# Image-Based 3D Documentation in Archaeology

Robert Wulff

Multimedia Information Processing Group
Department of Computer Science
Christian Albrechts University of Kiel, Germany
rwulff@mip.informatik.uni-kiel.de

**Abstract.** In this work, an image-based scene reconstruction algorithm for the 3D documentation of archaeological trenches is proposed. It follows the structure-from-motion approach and uses only equipment that is already part of the standard documentation procedure in archaeology, namely a digital camera and a total station. Points measured with the total station are used to transform the model into the reference coordinate system used at the excavation site, so that measuring and geo-referencing becomes possible.

## 1 Introduction

During archaeological excavations much time and effort is spent on documenting the site. While a lot of progress has been made in the fields of computer vision and 3D scene reconstruction, the documentation methods in archaeology mainly yield 2D representations of trenches. Even the photogrammetric methods—which survey 3D points—are mostly used to produce CAD drawings and to rectify photographs for gaining pseudo-orthoprojections. However, virtual 3D models bear a high potential to aid in the archaeological interpretation:

- Since the configuration of finds and features is usually destroyed when the next layer is unveiled, a 3D model provides a lasting representation of a trench's 3D geometry at a given point in time.
- Archaeologists can navigate the scene interactively and are no longer bound to the photographer's point of view. Furthermore, the ability to produce true orthographic projections from *any* direction may supersede the pseudo-orthoprojections currently used.
- The 3D models can be transformed into the coordinate system used at the excavation site and therefore provide a means for measuring and geo-referencing. This also enables archaeologists to correlate the model with other spatial data, e. g. with the models of different layers of the same trench so that the whole excavation can be reproduced virtually.
- Another benefit applies to popular science: Since laymen are not used to abstract representations, as the above mentioned CAD drawings, 3D models can help in presenting scientific results to the public.

For this reasons, the computation of 3D models in the archaeological documentation procedure is proposed. That this is indeed promising was shown in [1].

Section 2 relates this work in the context of other projects and states its goals. The single steps of the reconstruction process are described in Sect. 3. Section 4 presents results with synthetic as well as real data. The paper is concluded in Sect. 5.

## 2 Related Work and Goal of this Project

The idea of computing 3D models of archaeological sites is not new. Basically, there are two approaches to the problem: Laser scanning and image-based reconstruction. A variety of projects use laser scanners for documentation purposes (see for example [2], [3], and [4]). However, there are some drawbacks: Laser scanners are not part of the standard documentation procedure in archaeology and serve only a very special purpose. In some situations, they are not applicable, for example in underwater archaeology. Besides that, technically trained personnel is needed to operate the device and data acquisition can be tiresome and time consuming. But most importantly, they are still too expensive to be used widely.

These drawbacks are circumvented by the image-based approach for the cost of a lower accuracy. Two approaches for general scenes can be found in [5] and [6]. The projects 3D MURALE [7], 3D-Arch [8], and ARC3D [9] perform 3D scene reconstructions of archaeologically relevant sites. The scope of 3D MURALE is much broader than the one in this work: Besides 3D scene reconstruction of trenches, it focuses on artifact and terrain reconstruction and data integration using databases. 3D-Arch has a broader focus as well with the aim of reconstructing architectural sites like castles. The project ARC3D focuses on providing a web service where users can upload their images.

In contrast to the above mentioned projects, the goal here is to develop an image-based scene reconstruction procedure for the 3D documentation of archaeological trenches. The most important requirement for this approach is to use only technical equipment that is already part of the standard documentation procedure, namely a digital camera and a total station. Furthermore, existing data is reused: since the models should enable measuring and geo-referencing, they need to be explicitely transformed into the reference coordinate system from the excavation site (see Sect. 3.5). The photogrammetric points used to compute the transformation are surveyed in the standard documentation procedure to rectify images for the afore mentioned pseudo-orthoprojections. Reusing existing equipment and data ensures that the reconstruction algorithm fits well into the standard documentation procedure. Since archaeologists are familier with digital cameras and total stations, no additional technical training is necessary. Apart from that, some guidelines have to be followed for the data acquistion which are described in Sect. 3.1.

Since the algorithm focuses on a special field of application, knowledge about the expected camera movement can be exploited. The *LoopClosing* algorithm,

described in Sect. 3.4, assumes that the camera was moved in an orbit around the scene and distributes the accumulated drift error amongst all cameras to reduce the reprojection error.

## 3 Reconstructing an Archaeological Site

In this section, the reconstruction algorithm is described. It follows the structure-from-motion approach (see for example [6]). The reconstruction pipeline is visualised in Fig. 1.

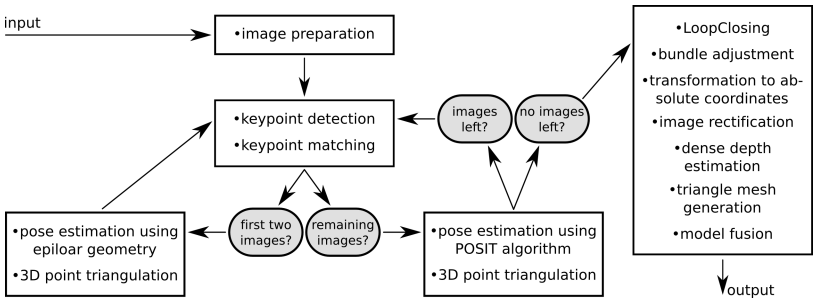


**Fig. 1.** Flowchart of the reconstruction process.

### 3.1 Data Acquisition

The algorithm requires an ordered sequence of images taken with a standard digital camera as input. While in principal reconstructions can be performed with unordered sequences as well (by using nearest-neighbour approximations for keypoint matching [10]), this knowledge is used here to limit the search area for keypoint matching as described in Sect. 3.2. Since the images are usually taken with a single camera in an ordered fashion, this means no restriction for the documentation procedure. The following principles should be followed when taking the images:

- The intrinsic camera parameters must be constant for all images.
- Consecutive images need to overlap by a large proportion (about 80%) to allow stable keypoint matching.
- The lighting conditions should be relatively constant.
- The scene should be opaque and static. Reflections are to be avoided. Especially the latter can be problematic in water-saturated sites. Here, pumps must be used to lower the water level, while small puddles can be dried temporarily with sponges.
- If the camera was moved in an orbit around the scene, this knowledge can be exploited to further reduce the reprojection error (see Sect. 3.4). However, this requirement is just optional.

Since the reconstruction is performed metrically (and not projectively), the intrinsic camera parameters have to be known and are determined using [11]. The calibration can be performed before or even after the excavation, as long as the camera parameters remain the same.

To allow measuring and geo-referencing, the model must be explicitely transformed into the reference coordinate system used at the excavation site. For this purpose, the coordinates of at least three 3D points in the scene must be known. Such points are already surveyed in the standard documentation procedure using a total station. They are called *photogrammetric points* and are reused here. The markers of the points are left in the scene while the images are taken, so that their 2D projections can be determined manually in the images.

### 3.2 Keypoint Detection and Matching

The first step of the algorithm is to detect keypoints in the images using the well-known SIFT algorithm [12]. In the next step, keypoint correspondences are established between successive images. As was shown in [13] 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 keypoints' descriptors, computing the similarity of two keypoints is not insignificant. To reduce the quadratic complexity of the naive approach—comparing each keypoint of one image with all keypoints in another image—a bin-based approach is employed. The image space is divided into bins and each keypoint is afterwards added into the corresponding bin. The matching is then performed only with the other image's keypoints in the bin at the same position and its surrounding neighbours.

### 3.3 Pose Estimation

Based on the keypoint correspondences, the camera poses can now be estimated. The reconstruction is initialised with the first two cameras of the scene using epipolar geometry. Since the intrinsic camera parameters are assumed to be known, the pose estimation can be performed with the essential matrix [5, 14]. To effectively deal with outliers, a RanSaC [15] approach is employed.

The poses of the remaining cameras are computed with the POSIT algorithm [16]. This algorithm requires 2D-3D correspondences, so 3D points need to be triangulated [17] from the already processed views.

### 3.4 The LoopClosing Algorithm

In the scenario of this work it is likely that the camera was moved in an orbit around the trench. This implies that the first and the last camera share a large proportion of their viewports. In this case, attaching the first image again at the end of the image sequence enables the keypoint matching and pose estimation to be performed between these cameras as well.

Let $n$ denote the number of input images, $n+1$ the index of the attached camera, $c_i$ the position of camera $i$ and $q_i$ the orientation of camera $i$ in quaternion representation, where $0 \leq i \leq (n+1)$. Since the first camera is aligned with the coordinate system, its position is $c_1 = (0,0,0)^{\mathrm{T}}$ and its orientation is $q_1 = (0,0,0,1)$. Ideally, $c_1 = c_{n+1}$ and $q_1 = q_{n+1}$ hold, but in practice errors in camera calibration, measuring, and rounding will lead to a discrepancy between these values. The *LoopClosing* algorithm distributes these discrepancies between all cameras according to a weighting function so that the poses of the first and the attached camera will match perfectly. Furthermore, the reprojection error is minimized—see Sect. 4.1 for results.

The first step is to compute the discrepancies. For the position, the difference vector is given by $\Delta_c := c_1 - c_{n+1}$. Since $c_1 = (0,0,0)^{\mathrm{T}}$ this simplifies to $\Delta_c = -c_{n+1}$.

Using the quaternion representation, the discrepancy in the orientation is given as the conjugate of the quaternion of the orientation of camera $n+1$, i. e. $\Delta_q := \overline{q_{n+1}} = (q_1, -q_2, -q_3, -q_4)$, where $(q_1, q_2, q_3, q_4) = q_{n+1}$.

In the second step, a weighting function $w : \{1, \ldots, n+1\} \rightarrow [0,1]$ is computed. Under the reasonable assumption that the discrepancies were accumulated over the sequence and grow with an increasing number of images, this function is choosed so that the following conditions are met:

- $w(1) = 0$, i. e. the first camera shall not be transformed at all
- $w(n+1) = 1$, i. e. the attached camera shall be fully transformed so that the poses of the first and the attached camera are equal
- $w(i) < w(j)$ for all $i < j$, i. e. the cameras at the front of the sequence are transformed less than the latter ones

The weights are computed recursively over the distance of each camera to the first camera along the camera path: let $L_1 := 0$ and $L_i := L_{i-1} + d(c_i, c_{i-1})$, for $1 < i \leq n+1$, where $d(.,.)$ is the Euclidean distance. The total length of the camera path is given by $L_{\mathrm{total}} := \sum_{i=1}^{n+1} d(c_i, c_{i-1}) = L_{n+1}$. Now the weighting function can be defined as $w(i) := L_i / L_{\mathrm{total}}$, for all $1 \leq i \leq n+1$, which meets the above mentioned conditions.

The third step is to transform the cameras. The location of camera $i$ is replaced by its adjusted position $c_i + w(i)\Delta_c$. For the orientation, the computation is slightly more complex. The angle of the rotation represented by the quaternion $\Delta_q = (\Delta_{q,1}, \Delta_{q,2}, \Delta_{q,3}, \Delta_{q,4})$ is given by $\alpha := 2\cos^{-1}(\Delta_{q,1})$ and the axis can be computed as $a := (1/\sin(\alpha/2))(\Delta_{q,2}, \Delta_{q,3}, \Delta_{q,4})^{\mathrm{T}}$. Using the weight $w(i)$, one computes the adjustment angle $\kappa_i := w(i)\alpha$ for camera $i$ and obtains the quaternion $r_i := (\cos(\kappa_i/2), \sin(\kappa_i/2)\, a)$. The orientation of camera $i$ is then replaced by its adjusted orientation $r_i q_i$.

Finally, the camera path is closed by correlating the correspondences of the keypoints of camera $n+1$ with the keypoints of camera 1. Afterwards, camera $n+1$ can be removed from the sequence. The loop closing procedure is concluded by a global bundle adjustment [18].

### 3.5 Absolute Coordinates

To enable measuring and geo-referencing, the scene has to be transformed into the reference coordinate system used at the excavation site. This coordinate system is clearly determined by the photogrammetric points that are surveyed during the standard documentation procedure. Given at least three 3D-3D correspondences, the algorithm from [19] computes a transformation between the two coordinate systems, precisely: translation, rotation, and scale. The corresponding 3D points in model coordinates are triangulated [17] from the 2D projections. To identify these 2D projections, the markers of the photogrammetric points need to be kept in the scene during image acquisition so that they are visible in the images.

### 3.6 3D Model Generation

After the camera poses have been estimated and transformed into the reference coordinate system, a dense 3D model of the scene can be generated. For image rectification method [20] is used and for dense depth estimation the approach from [21]. The output of the dense depth estimation is a depth map for each view. Based on these depth maps a triangle mesh can be computed for each view. To compensate for gaps and occlusions, the single meshes are displayed simultaneously to produce the final model.

## 4 Results

The algorithm was evaluated on synthetic as well as real data. This section summerises the results that were presented before in [13] and [22]. First, the results from the synthetic scene are discussed, followed by an evaluation with two real scenes.

### 4.1 Synthetic Scene

The synthetic scene was assembled similar to one of the real scenes considered in Sect. 4.2: The virtual trench is of the same size as the trench from the Bruszczewo scene, so that the virtual units can be interpreted as meters. One of the input images from the Bruszczewo scene is used as texture to provide structural information for the reconstruction algorithm. The experiments with the synthetic scene focused on the following three aspects.

**Matching Keypoints with two Preceeding Images.** Since is is assumed that the input images overlap by a large proportion, each image also shares its viewport with more than one predecessor. In this experiment the consequences of matching each image with its two preceeding images are evaluated. As the results in Fig. 2 (left) show, the drift of the cameras is clearly reduced by matching with two preceeding images. The reason for this is a more robust triangulation due to longer baselines.

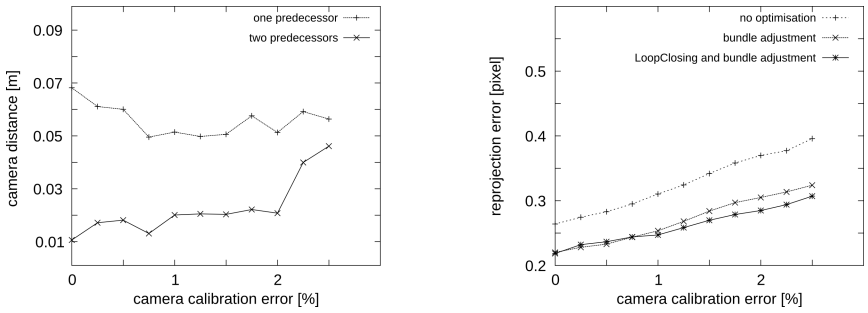


**Fig. 2.** Results for the synthetic scene. Left: Systematic noise was added to the ground-truth camera calibration ($x$-axis). The $y$-axis shows the distance between the first and the attached camera (see Sect. 3.4). The distance—which would ideally be zero—is lowered if keypoint matching is performed with two predecessors. Right: Again, the $x$-axis shows systematic noise. The reprojection error ($y$-axis) is lowered for increasingly erroneous camera calibrations.

**Evaluating the LoopClosing Algorithm.** To evaluate the LoopClosing algorithm the reprojection errors of the triangulated 3D points are considered. Figure 2 (right) shows that for an increasing error in the camera calibration the LoopClosing algorithm yields better results than a sole bundle adjustment.

**Analysing the Geometric Accuracy.** The algorithm was developed with measuring and geo-referincing in mind, so the geometric accuracy is of special interest. To analyse the accuracy the depth maps that were estimated during the reconstruction process are compared with the corresponding ground truth depth maps by computing the pixelwise difference. The mean depth error is 0.0263 [m] with a standard deviation of 0.0297 [m]. Compared to the trench's maximum extend of 4 m, the relativ error is about 0.6575%. See [22] for more details.

### 4.2 Real Scenes

The input images for the reconstruction were taken at two different excavation sites: Bruszczewo, Poland, and Priene, Turkey. The final models of the reconstruction are visualised in Fig. 3 and 4, respectivly. Since both of the models were transformed into the reference coordinate system used at the excavation site, the geometric accuracy can be analysed by comparing known distances in the models. The differences between the real distances and the ones from the models are shown in Table 1. The median error in both scene was below 6 mm, while in the Bruszczewo scene the worst error was much higher than in the Priene scene. This is caused by the dark wet soil in the Bruszczewo scene which led to nearly saturated pixel intensities. Therefore, keypoint detection and dense depth estimation were complicated. Again, see [22] for more details.

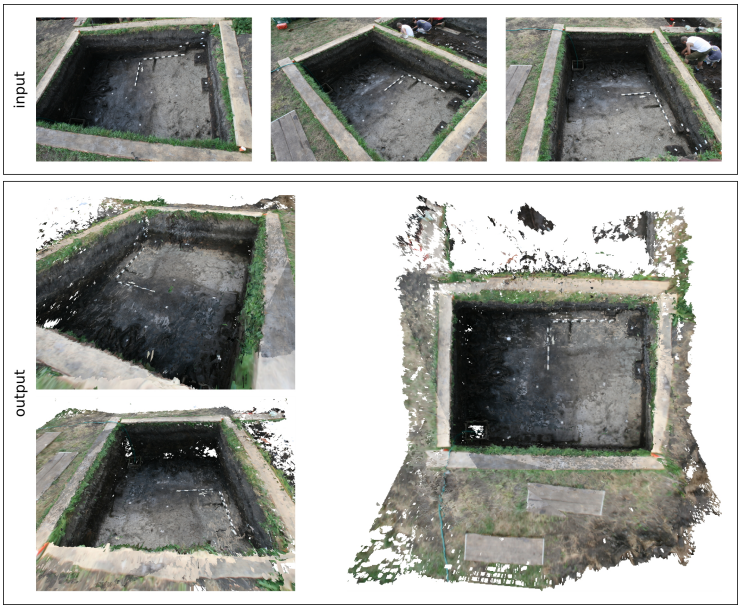


**Fig. 3.** Bruszczewo scene. Top: three out of sixteen input images. Bottom: resulting model.

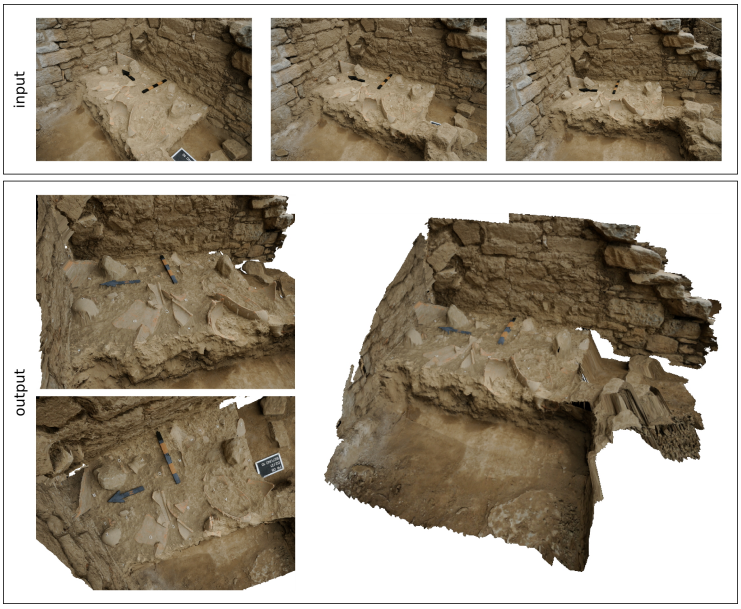


**Fig. 4.** Priene scene. Top: three out of twenty-one input images. Bottom: resulting model.

**Table 1.** Differences to real distances.

| 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 |

## 5 Conclusions and Outlook

The possibility to generate 3D models of an archaeological trench bears a high potential for archaeological investigations and interpretations. The proposed method yields 3D models that have absolute position, orientation, and scale and therefore allow measuring and geo-referencing. It is an image-based approach that only needs equipment that is already part of the standard documentation procedure in archaeology and known to archaeologists.

Currently, high-dynamic range imaging techniques are integrated into the reconstruction pipeline to handle difficult lighting conditions such as in the Bruszczewo scene. Besides that, a more elaborated method for fusing depth maps or triangle meshes is needed. Some promising methods that will be investigated further are [23], [24], and [25]. It is also planned to add semantics to the model by segmenting it into finds and features.

## Acknowledgements

The author would like to thank Dr. Jutta Kneisel of the Institute of Prehistoric and Protohistoric Archaeology at the University of Kiel for the possibility to gain insight into the archaeological work and to take images at excavation sites. In addition, Prof. Rumscheid and his staff of the Department of Classics at Kiel University provided images from the excavation in Priene, for which the author is very grateful.

## References

1. Wulff, R.: 3d-rekonstruktion archäologischer schnitte aus einzelbildsequenzen. Diploma thesis, Department of Computer Science, University of Kiel, Germany (July 2009)
2. Ioannides, M., Wehr, A.: 3d-reconstruction and re-production in archaeology. In: Proc. of the Int. Workshop on Scanning for Cult. Herit. Rec. (2002)
3. Marbs, A.: Experiences with laser scanning at i3mainz. Proc. of the Int. Workshop on Scanning for Cult. Herit. Rec. (2002)
4. Allen, P., Feiner, S., Troccoli, A., Benko, H., Ishak, E., Smith, B.: Seeing into the past: Creating a 3d modeling pipeline for archaeological visualization. In: Proc. of the 2nd Int. Symp. on 3D Data Processing, Vis., and Transm. (3DPVT'04). (2004)
5. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. 2nd edn. Cambridge University Press (2004)

6. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. Int. J. of Computer Vis. **59**(3) (2004) 207–232
7. Cosmas, J., Itagaki, T., Green, D., Grabczewski, E., Weimer, F., Van Gool, L.J., Zalesny, A., Vanrintel, D., Leberl, F., Grabner, M., Schindler, K., Karner, K.F., Gervautz, M., Hynst, S., Waelkens, M., Pollefeys, M., DeGeest, R., Sablatnig, R., Kampel, M.: 3d murale: A multimedia system for archaeology. In: Virtual Real., Archaeol. and Cult. Herit. (2001) 297–306
8. Remondino, F., El Hakim, S., Girardi, S., Rizzi, A., Benedetti, S., Gonzo, L.: 3d virtual reconstr and vis of complex architectures: The 3d-arch project. In: 3DARCH09. (2009)
9. Vergauwen, M., Van Gool, L.J.: Web-based 3d reconstruction service. Mach. Vis. Appl. **17**(6) (December 2006) 411–426
10. Beis, J.S., Lowe, D.G.: Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: Proc. of the 1997 Conf. on Computer Vis. and Pattern Recognition (CVPR '97). (1997)
11. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. on Pattern Anal. and Mach. Intell. **22**(11) (2000) 1330–1334
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. of Computer Vis. **60**(2) (2004) 91–110
13. Wulff, R., Sedlazeck, A., Koch, R.: 3d reconstruction of archaeological trenches from photographs. In: Proc. of Sci. Computing and Cult. Herit., Heidelberg, Germany (2009) (to appear).
14. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. on Pattern Anal. and Mach. Intell. **26**(6) (2004) 756–777
15. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. of the ACM **24**(6) (June 1981) 381–395
16. DeMenthon, D.F., Davis, L.S.: Model-based object pose in 25 lines of code. In: Eur. Conf. on Computer Vis. (1992) 335–343
17. Hartley, R., Sturm, P.: Triangulation. Computer Vis. and Image Underst. **68**(2) (November 1997) 146–157
18. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment – a modern synthesis. Lect. Notes in Computer Sci. (2000)
19. Horn, B.K.: Closed-form solution of absolute orientation using unit quaternions. J. of the Opt. Soc. of Am. **4**(4) (April 1987) 629 et seq.
20. Pollefeys, M., Koch, R., Van Gool, L.: A simple and efficient rectification method for general motion. IEEE Int. Conf. on Computer Vis. **1** (1999) 496–501
21. Van Meerbergen, G., Vergauwen, M., Pollefeys, M., Van Gool, L.: A hierarchical symmetric stereo algorithm using dynamic programming. Int. J. of Computer Vis. **47** (2002) 275–285
22. Wulff, R., Sedlazeck, A., Koch, R.: Measuring in automatically reconstructed 3d models. In: Proc. of Geoinformatik2010, Kiel, Germany (March 2010) 95–102
23. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. Spec. Interest Group on Graph. and Interact. Tech. (SIGGRAPH) (1996)
24. Pulli, K., Duchamp, T., McDonald, J., Shapiro, L., Stuetzle, W.: Robust meshes from multiple range maps. Proc. of the IEEE Int. Conf. on Recent Adv. in 3D Digit. Imaging and Modeling (1997)
25. Zach, C.: Fast and high quality fusion of depth maps. Int. Symp. on 3D Data Processing, Vis. and Transm. (2008)