

Figure 1: SfM derived point cloud of a rocky highway embankment, Long Lake, Alaska

EMERGING TECHNOLOGY

STRUCTURE FROM MOTION



Figure 2: Zoomed view of embankment

Structure from Motion (SfM) is a method of creating dense point clouds from sets of overlapping images where the point cloud scene (Structure) is created by the rectification of the images from multiple camera locations and orientations (from Motion). This technique combines the theory behind photogrammetric principles with advanced computer vision and image processing techniques to extract features. This combination provides a simple, easy, and rapid form of data collection. Contrasting LiDAR, in which instruments are moderately to extremely expensive, SfM reconstructions of sufficient accuracy for some

BY JOHN D. RAUGUST AND MICHAEL J. OLSEN

Table 1: Benefits and hindrances of SfM technology

Benefits	Hindrances
Ease of data collection	Real world scaling and geo-referencing is subject to human error
Low / no cost	Significant processing time may occur
Semi or fully automated processing	Lower accuracy than TLS
Good color mapping	Noise may need to be cleaned out of point cloud
Point cloud is visually appealing after little or no manual cleaning	Multiple models may be created on different reference frames due to inability to detect common features

applications, can be generated with common, handheld cameras.

SfM first came into use in the late 1970's in the field of computer science. Specifically, robotics and machine learning has made great strides in advancing this technology. Photo tourism has also played a large role with the development of SfM and subsequent marketing to the general public. As with any emerging technology SfM has found its way into several other niches, including: geosciences, geomorphology, archeology, architecture, construction, indoor mapping, and augmented reality, among others.

SfM is largely open-source, and hence, source code can often be downloaded and modified to further expedite, enhance, and tailor this process to specific user needs. An example of this hierarchal structure is shown with an easy to use interface; [Visual SfM](#), which is available as a free download and utilizes the software [Bundler](#), [CMVS/PMVS](#), and [CMPMVS](#); also having several embedded algorithms. However, when using the open source platforms, one needs to have some knowledge of writing code, some time to figure it out, or in most cases, both!

Free services are also available from [Arc3D Web](#), [Microsoft Photosynth](#) and [Cubify Capture](#), which enable a user to

upload images, visualize and, in the case of the latter, download the point cloud. These simple services do not require significant technical expertise. They are designed to be fully automated with very limited user input. Unfortunately, this does not give the user flexibility to “help” the processing or tweak results.

Commercial software packages such as [Photomodeler](#) and [Agisoft Photoscan](#) are also available for relatively low prices; ranging from \$1,150 – \$3,500 and \$180 – \$3,500, respectively. The important question that remains is; just how beneficial are these results?

An overview of some of the strengths and weaknesses of this technology are given in **Table 1**. A complete accuracy assessment of SfM is outside of the scope of this article; however, the reader is referred to the following recent research papers, offering accuracy analyses: [Westoby et al., 2012](#), [James and Robson 2012](#), and [Fonstad et al., 2013](#). In general, SfM is producing results with similar point densities and color mapping as terrestrial laser scanning—TLS. These are visually pleasing point clouds, well suited for presentation uses (see **Figures 1–5**). However, TLS still can

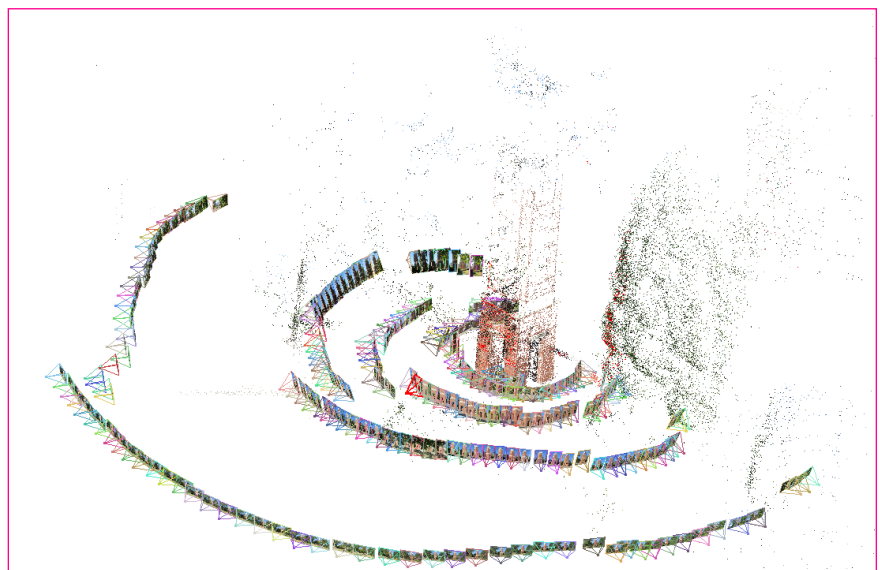


Figure 3: Camera locations, rotations, and resulting sparse point cloud.



Figure 4: Ground truth point cloud (TLS).



Figure 5: SfM derived point cloud.

provide improved accuracy (mm- cm level). The results in this article are derived from open-source packages and lie in the range of decimeter accuracy.


SfM allows the user to more easily cover important areas, when compared to TLS. Rather than setting up multiple scan positions, sometimes taking hours to fully cover occlusions, the user simply

walks around the area or zooms the camera and snaps more photos, where feasible. Use of the two systems together can provide the “best of both worlds.”

General Algorithm Workflow

Once an image set is uploaded into a compiled SfM software, the first step is to use the **SIFT** (Scale Invariant

Feature Transform) algorithm to detect keypoints in that image and log these keypoints in an accessible database ([Lowe, 1999](#)). Keypoints are determined by sharp changes in contrast between neighboring pixels, typically in two, orthogonal directions (object corners). The SIFT algorithm scales an image in both directions at several magnitudes.



Subtracting the lower scale from the higher, generates a Difference of Gaussian (DoG) image. These various DoG images can be compared to the actual scale of each candidate keypoint to determine the closest one, which is used in all subsequent calculations, thus creating scale invariance. Rotational invariance is created by an orientation histogram determined with the actual scale space of the image.

Once keypoints have been identified and logged with keypoint descriptors, keypoints are matched throughout images in the dataset, or, sparse point cloud creation. A combination of 3D triangulation and least squares adjustment to minimize residual errors can be used accomplish this multiple image, bundle adjustment. Accurate and robust keypoint descriptors are imperative to filter out false keypoints. For more information, see an in-depth discussion of the [SIFT](#) algorithm.

Although resection using matching keypoints throughout an image set gives the camera location, pose, and a general look at the scene, it does not fully describe it, as seen in **Figure 6**. To accomplish this, a multi-view stereo (MVS) technique must be employed. Several researchers have developed algorithms, with various approaches and erratic success. A comparison of some of these techniques can be found in a collective, interactive [website](#). To further the development of MVS, it is possible to download the datasets or upload your own code for addition to this comparison. In this article, the previously mentioned CMVS/PMVS, MVS software is used in the Visual SfM GUI. This step will increase the density of the sparse point cloud, sometimes by an order of magnitude or more.


Best Uses and Practices

There are a variety of methods of image collection and options that can evolve from the numerous applications of Structure from Motion. This section will summarize the key concepts from a literature review of uses of SfM and the authors' experiences and testing.


1. Significant Overlap

Probably the most important factor affecting SfM results is the amount of overlap in adjacent images. It is important

at different distances, or zoom levels, from the object. This helps to most accurately determine camera position and pose. However, to create a scene with a dense point cloud, it is also important to have close and distant views of the object or scene in question. To utilize these images effectively, the SfM process must be able to link at least one close-up image with farther away ones, to have it added to the image set. If they cannot link, independent models with different coordinate systems will be created.



“SfM allows the user to more easily cover important areas, when compared with TLS. Use of the two systems together can provide the ‘best of both worlds.’”



to get as much overlap as possible, while keeping in mind that the more overlap (i.e., larger image sets), means more computing time. When results with decimeter accuracy are desired, somewhere between 70-90% overlap is preferable. Less than 70% does not accurately describe the scene, while more than 90% runs into the law of diminishing returns for computing time and feature detection. When the output needed is quantification of simple objects, as low as 40-50% overlap will sufficiently build a 3D model.

2. Fully Link Scene

It is important to have each image in the set verifiable by at least one other image. This can be accomplished by (1) taking multiple vantage point images, spaced a minimal distance apart (to achieve significant overlap), and (2) taking images

3. Scale

Bear in mind that SfM is scale-invariant. Hence, your resulting model coordinates will be unitless and scaled arbitrarily (usually to a unit volume). Hence, in order to extract measurements, one needs to have distinct reference points in the images that have been measured to determine the appropriate scale factor. While some scenes have very clearly defined objects, it is helpful to place targets that can be surveyed in using a total station or GPS.

4. More Megapixels Does Not Guarantee Better Results, Better Cameras Do

It is easy and cheap to find point and shoot cameras with a high level of megapixels. However, with their smaller image sensors and slower ISO levels,



Figure 6: Close-up view of plaque.

resulting images do not contain the rich pixel information that DSLR's do. Therefore, DSLR cameras produce better accuracy than point and shoot ones, by way of less noise, and sharper feature edges. This improved accuracy comes at a high computing cost, however. One way to offset this is to process data with multiple computing cores. If these are not readily available, parallel computing in the cloud can be purchased on a pay as you go basis by [Amazon Web Services](#), or similar. Dense matching can also be set up to run at night, when the computer can be fully dedicated.

5. No Surface Texture = No Return

Features with very little or no surface texture or very low levels of variation do

not bode well for SfM results. Large areas containing materials such as glass, mirror, smooth dirt, and glossy, painted surfaces do not tend to offer point returns. Points are typically visible in areas close to intersections of varying surface textures; however, returns are sparse in the middle. Fully linking these areas to adjacent ones helps with this phenomenon. Some topography can also be difficult to capture since many SIFT algorithms search for sharp, distinct edges.

6. See the Light, But Not Too Much

Keep in mind that the main feature detection algorithm relies on changes in contrast, therefore saturating the scene with darkness or light does not correctly identify features. Timing data

collection with times of crisp, bright lighting will most completely depict features in the scene.

7. Keep the Site in Mind

Perhaps the most important idea to remember with SfM is that each site will have unique properties that must be taken into account. Hence, there is not a "one size fits all" workflow. As more independent and in-depth studies are performed into the various facets of SfM, best practices will be documented and adjusted accordingly.

Conclusion

Structure from Motion is indeed a viable and powerful method of 3D visualization. It has benefits such as ease of use, low cost, and eye-catching results. Before it is regarded as a proven technology, worthy of engineering grade accuracy and precision, SfM will have to grow out of its infancy by standing up to the rigors of the research and commercial communities. Even if you are rooted in historical survey methods, or a firm believer of laser scanning results, grab your camera and give SfM a try for yourself. It can be done for free and in less than an hour. You have nothing to lose but doubt! ■

John Raugust has a B.S. in Civil Engineering from Oregon State University, and is currently in his final term of a M.S. degree, with a project exploring Structure from Motion accuracy and practices. He hails from Bend, Oregon, where he has a background in construction, 3D CAD design, and business.

Michael Olsen is the Eric HI and Janice Hoffman Faculty Scholar in the School of Civil and Construction Engineering at Oregon State University. He currently is the Associate Editor for the ASCE Journal of Surveying Engineering and the Vice-Chair of the ASCE Geomatics Division.