ISSN 1756-753X

AARGnews

The newsletter of the Aerial Archaeology Research Group

Offprint from:
Number 44 March 2012

Contents

aarg
Aerial Archaeology Research Group

Computer vision techniques: towards automated orthophoto production

G. Verhoeven[1,2], M. Doneus[3,1] and Ch. Briese[1,4]

Archaeological practice has always been facing huge challenges in ways of fast and accurate three-dimensional recording, whether it is during excavations, artefact study or mapping of archaeological remains throughout the landscape. Of all archaeological remote sensing techniques, aerial photographic reconnaissance from a low-flying aircraft has been the workhorse since it is one of the most effective methods for site discovery. Once airborne, the archaeologist flies around in a certain area, trying to detect possible archaeological and palaeoenvironmental remains, usually indicated by visible marks. After detection, the marks are orbited and captured in more or less oblique photographs from various positions using a hand-held camera equipped with a lens that is commonly uncalibrated.

However, before aerial images can be used to map the partly-eroded and sub-surface archaeological and palaeoenvironmental features, they need to be **georeferenced**. This georeferencing process assigns spatial information to the imagery to explicitly define their location and rotation in respect to a specific Earth-related coordinate frame. Obviously, accurate airphoto georeferencing and subsequent orthophoto generation is an absolutely necessary prerequisite for the further study and integration of aerial images with other data sets (such as old maps and geophysical plots) or for a multi-temporal analysis in a GIS-like environment. As one can expect, there are a variety of georeferencing and orthophoto generation procedures and not all are equally-well suited for every type of aerial image. Since a photograph maps the geometric three-dimensional scene properties (x, y, z) to a two-dimensional plane (X, Y), the geometry of the scene gets seriously distorted. Amongst many other factors, the geometrical deformations induced by the topographical relief (called relief displacement), the tilt of the camera axis (called tilt displacement) and the distortion of the optics are most considerable. Although it is *sensu stricto* not covered by its definition, georeferencing often involves the necessary steps to remove these factors in order to place each image pixel on its true location on the Earth's surface. To do this, several approaches and software solutions exist.

In general, archaeologists georeference individual photographs using simple methods such as (planar) rectification, polynomial correction or piecewise affine warping embedded in dedicated low-cost packages such as Airphoto (Scollar 2002) and AERIAL (Haigh 2005) or almost any form of GIS-software. Although these approaches are popular and might deliver fairly good metrical information when the terrain variations are quite moderate, the methods are suboptimal in hilly areas or sites with considerable relief variations. To this end, advanced ortho-correction procedures embedded in more expensive photogrammetric packages such as Trimble INPHO Photogrammetric System or Leica Photogrammetry Suite must be applied (Doneus 2001). Although these rigorous corrections produce superior geometric quality because they consider all main geometric influencing factors, they suffer from the fact that calibrated camera information and an accurate, high-resolution digital surface model are essential: two prerequisites that are generally not met in aerial archaeology. Besides, the process is time-consuming while a varying degree of photogrammetric skills and experience are required. Thanks to some recent advances in the fields of computer vision and photogrammetry as well as the improvements in processing power of computer processors and graphical cards, it is currently possible to generate orthophotos of aerial imagery – collected during the previously described type of oblique aerial archaeological reconnaissance – with the straightforwardness of the former approach and the accuracy of the latter.

[1]LBI for Archaeological Prospection and Virtual Archaeology, Vienna, Austria. Geert.Verhoeven@archpro.lbg.ac.at
[2] Department of Archaeology, Ghent University, Belgium
[3] Vienna Institute for Archaeological Science, University of Vienna, Austria
[4] Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Austria

In a first phase, the process uses a technique called Structure from Motion (SfM; Ullman 1979). In essence, SfM allows the reconstruction of three-dimensional scene geometry and the exact position of the cameras during image acquisition from a sequence of two-dimensional imagery captured by a camera moving around the scene (Szeliski 2011 – Figure A1). To do so, SfM relies on algorithms that detect feature points for each image (Figure A2) and subsequently tries to match those 2D points throughout the image series (Figure A3). Using these obtained point correspondences SfM computes the locations of those feature points and renders them as a sparse 3D point cloud that represents the structure of the scene in a local coordinate frame (Figure B). As SfM greatly depends on the accurate knowledge of camera positions, estimating the latter is one of the core components in the whole process (Hartley and Zisserman 2003). More specifically, the complete projection geometry of all images gets computed: the interior camera calibration parameters (focal length, the principal point location plus lens distortion coefficients), the position of the camera projection centre and six exterior orientation parameters defining the camera orientation at the moment of image acquisition (Robertson and Cipolla 2009 – Figure 1B). At this stage, the reconstruction is still expressed in a local coordinate framework and equivalent of the original scene up to a global scale and rotation factor. To transform the camera positions and point cloud into an absolute coordinate system a Helmert similarity transformation, using at least three ground control points with known altitude values, is applied.

Recently, SfM has received a great deal of attention due to Bundler, Microsoft’s Photosynth and Autodesk’s Project Photofly (now called 123D Catch): three SfM implementations that are feely available on the Web. Besides, commercial SfM solutions are also available (such as AgiSoft’s PhotoScan or Pix4D’s cloud processing software). Most of these software solutions also come with additional functionality to yield a dense representation of the scene’s surface geometry using one or more multi-view stereo (MVS) algorithms. Because such dense MVS solutions operate on the pixel values instead of on the feature points (Seitz et al. 2006), this additional step enables the generation of detailed three-dimensional point clouds or triangular meshes (Figure C). When working with aerial images, the resulting model can be considered a digital surface model (DSM): a numerical representation of the topography and all its imposed structures such as trees and houses. As is known from conventional orthophoto generation, such a dense DSM is elementary when one wants to generate true orthophotos in which all objects with a certain height (such as houses, towers and trees) are also accurately positioned. Since all necessary information is available, a detailed and accurate orthophoto can now be produced (Figure D).

Although the presented algorithms are best run on computers with multicore processors, a decent amount of RAM (minimum 8 GB), a 64-bit operating system and a high-end graphical card, they offer the enormous advantage that they can be used with archaeologists’ usual oblique photographs. Apart from a sufficient number of sharp images covering the scene to be reconstructed and at least three GCPs to pin down the reconstruction, no other information is needed. Besides, only a minimal technical knowledge and user interaction are required. However, it has to be stressed that it is not all roses here: the method is not applicable for an individual image and the determination of the correct camera projection geometry can fail when dealing with blurred, noisy and badly exposed images or photographs that have a very dissimilar appearance (e.g. due to major underexposure or changing topographic terrain parameters). For a more elaborate overview and multiple examples of this orthophoto procedure applied on aerial archaeological imagery, consider Verhoeven et al. (2012a). Additionally, research by Doneus et al. (2011) proved how well this method holds up when compared with terrestrial laser scanning in an excavation context, while Verhoeven et al. (2012b) thoroughly evaluated the positional accuracy of the generated orthophotographs. This type of quality control and documentation is essential in order to ensure the proper quality of the final orthophoto.

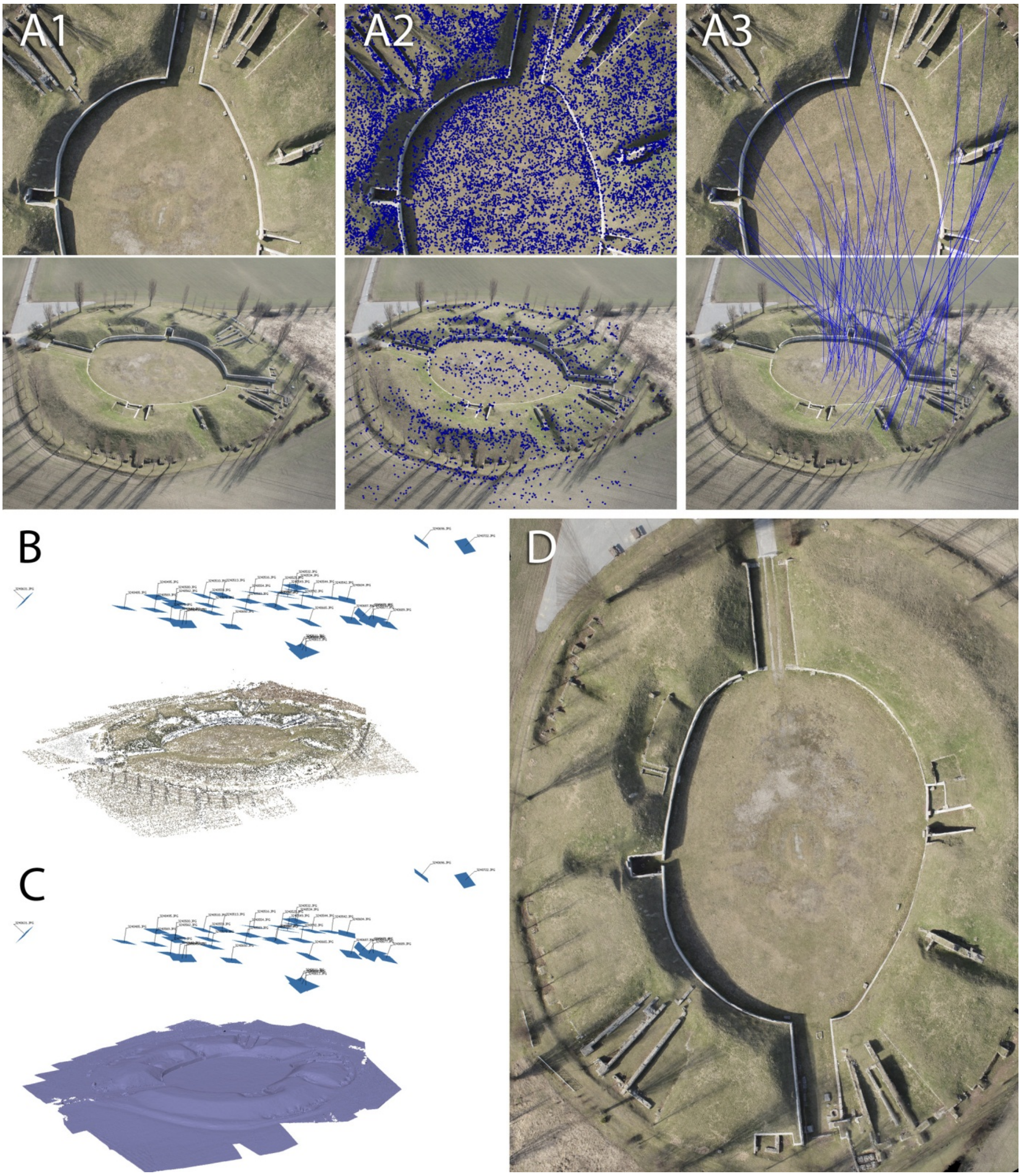

Figure 1 – All individual workflow steps used to create an orthophotograph of the 2nd century AD amphitheatre of the civil Roman town of Carnuntum. (A1) displays two out of the forty digital photographs used; (A2) indicates the feature points that were detected in those images and (A3) the matches found between those feature points. (B) shows the sparse point cloud and the camera positions provided by the SfM solution. After the dense reconstruction stage, a DSM is created (C). Using all these data, an orthophotograph (D) can be generated. All aerial images were acquired at the end of March 2011 around 9.30 h using a radio-controlled Microdrone md4-1000 quadrocopter and an Olympus PEN E-P2 (a 12.3 megapixel mirrorless Micro Four Thirds camera) equipped with an Olympus M.Zuiko Digital 17 mm f/2.8 lens.

Acknowledgements

The Ludwig Boltzmann Institute for Archaeological Prospection and Virtual Archaeology (archpro.lbg.ac.at) is based on an international cooperation of the Ludwig Boltzmann Gesellschaft (A), the University of Vienna (A), the Vienna University of Technology (A), the Austrian Central Institute for Meteorology and Geodynamic (A), the office of the provincial government of Lower Austria (A), RGZM-Roman- Germanic Central Museum Mainz (D), RAÄ-Swedish National Heritage Board (S), IBM VISTA-University of Birmingham (GB) and NIKU-Norwegian Institute for Cultural Heritage Research (N).

References

Doneus, M., 2001. Precision mapping and interpretation of oblique aerial photographs. *Archaeological Prospection* 8, 13–27.

Doneus, M., Verhoeven, G., Fera, M., Briese, Ch., Kucera, M. and Neubauer, W. 2011. From deposit to point cloud – A study of low-cost computer vision approaches for the straightforward documentation of archaeological excavations. *Proceedings of the XXIIIrd International CIPA Symposium, Prague, Czech Republic, September 12 - 16, 2011*.

Haigh, J.G.B., 2005. From Photographs to Maps. A Collaborative Development. In: Bourgeois, J., Meganck, M. (Eds.), *Aerial Photography and Archaeology 2003. A Century of Information*. Academia Press, Ghent, 201-212.

Hartley, R. and Zisserman, A., 2003. *Multiple view geometry in computer vision*. 2nd ed. Cambridge: Cambridge University Press.

Robertson, D.P. and Cipolla, R., 2009. Structure from Motion, in: M. Varga (Ed.), *Practical image processing and computer vision*. John Wiley and Sons Ltd: New York.

Scollar, I., 2002. Making Things Look Vertical. In: Bewley, R., Raczkowski, W. (Eds.), *Aerial Archaeology: Developing Future Practice*. NATO Science Series 337. IOS press: Amsterdam, 166-172.

Seitz, S.M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., 2006. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, in: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR’06)*. IEEE: Washington, 519–528.

Szeliski, R., 2011. *Computer vision. Algorithms and applications*. Springer, New York.

Ullman, S., 1979. The interpretation of structure from motion. *Proceedings of the Royal Society of London B* 203, 405–426.

Verhoeven, G., Doneus, M., Briese, C. and Vermeulen, F., 2012a. Mapping by matching - A computer vision-based approach to fast and accurate georeferencing of archaeological aerial photographs. *Journal of Archaeological Science*. (*in press*)

Verhoeven, G., Taelman, D. and Vermeulen, F., 2012b. Computer vision-based orthophoto mapping of complex archaeological sites: the ancient quarry of Pitaranha (Portugal-Spain). *Archaeometry*. (*in press*)