Abstract

Visual odometry (VO) is the process of determining camera position and orientation by analyzing associated camera input image sequences. In recent years, there has been an increase in research utilizing neural networks for an *ego-motion* estimation. DeepVO [1] is one of the first of such methods to use a deep recurrent convolutional network (RCNN) and achieves satisfiable ego-motion estimation. In this thesis, we introduce Hybrid VO, a variation of DeepVO architecture through the use of self-attention, bi-directional long short term memory (Bi-LSTM), and continuous representation of 3D rotation [2]. The introduction of Bi-LSTM and self-attention emphasizes local consistency and augments data with the utilization of both forward and backward sequences. DeepVO and other neural network-based VO methods parameterize rotation with Euler angles and quaternions, hindering the model from convergence with their discontinuities. The proposed model uses a continuous representation of 3D rotation that does not suffer from discontinuities. Similar to DeepVO, Hybrid VO is an end-to-end monocular VO framework without any conventional VO pipelines. It only attempts to model the rotation estimation and estimates translations independently with the geometric 3D-to-2D method. Experiments on the KITTI benchmark show that Hybrid VO outperforms DeepVO and the traditional 3D-to-2D method.