AutoDesk 123D Catch: How accurate is it?

By Jim Chandler and John Fryer

123D Catch is free software for creating 3D meshes from user-supplied digital imagery. In this article **Jim Chandler** and **John Fryer** assess its accuracy and applications.

In April 2011, Autodesk provided access to a free and simple to use package for creating 3D meshes from user's digital imagery, a technique which many of the more traditional readers of Geomatics World would perhaps recognise as "photogrammetry"! However, software to carry out photogrammetric measurement has traditionally neither been cheap nor simple to use. The ability therefore to automatically generate a 3D model from any digital imagery, for free, appears to be extremely significant. The question for readers is naturally, how accurate is it? This brief report outlines an initial approach adopted to answer this question.

How does it work?

Autodesk 123D Catch can be freely downloaded (http://www.123dapp.com/catch) and remains free to use for non-commercial purposes, at least currently. It simply requires the user to supply a minimum of three images of an object acquired from different viewpoints, which are then uploaded to a server for processing. Although exact details are not provided by Autodesk, the processing undoubtedly involves SfM (Structure from Motion) methods, which initially involves identifying common image features across images using a freely available algorithm called SIFT (scale invariant feature transform) or perhaps the more robust version ASIFT. A bundle adjustment is then used to compute parameters to represent the inner and exterior camera geometry, combined with a dense point cloud to represent the object, (James and Robson, 2012). 123D Catch evolved from Autodesk's project Photofly, originally launched in 2010.

No restrictions appear to be placed upon either camera type or focus setting, so some form of camera calibration is clearly being conducted. The processing strategy appears similar to the Microsoft Photosynth initiative, in that minimal user input is required other than providing the images, which critically makes the software very easy to use. 123D Catch appears to offer increased benefits, particularly for visualisation and options to extract data for subsequent use. Both these packages represent a significant development for those interested in photogrammetry or, perhaps more significantly, those simply interested in capturing three-dimensional information cheaply and easily. However, a key question is just how accurate are the data generated by such a wholly image-based approach that uses no external object control constraints and includes fully automated camera calibration.

"... minimal user input is required..."

Figure 1: Aboriginal cave

Aboriginal cave re-measurement

To answer this question, a past project conducted to record an aboriginal cave site was reprocessed. The cave (Figure 1) is approximately 9 m in length and is of interest because of engraved features which resemble an emu foot: a token for the local aboriginal community that lived in this area of the Blue Mountains, New South Wales, Australia. This site had been recorded in 2004 using a series of overlapping stereo-pairs acquired using a 6 megapixel DCS 460 digital camera, equipped with a 24 mm lens. Of significance for this latest study, was the inclusion originally of twenty 3D control points which were established using a Leica 1100 series reflectorless total station. These data were used to conduct the original photogrammetric survey, extracting high resolution DEMs/orthophotos and a fly-through visualisation (**http://www.youtube.com/watch?v=yLX2wCcEH50**). Further details concerning this earlier work are described more fully in a series of papers: (Chandler et al, 2005; 2007; Chandler and Fryer, 2005; Chandler and Bryan, 2007).

For this latest test, the sixteen original images were uploaded to the 123D Catch server and were processed successfully and automatically within just fifteen minutes. In 2004, this processing stage had required four days work, with a high level of user input and experience, particularly in generating initial estimates for the exterior orientation and using an external self-calibrating bundle adjustment to calibrate the camera. The reduction in time and immediate generation of results was a particular contrast. Visually, the mesh looked superb (Figure 2) and indeed a 3-D fly through visualisation (**http://www.youtube.com/watch?v=hyCmrG3YShI**) suggested a far wider area of successful measurement than had been achieved previously. The control points were then measured individually in 123D Catch as "reference points" and measured on a single frame, hence interpolated from the 3-D mesh generated by 123D catch. The model was scaled using a known distance using the "define reference distance" command. The model and measured data were then exported as an Autodesk FBX file, and embedded in this ASCII file are the measured control coordinates, mesh vertices and a variety of other data of relevance to the camera calibration/restitution process.

Accuracy assessment

The 123D Catch control points and original control coordinates derived from the reflectorless total station were then used in a 3-D similarity transformation to determine the optimum rigid body transform between the two coordinate systems. Seven parameters were estimated: 3 translation, 3 rotation and 1 scale. The residuals derived from this least-squares estimation are presented in Table 1.

As the overall standard deviations suggest, the fit to the original control is just 12 mm, 11 mm and 4 mm in XYZ respectively. Although such accuracy is comparatively low (1:600) compared to normal stereo close-range photogrammetry (1:1,000-1:10,000), the results are certainly acceptable for many applications, particularly when considering that the restitution includes camera calibration for each photo, the whole task was fully automated and 123D Catch reduces the resolution of each original image to just 3 megapixel. Finally, the process is solely image based and no control constraints have been applied other than applying an approximate scale factor. The original field work would have consisted of nothing more than acquiring the imagery and a single measured distance between two well identified features.

During the original data-processing conducted in 2004, a self-calibrating bundle adjustment (Erdas Imagine/ LPS/in-house software) had been used to derive a set of parameters to model the focal length, principal point offset and radial lens distortion, which was assumed to be stable for all frames. In the least-squares adjustment for this original restitution, the overall residual fit to the control was 3.5 mm, 1.7 mm, 3.4 mm in XYZ respectively. Clearly this earlier estimation achieved a higher accuracy (1:1,600) than Autodesk 123D Catch could manage, but a significantly greater effort had been required!

Pt Nr	X	Y	Z
101	-0.01	-0.03	-0.01
102	-0.004	-0.017	0.000
105	-0.016	-0.003	-0.005
106	0.003	0.009	0.004
107	0.000	0.018	-0.003
109	0.002	0.006	-0.001
110	0.008	0.006	0.007
111	0.004	0.009	-0.002
112	0.021	0.012	0.000
114	0.009	0.001	-0.001
115	0.017	0.006	0.005
116	0.013	-0.002	0.004
117	0.003	-0.003	0.000
118	0.009	-0.005	0.007
119	0.003	-0.005	0.000
120	-0.009	-0.004	-0.001
122	-0.011	-0.002	0.000
125	-0.016	-0.003	-0.004
126	-0.024	0.007	-0.005
Std.Dev.	**0.012**	**0.011**	**0.004**

Table 1: Residuals at control points.

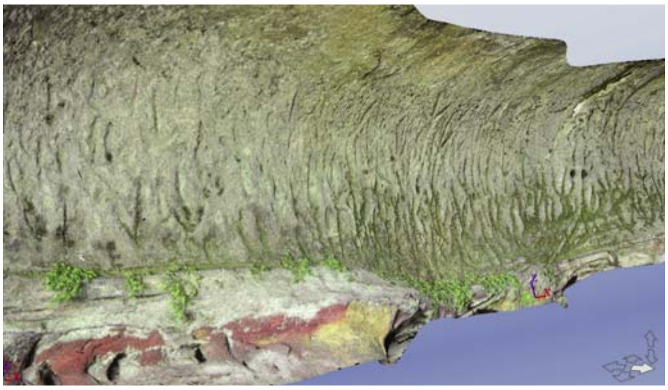

Figure 2: Aboriginal cave, meshed using 123D Catch

Discussion

Examining the residuals graphically and in three dimensions was revealing (Figure 3), the viewpoint being similar to the camera position adopted for Figure 1. Note also that viewing using standard red/green stereo glasses will enhance the three-dimensional effect! Figure 3 demonstrates a clear systematic pattern in which residuals are highest towards the edge and middle, but are in opposing directions to the approximate camera axes. This could be explained in two ways. First, the accuracy of the estimated focal lengths for each frame could be questioned, the inaccurate estimates creating the "push/pull" effect so graphically represented. Alternatively, the systematic pattern could be accounted for by considering classical principles associated with vertical aerial photography used for mapping. Although a series of stereo pairs were captured, they were effectively in the form of a classical aerial "strip", in which the normal end lap simply varied sequentially.

This is an inherently weak geometry, one that is recognized, tolerated and accepted because it is usually managed and minimised through the use of a series of ground control points. Such control would constrain each image individually, forcing it to fit the known object space. Without such a control constraint, any strip would have a tendency to wobble as small systematic errors make their presence known. Indeed the authors have

> *"Now, simply by acquiring a few additional frames, a 3-D record can be captured"*

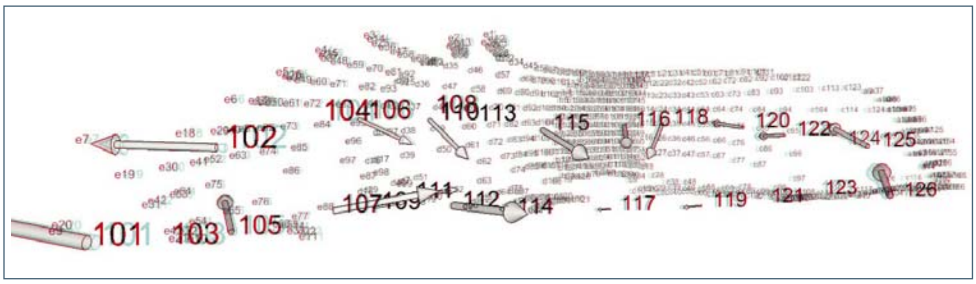

Figure 3: Residual fit following 3D similarity transformation.

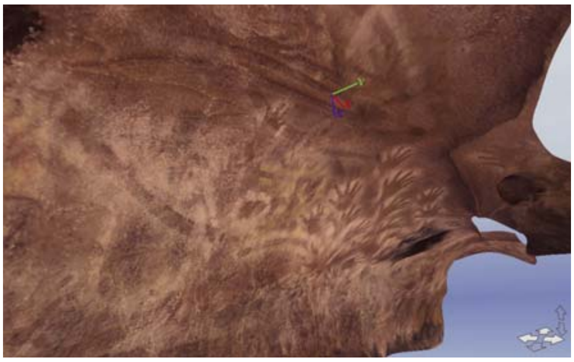

Figure 4: Additional cave survey

seen and modelled this type of effect before (Fryer et al, 1994). This earlier study revealed that a measurement error introduced into the centre of the block will propagate to the geometrically weaker periphery, as can be seen repeated for the Emu cave examined here. Autodesk 123D Catch is wholly image-based, and provides no opportunity to constrain individual frames in the manner required for this particular configuration. The simple solution would have been to strengthen the image block by including additional frames which capture larger areas of the cave from different positions. This would have no doubt prevented the wavering/drifting effect so graphically represented in figure 3, but unfortunately such imagery wasn't acquired at the time.

While the example of the Emu Cave worked extremely well, that is not the only strip or set of images submitted to 123D Catch by the authors. A cave with many more complicated 3D curved surfaces had been photographed in 2004 but proved almost too complicated for manual processing at that point. This imagery was similarly resolved by 123D Catch in a matter of minutes (Figure 4).

Applications in geosciences

The ability to create 3-D models from imagery freely will undoubtedly create new opportunities and applications. One particular area is in the geosciences, where there is a need to capture objects in the physical and natural environment effectively and at different scales. Geographers, geomorphologists , archaeologists and geologists are used to acquiring imagery to record the landscape in two dimensions. Now, simply by acquiring a few additional frames, a 3-D record can be captured. Levels of accuracy do not always need to be high but it is often simply necessary to capture the relative position of one feature in relation to another. The first author established ISPRS working group V6 "close range morphological measurement in the Earth sciences" in 2008 (http://isprsv6.lboro.ac.uk/). This was to both raise awareness of ISPRS within the geoscience community and to allow members of ISPRS to recognise the opportunities and needs of geoscience. At the recent ISPRS Congress in Melbourne, (August 2012) it was pleasing that ISPRS Council approved the continuation of the initiative. The new working group title for 2012 -2016 will be ISPRS V5 "Close-range measurements for bio- and geosciences" and will be chaired by Dr Dirk Rieke-Zapp.

Comparable to laser scanning?

Techniques like 123D Catch will undoubtedly provide geoscientists with a useful tool. Indeed, there is evidence the SfM techniques combined with the increased densification of point clouds through MVS (Multiview Stereo) is capable of producing datasets which are comparable to terrestrial laser scanning. This is significant as these high resolution datasets are generated at lower costs and with reduced data collection times (James and Robson, 2012). Despite this optimism it should be recognised that image-based methods do require significant image texture to be effective and accurate. Current testing using a range of imagery (digital SLR's and smart phones) of a cliff suggest the need to distinguish between the accuracy of individual measured points and subsequent point density necessary to capture an accurate surface representation. Work is ongoing!

This brief test has provided an assessment of the accuracy of Autodesk's 123D Catch. Although not at the level of accuracies routinely achieved in normal terrestrial photogrammetry using other packages, the accuracy achieved was certainly useful for many applications. If more imagery at a diverse range of scales had been acquired originally and used to provide a stronger configuration, accuracies would certainly have been improved. The real significance of Autodesk's 123D catch is that it is free, which will certainly help increase the number of applications that use this approach. It is hoped that this short article will help the wider surveying, geomatics and heritage recording community more fully to realise the potential of 123D Catch and of spatial measurement from images.

Jim Chandler

John Fryer

About the authors

Jim Chandler is the Professor of Geomatics at Loughborough University (http://www-staff.lboro.ac.uk/~cvjhc), responsible for teaching civil engineering students surveying and setting out. He has developed and applied geomatic methods and close range photogrammetry to a range of measurement problems over the last 25 years, particularly in the geosciences.

John Fryer is Emeritus Professor of Surveying and Photogrammetry at the University of Newcastle, Australia and retired as the Head of School of Engineering in 2004. He has had a long-term interest in close range photogrammetry.

References

Chandler, J.H. and Bryan, P., 2007. Cost-effective rock-art recording in a production environment: is there a wider message? CIPA XXI International Symposium, *AntCIPAting the future of the Cultural past, Athens, [CD-ROM], ISSN 1682 1777.*

Chandler, J.H. and Fryer, J.G., 2005.'Recording aborginal rock cut using cheap digital cameras and digital photogrammetry, *CIPA XX International Symposium, XX, International Cooperation to save the World's Cultural Heritage, Torino, pp. 193-8, ISSN 1682 1777.*

Chandler, J.H., Bryan, P. and Fryer, J.G., 2007. The development and application of a simple rock-art recording methodology based on consumer grade digital cameras", *The Photogrammetric Record*, 22(117): 10-21.

Chandler, J.H., Fryer, J.G. and Kniest, H.T., 2005. Non-invasive 3D recording of aboriginal rock art using cost effective digital photogrammetry, *Rock Art Research, 22(2): 119-130.*

Fryer, J.G., Chandler, J.H. and Cooper, M.A.R., 1994. On the Accuracy of Heighting from Aerial Photographs and Maps: Implications to Process Modellers", *Earth Surface Processes and Landforms, 19: 577-583, ISSN 0197-9337.*

James, M.R., and Robson, S., 2012. Straightforward reconstruction of 3-D surfaces and topography with a camera: Accuracy and geoscience application. *Journal of Geophysical Research, 117, F03017, doi: 10.1029/2011JF002289.*

Kersten, T. P., and Lindstaedt, M., 2012. Image-based low-cost systems for automatic 3-D recording modelling of archaeological finds of objects. *Proceedings of EuroMed 2012 to, International Conference on cultural heritage.*