(Semantic) 3D Reconstruction

Marc Pollefeys
ETH Zurich, Switzerland

Abstract

Obtaining 3D geometric information from images is one of the big challenges of computer vision. It is critical for applications such as robotics, autonomous vehicle navigation and augmented reality.

In the first part of this talk, I will provide an overview of the classical 3D reconstruction pipeline and briefly introduce the basic concepts of structure from motion and multi-view stereo. We will go over the tremendous progress in the last two decades and provide application examples, such as fully autonomous micro-aerial vehicles, driverless cars and 3D selfies on mobile phones.

While purely geometric models of the world can be sufficient for some applications, there are also many application that need additional semantic information. In the second part, I will focus on 3D reconstruction approaches which combine geometric and appearance cues to obtain semantic 3D reconstructions. Specifically, the approaches I will discuss are formulated as multi-label volumetric segmentation, i.e. each voxel gets assigned a label corresponding to one of the semantic classes considered, including free-space. We propose a formulation representing raw geometric and appearance data as unary or high-order (pixel-ray) energy terms on voxels, with class-pair-specific learned anisotropic smoothness terms to regularize the results. We will see how by solving both reconstruction and segmentation/recognition jointly the quality of the results for both subtasks can be improved and we can make significant progress towards 3D scene understanding.

Keywords

3D reconstruction, structure-from-motion, multi-view stereo, semantic 3D reconstruction, scene understanding