Reading Assignments: KinectFusion/NeRF

Submitted by: Shahram Najam Syed
Andrew ID: snsyed

KinectFusion Overview

KinectFusion is a real-time dense SLAM system that accomplishes tracking and mapping using depth data acquired from a single Kinect sensor. The system maintains a volumetric truncated signed distance function (TSDF) representation of the scene, fusing all depth measurements into a single implicit surface model. Camera tracking is achieved by aligning each new depth frame against a raycasted surface prediction from the TSDF using a coarse-to-fine ICP (Iterative Closest Point) algorithm, utilizing all depth measurements rather than sparse features. Real-time performance is achieved through highly parallel GPU implementations of tracking and mapping, operating at 30Hz on standard graphics hardware. Key innovations include frame-to-model tracking (as opposed to frame-to-frame tracking), efficient volumetric integration of depth data, and a surface prediction method based on TSDF raycasting, enabling constant-time operation within room-sized scenes.

NeRF Overview

NeRF (Neural Radiance Fields) represents 3D scenes as a continuous 5D neural radiance field using a fully-connected neural network that maps spatial location (x, y, z) and viewing direction (θ, ϕ) to RGB color and density. This model is optimized end-to-end using only 2D images with known camera poses by projecting the 5D radiance field into 2D images through differentiable volume rendering. NeRF introduces positional encoding of input coordinates to capture high-frequency details and employs hierarchical volume sampling with both coarse and fine networks to efficiently sample relevant spatial regions. The model achieves state-of-the-art results in novel view synthesis of complex scenes, accurately modeling both geometry and view-dependent appearance with only 5MB of network weights, a significant reduction compared to voxel-based methods requiring gigabytes of storage.