Measuring in Automatically Reconstructed 3D Models

Robert WULFF, Anne SEDLAZECK, and Reinhard KOCH
Multimedia Information Processing Group
Department of Computer Science
Christian Albrechts University of Kiel, Germany
{rwulff,sedlazeck,rk}@mip.informatik.uni-kiel.de

Abstract. In this paper a system for the 3D documentation of a scene within an absolute coordinate system is presented. It follows the idea of structure-from-motion and was designed to fit well into the standard archaeological documentation process. It only requires equipment that is already present at an excavation site, namely a digital camera and a total station. Besides that, data that is surveyed during the archaeological documentation process is reused to transform the model into the geo-referenced coordinate system used at the site. The paper discusses possible sources of error that influence the accuracy of reconstructed models in general. The achieved measuring accuracies are evaluated on synthetic as well as real data.

1. Introduction

In the fields of archaeology, geology, or geography, often the need to document the configuration of objects arises. This can be due to the need of having to destroy the configuration, for example during an excavation, or objects being subject to constant change through their environment or difficult to access.

One way to provide representations for documentation purposes is the generation of digital 3D models. These models provide an intuitive means for scientists not present at the site to examine the model. By using an interactive viewer the user is not restricted to the photographer's point of view, but can freely navigate in the scene. In addition to that, the models may provide a comprehensive overview of the complete scene. Finally, models that have absolute scale allow distance measuring even after the scene has been destroyed or changed. Other fields of application may be presentations, for example in museums.

Many methods for computing 3D models have been suggested. Prominent among them are for example laser scanners [9, 7]. However, in some scenarios laser scanners cannot be used or are too expensive. Furthermore, the data acquisition can be quite tiresome and time consuming. We therefore propose the computation of a 3D model based on images captured with a standard digital camera. Methods for scene reconstruction in archaeology are for example 3D Murale [1], 3D-ARCH [14] and ARC3D [16]. While the focus of 3D Murale and ARCH3D is broader than ours, ARC3D aims at offering a web service where users can upload their photos. The models computed by ARC3D offer no measuring capabilities. A structure-from-motion

approach for general rigid scenes was proposed by Pollefeys et al. [11]. Our work is based on that approach.

In this paper we present our reconstruction method, analyse possible sources of error, and discuss the accuracies we achieved with our method. It is structured as follows. Section 2 gives an overview of our reconstruction algorithm and discusses sources of error in measurements. Section 3 presents evaluations on synthetic and real data. The paper is concluded in Section 4.

2. 3D Modelling from Images

In our approach we compute 3D models from image data. To allow measuring in the model, we transform it into a reference coordinate system. The ability to measure in the model is especially useful in scenarios where the object of interest is destroyed or subject to change through its environment. In some case, the objects may be difficult to access, for example in underwater archaeology. In all these cases the analysis can be performed in retrospect using just a computer.

2.1. Our Approach

In order to compute 3D models from image sequences, we propose using a structure-from-motion approach [11] and extend it to meet the special requirements in archaeology [17]. Our algorithm performs the following steps:

1. *Detect keypoints in each image.* In contrast to [11] we use the SIFT keypoints [8], as they provide invariance against changes in lighting, rotation, and scale. Furthermore, they provide a high level of discrimination.
2. *Establish keypoint correspondences between successive images.* As we showed in [17] the reprojection errors can be decreased if the matching is performed with two predecessors, instead of only one. Because of the high dimensionality of the keypoint descriptor, comparing two keypoints regarding their similarity is computationally expensive. To circumvent the quadratic complexity for comparing every keypoint in one image with every keypoint in another image, we use a bin-based approach: we divide the image space into bins and add each keypoint to the corresponding bin. The matching is then performed only with the other image's keypoints in the bin at the same position and its neighbours.
3. *Estimate the camera poses.* In contrast to [11], we use the epipolar geometry only for the first two cameras of the sequence. Since we assume the camera calibration to be known, we perform the initialisation of the first two cameras using the essential matrix [10]. This allows the reconstruction to be performed metrically, instead of projectively. The poses of the remaining cameras are then estimated using the POSIT algorithm [3], so that all poses are reconstructed within the same coordinate frame. If the camera was moved in an orbit around the scene, we use the LoopClosing algorithm from [17] to distribute the drift error amongst all cameras. The pose estimation is concluded by a global bundle adjustment [15]. The reconstructed camera path is shown in figure 1.
4. *Transform the scene into an absolute coordinate system.* This step has to be handled explicitly, because the epipolar geometry allows the 3D reconstruction only in an arbitrary coordinate frame. To compute the transformation, at least three 3D world points and its 2D projections have to be known. In archaeology, such

points are already determined during the documentation procedure by photogrammetric techniques. Hence, we call them *photogrammetric points*. For each photogrammetric point a corresponding 3D point in the model's coordinate system can be computed by triangulating [5] from its 2D projections. We use the method described in [6] to compute the transformation. Transforming the scene into the absolute coordinate system not only allows to measure in the virtual model, but also to geo-reference it.

5. *Generate a dense 3D model for each view.* To compute a model for each viewpoint, we follow the original approach of [11], which consists of two steps. First, depth maps are computed for each view which hold the distance to the camera of each pixel. Second, a triangle mesh is generated from each depth map. These models then have to be merged into the final model. In this early stage of our project, we do this by simply combining the data into the same coordinate system. See figure 1 for the final model.

The whole computation is performed automatically and the resulting models are stored as triangle meshes, coloured with texture maps.

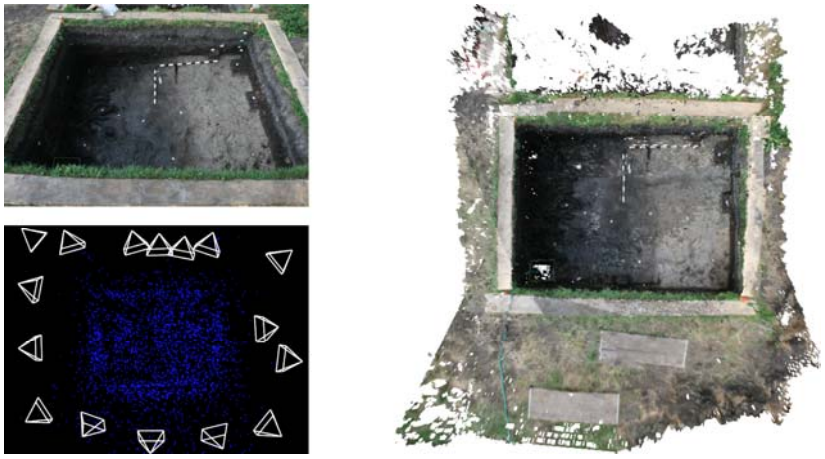

Figure 1. Bruszczewo scene. Top left: One out of sixteen input images. Lower left: estimated camera path. Right: Final model.

2.2. Sources of Error

Since we compute the models from image data, the achieved accuracy mainly depends on the quality of the images and the camera calibration. This section discusses the possible sources of error.

Obviously, the maximum accuracy is limited by the resolution of the images. However, the increased accuracy by using higher resolutions implies higher program execution times and memory consumption. So in practice, there is always a trade-off to choose from carefully. One way for lowering computational demands is to use images with reduced resolution for the pose estimation and to perform the dense depth estimation on higher resolutions.

Images with low contrast complicate keypoint extraction and dense depth estimation and, hence, influence the accuracy as well. In the Bruszczewo scene (see section 3.2.1), in the dark areas of the images the density of the keypoints was much lower than in the rest of the images. This means that there is less data available for statistically robust pose estimations. Furthermore, the dense depth estimation failed in some parts of the dark areas, so that gaps and errors in the models occur. The gaps can be compensated by fusing each viewpoint's model into one final model. In the future, we hope to better account for saturated regions by using high-dynamic range images.

The camera calibration has two impacts on the accuracy: first, the pose computation is influenced by the quality of the estimated camera parameters. Small errors in the parameters can be compensated with a bundle adjustment. Second, lens distortion effects in the texture maps for the final model need to be removed. If the distortion coefficients are not determined properly, the textures will be distorted with respect to the geometry and seams will be visible.

As we described in [17], a drift in the computation of the camera path is accumulated over the sequence. We compensate for this by applying the LoopClosing procedure from [17] if the camera was moved in an orbit around the scene.

Yet another source of error is quantisation noise in the depth maps. Because of the discretisation, steplike artifacts will occur in the model. These artifacts are called *iso-disparity surfaces* and they depend on the vergence of the cameras. For a detailed discussion on this topic see [12].

The survey of the photogrammetric points with a total station and the manual identification of their projections also bear a potential for error. Technically, very high accuracy can be achieved if special care is taken. For noisy data, a RanSaC [4] approach may be considered to eliminate outliers. However, in our experiments we didn't use this facility because of the quality of our measurements.

All afore mentioned sources of error influence the reconstruction of the final model. But there is a problem in determining the geometric accuracy for reconstructions of real scenes *per se*. This is due to the fact that one cannot locate points on the reconstructed model for which the real-world coordinates are known without considering its texture (e.g. by using special markers that are located in the reconstructed model). This implies that the (mis-)alignment of textures on the model will bias the accuracy analysis of reconstructed 3D models in general. Besides that, the perceived detail level strongly depends on the resolution of the model's texture. So even if images with reduced resolution were used for the computation we take the original images as textures. This enables the user to identify points of interest in the model more reliably.

To summarise, some sources of error can be compensated while others are inevitable. The achievable accuracy is bound by technical limitations, such as the image resolution. And in general it is difficult to measure the absolute accuracy of the estimated geometry. In the evaluation we will analyse the effects of the errors described above.

3. Evaluation

Our algorithm was tested on synthetic as well as real data. First, we analyse results from a synthetic scene and then discuss the evaluation with data from real scenes.

3.1. Synthetic Scene

The synthetic scene resembles the archaeological trench from the Bruszczewo scene which is considered in the next section. To provide realistic structural information, we used one of the input images from the Bruszczewo scene as texture. The virtual trench is about 3 x 4 units which is the size of the trench from the Bruszczewo scene. Hence, we can interpret the units as meters. Since the camera calibration is known, this experiment allows the evaluation of our method without being affected by noisy camera parameters.

In this experiment we focused on comparing the estimated geometry with the ground truth data available. For measuring the accuracy of the reconstruction we chose to compute the pixelwise differences of a set of corresponding depth maps. One depth map of each correspondence pair belongs to the input image and was generated from the ground truth model, the other one was estimated during the reconstruction process. Ideally, the pixelwise differences should be zero, because both models share the same coordinate frame. Figure 2 shows an exemplary depth map pair and its pixelwise difference. The mean depth error was 0.0263 [m] with a standard deviation of 0.0297 [m] for eight depth map pairs. Given the maximum extent of 4 m, the error is about 0.6575% relative to the trench's size. As figure 2 shows, most of the errors occur at discontinuities. We hope to further minimize these by an improved fusion of the depth maps during the reconstruction process.

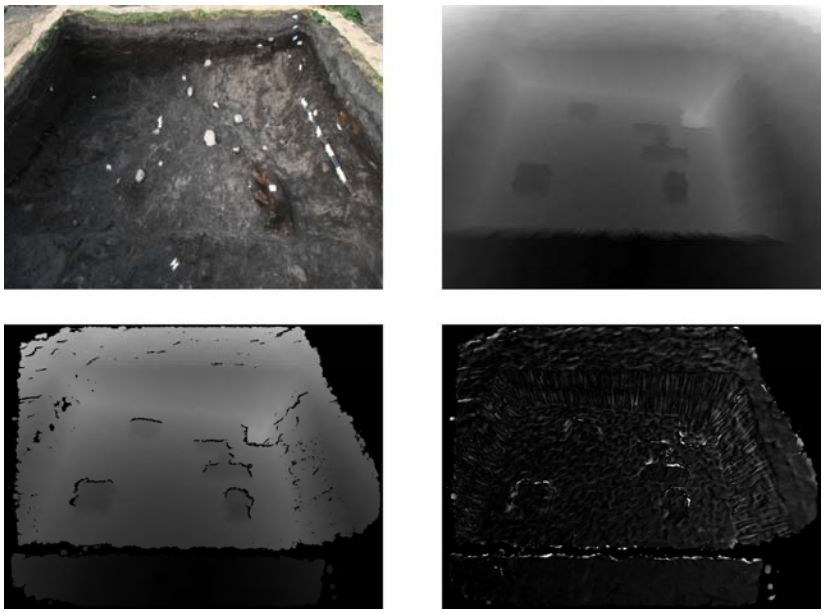

Figure 2. Synthetic scene. Top left: One input image. Note that the texture doesn't fit the geometry perfectly. Is is just used to provide structure in the images. Top right: The corresponding ground truth depth map. Pixels with higher intensity are further away from the camera than pixels with lower intensity. Lower left: Estimated depth map. Bottom right: Pixelwise difference of both depth maps. The intensity values were scaled for visualisation purposes. Higher intensity means higher difference.

3.2. Real Scenes

Figures 1 and 3 show reconstructions of archaeological trenches from Bruszczewo, Poland, and Priene, Turkey, respectively. In real scenes, one cannot compare depth maps because of the missing ground truth information. However, since both reconstructions were transformed into the coordinate system used at the excavation site, we can compare the Euclidean distances between known points. These encompass the distances between the photogrammetric points and the sizes of measuring sticks. The distances were identified manually in each model, and the absolute measurement error in mm was determined by comparing it with the known true values. For both reconstructions the errors were in about the same order of magnitude, except for a larger set of outliers in the Bruszczewo scene.

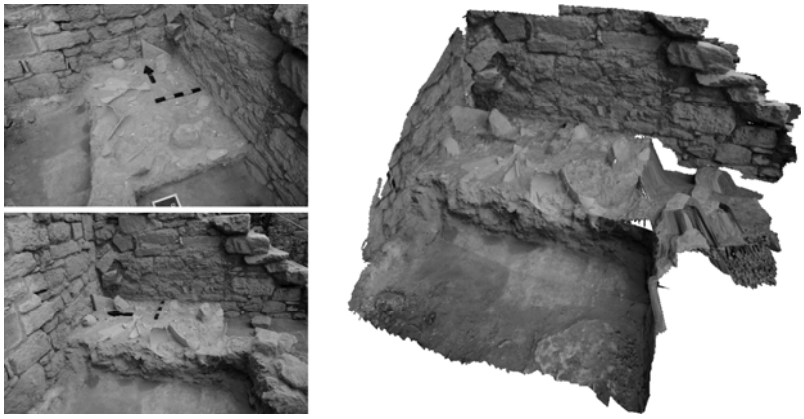

Figure 3. Priene scene. Left: Two out of 21 input images. Right: Final Model. The input images were provided by the Department of Classical Archaeology at Kiel University.

3.2.1. Bruszczewo Scene

The trench in the Bruszczewo scene has an extension of approximately 3 x 4 m. The measured distances vary between 0.1 m and approximately 1 m. Table 1 shows the resulting average measurement errors, which are very small (below 10 mm). Note that the experimental set-up is different from the one described in section 3.1. To enable comparison with ground truth data, the results for the synthetic scene are given in table 1 as well.

There are about 10 outliers out of the 312 measurements. These are caused by the artifacts of the iso-disparity surfaces and errors in the dense depth estimation due to difficult lighting conditions in the dark areas of the trench. Since our simple depth fusion approach performs no consistency checks yet, these become clearly visible in the model.

The resolution used for the reconstruction is 972 x 648 pixels. Given the camera's sensor size of 22.2 x 14.8 mm and the minimum and maximum distance of 2000–5000 mm between objects and camera, the lower bound for the geometric accuracy can be calculated as 3.1–7.8 mm using the intercept theorems.

3.2.2. Priene Scene

In the Priene scene, the trench is about 1 x 1.5 m and the measurements range from 0.1 m up to about 0.75 m. The measurement errors shown in table 1 are far lower than in the Bruszczewo scene because we used higher resolutions for the dense depth estimation (see below). This and the smaller distance of the camera with respect to the trench lowered the quantisation noise of the iso-disparity surfaces. Furthermore, the homogeneous lighting conditions led to a more accurate computation of the depth maps.

In this scenario, we used different resolutions for pose estimation and the dense depth estimation: while the images for the pose estimation had a resolution of 1072 x 712 pixels, the dense depth estimation was performed on images with a resolution of 2144 x 1424 pixels. The minimum and maximum distance between objects and the camera were 1000 and 3000 mm, respectively. Hence, the lower bound for the geometric accuracy varies between 0.7 mm and 2.0 mm for a sensor of size 23.6 x 15.8 mm.

Table 1: Average distance measuring errors.

scene	number of measurements	median error [mm]	mean error [mm]	std. dev. of mean	worst error [mm]
Bruszczewo	312	5.878	8.433	11.223	90.833
Priene	32	5.691	6.396	5.598	19.457
Synthetic	59	5.309	6.2155	4.5909	13.987

4. Conclusion

This paper presented a method for 3D scene reconstruction. As it mainly focuses on 3D documentation in archaeology, it only needs equipment that is already available at an excavation site, namely, a standard digital camera and a total station. The resulting models have absolute position, orientation, and scale, so that measuring and geo-referencing become possible. The achieved measuring accuracies were analysed and discussed.

In the future, we plan to extend the approach to volumetric measurement capabilities. For fusing several depth maps or triangle meshes the approaches [2, 13, and 18] seem promising to us and will be investigated further. Besides that, we hope to improve accuracy in high-contrast lighting situations by using exposure bracketing to produce high-dynamic range images. The automatic detection of the projections of the photogrammetric points is desirable as well.

Acknowledgements

The authors would like to thank Prof. Rumscheid and his staff of the Department of Classical Archaeology at Kiel University for providing images of the excavation in Priene, Turkey.

References

[1] J. Cosmas, T. Itagaki, D. Green, E. Grabczewski, F. Weimer, L. J. Van Gool, A. Zalesny, D. Vanrintel, F. Leberl, M Grabner, K. Schindler, K. F. Karner, M. Gervautz, S. Hynst, M. Waelkens, M. Pollefeys, R. DeGeest, R. Sablatnig, and M. Kampel. *3d murale: A multimedia system for archaeology.* In Virtual Reality, Archeology and Cultural Heritage, pages 297–306, 2001.

[2] B. Curless, and M. Levoy. *A Volumetric Method for Building Complex Models from Range Images.* Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH), 1996.

[3] D. F. DeMenthon and L. S. Davis. *Model-based object pose in 25 lines of code*. In ECCV, pages 335–343, 1992.

[4] M. A. Fischler and R. C. Bolles. *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*. Commun. of the ACM, 24(6):381–395, June 1981.

[5] R. Hartley and P. Sturm. *Triangulation*. CVIU, 68(2):146–157, November 1997.

[6] B. K. Horn. *Closed-form solution of absolute orientation using unit quaternions*. JOSA, 4(4):629 et seq., April 1987.

[7] M. Ioannides and A. Wehr. *3d-reconstruction and re-production in archaeology.* Proc. of the Int. Workshop on Scanning for Cultural Herit. Rec., 2002.

[8] D. G. Lowe. *Distinctive image features from scale-invariant keypoints*. IJCV, 60(2):91–110, 2004.

[9] A. Marbs. *Experiences with laser scanning at i3mainz*. International Workshop on Scanning for Cultural Heritage Recording. 2002.

[10] D. Nistér. *An efficient solution to the five-point relative pose problem*. PAMI, 26(6):756–777, 2004.

[11] M. Pollefeys, L. van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. *Visual modeling with a hand-held camera*. International Journal of Computer Vision, 59(3):207–232, 2004.

[12] M. Pollefeys, S. Sinha. *Iso-disparity surfaces for general stereo configurations*, T. Pajdla and J. Matas (Eds.), Computer Vision - ECCV 2004 (European Conference on Computer Vision), LNCS, Vol. 3023, pp. 509-520, Springer-Verlag, 2004.

[13] K. Pulli, T. Duchamp, J. McDonald, L. Shapiro, and W. Stuetzle. *Robust Meshes from Multiple Range Maps.* Proceedings of the IEEE International Conference on Recent Advances in 3D Digital Imaging and Modeling, 1997.

[14] F. Remondino, S. El Hakim, S. Girardi, A. Rizzi, S. Benedetti, and L. Gonzo. *3d virtual reconstruction and visualization of complex architectures: The 3d-arch project.* In 3DARCH09, 2009.

[15] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. *Bundle adjustment—a modern synthesis*. LNCS, 2000.

[16] M. Vergauwen and L. J. Van Gool. *Web-based 3d reconstruction service.* MVA, 17(6):411–426, December 2006.

[17] R. Wulff, A. Sedlazeck, and R. Koch. *3d reconstruction of archaeological trenches from photographs.* In Proc. SCCH '09, 2009.

[18] C. Zach. *Fast and high quality fusion of depth maps*. International Symposium on 3D Data Processing, Visualization and Transmission, 2008.