Dynamic Scene Understanding for Autonomous Driving

Autonomous navigation in complex urban environments requires the ability to recover a holistic model of the dynamic scene as it unfolds in time. Dynamic scene models include a 3D semantic map of the structures in the urban environment, such as buildings, roadway, and sidewalks, which define the substrate for the dynamic behaviors of vehicles and pedestrians. This static model is then augmented with a complete description of the dynamic scene elements via target detection and multi-target tracking. I will describe five recent advances which support the construction of dynamic scene models. Our starting point is a method for recovering a 3D volumetric semantic map of a static scene from monocular video. Given a sparse point cloud from structure-from-motion, we recover a dense 3D volumetric map through a global CRF optimization and the use of structural priors. The recovery of scene semantics is enabled by a novel and fast method for semantic video segmentation which uses feature optimization to enforce temporal continuity. We then describe a deep model for object detection in video frames which can serve as the starting point for multitarget tracking. Our detection architecture was the best-performing single model detector in the 2016 Microsoft COCO Challenge. We present a multiple-hypothesis tracking approach to multi-target tracking which incrementally processes video frames and uses efficient on-line appearance classification to differentiate object instances. We close by showing recent experimental results from our AutoRally project, a 1:5 scale open source vehicle platform which supports research in high-speed autonomous driving.