ETH zürich

SLAM II: from SLAM to robotic perception

Autonomous Mobile Robots

Margarita Chli
Roland Siegwart and Nick Lawrence

SLAM II | today's lecture

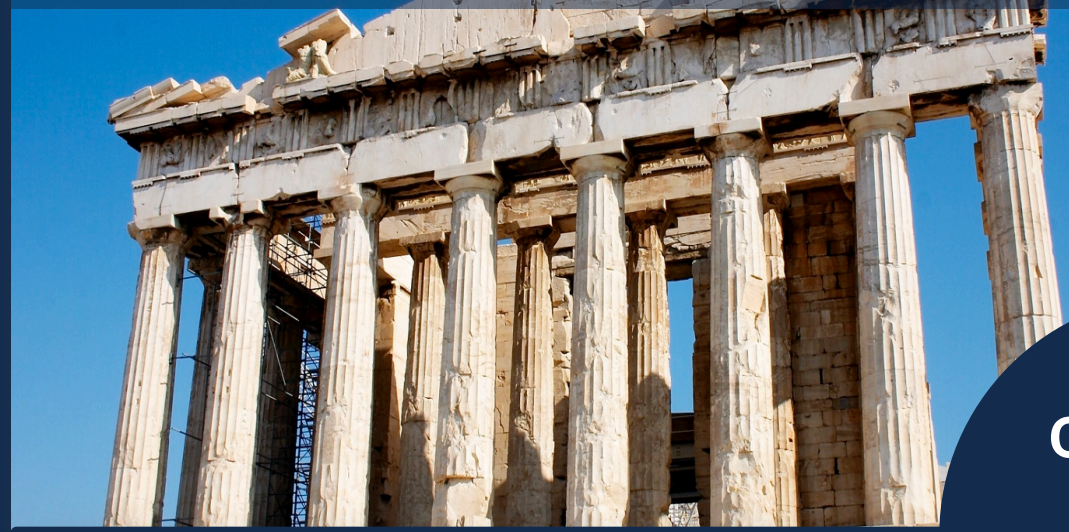
Last time: how to do SLAM?

Today: what to do with SLAM, and beyond.

- Vision-based SLAM – state of the art
- Vision-based Robotic Perception:
  - Current Challenges
  - Overview of Research Activities in V4RL


sky
clouds
statue
sunny
trees
monument
bus
people
motorist
car
roundabout
street

DIGITIZATION IN ARCHAEOLOGY



SEARCH & RESCUE

Computer Vision & Robotics

CROP MONITORING

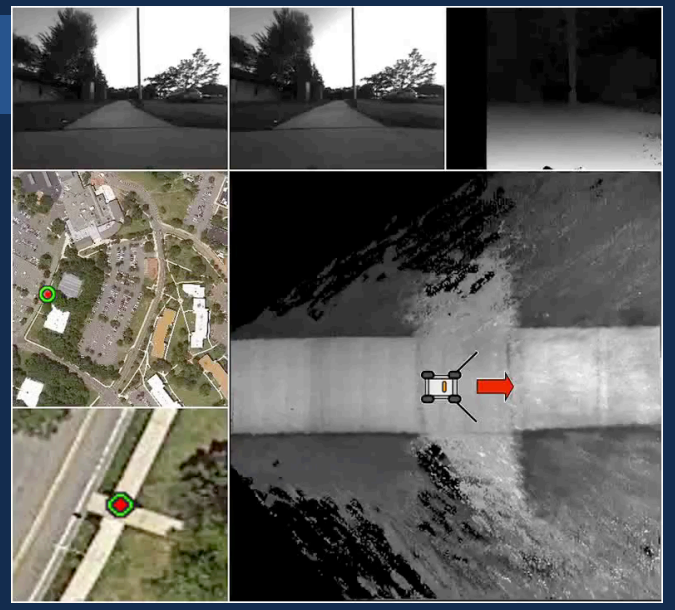
INDUSTRIAL INSPECTION

SLAM | Simultaneous Localization And Mapping

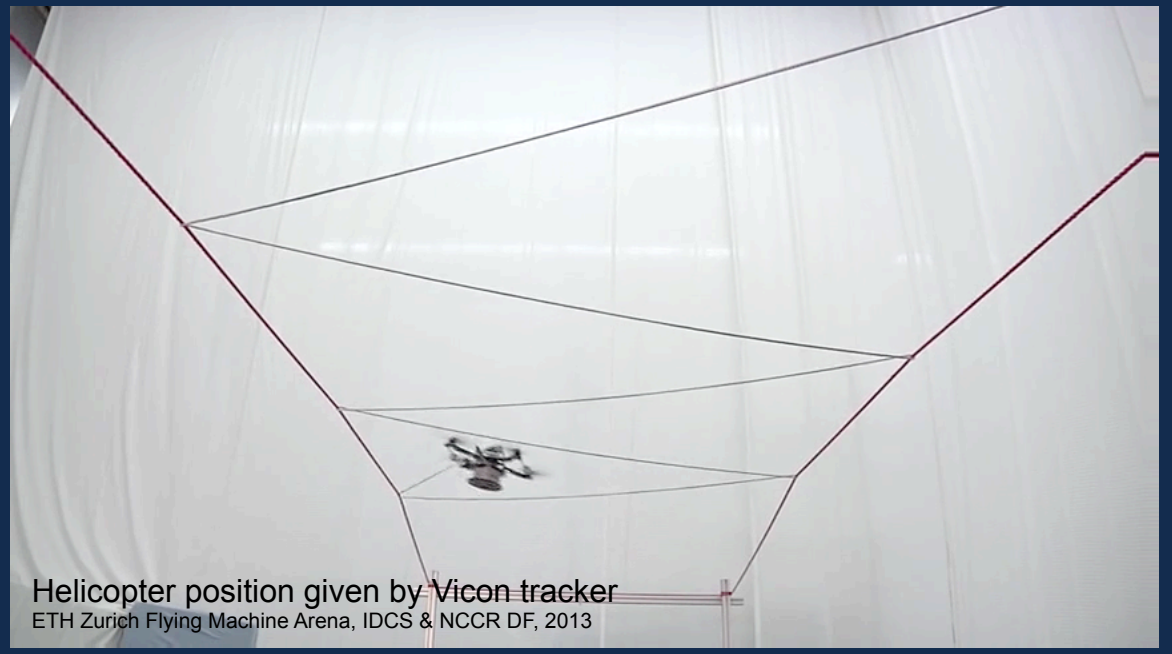
- **The backbone of spatial awareness of a robot**
- One of the most challenging problems in probabilistic robotics
- An unbiased map is necessary for localizing the robot
  Pure localization with a known map.

  SLAM: no a priori knowledge of the robot's workspace
- An accurate pose estimate is necessary for building a **map** of the environment
  Mapping with known robot poses.

  SLAM: the robot poses have to be estimated along the way

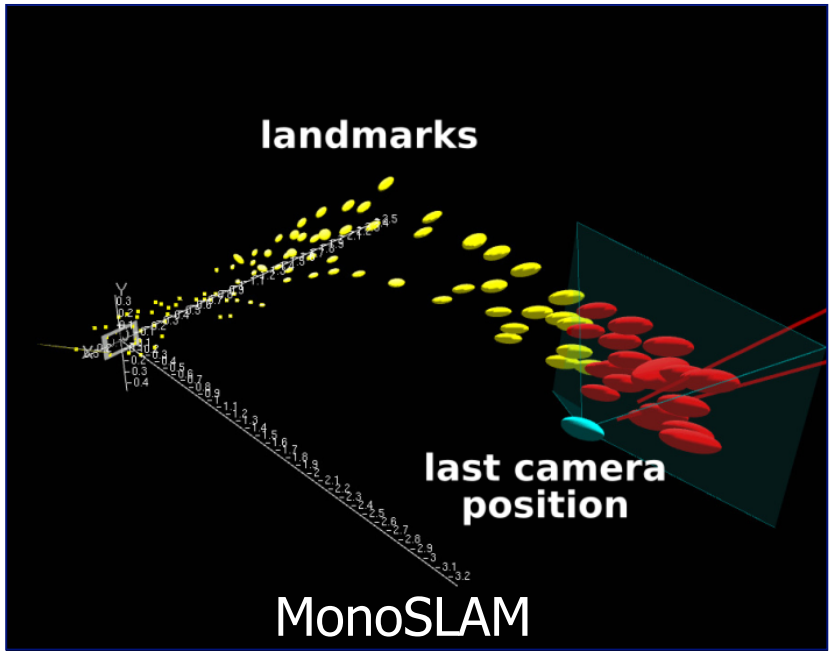


Robot localization using Satellite images
[Senlet and Elgammal, ICRA 2012]

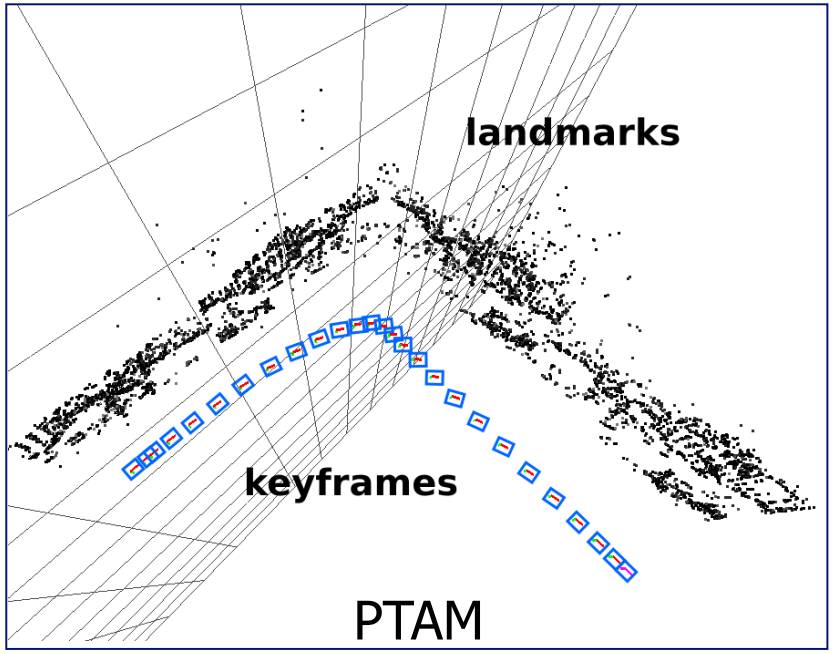


Helicopter position given by Vicon tracker
ETH Zurich Flying Machine Arena, IDCS & NCCR DF, 2013

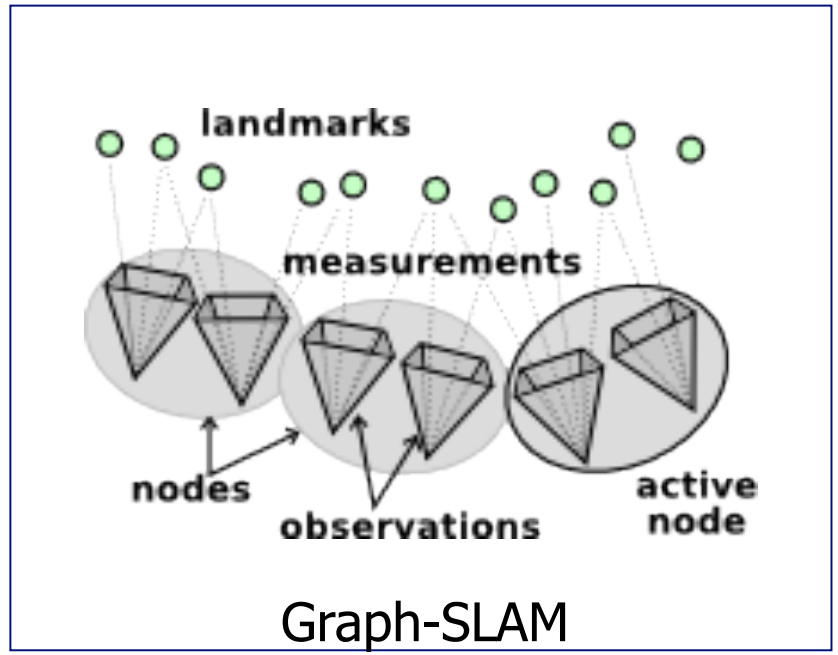
Monocular SLAM | milestone systems



[Davison et al. 2003, 2007]



[Klein, Murray 2007]



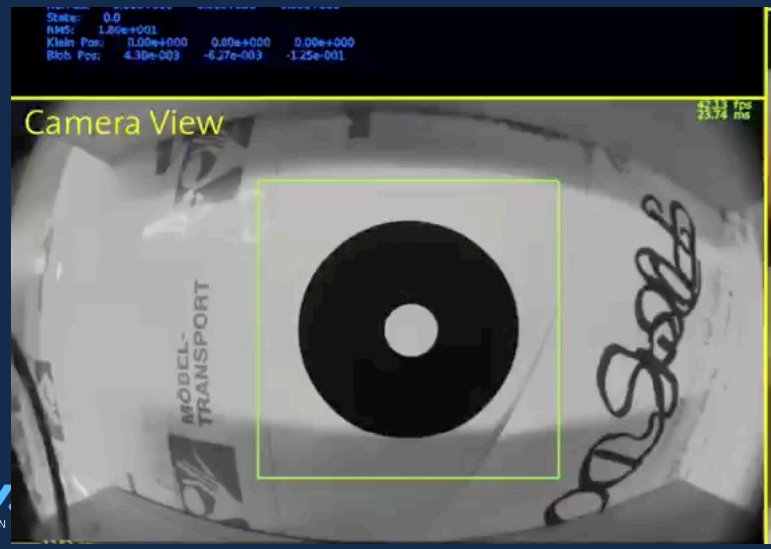
[Eade, Drummond 2007]

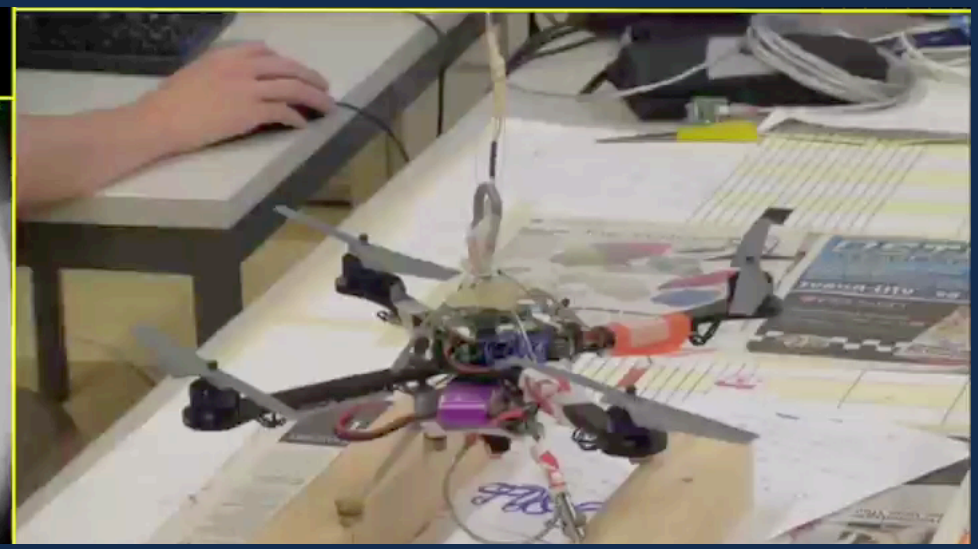
- ✓ revolutionary in the Vision & Robotics communities, but…
- ✘ not ready to perform tasks in general, uncontrolled environments

Computer Vision meets Robotics

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

2009: EU FP7, **sFly**



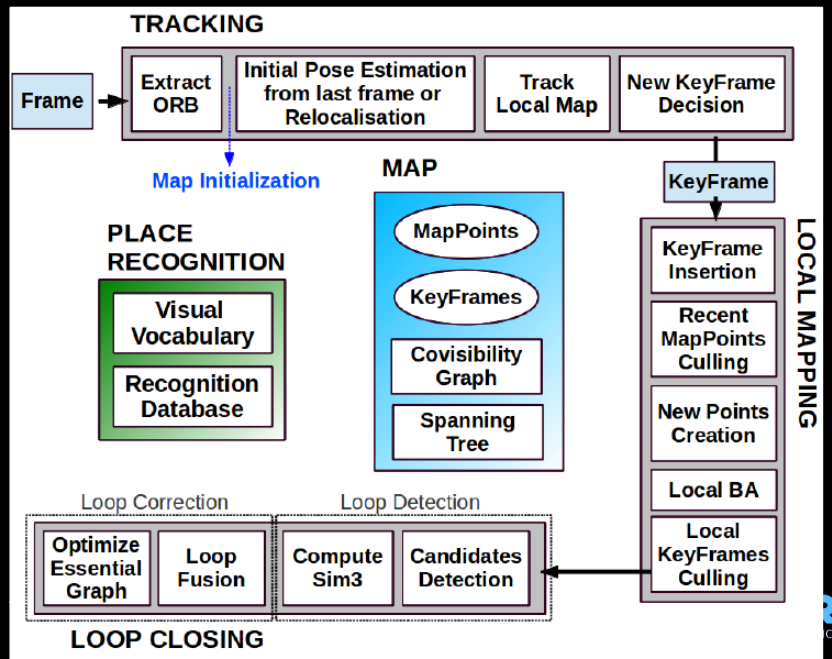


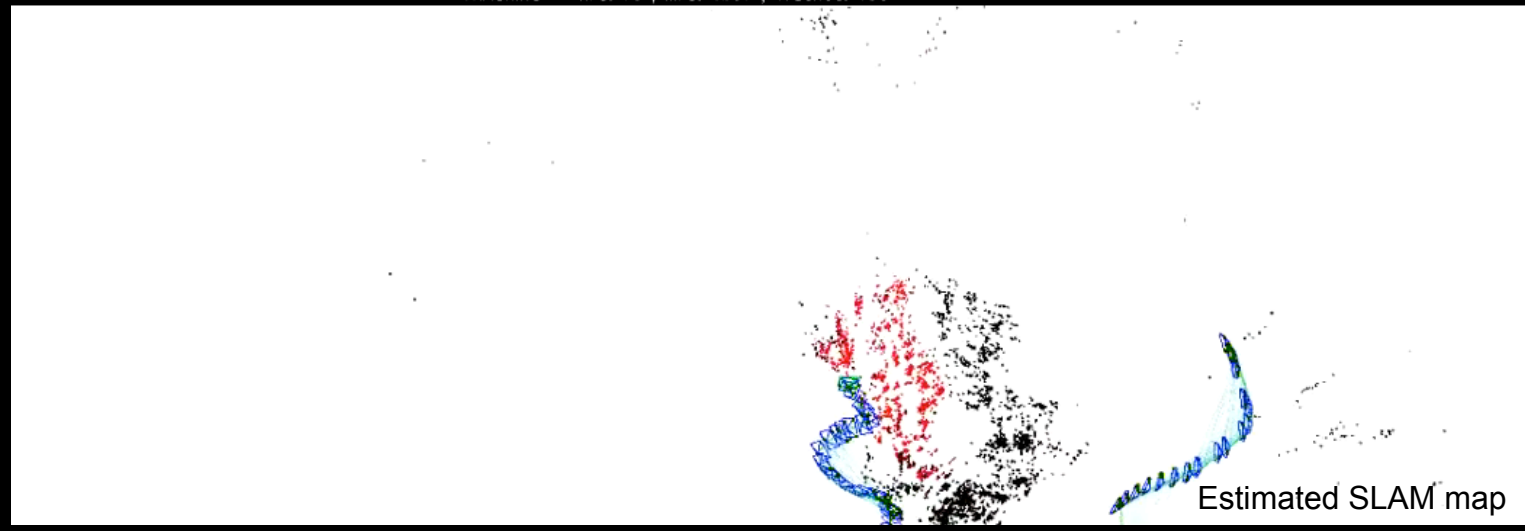
SLAM today | state of the art

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

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

- The most powerful open source **monocular** SLAM approach today
- Uses ORB features (binary) in a keyframe-based approach
- Binary place recognition





SLAM today | OKVIS [Leutenegger et al., RSS 2013 & IJRR 2015]

- **OKVIS: Open Keyframe-based Visual Inertial SLAM**

Code available on https://github.com/ethz-asl/okvis

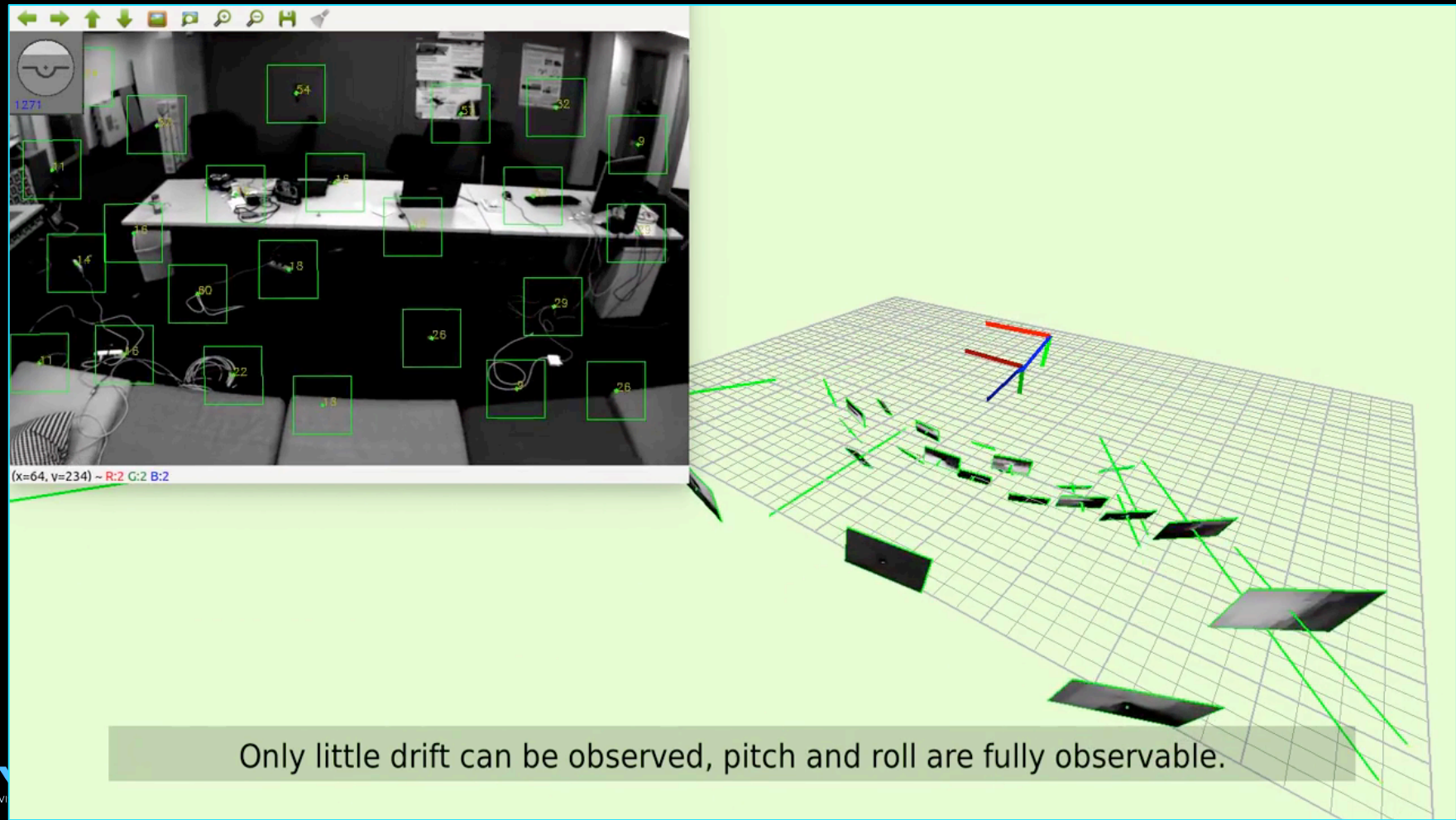
- One of the most powerful visual-inertial ~~SLAM~~ odometry approaches (i.e. no loop-closure)
- Uses BRISK features in a keyframe-based approach
- Tight visual-inertial fusion – handles both monocular and stereo vision

SLAM today | ROVIO [Bloesch et al., IROS 2015]

- **ROVIO: RObust Visual-Inertial Odometry using a direct EKF-based approach**

Code available on http://github.com/ethz-asl/rovio

- EKF-based
- Detects a variant of Shi-Tomasi features at different scale levels
- Tracks patches and uses the intensity errors in the innovation term
- Can only track a limited no. features, so ROVIO performs odometry, not SLAM

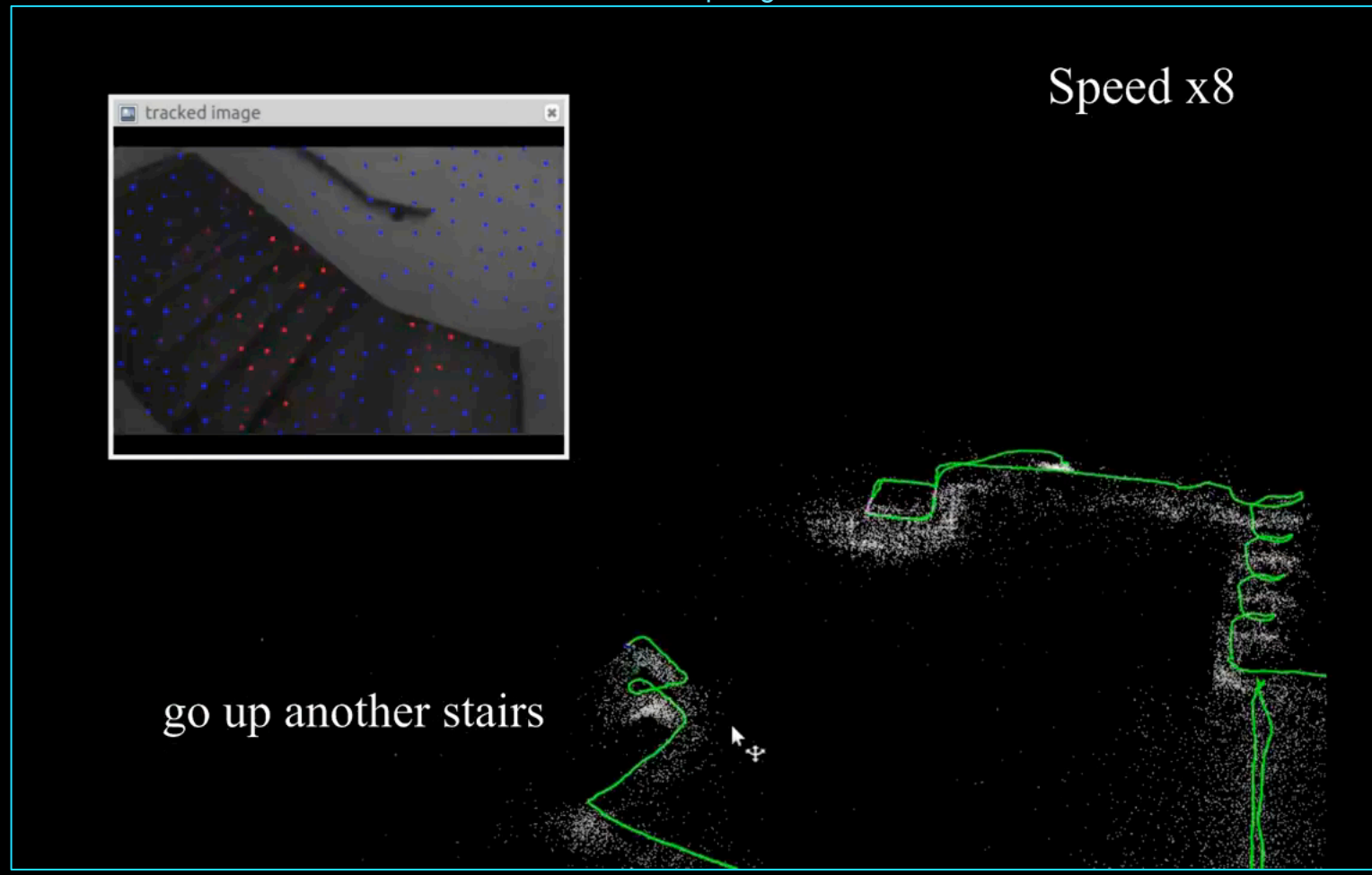


SLAM today | VINS-mono [Qin et al., TRO 2018]

- **VINS-mono: a robust and versatile monocular visual-inertial state estimator**

Code available on https://github.com/HKUST-Aerial-Robotics/VINS-Mono

- Keyframe-based approach
- Runs SLAM based on a tightly-coupled visual-inertial odometry with relocalization
- Shi-Tomasi features tracked using the KLT sparse optical flow tracker
- BRIEF descriptors for relocalization, using binary place recognition
- Extensions to stereo and IMU open-sourced



Autonomous Robots | how close are we?

Extract from the TED talk of Marc Railbert, founder of Boston Dynamics, August 2017

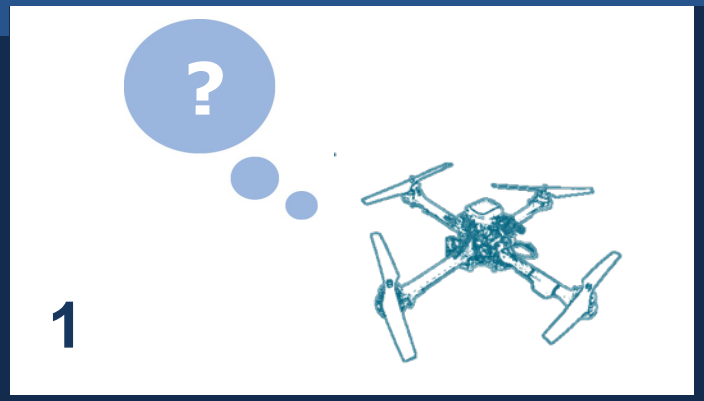
ETH zürich

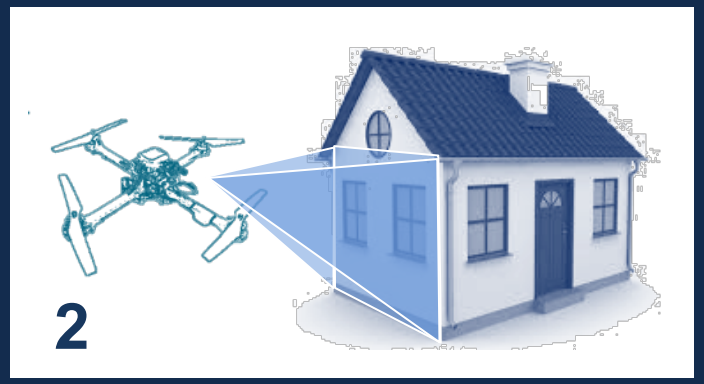
Research at V4RL: Vision-based Robotic Perception

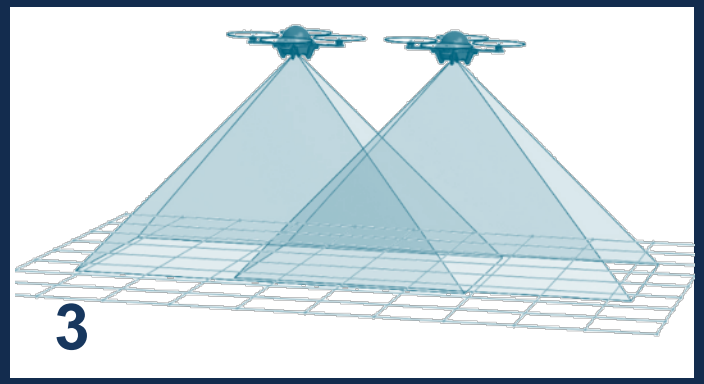
Margarita Chli

Vision for Robotics Lab (V4RL)
www.v4rl.ethz.ch

The Challenges







1. High fidelity ego-motion & scene estimation

2. Scene reconstruction for interaction and path-planning

3. Multi-agent collaboration

Vision-based Robotic Perception – what's missing

- Fast motion
- Large scales
- Rich maps
- Robustness
- Lower Computation

1ms Target Tracking Vision Chip by Ishikawa Komuro University

Handle **larger** amounts of data more **effectively**

- Competing goals:
  Precision vs. Efficiency

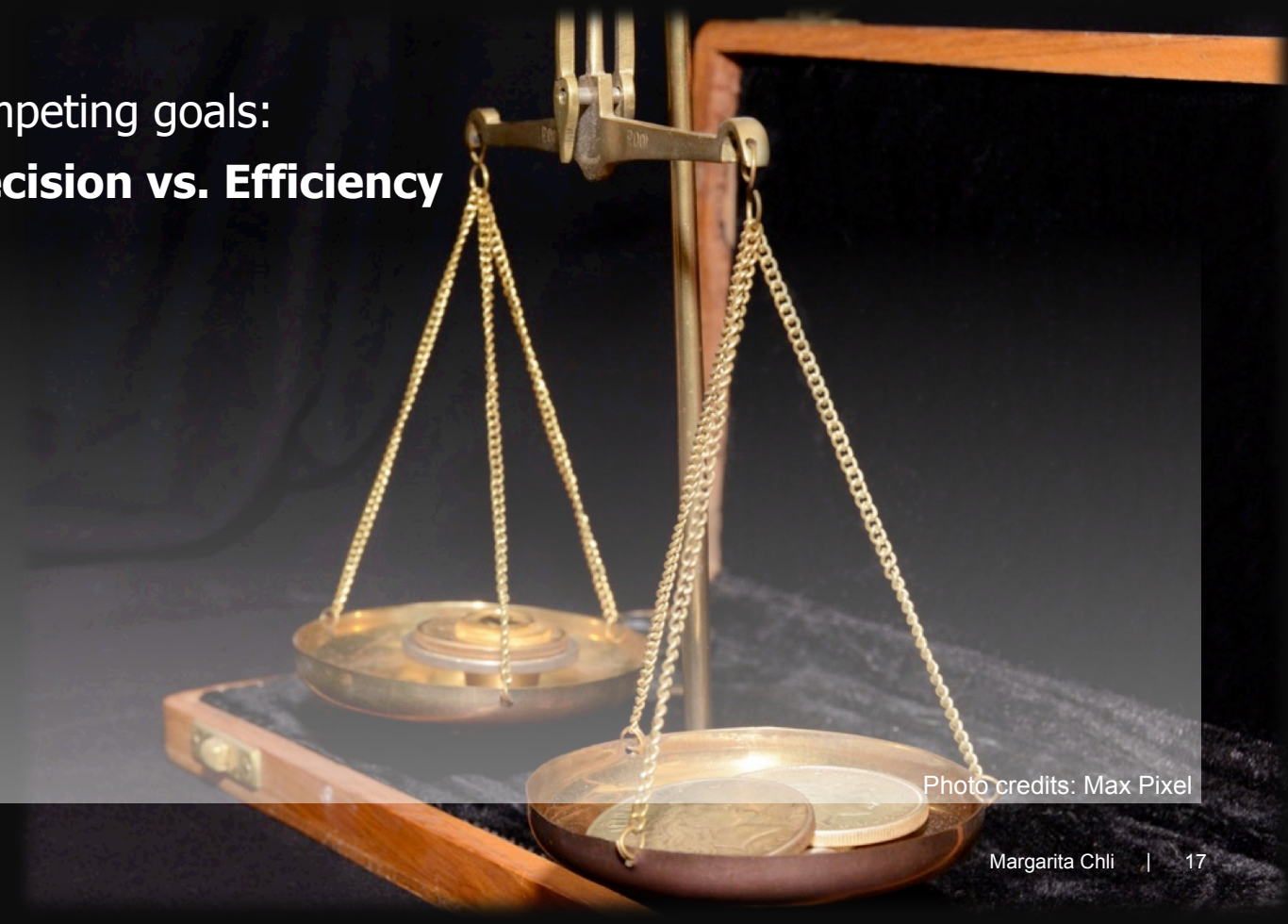


Photo credits: Max Pixel

Vision-based Drone Perception – UAV properties & challenges

Weight

✓ Lighter & safer than larger robots

✗ Limited resources

Agility

✓ Very agile

✗ Fast, unstable dynamics, cannot “stop”

Autonomy

✗ Battery, communication bandwidth

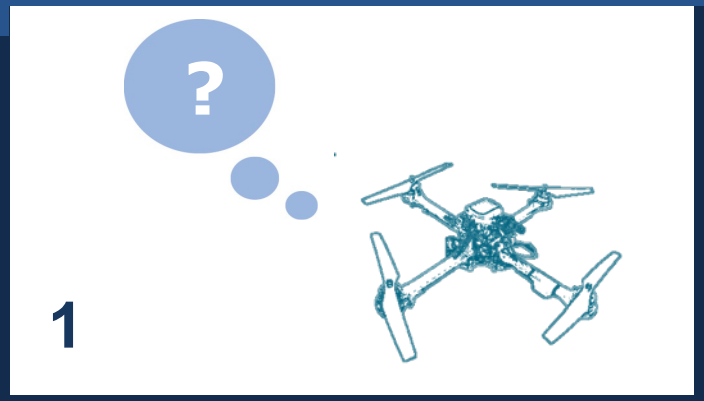
Photo credits: Francois Pomerleau

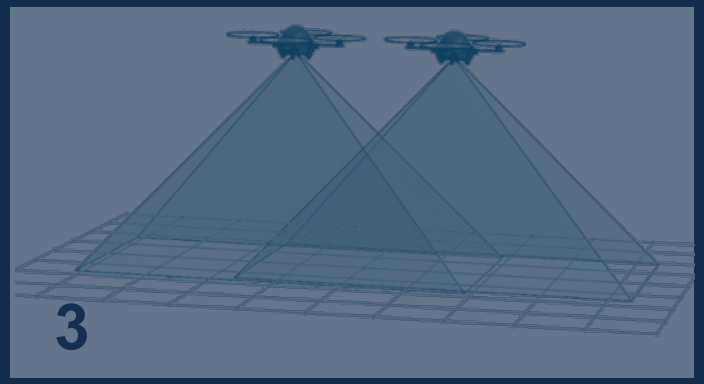
The Vision

Develop

visual **perception** & **intelligence**

The Challenges

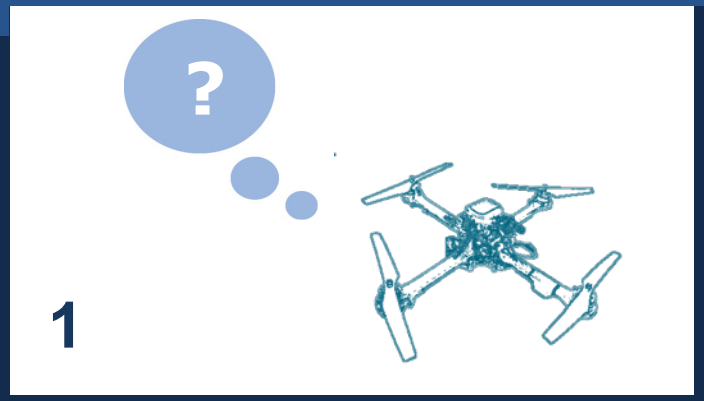




1. High fidelity SLAM

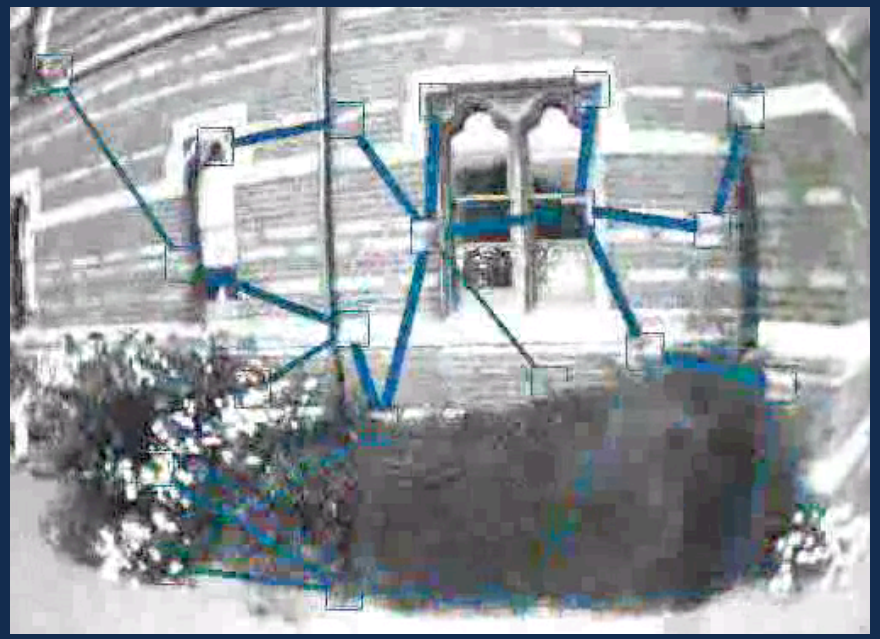
2. Scene reconstruction for interaction and path-planning

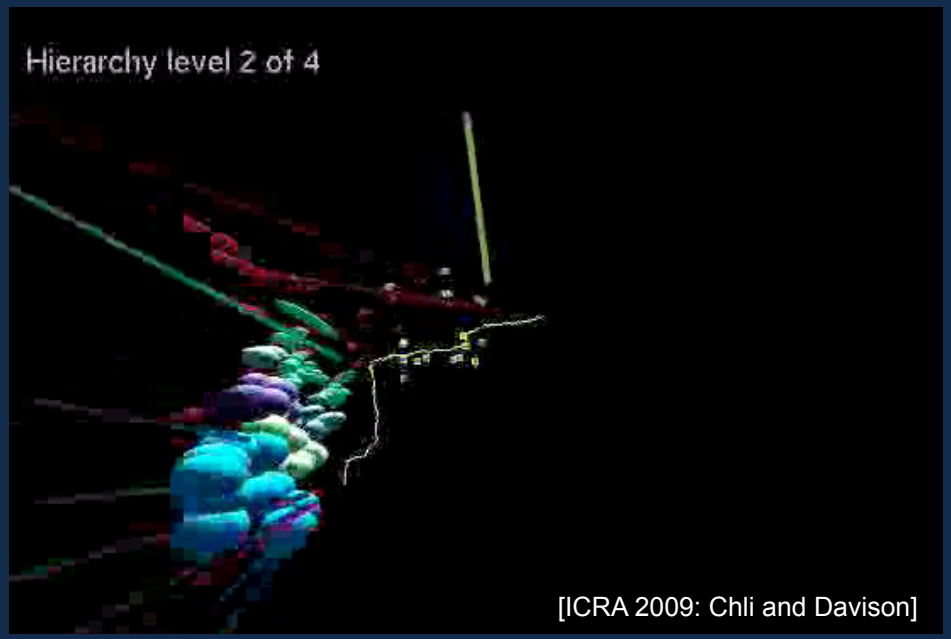
3. Multi-agent collaboration

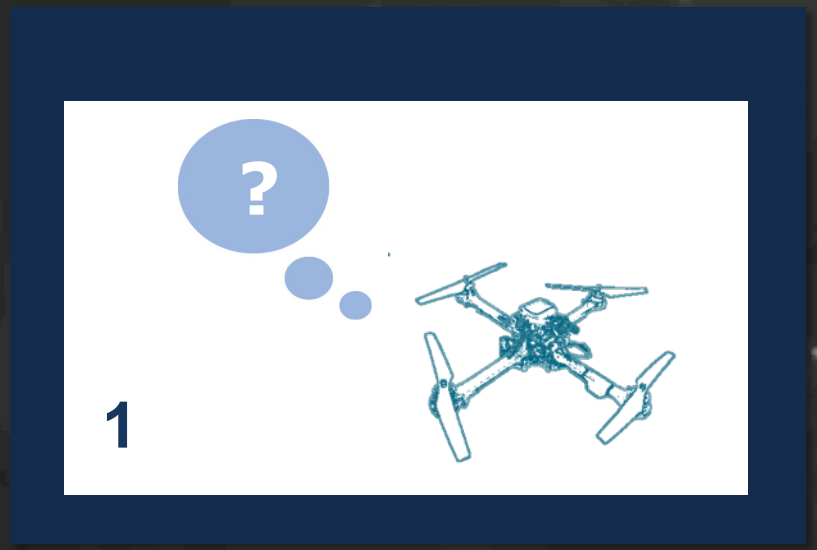


SLAM: Simultaneous Localization And Mapping

- How to track the motion of a robot while it is moving in an unkown environment using onboard sensors?
- Traditional SLAM:
  Pick natural scene features as landmarks, observe their motion & reason about robot motion







View Map Off Spacebar Reset

[IROS 2012: Achtelik, Lynen, Weiss, Kneip, Chli and Siegwart]

nar scene and press spacebar to start tracking for initial map.

[IROS 2012: Achtelik, Lynen, Weiss, Kneip, Chli and Siegwart]

Altitude: 14.19 m

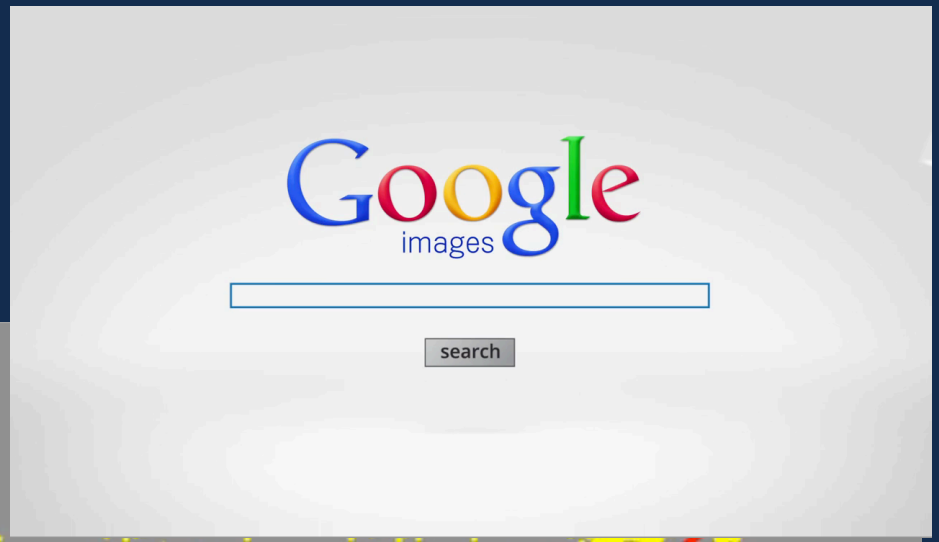
[IROS 2012: Achtelik, Lynen, Weiss, Kneip, Chli and Siegwart]

Visual-inertial SLAM downtown Zurich

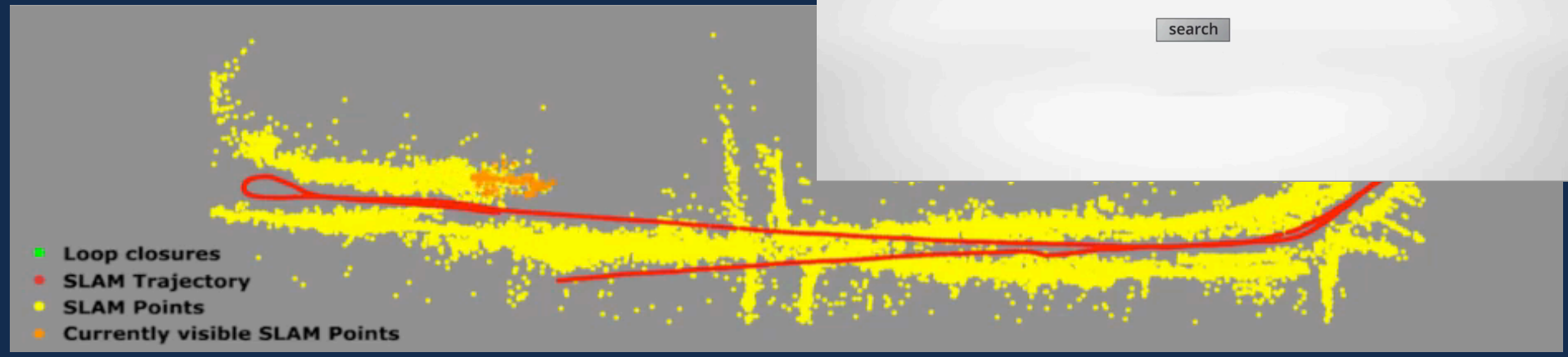
Current view:

Place Recognition

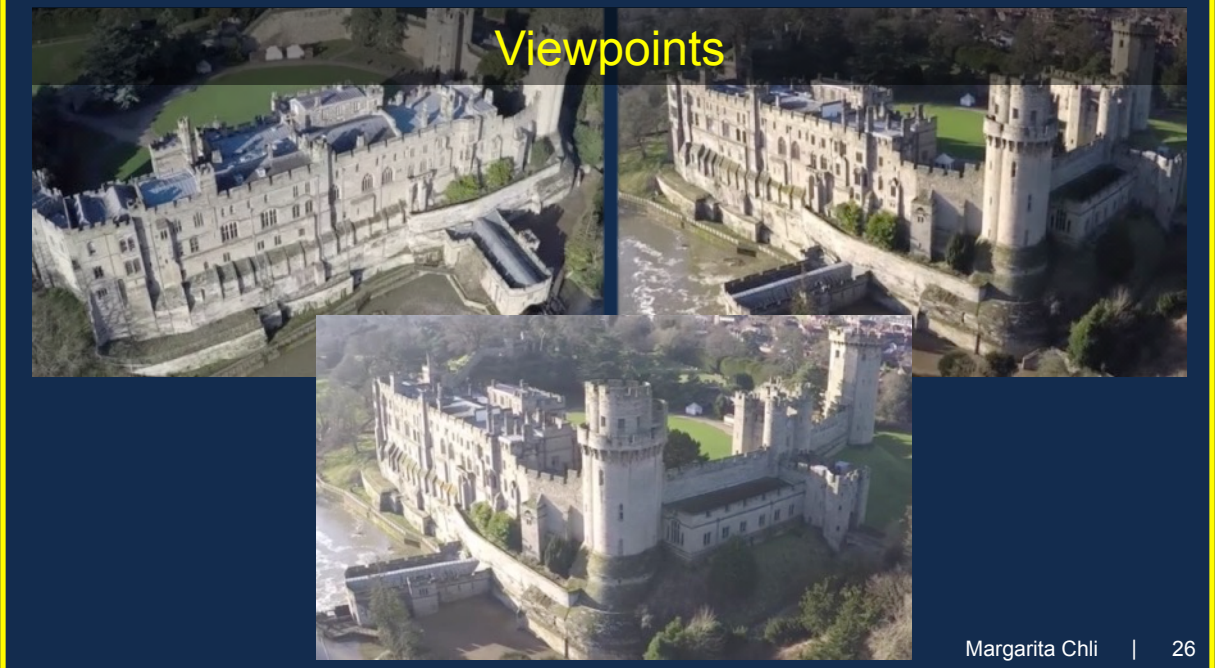
- Recognize when the robot visits a “known” location
- Build vocabulary of “visual words” & search for matching images

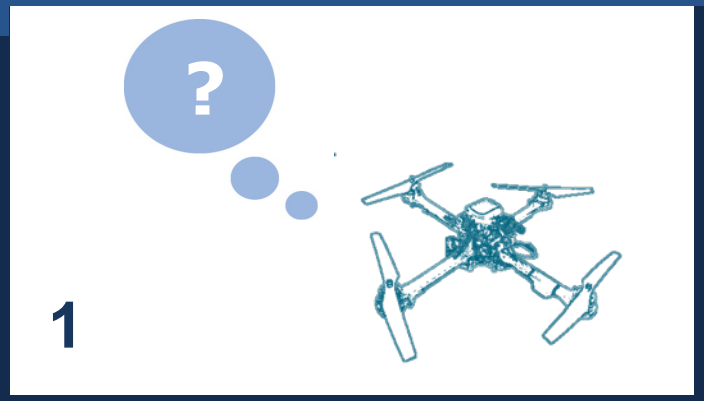


SLAM map:

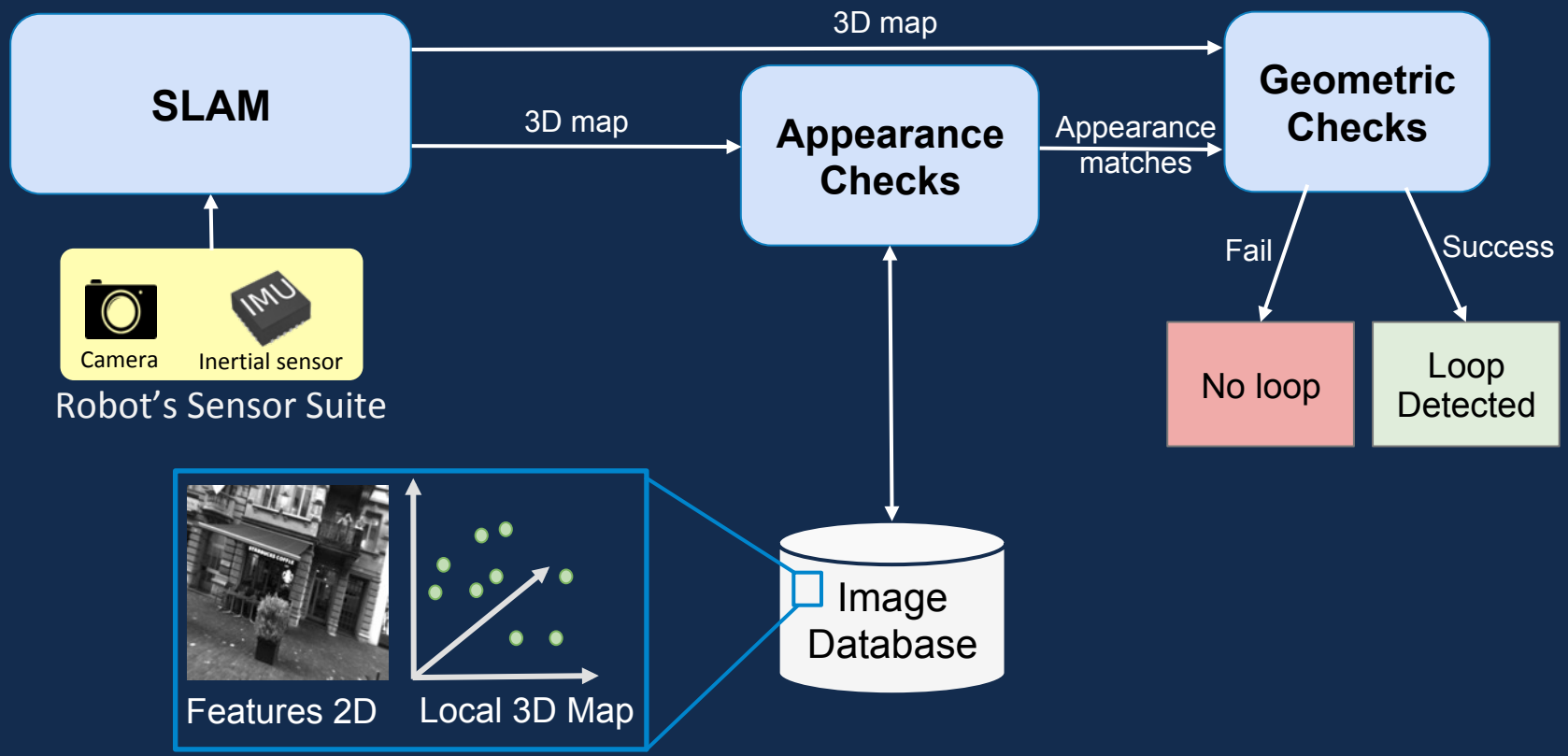


Place Recognition for High fidelity SLAM





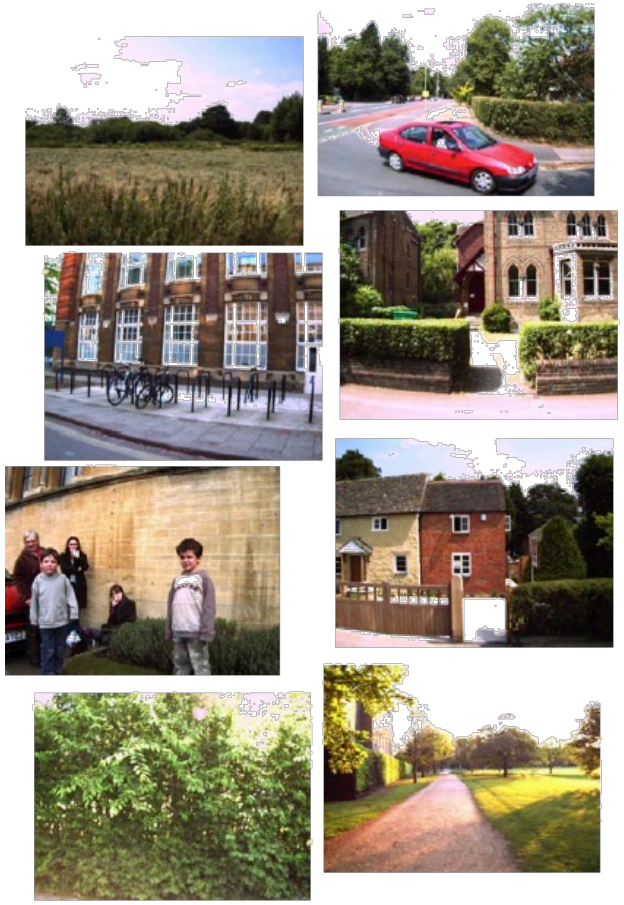
Viewpoint-tolerant place recognition



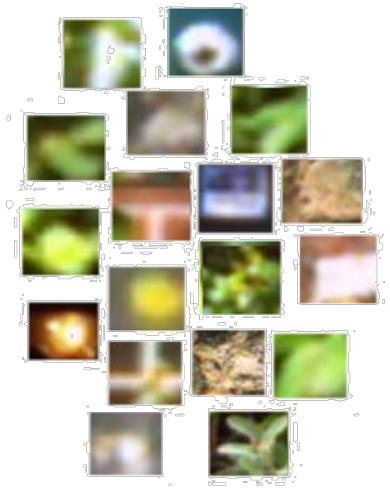
[ICRA 2018: Maffra, Teixeira, Chen and Chli]

Viewpoint-tolerant Place Recognition: Appearance Checks

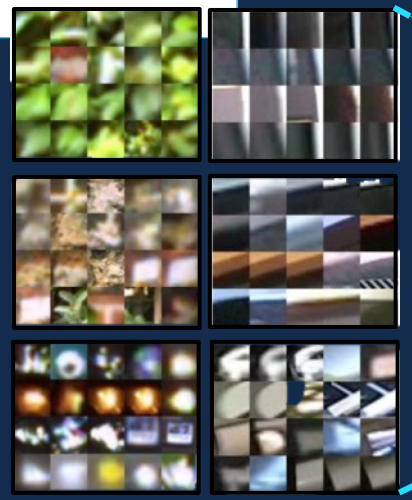
Image Collection



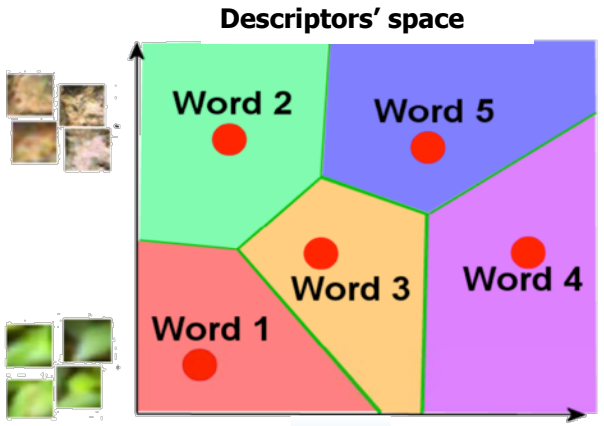
Extract Features

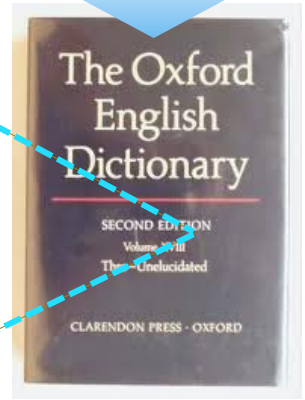


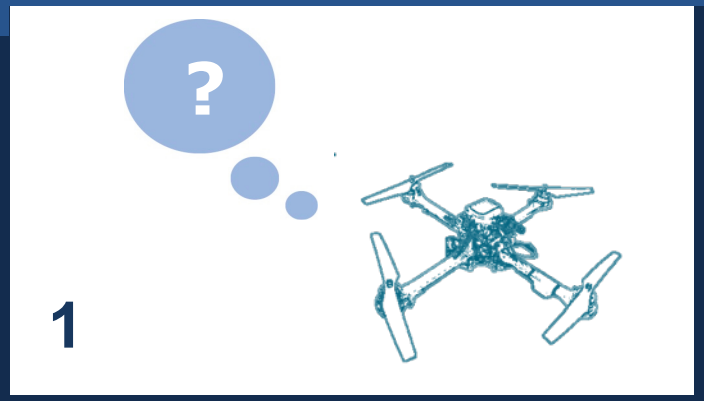
Examples of Visual Words:



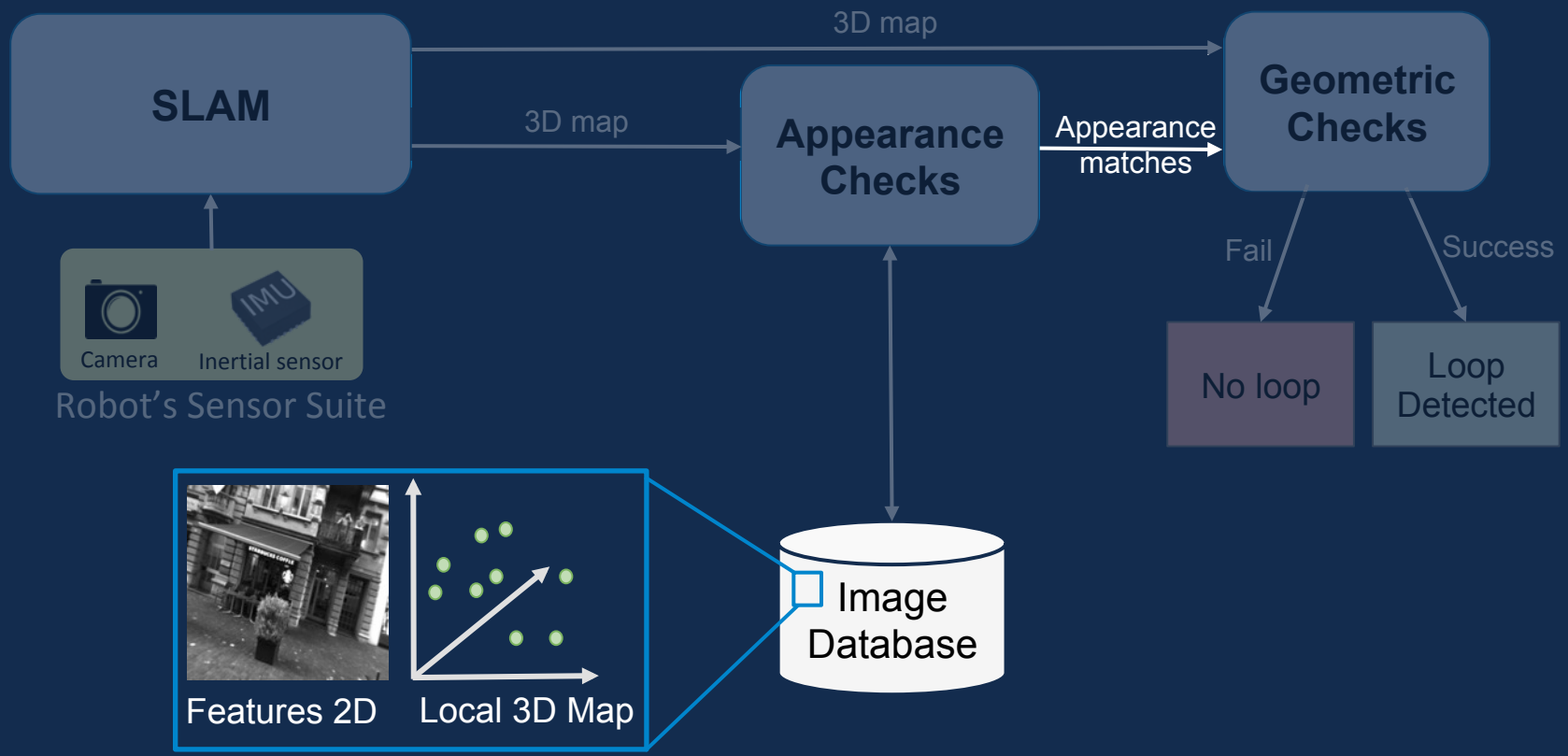
Cluster Descriptors







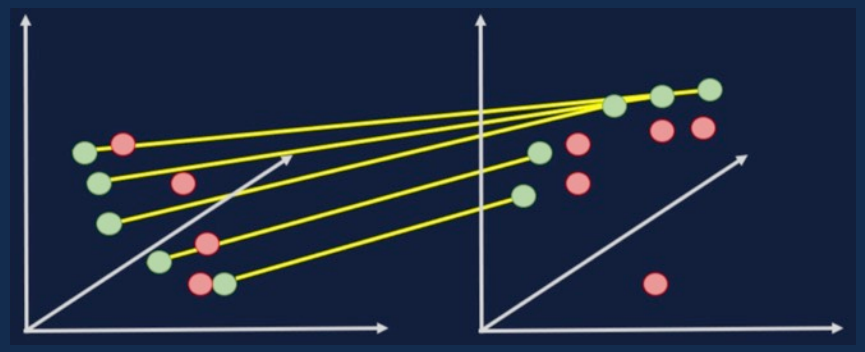
Viewpoint-tolerant place recognition



[ICRA 2018: Maffra, Teixeira, Chen and Chli]

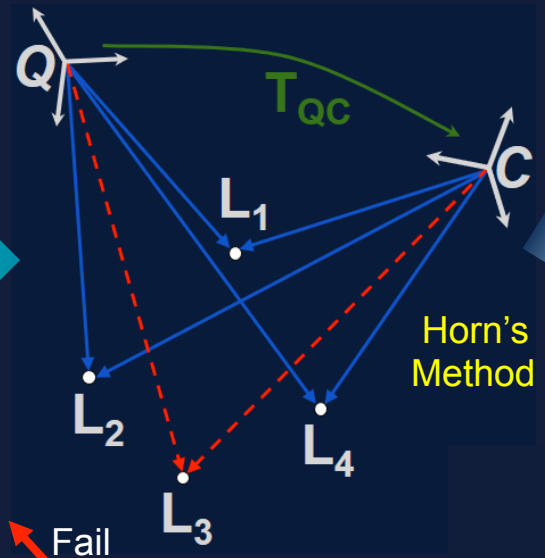
Viewpoint-tolerant Place Recognition: Geometric Checks

Find 3D-3D matches

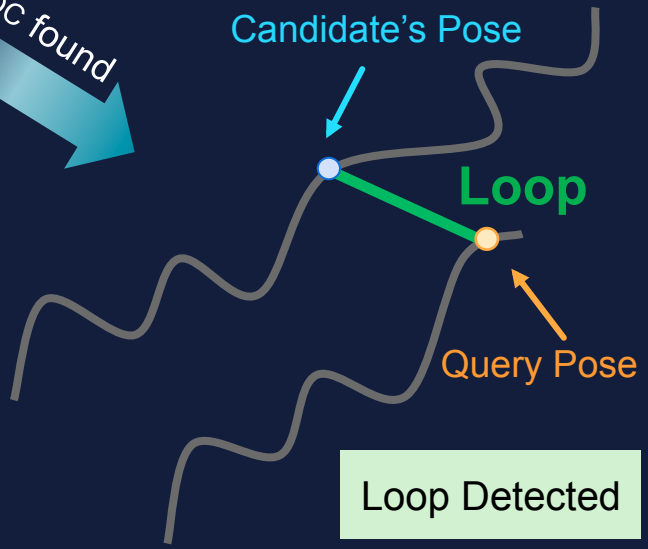


Query & Candidate's Point Clouds

Enough 3D-3D matches



T_{QC} found



Loop Detected

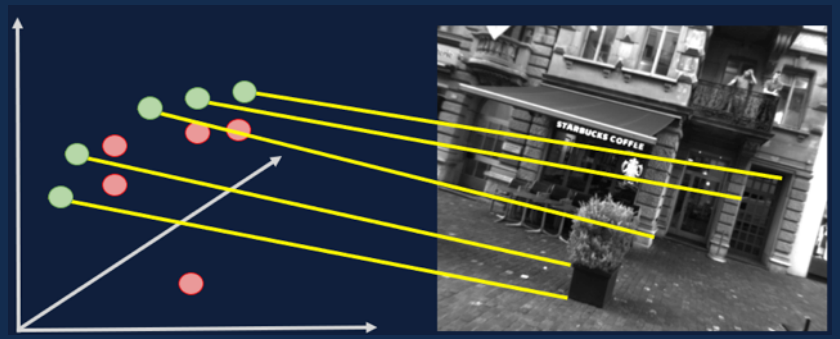
Not enough 3D-3D matches
OR
T_{QC} not found in 1st attempt

Fail

No Loop

Fail

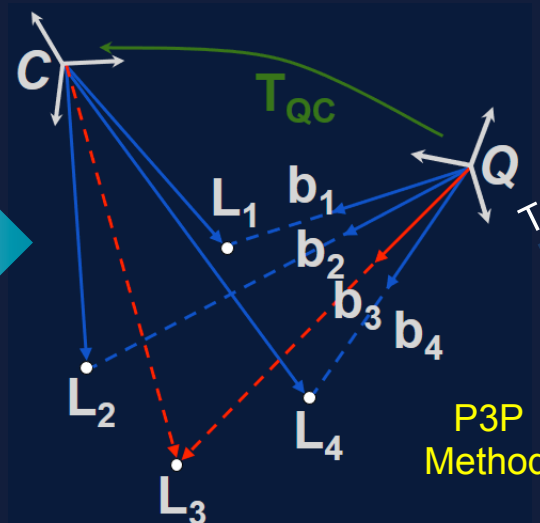
Find 3D-2D matches



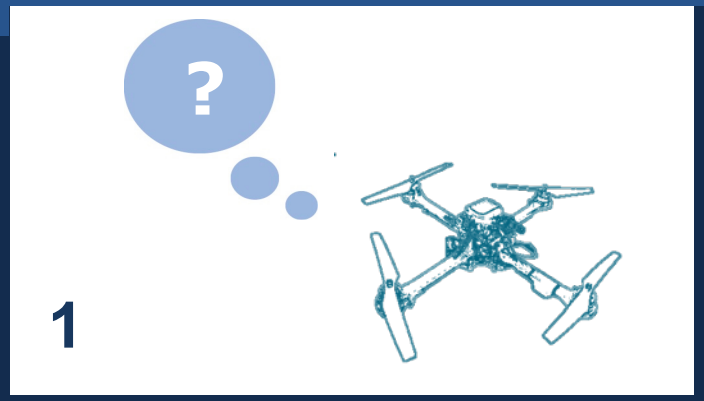
Candidate's Point Clouds

Query Image

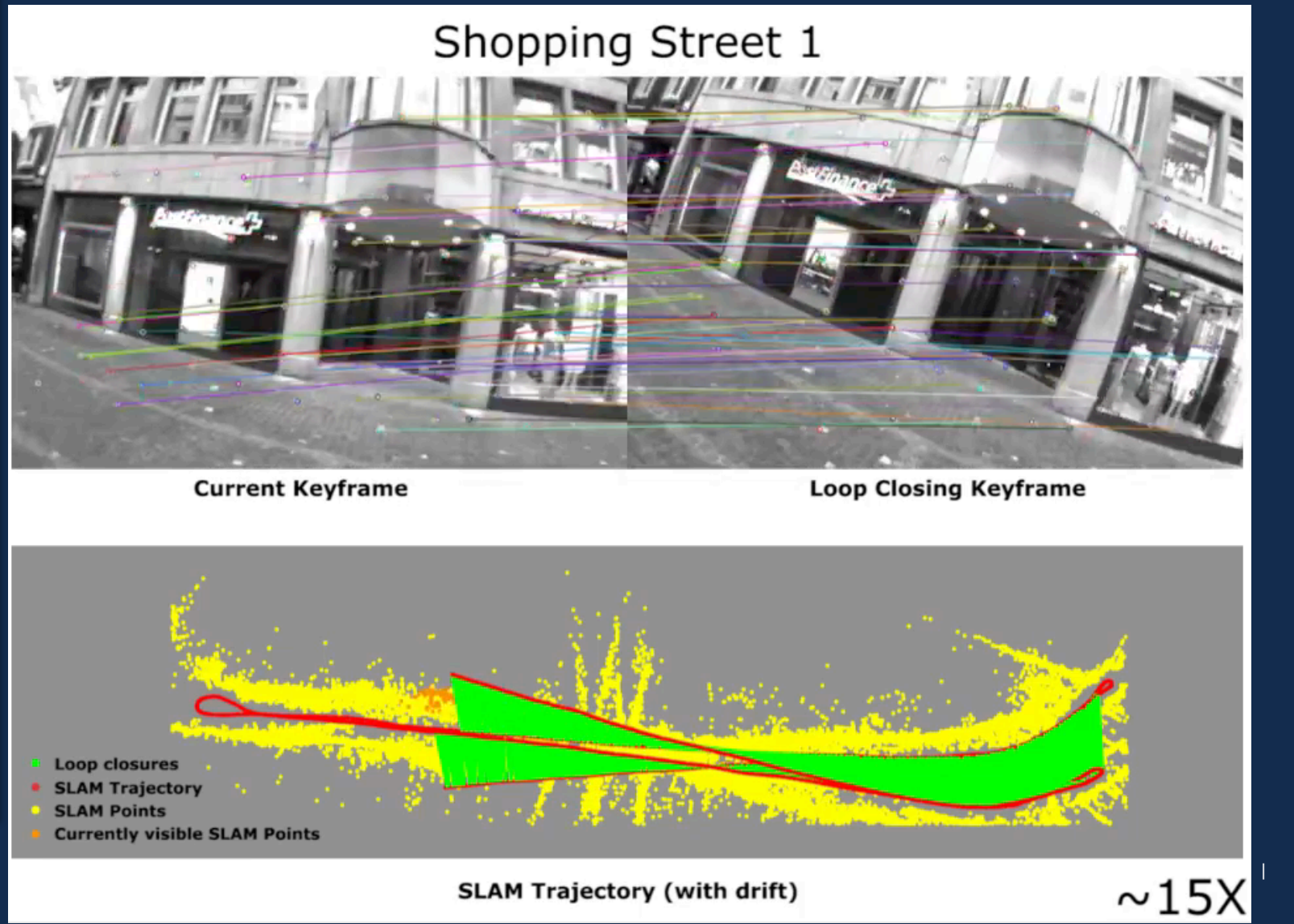
Enough 3D-2D matches



T_{QC} found

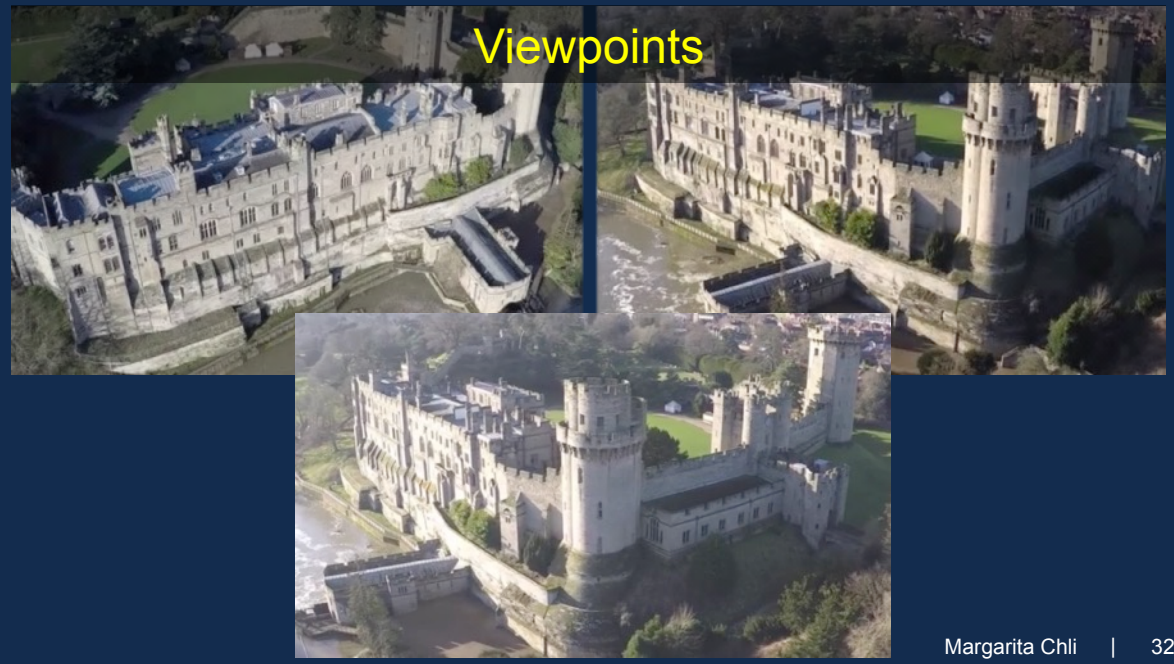


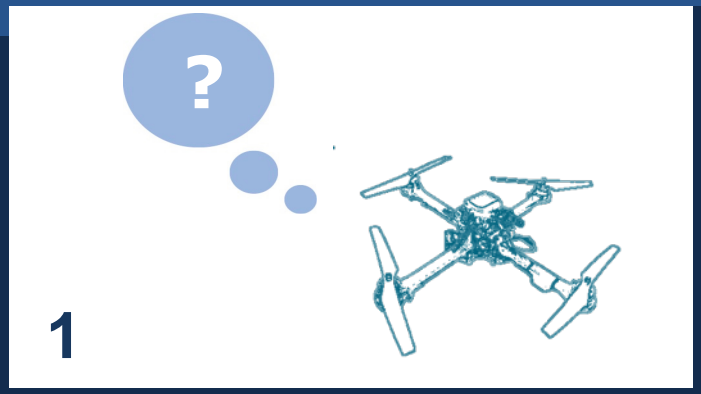
Viewpoint-tolerant place recognition



[RAL 2019: Maffra, Texeira, Chen and Chli]
[ICRA 2018: Maffra, Chen and Chli]

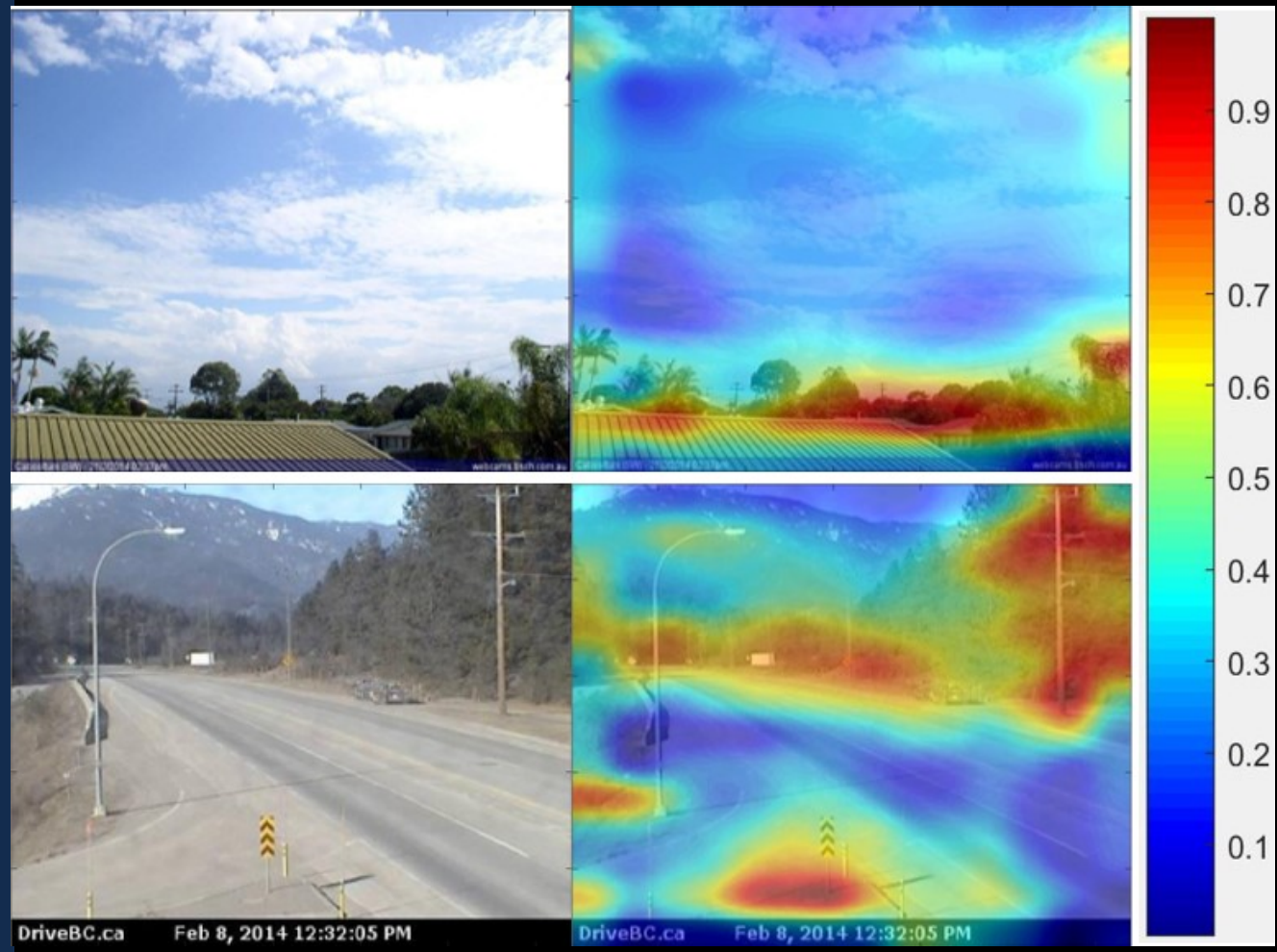
Place Recognition for High fidelity SLAM





Place Recognition deep learning

Learning an attention model:

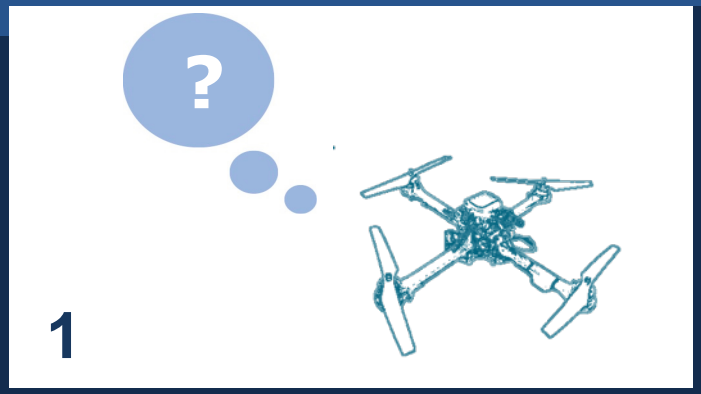


Most salient

Least salient

[IROS 2017: Chen, Maffra, Sa and Chli]

[RAL 2018: Chen, Liu, Sa, Ge and Chli]



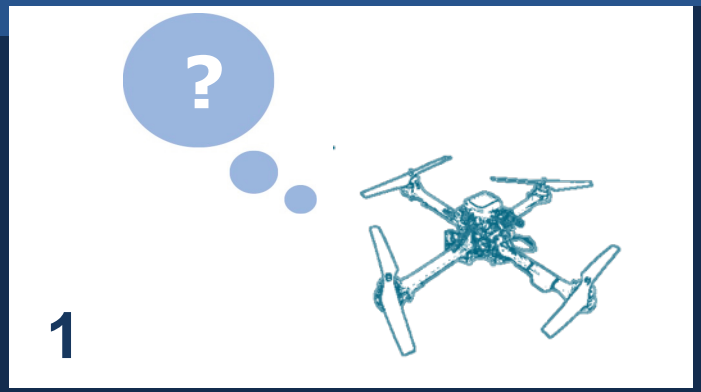
Place Recognition using deep learning

Query Image

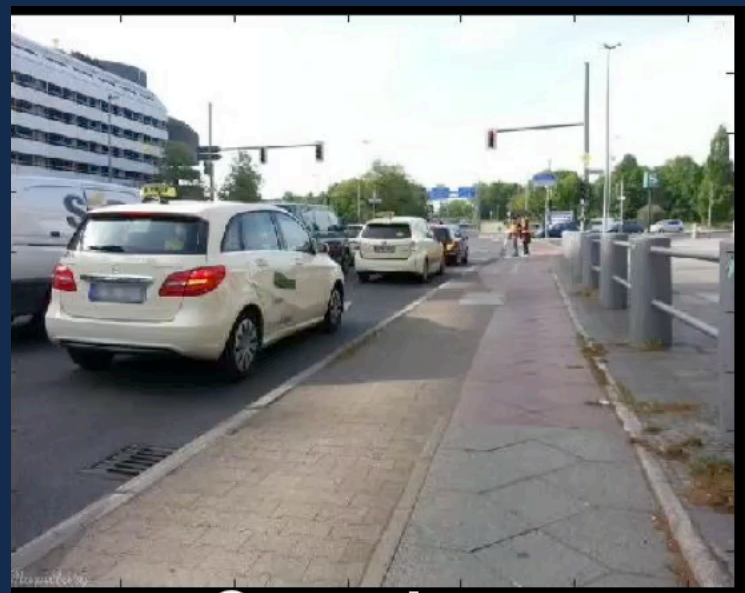
Our return

[IROS 2017: Chen, Maffra, Sa and Chli]

[RAL 2018: Chen, Liu, Sa, Ge and Chli]



Place Recognition using deep learning



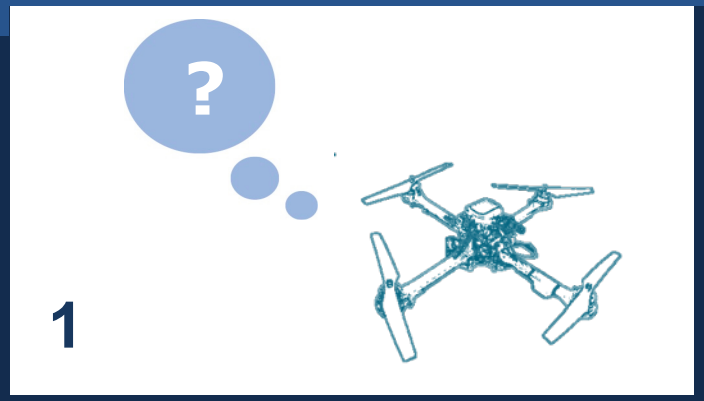
Query Image

Ground Truth

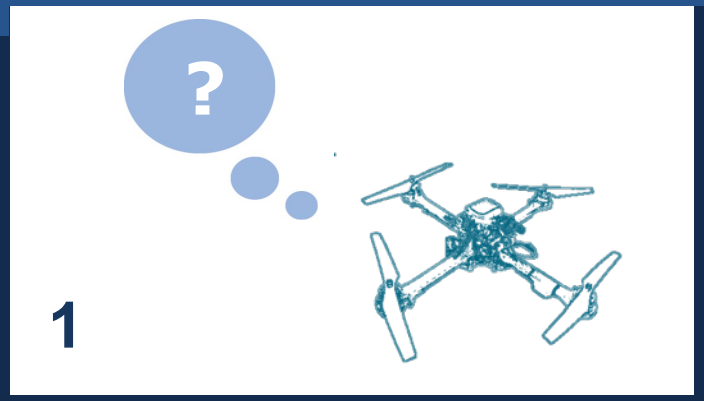
Our Return

[IROS 2017: Chen, Maffra, Sa and Chli]

[RAL 2018: Chen, Liu, Sa, Ge and Chli]



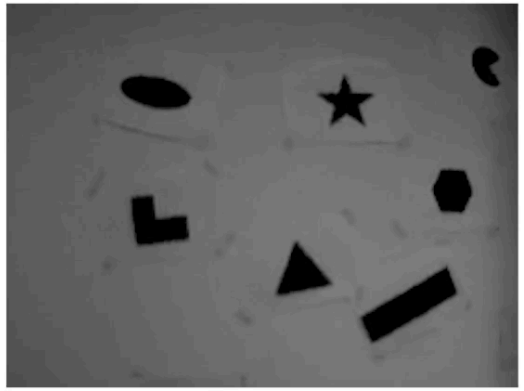
Event cameras for high-fidelity SLAM



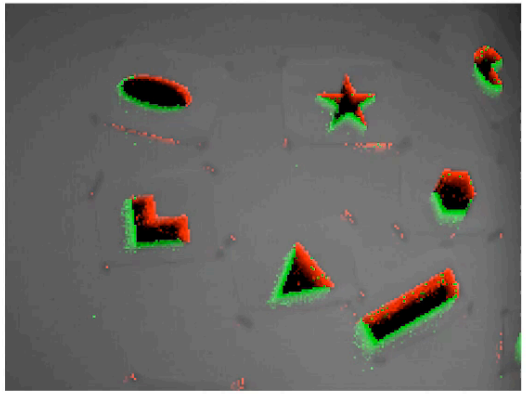
Event cameras for high-fidelity SLAM

Event cameras for high-speed tracking

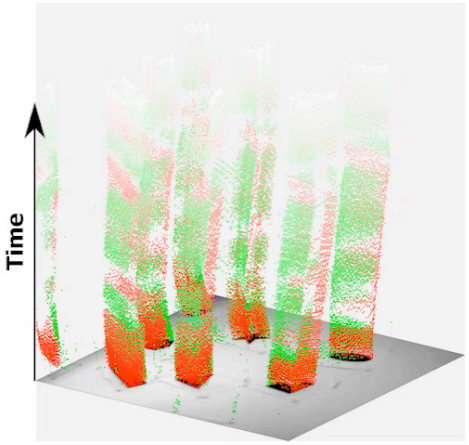
- **Each pixel** reacts independently generating an **event** when **intensity changes**
- **Events** are captured as an **asynchronous stream** with **high temporal resolution**



Intensity frames
(Not used - Only for illustration)



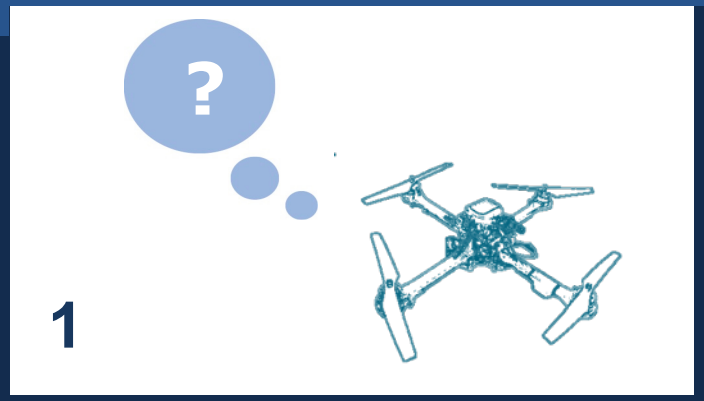
Events



Event Stream

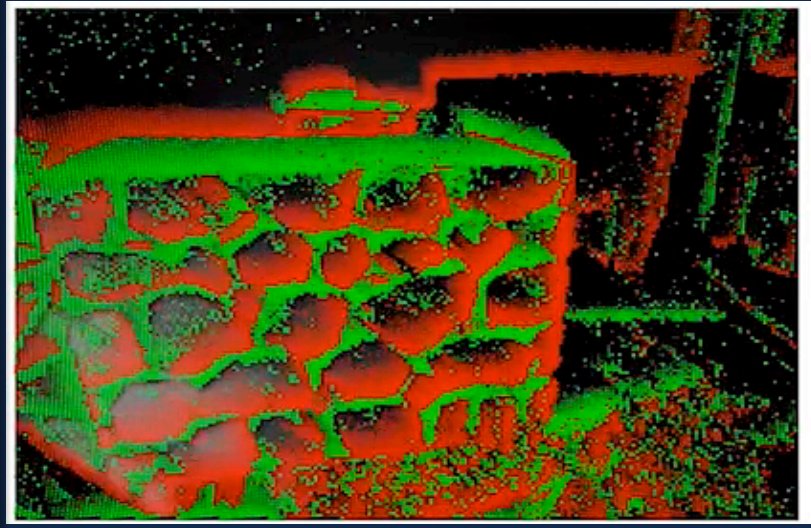
[3DV 2018: Alzugaray and Chli]

[RAL 2018: Alzugaray and Chli]

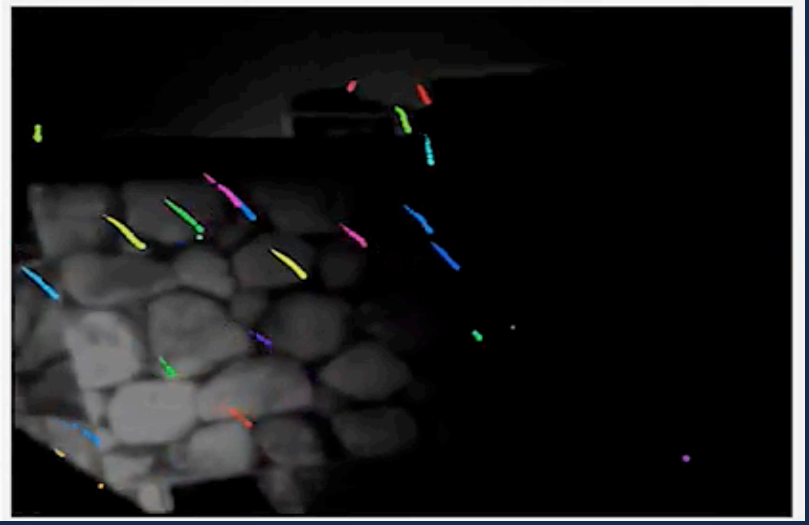


Event cameras for high-fidelity SLAM

Event cameras for high-speed tracking in darker scenes



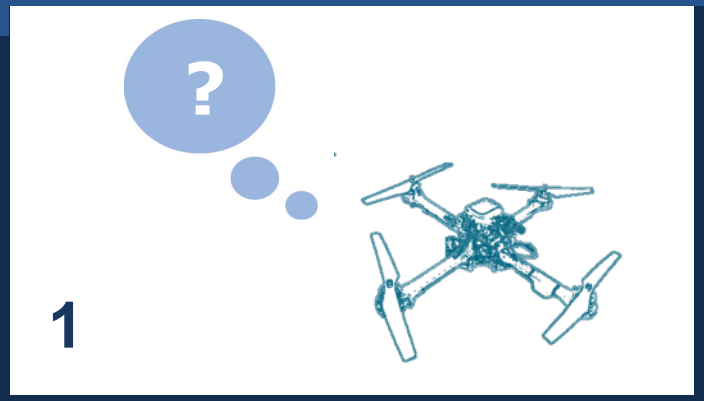
Events



Corner Tracks

[3DV 2018: Alzugaray and Chli]

[RAL 2018: Alzugaray and Chli]

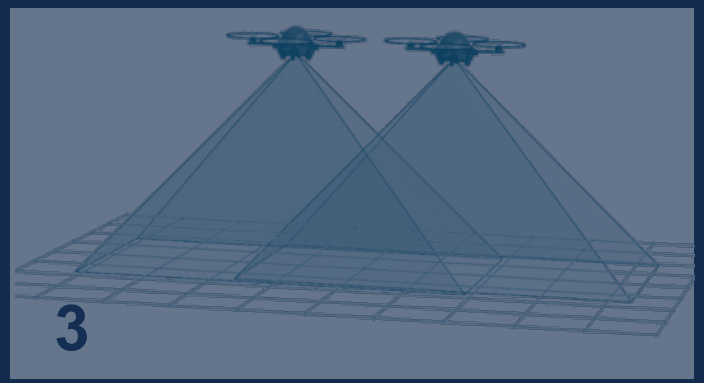


1. High fidelity SLAM

- Experimenting with event-cameras, sensor fusion, deep learning, place recognition, …
- The backbone of **spatial awareness** of a robot, giving the robot's **pose** and a **map** of its surroundings

sparse

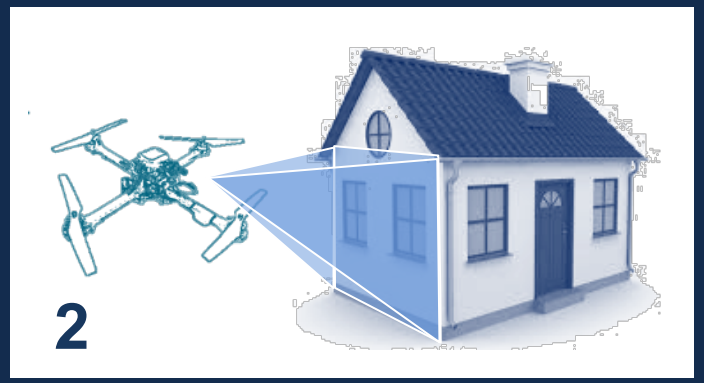
The Challenges



1. High fidelity SLAM

2. Scene reconstruction for interaction and path-planning

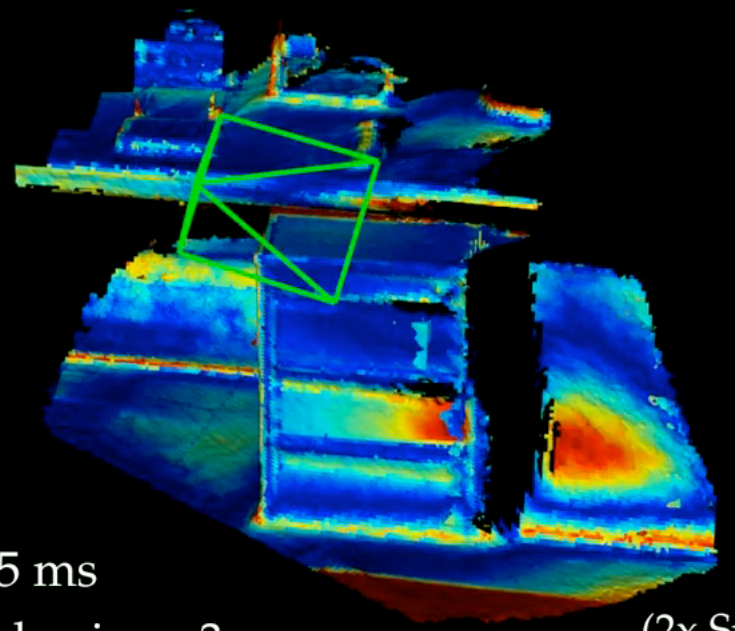
3. Multi-agent collaboration

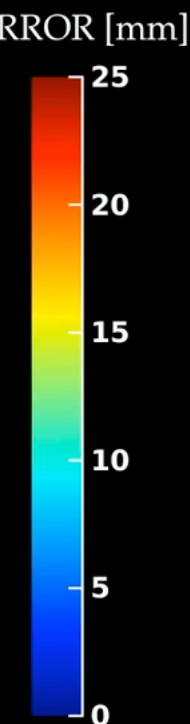


[IROS 2016: Karrer, Kamel, Siegwart and Chli]

Scene estimation from visual, inertial & depth data

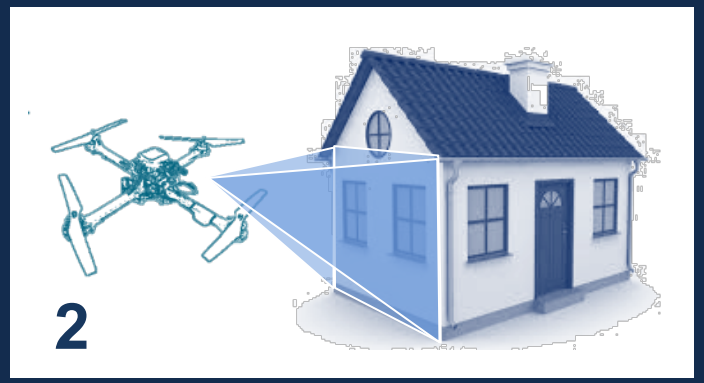
Real-time reconstruction from visual, inertial and depth data



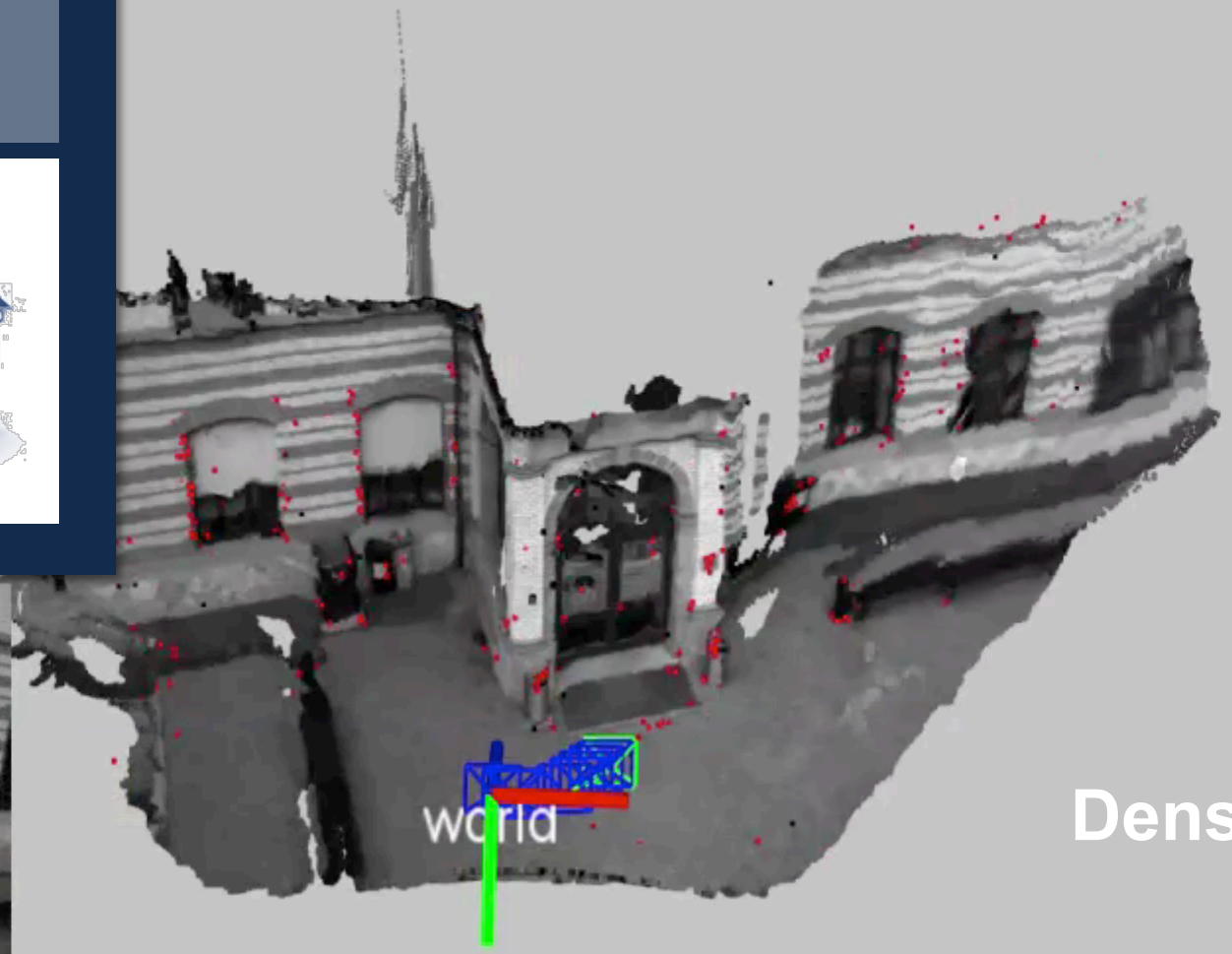


average time per frame: 5 ms
mean error: 8.5mm, time horizon 3s

(2x Speed)

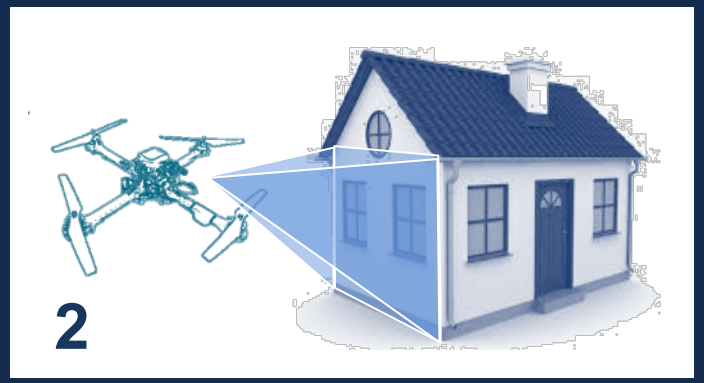


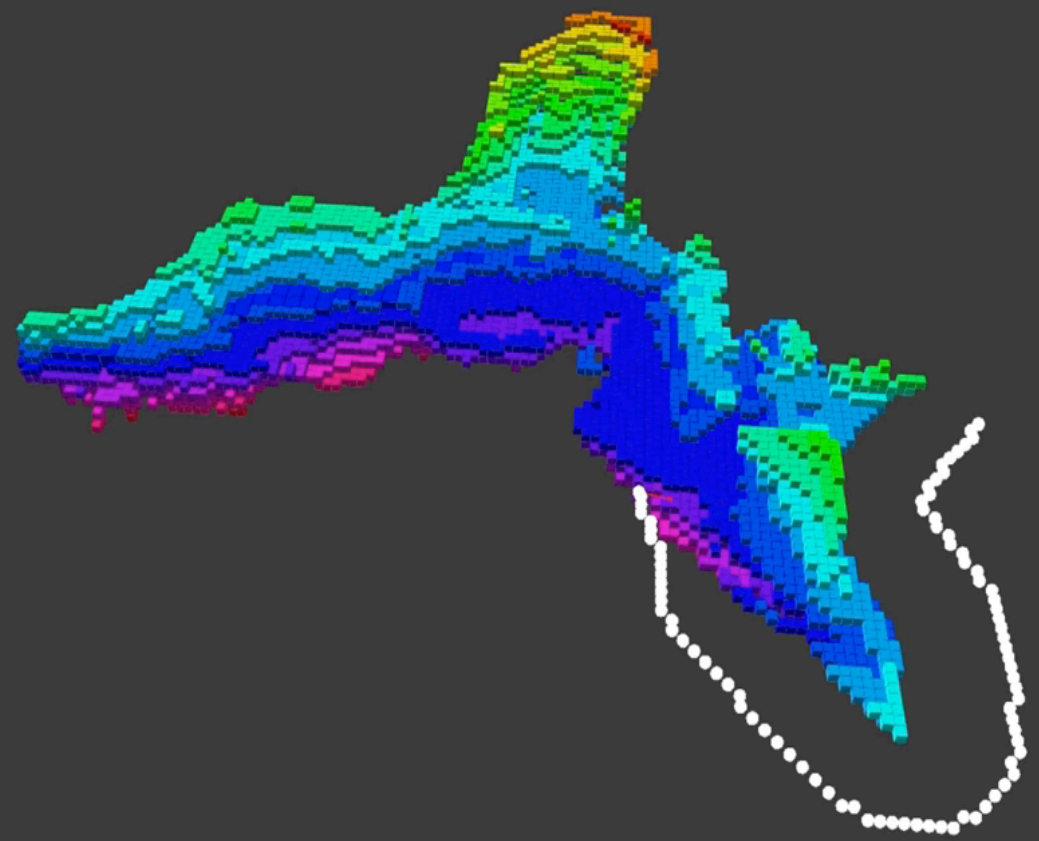
[ICRA 2017: Teixeira and Chli]



Denser mapping with a single camera

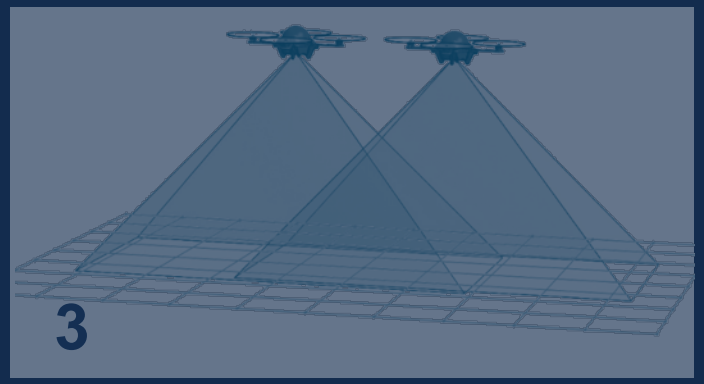
Dense mapping for collision-free path planning





[FSR 2017: Teixeira, Alzugaray and Chli]

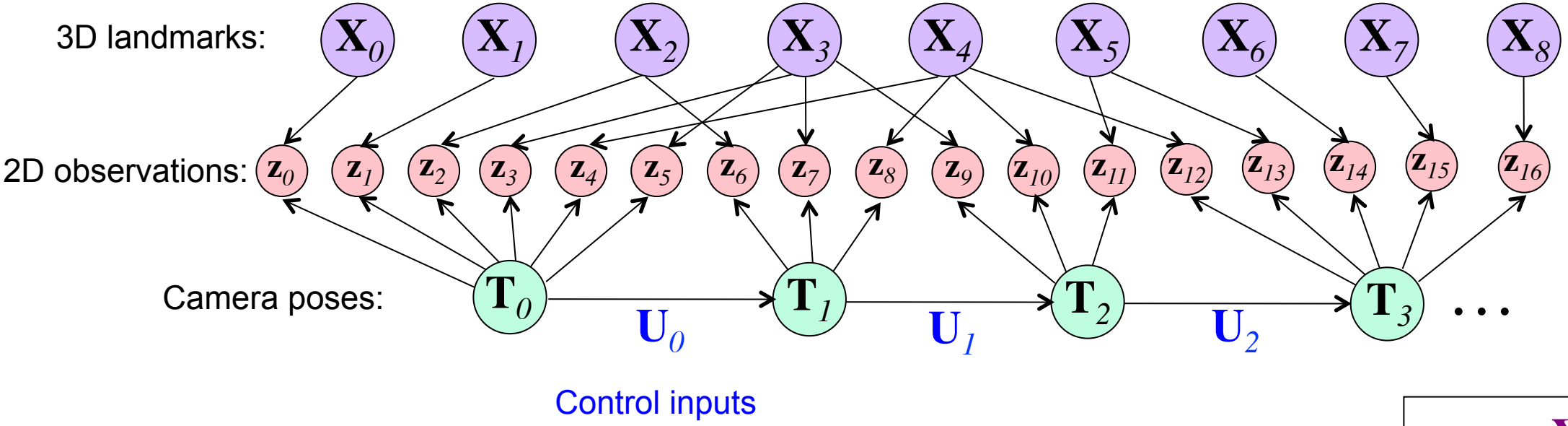
The Challenges

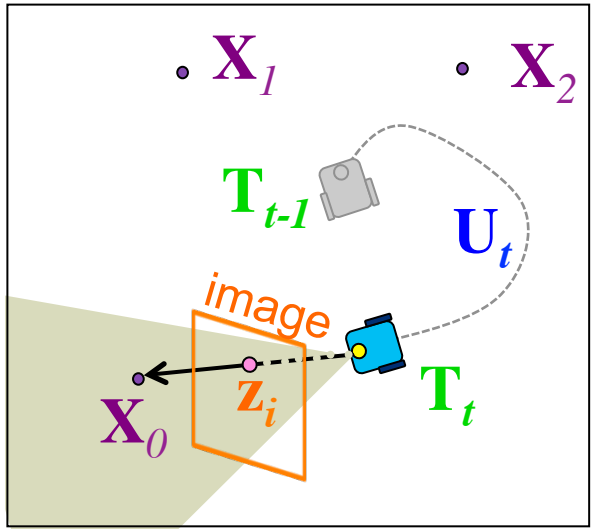


1. High fidelity SLAM

2. Scene reconstruction for interaction and path-planning

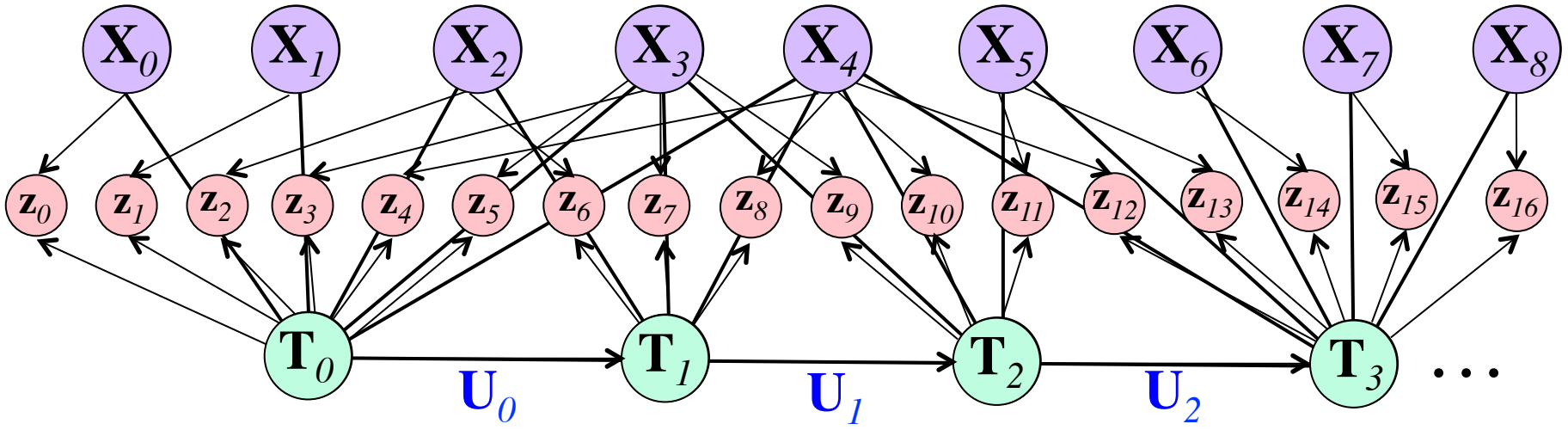
3. Multi-agent collaboration


3D landmarks:
2D observations:
Camera poses:
X_0 X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8
z_0 z_1 z_2 z_3 z_4 z_5 z_6 z_7 z_8 z_9 z_{10} z_{11} z_{12} z_{13} z_{14} z_{15} z_{16}
T_0 T_1 T_2 T_3 . . .
U_0 U_1 U_2
Control inputs


X_1
X_2
T_{t-1}
U_t
image
z_i
X_0
T_t

SLAM | approaches to SLAM

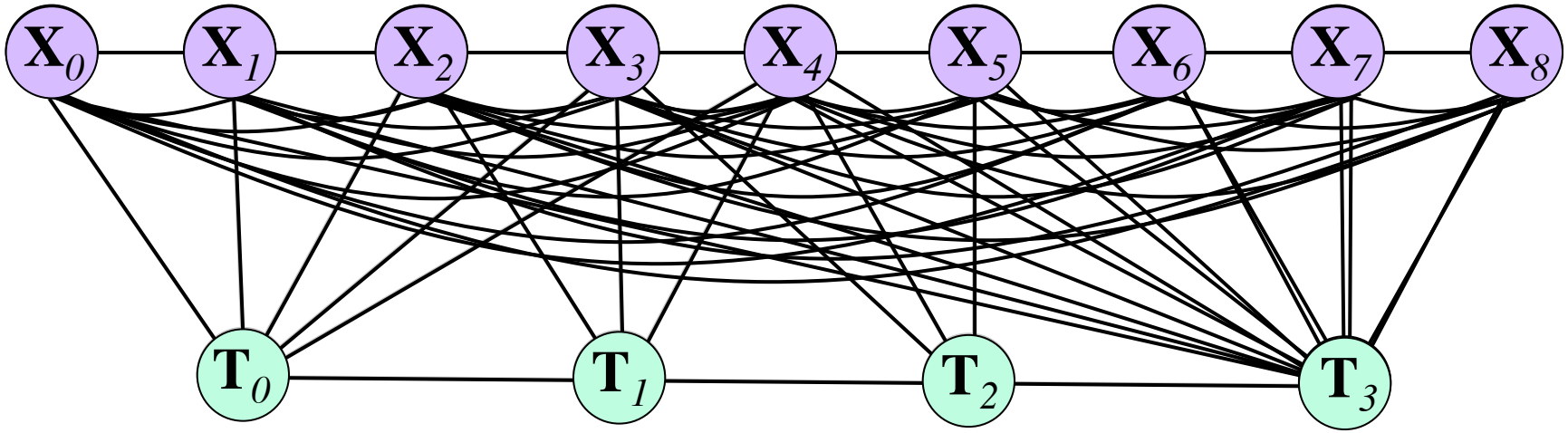
Full graph optimization (Bundle Adjustment) to estimate posterior: $p(\mathbf{T}_{0:t}, \mathbf{X}_{0:N} \mid \mathbf{z}_{0:k}, \mathbf{U}_{0:t})$



- Eliminate observations & control-input nodes and solve for the constraints between poses and landmarks.
- Globally consistent solution, but infeasible for large-scale SLAM

SLAM | approaches to SLAM

Filtering -- to estimate the posterior: $p(\mathbf{T}_t, \mathbf{X}_{0:N} \mid \mathbf{z}_{0:k}, \mathbf{U}_{0:t})$

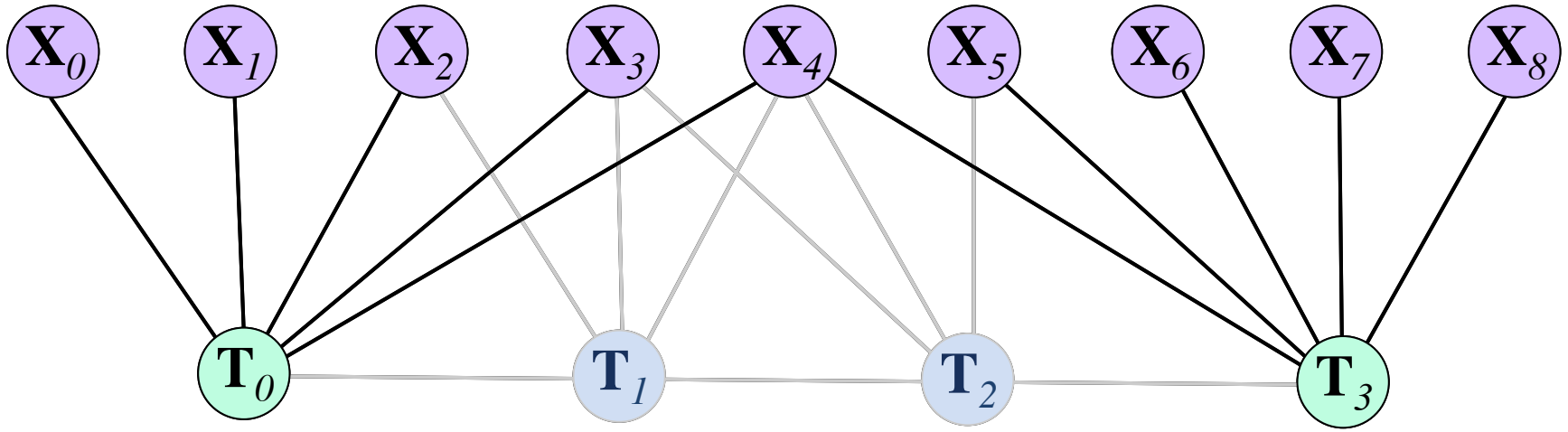


- Eliminate all past poses: 'summarize' all experience with respect to the latest pose, using a **state vector** and the associated **covariance matrix.**
- Unimodal estimate, $O(N^3)$ in the no. features

SLAM | approaches to SLAM

Keyframes -- to estimate the posterior: $p(\mathbf{T}_{KF(0:t)}, \mathbf{X}_{0:N} \mid \mathbf{z}_{0:k}, \mathbf{U}_{0:t})$,

↳ selects keyframe indices in the range [0:t]

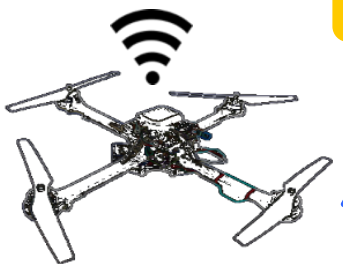


- Retain the most representative poses (keyframes) and their dependency links ⇨ optimize the resulting graph
- Non-linear optimization, $O(N)$ in the no. features

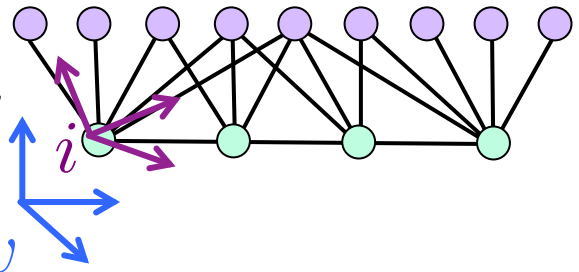
Collaborative SLAM | overview

[ICRA 2017: Schmuck and Chli]
[JFR 2018: Schmuck and Chli]

Robotic Agent



- Visual SLAM with **local BA**
- Limit local map of current environment



Landmark j's 3D location : $\mathbf{X}_{w,j} \in \mathbb{R}^3$

Keyframe i's pose : $\mathbf{T}_{iw} \in SE(3)$

$$\mathbf{T}_{iw} := \{\mathbf{R}_{iw}, \mathbf{t}_{iw}\}, \mathbf{R}_{iw} \in SO(3)$$

Objective:

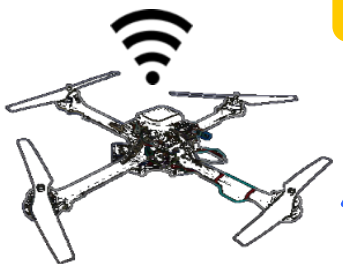
Optimize all $\mathbf{X}_{w,j}, \mathbf{T}_{iw}$ by minimizing the **reprojection error** w.r.t. all matched keypoints $\mathbf{z}_{i,j} \in \mathbb{R}^2$

$$\mathbf{e}_{i,j} = \mathbf{z}_{i,j} - \pi_i(\mathbf{T}_{iw}, \mathbf{X}_{w,j})$$

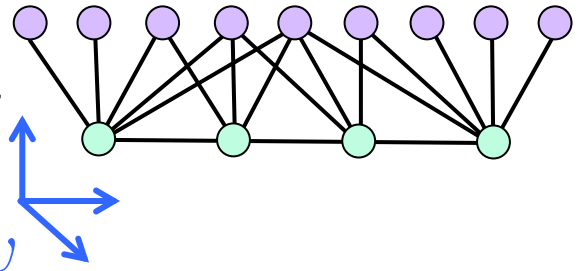
Collaborative SLAM | overview

[ICRA 2017: Schmuck and Chli]
[JFR 2018: Schmuck and Chli]

Robotic Agent



- Visual SLAM with **local BA**
- Limit local map of current environment



Landmark j's 3D location : $\mathbf{X}_{w,j} \in \mathbb{R}^3$

Keyframe i's pose : $\mathbf{T}_{iw} \in SE(3)$

$$\mathbf{T}_{iw} := \{\mathbf{R}_{iw}, \mathbf{t}_{iw}\}, \mathbf{R}_{iw} \in SO(3)$$

Objective:

Optimize all $\mathbf{X}_{w,j}, \mathbf{T}_{iw}$ by minimizing the **reprojection error** w.r.t. all matched keypoints $\mathbf{z}_{i,j} \in \mathbb{R}^2$

$$\mathbf{e}_{i,j} = \mathbf{z}_{i,j} - \pi_i(\mathbf{T}_{iw}, \mathbf{X}_{w,j})$$

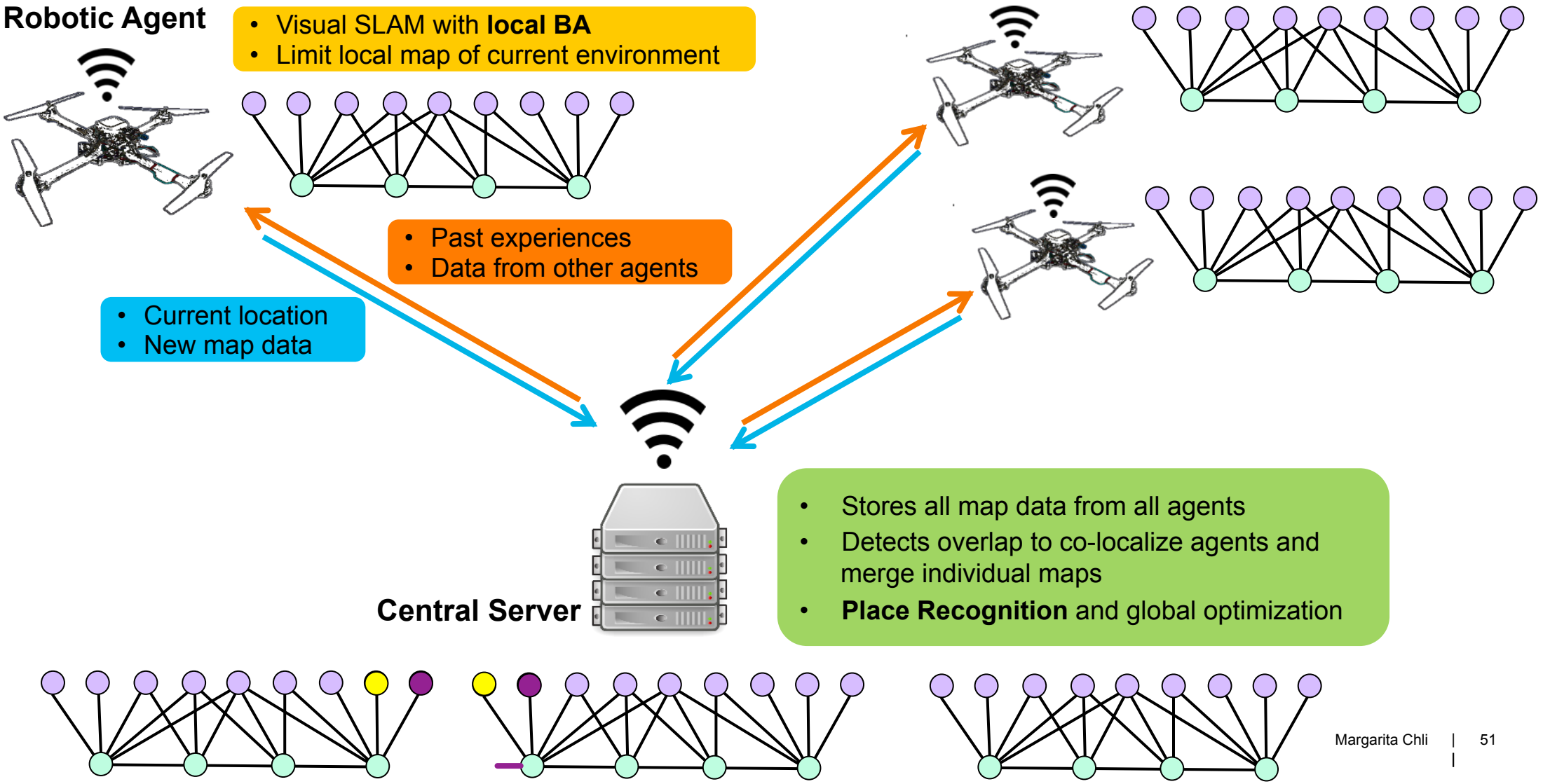
So: $$\underset{\mathbf{X}_{w,j}, \mathbf{T}_{iw}}{\operatorname{argmin}} \sum_{i,j} \rho_h \left(\mathbf{e}_{i,j}^T \Omega_{i,j}^{-1} \mathbf{e}_{i,j} \right) \quad , \; \Omega_{i,j} = \sigma_{i,j}^2 \mathbf{I}_{2\times 2}$$

Covariance matrix associated with the scale at which the keypoint was detected

Huber robust cost function

Collaborative SLAM | overview

[ICRA 2017: Schmuck and Chli]
[JFR 2018: Schmuck and Chli]



Collaborative SLAM | overview

[ICRA 2017: Schmuck and Chli]
[JFR 2018: Schmuck and Chli]

When two agents' maps trigger map fusion

1. **Pose-graph optimization** (binary graph) to distribute error along graph

$\mathbf{S}_{ij}$: the relative $Sim(3)$ transformation between the loop-closing keyframes $\mathbf{T}_{iw}$ and $\mathbf{T}_{jw}$

Error in a pose-graph edge:

$Sim(3)$ representation of $\mathbf{T}_{jw}$

$$\mathbf{e}_{i,j} = \log_{Sim(3)}(\mathbf{S}_{ij}\mathbf{S}_{jw}\mathbf{S}_{iw}^{-1})$$

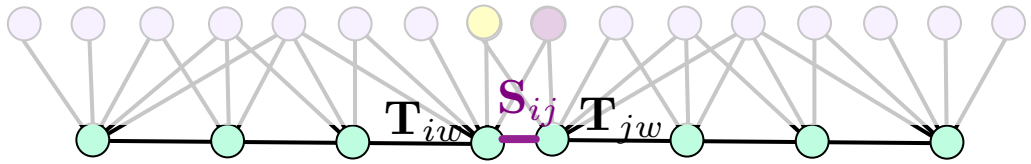
Transforms to the tangent space s.t. $\mathbf{e}_{i,j} \in \mathbb{R}^7$

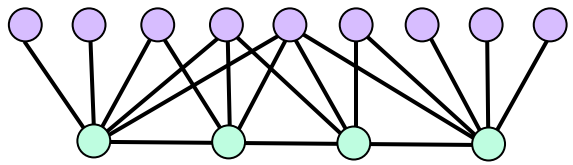
Optimize the $Sim(3)$ keyframe poses s.t.:

$$\underset{\mathbf{S}_{ij}}{\operatorname{argmin}} \sum_{i,j} \rho_h \left(\mathbf{e}_{i,j}^T \Lambda_{i,j} \mathbf{e}_{i,j}\right)$$

Information matrix on this edge

- Stores all map data from all agents
- Detects overlap to co-localize agents and merge individual maps
- **Place Recognition** and **global optimization**





Collaborative SLAM | overview

[ICRA 2017: Schmuck and Chli]
[JFR 2018: Schmuck and Chli]

When two agents' maps trigger map fusion

1. **Pose-graph optimization** (binary graph) to distribute error along graph

$\mathbf{S}_{ij}$: the relative $Sim(3)$ transformation between the loop-closing keyframes $\mathbf{T}_{iw}$ and $\mathbf{T}_{jw}$

Error in a pose-graph edge:

$Sim(3)$ representation of $\mathbf{T}_{jw}$

$$\mathbf{e}_{i,j} = \log_{Sim(3)}(\mathbf{S}_{ij}\mathbf{S}_{jw}\mathbf{S}_{iw}^{-1})$$

Transforms to the tangent space s.t. $\mathbf{e}_{i,j} \in \mathbb{R}^7$

Optimize the $Sim(3)$ keyframe poses s.t.:

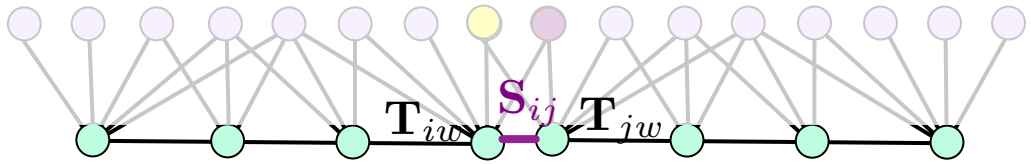
$$\underset{\mathbf{S}_{ij}}{\operatorname{argmin}} \sum_{i,j} \rho_h \left(\mathbf{e}_{i,j}^T \Lambda_{i,j} \mathbf{e}_{i,j}\right)$$

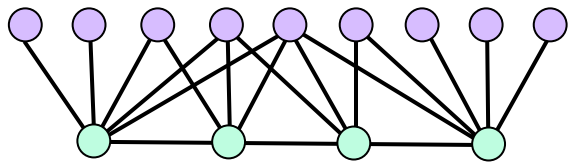
Information matrix on this edge

2. **Perform full global BA**

$$\underset{\mathbf{X}_{w,j}, \mathbf{T}_{iw}}{\operatorname{argmin}} \sum_{i,j} \rho_h \left(\mathbf{e}_{i,j}^T \Omega_{i,j}^{-1} \mathbf{e}_{i,j}\right)$$

- Stores all map data from all agents
- Detects overlap to co-localize agents and merge individual maps
- Place Recognition and **global optimization**





CCM SLAM: [JFR: 2018: Schmuck and **Chli**]

> CCM SLAM code is released

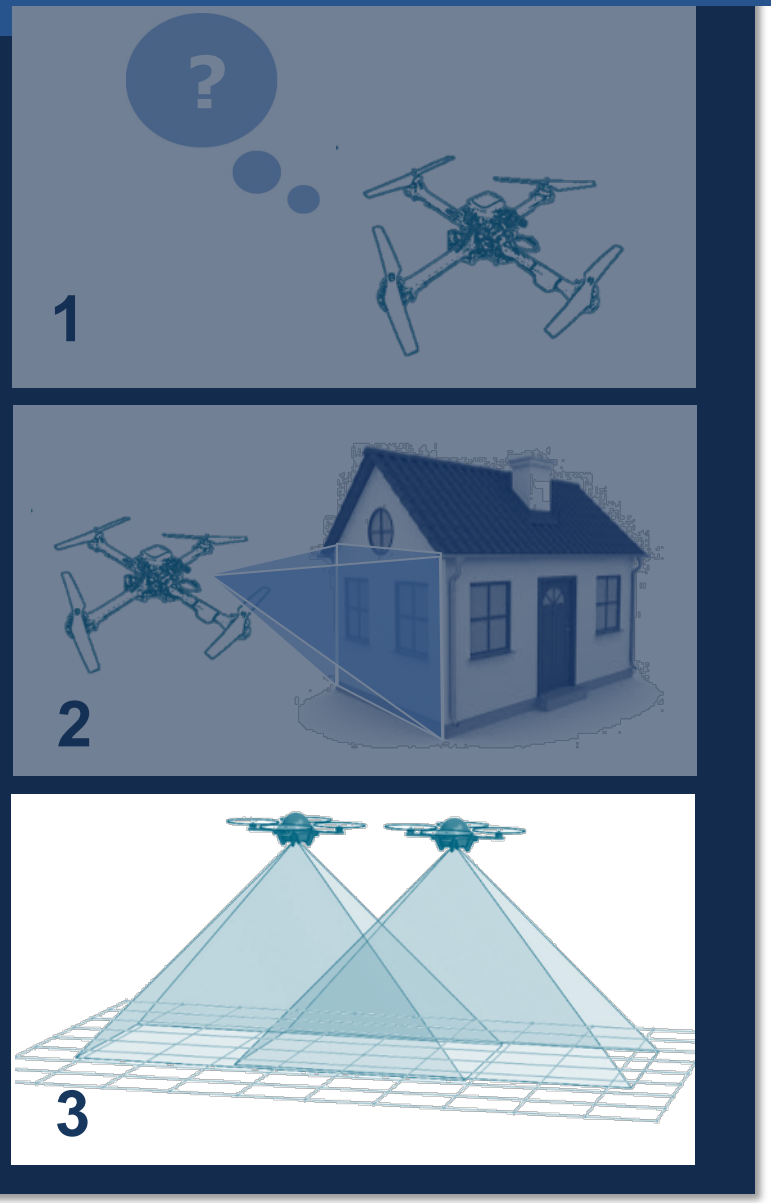
CVI-SLAM – Collaborative Visual-Inertial SLAM

[RAL 2018: Karrer, Schmuck and Chli]

- Visual-inertial odometry front-end
- New optimization back-end for Visual-Inertial SLAM [ICRA 2018: Karrer and Chli]

- Visual-inertial setup:
  - Metric scale
  - Gravity alignment
  - Information between camera frames

- Architecture & communication protocol from [Schmuck and **Chli**, JFR 2018]



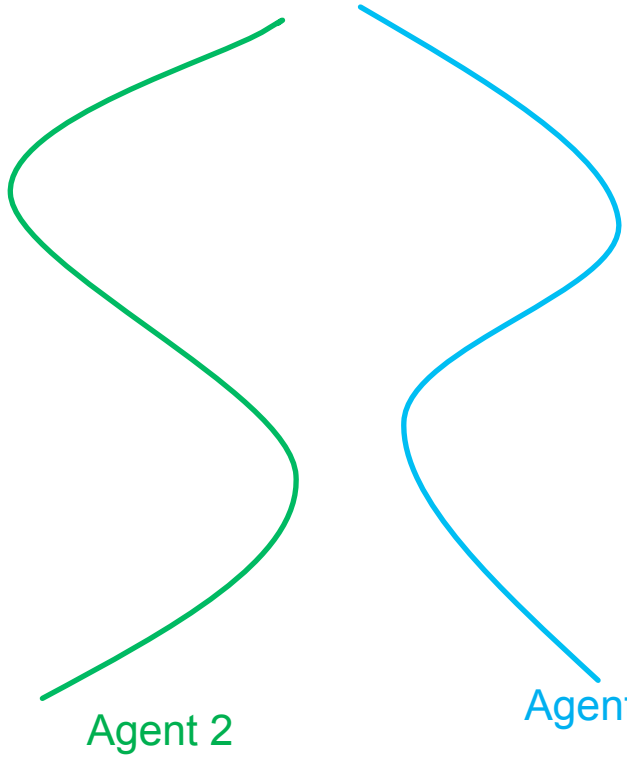
CVI-SLAM: Accuracy

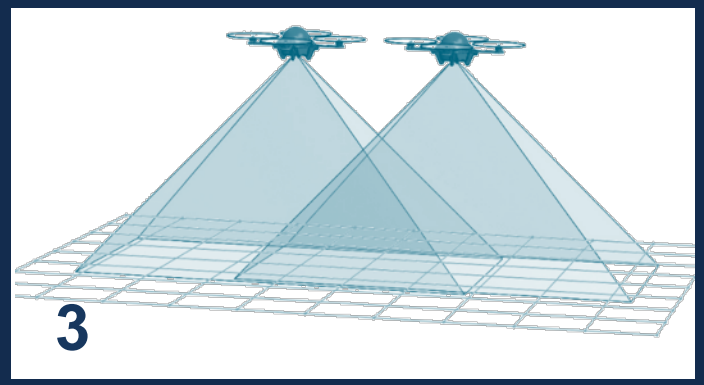
[RAL 2018: Karrer, Schmuck and Chli]

- Single-agent trajectory RMSE: comparable to state-of-the-art methods (VI-ORB-SLAM, VINS-mono)
- Pose RMSE:

Sequences	Single Agent
MH1 & MH2	2.24 cm
MH2 & MH3	2.95 cm
MH4 & MH5	4.12 cm

On the EuRoC datasets





CVI-SLAM: Accuracy

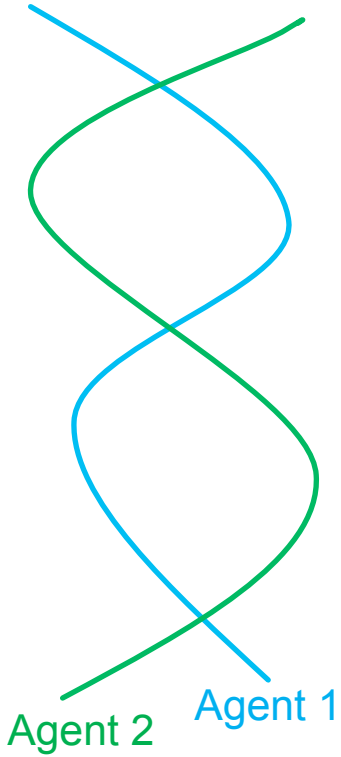
[RAL 2018: Karrer, Schmuck and Chli]

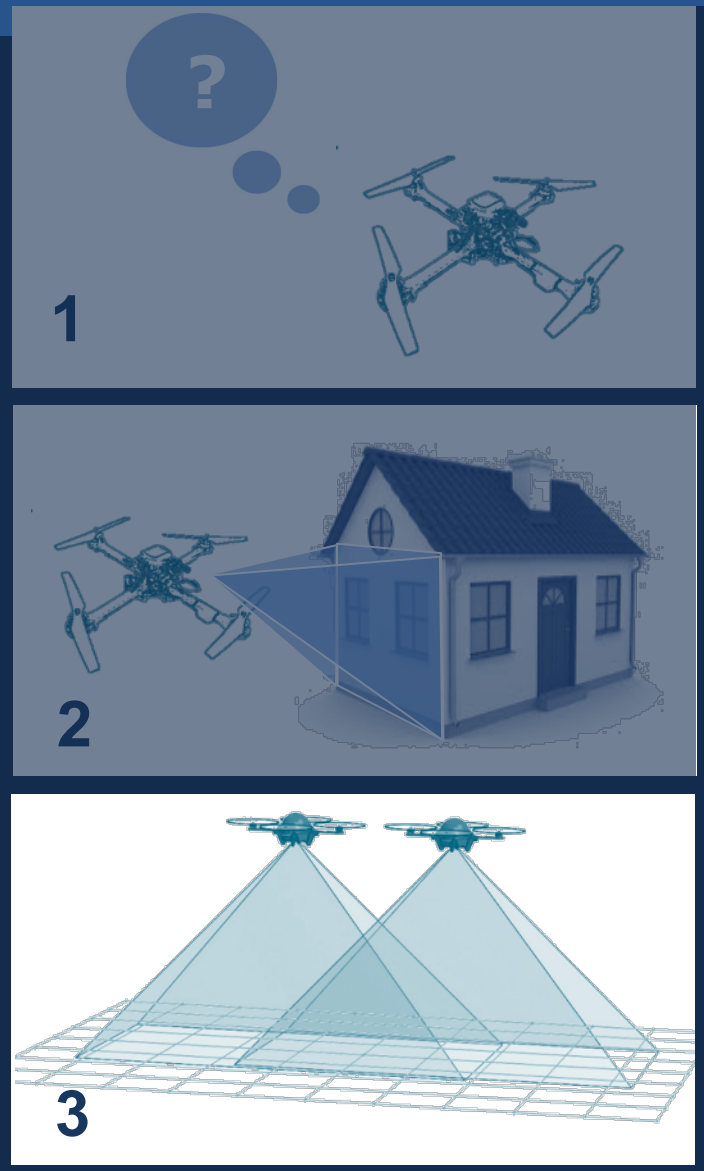
- Single-agent trajectory RMSE: comparable to state-of-the-art methods (VI-ORB-SLAM, VINS-mono)
- Pose RMSE:

Sequences	Single Agent	Collaboration
MH1 & MH2	2.24 cm	1.39 cm
MH2 & MH3	2.95 cm	2.56 cm
MH4 & MH5	4.12 cm	3.40 cm

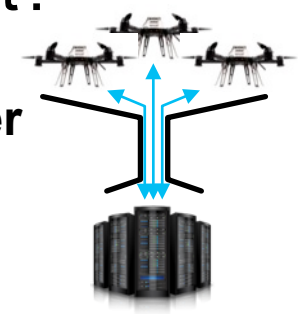
On the EuRoC datasets

✓ Multi-agent collaborative RMSE: improved accuracy not only after, but also ***during*** the mission



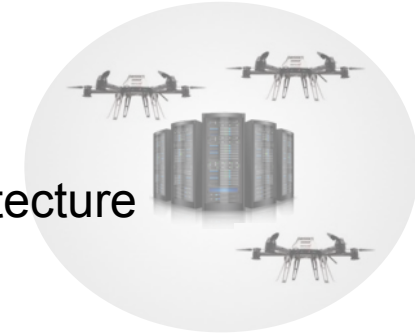


3. Multi-agent Collaboration: what's next?



Centralized: all data needs to pass through the server

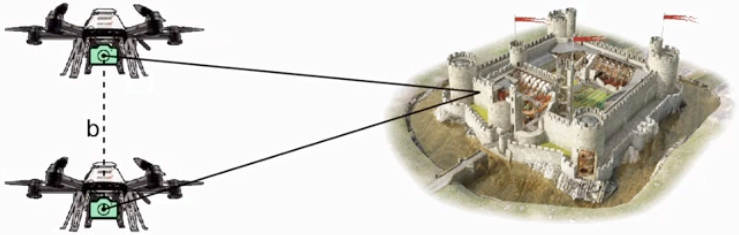
- Remove redundant data
- Work towards distributed collaboration



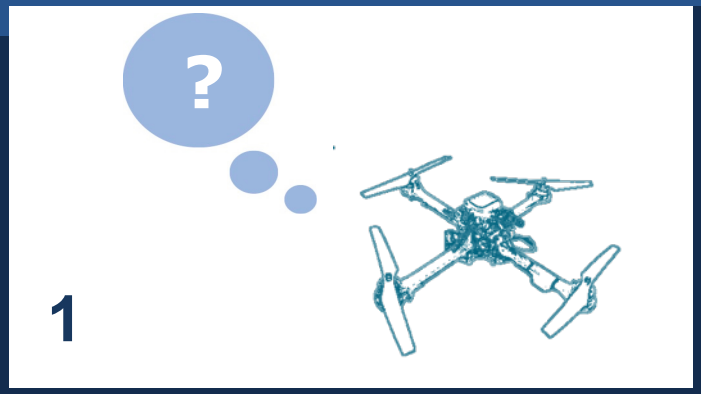
Server accessibility limits mission range

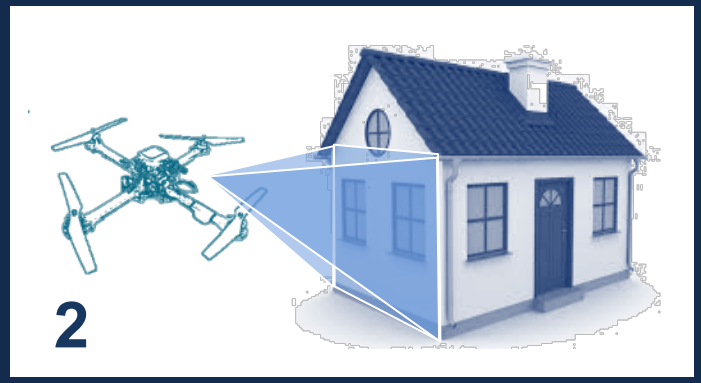
- Use cloud computing or a mobile server
- … or peer-to-peer communication in a distributed architecture

Enable stronger collaboration



[Karrer et al., ICRA 2018]





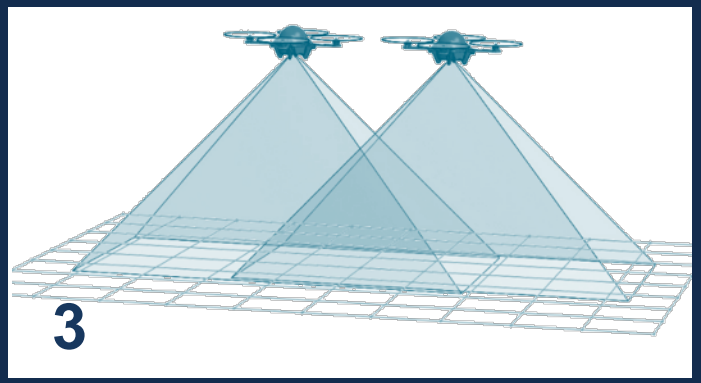


Photo credit: Mario Trimikliniotis

Robotic **Perception** & **Collaboration** are **key** to Robotic Autonomy

The Vision

Develop

collaborative

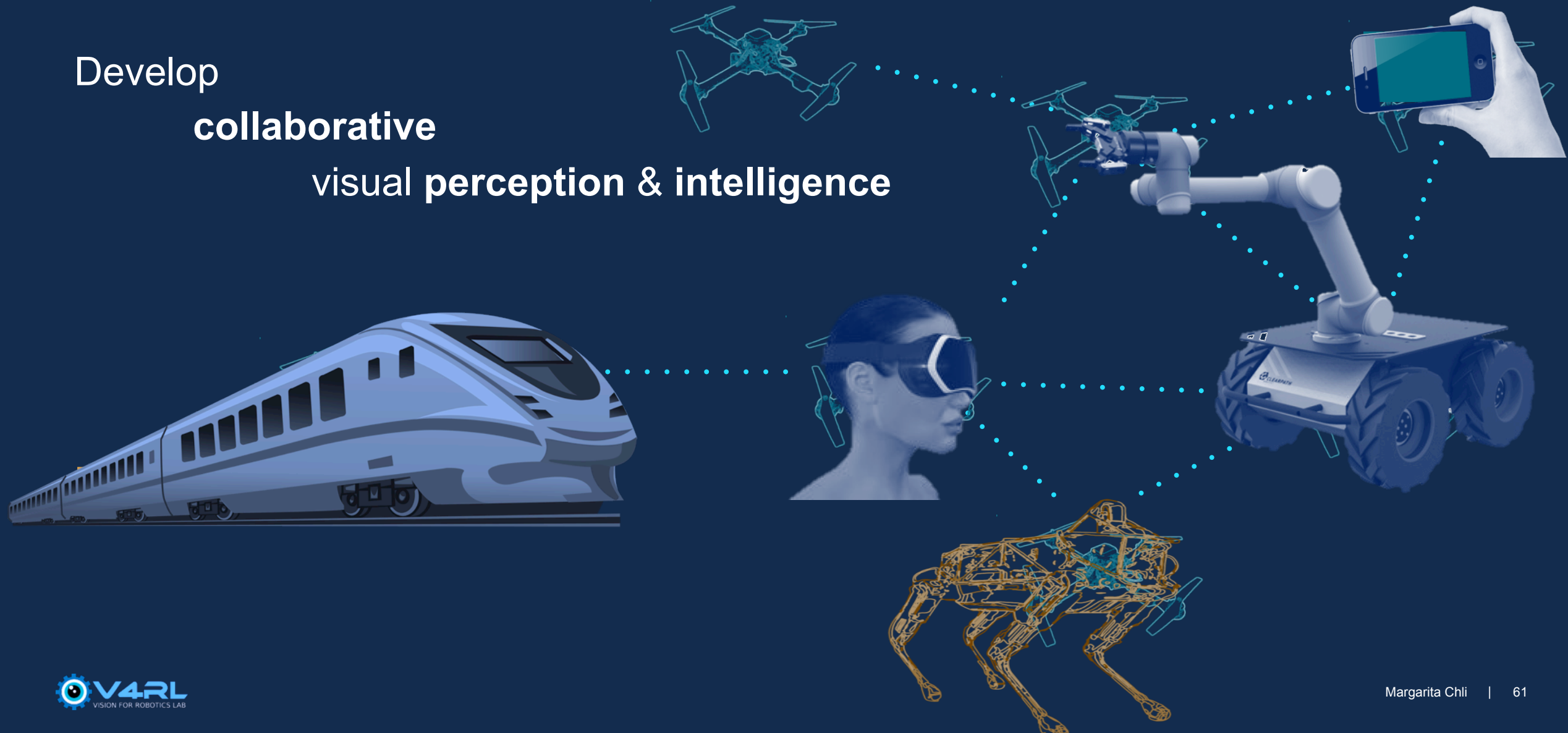
visual **perception** & **intelligence**

The Vision

Develop

collaborative

visual **perception** & **intelligence**



The Vision -- ongoing projects

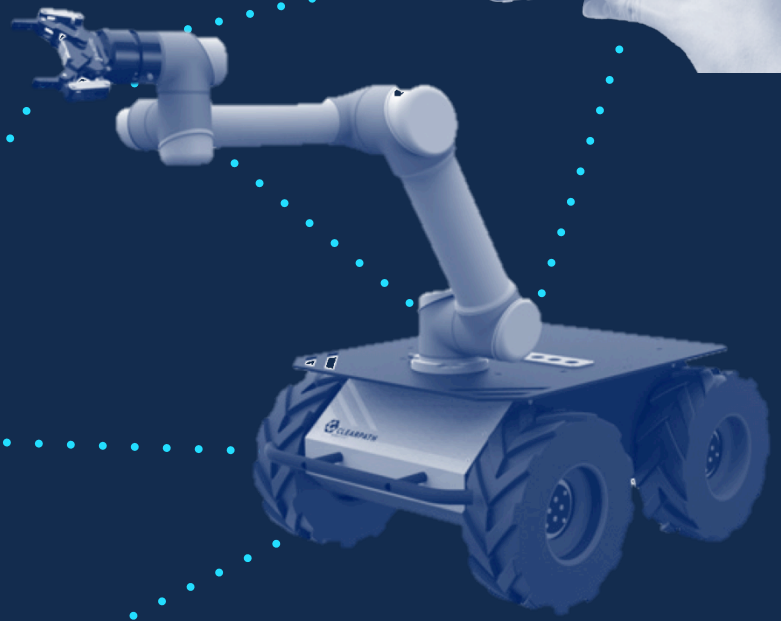
V4RL VISION FOR ROBOTICS LAB

armasuisse
Science + Techology

Develop

collaborative

visual **perception** & **intelligence**



SBB CFF FFS