Automated 3D Model Reconstruction from Photographs

Paul Bourke
iVEC @ UWA
The University of Western Australia
Perth, Western Australia 6009, Australia
Email: paul.bourke@uwa.edu.au

Abstract— **This workshop is intended to provide an overview of the state of the art as well as practical considerations in relation to the automatic reconstruction of textured 3D models solely from a number of photographs. Presented will be the theory, introduction to the current software solutions and possible pipelines as well as some of the post processing requirements and tools. Reference will be made to camera, lenses and photographing techniques that result in an optimal chance of successful reconstructions. In short, the workshop will aim to provide a complete introduction to the subject. The emphasis will be towards creating 3D assets for gaming and virtual environments as well as targeting the recording and archive of heritage objects or places. These later applications will be the topic of most of the examples based upon the practical deployment of this technology by the author.**

Keywords— Photogrammetry, 3D reconstruction, Bundler, Structure from Motion, SfM, archaeology, heritage, archives.

I. INTRODUCTION

Photogrammetry is the general term given to the process of deriving some 3D quality solely from photographs, typically two or more. While it is almost as old as photography itself, the initial applications were around deriving relative distances or dimensions of objects from a pair of photographs. These and other uses for surveying generally required careful camera positioning, lens calibration and often required markers or known reference points in the scene. If one can derive a number of distances and dimensions from a pair of images, one might imagine that additional distances could be acquired with more photographs, even enough to form a cloud of points in 3D space. More recently, due to improvements in algorithmic developments in computer science and in particular machine vision, it has become possible to create sufficiently dense point clouds that they can reasonably be draped with a mesh and subsequently textured to form highly realistic representations of an object or place. The main family of algorithms are those referred to as SfM [1,2,3] (Structure from Motion) where the Bundler algorithm is the most widely known.

These algorithms present an exciting new opportunity for various recordings in heritage and archaeology, creating a powerful digital representation of the three dimensional structure rather than just flat 2D photographs. Given such a 3D model, views can be derived from positions other than from where the original photographs were acquired. Additionally structural measurements and analysis can be performed and research questions answered subsequent to the recording. In the context of remote sites it offers significant advantages over more traditional approaches such as laser scanning, in particular, it does not require any heavy equipment other than a good quality camera. Laser scanning generally requires multiple scans, are often very time consuming and considerable rigor is required in order to join multiple scans. Compared to return of flight methods and depth cameras it more naturally deals with convoluted geometry and can be performed in a wider range of environmental conditions. Finally, the texture quality from 3D reconstruction methods is generally much higher than alternative approaches, not surprising since it is based upon a photographic process in the first place.

An area of exploration involves 3D reconstruction from ad-hoc photographs [4], such as those found in on-line photography collections. However in order to achieve undistorted, more complete and dimensionally correct models a more robust process is required, in particular, having constant and known camera parameters. Applying some care and experience to the photographic process can result in accurate and dimensionally correct models even when the scene is free of any in-scene markers, often not possible for heritage objects or valuable objects.

II. ALGORITHM

The algorithmic pipeline for 3D reconstructions of the sort presented here is as follows. The process as described is largely automatic except for the selection of various algorithm parameters. The main manual aspect is the last process that depends on the degree of model cleaning required.

- Feature point detection [5,6,26] generally between all pairwise combinations of the photographic collection. The performance of this process can be improved if something is known about the order of the photographs.
- Numerical process to estimate the camera positions and intrinsic properties of the cameras and at the same time deriving the 3D positions of the feature points. Traditionally this process would be seeded with the parameters from a lens calibration [7], increasingly this is not required except for very wide angle lenses or otherwise non-linear optics. This is generally referred to as Structure from Motion [8] (SfM) process, one implementation referred to as the Bundler algorithm [9,10].

- With the knowledge of the derived camera positions and the feature points one can now derive a denser point cloud. Generally this can be an order of magnitude more points than the sparse cloud. One solution is called CMVS - Clustering Views for Multi-view Stereo.
- Form a triangular mesh over the dense point cloud. There are a number of approaches, some options are called ball pivoting, Poisson Surface Reconstruction and Marching Cubes.
- Given this mesh and the known camera positions and their lens parameters it is possible to re-project the photographs from each camera onto the mesh and blend across the overlap regions to form a texture map for the mesh.
- Perform various post processing operations [11] such as removing unwanted reconstructed parts of the model, closing holes usually arising from non photographed regions, subsampling the mesh and/or texture resolution before exporting the model, noting the degree of subsampling depends on the intended application.

III. Photography

There are three general topological categories that require slightly different photographic techniques. They are:

- Single objects, either isolated 3D objects [12] or 2D surfaces [13,14]. These can typically be photographed from arbitrary positions around or across the object where the whole object is contained within each photograph, figure 1.
- Extended objects, figure 2, where the photographs are generally taken along paths roughly equal distant to the object and each photograph only contains a small region of the overall structure to be reconstructed. The key to photographing these objects is to ensure sufficient overlap between photographs and different perspective views of all portions of the object.
- The most challenging is a combination of the above where there may be large-scale structures as well as localised objects to be captured in more detail perhaps. Figure 3 is an example of a concave cave with additional internal structures.

In all cases the photographic principles are the same, they only differ in the detail and the number of photographs required. The general rule is that one aims to photograph every part of the object from both different camera positions and with different perspectives. In many regards photographing for the purpose of reconstruction is the opposite of panorama photography where one aims to move the camera about a single position, called the nodal point of the camera. In the context of 3D reconstruction one may well take multiple photographs from a single position in order to get object coverage, but these would be considered a single photograph from a single perspective.

The number of photographs acquired depends very much on the type of object or scene being captured. Single objects to be reconstructed as surfaces or fully 3D objects generally require the least number of photographs, sometimes as few as 6, rarely more than 40. For extended objects the number of photographs depends on the field of view of the camera relative to the object. One might average 6 to 8 photographs (from different positions) for each part of the object. Large-scale structures with internal detail can easily require many hundreds of photographs, see figure 3.

Fig. 1. Example of a single self contained object reconstructed from 25 photographs.

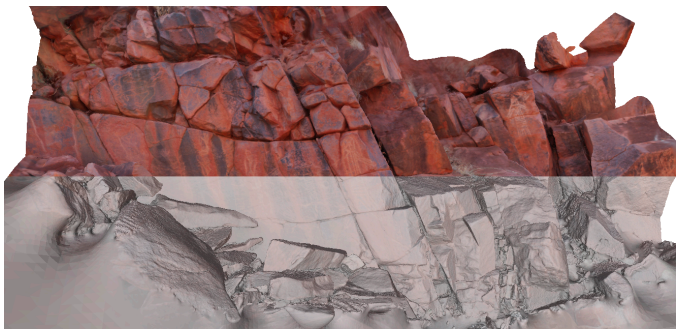

Fig. 2. Extended cliff face with Australian indigineous rock art reconstructed from 50 photographs. (Wanmanna, Western Australia).

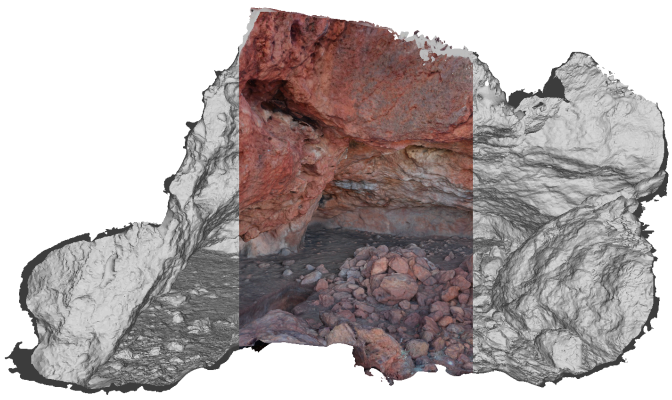

Fig. 3. Extended rock shelter scene with internal structures reconstructed from 350 photographs. (Weld ranges, Western Austalia)

IV. Resolution

An important consideration in such reconstructions is the actual geometric resolution as opposed to the apparent resolution. Apparent resolution is referring to the illusion of geometric detail conveyed visually through the high quality textures as opposed to structural detail. The relative importance of these two sources of detail depends on the intended application for the reconstructed object. For example as a digital asset [15] of the object/site then one strives for the highest resolution for both the visual detail (texture) and geometry. For measurement or structural analysis [16] one may not be interested in the texture resolution at all. For real time environments there can be constraints on the geometric resolution supported and providing apparent detail through good quality textures is acceptable, figure 4. For educational and practical delivery of models online, a compromise may be required. It should be noted that geometric detail is the most difficult to achieve and is currently an area of active research as well as the aspect that benefits from practical experience in the photographic acquisition. The relative importance of these two sources of detail is summarised in table 1.

Application	Geometric detail	Texture detail
Virtual environments	Low	High
Geometric analysis	High	None
Education	Medium	High
Archive	High	High
Online	Low/average	Average

Table 1. Relative importance of the two sources of detail.

V. Accuracy

A commonly asked question in relation to 3D reconstruction is how accurate are the models? Quantitative measures are problematic for a number of reasons. The reconstructed models usually do not have uniform fidelity, the areas/regions of interest may be more accurate than more distant parts. Localised regions of a model may have lower accuracy than others due to a range of factors, many stemming from poor photographic practice. Some surface textures/properties are more suited to 3D reconstruction than others. Measuring accuracy over a model can be problematic due to there being no ground truth. It should be noted that the claim by some that laser scanning is the ground truth is not as clear cut and obvious as one may think.

The author has focused on three methods for estimating accuracy, they are:

- Repeatability. Taking many more photographs than necessary and performing multiple reconstructions each based upon a subset of the photographs. The variability in the resulting models is an estimate of how accurate any one reconstructed model would be.
- Measurement. Performing a reconstruction and comparing measurements from the reconstruction to the actual object, figure 5. Since reconstructions may be of arbitrary scale and orientation, it may be necessary to fix scale by measuring one known feature and initially scaling the model to that.
- Modalities. Comparing reconstructed models with other scanning modalities, for example: laser scans, structured light scanners, CT scans.

Fig. 4. 1,000,000 triangle (left), 100,000 triangles (right).

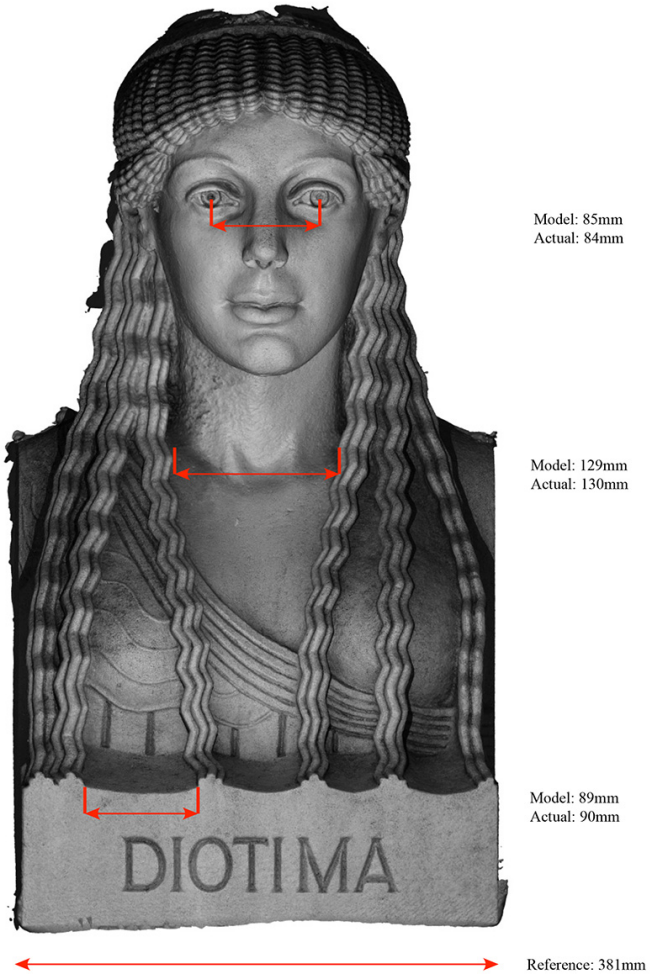

Fig. 5. Comparison of subsequent dimensions with the actual object after fixing the scale of the reconstructed object with a reference measurement.

VI. LIMITATIONS

There are a number of limitations; some are inherent in the process while others may still be solved with improved algorithms or technique.

Shadows from the environment are baked into the textures applied to the surfaces. This can include shadows cast from the photographer and camera.

Movement during the photography process will generally invalidate the first stage of the processing, namely the feature point detection [17]. The obvious solution is to use multiple cameras in a rig with a synchronised capture. While this is a viable solution for a certain class of object, for more complicated objects or environments the number of cameras required may be prohibitive.

Reflective surfaces are a problem and even highly specular surfaces can cause errors. Reflective surfaces "fold" the visible world about the reflective surface plane, as such the folded world is often reconstructed behind the reflective surface. Curved reflective surfaces provide a distorted nonlinear folding and thus the feature point detection usually fails.

Structures that are close to the resolving power of the camera system [18] may be visible in some photographs and not in others. This can confuse the feature point detection algorithms as well as lead to erroneous geometry during meshing.

Obviously one cannot hope to reconstruction what one does not photograph, constraints of access may mean one cannot capture the desired set of photographs. Related to this are foreground occluders that can limit the ability to reconstruct background surfaces. This however can be one of the advantages over laser scanning where there is an increased risk of shadow zones for highly convoluted objects. A laser scanner is significantly more time consuming to repeatedly reposition and calibrate for multiple scans.

CONCLUSION

Presented in the workshop and summarised in this paper is a comprehensive introduction to the automatic 3D reconstruction of models from photographs. The optimal hardware and software pipelines are described along with strategies for taking photographs that will give optimal results. Practical examples are illustrated in the workshop including fully worked pipelines. In addition the applications, challenges and limitations of the process are outlined. Photographic based reconstruction like all other capture modalities does not suit all applications, the key in understanding when it is the appropriate solution. As a result of the workshop attendees should fee more comfortable exploring this technology to their own practice.

ACKNOWLEDGMENTS

The work was supported by iVEC through the use of advanced computing resources located at iVEC@UWA.

The Weld Range Web of Knowledge Project is funded by the Federal Government via the Indigenous Heritage Program. 3D visualisation was undertaken in collaboration with Ethical Engagement Consultancy and with assistance from Sinosteel Midwest Corporation.

Rock art access in conjunction with the Centre for Rock Art Research and Management, The University of Western Australia, with financial support from BHP Billiton.

REFERENCES

[1] O. Faugeras, S. Maybank. “Motion from point matches: multiplicity of solutions”. International Journal of Computer Vision, 4(3):225-246, June 1990.

[2] C. Tomasi, T. Kanade. “Shape and motion from image streams under orthography: A factorization method”. International Journal of Computer Vision, 9(2):137-154, 1992.

[3] J. Shi, C. Tomasi. "Good Features to Track,". 9th IEEE Conference on Computer Vision and Pattern Recognition. Springer. June 1994.

[4] N. Snavely, S.M. Seitz, R. Szeliski. Modeling the World from Internet Photo Collections. International Journal of Computer Vision, 2007.

[5] D. G. Lowe. "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision 2. pp. 1150–1157, 1999.

[6] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.

[7] Y. Furukawa, J. Ponce. Accurate camera calibration from multi-view stereo and bundle adjustment. International Journal of Computer Vision 84, 257–268, 2009.

[8] C. Wu. "VisualSFM: A Visual Structure from Motion System", http://homes.cs.washington.edu/~ccwu/vsfm/, 2013.

[9] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon. Bundle adjustment—A modern synthesis. in: Triggs, W., Zisserman, A., Szeliski, R. (Eds.), Vision Algorithms: Theory and Practice, LNCS, Springer–Verlag, pp. 298–375, 2000.

[10] M. Lourakis, A. Argyros. SBA: A generic sparse bundle adjustment C/C++ package based on the Levenberg-Marquardt algorithm. /http://www.ics.forth. gr/lourakis/sbaS, 2008.

[11] P. Cignoni, M. Callieri, M. Corsini, M. Dellepaine, F. Ganovelli, G. Ranzuglia. MeshLab: an opensource mesh processing tool. Eurographics Italian Chapter Conference, The Eurographics Association, 129-136, 2008.

[12] G. Percoco. Digital close range photogrammetry for 3D body scanning for custom-made garments. The Photogrammetric Record, 26: 73–90. doi: 10.1111/j.1477-9730.2010.00605, 2011.

[13] L. Barazzetti, M. Scaioni, F. Remondino, Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation. The Photogrammetric Record, 25: 356–381. doi: 10.1111/j.1477-9730.2010.00599, 2010.

[14] B Jeffery. From Seabed to Computer Screen - digital mapping of submerged and shipwreck sites. Bulletin of the Australian Institute for Maritime Archaeology, 23:86-94, 2006.

[15] T. P. Kersten, M. Lindstaedt, Automatic 3D Object Reconstruction from Multiple Images for Architectural, Cultural Heritage and Archaeological Applications Using Open-Source Software and Web Services. Photogrammetrie - Fernerkundung - Geoinformation, Heft 6, pp. 727-740.

[16] M. Favalli, A. Fornaciai, I. Isola, S. Tarquini, L. Nannipieri. Multiview 3D reconstruction in geosciences. Computers & Geosciences, 44 168–176, 2012.

[17] S. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision. 60 (2), pp 91-110, 2004.

[18] D. Nister. Automatic passive recovery of 3D from images and video. In: Proceeding of the Second IEEE International Symposium on 3D Data Proces- sing, Visualization and Transmission, pp. 438–445, 2004.

[19] E.M. Mikhail, J.S. Bethel, J.C. McGlone. Introduction to Modern Photogrammetry. John Wiley & Sons, Inc., New York, 2001