SLAM II: SLAM for robotic vision-based perception

Autonomous Mobile Robots

Margarita Chli

Martin Rufli, Roland Siegwart

SLAM II | today's lecture

Last time: how to do SLAM?

Today: what to do with SLAM?

- Vision-based SLAM – state of the art
- Vision-based Robotic Perception:
  - Current Challenges &
  - Case Studies (past & ongoing EU projects)
  - Overview of Research Activities in V4RL

Computer Vision for Robotics | making an impact

V4RL VISION FOR ROBOTICS LAB

Computer Vision meets Robotics| the SLAM problem

SLAM (SIMULTANEOUS LOCALIZATION AND MAPPING):

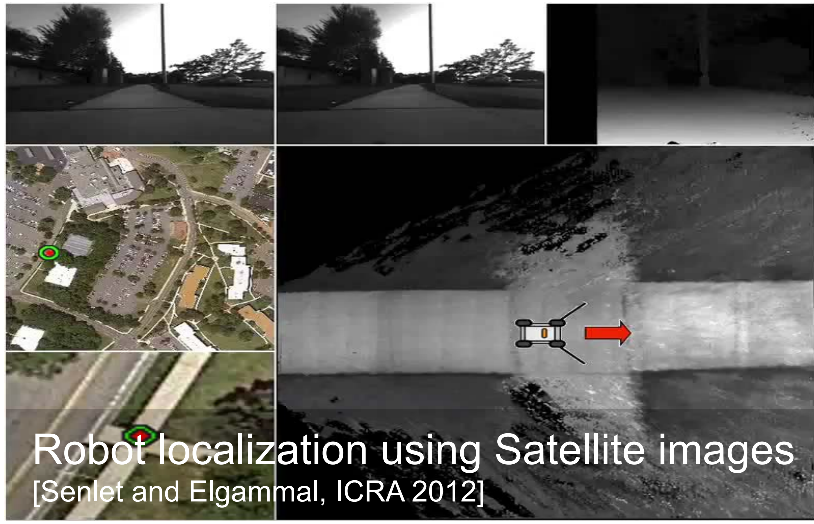
*"How can a body **navigate** in a previously unknown environment, while constantly building & updating a **map** of its workspace using onboard sensors & onboard computation only?"*

- **The backbone of spatial awareness of a robot**

- One of the most challenging problems in probabilistic robotics

  - **Pure localization with a known map.**
    SLAM: no a priori knowledge of the robot's workspace

  - **Mapping with known robot poses.**
    SLAM: the robot poses have to be estimated along the way



Robot localization using Satellite images
[Senlet and Elgammal, ICRA 2012]

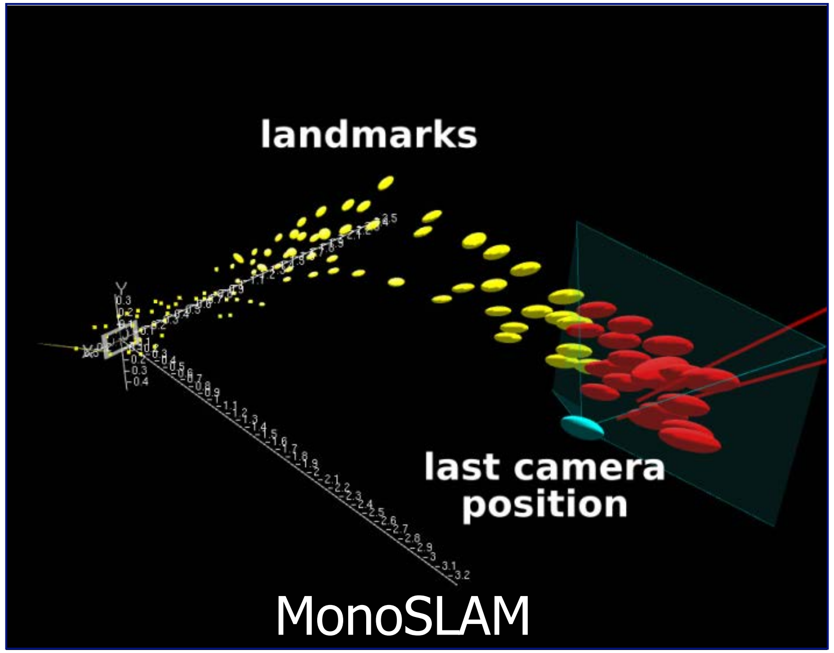
Helicopter position given by Vicon tracker
ETH Zurich Flying Machine Arena, IDCS & NCCR DF, 2013

SLAM | how does it work?

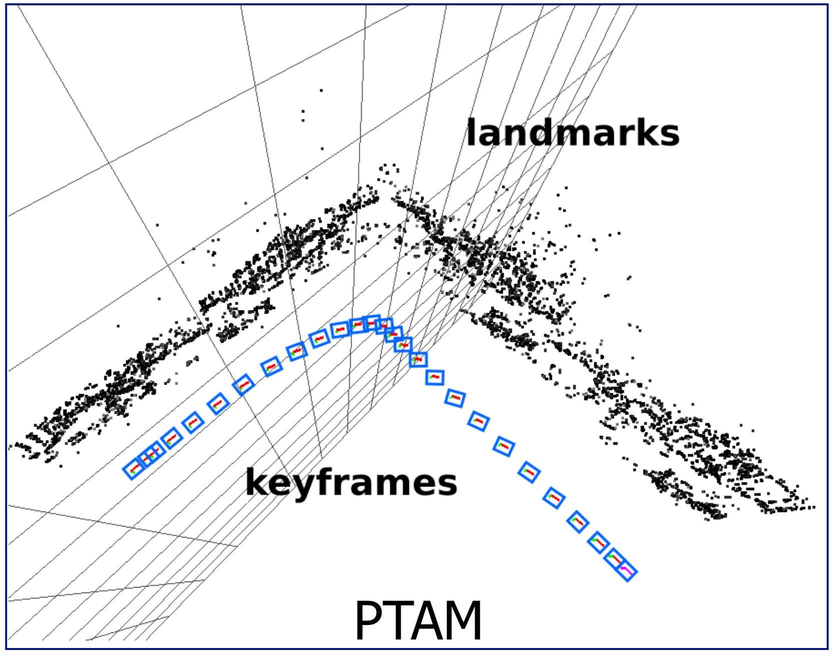
- Can we track the motion of a camera/robot while it is moving?
- Traditional SLAM:
  Pick natural scene features as landmarks, observe their motion & reason about robot motion
- Research into:
  - "Good" features to track, sensors, trackers, representations, assumptions
  - Ways of dealing with uncertainty in the processes involved

The videos are courtesy of A. Davison

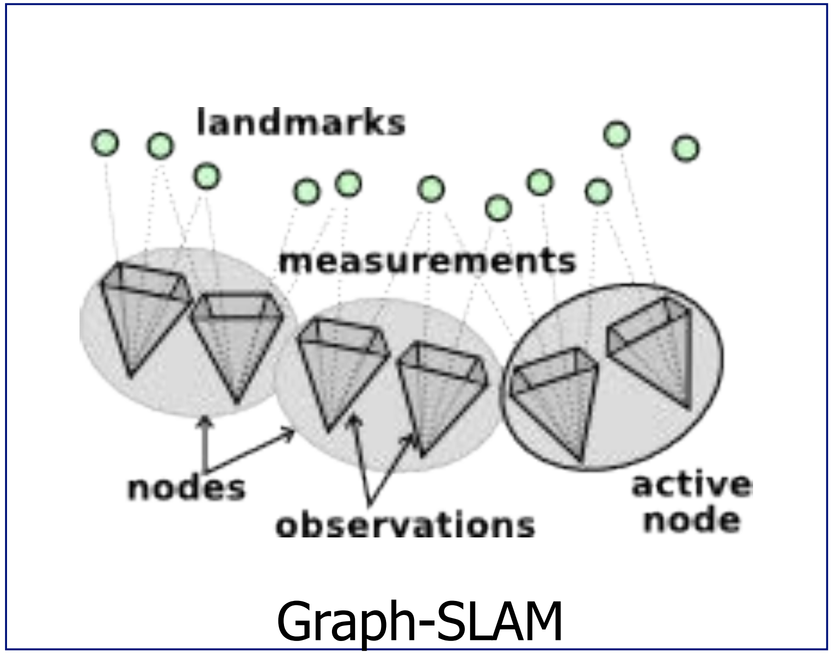
Monocular SLAM | milestone systems



[Davison et al. 2003, 2007]



[Klein, Murray 2007]

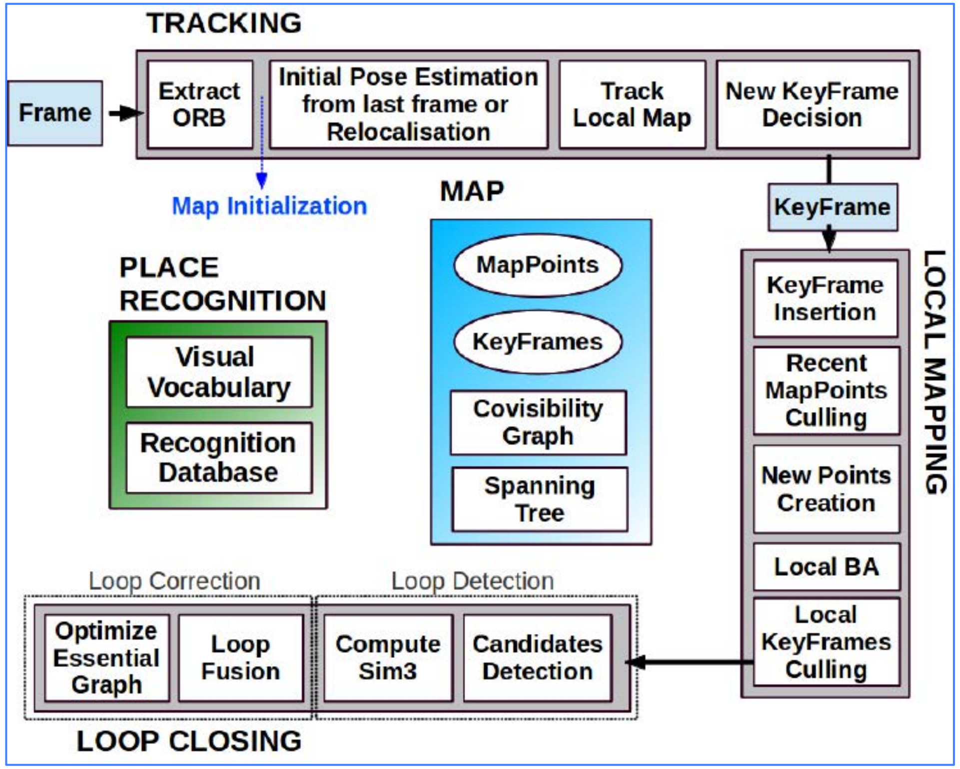


[Eade, Drummond 2007]

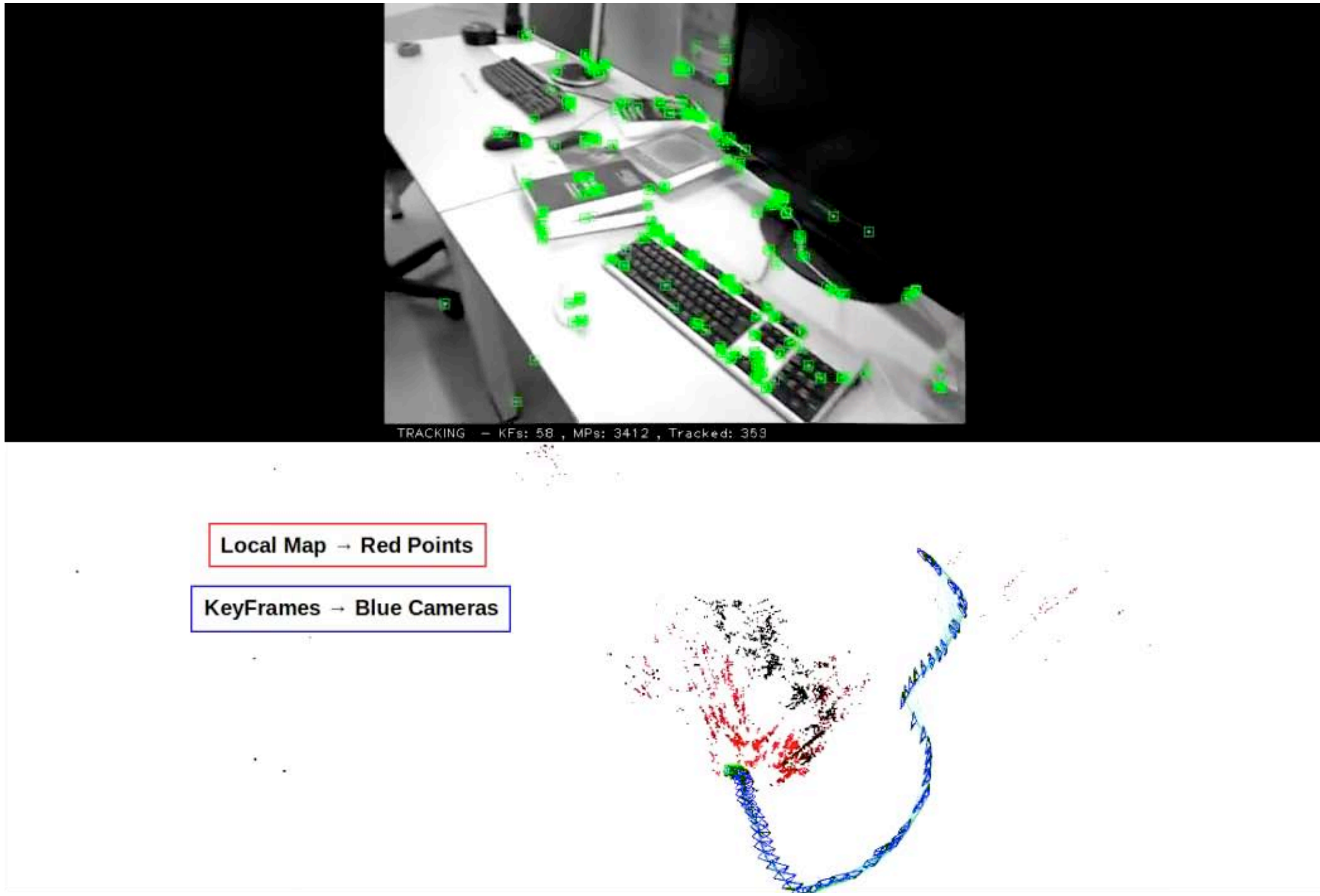
- ✓ revolutionary in the Vision & Robotics communities, but…
- ✗ not ready to perform tasks in general, uncontrolled environments

ORB-SLAM [Mur-Artal et al., TRO 2015]

- One of the most powerful vision-only SLAM approaches currently
- Uses ORB features (binary) in a keyframe-based approach
- Binary place recognition



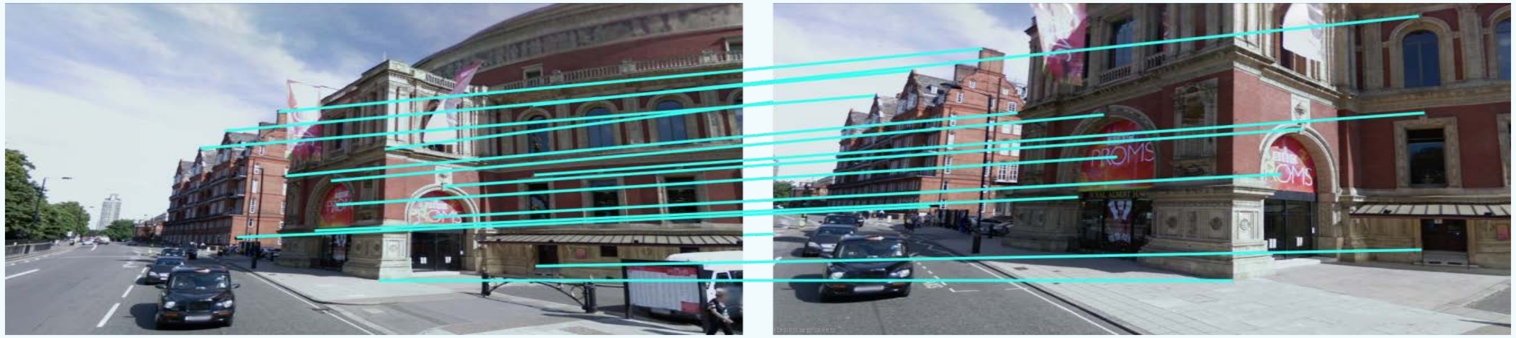
Code available on http://webdiis.unizar.es/~raulmur/orbslam/



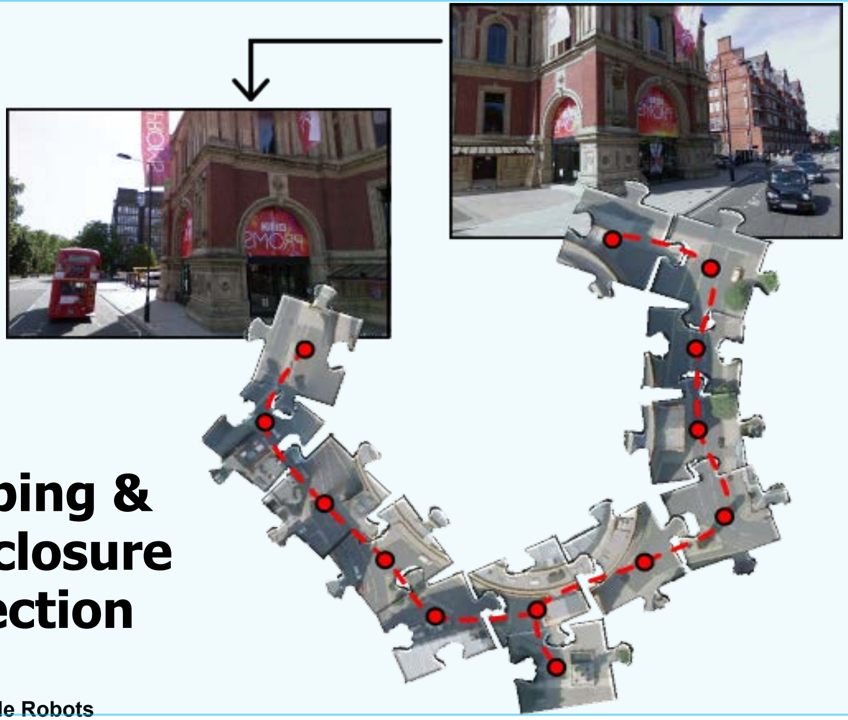
Scalable SLAM | the essentials

1

Robust local motion estimation



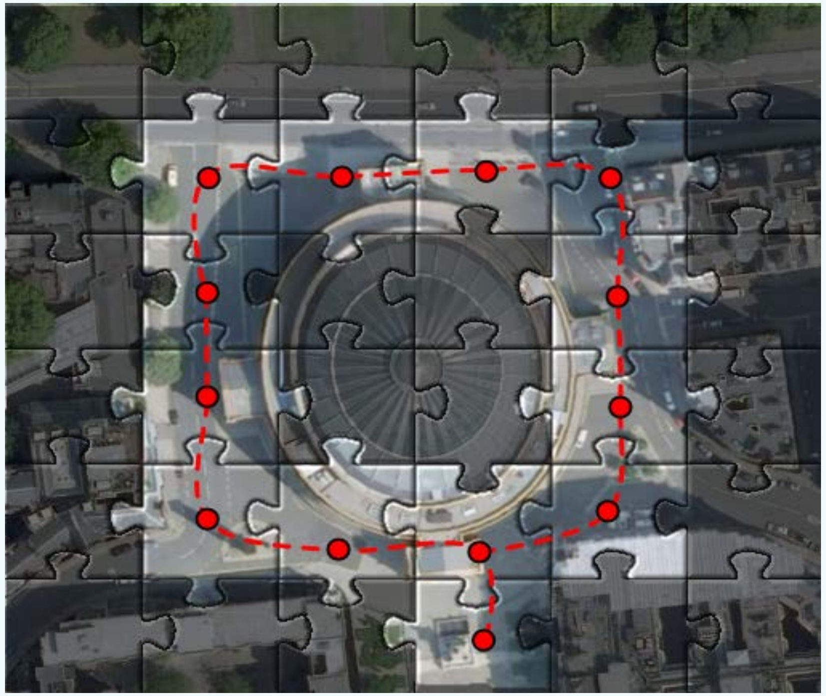
2



Mapping & loop-closure detection

3

Map management & optimisation



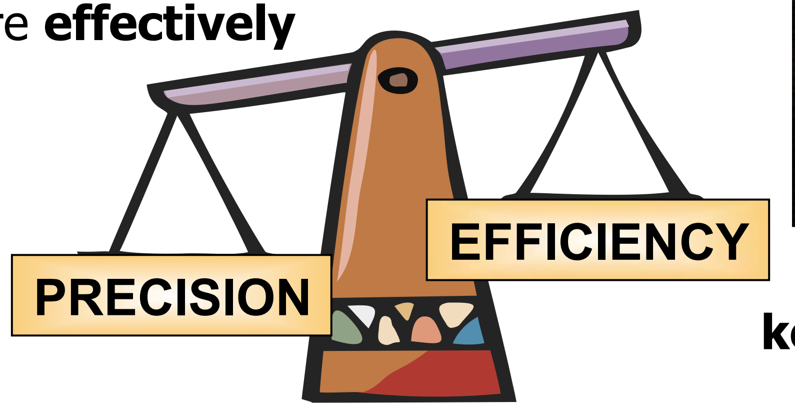
[Chli, PhD Thesis, 2009]

SLAM | current challenges

- Fast motion
- Large scales
- Robustness
- Rich maps
- Low computation for embedded apps
- Combination of multiple agents

↳ dynamic scenes, motion blur, lighting, …

- Handle **larger** amounts of data more **effectively**
- Competing goals:



- When is it worth to process an extra piece of information?
- Employ **Information Theory** to
  - Quantify amount of information gained
  - Understand the problems we are trying to solve

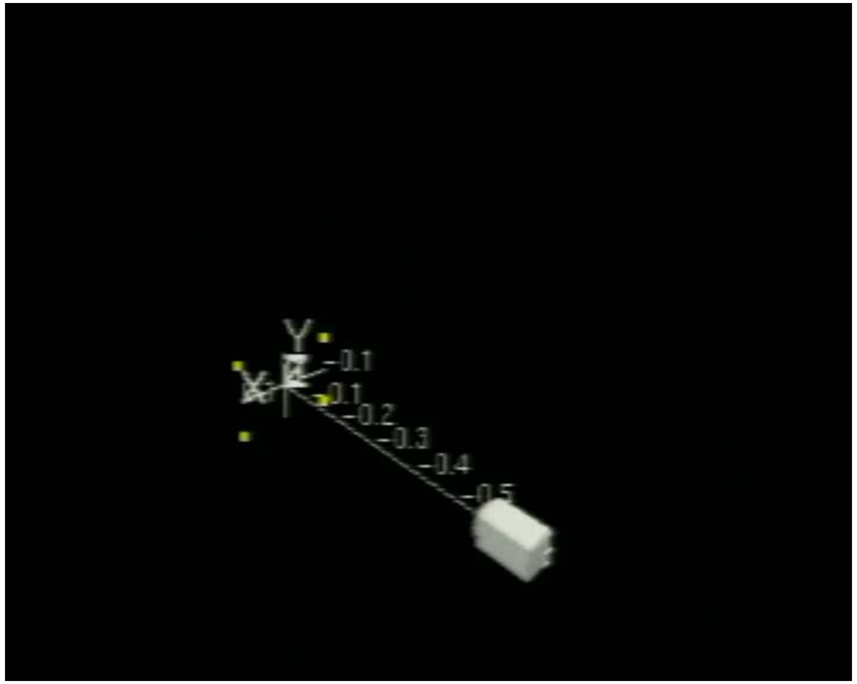
1ms Target Tracking Vision Chip
by Ishikawa Komuro University

key: agile manipulation of information

Computer Vision meets Robotics

| a very short history

2007: [**MonoSLAM**, Davison et al., PAMI]



2009: EU FP7, **sFly**

sFly | swarm of micro flying robots

aim:

Fully autonomous UAVs* to operate in and map an unknown environment in a search & rescue scenario.

*UAV= Unmanned Aerial Vehicle

Small UAVs | properties & challenges

Weight

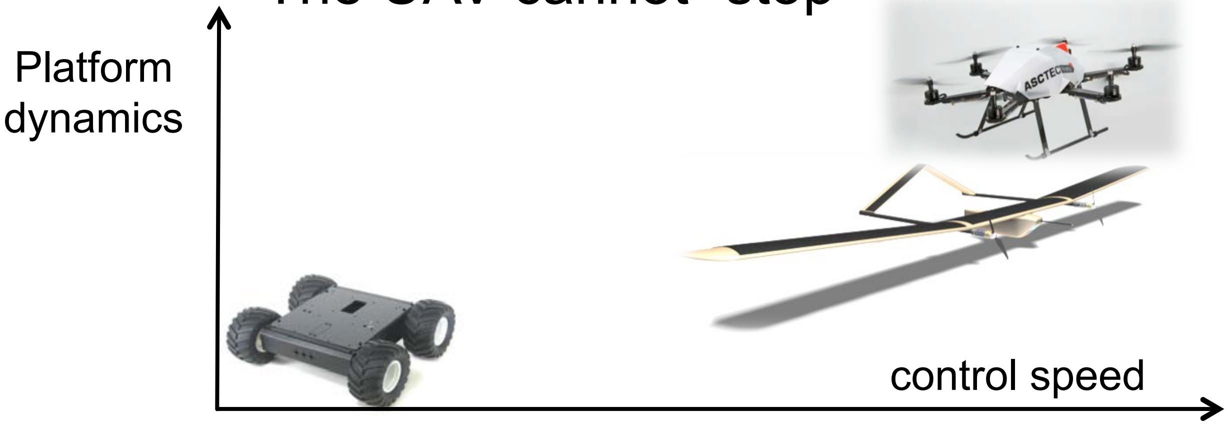
- Lightweight & safe(r) ⇨ easily deployable than larger robots
- Limited payload (<500g): 10g needs approx. 1W in hovering mode
  ➔ Limited computational power onboard ➔ choose sensors with high information density

Autonomy

- Low bandwidth / unreliable data links
  ⇨ onboard processing
- Limited battery life (~10mins)

Agility

- Highly agile (up to 8m/s)
- Fast, unstable dynamics
- High-rate real-time state estimation.
  The UAV cannot "stop"

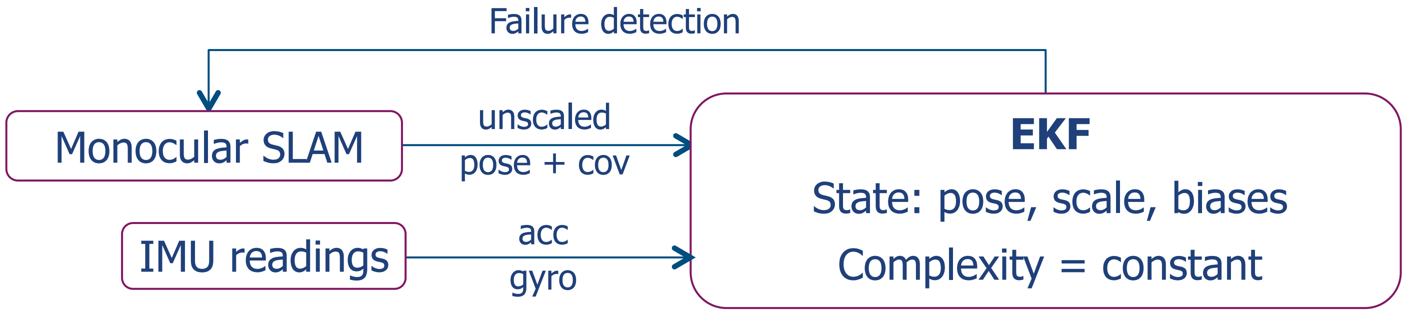


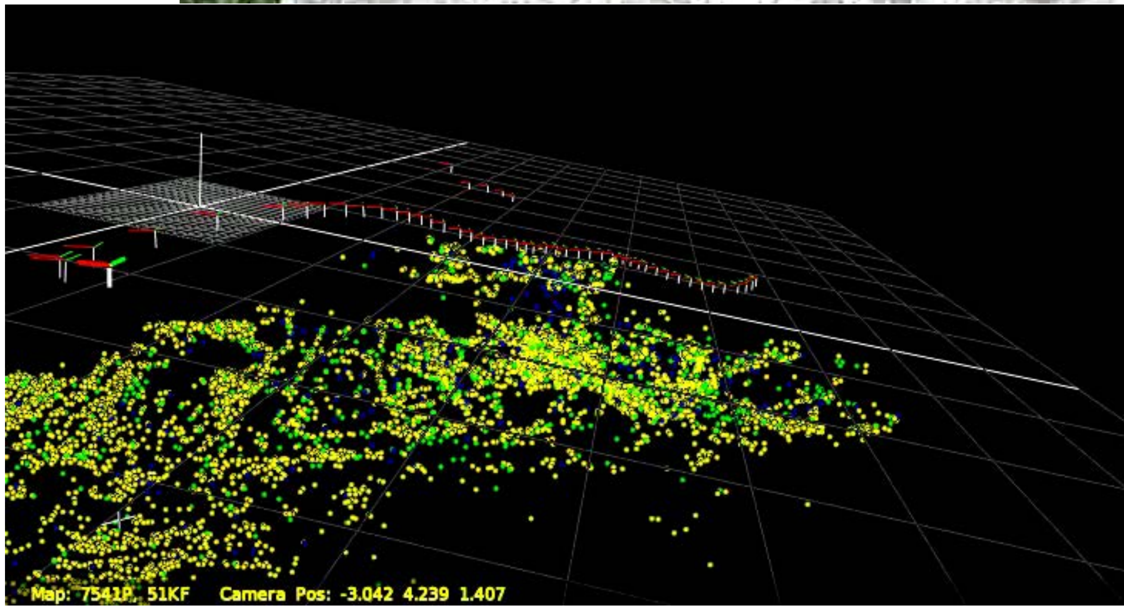
sFly | enabling UAV navigation

aim: autonomous vision-based flights in unknown environments

approach: minimal sensor setup
→ essentially fuse visual & inertial cues

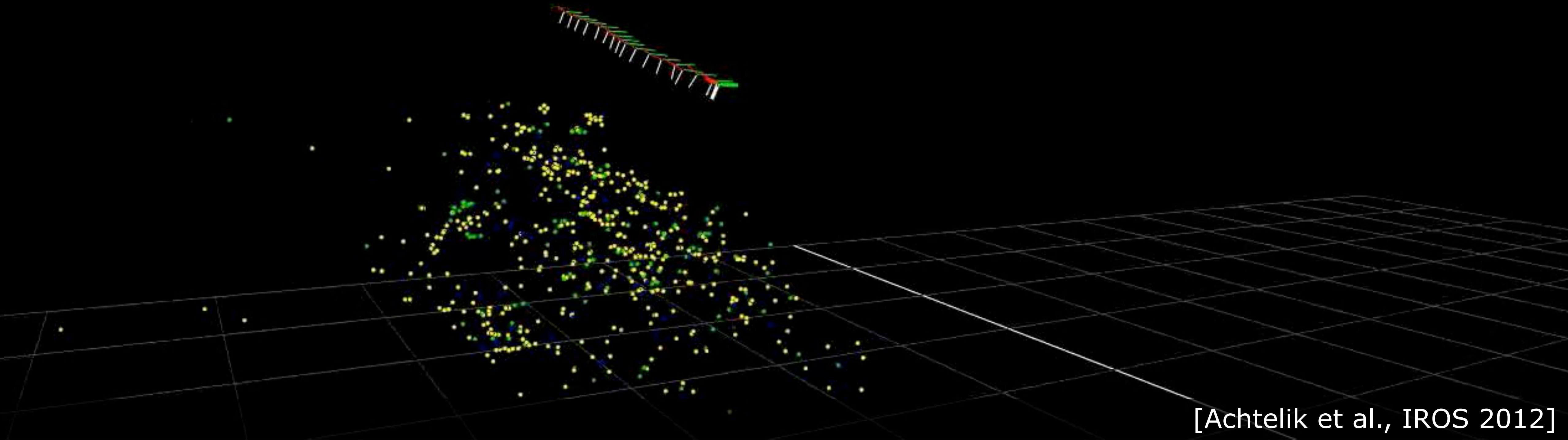
- **Downward-looking camera**: bearing only measurements
  → Monocular SLAM (based on PTAM)
- **IMU:** Acceleration & angular velocity
- Loosely-coupled visual-inertial fusion





[Weiss et al., JFR 2013]

Flights controlled using visual & inertial cues



[Achtelik et al., IROS 2012]

ETH zürich

Vision-based UAV navigation

V4RL VISION FOR ROBOTICS LAB

- First UAV system capable of vision-based flights in such real scenarios
- Publicly available framework used by NASA JPL, UPenn, MIT, TUM,…

Are we there yet?

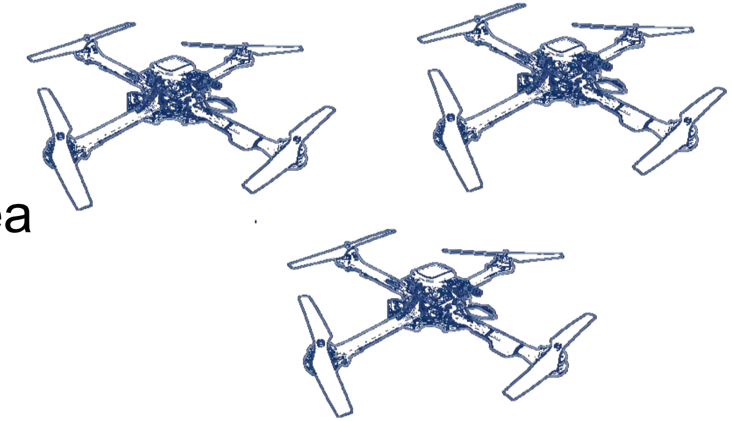
- Enhance efficiency, robustness for increased autonomy & deployability
- Follow-up directions to explore swarm and multi-robot behavior.

Photo credit: Francois Pomerleau

Vision-based Robotic Perception | what next?

V4RL VISION FOR ROBOTICS LAB

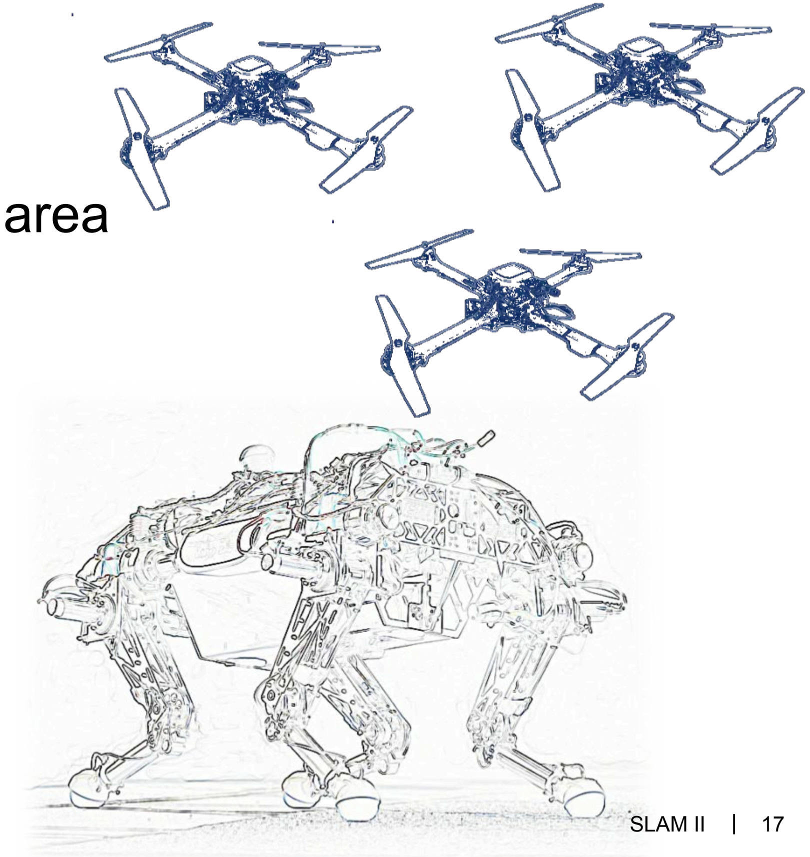
- Employ team of robots equipped with cameras
- Develop visual perception & intelligence to:
  - Navigate autonomously
  - Collaboratively build a 3D reconstruction of the surrounding area



Vision-based

Robotic Perception
| what next?

- Employ team of aerial robots equipped with cameras
- Develop visual perception & intelligence to:
  - Navigate autonomously
  - Collaboratively build a 3D reconstruction of the surrounding area

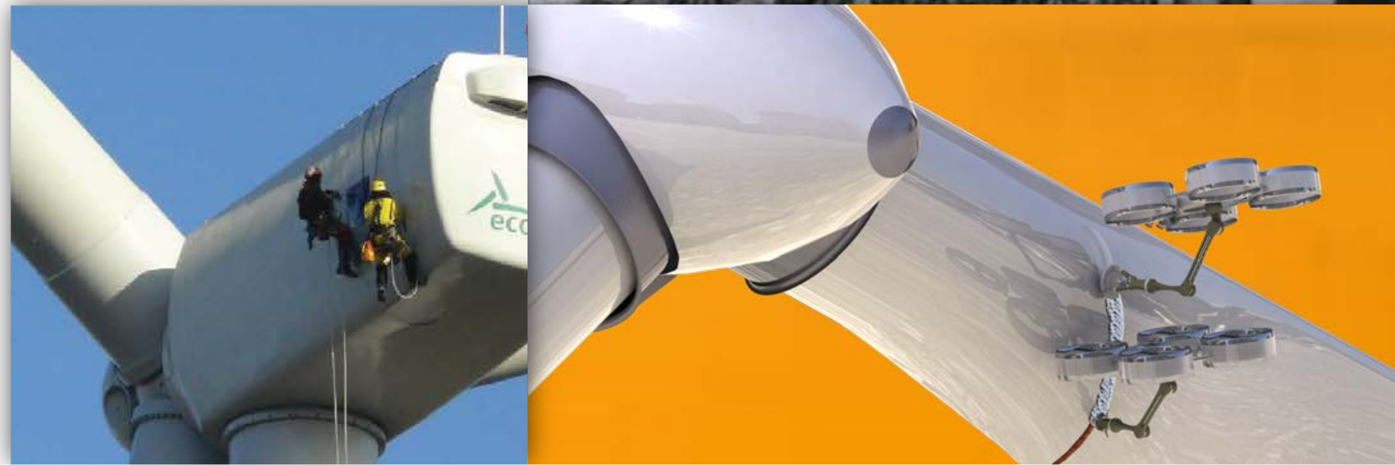


UAVs : Unmanned Aerial Vehicles

- Agile, easy access to scene overview & remote areas
- Dynamics hard to track, limited payload
  ⇨ *collaboration* is key to efficient sensing & processing
- Extension to additional platforms

AEROWORKS | EU project

- Team of small UAVs: each equipped with visual & inertial sensors and a manipulator
- **Aim**: collaboratively perceive the environment, develop autonomy in navigation and coordination to perform a common manipulation task
- V4RL: collaborative vision-based perception for navigation & 3D reconstruction
- 2015-2018, 9 partners



ICARUS| EU project

- **I**ntegrated **C**omponents for **A**ssisted **R**escue and **U**nmanned **S**earch operations (2012-2015), budget: 17.5 M€, 24 partners
- Search-and-rescue combining robotics for land, sea and air
- ETHZ: map generation, people detection, … from a UAV

SHERPA | EU project

- **S**mart collaboration between **H**umans and ground-a**E**rial **R**obots for im**P**roving rescuing activities in **A**lpine environments
- 11 M€, 10 partners, 2013-2017
- Sensor fusion (visible light and thermal cameras, IMU, …) for robust SLAM, environment reconstruction & victim localization

Vision-based Robotic Perception | the challenges

Challenge I: High-fidelity localization & mapping

- The backbone of perception of space & navigation autonomy

- Pioneering work in UAV navigation, but lacks scalability, robustness and deployability for application in real scenarios

Suitable keypoint detection/description

- Research into image keypoints suitable for robotics applications: for fast & robust detection and matching
- Rotation-, scale-invariant keypoints
- Binary descriptor: e.g. BRISK, ORB, BRIEF & variants

Descriptor	Run time [ms.]
SURF	117.1
SIFT	448.6
BRIEF	**3.8**
BRISK	**10.6**
ORB	**4.2**

BRISK descriptor

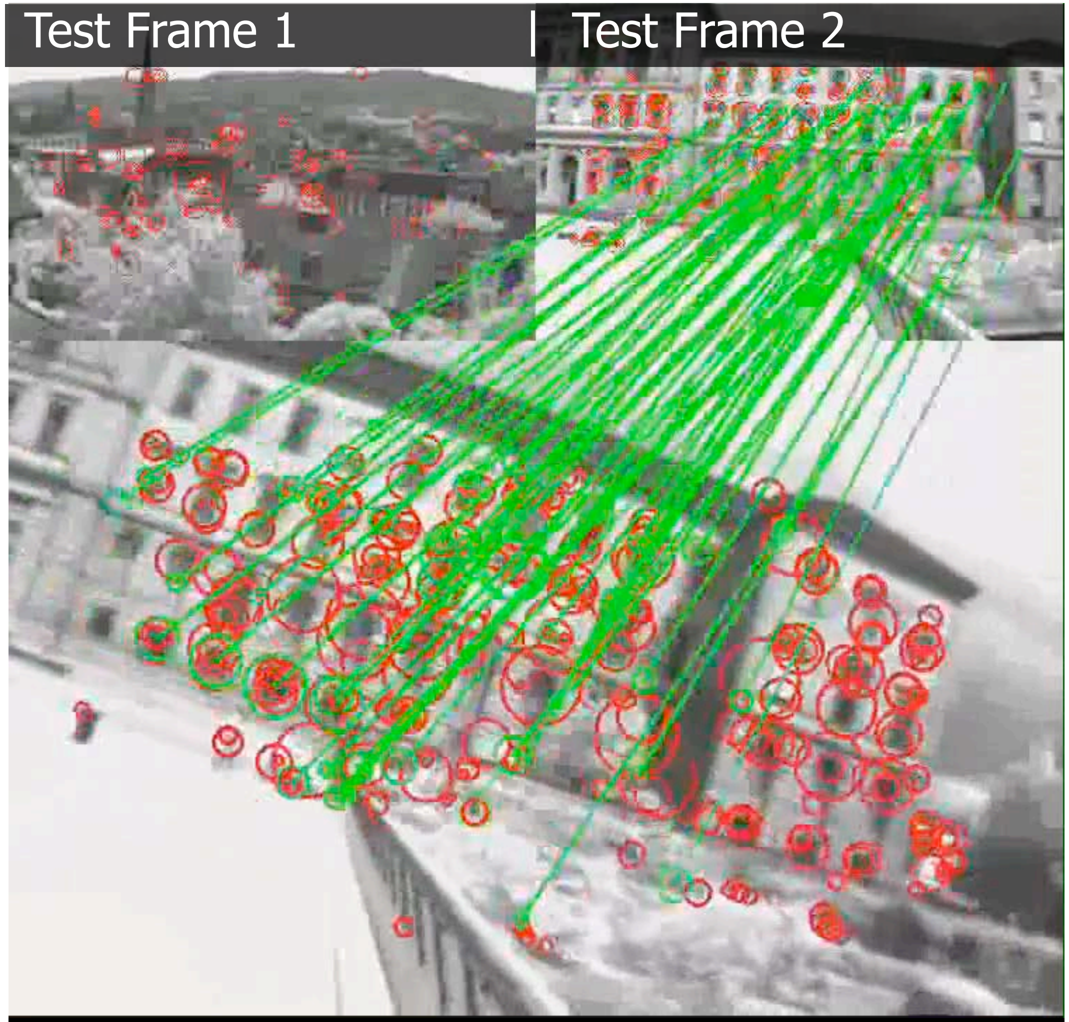
1	0	0	1	1	1	...	0	0	1	0	0	1	1	1	0

SURF descriptor

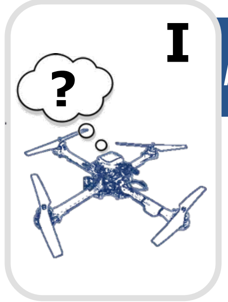
0.4	1.0	0.1	...	0.3	0.4	0.7	0.6	0.11	2	5	0.1	0.7

BRISK:

- Precision-Recall: comparable to SIFT & SURF
- ~10x faster than SURF
- Open-source, features in OpenCV



[Leutenegger et al., ICCV 2011]



Sensor fusion for SLAM

- Visual-Inertial sensor:
  HW-synced stereo camera (global shutter) + IMU

“OKVIS”: visual-inertial SLAM

- Tight visual & inertial fusion: replace motion model with IMU constraints on the actual motion
- Visual cues: very descriptive, but sensitive to motion blur, lighting conditions…
- Inertial cues: accurate estimates for short-term motions, unsuitable for longer-term
- Open-source: http://ethz-asl.github.io/okvis_ros/

[Leutenegger et al., RSS 2013]

Event-based Cameras for Robot Navigation

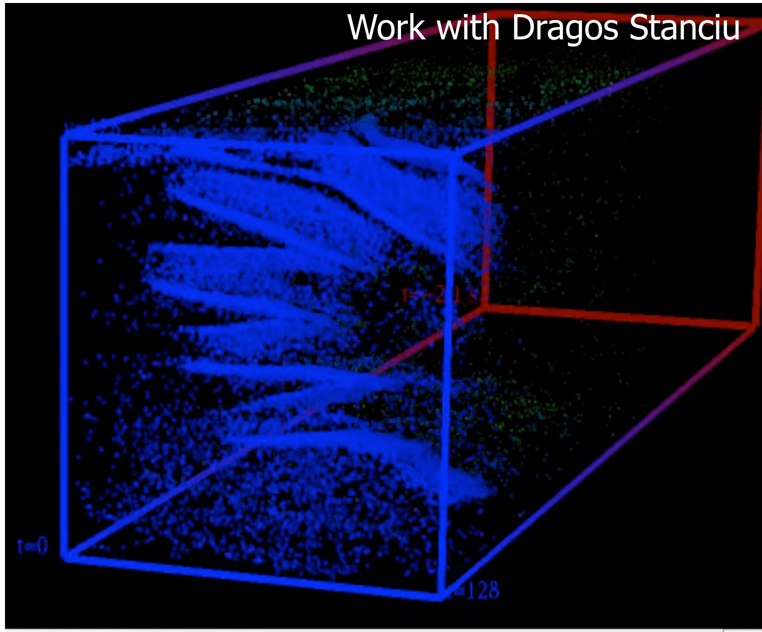
- Dynamic Vision Sensor (DVS)

iniLabs

- Similar to the human retina:

captures intensity changes asynchronously instead of capturing image frames at a fixed rate

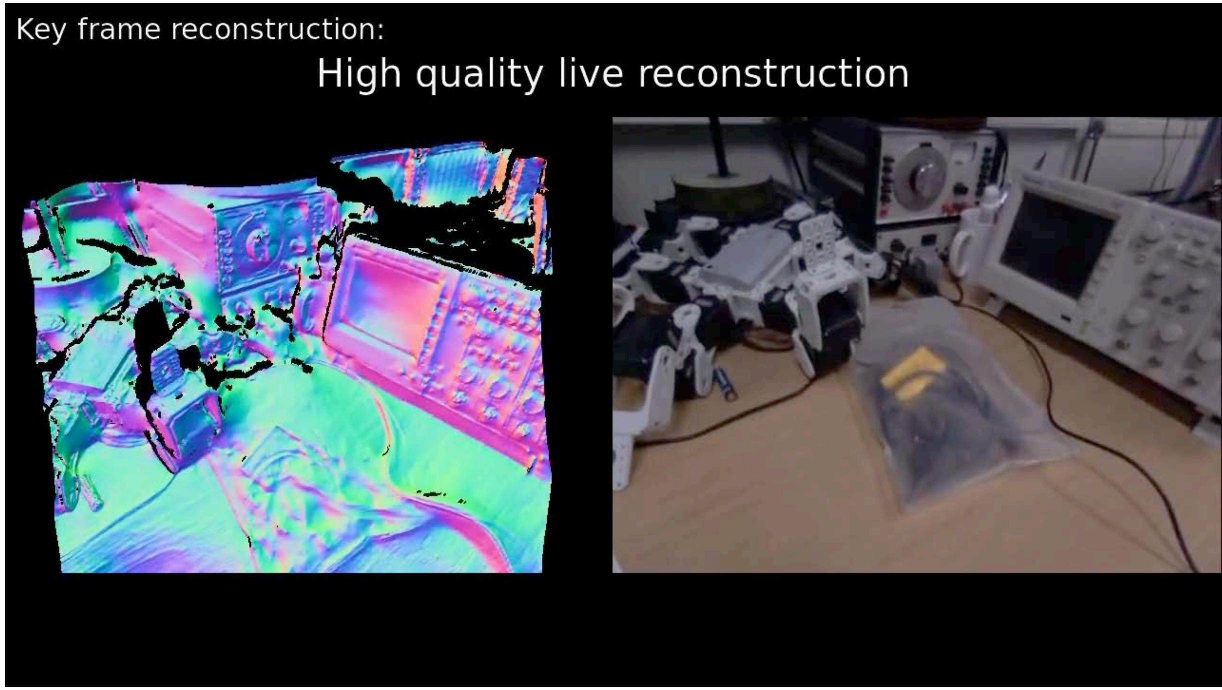
  - ✓ Low power
  - ✓ High temporal resolution ➔ tackle motion blur
  - ✓ High dynamic range



Vision-based Robotic Perception | the challenges

Challenge II: Dense scene reconstruction

- Vital for robot interaction with its environment

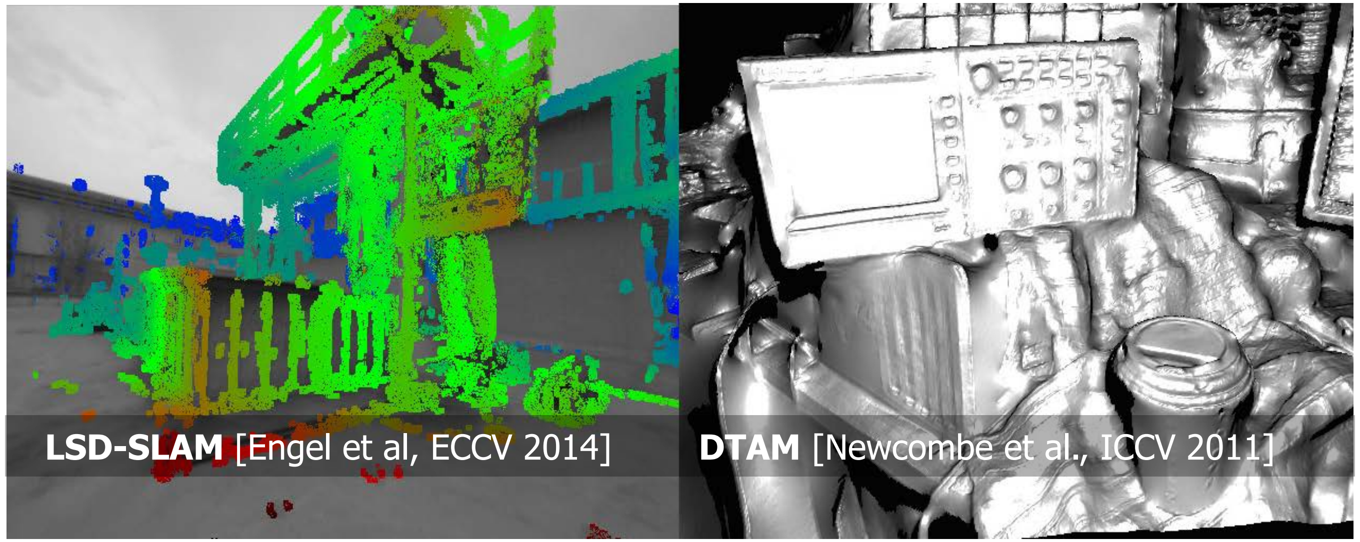


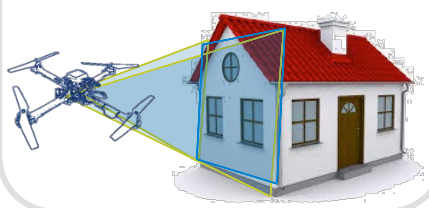
DTAM
[Newcombe et al.,
ICCV 2011]

Vision-based Robotic Perception | the challenges

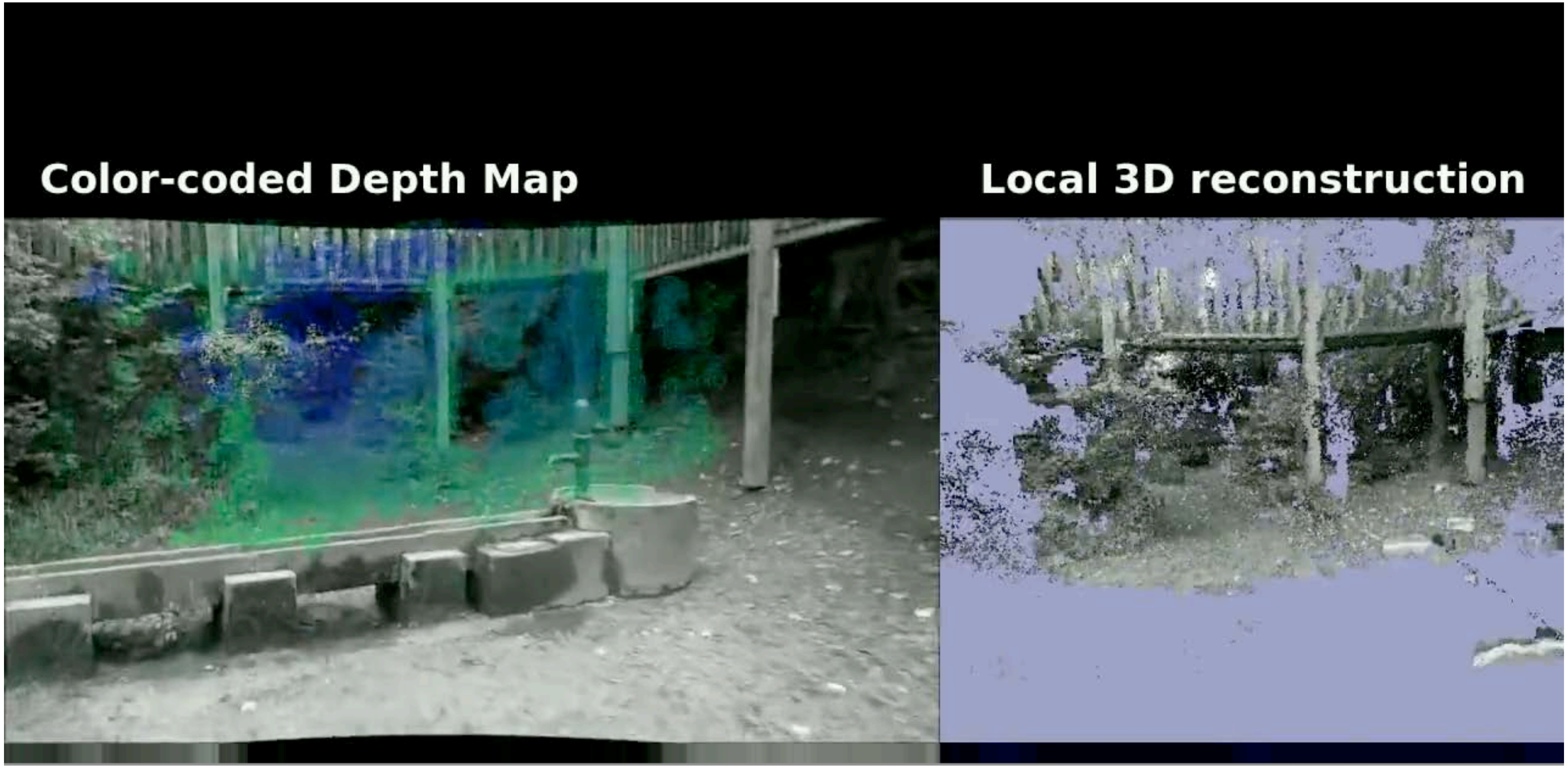
Challenge II: Dense scene reconstruction

- Vital for robot interaction with its environment
- Trade-off: level of detail vs. computational cost
- Work towards both
  (a) online onboard and
  (b) scalable offboard functionality

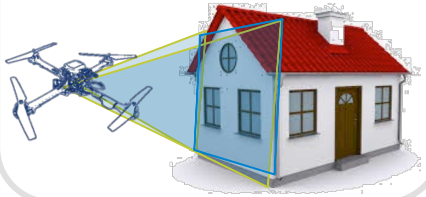




Towards low-cost, denser 3D reconstruction with a single camera

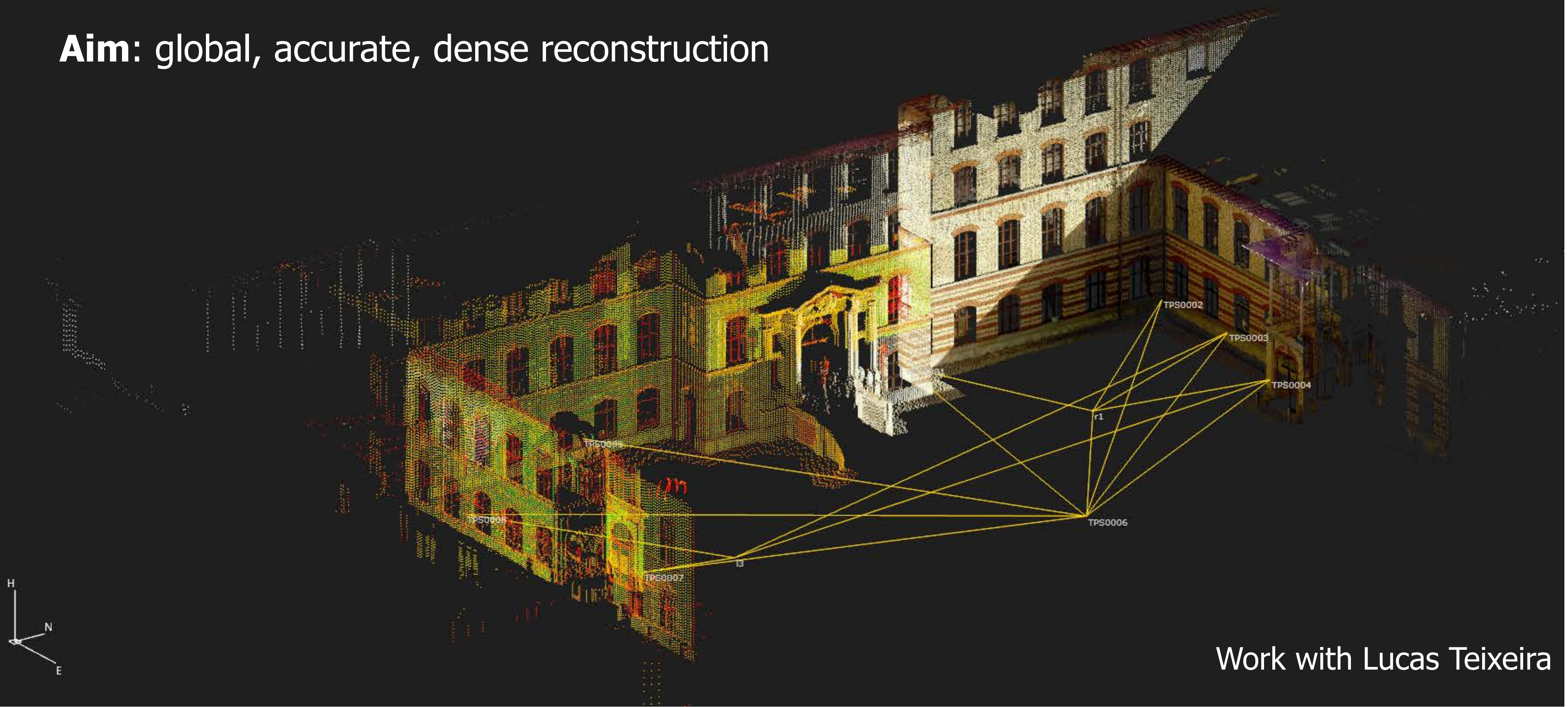


Work with Lucas Teixeira



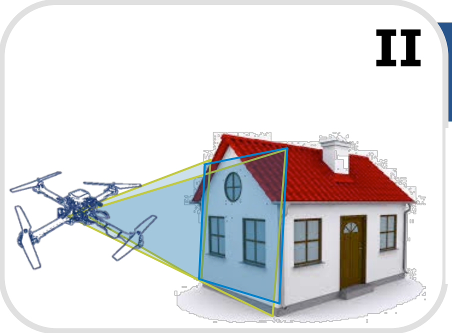
Towards low-cost, denser 3D reconstruction with a single camera

Aim: global, accurate, dense reconstruction

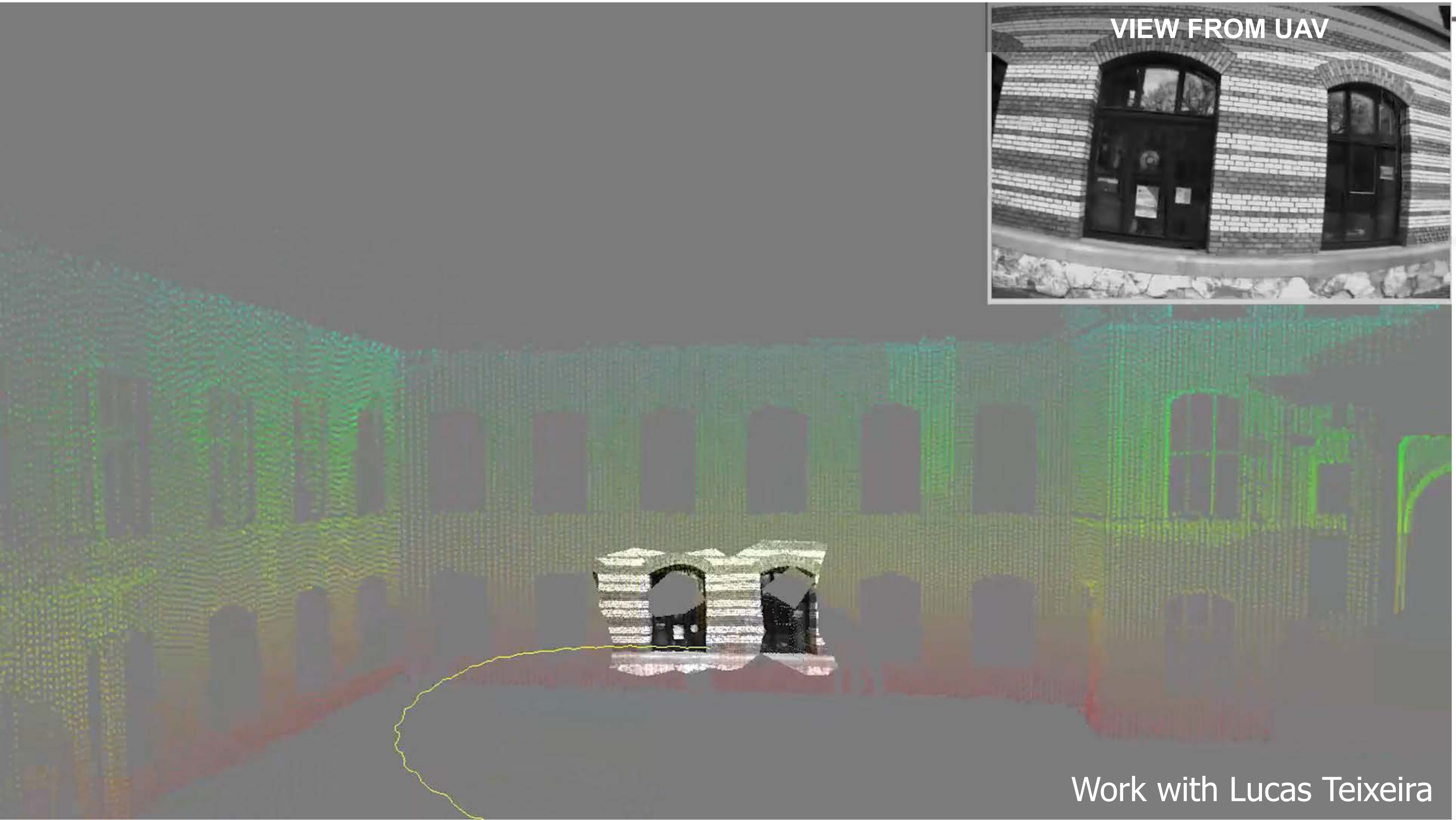


Work with Lucas Teixeira

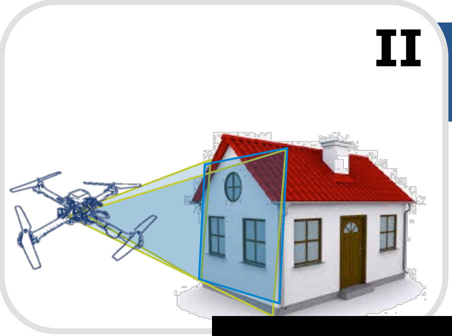
Leica Theodolite Total Station



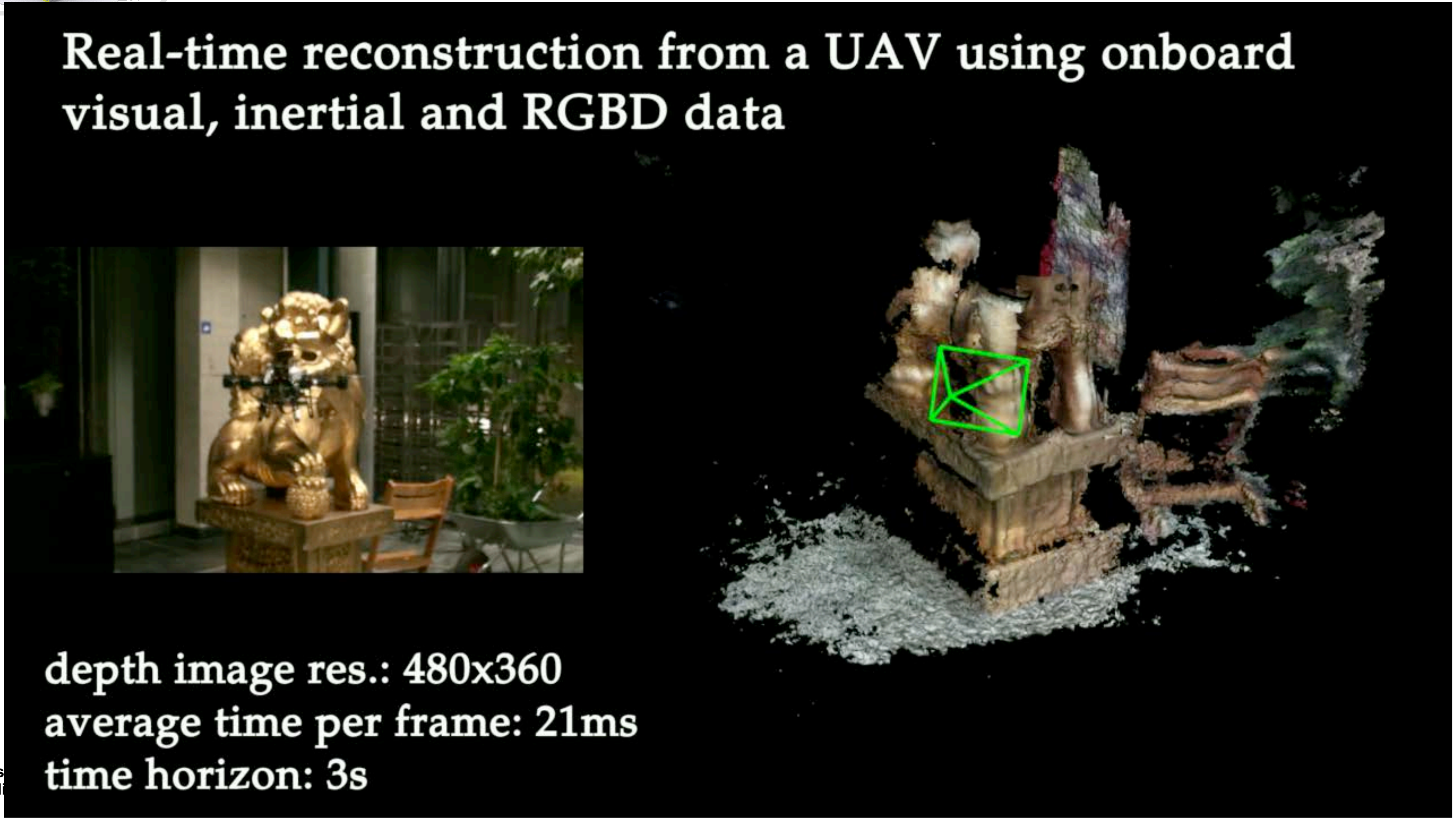
Towards low-cost, denser 3D reconstruction with a single camera from a small UAV



- Monocular-inertial SLAM
- Denser representation in < 8ms per frame



Real-time Dense Surface Reconstruction for Manipulation

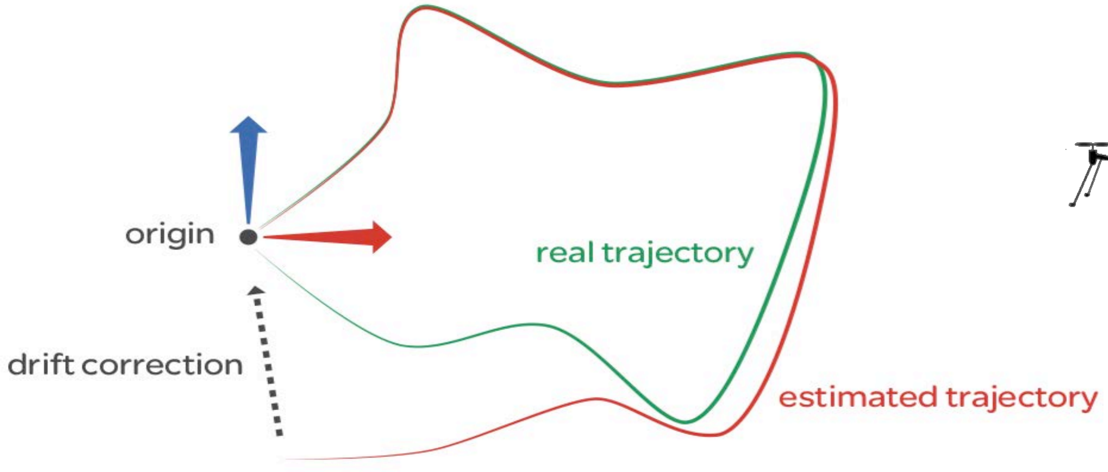


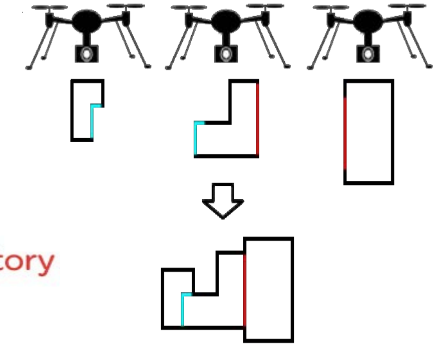
Work with
Marco Karrer &
Mina Kamel

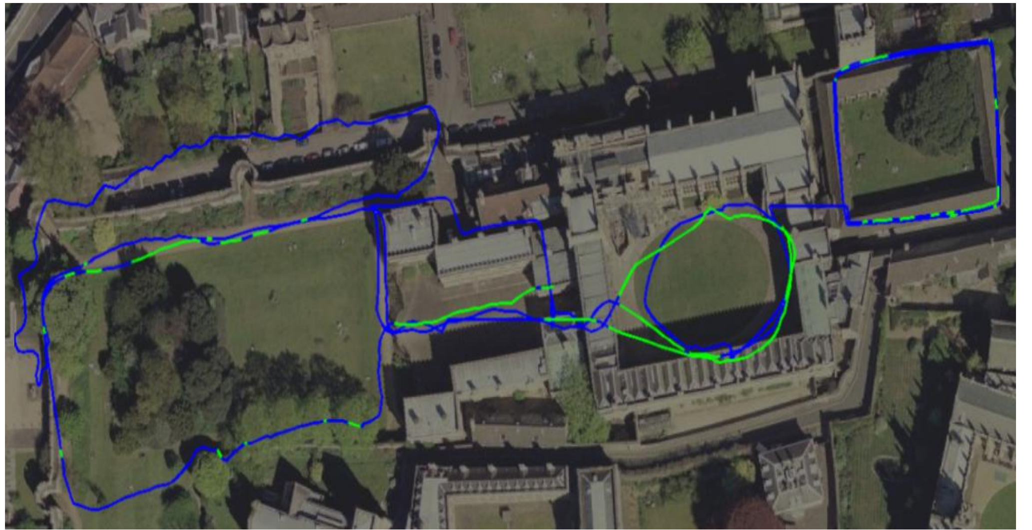
Vision-based Robotic Perception | the challenges

Challenge III: Place recognition

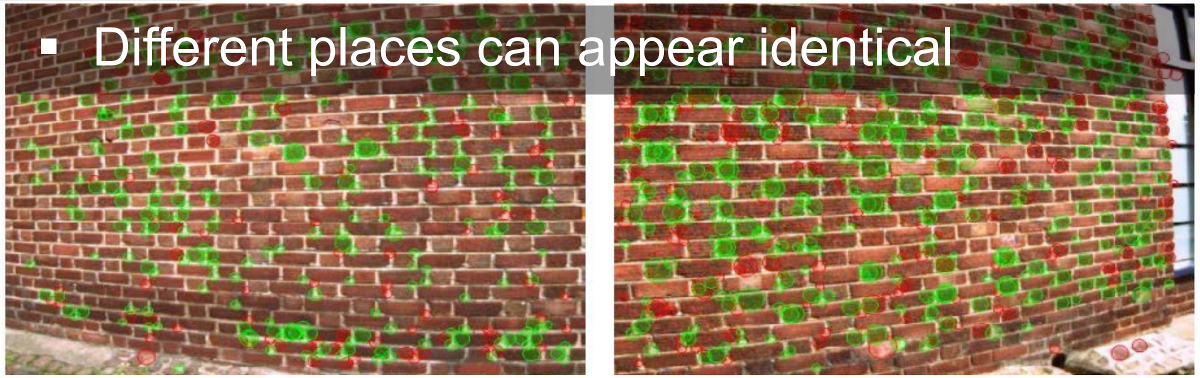
- Recognising when the robot visits a “known” location for:
  - Drift Correction
  - Trajectory / map merging





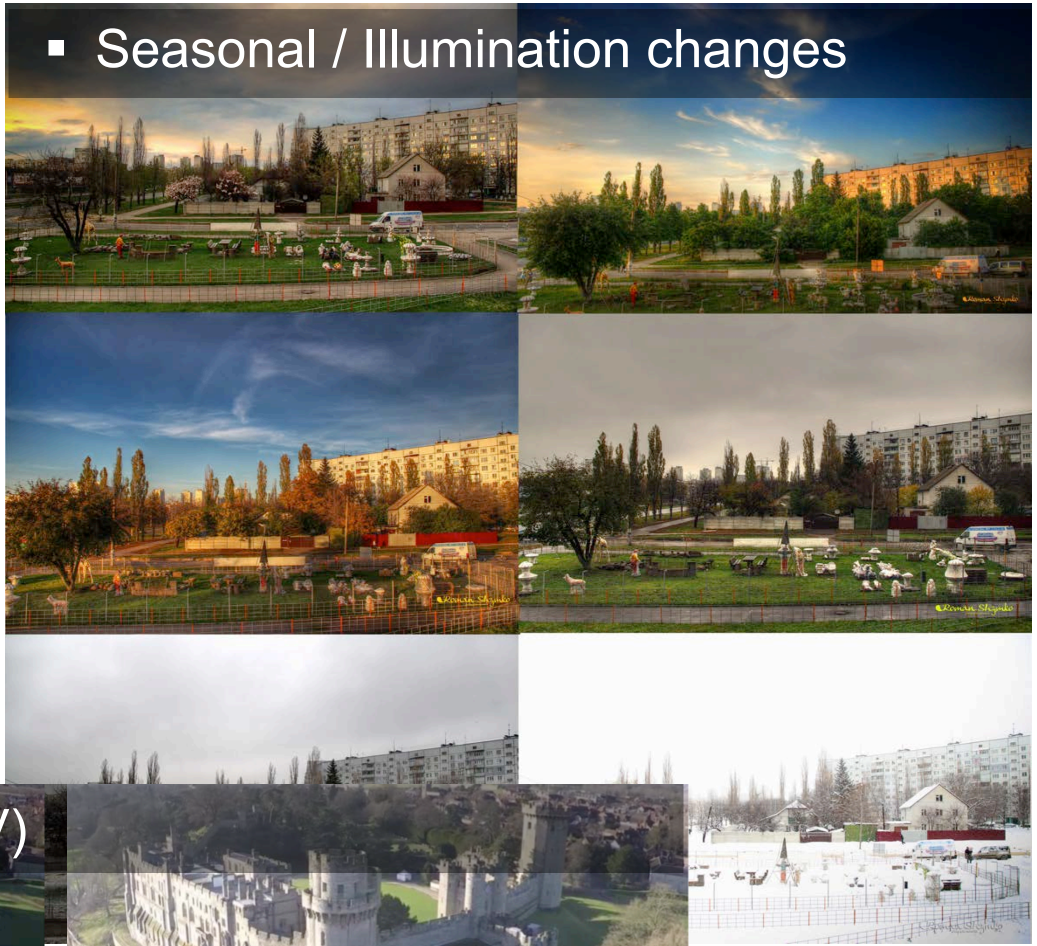


Vision-based Place recognition: common problems

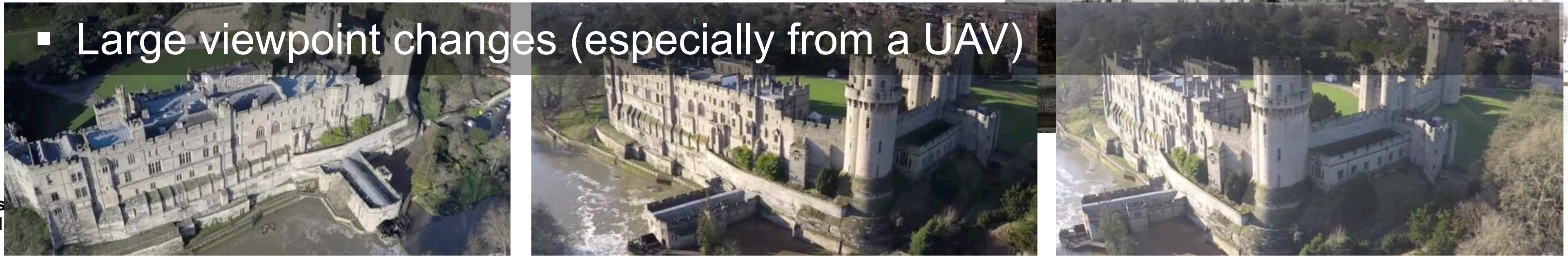


- Different places can appear identical

- Place appearance changes between visits



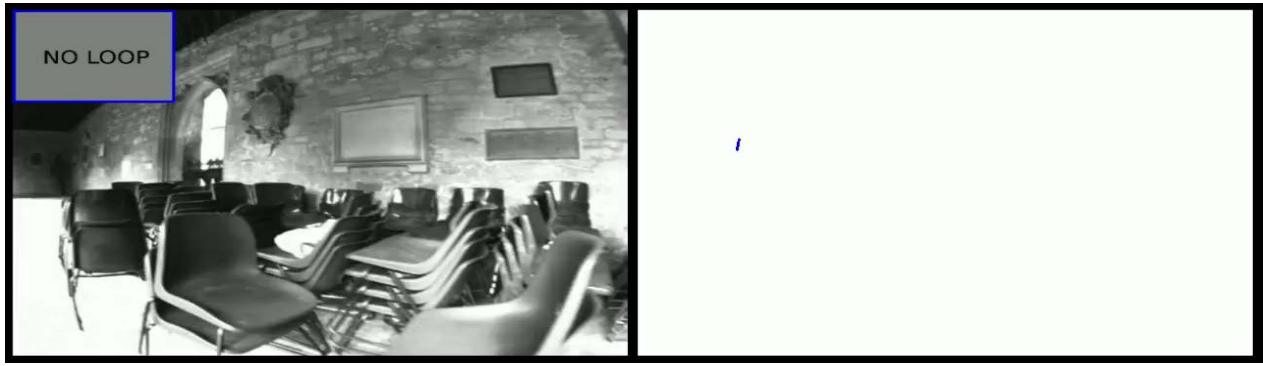
- Seasonal / Illumination changes



- Large viewpoint changes (especially from a UAV)

Place recognition for UAV navigation

- FABMAP 2.0 [Cummins, Newman IJRR 2011]
  Extracts features & Clusters: “Bag of Words”
  - ✓ Robust to perceptual aliasing
  - x Computationally expensive
  - x Low tolerance to viewpoint change



0.3	0.5	0.71	0.4	1.0	0.1	SURF descriptor
1	0	0	1	1	1	BRISK descriptor

Descriptor	Run time [ms.]
SURF	117.1
SIFT	448.6
BRIEF	**3.8**
BRISK	**10.6**
ORB	**4.2**

Boosting efficiency:

- Replace expensive features: use binary features (e.g. BRISK) – How to cluster binary strings?

Tackling recognition from different viewpoints:

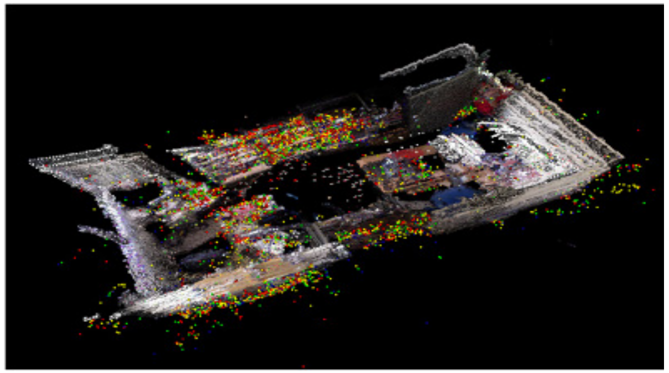
- Consider world as a continuous constellation of landmarks/visual words
- Use covisibility of feautures to identify potential location matches
- Use geometric constraints

Work with Fabiola Maffra & Elena Stumm

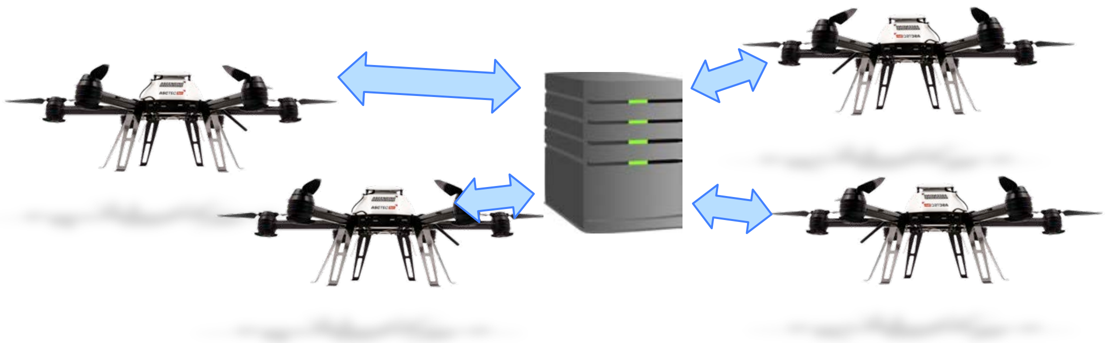
Vision-based Robotic Perception | the challenges

Challenge IV: Collaborative robot sensing & mapping

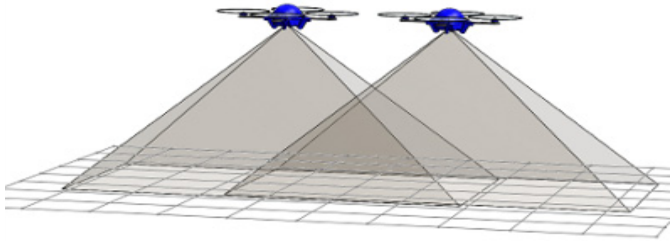
- Exploit presence of multiple UAVs (occlusions, accuracy, time efficiency)
- Very little existing research



Collaborative reconstruction using 2 RGBD cameras
[Riazuelo et al, RAS 2014]



- Flight-critical tasks on client
- Computationally expensive tasks on server
- What information needs to be shared?



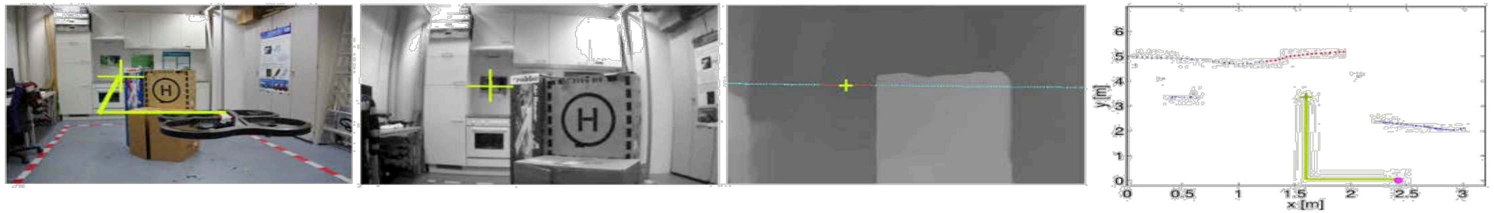
Variable-baseline stereo from 2 UAVs
[Achtelik et al, IROS 2011]

Work with Patrik Schmuck
& Marco Karrer

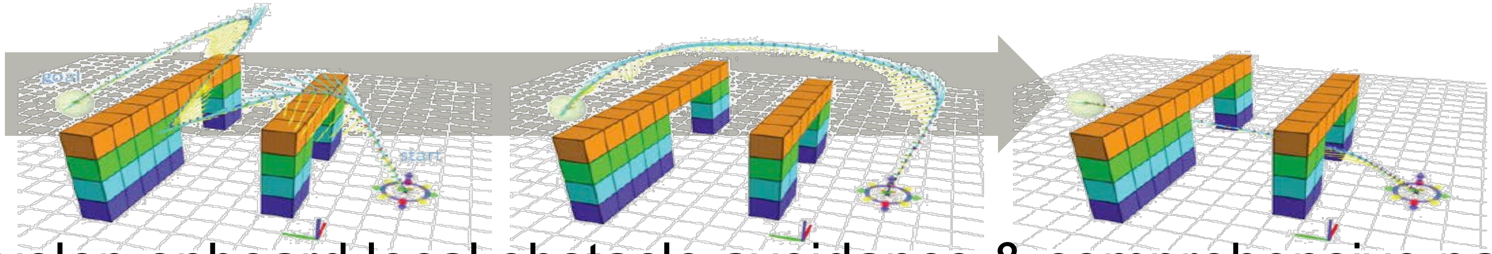
Vision-based Robotic Perception | the challenges

Challenge V: Navigation Strategies – obstacle avoidance & path planning

- Complete the navigation loop
- Existing: mostly off-board solutions



[Alvarez et al, ISER 2014]
Collision avoidance with a camera and offboard GPU processing



[Achtelik et al, JFR 2014]
Intermediate & final paths computed in simulation

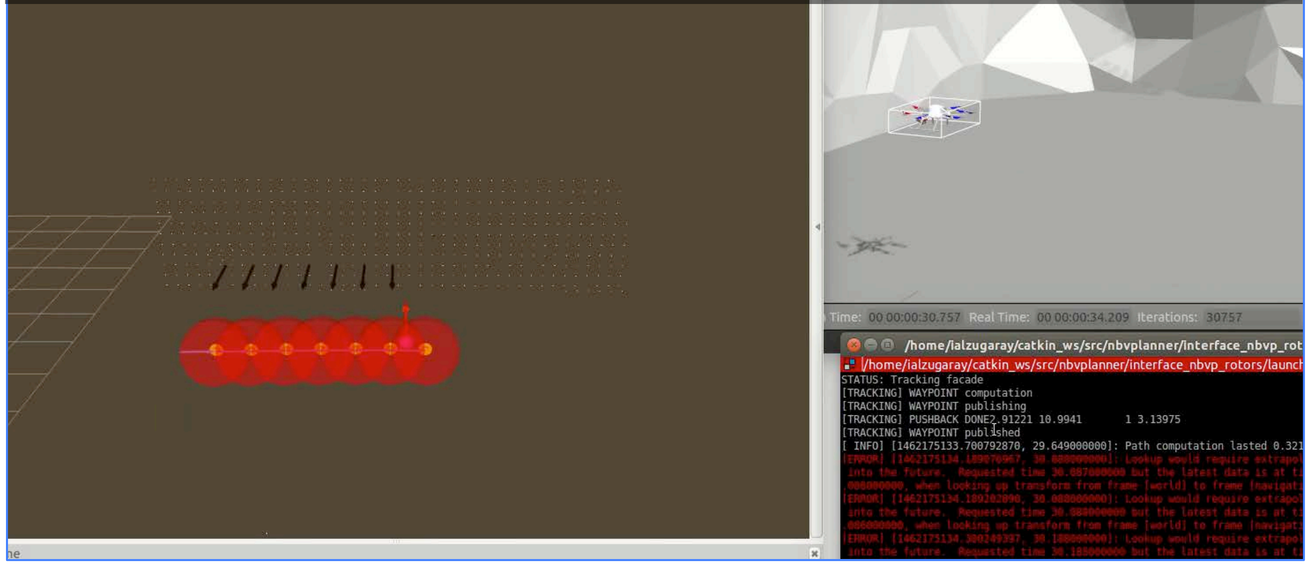
- Develop onboard local obstacle avoidance & comprehensive path planning

ETH zürich

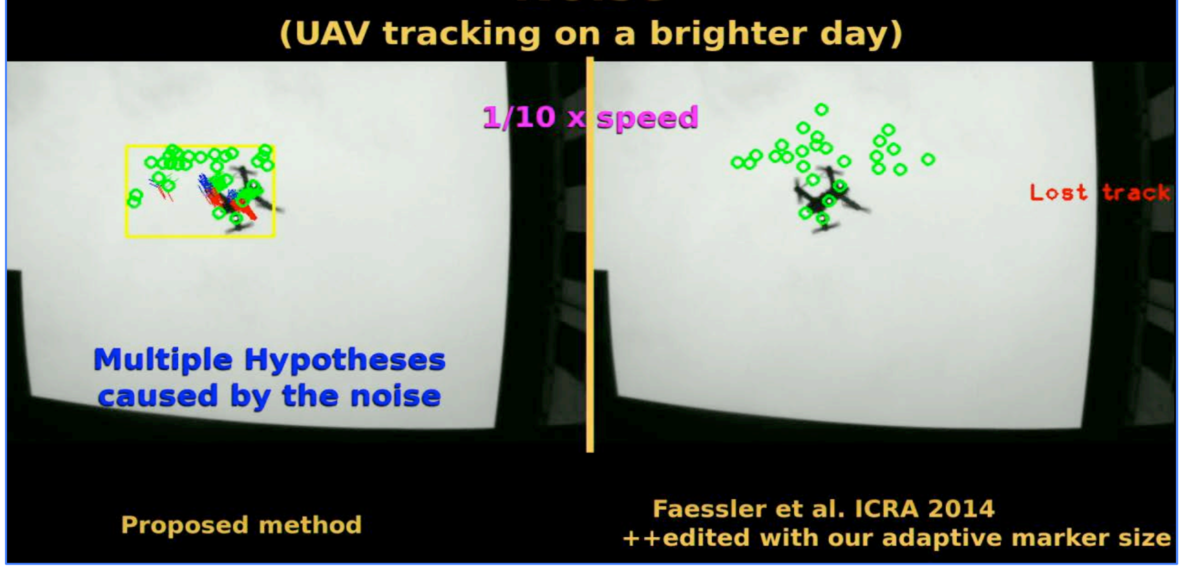
Master/Semester Projects @ V4RL

V4RL VISION FOR ROBOTICS LAB

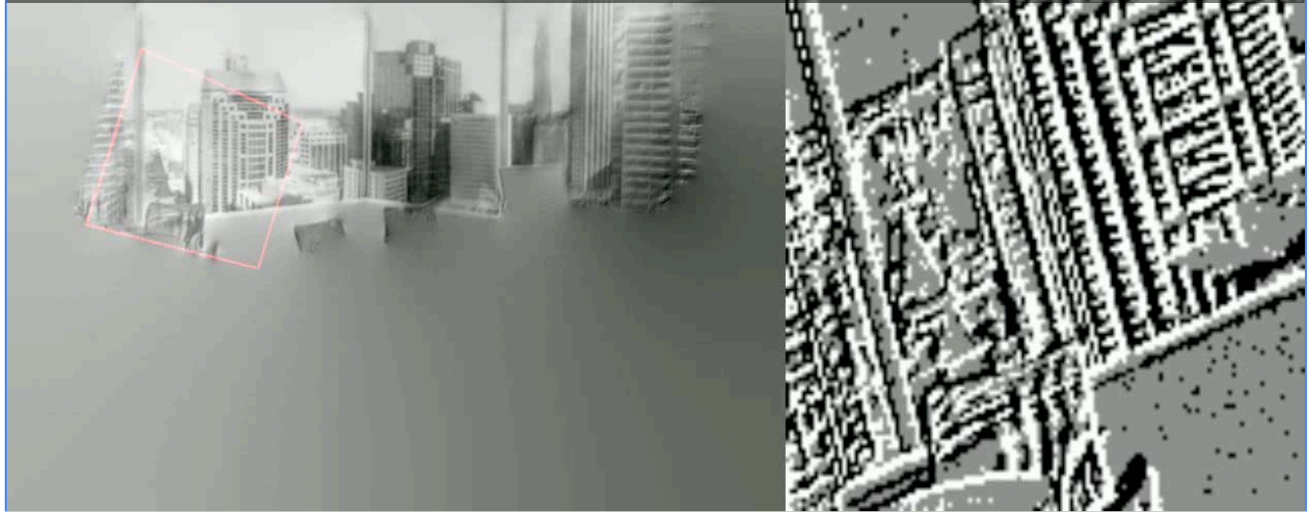
UAV Path planning in the loop of mono-inertial navigation (Ignacio Alzugaray)



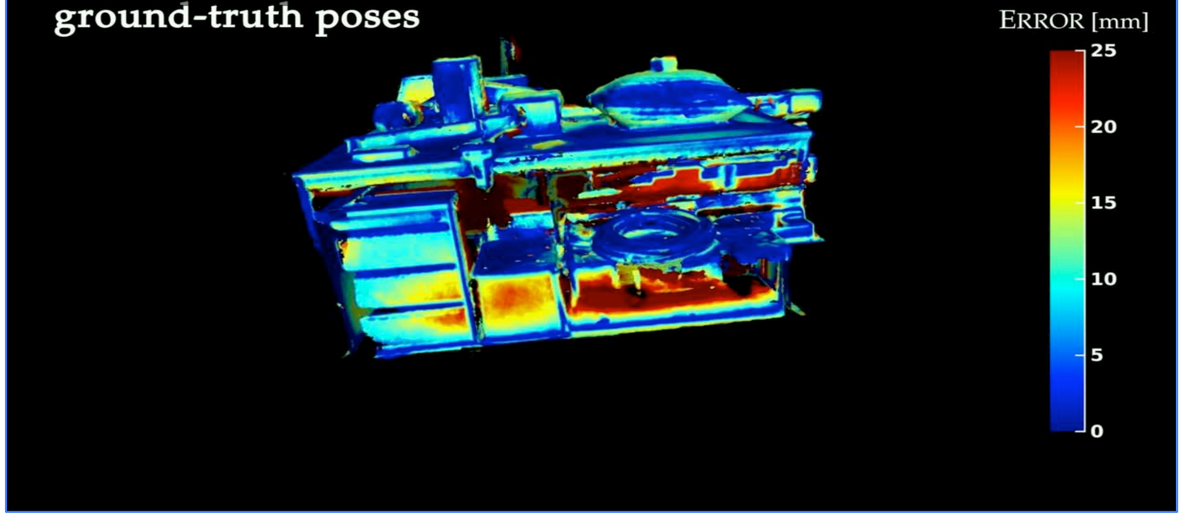
Real-time pose tracking with an external camera (Marco Moos)



Scene Reconstruction from a DVS camera (Wilko Schwarting)



Dense 3D Reconstruction for Aerial Manipulation (Marco Karrer)



Conclusion & Impact

Vision-based SLAM:

- has come a long way: from handheld to vision-stabilised flights of UAVs
- key to spatial awareness of robots ⇨ bridges the gap between Computer Vision and Robotics

Perception + Collaboration are central to Robotics today:

- Large sums of research funds in the area (e.g. SHERPA €11M, ICARUS €17.5M)

- Still work to be done before robots are ready for real missions
- Potential for great impact in the way we perceive/employ robots today