Efficient Image Selection for 3D Reconstruction

Julian Knodt, Michael Hu, Justin Yan

December 2019

(Note: In this proposal, we use the words "camera", "photo", and "image" interchangeably. For our use case, there is no difference. Each photo / image defines a camera orientation relative to the central object, and vice versa.)

Problem

How does the quality of a 3D object reconstruction from 2D images degrade as we remove images? Which 2D images are most vital for 3D reconstruction, and how can we select them?

Background

Photogrammetry is the field of extracting 3D information from 2D pictures. We wish to better understand how to choose which pictures to take of an object, such that we can produce the most accurate 3D reconstruction possible, given constraints on the number of photos that we can use in the reconstruction. This information is potentially valuable in fields where photo data is rare or expensive to obtain, such as radiology or cosmology.

Approach

Structure from motion (SfM) algorithms recover camera orientations and a 3D scene reconstruction from a stream of 2D images. In 2006, Snavely et. al. adapted the SfM algorithm proposed by Brown and Lowe [1] to create a famous landmark explorer for photo tourism [3]. An updated version of the Snavely algorithm is BigSFM. We plan to use BigSFM as a black box to compute 3D reconstructions from streams of 2D images. If a more modern, improved SfM algorithm exists, we would likely use that instead. We plan to use the 1DSfM dataset for reconstruction and evaluation. 1DSfM contains thousands of photos of famous landmarks. We plan to run experiments on the Nobel computer cluster. Julian Knodt already has access.

We seek to find a minimal subset of cameras that maximizes the 3D reproduction quality relative to the ground-truth reconstruction. A first approach would be to, for a given scene, choose a random subset of k cameras, perform and evaluate the reconstruction, and observe the behavior of the resulting loss as a function of k. We expect reconstruction quality to increase with k, but are also hoping to characterize loss function behavior across scenes.

The next step is to investigate the behavior of the reconstruction loss with respect to the camera processing order. A naive approach is simply to observe the results for a few permutations of image sequences for small k. It's possible that this has been previously investigated, but I was unable to find it mentioned in a paper. If the order makes a difference, we'll need to take a look at some math to see why.

Finally, we'll want to optimize our selection of camera scenes for any k. Intuitively, we would like to choose scenes that expose unique features of the target object. We may try to perform clustering on the entire image set, and then distribute our budget of k images evenly across the clusters. This is a k-means approach, but ultimately we want to take a PCA approach similar to [2] and select the k principal component images for reconstruction.

From these analyses, we hope to obtain some intuition or formal mathematical insight about where to take photos in an online manner. We are interested in the setting where photos are expensive to take. Thus, we hope to design an algorithm or heuristic that, given one photo, decides where to take the next photo.

A More Formal Setting

The Bundler algorithm takes in a 2D stream of n photos and outputs a cloud of points $\in \mathbb{R}^3$. Let p^* represent the point cloud constructed from all available photos of the object. We consider p^* as the optimal, or ground truth, 3D reconstruction. Let $p^{(k)}$ be a point cloud constructed from k photos. $p^{(k)}$ is computed from C, a subset of the n 2D photos of size k.

For each reconstruction $p^{(k)}$, we construct a mapping $f : p^* \to p^{(k)}$. (There likely exists such an algorithm that produces a minimal mapping; we will write one ourselves if one does not exist.) We then define our loss function for $p^{(k)}$ like so:

$$L(f, p^*) = \sum_i ||p_i^* - f(p_i^*)||_2$$

Essentially, we take the L2 norm between each point in p^* and its closest point in $p^{(k)}$. We will evaluate the effect of adding photos to the reconstruction by graphing $L(f, p^*)$ as a function of k, the number of photos used in the reconstruction.

Next, we wish to determine the subset of photos C^* for each k that minimizes $L(f, p^*)$. Concretely, let:

$$\begin{aligned} C^* &= \arg\min_{C} L(f, p^*) \\ \text{s.t. } C &\subseteq [n] \\ |C| &= k \end{aligned}$$

By computing C^*, we hope to obtain some insights about the positioning of cameras relative to the object. For example, should we position cameras orthogonal to one another? At what heights should we place the cameras?

Target Outcome & Fallback

We plan to have 3 main deliverables.

- A visualization of $L(f, p^*)$ as a function of the number of cameras. Through this result, we will better understand how the number of cameras affects 3D reconstruction.
- An algorithm or heuristic that computes where to take the next photo for 3D reconstruction, given all the photos taken thus far.
- An application of our algorithm / heuristic to faces. We plan to take pictures of our friends' faces by following our algorithm's instructions. Suppose we have an upper limit of k photos. After taking k photos, we will compute the 3D reconstruction of the face using our black box algorithm of choice.

In terms of fallback plan, we believe we can guarantee the first deliverable, as it mainly involves analysis of the output files of BigSFM, which we use as a black box. After performing the analyses outlined in **Approach**, we hope to have enough intuition to create some heuristic for where to take the next photo. As a fallback, we may look into stitching the point cloud together into a mesh, or doing some interpolation to refine the reconstruction.

Related Course Topics

- Lecture 4, Corners, Blobs, SIFT
- Lecture 14, 3D reconstruction

References

[1] Brown, M., and Lowe, D. G. Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets. 56–63.

[2] Qu, Y., Huang, J., and Zhang, X. Rapid 3D Reconstruction for Image Sequence Acquired from UAV Camera. *Sensors 18*, 1 (2018), 225.

[3] Snavely, N., Seitz, S. M., and Szeliski, R. Photo Tourism: Exploring Photo Collections in 3D. *ACM Trans. Graph. 25*, 3 (July 2006), 835–846.