VSO: Visual Semantic Odometry

Konstantinos-Nektarios Lianos [1,*], Johannes L. Schönberger [2], Marc Pollefeys [2,3], Torsten Sattler [2]

[1] Geomagical Labs, Inc., USA [3] Microsoft, Switzerland
[2] Department of Computer Science, ETH Zürich, Switzerland
nelianos@geomagical.com Fjsch,marc.pollefeys,sattlertG@inf.ethz.ch

Abstract. Robust data association is a core problem of visual odometry, where image-to-image correspondences provide constraints for camera pose and map estimation. Current state-of-the-art direct and indirect methods use short-term tracking to obtain continuous frame-to-frame constraints, while long-term constraints are established using loop closures. In this paper, we propose a novel visual semantic odometry (VSO) framework to enable medium-term continuous tracking of points using semantics. Our proposed framework can be easily integrated into existing direct and indirect visual odometry pipelines. Experiments on challenging real-world datasets demonstrate a significant improvement over state-of-the-art baselines in the context of autonomous driving simply by integrating our semantic constraints.

Keywords: visual odometry, SLAM, semantic segmentation

1 Introduction

Visual Odometry (VO) algorithms track the movement of one or multiple cameras using visual measurements. Their ability to determine the current position based on a camera feed forms a key component of any type of embodied artificial intelligence, e.g., self-driving cars or other autonomous robots, and of any type of intelligent augmentation system, e.g., Augmented or Mixed Reality devices.

At its core, VO is a data association problem, as it establishes pixel-level associations between images. These correspondences are simultaneously used to build a 3D map of the scene and to track the pose of the current camera frame relative to the map. Naturally, such a local tracking and mapping approach introduces small errors in each frame. Accumulating these errors over time leads to drift in the pose and map estimates. In order to reduce this drift, constraints between corresponding image observations are used to jointly optimize poses and map, e.g., using an extended Kalman Filter [33] or bundle adjustment [26, 46].

In general, there are two orthogonal approaches to reduce drift in VO. The first uses short-term correspondences between images to enable temporal drift correction by transitively establishing constraints between subsequent camera frames. This is especially useful in automotive scenarios where a car drives along

[*] This work was done while Konstantinos-Nektarios Lianos was at ETH Zürich.