

A NOVEL HIGH ACCURACY 3D SCANNING DEVICE FOR ROCK-ART SITES

Thomas Höll, Gert Holler, and Axel Pinz

Inst. of Electrical Measurement and Measurement Signal Processing, Graz University of Technology, Austria
 {thomas.hoell, gert.holler, axel.pinz}@tugraz.at

KEY WORDS: Stereo, Structure from Motion, drift compensation, tachymeter

ABSTRACT:

We are currently developing a novel 3D scanning device for rock-art. Within the European project 3D-Pitoti, this scanner shall be used to acquire 3D structure and radiometric surface properties of ancient rock-art sites in Valcamonica. Overall design goals include high spatial accuracy and precision, as well as radiometric quality beyond phototexture. This paper is devoted to the geometric measurement principle of the new scanner. We present a novel scanning scheme based on various constraints to Structure from Motion, that guarantees high accuracy of the resulting scans by combining tachymeter-based tracking of the scanner, stereo, and structure-from-motion. This method provides scale information (by calibrated stereo), and does not require ground control points, because outside-in tracking avoids the typical issues of drift in structure-from-motion. The system is designed for flexibility, high throughput, approx. 0.1mm precision, and an overall accuracy of the reconstructed 3D structure that conforms with the specifications of the tachymeter.

1 INTRODUCTION

This paper presents the first prototype of a 3D rock-art scanner that has been developed within the European project 3D-Pitoti¹. This project aims at scanning of large rock-art sites through all scales, applied to the ancient rock-art at the site of Valcamonica. Data is captured at three scales, i.e. airborne scanning of the valley, mid-range scanning of rock panels by unmanned aerial vehicles, and micro-range scanning up to 0.1mm spatial resolution produced by the scanner presented in this paper.

The 3D-Pitoti project aims not only at the 3D scanning that will result in point clouds and meshes, but also on novel use cases of the 3D data in various application scenarios. Most scenarios require the seamless transition between scales. Therefore, excellent registration of all the scanned 3D data is required. Our novel micro-range scanner achieves excellent registration, because it avoids the typical issues of drift of large, stitched models in Structure from Motion by outside-in tracking using a tachymeter (i.e. total station). Thus, the scanner can be considered a front-end to total station based surveillance in 3D rock-art scanning.

2 SCANNER PROTOTYPE

Figure 1 shows the main components of the scanner. Since major requirements include portability, ease of use, autonomous operation with long battery life, and affordability, the scanner can be used in a “walking stick” manner. The mini-tripod ensures a solid rest during the capture of images (< 1 sec per capturing position), and the one-button operation triggers all scanner components.

This general scanner concept can be reduced to the “essence” of the geometry measurement as shown in the T-shaped configuration in Figure 2, where we also introduce the local scanner coordinate system. Given the calibration of the T-shaped rig, we can calculate a 3D Euclidean reconstruction of the surface within the stereo model (overlapping red region) of the scanner in local scanner coordinates.

When operated in the 3D-Pitoti context introduced in section 1, each of the n scanner positions is recorded by a tachymeter, providing the 3D points $S_{j=1 \dots n}$ in the world coordinate system (W_X, W_Y, W_Z) established by the tachymeter (see Figure 3).

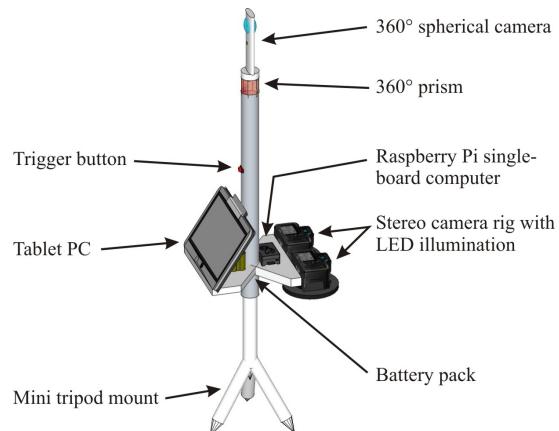


Figure 1: The “walking stick” scanner and its main components: A stereo-rig with custom LED illumination, a microcontroller based control unit, a tablet PC, a button to trigger one scan, a 360° prism for outside-in tracking by a tachymeter, and a spherical camera for inside-out verification of scanner pose.

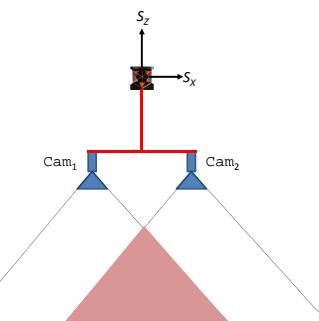


Figure 2: The “essence” of the geometry measurement: A calibrated T-shaped rig consisting of a calibrated stereo configuration and a 360° prism. The calibration includes interior camera parameters and the relative orientation between Cam_1 , Cam_2 , and the prism. The scanner coordinate system (S_X, S_Y, S_Z) is co-located with the prism center. S_X is aligned with the axis of the stereo rig, and S_Z coincides with the upright pole of the “T”.

¹<http://www.3d-pitoti.eu/>

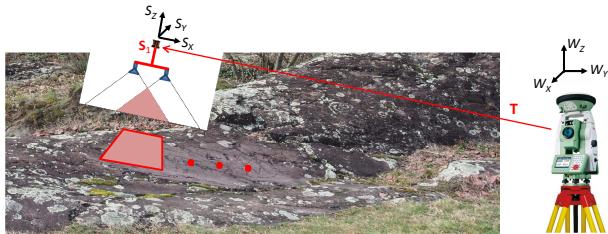


Figure 3: The scanner is positioned on a rock panel. The position of the prism is measured by a tachymeter (yielding the translation vector \mathbf{T}), and a stereo pair is captured.

3 CONSTRAINED STRUCTURE FROM MOTION

For one isolated scan, only the scanner position (three degrees of freedom – 3 DoF) can be reconstructed. This means, that three rotational DoF cannot be fixed, so that the resulting 3D surface reconstruction is tied to \mathbf{S}_1 ², but it can be oriented arbitrarily on a sphere around \mathbf{S}_1 . When several scans with partially overlapping scanner footprints are taken (see Figure 4), the situation changes. As soon as at least $n = 3$ scans are available, the position and orientation (3 positional and 3 orientational DoF) of the blue triangle in Figure 4 can be recovered. Next, the three scanner footprints can be rigidly stitched in 3D.

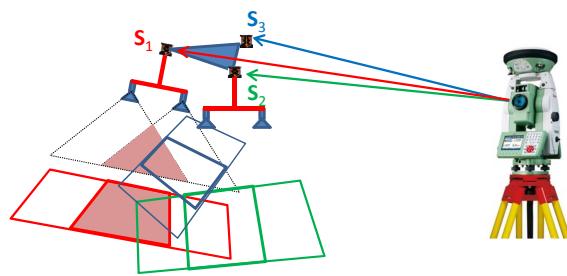


Figure 4: Scanning of larger areas requires to move the scanner around. Successive stereo pairs are captured, and prism positions $\mathbf{S}_j, j = 1 \dots n$ are recorded. This figure also shows the overall “footprint” per scan, i.e. a mapping of the two respective images to the surface.

This poses an interesting constrained Structure from Motion problem that can be tackled in various ways. The six constraints at hand are:

1. Known 3D points $\mathbf{S}_j, j = 1 \dots n$,
2. Calibrated cameras Cam_1 and Cam_2 , i.e. interior parameters \mathbf{K}_1 and \mathbf{K}_2 ,
3. Calibrated T-shaped rig, i.e. relative orientation between the prism, Cam_1 and Cam_2 ,
4. n 3D surface reconstructions w.r.t. n rig positions, i.e. shape from calibrated stereo, providing scale information, avoiding the need for ground control points,
5. $2n$ individual images with n known relative orientations,
6. For $n \geq 3$: Full 6 DoF of prism trajectory.

²Upper-case bold letters \mathbf{V} refer to 3D world coordinates, lower-case bold letters \mathbf{v} denote 2D image coordinates, and sans-serif upper-case letters \mathbf{M} denote matrices.

To integrate constraints in the bundle adjustment optimization, (Triggs et al., 2000) suggest to use a sequential quadratic programming method (SQP). This approach is for instance used by (Lhuillier, 2011) to fuse GPS-measurements and Structure from Motion reconstructions. (Kurz et al., 2011) use a stereo-pair constraint during the bundle adjustment, extending the sparse Levenberg-Marquardt algorithm presented in (Hartley and Zisserman, 2003).

Scanning a large rock-art site like Valcamonica typically requires hundreds of individual scans (i.e. acquisition of stereo pairs) per individual rock panel. During the scanning, the user needs feedback about the quality of the scans, covered area, and individual coverage (i.e. number and quality of images overlapping on the ground). Therefore, we require online and incremental Structure from Motion processing, as provided by the processing pipeline of (Hoppe et al., 2012). This method adds incoming images online, followed by bundle adjustment for a reasonable number of the most recently added images.

In general, bundle adjustment minimizes the reprojection error by altering the camera parameters and the position of a reconstructed world point. One can express this objective function as follows:

$$f(\mathbf{R}_i, \mathbf{C}_i, \mathbf{X}_l) = \sum_i^{2n} \sum_l^m D(\mathbf{x}_{il}, \mathbf{R}_i, \mathbf{C}_i, \mathbf{X}_l)^2, \quad (1)$$

where \mathbf{R}_i and \mathbf{C}_i are the orientation of camera reference frame and the center of camera i . The function $D(\cdot)$ expresses the reprojection error between the image measurement \mathbf{x}_{il} and a world point \mathbf{X}_l observed by the camera i . It is important to note that the objective function (equation 1) treats every camera individually. Incorporating additional constraints as our six constraints listed above is a subtle task that may be implemented in various ways. Here, we present a general approach and formulate our two key constraints, i.e. the stereo constraint and the known scanner positions \mathbf{S}_j as follows:

$$\underset{\mathbf{R}_i, \mathbf{C}_i, \mathbf{X}_l}{\text{minimize}} \quad f(\mathbf{R}_i, \mathbf{C}_i, \mathbf{X}_l) + \frac{1}{\mu_k} \sum_j^n \|\mathbf{G}_j\|_2^2 + \alpha \sum_j^n \|\mathbf{E}_{S_j}\|_2^2, \quad (2)$$

where the second term is based on the quadratic penalty method by (Nocedal and Wright, 2006), and $\|\mathbf{G}_j\|_2^2$ denotes the quadratic penalty term added for each stereo rig j . The third term is a regularization term that adds $\|\mathbf{E}_{S_j}\|_2^2$ for each tachymeter measurement j , and the weights α and μ control the influence of each constraint on the minimization process, as explained in more detail below.

Below, we investigate unconstrained, incremental SfM with subsequent scale enforcement (see section 3.1), followed by a detailed analysis of the stereo constraint (section 3.2), and of the constraint obtained by the tachymeter measurements (section 3.3).

3.1 Incremental Structure from Motion and subsequent scale enforcement

We treat the n stereo pairs as $2n$ independent images, and sequentially feed them into the online incremental SfM method of (Hoppe et al., 2012) to obtain a sparse point cloud as well as the poses of the cameras, resulting in a similarity reconstruction (up to an unknown scale factor). The next step is to enforce the true scale using the known, constant distance (i.e. stereo baseline) of the two cameras of the calibrated stereo-rig. In a straightforward manner, we calculate the average distance between left and right camera of all stereo pairs of our reconstruction, and rescale the

reconstruction such that this average distance equals the correct stereo baseline.

This simple method works excellent for cases where the online incremental SfM can produce a similarity reconstruction. Our experiments suggest this behaviour of the (Hoppe et al., 2012) for compact scenes with a good coverage of several images per surface point, taken from various view angles. Our scanner, however, scans with the camera's principal axes perpendicular to the surface and therefore, this method will produce good results only in cases, when compact regions (i.e. an individual pitot of typical size about $20 \times 20\text{cm}$, or a compact group of a few pitots less than 1m^2) are scanned. For scans of larger rock-panels, this method will tend to drift.

3.2 Incremental SfM with stereo constraints through a quadratic penalty method

In this approach, we want to directly obtain a Euclidean reconstruction from online incremental SfM by enforcing the stereo baseline constraint. In each incremental step of the online SfM reconstruction, the (Hoppe et al., 2012) SfM method performs a bundle adjustment optimization to increase the accuracy of each partial reconstruction. We investigate the effect of placing a stereo constraint during this bundle adjustment step, by treating the two cameras as one stereo-rig j :

$$\mathbf{G}_j = \hat{\mathbf{R}}_{j,1 \rightarrow 2} \mathbf{R}_{j,1} \mathbf{C}_{j,1} - \mathbf{R}_{j,2} \mathbf{C}_{j,2} + \mathbf{R}_s \mathbf{C}_s, \quad (3)$$

where $\mathbf{R}_{j,1}$ and $\mathbf{R}_{j,2}$ are the orientation of the primary and secondary camera of the stereo-rig j and $\hat{\mathbf{R}}_{j,1 \rightarrow 2}$ is the estimated rotation between the camera pair. The camera centers of the two stereo-rig cameras are denoted by $\mathbf{C}_{j,1}$ and $\mathbf{C}_{j,2}$. The calibration of the stereo-rig is given by \mathbf{R}_s , the orientation of the secondary camera with respect to the primary camera, and \mathbf{C}_s , the camera center of the secondary camera w.r.t. the primary camera. Using this constraint in combination with the quadratic penalty method (Nocedal and Wright, 2006), we enforce the correct stereo base and hence obtain Euclidean reconstruction. In this case, the complete minimization problem is given by:

$$\underset{\mathbf{R}_i, \mathbf{C}_i, \mathbf{X}_l}{\text{minimize}} \quad f(\mathbf{R}_i, \mathbf{C}_i, \mathbf{X}_l) + \frac{1}{\mu_k} \sum_j^n \|\mathbf{G}_j\|_2^2. \quad (4)$$

The quadratic penalty method decreases μ_k gradually to enforce that the final result fulfills the stereo-rig constraint. We implement this optimization scheme using the Ceres solver (Agarwal et al., 2014).

3.3 Bundle adjustment with incorporated tachymeter measurements

Finally, we need to develop a constraint to integrate tachymeter measurements in the reconstruction pipeline. First, we assume that we have a sparse point cloud $\{\mathbf{X}_1 \dots \mathbf{X}_m\}$ and the corresponding camera poses given by their orientations $\{\mathbf{R}_1 \dots \mathbf{R}_{2n}\}$ and positions w.r.t. a world reference frame $\{\mathbf{C}_1 \dots \mathbf{C}_{2n}\}$. Next, we obtain n measurements $\{\mathbf{S}_1 \dots \mathbf{S}_n\}$ from a tachymeter and a correspondence list so that we can assign one tachymeter measurement \mathbf{S}_j to the primary camera pose $\{\mathbf{R}_{S_j}, \mathbf{C}_{S_j}\}$ of the stereo-rig (i.e. \mathbf{C}_{S_j} denoting the center of the primary camera of stereo rig j).

Now, we can formulate an error function that evaluates the distance between the measurement and our reconstruction, given the calibration of the T-shaped rig. This error is calculated in the following way:

$$\mathbf{E}_{S_j} = \mathbf{S}_j - \hat{\mathbf{S}}(\mathbf{R}_{S_j}, \mathbf{C}_{S_j}; \mathbf{P}), \quad (5)$$

where \mathbf{S}_j is the tachymeter measurement and $\hat{\mathbf{S}}(\mathbf{R}_{S_j}, \mathbf{C}_{S_j}; \mathbf{P})$ is the corresponding reconstruction of the 360° -prism w.r.t. the primary camera of the stereo rig:

$$\hat{\mathbf{S}}(\mathbf{R}_{S_j}, \mathbf{C}_{S_j}; \mathbf{P}) = \mathbf{R}_{S_j}^{-1} \mathbf{P} + \mathbf{C}_{S_j}. \quad (6)$$

Here, \mathbf{R}_{S_j} is the orientation of the camera coordinate frame and \mathbf{C}_{S_j} is the center of the primary camera. The vector \mathbf{P} is the position of the 360° -prism w.r.t. the camera coordinate frame (see figure 5 for a sketch of this geometry).

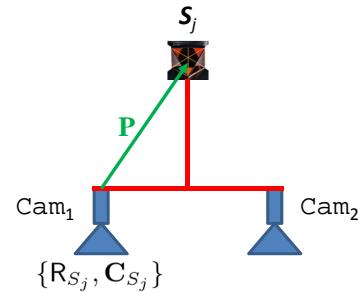


Figure 5: The geometry concerning the relationship of the primary camera and the 360° prism for the case of bundle adjustment with incorporated tachymeter measurements. The vector \mathbf{P} represents the position of the 360° prism w.r.t. the primary camera Cam_1 .

The final optimization problem can be formulated as:

$$\underset{\mathbf{R}_i, \mathbf{C}_i, \mathbf{X}_l}{\text{minimize}} \quad f(\mathbf{R}_i, \mathbf{C}_i, \mathbf{X}_l) + \alpha \sum_j^n \|\mathbf{E}_{S_j}\|_2^2, \quad (7)$$

where α controls the influence of the tachymeter measurements on the final result.

4 EXPERIMENTAL SETUP

For this paper, we conducted experiments in the lab to provide a controlled, reproducible setting for a thorough quantitative validation of the 3D geometry reconstruction method. This means that 3D-Pitoti rock-art, including ground truth had to be moved to the laboratory, and that outside-in tracking of the scanner prototype at an accuracy better than a tachymeter in the field was required.

4.1 Evaluation on ground truth data

To evaluate our scanner-prototype in the laboratory, we use high quality 3D-prints based on reconstructions from rock-art scans performed in Valcamonica to obtain a ground-truth dataset for the 3D-Pitoti project (Figure 6 shows the 3D-mesh of the ground truth dataset *rose*). The reconstructions contain a dense mesh of the area of interest as well as the texture information that is used to visualize the data and is also plotted as color onto the 3D prints. In a final step, the 3D prints were again scanned and reconstructed, obtaining the ground truth for the 3D prints as dense, textured 3D point clouds.

In contrast to the ground truth data, the results generated by the scanner prototype are sparse point clouds of the rock surface. Hence, after scanning the 3D print we first register the obtained sparse point cloud to the (dense) ground truth mesh and afterwards we calculate the Euclidean distance (mean and standard deviation) between our point cloud and the ground truth data.



Figure 6: The figure shows a visualization of the ground truth mesh *rose* that has been printed in 3D and is used as one of our test objects for laboratory measurements. The size of the print is approximately 15×15 cm.

4.2 Scanner setup

The measurement data presented in this paper was obtained by a 3D scanner prototype as depicted in Figure 1 reduced to the basic components required for geometry acquisition (i.e. no configurable lighting, microcontroller, tablet PC, and battery pack on the scanner).

The scanner is shown in Figure 7 and comprises two Canon EOS 100D DSLR cameras (C1, C2) with prime lenses (focal length 40 mm) positioned roughly 50 cm above ground. The stereo baseline was set to 17 cm (resulting in 171,4283 mm after high-accuracy camera and stereo rig calibration), yielding an overall footprint of the sensor on the ground of 40×15 cm with a comparably small stereo overlap region (6.5×15 cm). For outside-in tracking of the scanner by a stereo camera (see section 4.3), a small white sphere (A) is attached to the scanner as a replacement of the 360° prism planned for the final "walking stick" scanner. Moreover, D depicts the mount for a removable custom LED illumination component (not shown here).

4.3 Outside-in tracking

In the field, the required outside-in tracking will be performed by a tachymeter that tracks the prism mounted on top of the scanner mount. In the lab, the presented measurements were performed using a second stereo setup with large baseline to track the scanner position. For laboratory experiments, this approach offers several advantages. First, no absolute geoposition is required for the measurements, and hence the usage of a tachymeter is not mandatory. Second, at close range, the accuracy of the optical stereo setup is better than the accuracy of a tachymeter, so that the in-field accuracy can be simulated by adding noise with a magnitude corresponding to the measurement uncertainty of the tachymeter. In this way, the required accuracy of the outside-in tracking can be assessed for a given maximum uncertainty/drift of the reconstructed 3D surface.

The stereo rig used for outside-in tracking comprises two identical industrial cameras with C-mount lenses mounted on a common guide plate (stereo baseline roughly 63 cm). The setup is



Figure 7: Scanner prototype used for data acquisition: A stereo rig containing two DSLR cameras (C1, C2) and a white sphere for outside-in tracking (A) are mounted onto a carbon fibre mount (B). A mount (D) is included to properly fix a detachable custom LED illumination component (not shown here). The scanner observes the laboratory mockup consisting of a few rocks and two 3D prints (E).

used for both, tracking of the scanner position during the measurements, and for calibrating the T-shaped scanner structure prior to the measurements (compare Figure 8).

The purpose of the calibration is to exactly relate the position of the white sphere on the scanner (D in Figure 8) to the poses of the scanner's DSLR cameras. To achieve this, a target B is positioned in the field of view of the scanner prototype, and both the scanner prototype A and the external stereo rig C are used to obtain a measure of the target in their respective coordinate systems. The external stereo rig C is further used to measure the position of the white sphere D with respect to the position of the target B. Combining these measurements, the relative position of the white sphere D with respect to the cameras is computed. This relationship is required to relate the measured scanner positions (i.e. the tracked position of the white sphere D) to the positions of the acquired stereo images. Blob detection of the white sphere allows the reconstruction of its center at subpixel accuracy.

4.4 Laboratory setup

Two test scenes composed of two 3D prints of Valcamonica rock-art and several pieces of rock were set up in the laboratory to obtain the data required to assess the performance of the reconstruction algorithms under examination. Test scene I covers an area of roughly 90×30 cm and is shown in Figure 9, while test scene II is extended roughly 250×50 cm to assess the performance on larger scenes for which the standard SfM approaches typically reveal a significant drift.

The test scenes can be used to evaluate the quality of the 3D reconstruction (e.g. by comparing the resulting point cloud to the ground truth available for the two 3D prints positioned in the



Figure 8: Setup for calibration of the T-shaped scanner structure: The scanner prototype A tracks a target B positioned in its field-of-view, and the external stereo camera rig C simultaneously tracks the target B and the white sphere D on the the scanner structure that is afterwards used to track the scanner position.

scene), and to assess several parameters that are of practical interest for the scanner construction and adjustment (e.g. preferable aperture value to ensure a sufficient depth of focus, required lighting and exposure times to obtain well-exposed images for most of the rock materials to be expected).



Figure 9: Test scene I as used for measurements of a compact scene: Two 3D prints of Valcamonica rock-art and some rock samples are arranged to comprise a scene of roughly 90 cm by 30 cm.



Figure 10: Test scene II as used for measurements to assess the drift of the reconstruction for larger rock panels: Two 3D prints of Valcamonica rock-art, a concrete plate and several rock samples are arranged to comprise a scene of roughly 250 cm by 50 cm.

To allow an easy positioning of the scanner, no power or data cabling was used during the experiments. All images were stored on the internal storage cards of the cameras, and the shutter was released using an infrared remote control. A tungsten studio flood-light was used for indirect, diffuse illumination of scene I. This lighting method could not be applied for scene II due to the larger dimensions, and therefore an LED ring light was attached to the scanner to locally illuminate the scene during image acquisition.

Figure 12 shows a sample stereo image pair obtained by the scan-

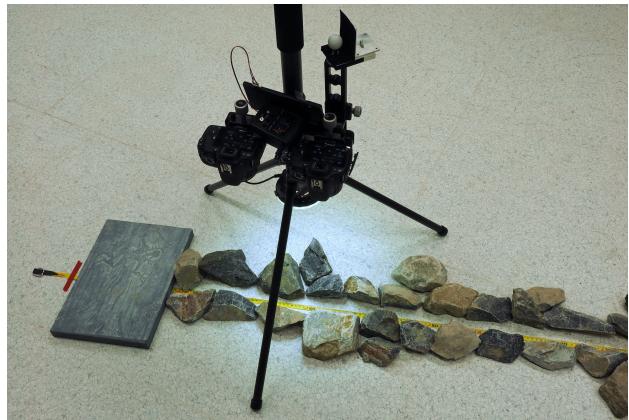


Figure 11: Application of the scanner to test scene II using LED lighting attached to the scanner in-between the cameras in order to achieve a nearly diffuse local illumination of the scene.

ner cameras showing the field of view (scanner footprint), and the small region of overlap due to the large baseline and comparably long focal length of the lenses used.



Figure 12: Typical pair of images acquired by the DSLR cameras of the scanner: The field of view of each camera is roughly 23×15 cm. Considering the small overlapping region of 6.5 cm width (marked by a box in both images), this yields an overall footprint of the scanner of 40×15 cm.

5 EXPERIMENTAL RESULTS

To evaluate the methods from section 3 we use two different error measures. To see how good our method performs locally, on a compact scene, we evaluate the mean Euclidean distance between our reconstruction and the ground-truth dataset `rose`. As a second error measure we evaluate the distance between the plane at the beginning of test scene II and the plane at the end of test scene II to validate our reconstruction on a larger scale.

5.1 Accuracy of SfM with subsequent scale enforcement

The method from section 3.1 achieves good performance with respect to the ground truth dataset `rose`. The results are presented in table 1, showing equal performances for both test scenes. This behaviour is to be expected, and is due to the fact that the dataset `rose` is compact and thus the error is evaluated locally, just inside a 15×15 cm region.

Test scene	Mean Euclidean distance [mm]	
I	$\mu_\epsilon = 0.13$	$\sigma_\epsilon = 0.20$
II	$\mu_\epsilon = 0.14$	$\sigma_\epsilon = 0.19$

Table 1: Mean Euclidean distance error μ_ϵ and standard deviation σ_ϵ for the method of section 3.1.

The method from section 3.1 scales the similarity reconstruction based on the known stereo baseline. Therefore, the standard deviation of the reconstructed stereo baselines for all stereo pairs

can also provide evidence on reconstruction accuracy. Table 2 presents the standard deviation of the reconstructed baselines for test scene I and test scene II. As expected, this measure increases with increasing size of the scene.

Test scene	Standard deviations of estimated baselines
I	$\sigma_b = 0.37$
II	$\sigma_b = 1.46$

Table 2: Standard deviation of the estimated baselines for the small test scene I and the larger test scene II.

5.2 Accuracy of SfM with enforced stereo constraint

In section 3.2 we suggest to use a quadratic penalty method to enforce the stereo-constraint. The quadratic penalty method iteratively increases the weight of the stereo constraint by iteratively decreasing μ_k . In our experiments, we observe that for decreasing μ_k the reconstruction quality drops, both in terms of reconstruction error and standard deviation. This can be seen in table 3 where the mean Euclidean distance error μ_e (between ground-truth dataset **rose** and our reconstruction) and the standard deviation σ_e are presented for different values of μ_k .

μ_k	1	$1e - 3$	$1e - 6$	$1e - 9$
μ_e [mm]	0.17	0.24	0.36	0.36
σ_e [mm]	0.28	0.39	0.53	0.53

Table 3: Mean Euclidean distance error μ_e and standard deviation σ_e for different values of μ_k .

Two further experiments show, how the stereo constraint is fulfilled. In the first experiment (see table 4), we obtain an initial scale by the method proposed in section 3.1, followed by an optimization according to equation 4. In the second experiment (see table 5), we enforce the stereo constraint during the SfM reconstruction (as proposed in section 3.2). As expected, the difference d_{baseline} between calibrated baseline and mean of the reconstructed baselines decreases with decreasing values of μ_k .

μ_k	1	$1e - 3$	$1e - 6$	$1e - 9$
d_{baseline}	0.0347	0.0172	0.0019	0.0017

Table 4: Absolute distance in mm between the length of the calibrated baseline and the mean of the reconstructed baselines for the case that the stereo constraint is enforced at the end of the reconstruction phase.

μ_k	1	$1e - 3$	$1e - 6$	$1e - 9$
d_{baseline}	3.1288	0.1741	0.0002	0.0002

Table 5: Absolute distance in mm between the length of the calibrated baseline and the mean of the reconstructed baselines for the case that the stereo constraint is enforced during the SfM reconstruction.

5.3 Influence of tachymeter measurements

This section presents the evaluation of the experiments concerning the influence of the tachymeter measurements. We first present the influence of the tachymeter measurement noise onto the final reconstruction result, followed by an assessment of the benefit of the tachymeter measurements when larger scenes are scanned.

In our first experiment on tachymeter influence, we add Gaussian noise to the measurements of our outside-in tracking system and observe the Euclidean distance error between our reconstruction

of test scene I and the ground truth data. The results of this experiment are summarized in table 6 in which σ_n is the standard deviation of the Gaussian noise that we add to the tachymeter measurements³, μ_e are the mean Euclidean distance errors between our reconstruction and the ground truth, and σ_e are the standard deviations of the Euclidean distance error. Each μ_e and σ_e was computed using different values for α and the Gaussian measurement noise σ_n . Compared to the ground-truth dataset **rose**, table

σ_n	α			
	0	100	1000	10000
0	$\mu_e = 0.13$ $\sigma_e = 0.20$	$\mu_e = 0.14$ $\sigma_e = 0.22$	$\mu_e = 0.14$ $\sigma_e = 0.21$	$\mu_e = 0.14$ $\sigma_e = 0.21$
5	$\mu_e = 0.14$ $\sigma_e = 0.21$	$\mu_e = 0.15$ $\sigma_e = 0.21$	$\mu_e = 0.32$ $\sigma_e = 0.50$	$\mu_e = 0.80$ $\sigma_e = 0.83$
10	$\mu_e = 0.13$ $\sigma_e = 0.18$	$\mu_e = 0.17$ $\sigma_e = 0.20$	$\mu_e = 0.39$ $\sigma_e = 0.48$	$\mu_e = \text{n.a.}$ $\sigma_e = \text{n.a.}$

Table 6: Influence of tachymeter measurements for test scene I. For each $\sigma_n = \{0, 5, 10\}$ and $\alpha = \{0, 20, 65, 100\}$ we calculate the mean Euclidean distance error μ_e and the standard deviation of the error σ_e . For $\sigma_n = 10$ and $\alpha = 10000$ no solution could be found.

6 shows that the mean Euclidean distance error grows with increasing measurement noise σ_n and higher weighting α of the regularization term associated with tachymeter measurements.

Structure from motion inherently suffers from drift in the final reconstruction, especially for large, elongated scenes. Our second experiment on test scene II explicitly addresses this issue, analyzing how the weighting α of the external tachymeter measurements can be used to reduce this drift. To assess the drift in our reconstruction of test scene II, we estimate the distance between two planes. The first plane is estimated using reconstructed points of the floor at the beginning of test scene II and the second plane is estimated in the same manner at the end of test scene II (see figure 10). The error measure d_{pl} is calculated as the Euclidean distance from the center of gravity of one of these plane patches, relative to the other plane, averaged over both patches. Table 7 shows the distances for different values of α .

α	0	30	400	1000	5000	10000
d_{pl}	6.78	10.89	9.62	8.42	4.37	2.47

Table 7: Mean Euclidean distance d_{pl} between the planes at the beginning and end of test scene II for different values of α .

As already mentioned in section 5.1, the standard deviation of the reconstructed baselines can also be used to assess reconstruction quality. For the experiment summarized in Table 7, we obtain $\sigma_b = 1.46$ for $\alpha = 0$, and $\sigma_b = 0.56$ for $\alpha = 10000$, indicating an improved reconstruction quality.

6 SUMMARY AND CONCLUSIONS

We have presented a novel measurement concept for the micro-range scanning of rock-art sites, that combines stereo, Structure from Motion, and outside-in tracking using a tachymeter, so that the resulting scanner prototype can be considered a high-resolution 3D reconstruction front-end to tachymeter-based surveillance in field archaeology. This paper has discussed in detail the various aspects and challenges of the geometric measurement principle, as well as first experimental 3D reconstruction results obtained with our first prototype setup in the lab.

³i.e., the measurements of the positions of the white sphere by the external stereo-rig as explained in section 4.3, see Figure 8.

In summary, most of our experimental findings meet our main expectations:

- We achieve a reconstruction accuracy better than 0.15 mm, which is already quite close to the goal of the 3D-Pitoti project. There are straightforward ways to improve this accuracy, by choosing a larger baseline, a smaller object distance, and by tilting the cameras to increase stereo overlap. However, there is a tradeoff between these obvious adaptations and various other requirements, for instance the size of the scanner footprint (the larger, the better), and its usability in terms of weight, compactness of the scanning device, and displacement from scan to scan (the larger, the more efficient scanning is possible).
- For large scenes up to complete rock-panels in Valcamonica, the drift of SfM can be significantly reduced by outside-in tracking (demonstrated by use of an external stereo-rig tracking a white sphere in the lab) that will be implemented by tracking of a 360° prism with a tachymeter.
- Enforcing the known stereo baseline for all known stereo image pairs already provides us with an excellent Euclidean reconstruction, at least for rather compact scenes.

Interestingly, however, we do not gain improved reconstruction accuracy by enforcing a stereo constraint using the quadratic penalty method as described in section 3.2. Our current explanation is twofold. First, our results seem to demonstrate the limitations of the quadratic penalty method that tends to enforce the stereo constraint in a manner that is quite rigid and strongly tends to diminish the overall reconstruction quality, both in terms of accuracy and measurement noise. Second, the incremental, online SfM method by (Hoppe et al., 2012) already produces excellent similarity reconstructions, even for larger scenes like our test scene II⁴, so that a simple scaling (using the known stereo baseline of the scanner) leads to Euclidean reconstruction up to 0.15 mm accuracy w.r.t. our ground truth. We therefore conclude that locally, at the level of the individual piece of rock-art, this method will already provide us with the required accuracy in the field.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the EC FP7 project 3D-PITOTI (ICT-2011-600545). The online incremental Structure from Motion software package was kindly provided by the authors of (Hoppe et al., 2012). The ground truth for selected items of rock art in Valcamonica has been collected, and corresponding 3D prints have been provided by *ArcTron^{3D}*.

REFERENCES

- Agarwal, S., Mierle, K. and Others, 2014. Ceres solver. <https://code.google.com/p/ceres-solver/>.
- Hartley, R. and Zisserman, A., 2003. Multiple View Geometry. 2nd edn, Cambridge University Press.
- Hoppe, C., Klopschitz, M., Rumpler, M., Wendel, A., Kluckner, S., Bischof, H. and Reitmair, G., 2012. Online feedback for structure-from-motion image acquisition. In: Proc. BMVC.
- Kurz, C., Thormählen, T. and Seidel, H.-P., 2011. Bundle adjustment for stereoscopic 3d. In: Proceedings of the 5th International Conference on Computer Vision/Computer Graphics Collaboration Techniques, MI-RAGE'11, Springer-Verlag, Berlin, Heidelberg, pp. 1–12.

⁴For example, the widely known VisualSfM method by (Wu, 2013) completely fails on test scene II, resulting in a large drift and even incorrect stitching of the 3D model.

Lhuillier, M., 2011. Fusion of gps and structure-from-motion using constrained bundle adjustments. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp. 3025–3032.

Nocedal, J. and Wright, S. J., 2006. Numerical Optimization. 2nd edn, Springer, New York.

Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A., 2000. Bundle adjustment – a modern synthesis. In: Vision Algorithms: Theory and Practice, LNCS, Springer Verlag, pp. 298–375.

Wu, C., 2013. Towards linear-time incremental structure from motion. In: Proc. IEEE Int. Conf. 3DV, pp. 127 – 134.