3D Point Cloud Generation Using Structure from Motion with Multiple View Images

Jaeryun Ko, Yo-Sung Ho
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST)
Gwangju, Republic of Korea
{jrko, hoyo}@gist.ac.kr

***Abstract*—Structure from motion is a method of computer vision techniques, which estimates camera motions and generates a three-dimensional point cloud of the object from multiple view images or a sequential set of images of a single camera. In this paper, we explain one of the structure from motion techniques, known as incremental structure from motion, and introduce open source software related to structure from motion briefly. Our experiments show 3D point cloud generation results from executing an incremental structure from motion algorithm of simplified bundle adjustment without outlier rejection, depending on the different number of multiple view images.**

***Index Terms*—Structure from motion, 3D reconstruction, point cloud, camera motion estimation, multiple view geometry**

I. Introduction

Structure from motion is a method of estimating the motion of the camera and the reconstructed three-dimensional (3D) structure of the photographed scene with images taken at two or more different viewpoints. The method can be categorized into two classes in terms of input image types; one is the method with images acquired sequentially through a single camera, such as video sequence, and another method is estimating with unordered image set with different views.

This paper deals with a method to generate a 3D point cloud through the existing structure from motion with images capture at various viewpoints. We describe the incremental structure from motion, a commonly known algorithm among the various versions of structure from motion methods. Then we introduce the related software in which the source code is freely released. In addition, SfM-Toy-Library, one of the publicly available software, was used to execute real structure from motion algorithm to experiment on 3D point cloud generation results.

This paper is extended from a domestic conference paper that was presented in the Korean Institute of Smart Media (KISM) Fall Conference in 2016 [1].

II. 3D Point Cloud Generation using SFM

A. Structure from Motion Pipeline

Structure from motion is a pipelined algorithm in which each sub-task is processed sequentially as shown in Fig. 1, and the incremental structure from motion which contains the iterative reconstruction process is most often used [2].

In the correspondence search step, feature points are extracted for each image. We use the feature descriptors which are invariant to size and rotation so that the structure from motion can easily recognized feature points extracted multiple view images. Basically SIFT and SURF, which are robust feature point descriptors, are widely used for this purpose. Matching process for those descriptors can be implemented with RANSAC for the accuracy of the correspondence. Nowadays ORB feature is famously used for feature extraction and description with lower computational cost than previous features [3]. The descriptor of ORB feature can be expressed into binary representation, so the correspondence matching can be done with Hamming distance comparison, which is faster than other matching methods.

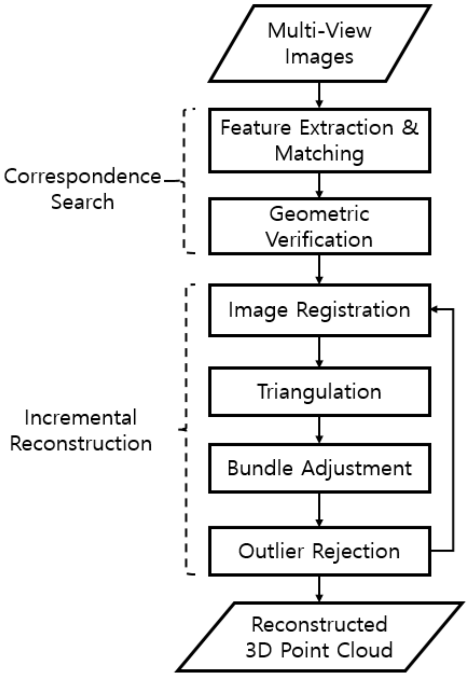


Figure 1. Common flowchart of incremental structure from motion algorithm

Feature point extraction and matching are only for the external shape and do not guarantee the same correspondence with real 3D scene points. Therefore, a geometric verification process is performed to confirm the accuracy of the correspondence point search by the method of estimating the homography between the images. It can be said that a sufficient amount of feature point pairs has been verified only if it is properly projected. A scene graph is created by connecting each image as a vertex with the output data and connecting the verified image pair with a line.

In the incremental reconstruction step, the motion of the camera and the reconstructed 3D point cloud are obtained based on the generated scene graph from the previous step. First, the initial 3D reconstruction model is initialized with the two-view images. The initial model must include the correspondence relation between the feature points on the two-dimensional viewpoint and the points corresponding to the actually observed 3D scene.

Thereafter, the remaining multiple view images are added to the model one by one. Since this process can acquire the extrinsic camera parameters by estimating camera motion information through the Perspective-n-Points (PnP) algorithm [4], the triangulation process acquires a new 3D scene point for the added viewpoint image [5]. These tasks can be easily expressed in Fig 2.

However, the camera motion information and 3D scene points obtained in the previous two processes are often contaminated with outliers due to external factors such as scene acquiring environment. A bundle adjustment process is performed to mitigate the accumulation of the outliers [6]. This improves the coordinates of the acquired camera parameters and 3D scene points using optimization algorithms such as Levenberg-Marquardt algorithm. Finally, an algorithm to remove statistical outliers can be added. After this final process is applied to all multiple view images, the structure from motion algorithm yields the reconstructed 3D point cloud groups and finishes itself.

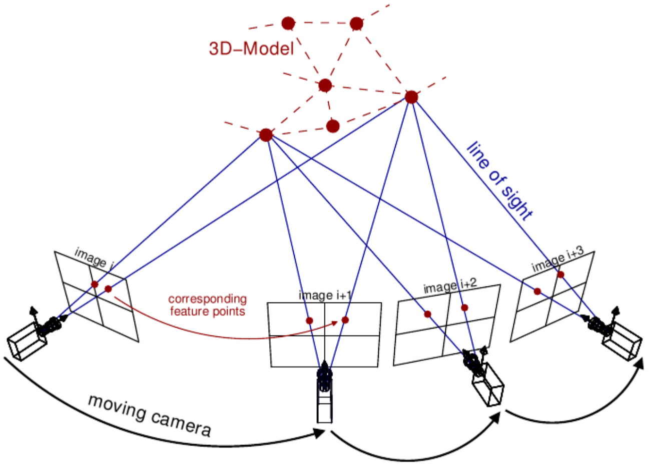


Figure 2. Iterative reconstruction of incremental structure from motion [7]

B. *Open source software related to Structure from motion*

Open source software that implements structure from motion algorithms are available with various versions online. Since it requires various functions, complex operations and reconstruction output into 3D point cloud, it often requires a large number of sub-library dependencies. So there are a lot of software implemented in C++ and we introduce some of the well-known open source software related to structure from motion.

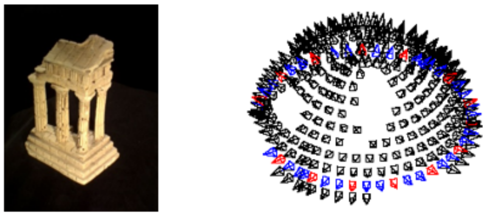
Bundler [8] is the representative software that provides all of the executable version and source code, which is implemented as a GUI and provides various functions for dense reconstruction. COLMAP implements a structure form motion algorithm especially for general purpose [9]. In addition, openMVG [10], a library for implementing structure from motion algorithms is well documented and accessible, including all the dependent libraries for simple installation. OpenCV, the most well-known library of computer vision projects, provides a SfM module as an external contribution since version 3.0, and is easy to install but only available in Linux OS.

III. Experiment Results

The algorithm used in the experiment is based on SfM-Toy-Library, one of the open source software [11]. The source code is written in C++ with OpenCV library.

Fig. 3 shows the dataset Temple and sampled positions of various Temple views as the input images for the experiment. The image set Temple is provided by Middlebury Multi-View Stereo Dataset [12]. Temple set is separated into three sub-sets each named Temple, TempleRing and TempleSparseRing. Temple has 312 views from all the sampled positions on a hemisphere. The subsets TempleRing and TempleSparseRing have 47 views and 16 views respectively, sampled on a ring around the object and the sampled positions are corresponding to the red and blue pyramids in Fig. 3.

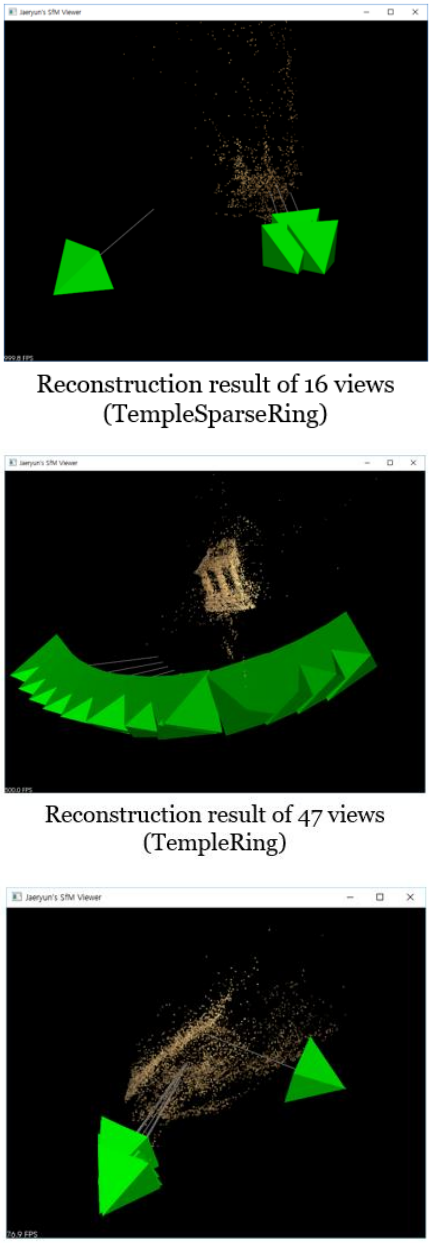
In the algorithm implementation, we used ORB for feature extraction and description and include the outlier removal procedure. We applied the simplified version of bundle adjustment with lower computational cost. Fig. 3 shows that the result of the camera motion and the 3D point cloud are quite distorted at 16 views and 312 views. However 3D point cloud and camera motions are well estimated and reconstructed at 47 views.



Dataset “Temple”

Sampled Position of “Temple” views

Figure 3. Input image set Temple and its view positions for the experiment



Reconstruction result of 16 views (TempleSparseRing)

Reconstruction result of 47 views (TempleRing)

Reconstruction result of 312 views (Temple)

Figure 3. The image set Temple results with different numbers of viewpoints

IV. Conclusion

In this paper, we described the method based on structure from motion to generate 3D point cloud of the acquired scene by using multiple view images as input data and introduced the related open source software. In the existing incremental structure from method, we experimented with inputting a multiple view image set with different number of viewpoints. We used ORB feature for correspondence matching between pairs of multiple view images, skipped the outlier rejection process and applied simplified version of the bundle adjustment to the algorithm.

In the experiment result, we can find out that the smallest number of input images, only 16 views makes bad results of reconstruction and camera motion estimations, since the viewpoints are not enough to match the correspondences of images accurately. The largest number of input images with 312 views yields the worst result with accumulated errors of iterative reconstruction. So the proper number of input images with 47 views only yields the best result with our experiment. Not only this experiment but also for improvements we should consider the characteristics of the image set and the effects of existing outliers.

References

[1] J. Ko and Y. S. Ho, “3D Point Cloud Generation Using Structure from Motion with Multiple View Images,” *The Korean Institute of Smart Media Fall Conference*, pp. 91-92, Oct. 2016.

[2] J. Schonberger and J. Frahm, “Structure-from-Motion Revisited,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[3] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, “ORB: An Efficient Alternative to SIFT or SURF,” *IEEE International Conference on Computer Vision*, pp. 2564-2571, Nov. 2011.

[4] V. Lepetit, F. Moreno-Noguer and P. Fua, “EPnP: An Accurate o(n) Solution to the PnP Problem,” *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155-166, 2009.

[5] R. Hartley and P. Strum, “Triangulation,” *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 146-157, 1997.

[6] B. Triggs, P. McLauchlan and R. Hartley, “Bundle Adjustment-A Modern Synthesis,” *International Workshop on Vision Algorithms*, pp. 298-372, Sep. 1999.

[7] http://www.theia-sfm.org/sfm.html

[8] https://www.cs.cornell.edu/~snavely/bundler/

[9] http://colmap.github.io/

[10] https://github.com/openMVG

[11] http://github.com/royshil/SfM-Toy-Library

[12] http://vision.middlebury.edu/mview/data/