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

Sergei Solonets, Daniil Sinitsyn
Computer Vision Group
Technical University of Munich
How can I access these slides?

- https://vision.in.tum.de/teaching/ws2023/seminar_realtime3d
Outline

◦ General Information
  ◦ About the Seminar
  ◦ Registration

◦ Possible Papers
  ◦ Bundle Adjustment
  ◦ Monocular Cameras

◦ Questions
Outline

◦ General Information
  ◦ About the Seminar
  ◦ Registration

◦ Possible Papers
  ◦ Bundle Adjustment
  ◦ Monocular Cameras

◦ Questions
How is the seminar organized?

- Seminar meetings: Talks and discussion
  - Day: Tuesday, approximately every second week (TBA)
  - Time: 14:00–16:00 (seminar) & 16:00–18:00 (supervisor meetings)
  - Location: 00.08.055, Seminarraum (5608.EG.055)
  - 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
  - Two weeks before talk: meet supervisor for questions
  - One week before talk: meet supervisor to go through slides

- Grading
  - 60% Presentation
  - 40% Tests: at the beginning of each session. 8/12 best scores are counted
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 → 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?

- 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
  - 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
  ◦ Bundle Adjustment
  ◦ Monocular Cameras

◦ Questions
How do you register for the seminar?

· **Step 1**: Official registration via TUM matching system
  - Go to https://matching.in.tum.de : from July 14
  - Register for: *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]”
  - List how you fulfill the lecture requirements: See next slide
  - Attach your transcript(s)
  - Registrations without email / emails with missing information will be ignored!

· **Deadline** for both registrations: July 19, 2023
Required Lectures for the Seminar

- To understand the content of the seminar well, we recommend students to have completed
  - A lecture on optimization, similar to Nonlinear Optimization: Advanced (MA3503)
  - A lecture on computer vision that includes geometry, similar to Computer Vision II: Multiple View Geometry (IN2228)

- You can name up to three lecture from your transcript that, in combination, fulfill the requirements for subdomain. Example:
  - CV & Geometry: Computer Vision I (computer vision), Projektive Geometrie 1 (for geometry)

- Please list the lectures and brief explanations in your e-mail and attach your transcript(s) as proof. We will not scan your transcript(s) for suitable lectures!

- If you don’t perfectly fulfill the lecture requirements you might still be able to join – this will depend on the other applicants! Thus: Consider applying anyways if you really want to take the seminar!
How do you register for the seminar?

Example registration email:

To:   realtime3d-wo23@vision.in.tum.de
Cc:   
Bcc:  

Reply To:
Subject: [Realtime3D] application [Lukas Koestler]
From:  Köstler, Lukas – lukas.koestler@tum.de
Signature: None

Message Size: 373 KB

Hi,

I would like to present the following papers:
1. ORB-SLAM
2. DSO
3. DTAM

I have taken the following related courses:
1. Optimization: Nonlinear Optimization: Advanced (MA2903)
2. Computer Vision & Geometry: Multiple View Geometry, Differential Geometry, Projektive Geometrie

Please find the attached transcript.

Best,
Lukas

transcript.pdf
361 KB
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 all criteria, selection will be random. Other students will be ranked according to the requirement fulfillment.
  - You will get notified by the matching system about the decision

• 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
  ◦ Bundle Adjustment
  ◦ Monocular Cameras

◦ Questions
Bundle Adjustment in the Large

Agarwal, Snavely, Seitz, Szeliski 2010

- Proposes to solve bundle adjustment problems with an inexact Newton method linked to a preconditioner rather than by a direct factorization
- Crucial for large-scale 3D scene reconstruction
Multicore Bundle Adjustment
Wu, Agarwal, Curless, Seitz 2011

- Presents multicore solutions to large-scale 3D scene reconstruction problems
- Based on a restructuring of the conjugate gradients solver into easily parallelizable operations
Stochastic Bundle Adjustment for Efficient and Scalable 3D Reconstruction

Zhou et al. 2020

- Integrates a clustering scheme into solving bundle adjustment
- Drastically reduces the per-iteration cost and allows distributed computing by decomposing the reduced camera matrix into subproblems
Square Root Bundle Adjustment for Large-Scale Reconstruction

Demmel, Sommer, Cremers, Usenko 2021

- Challenges the traditional Schur Complement trick
- Combines a very general theoretical derivation of nullspace marginalization with the specific structure of bundle adjustment problems
Pixel-Perfect Structure-from-Motion with Featuremetric Refinement

Lindenberger, Sarlin, Larsson, Pollefeys 2021

- Refinements of keypoint and bundle adjustments by using a featuremetric error based on dense features predicted by a neural network
- Significantly improves the accuracy of camera poses and scene geometry
Outline

◦ General Information
  ◦ About the Seminar
  ◦ Registration

◦ Possible Papers
  ◦ Bundle Adjustment
  ◦ Monocular Cameras

◦ 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

Michael 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
Use feature-metric Bundle Adjustment within a differentiable deep-learning pipeline

- Allows the end-to-end training of NNs for SLAM

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)
DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras

Teed and Deng 2021

Monocular, Stereo or RGB-D Video

Δpose  Δdepth

DROID-SLAM

Figure 1: DROID-SLAM can operate on monocular, stereo, and RGB-D video. It builds a dense 3D map of the environment while simultaneously localizing the camera within the map.

• Monocular, Stereo, and RGB-D visual SLAM based on optical flow estimation (RAFT by Teed and Deng, 2020 ECCV best paper) and bundle adjustment

• Shows impressive robustness and accuracy across a wide range of datasets while trained only on the TartanAir dataset
iMAP: Implicit Mapping and Positioning in Real-Time

Edgar Sucar, Shikun Liu, Joseph Ortiz, Andrew J. Davison 2021

- Uses a multilayer perceptron (MLP) as scene representation for an RGB-D camera
NICE-SLAM: Neural Implicit Scalable Encoding for SLAM

Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R. Oswald, Marc Pollefeys 2022

- Uses a multilayer perceptron (MLP) as scene representation for an RGB-D camera
DualRefine: Self-Supervised Depth and Pose Estimation Through Iterative Epipolar Sampling and Refinement Toward Equilibrium

Antyanta Bangunharcana, Ahmed Magd, Kyung-Soo Kim 2023

- Uses deep equilibrium model to estimate depth and pose
Propose your own

- You can propose a paper which is relevant to the topic
- acceptance is not guaranteed
Questions?

Reminder:

- Web page: https://vision.in.tum.de/teaching/ws2023/seminar_realtime3d
- Contact: realtime3d-ws23@vision.in.tum.de