Technical Report on ISPRS Scientific Initiative 2021

H3D - Hessigheim 3D: Benchmark on Semantic Segmentation of High-Resolution 3D Point Clouds and Meshes from Airborne LiDAR and Multi-View-Stereo-Image-Matching

Investigators

Norbert Haala (Principal Investigator)
University of Stuttgart, Germany

Franz Rottensteiner (Co-Investigator)
Leibniz Universität Hannover, Germany

Jan Dirk Wegner (Co-Investigator)
ETH Zurich, Switzerland

Michael Kölle (Co-Investigator)
University of Stuttgart, Germany

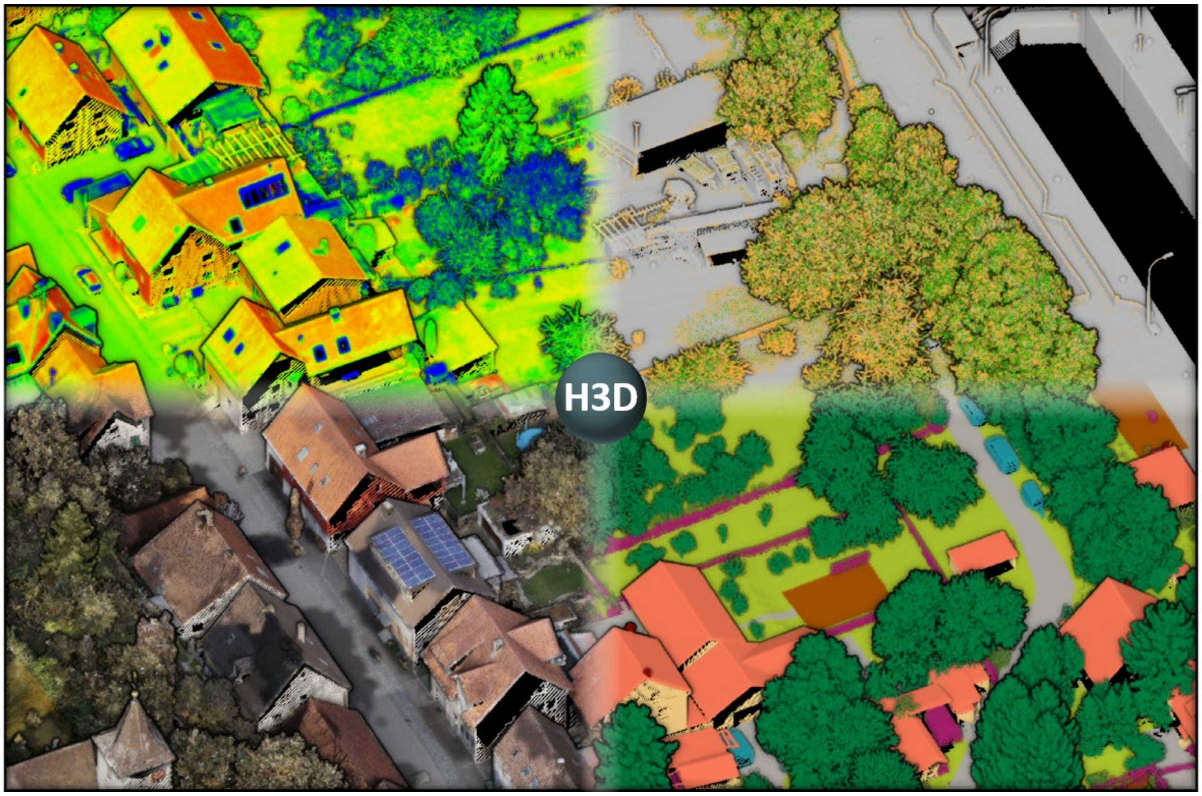
Dominik Laupheimer (Co-Investigator)
University of Stuttgart, Germany

Project Goals

The Hessigheim 3D (H3D) benchmark dataset provides an ultra-high resolution, fully annotated 3D dataset acquired from a LiDAR system and cameras integrated on the same Unmanned Aerial Vehicle (UAV) platform. This results in a unique multi-modal scene description by a LiDAR point cloud H3D(PC) and a textured 3D mesh H3D(Mesh), labelled into 11 classes. In addition to tests on semantic segmentation of LiDAR point clouds, H3D shall improve the dissemination and acceptance of the mesh representation in photogrammetry and remote sensing. So far, meshes are the state-of-the-art representation for small-scale datasets covering indoor scenes or single objects that are commonly treated by the computer vision community. In contrast to unordered point clouds, meshes are graph-based surface structures that provide explicit adjacency information. The surface description enables high-resolution texturing while efficiently storing geometry. With the help of H3D(Mesh), we want to foster semantic mesh segmentation and evaluate the community's interest in this kind of representation at the same time. Since data captured by the same sensor system is available for three epochs (March 2018, November 2018 and March 2019), investigations on multi-temporal semantic segmentation are also feasible.

Access to the data, including a more detailed description on data collection and labelling as well as the possibility to upload results for the participants is feasible by the benchmark webpage

https://ifpwww.ifp.uni-stuttgart.de/benchmark/hessigheim/default.aspx



References

Kölle, M., Laupheimer, D., Schmohl, S. , Haala, N. Rottensteiner, F. , Wegner, J.D. & Ledoux, H. (2021) The Hessigheim 3D (H3D) benchmark on semantic segmentation of high-resolution 3D point clouds and textured meshes from UAV LiDAR and Multi-View-Stereo, ISPRS Open Journal of Photogrammetry and Remote Sensing, Volume 1, 2021, https://doi.org/10.1016/j.ophoto.2021.100001.

Summary

Automated extraction of geographic objects from airborne data is an important research topic in photogrammetry and remote sensing since decades. In addition to images, 3D point clouds from airborne LiDAR and Multi-View-Stereo-Image-Matching became more and more important as basic data source. The aim of our project is to provide state-of-the-art data sets to the ISPRS community, which can be used by interested researchers to test own methods and algorithms on semantic segmentation for geospatial applications. We propose a benchmark consisting of highly dense LiDAR point clouds captured at three different epochs. The respective point clouds are manually labeled into 11 classes and are used to derive labeled textured 3D meshes as an alternative representation.

Current sensor developments in LiDAR technology provide 3D point clouds at densities which were unforeseeable until recently. This development to high-density point clouds is further amplified by the increasing availability of UAV-based LiDAR systems. Motivated by these developments, our test data consist of a LiDAR point cloud collected from a UAV-platform at a point density of 800 points/m^2, which potentially allows for applications that were not feasible in the past. Another current direction both in hardware and software development is the integrated capture and evaluation of imagery and LiDAR data. Thus, in our proposed benchmark the LiDAR point cloud is enriched by RGB colour as derived from the mesh texture. The texture is generated by images captured from a camera integrated into the system. In addition to the manually labeled point cloud, which provides a “standard” input for our benchmark on 3D semantic segmentation, a 3D textured mesh is included as an alternative type of representation. For this purpose, the manually labeled classes of the “standard” point cloud are transferred to meshed representation (Laupheimer et. al., 2020), i.e. both data sets are labeled to the same scheme. Such textured 3D meshes integrate information both from images and point clouds, which is beneficial for interpretation purposes. Also due to their advantages for visualisation applications, an increasing number of commercial software systems on multi-view-stereo image matching provide that type of representation as standard output. Our meshes are generated by SURE 3 from nFrames. Finally, multi-temporal data sets provide valuable additional information for semantic analysis in the context of geospatial analysis. While this results in the increasing availability of multi-temporal spaceborne and airborne imagery, there is a clear lack of multi-temporal LiDAR data available to the public. To close this gap, our benchmark consists of data sets collected in March 2018, November 2018, and March 2019 captured over the same area with the same sensor configuration.

Our benchmark data is subdivided into two parts. For one area we will provide reference information, while in the second area we will use the reference for the evaluation of participants' results. Participants are expected to deliver for each point of the point cloud and/or mesh data in the test area a list of XYZ coordinates and a label assigned. This will then be used to evaluate the respective results within the project.

Full project outline

1. Introduction

In recent years, machine learning techniques for automated semantic segmentation of 3D data advanced rapidly. While novel approaches only relying on few labels are currently pursued (e.g. Lin et.al., 2020) & Kölle et. al., 2020), typically, these systems require large pools of annotated data for training and evaluating. However, fully labeled data sets are scarce, which is a major obstacle for the application and acceptance of such systems. This is why existing data, like the ISPRS 3D Semantic Labeling benchmark of Vaihingen/Germany still receives considerable attention, despite the fact that the airborne laser scanning data was captured more than 10 years ago. In contrast, our Hessigheim 3D (H3D) benchmark on semantic segmentation of high-resolution 3D point clouds and textured meshes from airborne LiDAR and Multi-View-Stereo-Image-Matching will be based on data generated using most recent software and hardware technology. Our benchmark will be unique in that it consists of data covering multiple epochs i.e. multiple manually labeled data sets of the same area therefore, and will outperform existing data due to the very high point density of 800pts/m^2. Furthermore, we transfer the labels of the point cloud to a 3D textured mesh representing the same area. Consequently, each epoch of our benchmark consists of two parts:

- H3D(PC): a manually labeled point cloud
- H3D(Mesh): a semi-automatic labeled textured mesh where labels are transferred automatically from the manually labeled point cloud counterpart.

The following sections will explain the details of both our data set and the proposed workflow for building this benchmark set and the corresponding evaluation of results as provided by the participants.

2. Data Set(s)

Figure 1: Our area of interest captures the village of Hessigheim in Germany.

Imagery and LiDAR data for our proposed benchmark were originally captured in a joint project between the University of Stuttgart and the German Federal Institute of Hydrology (BfG) for detecting ground subsidence in the domain of sub-mm accuracy. For this monitoring application, the area of interest which is the village of Hessigheim, Germany (see Figure 1), was surveyed at multiple epochs in March 2018, November 2018, and March 2019. The process of high precision alignment and georeferencing of imagery and LiDAR data, which is also mandatory for joint semantic evaluation, is described in (Cramer et. al., 2018) and (Haala et. al., 2020). In all three epochs, our sensor setup is constituted of a Riegl VUX-1LR Scanner and two oblique-looking Sony Alpha 6000 cameras integrated on a RIEGL Ricopter platform. Considering a height above ground of 50 m, we achieved a laser footprint of less than 3 cm and a Ground Sampling Distance for the cameras of 1.5-3 cm. Georeferencing of acquired LiDAR strips of this highly dense LiDAR point cloud with 800 pts/m^2 is accomplished using the OPALS software (Pfeifer et.al., 2014). Both the LiDAR data and the imagery were additionally georeferenced by a hybrid adjustment (Haala, et.al., 2020). The 3D textured mesh was generated by the SURE software (Rothermel et al., 2012), which integrated the LiDAR data to its Multi-View-Stereo image matching pipeline. Both data types are visualized in Figure 2. Their inherent properties are discussed in detail in sections 2.1 and 2.2.

Figure 2: Subset of our proposed benchmark datasets. The high-resolution LiDAR data H3D(PC) (*left; coloured according to reflectance*) is manually annotated and labels are transferred to the 3D textured mesh H3D (Mesh) (*right*).

2.1 3D Point Cloud. Apart from the XYZ coordinates of each point, LiDAR-inherent features such as the echo number, number of echoes, and reflectance [1] were measured. The latter is derived by range correction of the raw intensity measurement by RIEGL. The LiDAR point cloud is furthermore colorized by transferring colors from the textured mesh. Additionally, we provide a class label for every point (classes will be discussed in section 3). As exchange format of data both plain ASCII files and Las files are provided.

[1] http://www.riegl.com/uploads/tx_pxpriegldownloads/Whitepaper_LASextrabytes_implementation_in-RIEGLSoftware_2017-12-04.pdf

2.2 3D Textured Mesh. We generate the 3D mesh by utilizing software SURE from nFrames (Rothermel et al., 2012). The geometric reconstruction is based on both LiDAR data and multi-view stereo-image matching (using the oblique Sony imagery) to benefit from the complementary information and thus achieve better completeness. These oblique images also provide texture for the resulting meshes and thus guarantee good texturing of vertical faces, e.g. facades (see Figure 2). Since manual labeling is very time-consuming, we opt for semi-automatic labeling of the textured mesh with the manual annotations being transferred automatically from the point cloud to the mesh (see section 3.2). The mesh data is provided in a tiled manner. Each tile is provided both as plain ASCII file (.txt) and Wavefront OBJ (.obj) file. The plain ASCII files provide the centers of gravity (CoGs) for each face along with the transferred label. The .obj files are delivered in a textured and labeled fashion.

3. Generating Ground Truth (GT) Data

3.1 Point Cloud Labeling: Manual Annotation. The main focus of our proposal is to provide labeled multi-temporal and multi-modal data sets for training and evaluation of machine learning systems aiming at semantic point cloud segmentation. For point cloud labeling, we established a manual process already used by student assistants to annotate parts of the epoch March 2018 as depicted in Figure 3. This classification was generated by extracting point cloud segments of unique class affiliation (i.e. the point cloud is cut into many small subsets of homogeneous class membership) and segments of each class are afterwards merged to form the semantic segmentation by the usage of the CloudCompare software (CloudCompare, 2020). We also plan to use this well-established method for annotating the follow-up epochs November 2018 and March 2019. Furthermore, in follow-up epochs, we plan to derive labels for the complete western shore, i.e. the top area in Figure 1.

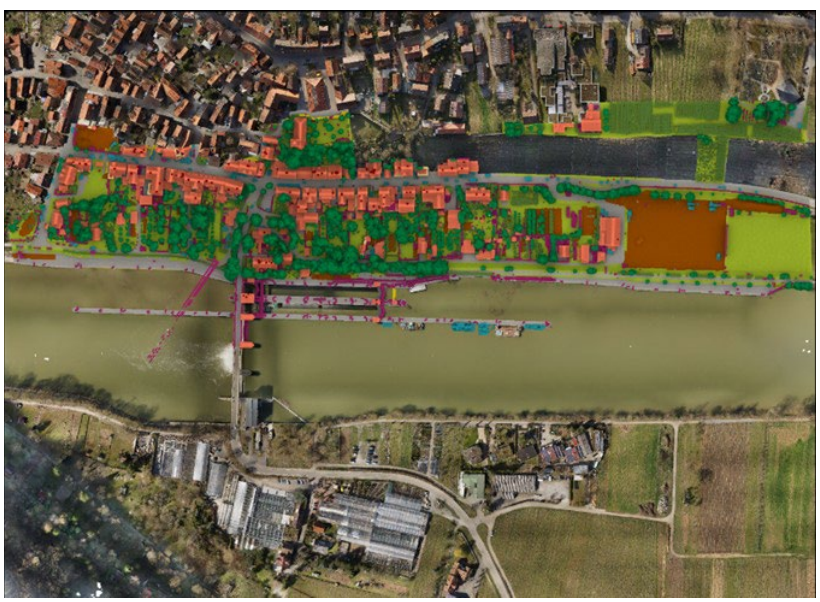


Figure 3: Already available labels in the epoch March 2018. North points to the right.

Figure 4: Detailed view of labels already available for epoch March 2018.

For the March 2018 data set, we manually generated reference labels for differentiating 11 classes (see also Figure 3 and Figure 4):

- <u>Class Catalogue March 2018:</u>

Low vegetation	Impervious Surface	Vehicle	Urban Furniture
Roof	Façade	Shrub	Tree
Soil/Gravel,	Vertical Surface (i.e. “walls”)	Chimney/Antenna	

We plan to generate refined reference considering class structure for the following epochs:

- <u>Class Catalogue November 2018 and March 2019:</u>

Ground:	Low Vegetation / Grassland (sports ground)
	Impervious Surface (Street, path in garden, terrace)
	Soil
	Gravel
Building:	Roof
	Roof Furniture (chimneys, antennas,...)
	Facade
	Facade Furniture (balcony,...)
	Solar panels
Vegetation:	Bush/Hedge
	Tree
Other Man-Made:	Powerline
	Vehicle
	Urban Furniture (trash bin, lantern, bench)
	Vertical Surface (Walls, Lock)
Other:	Clutter (fog/indoor points, humans)

Labeling for epoch November 2018 has already started. In total, 10 student assistants are currently working on this task or have worked on it in the past. Quality control is accomplished in a two-stage procedure. First, student assistants checking each other's labels and finally, the applicants check the labels as last instance. However, we want to stress, that despite a careful quality control, we will not be able to completely avoid label noise.

3.2 Mesh Labeling: Automatic Transfer of Point Cloud Labels to the 3D Mesh. Since manual labeling is very time-consuming, we opt for semi-automatic labeling of the textured mesh for each epoch. The manual point cloud annotations will be transferred automatically by a geometric-driven approach that associates the representation entities points and faces (Laupheimer et.al., 2020). Therefore, the mesh inherits the class catalogue of the manually labeled point cloud. In comparison to the point cloud representation, the mesh is an efficient non-uniform representation requiring only a small number of faces to represent flat surfaces. For this reason, the number of faces is significantly smaller than the number of LiDAR points. Consequently, several points are commonly linked to the same face. Hence, the per-face label is determined by majority vote. However, due to structural discrepancies, some faces remain unlabeled because no points can be associated to them (e.g. absence of LiDAR points or geometric reconstruction errors). These faces are marked by the class label -1. With the help of the labeled mesh data, we want to i) foster semantic mesh segmentation and ii) evaluate the community's interest in this kind of representation at the same time. Our mesh labeling by semi-automatic label transfer from the point clouds will also provide insights on the need for manual refinement during potential future extensions of the benchmark while limiting the effort for the time being.

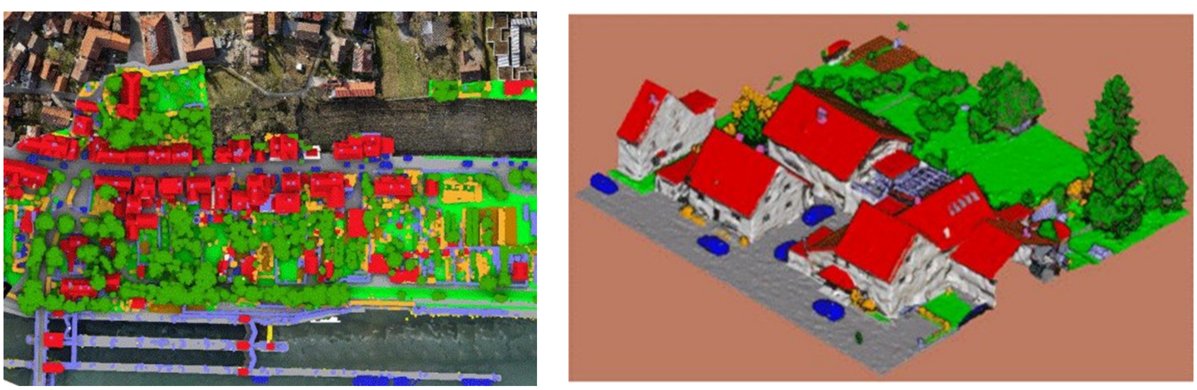


Figure 5: Automatically labeled mesh. Left: overview, right: close-up for a tile.

3.3 Evaluation of results from benchmark participants On top of providing annotated 3D data for research purposes, we will split the data in each epoch into a distinct training, validation and test area for both representations. The splits are congruent in both modalities and in accordance with the mesh tiling. Labels for the test set will be kept sealed, because we plan to continue the benchmark evaluation even after the funding period. We would like to encourage researchers to participate in this benchmark by testing their method on this data set. Precisely, the training and validation labels may be used for training their models. If participants intend to take part in the evaluation process, we ask them to submit their predicted labels for the test area as simple ASCII file either for H3D(PC) or H3D(mesh) (columns [X, Y, Z, classification]). To guarantee a structured exchange, the submission process is managed by the benchmark website, which will be set up and maintained by the Institute for Photogrammetry, University of Stuttgart, where the PI and 2 of the co-Is are affiliated. We will then evaluate the performance of the participants model by comparing the results to the ground truth labels. For this purpose, we will derive the normalized confusion matrix, overall accuracy, F1 scores and mean F1 score, which will be i) returned to the participants and ii) made publicly available in the context of benchmark ranking on our website. To foster joint evaluation, participants are also asked to provide contact details and information on their applied methods i.e. by a short description or link to a recent publication of their approach. This will also allow to track trends of methods, establish a sort of "spam-filter", and will be useful to plan workshops or conference sessions at ISPRS events and papers in ISPRS publications based on the respective outcomes. We would like to stress that we allow multiple submissions for the same authors only if the approaches are different, i.e. repeated submission of results from the same method with differing parametrization will be refused.

Expected outcomes

- Fully labeled point clouds captured by airborne LiDAR are scarce, which greatly hampers algorithmic development as well as application and acceptance of software systems for their automatic analysis due to lacking demonstration of their reliability and resilience. The availability of such data will stimulate research and development in the growing field of semantic information extraction for 3D geodata, which is a core task within ISPRS.
- Our scientific initiative is proposed by ISPRS officers representing three different working groups. The list of applicants is limited to the respective chairs for organizational reasons, however the proposal has been discussed with all officers of these groups. These discussions showed a very strong support for the proposal and already demonstrated the significant interest of researchers organised within ISPRS for our benchmark. We thus expect the link of researchers to ISPRS to be strengthened by our initiative since the benchmark will foster the collaboration under the umbrella of ISPRS. We thus expect the benchmark data being used for high quality publications and sessions during future ISPRS events.
- Labeled mesh data, which is provided in addition to the labeled point clouds will help to foster semantic segmentation of such data and evaluate the community's interest in this kind of representation at the same time. The promotion of such data is of high interest for ISPRS partners from industry, which are currently developing hardware and software systems for the generation of such high quality textured meshes from airborne data capture.
- Point cloud data used for the benchmark were generated by state-of-the-art software and hardware systems. Thus, the proposed benchmark provides 3D data at a coverage, density and quality not feasible until recently. From our point of view, this will also motivate researchers from neighbouring disciplines to use the data and contribute to the benchmark. Thus, the proposed scientific initiative will increase the awareness and visibility of ISPRS far beyond the current state.

Project Milestones and schedule

Point cloud labeling by student assistants including quality control will be organized at the Institute for Photogrammetry, University of Stuttgart. If the scientific initiative is accepted, we will release data for the first epoch (March 2018), where labeling has already been completed, via the first version of the benchmark website in January 2021. We will use this data set for testing the IT infrastructure and the evaluation process. If problems of any sort occur, we can avoid them in follow-up epochs, thus, feedback of participants will be valuable for improving the benchmark. Furthermore, we will also provide a ranking of participating methods via the hosted websites. Our general schedule is displayed in Table 1.

		2021											
	Objective	January	February	March	April	May	June	July	August	September	October	November	December
Labeling	Complete Labeling of epoch November 2018	■	■	■	■	■							
	Cross checks by student assistants and applicants						■	■					
	Transferring Labels to epoch March 2019								■	■			
	Cross checks by student assistants and applicants										■	■	
Milestones	Release of Epoch March 2018	■											
	Release of Epoch November 2018								■				
	Release of Epoch March 2019												■
Bench-marking	Evaluation of submitted results	■	■	■	■	■	■	■	■	■	■	■	■

Appendix – CVs of the PI and all co-investigators (Co-Is)

Norbert Haala is Professor at the Institute for Photogrammetry, University of Stuttgart, where he is responsible for research and teaching in photogrammetric computer vision and image processing. Currently he chairs the ISPRS working group on Point Cloud Generation and is head of EuroSDR commission on Modelling and Processing

Franz Rottensteiner received the Ph.D. degree and venia docendi in photogrammetry from the Vienna University of Technology. He became leader of the research group on Photogrammetric Image Analysis at the Institute of Photogrammetry and GeoInformation (IPI) at Leibniz Universität in Hannover, Germany, in 2008. Since 2014, he has been Associate Professor at IPI, focussing both his research and teaching on

Image Analysis and its applications in Photogrammetry. Franz is chair of ISPRS WG II/4 - 3D Scene Reconstruction and Analysis

Jan Dirk Wegner joined the Photogrammetry and Remote Sensing group at ETH in 2012. He is -founder and head of the EcoVision Lab (9 PhDs and 3 PostDocs), which does research at the frontier of machine learning and computer vision to solve ecological questions. Jan is founder and chair of the ISPRS II/WG 6 "Large-scale machine learning for geospatial data analysis" and (together with colleagues) organizer and chair of the CVPR EarthVision workshops.

Michael Kölle holds an M.Sc. in Geodesy & Geoinformatics from the University of Stuttgart. As a member of the geoinformatics group at the Institute for Photogrammetry, University of Stuttgart, he is currently working on his PhD. His main research interests focus on combining paid crowdsourcing and machine learning techniques such as Active Learning for generating high-quality training data, especially in the context of 3D point clouds.

Dominik Laupheimer holds an M.Sc. in Geodesy and Geoinformatics from the University of Stuttgart. Currently, he is a Ph.D. candidate at the Institute for Photogrammetry, University of Stuttgart. His main research interest is the semantic interpretation of 3D urban scenes as acquired by photogrammetric and LiDAR sensors. His work focuses on the semantic segmentation of meshes leveraging machine learning techniques.

Literature

CloudCompare (version 2.10) [GPL software]. (2020). Retrieved from http://www.cloudcompare.org

Cramer, M., Haala, N., Laupheimer, D., Mandlburger, G., and Havel, P. (2018): Ultra-high precision UAV-based LiDAR and Dense Image Matching, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-1, 115-120, https://doi.org/10.5194/isprs-archives-XLII-1-115-2018.

Haala, N., Kölle, M., Cramer, M., Laupheimer, D. Mandlburger, G. & Glira, P. (2020): Hybrid Georeferencing, Enhancement and Classification of Ultra-High Resolution UAV LiDAR and Image Point Clouds for Monitoring Applications ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-2-2020, 727–734, 2020 https://doi.org/10.5194/isprs-annals-V-2-2020-727-2020

Kölle, M., Walter, V., Schmohl, S., and Soergel, U.(2020): Hybrid Acquisition of High Quality Training Data for Semantic Segmentation of 3D Point Clouds using Crowd Based Active Learning, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-2-2020, 501–508, https://doi.org/10.5194/isprs-annals-V-2-2020-501-2020, 2020.

Laupheimer, D., Shams Eddin, M. H. & Haala, N. (2020) On the Association of LiDAR Point Clouds and Textured Meshes for Multi-Modal Semantic Segmentation ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-2-2020, 509–516, 2020 https://doi.org/10.5194/isprs-annals-V-2-2020-509-2020

Lin, Y., Vosselman, G., Cao, Y., and Yang, M. Y. (2020): Efficient training of semantic point cloud segmentation via active learning, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-2-2020, 243–250, https://doi.org/10.5194/isprs-annals-V-2-2020-243-2020, 2020.

Pfeifer, N. & Mandlburger, G. & Otepka, J. & Karel, W. (2013). OPALS – A framework for Airborne Laser Scanning data analysis. Computers, Environment and Urban Systems. 45. 10.1016/j.compenvurbsys.2013.11.002.

Rothermel, M., Wenzel, K., Fritsch, D., Haala, N., (2012). SURE: Photogrammet-ric Surface Reconstruction from Imagery, in: Proceedings LC3D Workshop,Berlin, p. 2.