3D ACQUISITION OF ARCHAEOLOGICAL HERITAGE FROM IMAGES

Marc Pollefeys, Maarten Vergauwen, Kurt Cornelis, Frank Verbiest, Joris Schouteden, Jan Tops, Luc Van Gool
Center for Processing of Speech and Images, K.U.Leuven,
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
Tel. +32 16 321064; Fax +32 16 321723
Marc.Pollefeys@esat.kuleuven.ac.be

KEY WORDS: photogrammetry, archaelogy, heritage conservation, image-based 3D reconstruction, structure from motion, self-calibration, dense stereo.

ABSTRACT
In this contribution an approach is proposed that can capture the 3D shape and appearance of objects, monuments or sites from photographs or video. The flexibility of the approach allows us to deal with uncalibrated hand-held camera images. In addition, through the use of advanced computer vision algorithms the process is largely automated. Both these factors make the approach ideally suited to applications in archaeology. Not only does it become feasible to obtain photo-realistic virtual reconstructions of monuments and sites, but also stratigraphy layers and separate building blocks can be reconstructed. These can then be used as detailed records of the excavations or allow virtual re-assemblage of monuments. Since the motion of the camera is also computed, it also becomes possible to augment video streams of ancient remains with virtual reconstruction. The proposed approach retrieves both the structure of a scene and the motion of the camera from an image sequence. In a first step features are extracted and matched over consecutive images. This step is followed by a structure-from-motion algorithm that yields a sparse 3D reconstruction (i.e. the matched 3D features) and the path of the camera. These results are enhanced through auto-calibration and bundle adjustment. To allow a full surface reconstruction of the observed scene, the images are rectified so that a standard stereo algorithm can be used to determine dense disparity maps. By combining several of these maps, accurate depth maps are computed. These can then be integrated together to yield a dense 3D surface model. By making use of texture mapping photo-realistic models can be obtained.

1 INTRODUCTION

In the context of archaeology measurement and documentation are very important. Not only to record endangered archaeological heritage or monuments, but also to record the excavations themselves. Archaeology is one of the sciences were annotations and precise documentation are most important because evidence is destroyed during work. Recently archaeologists are also becoming aware of the possibilities of 3D visualization for dissemination to the wider public. However, up to now the cost of photo-realistic virtual reconstruction remains prohibitive for most applications.

Many different approaches are being used for obtaining 3D measurements in the field of archaeology. Some approaches are based on specialized instruments such as theodolites, total stations or laser range scanners. These instruments are often very expensive, require careful handling and complex calibration procedures and are designed for a restricted depth range only. There are also still a large number of manual measurements being made using tapes, plumb-bobs, levels, etc. Although these approaches are very flexible, only a limited number of measurements can be obtained this way.

An alternative approach consists of using images of the site. This is the approach that has traditionally been followed by photogrammetry. However, in this case the emphasis has mostly been on accuracy and not on automation. Therefore, most of the available tools require a lot of manual interaction to obtain 3D measurements and models (e.g. PhotoModeler). An alternative approach consists of starting from a preliminary 3D model that is then refined based on images (Debevec et al. 1996). Both these approaches are in practice limited to relatively simple models.

During recent years a lot of effort has been going on in the computer vision community to obtain automatic solutions for constructing 3D models from images. Early work by Tomasi and Kanade (1992) allowed capturing simple 3D models using a hand-held camera under orthographic conditions. Since then approaches have been developed that automatically retrieve detailed 3D models from sequences of images acquired with an uncalibrated camera (Pollefeys et al. 1999). This approach has been developed further over the last few years, removing some of the limitations and yielding more accurate results by incorporating techniques such as bundle adjustment. The first part of this paper presents an overview of this approach. Then, different examples and applications in the field of archaeology are discussed.

2 RELATING IMAGES

Starting from a collection of images or a video sequence the first step consists in relating the different images to each other. This is not an easy problem. A restricted number of corresponding points is sufficient to determine the geometric relationship or multi-view constraints between the images. Since not all points are equally suited for matching or tracking (e.g. a pixel in a homogeneous region), the first step consists of selecting feature points (Harris and Stephens, 1988; Shi and Tomasi, 1994). Depending on the type of image data (i.e. video or still pictures) the feature points are tracked or matched and a number of potential correspondences are obtained. From these the multi-view constraints can be computed. However, since the correspondence problem is an ill-posed problem, the set of corresponding points can be contaminated with an

important number of wrong matches or outliers. In this case, a traditional least-squares approach will fail and therefore a robust method is used (Torr, 1995; Fishler and Bolles, 1981). Once the multi-view constraints have been obtained they can be used to guide the search for additional correspondences. These can then be used to further refine the results for the multi-view constraints.

3 STRUCTURE AND MOTION RECOVERY

The relation between the views and the correspondences between the features, retrieved as explained in the previous section, will be used to retrieve the structure of the scene and the motion of the camera. The approach that is used is related to the approach proposed by Beardsley et al. (1997) but is fully projective and therefore not dependent on the quasi-Euclidean initialisation. This is achieved by strictly carrying out all measurements in the images, i.e. using reprojection errors instead of 3D errors. To support initialisation and determination of close views (independently of the actual projective frame) an image-based measure to obtain a qualitative evaluation of the distance between two views had to be used. The proposed measure is the minimum median residual for a homography between the two views. At first two images are selected and an initial projective reconstruction frame is set-up (Faugeras, 1992; Hartley et al. 1992). Then the pose of the camera for the other views is determined in this frame and for each additional view the initial reconstruction is refined and extended. In this way the pose estimation of views that have no common features with the reference views also becomes possible. Typically, a view is only matched with its predecessor in the sequence. In most cases this works fine, but in some cases (e.g. when the camera moves back and forth) it can be interesting to also relate a new view to a number of additional views. Candidate views are identified using the image-based measure mentioned above. Once the structure and motion has been determined for the whole sequence, the results can be refined through a projective bundle adjustment (Triggs et al. 2000). Then the ambiguity is restricted to metric through auto-calibration (Triggs, 1997; Pollefeys, 1999b). Finally, a metric bundle adjustment is carried out to obtain an optimal estimation of the structure and motion.

4 DENSE SURFACE RECONSRUCTION

To obtain a more detailed model of the observed surface dense matching is used. The structure and motion obtained in the previous steps can be used to constrain the correspondence search. Since the calibration between successive image pairs was computed, the epipolar constraint that restricts the correspondence search to a 1-D search range can be exploited. Image pairs are warped so that epipolar lines coincide with the image scan lines. Dealing with images acquired with a freely moving hand-held camera, it is important to use a calibration scheme that works for arbitrary motions (Pollefeys et al., 1999a). In addition, this approach guarantees minimal image sizes. The correspondence search is then reduced to a matching of the image points along each image scan-line.

In addition to the epipolar geometry other constraints like preserving the order of neighbouring pixels, bi-directional uniqueness of the match, and detection of occlusions can be exploited. These constraints are used to guide the correspondence towards the most probable scan-line match using a dynamic programming scheme (Cox et al. 1996). The matcher searches at each pixel in one image for maximum normalized cross correlation in the other image by shifting a small measurement window along the corresponding scan line. Matching ambiguities are resolved by exploiting the ordering constraint in the dynamic programming approach (Koch, 1996). The algorithm was further adapted to employ extended neighbourhood relationships and a pyramidal estimation scheme to reliably deal with very large disparity ranges of over 50% of image size (Falkenhagen, 1997). The disparity search range is limited based on the disparities that were observed for the features in the structure and motion recovery.

The pairwise disparity estimation allows computing image-to-image correspondence between adjacent rectified image pairs and independent depth estimates for each camera viewpoint. An optimal joint estimate is achieved by fusing all independent estimates into a common 3D model using a Kalman filter. The fusion can be performed in an economical way through controlled correspondence linking. This approach was discussed more in detail in (Koch et al. 1998). This approach combines the advantages of small baseline and wide baseline stereo. It can provide a very dense depth map by avoiding most occlusions. The depth resolution is increased through the combination of multiple viewpoints and large global baseline while the matching is simplified through the small local baselines.

5 BUILDING VIRTUAL MODELS

In the previous sections a dense structure and motion recovery approach was given. This yields all the necessary information to build photo-realistic virtual models. The 3D surface is approximated by a triangular mesh to reduce geometric complexity and to tailor the model to the requirements of computer graphics visualization systems. A simple approach consists of overlaying a 2D triangular mesh on top of one of the images and then build a corresponding 3D mesh by placing the vertices of the triangles in 3D space according to the values found in the corresponding depth map. The image itself is used as texture map. If no depth value is available or the confidence is too low the corresponding triangles are not reconstructed. The same happens when triangles are placed over discontinuities. This approach works well on dense depth maps obtained from multiple stereo pairs. The texture itself can also be enhanced through the multi-view linking scheme. A median or robust mean of the corresponding texture values can be computed to discard imaging artefacts like sensor noise, specular reflections and highlights.

To reconstruct more complex shapes it is necessary to combine multiple depth maps. Since all depth-maps can be located in a single metric frame, registration is not an issue. In some cases it can be sufficient to load the separate models together in

the graphics system. For more complex scenes it can be interesting to first integrate the different meshes into a single mesh. This can for example be done using the volumetric technique proposed in (Curless and Levoy, 1996). Alternatively, when the purpose is to render new views from similar viewpoints image-based approaches can be used (Koch et al. 2001). This approach avoids the difficult problem of obtaining a consistent 3D model by using view-dependent texture and geometry. This also allows taking more complex visual effects such as reflections and highlights into account.

6 EXAMPLE AND APPLICATIONS

In this section a number of different examples and applications are presented. First, it is shown how starting from a small number of photographs a detailed 3D model is obtained. Next, different applications in the field of archaeology are discussed. By combining different 3D models together with CAD models that represent archaeological hypothesis, a whole archaeological site is reconstructed. Then more specific archaeological applications are discussed: 3D recording of stratigraphy and 3D recording of broken columns to generate and verify building hypothesis.

6.1 Acquiring 3D scenes

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 reconstruct. To allow the algorithms to yield good results viewpoint changes between consecutive images should not exceed 5 to 10 degrees. The following sequence was shot in Ranakpur (India) using a standard Nikon F50 photo camera and then scanned. The sequence seen at the left of Figure 1 was processed through the method presented in this paper. The results can be seen on the right of this figure.

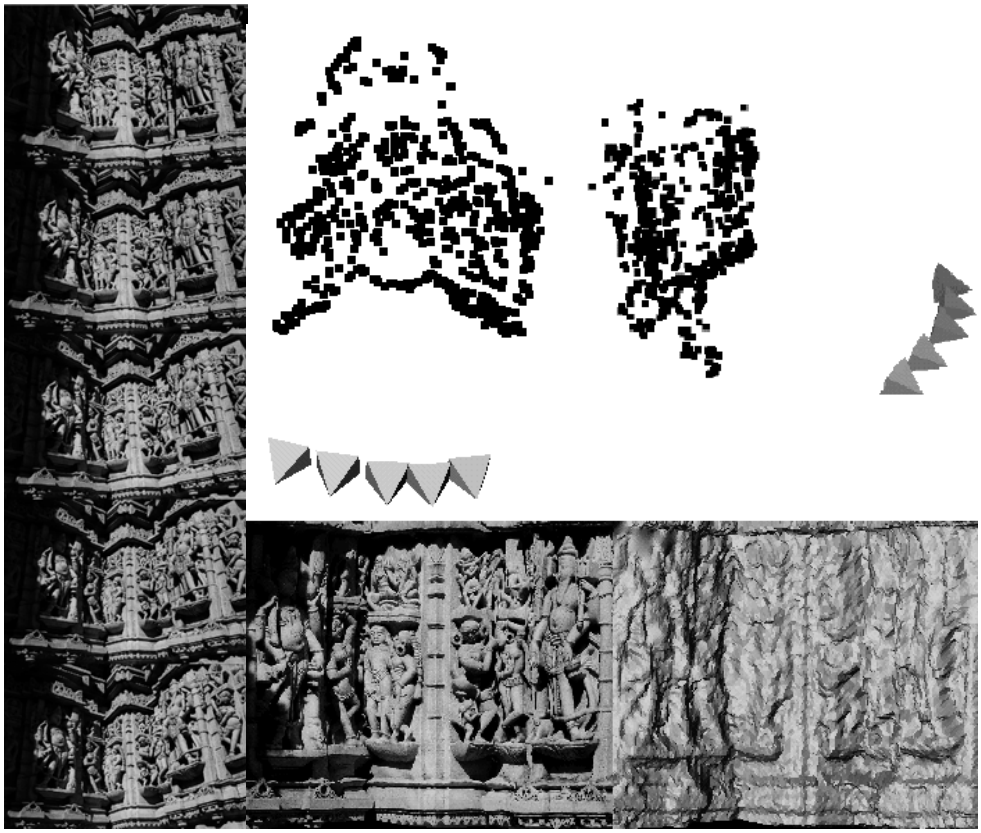

Figure 1Left: Indian temple sequence, Top-right: recovered sparse structure and motion, Bottom-right: textured and shaded view of the reconstructed 3D surface model.

A second example was recorded in Sagalassos (Turkey). This was an important Greco-roman city in ancient Pisidia. More specifically, the sequence of Figure 2 shows a corner of the Roman baths. From the 6 images at the top of the figure, the 3D model shown at the bottom was automatically computed. An important advantage is that details like missing stones, not perfectly planar walls or non-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 this high level of detail is also important in the field of archaeology and conservation, but is hard to achieve with techniques that require a lot of manual interaction.

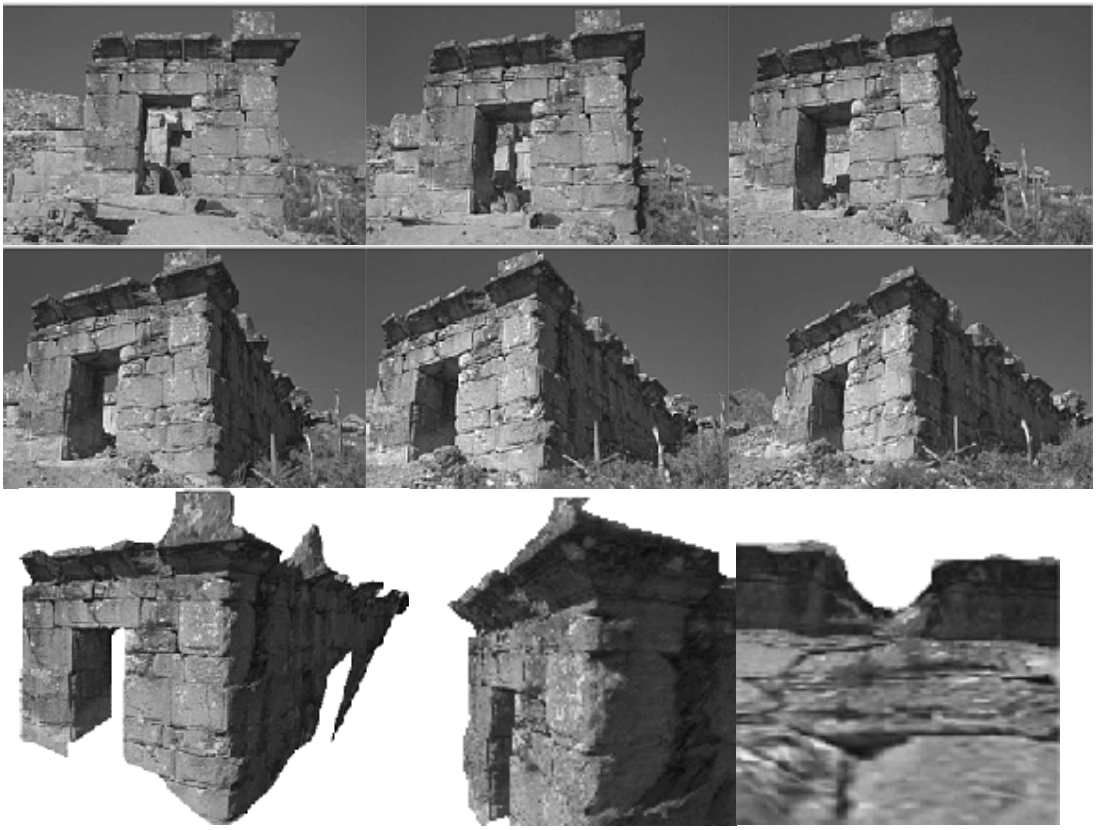

Figure 2 Corner of the Roman baths at Sagalassos. Top: 6 photographs used for modelling, Bottom: different views of the 3D reconstruction.

6.2 Building a complete 3D 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 3. The model was created from 9 images taken from a hillside near the excavation site. Using a restricted number of absolute coordinate measurements (3 or more) it is also possible to extract a digital terrain map or orthophotos from the global reconstruction of the site.

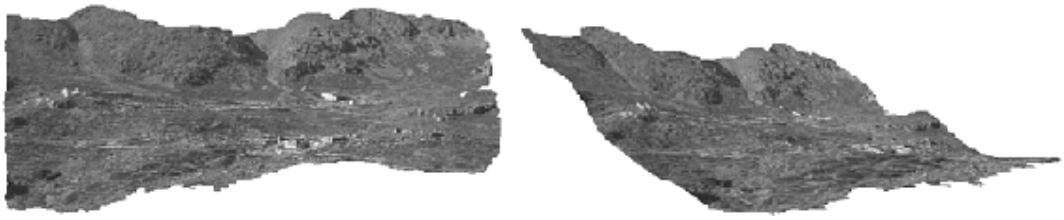

Figure 3 Overview model of Sagalassos.

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 that should be provided for optimal rendering. In Figure 4 reconstructions of the Roman baths are given for three different levels of details (site overview, complete Roman bath house and detail of right corner).

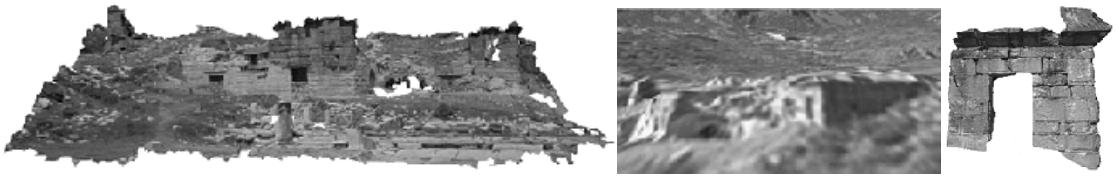

Figure 4 Models of the Roman baths at different scales. Left: complete baths, Middle: zoom onto the baths in the overview model of Figure 3, Right: detailed corner of the baths (see Figure 2).

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 5. Also other models obtained with different 3D acquisition techniques could easily be integrated.

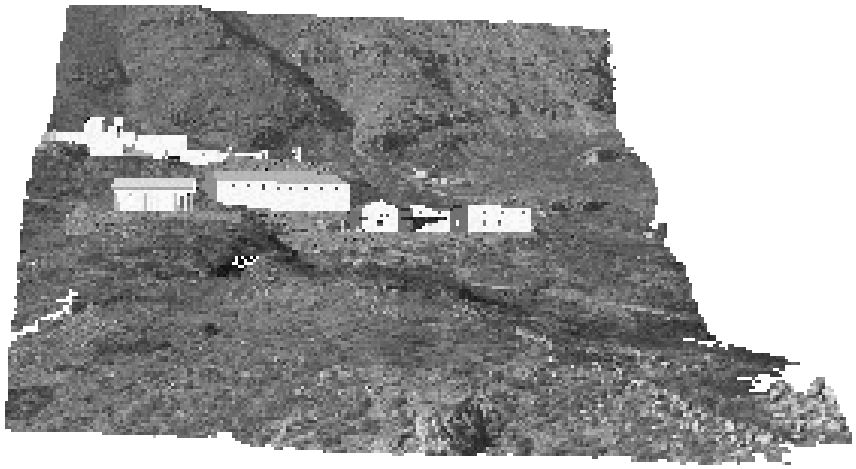

Figure 5 Virtual landscape of Sagalassos combined with CAD-models of reconstructed monuments.

6.3 Recording 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 artefacts which are found in these layers and to include the models in the final 3D stratigraphy. Having a 3D model of the excavations for the different time periods, one can start talking about 4D excavation records.

This concept is illustrated in Figure 6. The excavations of an ancient Roman villa in Sagalassos were recorded with our technique. In the figure a view of the 3D model of the excavation is provided for two different layers.

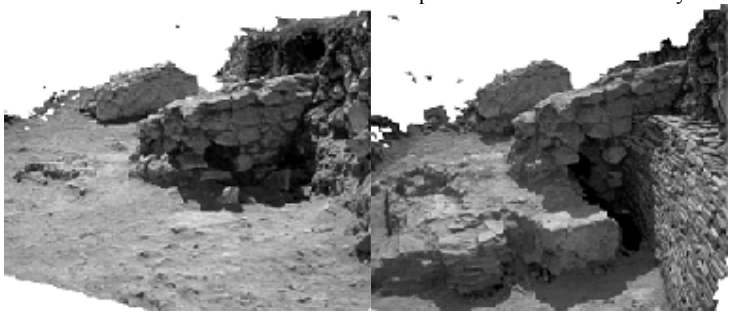

Figure 6 recording 4D stratigraphy: the excavation of a Roman villa at two different moments.

6.4 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 7 for example could be verified through a standard registration algorithm (Chen and Medioni 1992). An automatic procedure can be important when dozens of broken parts have to be matched against each other.

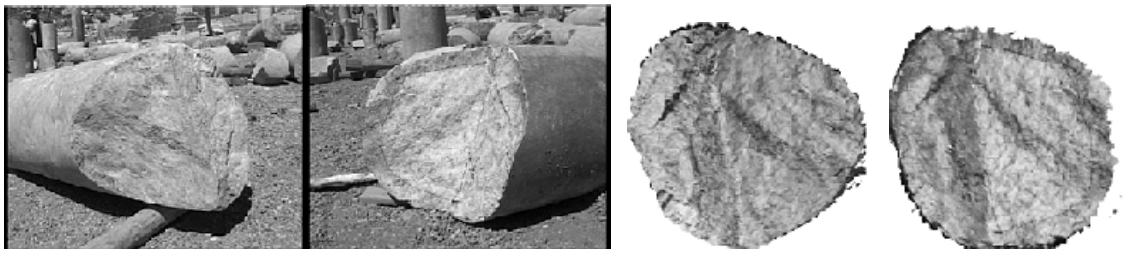

Figure 7 Two images of parts of broken pillars (top) and two orthographic views of the matching surfaces generated from the 3D models (bottom)

6.5 Mixing remains and virtual reconstructions

Another challenging application consists of seamlessly merging virtual objects with real video. In this case the ultimate goal is to make it impossible to differentiate between real and virtual objects. Several problems need to be overcome before achieving this goal. The most important of them is the rigid registration of virtual objects into the real environment. This can be done using the motion computation that was presented in this paper. A more detailed discussion of this application can be found in (Cornelis et al. 2001).

The following example was recorded at Sagalassos in Turkey, where footage of the ruins of an ancient fountain was taken. The *fountain* video sequence consists of 250 frames. A large part of the original monument is missing. Based on results of archaeological excavations and architectural studies, it was possible to generate a virtual copy of the missing part. Using the proposed approach the virtual reconstruction could be placed back on the remains of the original monument, at least in the recorded video sequence. The top part of Figure 8 shows a top view of the recovered structure before and after bundle-adjustment. Besides the larger reconstruction error it can also be noticed that the non-refined structure is slightly bent. This effect mostly comes from not taking the radial distortion into account in the initial structure recovery. In the rest of Figure 8 some frames of the augmented video are shown.

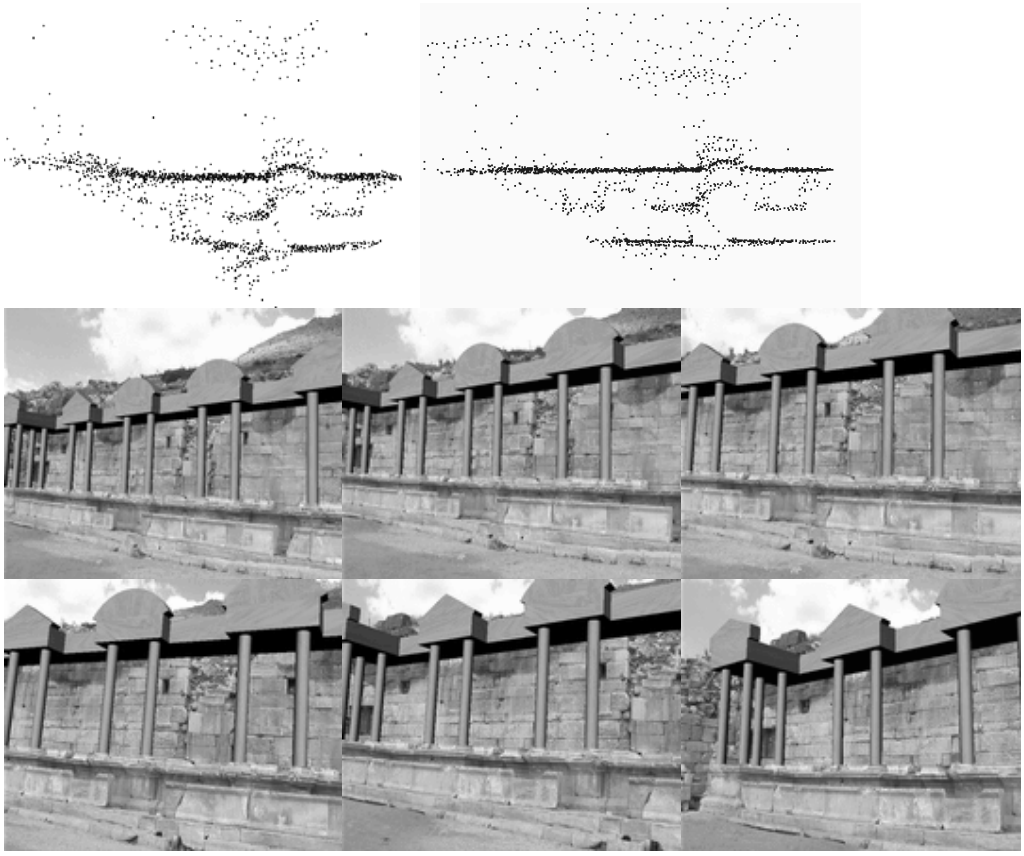

Figure 8 Fusion of real and virtual fountain parts. Top: Structure before and after bundle adjustment. Bottom: 6 of the 250 frames of the fused video sequence.

6 CONCLUSION

In this paper an approach for obtaining virtual models with a hand-held camera was presented. The approach utilizes different components that gradually retrieve all the information that is necessary to construct virtual models from images. Automatically extracted features are tracked or matched between consecutive views and multi-view relations are robustly computed. Based on this the projective structure and motion is determined and subsequently upgraded to metric through self-calibration. Bundle-adjustment is used to refine the results. Then, image pairs are rectified and matched using a stereo algorithm and dense and accurate depth maps are obtained by combining measurements of multiple pairs. From these results virtual models can be obtained or, inversely, virtual models can be inserted in the original video.

This technique was successfully applied to the acquisition of virtual models of archaeological sites. There are multiple advantages: 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 verification of building hypothesis and mixing archaeological remains with virtual reconstruction in video.

ACKNOWLEDGEMENT
Marc Pollefeys and Kurt Cornelis are respectively post-doctoral fellow and research assistant of the Fund for Scientific Research - Flanders (Belgium). The financial support of the FWO project G.0223.01, the ITEA BEYOND of the IWT and the IST-1999-20273 project 3DMURALE are also gratefully acknowledged.

References

P. Beardsley, A. Zisserman and D. Murray, 1997. Sequential Updating of Projective and Affine Structure from Motion, International Journal of Computer Vision (23), No. 3, pp. 235-259.

Y. Chen and G. Medioni, Object modelling by registration of multiple range images, Image and Vision Computing, vol. 10, no. 3, pp. 145-155, April 1992.

K. Cornelis, M. Pollefeys, M. Vergauwen and L. Van Gool, 2001. Augmented Reality from Uncalibrated Video Sequences, In M. Pollefeys, L. Van Gool, A. Zisserman, A. Fitzgibbon (Eds.), 3D Structure from Images - SMILE 2000, Lecture Notes in Computer Science, Vol. 2018, Springer-Verlag. pp.150-167.

I. Cox, S. Hingorani and S. Rao, 1996. A Maximum Likelihood Stereo Algorithm, Computer Vision and Image Understanding, Vol. 63, No. 3.

B. Curless and M. Levoy, 1996. A Volumetric Method for Building Complex Models from Range Images. Proc. SIGGRAPH. pp. 303-312.

P. Debevec, C. Taylor and J. Malik. 1996. Modelling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach. Proc. SIGGRAPH. pp. 11-20.

L. Falkenhagen, 1997. Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints, Proc. International Workshop on SNHC and 3D Imaging, Rhodes, Greece, pp.115-122.

O. Faugeras, 1992. What can be seen in three dimensions with an uncalibrated stereo rig, Computer Vision - ECCV'92, Lecture Notes in Computer Science, Vol. 588, Springer-Verlag, pp.563-578.

M. Fischler and R. Bolles, 1981. RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography, Commun. Assoc. Comp. Mach., 24:381-95.

C. Harris and M. Stephens, 1988. A combined corner and edge detector, Fourth Alvey Vision Conference, pp.147-151.

R. Hartley, R. Gupta, and T. Chang, 1992. Stereo from uncalibrated cameras. Proc. Conference Computer Vision and Pattern Recognition, pp. 761-764.

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

R. Koch, M. Pollefeys and L. Van Gool, 1998. Multi Viewpoint Stereo from Uncalibrated Video Sequences. Proc. European Conference on Computer Vision, pp.55-71.

R. Koch, B. Heigl, M. Pollefeys, 2001, Image-based rendering from uncalibrated lightfields with scalable geometry, in R. Klette, T. Huang, G. Gimel´farb (Eds.), Multi-Image Analysis, Lecture Notes in Computer Science, Vol. 2032, pp.51-66.

M. Pollefeys, R. Koch and L. Van Gool, 1999a. A simple and efficient rectification method for general motion, Proc.ICCV, pp.496-501.

M. Pollefeys, R. Koch and L. Van Gool. 1999b. Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters, International Journal of Computer Vision, 32(1). pp.7-25.

J. Shi and C. Tomasi, 1994. Good Features to Track, Proc. IEEE Conference on Computer Vision and Pattern Recognition. pp. 593 - 600.

C. Tomasi and T. Kanade, 1992, Shape and motion from image streams under orthography: A factorization approach, International Journal of Computer Vision, 9(2): 137-154.

P. Torr, 1995. Motion Segmentation and Outlier Detection, PhD Thesis, Dept. of Engineering Science, University of Oxford.

B. Triggs, 1997. The Absolute Quadric, Proc. Conference on Computer Vision and Pattern Recognition, pp.609-614.

B. Triggs, P. McLauchlan, R. Hartley, A. Fiztgibbon, 2000. Bundle Adjustment -- A Modern Synthesis, In B. Triggs, A. Zisserman, R. Szeliski (Eds.), Vision Algorithms: Theory and Practice, LNCS Vol.1883, pp.298-372, Springer-Verlag.