The proposed research regards the description of a scene with geometric primitives for the task of indoor global localization. Our previous research indicated that there only a few methods for scene decomposition and the existing methods are not well suited for the task of global localization. To be useful in the context of global localization, the method should describe the scene geometry accurate enough to enable to match it to the map geometry in order to recover the pose. On the other hand, the description should not contain too many parts to keep the computations feasible. In the presented project we chose to use planar segments as the geometric primitives because they are commonly present in the man-made environments. Therefore, the task can be stated as approximating as much as possible of the scene geometry with features in a form of planar segments. The problem unbreakably related to the detection of features is their comparison. This proposal also concerns planar segment feature matching in navigation and localization context. The initial research revealed that one of the most important aspects is the range of the sensor used. To assure that this requirement is met, we selected a stereo camera pair as our base sensor. To avoid manual parameters tuning we plan to exploit a data-driven approach, i.e. approach where values of parameter are derived from exemplary data. Furthermore, inspired by a recent success of deep neural networks (DNNs) in the field of computer vision regarding semantic segmentation and depth estimation, we chose to adapt this approach to our problem.

Deep learning, despite being a very effective tool for image processing, requires a large amount of human-labeled data for learning. Our goal is to organize learning in a way that minimizes the need for supervision and therefore the amount of labeled data. It is of pivotal importance because collection and annotation of real-world data are troublesome, tedious, and in this case, there is a more than one correct labeling. Methods to reduce this burden will facilitate application in real-world scenarios of global localization that demand training on real-world data, not the synthetic one. We proposed a semi-supervised scheme of learning the geometric primitive detection network. The scheme combines supervision using depth data from 3-D laser scanner with an unsupervised part that assesses image consistency between both cameras, considering detected segments. This research also tackles the problem of comparing planar segments which emerges when detected features are used for localization. Unfortunately, there is also a lack of datasets with proper annotations for this task. For this reason, we propose to use end-to-end learning by indirect supervision. This indirect supervision will be done on a level of agent pose, also verifying the usefulness of the researched methods in the context of finding a correct pose, i.e. localization.

Modern assistive systems are getting more and more autonomous, whether it is a smartphone for personal indoor navigation or a mobile robot that assists visitors at a museum. They are capable of navigating in simple cases or well-defined environments, yet to become truly autonomous they should have a reliable localization module. And the global localization is an essential part of localization because virtually always the localization that takes into account only consecutive displacements will drift away from the correct pose. The aim of global localization is therefore to find a pose of an autonomous agent with respect to the global map, i.e. a map of the environment that includes the whole investigated area, without any previous knowledge about the whereabouts. During the global localization, it is necessary to consider some objects of reference that will be matched between current observation and the map. Till now, most of the research considered salient points as the objects of reference. However, those simple features stand many difficulties during the matching due to their number and low descriptiveness. We, therefore, propose to use higher-abstraction-level features in a form of planar segments. Unfortunately, existing semantic scene segmentation methods are insufficient for localization purposes because they ignore scene geometry and rarely concern planar segments. The novelty of the proposed project lays in combining 3-D geometry, deep learning, and computer vision techniques by including knowledge stemming from known geometric relations into data-driven learning process in the context of global localization.