Robotic 3D Vision

Lecture 1: Introduction

WS 2017/18

Prof. Dr. Jörg Stückler
Computer Vision Group, TU Munich
http://vision.in.tum.de

Organization

Lecturer:

- Prof. Dr. Jörg Stückler (stueckle@in.tum.de)

Teaching Assistant:

- Rui Wang (rui.wang@in.tum.de)

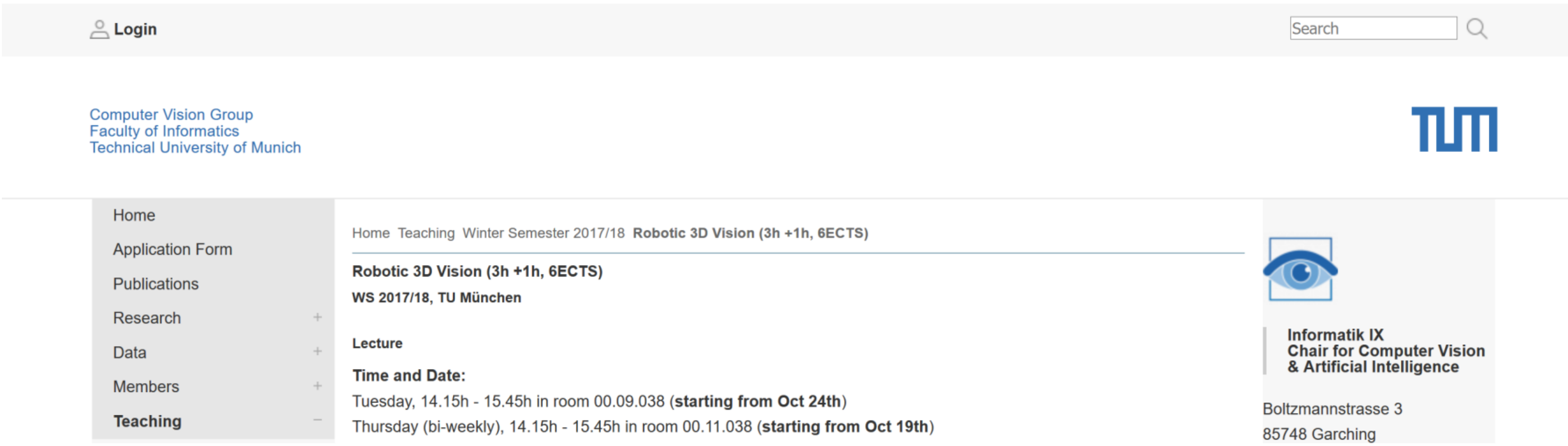
Course Webpage:

- https://vision.in.tum.de/teaching/ws2017/r3dv
- Slides will be made available on the webpage

Organization

- Structure: 3L (lecture) + 1E (exercises)
  - 6 ECTS credits
- Study programme: **M.Sc. Informatics**
- Place & Time
  - Lecture: Tue 14:15 – 15:45 00.09.038
  - Lecture/Exercises: Thu 14:15 – 16:00 00.11.038
- Exam
  - Planned as written exam
  - Date tba

Course Organization



- https://vision.in.tum.de/teaching/ws2017/r3dv
- A detailed course schedule will appear soon on the website

Exercises and Demos

- Exercises
  - Typically 1 exercise sheet every 2 weeks (theoretical and Matlab-based assignments)
  - Hands-on experience with the algorithms from the lecture
  - Send in your solutions the night before the exercise class
  - Handing in the exercises is not mandatory to take the exam
  - First exercise class: Thursday Nov. 2nd 2017, 14.15-16.00

- Teams are encouraged!
  - You can form teams of up to 3 people for the exercises
  - Each team should only turn in one solution
  - List the names and matriculation numbers of all team members in the submission
  - Each exercise will be demo'ed by a team during the exercise class

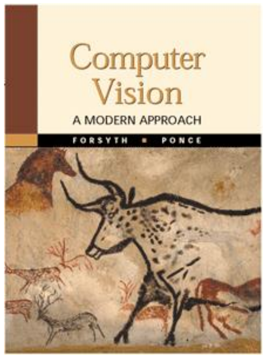
Course Requirements

- We will build on basics from previous lectures
  - Computer Vision II: Multiple View Geometry
    https://vision.in.tum.de/teaching/ss2017/mvg2017

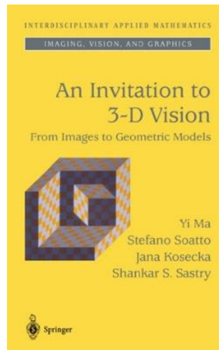
- Solid background in linear algebra and analysis

Textbooks

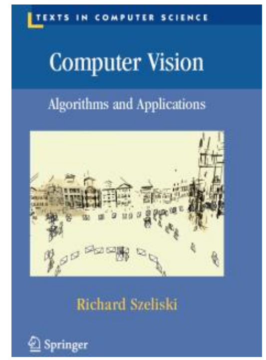
- No single textbook for the class, some basics can be found in



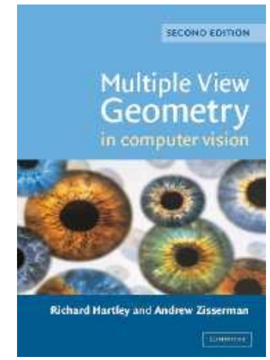
Computer Vision – A Modern Approach, D. Forsyth, J. Ponce, Prentice Hall, 2002



An Invitation to 3D Vision, Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, Springer, 2004



Computer Vision – Algorithms and Applications, R. Szeliski, Springer, 2006



Multiple View Geometry in Computer Vision, R. Hartley and A. Zisserman, Cambridge University Press, 2004

- We will also give out research papers
  - Tutorials for basic techniques
  - State-of-the-art research papers for current developments

How to Find Us

- Office:
  - TUM Math&CS Building
  - Boltzmannstr. 3, Garching, 2nd floor
  - I9, rooms 02.09.044 (Rui Wang), 02.09.059 (me)

- Office hours
  - If you have questions about the lecture, come to Rui Wang or me.
  - Our regular office hours will be announced
  - Send us an email before to confirm a time slot.

Questions are welcome!

Getting Involved

How can you get involved in scientific research during your study?

- Bachelor lab course (10 ECTS)
- Bachelor thesis (15 ECTS)
- Graduate lab course (10 ECTS)
- Interdisciplinary project (16 ECTS)
- Master thesis (30 ECTS)
- Student research assistant (10 EUR/hour, typically 10 hours/week)

Vision-based Navigation

- We also offer a practical course on Vision-based Navigation in this semester

- Participants will work on a project related to vision-based navigation for multicopters

- We still have participant slots available. If you are interested, please contact us until Friday, Oct. 20th via visnav_ws2017@vision.in.tum.de

- Further information on the course can be found at https://vision.in.tum.de/teaching/ws2017/visnav_ws2017

Robots in Complex Environments

Image credit: Amazon

Image credit: DHL

Image credit: Waymo

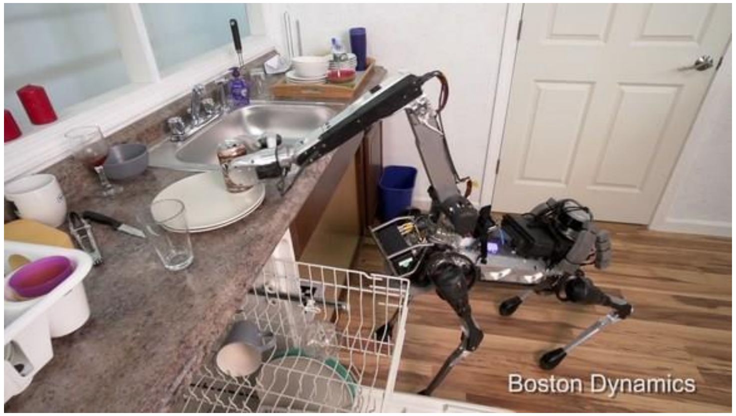


Image credit: Boston Dynamics

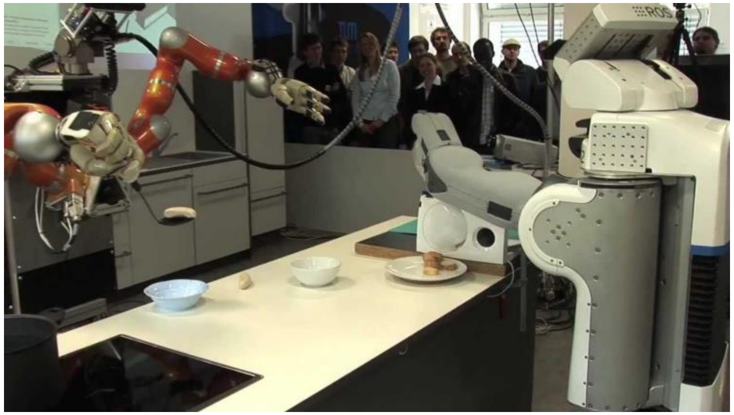


Image credit: IAS TUM / UBremen

Robotic Perception

(Usenko, von Stumberg, Pangercic, Cremers, IROS 2017)

Robotic Perception

(Stückler, Schwarz, Behnke, Frontiers 2016)

Robotic Perception

(Kappler et al., arXiv 2017)

What We Will Cover Today

- Why Vision for Robotic Perception?

- What is Robotic 3D Vision?

- Terminology of
  - Visual Odometry
  - Visual-Inertial Odometry
  - Visual Simultaneous Localization and Mapping
  - Map Representations
  - Dense vs. Sparse Reconstruction
  - Visual 3D Object Detection and Tracking
  - Indirect and Direct Methods

Sensors for Robotic Perception

Vision

+ low power consumption
+ dense 2D projection
+ appearance
+ high frame-rate
- indirect distance

Laser

+ accurate distance
- power consumption
- sparse
- low frame-rate
- scan plane

Inertial

+ linear acceleration
+ gravity
+ rotational velocity
+ high frame-rate
- noise & bias
- local

Proprioceptive

+ forward kinematics

(+ forward dynamics)

- only internal

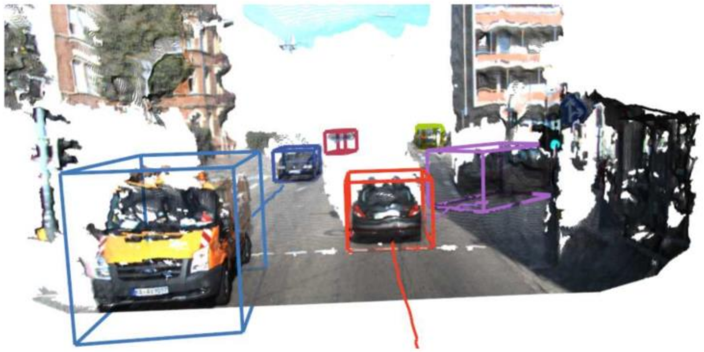
Tactile

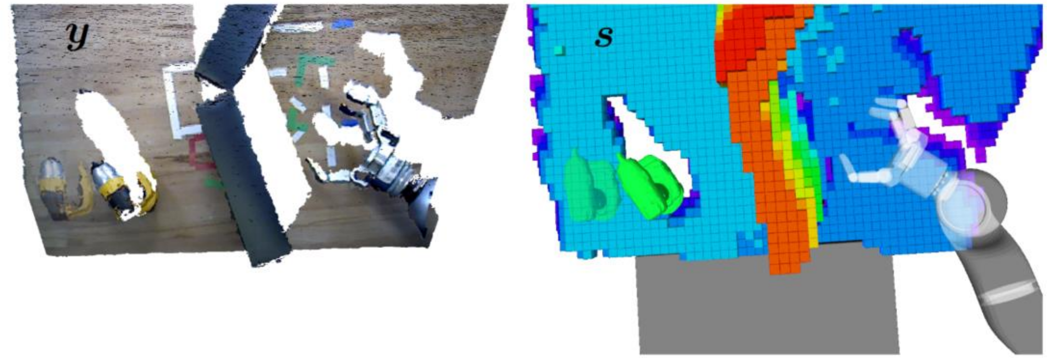
+ contact with environment

RGB-D

+ depth image
- power consumption

Robotic 3D Vision



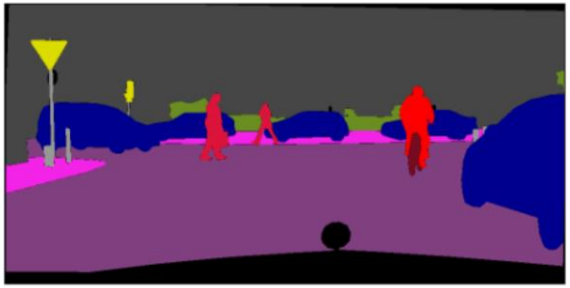
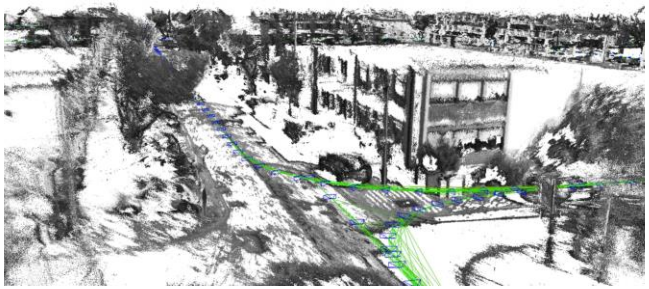


- Robots require 3D scene understanding
  - Where is the robot in the environment?
  - What is the shape (structure) of the environment?
  - Where are task-relevant objects?

- 3D Vision: 3D scene understanding from camera images

Images from: (Osep et al., ICRA 2016), (Kappler et al., arXiv 2017)

Why Vision?

Vision provides robots with rich information about the world

- Dense 2D measurements of the 3D world, in contrast to, for example, laser scanners or ultrasonic range scanners
- RGB/grayscale measurements of the appearance of objects available to detect and recognize objects
- Range (third dimension) assessable by stereo
- Lightweight and low power consumption (passive cameras)

Images from: (Pohlen et al., CVPR 2017), (Engel, Stückler, Cremers, IROS 2015)

Types of Cameras

Monocular camera

- Structure from motion (chicken-and-egg problem)
- Scale ambiguity

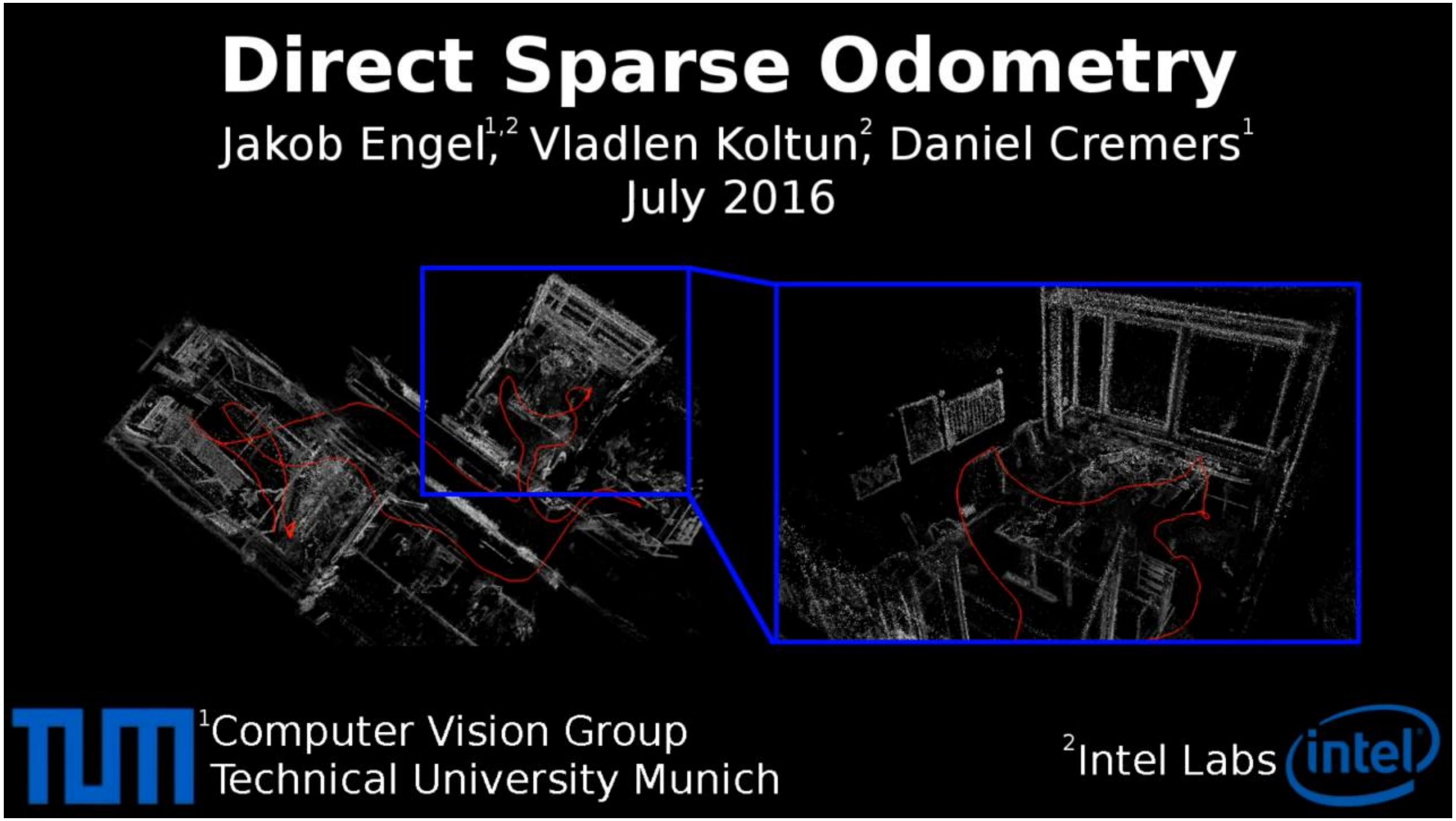
Stereo camera

- Depth from stereo in fixed configuration
- Scale observable
- Fixed baseline

RGB-D camera

- Directly measures per-pixel depth
- Active sensing

Visual Odometry



How does the robot move?

What is Visual Odometry?

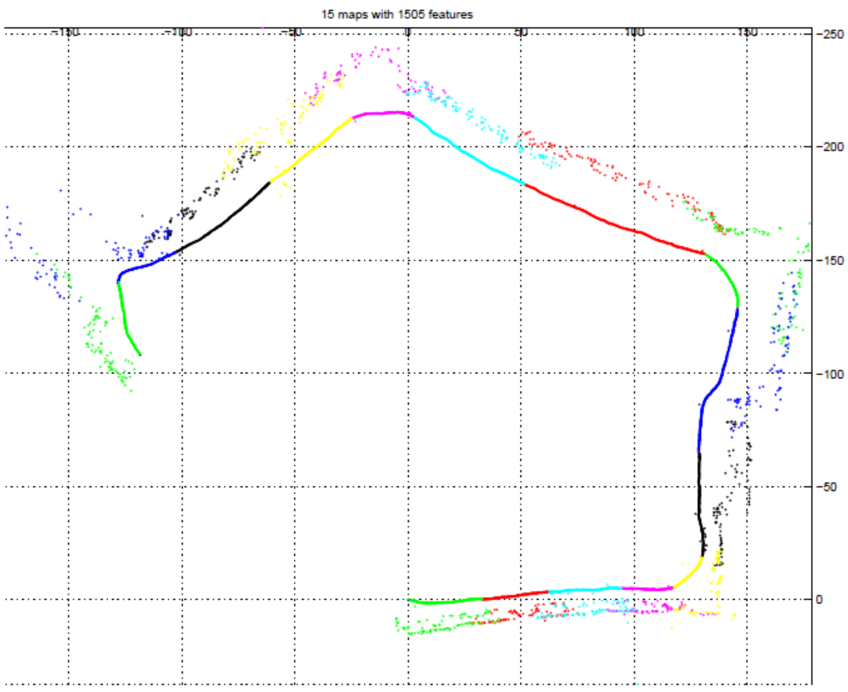
Visual odometry (VO)…

- … is a variant of tracking
  - Track the current pose, i.e. position and orientation, of the camera with respect to the environment from its images
  - Only considers a limited set of recent images for real-time constraints
- … involves a data association problem
  - Motion is estimated from corresponding interest points or pixels in images, or by correspondences towards a local 3D reconstruction

What is Visual Odometry?

Visual odometry (VO)…

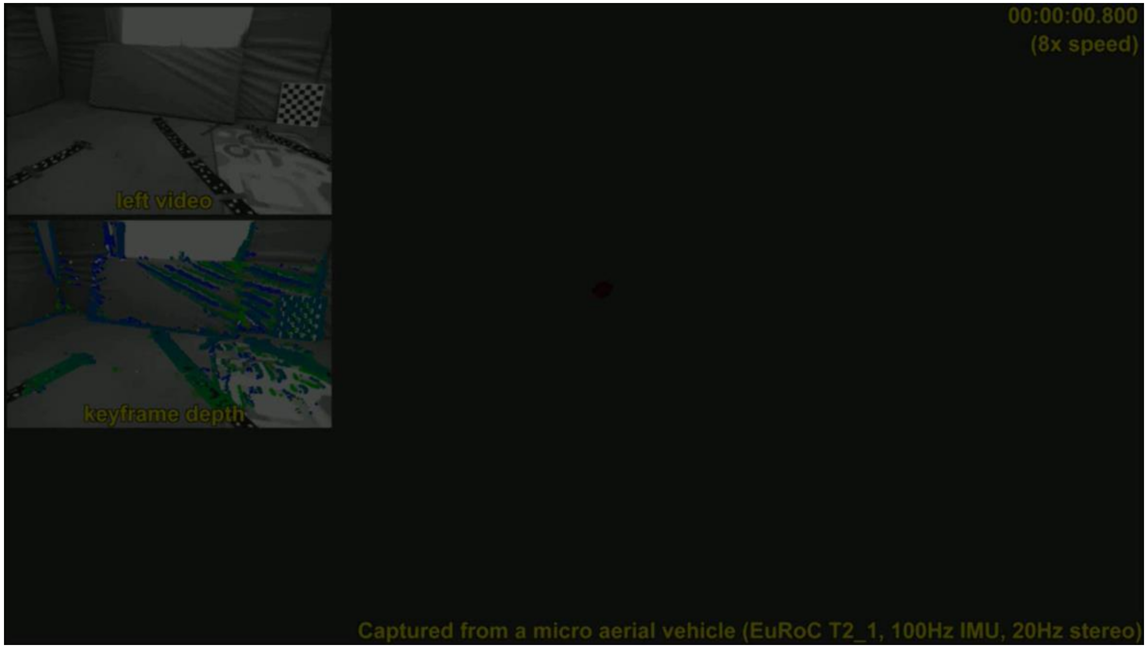
- … is prone to drift due to its local view
- … is primarily concerned with estimating camera motion
  - 3D reconstruction often a “side product”. If estimated, it is only locally consistent



Visual-Inertial Odometry

Sensor includes

- Stereo camera
- 3-axis accelerometer
- 3-axis gyroscope
- Time-synchronization



(Usenko, Engel, Stückler, Cremers, ICRA 2016)

What is Visual-Inertial Odometry?

Visual-inertial odometry (VIO)…

- … complements visual odometry with inertial measurements
  - Visual measurements provide up to 6-DoF relative motion using the environment as reference
  - Inertial sensors measure 3D linear accelerations and angular velocities, typically at much higher frame-rate than images
  - Gravity is also included in the acceleration measurements serving as an absolute external reference
  - Pure integration of gravity-compensated linear accelerations and angular velocities drifts
  - Vision helps to reduce integration drift, estimate sensor biases, discern gravity from motion-induced accelerations
  - Inertial measurements help to compensate degenerate cases of pure visual tracking (textureless areas, fast motion, etc.)

Simultaneous Localization and Mapping

(Engel, Stückler, Cremers, IROS 2015)

Where is the robot and what is the
3D structure of the environment?

What is Visual SLAM?

- Visual simultaneous localization and mapping (VSLAM)…
  - Tracks the pose of the camera in a map, and simultaneously
  - Estimates the parameters of the environment map (f.e. reconstruct the 3D positions of interest points in a common coordinate frame)
- Loop-closure: Revisiting a place allows for drift compensation
  - How to detect a loop closure?

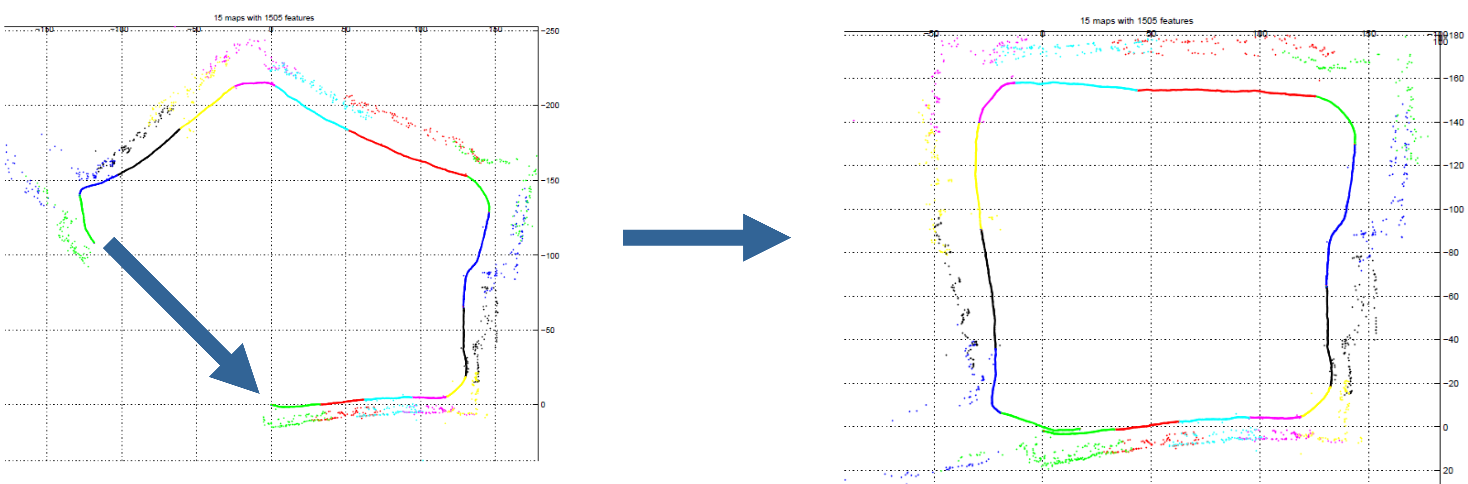
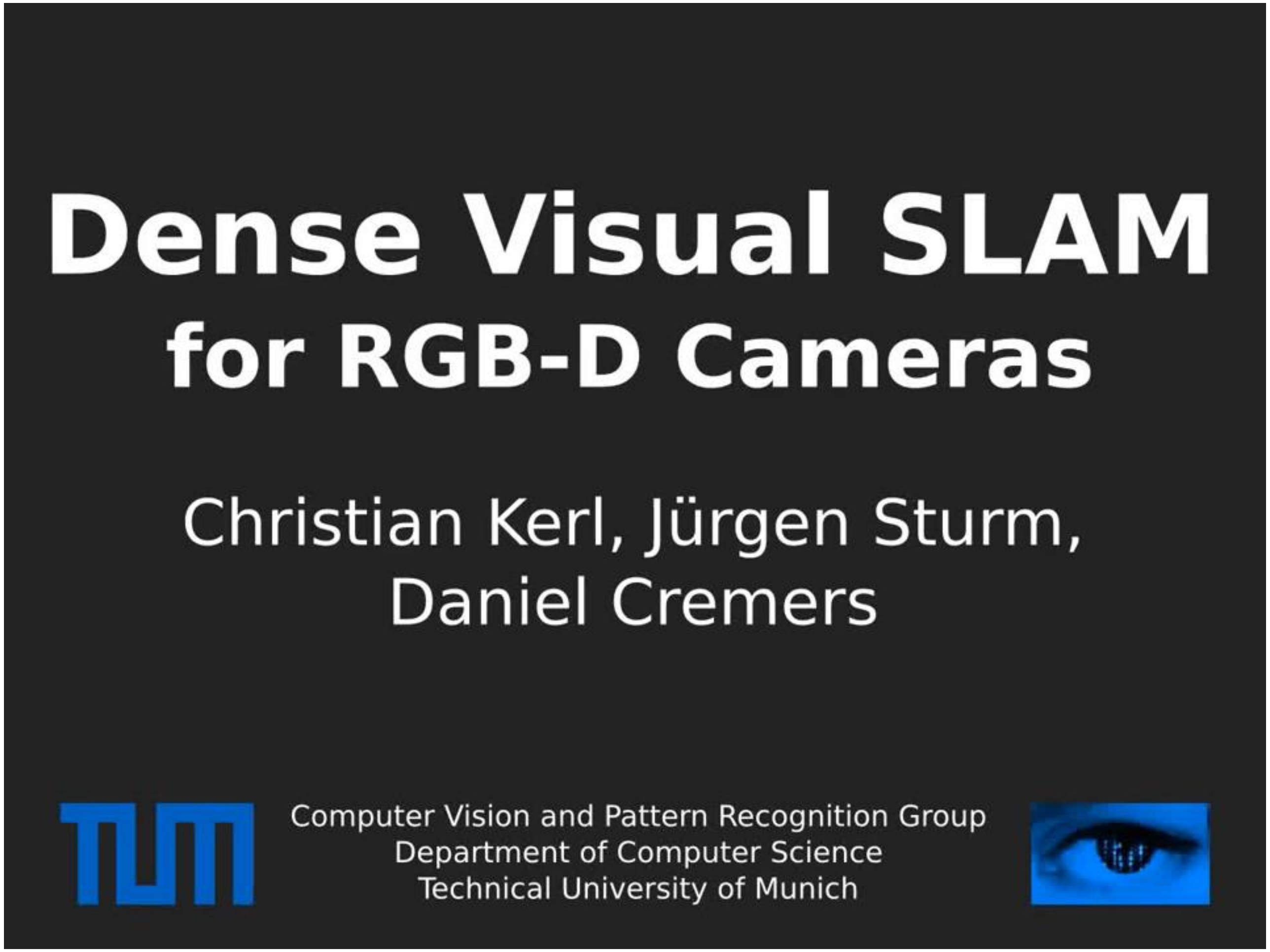


Image credit: Clemente et al., RSS 2007

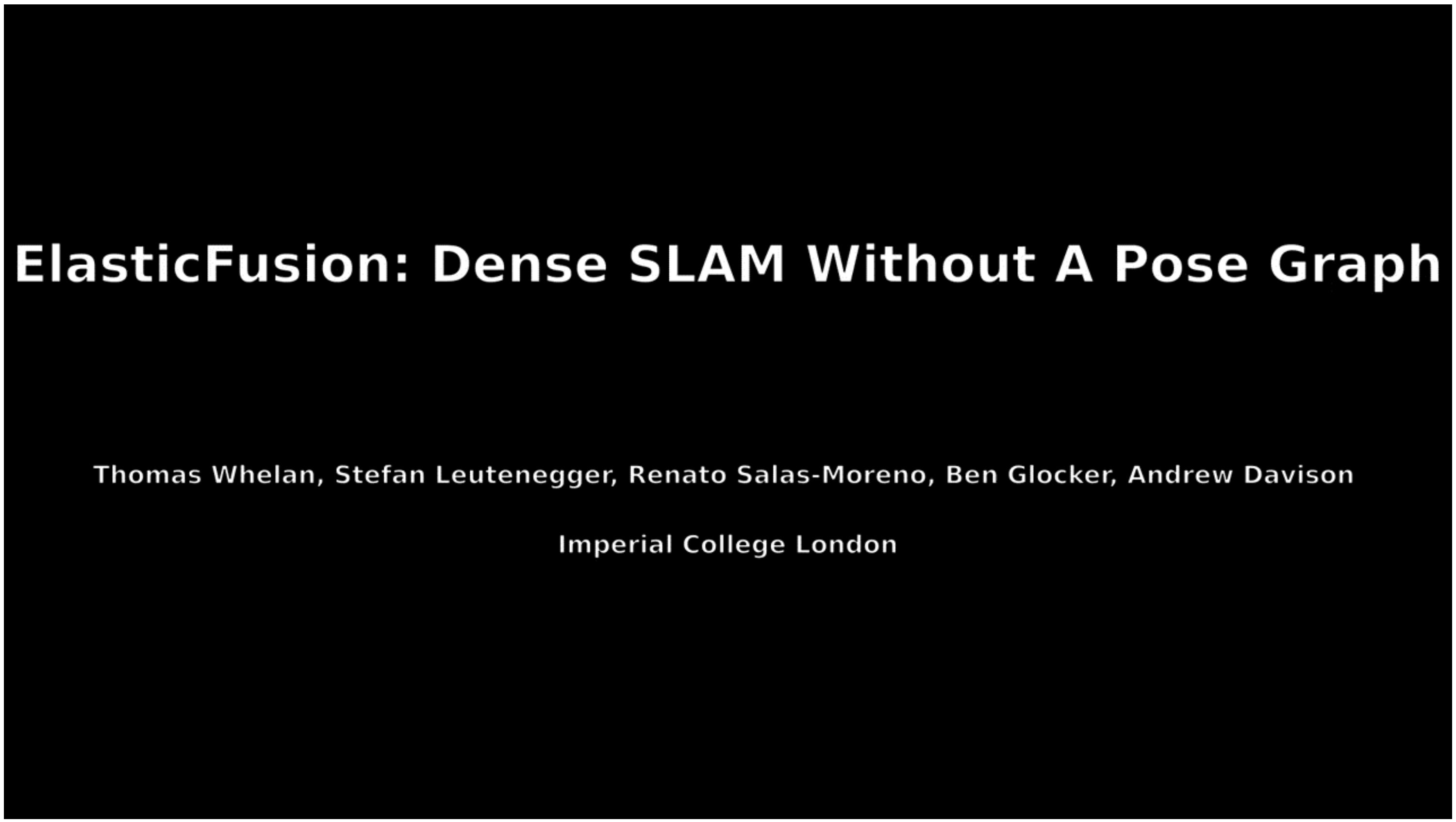
What is Visual SLAM?

- Visual simultaneous localization and mapping (VSLAM)…
  - Tracks the pose of the camera in a map, and simultaneously
  - Estimates the parameters of the environment map (f.e. reconstruct the 3D positions of interest points in a common coordinate frame)
- Loop-closure: Revisiting a place allows for drift compensation
  - How to detect a loop closure?
- Global and local optimization methods
  - Global: bundle adjustment, pose-graph optimization, etc.
  - Local: incremental tracking-and-mapping approaches, visual odometry with local maps. Often designed for real-time.
  - Hybrids: Real-time local SLAM + global optimization in a slower parallel process (f.e. LSD-SLAM)

Visual SLAM with RGB-D Cameras



RGB-D SLAM by Map Deformation



Visual SLAM using Bundle Adjustment

Universidad Zaragoza

1542

ROBOTICS

UNIVERSIDAD DE ZARAGOZA

Instituto Universitario de Investigación en Ingeniería de Aragón

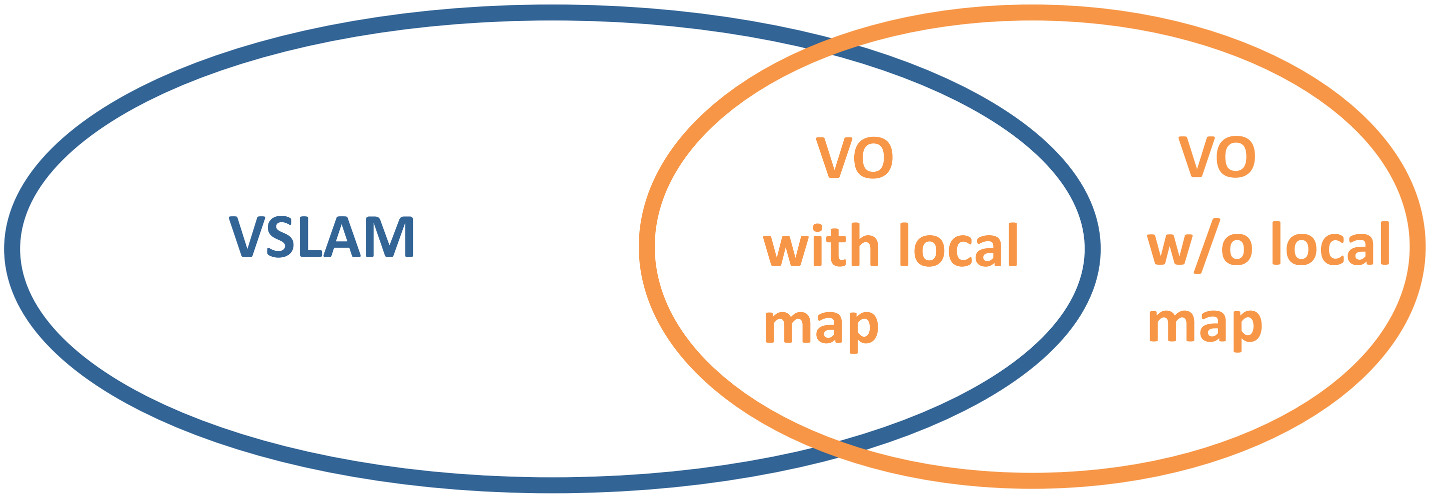
Universidad Zaragoza

ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras

Raúl Mur-Artal and Juan D. Tardós

raulmur@unizar.es tardos@unizar.es

VO vs. VSLAM



Structure from Motion

- Structure from Motion (SfM) denotes the joint estimation of
  - Structure, i.e. 3D reconstruction, and
  - Motion, i.e. 6-DoF camera poses,

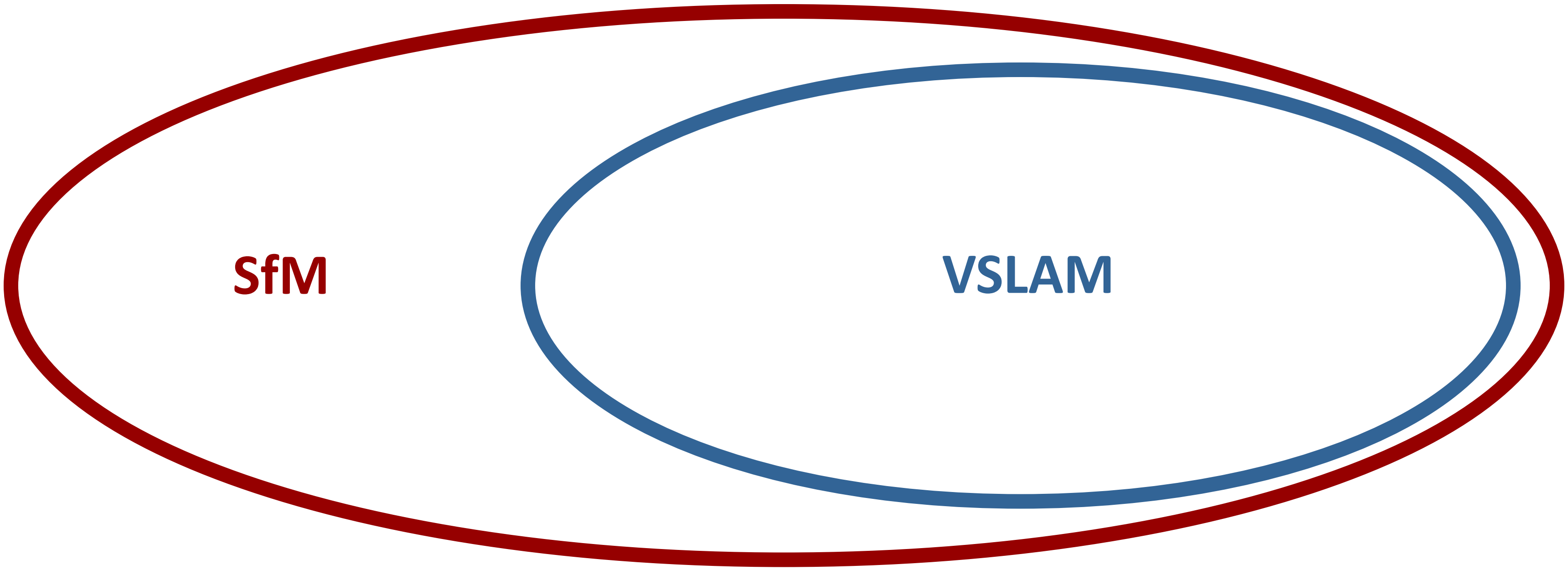
  from a collection (i.e. unordered set) of images

- Typical approach: keypoint matching and bundle adjustment

Structure from Motion

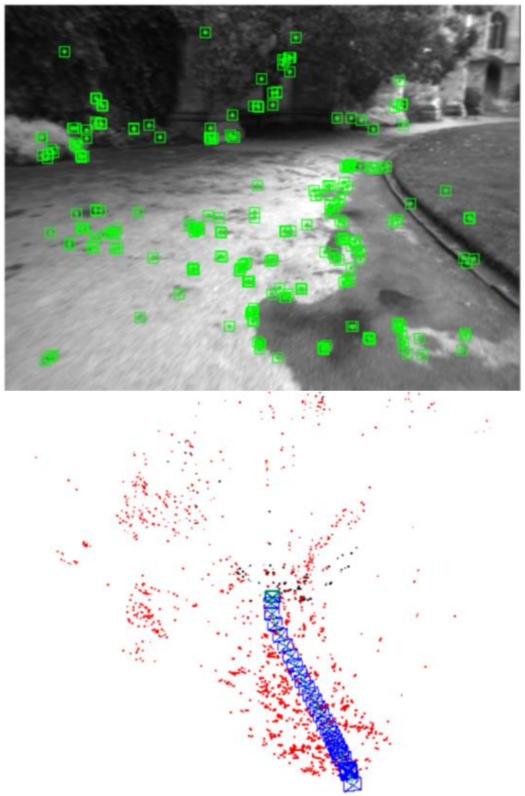
Agarwal et al., Building Rome in a Day, ICCV 2009, „Dubrovnik“ image set

VSLAM vs. SfM

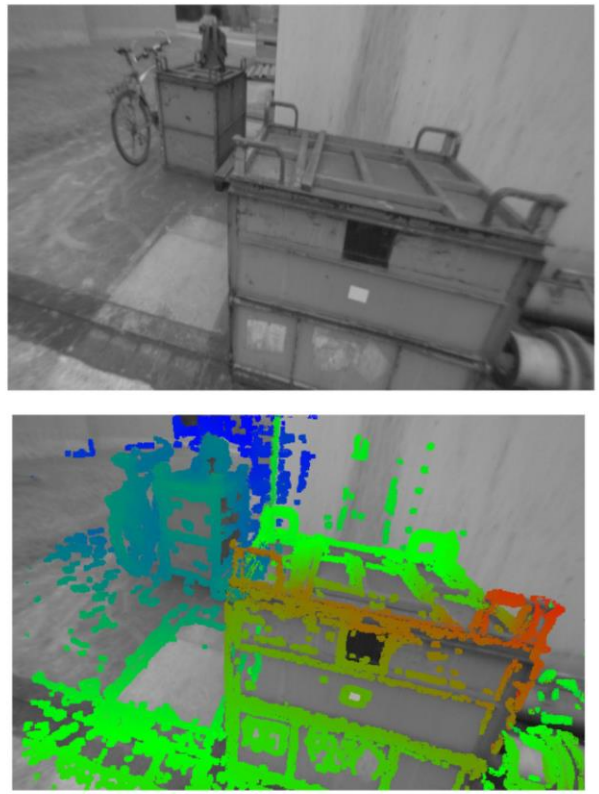


Sparse vs. Dense Reconstruction

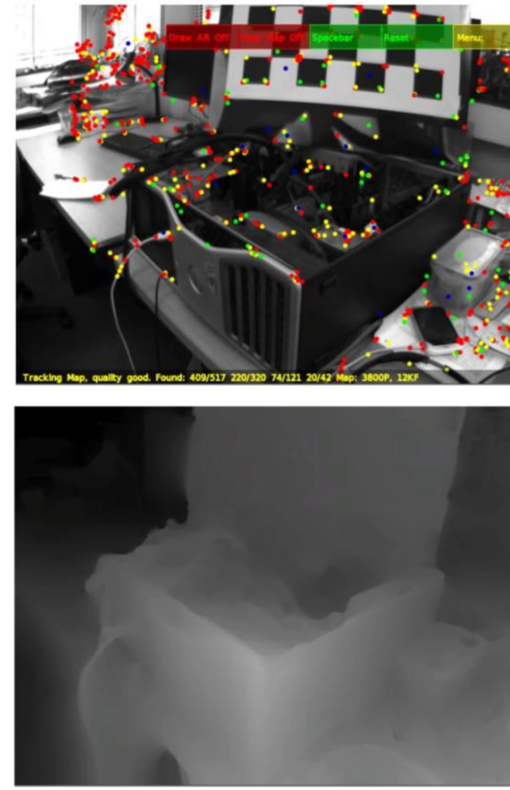
Sparse (ORB-SLAM)	Semi-Dense (LSD-SLAM)	Dense (DTAM)



(Mur-Artal and Tardós, TRO 2015)



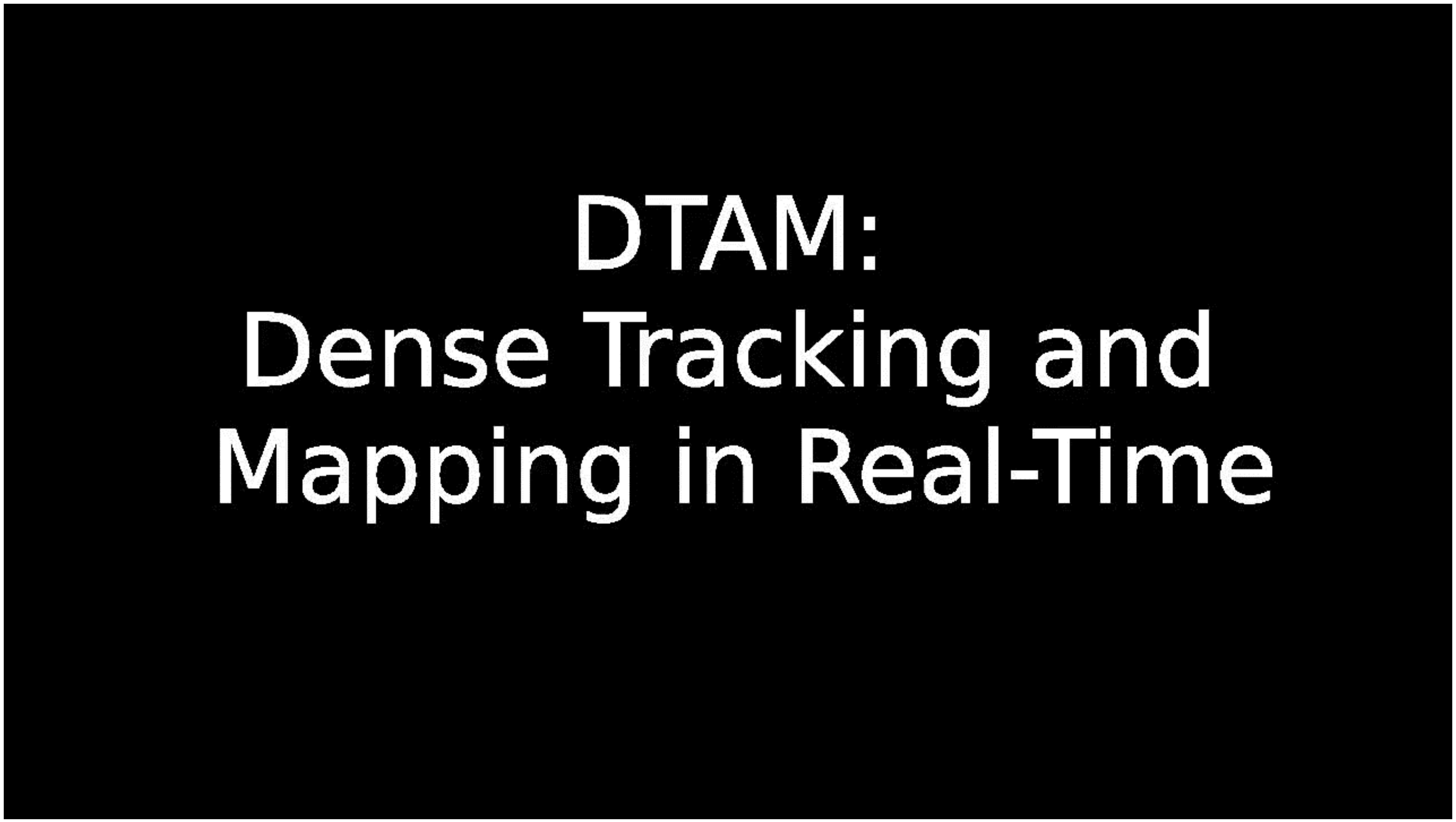
(Engel et al., ECCV 2014)



(Newcombe et al., ICCV 2011)

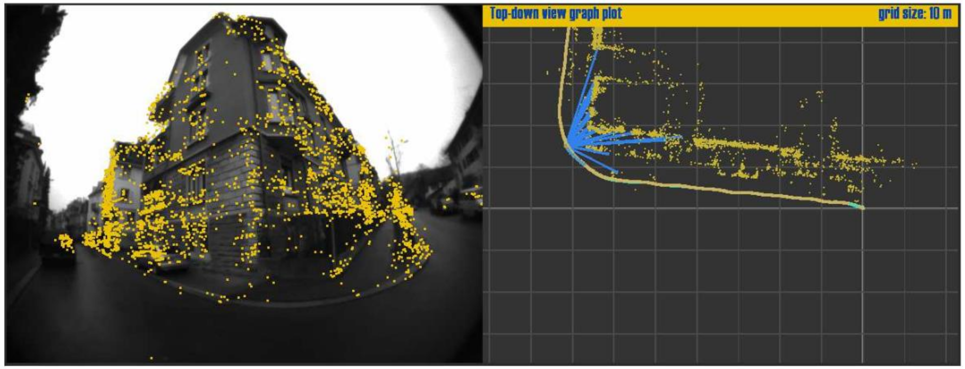
Good for VO/VSLAM = Good for robotic perception?

Dense VSLAM with a Single Camera

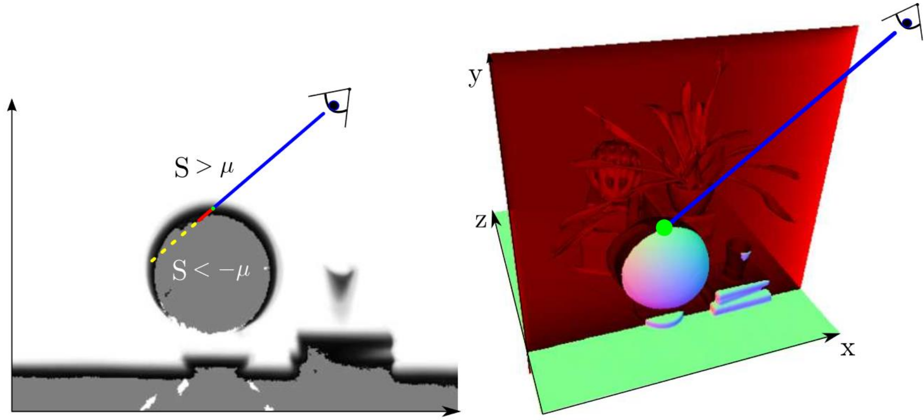


(Newcombe et al., DTAM: Dense Tracking and Mapping in Real-time, ICCV 2011)

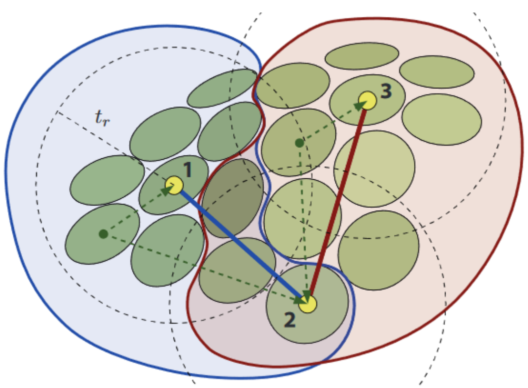
How Should We Represent The Map?



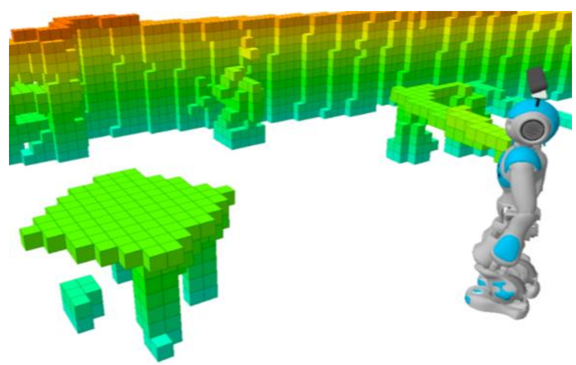
Sparse interest points



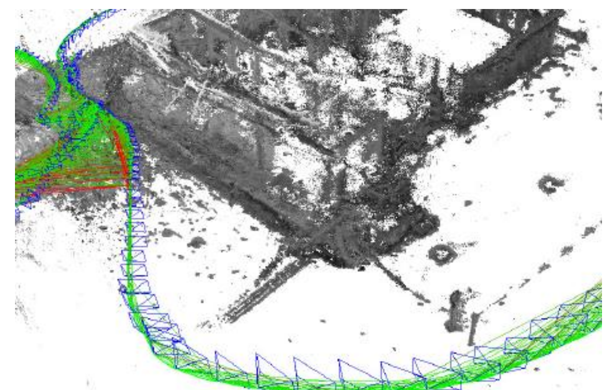
Volumetric, implicit surface



Explicit surface
(surfels, mesh,…)



Volumetric, occupancy

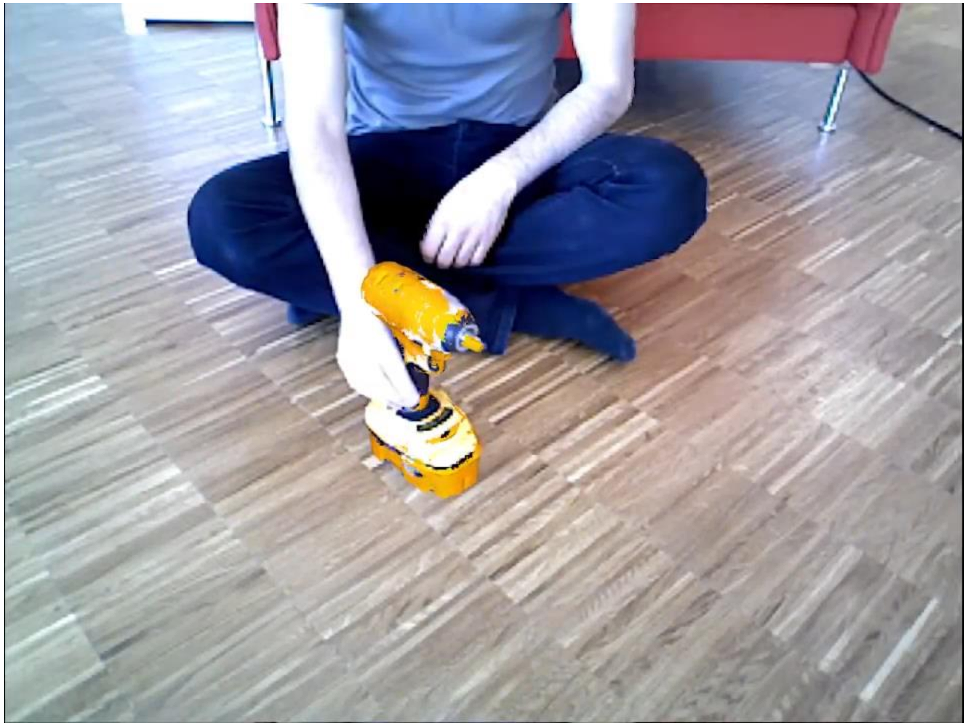


Keyframe-based maps

Good for VO/VSLAM = Good for robotic perception?

(Lynen et al., RSS 2015), (Newcombe, 2015), (Weise et al., 2009), (Maier et al., 2012), (Engel et al., ECCV 2014)

3D Object Detection and Tracking

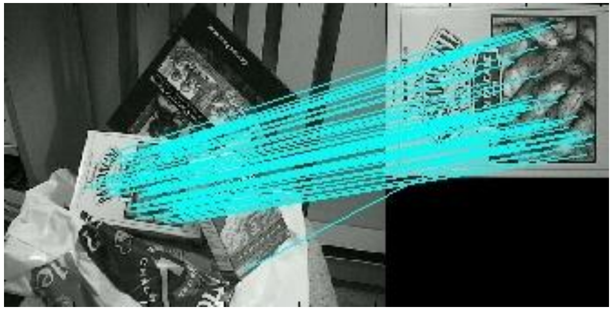


(Wüthrich et al., IROS 2013)

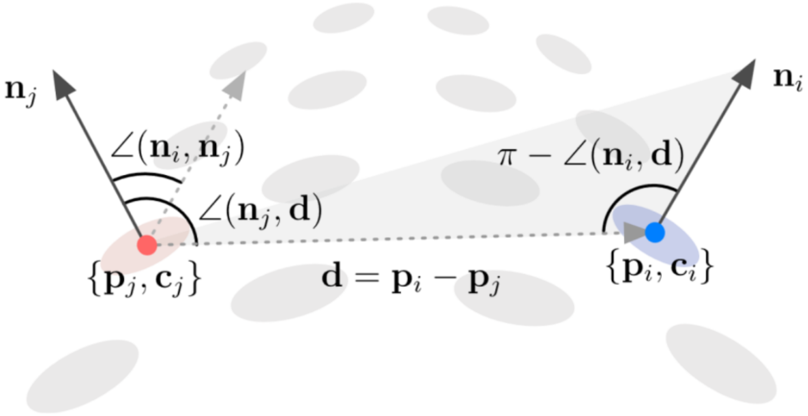
Where are objects in the robot's surrounding?

What is Visual 3D Object Detection?

- Visual 3D object detection…
  - …finds an object in an image and
  - …estimates its 6-DoF pose from the image

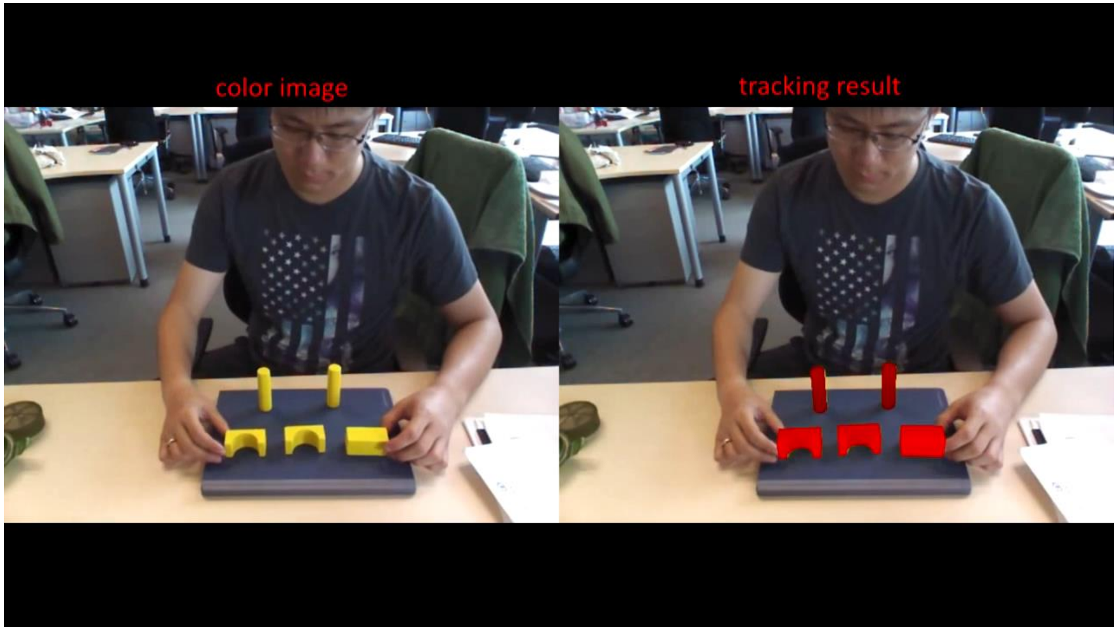


(Choi and Christensen, RAS 2016)



What is Visual 3D Object Tracking?

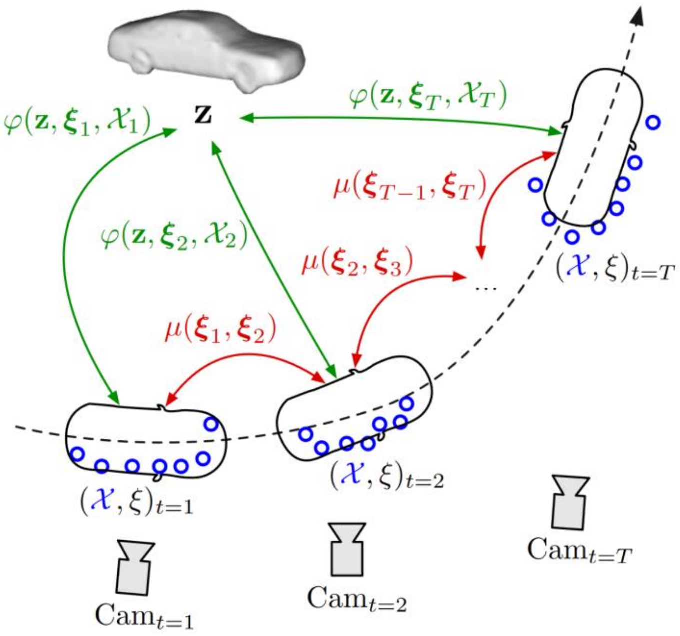
- Visual 3D object tracking…
  - …tracks the 6-DoF pose of an object in an image sequence
- Tracking-by-detection, incremental registration, …
- Multi-object tracking involves data association

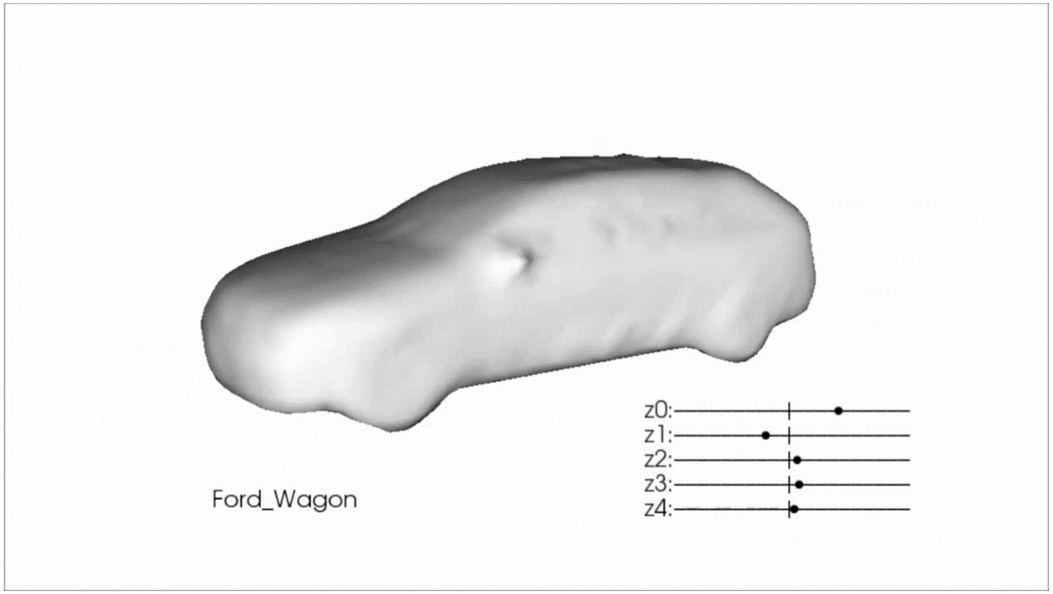


(Ren et al., Real-Time Tracking of Single and Multiple Objects from Depth-Colour Imagery Using 3D Signed Distance Functions, IJCV 2017)

Joint Object Shape Estimation and Tracking

Impose shape and motion priors for spatio-temporal reconstruction of vehicles





(Engelmann, Stückler, Leibe, WACV 2017)

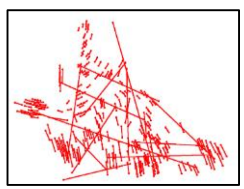
Indirect vs. Direct Methods

Indirect

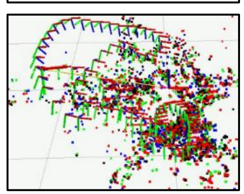
Input images


Extract and match keypoints (SIFT,BRIEF,…)

Track: min. reprojection error (point distances)



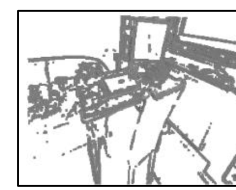
Map: estimate keypoint parameters (f.e. 3D coordinates)



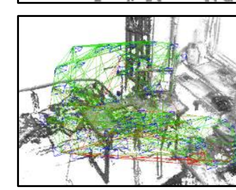
Direct

Input images


Track: min. photometric/ geometric error pixel-wise



Map: estimate per-pixel depth from photoconsistency



Indirect vs. Direct Methods

- Direct methods formulate image alignment objective in terms of photometric error (e.g. intensities)

$$E(\boldsymbol{\xi}) = \int_{\mathbf{u} \in \Omega} \left| \mathbf{I}_1(\mathbf{u}) - \mathbf{I}_2(\omega(\mathbf{u}, \boldsymbol{\xi})) \right| d\mathbf{u}$$

- Indirect methods formulate image alignment objective in terms of reprojection error of geometric primitives (e.g. points, lines)

$$E(\boldsymbol{\xi}) = \sum_i \left| \mathbf{y}_{1,i} - \omega(\mathbf{y}_{2,i}, \boldsymbol{\xi}) \right|$$

Indirect vs. Direct Methods

- Which of the approaches performs better is still in debate
- Indirect methods for VO and VSLAM have been investigated for a longer time by a broader research community
- Hence, indirect VO and VSLAM approaches are currently still more mature (f.e. ORB-SLAM2)
- However, recent methods such as direct sparse odometry (Engel et al., 2016) demonstrate better performance than several indirect visual odometry approaches
- Key to achieving high accuracy with direct methods is the proper treatment of camera properties such as vignetting, exposure times, rolling shutter etc.

Visual SLAM in Dynamic Scenes

- So far VO or VSLAM assumed static environments
- How to handle moving or deforming objects in SLAM?
- Recently impressive results with RGB-D cameras

*Dynamic***Fusion**:

Reconstruction & Tracking of Non-rigid Scenes in *Real-Time*

Richard Newcombe, Dieter Fox, Steve Seitz

Computer Science and Engineering,
University of Washington

Course Contents

- Image formation, multi-view geometry, SE3 (recap)
- Probabilistic filtering, non-linear least squares
- Visual odometry
- Visual-inertial odometry
- Visual SLAM
- Dense reconstruction
- Map representations
- 3D object detection and tracking
- Outlook: Visual SLAM in dynamic scenes

Thanks for your attention!