EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2014)
R. Klein and P. Santos (Editors)

Low-Cost Real-Time 3D Reconstruction of Large-Scale Excavation Sites using an RGB-D Camera

M. Zollhöfer, C. Siegl, B. Riffelmacher[1], M. Vetter[3], B. Dreyer[2], M. Stamminger and F. Bauer [1]

FAU Erlangen-Nuremberg - [1]Computer Graphics Group, [2]Ancient History
[3]Karlsruhe University of Applied Sciences

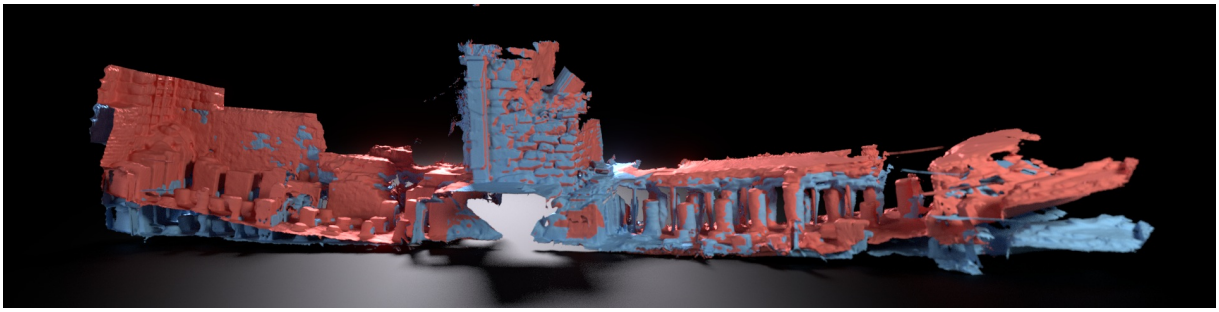

Figure 1: *A neronic bath in Metropolis, scanned (red version) and processed (blue version) using our pipeline.*

Abstract

In this paper, we present an end-to-end pipeline for the online reconstruction of large-scale outdoor environments and tightly confined indoor spaces using a low-cost consumer-level hand-held RGB-D sensor. While scanning, the user sees a live view of the current reconstruction, allowing him to intervene immediately and to adapt the sensor path to the current scanning result. After a raw reconstruction has been acquired, we interactively warp the digital model to fit a geo-referenced map using a handle based deformation paradigm. Even large sites can be scanned within a few minutes, and no costly postprocessing is required.

We developed our prototype in cooperation with researchers from the field of ancient history and geography and extensively tested the system under real world conditions on an archeological excavation in Metropolis, Ionia, Turkey. The quality of the acquired digitized raw 3D models is evaluated by comparing them to actual imagery and a geo-referenced map of the excavation site. Our reconstructions can be used to take virtual measurements that are often required in research and are the basis for a digital preservation of our cultural heritage. In addition, digital models are a helpful tool for teaching as well as for edutainment purposes making such information accessible to the general public.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Digitizing and scanning

1. Introduction

Currently, a digital revolution is happening in the humanities. Especially in the domain of cultural heritage computer vision and graphics techniques are gaining more and more acceptance. The acquisition of digital models of excavation sites and artifacts for preservation purposes is nowadays a standard procedure in the field of archeology and classical history. Having three-dimensional models simplifies both research and education as the in-depth analysis can be conducted directly on the excavation site as well as from the home base. It is also much easier to provide students and

DOI: 10.2312/gch.20141298

researchers with detailed information and measurements of objects and sites all around the world. This has the potential to lower the cost for archeological surveys while broadening the circle of researchers that can conduct objective research on the obtained findings without requiring access to the real world artifacts or sites.

On the other hand, the acquisition of three-dimensional models is still an expensive and time consuming process which requires a certain amount of technical knowledge. This is currently often handled by additional staff increasing the cost of carrying out an excavation. In order to obtain high-quality results, it is necessary to manually post-process the gathered data in a tedious and time consuming step. This cleanup stage often takes longer than actually capturing the data and has to be performed by trained specialists as well.

Another problem is that the reconstruction often is performed in an offline step after capturing the data. When parts of the scene have not been captured or are poorly sampled, it is hard to obtain and integrate reconstructions of those areas later on. With traditional special purpose scanning approaches it is often not possible to use the same technique when capturing large scale outdoor scenes and tightly confined indoor spaces, because of spatial restrictions.

With the advent of the Microsoft Kinect, a very cheap depth sensor became available. Compared to traditional 3D scanning devices, the resolution of a single depth frame is poor and contains a lot of sensor noise and holes. However, these devices capture a contiguous RGB-D stream (color and depth) at real-time rates (30Hz). In the last years, there was a large body of research on computing reconstructions from the Kinect's depth stream. These approaches are typically referred to as SLAM methods (simultaneous localization and mapping) and allow the user to freely move the depth sensor through the scene. Whenever a new depth frame is captured, its relative movement w.r.t. the previous frame is determined by registering the new depth image with the previous one or the current model (*tracking*). Next, the new depth information is fused into the model (*fusion*). The reconstruction is built online, while the user moves, and the current state of the reconstruction can be immediately visualized to give instantaneous feedback. The fusion of many noisy depth images results in super-resolution reconstructions with a higher amount of detail and much less noise than an individual input frame would permit. The relatively small size of these sensors makes them easy to carry and straight forward to use even in cluttered and spatially restricted spaces (i.e. indoor scenes with debris from decaying structures).

In this paper, we examine the application of such low-cost hand-held sensors for the reconstruction of archeological sites. We present an end-to-end pipeline that allows to reconstruct narrow indoor scenes as well as whole excavation sites. While digitizing an excavation site, the user is tightly integrated in the scanning loop by having access to a live view of the reconstruction, allowing him to intervene and directly close gaps or fix errors in the scan. Due to drift in the sensor tracking, the resulting reconstruction will contain large-scale deformations. We remove these in a post-process, using an intuitive manual deformation step, that warps the reconstructed model such that it best fits geo-referenced data of the excavation site. This step generates a geo-referenced, undistorted 3d reconstruction with geographic coordinates.

The resolution of the reconstructions cannot directly compete with the quality of expensive high resolution 3D scanners, but we see three main applications for our method:

- Our method is well suited to quickly generate large-scale overview scans that can be used as global reference for smaller high-resolution reconstructions.
- In addition, our method can be used to document an excavation, for example by making a quick scan of modified regions every evening.
- In training and education, teachers and trainers can quickly generate 3D-models with very little effort, which can be used to generate a lively and immersive teaching experience.

We developed and tested the prototype of our pipeline in cooperation with researchers from ancient history and geography. We attended an excavation in Metropolis, Turkey where we captured large portions of the site as well as a confined indoor areal (see Figure 1). While the basic reconstructions have been digitized using our live system, we corrected for geometric drift by warping the reconstructions to a geo-referenced cartographic map in a separate post-processing step. The map was recorded by geographers using differential GPS.

Comparing the digitized 3D models with actual images of the site and a cartographic excavation map (a map measured and drawn by archeologists on site) shows that our approach can provide reconstruction qualities suitable for use in eduction and allows to take virtual measurements often needed in research.

The main three contributions of this work are:

- A complete end-to-end digitization pipeline using only a single commodity RGB-D sensor: starting from low resolution depth input, we compute super-resolution reconstructions in real-time.
- A method to remove global deformations and to map the reconstruction to geographic coordinates by warping the model according to a set of geo-referenced landmarks or a cartographic map.
- An evaluation of the applicability and constraints of our system under the real-world conditions of an archeological excavation.

2. Pipeline Overview

We describe an end-to-end pipeline for the acquisition of three-dimensional models using only a low-cost hand-held

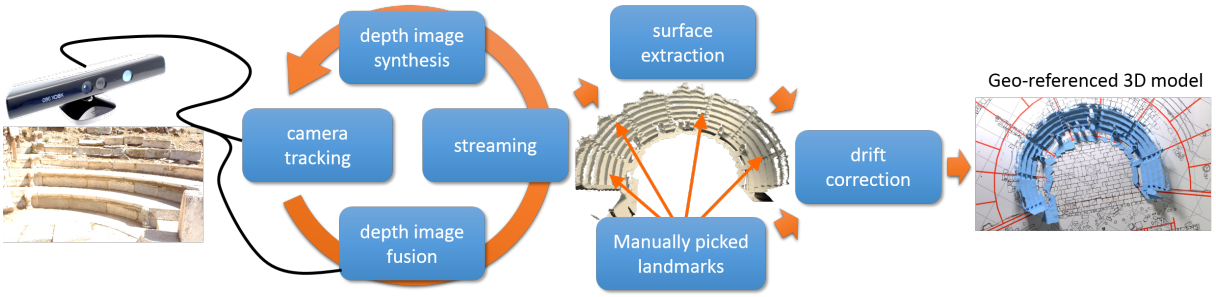

Figure 2: *Schematic view of the end-to-end pipeline we used for the online reconstruction of a large-scale excavation site.*

RGB-D sensor. Our pipeline is depicted in Fig. 2. A digital model of the excavation site is captured by walking around with a hand-held RGB-D sensor. The generated depth observations are continuously fused to obtain a high-quality 3D model that is immediately visualized. This process requires a loop of camera tracking, depth image fusion, streaming and depth image synthesis. In this work, we use the Voxel Hashing reconstruction pipeline [NZIS13], the source code of which is publically available.† In Section 5 we briefly describe the steps of this pipeline, for details we refer to the original papers.

In Section 6, we describe limitations relevant for the application in an archeological context. One major problem is that due to unavoidable inaccuracies in camera tracking (drift), the generated model will be globally deformed (compare Figure 1). The deformation becomes significant, in particular for large scenes. One way to solve this problem would be to remove or reduce sensor drift using additional acceleration sensors attached to the camera [NDF14], or to even use global tracking for the sensor (e.g. tacheometry or differential GPS). We deliberately avoided such techniques that make the scanning process much more complicated and expensive. Instead, we show that with a simple interactive post-processing step, the deformations can be easily removed. This step allows us to both compensate geometric drift and brings the model into the coordinate system of a geo-referenced map. Details of this process are described in Section 7. Our approach easily scales to whole excavation sites without compromising scanning quality, see Section 8.

3. Related Work

Depth Sensing Different types of sensors are used for the acquisition of three-dimensional data. These systems can be broadly categorized in two different paradigms: active and passive acquisition systems (for an in depth discussion pleases refer to [PLB12]). Passive systems do not interfere with the captured scene and simple RGB cameras are used for data acquisition. The reconstruction density of these methods depends on the amount of available texture features in the scene. Active systems introduce additional virtual features by actively illuminating the scene with projectors. The resulting reconstructions always have the same density. Since standard projectors would interfere with texture data, RGB-IR systems are preferred.

Noise Reduction Modern day low-cost RGB-IR sensors produce depth and color streams at real-time rates ($30Hz$), but have a small signal-to-noise ratio, making filtering a necessity [TM98]. Therefore a lot of work focuses on the analysis of the accuracy and resolution [KE] of such devices. [NIL12] explicitly learn the noise characteristic of such sensors and exploit this by using a custom filtering scheme.

Tracking The high amount of temporal coherence in the input from a moving depth sensor can only be leveraged if the exact camera pose is known for each frame. Therefore, all reconstruction techniques explicitly track the camera pose. Most of these algorithms use depth based tracking by variants of the Iterative Closest Points (ICP) algorithm [BM92, CM92a]. Methods using a point-point distance metric are based on Orthogonal Procrustes Analysis [Gow75, Ume91]. Since we have to perform an interactive registration at sensor rate, performance [RHHL02, RL01a, RL01b] and clever metrics that provide faster convergence [lL04, CM92b] play an important role. Other work employs global optimization to handle the loop closure problem [Pul99] by evenly distributing registration error. [HKH*12] present an offline solution for dealing with loop-closure using consumer grade sensors. The general process is known as Multi-View Simultaneous Localization and Mapping (SLAM) and there is a huge body of work in this research area from computer vision. [SCD*06] and [MWA*12] give a good initial overview of the general principles.

Fusion Fusion based algorithms can be roughly classified in two main categories: variants based on surfels [PZVBG00, WWL*09, WLG08, KLL*13] cluster the input depth observations based on their position and orientation. Methods based on implicit functions [CL96] fuse the observations

† https://github.com/nachtmar/VoxelHashing

into a consistent implicit representation. In our application scenario, this class of methods seems most appropriate, because it allows to easily obtain super-resolution reconstructions. The first system based on implicit functions [CL96] that demonstrated real-time performance was the KinectFusion framework [IKH*11, NIH*11]. While they demonstrate high quality reconstruction results, their approach is restricted to small scenes by the underlying uniform grid. The methods presented in [WJK*13, WJK*12] and [WKF*12] use a shifting window to circumvent this limitation. This allows for larger scale reconstructions, but their approach is still limited to a small active reconstruction volume. A different approach is Scalable Fusion [CBI13] which uses a hierarchy to store the reconstructed volumes and scales better for large scenes. In this paper, we build on the work of [NZIS13], they use a sparse scene representation based on a spatial hashing scheme. At the moment, this is the fasted system available.

Surface Reconstruction The method introduced in [TL94] can directly reconstruct a mesh from multiple depth images. However, the scanning result is normally either a point cloud or an implicit surface representation making it necessary to derive a surface mesh in a separate processing step. Most methods for high quality surface extraction first represent the surface as the zero-crossing of an implicit distance field. In the easiest case, the distance is directly computed from a point cloud based on nearest neighbor queries [HDD*92], other representations are based on radial basis functions (RBFs) [TO99, CBC*01, OBS06, SMG10], multi level partition of unity [OBA*03] or Poisson surface reconstruction [KBH06, MIPS14]. The final surface can then be extracted from this implicit representation using variants of the Marching Cubes Algorithm [LC87].

Non-rigid Mesh Deformation Handle based mesh deformation is the process of manipulating three-dimensional objects using a small number of user provided constraints. Often, these constraints are specified using an interactive editing environment. This intuitive and easy-to-use modeling metaphor allows even unexperienced users to perform complex modeling tasks. Linear methods [LSCO*04, LSA*05, ZHS*05, BS08] are fast, but have problems dealing with large rotational components in the deformation. To alleviate this restriction, non-linear deformation models [SA07] can be used. While this allows to better handle large rotations, these approaches are time consuming and do not scale to large models with millions of polygons. Therefore, approaches have been introduced that decouple the optimization problem [ZSGS12, SSP07] from the mesh by introducing a proxy, making interactive manipulations of large scale models possible.

Geo-Referencing In recent years, 3D-models have been produced using different airborne or terrestic scanning methods. However, most aquired data is not geo-referenced. Hence, the challenge is to bring the non-geo-referenced 3D-scans of an area into a geodetic reference frame. This could be realized by the application of very sophisticated, complex and expensive technologies, e.g. with geodetic instruments like 3D-Laserscanning by *total stations*. For low-cost alternatives (e.g. in our case a RGB-D sensor) geodetic information has to be acquired using further equipment. Therefore, standard GIS-Software and a lower cost differential GNSS device can be used. This instrument is based on satellite positioning technology and might produce an error of some meters which can be reduced by a signal correction of a differential station to less than a decimeter.

4. Archeological Background

The city Metropolis, where our field tests took place is a result of a Seleucid synoecism of the 3rd cent. BC. In this city indigenous inhabitants of the region as well as Seleucid-Macedonian mercenaries and Greeks were integrated. Two main building periods (as far as official buildings are concerned) can be verified, one in the 2nd cent. BC during the time of the Attalid hegemony in Asia Minor: At this time the buleuterion (see Figure 7, right), the gymnasion (at least at that time) and the theater (see Figure 4) in its first period were built. The second building period started at the beginning of the emperial period: The scene-building in the front of the theater was expanded, the bath (see Figure 1) of the neronic reign was completed, further the stoa (see Figure 3), the irrigation plan and the latrines were built. Additionally, in the second part of the 2nd cent. AD, during the reign of Antoninus Pius, a huge Thermae-building was erected [Mit83], [Mer92], [Mer04], [SA10], [Ayb14].

The theater had an important social role within the city's society. Its position within the city is sure, the present reconstruction is almost entirely a suggestion, built partly with modern concrete. The first building period belongs to the 2nd cent. BC, the second into the early 1st cent AD, especially as far as the scene building of this period is concerned. The theater of this period offers seats for about 3500 spectators. With 3D-scanning techniques, we can preserve the current geometry of the theater. In addition, this allows us to virtually undo the modern day reconstructions to determine the original, ancient shape which can be the basis for a new reconstruction based on new findings. A very important effort in archeology is applying computer based methods to avoid the need of damaging ancient remains.

The area of the gymnasion and the bath on the terrace below the double stoa walls of the Tiberian period has at least two phases, which are melted together. The period of the gymnasion in Hellenistic era - with the typical arrangements of buildings, adopted to the narrow preconditions on the terraces of the city - can hardly be discerned nowadays. Especially any traces of the obligatory place for exercise (palaestra) are missing, probably due to a lack of space. During the reign of Nero (54–68 AD) a bath (see Figure 3) was erected

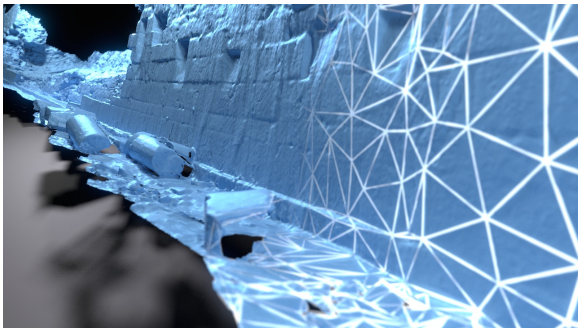

Figure 3: *The double stoa walls of the Tiberian period scanned with our system (without drift correction).*

upon parts of the gymnasion. The bath was supplied by a modern irrigation system including pipes, which also waters the latrines in the level below. It is quite common that hellenistic gymnasia are melting with Roman baths with the beginning of the Roman imperial period. But how this procedure took place in the narrow space on one terrace in the city of Metropolis, is a new research field, which can be pursued by using the 3D-models of this area to show the current state of the different building periods which can be separated thereafter and reconstructed separately.

5. Scalable Live 3D Reconstruction

Our end-to-end reconstruction pipeline has an online state-of-the-art GPU-based surface reconstruction algorithm [NZIS13] based on volumetric fusion at its heart. Leveraging the high computational power of modern consumer-level GPUs, 3D reconstruction and preview can be performed at real-time rates. This allows us to tightly integrate the staff at the excavation site into the scanning process by showing a live view of the reconstruction. In addition, this makes the technique more accessible to non-technical team members. Through this tight coupling of man and machine, reconstructions can be inspected interactively and proper adaptations of the scanning procedure can be made based on expert knowledge to improve the robustness of the acquisition process and the quality of the obtained reconstructions (i.e. yet unseen or badly sampled regions can be rescanned on-the-fly).

In the following, we summarize the main steps of the used reconstruction algorithm [NZIS13] that are also visualized in Fig. 2.

Depth-based Camera Tracking In our setup, the depth scanner is hand-held, without further tracking. For each frame, we thus have to estimate the position and orientation of the camera in a world space coordinate frame. This is achieved with a *Depth-based Odometry* approach which estimates the transformation Φ_k that maps camera space positions x at time frame k to world space positions, solely based on the current depth image. As long as intrinsic camera parameters do not change, Φ_k is a rigid transformation:

$$\Phi_k(x) = \mathbf{R}_k x + t_k,$$

with $\mathbf{R}_k$ being the camera orientation encoded as an orthogonal 3×3-matrix and t_k the absolute camera position.

We estimate the incremental rigid transform $\Phi^k_{k-1}(x) = \mathbf{R}^k_{k-1}x + t^k_{k-1}$ that best aligns the input depth map $\mathcal{D}^k$ with the current model $\mathcal{M}^{k-1}$ at time $k-1$. To this end, we compare the current depth map with a *synthesized* depth image of the current model. By using the entire model (model-to-frame) and not only the previous frame (frame-to-frame), precision of the tracking is improved, and geometric drift reduced. $(\mathbf{R}^k_{k-1}, t^k_{k-1})$ are computed in real-time using a fast GPU-based implementation of a dense point-to-plane Iterative Closest Point (ICP) algorithm that minimizes the following alignment objective:

$$\sum_{(i,j)\in\mathcal{D}} ||n^T_{\mathcal{C}(i,j)}(\Phi^k_{k-1}(\mathcal{D}^k_{i,j}) - \mathcal{M}^{k-1}_{\mathcal{C}(i,j)})||^2.$$

The mapping $\mathcal{C}(i,j)$ associates each input depth pixel (i,j) in the input image with the corresponding pixel in the model frame using fast projective correspondence association. To measure distance, we compute the distance of each sample point to the tangent plane defined by the model's surface normal $n_{\mathcal{C}(i,j)}$.

Finding the optimal solution is a non-linear least squares problem in the unknown transformation parameters that can be solved by explicitly linearizing the rotational components and iteratively solving for small linear updates. The corresponding linear $6x6$ least squares system is constructed fully in parallel on the GPU using scan operations and solved using singular value decomposition.

Depth Image Fusion After obtaining the current camera pose Φ^k we combine the depth information of the new input frame with the digital model we created thus far. We store the surface under reconstruction as the zero-crossing of a truncated signed distance field (TSDF) [CL96]. This representation has good denoising properties leading to super-resolution reconstructions. The internal representation can be visualized at real-time rates using raymarching. New depth observations m_i are fused with the model (old depth d_{i-1} and old confidence w_{i-1}) using a floating average:

$$d_i = \frac{w_{i-1}d_{i-1} + \delta w_i m_i}{w_{i-1} + \delta w_i},$$
$$w_i = \min(w_{i-1} + \delta w_i, 255).$$

Here, the weight δw_i can be used to balance noise and smoothness. We also store an RGB color value per voxel and update it using an exponential average. Compared to other fusion based approaches, we do not store a uniform dense discretization of the TSDF, but exploit the local nature of the TSDF (as proposed by [NZIS13]) to only store non-empty

space using a sparse memory efficient hash based representation. Collisions are resolved using buckets and linked lists. Note, that the number of actual collisions is small for moderate hash sizes and the incurred overhead of collision resolution is therefore negligible.

Streaming Despite the sparse nature of the scene representation, large-scale scenes might not entirely fit on the GPU. Therefore, we employ a fast streaming scheme which dynamically and transparently exchanges data between the GPU and CPU if required. This allows reconstructions with basically no limit. We use the location of the camera computed during the tracking step to only keep the relevant data in a delta neighborhood in GPU memory.

Depth Image Rendering For live visualization of the scene under reconstruction and to generate the synthetic views of the model $\mathcal{M}$ required for camera pose estimation, we use a fast GPU based raymarcher to find the zero-crossings of the implicit function along the viewing rays. We march along the ray with a step size equal to the truncation distance of the signed distance field. Because of the two level structure of the hashing scheme we have to sample 8 times when performing tri-linear interpolation. To obtain subvoxel accuracy, we use 3 bisection steps to compute refined surface positions and normals close to the zero-crossing.

High-Quality Surface Extraction To obtain a final triangle-based reconstruction result, we use a fast GPU-based implementation of the Marching Cubes algorithm that extracts the zero level set in parallel for all voxels. This extracted surface is then input for further post-processing steps and is used to store the captured datasets. If the complete scene does not fit into GPU memory, we use the described streaming approach to stream in and out all the data during the mesh extraction phase. We use append buffers to collect the extracted vertices and triangles of the mesh. After that we eliminate duplicate and close vertices per block and then perform a global pass to merge the separate meshes of all blocks. If the scene has been scanned completely the mesh based reconstruction is watertight and ready for further processing. We also extract color from the implicit function using nearest neighbor sampling at the triangle vertices.

6. Limitations

In general, the reconstruction setup turned out to be well applicable in the excavation site scenario, yet we experienced some limitations:

Camera Drift At the moment camera tracking exclusively relies on features in the captured depth data to compute the incremental updates of the camera pose. This can result in misalignments and geometric drift for feature-less or geometrically symmetric scenes (i.e aligning two parts of a sphere or segments one and seven in the theater in Figure 6, top). In these cases, the solution to the linearized system is not unique leading to camera drift. We can detect this case and warn the user by monitoring the eigenvalues of the system matrix, but cannot resolve the problem only relying on depth input. This limitation is common to all approaches that solely rely on depth data for camera tracking. Nevertheless, we show in the next section that camera drift can be well removed by deforming the reconstructed model using geo-referenced data in an easy-to-use, interactive post-process. An alternative would be to use position tracking devices for the sensor (as in [NDF14]), making the process more expensive and involved.

Lighting Conditions The above limitation of geometric drift can be fixed (as long as a sufficient amount of color information is available in the RGB input) by switching to a color based tracking scheme if instabilities in depth odometry arise. One further limitation of IR based RGB-D sensors is that scanning is restricted to indoor use and outdoor use on cloudy days, since bright sunlight interferes with the projected IR pattern and results in massive data drop-outs. Therefore, the available scanning time in Turkey was restricted to the early morning hours and night. This requirement conflicts with the acquisition of good RGB data requiring high-contrast color features.

Reconstruction Accuracy The reconstruction quality of our system is generally limited by the used virtual voxel resolution. For color information, the voxel resolution introduces a natural resolution threshold, but for depth, we store sub-voxel accurate distance values in the TSDF and thus effectively have a higher resolution than the voxel size. For good accuracy, a proper choice of the truncation range is important. In general, the truncation must be large enough (greater than the noise characteristic of the used sensor) to achieve proper denoising. Yet, too large truncation sizes degrades performance and limits the thickness of objects that can be scanned: thin objects scanned from both sides can disappear, when front and back side lie within the truncation range. This means the achievable resolution is also limited by the Kinect's signal-to-noise ratio and thus the required truncation size. For the best result, we have to choose a trun-

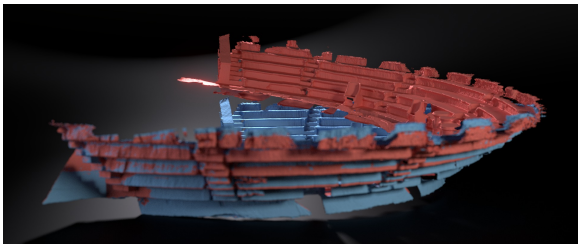

Figure 4: *Comparison of an unprocessed scan (red) with our drift corrected reconstruction (blue).*

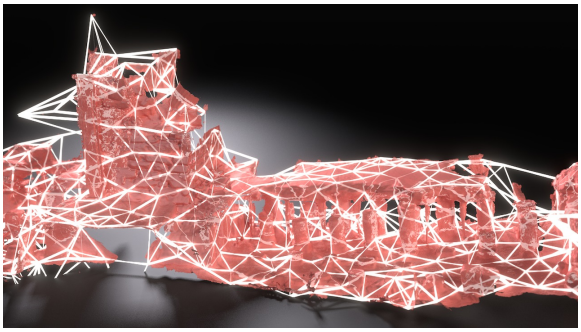

Figure 5: *The deformation graph of the neronic bath. Notice the connecting edges between unconnected parts (top left of the image).*

cation at least of the size such that we can efficiently average all sample points corresponding to the same surface. So at the moment the poor quality of a single depth frame remains the limiting factor and not the virtual resolution the presented pipeline can work at.

Loop Closure A well known problem of SLAM-based reconstruction methods is *loop closure*. If the user makes a 360°-turn the final estimated position of the sensor will be off by a few degrees, and objects that become visible again will not coincide with the first observation. As a result, gaps or ghost objects appear. Solving this problem requires global optimization procedures (e.g. [ZK13]), that are computationally too expensive for our scenario. In all shown reconstructions we thus avoided such turns and scanned using a zig-zag-pattern.

7. Geo-referencing

The previously generated models are usually globally deformed due to sensor drift, and a lack of any geo-references. We solve both issues using a geo-referenced map or geo-referenced landmarks combined with an interactive and intuitive handle-based mesh deformation metaphor allowing even unexperienced users to easily improve the reconstructions by dragging a few handles.

7.1. Geo-referenced Landmarks and Maps

To get geo-referenced landmarks of the excavation site, a differential station by Izmir was used. The accuracy of the measured points was validated with maps and GIS-Data of the Turkish survey institutes. For this purpose, a geodetic transformation has to be carried out, since the official spatial data from Turkey is in the TUREF geodetic system while the GNSS (Global Navigation Satellite System) Sensors typically work in the WGS-System. This transformation was realized using ArcGIS by ESRI. In general, most existing GIS-Software is not able to perform the geo-referencing of huge point clouds. We also used these landmarks to position maps of the excavation site in global space.

7.2. Deformation Model

The geo-referenced landmarks and maps are used to undistort and geo-reference the reconstructed model. To this end, the user can manually deform the model to match a geo-referenced landmark/map (see also accompanying video).

The three-dimensional reconstructions obtained with our digitization pipeline typically consists of millions of polygons. To deal with such a vast amount of data at interactive rates, we use a deformation graph $\mathcal{G}$ [SSP07] to decouple the optimization problem from the input resolution. The deformation nodes of the proxy deformation graph $\mathcal{G}$ are com-

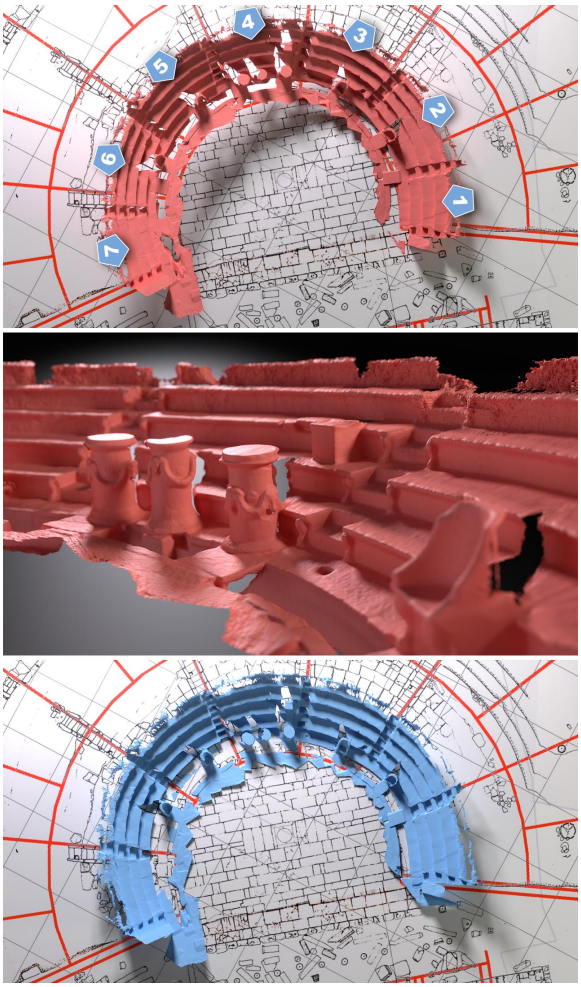

Figure 6: *From top to bottom: The raw scan of the theater over a geo-referenced map (red lines) and an excavation drawing (by S. Aybek). A closeup of segment four and three and the drift corrected theater using the geo-referenced map.*

puted via Poisson Disc Sampling. By choosing suitable radii, proxies of different resolution (for the shown examples we used between 150 and 250 nodes) can be generated. Vertices of the input model are encoded relative to the $k = 5$ nearest graph nodes. Two nodes in the graph are connected with an edge if and only if they influence a common vertex of the input model. Such a graph can be seen in Figure 5 for the bath. Each node in the graph encodes the deformation of its local neighborhood. Such a space deformation technique is oblivious to the type of input and can transparently handle triangle meshes, polygon soups and multi component models. This is a clear advantage, because the digitized models might contain multiple disconnected components.

7.3. Energy Function

The problem of finding the unknown local transformations attached to the graph nodes that best satisfy the user constraints while maintaining the local detail is cast as a non-linear optimization problem:

$$E_{total}(\mathcal{G}) = E_{fit}(\mathcal{G}) + E_{smooth}(\mathcal{G}) + E_{rigid}(\mathcal{G}).$$

Here, the total objective E_{total} is composed of an extrinsic term E_{fit} modeling the user constraints and two regularizers (E_{smooth} and E_{rigid}) that try to preserve the intrinsic properties of the model. The objective is non-linear in its unknowns and is solved using a Gauss-Newton solver. For detailed information on the energy terms, we refer to [SSP07]. Note, the user constraints are still placed on the input model not the proxy geometry,

8. Results and Evaluation

Using our inexpensive scanning system on the excavation site proved to be a huge help for the involved archeologists. However, the environment in Turkey was harsh and posed some additional difficulties. Especially the bright and hot sunlight during daytime was a challenge for the used depth sensor making it impossible to capture useful data while the sun was up. To work around this issue, we took most of our scans at night which had the added benefit of not interfering with the ongoing regular excavation.

As we expected, the scans taken with our on-line fusion based scanning software suffer from distortions. This is especially obvious in Figure 4. The red reconstruction depicts the raw model that is the direct output of our scanning software without any post-processing steps. The blue one shows the model after drift correction with our interactive deformation tool adjusting only a few handles. This step eliminated the screw like deformation of the original model.

We also evaluated the accuracy of our scans compared to a geo-referenced map (red lines in the background of Figure 6 top) of the theater. After a crude rigid alignment step on top of the map we were able to identify additional problems as the eight staircases in the theater should match the radial red lines on the background map and the staircases in the excavation drawing of that same structure. The measurements taken on the excavation map show that all staircases are equally spaced. However, in our scans we can observe that segments one and seven are much smaller (length of the secant is only 67% of the one measured in segment four). A similar (but smaller) effect is observed when we compare the remaining four segments with segment four (secant length is 97%).

This makes the theater in Figure 6 an interesting example for drifting problems. As we described in Section 5 our frame-to-model tracking is based on geometric features. If a captured frame only contains symmetrical geometry, the exact alignment of the new data can not be computed. In segments two through six there is some geometric detail on the bottom level (three columns in segment four and one elevated seat in the other four segments, see Figure 6 (middle) for a partial closeup of segments three and four). This ensures that the tracker finds always enough geometric detail to prevent the algorithm from drifting. The two outer segments do not contain such additional features. In fact, they are rotationally symmetric (for frames where we do not observe the staircases), which explains the faulty size. The best reconstructed segment is segment four, where the additional geometric information was visible in almost all captured frames.

Using the deformation graph with an additional sparse set of markers gathered from a geo-referenced map, we are able to compensate the drifting error (Figure 6, bottom). Since the correction was created using a geo-referenced map of the structure, we are able to transfer the correction to the entire model. We applied the same technique to other scenes like the neronic bath from Figure 1 with similar results. In addition, we tested our system on a variety of ancient structures on the excavation site. Some of these results are presented in Figures 7, 8 and 3 and in the accompanying video. All results show reconstructions with a voxel resolution between 5mm and 10mm.

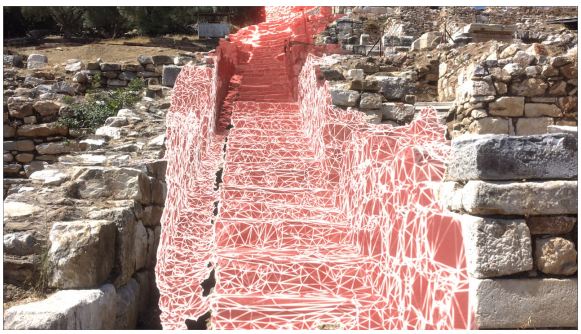

Figure 8: *Reconstruction of a staircase rendered on top of a photograph taken on the excavation site.*

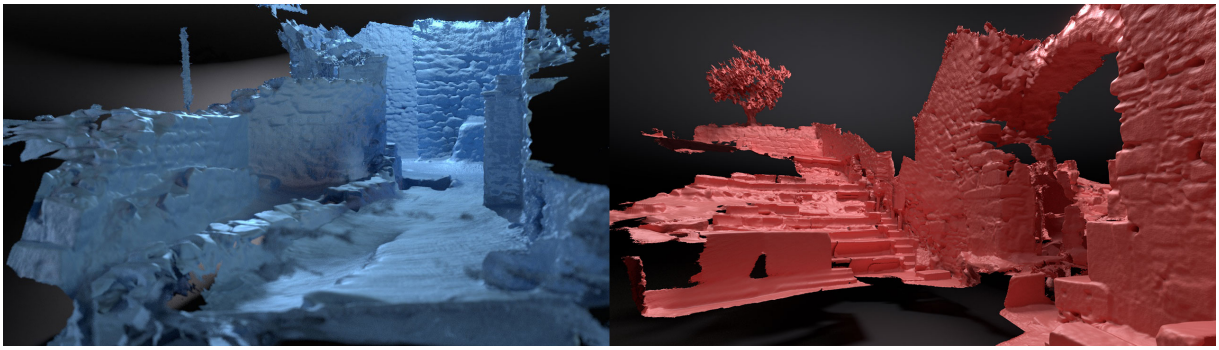

Figure 7: *Reconstruction of the backside of the neronic bath from Figure 1, as a demonstration of a tightly confined indoor scene (left). The buleuterion with a wall later built on top of the original structure (right).*

9. Conclusion

In this paper, we demonstrated an easy-to-use and robust end-to-end pipeline for digitizing large excavation sites with consumer grade affordable hardware. The scanning itself is performed using an online kinect fusion algorithm that shows a live reconstruction while walking the scene. Compared to traditional methods for capturing excavation sites (like geo-measurements or drawings), no additional setup time (like calibrating the equipment) is required. This makes scanning approachable even for non-technical staff members and allows the user to react to scanning errors directly while capturing the reconstruction.

The data gathered by the online scanning system is not geo-referenced and shows some distortions due to the low-cost hardware. Thus, we suggest a deformation graph based postprocessing step that allows to manually correct the recorded errors using an interactive and intuitive modeling metaphor. Using existing geo-referenced 2D-maps, we are able to transfer the scan into a geo-referenced coordinate system, while compensating for drift. This step produces results on par with traditional excavation maps.

References

[Ayb14] AYBEK S.: Ausgrabungen am unteren (han yikigi) römischen bad und an der palästra in metropolis (ionien): Ein kurzer bericht mit epigraphischem anhang. *Orient und Okzident in der Antike 1* (2014), 107–125. 4

[BM92] BESL P. J., MCKAY N. D.: A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell. 14*, 2 (Feb. 1992), 239–256. 3

[BS08] BOTSCH M., SORKINE O.: On linear variational surface deformation methods. *IEEE Trans. Vis. Comp. Graph 14*, 1 (2008), 213–230. 4

[CBC*01] CARR J. C., BEATSON R. K., CHERRIE J. B., MITCHELL T. J., FRIGHT W. R., MCCALLUM B. C., EVANS T. R.: Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of SIGGRAPH'01* (2001), pp. 67–76. 4

[CBI13] CHEN J., BAUTEMBACH D., IZADI S.: Scalable real-time volumetric surface reconstruction. *ACM TOG 32*, 4 (2013), 113. 4

[CL96] CURLESS B., LEVOY M.: A volumetric method for building complex models from range images. In *Proc. of the 23rd annual conference on Computer graphics and interactive techniques* (1996), ACM, pp. 303–312. 3, 4, 5

[CM92a] CHEN Y., MEDIONI G.: Object modelling by registration of multiple range images. *Image Vision Comput. 10*, 3 (Apr. 1992), 145–155. 3

[CM92b] CHEN Y., MEDIONI G.: Object modelling by registration of multiple range images. *Image Vision Comput. 10*, 3 (Apr. 1992), 145–155. 3

[Gow75] GOWER J.: Generalized procrustes analysis. *Psychometrika 40*, 1 (March 1975), 33–51. 3

[HDD*92] HOPPE H., DEROSE T., DUCHAMP T., MCDONALD J., STUETZLE W.: Surface reconstruction from unorganized points. *SIGGRAPH Comput. Graph. 26*, 2 (July 1992), 71–78. 4

[HKH*12] HENRY P., KRAININ M., HERBST E., REN X., FOX D.: RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *Int. J. Robotics Research 31* (Apr. 2012), 647–663. 3

[IKH*11] IZADI S., KIM D., HILLIGES O., MOLYNEAUX D., NEWCOMBE R., KOHLI P., SHOTTON J., HODGES S., FREEMAN D., DAVISON A., FITZGIBBON A.: KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. In *Proc. UIST* (2011), pp. 559–568. 4

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson surface reconstruction. In *Proc. EG Symp. Geometry Processing* (2006). 4

[KE] KHOSHELHAM K., ELBERINK E. O.: Accuracy and resolution of kinect depth data for indoor mapping applications. In *Sensors 2012, 12, 1437-1554*, p. 8238. 3

[KLL*13] KELLER M., LEFLOCH D., LAMBERS M., IZADI S., WEYRICH T., KOLB A.: Real-time 3d reconstruction in dynamic scenes using point-based fusion. In *Proc. of Joint 3DIM/3DPVT Conference (3DV)* (2013), IEEE, pp. 1–8. 3

[LC87] LORENSEN W. E., CLINE H. E.: Marching cubes: A high resolution 3d surface construction algorithm. In *Proc. of the 14th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1987), SIGGRAPH '87, ACM, pp. 163–169. 4

[lL04] LIM LOW K.: *Linear least-squares optimization for point-to-plane ICP surface registration*. Tech. rep., 2004. 3

[LSA*05] LIPMAN Y., SORKINE O., ALEXA M., COHEN-OR D., LEVIN D., RÖSSL C., SEIDEL H.-P.: Laplacian framework for interactive mesh editing. *International Journal of Shape Modeling (IJSM) 11*, 1 (2005), 43–61. 4

[LSCO*04] LIPMAN Y., SORKINE O., COHEN-OR D., LEVIN D., RÖSSL C., SEIDEL H.-P.: Differential coordinates for interactive mesh editing. In *Proceedings of Shape Modeling International* (2004), IEEE Computer Society Press, pp. 181–190. 4

[Mer92] MERIÇ R.: *Metropolis*. Tech. rep., Istanbul, 1992. 4

[Mer04] MERIÇ R.: *Metropolis: City of the Mother Goddess*. Metropolis Sevenler Derneği, 2004. 4

[MIPS14] MEDEIROS E., INGRID L., PESCO S., SILVA C.: Fast adaptive blue noise on polygonal surfaces. *Graph. Models 76*, 1 (Jan. 2014), 17–29. 4

[Mit83] MITCHELL S.: Recep meriç: Metropolis in ionien. ergebnisse einer survey-unternehmung in den jahren 1972–1975.(beiträge zur klassischen philologie, 142.) pp. xi+ 144; 2 maps, 113 figs, in 30 plates. königstein: Anton hain, 1982. paper, dm. 58. *The Classical Review (New Series) 33*, 02 (1983), 360–360. 4

[MWA*12] MUSIALSKI P., WONKA P., ALIAGA D., WIMMER M., VAN GOOL L., PURGATHOFER W., MITRA N., PAULY M., WAND M., CEYLAN D., ET AL.: A survey of urban reconstruction. In *Eurographics 2012-State of the Art Reports* (2012), The Eurographics Association, pp. 1–28. 3

[NDF14] NIESSNER M., DAI A., FISHER M.: Combining inertial navigation and icp for real-time 3d surface reconstruction. 3, 6

[NIH*11] NEWCOMBE R. A., IZADI S., HILLIGES O., MOLYNEAUX D., KIM D., DAVISON A. J., KOHLI P., SHOTTON J., HODGES S., FITZGIBBON A.: KinectFusion: Real-time dense surface mapping and tracking. In *Proc. ISMAR* (2011), pp. 127–136. 4

[NIL12] NGUYEN C., IZADI S., LOVELL D.: Modeling Kinect sensor noise for improved 3D reconstruction and tracking. In *Proc. Int. Conf. 3D Imaging, Modeling, Processing, Visualization and Transmission* (Oct. 2012), pp. 524–530. 3

[NZIS13] NIESSNER M., ZOLLHÖFER M., IZADI S., STAMMINGER M.: Real-time 3D reconstruction at scale using voxel hashing. *ACM TOG 32*, 6 (2013), 169. 3, 4, 5

[OBA*03] OHTAKE Y., BELYAEV A., ALEXA M., TURK G., SEIDEL H.-P.: Multi-level partition of unity implicits. *ACM Transactions on Graphics 22* (2003), 463–470. 4

[OBS06] OHTAKE Y., BELYAEV A., SEIDEL H.: Sparse surface reconstruction with adaptive partition of unity and radial basis functions. *Graphical Models 68*, 1 (2006), 15–24. 4

[PLB12] PEARS N., LIU Y., BUNTING P.: *3D imaging, analysis and applications*. Springer, 2012. 3

[Pul99] PULLI K.: Multiview registration for large data sets. In *Proc. of the 2Nd International Conference on 3-D Digital Imaging and Modeling* (Washington, DC, USA, 1999), 3DIM'99, IEEE Computer Society, pp. 160–168. 3

[PZVBG00] PFISTER H., ZWICKER M., VAN BAAR J., GROSS M.: Surfels: Surface elements as rendering primitives. In *In Proc. Computer graphics and interactive techniques* (2000), ACM Press/Addison-Wesley Publishing Co., pp. 335–342. 3

[RHHL02] RUSINKIEWICZ S., HALL-HOLT O., LEVOY M.: Real-time 3d model acquisition. *ACM Trans. Graph. 21*, 3 (July 2002), 438–446. 3

[RL01a] RUSINKIEWICZ S., LEVOY M.: Efficient variants of the ICP algorithm. In *Proc. of the Third Intl. Conf. on 3D Digital Imaging and Modeling* (2001), pp. 145–152. 3

[RL01b] RUSINKIEWICZ S., LEVOY M.: Efficient variants of the ICP algorithm. In *Third International Conference on 3D Digital Imaging and Modeling (3DIM)* (June 2001). 3

[SA07] SORKINE O., ALEXA M.: As-rigid-as-possible surface modeling. In *Proceedings of SGP'07* (2007), pp. 109–116. 4

[SA10] S. AYBEK R. M.: *Metropolis Ionia II. Land of the Crossroads*. Tech. rep., Istanbul, 2010. 4

[SCD*06] SEITZ S., CURLESS B., DIEBEL J., SCHARSTEIN D., SZELISKI R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. IEEE Conf. Comp. Vision and Pat. Rec.* (2006), vol. 1, IEEE, pp. 519–528. 3

[SMG10] SÜSSMUTH J., MEYER Q., GREINER G.: Surface reconstruction based on hierarchical floating radial basis functions. *Computer Graphics Forum 29*, 6 (2010), 1854–1864. 4

[SSP07] SUMNER R. W., SCHMID J., PAULY M.: Embedded deformation for shape manipulation. *ACM TOG 26*, 3 (2007), 80. 4, 7, 8

[TL94] TURK G., LEVOY M.: Zippered polygon meshes from range images. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1994), SIGGRAPH '94, ACM, pp. 311–318. 4

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Proc. of the Sixth International Conference on Computer Vision* (Washington, DC, USA, 1998), ICCV '98, IEEE Computer Society, pp. 839–. 3

[TO99] TURK G., O'BRIEN J. F.: *Variational Implicit Surfaces*. Tech. rep., 1999. 4

[Ume91] UMEYAMA S.: Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell. 13*, 4 (Apr. 1991), 376–380. 3

[WJK*12] WHELAN T., JOHANNSSON H., KAESS M., LEONARD J., MCDONALD J.: *Robust Tracking for Real-Time Dense RGB-D Mapping with Kintinuous*. Tech. rep., 2012. Query date: 2012-10-25. 4

[WJK*13] WHELAN T., JOHANNSSON H., KAESS M., LEONARD J. J., MCDONALD J.: Robust real-time visual odometry for dense rgb-d mapping. In *IEEE Intl. Conf. on Robotics and Automation, ICRA, Karlsruhe, Germany* (2013). 4

[WKF*12] WHELAN T., KAESS M., FALLON M., JOHANNSSON H., LEONARD J., MCDONALD J.: *Kintinuous: Spatially Extended KinectFusion*. Tech. rep., CSAIL, MIT, 2012. 4

[WLG08] WEISE T., LEIBE B., GOOL L. J. V.: Accurate and robust registration for in-hand modeling. In *CVPR* (2008), IEEE Computer Society. 3

[WWL*09] WEISE T., WISMER T., LEIBE B., , GOOL L. V.: In-hand scanning with online loop closure. In *IEEE International Workshop on 3-D Digital Imaging and Modeling* (October 2009). 3

[ZHS*05] ZHOU K., HUANG J., SNYDER J., LIU X., BAO H., GUO B., SHUM H.-Y.: Large mesh deformation using the volumetric graph laplacian. In *ACM SIGGRAPH 2005 Papers* (New York, NY, USA, 2005), SIGGRAPH '05, ACM, pp. 496–503. 4

[ZK13] ZHOU Q.-Y., KOLTUN V.: Dense scene reconstruction with points of interest. *ACM Transactions on Graphics (TOG) 32*, 4 (2013), 112. 7

[ZSGS12] ZOLLHÖFER M., SERT E., GREINER G., SÜSSMUTH J.: Gpu based arap deformation using volumetric lattices. In *Eurographics (Short Papers)* (2012), Andajar C., Puppo E., (Eds.), Eurographics Association, pp. 85–88. 4