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

NERFBK:
A HOLISTIC DATASET FOR BENCHMARKING NERF-BASED 3D RECONSTRUCTION

Z. Yan [1,2], G. Mazzacca [1,3], S. Rigon [1], E.M. Farella [1], P. Trybala [1], F. Remondino [1]

[1] 3D Optical Metrology Unit, Bruno Kessler Foundation (FBK), Trento, Italy
Email: (zyan, gmazzacca, srigon, elifarella, ptrybala, remondino)@fbk.eu

[2] DISI, University of Trento, Italy – Email: ziyang.yan@unitn.it

[3] Dept. Mathematics, Computer Science and Physics, University of Udine, Italy - 165480@spes.uniud.it

KEY WORDS: NeRFBK, Photogrammetry, Neural Radiance Fields (NeRF), 3D reconstruction, Benchmark.

ABSTRACT
Neural Radiance Field methods are innovative solutions to derive 3D data from a set of oriented images. This paper introduces new real and synthetic image datasets - called NeRFBK - specifically designed for testing and comparing NeRF-based 3D reconstruction algorithms. More and more reconstruction algorithms and techniques are available nowadays, raising the need to evaluate and compare the quality of derived 3D products currently used in various domains and applications. However, gathering diverse data with precise ground truth is challenging and may not encompass all relevant applications. The NeRFBK dataset addresses this issue by providing multi-scale, indoor and outdoor datasets with high-resolution images and videos and camera parameters for testing and comparing NeRF-based algorithms. This paper presents the design and creation of the NeRFBK set of data, various examples and application scenarios, and highlights its potential for advancing the field of 3D reconstruction.

1. INTRODUCTION

Generating high-quality 3D models is a central objective of many research and studies in computer vision and photogrammetry. Nowadays, 3D reconstructions are commonly produced in several sectors and for various applications (such as quality controls, 3D monitoring, 3D inspections, robotics, virtual and augmented reality, and / or medical imaging. Over the past decades, improvements in image-based 3D reconstruction algorithms have made it essential to evaluate and compare their performance. In particular, NeRF-based algorithms are increasingly attracting the attention of the research community, interested to explore their potential in 3D reconstruction.
A benchmark dataset is a set of data used by scientists to evaluate and compare the performance of sensors, platforms, or processing algorithms against a reliable and accurate ground truth (GT). However, obtaining enough diverse data poses a challenge due to the associated costs, time, and the need for precise annotations and accurate GT. The dataset should comprise data with various characteristics, such as different scale or environments, and should be accompanied by reliable and accurate annotations and GT data.
Since Mildenhall et al. (2021), vanilla NeRF was introduced, a novel 3D reconstruction approach based on view synthesis which can represent a scene through a continuous volumetric function and be parameterized by multilayer perceptions to generate the volume density and directional emitted radiance at each point (Hedman et al., 2021; Verbin et al., 2022; Xu et al., 2022). This sparked a major revolution in the 3D reconstruction field. Different from many other 3D neural representations (Vijayanarasimhan et al., 2017; Ibrahimli et al., 2023), NeRF models are self-supervised and can learn a scene starting from a set of multi-view images and poses, without the requirement of 3D/depth supervision. NeRF methods have found large potential applications in many fields, including robotics, autonomous navigation, virtual reality/augmented reality, industrial inspection, etc. (Gao et al., 2022). According to some studies (Yen-Chen et al., 2022; Mazzacca et al., 2023; Remondino et al., 2023; Jäger et al., 2023), NeRF-based methods are considered to have great potential and application prospects as they can potentially achieve better results with respect to traditional image-based 3D reconstruction techniques such as Multi-View stereo (Schönberger et al., 2016; Wang et al., 2021a; Wang et al., 2021b; Stathopoulou and Remondino, 2023) and RGB-D-based methods (Truong et al., 2020a; Truong et al., 2020b; Liu et al., 2021) when dealing with texture-less, metallic, highly reflective, transparent objects due to the view-dependent nature of the NeRF model. To accurately develop, evaluate, and compare NeRF-based techniques, it is essential to have access to high-quality data that includes precise ground truth. However, gathering synthetic and real-world data with different characteristics (such as featureless, well-textured, refractive, and reflective surfaces, in various sizes and shapes) can be difficult and may not encompass all relevant domains and applications. To overcome this issue, a new real and synthetic collection of datasets called NeRFBK is presented for testing and comparing NeRF-based algorithms. Our benchmark consists of multi-scale, indoor, and outdoor datasets, high-resolution images and videos, camera parameters (interior and exterior parameters), and ground truth. The dataset includes both real and synthetic images featuring different surfaces, size, and linked to three main domains (industry, cultural heritage, and geospatial sector). The NeRFBK datasets enables addressing different challenges in 3D reconstruction and advancing the field. The paper reports the usefulness of the NeRFBK datasets by presenting the results of experimental tests conducted with the shared data.

2. BENCHMARKING NERF

2.1 Repositories

In recent years, NeRF has emerged as a powerful technique for high-quality 3D reconstruction from 2D images. As the field is growing, the need for benchmarks to evaluate and compare their different performance is crucial. Some datasets for testing or evaluating NeRF-based methods are already available and hereafter reported.
Tanks and Temples[1] (Knapitsch et al., 2017): it includes various indoor and outdoor scenes of varying size and complexity captured under realistic conditions using high-resolution video sequences, as well as ground truth camera poses. The GT data of each dataset is derived from terrestrial laser scanner acquisitions

[1] https://www.tanksandtemples.org/

and COLMAP (Schönberger and Frahm, 2016) camera poses estimation. From the video sequences, high-frame rate datasets can be extracted and used for NeRF processing.

Scannet[2] (Dai et al., 2017): Containing 2.5Mil RGB-D images from 1513 scans acquired in different space, this dataset covers various kinds of indoor scenes. In addition to the calibration parameters and camera poses, it also provides instance-level object category labels for 3D object classification and segmentation, and CAD models for matching the objects in different scans.

BlendedMVS[3] (Yao et al., 2019): it is a large-scale synthetic dataset for multi-view stereo training, consisting of about 17000 rendered images with a maximum resolution of 2048 x 1536 pixels. It contains 113 scenes, each with 20-1000 blended images and includes GT camera poses, depth maps, and 3D surface models. It also includes BlendedMVG, a superset multi-purpose large-scale dataset for solving multi-view and geometry-related problems.

NeRF[4] (Mildenhall et al., 2021): it contains three parts: Diffuse Synthetic 360◦ with simple geometry, Realistic Synthetic 360◦ with complicated geometry and realistic non-Lambertian materials, and real images of complex scenes captured with a smartphone. The synthetic renderings are captured at 512x512 or 800x800 pixels from viewpoints sampled on the upper hemisphere or full sphere, while the real images are captured at 1008x756 pixels. GT is not available for these datasets.

Shiny dataset[5] (Wizadwongsa et al., 2021): it is comprised of 8 scenes captured with a smartphone that exhibit a range of complex view-dependent effects, including reflections, refractions, and specular highlights on metallic and ceramic materials, as well as detailed thin structures. All images in the dataset have a resolution of 1008 x 756 pixels. No GT is available.

UrbanScene3D [6] (Lin et al., 2022): it is designed for research of urban scene perception and reconstruction. It has 128K high-resolution images generated by 10 synthetic scenes and 6 real urban scene using drone. The images and GT of synthetic scenes are produced by CAD while for the real scenes, no ground truth for the whole scene is available but the ground-truth meshes of some buildings in the scenes which are generated by Trimble X7 LiDAR scanners loaded with GPS localization devices are provided. The dataset also provides manually annotated instance labels for each building that can be used for segmentation.

Mip-NeRF 360[7] (Barron et al., 2022): it comprises 9 scenes, encompassing 5 outdoor and 4 indoor settings, each exhibiting a complex central object or area along with a detailed background. The dataset was acquired using two mirrorless digital cameras, with the initial camera position serving as a reference view. The images are obtainable at variable resolutions ranging from 1 Mpx to 15 Mpx. In order to mitigate color harmonization problems, the outdoor scenes were captured when the sky was overcast, while the indoor scenes employed large diffuse light sources to avert casting shadows. The dataset lacks any GT.

X-NeRF dataset[8] (Poggi et al., 2022): it includes 16 forward-facing scenes captured by different sensors, one high-resolution (12.4Mpx) RGB camera and two low-resolution (1Mpx) IR and MS cameras. For each camera, it took approximately 30 viewpoints, so that around 90 views are obtained per scene. IR and MS images encoded with colormaps and RGB images with corresponding camera poses estimated by COLMAP have been released to the public.

Mill 19[9] (Turki et al., 2022): it contains two large-scale scenes recorded by drone around CMU. The scene one is an industrial building with a 500x250 m^2 square and another scene is selected nearby a construction area where is full of debris. 1940 and 1768 images with 4608x3456 px resolution are captured from scene one and scene two respectively, and the camera poses are refined by PixSFM (Lindenberger et al., 2021). No 3D GT is available.

Block-NeRF[10] (Tancik et al., 2022): it consists of 13.4 hours driving record from 1330 different data collection runs on different public roads in San Francisco. The videos are captured by 12 cameras (8 cameras provide a complete surround view on the top of the car, 4 additional cameras fixed at the front of vehicles pointing forward and sideways) mounted on data collection vehicles, and totally more than 2.8M images and corresponding camera poses are generated. This dataset is suitable for the research about city-scale reconstruction and autonomous driving.

OMMO[11] (Lu et al., 2023): it is a large-scale outdoor multi-model dataset consisting of 33 scenes with 14000 calibrated images. The highest resolution of the images can be up to 4K and all of them are generated by the videos released in YouTube or captured from drones. Camera poses generated by COLMAP and prompt annotations for multi-model NeRF labelled by manual and CLIP (Radford et al., 2021) are provided, but GT is not available.

Relighting NeRF[12] (Toschi et al., 2023): it is a public benchmark created for view synthesis and relighting method. Images are captured by two robotic arms (one controls the camera and the other controls the light source) about 20 scenes with different challenging objects. The data acquisition is designed based on *one-light-at-time* (OLAT) (Zhang et al., 2021) illumination, which used 50 camera viewpoints and took 2000 images per scene under 40 OLAT light conditions. This dataset provides a totally of 40000 images with the pose of both camera and light source, but lacks GT data for 3D evaluations.

All (and other) described datasets feature different characteristics but most of them lack 3D GT data for evaluating the 3D results of NeRF methods. The proposed NeRFBK set of data tries to fill this gap (Section 3).

2.2 Metrics

Besides data for benchmarking algorithms, the use of standard metrics for evaluating methods plays an important role. Common metrics (Mousavi et al., 2018; Mohammadi et al., 2021) used in the 3D reconstructions field include:

$$STD = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(P_j - \underline{P}\right)^2} \quad (1)$$

$$Mean\_E = \frac{\left(P_1 + P_2 + \cdots P_j\right)}{N} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N}\left(P_j\right)^2}{N}} \quad (3)$$

$$MAE = \frac{\sum_{j=1}^{N}\left|P_j\right|}{N} \quad (4)$$

[2] http://www.scan-net.org/
[3] https://github.com/YoYo000/BlendedMVS
[4] https://paperswithcode.com/dataset/nerf
[5] https://nex-mpi.github.io/
[6] https://vcc.tech/UrbanScene3D
[7] https://jonbarron.info/mipnerf360/
[8] https://amsacta.unibo.it/id/eprint/7142/
[9] https://meganerf.cmusatyalab.org/
[10] https://waymo.com/research/block-nerf/
[11] https://ommo.luchongshan.com/
[12] https://eyecan-ai.github.io/rene/

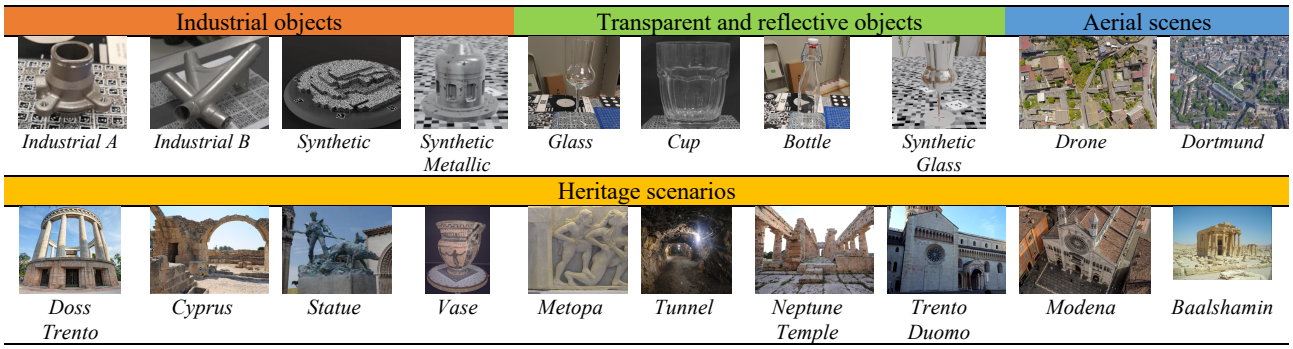


Table 1: Visual summary of the NeRFBK datasets, data and more info available at https://github.com/3DOM-FBK/NeRFBK

where N is the number of observed vertices, P_j denotes the closest distance of each vertex to the corresponding reference vertex, and P denotes the average observed distance.

Accuracy and completeness, sometimes also called precision and recall (Nocerino et al., 2020; Remondino et al., 2023) are also used. The accuracy can measure the percentage of overlap between the reconstructed results and GT, while completeness reflects how much percentage of points in GT have been reconstructed. A threshold distance is used and decided based on the data density and noise levels to calculate the fraction or percentage of points fall within the threshold. The calculation of accuracy and completeness is shown as follows:

$$ACCURACY = \frac{\sum_{i=1}^{S}(DisT_i < Th)}{S} \quad (5)$$

$$COMPLETENESS = \frac{\sum_{i=1}^{T}(DisS_i < Th)}{T} \quad (6)$$

where DisT denotes the point distance between source mesh to corresponding point in ground truth and DisS represent the opposite. S and T are the total number of points in source mesh and ground truth respectively, while Th is the threshold distance, used to filter the points outside the range we set.

3. THE PROPOSED NERFBK

The NeRFBK datasets are composed by real and synthetic scenes. In the real-world datasets, high-resolution images are captured in various scenarios, under different lighting conditions, cameras, scales and Ground Sample Distance (GSD), as well as 3D GT data. Images and GT for synthetic scenes are created in Blender, by modelling different-shape and different-size objects and defining various camera paths for the simulated image acquisition. As shown in Table 1, NeRFBK consists of different types of datasets, including:

- *Industrial*: two metallic real objects and two synthetic datasets are available. These objects have complex geometry, poor texture and reflective surfaces, causing challenges for 3D reconstruction using traditional and learning-based methods.
- *Transparent and reflective*: three real and one synthetic textureless objects are available. The main processing challenge is related to image matching. Their appearance depends on the object's shape, surrounding background and lighting conditions. Refraction and specular reflections, with light travelling through the surface, are normally present.
- *Heritage*: ten different scenarios are available, including the temple of Baalshamin in Syria, which was tragically destroyed in 2015, whose images were collected from the REKREI online repository (Vincent et al., 2016).
- *Aerial*: two datasets are available, one from a UAV flight and another acquired with an aerial oblique camera over the city of Dortmund (Nex et al., 2015).

Each dataset consists of sets of images, original video (for some scenes), point clouds or mesh models as ground truth.

In Section 4, some of these datasets are leveraged for testing some NeRF-based algorithms, reporting performances and quantitative results.

4. EXPERIEMENTS AND RESULTS

This section presents experimental findings to evaluate and compare the performance of different NeRF-based techniques. Experiments are based on SDFStudio (Yu et al., 2022a) and Nerfstudio (Tancik et al., 2023), two comprehensive collections of NeRF methods, including Instant-NGP (Müller et al., 2022), Nerfacto (Tancik et al., 2023), MonoSDF (Yu et al., 2022b), Tensorf (Xu et al., 2022), VolSDF (Yariv et al., 2021), Neus (Wang et al., 2021c), Unisurf (Oechele et al., 2021), and some variants such as Neus-Facto and Mono-Neus. Experiments are performed using some of the datasets available in the NeRFBK repository using an NVIDIA GeForce A40 GPU.

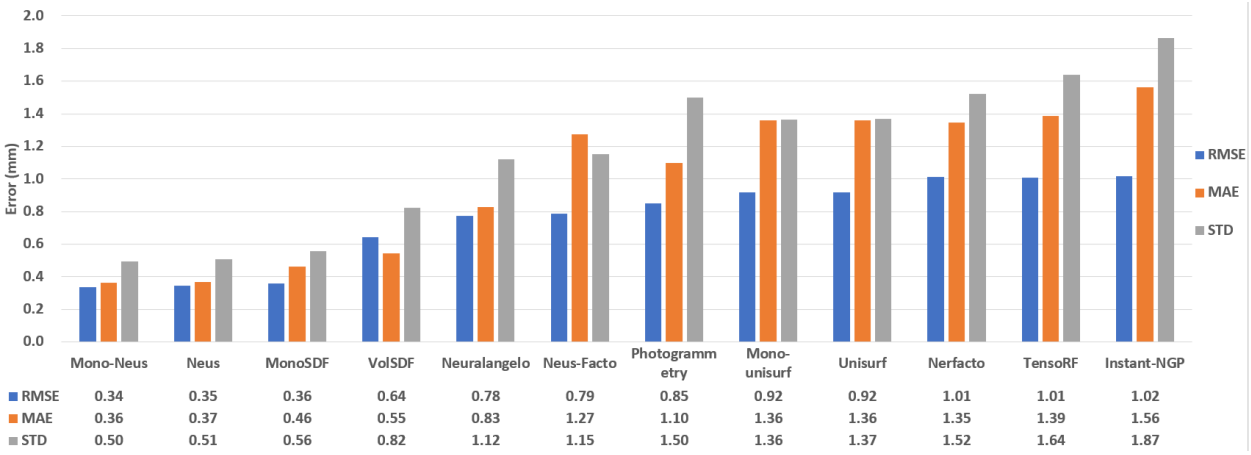
4.1 Industrial objects

The 3D reconstruction of textureless shiny metallic surfaces is problematic for many active and passive sensors (Karami et al., 2021). NeRF could offer a valuable alternative for this task. We evaluated the performance of various NeRF-based methods using the images available for the Industrial B object (Table 1): it consists of 220 sequential images extracted from a smartphone video with a resolution of 1920x1080 pixels. The comparison results (Figure 1) show that Mono-Neus achieved the best results, with Neus as second among all methods. Their RMSE were only 0.34 mm and 0.35 mm, respectively. Accuracy and completeness with respect to the ground truth (GT) data were also computed for all methods (Figure 2).

As shown in Figure 2, Mono-Neus outperforms the other methods. Instant-NGP rank low in terms of completeness as well as accuracy. This is because the generated mesh using Instant-NGP is quite noisy compared to Mono-Neus and can sparsely cover the GT. It is interesting to notice that the estimated accuracy for Nerfacto is similar to TensoRF, but the completeness is much worse.

4.2 Reflective objects

Reflective and transparent surfaces are challenging for conventional 3D reconstruction methods due to the lack of diffuse reflection and surface texture. Photogrammetric 3D reconstruction methods often produce incomplete or noisy results in such cases (Karami et al., 2022).



	Mono-Neus	Neus	MonoSDF	VolSDF	Neuralangelo	Neus-Facto	Photogrammetry	Mono-unisurf	Unisurf	Nerfacto	TensoRF	Instant-NGP
RMSE	0.34	0.35	0.36	0.64	0.78	0.79	0.85	0.92	0.92	1.01	1.01	1.02
MAE	0.36	0.37	0.46	0.55	0.83	1.27	1.10	1.36	1.36	1.35	1.39	1.56
STD	0.50	0.51	0.56	0.82	1.12	1.15	1.50	1.36	1.37	1.52	1.64	1.87

Figure 1. Metrics for the mesh-to-mesh comparisons of several NeRF-based methods applied to the Industrial B object [unit: mm].

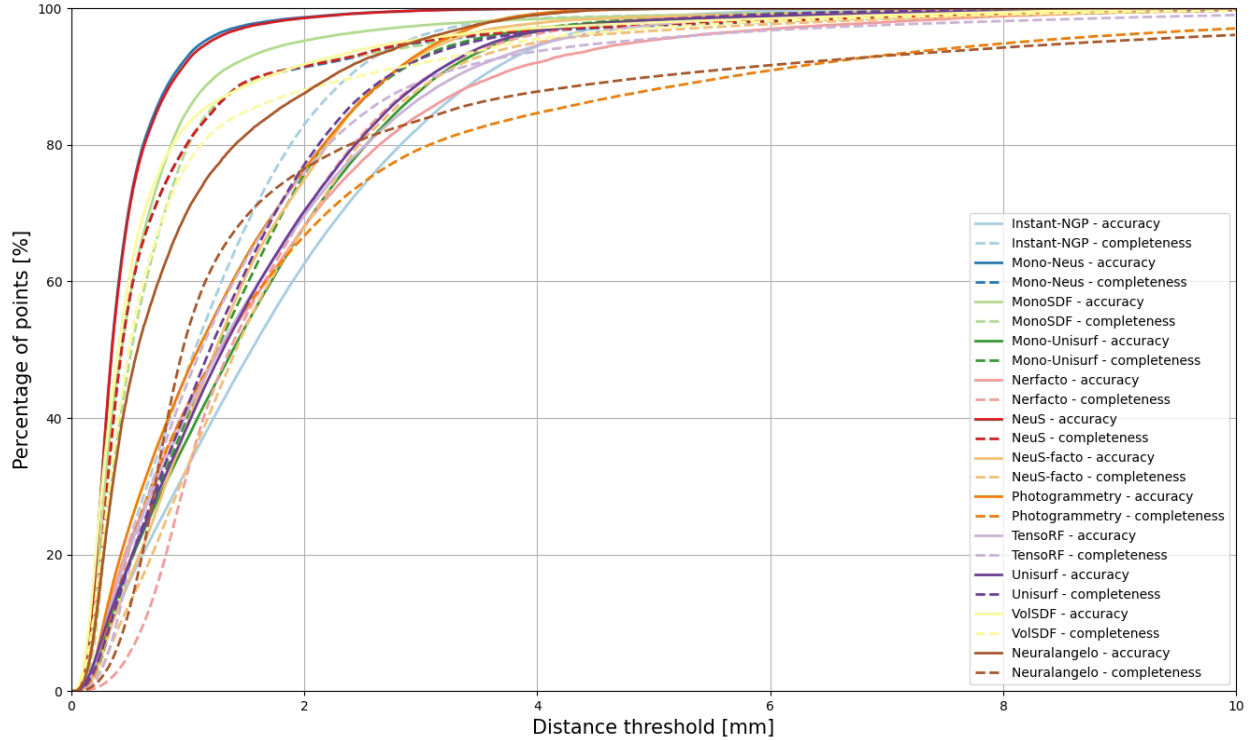


Figure 2. Estimated accuracy and completeness for NeRF-based methods using the Industrial B object of NeRFBK [unit: mm].

However, NeRF-based methods can learn to generate geometric information by leveraging the view-dependent nature of the NeRF model.

The Synthetic Glass dataset of NeRFBK repository is considered: it consists of 300 sequential images extracted from a rendered video with a resolution of 1080x1920 pixels. Neuralangelo (Li et al., 2023) and Nerfacto are the only methods that could successfully reconstruct the object (Figure 3 and Table 2). We selected 3σ as the min/max errors of the comparison analysis with respect to the ground truth. As shown in Figure 3, Neuralangelo achieves a better result among RMSE, MAE and STD compared to Nerfacto as the mesh for former is obviously cleaner while another one is full of noisy. The result for completeness in Figure 4 also verified the previous conclusion as we can see Neuralangelo outperforms than Nerfacto in accuracy. However even if the result generated by Nerfacto is noisy, its reconstruction result is more complete while the other one is full of holes.

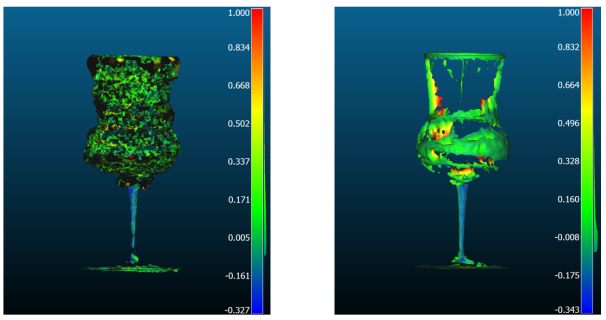


Figure 3. Mesh-to-mesh comparison [mm] for Nerfacto (left) and Neuralangelo (right) on the Synthetic Glass object.

Method	*RMSE*	*MAE*	*STD*	*Mean_E*
Nerfacto	2.34	2.25	2.48	2.10
Neuralangelo	1.57	1.70	1.91	1.30

Table 2. Metric assessment [mm] of the tested NeRF methods.

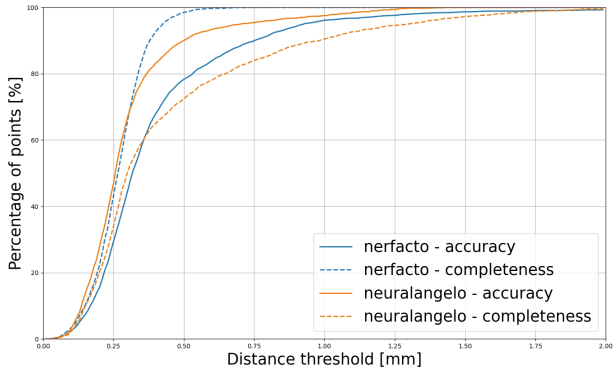


Figure 4. Accuracy and completeness for NeRF-based methods on the Synthetic Glass object.

4.3 Heritage objects

The synthetic Statue dataset (Marelli et al., 2023) is selected to compare Instant-NGP and conventional photogrammetry with respect to the available GT. The Statue is approximately 2x1x5 m, and 50 images were used for the NeRF 3D reconstruction. The metric assessment (Figure 5) shows that photogrammetry performed better than NeRF with a subset of 50 images captured from camera 1. RMSE and STD for the photogrammetric reconstruction are respectively 5.71 mm and 10.81 mm, while for NeRF are 10.47 mm and 15.35 mm.

To check whether small geometric details are reconstructed, cross-sections are extracted from the reconstructed geometries and compared (Figure 6). The photogrammetric profile (blue line) results in smaller distances from the with respect to the NeRF one (red line).

The Vase object, measuring approximately 40 x 30 cm, was acquired using a smartphone Google Pixel2, for a total of about 50 images at a resolution of 4024x3016 pixels. Instant-NGP, Nerfacto and Neuralangelo methods were tested with this dataset, and results compared to the available ground truth (a photogrammetric reconstruction performed with a Reflex camera). All three methods exhibit high levels of accuracy when compared to the ground truth mesh, as seen in the metrics (Figure 7, Table 4), with Nerfacto and Neuralangelo considerably outperforming Instant-NGP.

The completeness analysis (Figure 8) reveals similar performance for Neuralangelo and Nerfacto, while Instant-NGP reaches higher levels of completeness but a much lower accuracy with respect to the other methods. Neuralangelo mesh has the least amount of noise, as visible in Figure 8, and exhibits the best results in terms of both completeness and accuracy.

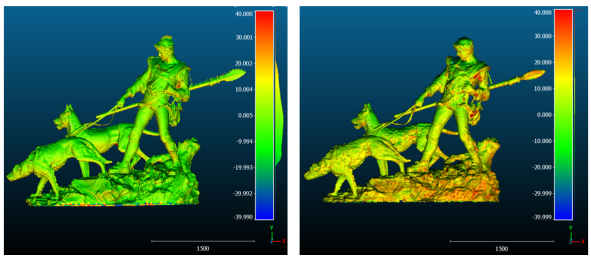


Figure 5. Comparison results [mm] of Photogrammetry (left) and NeRF (right) on the Statue object.

Method	*RMSE*	*MAE*	*STD*	*Mean_E*
Photogrammetry	5.78	8.58	10.26	-1.28
Instant-NGP	10.47	12.85	15.35	6.27

Table 3. Evaluation metrics [mm] for the Statue object.

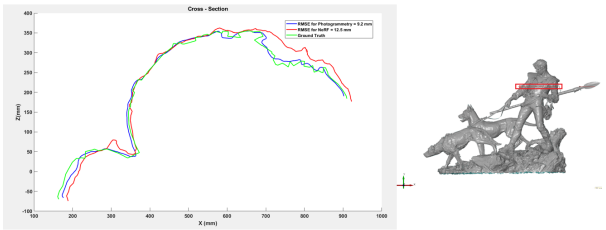


Figure 6. Cross-section profiles on the Status object.

Method	*RMSE*	*MAE*	*STD*	*Mean_E*
Neuralangelo	0.30	0.31	0.42	-0.01
Nerfacto	0.28	0.31	0.37	0.19
Instant-NGP	0.79	0.86	1.11	0.36

Table 4. Metric assessment [mm] for the Vase object.

4.4 Aerial scene

Aerial scene is another challenging dataset for NeRFBK due to the large-scale scenario, varying camera parameters and large image size. The Drone dataset is chosen: 244 nadir high resolution images with 7952x5304 px over an urban and rural area. The only NeRF method able to deliver successful results was Neuralangelo (Li et al., 2023). Visuals and metrics are reported in Figure 9 and Table 5.

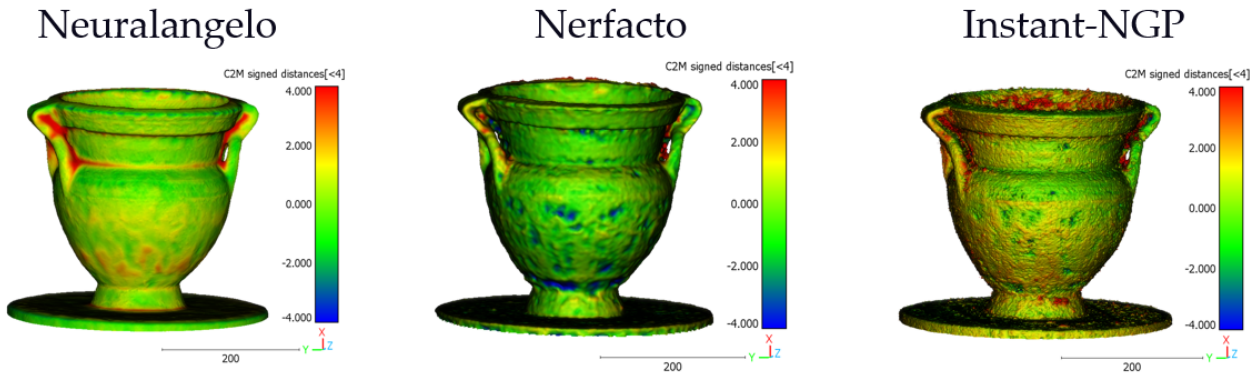


Figure 7. Mesh-to-mesh comparison [mm] for the tested NeRF-based methods on the Vase object. Metrics in Table 4. Instant-NGP apparently has more noisy results with respect to the other two methods.

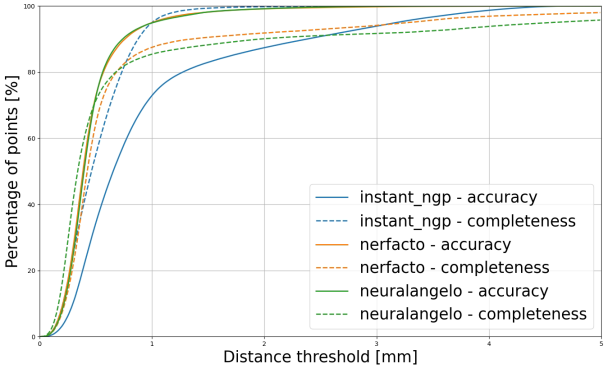


Figure 8. Accuracy and completeness for NeRF-based methods on the Vase object.

Figure 9. GT (left), Neuralangelo textured mesh model (centre) and mesh-to-mesh comparison [m] for the Drone dataset.

Method	*RMSE*	*MAE*	*STD*	*Mean E*
Neuralangelo	0.90	0.96	1.29	-0.26

Table 5. Metric assessment [m] for the Drone dataset.

5. CONCLUSIONS

The paper presented a new set of data - called NeRFBK - comprising real and synthetic data specifically designed for testing and comparing NeRF-based 3D reconstruction methods. In order to address the issue of gathering precise GT data for the evaluation of NeRF-based algorithms, NeRFBK provides multi-scale, indoor and outdoor datasets with high-resolution images, videos and camera parameters. The datasets in NeRFBK contain both real and synthetic data, representing objects from different domains (industry, cultural heritage, geospatial), with different surface characteristics (featureless, well-textured, refractive and reflective), and various objects size and shape.

The presented experiments evaluate and compare the performance of some NeRF-based techniques with metrics and 3D comparisons. The results show that there is not a winner and performances vary according to the scene/object. More investigations and developments are surely necessary to make NeRF more efficient and competitive with conventional photogrammetric methods but for some types of surfaces NeRF methods are relay promising.

In the future, the NeRFBK benchmark will be constantly enriched with more datasets in various challenging and complicated scenarios to encompass as more relevant applications as possible.

REFERENCES

Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P. and Hedman, P., 2022. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. Proc. *CVPR*, pp. 5470-5479.

Chen, A., Xu, Z., Geiger, A., Yu, J. and Su, H., 2022. Tensorf: Tensorial radiance fields. Proc. *ECCV*, pp. 333-350.

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., & Nießner, M. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. Proc. *CVPR*, pp. 5828-5839.

Gao, K., Gao, Y., He, H., Lu, D., Xu, L. and Li, J., 2022. NeRF: Neural radiance field in 3D vision, a comprehensive review. *arXiv preprint arXiv*:2210.00379.

Hedman, P., Srinivasan, P. P., Mildenhall, B., Barron, J. T., & Debevec, P. 2021. Baking neural radiance fields for real-time view synthesis. Proc. *ICCV*, pp. 5875-5884.

Ibrahimli, N., Ledoux, H., Kooij, J. F., & Nan, L. 2023. DDL-MVS: Depth Discontinuity Learning for Multi-View Stereo Networks. *Remote Sensing*, 15(12), 2970.

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. *arXiv preprint arXiv*:2304.10664.

Karami, A., Menna, F. and Remondino, F., 2021. Investigating 3D reconstruction of non-collaborative surfaces through photogrammetry and photometric stereo. *ISPRS Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, pp.519-526.

Karami, A., Battisti, R., Menna, F. and Remondino, F., 2022. 3D digitization of transparent and glass surfaces: state of the art and analysis of some methods. *ISPRS Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2022, 695–702.

Knapitsch, A., Park, J., Zhou, Q.Y. and Koltun, V., 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. ACM *Transactions on Graphics* (ToG), 36(4), pp.1-13.

Li, Z., Müller, T., Evans, A., Taylor, R. H., Unberath, M., Liu, M. Y., & Lin, C. H. 2023. Neuralangelo: High-Fidelity Neural Surface Reconstruction. Proc. *CVPR*, pp. 8456-8465.

Lin, L., Liu, Y., Hu, Y., Yan, X., Xie, K., & Huang, H. 2022. Capturing, reconstructing, and simulating: the urbanscene3d dataset. Proc. *ECCV*, pp. 93-109.

Lindenberger, P., Sarlin, P. E., Larsson, V., & Pollefeys, M. 2021. Pixel-perfect structure-from-motion with featuremetric refinement. In Proc. *ICCV*, pp. 5987-5997.

Liu, D., Zhang, Y., Luo, L., Li, J., & Gao, X. 2021. PDC-Net: robust point cloud registration using deep cyclic neural network combined with PCA. *Applied optics*, 60(11), 2990-2997.

Lu, C., Yin, F., Chen, X., Chen, T., Yu, G., & Fan, J. 2023. A Large-Scale Outdoor Multi-modal Dataset and Benchmark for Novel View Synthesis and Implicit Scene Reconstruction. *arXiv preprint arXiv*:2301.06782.

Mazzacca, G., Karami, A., Rigon, S., Farella, E. M., Trybala, P., & Remondino, F. 2023. NERF for heritage 3d reconstruction. *ISPRS Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 1051-1058.

Marelli, D., Morelli, L., Farella, E. M., Bianco, S., Ciocca, G., & Remondino, F. 2023. ENRICH: Multi-purposE dataset for beNchmaRking In Computer vision and pHotogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 198, 84-98.

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. Proc. *ECCV*.

Mohammadi, M., Rashidi, M., Mousavi, V., Karami, A., Yu, Y. and Samali, B., 2021. Quality evaluation of digital twins generated based on UAV photogrammetry and TLS: Bridge case study. *Remote Sensing*, 13(17), p. 3499.

Mousavi, V., Khosravi, M., Ahmadi, M., Noori, N., Haghshenas, S., Hosseininaveh, A. and Varshosaz, M., 2018. The performance evaluation of multi-image 3D reconstruction software with different sensors. *Measurement*, 120, pp.1-10.

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

Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S., Geiger, A. and Radwan, N., 2022. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. Proc. *CVPR*, pp. 5480-5490.

Nocerino, E., Stathopoulou, E.K., Rigon, S., Remondino, F., 2020. Surface reconstruction assessment in photogrammetric applications. *Sensors*, 20, 5863.

Oechsle, M., Peng, S. and Geiger, A., 2021. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. Proc. *ICCV*, pp. 5589-5599.

Poggi, M., Ramirez, P. Z., Tosi, F., Salti, S., Mattoccia, S., & Di Stefano, L. 2022. Cross-Spectral Neural Radiance Fields. Proc. *3DV*, pp. 606-616.

Radford, A., Wook L.K., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I., 2021. Learning transferable visual models from natural language supervision. *Proc. Machine Learning Research*, pp 8748–8763.

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), 3585.

Stathopoulou, E.K., Remondino, F., 2023. A survey of conventional and learning-based methods for multi-view stereo. *The Photogrammetric Record*, DOI: 10.1111/phor.12456.

Schönberger, J. L., Zheng, E., Frahm, J. M., & Pollefeys, M., 2016. Pixelwise view selection for unstructured multi-view stereo. Proc. *ECCV*.

Schonberger, J.L. and Frahm., J.M., 2016. Structure-from-motion revisited. Proc. *CVPR*, pp. 4104-4113.

Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P., Barron, J.T., Kretzschmar, H. 2022. Block-nerf: Scalable large scene neural view synthesis. Proc. *CVPR*, pp. 8248-8258.

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., Kanazawa, A. 2023. Nerfstudio: A modular framework for neural radiance field development. Proc. *ACM SIGGRAPH*, pp. 1-12.

Toschi, M., De Matteo, R., Spezialetti, R., De Gregorio, D., Di Stefano, L., & Salti, S. 2023. ReLight my NeRF: A Dataset for Novel View Synthesis and Relighting of Real World Objects. Proc. *CVPR*. pp. 20762-20772.

Truong, P., Danelljan, M., & Timofte, R. 2020a. GLU-Net: Global-local universal network for dense flow and correspondences. Proc. *CVPR*, pp. 6258-6268.

Truong, P., Danelljan, M., Gool, L. V., & Timofte, R. 2020b. GOCor: Bringing globally optimized correspondence volumes into your neural network. *Advances in Neural Information Processing Systems*, 33, 14278-14290.

Turki, H., Ramanan, D., & Satyanarayanan, M. 2022. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. Proc. *CVPR*, pp. 12922-12931.

Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T. and Srinivasan, P.P., 2022. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. Proc. *CVPR*, pp. 5481-5490.

Vincent, M., Coughenour, C., Remondino, F., Gutierrez, M.F., Lopez-Menchero Bendicho, V.M., Fritsch, D., 2016. Rekrei: A public platform for digitally preserving lost heritage. Proc. 44th *CAA Conference.*

Vijayanarasimhan, S., Ricco, S., Schmid, C., Sukthankar, R., & Fragkiadaki, K. 2017. Sfm-net: Learning of structure and motion from video. *arXiv preprint arXiv*:1704.07804.

Wang, F., Galliani, S., Vogel, C., Speciale, P., Pollefeys, M., 2021a. PatchmatchNet: Learned Multi-View Patchmatch Stereo. Proc. *CVPR*.

Wang, X., Wang, C., Liu, B., Zhou, X., Zhang, L., Zheng, J., Bai, X., 2021b. Multi-view stereo in the Deep Learning Era: A comprehensive review. *Display*, Vol. 70.

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., & Wang, W. 2021c. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:*2106.10689.

Wizadwongsa, S., Phongthawee, P., Yenphraphai, J., & Suwajanakorn, S. 2021. Nex: Real-time view synthesis with neural basis expansion. Proc. *CVPR*, pp. 8534-8543

Xu, Q., Xu, Z., Philip, J., Bi, S., Shu, Z., Sunkavalli, K., & Neumann, U. 2022. Point-nerf: Point-based neural radiance fields. Proc. *CVPR*, pp. 5438-5448.

Yao, Y., Luo, Z., Li, S., Shen, T., Fang, T. and Quan, L., 2019. Recurrent MVSet for high-resolution multi-view stereo depth inference. Proc. *CVPR*, pp. 5525-5534.

Yariv, L., Gu, J., Kasten, Y., & Lipman, Y. 2021. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34, 4805-4815.

Yen-Chen, L., Florence, P., Barron, J. T., Lin, T. Y., Rodriguez, A., & Isola, P. 2022. Nerf-supervision: Learning dense object descriptors from neural radiance fields. Proc. *ICRA*, pp. 6496-6503.

Yu, Z., Chen, A., Antic, B., Peng, S. P., Bhattacharyya, A., Niemeyer, M., Tang, S., Sattler, T., & Geiger, A., 2022a.

SDFStudio: A Unified Framework for Surface Reconstruction. Retrieved from https://github.com/autonomousvision/sdfstudio.

Yu, Z., Peng, S., Niemeyer, M., Sattler, T., Geiger, A., 2022b. MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. Proc. *NIPS*.

Zhang, X., Fanello, S., Tsai, Y. T., Sun, T., Xue, T., Pandey, R., Freeman, W. T. 2021. Neural light transport for relighting and view synthesis. ACM *Transactions on Graphics*, 40(1), 1-17.