Geometric and Visual SLAM: The accuracy of modern handheld LiDAR scanners

Christoph Strecha[1], Ryan Hughes[1], Davide A. Cucci[1], Andrea Dotti[1], Marta Coelho Lopes[2], Pierangelo Rothenbühler[1]

[1] PIX4D SA, Route de Renens 24, 1008 Prilly, Switzerland (firstname.lastname@pix4d.com)
[2] École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (firstname.lastname@epfl.ch)

Keywords: handheld laser scanner, iPhone, SLAM, Visual SLAM, LiDAR, RTK, GNSS, Surveying, 3D-modeling

Abstract

Recent handheld scanners increasingly integrate geometric (LiDAR-based) and visual (image-based) SLAM (Simultaneous Localization And Mapping), promising low-cost and flexible solutions for surveying tasks. This paper evaluates the accuracy of three such systems: the XGRIDS Lixel K1, the SHARE S20, and a Pix4D solution pairing an iPhone Pro with an Emlid Reach RX GNSS (Global Navigation Satellite System) antenna. We conducted experiments in two distinct environments: Scene 1, with continuous, high-quality RTK (Real-Time Kinematic) coverage, and Scene 2, which included an indoor trajectory resulting in a temporary loss of the RTK fix. Accuracy was validated against independent GNSS check points. In Scene 1, the Pix4D solution delivered survey-grade results, achieving a RMSE (Root Mean Square Error) below 3 cm in the X, Y, and Z directions. The XGRIDS and SHARE scanners yielded larger maximum errors, around 10 to 15 cm. In Scene 2, accuracy degraded; the Pix4D solution's maximum error increased to approximately 12 cm , while the Share S20's maximum error exceeded 25 cm. We conclude that while the fusion of visual and geometric SLAM is powerful, a stable RTK fix remains critical for achieving consistent survey-grade accuracy with current low-cost handheld scanners.

1. Introduction

The use of digital models in surveying applications is becoming increasingly popular. These tools have the potential to replace, at least partially, traditional point-surveying in many applications, due to the increased flexibility and the possibility to reach survey-grade accuracy standards. As an example, most dominantly in open pit mining applications, the use of drones allows surveyors to avoid the need to physically enter potentially dangerous mining areas to collect single points, breaks, or volumes. The photogrammetric digital models obtained with drones allow one to accurately perform measurements remotely. In these contexts the use of GCPs (Ground Control Points) and/or RTK (Real-Time Kinematic) is essential to georeference the digital model into the required coordinate system with the needed accuracies. These workflows are well understood and are routinely used. A limiting factor for the adoption of drone technologies is the specific knowledge required to perform data acquisition with the best possible parameters, an important aspect that characterizes the overall quality of the digital model.

With the advancement of the accuracy and precision of data processing algorithms for the reconstruction of the acquired scenes, it is now possible to explore the use of consumer-grade hardware (e.g., phones, handheld scanners) in the field of high-precision surveying. This trend is clearly visible with the availability of data acquisition devices that target the sector. The benefit of this category of devices is not only in economics, but also in the simplified operational considerations.

All the modeling methods require, as a key input, the accurate knowledge of the trajectory where the image and LiDAR (Light Detection And Ranging) data have been acquired. The reconstruction of such acquisition path, in space and time, from hardware sensors is the domain of SLAM (Simultaneous Localization And Mapping) algorithms.

1.1 Geometric and Visual SLAM

SLAM is one of the key technologies to a precise trajectory estimation of the scanner. There are two main families of algorithms, used in different domains, and they differentiate mainly on the availability of LiDAR data. We will here reference these techniques as geometric SLAM, where time-of-flight information is used to create a spatial representation of the scene. If LiDAR data are not available, algorithms need to use the information in images to reconstruct the trajectory via computer vision techniques. We will reference this category of techniques as visual SLAM. Both techniques are usually integrated with inertial IMU (Inertial Measurement Unit) sensor information or / and absolute positioning information from GNSS (Global Navigation Satellite System) (Zhu et al., 2024).

Handheld laser scanning has been available for some time. The main accuracy challenge of those scanners lies in the exact determination of the position and orientation of the device, usually referred to as trajectory optimization. The fundamental principle is the use of a high accurate clock and a laser to measure distances, reconstructed as absolute $3D$ coordinates. The trajectory adjustment use geometric SLAM to integrate the positioning (GNSS) and orientation sensors (IMU) with the constraint of matching surfaces or distinct features in the point cloud to obtain a consistent trajectory. One of the main limitation of SLAM reconstructions is the *drift*, positioning errors that accumulate along the trajectory. Much work has been done to mitigate drifts, as in (Keitaanniemi et al., 2023, Chen et al., 2022, Kuželka, 2025, Karam et al., 2020) just to name a few. Those authors study the integration in the SLAM pipeline of IMU (Inertial Measurement Unit) data and, for outdoor applications, GNSS data. Geometric SLAM requires, however, distinct geometric features which may be lacking especially in indoor environments. In practice geometric SLAM workflows often rely on the use of GCPs to improve the quality of the reconstruction, for instance (NavVis, 2025,

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-193-2026

Figure 1. From left to right: XGRIDS Lixel K1, SHARE S20 and iPhone with Emlid Reach RX running PIX4Dcatch.

Faro, 2025).

Visual SLAM, or VSLAM, is very popular in the robotics, computer vision and Augmented Reality community (Macario Barros et al., 2022). VSLAM requires, compared to LiDAR sensor, lower energy, weight, and cost for a pure imaging sensor, making this technology very attractive in applications where these characteristics are of value. Similarly to geometric SLAM, it relies on the use of features (in this case extracted from images) that are tracked to estimate the acquisition trajectory. Often IMU data are used in the process to better constrain the scene (Vidal et al., 2017). This class of VSLAM algorithms are very popular and are available in the majority of smartphones and other consumer-grade devices. ARcore (for Android-based devices) and ARkit (for iOS-based devices) are libraries that are used by applications in the AR/VR and $3D$ modeling fields (Nowacki and Woda, 2020). For the iPhone pro, which comes with a LiDAR sensor, the trajectory estimation makes use of a mixed RGB-Depth SLAM approach (Shan et al., 2019). VSLAM performs best in areas where images have rich textures, where many distinct features can be extracted from images and tracked consistently throughout the image sequence. To increase the recognition of distinct visual features it is possible to use fisheye cameras (El-Alailyi et al., 2024) that increase substantially the field-of-view. Using a pair of synchronized fisheye cameras (Zhang et al., 2023) it is possible to acquire a full scene in all directions (*360-degree cameras*).

1.2 Integrating Visual and Geometric SLAM

Geometric and visual SLAM, in combination, have the potential to provide the best results, removing the limitations of the single technologies. LiDAR data can compensate the lack of texture in interior scenes; strong visual constraints in images (key points) can reduce the effect of LiDAR drift. It seems obvious to integrate the constraints of geometric and visual SLAM into a single optimization, further refined by IMU and GNSS data when available. Combined optimization approaches have a larger potential to remove the need for GCPs which prevents currently a more widespread use of handheld scanners for surveying applications.

In the robotics community, we have integrated visual and geometric SLAM (Graeter et al., 2018, Zuo et al., 2020, Zhao et al., 2021, Lin and Zhang, 2022) and recently a number of manufacturers in the surveying world have combined in single devices image-based VSLAM with the geometric SLAM (Pix4D, 2024, xgrids, 2025, Share, 2025, NavVis, 2025, Leica Geosystems, 2024). Those devices are aided by data from internal IMU sensors. Some device categories are equipped with GNSS antennas that allow cm-level absolute positioning of the device and reduce drift even in long captures (Pix4D, 2024, xgrids, 2025, Share, 2025). In case of good satellite visibility and rich image content, the producers claim to reach survey-grade accuracy even without the use of GCPs.

In this paper, we compare three scanning solutions that combine Photogrammetry, LiDAR, global positioning, and orientation sensors for a precise trajectory estimation.

The three considered solutions (see Fig. 1) are:

- XGRIDS Lixel K1, equipped with two 1/2-inch sensors with 12MP images with $180°$ fisheye lenses and a LiDAR with a range of 0.1 to 40 m. According to specifications, it reaches an accuracy of $\leq$ 3 cm (absolute) and $\leq$ 1.2 cm (relative). An RTK antenna is included (xgrids, 2025).
- SHARE S20, equipped with two 1-inch 16MP image sensors with $180°$ fisheye lenses and a LiDAR with a range of 0.1 to 40 m. According to specifications, it reaches an accuracy of $\leq$ 5 cm (absolute) and $\leq$ 1 cm (relative). An RTK antenna is included (Share, 2025)
- Pix4D provides a solution with the iPhone pro as the imaging and ToF LiDAR sensor. The iPhone can be connected with various GNSS antennas (Pix4D, 2025) in rigid manner such that the relative position between

GNSS antenna and image sensors can be pre-calibrated. For the experiments in this paper we use the Emlid Reach RX for the RTK corrections (Emlid, 2024).

Other scanners available in the market, e.g., Leica's BLK2Go (Leica Geosystems, 2024) and NavVis (NavVis, 2025) scanners, advertise integrated photogrammetry and LiDAR capabilities. These devices are typically in a different price range or provide the final result *as-a-service*, as is the case for NavVis. In the scope of this paper we have instead focused our attention on low-cost devices that allow some control on the data acquisition and processing.

2. Scanning Workflow

The workflow for the handheld scanners is divided into two parts described in the next sections. The details are not shared by manufacturers and are inferred from our experience.

2.1 Data Acquisition and Online Data Processing

The three devices all use a cellular phone to act as the user interface and to manage the collection of the image and LiDAR data. During acquisition a real-time visualization of the LiDAR points is shown on the screen. This helps the operator to direct the scanners towards areas that have not yet been scanned or need more data.

According to the specifications of the manufacturers, all three integrate IMU, RTK and visual tracking information in real-time to compute the scanner trajectory, allowing for the visualization of the LiDAR distance measurements directly on the phone application. Specific details are not provided. From our experience this process works reasonably well, small misalignments due to drift are visible but do not influence the ability to scan a scene as complete as desired. For the iPhone scanner, we use the *PIX4Dcatch* application. Similar to the other two devices it provides a real-time $3D$ LiDAR visualization, in addition it displays a semi-transparent image overlay, with the history of the acquired LiDAR data. This feature additionally helps guiding the user towards the desired scanning area requiring more coverage. However, the LiDAR only has a limited range of approximately 1 to 7 m. For XGRIDS Lixel K1 and SHARE S20 users are provided with *LixelGo* and *SHARE capture* applications. Available both for the Android and iOS operating systems. For our experiments, we used the Android versions.

After the data acquisition, all solutions provide a preview of the LiDAR data. All raw acquisition data (LiDAR, images, geolocation, and IMU) can be downloaded to fix potential drift during a trajectory refinement stage in an offline post-processing software.

2.2 Trajectory Refinement and Post-processing

Each of the manufacturers of the three tested devices provides an offline software to refine the trajectories by integrating geometric and VSLAM outputs with the external orientation provided by GNSS and IMU. We use the following software for post-processing (all available for the Windows operating system):

- For XGRIDS Lixel K1 data: LixelStudio v3.0.2.1
- For SHARE S20 data: PointCloud Studio v2.2.1
- For the PIX4Dcatch data: PIX4Dmatic v1.81

The raw data of the iPhone acquisition with PIX4Dcatch can be processed, in principle, with other photogrammetry software suites, but we restrict the evaluation of the accuracy within the Pix4D ecosystem, via PIX4Dmatic as post-processing software. For the other two devices, since their raw data is not an open format, only the software provided by the respective companies can be used.

3. Experimental Setup and Accuracy Validation

For the accuracy estimation of the three devices, we used two scenes. The first has good satellite visibility and, therefore, good RTK accuracy throughout the entire trajectory. The second scene did include an indoor space. The data of the three devices was therefore collected with lost RTK fix for about 20% of the overall trajectory. The two reconstructed $3D$ scenes are shown for reference in Fig. 2. In addition to the scans, we collected independent GNSS CPs (Check Points) using the Emlid Reach RX RTK rover (the points are shown in purple). The measurement of these control points are used as ground truth, against which we measure the accuracy of the three setups. A total of 29 and 18 reference points are collected for the two scenes, respectively.

To evaluate the accuracy of the scans with respect to our independent ground truth points, we marked CPs in the images and triangulated their $3D$ - position using externals provided by the respective post-processing solutions. Accuracy is defined as the RMSE (Root Mean Square Error) of the distance between the triangulated $3D$ - position and the ground truth measurement.

Marking the CPs in images and then triangulating, instead of directly marking the LiDAR points, increases the precision of the marks. This is possible because all the post-processing solutions provide externals of the captured images. Operations on images can be done with a pixel level precision, a more accurate approach than the manual selection of single $3D$ LiDAR points.

4. Results

The bar plots in Fig. 3 and Fig. 4 report the RMSE, in cm, for each check point of the two scenes. The errors for the Pix4D solution are the lowest, with about 5 cm maximal error. The largest error of the two other scanners is around 10 to 15 cm. For the second scene, the RTK fix is missing in about 20% of the trajectory and the trajectory refinement calculation cannot benefit from accurate geolocation constraints in all points. The maximum error increases, in this case, to about 12 cm, 17 cm, and more than 25 cm for Pix4D, XGRIDS and SHARE solutions, respectively.

Similar conclusions can be drawn for the three spatial components RMSE measured separately (Northing, Easting, Elevation: X, Y, Z), as shown in Fig. 5, and Fig.6 for the two scenes.

Pix4D solution performs very well for both scenes. The iPhone with post-processing in PIX4Dmatic yields

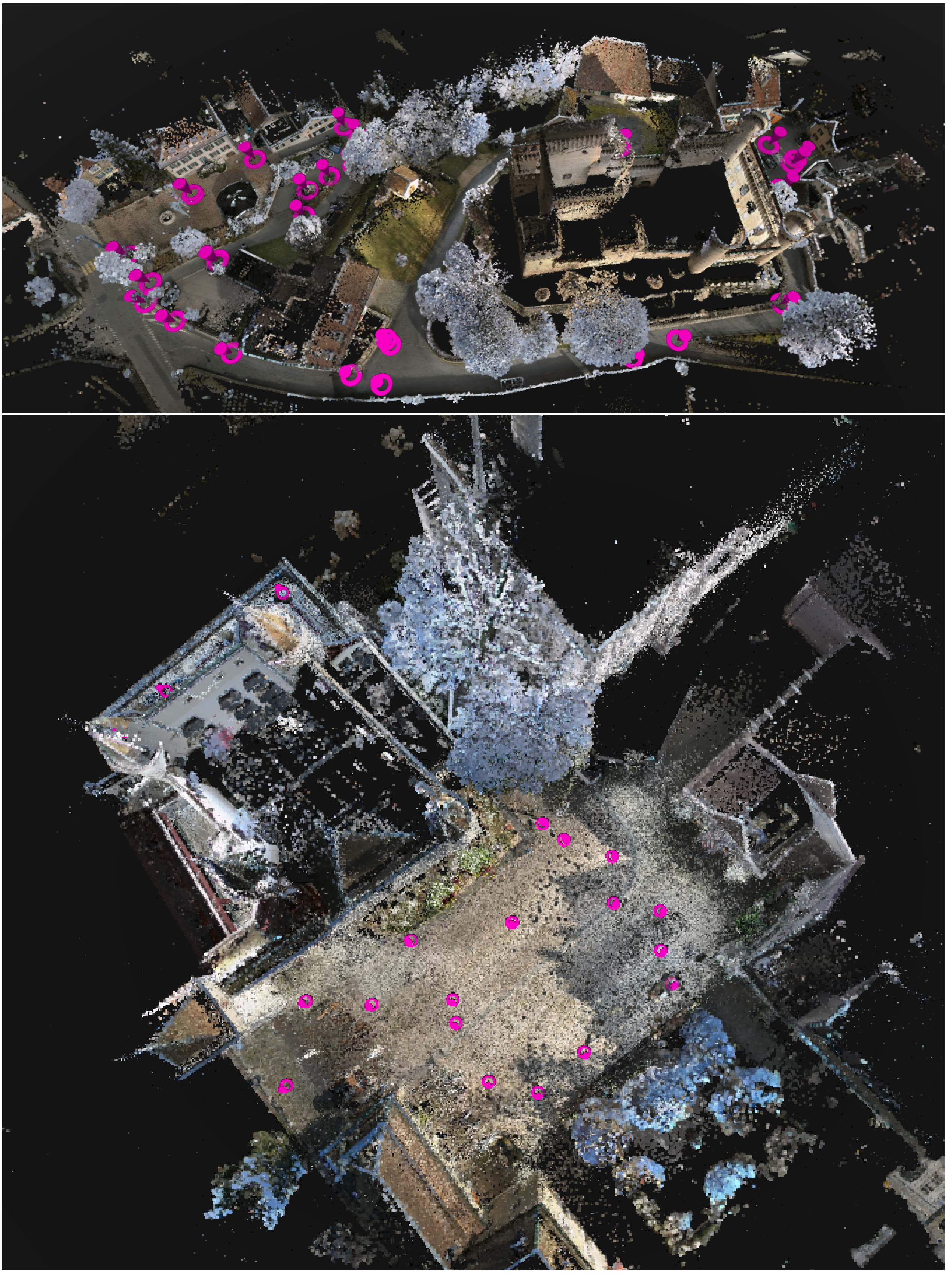

Figure 2. LiDAR point cloud and validation points (purple). **Top**: Scene 1, with good RTK accuracy. **Bottom**: Scene 2, which included an indoor trajectory where RTK fix was lost.

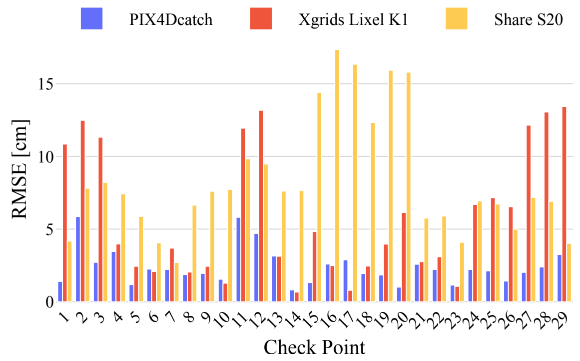

Figure 3. RMSE of all 29 individual CPs for the first scene and for each evaluated device.

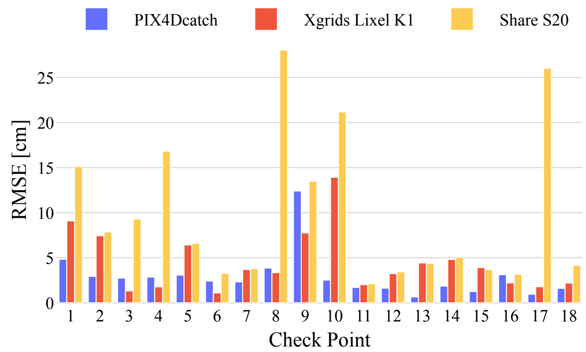

Figure 4. RMSE of all 18 individual CPs for the second scene and for each evaluated device.

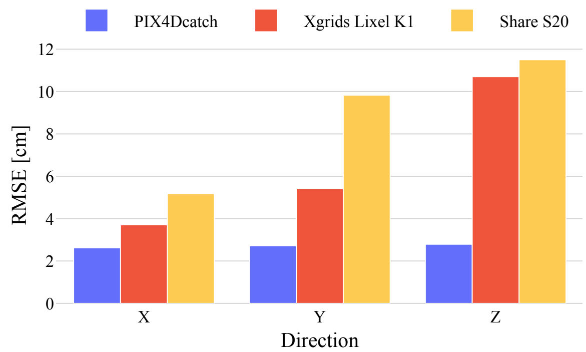

Figure 5. RMSE of the three components $(X,\ Y,\ Z)$ measured in the first scene.

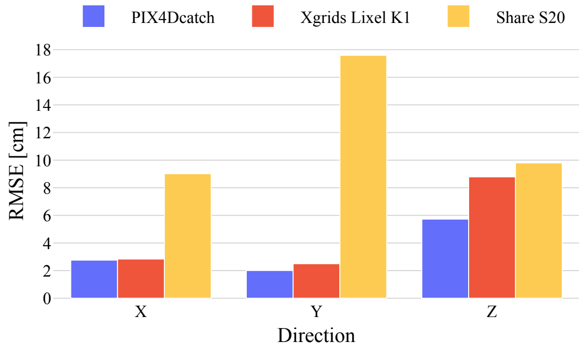

Figure 6. RMSE of the three components $(X,\ Y,\ Z)$ measured in the second scene.

survey-grade measurements with an RMSE below 3 cm in each single direction for the first scene. The XGRIDS and SHARE scanners have much larger errors. In the second scene the RMSE of the Pix4D solution degrades to 5 cm in the Z direction (Fig. 6).

As an additional check, we have visually inspected the reconstructed point clouds and verified the absence of double surfaces. Visualization of the point clouds and trajectories, are shown in Fig. 7 and Fig. 8. Even though the iPhone LiDAR has a maximum range of only 7 m, substantially lower than the handheld scanners case, it is possible to notice that the point cloud is complete. This is possible thanks to the combined generation of LiDAR and Photogrammetric point clouds in the PIX4Dmatic software. In the first scene (Fig. 7) one can appreciate points even on the tallest tower, not present in the pure LiDAR data from the handheld devices, because the top of the tower is outside the 40 m range from any of the possible trajectory locations. In the second scene (Fig. 8), it is possible to appreciate parts of the city, far beyond the scanning area. The Pix4D solution's technique of combining dense LiDAR and photogrammetry is an advantage, however, this is likely only a temporary advantage, since a dense photogrammetric points cloud could also be computed in the post-processing software of the handheld devices.

5. Discussion and Conclusion

5.1 Discussion of Results

Our experimental results provide a clear snapshot of the current state of low-cost handheld scanners. The most significant finding is the critical importance of a stable GNSS signal to achieve survey-grade accuracy. The difference in error magnitudes between Scene 1 (complete RTK) and Scene 2 (20% RTK loss) demonstrates this point conclusively. Although the integration of visual and geometric SLAM aids in trajectory estimation, a continuous RTK fix remains the dominant constraint for absolute accuracy. When the RTK fix was lost, accuracy degraded across all tested devices (e.g. check point 9).

Among the three solutions, the Pix4D workflow (combining an iPhone Pro, PIX4Dcatch, and PIX4Dmatic) yielded the highest accuracy in both scenarios. In the ideal conditions of Scene 1, it was the only system to achieve survey-grade results, with an RMSE below 3 cm in all three directions. The XGRIDS and

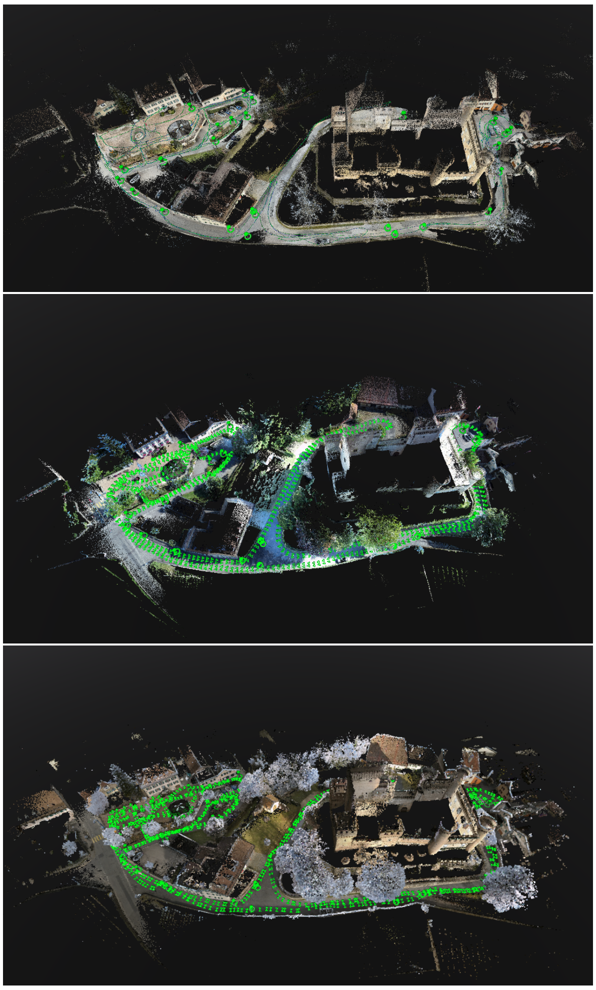

Figure 7. Final point cloud for Pix4D (top), SHARE S20 (middle) and XGRIDS (bottom) for the first scene. In each of those images one can see the check points and image positions in green.

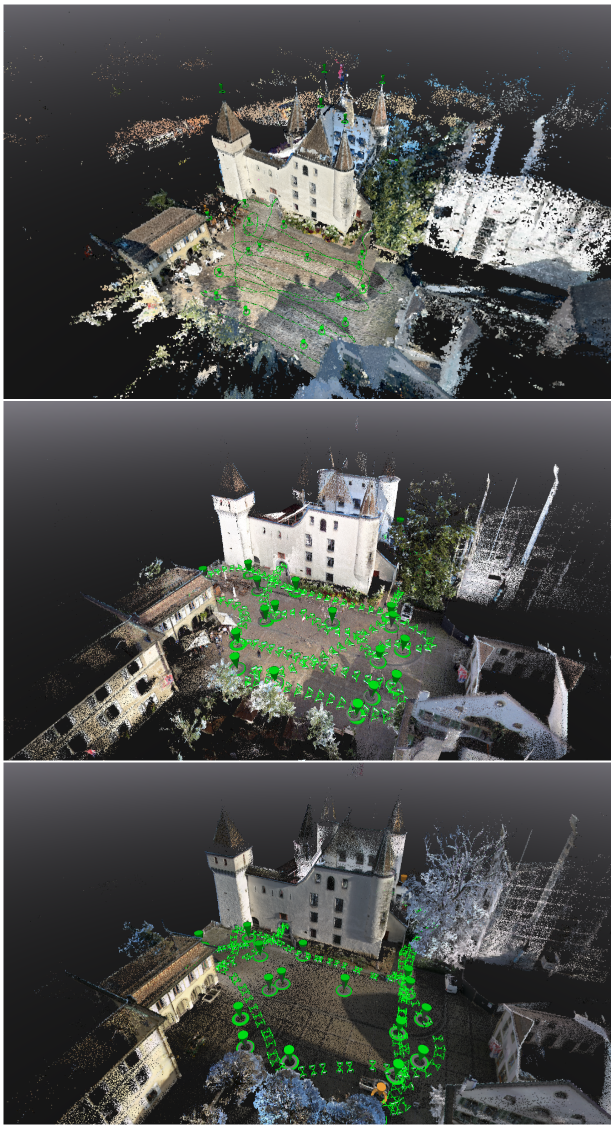

Figure 8. Resulting point cloud for Pix4D (top), SHARE S20 (middle) and XGRIDS (bottom) for the second scene. In each of those images one can see the check points and image positions in green.

SHARE scanners, while functional, produced larger errors, with maximums reaching 10 to 15 cm. In the more challenging Scene 2, the Pix4D solution's robustness was evident, as it maintained the lowest error, with a maximum of $\approx$ 12 cm, compared to $\approx$ 17 cm for XGRIDS and over 25 cm for the SHARE S20. Our PIX4D results in this work are consistent with their own earlier published results (Rehak et al., 2025, Strecha et al., 2024).

Furthermore, the Pix4D solution generated a complete point cloud even in distant areas. This is due to the software's fusion of dense photogrammetric points with the LiDAR data. This approach overcomes the iPhone's limited LiDAR range (approx. 7 m), successfully reconstructing distant objects like the tower in Scene 1 and background elements in Scene 2. This is currently a significant software advantage, though other vendors could theoretically implement similar fusion techniques.

5.2 Conclusion and Future Outlook

Handheld LiDAR scanners are developing rapidly. The integration of geometric and visual SLAM for post-processing, as seen in all three tested devices, clearly has the potential to achieve survey-grade accuracy. However, our results show this is not yet universally achieved, especially under challenging conditions.

The major bottleneck appears to be the quality and robustness of the SLAM and trajectory refinement algorithms. This is a software and algorithmic challenge that the scientific community could help solve much faster if raw sensor data were made openly available. The Pix4D solution is a notable exception, as it provides all information in its Open Photogrammetry Format (OPF) (Pix4D, 2023). The other systems rely on proprietary formats, locking users into their provided software.

The breakthrough and widespread adoption of integrated geometric and visual SLAM solutions will depend on this flexibility. Open data formats are essential. They allow for transparent and independent evaluation of accuracy and, more importantly, create the possibility for developing and testing new SLAM algorithms that are independent of any specific hardware platform.

References

Chen, Z., Li, Q., Li, J., Zhang, D., Yu, J., Yin, Y., Lv, S., Liang, A., 2022. IMU-Aided Registration of MLS Point Clouds Using Inertial Trajectory Error Model and Least Squares Optimization. *Remote Sensing*, 14(6). https://www.mdpi.com/2072-4292/14/6/1365.

El-Alailyi, A., Perfetti, L., Fassi, F., Remondino, F., 2024. V-SLAM-AIDED PHOTOGRAMMETRY TO PROCESS FISHEYE MULTI-CAMERA SYSTEMS SEQUENCES. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W4-2024, 189-195.

Emlid, 2024. Reach RX. Available at `https://emlid.com/reachrx`, Accessed: 2024-10-15.

Faro, 2025. Faro orbis product specs. Available at `https://www.faro.com/en/Products/Hardware/FARO-Orbis-Mobile-Laser-Scanner` Accessed 2025-10-15.

Graeter, J., Wilczynski, A., Lauer, M., 2018. Limo: Lidar-monocular visual odometry. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7872–7879.

Karam, S., Lehtola, V., Vosselman, G., 2020. STRATEGIES TO INTEGRATE IMU AND LIDAR SLAM FOR INDOOR MAPPING. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020, 223-230.

Keitaanniemi, A., Rönnholm, P., Kukko, A., Vaaja, M. T., 2023. Drift analysis and sectional post-processing of indoor simultaneous localization and mapping (SLAM)-based laser scanning data. *Automation in Construction*, 147, 104700.

Kuželka, K., 2025. Trajectory design for handheld mobile laser scanning in complex natural forests: a simulation approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 230, 524-546.

Leica Geosystems, 2024. BLK2GO. Available at `https://shop.leica-geosystems.com/leica-blk/blk2go/overview`, Accessed: 2024-10-15.

Lin, J., Zhang, F., 2022. R3live: A robust, real-time, rgb-colored, lidar-inertial-visual tightly-coupled state estimation and mapping package. *2022 International Conference on Robotics and Automation (ICRA)*, 10672–10678.

Macario Barros, A., Michel, M., Moline, Y., Corre, G., Carrel, F., 2022. A Comprehensive Survey of Visual SLAM Algorithms. *Robotics*, 11.

NavVis, 2025. Navvis product specs. Available at `https://www.navis.com` Accessed 2025-10-15.

Nowacki, P., Woda, M., 2020. Capabilities of arcore and arkit platforms for ar/vr applications. W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, J. Kacprzyk (eds), *Engineering in Dependability of Computer Systems and Networks*, Springer International Publishing, Cham, 358–370.

Pix4D, 2023. Available at `https://pix4d.github.io/opf-spec/`, Accessed 2025-10-15.

Pix4D, 2024. PIX4Dcatch. Available at `https://www.pix4d.com/product/pix4dcatch`, Accessed: 2024-10-15.

Pix4D, 2025. Gnss support. Available at `https://www.pix4d.com/product/rtk/#devices`.

Rehak, M., Cucci, D., Magnin, J.-B., Strecha, C., 2025. Accurate Mapping of Subterranean Structures with Mobile Phones. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-G-2025, 1269–1275.

Shan, Z., Li, R., Schwertfeger, S., 2019. RGBD-inertial trajectory estimation and mapping for ground robots. *Sensors*, 19(10), 2251.

Share, 2025. Share s20 product specs. Available at `https://www.shareuavtec.com/S20` Accessed 2025-10-15.

Strecha, C., Rehak, M., Cucci, D., 2024. The accuracy of RTK enabled mobile phones in RTK denied areas. *ISPRS Technical Commission IV Symposium 2024*, Perth, AUS.

Vidal, A. R., Rebecq, H., Horstschaefer, T., Scaramuzza, D., 2017. Hybrid, Frame and Event based Visual Inertial Odometry for Robust, Autonomous Navigation of Quadrotors. *CoRR*, abs/1709.06310. http://arxiv.org/abs/1709.06310.

xgrids, 2025. Xgrids lixel k1 product specs. Available at `https://xgrids.com/lixelk1` Accessed 2025-10-15.

Zhang, W., Wang, S., Dong, X., Guo, R., Haala, N., 2023. BAMF-SLAM: Bundle Adjusted Multi-Fisheye Visual-Inertial SLAM Using Recurrent Field Transforms.

Zhao, S., Zhang, H., Wang, P., Nogueira, L., Scherer, S., 2021. Super odometry: Imu-centric lidar-visual-inertial estimator for challenging environments. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 8729–8736.

Zhu, J., Li, H., Zhang, T., 2024. Camera, LiDAR, and IMU Based Multi-Sensor Fusion SLAM: A Survey. *Tsinghua Science and Technology*, 29(2), 415-429. https://www.sciopen.com/article/10.26599/TST.2023.9010010.

Zuo, X., Yang, Y., Geneva, P., Lv, J., Liu, Y., Huang, G., Pollefeys, M., 2020. Lic-fusion 2.0: Lidar-inertial-camera odometry with sliding-window plane-feature tracking. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5112–5119.