APPLICABILITY OF 3D-SCAN METHODS USING SMARTPHONES FOR HRTF SYNTHESIS

Benedikt Bugl[1], Azamat Musaev[1], Armin Sehr[1]
[1] *OTH Regensburg, Regensburg, Germany*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Individual head-related transfer functions (HRTFs) are an essential tool for accurate spatial audio reproduction, but traditional measurement procedures are time-consuming and require specialized laboratory equipment. Numerical approaches enable the computation of individual HRTFs from 3D models of the torso, head, and pinnae. Although capturing these 3D models can present a challenge in itself, recent advances in smartphone-based 3D scanning technologies offer a more accessible alternative to professional scanning systems.

This paper investigates the applicability of three smartphone-based 3D scanning methods, namely Photogrammetry, LiDAR, and TrueDepth, for HRTF synthesis using Mesh2HRTF. The contribution focuses on evaluating the methods with respect to ease of use and achievable spatial resolution. Therefore, 3D scans of an artificial head are compared against reference measurements based on anthropometric parameters, and the resulting mesh quality is analyzed.

The results show that smartphone-based LiDAR is not suitable due to its limited spatial resolution, while Photogrammetry and TrueDepth can provide sufficient geometric detail for HRTF synthesis. Among the evaluated methods, TrueDepth offers the best compromise between acquisition effort and mesh resolution, whereas Photogrammetry provides high flexibility and device compatibility.

Keywords: *Virtual Acoustics, Photogrammetry, LiDAR, TrueDepth*

1. Introduction

Head-related transfer functions (HRTFs) describe the direction-dependent filtering of sound caused by the anatomical features of a listener, in particular the torso, head, and pinnae [1]. They are a fundamental component for the synthesis of immersive auditory experiences in virtual and augmented reality, as they contain binaural and monaural cues required for three-dimensional sound localization [2]. Since HRTFs are highly individual due to anatomical variability, individual datasets are considered the gold standard for accurate spatial audio reproduction [3].

Traditionally, individual HRTFs are obtained by acoustic measurements using microphones placed in the ear canals of a subject [4] within controlled environments such as anechoic or semi-anechoic chambers, fitted with loudspeaker arrays and positioning systems. The measurement procedure is time-consuming and demanding for participants, who must remain largely motionless during the acquisition process. Reported acquisition times vary depending on the measurement setup and spatial resolution: high-resolution measurements can take approximately 20–40 minutes per subject [5]. These constraints limit the use of HRTFs and spatial audio to specialized institutions.

To overcome these limitations, numerical approaches for HRTF synthesis based on geometric models of the listener have gained increasing attention. Methods such as the boundary element method (BEM) [6], as implemented in frameworks such as Mesh2HRTF, enable the computation of individual HRTFs from detailed 3D meshes of the head and ears [7]. The quality of such approaches strongly depends on the geometric fidelity of the underlying models, particularly with respect to the pinnae, where small structural variations significantly affect acoustic localization cues [8]. Accordingly, the literature indicates that geometric deviations should remain within approximately 1 mm in order to preserve the acoustic validity of the synthesized HRTFs [9].

Recent advances in consumer-grade sensing technologies have made it possible to capture 3D geometry using widely available devices such as smartphones or tablets, offering low-cost and accessible alternatives to dedicated scanning systems [10]. These approaches have the potential to reduce acquisition time for the test subject and simplify the workflow for generating

Corresponding author: *benedikt.bugl@oth-regensburg.de.*

individual HRTFs. Without the need for expensive and specialized hardware, the barrier to entry for personalized 3D audio can be significantly reduced, potentially making individualized spatial audio accessible to the general public.
However, the suitability of smartphone-based scanning techniques for HRTF synthesis remains an open question. While recent datasets demonstrate the feasibility of combining 3D scans with numerical HRTF computation [11], a systematic comparison of different consumer-grade acquisition methods with respect to their applicability for Mesh2HRTF is still lacking, in contrast to industrial-grade scanning methods, which have already been examined in previous studies [12].
In this work, we present a comparative evaluation of three smartphone-based 3D scanning approaches – LiDAR (LD), Photogrammetry (PG), and TrueDepth (TD) – for capturing head and ear geometries. The resulting meshes are assessed in terms of geometric quality, usability, and their potential applicability for numerical HRTF synthesis.

2. Workflow for HRTF-Synthesis

2.1. Scanning workflow

The intended workflow follows the practical recommendation in the Any2HRTF scanning workflow, a rough base head mesh is combined with detailed ear scans to preserve the geometric fidelity of the pinna region while keeping the remaining head geometry comparatively coarse [13]. A similar multi-scan photogrammetric strategy was also used by Di Giusto et al. to investigate the influence of scanning accuracy on numerically determined HRTFs of a dummy head [14]. This hybrid acquisition strategy reduces the overall scanning effort while retaining the pinna detail required for reliable numerical HRTF simulation. A fixed acquisition time of 10 minutes was set for each scanning method.

2.2. Scan Processing

Before numerical HRTFs can be calculated, the scanned head-and-pinna geometry must be converted into a watertight, manifold surface mesh without overlapping or self-intersecting geometry. In the workflow considered in the context of this contribution, the three recorded scans are first cleaned, aligned, and merged into a single model. Alignment is performed in *MeshLab*, and the merged surface is reconstructed using *Poisson Surface Reconstruction* [15]. The *Poisson Surface Reconstruction* was performed according to the Any2HRTF tutorial, using `Pre-Clean` and a `Reconstruction Depth` of 10 [16]. The resulting mesh is then inspected and repaired in *Meshmixer*; remaining defects and openings, e.g. in the lower neck region, are closed where necessary [17]. In a subsequent modeling step, the geometry is adapted to the simulation requirements by correcting major scan artefacts, smoothing residual surface defects, and, if required, modifying the ear-canal entrances [16], [18]. Finally, the mesh is graded before export to Mesh2HRTF in order to reduce computational cost while preserving accuracy in acoustically relevant regions: fine elements are retained around the simulated ear, whereas less critical regions can be discretized more coarsely [7], [8]. Based on this processing workflow, the criteria for the following comparison of scanning methods are derived.

3. Comparison of Smartphone-based 3D-Scan methods

This section reviews and compares the three most commonly used smartphone-based 3D scanning methods in terms of their suitability for capturing the anatomical features required for HRTF synthesis using the procedure described above.

3.1. Photogrammetry (PG)

PG is a passive 3D reconstruction technique that derives geometry from multiple overlapping 2D images taken from different viewpoints, as visualized in Figure 1. By detecting and matching visual features across images, algorithms estimate camera poses and reconstruct the scene using structure-from-motion (SfM) and multi-view stereo (MVS) methods [19]. The resulting point clouds are converted into dense meshes, which can be optionally textured. For the PG scan, 120 overlapping images were acquired with an iPhone 16 Pro using its 48 MP camera system and reconstructed in KIRI Engine, version 4.0.4, using the Photo Scan mode [20], [21].

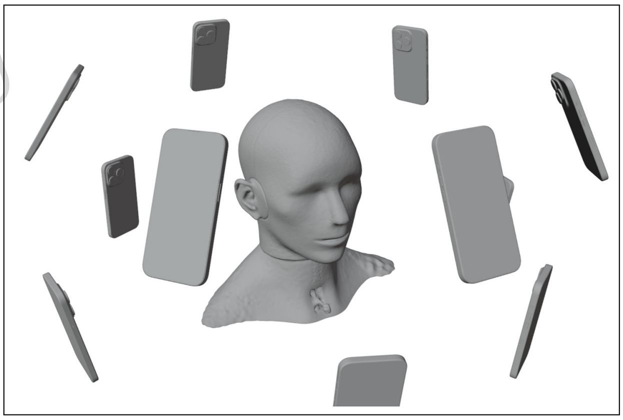


Figure 1: Operating principle of Photogrammetry (PG): Multiple pictures of the object are taken from multiple angles.

3.2. LiDAR (LD)

LiDAR (LD), short for Light Detection and Ranging, is an active depth-sensing technique that measures distances by emitting laser pulses and recording their time-of-flight to the scene and back [22]. Figure 2 shows how in smartphone implementations, a near-infrared laser projects structured light or pulse patterns onto nearby surfaces, and a sensor measures the return time or phase shift to estimate depth. By scanning across the scene, a depth map is generated, which can be fused with RGB images to produce textured

3D meshes. In this work, the LD scan was acquired with an iPhone 16 Pro and reconstructed in KIRI Engine, version 4.0.4, using the LiDAR Scan mode [20], [23].

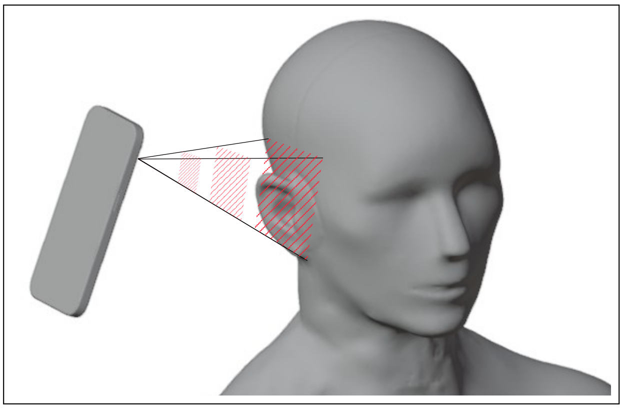


Figure 2: Ooperating principle of LiDAR (LD): light pulses are projected on to the object and distances are calculated from the time of flight of different areas.

3.3. TrueDepth (TD)

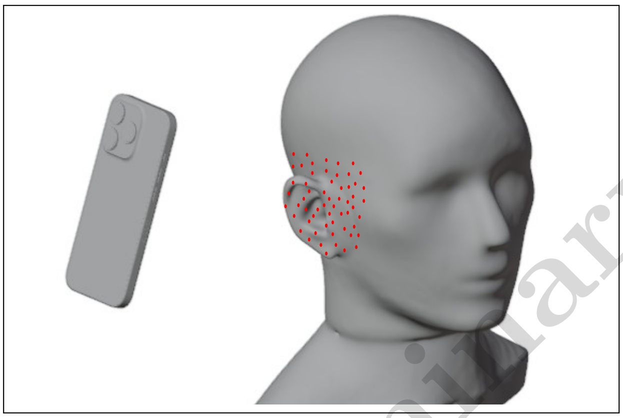


Figure 3: Operating principle of TrueDepth (TD): An dot pattern of infrared light is projected onto the object.

TD is an active, structured-light-based depth sensing system, primarily used for face recognition in mobile devices. It projects a dense pattern of infrared dots onto the scene, as shown in Figure 3, and captures their deformation using an infrared camera to infer depth through triangulation. Combined with RGB imaging and onboard processing, it produces real-time depth maps and 3D representations of nearby objects [24]. The TD scans were acquired with an iPhone 11 using Heges 3D Scanner, version 1.8.2, with the `0.5 mm` precision setting [25], [26]. An additional TD scan was also performed with an iPhone 16 Pro; however, due to the lower surface quality observed in this setup, the iPhone 11 scan was used for the comparison of the scanning methods. The iPhone 16 Pro result is shown separately to illustrate the device-dependent performance of TD scanning.

3.4. Comparison and Discussion

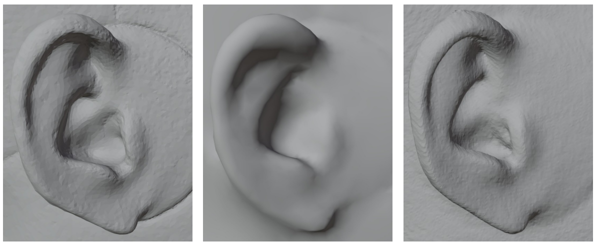
Table 1 shows a direct comparison of the three scanning methods. For this comparison, PG images and LD scans were acquired with an iPhone 16 Pro smartphone, while TD scans were acquired with an iPhone 11 smartphone.

Table 1: Comparison of smartphone-based 3D scanning technologies, accumulated data from [9], [10], [27], [28].

	PG	**LD**	**TD**
Principle	image-based reconstruction	time-of-flight	structured IR light
Acquisition Effort	high	low	low
Comput. Effort	high	low	low
Lighting	critical	not critical	not critical
Scaling	reference geometry	direct	direct
Range	flexible	$\leq$ 5 m	0.2–1 m
Resolution	$<$ 1 mm	$\sim$10 mm	$\sim$0.5 mm

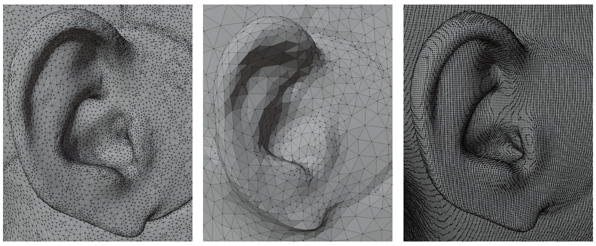
PG, LD, and TD differ significantly in their acquisition principles, effort, and suitability for capturing fine anatomical detail. PG exhibits the highest acquisition and computational effort, as it requires lots of images to be taken and a large processing effort to generate the 3D-model. In contrast, for both LD and TD the 3D-model is being built in real time on the smartphone. As a passive method, PG is highly sensitive to lighting conditions and surface texture, whereas LD and TD are active methods and largely independent of ambient illumination. A key distinction lies in scale recovery: PG requires an external reference to obtain an absolute scale, while both LD and TD inherently provide measurements in real-world units.

Regarding spatial resolution, PG and TD outperform LD. PG can achieve sub-millimeter detail under favorable conditions, and TD typically reaches approximately 0.5 mm resolution at close range. The difference in detail is clearly visible in Figure 4. TD and PG both show very detailed scans, while the pinna geometry in the LD-Scan is visibly distorted. A complementary perspective is provided by Figure 5 and Table 2. For the same acquisition time, TD produces the finest mesh, followed by PG. In contrast, the LD scan exhibits large facets and a very sparse point cloud, showing that this method is clearly unsuitable for the intended application. To put the values into context, Tables 2 and 4 also report the results of an Einstar handheld 3D scanner (HS) from SHINING 3D as a reference. The HS scan was acquired using the manufacturer-provided software, version 1.2.2.0, in *Portrait Scan* mode with *Texture Alignment* enabled and a resolution setting of `0.2 mm` [29].



(a) PG (b) LD (c) TD

Figure 4: Comparison of Pinna 3D-Scan after smoothing



(a) PG (b) LD (c) TD

Figure 5: Comparison of scans methodes

Table 2: Measurements in the region of the right ear for Photogrammetry (PG), LiDAR (LD), and TrueDepth (TD) and a handheld scanner (HS).

	PG	**LD**	**TD**	**HS**
Number of points	5 846	707	95 602	132 331
Mean point spacing	1.1 mm	3.0 mm	0.3 mm	0.2 mm
Max point spacing	4.9 mm	8.5 mm	0.9 mm	0.7 mm

Table 4 summarizes the absolute deviations of the anthropometric parameters defined by Algazi et al. [30], see Figure 6 and Table 3. Due to the complex geometry and the absence of clearly defined landmarks, both the physical measurements and the measurements derived from the 3D models are associated with measurement uncertainty.
Nevertheless, the quantitative results are consistent with the visual assessment of the scans and indicate that LD does not provide sufficient geometric detail for the intended application. The TD scan achieves the highest overall accuracy among the evaluated methods. The comparatively poorer performance of PG can largely be attributed to the limited acquisition time, suggesting that its theoretically higher accuracy cannot be fully realized under time-constrained conditions and sub-optimal lighting.
These findings are consistent with the observations of Di Giusto et al. [14]. Their work emphasizes that geometric scan quality is not only a visual or metrological requirement, but directly relevant for the acoustic validity of numerical HRTF simulations. In this context, the present results provide a practical extension to their findings under smartphone-based acquisition conditions. The poor LD performance indicates that a fast and robust scanning procedure alone is not sufficient if the acoustically relevant pinna structures are not captured with adequate spatial detail. By contrast, TD preserves substantially more pinna detail within the fixed acquisition time and therefore better matches the geometric requirements of the established Mesh2HRTF/Any2HRTF workflow [13], [16]. PG remains a viable and flexible alternative, but its performance in this study shows that its achievable accuracy depends strongly on acquisition time, lighting, image quality, and scaling.

Table 3: Anthropometric parameters adapted from Algazi et al. [30].

Parameter ID	**Anthropometric Parameter**
d1	cavum concha height
d2	cymba concha height
d3	cavum concha width
d4	fossa height
d5	pinna height
d6	pinna width
d7	intertragal incisure width
d8	cavum concha depth

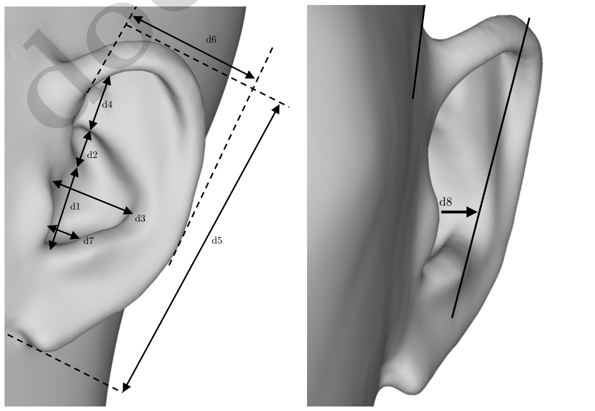


Figure 6: Illustration of the measured anthropometric parameters, adapted from Algazi et al. [30].

Table 4: Absolute deviations of the measured anthropometric parameters (Adapted from [30]) for Photogrammetry (PG), LiDAR (LD), and TrueDepth (TD) and a handheld scanner (HS) from physically measured reference values.

ID	**PG [mm]**	**LD [mm]**	**TD [mm]**	**HS [mm]**
d1	1.13	2.71	0.02	0.02
d2	0.67	0.20	0.19	0.01
d3	1.77	2.73	0.35	0.29
d4	1.28	3.74	0.65	0.60
d5	0.23	4.01	0.07	0.15
d6	0.54	3.30	0.25	0.03
d7	0.45	2.11	0.01	0.01
d8	0.35	3.69	0.09	0.02
mean	**0.80**	**2.81**	**0.20**	**0.14**

The effective working range shows substantial differences between the methods. PG is flexible and mainly constrained by camera properties, LD in smartphones supports ranges of several meters, and TD is restricted to short distances of approximately 0.2–1 m. In the context of this work however, the working range is not critical, as long as the head geometry can still be adequately captured.
Regarding TD, Figure 7 illustrates that scan quality can vary considerably between different smartphone models. The iPhone 11, one of the earlier Apple devices equipped with the TrueDepth sensor, produced substantially better results than the newer iPhone 16 Pro. Similar variations in TD performance across different Apple device generations have also been reported for iPads [24]. This device dependency highlights an important advantage of PG: since it relies solely on RGB images, it can be used with practically any smartphone camera. In contrast, TD and LD require dedicated depth sensors that are only available on selected devices, and their performance may vary considerably between hardware generations.

(a) iPhone 16 Pro **(b)** iPhone 11

Figure 7: Comparison of TrueDepth (TD) scans captured with two iPhone models

4. Conclusion

The comparative evaluation of smartphone-based 3D scanning technologies reveals distinct trade-offs for HRTF generation.
LiDAR (LD), although robust and fast, was unsuitable in the evaluated setup due to its limited ability to capture fine anatomical structures, particularly within the pinna.
Photogrammetry (PG) demonstrates high potential due to its device independence but requires increased acquisition and processing effort and is sensitive to lighting conditions.
TrueDepth (TD) showed the most favorable results in this evaluation, combining the highest geometric accuracy with low acquisition effort. Its main drawback is the restriction to devices supporting the TrueDepth sensor.
Overall, both PG and TD are viable approaches for generating scans suitable for numerical HRTF synthesis, with TD currently offering the most favorable balance between usability and geometric fidelity. However, further investigations with human subjects and direct comparisons between HRTFs synthesized from smartphone-based scans and measured HRTFs remain necessary to validate the practical applicability of these methods.

References

[1] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, " Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300–321, May 1995.

[2] F. Brinkmann and C. Pike, "Binauraltechnik," in *Handbuch der Audiotechnik*, S. Weinzierl, Ed., Berlin, Heidelberg: Springer Vieweg, 2023, pp. 1–23. DOI: 10.1007/978-3-662-60357-4_27-2.

[3] S. Li and J. Peissig, "Measurement of Head-Related Transfer Functions: A Review," *Applied Sciences*, vol. 10, no. 14, p. 5014, 2020. DOI: 10.3390/app10145014.

[4] H. Møller, "Fundamentals of Binaural Technology," *Applied Acoustics*, vol. 36, no. 3–4, pp. 171–218, 1992. DOI: 10.1016/0003-682X(92)90046-U.

[5] E. A. Torres-Gallegos, F. Orduña-Bustamante, and F. Arámbula-Cosío, "Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database," *Applied Acoustics*, vol. 97, pp. 84–95, 2015. DOI: 10.1016/j.apacoust.2015.04.009.

[6] L. Gaul, M. Kögl, and M. Wagner, *Boundary Element Methods for Engineers and Scientists: An Introductory Course with Advanced Topics*. Springer Berlin Heidelberg, 2003, ISBN: 978-3-540-00463-9.

[7] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "Mesh2HRTF: An Open-Source Software Package for the Numerical Calculation of Head-Related Transfer Functions," in *Proceedings of the 22nd International Congress on Sound and Vibration (ICSV 22)*, Florence, Italy, Jul. 2015, pp. 1–8. DOI: 10.13140/RG.2.1.1707.1128.

[8] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "A priori mesh grading for the numerical calculation of the head-related transfer functions," *Applied Acoustics*, vol. 114, pp. 99–110, 2016, ISSN: 0003-682X. DOI: https://doi.org/10.1016/j.apacoust.2016.07.005.

[9] M. Dinakaran et al., "Perceptually Motivated Analysis of Numerically Simulated Head-Related Transfer Functions Generated By Various 3D Surface Scanning Systems," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018,

pp. 551–555. DOI: 10.1109/ICASSP.2018.8461789.

[10] M. Vogt, A. Rips, and C. Emmelmann, “Comparison of iPad Pro®’s LiDAR and TrueDepth Capabilities with an Industrial 3D Scanning Solution,” *Technologies*, vol. 9, no. 2, p. 25, 2021. DOI: 10.3390/technologies9020025.

[11] K. Poole et al., “The Extended SONICOM HRTF Dataset and Spatial Audio Metrics Toolbox,” in *Forum Acusticum / EuroNoise 2025 Málaga (Spain)*, Jun. 2025, pp. 6483–6486. DOI: 10.61782/fa.2025.0864.

[12] A. Reichinger, P. Majdak, R. Sablatnig, and S. Maierhofer, “Evaluation of Methods for Optical 3-D Scanning of Human Pinnas,” in *2013 International Conference on 3D Vision*, 2013, pp. 390–397. DOI: 10.1109/3DV.2013.58.

[13] Any2HRTF Project. “Basic tutorial: 3D scanning,” Accessed: Mar. 28, 2026. [Online]. Available: https://github.com/Any2HRTF/Mesh2HRTF/wiki/Basic_tutorial_3d_scanning.

[14] F. Di Giusto, S. van Ophem, W. Desmet, and E. Deckers, “Analysis of laser scanning and photogrammetric scanning accuracy on the numerical determination of Head-Related Transfer Functions of a dummy head,” *Acta Acustica*, vol. 7, p. 53, 2023. DOI: 10.1051/aacus/2023049.

[15] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, “MeshLab: An Open-Source Mesh Processing Tool,” in *Eurographics Italian Chapter Conference*, The Eurographics Association, 2008, pp. 129–136. DOI: 10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136.

[16] Any2HRTF Project. “Basic 3D mesh clean-up and merging,” Accessed: Mar. 28, 2026. [Online]. Available: https://github.com/Any2HRTF/Mesh2HRTF/wiki/Basic_3D_Mesh_cleanup_and_merging.

[17] Autodesk Inc., *Meshmixer*, Computer software, Autodesk Help, 2019. [Online]. Available: https://help.autodesk.com/view/MSHMXR/2019/ENU/.

[18] Blender Foundation, *Blender Manual*, Accessed 2026-03-22, 2026. [Online]. Available: https://docs.blender.org/manual/en/latest/.

[19] E. Stathopoulou, M. Welponer, and F. Remondino, “Open-source image-based 3d reconstruction pipelines: Review, comparison and evaluation,” in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W17, 2019, pp. 331–338. DOI: 10.5194/isprs-archives-XLII-2-W17-331-2019.

[20] Apple Inc., *iPhone 16 Pro – Tech Specs*, 2024. Accessed: Jan. 23, 2025. [Online]. Available: https://support.apple.com/en-en/121031.

[21] KIRI Team. “Photo Scan Mode,” Accessed: Jan. 23, 2026. [Online]. Available: https://www.kiriengine.app/blog/explained/photo-scan-mode.

[22] G. Luetzenburg, A. Kroon, and A. A. Bjørk, “Evaluation of the Apple iPhone 12 Pro LiDAR for an Application in Geosciences,” *Scientific Reports*, vol. 11, p. 22 221, 2021. DOI: 10.1038/s41598-021-01763-9.

[23] KIRI Team. “LiDAR Scan Mode,” Accessed: Jan. 23, 2026. [Online]. Available: https://www.kiriengine.app/blog/explained/lidar-scan-mode.

[24] S. Urban, T. Lindemeier, D. Dobbelstein, and M. Haenel, *On the Issues of TrueDepth Sensor Data for Computer Vision Tasks Across Different iPad Generations*, 2022. arXiv: 2201.10865 [cs.CV].

[25] Apple Inc. “iPhone 11 – Technical Specifications,” Accessed: Jan. 23, 2025. [Online]. Available: https://support.apple.com/en-gb/111865.

[26] M. Simonik. “Heges 3D Scanner – Features (Frequently Asked Questions),” Accessed: Jan. 23, 2025. [Online]. Available: https://hege.sh/faq.

[27] T. Luhmann, S. Robson, S. Kyle, and J. Boehm, *Close-Range Photogrammetry and 3D Imaging*, 2nd ed. De Gruyter, 2013. DOI: 10.1515/9783110302783.

[28] B. Plaß and T. Klauer, “Potenziale von LiDAR-Smartphones und Machine Learning zur Bewertung der altersgerechten Nutzbarkeit von Wohnraum,” *Allgemeine Vermessungs-Nachrichten*, vol. 131, no. 1, pp. 28–38, 2024. DOI: 10.14627/avn.2024.1.4.

[29] F. Yang, *Einstar User Manual*, version v1.0.6.0, Modified: Jan. 12, 2024. Accessed: 2026-07-07, SHINING 3D, 2024. [Online]. Available: https://support.einstar.com/helpdesk/attachments/60104900453.

[30] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The CIPIC HRTF Database,” in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 99–102. DOI: 10.1109/ASPAA.2001.969552.