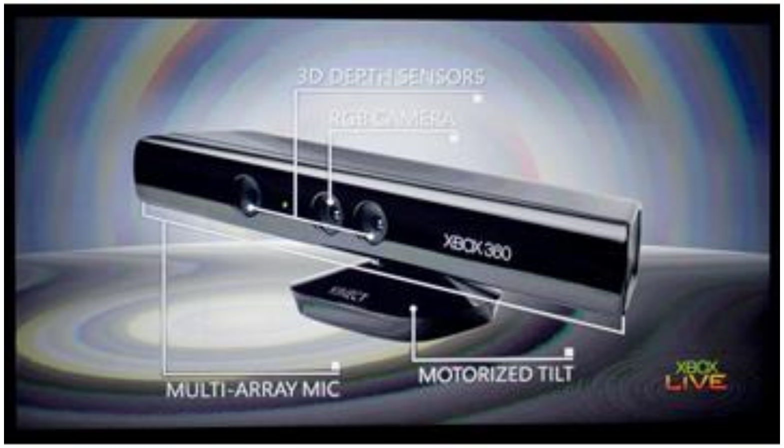
Pix4D

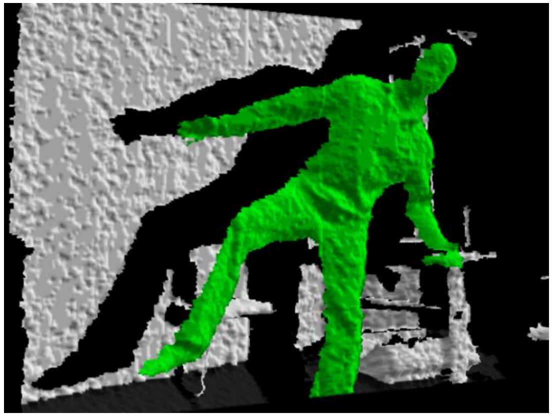
- EPFL startup – Now a company

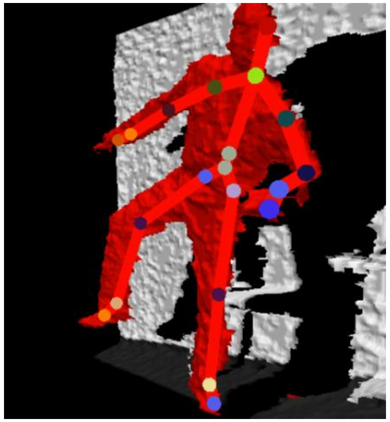
Automotive safety

- Mobileye: Vision systems in high-end Tesla, BMW, GM, Volvo models. Bought by **Intel in 2017 for 15 billion USD**!
  - Pedestrian collision warning
  - Forward collision warning
  - Lane departure warning
  - Headway monitoring and warning

Vision-based interaction: Xbox Kinect

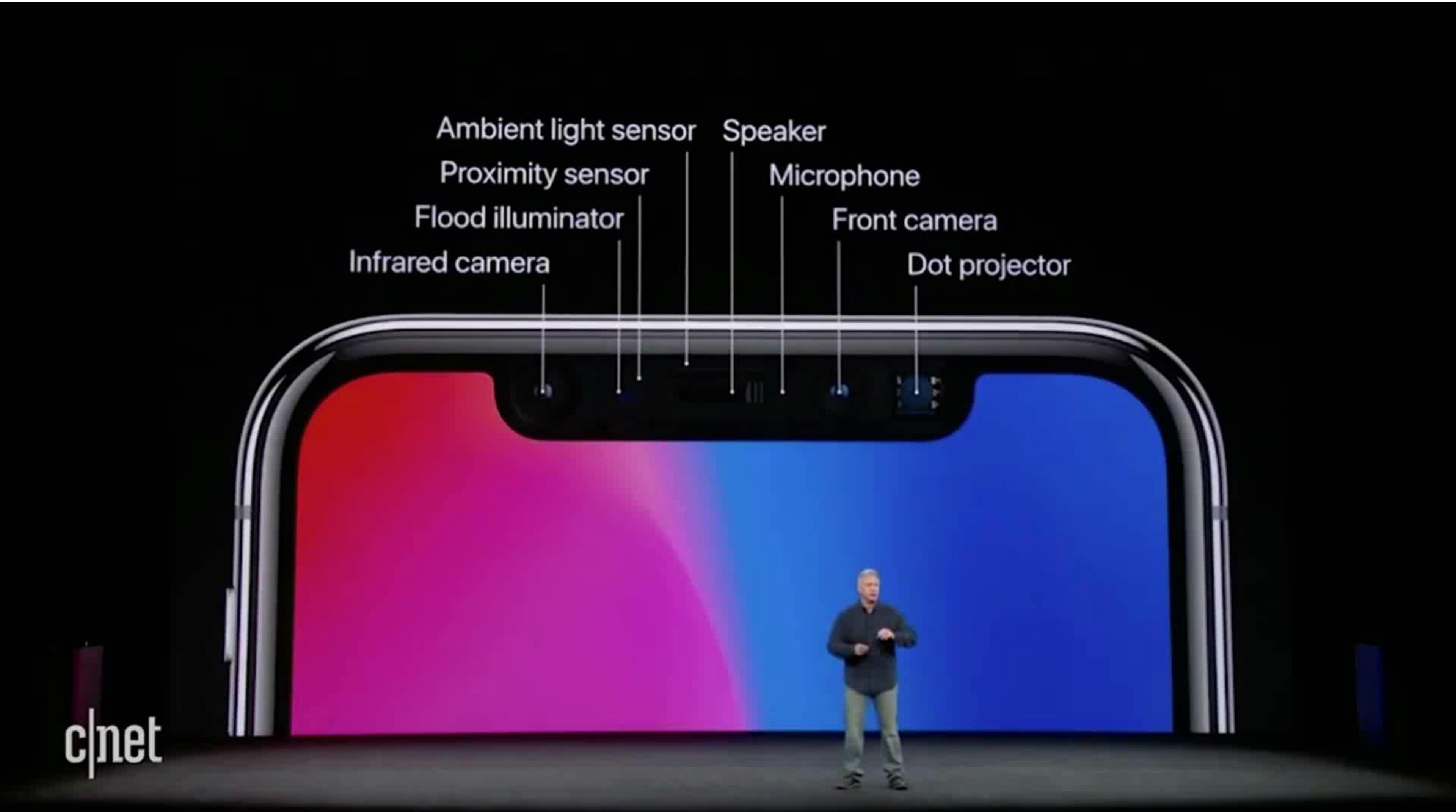






Lot of Computer Vision in Modern Smartphones

iPhone X



Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (made by JPL) used for several tasks

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

Vision-based Autonomous Drone Navigation

Works in GPS-denied Environments (EU project SFLY)

[Scaramuzza et al., Vision-Controlled Micro Flying Robots: from System Design to Autonomous Navigation and Mapping in GPS-denied Environments, IEEE RAM, September, 2014

Dacuda's mouse scanner

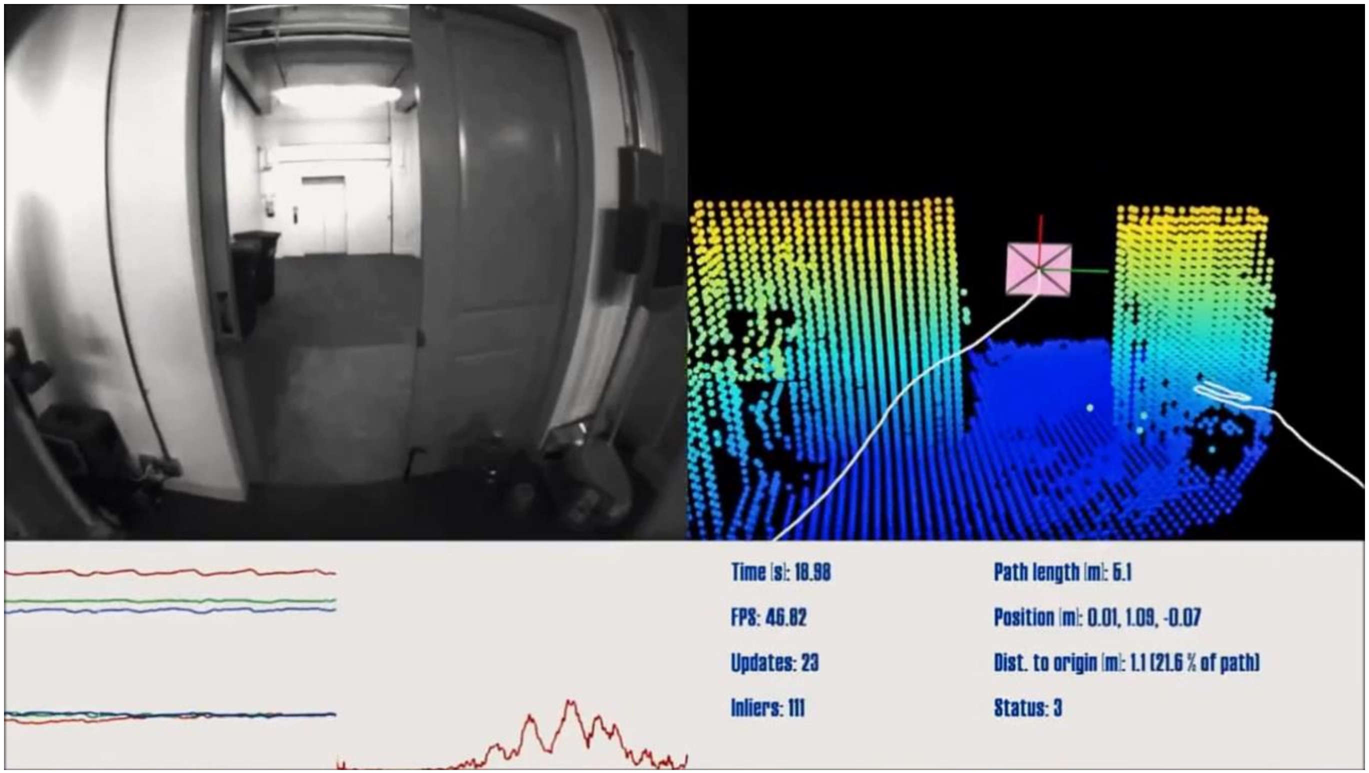
Dacuda

- World's first mouse scanner,
  Distributed by LG, Logitech, etc.
- Dacuda was bought by Magic Leap in 2017 and is now Magic Leap Zurich (focusing on Augmented Reality)

LG

Microsoft HoloLens

Google Tango



Project Tango

Augmented Reality with Google Tango and Apple ARKit

Current state of the art

- These were just few examples of current systems
  - Many of these are less than 5 years old
- Computer Vision is a very active field of research, and rapidly changing
  - Many new applications and phone apps in the next few years
- To learn more about vision applications and companies
  - David Lowe maintained an excellent overview of vision companies until 2015
    - http://www.cs.ubc.ca/spider/lowe/vision.html

Let’s have a 10 minute break with Google Tango and Microsoft Hololens Demos

Today's Class

- About me
- What is Computer Vision?
- Example of Vision Applications
- Specifics of this course
- Overview of Visual Odometry

Organization of this Course

- **Lectures**:
  - 10:15 to 12:00 every week
  - Room: ETH LFW C5, Universitätstrasse 2

- **Exercises:**
  - 13:15 to 15:00: Starting from the 3rd week. Then almost every week.
  - Room: ETH HG E 1.1

- **Official course website: http://rpg.ifi.uzh.ch/teaching.html**
  - Check it out for the PDFs of the lecture slides and updates

Learning Objectives

- **High-level goal:** learn to implement current visual odometry pipelines used in mobile robots (drones, cars, Mars rovers), and Virtual-reality (VR) and Augmented reality (AR) products: e.g., Google Tango, Microsoft HoloLens
- You will also learn **to implement the fundamental computer vision algorithms** used in mobile robotics, in particular: feature extraction, multiple view geometry, dense reconstruction, object tracking, image retrieval, visual-inertial fusion, event-based vision.
- In order to learn these competences, **participation in the exercise sessions is critical although not mandatory**!

Course Schedule

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

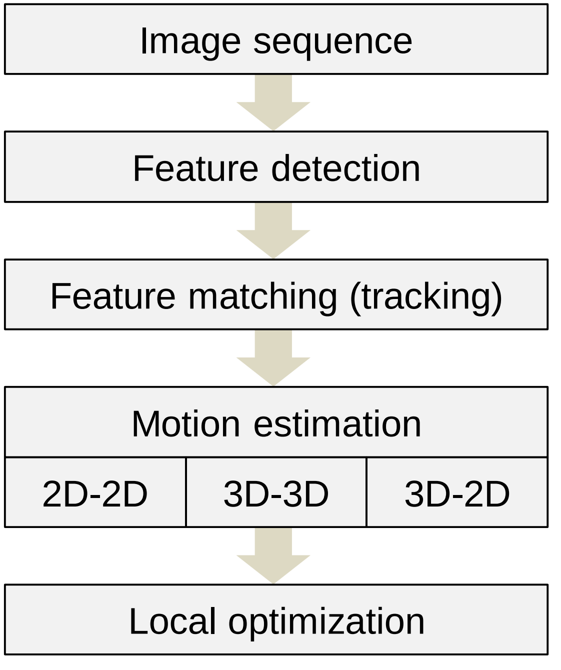
Date	Time	Description of the lecture/exercise	Lecturer
21.09.2017	10:15 - 12:00	01 – Introduction	Davide Scaramuzza
28.09.2017	10:15 - 12:00	02 - Image Formation 1: perspective projection and camera models	**Guillermo Gallego**
05.10.2017	10:15 - 12:00	03 - Image Formation 2: camera calibration algorithms	Davide Scaramuzza
	13:15 – 15:00	Exercise 1: Augmented reality wireframe cube	T. Cieslewski/H. Rebecq/A. Loquercio
12.10.2017	10:15 - 12:00	04 - Filtering & Edge detection	Davide Scaramuzza
	13:15 – 15:00	Exercise 2: PnP problem	T. Cieslewski/H. Rebecq/A. Loquercio
19.10.2017	10:15 - 12:00	05 - Point Feature Detectors 1: Harris detector	Davide Scaramuzza
	13:15 – 15:00	Exercise 3: Harris detector + descriptor + matching	T. Cieslewski/H. Rebecq/A. Loquercio
26.10.2017	10:15 - 12:00	06 - Point Feature Detectors 2: SIFT, BRIEF, BRISK	Davide Scaramuzza
02.11.2017	10:15 - 12:00	07 - Multiple-view geometry 1	**Guillermo Gallego**
	13:15 – 15:00	Exercise 4: Stereo vision: rectification, epipolar matching, disparity, triangulation	T. Cieslewski/H. Rebecq/A. Loquercio
09.11.2017	10:15 - 12:00	08 - Multiple-view geometry 2	Davide Scaramuzza
	13:15 – 15:00	Exercise 5: Eight-point algorithm and RANSAC	T. Cieslewski/H. Rebecq/A. Loquercio
16.11.2017	10:15 - 12:00	09 - Multiple-view geometry 3	Davide Scaramuzza
	13:15 – 15:00	Exercise 6: P3P algorithm and RANSAC	T. Cieslewski/H. Rebecq/A. Loquercio
23.11.2017	10:15 - 12:00	10 - Dense 3D Reconstruction (Multi-view Stereo)	Davide Scaramuzza
	13:15 – 15:00	Exercise 7: Intermediate VO Integration	T. Cieslewski/H. Rebecq/A. Loquercio
30.11.2017	10:15 - 12:00	11 - Optical Flow and Tracking (Lucas-Kanade)	Davide Scaramuzza
	13:15 – 15:00	Exercise 8: Lucas-Kanade tracker	T. Cieslewski/H. Rebecq/A. Loquercio
07.12.2017	10:15 - 12:00	12 – Place recognition	Davide Scaramuzza
	13:15 – 15:00	Exercise 9: Recognition with Bag of Words	T. Cieslewski/H. Rebecq/A. Loquercio
	10:15 - 12:00	13 – Visual inertial fusion	Davide Scaramuzza
14.12.2017	13:15 – 15:00	Exercise 10: Pose graph optimization and Bundle adjustment	T. Cieslewski/H. Rebecq/A. Loquercio
21.12.2017	10:15 - 12:00	14 - Event based vision + **lab visit and live demonstrations**	Davide Scaramuzza
	13:15 – 15:00	Exercise 11: final VO integration	T. Cieslewski/H. Rebecq/A. Loquercio

Exercises

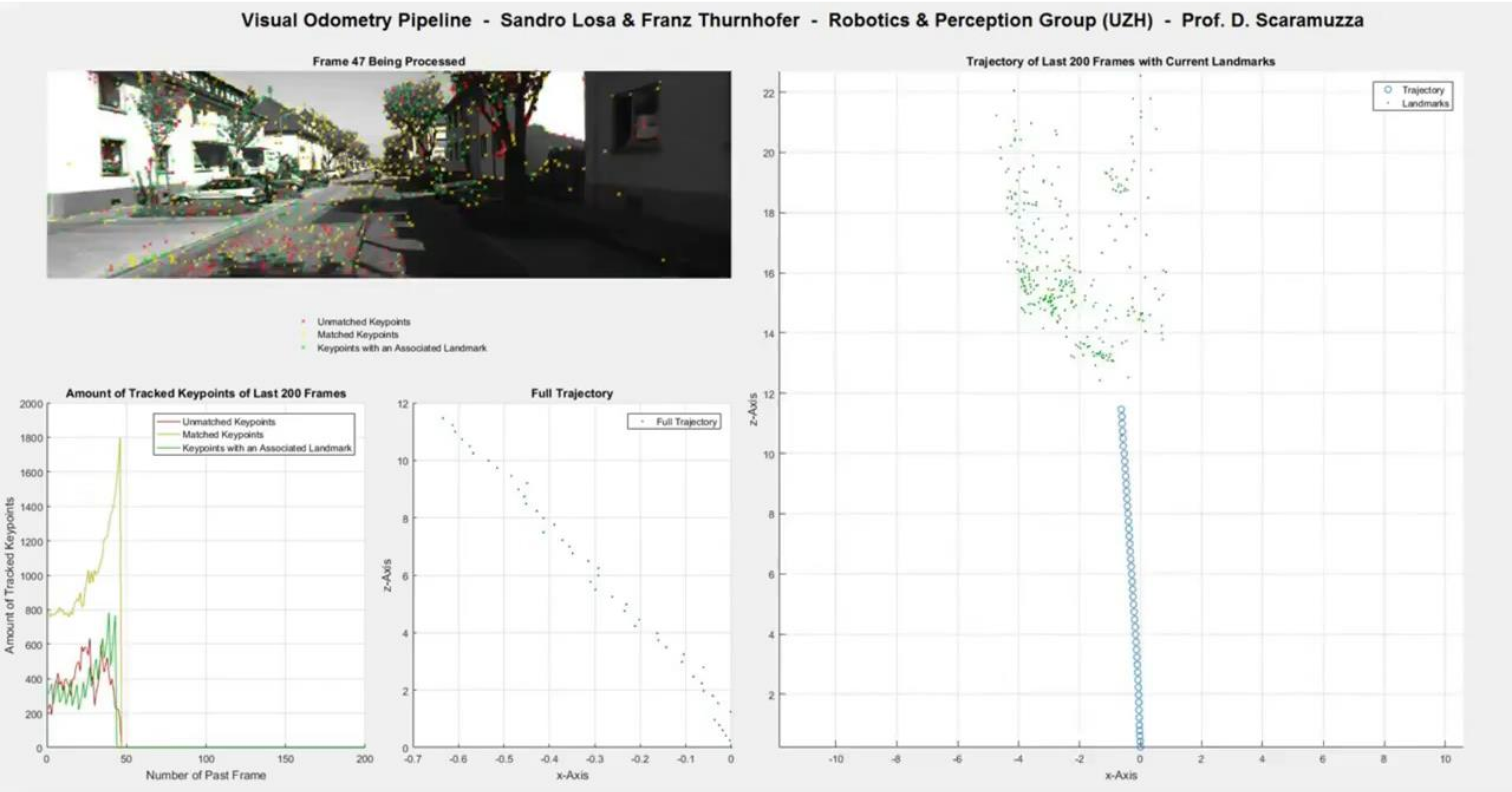
- Almost every week starting from the 3rd week (check out course schedule)
- Bring **your own laptop**
- Each exercise will consist of coding a building block of a visual odometry pipeline. At the end of the course there will be one additional exercise dedicated to assembling all the blocks together into a full pipeline.
- Have **Matlab** pre-installed!
  - **ETH**
    - Download: https://idesnx.ethz.ch/
  - **UZH**
    - Download: http://www.id.uzh.ch/dl/sw/angebote_4.html
    - Info on how to setup the license can be found there.
  - An introductory tutorial on Matlab can be found here: http://rpg.ifi.uzh.ch/docs/teaching/2017/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 and on current AR/VR devices (but actually much better ☺)).
- **Each week** you will learn how to implement a **building block** of visual odometry. The building blocks are:

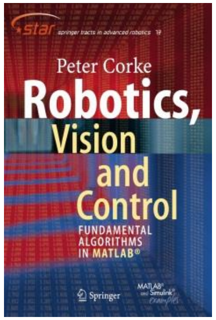


Outcome of last year exercises

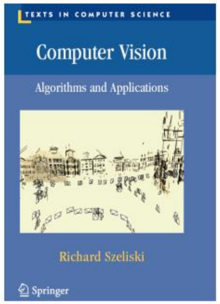


Recommended Textbook

- **Robotics, Vision and Control: Fundamental Algorithms**, by Peter Corke 2011. The PDF of the book can be freely downloaded (only with ETH VPN) from Springer or alternatively from Library Genesys



- **Computer Vision: Algorithms and Applications**, by Richard Szeliski, 2009. Can be freely downloaded from the author webpage: http://szeliski.org/Book/



- Other books:
  - *An Invitation to 3D Vision*: Y. Ma, S. Soatto, J. Kosecka, S.S. Sastry
  - *Multiple view Geometry*: R. Hartley and A. Zisserman

Instructors

- Lecturer
  - Davide Scaramuzza: sdavide (at) ifi (dot) uzh (dot) ch
  - Receving hours: Thursday afternoon (announce yourself by email)

- Exercises

Henri Rebecq
rebecq (at) ifi (dot) uzh (dot) ch

Titus Cieslewski
titus (at) ifi (dot) uzh (dot) ch

Antonio Loquercio
loquercio (at) ifi (dot) uzh (dot) ch

Prerequisites

- Linear algebra
- Matrix calculus
- No prior knowledge of computer vision and image processing required

Grading and Exam

- The **final grade is based on the oral exam** (30 minutes)
- In addition, **strong class participation can offset negative performance** at the oral exam.
- Optional mini project: you have the **option** (not mandatory) to do a **mini project**, which consists of implementing a working visual odometry algorithm in Matlab. If the algorithm runs properly 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 quite time consuming! The deadline to hand the mini project is 07.01.2018. Group work (up to 4) possible.

Class Participation

- Class participation includes
  - showing up
  - being able to articulate key points from last lecture

Today's Class

- About me
- What is Computer Vision?
- Example of Vision Applications
- Specifics of this course
- Overview of Visual Odometry

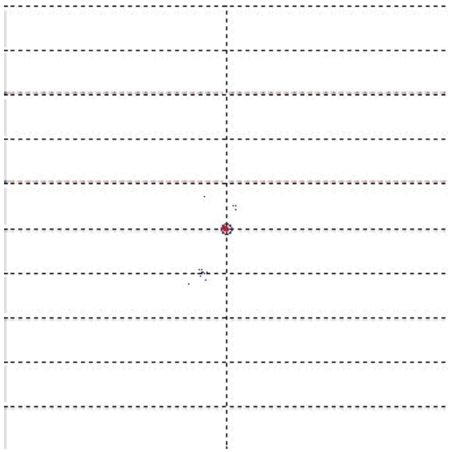
What is Visual Odometry (VO) ?

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)

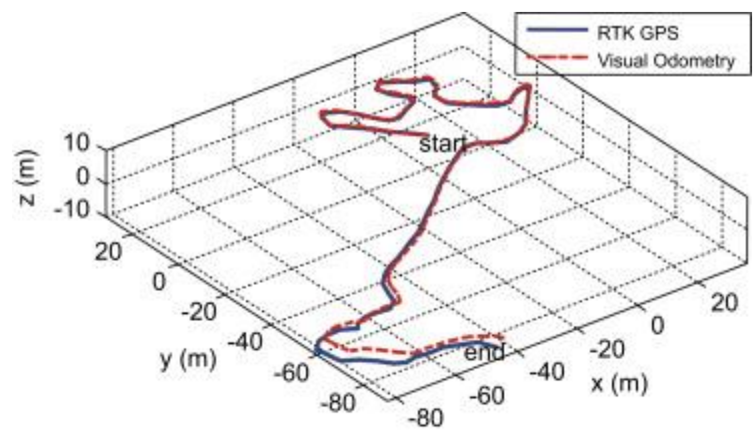
Why VO?

- Contrary to wheel odometry, VO is **not affected by wheel slippage** on uneven terrain or other adverse conditions.

- More accurate trajectory estimates compared to wheel odometry (**relative position error 0.1% – 2%**)

- VO can be used as a complement to
  - wheel encoders (wheel odometry)
  - GPS
  - inertial measurement units (IMUs)
  - laser odometry

- Crucial for flying, walking, and underwater robots

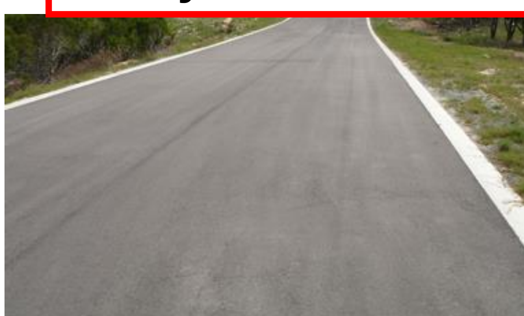


Assumptions

- **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.

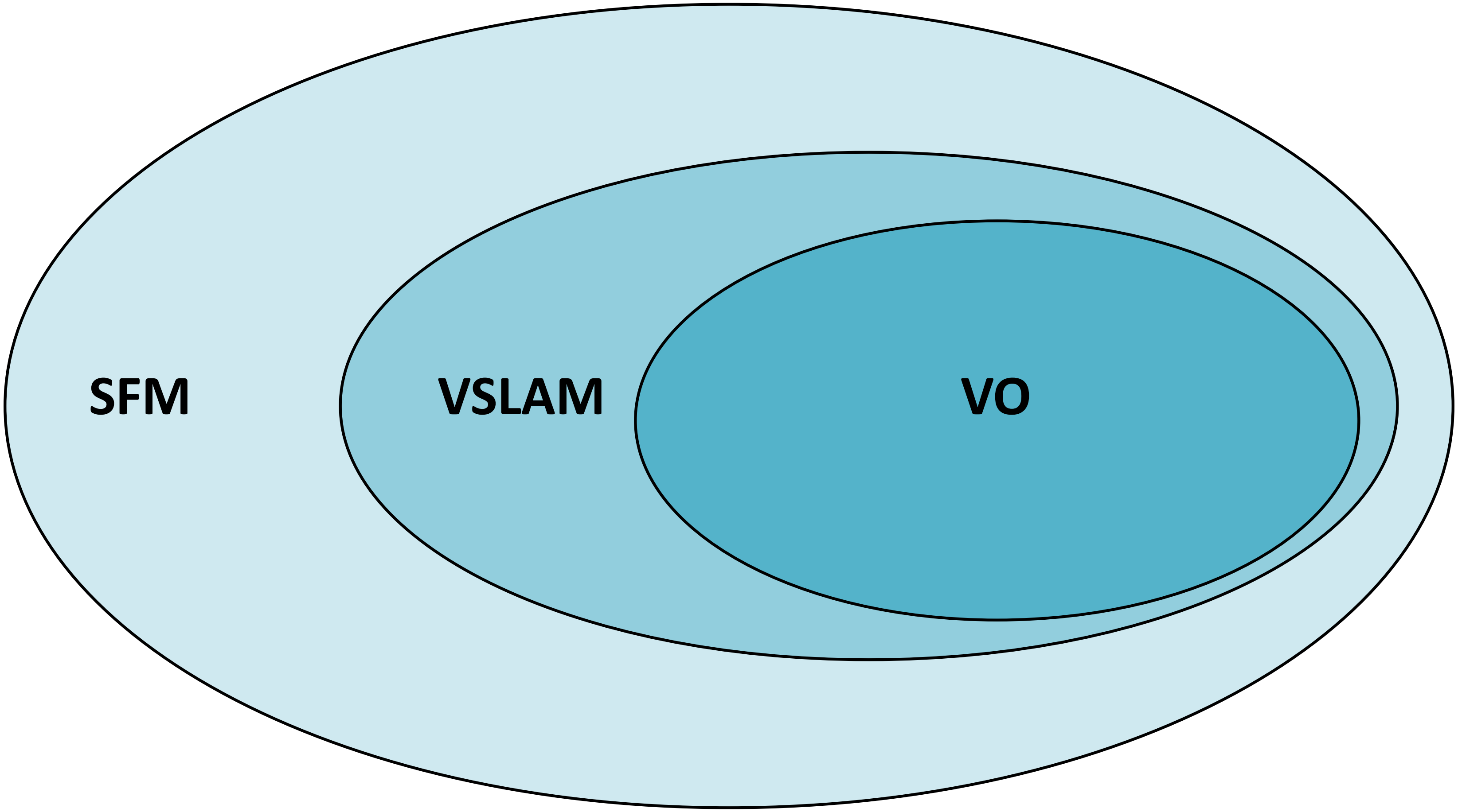
More about history and tutorials

- 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**

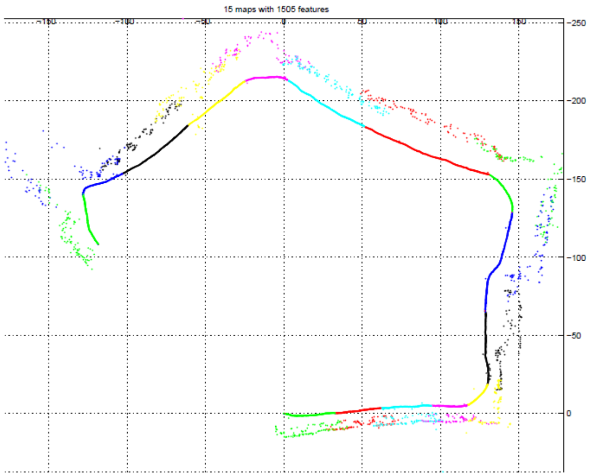
Reconstruction from 3 million images from Flickr.com
Cluster of 250 computers, 24 hours of computation!
Paper: “Building Rome in a Day”, ICCV’09

VO vs SFM

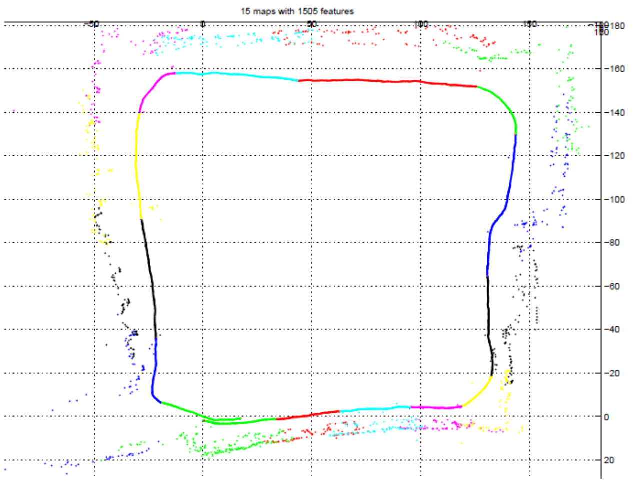
- VO is a **particular case** of SFM
- VO focuses on estimating the 3D motion of the camera **sequentially** (as a new frame arrives) and in **real time**.
- Terminology: sometimes SFM is used as a synonym of VO

VO vs. Visual SLAM

- **Visual Odometry**
  - Focus on incremental estimation/**local consistency**
- **Visual SLAM**: Simultaneous Localization And Mapping
  - Focus on **globally consistent** estimation
  - **Visual SLAM = visual odometry + loop detection + graph optimization**
- The choice between VO and V-SLAM depends on the **tradeoff between performance and consistency**, and simplicity of implementation.
- VO **trades off consistency for real-time performance**, without the need to keep track of all the previous history of the camera.



Visual odometry



Visual SLAM

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

VO Working Principle

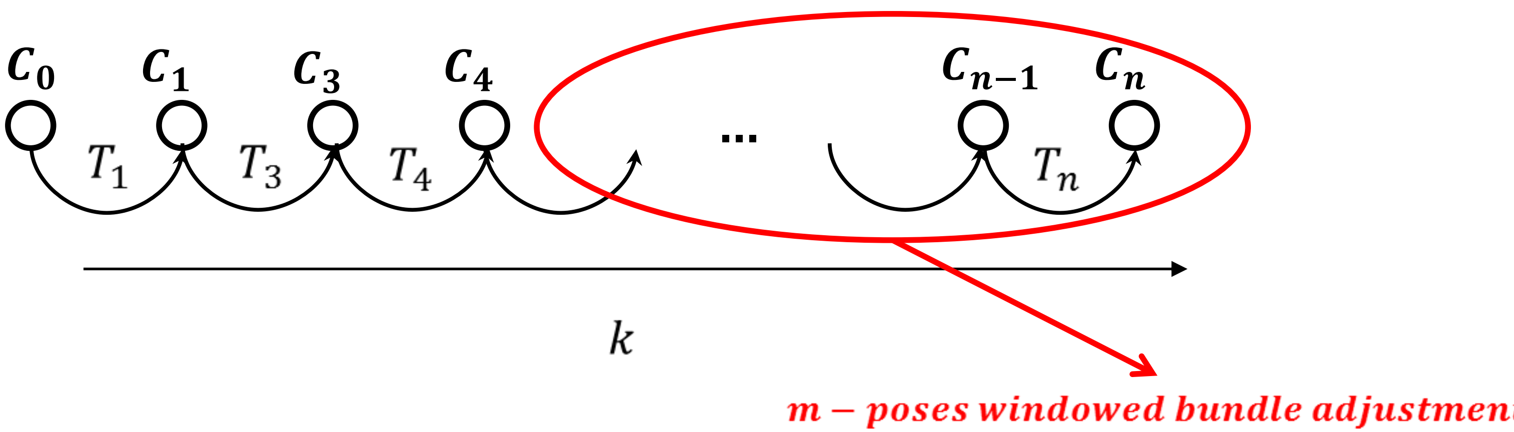
1. Compute the relative motion T_k from images I_{k-1} to image I_k

$$T_k = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ 0 & 1 \end{bmatrix}$$

2. Concatenate them to recover the full trajectory

$$C_n = C_{n-1} T_n$$

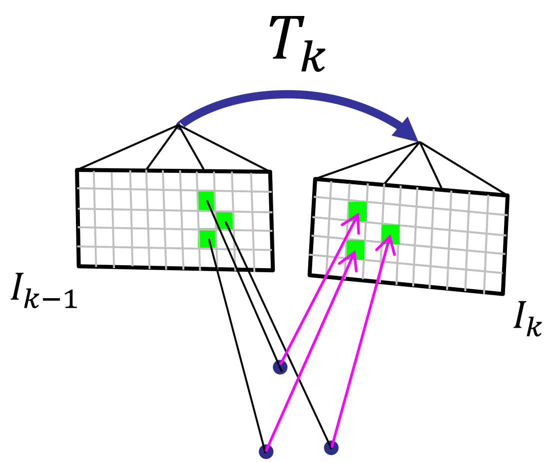
3. An optimization over the last *m* poses can be done to refine locally the trajectory (Pose-Graph or Bundle Adjustment)



How do we estimate the relative motion T_k ?

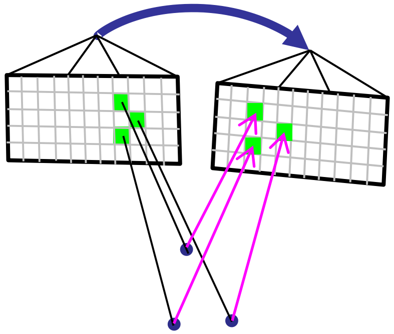
Image I_{k-1}

Image I_k



$$T_k = \arg\min_{\mathbf{T}} \iint_{\bar{\mathcal{R}}} \rho \left[I_k\Big(\pi\big(\mathbf{T} \cdot \pi^{-1}(\mathbf{u}, d_{\mathbf{u}})\big)\Big) - I_{k-1}(\mathbf{u}) \right] d\mathbf{u}$$

Direct Image Alignment



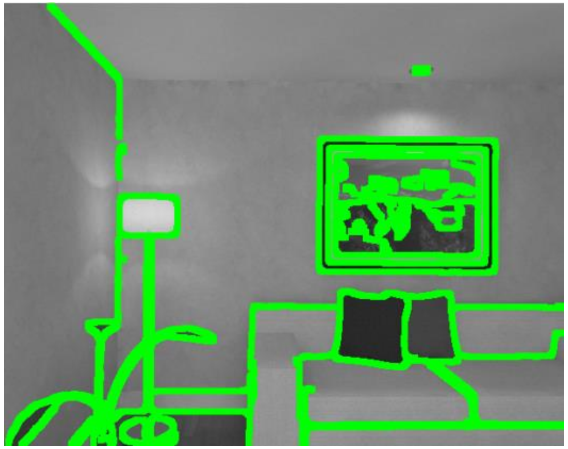
It minimizes the **per-pixel intensity difference** [1]

$$T_{k,k-1} = \arg\min_{T} \sum_{i} \| I_k(\boldsymbol{u}'_i) - I_{k-1}(\boldsymbol{u}_i) \|_\sigma^2$$

Dense

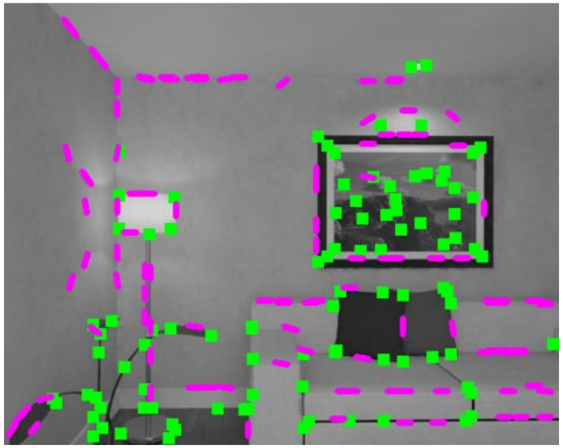
DTAM [Newcombe et al. '11]
300'000+ pixels

Semi-Dense



LSD [Engel et al. 2014]
~10'000 pixels

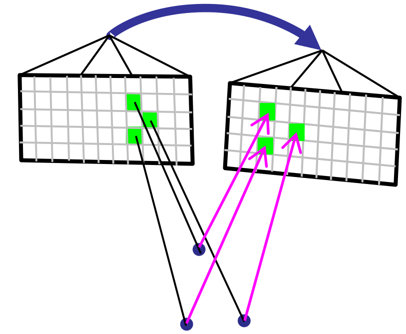
Sparse



SVO [Forster et al. 2014]
100-200 features x 4x4 patch
~ 2,000 pixels

D. Cremers, Direct methods for 3D reconstruction and visual SLAM, International Conference on Machine Vision Applications, 2017, PDF

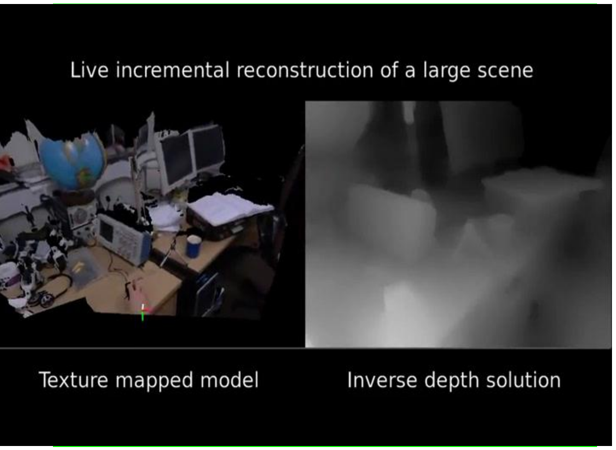
Direct Image Alignment



It minimizes the **per-pixel intensity difference** [1]

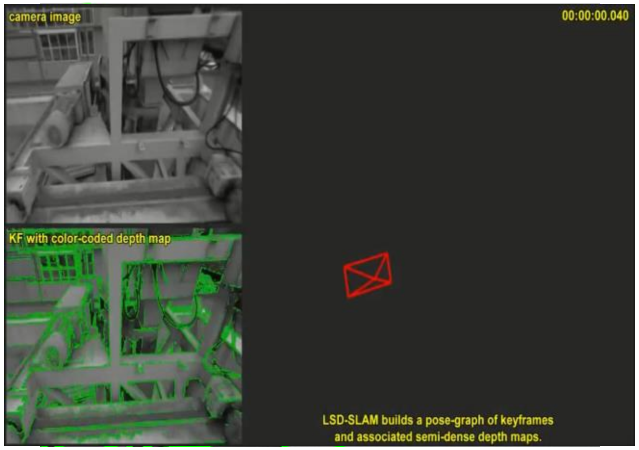
$$T_{k,k-1} = \arg\min_{T} \sum_{i} \| I_k(\boldsymbol{u}'_i) - I_{k-1}(\boldsymbol{u}_i) \|_{\sigma}^{2}$$

Dense



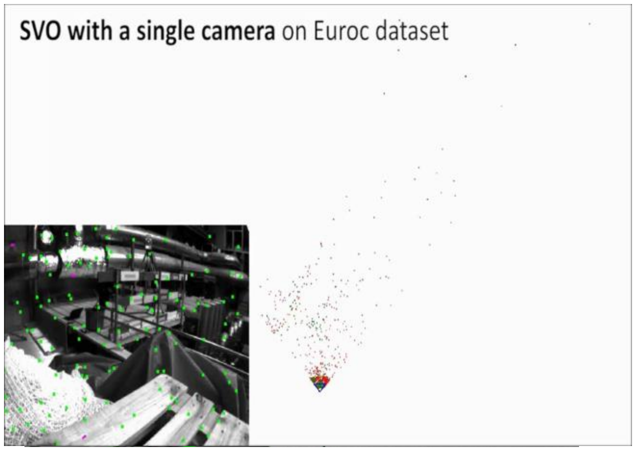
DTAM [Newcombe '11] REMODE [Pizzoli'14]
300'000+ pixels

Semi-Dense



LSD-SLAM [Engel'14]
~10,000 pixels

Sparse

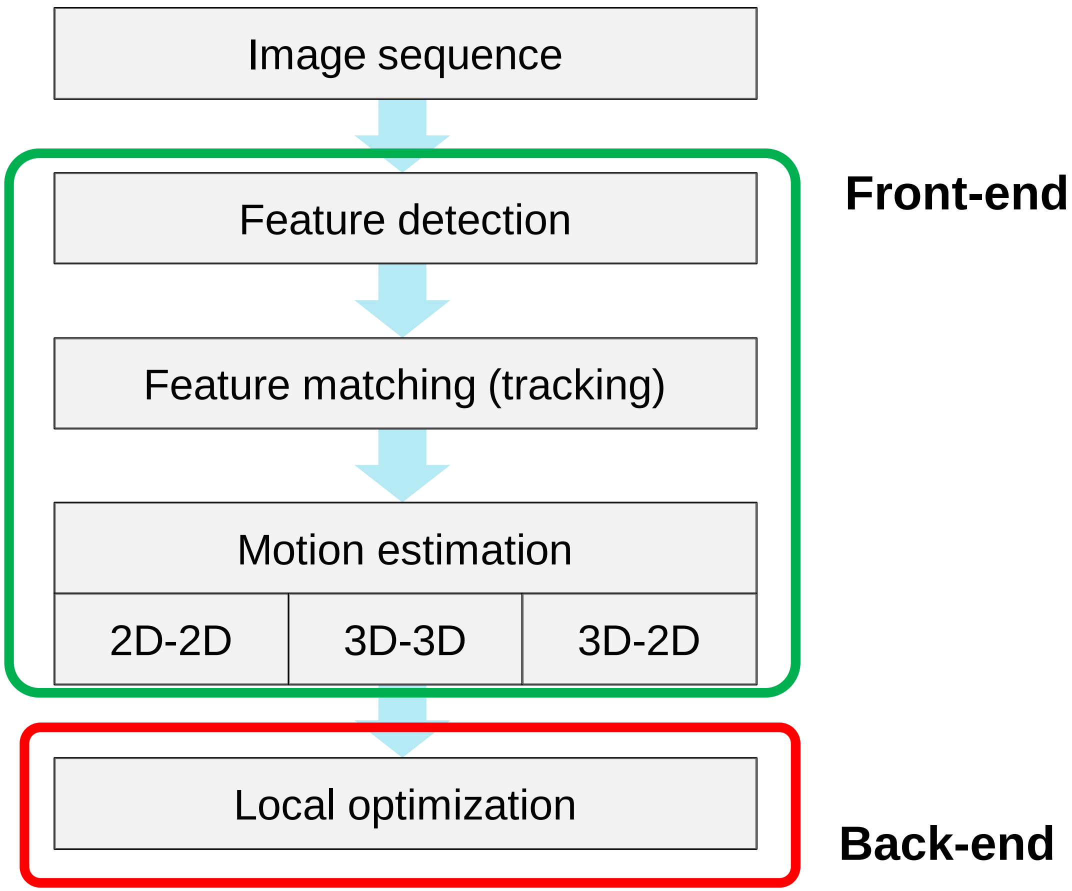


SVO [Forster'14]
100-200 x 4x4 patches ≅ 2,000 pixels

D. Cremers, Direct methods for 3D reconstruction and visual SLAM, International Conference on Machine Vision Applications, 2017, PDF

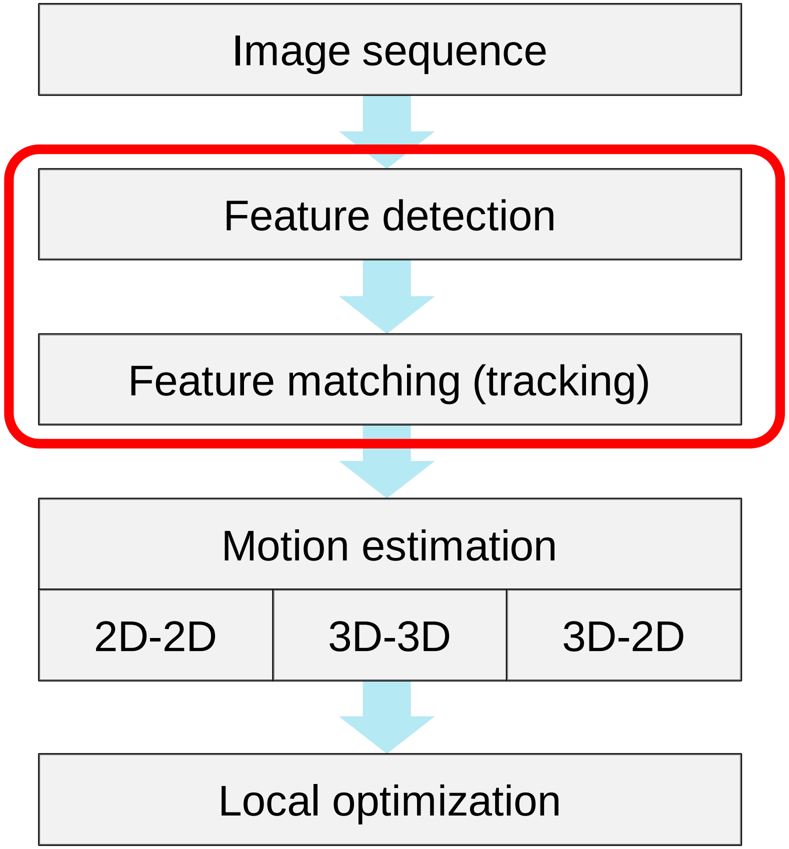
VO Flow Chart

VO computes the camera path incrementally (pose after pose)



VO Flow Chart

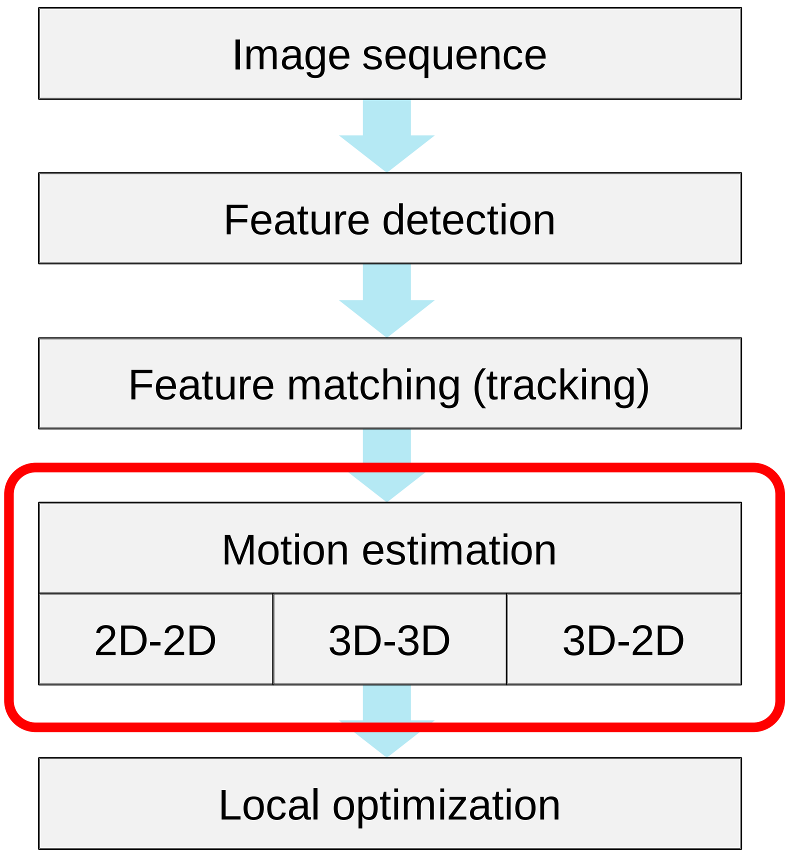
VO computes the camera path incrementally (pose after pose)

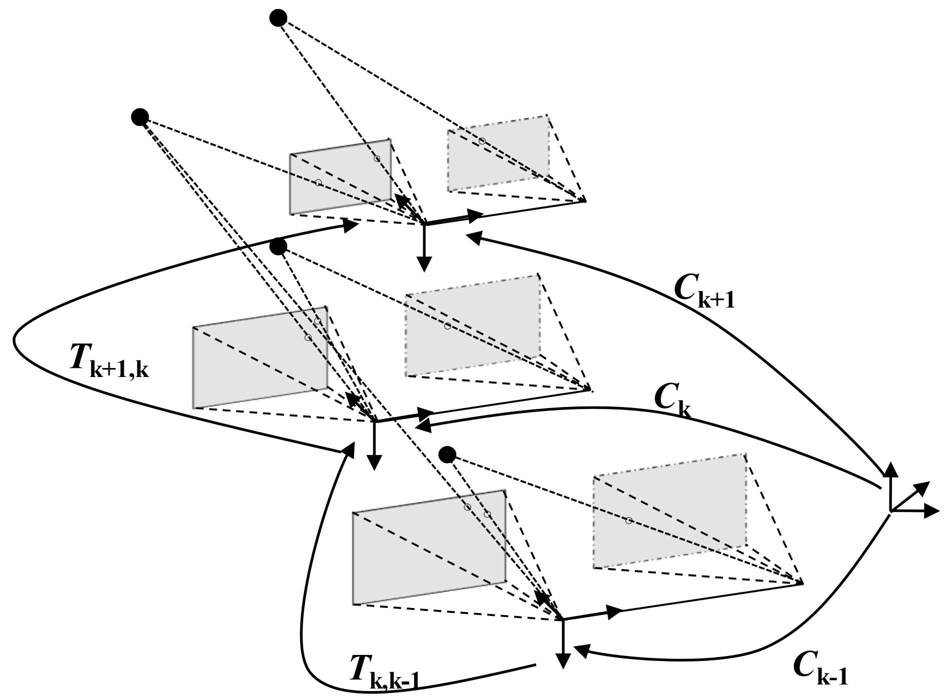


Example features tracks

VO Flow Chart

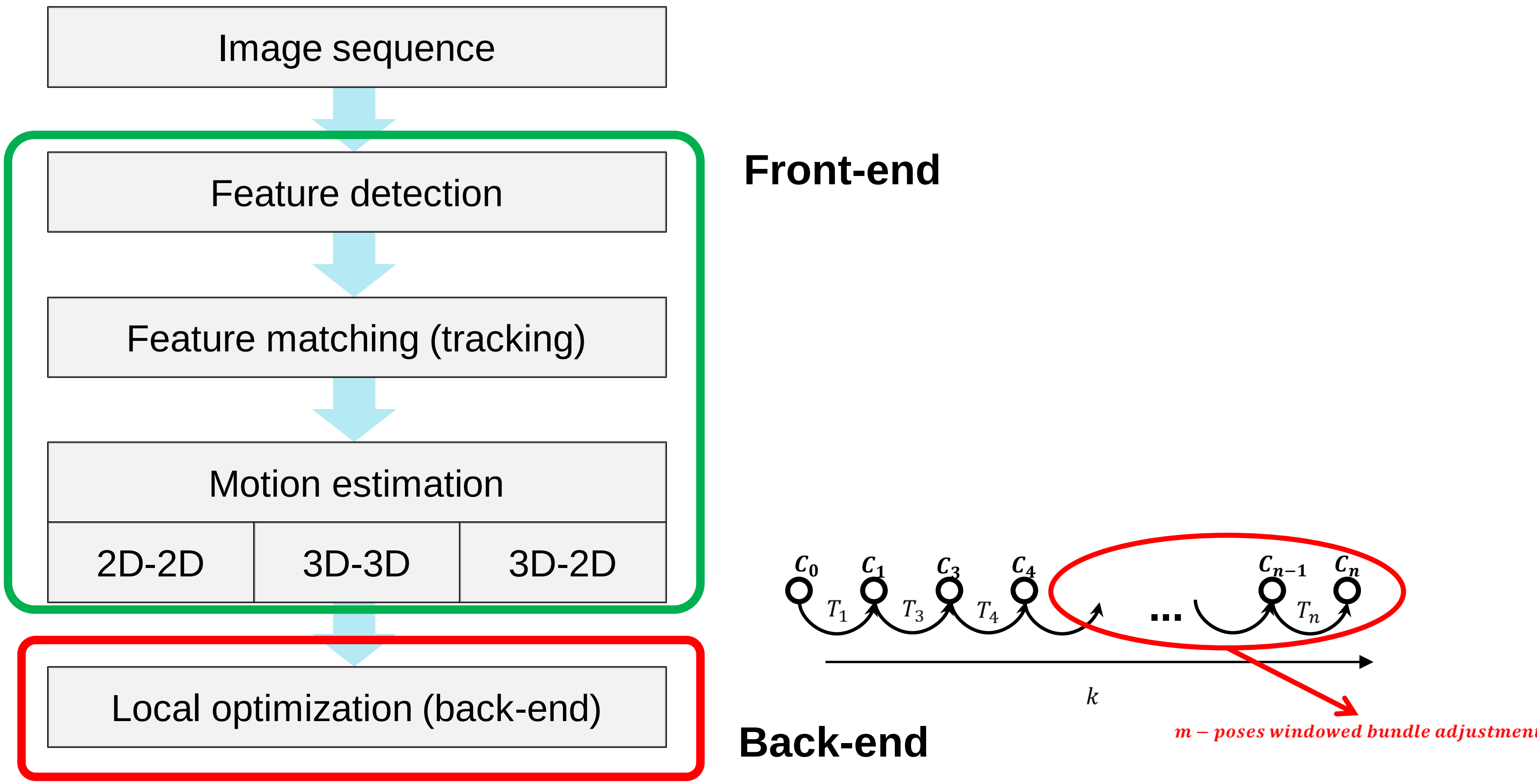
VO computes the camera path incrementally (pose after pose)





VO Flow Chart

VO computes the camera path incrementally (pose after pose)

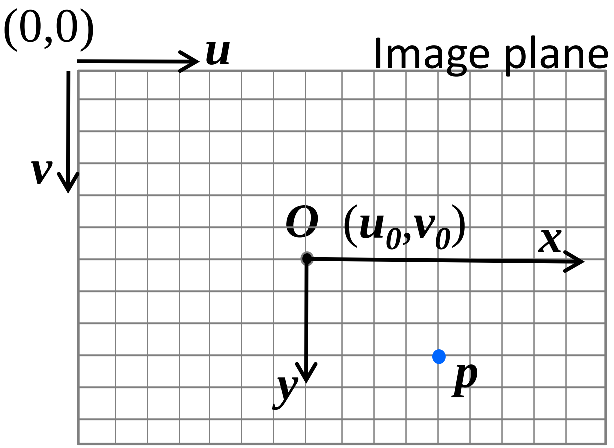


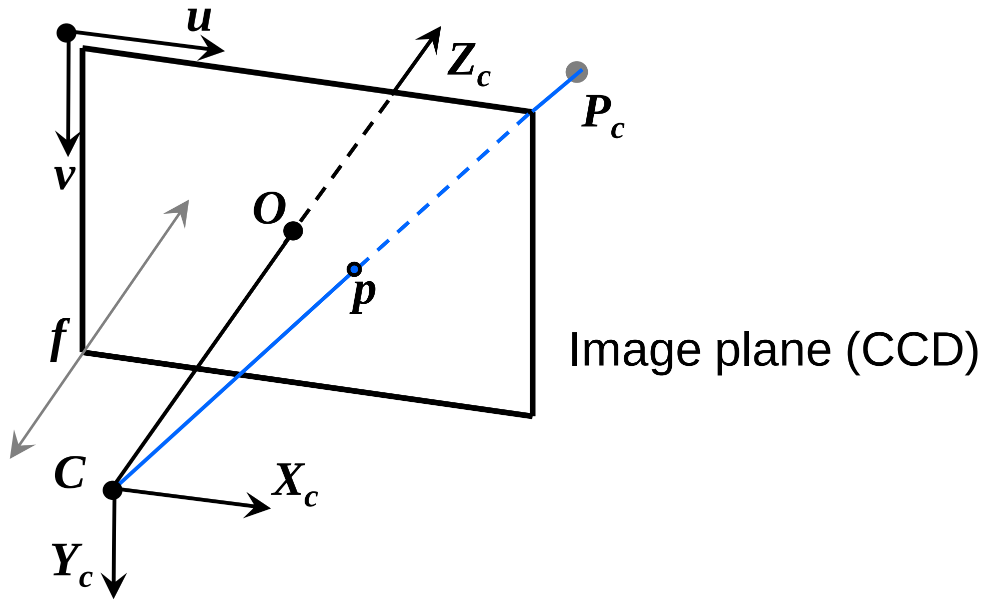
Course Topics

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

Course Topics

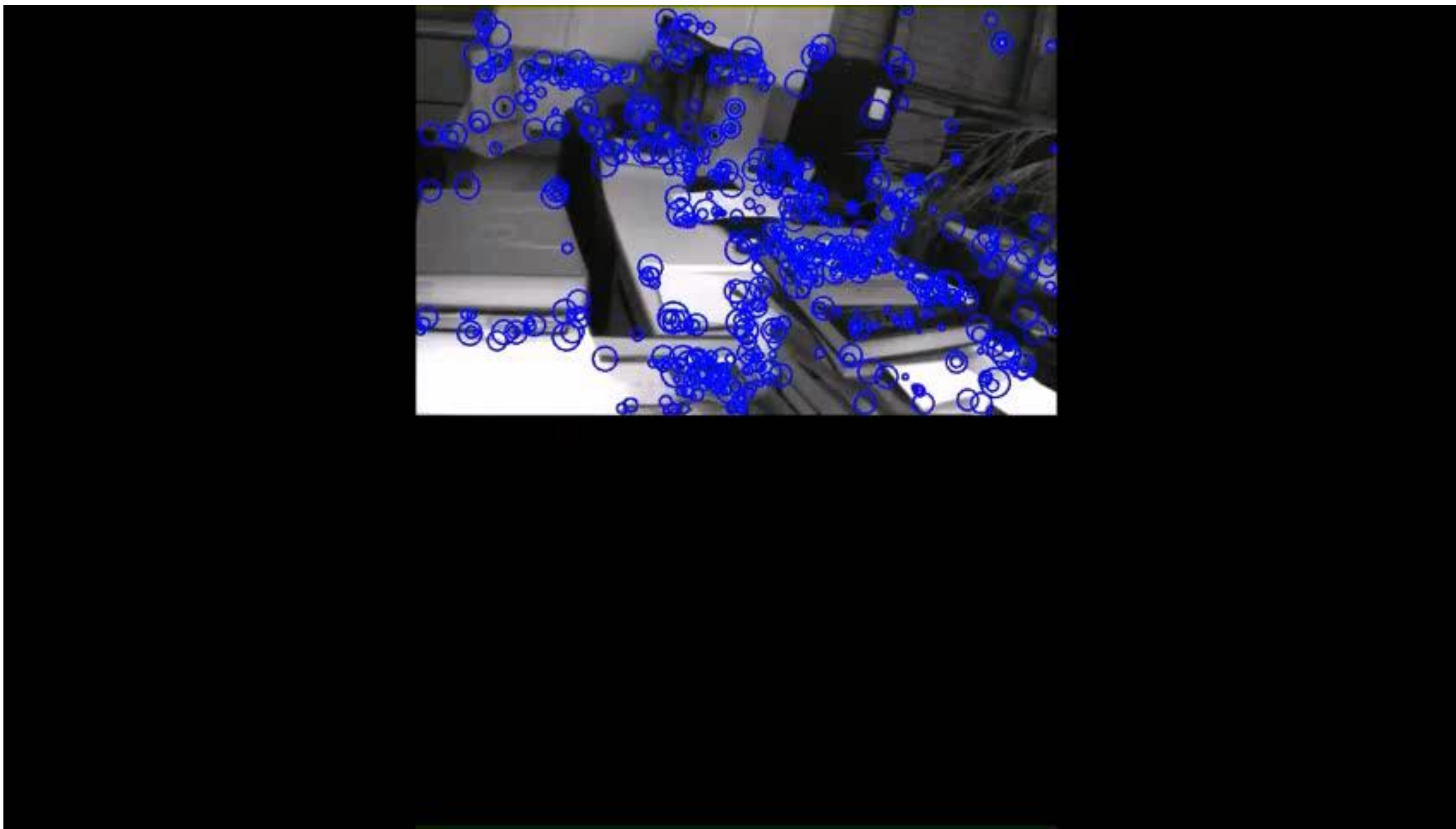
- Principles of image formation
  - Perspective projection
  - Camera calibration





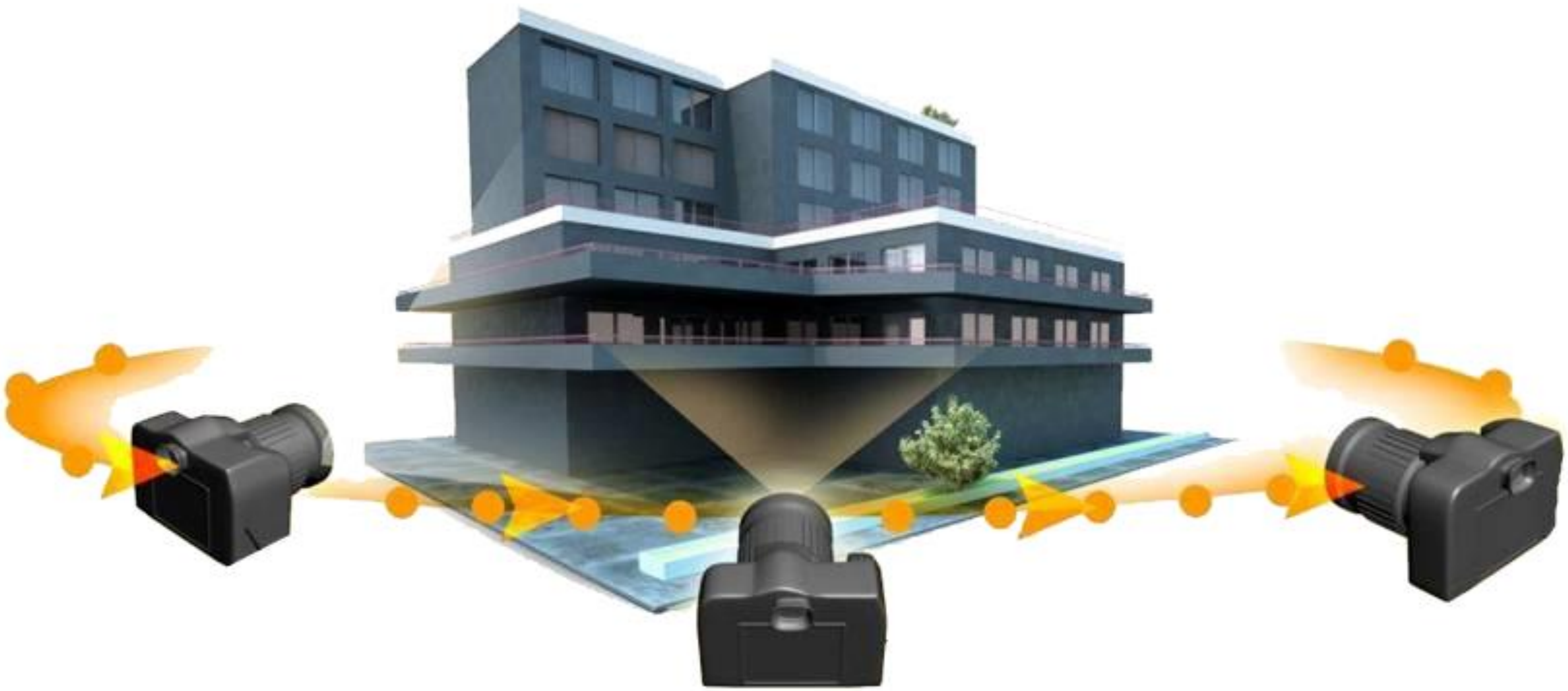
Course Topics

- Feature detection and matching



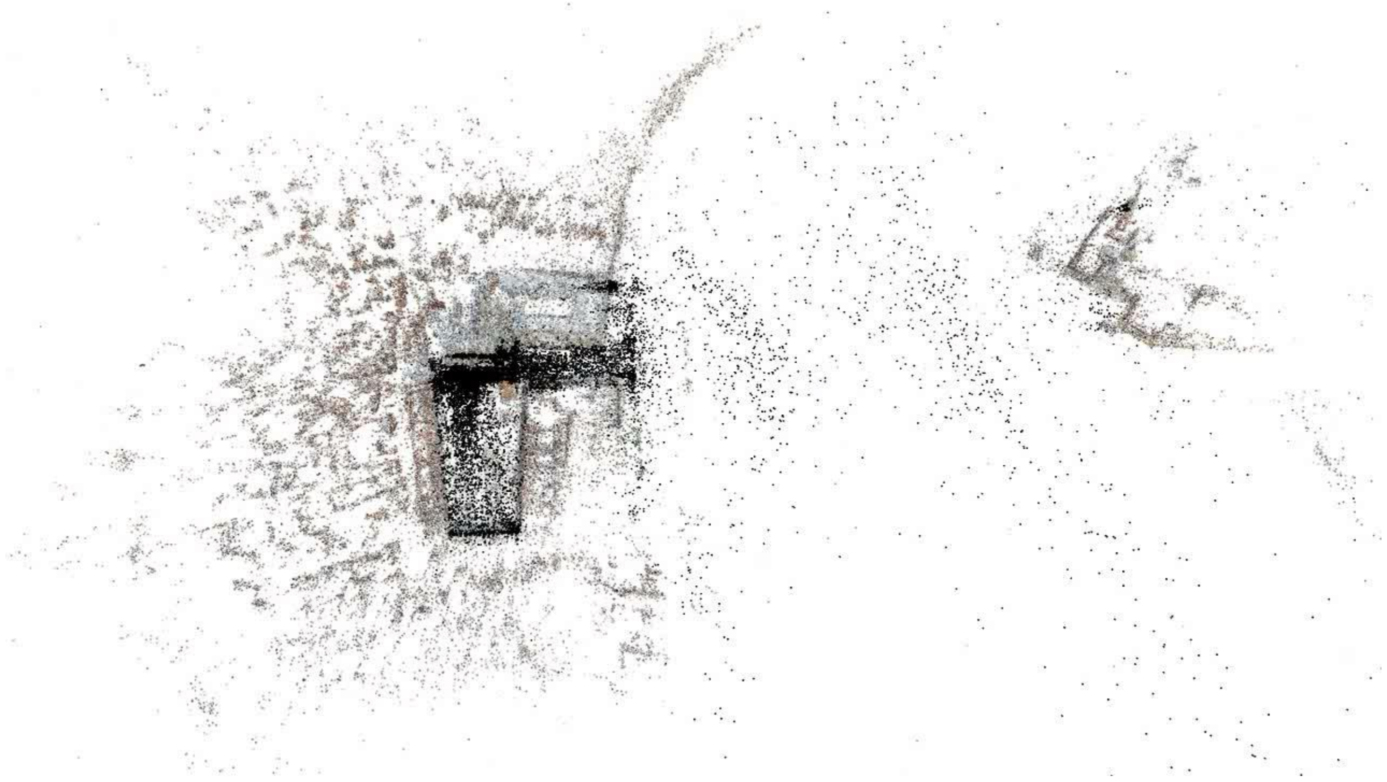
Course Topics

- Multi-view geometry and 3D reconstruction



Course Topics

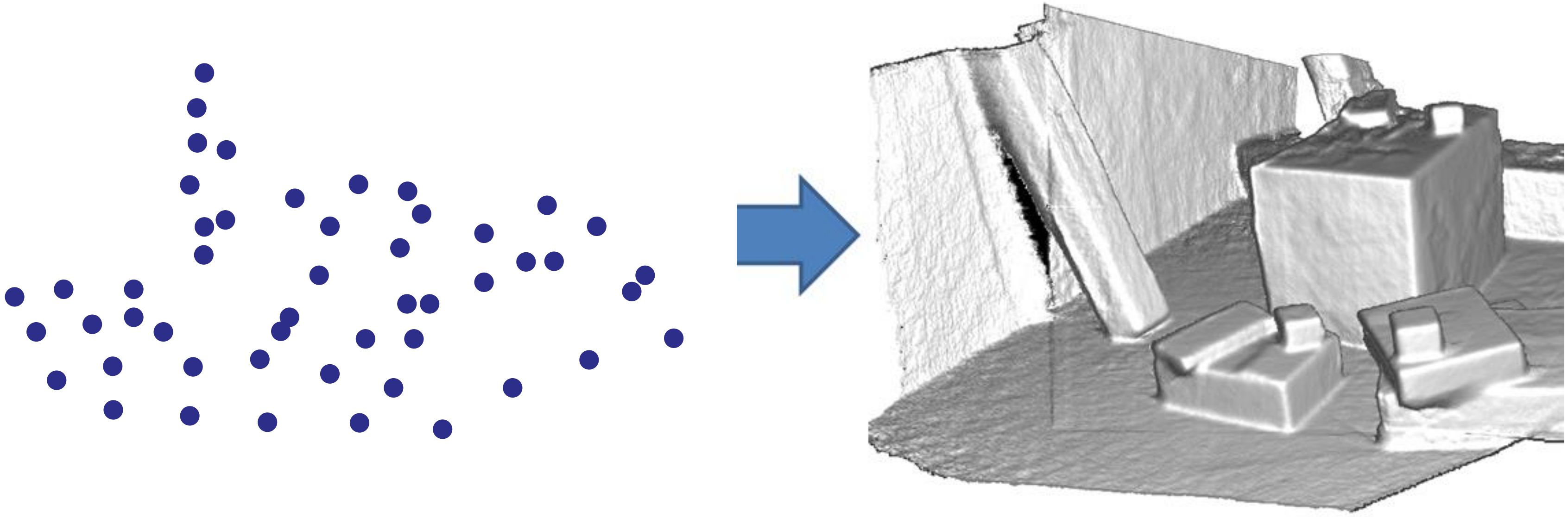
- Multi-view geometry and 3D reconstruction



San Marco square, Venice
14,079 images, 4,515,157 points

Course Topics

- Dense reconstruction



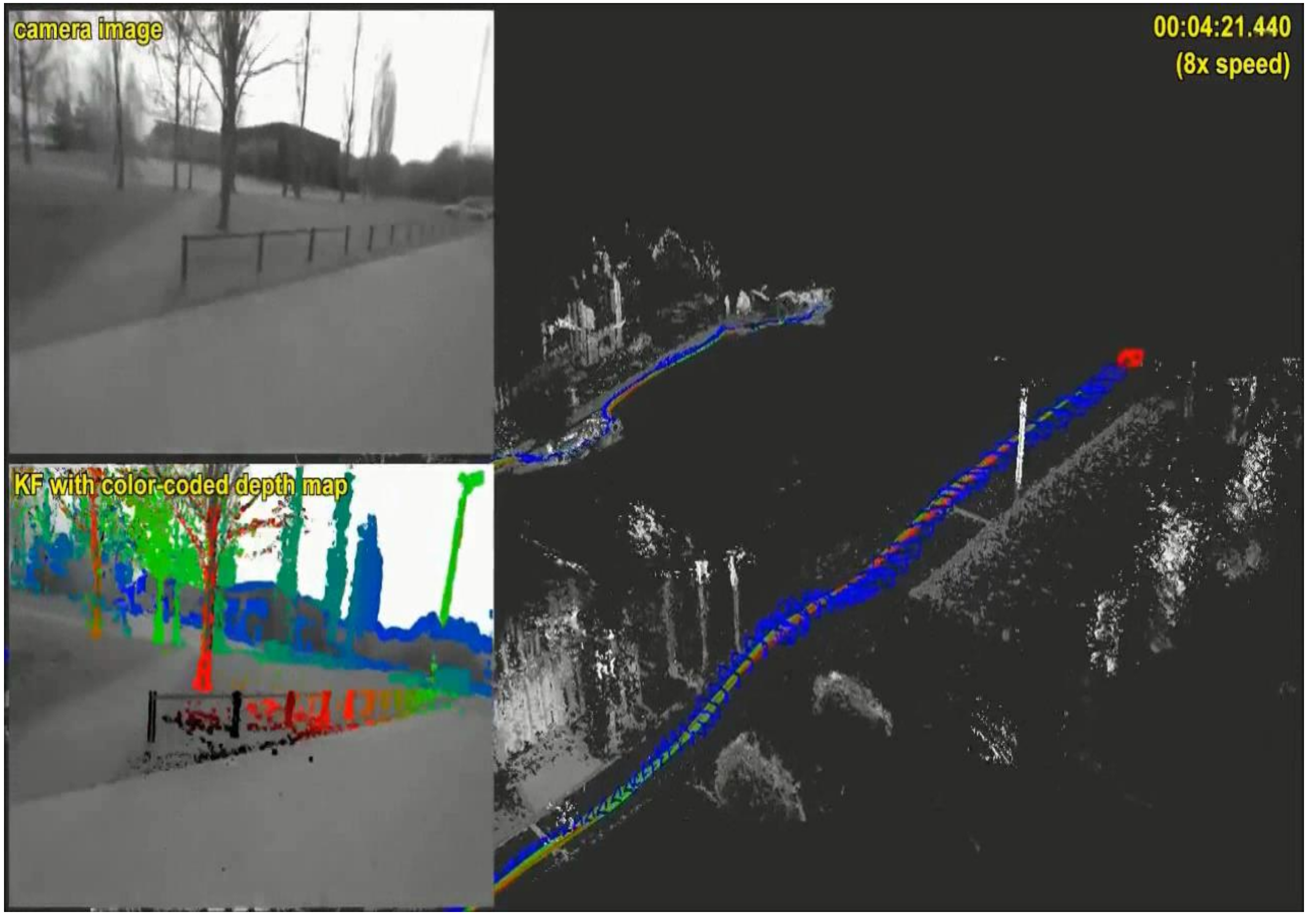
Course Topics

- Dense reconstruction

M. Pizzoli, C. Forster, D. Scaramuzza, REMODE: Probabilistic, Monocular Dense Reconstruction in Real Time, ICRA'14

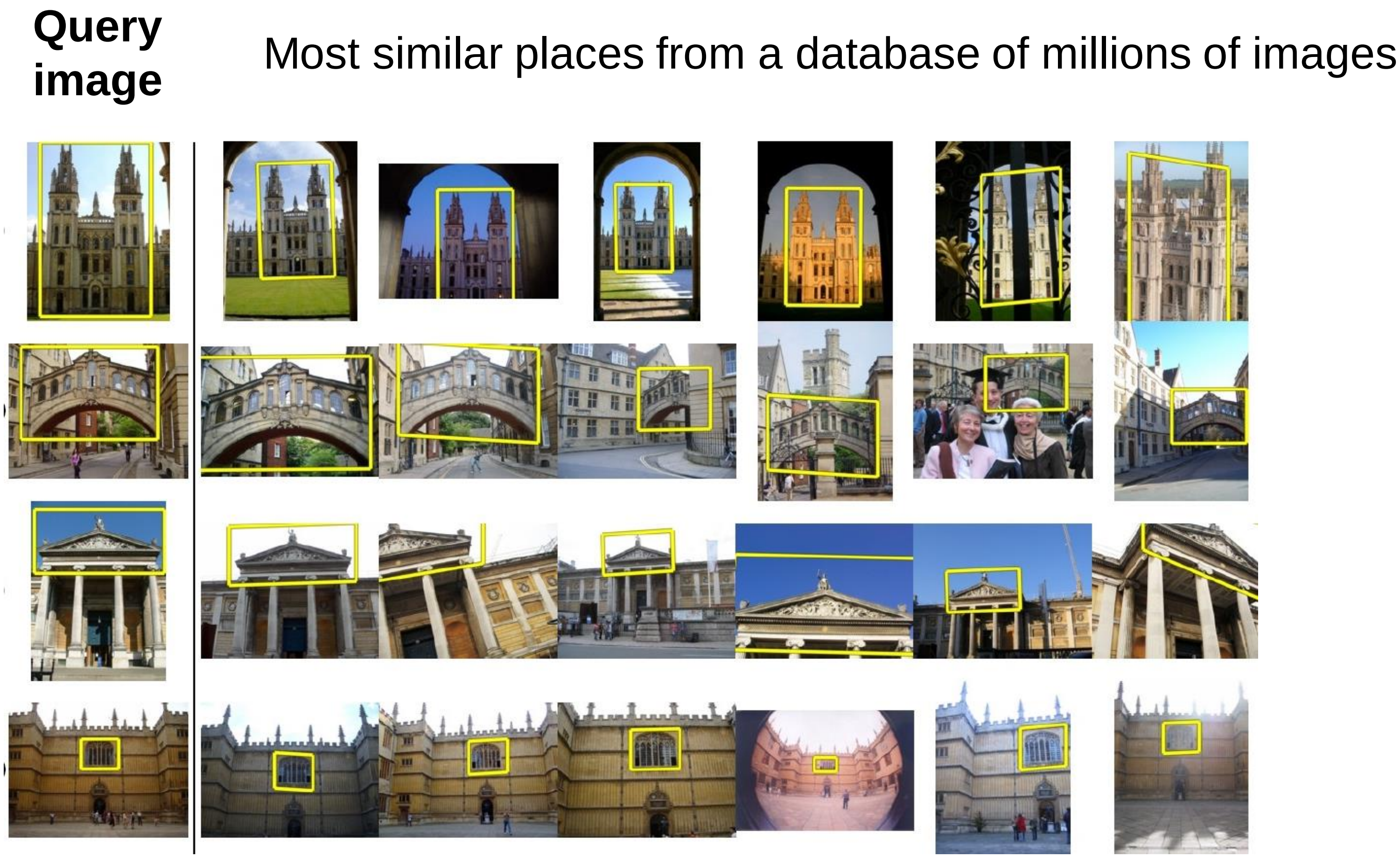
Course Topics

- Visual place recognition



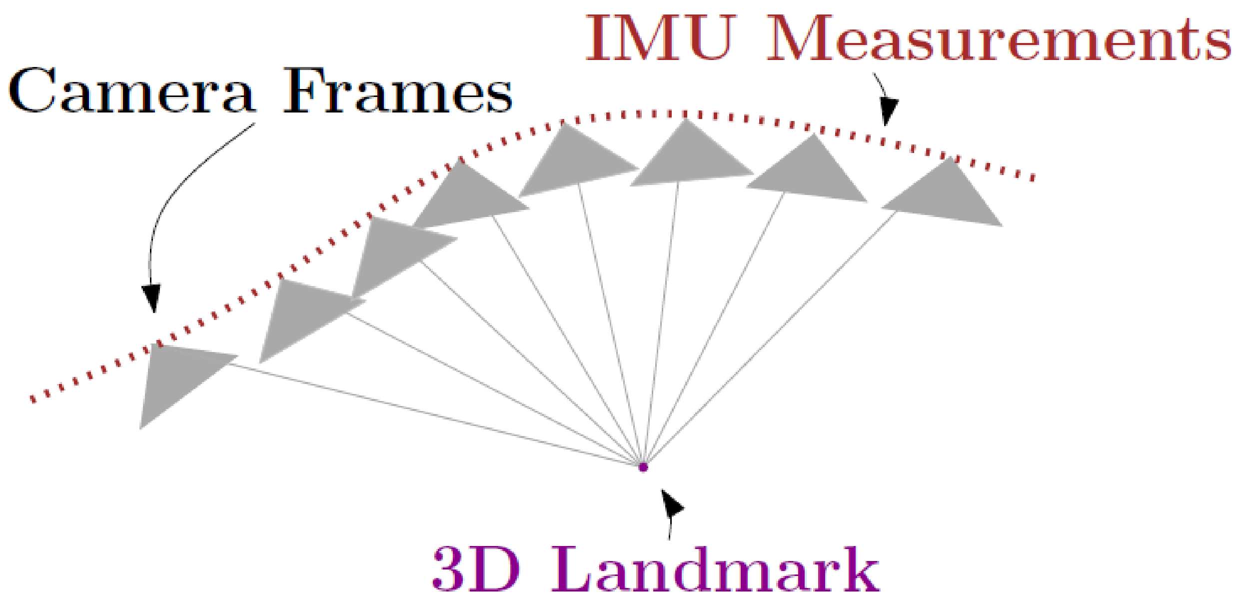
Course Topics

- Visual place recognition



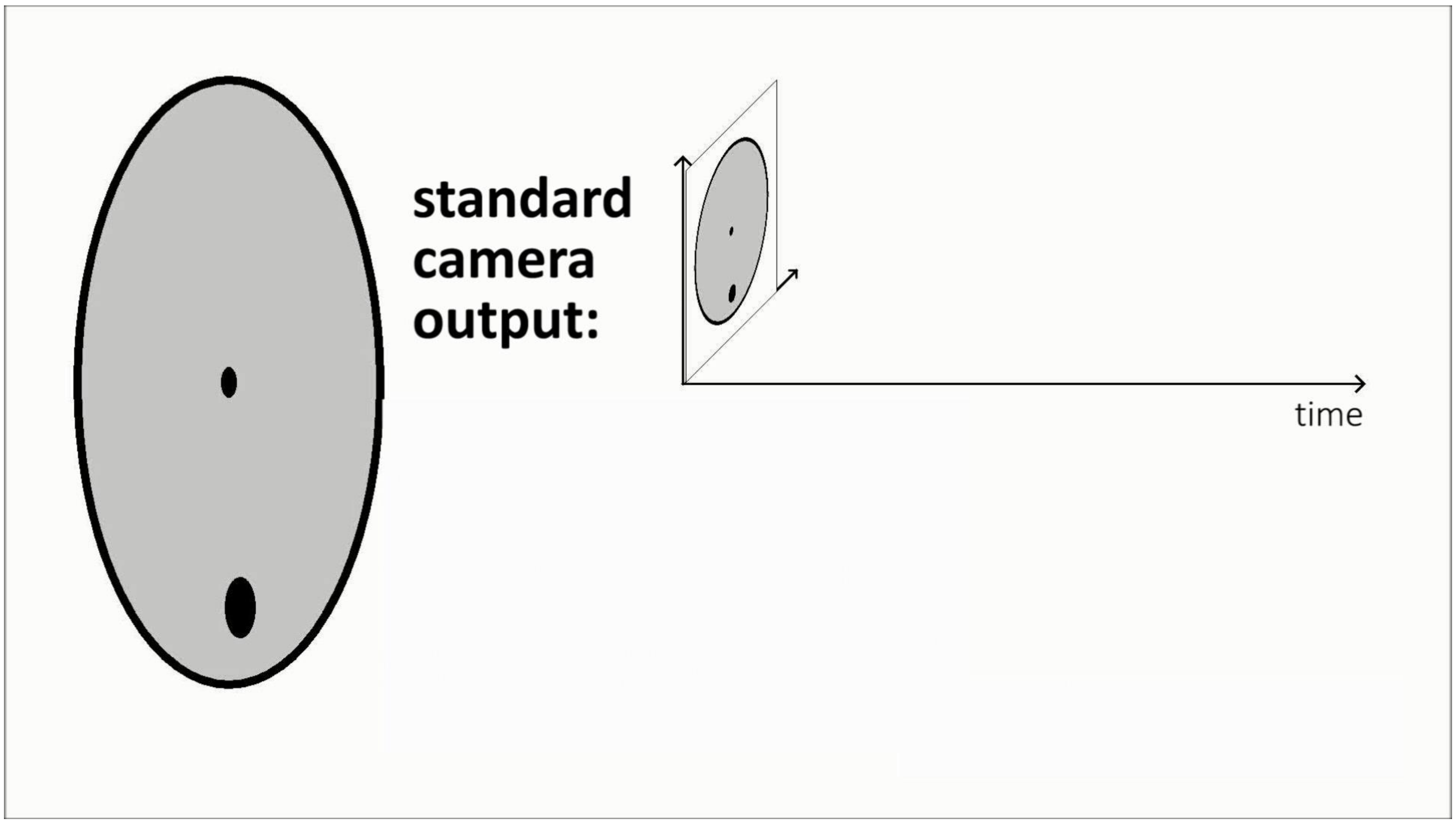
Course Topics

- Visual-inertial fusion



Course Topics

- Event-based vision



Course Topics

- Visual odometry