**PFG 2014 / 3, 0189–0195**
Stuttgart, June 2014

**Report**

# A Method of Evaluating the Internal Precision of Multi-View Stereo Dense Reconstruction, Applied on Parthenon Frieze

Dimitrios Skarlatos, Limassol, Cyprus, Stavroula Kiparissi, Athens, Sofia Theodoridou, Thessaloniki & Dimitrios Pandermalis, Athens, Greece

**Keywords:** MVS, precision, computer vision, Parthenon frieze

**Summary:** Recent advances in Structure-from-Motion (SfM) and multi-view stereo (MVS) techniques are getting a lot of attention in the 3D modeling community, as they provide a sophisticated, versatile and rapid way to compute dense 3D point clouds. Many applications in archeology and aerial, surface or underwater surveys have been reported based in the combination of SfM and MVS. In addition, they become more popular and widespread as commercial applications and Internet services are emerging. Nevertheless, there is no known way to evaluate the precision of the extracted point cloud. This paper presents a method for estimating the precision of each point in the final point cloud generated by open source MVS, as it is applied to the Parthenon frieze in the Acropolis Museum.

**Zusammenfassung:** *Methode zur Bewertung der internen Genauigkeit einer Dense-Stereo-Messung, dargestellt am Parthenonfries.* Die neueren Entwicklungen bei Structure-from-Motion (SfM) und multi-view stereo (MVS) sind von großem Interesse, weil sie eine differenzierte, praktische und schnelle Methode zur Erzeugung von 3D-Punktwolken darstellen. Man findet viele Anwendungen, z.B. in der Archäologie und bei der Datenerfassung aus der Luft, auf dem Boden und unter Wasser. Verstärkt wird die Entwicklung durch die vielen neuen kommerziellen Lösungen, aber auch die freien Angebote, besonders im Internet. Es fehlen aber noch allgemein anerkannte Methoden zur Bewertung der Genauigkeit der erzeugten Punktwolke. Dieser Artikel stellt eine Methode zur Bestimmung der Genauigkeit jedes einzelnen Punktes einer Punktwolke vor, die mit MVS erzeugt wurde. Der Datensatz wurde am Parthenonfries gewonnen.

## 1 Introduction

Computer vision lacks a solid definition, as it endorses many subjects, among which 3D reconstruction algorithms hold a prominent position. Such algorithms, focusing on the domain of machine vision, advanced quickly starting from zero, using linear models as mathematical background. Most of them were developed for speed and real time processing rather than metric accuracy (Szeliski 2011). It soon become apparent that photogrammetry could be benefited by computer vision algorithms or vice versa. Nowadays, image based modelling (IBM) emerges as a separate branch in computer vision involving algorithms for Structure-from-Motion (SfM) and multi-view stereo (MVS). Many applications were reported using IBM to record archaeological sites or objects, using open source or commercial software. Nevertheless, there is no report on the estimation of the precision of IBM, while this method is widely adopted by experts and non-experts in a variety for applications from simple shape capturing up to 3D printing reconstruction. Addressing both non-experts to help them avoid misleading results and overestimating technology, as well as experts who

DOI: 10.1127/1432-8364/2014/0219 1432-8364/14/0219 $ 2.00

wish to evaluate results, this paper introduces a method to estimate precision of a point cloud generated using SfM and MVS methodology, so that users may have an estimation of their results. The developed method was applied to 3D point clouds of Parthenon Frieze marble blocks, as this work was a part of the "Digitization, renovation and reconstruction of the Parthenon frieze and offerings" research project, where the precise three-dimensional modeling of Parthenon frieze within an accuracy of some tens of microns was required.

As Parthenon's frieze is a unique world document, which lasted the test of time, many 3D recording attempts have been made in previous years. Attempts have been made to digitally reunify all marbles (Stumpfel et al. 2003) using structured light scanning. The project was involving scanning of all marbles hence the adopted method was focused on productivity, speed, flexibility and resolution of about 1 mm. Remondino et al. (2008) have adopted IBM methods, mostly multi image matching, to produce a 3D model of the Erechtheion temple in Acropolis. Attempts using laser scanners report on the density of points and the accuracy of the model based on the scanner's specifications data sheet. While this as a fact is arguable, the precision on points depends on scanner-object distance as well as other factors which are different for each point. Reports using IBM mention an accuracy based on the block adjustment figures and on the sensor pixel size, while in the case of a scanner many more factors affect the final precision. Nevertheless, none of them, laser scanning or IBM, has metadata information about the accuracy of the model or the raw data (points) themselves.

In this study, the quality of the point cloud was of most importance and the highest possible precision was required. Therefore, the proposed method was developed to estimate the precision of each point of the point cloud using standard photogrammetric techniques. The method reads Bundler (Snavely et al. 2007) and patched based multi-view stereo (PMVS) (Furukawa et al. 2010) geometry output files to calculate the precision of all rays intersection, from images to the point, using least-squares (LSQ) estimation.

## 2 Methodology

### 2.1 *Tools*

When the project started, November 2011, the open source Bundler followed by CMVS/PMVS work flow were available (CMVS = clustering views for MVS). The combination of Bundler and CMVS/PMVS allows the processing of huge sets of random photographs, taken with uncalibrated cameras to produce a dense 3D colour point cloud, with arbitrary scale. The latter was recovered within Geomagic software using scale bar measurements. During the photo-takes, scale bars were placed above and below the marble block, thus they were reconstructed in the final point cloud. In order to compute a single scale factor, multiple 3D distance measurements were taken and the mean scale factor was applied to the point cloud. The photo acquisition process is very important as it ensures multi photo coverage of the object, undercut visibility and density of the point cloud. Differences in scale and rotations of the photos were subtle enough to be recovered using Bundler in our implementation. Bundler is a rather black box operation, which does not provide much functionality while CMVS\PMVS, which were used for dense multi image matching, allow for a number of parameters to be selected so that the user can set and adjust processing with respect to object's needs. Among these factors are the density of the collected points in terms of pixels on the photos, the minimum number of images on which the point must be visible to be accepted and the patch template size.

### 2.2 *Application of Analytical LSQ Model*

The main principle for computing 3D points behind both traditional photogrammetry and MVS algorithms is the same and is well known as stereo triangulation. The 3D location of each point is estimated by modern MVS algorithms as the optimal intersection of multiple pairs of optical rays. Thus, the final position is influenced by errors due to many factors, such as the quality of camera calibration, external orientation parameters of the photos, the image matching quality, the number and geometry of the line intersections in 3D space. Not all of these factors were available through the Bundler and PMVS exported files and therefore estimations had to be made to formulate the design and weight matrices of the least squares. For each reconstructed point, the mean value of the correlation coefficients of all matched projections, is calculated and available in PMVS log file. This value is an indication of how well this point is matched across the images, thus is a good quality indicator which can be used to weight image observations. Low correlation scores indicate areas with no texture, occlusions or strong angles; hence the 2D imprecise measurement will affect the final 3D point localization. For each point the mean correlation value was used as the element of the diagonal of the weight matrix (**W**). Note that the use of different correlation values for each point correspondence would be more appropriate, unfortunately such information was not available.

The camera calibration quality and network geometry of cameras viewpoints, also affect the final precision. The principal point and distortion coefficients affect the precision of point 2D measurements, while focal length error affects the scale factor. Small errors in the latter may be absorbed by the block adjustment, especially if all photos are treated as being taken by separate cameras. By default Bundler treats all photos as they were taken by different cameras, hence automatically calibrates all of them. Since all photos were taken in auto focus mode in order to avoid unfocused images due to variations in the object distance, the Frieze's strong relief and different viewing angles, a different focal length value was computed for each one of the images. The parameters and the accuracy of the exterior orientation of the photos contribute to the final point cloud accuracy as well. 3D positions computed from the intersection of optical rays with a short base may suffer from those errors. Unfortunately, there is no information about the exterior orientation parameters within the Bundler files, hence there is no way to extract meaningful information for **W**. However, even if such accuracy information is not involved in the precision estimation, the values of the above mentioned interior orientation parameters are. Thus, the network geometry is taken into consideration for the computation of the $\sigma_X$, $\sigma_Y$, $\sigma_Z$ values. Finally one single overall accuracy value is computed as the length of a 3D error vector.

The method is fully automated and the only value given by the user is the scale factor of the point cloud, so that precision values are converted in real world units. In particular, the input data of the algorithm are: exterior orientation parameters of all images and their focal length, which are included in the output file of Bundler, the 3D point cloud in the ply-format, a scale factor and additional information about 3D points such as the number of images where each point was visible and the mean correlation coefficient for each point included in the PMVS output patch-file. For each 3D point, the algorithm collects all the information needed and computes an error value as described above. The new ply-file of the original point cloud is exported, where an extra column with the quality measure is added. After that, a filtering approach is applied and all points with an error lower than $2\sigma_o$ are excluded. In both stages, before and after the removal of outliers, statistics are computed. For the visualization of the results the open source software CloudCompare was used.

## 3 Applications

In the context of this work, the Bundler – PMVS pipeline was applied to original marble blocks of the Parthenon Frieze, exhibited in the Acropolis Museum. Before applying the precision evaluation algorithm to an entire block, it was crucial the method be validated in a smaller dataset under controlled conditions, in order to check whether the absence of focal length and exterior orientation standard deviations affect the proposed mathematical model. For this reason a small part of the second marble block from the North Frieze (NF II) was used. After this initial test, a dense point cloud of the sixth marble block from the East Frieze (EF VI) was generated and its precision was estimated using the developed method.

## 3.1 *Verification*

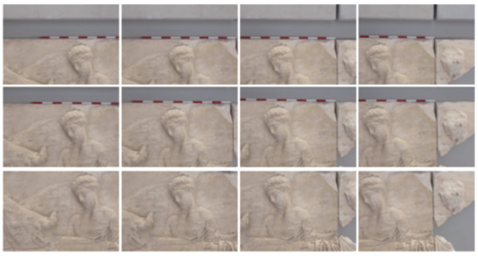

**Fig. 1:** The TST image dataset.

The aim of this test was the validation of the proposed algorithm. Thus the relative precision among controlled datasets was compared. It was expected that the use of more photos and a higher image resolution should improve the precision. A part of the NF II marble block has been used as a test area, forming the TST dataset. Twelve vertical photos (Fig. 1) were taken with a full frame Canon 6D camera (5472 × 3648 pixels, 36 mm × 24 mm CMOS, 6.55 µm pixels size) with a fixed 50 mm lens, from a distance of approximately 1 m.

Using the Bundler output, four point clouds were extracted with different PMVS parameters. The pixel size varied, using the levels 1 and 2, with a minimum of two or four images for the extraction of the point cloud. The latter means that all points which can be correlated in at least two or four photos will be recorded of the final point cloud. The value two means that in the final point cloud points that were found in 2 or more images will be used for the 3D point computation in the final point cloud. All extracted point clouds were processed with the proposed method for the precision calculation (Tab. 1). From the very beginning advantages of the proposed method became apparent, when correlation failed on the featureless background wall, producing spurious points which should be eliminated to produce a meaningful result. Therefore, noise reduction took place by eliminating points with an error larger than $2\sigma_o$. Visualization results for the raw point clouds are presented in Fig. 2, while statistics over precision estimation after noise reduction are presented in Tab. 1. Observing both Fig. 2 and Tab. 1, it becomes apparent that the proposed method works well, since the results are meeting the expectations. In more detail, the error values show that a multiple intersection with four images is more accurate than the simple case with only two images. The above is presented in Fig. 2, where areas with many overlapping images are marked by colours as higher precision areas. The standard deviation values seem to be stable in the case of multi-view stereo while in the simple case they seem to be correlated to the pixel size. Additionally, the maximum estimated error after noise reduction deteriorates fast with respect to a larger pixel size, while the same measure on multi-view stereo remains stable. Once again multi-view stereo filters outliers internally much better. Finally, it has to be noted that in all scenarios, the largest error values appear in the flat grey featureless area above the marble stone, because in this part the image matching may return wrong correspondences. The fact that PMVS computed 3D points in this area, means that the correlation values were near 1.00, as the texture of two windows may be identical but the windows are still not be the same. The above is an example of outliers that

**Tab. 1:** Estimated precision on TST dataset. Window size decreases along with subsampling level so that the correlation area remains the same on every test.

| | | | | | |
|---|---|---|---|---|---|
| pmvs parameters | level | 0 | 0 | 1 | 1 |
| | cell size (pix) | 2 | 2 | 1 | 1 |
| | window size (pix) | 15 | 15 | 7 | 7 |
| | min no of images | 2 | 4 | 2 | 4 |
| | no of 3D points x103 | 1468 | 1270 | 1529 | 1255 |
| filtered | mean σXYZ (mm) | 0.385 | 0.346 | 0.413 | 0.346 |
| | std σXYZ (mm) | 0.244 | 0.108 | 0.315 | 0.111 |
| | max σXYZ (mm) | 3.969 | 1.056 | 4.020 | 1.078 |

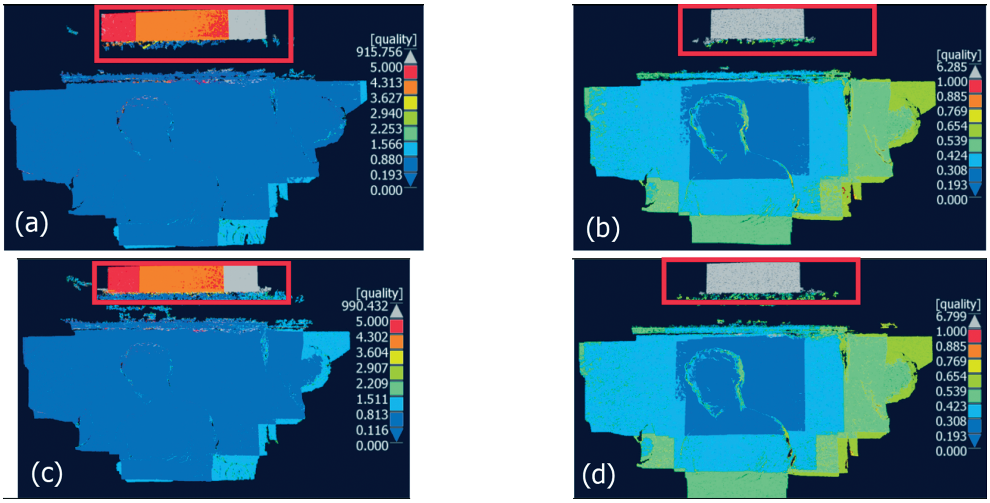

**Fig. 2:** Visualization results from TST dataset before noise reduction. Marked in red, the wall where correlation fails. a) level 0, min no of images 2, b) level 0, min no of images 4, c) level 1, min no of images 2, d) level 1, min no of images 4.

PMVS failed to filter but the proposed method detects them as uncertain 3D points.

## 3.2 *Application on a Frieze Block*

EF VI is a long fragmented block and its left part which depicts the gods Poseidon, Apollo and Aphrodite has a length of 1.40 m. This part of carved stone is a well-preserved scene showing the gods cloths with nice sharp folds and many surprising details. During the project a very dense point cloud of this block was generated and the proposed method was used in order to compute its precision. For the 3D digitization, 330 images with convergent geometry were taken with the Canon 6D camera with 50 mm lens, from an average distance of 0.70 m. In order to ensure the adjustment in the processing phase, additional photos were taken from longer distances, i.e. up to 2 m, to act as bridges for the more detailed images. A dense aerial imagery layout was used to ensure full coverage of the marble block. Additional convergent photos were employed from approximately 0.7 m distance and longer to ensure that undercuts would be visible in at least five images. The data acquisition phase is very important to the results of MVS both in terms of final product density, completeness and precision. Although the main guidelines for the block layout are similar to what is known from aerial photography, the solution of undercuts and complex geometry with convergent or oblique photos cannot be defined by strict rules as it is depended on the photographer’s experience. Unfortunately, MVS is a post processing technique, so that incomplete areas may reveal afterwards. In this application, the following PMVS parameters were used: image pyramid level 1, cell size 2, window size 7. The final result contained a dense point cloud with 17 million 3D points, corresponding roughly to 20 points per square mm. At a first glance, the resolution of the reconstructed point cloud meets the accuracy and completeness requirements.

# 4 Results and Conclusions

The proposed method applied to EF VI returns a mean error of 0.193 mm, standard deviation equal to 0.118 mm and maximum error 9.556 mm. In Fig. 3, where the initial error values are visualized, it can be seen that points with such large errors (~ up to 1 mm coloured with red) are very few and that almost the en-

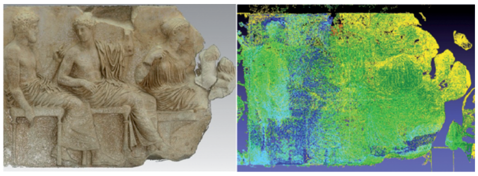

**Fig. 3:** EF VI point cloud (left) and visualization of precision results before noise reduction (right).

tire model has a precision better than 0.5 mm (points coloured with blue, green and yellow). Fig. 4 shows that 54% of the 3D points have estimated precision of 100 µm – 200 µm, 21% of 50 µm – 100 µm and 20% of 200 µm – 300 µm which means that 90% of the point cloud has a precision between 50 µm and 300 µm, which is acceptable.

In this case, a filtering approach was applied. Once again, points with an error larger than two times the sigma zero were detected and rejected considered as outliers. After this step, error values decreased significantly, especially the maximum. Specifically, the mean error was 0.172 mm, the standard deviation was 0.060 mm and the maximum error fell to 0.658 mm, ensuring that the objective of digitizing Parthenon Frieze at the tens microns level was reached. Note that after the noise reduction no 3D point has an error value over 1 mm. The mean precision has not changed significantly while the fact that the new standard deviation value is equal to 60 µm means that all points enjoy a similar precision and no areas are vastly better or worse.

The proposed precision evaluation method has been tested under controlled conditions and applied to a Parthenon's frieze marble block as a proof of concept. The algorithm does not take into consideration external and internal orientation errors as they were not available. Nevertheless, the proposed model uses all available information about the precision of correlation among images and the geometry of intersected 3D lines in a point to point basis in order to calculate the quality of the extracted point cloud. The final result proved to be an excellent evaluation tool for comparing a number of factors affecting overall precision and accuracy. In the Parthenon frieze project in particular it was used to optimize data acquisition in terms of distance, camera resolution and photo layout, realize strong and weak points of MVS, compare it to the main laser triangulation technique used and calculate overall precision of our modelling techniques and document with metadata information the digital frieze.

Further research should focus on the evaluation and the use of the standard deviation of the external orientation results as well as a better documentation of the 2D precision evaluation, in order to improve the results. The use of external reference data for an overall accuracy estimation and a comparison to precision estimation is also under investigation.

## Acknowledgements

This research was supported by the 09SYN-62-629 project "COOPERATION 2009 – Partnerships of Production and Research Institutions in Focused Research and Technology Sectors". The public expenditure is co-

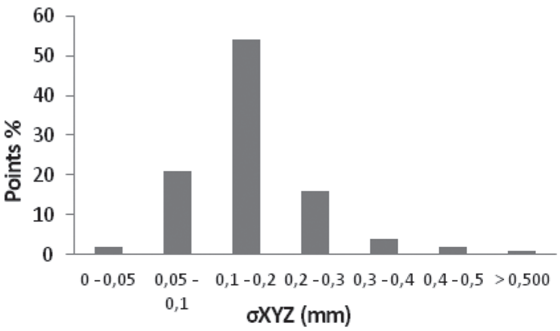

**Fig. 4:** Estimated precision of EF VI dataset.

financed by the European Regional Development Fund (ERDF) of the European Union and by National Resources. We are grateful to the senior conservators of the Acropolis Museum D. Maraziotis, C. Vassiliadis and others in team. We are also grateful to the Director of the Laboratory for Machine Tools & Manufacturing Engineering in AUTH Prof. Dr.-Ing. habil. Dr.-Ing. E.h. K.-D. Bouzakis.

## References

Furukawa, Y., Curless, B., Seitz, S.M. & Szeliski, R., 2010: Towards Internet-scale Multi-view Stereo. – IEEE Conference on Computer Vision and Pattern Recognition: 1434–1441.

Remondino, F., El-Hakim, S., Baltsavias, E., Picard, M. & Grammatikopoulos, L., 2008: Image-based 3D modeling of the Erechteion, Acropolis of Athens. – International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **37** (B5): 1083–1091.

Snavely, N., Seitz, S.M. & Szeliski, R., 2007: Modeling the World from Internet Photo Collections. – International Journal of Computer Vision **80** (2): 189–210.

Stumpfel, J., Tchou, C., Yun, N., Martinez, P., Hawkins, T., Jones, A., Emerson, B. & Debevec, P., 2003: Digital reunification of the Parthenon and its sculptures. – 4th International conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage: 41–50.

Szeliski, R., 2011: Computer Vision, Algorithms and applications. Springer.

Addresses of the Authors:

Prof. asst. Dimitrios Skarlatos, M.Sc. Stavroula Kiparissi & M.Sc. Sofia Theodoridou, GeoAnalysis S.A., G. Genimata Str. 54, GR-55134 Thessaloniki, Greece, Tel. +30-231-049-6800, Fax. +30-231-049-6880, e-mail: dimitrios.skarlatos@cut.ac.cy, {skiparissi}{sofia}@geoanalysis.gr

Prof. em. Dimitrios Pandermalis, The Acropolis Museum, Dionysiou Areopagitou Str. 15, GR-11742 Athens, Greece, Tel. +30-210-900-0900, e-mail: presidentsoffice@theacropolismuseum.gr

Manuskript eingereicht: Januar 2014
Angenommen: März 2014