ETH zürich

Feature-based target tracking

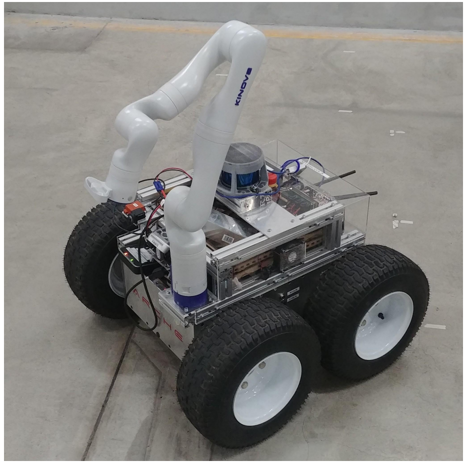
Bachelor Thesis

Theresa Wakonig

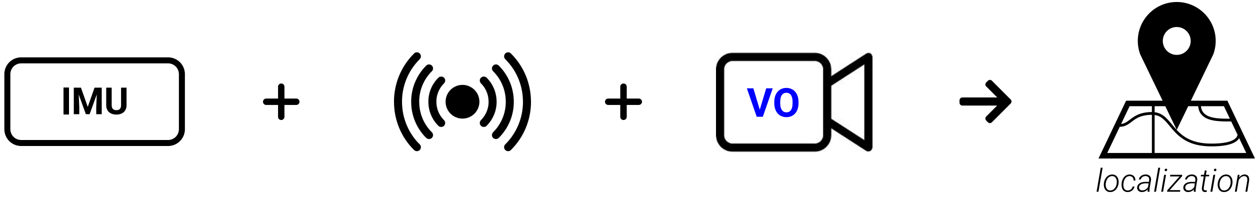
Prof. Dr. Margarita Chli
Dr. Timothy Sandy, Luca Bartolomei

VISION FOR ROBOTICS LAB

Motivation



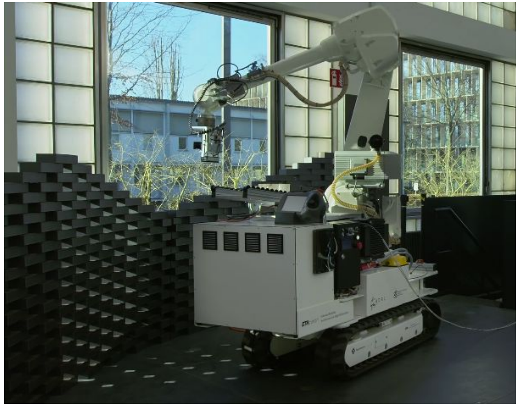
SWACO



Motivation

Aim: Tight coupling of **visual odometry** with robot's full state estimation

- Feature-tracking algorithm alongside state estimation
- Jointly estimate all sensor states
- Smoother state estimates



[1] In-situ Fabricator

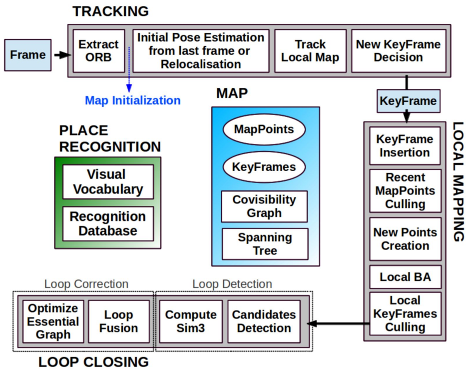
Related work

- Big variety of approaches to odometry and SLAM algorithms
- Common objective: find *best* estimate for model parameters

Successful visual SLAM systems in relation to my work ...

- PTAM [2]: tracking and mapping in parallel, keyframe-based SLAM
- ORB-SLAM [3]: tracking, local mapping and loop closing in parallel

- **My framework:** feature-based, batch optimization



[4] ORB-SLAM system overview

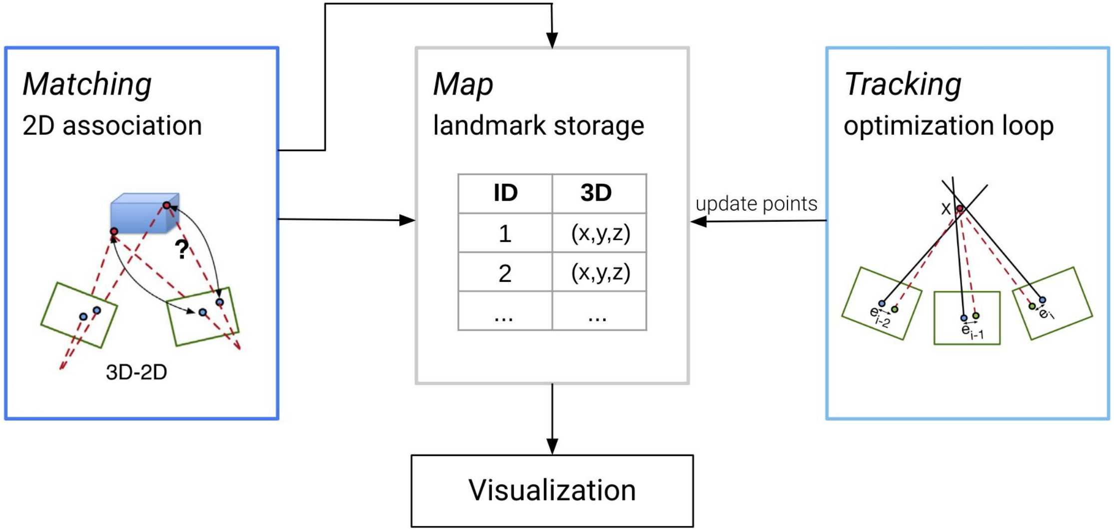
Related work

Typical building blocks of keyframe-based SLAM frameworks...

Initialization	Data association	Map generation	Map maintenance	Loop detection	Failure recovery

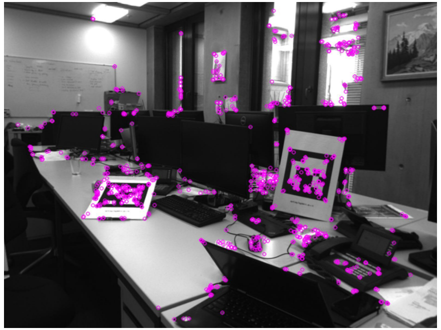
Visual Odometry Structure

- Two threads run in parallel: *matching* and *tracking*
- *Map* serves as point storage

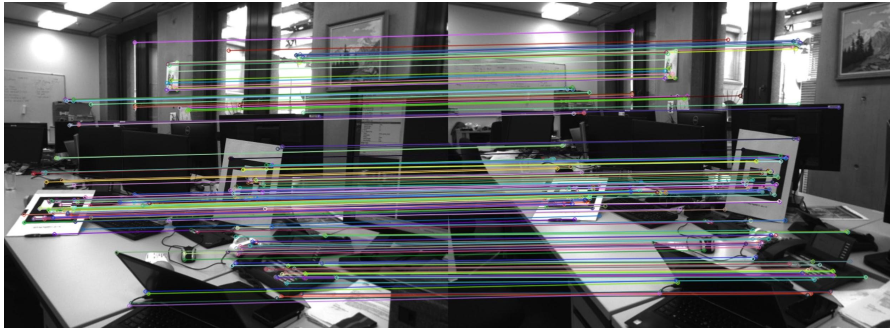


Matching Thread

- Map point generation via triangulation
- Identifying matches to map points
- Lifetime of a map point: triangulation ⟶ tracking ⟶ removal



Feature extraction

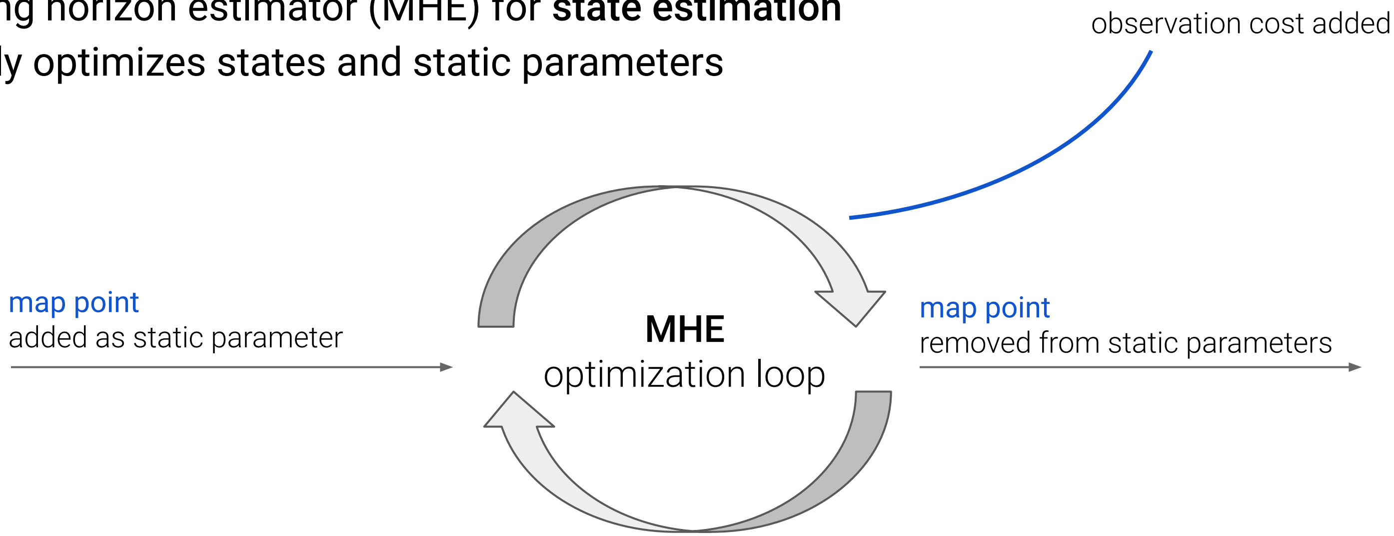


Feature matching

Tracking Thread

- Feature observations are fused over time - “optimization loop”

- Moving horizon estimator (MHE) for **state estimation**
- Jointly optimizes states and static parameters



State Estimation Framework - ConFusion [4]

- Online sensor fusion framework
- Moving horizon estimator (MHE), active batch of estimated states
- Jointly optimize state and static parameters

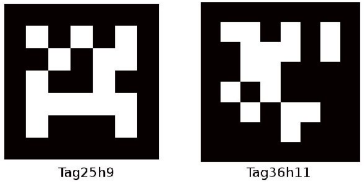
- Map point: reprojection error minimized

Objective function:

$$\{\boldsymbol{x}^*_{t_i:t_{i+N}}, \boldsymbol{s}^*\} = \underset{\boldsymbol{x}_{t_i:t_{i+N}},\boldsymbol{s}}{\arg\min} \underbrace{\|\boldsymbol{e}_p(\check{\boldsymbol{x}}_{t_i}, \boldsymbol{x}_{t_i}, \check{\boldsymbol{s}}, \boldsymbol{s})\|^2}_{\textit{prior constraint}} + \underbrace{\sum_{j=i}^{i+N}\left(\sum_{\hat{\boldsymbol{u}}\in\boldsymbol{\mathcal{U}}_{t_j}} \|\hat{\boldsymbol{u}} \boxminus g(\boldsymbol{x}_{t_j},\boldsymbol{s})\|^2_{W_{\hat{u}}}\right)}_{\textit{update measurements cost}} + \underbrace{\sum_{j=i}^{i+N}\left(\sum_{\hat{\boldsymbol{p}}\in\boldsymbol{\mathcal{P}}_{t_{j-1}:t_j}} \|\boldsymbol{x}_{t_j} \boxminus h(\boldsymbol{x}_{t_{j-1}},\hat{\boldsymbol{p}},\boldsymbol{s})\|^2_{W_{\hat{p}}}\right)}_{\textit{process measurements cost}}$$

Setup and Testing

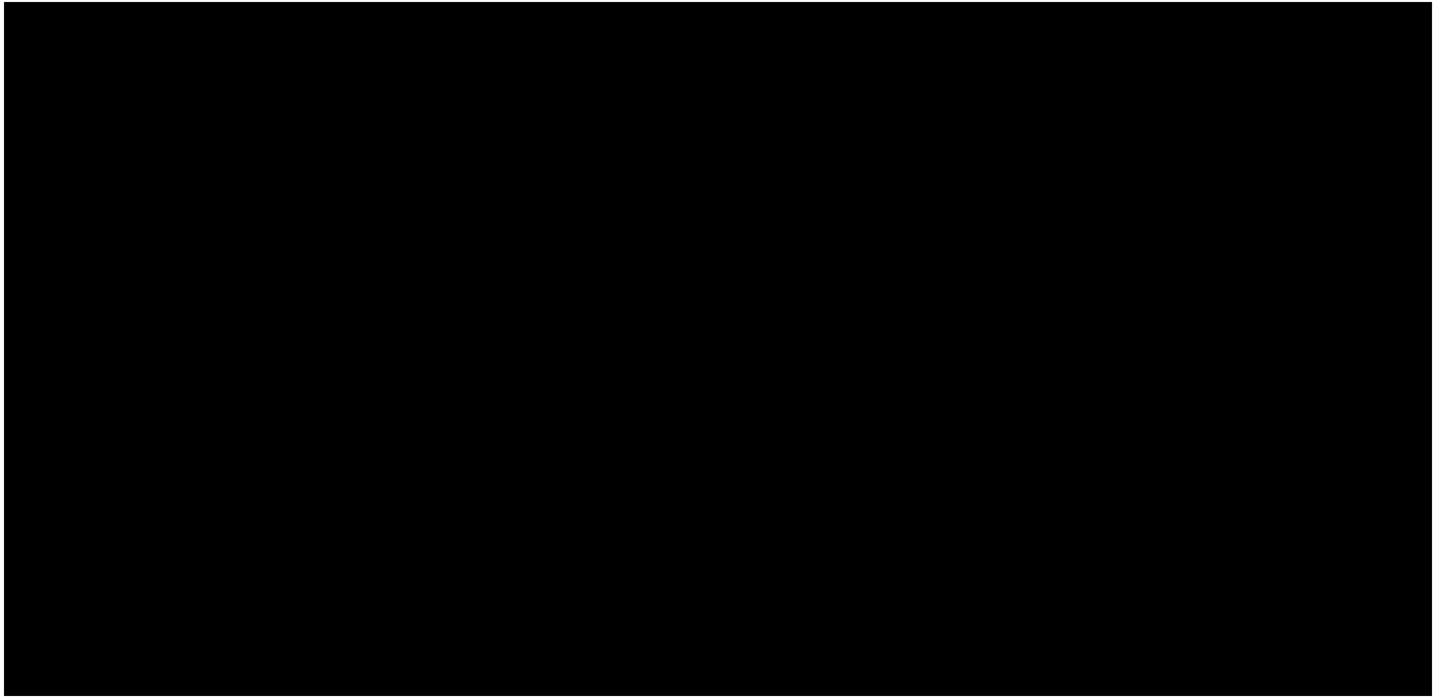
- AprilTags and IMU fusion as state estimator of robot
- Tag tracking vs. tag tracking **+ visual odometry**
- Ground truth given by Vicon system



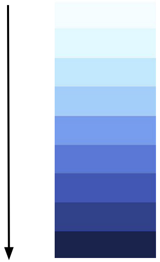
AprilTag fiducial markers


Sample bag file

Results - Map Point Refinement



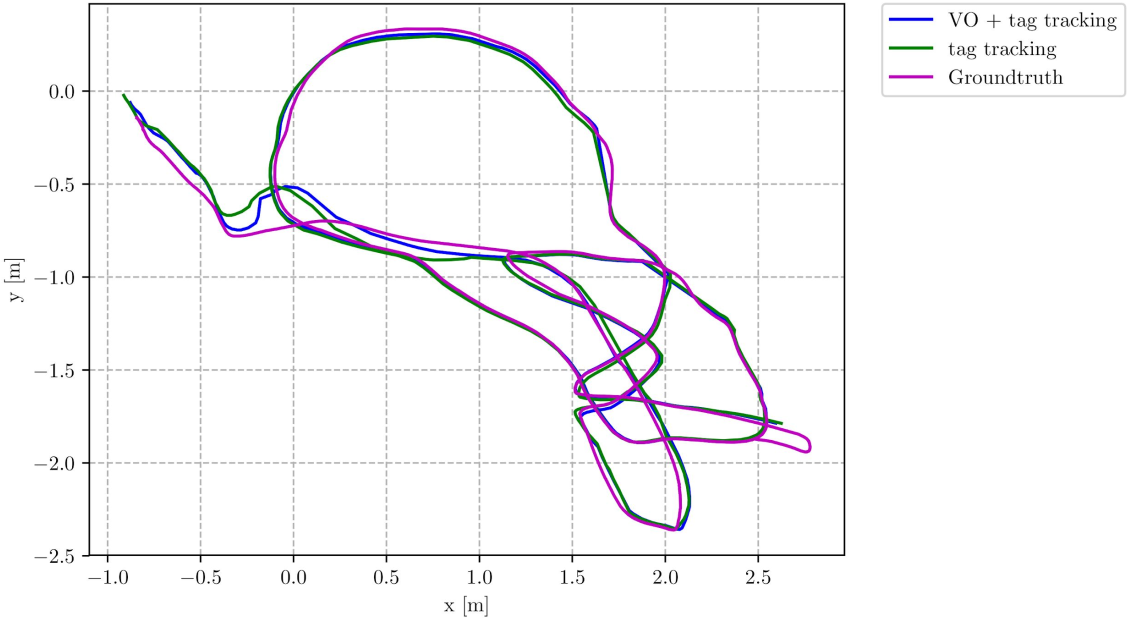
initial guess



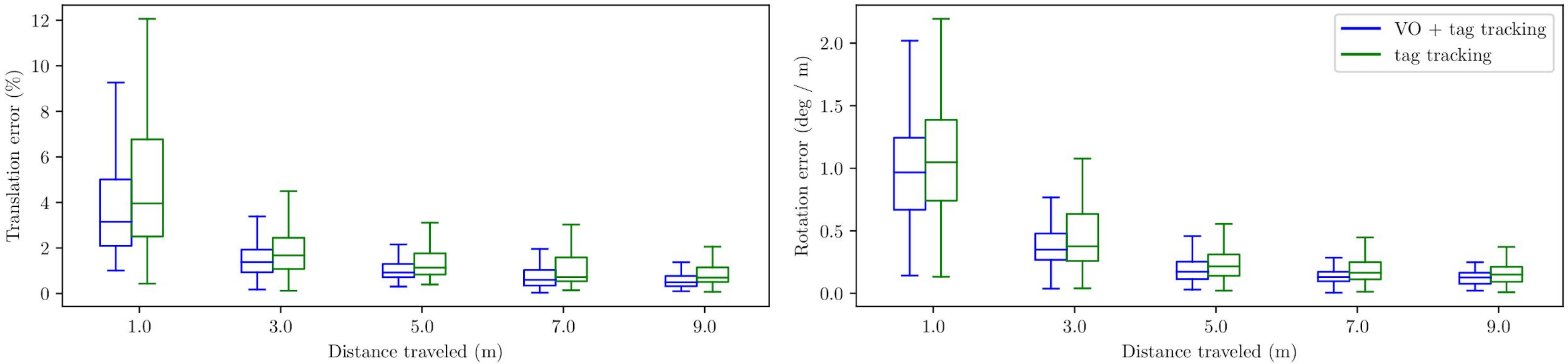
number of observations

Results - Pose Estimation

- Top view of estimated trajectory compared to Vicon ground truth



Results - Relative Pose Error



- Error slightly reduced when features tracked (visual odometry + tag tracking)
- Estimates generated using a batch size of N = 8

Conclusion

Implementation of visual odometry framework …

- Abstraction of visual measurements
- Point cloud generation
- Added feature observations to state estimator

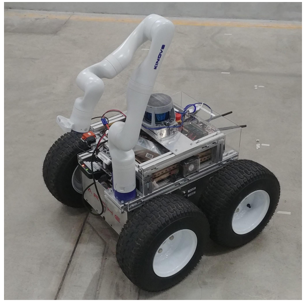
Results …

- Point cloud refinement satisfactory
- Pose error slightly reduced

Outlook

Continuation of the project and open tasks ...

- Implement system initialization → remove tags
- Visual odometry framework as built in module into ConFusion
- Test with real state estimation of SWACO robot
- Work on performance of algorithm (e.g. use epipolar search for matching)



SWACO

References

[1] T. Sandy, M. Giftthaler, K. Dörfler, M. Kohler and J. Buchli, "Autonomous repositioning and localization of an in situ fabricator," 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, 2016, pp. 2852-2858, doi: 10.1109/ICRA.2016.7487449.

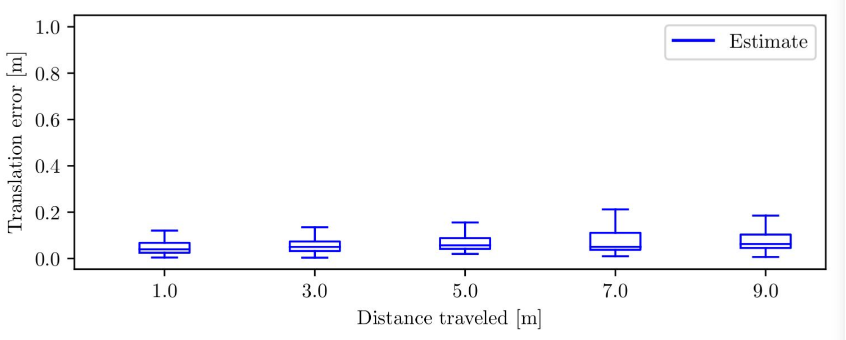
[2] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07), Nara, Japan, November 2007.

[3] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," in IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, Oct. 2015, doi: 10.1109/TRO.2015.2463671.

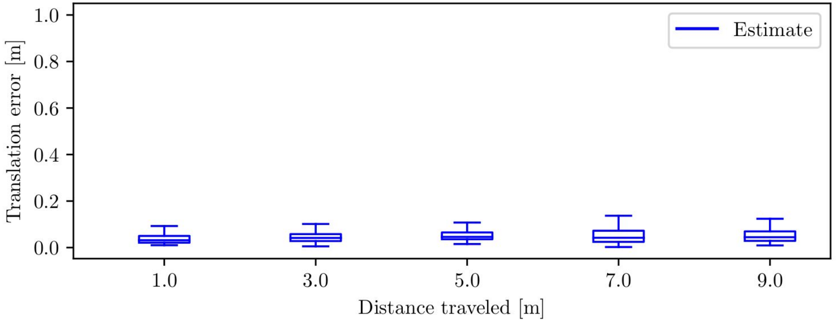
[4] T. Sandy, L. Stadelmann, S. Kerscher and J. Buchli, "ConFusion: Sensor Fusion for Complex Robotic Systems Using Nonlinear Optimization," in IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1093-1100, April 2019, doi: 10.1109/LRA.2019.2894168.

Relative Translation Error

- Tag tracking:

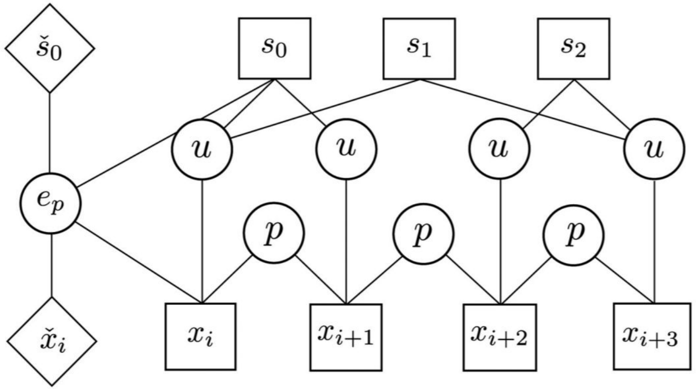


- Tag tracking + VO:

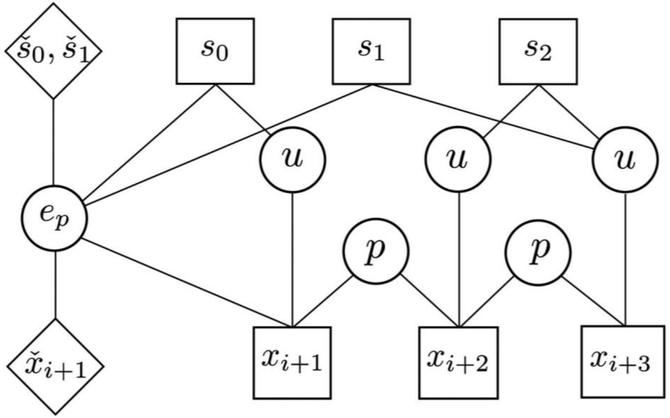


Marginalization

- Before marginalization

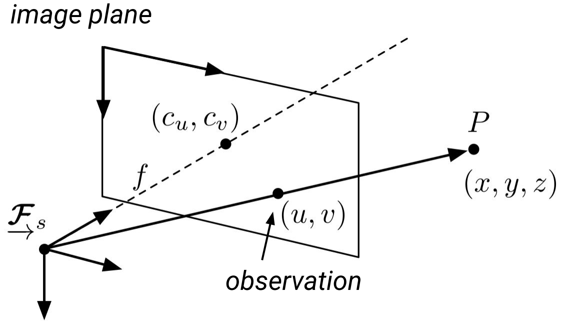


- After marginalization



Feature Observation Cost

- Joint optimization of camera poses and landmarks
- Minimization of the reprojection error
- Pinhole camera model



- Residual vector: $$\mathbf{r_i} = w \cdot \underbrace{\begin{bmatrix} p_{i,u_{obs}} - p_{i,u_{est}} \\ p_{i,v_{obs}} - p_{i,v_{est}} \end{bmatrix}}_{\text{error}}$$

- Projected point: $$\mathbf{p_{i,est}} = \begin{bmatrix} f_u \cdot \frac{x_i}{z_i} + c_u \\ f_v \cdot \frac{y_i}{z_i} + c_v \end{bmatrix}$$

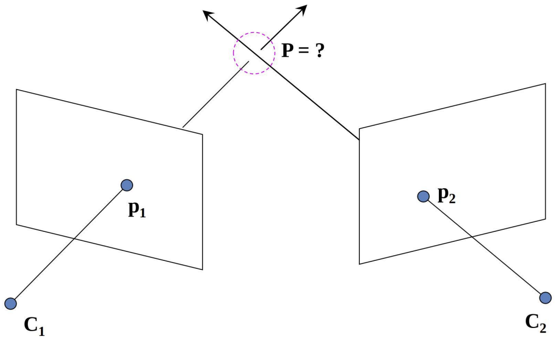
Triangulation

Linear approximation → 3D point initialization

$$\underbrace{\lambda \cdot p_i}_{\text{parallel}} = \underbrace{M_i \cdot P}_{} \quad \rightarrow \quad p_i \times M_i \cdot P = 0 \quad \rightarrow \quad [p_i]_\times \cdot M_i \cdot P = 0$$

p	*projected point on image plane*
M	*projection matrix*
K	*camera matrix (intrinsic)*
R	*rotation matrix*
t	*translation vector*
P	*3D point coordinates*
i	*keyframe*

$$M_i = \underbrace{K}_{\text{camera to image}} \cdot \underbrace{[R|t]}_{\text{world to camera}}$$



[2]