SplattingTAPIR: Utilizing TAPIR and 3D Gaussian Splatting for Efficient Robust Reconstruction

Nguyen, Tim (School: Boston Latin School)

The recent rise of foundation models has been shown to have immense power in pattern recognition and performing "zero-shot" predictions. A key field that could greatly benefit from leveraging these pattern recognition techniques is computer perception. Whether it is the Simultaneous Localization And Mapping (SLAM) algorithms found in our autonomous vehicles or Structure from Motion (SfM) algorithms found in 3D mapping, computer perception has played a pivotal in bridging the gap between computers and our 3D world. While these classical techniques are generally robust in perfect scenarios, the imperfect conditions present in the real world, such as texture-less materials, often create critical failure points for these algorithms. We present SplattingTAPIR, a deep-learning twist on a classical visual odometry problem. SplattingTAPIR is a robust and efficient end-to-end pipeline that can transform a set of 2D images into a 3D scene using Google Deepmind's Tracking Any Point with per-frame Initialization and temporal Refinement (TAPIR) network for 2D keypoint detection with a modern twist on monocular pose estimation and 3D reconstruction through the use of the Depth Anything foundation model, SIM-Sync, and 3D Gaussian Splatting.

Awards Won: