ELSEVIER

ISPRS Journal of Photogrammetry & Remote Sensing 55 (2000) 251–267

PHOTOGRAMMETRY & REMOTE SENSING

www.elsevier.nl/locate/isprsjprs

# Automated reconstruction of 3D scenes from sequences of images

M. Pollefeys [a,*], R. Koch [b], M. Vergauwen [a], L. Van Gool [a]

[a] *ESAT-PSI, K.U. Leuven, Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium*

[b] *Institut für Informatik und Praktische Mathematik, Christian-Albrechts-Universität Kiel, Olshausenstr. 40, D-24118 Kiel, Germany*

**Abstract**

Modelling of 3D objects from image sequences is a challenging problem and has been an important research topic in the areas of photogrammetry and computer vision for many years. In this paper, a system is presented which automatically extracts a textured 3D surface model from a sequence of images of a scene. The system can deal with unknown camera settings. In addition, the parameters of this camera are allowed to change during acquisition (e.g., by zooming or focusing). No prior knowledge about the scene is necessary to build the 3D models. Therefore, this system offers a high degree of flexibility. The system is based on state-of-the-art algorithms recently developed in computer vision. The 3D modelling task is decomposed into a number of successive steps. Gradually, more knowledge of the scene and the camera setup is retrieved. At this point, the obtained accuracy is not yet at the level required for most metrology applications, but the visual quality is very convincing. This system has been applied to a number of applications in archaeology. The Roman site of Sagalassos (southwest Turkey) was used as a test case to illustrate the potential of this new approach. 

*Keywords:* 3D reconstruction; self-calibration; image matching; virtual reality; uncalibrated camera; image sequences; archaeology

## 1. Introduction

Obtaining 3D models from objects is an ongoing research topic in computer vision and photogrammetry. A few years ago, the main applications were visual inspection and robot guidance. However, nowadays the emphasis is shifting. There is more and more demand for 3D models in computer graphics, virtual reality and communication. This results in a change in emphasis for the requirements. The visual quality becomes one of the main points of attention. The acquisition conditions and the technical expertise of the users in these new application domains can often not be matched with the requirements of existing systems. These require intricate calibration procedures every time the system is used. There is an important demand for flexibility in acquisition. Calibration procedures should be absent or restricted to a minimum. Additionally, the existing systems are often built around specialised hardware (e.g., Sequeira et al., 1999; El-Hakim et al., 1998) resulting in a high cost for these systems. Many new applications however require simple low-cost acquisition systems. This stimulates the use of consumer photo or video cameras.

Other researchers have presented various systems for extracting 3D shape and texture from image sequences acquired with a freely moving camera. The approach of Tomasi and Kanade (1992) used an

* Corresponding author. Tel.: +32-16-321064; fax: +32-16-321723.

*E-mail address:* Marc.Pollefeys@esat.kuleuven.ac.be (M. Pollefeys).

0924-2716/00/$ - see front matter 
PII: S0924-2716(00)00023-X

affine factorisation method to extract 3D from image sequences. An important restriction of this system is the assumption of orthographic projection. Debevec et al. (1996) proposed a system that starts from an approximate 3D model and camera poses and refines the model based on images. View dependent texturing is used to enhance realism. The advantage is that only a restricted number of images are required. On the other hand, a preliminary model must be available and the geometry should not be too complex. A similar approach was described by Dorffner and Forkert (1998).

In this paper, we present a system that retrieves a 3D surface model from a sequence of images taken with off-the-shelf consumer cameras. The user acquires the images by freely moving the camera around the object. Neither the camera motion nor the camera settings have to be known. The obtained 3D model is a scaled version of the original object (i.e., a *metric* reconstruction), and the surface texture is obtained from the image sequence as well. Our system uses full perspective cameras and does not require prior models or calibration. In contrast to existing systems such as PhotoModeler (2000), our approach is fully automatic. The complete system combines state-of-the-art algorithms of different areas of computer vision: *projective reconstruction*, *self-calibration* and *dense depth estimation*.

### 1.1. Projective reconstruction

It has been shown by Faugeras (1992) and Hartley et al. (1992) that a reconstruction up to an arbitrary projective transformation was possible from an uncalibrated image sequence. Since then, a lot of effort has been put in reliably obtaining accurate estimates of the projective calibration of an image sequence. Robust algorithms were proposed to estimate the fundamental matrix from image pairs (Torr, 1995; Zhang et al., 1995). Based on this, algorithms that sequentially retrieve the projective calibration of a complete image sequence were developed (e.g., Beardsley et al., 1997).

### 1.2. Self-calibration

Since a projective calibration is not sufficient for most applications, researchers tried to find ways to automatically upgrade projective calibrations to metric (i.e., Euclidean up to scale). When this is done based on some constraints on the camera intrinsic parameters, this is called self-calibration. Note that this definition is different from the usual definition in photogrammetry where self-calibration consists of the refinement of an initial calibration based on the available measurements (Fraser, 1997). Typically, it is assumed that the same camera is used throughout the sequence and that the intrinsic camera parameters are constant. This proved a difficult problem and many researchers have worked on it (e.g., Faugeras et al., 1992; Hartley, 1994; Triggs, 1997; Pollefeys and Van Gool, 1999). One of the main problems is that critical motion sequences exist for which self-calibration does not result in a unique solution (Sturm, 1997). We proposed a more pragmatic approach (Pollefeys et al., 1999a) which assumes that some parameters are (approximately) known but which allows others to vary. Therefore, this approach can deal with zooming/focusing cameras.

### 1.3. Dense depth estimation

Once the calibration of the image sequence has been estimated, stereoscopic triangulation techniques of image correspondences can be used to estimate depth. The difficult part in stereoscopic depth estimation is to find dense correspondence maps between the images. The correspondence problem is facilitated by exploiting constraints derived from the calibration and from some assumptions about the scene. We use an approach that combines local image correlation methods with a dynamic programming approach to constrain the correspondence search (see Falkenhagen, 1995; Cox et al., 1996; Koch, 1996). The results are further refined through a multi-view approach (Koch et al., 1998).

This paper is organised as follows. In Section 2, a general overview of the system is given. In the subsequent sections, the different steps are explained in more detail: projective reconstruction (Section 3), self-calibration (Section 4), dense depth estimation (Section 5) and model generation (Section 6). In Section 7, some results that were obtained from the archaeological site of Sagalassos are presented. Section 8 concludes the paper.

## 2. Overview of the method

The presented system gradually retrieves more information about the scene and the camera setup. The first step is to relate the different images. This is done pairwise by retrieving the epipolar geometry. An initial reconstruction is then made for the first two images of the sequence. For the subsequent

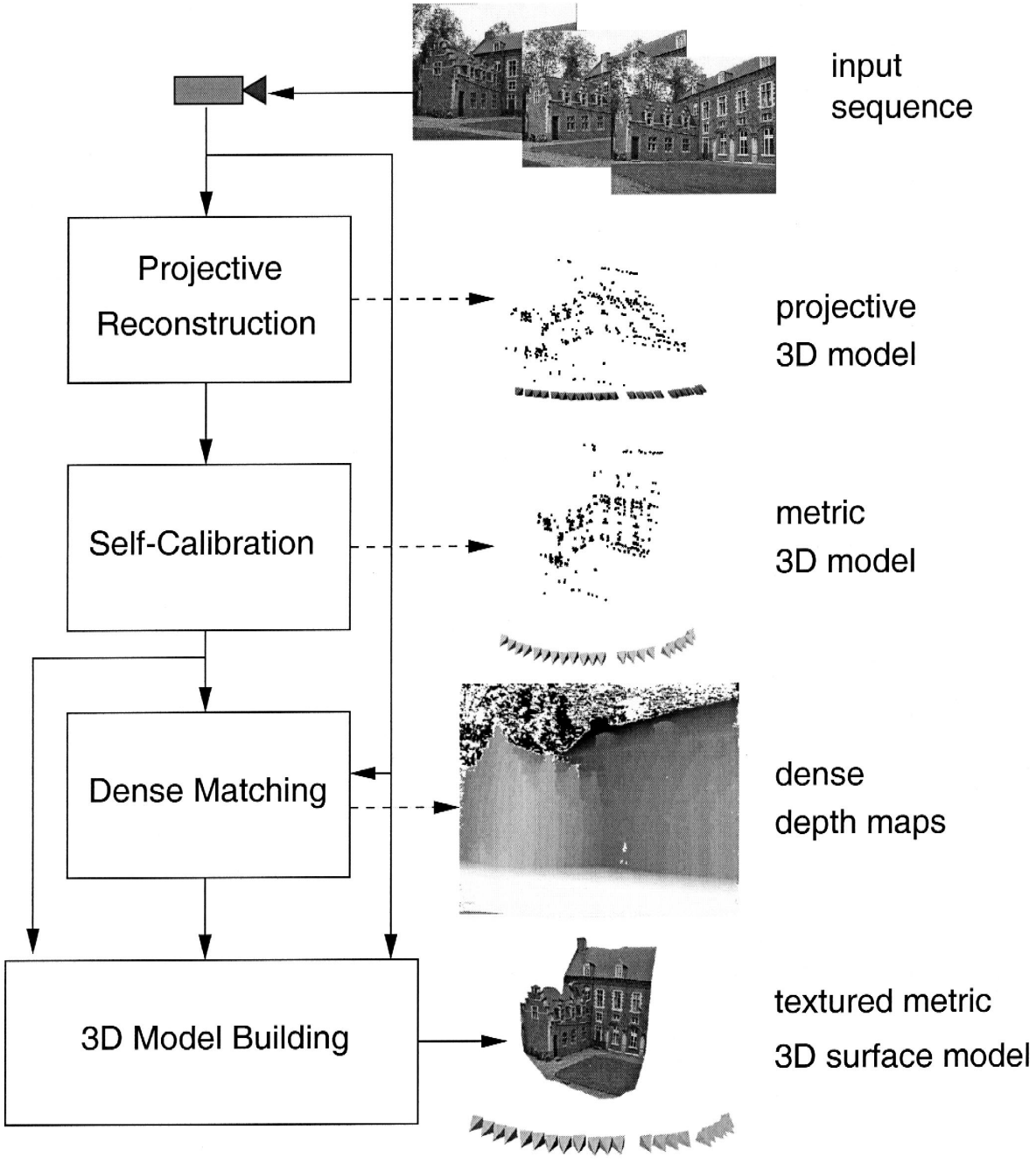


Fig. 1. Overview of the system (the cameras are represented by little pyramids).

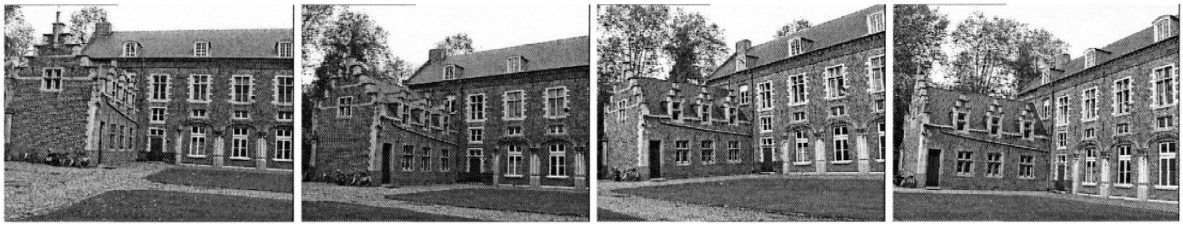

Fig. 2. Some images of the Arenberg castle sequence. This sequence is used throughout this paper to illustrate the different steps of the reconstruction system.

images, the camera pose is estimated in the projective frame defined by the first two cameras. For every additional image that is processed at this stage, the interest points corresponding to points in previous images are reconstructed, refined or corrected. Therefore, it is not necessary that the initial points stay visible throughout the entire sequence. The result of this step is a reconstruction of typically a few hundred to a few thousand interest points and the (projective) pose of the camera. The reconstruction is only determined up to a projective transformation.

The next step consists of restricting the ambiguity of the reconstruction to a metric one. In a projective reconstruction not only the scene, but also the cameras are distorted. Since the algorithm deals with unknown scenes, it has no way of identifying this distortion in the reconstruction. Although the camera is also assumed to be unknown, some constraints on the intrinsic camera parameters (e.g., rectangular or square pixels, constant aspect ratio or principal point close to the centre of the image) can often be assumed. A distortion on the camera mostly results in the violation of one or more of these constraints. A metric reconstruction/calibration is obtained by transforming the projective reconstruction to a solution for which all the constraints on the camera intrinsic parameters are satisfied.

At this point, the system effectively disposes of a calibrated image sequence. The relative position and orientation of the camera is known for all viewpoints. This calibration facilitates the search for corresponding points and allows the use of a stereo algorithm that was developed for a calibrated system and which allows to find correspondences for most of the pixels in the images. To allow the use of efficient algorithms, the images are first rectified. Since the standard planar rectification scheme does not work for all types of camera motion, a new procedure was proposed.

From the correspondences, the distance from the points to the camera centre can be obtained through triangulation. These results are refined and completed by combining the correspondences from multiple images. A dense metric 3D surface model is obtained by approximating the depth map with a triangular wireframe. The texture is obtained from the images and mapped onto the surface.

In Fig. 1, an overview of the system is given. It consists of independent modules, which pass on the necessary information to the next modules. The first module computes the projective calibration of the sequence together with a sparse reconstruction. In the next module, the metric calibration is computed from the projective camera matrices through self-calibration. Then, dense correspondence maps are estimated. Finally, all results are integrated in a textured 3D surface reconstruction of the scene under consideration.

In the following sections of this paper, the different steps of the method will be explained in more detail. An image sequence of the Arenberg castle in Leuven will be used for illustration. Some images of this sequence can be seen in Fig. 2. The full sequence consists of 24 images recorded with a video camera (resolution $768 \times 576$ pixels).

## 3. Projective reconstruction

At first, the images are completely unrelated. The only assumption is that the images form a sequence in which consecutive images do not differ too much. Therefore, the local neighbourhood of image points originating from the same scene point should look

similar, if images are close in the sequence. This allows for automatic matching algorithms to retrieve correspondences. The approach taken to obtain a projective reconstruction is very similar to the one proposed by Beardsley et al. (1997).

### 3.1. Relating the images

It is not feasible to compare every pixel of one image with every pixel of the next image. It is therefore necessary to reduce the combinatorial complexity. In addition, not all points are equally well suited for automatic matching. The local neighbourhoods of some points contain a lot of intensity variation and are therefore naturally more suited to differentiate from others. The Harris corner detector (Harris and Stephens, 1988) is used to select a set of such points. Correspondences between these image points need to be established through a matching procedure.

Matches are determined through normalised cross-correlation of the intensity values of the local neighbourhood. Since images are supposed not to differ too much, corresponding points can be expected to be found back in the same region of the image. Therefore, at first only interest points, which have similar positions, are considered for matching. When two points are mutual best matches, they are considered as potential correspondences.

Since the epipolar geometry describes the complete geometry relating two views, this is what should be retrieved. Computing it from the set of potential matches through least squares generally does not give satisfying results due to its sensitivity to outliers. Therefore, a robust approach should be used. Several techniques have been proposed (Torr, 1995; Zhang et al., 1995). Our system incorporates the method developed by Torr (1995) which is based on RANSAC (RANdom SAmpling Consesus) from Fischler and Bolles (1981). This approach is briefly sketched here:

```
repeat
  take minimal sample (7 matches)
  compute epipolar geometry
  estimate percentage of inliers
until satisfying solution
refine epipolar geometry (using
all inliers)
```

Once the epipolar geometry has been retrieved, one can start looking for more matches to refine this geometry. In this case, the search region is restricted to a few pixels around the epipolar lines.

### 3.2. Initial reconstruction

The two first images of the sequence are used to determine a reference frame. The world frame is typically aligned with the first camera. The second camera is chosen so that the epipolar geometry corresponds to the one retrieved in the previous section. In fact, 4 degress of freedom corresponding to the position of the plane at infinity and the global scale of the reconstruction remain (remember that for a projective reconstruction the plane that corresponds to the plane at infinity can be located anywhere). Although the precise position of this plane will only be determined in the self-calibration stage, it is interesting to avoid that it passes through the scene. This can be done by choosing its position so that all points are reconstructed in front of the camera (see Hartley, 1994 or Laveau and Faugeras, 1996). Once the camera projection matrices have been determined, the matches can be reconstructed through triangulation. The optimal method for this is given by Hartley and Sturm (1997). This gives us a preliminary reconstruction, which can be used as a calibration object for the next views.

### 3.3. Adding a view

For every additional view, the pose towards the pre-existing reconstruction is determined, then the reconstruction is updated. This is illustrated in Fig. 3. The first step consists of finding the epipolar geometry as described in Section 3.1. Then, the matches that correspond to already reconstructed points are used to compute the projection matrix. This is done using a robust procedure similar to the one laid out in Section 3.1. In this case, a minimal sample of six matches is needed to compute the camera projection matrix. Once it has been determined, the projection of already reconstructed points can be predicted. This allows to find some additional matches to refine the estimation of projection matrix. This means that the search space is gradually reduced from a large neighbourhood around the origi-

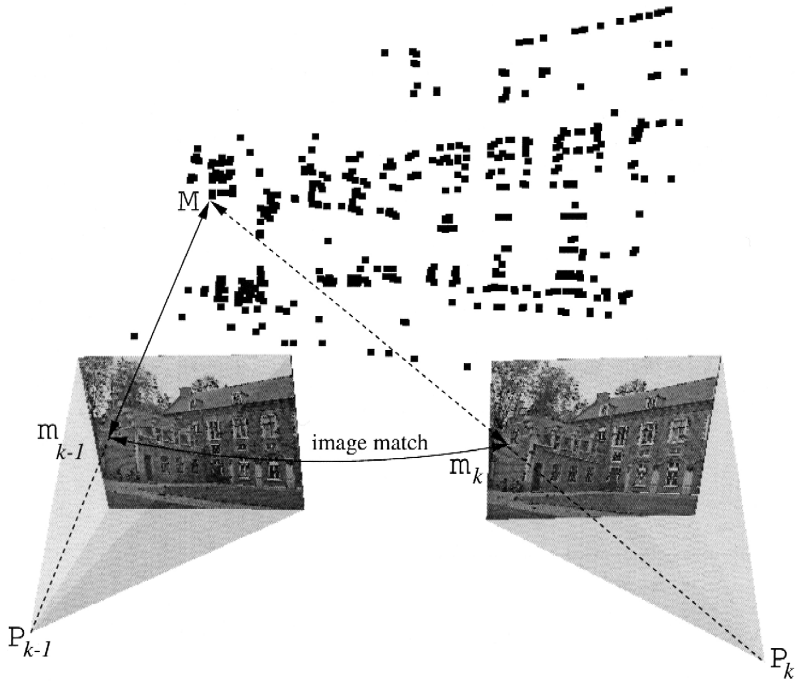


Fig. 3. Image matches $(m_{k-1}, m_k)$ are found as described before. Since the image points, $m_{k-1}$, relate to object points, $M_k$, the pose for view $k$ can be computed from the inferred matches $(M, m_k)$. $P_i$ represent the camera stations.

nal coordinates to the epipolar line and then to the predicted projection of the point. This is illustrated in Fig. 4.

Once the camera projection matrix has been determined, the reconstruction is updated. This consists of refining, correcting or deleting already reconstructed points and initialising new points for new matches. Therefore, this approach does not require that the initial feature points stay visible, since the calibration object is extended for every new view. After this procedure has been repeated for all images, one disposes of camera poses for all views and the reconstruction of the interest points. In further modules, mainly the camera calibration is used. The reconstruction itself is used to obtain an estimate of the disparity range for the dense stereo matching.

This procedure to add a view only relates the image to the previous image. In fact, it is implicitly assumed that once a point gets out of sight, it will not reappear. Although this is true for many sequences, this assumption does not always hold. When the system is used to record the appearance of an object from all around, the camera is often moved back and forth. In this case, interest points will continuously disappear and reappear, leading to a single point having multiple instances, which are not

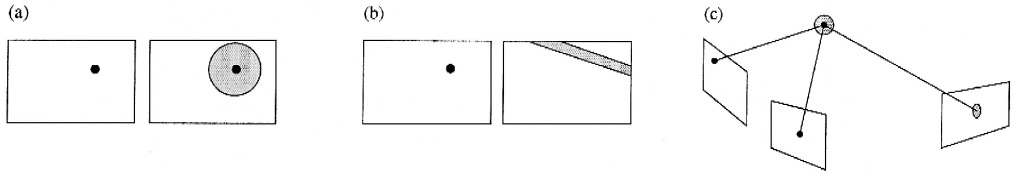


Fig. 4. (a) A priori search range, (b) search range along the epipolar line and (c) search range around the predicted position of the point.

consistent. This allows the error to accumulate over longer sequences. A possible solution (see Pollefeys, 1999) consists of not only relating the actual view with the previous one, but also with other close views. The proposed procedure goes as follows. An initial position of the camera for the actual view is estimated based on the previous view. This is then used to determine which other camera poses are close, then the matching procedure of the previous section is repeated for each one of these views. This allows the viewer to relate reappearing points to the reconstruction that was previously obtained for them. This approach was successfully applied in the context of the acquisition of image-based scene models (Koch et al., 1999).

## 4. Self-calibration

The reconstruction obtained as described in the previous paragraph is only determined up to an arbitrary projective transformation. This might be sufficient for some applications but certainly not for visualisation. The system uses the self-calibration method described by Pollefeys et al. (1999a) (see also Pollefeys, 1999) to restrict the ambiguity on the reconstruction to metric (i.e., Euclidean up to scale). This flexible self-calibration technique allows the intrinsic camera parameters to vary during the acquisition. This feature is especially useful when the camera is equipped with a zoom or with auto-focus.

It is outside the scope of this paper to discuss this method in detail. The approach is based on the concept of the absolute conic. This virtual conic is present in every scene. Once located, it allows to viewers to carry out metric measurements and thus upgrade the reconstruction from projective to metric. The problem, however, is that the absolute conic can only be observed through constraints on the intrinsic camera parameters. The approach described in Pollefeys et al. (1999a) consists of translating constraints

Fig. 5. Reconstruction before (top) and after (bottom) self-calibration.

on the intrinsic camera parameters to constraints on the absolute conic.

Some reconstructions *before* and *after* the self-calibration stage are shown. The top part of Fig. 5 gives the reconstruction before self-calibration. Therefore, it is only determined up to an arbitrary projective transformation and metric properties of the scene cannot be observed from this representation. The bottom part of Fig. 5 shows the result after self-calibration. At this point, the reconstruction has been upgraded to metric and properties such as parallelism and orthogonality can be verified.

## 5. Dense depth estimation

Only a few scene points are reconstructed from matched features. Obtaining a dense reconstruction could be achieved by interpolation, but in practice this does not yield satisfactory results. Small surface details would never be reconstructed in this procedure. Additionally, some important features are often missed during the corner matching and would therefore not appear in the reconstruction. These problems can be avoided by using algorithms that estimate correspondences for almost every point in the images. Because the reconstruction was upgraded to metric, algorithms that were developed for calibrated stereo rigs can be used.

### *5.1. Rectification*

Since we have computed the calibration between successive image pairs, we can exploit the epipolar constraint that restricts the correspondence search to a 1D search range. It is possible to re-map the image pair so that the epipolar lines coincide with the image scan lines (Koch, 1996). The correspondence search is then reduced to a matching of the image points along each image scan-line. This results in a dramatic increase of the computational efficiency of the algorithms by enabling several optimisations in the computations.

For some motions (i.e., when the epipole is located in the image), standard rectification based on planar homographies is not possible because it would result in infinitely large images. Since, in our case, the motion is not restricted, another approach should be used. The system described in this paper uses a new approach (Pollefeys et al., 1999b). The method combines simplicity with minimal image size and works for all possible motions. The key idea consists of reparameterising the image in polar coordinates around the epipole. Minimal image size is guaranteed by selecting the distance between two consecutive epipolar lines so that the area of every pixel is at least preserved. Since the matching ambiguity is restricted to half epipolar lines (Laveau and Faugeras, 1996), all that is needed is the oriented epipolar geometry. This can be easily extracted from the computed camera projection matrices. As an example, a rectified image pair from the castle is shown for both the standard technique and our new generalised technique. Fig. 6 shows the rectified image pair for both methods. A second example shows that the methods work properly when the epipole is in the image. In Fig. 7, the original and the rectified image pairs are shown. Note that in this case the standard rectification procedure cannot deliver rectified images.

Fig. 6. Rectified image pair for both methods: standard homography based method (left), new method (right).

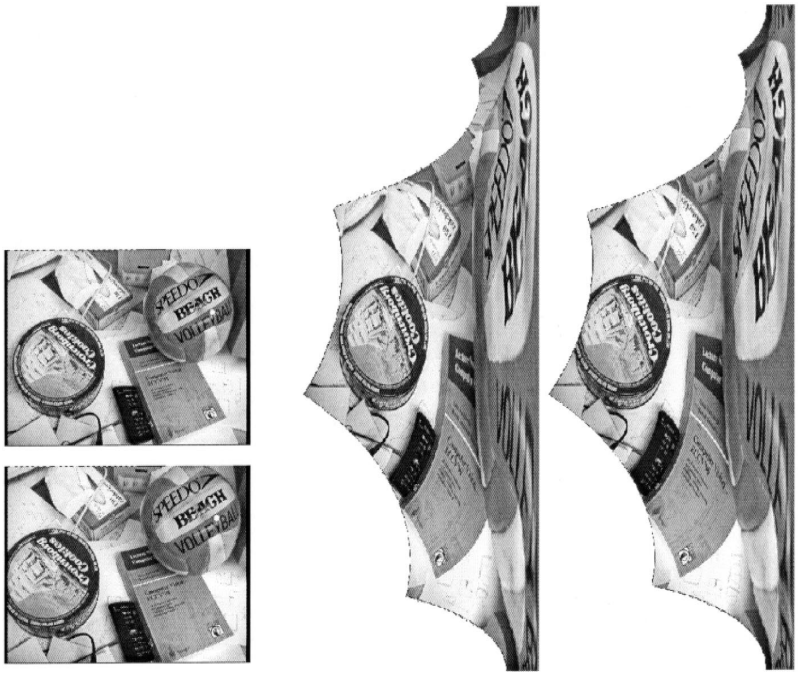


Fig. 7. Original image pair (left; epipole marked by white dot) and rectified image pair (right; epipole corresponds to right image borders).

### 5.2. Dense stereo matching

In addition to the epipolar geometry, other constraints like preserving the order of neighbouring pixels, bidirectional uniqueness of the match, and detection of occlusions can be exploited. These constraints are used to guide the correspondence towards the most probable scan-line match using a dynamic programming scheme (Cox et al., 1996). This approach operates on rectified image pairs and is illustrated in Fig. 8. The matcher searches at each pixel in one image for maximum normalised cross-correlation in the other image by shifting a small measurement window (kernel size $5 \times 5$ to $7 \times 7$ pixels) along the corresponding scan line. Matching ambiguities are resolved by exploiting the ordering constraint in the dynamic programming approach (see Koch, 1996). The algorithm was further adapted to

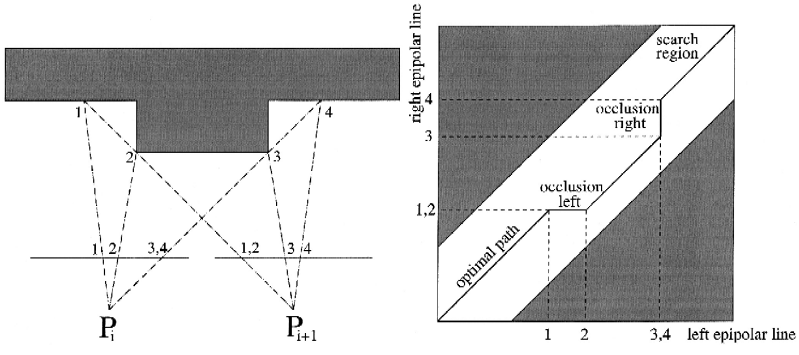


Fig. 8. Illustration of the ordering constraint (left). Dense matching as a path search problem (right).

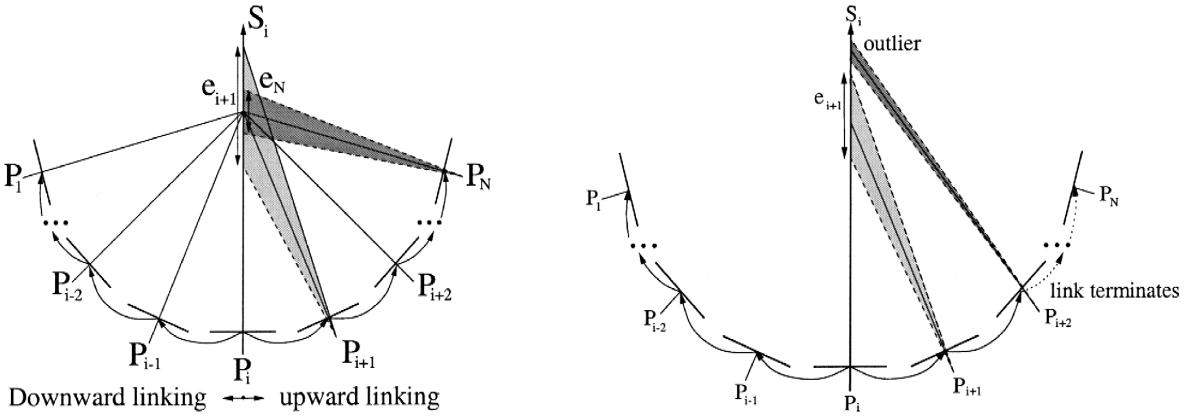


Fig. 9. Depth diffusion and uncertainty reduction from correspondence linking (left), linking stops when an outlier is encountered (right). $P_i$ represent the camera stations, $e_i$ the uncertainty range and $S_i$ the line of sight for a specific pixel. Downward and upward linking corresponds to images before and after the current image, respectively.

employ extended neighbourhood relationships and a pyramidal estimation scheme to reliably deal with very large disparity ranges of over 50% of image size.

### *5.3. Multi-view matching*

The pairwise disparity estimation allows to compute 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. The fusion can be performed in an economical way through controlled correspondence linking (see Fig. 9). The approach utilises a flexible multi-viewpoint scheme, which combines the advantages of small baseline and wide baseline stereo (see Koch et al., 1998). The result of this procedure is a very dense depth map (see Fig. 10). Most occlusion problems are avoided by linking correspondences from preceding and succeeding images in the sequence.

## 6. Building the model

The dense depth maps, as computed by the correspondence linking, must be approximated by a 3D

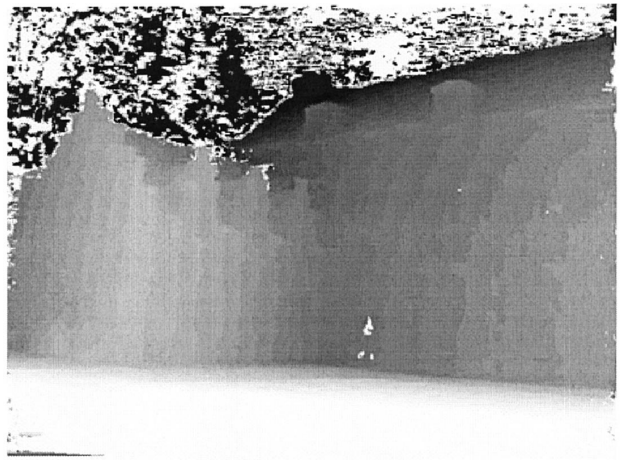

Fig. 10. Dense depth map (cf. second image of Fig. 2; light means near and dark means far).

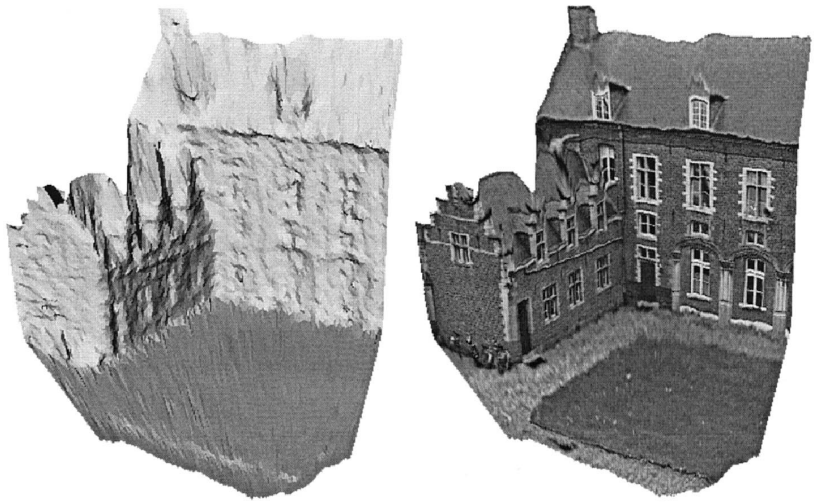

Fig. 11. Resulting 3D surface model, shaded (left), textured (right).

surface representation suitable for visualisation. So far, each object point was treated independently. To achieve spatial coherence for a connected surface, the depth map is spatially interpolated using a parametric surface model. The boundaries of the objects to be modelled are computed through depth segmentation. In a first step, an object is defined as a connected region in space. Simple morphological filtering removes spurious and very small regions. Then, a bounded thin plate model is employed with a second-order spline to smooth the surface and to interpolate small surface gaps in regions that could not be measured.

The spatially smoothed surface is then approximated by a triangular wire-frame mesh to reduce geometric complexity and to tailor the model to the requirements of computer graphics visualisation systems. The mesh triangulation currently utilises the reference view only to build the model. The surface fusion from different viewpoints to completely close the models remains to be implemented.

Texture mapping onto the wire-frame model greatly enhances the realism of the models. As texture map, one could take the reference image texture alone and map it to the surface model. However, this creates a bias towards the selected image and imaging artifacts like sensor noise, unwanted specular reflections or the shading of the particular image are directly transformed onto the object. A better choice is to fuse the texture from the image sequence in much the same way as depth fusion. The viewpoint linking builds a controlled chain of correspondences that can be used for texture enhancement as well. The estimation of a robust mean texture will capture the static object only while the artifacts (e.g., specular reflections or pedestrians passing in front of a building) are suppressed. An example of the resulting model can be seen in Fig. 11.

## 7. Some examples

The 3D surface acquisition technique that was presented in the previous sections, can readily be applied to archaeological sites. This is illustrated in this section with some results from the archaeological site of Sagalassos (Turkey). The on-site acquisition procedure consists of recording an image sequence of the scene that the viewer desires to *virtualise*, making sure that everything that should be modelled is seen in at least two images. To allow the algorithms to yield good results, viewpoint changes between consecutive images should not exceed 5–10°. An example of such a sequence is given in Fig. 12. The subsequent processing is fully automatic. The result for the image sequence under consideration can be seen in Fig. 13. An important advantage is that details like missing stones, not

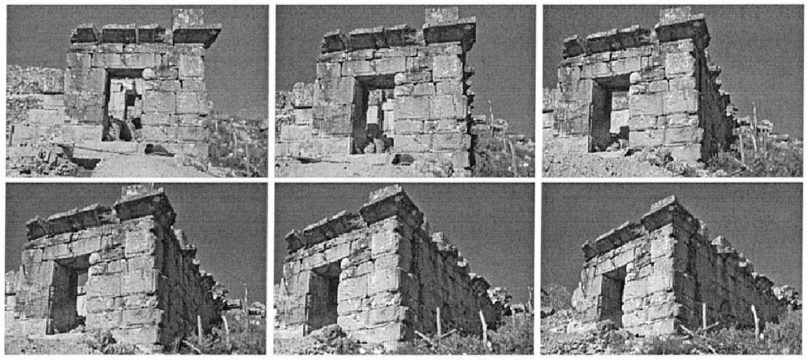

Fig. 12. Image sequence which was used to build a 3D model of the corner of the Roman baths.

perfectly planar walls or symmetric structures are preserved. In addition, the surface texture is directly extracted from the images. This does not only result in a much higher degree of realism, but is also important for the authenticity of the reconstruction. Therefore, the reconstructions obtained with this system can also be used as a scale model on which measurements can be carried out or as a tool for planning restorations.

### *7.1. Building a virtual 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 Fig. 14. The model was created from nine 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. Absolute localisation could be achieved by localising 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 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 recon-

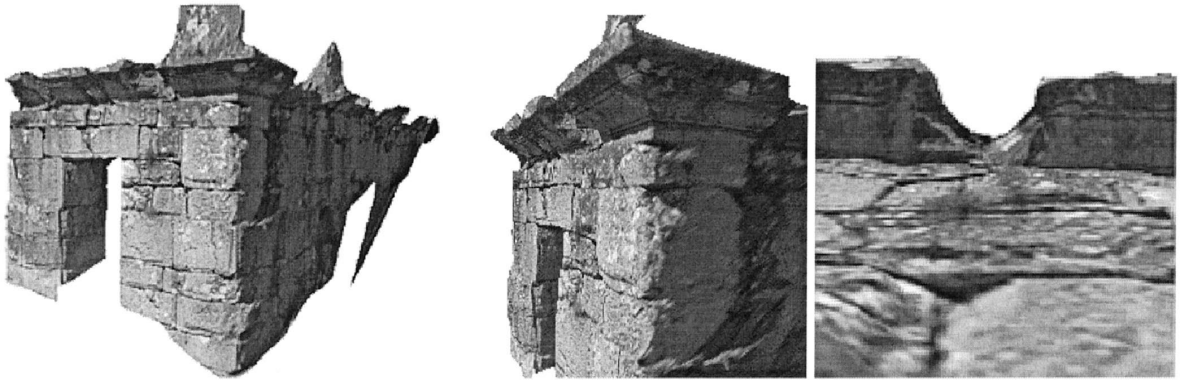

Fig. 13. Virtualised corner of the Roman baths. On the right panel, some details are shown.

Fig. 14. Overview model of Sagalassos.

structions of the site at different scales, going from a global reconstruction of the whole site to a detailed reconstruction for every monument. These reconstructions thus naturally fill-in the different levels of details, which should be provided for optimal rendering.

An interesting possibility is the combination of these models with other types of models. In the case of Sagalassos, some building hypothesis were translated to CAD models (Martens et al., 2000). These were integrated with our models. The result can be seen in Fig. 15. Other models obtained with different 3D acquisition techniques could also be easily integrated. These reconstructions are available on the Internet (see Virtual Sagalassos, 2000).

### *7.2. Other applications*

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

#### *7.2.1. 3D stratigraphy*

Archaeology is one of the sciences where annotations and precise documentation are most important because evidence can be destroyed during work. An

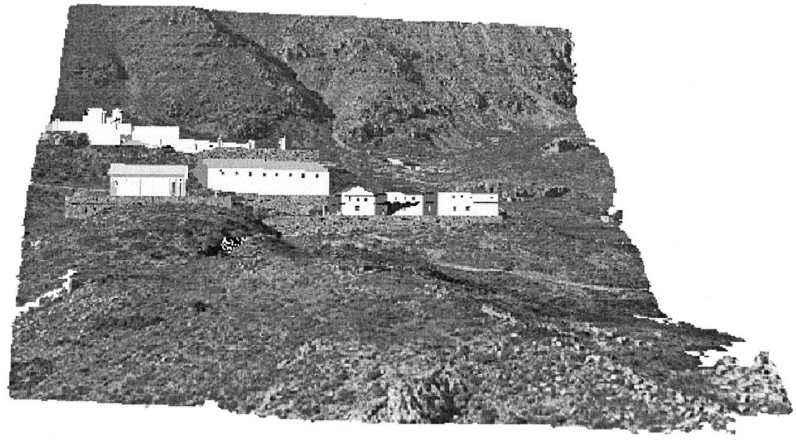

Fig. 15. Virtualised landscape of Sagalassos combined with CAD models of reconstructed monuments.

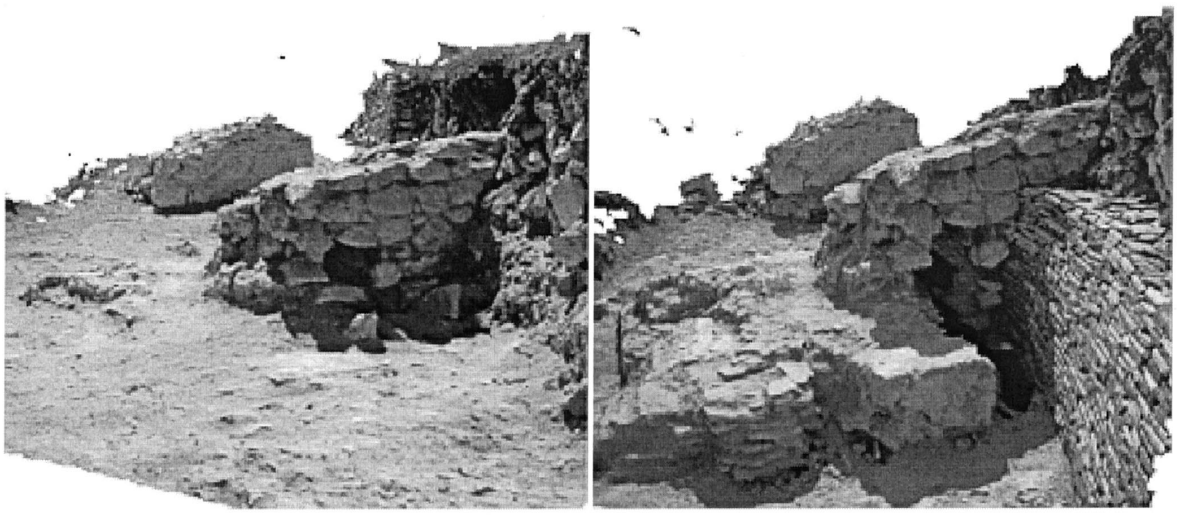

Fig. 16. 3D stratigraphy, the excavation of a Roman villa at two different moments.

important aspect of this is 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 for a more optimal approach. For every layer, a complete 3D model of the excavated sector can be generated. Since this only involves taking a series of pictures, this does not slow down the progress of the archaeological work. In addition, it is possible to model separately artifacts found in these layers and to include the models in the final 3D stratigraphy. The excavations of an ancient Roman villa at Sagalassos were recorded with our technique. In Fig. 16, a view of the 3D model of the excavation is provided for two different layers. The on-site acquisition time was around 1 min per layer.

*7.2.2. Generating and testing building hypothesis*

The technique proposed in this paper has also 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. Then, the different construction hypothesis can be verified interactively at a virtual building site. In addition, registration algorithms (Chen and Medioni, 1991; Vanden Wyngaerd et al., 1999) could be used to automate this process. Fig. 17 shows an example.

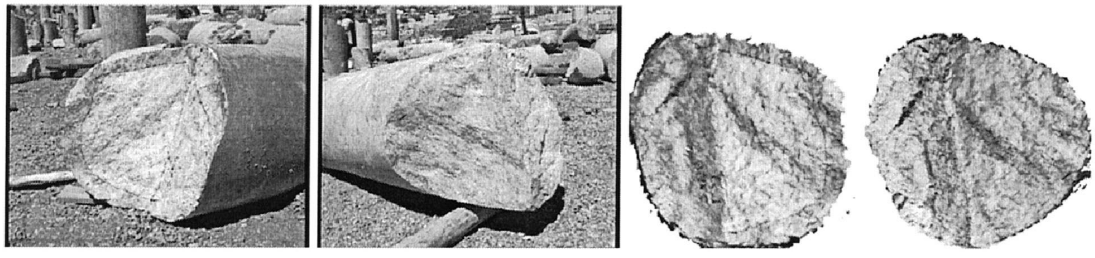

Fig. 17. Two images of a broken pillar (left) and the orthographic views of the matching surfaces generated from the obtained 3D models (right).

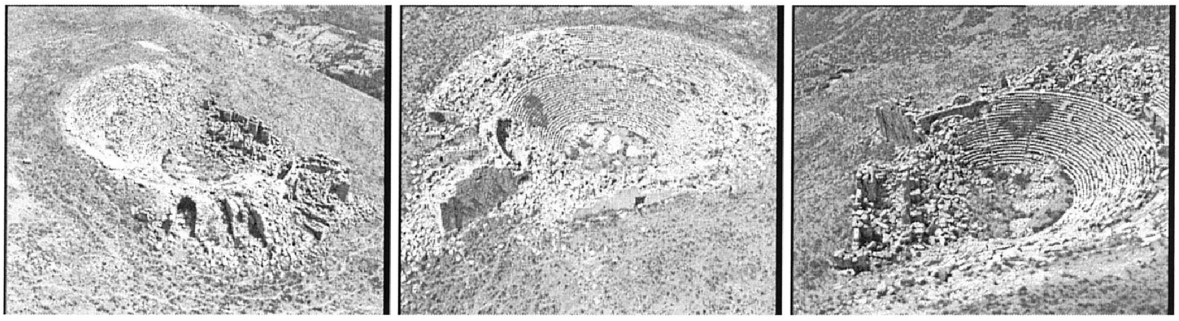

Fig. 18. Three images of the helicopter shot of the ancient theater of Sagalassos.

*7.2.3. Reconstruction from archives*

The flexibility of the proposed approach makes it possible to use existing photo or video archives to reconstruct objects. This application is very interesting for monuments or sites that have been destroyed due to war or natural disasters. The feasibility of this type of reconstruction is illustrated with a reconstruction of the ancient theater of Sagalassos based on a sequence filmed by the Belgian TV in 1990 to illustrate a documentary on Sagalassos. From the 30-s helicopter shot, approximately hundred images were extracted. Because of the motion, only image fields—not frames—could be used, which restricted the vertical resolution to 288 pixels. Three images of the sequence are shown in Fig. 18. The reconstruction of the feature points together with the recovered camera poses are shown in Fig. 19.

## 8. Conclusion and further research

An automatic 3D scene modelling technique, which is capable of building models from uncalibrated image sequences, is presented and discussed. The technique is able to extract detailed metric 3D models without prior knowledge about the scene or the camera. This technique was successfully applied to the acquisition of virtual models of archaeological

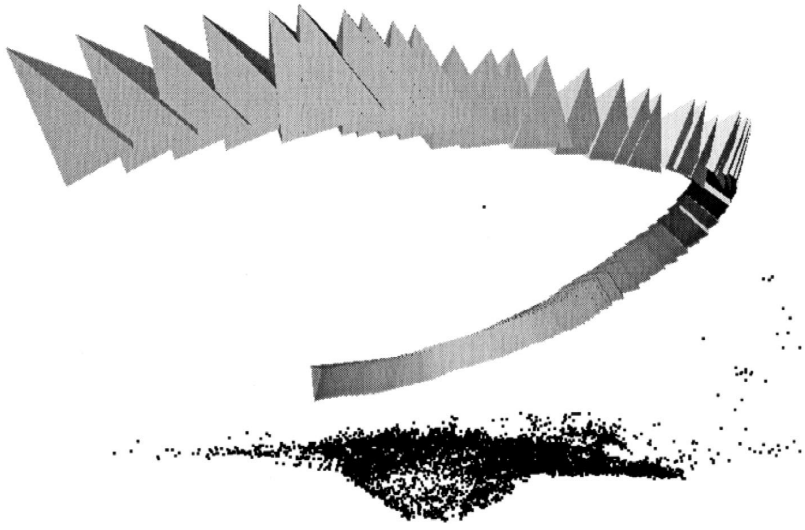

Fig. 19. The reconstructed feature points and camera poses recovered from the helicopter shot.

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

Our main goal, while developing this approach, was to show the feasibility of 3D reconstruction from a hand-held camera. Future research will consist of improving the approach and developing it further. To increase the accuracy of this technique, the simple pinhole camera model will be extended to take distortions into account. Another improvement will consist of integrating partial models into a single global reconstruction through volumetric techniques.

## Acknowledgements

We would like to thank Prof. Marc Waelkens from the Sagalassos Archaeological Research Project (K.U. Leuven) for inviting us to Sagalassos and for his collaboration, Pol and Jos Legrand for the CAD models and Joris Vanden Wyngaerd for the integration of the CAD models with our models. The first author acknowledges the Fund for Scientific Research—Flanders for a Postdoctoral Fellowship. Financial support of the IWT (Institute for the Promotion of Innovation by Science and Technology, Flanders) STWW VirtErf and the European Murale IST-1999-20273 project are also gratefully acknowledged.

## References

Beardsley, P., Zisserman, A., Murray, D., 1997. Sequential updating of projective and affine structure from motion. Int. J. Comput. Vision 23 (3), 235–259.

Chen, Y., Medioni, G., 1991. Object modeling by registration of multiple range images. Proc. IEEE Int. Conf. Rob. Autom. 2724–2729.

Cox, I., Hingorani, S., Rao, S., 1996. A maximum likelihood stereo algorithm. Comput. Vision Image Understanding 63 (3), 542–567.

Debevec, P., Taylor, C.J., Malik, J., 1996. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. Proc. of SIGGRAPH 96, 11–20.

Dorffner, L., Forkert, G., 1998. Generation and visualization of 3D photo-models using hybrid block adjustment with assumptions on the object shape. ISPRS J. Photogramm. Remote Sens. 53 (6), 369–378.

El-Hakim, S., Brenner, C., Roth, G., 1998. A multi-sensor approach to creating accurate virtual environments. ISPRS J. Photogramm. Remote Sens. 53 (6), 379–391.

Falkenhagen, L., 1995. Depth estimation from stereoscopic image pairs assuming piecewise continuous surfaces. In: Paker, Y., Wilbur, S. (Eds.), Image Processing for Broadcast and Video Production. Springer Ser. Workshops Comput. Springer-Verlag, London, pp. 115–127.

Faugeras, O., 1992. What can be seen in three dimensions with an uncalibrated stereo rig. Computer Vision–ECCV'92. Lect. Notes Comput. Sci. vol. 588, Springer-Verlag, Berlin, pp. 563–578.

Faugeras, O., Luong, O., Maybank, Q.-T., 1992. Camera self-calibration: theory and experiments. Computer Vision–ECCV'92. Lect. Notes Comput. Sci. vol. 588, Springer-Verlag, Berlin, pp. 321–334.

Fischler, M., Bolles, R., 1981. RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography. Commun. Assoc. Comp. Mach. 24 (6), 381–395.

Fraser, C., 1997. Digital camera self-calibration. ISPRS J. Photogramm. Remote Sens. 52 (4), 149–159.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. Proc. 4th Alvey Vision Conference, 147–151.

Hartley, R., 1994. Euclidean reconstruction from uncalibrated views. Applications of Invariance in Computer Vision. In: Mundy, J.L., Zisserman, A., Forsyth, D. (Eds.), Lect. Notes Comput. Sci. vol. 825, Springer-Verlag, Berlin, pp. 237–256.

Hartley, R., Sturm, R., 1997. Triangulation. Comput. Vision Image Understanding 68 (2), 146–157.

Hartley, R., Gupta, R., Chang, T., 1992. Stereo from uncalibrated cameras. Proc. International Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Press, Los Alamitos, CA, pp. 761–764.

Koch, R., 1996. Automatische Oberflachenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten. PhD thesis, Faculty of Electrical Engineering and Information Technology, University of Hannover, Germany. Also published as Fortschritte-Berichte VDI, Reihe 10, Nr. 499, VDI Verlag, 1997.

Koch, R., Pollefeys, M., Van Gool, L., 1998. Multi Viewpoint stereo from uncalibrated video sequences. Computer Vision–ECCV'98. Lect. Notes Comput. Sci. vol. 1406, pp. 55–71.

Koch, R., Pollefeys, M., Heigl, B., Van Gool, L., Niemann, H., 1999. Calibration of hand-held camera sequences for plenoptic modeling. Proc. International Conference on Computer Vision. IEEE Computer Society Press, Los Alamitos, CA, pp. 585–591.

Laveau, S., Faugeras, O., 1996. Oriented projective geometry for computer vision. Computer Vision–ECCV'96. Lect. Notes Comput. Sci. vol. 1064, Springer-Verlag, Berlin, pp. 147–156.

Martens, F., Legrand, J., Legrand, P., Loots, L., Waelkens, M., 2000. Computer aided design and archeology at Sagalassos: methodology and possibilities of CAD reconstructions of ar-

chaeological sites. In: Barcelo, J.A., Forte, M., Sanders, D. (Eds.), Virtual Reality in Archaeology. ArcheoPress, Oxford, pp. 205–212.

PhotoModeler. (accessed 25 September 2000).

Pollefeys, M., 1999. Self-calibration and metric 3D reconstruction from uncalibrated image sequences. PhD Thesis, Center for Processing of Speech and Images (PSI), Dept. of Electrical Engineering, Faculty of Applied Sciences, K.U. Leuven, available at http://www.esat.kuleuven.ac.be/~pollefey/publications/PhD.html ( accessed 25 September 2000).

Pollefeys, M., Van Gool, L., 1999. Stratified self-calibration with the modulus constraint. IEEE Trans. Pattern Anal. Mach. Intell. 21 (8), 707–724.

Pollefeys, M., Koch, R., Van Gool, L., 1999a. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. Int. J. Comput. Vision 32 (1), 7–25.

Pollefeys, M., Koch, R., Van Gool, L., 1999b. A simple and efficient rectification method for general motion. Proc. International Conference on Computer Vision. IEEE Computer Society Press, Los Alamitos, CA, pp. 496–501.

Sequeira, V., Ng, K., Wolfart, E., Gonçalves, J.G.M., Hogg, D., 1999. Automated reconstruction of 3D models from real environments. ISPRS J. Photogramm. Remote Sens. 54 (1), 1–22.

Sturm, P., 1997. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. Proc. International Conference on Computer Vision and Pattern Recognition. IEEE Computer Soc. Press, Los Alamitos, CA, pp. 1100–1105.

Tomasi, C., Kanade, T., 1992. Shape and motion from image streams under orthography: a factorization approach. Int. J. Comput. Vision 9 (2), 137–154.

Torr, P., 1995. Motion segmentation and outlier detection. PhD Thesis, Dept. of Engineering Science, University of Oxford, 1995.

Triggs, B., 1997. The absolute quadric. Proc. International Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Press, Los Alamitos, CA, pp. 609–614.

Vanden Wyngaerd, J., Van Gool, L., Koch, R., Proesmans, M., 1999. Invariant-based registration of surface patches. Proc. International Conference on Computer Vision. IEEE Computer Society Press, Los Alamitos, CA, pp. 301–306.

Virtual Sagalassos. (accessed 25 September 2000).

Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.-T., 1995. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artif. Intell. J. 78 (1-2), 87–119.