AUTOMATIC MULTI-VIEW TEXTURE MAPPING OF 3D SURFACE PROJECTIONS

L. Grammatikopoulos[1,2], I. Kalisperakis[1,2], G. Karras[2], E. Petsa[1]

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

ABSTRACT

Texture-mapping in close-range photogrammetry focuses mostly on the generation of large-scale projections of 3D surfaces, the most common instance being that of orthoimaging. Today, photogrammetry increasingly uses terrestrial laser scanning as basis for generation of 3D models. For a full exploitation of fully 3D data, however, typical shortcomings of conventional orthorectification software (inability to handle both surface self-occlusions and image occlusions; single-image texturing) must be addressed. Here, the authors elaborate on their approach for the automated generation of orthoimages and perspective views, based on fully 3D models from laser scanning and multi-image texture interpolation. The problem of occlusion is solved by first identifying all surface points visible in the direction of projection; texture is then interpolated through blending using all images which actually view each particular surface point. Texture outliers from individual images are automatically filtered out with a statistical test. Yet, further means for excluding outlying colour values are needed. Rather than using the depth maps of source images to identify possible occlusion borders, these borders are automatically extracted directly on the orthoprojection of each image. Their back-projection on the corresponding image, suitably processed with a morphological operator (dilation), defines 'risk buffers' on each image and suppresses their participation in colour interpolation. Combined with a statistical test, this procedure has proved beneficial for the quality of the results. Practical tests based on image sets with considerable variations in image scale have indicated that such new features of the algorithm facilitate the cooperation of laser scanning with photogrammetry for the automatic multi-view synthesis of textured projections.

1. INTRODUCTION

In a context analogous, yet not identical, to that of 'true ortho-photo' generation in aerial cases, investigations are taking place in terrestrial applications (notably in the field of cultural heritage) regarding photo-texturing of 3D models (Alshawabkeh & Haala, 2005; Zhang *et al.*, 2006). Due to the complex shapes of cultural items, the accuracy of 3D models and their completeness are both directly reflected in the quality of results. Merits and limitations of interactive or automatic image-based methods for 3D modeling and visualisation are discussed in El-Hakim *et al.* (2003). However, it may be safely assumed that in numerous instances terrestrial laser scanning permits now accurate surface modeling, thus offering a fully 3D basis for texture mapping or for creating 2D projections, and orthoprojections in particular.

Even so, the main shortcomings of conventional orthorectification software as regards the question of occlusion remain (blind areas, double-projection etc.). This problem is twofold. On one hand, surface self-occlusions in the direction of projection have to be established, a problem irrelevant in typical cases of a 2.5D model but clearly crucial for 3D models. Besides, conventional software does not respond properly even in the simple case of a 2.5D model by failing to recognize image occlusions. Therefore, rigorous tools for digital orthoprojection must face both aspects of occlusion. The visibility issue extends well beyond the strict limits of orthoprojection to also include creation of perspective views and, ultimately, place the matter in a wider framework of texture-mapping. In fact, methods using multiple oblique aerial images for texturing 3D city models derived through aerial and terrestrial laser scanning (Früh *et al.*, 2004) might be viewed as a generalisation of orthoimaging.

On the other hand, modeling and registration errors, as well as lighting conditions, lead to unequal colour values from different images, thus producing abrupt changes in texture (discontinuity artefacts). Therefore it is advisable, in both view-dependent and view-independent texture mapping, to blend textures from more images to generate a weighted average of corresponding values

(Baumberg, 2002; Orzan & Hasenfratz, 2005). It is believed that a single texture map in 3D models is usually sufficient (Wang *et al.*, 2001). Indeed, for purposes of photogrammetric mapping it seems reasonable to adopt the view-independent alternative and thus create a single texture map by regulating, after Poulin *et al.* (1998), the contribution of each source image according to its fixed spatial relation to the object. Despite a certain smoothing function of the colour blending process, however, existing error sources still call for the introduction of additional tools for removing outlying colour values. As a preliminary precaution it is advisable to incorporate all participating images in a single self-calibrating bundle adjustment, including lens distortion.

The authors have presented a multi-view algorithm for automatically creating orthographic or perspective views based on existing 3D models (Grammatikopoulos *et al.*, 2004, 2005; Karras *et al.*, 2007). In this contribution the basic algorithm is briefly outlined, followed by a more detailed report on the innovations which focus on the elimination of faulty colour values.

2. OCCLUSION CHECKING

Although for other purposes this time-consuming checking for occlusions might be by-passed via a basic statistical test (Orzan & Hasenfratz, 2005), it is considered that the occlusions need to be explicitly identified if large-scale projections of high visual and geometric quality are to be expected. Hence, the algorithm establishes: a) those surface parts which should be visible in the specific orthogonal (or perspective) projection; and b) the areas of each individual source image which are entitled to contribute texture to the model parts visible in the projection.

First, the 3D mesh is orthogonally projected onto the specified projection plane q (Fig. 1). All projected model triangles which contain an orthoimage pixel P can be thereupon identified. As a consequence, all 3D triangles intersected by the projection ray of P are known. Among all intersection points (A, B in Fig. 1), that closest to the projection plane q (point A) should be visible.

Thus, a depth map of the orthoimage is generated, namely each orthoimage pixel P is uniquely associated with a model point.
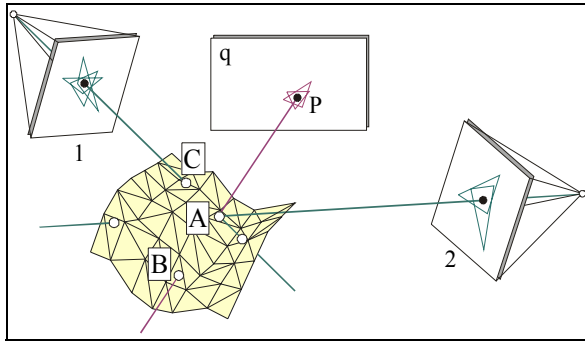

Figure 1. Search for model point and image points which correspond to an orthoimage pixel P (see text).

In a similar fashion, model triangles are projected centrally onto all participating images. For each orthoimage pixel P its homologue point on the surface (A in Fig. 1) is also projected onto all images. The 2D triangles containing this image point, and hence the model triangles intersected by this projection ray, are identified. The intersection points closest to the respective projection centres are those actually seen on each image. In this mode it is established which images will contribute colour for a particular pixel P (image 2 in Fig. 1) and which will not (image 1).

This procedure decides which model parts should appear on the orthoprojection and which among these are also visible on each individual source image. Consequently, from each image a corresponding orthoimage, in which occluded areas are left blank, may be generated. Details about techniques for speeding up the search process are given in Grammatikopoulos et al. (2004).

3. PHOTO-TEXTURING

3.1 Colour interpolation

The colour value from each image corresponding to a particular orthoimage pixel is interpolated via bicubic convolution due to its smoothing effect. If images have significant scale differences it seems reasonable to fix the interpolation window size not in image space but rather in the space of the projection; besides, pixel adjacency does not necessarily imply adjacency in space, e.g. at occlusion borders. Such issues are currently under study (Karras et al., 2007). For a model point visible on no image, the corresponding orthoimage pixel is left blank. This may be seen beforehand, as the algorithm creates a map showing how many images view each orthoimage pixel. Such gaps might be treated with 'hole-filling' (Debevec et al., 1998; Poulin et al., 1998).

3.2 Colour blending

Final texture for each orthoimage pixel is the weighted mean of corresponding colour values from all images which are entitled to contribute. The 'weight' of colour values depends on viewing angle (intersection angle of image ray with surface triangle) and resolution in space (result of image resolution, camera constant and imaging distance) which determine the size of the 2D image triangle. Thus, each participating colour value is weighted, after Poulin et al. (1998), with the surface area of its corresponding 2D triangle (this is also adopted in Visnovcova et al., 2001).

However this approach is suitable for images with similar scales (i.e. similar groundel size), such as those used in Grammatikopoulos et al. (2004, 2005). In the tests performed here, the scale

differences among images were significant. This has a blurring effect on the final blended texture (Neugebauer & Klein, 1999). The problem may be handled either by modifying the weighting scheme, e.g. by using the squared surface areas of 2D triangles, or by selecting the N images with the best geometry. Both these approaches, implemented here, produced a certain improvement in image sharpness while retaining radiometric continuity.

3.3 Elimination of colour outliers

But among colour values from different images, which are to be blended for texturing an orthoimage pixel, outlying values may also be included. These emanate from view-dependent features (e.g. obstacles) and orientation or modeling errors, particularly in the vicinity of occlusion borders, resulting in image artefacts (Poulin et al., 1998; Neugebauer & Klein, 1999). Hence, such a value is to be discarded before texture is assigned to orthoimage pixels. Bornik et al. (2001) apply a median filter for detecting values outside a user-specified range. Grammatikopoulos et al. (2004) compute the mean μ and standard deviation σ from all valid colour values for each orthoimage pixel and discard those outside the range $\mu \pm \sigma$. In the present implementation, σ may also refer to the median which is more sensitive to outliers. It is to note that the effect of lens distortion cannot be treated in the sense of 'blunders'; if not corrected directly it will ruin the final product (see Grammatikopoulos et al., 2005, where successful removal of occlusion-induced artefacts and of obstacles are also shown). It is clear that this test necessitates at least three colour values. If for an orthoimage pixel textures from only two source images are available, image depth maps may be useful.

3.4 Use of image depth maps

Besides to the orthoimage depth map, similar maps may also be produced for each source image. Such maps are seen in Fig. 3. The 'occlusion risk' of a pixel may be evaluated by comparing its imaging distance with those of adjacent pixels.
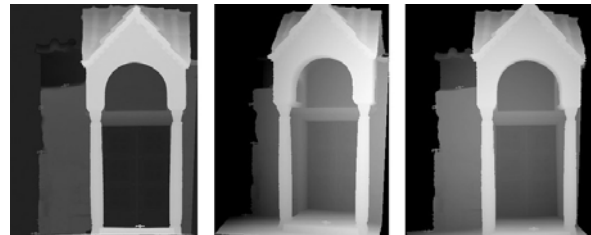

Figure 2. Depth maps of orthoimage (left) and of two images.

Thus, if only two colour values are available and differ by more that a limit set according to the σ-value of neighbouring pixels, that pixel whose imaging distance differs more from that of its neighbours can be discarded (Karras et al., 2007). Furthermore, it appears that this basic statistical test is not always fruitful in cases of scattered colour values. In this sense, it is advisable to discard in advance – or introduce with a reduced weight, e.g. in a way similar to the image 'weight map' of Bamberg (2002) – pixels with depth values which differ by more than a threshold from their neighbours. However, changes in depth maps do not necessarily imply the existence of occlusion borders. To resolve this, a comparison with the orthoimage depth map or the depth maps of other images would be necessary.

3.5 Direct treatment of occlusion borders

Here, a direct single-image treatment of occlusion borders has been implemented based on morphological transformations and edge extraction. From each individual source image the corresponding orthoimage is generated, in which image occlusions are

marked with a specific colour. The orthoimage is then binarized to single out these occluded areas. An opening operator, applied to the binary image, removes small areas of a few pixels. Next, image occlusion edges of this orthoimage are extracted using a Laplace filter. These 3D entities are, thereupon, back-projected on the original source image and then expanded with a dilation operator. In this mode, the occluded areas of the original image have slightly grown. The range of the dilation process depends on the assumed precision of 3D modeling and image orientation. The contribution of the image in the final orthoimage will not involve texture from the regions added along edges through this operation. This procedure is illustrated in Figure 3.

Figure 3. From top to bottom: orthoimage from a source image (image occluded areas in green); occlusion edges extracted after opening; original source image; orthoimage occlusion edges projected onto the source image and subjected to dilation.

After the above steps, texture blending involves all valid colour values which pass the $\mu \pm \sigma$ outlier test. The described dilation process of occlusion borders improves the final result compared to the exclusive application of the σ-test, as seen in Fig. 4.

4. PRACTICAL APPLICATIONS

The algorithm was experimentally applied to the 3D model of a Roman archaeological site. The Mensi GS200 laser scanner was used with 1 cm resolution, covering the object in 9 scans.
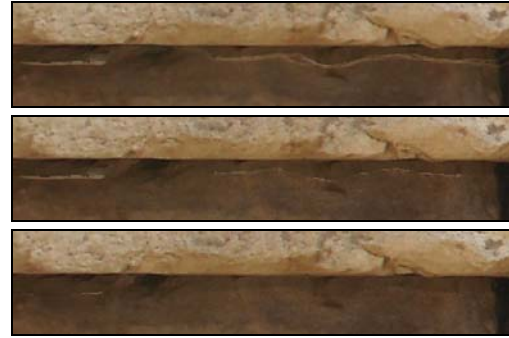

Figure 4. Orthoimage detail. Top: without exclusion of outliers. Middle: outlier exclusion with σ-test. Bottom: dilation of occlusion edges and outlier exclusion with σ-test.

To overcome difficulties in scan registration with the available software, the 3D similarity transformation

$$\mathbf{X}_i = \mathbf{R}_k\mathbf{x}_{ik} + \mathbf{t}_k$$

was used to transform in a one-step solution all control and tie points $\mathbf{x}_i$ of scans k in the geodetic system ($\mathbf{X}_i$) through rotations $\mathbf{R}_k$ and translations $\mathbf{t}_k$ of each scan system ($\sigma_o = \pm 7$ mm).

From the final model three parts (two of them with large depth differences) were selected, consisting of 0.8, 3.5 and 3.6 million triangles. The first part, which is seen in Fig. 5, was recorded on 10 images with large differences in imaging distances (Fig. 6).
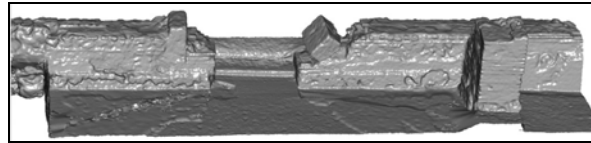

Figure 5. Surface model of part I.

Figure 6. Image set of part I.

Part II (Fig. 7) includes a semi-cylindrical surface, a demanding shape for orthoprojection which puts the algorithm to the test. A large number of images were used (16 and 23) taken with two cameras and largely varying orientations and imaging distances. Plan views of the imaging configurations are seen in Fig. 8.
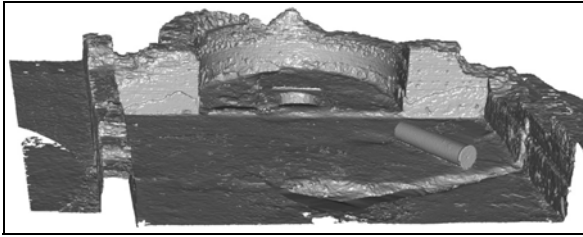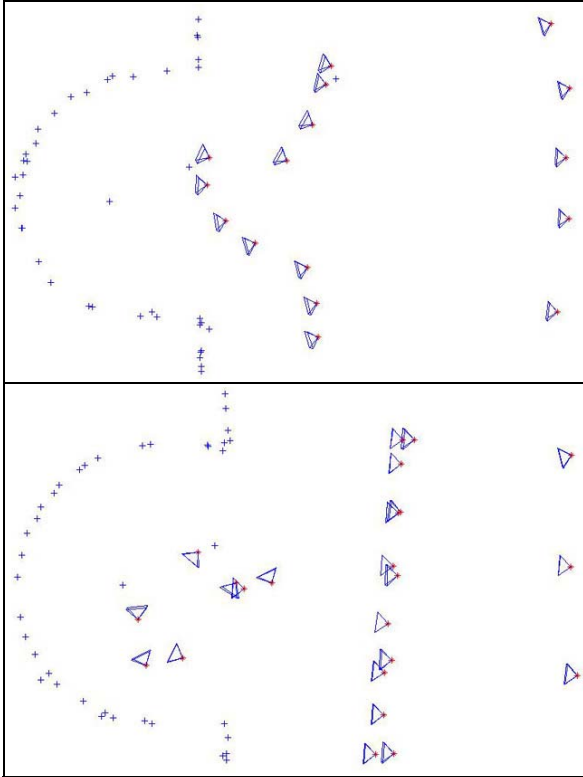
Figure 7. Surface model of part II.

Figure 8. The two imaging configurations for part II.

The third part (Fig. 9) had two occluding planes and hence was suitable also for the generation of a cross-section. Provided that 3D data and imagery are available, the algorithm has the capability of generating such sections by a simple depth thresholding (Grammatikopoulos *et al.*, 2004). It is noted that a cross-section was also created for part II by 'jumping' over the column which lies in front of the façade. A total of 17 images, acquired from different imaging distances, participated in the texture-mapping of part III. Figure 10 is a plan view of the image configuration. This image set is seen in Fig. 11.
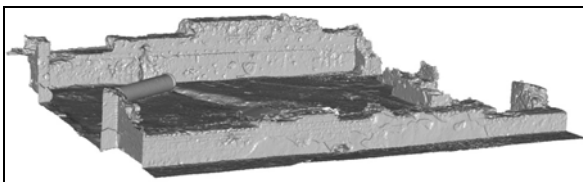

Figure 9. Surface model of part III.

In texture-mapping – and particularly for the generation of large scale products by colour blending – it is evidently imperative to secure a tight registration of images to each other and to the 3D model. Hence all images have to participate in a self-calibrating

bundle adjustment. This is necessary, despite the fact that configurations in such cases might be weak or unconventional (lack of convergence between image axes, similar image strips taken from different distances, limited variations in object depth as in the case of part I) producing parameter values of low precision.

All parts were imaged with a 7.2 Megapixel camera; part II was additionally recorded with a 10 Megapixel camera. All standard errors of the adjustments were 0.6 pixel. This is attributed to the fact that all control points were extracted from the point cloud.
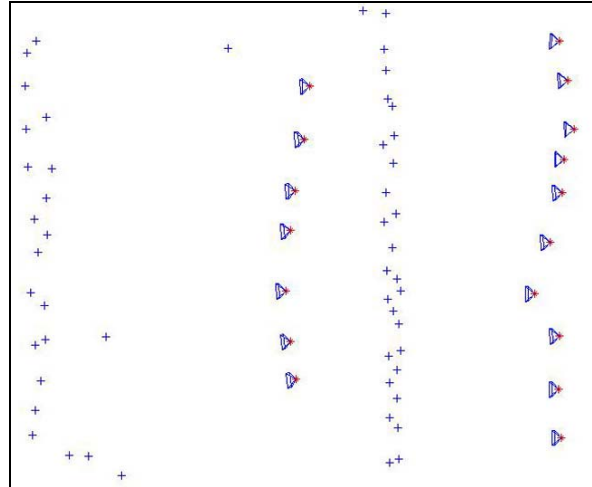

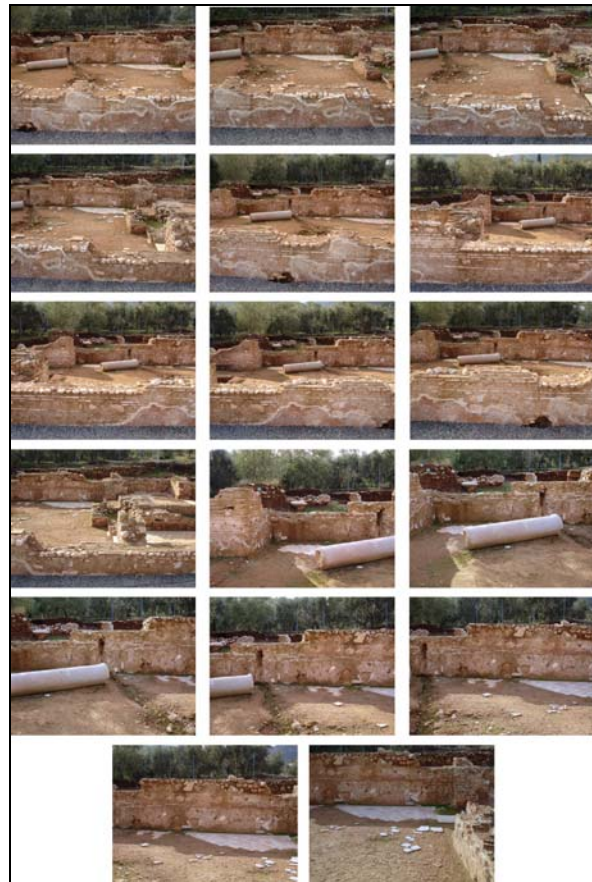Figure 10. Imaging configuration for part III.

Figure 11. Image set of part III.

Figure 12. Orthoimages of part II from the two image sets (top: 16 images, bottom: 23 images).

Figure 13. Orthoimages of part III from 17 images (top). Below is the projection of the posterior wall from the same images.

The principal point location of the common camera as estimated in the bundle adjustment for part I (small object relief) differed considerably from those estimated from the adjustments of parts II and III, and had a large uncertainty which, however, does not affect ray intersections significantly. The values for the camera constant deviated by no more than 1.3‰ from their mean, while radial distortion curves were coincident. The final products indicate that for the purpose of precise texture-mapping one might trust control points extracted directly from the 3D data.

Figure 14. Orthoimage of part I from 10 images.

After histogram matching, all images of each block participated in orthoimage generation with pixel size fixed to 2 mm in object space. In Fig. 14 the orthoimage of part I is seen. Final products for part II are presented in Fig. 12. The orthoimages refer to two image sets taken at different times. Between the two recordings a part of the monument had been covered for protection, which accounts for the deformations in the top left corner of the lower orthoimage. It is also to note that in this orthoimage of Fig. 12 the column lying in the foreground has been removed. Finally,

Fig. 13 shows the orthoimage of part III, along with the projection of the hind wall created by depth thresholding.

5. CONCLUDING REMARKS

Given an accurate 3D surface mesh and precise image calibration and orientation data for overlapping images, the presented algorithm identifies object and image visibilities to allow a fully automatic multi-view synthesis of textured projections, namely orthographic as well as perspective (for the latter see Grammatikopoulos *et al.*, 2005), but also cross-sections. Pixel colouring succeeds by the weighted contribution from all viewing images. A standard blunder detection tool anticipates most cases of mis-texturing. This issue has been further studied here by using (e.g. instead of depth maps) morphological operators for minimizing possible effects of image occlusion borders on colour-blending.

The presented results are considered as quite satisfactory. The algorithm behaves well in the case of edges, as illustrated in the example of Fig. 15. It is seen there that a clear edge is generated from 9 images, i.e. from all images among the 23 of the image set which view this particular object area. The imperfections are mainly small holes, which witness lack of texture. This problem is artificially aggravated by the dilation operator, which – next to expanding occlusion edges – also expands all existing small

clusters of occluded pixels. An alternative treatment would be necessary (e.g. by connected components labeling) for the removal of such occlusion 'blobs'. Possible further elaborations of the approach have been indicated in Grammatikopoulos *et al.* (2005) and Karras *et al.* (2007).

Figure 15. Detail of orthoimage, generated by blending texture from 9 images, in the area of an object edge.

Closing, it is stressed that, despite all possible tools for handling texturing error sources, the precision of the original input data is of primary importance. This refers not only to the accuracy of 3D modeling but also to the coherence of reconstructed imaging configurations, i.e. image blocks must be processed in a unique self-calibrating adjustment. More than the estimated precision of the unknowns, one may say that it is in fact their correlations which are mainly valuable in the present context.

REFERENCES

Alshawabkeh Y., Haala N., 2005. Automatic multi-image photo-texturing of complex 3D scenes. *XVIII CIPA Int. Symposium*, Torino, 27 September - 1 October, pp. 68-73.

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

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), pp. 263-276.

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

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

Früh C., Sammon R., Zakhor A., 2004. Automated texture mapping of 3D city models with oblique aerial imagery. *2nd Int. Symposium on 3D Data Processing, Visualization and 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), pp. 360-5.

Grammatikopoulos L., Kalisperakis I., Karras G., Petsa E., 2005. Data fusion from multiple sources for the production of orthographic and perspective views with automatic visibility checking. *XVIII CIPA Int. Symposium*, Torino, pp. 819-824.

Karras G., Grammatikopoulos L., Kalisperakis I., Petsa E., 2007. Generation of orthoimages and perspective views with automatic visibility checking and texture blending. *Photogrammetric Engineering and Remote Sensing*, 73(4), pp. 403-411.

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

Orzan, A., Hasenfratz, J.-M., 2005. Omnidirectional texturing of human actors from multiple view video sequences. *IEEE Int. Workshop on Human Computer Interaction*, Beijing.

Poulin P., Ouimet M., Frasson M.-C., 1998. Interactively modeling with photogrammetry. *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. *Int. Workshop on Recreating the Past: Visualization & Animation of Cultural Heritage*, Ayuttaya, Thailand (unpaginated CD).

Wang L., Kang S. B., Szeliski R., Shum H.-Y., 2001. Optimal texture map reconstruction from multiple views. *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR '01*, vol. 1.

Zhang Y., Yana L., Lu X., 2006. Precise texture modeling with 3D laser scanning. *Geoinformatics 2006: Remotely Sensed Data and Information*, Proc. SPIE, vol. 6419, paper no. 64191Z.