Mediterranean Archaeology and Archaeometry, Vol. 14, No 4, pp. 125-133

THE ART OF IMPLEMENTING SFM FOR RECONSTRUCTION OF ARCHAEOLOGICAL SITES IN GREECE: PRELIMINARY APPLICATIONS OF CYBER-ARCHAEOLOGICAL RECORDING AT CORINTH

Thomas E. Levy[1], Matthew L. Vincent[1], Matthew Howland[1], Falko Kuester[2], and Neil G. Smith[3]

[1]Department of Anthropology and Qualcomm Institute, University of California, San Diego, USA
[2]Department of Structural Engineering and Qualcomm Institute, University of California, San Diego, USA
[3]Visual Computing Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Received: 18/12/2013
Accepted: 25/06/2014
Corresponding author: Thomas E. Levy (tlevy@ucsd.edu)

ABSTRACT

Cyber-archaeology represents the marriage of archaeology, computer science, engineering, and the natural sciences with the aim of taking advantage of constantly evolving technologies for digital data capture, curation, analyses and dissemination. Digital data collection tools are perhaps the most rapidly changing arenas of development in cyber-archaeology and are becoming affordable tools for every archaeologist. In this paper, we examine two users' approaches to produce point cloud models of archaeological sites using structure from motion (SfM) photography. The experiment took place at the Fountain of Peirene in ancient Corinth, Greece. Their implementation of the technology and their results are compared to highlight the very important role the photo-shooting session can play in the final outcome of the SfM reconstruction. We correlate the users' approaches to the applied algorithms' robust features and known limitations to provide a technical explanation of how archaeologists can significantly improve their success in SfM. As new algorithms and software emerge making SfM a common tool in archaeological documentation the methodology presented in this paper will enable archaeologists to meet the high demand for digital documentation on a global scale.

KEYWORDS: Cyber-archaeology, digital data collection, Structure from Motion, SfM, Corinth

1. INTRODUCTION

The past decade has seen an exponential growth in interest by researchers in the application of digital technologies to archaeological research. The reason for this has to do with the vast quantities of data collected during the archaeological process; the relatively low cost of digital tools such as surveying equipment, computers, portable storage units, cameras, and more; and the need to find ever more efficient ways of collecting, storing, analyzing and sharing those data. Cyber-archaeology provides a solution by developing an integrated system for data capture, curation, analyses and dissemination using traditional print-based systems, 3D visualization and the internet (Figure 1; Levy 2013). The University of California, San Diego Levantine and Cyber-Archaeology Laboratory has contributed to the development of cyber-archaeology since its Edom Lowlands Regional Archaeology Project (ELRAP) in Jordan 'went digital' in 1998 (Levy, et al. 2001) developing an integrated geo-spatial collection and curation for field archaeology (Levy and Smith 2007). The Greece Cyber-Archaeology Collaboratory Project was initiated in September 2013 to compare, contrast and improve digital methodologies used in projects with long-term, sometimes full-time, research endeavors with those developed by ELRAP that is characterized by two-month long excavations over one to three seasons. The American School of Classical Studies in Athens (ASCSA) is an ideal partner to organize this collaboration. Founded in 1881 by a consortium of American Universities, the ASCSA began excavations at Ancient Corinth in 1896, which are one of the oldest continuing excavations in the world. Excavations at Corinth have revealed a vast Roman metropolis that for five centuries was one of the most important cities in the ancient Mediterranean world. As our field season in Greece was only five days, we deployed two data capture technologies – Structure from Motion for rapid 3D documentation of ancient monuments and CAVEcam stereo photography that are described below. Here we report only on the SfM results from Corinth.

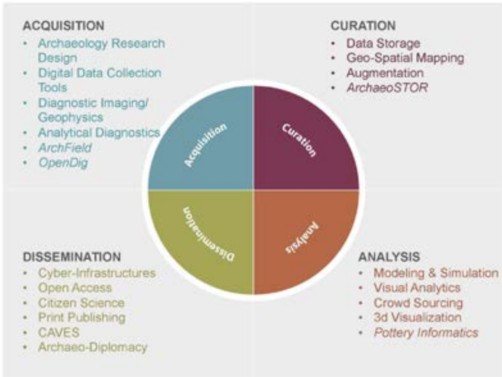

Figure 1 Model of Cyber-Archaeology system.

2. EXPEDIENT 3D DATA CAPTURE – STRUCTURE FOR MOTION (SFM) AND BEST PRACTICES

We used Structure from Motion (SfM) to create rapid 3D digital models of architectural complexes on the Greek mainland using a variety of DSLR cameras (Nikon D80, Canon 50D, and Canon 30D). SfM reconstructed models provide a situated 3D context for all the artifacts, architecture and loci we record using *ArchField* (Smith and Levy 2012) or other GIS-based recording systems.

'Structure from Motion' refers to the method of extracting a 3D structure from many overlapping digital images. Rather than standing in a fixed position and capturing 3D data, the method uses a change in camera position for each image to find the distance (motion) between them and at the same time triangulate the 3D positions of pixels matched in overlapping images. The more motion and movement around the site, the more complete the 3D model becomes. SfM is composed of several computer vision algorithms.[1] Scale-

[1] There are many variants of the computer vision algorithms used in the SfM pipeline. Proprietary software such as AgiSoft Photoscan ((http://www.agisoft.ru/) and Pix4D (http://pix4d.com/) use their own algorithms.

invariant feature transform is used to automatically detect unique features in the images (Lowe 2004). SIFT creates descriptors for each feature that enables it to locate the same feature in other images even when there is a change in scale, rotation, position, or lighting. Bundle Adjustment uses these matched SIFT features and an estimate on the camera focal length to solve for the camera position, rotation, radial distortion, and actual focal length for each image using least squares approximation (see Snavely et al. 2006). As it estimates camera positions for each image it is estimating the 3D position of the matched SIFT features. These features become a sparse set of 3D points representing the general structure of the captured scene. As a final stage a MultiView stereo algorithm like PMVS (Furukawa and Ponce 2007) is used to generate a dense collection of 3d points using the now calibrated position of the images and the sparse 3D points.[2]

The collection of matched pixels and their calculated 3D positions become a cloud of millions of 3D points, called a point cloud. From a distance the point cloud appears as a solid model similar to 3D models seen in CAD programs or video games, but as you zoom in it becomes clear it is actually a collection of millions of points.

With SfM, between 5-20 million 3D points are captured, enabling the recreation of excavation surfaces and architecture digitally. Although the resolution is much lower than a laser scan, it is much faster, easier to perform, and vastly more accurate than hand illustrated plans. In order to meet the demands of archaeological documentation, we have been working at UC San Diego and KAUST to push SfM's capabilities to its limits (Levy, et al. 2012).

2.1 Ancient Corinth

The site is located in the northeast corner of the Peloponnese at the head of the Gulf of Corinth and was referred to in antiquity as one of the fetters of Greece, guarding as it did the narrow land bridge that connects the Peloponnese with the Greek mainland, and providing access to both the Gulf of Corinth to the north and the Saronic Gulf to the east. This strategic position was one of the keys to its prosperity, especially as a Roman city. Excavations by the American School of Classical Studies at Athens have continued for over a century with little interruption until today. The primary focus of excavations has been on the area of the Roman Forum, located within the west side of the village and south of the hill surmounted by the mid-sixth c. B.C.E. Temple of Apollo. This dominating monument has been one of the only features of the site visible since antiquity. For our SfM experiments, we focused on the Fountain of Peirene, an impressive monument described by the Greek historian Pausanias, and according to Greek myth, was a favorite watering hole for Pegasus, the winged horse that was the offspring of Poseidon, the gods of water and earthquakes, and Medusa, the Gorgon female creature. The Fountain is east of the Agora at Corinth and one of the most important fountains of ancient Greece. It represents a challenge for any 3D imaging project because it contains remains from many construction periods with different features ranging from a façade of natural rock to free-standing marble columns from the Byzantine period (Robinson 2011) so it is difficult to parse out these metadata in a scan.

Although SfM is being rapidly adopted by archaeologists and cultural historians due to the development of user-friendly software such as Agisoft Photoscan (http://www.agisoft.ru/), the quality and

In this paper we used VisualSfM, an open source software that uses siftGPU (Wu 2007) and MPBA (Wu et al. 2011) for accelerated GPU processing.

[2] There are two main approaches to MVS: 1. Patched based PMVS (Furukawa and Ponce 2007), CMPMVS (Jancosek et al. 2011); 2. Semi-Global Matching (SGM) (Hirschmüller 2008) used by SURE (Rothermel et al. 2012).

accuracy of the results is still predominately dependent on the quality of the images and the capture methodology employed. As an experiment we captured the Fountain of Peirene with two different cameras (Nikon D90 and Canon EOS 30D) and two users. In this way, we can compare the two final reconstructions, pinpoint how different users approach the application of the method and determine what practices lead to the best results in light of the known unresolved computer vision problems in SfM.

3. RESULTS

The results of both users' acquisition of the fountain of Peirene using SfM appear at first glance to have captured a significant portion of the fountain's architecture (figs 1-2). However, under close examination it becomes clear that User A's reconstruction was much more complete but still had several large gaps and point cloud coloring issues. Below we detail the main differences between the two users and explain how the user's capturing methodology could have achieved better results.

When the users cameras and chosen settings are compared it would appear that User B would acquire a much higher quality scan. In some respects this turned out to be the case. User B had a higher focal length and higher resolutions sensor compared to User A (Table 1). We can directly compare ground sampling density of the two users' camera setup by calculating their field of view taking into consideration each camera's crop factor and sensor resolution. At five meters User B's ground sampling density is 8.5 pixels/cm which is ca. 1.5x the density in pixels of User A's camera setup. However, the wider field of view allows User A to capture more area in each picture. In figure 2, where both user's photographed from the same position it is clear that User A was able to capture more area.

Table 1: Comparison Chart of Cameras and SfM Results

Category	User A	User B
Camera	Canon EOS 30D	Nikon D90
Focal Length	18mm (28.8mm)*	24mm (36mm)*
Resolution	3522x2348	4310x2868
Pixel Density at 5m	5.6px/cm 31.6px/cm²	8.5px/cm 72.25px/cm²
Camera Orientation	Portrait	Portrait
Camera Setting	Aperture Priority f/8	Aperture Priority f/4
Pictures Taken	478	548
Pictures Matched	473	349
Picture Efficiency	98.95%	63.69%
Sparse Point Features	233,003	203,139
Dense Point Cloud (PMVS)	11,216,589	12,821,729
Dense to Sparse Points[1]	48	63
Dense Points after Cleaning	10,955,837	12,289,853
Excess Points Removed	260,752	531,876
Qualitative Completeness[2]	95%	75%

[1]Calculates how many dense points could be extracted from images given found sparse points. A product of MP of camera.
[2]Area Captured/Total area of structure. *Equivalent Full frame focal length

This translates into more features that can be automatically detected for matching across images but at a slightly sparser density. User A's photograph resulted in 39,669 SIFT features, while User B's photograph had 46,780 SIFT features. User B had a denser count of features within a smaller field of view. User A's camera setup is best geared towards capturing more area which will help during matching and the final bundle adjustment of the reconstruction. In contrast User B's camera setup will have a higher density of extracted SIFT features per image, but User B will have less area to match between images unless they take more pictures than User A with more overlap.

Figure 2 First images with starting position for both Users. This figure shows a good comparison of differences between lenses and camera sensors.

Figure 3 User A (Top), note more complete reconstruction of surfaces. User B (Bottom), more lost data but areas appear sharper.

An analysis of the final reconstructions for both users highlights the trade-off between field of view and ground sampling density (GSD). For example, User B's reconstruction had a slightly higher point cloud density (1,334,016 more points) due to the higher image GSD (see table 1). In figure 3, where User B captured the same scene in almost the exact same positions the 3D point cloud appears sharper with brighter and more defined colors because the point cloud is much denser in this area for User B. In figure 4, a close-up of the steps, shows that User B's point cloud has a greater density. However, despite the greater point cloud density the coverage and qualitative results of User B's to User A's scan are much lower in part due to the smaller FOV and other acquisition mistakes discussed further below.

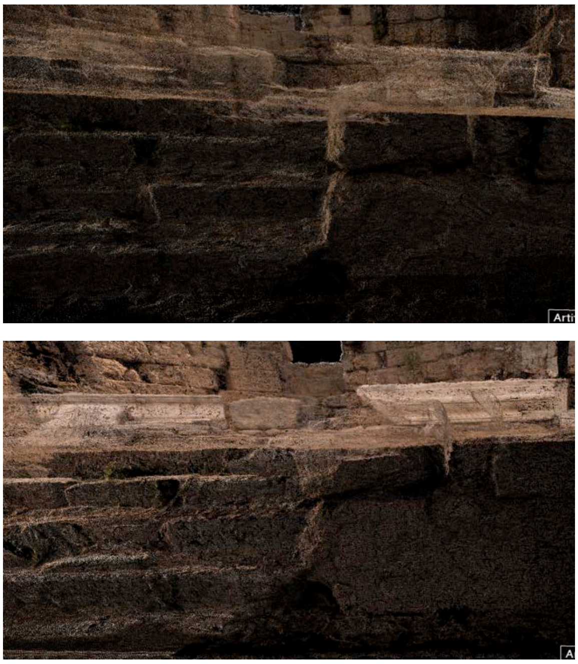

Figure 4 A close-up of the steps and white marble in background shows that User B (Bottom) acquired a denser point cloud due to a higher megapixel camera and higher focal length.

When closely examining the individual input images and resulting point clouds it becomes clear that neither user compensated for light changes and the effects of their cameras' built in white balance meter. First, many of the images where depth-of-field played a role were not pixel sharp, the combination of aperture and shutter speed resulted in a shallow focus. Second, both users periodically had blurred pictures due to low shutter speeds as they were capturing and moving at the same time.

Lighting problems are most apparent when they faced the dark clouds illuminated by the sun. The result for both users was dark under exposed images (figures 5-6).

Figure 5 Under exposed areas occur for both Users. User B's (Right) automatic settings led to much more under exposure. User A: 18mm, 1/60, f/8, ISO-100. User B: 24mm, 1/400, f/4, ISO-200.

Figure 6 The effects of under exposed images can be seen in these figures, where the pillars and back wall appear to have a false shadow. Note User B's reconstruction of this area is sparser and missing sections of the recessed courtyard.

This again directly affects the amount of SIFT features that can be detected and matched. SIFT and other related algorithms are typically quite robust to minor light change but extreme illumination changes or lack of sufficient light are quite disruptive (c.f. Lowe 2004). Dark shadows and under exposed images will have significantly less features and in turn fewer successful matches. Examining the amount of sift features extracted for the underexposed images in Figure 5, User A's photograph had 30,988 SIFT features, while User B's had 11,861 SIFT features. The significantly reduced count of SIFT features compared to User B's average (ca. >45,000) is directly related to the underexposure of the image. Although these images were still included in the complete reconstruction, the affect of the lighting can be seen especially in some of the pillars, which in the case of User B were almost black (where no points were extracted). In general, lighting for both users was highly variable; this is likely due to over-reliance on Matrix Metering (Nikon) or Evaluative Metering (Canon) that processes the entire scene's lighting. Spot metering, relying on a single, or small, area helps prioritize the exposure to the user's desired target.

A qualitative comparison between the two users indicates that although User B had a significant advantage in the selection of camera, a number of mistakes were made leading to User A being much more successful in conducting a thorough capture of the fountain (figures 7-8). Although both users maintained a fixed focal length for the majority of capture and User B roughly followed User A, their results differed significantly. Since SfM algorithms use a positional change in camera position for each image to find the distance (motion) between them it is critical to move and not stay in one place, maintain significant overlap and not make large rotational changes all at once. When capturing complex archaeological sites a systematic approach should be planned before hand to ensure the capture session achieves full coverage and mistakes in maintaining image overlap are not made when transitioning from one area to another.

User B failed to achieve enough overlap to fully reconstruct the entire fountain due to a lack of planning and not fully understanding the limits of the SfM algorithms. First, the southern section had few overlapping images and resulted in it being processed as a different model (Figure 9). Second, User B often stood in the same position and rotated the camera (panoramas): these were found during the bundle adjustment but did not contribute significantly to key feature matches since they had little parallax. Third, User B's images overall were much darker resulting in a poorer quality reconstruction and possibly led to the loss of key tie points. User B's reconstruction had more points in the model due to the higher GSD, but qualitatively it is difficult to determine where this higher density paid off. Certain areas are more detailed in User B's reconstruction, but overall it had many more holes and areas too underexposed to be clearly seen in the final point cloud. User B moved parallel to each side of the square courtyard (figure 9). Rarely were shots taken at oblique angles, another possible cause of poorer and more occluded capture.

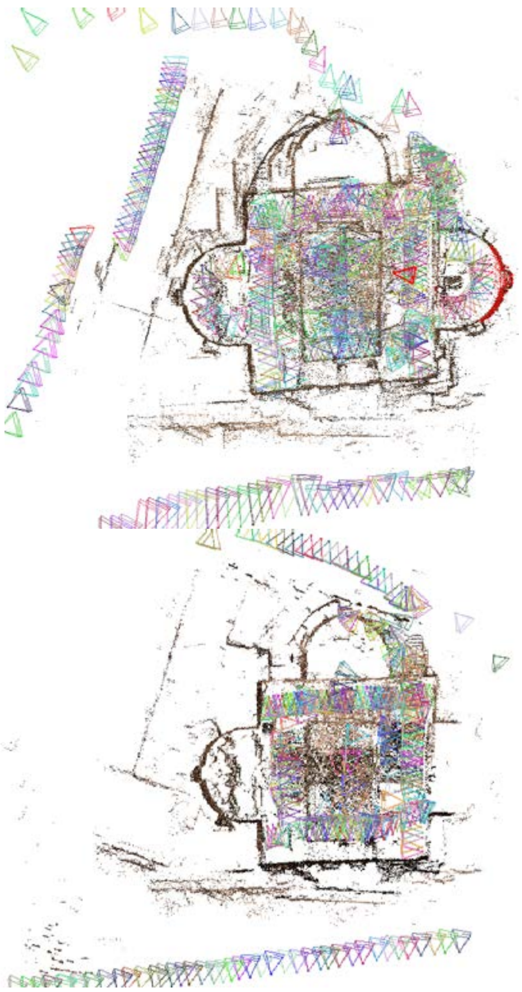

Figure 7 Sparse Reconstruction and calibrated camera positions for Fountain of Peirene. User A (Top), and User B (Right), processed in VisualSfM.

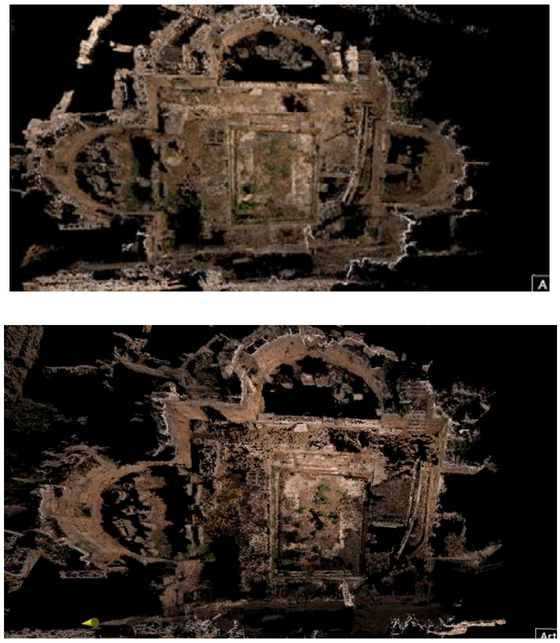

Figure 8 Dense Reconstruction of Fountain of Peirene. Note many more sparse areas in User B (Bottom) reconstruction and missing right fountain (Processed by PMVS).

In contrast, user A appears to be more experienced with SfM acquisition and had a specific plan of how they would sufficiently capture the entire site. Matching of image features is robust to 30 degrees in any direction (Lowe 2004). The user had very thorough coverage paying attention to not turn sharp angles (>30 degrees) and insuring significant overlap between each image (figure 9). At corners extra shots were taken and the user appears to have turned 90 degrees to shoot down the path they came and also about faced to get close-ups of the corners to better tie them in. User A spent special attention to difficult areas and took many detailed shots of areas with high occlusion or windows/pits.

In summary, even though User A took fewer pictures with a lower GSD the method of documentation resulted in a much better capture than that of User B. Both users' results could have been even better if they paid greater attention to lighting, shutter speed, and aperture.

Finally, both users failed to adequately capture the floors to reconstruct them properly using Patch-based Multi-view Stereo (PMVS) software and would be an ideal candidate for an improved Multi-view stereo algorithm called CMPMVS (Jancosek et al. 2011). The ability to point the camera down into the inset courtyard enabled this area to be captured but the other areas horizontal to the camera were sparsely documented (see Figs. 6-8). Especially in the most important area (the fountains), neither user angled the camera down to the floor to capture this tricky area.

Figure 9 Close-up of the sparse reconstructions and calibrated camera positions for Fountain of Peirene, User A (Top Left); User B (Top Right) (Processed in VisualSfM (Wu et al. 2011)).

The results show that opportunistic capture is not as much a threat to Cultural Heritage as one might think. Rouge SfM photographers cannot be compared to a trained surveyor with ample time to wait for the best lighting conditions, plan a detailed approach for full coverage of the site, and conduct follow-up visits to address mistakes in their photo shooting session.

4. CONCLUSION

SfM provides an ideal 3D solution to rapidly recording monuments uncovered by long-term excavation projects such as the ASCSA project at Corinth. Implementing proper capture of cultural heritage sites using SfM is a learned art. It has been argued here that the proper setup and methods applied during an SfM photo-shooting session play a very critical role in the final outcome. Archaeologists seeking to apply this method must take into consideration the limitations of the algorithms used in SfM and apply a systematic approach to full site coverage. A developed methodology for applying SfM in the field plays as much if not more influence on the final reconstruction than the specific SfM program used.

In the near future, we will work toward using SfM as a 'digital scaffold for embedding many years of metadata collected by the generations of excavators who have worked at this remarkable site.

While it took less than two hours to capture the Fountain of Peirene using SfM technology, the ease of capturing such data does not make it right to carry out such work without permission of the authorities, whether in Greece or any other country. Ease of data capture raise hard ethical issues about who owns the cultural heritage digital datasets that are becoming so easy to collect. Ultimately, we believe the same ethical standards apply to those professionals wishing to capture digital data at cultural heritage sites as to those researchers that wish to study any aspect of the patrimony of a country. Consequently, this will always begin with obtaining a permit from the authorities of the country where such work is to be carried out.

5. ACKNOWLEDGEMENTS

We are grateful to James C. Wright, Director, of the American School of Classical Studies in Athens for securing permissions and travel to Corinth to carry –out the SfM experiment described here. Thanks to Guy Saunders, Director of the American School of Classical Studies' Excavations at Corinth for facilitating our work at the site. Special thanks to the Institute for Aegean Prehistory (INSTAP) for a travel grant awarded to Thomas E. Levy that facilitated group travel to Greece for this project. Matthew L. Vincent was supported by the National Science Foundation under IGERT Award #DGE-0966375, "Training, Research and Education in Engineering for Cultural Heritage Diagnostics."

REFERENCES

Furukawa, Y., and J. Ponce. (2007) Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, July 2007. http://grail.cs.washington.edu/software/pmvs/bibtex.txt

Hirschmüller, H. (2008) Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, pp. 328–341.

Jancosek, M., T. Pajdla. (2011) Multi-View Reconstruction Preserving Weakly-Supported Surfaces. *CVPR 2011 - IEEE Conference on Computer Vision and Pattern Recognition 2011*.

Levy, T. E., Anderson, J.D., Waggoner, M., Smith, N., Muniz, A., and Adams, R.B. (2001) Interface: Archaeology and Technology - Digital Archaeology 2001: GIS-Based Excavation Recording in Jordan. *The SAA Archaeological Record* 1(3):23 - 29.

Levy, T.E., Tuttle, C.A., Vincent, M., Howland, M., Richter, A., Petrovic, V., and Vanoni, D. (2013) The 2012 Petra Cyber-Archaeology Cultural Conservation Expedition: Temple of the Winged Lions and Environs, Jordan. *Antiquity (Project Gallery)* http://antiquity.ac.uk/projgall/levy335/.

Levy, T.E, and N.G Smith (2007) On-Site Digital Archaeology: GIS-Based Excavation Recording in Southern Jordan. In *Crossing Jordan – North American Contributions to the Archaeology of Jordan*. T.E. Levy, M. Daviau, R. Younker, and M. M. Shaer, eds. Pp. 47-58. London: Equinox.

Levy, T.E. 2013 Cyber-Archaeology and World Cultural Heritage: Insights from the Holy Land. *Bulletin of the American Academy of Arts & Sciences* LXVI:26-33.

Lowe, D.G. (2004) Distinctive image features from scale invariant keypoints. *IJCV*, 60(2), pp. 91-110.

Snavely, N., S. M. Seitz, R. Szeliski. (2006) Photo Tourism: Exploring image collections in 3D. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006).*

Robinson, B. A. (2011) *Histories of Peirene: A Corinthian Fountain in Three Millennnia*. Princeton: American School of Classical Studies at Athens.

Rothermel, M., K., Wenzel, D. Fritsch, N. Haala. (2012). SURE: Photogrammetric Surface Reconstruction from Imagery. *Proceedings LC3D Workshop*, Berlin ,December 2012.

Smith, N.G., and T.E. Levy 2012 Real-time 3D archaeological field recording: ArchField, an open-source GIS system pioneered in Jordan. *Antiquity* 85(331):on-line http://antiquity.ac.uk/projgall/smith331/.

Wu, ChangChang. (2007) SiftGPU: A GPU Implementation of Scale Invariant Feature Transform SIFT. *http://cs.unc.edu/~ccwu/siftgpu*.

Wu, Changchang, S. Agarwal, B. Curless, and S. M. Seitz. (2011) Multicore Bundle Adjustment. *CVPR 2011, (poster ,supplemental material)*.