Popular science summary

Transforming 2D Images into 3D Reality: Recovering Objects and Camera Positions

From multiple 2D images, we can recreate a vivid 3D object. Amazing, isn't it? The real challenge lies in the accuracy of this reconstruction. We measure this accuracy using something called 'reprojection errors', which simply put, tells us how closely our 3D reconstruction matches the original 2D images.

The objective of this project is to reconstruct 3D objects from multiple 2D images. This task can be complex, especially when images contain both stationary and moving objects, such as buildings and vehicles. So, the project is divided into two parts based on the object types.

In the first part, we worked with images containing only stationary objects like buildings, parked cars, and trees. We used a new dataset named 'BlendedMVS'. With the help of advanced image processing techniques, we extracted 'point matches' from these images. A 'point match' represents the same 3D point viewed from different images. After compiling these points into an input for our model, we used the network proposed by researchers Moran et al. to automatically generate the 3D objects and camera positions in space. We also add some improvements based on the network to increase the reconstruction quality. The results were promising; we managed to improve the accuracy of the reconstructions, as indicated by lower reprojection errors.

In the second part, we worked with video files, which can be seen as a series of images. This time, the data contained both stationary and moving objects. For this task, we used a dataset called 'Hopkins 155'. In each video, some 2D point matches from both the stationary scene and moving cars were already provided. Since we couldn't reconstruct all of them at once, we segmented them and reconstructed each individually. To do this, we used tools like YOLOV7 and Deep Sort to detect and isolate different objects, such as cars and trucks. By seeing which bounding box a point was located in, we could roughly determine which object it came from. This method gave us highly accurate segmentation results, close to 100%. Using our improved architecture, we achieved reprojection errors below 1 pixel for 'Hopkins 155' dataset, indicating high-quality 3D reconstructions. Figures 1 and 2 are provided here to illustrate the process.

Figure 1: One image from the input image set: multiple angle views of an ox sculpture.

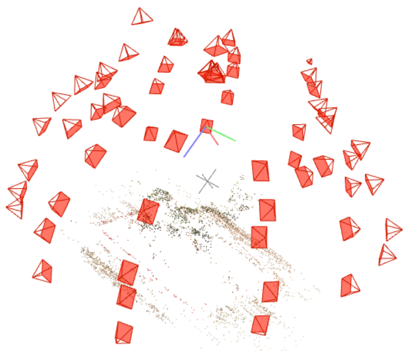

Figure 2: This image illustrates both the 3D point cloud and cameras. The red signs denote camera locations and orientations. The point cloud is surrounded by those cameras. The point cloud is a set of 3D points on the surface of a 3D object.