The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B1-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission I, 4–11 July 2026, Toronto, Canada

Calibration and Georeferencing for GNSS-Equipped Vehicle Video Mapping Using a Tesla Model Y

Rami Tamimi[1], Charles Toth[1]

[1] SPIN Laboratory, Dept. of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA
(tamimi.12, toth.2)@osu.edu

Keywords: Camera Calibration, Georeferencing, Consumer Vehicle, Video Mapping, PnP Localization, Photogrammetry

Abstract

The evolution of mapping platforms has followed a consistent pattern: professional instruments are complemented by consumer devices that trade precision for scalability. Unmanned aerial systems transformed aerial photogrammetry by making it accessible beyond traditional aircraft, and smartphones, especially when equipped with high accuracy GNSS positioning, have demonstrated viable terrestrial mapping. This paper extends that progression to vehicle-based mapping by presenting SurveyXR, a web-based calibration and georeferencing framework that converts dashcam video from a GNSS-equipped vehicle into georeferenced imagery suitable for Structure-from-Motion (SfM) processing. By providing accurate per-frame exterior orientation parameters, the system enables direct georeferencing of the SfM output, eliminating the need for ground control points in the photogrammetric workflow. The pipeline implements checkerboard-based intrinsic calibration with automated quality diagnostics, Perspective-n-Point exterior orientation solving with automatic boresight detection, GNSS-synchronized frame extraction, and lever arm correction between the GNSS antenna and each camera. All computation runs in a browser or lightweight cloud backend, requiring no local software installation. The framework was evaluated on a 2026 Tesla Model Y equipped with a roof-mounted Emlid Reach RS4 Pro PPK GNSS receiver on the Ohio State University campus. Georeferencing accuracy was assessed against 71 independently surveyed RTK check points in two configurations: direct georeferencing only (no ground control) and GCP-constrained bundle adjustment. The paper documents the calibration methodology, time synchronization model, error budget analysis, and quantitative accuracy assessment.

1. INTRODUCTION

The history of geospatial data acquisition shows a recurring pattern of democratization. Aerial photogrammetry, once limited to manned aircraft carrying metric cameras, was transformed by the introduction of unmanned aerial systems carrying consumer-grade sensors (Colomina and Molina, 2014; Nex and Remondino, 2014). What required a survey aircraft and trained crew could suddenly be accomplished with a sub-kilogram platform and an off-the-shelf camera. Similarly, terrestrial mobile mapping, traditionally the domain of dedicated survey vehicles with tightly coupled GNSS/IMU and calibrated sensors (Toth and Grejner-Brzezinska, 2004), has been extended to smartphones equipped with RTK positioning (Tamimi and Toth, 2023). In each case, the enabling step was not a breakthrough sensor but a calibration and georeferencing workflow that bridged the gap between consumer hardware and mapping-grade output.

Consumer vehicles represent the next logical platform in this progression. Modern production vehicles are equipped with multiple cameras. Tesla models carry six surround-view cameras that operate continuously while driving. Unlike UAS or smartphones, these platforms collect data incidentally: every trip generates hours of roadway video without any deliberate survey effort. The raw video is not unlike early dashcam footage, useful for documenting events but without spatial reference. The challenge is transforming this footage into data: calibrating cameras that were never characterized for photogrammetric use, synchronizing video timelines with a navigation trajectory, and assigning each frame a position and orientation in a geodetic reference frame.

Recent work on Tesla cameras has demonstrated their photogrammetric potential. Pepe et al. (2023) calibrated Tesla front and rear cameras and applied reverse-projection photogrammetry to determine vehicle speeds from dashcam video, achieving 2.45% average error. Wu et al. (2020) presented image-based camera localization methods including PnP-based pose estimation from 2D–3D correspondences, establishing the theoretical foundation for the exterior orientation approach used here. These studies confirm that consumer automotive cameras, while not designed for survey work, can yield quantitative spatial measurements when properly calibrated.

This paper presents a framework that addresses the full calibration-to-georeferencing workflow. The system accepts vehicle camera video and a GNSS trajectory, and outputs georeferenced images with embedded camera model and orientation metadata suitable for direct import into SfM software. The overall system architecture is shown in Figure 1. The implementation follows the same principle that made UAS photogrammetry where accessibility is as important as accuracy for enabling widespread adoption.

2. CAMERA CALIBRATION

2.1 Intrinsic Calibration

By implementing Zhang’s checkerboard calibration method (Zhang, 2000), the system determines the camera intrinsic parameters: focal length (f), principal point (cx, cy), and lens distortion coefficients. The choice of distortion model depends on the camera’s field of view. For narrow-to-moderate FOV cameras (the Tesla front camera, ~45°), the Brown-Conrady model with radial coefficients k1, k2 and tangential coefficients p1, p2 is used, with k3 fixed to zero (CALIB_FIX_K3) to prevent overfitting. For wide-angle and fish-eye cameras (the Tesla rear camera, ~130°), the Kannala-Brandt equidistant model (cv2.fisheye.calibrate) is used instead, as the standard Brown model cannot represent the severe distortion of near-fish-eye lenses. Not all parameters of the Brown model may be statistically significant for a given camera; the system applies the minimal parameter set needed to achieve sub-pixel residuals. Calibration frame selection follows established best practices: frames must include 90-degree rotated board orientations, oblique angles (board not parallel to the sensor plane), and coverage of all nine zones of the image field (edges and corners, not only the center). Consumer cameras may exhibit instability due to temperature changes and vehicle vibration, which limits the achievable calibration precision regardless of the number of frames used. Each image is searched for checkerboard corners using adaptive thresholding, detected corners are refined to sub-

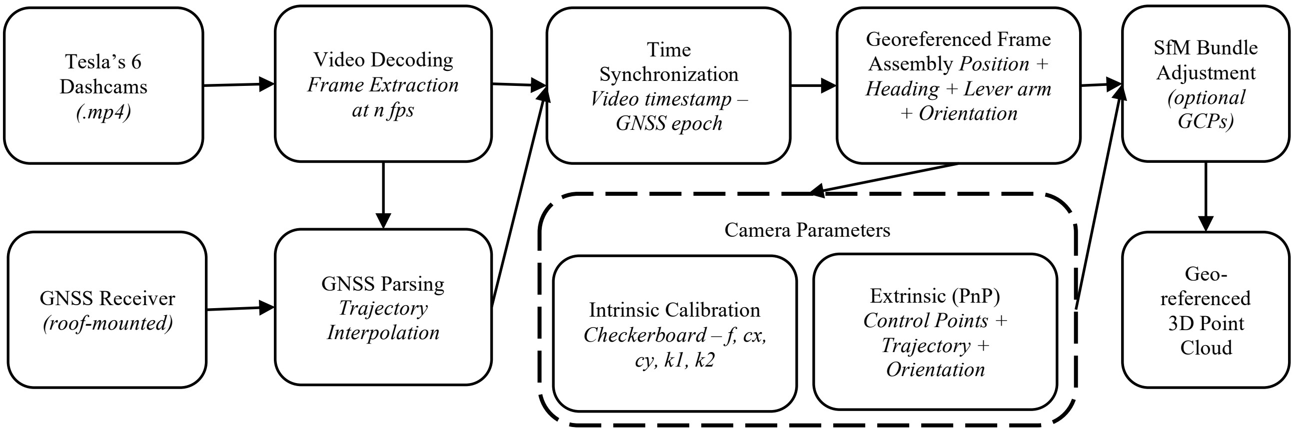

Figure 1. System overview. Video streams feed directly to frame decoding; GNSS provides a trajectory used for time synchronization and exterior orientation. Intrinsic calibration (dashed box) is performed offline and reused across sessions. The extrinsic computation module receives both decoded frames and the interpolated trajectory to solve per-frame position and orientation via PnP.

pixel accuracy using cornerSubPix with an 11×11 search window, and a diversity-selection algorithm chooses the most spatially varied subset of frames for the final calibration solve.

The camera matrix K encodes the focal lengths fx, fy in pixels and the principal point (cx, cy):

$$K = [[fx, 0, cx], [0, fy, cy], [0, 0, 1]] \tag{1}$$

The distortion model follows the Brown–Conrady formulation with radial coefficients ($k1$, $k2$) and tangential coefficients ($p1$, $p2$):

$$x' = x(1 + k1 \cdot r^2 + k2 \cdot r^4) + 2 \cdot p1 \cdot xy + p2 \cdot (r^2 + 2x^2) \tag{2}$$

$$y' = y(1 + k1 \cdot r^2 + k2 \cdot r^4) + p1 \cdot (r^2 + 2y^2) + 2 \cdot p2 \cdot xy \tag{3}$$

where $r^2 = x^2 + y^2$ and (x, y) are normalized image coordinates.

2.2 Calibration Quality Assessment

We implement automated quality assessments to prevent users from accepting poor calibrations. Three metrics are evaluated. First, the mean reprojection error (RMSE in pixels) across all detected corners in all frames. Second, a coverage score: the image is divided into a 3×3 grid (9 zones) and the score is the percentage of zones containing at least one checkerboard detection. Third, a rule-based health assessment flags three failure modes: extreme distortion ($|k1| > 1.0$, indicating mathematical failure), overfitting (low RMSE with low coverage), and insufficient coverage (< 50% zones filled). The system produces a rectification preview and a coverage heatmap for visual inspection, as illustrated in Figure 2.

a)

b)

c)

Figure 2. Calibration quality assessment: (a) checkerboard detection with color-coded corner rows, (b) 3×3 zone coverage matrix showing full-field coverage including rotated and oblique frames and the reprojection error, (c) undistorted rectification preview confirming lens model accuracy.

2.3 Extrinsic Orientation (PnP Solving)

For cameras where control points with known 3D coordinates are visible, we compute the exterior orientation using the Perspective-n-Point (PnP) algorithm. Given $n \geq 4$ correspondences between 3D world points and their 2D image projections, the system solves the rotation R and translation t:

$$s \cdot [u, v, 1]^T = K \cdot [R \mid t] \cdot [X, Y, Z, 1]^T \qquad (4)$$

where (u, v) are image coordinates, (X, Y, Z) are world coordinates, and s is a scale factor. The implementation uses solvePnPGeneric with the SQPNP algorithm and includes chirality checking (verifying all object points project in front of the camera) and outlier filtering (rejecting solutions with RMSE > 100 pixels). A coordinate centering strategy shifts world coordinates to their centroid before solving to improve numerical stability, which is typically necessary for projected coordinate systems where easting/northing values can exceed 10^6 units. Figure 3 illustrates the PnP georeferencing process. The Euler angles $(\omega, \varphi, \kappa)$ are extracted using the ZYX convention.

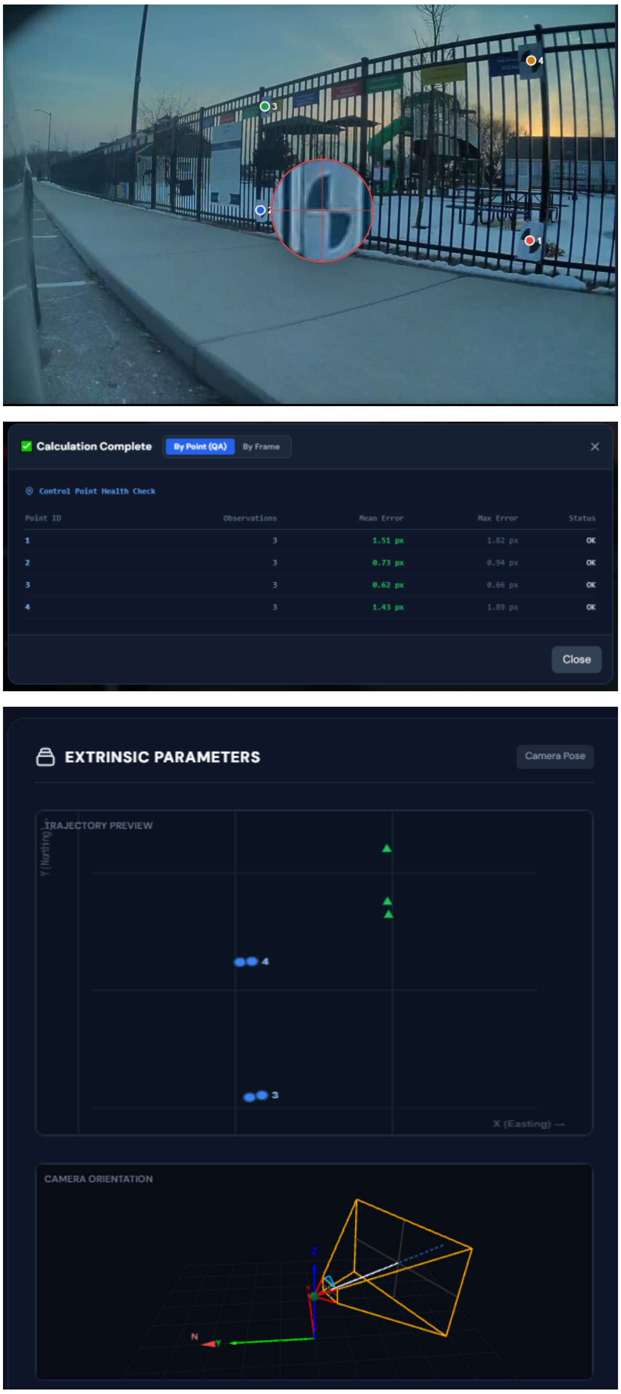

Figure 3. Perspective-n-Point georeferencing for exterior orientation solving.

3. FRAME EXTRACTION AND GNSS SYNCHRONIZATION

3.1 Time Synchronization

The core challenge in georeferencing vehicle camera video is establishing the temporal relationship between video frames and the GNSS trajectory. The Tesla camera system does not expose hardware-level timing signals, so the video file's filesystem timestamp (the lastModified property) is used to anchor the video timeline to UTC. This timestamp corresponds to the end of recording; the start time is computed by subtracting the video duration. Each frame's UTC time is then computed as:

$$t\_frame(i) = t\_file\_end - (duration - t\_video(i)) \qquad (5)$$

where t_file_end is the file's last-modified timestamp (corresponding to the end of recording), duration is the video duration in seconds, and $t\_video(i)$ is the time of frame i within the video. The synchronization accuracy is limited by two factors: the video frame rate and the GNSS sample rate. At 5 fps, each frame has a temporal resolution of 200 ms. With a 5 Hz GNSS receiver, positions are available every 200 ms. Linear interpolation between GNSS epochs reduces the effective timing uncertainty to approximately 20–50 ms, which at 25 mph (11.2 m/s) translates to 2–6 cm of positional uncertainty per frame. The camera clock is assumed to have negligible drift over the recording duration (typically 1–10 minutes). The GNSS trajectory file is parsed to extract timestamped positions in a compatible time base.

3.2 Frame Extraction and Georeferencing

Frame extraction operates in the browser using the HTML5 Video API. For each camera profile and its associated video files, the video is loaded and sought to target times at the user-specified frame rate (typically 1–5 fps for mapping). Each frame is drawn to an HTML5 Canvas and exported as JPEG. The frame's UTC timestamp is computed using Equation 5, and the GNSS position at that timestamp is obtained by linear interpolation on the trajectory. Vehicle heading (κ) is estimated from the GNSS trajectory tangent. Raw GNSS positions may fluctuate, particularly at low speeds, so a moving-average smoothing filter (3-point window) is applied before computing the bearing. The heading at time t is computed as the circular mean of bearings at multiple lookahead intervals (200, 400, 600 ms) to improve robustness:

$$\kappa = atan2(sin(\Delta\lambda)\cdot cos(\varphi_2),\ cos(\varphi_1)\cdot sin(\varphi_2) - sin(\varphi_1)\cdot cos(\varphi_2)\cdot cos(\Delta\lambda)) \qquad (6)$$

The lever arm, the 3D vector from the GNSS antenna phase center to each camera's sensor center, is a critical correction for multi-sensor vehicle platforms. The lever arm components $(\Delta X, \Delta Y, \Delta Z)$ are determined by surveying both the GNSS antenna and camera positions with RTK and computing the difference. Because the lever arm is defined in the vehicle body frame, it must be rotated by the vehicle heading at each epoch before being applied to the GNSS position. For a vehicle heading θ, the rotated offset in the ground frame is:

$$\begin{aligned} \Delta E &= \Delta X \cos(\theta) - \Delta Y \sin(\theta) \\ \Delta N &= \Delta X sin(\theta) + \Delta Y cos(\theta) \end{aligned} \qquad (7)$$

Each camera has its own lever arm and boresight direction (0° for front, 180° for rear, etc.), which is automatically determined from the PnP localization trajectory. Each extracted frame is written as a JPEG with embedded EXIF metadata containing GPS position (WGS84), UTC timestamp, camera orientation (ω, φ, κ), focal length, and accuracy indicators (horizontal and vertical RMS from the GNSS solution). This metadata enables direct import into photogrammetric software with position and orientation priors.

4. PLATFORM AND TEST DATA

4.1 Vehicle Configuration

The system was developed and tested using a 2026 Tesla Model Y (HW4) equipped with six onboard cameras (approximately 36 fps each) and a roof-mounted Emlid Reach RS4 Pro GNSS receiver operating at 5 Hz with PPK processing via Emlid Studio. The Tesla camera system provides surround-view coverage: front-facing, left/right pillar, left/right repeater (B-pillar), and rear-facing cameras. All sensor positions were measured relative to the GNSS antenna in the vehicle body frame to establish the lever arm offsets required for georeferencing.

4.2 Test Area and Results

Data collected across approximately 1.2 km of campus roadway at The Ohio State University, traversing open-sky, tree canopy, and urban canyon environments. The PPK solution achieved a fixed-ambiguity (Q=1) rate of 90.5% with precision of $\sigma\_N = 3.4$ mm, $\sigma\_E = 2.9$ mm, and $\sigma\_U = 9.0$ mm. The resulting georeferenced frame positions are shown in Figure 4.

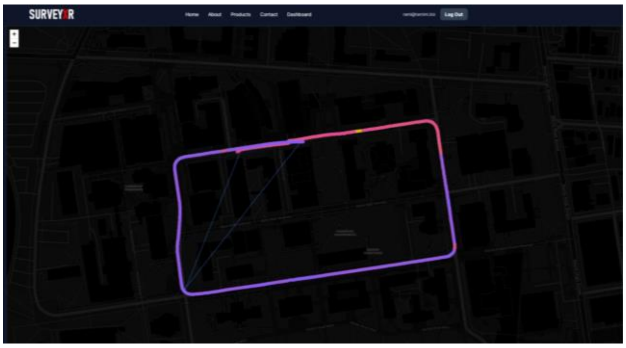

Figure 4. Georeferenced frame map view. Each marker represents a decoded video frame from the Tesla front camera, positioned along the driving trajectory on the OSU campus.

5. RESULTS

5.1 Front Camera Evaluation

Georeferenced front-camera frames (1665 images at 5 fps) were processed in Pix4Dmatic to generate two point clouds: a direct georeferencing solution using our camera positions without ground control, and a GCP-constrained solution using 10 GCPs. Of the 1665 images, 1391 (83.5%) were calibrated in both reconstructions, producing 748,922 tie points and an average GSD of 5.2 cm.

A total of 71 surveyed points were available for evaluation. Of these, 10 served as GCPs in the constrained solution and 15 were identifiable on well-defined surfaces (curb edges, pavement markings, utility covers) and used as independent check points. The remaining points fell on ambiguous surfaces (grass, unmarked concrete) where precise photo identification was not feasible and were excluded from quantitative assessment.

The direct georeferencing solution showed check point residuals ranging from 20–40 cm in 3D. The GCP-constrained solution reduced check point residuals to 7–20 cm in 3D, with GCP RMSE of 0.037 m, confirming the control points constrained the bundle adjustment. Table 1 summarizes the results.

Table 1 summarizes the direct georeferencing accuracy and check point residuals.

Table 1. Georeferencing accuracy: direct vs. GCP-constrained bundle adjustment.

Metric	Direct Georef.	10 GCPs
Images calibrated	1391 / 1665 (83.5%)	
Tie points	748,922	
GCP RMS (m)	—	0.037
Check Point residual range, 3D (m)	0.20 – 0.40	0.07 – 0.20
Check Points used	25	15

5.2 Error Budget Analysis

The total georeferencing error decomposes into independent components. PPK GNSS positioning with a local base station contributes 0.4–1.2 cm (sdn=4 mm, sde=4 mm, sdu=11 mm). Time synchronization between the 5 fps frame rate and 5 Hz GNSS, after linear interpolation at 11.2 m/s driving speed, introduces 2–6 cm. The RTK-measured lever arm between GNSS antenna and camera adds 1–3 cm. Camera calibration error projects to 2–8 cm depending on range. PnP localization contributes 3–10 cm depending on control distribution. Rolling shutter, compensated in Pix4Dmatic, leaves a residual 1–2 cm.

The root sum of squares predicts a combined uncertainty of 20–40 cm for direct georeferencing, which matches the observed check point residuals. GCP constraint absorbs the systematic components (timing bias, lever arm residuals, calibration drift) and reduces the residual to 7–20 cm, leaving primarily PnP localization and check point identification error.

Table 2. Error budget for vehicle-based consumer video mapping.

Error Source	Magnitude	Notes
GNSS (PPK)	0.4–1.2 cm	RS3 PPK, local base, sdn=4mm
Time synchronization	2–6 cm	5 fps + 5 Hz GNSS, 20–50 ms eff.
Lever arm	1–3 cm	RTK-measured, rotated by heading
Camera calibration	2–8 cm	1.8 px RMSE at 5–20 m range
PnP localization	3–10 cm	Control point distribution dependent
Rolling shutter	1–2 cm	Compensated in Pix4Dmatic
Combined (RSS)	**5–15 cm**	**Root sum of squares**

The direct georeferencing solution produced check point residuals of 20–40 cm in 3D, driven primarily by time synchronization and lever arm calibration errors. GCP-constrained bundle adjustment with 10 control points reduced check point residuals to 7–20 cm, demonstrating that the consumer-vehicle imagery supports photogrammetric processing when adequate ground control is provided. This places the GCP-constrained solution in the 10–20 cm class, within the accuracy requirements for DOT asset inventory applications (10–30 cm), though above the 1–3 cm achieved by professional vehicle-mounted mobile mapping systems. The resulting SfM point cloud and GCP distribution are shown in Figure 5. Hardware-level time synchronization and tighter IMU integration are expected to reduce the direct georeferencing bias and lessen reliance on GCPs.

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-341-2026

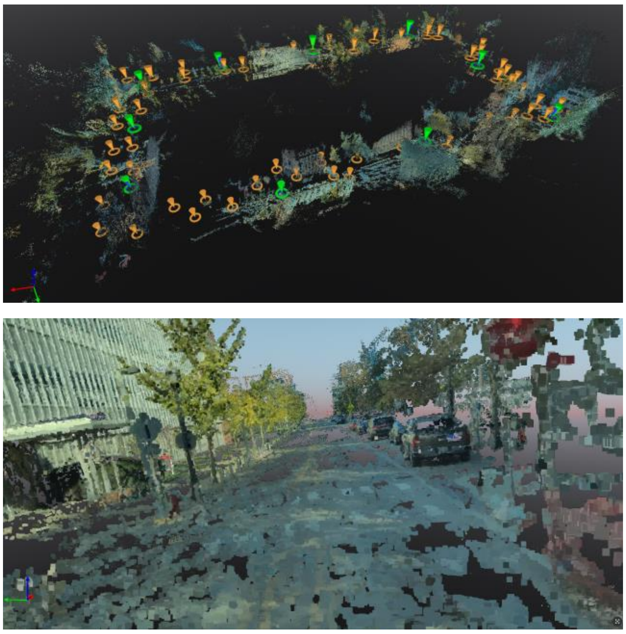

Figure 5. SfM point cloud from Pix4Dmatic: (top) plan view with GCP distribution (green = control, orange = check points), (bottom) oblique 3D view of campus corridor showing road surface and building reconstruction.

6. DISCUSSION & CONCLUSIONS

6.1 Advantages and Limitations

The consumer vehicle approach enables continuous, high-frequency data collection on a scale. Tesla's millions of deployed vehicles equipped with cameras represent an untapped crowdsourcing platform for geospatial data. Unlike traditional crowdsourcing, calibrated georeferenced video frames can produce metric-quality 3D reconstructions, not just 2D annotations.

Error sources include file-timestamp synchronization (sub-second errors translated to meter-level position errors at highway speeds), camera calibration stability (thermal effects, vibration, windshield refraction), precise lever arm knowledge, and camera design trade-offs (rolling shutter, auto-exposure, compression artifacts, dynamic range).

6.2 System Performance and Applicability

We evaluated the front camera alone on a 1.2 km campus test route using a Tesla Model Y with roof-mounted PPK GNSS. From 1665 front-camera frames at 5 fps, 83.5% calibrated successfully, yielding 748,922 tie points and 5.2 cm GSD. Direct georeferencing exhibited 20–40 cm residuals. GCP-constrained bundle adjustment with 10 control points reduced residuals to 7–20 cm (3D), placing the system within DOT asset inventory requirements (10–30 cm). This represents a decimeter-level accuracy class suitable for road sign inventory, pavement assessment, lane marking detection, and vegetation monitoring—applications where consumer-vehicle cost and collection frequency offset accuracy trade-offs versus professional MMS (1–3 cm).

6.3 Future Work

Planned improvements include leveraging additional vehicle cameras (pillar, repeater, and rear) for multi-view georeferencing and expanded spatial coverage. Multi-antenna GNSS configurations will enable attitude estimation by constraining a rigid baseline geometry; as the vehicle maneuvers, the preserved triangle formed by three or more receivers provides direct orientation measurement independent of camera calibration or inertial data. Integration with our custom mobile mapping platform, equipped with dual VLP-16 LiDAR and a SPAN GNSS/INS system, will provide direct comparison of camera-based SfM against survey-grade tightly coupled inertial navigation, benchmarking the consumer-vehicle approach against production-grade mobile mapping standards.

ACKNOWLEDGEMENTS

This research was supported by the SPIN Laboratory at The Ohio State University. Special thank you to Emlid and Pix4D for providing the software and hardware used in this research. Their support for this research is gratefully acknowledged.

REFERENCES

Colomina, I., Molina, P., 2014: Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS J. Photogramm. Remote Sens., 92, 79–97.

Elhashash, M., Albanwan, H., Qin, R., 2022: A review of mobile mapping systems: From sensors to applications. Sensors, 22(11), 4262.

Krausková, D., Mikita, T., Hrůza, P., Kudrnová, B., 2025: Accuracy assessment of iPhone LiDAR for mapping streambeds and small water structures in forested terrain. Sensors, 25(19), 6141.

Kurşun, H., 2023: Accuracy comparison of mobile mapping system for road inventory. Mersin Photogramm. J., 5(2), 55–66.

Lepetit, V., Moreno-Noguer, F., Fua, P., 2009: EPnP: An accurate O(n) solution to the PnP problem. Int. J. Comput. Vision, 81(2), 155–166.

Nex, F., Remondino, F., 2014: UAV for 3D mapping applications: A review. Applied Geomatics, 6(1), 1–15.

Novak, K., 1995: Mobile mapping technology for GIS data collection. Photogramm. Eng. Remote Sens., 61(5), 493–501.

Pepe, M. D., Sobek, J. S., Zimmerman, D. A., 2023: Evaluation of Tesla Dashcam Video System for Speed Determination Via Reverse Projection Photogrammetry. SAE Technical Paper 2023-01-0629.

Scolamiero, V., Boccardo, P., La Riccia, L., 2025: Mobile Mapping System for Urban Infrastructure Monitoring: Digital Twin Implementation in Road Asset Management. Land, 14(3), 597.

Tamimi, R., Toth, C., 2023: Performance Assessment of a Mini Mobile Mapping System: iPhone 14 Pro Installed on an E-Scooter. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 48, 307–315.

Toth, C., Grejner-Brzezinska, D., 2004: Redefining the Paradigm of Modern Mobile Mapping. Photogramm. Eng. Remote Sens., 70(6), 685–694.

Wu, Y., Tang, F., Li, H., 2020: Image-based camera localization: an overview. Visual Comput. Ind. Biomed. Art, 1(1), 8.

Zhang, Z., 2000: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell., 22(11), 1330–1334.