A Guided Tour to Virtual Sagalassos

Marc Pollefeys Luc Van Gool Ive Akkermans Dirk De Becker Kris Demuynck

Center for Processing of Speech and Images
Dept. of Electrical Engineering
K.U.Leuven

Abstract

The system presented in this paper allows to take a guided tour to a virtual reconstruction of the ancient site of Sagalassos. The site has been reconstructed by combining 3D models of the landscape and remains with CAD reconstructions of monuments based on archaeological hypothesis. The 3D models were obtained using advanced image-based modelling techniques developed over the last few years in computer vision. The visitor can communicate with the virtual guide through natural speech. The guide itself is represented by a 3D face model.

Keywords: Virtual Tour Guides, Large scale terrain modelling, Site reconstruction, Speech recognition.

1 Introduction

There is a lot of interest of the public for visiting ancient archaeological sites and often the excavations are financed by public funding. Therefore, making a site accessible to an interested public is also one of the major tasks of archaeologists. This often includes (partially) rebuilding some of the monuments. On the other hand, it is also the responsability of the archaeologist to preserve the remains of the past for future generations. Theoretically, reburial is one of the safest methods to preserve excavated monuments for the future, but since non-specialists in that case have hardly any access to their past, it should be avoided as much as possible. These two requirements are often contradictory. Another important aspect is that archaeological sites are often located at remote places, making them unaccessible to tourists, students and scholars.

In this area virtual reality is a technology that offers promising perspectives for archaeologists. Having a virtual reconstruction of the site, reduces the need for physical reconstructions. In addition, new insights can be gained by immersion in ancient worlds, unaccesible sites can be made available to a global public, courses can be given in situ and different periods or building phases can coexist.

One of the important limitations up to now was the high cost and the required expertise to reconstruct virtual models of a complete site. Nowadays 3D modelling tools allow to recreate monuments based on archaeological hypothesis [7]. Capturing the detailed geometry and appearance of the remains that are still standing is, however, much harder. Tools such as PhotoModeler [9] and Facade [2] allow to easily generate simple models of reality, but are not suited for generating objects with complex geometry or that require high levels of detail, like large archaeological sites. For modelling the remains of Sagalassos, advanced automatic 3D modelling tools were used [10, 11, 12] (see Section 3).

Sagalassos is one of the largest archaeological projects in the Mediterranean dealing with a Greco-Roman site over a period of more than a thousand years (4^{th} century BC - 7^{th} century AD). Sagalassos, one of the three great cities of ancient Pisidia, lies 7km north of the village Aglasun in the province of Burdur, Turkey.

The ruins of the city lie on the southern flank of the Aglasun mountain ridge (a part of the Taurus-mountains) at a height between 1400 and 1650 metres.

Figure 1 shows Sagalassos against the mountain flank. A team of the University of Leuven under the supervision of professor Marc Waelkens has been excavating the whole area since 1990.

As on a real site, allowing a visitor to freely walk around is in general not the best idea. In this way, the visitors often quickly gets bored and typically misses important monuments on the site. It is also hard to get the right information across. Therefore, it is better to provide the visitor with a virtual guide. In our case the guide consists of an animated face that can understand the requests of the user through speech recognition. The visitor can e.g. ask to go to a certain building or ask questions about finds that have been made at different places.

In the next section an overview of the system is presented. Then, some of the components of the system are discussed more in detail: 3D modelling (Section 3), 3D visualization (Section 4) and speech recognition (Section 5). The last two sections contain some future directions for research and the conclusions.

Figure 1: Overview of the Sagalassos site.

2 System overview

The complete system integrates a diversity of techniques to create a interactive virtual tour through the ancient city of Sagalassos. The central module of the virtual guide system consists of a finite state machine that describes the actual location during the visit and some other state variables (e.g. guide visible or not). Based on the speech input, state transitions can be triggered. Some other transitions are triggered automatically e.g. after time-out. This module also controls the motion and animation of the guide.

The viewpoint of the user is automatically adjusted to the position of the guide, i.e. the user follows the guide. However, the user still has some local control over his viewpoint. When head tracking is enabled motions of the head causes the viewpoint to change. The viewpoint can also be modified using mouse or joystick.

The 3D visualization is done using OpenGL Performer [8]. The system runs on high-end SGI hardware as well as on linux PCs. Using SGI hardware high-end 3D stereo visualization screens can be used (see e.g. [1]).The architecture of the virtual guide system is illustrated in Figure 2.

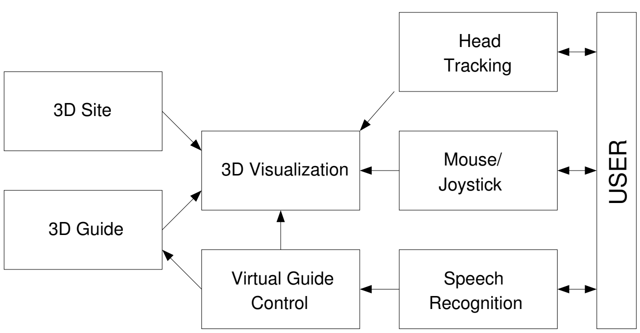

Figure 2: Overview of the Virtual Guide system

The virtual guide consists of a 3D mask. Animations are used to provide some feedback to the user e.g. the guide frowns the eyebrows when he can't understand what the user is saying. This allows to obtain a relatively natural communication between the user and the system. This is important since it allows untrained users to react correctly to the system e.g. by repeating a command if necessary. In Figure 3 a close-up of the virtual guide is shown. In fact this is already the second iteration of the Sagalassos Virtual Guide. The first system was build several years ago and was based on SGI's OpenInventor for visualization. A screenshot of this system can be seen in Figure 4.

Figure 3: Close-up of the virtual guide

Figure 4: Snapshot of the previous version of the virtual guide.

3 Site Modelling

The 3D surface acquisition technique that we have developed [10, 11, 12] 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 capture. To allow the algorithms to yield good results, viewpoint changes between consecutive images should not exceed 5 to 10 degrees.

The further processing is fully automatic. The first step consists of calibrating the image sequence (i.e. obtaining relative position and orientation of the camera up to scale for the different viewpoints). This is done based on features which are tracked over consecutive images. The next part is the computation of the surface geometry itself. For every pixel in an image the corresponding points in other images are searched through crosscorrelation and an optimal solution is found through dynamic programming. Once corresponding points are known, the actual surface is reconstructed in 3D through triangulation. The images can be used as texture maps. An example of a 3D model and the image sequence it was exctrated from is shown in Figure 5.

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 could also be used as

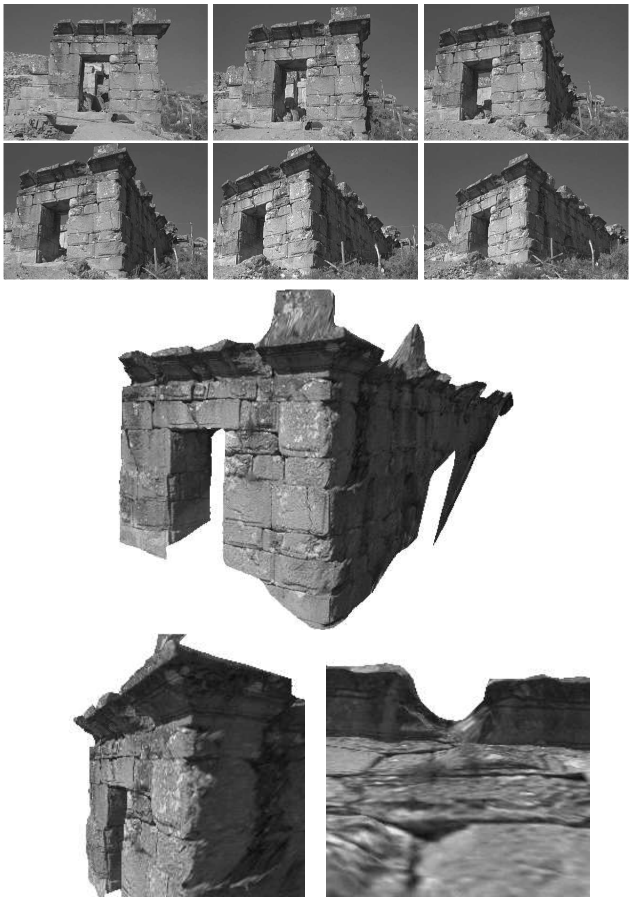

Figure 5: Corner of Roman baths in Sagalasos: Top: 6 images used for the reconstruction, Bottom: different views of the reconstruction.

a scale model on which measurements can be carried out or as a tool for planning restaurations.

A number of other techniques have also been used to reconstruct parts of the site. For highly detailed objects such as statues the ShapeSnatcher technology from Eyetronics [14] was used. This technology is based on a structured light approach and can record both shape and texture from a single image. The first processing step automatically extracts the projected grid from the image. The shape is then obtained by observing the deformation of the grid and the texture is obtained by removing the grid using an advanced non-linear diffusion filter. An example is shown in Figure 6. Another

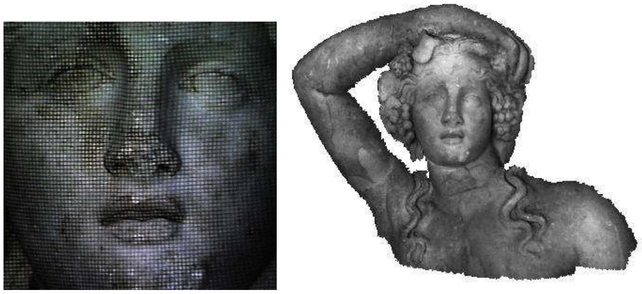

Figure 6: Dionysos statue: Left: image with projected grid, Right: reconstructed model.

recent technique was also considered [5].

Since some of the monuments and buildings are only partially standing, it is not possible to capture their shape directly. In this case 3D modelling tools have to be used to reconstruct the buildings based on archaeological hypothesis. The procedure can be sketched as follows. During excavations archaeologists take notes and make measurements and drawing of their findings. After a careful study of the monument it is possible to make reconstruction drawings based on all the collected evidence and by comparison with other related monuments. These then represent construction hypothesis for the considered monument. An example is shown on the left of Figure 7. Using a 3D modelling package it is then possible to build a 3D representation of that monument. The corresponding reconstruction is shown on the right side of Figure 7. More details on this subject can be found in [7].

Figure 7: Reconstruction drawing and 3D model of the Heroon of Sagalassos.

Figure 8: 3D representation of Sagalassos containing 3D models obtained through image-basd techniques and interactive 3D modelling tools.

A view of the integrated 3D representation of the site can be seen in Figure 8 Some more views of details of the integrated virtual Sagalassos site can be seen in Figures 9, 10 and 11.

4 3D visualization

An important component of the system is the 3D visualization. For this purpose OpenGL Performer is used. This makes it possible to

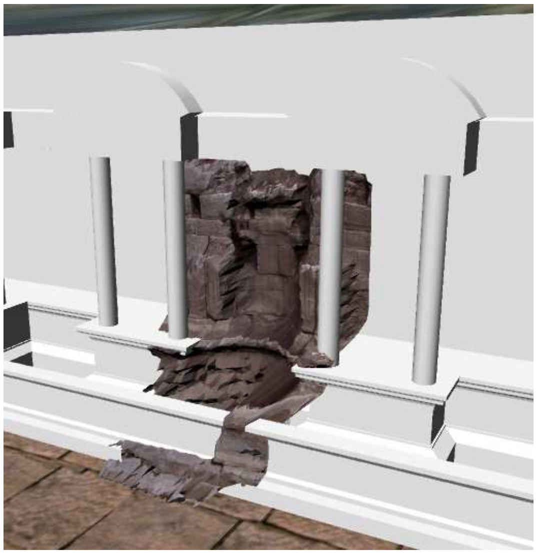

Figure 9: Reconstruction of the fountain of the upper agora with recorded central niche.

Figure 10: Excavations of the roman baths recorded at three different levels of resolution.

have the Virtual Guide running on linux PCs as well as high-end SGI visualization systems. The system has been developed in such a way that different types of displays can be used.

Besides a standard PC screen the system will also be running on high-end BARCO displays such as the BARCO Baron VR workbench and the BARCO Reality Center cylindrical display (See Figure 12). In this case it also makes sense to use headtracking so that the perspective projection adapts to the position of the user. This is important to increase the immersion for the user. However, this also limits the full experience of the visit to a single user.

5 Speech recognition

For the purpose of speech recognition two different systems have been used. In one version our in-house large-vocabulary speaker-independent continuous speech recognizer was used [3, 4]. This system is, similar to all modern large vocabulary speech recognizers, based on statistical pattern recognition and integrates all knowledge concerning the spoken input according to the scheme depicted in figure 13. The acoustic models describe the time-frequency characteristics of all basic sounds (phonemes) in a language by means of hidden Markov processes. These time-frequency properties are learned from a large database containing a few hundred speakers, allowing the guide to cope with almost any voice or accent. By relying on phonemes as fundamental speech units, all words relevant to the task can be easily added to the system by just providing their phonetic transcription, as can be found in pronouncing dictionaries. To cope with co-articulation effects within and between words, context dependent phoneme models are used, i.e. not the phonemes themselves are modelled but the phonemes given their left and right neighboring phonemes. The final component, the language model, describes the format of typical user input.

Figure 11: Reconstruction of the Heroon located in the virtual Sagalassos terrain.

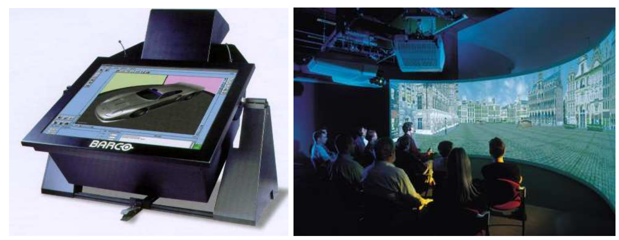

Figure 12: High-end VR displays considered for Sagalassos Virtual Guide.

Since speech recognizers have –in contrast to human guides– no world knowledge, every aspect of the user input, such as the function and meaning of the words and how sentences are formed, must be incorporated in this model. Automatic techniques that extract this kind of knowledge from large text corpora exist, but for new tasks such as the virtual guide, most knowledge must be added manually. To be useful as an interface to the virtual guide, the recognizer not only has to make few and preferably no recognition errors irrespective of the speaker, but it also has to react promptly.

Due to the smart representation of the lexicon [3] and the fast evaluation of the acoustic models [4], our recognizer provides real-time decoding in combination with a low latency, even for continuous speech and large vocabularies (20000 or more words). The low-latency of the recognizer allows for example direct visual feedback to the user by means of frowning eyebrows whenever a language construct that is likely to be misunderstood by the virtual guide, is heard. Although primarily designed to be a flexible research tool, our in-house speech recognizer was able to provide all functionality required for a natural communication with the virtual guide.

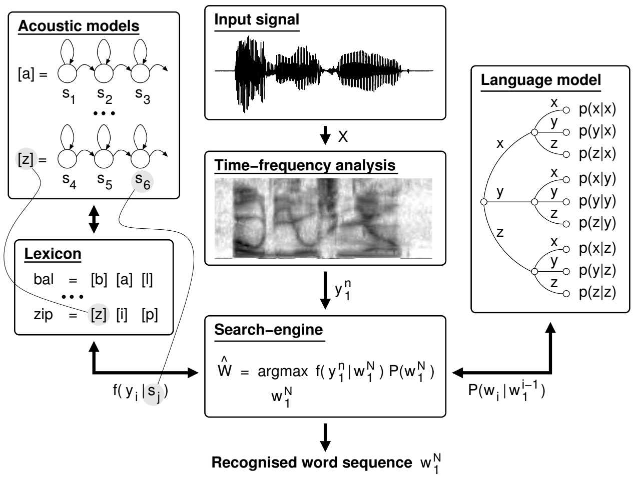

Figure 13: Overview of the large vocabulary speaker independent recognition system.

In another version of our system a small vocabulary on-chip system was used [6]. This last system provides adaptive stereo echo cancelling and noise reduction. This allows to provide effective speech recognition from a distant microphone, while allowing the system itself to have an audio output. In the future, similar techniques will be added to our in-house recognizer, allowing us to combine the best of both solutions: a natural interface using fluent speech with the possibility to interrupt the guide while he is providing audio output.

6 Future developments

Future work consists of using a real face for the virtual guide. The goal is to capture both the shape and the animation from a real person. For this purpose the technology of Eyetronics would be used for the capture [14]. A preliminary example is shown in Figure 14. Four different phases of an animation have been captured. The goal is to automatically build animation models by analysing the captured data. The ultimate goal in this context is to build a speech-to-animation (or even text-to-animation) package that can be used to automatically generate realistic facial animation for the guide based on data captured from a real person.

7 Conclusions

In this paper we presented our virtual guide system. This system allows to visit a virtual reconstructin of the ancient city of Sagalassos. High-end visualization is combined with speech recognition to provide a convincing experience to the visitor. Facial animation is succesfully used to provide the user with feedback. The 3D representation of the site is mostly obtained using advanced image-based techniques.

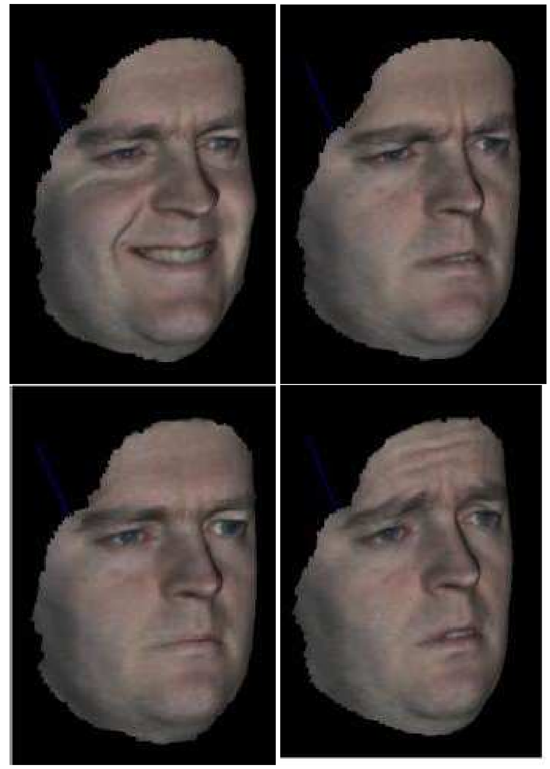

Figure 14: Four frames of a captured animation.

Acknowledgement

Marc Pollefeys is a post-doctoral fellow of the Fund for Scientific Research - Flanders (Belgium). The financial support of the KULeuven research fund (GOA VHS), the ITEA BEYOND project of the IWT and the IST-1999-20273 project 3DMURALE are also gratefully acknowlegded. We would also like to acknowledge the following people for their contributions: Marc Waelkens, Pol Legrand, Jos Legrand, Joris Vanden Wyngaerd, Danny Martens, Marc Proesmans, Jacques Duchateau, Konrad Schindler, Markus Grabner, Peter Ramon, Filip Louwet, Jean-Pierre Jallet, Bart Motmans, Piotr Wilinski and Fabian Ernst.

References

[1] BARCO projection systems, (more info http://www.barco.com/projection_systems/).

[2] P. Debevec, C. Taylor and J. Malik, Modelling and rendering architecture from photographs: A hybrid geometry- and image-based approach, *SIGGRAPH'96*, August 1996.

[3] K. Demuynck, J. Duchateau, D. Van Compernolle and P. Wambacq. An Efficient Search Space Representation for Large Vocabulary Continuous Speech Recognition. Speech Comm., volume 30, No. 1, pages 37-53, January 2000.

[4] J. Duchateau, K. Demuynck and D. Van Compernolle. Fast and Accurate Acoustic Modelling with Semi-Continuous HMMs. Speech Comm., volume 24, No. 1, pages 5-17, April 1998.

[5] F. Ernst, C. van Overveld, P. Wilinski. Efficient generation of 3-D models out of depth maps To appear at Vision, Modelling and Visualization (VMV01), Stuttgart, Germany, November 21 - 23, 2001.

[6] Hello IC, by Philips semiconductors, (for more info: http://www.semiconductors.philips.com/news/content/file_690.html, visited september 2001).

[7] F. Martens, P. Legrand, J. Legrand, L. Loots and M. Waelkens, "Computer-aided design and archaeology at Sagalassos: methodology and possibilities of reconstructions of archaeological sites", *Virtual Reality in Archaeology*, J. Barceló, M. Forte and D. Sanders (Eds.), ArcheoPress, Oxford (British Archaeological Reports, International Series # 843).

[8] OpenGL Performer, by SGI, (for mroe information: http://www.sgi.com/software/performer/)

[9] PhotoModeler, by Eos Systems Inc., (for more information: http://www.photomodeler.com)

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

[11] M. Pollefeys, R. Koch, M. Vergauwen, L. Van Gool. Automated reconstruction of 3D scenes from sequences of images, Isprs Journal Of Photogrammetry And Remote Sensing (55)4 (2000) pp. 251-267.

[12] Marc Pollefeys, Luc Van Gool, Maarten Vergauwen, Kurt Cornelis, Frank Verbiest, Jan Tops, Image-based 3D Acquisition of Archaeological Heritage and Applications, Proc. VAST 2001.

[13] Marc Proesmans, Luc Van Gool and A. Oosterlinck, One-shot active range acquisition, Proc. Internat. Conf. Pattern Recognition (ICPR), Vienna, Vol. C, pp. 336-340, 1996

[14] ShapeSnatcher, by Eyetronics. (more information http://www.eyetronics.com).