European Spatial Data Research

From Traditional to AI-based 3D Scene Capture and Modeling

Instructors:
Michael Weinmann, Delft University of Technology (The Netherlands)
e-mail: m.weinmann@tudelft.nl
Dennis Haitz, Karlsruhe Institute of Technology (Germany)
e-mail: dennis.haitz@kit edu
Martin Weinmann, Karlsruhe Institute of Technology (Germany)
e-mail: martin.weinmann@kit.edu

Dates: March 17-28, 2025

Target audience: Staff of national mapping agencies, researchers, academics, students, private companies

Prerequisites: Participants are expected to be familiar with the Python programming language

Course objectives: Deep learning has led to significant breakthroughs in various fields. The advent of learning-based scene representations such as Neural Radiance Fields (NeRFs) and 3D Gaussian Splatting (3DGS) marks a significant leap in photogrammetric computer vision and novel view synthesis as well as respective applications in robotics, urban mapping, autonomous navigation, virtual/augmented reality, etc.. Learning-based approaches have been demonstrated to allow the precise and efficient encoding of high-resolution scene information, while additionally being more compact than conventional approaches based on scene representations in terms of point clouds, meshes or voxel block models. Through a blend of theoretical insights, visual illustrations and practical exercises, this course will delve into core concepts, implementation strategies, and advanced applications of traditional as well as recent AI-based 3D scene capture and visualization, providing you with the skills and knowledge to reflect on the strengths, innovation potential and limitations of current approaches.

Topics tackled: Photogrammetry, point cloud generation, machine/deep learning, conventional and learning-based approaches for 3D reconstruction from imagery, Structure-from-Motion (SfM), Multi-View Stereo (MVS), Neural Radiance Fields (NeRFs), 3D Gaussian Splatting (3DGS).

3D reconstruction from imagery: given a set of input images (left), the main objective is to derive a 3D scene representation in terms of geometry and colour information

Course outline: The course provides an overview of traditional and advanced AI-based approaches for 3D scene reconstruction from imagery. The pre-course seminar will introduce the underlying problem statement as well as the respective challenges. Furthermore, it will cover basic principles, best practices and recent trends related to the addressed topics. The e-learning part of the course is structured into different modules. These modules will include more in-depth theoretical aspects as well as practical assignments to deepen the studied materials. To benefit from the experimental work, a basic level of Python programming skills is required.

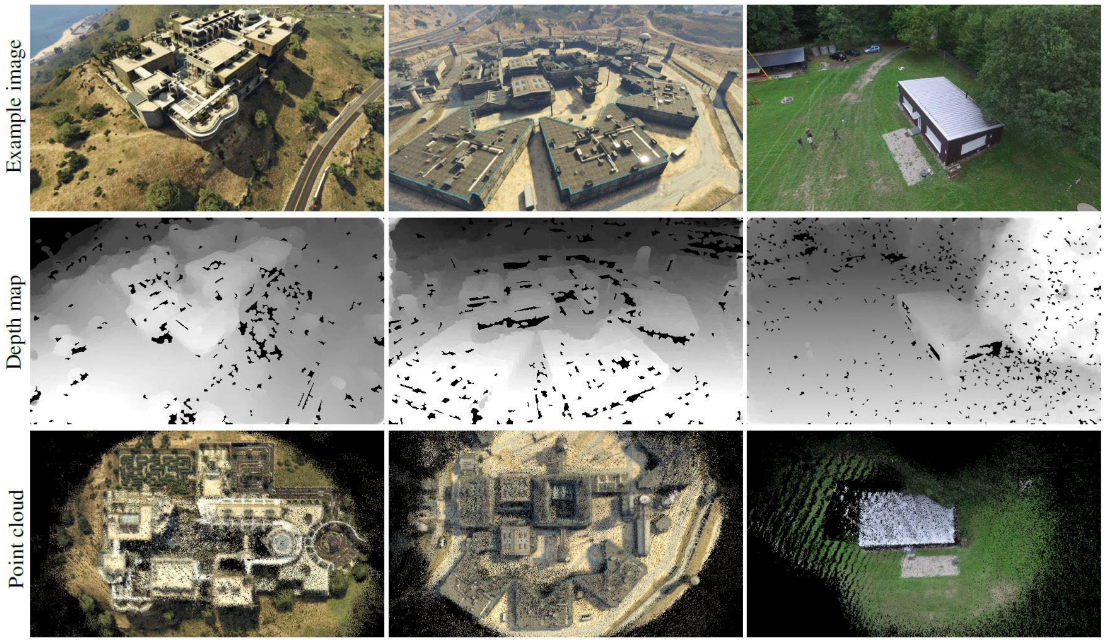
Module 1: Traditional Photogrammetry

This module provides an overview of basic concepts of geometry acquisition via (passive and active) optical 3D sensing techniques. A particular focus will be put on traditional approaches for the generation of 3D scene representations from acquired imagery using low-cost consumer hardware, which is, for instance, commonly applied for large-scale 3D reconstruction from aerial imagery or for indoor 3D mapping. In this regard, the organization of large unstructured (e.g. crowdsourced) image collections and well-established 3D reconstruction techniques (from either imagery or RGB-D data) will be considered in detail. In this context, representations of acquired data in the form of 3D point clouds, voxel occupancy grids and 3D meshes will be discussed as well as different approaches for closed surface reconstruction from 3D point clouds.

Exemplary 3D point cloud obtained with a traditional approach for 3D reconstruction from imagery

Module 2: Deep Learning meets Photogrammetry

This module addresses developments that leverage the potential of deep learning for 3D reconstruction from imagery, while still relying on conventional scene representations in the form of 3D point clouds, voxel occupancy grids, 3D meshes or depth maps. After an introduction to deep learning including standard network architectures, gradient-based optimization and regularization schemes, we will consider how different approaches can benefit from involving deep learning principles. In addition to learning-based multi-view 3D reconstruction, we will also revisit approaches for the ill-posed scenario of single-image-based 3D scene reconstruction.



Exemplary depth maps and 3D point clouds obtained with a learning-based approach for 3D reconstruction from imagery (Hermann et al., 2020)

Module 3: Advanced Scene Representations for 3D Scene Reconstruction

This module covers recent developments regarding advanced, learning-based scene representations for 3D scene reconstruction. Based on combining principles of machine learning, image formation and image synthesis, recently emerging scene representations such as Neural Radiance Fields (NeRFs) or 3D Gaussian Splatting (3DGS) are designed to optimize a scene representation in terms of weights of a neural network (in the case of NeRFs) or primitives like 3D Gaussians (in the case of 3DGS) in a supervised manner to match its predicted scene appearance under certain views to the respectively observed input photographs for the corresponding view configurations. Furthermore, we will focus on extensions towards improved model quality, acceleration of the involved training process, sparse input data, handling photo collections taken in-the-wild, and handling large-scale scenarios.

Exemplary visualization of an advanced 3D scene representation inferred from imagery