3D Recording for Archaeological Fieldwork

Marc Pollefeys
University of North Carolina, Chapel Hill

Luc Van Gool
Katholieke Universiteit Leuven and ETH Zürich

Maarten Vergauwen, Kurt Cornelis, Frank Verbiest, and Jan Tops
Katholieke Universiteit Leuven

Until recently, archaeologists have had limited 3D recording options because of the complexity and expense of the necessary recording equipment. We outline a system that helps archaeologists acquire 3D models without using equipment more complex or delicate than a standard digital camera.

In archaeology, measurement and documentation are both important, not only to record endangered archaeological sites, but also to record the excavation process itself. Annotation and precise documentation are important because evidence is actually destroyed during archaeological work. On most sites, archaeologists spend a large amount of time drawing plans, making notes, and taking photographs. Because of the publicity that accompanied some recent archaeological research projects, such as Stanford's Digital Michelangelo project[1] or IBM's Pieta project,[2] archaeologists are becoming aware of the advantages of using 3D visualization tools.

Archaeologists can now use the data recorded during excavations to generate virtual 3D models suited for project report presentation, restoration planning, or even digital archiving, although many issues remain unresolved. Until recently, the cost in time and money to generate virtual reconstructions remained prohibitive for most archaeological projects. At a more modest level, some archaeologists use commercially available software, such as PhotoModeler (http://www.photomodeler.com), to build simple virtual models. These models can suffice for some types of presentations, but typically lack the detail and accuracy needed for most scientific applications.

Clearly, archaeologists need more flexible measurement techniques, especially for fieldwork. Archaeologists should be able to acquire their own measurements simply and easily. Our image-based 3D recording approach offers several possibilities.[3-8] To acquire a 3D reconstruction, our system lets archaeologists take several pictures from different viewpoints using a standard photo or video camera. In principle, using our system means that archaeologists need not take additional measurements of the scene to obtain a 3D model. However, a reference length can help in obtaining the reconstruction's global scale. Archaeologists can use the resulting 3D model for measurement and visualization purposes. Figure 1 shows an example of the types of pictures possible with a standard camera.

In developing our system, we regularly visited Sagalassos, a site that is one of the largest archaeological projects in the Mediterranean. The site consists of elements from a Greco-Roman period spanning more than a thousand years from the 4th century BC to the 7th century AD. Sagalassos, one of the three great cities of ancient Pisidia, lies a few miles north of the village Aglassun in the province of Burdur, Turkey. The ruins of the city lie on the southern flank of the Aglassun mountain ridge (a part of the Taurus mountains) at an elevation of several thousand feet. Figure 2 shows Sagalassos against the mountains. A team from the University of Leuven, under the supervision of Marc Waelkens, has been excavating the area since 1990. Because of the different measurement problems, Sagalassos has been an ideal test field for our algorithms.

Image-based 3D recording

The first step in our 3D recording system recovers the relative motion between images taken consecutively. This process involves finding corresponding features between these images—image points that originate from the same 3D features. The process happens in two phases. First, the reconstruction algorithm generates a reconstruction containing a projective skew so that initially parallel lines are not parallel, angles are not correct, and distances are too long or too short. Next, using a self-calibration algorithm,[3,9] our system removes these distortions, yielding a reconstruction equivalent to the original.

The reconstruction only contains a sparse set of 3D points. Although interpolation might be one solution, it yields models with poor visual quality. Therefore, the next step attempts to match all of an image's pixels with those from neighboring images so that the system can

Published by the IEEE Computer Society

(a)

(b)

1 Reconstruction of a corner of the Roman baths at the Sagalassos archaeology site. (a) Our system used the six images (b) and automatically created the model.

reconstruct these points. A pixel in the image corresponds to a ray in space. Because we can predict the projection of this ray in other images from the camera's recovered pose and calibration, we can restrict the search for a corresponding pixel in other images to a single line. We also employ additional constraints, such as the assumption of a piecewise continuous 3D surface, to constrain the search even further.

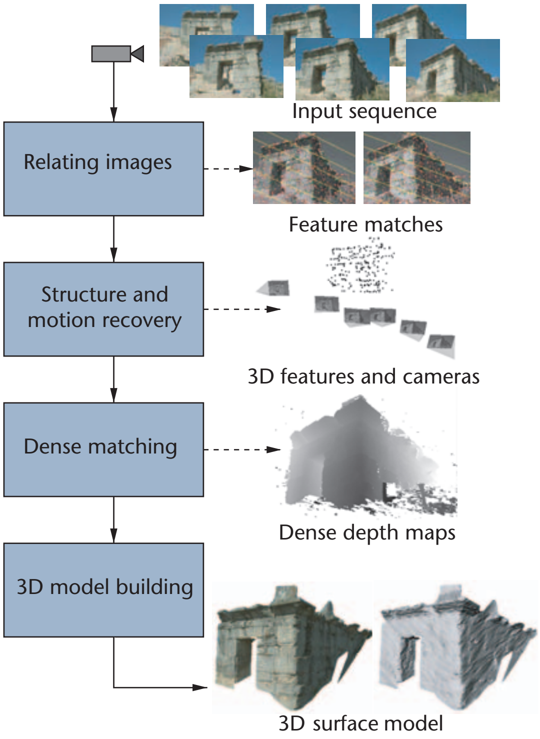
2 Overview of the Sagalassos site.

It's possible to warp the images so that the search range coincides with the horizontal scan-lines, letting us use an efficient stereo algorithm to compute an optimal match for the whole scan-line at once.[6] Using this algorithm, we can obtain a depth estimate—the distance from the camera to the object surface—for almost every pixel of an image. Fusing all the images' results gives us a complete surface model. To achieve a photorealistic result, we can apply the images used in the reconstruction as texture maps. Figure 3 (next page) illustrates the four steps of the process. The following sections describe the steps in more detail.

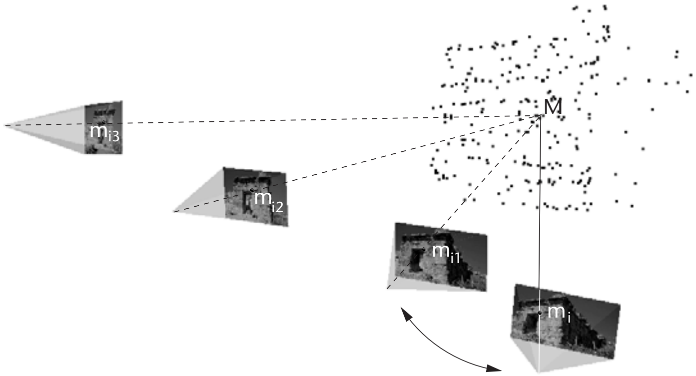
Relating images

Starting from a collection of images or a video sequence, our system's first step relates the different images to each other. A restricted number of corresponding points helps determine the images' geometric relationships. Our system selects the feature points suited for matching or tracking. Depending on the type of image data—such as video or still pictures—our system tracks the feature points to obtain several potential correspondences. From these correspondences, we compute the multiview constraints.

However, the set of corresponding points can be—and almost certainly will be—contaminated with several wrong matches. In light of this potential trouble, a traditional least-squares approach will fail; we therefore use a more robust method. The system uses the mul-



3 Overview of our image-based 3D recording approach.



4 Estimation of a new view using inferred structure-to-image matches.

5 Example of a rectified stereo pair.

tiview constraints to guide the search for additional correspondences, which it can in turn employ to refine results for the multiview constraints.

Structure and motion recovery

Our system uses the relation between the views and the correspondences between the features to retrieve the scene's structure and the camera's motion. Our approach doesn't depend on the initialization because we carry out all measurements in the images using reprojection errors instead of 3D errors. The system selects two images to set up an initial projective reconstruction frame and then reconstructs matching feature points through triangulation. The system then refines the initial reconstruction and extends it. By sequentially applying the same procedure, the system can compute the structure and motion of the whole sequence. Figure 4 illustrates the pose-estimation procedure.

The system can refine these results through a global least-squares minimization of all reprojection errors. Efficient bundle-adjustment techniques work well for this process. The ambiguity is then restricted further through self-calibration. Finally, the system carries out a second bundle adjustment, taking the self-calibration into account to obtain an optimal estimation of the images' structure and motion.

Dense surface estimation

To obtain a more detailed model of the observed surface, we use a dense-matching technique. The system can use the structure and motion obtained previously to constrain the correspondence search. Because we compute the calibration between successive image pairs, we can exploit the epipolar constraint that restricts the correspondence search to a one-dimensional search range. The system warps image pairs so that epipolar lines coincide with the image scan-lines. For this purpose, we use a rectification scheme[5] that deals with arbitrary relative camera motion. We then reduce the correspondence search to a matching of the image points along each image scan-line, which increases the algorithm's computational efficiency.

Figure 5 shows an example of a rectified stereo pair. The system has located all corresponding points on the same horizontal scan-line in both images. In addition to the epipolar geometry, we can exploit other constraints, such as the neighboring pixels' order and the match's bidirectional uniqueness. We use these constraints to guide the correspondence search toward the most probable scan-line match using dynamic programming.[6] 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. The algorithm's pyramidal estimation scheme deals with large disparity ranges, but the system limits the disparity search range according to observed feature disparities from the previous reconstruction stage.

The pairwise disparity estimation lets us compute image-to-image correspondence between adjacent rec-

(a) (b) (c) (d)

6 3D reconstruction of Dionysus showing (a) one of the original video frames, (b) the corresponding depth map, (c) a shaded view of the 3D reconstruction, and (d) a view of the textured 3D model with the original images.

tified image pairs and independent depth estimates for each camera viewpoint. We obtain an optimal joint estimate by fusing all independent estimates into a common 3D model using a Kalman filter. The system can perform the fusion economically through controlled correspondence linking,[4] which combines the advantages of small- and wide-baseline stereo and provides a dense depth map by avoiding most occlusions. Multiple viewpoints increase the depth resolution while small local baselines simplify the matching.

Building virtual models

Our dense structure and motion recovery approach yields all the necessary information to build textured 3D models. We approximate the 3D surface with a triangular mesh to reduce geometric complexity and tailor the model to the computer graphics visualization system requirements. A simple approach consists of overlaying a 2D triangular mesh on one of the images, then building a corresponding 3D mesh by placing the triangle vertices in 3D space according to the values found in the corresponding depth map. We use the image itself as the texture map. If no depth value is available or the confidence is too low, our system doesn't reconstruct the corresponding triangles. This approach works well on dense depth maps obtained from multiple stereo pairs.

A multiview linking scheme can enhance the texture itself. The system computes a median or robust mean of the corresponding texture values to discard imaging artifacts such as sensor noise, specular reflections, and highlights. To reconstruct more complex shapes, the system must combine multiple depth maps. Because all depth maps reside in a single metric frame, registration is not an issue. To integrate the multiple depth maps into a single surface representation, we use a volumetric technique.[10] Alternatively, to render new views from similar viewpoints, we use image-based approaches[11] that avoid the difficult problem of obtaining a consistent 3D model by using view-dependent texture and geometry. Doing so also helps us take into account more complex visual effects, such as reflections and highlights.

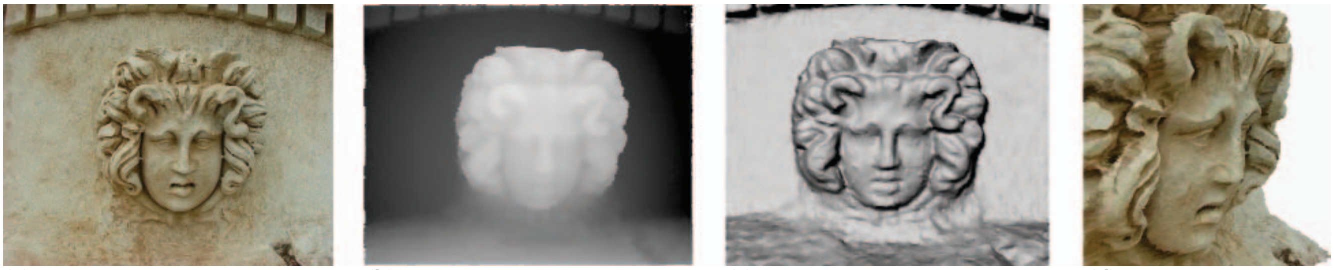
Applications to archaeological fieldwork

The techniques described here have many applications in the field of archaeology. The on-site acquisition procedure consists of recording an image sequence of the scene. So the algorithms can yield good results, although the viewpoint changes between consecutive images should not exceed 5 to 10 degrees. An example is the Dionysus statues found in Sagalassos on the upper market square. The statue is now located in the garden of the museum in Burdur.

It was simple to record a one-minute video of the statue. Bringing in more advanced equipment, such as a laser range scanner, would not only be logistically more complicated but would also require special authorization. Figure 6 illustrates different steps of the reconstruction process. We obtained the 3D model from a single depth map. We could have obtained a more complete and accurate model by combining multiple depth maps. And we could have obtained a smoother look for the shaded model by filtering the 3D mesh in accordance with the standard deviations obtained as a byproduct of the depth computation. This type of result is not so important when the model is texture mapped.

Figure 7 shows a second example, a Medusa head located on the entablature of a fountain. We obtained the 3D model from a short video sequence and used a single depth map to reconstruct the 3D model. Errors on the camera motion and calibration computations can result in a global bias on the reconstruction. From the results of the bundle adjustment, we estimate this error to be of the order of just a few millimeters for points on the reconstruction. The depth computations indicate that 90 percent of the reconstructed points have a relative error of less than 1 mm. The stereo correlation uses a window that corresponds to the object and therefore the measured depth will typically correspond to the dominant visual feature within that patch.

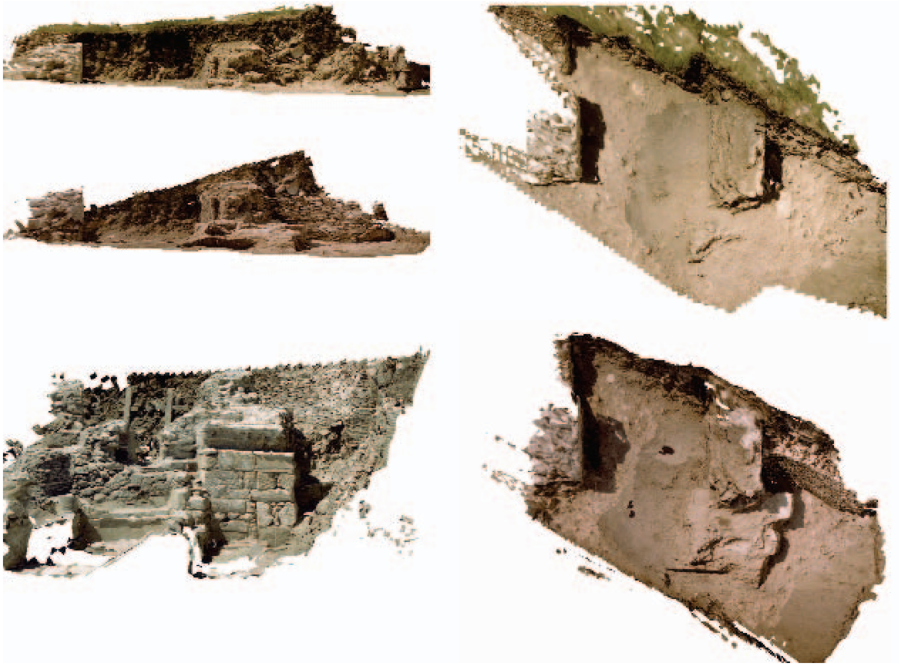
An important advantage of our approach compared to more interactive techniques[12] is that it can deal with more complex objects. Compared to non-image-based techniques, we can extract surface texture directly from the images, resulting in a much higher degree of realism and contributing to the authenticity of the recon-



(a) (b) (c) (d)

7 3D reconstruction of a Medusa head showing (a) one of the original video frames, (b) the corresponding depth map, (c) a shaded view of the 3D model, and (d) a textured view of the 3D model.

8 3D stratigraphy showing the excavation of a Roman villa at three different moments. The left image shows a front view of three stratigraphic layers. The right image shows a top view of the first two layers.



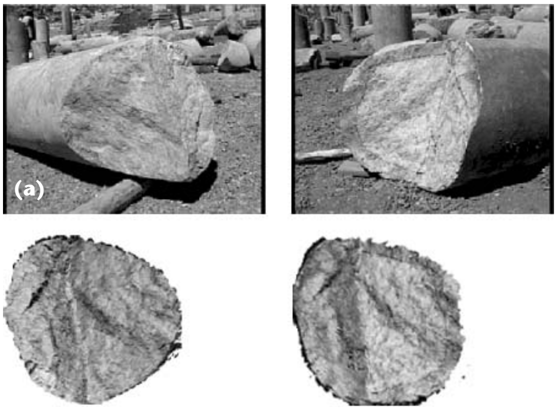
struction. Archaeologists can use reconstructions obtained with this system as scale models on which they can carry out measurements or plan restorations.

A disadvantage of our technique is that it can't directly capture the photometric properties of an object. It's therefore not immediately possible to rerender the 3D model under different lighting. We could possibly combine recent work[13] on recovering the radiometry of multiple images with our approach so that we could decouple reflectance and illumination. However, doing so would require us to record the scene under different illuminations or lighting conditions.

Recording 3D stratigraphy

An important aspect of archaeological annotation and documentation is stratigraphy, a process that reflects the different layers of soil that correspond to different time periods in an excavated sector. Because of practical limitations, stratigraphy is often only recorded for certain soil slices, not for the whole sector. Our technique allows a more optimal approach to this documentation problem. We can generate a complete 3D model of the excavated sector for every layer. Because the technique only involves taking a series of pictures, it does not slow down the progress of the archaeological work.

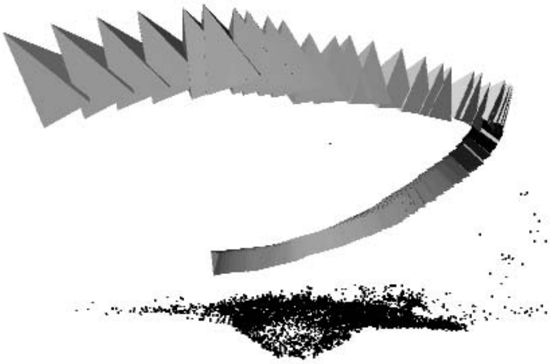
In addition, our system enables modeling all found artifacts separately and including the models in the final 3D stratigraphy, which makes it possible to use the 3D record as a visual database. For example, we recorded the excavations of an ancient Roman villa at Sagalassos using our technique. Figure 8 shows several layers of the excavation's 3D model. It took about one minute per layer to acquire the images at the site. From the results of the bundle adjustment, we can estimate the global error to be of the order of 1 cm for points on the reconstruction. Similarly, the depth computations indicate that the depth error of most of the reconstructed points should be within 1 cm. The correlation window corresponds to an area of approximately five square centimeters in the scene. This means that some small details might not appear in the reconstruction, but this accuracy level is more than sufficient to satisfy the requirements of the archaeologists. To obtain a single 3D representation for each stratigraphic layer, we used a volumetric integration approach.



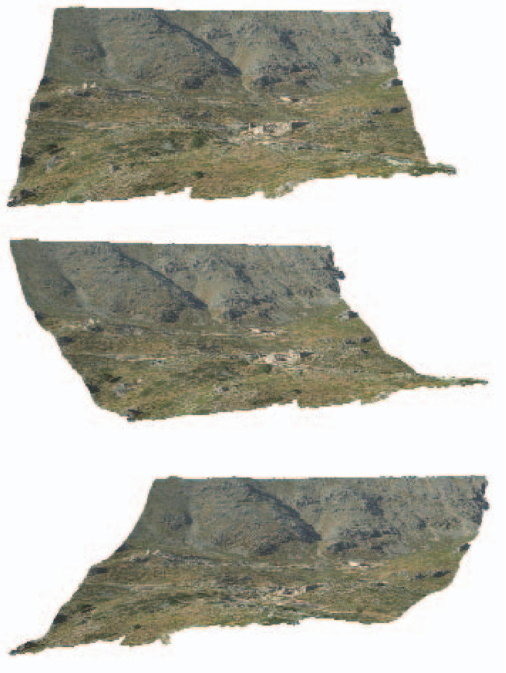
(b)

9 (a) Two images of a broken pillar. (b) The orthographic views of the matching surfaces generated from the obtained 3D models. The surface on the right is observed from the inside of the column.

10 Three images of the helicopter shot of the ancient theater of Sagalassos.



11 The reconstructed feature points and camera poses recovered from the helicopter shot.



12 Overview model of Sagalassos.

Construction and reconstruction

Our technique also offers many advantages in terms of generating and testing construction hypotheses. Ease of acquisition and the level of detail we can obtain make it possible to reconstruct every building block separately. Archaeologists could then interactively verify different construction hypotheses on a virtual reconstruction site. We could even use registration algorithms[14,15] to automate this process. Figure 9 shows two segments of a broken column. The whole monument contains 16 columns that were all broken in several pieces by an earthquake. Because each piece can weigh several hundreds kilograms, trying to fit the pieces together is very difficult. Traditional drawings also do not offer a proper solution.

Our approach's flexibility lets us use existing photo or video archives to reconstruct a virtual site. This application is suited for monuments or sites destroyed through war or natural disaster. We illustrated the feasibility of this type of approach with a reconstruction of the ancient theater of Sagalassos based on a video sequence filmed by Belgian TV as part of a documentary on Sagalassos. From the 30-second helicopter shot, we extracted about one hundred images. Because of the motion in the images, we could only use fields, not frames, restricting the vertical resolution to 288 pixels. Figure 10 shows three images from the sequence. Figure 11 shows the reconstruction of the feature points together with the recovered camera poses.

Obtaining a virtual reality model for a whole site consists of taking a few overview photographs from a distance. Because our technique is independent of scale, it can yield an overview model of the whole site. The only difference is the distance needed between two camera poses. Figure 12 shows an example of the results obtained for Sagalassos. We created the model from nine images taken from a hillside near the excavation site. It's a relatively straightforward process to extract a digital terrain map from the global site reconstruction. We could achieve absolute localization by localizing as few as three reference points in the 3D reconstruction.

The problem is that this kind of overview model is too coarse for use in realistic walkthroughs or for close-up views at monuments. For these purposes, archaeologists would need to integrate more detailed models into the overview model by taking additional image sequences for all the interesting areas on the site. The system would use these additional images to generate reconstructions

13 **<<Originally figure 15>>Architect contemplating a virtual reconstruction at Sagalassos.**

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 detail levels. Seamlessly merging reconstructions obtained at different scales remains an issue for further research.

Fusing real and virtual

Another potentially interesting application is combining recorded 3D models with other model types. In the case of Sagalassos, we translated some reconstruction drawings to CAD models[16] and integrated them with our Sagalassos models. This reconstruction is available at http://www.esat.kuleuven.ac.be/sagalassos/ as an interactive virtual reality application that lets users take a virtual visit to Sagalassos.[17]

Another challenging application consists of seamlessly integrating virtual objects in video. In this case, the ultimate goal is to make it impossible to differentiate between real and virtual objects. But to do this, we need to overcome several problems first. Among them are the rigid registration of virtual objects into the real environment, the mutual occlusion of real and virtual objects, and the extraction of the real environment's illumination distribution to render virtual objects with the illumination model. Accurate registration of virtual objects into a real environment, as shown in Figure 13, is a challenging problem. Systems that fail to do so will also fail to give the user a real-life impression of the augmented outcome.

Because our approach does not use markers or a priori knowledge of the scene or the camera, it lets us deal with video footage of unprepared environments or archived video footage. More details on our approach can be found elsewhere.[18] To successfully insert a large virtual object in an image sequence, the 3D structure should not be distorted. For this purpose, it's important to use a camera model that takes nonperspective effects into account and to perform a global least-squares minimization of the reprojection error through a bundle adjustment.

Conclusions

Our approach uses several different components that gradually retrieve all information necessary to construct virtual models from images. There are multiple advantages to using our 3D modeling technique: The on-site acquisition time is brief, the construction of the models is automatic, and the generated models are realistic. Our technique supports some promising applications, such as recording 3D stratigraphy, generating and verifying construction hypotheses, reconstructing 3D scenes based on archive photographs or video footage, and integrating virtual reconstructions with archaeological remains in video footage.

Our future research plans consist of increasing the reliability and flexibility of our approach. One important topic is the development of wide-baseline matching techniques so that pictures can be taken further apart. Another aspect consists of being able to take advantage of auto-exposure modes without degrading the visual quality of the models. In terms of applications, we are exploring possibilities in different fields, including architectural conservation, geology, forensics, movie special effects, and planetary exploration. ■

Acknowledgments

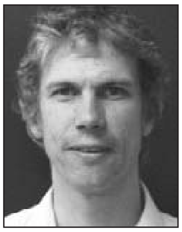
We thank Marc Waelkens and his team for making it possible for us to do experiments at the Sagalassos (Turkey) site. We also thank the FWO project G.0223.01, the IST-1999-20273 project 3DMURALE and the NSF project IIS 0237533 for their financial support.

References

1. M. Levoy et al., "The Digital Michelangelo Project: 3D Scanning of Large Statues," *Proc. Siggraph*, ACM, 2000, pp. 131-144 .
2. H. Rushmeier et al., "Acquiring Input for Rendering at Appropriate Levels of Detail: Digitizing a Pieta," *Proc. 9th Eurographics Rendering Workshop*, Springer-Verlag, 1998, pp. 81-92.
3. M. Pollefeys, R. Koch, and L. Van Gool, "Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters," *Int'l J. Computer Vision*, vol. 32, no. 1, 1999, pp. 7-25.
4. R. Koch, M. Pollefeys, and L. Van Gool, "Multi-Viewpoint Stereo from Uncalibrated Video Sequences," *Proc. European Conf. Computer Vision*, Springer-Verlag, 1998, pp.55-71.
5. M. Pollefeys, R. Koch, and L. Van Gool, "A Simple and Efficient Rectification Method for General Motion," *Proc. ICCV*, IEEE Computer Society Press, 1999, pp. 496- 501.
6. G. Van Meerbergen et al., "A Hierarchical Symmetric Stereo Algorithm Using Dynamic Programming," *Int'l J. Computer Vision*, vol. 47, nos. 1-3, 2002, pp. 275-285.
7. M. Pollefeys et al., "Virtual Models from Video and Vice-Versa," *Proc. Int'l Symp. Virtual and Augmented Architecture*, Springer-Verlag, 2001, pp.11-22.
8. M. Pollefeys, "Obtaining 3D Models with a Handheld Camera," *Siggraph Course*, ACM, 2001, Course notes CD-ROM; http://www.cs.unc.edu/~marc/tutorial/.
9. M. Pollefeys, F. Verbiest, and L. Van Gool, "Surviving Dominant Planes in Uncalibrated Structure and Motion Recov-

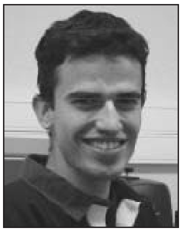
ery," *Proc. European Conf. Computer Vision*, Springer-Verlag, LNCS, vol. 2351, 2002, pp. 837-851.
10. B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images," *Proc. SIGGRAPH 96*, ACM, 1996, pp. 303-312.
11. R. Koch, B. Heigl, and M. Pollefeys, *Image-Based Rendering from Uncalibrated Light-fields with Scalable Geometry*, LNCS 2032, Springer-Verlag, 2001, pp.51-66.
12. P. Debevec, C. Taylor, and J. Malik, "Modeling and Rendering Architecture from Photographs: A Hybrid Geometry and Image-Based Approach," *Proc. SIGGRAPH 96*, ACM, 1996, pp. 11-20.
13. Q.-T. Luong, P. Fua, and Y. Leclerc, "The Radiometry of Multiple Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, Jan. 2002, pp. 19-33.
14. Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," *Proc. IEEE Int'l Conf. on Robotics and Automation*, IEEE CS Press, 1991, pp 2724-2729.
15. J. Wyngaerd et al., "Invariant-Based Registration of Surface Patches," *Proc. Int'l Conf. Computer Vision*, IEEE CS Press, 1999, pp. 301-306.
16. F. Martens et al., "Computer-Aided Design and Archaeology at Sagalassos: Methodology and Possibilities of Reconstructions of Archaeological Sites," *Virtual Reality in Archaeology*, J. Barcelo, M. Forte, and D. Sanders, eds., ArcheoPress, 2000, pp. 205-212.
17. M. Pollefeys et al., "A Guided Tour to Virtual Sagalassos," *2nd Int'l Symp. Virtual Reality, Archaeology, and Cultural Heritage*, ACM Siggraph, 2001, pp. 213 - 218.
18. K. Cornelis et al., Augmented Reality from Uncalibrated Video Sequences, LNCS 2018, Springer-Verlag, 2001, pp.150-167.

Marc Pollefeys *is an assistant professor of computer vision in the department of computer science at the University of North Carolina, Chapel Hill. His research interests include developing flexible approaches to capture visual representations of real-world objects, scenes, and events. He received a PhD in electrical engineering from the Katholieke Universiteit Leuven.*

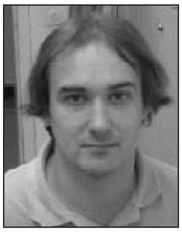


Luc Van Gool *is professor of computer vision at the Katholieke Universiteit Leuven and the Swiss Federal Institute of Technology. His research interests include 3D reconstruction, animation, texture synthesis, object recognition, robot vision, and motion capture. He received a PhD in Electrical Engineering from the Katholieke Universiteit Leuven.*

Maarten Vergauwen *is PhD student in the electrical engineering department at the Katholieke Universiteit Leuven. His research interests include 3D reconstruction from images, camera calibration, and vision in space applications. He received an MS in electrical engineering from the Katholieke Universiteit Leuven.*



Kurt Cornelis *is a PhD student in the electrical engineering department at the Katholieke Universiteit Leuven. His research interests include 3D reconstruction, recognition, and the use of computer vision in augmented reality. He received an MS in electrical engineering from the Katholieke Universiteit Leuven.*



Frank Verbiest *is a PhD student in the department of electrical engineering at the Katholieke Universisteit Leuven. His research interests include computer vision, especially 3D reconstruction from images. He received an MS in electrical engineering and an MS in artificial intelligence from the Katholieke Universiteit Leuven.*

Jan Tops *is a research assistant in the department of electrical engineering at the Katholieke Universisteit Leuven. His research interests include computer vision, 3D modeling, and visualization. He received an MS in computer science from the Katholieke Universiteit Leuven.*

Readers may contact Marc Pollefeys at the Univ. of North Carolina at Chapel Hill, Dept. of Computer Science, CB #3175, 205 Sitterson Hall, Chapel Hill, NC 27599-3175; marc.pollefeys@cs.unc.edu.