Proc. XX CIPA International Symposium, Torino, 27 September - 1 October, pp. 819-824.

DATA FUSION FROM MULTIPLE SOURCES FOR THE PRODUCTION OF ORTHOGRAPHIC AND PERSPECTIVE VIEWS WITH AUTOMATIC VISIBILITY CHECKING

L. Grammatikopoulos, I. Kalisperakis, G. Karras, E. Petsa*

Department of Surveying, National Technical University of Athens (NTUA), GR-15780 Athens, Greece
* Department of Surveying, Technological Educational Institute of Athens (TEI–A), GR-12210 Athens Greece
E-mail: lazaros@central.ntua.gr, ilias_k@central.ntua.gr, gkarras@central.ntua.gr, petsa@teiath.gr

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

ABSTRACT

Orthophotography – and photo-textured 3D surface models, in general – are most important photogrammetric products in heritage conservation. However, it is now common knowledge that conventional orthorectification software accepts only surface descriptions obtained via 2D triangulation and cannot handle the question of image visibility. Ignoring multiple surface elevations and image occlusions of the complex surface shapes, typically met in conservation tasks, results in products visually and geometrically distorted. Tiresome human intervention in the surface modeling and image orthorectification stages might partly remedy this shortcoming. For surface modeling, however, laser scanners allow now collection of numerous accurate surface points and creation of 3D meshes. The authors present their approach for an automated production of correct orthoimages (and perspective views), given a multiple image coverage with known calibration/orientation data and fully 3D surface representations derived through laser scanning. The developed algorithm initially detects surface occlusions in the direction of projection. Next, all available imagery is utilised to establish a colour value for each pixel of the new image. After back-projecting (using the bundle adjustment data) all surface triangles onto all initial images to establish visibilities, texture 'blending' is performed. Suitable weighting controls the local radiometric contribution of each participating source image, while outlying colour values (due mainly to registration and modeling errors) are automatically filtered with a simple statistical test. The generation of a depth map for each original image provides a means to further restrict the effects of orientation and modeling errors on texturing, mainly by checking closeness to occlusion borders. This 'topological' information may also allow establishing suitable image windows for colour interpolation. Practical tests of the implemented algorithm, using images with multiple overlap and two 3D models, indicate that this fusion of laser scanner and photogrammetry is indeed capable to automatically synthesize novel views from multiple images. The developed approach, combining an outcome of geometric accuracy and visual quality with speed, appears as a realistic approach in heritage conservation. Further necessary elaborations are also outlined.

1. INTRODUCTION

The shortcomings of conventional aerial orthoimages are widely discussed in the context of 'true orthophoto' generation. But an equally important parallel investigation is taking place in the domain of close-range photogrammetry, notably in architectural applications where orthoimages and other digital projections indeed constitute the main documentation tool. Leaving important issues of radiometry aside (quality or radiometric compatibility of source images, shadows), it is mainly two geometric sources of distortion which need to be addressed: surface modeling and orthorectification algorithm.

Notwithstanding certain additional issues typically facing orthoimaging in close-range tasks, e.g. use of non-metric cameras on unstable platforms and complex bundle configurations, accurate surface modeling is undoubtedly the key factor in ensuring geometric and visual quality. El-Hakim et al. (2003) have discussed the respective merits and limitations of interactive or automatic image-based methods for 3D model capturing and visualisation. In view of the complex shape of many cultural items, outcomes depend not solely on the actual accuracy of the models but, primarily, on their completeness. In general, commercial matching software appears as rather unsuitable for archaeological objects. Mavromati et al. (2003) indicated that manual photogrammetric approaches, relying on a suitable point and breakline collection strategy, is indeed capable of providing high quality results – at the unavoidable cost in both effort and time.

Today, however, laser scanners allow the rapid collection of innumerable surface points. Without overlooking the problems – including cost and difficulties in post-processing large volumes of data for mesh triangulation (Böhler et al., 2003) – scanning, among its other merits, offers itself to a fusion with photogrammetry, in particular as a fully 3D basis for orthoprojection. Even after successful post-processing of raw data and a 3D triangulation, however, a typical orthorectification software will still not yield the desired outcome, being incapable of handling the two questions illustrated in Figure 1: image occlusions; and model occlusions (since it is limited to 2.5D surface descriptions).

In the aerial case (notably regarding urban areas) orthorectification might mostly be satisfied with a 2.5D digital surface model. But conventional software cannot cope properly even with such a case (visibility of all model points on the image is assumed), giving rise to algorithms for handling image occlusion (e.g. Oda et al., 2004). More than often, furthermore, object complexity in terrestrial projects dictates the use of fully 3D models. Rigorous approaches of digital orthorectification, therefore, must face the questions of both image visibility as well as model occlusion in the direction of projection (Wanshou & Yixuan, 1999; Boccardo et al., 2001; for the aerial case see Kuzmin et al., 2004). Obviously, thus putting the issue of visibility extends well beyond the strict limits of orthoimaging to also include e.g. generation of perspective views and, ultimately, place the discussion in the wider context of texture-mapping. Indeed, an approach like that of Früh et al. (2004), who use multiple oblique aerial images for texturing 3D city models derived from aerial and terrestrial laser scanning, might be viewed as a generalisation of orthoimaging.

Finally, how to choose texture for each output pixel is a crucial question in itself. Due to lighting conditions and errors in camera calibration, bundle adjustment, registration or modeling, one may not expect that corresponding colour and intensity values

from different images will actually coincide, unavoidably leading to the emergence of 'discontinuity artifacts' or radiometric distortion. Thus, it is expedient to use weighted combinations of corresponding texture values from different images, i.e. 'colour-blending' (e.g. Neugebauer & Klein, 1999; Baumberg, 2002).

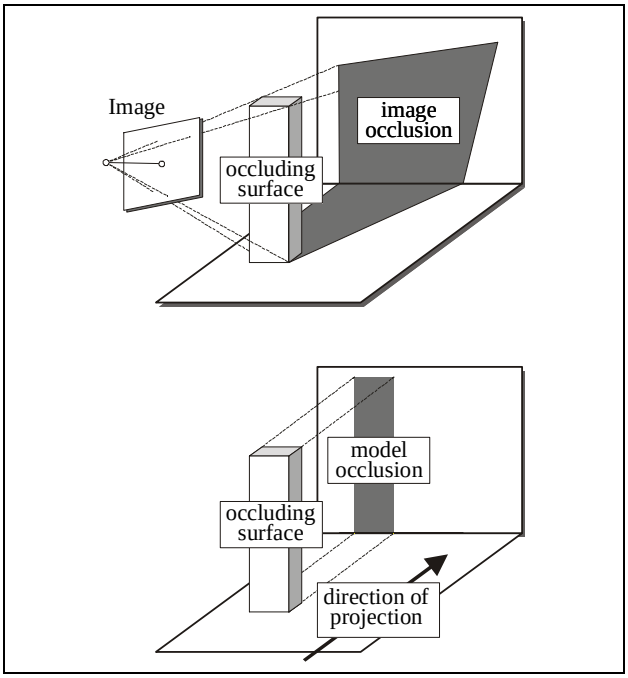

Figure 1. Image and model occlusions.

Stemming from the area of computer graphics, colour weighting is mostly understood in the context of 'view-dependent' texture mapping (Debevec et al., 1996). Yet, for the purposes of photogrammetric mapping it appears as more reasonable to produce a unique 'texture map' by adjusting, after Poulin et al. (1998), the contribution of each source image according to its fixed spatial relation to the object itself (resolution in object space, angle of view). Despite a smoothing effect of blending during texture interpolation, however, existing error sources still call for the introduction of further tools for detecting outlying colour values.

Here, the authors present an algorithm, first reported in Grammatikopoulos et al. (2004), for the automatic generation of orthographic or perspective views from an existing 3D model and multiple image coverage. Further aspects are also discussed and new examples are reported.

2. TEST OBJECTS

A short description of the two test objects will be given here, to allow illustrating the steps of the implemented algorithm. First, the entrance of a distinguished 11th century Byzantine church in Athens was sampled in 3 scans, at a resolution of ~3 mm, using the Mensi GS200 laser scanner (Grammatikopoulos et al., 2004). Based on 6 target spheres, the precision of scan registration was ~2.5 mm. The merged point cloud, initially consisting of 7 million points, was edited with the Geomagic Studio software. The final 3D model comprised 3 million triangles (a view is given in Figure 2). Relying on 18 signalised control points, the 7 images from a 5 Megapixel camera were used for a full self-calibration (including two coefficients of radial symmetric lens distortion) with our bundle adjustment software, producing a standard error of $\sigma_o = \pm 0.28$ pixel.

The second object is the entrance of the cafeteria of the School of Rural & Surveying Engineering of NTUA which was sampled in 2 scans at a ~6 mm resolution using the Cyrax 2500 scanner. A total of 10 signalised control points were manually measured on the intensity images to give a 4 mm precision of registration. The final 3D model was down-sampled to 1.2 million triangles (Figure 2). Bundle adjustment, again with 7 images (Figure 3), from the same camera gave a precision of $\sigma_o = \pm 0.25$ pixel. It is noted that the particular lens exhibits a strong radial distortion, which must necessarily be taken into account (ignoring the distortion increased the σ_o value by a factor of 9). The effect of uncorrected distortion will be illustrated later.

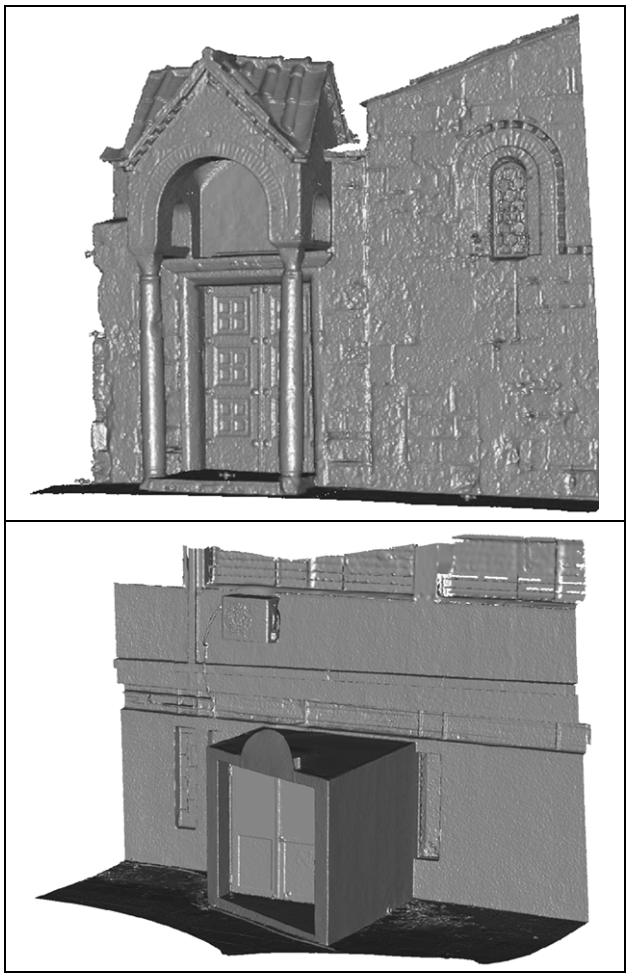

Figure 2. Views of the two 3D models.

3. ALGORITHM AND APPLICATION

3.1 The question of occlusions

Referring to Fig. 1, the task is to establish occlusions, namely: a) which model parts should in fact be visible in a specific new projection (orthogonal or perspective); and b) which regions of each individual source image are entitled to contribute texture.

Regarding the first issue, the blank array of the novel image is initially tessellated into a rectangular grid, in order to speed up the search process (see Grammatikopoulos et al., 2004). Subsequently, the 3D mesh is orthogonally (or perspectively) projected onto the specified projection plane. The circumscribing orthogonal parallelogram of each projected triangle covers certain adjacent cells of the grid, to which the ID number of the particular triangle is assigned. The total of projected triangles containing a certain orthoimage pixel can be, thereupon, established by

checking the triangle ID numbers ascribed to the corresponding grid cell. Among these, the one with the shortest distance to the projection plane provides the orthoimage pixel with a Z-value, thus leading to the generation of an orthoimage depth map.

Figure 3. All images of the second test object.

A similar procedure is adopted as regards image visibility, with all model triangles being, of course, projected centrally onto all participating images. On each image, the image coordinates are calculated for the 3D object space coordinates which are already uniquely associated with each orthoimage pixel. Among model triangles intersected by a particular image ray, that closer to the projection center is the one actually recorded on the image. The coincidence of its ID number with that ascribed at the previous stage to the particular orthoimage pixel implies that the model point associated with this orthoimage pixel is also visible on the image; thus, the RGB values can be interpolated at the particular image location and stored. On the contrary, disagreement of the triangle ID numbers means that the model point is occluded on this image and is not entitled to offer texture to the orthoimage.

The preceding steps have distinguished those model parts which have to appear on the orthoimage and which, among these, are indeed also visible on each individual source image. Hence, for each image a corresponding orthoimage, showing the occluded regions, may be produced. Figure 4 gives such an example.

3.2 Colour interpolation and weighted texture blending

At this stage, the local contribution of each image to a particular orthoimage pixel is determined with standard bicubic convolution. Cases with considerable differences in scale among images need to be further investigated regarding the image window size used in interpolation. In addition, image adjacency does not necessarily correspond to model adjacency, notably in the vicinity of occlusion borders. Since our algorithm is also capable of producing a depth map for each individual source image (Figure 5 gives an example), this information in combination with a depth threshold could allow rejecting border pixels from interpolation (for instance, in the mode outlined in Früh et al., 2004).

Figure 4. Orthoprojection of one image and occluded areas.

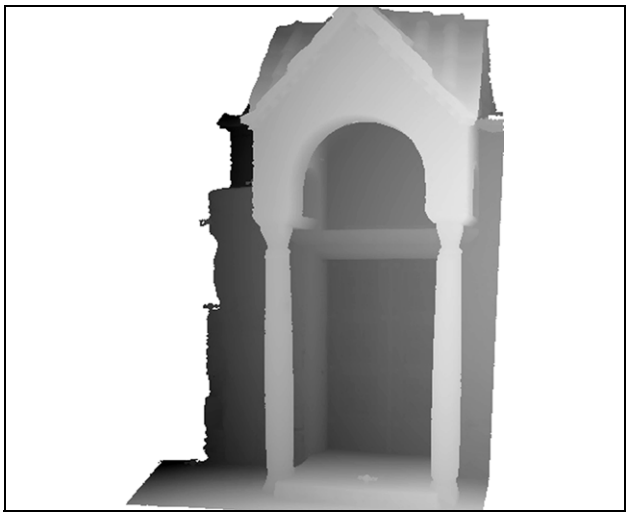

Figure 5. Depth map of a source image.

Thus, interpolated colour data from several images are available for each orthoimage pixel. However, among them outliers may also be present, emanating from orientation or modeling errors; view-dependent features (obstacles or specular highlights); and, of course, proximity to occlusion borders, in which cases even a small registration or modeling error may provide wrong colour values from occluded or occluding model points and, thus, produce image artifacts (Poulin et al., 1998; Neugebauer & Klein, 1999). Such outliers must be discarded before assigning texture to the orthoimage pixels, a question addressed by Bornik et al. (2001) and Baumberg (2002). Grammatikopoulos et al. (2004) compute the mean μ and standard deviation σ from all existing colour values for an orthoimage pixel and discard values falling outside the range $\mu \pm \sigma$. If for some orthoimage pixel there exist colour values from only two source pixels and they differ more than a threshold, the image depth maps allow omitting that pixel in whose vicinity the largest depth differences are met. Figure 6 gives an example for the removal of outliers originating from an occlusion border. Figure 7 illustrates a different case of outlier, that caused by an obstacle (here a tree) depicted in one image.

It is needless to emphasise that the effect of radial distortion is an altogether different matter and cannot be treated in the sense of 'blunders'. If uncorrected, its influence will damage the final product, as clearly seen in the example of Figure 8.

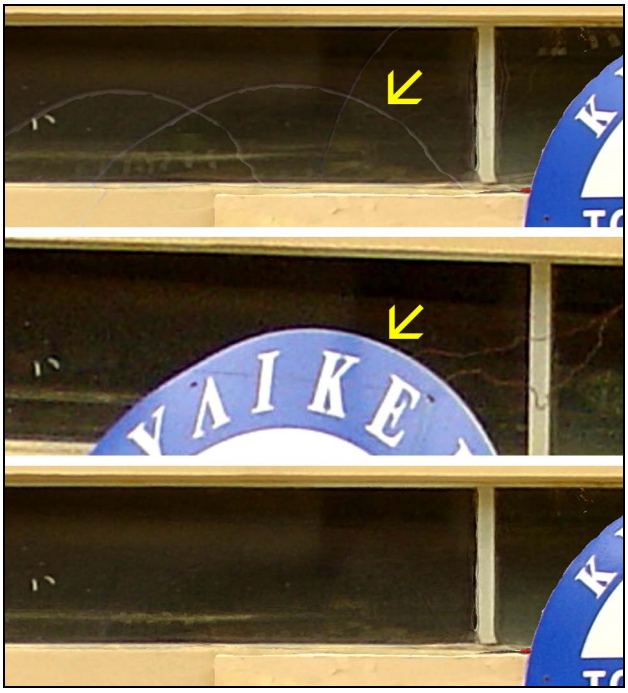

Figure 6. Above, orthoimage detail from 7 images with artifacts due to occlusion borders (e.g. of the image in the middle). In the orthoimage below the blunders have been removed.

Figure 7. Orthoimage detail from 7 images with radiometric distortion (left), due to a tree obstructing the view of the source image (centre), which have been filtered out on the right.

Figure 7. Orthoimage details from 7 images without (left) and with correction of radial lens distortion (right).

Finally, the texture for all pixels of the orthoimage is calculated as the weighted mean of all valid corresponding colour values. The 'strength' of colour values depends on resolution in object space (a result of imaging distance, camera constant and image resolution) and viewing angle (intersection angle of image ray and surface triangle), which determine the size, in pixel dimensions, of the 2D image triangle. Hence – following Poulin et al. (1998) – each contributing colour value is weighted through the surface area of its corresponding image triangle (Visnovcova et al., 2001, also use a similar approach).

3.3 Final products

For orthoprojection, the orthoimage pixel size was set to 2 mm (adequate for the scale 1:20). Figure 9 presents the final result for the second test object, created automatically with the contribution of all 7 images. As in the case of the first object (which is seen in Figure 11), the outcome is considered as satisfactory, notwithstanding certain small gaps caused by lack of texture or unmodeled details. The existence of an aliasing effect, observed with strong zooming at certain edges, is an issue for further investigation (Grammatikopoulos et al., 2004).

Figure 9. Orthoimage from 7 images.

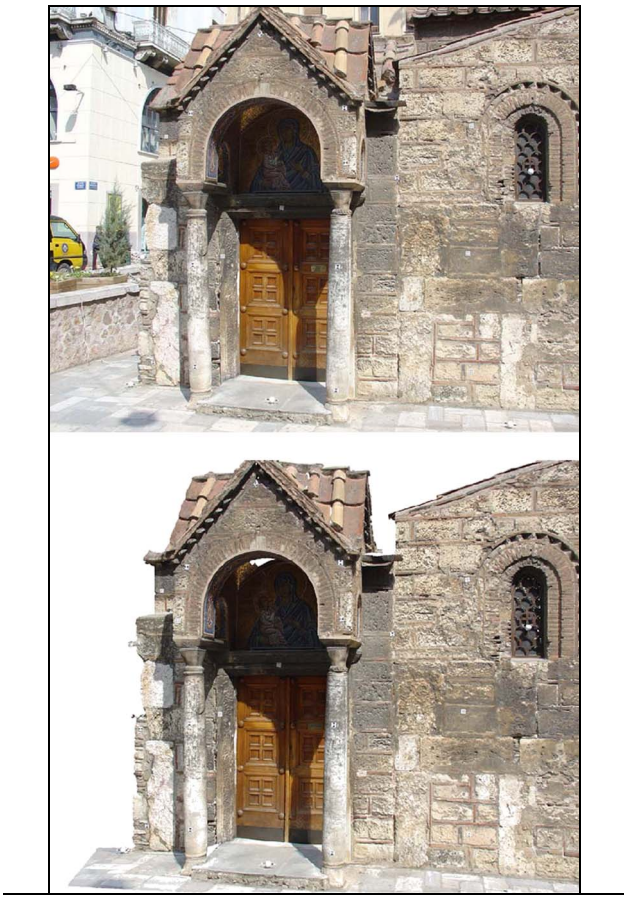

Figure 10. Original (above) and synthetic image (below).

In addition, perspective rather than orthographic views can also be generated as central projections of the surface model, using freely selected camera and orientation. In Figure 10 a synthetic

perspective image is seen along with its corresponding original (which abstained from image synthesis). Finally, with a fully 3D surface model and corresponding texture, the algorithm can also produce cross-sections by simple depth thresholding, as illustrated in Figure 12 (Grammatikopoulos et al., 2004).

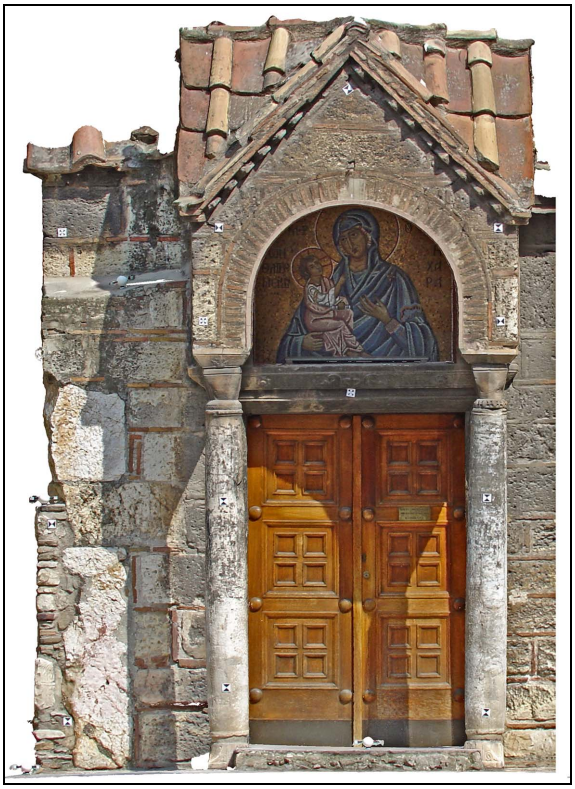

Figure 11. Orthoimage from 7 images.

Figure 12. Vertical cross-section from 7 images.

4. CONCLUDING REMARKS

The presented algorithm, situated in the context of photogrammetric texture mapping, results in the automatic composition of high quality orthographic or perspective views of an existing 3D surface mesh. Relying on a two-fold visibility check and texture blending, this approach produced here very satisfactory images. Discarding of outlying colour values suppressed artifacts, while at the same time contributing to image sharpening.

It must be admitted, however, that in the present instances the input data were of high geometric and radiometric quality. This refers to carefully edited and triangulated 3D models, but also to the employment of signalised control points and self-calibrating bundle adjustment. Furthermore, the merit of all images having been acquired under the same lighting conditions was that there existed no actual need for pre-processing (on this see e.g. Visnovcova et al., 2001). The difficulty encountered when attempting to use images of the second test object taken with no fixed camera settings and in windy weather causing tree shadows on the object surface to move were insurmountable. A further field of investigation is indeed shadow removal. For the detection of shadows a geometric approach, similar to that regarding occlusions, might be applied using a parallel rather than perspective projection. An additional issue to be dealt with is 'hole-filling', based on colour extraction from adjacent model regions (Poulin et al., 1998). Measures already mentioned here – such as using a image depth map to weaken the participation of colour values near occlusion borders and taking into account significant variations in image scale – are also to be further elaborated. Finally, two more general aspects are open to study: the generation of fully textured 3D models rather than single views of them; and the introduction of image matching techniques in the direction of model and registration refinement (Debevec et al., 1996).

Acknowledgements

This work was supported financially within the framework of programme ARCHIMEDES II (EU and Greek Ministry of Education). The authors sincerely thank Drs. V. Balis and C. Liapakis of GEOTECH Ltd. (Greek representatives of Trimble), for providing scans and ground control, and Dr. M. Tsakiri (NTUA). We are also grateful to INFOCAD Ltd. (representative of Geomagic Studio in Greece) for kindly providing the license to use this software.

REFERENCES

Baumberg, A., 2002. Blending images for texturing 3D models. Proc. British Machine Vision Conference, pp. 404-413.

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. XIX CIPA Int. Symposium, Antalya, Turkey, pp. 474-479.

Bornik, A., Karner, K., Bauer, J., Leberl, F. Mayer, H., 2001. High quality texture reconstruction from multiple views. Journal of Visualisation & Computer Animation, 12(5):263-276.

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.

El-Hakim S. F., Beraldin, J.- A., Picard, M., Vettore, A., 2003. Effective 3D modeling of heritage sites. Proc. 4th International Conference of 3D Imaging & Modeling, Banff, pp. 302-309.

Früh, C., Sammon, R. Zakhor, A., 2004. Automated texture mapping of 3D city models with oblique aerial imagery. Proc. 2nd International Symposium on 3D Data Processing, Visualization & Transmission, Thessaloniki, pp. 396-403.

Grammatikopoulos, L., Kalisperakis, I., Karras, G., Kokkinos, T., Petsa, E., 2004. On automatic orthoprojection and texture-mapping of 3D surface models. Int. Arch. Photogrammetry, Remote Sensing & Spatial Information Sciences, 35(5):360-365.

Kuzmin, Y. P., Korytnik, S.A., Long, O., 2004. Polygon-based true orthophoto generation. Int. Arch. Photogrammetry, Remote Sensing & Spatial Information Sciences, 35(3):529-531.

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

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

Oda, K., Lu, W., Uchida, O., Doihara, T., 2004. Triangle-based visibility analysis & true orthoimage generation. Int. Arch. Photogrammetry, Remote Sensing & Spatial Information Sciences, 35(3):623-628.

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

Visnovcova, J., Li, Z., Grün, A., 2001. Generating a 3D model of a Bayon tower using non-metric imagery. Proc. Int. Workshop on Recreating the Past: Visualization & Animation of Cultural Heritage, Ayuttaya, Thailand (in CD).

Wanshou, J., Yixuan, Z., 1999. The making of high precision orthoimage of ancient buildings. Asian Conf. of Remote Sensing (www.gisdevelopment.net/aars/acrs/1999/ps4/ps4245.shtml).