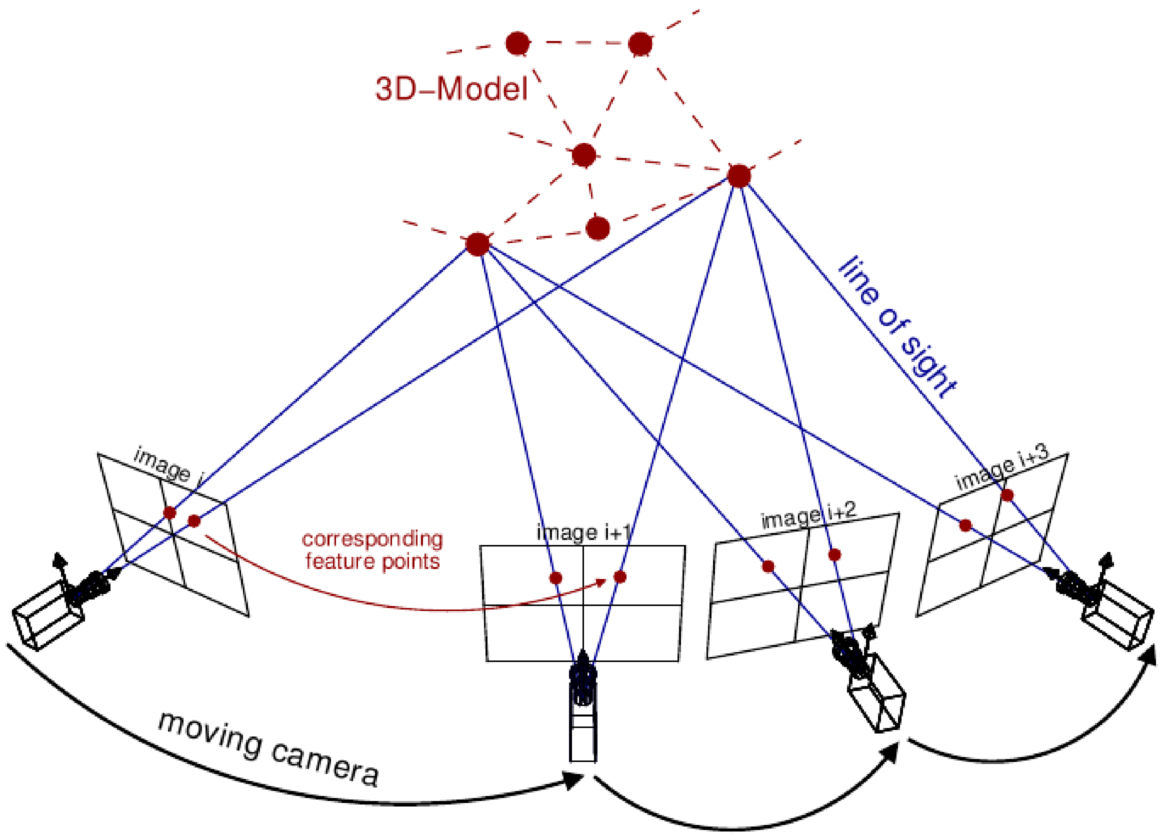
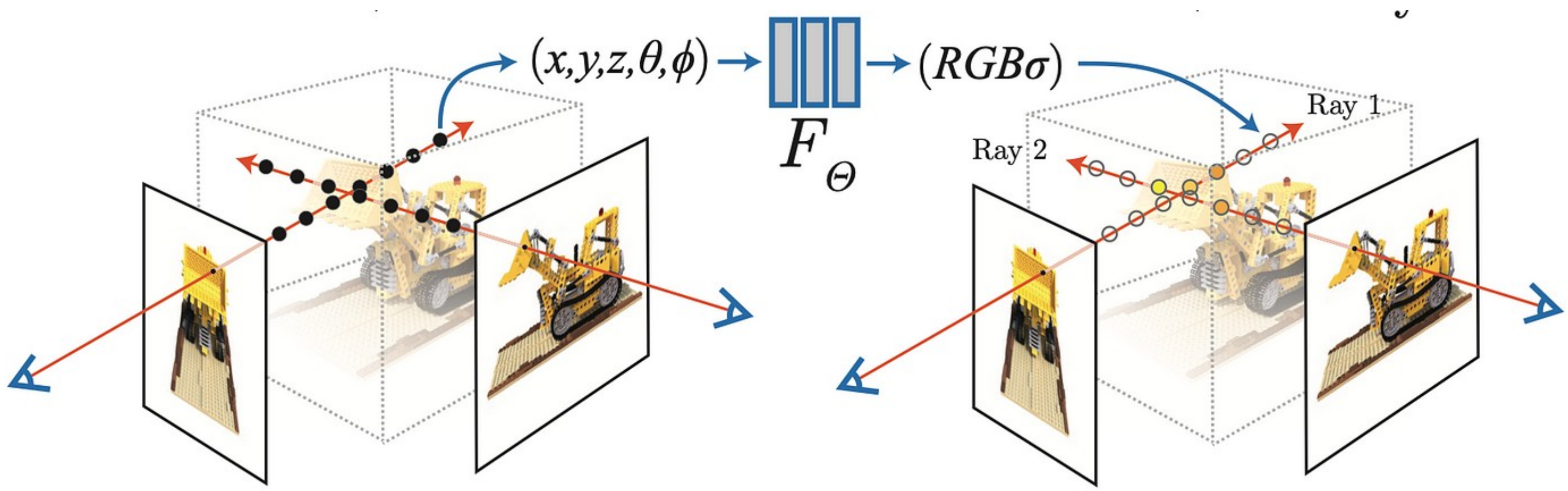
SFM (Structure From Motion)

A set of images captured around an object are processed by bundle adjustment algorithm. The result is a sparse point cloud + images that have 6-DoF poses. One can densify the point cloud, create a mesh, and superimpose RGB data on the generated mesh as post-processing. Rendering a view from an arbitrary location requires projecting the colored mesh into the desired location.



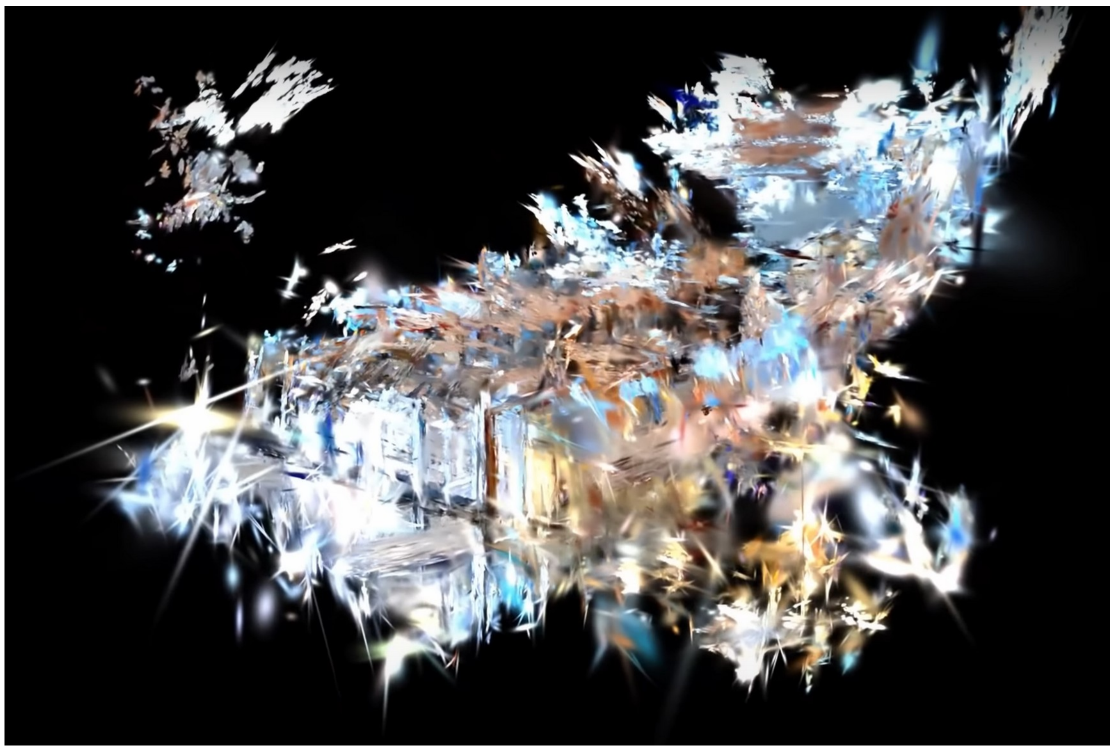
NeRF (Neural Radiance Fields)

A set of images captured around an object are used to overfit a dense neural network. To be more precise, each image must have a pose (often achieved through SFM). For each pixel in each image, a ray is shot and 3d points are sampled in different depths along the ray. The sampled points and the orientation of the ray are fed to a network that outputs RGB and the density. So unlike SFM, NeRF doesn't generate a point cloud. It just encodes the object's geometry as weights of the network (which takes much less disk space!). Rendering a view from an arbitrary location requires 1) shooting a ray for each pixel 2) sampling points along the ray and 3) running an inference. Compared to SFM, the results are much more photo-realistic but the inference time is a bottleneck.



3D Gaussian Splatting

A set of images captured around an object are first fed to SFM to generate a sparse point cloud. A 3D Gaussian (an ellipsoid) is initialized per each 3D point. Each Gaussian encodes the location, shape, color, and other data. We can project the Gaussians to an arbitrary pose and render a new image. To improve the rendering quality, a neural network is trained to fine-tune the 3D Gaussians and minimize the error between the rendered and the input images. Unlike NeRF, this method has an explicit 3D representation of the scene which makes rendering much faster.



Visualized Gaussians obtained from scanning a room