Hybrid Calibration between a Laser Scanner and Smartphone Camera Using hourglass targets and Deep Learning

Lorenz Krefft[1], Ludwig Hoegner[2]

Department of Geoinformation, Munich University of Applied Sciences, Munich, Germany
[1]lorenz.krefft@hm.edu , [2] ludwig.hoegner@hm.edu

Keywords: Sensor data fusion, Laser scanner, Laser Scanner Camera Calibration, Neural Networks, Time synchronization.

Abstract

This paper presents a novel hybrid calibration pipeline that jointly estimates the spatial and temporal alignment between a handheld laser scanner and a smartphone camera without any hardware synchronization. The method combines deep-learning-based target detection with classical geometric calibration using 2D-3D correspondences derived from black and white hourglass planar targets. Target centers are precisely localized in both the RGB images and the 3D point cloud using a symmetric templatematching scheme, enabling robust solution of the perspective-n-point (PnP) problem for spatial calibration. To address the lack of hardware synchronization, we introduce a temporal calibration method that exploits geometric correspondences between rendered intensity images and camera frames. On a Lixel L2 Pro scanner with a Huawei P20 Pro camera, the pipeline achieves a median Reprojection error of 0.76 px for static calibration and 2.19 px across 91 dynamic evaluations. The approach enables accurate image-pointcloud fusion for scanners without syncronisation interfaces and provides a foundation for colorization, image analysis, and redensification of laser data.

1. Introduction

The automated detection and quantification of cracks in masonry have been demonstrated in (Krefft and Hoegner, 2025). This paper presents an approach that combines 3D point clouds with image data to analyze cracks. The 3D point clouds are generated using photogrammetric methods. Although photogrammetric methods are powerful, they are limited in complex indoor environments, particularly due to insufficient texture, difficult lighting conditions, or limited lines of sight.

Laser scanners offer an alternative, providing higher geometric accuracy in indoor areas. However, they primarily provide geometric information, which can be a disadvantage. Although modern systems have intensity values or RGB data, their quality is often insufficient to capture fine structures like masonry cracks with a width of 0.1 mm.

A handheld Lixel L2 Pro laser scanner with two integrated cameras was used in this study. However, the strong distortion of the images caused by the panoramic view of the lenses limits the detection of small objects like fine cracks.

To address this problem, this paper presents a method that enhances the laser scanner system with an additional camera to capture even the smallest surface features reliably. The focus is on determining the external calibration parameters between a smartphone camera and the laser scanner to enable precise fusion of image and point cloud data.

This article presents a calibration method in which virtual images are generated based on the intensity values of laser scanner points. These images are used to perform 2D-based matching with RGB images from a smartphone camera. The findings obtained from this process are ultimately used to determine the registration of the two systems in relation to each other via 2D-3D correspondences. Black-and-white hourglass targets with a diameter of 10 cm are used to identify these correspondences. A method is presented to precisely detect the centers of these targets in both images and point clouds. Finally, a method is presented to ensure temporal synchronization between the images and the laser scanner system.

2. Related Work

The precise fusion of laser scanner data and camera images is essential for applications such as autonomous navigation, 3D reconstruction, and environmental awareness in robotics and computer vision (Fuersattel et al., 2017). This requires the extrinsic calibration of the laser-camera system, where the spatial transformation between the two sensors is determined. This transformation involves both rotation and translation (An et al., 2024).

Various methods exist for determining these parameters, all of which rely on identifying correspondences between camera images and the point cloud of the laser scanner. According to (An et al., 2024), these approaches can be broadly categorized into explicit and implicit methods.

Explicit methods can be further divided into three subcategories:

- 2D-3D correspondences
- 3D-3D correspondences
- Odometry-based correspondences

In odometry-based methods, the movement trajectories of the laser scanner are compared with those of the camera system. The camera’s trajectory is often determined through hand-eye calibration. A significant advantage of this approach is that the fields of view of the two sensors do not need to overlap (An et al., 2024).

The 2D-3D and 3D-3D correspondence methods can be further distinguished based on the use of artificial targets or natural environmental structures (Wang et al., 2022). Planar boards, printed with checkerboard patterns or featuring holes, serve as targets, for example. Other examples include AprilTags or ArUco markers. In addition to planar targets, 3D calibration objects such as spheres or boxes with known dimensions can also be employed (An et al., 2024).

If no artificial objects are used, geometric features in the natural scene must be identified. Corresponding planes and lines are often used for this purpose (Wang et al., 2022).

Many methods are based on 3D-3D calibration. Planar checkerboard patterns can also be used for this purpose, as they can be reliably detected in both images and point clouds. With knowledge of the intrinsic camera parameters and the actual dimensions of the checkerboard, the 3D coordinates of the checkerboard corners can be calculated from the images. By mapping them to the corresponding points in the point cloud, the transformation parameters can be calculated.

Another approach to 3D-3D calibration is point-to-plane calibration. Here, the plane of the checkerboard, including the normal vector, are determined from the image. The corresponding points are extracted from the point cloud, with the condition that they must lie on the plane. Optimization is performed until the sum of the distances of the points to the plane reaches a minimum (Pandey et al., 2010). Extensions of this method also take into account line correspondences, such as the edges of the chessboard in the image and point cloud (Zhou et al., 2018; Mishra et al., 2020).

An interesting 3D-3D approach is presented in (Tóth et al., 2020), in which spheres are used as target markers. These can be reliably detected in both point clouds and images, for example, by ellipse fitting. With the known sphere radius, the 3D centers can be determined and the transformation is calculated from them.

There are also 3D-3D calibration methods that do not require target markers. In (Gong et al., 2013), a method is presented in which three intersecting planes serve as calibration targets, for example, the corner of a room (two walls and the floor). Another option is to register two point clouds with each other. The image-based point cloud can be generated using Structure from Motion (SfM), for example. The transformation is then determined using algorithms such as Iterative Closest Point (ICP) (An et al., 2024).

Implicit calibration methods include edge-based methods, deep learning approaches, and hybrid methods that combine both approaches (An et al., 2024). Edge-based methods mostly maximize the overlap of image and laser edges (An et al., 2024). An exciting approach is presented in (Pandey et al., 2014), which maximizes the mutual information (MI) between the intensity values of the laser scanner and the camera. The idea is that the reflectivity of a laser point is statistically correlated with the intensity of the corresponding image point, provided that the transformation is correct.

A deep learning-based approach is described in (Koide et al., 2023). Here, virtual images are generated from the intensity values of the laser points. Correspondences between these and the real camera images are determined using the SuperGlue Network. The transformation is then calculated using the RANSAC algorithm.

All of these studies describe how the geometric relationship between a laser scanner and a camera can be determined. However, they do not address the calibration of the synchronization between the laser scanner and the camera. The method presented in this study allows the determination of the geometric relationship between a laser scanner and a camera, as well as the synchronization of the two systems with each other. Temporal synchronization is achieved using an image-based approach. Geometric calibration is achieved using corresponding 2D-3D points. For this purpose, a method is presented that allows the precise detection of the centers of black and white hourglass targets in 3D space as well as in 2D images.

3. Methodology

A handheld Lixel L2 Pro laser scanner was used in this study. Image data were acquired with a Huawei P20 Pro smartphone, which was rigidly mounted on the scanner. This fixed configuration ensures that the relative position and orientation between the smartphone and the laser scanner remain constant.

Since the laser scanner hardware does not support direct synchronization with the smartphone camera, both sensors were operated independently. During data acquisition, the laser scanner recorded point clouds while the smartphone simultaneously captured video. Individual frames were extracted from these videos and subsequently registered within the scanned scene. The laser scanner’s trajectory serves as the spatial reference for this registration.

To achieve a successful registration, two unknowns must be determined:

1. the spatial transformation between the two sensors, and
2. the temporal alignment between the image frames and the laser scanner trajectory.

For the spatial transformation, corresponding points between the two datasets were identified using black and white hourglass targets. The centers of these targets can be precisely detected in both, the smartphone images and the laser scanner point clouds. The subsequent sections describe the procedure for identifying these correspondences and for estimating the transformation parameters.

Unlike previous calibration pipelines, described in the section before, our approach combines deep-learning-based feature detection for temporal alignment without any hardware synchronization. For the synchronization, a method is proposed in which both datasets are treated as time series. Linear interpolation parameters are estimated to align the images with the scanner trajectory. This approach relies solely on natural scene features, eliminating the need for artificial targets.

Figure 1 illustrates the mobile platform used in this work. The core component is the Lixel L2 Pro laser scanner, mounted on a steel plate together with the smartphone. Both sensors are fixed-centered, ensuring that the relative geometry remains stable even after dismounting and remounting the system.

Figure 1. Handheld Lixel L2 Pro laser scanner with fixed mounted camera.

3.1 Laser scanner camera calibration

To estimate the transformation parameters between the laser scanner and the camera, 2D–3D correspondences between the image data and the laser point cloud are established. These correspondences are defined by the centers of hourglass targets, which can be precisely identified in both datasets. Based on these correspondences, the parameters of the spatial transformation between the two sensor coordinate systems are derived. Furthermore, this section describes the approach used to synchronize the image and laser scanner data.

3.1.1 Targetfinder in images

The centers of the targets are determined in the RGB images in a two-step process. First, the targets are detected in the images using a YOLO11 object detector (Jocher and Qiu, 2024), and then the exact centers of the targets are determined using a correlation-based method described in (Abmayr et al., 2008).

The YOLO11 object detector was trained with a resolution of 1280 px x 1280 px over 200 epochs. The underlying dataset has 262 training instances and 112 validation instances. On the validation dataset, with a confidence level of 0.5, a precision of 1.0 was achieved, with a recall of 0.99 and an F1 score of 0.99.

The core idea behind the second step, the correlation-based approach, is to use symmetrical targets that can be reliably detected even when the input image is mirrored. In this procedure, the position of the extracted image patch, detected by the YOLO11 object detector, is first located within the original image by template matching. The same operation is then repeated with a mirrored version of the patch. The final target center is computed from the mean position of both detections, as defined in Equation 1.

$$(x_n, y_n) = \frac{1}{2}((x_1, y_1) + (x_2, y_2)) \tag{1}$$

where (x_n, y_n) = Pixel coordinates of the target center.
(x_1, y_1) = Pixel coordinates of the center of the first template matching.
(x_2, y_2) = Pixel coordinates of the center of the second template matching.

The described procedure is illustrated in Figure 3. In this example, the center of the target is determined based on an intensity image instead of an RGB image. However the underlying methodology remains identical. The green rectangle indicates the YOLO11 object detection, the red rectangle shows its mirrored counterpart, and the blue point represents the detected center.

3.1.2 Generate intensity image

This section is included because it is essential for the subsequent two sections. The objective is to generate synthetic images from the intensity values of the laser scanner point cloud. These images are used in Chapter 3.1.3 to locate the centers of target markers in the 3D point cloud and in Chapter 3.2 to temporally synchronize the laser scanner with the smartphone camera.

A virtual camera view is rendered to match the viewing direction of the actual smartphone camera. The 3D points of the point cloud are projected onto a virtual image plane using the smartphone’s intrinsic camera parameters, which account for lens distortion. The projection procedure follows (Krefft and Hoegner, 2025), and a visibility check according to (Katz et al., 2007) ensures that only points visible from the virtual camera are projected.

Intensity values from the 3D points are assigned as pixel values on the image plane. Any resulting gaps in the synthetic image are filled using linear interpolation. An example of such an intensity image is shown in Figure 2.

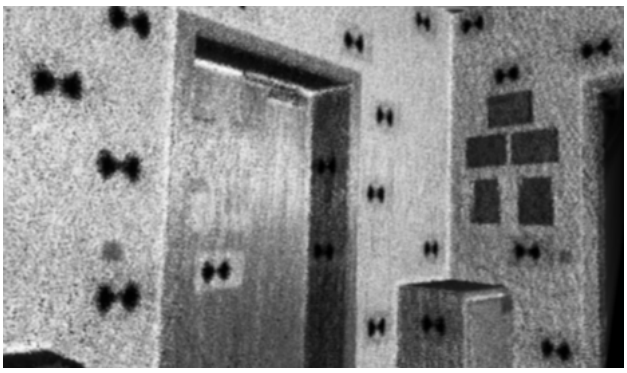

Figure 2. Synthetic projected intensity image.

3.1.3 Target finder in point cloud

This section describes the automatic detection of black and white hourglass target centers in a laser scanner point cloud. The method builds on the image-based procedure presented in

Chapter 3.1.1. An artificially generated intensity image (see Section 3.1.2) serves as input.

For the coarse localization of targets, a YOLO11 (Jocher and Qiu, 2024) object detector was trained to detect targets in these intensity images. The model was trained over 200 epochs at a resolution of 1280 px x 1280 px with 551 labeled instances. On a validation dataset with 100 instances, a precision of 0.98, a recall of 0.97, and an F1 score of 0.93 were achieved at a confidence level of 0.5.

Each target detected by YOLO11 is processed individually. The 2D image coordinates of the target center are determined, and the nearest back-projected 3D point is identified to provide an approximate 3D location of the target center. All 3D points within a defined radius around this location are extracted to isolate the point cloud of the target. Since the markers are planar, RANSAC is applied to robustly fit a plane, and only points within a specified distance from this plane are retained.

For further analysis, the 3D points are projected into 2D. The point cloud is first centered, and singular value decomposition (SVD) is performed to determine the principal axes. The first two singular vectors define the plane of the target, while the third vector corresponds to the normal vector. The centered point cloud is then projected onto the plane, resulting in 2D coordinates.

Intensity values from the laser point cloud are assigned to the projected points, and linear interpolation is applied to fill gaps, producing a frontal top-view image of the target. The template matching procedure described in Section 3.1.1 is then applied. An initial patch for the first template matching is derived from the YOLO11 detection. The mirrored patch is used for the second matching. The final target center is computed as the mean of the two template centers (Equation 1), as visualized in Figure 3 (blue).

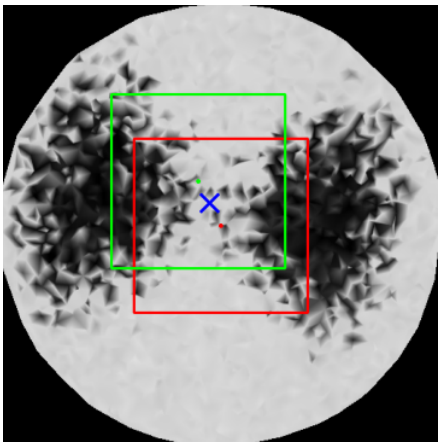

Figure 3. Detecting the center of the target with double template matching.

Finally, the 2D target center is converted to a 3D coordinate. To improve accuracy, the 3D center is obtained by linear interpolation of the 3D coordinates corresponding to the projected image points.

The results of this procedure are shown in Figure 4, where the detected target centers are marked in red.

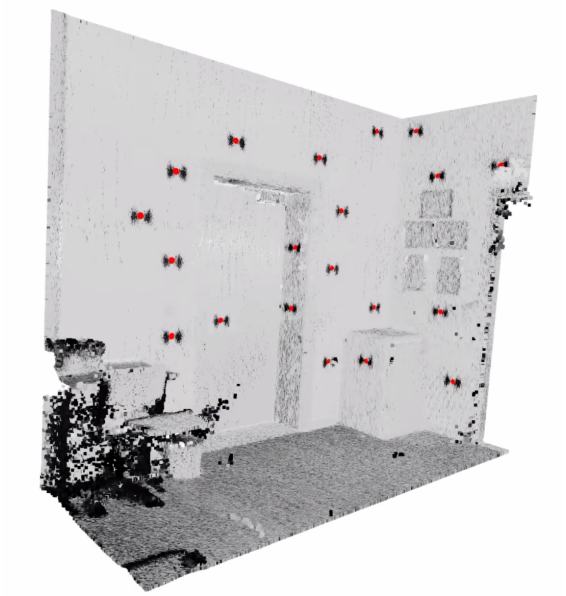

Figure 4. Detected centers of targets in the point cloud.

3.1.4 Co-registration via corresponding points

To determine the spatial transformation $T_{\text{scanner}\rightarrow\text{camera}}$ between the laser scanner and the smartphone camera, the handheld system is positioned statically, ensuring that the targets are visible in both the camera images and the laser scanner point cloud. The targets are distributed spatially in both depth and width to enable robust estimation.

The camera pose T_{camera} is estimated from 2D–3D correspondences using the PnP (Perspective-n-Point) method in combination with RANSAC (Marchand et al., 2016). RANSAC increases robustness by iteratively selecting random subsets of correspondences, estimating a pose, and evaluating the number of remaining correspondences consistent with this pose within a defined tolerance. The subset with the highest consensus is selected as an initial Camera pose (Marchand et al., 2016).

The final camera pose is refined using the SOLVEPNP_ITERATIVE method, which performs a Levenberg-Marquardt optimization to minimize the joint Reprojection error (Bradski, 2000a).

The transformation parameters are derived by combining the camera pose T_{Camera} with the laser scanner pose $T_{Scanner}$ obtained from its trajectory. Both poses describe position and orientation in a global coordinate frame. The transformation from the scanner to the camera is then computed as shown in the following Equation 2.

$$T_{scanner\rightarrow camera} = T_{Scanner}^{-1} \cdot T_{Camera}^{-1} \tag{2}$$

T are the homogeneous transformation matrices consisting of the rotation matrices R and the translation vectors p.

$$T_{Scanner} = \begin{bmatrix} R_{Scanner} & p_{Scanner} \\ 0 & 1 \end{bmatrix}$$
$$T_{Camera} = \begin{bmatrix} R_{Camera} & p_{Camera} \\ 0 & 1 \end{bmatrix}$$

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-1-2026-167-2026 | © Author(s) 2026. CC BY 4.0 License.

3.2 Temporal synchronization without hardware trigger

In contrast to previous laser-camera calibration pipelines that rely on hardware triggering, the method presented in this paper estimates a temporal correction factor purely from geometric scene consistency, leveraging SuperGlue feature matching (Sarlin et al., 2020) between virtual intensity views and real images. The precise projection of 3D points onto the image plane requires temporal synchronization between the laser scanner and the smartphone camera. Hardware based synchronization is not available for the scanner used. Therefore an alternative approach is applied.

The method exploits the scanner trajectory and the known camera intrinsics along with the spatial transformation $T_{scanner \rightarrow camera}$ determined in Section 3.1.4. During scanning, a video was recorded simultaneously with the smartphone rigidly mounted on the scanner. Both the relative position and orientation remain constant. It is assumed that the scanner poses and video frames are available at constant time intervals. The scanner poses at 0.1 sec intervals and video frames at 30 Hz. Under this assumption, a linear mapping between scanner poses and video frames can be established.

$$PoseIndex_n = a \cdot FrameIndex_m + b \tag{3}$$

where $PoseIndex_n$ = Pose n of the laser scanner
a = Slope
b = Axis section
$FrameIndex_m$ = Image number m, taken from the video

The unknown parameters a and b are determined as follows.

For each scanner pose, a virtual image is rendered from the 3D point cloud by projecting points onto an image plane using the smartphone camera intrinsics and the external pose derived from the scanner trajectory. Only points within the camera field of view are considered.

To find the best correspondence between virtual and real images, 2D image points are matched. The SuperGlue Network (Sarlin et al., 2020) is employed to detect corresponding points between the rendered intensity image and the original video frame, as illustrated in Figure 5. An affine transformation is computed from these correspondences (Bradski, 2000b). Minimal transformation parameters indicate a good match between a scanner pose and a video frame.

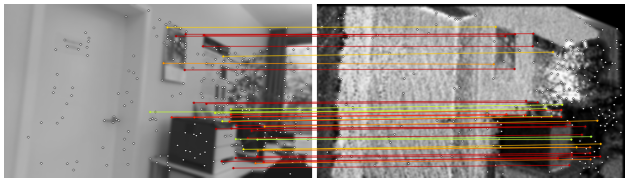

Figure 5. Using the Super Glue Network to search for corresponding points.

This procedure is repeated for all video frames. The resulting correspondences exhibit a linear relationship, as shown in Figure 6. Based on this correspondence, a straight line is derived using the RANSAC algorithm. In Figure 6 the found straight line is shown in green, inliers in blue and outliers in red. The individual frames of the video are shown on the x-axis. The scanner trajectory is listed on the y-axis.

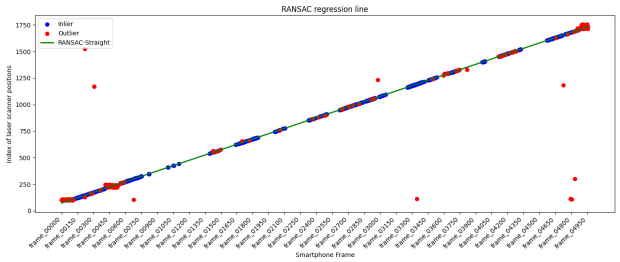

Figure 6. RANSAC regression line for temporal synchronization of the laser scanner with the smartphone images.

Using the determined linear equation, it could now be determined, for example, that image number 100 correlates with laser scanner pose 200.4. To assign the image to a laser scanner pose accurately, the resolution of the laser scanner poses is increased by a factor of 100 using Akkima interpolation. This enables more precise alignment between the images and the laser scanner poses.

4. Reprojecting the 3D point cloud onto the image plane

With all calibration and synchronization information available, 3D points can be projected onto the image plane according to the camera model by (Zhang, 2000):

$$p = A \cdot (T_{Scanner_n} \cdot T_{scanner \rightarrow camera}) \cdot P_w \tag{4}$$

where p = Pixels on the image plane
A = Internal camera parameters
$T_{Scanner_n}$ = Scanneropose at time n
$T_{scanner \rightarrow camera}$ = Transformation parameters
P_w = 3D coordinates in the world coordinate system

The transformation $T_{scanner \rightarrow camera}$ is known from the calibration procedure described in Section 3.1.4. The scanner pose $T_{Scanner_n}$ is derived using the linear mapping from Section 3.2.

$$T_{Scanner_n} = Scannerposes_{PoseIndex_n} \tag{5}$$

In order to use only points that are actually visible from the camera when projecting the point cloud, a visibility check, like in chapter 3.1.2, is performed according to (Katz et al., 2007). This allows 3D points to be determined that are occluded by other objects in the scene. In Figure 7 the visibility check is illustrated. The Figure shows the point cloud of a room, where all 3D points within the camera’s field of view are colored according to the corresponding camera image. Points that are visible but occluded by other objects are colored in red. The location of the scanner is also displayed in the image with a coordinate system.

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-1-2026-167-2026 | © Author(s) 2026. CC BY 4.0 License.

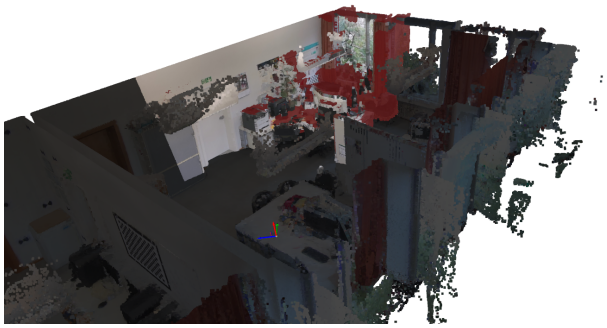

Figure 7. Point cloud in which the 3D points are colored according to the color values of the corresponding camera image. Hidden points are colored red.

5. Discussion

The method presented in this work demonstrates how the parameters of a spatial transformation between a handheld laser scanner and a smartphone camera can be precisely determined. Natural features in the scene are used to assign individual frames of a smartphone video to the corresponding poses of the laser scanner. Temporal calibration parameters are subsequently derived from these features.

The SuperGlue Network for correspondence search has proven to be particularly powerful. It can recognize identical points even when the images to be matched are very different from each other. Studies using a SIFT-based feature matching were unable to show these characteristics.

To validate the method, a recording of approximately three minutes was performed, in which the hourglass targets were visible in the smartphone images at multiple time points. Reprojection errors were computed based on these targets and the determined transformation parameters. The results are summarized in Table 1, with columns corresponding to the respective time points during the scan.

The results in the table below show that the presented method can achieve spatial and temporal alignment without any hardware synchronization. Across all 91 Reprojection errors, the median error is 2.19 pixels with a standard deviation of 1.41 pixels. The transformation parameters between the smartphone camera and the laser scanner were determined based on the correspondences in frame 40, prior to the start of any movement (movement began around frame 100). For the static calibration, all visible marks were used. This includes all marks except for mark 207 (This is the only target that is not in the camera’s field of view.). For this calibration, the average Reprojection error is 0.76 pixels.

The behavior of the Reprojection error was further analyzed by varying the number of targets used to estimate the geometric transformation parameters. When all 19 visible targets are used for calibration, the median Reprojection error amounts to 2.19 pixels, as reported in Table 1. If the number of targets employed for geometric calibration is randomly reduced to 10, the median Reprojection error across the six test images slightly increases to 2.28 pixels. A further reduction to five targets results in a median Reprojection error of 2.26 pixels.

Target ID	Frame 40 Error [px]	Frame 209 Error [px]	Frame 330 Error [px]	Frame 3307 Error [px]	Frame 4183 Error [px]	Frame 4716 Error [px]
100	1.20	2.90	1.77	2.70	4.86	1.46
101	1.42	1.16		0.96		2.75
102	1.70	0.71	0.83	1.07	2.64	3.23
103	2.08			5.86		0.83
104	1.14					3.36
105	2.49	2.12	0.72	1.99	2.95	2.07
106	3.17	2.89	0.80	4.16	2.29	1.80
107	0.67	2.44		2.19		5.33
108	1.44	5.98		5.08		2.05
109	2.50	2.35	0.82	2.47	4.65	4.90
200	2.42	0.92	0.50	0.76	4.21	3.02
201	3.18			3.64		3.26
202	2.69	1.40	0.85	1.45	4.66	1.93
203	3.29					4.59
204	0.69	3.59	3.03	5.33		3.00
205	1.78	3.39	1.27	2.88		1.41
206	2.01	2.15	0.55	4.41	1.62	2.83
207		2.93	0.58	4.16		1.87
208	0.44			1.19		4.28
209	1.85		1.33	0.54		5.14

Table 1. Reprojection error (in pixels) of the reprojected 3D coordinates of the centers of the target onto the image plane. The frame number refers to the corresponding frame in the smartphone video recording.

The targets required for the evaluation were reliably detected by the YOLO11 networks. There were no false detections, and no targets were missed. This indicates that the detection approach performs reliably under the evaluated conditions. However, it must be taken into account that the training data was recorded in the same scene as the test run. It is to be expected that the targets cannot be detected with the same quality in a different environment. To achieve this, the networks would probably need to be retrained with additional data.

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-1-2026-167-2026 | © Author(s) 2026. CC BY 4.0 License.

In addition to the Reprojection errors, the trajectory of the smartphone camera was evaluated. For this purpose, the trajectory determined using the method presented here was compared with the trajectory of a photogrammetric evaluation. A total of 456 images were used for this experiment. The trajectories are illustrated in graphic 8 below. The trajectory estimated using the method presented here is shown in blue, while the photogrammetric trajectory is shown in red.

The positions of the two trajectories were used for the evaluation. To eliminate systematic errors, the residual vectors were corrected by subtracting a constant coordinate offset. This offset was determined as the mean coordinate difference between the two trajectories. Based on the corrected residuals, the Euclidean 3D distances were calculated and used for the evaluation. The resulting root mean square error (RMSE) of the residual distances is 0.078 m. The RMSE values of the individual x, y, and z coordinates are 0.042 m, 0.062 m, and 0.021 m, respectively. The standard deviation of the 3D residual distances is 0.039 m, while the standard deviations of the x, y, and z residual components are 0.041 m, 0.062 m, and 0.021 m. The median of the corrected 3D residual distances is 0.06 m.

Figure 8. Visualization of movement trajectories. The trajectory of the smartphone was determined using the method presented here and is shown in blue, while the trajectory of the smartphone determined using photogrammetry is shown in red.

The results are also very impressive when viewed visually. Figure 9 shows, for example, that the projected 3D points correspond well with the edges in the image. This can be clearly seen at the edges of the door or the cupboard in the image.

Figure 9. The point cloud shown in Figure 7 was projected onto the image plane.

The calibration field, shown in Figure 10, visualizes the results of the static calibration. The centers of the targets in the images are marked in green, while the corresponding back-projected 3D points are shown in red. Due to the small Reprojection errors, these points nearly overlap, making them difficult to distinguish. The error vectors in these image are amplified by a factor of 50 and shown in blue, providing a more informative visualization. Their directions vary significantly, indicating the absence of systematic errors in the Reprojection process.

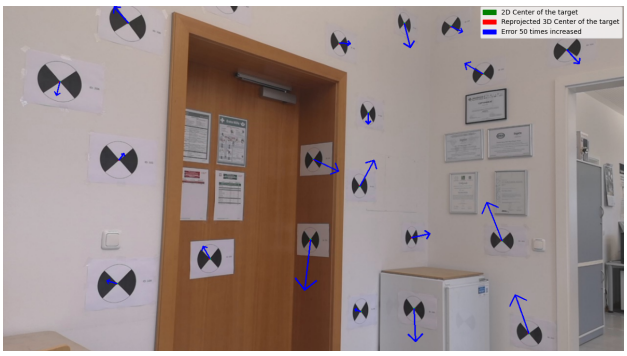

Figure 10. The calibration field for determining the transformation parameters between the laser scanner and smartphone camera.

6. Conclusions and FutureWork

The hardware used in this study is limited to the Lixel L2 Pro handheld scanner and a standard smartphone. For a more precise evaluation of temporal synchronization, future work should employ a laser scanner that allows hardware-based camera triggering. This would enable a more accurate assessment of how closely image frames can be synchronized with scanner poses using the proposed method. With the current median Reprojection error of 2.19 pixels, the error likely results from a combination of scanner pose uncertainty and temporal interpolation effects. A scanner with hardware synchronization would allow for a more detailed separation of these error sources.

Future research will also investigate alternative synchronization technologies, such as GNSS-based timestamps, to improve temporal alignment between the sensors. It might also be worth investigating replacing the virtual intensity images from the laser

scanner with Gaussian splatting images. Calculating the intensity images is a very time-consuming process, which could be significantly reduced by using Gaussian splatting methods.

Another important aspect is the redensification of the point cloud. The main motivation for fusing image and laser data is to combine their respective strengths. The laser scanner provides geometrically accurate and globally consistent 3D points, while the images offer high-resolution texture and detailed visual information. The current point cloud has an average point spacing of approximately 5 mm. By using back-projected image points, depth maps could be computed and used to redensify the point cloud, either via classical interpolation or with neural network-based methods. However, the findings of the method presented here are the basis of redensification, since the external camera parameters must be known for this.

Finally, further work could investigate the extent in which the targets can be avoided in static calibration. A method such as presented in (Sarlin et al., 2020) could be used to determine the geometric relationship between the camera and the laser scanner. In this method, the transformation parameters are determined using the SuperGlue network.

In summary, it can be concluded that the method presented provides a good way of equipping a laser scanner with an additional camera without the need for hardware synchronization. This can be particularly useful when existing systems are to be expanded with a camera. As mentioned at the beginning, this can be particularly important when images are used, for example, to detect, evaluate, and locate structural damage. However, this requires knowledge of the external camera parameters of an image. The method presented here allows the necessary information to be obtained. However, the limiting factor is that temporal synchronization can only be performed in post-processing. This means that this method is not suitable for real-time applications.

Acknowledgments

This work was supported by the NEMETSCHEK Innovationsstiftung.

References

Abmayr, T., Härtl, F., Burschka, D., Fröhlich, C., Hirzinger, G., 2008. A correlation based target finder for terrestrial laser scanning. *Journal of Applied Geodesy*, 2, 131-138.

An, P., Ding, J., Quan, S., Yang, J., Yang, Y., Liu, Q., Ma, J., 2024. Survey of Extrinsic Calibration on LiDAR-Camera System for Intelligent Vehicle: Challenges, Approaches, and Trends. *IEEE Transactions on Intelligent Transportation Systems*, 25(11), 15342-15366.

Bradski, G., 2000a. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Bradski, G., 2000b. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Fuersattel, P., Plank, C., Maier, A., Riess, C., 2017. Accurate laser scanner to camera calibration with application to range sensor evaluation. *IPSJ Transactions on Computer Vision and Applications*, 9.

Gong, X., Lin, Y., Liu, J., 2013. 3D LIDAR-Camera Extrinsic Calibration Using an Arbitrary Trihedron. *Sensors*, 13, 1902-1918.

Jocher, G., Qiu, J., 2024. Ultralytics yolo11.

Katz, S., Tal, A., Basri, R., 2007. Direct visibility of point sets. 26.

Koide, K., Oishi, S., Yokozuka, M., Banno, A., 2023. General, single-shot, target-less, and automatic lidar-camera extrinsic calibration toolbox. 11301–11307.

Krefft, L., Hoegner, L., 2025. A Method for Crack Detection and Quantification in Masonry Using Neural Network-Based Image Analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W6-2025, 113–119. https://isprs-annals.copernicus.org/articles/X-4-W6-2025/113/2025/.

Marchand, E., Uchiyama, H., Spindler, F., 2016. Pose Estimation for Augmented Reality: A Hands-On Survey. *IEEE Transactions on Visualization and Computer Graphics*, 22(12), 2633-2651.

Mishra, S., Pandey, G., Saripalli, S., 2020. Extrinsic calibration of a 3d-lidar and a camera.

Pandey, G., McBride, J., Savarese, S., Eustice, R., 2010. Extrinsic Calibration of a 3D Laser Scanner and an Omnidirectional Camera. *IFAC Proceedings Volumes*, 43(16), 336-341. https://www.sciencedirect.com/science/article/pii/S1474667016350790. 7th IFAC Symposium on Intelligent Autonomous Vehicles.

Pandey, G., McBride, J., Savarese, S., Eustice, R., 2014. Automatic Extrinsic Calibration of Vision and Lidar by Maximizing Mutual Information. *Journal of Field Robotics*, 32.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. SuperGlue: Learning feature matching with graph neural networks. *CVPR*.

Tóth, T., Pusztai, Z., Hajder, L., 2020. Automatic lidar-camera calibration of extrinsic parameters using a spherical target. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 8580–8586.

Wang, Y., Li, J., Sun, Y., Shi, M., 2022. *A Survey of Extrinsic Calibration of LiDAR and Camera*. 933–944.

Zhang, Z., 2000. A Flexible New Technique for Camera Calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22, 1330 - 1334.

Zhou, L., Li, Z., Kaess, M., 2018. Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5562–5569.