Zachary Bunch

An Overview SfM & NeRF

SfM

Structure from Motion is a technique that uses two two-dimensional images to estimate depth based on the shift in perspective. This technique takes advantage of a similar method that we perceive depth (stereo vision). We are able to determine the depth of objects since we have two eyes that view their environment from slightly different perspectives. SfM works in a similar way, it takes the focal length of the camera, the baseline difference between the camera positions, and a left and right image as input. SfM uses the difference in pixel matrices and converts them to a disparity map. The values of the disparity map depends on how much the pixels have moved between images (objects that are closer would move more). Using a simple formula, a depth map can be created. Similar to 3D computer graphics, we can estimate how the resulting scene would look in 3D by applying the depth map to a projection matrix. The aspect that interests me the most about this approach to approximating depth, is the overall approachability. No special tools or hardware are necessary. It also takes advantage of human biology principles in a similar fashion to machine learning.

NeRF

NeRF is a computer vision technique that utilizes machine learning, camera tracking (typically using SLAM), and ray marching to compute real life objects from new camera perspectives. NeRF programs would typically take the camera position, camera orientation, and ray vector data as input. The output would be an image that is rendered from the new camera position and orientation with accurate lighting, material (including both color and opacity), and depth detail. The user would collect data by taking pictures of a scene while taking note of the camera's position and orientation relative to the scene (typically by using a method like SLAM). The program will then use ray marching techniques (in a similar way to ray tracing) to cast rays from the camera to append 3D vertices using a machine learning MLP algorithm to predict the depth and radiance (amount of light) of each point. One thing that I find particularly interesting about this method is the utilization of machine learning in conjunction with ray tracing to make calculations on complex graphics details such as the normal maps and lighting calculations.