The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W3-2023
2nd GEOBENCH Workshop on Evaluation and BENCHmarking of Sensors, Systems and GEOspatial Data
in Photogrammetry and Remote Sensing, 23–24 October 2023, Krakow, Poland

# GEOMETRIC ACCURACY ANALYSIS BETWEEN NEURAL RADIANCE FIELDS (NERFS) AND TERRESTRIAL LASER SCANNING (TLS)

I. Petrovska*, M. Jäger, D. Haitz, B. Jutzi

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
(ivana.petrovska, miriam.jaeger, dennis.haitz, boris.jutzi)@kit.edu

**KEY WORDS:** Neural Radiance Fields, 3D Reconstruction, Accuracy Assessment, Laser Scanning, Point Cloud Comparison.

**ABSTRACT:**

Neural Radiance Fields (NeRFs) use a set of camera poses with associated images to represent a scene through a position-dependent density and radiance at given spatial location. Generating a geometric representation in form of a point cloud is gained by ray tracing and sampling 3D points with density and color along the rays. In this contribution we evaluate object reconstruction by NeRFs in 3D metric space against Terrestrial Laser Scanning (TLS) using ground truth data in form of a Structured Light Imaging (SLI) mesh and investigate the influence of the density to the reconstruction's accuracy. We extend the accuracy assessment from 2D to 3D space and perform high resolution investigations on NeRFs by using camera images with 36MP resolution as well as comparison among point clouds of more than 20 million points against a 0.1mm ground truth mesh. TLS achieves the highest geometric accuracy results with a standard deviation of 1.68mm, while $\text{NeRF}_{\delta_{t=300}}$ diverges 18.61mm from the ground truth. All NeRF reconstructions contain 3D points inside the object which have the highest displacements from the ground truth, thus contribute the most to the accuracy results. NeRFs accuracy improves with increasing the density threshold as a consequence of completeness, since beside noise and outliers the object points are also being removed.

## 1. INTRODUCTION

3D geometric and radiometric scene reconstruction is one of the most essential challenges in computer graphics because high geometric accuracy and photo-realism must be satisfied. Recently, due to the rapid development of acquisition instruments and processing methods, point clouds have become a fundamental data type for realistic representation of 3D objects and scenes in applications that require highly accurate geometric information. A point cloud is represented through a collection of non-uniform distributed points where each point has its geometric coordinates (x,y,z), but may also contain other attributes such as density, color, reflectance and surface normal (Liu et al., 2023).

Generating a dense point cloud of a scene based on camera parameters and poses in 3D space is initially addressed by Multi-View Stereo (MVS) algorithms. MVS allows high reconstruction accuracy when the surfaces are textured, thus the main challenge lies in computing precise corresponding points between images of the scene (Stathopoulou et al., 2021).

Another method for fast and highly accurate capturing the geometry of objects is laser scanning. Using reflected light, Terrestrial Laser Scanning (TLS) systems capture dense point clouds of objects and scenes through acquiring the coordinate and intensity value for each 3D point (Kermarrec et al., 2022).

Beside classical photogrammetric approaches, over the past years, deep learning algorithms have become an alternative tool for image processing and spatial analysis (Zhu et al., 2021) leading to a number of improvements in terms of processing time and accuracy. Resolving the challenges of novel view synthesis and representing detailed scene geometry with complex occlusions is introduced by Neural Radiance Fields (NeRFs) (Mildenhall et al., 2021). Using a set of camera poses with associated images as input, the scene is represented through the weights of a fully-connected neural network to a position-dependent density and radiance which additionally depends on the viewing direction at that spatial location. Generating a scene representation in form of a point cloud with densities and colors is gained by sampling rays from the camera poses, resulting in rendering of depth maps.

NeRFs and its variants (Barron et al., 2021, Oechsle et al., 2021, Tao et al., 2023, Chen et al., 2023a, Chen et al., 2023b) have demonstrated impressive performance (Müller et al., 2022b) of learning 3D representations from images. Nonetheless, quantifying the confidence of any spatial representation implies an accuracy assessment analysis where the geometrical similarity is the key factor.

To our knowledge, this is the first attempt to evaluate NeRFs reconstructive geometry in 3D space against TLS using a ground truth data in form of Structured Light Imaging (SLI) mesh. We explore the influence of the density to the reconstruction's accuracy and investigate if NeRFs can compete against classical photogrammetric methods for representing ground truth locations.

Our main contributions can be summarized as follows:

- we extend the accuracy assessment from 2D to 3D space, unlike current pipelines (Martin et al., 2023) that only calculate the quantitative metrics that refer to the image quality (Azzarelli et al., 2023).
- we perform high resolution investigations on NeRFs by using camera images with 36MP resolution as well as comparison among point clouds of more than 20 million points against a ground truth mesh with 0.1mm accuracy.
- we demonstrate the benefits of an accuracy evaluation performed using a ground truth data, which besides providing quantitative results it makes them spatially significant.

* Corresponding author

After briefly summarizing the related work in Section 2, in the following, in Section 3 we describe the processing steps for the reconstruction and extraction of the point clouds from NeRFs and the accuracy assessment method. In Section 4 the data acquisition is described, followed by our experiments and results in Section 5. The discussion is presented in Section 6 and finally Section 7 concludes this contribution.

## 2. RELATED WORK

In this contribution we evaluate NeRFs reconstructive geometry by analysing point clouds, thus we hereafter review the respective literature.

**NeRFs** Differing from the traditional 3D reconstruction with MVS, NeRFs represent scenes as a continuous volumetric field consisting of position-dependent density and radiance which additionally depends on viewing direction. Still, the time-consuming processing and rendering time is a significant drawback which sets limitations to its applicability.

AligNeRF (Jiang et al., 2023) explores the possibilities for learning high resolution 3D geometric representations, while EventNeRF (Rudnev et al., 2023) and Ev-NeRF (Hwang et al., 2023) introduce a method for dense photo realistic RGB view synthesis of a static scene using event cameras.

Triggered by NeRFs necessity for accurate camera parameters, F2-NeRF (Wang et al., 2023) enables arbitrary camera trajectories as input, but still can not recover camera poses from scratch. NeRF– (Wang et al., 2021) optimises camera poses and scene representation, eliminating the need of pre-extracting the camera parameters.

Modelling a volumetric radiance field with a neural point cloud is the core concept in Point-NeRF (Xu et al., 2022) and Points2NeRF (Zimny et al., 2022) which take a 3D point cloud with the associated color values to obtain colored mesh representation, on account of the computational cost required to store NeRFs architecture.

Significantly reducing the processing time without sacrificing rendering quality is established in Instant Neural Graphic Primitives (Instant-NGP) (Müller et al., 2022b), using a small neural network augmented by a multi-resolution hash encoding. The network learns to disambiguate hash collisions and it's implemented on fully-fused CUDA kernels, hence reconstructing scenes in a few seconds. Although Instant-NGP can be a base framework for a few other methods, specific implementation developments for NeRFs have been applied with pose refinement (Lin et al., 2021), which is the used pipeline for our research purposes.

**Point Cloud Accuracy Metrics** Unlike 2D content, point clouds are distributed in 3D space which makes the accuracy assessment more challenging. Thus, a 3D-to-2D projection-based mechanism (Yang et al., 2020) enables a simplified objective comparison between point cloud quality evaluation and conventional image-based measurements. Consequently, with aim to build an open standard for compactly representing 3D point clouds, the Moving Picture Experts Group (MPEG) proposes establishing quality metrics, such as point-to-point (p2point), point-to-plane (p2plane) and point-to-mesh (p2mesh) (Marvie et al., 2023). The p2point metric quantifies the distances between corresponding points to measure the degree of distortion, p2plane projects the obtained p2point distances along the surface normal direction, while p2mesh reconstructs the surface and then measures the distance from a point to the surface, but the efficiency is strongly dependent on the accuracy of the surface reconstruction algorithm (Liu et al., 2023).

For detecting object variations, a comparison among Cloud-to-Cloud (C2C), Cloud-to-Mesh (C2M) and Multiscale Model to Model Cloud Comparison (M3C2) point cloud evaluation methods is presented (Kharroubi et al., 2022). Furthermore, the same methods and TLS as ground truth data are used to analyze the similarity between point clouds for the purposes of cultural heritage (Di Stefano et al., 2021). Both C2C and C2M methods usually consider either the closest point or the neighboring points within a fixed searching radius as ground truth, which may not reflect the actual accuracy. Therefore, an Adaptive Cloud-to-Cloud (AC2C) comparison method (Huang et al., 2022) searches the potential ground truth in theoretical error space of each point, which is estimated according to the position of the corresponding visible cameras and their distances to a target point.

We can conclude that NeRFs are a driving force for vast contributions in the field of view synthesis, exploring new possibilities and expanding the views of spatial reconstruction and 3D modelling. However, the linking point of all this research is that the quantitative accuracy metrics refer to the image quality and not to the volume density itself, from which the geometry can further be evaluated. In that regard, in this contribution we bring the evaluation from image to 3D metric space, by comparing NeRFs and TLS method against a ground truth mesh with aim to assess NeRFs degree of confidence in obtaining reliable information about physical objects.

## 3. METHODOLOGY

As depicted in Figure 1, in Section 3.1 the principles of refined camera pose estimation are introduced. Subsequently, in Section 3.2 the methodology for the reconstruction and extraction of the dense point cloud from NeRFs is described. At last, in Section 3.3 background of the accuracy assessment method is laid out.

### 3.1 Pose estimation

Misaligned camera poses result in cloudy artifacts and a reduction of sharpness and details of the reconstructed scene. Considering NeRFs ability to reliably reconstruct a scene only with a-priori known accurate intrinsic and extrinsic camera parameters, the refined pose estimation is implemented as it leads to better reconstruction (Jäger et al., 2023). The method allows to back propagate loss gradients to the input pose calculations that are used to optimize and refine the poses (Müller et al., 2022a). The pose estimation process is executed independently and only once for creating a set of input poses through Structure from Motion (SfM) (Schonberger and Frahm, 2016). Namely, the algorithm expects as input a set of overlapping images of the scene captured from different camera positions. The camera poses get determined during the reconstruction process along with the 3D geometry that starts with detecting visually distinctive key points in all images. Then the key points are matched over all images based on similarity depicting the same 3D point. These corresponding key points are input to a bundle adjustment that determines the camera poses and intrinsic parameters, as well as the 3D coordinates for all key points depicted in multiple images, resulting in a sparse point cloud.

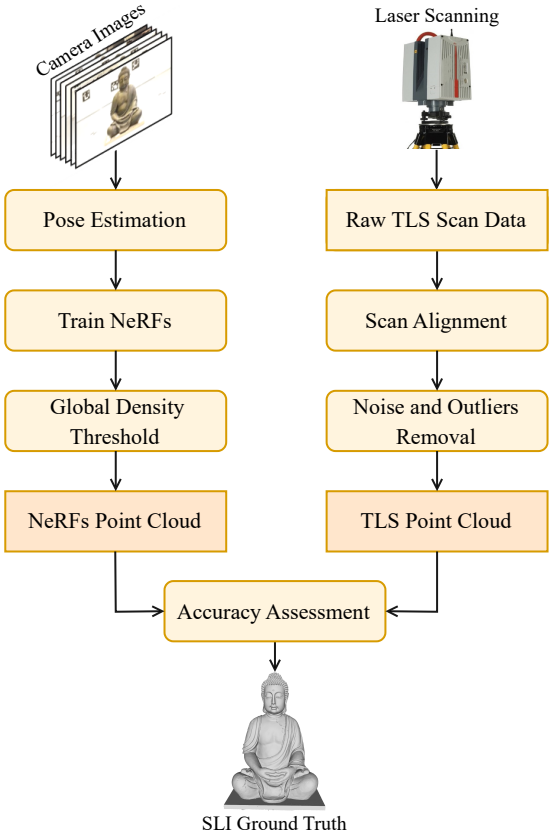



Figure 1. Illustration of the processing steps. The images with estimated camera poses are input for training NeRFs (Section 3.1). We filter NeRFs volume density using a global density threshold (Section 3.2). TLS scans are aligned and cleaned of noise and outliers to complete the whole object. All point clouds are registered in the same metric space as the ground truth mesh for evaluation (Section 3.3).

### 3.2 3D Reconstruction by NeRFs

In order to obtain a scene representation in form of a dense point cloud with density and color, NeRFs sample rays from the camera poses and capture the 3D coordinates resulting in rendering of depth maps (Müller et al., 2022b). Although rendering new views of a scene is the primary task of NeRFs, for computer vision and photogrammetric purposes reliably representing the geometry in 3D space is a more valuable factor. Since the position-dependent density is a differential opacity of accumulated radiance by a ray passing through, positions with higher density values indicate a higher probability to be an object point (Jäger et al., 2023). To determine the influence of the density to the point cloud accuracy, we filter the volume density using a global density threshold $\delta_t$.

### 3.3 Accuracy assessment

To estimate the geometric accuracy of NeRFs in 3D space, we use point cloud accuracy metrics based on the calculation of the distance between two points which establish displacements based on proximity in Euclidean space. All point clouds are given spatial significance and aligned in the same metric space as the ground truth mesh for evaluation using Iterative Closest Point (ICP) (Besl and McKay, 1992) which finds an optimal rigid transformation to align two point sets.

C2M computes the displacements between each point in the compared point cloud and the nearest facet of the reference mesh by using the nearest Euclidean distance. For each point of the compared cloud the nearest triangle in the reference mesh is searched (Kharroubi et al., 2022). As meshes generally provide a side information by looking at the normal of the triangle, Cloud-to-Mesh distances can be signed.

If the scalar field corresponds to the calculated displacements and the distribution of the measurement noise is known, then the points for which the local scalar values seem to fit the noise distribution can be filtered. Measuring how dispersed the data is in relation to the ground truth mesh is provided through: Mean Error (Mean_E), Standard Deviation (SD), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) (Remondino et al., 2023) accordingly:

$$Mean\_E = \frac{\sum_{i=1}^{n}(x_i)}{n} \tag{1}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \tag{2}$$

$$MAE = \frac{\sum_{i=1}^{n}|x_i - \hat{x}|}{n} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x})^2}{n}} \tag{4}$$

where n is the number of points of the compared point cloud, $x_i$ stands for the closest distance between each point in the compared point cloud and the nearest facet of the reference mesh, while $\overline{x}$ denotes the mean value of the distances.

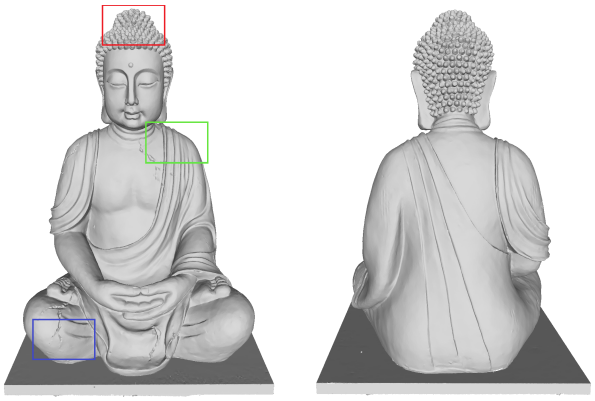
## 4. DATA

NeRFs are trained by using images with estimated camera poses as input to predict the density and radiance for each position. Moreover, evaluating NeRFs reconstructive geometry by analysing the volume density requires a reference comparable standard. Consequently, our investigations are based on real data of an indoor scene of 0.7m tall Buddha statue (further on named as object) placed on a 0.48m x 0.38m x 0.02m rectangular plate.

(a) (b)

Figure 2. Data acquisition setup: (a) Recording 3D point cloud with TLS (Leica HDS6000). Twelve scans from different positions are recorded for full object coverage. (b) Capturing images with a professional digital camera (Nikon D810). The camera is mounted on a tripod to prevent vibrations during acquisition.

**SLI** For generating the ground truth data, Structured Light Imaging is used because it can capture more accurate point clouds than typical laser scanning methods especially in close range. Strictly speaking, it captures the geometry of the object by illuminating the surface using structured light projection. The light source projects a coded pattern of parallel light stripes onto the object, the cameras capture these patterns from a known position resulting in a specific sequence of gray values for each pixel of an image, from which the range can be calculated. For this purpose we utilize the stereoSCAN3D-HE scanning device [1] with two digital cameras and a projector in between. The object is placed on a turntable and the rotations are controlled automatically by a workstation. Any remaining holes are closed (Püschel, 2011) resulting in a smoothed mesh with 0.1mm accuracy (Figure 3).



(a) (b)

Figure 3. Ground truth data captured by Structured Light Imaging (SLI). Visualization with 3D mesh representation, (a) front and (b) back view. The colored rectangles are enlarged in Figure 5.

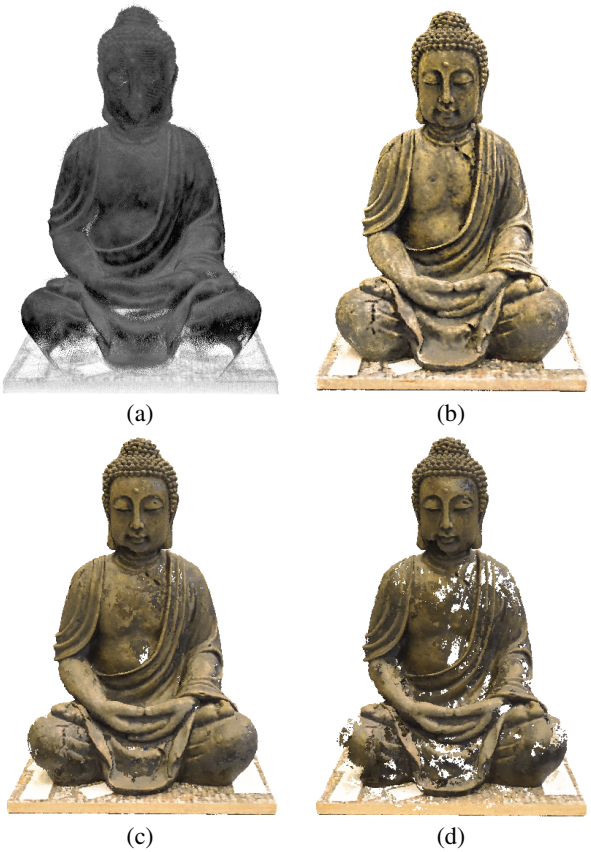
**TLS** Laser scanning systems provide fast and highly accurate generation of millions of 3D points and can capture the depth of complex scenes. For the purpose of high resolution comparison we use Leica HDS6000 (Figure 2) with scan rate of up to 500.000 points/sec. and 0.7mm RMS (Voegtle and Wakaluk, 2009). Recognizing that regions of the scanned scene occluded from one scan position are almost always visible from another, twelve scans from 1m distance from the object and 0.3m, 1.4m and 1.6m height respectively have been recorded. All scans are aligned to complete the whole object. Usually more than one surface entity is observed in a single measurement because the physical behaviour of the sensor creates spatial noise in the data which prevent the objects to be accurately modeled. The noise and outliers are removed leading to the final point cloud of 21 million points.

**Image capturing** The image dataset to train NeRFs is captured using 24 bit depth Nikon D810 with a camera sensor of 36MP and image resolution of 7360x4912 pixels. Due to its high resolution the camera is mounted on a tripod (Figure 2) to prevent shivering and vibrations during acquisition. The images are captured in a hemispheric trajectory around the object from three different camera heights of 0.7m, 1.1m and 1.5m accordingly, to achieve full coverage of the object. Since eliminating mislabelled camera data and blurry frames is a prerequisite for a reliable reconstruction, all inconsistent quality images are removed and NeRFs are trained on the remaining 125 images.

Hence reducing the reconstruction time within a few seconds without sacrificing the rendering quality, we use the Instant-NGP NeRF implementation. Moreover, we have integrated a PLY-writer in order to extract the point clouds. All results are computed on Intel i9 10850K CPU, 32GB RAM as well as a Nvidia Geforce RTX 3090 GPU, which are needed for training NeRFs efficiently.

## 5. EXPERIMENTS AND RESULTS

The results are categorized into qualitative (Section 5.1), where the geometric reconstructions as point clouds are visualized and quantitative (Section 5.2).



(a) (b) (c) (d)

Figure 4. Comparison of the various point cloud reconstructions. (a) TLS contains noise on the shoulders and top of the head and gaps below the knees, (b) NeRF$_{\delta_{t=30}}$ shows strong noise, but the object is completely reconstructed, (c) NeRF$_{\delta_{t=150}}$ eliminates the noise but also the object points and (d) NeRF$_{\delta_{t=300}}$ achieves the sharpest result on account of completeness.

### 5.1 Qualitative results

**TLS** The TLS point cloud, although filtered from significantly different points, still contains noise on the shoulders and the top of the head. In addition, it has gaps below the knees (Figure 4 and 5). When it comes to accuracy, TLS achieves the best results which correspond to the displacement values in Table 1. The biggest errors occur due to noise, otherwise the object shows high correspondence with the ground truth mesh (Figure 6).

[1] https://used.exactmetrology.com/used-equipment/breuckmann-stereoscan-5-0mp-3d-scanner/ (last access 12/09/2023)

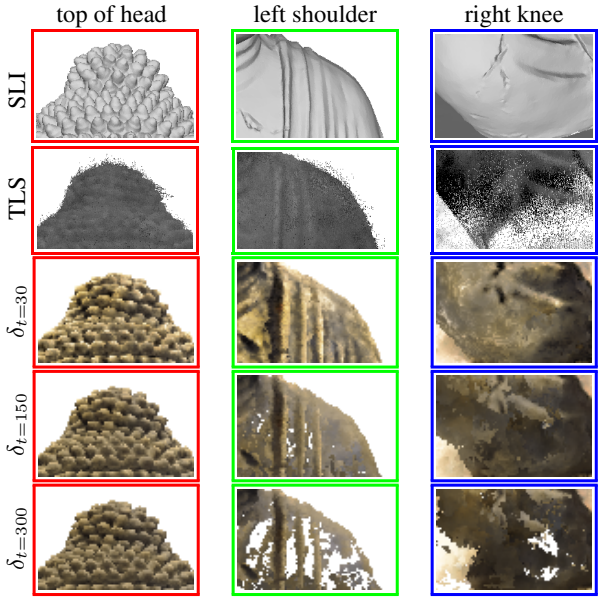



Figure 5. Here we visualize different types of surfaces against the ground truth (Figure 3). TLS has noise on the top of the head and shoulders and gaps under the knee. NeRFs show good reconstruction of complex geometry (top of head). Increasing the density threshold, besides noise, removes also object points leading to incomplete reconstruction (shoulder and knee).

**NeRFs** For NeRFs gaps and noise can be observed in the reconstructions. Increasing the density threshold removes noise points, on account of object's completeness (Figure 4). With $\delta_{t=30}$, a large amount of noise points can be identified on the object, however the geometry is well reconstructed. Increasing the density threshold to $\delta_{t=150}$ eliminates the noise but also the object points. $\delta_{t=300}$ achieves the sharpest result, nevertheless resulting in a least complete reconstruction (Figure 5). All NeRF reconstructions have points inside the object and under the plate (Figure 6) since points are sampled along camera rays to render the volume density field, unlike TLS and SLI which only capture the object's surface. Those points have the biggest error values and contribute the most to the accuracy analysis.

### 5.2 Quantitative results

**TLS** The 21 million TLS point cloud shows the highest geometric accuracy (Table 1). The MAE, which is an indicator of the variance, is just -0.22mm, while the deviation from the ground truth is in range of 1.68mm, mostly due to noise and outliers.

Table 1. Numerical representation of the C2M comparison results (in mm). TLS achieves the most reliable accuracy results. NeRFs accuracy improves with increasing the density threshold due to noise and outlier points removal.

| | 3D points (million) | Mean_E (mm) | SD (mm) | MAE (mm) | RMSE (mm) |
|---|---|---|---|---|---|
| TLS | 21.0 | -0.22 | 1.68 | 1.29 | 1.69 |
| NeRF$_{\delta_{t=30}}$ | 19.0 | -25.26 | 27.92 | 28.41 | 37.66 |
| NeRF$_{\delta_{t=150}}$ | 9.1 | -20.58 | 22.24 | 22.48 | 30.31 |
| NeRF$_{\delta_{t=300}}$ | 5.4 | -18.25 | 18.61 | 19.58 | 26.06 |

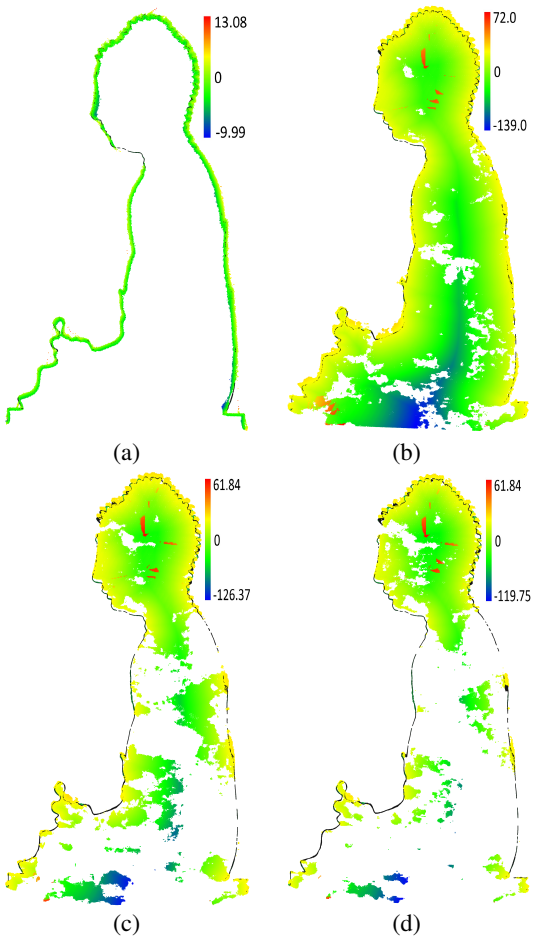



Figure 6. C2M comparison by visualization of the reconstructed point clouds against the ground truth accordingly: (a) TLS, (b) NeRF$_{\delta_{t=30}}$, (c) NeRF$_{\delta_{t=150}}$ and (d) NeRF$_{\delta_{t=300}}$. The ground truth mesh is presented as a black thin line and the colors correspond to the error displacements (in mm) in a longitudinal section of the object.

**NeRFs** All NeRF point clouds demonstrate lower accuracy compared to TLS. The number of points obviously decreases with increasing the density threshold. NeRF$_{\delta_{t=30}}$ has 19 million points, thus the lowest accuracy because it contains large amount of noise points, as it diverges from the ground truth 28.41mm. Subsequently, with increasing the density threshold to NeRF$_{\delta_{t=150}}$ the number of points drops significantly to 9.1 million, however the accuracy increases as it deviates 22.24mm from the ground truth. With MAE lower than 20mm and SD of 18.61mm, NeRF$_{\delta_{t=300}}$ achieves the highest accuracy, but it has just 5.4 million points, insufficient to completely reconstruct the object.

## 6. DISCUSSION

Since we evaluate NeRFs geometric accuracy against TLS using a ground truth mesh by analysing the density in a 3D metric space, in this section we discuss the generated results.

**TLS** Despite recording multiple scans from different position and height to achieve full object coverage, the TLS point cloud still contains gaps below the knees (Figure 4). Moreover, the details of the facial features (nose and mouth) are not reliably

represented and the neck is failed to be reconstructed (Figure 6).

With a view to eliminate noise, a statistical analysis on point's neighborhood is performed and those points which do not meet a certain criteria are removed. For each point the mean distance to all its neighbors is computed. Considering the mean and standard deviation, all points whose mean distances are outside an interval are considered as outliers and trimmed from the point cloud. Nevertheless, the point cloud still contains outliers on the shoulders and on the top of the head (Figure 5) which have the highest error values. In spite of that, TLS achieves the highest accuracy results compared to SLI with divergence of just a few millimeters from the ground truth (Table 1).

**NeRFs** As demonstrated, increasing the density threshold results in increasing the accuracy as a consequence of completeness, since beside noise and outliers the object points are also being removed. It should also be noted that higher density threshold leads to significantly dropping the number of points, which explains the gaps and incomplete object reconstruction (Figure 4). In addition, when using different density thresholds, color differences within the NeRF reconstructions can be observed. With $\delta_{t=30}$ the object is brighter (Figure 5) because it contains a large amount of noise, which can indicate that the ray does not terminate within the object leading to inaccurate geometric representation (Haitz et al., 2023). However, the color differences don't affect the geometric reconstruction.

Opposite of TLS and SLI which only capture the object's outer surface, NeRFs use deep learning to infer a continuous volumetric representation of a scene from a set of 2D images by sampling points through the rays. For that reason, all NeRF reconstructions contain points inside the object (Figure 6) with highest divergence from ground truth, thus they have the highest influence on the accuracy results.

Considering that density represents a position-dependent parameter, positions with higher density values indicate a higher probability to be an object point. Increasing the density threshold doesn't result in a proportional removal of the points, since the points on the head are not being removed almost at all. It is likely that those points have higher density values, since NeRFs are optimized to capture complex geometry with high accuracy and detail. We can also notice that NeRFs mean and standard deviation values keep the same trend, the values decrease with increasing the global density threshold, due to the flexibility of the volume density which is unable to sufficiently constrain the 3D geometry (Oechsle et al., 2021).

## 7. CONCLUSION

In this contribution, we present the geometric accuracy of NeRFs against TLS regarding a ground truth SLI mesh. We evaluate the reconstructive geometry of NeRFs by analysing the volume density in a 3D metric space and explore the influence of the density to the reconstruction's accuracy. For that we use C2M method based on displacements between each point in the compared point cloud and the nearest facet in the reference mesh using Euclidean distance.

NeRFs accuracy improves with increasing the density threshold, on account of completeness since beside noise and outliers object points are also being removed. Due to that, NeRFs show lower accuracy results than TLS. Therefore, on top of their usefulness in applications where novel views and realism must be satisfied, in future work the focus should be on increasing NeRFs geometric accuracy for 3D reconstruction, by localizing the density just on object's outer surface and filtering the noise without removing the object points.

## REFERENCES

Azzarelli, A., Anantrasirichai, N., Bull, D. R., 2023. Towards a Robust Framework for NeRF Evaluation. *arXiv e-prints*, arXiv:2305.18079.

Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P. P., 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5855–5864.

Besl, P. J., McKay, N. D., 1992. Method for registration of 3-d shapes. *Sensor fusion IV: control paradigms and data structures*, 1611, Spie, 586–606.

Chen, Y., Chen, X., Wang, X., Zhang, Q., Guo, Y., Shan, Y., Wang, F., 2023a. Local-to-global registration for bundle-adjusting neural radiance fields. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8264–8273.

Chen, Z., Funkhouser, T., Hedman, P., Tagliasacchi, A., 2023b. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16569–16578.

Di Stefano, F., Torresani, A., Farella, E. M., Pierdicca, R., Menna, F., Remondino, F., 2021. 3D surveying of underground built heritage: Opportunities and challenges of mobile technologies. *Sustainability*, 13(23), 13289.

Haitz, D., Jutzi, B., Ulrich, M., Jäger, M., Hübner, P., 2023. Combining Hololens with Instant-NERFs: Advanced Real-Time 3D Mobile Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W1-2023, 167–174.

Huang, H., Ye, Z., Zhang, C., Yue, Y., Cui, C., Hammad, A., 2022. Adaptive Cloud-to-Cloud (AC2C) Comparison Method for Photogrammetric Point Cloud Error Estimation Considering Theoretical Error Space. *Remote Sensing*, 14(17), 4289.

Hwang, I., Kim, J., Kim, Y. M., 2023. Ev-nerf: Event based neural radiance field. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 837–847.

Jäger, M., Hübner, P., Haitz, D., Jutzi, B., 2023. A Comparative Neural Radiance Field (NERF) 3D Analysis Of Camera Poses From Hololens Trajectories And Structure From Motion. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W1-2023, 207–213.

Jiang, Y., Hedman, P., Mildenhall, B., Xu, D., Barron, J. T., Wang, Z., Xue, T., 2023. Alignerf: High-fidelity neural radiance fields via alignment-aware training. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 46–55.

Kermarrec, G., Yang, Z., Czerwonka-Schröder, D., 2022. Classification of terrestrial laser scanner point clouds: A comparison of methods for landslide monitoring from mathematical surface approximation. *Remote Sensing*, 14(20), 5099.

Kharroubi, A., Poux, F., Ballouch, Z., Hajji, R., Billen, R., 2022. Three Dimensional Change Detection Using Point Clouds: A Review. *Geomatics*, 2(4), 457–485.

Lin, C.-H., Ma, W.-C., Torralba, A., Lucey, S., 2021. Barf: Bundle-adjusting neural radiance fields. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5741–5751.

Liu, Y., Yang, Q., Xu, Y., Yang, L., 2023. Point cloud quality assessment: Dataset construction and learning-based no-reference metric. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2s), 1–26.

Martin, P., Rodrigues, A., Ascenso, J., Queluz, M. P., 2023. NeRF-QA: Neural Radiance Fields Quality Assessment Database. *arXiv preprint arXiv:2305.03176*.

Marvie, J.-E., Nehmé, Y., Graziosi, D., Lavoué, G., 2023. Crafting the MPEG metrics for objective and perceptual quality assessment of volumetric videos. *Quality and User Experience*, 8(1), 4.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99–106.

Müller, T., Evans, A., Schied, C., Foco, M., Bódis-Szomorú, A., Deutsch, I., Shelley, M., Keller, A., 2022a. Instant neural radiance fields. 1–2.

Müller, T., Evans, A., Schied, C., Keller, A., 2022b. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4), 1–15.

Oechsle, M., Peng, S., Geiger, A., 2021. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5589–5599.

Püschel, J., 2011. Vergleich eines 3d-Modells zwischen Bundler und Breuckmann. *Bachelor Thesis*, Institute for Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology - KIT.

Remondino, F., Karami, A., Yan, Z., Mazzacca, G., Rigon, S., Qin, R., 2023. A Critical Analysis of NeRF-Based 3D Reconstruction. *Remote Sensing*, 15(14). https://www.mdpi.com/2072-4292/15/14/3585.

Rudnev, V., Elgharib, M., Theobalt, C., Golyanik, V., 2023. Eventnerf: Neural radiance fields from a single colour event camera. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4992–5002.

Schonberger, J. L., Frahm, J.-M., 2016. Structure-from-motion revisited. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Stathopoulou, E. K., Rigon, S., Battisti, R., Remondino, F., 2021. Enhancing Geometric Edge Details in MVS Reconstruction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021, 391–398.

Tao, T., Gao, L., Wang, G., Chen, P., Hao, D., Liang, X., Salzmann, M., Yu, K., 2023. LiDAR-NeRF: Novel LiDAR View Synthesis via Neural Radiance Fields. *arXiv preprint arXiv:2304.10406*.

Voegtle, T., Wakaluk, S., 2009. Effects on the measurements of the terrestrial laser scanner HDS 6000 (Leica) caused by different object materials. *Proceedings of ISPRS work*, 38(2009), 68–74.

Wang, P., Liu, Y., Chen, Z., Liu, L., Liu, Z., Komura, T., Theobalt, C., Wang, W., 2023. F2-nerf: Fast neural radiance field training with free camera trajectories. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4150–4159.

Wang, Z., Wu, S., Xie, W., Chen, M., Prisacariu, V. A., 2021. NeRF–: Neural Radiance Fields Without Known Camera Parameters. *arXiv e-prints*, arXiv:2102.07064.

Xu, Q., Xu, Z., Philip, J., Bi, S., Shu, Z., Sunkavalli, K., Neumann, U., 2022. Point-nerf: Point-based neural radiance fields. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5438–5448.

Yang, Q., Chen, H., Ma, Z., Xu, Y., Tang, R., Sun, J., 2020. Predicting the perceptual quality of point cloud: A 3d-to-2d projection-based exploration. *IEEE Transactions on Multimedia*, 23, 3877–3891.

Zhu, Q., Min, C., Wei, Z., Chen, Y., Wang, G., 2021. Deep learning for multi-view stereo via plane sweep: A survey. *arXiv preprint arXiv:2106.15328*.

Zimny, D., Trzciński, T., Spurek, P., 2022. Points2NeRF: Generating Neural Radiance Fields from 3D point cloud. *arXiv e-prints*, arXiv:2206.01290.