AN AUTOMATIC METHOD FOR ACQUIRING 3D MODELS FROM PHOTOGRAPHS: APPLICATIONS TO AN ARCHAEOLOGICAL SITE

POLLEFEYS M., KOCH R., VERGAUWEN M., VAN GOOL L.
ESAT-PSI, K.U.Leuven
E-mail: Marc.Pollefeys@esat.kuleuven.ac.be

KEY WORDS: 3D reconstruction, 3D measurements, calibration, photogrammetry, archaeology

ABSTRACT:

In this paper a system is presented which automatically extracts a 3D surface model from a sequence of photographs of a scene. The system can deal with unknown camerasettings. In addition the parameters of this camera are allowed to change during acquisition (e.g. by zooming or focussing). No prior knowledge about the scene is necessary to build the 3D models. This system therefore offers a high degree of flexibility. The system is based on state-of-the-art algorithms recently developed in computer vision. The 3D modelling task is decomposed into a number of succesive steps. Gradually more knowledge of the scene and the camera setup is retrieved. This system has been applied to a number of applications in archaeology. The Roman site of Sagalassos (south-west Turkey) was used as a test case to illustrate the potential of this new approach. Besides the construction of a virtual site consisting of different level of details, some more applications to archaeology and conservation of heritage sites are presented.

1 INTRODUCTION

In the last few years the interest in 3D models has dramatically increased. More and more applications are using computer generated models. The main difficulty lies with the model acquisition. Although more tools are at hand to ease the generation of models, it is still a time consuming and expensive process. In many cases models of existing scenes or objects are desired. Traditional solutions include the use of stereo rigs, laser range scanners and other 3D digitizing devices. These devices are often very expensive, require careful handling and complex calibration procedures and are designed for a restricted depth range only.

In this work an alternative approach is proposed which avoids most of the problems mentioned above. The scene which has to be modeled is photographed from different viewpoints. The relative position and orientation of the camera and its calibration parameters will automatically be retrieved from image data. Hence, there is no need for measurements in the scene or calibration procedures whatsoever. There is also no restriction on range, it is just as easy to model a small object (using a macro lens), as to model a complete building or even a whole landscape. The proposed method thus offers a previously unknown flexibility in 3D model acquisition. In addition, no more than a photo camera is needed for scene acquisition. Hence, increased flexibility is accompanied by a decrease in cost.

This flexibility opens the way to new applications. Scenes can be reconstructed from a sequence of photographs. Models can be generated from archive images (e.g. from monuments destroyed during the war). It becomes possible to generate realistic 3D models of complete sites (e.g. archeological sites). Besides this, 3D modeling of objects (e.g. for tele-shopping applications or virtual exhibitions) is eased a lot.

The paper is organized as follows. Section 2 describes the approach that is followed for the acquisition of 3D models from photographs. The subsequent sections describes different applications in the field of archaeology.

2 MODEL ACQUISITION

Two things are needed to build a 3D model from an image sequence: (1) the calibration of the camera setup[1] and (2) the correspondences between the images. Starting from an image sequence acquired by an uncalibrated photo or video camera, both these prerequisites are unknown and therefore have to be retrieved from image data. At least a few correspondences are needed to retrieve the calibration of the camera setup, but on the other hand this calibration facilitates the search for correspondences a lot.

In Figure 1 an overview of the systems is given. It consists of independent modules which pass on the necessary information to the next modules. The first module computes the projective calibration of the sequence together with a sparse reconstruction. In the next module the metric calibration is computed from the projective camera matrices through self-calibration. Then dense correspondence maps are estimated. Finally, all results are integrated in a tex-

---

[1] By *calibration* we mean the actual internal calibration of the camera as well as the relative position and orientation of the camera for the different views.

tured 3D surface reconstruction of the scene under consideration. A more detailed description of this system can be found in (Pollefeys, 1999).

2.1 Retrieving the Projective Framework

The first correspondences are found by extracting intensity corners in different images and matching them using a robust tracking algorithm. In conjunction with the matching of the corners the projective calibration of the setup is calculated. This allows to eliminate matches which are inconsistent with the calibration. Using the projective calibration more matches can easily be found and used to refine this calibration.

At first corresponding corners in two consecutive images are matched. This defines a projective framework in which the projection matrices of the other views are retrieved one by one (Beardsley et. al., 1996). We therefore obtain projection matrices (3 × 4) of the following form:

$$\mathbf{P}_1 = [\mathbf{I}|0] \text{ and } \mathbf{P}_k = [\mathbf{H}_{1k}|e_{1k}] \tag{1}$$

with $\mathbf{H}_{1k}$ the homography for an arbitrary reference plane from view 1 to view k and e_{1k} the corresponding epipole.

2.2 Retrieving the Metric Framework

Such a projective calibration is certainly not satisfactory for the purpose of 3D modeling. A reconstruction obtained up to a projective transformation can differ very much from the original scene according to human perception: orthogonality and parallellism are in general not preserved, part of the scene can be warped to infinity, etc. To obtain a more complete calibration, constraints can be obtained by imposing some restrictions on the internal camera parameters (e.g. square pixels). By exploiting these constraints, the projective reconstruction can be upgraded to metric (Euclidean up to scale) (Pollefeys et. al., 1998).

2.3 Dense Correspondences

At this point we dispose of a sparse metric reconstruction. Only a few salient points are reconstructed. Obtaining a dense reconstruction could be achieved by interpolation, but in practice this does not yield satisfactory results. Often some salient features are missed during the corner matching and will therefore not appear in the reconstruction. If for example the corner of the roof is missing, this could result in a whole part of the roof missing when using interpolation.

These problems can be avoided by using algorithms which estimate correspondences for almost every point in the images. At this point algorithms can be used which were developed for calibrated 3D systems like stereo rigs. Since we have computed the projective calibration between successive image pairs we can exploit the epipolar constraint that restricts the correspondence search to a 1-d search range. In particular it is possible to remap the image pair to standard geometry where the epipolar lines coincide with

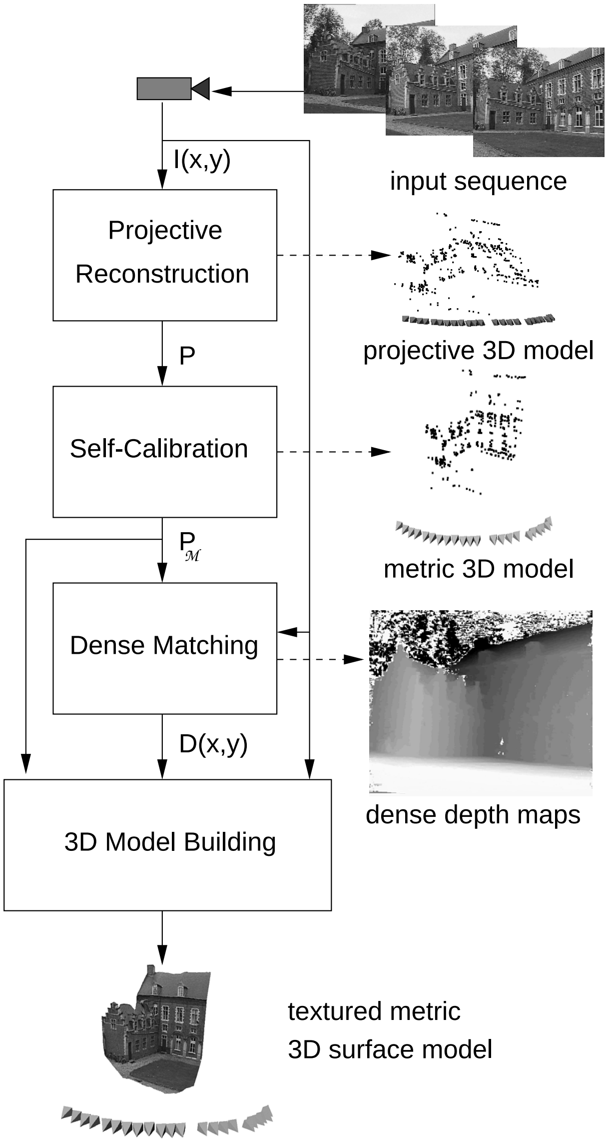

Figure 1: Overview of the system: from the image sequence ($I(x, y)$) the projective reconstruction is computed; the projection matrices P are then passed on to the self-calibration module which delivers a metric calibration $P_{\mathcal{M}}$; the next module uses these to compute dense depth maps $D(x, y)$; all these results are assembled in the last module to yield a textured 3D surface model. On the right side the results of the different modules are shown: the preliminary reconstructions (both projective and metric) are represented by point clouds, the cameras are represented by little pyramids, the results of the dense matching are accumulated in dense depth maps (light means close and dark means far).

Figure 2: Image sequence which was used to build a 3D model of the corner of the Roman baths

the image scan lines. The correspondence search is then reduced to a matching of the image points along each image scanline. In addition to the epipolar geometry other constraints like preserving the order of neighboring pixels, bidirectional uniqueness of the match, and detection of occlusions can be exploited. These constraints are used to guide the correspondence towards the most probable scanline match using a dynamic programming scheme (Falkenhagen, 1997). The most recent algorithm (Koch et. al.,1998) improves the accuracy by using a multibaseline approach.

2.4 Building the Model

Once a dense correspondence map and the metric camera parameters have been estimated, dense depth maps are computed using depth triangulation. The 3D model surface is constructed of triangular surface patches with the vertices storing the surface geometry and the faces holding the projected image color in texture maps. The texture maps add very much to the visual appearance of the models and augment missing surface detail.

The model building process is at present restricted to partial models computed from single view points and work remains to be done to fuse different view points. Since all the views are registered into one metric framework it is possible to fuse the depth estimate into one consistent model surface (Koch, 1996).

Sometimes it is not possible to obtain a single metric framework for large objects like buildings since one may not be able to record images continuously around it. In that case the different frameworks have to be registered to each other. This will be done using available surface registration schemes (Chen and Medioni, 1991).

3 VIRTUALIZING SCENES FROM IMAGES

The 3D surface acquisition technique that was presented in the previous section, can readily be applied to archaeological sites. The on-site acquisition procedure consists of recording an image sequence of the scene that one desires to *virtualize*. To allow the algorithms to yield good results viewpoint changes between consecutive images should not exceed 5 to 10 degrees. An example of such a sequence is given in Figure 2. The further processing is fully automatic. The result for the image sequence under consideration can be seen in Figure 3. An important advantage is that details like missing stones, not perfectly planar walls or symmetric structures are preserved. In addition the surface texture is directly extracted from the images. This does not only result in a much higher degree of realism, but is also important for the authenticity of the reconstruction. Therefore the reconstructions obtained with this system can also be used as a scale model on which measurements can be carried out or as a tool for planning restaurations.

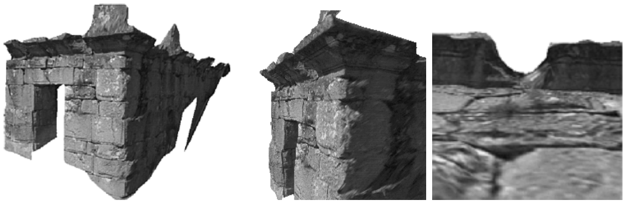

Figure 3: Virtualized corner of the Roman baths, on the right some details are shown

4 VIRTUALIZING A WHOLE SITE

A first approach to obtain a virtual reality model for a whole site consists of taking a few overview photographs from the distance. Since our technique is independent of scale this yields an overview model of the whole site. The only difference is the distance needed between two camera poses. An example of the results obtained for Sagalassos are shown in Figure 4. The model was created from 9 images taken from a hillside near the excavation site. Note that it is straightforward to extract a digital terrain map or orthophotos from the global reconstruction of the site.

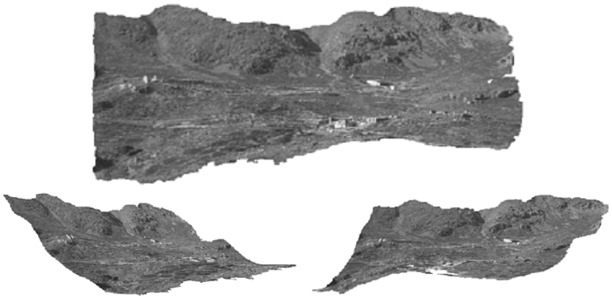

Figure 4: Overview model of Sagalassos

4.1 Integration of models at different scales

The problem is that this kind of overview model is too coarse to be used for realistic walk-throughs around the site or for looking at specific monuments. Therefore it is necessary to integrate more detailed models into this overview model. This can be done by taking additional image sequences for all the interesting areas on the site. These are used to generate reconstructions of the site at different scales, going from a global reconstruction of the whole site to a detailed reconstruction for every monument.

These reconstructions thus naturally fill in the different levels of details which should be provided for optimal rendering. In Figure 5 reconstructions of the Roman baths are given for three different levels of details (site overview, complete Roman bath house and detail of right corner).

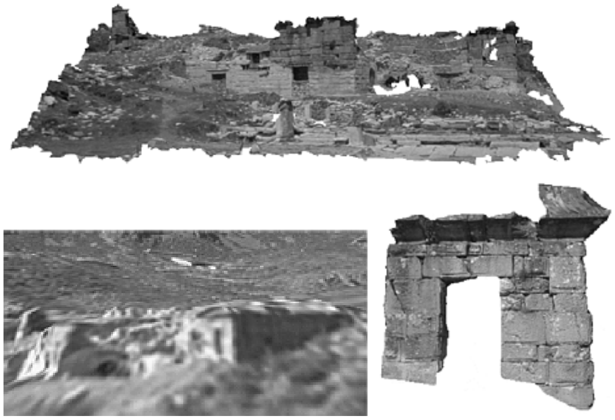

Figure 5: Models of the Roman baths at different scales: complete baths (top), zoom onto the baths in the overview model of Figure 4 (bottom left), detailed right corner of the baths (bottom right)

4.2 Combination with other models

An interesting possibility is the combination of these models with other type of models. In the case of Sagalassos some building hypothesis were translated to CAD models. These were integrated with our models. The result can be seen in Figure 6. Also other models obtained with different 3D acquisition techniques could easily be integrated.

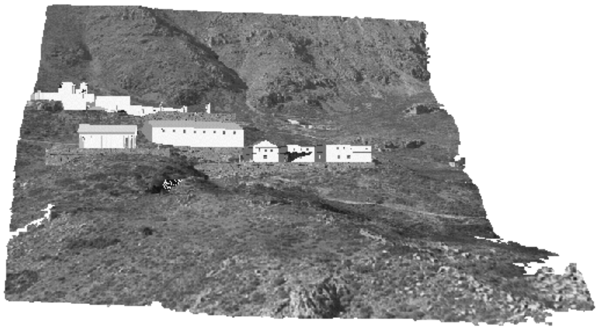

Figure 6: Virtualized landscape of Sagalassos combined with CAD-models of reconstructed monuments

5 OTHER APPLICATIONS

Since these reconstructions are almost completely automatic and the on-site acquisition time is very short, several new applications come to mind. In this section a few possibilities are illustrated.

5.1 3D Stratigraphy

Archaeology is one of the sciences were annotations and precise documentation are most important because evidence is destroyed during work. An important aspect of this is the stratigraphy. This reflects the different layers of soil that corresponds to different time periods in an excavated sector. Due to practical limitations this stratigraphy is often only recorded for some slices, not for the whole sector.

Our technique allows a more optimal approach. For every layer a complete 3D model of the excavated sector can be generated. Since this only involves taking a series of pictures this does not slow down the progress of the archaeological work. In addition it is possible to model separately artifacts which are found in these layers and to include the models in the final 3D stratigraphy.

This concept is illustrated in Figure 7. The excavations of an ancient Roman villa at Sagalassos were recorded with our technique. In the figure a view of the 3D model of the excavation is provided for two different layers.

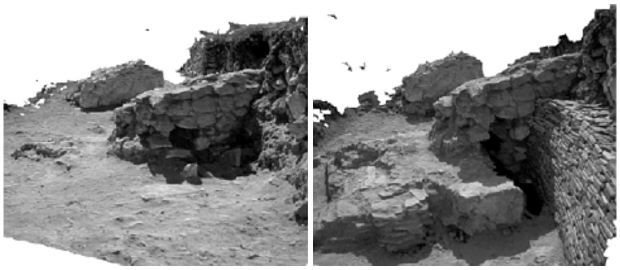

Figure 7: 3D stratigraphy, the excavation of a Roman villa at two different moments.

5.2 Generating and testing building hypothesis

The technique proposed in this paper also has a lot to offer for generating and testing building hypothesis. Due to the ease of acquisition and the obtained level of detail, one could reconstruct every building block separately. The different construction hypothesis can then interactively be verified on a virtual building site. Some testing could even be automated. The matching of the two parts of Figure 8 for example could be verified through a standard registration algorithm (Chen and Medioni, 1991). An automatic procedure can be important when dozens of broken parts have to be matched against each other.

5.3 Reconstruction from archives

Here the reconstruction of the ancient theater of Sagalassos is shown. The reconstruction is based on a sequence filmed by a cameraman from the BRTN (Belgische Radio en Televisie van de Nederlandstalige gemeenschap) in 1990. The sequence was filmed to illustrate a TV program about Sagalassos. Because of the motion only fields –and not frames– could be used. The resolution of the images we could use was thus restricted to 768×288. The sequence consisted of about hundred images, three of them are shown in Figure 9. We recorded approximately 3 images per second.

In Figure 10 the reconstruction of interest points and cameras is given. This shows that the approach can deal with long image sequences.

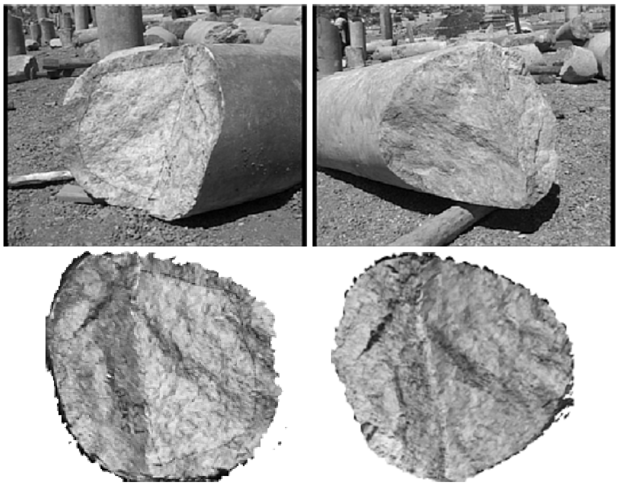

Figure 8: Two images of parts of broken pillars (top) and two orthographic views of the matching surfaces generated from the 3D models (bottom)

Figure 9: This sequence was filmed from a helicopter in 1990 by a cameraman of the BRT (Belgische Radio en Televisie) to illustrate a TV program on Sagalassos (an archaeological site in Turkey).

Dense depth maps were generated from this sequence and a dense textured 3D surface model was constructed from this. Some views of this model are given in Figure 11.

6 CONCLUSION

An automatic 3D scene modelling technique was discussed that is capable of building models from uncalibrated image sequences. The technique is able to extract metric 3D models without any prior knowledge about the scene or the camera. The calibration is obtained by assuming a rigid scene and some constraints on the intrinsic camera parameters (e.g. square pixels).

This technique was successfully applied to the acquisition of virtual models of archaeological sites. The advantages are numerous: the on-site acquisition time is restricted, the construction of the models is automatic and the generated models are realistic. The technique allows some more promising applications like 3D stratigraphy, the (automatic) generation and testing of building hypothesis and the virtual reconstruction of (destroyed) monuments from archive images.

ACKNOWLEDGMENTS

We would like to thank prof. Marc Waelkens from the Sagalassos Archaeological Research Project (K.U.Leuven) for inviting us to Sagalassos and for his collaboration, Pol and Jos Legrand for the CAD models and Joris Vanden Wyngaerd for the integration of the CAD models with our models. Financial support from the EU ESPRIT project LTR 23515 IMPROOFS is also gratefully acknowledged.

Figure 10: The reconstructed interest points and camera poses recovered from the BRT sequence.

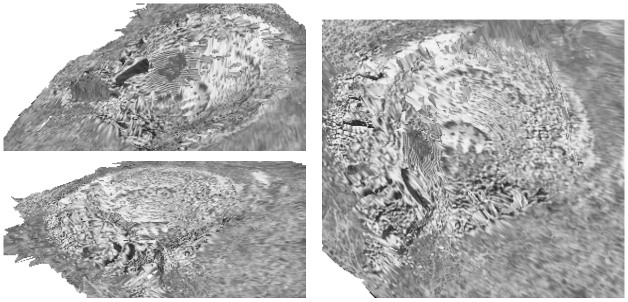

Figure 11: Some views of the reconstructed model of the ancient theater of Sagalassos.

REFERENCES

P. Beardsley, P. Torr and A. Zisserman, "3D Model Acquisition from Extended Image Sequences", *Proc. European Conference on Computer Vision*, 1996.

Y. Chen and G. Medioni, Object Modeling by Registration of Multiple Range Images, IEEE. International Conference on Robotics and Automation, pp 2724-2729, 1991.

L. Falkenhagen, "Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints". International Workshop on SNHC and 3D Imaging, 1997.

R. Koch, "Automatische Oberflachenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten", PhD thesis, University of Hannover, 1996.

R. Koch, M. Pollefeys and L. Van Gool, "Multi Viewpoint Stereo from Uncalibrated Video Sequences", *Proc. European Conference on Computer Vision*, 1998.

M. Pollefeys, R. Koch and L. Van Gool, "Self-Calibration and Metric 3D reconstruction in Spite of Varying and Unknwon Internal Parameters", *Proc. International Conference on Computer Vision*, 1998.

M. Pollefeys, *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*, Ph.D. thesis, K.U.Leuven, 1999.