Automatic 3D Model Acquisition from Uncalibrated Image Sequences

Reinhard Koch, Marc Pollefeys, and Luc Van Gool
K.U.Leuven, Dept. Elektrotechniek, ESAT-PSI
Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium
{firstname.lastname}@esat.kuleuven.ac.be

Abstract

In this paper the problem of obtaining 3D models from image sequences is addressed. The proposed method deals with uncalibrated monocular image sequences. No prior knowledge about the scene or about the camera is necessary to build the 3D models. The only assumptions are the rigidity of the scene objects and opaque object surfaces.

The modeling system uses a 3-step approach. First, the camera pose and intrinsic parameters are calibrated by tracking salient feature points throughout the sequence. Next, consecutive images of the sequence are treated as stereoscopic image pairs and dense correspondence maps are computed by area matching. Finally, dense and accurate depth maps are computed by linking together all correspondences over the viewpoints. The depth maps are converted to triangular surfaces meshes that are texture mapped for photo-realistic appearance. The feasibility of the approach has been tested on both real and synthetic data and is illustrated here on several outdoor image sequences.

1. Introduction

The use of three-dimensional surface models for the purpose of visualization is gaining importance. Highly realistic 3D models are readily used to visualize and simulate events, like in flight simulators, in the games and film industry or for product presentations. The range of applications span from architecture visualization over virtual television studios, virtual presence for video communications to general "virtual reality" applications.

A limitation to the widespread use of these techniques is currently the high cost of such 3D models since they have to be produced manually. Especially if existing objects are to be reconstructed the measurement process for obtaining the correct geometric and photometric data is tedious and time consuming. Traditional solutions include the use of stereo rigs, laser range scanners and other 3D digitizing devices. These devices are often very expensive, require careful handling and complex calibration procedures and are designed for a restricted depth range only.

In this work an image based approach is proposed which avoids most of the problems mentioned above. The scene which has to be modeled is recorded from different viewpoints by a video camera. The relative position and orientation of the camera and its calibration parameters will automatically be retrieved from the image data by the algorithms. Hence, there is no need for measurements in the scene or calibration procedures whatsoever. There is also no restriction on range, it is just as easy to model a small object, as to model a complete landscape. The proposed method thus offers a previously unknown flexibility in 3D model acquisition. In addition, any photographic recording device - e.g. camcorder, digital camera, or even standard photographic film camera - is sufficient for scene acquisition. Hence, increased flexibility is accompanied by a decrease in cost.

The proposed method is placed in the framework of uncalibrated scene reconstruction that is a recent research topic in the computer vision society. In the uncalibrated case all parameters, camera pose and intrinsic calibration as well as the 3D scene structure have to be estimated from the 2D image sequence alone. Faugeras and Hartley first demonstrated how to obtain uncalibrated projective reconstructions from image sequences alone [6, 8]. Since then, researchers tried to find ways to upgrade these reconstructions to metric (i.e. Euclidean but unknown scale, see [7, 20, 17]). Newest results report full self-calibration methods even for varying intrinsic parameters like focal length, which allows the unrestricted use of the camera, for example zooming [9, 16, 18].

To employ these self-calibration methods for sequence analysis they must be embedded in a complete scene reconstruction system. Beardsley et al. [1] proposed a scheme to obtain projective calibration and 3D structure by robustly tracking salient feature points throughout an image sequence. This sparse object representation outlines the object shape, but gives not sufficient surface detail for visual

reconstruction. Highly realistic 3D surface models need the dense depth estimation and can not rely on few feature points alone.

In [18] the method of Beardsley e.a. was extended in two directions. On the one hand the projective reconstruction was updated to metric even for varying internal camera parameters, on the other hand a dense stereo matching technique [4] was applied between two selected images of the sequence to obtain a dense depth map for a single viewpoint. From this depth map a triangular surface wire-frame was constructed and texture mapping from one image was applied to obtain realistic surface models. In [14] the approach was further extended to multi viewpoint sequence analysis.

In this contribution we will discuss the complete and automatic modeling system that is capable to compute accurate and dense 3D surface models from uncalibrated image sequences. Section 2 discusses the steps needed for depth estimation from image sequences. It involves feature point tracking for projective calibration, metric self-calibration, dense depth map estimation and sequence fusion. Section 3 deals with the 3D model generation and the creation of textured surfaces. In sect. 4 several experiments on real outdoor sequences are performed. Objects of different scale are modeled and different imaging sensors are used to demonstrate the quality and flexibility of the proposed reconstruction system.

2. Dense Depth Estimation from Image Sequences

Accurate and robust depth estimation is the key problem in a 3D modeling system. It is solved using a 3-step approach:

- Camera self-calibration and metric structure is obtained by robust tracking of salient feature points over the image sequence.
- Dense correspondence maps are computed between adjacent image pairs of the sequence.
- All correspondence maps are linked together by multiple view point linking to fuse depth measurements over the sequence.

2.1. Camera Calibration through Feature Point Tracking

Camera calibration[1] is obtained by tracking salient image features throughout the sequence. The difficulty of this step is to robustly find at least a few but very reliable correspondences that are needed for camera calibration. Salient feature points like strong intensity corners are matched using robust (RANSAC) techniques for that purpose. In a two-step procedure a projective calibration and feature point reconstruction is recovered from the image sequence which is then updated to metric calibration with a self-calibration approach.

Retrieving the projective framework. At first feature correspondences are found by extracting intensity corners in different images and matching them using a robust corner matcher [19]. In conjunction with the matching of the corners a restricted calibration of the setup is calculated (i.e. only determined up to an arbitrary projective transformation). This allows to eliminate matches which are inconsistent with the calibration. The 3D position of a point is restricted to the line passing through its image point and the camera projection center. Therefore the corresponding point is restricted to the projection of this line in the other image. Using this constraint, more matches can easily be found and used to refine this calibration.

The matching is started on the first two images of the sequence. The calibration of these views define a projective framework in which the projection matrices of the other views are retrieved one by one. In this approach we follow the procedure proposed by Beardsley *et al* [1]. We therefore obtain projection matrices (3×4) of the following form:

$$\mathbf{P}_1 = [\mathbf{I}|0] \text{ and } \mathbf{P}_k = [\mathbf{H}_{1k}|e_{1k}] \tag{1}$$

with $\mathbf{H}_{1k}$ the homography for some reference plane from view 1 to view k and e_{1k} the corresponding epipole.

Retrieving the metric framework. Such a projective calibration is certainly not satisfactory for the purpose of 3D modeling. A reconstruction obtained up to a projective transformation can differ very much from the original scene according to human perception: orthogonality and parallelism are in general not preserved, part of the scene can be warped to infinity, etc. To obtain a better calibration, constraints on the internal camera parameters can be imposed (e.g. absence of skew, known aspect ratio, ...). By exploiting these constraints, the projective reconstruction can be upgraded to metric (Euclidean up to scale). In the metric case the camera projection matrices have the following form:

$$\mathbf{P}_i = \mathbf{K}_i[\mathbf{R}_i|\text{-}\mathbf{R}_i\mathbf{t}_i] \text{ with } \mathbf{K}_i = \begin{bmatrix} f_x & s & u_x \\ & f_y & u_y \\ & & 1 \end{bmatrix} \tag{2}$$

where $\mathbf{R}_i$ and $\mathbf{t}_i$ indicate the orientation and position of the camera for view i and $\mathbf{K}_i$ contains the internal camera parameters: f_x and f_y stand for the horizontal and vertical

[1] By *calibration* we mean the actual internal calibration of the camera as well as the relative position and orientation of the camera for the different views with respect to an arbitrary coordinate system.

focal length (in pixels), $\mathbf{u} = (u_x, u_y)$ is the principal point and s is a measure of the image skew.

A practical way to obtain the calibration parameters from constraints on the internal camera parameters is through application of the concept of the absolute quadric [20, 18]. In space, exactly one degenerate quadric of planes exists which has the property to be invariant under all rigid transformations. In a metric frame it is represented by the following 4×4 symmetric rank 3 matrix $\Omega = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix}$. If $\mathbf{T}$ transforms points $M \rightarrow \mathbf{T}M$ (and thus $\mathbf{P} \rightarrow \mathbf{P}\mathbf{T}^{-1}$), then it transforms $\Omega \rightarrow \mathbf{T}\Omega\mathbf{T}^{\top}$ (which can be verified to yield Ω when $\mathbf{T}$ is a similarity transformation). The projection of the absolute quadric in the image yields the intrinsic camera parameters independent of the chosen projective basis[2]:

$$\mathbf{K}_i\mathbf{K}_i^{\top} \propto \mathbf{P}_i\Omega\mathbf{P}_i^{\top} \tag{3}$$

where $\propto$ means equal up to an arbitrary non-zero scale factor. Therefore constraints on the internal camera parameters in $\mathbf{K}_i$ can be translated to constraints on the absolute quadric. If enough constraints are at hand, only one quadric will satisfy them all, i.e. the *absolute quadric*. At that point the scene can be transformed to the metric frame (which brings Ω to its canonical form).

2.2. Dense Correspondence Matching

Only a few scene points are reconstructed from feature tracking. Obtaining a dense reconstruction could be achieved by interpolation, but in practice this does not yield satisfactory results. Often some important features are missed during the corner matching and will therefore not appear in the reconstruction.

These problems can be avoided by using algorithms which estimate correspondences for almost every point in the images. At this point algorithms can be used which were developed for calibrated stereo rigs. Since we have computed the calibration between successive image pairs we can exploit the epipolar constraint that restricts the correspondence search to a 1-D search range. In particular it is possible to re-map the image pair to standard geometry with the epipolar lines coinciding with the image scan lines [13]. The correspondence search is then reduced to a matching of the image points along each image scan-line. In addition to the epipolar geometry other constraints like preserving the order of neighboring 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 [4].

[2]Using Equation 2 this can be verified for a metric basis. Transforming $\mathbf{P} \rightarrow \mathbf{P}\mathbf{T}^{-1}$ and $\Omega \rightarrow \mathbf{T}\Omega\mathbf{T}^{\top}$ will not change the projection.

For dense correspondence matching a disparity estimator based on the dynamic programming scheme of Cox *et al.* [3], is employed that incorporates the above mentioned constraints. It operates on rectified image pairs (I_i, I_k) where the epipolar lines coincide with image scan lines. The matcher searches at each pixel in image I_i for maximum normalized cross correlation in I_k by shifting a small measurement window (kernel size 5x5 to 7x7 pixel) along the corresponding scan line. The selected search step size ΔD (usually 1 pixel) determines the search resolution. Matching ambiguities are resolved by exploiting the ordering constraint in the dynamic programming approach [13]. The algorithm was further adapted to employ extended neighborhood relationships and a pyramidal estimation scheme to reliably deal with very large disparity ranges of over 50% of image size [4].

2.3. Sequence Linking

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. The approach utilizes a flexible multi viewpoint scheme by combining the advantages of small baseline and wide baseline stereo [14].

As *small baseline stereo* we define viewpoints with a baseline much smaller than the observed average scene depth. This configuration is usually valid for image sequences were the images are taken as a spatial sequence from many slightly varying view points. The advantages are an easy correspondence estimation and small regions of occlusion[3] between adjacent images. Disadvantage is clearly the limited depth resolution due to the small triangulation angle between the view points.

The *wide baseline stereo* in contrast is used mostly with still image photographs of a scene where few images are taken from a very different viewpoint. Here the depth resolution is superior but correspondence and occlusion problems appear, because the views are very different and large image regions without correspondence may occur.

The *multi viewpoint linking* combines the virtues of both approaches by concatenating corresponding points over multiple images. In addition it will produce denser depth maps than either of the other techniques, and allow for additional features during depth and texture fusion [14]. In the linking process care is taken to deal with occlusions and to check for measurement outliers.

[3]As occlusions we consider those parts of the object that are visible in one image only, due to object self-occlusion.

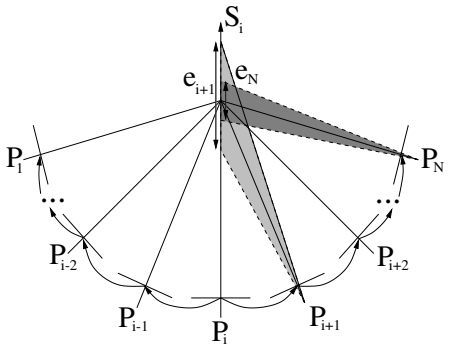

Figure 1. Depth fusion and uncertainty reduction from correspondence linking.

2.3.1 Depth fusion

Assume an image sequence with $i = 1 \rightarrow N$ images. Starting from a reference view point i the correspondences between adjacent images $(i+1, i+2, ..., N)$ and $(i-1, i-2, ..., 1)$ are linked in a chain. The depth for each reference image point $\mathbf{x_i}$ is computed from the correspondence linking that delivers two lists of image correspondences relative to the reference, one linking down from $i \rightarrow 1$ and one linking up from $i \rightarrow N$. For each valid corresponding point pair $(\mathbf{x_i}, \mathbf{x_k})$ we can triangulate a depth estimate $d(x_i, x_k)$ along S_{xi} with e_k representing the depth uncertainty. Figure 1 visualizes the decreasing uncertainty interval during linking.

While the disparity measurement resolution ΔD in the image is kept constant (at 1 pixel), the reprojected depth error e_k decreases with the baseline. Outliers are detected by controlling the statistics of the depth estimate computed from the correspondences. All depth values that fall within the uncertainty interval around the mean depth estimate are treated as inliers. They are fused by a 1-D kalman filter to obtain an optimal mean depth estimate. Outliers are undetected correspondence failures and may be arbitrarily large. As threshold to detect the outliers we utilize the depth uncertainty interval e_k.

2.3.2 Occlusions

If an object region is visible in view point i but not in k, we speak of an occlusion. Occlusions are eliminated by incorporating a multi viewpoint matcher that operates symmetrically to a particular viewpoint i. Points that are occluded in the view $i+1$ are normally visible in the view $i-1$ and vice versa. The exploitation of links starting up and down from viewpoint i will resolve for most of the occlusions and will produce a very dense depth map.

3. Surface Modeling

The dense depth maps as computed by the correspondence linking must be approximated by a 3D surface representation suitable for visualization. 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 modeled 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. We then employ a bounded thin plate model with a second order spline to smooth the surface and to interpolate small surface gaps in regions that could not be measured. If the object consist of dominant planar regions, the local surface normal may be exploited to segment the object into planar parts [12].

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 visualization systems. The mesh triangulation currently utilizes the reference view only to build the model. The surface fusion from different view points to completely close the models remains to be implemented [11]. Sometimes it is not possible to obtain a single metric framework for large objects like buildings since one may not be able to record images continuously around it. In that case the different frameworks have to be registered to each other. This will be done using available surface registration schemes [2].

3.1. Texture Fusion

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 is 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. A texture map in this context is defined as the color intensity values for a given set of image points, usually the pixel coordinates. While depth is concerned with the *position* of the correspondence in the image, texture uses the *color intensity value* of the corresponding image point. For each reference image position one may now collect a list of color intensity values from the corresponding image positions in the other viewpoints. This allows to enhance the original texture in many ways by accessing the color statis-

tics. Some features that can be derived naturally from the texture linking algorithm are described below.

Specular reflection and artifact removal. The surface reflectance of the object is modeled by a viewpoint independent diffuse and a viewpoint dependent specular reflection. In this case the color intensity statistics can be modeled as Gaussian noise contaminated with an outlier tail distribution that contains the reflection. By collecting the corresponding color intensities over a series of different viewpoints one can detect the specular reflectance as outlier and retain the diffuse reflection using median filtering. The same statistics hold if a fast moving object temporarily occludes the observed object, like a pedestrian passing in front of a building to be modeled. The exploitation of a robust mean texture will therefore capture the static object only and the artifacts are suppressed [10].

Super-resolution texture. The correspondence linking is not restricted to pixel-resolution, since each between-pixel-position in the reference image can be used to start a correspondence chain as well. Color intensity values will then be interpolated between the pixel grid. If the object is observed from many different view points and possibly from different object distances, the finite pixel grid of the images for each viewpoint is generally slightly displaced. This displacement can be exploited to create super-resolution texture by fusing all images on a finer resampling grid. The super-resolution grid in the reference image can be chosen arbitrarily fine, but the measurable real resolution is of course depending on the displacement and resolution of the corresponding images [15].

4. Experiments

In this section the performance of the modeling system is tested on the outdoor sequences *Castle*, *Pillar*, *Fountain*, *Roman Bath*, and *Site*.

4.1. Castle sequence

The *Castle* sequence consists of 22 images of 720x576 pixel resolution taken with a standard semi-professional camcorder that was moved freely in front of a building. Figure 2 shows the images 1,8,14, and 22 of the sequence.

To judge the geometric and visual quality of the reconstruction, different perspective views of the model were computed and displayed in Figure 3. In the shaded view, the geometric details like the window and door niches are seen. A close-up look from a position that a human observer would take reveals the high visual quality of the model.

A more quantitative evaluation was obtained by measuring angles in the reconstructed scene between parallel lines

Figure 2. Images 1, 8, 14, and 22 of the castle sequence.

(1.0 ± 0.6 degrees) and orthogonal lines (92.5 ± 0.4 degrees). These results[18] confirm the good metric reconstruction obtained by the method.

4.2. Pillar sequence

As an example for varying camera parameters 8 images of a stone pillar with curved surfaces were taken. Figure 4 show 2 of the recorded images. While filming and moving

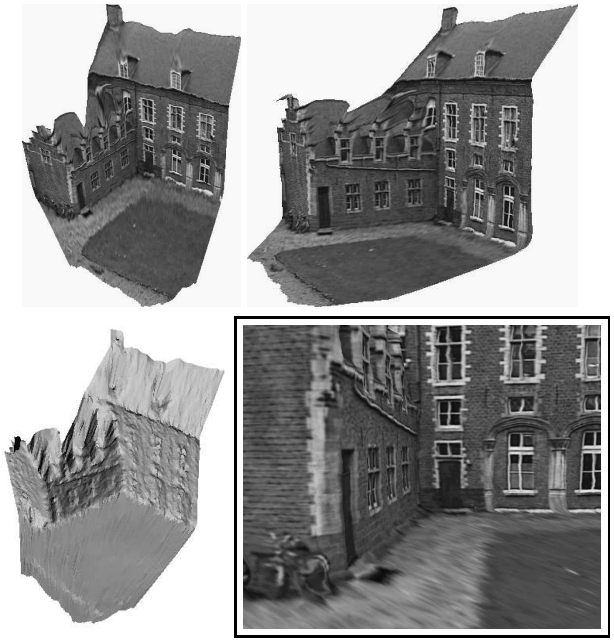

Figure 3. Top: textured views. Bottom left: shaded view. Bottom right: close-up view.

Figure 4. Images 1 and 8 of pillar sequence.

away from the object the zoom was changed ($2\times$) to keep the image size of the object constant. In spite of the changes in focal length the metric frame could be retrieved through self-calibration. In Figure 5 some perspective views of the reconstruction are given, rendered both shaded and with surface texture mapping. The shaded view shows that even most of the small details of the object are modeled.

To assess the metric properties for the pillar, 27 different lengths were measured on the real object and compared with the metric model to obtain the scale factor. Averaging all measured distances gave a consistent scale factor of 40.25 with a standard deviation of 5.4% overall. For the interior distances (avoiding the inaccuracies at the boundary of the model), the reconstruction error dropped to 2.3%. These results demonstrate the metric quality of the reconstruction even for complicated surface shapes and varying focal length.

Figure 5. Perspective views of the reconstruction (with texture and shading).

Figure 6. Left: Image 1 and 5 of the fountain sequence.

4.3. 3-D Modeling at Sagalassos: A Test Case

The system was tested on a variety of scenes with different cameras of varying quality (35 mm photo camera on Photo-CD, digital still camera, cam-corders) and was found to work even in difficult acquisition circumstances. As special test case field trials were carried out at the archaeological excavation site of Sagalassos in Turkey. This is a challenging task since the archaeologists want to reconstruct even small surface details and irregular structures. Measurements with highly calibrated photogrammetric worksta-

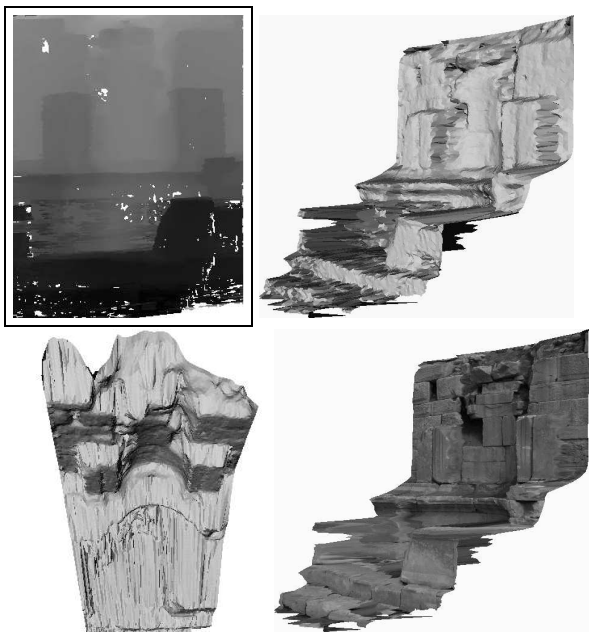

Figure 7. Depth map, intensity coded (dark = near, light = far). Textured and shaded side views of model, shaded top view.

Figure 8. Images 1, 3 and 6 of Roman Bath sequence. Lower right: estimated depth map (dark = near, light = far).

tions failed since those systems could not withstand the high temperatures at the site. We show here three of the different objects selected for modeling.

Fountain. The *Fountain* sequence consists of 5 images of an old fountain at Sagalassos, taken with a digital camera with 573x764 pixel resolution. Figure 6 shows two images of the sequence and fig. 7 the fused dense depth map and views of the 3-D reconstruction. The geometric and visual reconstruction quality for the fountain is very high. The depth map has a fill rate (percentage of valid depth estimates over all image pixels) of 96% and an average relative depth error of 0.1% [14]. Even fine details like the relief carved into the stones are preserved, and the side and top views of the overall model show a detailed 3D structure.

Roman Bath. The next example shows the reconstruction of parts of the *Roman bath* from 8 uncalibrated images taken with a standard photo camera. Figure 8 shows 3 of the images and the fused depth map. The relative depth error was estimated to 0.8% and the depth map is very dense. Figure 9 reveals the high reconstruction quality which gives a realistic impression of the object. The close-up view confirms that each stone is modeled, including relief and small indentations. The indentations belong to erosion gaps between the stones.

Sagalassos Site. The *Site* sequence in figure 10 is a good example of a large scale modeling using off-the-shelf equipment. 9 images of the complete excavation site of Sagalassos in Turkey (extension a few km^2) were taken with a conventional photographic camera while walking along the valley rim. The film was then digitized on PhotoCD.

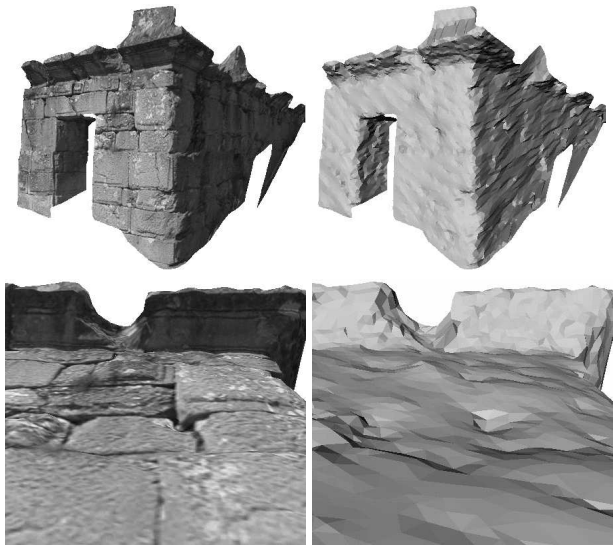

Figure 9. Textured and shaded views of Roman bath model. The close-up view shows that even small details like single stones are modeled.

The *Site* model in figure 11 gives a good reconstruction of the valley relief. Some of the dominant objects like the Roman Bath and the Agora, as well as landmarks like big trees or stones are already modeled at this coarse scale. It is intended to register the detailed object models like the Fountain or the Roman Bath to the Site and to build a virtual excavation site that one can visit as part of a virtual archaeological show case.

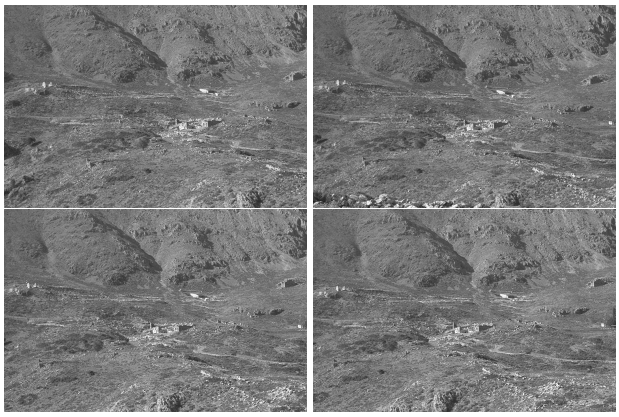

Figure 10. 4 of 9 Images of the site sequence.

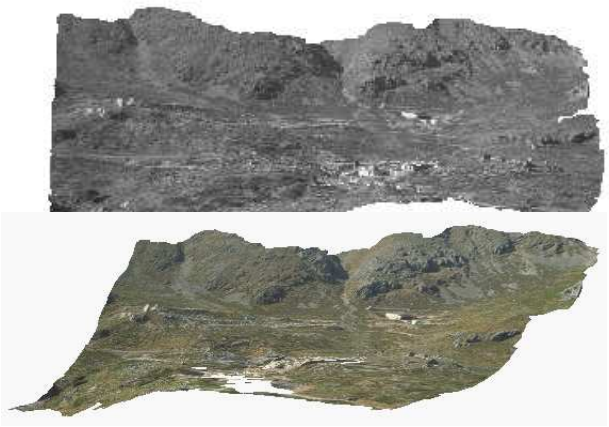

Figure 11. Textured views of the site reconstruction.

5. Conclusion

An automatic 3D scene reconstruction system was described that is capable of building metric textured 3D models from uncalibrated image sequences. The technique is capable of extracting metric surface models without prior knowledge about the scene or the camera other than assuming rigid objects. The approach was tested with different off-the-shelf camera types and for scenes of varying scale and complexity. Typically a depth accuracy of below 1% of scene depth could be computed for dense depth maps with a fill rate of over 70%. The high quality of the reconstructed objects, the different scene types, and the use of off-the-shelf equipment prove the versatility and flexibility of the proposed scene reconstruction approach.

Acknowledgments

We would like to thank Andrew Zisserman and his group from Oxford for supplying us with robust projective reconstruction software. I. Cox and L. Falkenhagen supplied versions of their disparity matching code. A specialization grant from the Flemish Institute for Scientific Research in Industry (IWT) and the financial support from the EU ACTS project AC074 VANGUARD are also gratefully acknowledged.

References

[1] P. Beardsley, P. Torr and A. Zisserman: 3D Model Acquisition from Extended Image Sequences. In: B. Buxton, R. Cipolla(Eds.) Computer Vision - ECCV 96, Cambridge, UK., vol.2, pp.683-695. Lecture Notes in Computer Science, Vol. 1064. Springer Verlag, 1996.

[2] Y. Chen and G. Medioni: Object Modeling by Registration of Multiple Range Images. Proc. Int. Conf. on Robotics and Automation, Sacramento CA, pp. 2724-2729, 1991.

[3] I. J. Cox, S. L. Hingorani, and S. B. Rao: A Maximum Likelihood Stereo Algorithm. Computer Vision and Image Understanding, Vol. 63, No. 3, May 1996.

[4] L.Falkenhagen: Hierarchical Block-Based Disparity Estimation Considering Neighborhood Constraints. Intern. Workshop on SNHC and 3D Imaging, Rhodes, Greece, Sept. 1997.

[5] O.Faugeras: Three-Dimensional Computer Vision - a geometric viewpoint. MIT-Press, 1993.

[6] O. Faugeras: What can be seen in three dimensions with an uncalibrated stereo rig. *Proc. ECCV'92*

[7] O. Faugeras, Q.-T. Luong and S. Maybank: Camera self-calibration - Theory and experiments. *Proc. ECCV'92*, pp.321-334.

[8] R. Hartley: Estimation of relative camera positions for uncalibrated cameras. *Proc. ECCV'92*, pp.579-587.

[9] A. Heyden and K. Åström: Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point. *Proc. CVPR'97.*

[10] M. Irani and S. Peleg: Super Resolution from Image Sequences. Tenth International Conference on Pattern Recognition (Atlantic City, NJ, June 16-21, 1990), IEEE Catalog No. 90CH2898-5, 1990, subconference C, 115-120.

[11] R. Koch: 3-D Surface Reconstruction from Stereoscopic Image Sequences, *Proc. ICCV'95*, Cambridge, USA, June 1995.

[12] R. Koch: Surface Segmentation and Modeling of 3-D Polygonal Objects from Stereoscopic Image Pairs. *Proc. ICPR'96*, Vienna 1996.

[13] R. Koch: Automatische Oberflächenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Bildfolgen. Ph.D. Thesis, VDI Verlag, Fortschritt-ber. 10/499, Düsseldorf, 1997.

[14] R. Koch, M. Pollefeys, and L. Van Gool: Multi Viewpoint Stereo from Uncalibrated Video Sequences. *Proc. ECCV'98*, Freiburg, June 1998.

[15] E. Ofek, E. Shilat, A. Rappoport, M. Werman: Highlight and Reflection-Independent Multiresolution Textures from Image Sequences. IEEE Computer Graphics and Applications vol. 17 (2), March-April 1997.

[16] M. Pollefeys, L. Van Gool and M. Proesmans: Euclidean 3D Reconstruction from Image Sequences with Variable Focal Lengths. *Proc. ECCV'96*, vol.1, pp. 31-42.

[17] M. Pollefeys and L. Van Gool: Self-calibration from the absolute conic on the plane at infinity. *Proc. CAIP'97.*

[18] M. Pollefeys, R. Koch and L. Van Gool: Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters. Proc. ICCV'98, Bombay, India, Jan. 1998. Also accepted for publication in: International Journal on Computer Vision, Marr Price Special Issue, 1998.

[19] P.H.S. Torr: Motion Segmentation and Outlier Detection. PhD thesis, University of Oxford, UK, 1995.

[20] B. Triggs: The Absolute Quadric. *Proc. CVPR'97.*