Synthetic training data
Accelerating the training of artificial neural networks

Infrastructure mapping and planning: Synthetic training data is evolving into an indispensable prerequisite for the automatic interpretation of 3D geo data.

Automatic generation of data sets for 3D environments

Generating data for training artificial neural networks (ANN) is a time-consuming process. The lack of comprehensive training data sets often proves to be the main obstacle for the use of deep learning in automated data interpretation. Based on a strong background in 3D measuring technology and automated data interpretation, Fraunhofer IPM has developed SynTra, a software tool for generating synthetic training data for specific mobile mapping application scenarios.

Manual generation of training data impedes wide use of machine learning

Machine learning based on ANNs is increasingly used for automatically interpreting large amounts of mobile mapping data. Today, training deep learning algorithms for an interpretation task is a matter of hours and interpretation of 3D or image data is often carried out in real time.

By contrast, the data needed for training an ANN is normally still generated in a laborious process where thousands of data sets must be annotated manually. The effort of generating training data sets for a specific interpretation task, or adding a new object class to an existing data set, often turns out to be disproportionate to the efficiency gain in data interpretation.

### No need for field measurements and labelling

Massive data is needed for training AI models for data interpretation tasks in a geodetic context. High-quality training data for ANNs should include a bulk of typical scenes with a large variety of environmental parameters such as different seasons or lighting conditions. Furthermore, all possible varieties of an object class to be identified must be considered. Procuring this data in field measurement campaigns is time-consuming and incurs high costs. For training an ANN, several thousands of images must be labeled according to the object classes to be identified. Manually annotating one single image takes around 30 minutes, and it bears a high risk of errors. Synthetic training data generation is

The benefits of synthetic training data

- Large amounts of training data in short time
- No need for costly field measurements
- Automatic annotation
- Less error-prone than manual annotation
- Expandable to further object classes
- Fast generation of different urban scenes
- Possibility of generating annotated RGB images and annotated point clouds

Rendered image (left) and automatically annotated image (right)
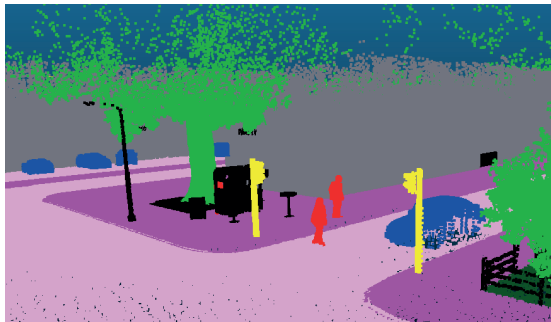
Simulated point cloud (left) and automatically annotated point cloud (right)

comparatively inexpensive and less error-prone, providing a smart alternative to the manual collection of training data.

Use case: suburban scenarios

Fraunhofer IPM has developed tools for artificially creating training data sets with a first use case for suburban scenarios. Urban road environment plays a key role in a broad range of 3D mapping applications, from planning of infrastructure construction to generating urban land registers. The SynTra (Synthetic Training Data) tool creates data sets of urban scenes where target objects are already labeled according to predefined categories. Using the tools, enables creating a multitude of scene variations, which are difficult to get hold of in real life. Different types of data sources or sensors can be represented so that data sets can be created as RGB images or as point clouds including their annotated counterpart.

The tools use open street map (OSM) data as a basis. More than 50 different types of houses have been implemented so far along with a multitude of road surfaces, vegetation types, lightning, sidewalks, etc. Scenes can be created from different perspectives, just like real mobile mapping data is acquired from road vehicles or airborne platforms. The topography of the terrain can be integrated as well. Regional characteristics, weather and lighting conditions, or seasonal variations can be simulated.

Valuable tool for infrastructure planning and administration

Synthetic training data gives access to automatic data interpretation even for users who, up to now, eschewed the effort of annotating large amounts of data and training their own ANN

for their specific interpretation task. This is true for construction companies just as much as for survey providers, local authorities or land registry offices who increasingly use geodetic data for planning infrastructure or generating digital cadastral maps.

Synthetic data for optimizing mobile mapping systems

Whereas using synthetic data will in future make real-life measurement and effortful data annotation for training ANN superfluous to a certain extent, this data may in turn be used as a valuable source for optimizing measuring systems.

Contact

Prof. Dr. Alexander Reiterer
Head of Department Object and Shape Detection
Phone +49 761 8857-183
alexander.reiterer@ipm.fraunhofer.de

Benedikt Rombach
Group Manager 3D Geodata Analytics
Phone +49 761 8857-228
benedikt.rombach@ipm.fraunhofer.de

Fraunhofer Institute for Physical Measurement Techniques IPM
Georges-Köhler-Allee 301
79110 Freiburg, Germany
www.ipm.fraunhofer.de/en