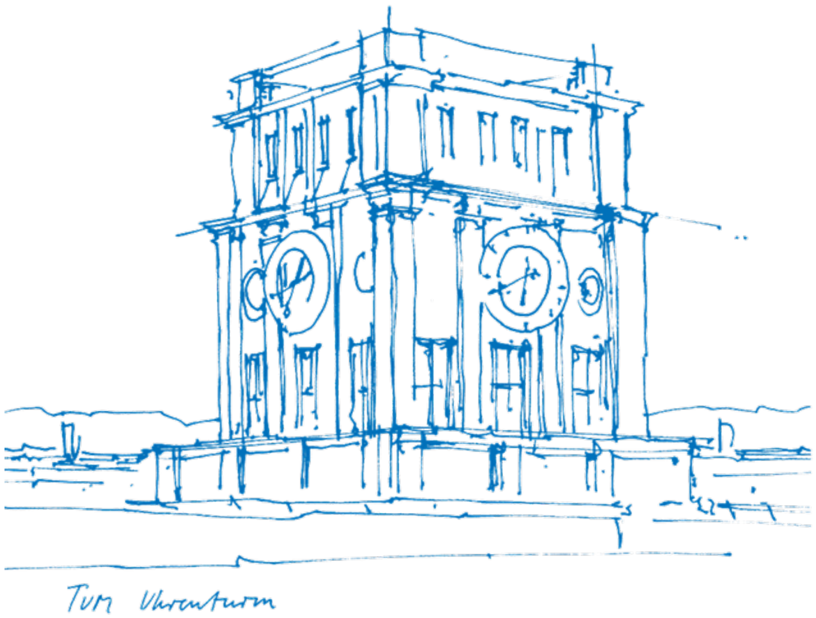
Seminar: The Evolution of Motion Estimation and Real-time 3D Reconstruction

Christiane Sommer, Lukas Köstler

Computer Vision Group

Technical University of Munich



How can I access these slides?

- Option 1 (preferred): seminar web page
  - `https://vision.in.tum.de/teaching/ws2020/seminar_realtime3d`
  - Password for material page: `ws20-realtime3d`
  - Material page will go online after this pre-meeting
- Option 2: contact organizers
  - `realtime3d-ws20@vision.in.tum.de`
  - **Only use this option if you forgot the password**

Outline

- General Information
  - About the Seminar
  - Registration
- Possible Papers
  - Depth Sensors
  - Monocular Cameras
  - Deep-learning-based Scene Representation
- Questions

Outline

- General Information
  - About the Seminar
  - Registration
- Possible Papers
  - Depth Sensors
  - Monocular Cameras
  - Deep-learning-based Scene Representation
- Questions

How is the seminar organized?

- Seminar meetings: Talks and discussion
  - Day: Tuesday, approximately every second week (TBA)
  - Time: 14:00–16:00
  - Location: MI 02.05.014 or **virtual** (TBA)
  - In case of special circumstances please let us know and we will find a solution
  - Each session will consist of two talks which are held in English
  - **Attendance is mandatory!**
- Talk preparation / contact with supervisor
  - One month before talk: meet supervisor for questions (optional, but recommended)
  - Two weeks before talk: meet supervisor to go through slides (optional, but recommended)
  - One week before talk: send slides to your supervisor (mandatory)
  - Two weeks after talk: submit your report via email (mandatory)

What about the presentation?

- General set-up:
  - Duration: 20–25 minutes talk + 10–15 minutes discussion
  - Make sure to finish on time - not too early and not too late!
  - Rule of thumb: 1–2 minutes per slide $\rightarrow$ 10–20 slides
  - Do not put too much information on the slides!
- Recommended structure (talk):
  - Introduction
  - Overview / Outline
  - Method description
  - Experiments and results
  - Personal comments
  - Summary

What about the discussion after each talk?

- Discussion afterwards **will** influence your grade
- Ask questions!
- There are **no** stupid questions!

What about the final report?

- General set-up:
  - Use LaTeX template provided on web page
  - Length: 3-4 pages
  - Send final report as pdf by email to `realtime3d-ws20@vision.in.tum.de`
  - Submission deadline: **Two weeks after talk**
- Recommended structure (main text only):
  - Introduction
  - Method description
  - Experiments and results
  - Discussion of results
  - Summary

Outline

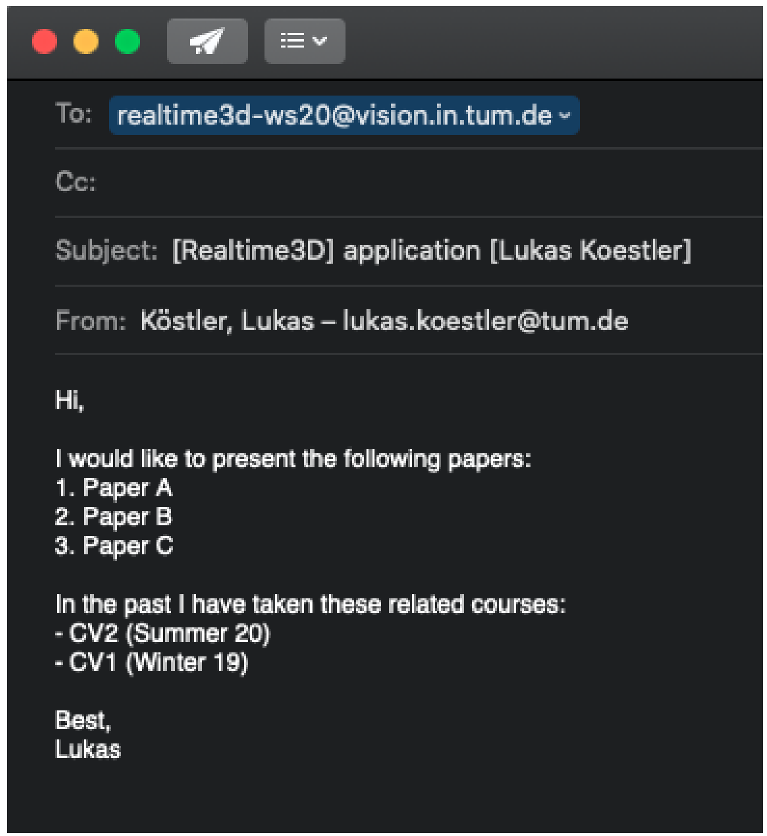
- General Information
  - About the Seminar
  - Registration
- Possible Papers
  - Depth Sensors
  - Monocular Cameras
  - Deep-learning-based Scene Representation
- Questions

How do you register for the seminar?

- Step 1: Official registration via TUM matching system
  - Go to `https://matching.in.tum.de`
  - Register for seminar with the title *The Evolution of Motion Estimation and Real-time 3D Reconstruction*
- Step 2: Personal registration via email
  - In the list of papers on the web page, select your three favorites
  - Write an email ranking these three favorites to the seminar email address
  - Email subject: “[Realtime3D] application [your name]”
  - Include information about related lectures / courses you have taken so far
  - We do **not** need your CV or a motivation letter!
  - Registrations without email / emails with missing information will be ignored!
- Deadline for both registrations: July 21st, 2020

How do you register for the seminar?

Example registration email:



How do we select candidates and assign papers?

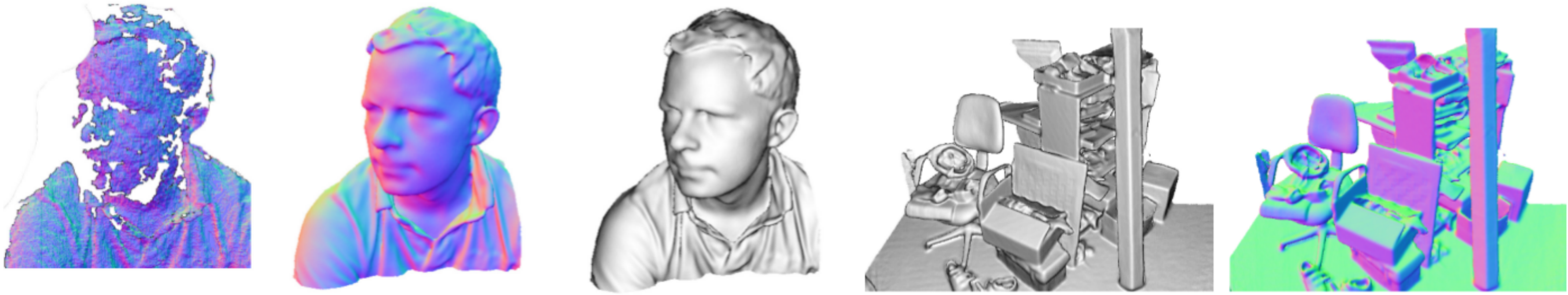
- Candidate selection
  - Only students registered in the matching system **AND** emails containing all required information will be considered
  - Among students meeting the formal criteria, selection will be random
  - You will get notified by the matching system about the decision (July 30th, 2020)
- Paper assignment
  - Papers are assigned after the participant list is finalized
  - We give our best to accommodate your preference list in the assignment

Outline

- General Information
  - About the Seminar
  - Registration
- Possible Papers
  - Depth Sensors
  - Monocular Cameras
  - Deep-learning-based Scene Representation
- Questions

KinectFusion: Real-Time Dense Surface Mapping and Tracking

Newcombe, Izadi, Hilliges, Molyneaux, Kim, Davison, Kohli, Shotton, Hodges, Fitzgibbon 2011



- First paper to generate dense 3D models in real-time using depth sensor and GPU
- Highly cited, impactful, baseline method for 3D reconstruction using RGB-D cameras

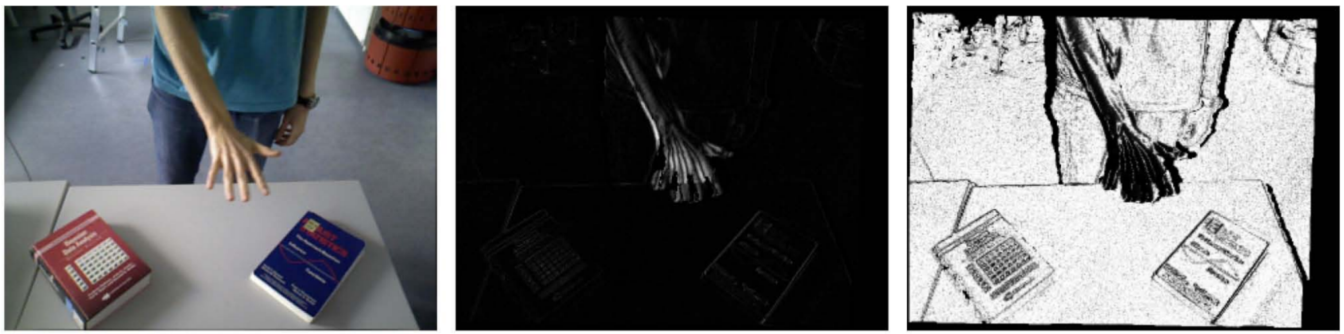
Real-Time Camera Tracking and 3D Reconstruction Using Signed Distance Functions

Bylow et al. 2013, RSS

- Nice introduction to SDFs using RGB-D cameras

Robust Odometry Estimation for RGB-D Cameras

Kerl, Sturm, Cremers 2013



- Odometry method that minimizes photometric cost using depth measurements
- Improved weighting function for robustness in the presence of outliers

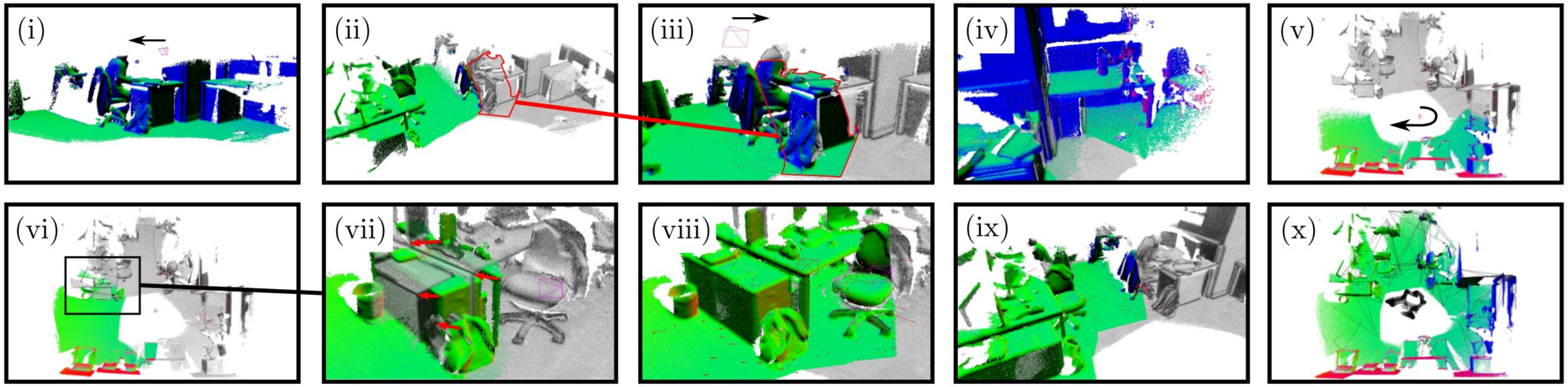
Real-time 3D Reconstruction at Scale using Voxel Hashing

Nießner, Zollhöfer, Izadi, Stamminger 2013

- Uses hashing to store TSDF grid efficiently
- Used within many state-of-the-art voxel-based reconstruction methods

ElasticFusion: Dense SLAM Without A Pose Graph

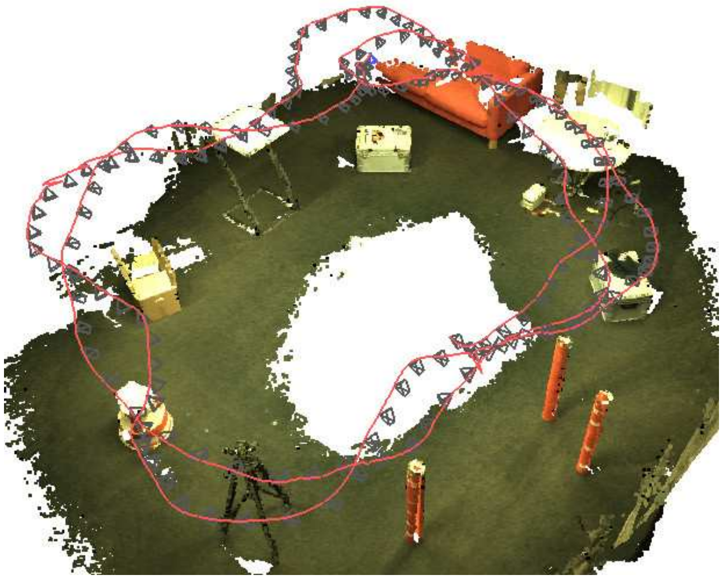
Whelan, Leutenegger, Salas-Moreno, Glocker, Davison 2015



- Uses surfels instead of a TSDF to represent the 3D model
- First method to update the surface in online manner

BAD SLAM: Bundle Adjusted Direct RGB-D SLAM

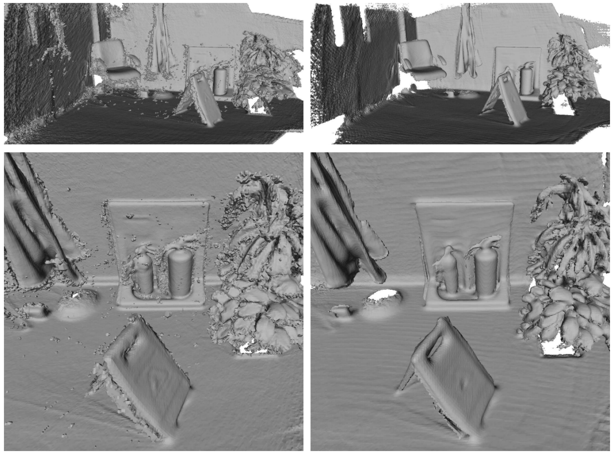
Schöps, Sattler, Pollefeys 2019



- Published on CVPR 2019
- Perform bundle adjustment on surfels to get a high quality pose

RoutedFusion: Learning Real-time Depth Map Fusion

Weder, Schönberger, Pollefeys, Oswald 2020



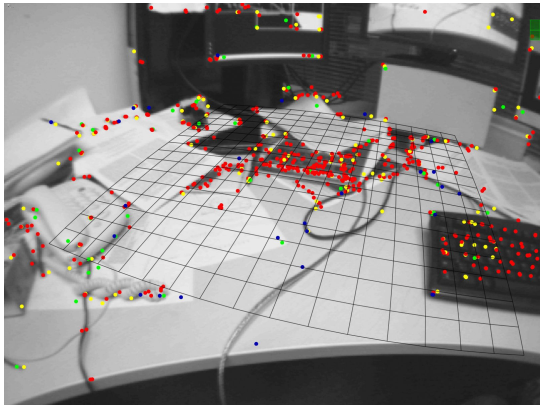
- Published on CVPR 2020
- Learning-based approach to depth map fusion

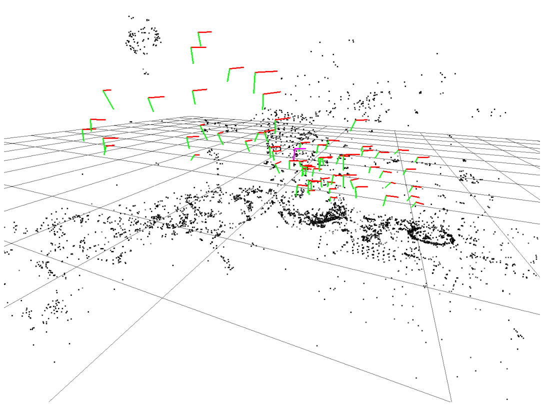
Outline

- General Information
  - About the Seminar
  - Registration
- Possible Papers
  - Depth Sensors
  - Monocular Cameras
  - Deep-learning-based Scene Representation
- Questions

PTAM: Parallel Tracking and Mapping

Klein, Murray 2007

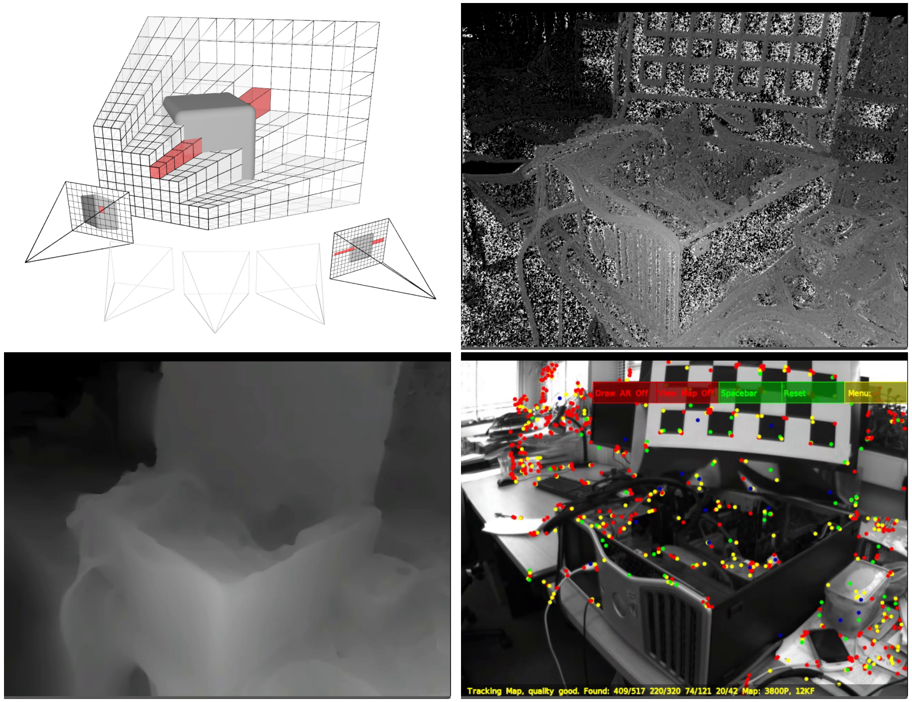




- One of the first systems capable of estimating both pose and geometry in real-time for handheld cameras
- Simple AR applications

DTAM: Dense Tracking and Mapping in Real-Time

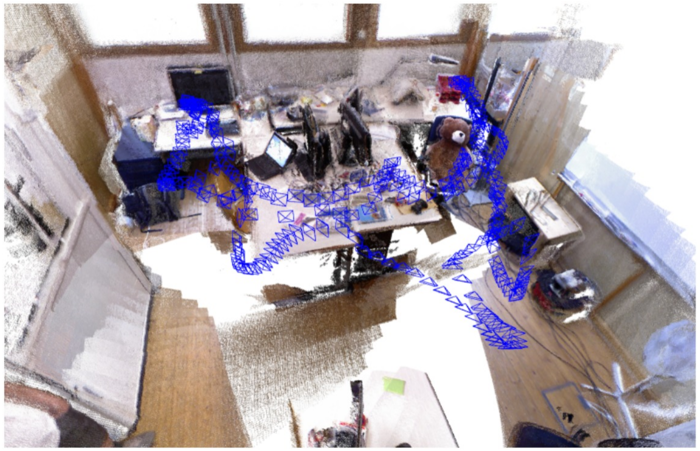
Newcombe, Lovegrove, Davison 2011



- One of the first monocular systems to create dense 3D models

ORB-SLAM: a Versatile and Accurate Monocular SLAM System

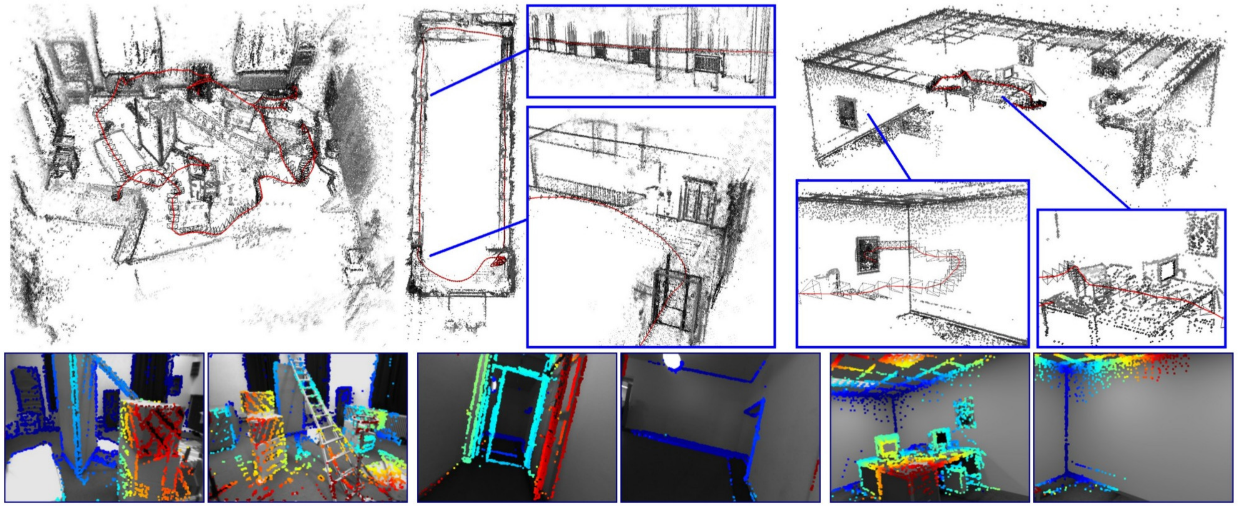
Mur-Artal, Montiel, Tardós 2015



- Use all depth and color data to obtain consistent mapping

Direct Sparse Odometry

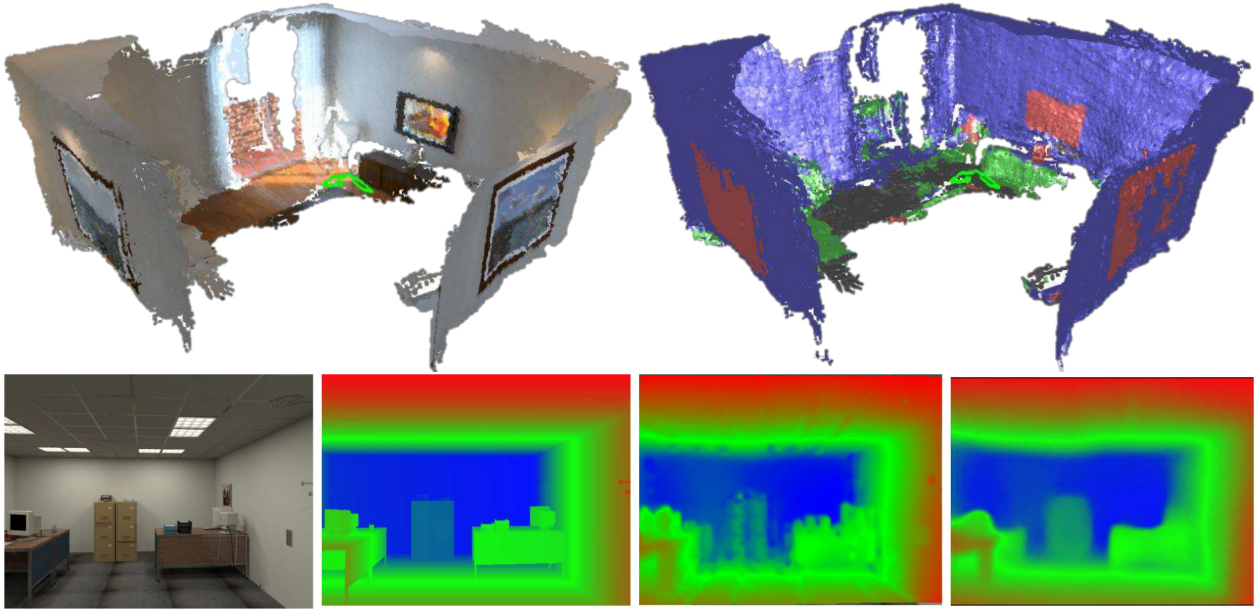
Engel, Koltun, Cremers 2016



- Large-scale odometry
- Does not rely on keypoint detections

CNN-SLAM

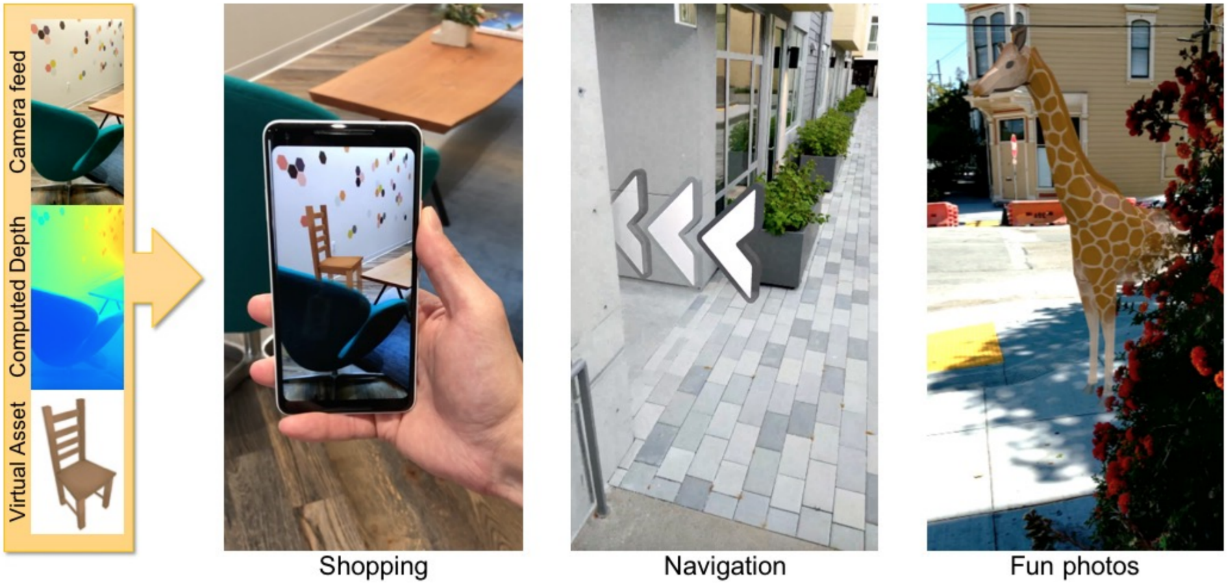
Tateno et al. 2017



- Dense monocular SLAM
- Use depth map predicted from CNN

Depth from Motion for Smartphone AR

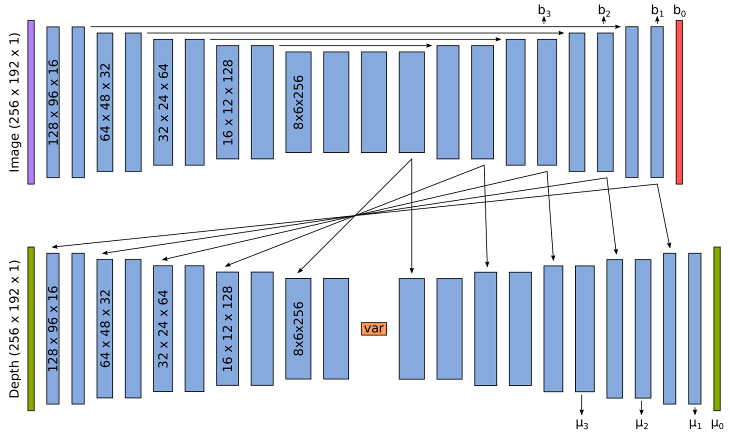
Valentin et al. 2018



- Uses poses predicted by Visual-Inertial Odometry in a Multi-View-Stereo pipeline to predict depth
- More engineering focused work that shows impressive results on CPU

CodeSLAM

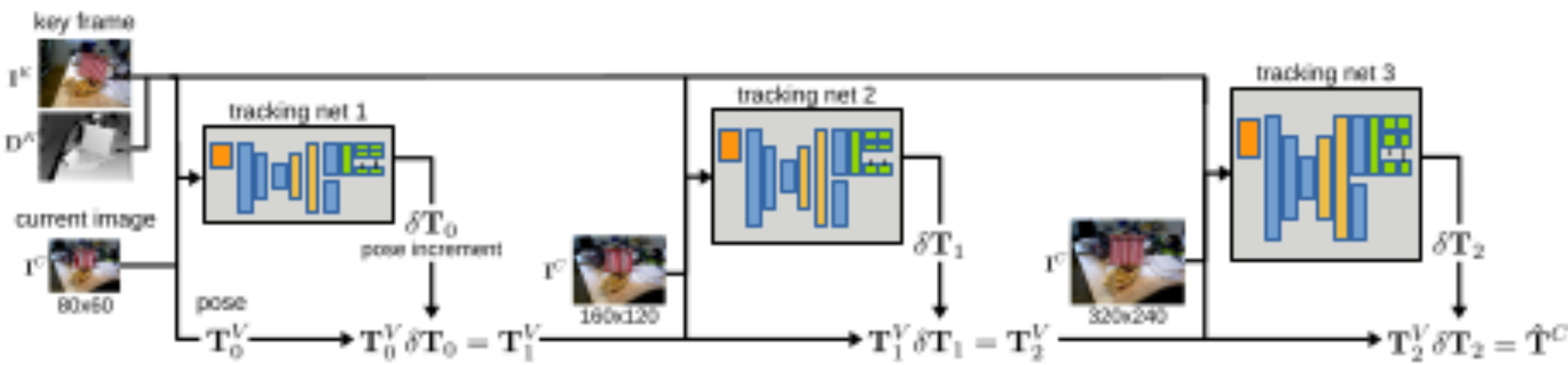
Bloesch et al. 2018



- Learning a compact, optimisable representation of the scene geometry

DeepTAM: Deep Tracking and Mapping

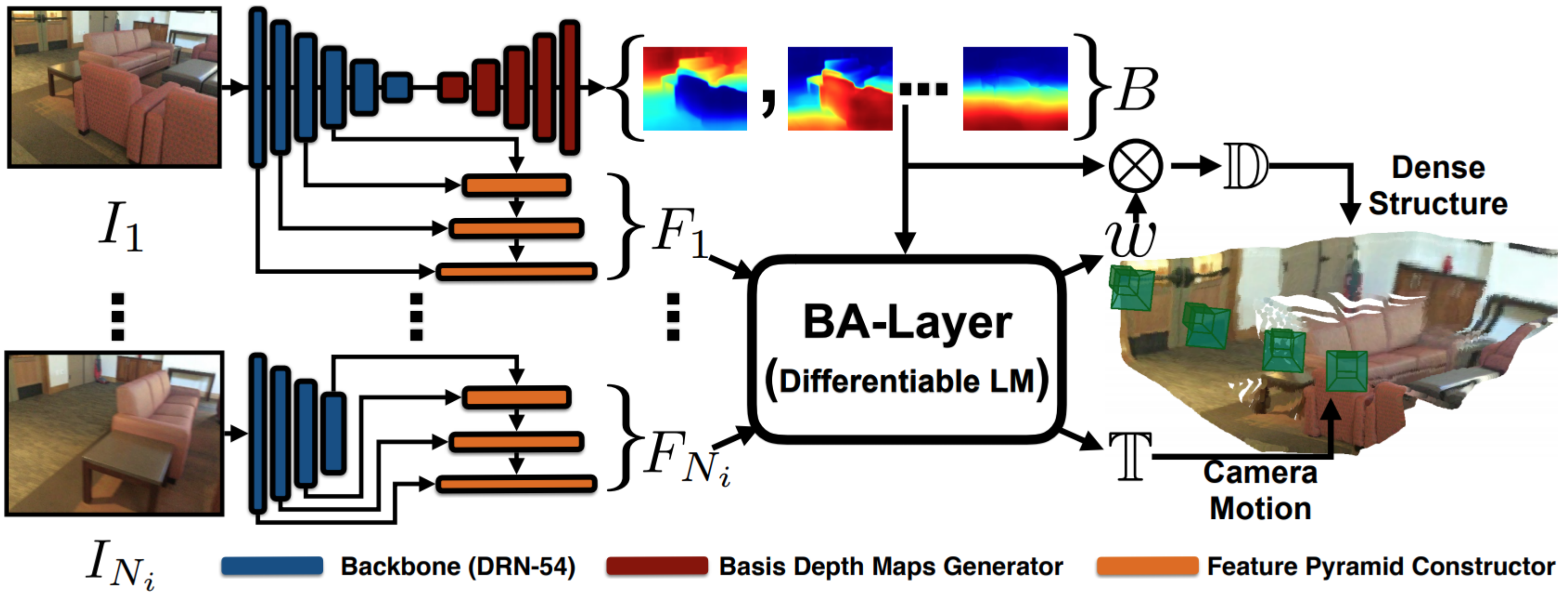
Zhou, Ummenhofer, Brox 2018



- Learn a network to predict the pose and generate depth images

BA-Net: Dense Bundle Adjustment Networks

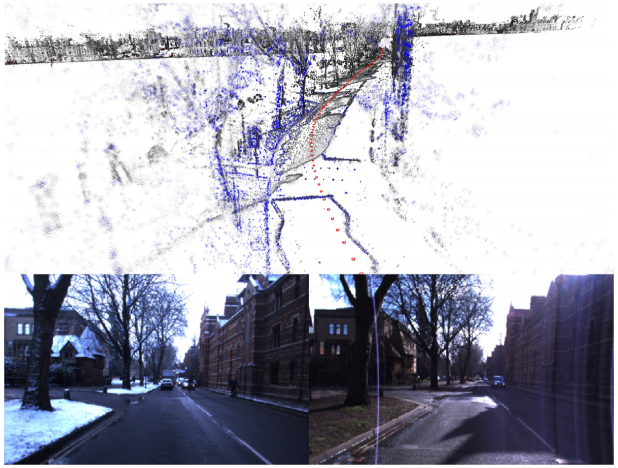
Tang, Tan 2019



- Use feature-metric Bundle Adjustment within a differentiable deep-learning pipeline
- Allows the end-to-end training of NNs for SLAM

GN-Net: The Gauss-Newton Loss for Multi-Weather Relocalization

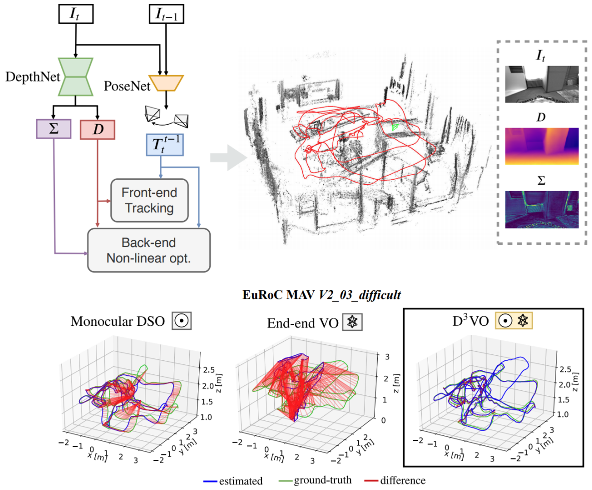
von Stumberg, Wenzel, Khan, Cremers 2020



- Use feature-metric Bundle Adjustment for multi-weather relocalization
- Propose the Gauss-Newton loss to train NNs which generate feature maps that are suitable for direct image alignment

D3VO: Deep Depth, Deep Pose and Deep Uncertainty for Monocular Visual Odometry

Yang, von Stumberg, Wang, Cremers 2020



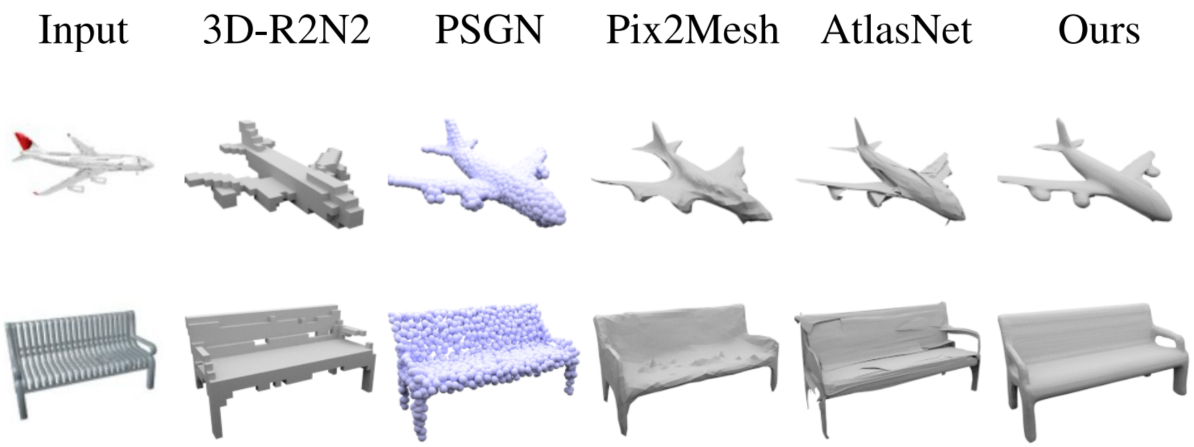
- Monocular visual odometry framework that uses deep-learning on three levels: deep depth, pose and uncertainty estimation
- Shows impressive performance improvements in comparison to traditional methods (DSO, ORB)

Outline

- General Information
  - About the Seminar
  - Registration
- Possible Papers
  - Depth Sensors
  - Monocular Cameras
  - Deep-learning-based Scene Representation
- Questions

Occupancy Networks: Learning 3D Reconstruction in Function Space

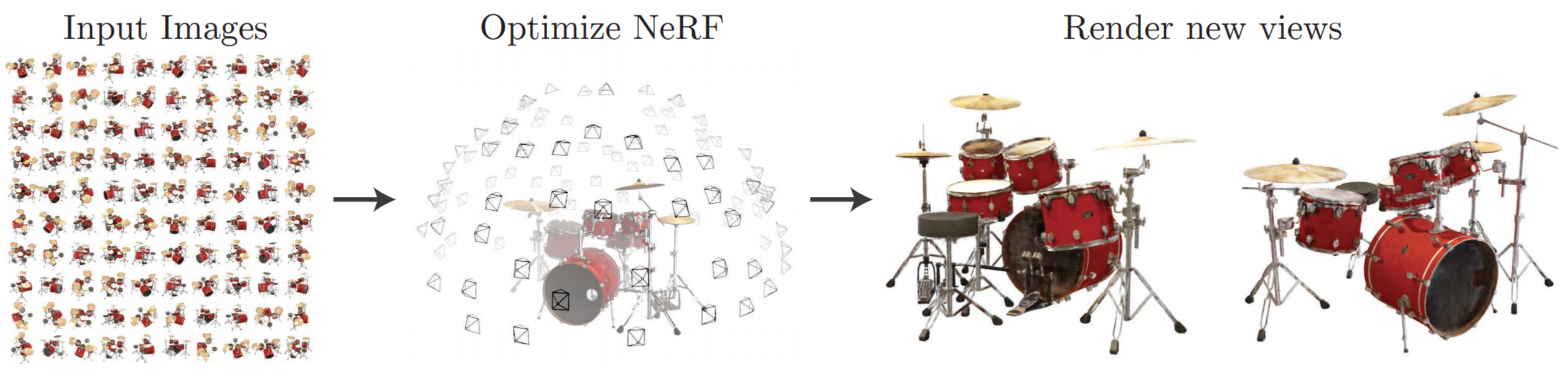
Mescheder, Oechsle, Niemeyer, Nowozin, Geiger 2019



- Train a NN to predict the occupancy probability for a point in space and extract the isosurface
- The above representation can be condition on an additional input which allows for: Single Image 3D Reconstruction, Super Resolution, etc.

NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

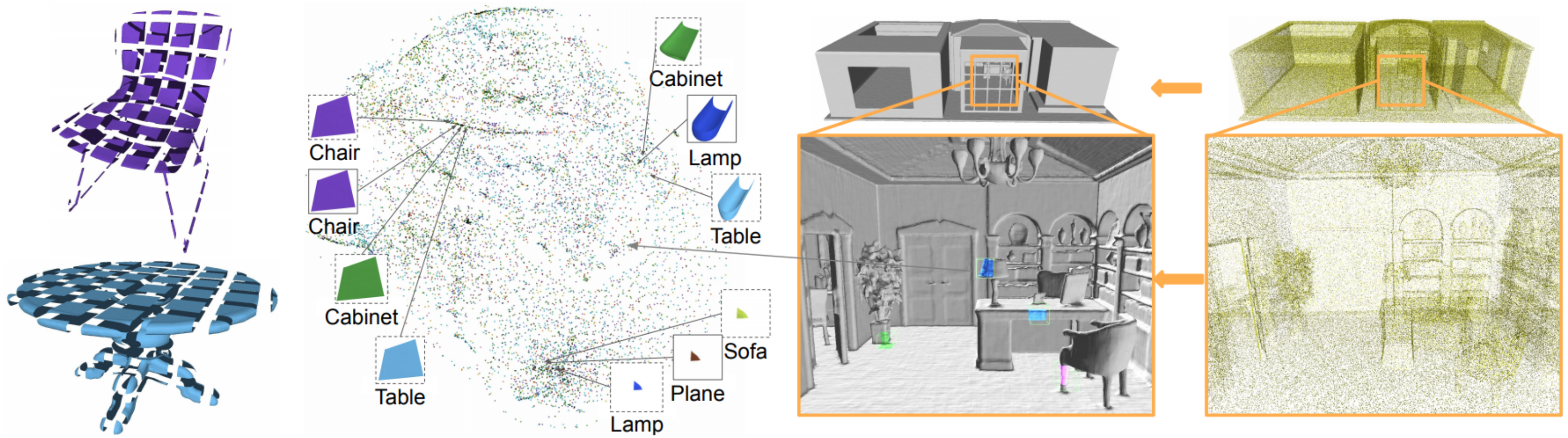
Mildenhall et al. 2020



- For a set of images of one scene train one NN that maps $(x, y, z, \theta, \phi) \mapsto (RGB\sigma)$ using volume rendering and image based losses
- The NN can be used to generate novel views and gives very impressive results

Local Implicit Grid Representations for 3D Scenes

Jiang, Sud, Makadia, Huang, Nießner, Funkhouser 2020



(a) Training parts from ShapeNet. (b) t-SNE plot of part embeddings. (c) Reconstructing entire scenes with Local Implicit Grids

- Scales implicit (neural) representation to room scale by using a grid of implicit functions
- The main idea is that 3D surfaces share geometric details at part scale (ca. 25 cm – 100 cm)

Questions?

Reminder:

- Web page: https://vision.in.tum.de/teaching/ws2020/seminar_realtime3d
- Password: ws20-realtime3d
- Contact: realtime3d-ws20@vision.in.tum.de