Acquisition of Detailed Models for Virtual Reality

M. Pollefeys[*], M. Proesmans, R. Koch, M. Vergauwen and L. Van Gool
ESAT-PSI, K.U.Leuven, Belgium
Marc.Pollefeys@esat.kuleuven.ac.be

Abstract

More and more archaeological sites are being reconstructed in virtual reality. The problem remains the huge effort that has to be made to obtain realistic models. Besides on-site measurements, much time is often spent in manually rebuilding the whole site with a CAD package or a 3D-modelling tool. This limits the tractable complexity. In this paper two flexible automatic 3D surface acquisition systems are used to "virtualise" archaeological sites. The Roman site of Sagalassos (Southwest Turkey) is used as a test case to illustrate the potential of these new approaches. 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.

Introduction

Virtual reality is a technology that offers promising perspectives for archaeologists. It can help in many ways. New insights can be gained by immersion in ancient worlds, inaccessible sites can be made available to a global public, courses can be given "on-site" and different periods or building phases can coexist.

One of the main problems is however the generation of these virtual worlds. They require a huge amount of on-site measurements. In addition the whole site has to be reproduced manually with a CAD- or 3D modelling system. This requires a lot of time. In addition it is difficult to model complex shapes and to take all the details into account. Obtaining realistic surface texture is also a critical issue. As a result walls are often approximated by planar surfaces, stones often all get the same texture, statues are only crudely modelled, small details are left out, etc.

An alternative approach consists of using images of the site. Some software tools exist, but require a lot of human interaction (PhotoModeler) or preliminary models (Debevec, 1996). In this contribution some tools are presented which allow generating dense surface models of statues, monuments, buildings or archaeological sites using only consumer products as photo- or video cameras. With these techniques the complexity of the acquisition is moved from hardware to software. This involves many advantages. A camera can work in demanding circumstances. It is cheap, easy to transport and can be handled by non-specialist. The on-site acquisition time of the presented technique is short (i.e. it only involves taking some images).

In addition these models which are immediately extracted from images contain much more detail than models which are build from a restricted number of measured points. In addition the images themselves are also used to generate the texture. These are important advantages for generating realistic VR walk-throughs.

These detailed models open the way to new applications of VR in archaeology. Once it is easy to obtain detailed 3D models of almost any part of a building, it becomes possible to build ``Virtual Reconstructions''. It is much easier to test building hypothesis on a graphics workstation than on the site itself. Broken pillars can be reassembled using VR tools or even using automatic registration software to find corresponding parts.

Another promising application is the full 3D recording of the excavations for every stratigraphic layer. Therefore the archaeologists would not only dispose of a 3D model of the excavation, but of its evolution over time. An associated VR tool would not only allow travelling in space, but also in time.

All the examples in this contribution were taken at the archaeological site of Sagalassos in Southwest Turkey. The images were obtained with a consumer photo camera (digitised on photo-CD) and with a consumer digital video camera.

High-quality surface reconstruction with single-shot range sensor

The technique presented in this section is an active technique (Proesmans et al., 1996; Proesmans and Van Gool, 1998). It is very flexible and only requires a minor calibration step. A pattern is projected with a standard slide projector onto an object. A single image of the illuminated object is then sufficient to obtain a very accurate textured 3D reconstruction of that object.

---

[*] Postdoctoral fellow of the Fund for Scientific Research Flanders (Belgium) (F.W.O. Vlaanderen)

This technique was used to generate models of statues and masks found at Sagalassos. These can be found on the companion CD-ROM.

The advantage of this technique is the fast delivery of accurate 3D models of objects at a fraction of the price of a laser range scanner. This is due to the use of off-the-shelf components –slide projector and camera-- complemented with software.

Overview of the method

Starting from an image that contains the projected grid the first step consists of extracting horizontal and vertical lines. From this an initial grid is constructed. In general this grid still contains some inconsistencies. These are detected and corrected. Then the grid is refined to obtain subpixel accuracy. From the deformation of this final grid the shape is computed. At the same time the lines are filtered away to obtain the texture. In Figure 1 an overview of the different steps of the method is given.

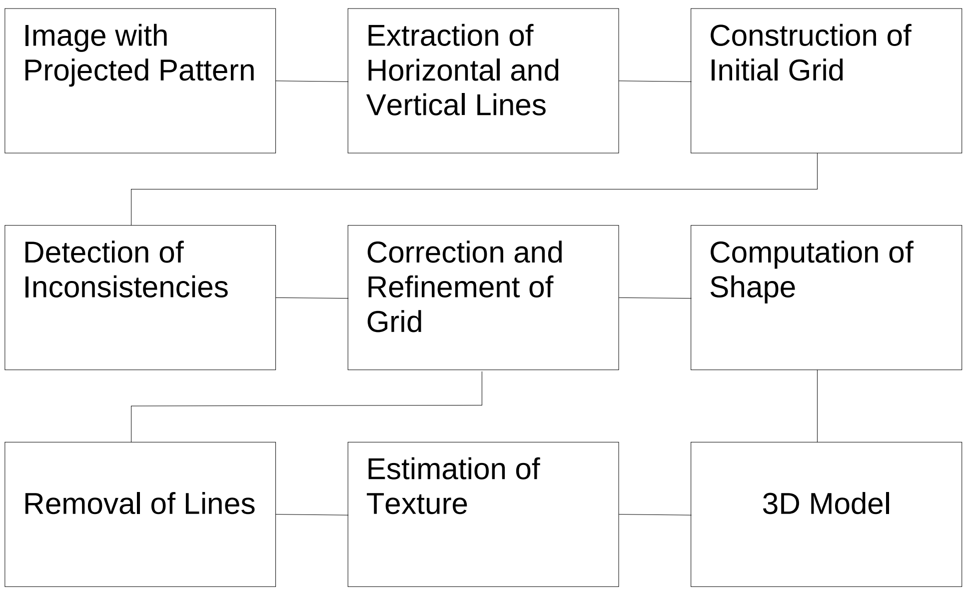

Figure 1 Overview of the single-shot range sensor method

System Calibration

Given the restriction of pseudo-orthographic projection, one can verify (Proesmans et al. 1996) that the shape can be recovered up to a scale, if the angle between the directions of projection and viewing is known. In order to find this angle, the system is initialized or "calibrated". For calibration it suffices to show two planar patches and to give their subtended angle. In practice, showing a box or corner of a room and specifying the angle is all it takes to perform this calibration. The system will extract the grid as explained below, looks for the two largest planar regions within that grid and using the knowledge about their orientation (typically 90 degrees) gives the angle between directions of viewing and projection.

Extraction of the grid

Once the system is calibrated, the object to be modeled in 3D can be shown. The system first extracts the grid lines in the image, through a set of horizontally and vertically aligned line detectors. The different edge segments are then linked to two sets of crossing contours. The crossings are used to generate a complete grid representation of the pattern. Each grid point or "node" has at most 4 neighbors which are labeled *N*, *E*, *S* and *W*. Note that nodes that are vertices of quadrangles are of special interest, since they indicate how the square pattern is deformed under projection. In order to determine the lines with sub-pixel precision, a snake-like process is introduced. The problem is stated as the minimization of the energy functional that relates both the intensity along the grid connections as well as its smoothness. This makes sure that the grid will move towards the darker parts of the image and let it more closely follow the pattern lines.

Correction of the extracted grid

Ideally, we would now have a complete set of all the line intersections or nodes, together with their connections. Of course, the extracted grid will not be perfect yet. Nodes or their connections may have gone undetected, and similarly spurious nodes and connections may have been found. Furthermore, in order to arrive at a 3D reconstruction the connectivity of the lines has to be known. The important difference with classical techniques lies in the fact that no absolute correspondences between the lines in the grid and the lines in the image are (nor can be) derived. Finding the correct connectivity between lines is the crux of the matter (see Proesmans et al. 1996). To that end, iterative procedures mend the initially extracted grid. Two principles underlie this mending. First, the grid should look square when viewed from the direction of the projector. Secondly, the mending should lead to a consistent numbering for the two sets of grid: whatever path is taken between two nodes, the grid based city block distance should be the same. Based on these principles, connections are broken and created. These procedures allow the system to work its way around local discontinuities in the grid, e.g. due to sudden changes in depth. A limitation of the system as described is that the grid patches should be connected somewhere in the image to be reconstructed as a single surface.

Extraction of 3D shape

As soon as the grid has been extracted and the connectivity has been established, the extraction of the 3D shape is quite straightforward. The grid in the image plane can be modeled as the cross-section of two sets of planes with the object's surface. To each node a depth value can be assigned (see Proesmans et al. 1996). Note that the reconstruction is up to a scale factor. The knowledge of a single distance between two points would fix the scale.

As mentioned, the ability to deal with discontinuities is restricted to connected sets of data. If two or more separated objects are present in the scene, they will be reconstructed independently in the sense that their relative positions are not extracted.

For many applications in visualization, this is not necessarily a problem, because objects are either shown separately or the scene is synthesized starting from individual object models. Discontinuities by self-occlusion need not be a problem, however. The system recovers from self-occlusions. Of course, there are gaps in the reconstructions for those parts not visible to the camera.

It is noteworthy that the system yields a genuine surface description, not a mere cloud of points. The connections found between the nodes are used in the description, that consists of small bilinear patches (the little squares of the pattern). Thus, problems with retrieving the correct topology are largely eliminated.

Extraction of surface texture

With the given setup, shape and texture extraction could in principle be very simple by just alternating between images with and without the grid. The former would yield shape and the latter texture. Alignment would be perfect as the same, fixed camera takes both images. However, the need for switching the pattern on and off would complicate system operation. For one thing, it would require additional provisions to block the pattern. Moreover, if the alternations should follow each other quickly and automatically, this would require the synchronization of this blocking with the image acquisition by the camera. All this much to the detriment of the simplicity gained by using nothing but a simple camera and slide projector.

Therefore, a method was developed (and patented as the rest of the system) that extracts surface texture with the pattern on. It does away with the need to alternate between shape and texture frames altogether. The underlying principle literally is to read between the lines. This might seem a weird approach at first, but actually the human visual system solves a similar problem. The blood vessels on our retina are positioned in front of the photoreceptors. As with the blind spot, we are hardly aware of these obstacles because they are filtered out.

In Figure 2 an example of an image of an object with the grid pattern projected onto it can be seen. Figure 3 shows a view of the reconstructed 3D model.

Figure 2 Grid projection on 3D object

Figure 3 Reconstructed 3D model of Dionysos

3D reconstruction from uncalibrated image sequences

The second technique (Pollefeys et al. 1998; Koch et al. 1998) starts from an uncalibrated image sequence of a scene. Using nothing else than the images a textured 3D reconstruction of the recorded scene is obtained in an automated way. The sequence can be taken with a simple hand-held video- or photo camera. The camera need not be calibrated and zoom and focus can be used freely. The motion is unconstrained and we do not make use of any reference points. In addition the method is just as easy to use for small objects as for complete sites. Our method therefore offers a lot of flexibility and is easy to use. We have for example obtained a global reconstruction of the whole site of Sagalassos (Turkey), but also of separate monuments.
Some examples can also be found on the companion CD-ROM.

This technique could therefore be an alternative for a wide variety of measurement techniques used nowadays by archeologists, which range from drawings and measurements done manually, over traditional photogrammetry to the computation of a DTM (digital terrain map) from aerial photographs.

Overview of the method

In this paragraph a brief overview of 3D acquisition from multiple views is given. Starting from the images the knowledge about the scene and the cameras is gradually increased. At first a partial camera calibration is obtained together with a sparse reconstruction of the scene. This is all done using robust algorithms to insure that the algorithms can work on real images. The next step consists of upgrading the calibration of the cameras. Imposing some constraints that are in general valid for standard cameras does this. The next step consists of determining the complete surface (i.e. not only a few feature points).

This is done through dense correspondence algorithms. Finally a 3D-model is build which contains the geometric (i.e. the shape) and photometric (i.e. the color and the texture) information of the scene. These models can then be used for visualization or measurement purposes.
In Figure 2 an overview of the method is shown.

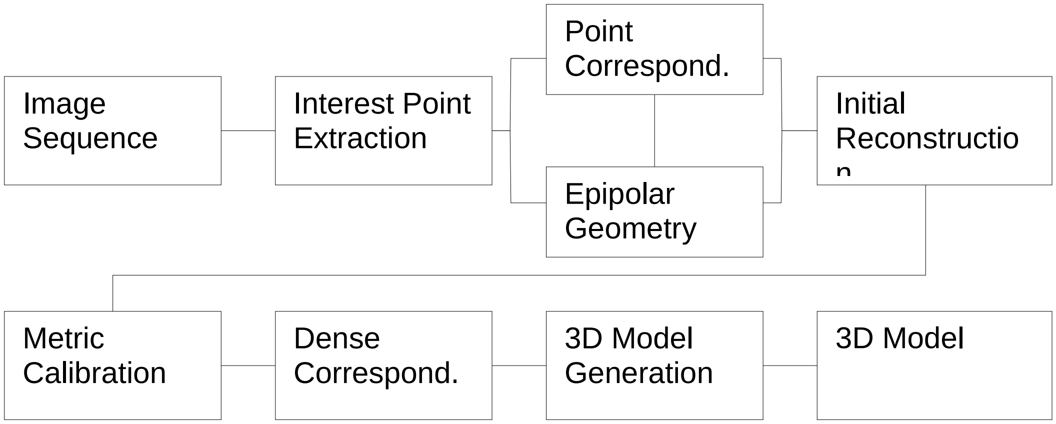

Figure 4 Overview of 3D-model acquisition from multiple views

The acquisition of the image sequence

The first step is to take the images from which the 3D model of the scene will be generated. It is important to take some restrictions of our method into consideration at this point. Since the images are to be matched automatically consecutive images of the sequence shouldn't be too different. Consecutive viewpoints should be targeted at more or less the same point and shouldn't be separated by more than 5 to 10 degrees. Illumination conditions are also important. If shadows are present at the time of the image acquisition, these will be found back in the final model. Additionally the reconstruction of parts that are in the shadow will be more complicate due to the restricted range of the intensity values in these parts of the images. Therefore the ideal circumstance for image acquisition is an overcast sky. Quite good results can also be obtained in other circumstances.

The images can be taken with different types of cameras. We have already used video cameras, digital still cameras and standard photo cameras. In this last case the images can either be scanned in or put on photo-CD. Typically images of 700x500 were used. When a higher resolution was available it was used to get a better texture map for the final model. An example of an image sequence can be seen in Figure 5.

Figure 5 Image sequence of a part of the Roman baths in Sagalassos

Relating the images

A first step in the reconstruction procedure is to relate the different images to each other. Therefore it is necessary to put points from different images into correspondence (i.e. finding the points which originate from the same 3D point). When the

images are not too different the neighborhood of these points will have the same appearance and they can therefore be matched through normalized intensity correlation.
To avoid a combinatorial explosion of the problem only a restricted amount of so-called interest points will be considered at this stage. These are extracted through a special filter which selects corner-like features from the images (Harris, 1988). In fact these features are often not real corners, but this filter tends to extract the same points in consecutive images (i.e. points corresponding to the same 3D points).

An important concept to simplify the matching process between two images is the epipolar geometry. This expresses the fact that a point which is situated on a specific plane through the two camera centers is bound to be projected on the intersection lines between that plane and the respective image planes. These lines are therefore in so-called epipolar correspondence (i.e. the corresponding point of every point on an epipolar line in one image should be found back on the corresponding epipolar line in the other image). All these epipolar lines pass through the epipole. This point is the intersection of the image plane with the line through the two camera positions.

If this geometry is known the matching complexity is reduced from a 2D (whole image) to a 1D search (epipolar line). In uncalibrated circumstances the only way to obtain this epipolar geometry is by computing it from a number of matches. We are therefore confronted with a chicken and egg problem. The approach that is followed first obtains a set of initial matches, then computes some candidate epipolar geometry from a minimal subset of matches. If this epipolar geometry receives enough support from the rest of the matches, it is accepted, else another subset is selected and the procedure is repeated. Once some satisfactory epipolar geometry has been obtained it can be used to generate more matches who in turn allow refining the epipolar geometry (see Torr (1995)).

A first reconstruction

Once the epipolar geometry and a set of corresponding points are known it is possible to generate a first reconstruction. This reconstruction is only determined up to a projective ambiguity. This means that besides not knowing the absolute position, orientation and scale of the scene some additional parameters are missing. In fact parallelism and orthogonality can not be verified at this stage and metric properties like relative distances and angles can not yet be measured.

This initial reconstruction is only based on the two first images of a sequence and is therefore far from complete. For additional images the epipolar geometry (with the previous image) is determined as described earlier. Then the relative position and orientation of the camera in the frame determined from the first two views is determined in a similar way through the already reconstructed points. At this point the reconstruction is refined, extended and corrected. A technical description of this approach can be found in Beardsley et al. (1996).

Obtaining metric information

The projective reconstruction obtained in the previous paragraph is clearly not satisfactory for our purposes since both visualization and measurement applications require at least a metric reconstruction. Luckily algorithms have been developed which can upgrade projective reconstruction to metric reconstructions. The approach presented here is based on the work described in Pollefeys et al. (1998).

It is always possible to factorize the cameras in intrinsic (e.g. focal length) and extrinsic parameters (i.e. position and orientation of the camera). In the case of a projective reconstruction the obtained parameters do not have a physical interpretation. In the metric case however they correspond to actual physical entities that are subject to some constraints. Self-calibration therefore consists of finding the projective transformation that transforms the initial reconstruction into a reconstruction for which all these constraints are satisfied.

The most successful methods to achieve this are based on some abstract geometric concepts. When a camera is moved through space, its relative position to the *plane at infinity* and the *absolute conic* does not change. This property allows identifying both entities in projective space. These entities encode the metric properties of space and are therefore equivalent to a metric calibration.

Once the metric calibration is retrieved, the initial reconstruction can be transformed to a metric reconstruction. This reconstruction is now a true scale model of the initial scene that was recorded. The absolute scale can be determined through a single measurement in the scene.

Dense correspondences

At first, only the correspondences for specific feature points were extracted. In this paragraph it will be discussed how the correspondences can be determined for almost all the points of the images. In this way a full surface model of the recorded scene can be reconstructed. This is not only important for visualization, but also for measurements since not all the interesting points are always retrieved as feature points in our sparse reconstruction.

The epipolar geometry greatly simplifies the matching problem. As described previously corresponding points have to be found on each other's epipolar lines. The matching problem can therefore be solved line by line.

To obtain efficient algorithms the images are first warped so that corresponding image rows are also corresponding epipolar lines. This procedure is called rectification.

The matching proceeds row by row over both images. Not only the correlation score is taken into account, but a cost is also included for discontinuities in the surface. Some other constraints are also taken into account (e.g. uniqueness, ordering). An optimal solution can then be obtained through dynamic programming.

The result is a disparity map. This is an image were the value associated with each pixel indicates the position of the corresponding pixel in the other. Since at this point the cameras are calibrated this can immediately be translated to a dept map. This is an image that contains the distance from the camera center to the scene along direction corresponding every pixel. This is often called a 2.5D representation.

To obtain a more accurate and more complete depth-information more than two images should be used. We have implemented an approach, which is described in detail in Koch et al. (1998). At first the depth maps are computed pair-wise for the whole sequence. Then, to compute accurate depths for a specific image, the disparities are used to create a correspondence chain that goes over multiple images. Taking into account the other correspondences of the chain using an extended Kalman filter refines the initial depth. The chain is interrupted once the back-projected ray falls outside the 95% probability bounds.

An important advantage of this approach is that one can go both backward and forward in the sequence. This allows avoiding most of the occlusion problems one typically encounters with stereo. In addition the inaccuracy of small baseline and the difficulty of matching for wide baseline stereo are both solved through this technique.

Additionally this multi-viewpoint stereo also allows improving the quality of the texture. One 3D point is observed in many images and the texture can therefor robustly be estimated. Highlights and other artifacts can easily be removed.

Building the model

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 viewpoint and work remains to be done to fuse different viewpoints. Since all the views are registered into one metric framework it is possible to fuse the depth estimate into one consistent model surface.

As an example, a few views of the reconstruction of a part of the Roman baths are shown in Figure 6.

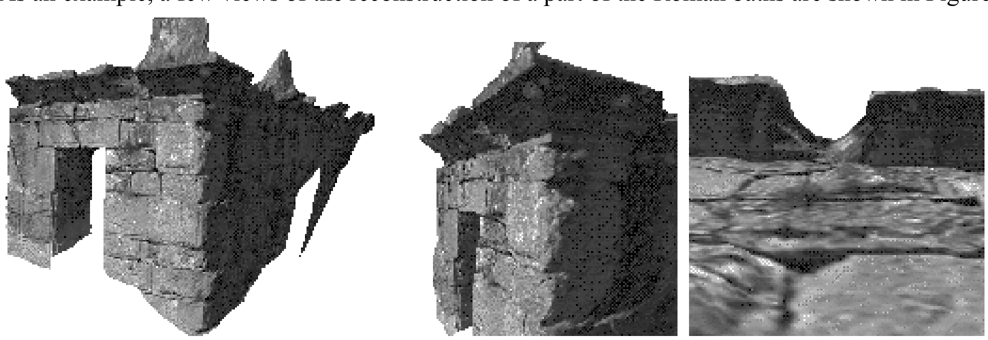

Figure 6 Reconstruction of a part of the Roman baths in Sagalassos

Constructing VR models of archaeological sites

In the previous sections it was explained how textured 3D surface models could be extracted from one or more images. In this section this will be applied to the construction of VR models of archaeological sites.

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 on the companion CD-ROM. 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.

The problem is that this kind of overview model is too coarse to be used for a realistic walk-through around the site or for looking at specific monuments. Therefore it is necessary to integrate more detailed models into this overview model. Taking additional image sequences for all the interesting areas on the site can do this. 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. On the CD-ROM the reconstructions of the Roman baths are given for three different levels of details (site overview, complete Roman bathhouse and detail of right corner).

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

The models delivered by the different techniques presented in this paper are thus well suited for generating VR models for virtual exhibitions. The models have a high visual quality since they are immediately generated from real images. Complex shapes like statues, masks or pottery can accurately be modeled. An important advantage of these virtual exhibitions is that it allows placing the object back in their natural environment. In this respect the method based on multiple views offers the advantage that the environment can also be “virtualized” without too much effort.

Figure 7 illustrates how 3D models can be combined with CAD models.

Figure 7 Combination of 3D model of Sagalassos with CAD models

Other applications

Since these 3D models are almost completely automatic and the on-site acquisition time is very short, several new applications come to mind. In this section a few possible applications of the presented techniques are described.

3D Measurements

Once a 3D surface model of an object is obtained, the model can be used to obtain 3D measurements. Our models are up to scale and can therefore immediately be used to obtain relative length measurements. Carrying out a single length measurement in the real scene allows getting absolute lengths from the model. Measuring lengths on the 3D model is much simpler than in the real scene. It is sufficient to click on two points on the model surface and the computer immediately computes the distance between the two points. When the model coordinates are aligned with the world coordinates a simple click on the model surface results in the absolute coordinates of the indicated point. It is clear that these procedures result in an important gain of efficiency for generating plans of the archeological sites.

Testing building hypothesis through VR

The presented techniques do not only allow acquiring the 3D-model of the remains of a building, but also of the broken down parts which are retrieved on the site. It becomes therefore possible to carry out a “virtual reconstruction”. The different parts of a monument can be reassembled on the screen of a computer.

This offers a flexible method to test building hypothesis without the problems of carrying this out with the real building blocks. Based on this additional tools could even be developed to simplify reconstruction even more. Broken parts of pillars could for example automatically be matched using standard registration software (see e.g. Chen and Medioni, 1991).

3D stratigraphy

Since much evidence is destroyed during diggings it is very important for archaeologists to annotate all the findings. Traditionally a lot of pictures are taken many plans and notes are made and the stratigraphic structure of slices of the terrain is drawn.

An alternative would be to do a 3D recording of a location after the removal of every stratigraphic layer. In this 3D reconstruction specific models of pieces of interest could be inserted at the exact location were they were discovered. The whole could then be assembled to a complete annotation model were a slider would allow to travel through time by adding or removing stratigraphic layers at will.

Some of these concepts are also illustrated through examples on the CD-ROM.

Conclusion

In this paper two techniques for 3D acquisition were presented. They both heavily rely on state-of-the-art computer vision and image processing algorithms. Compared to traditional acquisition techniques the complexity of the acquisition has been moved from hardware to software. Additionally, geometric insights have been used to avoid complicate calibration procedures and provide a maximal flexibility in acquisition. These different factors make the presented techniques especially suited for use in archaeology where often non-technical people have to carry out 3D acquisition in difficult circumstances.

Additionally the presented techniques offer the possibility to develop new applications or to improve existing approaches. A few examples of these applications were given in this paper.

Acknowledgement

We would like to thank Prof. Dr. Marc Waelkens and his team for making it possible for us to do experiments on the archeological site of Sagalassos (Turkey). We would also like to thank Dr. Andrew Zisserman and his team in Oxford for providing us with robust projective reconstruction algorithms. The financial support of the Belgian IUAP project IMechS is gratefully acknowledged.

References

BEARDSLEY, P., TORR, P. AND ZISSERMAN, A., 1996, “3D Model Acquisition from Extended Image Sequences”, *Computer Vision - ECCV’96*, Buxton, B. and Cipolla, R. (eds.), Lecture Notes in Computer Science, Vol.1065, pp. 683-695. Springer-Verlag.

CHEN, Y. AND MEDIONI, G., 1992, “Object Modeling by Registration of Multiple Range Images”, *Image and Vision Computing,* No.3, pp.145-155.

DEBEVEC, P., TAYLOR, C. AND MALIK, J., 1996, “Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach”, *Proceedings of SIGGRAPH’96,* pp.11-20, ACM.

HARRIS, C. AND STEPHENS, M., 1988. “A combined corner and edge detector”, *Proceedings Fourth Alvey Vision Conference*, pp.147-151.

KOCH, R., POLLEFEYS, M. AND VAN GOOL, L., 1998, “Multi Viewpoint Stereo from Uncalibrated Video Sequences”, *Computer Vision - ECCV98*, Burkhardt, H. and Neumann, B. (eds.), Lecture Notes in Computer Science, Vol.1406, Springer-Verlag.

KOCH, R., POLLEFEYS, M. AND VAN GOOL, L., 1998, Automatic 3D Model Acquisition from Uncalibrated Image Sequences, proceedings CGI'98 (Computer Graphics International)

MARTENS, F., LEGRAND, P., LEGRAND, J., LOOTS, L. AND WAELKENS, M., “Computer aided design and archaeology at sagalassos: methodology and possibilities of 3D computer reconstructions of archaeological sites”, *this book*.

PHOTOMODELER, by Eos Systems Inc., (for more information: http://www.photomodeler.com).
POLLEFEYS, M., KOCH, R., VERGAUWEN, M. AND VAN GOOL, L., 1998, “Metric 3D Surface Reconstruction from Uncalibrated Image Sequences”, *3D Structure from Multiple Images of Large-Scale Environments*, Koch, R. and Van Gool, L. (eds.), Lecture Notes in Computer Science, Vol. 1506, Springer-Verlag.
POLLEFEYS, M., KOCH, R. AND VAN GOOL, L., 1999, “Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters”, *International Journal of Computer Vision,* 32(1), pp.7-25.
PROESMANS, M., VAN GOOL, L. AND OOSTERLINCK, A., 1996, “One-shot active range acquisition”, *Proceedings International Conference on Pattern Recognition,* Vol. C, pp.336-340, IEEE Computer Society Press.
PROESMANS, M., AND VAN GOOL, L., 1998, “Reading between the lines -- a method for extracting dynamic 3D with texture”, *Proceedings International Conference on Computer Vision*, pp.1081-1086.
TORR, P., 1995, *Motion Segmentation and Outlier Detection*, Ph.D. Thesis, Dept. of Engineering Science, University of Oxford.