

ORIGINAL PAPER

Improved Low-cost 3D Reconstruction Pipeline by Merging Data From Different Color and Depth Cameras

Eberty Alves da Silva ¹ and Karl Philips Apaza-Agüero ¹

¹Department of Computer Science, Federal University of Bahia (UFBA)

*eberty.silva@hotmail.com; kaguero@ufba.br

Received: 2021-10-11. Revised: 2022-07-13. Accepted: 2022-07-22.

Abstract

The performance of traditional 3D capture methods directly influences the quality of digitally reconstructed 3D models. In order to obtain complete and well-detailed low-cost three-dimensional models, this paper proposes a 3D reconstruction pipeline using point clouds from different sensors, combining captures of a low-cost depth sensor post-processed by Super-Resolution techniques with high-resolution RGB images from an external camera using Structure from Motion and Multi-View Stereo output data. The main contribution of this work includes the description of a complete pipeline that improves the stage of information acquisition and merges data from different sensors. Several phases of the 3D reconstruction pipeline were also specialized to improve the model's visual quality. The experimental evaluation demonstrates that the developed method produces good and reliable results for low-cost 3D reconstruction of an object.

Keywords: Depth Sensor, Low-Cost 3D Reconstruction, Photogrammetry.

Resumo

O desempenho dos métodos tradicionais de captura 3D influenciam diretamente na qualidade dos modelos 3D reconstruídos digitalmente. Com o objetivo de obter modelos tridimensionais de baixo custo completos e bem detalhados, este trabalho propõe um pipeline de reconstrução 3D usando nuvens de pontos de diferentes sensores, combinando capturas de um sensor de profundidade de baixo custo pós-processadas por técnicas de Super-Resolução com imagens RGB de alta resolução adquiridas por uma câmera externa aplicadas sobre algoritmos de Structure from Motion e Multi-View Stereo. A principal contribuição deste trabalho inclui a descrição de um pipeline completo que aprimora o estágio de aquisição de informações e mescla dados de diferentes sensores. Várias fases do pipeline de reconstrução 3D também foram especializadas para melhorar a qualidade visual do modelo. A avaliação experimental demonstra que o método desenvolvido produz resultados bons e confiáveis para a reconstrução 3D de baixo custo de um objeto.

Palavras-Chave: Fotogrametria, Reconstrução 3D de Baixo Custo, Sensor de Profundidade.

1 Introduction

3D reconstruction makes it possible to capture the geometry and appearance of an object or scene, allowing us to inspect details, measure properties, and reproduce 3D models in different materials (Raimundo et al., 2018). In recent years, numerous advances in 3D digitization have been observed, mainly by applying pipelines for three-dimensional reconstruction using costly high-precision 3D scanners. In addition, recent researches have sought to reconstruct objects or scenes using depth images from

low-cost acquisition devices (e.g., the Microsoft Kinect sensor (Newcombe et al., 2011)) or using Structure from Motion (SfM) (Schonberger and Frahm, 2016) combined with Multi-View Stereo (MVS) (Cerneja, 2020) from RGB images.

Good quality 3D reconstructions require a large number of financial resources, as they require state-of-the-art equipment to capture object data in high precision and detail. On the other hand, low-resolution equipment implies lower quality captures, even being financially more viable. Even with the ease of operation, lightweight,

and portability, hardware low-cost approaches must consider the limitations of the scanning equipment used (Raimundo, 2018).

The acquisition step of a 3D reconstruction pipeline refers to the use of devices to capture data from objects in a scene, such as their geometry and color (Raimundo and Apaza-Agüero, 2020). One result of 3D geometry capture is the production of discrete points collection that demonstrates the model shape. We call it point cloud. The data obtained by this step will be used in all other phases of the 3D reconstruction process (Bernardini and Rushmeier, 2002).

Active capture methods use equipment such as scanners to infer an object's geometry through a beam of light, inside or outside the visible spectrum. The scanner sensor has the advantages of fast measuring speed, robustness regarding external factors, and ease of acquiring information. Active sensors also have good performance in reconstructing texture-less and featureless surfaces (Chen et al., 2019, Raimundo and Apaza-Agüero, 2020). The sensors need to be sensitive to small variations in the information acquired, since, for small differences in distance, the variation in the time it takes to reach two different points is very low, requiring low equipment latency and good response time. For this reason, these systems tend to be slightly noisy (Raimundo et al., 2018). Considering low-cost reconstruction approaches, difficulties to capture color in high precision are a disadvantage (Hernández and Vogiatzis, 2010).

Passive methods are based on optical imaging techniques. They are highly flexible and work well with any modern digital camera. Image-based 3D reconstruction is practical, non-intrusive, low-cost and easily deployable outdoors. Various properties of the images can be used to retrieve the target shape, such as material, viewpoints and illumination. As opposed to active techniques, image-based techniques provide an efficient and easy way to acquire the color of a target object (Hernández and Vogiatzis, 2010). Although passive reconstructions mainly using SfM and MVS produce excellent results, they have limitations like the difficulty of distinguishing the target object from the background (Sergeeva and Sablina, 2018) and require the target object to having detailed geometry (Chen et al., 2019). A controlled environment is needed to obtain better reconstruction results (Hosseininaveh Ahmadabadian et al., 2019, Schonberger and Frahm, 2016).

Considering the limitations imposed by the presented approaches, it is important to note that a target whose geometry has been described by only a low-cost capture method has a real challenge in expressing its completeness, with rich and small details (Chen et al., 2019).

This paper proposes a hybrid pipeline from a low-cost depth camera (low-resolution images) and an external color capture camera (digital camera with high-resolution RGB images) to estimate and reconstruct the surface of an object and apply a high-quality texture. The individual limitations imposed by each presented low-cost capture approach are overcome by the proposed pipeline, generating a complete and well-detailed replica of the

target model with high visual quality. To achieve this effect, this project uses a variation and combination of Structure from Motion, Multi-View Stereo and depth camera capture techniques.

The main contribution of this work is the description of a low-cost and complete pipeline that makes use of post-processed depth captures and merging data from different sensors, in which depth sensor data and high-resolution color images do not need to be synchronized.

In addition to this introductory section, this work is organized as follows: Section 2 presents related works, while Section 3 describes the proposed pipeline. The experiments and evaluation of the pipeline are presented in Section 4. Finally, Section 5 discusses the final considerations and results achieved by this research.

2 Related work

Prokos et al. (2009) proposed a hybrid approach combining shape from stereo (with additional geometric constraints) and laser scanning techniques. Using two cameras and a portable laser beam, they achieved accuracy as good as some high-end laser triangulation scanners. They do not include automatically detecting outliers in their results.

The KinectFusion system (Newcombe et al., 2011) tracks the pose of portable depth cameras (Kinect) as they move through space and perform good three-dimensional surface reconstructions in real-time. The Kinect sensor has considerable limitations, including temporal inconsistency and the low resolution of the captured color and depth images (Raimundo and Apaza-Agüero, 2020). This approach does not include the texturing step.

Silva et al. (2013) provides a guided reconstruction process using Super-Resolution (SR) techniques, helping to increase the quality of the low-resolution data captured with a low-cost depth sensor. The method of data acquisition using low-cost depth cameras and SR is also improved by Raimundo and Apaza-Agüero (2020). Even with depth image improvements, a poor registration of captures can affect the final model's shape.

Falkingham (2013) demonstrates the potential applications of low-cost technology in the field of paleontology. The Microsoft Kinect was used to digitalize specimens of various sizes, and the resulting digital models were compared with models produced using SfM and MVS. The work pointed out that although Kinect generally registers morphology at a lower resolution capturing less detail than photogrammetry techniques, it offers advantages in the speed of data acquisition and generation of the 3D mesh completed in real-time during data capture. Also, they did not use Super-Resolution to improve captures from low-cost devices, and the models produced by the Kinect lack any color information.

Zollhöfer et al. (2014) used a Kinect sensor to capture the geometry of an excavation site and took advantage of a topographic map to distort the reconstructed model, significantly increasing the quality of the scene. The global distortion, with Super-Resolution techniques applied to raw scans, significantly increased the fidelity and realism

of its results but is too specialized for large scale-scenes.

Di Paola and Inzerillo (2018), in order to digitally produce the Egyptian stone from Palermo, proposed a method with a structured light scanner, smartphones and SfM to apply texture in the highly accurate mesh generated by the scanner. The main challenges were the dark color of the material and the superficiality of the groove of the hieroglyphs that some capture approaches have difficulty recognizing. The level of detail of the texture application showed up quite accurately. This reference work used a high-resolution 3D scanner, not aiming for a low-cost reconstruction.

Jo and Hong (2019) use a combination of terrestrial laser scanning and Unmanned Aerial Vehicle (UAV) photogrammetry to establish a three-dimensional model of the Magoksa Temple in Korea. The scans were used to acquire the perpendicular geometry of buildings and locations, being aligned and merged with the photogrammetry output, producing a hybrid point cloud. The photogrammetry adds value to the 3D model, complementing the point cloud with the upper parts of buildings, which are difficult to acquire through laser scanning.

Chen et al. (2019) proposes a registration method to combine the data of a laser scanner and photogrammetry to reconstruct the real outdoor 3D scene. They managed to greatly increase the accuracy and convenience of the operation. The two sensors can work independently, as the method fuses their data even if in different scales. Mesh reconstruction and texturing were not explored by this work, they also don't use MVS point clouds in their experiments.

Unlike the related works, the pipeline described in this article includes all reconstruction steps from capture to texturing, focusing on data merging using low-cost equipment.

3 Pipeline proposal

To overcome limitations of the low-cost three-dimensional data acquisition process, such as the low-resolution of depth captures with a low-cost sensor and the need for features for reconstruction by photogrammetry, and taking the advantages of each method individually, the following pipeline is proposed¹: acquisition of depth and color images (using a low-cost depth sensor and a digital camera); generation of point clouds from low-cost RGB-D camera depth images (using SR techniques (Raimundo and Apaza-Agüero, 2020)); shape estimation from RGB images (using SfM (Schonberger and Frahm, 2016) and MVS (Cernea, 2020)); aligning and merging of data from these different capture techniques; surface reconstruction; and texturing with high-quality photos (Fig. 1).

Several phases of the pipeline were specialized to achieve better accuracy and visual quality of 3D reconstructions of small and medium scale objects. The proposed pipeline works offline, allowing greater freedom in the execution of individual steps.

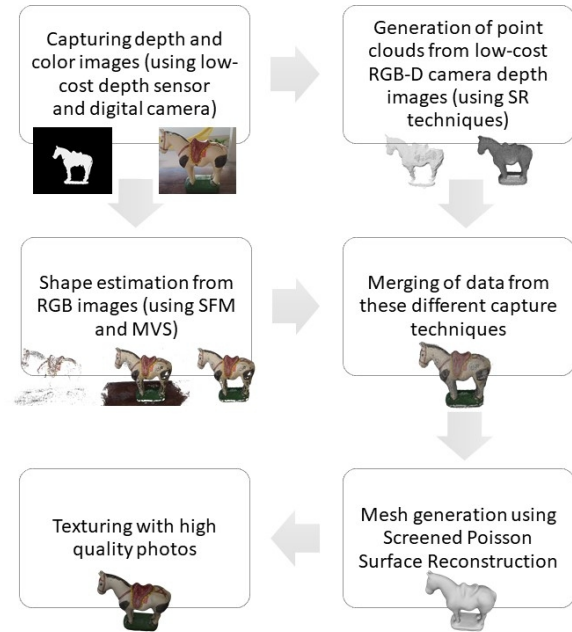


Figure 1: Schematic diagram for the proposed pipeline and the 3D reconstruction processes of an object.

3.1 Data acquisition

The data acquisition step comprises the capture of depth and color images (raw data), generation of point clouds from low-cost camera depth images and the shape estimation from RGB images (processed data). As the output, this step provides the point clouds used in the next steps of the pipeline.

3.1.1 Low-cost depth captures

For the captures using a low-cost depth sensor, we established the following acquisition procedure: take several depth captures, moving the sensor around the object, and defining the limits of the capture volume. The number of views captured is less than that of real-time approaches due to the additional processing required to ensure the quality of each capture (Raimundo and Apaza-Agüero, 2020). Considering the quality requirements for this proposed work, an interactive tool (Raimundo, 2018) is used to acquire the raw data from the depth sensor (Fig. 2).

The depth capture method will present results proportional to the best captures of the device (less noise incidence and best depth accuracy). To achieve this, each depth image, acquired by a low-cost depth sensor, goes through a filtering step with the application of Super-Resolution (Raimundo and Apaza-Agüero, 2020). In order to provide high-resolution information beyond what is possible with a specific sensor, several low-resolution captures are merged, recreating as much detail as possible.

3.1.2 Photogrammetry

In order to add 3D information in greater detail and apply a simple high-quality texturing process, photographs are

¹<https://github.com/Eberty/LowCost3DReconstruction>

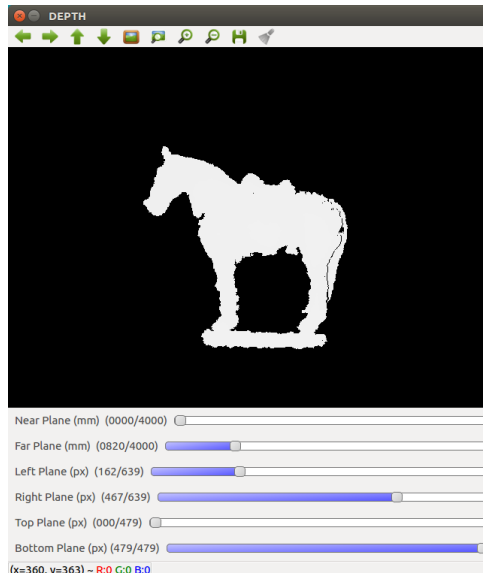


Figure 2: Software to acquire and process depth images. The slider controls the capture limits (in millimeters) and the cut limits (in pixels), effectively determining the capture volume.

taken from a digital camera around the target object. In our pipeline, these captures are independent of the depth sensor. We need just to take pictures with the fixed object, in a free movement of the camera. The set of captured images must be sufficient to cover most of the object's surface and the images must portray, in pairs, common parts of it. The color images will be used in the SfM pipeline.

The SfM pipeline detects characteristics in the images (feature detection), mapping these characteristics between images and finding descriptors capable of representing a distinguishable region (feature matching). These descriptors represent vertices of the reconstruction of the 3D scene (sparse reconstruction). The greater the number of matches found between the images, the greater the degree of accuracy of calculating a 3D transformation matrix between the images, providing the estimation of the relative position between camera poses (Hernández and Vogiatzis, 2010, Bianco et al., 2018).

Photographs with good resolution and objects with a high level of detail tend to bring greater precision to the photogrammetry algorithms. For objects with fewer details and features, the environment can be used to achieve better results (Schonberger and Frahm, 2016, Chen et al., 2019). In addition to the estimated structure to improve the depth sensor captured geometry, we use these cameras' pose estimation to apply directly texture over the final model surface.

The Multi-View Stereo process is used to improve the point cloud obtained by SfM, resulting in a dense reconstruction. As the camera parameters such as position, rotation, and focal length are known from SfM, the MVS computes 3D vertices in regions not detected by the descriptors. Multi-View Stereo algorithms generally have good accuracy, even with few images (Hernández and

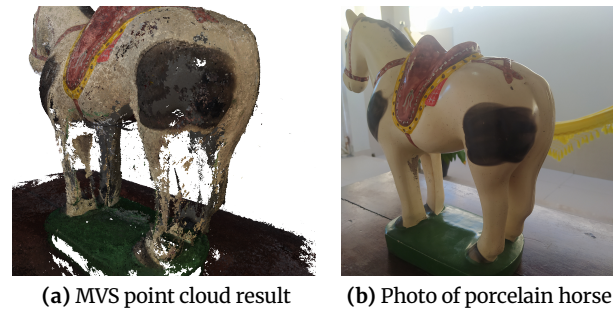


Figure 3: Some parts of the surface may not be estimated by the photogrammetry process. In (a) the white and smooth painting of the object (in (b)) prevents the MVS algorithm from obtaining a greater number of points that define this part of the structure of the model, leaving this featureless surface region with a fewer density of points than others.

Vogiatzis, 2010). A good evaluation of the performance of different state-of-the-art SfM and MVS implementations is presented by Bianco et al. (2018).

For this image-based point cloud result, a crop box filter and a euclidean cluster extraction can be used to highlighting the target object. If the floor below the object is discernible, it is also possible to use a planar segmentation algorithm to remove the plane. A statistical removal algorithm can also be used to remove outliers. Most of the discrepancies and the background are removed using the proposed steps, minimizing working time and human intervention. Details of implementation and application of the algorithms described in this paragraph are presented by Rusu and Cousins (2011).

Although image-based 3D reconstructions get greater detail than using low-cost depth sensors (Falkingham, 2013), this approach may not be able to estimate the completeness of the object (Fig. 3). This is a common result when the captures do not fully describe the target model, or it does not have a very distinguishable texture or detail (Chen et al., 2019).

3.1.3 Normal estimation

The algorithms used in the next steps require a guided set of data, thus, the normals of the point clouds are estimated before performing the alignment step. A normal estimation k-neighbor algorithm (Rusu and Cousins, 2011) is used for this task.

3.2 Alignment

The alignment task, usually called Point Cloud Registration, seeks to find the transformations which align two (or more) point sets, placing all captures in a global coordinate system. To find these transformations, the algorithm needs to establish correct correspondences between features present in each point cloud.

The registration is usually performed in a coarse and fine alignment steps. To perform the coarse alignment, we use global alignment algorithms where the pairs of

three-dimensional captures are roughly aligned (pairwise incremental registration). A good algorithm for global registration is the Super 4PCS (Mellado et al., 2014). With the captures positioned closer to the correct spot in the real-world representation, a fine adjustment step aims to align the geometric features of the objects. To do this, the Iterative Closest Point (ICP) algorithm (Holz et al., 2015) is used due to its satisfactory performance for the registration problem. This step needs to be carefully parameterized to produce good alignment results due to the nature of the depth data utilized, otherwise it may lead to drifts in the registration (Wang et al., 2016, Raimundo, 2018).

SfM approaches use a geometric verification strategy to improve the triangulation method, responsible for finding multiple planes relationship. This strategy performs a more robust camera position estimation, improving the 3D reconstruction and images projection (Schonberger and Frahm, 2016, Bianco et al., 2018). With this and the results from MVS (Cernea, 2020), we use the point cloud obtained by photogrammetry as an auxiliary to apply a new alignment over the depth sensors point clouds, distorting the initial transformation, propagating the accumulation of errors between consecutive alignments and avoiding loop closure problems (Li et al., 2013).

It is important to note that the point cloud generated by the image-based 3D reconstruction pipeline and the ones obtained with the depth sensor captures are created from different image spectrums and are very common to have different scales (Chen et al., 2019). As the depth sensor captures are already in a global coordinate system, to carry out the aligned with the corresponding points of the object in the photogrammetry point cloud, it is sufficient to find a transformation matrix to a single initially aligned depth sensor capture over the MVS point cloud by a manual process or using a scale-based iterative closest point algorithm (scale-based PCA-ICP) (Chen et al., 2019). After finding this matrix, we apply the transformation to all depth sensor point clouds. For better results, the ICP (Holz et al., 2015) algorithm can be applied for each depth sensor point cloud over the photogrammetry output point cloud. The camera positions that we will utilize for texturing will use the photogrammetry model's coordinate system.

The merging of point clouds from both data capture approaches, using an algorithm to accumulate all 3D coordinates described by each point clouds and save as a single cloud, should increase the information that defines the object geometry.

The merged point clouds are also filtered, using a statistical outlier removal algorithm (Rusu and Cousins, 2011) and down-sampled to facilitate visualization, meshing generation, and processing, since the aligned and combined point clouds may have an excessive and redundant number of vertices and there is no guarantee that the sampling density is sufficient for proper reconstruction (Bernardini and Rushmeier, 2002). A voxel grid filter Rusu and Cousins (2011) is used to downsampling the point cloud, joining points close enough. The resulting point cloud is used in the next steps of the pipeline.

3.3 Surface reconstruction

The mesh generation step is characterized by the reconstruction of the surface, a process in which a 3D continuous surface is inferred from a collection of discrete points that prove the object's shape (Berger et al., 2017).

For this step, we use the algorithm Screened Poisson Surface Reconstruction (Kazhdan and Hoppe, 2013) from MeshLab. This algorithm seeks to find a surface in which the gradient of its points is the closest to the normals of the vertices of the input point cloud. The choice of this parametric method for the surface reconstruction is justified by the robustness in the geometric fidelity and the possibility of using numerical methods to improve the results. Also, the resulting meshes are almost regular and smooth.

3.4 Texture synthesis

Applying textures to reconstructed 3D models is one of the keys to realism (Waechter et al., 2014). High-quality texture mapping aims to avoid seams, smoothing the transition of an image used for applying texture and its adjacent one (Muratov et al., 2016).

The texture synthesis phase of the proposed pipeline comprises the combination of the high-resolution pictures captured with an external digital camera with the integrated model obtained from the surface reconstruction.

The high-resolution photos taken with a digital camera with the poses calculated using SfM, will be used to perform the generation of texture coordinates and atlas of the model, avoiding a time-consuming manual process. For this texturing stage, we used the algorithm proposed by Waechter et al. (2014).

The images with poses from SfM may not be able to apply a texture on faces not visible by any image used for the reconstruction, causing non-textured mesh surfaces in the three-dimensional model. To overcome this limitation, we post-apply the texture, merging camera relative poses result from SfM with new photos, calculating the new poses using photogrammetry result relative coordinate system.

The need for additional photos is determined by non-textured surfaces in the final texturing result and is manually solved using a user interface program like MeshLab's (Cignoni et al., 2008), followed by a mutual information filter (Corsini et al., 2009) for fine adjustments finding and a transformation matrix for the new photo. Note that as this is a post-processed step, the new image of the object can be inserted even if the photo is taken from another environment.

Using a suitable output format, the new camera's pose is added to the output of the SfM module, and the texturing algorithm must be run again.

4 Experiments and evaluation

For evaluation, we run the proposed pipeline on tabletop objects of varying size and complexity. We present the results for a porcelain horse-shaped object ("Porcelain

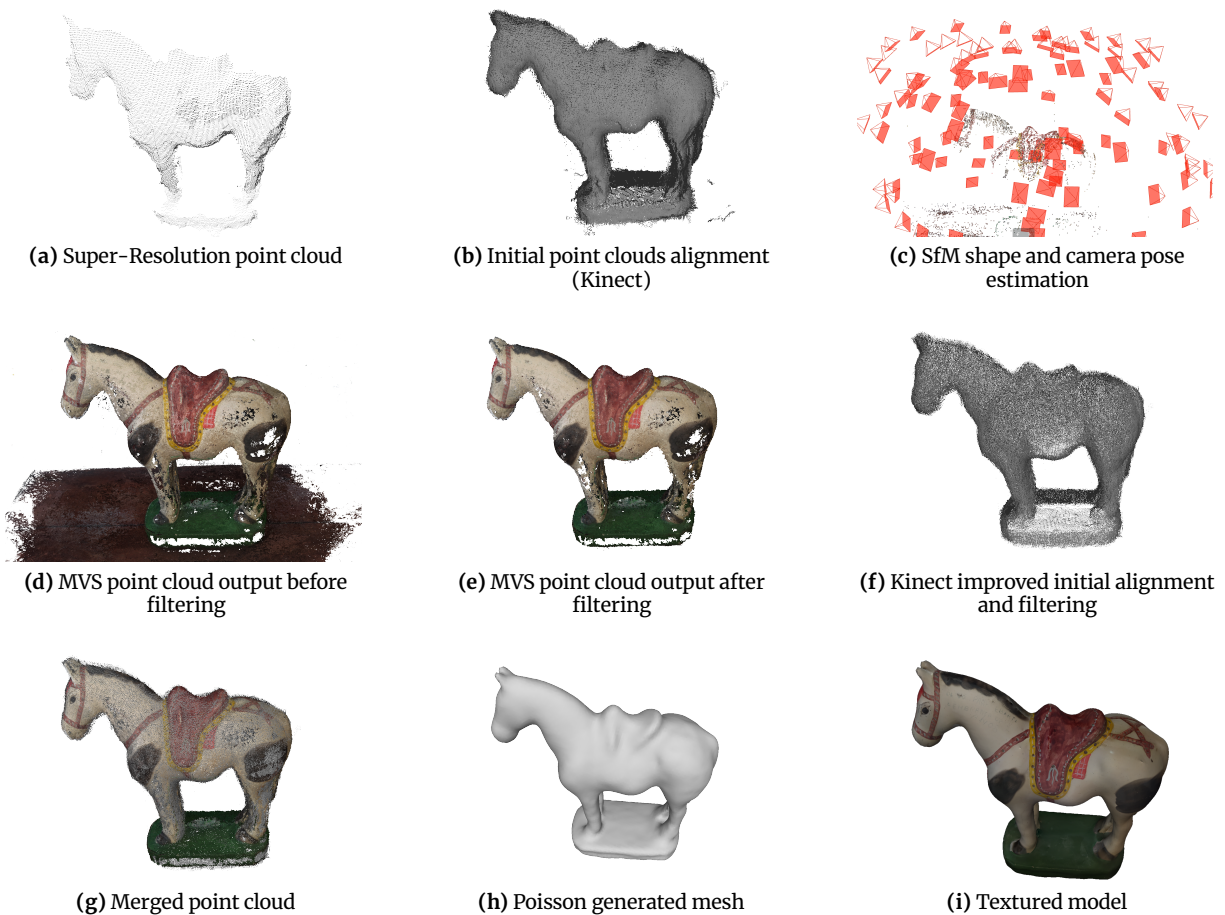


Figure 4: Porcelain horse. With the richness of details that this object has, as in the head and saddle, we use the photogrammetry method for distinguishing them with the highest level of detail. At the same time it has a low number of characteristics in predominantly smoothness regions, as the base of the structure and the body of the animal, we use the depth sensor capture approach where this factor does not influence the 3D acquisition process.

horse”, Fig. 4), a jaguar and a turtle-shaped clay pan replicas (“Jaguar pan”, Fig. 5 and “Turtle pan”, Fig. 6 respectively). These last two objects mentioned are replicas of cultural objects from the collection of Federal University of Bahia’s Brazilian Museum of Archaeology and Ethnology (MAE/UFBA) and were three-dimensionally reconstructed and 3D printed. In addition, the turtle replica was colored by hydrographic printing.

In our experiments we used Microsoft Kinect version 1, however, any other low-cost sensor can be used to capture depth images. This sensor is affordable and captures color and depth information with a resolution of 640x480 pixels. To produce point clouds from the low-cost 3D scanner, we used the Super-Resolution approach proposed by Raimundo and Apaza-Agüero (2020) with 16 Low-Resolution (LR) depth frames.

The photos used as input to the passive 3D reconstruction method were taken with a Redmi Note 8 camera for all evaluated models. The number of photos was arbitrarily chosen to maximize the coverage of the object. For the SfM pipeline, the RGB images were processed using COLMAP (Schonberger and Frahm,

2016) to calculate camera poses and sparse shape reconstruction. OpenMVS (Cerne, 2020) was used for dense reconstruction.

Some software tools were developed from third-party libraries for various purposes. For instance, OpenCV (Bradski, 2000) and PCL (Rusu and Cousins, 2011) were used to handle and process depth images and point clouds, libfreenect (OpenKinect, 2012) was used for the depth acquisition application to access and retrieve data from the Microsoft Kinect. MeshLab’s system (Cignoni et al., 2008) has been used for Poisson reconstruction and adjustments in 3D point clouds and meshes when necessary.

Details for reproducing the results can be found in the project repository.

Figs. 4 and 5 show the acquisition, merging, and reconstruction steps proposed by this pipeline for the Porcelain Horse and Jaguar Pan respectively. The figures also bring the discussion of the main challenges for each reconstruction and how they were handled by the pipeline. The algorithms and main components of each experiment are described in Table 1.

The resolution of clouds obtained by the low-cost

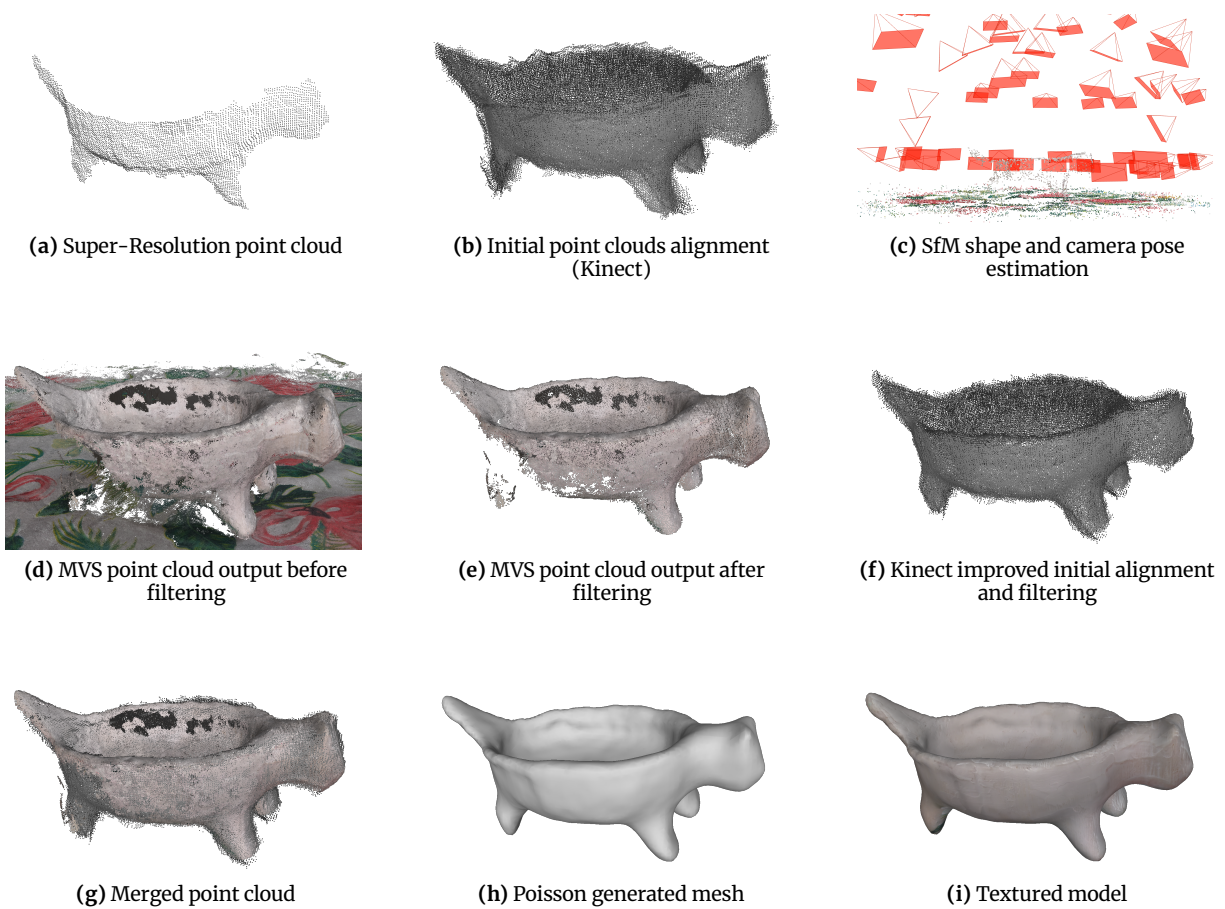


Figure 5: Jaguar pan replica. Even with some visual characteristics generated by the 3D printing process, the object has very few distinguishable features. This factor makes the reconstruction process by SfM and MVS difficult. With this, we use the environment to assist in detecting the positions and orientation of the cameras. The data captured by the low-cost depth sensor aggregated information where there are no relevant representations in the photogrammetry method, as can be seen at the legs of the jaguar.

sensor, even with SR, is considerably lower than the clouds obtained by photogrammetry and therefore these point clouds, although represent geometry well, don't describe with good precision small object details. The low-cost sensor captures also presented a scale limitation, making it difficult to retrieve the geometry of small objects such as the Turtle replica (Fig. 6). However, it has the advantage of making new captures of the object even if it has moved in the scene. The photogrammetry presented limitations when it tries to describe featureless regions of any object (as shown in Fig. 3 and Fig. 5e). This does not happen with the depth sensor, since the coloring does not influence on the captures. The point clouds obtained by photogrammetry were capable of representing, with good quality, distinguishable details on a millimeter scale. The merging of point clouds was helpful to express in greater detail the objects that were reconstructed, taking the advantages of both captures.

Point clouds were meshed using the Screened Poisson Surface Reconstruction feature in MeshLab (Cignoni et al.,

2008) using reconstruction depth 7 and 3 as the minimum number of samples. It is important to note that the production of a mesh is a highly dependent process on the variables used to generate the surface. We consider as standard for all reconstructions the Poisson Surface Reconstruction, the parameters defined in this paragraph.

For quantitative validation, the 3D surfaces reconstructions of the Jaguar were compared with a ground truth, a mesh 3D which was printed and used as the target of our reconstructions (Fig. 7f), making this analysis possible. In order to make an evaluation with previous works, this paper also provides a comparison with the results produced using publicly available implementations based in KinectFusion system, the PCL Kinfu (Rusu and Cousins, 2011) and ReconstructMe (Heindl et al., 2015). For the comparison, we used the one-sided Hausdorff Distance tool of MeshLab (Cignoni et al., 2008). The results are graphically represented on Fig. 7 and discussed on Table 2.

It was observed that the capture approach with super-

Table 1: Algorithms and main components of each experiment.

Object	Porcelain horse	Jaguar pan	Turtle pan
Dimensions (cm)	35 x 12 x 31	21.5 x 15 x 7	9 x 6.5 x 3.5
Texture	Handmade	Predominantly white	Hydrographic printing
Num. of RGB images	108	65	29
RGB images resolution	8000 x 6000px	4000 x 1844px	8000 x 6000px
SfM algorithm	COLMAP*	COLMAP*	COLMAP*
MVS algorithm	OpenMVS+	OpenMVS+	OpenMVS+
Depth sensor	Kinect V1	Kinect V1	Kinect V1
LR frames per capture	16	16	16
SR point clouds	26	22	20

* (Schonberger and Frahm, 2016)

+ (Cernea, 2020)

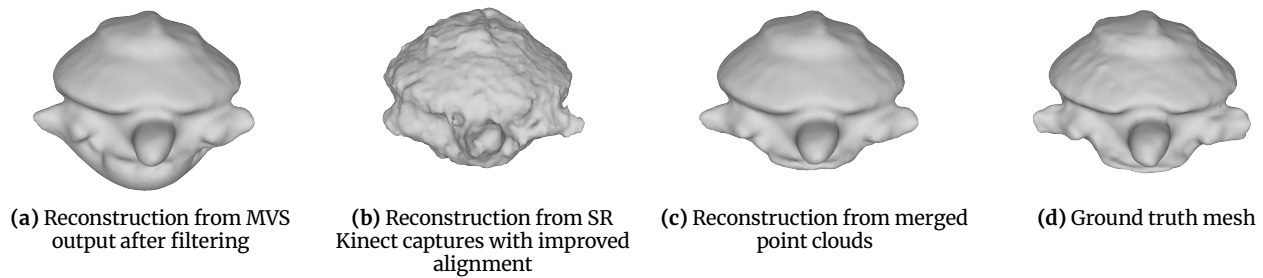


Figure 6: Screened Poisson Surface Reconstruction results for the Turtle point clouds. In (a) the limiting factor was the bottom part of the object that is not inferred by the photogrammetry process. (b) shows that the low-cost depth sensor was unable to identify details of the model, this is due to the small size of the object, making it difficult to obtain details, however, this mesh was able to represent the model in all directions, including the bottom. The merged mesh (c) was able to reproduce all the small details found by photogrammetry and include regions that were represented only by depth sensor captures. For comparison, (d) presents the model's ground truth used for 3D printing.

Table 2: One-sided Hausdorff distances for 3D surface reconstructions of the Jaguar. Each sampled face from the reconstructed models is searched to the closest point on the ground truth. Values in the mesh units and concerning the diagonal of the bounding box of the ground truth mesh. It was used a denser sampling approach over the meshes.

Mesh	MVS	Kinect SR	Merged	PCL KinFu	ReconstructMe
Vertices	12513	13034	13147	70812	27505
Faces	25022	26064	26290	23390	54500
Samples	42000	42000	42000	42000	42000
Minimum	0.000004	0.000000	0.000000	0.000000	0.000000
Maximum	0.775509	0.172886	0.136019	0.169612	0.194958
Mean	0.124142	0.017680	0.019820	0.023655	0.031021
RMS	0.208610	0.028585	0.026591	0.031335	0.043607
Reference	Fig. 7a	Fig. 7b	Fig. 7c	Fig. 7d	Fig. 7e

resolution manages to get very close to the real geometry of the object, even losing details, that can be recovered by the MVS approach. The KinectFusion based systems, once not having an adequate treatment of the data, as SR, despite representing the geometry well, is also unable to reproduce small details because of the temporal inconsistency of the Kinect.

All objects evaluated were benefited from the merging of point clouds. For the jaguar pan, the captures with the depth sensor added information in the legs of the jaguar and the belly (bottom) not acquired by photogrammetry. Poisson's surface reconstruction identifies and differentiates nearby geometric details, some of them are added by the merging. For the horse, small depressions in the mouth and eye of the original

model were not well recovered in the reconstructed model. Nonetheless, small reliefs of the saddle and mane were well-preserved.

Texturing results using surfaces from merged point clouds are shown in Figs. 4i, 5i and 8. This stage is satisfactory due to the high quality of the images used and from the camera positions correctly aligned and undistorted with the target object from SfM results.

Every procedure described in this section was performed on a notebook Avell G1550 MUV, Intel Core i7-9750H CPU @ 2.60GHz x 12, 16 GB of RAM, GeForce RTX 2070 graphics card, on Ubuntu 16.04 64-bits (Windows 10 for ReconstructMe).

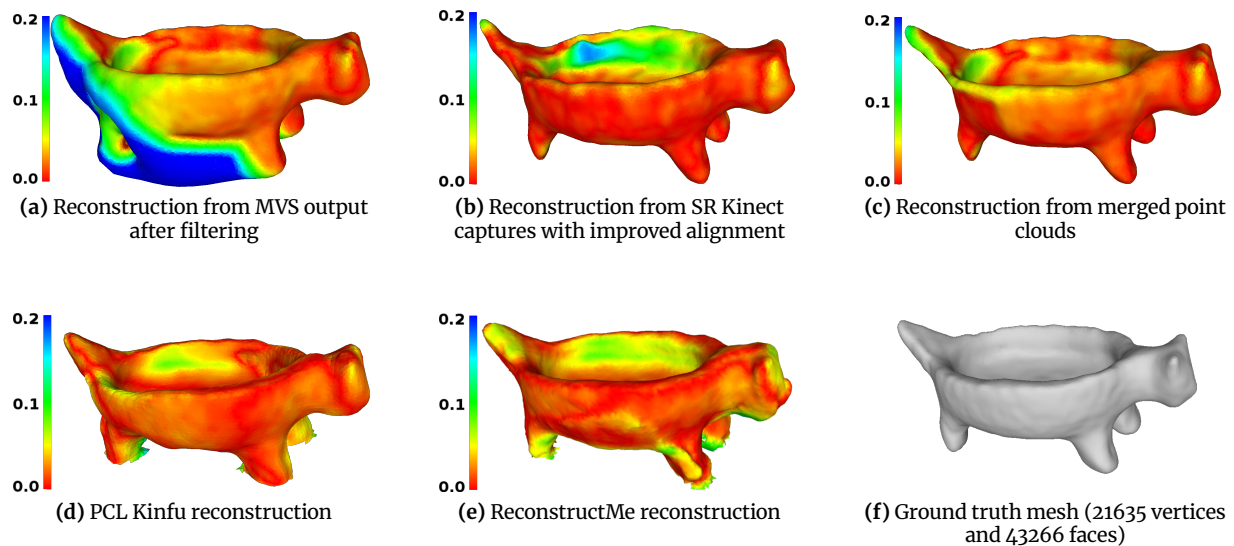


Figure 7: One-sided Hausdorff Distances for Jaguar' meshes. For each face sampled over the reconstructed meshes, were searched the closest point on ground truth (f). The red-green-blue color map was truncated to facilitate the comparison between results using a minimum of 0.0 (red, good) and a maximum of 0.2 (blue, bad). Values in the mesh units and concerning the diagonal of the bounding box of the ground truth mesh. Details of the Hausdorff Distances for each evaluated mesh are presented on Table 2. Color created using the tool "Colorized by vertex quality" from MeshLab (Cignoni et al., 2008). The meshes (d) and (e) were made using PCL Kinfu (Rusu and Cousins, 2011) and ReconstructMe (Heindl et al., 2015) respectively, with the environment and outliers manually removed.

5 Conclusion

With the proposed pipeline, it is possible to add 3D capture information, reconstructing details beyond what a single low-cost capture method initially provides. A low-cost depth sensor allows preliminary verification of data during acquisition. The Super-Resolution methodology reduces the incidence of noise and mitigates the low amount of details from depth maps acquired using low-cost RGB-D hardware. Photogrammetry, despite capturing a higher level of detail, has certain limitations related to the number of resources, like geometric and feature details.

The texturing process that uses high definition images from SfM output and adds possible missing parts, if needed, also helps to achieve greater visual realism to the reconstructed 3D model.

The proposed techniques are not novel, but adapted from known methods. Their combination to compose a low-cost 3D reconstruction is a novelty, and it demonstrates a good surface representation of tabletop objects, even for small details.

The pipeline, despite being robust, has some limitations. If after merging point clouds between different capture methods, it is still not possible to recognize part of the object, the surface reconstruction can stay too smoothed or distorted. If the registration between captures does not reach a desirable alignment result, the surface reconstruction may become bad. The photogrammetry model is usually denser than the Kinect one, which can make it difficult in the alignment and surface reconstruction process. Texturing for regions not covered by photogrammetry needs manual intervention.

Future research involves a quantitative analysis of the 3D reconstruction after the texturing step. It is also projected a better evaluation of the automated alignment of point clouds using the scale-based iterative closest point algorithm (scaled PCA-ICP) and the application of this pipeline to digital preservation of artifacts from the cultural heritage of the MAE/UFBA.

Acknowledgments

The authors would like to thank CNPq for partially financing this research through scholarship 373301/2019-7 related to the project 406719/2016-0. We also would like to thank the MAE/UFBA for making the Turtle and Jaguar pans available during our research.

References

- Berger, M., Tagliasacchi, A., Seversky, L. M., Alliez, P., Guennebaud, G., Levine, J. A., Sharf, A. and Silva, C. T. (2017). A survey of surface reconstruction from point clouds, *Computer Graphics Forum* 36(1): 301–329. <https://doi.org/10.1111/cgf.12802>.
- Bernardini, F. and Rushmeier, H. (2002). The 3d model acquisition pipeline, *Computer Graphics Forum* 21(2): 149–172. <https://doi.org/10.1111/1467-8659.00574>.
- Bianco, S., Ciocca, G. and Marelli, D. (2018). Evaluating the performance of structure from motion pipelines,

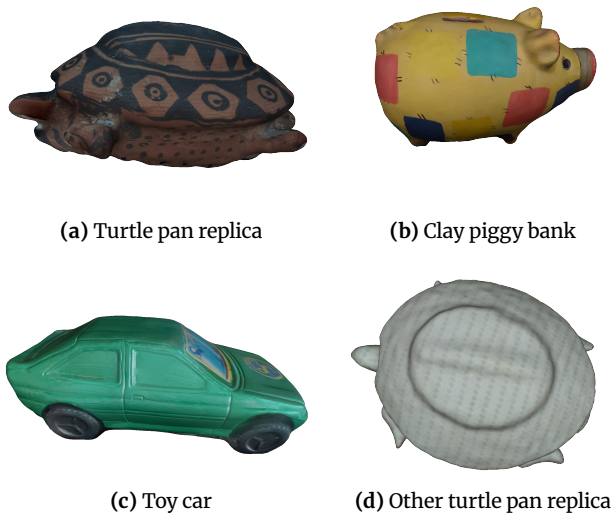


Figure 8: Texturing results using SfM cameras poses estimation over Screened Poisson Reconstruction models of the merged point clouds. In addition to the objects previously evaluated, other models generated through our pipeline are presented in this figure.

Journal of Imaging 4(8). <https://doi.org/10.3390/jimaging4080098>.

Bradski, G. (2000). The OpenCV Library, *Dr. Dobb's Journal of Software Tools*. Available at <https://opencv.org/>.

Cernea, D. (2020). OpenMVS: Multi-view stereo reconstruction library. Available at <https://cdcseacave.github.io/openMVS>.

Chen, H., Feng, Y., Yang, J. and Cui, C. (2019). 3d reconstruction approach for outdoor scene based on multiple point cloud fusion, *Journal of the Indian Society of Remote Sensing* 47(10): 1761–1772. <https://doi.org/10.1007/s12524-019-01029-y>.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F. and Ranzuglia, G. (2008). MeshLab: an Open-Source Mesh Processing Tool, in V. Scarano, R. D. Chiara and U. Erra (eds), *Eurographics Italian Chapter Conference, The Eurographics Association*. <https://doi.org/10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136>.

Corsini, M., Dellepiane, M., Ponchio, F. and Scopigno, R. (2009). Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties, *Computer Graphics Forum* 28(7): 1755–1764. <https://doi.org/10.1111/j.1467-8659.2009.01552.x>.

Di Paola, F. and Inzerillo, L. (2018). 3d reconstruction-reverse engineering – digital fabrication of the egyptian palermo stone using by smartphone and light structured scanner, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and*

Spatial Information Sciences XLII-2: 311–318. <https://doi.org/10.5194/isprs-archives-XLII-2-311-2018>.

Falkingham, P. (2013). Low cost 3d scanning using off-the-shelf video gaming peripherals, *Journal of Paleontological Techniques* 11: 1–9. Available at http://jpaleontologicaltechniques.org/pasta3/JPTN11/Pdf/JPT_n011_Jun.pdf.

Heindl, C., BAUER, H., ANKERL, M. and PICHLER, A. (2015). Reconstructme sdk: a c api for real-time 3d scanning, *6th International Conference and Exhibition on 3D Body Scanning Technologies*. <https://doi.org/10.15221/15.185>.

Hernández, C. and Vogiatzis, G. (2010). *Shape from Photographs: A Multi-view Stereo Pipeline*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 281–311. https://doi.org/10.1007/978-3-642-12848-6_11.

Holz, D., Ichim, A. E., Tombari, F., Rusu, R. B. and Behnke, S. (2015). Registration with the point cloud library: A modular framework for aligning in 3-d, *IEEE Robotics Automation Magazine* 22(4): 110–124. <https://doi.org/10.1109/MRA.2015.2432331>.

Hosseiniaveh Ahmadabadian, A., Karami, A. and Yazdan, R. (2019). An automatic 3d reconstruction system for texture-less objects, *Robotics and Autonomous Systems* 117: 29 – 39. <https://doi.org/10.1016/j.robot.2019.04.001>.

Jo, Y. H. and Hong, S. (2019). Three-dimensional digital documentation of cultural heritage site based on the convergence of terrestrial laser scanning and unmanned aerial vehicle photogrammetry, *ISPRS International Journal of Geo-Information* 8(2). <https://doi.org/10.3390/ijgi8020053>.

Kazhdan, M. and Hoppe, H. (2013). Screened poisson surface reconstruction, *ACM Trans. Graph.* 32(3): 29:1–29:13. <http://doi.acm.org/10.1145/2487228.2487237>.

Li, H., Vouga, E., Gudym, A., Luo, L., Barron, J. T. and Gusev, G. (2013). 3d self-portraits, *ACM Trans. Graph.* 32(6). <https://doi.org/10.1145/2508363.2508407>.

Mellado, N., Aiger, D. and Mitra, N. J. (2014). Super 4pcs fast global pointcloud registration via smart indexing, *Proceedings of the Symposium on Geometry Processing, SGP '14*, Eurographics Association, Goslar, DEU, p. 205–215. <https://doi.org/10.1111/cgf.12446>.

Muratov, O., Slynko, Y., Chernov, V., Lyubimtseva, M., Shamsuarov, A. and Bucha, V. (2016). 3dcapture: 3d reconstruction for a smartphone, *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 893–900. <https://doi.org/10.1109/CVPRW.2016.116>.

Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohi, P., Shotton, J., Hodges, S. and Fitzgibbon, A. (2011). Kinectfusion: Real-time dense surface mapping and tracking, *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pp. 127–136. <https://doi.org/10.1109/ISMAR.2011.6092378>.

- OpenKinect (2012). libfreenect. Available at <https://github.com/OpenKinect/libfreenect>.
- Prokos, A., Karras, G. and Grammatikopoulos, L. (2009). Design and evaluation of a photogrammetric 3d surface scanner, *hand* 2: 2.
- Raimundo, P. (2018). *Low-cost 3d reconstruction of cultural heritage*, Master's thesis, Universidade Federal da Bahia, Salvador.
- Raimundo, P. and Apaza-Agüero, K. (2020). Improved point clouds from a heritage artifact depth low-cost acquisition, *Revista Brasileira de Computação Aplicada* 12(1): 84–94. <https://doi.org/10.5335/rbca.v12i1.10019>.
- Raimundo, P., Apaza-Agüero, K. and Apolinário, A. (2018). Low-cost 3d reconstruction of cultural heritage artifacts, *Revista Brasileira de Computação Aplicada* 10(1): 66–75. <https://doi.org/10.5335/rbca.v10i1.7791>.
- Rusu, R. B. and Cousins, S. (2011). 3D is here: Point Cloud Library (PCL), *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China. <https://doi.org/10.1109/ICRA.2011.5980567>.
- Schonberger, J. L. and Frahm, J. (2016). Structure-from-motion revisited, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Los Alamitos, CA, USA, pp. 4104–4113. <https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.445>.
- Sergeeva, A. D. and Sablina, V. A. (2018). Using structure from motion for monument 3d reconstruction from images with heterogeneous background, *2018 7th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–4. <https://doi.org/10.1109/MECO.2018.8406058>.
- Silva, J. W., Gomes, L., Agüero, K. A., Bellon, O. R. P. and Silva, L. (2013). Real-time acquisition and super-resolution techniques on 3d reconstruction, *2013 IEEE International Conference on Image Processing*, pp. 2135–2139. <https://doi.org/10.1109/ICIP.2013.6738440>.
- Waechter, M., Moehrle, N. and Goesele, M. (2014). Let there be color! large-scale texturing of 3d reconstructions, in D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars (eds), *Computer Vision – ECCV 2014*, Springer International Publishing, Cham, pp. 836–850. https://doi.org/10.1007/978-3-319-10602-1_54.
- Wang, R., Wei, L., Vouga, E., Huang, Q., Ceylan, D., Medioni, G. and Li, H. (2016). Capturing dynamic textured surfaces of moving targets, in B. Leibe, J. Matas, N. Sebe and M. Welling (eds), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, pp. 271–288. https://doi.org/10.1007/978-3-319-46478-7_17.
- Zollhöfer, M., Siegl, C., Riffelmacher, B., Vetter, M., Dreyer, B., Stamminger, M. and Bauer, F. (2014). Low-Cost Real-Time 3D Reconstruction of Large-Scale Excavation Sites using an RGB-D Camera, in R. Klein and P. Santos (eds), *Eurographics Workshop on Graphics and Cultural Heritage*, The Eurographics Association. <https://doi.org/10.2312/gch.20141298>.