Evaluating iPhone-Based 3D Scanning Applications for Heritage Documentation: Controlled Experiments and Future Directions

Abdelrahman A. Abdelghany[1, *], Derek. D. Lichti [1], Peter Dawson [2], Shabnam Jabari [3]

[1] Dept. of Geomatics Engineering, University of Calgary, Calgary, AB, Canada - (abdelrahman.abdelgha@ucalgary.ca , ddlichti@ucalgary.ca)

[2] Dept. of Anthropology and Archaeology, University of Calgary, Calgary, AB, Canada, (pcdawson@ucalgary.ca)

[3] Dept. of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, NB, Canada (sh.jabari@unb.ca)

Keywords: 3D Scanning, Mobile LiDAR, Heritage Preservation, iPhone Applications, Controlled Evaluation, Digital Twins

Abstract:

Remote heritage cabins in Canada's national parks require regular dimensional recording to support maintenance and conservation, yet professional laser scanners remain impractical for routine backcountry use due to cost and logistics. LiDAR-equipped iPhones offer a low cost, lightweight alternative, but their geometric reliability across different applications is not well understood. This study evaluates five iPhone-based 3D scanning applications — Polycam, PIX4Dcatch, KIRI Engine, Modelar, and UC3D (a research prototype developed at the University of Calgary) — using controlled indoor and real-world outdoor tests. In an indoor calibration field, 253 coded targets surveyed with a Trimble X9 terrestrial laser scanner serve as reference. Each application was used to capture three independent scans to assess accuracy and repeatability. Outdoor validation was conducted on two heritage log cabins in Yoho National Park using a GeoSLAM ZEB Horizon reference. A semi-automated pipeline was developed to extract targets and compute positional accuracy, inter-target distance error, and cross-surface consistency. Results indicate that UC3D achieved the lowest positional RMSE (49.8 mm) and demonstrated stable geometric consistency across repeated scans. PIX4Dcatch showed strong performance in within-surface distance accuracy but exhibited reduced consistency across surfaces. The remaining applications produced comparable reconstructions, with increased levels of geometric distortion observed particularly between surfaces. Variations in performance are likely linked to differences in reconstruction strategies and sensor fusion approaches. The results demonstrate that smartphone LiDAR can achieve centimetre-level accuracy for heritage documentation, though performance varies significantly between applications.

1. Introduction

Thousands of historic structures dot Canada's national parks — log cabins, warden patrol shelters, and backcountry lodges, many of them more than a century old. Keeping dimensional records of these buildings matters for conservation planning, structural monitoring, and creating digital archives before they are lost. The 2024 Jasper wildfire destroyed several heritage buildings overnight, a stark reminder that wooden structures in fire-prone forests can vanish between survey campaigns. What park managers need are updateable Digital Twins: three-dimensional digital copies of each cabin that can be refreshed annually by patrol officers carrying nothing more than a phone and a tripod (Jouan and Hallot, 2020).

The problem is that professional survey tools and heritage recording workflows were not designed with this use case in mind. Terrestrial laser scanners like the Trimble X9 achieve millimetre precision, but a single instrument costs more than most park budgets allow, weighs upward of 5 kg before batteries and tripods, and typically requires a trained operator and multi-day field campaigns to deploy at backcountry sites (Vosselman and Maas, 2010; Boardman and Bryan, 2018; Heritage and Large, 2009). Helicoptering a TLS crew to a remote cabin for an annual check is simply not practical at scale.

Structure-from-Motion photogrammetry went partway toward solving this. By extracting geometry from overlapping photographs, SfM cut the equipment cost to nearly zero (Westoby et al., 2012). Researchers showed it could produce sub-centimetre models of heritage facades under good conditions (Dhonju et al., 2017; Murtiyoso et al., 2018). But SfM has well-known blind spots: surfaces with little visual texture, shiny or wet materials, and the low light beneath a forest canopy all cause matching failures (James and Robson, 2012). For the kind of rough-sawn log walls found in national park cabins, these limitations are not edge cases — they are the norm.

When Apple released a direct time-of-flight LiDAR sensor on the iPad Pro in 2020 and the iPhone 12 Pro later that year, the landscape shifted. Suddenly a consumer device that fits in a jacket pocket could capture both depth and colour at walking speed. Early testing was encouraging. Luetzenburg et al. (2021) measured decimetre-level accuracy to 5 m range with the iPhone 12 Pro, noting systematic bias that increased with distance.Teppati Losè et al. (2022) found centimetre accuracy for small heritage objects but cautioned that drift grew quickly over larger volumes. Gollob et al. (2021) tried iPad LiDAR for forest inventory and reported usable tree diameter estimates, though dense canopy confused the sensor.

Several studies have evaluated the sensor for building documentation. Paukkonen (2023) paired iPhone LiDAR with video photogrammetry and reported centimetre accuracy for rapid building dimension checks. Askar and Sternberg (2023) compared several iPhone 13 Pro scanning apps on building facades and found that performance varied a lot depending on which app was used and how the user moved the phone. Spreafico et al. (2021) tested iPad LiDAR for indoor heritage mapping and confirmed its speed advantage, while also calling for more controlled testing. Murtiyoso and Grussenmeyer (2021) showed that simply recording video from a smartphone and processing it photogrammetrically could produce heritage models of reasonable geometric quality. Nocerino et al. (2017)

* Corresponding author

had earlier established a baseline for what smartphone cameras could do photogrammetrically. Sammartano and Spanò (2018) explored SLAM-based rapid mapping for heritage at risk and found that combining active and passive sensors yielded complementary advantages. More recently, Costantino et al. (2022) assessed smartphone LiDAR for urban surveying, concluding that the processing pipeline matters as much as the hardware. Spreafico et al. (2021) investigated iPad LiDAR for quick cultural heritage recording, confirming that while speed is a clear win, geometric quality still lags behind dedicated photogrammetric workflows. Tanduo et al. (2022) benchmarked iPhone LiDAR against TLS for archaeological documentation and found centimetre accuracy within the sensor's comfortable range, degrading rapidly beyond 5 m.

Despite all this work, a few things remain unclear. Most published studies test only one or two apps at a time, making it hard to compare results across studies because every team uses a different test room, different reference scanner, and different evaluation criteria. Very few studies use a proper indoor calibration field with pre-surveyed targets — instead they rely on cloud-to-cloud comparisons that mix registration noise with reconstruction error and cannot isolate geometric distortion.

The concept of inter-target distance accuracy, which tests how accurately the model represents shape independent of the registration, has barely been explored. And nobody, to our knowledge, has put a purpose-built research scanning app head-to-head with commercial products under the same controlled conditions to see where the commercial tools fall short and why. There is also a question of repeatability. If a park officer scans the same cabin three visits in a row, will the three models agree with each other closely enough to detect a leaning wall or a sagging roof beam This matters at least as much as absolute accuracy for a monitoring programme, and it has received almost no attention in the smartphone LiDAR literature.

This study addresses these gaps by evaluating five iPhone-based scanning applications — four commercial and one research prototype — under controlled indoor conditions. Field validation was subsequently carried out on two heritage log cabins in Yoho National Park using a subset of the applications (Polycam and UC3D).

The evaluation goes beyond the usual cloud-to-cloud comparison. We developed a semi-automated target extraction pipeline that lets us compute target-based positional accuracy, within-surface distance accuracy (local shape fidelity), and cross-wall distance accuracy (global geometric consistency) the last of which turns out to be the most revealing metric. We also examine point density, ICP registration quality, and practical factors like processing stability and capture effort.

The paper is organised as follows. Section 2 describes the test environments, equipment, scanning protocol, target extraction pipeline, and accuracy metrics. Section 3 presents results for both the indoor and outdoor experiments, including the repeatability analysis. Section 4 discusses what the results mean for heritage documentation practice, and Section 5 wraps up with practical recommendations and directions for future work.

2. Methodology

2.1 Study Sites

Indoor calibration field. A multi-surface calibration room occupied an area of approximately 11.0 m x 11.25 m, in the Geomatics Engineering laboratory at the University of Calgary. The room shown in (Figure 1) has five vertical walls and a ceiling, with 253 coded circular targets attached at various heights and orientations. All target centres were surveyed with a Trimble X9 TLS to sub-millimetre precision. Because the coordinates are known independently and of substantially higher accuracy, any error in the smartphone data can be attributed directly to the scanning application rather than to the registration process.

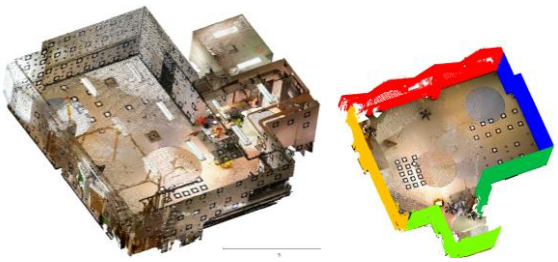

Figure 1. Indoor calibration field (left) measures approximately 11.0 m x 11.25 m and corresponding wall distribution (right).

Outdoor heritage cabins. Two log cabins near Takakkaw Falls in Yoho National Park served as our field test sites. Both are federally designated heritage structures dating to the 1930s, built in the rustic Parks Canada style — exactly the kind of building the parks service wants to monitor. Cabin 1 occupied an area of approximately 7.25 m x 4.85 m, was scanned with Polycam and referenced against a GeoSLAM ZEB Horizon scan. Cabin 2 occupied an area of approximately 10.50 m x 7.25 m, was scanned with both Polycam and UC3D against the same ZEB Horizon reference as shown in (Figure 2), allowing a direct comparison of two apps on the same building under field conditions.

Figure 2. Comparison of point clouds for Cabin 2: GeoSLAM ZEB Horizon reference (top row), Polycam (middle row), and UC3D (bottom row), with corresponding cross-sections shown on the right.

2.2 Sensors

We used three instruments for the study as shown in Figure 3.

Figure 3. Instruments used in this study. Left: iPhone 12 Pro with integrated dToF LiDAR, wide, ultra-wide, and telephoto cameras; Middle: GeoSLAM ZEB Horizon handheld SLAM scanner; Right: Trimble X9 terrestrial laser scanner.

The iPhone 12 Pro carries a direct Time-of-Flight (dToF) LiDAR sensor (940 nm, up to ~5 m range) alongside a triple-camera array. All five scanning applications were run on the same device, sharing the same LiDAR sensor and camera hardware, to eliminate hardware variability. The GeoSLAM ZEB Horizon is a handheld SLAM scanner that achieves 1–3 cm accuracy in typical indoor/outdoor environments. It produced the reference point clouds for the outdoor cabins. The Trimble X9 TLS provided the indoor reference. With 2.5 mm accuracy at 50 m range, it delivered a unified reference dataset with target-centre precision better than 1 mm after multi-station registration.

2.3 Applications Tested

Five applications have been evaluated in the indoor calibration field:

1. **Polycam** is a commercial cross-platform 3D scanning application for spaces, objects, floor plans, and aerial capture, with support for LiDAR- and image-based capture workflows (Polycam, 2024).
2. **PIX4Dcatch**, is a mobile 3D capture application developed by Pix4D for photogrammetric and LiDAR-assisted surveying workflows, with processing through the Pix4D ecosystem (Pix4D, 2024).
3. **KIRI_Engine** is a mobile and web-based 3D scanning platform offering both photo-based and LiDAR-assisted capture modes (KIRI Engine, 2024).
4. **Modelar** Modelar is a LiDAR-based 3D scanning application for Apple devices intended for 3D model and point-cloud capture (Modelar Technologies, 2024).
5. **UC3D** is a prototype developed at the University of Calgary and is described here based on its implementation in this study. Additional comments on workflow behaviour are based on observations made during testing..

All applications produce metric point clouds in their own coordinate systems. These were subsequently registered to the reference point cloud using Iterative Closest Point (ICP) alignment to enable comparison. All indoor scans were captured using the same iPhone 12 Pro under consistent lighting. Each application was operated according to its own recommended scanning procedure to keep the comparison fair. Critically, each application was used to capture three independent scans of the calibration room on separate occasions. These repeated captures form the basis of the repeatability analysis.

2.4 Target Extraction Pipeline

An automated pipeline was developed to find and localise the coded targets in every point cloud — both the reference and the smartphone data. Having an automated, objective workflow was essential because manually picking 253 targets in five (or more) point clouds would be prohibitively slow and prone to operator bias.

The pipeline shown in (Figure 4) works as follows:
Step 1 — Plane segmentation. Each surface (all walls ceiling, floor) is extracted from the point cloud using RANSAC plane fitting.
Step 2 — 2D projection. Points on each fitted plane are projected to create a 2D intensity raster.
Step 3 — Dark-blob detection. The intensity image is inverted and a size-filtered blob detector identifies circular dark features (the targets) based on area, roundness, and contrast to pick the bigger circles.
Step 4 — Candidate refinement. False positives are pruned using interpolation, edge analysis, ring-pattern verification, and geometric constraints (minimum spacing, expected diameter).
Step 5 — Circle fitting and back-projection. Each surviving candidate is fitted with a 2D least-squares circle to find its sub-pixel centroid, which is then back-projected to 3D using the known plane parameters. A sample of data processing is shown in (Figure 5).
Step 6 — Matching. Smartphone target coordinates are matched to reference coordinates by nearest-neighbour search with a 0.1 m threshold in addition to use ICP to have the highest number of target matching.

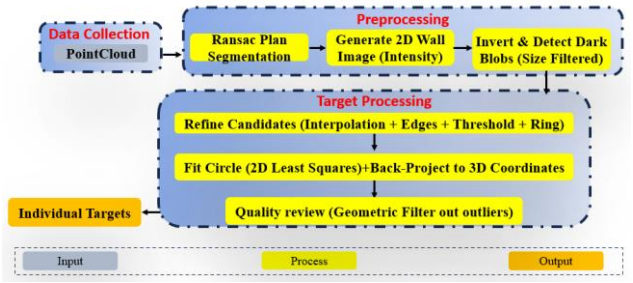

Figure 4. Target detection pipeline

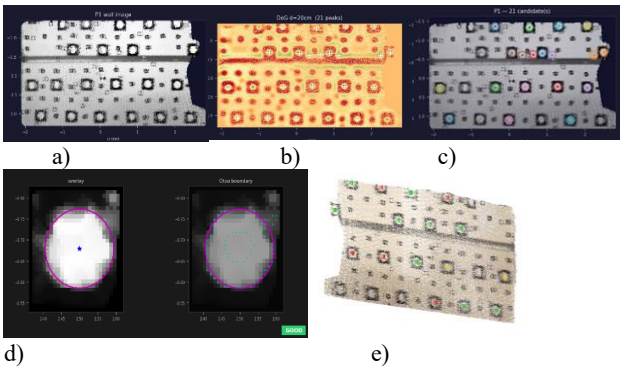

Figure 5. Pipeline output for a single wall surface: (a) 2D intensity raster generated by projecting on-plane points; (b) coded targets visible as dark circles against a lighter wall; (c) initial candidate detections; (d) refined candidate selection; (e) 3D point cloud with detected target centers highlighted.

2.5 Accuracy Metrics

Performance is evaluated using three groups of metrics. The first, target-based 3D positional accuracy, compares each matched

target's smartphone-derived coordinates to its known reference position, yielding Root Mean Square Error ($RMSE_{3D}$), Mean Absolute Error (MAE_{3D}), 95th Percentile Error ($P95_{3D}$), maximum error, and axis-wise mean bias. The detection rate is also reported, defined as the fraction of the 253 reference targets successfully identified by each application.

Let $p_i = (x_i, y_i, z_i)$denote the estimated point and $p_i^{ref}$the corresponding reference point. The 3D positional error magnitude e_i is defined as the Euclidean distance between the estimated and reference coordinates as shown in Eq.1.

$$e_i = \sqrt{\left(\left(x_i - x_i^{ref}\right)^2 + \left(y_i - y_i^{ref}\right)^2 + \left(z_i - z_i^{ref}\right)^2 \right)} \quad (1)$$

From these target-wise errors, the following summary statistics were computed:

$$RMSE_{3D} = \sqrt{\left(\left(\frac{1}{N}\right) sum_{i=1}^{N} e_i^2 \right)} \quad (2)$$

$$MAE_{3D} = \left(\frac{1}{N}\right) sum_{i=1}^{N} e_i \quad (3)$$

$$P95_{3D} = percentile_{95(e_1, e_2, \ldots, e_N)} \quad (4)$$

The second group assesses within-surface inter-target distance accuracy. For every pair of matched targets on the same surface, $\Delta D = D_{app} - D_{ref}$ is computed, representing the difference between distances measured in the smartphone point cloud and the reference. As this metric relies solely on relative positions within a single plane, it is invariant to rigid-body transformations and isolates local geometric fidelity.Because individual ΔD values can be positive or negative, the signed Mean ΔD is reported to reveal systematic bias (e.g., consistent over- or under-estimation of distances), while the Mean Absolute Error (MAE ΔD) quantifies the typical magnitude of the error regardless of sign direction.

The third group evaluates cross-surface inter-target distance accuracy, applying the same formulation to target pairs located on different surfaces. This represents the most stringent test, capturing wall-to-wall distance discrepancies, inter-surface drift, and systematic scale distortions that within-surface metrics cannot reveal. In practice, this metric most clearly exposes structural limitations in commercial applications.

For the outdoor data, Completeness was calculated using Eq. (5) as the proportion of reference points with a corresponding point in the smartphone cloud within a distance threshold.

$$C(\tau) = \frac{1}{N_R} \sum_{i=1}^{N_R} M(d_i \leq \tau) \quad (5)$$

where N_R is the number of points in the reference cloud, $d_i$is the nearest-neighbour distance between each reference point and the evaluated cloud, τ is the defined distance threshold, and $M(d_i \leq \tau)$ is an indicator function (1 if the condition is met, otherwise 0). Completeness is reported as $100 \times C(\tau)$.

For repeatability, each application was used to capture three independent scans of the indoor calibration field. Scan-to-scan consistency was evaluated using Cloud-to-Cloud (C2C) distance analysis and ICP registration RMSE. Each scan was aligned via ICP, and the resulting registration RMSE and C2C deviation profiles (mean, standard deviation, and percentile distributions) were recorded. This approach directly quantifies how much the reconstructed geometry shifts between independent captures, without requiring target identification in every session.

3. Results

3.1 Indoor Registration and Point Density

Each application captured three independent scans of the calibration room. All were registered to the same Trimble X9 reference (183.6M points) via ICP. Table 1 reports the registration RMSE and point count for every scan.

App	ICP RMSE (m)			Mean RMSE (m)	No of Points (M)		
	Scan Number				Scan Number		
	1	2	3		1	2	3
UC3D	0.069	0.094	0.103	0.089	9.5	17.5	10.1
PIX4Dcatch	**0.068**	0.164	0.085	0.106	2.8	2.7	2.4
Polycam	0.074	0.104	0.098	0.092	4.1	3.3	4.0
KIRI Engine	0.097	**0.084**	**0.084**	**0.088**	1.4	2.9	3.0
Modelar	0.095	0.103	0.099	0.099	**123.0**	**124.3**	**165.1**

Table 1. ICP registration RMSE and point density for all three indoor scans.

Several patterns emerge from the three independent scans, each captured on separate occasions under consistent indoor lighting and following the same scanning protocol. First, no application produces the same result twice — registration RMSE fluctuates across scans for every app, which is itself a finding about smartphone scanning reliability. PIX4Dcatch showed the widest swing: its Scan 1 registered at 0.068 m (the best single result of any app) but Scan 2 jumped to 0.164 m, more than double. That kind of variability is a concern for anyone relying on a single scan without a quality check.

UC3D was more stable, ranging from 0.069 m to 0.103 m across the three scans, with a mean of 0.089 m. It also consistently produced the highest point density among the LiDAR-based apps (9.5–17.5 M points), peaking at 17.5 M in Scan 2. The denser clouds give the target extraction pipeline more data to work with, which partly explains UC3D's higher detection rate.

Modelar stands out for its high point counts (61–165 M), which in some scans approach or exceed the Trimble X9 reference. However, qualitative inspection suggests that this increased density does not always correspond to finer geometric detail. Instead, the reconstruction appears to rely on a densification strategy with relatively relaxed geometric constraints, which may introduce redundant or overlapping surface representations. Consequently, the algorithm introduces structural redundancies, such as duplicated or misaligned walls, which explains why the high point count does not inherently translate to improved accuracy (see Section 3.2). Notably, applying a Statistical Outlier Removal (SOR) filter with a stringent threshold effectively isolates and removes most of these artifact points, mitigating the redundancies.

KIRI Engine showed an interesting pattern: its Scan 1 was notably worse (0.097 m) than Scans 2 and 3 (both 0.084 m), and

the point count roughly doubled from Scan 1 to Scans 2–3. This is because the main output is mesh format which can be sampled to any number of points.

A key takeaway from Table 1 is that a low ICP RMSE on a single scan does not guarantee consistent performance. PIX4Dcatch achieved the single-best registration (0.068 m) but also the single-worst (0.164 m). For monitoring applications, where repeat scans must be comparable, the mean and range of registration RMSE across sessions matter more than the best-case number.

3.2 Target-Based 3D Accuracy

Table 2 summarises the target-based accuracy for all five apps, evaluated against the 253 reference targets. Figure 6 illustrates the extracted targets along with a sample trajectory from UC3D. The acquisition followed a consistent scanning strategy recommended across applications: starting from the left wall, capturing the lower regions first, then progressively covering the middle and upper areas, before moving forward and repeating the sequence. This acquisition path is illustrated in Figure 7 through both the calibration field plan and a corresponding cross-section.

App	Matched (%)	$RMSE_{3D}$ (mm)	MAE_{3D} (mm)	$P95_{3D}$ (mm)
UC3D	**62.5**	**49.8**	**43.6**	**93.2**
PIX4Dcatch	56.1	103.4	96.8	151.4
Polycam	53.8	96.7	92.4	128.9
KIRI Engine	24.5	194.2	189.9	243.9
Modelar	51.0	216.2	196.6	322.2

Table 2. Target-based 3D accuracy comparison. Reference: Trimble X9 TLS. Targets were matched to reference coordinates using nearest-neighbour search after ICP registration, with a 0.1 m distance threshold.

The applications show clear differences in geometric accuracy and target detectability. UC3D performed best, achieving the lowest $RMSE_{3D}$ (49.8 mm) and the highest matching rate (62.5%). Polycam and PIX4Dcatch showed comparable mid-level performance, with moderate detection rates (53.8% and 56.1%) and higher $RMSE_{3D}$ values (96.7 mm and 103.4 mm). KIRI Engine recovered few targets (24.5%), consistent with its lower point density, and showed correspondingly larger errors. Modelar produced dense point clouds but the largest positional deviations overall ($RMSE_{3D}$ = 216.2 mm).

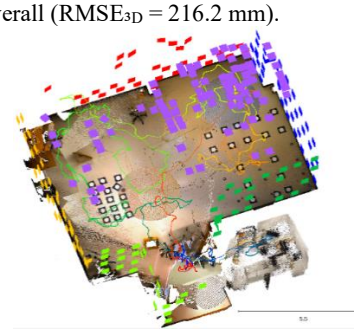

Figure 6. Extracted targets from the reference and trajectory of UC3D scan colored with time. Reference targets are colour-coded by surface.

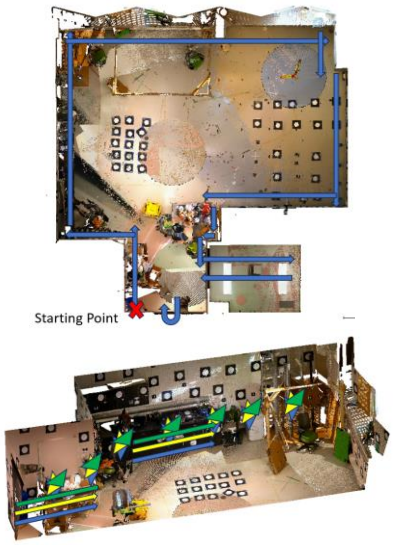

Figure 7. The planned trajectory for data collection (Top). Cross section view (Bottom)-Color triangles represent different observation angles.

Figure 8 presents the cumulative distribution of 3D target position errors for all five applications. UC3D achieves 95% of targets within 93 mm, with a steep curve concentrated below 60 mm, indicating that most errors remain relatively small. PIX4Dcatch and Polycam exhibit similar distribution patterns, though slightly shifted toward higher error values, suggesting moderately larger deviations that remain below 150 mm for the majority of targets.

KIRI Engine and Modelar display broader error distributions. Their curves are more gradual and extend beyond 200 mm, with Modelar reaching a P95 of 322 mm. The plateau observed near 50% in KIRI Engine's curve suggests that approximately half of the matched targets fall within a reasonable accuracy range, while the remainder exhibit higher deviations, which may indicate variability in reconstruction performance across different surfaces.

Modelar's progressively increasing curve, without a distinct inflection point, suggests a more uniform spread of errors across the full range rather than a concentration in a limited number of outliers.

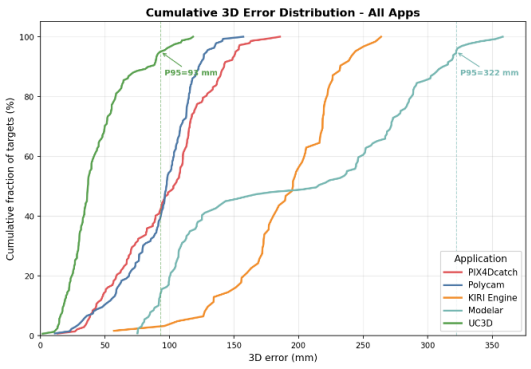

Figure 8. Plan-view position-error quiver for UC3D, Pix4D and Modelar apps.

https://doi.org/10.5194/isprs-archives-XLIX-M-1-2026-1-2026

3.3 Inter-Target Distance Accuracy

Tables 3A through 3C break down the inter-target distance metric by within-surface pairs, cross-wall pairs, and all pairs combined. Where N Pairs is the total number of unique within-surface target pairs, C(n,2), summed across surfaces; all pairs are used to assess accuracy over different baselines, although only n-1 are geometrically independent per surface.

App	N Pairs	Mean ΔD (mm)	MAE ΔD (mm)	RMSE ΔD (mm)	RMSE ΔD (%)
UC3D	**1853**	-11	26.6	37.3	**1.74**
PIX4Dcatch	1536	**4.8**	**24.9**	**33.5**	1.97
Polycam	1412	-22.4	36.5	48	2.3
KIRI Engine	304	-35.1	44.3	55.5	2.96
Modelar	1381	-20.7	29.6	41.5	2.03

Table 3A. Within-surface distance accuracy (**ΔD**) (local shape fidelity).

App	N Pairs	Mean ΔD (mm)	MAE ΔD (mm)	RMSE ΔD (mm)	RMSE ΔD (%)
UC3D	**10550**	**2.9**	**29.2**	**36.8**	**0.71**
PIX4Dcatch	8475	-44.9	75.1	99.4	1.35
Polycam	7768	-75.5	81.4	101.7	1.35
KIRI Engine	1587	-160.4	185.9	215.6	3.46
Modelar	6875	-207.8	211	254.5	3.21

Table 3B. Cross-wall distance accuracy (global geometric consistency).

Application	N Pairs	Mean ΔD (mm)	MAE ΔD (mm)	RMSE ΔD (mm)	RMSE ΔD (%)
UC3D	**12403**	**0.8**	**28.8**	**36.8**	**0.94**
PIX4Dcatch	10011	-37.2	67.4	92.4	1.46
Polycam	9180	-67.3	74.5	95.4	1.54
KIRI Engine	1891	-140.3	163.2	198.7	3.39
Modelar	8256	-176.5	180.7	232.9	3.04

Table 3C. Combined (all pairs) — overall geometric consistency.

Tables 3A-3C show that local surface geometry was generally preserved more consistently than the overall geometry of the room. In Table 3A, within-surface RMSE values range from 33.5 mm to 55.5 mm, whereas in Table 3B the cross-wall RMSE values range more widely, from 36.8 mm to 254.5 mm. This indicates that reconstructing the relationship between different surfaces was more difficult than maintaining the shape of individual walls. Table 3C shows the same overall trend in the combined results, where mean ΔD values range from 0.8 mm to -176.5 mm. The negative mean values indicate that, for some applications, reconstructed distances tend to be shorter than the reference distances, suggesting a degree of global geometric contraction.

Overall, the tables suggest that the main difference between the applications lies not only in the size of the error, but also in how well each method maintains geometric consistency across the full scene.

Figure 9 shows that distances within the same surface are generally preserved better than distances across different surfaces. The within-surface error distributions are typically narrower and closer to zero, whereas the cross-wall distributions are broader and often negatively biased. This pattern suggests that local wall geometry is reconstructed more consistently than the overall room geometry and likely reflects the effect of accumulated SLAM drift as pose errors increase along the scanning trajectory. In other words, some applications can model individual surfaces reasonably well but have more difficulty maintaining correct spacing between walls. Overall, the figure reinforces the trend seen in the tables: the main difference between applications lies not only in the magnitude of error, but also in how well accuracy is preserved across the full scene.

3.4 Scan-to-Scan Repeatability

If smartphone scanning is to be useful for structural monitoring, repeated scans of the same scene must show a stable relationship to a common reference. To evaluate this, three independent scans from each application were registered to the same Trimble X9 point cloud, and repeatability was assessed in two complementary ways: first through the ICP registration RMSE (Table 1), and second through cloud-to-cloud (C2C) deviation profiles computed between each registered scan and the reference. For clarity, the comparison was limited to the retained structural surfaces used in the analysis, rather than the full raw scene, so that the profiles reflect repeatability over the same comparable geometry in each scan. In Figures 10 and 11, panel (a) shows the reference and panels (b)-(d) show the three repeated smartphone scans after registration to that reference.

The C2C profiles support the interpretation from the registration statistics. In Figure 10, the three UC3D scans show broadly similar deviation patterns, with most points remaining within a comparable range and with the larger deviations occurring in approximately the same parts of the scene from one scan to the next. This suggests that the remaining error is relatively stable and spatially repeatable, which is an encouraging result for monitoring applications because consistent biases are generally easier to characterise and account for over time.

Figure 11 shows a different behaviour for PIX4Dcatch. Although one scan appears comparatively close to the reference, the deviation patterns vary more noticeably between repetitions, and one scan in particular shows larger departures across substantial portions of the scene, which is consistent with its higher registration RMSE. From a monitoring perspective, this greater scan-to-scan variability is important, because it suggests that repeatability may depend not only on average accuracy, but also on how consistently the same surfaces are reconstructed across repeated acquisitions.

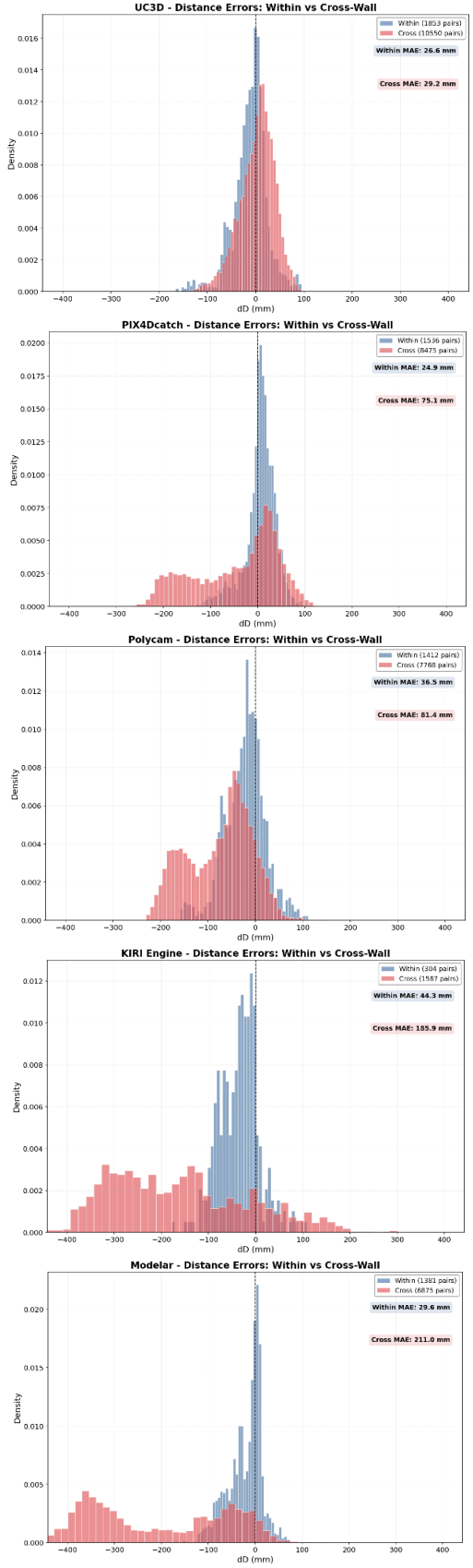

Figure 9. Within-surface vs cross-wall histogram overlay per app, with MAE values annotated. The gap between blue and red directly shows each app's SLAM drift magnitude.

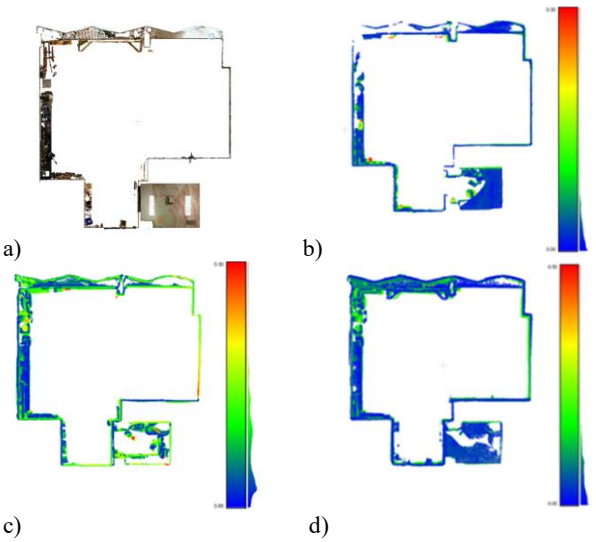

Figure 10. UC3D-C2C deviation profiles between each repeated scan (b, c, d) and the reference point cloud (a).

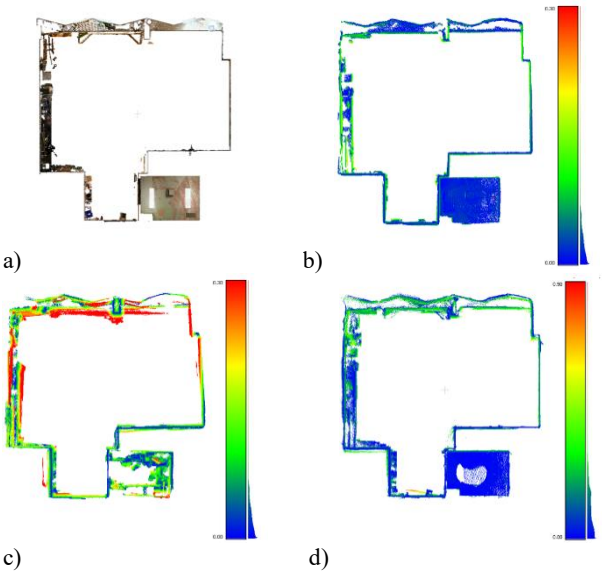

Figure 11. PIX4Dcatch-C2C deviation profiles between each repeated scan (b, c, d) and the reference point cloud (a).

Polycam and KIRI Engine showed intermediate behaviour. Polycam's three scans were moderately consistent, although the cross-wall distortion identified earlier (Section 3.3) was also visible across repeated sessions.

KIRI Engine showed the second-smallest spread in registration RMSE (0.013 m), suggesting relatively stable alignment across scans; however, its comparatively low point density reduces the amount of geometric detail available in the C2C comparison and therefore limits how informative the repeatability assessment can be.

Modelar showed a different type of behaviour. Its registration RMSE remained within a relatively narrow range (0.095-0.103 m), but the total point count varied substantially between scans, from 61 M to 165 M. Visual inspection indicates that some of the observed variation in point density may be associated with overlapping surface representations in the reconstructed point clouds. Consequently, increases in density were not always

accompanied by proportional improvements in geometric definition.

As a result, the apparent increase in density did not necessarily correspond to improved geometric definition. In practice, applying a Statistical Outlier Removal (SOR) filter improved the point cloud quality and produced more reasonable registration behaviour. This indicates that point count alone should not be interpreted as a direct indicator of reconstruction quality, particularly when repeated or redundant geometry may be present.

From a monitoring perspective, the practical implication is that repeatability matters as much as nominal accuracy. The most suitable system is not simply the one that performs well in a single scan, but the one that behaves consistently across repeated surveys. In that sense, the more repeatable applications provide a stronger basis for comparing models over time, whereas systems with greater scan-to-scan variability may make it more difficult to distinguish true structural change from differences introduced during acquisition or reconstruction.

3.5 Outdoor Field Results

Table 4 summarises the field results at Yoho National Park. Both apps achieved usable completeness — Polycam slightly higher (92–98%) thanks to its denser point clouds. But visual inspection tells a different story from the numbers. Polycam's reconstructions show noticeable roof sagging and wall curvature (Figure 12, middle row), artefacts consistent with the cross-wall drift we measured indoors. UC3D's model has fewer points but the structural lines are noticeably straighter. Cross-sections cut perpendicular to the cabin's long axis confirm that UC3D maintained more consistent wall verticality and ridge-to-eave angles compared to Polycam.

Test Cabin	Scanner	Pts (M)	Completeness (%)	RMSE (m)
1	GeoSLAM	35.9	100	–
	Polycam	2.5	92	0.04
2	GeoSLAM	57.4	100	–
	Polycam	3.8	98	0.07
	UC3D	1.0	91	0.07

Table 4. Outdoor field comparison against GeoSLAM ZEB Horizon reference.

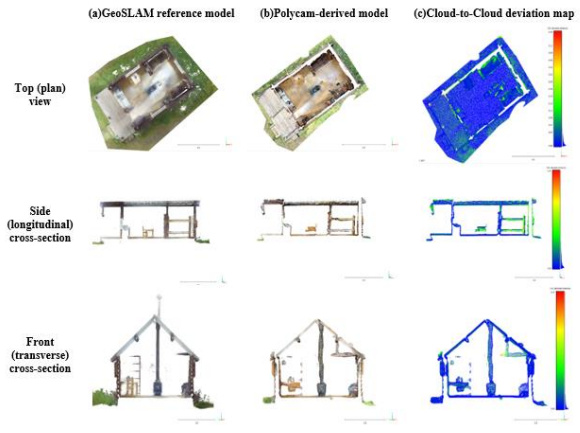

Figure 12. C2C deviation analysis between Polycam and GeoSLAM scans of Cabin 1.

UC3D did produce some data gaps in areas with uniform log textures, where the camera-based tracking had little visual information to work with. This is a known trade-off of the prototype's more conservative fusion strategy — it rejects uncertain frames rather than interpolating through them. For monitoring purposes, a few gaps in a uniform wall are preferable to a smooth but geometrically warped surface

4. DISCUSSION

4.1 What the Numbers Mean

Taken together, the results show a consistent pattern across the different metrics. The values span a wide range, especially for the cross-wall distance measures, which are the most informative for judging how well the overall room geometry is preserved. In general, within-surface results were more stable than cross-wall results, indicating that reconstructing the shape of an individual wall or surface was less challenging than maintaining the correct geometric relationship between multiple surfaces across the room.

This difference can be seen clearly in the inter-target distance results. The within-surface RMSE values fall within a narrower range, while the cross-wall RMSE values increase substantially in some cases. This suggests that good local surface reconstruction does not necessarily translate into equally strong room-scale geometric consistency. The sign of the mean ΔD values is also informative: negative values indicate a tendency for reconstructed distances to be shorter than the corresponding reference distances, suggesting a degree of geometric contraction in the reconstructed scene.

The ICP registration results (Table 1) provide an important point of comparison. In some cases, scans that registered well to the reference surface still showed noticeably larger target-based or inter-target distance errors. This indicates that ICP alone, because it minimizes overall surface-to-surface deviation, does not necessarily capture how well internal geometry is preserved. As a result, evaluation based only on registration RMSE or cloud-to-cloud distance may give an incomplete picture of metric performance.

4.2 Repeatability and What It Means for Monitoring

The repeatability analysis addresses a practical question that single-scan accuracy alone cannot answer whether differences observed between scans collected at different times can be interpreted with confidence. For monitoring applications, it is not enough for a scan to be accurate once; it also needs to behave consistently from one session to the next. If deviation patterns shift substantially between repeated scans, it becomes more difficult to distinguish genuine structural change from variation introduced during acquisition or reconstruction.

The repeated-scan results suggest that the tested systems do not all behave with the same level of stability. In the more repeatable cases, both the registration residuals and the spatial deviation patterns remained comparatively consistent across sessions, which implies a more predictable scan-to-scan noise floor. In other cases, larger variation was observed between repeated scans, making the interpretation of change less straightforward. From a monitoring perspective, this distinction is important, because the usefulness of a system depends not only on nominal accuracy, but also on whether its residual error remains stable enough to support meaningful comparison over time.

At the same time, the repeatability achieved by smartphone-based systems remains below that typically expected from terrestrial laser scanning. These methods should therefore not be interpreted as direct replacements for survey-grade instruments where sub-centimetre precision is required. However, where conventional survey is impractical or unavailable, a repeatable smartphone workflow may still offer useful comparative information, especially if scans are always registered to the same reference baseline and supported by clear quality-control criteria.

4.3 Tolerances and Heritage Documentation Standards

Published guidance from ICOMOS and ISPRS generally places acceptable tolerances in the range of 1-5 cm for measured drawings and roughly 2-10 cm for general heritage recording (Letellier et al., 2007). When viewed against those benchmarks, the lower-error results in this study fall within the broader range expected for general heritage documentation and, in some cases, approach the lower end of measured-drawing requirements. The larger-error results, particularly in the cross-wall metrics, would be more difficult to justify where reliable dimensional accuracy is essential.

For the cabin-monitoring context considered here, relative geometry is especially important. Measures such as wall-to-wall, corner-to-corner, and other room-scale relationships are often more relevant than absolute point position alone. From that perspective, the cross-wall distance metric is particularly informative, because it reflects how well the full spatial structure of the room is maintained rather than only the shape of isolated surfaces.

4.4 Why the Commercial Apps Struggle with Cross-Wall Accuracy

The mostly negative mean ΔD values in the cross-wall results suggest that larger room-scale distances are often reconstructed slightly shorter than they are in the reference. A likely explanation is accumulated SLAM trajectory drift, where small pose errors build up over the scan path and are not fully corrected during loop closure or global adjustment (Strasdat et al., 2010). In practical terms, this creates a mild compression effect, so that distant surfaces, such as opposite walls, appear closer together than they should. Similar behaviour has been reported in previous studies of SLAM-based mobile mapping and handheld LiDAR surveying (Sammartano and Spanò, 2018; Díaz-Vilariño et al., 2022; Teppati Losè et al., 2022).

The lower-bias results suggest that some reconstruction pipelines are better at limiting this effect than others. This means that the main difference is not only how well a single wall is reconstructed, but how well the full room geometry is maintained once all surfaces are combined into one model. For that reason, cross-wall distance error is a useful way to assess how well a smartphone scanning workflow controls drift over the full scene.

4.5 Limitations

Several limitations should be acknowledged. The indoor calibration room used here was more regular and controlled than a typical heritage cabin interior, where furniture, rough timber surfaces, clutter, and variable lighting may introduce additional difficulty. The field testing was limited in scope, and the repeatability analysis was based on three scans per application, which is sufficient to show general trends but not to fully characterize variability under all conditions. In addition, the effective range of the iPhone LiDAR sensor limits coverage in larger spaces, particularly for ceilings and more distant surfaces. These factors mean that the present results should be interpreted as a controlled comparative assessment rather than a complete representation of all field situations.

5. CONCLUSION

This study examined whether iPhone-based scanning applications can provide metrically useful 3D data for heritage cabin documentation and monitoring. The results suggest that they can offer useful geometric information, but that the level of performance varies substantially depending on the application and on the metric being considered.

Across the controlled indoor testing, the lower-error cases showed not only smaller positional and distance errors, but also better preservation of room-scale geometry and more stable behaviour across repeated scans. Other cases produced acceptable local surface reconstruction but showed larger cross-wall errors or greater scan-to-scan variability, which would make them less suitable for dimensional interpretation over time. This distinction is important, because it suggests that evaluating smartphone scanning only in terms of local surface quality may overlook more meaningful differences in global geometric consistency.

The field observations were broadly consistent with the indoor findings. Both systems tested in the cabins were able to produce highly complete models, but visible geometric distortion remained more apparent in some reconstructions than in others. This agreement between controlled and field-based observations strengthens confidence in the general interpretation of the results.

Overall, the findings suggest that smartphone scanning can serve as a practical supplementary tool for heritage recording, particularly where access, logistics, or cost limit the use of terrestrial laser scanning. Its role is best understood as an accessible method for documentation, screening, and repeated comparison, rather than as a replacement for survey-grade instruments. Future work should expand testing to a wider range of heritage sites and conditions, investigate repeatability under field use more extensively, and continue refining workflows so that non-specialist users can collect data suitable for consistent comparison over time.

Acknowledgment

The author would like to thank the supervisors and collaborators who provided guidance and feedback throughout this study. Appreciation is also extended to Parks Canada and Yoho National Park for supporting the field component of the work and for facilitating access to the heritage cabin sites. Special thanks are extended to Dr. Peter Dawson for assistance with GeoSLAM data collection, and to Rhonda Owchar of Parks Canada for facilitating access to the heritage cabin sites in Yoho National Park. Funding for this work was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2024-03793) and the UCalgary Research Excellence Chairs program.

REFERENCES

Askar, C., Sternberg, H., 2023. Use of smartphone LiDAR technology for low-cost 3D building documentation with iPhone 13 Pro: a comparative analysis of mobile scanning applications. Geomatics, 3(4), 563–579. https://doi.org/10.3390/geomatics3040030

Boardman, C., Bryan, P., 2018. 3D laser scanning for heritage: advice and guidance on the use of laser scanning in archaeology and architecture (3rd ed.). Historic England, Swindon, UK. https://historicengland.org.uk/images-books/publications/3d-laser-scanning-heritage/heag155-3d-laser-scanning/

Costantino, D., Pepe, M., Restuccia, A.G., 2022. Smartphone LiDAR technologies for surveying and reality modelling in urban scenarios: evaluation methods, performance and challenges. Applied System Innovation, 5(4), 63. https://doi.org/10.3390/asi5040063.

Dhonju, H.K., Xiao, W., Sarhosis, V., 2017. Feasibility study of low-cost image-based heritage documentation in Nepal. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W3, 237–242. https://doi.org/10.5194/isprs-archives-XLII-2-W3-237-2017

Díaz-Vilariño, L., Tran, H., Frías, E., Balado, J., Khoshelham, K., 2022. 3D mapping of indoor and outdoor environments using Apple smart devices. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B4-2022, 303–308. https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-303-2022

Gollob, C., Wieser, M., Briese, C., 2021. Measurement of forest inventory parameters with Apple iPad Pro and integrated LiDAR technology. Remote Sensing, 13(16), 3129. https://doi.org/10.3390/rs13163129

Heritage, G.L., Large, A.R.G. (Eds.), 2009. Laser Scanning for the Environmental Sciences. Wiley-Blackwell, Oxford, UK. https://doi.org/10.1002/9781444311952

James, M.R., Robson, S., 2012. Straightforward reconstruction of 3D surfaces and topography with a camera: accuracy and geoscience application. Journal of Geophysical Research: Earth Surface, 117, F03017. https://doi.org/10.1029/2011JF002289

Jouan, P., Hallot, P., 2020. Digital twin: research framework to support preventive conservation policies. ISPRS Int. J. Geo-Inf., 9(4), 228. https://doi.org/10.3390/ijgi9040228

KIRI Engine, 2024. KIRI Engine: 3D Scanner App for iPhone, Android, and Web. https://www.kiriengine.app/ (accessed 15 March 2026).

Letellier, R., Schmid, W., LeBlanc, F., 2007. Recording, Documentation, and Information Management for the Conservation of Heritage Places: Guiding Principles. Getty Conservation Institute, Los Angeles, CA, USA. https://hdl.handle.net/10020/gci_pubs/recordim

Luetzenburg, G., Kroon, A., Bjørk, A.A., 2021. Evaluation of the Apple iPhone 12 Pro LiDAR for an application in geosciences. Scientific Reports, 11, 22221. https://doi.org/10.1038/s41598-021-01763-9

Modelar Technologies, 2024. Modelar - 3D LiDAR Scanner. Apple App Store. https://apps.apple.com/us/app/modelar-3d-lidar-scanner/id1572844190 (accessed 15 March 2026).

Murtiyoso, A., Grussenmeyer, P., 2021. Experiments using smartphone-based videogrammetry for low-cost cultural heritage documentation. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVI-M-1, 487–491. https://doi.org/10.5194/isprs-archives-XLVI-M-1-2021-487-2021

Murtiyoso, A., Grussenmeyer, P., Suwardhi, D., Awalludin, R., 2018. Multi-scale and multi-sensor 3D documentation of heritage complexes in urban areas. ISPRS Int. J. Geo-Inf., 7(12), 483. https://doi.org/10.3390/ijgi7120483

Nocerino, E., Poiesi, F., Loescher, A., Remondino, F., Rieke-Zapp, D., 2017. 3D reconstruction with a collaborative approach based on smartphones and a cloud-based server. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W8, 187–194. https://doi.org/10.5194/isprs-archives-XLII-2-W8-187-2017

Paukkonen, N., 2023. Towards a mobile 3D documentation solution: video-based photogrammetry and iPhone 12 Pro LiDAR. Journal of Computer Applications in Archaeology, 6(1), 143–154. https://doi.org/10.5334/jcaa.135

Pix4D, 2024. PIX4Dcatch documentation. https://support.pix4d.com/hc/pix4dcatch (accessed 15 March 2026).

Polycam, 2024. Polycam: cross-platform AI 3D scanning, floor plans and drone mapping. https://poly.cam/ (accessed 15 March 2026).

Sammartano, G., Spanò, A., 2018. Point clouds by SLAM-based mobile mapping systems: accuracy and geometric content validation in multisensor survey and stand-alone acquisition. Applied Geomatics, 10, 317–339. https://doi.org/10.1007/s12518-018-0221-7

Spreafico, A., Chiabrando, F., Teppati Losè, L., Tonolo, F.G., 2021. The iPad Pro built-in LiDAR sensor: 3D rapid mapping tests and quality assessment. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B1-2021, 63–69. https://doi.org/10.5194/isprs-archives-XLIII-B1-2021-63-2021.

Strasdat, H., Montiel, J.M.M., Davison, A.J., 2010. Scale drift-aware large scale monocular SLAM. In: Robotics: Science and Systems VI. https://doi.org/10.15607/RSS.2010.VI.010

Tanduo, B., Malinverni, E.S., Pierdicca, R., Paolanti, M., Frontoni, E., 2021. 3D surveying of underground built heritage: opportunities and challenges of mobile technologies. Sustainability, 13(23), 13289. https://doi.org/10.3390/su132313289.

Teppati Losè, L., Chiabrando, F., Dalmasso, S., 2022. Apple LiDAR sensor for 3D surveying: tests and results in the cultural heritage domain. Remote Sensing, 14(17), 4157. https://doi.org/10.3390/rs14174157

Vosselman, G., Maas, H.-G. (Eds.), 2010. Airborne and Terrestrial Laser Scanning. Whittles Publishing, Dunbeath, Scotland, UK. [no DOI] https://www.worldcat.org/title/airborne-and-terrestrial-laser-scanning/oclc/422689997

Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M., 2012. "Structure-from-Motion" photogrammetry: a low-cost, effective tool for geoscience applications. Geomorphology, 179, 300–314. https://doi.org/10.1016/j.geomorph.2012.08.021