Project: Structure from Motion

1 Project Goal

Multiview 3D reconstruction is divided into 2 stages:

1. Finding the pose of each image. This is the main objective of Structure from Motion.
2. Recovering dense 3D model given posed images. Multiview stereo / NeRF apply here.

This project will focus on SfM. We have another project on NeRF. Check it out and decide which one you want to work on.

In our assignments we have laid out the ground work for SfM. You have reconstructed a robot statue of "Optimus Prime" from 2 viewpoints (10 if you have attemped the bonus). For this project we will furnish the remaining pieces of the SfM pipeline: keypoint extraction, matching, RANSAC, and handling multiple views (if you didn't do the bonus problem). We will move beyond synthetic setup and apply SfM on a scene of your choice. It can be a piece of our campus architecture, your apartment, a keyboard, a vase, etc. Use your smartphone camera. TA has a drone. Matt's Robotics lab has a huge drone. Let us know if you have a plan for these.

The most popular open-source software package for SfM is COLMAP [2]. It can be helpful to compare your reconstructed camera poses and point cloud against COLMAP's output. You can refer to its documentation for installation instructions. It is somewhat tricky and not required.

Projects are more open-ended and less guided than homeworks. Expect some rough edges. We will keep posting updated resources on canvas as issues arrive. The final report is evaluated more on efforts than a certain outcome. The spirit is to enjoy the creative possibilities of this technology.

2 Steps Outline

1. Calibrate your camera intrinsics. This can be your phone or a more professional device. Look up the specs and find out what the camera field of view is. It's safe to assume that principal points are at the image center and pixels are perfect squares. As a rule of thumb many smartphone cameras use 60 degrees vertical field of view.
2. In principle we can use Zhang's method for calibration, but printing the checkerboard pattern on letter paper is a logistically formidable task. Let us know if you want to try it. If one volunteer does it other people can share.
3. Extract the image frames if you are taking a video. Taking a video might be easier than separate snapshots. **FFMPEG** is the most widely used tool for video processing. Beware that there will be motion blur. It might not affect things too much. Select the frames wisely.

4. Extract keypoints and descriptors using SIFT. Even in the deep learning era SIFT keypoints are still the gold standard. Descriptors might have evolved with neural net features [1], but the notion of keypoints as scale-space extrema endures.

5. Filter erroneous matchings with RANSAC and normalized eight-point algorithm. We saw in ps3 that 2% incorrect matchings can poison SfM. The bad influence lingers on even after bundle adjustment; the reconstructed statue looks bent. Discarding outliers with RANSAC is critical.

6. Run the rest of your SfM pipeline. Integrate keypoints from multiple views.

7. As we discussed extensively in the NeRF project spec, the main goal of SfM is pose, and the reconstructed geometry is sparse / visually impoverished. If you want something denser, you can use COLMAP's multiview stereo. Contact us if you are interested in getting help with this. It needs to run on GPU.

8. Optional: make a video or a small website to display the findings. There is a web-based point cloud visualizer called *potree*.

3 Project submission

Please describe the pipeline you end up implementing in the PDF writeup. We recommend LaTeX, but other typesetting environments like Word or Google Docs are fine, as long as the submitted document is in PDF. Please submit this PDF (try to keep it shorter than 4 pages) along with the full source code files, and a .tar file containing any additional material we need to replicate your experiments. This may include your own images, etc.

Your writeup should include:

- Precise description of the implementation, written in a way that would allow a fellow student in the course to replicate your design.
- Discussion of any choices you had to make, and an explanation of how and why you made a particular choice there. (E.g., things that can be parameterized in different way; settings for hyperparameters; etc.)
- Qualitative evaluation of the performance of your method, and any thoughts on potential improvements one could make with more time/effort. This includes improvements to the method per se, as well as improvements to implementation (such as speeding it up).
- Discussion of the fundamental limitations of the method as designed and implemented, including any specific failure modes (types of input/reference combinations in which the method is likely to fail).

References

[1] Philipp Lindenberger, Paul-Edouard Sarlin, Viktor Larsson, and Marc Pollefeys. Pixel-perfect structure-from-motion with featuremetric refinement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5987–5997, 2021.

[2] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016.