ON AUTOMATIC ORTHOPROJECTION AND TEXTURE-MAPPING OF 3D SURFACE MODELS

L. Grammatikopoulos[a], I. Kalisperakis[a], G. Karras[a], T. Kokkinos[a], E. Petsa[b]

[a] Laboratory of Photogrammetry, Department of Surveying,
National Technical University of Athens (NTUA), GR-15780 Athens, Greece
[b] Department of Surveying, The Technological Educational Institute of Athens (TEI-A),
Ag. Spyridonos Str., GR-12210 Athens, Greece
E-mail: lazaros@central.ntua.gr, ilias_k@central.ntua.gr, gkarras@central.ntua.gr, tk97010@survey.ntua.gr, petsa@teiath.gr

KEY WORDS: Orthorectification, DEM/DTM, Laser scanning, Texture, Visualization, Automation, Heritage Conservation

ABSTRACT

Photo-textured 3D surface models, and orthophotography in particular, are most important photogrammetric products, notably in heritage conservation. However, conventional software typically uses surface descriptions obtained via 2D triangulation; additionally, it cannot handle image visibility. Ignoring multiple elevations and image occlusions is clearly too restrictive for a complex surface shape. Geometric accuracy and visual quality are then possible only with tedious human interaction during surface modeling but also orthoprojection. Yet, laser scanning allows today fast collection of accurate, dense surface point clouds and creation of 3D meshes. Close-range photogrammetry is obviously expected to take full advantage of this.
The authors present their approach for an automated production of orthoimages from fully 3D surface representations derived from laser scanning. In a first step, the algorithm detects surface occlusions for the novel view. While common photogrammetric software needs operator-defined patches on individual original images as the source for image content, here all available images are combined for 'viewer-independent' texturing of the new image. To this end, bundle adjustment data allow all surface triangles to be back-projected onto all initial images to establish visibilities. Texture blending is performed with suitable weighting, which controls the local radiometric contribution of each original image involved. Given more than two values, a statistical test allows to automatically exclude outlying colour data. The implemented algorithm was tested at the example of a Byzantine church in Athens to indicate that this coupling of laser scanning with photogrammetry is capable to automatically create novel views from several images, while combining geometric accuracy and visual quality with speed. Finally, future tasks and further elaborations are outlined.

1. INTRODUCTION

Among all photogrammetric products for the documentation of cultural heritage, digital orthomosaics – a combination of geometric accuracy with textured detail – are perhaps the most prominent. This, of course, is not intended to understate the significance of other related products, such as digital developments or cartographic projections, drapings or photorealistic visualisation and animation. In fact, orthophoto generation stands here as the paradigm for a core problem of photogrammetry, which incorporates both surface modeling and photo-texturing.

Compared to conventional aerial mapping, orthoimaging of cultural monuments often faces a number of significant problems. For instance, as discussed in Mavromati et al. (2002), these may include use of amateur cameras on unstable camera platforms; related problems concerning control over image configurations; resulting difficulties in bundle adjustment. However, a matter of primary importance is accurate surface modeling. It needs to be underlined here that a 3D model is not simply a prerequisite for orthoprojection or realistic rendering. Actually, in cases where only photo-realism or animated visualizations are required, then image-based rendering techniques may provide a direct solution (Beraldin et al., 2002). But photogrammetry typically relies on model-based texturing, as it is mostly asked to also produce explicit 3D data and representations for the purposes of geometric or morphological documentation and analysis.

In many close-range applications object shapes may indeed be complex. As a rule, this implies significant occlusion problems. Thus, surface modeling is a key factor for producing orthophoto results, which will be geometrically reliable and visually correct (no 'melting' or 'stretching'). Conventionally, all surface points are collected manually with stereoscopic viewing (the commercial matching algorithms usually require considerable editing in

the case of an archaeological object). It has been demonstrated by Mavromati et al. (2003) that suitable collection strategies, as regards breaklines in particular, are capable of providing results of high quality. Notwithstanding its merits, however, this course is indeed tedious and time-consuming. Its limitations also include registration problems among stereopair-based 3D models in the case of images all around the object. At the other far end of image-based modeling, powerful techniques are being developed, notably in computer vision, for the automatic extraction of 3D surface models from an image sequence without any prior information about objects or camera. Although models of high visual quality can be thus produced, it appears that the obtained accuracies are not yet in position to meet the requirements for most mapping applications (Pollefeys et al., 2000). The metric potential of advanced techniques for an automatic dense reconstruction from small numbers of multiple wide-baseline images (Strecha et al., 2003) also remains to be further assessed. For certain objects classes, semi-automatic (hybrid) methods, based on a basic volumetric model of the scene which is subsequently exploited to constrain stereo matching, have also been presented (Debevec et al., 1996).

On the other hand, range-based modeling (notably through laser scanning) represents a powerful technology capable of sampling vast numbers of surface points at very fast rates. In this sense, it may well provide the required 3D support for orthorectification (Monti et al., 2002). In a wider sense, the same also holds true for creating photo-textured virtual models of real-world scenes, chiefly in computer graphics applications, where visual quality is a major concern (Bernardini et al., 2001; Corrêa et al., 2002). In fact, high-resolution recording of cultural sites and possibilities to promote them through virtual 3D visits, for instance, stimulates research, notably regarding fusion of laser scanning and colour imagery (Beraldin et al., 2002). Certain commercial 3D systems provide model-registered colour texture but, neverthe-

less, the limited image quality usually makes the acquisition of separate images with higher resolution necessary. Clearly, this is equally true for the high requirements of orthoimaging.

However, the laser scanning approach is faced with several problems. Leaving high cost aside, a major question is that of post-processing vast volumes of data for mesh triangulation (including noise removal and hole-filling), which is a very demanding procedure indeed (Böhler et al., 2003). Yet, the next problem is that commercial orthoprojection software handles only surfaces described as a DTM with a unique elevation at each planimetric XY location. Thus, all scanned 3D points are typically processed by 2D triangulation into a 2.5D mesh – with predictable consequences on the final orthoimage (Mavromati et al., 2003, give such an example). Unless one finds suitable projections yielding single-valued 'depth' functions for particular surface types (as done by Knyaz & Zheltov, 2000), orthoimaging algorithms for fully 3D models must necessarily be introduced.

The main task of such algorithms is to avoid the common flaws of orthoimaging (displacement, blind areas, double-projection) by handling the problem of visibility, which is twofold. On the one hand, every individual surface unit (groundel) which is visible in the direction of the orthoimage should be established. In this way, to each orthoimage pixel a unique elevation is assigned. Next is to check whether these surface points are in fact visible from the perspective centre of the original image, too. In case of occlusion, a colour value can be extracted from an adjacent image. Such approaches, based on dense regular DTMs derived from laser scanning, have been implemented in aerial and close-range projects (Kim et al., 2000; Boccardo et al., 2001).

Following the usual photogrammetric practice, in the above and other cases the texture for each visible surface unit is extracted from a corresponding single original image. Provided that overlapping images exist, the source image can be selected according to different possible criteria, for instance: imaging distance, angle formed by the projective ray and the surface; size of the imaged surface triangle. Evidently, this 'single-image' texturing approach can lead to adjacent surface triangles receiving colour from different images with varying radiometric characteristics. The consequences on triangle borders can be radiometric distortion and discontinuity artifacts (El-Hakim et al., 2003). Alternative elaborate responses to this, in the area of computer graphics and computer vision, rely on colour interpolation, or 'blending'. For every point, appropriately weighted combinations of corresponding triangle textures from all available images – or from a suitably selected image subset – on which this point appears are used (Neugebauer & Klein, 1999; Bernardini et al., 2001; Buehler et al., 2001; Wang et al., 2001; Rocchini et al., 2002). In this way of smoothing radiometric difference, seamless texture with no jumps in colour appearance can be obtained – at the possible cost of a certain blurring effect (El-Hakim et al., 2003).

The approaches referred to above have been developed in the field of computer graphics, where life-like animations, realism or illumination are evidently important. A weighting strategy is thus formulated mostly in the context of view-dependent texture mapping, where interpolation schemes favour images observing the object or scene closest in angle to the current viewing direction. In this way, surface specularities and incorrect model geometry may be better captured (Debevec et al., 1996, 1998).

However, it has been pointed out that using a single texture map in 3D models is usually sufficient (Wang et al., 2001). In this sense – though one obviously has much to benefit from research in this field – it appears that static rather than dynamic texturing is preferable for most instances of photogrammetric mapping. A view-independent algorithm weights the contribution of partici-

pating original images according to their spatial relation to the model – e. g. distance, angle of view – and their characteristics – camera constant and resolution – in order to assign a unique colour value to each surface unit (see, for instance, Poulin et al., 1998; Grün et al., 2001).

Although colour blending may be regarded, to some extent, also as an 'error filtering' process, existing error sources may cause geometric and radiometric distortions. Obviously, the final product is significantly affected by the accuracy, with which image orientations – relative to each other as well as in object space – have been recovered. This holds also true for camera calibration parameters. Consequently, a self-calibrating bundle adjustment, including lens distortion, is indispensable. Further obvious error sources causing misalignments include accuracy of 3D recording, quality of surface description by 3D faces and model registration. Finally, though such problems are rather uncommon in photogrammetric applications, significant differences in resolution of the source images, which can blur texture, are also to be considered (Neugebauer & Klein, 1999; Buehler et al., 2001).

Here, an approach is presented for the automated generation of orthoimages from a 3D mesh, derived from laser scanning. The implemented algorithm identifies all surface triangles which are seen in the viewing direction and then establishes whether these appear or not on every available image. Each orthoimage pixel is coloured through weighted blending of texture from all viewing images, whereby outlying colour data are automatically excluded. Results of experimental applications are also given.

2. PROJECTION AND TEXTURING ALGORITHM

For this procedure, the following input data are required:
- a triangulated 3D mesh in the form of successive XYZ triplets describing the object surface;
- grayscale or colour images along with their interior and exterior orientation parameters;
- the equation in space of the projection plane;
- the endpoints in object space, if necessary, of the area to be projected;
- the pixel size of the new digital image.

It is seen that, besides orthogonal, oblique projections may also be accommodated.

2.1 Model visibility and occlusion

In the first step, the triangulated 3D mesh is projected orthogonally onto the specified plane of projection. In order to speed up the search process, the area of the orthoimage is tessellated into a rectangular grid, whose cell is larger than the one of the orthoimage, e.g. by 5 times (its size depends on factors such as the available computer memory, the model size and that of the new image). For each 2D triangle, the circumscribing orthogonal parallelogram is formed. This occupies a number of adjacent grid cells, to which the identity number (ID) of the particular triangle is assigned.

This procedure is repeated for all triangles, resulting into a table containing all triangle IDs ascribed to each individual grid cell. In this way, all projected triangles actually containing a particular pixel of the orthoimage may be established by checking only a limited number of triangles (namely, those ascribed to the corresponding grid cell). Among these model triangles intersected in space by the projecting line of a particular orthoimage pixel, the one whose intersection yields the largest elevation value is selected; the elevation value, which provides the Z-value of the orthoimage pixel, and the triangle ID number are stored. In this mode, the model visibility/occlusion question has been handled.

2.2 Image visibility and occlusion

Coming next to the image visibility aspect, all 3D triangles are now centrally projected, via the corresponding orientation data, onto all images involved. For every planimetric XY value of the orthoimage and its attributed Z-value, the corresponding image xy coordinates on all images are calculated. A similar technique as before is followed: among the model triangles intersected by a particular image ray, the one closer to the projective center is the triangle which has been actually recorded on the image. If the ID number of this triangle is not identical with that already assigned in the previous phase to the orthoimage pixel, it is in fact established that the model point corresponding to this particular ortho pixel is occluded on the examined image. If, on the contrary, the two triangle IDs coincide, then the model point is visible on the particular image, and the RGB values are stored.

Despite the computational burden, colour values are interpolated in the present implementation by bicubic convolution, since it provides an obviously smoother result. However, it is evident that adjacent pixels do not necessarily relate to adjacent model points. Although no discernible effects emerged in the applications, checks may possibly be considered to omit such pixels.

2.3 Texture interpolation

Result of the preceding step for all orthoimage pixels is colour values from several images – unless, of course, the corresponding model point is occluded on all images. In this latter case, a specific colour value marks the particular orthoimage pixels as undefined. For such regions, 'hole-filling' processes can extract colour values from the surrounding model areas (Debevec et al., 1998; Poulin et al., 1998), which has not been examined here. If asked, however, the algorithm can create a map which displays all orthoimage areas visible on 0, 1, 2 and > 2 source images. In this way, additional images, if available, could be introduced in the process to fill the gaps. It is remarked that it is also useful to know which orthoimage areas are visible in more that 2 images, as this allows a test to detect and exclude outliers.

Indeed, in order to assign a final colour value to the orthoimage pixels, outlying values must first be excluded. Generally, these could originate not only from model faults, but also from view-dependent features – such as specular highlights, transparencies, mirrors, refractions, obstacles etc. (Poulin et al., 1998; Rocchini et al., 2001). However, more significant for photogrammetry is probably the case when one approaches model parts not seen by a camera, i. e. borders of occlusion (Neugebauer & Klein, 1999; Buehler et al., 2001). In these instances artifacts might appear, since even very small orientation or registration – or modeling – errors can lead to colour mistakenly derived from an occluding or, respectively, an occluded model point (Fig. 1 shows such an example; see also Fig. 5 but also Fig. 4). One may possibly evaluate the 'occlusion risk' of pixels – for instance, by a comparison of the imaging distance with those of adjacent pixels from their own visible surface point. This is a topic of future study.

Here, a basic statistical test was adopted, provided that a sufficient number (> 2) of colour values are available for a particular orthoimage pixel. Mean (μ) and standard deviation (σ) of colour values are computed each time; individual colour values falling outside the range $\mu \pm \beta \times \sigma$ are excluded. It is estimated that the value of factor β could be around 1 (indeed, in the test presented in the next section using 7 images, it was set $\beta = 1$). After this procedure, the valid contributing colour values from all images are used to generate the final texture of each orthoimage pixel.

A weighted mean of all contributing images is finally used for texturing each particular orthoimage pixel. In view-independent texture mapping, the main factors influencing colour quality are scale (i. e. imaging distance and camera constant) of the source image; its viewing angle (i. e. the angle formed by the intersection of the image ray and the model triangle); and image resolution. In fact, these factors are all combined to yield the size (in pixels) of the 2D triangle on each image, which is regarded as a good indication of the quality of the extracted colour. Hence, as suggested by Poulin et al. (1998), the contribution of all participating colour values are weighted here as relative functions of the corresponding 2D triangle areas (this weighting scheme has also been used by Grün et al., 2001).
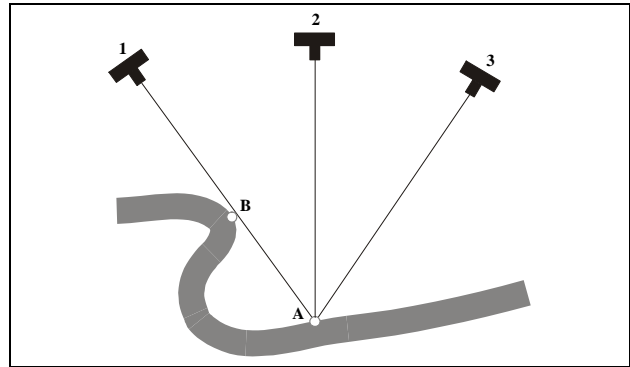

Figure 1. Due to small modeling, calibration and orientation errors, the texture of point B on image 1 may be assigned to A.

The algorithm was initially developed in MatLab and was finally implemented in C. In order to verify its performance and also increase speed, tests were carried out with synthetic data, using images extracted from an existing photo-textured 3D model of a building (Kokkinos, 2004).

3. APPLICATION OF THE ALGORITHM

The object of the experimental application was the entrance of the 11th century church of 'Kapnikarea', an important Byzantine monument in the centre of Athens. Mainly due to its columns, the object is sufficiently complex for the task.

3.1 Scanning and modeling

For surface recording, the Mensi GS200 laser scanner was used. The device scans at a rate of 5000 points/sec, having a 60° vertical field of view. Three separate scans were carried out from a distance of about 5 m, for which a typical value ±1.4 mm of the standard deviation is given (the resolution is 4 mm at a distance of 10 m). For registration, 6 well distributed target spheres were employed, also measured geodetically. The RealWorks Survey 4.1.2 software was used for a target-based registration of scans. The precision of geo-referencing was about ±2.5 mm. In total, 7 million surface points were obtained. Subsequently, these were spatially down-sampled to provide a final 3D mesh, which consisted of about 3 million triangles. Special features of the software (smoothing and peak removal) were used to improve the 3D model. A grayscale intensity map was also obtained (in Fig. 2 the intensity map of the central scan is seen).

3.2 Bundle adjustment

The object has been recorded employing different analogue and digital cameras, which will be used in future tests. Here, results are given for images (2592×1944) from a Sony 5 MegaPixel camera. A total of 7 overlapping images were selected, taken with fixed focusing, to keep interior orientation invariant. All images used are seen in Fig. 3.
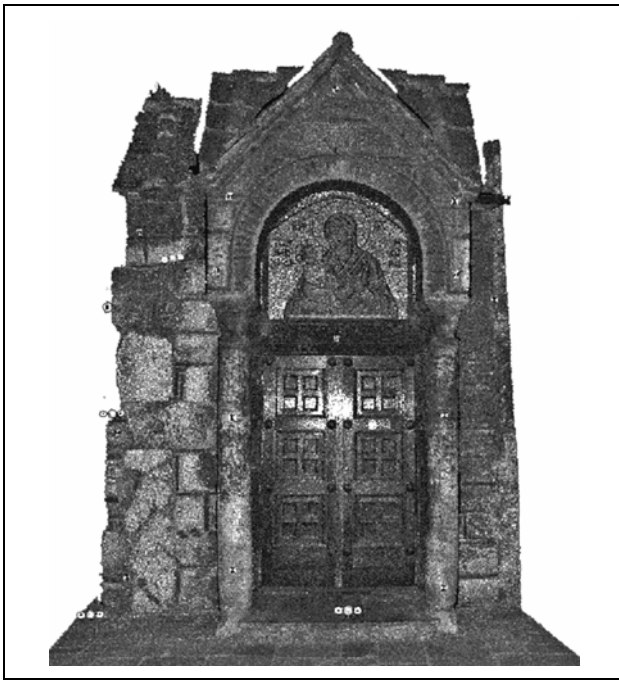
Figure 2. Intensity image of part of the central scan.

A set of 18 signalised points served as geodetic ground control, most of which appeared on all images. Using our own software, a bundle adjustment with self-calibration (including coefficients k_1 and k_2 of radial symmetric lens distortion) was carried out involving 8 tie points. It was possible to achieve a high accuracy, as seen in Table 1.

Table 1. Bundle Adjustment *(7 images; 18 GCPs; 8 tie points)*	
$\sigma_o = \pm 0.28$ pixel	
c (pixel)	2573.94 ± 1.36
x_o (pixel)	10.27 ± 1.03
y_o (pixel)	7.35 ± 0.87
$k_1 (\times 10^8)$	-2.98 ± 0.04
$k_2 (\times 10^{15})$	4.90 ± 0.19
rms (GCP) = 0.3 mm	
σ (TIE) = 1.3 mm	

Using the data for camera calibration and exterior orientations, the images could be used to colour the projections of the model.

3.3 Model texturing

All images used here had been acquired successively, i .e. under the same lighting conditions. Thus, as no radiometric variations among images were observed, texture averaging was considered as adequate (for pre-processing see e. g. Grün et al., 2001). The plane of the church façade was selected as projection plane. The orthoimage pixel size was fixed to 2 mm. As already seen, the algorithm identifies image visibility of all model parts previously established as being visible in the direction of projection. The results for an individual image are shown in Fig. 4.

A next question to be noted, also noticed in Fig. 4, concerns the matter of outlying colour values, particularly close to an occlusion border (cf. section 2.3 and Fig. 1), a problem illustrated in Fig. 5. Above, an extract is seen of the orthoimage derived from all source images without blunder-filtering. The white artifact – clearly originating from the occlusion border of the image in the middle – disappears if colour values deviating more than $\pm\sigma$ from the mean are automatically ignored (below).

The final, automatically generated, orthoprojection is presented in Fig. 6, left. The result is essentially satisfactory. A main imperfection are small holes due to lack of texture (the camera has not been elevated). Most other defects – for example, a certain aliasing at edges of some parts which, however, may not be perfectly perpendicular to the projection plane – are observed only with considerable zooming. On the right, a section is presented utilising all available model and texture information, including parts which are not visible on the left (e. g. a part of the arches).
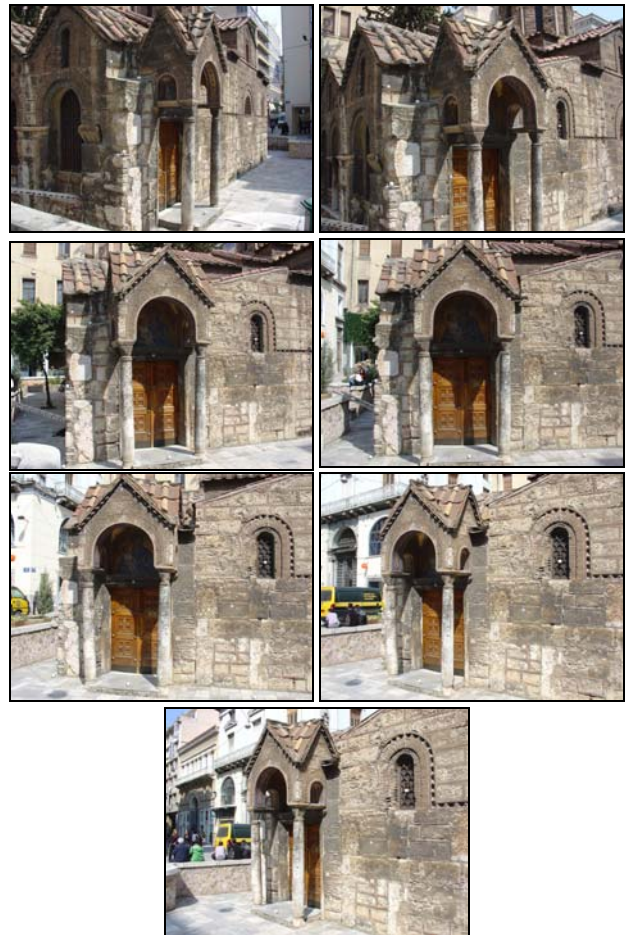

Figure 3. The seven images used in the project.

Regarding speed, a PC (CPU Athlon XP 2.4 GHz, 512 MB RAM) required 5 min to project 3 million triangles onto the images, in a suitable order to facilitate the last step; 1 min to establish the Z-values for each orthoimage pixel; and 13 minutes for the production of a final image 3720×2775 from 7 images 2592×1944.

5. CONCLUDING REMARKS

An algorithm has been implemented for the automatic synthesis of textured views, given a 3D triangulated mesh and precise calibration/orientation information for overlapping images. Model and image visibility are identified, to allow pixel colouring with the weighted average from all viewing images, whereby a basic blunder detection tool allows to avoid artifacts. The product for the paticular object studied here is indeed very satisfactory.

Yet, further enhancements of the technique are both necessary and feasible. Thus, some hole-filling tools should be introduced. Besides, more robust means for outlier avoidance need to be experimented with. For instance, closeness of a source pixel to an occlusion border can be taken into account, probably combined

with the viewing angle. Further, image matching has been used for model and registration refinement (Debevec et al., 1996; see also Bernardini et al., 2001); indeed, a multiple image coverage and precise starting values supplied by the 3D model may allow to refine the model and/or the textured output. The above points indicate possible future tasks of the work presented here.

Figure 5. Part of the orthoimage from 7 images without (above) and with (below) outlier filtering. In the middle is the source image which causes the artifact with its occlusion border.

Figure 4. Orthoprojection (below) of the image above. At the right silhouettes of the columns texturing problems are seen.

REFERENCES

Beraldin J.-A., Picard. M., El-Hakim, S. F., Godin, G., Latouche, C., Valzano, V., Bandiera, A., 2002. Exploring a Byzantine crypt through a high-resolution texture mapped 3D model: combining range data and photogrammetry. Proc. CIPA Int. Workshop on Scanning for Cultual Heritage Recoding, pp. 65-70.

Bernardini, F., Martin, I. M., Rushmeier, H., 2001. High-quality texture reconstruction from multiple scans. IEEE Trans. Visualization & Computer Graphics, 7(4):318-332.

Boccardo, P., Dequal, S., Lingua, A., Rinaudo, F., 2001. True digital orthophoto for architectural and archaeological applications. Proc. Int. Workshop on Recreating the Past: Visualization & Animation of Cultural Heritage, Ayuttaya, Thailand (in CD).

Böhler, W., Bordas Vicent, M., Hanke. K., Marbs, A., 2003. Documentation of German Emperor Maximilian I's tomb. Proc. XIX CIPA Int. Symposium, Antalya, Turkey, pp. 474-479.

Figure 6. Left: Extract of the orthoimage from 7 images. Right: orthoprojection of a model section defined by a maximum Z-value.

Buehler, C., Bosse, M., McMillan, L., Gortler, S., Cohen, M., 2001. Unstructured lumigraph rendering. Proc. ACM SIGGRAPH Annual Conference Series, pp. 425-432.

Corrêa, W., Oliveira, M., Silva, C., Wang, J., 2002. Modeling and rendering of real environments. RITA, IX(1), pp. 1-32.

Debevec, P., Taylor, C.J., Malik, G., 1996. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. ACM SIGGRAPH, pp. 11–20.

Debevec P., Borshukov G., Yu Y., 1998. Efficient view-dependent image-based rendering with projective texture-mapping. Proc. 9th Eurographics Rendering Workshop, Rendering Techniques '98, Springer, pp. 14-26.

El-Hakim, S., Gonzo, L., Picard, M., Girardi, S., Simoni, A., 2003. Visualization of frescoed surfaces: Buonconsiglio Castle – Aquila Tower, 'Cycle of the Months'. Int. Workshop on Visualization & Animation of Reality-Based 3D Models (in CD).

Grün, A., Zhang, L., Visnovcova, J., 2001. Automatic reconstruction and visualization of a complex Buddha Tower of Bayon, Angkor, Cambodia. Proc. 21.Wissenschaftlich-Technische Jahrestagung der DGPF, pp. 289-301.

Kim, Y.-I., Kwon, O.-H., Kim, H.-T., 2000. Detecting and restoring the occlusion area for generating a digital orthoimage. Proc. ASPRS Annual Conference (in CD).

Knyaz, V. A., Zheltov, S. Y, 2000. Approach to accurate photo-realistic model generation for complex 3D objects. Int. Arch. Phot. & Rem. Sens., 33(5), pp. 428-433.

Kokkinos, T., 2004. Automatic Photo-Texturing of Projections of 3D Models. Diploma Thesis, Dept. of Surveying, NTUA.

Mavromati, D,. Petsa, E., Karras, G., 2002. Theoretical and practical aspects of archaeological orthoimaging. Int. Arch. Phot. & Rem. Sens., 34(B5), pp. 413-418.

Mavromati, D., Petsa, E., Karras, G., 2003. Experiences in photogrammetric archaeological recording. Proc. XIX CIPA Int. Symposium, Antalya, Turkey, pp. 666-669.

Monti C., Brumana R., Fregonese L., Savi C., Achille C., 2002. 3D models generation from laser canner data to support digital orthophoto. Int. Arch. Phot. & Rem. Sens., 34(5), pp. 423-428.

Neugebauer, P., Klein, K., 1999. Texturing 3D models of real world objects from multiple unregistered photographics views, Proc. Eurographics '99, Computer Graphics Forum, 18(3).

Pollefeys, M., Koch, R., Vergauwen, M., van Gool, L., 2000. Automated reconstruction of 3D scenes from sequences of images. ISPRS J. Phot. & Rem. Sens., 55:251–267

Poulin, P., Ouimet, M., Frasson, M.-C., 1998. Interactively modeling with photogrammetry. Proc. Eurographics Workshop on Rendering '98, pp. 93-104.

Rocchini, C., Cignoni, P., Montani, C., Scopigno, R., 2002. Acquiring, stitching and blending diffuse appearance attributes on 3D models. The Visual Computer, 18:186–204.

Strecha, C., Tuytelaars, T., van Gool, L., 2003. Dense matching of multiple wide-baseline views. Proc. 9th Int. Conf. on Computer Vision, ICCV '03, Vol. 2, pp. 1994-1201.

Wang, L., Kang, S.B., Szeliski, R., Shum, H.-Y., 2001. Optimal texture map reconstruction from multiple views, Proc. Computer Vision & Pattern Recognition, CVPR '01, vol. 1.