Rough estimation of interior dimensions using structure from motion techniques

Sander Tiganik
Institute of Computer Science, University of Tartu
Supervisor: Artjom Lind

November 18, 2016

Abstract

This article gives insights into how one might implement a structure from motion pipeline with the goal of extracting a rough estimation of a room. The secondary goal will be to extract the room in such a manner that the walls, floor and objects (simplified geometry) are all complete and do not have unexpected missing or redundant faces.

1 Introduction

Strucutre from motion is a range imaging technique for estimating three-dimensional structures from two-dimensional image sequences [1]. Structure from motion comes in a wide variety of possibilities, starting from image sequences with lots of metadata attached, such as camera position, angle, velocity, laser scanning results etc. up to image sequences that only contain the image and where motion vectors have to be calculated. In this paper we will focus on the latter. In simple terms structure from motion (SfM) is the science on how to reconstruct a three-dimensional object using only images of the object. Structure from motion has a very wide application base, starting from creating three-dimensional street maps all the way to archiving archelogical findings using 3D scanning technology and everything in between. The technology can not only map objects from the outside in (E.g table, chair, house), but also from the inside out (E.g rooms, caves etc.). This is the part that we will focus on in this article. The main goal of the article is to present a structure from motion pipeline that will allow us to extract a rough estimation of a room from two-dimensional raw images without any extra data. There is the extra constraint that the estimation of the room has to be complete (I.e it cannot contain any missing or redundant faces on any object, wall or floor in the room). The term "rough" in this context means that the objects in the room do not have to be perfectly represented, but may be approximated by simpler geometrical shapes (E.g boxes, cilinders etc.). The walls and floorplan of the estimation do have to be to scale with the original room where the images were made.

2 Idea

Structure from motion as an idea can be split into two separate processes. The first being the extraction of the point cloud and the second being the extraction of faces from the point cloud.

A point cloud is a dense set of points. In SfM it is usually a set of three-dimensional points. The points are generated using the input image sequence of the object that one wishes to model. The images are processed to find features on them and then match the features of one picture to the features of another. All similar features are assumed the same and kept. This process is repeated for

all pairs of input images. Once all images are processed, one can use linear algebra to calculate the position of the cameras relative to each other, when the images were takes. By knowing the relative position of the cameras and the overlapping features of all images, one can calculate the relative position of the feature points in three-dimensional space. This is how a point cloud is extracted from a sequence of images.

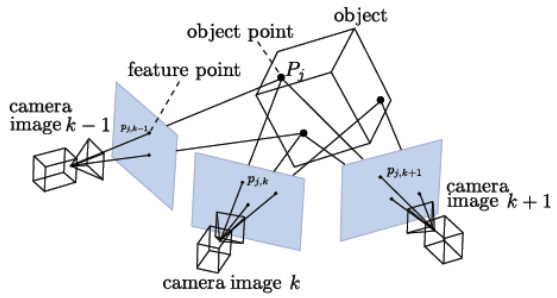

Figure 1: Point cloud generation using matching of features and camera positions [2]

The second part of the process is the extraction of faces. In simple terms one takes the generated point cloud and tries to find polygons on it. By connecting points in the pointcloud in a deterministic manner (Some proposed algorithm) one can reconstruct the object that was photographed. Later textures from the original image sequences can be applied to the polygons to create a model of the object.

3 Structure from motion pipeline

Structure from motion can be layed out by the following pipeline:

1. Collecting the input images
2. Detect features on input images (E.g SIFT [3], SURF [4], Corner detection etc.)
3. Group images together by similarity (E.g nearest neighbor algorithm [5])
4. Match features of pictures with each other (E.g brute-force matching etc.)
5. Calculate camera positions based on matched features
6. Generate point cloud
7. Extract faces from point cloud (E.g Manhattan-world stereo algorithm [6], multi-view stereo reconstruction algorithm etc.)
8. Apply textures from input images.

4 Face extraction

The main focus of this article is point cloud face extraction. Many algorithms have been created (E.g Manhattan-world, multi-view setereo reconstruction [7] etc.) for face extraction. Usually each new algorithm serves a different purpose and has different strengths and weaknesses, but searching the web for articles it is hard to find face extraction algorithms whose strength lie in completeness. There are algorithms for nature objects and algorithms for high-rise buildings, also for interior rooms, but they always try to go for perfection and do not mind if the model has missing or redundant faces. This is something that we in this article are trying to avoid.

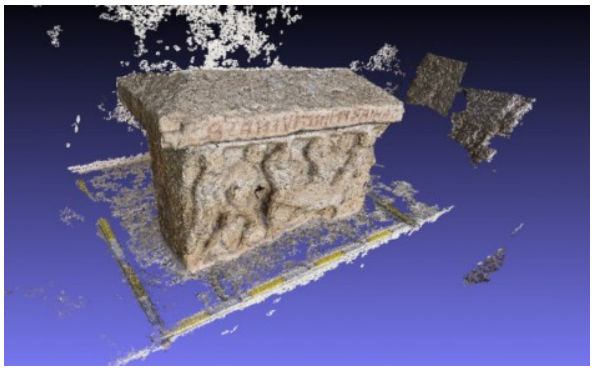

Figure 2: An incomplete model of an object created using SfM techniques [8]

The idea is very simple. Try to simplify the face extraction by making assumptions based on what

we want to achieve. We want to create a model of a room. So a natural assumption would be that all floors (no matter how many levels there are) are all horizontal. Also, although this is a but stretching it, we can assume that all walls are perpendicular in relation to the floor (That assumption may not always hold for all rooms). That gives us a box. Now for the details, all that needs to be done is find clusters of points that represent objects. Working under the rule that objects in the room are represented by simple geometric figures, we can map all point cloud subgroups to rectangles, cilinders or other such simple shapes. Extending the all the objects in such a manner that they touch the floor, or other objects (E.g in case a cup is on a desk). After completing the processing of the pointcloud the result should be satisfactory.

5 Expected results

The expected results of following the plan outlined in this article would be a rough estimation of an interior room which should have no missing or redundant faces and should be to scale with the original room from which the model was created. There probably will be some fine tuning and tweaking that will need to be done in order to make it work exactly as required, but that has to be handled on a case by case basis. The reason why we are talking about expected results is because the implementation for this algorithm has yet to be tested, but will be published in a master's thesis mid next year.

6 Conclusion

As a conclusion we can say that a model created following the requirements specified in this article would be of significant benefit to areas that require a complete model with proper scaling, such as robotics, pathfinding etc.

References

[1] Structure from motion, address: https://en.wikipedia.org/wiki/Structure_from_motion

[2] Structure from motion point cloud calculating, address: http://openmvg.readthedocs.io/en/latest/_images/structureFromMotion.png

[3] SIFT feature detection, address: http://www.scholarpedia.org/article/Scale_Invariant_Feature_Transform

[4] SURF feature detection, address: http://link.springer.com/chapter/10.1007%2F11744023_32

[5] Nearest neighbor algorithm, address: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[6] Manhattan-world stereo, address: http://ieeexplore.ieee.org/document/5206867/

[7] Multi-view stereo reconstruction, address: http://ieeexplore.ieee.org/document/4587792/

[8] Structure from motion, address: http://www.digitalmeetsculture.net/wp-content/uploads/2013/09/Sofia-Menconero-High-resolution-dots-cloud-created-b
jpg