Evaluation of 3D Reconstruction for Cultural Heritage Applications

Cristián Llull[1], Nelson Baloian[1], Benjamin Bustos[1], Kornelius Kupczik[2], Ivan Sipiran[1], and Andrés Baloian[1]

[1]Department of Computer Science, University of Chile
[2]Departamento de Antropología, Facultad de Ciencias Sociales, Universidad de Chile

Abstract

In recent years, we have seen the emergence of methods for creating 3D digital reproductions of objects using photos. These techniques, particularly when combined with handheld video devices like smartphones, have significant applications in various fields such as medicine, museology, mechanics, and archaeology. However, previous works often lack an objective assessment of the resulting models' quality. To address this issue, the paper focuses on the systematic evaluation of reconstruction methods. This paper investigates the principles and application of the Chamfer distance, specifically the average, forward, and backward variants, for evaluating reconstructions produced by different methods: Photogrammetry, NeRF, and NVDiffrec. We also explore the impact of background filtering on the reconstructions. The ground truth for comparison is a reconstruction obtained with a structured light scanner, considered the best possible reconstruction with current technology. The results demonstrate that a comprehensive evaluation of reconstruction methods requires considering multiple measures, as they provide information about different aspects of reconstruction quality. By utilizing the Chamfer distance and comparing against the ground truth, we highlight the importance of assessing various aspects when analyzing the performance of different reconstruction methods.

1. Introduction

In recent years, there has been significant interest in developing techniques for creating digital twins using photos and videos to generate 3D representations [3]. Neural network-based algorithms utilizing multiple images from different angles have gained prominence due to their cost-effectiveness compared to high-precision scanners [39]. The applications of digital twins in various fields, including virtual museums, archaeology, palaeontology, and physical anthropology, have demonstrated their potential for improving processes and facilitating research, education, and preservation efforts.

The evaluation of reconstruction techniques in cultural heritage applications is crucial to ensure the reliability and effectiveness of the resulting 3D models. Challenges such as specularity, lack of distinctive features, and difficult lighting conditions make image-based 3D reconstruction in these contexts particularly challenging. Researchers have addressed these challenges by developing lighting models [17, 23, 34] and reconstruction techniques based on differentiable rendering [37, 36, 33], which focus on estimating scene geometry and modeling light emission and opacity.

Various 3D technologies, such as X-ray computed tomography, laser scanners, and photogrammetry, are employed in creating digital twins of human remains [11, 14, 15]. While CT offers non-invasive capturing of the complete volume, it is costly and not always accessible. Photogrammetry, particularly Structure-from-Motion, has gained popularity due to its affordability and portability using standard cameras or mobile phones. Studies comparing imaging techniques in anthropology have shown that while photogrammetry accurately represents the overall geometry of bones and teeth, laser scanner-derived models exhibit higher accuracy, finer surface details, and smaller surface features [4, 12, 28].

To address the need for evaluating and comparing 3D reconstruction methods in cultural heritage applications, this paper proposes a benchmark and evaluation methodology. We carefully select scenes, provide images and videos as inputs, and use 3D laser scanning as ground truth for comparison. Our findings reveal the challenges faced by 3D reconstruction methods in cultural heritage applications. The proposed benchmarking approach establishes a symbiotic relationship between vision and graphics technology and social sciences, such as archaeology and anthropology, paving the way for interdisciplinary advancements in these fields.

2. State of the art

Three-dimensional reconstruction is a fundamental problem in computer vision, and numerous techniques have been proposed to address it. To evaluate the performance of reconstruction algorithms, several benchmarks and methodologies have been developed (see Table 1). The Middlebury [27] and Strecha [29] benchmarks were early efforts in evaluating 3D reconstruction algorithms. They provided scenes with multi-view images and corresponding 3D reference models, focusing on Lambertian materials. These benchmarks introduced objective measures of accuracy and completeness to compare algorithms, considering the distance between points in the computed and reference models.

As learning-based methods gained popularity, the need for large-scale and high-quality data became apparent. Aanaes et al. [2] proposed a benchmark containing 80 scenes captured from multiple viewpoints under various lighting conditions, utilizing structured-light scanners to capture 3D data. The Tanks and Temple benchmark [13] introduced high-resolution geometry data captured with a Faro Focus 3D X 330 HDR scanner, offering scenes of different complexity levels. The ETH3D dataset [26] focused on providing high-resolution images and videos recorded from multiple calibrated cameras, paired with laser-scanned geometry.

Creating large-scale benchmarks using physical methods is time-consuming, leading to the exploration of synthetic data as a viable option [16, 31]. BlendedMVS [35] and PASMVS [5] employed photogrammetric methods and rendering techniques to reconstruct scenes and generate large-scale datasets. The MVImgNet dataset [38] comprises a vast number of videos with reconstructed scenes, allowing the evaluation of depth maps and direct assessment of 3D geometry.

These benchmarks and datasets play a crucial role in evaluating and advancing reconstruction algorithms. They provide standardized protocols, realistic scenes, and diverse data, enabling researchers to compare and improve the performance of their algorithms. The availability of large-scale and high-quality benchmarks is essential for driving progress in 3D reconstruction.

3. Methodology

We present a methodology for quantitatively measuring the quality of a reconstruction mesh. For this purpose, we compute how similar two distinct meshes are using the Chamfer distance [21]. One of these meshes is a reconstruction of a real-world object. The other mesh is also a reconstruction of the same real-world object obtained with a Calibry Scanner [1], which we define as the ground truth. By using the Chamfer distance, we can rank the quality of the reconstruction mesh obtained by different state-of-the-art methods from images of an object. We use the Chamfer distance in the same way recent works have used it, but unlike them, we take advantage of the asymmetric nature of this distance for presenting two different ways of interpreting the obtained results.

3.1. Chamfer Distance

This section discusses the Chamfer distance and its use in ranking the quality of reconstruction methods in Multi-view 3D reconstruction. The Chamfer distance is a measure used for quantitative evaluation and is often employed as a loss function for training deep neural networks. It involves computing the squared distances between nearest neighbor correspondences of two point clouds.

Several works, such as point set generation [8] and Photometric Mesh Optimization [18], have utilized the Chamfer distance as a loss function to train their networks. Additionally, the Chamfer distance has been used for comparing the quality of 3D reconstructions in the 3D MoMa project [20]. However, it is important to note that the Chamfer distance is primarily used to rank methods and is commonly applied to synthetic datasets.

The computation of the Chamfer distance involves summing the squared distances between nearest neighbor correspondences in two point clouds. The forward distance considers vertices from the source mesh and finds the minimum distance to the target mesh, while the backward distance reverses the process.

$$d_{CD}(S,T) = \frac{1}{2}\sum_{x\in S}\min_{y\in T}||x-y||_2^2 + \frac{1}{2}\sum_{y\in T}\min_{x\in S}||x-y||_2^2 \quad (1)$$

The Chamfer distance is sensitive to the number of points in the point cloud, so normalizing the distances to a mean distance is proposed to address this issue, as follows:

$$d_{NCD}(S,T) = \frac{1}{2|S|}\sum_{x\in S}\min_{y\in T}||x-y||_2^2 + \frac{1}{2|T|}\sum_{y\in T}\min_{x\in S}||x-y||_2^2 \quad (2)$$

To further address discrepancies in object sizes when comparing Chamfer distances, the reconstruction meshes are normalized to the size of their respective ground truth objects before measuring the Normalized Chamfer distance.

3.2. Generating models

We scanned real-world objects inside a room with enough space to use a Calibry scanner, a hand-held 3D scanner meant to capture objects from 30 cm to 10 m in length.

Benchmark	# scenes	Input	3D acquisition
Middlebury [27]	2	multi-view images	laser scanner
Strecha et al. [29]	6	multi-view images	LIDAR
DTU [2]	80	multi-view images	structured-light scanner
Tanks and temples [13]	14	multi-view images	laser scanner
ETH3D [26]	82	multi-view images and videos	laser scanner
BlendedMVS [35]	113	multi-view images	3D reconstruction
PASMVS [5]	400	multi-view images	synthetic
MVImgNet [38]	80,000	multi-view images	3D reconstruction

Table 1. Table

The room is illuminated with a mixture of artificial and natural light. Firstly, the Calibry scanner produces the reconstruction of an object. Secondly, we shot a video around the object, using FFmpeg[30]with 6 FPS frame rate for taking the images. Finally, we remove the images' background using Daniel Gatis' Rembg framework [9], based on U^2-Net [22], without any additional work. The videos were all shot with the same smartphone and of resolution 720×1280 p., with smartphone in vertical view.

We processed the obtained images with COLMAP [25, 24] using a NeRF script provided by Instant NGP [19] that gives camera positions in the format NeRF requires. This process generates a json file, normally called "transforms.json", that can be used both by NeRF and NVDiffrec methods.

For performing the NeRF reconstruction, we use the Instant NGP framework [19]. We set a minimum of $3,000$ steps and a maximum of $10,000$ in case the loss does not reach a value smaller than 0.0015 constantly through iterations. We apply the NeRF reconstruction method to both 6 FPS images and 6 FPS images without background. We found that the COLMAP process over the images without background was having problems finding good camera positions. To tackle this issue, the cameras' file (transforms.json) for 6 FPS was given as input to the images without background. We call this reconstruction as "NeRF without background with inherited cameras". In summary, we obtain three different reconstruction meshes for each object with the NeRF method. The capabilities of the PC allowed to generate meshes with resolution $404 \times 404 \times 404$ tetrahedrons.

NVDiffrec takes images and masks of the cropped object as input, but it also accepts images already cropped as masks. So, the 6 FPS images without background used for NeRF without background method are used, as they have the object already cropped. NVDiffrec also uses the same "transforms.json" file generated with COLMAP, but with some modifications. The maximum batch possible for the image's resolution and hardware capabilities was 6. The maximum reconstruction resolution for a reasonable result is $128 \times 128 \times 128$ tetrahedrons.

We computed a photometric reconstruction using Meshroom [10]. This method automatically calculates scene cameras, so there was no need to give this information as input. We generated two different models with this method: one using the 6 FPS images, and another one using the images without background.

Finally, we made a Control reconstruction. It is a simplification using MeshLab [7] of the ground truth, with the minimum amount of points without losing the shape of the model (quantitatively). It is used for comparing the reconstruction results against a reconstruction with good shape but poor definition. It is expected for the other reconstruction methods to do reconstructions with lower Chamfer distance.

The hardware used for Photogrammetry and all types of NeRF reconstruction is an AMD Ryzen 7 3750H CPU with 13 GB RAM and an NVIDIA GeForce GTX 1650 GPU with 4 GB RAM. NVDiffrec uses more resources, so an external server was used, with an Intel Core i7-9700F CPU with 128 GB RAM and a GeForce RTX 3090 GPU with 24 GB RAM.

3.3. Data collection

We collected a dataset of 16 different real-world objects for scanning and evaluating purposes. There are 14 Khachkar samples on small-scale models, each one different from each other in material, size, and color. Khachkars consist of a parallelepiped-shaped stone, with two wider faces, where a cross is usually sculpted on one of them. They are part of the Armenian cultural heritage, and now they are inscribed in the Representative List of the Intangible Cultural Heritage of Humanity[32]. There is also a small-scale model of an Armenian church. Finally, there is a real scale polyurethane resin model of a *Homo erectus* fossil skull.

The different materials, sizes, and colors of the real-world objects create different conditions for processing the images from the video, as they may affect the light on camera, the level of detail, and the contrasts. Then, only robust pipelines could give consistent good results, as the images could have better quality and sometimes cropping the background could be more difficult. The idea behind taking different models is to imitate real-world conditions when shooting videos of archaeological objects of interest. Some details on selected objects follow:

Wooden Khachkar models. Tend to have many details, as it is easier to sculpt them, and to be in an equilibrium between absorbing and reflecting light. See Khachkars 02, 03, 06, 08, 12, 13 and 14 in Figure 1.

Stone Khachkar models. They have sharp edges and shallow details and do not shine. See Khachkars 01, 04, 05, 10 and 11 in Figure 1.

Plastic Khachkar models. They have smoother edges and tend to reflect more light and present some shininess. See Khachkars 07 and 09 in Figure 1.

Church. Made out of wood. Its shape is more complex than of Khachkars, presenting great detail and overlapping edges. See Church in Figure 1.

***Homo erectus* skull replica.** The only bioanthropological object used for the experimental evaluation. It presents a lot of complex details. See the skull in Figure 1.

4. Experimental evaluation

To compare the 3D reconstructions given by the aforementioned methods, we do a manual scaling of the objects by aligning information rich points of the reconstruction to the ground truth's. This process is done with MeshLab [6], which outputs the translation matrix for scaling the mesh. By using Open 3D framework [40], an iteration over all vertices of the mesh is done and the new vertices after applying the translation matrix are stored. Please note that this process is prone to human error.

After that, we compute the normalized Chamfer distance using the ground truth. In addition, we store the values of the forward and backward distances obtained for each mesh, computed as shown at Sec. 3.1.

If a reconstruction mesh presents low forward and backward distances, which also implies a low total Chamfer distance, it means that the evaluated reconstruction method is a good approach in terms of having a model that takes good shape representation.A method with a high forward distance could mean that the reconstruction mesh contains too many vertices with respect to the ground truth, or that it contains a high amount of noise. A method with a high backward distance could mean that the reconstruction mesh is not similar to the ground truth in its details.

4.1. Results

Tables 2, 3, and 4 present the results of the normalized Chamfer distance measurements (Equation 2). Non normalized results were also stored, but to maintain a narrow scope, we exclude them. For each model, the tables report the Chamfer distance between the reconstruction and its correspondent ground truth for each technique used.

Each reconstruction takes as input the extracted images of a recorded video taken around the object, at 6 frames per second rate. "w/o BG" means extracting the background of the images and computing the camera poses with these new ones, and "inherited cameras" means computing the camera poses for the images with background and passing them to the ones without.

Total normalized Chamfer distances show Photogrammetry as the technique that produces the lowest distances, but in some cases, NeRF has lower, as in Khachkars 8, 10 and 11. Photogrammetry without background and NeRF without background have the largest Chamfer distances in general.

Photogrammetry obtained the lowest distances for normalized forward, only after Control. However, Control has the largest backward distances, probably setting a limit to how large can be the backward distance of an object. In this sense, NeRF without background for Khachkars 5 and 11 are outliers, as these distances are larger than Control distances. In general, NeRF has the lowest normalized backward distances, followed by NVDiffrec and Photogrammetry.

One interesting case is Khachkar 13 (Figure 2), being one of the three objects having Photogrammetry without background as the lowest value for total Chamfer distance. In this case, NVDiffrec presents one of its lowest results, too. Khachkar 13 also has the lowest normalized median and mean distances for each method, for Total and Backward distances, as shown in Table 5.

For the skull, Photogrammetry has the lowest forward and NeRF without background, inherited cameras, has the lowest backward. For total Chamfer distance, Photogrammetry without background is the lowest. According to the image of the reconstruction in Figure 3, the normalized Chamfer distance represents the reality.

4.2. Discussion

Normalized Chamfer distance, even though it masks the different with respect to density between the point clouds correctly, it cannot mask the differences with respect to the size of the models. This is illustrated in Table 5, where Khachkars 01 to 06, the Church and the Skull present 7 out of 8 reconstructions with 15.00 or more normalized total Chamfer distance v/s 5 out of 8 in the Khachkars 07 to 14. This is correlated with the size, as the first group is bigger. However, more data is needed to confirm this tendency.

NeRF tends to have larger distance values than Photogrammetry in total distances. This is mainly because forward distances are very high in the case of NeRF, and backward distances are lower, but not as much to counter the forward distances. This could be because NeRF reconstructions tend to have more artifacts and to be more dense than Photogrammetry. Also, NeRF prioritizes having a good visual representation, leaving the inside of the reconstruction with extra points

Figure 1. Image samples of the videos taken for Khachkars 01 to 16, the Church and the *Homo erectus* skull replica.

Table 2. Total Normalized (average of normalized forward and backward) Chamfer distances for each object and reconstruction method. The minimum distances are highlighted.

	Photogramm.	Photogramm. w/o BG	NeRF	NeRF w/o BG	NeRF w/o BG, inherited cameras	NVDiffrec	Control
Khachkar 01	4.06	41.94	6.45	35.49	23.88	26.38	241.22
Khachkar 02	3.30	42.83	3.34	14.85	13.10	11.52	197.74
Khachkar 03	7.01	50.12	9.95	21.15	12.83	13.06	256.34
Khachkar 04	3.48	27.45	3.00	19.79	18.77	18.91	109.82
Khachkar 05	3.76	30.98	9.42	1511.84	17.09	23.68	118.12
Khachkar 06	9.07	16.08	17.34	30.76	39.49	57.08	41.07
Khachkar 07	2.90	40.58	8.87	11.24	5.03	7.32	43.77
Khachkar 08	5.44	22.23	5.15	53.13	5.80	9.68	29.69
Khachkar 09	2.85	48.52	3.16	4.03	2.29	4.52	18.83
Khachkar 10	5.01	31.79	3.73	62.48	17.91	17.48	38.82
Khachkar 11	5.63	14.76	2.60	161.99	2.01	3.81	20.88
Khachkar 12	3.06	14.15	6.32	98.28	16.36	22.88	24.27
Khachkar 13	1.48	1.34	4.35	24.11	8.50	9.48	21.70
Khachkar 14	2.60	11.40	4.79	98.57	7.91	10.59	26.59
Church	4.43	4.42	51.51	41.97	49.02	17.53	50.87
Skull, upper side	62.88	61.84	72.44	103.80	88.35	92.53	19.55
Skull, full reconstruction	9.03				102.22		19.55
Mean	8.00	28.78	13.28	143.34	25.33	21.65	75.22
Median	4.06	29.21	5.73	38.73	16.36	15.27	38.82

NVDiffrec does not exhibit the last problem. However, as shown in Table 3, NVDiffrec tends to obtain larger forward distance values than NeRF. This could be explained because the background may not be removed precisely due to the use of a non-specifically trained neural network to crop the object (Rembg [9]). In this sense, NVDiffrec is

Table 3. Forward Normalized Chamfer distance for each object and reconstruction method. The minimum distances are highlighted.

	Photogramm.	Photogramm. w/o BG	NeRF	NeRF w/o BG	NeRF w/o BG, inherited cameras	NVDiffrec	Control
Khachkar 01	2.32	5.64	11.01	68.58	46.36	49.64	1.07
Khachkar 02	1.85	10.73	4.14	26.41	25.37	19.05	0.62
Khachkar 03	7.60	5.31	17.35	40.14	24.03	19.49	1.48
Khachkar 04	1.71	2.21	1.91	33.62	32.82	31.21	1.26
Khachkar 05	2.98	10.04	14.28	24.91	27.17	43.15	0.77
Khachkar 06	7.80	16.06	27.16	58.48	76.92	101.54	1.83
Khachkar 07	2.03	60.49	13.89	16.34	8.00	10.78	0.61
Khachkar 08	4.63	23.47	8.24	102.63	9.82	14.27	0.76
Khachkar 09	3.47	61.03	4.86	6.10	3.32	6.20	0.33
Khachkar 10	4.82	5.21	5.55	102.69	31.84	28.54	0.42
Khachkar 11	2.44	1.40	2.19	7.07	3.10	4.93	0.91
Khachkar 12	1.98	8.01	9.35	148.55	31.56	29.68	0.75
Khachkar 13	1.19	1.06	7.65	44.48	16.24	16.21	0.75
Khachkar 14	1.79	3.01	7.40	141.80	14.36	18.78	1.19
Church	1.94	3.46	102.09	81.58	96.94	32.12	0.49
Skull, upper side	4.67	7.89	138.94	203.00	172.90	174.76	0.49
Skull, full reconstruction	7.94				202.78		0.49
Mean	3.60	14.06	23.50	69.15	48.44	37.52	0.84
Median	2.44	6.77	8.79	51.48	27.17	24.02	0.75

Table 4. Backward Normalized Chamfer distance for each object and reconstruction method. Minimum distances are highlighted.

	Photogramm.	Photogramm. w/o BG	NeRF	NeRF w/o BG	NeRF w/o BG, inherited cameras	NVDiffrec	Control
Khachkar 01	5.79	78.23	1.89	2.39	1.40	3.12	481.36
Khachkar 02	4.75	74.92	2.53	3.29	0.83	3.99	394.87
Khachkar 03	6.42	94.92	2.56	2.17	1.62	6.62	511.21
Khachkar 04	5.26	52.69	4.10	5.96	4.72	6.62	218.38
Khachkar 05	4.54	51.92	4.55	2998.77	7.01	4.21	235.48
Khachkar 06	10.35	16.11	7.53	3.05	2.05	12.61	80.31
Khachkar 07	3.76	20.68	3.86	6.14	2.05	3.87	86.93
Khachkar 08	6.24	21.00	2.06	3.62	1.79	5.08	58.61
Khachkar 09	2.22	36.02	1.45	1.96	1.25	2.83	37.33
Khachkar 10	5.20	58.37	1.90	22.26	3.97	6.41	77.21
Khachkar 11	8.83	28.11	3.02	316.91	0.93	2.70	40.86
Khachkar 12	4.14	20.28	3.30	48.02	1.17	16.08	47.79
Khachkar 13	1.76	1.62	1.04	3.75	0.76	2.76	42.64
Khachkar 14	3.41	19.80	2.19	55.33	1.47	2.41	51.99
Church	6.91	5.38	0.94	2.36	1.11	2.94	101.24
Skull, upper side	121.10	115.79	5.95	4.59	3.81	10.31	38.60
Skull, full reconstruction	10.12				1.65		38.60
Mean	12.40	43.49	3.05	217.54	2.21	5.78	149.61
Median	5.26	32.06	2.55	4.17	1.62	4.10	77.21

Table 5. Mean and median normalized Chamfer Distance (total, forward, and backward) for each object. Minimum values are highlighted in yellow and maximums in green.

	Total Chamfer		**Forward Chamfer**		**Backward Chamfer**	
	Mean	**Median**	**Mean**	**Median**	**Mean**	**Median**
Khachkar 01	23.03	25.13	30.59	25.98	15.47	2.76
Khachkar 02	14.82	12.31	14.59	10.45	15.05	3.64
Khachkar 03	19.02	12.94	18.99	13.54	19.05	4.49
Khachkar 04	15.23	18.84	17.24	16.46	13.22	5.61
Khachkar 05	266.13	20.38	20.42	23.07	511.83	5.78
Khachkar 06	28.30	24.05	47.99	54.67	8.62	8.94
Khachkar 07	12.66	8.10	18.59	6.40	6.73	3.86
Khachkar 08	16.90	7.74	27.18	9.45	6.63	4.35
Khachkar 09	10.89	3.59	14.16	4.84	7.62	2.09
Khachkar 10	23.06	17.69	29.78	16.68	16.35	5.81
Khachkar 11	31.80	4.72	3.52	3.68	60.08	5.92
Khachkar 12	26.84	15.25	38.19	15.83	15.50	10.11
Khachkar 13	8.21	6.42	14.47	8.70	1.95	1.69
Khachkar 14	22.65	9.25	31.19	10.28	14.10	2.91
Church	28.15	29.75	53.02	17.03	3.27	2.65
Skull, upper side	80.31	80.40	117.03	89.71	43.59	8.13
Skull, full reconstruction	55.62	55.62	105.36	7.94	5.88	5.88

Figure 2. Khachkar 13 reconstructions. The minimum for total CD is Photogramm. w/o BG. The minimum for forward is Photogrammetry w/o BG and for backward is NeRF w/o BG, inherited cams.

comparable with NeRF w/o BG, inherited cameras. Indeed, results for both methods are similar for forward distances, but not for backward.

The backward normalized Chamfer distance results implies that Photogrammetry methods have more problems capturing the roughness of the models. Photogrammetry without background, specially, struggles to capture it. On the contrary, NeRF presents the lowest distances in general, competing directly with NeRF without background, inherited cameras. Despite this, the latter technique presents some outliers like Khachkar 11, where the distance grows one order of magnitude. This implies that Photogrammetry methods, despite not being the best results on backward

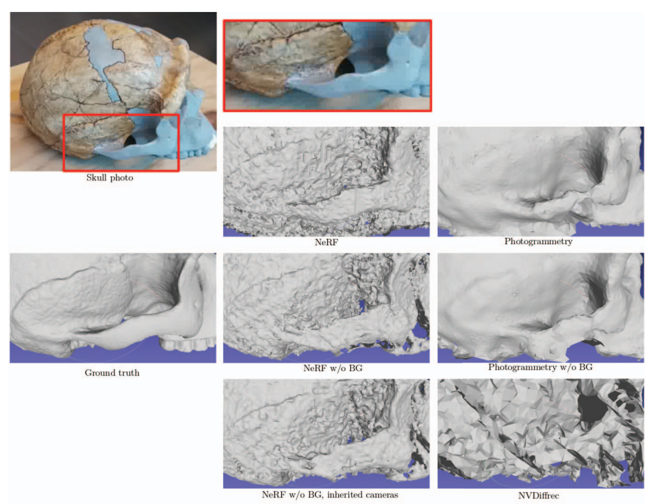

Figure 3. Reconstructions of the right side of the skull. Photogrammetry w/o BG has the lowest normalized Chamfer distance.

distance, are more reliable.

NeRF is better at capturing the complexities of the object than Photogrammetry techniques. Note the Church results shown in Table 4 for normalized backward distances. For this object, Photogrammetry has a distance around seven times larger than NeRF, on average, of the point cloud. The same is true for the Skull, as seen that NeRF has a distance two orders of magnitude lower than the Photogrammetry method. Therefore, as the scanned object complexity grows, the difference in backward distance for NeRF and Photogrammetry also grows.

Furthermore, when NeRF without background distances are too large, it is mainly due to the cameras not being well computed because of lack of information about the space where the Khachkar is located. Photogrammetry with images without background presents the same problem. Therefore, the algorithms that analyze the scene to create camera poses benefit from the background.

The case of Khachkar 13 suggests it has some ideal characteristics for background to be extracted well without additional work, and probably the details are easier to capture. This is because it has the lowest normalized results, in general. It also has low distances in all those techniques that require cropping the background, coinciding with the results in Figure 2.

5. Conclusion

We propose an objective assessment of the quality of resulting meshes from different image from video based reconstruction techniques, by decomposing the Chamfer distance in forward, backward, and total or average distances. The videos taken as input for the reconstruction methods were shot in the most possible real conditions to test the reconstruction techniques with real input. To increase the challenge, a bioarcheological object was also used. The results shown that the forward, backward, and total Chamfer distances need to be taken into account when analyzing the performance of the reconstruction methods, as some of these metrics are more influenced by the size of the model or the amount of detail present on it.

As stated in this study, there are some new challenges that need to be tackled when evaluating the quality of the models. In particular, some open questions are how to evaluate the impact of artifacts inside a point cloud that cannot be seen from outside the mesh and how to evaluate the resilience of the reconstruction technique to input with interference.

Acknowledgment

This work was funded by: ANID FONDECYT Regular N° 1230448 (B. Bustos and I. Sipiran); ANID - Millennium Science Initiative Program - Code ICN17_002 (B. Bustos).

References

[1] Thor 3D. Calibry 3d scanner, 2022.

[2] Henrik Aanæs, Rasmus Ramsbøl Jensen, George Vogiatzis, Engin Tola, and Anders Bjorholm Dahl. Large-scale data for multiple-view stereopsis. *Int. J. Comput. Vis.*, 120(2):153–168, 2016.

[3] M. Aharchi and M. Ait Kbir. A review on 3d reconstruction techniques from 2d images. In Mohamed Ben Ahmed, Anouar Abdelhakim Boudhir, Domingos Santos, Mohamed El Aroussi, and İsmail Rakıp Karas, editors, *Innovations in Smart Cities Applications Edition 3*, pages 510–522, Cham, 2020. Springer International Publishing.

[4] H. Bennani, B. McCane, and J. Cornwall. Three dimensional (3d) lumbar vertebrae data set. *Data Science Journal*, 15(0):9, 2016.

[5] André Broekman and Petrus Johannes Gräbe. Pasmvs: A perfectly accurate, synthetic, path-traced dataset featuring specular material properties for multi-view stereopsis training and reconstruction applications. *Data in Brief*, 32:106219, 2020.

[6] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. MeshLab: an Open-Source Mesh Processing Tool. In Vittorio Scarano, Rosario De Chiara, and Ugo Erra, editors, *Eurographics Italian Chapter Conference*. The Eurographics Association, 2008.

[7] Massimiliano Corsini, Paolo Cignoni, and Roberto Scopigno. Efficient and flexible sampling with blue noise properties of triangular meshes. *IEEE Transaction on Visualization and Computer Graphics*, 18(6):914–924, 2012.

[8] Haoqiang Fan, Hao Su, and Leonidas J. Guibas. A point set generation network for 3d object reconstruction from a single image. *CoRR*, abs/1612.00603, 2016.

[9] Daniel Gatis. Rembg, 2020.

[10] Carsten Griwodz, Simone Gasparini, Lilian Calvet, Pierre Gurdjos, Fabien Castan, Benoit Maujean, Gregoire De Lillo, and Yann Lanthony. Alicevision Meshroom: An open-source 3D reconstruction pipeline. In *Proceedings of the 12th ACM Multimedia Systems Conference - MMSys '21*. ACM Press, 2021.

[11] Jean-Jacques Hublin, Abdelouahed Ben-Ncer, Shara E. Bailey, Sarah E. Freidline, Simon Neubauer, Matthew M. Skinner, Inga Bergmann, Adeline Le Cabec, Stefano Benazzi, Katerina Harvati, and Philipp Gunz. New fossils from jebel irhoud, morocco and the pan-african origin of homo sapiens. *Nature*, 546(7657):289–292, 2017.

[12] David Katz and Martin Friess. Technical note: 3d from standard digital photography of human crania—a preliminary assessment. *American Journal of Physical Anthropology*, 154(1):152–158, 2014.

[13] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Trans. Graph.*, 36(4), jul 2017.

[14] Kornelius Kupczik, Lucas K. Delezene, and Matthew M. Skinner. Mandibular molar root and pulp cavity morphology

in homo naledi and other plio-pleistocene hominins. *Journal of Human Evolution*, 130:83–95, 2019.

[15] Kornelius Kupczik and Jean-Jacques Hublin. Mandibular molar root morphology in neanderthals and late pleistocene and recent homo sapiens. *Journal of Human Evolution*, 59(5):525–541, 2010.

[16] Andreas Ley, Ronny Hänsch, and Olaf Hellwich. Syb3r: A realistic synthetic benchmark for 3d reconstruction from images. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 236–251, Cham, 2016. Springer International Publishing.

[17] Junxuan Li and Hongdong Li. Self-calibrating photometric stereo by neural inverse rendering. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 166–183, Cham, 2022. Springer Nature Switzerland.

[18] Chen-Hsuan Lin, Oliver Wang, Bryan C Russell, Eli Shechtman, Vladimir G Kim, Matthew Fisher, and Simon Lucey. Photometric mesh optimization for video-aligned 3d object reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[19] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022.

[20] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Mueller, and Sanja Fidler. Extracting Triangular 3D Models, Materials, and Lighting From Images. *arXiv:2111.12503*, 2021.

[21] PDAL contributors. *PDAL: The Point Data Abstraction Library*, 2022.

[22] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. volume 106, page 107404, 2020.

[23] L. Sang, B. Hafner, X. Zuo, and D. Cremers. High-quality rgb-d reconstruction via multi-view uncalibrated photometric stereo and gradient-sdf. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3105–3114, Los Alamitos, CA, USA, jan 2023. IEEE Computer Society.

[24] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[25] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

[26] Thomas Schöps, Johannes L. Schönberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2538–2547, 2017.

[27] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 519–528, 2006.

[28] Christopher Martin Silvester and Simon Hillson. A critical assessment of the potential for structure-from-motion photogrammetry to produce high fidelity 3d dental models. *American Journal of Physical Anthropology*, 173(2):381–392, 2020.

[29] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[30] Suramya Tomar. Converting video formats with ffmpeg. *Linux Journal*, 2006(146):10, 2006.

[31] Jonathan Tremblay, Moustafa Meshry, Alex Evans, Jan Kautz, Alexander Keller, Sameh Khamis, Charles Loop, Nathan Morrical, Koki Nagano, Towaki Takikawa, and Stan Birchfield. Rtmv: A ray-traced multi-view synthetic dataset for novel view synthesis. *IEEE/CVF European Conference on Computer Vision Workshop (Learn3DG ECCVW), 2022*, 2022.

[32] Unesco. Armenian cross-stones art. Symbolism and craftsmanship of Khachkars, 2010.

[33] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Neural Information Processing Systems*, 2021.

[34] X. Wang, Y. Guo, B. Deng, and J. Zhang. Lightweight photometric stereo for facial details recovery. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 737–746, Los Alamitos, CA, USA, jun 2020. IEEE Computer Society.

[35] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1787–1796, 2020.

[36] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 4805–4815. Curran Associates, Inc., 2021.

[37] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 2492–2502. Curran Associates, Inc., 2020.

[38] Xianggang Yu, Mutian Xu, Yidan Zhang, Haolin Liu, Chongjie Ye, Yushuang Wu, Zizheng Yan, Tianyou Liang, Guanying Chen, Shuguang Cui, and Xiaoguang Han. Mvimgnet: A large-scale dataset of multi-view images. In *CVPR*, 2023.

[39] Anny Yuniarti and Nanik Suciati. A review of deep learning techniques for 3d reconstruction of 2d images. In *2019 12th*

International Conference on Information & Communication Technology and System (ICTS), pages 327–331, 2019.

[40] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018.