Abstract

The landscape of 3D sensing has undergone a significant transformation in recent years. While dedicated 3D depth sensing technologies, such as structured-light scanners, are providing accurate depth measurements, the rise of foundational models is blurring the distinction between RGB and RGB-D in many applications. Today, powerful deep learning methods can tackle 3D scene understanding problems directly from standard RGB images, including depth data inference, reducing the need for a RGB-D camera. This shift has expanded the applicability of 3D sensing while introducing new challenges and opportunities. This thesis explores the effective design, implementation, and evaluation of neural models for 3D data processing, considering both traditional and modern depth-sensing approaches. It covers a wide spectrum of real-world tasks, ranging from low-level data refinement to high-level scene understanding. Furthermore, given that depth sensors are still less ubiquitous than RGB cameras, synthetic data generation remains essential for training robust models. Alongside an analysis of the evolving state of the art, this thesis presents novel datasets and methodologies addressing key 3D sensing challenges. Specifically, this work introduces a new neural model and dataset for 6D bin pose estimation from structured 3D data, as well as a pipeline for auto-labeling glass objects for instance segmentation. Moreover, the thesis analysis synthetic data generation with added noise and deployment guidelines for embedded devices, extending the applicability of deep learning solutions to challenging industrial scenarios.

Keywords: machine learning, RGB-D data, point clouds, 3D scanners