

PAPER • OPEN ACCESS

## 3D Point Cloud Generation Using Incremental Structure-from-Motion

To cite this article: Qi Hu *et al* 2018 *J. Phys.: Conf. Ser.* **1087** 062031

View the [article online](#) for updates and enhancements.

You may also like

- [CDF GlideinWMS usage in Grid computing of high energy physics](#)  
Marian Zvada, Doug Benjamin and Igor Sfiligoi
- [To the question of the possibilities of reducing the microbiological contamination of spices](#)  
M Osipova, A Petrova, K Laricheva et al.
- [High-precision camera pose estimation and optimization in a large-scene 3D reconstruction system](#)  
Lei Yu, Xiaofan Fu, Haonan Xu et al.



The Electrochemical Society  
Advancing solid state & electrochemical science & technology

### 241st ECS Meeting

Vancouver, BC, Canada. May 29 – June 2, 2022



ECS Plenary Lecture featuring  
**Prof. Jeff Dahn,**  
**Dalhousie University**



Register now!



# 3D Point Cloud Generation Using Incremental Structure-from-Motion

Qi Hu<sup>1</sup>, Jianxin Luo<sup>1\*</sup>, Guyu Hu<sup>1</sup>, Weiwei Duan<sup>1</sup>, Hui Zhou<sup>2</sup>

<sup>1</sup>. The College of Command Information Systems, The Army Engineering University of PLA, Nanjing, Jiangsu 210001, China

<sup>2</sup>. The Planning Coordination Apartment, The 31107 Army of PLA, Nanjing, Jiangsu210001, China

luojianxin555@163.com

**Abstract.** Structure from motion is a computer vision technique of estimating camera motions and generating 3D models of the object from a sequence of multiple view images. In this paper, we describe the pipeline of structure from motion strategy especially the incremental approach and also introduce some prevalent open source software implemented structure from motion algorithms briefly. The results of our experiment depending on two different datasets show 3D point cloud and camera motions obtained from a 3D reconstruction program developed by applying bundle adjustment from scratch with the use of OpenCV library.

## 1. Introduction

Structure from motion is a technique of estimating the motion of the camera and recovering three-dimensional (3D) scenes from 2-dimensional (2D) image sequences taken from two or multiple different views of one camera. All the recent work about multi-view reconstruction can be divided into two types: incremental and global approach while the first one is mainly discussed in this paper. The incremental SfM strategy has advantages in accuracy, completeness and robustness which including iteration process to add images one by one. At the last step of the algorithm pipeline, bundle adjustment is implemented to improve the global performance.

We then describe several open source applications, such as Bundler, VisualSFM and OpenMVG, which released the full incremental structure from motion pipeline publicly. Their own characters of each program are also introduced simply. In the experiment part, we adopt a method to execute the real 3D point cloud and estimate camera poses through using the incremental structure from motion algorithm on two image datasets, Fountain-P11 and Crazy Horse sequences. The ORB feature for points extracting and matching, RANSAC for robustness and bundle adjustment for optimization are all implied to generate the 3D model results.

## 2. 3D Point Cloud Generation Using SFM

### 2.1. The Incremental Structure from Motion Pipelines

The incremental reconstruction algorithm [1] as a practical interactive pipeline reconstruct the three dimensional (3D) model depending on the corresponding image points. This incremental SfM strategy is widespread use nowadays as global methods need many images which will allow related estimations create closed loops. In addition, the accurate algorithms for triangulation and camera motions



extraction are contributed to the good performance too. The basic incremental pipeline is shown in Fig1.

The first step is to solve the correspondence problem, which is to extract and match feature points in the series of input images corresponding to the projections of the same point in space. For detection and matching, our main principle of choosing feature descriptors are invariant to size and rotation and meanwhile make a balance of computationally-efficiency so that can suit for real-time performance. Based on FAST and BREIF features [2], ORB feature [3] which has an obvious advantage in computational cost is the most fast binary descriptor nowadays. We know that FAST feature [4] points have no scale invariance, so ORB achieve it by constructing Gauss Pyramid and detecting the corners on each layer of Pyramid images. As for orientation invariance, an intensity centroid method [5] is used by ORB to represent direction through a vector. In order to reduce the loss of Steered BRIEF variance and the correlation between binary code strings, ORB also proposed the descriptor rBRIEF, which uses a learning method to select a smaller set of points, proved to gain better diversity and lower correlation.

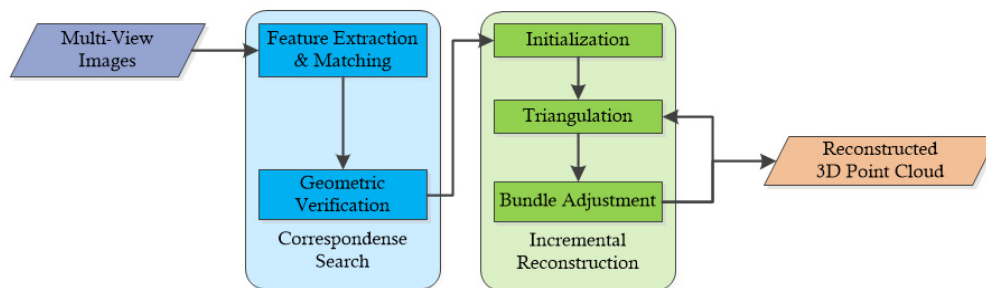


Figure 1 Flowchart of 3D reconstruction pipeline

In the incremental reconstruction step, the stereo initialization is first performed through two-view geometry expressed in Fig2. The initial pair should be selected critically which at least have enough well-matched features in order to avoid a bad initialization. After the two initial views are selected, Longuet-Higgins' work [6] explained how an essential matrix [7] could be calculated using RANSAC algorithm [8] from eight or more corresponded points through solving linear equations. This implementation can be found in OpenCV. And then the essential matrix is decomposed using SVD to estimate the camera positions.

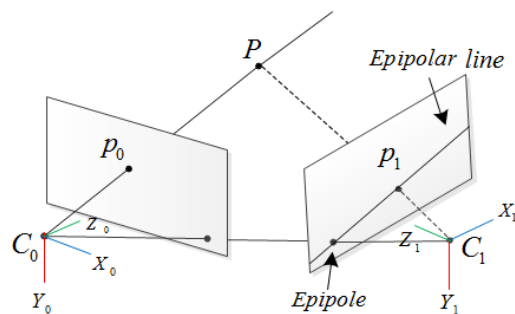


Figure 2 Epipolar constraints between two-view graph.

Once the transformation and orientation matrix between adjacent cameras are calculated and each pair of points' plane coordinates are known, triangulation method is used to obtain the spatial coordinates of 3D points. An implementation of Iterative Linear Least Squares Triangulation in [9] was found to perform better than the OpenCV default implementation.

Thereafter, new images in the remaining multiple-view set can be added to the 3D model scene iteratively by solving the Perspective-n-Point (PnP) problem [10]. The problem can be clearly explained in Fig3. Moreover, RANSAC approach [8] is also used in PnP algorithm for making it more robust to noise.

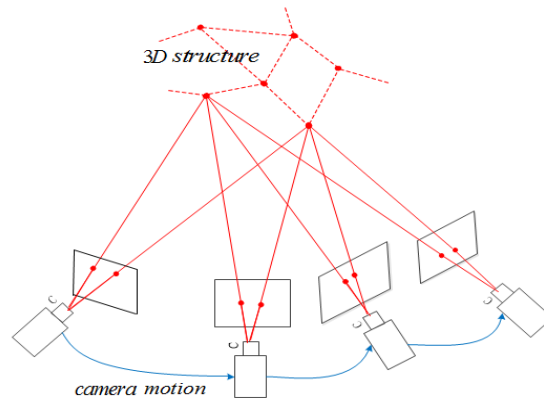


Figure 3 PnP problem in reconstruction process

In the last step, the 3D reconstruction is successfully extended to any number of images, but with the increase of new added graph, the accumulated errors will become a problem affecting the final result. Hence we perform Bundle Adjustment (BA) [11], which is essentially a non-linear optimization, to refine the camera internal parameter  $K_i$ , external parameters  $R_i$  and  $T_i$  of the  $i$ th image as well as point parameters  $X_j$  by minimizing the reprojection error  $E$ .

$$E = \sum_j \rho_j \left( \left\| \pi \left( K_i [R_i T_i] X_j \right) - x_j^i \right\|_2^2 \right)$$

Function  $\pi$  is used to project scene points to image space while loss function  $\rho_j$  such as Huber or Turkey function is designed to enhance the robustness of the algorithm so as not to be influenced by outliers. The general idea to solve BA is implementing the gradient descent method, for instance, Gauss Newton iterative or Levenberg-Marquardt algorithm [12] with its own sparse characteristic. After all multiple view images finished the process, the 3D point cloud scene will be yielded finally.

## 2.2. Open Source Software Related to SfM

Open source software implemented incremental structure from motion technique could be found through the internet and tested on the consistent datasets. We then introduce some of the representative programs which require a large number of sub-library dependencies and mostly use in C++.

Noah Snavely created the software called Bundler in 2007 [13] which is the first open source software implemented incremental structure from motion pipeline. It is designed for large scale datasets that works impressively robust and accurate. Bundler uses an efficient approach to filter reconstructed points by removing too close one. In addition, Bundler only performs bundle adjustment when enough sufficient feature points are matched with the existing model. Actually this method is clever than running optimization every processing time.

Since the target of VisualSfM [14] is to find a more efficient approach without changing the basic pipeline of the incremental SfM algorithm, the processing time is saved from two contributions. Firstly, feature matching is preemptive which only some top-scale features are first matched. Secondly, VisualSfM was one of the first free photogrammetry programs to really utilize the power of GPU (Graphics Processing Unit). In addition, a new re-triangulation step is also introduced to improve the reconstruction quality.

PMVS/CMVS (Patch-based Multi-View Stereo Software and Clustering views for Multi-View Stereo) handles the dense reconstruction after VisualSfM has matched the images. In fact, PMVS and CMVS were the follow up to Bundler as used in Falkingham 2012.

OpenMVG (Open Multiple View Geometry) library [15] has both incremental and global SfM pipeline as the incremental one is based on Bundler. Anyway, Pierre Moulon has proposed an important difference about the RANSAC algorithm which called AC-RANSAC. The main advantage of the AC-RANSAC method is to set the select error thresholds automatically during estimating the essential matrix and camera poses. Besides, OpenMVG is also available on Windows 10 by using the new Bash terminal.

As the most well-known library of computer vision, OpenCV provides a SfM module which is easy to be installed since version 3.0.

### 3. Experiment Results

The results of the experiment were obtained based on the open source software with OpenCV library written in C++. We performed the tests on two different datasets. The first dataset is the Fountain-P11 sequence downloaded at [16] where many standard datasets for SfM can be found. The second one we use is the Crazy Horse in SfM-Toy-Library [16], one of the open source software, including images of the same scene capturing from various camera views.

In the implementation of the algorithm, we used ORB for feature extraction and description and excluded points that did not satisfy Ratio Test. Since its high efficiency and obvious advantages, especially in large-scale problem, the ceres-solver library from Google is applied in the bundle adjustment step. Finally, the results of the camera motion and 3D point cloud are displayed in Fig 4 and Fig5 with two input datasets.



Figure 4 Obtained 3D models and camera poses with the Fountain-P11 dataset



Figure 5 Obtained 3D models and camera poses with Crazy Horse database

#### 4. Conclusion

In this paper, we described the method of incremental structure from motion to generate 3D point cloud and camera poses from a series of multi-view images. We made the brief introduction of some representative open source software related to 3D reconstruction. Based on the publicly SfM algorithm, we experimented two tests with different datasets, including using ORB feature as extracting and matching approach as well as applying sufficient bundle adjustment algorithm.

#### References

- [1] Schönberger J L, Frahm J M. Structure-from-Motion Revisited[C]// Computer Vision and Pattern Recognition. IEEE, 2016.
- [2] Calonder M, Lepetit V, Strecha C, et al. BRIEF: binary robust independent elementary features[C]// European Conference on Computer Vision. Springer-Verlag, 2010:778-792.
- [3] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]// IEEE International Conference on Computer Vision. IEEE, 2012:2564-2571.
- [4] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]// European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2006:430-443.
- [5] Rosin P L. Measuring Corner Properties[J]. Computer Vision & Image Understanding, 1999, 73(2):291-307.
- [6] Longuet-Higgins H C. A computer algorithm for reconstructing a scene from two projections[J]. Nature, 1981, 293(5828):133-135.
- [7] Faugeras O. Three-dimensional computer vision, a geometric viewpoint[J]. Computer Vision & Sensor Based Robots, 1993:187-205.
- [8] Lepetit V, Moreno-Noguer F, Fua P. EP n P: An Accurate  $O(n)$  Solution to the P n P Problem[J]. International Journal of Computer Vision, 2009, 81(2):155-166.
- [9] Hartley R I, Sturm P. Triangulation[C]// International Conference on Computer Analysis of Images and Patterns. Springer, Berlin, Heidelberg, 1995:190-197.
- [10] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[M]. ACM, 1981.



- [11] Triggs, Bill, McLauchlan, et al. Bundle Adjustment - A Modern Synthesis[J]. Lecture Notes in Computer Science, 2000, 1883(1883):298-372.
- [12] Hartley R, Zisserman A. Multiple view geometry in computer vision[J]. Kybernetes, 2003, 30(9/10):1865 - 1872.
- [13] Snavely N, Seitz S M, Szeliski R. Modeling the World from Internet Photo Collections[J]. International Journal of Computer Vision, 2008, 80(2):189-210.
- [14] Wu C. Towards Linear-Time Incremental Structure from Motion[C]// International Conference on 3dtv-Conference. IEEE, 2013:127-134.
- [15] Moulon P, Monasse P, Marlet R. Adaptive structure from motion with a contrario, model estimation[C]// Asian Conference on Computer Vision. Springer-Verlag, 2012:257-270.
- [16] <http://cvlabwww.epfl.ch/data/multiview/denseMVS.html>
- [17] <https://github.com/royshil/SfM-Toy-Library>