COMBINING LASER SCANNING AND PHOTOGRAMMETRY - A HYBRID APPROACH FOR HERITAGE DOCUMENTATION

Norbert Haala, Yahya Alshawabkeh

Institute for Photogrammetry, Universitaet Stuttgart, Germany

Abstract

High quality 3D models of cultural heritage sites can be generated efficiently by laser scanning, which allows the accurate and dense measurement of surface geometry. In addition to the geometric data collection, texture mapping based on additionally collected digital imagery is particular important for this type of application. This requires a combined processing of range and image data sets. For this purpose, they have to be registered or aligned by a suitable transformation to a common reference coordinate system. The involved transformation parameters can be determined based on corresponding elements to be extracted from the different data sets. In the paper an efficient edge detection algorithm is presented, which allows for the automatic segmentation of such primitives even in complex scenes. In order to achieve a high quality 3D photo-realistic mode,l this alignment process has to be followed by an automatic texture mapping, which is discussed in the second part of the paper. The presented algorithms are demonstrated in the framework of a project aiming at the generation of a 3D virtual model of the Al-Khasneh, a well-known monument in Petra, and a Roman Theatre in ancient Jerash city, Jordan.

Categories and Subject Descriptors: I.4.8 [Scene Analysis]:Range data, I.4.7 [Feature Measurement]: Texture

1. Introduction

Terrestrial laser scanning is frequently used to provide high quality 3D models of cultural heritage sites and historical buildings. Based on the run-time of reflected light pulses, these sensor systems allow for the fast, reliable and area covering measurement of millions of 3D points. While this enables an effective and dense measurement of surface geometries, the provision of image data is frequently additionally required for a number of applications. This is especially true in the context of heritage documentation, since the complete documentation of heritage sites usually implies a high quality texture mapping based on supplementary image data. For this reason, some commercial 3D systems directly integrate a digital camera in order to simultaneously collect corresponding RGB values for each LIDAR point. This induces camera viewpoints, which are identical to the laser scanning stations. However, these camera viewpoints might not be optimal for the collection of high quality imagery as they are required for texture mapping. Additionally, laser scanning for the documentation of complex object structures and sites frequently has to be realised from multiple viewpoints. This can result in a relatively time consuming process. For outdoor applications these large time differences will cause varying conditions of illumination and changing shadows. Thus, the captured images will be subject to considerable radiometric differences, which will disturb the visual appearance of the resulting textured model. For these reasons, the acquisition of object geometry and texture by two independent sensors and processes will be advantageous. This allows for an image collection at optimal positions and time for texturing, which is especially important for the high requirements to be met during the realistic documentation and subsequent visualisation of heritage sites.

While an independent image collection considerably improves the quality of the available texture images, data processing and evaluation will complicate. If a camera is directly integrated with the laser scanning system, the captured images can be directly linked to the 3D point cloud, if a proper calibration of the complete system is available. In contrast, for independent camera and laser scanner stations, the combined evaluation requires a co-registration or alignment of the collected range and image data sets as a first processing step. This co-registration is usually realised based on corresponding primitives within

the different data sets. For this purpose, primitives like 3D points or lines have to be extracted from the range data and matched against their 2D projections in the image. While a number of algorithms are available to determine the camera pose from such 2D-3D correspondences, the automatic provision of such point or line features is difficult. Within section 2 an algorithm, which can generate suitable primitives based on efficient edge detection also in complex scenes is presented.

If the camera pose is available, the texture images can be directly mapped to the corresponding 3D surface patches for visualisation. However, texture should of course only be assigned to those parts of the 3D object model, which are actually visible. This requires the detection of occlusions in the respective texture images. While this can be solved relatively easy for simple object geometries, the effort for visibility analyses increases considerably for geometrically complex object representations as they are usually provided from laser scanning in the context of cultural heritage documentation. In these scenarios the 3D object model is frequently represented by meshed surfaces. Since these meshes are derived from dense 3D point clouds, a huge amount of surface patches is generated. Thus, efficient approaches have to be made available to solve the visibility problem. As it will be discussed in section 3 this can be realised based on approaches originally developed in computer graphics.

2. Feature Extraction and Registration

Registration, or alignment, is the first step to be solved for the combined processing of range and image data collected from different viewpoints. Frequently, different laser scans are combined using measurements at artificial targets like spheres or signals, which can be detected and identified easily in the range data. Since such targets are also clearly visible in images, they can additionally be used for the registration of the image data. However, the provision and measurement of such targets requires additional effort. This is especially true, if a large number of potential texture images have to be collected. Additionally, these targets may occlude important parts of the object within the texture images [LHS00]. For these reasons, it is advantageous to use natural features to provide the required corresponding primitives for registration.

Traditionally, in photogrammetry corresponding points are used for determination of the image pose. As it is demonstrated exemplarily for a cultural heritage application in Figure 1 (left), suitable points can be identified well in an image. In contrast, the exact measurement of corresponding points within data from laser scanning is almost impossible. This is deomnstrated for the captured 3D point cloud (Figure 1 middle) as well as the range image (Figure 1 right), which was derived from laser measurements by re-interpolation to a regular grid. Due to the difficulties in exactly identifying such point structures, the measurement accuracy frequently is not sufficient for registration [LS05].

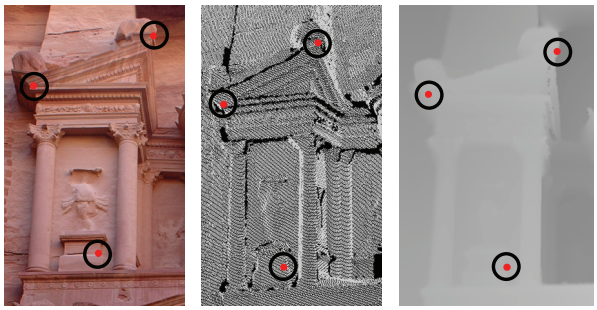

Figure 1: Corresponding points in image, point cloud and range image.

In contrast, linear features can usually be extracted and measured more accurately and reliably [CH99]. Such edge structures are frequently available within scenes of man made environments and allow for the reliable pose estimation by spatial resection. Thus, an automatic extraction of such edge structures is beneficial compared to a measurement of distinct points [LH02]. For this reason an efficient segmentation algorithm was developed to automatically extract such features of interest from the collected range and image measurements.

2.1. Curvature based range image segmentation

Usually, range data segmentation does not use unordered 3D point clouds (Figure 1 middle), but is based on 2.5D representations by raster grids (Figure 1 right). These range images maintain the original topology of the original laser scans. Thus, neighborhood relations are available implicitly and tools from image processing can be adopted. For this reason an easier implementation of segmentation algorithms is feasible compared to unordered 3D point clouds.

Similar to image segmentation, existing approaches for range data processing can be categorized in region-based and edge based techniques. Region based approaches group range pixels into connected regions by some homogeneity measure. A survey is e.g. given in [HJJ*96] and [MLM02].

While region based approaches allow for a reliable extraction of smooth or planar surface patches from range data, the direct segmentation of edges is difficult. This results from the limited spatial resolution of range data. Additional problems result from the large amount of noise at such height discontinuities, frequently occurring due to multipath effects. Thus, only a few edge based segmentation algorithms have been developed [SD01]. However, since edges are very well suited for co-registration of range and image data, this was our motivation for the development of an algorithm for the extraction of such structures from range images. There, the range image edges are extracted based on the analysis of the mean curvature values. The surface is approximated locally by an analytic representation, which is used to calculate the different properties of the respective patches.

By these means, the mean curvature at edges is used to detect local maxima or zero crossings in the range image. Within further processing steps, a multi-scale edge detection and a subsequent skeletonization is used to increase the reliability and accuracy of edge detection and localization.

2.2. Computation of Mean Curvature Values

Range data segmentation requires an appropriate surface description. This description should be rich, so that matches of similar elements can be detected, stable so that local changes do not radically alter the descriptions, and it should have a local support so that the visible objects can be easily identified. These characteristics are provided by the mathematical properties of the mean curvature, which is closely related to the first variation of a surface area. Unlike the Gaussian curvature, the mean curvature depends on the embedding, for instance, a cylinder and a plane are locally isometric but the mean curvature of a plane is zero while that for a cylinder is non-zero. Mean curvature is invariant to arbitrary rotations and translation of a surface, which is important for surface shape characterization. Since mean curvature is the average of the principal curvatures, it is slightly less sensitive to noise during numerical computations. Due to these characteristics, mean curvature values can provide stable and useful measures for detecting surface features in range and intensity images.

Several techniques are known for the efficient estimation of the mean curvature. As an example analytical methods fit a surface to a local neighbourhood of the point of interest. This surface approximation is then used to compute the partial derivatives needed to calculate the curvature values. Our approach is based on the work of [BJ88], who proposed an analytical technique for estimating the mean and Gaussian curvature. The advantage of this approach is its flexibility to estimate the curvature values at multiple scales, and the efficient computation of the values by optimized convolution operations. The approach and can be summarized as follows: For a given odd $N \times N$ window, each data point is associated with a position (u, v) from the set $U \times U$ where

$$U = \{-(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2\}$$

The local biquadratic surface fitting capability is provided using the following discrete orthogonal polynomials:

$\emptyset_0(u)=1,\ \emptyset_1(u)=u,\ \emptyset_2(u)=(u^2 - M(M+1)/3);\ M=(N-1)/2$

To estimate the first and second partial derivatives, an orthogonal set of $d_i(u)$ functions using the normalized versions of the orthogonal polynomials $\emptyset_i(u)$ is used:

$$\vec{d}_i(u) = \frac{\emptyset_i(u)}{P_i(M)} : P_0(M) = N,\ P_1(M) = \frac{2}{3}M^3 + M^2 + \frac{1}{3}M.$$

$$P_2(M) = \frac{8}{45}M^5 + \frac{4}{9}M^4 + \frac{2}{9}M^2 - \frac{1}{9}M^2 - \frac{1}{15}M.$$

Since the discrete orthogonal quadratic polynomials over the 2D window are separable in u and v, partial derivative estimates can be computed using separable convolution operators. These derivatives estimates can then be plugged into the equation for mean curvature. The equally weighted least squares derivative estimation window operators are then given by:

$$[D_u] = \vec{d}_0 \vec{d}_1^T,\ [D_v] = \vec{d}_1 \vec{d}_0^T,$$

$$[D_{uu}] = \vec{d}_0 \vec{d}_2^T,\ [D_{vv}] = \vec{d}_2 \vec{d}_0^T,\ [D_{uv}] = \vec{d}_1 \vec{d}_1^T$$

$\tilde{g}(i,j)$ represents the noisy, quantized discretely sampled version of a piecewise-smooth graph surface. Afterwards, the partial derivative estimate images are computed via appropriate 2D image convolutions.

$$\tilde{g}_u(i,j) = D_u \otimes \tilde{g}(i,j), \tilde{g}_v(i,j) = D_v \otimes \tilde{g}(i,j),$$
$$\tilde{g}_{uu}(i,j) = D_{uu} \otimes \tilde{g}(i,j), \tilde{g}_{uv}(i,j) = D_{vv} \otimes \tilde{g}(i,j),$$
$$\tilde{g}_{uv}(i,j) = D_{uv} \otimes \tilde{g}(i,j)$$

The mean curvature is then computed using the partial derivatives estimates as the following:

$$H(i,j) = \frac{(1+\tilde{g}_v^2(i,j))\,\tilde{g}_{uu}(i,j) + (1+\tilde{g}_u^2(i,j))\tilde{g}_{vv}(i,j) - 2\tilde{g}_u(i,j)\,\tilde{g}_v(i,j)\tilde{g}_{uv}(i,j)}{2(\sqrt{1+\tilde{g}_u^2(i,j)+\tilde{g}_v^2(i,j)})^3}$$

2.3. Mean Curvature Analysis

The behaviour of the mean curvature for specific object properties can be demonstrated well by the filter results for synthetic range images. Thus the mean curvature was computed for range images of a block and a wye, which are depicted in Figure 2. The curvature values were then extracted at the horizontal profile represented by the line overlaid to the respective range image. From the analysis of these curvature values as they are depicted in the bottom of Figure 2 one can conclude that:

a) For jump edge boundaries (J) where surface depths are discontinuous, the mean curvature exhibits a zero crossing. Two distinct peaks of opposite algebraic sign are clearly visible in the profile of computed curvature values.
b) For crease edges (C) at discontinuities in the surface normal direction, the curvature response is a smooth peak. Concave (Cv) and convex (Cx) edges can be discriminated by the different algebraic sign of the curvature values. The exact position of a convex crease edge is defined by the maximum curvature value, while the concave crease edge is given at a minimum.

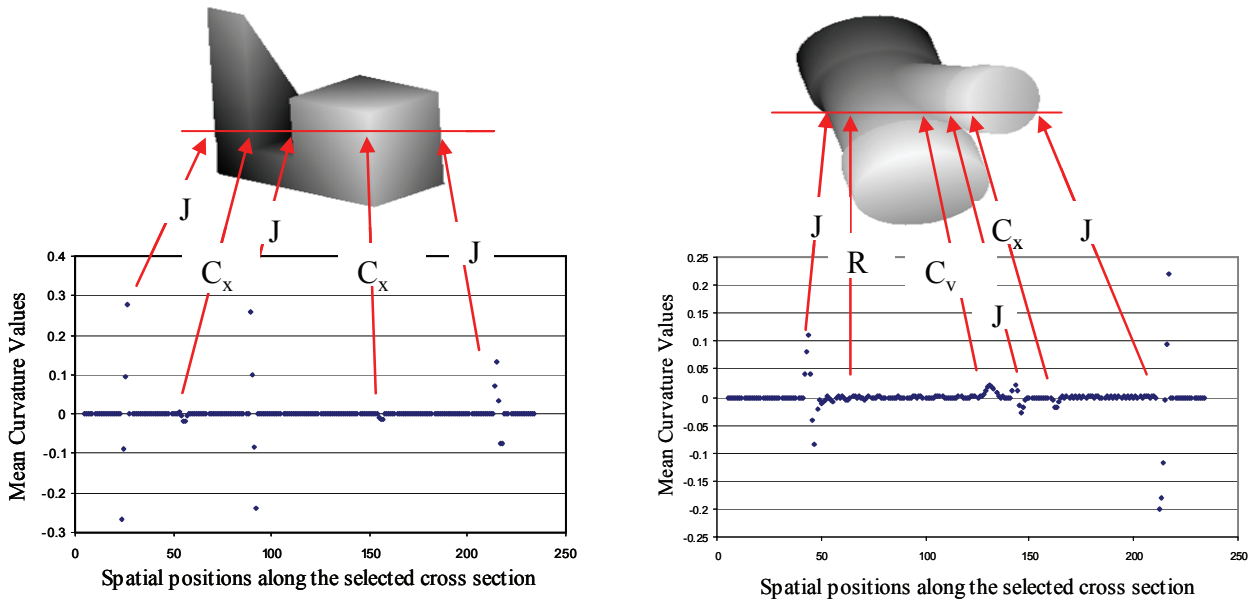

Figure 2: *Spatial distribution of the mean curvature values for block and wye range image*

c) At ridges (R) the mean curvature also indicates a change in the orientation of the surface normal, however, the response is smaller compared to crease edges.
d) Compared to crease edges, the values of the mean curvature are larger at jump edges. Their value mainly depends on the magnitude of the depth discontinuity.
e) For jump edges, the exact position is defined at a zero crossing between two peaks of the mean curvatures, whereas for both crease and ridge edges the true edge is given by the maximum and minimum value of the peaks.

After computation of the mean curvature values $H(x,y)$ a pixel represents an edge location $\{(x,y): E(x,y)=1\}$ if the value of the gradient exceeds some threshold. Thus:

$$E(x,y)=\begin{cases}1 \text{ if } \| H \| > T \text{ for some threshold } T \\ 0 \text{ otherwise}\end{cases}$$

In order to locate the position of crease, ridge and step edges, zero crossings as well as smooth peak values are searched within the computed mean curvature values. Since the value of the mean curvature is smaller for crease edges than for jump edges, the edge types of an object can be classified easily by applying different threshold values. Low threshold values are used to detect the small peaks of crease edges while larger values can be used for step edge detection. This ability of our algorithm to reliably characterize these edge types is demonstrated exemplarily in Figure 3.

Figure 3: Curve block segmentation using different thresholds to detect step edges (red) and crease edges (blue).

Figure 4 displays the segmentation results for two range data sets for the 3D model of Al-Kahsneh monument in Petra, Jordan. The data were collected by a Mensi GS100 laser scanner [AH05]. As already mentioned, the segmentation is based on range images, which, in contrast to unordered 3D point clouds, maintain the original topology of a laser scan and thus allow an easier implementation of the segmentation algorithms. The top row of Figure 4 depicts the outer façade of Al-Kahsneh, while the bottom row shows data collected for one of the interior rooms. The second column, referred as Figure 4b shows the binary edge maps as they are generated using the curvature based segmentation. As it is visible, most of the main features are detected. Since a large filter mask size was used, the edges are rather blurred. For this reason, an edge thinning is applied. Figure 4c depicts the resulting skeletons overlaid to the original range image.

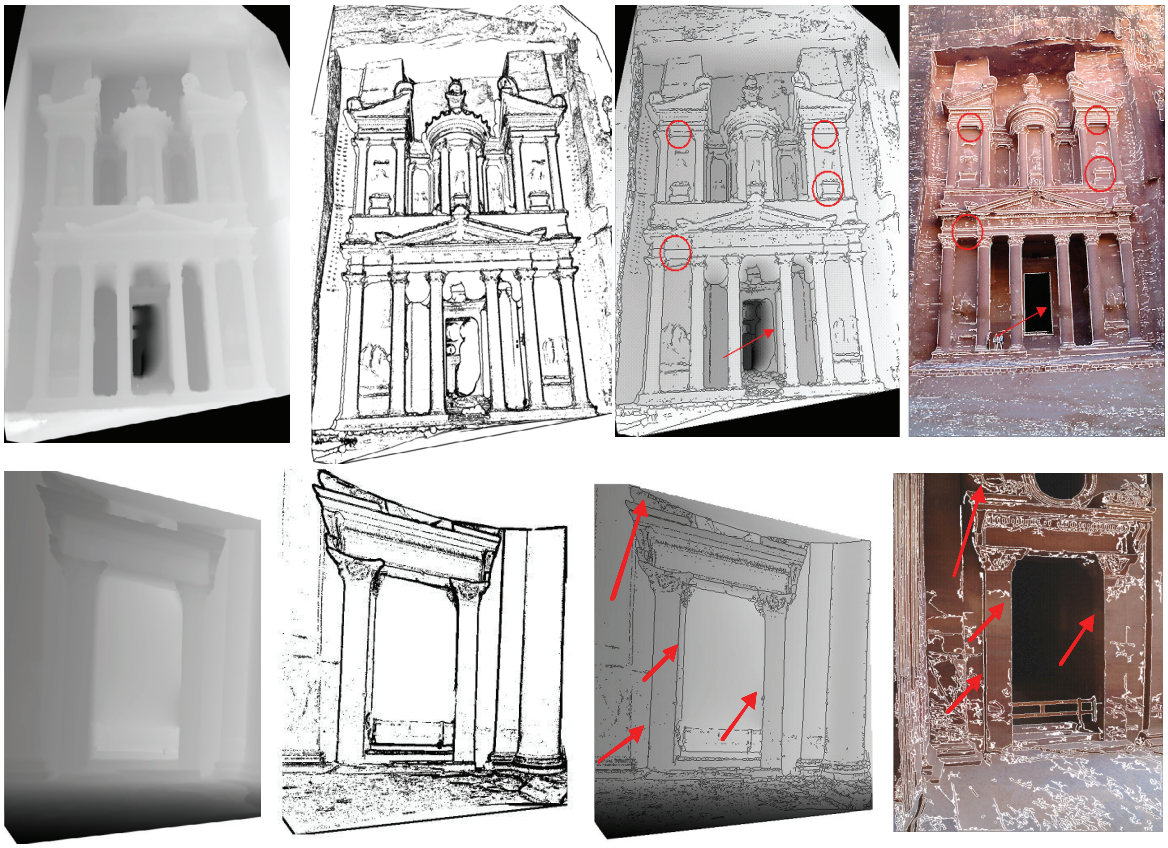

Figure 4: *a) Range images for 3D model of Al-Khasneh. b) Binary range edge images. c) Segmentation results after thinning overlaid to corresponding range image. d) Segmentation results of colored image. Red arrows and circles show selected edges used for registration of the 2D-3D data sets.*

As it is depicted in Figure 4d our segmentation is not limited to range images, but can also be applied for feature extraction from intensity images. In addition to the segmentation results, Figure 4c and Figure 4d also show some manually selected corresponding lines, which were used to register both colour and range images based on an algorithm developed by [KF03]. Thus, the linear features as they are extracted by our segmentation process allow for a precise and reliable co-registration of range and image data sets.

3. Multi-image texturing of complex 3D scenes

After this registration, texture mapping can be realized by warping the images onto the collected object surfaces. For this purpose, the transformation parameters as determined in the preceding step are used. Additionally, it has to be guaranteed that the image is only mapped onto parts of the 3D model that are actually visible from the respective camera viewpoints. Thus, a occlusion detection is additionally required before texture mapping. While such a visibility checking within the texture images can be realised without problems for relatively simple object representations, it can become very tedious and time consuming within the context of cultural heritage documentation. As an example, for documentation of the Al-Kasneh monument shown in Figure 4, laser scans were collected for three different viewpoints. After registration and meshing, this resulted in a 3D model at an average resolution of 5 cm with more than 2 million triangles. During texture mapping in principle every triangle node has to be mapped from object space to the corresponding point on the texture image and the visibility of each mesh in the respective images has to be checked.

Efficient approaches to solve the visibility problem have already been developed in computer graphics. There occlusion detection is required within the virtual images during scene rendering. These approaches can be modified for efficient texture extraction and placement in the context of cultural heritage representation. As an example, the algorithm proposed by (Grammatikopoulos L. et al 2005) is based on z-buffering, which computes the visibility of each pixel by comparing the surface depths at each position in image space and selects the one closer to the observer. (Alshawabkeh & Haala 2005) combine such an approach with techniques originating from the painter's algorithm. In

this approach, also well known in computer graphics for occlusion detection, the model polygons are sorted by their depth in object space with respect to the collected texture image. This algorithm efficiently detects ambient, back-face and view frustum occlusions based on computations both in image and object space.

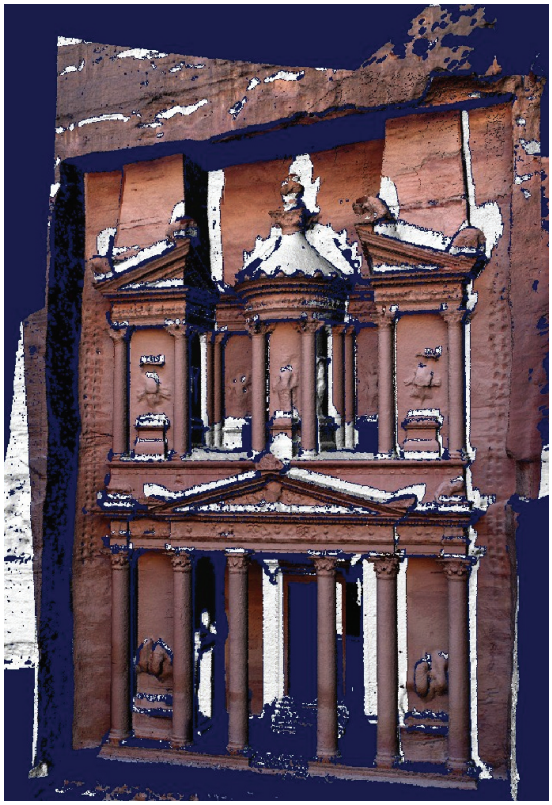

Figure 5: *Result of texture mapping with occlusions detection.*

An intermediate result of this texture mapping process for the meshed 3D model of the Alkasneh monument is depicted in Figure 5**.** In this example, the image already shown in Figure 4d is used to provide the required surface texture. In Figure 5 the parts of the 3D model, which are not visible in this texture image are marked as grey regions. These regions are separated from the original meshed 3D model and then used as input surfaces for a second texture image. The process for visibility checking and model separation is repeated until the 3D model is textured using all available images. Finally, the different parts are merged again to provide the overall model.

This process is demonstrated additionally in Figure 6 to Figure 9 for a data set showing a Roman Theatre in Jerash, Jordan. For this project, laser data was collected from 6 stations by a Mensi GS100 scanner, resulting in a 3D model consisting of 4.6 million triangles. Figure 6 shows the digital image, which was used for the texture mapping depicted in Figure 7. Similar to Figure 5**,** the parts of the 3D model, which are not visible, are marked in grey.

Figure 6: First texture image for North theatre –Jerash

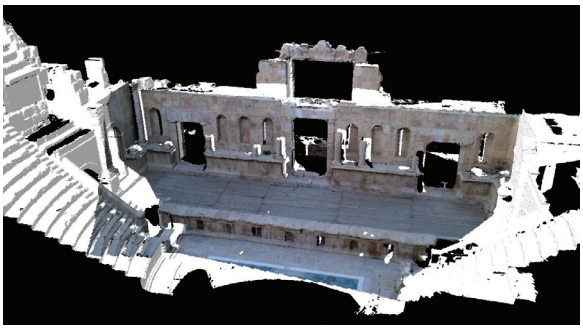

Figure 7: 3D model textured by first image after occlusion detection

As it is shown in Figure 8 these parts are then extracted from the original model and used for further processing. In Figure 8, the image depicted in Figure 9 has already been used for this purpose. After this step, those meshes of the 3D model, which are not visible in the image are again extracted and have to be textured from another available image in subsequent steps. Finally, the separated parts are merged again to generate the overall model depicted in Figure 11**.**

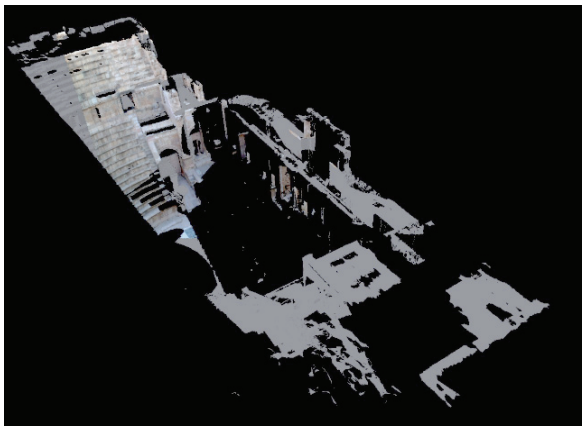

Figure 8: Occluded parts in first image as extracted from the 3D model after texture mapping from second image.

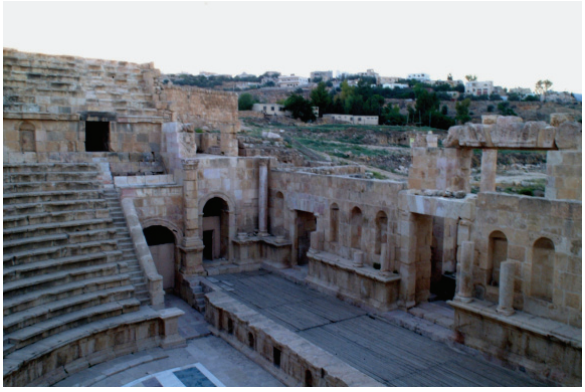

Figure 9: *Second texture image.*

Figure 10: *Additional parts of textured 3D model.*

Additional parts of the 3D model, which have been textured and separated by different images, are depicted in Figure 10. In addition to the correct geometric processing of the images, which is realised by the co-registration process discussed in section 2, homogenous radiometry between the different texture images has to be guaranteed to allow for realistic visualisation of the collected 3D model. Optionally, artefacts due to illumination changes have to be removed in an additional pre-processing step. In our application an off-the-shelf remote sensing image processing package was used for histogram equalization and stitching of the texture images. The final texture model is depicted in Figure 11.

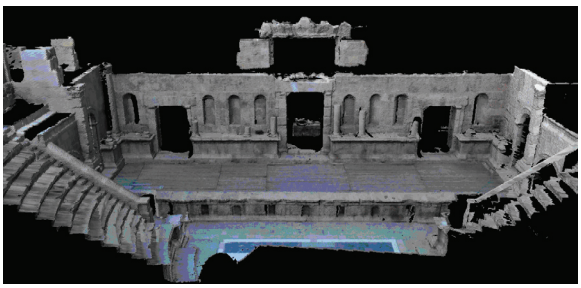

Figure 11: *3D textured model using 4 images after colour correction.*

4. Conclusion

Especially for the documentation of complex terrestrial scenes, the 3D geometric model of real world objects has to be enhanced with texture as provided from separate sets of photographs. Thus, the generation of 3D virtual models in the context of cultural heritage documentation frequently requires the combination of terrestrial LIDAR and image data in order to optimize their geometric accuracy and the visual quality.

The combination of the different data sets has to be realized by an exact co-registration, which is implies a 2D-3D pose estimation algorithm. The most common methods for solving such registration problems between two datasets are based on the identification of corresponding points. However, such methods usually are not applicable when dealing with surfaces derived from 3D point clouds. This type of data is derived from laser footprints with limited scanning resolution rather than distinct points that could be identified in the imagery. Thus, within point clouds and range images the perception of objects structure is limited and not very appropriate for registration.

To allow for a precise and reliable co-registration of the data sets, linear features were extracted by a suitable segmentation process. Since these lines possess a considerable amount of semantic information, the correspondence problem between the image and object space can then be solved easier. Currently, the segmented lines are used as input for a following manual selection of correspondences, while future work will aim on an automatic matching process.

The final goal of our work is the generation of photo-realistic models of complex shaped heritage sites with optimal efficiency. In addition, to the co-registration process, this is supported by a fast algorithm to verify the visibility of the available texture images. While these algorithms are currently used in a post-processing mode, an easy visibility detection of 3D models and 2D images should also be beneficial to guarantee the completeness and sufficient coverage during data collection.

References

[AH05] ALSHAWABKEH, Y. & HAALA N.: Automatic Multi-Image Photo Texturing of Complex 3D Scenes. *CIPA IAPRS Vol. 34-5/C34* (2005), pp. 68-73.

[BJ88] BESL, P. & JAIN, R. Segmentation Through Variable-Order Surface Fitting: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2), (1998) pp. 167–192.

[CH96] CHRISTY, S. & HORAUD, R.: Iterative Pose Computation from Line Correspondences. *Computer Vision and Image Understanding* Vol. 73, No. 1, January, (1996) pp. 137–144.

[GKK*05] GRAMMATIKOPOULOS L., KALISPERAKIS I., KARRAS G. & PETSA E.: Data Fusion from Multiple Sources for the production of Orthographic and Perspective Views with Automatic Visibility Checking. *CIPA IAPRS Vol. 34-5/C34* (2005)

[HJJ*96], HOOVER, A., JEAN-BAPTISTE, G., JIANG, X, FLYNN,P. & BUNKE, H.: An Experimental Comparison of Range Image Segmentation Algorithms. IEEE *Transactions on Pattern Analysis and Machine Intelligence,* 18(7) (1996) pp. 673-689.

[KF03] KLINEC, D. & FRITSCH, D.: Towards pedestrian navigation and orientation. *Proceedings of the 7th South East Asian Survey Congress: SEASC'03*, Hong Kong, (2003).

[LH02] LEE, Y. & HABIB, A.: Pose Estimation of Line Cameras Using Linear Features, *ISPRS Symposium of PCV'02 Photogrammetric Computer Vision,* Graz, Austria (2002).

[LHS00] LENSCH, H., HEIDRICH, W. & SEIDEL, H.: Automated texture registration and stitching for real world models. In *Proceedings of Pacific Graphics* (2000), pp. 317–326.

[LS05] LIU, L. & STAMOS, S.: Automatic 3D to 2D registration for the photorealistic rendering of urban scenes. *Computer Vision and Pattern Recognition. IEEE Computer Society.* (2005)

[MLM01] MARSHALL, D., LUKACS, G. & MARTIN, R.: Robust segmentation of primitives from range data in the presence of geometric degeneracy. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23(3), (2001), pp. 304-314.

[SD01] SAPPA, A. & DEVY, M.: Fast Range Image Segmentation by an Edge Detection Strategy: *Proc. IEEE Conf. 3D Digital Imaging and Modelling*, (2001), pp. 292-299.