# Fast and accurate point surveying using the PIX4Dcatch mobile app

Giulia Rovelli[1], Marta Coelho Lopes[2], Gaia A. Taberna[1], Adrian Fernandez[1], Paloma Pomares[1], Jean-Baptiste Magnin[1], Andrei Mitache[1], Davide A. Cucci[1], Christoph Strecha[1], 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:** Point surveying, Mobile application, RTK, GNSS, Sensor fusion

## Abstract

The digitalization of the architecture, construction and subsurface utility engineering sectors demands efficient, accurate and flexible 3D point surveying methods. Established ones based on Global Navigation Satellite System (GNSS) rovers or total stations suffer from significant limitations, such as requiring open-sky visibility, high costs and complex setups. This paper introduces a novel method for georeferencing 3D points using the PIX4Dcatch mobile application coupled with an external Real-Time Kinematic (RTK) GNSS receiver. The method enables to survey a point of interest by just aiming the smartphone and tapping on the screen during a capture. A lightweight, modified Bundle Adjustment algorithm runs on the device, delivering accurate 3D coordinates in seconds without any post-processing. We evaluated the method by surveying several known cadaster points for hundreds of times across diverse field conditions, achieving a mean planimetry error norm of approximately 3 cm and 97% of errors below 10 cm. Similar statistics are achieved with single-point measurements using an RTK rover. Although not intended to replace millimeter-precision instruments, the accuracy profile of our method is perfectly suited for many applications, such as subsurface utility mapping, which often have decimeter-level regulatory requirements. Given its high efficiency, low cost and ease of use, we believe that our method has the potential to transform as-built documentation workflows in diverse engineering sectors.

## 1. Introduction

The rapid expansion of digitalization in sectors such as architecture, engineering, and construction is driving an unprecedented demand for accurate 3D as-built models. One example is found in subsurface utility management and documentation, where the integration of field-surveyed data into Building Information Modeling (BIM) (Wang et al., 2019) and Geographic Information System (GIS) (Sharafat et al., 2021) platforms is becoming essential for planning, simulation, monitoring, and maintenance workflows.

A fundamental prerequisite for creating and updating these models is the ability to efficiently and accurately survey the real-world 3D coordinates of key features, assets, and control points. This demand places new pressure on traditional surveying methods, creating a need for solutions that are faster, more flexible, and accessible to a wider range of field personnel. However, acquiring these point measurements in dynamic and challenging field environments, such as dense urban canyons, active construction sites, or cluttered areas with poor satellite visibility, remains a significant operational hurdle. At the same time, not all applications require the highest level of position accuracy. For instance, at the time of writing, the accuracy required for surveying new underground installations in France is 40 cm (MEDDE, 2012), while in Germany 20 cm are recommended (BMVg, 2025). This situation forces a persistent trade-off between the required accuracy, the speed of acquisition, the cost of equipment, and the operational constraints of the job site.

One of the most widely employed instruments to measure points is the Global Navigation Satellite System (GNSS) rover: the receiver is mounted on a pole with a sharp bottom tip and a leveling indicator (e.g., a bubble), and GNSS data are recorded while the receiver is held statically on the point, either by hand

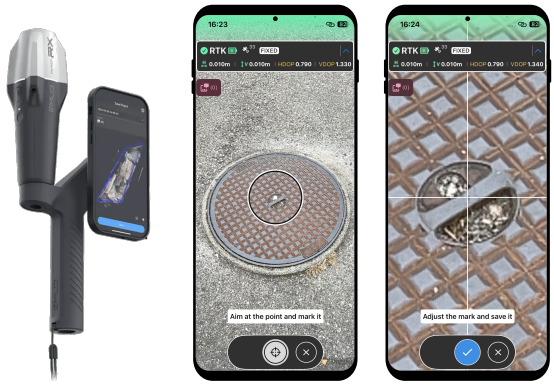


Figure 1. Left: PIX4Dcatch paired with the Emlid Reach RX RTK receiver. Center and right: Coarse and fine point marking.

or with the help of a tripod. The raw GNSS data are processed in real-time or after the capture to obtain the final 3D position of the point. Sub-centimeter accuracy can be achieved with this method if long acquisitions and careful static post-processing are performed (Leick et al., 2015). GNSS receivers with tilt compensation are also available: they employ an inertial measurement unit and a sensor fusion algorithm (Lin, 2021) to determine the orientation of the receiver and correct for any tilt of the pole. This enables the measurement of points that are not accessible directly from above, e.g., the corners of a building. However, determining the reveiver heading angle is challenging, special initialization maneuvers must be performed, and some accuracy degradation must be expected at high tilt angles (Gučević et al., 2024).

An important limitation of single-point measurements with GNSS rovers is the need for sufficient sky visibility. This requirement cannot be satisfied in many operational environments, for instance, in construction sites, underground

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-455-2026 | © Author(s) 2026. CC BY 4.0 License.

structures, and areas beneath bridges. Furthermore, in urban or cluttered environments, multipath effects (Xue et al., 2022) can severely compromise the accuracy of the measurement even if the number of satellites in view is sufficient. A method that overcomes these limitations involves the use of one of the most traditional and established survey instruments: a total station or a theodolite (Kahmen and Faig, 1988). Although in many cases total stations provide more precise results with respect to single-point GNSS observations, they are expensive, their usage requires training and experience, and the setup is nontrivial and long. Moreover, the coordinates of some other points must already be known in the area (e.g., from the cadaster service), relative to which the new points can then be measured. If none is available, one typically has to resort again to single-point GNSS measurements.

In this work, we introduce a new method for quick and accurate georeferencing of three-dimensional points. It is based on PIX4Dcatch, a mobile application that enables the creation of accurate 3D models of terrestrial assets using a conventional smartphone (PIX4D, 2020): the user scans the area of interest while the application acquires images and position fixes from an external Real-Time Kinematic (RTK) receiver, which is rigidly mounted with the phone. After the capture, the state-of-the-art Structure from Motion (SfM) algorithms developed by PIX4D are used to generate the final three-dimensional model, geo-referenced to centimeter-level accuracy thanks to the GNSS positions. Recently, the introduction of Gaussian Splatting further increased the geometric accuracy of the final point clouds and meshes, and enabled stunning photorealistic renderings (PIX4D, 2025).

The proposed method works as follows: during the capture, the user aims the phone at the Point Of Interest (POI) with the aid of a crosshair and marks it in one image, see Fig. 1. At the end of the capture, the location pinpointed by the user is automatically marked on all images by a tracking algorithm. A lightweight version of the Bundle Adjustment algorithm, including absolute and relative external control (Blázquez and Colomina, 2012b), is used to estimate accurate 3D coordinates of the POI in a few seconds. This method presents several advantages over the established solutions:

- It allows to survey points that are not accessible from above or that are located in GNSS denied areas. As we will discuss in Sec. 3.4, the error statistics degrade smoothly with respect to the distance to the closest GNSS position fix available.
- It is highly efficient and enables the user to georeference several points per minute. The acquisition throughput is primarily limited by the operator’s transit time between POIs while capturing images suitable for visual odometry and SfM algorithms. No post-processing is mandatory.
- No special setup is needed. The method requires no particular experience, background, or training.
- The cost of the hardware, as low as four thousands USD, is cheaper compared to more established surveying solutions with similar capabilities, such as total stations.

Lastly and most importantly, the method is accurate. To support this claim, we measured hundreds of points in different locations with ideal GNSS reception, obtaining a planimetry error norm below $10$ cm in $97\%$ of the cases, with a mean value of approximately $3$ cm. Furthermore, we collected points in GNSS-denied areas at a distance of up to $8$ m from an RTK fix, achieving a median error norm below $20$ cm. This result improved considerably after running the PIX4D SfM pipeline, yielding a median error norm below $7$ cm even at a distance of $8$ m from the closest RTK fix. In addition, we compared the proposed solution with single-point measurement using a GNSS rover. We obtained similar error statistics, with the GNSS rover achieving a planimetry error norm below $10$ cm in $98\%$ of the cases and a mean of approximately $2.3$ cm. Overall, these experiments demonstrate that our method achieves centimeter-level accuracy, similarly to the more traditional GNSS rover approach, while maintaining the advantages of ease of use and cost-effectiveness. It also achieves an acceptable accuracy in GNSS-denied areas.

The proposed method is described in Sec. 2, followed by the experimental evaluation in Sec. 3 and the conclusions in Sec. 4.

## 2. Instrumentation and Components

In this section, we present the core elements of the proposed method, starting from the PIX4Dcatch mobile application, the point surveying workflow and the sensor fusion algorithm that georeferences the points of interest.

### 2.1 PIX4Dcatch

PIX4Dcatch is a photogrammetry application running on a conventional smartphone. It allows the user to scan the area of interest, while it repeatedly captures images with the device camera, ensuring optimal overlap between consecutive ones. After the capture, PIX4D Structure from Motion (SfM) algorithms can be used to reconstruct a three-dimensional model of the scene, either uploading the data to PIX4Dcloud or on a desktop computer.

It is well known that SfM algorithms i) are locally precise but globally inaccurate; see, for instance, (Förstner and Wrobel, 2016, Section 15.3.4) and ii) cannot provide georeferenced models since no absolute position information can be extracted from images alone. One way to solve both issues is to introduce camera position priors coming from a GNSS receiver. This approach is now customary in aerial photogrammetry and allowed to eliminate or greatly reduce the need for ground control points (Rehak and Skaloud, 2016). Although the phone internal GNSS receiver can partly mitigate the second issue, it is not sufficient to solve the first. Therefore, PIX4Dcatch is typically coupled with an external RTK-enabled GNSS receiver rigidly mounted to the phone. In this study, the receiver used is the Emlid Reach RX, visible in Fig. 1. The phone connects to an NTRIP server via the mobile broadband connection and feeds real-time corrections to the receiver, so that RTK processing can be done onboard and the position of the antenna can be determined within few centimeters (Hamza et al., 2023).

Unfortunately, in terrestrial environments, the availability and the reliability of position fixes from an RTK receiver, especially in the case of a low-cost one, are not always satisfactory: the sky visibility may be limited by surrounding structures, causing fix loss or degraded dilution of precision. Furthermore, multipath effect may hinder the quality of the position fix, introducing shifts or outliers. Therefore, in PIX4Dcatch, the information coming from the GNSS receiver

is not used as is, but it is first integrated with other sensor information available on the phone, thanks to the `geofusion` sensor fusion algorithm (Strecha et al., 2024). The position and orientation of the camera determined by `geofusion` are used by the PIX4D SfM pipeline as position and orientation control, as in the Integrated Sensor Orientation (ISO) approach (Schwarz et al., 1993), allowing to produce accurately georeferenced 3D models.

The on-device sensor fusion step implemented in `geofusion` allows to i) detect and correct for GNSS outliers, ii) bridge gaps in GNSS position fixes, e.g., to map underground structures with centimeter-level accuracy (Rehak et al., 2025) and iii) georeference user-marked POIs directly on the phone, which is the subject of this work. The `geofusion` algorithm will be presented in detail in Sec. 2.4.

### 2.2 Point surveying workflow in PIX4Dcatch

In the following, we discuss in detail the functioning of the point surveying feature of PIX4Dcatch introduced in this work, from a user perspective.

In order to survey points the user proceeds as follows:

1. **Start the capture.** PIX4Dcatch starts acquiring images, ensuring optimal overlap and coverage of the scene content while the user moves around.
2. **Walk to one point of interest.**
3. **Activate the point surveying functionality.** A crosshair appears on the screen, see Fig. 1 center. The user aims at the POI and confirms. Next, a static zoomed-in image is presented on the screen, allowing the user to make a mark with pixel precision by dragging the crosshair.
4. **Resume the capture.** The user moves around the point, keeping it in the field of view so that further images with sufficient parallax can be acquired.
5. (optional) **Mark the POI additional times**. It is possible to mark the same point multiple times. This provides more seeds to the automated object detector, see Sec. 2.3.
6. (optional) **Acquire more POIs**. Go back to step 2.
7. **Stop the capture.** Automated object tracking is performed, see Sec. 2.3, followed by the `geofusion` algorithm, see Sec. 2.4. The 3D coordinates of the points are made available within seconds and can be exported.
8. (optional) **Run SfM algorithms**. The capture can be post-processed either in PIX4Dcloud or on a desktop computer. Apart from yielding a full 3D model of the scene (e.g., a point-cloud), the additional photogrammetric processing can, in certain situations, yield more accurate coordinates for the surveyed POIs, see Sec. 3.4.

### 2.3 Automated Point of Interest tracking

In this section, we briefly discuss how an automated object detector algorithm can be used to track a POI in all available images, starting from one (or multiple) image(s) manually marked by the user.

After the capture, PIX4Dcatch attempts to find the precise location of the POIs marked by the user in all acquired images,

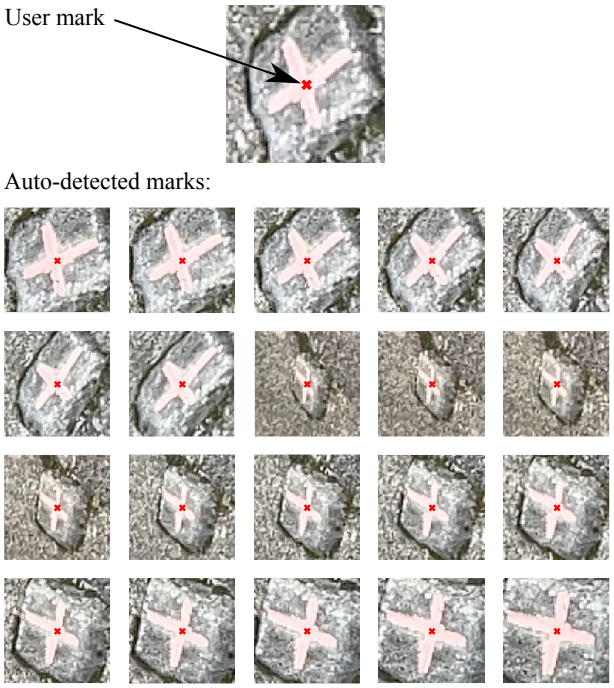


Figure 2. Example of a POI marked by the user, with the auto-detected marks from the object tracking algorithm.

similarly to what is done in the keypoint detection and matching step of an SfM algorithm. In particular, an object detector is seeded with the image patch around each user mark and it is run on all images predicted to contain the POI, based on the initial solution for the camera poses. The precise image coordinates in the new images are those that maximize the cross-correlation of the corresponding patch in the query image and the input patch (around the user mark).

The object detector is very robust to perspective, illumination and scale change. Please see Fig. 2 for an example of a POI used in the study, along with (some of) the automated detections found by the tracking algorithm.

### 2.4 The `geofusion` algorithm

In the following, we present in detail the mathematical formulation behind the `geofusion` algorithm. We will clarify how the 3D coordinates of the user-marked POI can be determined with great accuracy, together with the camera poses (the geotags): this is done by fusing GNSS observations, automatically extracted image measurements (in this study, the POI marks) and the real-time localization provided by an onboard visual-inertial odometry (VO) algorithm.

Let us first define some important 3D reference frames used in the following. $w$ is a user-selected projected Coordinate Reference System (CRS). A shift is applied such that the origin is moved near the area of operation. All output coordinates are computed with respect to $w$. On the other hand, $\tilde{w}$ is an arbitrary CRS with respect to which the real-time visual-inertial odometry algorithm provided by the phone manufacturer localizes[1]. Finally, the camera frame $c$ is placed at the optical center of each camera.

At the end of the capture, the following streams of partially redundant measurements are available:

[1] On Apple devices, PIX4Dcatch relies on ARKit (Apple, 2017).

1. Images, and, optionally, the depth maps derived from the LiDAR sensor on the phone. Images are indexed with $i$.
2. For each image $i$, the position fix from the external GNSS receiver, expressed with respect to $w$, $\text{GNSS}_i^w$.
3. For each image $i$, the pose of the camera provided in real-time by the onboard VO framework, $\text{VO}_{c,i}^{\tilde{w}}$.
4. For each user-marked POI $j$ and for each image $i$ in which this point has been detected, the image coordinates of the point, $[u, v]_{j,i}$, in pixels.

The 3D coordinates of the user-marked POI, along with the camera geotags, are computed by `geofusion` after the capture by solving a non-linear robust least-squares problem, similarly as in (Cucci et al., 2017). The unknowns in this problem are:

- $\Gamma_{c,i}^w \in \text{SE}(3)$, the pose of the $i$-th camera in $w$,
- $p_j^w \in \mathbb{R}^3$, the position of the $j$-th POI in $w$.

For each of the available measurements we define a residual, relating the observed value with the problem unknowns. For the GNSS position measurements, the residual reads as:

$$r_i^{\text{GNSS}} = \Gamma_{c,i}^w \, l^c - GNSS_i, \tag{1}$$

where $l^c$ is the GNSS antenna lever-arm in camera frame.

The onboard visual-inertial odometry framework does not provide absolute positioning in $w$, but only in $\tilde{w}$, and it is subject to non-negligible drift, therefore we extract only differential information from $\text{VO}_{c,i}^{\tilde{w}}$ as follows:

$$r_i^{\text{VO}} = \log\left[\text{VO}_{c,i}^{\tilde{w}} \left(\Gamma_{c,i}^w\right)^{-1} \Gamma_{c,i+1}^w \left(\text{VO}_{c,i+1}^{\tilde{w}}\right)^{-1}\right], \tag{2}$$

where $\log[\cdot]$ is the logarithmic map in $\text{SE}(3)$.

Finally, the POI marks are modeled with the usual collinearity equations:

$$\begin{bmatrix} r_{j,i}^{\text{M}} \\ 0 \end{bmatrix} = \lambda K_i \left[\Gamma_{c,i}^w\right]^{-1} p_j^w - \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}_{j,i}, \tag{3}$$

where $K_i$ is the camera internal calibration matrix for camera $i$ and $\lambda$ is obtained from the third component and eliminated, as usual. Note that in this context, $K_i$ is made available in real-time by the VO framework, accounting for auto-focus and image stabilization. Of course, real-time values are refined in case PIX4D SfM pipeline is run after the capture.

The solution for the camera poses and the 3D points is obtained by minimizing the residuals $r_i^{\text{GNSS}}$, $r_i^{\text{VO}}$ and $r_{j,i}^{\text{M}}$ as a function of the unknowns $\Gamma_{c,i}^w$ and $p_j^w$ while accounting for their assumed probability distribution, often heavy-tailed, see (Triggs et al., 2000, Section 3.3). The solution uncertainty is computed from the Hessian of the least-squares problem.

The approach presented in this section consists of a modified Bundle Adjustment with both relative and absolute position control and relative orientation control, see (Blázquez and Colomina, 2012b). It also shares many ideas with Fast AT (Blázquez and Colomina, 2012a), in the sense that a Bundle Adjustment is run, but very few image measurements are available, and the camera poses are constrained, at least in a relative sense, by strong position and orientation priors.

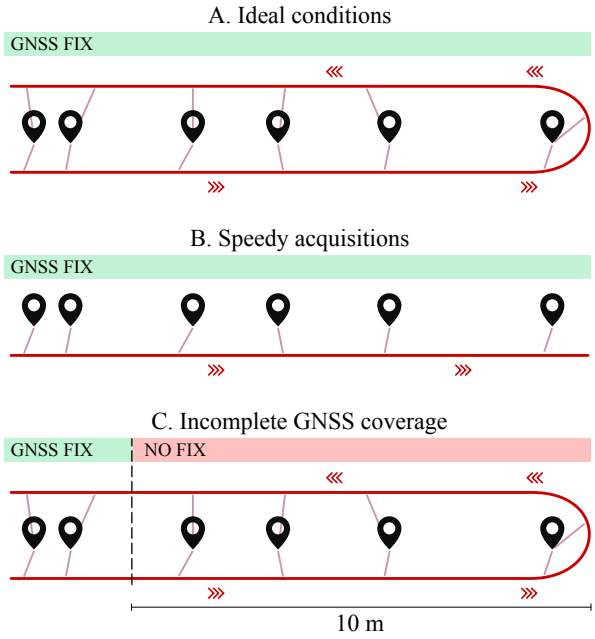


Figure 3. Illustrations of the three acquisition scenarios. The red line represents the acquisition path; the black pins represent the points of interest (POIs) surveyed; each pink segment shows the position on the path where a given POI is marked.

## 3. Experimental evaluation

In this section, we present the results of an extensive evaluation of the accuracy of the PIX4Dcatch workflow. First, we briefly describe the general methodology and how we assess measurement accuracies; then, we present three different experimental setups and discuss the main findings; finally, we compare our method with the widely adopted static single-point observation technique.

### 3.1 Methodology

In order to assess the accuracy of the presented method, we perform extensive field evaluation by simulating a typical workflow of a PIX4Dcatch user interested in asset surveying (e.g., pipes and valves in a trench, property boundaries, building footprints). We measure several known POIs and compare the estimated coordinates with the reference ones. Specifically, we consider 38 cadastral points physically signalized on the ground with bolts or crosses, as in Fig. 2, belonging to 4 different sites in the Lausanne (CH) area. The coordinates of these points have been established by the local cadaster authority (Canton de Vaud, 2014).

We conduct a total of 83 acquisitions, surveying each site on different days and at different times of the day, between 5 and 15 points per site. For each POI $j$ in acquisition $k$, we compare the estimated coordinates $\hat{p}_{j,k}$ with the reference coordinates $p_j$ from the cadaster. Inevitably, our estimations are affected by user error arising from inconsistencies in POI marking, which we have minimized as much as possible. As altitude data are not available for every point in the study, we only consider planimetry coordinates. The error is simply defined as:

$$e_{j,k} = \hat{p}_{j,k} - p_j. \tag{4}$$

Additionally, we can also define the Normalized Estimation

Error Squared (NEES) as:

$$z_{j,k} = e_{j,k}^T (\hat{\Sigma}_{j,k} + \Sigma_{\text{ref},j})^{-1} e_{j,k} \tag{5}$$

where $\hat{\Sigma}_{j,k}$ is the uncertainty of point $j$ in acquisition $k$ as estimated by `geofusion` and $\Sigma_{\text{ref},j}$ is the uncertainty of the reference coordinates for point $j$. The quantity $z_{j,k}$ provides another way to assess the quality of the results, since it expresses the error of a measurement relative to the estimated uncertainty. Large normalized errors help identify outliers or issues in the uncertainty estimation.

Our evaluation aims at establishing the error statistics for the asset surveying workflow. We consider the three different acquisition scenarios depicted in Fig. 3:

- **Ideal conditions** (Sec. 3.2). The RTK-corrected GNSS position is largely available, and each POI is marked twice while following a U-loop acquisition path.
- **Speedy acquisitions** (Sec. 3.3). The conditions are similar to the previous scenario, but each POI is marked only once during the capture, thus speeding up the workflow.
- **Incomplete GNSS coverage** (Sec. 3.4). The RTK corrected GNSS position is only available at the beginning and at the end of the capture, while in between the RTK fix is lost; POIs are marked twice. This scenario simulates, e.g., surveying points in a tunnel.

Lastly, we compare the surveying workflow using PIX4Dcatch with the single-point measurement workflow using a GNSS rover (see Sec. 3.5). Nowadays, this is the only alternative technique available that is coarsely comparable in terms of time and cost to the proposed method.

### 3.2 Ideal conditions

In the ideal conditions scenario, we mark each POI twice, from different viewing angles, and we keep the point in the field of view as much as possible during the acquisition, so that the automated object detector can then track it in many images with different perspectives. The capture follows a U-loop path, illustrated in Fig. 3A: we survey one POI after the other, then turn back and survey all POIs once again in the opposite order and from a different perspective. In total, we perform 56 acquisitions (7 to 18 acquisitions per location) and survey 435 points. Fig. 4 shows the distribution of the component-wise errors for the coordinates estimated by `geofusion`, with a standard deviation of approximately 2.5 cm on both the $x$- and $y$-axis. We expect the $z$ component to be slightly less accurate, as happens in general for GNSS position measurements.

We also post-process all the datasets on the PIX4Dcloud, therefore running the full SfM pipeline. Fig. 5 compares the distributions of the error norms, with and without post-processing. In this ideal conditions scenario, we observe minimal differences between the two distributions: the results obtained from the `geofusion` algorithm have a mean error norm of 3.0 cm, with 97% of the points having an error norm below 10 cm; after post-processing, the mean error norm slightly decreases to 2.9 cm, with 98% of points with an error norm below 10 cm. These results show that the proposed method is sufficiently accurate for a wide range of surveying applications, as discussed in Sec. 1: the centimeter-level accuracy achieved with PIX4Dcatch satisfies the requirements for urban asset surveying, and, furthermore, our solution is highly time and cost effective.

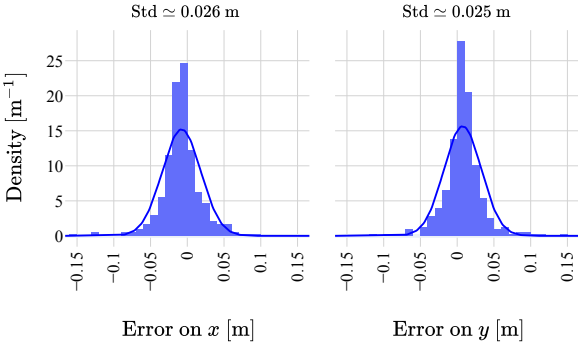


Figure 4. Histograms of the errors on the $x$ and $y$ coordinates with a normal distribution with sample mean and variance.

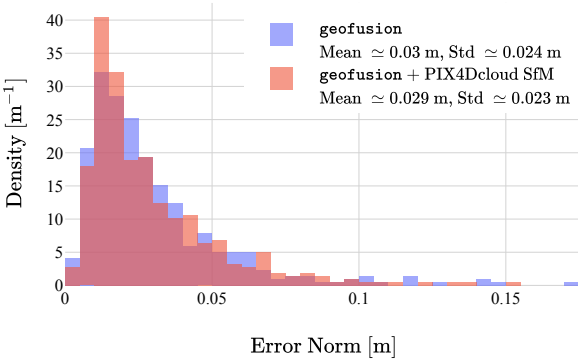


Figure 5. Histograms of the norms of the planimetry errors, before (blue) and after (red) PIX4Dcloud SfM post-processing.

Since the `geofusion` algorithm also estimates the uncertainties of the POI coordinates, we are able to compute the normalized planimetry errors $z_{j,k}$ as in Equation 5. The *true* reference uncertainty $\Sigma_{\text{ref},j}$ is unknown and is provisionally set to zero. Fig. 6 compares the empirical distribution of the normalized errors obtained with `geofusion`, with the theoretical $\chi^2_2$ distribution that they are expected to follow (Bar-Shalom et al., 2001, Section 5.4). Globally, the empirical histogram matches the theoretical curve. The fact that this close fit is achieved while assuming $\Sigma_{\text{ref},j} = 0$ suggests that `geofusion` slightly overestimates the uncertainty of the POI coordinates, as this estimated uncertainty is effectively also accounting for the missing reference uncertainty. Additionally, the empirical distribution contains an excess of large values, which correspond to samples for which the error is large relative to its estimated uncertainty. Quantile analysis confirms this observation: the theoretical 0.68-quantile of the $\chi^2_2$ distribution (corresponding to $z \approx 2.28$) covers only 44% of our data, instead of the expected 68%. Similarly, the 0.99-quantile ($z \approx 9.2$) covers only 82% of the data. Although the overall deviation is small, this discrepancy suggests that the distribution of the error $e_{j,k}$ is slightly heavier-tailed than the Gaussian distribution.

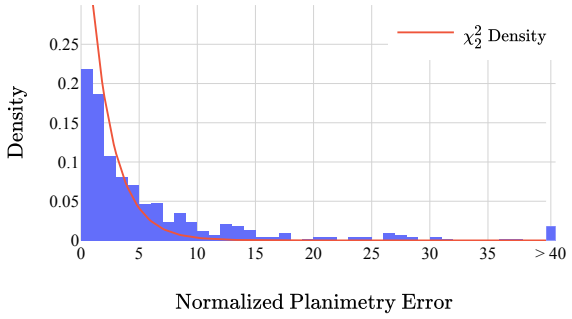


Figure 6. Histogram of the normalized planimetry errors. The density of a $\chi^2_2$ is also shown in red.

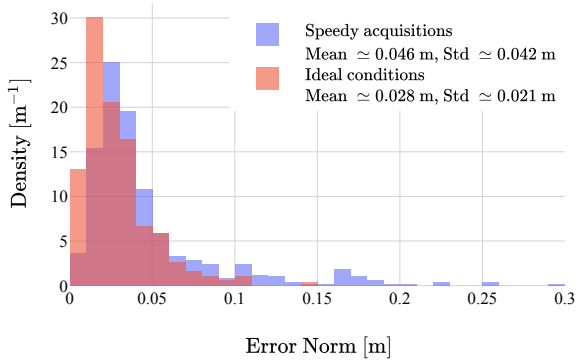


Figure 7. Histograms of the norms of the planimetry errors in the ideal conditions (red) and speedy acquisitions (blue) scenarios, obtained with `geofusion`.

### 3.3 Speedy acquisitions

In this scenario, we reproduce the workflow of a user who needs to survey as many assets as possible in a short amount of time. As illustrated in Fig. 3B, this can be achieved by marking POIs only once, without coming back to mark them a second time. To mimic such workflow, we reprocess the same 56 datasets as in Sec. 3.2, but we consider only one mark per POI at a time. In principle, this gives us twice as many point measurements as in the previous use case. Since the original acquisition follows a U-loop path, it can happen that, after feeding our automated object detector with a user mark, the related POI is also tracked on the opposite side of the U-path. If this is the case, we discard the POI for that acquisition, ending up with a total of 492 measurements. Fig. 7 compares the norm of the planimetry error obtained with `geofusion`, when the POIs are marked twice from different perspectives, as in the ideal scenario, or only once for a faster workflow. Notably, the mean and the variability of the error norm are slightly higher in the single-mark case, and more outlier measurements are present: when marking once we obtain error norms below 10 cm for 90% of the samples, compared to 99% of the samples in the case of double-marked POIs.

Having a closer look at the general outcome of our automated algorithm for object tracking (see Sec. 2.3), starting from either one or two user marks per POI, we can identify an important factor that influences the reliability of the results: the angular occupancy of the POI marks, i.e. the range of viewing angles where the POI is tracked along the acquisition path. For example, two extreme cases of angular occupancy

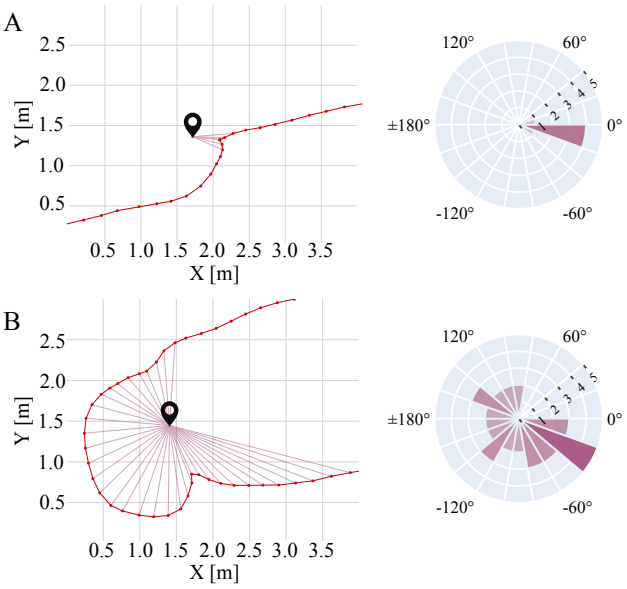


Figure 8. Examples of point of interest (POI) tracking. (A) The POI is tracked on a few adjacent images (left), with narrow angular occupancy of its marks (right). (B) The POI is tracked on several images (left), with wide angular occupancy (right).

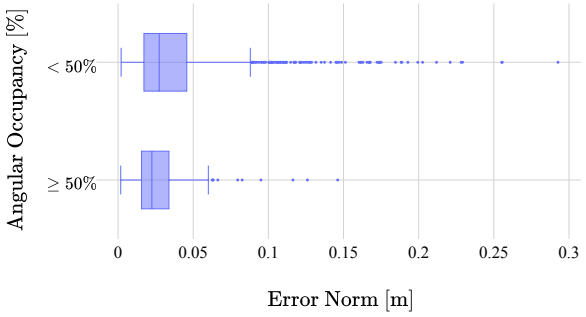


Figure 9. Box plots of the norms of the planimetry errors, considering two levels of the angular occupancy of the marks.

are represented in Fig. 8. When a POI is only visible in a few adjacent images in the capture, as depicted in Fig. 8A, the object tracking algorithm can span a narrow range of viewing angles; in contrast, when the POI is kept in the field of view for longer time, therefore spanning a wider range of viewing angles as in Fig. 8B, our algorithm can track it in more images with sufficient parallax. To investigate the impact of the angular occupancy on the accuracy of the measurements, we consider all the collected POIs with either one or two user marks, including the POIs that the automated object detector can track on both sides of the U-path. This yields a total of 1280 point measurements. Fig. 9 shows that a wider angular occupancy statistically gives more reliable measurements: for example, when the angular occupancy is less than 50%, the median error norm is about 2.7 cm, with 92% of the samples having an error below 10 cm. In contrast, when the angular occupancy is at least 50%, the median error norm decreases to 2.2 cm, with an error below 10 cm for 99% of the samples. This result is fully expected from photogrammetric theory.

We can conclude that keeping the POI in the field of view while doing the acquisition helps improve the accuracy of the georeferencing. This can be achieved, for instance, by walking

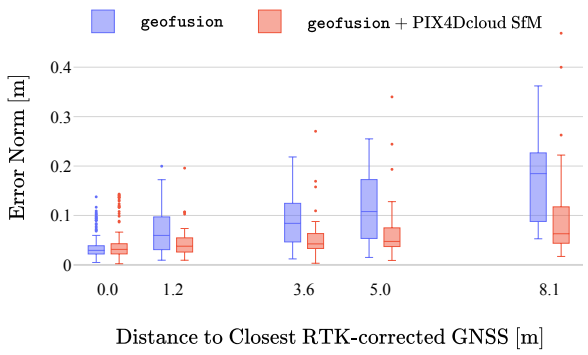


Figure 10. Box plots of the norms of the planimetry errors, at different distances of the point of interest to the closest RTK-corrected GNSS sample, before (blue) and after (red) PIX4Dcloud SfM post-processing.

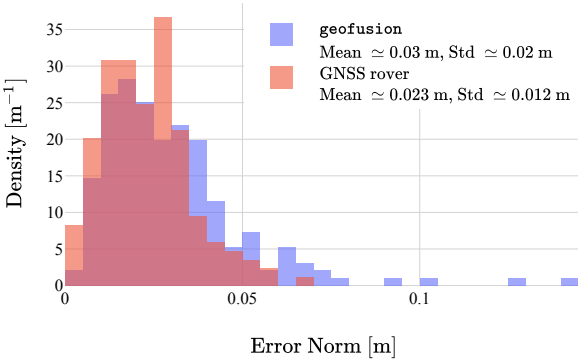


Figure 11. Histograms of the norms of the planimetry errors after `geofusion` processing (blue) and after single-point measurement using a GNSS rover (red).

around the point to capture a change in perspective and a sufficient diversity in viewing angles. Nevertheless, our study shows that a smaller parallax does not necessarily lead to a substantial degradation of error statistics.

### 3.4 Incomplete GNSS coverage

In the third scenario, we simulate asset surveying with limited sky visibility, e.g. when mapping under a bridge, through a tunnel, or under canopy. The aim is to characterize how far a user can extend in GNSS denied areas and still obtain acceptable error statistics. We choose one of the four locations of the study where 8 cadastral points are spread within approximately 20 meters, and we design the experiment as shown in Fig. 3C: following a U-loop path, we start the acquisition in ideal conditions with RTK-corrected GNSS signal and we mark 4 POIs; afterwards, we cover the GNSS antenna with a metal cap to prevent the RTK fix, and we continue marking the remaining POIs one after the other, turning back and marking once more in the opposite direction; finally, when we reach again the location where we covered the antenna, we remove the metal cap to acquire the RTK fix again, and we mark for the second time the last few POIs. We repeat the capture 27 times, for a total of 216 points, half of which are surveyed in GNSS denied conditions. Fig. 10 shows the statistics for the norm of the planimetry error as a function of the distance to the closest RTK-corrected GNSS sample. As expected, the results degrade smoothly with distance, ranging from a median error norm of 6.0 cm at 1.2 m distance, to 8.4 cm at 3.6 m, and 18.5 cm at 8.1 m before post-processing.

The observed trend in error statistics can be explained by the behavior of the `geofusion` sensor fusion algorithm. In the absence of GNSS fix, the algorithm can rely only on the poses estimated by the onboard visual-inertial odometry framework and on the few marks collected by the user. However, these poses are prone to drift during the capture, given the sequential and real-time nature of the VO framework itself. In contrast, the errors are expected to improve when running the full photogrammetric post-processing. Fig. 10 also shows the error statistics after processing each dataset on PIX4Dcloud, running the full proprietary SfM pipeline. As expected, in this case the degradation of the statistics is substantially lower, with a median error norm dropping to 6.3 cm at 8.1 m.

We conclude that our method would meet the requirements of the applications that do not need centimeter-level accuracy, even when the user partially extends in GNSS-denied areas. In the cases where the precision achieved by the sole `geofusion` algorithm is not sufficient, one can take advantage of the full automated post-processing provided by PIX4Dcloud or desktop solution to obtain an improved accuracy. This supports our claim that one of the main advantages of the PIX4Dcatch solution is the possibility to also operate in areas without full open sky visibility.

### 3.5 Comparison with static single-point observations

In the last study, we mount the Emlid Reach RX RTK-GNSS antenna on a pole and record single-point measurements while holding the pole statically on each POI for 15 seconds, as described in Sec. 1. We also ensure that the receiver reports an RTK fix with optimal accuracy before starting the recording. We measure 13 cadastral points among two of the chosen locations, repeating the collection 13 times on different days and times of the day, for a total of 169 measurements. We compare these measurements with 191 measurements collected with PIX4Dcatch in the same two locations (34 acquisitions in total). As previously discussed in Sec. 1, this method suffers from some limitations, such as the need for open sky visibility and the requirement to ensure proper leveling of the pole, while carefully performing a static point collection. On the contrary, these limitations do not affect the surveying workflow using the PIX4Dcatch mobile application.

Fig. 11 shows the distribution of the norm of the planimetry error for the two techniques. It is possible to see that both solutions provide centimeter-level accuracy: a mean error of about 3 cm is achieved with the `geofusion` algorithm, compared to a mean error of about 2.3 cm in the case of static single-point measurement. However, `geofusion` presents some more outliers compared to the GNSS rover approach; moreover, while in this experiment the rover always gave errors below 10 cm, `geofusion` gives 1.6% of errors larger than 10 cm. Our solution has the advantage of being more time-effective and easier to use for non-professional users, requiring almost no setup for asset surveying. Moreover, PIX4Dcatch allows one to survey points in hard-to-reach locations, such as a point on a wall or an asset in a trench.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-455-2026 | © Author(s) 2026. CC BY 4.0 License.

## 4. Conclusions

In this work, we introduced a novel method for accurately georeferencing 3D points using the PIX4Dcatch mobile application together with an external RTK-enabled GNSS receiver. Unlike conventional surveying with a GNSS rover alone, the proposed approach enables the acquisition of points that cannot be accessed vertically from above. Moreover, the method allows surveying in GNSS-denied environments, provided that position fixes can be obtained at another stage during data acquisition. The accuracy of the proposed method has been rigorously and extensively evaluated, demonstrating that, while it cannot fully replace total stations, its error statistics satisfy the requirements of all but the most demanding georeferencing applications. Owing to its flexibility, ease of use, and accuracy, we believe that the proposed method can become a cornerstone technology in the digital transformation of construction and infrastructure management, enabling scalable, high-throughput workflows that are beyond the reach of traditional surveying techniques.

## References

Apple, 2017. ARKit - Augmented Reality. Available at https://developer.apple.com/augmented-reality/arkit.

Bar-Shalom, Y., Li, X. R., Kirubarajan, T., 2001. *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons.

Blázquez, M., Colomina, I., 2012a. Performance Analysis Of Fast At For Corridor Aerial Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B1, 97–102. 10.5194/isprsarchives-XXXIX-B1-97-2012.

Blázquez, M., Colomina, I., 2012b. Relative INS/GNSS aerial control in integrated sensor orientation: Models and performance. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 120-133. 10.1016/j.isprsjprs.2011.11.003.

BMVg, 2025. A-5.1 ortung unterirdischer leitungen. Section “A-5.1 Ortung unterirdischer Leitungen” of *BFR Vermessung – Bauwerks- und Freianlagen-Richtlinie Vermessung*. https://www.bfrvermessung.de/bfr-verm/anlagen/a-5-fachspartenspezifische-anforderungen/a-51-ortung-unterirdischer-leitungen.

Canton de Vaud, 2014. Guichet cartographique cantonal. Available at https://www.geo.vd.ch.

Cucci, D. A., Rehak, M., Skaloud, J., 2017. Bundle adjustment with raw inertial observations in UAV applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 1-12. 10.1016/j.isprsjprs.2017.05.008.

Förstner, W., Wrobel, B. P., 2016. *Photogrammetric computer vision*. 6, Springer.

Gučević, J., Delčev, S., Vasović Šimšić, O., 2024. Practical Limitations of Using the Tilt Compensation Function of the GNSS/IMU Receiver. *Remote Sensing*, 16(8). https://doi.org/10.3390/rs16081327.

Hamza, V., Stopar, B., Sterle, O., Pavlovčič-Prešeren, P., 2023. Low-Cost Dual-Frequency GNSS Receivers and Antennas for Surveying in Urban Areas. *Sensors*, 23(5). 10.3390/s23052861.

Kahmen, H., Faig, W., 1988. *Surveying*. 3rd edn, De Gruyter, Berlin, Boston.

Leick, A., Rapoport, L., Tatarnikov, D., 2015. *GPS Satellite Surveying*. John Wiley & Sons.

Lin, H., 2021. High-precision RTK positioning with tilt compensation: Data fusion algorithm. *Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021)*, 2681–2695. 10.33012/2021.17946.

MEDDE, 2012. Arrêté du 15 février 2012 [...] relatif à l’exécution de travaux à proximité de certains ouvrages souterrains, aériens ou subaquatiques de transport ou de distribution. Journal officiel de la République française, n° 0417, 16 février 2012, texte n° 3. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000025391351/.

PIX4D, 2020. PIX4Dcatch. Available at https://www.pix4d.com/product/pix4dcatch.

PIX4D, 2025. Gaussian splatting at PIX4D: a new era of 3D visualization. https://www.pix4d.com/blog/pix4d-gaussian-splatting-3D-visualization.

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. 10.5194/isprs-archives-XLVIII-G-2025-1269-2025.

Rehak, M., Skaloud, J., 2016. Applicability Of New Approaches Of Sensor Orientation To Micro Aerial Vehicles. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-3, 441–447. 10.5194/isprs-annals-III-3-441-2016.

Schwarz, K.-P., Chapman, M., Cannon, M., Gong, P., 1993. An integrated INS/GPS approach to the georeferencing of remotely sensed data. *Photogrammetric Engineering & Remote Sensing*.

Sharafat, A., Khan, M. S., Latif, K., Tanoli, W. A., Park, W., Seo, J., 2021. BIM-GIS-Based Integrated Framework for Underground Utility Management System for Earthwork Operations. *Applied Sciences*, 11(12). 10.3390/app11125721.

Strecha, C., Rehak, M., Cucci, D., 2024. Mobile Phone Based Indoor Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2-2024, 415–420. 10.5194/isprs-archives-XLVIII-2-2024-415-2024.

Triggs, B., McLauchlan, P. F., Hartley, R. I., Fitzgibbon, A. W., 2000. Bundle adjustment — a modern synthesis. *Vision Algorithms: Theory and Practice*, Springer Berlin Heidelberg, Berlin, Heidelberg, 298–372.

Wang, M., Deng, Y., Won, J., Cheng, J. C., 2019. An integrated underground utility management and decision support based on BIM and GIS. *Automation in Construction*, 107, 102931. 10.1016/j.autcon.2019.102931.

Xue, Z., Lu, Z., Xiao, Z., Song, J., Ni, S., 2022. Overview of multipath mitigation technology in global navigation satellite system. *Frontiers in Physics*, 10, 1071539. 10.3389/fphy.2022.1071539.