*Archaeological Prospection*
*Archaeol. Prospect.* **18**, 67–73 (2011)
Published online 20 January 2011 in Wiley Online Library
(wileyonlinelibrary.com/journal/arp) DOI: 10.1002/arp.399

# Software Review

# Taking Computer Vision Aloft – Archaeological Three-dimensional Reconstructions from Aerial Photographs with PhotoScan

**GEERT VERHOEVEN***

*LBI for Archaeological Prospection & Virtual Archaeology, Ludwig Boltzmann Gesellschaft GmbH, Franz-Klein-Gasse 1, A-1190 Wien, Austria; Ghent University (UGent), Faculty of Arts and Philosophy, Department of Archaeology, Sint-Pietersnieuwstraat 35 (UFO - 110.016), B-9000 Gent, Belgium*

ABSTRACT Structure from motion (SFM) algorithms are known for their ability to reconstruct a sparse point cloud of scenes that were imaged by a series of overlapping photographs. When complemented by stereo-matching algorithms, detailed three-dimensional models can be built from such photograph collections in a fully automated way. Since 2010, a computer vision software package called PhotoScan has been available from the Russian manufacturer AgiSoft LLC. Even though the programme has a straightforward and simple interface, state-of-the-art SFM and other computer vision algorithms are implemented that enable PhotoScan to generate very dense and accurate three-dimensional meshes of various (archaeological) settings captured in two-dimensional imagery. Using a mixture of archaeological case studies, it will be shown that this low-cost application produces excellent results from aerial imagery in a minimum of time with almost no effort needed. Because of its solid and uncomplicated operation, it looks like this application could become an essential part of the aerial archaeologist's toolkit. Copyright © 2011 John Wiley & Sons, Ltd.

*Key words:* three-dimensional reconstruction; computer vision; structure from motion; multiview stereo reconstruction; aerial photography; PhotoScan

## Introduction

Let us face it – most people like three-dimensional visualizations. Whether it is in movies, holograms or games, three-dimensional visualization (literally) adds an extra dimension to pictures. However, three-dimensional data and their visualization can also have scientific archaeological benefits as the appraisal for structures and features can change quite drastically when three-dimensional information is added: this information helps in removing relief distortions from aerial photographs, aids in visualizing a landscape that might be gone long ago or helps to interpret a scene or its features. These are exactly the reasons why stereoscopy has always been very important to aerial archaeologists. This report looks at PhotoScan, a low-cost software package released in mid-2010 that allows extraction of three-dimensional data from aerial images in a very straightforward way. The embedded technology comes from the computer vision research field and uses so-called structure from motion (SFM) and dense stereo-reconstruction techniques. An enormous advantage of these tools is the fact that there are no prerequisites: no information is needed on time and location of image acquisition, nor about the instrument the imagery was acquired with.

## PhotoScan

### *General*

PhotoScan is produced by the Russian-based company AgiSoft LLC. As they advertise, the software is 'an

* Correspondence to: G. Verhoeven, LBI for Archaeological Prospection & Virtual Archaeology, Ludwig Boltzmann Gesellschaft GmbH, Franz-Klein-Gasse 1, A-1190 Wien, Austria.
E-mail: Geert.Verhoeven@archpro.lbg.ac.at



*Received 9 November 2010*
*Accepted 21 December 2010*

advanced image-based solution for creating professional quality three-dimensional (3D) content from still images' (AgiSoft LLC, 2010a). Built to operate on Windows systems (from XP onwards), PhotoScan uses a multitude of JPEG, TIFF, PNG, BMP or MPO files to generate three-dimensional meshes and accompanying mesh textures in an automatic way. Even though the user can set a large number of input parameters, the reconstruction itself is a painless three-step process. At any stage, it is possible to intervene and disable/enable individual photographs, mask parts of the images or import textures and meshes created in other applications. Even though it can be used for reconstructing various objects, several tests have revealed that PhotoScan excels in processing aerial frame imagery. The examples later in this article will illustrate this. The only assumption for a good reconstruction is that the scene to be reconstructed is visible on at least two photographs (AgiSoft LLC, 2010a). The implemented computer vision algorithms will do their very best to estimate the internal parameters of the optical system and the spatial positioning of the image acquisition stations.

## *Processing procedure*

PhotoScan is a computer vision application, often defined as the science that develops mathematical techniques to recover the three-dimensional shape and appearance of objects in imagery (Szeliski, 2010). In general, PhotoScan processes the images in three steps to reconstruct a scenes three-dimensional content.

First, an alignment of the photographs is executed. PhotoScan does this by a technique called structure from motion (SFM; Ullman, 1979). Recently, SFM received a great deal of attention due to Bundler (Snavely, 2010) and Microsoft's Photosynth (Microsoft Corporation, 2010): two freely accessible SFM implementations. Structure from motion allows the reconstruction of three-dimensional scene geometry and camera motion from a sequence of two-dimensional imagery captured by a camera moving around the scene (Fisher *et al.*, 2005; Szeliski, 2010). To do this, the SFM algorithms detect image feature points (i.e. various geometrical similarities such as object edges or other specific details) and subsequently monitor the movement of those points throughout the sequence of multiple images. Using this information as input, the locations of those feature points can be estimated and rendered as a sparse three-dimensional point cloud. As the SFM greatly depends on the accurate knowledge of camera positions, estimating the latter is one of the core components in SFM (Hartley and Zisserman, 2004; Szeliski, 2010). Finally, PhotoScan will end up with three datasets after this first processing stage: (i) a point cloud of typically a few thousand three-dimensional points representing the geometry/structure of the scene (Figure 1B and C); (ii) the camera positions at the moment of image acquisition (Figure 1B and C); (iii) the internal calibration parameters, being focal length, principal point location as well as three radial and two tangential distortion coefficients. Additionally this means that there is no real need to apply calibrated cameras and optics during the image acquisition stage. Since the input of custom calibration parameters is supported, PhotoScan also enables the use of available metric camera information, which might even improve the final reconstruction accuracy.

In a second step the majority of geometric scene details are built by applying a dense, multiview stereo-reconstruction on the aligned image set. Whereas SFM algorithms operate on a sparse set of feature points extracted from the source photographs, these dense reconstruction algorithms operate on the pixel values (Scharstein and Szeliski, 2002; Seitz et al., 2006). As all pixels are utilized, this multiview reconstruction step enables proper handling of fine details present in the scenes and represents them as a mesh (Figure 1D). PhotoScan supports several dense stereo-matching algorithms (called Exact, Smooth, Height Field and Fast), enabling an optimal result for a particular task. In most cases, the Height Field and Exact method are best suited for aerial photographs. Because of its superiority in reconstructing terrain-like features, the Height Field approach can be recommended as the default method for aerial photography. The Exact method, on the other hand, often recovers most terrain detail. This, however, also has implications on processing time as complex integration steps are included. In contrast to the other three methods, the exact reconstruction also requires some hole filling in post-processing. Irrespective of construction method, undistorted photographs are always used. Using the camera calibration parameters obtained from the self-calibration during the alignment stage, PhotoScan removes the lens distortion from the photographs for further processing. Besides parameters such as face count and some filter thresholds, all four reconstruction methods allow the reconstruction quality to be set. Obviously, the high quality settings yield more detailed geometry than the low and medium settings, but they come with a computation time penalty.

Finally, the mesh can be textured with the photographs (Figure 1E). The most important option to be found here is the mapping mode: Generic,



DOI: 10.1002/arp

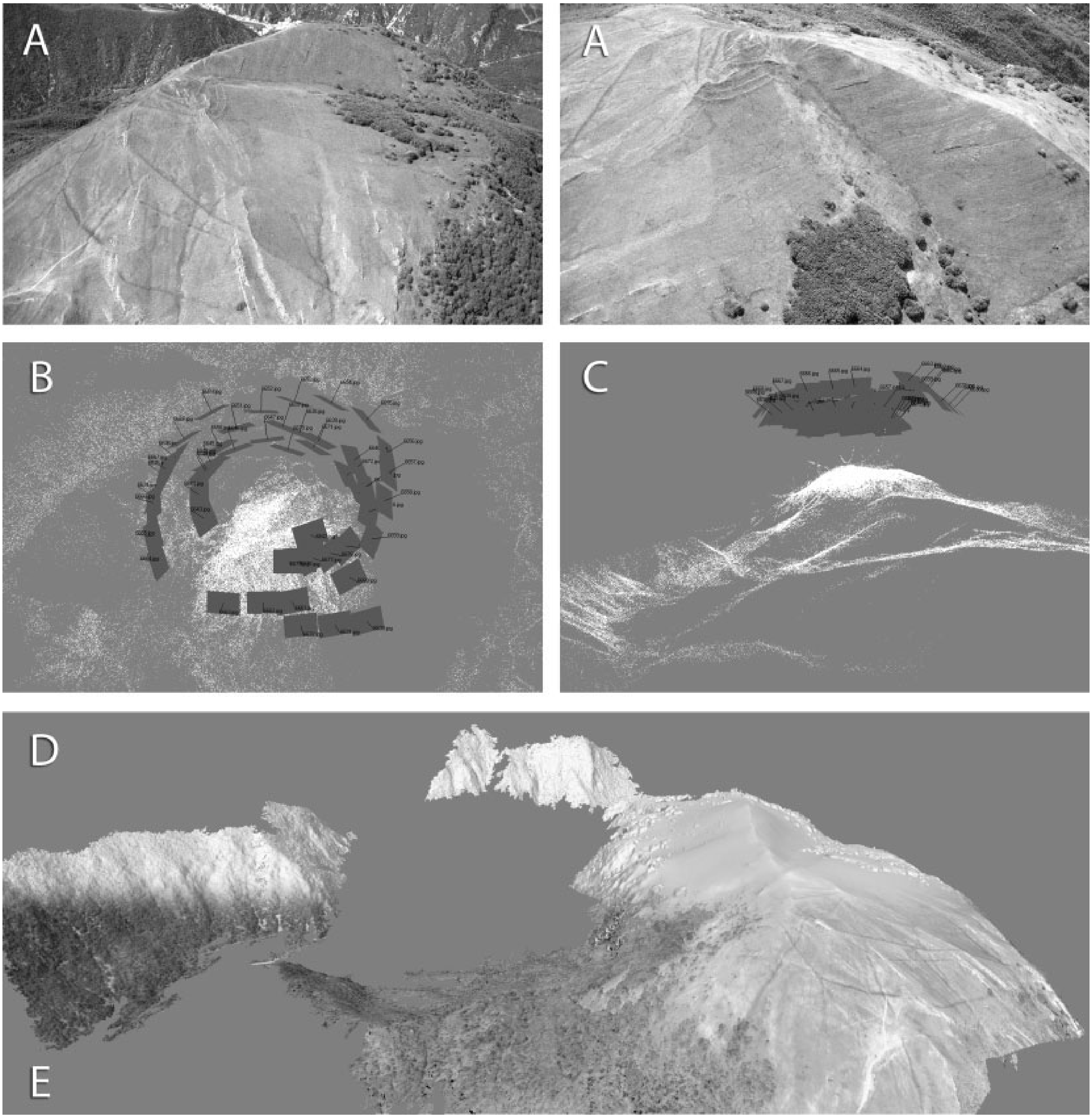


Figure 1. Out of a series oblique photographs (A), PhotoScan calculated the camera positions (B and C), a point cloud (B and C), a (Digital Surface Model; D) and textured DSM (E). (Photographs (A) by R. Goossens, University of Gent, Department of Geography.)

Orthophoto and Single photograph. By default, the generic mapping mode is selected as the texture atlas for arbitrary geometry can be parameterized, i.e. the process of calculating the necessary parameters for a specification of a geometric object. No assumptions are made regarding the geometry and the method will generate textures as uniformly as possible. When processing nearly planar geometry, it might be more desirable to use the orthographic projection. Since a standard orthographic projection produces poor results when vertical surfaces like walls have to be textured, PhotoScan implements an adaptive parameterization approach in its orthophoto mapping mode (Dmitry Semenov, Agisoft LLC, personal



DOI: 10.1002/arp

communication, 2010). It will therefore map horizontal surfaces with the orthophoto parameterization, while vertical regions are textured using the generic mapping mode. Due to this dual approach and the fact that the orthophoto mapping mode generally produces more compact textures, this should be the first option in processing aerial photographs. In almost all cases, the orthographic projection delivered the best results in a shorter time span than the generic mapping mode.

## Guidelines, considerations and requirements

### *Hardware*

By itself, PhotoScan does not impose any limits on the amount of photographs used in the reconstruction, the only exception being the Fast geometry reconstruction mode which is limited to 1024 images (AgiSoft LCC, 2010b). As it is obvious that high-quality reconstructions with large image files are very resource intensive, it is advisable to work with a multicore processor, a decent amount of RAM (minimum 6–8 GB) and a 64-bit operating system. Generally speaking, every quality step comes with the penalty of eight times longer processing. This figure is, however, valid as long as only RAM is applied. As soon as the computer runs out of its main memory, it switches to virtual memory usage and starts to address the page file. At this moment, it is best to quit the application as the whole reconstruction process will slow down enormously. This means that in practice, the amount of photographs to be processed in one run is limited by the amount of RAM available. The general advised strategy is to let PhotoScan solve the complex SFM math of an image set as large as possible, without having to rely on virtual memory. Afterwards, one can 'disable photos' and perform the dense reconstruction in parts. Only when RAM constraints prohibit the alignment of all images at once in a reasonable time, should the collection be split in several subsets called chunks. After all three processing steps are applied per chunk, PhotoScan has a tool to align these partial models. In most situations, an overlap of one to two images is sufficient for chunk alignment, although it might be more if the images lack any good texture details (AgiSoft LCC, 2010b). Finally, a high-end graphical card is extremely helpful as well, because PhotoScan supports OpenCL (Open Computing Language), which is a programming platform that allows access to the Graphics Processing Unit (GPU) for general purpose computing, i.e. not purely graphical. By exploiting the full power of the GPUs on the video card, three-dimensional reconstruction can be sped up considerably using this multi-GPU + CPU approach.

### *Programme settings*

PhotoScan sets an upper limit of image feature points that should be used in the alignment of the dataset. By default, 40 000 is set as a reasonable value. Only in very specific cases would increasing this parameter to a higher value improve the alignment. However, the computation time needed also increases and one even risks less accurate alignment and/or a degraded reconstruction quality. So practically, 5000 to 40 000 feature points is a good working range.

### *Image acquisition*

Even though PhotoScan works with ordered and unordered image collections, some considerations can hugely improve the alignment and reconstruction accuracy. During image acquisition, it is important to make sure that (at least the majority of) the cameras do not observe each other (AgiSoft LCC, 2010b). As PhotoScan can deal with incidence angles from 45° onwards, it is obvious that both vertical and low-oblique aerial imagery can be handled very well. Furthermore, there are no real restrictions on the camera and lens applied. If the focal length information cannot be extracted from the file's EXIF data, the user can always override the programmés standard value of 50 mm, thereby considerably improving the final alignment and helping PhotoScan to prevent false alignments. As can be expected from a computer vision program, this means that everything from a cell phone to a professional still camera is useable. Additionally, video capture can be processed as PhotoScan can import individual video frames.

## Case studies

In order to test the software in controlled and uncontrolled conditions, three-dimensional geometry was recovered from a wide variety of old and new imagery.

### *Oblique imagery*

The photographs acquired from the cabin of a low-flying airplane still form a very large part of all aerial archaeological input data. In general, those images are generated by uncalibrated hand-held cameras and are quite oblique in nature. Exactly this type of source data were used as input for the first case study. The images were acquired in May 2005 using a 35-mm lens fitted onto a Canon EOS-1Ds. The photographs show the Monte Primo, located in central Adriatic Italy, where a Protohistoric hilltop site overlooks the crucial passage of the Potenza river valley through the Apennines (Bonomi Ponzi, 1992). On the photographs (Figure 1A), several earthworks are discernable on the summit of the mountain. In total, 49 photographs were used and Figure 1 displays the results after each individual step. The software produced a good representation of this quite complex mountainous landscape.

Besides aeroplanes, all kinds of unmanned platforms can be used to gather archaeological imagery from relatively low altitudes, enabling so-called low-altitude aerial photography (LAAP; Verhoeven, 2009). In a second example, an unordered collection of images acquired by means of a Helikite (Verhoeven *et al*., 2009) was processed. The end result (Figure 2A and C) neatly displays the various excavated structures of the 'Casa Valentini' amphora kiln site in central Adriatic Italy (Vermeulen *et al*., 2009). Using orthophoto mapping, one can generate orthophoto-like textures. Nonetheless, this needs to be taken into account and PhotoScan uses the principal plane of the data as the projection plane, which can differ from the *XY* plane in the case of a slant in the reconstructed surface region. As vertical surfaces are textured separately, there might also be some holes in the orthophoto part of the texture. However, the model and its texture can be saved as an *.OBJ file and subsequently opened by a dedicated meshing package such as Geomagic Studio, MeshLab or VRMesh. In there, tools are available to redefine the *XY* plane, which means that a nadir view of the model will result in an orthophoto (Figure 2B).

## *Conventional stereo-coverage*

In most computer vision applications, an image point needs to be visible on at least three images in order to be reconstructed. PhotoScan is not limited by this threshold, as can be seen in Figure 3 where only two input images were used. Even though all points are maximally imaged twice, a three-dimensional model could be generated from this ordered image set. The final reconstruction was also very easy because the images were nadir (i.e. vertical) photographs, taken during a stereo-coverage flight above central Adriatic Italy in the 1960 s. Notwithstanding the fact that PhotoScan can deal with incidence angles from 45° onwards, perpendicular incidence angles (i.e. 90°) are still preferable (AgiSoft LCC, 2010b). Besides aligning scans, PhotoScan has tools to decimate meshes and

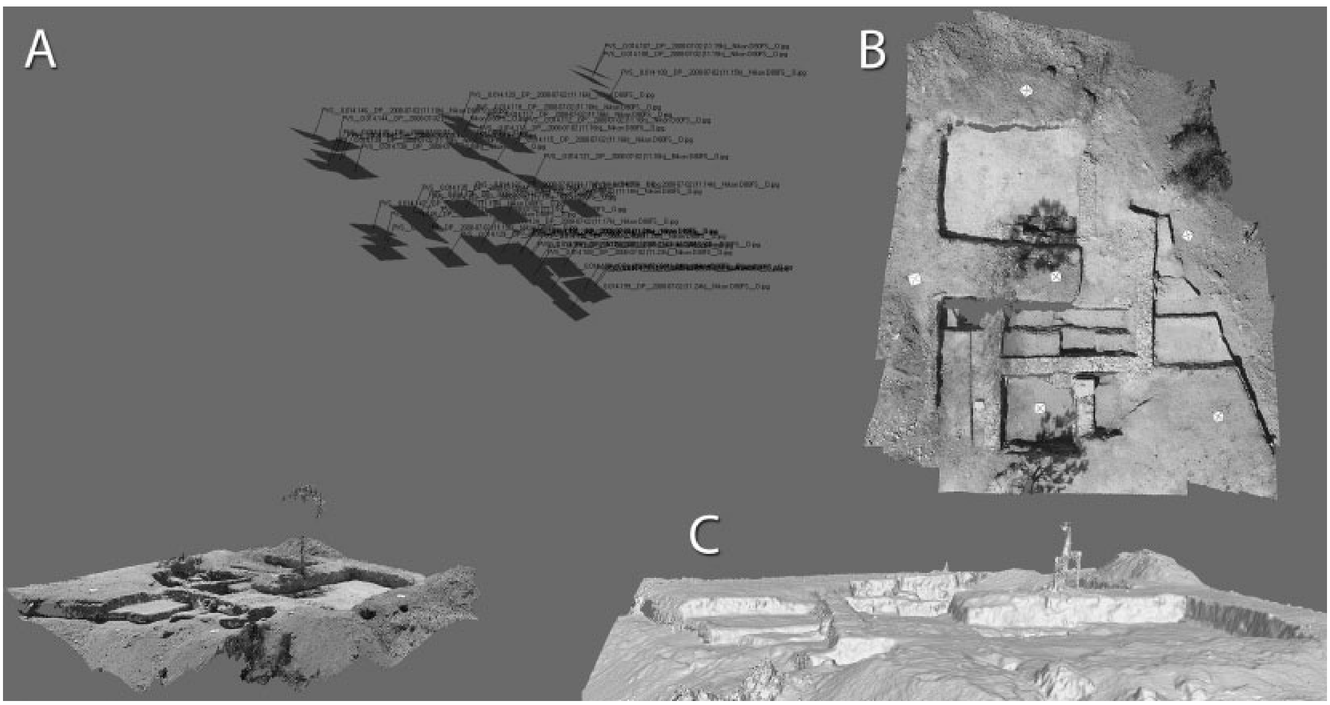


Figure 2. PhotoScan solved all the camera positions of the Helikite image set (A). An orthophoto could be generated (B) because the excavated structures were recovered quite well (C).



DOI: 10.1002/arp

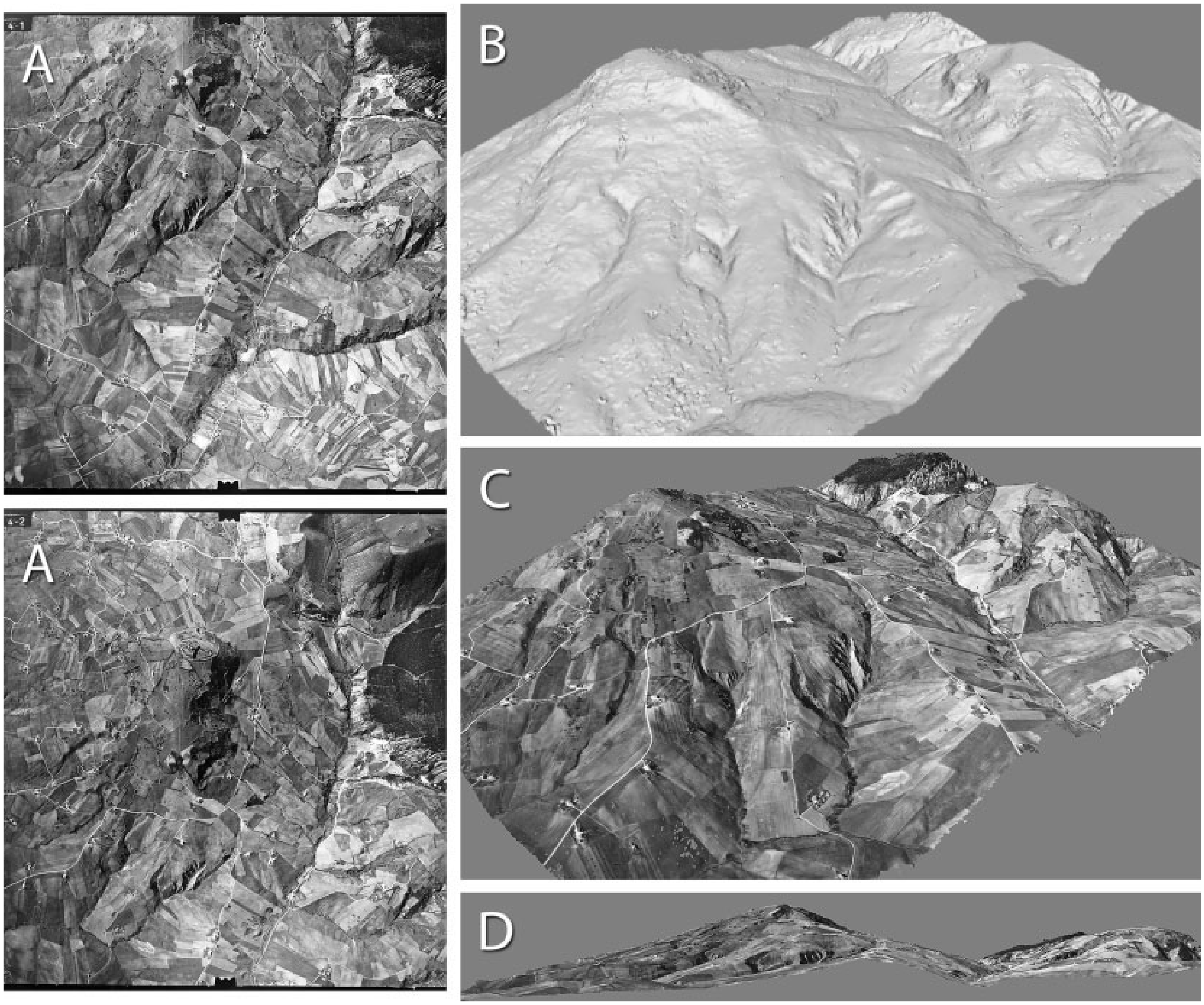


Figure 3. (Digital Surface Model; B) and textured DSM (C) of an old stereo-pair (A). A frontal view on the model is displayed in (D).

mask photographs. By masking the image parts that do not belong to the scene under study, the texture does not display the camera frame and fiducial marks, but only takes the scene pixels into consideration.

## Drawbacks and future approaches

PhotoScan works well, but it is not perfect of course. When dealing with very oblique images or images that have a dissimilar appearance (e.g. shot at different moments), PhotoScan might create erroneous links or is sometimes even unable to align the photographs. In those situations, Bundler and Photosynth often achieve better results. Nevertheless, one can always delete the problem photographs and perform realignment. Currently, the program does not have any tools to georeference the mesh and create an orthophotograph. Because of its wide variety in output formats (including PDF for easy sharing), the models can be opened in many other applications. As the resulting reconstructions are Euclidian, it means that only the scale of the reconstruction is undefined. By defining the distance between two known points, one can measure inside the model, while three ground-control points can pin down the reconstruction in a predefined geographical frame. Currently, ongoing tests aim to assemble a package of other (low-cost and open source) software solutions to create a simple and affordable workflow that enables both georeferencing of the calculated meshes as well as orthophoto production. This implies that in the very near future,



DOI: 10.1002/arp

orthophoto production should become available to literally everybody, and at small cost. Furthermore, accuracy studies are planned to compare and quantify the differences with the more common rigorous photogrammetric approaches.

## Conclusion

By applying the latest multiview algorithms, PhotoScan proves to be an advanced computer vision solution that enables the creation of high-quality three-dimensional content from a series of overlapping aerial images. Imagery can be generated in both controlled and uncontrolled situations with cameras ranging from simple cell phone cameras to highly professional medium-format imagers. Even though other computer vision solutions sometimes deal better with very oblique imagery, PhotoScan has the advantage of offering an integrated and still affordable toolbox that works with both still and motion pictures. Any part that is visible on at least two frames has a high potential to be reconstructed. Afterwards, one run through a three-step processing chain is all it needs. Later, new photographs can always be added and old ones deleted or disabled. Even though it seems too good to be true, creating three-dimensional visualizations for virtual displays or realistic models for site monitoring or publications has never been so easy.

## References

AgiSoft LLC. 2010a. *AgiSoft PhotoScan*. http://www.agisoft.ru/products/photoscan/ (date of access: 30 October 2010).

AgiSoft LCC. 2010b. *AgiSoft PhotoScan Forum*. http://www.agisoft.ru/forum.php?forum=7 (date of access: 03 November 2010).

Bonomi Ponzi L. 1992. Occupazione del territorio e modelli insediativi nel territorio Plestino e Camerte in età protostorica. In *La civiltà Picena nelle Marche. Studi in onore di Giovanni Annibaldi, Ancona 10-13 luglio 1988*. Gianni Maroni Editore: Ripatransone; 204–241.

Fisher R, Dawson-Howe K, Fitzgibbon A, Robertson C, Trucco E. 2005. *Dictionary of Computer Vision and Image Processing*. John Wiley & Sons: Hoboken.

Hartley R, Zisserman A. 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press: Cambridge.

Microsoft Corporation. 2010. *Microsoft Photosynth*. http://photosynth.net (date of access: 28 October 2010).

Scharstein D, Szeliski R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* **47**: 7–42.

Seitz S, Curless B, Diebel J, Scharstein D, Szeliski R. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. *Proceedings of the CVPR '06 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Volume 1*. IEEE Computer Society: Washington, DC; 519-526.

Snavely N. 2010. *Bundler: Structure from Motion for Unordered Image Collections*. http://phototour.cs.washington.edu/bundler/ (date of access: 28 October 2010).

Szeliski R. 2010. *Computer Vision: Algorithms and Applications*. Springer-Verlag: London.

Ullman S. 1979. The interpretation of structure from motion. *Proceedings of the Royal Society of London* **B203**: 405–426.

Verhoeven G. 2009. Providing an archaeological bird's-eye view – an overall picture of ground-based means to execute low-altitude aerial photography (laap) in archaeology. *Archaeological Prospection* **16**: 233–249.

Verhoeven G, Loenders J, Vermeulen F, Docter R. 2009. Helikite aerial photography (HAP) – a versatile means of unmanned, radio-controlled, low-altitude aerial archaeology. *Archaeological Prospection* **16**: 125–138.

Vermeulen F, De Dapper M, Mušič B, Monsieur P, Vereyke H, Carboni F, Dralans S, Verhoeven G, Verdonck L, Hay S, Sterry M, De Paepe P, De Seranno S. 2009. Investigating the impact of Roman urbanisation on the landscape of the Potenza Valley: a report on fieldwork in 2007. *BABESCH* **84**: 85–110.



DOI: 10.1002/arp