Seminar: Recent Advances in 3D Computer Vision

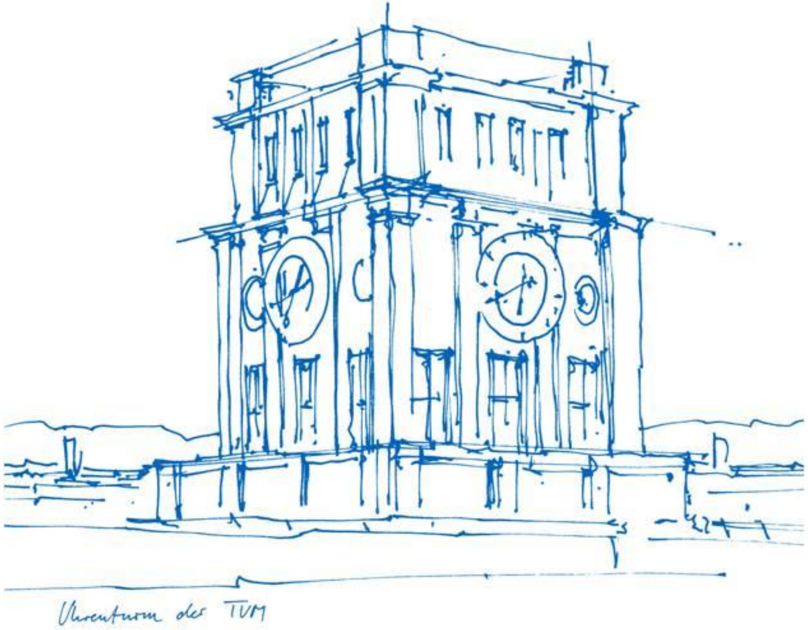
Jannik Peters

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Computer Vision & Artificial Intelligence

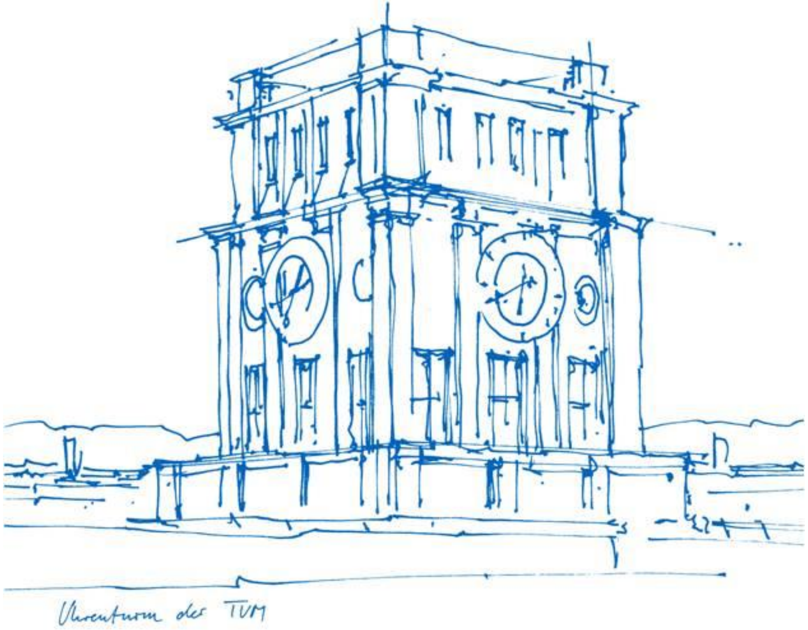
Garching, 20. November 2019



Paper: Parallel Tracking and Mapping for Small AR Workspaces

By Georg Klein and David Murray

2007



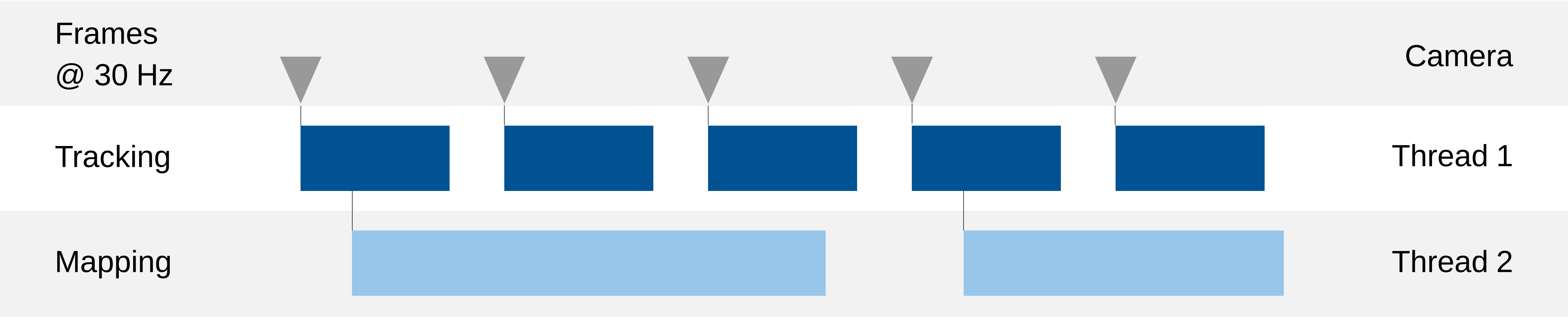
Agenda

1. Introduction
2. Overview
3. Tracking method description
4. Mapping method description
5. Experiments and results
6. Summary

Introduction

- SLAM: Camera pose at frame rate
- Single Monocular Camera (Greyscale data)
- Parallelized Tracking and Mapping
  - Tracking with hard real time constraints at frame rate
  - Mapping slower with sophisticated, global optimizations
  - Utilize multi core processors

Overview

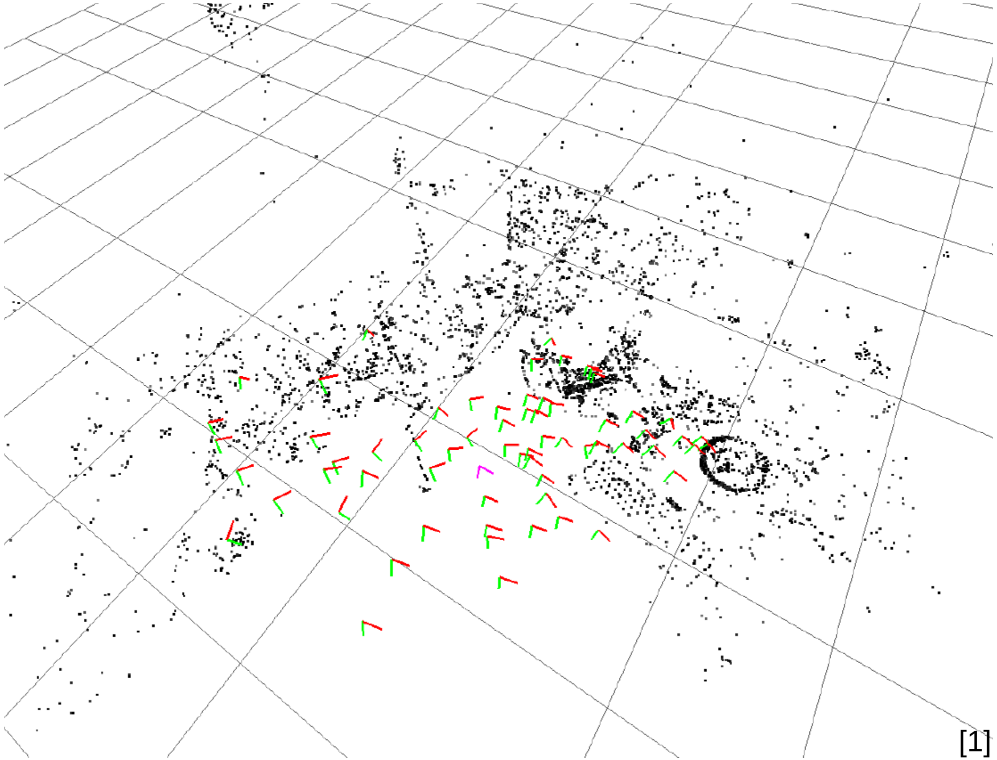


Overview

```
struct Map {
        List<MapPoint>,
        List<KeyFrame>,
}

struct KeyFrame {
        List<KeyPoint>,
        SE3 camera2world
}

struct MapPoint {
        Vector<3> worldPoint,
        Set<KeyFrame>,
        ImageData,
        QualityData,
}
```



[1]

Tracking

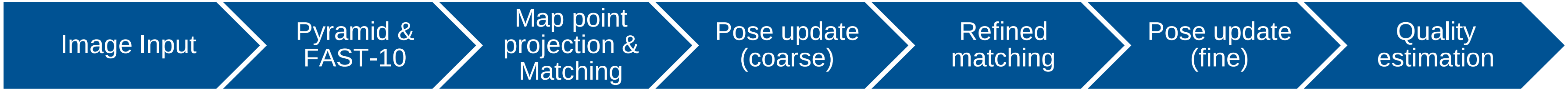


Image Input
Pyramid & FAST-10
Map point projection & Matching
Pose update (coarse)
Refined matching
Pose update (fine)
Quality estimation

Tracking

Image Input → Pyramid & FAST-10 → Map point projection & Matching → Pose update (coarse) → Refined matching → Pose update (fine) → Quality estimation

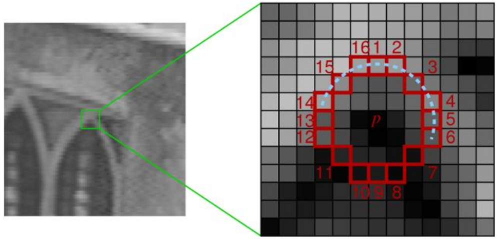
[1]

- RGB converted to grayscale
- 640 x 480 @ 30 Hz

Tracking

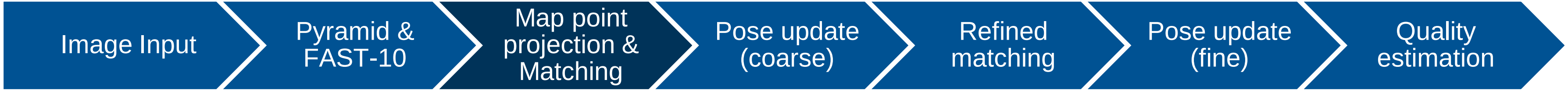
Image Input → Pyramid & FAST-10 → Map point projection & Matching → Pose update (coarse) → Refined matching → Pose update (fine) → Quality estimation

- 4 Level image pyramid
- FAST-10 corners at each level
- *Without* non-maximal suppression → Patches



[3]

Tracking



- Project all map points into image
- Match

Tracking

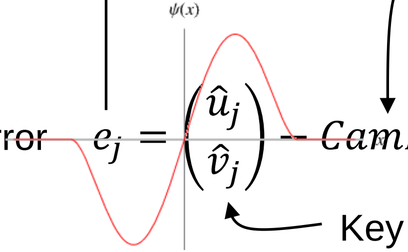
Image Input → Pyramid & FAST-10 → Map point projection & Matching → **Pose update (coarse)** → Refined matching → Pose update (fine) → Quality estimation

$$\mu' = \underset{\mu}{\operatorname{argmin}} \sum_{j\in S} Obj\left(\frac{|e_j|}{\sigma_j}, \sigma_T\right)$$

Optimize camera pose (6 DoF)

50 best matches

Tukey weights

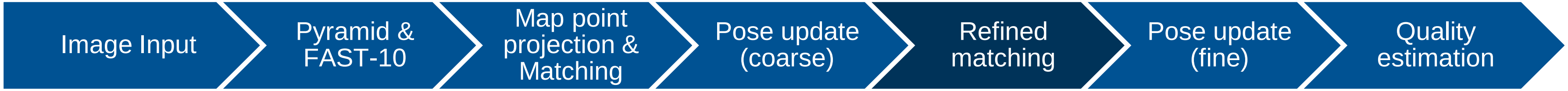


Frame transform
Pinhole camera model
Barrel distortion

Reprojection error $e_j = \begin{pmatrix}\hat{u}_j \\ \hat{v}_j\end{pmatrix} - CamProj\big(exp(\mu)E_{cw}p_j\big)$

Key point

Tracking



- Match map points to corners again
- **1000** best matches
- Sub pixel refinement for high level subset

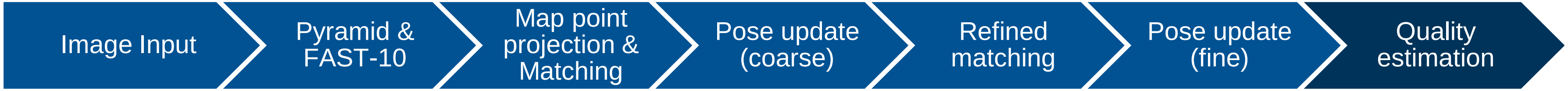
Tracking

Image Input → Pyramid & FAST-10 → Map point projection & Matching → Pose update (coarse) → Refined matching → **Pose update (fine)** → Quality estimation

$$\mu' = \underset{\mu}{\operatorname{argmin}} \sum_{j \in S} Obj\left(\frac{|e_j|}{\sigma_j}, \sigma_T\right)$$

Up to **1000** matches → $j \in S$

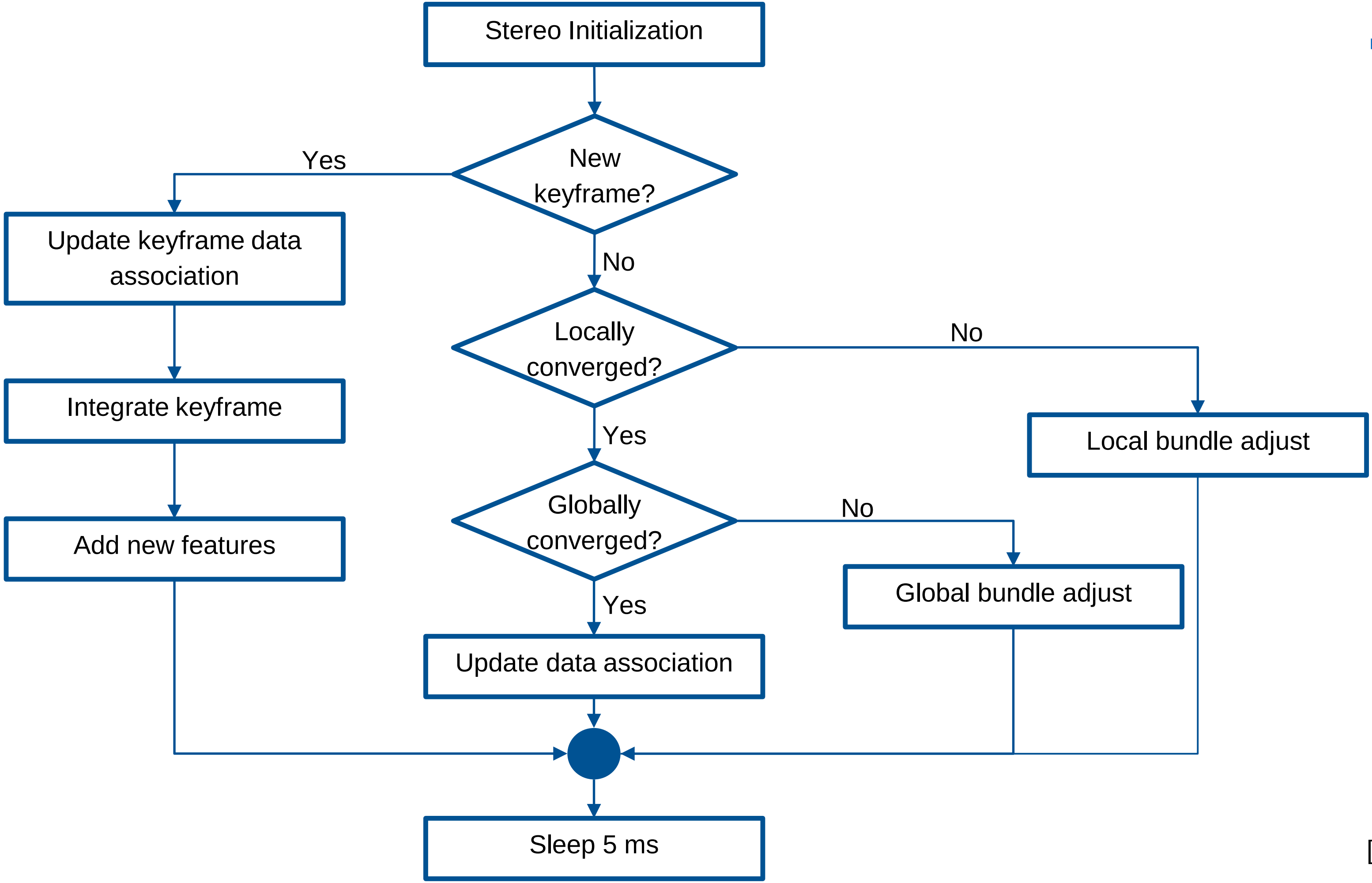
Tracking



- Ratio of successful matches as quality estimate
- < threshold → Tracking poor
- < even lower threshold for several frames → Tracking lost → Recovery

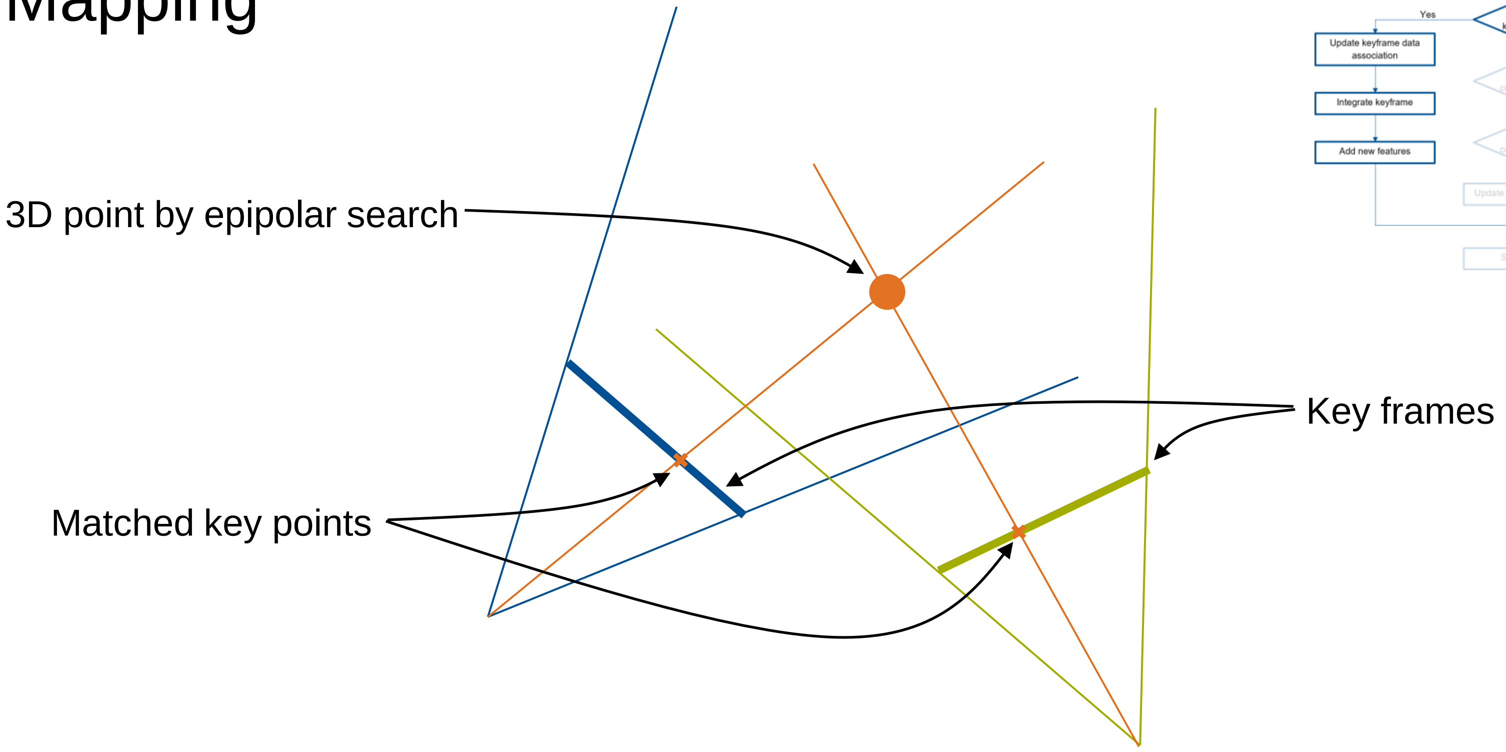
Mapping

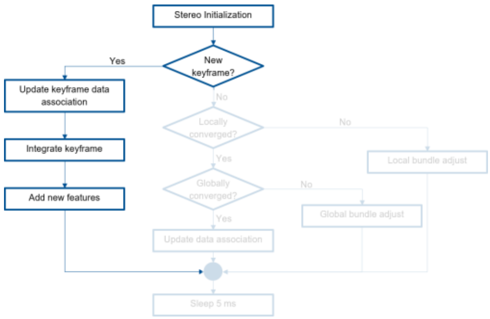
- Separate Thread
- Uses FAST-10 corners from Tracking


Stereo Initialization
New keyframe?
Yes
No
Update keyframe data association
Locally converged?
No
Yes
Integrate keyframe
Local bundle adjust
Globally converged?
No
Yes
Add new features
Global bundle adjust
Update data association
Sleep 5 ms

[1]

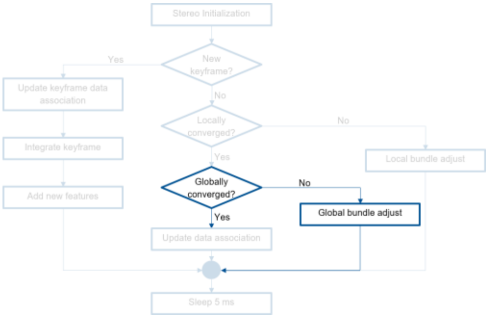
Mapping


3D point by epipolar search
Key frames
Matched key points


Stereo Initialization
New keyframe?
Yes
Update keyframe data association
Integrate keyframe
Add new features
No
Locally converged?
No
Local bundle adjust
Yes
Globally converged?
No
Global bundle adjust
Yes
Update data association
Sleep 5 ms

Mapping

- Global bundle adjustment
- All* key frames, all map points



$$\{\{\mu_2..\mu_N\},\{p'_1..p'_M\}\} = \underset{\{\{\mu\},\{p\}\}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{j \epsilon S_i} Obj\left(\frac{|e_{ij}|}{\sigma_{ij}}, \sigma_T\right)$$

All keyframes

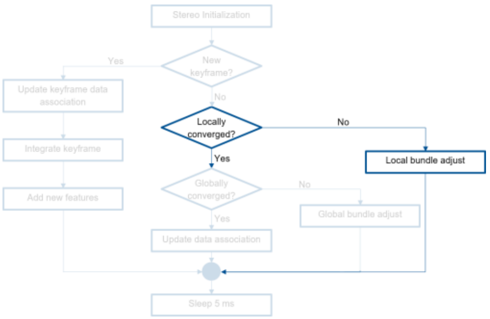
All map points

Tukey weights

Reprojection error, weighted by quality

- Expensive! → Main limitation of map size. ($\mathcal{O}\left((N+M)^3\right)$ in theory, but $\mathcal{O}\left(N^2M\right)$ in practice)

Mapping



- Local bundle adjustment → Speed optimization
- Key frames to adjust X
- Fixed key frames Y
- Some map points Z

$$\left\{\{\mu_{x\in X}\},\{p'_{z\in Z}\}\right\} = \underset{\{\{\mu\},\{p\}\}}{\operatorname{argmin}} \sum_{i\in X\cup Y} \sum_{j\epsilon Z\cap S_i} Obj\left(\frac{|e_{ij}|}{\sigma_{ij}}, \sigma_T\right)$$

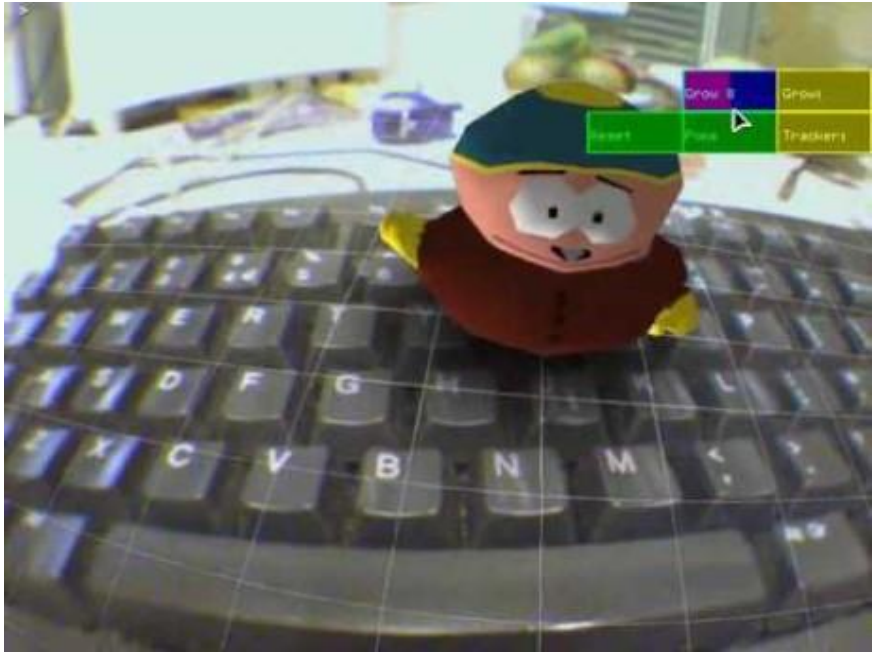
All relevant key frames

Selected keypoints

Tukey Weights

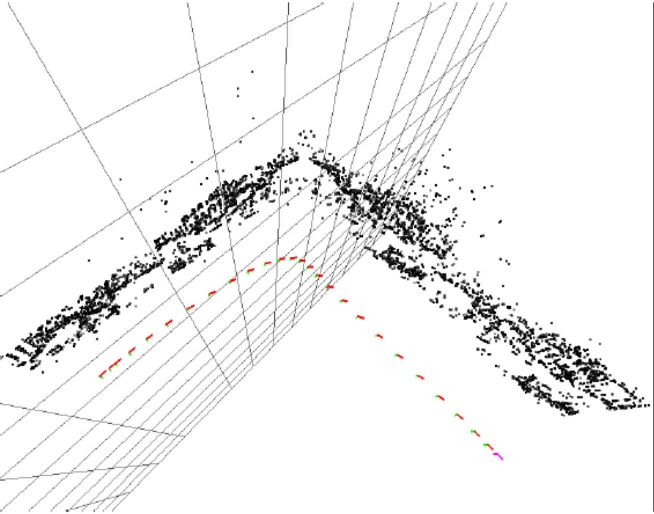
Reprojection error, weighted by quality

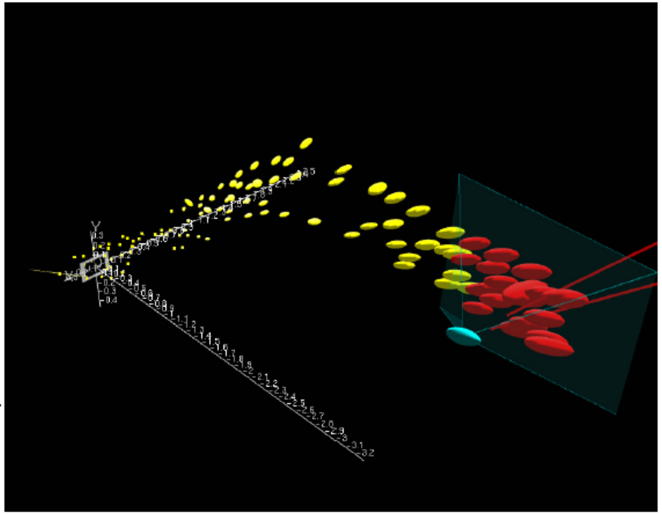
Experiments and Results

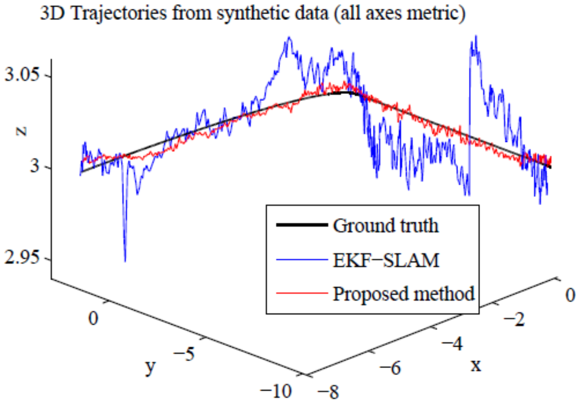


[2]

Experiments and Results

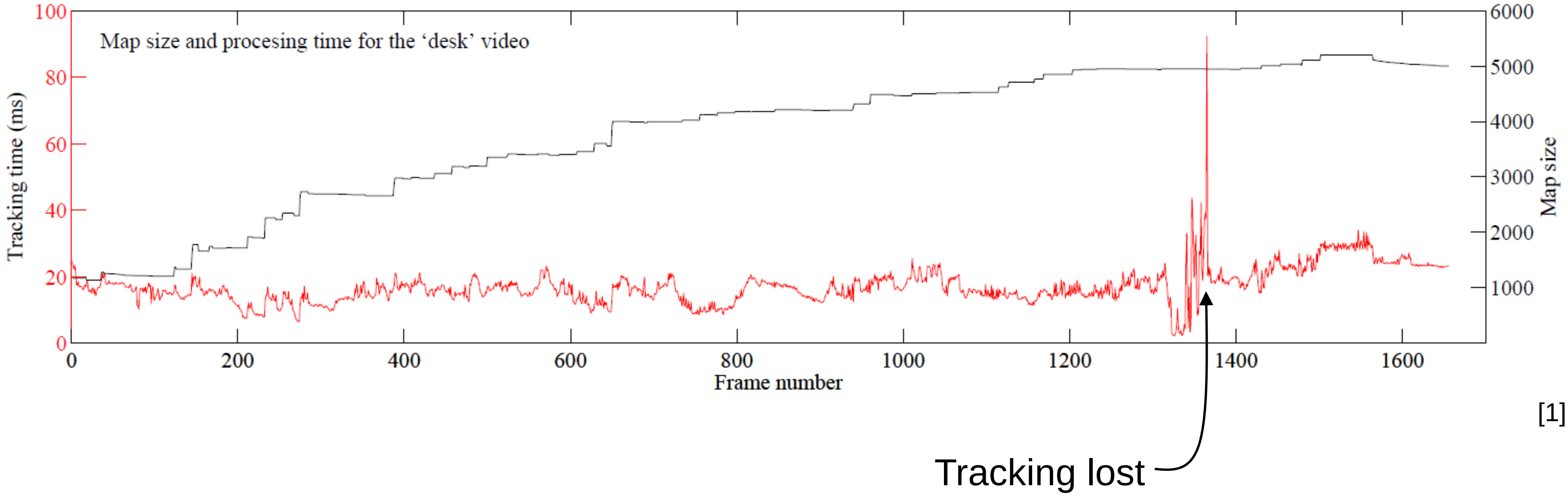






[1]

Experiments and Results



[1]

Experiments and Results

+ Pro
- Fast (real time tracking @ 30 Hz)
- Robust
- Accurate
- Simple, monocular camera

- Drawbacks
- Limited to small workspace
- No occlusion reasoning
- Non semantic map (Inadequate for AR)

Summary

- Parallelize SLAM
  - Real time tracking
  - Slower, sophisticated mapping (local and global Bundle adjust)
- Monocular
  - Grayscale images
  - FAST-10 corners as key points
  - 3D location by using >2 view angles
- Fast, robust, accurate
- Limited to small maps
  - ~6000 key points
  - $\mathcal{O}(N^2M)$ Global bundle adjust

“Parallel Tracking and Mapping for Small AR Workspaces”
- By Georg Klein and David Murray (2007)

Jannik Peters
jannik.peters@tum.de