BioMedical Engineering OnLine

https://doi.org/10.1186/s12938-026-01515-8

Article in Press

# Smartphone photogrammetry for prosthetics and orthotics: accuracy and reliability across upper-limb, lower-limb, and AFO casts

Received: 25 Sep 2025

Accepted: 06 Jan 2026

Published online: 16 January 2026

Cite this article as: Ellis, A., Pendse, V., Ngan, C. *et al.* Smartphone photogrammetry for prosthetics and orthotics: accuracy and reliability across upper-limb, lower-limb, and AFO casts. *BioMed Eng OnLine* (2026). https://doi.org/10.1186/s12938-026-01515-8

Arthur Ellis, Vishal Pendse, Calvin Ngan & Jan Andrysek

**Title: Smartphone Photogrammetry for Prosthetics and Orthotics: Accuracy and Reliability Across Upper-Limb, Lower-Limb, and AFO Casts**

Arthur Ellis[1,3], Vishal Pendse[1,2], Calvin C Ngan[1], Jan Andrysek[1,2*]

[1]Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON M4G 1R8, Canada.
[2]Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada.
[3]School of Biomedical Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.
*Correspondence: jan.andrysek@utoronto.ca

**Abstract**

**Background:** Conventional shape capture in prosthetics and orthotics (P&O) relies on plaster casting of a positive mould, which is then hand-rectified to the desired shape. While effective under expert practice, this workflow is labor-intensive, equipment-dependent, and difficult to archive or share. Digital approaches using structured-light scanners address some of these limitations but remain costly and require dedicated training. This study evaluated whether smartphone photogrammetry can accurately and reliably capture prosthetic and orthotic cast geometries, and assessed its usability, in comparison to a clinical structured-light scanner for integration into clinical workflows.

**Results:** A clinical-grade structured-light scanner (EinScan H2) served as the reference and demonstrated small volumetric and dimensional errors, at 0.21 ± 0.15% and 0.35 ± 0.18 mm, with intraclass correlation coefficients (ICCs) greater than 0.9999. Relative to this reference, across 12 cast models (upper limb, lower limb, and ankle–foot orthosis), smartphone photogrammetry achieved a volumetric error of 0.89 ± 0.68% and a dimensional error of 0.89 ± 0.51 mm; the mean surface point-to-point distance was 0.24 ± 0.19 mm. Reliability across operators was near-perfect (ICCs ≥ 0.9997). Usability data showed approximately 62 photographs and 88 seconds per capture for photogrammetry (about 12 minutes cloud processing) versus 34 seconds capture (about 85 seconds desktop processing) for the reference scanner. Photogrammetry scored higher on the System Usability Scale (79 versus 58).

**Conclusions:** On casts, smartphone photogrammetry produced accurate and reliable meshes with favorable perceived usability and minimal hardware demands. These findings support its integration into digital workflows in P&O, particularly for scanning rectified positive casts and other stable geometries. Further multi-site evaluations on live limbs should determine acceptable capture-time thresholds and effective stabilization strategies to ensure clinical feasibility in routine practice.

**Keywords:** 3D scanning; prosthetics; orthotics; photogrammetry; low-cost; smartphones; scanning

# 1. Background

In the field of prosthetics and orthotics (P&O), the design of customized components, such as prosthetic sockets and orthoses, significantly influences rehabilitation outcomes and device usage (1). The design process typically involves three steps: (i) capturing the limb shape, (ii) rectifying that shape to the desired device geometry, and (iii) fabricating the device (2,3). As the first step, shape capture defines the baseline model for rectification and fabrication; errors here propagate through the workflow, making accuracy critical.

Most clinics currently employ cast-based workflows for device design, where a positive mould of the limb shape is taken using Plaster of Paris, rectified by hand, and a thermoplastic is draped over it to yield the finalized device. This process, although well-established and effective under skilled hands, is labor- and time-intensive, in addition to being dependent on individual expertise (3,4). Further, because the mold is typically destroyed during fabrication, the casting process may need to be repeated if a device breaks or is uncomfortable to use, adding further burden for both the clinician and client (5).

In recent years, digital shape capture has emerged to address these limitations. This process typically uses structured-light or laser 3D scanners to convert limb and cast geometries into digital meshes that can be archived, shared, rectified through software, and fabricated using additive manufacturing (6,7). In addition to fabrication, these meshes serve as digital records for longitudinal tracking, audit, second opinions, and better collaboration with centralized fabrication hubs (8). Despite demonstrated reliability and sub-millimeter accuracy, adoption is limited by hardware costs and maintenance, licensing fees, and training demands (6,9–11). These barriers disproportionately affect low-resource and rural clinics, where access to advanced equipment is limited. Furthermore, clinical adoption is particularly sensitive to usability and workflow fit; tools must minimize capture and processing time, reduce labor, and integrate with existing clinical routines to be viable at larger scales (3,12,13).

At the same time, end-to-end digital workflows have matured in recent years, with some studies finding that digitally produced sockets (14) and orthoses (15,16) can meet clinical expectations. This strengthens the case for exploring lower-cost, simpler shape capture options, not to replace clinical input, but to reduce the equipment, cost, and usability barriers that limit deployment of otherwise viable digital pipelines (13,17). As a practical compromise within these digital workflows, some clinics have begun to scan rectified positive casts rather than live limbs. For example, the US Department of Veterans Affairs has implemented a model in which diagnostic sockets are scanned and sent to centralized fabrication units for 3D printing of the final device (18).

Smartphone photogrammetry, which reconstructs 3D geometries by stitching together overlapping photographs from different angles taken using a smartphone camera, offers a low-cost, low-infrastructure alternative to clinical-grade scanners. Given the widespread availability of smartphones, minimal hardware needs, and cloud-based processing software, photogrammetry has the potential to democratize P&O care by lowering costs, reducing labor, and enabling remote care (19). Early case-study-based research has demonstrated its feasibility for scanning foot insoles (20), spinal orthoses (21), transtibial sockets (22), and transradial prosthetic sockets (23). Despite this promise, most studies are small and focus on lower-limb applications, leaving a gap in understanding how photogrammetry performs across a breadth of P&O applications.

Accordingly, this study has evaluated smartphone photogrammetry across a wide range of casts used in P&O. First, the accuracy and reliability of a clinically used structured-light scanner were benchmarked by

comparing scans of a 3D-printed object to its original CAD model. Second, the accuracy and inter-/intra-rater reliability of smartphone photogrammetry were evaluated on 12 P&O-relevant cast models: five upper-limb, three lower-limb, and four orthoses. Third, usability and workflow fit were assessed using the capture time and image counts. We expect that, for these stable cast geometries, photogrammetric errors would be small relative to clinically used scanners, and are unlikely to constrain clinical applicability. Collectively, these analyses are intended to provide an evidence base for integrating photogrammetry into P&O workflows to enable lower-cost, faster, and potentially remote care.

# 2. Results

## 2.1 EinScan H2 Accuracy and Reliability

Across all raters, the signed average volumetric and dimensional differences between the EinScan H2 scans and the known-volume model were -0.00 ± 0.26% and -0.31 ± 0.20 mm, respectively. The largest difference was observed along the Z-axis, -0.67 ± 0.35 mm, then the Y-axis, -0.22 ± 0.27 mm, and the smallest was along the X-axis, -0.04 ± 0.20 mm (Figure 1). The absolute average volumetric and dimensional differences were 0.21 ± 0.15% and 0.35 ± 0.18 mm, respectively. The inter- and intra-rater reliability both achieved ICC values over 0.9999, with the 95% CIs ranging between 0.9998 and 1.000 (Table 1).

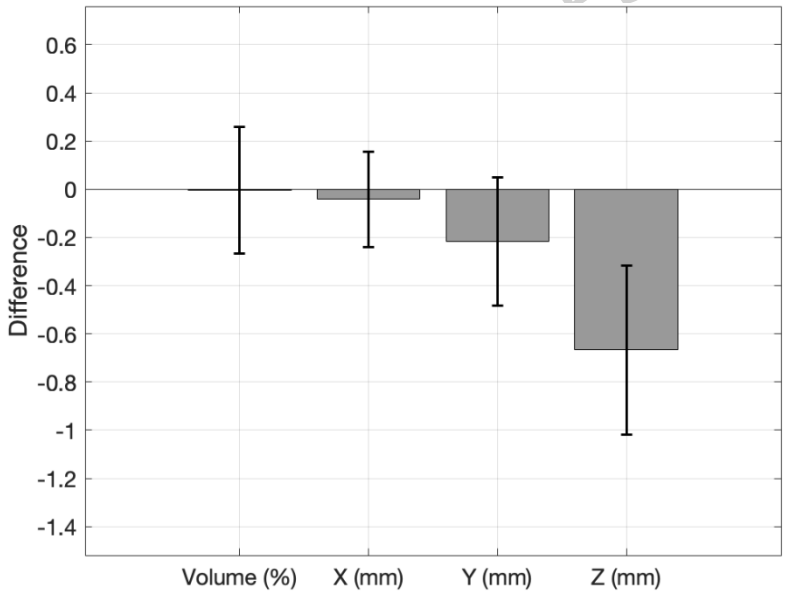



*Figure 1. Signed average volume (%) and dimensional difference (mm) for the EinScan H2 reference scanner against the known-volume surrogate model across three raters. These results benchmark the accuracy and reliability of the EinScan H2, used as a reference device in this study.*

*Table 1. Inter- and intra-rater reliability of the EinScan H2 measurements*

| | Rater | MD (SD) | ICC | 95% CI |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Intra-rater | 1 | 0.2 (0.12) | 1.0000 | 0.9998 - 1.0000 |
| | 2 | 0.2 (0.14) | 1.0000 | 0.9999 - 1.0000 |
| | 3 | 0.2 (0.12) | 1.0000 | 1.0000 |
| Inter-rater | 1 vs. 2 vs. 3 | 0.2 (0.11) | 1.0000 | 1.0000 |

2.2 KIRI Engine Accuracy and Reliability

Relative to the EinScan H2 references, the signed average volumetric and dimensional differences were -0.31 ± 1.08% and 0.51 ± 0.64 mm, respectively. The largest difference was observed along the Z-axis, 0.66 ± 1.45 mm, then the X-axis, 0.54 ± 0.81 mm, and the smallest was along the Y-axis, 0.34 ± 0.92 mm (Figure 2). The absolute percentage volume difference for KIRI Engine models was 0.89 ± 0.68%, and the overall absolute dimensional error was 0.89 ± 0.51 mm. Surface point-to-point distances averaged 0.24 ± 0.19 mm, ranging 0.08–0.72 mm across casts (Table 2), with the largest values observed for cast 9 (0.72 mm) and cast 12 (0.54 mm), both of which were AFOs. Figure 2 shows that KIRI Engine, on average, overestimated the dimensional measurements, whereas the EinScan H2 underestimated them. Additionally, it shows that KIRI Engine typically underestimated the volumes of the models.

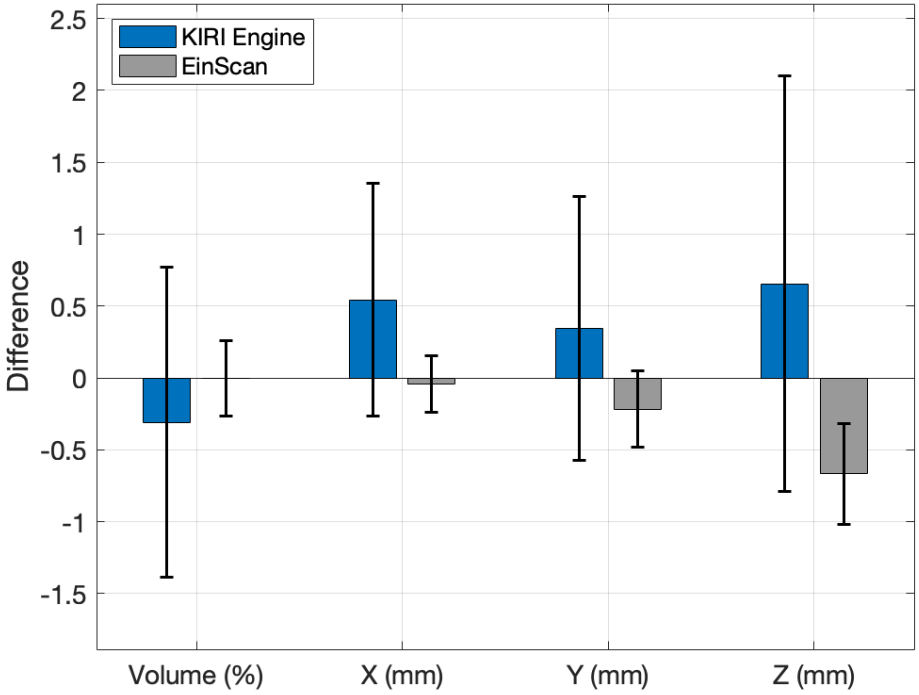



*Figure 2. Signed average volumetric (%) and dimensional (mm) differences for KIRI Engine relative to the EinScan H2 reference across all casts. Blue bars indicate KIRI Engine vs EinScan H2 differences, while grey bars reproduce the EinScan H2 vs known-volume surrogate differences from Figure 1 for comparison.*

*Table 2. KIRI Engine average point distances (mm) per cast*

| Cast | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Point differences | 0.13 | 0.10 | 0.20 | 0.27 | 0.08 | 0.13 | 0.27 | 0.21 | 0.72 | 0.10 | 0.13 | 0.54 |

As shown in Table 3, KIRI Engine had overall inter- and intra-rater reliability ICC values greater than 0.9997 for volume and all three dimension measurements. Furthermore, the 95% CIs ranged from 0.9990 to 1.0000. Across repeated scans, within-subject coefficients of variation (CV) were low; the worst-case CV was 1.72% for volume and 1.39% for linear dimensions, with most casts below 1%, indicating small trial-to-trial variability. Reliability was high across all measures: for volume, intra-rater ICC(2,1) values were 0.9999 for all three raters, and the inter-rater ICC was 1.0000. For the dimensions, intra-rater ICCs were all over 0.9998 (0.9990–1.000).

*Table 3. Inter- and intra-rater reliability of KIRI Engine volume measurements*

| | Rater | MD (SD) | ICC | 95% CI |
|---|---|---|---|---|
| Volume Measurements | | | | |
| Intra-rater | 1 | 0.7 (0.42) | 0.9999 | 0.9998 - 1.0000 |
| | 2 | 0.8 (0.59) | 0.9999 | 0.9998 - 1.0000 |
| | 3 | 0.8 (0.70) | 0.9999 | 0.9997 - 1.0000 |
| Inter-rater | 1 vs. 2 vs. 3 | 0.6 (0.50) | 1.0000 | 0.9999 - 1.0000 |
| X Dimension Measurements | | | | |
| Intra-rater | 1 | 0.6 (0.33) | 0.9999 | 0.9997 - 1.0000 |
| | 2 | 0.7 (0.64) | 0.9998 | 0.9994 - 0.9999 |
| | 3 | 0.8 (0.61) | 0.9998 | 0.9990 - 0.9999 |
| Inter-rater | 1 vs. 2 vs. 3 | 0.7 (0.54) | 1.0000 | 1.0000 |
| Y Dimension Measurements | | | | |
| Intra-rater | 1 | 0.4 (0.25) | 0.9999 | 0.9999 - 1.0000 |
| | 2 | 0.4 (0.34) | 0.9997 | 0.9994 - 0.9999 |
| | 3 | 0.4 (0.18) | 0.9999 | 0.9998 - 1.0000 |
| Inter-rater | 1 vs. 2 vs. 3 | 0.3 (0.27) | 1.0000 | 0.9999 - 1.0000 |
| Z Dimension Measurements | | | | |
| Intra-rater | 1 | 0.5 (0.47) | 0.9998 | 0.9995 - 0.9999 |
| | 2 | 0.6 (0.57) | 0.9999 | 0.9996 - 1.0000 |
| | 3 | 0.5 (0.38) | 0.9998 | 0.9995 - 0.9999 |
| Inter-rater | 1 vs. 2 vs. 3 | 0.5 (0.47) | 1.0000 | 1.0000 |

Note: MD = mean difference, SD = standard deviation, ICC = intra-class correlation, CI = confidence interval.

In the Bland–Altman plots (Figure 3), the center dashed line (mean difference) is slightly above zero for rater 1 vs 2 and for rater 1 vs 3, and slightly below zero for rater 3 vs 2. This indicates that rater 1 tended to report marginally larger volumes than raters 2 and 3, whereas rater 3 tended to report marginally smaller volumes than rater 2. The magnitude of these differences is small, with a bias of less than 1% of the mean volume for most casts, and remains within the 95% limits of agreement, aside from a single outlier (cast 8).

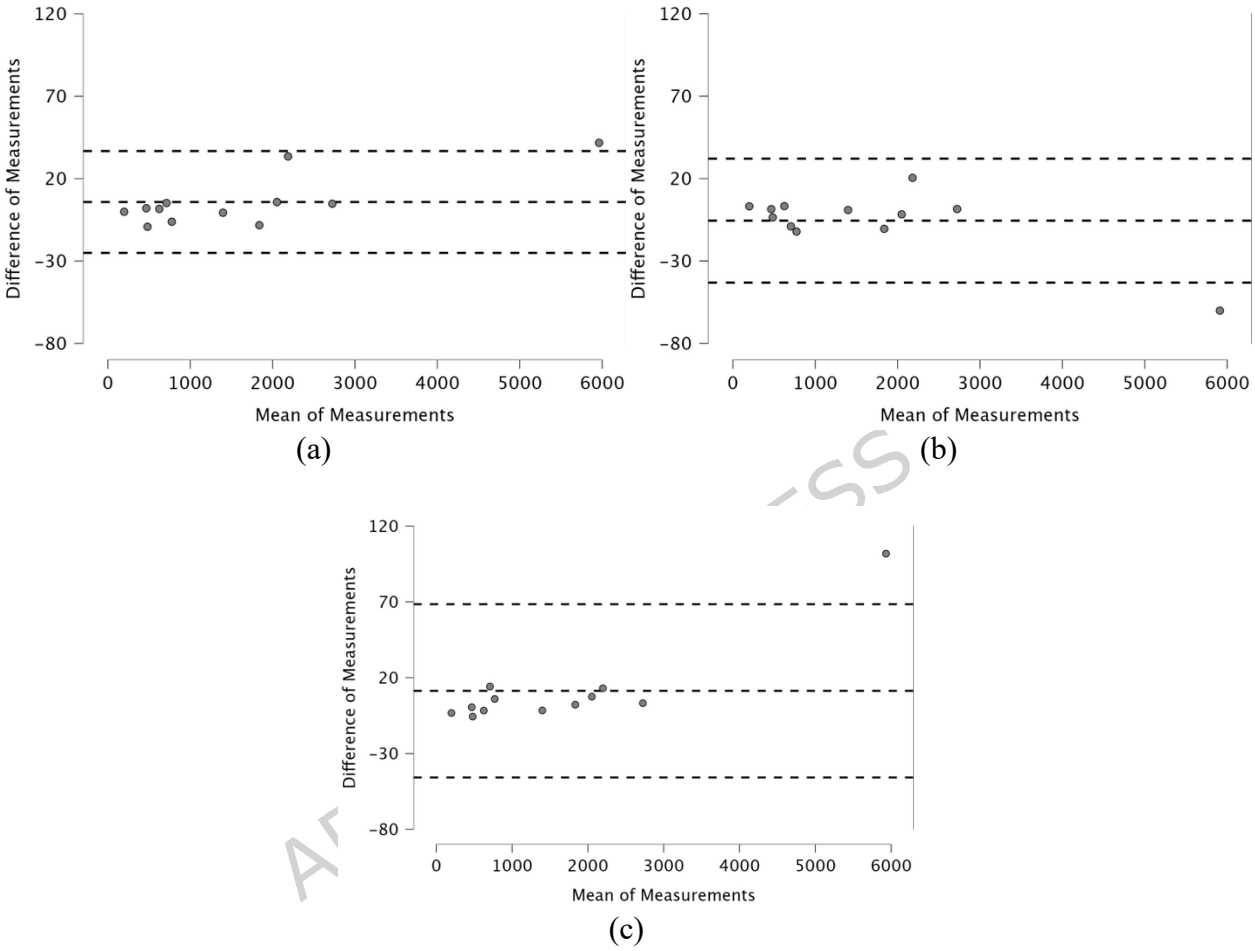



*Figure 3. Bland–Altman plots of mean volume difference: (a) rater 1 vs rater 2; (b) rater 3 vs rater 2; and (c) rater 1 vs rater 3*

## 2.3 Scanner Usability

Across all casts, KIRI Engine required 62 ± 13 photos per trial (mean ± SD): upper limb 55 ± 9, lower limb 60 ± 11, and AFO 73 ± 13. As shown in Figure 4, the capture time was 88 ± 17 s for KIRI Engine and 34 ± 17 s for the EinScan H2, while the processing time was 12 ± 2 min for KIRI Engine (cloud) and 85 ± 62 s for EinScan H2. There was no evidence of a linear or monotonic association between capture time or photo count and either volume error or dimensional error. Exploratory Spearman correlations were small (pooled $|\rho| \leq 0.25$) and simple linear fits explained little variance ($R^2 \leq 0.056$). Although the two correlations involving dimensional error reached statistical significance ($p = 0.013$ and $p = 0.009$), their

effect sizes were minimal ($\rho \approx 0.24$, $R^2 \approx 0.05$), indicating that these p-values likely reflect the larger pooled sample size rather than a meaningful relationship. Volume-error correlations were not significant. One participant showed a moderate correlation with dimensional error ($|\rho| \approx 0.42$), but this did not replicate across participants and did not appear in the pooled results (Figure 5). The average usability score (SUS) for KIRI Engine was 79 and was 58 for the EinScan H2 (Table 4. SUS scores for KIRI Engine and the EinScan H2).

*Table 4. SUS scores for KIRI Engine and the EinScan H2*

| | Rater 1 | Rater 2 | Rater 3 | Average |
|---|---|---|---|---|
| KIRI Engine | 90 | 82.5 | 65 | 79 |
| EinScan H2 | 55 | 60 | 60 | 58 |

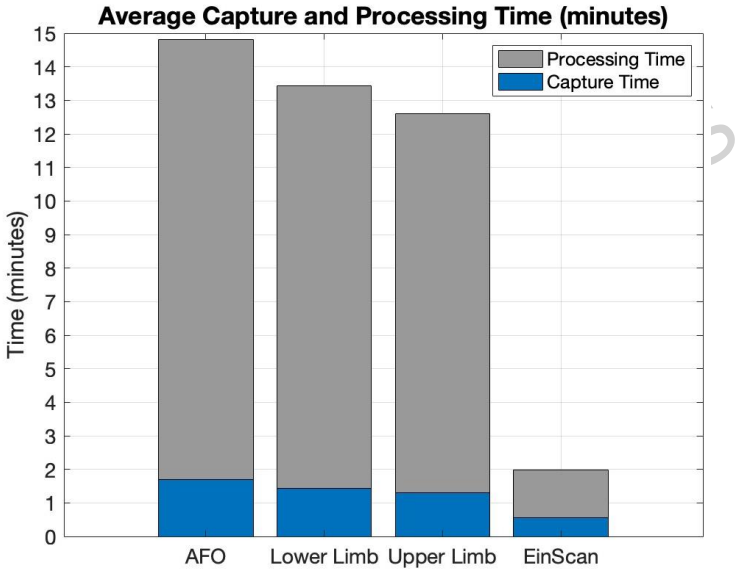



*Figure 4. Scanning and processing times per cast*

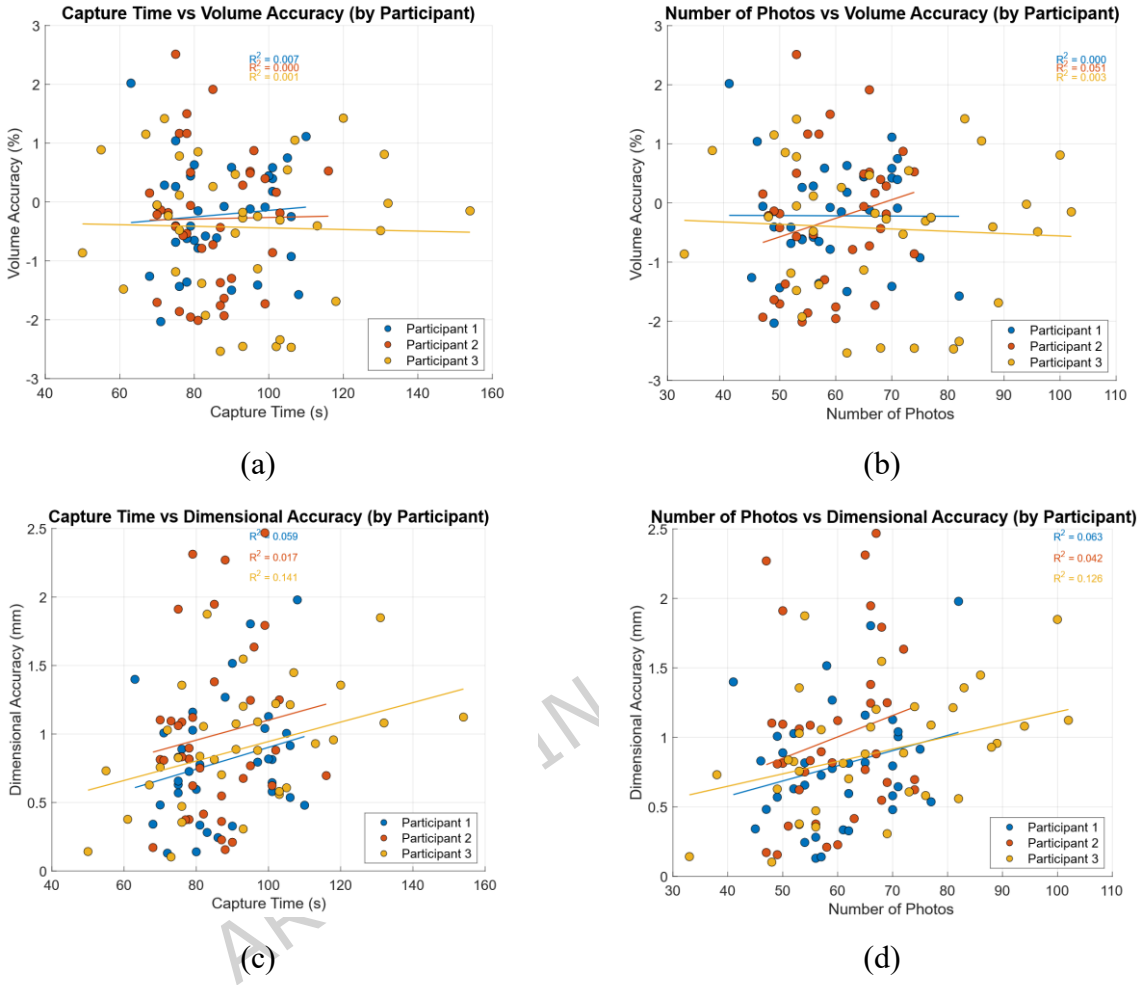



Figure 5. Graphs of volume accuracy versus (a) shape capture time and (b) number of photos; dimensional accuracy versus (c) shape capture time and (d) number of photos. Lines show least-squares fits (annotated $R^2$).

# 3. Discussion

Previous research on photogrammetry in P&O has primarily focused on small, case-study demonstrations in specific populations (particularly lower-limb applications), with limited sample sizes and mixed evidence on accuracy and reliability (5,13,22–24). To provide a broader and more actionable overview of digital scanning across P&O, this study (i) verified the accuracy and reliability of a clinical-grade structured-light scanner to serve as a gold-standard reference, (ii) validated the performance of a widely used smartphone photogrammetry app across a heterogeneous set of cast models in P&O, and (iii) assessed the usability.

## 3.1 EinScan H2 Accuracy and Reliability

The reference EinScan H2 had a mean absolute volumetric error of 0.21 ± 0.15% and a mean absolute dimensional error of 0.35 ± 0.18 mm, with excellent inter- and intra-rater ICCs (>0.9999) across three raters. One study on the EinScan H2 found ICCs >0.999 and volumetric/dimensional errors of approximately 2.5% on surrogate 3D-printed limb models (25). The ten-fold difference in volumetric error between the aforementioned study and ours can be attributed to methodological differences: the study by Cullen used the EinScan H2 in infrared mode (0.1 mm spec), and post-processing involved hole filling and mesh decimation, all of which inflate volumetric error. In contrast, our study matte-coated the known-volume model and used white-light mode (0.05 mm spec), with minimal post-processing before comparison to the digital reference.

Another study on lower-limb foam and plaster models found that the volumetric error between a structured-light scanner (Artec Eva; Artec 3D, California, USA) and high-precision laser scanner was 1.4%, the mean absolute dimensional error was 0.92 ± 0.13 mm, and the inter-and intra-rater ICCs were over 0.99 (11). Other studies on 15 live transradial residual limbs and 20 transtibial casts (3,26) found that the inter- and intra- rater ICCs were above 0.99 for the Spectra (Vorum, BC, Canada) and Go!SCAN (Creaform, Texas, USA) structured-light scanners, respectively, but focused mainly on the reliability of the scanners rather than their accuracy (6). In combination, these studies show that the performance of the EinScan H2 in this study falls well within these bounds and supports its use as a stable reference comparator for evaluating photogrammetry. Axis-wise differences observed in the results were sub-millimeter, but were largest along the Z-direction, likely indicating that operator-dependent trimming influences the error rather than scanner instability.

3.2 KIRI Engine Accuracy and Reliability

In our study, relative to the reference scanner, KIRI Engine achieved a mean absolute volume error of approximately 1% and mean point-to-point distances of 0.24 ± 0.19 mm across the whole mesh.

Our estimates align with previous research on photogrammetry. Cullen et al. (2023) reported a volumetric error of 0.87% and a segment-wise mean radial error (MRE) along the Z-axis of approximately 0.2 mm on lower-limb models; however, the MRE was reported only over the central region and does not include proximal/distal effects, whereas our ~0.24 mm point-to-point distances cover the whole mesh (5). Walters et al. (2024) found a volumetric difference of 1–2.9% and a point-to-point error of 1.99–2.36 mm across two photogrammetry methods versus a reference structured-light scanner in 10 residual limb models (13). Another study found that cast-based photogrammetry could achieve ~1% error in volumes/surface agreement, consistent with our findings. However, when extended to home settings without limb stabilization, accuracy decreased significantly, implying that our cast results should not be overgeneralized to live-limb workflows (25).

From a clinical perspective, these errors are small relative to existing sources of variation in socket and orthosis fabrication. Our ~1% volume error is within the “good” fit range (0–5%) (27) and below thresholds known to affect fit and comfort (>6%) (28). Further, work on cast rectification and fabrication consistency, including reports of inter- and intra-clinician variability, has shown that millimetre-level adjustments are common (3,29). Point-to-point distances among manually casted transtibial sockets produced by the same prosthetist have ranged from 2.79–4.41 mm (3), and inter-prosthetist variability from 1–3 mm (29), substantially larger than the sub-millimeter deviations observed here on casts. Taken together, these data define a conservative error budget in which our observed ~1% volumetric error and 0.24 ± 0.19 mm surface deviations occupy only a small fraction of the overall allowable variability, supporting the conclusion that, for rectified positive casts and other stable geometries, photogrammetry-induced error is unlikely to be the limiting factor for clinical fit.

Two outliers in point-to-point distances (12 and 9 (AFOs); Table 2) largely skewed the average. Both exhibited small holes in the meshes for one rater, which were filled in using meshing tools. These holes are likely attributed to self-occluding geometries (such as deep trimlines and undercuts), leading to insufficient photo coverage. Both casts were AFOs; this indicates that complex shapes with sharp curvature require careful capture protocols to avoid local deviations in geometry. Despite higher surface complexity, these localized deviations did not translate into substantial volumetric error, showing that global volume metrics remain stable even when pointwise surface error increases.

KIRI Engine had near-perfect reliability on casts (ICCs ≥ 0.9997; narrow 95% CIs), exceeding the X- and Y-dimension ICCs reported for *in-vivo* structured-light scans (ICC 0.918–0.99) and matching or surpassing their volume ICCs (0.998–0.999) (6). On casts, our results are an improvement on prior photogrammetry studies, which found two-rater ICCs of 0.81 (5). Bland–Altman analyses showed tight agreement with a single outlier (cast 8, which had the largest volume by threefold), consistent with observations that surface error tends to rise with limb volume (13).

Small, systematic inter-rater biases were also evident in the plots: rater 3 tended to underestimate relative to raters 1 and 2, and rater 1 slightly overestimated relative to the others. These patterns likely reflect subtle but consistent differences in capture technique, such as photo coverage and viewing distance, that translate into small, systematic surface reconstructions (30). The magnitude of these biases was small and did not change the overall conclusions on reliability.

### 3.3 Scanner Usability

KIRI Engine required ~60 photos and ~90 s per capture; EinScan H2 captures averaged ~35 s. Processing times were ~12 min (KIRI Engine, cloud) and ~1.5 min (EinScan H2). Despite the longer processing time, KIRI Engine scored higher on perceived usability (SUS ≈ 79, “Good”) than EinScan (SUS ≈ 58, “OK”) under the training and workflow conditions in this study (31). This is likely because, in our workflow, KIRI Engine used a familiar device (iPhone 11) with auto-capture, along with step-by-step prompts and minimal setup requirements. In contrast, the EinScan H2 necessitated wiring to a computer, maintaining appropriate distances from the object, and occasional rescans after tracking loss. These factors increase the perceived user complexity and the sense that technical support might be needed. However, with more extensive training and integration into routine workflows, this gap in usability could narrow. The cast type influenced the capture time: AFOs required more images due to overhangs and increased curvature, followed by lower-limb casts (relatively large) and upper-limb casts (smallest in size). Furthermore, the ~1-minute difference in capture time is unlikely to constrain routine clinical practice. However, during live-limb scans, longer capture windows may exacerbate motion/postural drift and patient burden, although neither capture time nor photo count was correlated with accuracy (Figure 5). These findings support the feasibility of smartphone photogrammetry in multi-operator workflows, particularly where hardware cost or portability matters.

Table 5 compares our results with those in recent reports: smartphone photogrammetry typically needs 2–5 min of capture, with optimized protocols reaching 1–2 min (5,13,25). Our ~1.5 min capture lies at the fast end of this range. Further, processing spans from 5–10 min to over 1 h; our ~12 min is an intermediate compromise: slower than on-device processing but far faster than the longer cloud workflows (13,25). Structured-light scanning remains faster end-to-end but requires dedicated hardware and positioning setups (6,25). Photo counts in optimized photogrammetry are lower (≈26–38), yet our data shows that higher image counts do not degrade accuracy (Figure 5), and the auto-capture feature of KIRI Engine makes the capture time equivalent even with increased photo counts.

*Table 5. Comparison of scanner usability across recent reports and this study*

| Study | Modality | Objects | No of photos | Capture time | Processing time | Notes |
|---|---|---|---|---|---|---|
| **This study** | Smartphone photogrammetry (KIRI Engine) vs structured-light scanner (EinScan H2) | 5 upper-limb + 3 lower-limb + 4 AFO casts | ~62 (KIRI) | KIRI: ~1.5 min<br>EinScan: ~0.5 min | KIRI: ~12 min (cloud)<br>EinScan: ~1.5 min | **SUS**<br>KIRI 79.2 ("Good")<br>EinScan 58.3 ("OK") |
| **Cullen et al. (2023)** | Smartphone photogrammetry (ReCap Photo) | 2 positive lower-limb/socket casts | ~26 (optimized method) | ~ 5 min | Tens of minutes to hours (cloud) | — |
| **Cullen et al. (2025)** | Smartphone photogrammetry (ReCap) vs structured-light scanner (EinScan H2) | 4 3D printed transtibial surrogate limbs | ~38 photos (optimized method) | 1–2 min for both methods | ReCap<br>~1 h 46 min<br>EinScan<br>4 min 35 sec | |
| **Ngan et al. (2022)** | Structured-light scanner (Vorum Spectra) | 15 transradial residual limbs | — | 0.5–1 min | *Not reported* | Dedicated positioning setup for high ICCs |
| **Walters et al. (2024)** | Smartphone photogrammetry (Polycam, Luma) vs structured-light scanner (Artec EVA) | 10 residual limbs (8 lower-limb & 2 upper-limb) | ~75 (optimized method) | 2–4 min | 5–10 min (both apps) | Did not measure usability for Artec EVA |

The study has a number of limitations. First, the accuracy and reliability were quantified on casts; caution should be exercised when extending these results to the scanning of actual residual limbs. At the same time, clinical practice recognizes that live-limb scanning may not be feasible or appropriate in all contexts. For many clinicians, physical casting remains the preferred way to assess soft-tissue behavior; in such situations, scanning the rectified positive cast preserves the benefits of hands-on casting while providing the advantages of digital manufacturing. Further, previous studies have shown reduced scanner performance when scanning limbs directly due to factors such as limb movement, hair and skin reflectance, soft-tissue deformation, and size (6).

More importantly, longer scan times increase the likelihood of postural drift, especially in children who may struggle to hold still. Clinical feasibility for live-limb scanning depends on minimizing capture time and formulating stabilization strategies (e.g., supports, seating, etc.), alongside standardized scanning and positioning protocols. Furthermore, appropriate consent, de-identification, and data-processing agreements with app vendors would be required for live-limb scanning. Secondly, this is a single-site, single-method study. It is important that future research focuses on multi-site data collection to capture a broader range of practices and abilities, as well as different phones and photogrammetry apps.

Methodologically, several factors could have introduced measurement error. Metric scaling in photogrammetry depends on a physical scale visible in the images; any occlusion or parallax would produce a multiplicative error that propagates to volume and all three dimensions. For larger objects, a longer, stiffer, and more conspicuous scale may be required to maintain accuracy. Further, operator-introduced variation along the cutting (Z) axis may affect comparisons; meshing and processing choices, as well as registration steps, can also influence surface quality.

Finally, several real-world factors not varied here could affect consistency: lighting (intensity, direction, specular glare), background (featureless vs patterned), phone/camera characteristics, and software differences (app/OS versions, reconstruction parameters, cloud vs on-device processing). Future work should systematically vary these factors and include multiple devices and sites.

# 4. Conclusion

This study benchmarked a clinically used structured-light scanner (EinScan H2) and evaluated a smartphone photogrammetry workflow across 12 P&O-relevant casts. The reference scanner demonstrated very high accuracy (absolute volume error 0.21 ± 0.15%; absolute dimensional error 0.35 ± 0.18 mm), supporting its role as a stable comparator. Relative to this reference, smartphone photogrammetry achieved ~1% absolute volume error, mean surface point-to-point distances of 0.24 ± 0.19 mm, and near-perfect inter-/intra-rater reliability, indicating that, on casts, it can produce geometries suitable for downstream rectification and fabrication.

From a practical standpoint, the photogrammetry workflow required ~60 photos and ~90 s of capture (≈12 min processing, cloud), scored "Good" on SUS, and showed no correlation between capture time/photo count and accuracy. These features are conducive to multi-operator use, especially where portability and cost matter. However, time sensitivity will be greater for live-limb scans, where even modestly longer capture windows can increase motion/postural drift and patient burden; future work should quantify acceptable capture-time thresholds and evaluate stabilization strategies in clinical settings.

In a broader context, prior work has established the advantages of end-to-end digital workflows in P&O. Our results add to this evidence by demonstrating that smartphone photogrammetry is a cost-effective, highly accessible approach to digital scanning—the crucial first step in digital workflows—that can lower barriers to wider adoption. Our findings support integrating smartphone photogrammetry into P&O practice, especially for scanning rectified positive casts and other stable geometries. Further multi-site, live-limb studies across different devices and apps are critical to establish best-practice protocols for routine care. In the interim, clinicians can consider smartphone photogrammetry as a practical, low-cost option for digitizing rectified positive casts and other stable geometries, particularly in settings where access to dedicated 3D scanning hardware is limited.



# 5. Methods

This section describes the shape-capture setup, instrumentation, scan objects, and workflows used to implement the three study phases: benchmarking accuracy and reliability using a clinical-grade scanner, evaluating the performance of photogrammetry, and assessing the usability of both systems.

## 5.1 Shape Capture

### 5.1.1 Setup

The shape-capture environment was standardized to minimize reflections, shadows, and occlusion while still allowing full 360° coverage of each object. For photogrammetry (Figure 6a), each object was mounted on a stand placed on a stool in a well-lit location, without any direct sunlight. A similar scanning setup was used for the structured-light scanner (Figure 6b).

(a) (b)

*Figure 6. Setup for (a) photogrammetry shape capture and (b) structured-light shape capture*



5.1.2 Instruments

A clinically used structured-light scanner, the EinScan H2 (Shining 3D, Hangzhou, China), was used as the reference device. The scanner has a resolution of 5 MP and an accuracy of 0.05 mm in white-light mode (32). Scans were processed in the EXScan H software (Shining 3D, Hangzhou, China); the user interface is shown in Figure 7. For photogrammetry, the standard 1× rear camera of an iPhone 11 (Apple Inc., Cupertino, USA) was used, with a resolution of 12 MP (33). Image capture, upload, and mesh generation were performed in the KIRI Engine 4.0.4 (KIRI Innovation, Shenzhen, China) mobile application, with mesh reconstruction executed on KIRI's cloud servers (34); the user interface is shown in Figure 8. Camera intrinsics and lens distortion were handled by KIRI Engine's internal calibration process. KIRI Engine was selected because image capture, upload, and processing (cloud) can be performed entirely on a smartphone, unlike other tools requiring desktop processing.

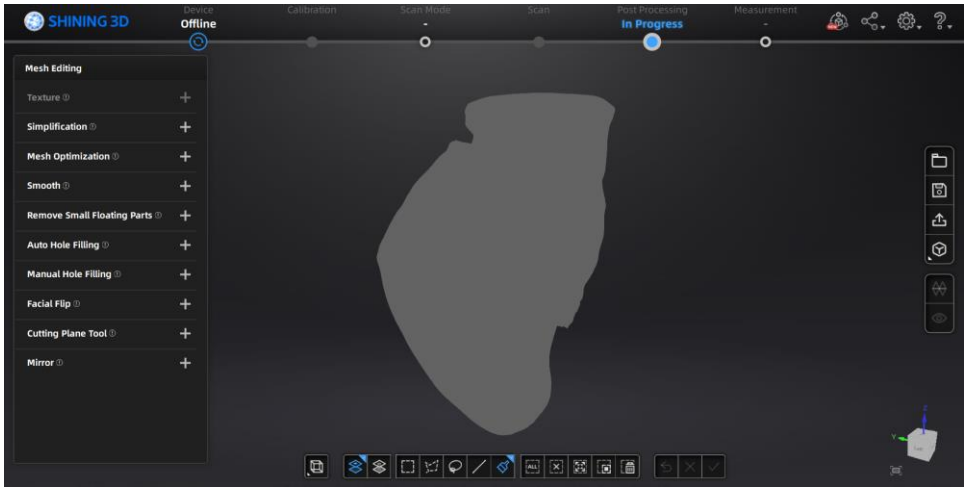



*Figure 7. Post-processing using the ExScan H software*

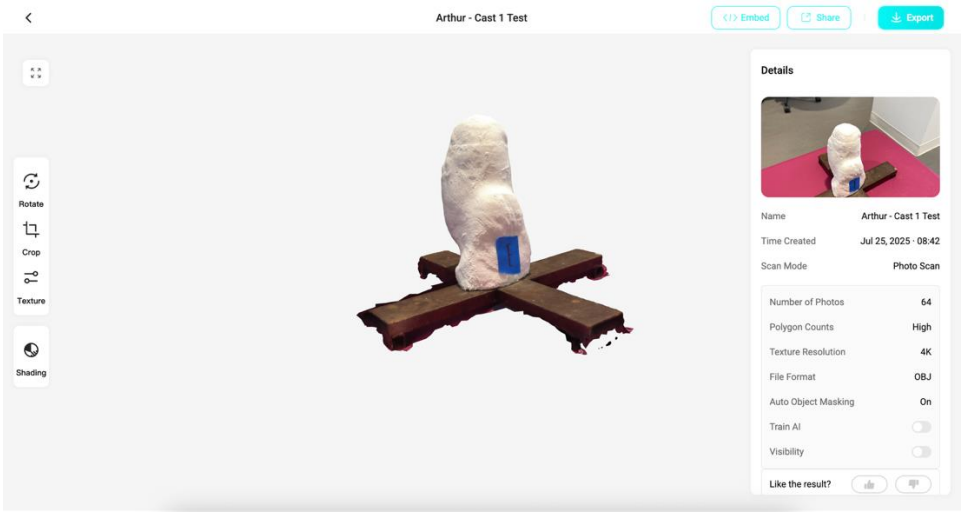



*Figure 8. KIRI Engine post-processing interface*

5.1.3 Shape Capture Procedure

In phase 1, we evaluated the accuracy of the EinScan H2 by scanning a 3D-printed lower-limb cast surrogate with known volume (Figure 9), calculated from the "STL" file used for printing. The object was printed in PETG on the Original Prusa XL (Prusa Research, Prague, Czech Republic), which can print to a resolution of 0.05 mm (35).

*Figure 9. 3D-printed lower-limb cast surrogate*

The 3D-printed surrogate was coated with matte spray to suppress reflections. The three raters scanned the object in white-light mode at the highest resolution available (0.05 mm). Raters were instructed to circle the object until it was fully captured and visible on the screen. Scanning and processing times were recorded; setup and calibration times were excluded as these can vary between clinical workflows and clinician preferences. The raters had basic previous experience in 3D scanning, received no formal training, and completed approximately 15 minutes of familiarization per device before data collection.

For the EinScan H2, the familiarization process included one trial scan of a non-study cast in white-light mode to help assess tracking stability, coverage, and scanner distance. For KIRI Engine, raters captured a full image set for a non-study cast using the auto-capture mode and the recommended multi-angle orbit

around the object. A second non-study cast was used if the rater was unable to complete the procedure without assistance.

In phase 2, we compared KIRI Engine to the reference EinScan H2 using 12 anonymized upper limb, lower limb, and AFO casts (Table 6). To ensure a statistical power of 0.9 ($\alpha = 0.05$, $w = 0.2$), we used a sample size of 12 ($n = 12$) casts and three raters ($k = 3$) (36). The cast volumes ranged from 200 mL to 6,000 mL (Table 7), spanning a wider range than prior P&O scanning studies (11). Selection emphasized surface features widespread in clinical practice and known to affect photogrammetry (encircled in red in Table 6): fine detail (upper limb); smooth, large areas (lower limb); and overhangs (AFO).

*Table 6. Pictures of the cast models used in this study. Examples of fine detail; smooth, large areas; and overhangs have been encircled in red*

| Type | Casts |
|---|---|
| Upper Limb socket |  |
| Lower Limb socket |  |
| AFO |  |

*Table 7. Volumes (mL) and dimensions (mm) of casts (calculated from the EinScan H2 meshes)*

| Casts | Type | Volume (mL) | X-axis (M-L) (mm) | Y-axis (A-P) (mm) | Z-axis (P-D) (mm) |
|---|---|---|---|---|---|
| 1 | TR | 482.53 | 79.27 | 89.73 | 147.75 |
| 2 | TR | 469.68 | 91.04 | 86.00 | 144.52 |
| 3 | TR | 764.60 | 93.36 | 99.71 | 156.10 |
| 4 | TR | 199.99 | 50.41 | 54.27 | 129.43 |
| 5 | TR | 628.78 | 74.82 | 66.95 | 278.24 |
| 6 | TT | 1860.36 | 120.31 | 161.56 | 252.84 |

| | | | | | |
|---|---|---|---|---|---|
| 7 | TT | 2207.54 | 121.31 | 142.79 | 238.95 |
| 8 | TF | 6029.60 | 219.62 | 210.41 | 306.41 |
| 9 | AFO | 710.64 | 68.78 | 196.54 | 247.56 |
| 10 | AFO | 2053.44 | 88.72 | 261.66 | 356.69 |
| 11 | AFO | 2716.03 | 95.34 | 252.85 | 352.01 |
| 12 | AFO | 1391.51 | 80.09 | 229.14 | 327.12 |

Note: TR = transradial (below-the-elbow), TT = transtibial (below-the-knee), TF = transfemoral (above-knee), AFO = (ankle–foot orthosis). Clinically, the X-, Y-, and Z-axes correspond to the medial–lateral (M-L), anterior–posterior (A-P), and proximal–distal (P-D) directions (Figure 11b).

All casts were scanned once by each rater with the EinScan H2 to create reference models (12 trials per rater for reference scanning). For photogrammetry, a 3-cm adhesive scale (blue tape with 0.5-cm black dot increments) was attached to a flat region of each cast, avoiding under-overhang areas, to permit metric scaling (Figure 10). Prior to scanning the casts, each rater became familiarized with KIRI Engine and the EinScan H2 by using each system for at least 15 minutes. Each cast was then captured three times by each rater, for a total of 36 trials per rater for photogrammetric scanning.

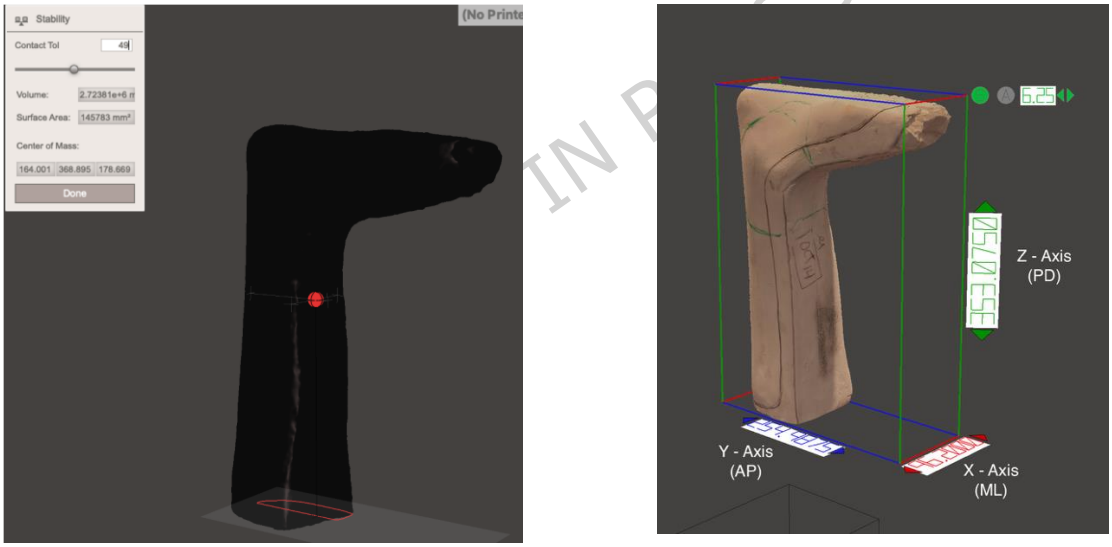
*Figure 10. Physical 3-cm scale*

Photogrammetry image capture followed common guidelines: raters orbited the cast in concentric paths, taking images at four viewing angles: 0° (level with the cast), 30°, 60°, and 90° (overhead view) (22–24). Distance to the object was not fixed, but raters were instructed to keep the cast centered and occupying most of the frame (37). For AFO casts, additional photos were taken beneath the overhang. KIRI Engine's auto-capture mode was used throughout, and the cast order was randomized. After upload, the highest polygon count and texture resolution settings were selected for reconstruction. No mesh decimation options were enabled, so the exported meshes retained the full resolution of the reconstruction. Meshes were exported in OBJ format. For every trial, the number of photos, capture time, and processing time were recorded.

## 5.2 Data Processing

First, all of the models were trimmed in MeshMixer 3.5 (Autodesk, Inc., San Rafael, CA, USA) along the Z-axis to remove all non-anatomical material, such as the mandrel and excess plaster, maintaining the volume of interest. Small non-physical gaps at open boundaries were closed using the standard hole-filling tool when required. No global smoothing, denoising, or remeshing operations were applied beyond the default behavior of the respective software.

Next, photogrammetry meshes were scaled using the 3-cm tape scale attached to the casts. Reference EinScan H2 models were aligned such that the base of each model was flat on the X-Y plane in Canfit O&P CAD Software (Vorum, Vancouver, BC, Canada), and the photogrammetry models were aligned to these reference models using the iterative closest points algorithm in CloudCompare. The volumes of the models were then recorded (Figure 11a) and the X, Y, and Z dimensions of the bounding box were extracted (Figure 11b). Clinically, the X-, Y-, and Z-axes correspond to the medial–lateral, anterior–posterior, and proximal–distal directions (6). All reference volumes and linear dimensions were extracted digitally from the EinScan H2 meshes to avoid human error in caliper placement and reading variability. The EinScan H2 has been independently reported to achieve sub-millimetre precision for surface capture (38,39), consistent with our own benchmark results (see Section 2.1).




(a) (b)

*Figure 11. Extraction of the (a) volume and (b) bounding-box dimension measurements*

## 5.3 Analysis

### 5.3.1 EinScan H2 Accuracy and Reliability

Volumetric and dimensional accuracy were quantified against the known-volume model. For each rater, the percentage volume differences and the dimensional errors for each axis (mm) were computed and then averaged across raters. Additionally, the overall dimensional error was computed by averaging the X, Y, and Z dimension differences. Intra- and inter-rater reliability were assessed using a two-way random-effects intraclass correlation, single-measure form (ICC(2,1)) with 95% confidence intervals, computed in JASP

(University of Amsterdam, version 0.18.3) (6). The choice of model and reporting of reliability indices were informed by the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (40). Summary plots report the overall means and standard deviations; no smoothing or additional post-processing was applied beyond the standardized trimming and alignment described in Section 5.2.

5.3.2 KIRI Engine Accuracy and Reliability

Photogrammetry accuracy was evaluated against the EinScan H2 reference for each cast. Metrics included percentage volume difference, dimensional error (X, Y, Z; mm), and mean unsigned surface point-to-point distances (the Euclidean distance from each vertex of the photogrammetry mesh to the nearest point on the aligned reference surface; mm) to capture overall surface deviations (7,30). Reliability was tested using the same method as that used for the EinScan H2. Bland–Altman plots (mean difference and 95% limits of agreement) were generated for volume for each rater pair to visualize bias and dispersion.

5.3.3 Scanner Usability

Workflow feasibility was summarized using the photo count, capture time, and processing time for each device, averaged across casts and raters. Perceived usability was measured using the System Usability Scale (SUS), a 10-question, 5-point Likert scale widely used in industry and research (31). To test whether workload affected performance, the average capture time and processing time of KIRI Engine for each cast type (i.e. upper limb, lower limb, and AFO), and the capture time and number of required photos versus the volume and dimension accuracy were measured and plotted.

5.3.4 Statistical Analysis

Descriptive statistics are reported as mean ± SD. Accuracy metrics include signed and absolute percentage volume differences and axis-wise dimensional errors (mm). Reliability was quantified using a two-way random-effects, absolute-agreement, single-measure ICC (ICC(2,1)) with 95% CIs**.** Bland–Altman analysis (mean bias; 95% limits of agreement) was used to visualize agreement for volumes between rater pairs. Analyses were performed in JASP (University of Amsterdam; version 0.18.3).



**Declarations**

Ethics approval: The study protocol was approved by the Research Ethics Board at Holland Bloorview Kids Rehabilitation Hospital (File #19-864; approved on 21 January 2020).

Consent for publication: *Not applicable.*

Data availability: Data cannot be shared publicly due to their potentially identifiable nature. Data are available through the Holland Bloorview Kids Rehabilitation Hospital Institutional Data Access/Ethics Committee (contact via Deryk Beal, Research Ethics Board Chair, Tel: (416) 425-6220, ext 3582, e-mail: dbeal@hollandbloorview.ca) for researchers who meet the criteria for access to confidential data.

Competing interests: The authors declare no competing interests.

Funding: This research was supported by funding from the Ontario Graduate Scholarship and the Kimel Family Fund at Holland Bloorview Kids Rehabilitation Hospital.

Authors' contributions: AE and VP contributed equally to this study. AE, VP, CCN, and JA collectively conceptualized the research. AE and VP curated the research data. AE, VP, and CCN developed the research methodology. AE, VP, CCN, and JA performed data analysis, and AE created the visualizations. AE and VP drafted the manuscript, while AE, VP, CCN, and JA collectively reviewed and edited the text. The project was supervised by JA. All authors approved the final manuscript.

Acknowledgements: The authors would like to thank Justin Fung for his help with data collection. We would also like to thank the clinicians and technicians at Holland Bloorview for their assistance with obtaining the data.

# References

1. Wang M, Nong Q, Liu Y, Lu H. Design of lower limb prosthetic sockets: a review. Expert Rev Med Devices. 2021;19(1):63–73.

2. Suyi Yang E, Aslani N, McGarry A. Influences and trends of various shape-capture methods on outcomes in trans-tibial prosthetics. Prosthet Orthot Int. 2019 Oct;43(5):540–55.

3. Dickinson AS, Donovan-Hall MK, Kheng S, Bou K, Tech A, Steer JW, et al. Selecting appropriate 3D scanning technologies for prosthetic socket design and transtibial residual limb shape characterization. JPO J Prosthet Orthot. 2022 Jan;34(1):33–43.

4. Ngan CC, Pendse V, Sivasambu H, Ouellette E, Ready N, Andrysek J. Preliminary characterization of rectification for transradial prosthetic sockets. Sci Rep. 2024 Mar 8;14(1):5759.

5. Cullen S, Mackay R, Mohagheghi A, Du X. Low-cost smartphone photogrammetry accurately digitises positive socket and limb casts. Prosthesis. 2023 Dec 18;5(4):1382–92.

6. Ngan CC, Sivasambu H, Ramdial S, Andrysek J. Evaluating the reliability of a shape capturing process for transradial residual limb using a non-contact scanner. Sensors. 2022 Sep 10;22(18):6863.

7. Cabrera IA, Pike TC, McKittrick JM, Meyers MA, Rao RR, Lin AY. Digital healthcare technologies: Modern tools to transform prosthetic care. Expert Rev Med Devices. 2021;18:129–44.

8. Sanders JE, Severance MR, Allyn KJ. Computer-socket manufacturing error: How much before it is clinically apparent? J Rehabil Res Dev. 2012;49(4):567.

9. Kofman R, Beekman AM, Emmelot CH, Geertzen JHB, Dijkstra PU. Measurement properties and usability of non-contact scanners for measuring transtibial residual limb volume. Prosthet Orthot Int. 2018 Jun;42(3):280–7.

10. Armitage L, Kwah LK, Kark L. Reliability and validity of the iSense optical scanner for measuring volume of transtibial residual limb models. Prosthet Orthot Int. 2019 Apr;43(2):213–20.

11. Seminati E, Canepa Talamas D, Young M, Twiste M, Dhokia V, Bilzon JLJ. Validity and reliability of a novel 3D scanner for assessment of the shape and volume of amputees' residual limb models. PLoS One. 2017 Sep 8;12(9):e0184498.

12. Ngan CC, Sivasambu H, Kelland K, Ramdial S, Andrysek J. Understanding the adoption of digital workflows in orthotic and prosthetic practice from practitioner perspectives: a qualitative descriptive study. Prosthet Orthot Int. 2022 Jun 1;46(3):282–9.

13. Walters S, Metcalfe B, Twiste M, Seminati E, Bailey NY. Smartphone scanning is a reliable and accurate alternative to contemporary residual limb measurement techniques. PLoS One. 2024 Dec 2;19(12):e0313542.

14. Mbithi F, Donovan-Hall M, Bramley J, Steer J, Rossides C, Worsley P, et al. Evidence-generated sockets for transtibial prosthetic limbs compared with conventional computer-aided designs: a multiple-methods study from the patient's perspective. JMIR Rehabil Assist Technol. 2025 Aug 21;12:e69962–e69962.

15. Xu R, Wang Z, Ma T, Ren Z, Jin H. Effect of 3D printing individualized ankle-foot orthosis on plantar biomechanics and pain in patients with plantar fasciitis: a randomized controlled trial. Med Sci Monit. 2019 Feb 21;25:1392–400.

16. Wojciechowski EA, Cheng TL, Hogan SM, Mudge AJ, Balassone D, Menezes MP, et al. Replicating and redesigning ankle-foot orthoses with 3D printing for children with Charcot–Marie–Tooth disease. Gait Posture. 2022 Jul;96:73–80.

17. Kim S, Yalla S, Shetty S, Rosenblatt NJ. 3D printed transtibial prosthetic sockets: a systematic review. PLoS One. 2022 Oct 10;17(10):e0275161.

18. Begasse T. VA Puget Sound Veteran receives first personalized 3D printed prosthetic socket. U.S. Department of Veterans Affairs; 2025. [Webpage]. URL: https://www.va.gov/puget-sound-health-care/stories/va-puget-sound-veteran-receives-first-personalized-3d-printed-prosthetic-socket. Accessed September 24, 2025.

19. Li G, Gao F, Wang Z. A photogrammetry-based system for 3D surface reconstruction of prosthetics and orthotics. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2011. p. 8459–62.

20. Ciobanu O, Ciobanu G, Rotariu M. Photogrammetric scanning technique and rapid prototyping used for prostheses and orthoses fabrication. Appl Mech Mater. 2013 Aug 30;371:230–4.

21. Grazioso S, Selvaggio M, Caporaso T, Di Gironimo G. A digital photogrammetric method to enhance the fabrication of custom-made spinal orthoses. JPO J Prosthet Orthot. 2019 Apr;31(2):133–9.

22. Cullen S, Mackay R, Mohagheghi A, Du X. The use of smartphone photogrammetry to digitise transtibial sockets: optimisation of method and quantitative evaluation of suitability. Sensors. 2021 Dec 16;21(24):8405.

23. Ismail R, Taqriban RB, Ariyanto M, Atmaja AT, Sugiyanto, Caesarendra W, et al. Affordable and faster transradial prosthetic socket production using photogrammetry and 3D printing. Electronics (Basel). 2020 Sep 7;9(9):1456.

24. Hernandez A, Lemaire E. A smartphone photogrammetry method for digitizing prosthetic socket interiors. Prosthet Orthot Int. 2017 Apr;41(2):210–4.

25. Cullen S, Mohagheghi A, Mackay R. Investigating the accuracy of smartphone photogrammetry for remote 3D scanning transtibial amputees. J Med Eng Technol. 2025 Aug 10;1–11.

26. Dickinson AS, Steer JW, Woods CJ, Worsley PR. Registering a methodology for imaging and analysis of residual-limb shape after transtibial amputation. J Rehabil Res Dev. 2016;53(2):207–18.

27. Fernie GR, Holliday PJ. Volume fluctuations in the residual limbs of lower limb amputees. Arch Phys Med Rehabil. 1982 Apr;63(4):162–5.

28. Sanders JE, Youngblood RT, Hafner BJ, Cagle JC, McLean JB, Redd CB, et al. Effects of socket size on metrics of socket fit in trans-tibial prosthesis users. Med Eng Phys. 2017 Jun;44:32–43.

29. Convery P, Buis AWP, Wilkie R, Sockalingam S, Blair A, McHugh B. Measurement of the consistency of patellar-tendon-bearing cast rectification. Prosthet Orthot Int. 2003 Dec;27(3):207–13.

30. Karantanellis E, Arav R, Dille A, Lippl S, Marsy G, Torresani L, et al. Evaluating the quality of photogrammetric point-clouds in challenging geo-environments – a case study in an alpine valley. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2020 Aug 14;XLIII-B2-2020:1099–105.

31. Brooke J. A "quick and dirty" usability scale. In: Jordan P, Thomas B, Weerdmeester B, McClelland I, editors. Usability evaluation in industry. London: Taylor & Francis; 1996. p. 189–94.

32. Shining 3D. EinScan H2 – SPECS. [Webpage]. URL: https://www.einscan.com/handheld-3d-scanner/einscan-h/einscan-h2-specs/. Accessed September 24, 2025.

33. Apple. iPhone 11 – technical specifications. [Webpage]. URL: https://support.apple.com/en-ca/111865. Accessed September 24, 2025.

34. KIRI Innovation. KIRI Engine app. [Webpage]. URL: https://www.kiriengine.app/. Accessed September 24, 2025.

35. Prusa Research. Original Prusa XL – 3D printer technical specifications. [Webpage]. URL: https://www.prusa3d.com/product/original-prusa-xl-semi-assembled-single-toolhead-3d-printer/. Accessed September 24, 2025.

36. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med. 2002 May 15;21(9):1331–5.

37. Taqriban RB, Ismail R, Ariyanto M, Yaya Syah Putra AF. 3D model of photogrammetry technique for transtibial prosthetic socket design development. In: 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). IEEE; 2019. p. 456–61.

38. Pellitteri F, Calza M, Baldi G, De Maio M, Lombardo L. Reproducibility and accuracy of two facial scanners: a 3D in vivo study. Appl Sci. 2025 Jan 24;15(3):1191.

39. Tangthaweesuk N, Raocharernporn S. The accuracy of three-dimensional facial scans obtained from three different 3D scanners. PLoS One. 2025 May 23;20(5):e0322358.

40. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS). J Clin Epidemiol. 2011 Jan;64(1):96–106.