Mobile Phone Based Indoor Mapping

Christoph Strecha[1], Martin Rehak[1], Davide Cucci [1]

[1] Pix4D SA. Route de Renens 24 1008 Prilly, Switzerland (firstname.lastname@pix4d.com)

Abstract

We presented a mobile phone scanning solution that offers a workflow for scanning not only small spaces, where drift can be neglected, but also larger spaces where it becomes a major accuracy issue. The LiDAR and image data is combined to build 3D representations of indoor spaces. The paper does focus on the drift compensation for larger scans on the mobile phone by using AutoTags detections. We show that those can also be used to combine scans from multiple independent scans.

1. Introduction

Mobile phones are increasingly becoming an important tool in geoscience. Various studies have assessed their accuracy and applicability across dedicated industries and applications. Previous research has focused on: 1) photogrammetry techniques using mobile phone imagery (Jaud et al., 2019, Dzelzkaleja et al., 2021, Bessin et al., 2023), 2) the utilization of LiDAR sensors found in modern Apple phones (Pluta and Siemek, 2024, Guenther et al., 2024, Kottner et al., 2023, Rutkowski and Lipecki, 2023, Hakim et al., 2023), and 3) the integration of both LiDAR and photogrammetry from these devices (Barrile et al., 2022, Blaszczak-Bak et al., 2023), along with comprehensive overview research (Corradetti et al., 2022). Additionally, studies have explored combining UAV imagery with ground-based photogrammetry from mobile phones (Kovanič et al., 2023).

One major difficulty with mobile phone-based scanning is drift, especially during longer captures, which results in the accumulation of errors in relative position and orientation. These errors render the results unusable for accuracy purposes. Such inaccuracies typically manifest as duplicated surfaces in point or meshing results. To manage drift in indoor scenarios, two common methods are used: the implementation of Ground Control Points (GCPs) or a self-calibration technique that utilizes plane alignment, see for instance (Chow et al., 2013)). Those are usually applied in an offline post-processing stage.

In this study we evaluate an interesting method that is based on the automatic detection of AutoTags as shown in figure 1.

- The AutoTags are detected in real-time during the data acquisition.

- Their position does not need to be known.

- The drift compensation is done on the mobile phone after finishing the data acquisition in seconds.

- The AutoTags are coded and can be used to define an internal coordinate system, that is used to align multiple scans to each other.

In sections 2, 3 and 4 we describe the first three points and discuss the potential for combining scans in section 5.
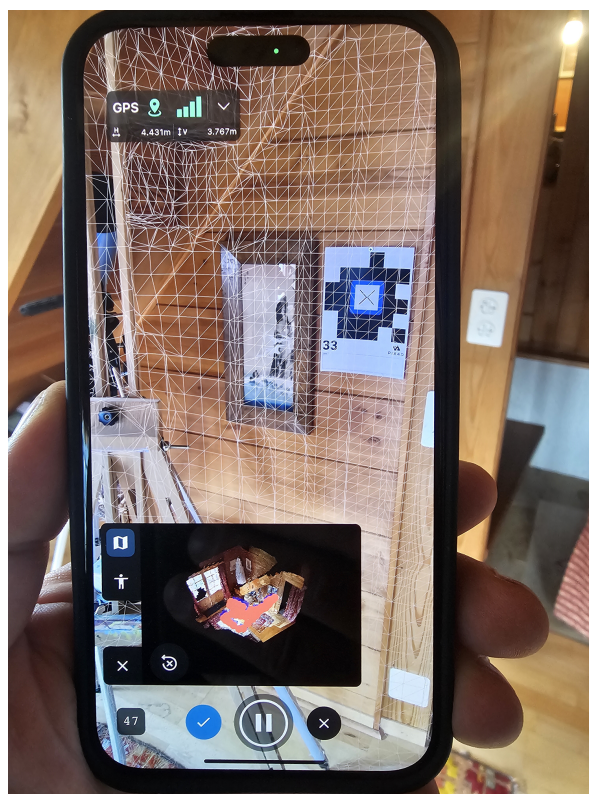

Figure 1. iPhone with PIX4Dcatch.The acquired LiDAR data of the phone can be overlaid to the current image, providing a real-time feedback on the completeness of the acquisition. AutoTags are detected in real-time (here indicated in blue.

Consumer devices, used in professional workflows, are intriguing due to their ease of use and affordability. Drones, originally designed for consumer use, are now integral to surveyors' toolboxes. Similarly, mobile phones offer an even simpler and cheaper option, allowing operation without the flight restrictions typically associated with drones.

2. Mobile Phone Tracking

PIXDcatch is a mobile phone application for iOS and Android that allows tracking the position and orientation of the phone

in its minimalistic mode. The application uses the live video feed from the camera and tracks the phone with the help of phone sensors (accelerometers, GPS, and compass) and based on image features. The underlying technology is similar to the well-known Simultaneous Localization and Mapping (SLAM). See (Macario Barros et al., 2022) for an overview.

Since the computational resources on a phone are limited, SLAM optimization usually involves windowed bundle adjustment, where the optimization runs on a window of n subsequent images. Consequently, its global accuracy is limited, and the reconstruction may suffer from drift if the sequence of images grows large. However, the results can be obtained in real-time, and the accuracy of the relative positions and orientations can be considered reasonably precise.

Let $\mathbf{p}$ be the vector of positions and orientations $\mathbf{p} = (\mathbf{X}, \omega, \kappa, \phi) = (X, Y, Z, \omega, \kappa, \phi)$ of the mobile phone camera center, i.e. the external camera parameters. Then, the SLAM procedure running in real-time inside PIX4Dcatch provide reasonable estimates for the relative change of positions and orientations $\triangle\mathbf{p}_{i,i+1}$ between frame i and $i+1$.

Given those relative changes in position and orientation, one can easily compute the position and orientation of a particular frame i:

$$\mathbf{p}_0 = (0,0,0,0,0,0) \tag{1}$$
$$\mathbf{p}_{i,i>1} = \mathbf{p}_i \times \triangle\mathbf{p}_{-1,i} \tag{2}$$

Mobile phone tracking is the first contribution that is available to compute the camera's external parameters $\mathbf{p}_i$. In the next section we are going to see how to integrate tracking with real time AutoTag detection.

3. Mobile Phone loop closure with AutoTags

The mobile phone tracking suffers from drift during long captures. AutoTags are an effective method to close loops and correct drift directly on the device. They can be detected in real-time during data acquisition (the catch) and considered together with the tracking information immediately after data collection is completed. There are 55 AutoTags available, two of them
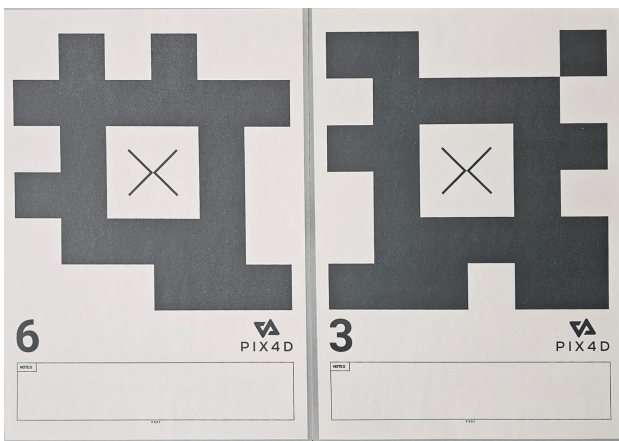

Figure 2. Two, out of 55 coded targets, called AutoTags

are show in figure 2. The tags can be placed at arbitrary positions and their 3D location does not necessarily need to be known. The ideal scenario is to place them at positions where one would regularly return, ensuring that loops in the data capture can be closed.

Let $\mathbf{m_{ij}}$ be the 2D image location of the AutoTag in frame i, where $j = 1..5$ represents the four corners of the AutoTag and its center. Let $\mathbf{X}_j$ be the unknown 3D positions of the AutoTag's corners and center. Then, we can formulate an additional constraint on the camera externals $\mathbf{p}_i$.

$$E[\mathbf{p}] = \sum_{i,j}(\mathbf{m_{ij}} - \mathbf{P}(\mathbf{p}_i)\mathbf{X}_j)^2 \tag{3}$$

with $\mathbf{P}(\mathbf{p_i})$ being the projection matrix that projects $\mathbf{X}_j$ into the image by using the current camera external parameters and the know camera internals.

4. Pix4D's Geofusion Algorithm

This algorithm runs on a mobile phone after data acquisition. Its purpose is to compute better external camera positions and orientations by integrating all available constraints, as described in Sections 2 and 3. These constraints are formulated into an energy minimization problem:

$$\begin{aligned} E[\mathbf{p}, \mathbf{X}] = {} & \alpha_t \sum_i (\mathbf{p}_i \times \triangle\mathbf{p}_{i-1,i} - \mathbf{p}_{i-1})^2 \\ & + \alpha_a \sum_{i,j}(\mathbf{m_{ij}} - \mathbf{P}(\mathbf{p}_i)\mathbf{X}_j)^2, \end{aligned} \tag{4}$$

with α_t and α_a being the weights both contributions, tracking and AutoTags. Equation 4 can be optimized efficiently on the mobile phone by non-linear optimization methods with respect to $\mathbf{p}, \mathbf{X}$. The solutions represents the optimal external camera parameters of all frames, with reduced drift if the AutoTags are placed such that loops on the acquisition are closed.

4.1 Validation

To validate the minimization approach described in Section 4, we conducted an indoor loop test using two mobile phones with identical settings, except that one phone utilized real-time AutoTag detection while the other did not. Consequently, one phone had its camera positions corrected by minimizing Equation 4, whereas the other relied solely on the tracking method detailed in Section 2. Both projects were set up in PIX4Dmatic with all AutoTags manually marked. No photogrammetry processing was applied in PIX4Dmatic, as the objective was to solely assess the Geofusion algorithm. The re-projection errors of the AutoTags were analyzed based on the camera externals provided by Geofusion, with results presented in Table 1. By integrating AutoTags into the Geofusion optimization, we significantly reduced drift, with the average re-projection error decreasing from 41 pixels to 7 pixels. This improvement was especially notable in areas with substantial initial drift, such as the locations of the first AutoTags.

With this result, the external camera positions are significantly improved, allowing the full photogrammetry processing to commence from a much better set of external camera parameters.
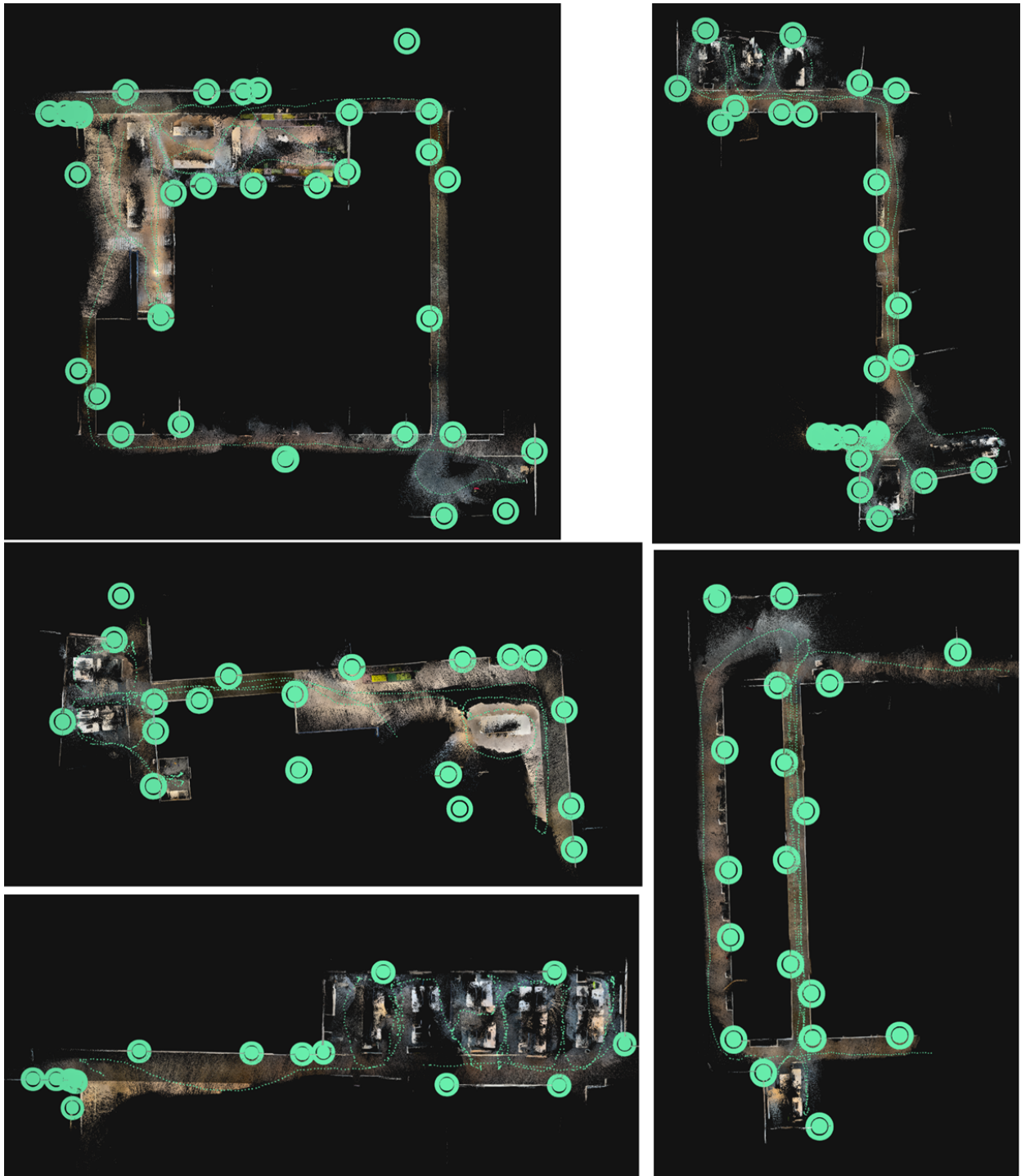
Figure 3. Screenshots from PIX4Dmatic show a 2000 m^2 office space from a top view across five different scans. In each image, one can observe the AutoTags marked in green and the iPhone LiDAR point cloud, which is colored with the images.
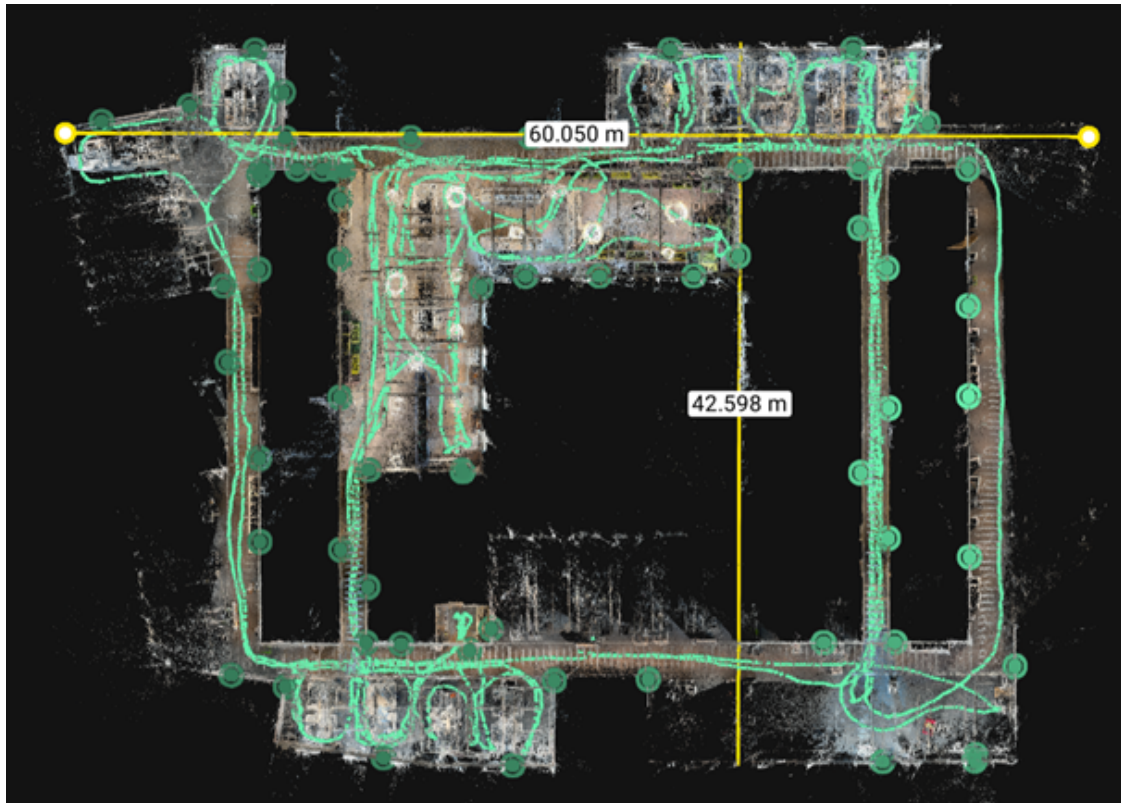
Figure 4. Screenshots from PIX4Dmatic of the merged datasets from Figure 3 are shown. One can observe the AutoTags in green, the 5000 camera positions as small green dots, the iPhone LiDAR point cloud colored with the images, and the dense photogrammetry point cloud.

5. Combining multiple phone scans by AutoTags

The advantage of the AutoTags in substantially reducing drift is a key feature of PIX4Dcatch, enabling larger indoor scans. AutoTags can also be used to combine scans for larger projects. As an example we show here two datasets: one is a $2000m^2$ office space and the other a family house. For the office space, we placed a set of 55 AutoTags around the office and acquired 5 scans with AutoTags and improved external camera parameters. As there is no GPS indoors, each of the 5 scans is in a different coordinate system. Although the 3D location of the AutoTags is unknown, we can use them to align all scans automatically, provided there are at least 3 common AutoTags between pairs of scans.

After capturing and performing geofusion, we conducted the full photogrammetric processing in PIX4Dmatic. The locations of the AutoTags - one of the outputs of the Geofusion, were used as manual tie points. The results can be seen in Figure 3. One can observe the optimized camera positions (represented by little green dots), the AutoTags with their optimized 3D locations (green circles), and the iPhone LiDAR point cloud. The reprojection error of the AutoTags is further reduced due to the full bundle adjustment, which includes automatically extracted tie points.

To obtain the full merged scan of all 5 projects, PIX4Dmatic offers the option to merge based on common tie point names, in this case, the AutoTags can be used. This process is performed sequentially. After the final merge, we optimized the entire image block via a final bundle block adjustment, used the camera positions to recompute the iPhone LiDAR point cloud,

AutoTag1	10.4	148.5
AutoTag2	9.7	127.4
AutoTag3	9.2	86.9
AutoTag4	8.8	60.6
AutoTag5	8.5	45.8
AutoTag6	6.3	19.6
AutoTag7	6	16.5
AutoTag8	5.9	14.8
AutoTag9	5.8	14.5
AutoTag10	5.7	13.3
AutoTag11	5.5	10.7
AutoTag12	4.7	7.7
AutoTag13	3.9	3.6
AutoTag14	3.3	3.5
average	6.69	40.96

Table 1. Reprojection errors of the individual AutoTags with AutoTag enabled in Geofusion (left) and without (right), as well as the average for both cases (last row).
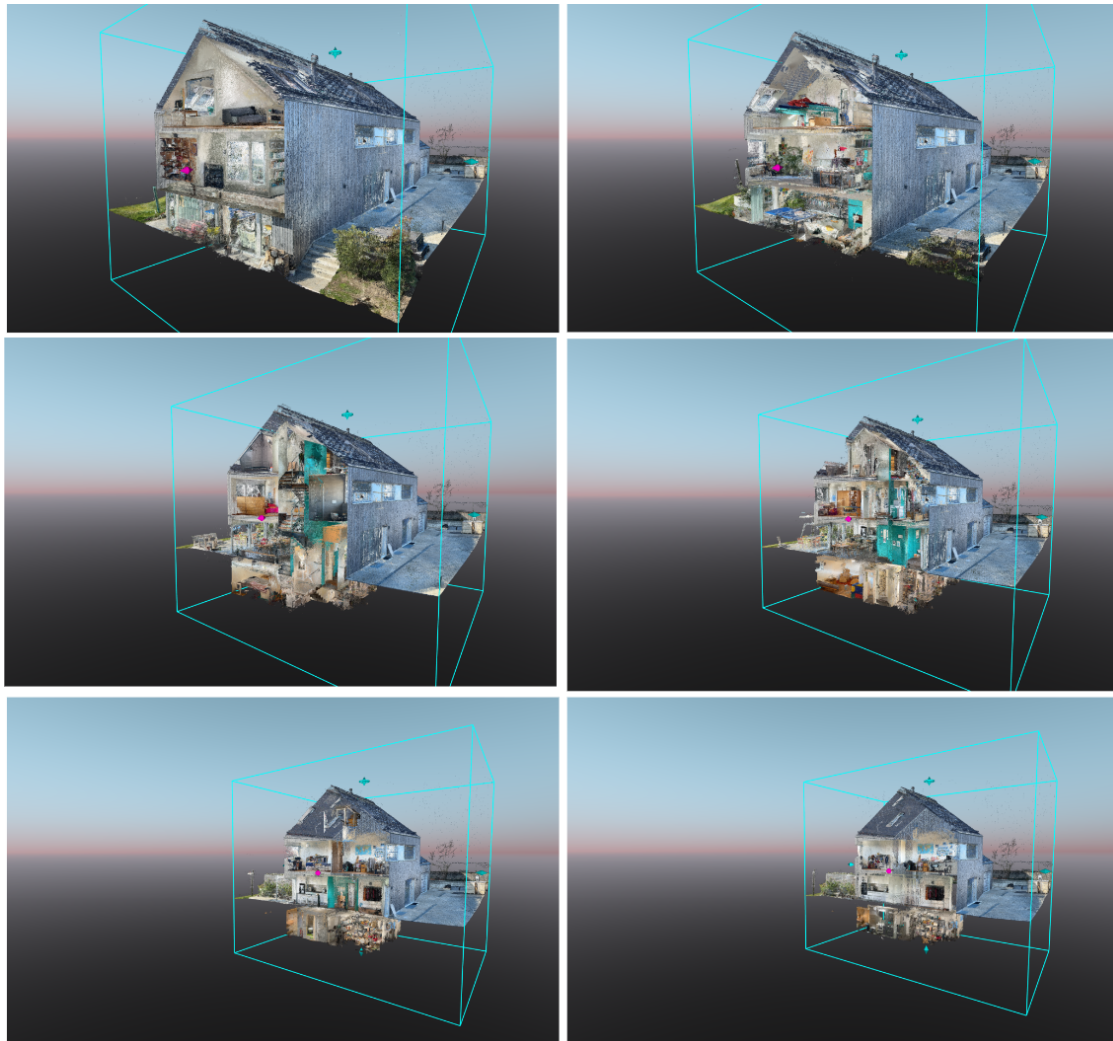
Figure 5. Screenshots from PIX4Dmatic of a house from 15 different PIX4Dcatch scans that have been merged with AutoTags. It is evident that AutoTags facilitate the modeling for both indoor and outdoor environments.

and conducted a photogrammetric dense matching. The result is shown in Figure 4, consisting of about 5000 iPhone images and LiDAR scans. This figure also indicates the size of the office space, which is about 60×42 m.

The second combined scan was conducted in a family house, where data from both the exterior and various floors inside were collected. A total of 15 PIX4Dcatch scans were merged in PIX4Dmatic after each scan was processed individually. The result is shown in figure 5. One can see six screenshots of PIX4Dmatic showcasing the final point cloud in different cutting planes.

6. Conclusion

In this article, we presented a mobile phone scanning solution that offers a workflow for scanning not only small spaces, where drift can be neglected, but also larger spaces where it becomes a major accuracy issue. It is based on combining mobile phone tracking with the use of AutoTags. We showed that these can also be used to combine scans from multiple independent projects. Future work will focus on including reference measurements and ground truth to fully assess the accuracy of the presented method.

References

Barrile, V., Bernardo, E., Fotia, A., Bilotta, G., 2022. Integration of Laser Scanner, Ground-Penetrating Radar, 3D Models and Mixed Reality for Artistic, Archaeological and Cultural Heritage Dissemination. *Heritage*, 5, 1529-1550.

Bessin, Z., Jaud, M., Letortu, P., Vassilakis, E., Evelpidou, N., Costa, S., Delacourt, C., 2023. Smartphone Structure-from-Motion Photogrammetry from a Boat for Coastal Cliff Face Monitoring Compared with Pléiades Tri-Stereoscopic Imagery and Unmanned Aerial System Imagery. *Remote Sensing*, 15, 3824.

Blaszczak-Bak, W., Suchocki, C., Kozakiewicz, T., Janicka, J., 2023. Measurement methodology for surface defects inventory of building wall using smartphone with light detection and ranging sensor. *Measurement*, 219, 113286.

Chow, J., Lichti, D., Glennie, C., Hartzell, P., 2013. Improvements to and Comparison of Static Terrestrial LiDAR Self-Calibration Methods. *Sensors (Basel, Switzerland)*, 13, 7224-7249.

Corradetti, A., Seers, T., Mercuri, M., Calligaris, C., Busetti, A., Zini, L., 2022. Benchmarking Different SfM-MVS Photogrammetric and iOS LiDAR Acquisition Methods for the Digital Preservation of a Short-Lived Excavation: A Case Study from an Area of Sinkhole Related Subsidence. *Remote Sensing*, 14, 5187.

Dzelzkaleja, L., Knēts, J., Rozenovskis, N., Sīlītis, A., 2021. *Mobile Apps for 3D Face Scanning*. 34–50.

Guenther, M., Heenkenda, M., Morris, D., Leblon, B., 2024. Tree Diameter at Breast Height (DBH) Estimation Using an iPad Pro LiDAR Scanner: A Case Study in Boreal Forests, Ontario, Canada. *Forests*, 15, 214.

Hakim, N. A., Razali, R., Mohd Said, Abdul Rahim, M., 2023. Apple LiDAR Sensor for 3D Surveying: Tests and Results in the Cultural Heritage Domain. *International Journal of Geoinformatics*, 19, 79.

Jaud, M., Kervot, M., Delacourt, C., Bertin, S., 2019. Potential of Smartphone SfM Photogrammetry to Measure Coastal Morphodynamics. *Remote Sensing*, 11, 2242.

Kottner, S., Thali, M., Gascho, D., 2023. Using the iPhone's LiDAR technology to capture 3D forensic data at crime and crash scenes. *Forensic Imaging*, 32, 200535.

Kovanič, , Štroner, M., Blistan, P., Urban, R., Boczek, R., 2023. Combined ground-based and UAS SfM-MVS approach for determination of geometric parameters of the large-scale industrial facility – Case study. *Measurement*, 216, 112994.

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

Pluta, P., Siemek, D., 2024. Possibilities and limitations of using iPhone 13 Pro with built-in LiDAR sensor in cave research -on the example of paleoflow analysis in Mylna Cave (Western Tatra Mts, Poland). *Landform Analysis*, 42, 51-62.

Rutkowski, W., Lipecki, T., 2023. Use of the iPhone 13 Pro LiDAR Scanner for Inspection and Measurement in the Mineshaft Sinking Process. *Remote Sensing*, 15, 5089.