IMAGE COMBINATION INTO LARGE VIRTUAL IMAGES FOR FAST 3D MODELLING OF ARCHAEOLOGICAL SITES

Maria Pateraki[a], Emmanuel Baltsavias[a], Petros Patias[b]

[a]Institute of Geodesy and Photogrammetry, ETH-Hoenggerberg, CH-8093, Zurich, Switzerland
(maria, manos)@geod.baug.ethz.ch

[b]Department of Cadastre Photogrammetry and Cartography, Aristotle University of Thessaloniki, Univ.Box 473, GR 54006, Thessaloniki, Greece
patias@topo.auth.gr

WG V/4

KEY WORDS: documentation, archaeology, combination, matching, DSM, pre-processing.

ABSTRACT:

In archaeological sites there are often several levels of ruins when the site has been constructed upon other ancient or prehistorical sites. It is very important for the archaeological research to map and monitor each step of the excavation. The traditional way for mapping of excavation is time consuming. Digital photogrammetry enables instead fast acquisition and processing of the data. Continuous excavation reveals different levels of findings coming from different chronological periods, images have to be acquired before findings from each level are removed and excavation continues. Therefore, the storage of data and processing time increases due to the number of images acquired from the excavation site. Apart from that, the images are acquired from amateurs, with less experience in acquisition techniques. A method has therefore been developed to transform many individual images into larger *virtual images* which can be considered of being of a normal central projection. Images can be combined under the assumption that the accuracy is sufficient for archaeological documentation. For the generation of each undistorted *virtual image* the Brown distortion model has been used. The model gives the corrections from the distorted to the undistorted image, therefore the invert transformation has to be calculated through an iterative approach, using the non-linear set of Brown equations. For the generation of the *virtual images* 2 strips were used, each consisting of 3-4 images with 70-80% overlap. The strips form two virtual images with sufficient overlap to extract a DSM. The images had to be enhanced and be radiometrically balanced. In this way more features can be extracted and mapped, even small structures that are covered with dust or lie in shadowed or overexposed areas. Methods are presented to reduce noise, enhance edges and contrast. They are essential for higher image quality and further processing. Results from automatic DSM generation using virtual images are presented, using a new multi-pass adaptive matching algorithm. The algorithm uses features as primitives. Through a multi-pass technique at each extracted feature and computed quality measures, a more robust quality control ensures higher reliability of the final result. A comparison of automatic DSM generation from the separate models of raw images and the two *virtual images* is done. All results are compared with a manual extracted DSM and presented.

1. INTRODUCTION

In the last years, digital photogrammetry became a major tool in archaeology. Photogrammetric techniques are used in order to record and document the findings of an archaeological site area. In some archaeological sites there are different levels of ruins, coming from different chronological periods, since in most cases the site has been constructed upon other older sites. The different levels of ruins have to be recorded as the excavation progresses and therefore a significant amount of data has to be stored for further processing and mapping. In each excavation site, there are several excavation holes and for each one of these different levels of ruins may exist. Additionally the number of images acquired depends on the size of the excavation hole, increasing significantly the amount of acquired images. The method presented in this paper aims at reduction of the data storage and easier handling by combining the acquired images into large virtual images under the assumption that the accuracy is sufficient for archaeological documentation. Consequently, the virtual image can substitute a strip of images and can be used for further processing, as DSM generation and 3D mapping. Generation of virtual images through projective transformation is discussed in (Stephani, 1999), thus one image is transformed and not sequence of images. Pöntinen (2000) and Koistinen (2000) use concentric image sequences, which can be projected onto the cylinder surface and adjacent frames combined to a panoramic image.

The excavation site is the ancient city of Eleftherna in Crete (Fig. 1, 2), Greece. Among few excavation holes at the archeological site a specific one has been selected for processing (4m x 4m). The images have been acquired from amateurs with less experience in the acquisition techniques and therefore the geometrical configuration of the block of images and the camera positions and rotations were not optimal for precise bundle adjustment and self calibration and consequently for extraction of 3D information.

Figure 1. Location of test site in Crete (highlighted)

Figure 2. Excavation site of Eleftherna

2. ACQUISITION – FIELD WORK

The images have been acquired in the early morning avoiding in this way the generation of saturated spots that can appear at strong noon sunlight. A digital camera Mavica FD81, with pixel size 4.7 μm, focal length 4.708 mm and dimensions 1024 x 768 pixels (Georgiadis et al., 2000). Colour images were acquired but B/W were finally used. As mentioned previously, the geometry of the block of images had some deficiencies. The ratio base-to-height and the rotations were not optimally configured (bases between the camera stations were rather small, close to 4 cm, and differences in omega angle, were up to 36 grad). In total, 2 strips were used, each consisting of four images with 70-80 % overlap. Additionally, control points have been established by topographic methods with an estimated accuracy of 1-2 cm. Only 10 out of the 17 measured GCP’s were finally used due to unsatisfying visibility and radiometric quality. An example of an acquired image with GCP’s overlaid is shown in Figure 3.

Figure 3. Example of image with control points marked

The estimated height error for the given imaging configuration was σ_Z= 0.04. Even though for archaeological documentation it can be accepted.

3. BUNDLE ADJUSTMENT

The epipolar geometry has been checked using the coordinates of the GCP’s, the exterior and interior orientation of the images given by an adjustment performed on the field. A deviation of approximately 15 - 20 pixels from the true position of the epipolar line was found. Therefore, a bundle adjustment with simultaneous self calibration had to be performed in order to improve the accuracy of the orientation and model the lens distortion parameters. This is necessary since radial symmetric distortion is the largest systematic error source when using solid-state cameras with low cost CCTV – type lenses and short focal lengths (5-20 mm) (Beyer,1992). In the bundle adjustment 7 images were used (3 from the first strip and 4 from the second), 14 tie points and 10 manually measured GCP’s. The in-house developed program SGAP was used and an iterative adjustment was performed. In a first step orientations and tie point coordinates were computed, excluding points with residuals greater than 3 cm. In a second step, camera constant, principal point coordinates and the Brown distortion parameters (Brown, 1971) were computed. Finally epipolar lines have been checked for the existing GCP’s. In Figure 4 the epipolar lines before and after the adjustment for 2 GCP’s are shown. The largest standard deviations computed in the bundle adjustment were (in m): σ_x= 0.0244, σ_y= 0.0193 and σ_z=0.0301. The overall theoretical precision calculated from the 24 points was (in m): σ_x = 0.0171, σ_y = 0.0142, σ_z = 0.0210. The camera constant was computed as 5.164 mm (initially it was 4.708). The precision achieved was within the given requirements and therefore the extraction of the 3D information more accurate and reliable.

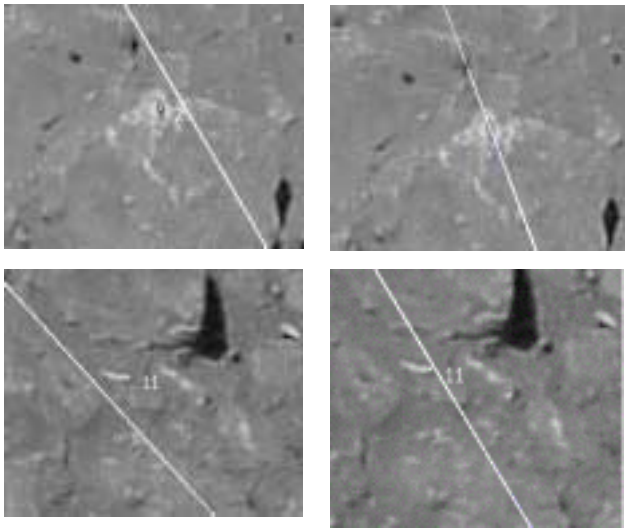

Figure 4. Deviation of epipolar lines before and after bundle adjustment with self calibration. The epipolar lines are shown for 2 GCP’s. On the left column the epipolar lines before bundle adjustment and on the right column after bundle adjustment. Point number 2 is close to the centre of the image, while point 11 near the border where the influence of the additional parameters, modelling lens distortion, is larger.

4. VIRTUAL IMAGE

The *virtual image* can be defined as an image that substitutes a sequence of images acquired from different stations, the sub-images. In case of sub-images acquired with the same camera, the virtual image has the same camera constant but a new virtual size, field of view and camera exterior orientation. Each virtual image substituted 3-4 images, without change in pixel size dimensions (4.7 μm). A reference plane is defined, the camera exterior orientation is set to the average exterior orientation of all 3 or 4 images, the focal length is kept the same and the principal point is set to 0,0. By projecting the 4 corners of each image on to the predefined plane (see Fig.5) and back-projecting through the orientation of the virtual image, the

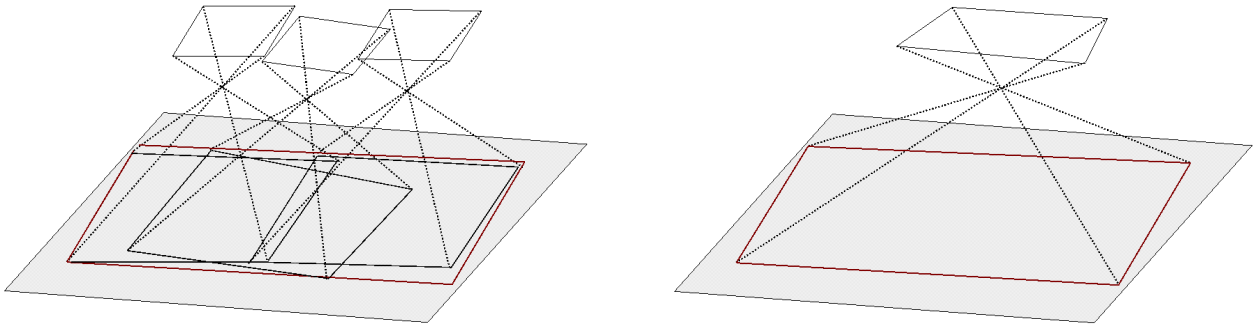

Figure 5. Left image : corners of sub images are projected on the predefined plane and the enclosed rectangle in object space is found. Right image: backprojections of the corners of the rectangle define the size of the virtual image.

enclosed rectangle of the images is defined. Finally, all pixels are resampled and the *virtual image* is created. Since demanding accuracy is less important for archaeological documentation and the terrain height differences are not large, the influences in planimetry and in height are generally small.

The reference plane may be defined and approximated in two different ways: in the first case, a given horizontal plane at a specified height or a best fitted plane from a group of 3D points. The equation of the plane is given as $a_1*x + a_2*y + a_3 = z$ and with least squares adjustment the best fitting plane is calculated. In the second case the intersection of the ray with the plane is calculated through Eq. 1. For nonplanar surfaces, a higher order polynomial surface may be fitted e.g. $a_1 + a_2*x + a_3*x^2 + a_4 *x*y + a_5 *y^2 + a_6 = z$ and the intersection point of the ray with the plane may be found by iterative methods.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} Xo \\ Yo \\ Zo \end{bmatrix} + t \cdot \begin{bmatrix} a_l \\ b_l \\ c_l \end{bmatrix} \qquad (1)$$

where $t = \dfrac{a_P(x_P - Xo) + b_P(y_P - Yo) + c_P(z_P - Zo)}{a_P a_l + b_P b_l + b_P b_l}$

(a_i, b_i, c_i) , i=l,p direction vector of ray and normal vector of plane

Xo, Yo, Zo camera position

x_p, y_p, z_p point that lies on the plane

The version using a best fitted plane was selected as a compromise between accuracy and computation time. Compared to the horizontal plane the surface was better approximated and compared to the higher polynomial surface it was faster. The plane has been selected, so that the residuals of all 3D points were minimum (Fig. 6).

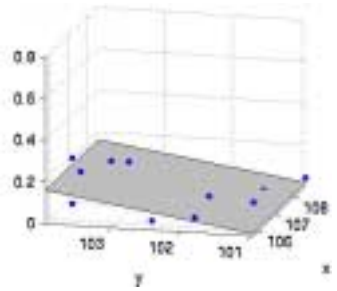

Figure 6. Plane fitted over the 3D points.

When generating the virtual image the corrections resulting from lens distortion have to be calculated. The Brown model (Eq. 2), modeling radial and descentering distortion gives the correction from distorted to the undistorted image, as the input are the coordinates of the distorted image. The inverse transformation corrections have to be computed by solving a two dimensional non-linear system of equations with an iterative method, e.g Newton Raphson. The corrections are given regarding to the distorted coordinates, therefore undistorted coordinated can not be used directly as input in the equation (see below).

$$\Delta x = \bar{x}r^2K_1 + \bar{x}r^4K_2 + \bar{x}r^6K_3 + (2\bar{x}^2 + r^2)P_1 + 2\bar{x}\bar{y}P_2 \qquad (2)$$
$$\Delta y = \bar{y}r^2K_1 + \bar{y}r^4K_2 + \bar{y}r^6K_3 + 2\bar{x}\bar{y}P_1 + (2\bar{y}^2 + r^2)P_2$$

where $\bar{x} = x - x_o$ and $\bar{y} = y - y_o$

$r = \sqrt{\bar{x}^2 - \bar{y}^2}$

x_o, y_o principal point coordinates

Both x_c, y_c undistorted coordinates have to be corrected by Δx and Δy after projecting from ground to image, give distorted coordinates and resample the grey value from that position. If the distortions of the camera are small the corrections can be approximated by using Eq. 2 with opposite signs in the parameters, but in case of the Mavica, the distortions are large and therefore a more accurate computation is needed. In Fig. 7 the influence of lens distortions for Mavica FD81 is shown. The two-dimensional system of equations (Eq. 3) is solved by calculating the Jacobian matrix, inverting and multiplying with the observation vector. The solution vector updated and the iterations continue till the convergence of the solution. As initial approximation the undistorted coordinates are used.

$$f_{\bar{x}}(\bar{x}, \bar{y}) = 0 \qquad (3)$$
$$f_{\bar{y}}(\bar{x}, \bar{y}) = 0$$

where

$$f_{\bar{x}}(\bar{x}, \bar{y}) = \bar{x} + \bar{x}r^2\,\mathrm{K}_1 + \bar{x}r^4\,\mathrm{K}_2 + \bar{x}r^6\mathrm{K}_3 + (r^2 + 2\bar{x}^2)P_1 + 2\bar{x}\bar{y}P_2 - x_c + x_o$$
$$f_{\bar{y}}(\bar{x}, \bar{y}) = \bar{y} + \bar{y}r^2\,\mathrm{K}_1 + \bar{y}r^4\,\mathrm{K}_2 + \bar{y}r^6\mathrm{K}_3 + (r^2 + 2\bar{y}^2)P_2 + 2\bar{x}\bar{y}P_1 - y_c + y_o$$

with x_c, y_c the undistorted coordinates

$$\begin{bmatrix} \frac{\partial f_{\bar{x},i}}{\partial \bar{x}} & \frac{\partial f_{\bar{x},i}}{\partial \bar{y}} \\ \frac{\partial f_{\bar{y},i}}{\partial \bar{x}} & \frac{\partial f_{\bar{y},i}}{\partial \bar{y}} \end{bmatrix} \begin{bmatrix} \bar{x}_{i+1} \\ \bar{y}_{i+1} \end{bmatrix} = \begin{bmatrix} -f_{\bar{x},i} \\ -f_{\bar{y},i} \end{bmatrix} + \begin{bmatrix} \frac{\partial f_{\bar{x},i}}{\partial \bar{x}} & \frac{\partial f_{\bar{x},i}}{\partial \bar{y}} \\ \frac{\partial f_{\bar{y},i}}{\partial \bar{x}} & \frac{\partial f_{\bar{y},i}}{\partial \bar{y}} \end{bmatrix} \begin{bmatrix} \bar{x}_i \\ \bar{y}_i \end{bmatrix}$$

where $i+1$ the solution in the $i-th$ iteration

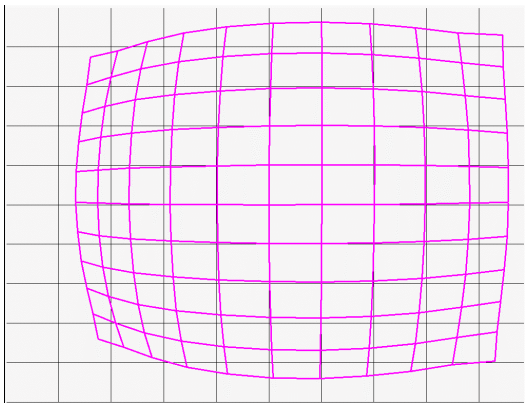

Figure 7. Influence of distortion parameters to the ideal undistorted image – distortions have been amplified 4.5 times for better visualisation

The mosaicking of the virtual image is done by resampling the grey value from the sub image with the minimum distance to the ground point. Alternatively, the weighted average of the grey values from all sub images can be used, using 1/distance as weight. Since the images were radiometrically balanced the differences in the seam lines were minimum; therefore the weighting could be avoided. The size of the 2 virtual images was 1562 x 986 and 1374 x 1662. The difference in the sizes is due to the different number of sub-images and enclosed area.
The created virtual image (Fig. 8) can be used for mapping and archaeological documentation without great loss in accuracy. It is significant though for the mosaicking and further more for the matching the images to be radiometrically balanced. The advantages are twofold: features can be extracted and mapped that are covered with dust or lie in shadowed or overexposed areas and also less radiometric differences exist in the image that is mosaicked.

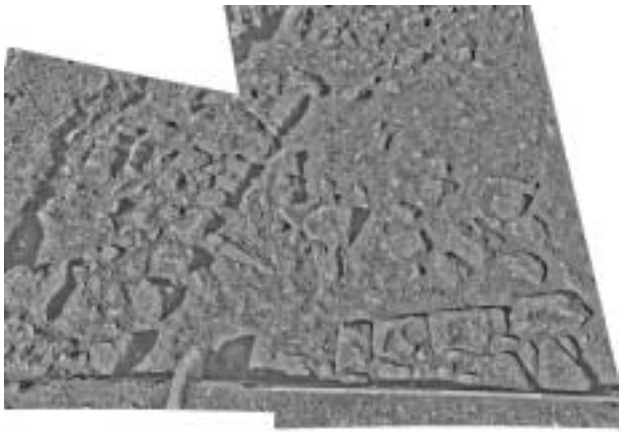

Figure 6. Virtual image created from first strip, using 3 sub-images. The sub-images prior to generation have been radiometrically equalized using Wallis filter (Baltsavias, 1991).

5. DSM GENERATION

5.1 General

Methods for automatic DTM and DSM generation exist already in various commercial digital photogrammetric systems. Many of them can not handle close range imagery, since they are developed mostly for processing aerial imagery. Moreover they show to have problems when dealing with images that have large tilts and bases between the image pairs are small. Therefore we used own developed algorithms in both virtual and normal case of images. In commercial system the virtual images may be used instead, free of distortions and large tilts and a DSM may be extracted.

5.2 Preprocessing

The quality of the images was poor and many features lied in shadows, also some GCP’s. To optimise the images for subsequent processing, filtering has been applied to reduce noise, while preserving even fine detail such as one-pixel wide lines, corners and line end-points. The filter employs a fuzzy method and require as input an estimate of the noise, which may be known or estimated by the method mentioned in Baltsavias et al. (2001). As an estimation of noise the standard deviation has been computed and the average noise level of the images was 1.2 grey values, after the filtering the noise level reduced 50%. In the next step, Wallis filter has been applied to enhance the content of the image and reveal small structures (Fig. 7). It is important prior to Wallis filtering to reduce the noise level otherwise the existing noise will be enhanced.

5.3 Extraction of features

The Canny operator was used to extract edge features in all images, both original and virtual ones. A dense matching can be achieved and the result can be accurately interpolated into a grid. Canny operator was chosen among others because it performs better than others in terms of processing time and the results are of reasonable quality.

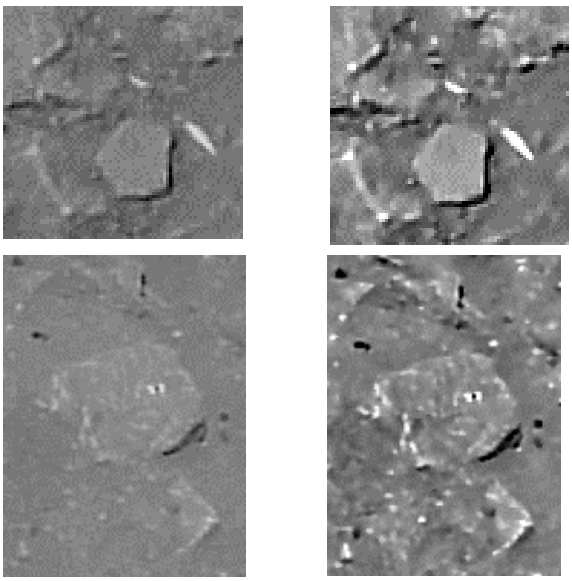

Figure 7. Details of image before (left) and after (right) preprocessing. Features are enhanced even in areas with low contrast

5.4 Matching strategy

A strategy is followed in order to get first some reliable and accurate match results in 2 pyramid levels. These can provide better surface approximation, reduction of search space, and support for weaker match features. Raw match features are matched as dense as possible, thus making the blunder detection and correction easier, avoiding mismatches being propagated to the lower levels. Due to the fact that rotations and shifts existed in the images, plus the bases were not large enough, matching used object space information to get some first approximate values. The approximate 3D information was used also by the forward intersection performed at the end of the algorithm in the least squares adjustment.
For the extracted edgels approximate heights have been interpolated from the coarse surface generated from the existing GCP’s. The interpolation of heights is done only for the template image and by back-projecting on to the search image a first approximation of the point is found.
Furthermore, a multi-patch approach is adopted where up to 3 passes of matching are performed with different parameters (such as search range and patch dimensions). Larger patches are less sensitive to noise, occlusions, multiple solutions etc. while smaller ones are more accurate and better preserve height discontinuities. The algorithm employs area-based correlation in the first stage, and least squares matching at the end for higher accuracy. Since least squares matching performs slower than cross correlation, it is used in a second stage as the last step in the quality control, where most of the mismatches have been excluded in previous steps in the quality control. It can estimate the position with sub-pixel accuracy or reveal suspicious points and flag them and thus increase robustness of the algorithm. Additionally, epipolar constraints can be used with a small weighting which enables the search area to be extended 1-2 pixels perpendicular to the epipolar line, since the estimation of the points that lie near the borders of the image can be less accurate.
During run time, quality measures are calculated, which play important role in the elimination of blunders and false matches. The cross correlation coefficient, the 2nd best similarity score and its distance to 1st one, the size of search window, the change of similarity measure between the 3 patch sizes, the change of position between patch sizes, the angle of dominant edge direction with the epipolar line, the residuals from 3D forward intersection, the change of final point position from the starting position, calculated standard deviations for x and y pixel coordinates are the quality criteria calculated. The quality criteria are combined according to the possible occurring error. E.g. in case of an occlusion, the cross-correlation coefficient would be small and the similarity measure would be generally decreasing from the largest to the smaller mask. Additionally the consistency of height in the local neighbourhood is checked. The matched points are assigned to error groups (e.g. occlusion, multiple solution, etc.) depending on their quality measures. Thresholds for each group of errors are computed from a statistical analysis of the quality measures extracted in each given pyramid level.

5.5 Results

Constrained and unconstrained method gave similar matching results. The constrained method was significantly faster than the unconstrained, because of the reduction of search space. Matching results were subsequently filtered to remove remaining suspect points, using a median filtering. In average 18.000 points per model were successfully matched. The matched 3D edges were finally converted to 3D gridded points through interpolation, with 0.02 m grid spacing. The gridded DSM’s of the normal images with best overall accuracy(see below) were merged through averaging in the overlap regions. The quantitative analysis of the DSM was done using reference 3D points (16 GCP’s), collected with topographic methods with an accuracy of 2 cm. The points were interpolated from the raw matched data and compared. The differences in height were calculated and as estimation of the accuracy their standard was used (Table 1).

Image pairs	Average std. dev. (in m)	RMS	Max abs. error
Strip 1: pair 1-2	0.04	0.05	0.12
Strip 1: pair 2-3	0.06	0.08	0.18
Strip 2: pair 1-2	0.03	0.04	0.09
Strip 2: pair 2-3	0.05	0.06	0.15
Strip 2: pair 3-4	0.09	0.1	0.18

Table 1. Accuracy of DSM’s extracted from image pairs using constrained approach (Normal images).

Since the different DSM’s were overlapping, the ones with the best accuracy, resulting from the comparison with the manually extracted DSM, were selected for merging (Fig. 8). The large difference in accuracy for pair 5-4 depends on the orientation data (small base). The orientation was significantly improved with the bundle adjustment but small errors remained. Therefore from strip 1 the pair 1-2 was selected and from strip 2 the pairs 1-2 and 2-3. The average standard deviation in these 3 models was 0.04.

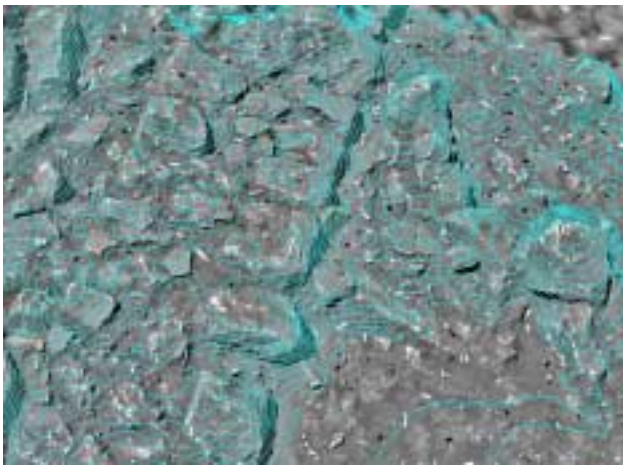

Figure 8. Part of the DSM. Contours with an interval of 0.01 m overlaid on the image.

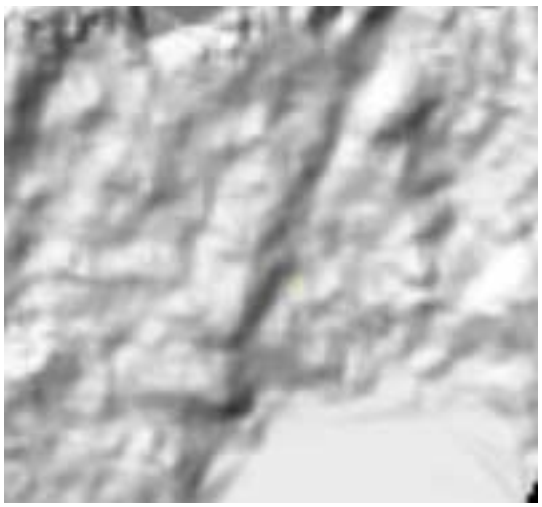

Figure 9. Shaded relief of same are as in figure 8

The matching performed on the 2 virtual images, generated from the images of the 2 strips gave an average accuracy of 0.06 m. The DSM was filtered and gridded to 0.02 m grid. The accuracy suffices for archaeological documentation; therefore virtual images can substitute sequences of images.

5.6 Conclusions

Virtual images can substitute sequences of images if the accuracy degraded by assuming that the object space is a plane, is sufficient for archaeological documentation. In case of higher accuracy, a DSM should be used in the algorithm. Therefore there are two possible ways to work with virtual images. Approximating the terrain with a plane and generating virtual images that can be used for the extraction of the 3D surface or using the normal images, extracting the 3D surface and generating the virtual image using this surface. Then, the virtual images can be stored and used for further processing in the future. The extraction of the 3D surface depends on the accuracy of the estimation of exterior and interior orientation. The use of virtual images leads to image data size reduction and easier handling.

REFERENCES

Baltsavias, E.P., 1992. Multiphoto geometrically constrained matching. Ph.D. dissertation, Institute of Geodesy and Photogrammetry, ETH Zurich, Report No. 49.

Baltsavias, E., Pateraki, M., Zhang, L., 2001. Radiometric and Geometric Evaluation of IKONOS Geo Images and their use for 3D building modelling. Joint ISPRS Workshop "High Resolution Mapping from Space 2001", Hannover, Germany, 19-21 September (on CD-ROM).

Beyer Horst, 1992. Geometric and Radiometric Analysis of a CCD - Camera based Photogrammetric Close-Range System, Institute of Geodesy and Photogrammetry, ETH Zurich, Report No. 51.

Brown, D.C., 1971. Lens Distortion for Close-Range Photogrammetry. Photometric Engineering, pages 855-866, Vol. 37, No. 8, 1971.

Georgiadis, C., Tsioukas, V., Sechidis, L., Stylianidis, E., Patias, P., 2000. Fast and accurate documentation of archaeological sites using in the field photogrammetric techniques. International Archives of Photogrammetry and Remote Sensing, Vol XXXIII, Supplement B5, Amsterdam.

Koistinen, K., 2000. 3D Documentation for archaeology during Finnish Jabal Haroun project. Vol. XXXIII, Part B5, Amsterdam 2000, pp. 440-445.

Pöntinen, P., 2000. On the creation of panoramic images from image sequences. Vol. XXXIII, Part B5, Amsterdam 2000, pp. 635- 641.

Stephani, M., 1999 Digitale Abbildung durch projektive Transformation eines fiktiven Bildes. Festschrift für Prof. Dr.-Ing. Heinrich Ebner zum 60. Geburtstag, Hsg.: C. Heipke und H. Mayer, Technische Universität München, Lehrstuhl für Photogrammetrie und Fernerkundung, 1999, pp. 303-311.