Autonomous Multi-Modal Localization and Mapping: Fundamentals and the State-of-the-Art

Christos Papachristos

Autonomous Robots Lab, University of Nevada, Reno

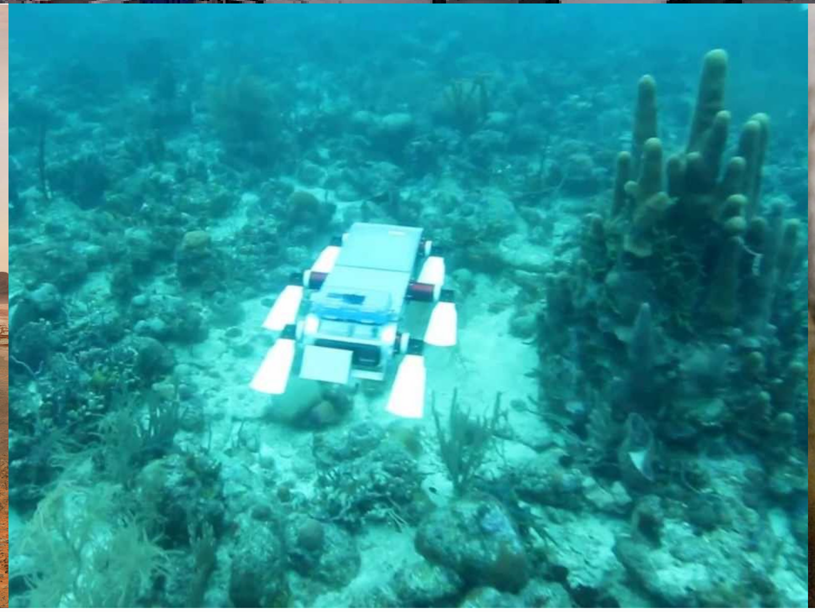
Introduction

Autonomous Multi-Modal Localization and Mapping: Fundamentals and the State-of-the-Art

The base questions

- Where am I now?

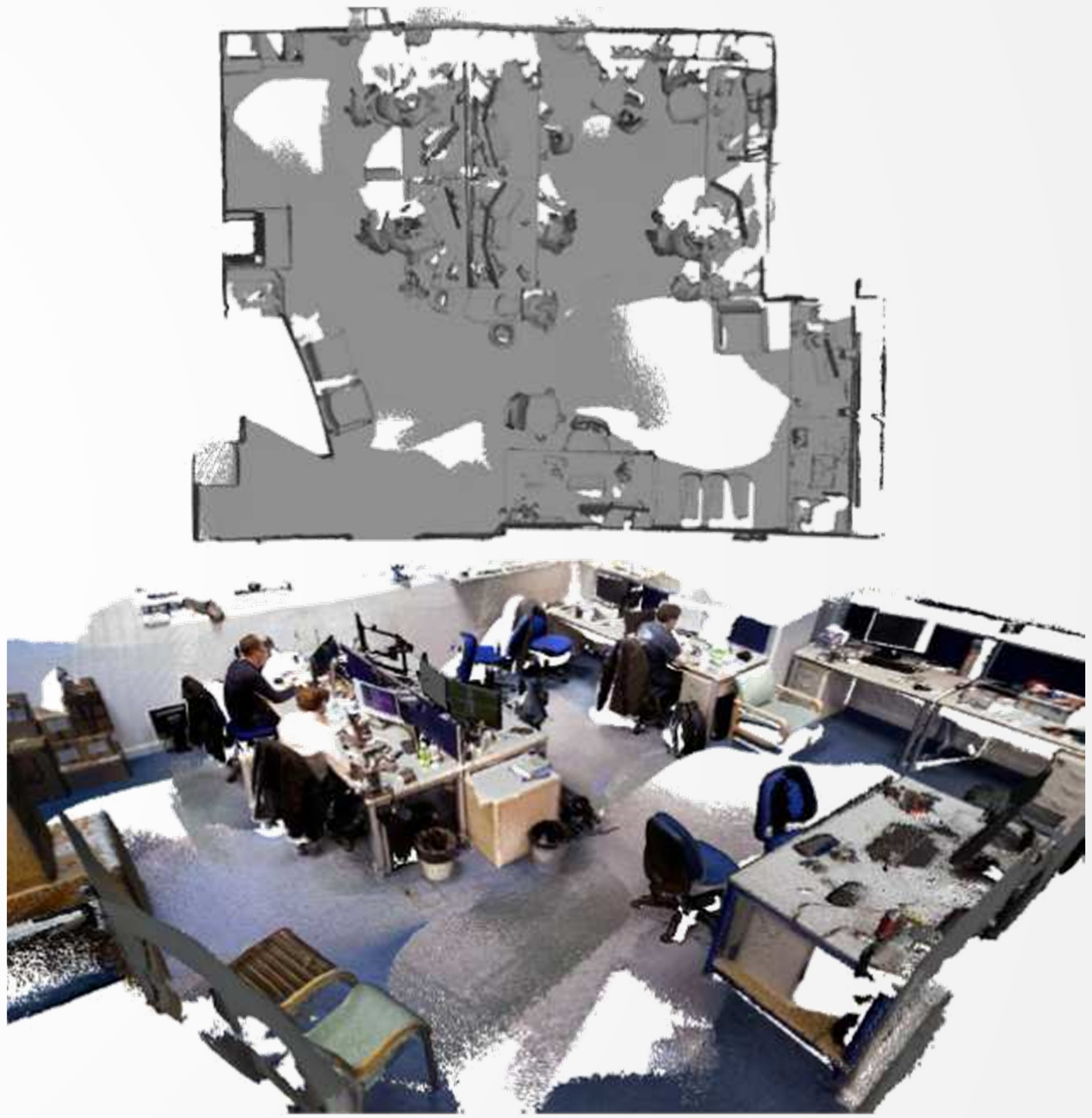
- What am I looking at?

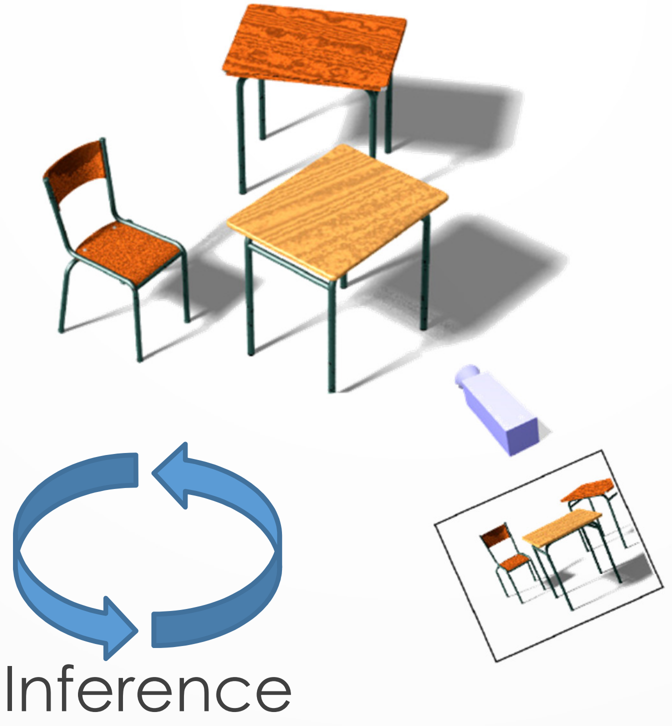

- Common to all mobile robots that "want" to interact (manipulate, navigate, actively observe, etc.) with their environment.

The base problem

- "Real" Autonomy
- The chicken & egg challenge:
  - Localize against what? A map is needed!
  - Map where? A pose is needed!

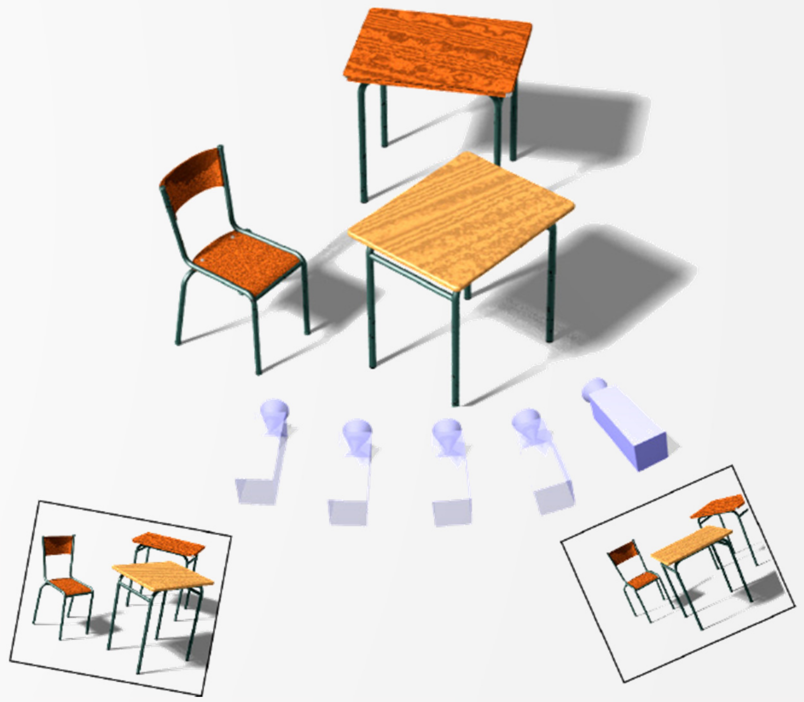


- Where am I now?
- What am I looking at?



The history

- **S**tructure **F**rom **M**otion
- The general problem of recovering sensor poses and 3-dimensional structure from a set of sensor snapshots.
  - Potentially unordered.
  - Typically refers to passive cameras - minimal SWaP footprint, nature- inspired.
- Early works date back to the first decades of mobile robot research. Field carries influence from Photogrammetry:
  - H. Longuet-Higgins, “A computer algorithm for reconstructing a scene from two projections,” Nature, 1981.
  - C. Harris and J. Pike, “3d positional integration from image sequences,” in Proc. Alvey Vision Conference, 1988.

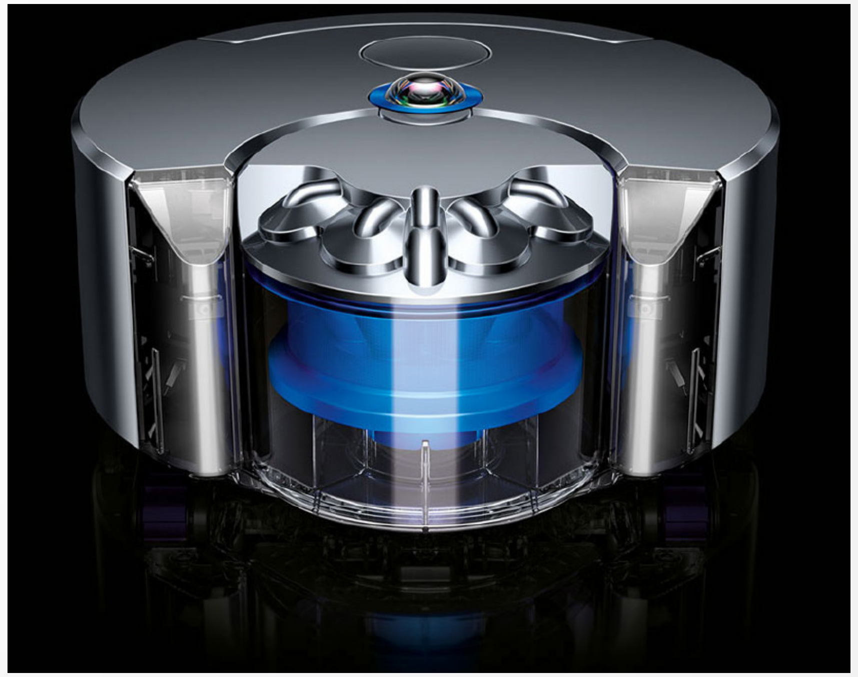


The applications

- **S**imultaneous **L**ocalization **A**nd **M**apping
- SLAM is more of a concept rather than a single algorithm. Can be implemented using:
  - Different hardware, sensor types, sensor configurations.
  - Different methods, algorithms, processing schemes.
- Is it important?

- Autonomous vehicles

- Augmented Reality



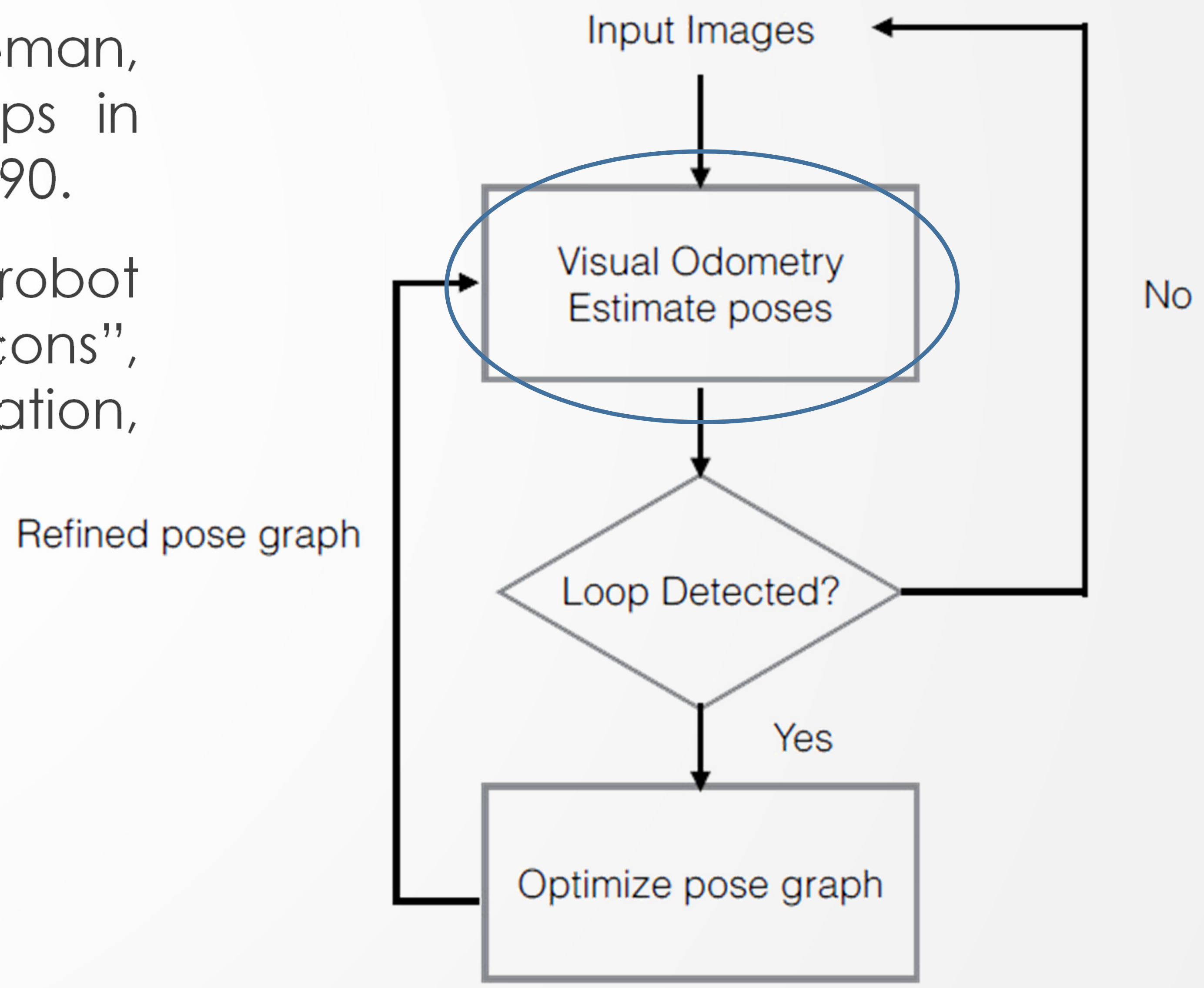
- "Simple" Appliances

The basics

Autonomous Multi-Modal Localization and Mapping:
Fundamentals and the State-of-the-Art

The basics

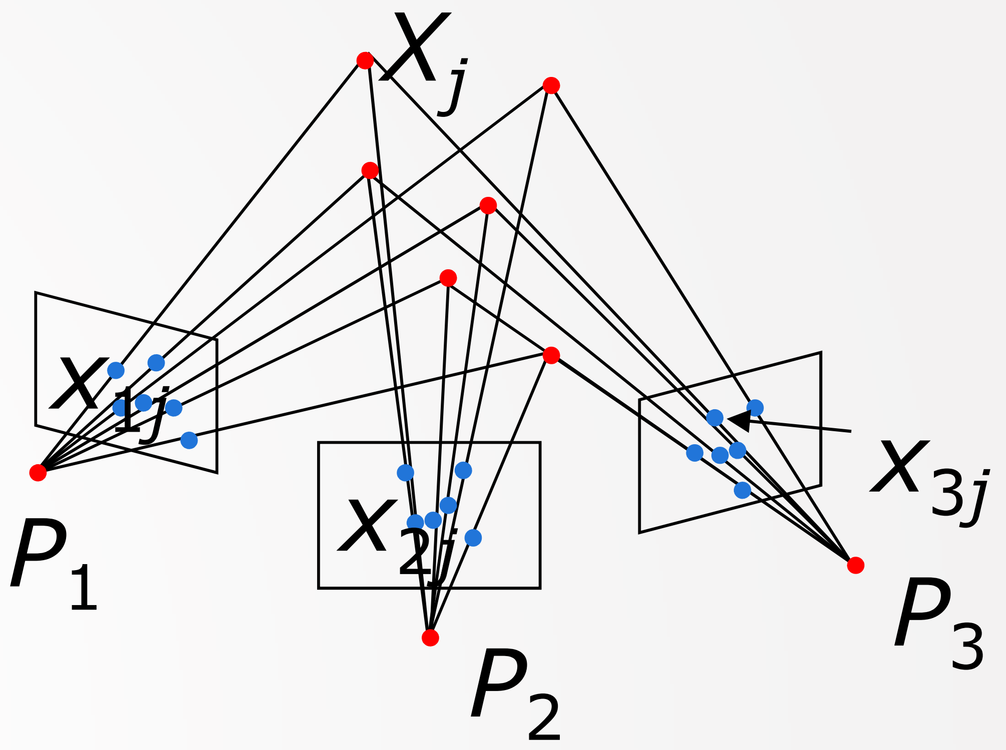
- **S**imultaneous **L**ocalization **A**nd **M**apping
- Term first coined decades ago.
  - Randall Smith, Matthew Self, Peter Cheeseman, "Estimating Uncertain Spatial Relationships in Robotics", Autonomous Robot Vehicles, 1990.
  - Leonard, Durrant-Whyte, "Mobile robot localization by tracking geometric beacons", IEEE Transaction on Robotics and Automation, 1991.
- Sensor-based inference.
  - proximity (sonar)?
- Generalized case of Visual-SLAM:
  - Front-end tracking
  - Back-end mapping



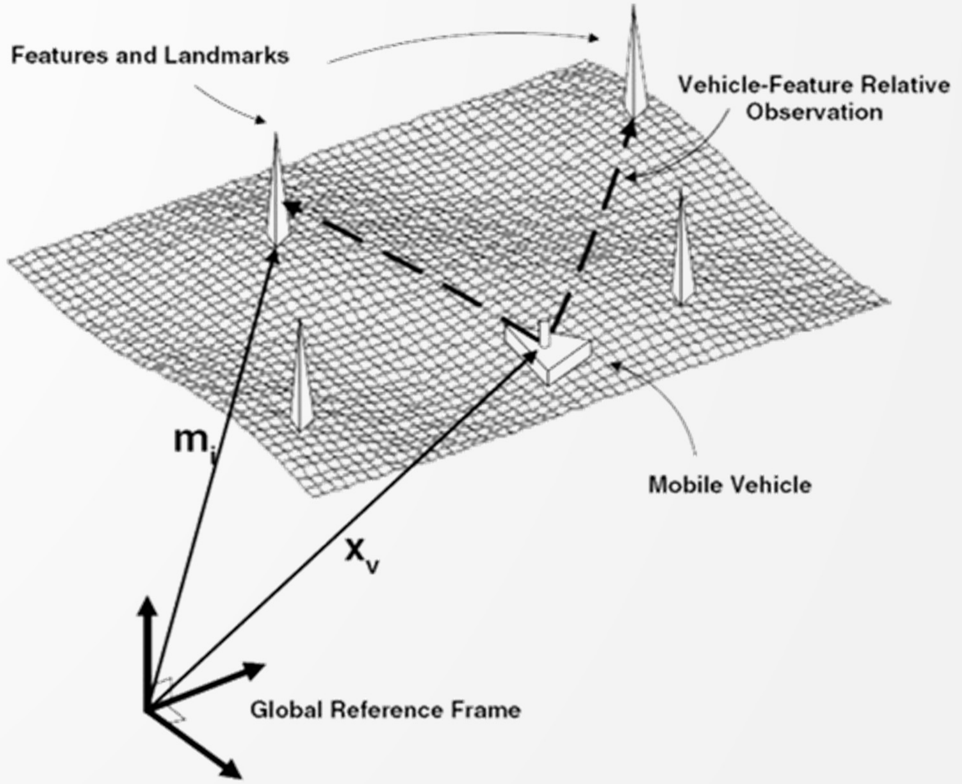
The basics

- **V**isual **O**dometry
- The problem of estimating a vehicle's egomotion from Visual input alone.
  - "Odometry" term inspired by wheeled robots.
  - Actually to address problems of wheel slippage on NASA Mars rovers (eneven / rough terrain).
  - Generalized 6-DoF motion estimation.
  - The first Motion Estimation Pipeline & the earliest Corner Detector:

    H. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," Ph.D. dissertation, Stanford Univ., 1980.
- Visual Odometry
  - Estimate 6-DoF pose [**R** | **T**] "incrementally"



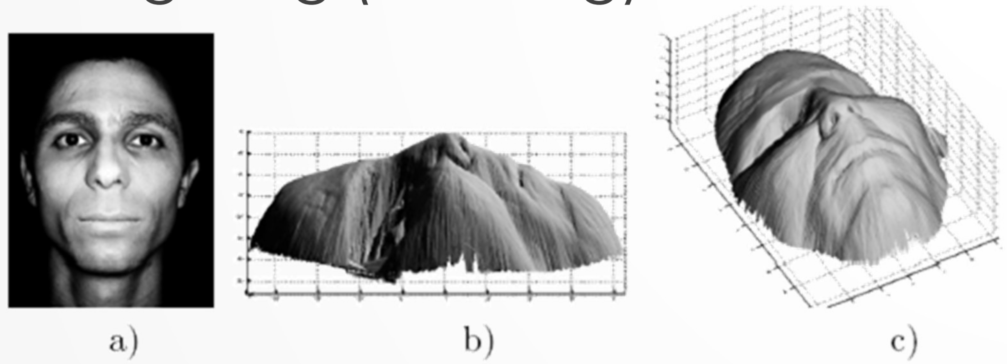
- Features and Landmarks



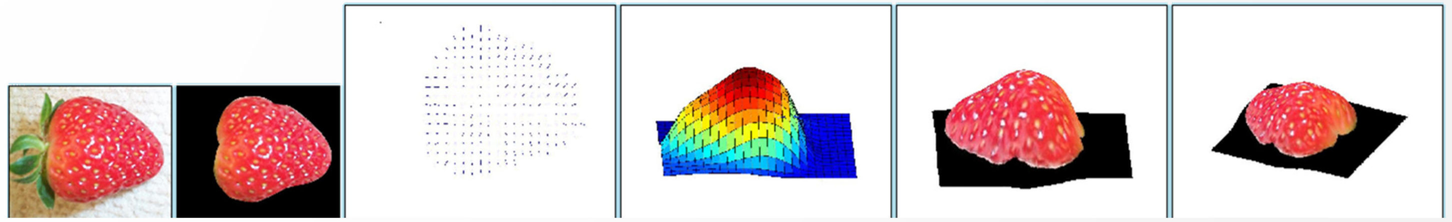
The basics

- Camera
- Basically a bearing sensor:
  - Structure and depth are ambiguous from single snapshots.
- But image is very rich in additional cues:

- Lighting (Shading)

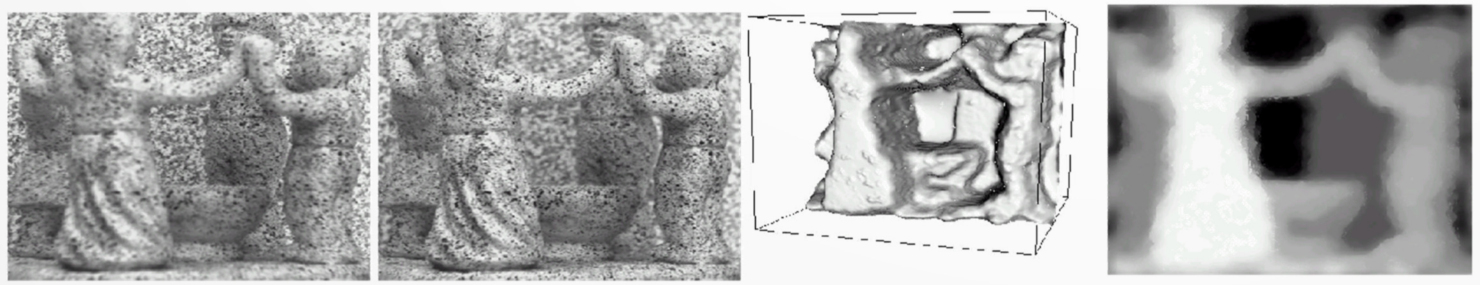


- Texture



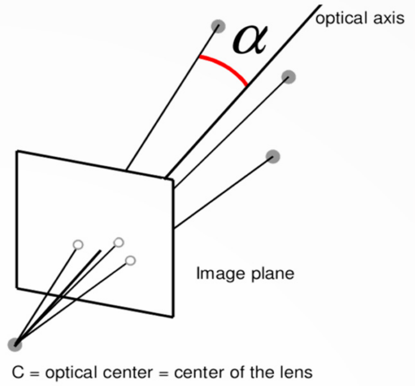
- Perspective

- Camera Parameters (Focus / Defocus)

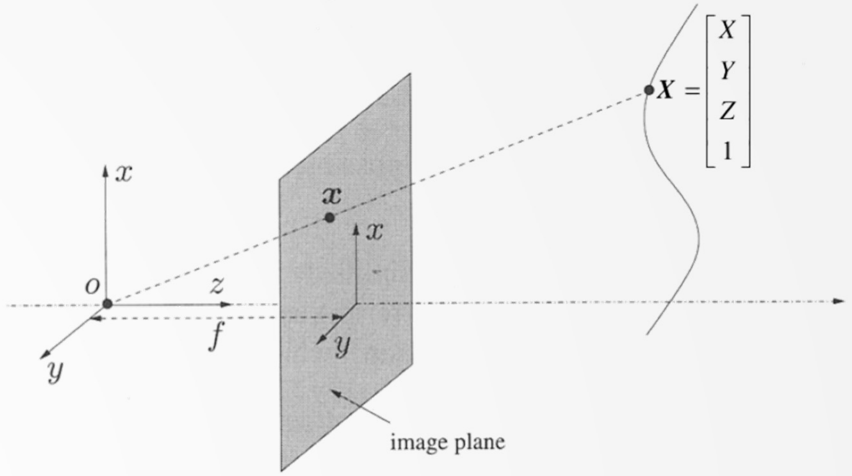


The basics

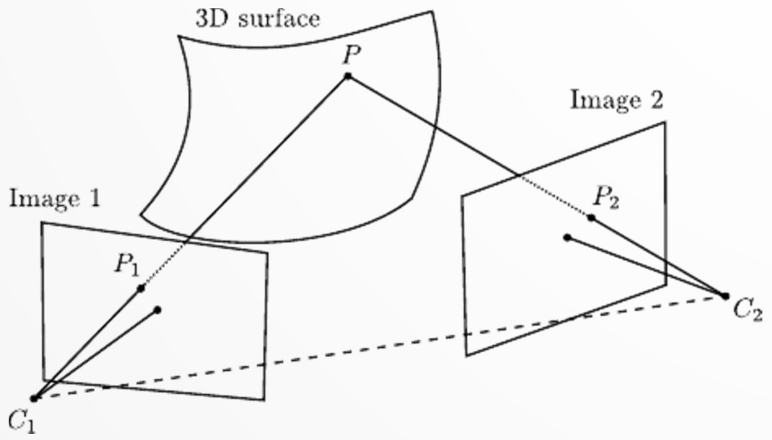
- Camera
- Basically a bearing sensor:

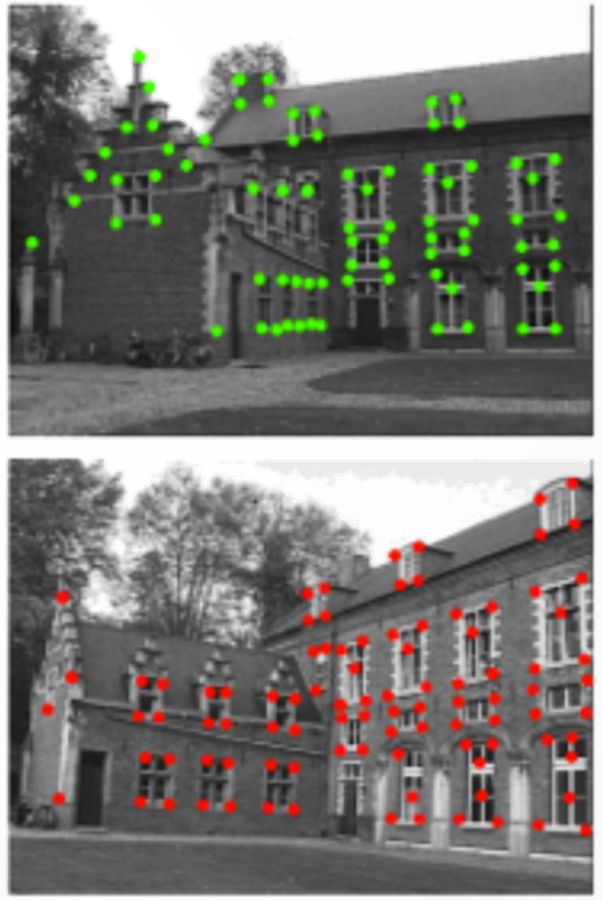


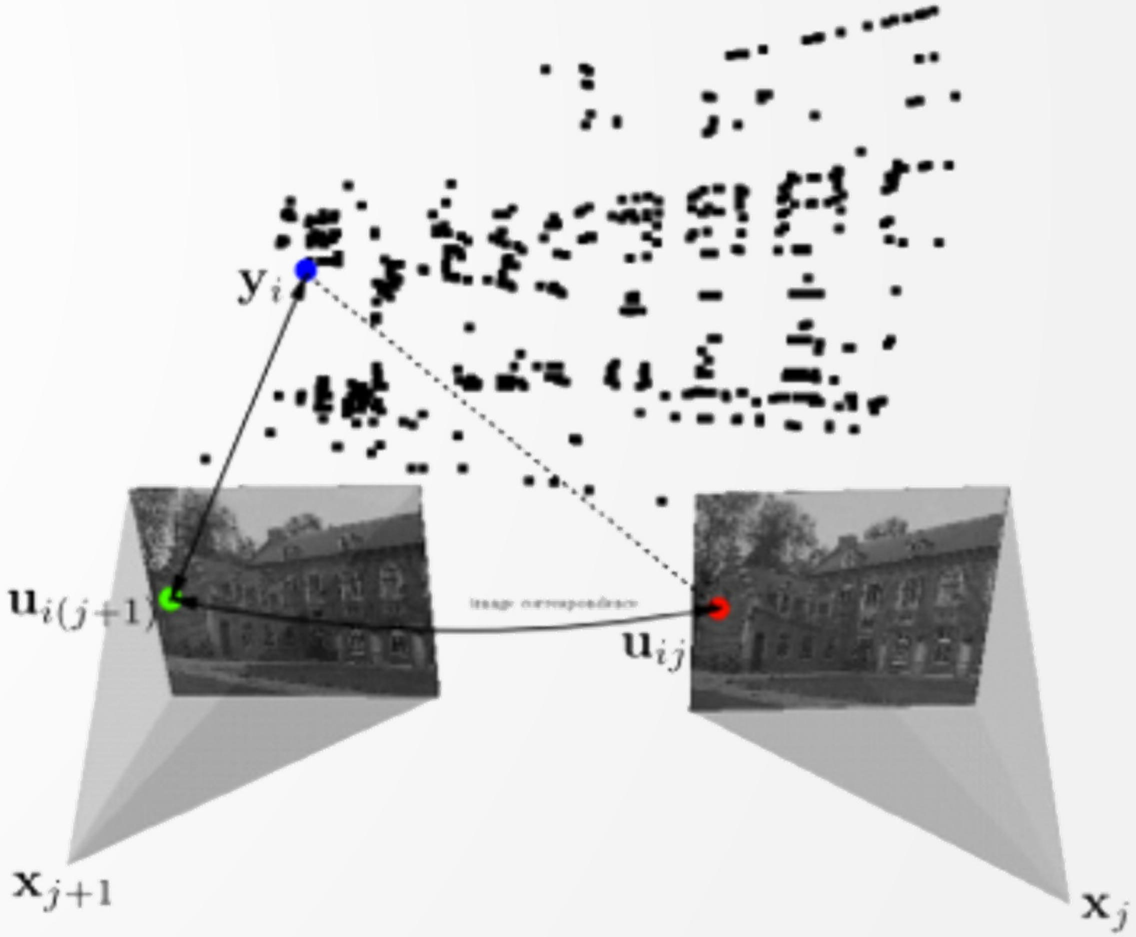
- Even with motion pose is recoverable up to a scale !



- Mobile Robotics !
  - Structure From **Motion**
  - Triangulation
  - Epipolar Geometry



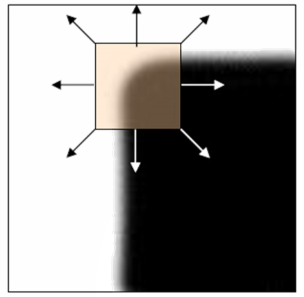




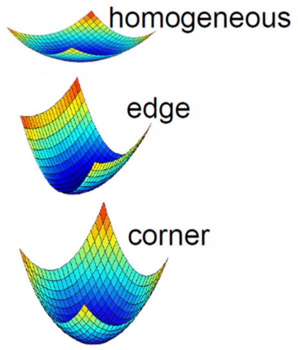
- More: Hartley, R.I. and Zisserman, A., "Multiple View Geometry in Computer Vision", Cambridge University Press

The basics

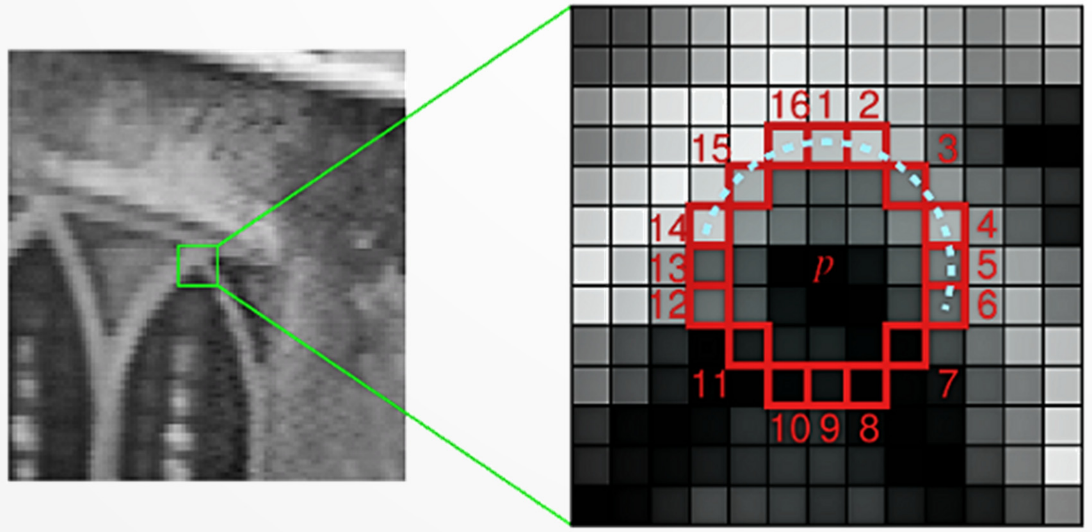
- Landmark Detection & Tracking from Image Features
  - Corner Detection:
    - FAST / AGAST
    - SIFT
    - SURF



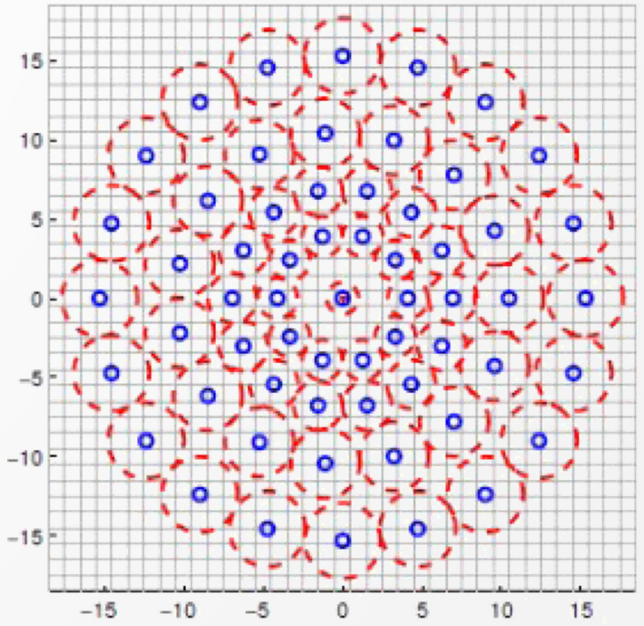
shift at least
2 directions



  - FAST/AGAST:
    - Features from Accel. Segment Test (9/16)
    - Particular efficiency for Real-Time application
    - Further Acceleration with Machine Learning

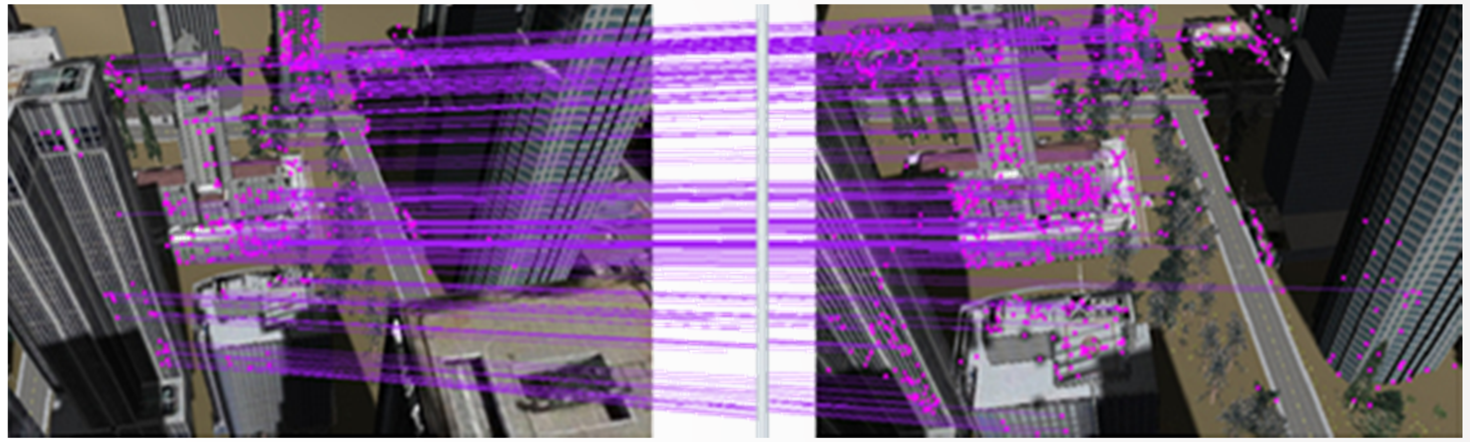


  - Descriptor Computation:
    - SIFT
    - SURF
    - BRISK
  - Invariance (Scale – Rotation – Affine T)
    - BRISK Sampling Pattern

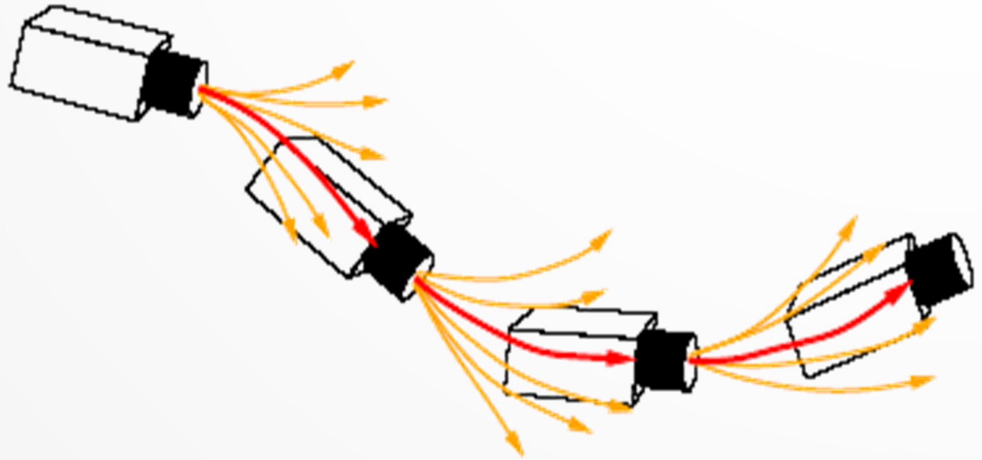


The basics

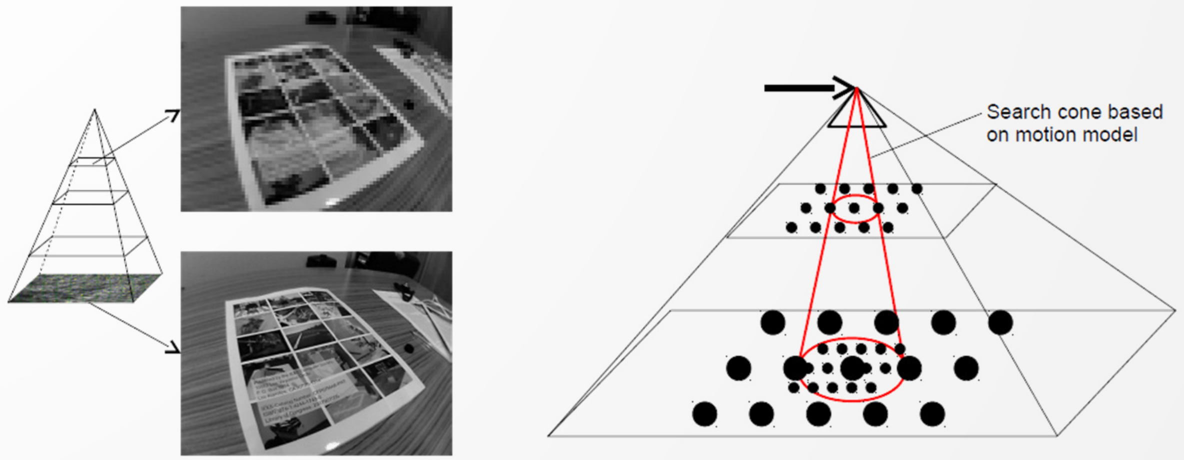
- Image Features – the Bottleneck
- Detection – Descriptors – Matching
  - Need to iterate 100s of times per frame
  - Need to happen in the order of [ms]



- Mobile Robotics ! Constrain the search region
  - Apply motion model



  - Apply pyramidal search

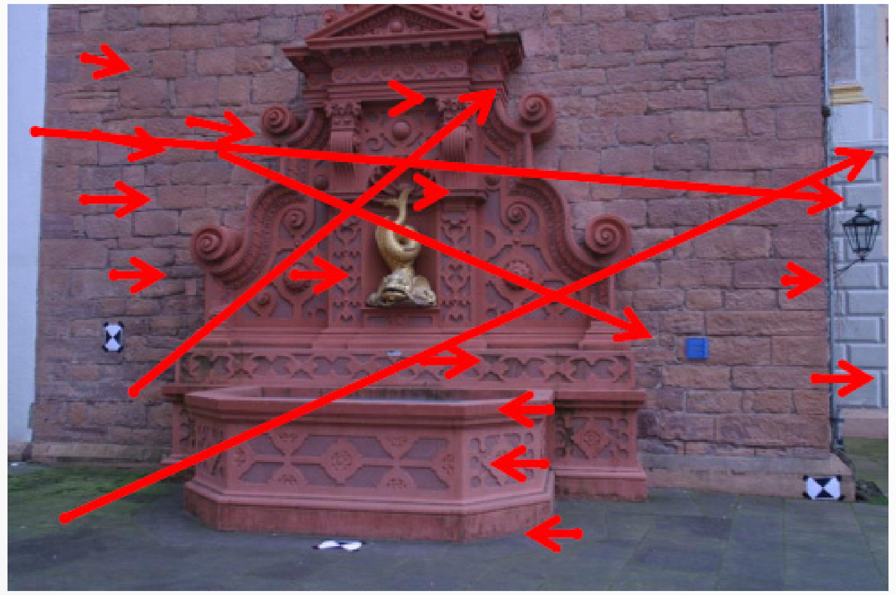


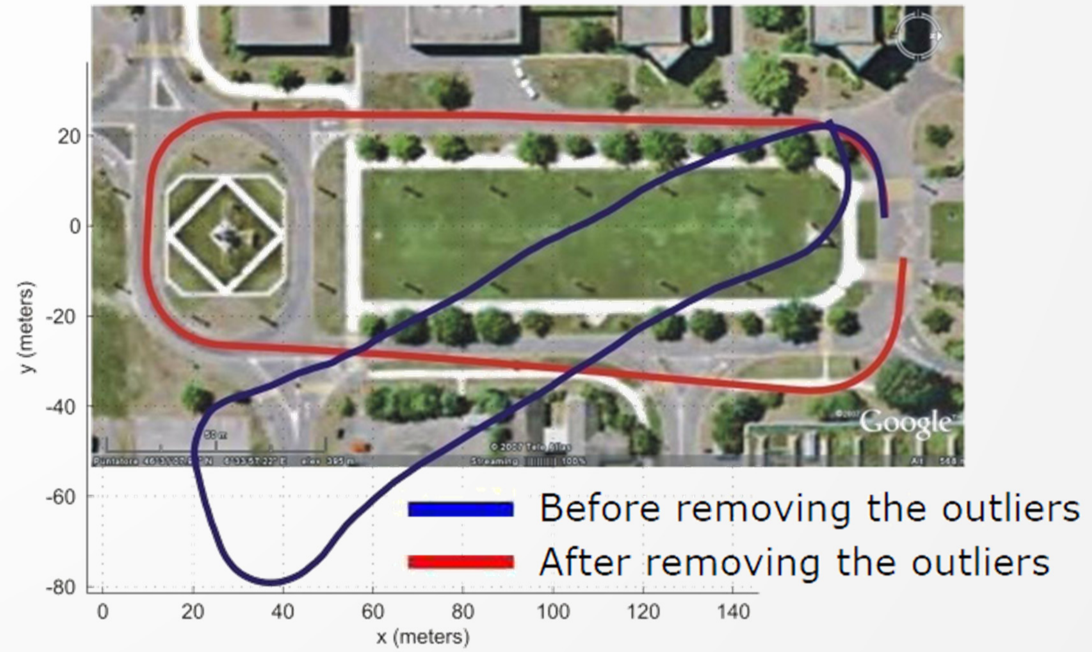
The basics

- Image Features:
  - Feature Matching is not perfect
    - Detection – Descriptors – Matching

Detector	Descriptor	Accuracy	Relocalization & Loop closing	Efficiency
Harris	Patch	++++	-	+++
Shi-Tomasi	Patch	++++	-	+++
SIFT	SIFT	++	++++	+
SURF	SURF	++	++++	++
FAST	BRIEF	++++	+++	++++
ORB	ORB	++++	+++	++++
FAST	BRISK	++++	+++	++++

  - Robust Estimation – **RANSAC**
    - Outlier rejection (actually "model fitting")
    - Robust outlier rejection **over** sophisticated features.

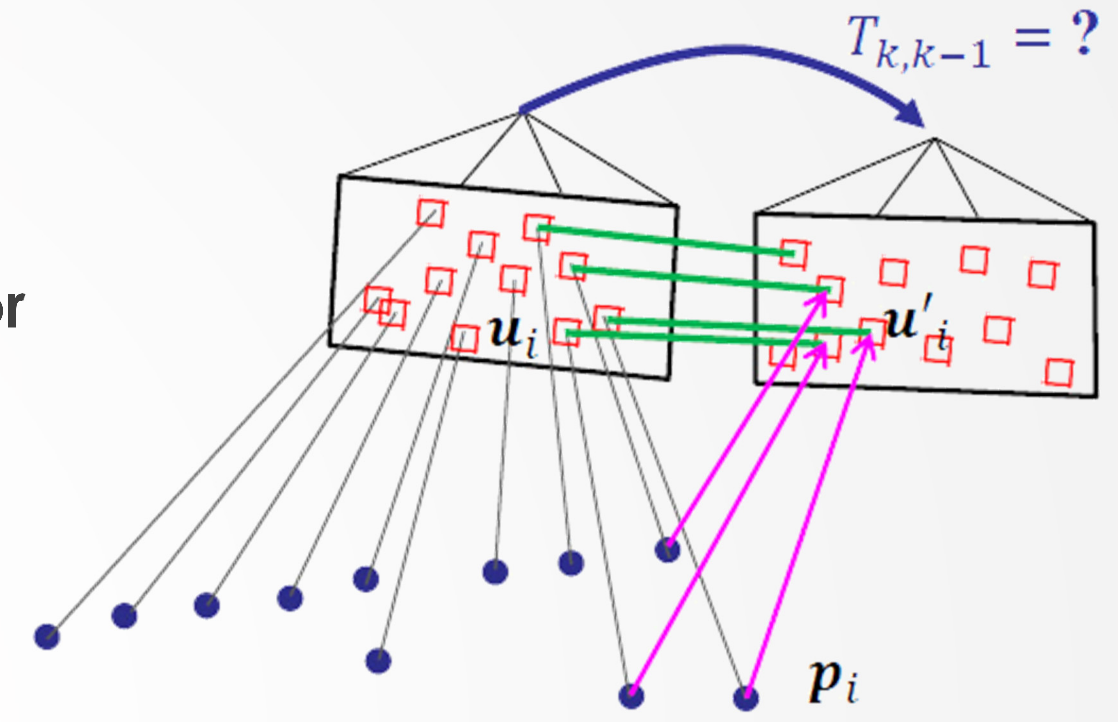




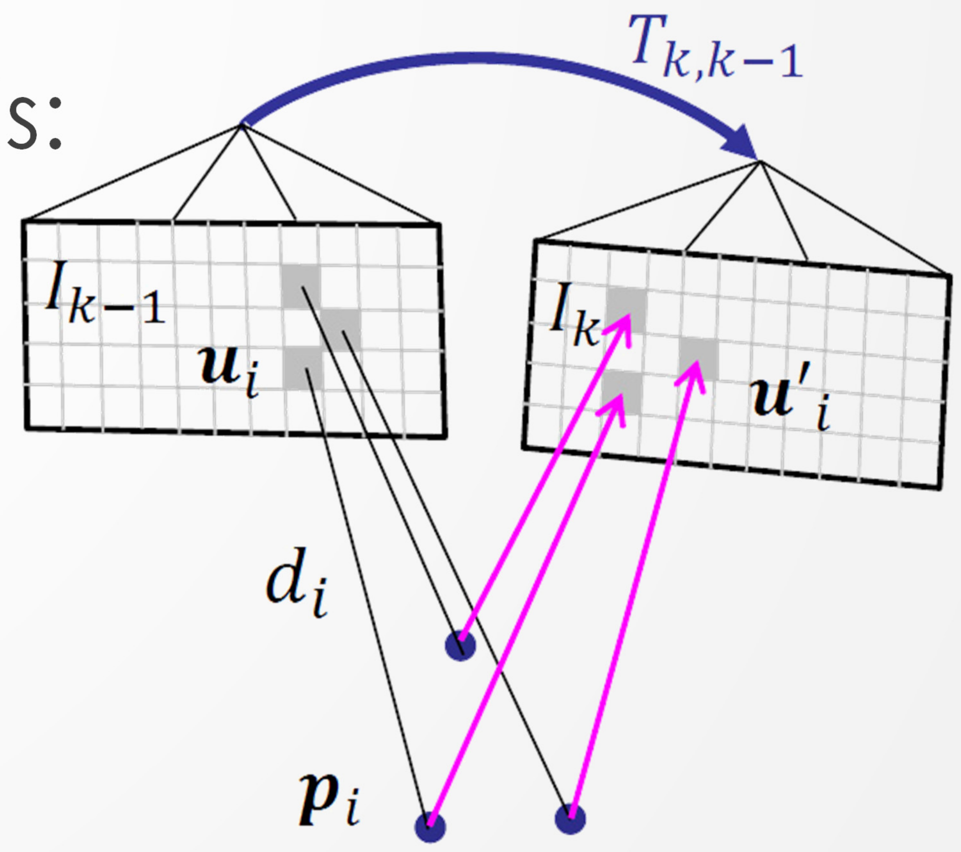
  - Critically affect eigenvalue decomposition

The basics

- Image Feature-based Methods:
  - (Non-linear) Minimization of **Reprojection Error**
  - ✓ Large frame-to-frame motions
  - ✓ Accuracy: Efficient optimization of structure and motion (Bundle Adjustment)
  - × Slow due to costly feature extraction and matching
  - × Matching Outliers (RANSAC)



- Image Appearance-based Methods:
  - Minimization of **Photometric Error**
  - ✓ All information in the image can be exploited (precision, robustness)
  - ✓ Increasing camera frame-rate reduces computational cost per frame
  - × Limited frame-to-frame motion
  - × Joint optimization of dense structure and motion too expensive

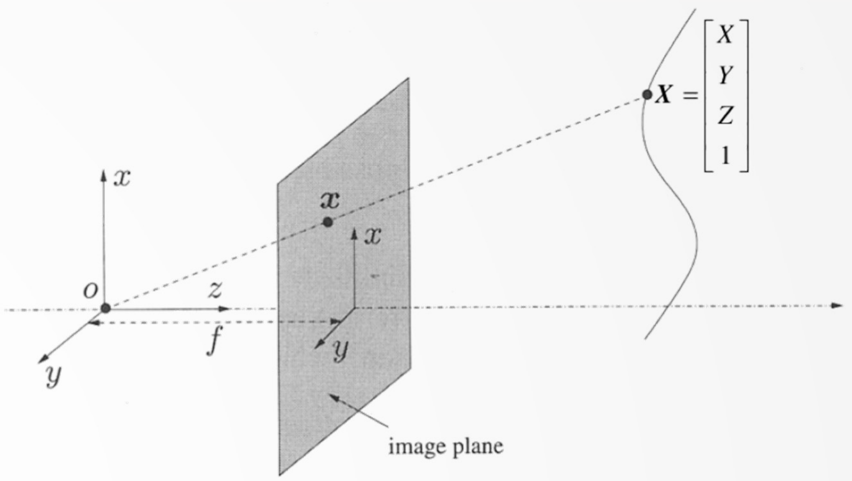


(Sparse) Feature Approaches

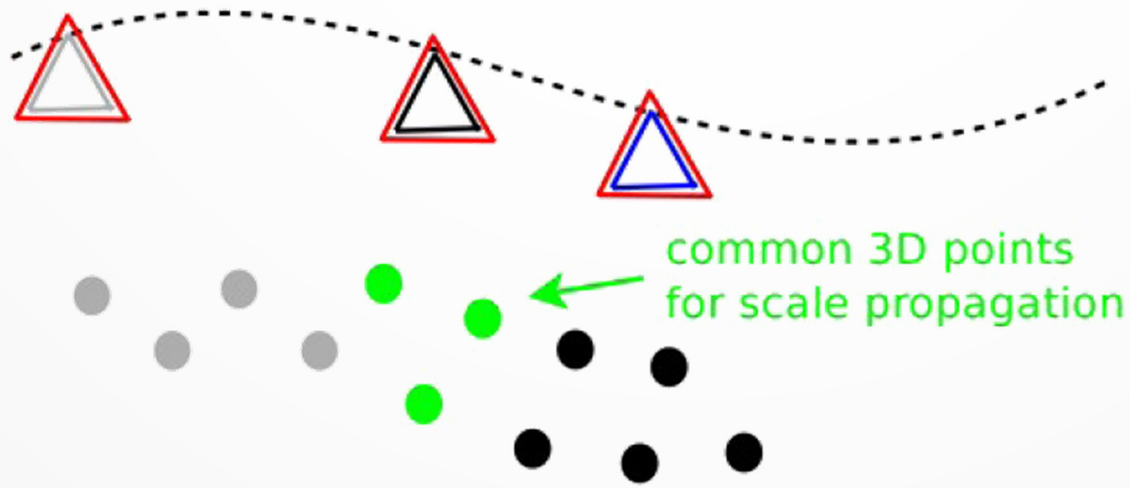
Autonomous Multi-Modal Localization and Mapping:
Fundamentals and the State-of-the-Art

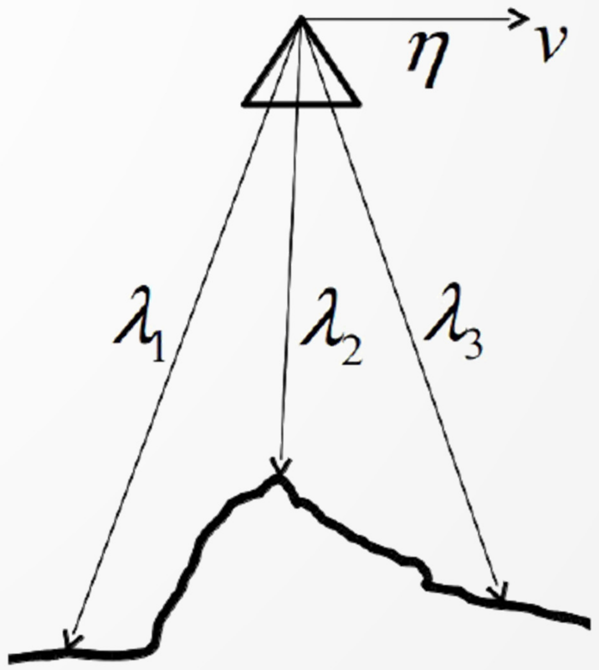
Landmark SLAM

- Monocular VO:
  - Motion is recoverable up to a scale factor



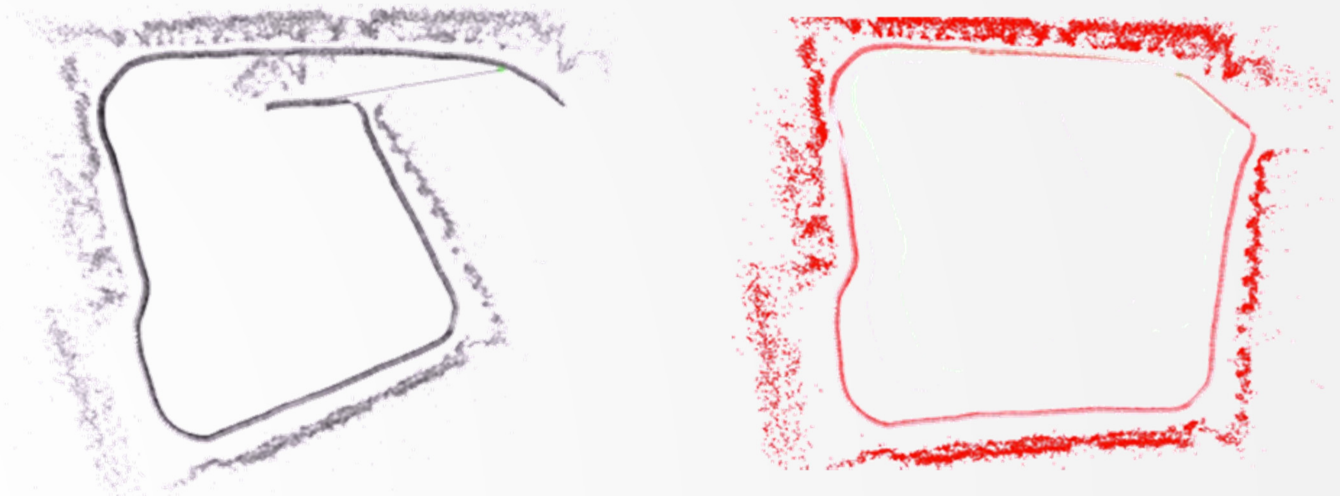
- 3D Landmark (point)-pairs triangulation
  - No image-to-image pair absolute scale
  - Transformation-to-transformation relative scale
  - Common point-pair distance ratio





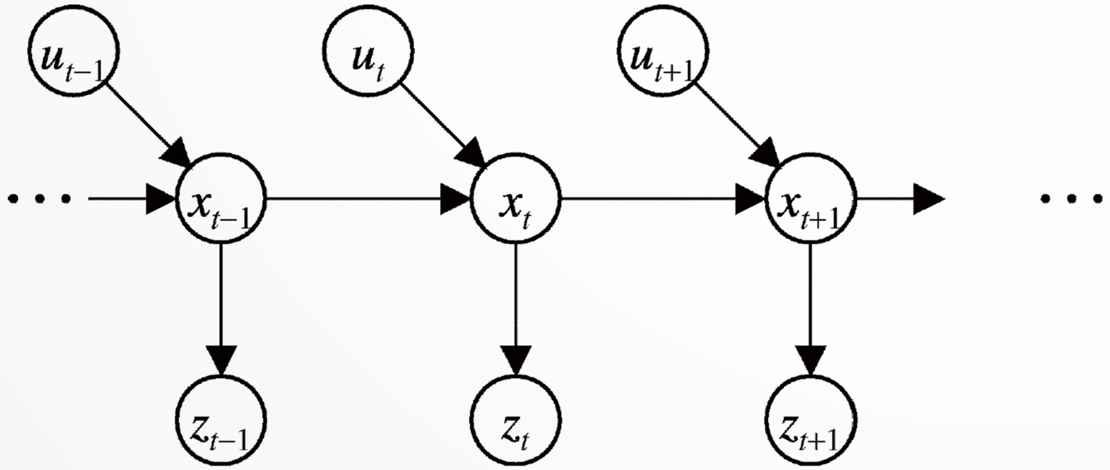
Landmark SLAM

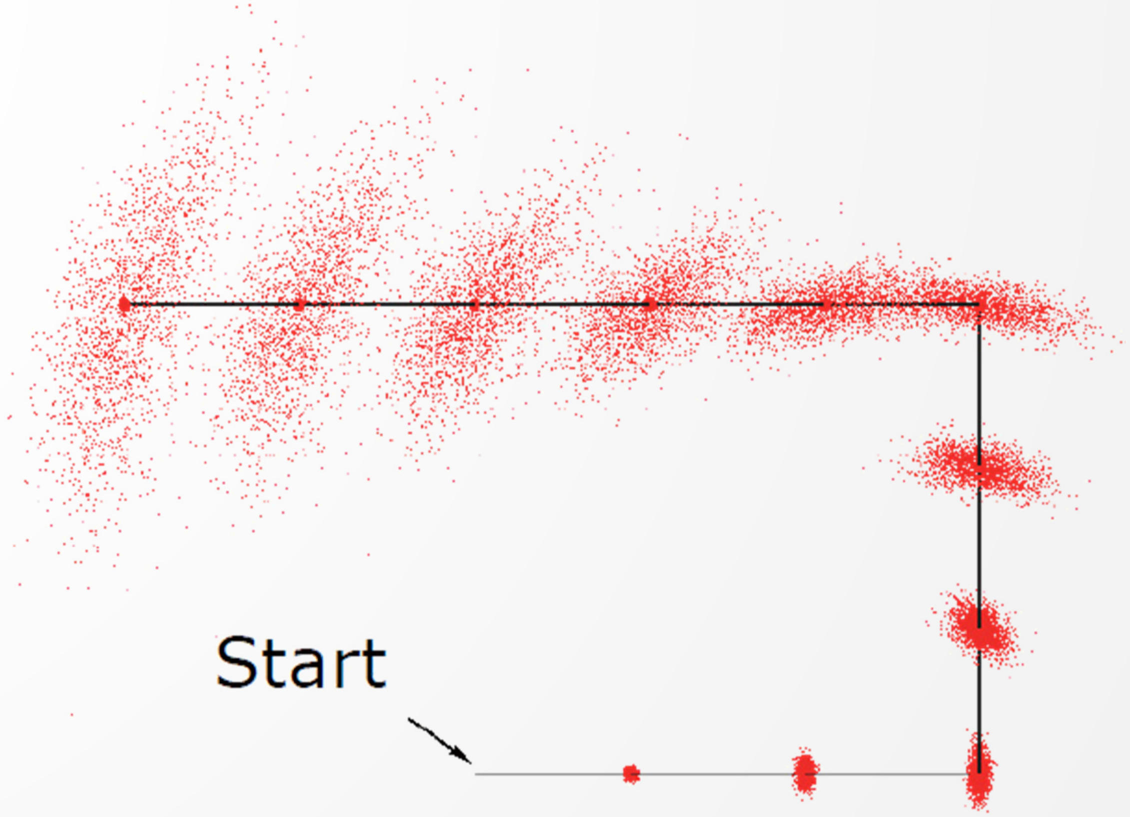
- Monocular VO:
  - Incremental pose “Belief Propagation“
    - Uncertainty **will** increase
    - Odometry **will** drift



- Incremental estimation

- Filter-based approach:

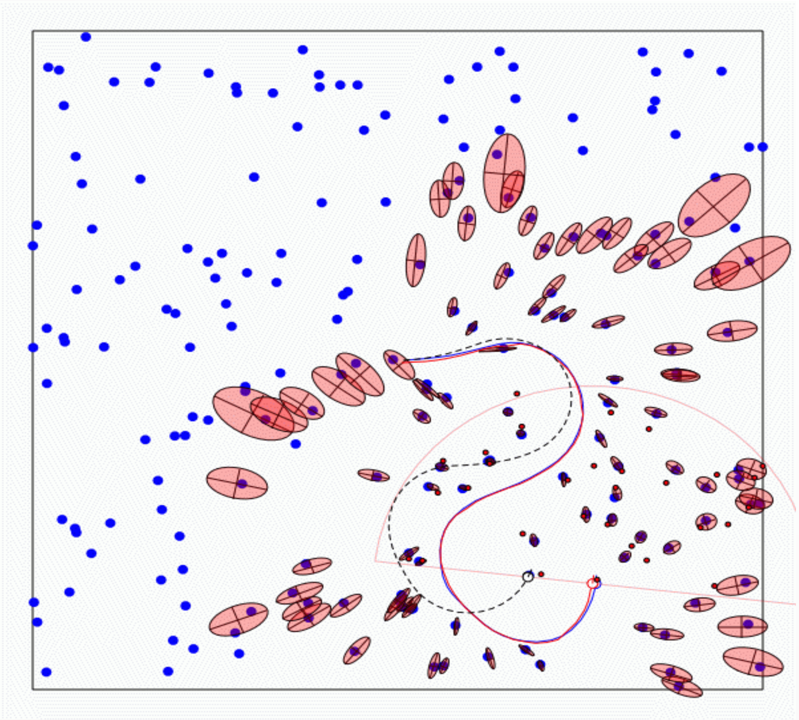




- Motion-model only

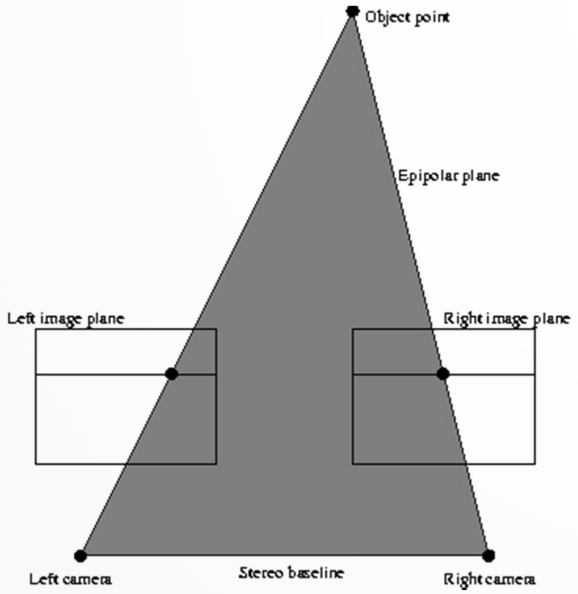
Landmark SLAM

- Monocular VO:
  - A. J. Davison, I. D. Reid, N. D. Molton and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007.
- Extended Kalman Filter
  - Includes **Landmarks** as filter states
  - "**Append**" - Bottleneck becomes map size

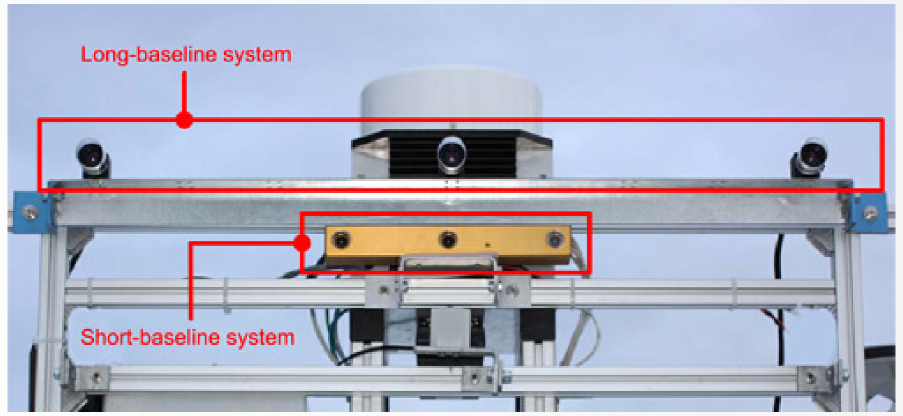


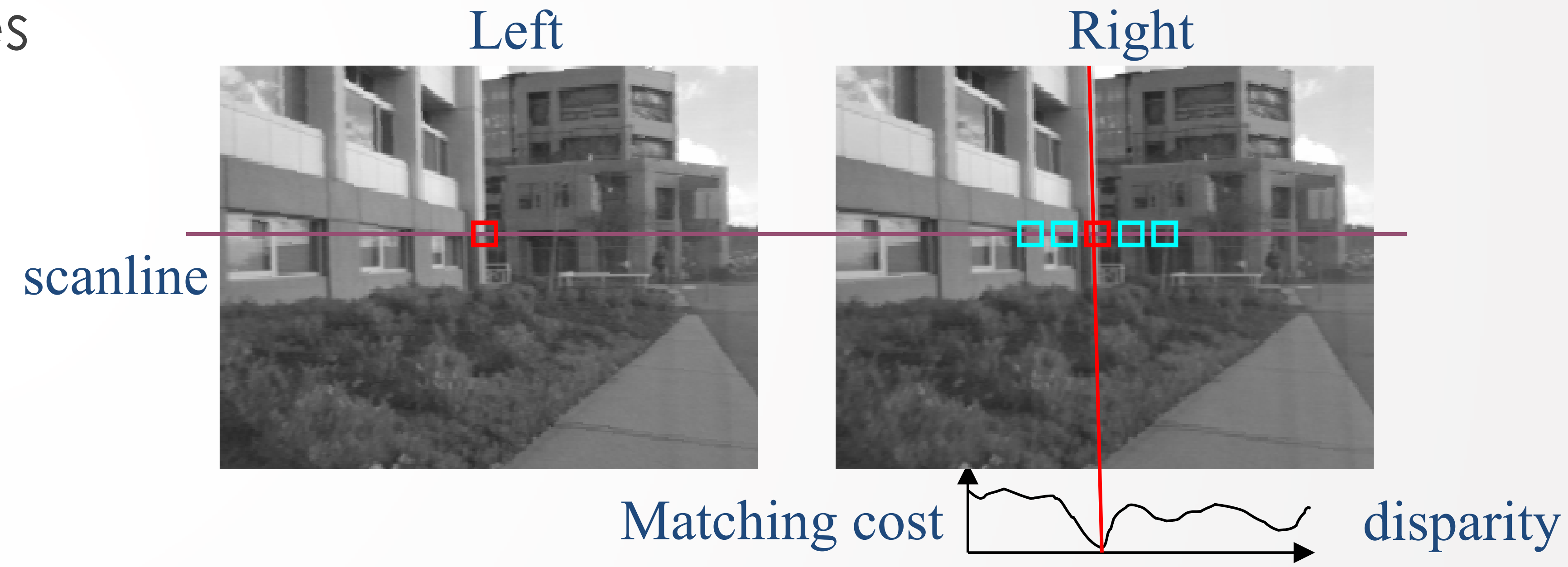
Landmark SLAM

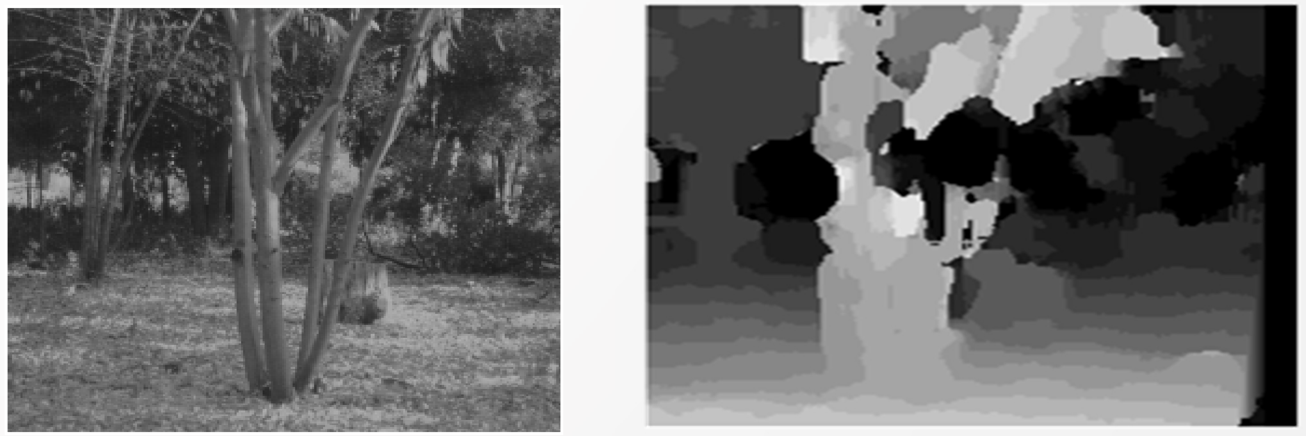
- Stereo VO:
  - Sliding stereo , Binocular stereo
- The known baseline advantages
  - Estimation of absolute scale



  - Estimation of scene depth (Mapping)

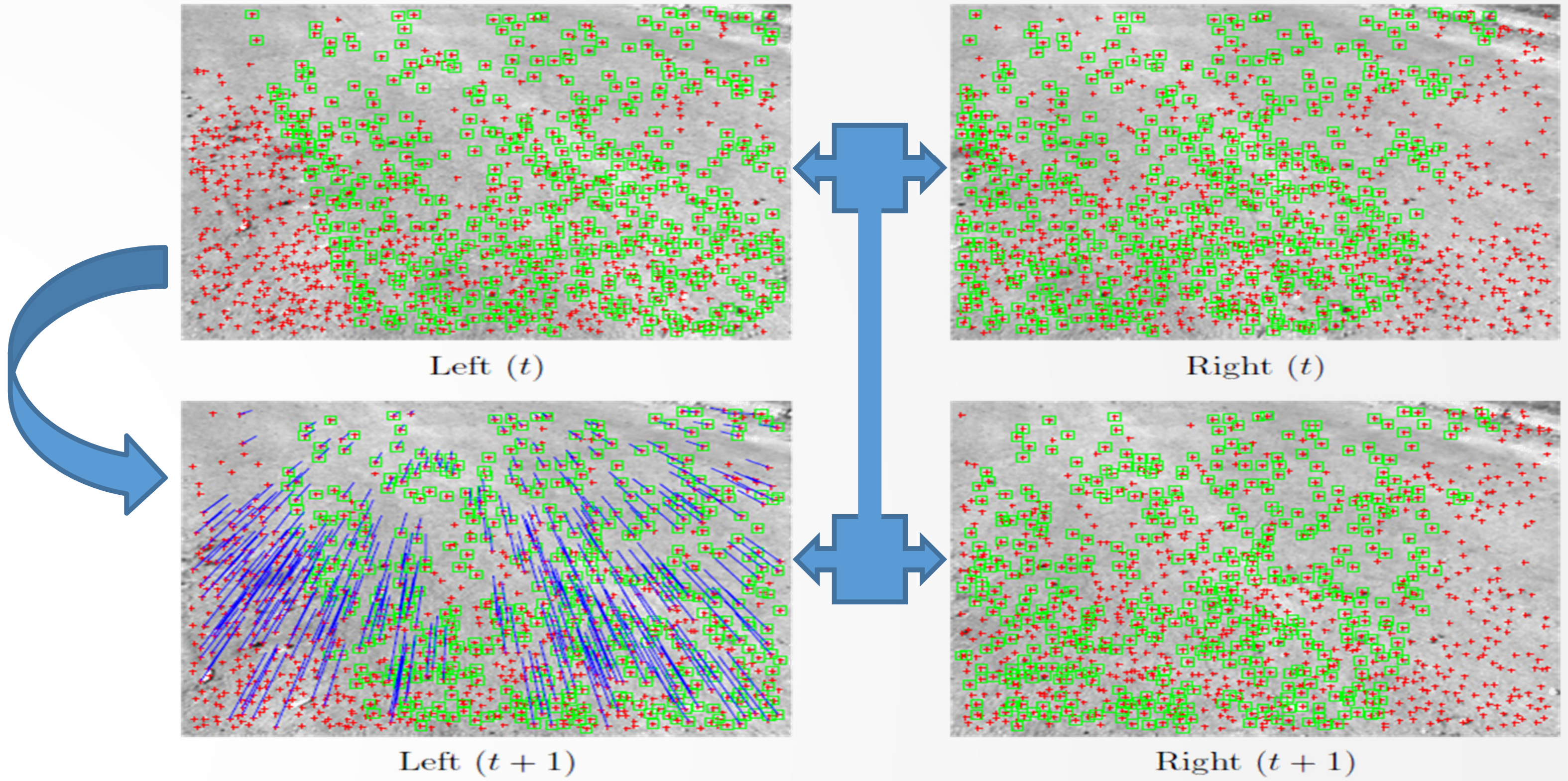


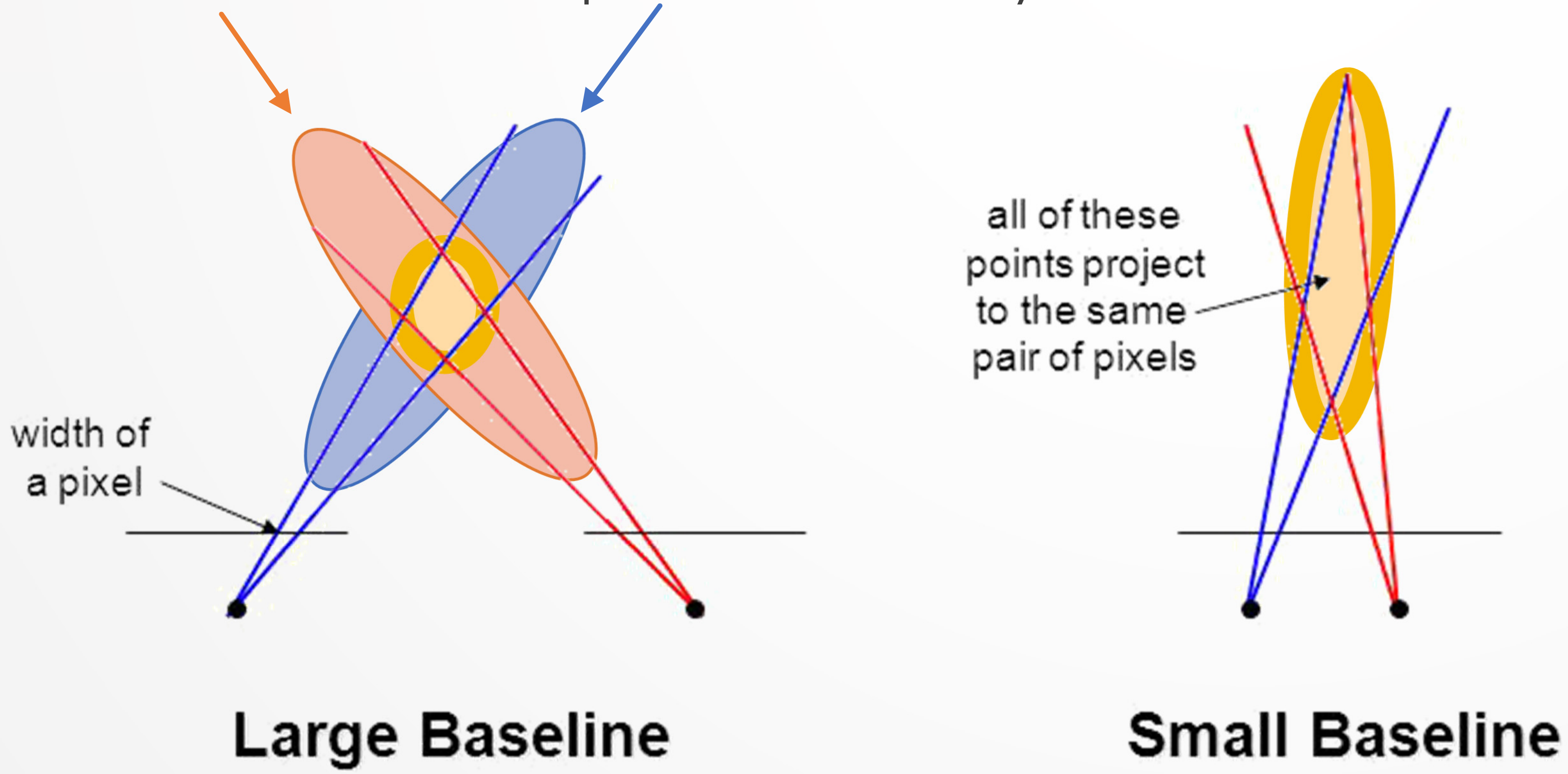


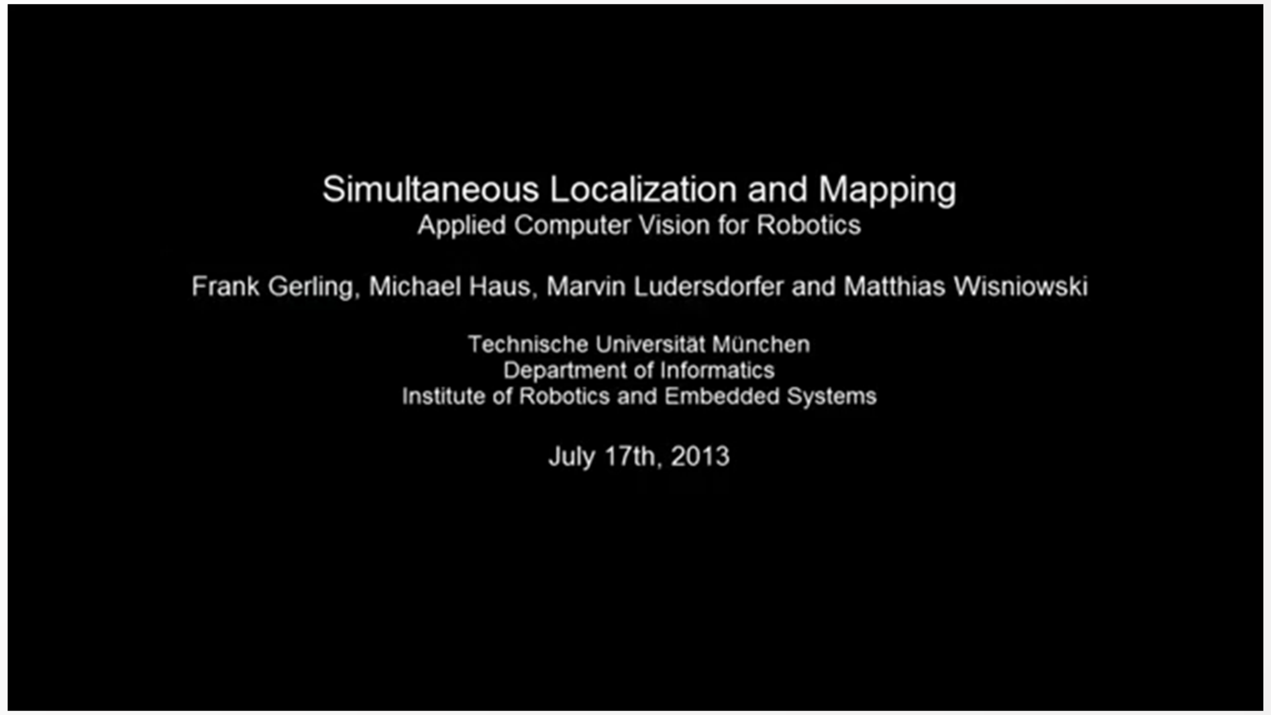


Landmark SLAM

- Stereo VO:
  - A naïve exploitation
- The known baseline advantages
  - Information Filter benefits
- Monocular depth uncertainty

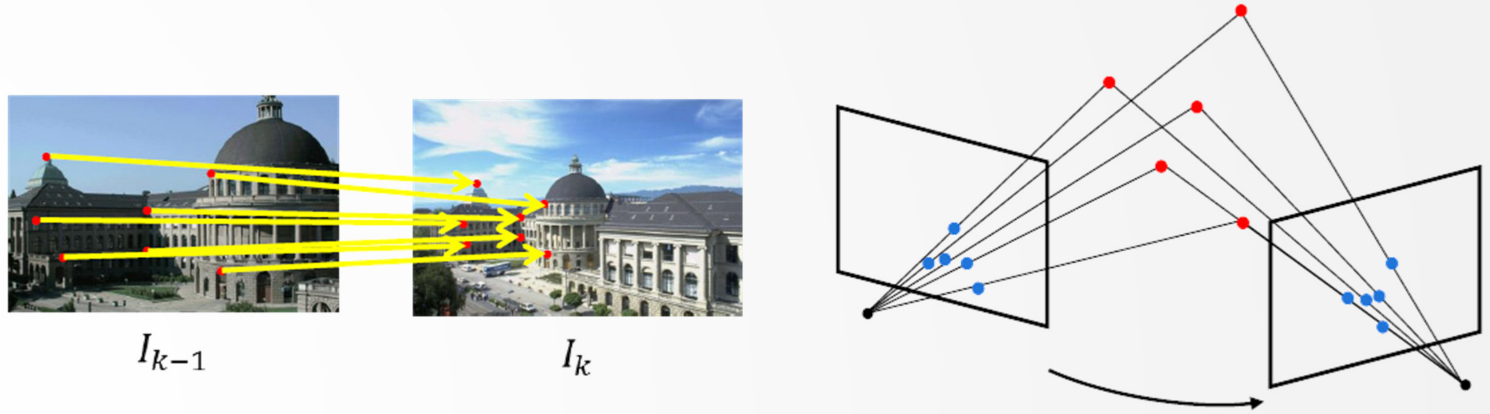




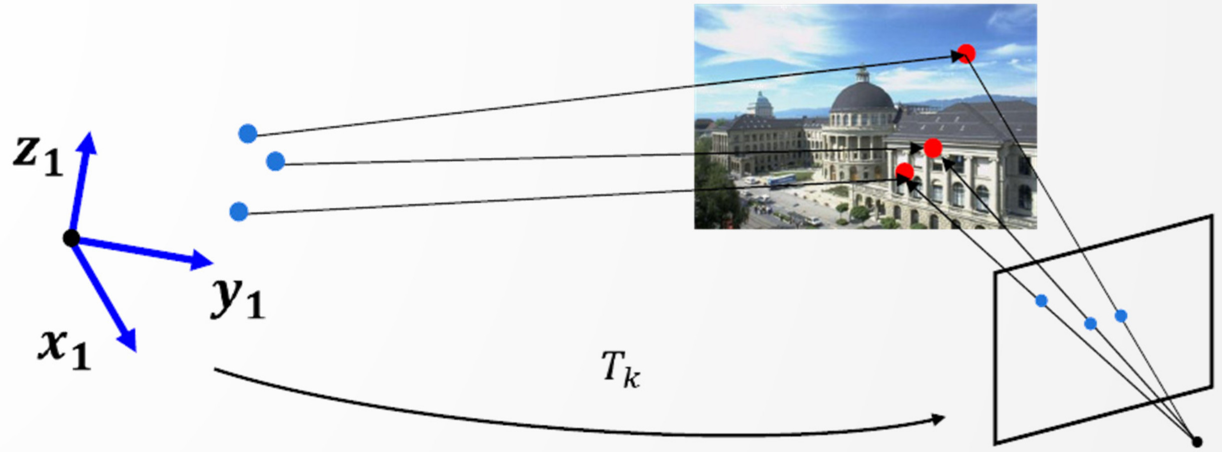


Landmark SLAM

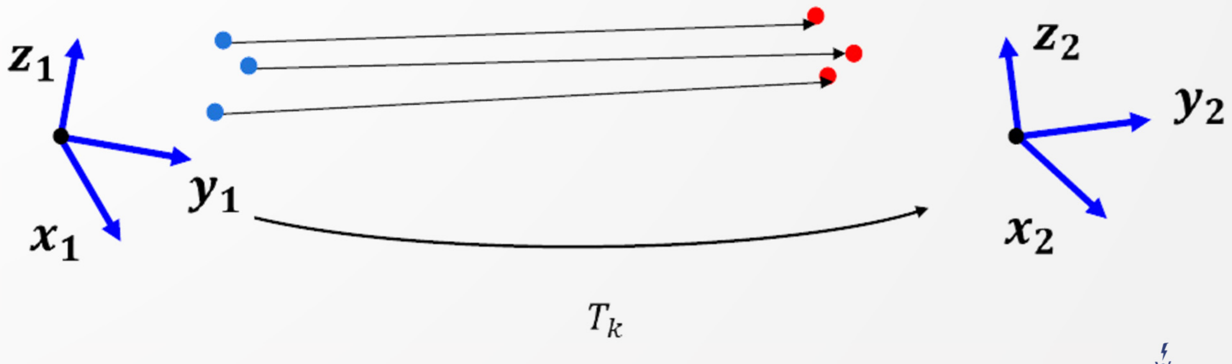
- Correspondences for VO:
- 2D-to-2D
  - Minimally requires (Nister's)5-points
  - (Higgins') 8-point simpler solution – stacking & decomposition suffers



- 3D-to-2D
  - Perspective-n-Points
  - (Gao's) 3-point solution of calibrated camera +1 disambiguates 4 solutions



- 3D-to-3D
  - 3 (non-collinear) correspondences
  - ICP



Landmark SLAM

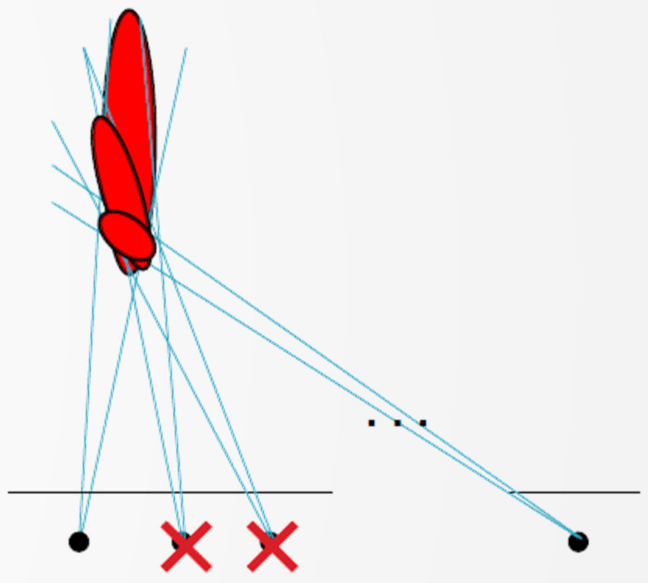
- Correspondences for VO:
  - Generally, Image Reprojection Error minimization is more accurate

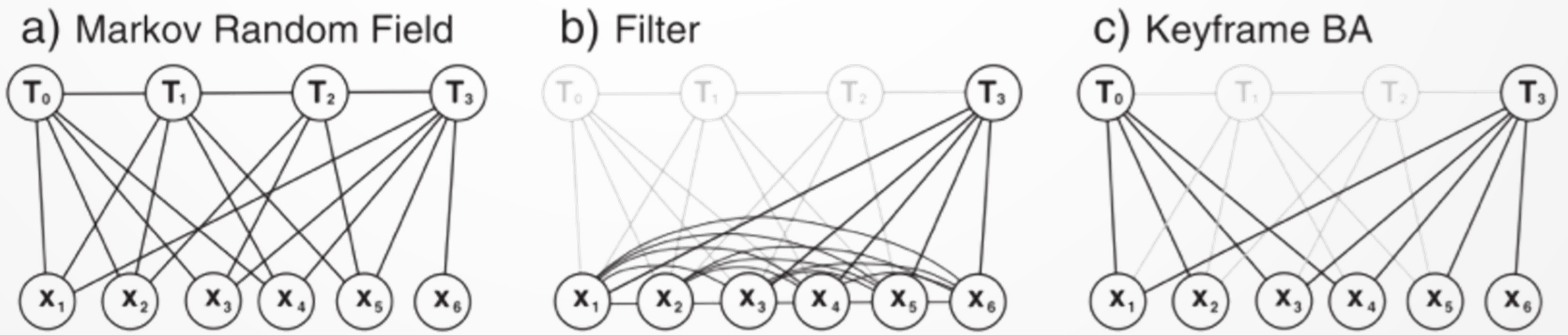
Type of correspondences	Monocular	Stereo
2D-2D	X	X
3D-3D		X
3D-2D	X	X

- Why is R&D still considering monocular?
  - A point at infinity will exhibit no parallax.
  - Stereo VO degenerates to Monocular.
    - What can be done in this case?

Landmark SLAM

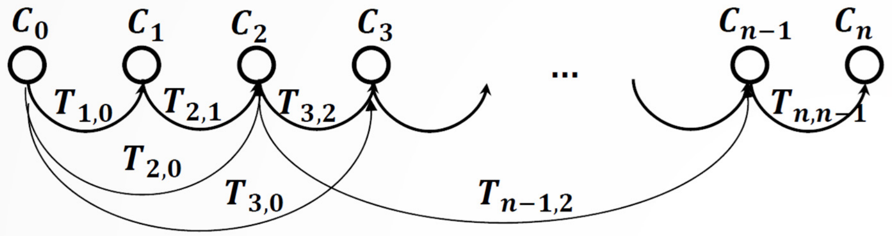
- Visual SLAM:
  - **Goal:** Global, consistent estimate of the robot path.
  - Requires: Optimization of VO pipeline. How?
  - Skip Data - Take Keyframes.
    When? (3D feature uncertainty-driven)
  - Perform Optimization
    When? (Last *m* keyframes)





Landmark SLAM

- Visual SLAM:
  - Pose Graph Optimization.

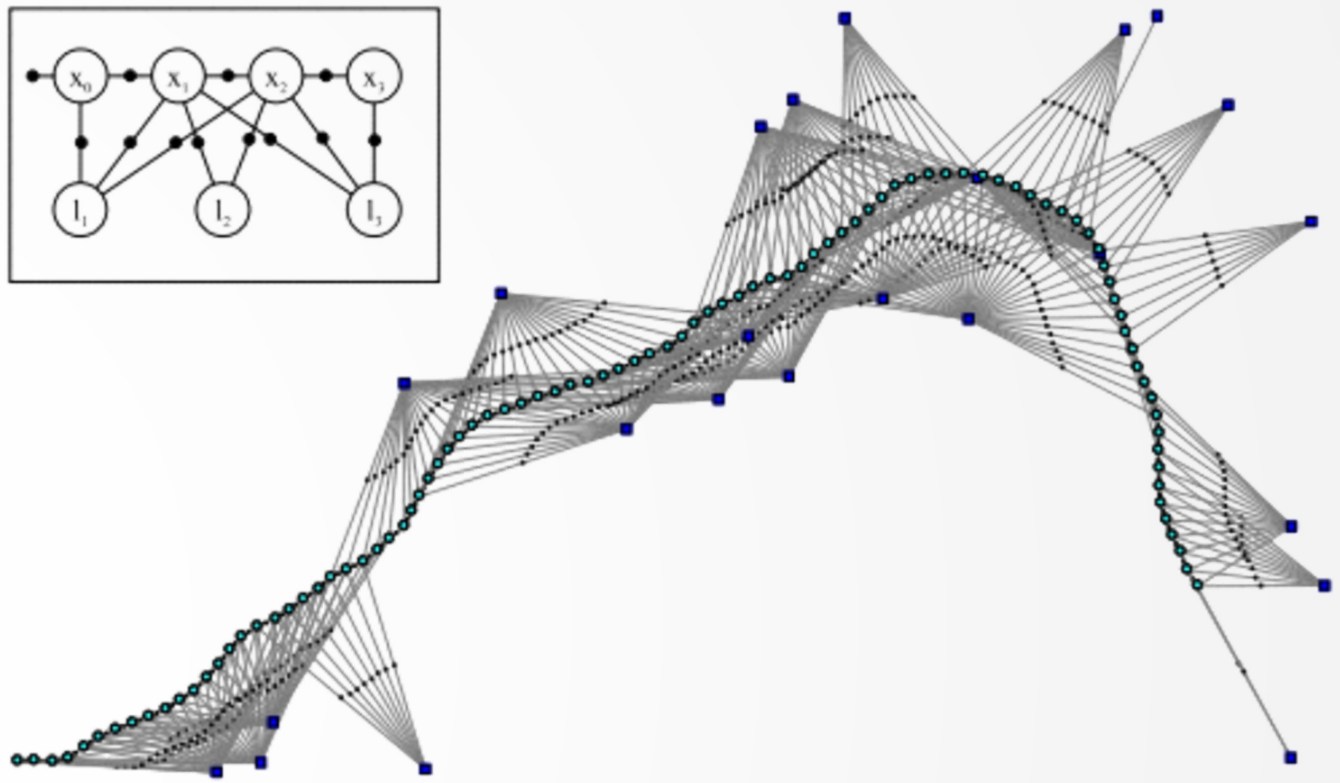


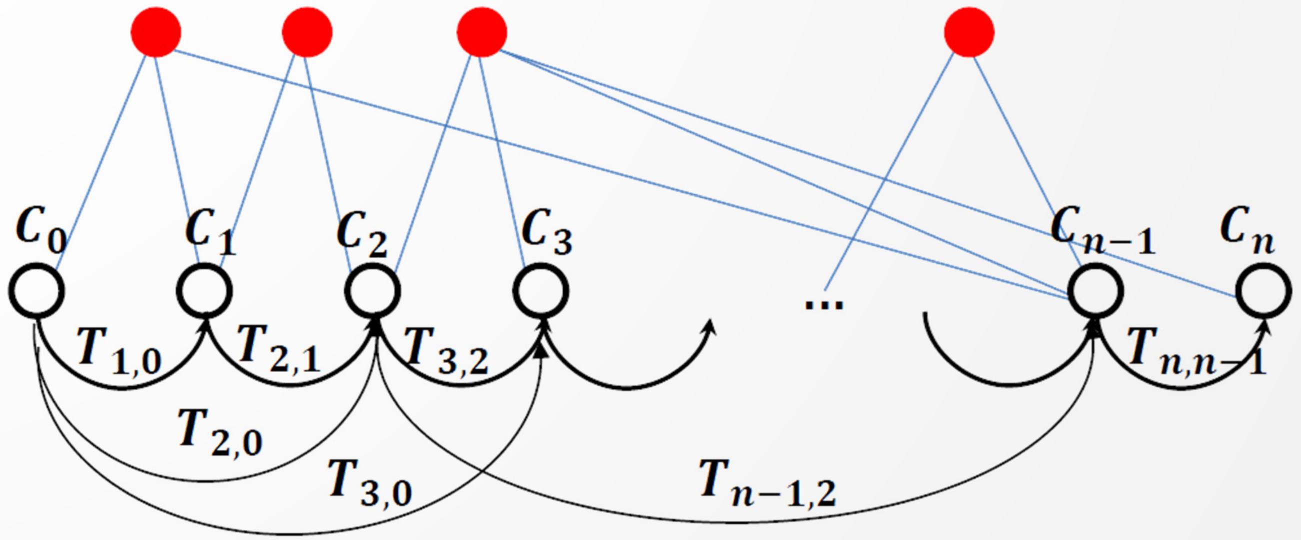
- Non-adjacent frames come into play.
- Gauss-Newton / Levenberg-Marquadt g2o , GTSAM , Ceres

  - Bundle Adjustment

- 3D- Features are considered too.
- Optimization of 3D structure, Camera motion, Camera parameters – Costly.
- Gauss-Newton / Levenberg-Marquadt g2o , GTSAM , Ceres

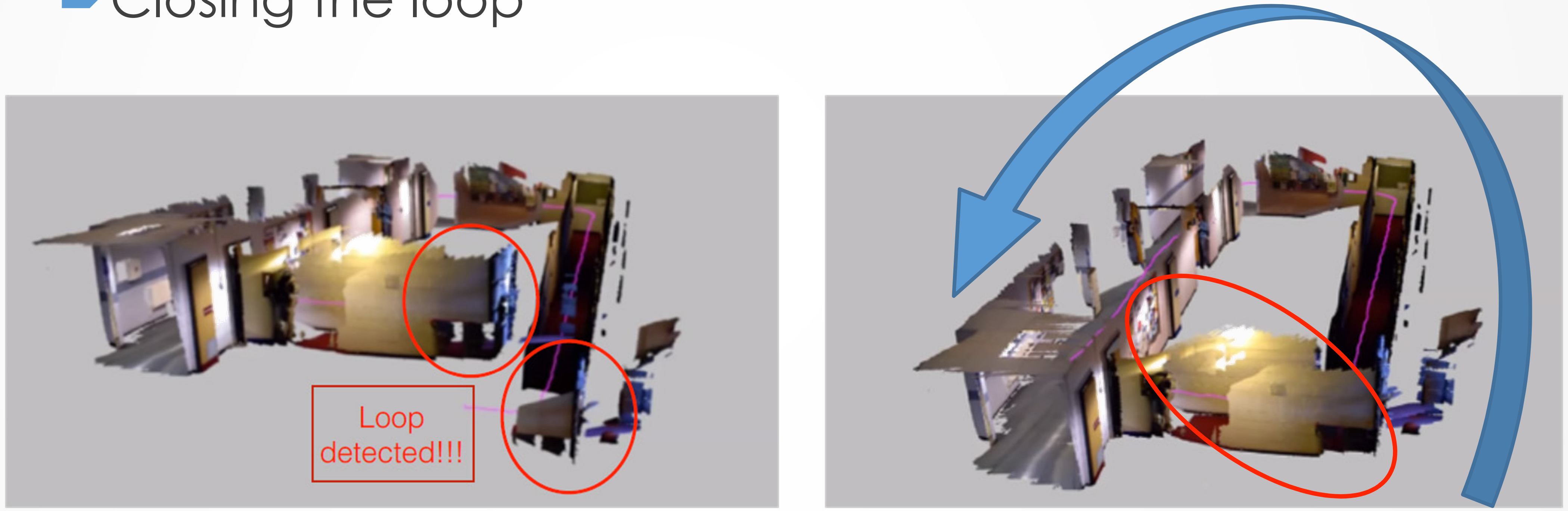
- Keyframe-based pose graph





Landmark SLAM

- Visual SLAM:
  - Global Landmark Correspondence – Global Optimization
    - Closing the loop



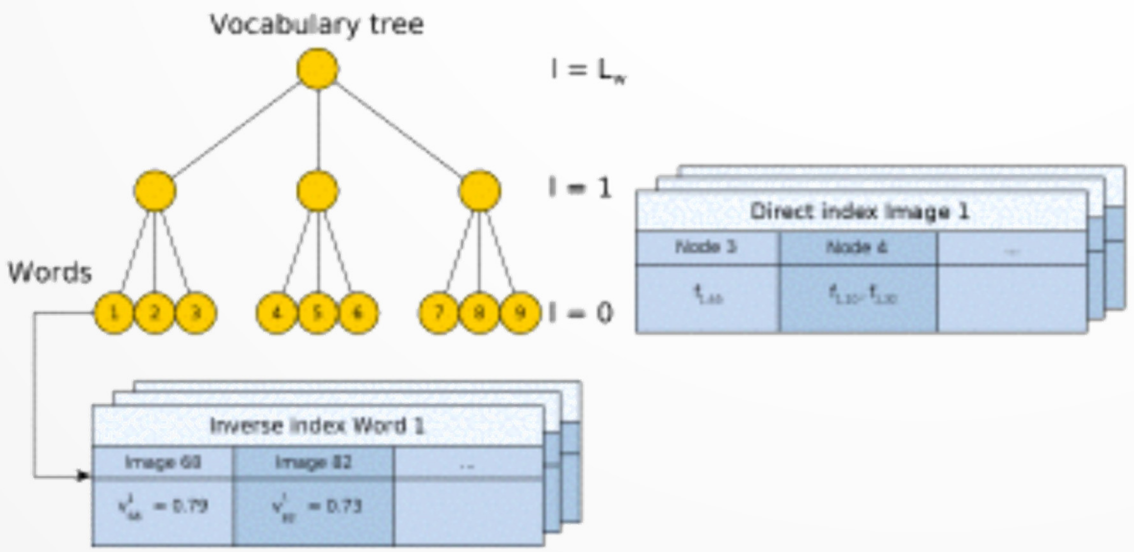
- Global Refinement

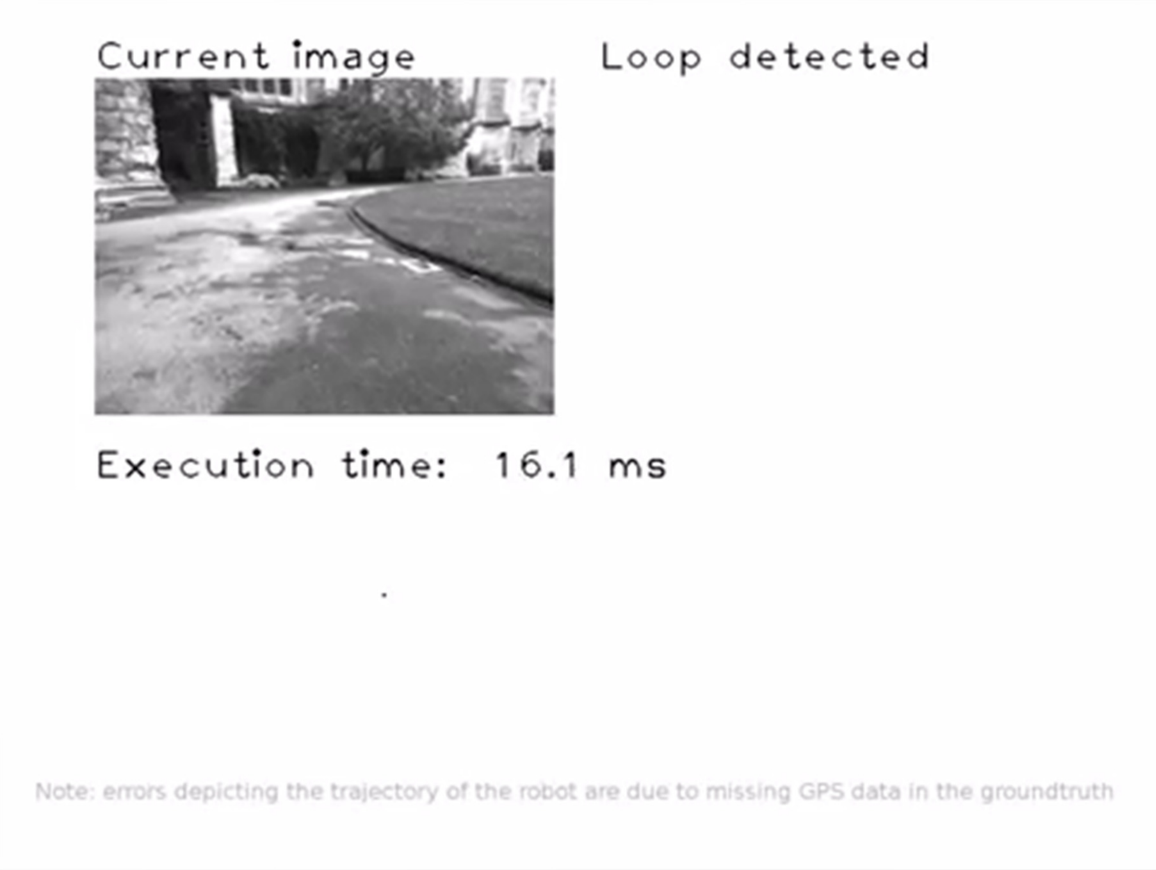
    - Avoid duplication of the map.
    - Compensate for accumulated drift
    - Relocalization capacity.

Landmark SLAM

- Visual SLAM:
  - Place Recognition (vision-based).
    - D. Galvez-López and J. D. Tardos, "Bags of Binary Words for Fast Place Recognition in Image Sequences," IEEE Transactions on Robotics, 2012.

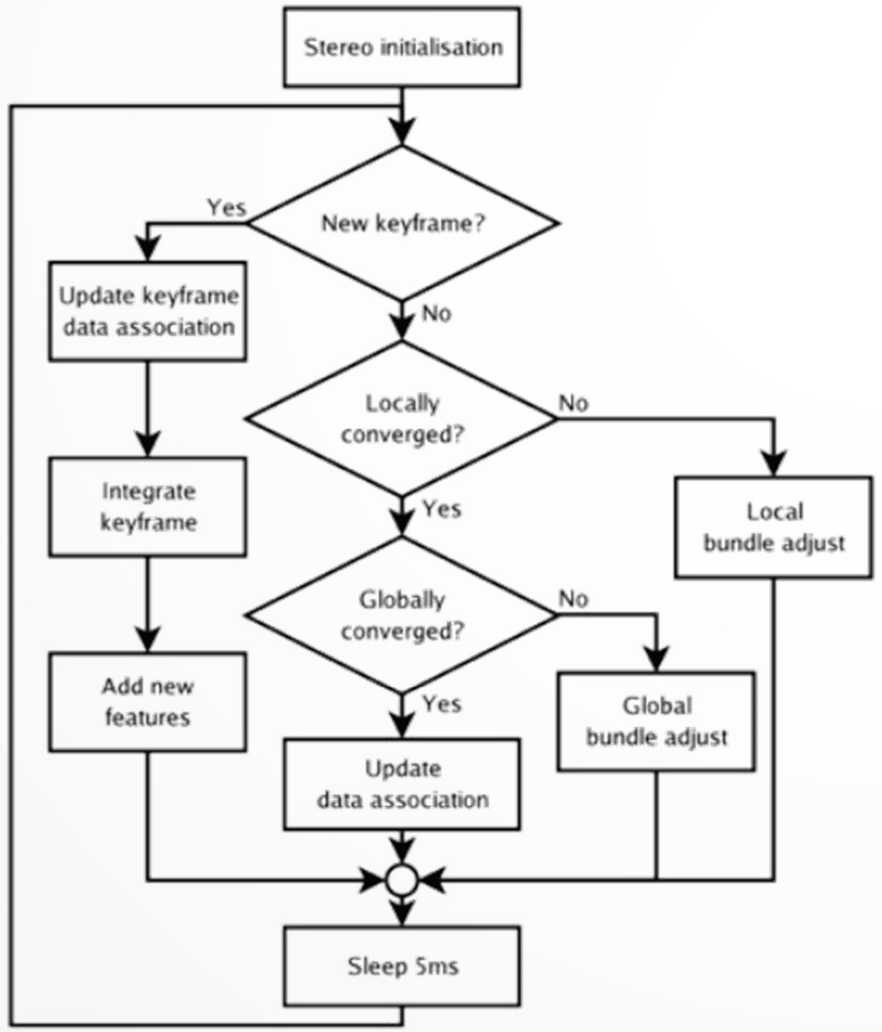
- Find most similar images in query set.
  - Maintain “Visual Word” dictionary tree.
  - Inverted text file logic.





Landmark SLAM

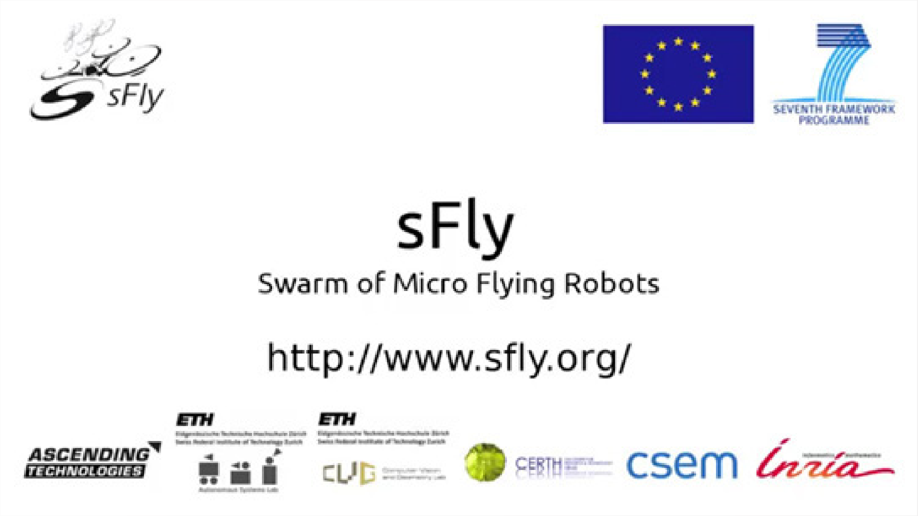
- Sparse Feature-based SLAM:
  - Parallel Tracking And Mapping (PTAM)
    a complete implementation



- G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007.

Landmark SLAM

- Sparse Feature-based SLAM:
  - Parallel Tracking And Mapping (PTAM)
  - Tracking and Mapping done in separate threads.
  - Designed for small workspaces
    - Requires Initialization
    - No-drift, efficient P3P localization with known landmarks
    - BA optimized known landmarks & keyframes
    - Reduced to windowed VO for large environments



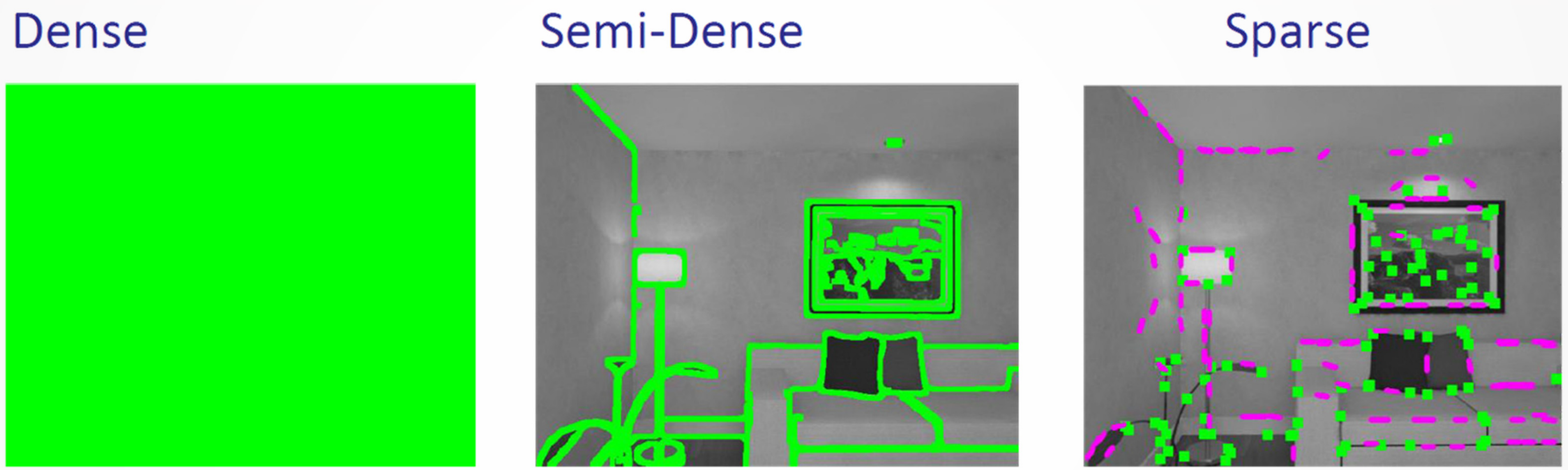
Dense Approaches

Autonomous Multi-Modal Localization and Mapping:

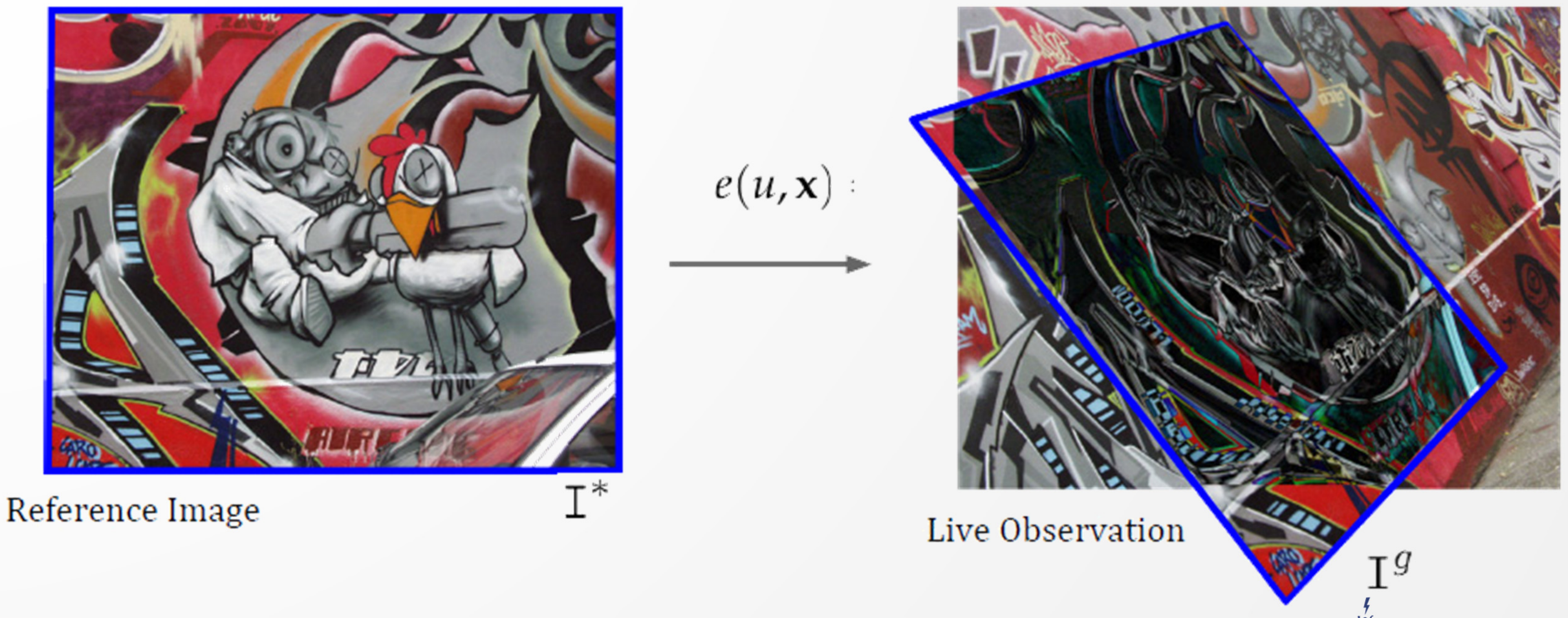
Fundamentals and the State-of-the-Art

Direct Image Alignment

- Appearance-based VO:
  - Per-pixel intensity error minimization.

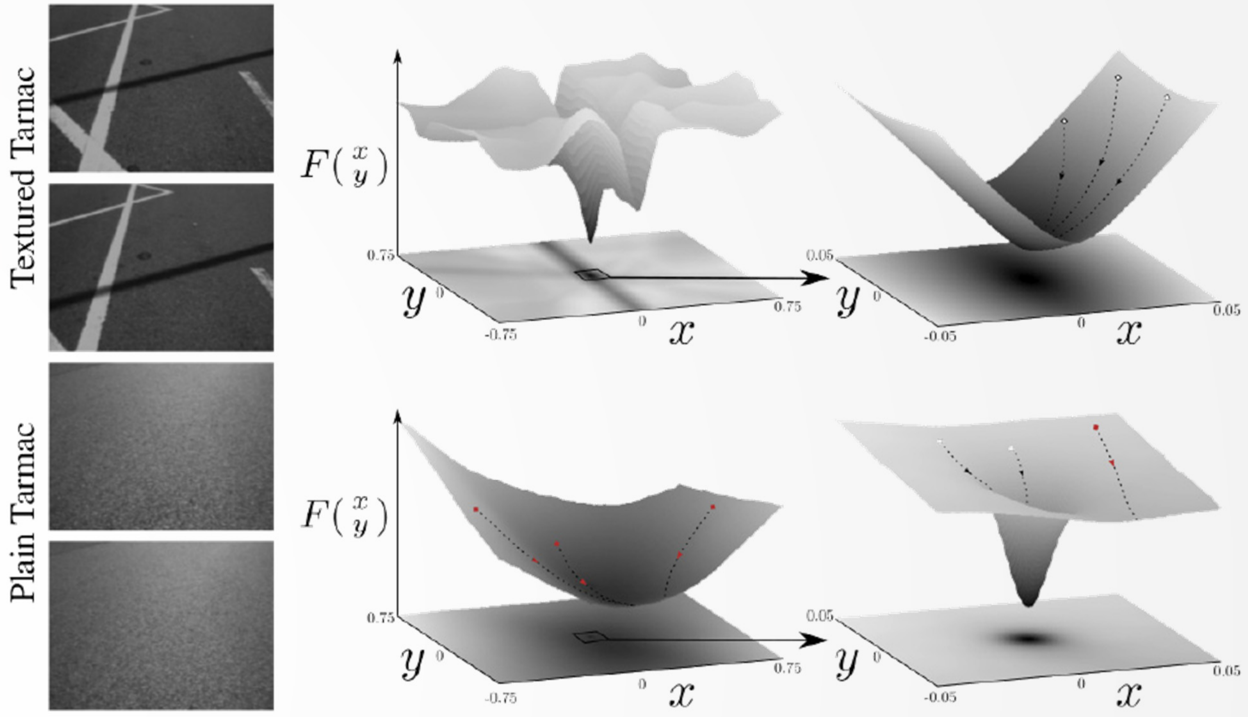


- Dense
  - Per-pixel intensity error minimization.
  - Given a dense, textured surface model, predict what should be seen

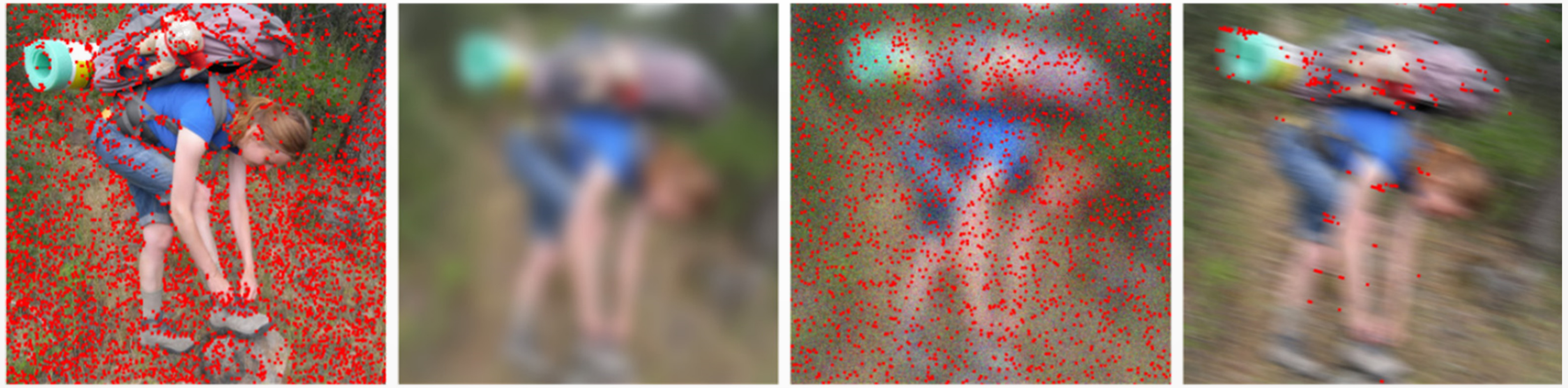


Direct Image Alignment

- Dense VO:
  - Whole image alignment.
  - Minimization of Photometric Error cost function:



- Why bother?
  - Sparse pipeline needs image features



Reference Image

Geometric transformation and blur

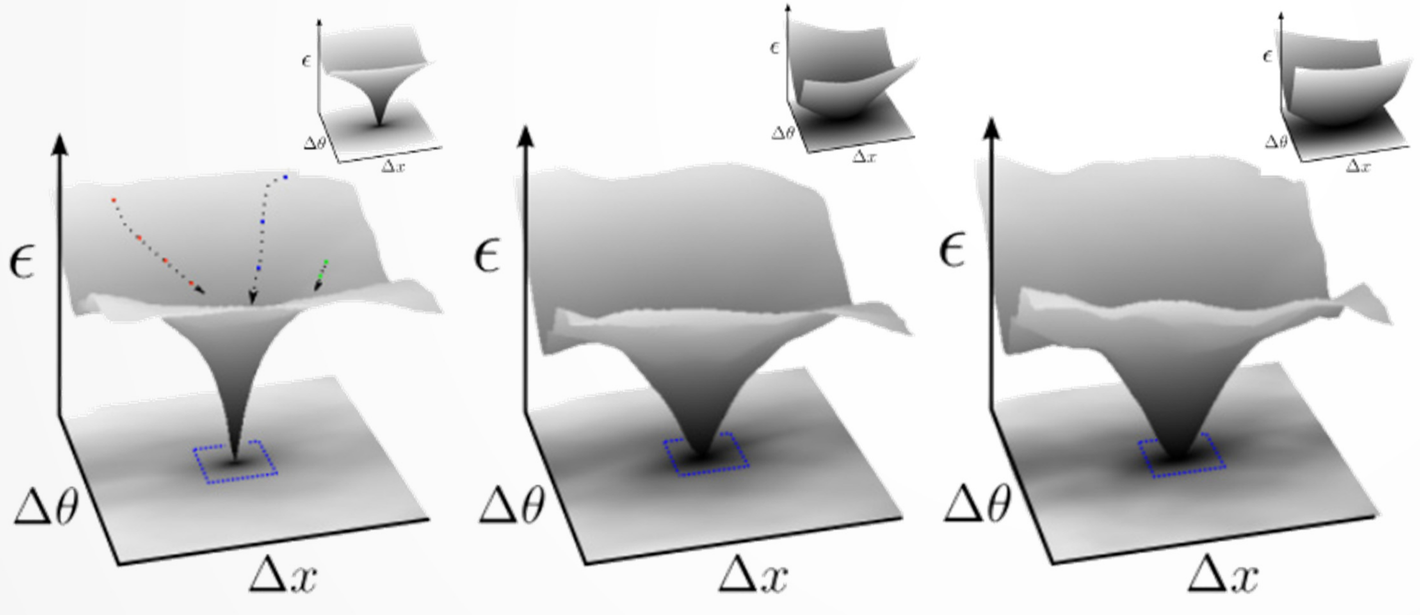
Geometric, blur and noise

Geometric, motion blur

- Example FAST detections in degradation

Direct Image Alignment

- Dense VO:
  - Clear global minimum despite single-pixel error term.
  - Local minima!



Geometric only

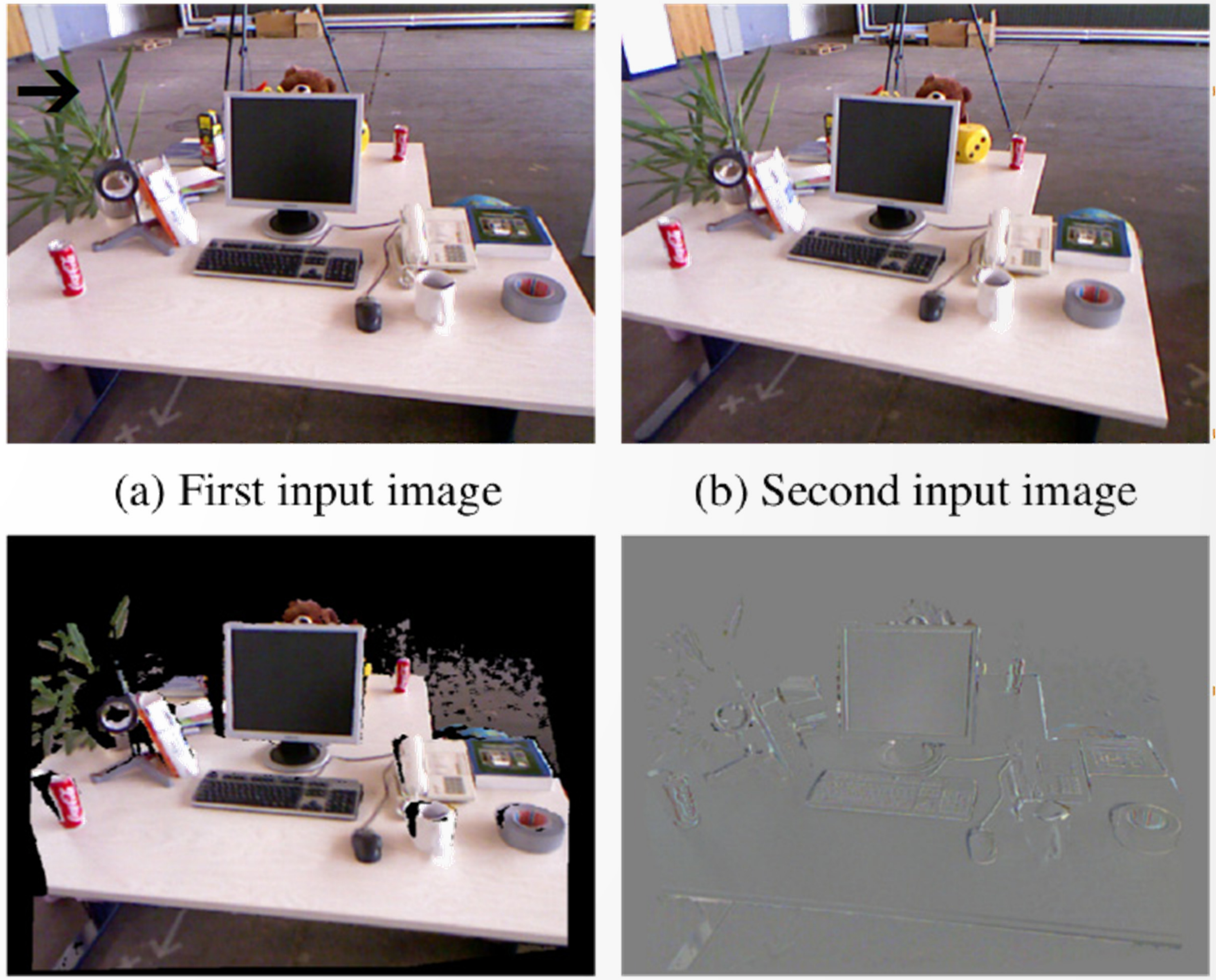
Geometric, blur, noise

Geometric, blur, noise, occlusion

- Whole image alignment: Redundancy for few estimated parameter. Robustness.
- Gradient Descent for cost function requires initialization near global minimum.
- Errors & Derivatives Optimization: Gauss-Newton.
- Paradox: Given trivially parallelizable nature, framerate increase reduces requirements.

Direct Image Alignment

- Dense VO:
  - Tools:
  - GPU
  - Rendering engines (OpenGL)
  - Whole image: RGB ++ (**D**)
  - Dense Pixel Transfer Assets:
  - Mesh surface representation. Can predict self-occlusion.



(a) First input image (b) Second input image

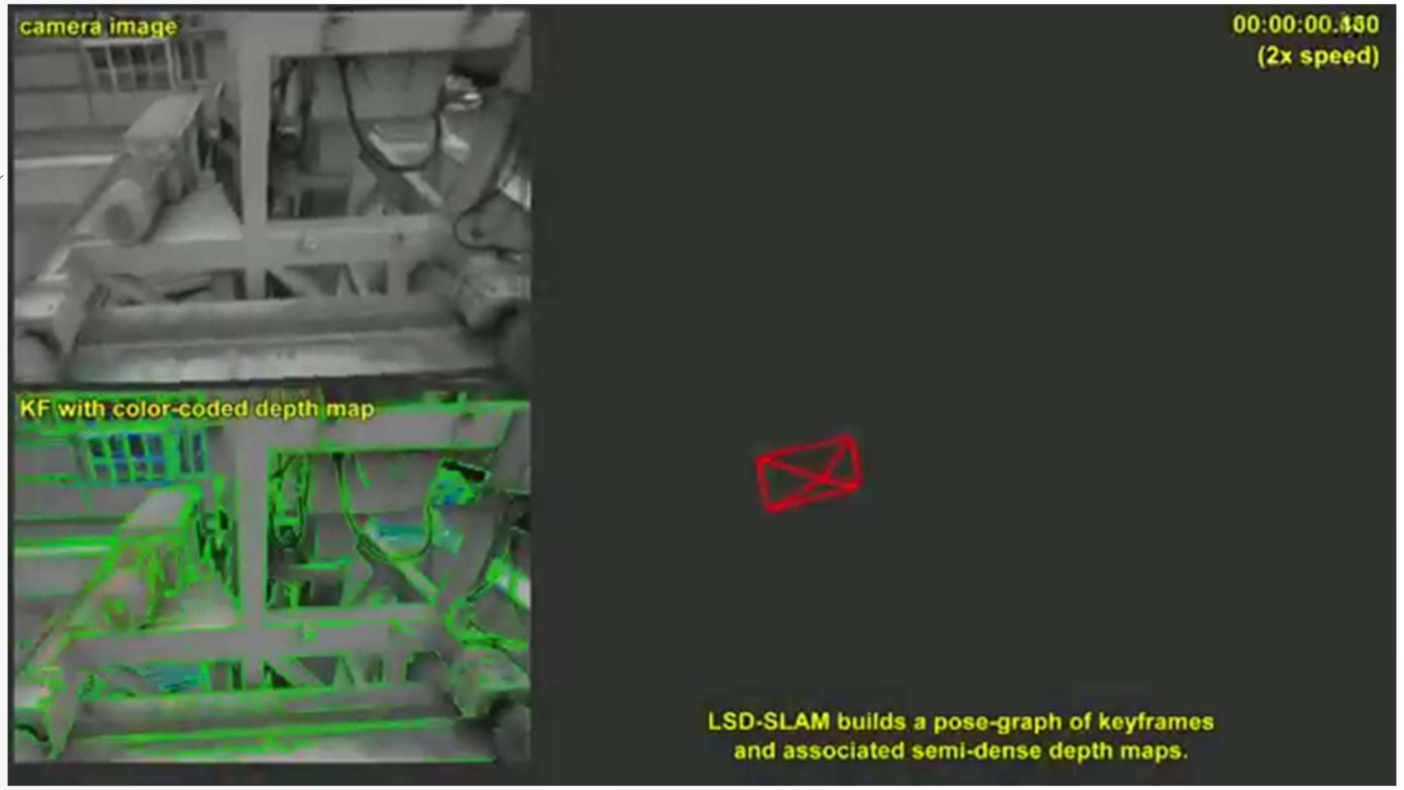
(c) Warped second image (d) Difference image

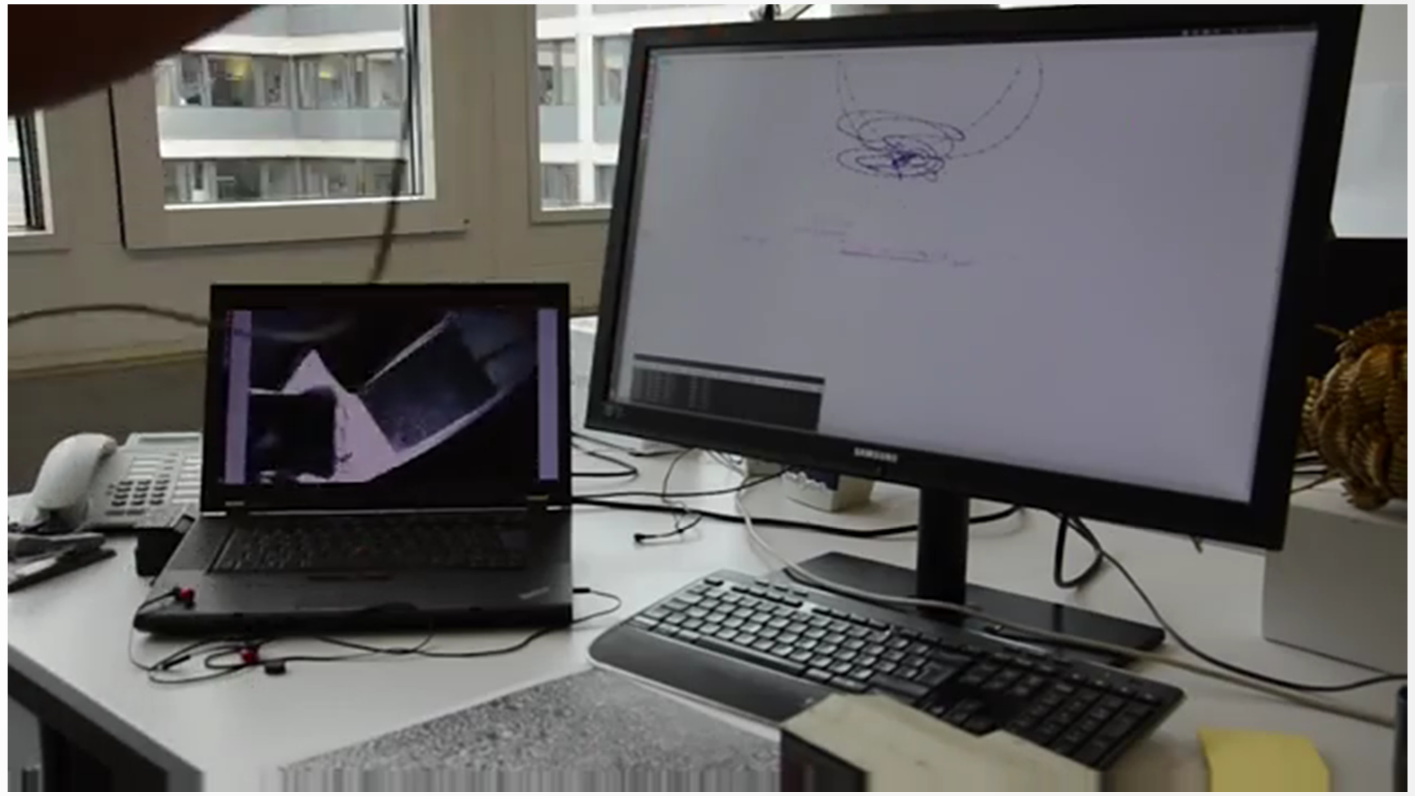
Direct Image Alignment

- Dense VO:
  - R. A. Newcombe, S. J. Lovegrove, A. J. Davison, "DTAM: Dense tracking and mapping in real-time," International Conference on Computer Vision, 2011
  - Dense Tracking & Mapping (DTAM)
    - 3D reconstruction
    - Robustness

Direct Image Alignment

- Semi-Dense:
  - Jakob Engel, Thomas Schöps, Daniel Cremers, "LSD-SLAM: Large-Scale Direct Monocular SLAM," European Conference on Computer Vision, 2014.

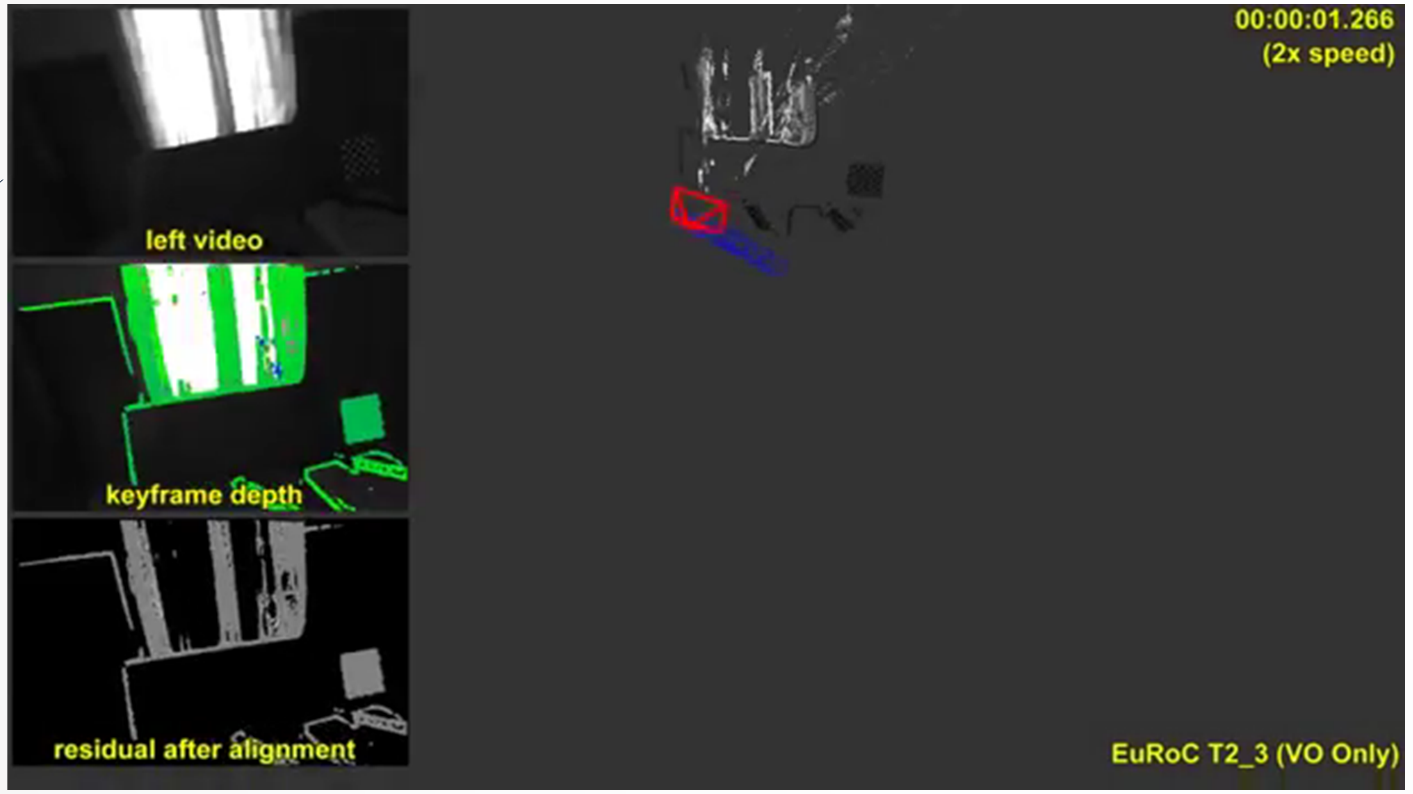


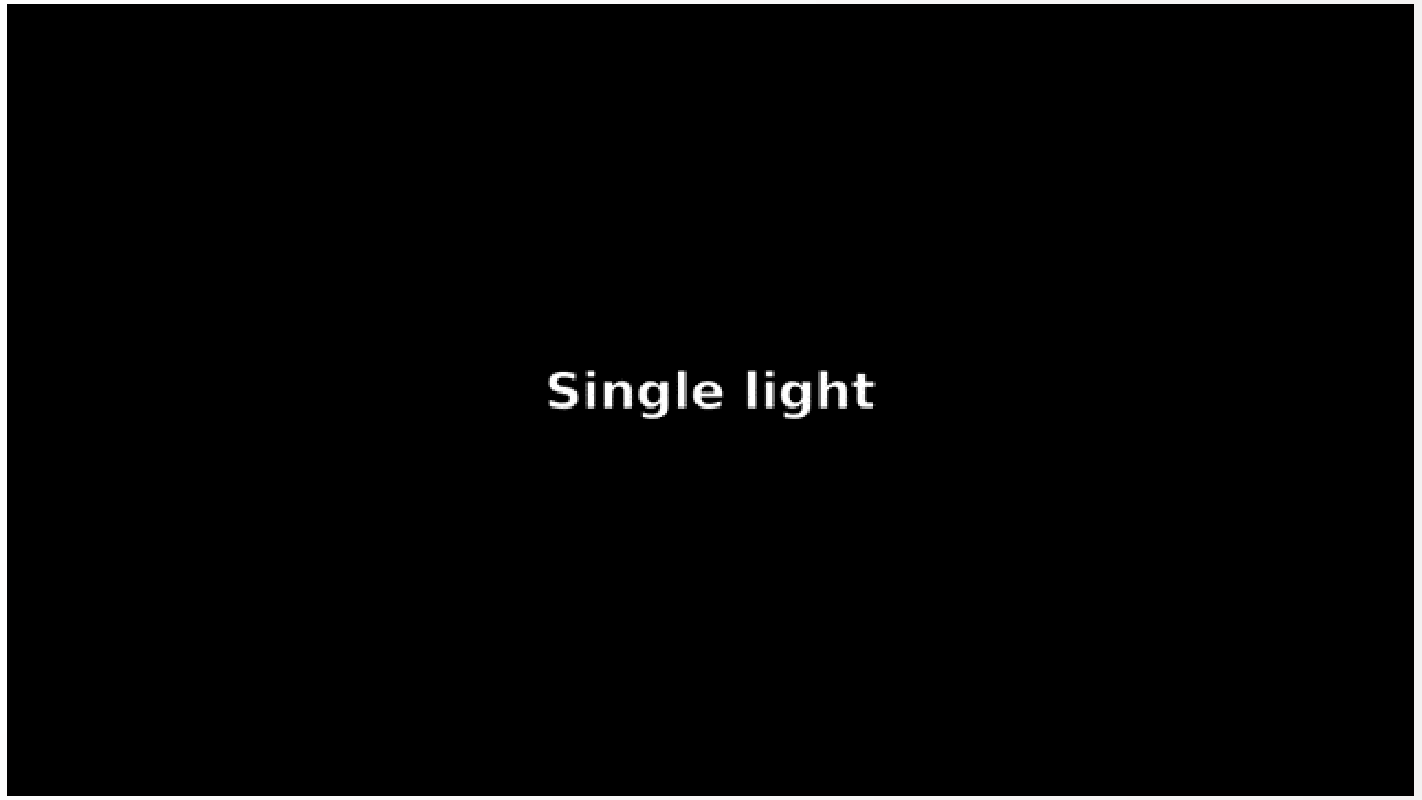


- Sparse:
  - C. Forster, M. Pizzoli and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," IEEE International Conference on Robotics and Automation (ICRA), 2014.

Direct Image Alignment

- Extras:
  - **Stereo (Semi)-Direct**: J. Engel, J. Stückler and D. Cremers, "Large-scale direct SLAM with stereo cameras,“ IROS 2015.





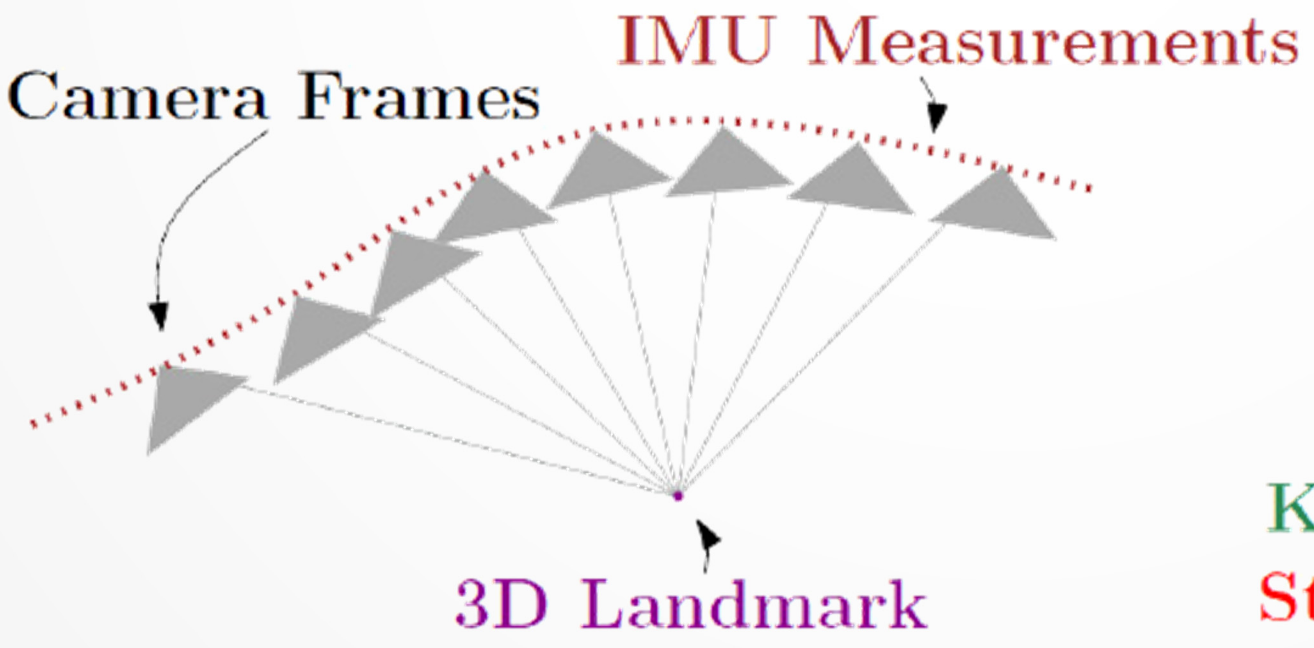
  - **Light Source Detection**: T. Whelan, S. Leutenegger, R. F. Salas-Moreno, B. Glocker and A. J. Davison, “ElasticFusion: Dense SLAM Without A Pose Graph,” RSS 2015.

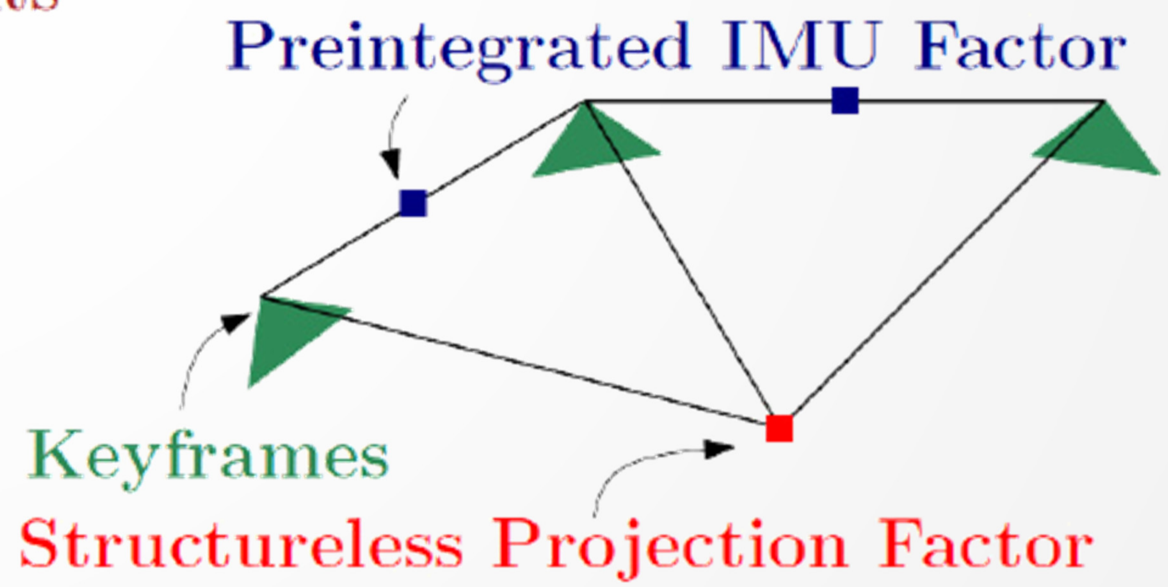
Multi-Modal Approaches

Autonomous Multi-Modal Localization and Mapping: Fundamentals and the State-of-the-Art

Visual-Inertial Fusion

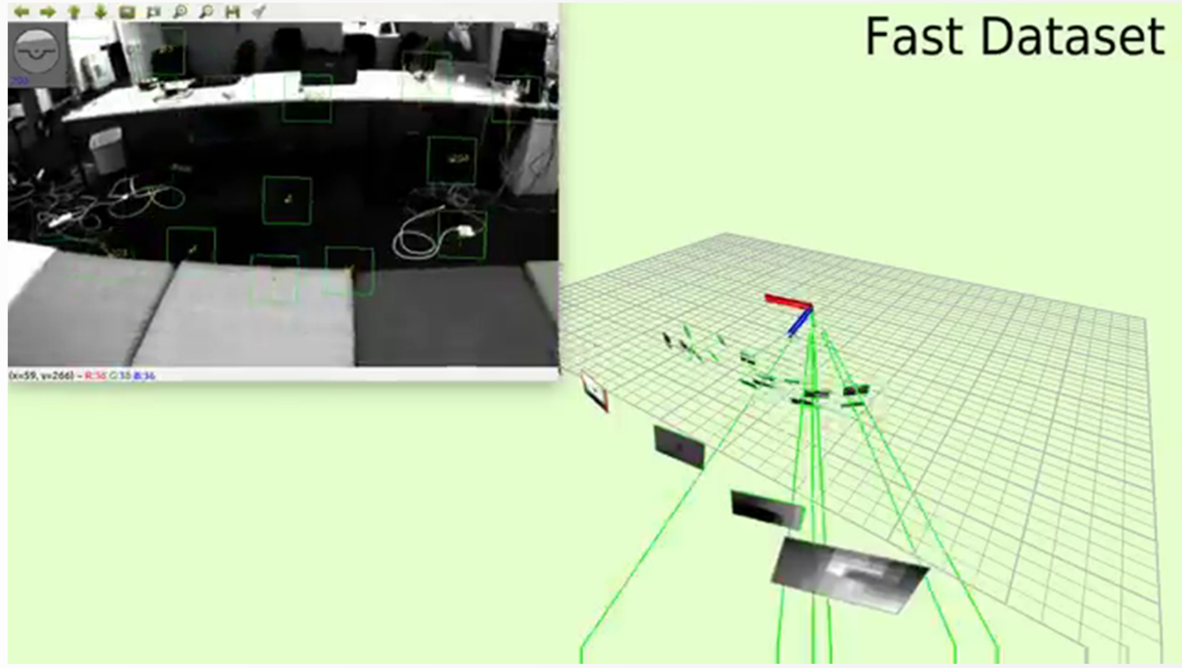
- Monocular Vision (issues)
  - Absolute pose is known up to a scale
  - Inertial Measurement Unit (IMU) provides accelerations.
    - Velocity, scale recoverable from 1 feature, 3 observations.
    - Better-than constant velocity model in propagation.

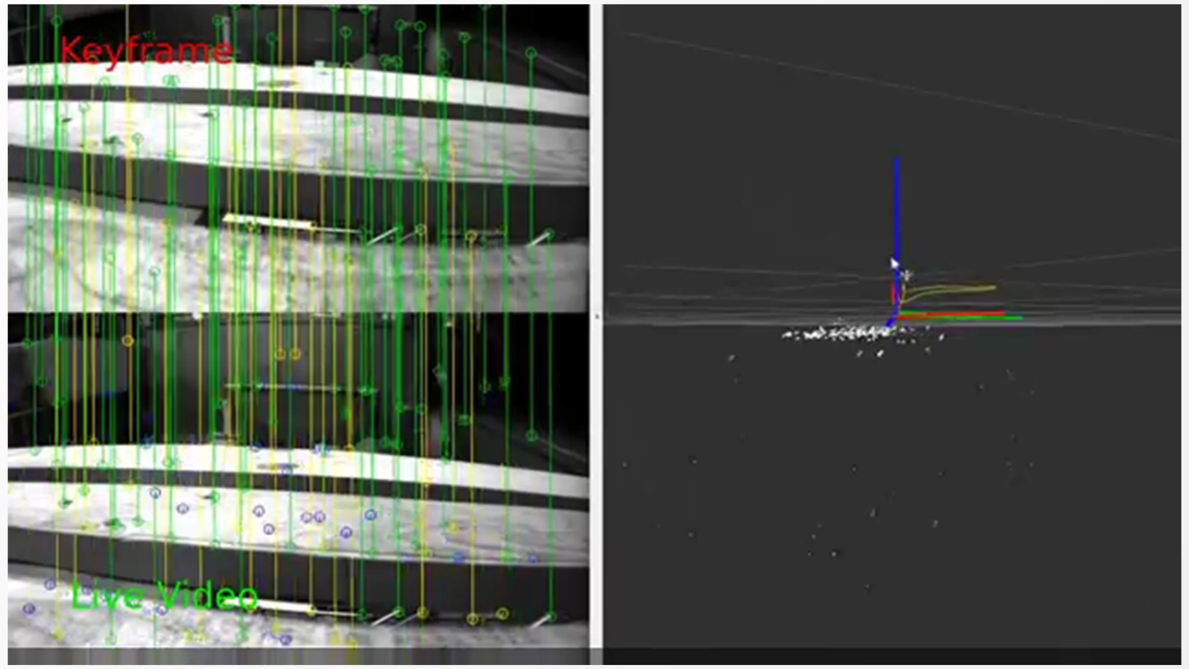




Visual-Inertial Fusion

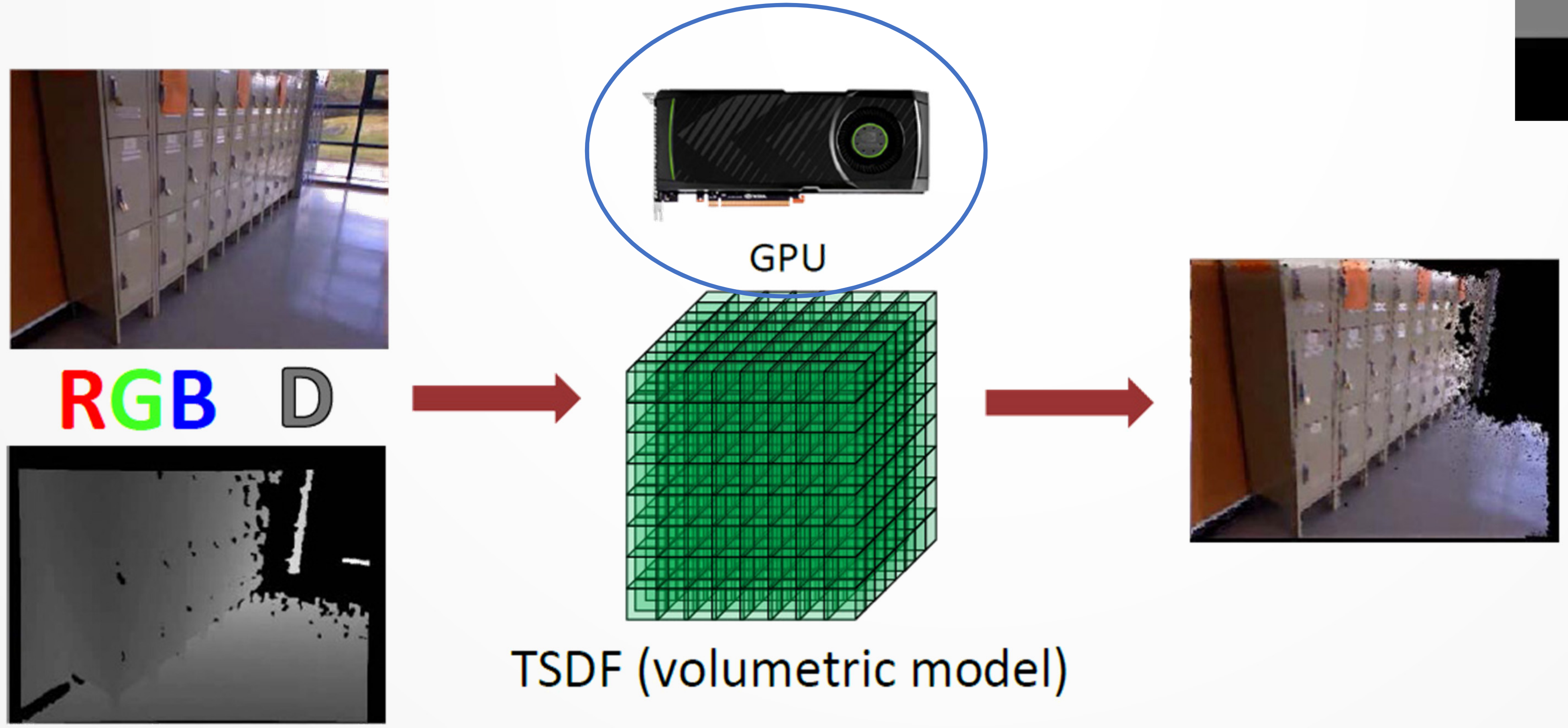
- Visual-Inertial Odometry:
  - Filter-based.
  - Al Mourikis, SI Roumeliotis, “A multi-state constraint Kalman filter for vision-aided inertial navigation,” ICRA, 2007.
  - M. Bloesch, S. Omari, M. Hutter and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," IROS, 2015.
  - Non-linear Optimization-based
  - Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart and Paul Timothy Furgale, “Keyframe-based visual–inertial odometry using nonlinear optimization”, IJRR, 2015.



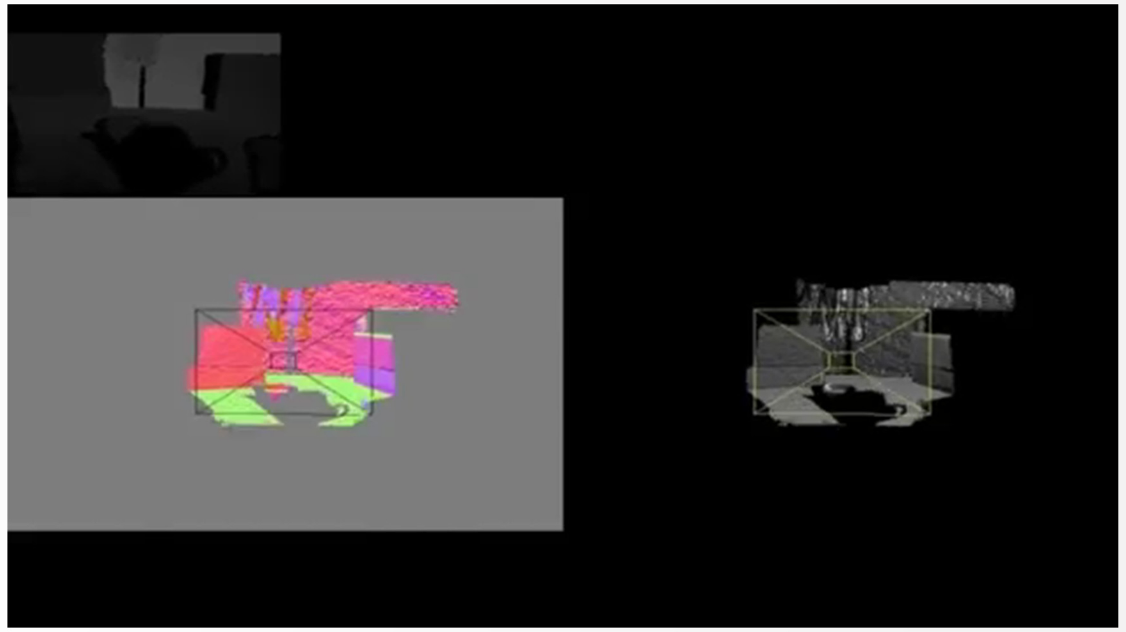


Depth – Time-of-Flight

- Depth sensors – ICP:
- How much space fits into the volume?
  - Depends on resolution:
    2GB GPU: 512x512x512 voxels
    5mm/voxel: 2.5m side length

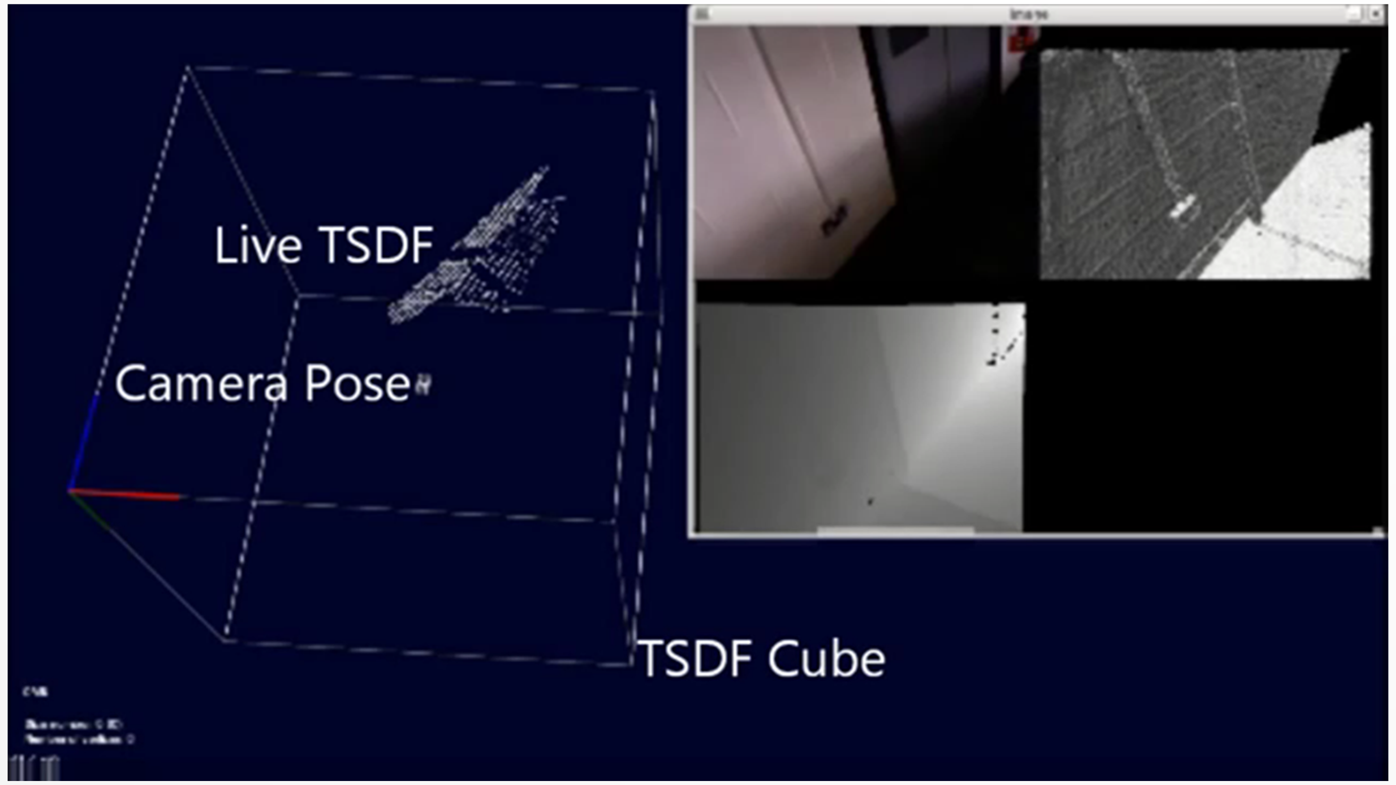


- KinectFusion

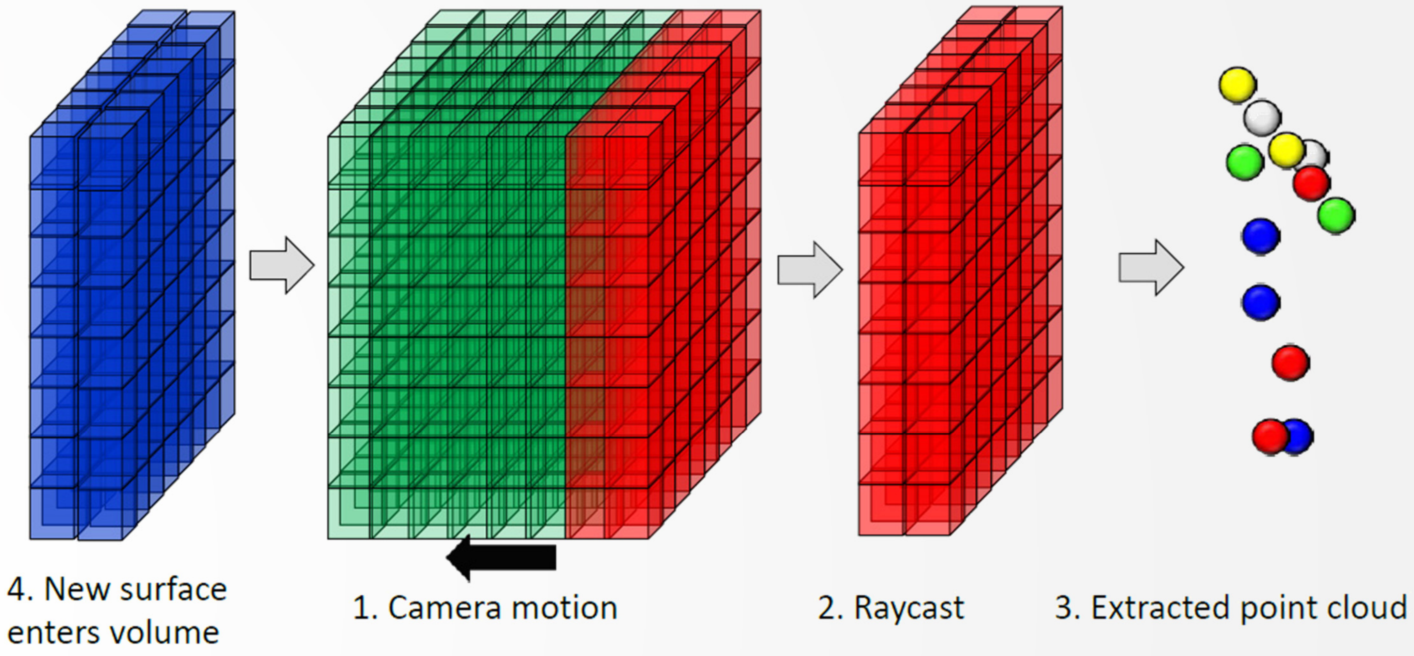


Depth – Time-of-Flight

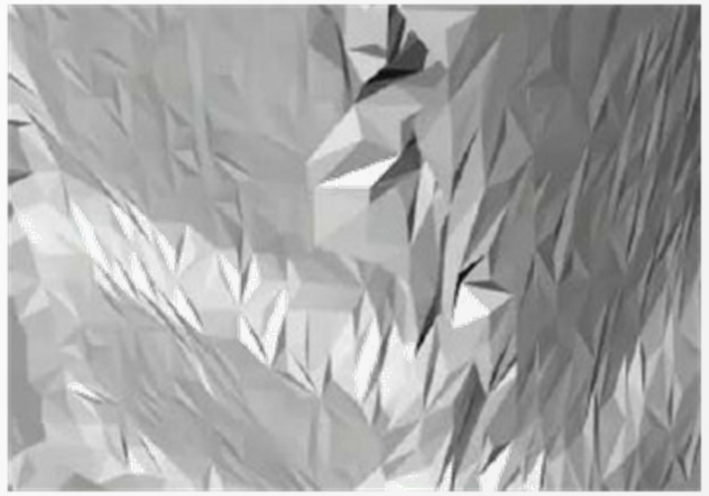
- Depth sensors – ICP:
- For arbitrarily large exploration volumes, treat TSDF as circular buffer.
- Whelan, Thomas; Kaess, Michael; Fallon, Maurice; Johannsson, Hordur; Leonard, John; McDonald, John, CSAIL 2012.



- Kintinuous: Spatially Extended KinectFusion



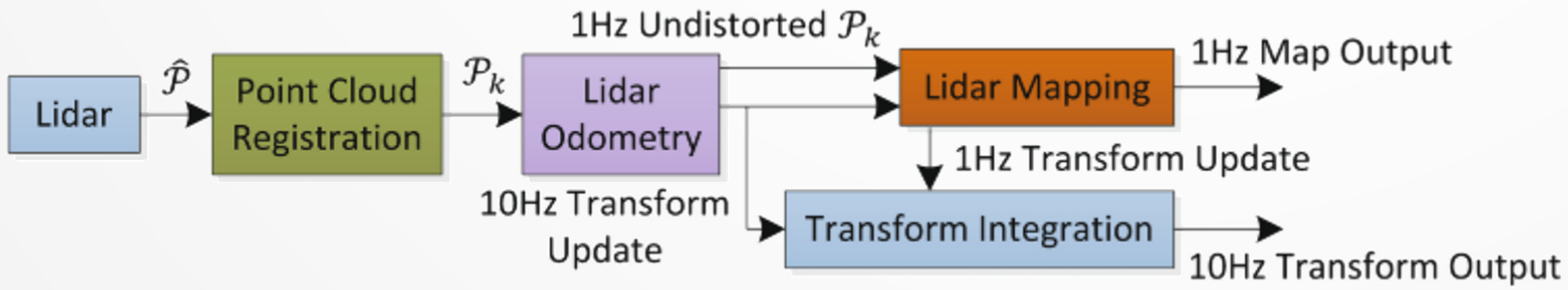
- Mesh Triangulation: Pointcloud "slices" of TSDF



Dense triangular mesh

Depth – Time-of-Flight

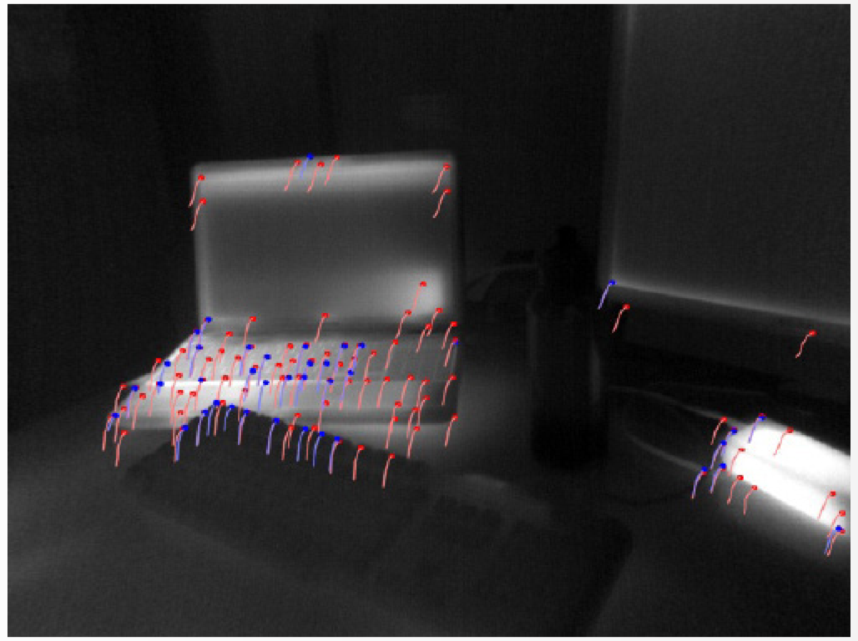
- LIght Detection And Ranging:
  - Extreme accuracy over long distances.
  - Local surface smoothness
  - Find planes (pathes), edges, track across LiDAR sweeps
  - Optimization-based (Levenberg-Marquadt)
  - Integration of pointclouds, Transform estimation at different rates.



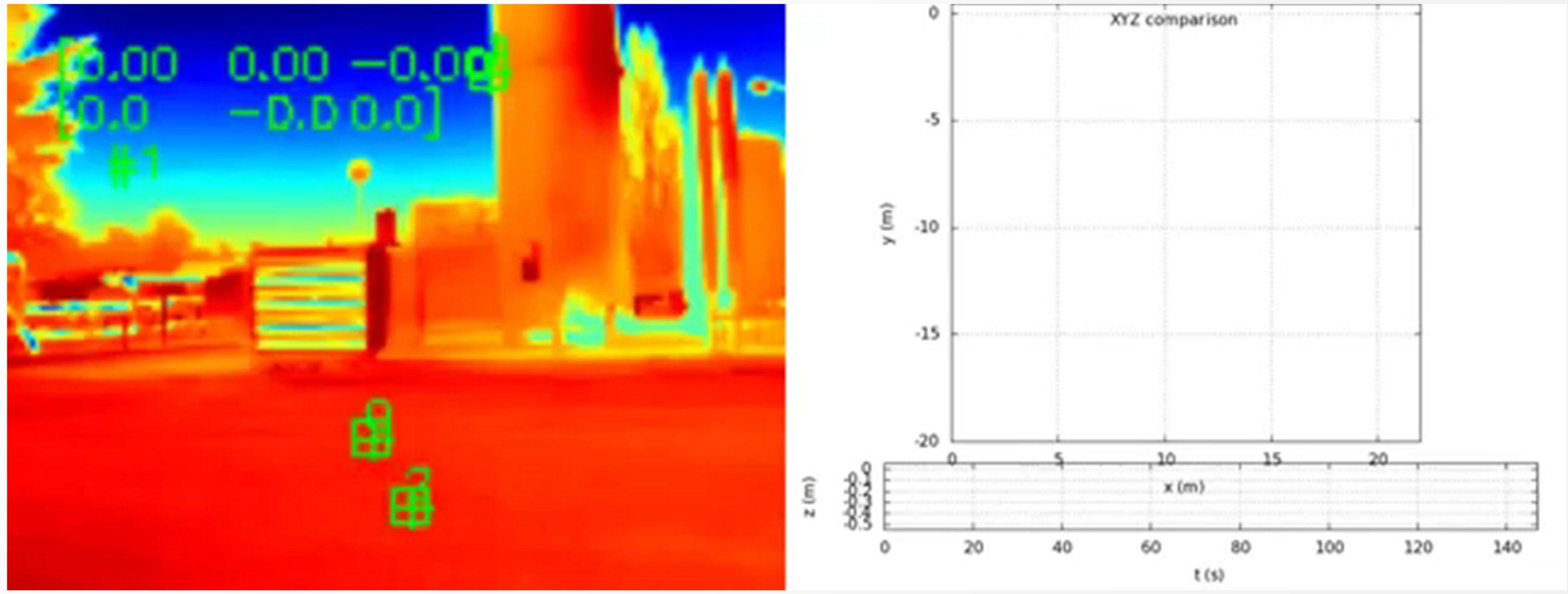
- Zhang, Ji, and Sanjiv Singh. "Low-drift and real-time LiDAR Odometry And Mapping." Autonomous Robots, 2016.

Thermal Cameras

- Monocular Vision (non-visible spectrum)
  - Feature-based
    - FAST, GFFT (Shi-Tomashi)
    - S. Vidas and S. Sridharan, "Hand-held monocular SLAM in thermal-infrared," ICARCV, 2012.

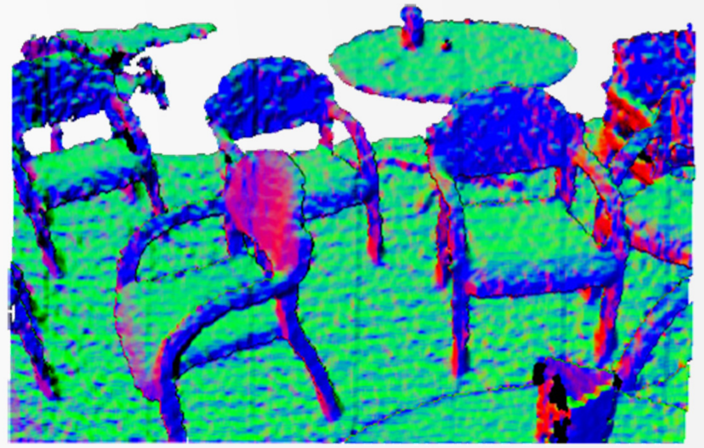


- Benefits:
  - Unique Invariance !



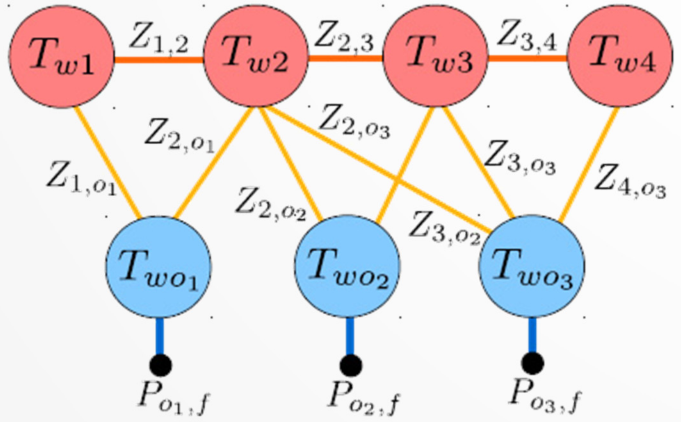
Extras

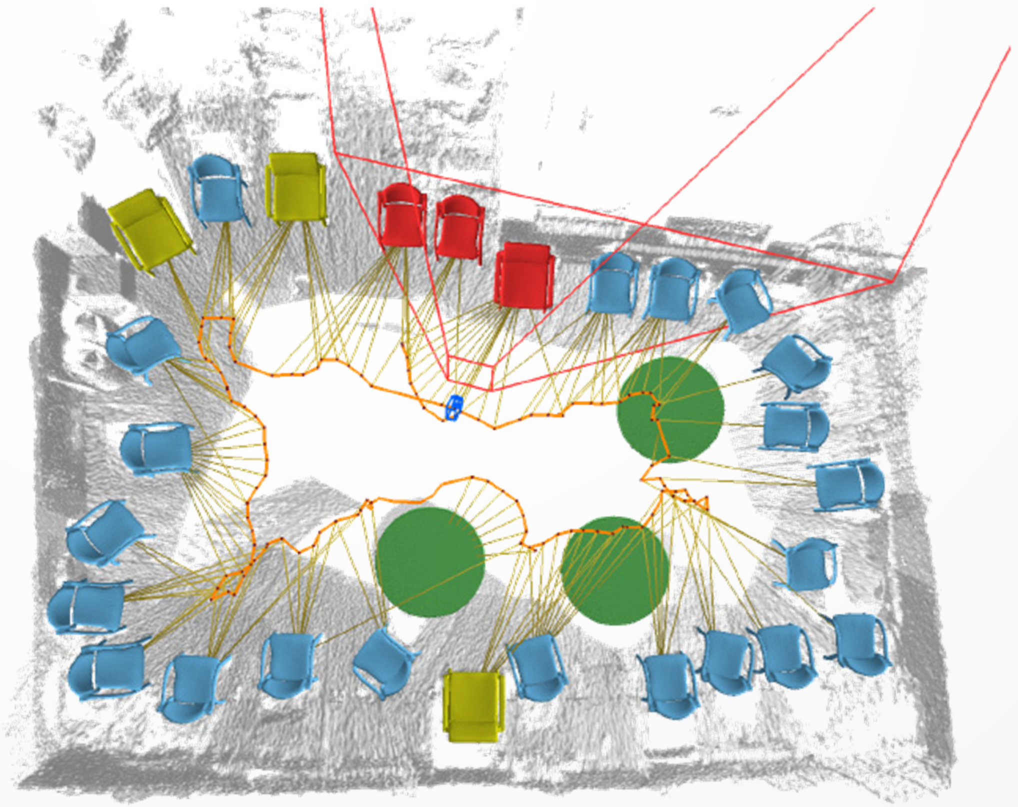
- Semantic SLAM:
  - Renato F. Salas-moreno , Richard A. Newcombe , Hauke Strasdat , Paul H. J. Kelly , Andrew J. Davison, “SLAM++ : Simultaneous Localisation and Mapping at the Level of Objects”, CVPR 2013



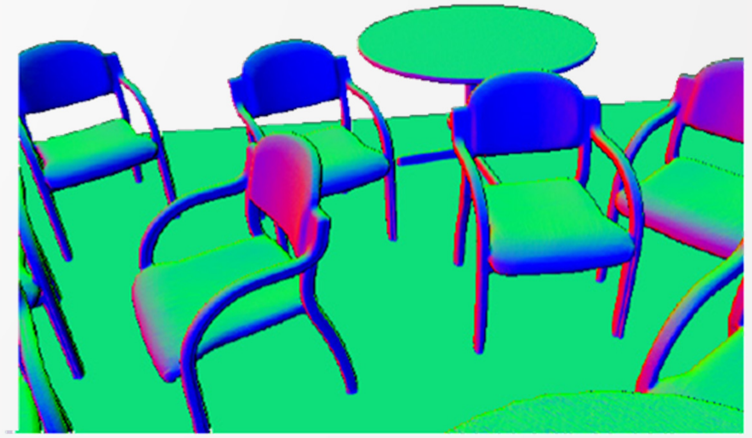
- Relies on Database of known objects.

- Map is a pure graph of objects.





- ICP between measurement and Rendered World



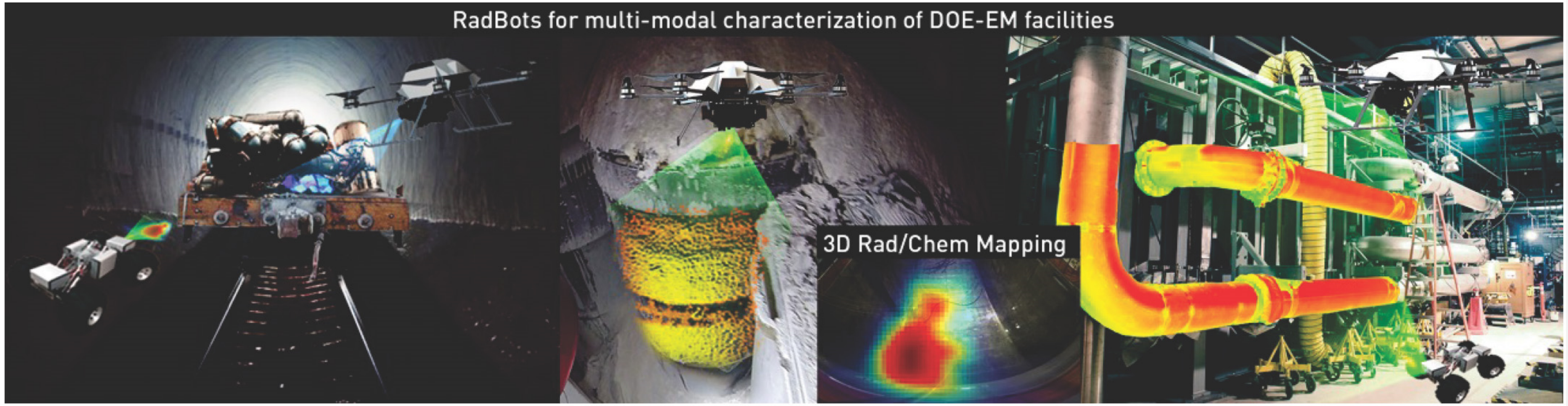
Research & Development at ARL

Autonomous Multi-Modal Localization and Mapping:

Fundamentals and the State-of-the-Art

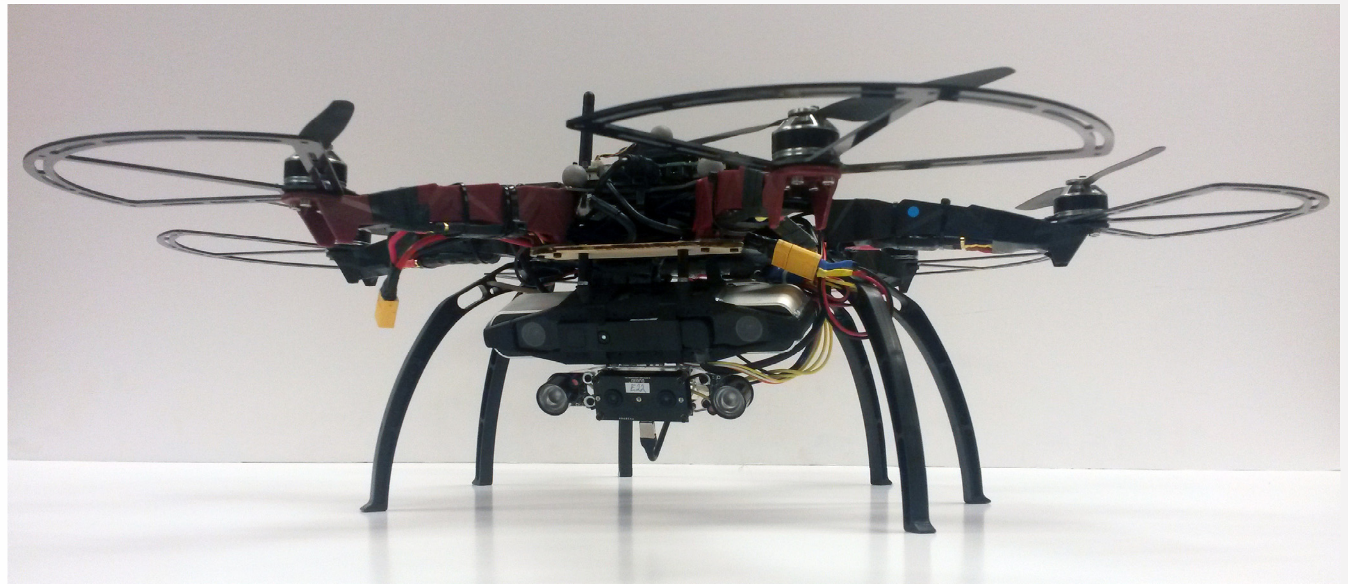
Multi-Modal Characterization of DOE-EM Facilities

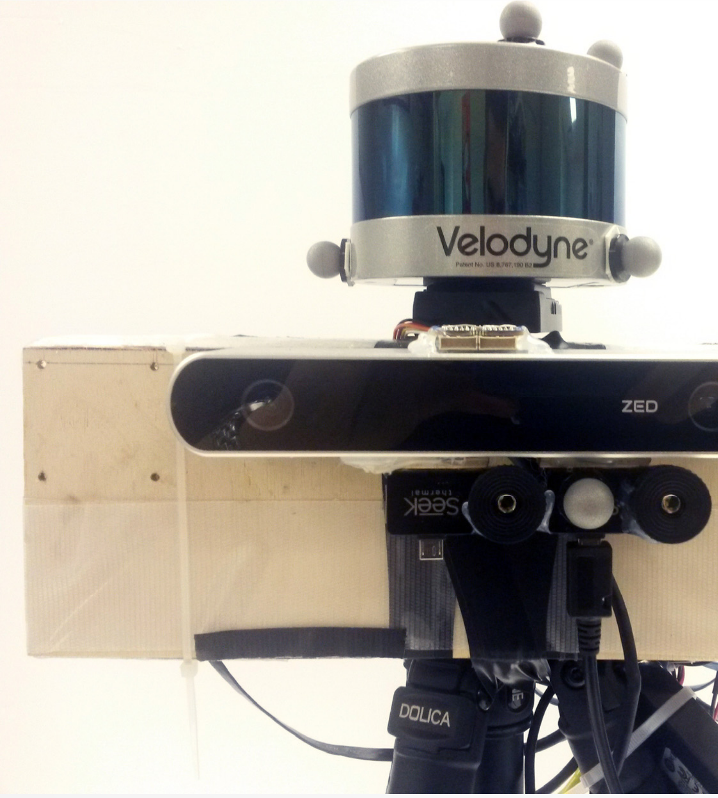
- Challenges:
  - Unknown Maps.
  - Ambiguous / Degraded-structure subsets.
  - Visually Degraded Environment.
  - Tight clearances.

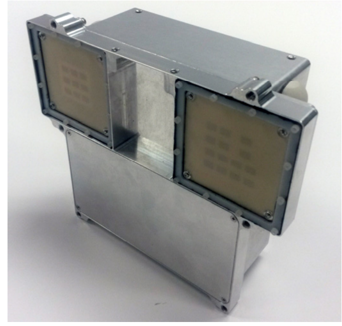


Multi-Modal Characterization of DOE Nuclear Facilities

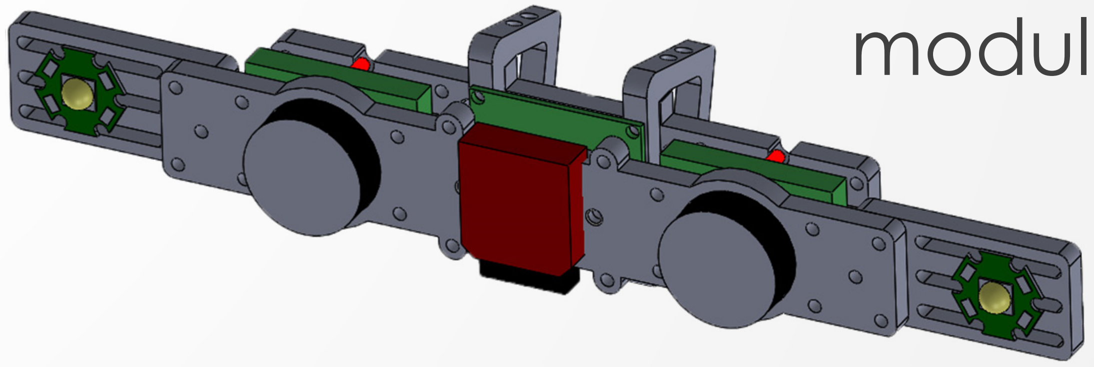
- Multi-Modal sensing.
  - Stereo Vision.
    - Visible light
    - NIR Spectrum
  - Visual-Inertial Fusion.
  - Time-of-Flight.







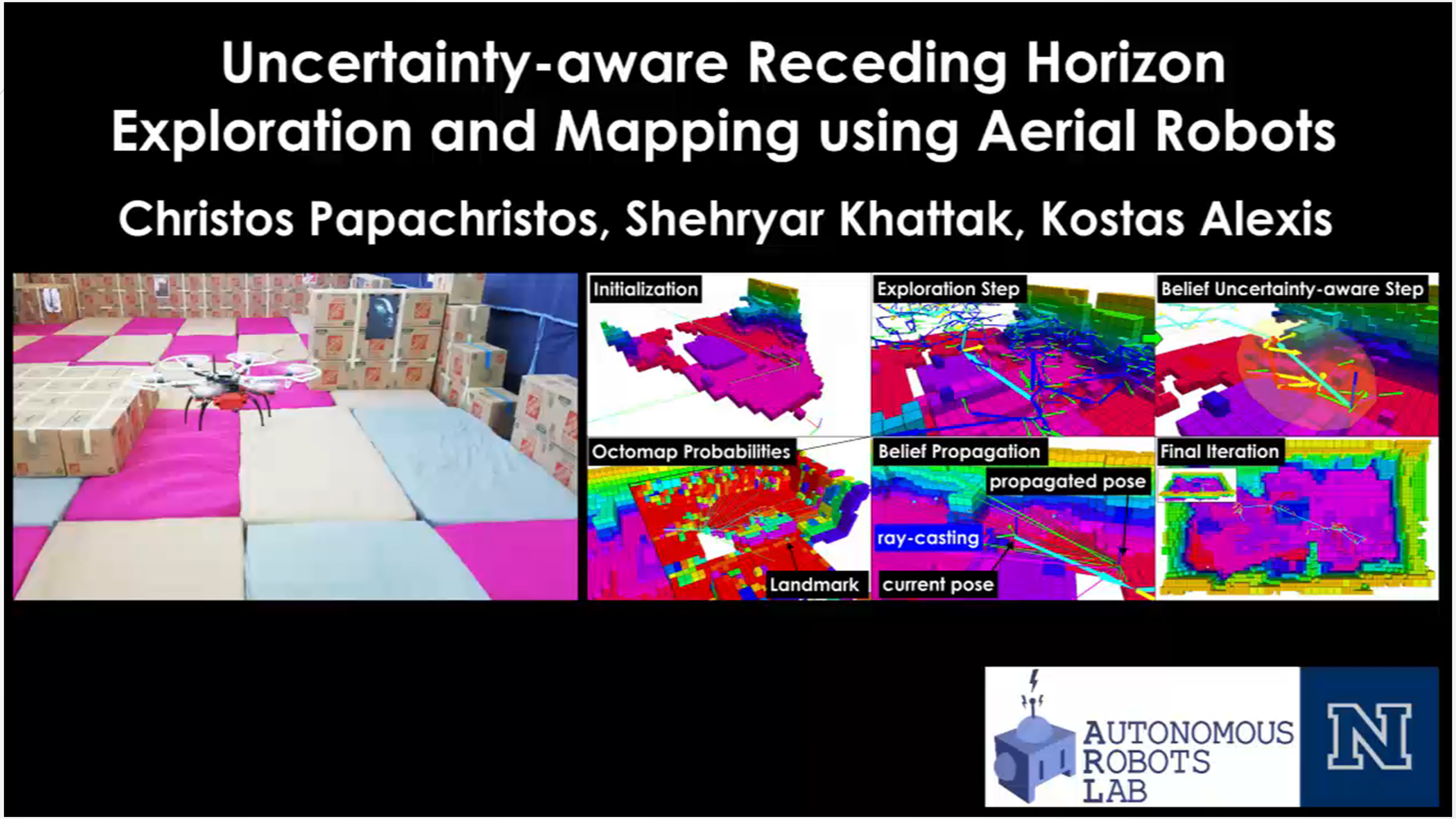
- CamSync module



- LiDAR unit.
- Thermal cameras.
- RADAR.

- Active Illumination

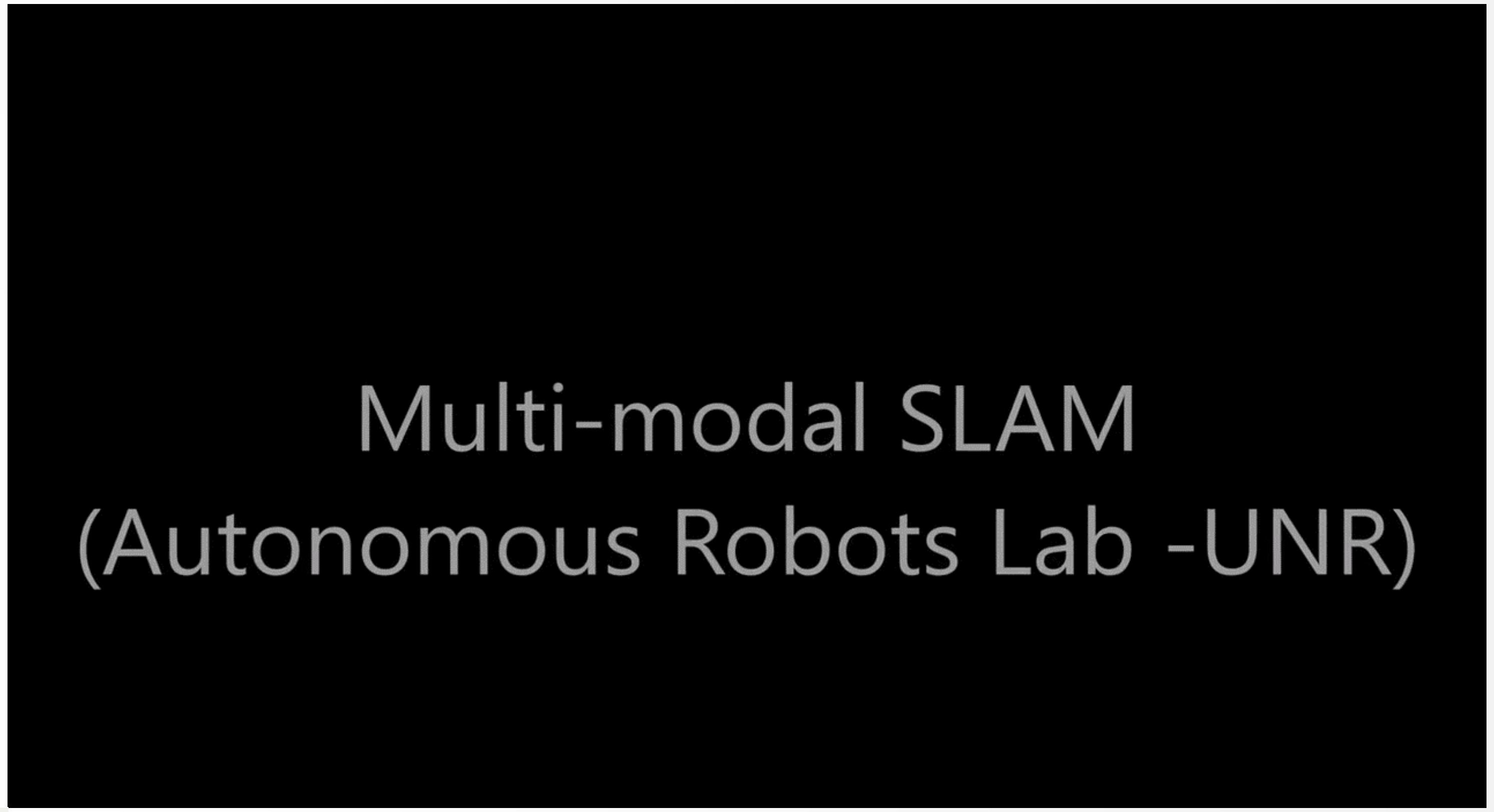
Consistent Localization & Mapping

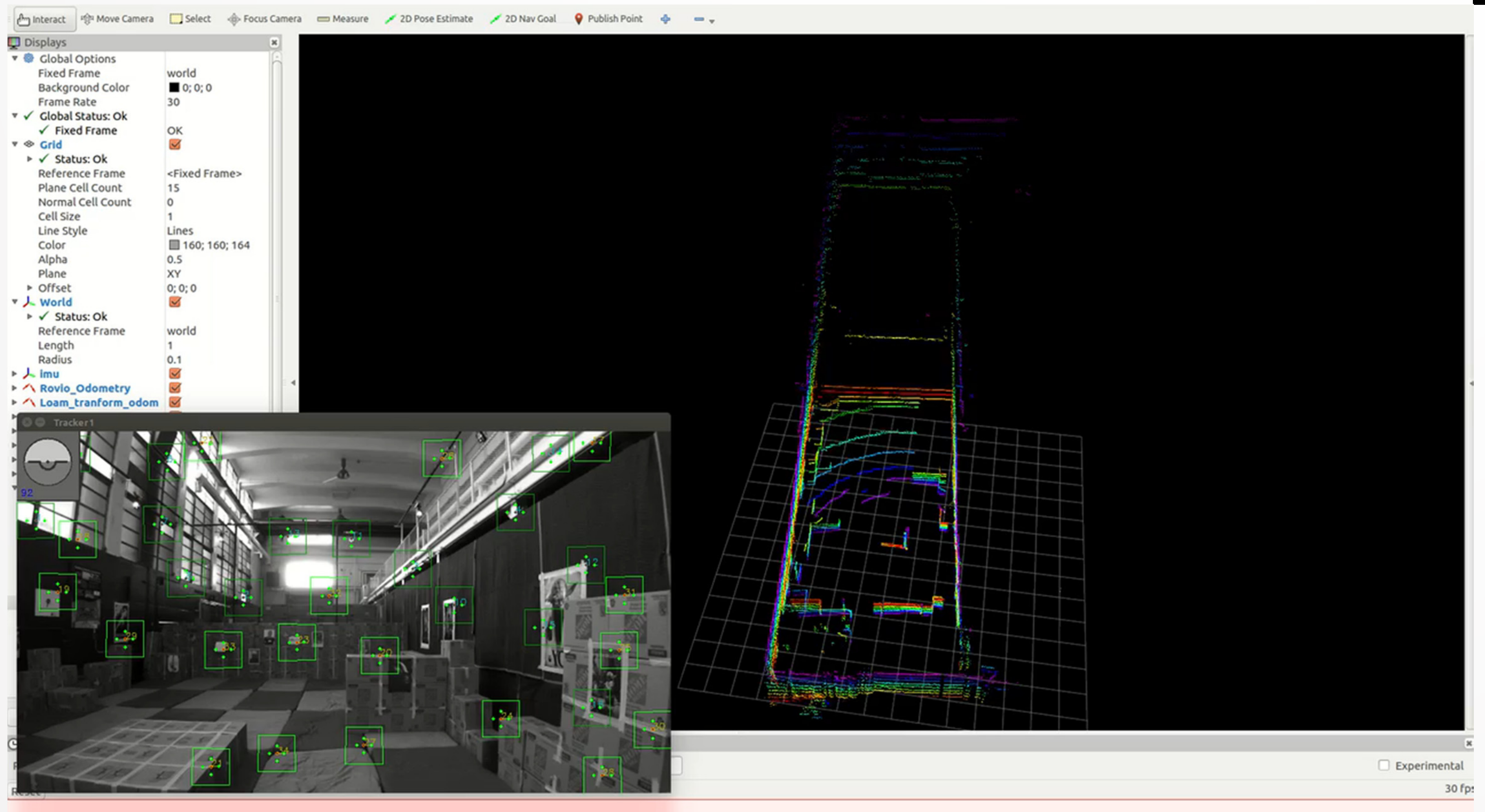


Localization & Mapping in VDE

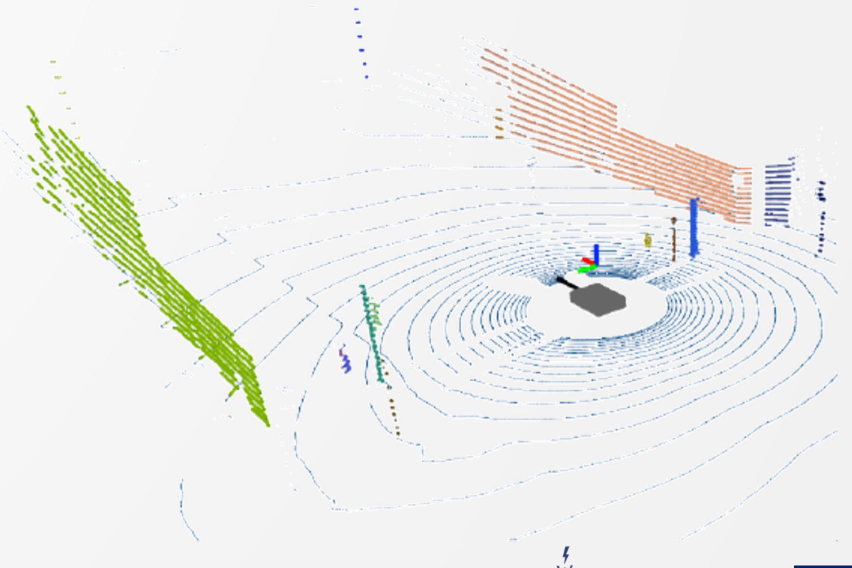
Multi-Modal Localization & Mapping

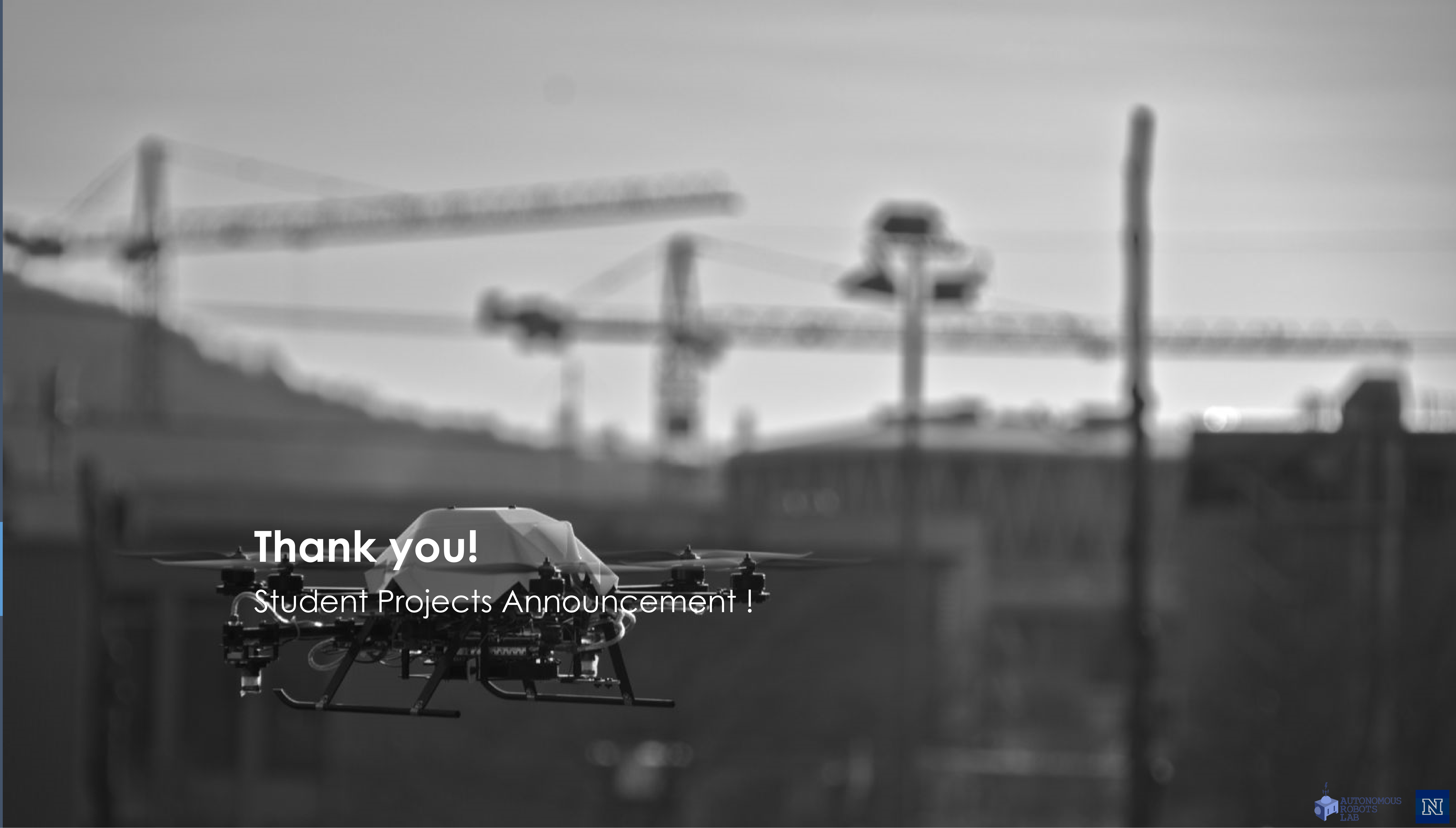
- Fusion of Multiple sensor Modalities.
  - Filter-based fusion.
  - Calibrated MM sensors package.





- Tight-fusion research.
- 3D Features



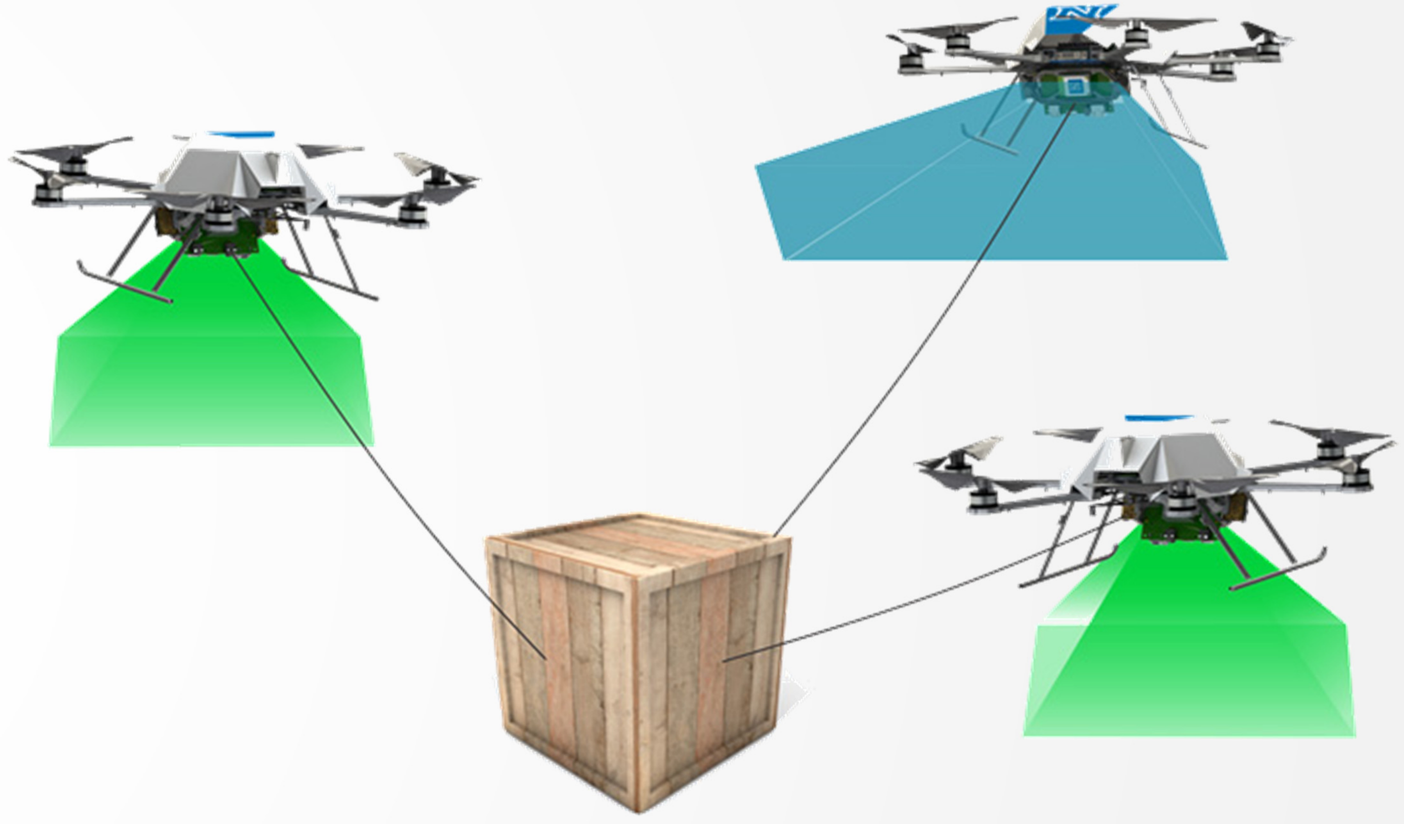


Thank you!

Student Projects Announcement !

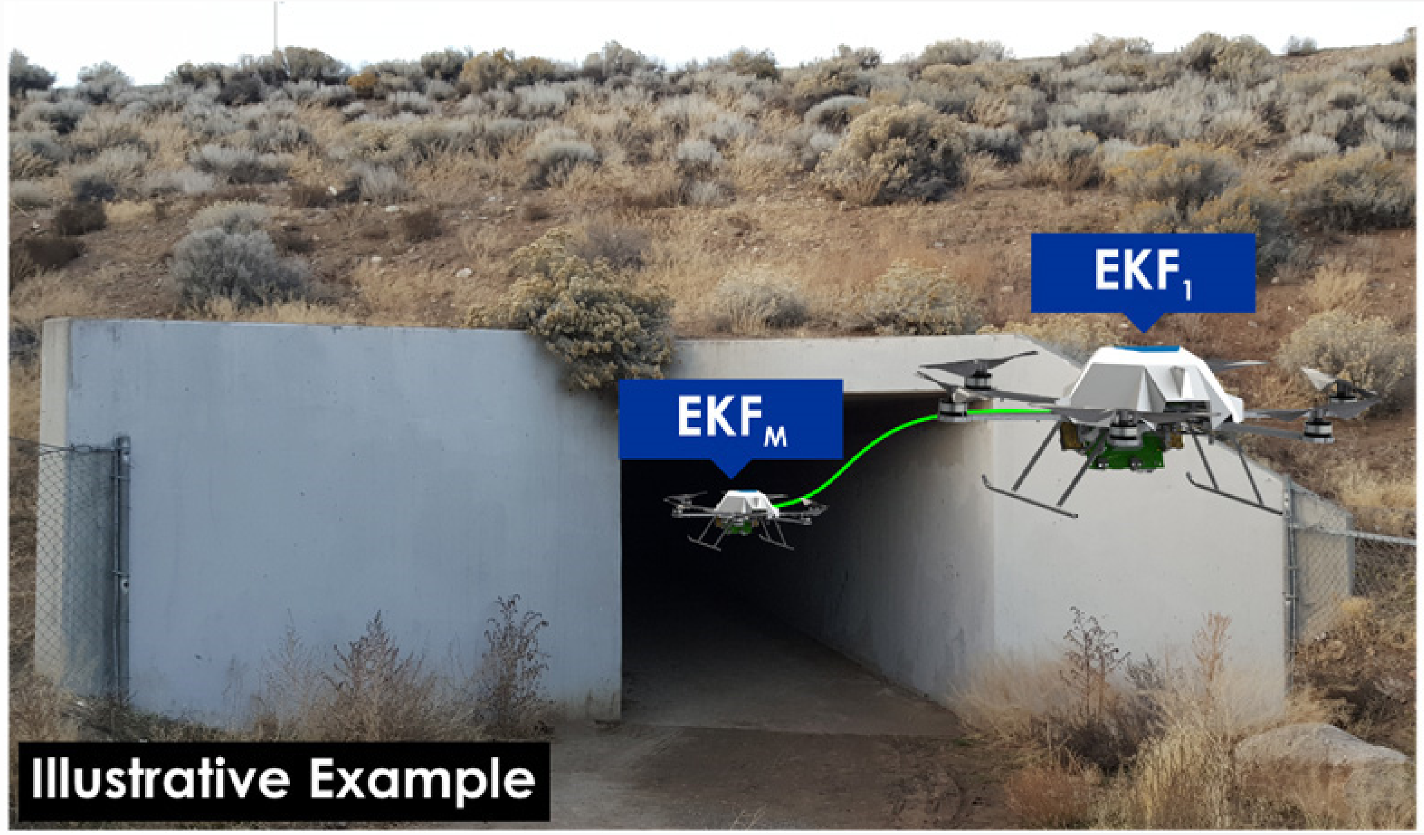
Student Projects

- **Project #1: Flying and Acting Together**
  - Perceive the world together.
  - Distributed state estimation between collaborative aerial robotic systems.
  - Collaborative navigation and mapping.
  - Collaborative physical action for tasks such as aerial transportation.
  - Constructive development and testing using the facilities of the Autonomous Robots Arena.
  - **Indicative example:** Rapid beachhead building in disaster areas.



Student Projects

- **Project #2: Adapting to the Environment**
  - Learn improved localization and planning behaviors by evaluating different active perception or multi-modal fusion strategies in different environment subsets.
  - Identify the map between environment types, optimize active perception and multi-modal fusion strategies.
  - Constructive development and testing using the facilities of the Autonomous Robots Arena.
  - **Indicative example:** Robot that operates in partially well-lit & dark. *Learn* best behavior, in first steps. *Adapt* automatically to different cases.



Thank you!

Please ask your question!