

# Automatic 3D reconstruction: An exploration of the state of the art

Paul Bourke  
iVEC@UWA

The University of Western Australia, 35 Stirling Hwy, Crawley, Perth, West Australia, 6009, Australia.  
E-mail: paul.bourke@uwa.edu.au

**Abstract**—Presented here in the form of case study examples are the results from a number of practical exercises to explore the state of the art of automatic 3D reconstructions. That is, deriving the underlying geometry of an object or place based only upon photographs. There is a wide range of applications for this technology; traditionally it has been used for landscape/terrain modeling, geology and by the mining industry. The interest here is in capturing geometric data in archaeology, providing a new data format suited to a richer exploration compared to the more traditional photography. Examples of the use of this 3D geometric representation include the population of virtual worlds and gaming engines. The manual generation of such assets is normally both time consuming and can involve an element interpretation on the part of a human modeler.

**Index Terms**—Reconstruction, photogrammetry, photography, gaming, virtual environments.

## I. INTRODUCTION

Photogrammetry is the name given to a range of techniques by which 3D properties of an object are derived from 2D images. It is most often associated with the derivation of topology using aerial photography, this is now more correctly referred to as a 2.5D reconstruction since it usually involves only convex surfaces. The automated creation of full 3D models from a series of photographs has been an active area of research for many years in computer science and vision research, it is often referred to as SfM (Structure from Motion) [1]. The techniques would appear to be maturing, this is reflected in a number of stable software tools being released from research laboratories both as commercial products but also appearing in the public domain. This opens the way for wider adoption since the algorithms can be used with the need for less specialist knowledge.

The following presents, by way of case studies, some recent exploration of the status of the technology in this area. A number of applications at The University of Western Australia for automatic 3D reconstruction from photographs have been identified, these include the population of game engines with accurate models (as opposed to manually created models involving artistic interpretation), creation of virtual worlds with realistic models [2] without the need for time consuming manual modeling, and generating databases of 3D models for research and documentation in archaeology [3] and heritage.

Semi-automated 3D surveys are often performed with "return of flight" methods such as LIDAR (Light Detection and ranging). These scan an object or scene with a laser, the phase of the reflected light results in a point cloud of depths (distance from the laser/received position). This is often supplemented with a (almost) coincident camera to assign colour to the points and optionally a texture to a reconstructed mesh representation.

Structured light scanners [4], used for example in consumer products such as the Kinect, project a known pattern onto an object and derive surface properties and depth by analysing an image of the result.

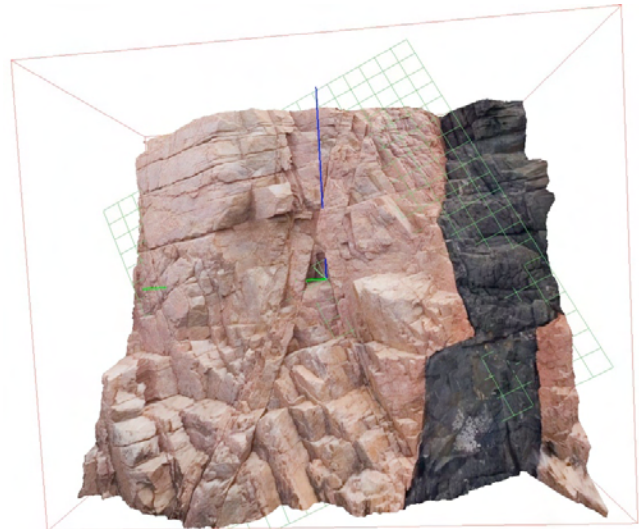


Figure 1: Reconstruction of an outcrop face from just two photographs using SiroVision [4]. Approximately 50m across.

An early commercial product using solely images and developed was SiroVision [5]. It targeted the mining industry as a solution for mapping 3D structures and determining, for example, the volume of rock extracted during mining operations. While the software generally only required two photographs and was aimed at creating a high degree of dimensional accuracy it was limited by the need for some rigor in the image capture process. For example there is often the need for in-scene reference points and a high degree of control and knowledge of the camera hardware, for example lens information and calibration.

The algorithms that are of more interest and will be explored and discussed here are primarily aimed at "ad hoc"

image capture. That is, reconstruction using standard cameras or, reconstruction of objects one may not be able to reach or add markers to for heritage/preservation reasons. There are obviously some guiding principles that can affect the quality of the final reconstruction. For example, it is fairly obvious that one is not going to be able to reconstruct parts of a model not visible to the camera, this is particularly relevant for concave regions of the object.

A number of the current software tools have been investigated [6], these include

1. 123D Catch by Autodesk [7]
2. PhotoScan by Agisoft [8]
3. PhotoSynth by Microsoft [9]
4. PhotoModeller by Eos Systems [10]
5. Bundler (Open Source) [11]
6. Visual SFM (Open Source) [12]

Note that the last two are based upon the open source tool Bundler. PhotoSynth and 123D Catch are both cloud based solutions, the results shown here are from either 123D Catch or PhotoModeller. The former is by far the simplest to use (almost no user control) at the price of potentially not providing the same quality results as PhotoModeller. This article is not intended to serve as an evaluation of these packages, in particular, they often have their own strengths and so the best choice can be application specific.



Figure 2: Four photographs of an embossed wall at the 1000 column temple, approximately 10cm across.

## II. EXAMPLE 1: CULTURAL HERITAGE (KARNATAKA, INDIA)

This first example is from the Moodabridi Jain temple [14], also known as the 1000 pillar temple located in the state of Karnataka in India. The intention was to capture the

many reliefs carved into the walls and columns of the temple and to subsequently document the associated deities and stories. Due to the large number of motifs involved an advantage of this technique is the rapid acquisition time. Since the reliefs are generally only 2.5D only between three and six photographs are typically required. See photographs in figure 2 and the final reconstruction illustrating the mesh in figure 3.

Feature point extraction is the process of identifying equivalent points in each image, or at least between pairs of images in a sequence. This phase of the reconstruction pipeline is similar to the feature point extraction used in stitching panorama or gigapixel photographs. The use of special markers, required for an accurate result using some packages, is a means of ensuring high quality matched feature points. Of note in this example is the lack of strong colour information, often assumed to be required in the process of feature point extraction between the images.



Figure 3: Reconstructed and textured mesh from the four images presented in figure 2.

## III. EXAMPLE 2: ROCK ART ARCHAEOLOGY (PILBARRA, WEST AUSTRALIA)

As the object to be captured becomes more convoluted a higher number of photographs are required. In this example from a rock art archeological survey in Western Australia [16] typically six to a dozen photographs were used, see figure 4 for a representative set. The larger number of photographs is a reflection of the need to photograph from multiple vertical positions rather than just horizontally. This example illustrates the ad-hoc nature of the photography. For example in this case there are a mixture of landscape



and portrait images as well as images taken at different times of the day and thus under different lighting conditions. This insensitivity of the algorithms raises the exciting possibility in the future of reconstructing models from public image collections such as Flickr [17]. It should also be pointed out that these photographs are taken with a modest ten MPixel "point and click" camera. This example is also interesting given the rock surface textures are very noisy surfaces without clear distinguishing geometric or colour features. Views of the reconstructed surface are given in figure 5.



Figure 4: Six photographs of a rock art structure, approximately 3m across.

#### IV. EXAMPLE 3: RECONSTRUCTION OF A FULL 3D OBJECT

This final example attempts to reconstruct a full 3D object, namely the bust of a statue, see figure 6. Due to the photographic coverage requirement such reconstructions require significantly more photographs, in this case 24 were used for the reconstruction.

A consideration in all the examples presented here is the degree of geometric detail versus the apparent detail conveyed by the high quality texture maps. Figure 3 gives a sense of the mesh resolution, and the close-up images in figure 6 show the true 3D nature of the mesh. The engraving on the statue is indeed resolved as geometry rather than just

texture, as are the cracks in the rock in figure 5. From the perspective of creating assets for virtual environments it is acceptable, even desirable, for the apparent geometry to be represented by the texture rather than the more computationally expensive geometry. For the archaeological record and creating objects as research data the true geometric detail is more important.



Figure 5: Two views of the reconstructed model from the six images in figure 4.

#### V. SUMMARY

Presented are three representative case studies from a larger set [15] aimed at exploring the state of the art of 3D reconstruction from photographs, photographs taken without rigorous procedures or specialist hardware. A number of software packages have been explored, these include high-end commercial packages (> \$10,000), public domain software, and cloud based services such as PhotoSynth and 123D Catch. It should be noted that the latter being vastly more useful since the 3D geometry can be extracted rather than remaining "in the cloud".

The technologies and algorithms in this area are increasingly moving out of the research laboratories and into

the consumer market. As such it is expected that their performance and accuracy will continue to improve. The time saving in creating assets suitable for serious gaming and as research objects in heritage and archaeology is significant as it has the added benefit of removing interpretation by a modeler/animater.



Figure 6: A full 3D reconstruction from 24 photographs.

#### ACKNOWLEDGMENT

The work was supported by iVEC through the use of advanced computing resources located at the University of Western Australia.

3D model and mesh representations presented here used the Meshlab software [13], this includes various mesh decimation algorithms for simplifying meshes to increase suitability for realtime interaction.

#### REFERENCES

- [1] Barazzetti, L., Scaioni, M. and Remondino, F. (2010), Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation. *The Photogrammetric Record*, 25: 356–381.
- [2] Percoco, G. (2011), Digital close range photogrammetry for 3D body scanning for custom-made garments. *The Photogrammetric Record*, 26: 73–90.
- [3] Remondino, F., Rizzi, A., Girardi, S., Petti, F. M. and Avanzini, M. (2010), 3D Ichnology— recovering digital 3D models of dinosaur footprints. *The Photogrammetric Record*, 25: 266– 282.
- [4] Fofi, D., Silwa, T., Voisin, Y.. A comparative Survey on Invisible Structured Light. *SPIE Electronic Imaging - Machine Vision Applications in Industrial Inspection XII*. San Jose, USA, 2004.
- [5] SiroVision, <http://sirovision.com/>
- [6] Beven, G. A Look at Photogrammetry using Images from the 2008 HMAS Sydney II Survey. iVEC Technical Report iVEC-TR001-2012, 2012.
- [7] 123D Catch <http://www.123dapp.com/catch>
- [8] PhotoScan <http://www.agisoft.ru/products/photoscan/>
- [9] PhotoSynth <http://photosynth.net>
- [10] Photomodeller <http://www.photomodeler.com>
- [11] Bundler <http://phototour.cs.washington.edu/bundler/>
- [12] Visual SFM <http://www.cs.washington.edu/homes/ccwu/vsfm/>
- [13] Meshlab <http://meshlab.sourceforge.net/>
- [14] Moodabidri Jain Temple <http://www.jaindharmonline.com/pilgri/mbidri.htm>
- [15] Further examples by author <http://paulbourke.net/miscellaneous/areconstruction/>
- [16] Paterson, A., McDonald, J. Centre for Rock Art Research and Management, University of Western Australia.
- [17] Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R. Building Rome in a Day. *International Conference on Computer Vision*, 2009.





*Paul Bourke is the director of the iVEC facility at the University of Western Australia. The facility hosts high performance computing resources on the campus and acts as an interface to the other high performance capabilities provided by iVEC. The facility also hosts display system in support of visualisation, scalable conferencing system, a teaching laboratory, high end workstations, and a video production unit.*

*Paul provides scientific visualisation services to researchers within university and externally. During his career he has worked in organisations where he concentrated on architectural, brain/medical, and astronomy visualisation. Of particular interest are engaging and immersive display technologies and how they may be used to support scientific visualisation, employed for public outreach and education, create environments for exhibitions in virtual heritage, and used to enhance more traditional forms of digital entertainment.*