Quality Analysis of Structure from Motion in Archaeological Airborne Applications

Cornelis Stal\textsuperscript{1}, Jean Bourgeois\textsuperscript{2}, Philippe De Maeyer\textsuperscript{1}, Guy De Mulder\textsuperscript{2}, Alain De Wulf\textsuperscript{1}, Rudi Goossens\textsuperscript{1}, Marijn Hendrickx\textsuperscript{1}, Timothy Nuttens\textsuperscript{1} and Birger Stichelbaut\textsuperscript{2}

\textsuperscript{1}Ghent University, Department of Geography, Ghent, Belgium; \{Cornelis.Stal, Philippe.DeMaeyer, Alain.DeWulf, Rudi.Goossens, Marijn.Hendrickx, Timothy.Nuttens\}@UGent.be

\textsuperscript{2}Ghent University, Department of Archaeology, Ghent, Belgium; \{Jean.Bourgeois, Guy.DeMulder, Birger.Stichelbaut\}@UGent.be

Abstract. For a few years, structure from motion has been commercially applied and interesting 3D models are discussed in recent literature. Structure from motion is known to be a relatively fast and low-cost technique for the generation of photo-realistic 3D models. This technique uses a series of digital images of an object or site, taken from different positions. In this image series, each part of the object is recorded on at least three images. Combining these images with information from the metadata file allows to determine the different camera positions and orientations and to calculate a textured mesh or colored point set in 3D. These calculations are based on solving a system of geometric matrices and a least squares 2D to 3D projective problem.

In this article, the geometric quality of an airborne digital surface model, based on structure from motion, is assessed, in order to use this model for archaeological research. An Iron Age hillfort site in the municipality of Kooigem (Belgium) is used as test case for this research. In this area, some interesting archaeological relicts, such as walls and ditches, are known and documented. These objects have a sub-meter height difference with the surrounding topography and are therefore very useful for visual validation and geometrical analysis of the digital elevation models. Although this fast and low-cost method results in a fair visual representation of the environment of the relicts, a special focus is needed on less manifest objects of interest in the landscape. Besides, more research on the geometrical accuracy of the final products will be discussed. A digital elevation model, acquired by airborne laser scanning, with an average point density of 4 points per square meter, is used for the statistical quality assessment of the elevation model generated by structure from motion. This statistical and visual analysis of the digital elevation model generated with structure from motion, allows to evaluate this new technique, its results and its applicability for archaeological research.

Keywords. Structure from motion, DSM, geometric quality, archaeology, airborne laser scanning.
1. Introduction

Spatial 3D data are required by a various range of disciplines, like environmental research [1, 2], city modeling [3, 4] and archaeology. The number of acquisition sensors is numerous and can be divided in active optical sensors (e.g. laser scanners), passive optical sensors (e.g. cameras or satellite sensors) and synthetic aperture RaDAR [5]. With the large number of sensors and accompanying processing techniques, the question arises which sensor and technique to deploy for which application. In archaeology, airborne laser scanning (ALS) [6, 7] and image based modeling [8] are applied for the calculation of virtual 3D reconstructions. ALS results in highly accurate digital surface models (DSMs) and, especially for forested areas, powerful filtering algorithms are developed for bare earth extraction and vegetation elimination [9], enabling the construction of digital terrain models (DTMs). However, the acquisition of these kinds of data is rather expensive in comparison to the price of airborne imagery, especially for small scale archaeological projects. In most cases, the generation of digital models based on imagery is preferred for practical and financial reasons, in particular with the introduction of structure from motion (SfM).

During the last decades, SfM is developed in the field of computer vision [10, 11]. The introduction of this technique in different algorithms and the commercialization of software packages containing these implementations, resulted in the availability of SfM as alternative 3D modeling technique for many research areas. In archaeology, the technique is already applied for the documentation and modeling of excavation sites, using huge amounts of terrestrial images. The results of these projects are good for visual analysis of the study objects but the geometric quality is not assessed intensively. Hence, the statistical comparison of SfM-based models with a reference data set is required in order to use these models for further geometric analysis.

The research presented in this article focuses on the geometric quality of a DSM, derived from a SfM-based 3D model. A 3D model of the environment of the forest of ‘Kooigem’ (in Dutch: ‘Kooigembos’) was generated for this research, using Agisoft PhotoScan software. In and around this forest, some Iron Aged archaeological relicts were previously detected, as discussed in section 2. A series of aerial images with a ground spacing distance (GSD) of 30 cm was provided by the Belgian National Mapping Agency (NGI/IGN) and compared with an ALS data set. In many applications, only regular aerial images of an area of interest are available for digital elevation modeling, instead of more expensive ALS data. The used data sets are presented in section 3. Although the acquisition coordinates of the used images were known, a set of ground control points (GCPs) was measured in the study area, using real time kinematic (RTK) global navigation satellite system (GNSS). The actual 3D reconstruction is discussed in section 4. The geometric quality is assessed by the construction of a random point set in the study area and by analyzing the height values of the SfM-based model and with the corresponding height values of the ALS-based model. Hence, the used analysis procedure is actually based on 2.5D geometry. The results of this statistical analysis, as well as a visual analysis, are presented and discussed in section 5, followed by the final conclusions in section 6.

2. Study area

The study area is located at the top of a hill of the forest of ‘Kooigem’ (in Dutch: ‘Kooigembos’), situated between the villages of Kooigem, Sint-Denijs and Spiere-Helkijn (West Flanders, Belgium, Figure 1). The dominance of the hill results in a good coverage of the river Scheldt valley in the south and the old road between Kortrijk and Tournai in the east [12]. The archaeological finds on the site cover a time period from the Hallstatt/La Tène period to the Gallo-Roman period and concern a large number of shards around a Celtic settlement and an early Roman villa [13]. These finds
are intensively discussed under supervision of the provincial society of archaeological heritage [12, 14].

Most relics are located in the forest itself. The direct neighborhood of this forest consists of intensively used farmland. Using SfM, the detection of relics around the forest in this farmland is therefore unlikely, but not impossible. Hence the motivation to use this area as case study for the generation of high quality SfM-based 3D models and the comparison with ALS data.

![Figure 1: Overview of the study area (source: NGI)](image)

### 3. Data sets

#### 3.1. Airborne images

On April 9, 2011 (around 12 AM), a stereoscopic aerial flight has been performed over the study area by *Hansa Luftbild*, commissioned by the Belgian National Mapping Agency (NGI/IGN). The image series covers the entire southern part of the province of West-Flanders, Belgium. The plane was equipped with a *ZI-Imaging DMC-01* pan-chromatic digital camera and a 120 mm lens. The sensor has a size of 13 824 x 7 680 pixels and a pixel size of 12 micron. The images were acquired with an average flight height of approximately 3 000 m above ground, resulting in a resolution of 0.30 m GSD, or a coverage of approximately 4 150 x 2 300 m. With a longitudinal overlap between 57 % and 63 % and a lateral overlap of approximately 25 %, nine images were required to generate a 3D model of the study area. Each of these images was accompanied by the acquisition coordinates,
measured by differential GNSS during the flight. These measurements were supplemented with RTK GNSS based GCPs, as discussed in the next section.

3.2. Airborne laser scanning

ALS or Airborne LiDAR (Light Detection and Ranging) is a very useful technique for the fast and accurate topographic recording of the terrain and the surface. Many governmental and non-governmental institutions have already used the technique to acquire 2.5D and 3D models in favor of modeling of river overflow areas, city modeling, archaeological modeling or general topographic mapping. The ALS system uses electro-magnetic signals in the near-infrared field. A signal is emitted from an aerial platform and is reflected by an object on the earth’s surface. Using common ALS sensors, the reflection occurs on both ground- and non-ground objects, resulting in a single point per signal. In many cases, where a bare-earth model is required for further analysis, the filtering of these ground points and non-ground points is necessary for the generation of a DTM. The biggest advantages of ALS are the fast and relatively accurate acquisition of point sets, with a wide range of possible point densities. Besides, recent research on processing ALS data enabled a reliable and straightforward workflow for the generation of DTMs and DSMs. However, for relatively small study areas, the deployment of the system is rather expensive.

The areas of the Spiere-Helkijn and the Mount Kemmel, both in the Province of Western Flanders (Belgium) were measured during an ALS campaign in 2008. This campaign was ordered by the deputy of the province of West-Flanders, on behalf of archaeological research by Ghent University [7, 15]. A LiteMapper 5600 was used at a flying height of 400 m above ground, with an average point density of 4.5 points per m². All points are labeled as either ground point or non-ground point and are interpolated by inverse distance weighting (IDW) interpolator, respectively resulting in a DTM and a DSM with a resolution of 0.50 m.

4. 3D model generation

4.1. Data processing

Conventional digital photogrammetric software packages rely on calibrated images and will only process single stereo couples [16], while new 3D photo modeling software, like Agisoft PhotoScan (www.agisoft.ru), AutoDesk PhotoFly (www.labs.autodesk.com) or Microsoft Photosynth (www.photosynth.net) are able to generate 3D models based on a large series of un-calibrated images, using structure from motion (SfM). SfM is a rather new technique to acquire the 3D geometry of an object or a series of objects using 2D images [17]. 3D coordinates of an object are calculated when each surface of the object is covered by at least three images. Characteristic points or feature points have to be detected on different images and matched with each other. In order to recover the position and orientation of the different camera positions, a system of geometrical projective matrices has to be solved, based on the 3D coordinates of these points [18]. The first step in this process is the (automatic) detection of these feature points. Different methods have been developed for image matching, e.g. methods based on region growing algorithms using seeding points [19] or the use of a GCP database [20]. The algorithms are based on least squares estimation, so a solution is obtained by an error minimization estimator [21]. Using the resulting parameters, the extrinsic transformation parameters of the images in a virtual 3D space can be calculated.

In most cases, as with using Agisoft PhotoScan, and when the geometric properties of the images and used cameras are unknown, at least three images are required to make a 3D reconstruction of an object. The projection of a real object in 3D on a 2D image plane and the inverse transformation of these 2D image coordinates to a virtual model in a 3D space, require the extrinsic and intrinsic
camera parameters as well as the focal length [11]. The extrinsic parameters are calculated by the image matching process, the intrinsic parameters and focal length of the images are extracted from the metadata of the image. The final 3D model is generated by the projection of pixel rays from different positioned and oriented images in the 3D space. The intersection of these rays results in the final 3D point, which is triangulated in order to get a 3D mesh or DSM.

For the construction of the DSM discussed in this paper, Agisoft PhotoScan is used. The software uses unprocessed images and accompanied metadata for the 3D reconstruction by SfM. The entire process from raw images to fully textured 3D meshes is covered by this software, and the workflow is already used in different archaeological projects, where the technique was used for virtual modeling without a focus on geometric quality [22-24]. Optionally, the images can be complemented with GCPs and with image acquisition coordinates. Thereafter, the modeling process is split into three steps [8]:

1. photo alignment;
2. geometry building;
3. texture building.

4.2. Preparations

The best modeling results are obtained when the coordinates of the acquisition points of all separate images are known. Besides, for the acquisition of GCPs, an equal spatial distribution is advisable to obtain good results for the image triangulation and absolute orientation [25]. For the DSM discussed in this paper, a minimum of twelve GCPs per stereo couple were selected. This selection was performed by the manual identification of unambiguously determinable features on the images. A total number of forty-eight points were selected, measured with RTK GNSS. Using the Flemish Flepos RTK system in this area, with a baseline of 10-15 km, an accuracy better than 2.9 cm in planimetry and 6.3 cm in altimetry (99.9% or 3σ) was achieved [26]. This accuracy is higher when smaller baselines are present, like during applications in the city of Ghent (Belgium) [27], but the presented accuracy is still ten times smaller than the GSD and therefore sufficient for this modeling application.

4.3. 3D reconstruction

As discussed above, the position and orientation parameters of each image are estimated based on an image matching algorithm. First, a set of feature points is detected in each image and these points are matched with feature points from other images, following an optimization algorithm. The final scene structure and the orientation and position of the used images are illustrated in Figure 2. The numbered squares, representing the images, and the feature points give an impression about the correctness of the image alignment. This first scene visualization also enables the selection of the study area by defining a bounding box around this area. In Figure 2, the GCPs are visualized as flags with numbers. The rather equal spatial distribution of the GCPs becomes clear in this figure.
Instead of using the 3D feature points, the positioned and oriented images are used for the actual 3D reconstruction after the image alignment. A 3D mesh is generated based on the intersection of perspective pixel rays. The linear projection parameters of these 2D image pixels in a 3D space are defined by the focal length of the used camera and result in a series of depth maps, representing the distance between the intersection of perspective rays and the focal center of the camera. Combining depth maps from differently oriented positions enables the creation of a dense point set, which will be triangulated into a mesh. This mesh can then be simplified for computational performance reasons. This 3D reconstruction technique is extensively discussed in [11, 28] and the results of the reconstruction of the study area are illustrated in Figure 3.

During the geometric 3D reconstruction, one single color value is assigned to each face in the 3D mesh, based on averaged color values of all corresponding pixels from different images. A more detailed and photorealistic mesh texturing is obtained by the projection of the original images on the
mesh, resulting in a 2D color map or uv-map. Each position on the 3D mesh is assigned to a pixel of this color map or uv-map, which is the average of all pixel values from the images. The level of detail or degree of photorealism is determined by the resolution of this uv-map. A screenshot of the textured study area is presented in Figure 4.

When the construction of the 3D model is finished and the mesh is optionally draped with a texture, the model can be exported. For visualization purposes, the texture map and geometry can be exported to a Wavefront object file (obj), a Stanford polygon file (ply) or a Collada file (dae), in which the texture map is separately stored in a compressed image file (jpeg or png). It is also possible to export only the geometry without textures, the feature points or the orthoimage of the model. In the context of this research, the model is saved as a raster, where each cell represents the local height value. Saving this elevation model as a tagged image file (tiff) enables further analysis in GIS-based software.

![Figure 4: Textured 3D model of the study area](image)

5. Quality analysis of the results

5.1. Statistical analysis

The final SfM-based DSM covers an area of 1,400 x 1,400 m, with a height range of 15.34 m to 73.52 m. A reference ALS-based DSM was generated with the same extends and both models were imported in ESRI ArcGIS for the geometric quality assessment. Using this GIS software, a new raster is calculated with the same resolution of the SfM-based DSM, where each cell represents the difference between the height value of the SfM-based DSM and the ALS-based DSM:

\[ \Delta h = h_{SfM} - h_{ALS} \]

Thereafter, a sample point set, with a variable spatial resolution, was randomly generated, resulting in a point set of 100 points. For each of these points, the difference value of the underlying cell is assigned to new point. The values in this point set will contain some no-data values, because
of gaps in the ALS-based DSM, but also some erroneous points, caused by vegetation. Hence, an outlier threshold needs to be applied, using the following criterion [29]:

\[ |\Delta h| \leq 2.5 \times \sigma \times \sqrt{2} \]

with \( \sigma = 0.30 \) m, which is equal to the GSD. After removing points with no-data values and values above the presented threshold, the resulting 92 points are supposed to represent the entire population of height differences in the study area. Some statistical values of the sample are presented in Table 1:

<table>
<thead>
<tr>
<th>Table 1. Statistical values of the height differences from the random sample set (values in meter)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td><strong>95% Confidence Interval for Mean</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Std. Deviation</strong></td>
</tr>
<tr>
<td><strong>Median</strong></td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
</tr>
<tr>
<td><strong>MAE</strong></td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
</tr>
</tbody>
</table>

The mean absolute error (MAE) and root mean square error (RMSE) are calculated as discussed in [30]. Based on these values, the differences between the SfM-based model and the ALS-based model are statistically significant, since the 95\% confidence interval for the mean does not include zero. However, the systematic error or algebraic mean error is below the GSD value of 0.30 m, which suggests that the height values in the SfM-based model fit the requirements of an acceptable height error equal to the GSD. The distribution of the height differences are presented in Figure 5, in which the approximation of a normal distribution of the differences can be confirmed.
5.2. Visual analysis

As demonstrated in Figure 4, the first visual impression of the model is quite good, in correspondence with the statements of the authors from earlier mentioned archaeological research. However, a more detailed comparison of the SfM-based hillshaded model (Figure 6, right) and the ALS-based hillshaded model (Figure 6, left) demonstrates that the SfM-based model suffers from a rather noisy outlook. Nevertheless, the main morphological structures are clearly visible and the orthoimage (Figure 6, middle) shows no major errors and the facades of the present buildings look straight.

The small irregularities, made visible by hillshading of the DSM, are caused by erroneous pixel ray intersections. These irregularities seem to be coherent with the used 3D reconstruction technique and appear in different other – yet unpublished – test cases. Nevertheless, based on the statistical analysis, the size of this noise does not have influence on the overall geometric quality of the DSM. Another drawback of the SfM-based model is the disability to detect terrain points under the
vegetation canopy, in contrast with the use of ALS. As a result, the generation of DTMs is not possible using SFM, and the Iron Aged structures in this test case, situated in the forest of ‘Kooigem’, are not visible in the generated model. Notwithstanding these drawbacks, the main structures outside the forested areas are clearly visible, like the street and river. This still enables the use of SFM-based models for virtual field prospection.

6. Conclusions

In this paper, the construction and analysis of 3D models was discussed, based on SFM in an archaeological context. This technique was used for a study area around the forest of ‘Kooigem’ (Belgium), where some Iron Aged relics are situated. Aerial images, with a GSD of 0.30 m, were combined with acquisition coordinates and a with series of RTK GNSS based GCPs. These points have a sub decimeter accuracy. The final model has a systematic error of less than the GSD, which enables the usage of this technique for high quality DSMs. This error has been calculated by the comparison of a SFM-based DSM with a ALS-based DSM. The correlation between these two data sets has been demonstrated and resulted in the complementarily of SFM in a high accurate modeling workflow. Moreover, the generated orthoimage, accompanying the 3D model, was of good quality. Hence, the possibilities to use SFM as a less time and funds consuming reconstruction technique in comparison with conventional techniques, had been suggested.

Acknowledgements

Our special thanks goes to the province of West Flanders, and provincial permanent delegate G. Perty in particular, for the use of the airborne laser scanning data, and to the Belgian National Geographic Institute for the airborne images. The Research Foundation Flanders (FWO) is also gratefully acknowledged for funding the work presented in this paper.

References

[22] Dowling, T., Read, A. & Gallant, J. Very high resolution DEM acquisition at low cost using a digital camera and free software, 18th World IMACS/MODSIM Congress, Cairns, Australia, 2009, pp. 2479-2485.
[23] Doneus, M., Verhoeven, G., Fera, M., Briese, C., Kucera, M. & Neubauer, W. From deposit to point cloud: a study of low-cost computer vision approaches for the straightforward documentation of archaeological excavations, 22th CIPA Symposium, Prague, Czech Republic, 2011, pp. 8 (on CD-ROM).