Out-of-Core Bundle Adjustment for 3D Workpiece Reconstruction

Robert Maier

Master's Thesis in Computer Science

In collaboration with Siemens Corporate Research Inc.
Princeton, NJ, USA

Advisor: Dr. Jürgen Sturm

September 24, 2013

Outline

1. Introduction
2. Related Work
3. 3D Reconstruction System
   - RGB-D data acquisition
   - Feature-based 3D alignment
   - Mapping
   - Out-of-core bundle adjustment
   - Dense 3D model representation
4. Evaluation and Experimental Results
   - Performance evaluation
   - 3D workpiece reconstruction
5. Conclusion and Future Work

Outline

Motivation: 3D reconstruction

- Reconstruction of digital 3D models from real objects
  - Fuse multiple camera views into global representation
  - Use of novel RGB-D sensors
  - Simultaneously estimate camera trajectory and 3D model
    $\rightarrow$ Simultaneous Localization And Mapping (SLAM)
- Application scenarios
  - Robot navigation, gaming, physics, etc.
  - Reverse-engineering

Motivation: 3D workpiece reconstruction

- Special case of reverse-engineering
- Practical advantages:
  - Visual inspection
  - Exact measurements
  - Detection of deformations
  - Construction of customized tools
- Challenges:
  - Large amount of data
  - High metric accuracy
  - Efficient optimization

Objectives of this thesis

- Reconstruction of accurate dense 3D models of workpieces
- Flexible and modular RGB-D-based SLAM system
- Global drift and inaccuracies in 3D model
  $\rightarrow$ Novel bundle adjustment approach:
  - Minimization of 3D alignment error
  - Out-of-core bundle adjustment using submaps

Outline

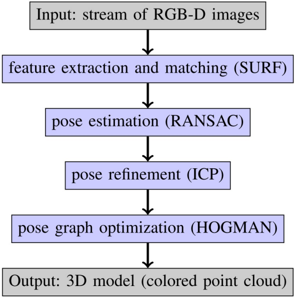
Related work: RGB-D-based 3D reconstruction

KinectFusion [Izadi et al., 2011]

- TSDF volume representation
- Real-time camera tracking based on ICP
- Limited scene size

RGB-D SLAM [Endres et al., 2012]

- Flexible processing pipeline
- Robust feature-based 3D alignment
- Pose-graph optimization

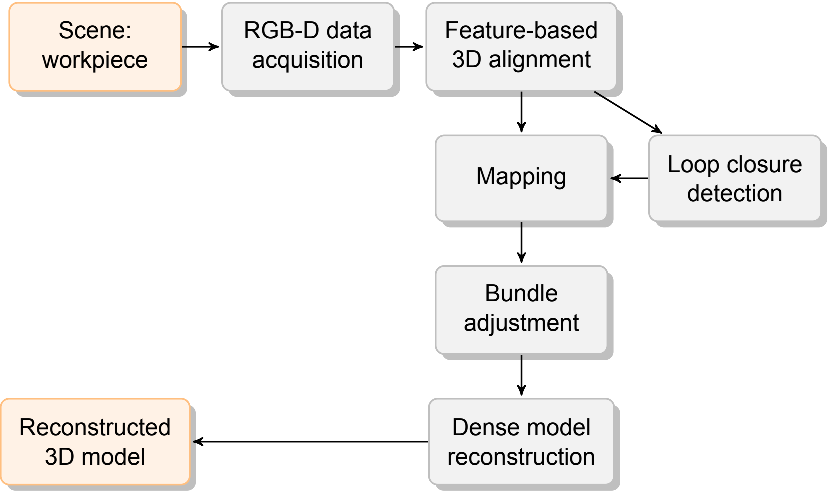


Related work: bundle adjustment

- Bundle adjustment (BA): Adjust light rays from landmarks into cameras
- Full bundle adjustment [Triggs et al., 2000]
  - Full graph of camera poses, landmarks and observations
  - Non-linear Least Squares (NLS) $\rightarrow$ Levenberg-Marquardt
  - High computational complexity
- Pose-graph optimization [Endres et al., 2012]
  - Only camera poses and pose-pose-connections
  - Efficient, but approximation per se
- Submap-based approaches [Ni et al., 2007]
  - Partition BA problem into submaps (optimized independ.)
  - Merge submaps after global optimization
  - Approaching accuracy of full BA, but more efficient

Outline

Basic approach

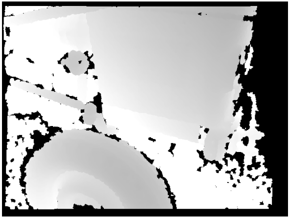


RGB-D data acquisition

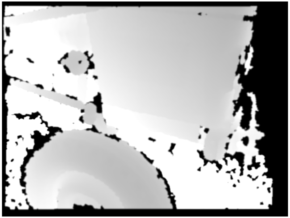
- RGB-D frame: RGB image + depth map
- Hand-held ASUS Xtion Pro Live
- Accuracy of depth measurements depend on distance to surface $\rightarrow$ between 0.70 m and 1.80 m
- Two loops around workpiece (lower and upper half)

RGB-D frame preprocessing

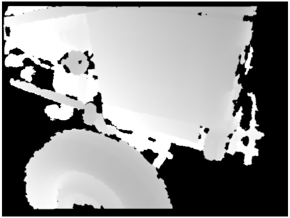
Input RGB image



Input depth map



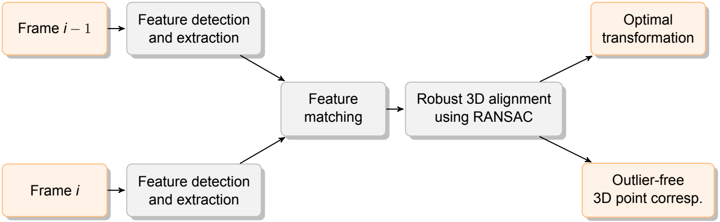
Depth map after bilateral filter



Depth map after threshold

Camera tracking

- Determine camera pose for every RGB-D frame
- Estimate relative camera motion between two frames: Feature-based 3D alignment



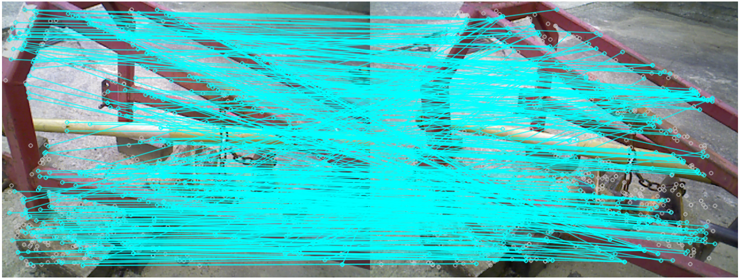
- Compute absolute poses by combining relative poses

Feature detection

- Detect distinctive feature points in RGB images
- Extract compact descriptors for the feature points
- SIFT, SiftGPU, SURF, ORB

Feature matching

- Match feature descriptors across two images
- Matching strategies: Brute-force, FLANN
- Result: 512 best 2D correspondences per frame pair
- But: many false positives

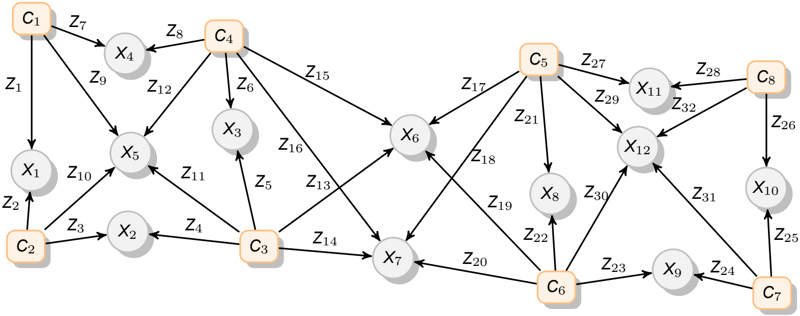


Robust 3D alignment using RANSAC

- 2D correspondences + depth $\rightarrow$ 3D correspondences
- Robust 3D alignment using RANSAC:
  - Select sample sets $\rightarrow$ determine largest consensus set $\rightarrow$ Outlier-free 3D correspondences
  - Optimal transformation

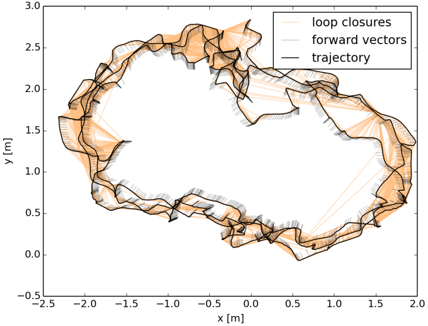
3D map representation

- SLAM graph:
  - M camera poses $C_i \in SE(3)$
  - N 3D landmarks $\mathbf{X}_j \in \mathbb{R}^3$
  - K observations $\mathbf{z}_{k_{ij}} = (u_{k_{ij}}, v_{k_{ij}}, d_{k_{ij}})^\top \in \mathbb{R}^3$
- Absolute estimates from frame-to-frame tracking



Loop closure detection

- Detect when current frame shows same scene as a previous frame
- 3D alignment with 20 uniformly sampled previous frames
  $\rightarrow$ Loop closure detected if alignment successful
  $\rightarrow$ Integrate redundancy for optimization into 3D map



Bundle adjustment using 3D alignment error

- Reduce global drift in map $\rightarrow$ bundle adjustment
- Full 2D bundle adjustment:
  - Measurement $\bar{\mathbf{z}}_{k_{ij}} = (u_{k_{ij}}, v_{k_{ij}})^\top \in \mathbb{R}^2$
  - Minimization of 2D reprojection error (w.r.t. C_{i_k} and $\mathbf{X}_{j_k}$):

$$\sum_{k=1}^{K} ||h_k(C_{i_k}, \mathbf{X}_{j_k}) - \bar{\mathbf{z}}_k||^2 \tag{1}$$

- Full 3D bundle adjustment: integrate depth constraints
  - Measurement $\mathbf{Z}_k = \rho(u_{k_{ij}}, v_{k_{ij}}, d_{k_{ij}}) \in \mathbb{R}^3$
  - Minimization of 3D alignment error (w.r.t. C_{i_k} and $\mathbf{X}_{j_k}$):

$$\sum_{k=1}^{K} ||\hat{h}_k(C_{i_k}, \mathbf{X}_{j_k}) - \mathbf{Z}_k||^2 \tag{2}$$

- Non-linear least squares optimization: Solution using sparse Levenberg-Marquardt (Ceres Solver & CXSparse)

Submap-based bundle adjustment

- Disadvantages of full BA:
  - High computational complexity
  - Inefficient for increasing amount of data
- Solution: out-of-core techniques
  - Process only portion of a large problem at once
  - Combine results from subparts

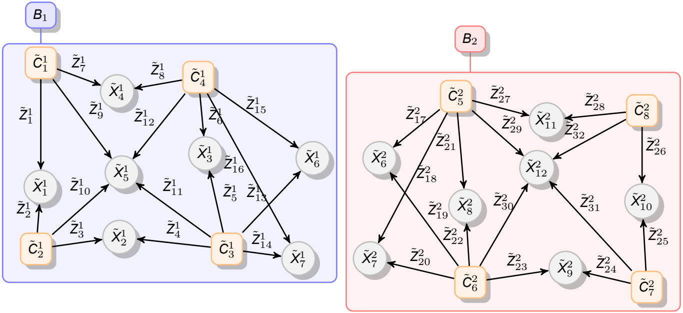
$\rightarrow$ Maintain accuracy, improve efficiency

- Submap-based BA approach:
  1. Partition SLAM graph into several submaps
  2. Optimize each submap internally
  3. Align submaps globally
  4. Optimize each submap internally with fixed separators
- Minimizations in all stages use 3D alignment error

Graph partitioning into submaps

- L submaps of size $\tilde{M} = M/L$ (no advanced graph partitioning)
- Assign base nodes B_l $(l \in 1 \dots L)$ to submaps
- Initialize base nodes: $B_l = C_{(l-1)\tilde{M}+1}$
- Express submap contents relative to base node:

$$\tilde{C}^l_i = \mathfrak{T}^{-1}(B_l, C_i), \qquad \tilde{\mathbf{X}}^l_j = \mathcal{T}^{-1}(B_l, \mathbf{X}_j), \qquad \tilde{\mathbf{Z}}^l_{k_{ij}} = \mathbf{Z}_{k_{ij}}$$



Submap optimization

- Optimize submaps independently:

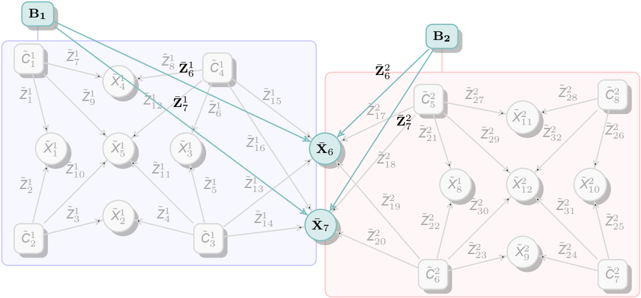
$$\sum_{k=1}^{K_l} ||\hat{h}_k(\tilde{C}^l_{i_k}, \tilde{\mathbf{X}}^l_{j_k}) - \tilde{\mathbf{Z}}^l_k||^2 \tag{3}$$

$\rightarrow \tilde{C}^l_i$ and $\tilde{\mathbf{X}}^l_j$ in all submaps optimal relative to B_l

- Landmarks connected to another submap:
  Separator landmarks: $\bar{\mathbf{X}}^l_j = \mathcal{T}(B_l, \tilde{\mathbf{X}}^l_j)$
- Locations of $\bar{\mathbf{X}}^l_j$ relative to B_l: inter-measurements $\bar{\mathbf{Z}}^l_k = \tilde{\mathbf{X}}^l_{j_k}$

Global submaps alignment

- Optimization graph: base nodes and separator landmarks as vertices, inter-measurements as edges

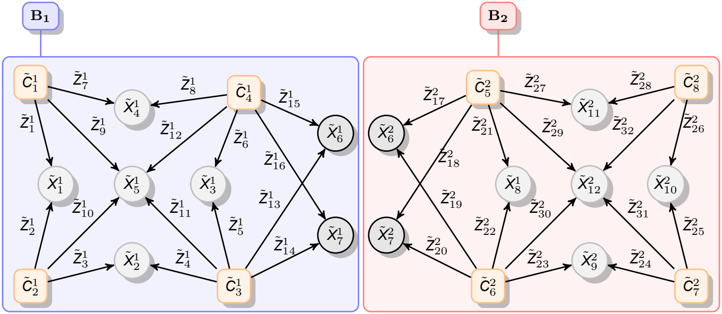


- Eliminate global drift by moving the base nodes
- Optimization for global alignment (w.r.t. B_l, $\bar{\mathbf{X}}_j$):

$$\sum_{k=1}^{\bar{K}} ||\hat{h}_k(B_{l_k}, \bar{\mathbf{X}}_{j_k}) - \bar{\mathbf{Z}}^l_k||^2 \qquad (4)$$

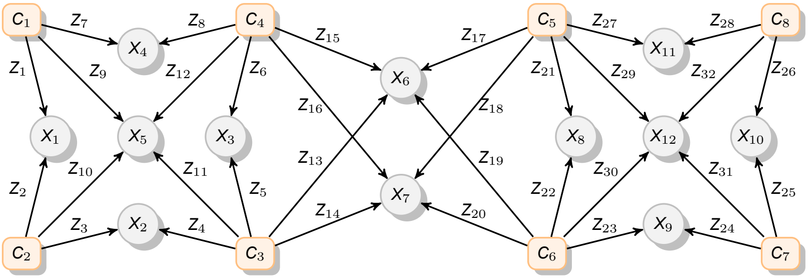
Internal submap update

- Base nodes and separator landmarks globally optimal
- Update separator landmarks in submaps: $\tilde{\mathbf{X}}_k^l = \mathcal{T}^{-1}(B_l, \bar{\mathbf{X}}_k^l)$
- Set separator landmarks fixed
- Optimize each submap independently (see stage 1)



Final optimized SLAM graph

- Final SLAM graph after submap-based BA:

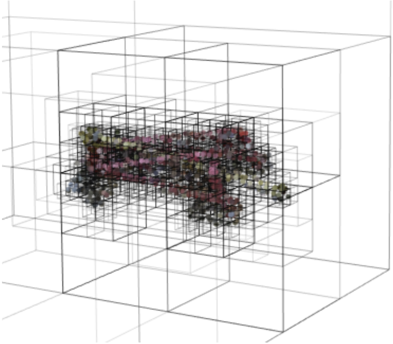


- Refined absolute camera poses and landmark locations:

$$C_i = \mathfrak{T}(B_l, \tilde{C}_i^l) \quad \text{and} \quad \mathbf{X}_j = \mathcal{T}(B_l, \tilde{\mathbf{X}}_j^l) \tag{5}$$

Dense 3D model representation

- Frame $\rightarrow$ 3D point cloud
  $\rightarrow$ Transformed using C_i
- Tree-based volumetric representation: Octree
- Voxels: occupancy, color, frames visible
- Integration using recursive subdivision



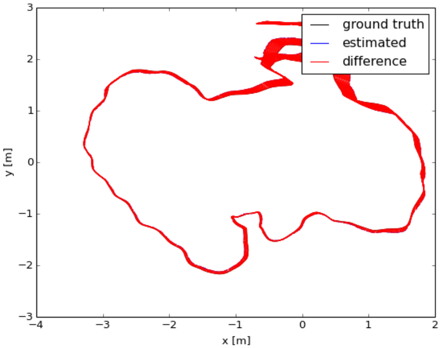
- Post-processing: remove voxels seen in < 5 frames
- Occupied octree leaves $\rightarrow$ colored 3D point cloud
- Adv.: extensible volume and limited memory consumption

Outline

1. Introduction
2. Related Work
3. 3D Reconstruction System
   - RGB-D data acquisition
   - Feature-based 3D alignment
   - Mapping
   - Out-of-core bundle adjustment
   - Dense 3D model representation
4. Evaluation and Experimental Results
   - Performance evaluation
   - 3D workpiece reconstruction
5. Conclusion and Future Work

Quantitative evaluation

- TUM RGB-D benchmark [Sturm et al., 2011]: Selected subset of 10 sequences
- Measurement of Absolute Trajectory Error (ATE) between estimated and ground-truth camera trajectory



Evaluation: 3D reconstruction system

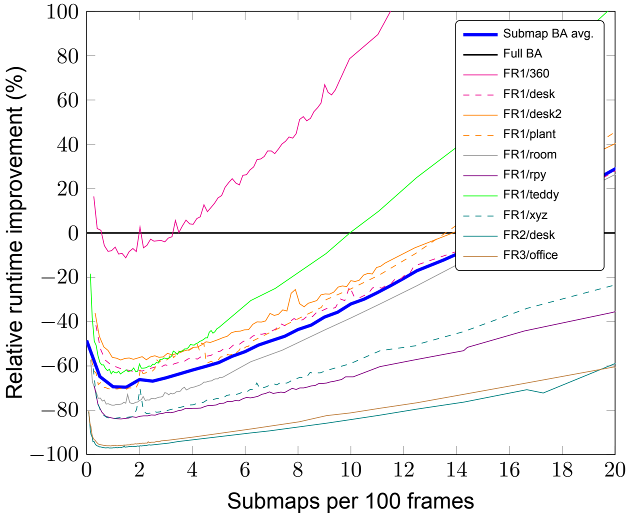
- Feature detectors:

	ORB	SIFT	SiftGPU	SURF
ATE [m]	0.158	0.145	**0.129**	0.162
Runtime [s]	**0.0091**	0.1012	0.0361	0.1676

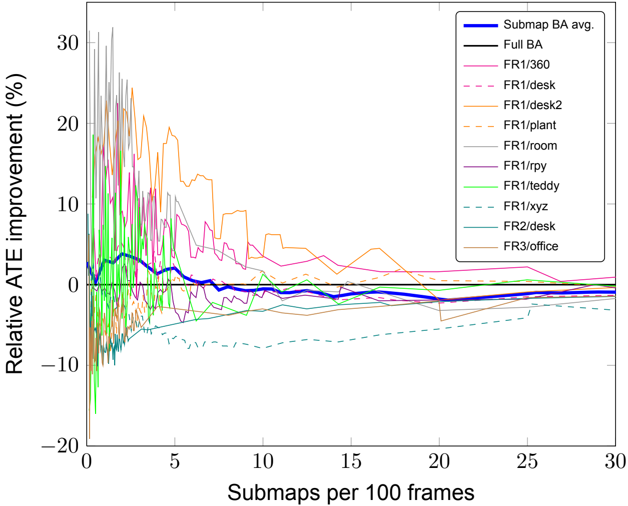
  - ORB: fastest feature detector, but increased drift
  - SiftGPU: best combination of speed and accuracy

- Average runtime of 0.5724 s per frame ($\sim$ 2 Hz):
  - Preprocessing: 0.0237 s
  - Feature detection: 0.0361 s
  - Feature matching: 0.3206 s
  - Pose estimation: 0.1919 s

Evaluation: bundle adjustment runtime



Evaluation: bundle adjustment accuracy



Evaluation: absolute results and comparison

- Best tradeoff between efficiency and accuracy: $\sim$ 10 submaps per 100 frames (i.e. $L \sim 0.10\,M$)

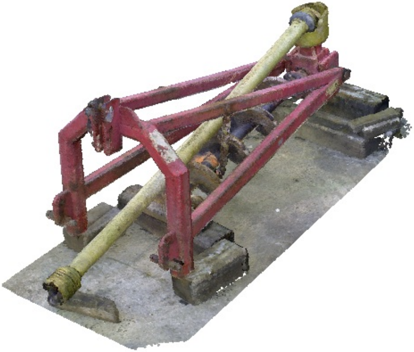
Sequence	No BA	Full 2D	Full 3D		Submap-based					RGB-D SLAM
	ATE	ATE	ATE	time	submaps	ATE	$\pm(\%)$	time	$\pm(\%)$	ATE
FR1/360	0.108	0.099	0.077	12.66	74	0.079	+3.6	22.62	+78.6	0.079
FR1/desk	0.047	0.021	0.022	28.97	57	0.022	-1.5	21.96	-24.2	0.023
FR1/desk2	0.098	0.044	0.030	27.23	62	0.031	+3.4	21.36	-21.5	0.043
FR1/plant	0.048	0.023	0.042	66.27	112	0.043	+1.7	49.36	-25.5	0.091
FR1/room	0.275	0.228	0.085	125.46	135	0.086	+1.7	77.30	-38.4	0.084
FR1/rpy	0.046	0.058	0.027	67.56	69	0.027	-1.6	23.69	-64.9	0.026
FR1/teddy	0.277	0.060	0.056	67.88	140	0.057	+1.3	68.06	+0.3	0.076
FR1/xyz	0.015	0.013	0.013	96.87	79	0.013	-7.9	39.72	-59.0	0.014
FR2/desk	0.201	0.080	0.079	2355.26	289	0.076	-3.3	372.20	-84.2	-
FR3/office	0.176	0.039	0.036	1290.24	248	0.035	-3.0	242.88	-81.2	-
average	**0.129**	**0.066**	**0.047**			**0.047**	**-0.5**		**-32.0**	**0.054**

- Submap-based BA approaches accuracy of full 3D BA, but is more efficient
- Our method outperforms RGB-D SLAM regarding accuracy

Soil auger




- Dimensions:
  0.85 m x 1.16 m x 2.80 m
- Map:
  - 2349 camera poses
  - 156974 landmarks
  - 1086734 observations
- Submap-based BA with
  230 submaps in 195 s



Lawn tractor

- Dimensions:
  0.94 m x 1.25 m x 2.23 m
- Map:
  - 2167 camera poses
  - 179616 landmarks
  - 1124111 observations
- Submap-based BA with 216 submaps in 184 s

Farm tractor



- Dimensions:
  0.99 m x 1.30 m x 3.19 m
- Map:
  - 2087 camera poses
  - 137657 landmarks
  - 1063204 observations
- Submap-based BA with
  208 submaps in 178 s

Outline

Conclusion and Future Work

- RGB-D-based 3D reconstruction system for 3D workpiece reconstruction
- Out-of-core BA: 3D alignment error + submaps
- Quantitative evaluation:
  - 3D reconstruction system: frame rate of 2 Hz
  - Submap-based vs. full BA:
    - Avg. runtime improvement of 32% (large datasets: 80%)
    - ATE approaches full BA and outperforms RGB-D SLAM
- Workpieces: soil auger, lawn tractor and farm tractor

- Future Work:
  - Improve efficiency (GPU programming, PROSAC [Chum and Matas, 2005], FABMAP [Cummins and Newman, 2008])
  - Mesh-based model representation; probabilistic approach
  - Submap-based BA: fully hierarchical tree of submaps [Ni and Dellaert, 2012]

Bibliography I

[Chum and Matas, 2005] Chum, O. and Matas, J. (2005).
Matching with PROSAC-progressive sample consensus.
In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 220--226. IEEE.

[Cummins and Newman, 2008] Cummins, M. and Newman, P. (2008).
FAB-MAP: Probabilistic localization and mapping in the space of appearance.
The International Journal of Robotics Research, 27(6):647--665.

[Endres et al., 2012] Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., and Burgard, W. (2012).
An evaluation of the RGB-D SLAM system.
In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1691--1696. IEEE.

[Izadi et al., 2011] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., et al. (2011).
KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera.
In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559--568. ACM.

[Ni and Dellaert, 2012] Ni, K. and Dellaert, F. (2012).
HyperSfM.
In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 144--151. IEEE.

[Ni et al., 2007] Ni, K., Steedly, D., and Dellaert, F. (2007).
Tectonic SAM: Exact, out-of-core, submap-based SLAM.
In *Robotics and Automation, 2007 IEEE International Conference on*, pages 1678--1685. IEEE.

Bibliography II

[Sturm et al., 2011] Sturm, J., Magnenat, S., Engelhard, N., Pomerleau, F., Colas, F., Burgard, W., Cremers, D., and Siegwart, R. (2011).
Towards a benchmark for RGB-D SLAM evaluation.
In *Proc. of the RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf.(RSS)*, pages 1--3.

[Triggs et al., 2000] Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon, A. (2000).
Bundle adjustment - a modern synthesis.
Vision algorithms: theory and practice, pages 153--177.