Praktikum on 3D Computer Vision

F. Tombari
N. Brasch, K. Li, A. Savkin, J. Huang, S.R. Vutukur, F. Tristram, L. Bastian, S. Wang, M. Kiray, C. Li, K. Reichard, C. Kapeller, N. Morbitzer

Introduction

- 3D Computer Vision

  Scene understanding/Novel View Synthesis
  - 6D object pose estimation
  - SLAM, Structure from Motion
  - 3D reconstruction
  - Camera pose / re-localization
  - 3D Scenes/Objects Generation
  - Neural Radiance Fields
  - Semantic segmentation / understanding
  - Depth prediction, stereo

  Human understanding
  - 3D body / hand / face tracking/pose
  - 3D modeling of body and body parts

- Application in Robotics
  - Grasping and Manipulation
  - Navigation
  - Obstacle avoidance
  - Mapping

- Augmented/Mixed Reality
  - Render virtual/augmented content on real objects/scenes

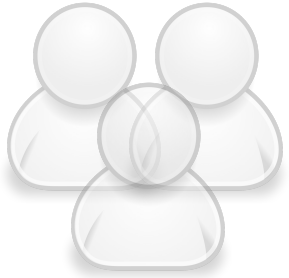
Goals of the Praktikum

- Learn about the **state of the art in 3D computer vision**
- Familiarize with **practical aspects and use cases** of typical 3D perception tasks (3D feature extraction and learning, surface matching and 3D reconstruction, 3D object localization and pose estimation, SLAM, ..)
- **Plan & develop an end-to-end project in a team** aiming to solve a relevant and challenging problem in 3DCV
- Learn to **explain and disseminate your work** in tech talks

Teams

Setup

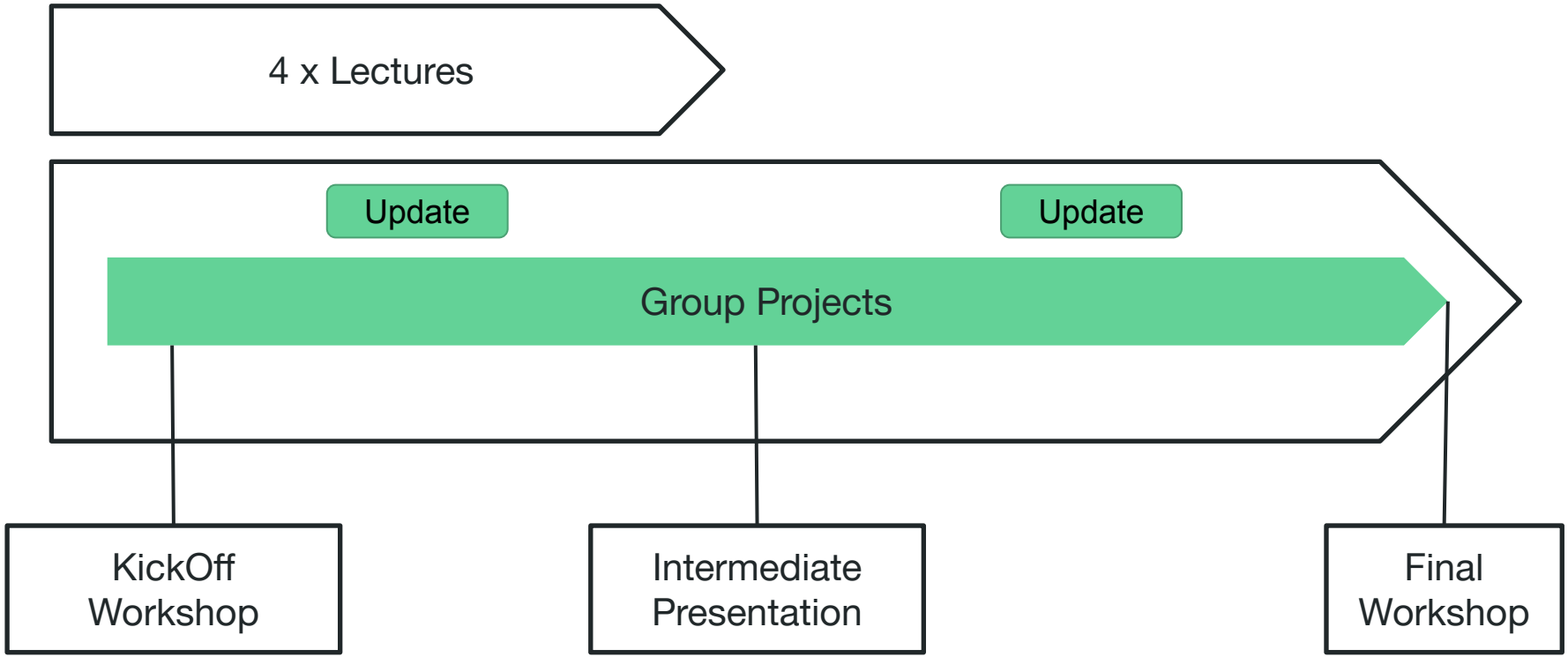
- There will be up to **5 projects** with **teams of 4-6 students**
- Each project has one or more **advisors** that will support the team during the project



Student-to-Project Matching

- Registered students can **indicate project preference** after project announcements
- Students will be assigned to a team and project that **best fits the indicated preference & background**

Course structure



Schedule (tentative)

Lecture period: 13.10.2025 – 06.02.2026

17.10.	Introductory talk & Project presentations
24.10.	Lecture I & Lecture II
	Project assignments
31.10.	Project KickOffs
	CVPR Break
07.11.	Lecture III & Lecture IV
28.11.	Project Update I
12.12.	Mid-term Presentations
	Christmas Break
16.01.	Project Update II
30.01.	Final Workshop

Time: Fridays 14.00 - 15.30
Place: Seminar Room 03.13.010

In-person attendance in each session is mandatory. Missing more than one session without a valid excuse can lead to failing the course.

Evaluation

Project work (75 %)

- Activity & engagement
- Teamwork & communication
- Scientific understanding & depth
- Methodology, implementation & evaluation
- Project management

Presentations (25%)

- Scientific understanding & explanations
- Structure of the presentation
- Presentation style
- Quality of slides
- Q&A

Prerequisites

- Required: 1+ computer vision-related course
  - Tracking and Detection in CV (IN2357)
  - Computer Vision I: Variational Methods,
  - Computer Vision II: Multiple View Geometry (IN2228)
  - Robotic 3D Vision, Convex Optimization for ML and CV, Probabilistic Graphical Models in CV
  - ...
- Required: 1+ deep-learning-related course
  - Introduction to Deep Learning (I2DL) (IN2346)
  - Machine Learning (IN2064)
  - Machine Learning for 3D Geometry (IN2392)
  - ...
- Suggested:
  - 1+ projects in the domain of CV/ML

Registration

- Register in Matching-System: https://matching.in.tum.de (until 22.07.)
- Send motivation letter, CV & transcript (not mandatory, but highly recommended) to: **p3dcv@camp.cit.tum.de** (until 22.07)
  - Why do you want to participate in this course in particular?
  - What background related to computer vision & machine learning do you have?
- Ranking based on
  - Motivation and fit into degree
  - Background and previous experiences
  - Study program & semester

Projects from last year

Project 1: Hypernetwork for Hash-Grid Representations

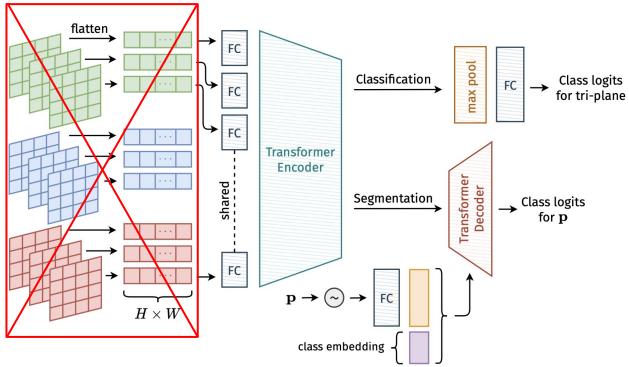
Felix Tristram, Nils Morbitzer

Motivation

- Various 3D Representations:
  - Explicit: Point clouds, meshes, voxel
  - Implicit: SDFs, NeRFs
- Unclear how to handle Implicit Representations (NeRF) for downstream tasks [2]
- How can we get good latent representations of Fast NeRF architectures? [1]

Objectives

- Train multiple InstantNGP's [3] as dataset (first 2D & later 3D)
- Autoencoder / MAE for InstantNGP [3] architectures
- T-SNE visualization of latent space
- Downstream tasks: Scene completion, Segmentation



[1] A. Cardace, P. Z. Ramirez, F. Ballerini, A. Zhou, S. Salti, and L. Di Stefano, 'Neural processing of tri-plane hybrid neural fields', arXiv preprint arXiv:2310. 01140, 2023.
[2] P. Z. Ramirez et al., 'Deep learning on 3D neural fields', arXiv preprint arXiv:2312. 13277, 2023.
[3] T. Müller, A. Evans, C. Schied, and A. Keller, 'Instant Neural Graphics Primitives with a Multiresolution Hash Encoding', ACM Trans. Graph., vol. 41, no. 4, p. 102:1-102:15, 2022.

Project 2: Sensor Cross Calibration by Neural Reconstruction

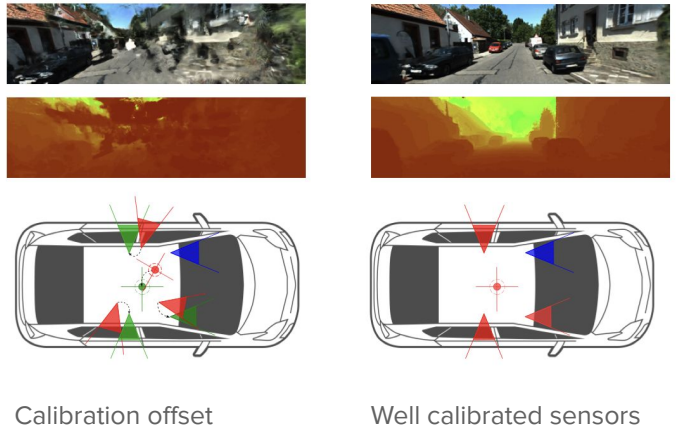
Markus Herb, Nikolas Brasch

Motivation

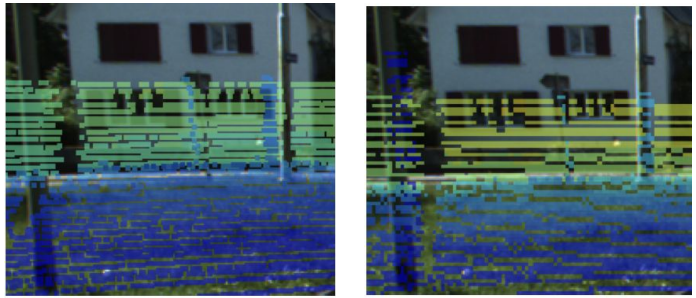
- Well calibrated sensors are crucial for reliable perception in AR, robotics, autonomous vehicles
- Manual calibration using calibration targets is tedious and error-prone
- NeRF [1] or 3DGS [2] can jointly reconstruct the scene *and* optimize for poses, which can be used for fully automatic, targetless sensor-calibration

Objectives

- Train 3DGS or NeRF scene reconstructions
- Optimize vehicle trajectory and relative sensor poses jointly with scene reconstruction to find calibration
- Stretch Goal: Optimize sensor intrinsics



Calibration offset | Well calibrated sensors



[1] Herau, Q., Piasco, N., Bennehar, M., Roldão, L., Tsishkou, D., Migniot, C., Vasseur, P., & Demonceaux, C. *SOAC: Spatio-Temporal Overlap-Aware Multi-Sensor Calibration using Neural Radiance Fields*. CVPR 2024
[2] Herau, Q., Bennehar, M., Moreau, A., Piasco, N., Roldao, L., Tsishkou, D., Migniot, C., Vasseur, P., & Demonceaux, C.. *3DGS-Calib: 3D Gaussian Splatting for Multimodal SpatioTemporal Calibration*. IROS 2024

Project 3: Feature Gaussian SLAM

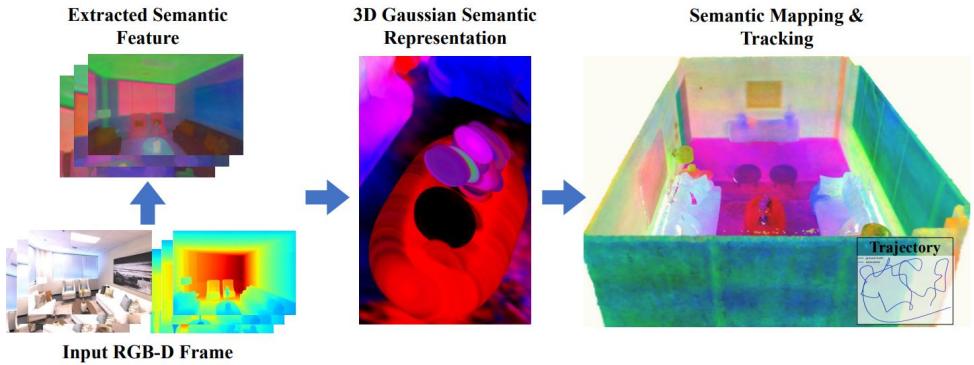
Kunyi Li, Sen Wang

Motivation

- Simultaneous localization and mapping
- Distill feature fields from 2D foundation model

Objectives

- Good camera pose estimation
- Distill feature fields from 2D foundation model(SAM+CLIP)



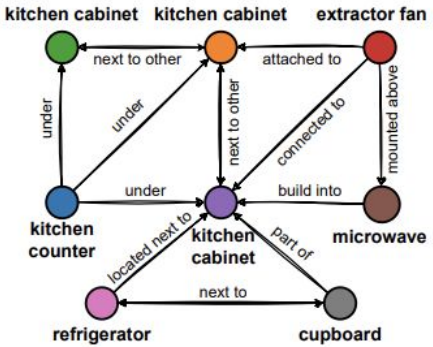
Keetha N, Karhade J, Jatavallabhula K M, et al. SplaTAM: Splat Track & Map 3D Gaussians for Dense RGB-D SLAM[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 21357-21366.
Zhou S, Chang H, Jiang S, et al. Feature 3dgs: Supercharging 3d gaussian splatting to enable distilled feature fields[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 21676-21685.
Zhu S, Wang G, Blum H, et al. Sni-slam: Semantic neural implicit slam[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 21167-21177.
Zhu S, Qin R, Wang G, et al. Semgauss-slam: Dense semantic gaussian splatting slam[J]. arXiv preprint arXiv:2403.07494, 2024.
Ren T, Liu S, Zeng A, et al. Grounded sam: Assembling open-world models for diverse visual tasks[J]. arXiv preprint arXiv:2401.14159, 2024.

Project 4: 3D-aware Open Vocabulary Scene Graphs

Mert Kiray, Klara Reichard

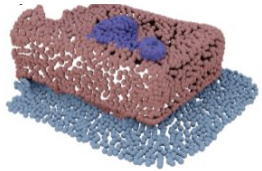
Motivation

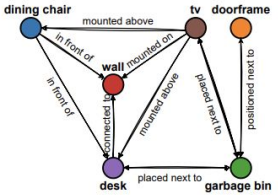
- Scene Graphs have difficulties with 3D aware relationships
- e.g. “to the left of” is not uniquely defined
- Use 3D-aware Vision Language Models like LLaVA-3D [1] or LL3DA [2] for relationships



Objectives

- Generate a 3D Open Vocabulary Scene Graph with improved 3D aware relationships between nodes
  - Reproduce Results from Open3DSG [3]
  - Compare Different 2D Features from Vision Language Models
  - Use LLaVA-3D [1] or LL3DA [2] Features instead of BLIP
- Open next steps from here





[1] Zhu, Chenming, Tai Wang, Wenwei Zhang, Jiangmiao Pang, and Xihui Liu. ‘LLaVA-3D: A Simple yet Effective Pathway to Empowering LMMs with 3D-Awareness’. arXiv Preprint arXiv:2409. 18125, 2024.
[2] Chen, Sijin, Xin Chen, Chi Zhang, Mingsheng Li, Gang Yu, Hao Fei, Hongyuan Zhu, Jiayuan Fan, and Tao Chen. ‘LL3DA: Visual Interactive Instruction Tuning for Omni-3D Understanding Reasoning and Planning’. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 26428–38, 2024.
[3] Koch, Sebastian, Narunas Vaskevicius, Mirco Colosi, Pedro Hermosilla, and Timo Ropinski. ‘Open3DSG: Open-Vocabulary 3D Scene Graphs from Point Clouds with Queryable Objects and Open-Set Relationships’. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.

Project 5: Language Enhanced Reconstruction with Reflection

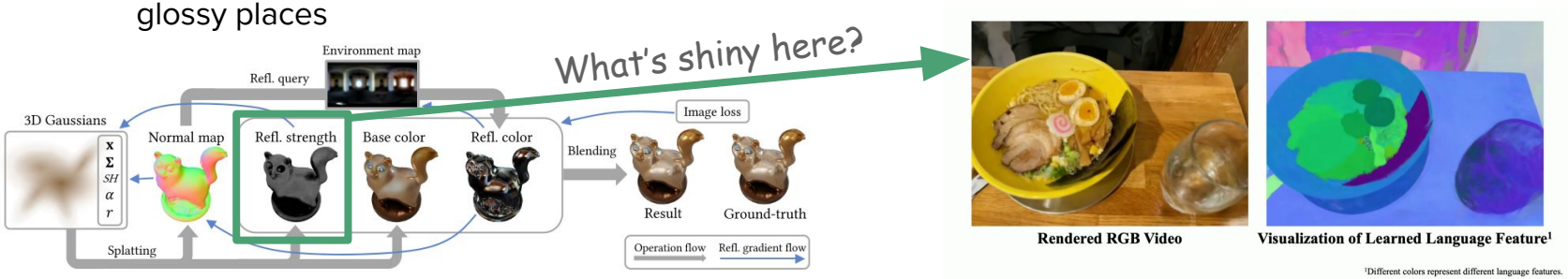
Advisors: Christian Kapeller, Changxuan Li

Motivation

- Reconstruction with 3DGS suffers issues for glossy materials prevalent in the real world
- Complex illumination scenarios exacerbate the issue
- Semantic embedded Gaussians [2] help scene understanding and query desired regions
- Join the advantages from both world for 3D scene understanding with complexity

Objectives

- Take a gaussian splatting algorithm [1] that provides material appearance information, such as roughness
- Incorporate the available appearance information into a semantic segmenter and CLIP [2]
- Use the resulting models to augment the available LangSplat algorithm [3] and enable it to query scenes for glossy places



[1] Ye, K., Hou, Q., & Zhou, K. (2024). 3D Gaussian Splatting with Deferred Reflection. *SIGGRAPH '24*, 1–10. https://doi.org/10.1145/3641519.3657456

[2] Qin, M., Li, W., Zhou, J., Wang, H., & Pfister, H. (2023). *LangSplat: 3D Language Gaussian Splatting*. http://arxiv.org/abs/2312.16084

[3] Radford, A.,et al. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. http://arxiv.org/abs/2103.00020

Questions?

Web. https://www.cs.cit.tum.de/camp/teaching/practical-courses/praktikum-on-3d-computer-vision-ws-2025-26/

E-Mail us: tombari@in.tum.de, nikolas.brasch@tum.de