NGI EXPLORERS

CHALLENGE 46
CMU-TVO-03

Extending TartanVO and Implementing Learning-based SLAM Algorithms & Improving Detect & Avoidance Algorithms and Systems

Air Lab Website: theairlab.org/research/

Meet the expectations of this US Node through the technology challenge described below

GOALS

Visual odometry remains a challenging problem in real-world applications. Geometric-based methods are not robust enough to many real-life factors, including illumination change, bad weather, dynamic objects, and aggressive motion. Learning-based methods do not generalize well and have only been trained and tested on the same dataset. It is widely accepted that by leveraging a large amount of data, deep-neural-network-based methods can learn a better feature extractor than engineered ones, resulting in a more capable and robust model. But why haven't we seen the deep learning models outperform geometry-based methods and work on all kind of datasets yet? We argue that there are two main reasons. First, the existing VO models are trained with insufficient diversity, which is critical for learning-based methods to be able to generalize. By diversity, we mean diversity both in the scenes and motion patterns. For example, a VO model trained only on outdoor scenes is unlikely to be able to generalize to an indoor environment. Similarly, a model trained with data collected by a camera fixed on a ground robot, with limited pitch and roll motion, will unlikely be applicable to drones. Second, most of the current learning-based VO models neglect some fundamental nature of the problem which is well formulated in geometry-based VO theories. From the theory of multi-view geometry, we know that recovering the camera pose from a sequence of monocular images has scale ambiguity. Besides, recovering the pose needs to take account of the camera intrinsic parameters. Without explicitly dealing with the scale problem and the camera intrinsics, a model learned from one dataset would likely fail in another dataset, no matter how good the feature extractor is.

DETAILS

To this end, we propose a learning-based method that can solve the above two problems and can generalize across datasets. Our team's contributions come in three folds. 1) We demonstrate the crucial effects of data diversity on the generalization ability of a VO model by comparing performance on different quantities of training data. 2) We design an up-to-scale loss function to deal with the scale ambiguity of monocular VO. 3) We create an intrinsic layer (IL) in our VO model enabling generalization across different cameras.

SKILLS REQUIRED

Research experience, C++/Python, Linux, OpenCV, Pytorch, Linear Algebra, Machine Learning.

FUNDED BY NGI

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825183