Journal of Archaeological Science 39 (2012) 2060–2070

Contents lists available at SciVerse ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Mapping by matching: a computer vision-based approach to fast and accurate georeferencing of archaeological aerial photographs

G. Verhoeven [a,b,d,*], M. Doneus [b,d], Ch. Briese [c,d], F. Vermeulen [a]

[a] Department of Archaeology, Ghent University, Belgium
[b] VIAS – Vienna Institute for Archaeological Science, University of Vienna, Austria
[c] Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Austria
[d] LBI for Archaeological Prospection and Virtual Archaeology, Vienna, Austria

ARTICLE INFO

Article history:
Received 28 July 2011
Received in revised form
16 February 2012
Accepted 21 February 2012

Keywords:
3D model
Aerial archaeology
Aerial photograph
Computer vision
Dense multi-view stereo reconstruction
Orthophoto
Structure from motion

ABSTRACT

To date, aerial archaeologists generally apply simple rectification procedures or more expensive and time-consuming orthorectification algorithms to correct their aerial photographs in varying degrees for geometrical deformations induced by the topographical relief, the tilt of the camera axis and the distortion of the optics. Irrespective of the method applied, the georeferencing of the images is commonly determined with ground control points, whose measurement and identification is a time-consuming operation and often limits certain images from being accurately georeferenced. Moreover, specialised software, certain photogrammetric skills, and experience are required. Thanks to the recent advances in the fields of computer vision and photogrammetry as well as the improvements in processing power, it is currently possible to generate orthophotos of large, almost randomly collected aerial photographs in a straightforward and nearly automatic way. This paper presents a computer vision-based approach that is complemented by proven photogrammetric principles to generate orthophotos from a range of uncalibrated oblique and vertical aerial frame images. In a first phase, the method uses algorithms that automatically compute the viewpoint of each photograph as well as a sparse 3D geometric representation of the scene that is imaged. Afterwards, dense reconstruction algorithms are applied to yield a three-dimensional surface model. After georeferencing this model, it can be used to create any kind of orthophoto out of the initial aerial views. To prove the benefits of this approach in comparison to the most common ways of georeferencing aerial imagery, several archaeological case studies are presented. Not only will they showcase the easy workflow and accuracy of the results, but they will also prove that this approach moves beyond current restrictions due to its applicability to datasets that were previously thought to be unsuited for convenient georeferencing.

1. Introduction

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. On the one hand, the non-invasive approach yields easily interpretable imagery with abundant spatial detail. On the other hand, the method is driven by the specific nature of the partly eroded or sub-surface archaeological features, which show up on the surface under certain conditions as “visibility marks”: i.e. indirect indicators of archaeological residues due to the changed properties of the local topography or the soil matrix (Bewley and Rączkowski, 2002; Brophy and Cowley, 2005; Crawford, 1924; Scollar et al., 1990; Wilson, 2000). According to their nature, archaeologists distinguish (i) shadow marks (when earthworks are thrown into relief by low slanting sunlight), (ii) soil marks (due to varying chemical and physical properties affecting the soil colour on the surface), (iii) crop marks (due to variable growth of the vegetation), (iv) snow/frost marks (due to differential snow accumulations and differential melting of snow or frost) as well as the less frequent flood and wind marks. Once detected from the air, they are orbited and documented from various positions mostly using small and medium-format Single-Lens Reflex (SLR)-cameras.

While discovery and photographic documentation of archaeological sites is commonly seen as the prime aim of aerial archaeology, the need for detailed interpretation (interpretative mapping

* Corresponding author. Department of Archaeology, Faculty of Arts and Philosophy, Ghent University (UGent), Sint-Pietersnieuwstraat 35 (UFO – 110.016), B-9000 Gent, Belgium. Tel.: +32 485 65 05 05.
E-mail address: Geert.Verhoeven@UGent.be (G. Verhoeven).

0305-4403/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jas.2012.02.022

– cf. Doneus, 2001) of the photographs is rarely emphasized and executed (see also Bewley and Rączkowski, 2002). As a result, millions of aerial photographs are just stored in archives where their archaeological information cannot (or will not) be exploited efficiently. This is unfortunate, since a thorough understanding of archaeological landscapes is based on combining the interpreted evidence from various prospection methods.

The neglect of this interpretative mapping may have multiple reasons (such as available resources), but one of the most important ones is likely the time-consuming georeferencing process, which is a necessary prerequisite for any kind of accurate mapping and data integration (Doneus and Neubauer, 1998). This process, which is also known as ground registration, assigns spatial information to the imagery to explicitly define their location and rotation in respect to a specific Earth-related coordinate frame. After georeferencing, several multi-temporal aerial observations can be mosaicked into an extensive overall view of an archaeological region that can serve as a basic information layer for further prospection, excavation, protection measures and heritage management.

Obviously, aerial archaeology is in need of fast and accurate georeferencing approaches that allow straightforward orthophoto production of a wide variety of images. This article elaborates on such an approach and presents a method to automate the important, but recurring task of orthophoto generation. The proposed methodology tries to overcome the conventional georeferencing problems related to archaeological aerial frame images. To this end, the routine exploits some of the technological improvements in hardware configurations as well as state-of-the-art algorithms mainly developed in the field of computer vision and photogrammetry. Before illustrating this approach by means of three case studies (Section 4), the generally applicable approaches to the georeferencing of aerial photos (Section 2) and the components of the newly presented approach (Section 3) will be shortly reviewed below.

2. Georeferencing and orthophoto generation in aerial archaeology

In the case of aerial imaging, it is important to note that an optical remote sensing instrument such as a (digital) photographic camera generates images of the spatial distribution of the Earth's upwelling radiance, but always degrades this analogue signal to a certain extent. Consequently, the final image is never a very faithful reproduction of the real-world scene. Besides the radiometric and spectral transformations that occur, the geometric three-dimensional (3D) properties are mapped to a 2D plane (Fig. 1). This mapping result (i.e. the photograph) is influenced by a wide variety of factors, of which those induced by the topographical relief, the tilt of the camera axis 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, a wide variety of approaches and software solutions exist.

In general, archaeologists often fit tilted images to a flat surface, a process which is called (planar) rectification and which requires a projective transformation (Wolf and Dewitt, 2000). Although these rectified images do not suffer any more from tilt displacements, they still contain scale variations and image displacements due to the topographic relief (hills, buildings etc.). Consequently, projective transformations can only be considered "archaeologically sufficient" when dealing with completely flat areas. If the aerial view suffers from topographically induced deformations (i.e. the relief displacements), georeferencing often employs polynomial corrections, spline algorithms or piecewise affine warpings embedded in archaeologically-dedicated tools such as AERIAL (Haigh, 2005) and Airphoto (Scollar, 2002). Although these approaches are very popular and might deliver fairly good metrical information when the terrain variations are quite moderate, the methods are often sub-optimal because they do not (or only partly) eliminate all the image displacements, the distortion of the optics and – to a lesser extent – the atmospheric refraction. To deal with those issues and create planimetrically correct true orthophotographs, more advanced ortho-correction offered by programs such as Leica Photogrammetry Suite or Trimble INPHO Photogrammetric System needs to be applied. Although these more expensive packages embed rigorous orthorectification algorithms to produce superior geometric quality, they suffer from the fact that photogrammetric skills, calibrated camera information and an accurate, high-resolution DSM are essential, three conditions that are generally not met in aerial archaeology.

Irrespective of the method applied, the georeferencing of the images is commonly determined with ground control points

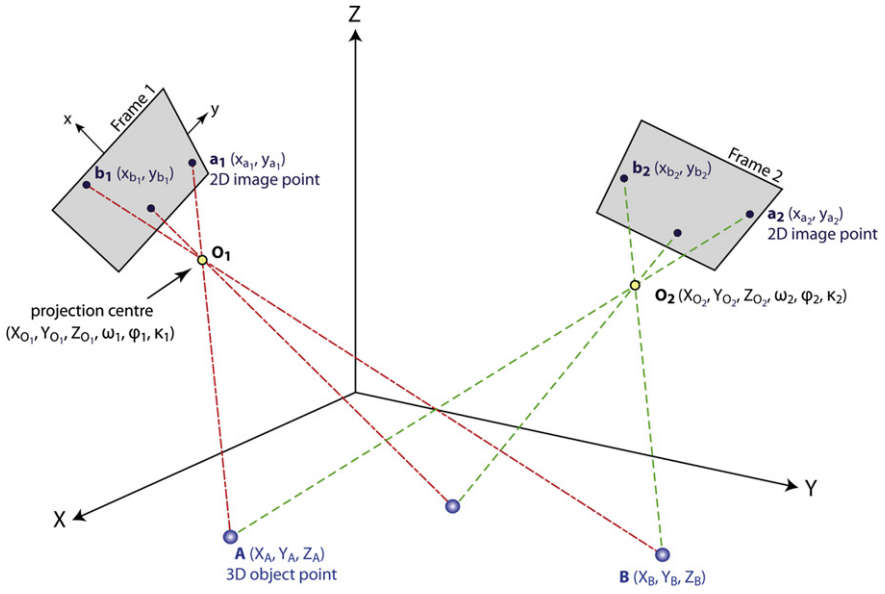

Fig. 1. Mapping of 3D object points onto 2D points in two aerial frame images.

(GCPs), whose manual measurement and identification is a time-consuming operation that requires experience while being bound to certain prerequisites. As a result of all these issues, many valuable aerial images never get properly georeferenced. Often, the georeferencing is suited for small-scale mapping but inadequately accurate for a detailed multi-temporal and multi-method analysis.

3. Structure from motion and multi-view stereo

A lot of tools and methods exist to obtain information about the 3D geometry of scenes from 2D views. One of the possibilities is to use multiple images from the same scene. Using techniques mainly developed in the field of photogrammetry, an image point occurring in at least two views can be reconstructed in 3D (Fig. 1). However, this can only be performed if the projection geometries of the images are known. These are expressed by the interior or inner calibration parameters of the camera (also called camera intrinsics and comprising the focal length, the principal point location plus lens distortion coefficients) and camera poses (i.e. the projection centre location and image orientation defined by the six exterior/outer orientation parameters or camera extrinsics – Fig. 1) at the moment of image acquisition (Robertson and Cipolla, 2009).

While the discipline of photogrammetry derives these parameters mainly from well-distributed GCPs and tie points, a Structure from Motion (SfM; Ullman, 1979) approach allows simultaneous computation of both this relative projection geometry and a set of 3D points, using only corresponding image features occurring in a series of overlapping photographs captured by a camera moving around the scene (Fisher et al., 2005; Quan, 2010; Szeliski, 2011). To do this, the SfM relies on algorithms that detect and describe local features for each image and subsequently match those 2D points throughout the multiple images. Afterwards, a number of potential correspondences are obtained. Using this set of correspondences as input, SfM computes the locations of those interest points and renders them as a sparse 3D point cloud that represents the geometry/structure of the scene in a local coordinate frame. As mentioned previously, the camera pose and internal camera parameters are also retrieved (Hartley and Zisserman, 2003; Szeliski, 2011). There is thus no real need to apply calibrated cameras and optics during the image acquisition stage (Quan, 2010), which makes the procedure very flexible and well-suited for almost any kind of imagery.

SfM algorithms were developed in the field of computer vision, which is often defined as the science that develops mathematical techniques to recover a variety of information from images. This image data can take many forms, such as multi-dimensional imagery from medical scanners, stereo photographs, video sequences or views from multiple still cameras. Initially, many computer vision applications were focused on robotic vision and automation. However, the last decade witnessed a shift of focus to 3D visualisations and virtual reality. Additionally, many new insights in the geometry of multiple images were obtained (see Hartley and Zisserman, 2003 for a good overview). Recently, SfM has received a great deal of attention due to Bundler (Snavely, 2010) and Microsoft's Photosynth (Microsoft Corporation, 2010): two SfM implementations that are freely available on the Web. To date, several SfM-based packages can be applied to obtain a (semi-)automated processing pipeline for image-based 3D visualisation. In this study, the commercial package PhotoScan (version 0.8.1 build 877) from AgiSoft LLC is used.

Since it is tailored towards high-level 3D model generation, PhotoScan complements the SfM approach with a variety of dense multi-view stereo (MVS) algorithms. One could for example interpolate the sparse set of 3D SfM points, but this would yield a far from optimal result. Therefore, a MVS reconstruction is used to compute a dense estimate of the surface geometry of the observed scene. Because these solutions operate on the pixel values instead of on the feature points (Scharstein and Szeliski, 2002; Seitz et al., 2006), this additional step enables the generation of detailed three-dimensional meshed models from the initially calculated sparse point clouds, hence enabling proper handling of fine details present in the scenes. More specifically, PhotoScan (version 0.8.1) offers three reconstruction methods (besides its "Fast" method). Each of those methods uses a pair-wise binocular stereo approach (Bradley et al., 2008) to compute a depth estimate (i.e. the distance from the camera to the object surface) for almost every image pixel of each view (Mellor et al., 1996; Pollefeys et al., 2004). Fusing the resulting independent depth maps of all the images yields a single 3D model which is afterwards approximated by a triangular mesh. PhotoScan's different reconstruction methods (Exact, Smooth and Height-field) differ by the way how these individual depth maps are merged into the final digital model.

When working with aerial images, the model calculated can be considered a Digital Surface Model (DSM): a numerical representation of the topography and all its imposed structures such as trees, houses and any other object that did not move. As is known from conventional orthorectification, such a dense DSM is elementary when one wants to generate so-called true orthophotos in which all objects with a certain height (such as houses, towers and trees) are also accurately positioned (Braun, 2003; Kraus, 2002).

In a final step, a textured 3D mesh can be created by a texture mapping using a particular selection of the initial images. At this stage, the reconstructed 3D scene is still expressed in a local coordinate framework and equivalent of the original scene up to a global scale and rotation factor. In order to transform the surface model into an absolute coordinate system, a Helmert similarity transformation is applied. Since seven parameters (three translations, one scale, three rotations) must be determined for this spatial transformation, at least three GCPs with known altitude values are needed (although the use of more GCPs is always advisable).

Because high-quality reconstructions with large image files are very resource intensive, multicore processors, a decent amount of RAM (minimum 8 GB) and a 64-bit operating system are essential. Since PhotoScan supports the OpenCL (Open Computing Language) programming platform, it can access the Graphics Processing Units (GPUs) of the video card when intensive computing tasks have to be executed, additionally helping in shortening processing times.

4. Case studies

When archaeologists fly around in small aeroplanes to acquire images of the visibility marks seen on the Earth's surface, they acquire single images that are mostly oblique in nature. Often, the aircraft may orbit the area of interest to maximise the photographer's options in getting several shots from a variety of angles, enabling the archaeological site to be documented as completely as possible with one camera. The acquisition of such a series of photographs is perfectly suited for the presented approach: since the camera pose and 3D structure are calculated from the overlap of single images, best results can be acquired when the site is equally well covered from all directions. However, also very ordered collections of vertical imagery and randomly acquired images over multiple years can also be successfully processed into true orthophotos.

In total, three case studies (Fig. 2) show the potential of this method in a wide variety of topographic settings, but also when dealing with various archaeological marks and imagery (e.g. analogue visible frames, digital near-infrared coverage etc.). Since

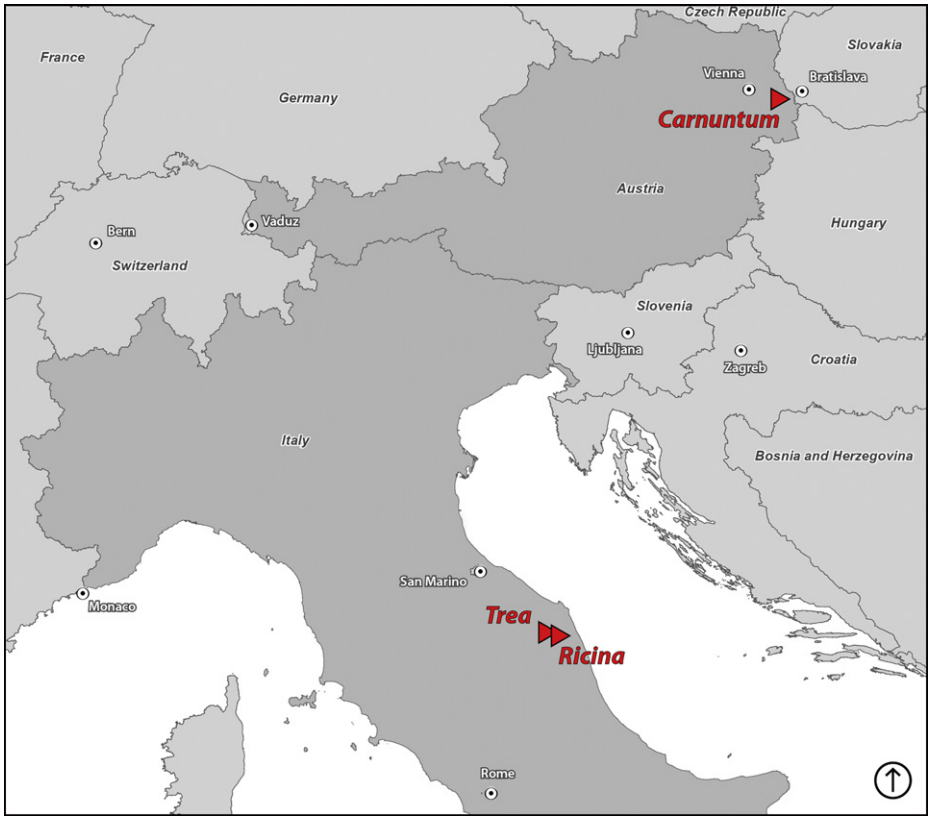

Fig. 2. The location of the three presented case studies: *Carnuntum*, *Ricina* and *Trea*.

all the images were acquired before the method was developed, the true potential of the method can be assessed and its possibilities concerning older imagery showcased. The case studies are presented in a common format: first a short introduction to the site and the acquisition of the photographs is presented; secondly, the building of the orthophoto and possible drawbacks are addressed; and thirdly, each case study will also highlight some very specific advantages of this approach.

4.1. *Carnuntum (Austria)*

In this first case study, a very simple topographic setting is chosen to go overall the individual processing steps. The data set used here consists of sixteen images of the central part of the Roman town of *Carnuntum*, located approximately 40 km south-east of Vienna on the southern bank of the Danube river (N 48°6′41″, E 16°51′57″ – WGS84; Fig. 2). The Roman legionary camp with attached *canabae* and a civil town covers some 650 ha, thus constituting the "largest archaeological landscape in Austria" (Jobst, 1983; Vorbeck and Beckel, 1973). As the Roman capital of the province Pannonia, *Carnuntum* was an important town during the first four centuries AD. Since aerial archaeology and geophysical archaeological prospection have proven to be ideally suited survey methods for the mapping and documentation of such large Roman city sites, those methods have been intensively applied at *Carnuntum* during the last fifteen years (Doneus and Neubauer, 2005; Doneus et al., 2001; Neubauer and Eder-Hinterleitner, 1997; Neubauer et al., 1999, 2002).

The aerial imagery used in this reconstruction was captured at the end of May 2000 around 11.00 h with a Hasselblad 205 FCC medium-format camera and an 80 mm lens from a Cessna 172 Skyhawk. For storage and data processing purposes, the 6 × 6 frames were digitised with Nikon®'s Super COOLSCAN 9000 ED, yielding images of circa 7000 × 7000 pixels. In Fig. 3A, three of those images are displayed. After uploading all sixteen views in PhotoScan, image feature points are detected in all the source images (Fig. 3A-3). The approach is similar to the well-known SIFT (Scale Invariant Feature Transform) algorithm developed by David Lowe (Lowe, 2004), since the features are also stable under viewpoint, scale and lighting variations. Based on its local neighbourhood, each point has its own local descriptor which is subsequently used to detect point correspondences across the complete image set. To do this, robust methods such as a modified version of RANSAC (Fischler and Bolles, 1981) are used. These algorithms can deal with wrong matches, which would otherwise seriously contaminate the retrieved geometric relationships between the images (the latter relations are called the multi-view constraints – Pollefeys et al., 2004). Fig. 3A-2 shows that it is also possible to mask certain image parts that should not be considered during any of the reconstruction steps. Using the correspondences between the image features as input, the locations of those feature points can be estimated and rendered as a sparse 3D point cloud. During the latter process – which is the aforementioned SfM approach – the camera intrinsic and extrinsic parameters are also computed. PhotoScan uses an algorithm to initially compute approximate camera locations, which are afterwards refined by a bundle-adjustment algorithm (pers. comm. PhotoScan support). At this stage, PhotoScan will produce three datasets: (i) a point cloud of typically a few thousand 3D points representing the geometry/structure of the scene (Fig. 3B); (ii) the relative camera poses at the moment of image acquisition (i.e. the position of the projection centre and the camera's orientation – Fig. 3B); (iii) the interior orientation parameters, being the focal length f, the principal point location (c_x, c_y), a skew factor s as well as three radial (k_1, k_2, k_3) and

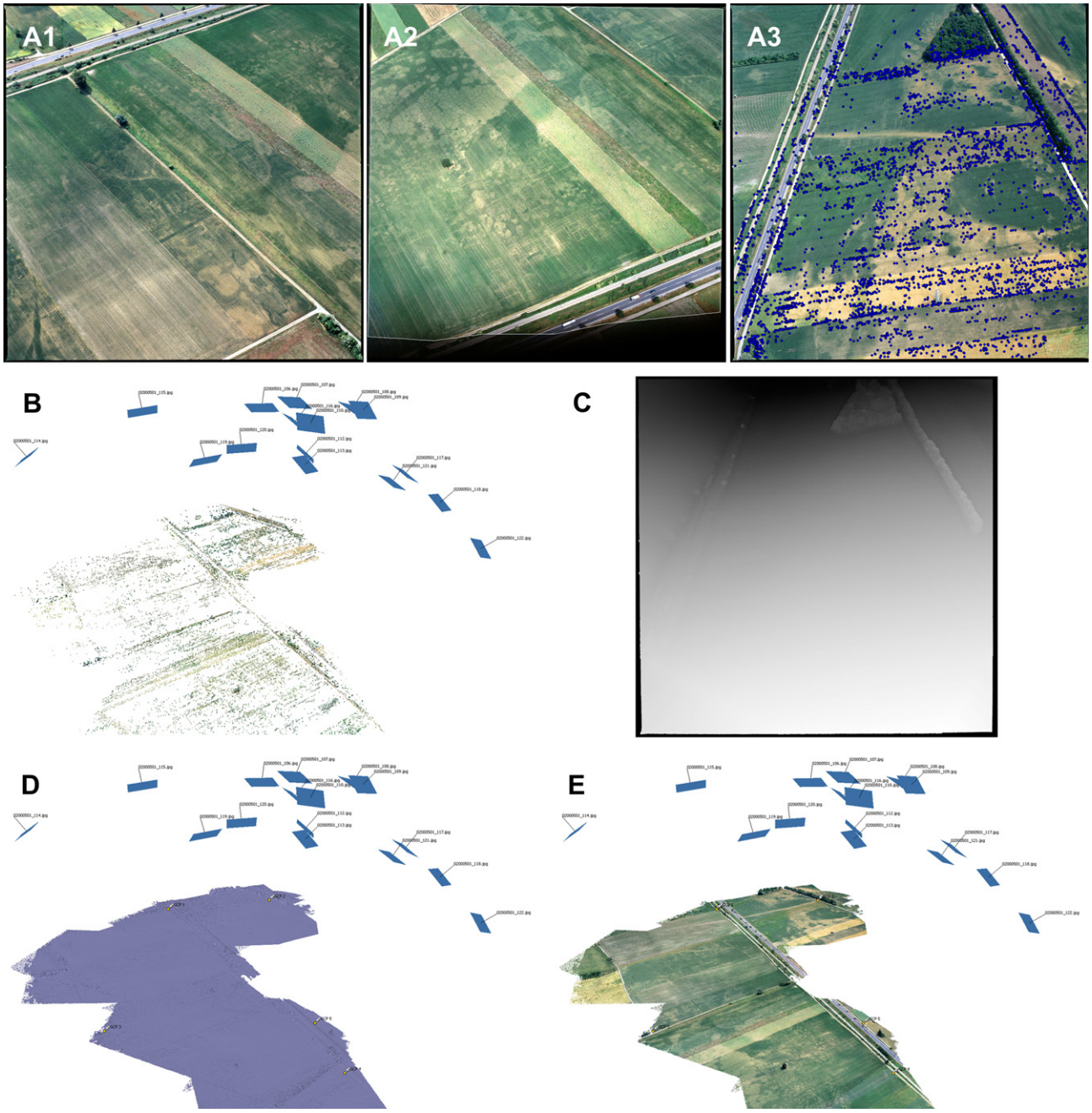

Fig. 3. (A1–3) Three of the oblique images used in a 3D reconstruction of the scene (A2 indicate a masked area and A3 shows the detected features of this photo). B shows the sparse point cloud and the camera positions while C depicts a depth map. D displays the 3D model with GCP markers, while E is a textured version of D.

two tangential distortion coefficients (p_1 and p_2). The first three parameters (f, c_x, c_y) are expressed in pixels (Agisoft LLC, 2011).

In a second step, a dense surface reconstruction is computed. Instead of interpolating the sparse 3D point cloud, a dense stereo matching is executed. Whereas SfM algorithms operate on a sparse set of feature points extracted from the source photographs, these dense reconstruction algorithms operate on the individual pixel values. As all pixels are utilized, this reconstruction step (which is based on a pair-wise depth map computation – Fig. 3C) enables proper handling of fine details present in the scenes and represents them as a 3D mesh (Fig. 3D). Several algorithms are available to do this, but the height-field method has previously been proven to be very suited for aerial images (see Verhoeven, 2011b).

In a third stage, the absolute orientation of the model can be derived by the manual measurement of GCPs in the photographs (or the model – Fig. 3D) or by automatic extraction of the geographical coordinates embedded in the Exif (EXchangeable Image File)-defined metadata tags of the images. Although the latter approach can be ideal for a very fast and straightforward geolocation of the 3D model, it is (currently) not advised considering the accuracy of most GNSS (Global Navigation Satellite System) receivers. Using one of the two approaches, PhotoScan will use a seven parameter similarity transformation to execute the georeferencing. At this stage, both an orthophoto and a DSM can be exported. If necessary, the initial imagery can be blended with the 3D model to form a texture atlas for the 3D scene (Fig. 3E). It needs

to be stressed that this texturing is not necessary for orthophoto production, although it might aid in a location of the positions of the GCP markers. In the latter case, the texturing needs to be applied directly after the dense estimation of the surface geometry.

Even though the terrain is quite homogeneous and flat, this first case study already proves some of the potential of the aforementioned algorithms for aerial archaeology. First of all, many photographs look quite dissimilar due to the Bidirectional Reflectance Distribution Function (BRDF) effects of vegetation in the visible spectrum (Guyot, 1990; Sandmeier and Itten, 1999). Despite the fact that archaeological features and meadows look brighter or fainter according to a specific viewing and illumination geometry, all images were accurately matched. Secondly, this alignment has been executed without previous knowledge on the location of image acquisition, nor about the instrument the imagery was acquired with. Although the camera and lens parameters are known in this specific case, they were not stored in the metadata after scanning the negatives. As a consequence, there were no initial values for the SfM algorithms to take into account. Although it is obvious that information on focal length can cure certain alignment problems, this case proves that the generation of 3D models and orthophotos can succeed even in the absence of these parameters.

4.2. *Ricina (Italy)*

For the next two case studies, the focus is shifted towards Italy. In January 2000, Ghent University initiated the geoarchaeological Potenza Valley Survey (PVS) project in the central Adriatic Region Marche. This interdisciplinary project has mainly aimed at reconstructing the changing physical and human landscape along the Potenza river, one of Marche's major rivers. Aerial archaeological reconnaissance was from the start identified as one of the main survey techniques (Vermeulen, 2002, 2004). In the beginning of June 2009, a series of oblique aerial photographs were acquired above the lower valley Roman town of *Ricina* (N 43°19′41″, E 13°25′26″ – WGS84; Fig. 2) with an uncalibrated compact digital camera (Panasonic DMC-FZ18), yielding photographs of eight megapixels. The total series consisted of 33 images, of which six were withheld for reasons of coverage and lack of sharpness. All photographs depict positive and negative crop marks related to this imperial Roman town, of which only a rather well-preserved theatre building is fully visible above ground level today (Capodoglio and Cipoletta, 1996). Although there have been a series of investigations to understand the character and extent of this city, almost nothing was known about its general layout and organization (Alfieri, 1937; Mercando, 1971; Percossi Serenelli, 1989; Cecchi, 1968) before aerial photographs were taken in the PVS framework (Vermeulen and Verhoeven, 2004, 2006). The specific image series under study here proved again the immense value of archaeological aerial photography, since it became for the first time possible to spot the hidden remains of a Roman amphitheatre, visible just above the indicated area 1 in Fig. 4 (Vermeulen, 2011).

Using four GCPs whose coordinates were extracted from a 1: 10 000 topographical map, the georeferencing of the model was performed with a final horizontal Root-Mean-Square Error (RMSE) of 0.31 m and a vertical RMSE of 0.15 m. To compute the orthophoto of the image, a grid space of 10 cm was used (which is the value that was also proposed by PhotoScan). After applying some additional image enhancement to make the crop marks stand out more clearly, the final orthophotomap nicely displays all archaeologically relevant features in their accurate position (Fig. 4).

The total processing time was about 38 min. After adding 5 min for image selection, about 25 min for GCP marking plus tweaking and a few additional minutes for contrast enhancement, the total orthophoto production time for these 27 images reached

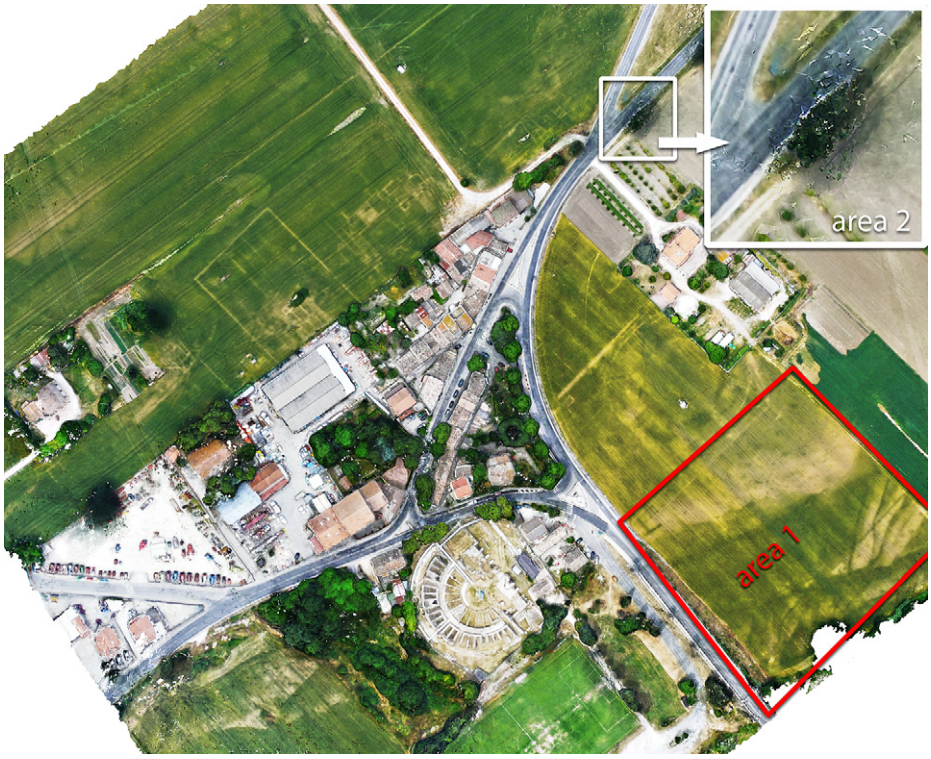

Fig. 4. The *Ricina* orthophoto calculated out of 27 oblique images and enhanced for crop mark display. The area indicated in red is a down sloping zone which was previously impossible to accurately georeference due to the lack of GCPs and a detailed elevation model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

approximately 70 min. Compared to conventional georeferencing on an image-by-image basis, this shows the huge saving in total processing time with the additional benefit of having a complete orthorectified overview image with a positional accuracy that might be hard to attain using the conventional archaeologically-dedicated low-cost packages.

Furthermore, it becomes apparent that imagery can be used which would have been considered "unusable" before (see Palmer, 2005 for some consideration on this issue). Some images of this and the previous dataset only record three or ill-distributed GCPs, neither of which are considered adequate for most georeferencing approaches. Furthermore, photographs that suffer from quite some topographic image displacement cannot fully be corrected using simple rectification algorithms such as a projective or polynomial transformation. Such a problematic area is indicated in Fig. 4 (area 1). Although the borders of this field (respectively a road and another field) are on the topographical map, no single point along these lines can be used as a possible control point. Besides this lack of potential GCPs, this south-east town area is slightly undulating and sloping downwards. Since the presented approach uses a detailed DSM for orthophoto production, the output is a true orthophoto in which all possible tilt and terrain displacements are taken into account. The fact that this orthopohoto procedure integrates all imagery into one photomosaic, also makes the search for suitable ground control much easier (hence solving all the issues mentioned).

Finally, it needs to be stressed that the generated orthophoto might still suffer from some artefacts. Area 2 of Fig. 4 displays an area which suffers from over-triangulation, most likely due to the image noise produced by the compact camera and insufficient pixel data to compute this part of the DSM. However, apart from being a visual artefact, such defects do in most cases not alter the general appraisal and global positional accuracy of the computed orthophoto. Obviously, such artefacts are more likely with reconstructions based on imagery from sub-optimal optical quality (such as those from smaller cameras or blurry images), but they can always be fixed using third-party software.

4.3. Trea (Italy)

The general advised strategy in using PhotoScan is to solve the complex SfM math of as large as possible a set of images, without having to rely on virtual memory. Later, one can 'disable photos' and perform the dense reconstruction in parts (Verhoeven, 2011b). Although this approach was meant to tackle limited hardware resources, it opens up a completely new application field for aerial archaeologists. To illustrate this, a time series covering six years of aerial PVS research on the Roman town of *Trea* (central Adriatic Italy, 43°18′40″ N, 13°18′42″ W – WGS84; Fig. 2) will be used. In contrast to the former case studies, the Roman city of *Trea* lies on a hill, surrounded by a seriously undulating landscape of the middle Potenza valley. The scene can thus be considered quite complex and the terrain displacement in the aerial images very substantial. Like *Ricina*, the city layout of *Trea* was only known summarily (see Moscatelli, 1985) before this systematic aerial photography campaign. The latter now allows a near complete mapping of the main urban structures of this abandoned Roman city, such as the town defences, the internal street network and the main public and private buildings.

From the 208 initially selected images, 203 were aligned correctly (Fig. 5A). This number is extremely high given the circumstances: a wide variety of cameras and lenses were used during the reconnaissance flights; the land cover varied from bare soil to crops in various phenological states; 39 images only

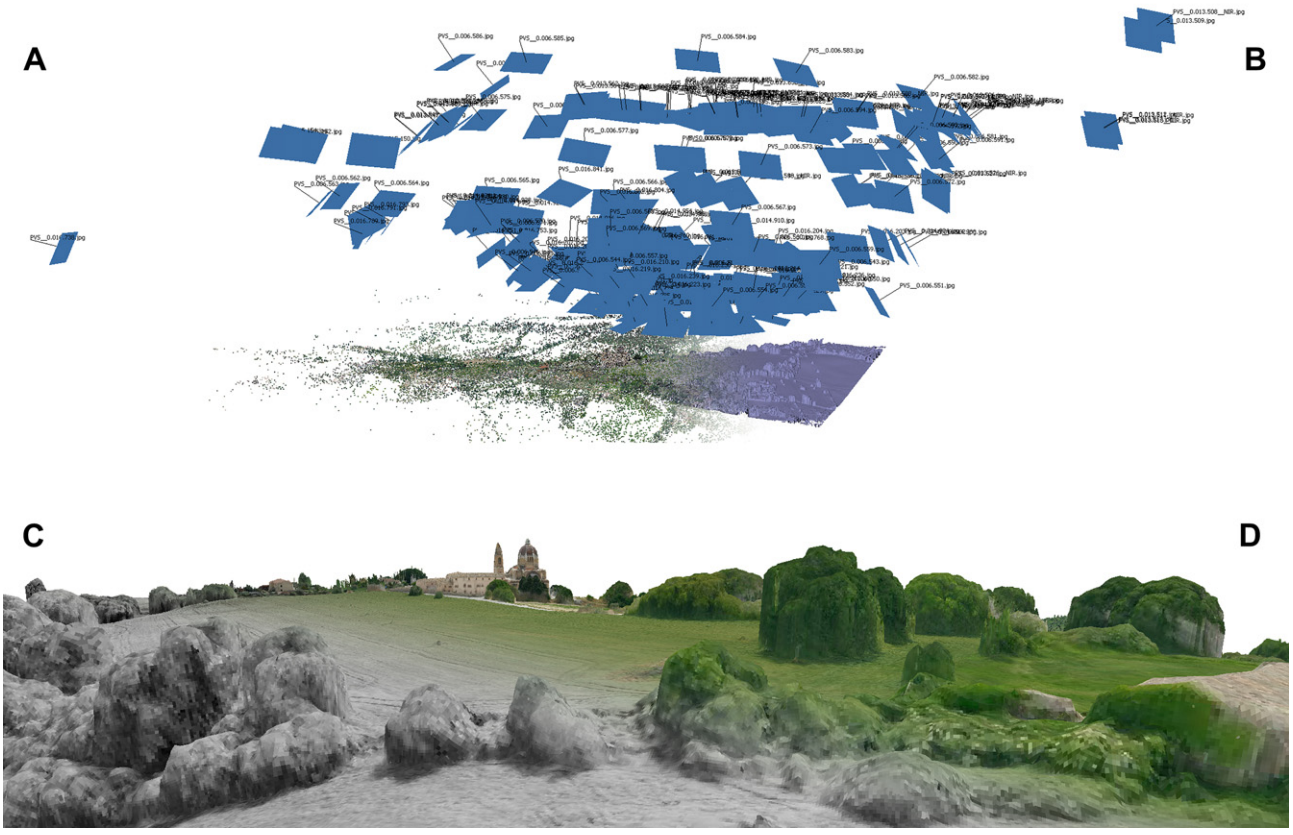

Fig. 5. (A) The relative position of all 203 camera stations and the extracted DSM of *Trea* (B); (C) shows a Near-Infrared and (D) a visible rendering of the DSM.

recorded the reflectance in the Near-InfraRed (NIR) spectral band (see Verhoeven, 2008, 2011a; Verhoeven et al., 2009 for details on this). Unquestionably, this alignment result was facilitated by the fact that all images still had information on the focal length embedded in the Exif metadata tags, so that these values could be used for initializing the SfM step. To execute the dense reconstruction stage, only a subset of images was used. Since all images have been aligned in Euclidian space in the SfM step, one can choose whatever image that is deemed useful for the reconstruction. Since too much pixel data increases computation time and does not necessarily enhance the accuracy and detail of the DSM, only a selection of 143 suitable images – largely based on image scale, scene coverage and sharpness – was used as input for the modelling step. In total, alignment and DSM processing took 4.5 h and could easily be run over night. Once an accurate 3D model of the terrain is generated (Fig. 5B), every image or combination of images can be applied in a texture mapping or orthophoto generation. The former can for example be interesting to create a new view on the scene. As an example, Fig. 5C and D respectively show a NIR and visible rendering of the DSM from a viewpoint that is very different from all the original aerial views.

Because the orthophoto generation is characterised by the same freedom in image selection, it is possible to use only the NIR images (Fig. 6A-3) or those that best illustrate the crop marks (Fig. 6A-1) or soil mark state (Fig. 6A-2). Not only does this approach speed up the processing of individual images (or related photo sets) considerably, but the final interpretation is more trustworthy as well. Due to the heavy undulating nature of the terrain and the very steep slopes bordering the central plateau, the archaeologically-dedicated tools (such as AERIAL or Airphoto) andmost GIS packages will typically fail to accurately georeference these images. Although this might not seem to be a big issue when dealing with vague soil marks, the nature of the crop marks (faint and small) as well as the type of site (a complex Roman town with different phases) makes the accurate mapping of the features of the utmost importance for comparison of aerial footage from different years or interpret the data with respect to a geophysical survey (for this case study, the georeferencing delivered a planar RMSE of 6.2 cm and an RMSE of 4.6 cm for the altitude component). Additionally, the whole process of orthophoto production is straightforward, fast and can deal with a variety of frame imaging sensors of which no calibration parameters need to be supplied. Besides, Fig. 6A-4 indicates that even individual images without any GCP can be transformed into orthophotos. The combination of these advantages largely overcomes the current drawbacks that archaeologists encounter in most image (ortho)rectification approaches, certainly when dealing with larger areas (features of a palaeolandscape, extensive sites) or terrain undulation.

However, it must be stressed that processing this amount of imagery requires significant computing power and DSM and orthophoto artefacts can occur. It should also be noted that such an integrated approach only works when no major scene changes have taken place during the years of image acquisition. In the presented case study, the biggest surface difference was related to the phenological state of the vegetation: sometimes the fields were just harvested, while on other moments the camera recorded the full canopy. Although it did not hamper the image alignment, the DSM will obviously be influenced by this. Therefore, one can best use a set of images displaying the most common surface condition, after which a numerical form of the latter can be used to compute the orthophotos of more or less all images. This approach was used in this case study and did not result in archaeologically relevant positional differences of the computed orthophotos. In case the difference between different topographical conditions is too big, a multitude of DSMs should be computed to cover all possible surface states. In the worst case scenario, the landscape has changed so drastically that the image alignment fails.

Although it is not the aim of this paper to provide a rigorous assessment of the positional accuracy of orthophotos (despite the RMSE values given so far), Fig. 6B give some visual clues about the horizontal accuracy of the orthophotos. In this illustration, total station measurements of the road and some building features were placed on top of the orthophotograph. This result clearly illustrates that the generated PhotoScan output is more than sufficient for

Fig. 6. (A) Integration of several orthophotos, showing crop marks (1) and soil marks (2) in the visible domain, the NIR terrain reflectance (3) and the orthorectification of an image (4) without any useable GCP; (B) shows the comparison between the generated orthophoto and total station measurements of the road and other building structures.

large-scale (e.g. 1:500) archaeological mapping from aerial imagery. For a more elaborate accuracy analysis of SfM-based orthophotos computed out of randomly acquired aerial imagery, consider Verhoeven et al. (2012). Additionally, research by Doneus et al. (2011) proved how well this method holds up when compared with terrestrial laser scanning in an excavation context.

5. Considerations and improvements

Aerial photography provides a basis for gathering spatial data. Before archaeological information can be extracted from these data in a way that is useful for mapping and further analysis, the aerial images must be georeferenced in an absolute manner. This process, which aims at placing each image pixel on its true location on the Earth's surface, should also try to take care of all geometric transformations that occur during the imaging process. By applying the approach presented in this article, simplicity is combined with geometrical quality. Besides the automatic estimation of the inner camera calibration parameters, a dense DSM is calculated, which is used in a final phase to generate true orthophotos. As a result, this method largely accounts for most relevant kinds of geometrical degradations and is capable of generating 3D models and orthophotos that are perfectly suited for archaeological purposes. This option of fast and accurate orthophoto production is very welcome for aerial archaeologists, given their current approaches which are not tailored to deal either with aerial frame imagery lacking sufficient ground control or with large amounts of photographs from different cameras shot in different seasons. This method offers the enormous advantage that there are just standard photographic recording prerequisites. Apart from a sufficient number of sharp images covering the scene to be reconstructed and at least three GCPs to georeference the reconstruction in respect to a common coordinate frame, no other information is needed. Besides, only a minimal technical knowledge and user interaction are required. Finally, this approach can also work in the total absence of any information about the instrument the imagery was acquired with, although the Exif-defined metadata tags, including even GPS coordinates, can be utilized. So, it really is obvious that this methodology can be integrated seamlessly into any existing aerial archaeological workflow. The extra investments needed for software and computing hardware are recovered easily when taking the time and cost savings of map production into account.

Although the presented case studies prove this semi-automatic, integrated approach to be a valid alternative to the current variety of (ortho)rectification methods, it has to be stressed that this methodology is not perfect. First of all, the method is not applicable for the individual image. At least two – but preferably more images – are needed for accurate DSM computation. In addition, erroneous alignment of the imagery can occur when dealing with very large photo collections, images that suffer from excessive noise or blur, highly oblique photographs or photographs that have a very dissimilar appearance (e.g. due to major underexposure or changing topographic terrain parameters). Currently, PhotoScan misses tools that allow straightforward and solid inspection of the quality and accuracy of this alignment step (although a reprojection error is provided in later PhotoScan versions), so the user can only visually assess any alignment problems. These can always be tackled by disabling the images that cause the error or inserting accurate camera calibration parameters before the SfM step. The latter can also prevent the reconstructed surface from being too curved, since such distortion is often related to a poor estimation of the camera intrinsics when dealing with large image sequences. Because absolute georeferencing in the current approach occurs after building the 3D model, any previous distortion will also remain after assigning absolute spatial coordinates to the DSM. Apart from the initial introduction of previously determined internal camera calibration parameters, a possible solution to this problem involves excessive coverage of the area using a larger overlap between the photos. However, it should be clear that high-quality reconstructions with many digital frames are extremely taxing for the computing hardware. Hence, computer components that are a few years old will most likely not be up to this task. Since PhotoScan and many similar applications support the OpenCL (Open Computing Language) programming platform, the graphical-card is one of the most important hardware components for speeding up the reconstructions. Still, better and more optimised algorithms are needed to permit the time-efficient processing of large image sets on standard computers. For instance, it would be interesting to fix the alignment of the initially used images. Currently, registering new photographs to an image set means that both the SfM and dense reconstruction steps have to be executed again. Being able to just align new imagery based on the previously calculated sparse 3D point cloud, would be a huge timesaver and enable on-the-fly production of orthophotos based on newly shot aerial imagery.

Another, rather hardware-based, improvement would certainly be the possibility to merge existing exterior camera parameters with the computed camera extrinsics, as was already performed by Wischounig-Strucl et al. (2011). Although PhotoScan already supports the direct import of GPS coordinates, the positional accuracy of current handheld GNSS devices is too low for very accurate positioning using only the raw latitude, longitude and altitude metadata. In case the relative camera extrinsics resulting from the SfM stage could be linked to absolute position and orientation data acquired by a differential GPS and IMU positioning system accompanying the digital camera, an iterative approach using the SfM data could be used to improve the final accuracy of the camera pose parameters. Since the whole 3D model is expressed in relative Euclidian coordinates, this matching would directly georeference the computed model with high accuracy, while the optimized absolute camera exterior parameters replace the first raw estimation and are subsequently stored in the metadata of the imagery. Moreover, if a few absolute camera extrinsics are added as constraints in the SfM process, it could prevent possible drift in the recovered camera and point locations (Snavely et al., 2006). For those cases where no accurate camera extrinsics are available, the georeferencing of the images is still determined with time-consuming GCPs identification.

6. Conclusion and outlook

Straightforward orthophoto production is very important in the discipline of aerial archaeology. In this article, computer vision algorithms (Structure from Motion and Multi-View Stereo) complemented by proven photogrammetrical principles (such as bundle adjustments) were exploited to present an integrated, low-cost, semi-automated orthophoto production of archaeological aerial (uncalibrated) frame images. This approach is straightforward and requires no assumptions on the camera projection matrix, existing photogrammetric and computer vision knowledge or the topography of the scenes. One just needs to make sure that enough overlapping aerial images are acquired. Even though this might involve flying one or more orbits of the scene of interest, this method will afterwards prove itself in terms of orthophoto quality and – in most occasions – processing speed, certainly when a larger area must be mapped or uneven terrain is involved. Furthermore, the case studies proved that the orthophoto output is more than sufficiently accurate for archaeological large-scale photo mapping, while the results are visually appealing as well.

Finally, drawbacks were evaluated and possible improvements proposed. First of all, it was indicated that the processing is very computer resource intensive, while image alignments are not completely optimal and solid accuracy measures still lacking. Also, the approach is presently semi-automatic and automation only makes sense when it seriously reduces or completely eliminates steps in a process. In the case of archaeological orthophoto generation, these are the recurring steps of visualizing the images, selecting the essential geodata (GCPs), and setting all the parameters for the subsequent execution of the algorithms. Since this is currently considered to be the bottleneck in large-scale archaeological projects with thousands of images, a project has been launched in 2012 by the first three authors (funded by the Austrian Science fund, P 24116-N23). It aims at the creation of completely automatic solutions for the GCP selection of archaeological aerial photographs. This would offer possibilities for the consistent creation and fast updating of archaeologically relevant cartographic data in our rapidly changing landscapes.

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), Airborne Technologies GmbH (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). The case studies from the Potenza Valley Survey project were made possible thanks to support from Belgian Science Policy (Interuniversity Attraction Poles, project P6/22).

References

Agisoft LLC, 2011. AgiSoft PhotoScan User Manual. Professional Edition, Version 0.8.0 (accessed 03.05.2011.). http://www.agisoft.ru/pscan/help/en/pscan_pro.pdf.

Alfieri, N., 1937. Ricina. In: Atti e Memorie della Deputazione di Storia Patria per le Marche, 5, pp. 21–37.

Bewley, R.H., Rączkowski, W., 2002. Past achievements and prospects for the future development of aerial archaeology: an introduction. In: Bewley, R.H., Rączkowski, W. (Eds.), Aerial Archaeology: Developing Future Practice. IOS Press, Amsterdam, pp. 1–8 (NATO Science Series. Series I: Life and Behavioural Sciences Vol. 337).

Bradley, D., Boubekeur, T., Heidrich, W., 2008. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In: CVPR 2008. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Anchorage, AK, pp. 1–8.

Braun, J., 2003. Aspects on true-orthophoto production. In: Fritsch, D. (Ed.), Photogrammetric Week'03. Wichmann Verlag, Heidelberg, pp. 205–214.

Brophy, K., Cowley, D., 2005. From the Air. Understanding Aerial Archaeology. Stroud, Tempus.

Capodoglio, G., Cipoletta, F., 1996. Ricina. Il Teatro, la città. S. Giuseppe srl, Pollenza.

Cecchi, D., 1968. Helvia Recina e il Piceno nell'età romana. Studi Maceratesi 4, 126–214.

Crawford, O., 1924. Air Survey and Archaeology. Ordnance Survey, Southampton.

Doneus, M., 2001. Precision mapping and interpretation of oblique aerial photographs. Archaeol. Prospect 8, 13–27.

Doneus, M., Neubauer, W., 1998. 2D combination of prospection data. Archaeol. Prospect 5, 29–56.

Doneus, M., Neubauer, W., 2005. Multiple survey techniques at Roman Carnuntum. Integrated prospection of the largest archaeological landscape in Austria. In: Musson, C., Palmer, R., Campana, S. (Eds.), In volo nel passato. Aerofotografia e cartografia archaeological. Biblioteca del dipartimento di archaeologia e storia delle arti – sezione archaeologica, Università di Siena, Sienna, pp. 272–279.

Doneus, M., Eder-Hinterleitner, A., Neubauer, W., 2001. Roman Carnuntum: prospecting the largest archaeological landscape in Austria. In: Doneus, M., Eder-Hinterleitner, A., Neubauer, W. (Eds.), Archaeological Prospection: Fourth International Conference on Archaeological Prospection. Austrian Academy of Sciences, Vienna, pp. 47–59. Vienna, 19–23 September 2001.

Doneus, M., Verhoeven, G., Fera, M., Briese, Ch., Kucera, M., Neubauer, W., 2011. From deposit to point cloud – A study of low-cost computer vision approaches for the straightforward documentation of archaeological excavations (XXIIIrd International CIPA Symposium). Geoinformatics 6, 81–88.

Fischler, M., Bolles, R., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. ACM 24, 381–395.

Fisher, R.B., Dawson-Howe, K., Fitzgibbon, A., Robertson, C., Trucco, E., 2005. Dictionary of Computer Vision and Image Processing. Wiley, Chichester.

Guyot, G., 1990. Optical properties of vegetation canopies. In: Steven, M.D., Clark, J.A. (Eds.), Applications of Remote Sensing in Agriculture. Butterworths, London, pp. 19–43.

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, Belgium, Ghent, 201–212 pp.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision, second ed. Cambridge University Press, Cambridge.

Jobst, W., 1983. Provinzhauptstadt Carnuntum. Österreichs größte archäologische Landschaft. Wiener Verlag, Himberg.

Kraus, K., 2002. Zur Orthophoto-Terminologie. Photogrammetrie. Fernerkundung. Geoinformation 6, 451–452.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.

Mellor, J.P., Teller, S., Lozano-Pérez, T., 1996. Dense Depth Maps from Epipolar Images. Memo 1593, Artificial Intelligence Laboratory, MIT.

Mercando, L., 1971. Villa Potenza (Macerata). – Rinvenimento di edificio romano con pavimento a mosaico. In: Atti della Accademia Nazionale dei Lincei. Notizie degli scavi di antichitá Serie 8, 25, pp. 381–401.

Moscatelli, U., 1985. Municipi romani della V regio Augustea: problemi storici ed urbanistici del Piceno centro-settentrionale (III – I sec. a.C.). PICUS. Studi e ricerche sulle Marche nell'antichità 5, 51–97.

Neubauer, W., Eder-Hinterleitner, A., 1997. Resistivity and magnetics of the Roman town Carnuntum, Austria: an example of combined interpretation of prospection data. Archaeol. Prospect 4, 179–189.

Neubauer, W., Eder-Hinterleitner, A., Seren, S., Doneus, M., Melichar, P., 1999. Kombination archäologisch-geophysikalischer Prospektionsmethoden am Beispiel der römischen Zivilstadt Carnuntum. Archaeologia Austriaca 82/83, 493–551.

Neubauer, W., Eder-Hinterleitner, A., Seren, S., Melichar, P., 2002. Georadar in the Roman civil town Carnuntum, Austria: an approach for archaeological interpretation of GPR data. Archaeol. Prospect 9, 135–156.

Palmer, R., 2005. If they used their own photographs they wouldn't take them like that. In: Brophy, K., Cowley, D. (Eds.), From the Air. Understanding Aerial Archaeology. Stroud, Tempus, pp. 94–116.

Percossi Serenelli, E., 1989. Rinvenimenti ed emergenze archeologiche nel territorio dell'antica Ricina. PICUS. Studi e ricerche sulle Marche nell'antichità 9, 65–117.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R., 2004. Visual modeling with a hand-Held camera. Int. J. Comput. Vis. 59, 207–232.

Quan, L., 2010. Image-based Modeling. Springer, New York.

Robertson, D.P., Cipolla, R., 2009. Structure from motion. In: Varga, M. (Ed.), Practical Image Processing and Computer Vision. John Wiley and Sons Ltd, New York.

Sandmeier, S.R., Itten, K.I., 1999. A field Goniometer system (FIGOS) for acquisition of hyperspectral BRDF data. IEEE Trans. Geosci. Remote Sensing 37, 978–986.

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

Scollar, I., 2002. Making Things look vertical. In: Bewley, R., Raczkowski, W. (Eds.), Aerial Archaeology: Developing Future Practice. IOS Press, Amsterdam, pp. 166–172 (= NATO science series Series 1: Life and behavioural sciences).

Scollar, I., Tabbagh, A., Hesse, A., Herzog, I., 1990. Archaeological Prospecting and Remote Sensing. Cambridge University Press, Cambridge.

Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., 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 (CVPR'06), vol. 1. IEEE, Washington, pp. 519–528.

Snavely, N., 2010. Bundler: structure from motion for unordered image collections (accessed 28.10.10.). http://phototour.cs.washington.edu/bundler/.

Snavely, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. ACM Trans. Graphics 25, 835–846.

Szeliski, R., 2011. Computer Vision. Algorithms and Applications. Springer, New York.

Ullman, S., 1979. The interpretation of structure from motion. Proc. Royal Soc. London B 203, 405–426.

Verhoeven, G., 2008. Imaging the invisible using modified digital still cameras for straightforward and low-cost archaeological near-infrared photography. J. Archaeological Sci. 35, 3087–3100.

Verhoeven, G., 2011a. Near-Infrared aerial crop mark archaeology: from its historical use to current digital implementations. J. Archaeol. Method Theory.

Verhoeven, G., 2011b. Taking computer vision aloft – archaeological three-dimensional reconstructions from aerial photographs with PhotoScan. Archaeol. Prospect 18, 67–73.

Verhoeven, G., Smet, P., Poelman, D., Vermeulen, F., 2009. Spectral characterization of a digital still camera's NIR modification to enhance archaeological observation. IEEE Trans. Geosci. Remote Sensing 47, 3456–3468.

Verhoeven, G., Taelman, D., Vermeulen, F., 2012. Computer vision-based orthophoto mapping of complex archaeological sites: the ancient quarry of Pitaranha (Portugal-Spain). Archaeometry. doi:10.1111/j.1475-4754.2012.00667.x.

Vermeulen, F., 2002. The Potenza valley survey (Marche). In: Attema, P.A.J., Burgers, G.-J., van Joolen, E., van Leusen, M., Mater, B. (Eds.), New Developments in Italian Landscape Archaeology. Theory and Methodology of Field Survey. Land Evaluation and Landscape Perception. Pottery Production and Distribution. In: Proceedings of a Three-Day Conference Held at the University of Groningen. Archaeopress, Oxford, pp. 104–106.

Vermeulen, F., 2004. Fotografi a aerea fi nalizzata nelle Marche centrali: un progetto integrato. Archeologia Aerea. Studi di Aerotopografia Archeologica 1, 91–118.

Vermeulen, F., 2011. Reviewing 10 years of aerial photography in the valley of the River Potenza (Le Marche). Archeologia Aerea. Studi di Aerotopografia Archeologica 4, 259–264.

Vermeulen, F., Verhoeven, G., 2004. The Contribution of aerial photography and field survey to the study of urbanization in the Potenza valley (Picenum). J. Roman Archaeol., 57–82.

Vermeulen, F., Verhoeven, G., 2006. An integrated survey of Roman urbanization at *Potentia*, central Italy. J. Field Archaeol. 31, 395–410.

Vorbeck, E., Beckel, L., 1973. Carnuntum – Rom an der Donau. Otto Müller Verlag, Salzburg.

Wilson, D., 2000. Air Photo Interpretation for Archaeologists. Tempus, Stroud.

Wischounig-Strucl, D., Quartisch, M., Rinner, B., 2011. A structure based mosaicking approach for aerial images from low altitude of non-planar scenes. In: Wendel, A., Sternig, S., Godec, M. (Eds.), 16th Computer Vision Winter Workshop.

Wolf, P.R., Dewitt, B.A., 2000. Elements of Photogrammetry with Applications in GIS, third ed. McGraw-Hill, Boston.