The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B2-2026
XXV ISPRS Congress 2026 "From Imagery to Understanding", Commission II, 4–11 July 2026, Toronto, Canada

From Image Space to Geospatial Space: A Camera Calibration Methodology for Video-Based Traffic Monitoring

D. Shokri[1,2], C. Larouche[1,2], S. Homayouni[2,3]

[1] Département des Sciences Géomatiques, Université Laval, Québec, Canada - (danesh.shokri.1, christian.larouche)@ulaval.ca
[2] Centre de Recherche en Données et Intelligence Géospatiales (CRDIG), Université Laval, Québec, Canada
[3] Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Québec, Canada - saeid.homayouni@inrs.ca

Keywords: Camera Calibration, Traffic Monitoring, Photogrammetry, Drone Imagery, Vehicle Detection, Geospatial Analysis, Point Cloud, Orthophoto

Abstract

This paper presents a novel methodological framework for georeferenced traffic monitoring that bridges the gap between image-based vehicle detection and geospatial analysis. Traditional video-based traffic monitoring systems operate exclusively in image space, limiting their utility for applications requiring physical measurements and integration with other geospatial datasets. We address this limitation by developing a comprehensive camera calibration approach that leverages readily available geospatial data—including smartphone video ground control point selection, a hierarchical calibration algorithm for camera parameter estimation, and a coordinate transformation approach for mapping image-space vehicle detections to geographic space. Experimental results demonstrate the effectiveness of our strategy, achieving a mean reprojection error of 3 pixels across the calibration points. We showcase the practical utility of the framework through a case study of multi-lane traffic monitoring, where vehicle detections are successfully transformed from image coordinates to geographic coordinates, enabling lane-specific traffic analysis and potential integration with traffic simulation models. The methodology presented in this paper allows for precise mathematical relationships between image coordinates, drone-derived orthophotos, and 3D point cloud data, thereby establishing exact correspondences between image coordinates and real-world geographic coordinates. The proposed methodology includes a robust workflow for urban planning by connecting conventional video surveillance with the rich analytical capabilities of geographic information systems, using only commonly available data sources and equipment.

1. Introduction

Traffic monitoring and analysis constitute fundamental components of modern intelligent transportation systems (ITS), providing essential data for urban planning, transportation management, and infrastructure development (Devarshi et al., 2025, Shokri et al., 2023). With increasing urban population and vehicle density, the need for accurate, scalable, and cost-effective traffic monitoring solutions has become more pressing than ever (de França et al., 2012). Traditional approaches to traffic monitoring have evolved from manual counting methods to sophisticated sensor networks, including inductive loop detectors, microwave radar sensors, and specialized traffic cameras. While these systems provide valuable data, they often require significant infrastructure investment, regular maintenance, and face challenges related to coverage limitations and deployment costs (Divasson-J. et al., 2025).

1.1 Evolution of Traffic Monitoring Approaches

The evolution of traffic monitoring technologies reflects a continuous effort to balance accuracy, coverage, and cost-effectiveness. Early traffic monitoring systems relied primarily on physical sensors installed in roadways, such as inductive loops that detect vehicle presence via electromagnetic induction (Harrou et al., 2024). These systems, while reliable for point measurements, provide limited spatial coverage and require invasive installation procedures that disrupt traffic flow and damage road surfaces. More recent approaches have shifted toward non-intrusive sensing technologies, including radar, lidar, and video-based systems that can monitor multiple lanes simultaneously without physical road modifications (Mei et al., 2025).

Video-based traffic monitoring has gained particular attention due to its rich information content, declining hardware costs, and improving computer vision capabilities (Nijhoff, 1986). Early video processing systems focused primarily on basic detection and counting, using techniques such as background subtraction and blob analysis. More recent approaches leverage advanced computer vision and machine learning techniques, particularly deep learning-based object detection and tracking, to identify vehicles with increasing accuracy even in challenging conditions (Nocua M. et al., 2025, Shokri et al., 2024).

Despite these advancements, most video-based traffic monitoring systems operate exclusively in image space, detecting and tracking vehicles in pixel coordinates without establishing a connection to real-world geographic coordinates. This limitation reduces the utility of video data for applications that require physical measurements, such as speed estimation, trajectory analysis, and integration with other geospatial datasets commonly used in transportation planning and management.

1.2 The Role of Orthophotos and Point Clouds in Camera Calibration

Transforming video-based vehicle detections into geographic coordinates requires determining the camera's intrinsic parameters and its pose (position and orientation) relative to a known coordinate system. Traditional camera calibration approaches rely on checkerboard patterns or surveyed ground control points, which require dedicated field procedures and specialized equipment (Zhang, 2000).

An alternative approach leverages existing geospatial products, specifically, orthophotos and 3D point clouds—that are increasingly available for urban areas through aerial surveys, lidar campaigns, and drone-based photogrammetry (Oluwatobi et al., 2021). Orthophotos provide accurate planimetric (XY) coordinates of identifiable features, while point clouds provide corresponding elevation (Z) values. Together, these enable the extraction of 3D ground control points visible in both the reference data and target video frames.

Several researchers have explored this approach for surveillance (Milosavljević et al., 2017), traffic monitoring (Dubská et al., 2014), and autonomous driving applications. However, most methods require specialized knowledge or dedicated calibration procedures that limit accessibility for routine traffic monitoring. A practical workflow that integrates readily available drone-derived products with smartphone video for traffic applications remains lacking.

1.3 Bridging Simulation and Real-World Traffic Monitoring

An important yet often overlooked challenge in traffic research involves the gap between simulation models and real-world traffic data. Traffic simulation software such as SUMO (Simulation of Urban Mobility), VISSIM, and Aimsun are widely used for planning, designing, and evaluating transportation systems (Pereira et al., 2024). These tools generate detailed vehicle trajectories in geographical coordinates that can be readily integrated with other geospatial data. However, validating these simulations against real-world observations typically requires a manual and often qualitative process that lacks spatial precision.

Establishing a methodological bridge between simulation environments and real-world video observations requires two key capabilities: (1) accurate detection and tracking of vehicles in video data, and (2) precise transformation of these detections from image space to geographic coordinates. While significant progress has been made on the first challenge through advances in computer vision, the second challenge of accurate georeferencing remains an active area of research with substantial practical implications.

1.4 Research Objectives and Contributions

This study addresses these challenges by developing a comprehensive methodological framework that links conventional video-based traffic monitoring with geospatial analysis through camera calibration and coordinate transformation. The approach integrates multiple readily available data sources—including smartphone video footage, drone-derived orthophotos, and 3D point cloud data—to establish accurate camera calibration parameters and transform image-space observations into geographic coordinates.

The primary contributions of this research include:

1. A robust camera calibration methodology that leverages readily available geospatial data to determine intrinsic and extrinsic camera parameters without requiring specialized calibration targets or equipment;

2. A practical workflow for selecting optimal ground control points from orthophoto and point cloud data, emphasizing points that are both precisely identifiable and geometrically advantageous for calibration;

3. An approach for transforming image-space vehicle detections to geographic coordinates, enabling physical measurements, spatial analysis, and direct comparison with simulated traffic data;

4. A quantitative assessment of calibration accuracy and its implications for various traffic monitoring applications, including vehicle counting, speed estimation, and trajectory analysis;

5. A demonstration of how georeferenced video analysis can serve as a validation mechanism for traffic simulation models, providing a more rigorous and quantitative approach to simulation verification.

The proposed methodology offers several advantages over existing approaches. Unlike methods that require specialized calibration targets or dedicated survey equipment, our approach uses geospatial data increasingly available in many urban areas. The methodology applies to both dedicated traffic cameras and opportunistic video sources such as smartphone recordings, offering flexibility for various deployment scenarios. Additionally, by establishing a direct link to geographic coordinates, our approach facilitates seamless integration with geographic information systems (GIS) and traffic simulation environments, enabling more comprehensive analysis and validation processes.

2. METHODOLOGY

2.1 Camera Calibration and Pose Estimation Framework

This section presents a robust framework for camera calibration and pose estimation leveraging orthophoto and point cloud data production. The methodology establishes precise mathematical relationships between three-dimensional (3D) world coordinates and two-dimensional (2D) image coordinates, which is fundamental for accurate spatial georeferencing of detected vehicles in traffic monitoring applications.

2.2 Mathematical Formulation of Camera Model

The proposed approach employs the standard pinhole camera model augmented with lens distortion correction. This model decomposes the imaging process into intrinsic parameters that characterize the camera's internal properties and extrinsic parameters that define its pose within the world coordinate system.

2.2.1 Intrinsic Parameter Formulation The camera's intrinsic properties are encapsulated in the calibration matrix K:

$$K = \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

where f_x and f_y denote the focal lengths in pixel units along the respective axes, s represents the skew coefficient (typically approximating zero in contemporary imaging sensors), and (c_x, c_y) defines the principal point coordinates.

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-1199-2026 | © Author(s) 2026. CC BY 4.0 License.

2.2.2 Extrinsic Parameters and Pose Estimation

The extrinsic matrix parameterizes the camera's pose within the global coordinate system $[R|t]$:

$$[R|t] = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{pmatrix} \tag{2}$$

The rotation matrix $R \in SO(3)$ describes the camera's orientation, while the translation vector $t \in \mathbb{R}^3$ specifies its position. These parameters jointly constitute the camera pose, which can be alternatively represented using quaternions or Euler angles for rotation to mitigate singularity issues.

The camera's position in world coordinates can be derived as $C = -R^T t$, providing the exact spatial location of the optical center. The camera's viewing direction is determined by the third row of R, which corresponds to the optical axis direction.

2.2.3 Comprehensive Projection Model

The transformation pipeline from a 3D world point $\mathbf{X}_w = [X_w, Y_w, Z_w]^T$ to a 2D image point $\mathbf{x} = [u, v]^T$ follows a sequential process:

Step 1: Coordinate transformation from world to camera reference frame:

$$\mathbf{X}_c = R\mathbf{X}_w + t \tag{3}$$

Step 2: Perspective projection to normalized image coordinates:

$$\mathbf{x}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} = \begin{pmatrix} X_c/Z_c \\ Y_c/Z_c \end{pmatrix} \tag{4}$$

Step 3: Application of lens distortion model:

$$\begin{aligned} \mathbf{x}_d &= \begin{pmatrix} x_d \\ y_d \end{pmatrix} \\ x_d &= x_n(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 x_n y_n + p_2(r^2 + 2x_n^2) \\ y_d &= y_n(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y_n^2) + 2p_2 x_n y_n \end{aligned} \tag{5}$$

where $r^2 = x_n^2 + y_n^2$ denotes the squared radial distance, $\{k_1, k_2, k_3\}$ are the radial distortion coefficients, and $\{p_1, p_2\}$ are the tangential distortion coefficients accounting for potential misalignment of the lens components.

Step 4: Transformation to pixel coordinates via the intrinsic matrix:

$$\mathbf{x} = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f_x x_d + c_x \\ f_y y_d + c_y \end{pmatrix} \tag{6}$$

The complete projection can be expressed as the mapping function:

$$\mathbf{x} = \Pi(K, R, t, \mathbf{d}, \mathbf{X}_w) \tag{7}$$

where $\mathbf{d} = [k_1, k_2, k_3, p_1, p_2]^T$ is the distortion coefficient vector and Π represents the composite projection operator.

2.3 Rigorous Parameter Estimation Using Geospatial Data

This study presents a novel methodology for camera parameter estimation that leverages the synergistic integration of ortho-images and point cloud data to achieve high-precision calibration results.

2.3.1 Strategic Ground Control Point Selection

The accuracy and reliability of the calibration fundamentally depend on the quality and distribution of Ground Control Points (GCPs). We employ a systematic approach for GCP selection based on the following criteria:

- **Geospatial Precision:** Points must be unambiguously identifiable in both the ortho-image and video frames with sub-pixel accuracy;
- **Spatial Distribution:** Points should be optimally distributed to span the entire field of view and incorporate varying depths to enhance parameter estimation stability.
- **Temporal Invariance:** Selected features must exhibit minimal temporal variation to ensure consistent identification across sequential frames.
- **Radiometric Distinctiveness:** Points should possess high contrast and distinctive textural properties to facilitate precise localization.

Road markings, permanent infrastructure elements, and invariant terrain features constitute ideal candidates for GCPs. For each selected point, we establish a correspondence between:

- Georeferenced 3D coordinates (X_w, Y_w, Z_w) extracted from the orthophoto and point cloud
- Corresponding 2D image coordinates (u, v) identified in the video frames

The elevation component (Z_w) derived from point cloud data provides critical depth information that significantly improves the accuracy of camera pose estimation, which cannot be determined adequately from the orthophoto alone.

2.3.2 Hierarchical Calibration Algorithm

The proposed calibration methodology follows a hierarchical optimization strategy:

Stage 1: Initial Pose Estimation

The camera pose is initially estimated using the Direct Linear Transform (DLT) algorithm (Abdel-Aziz et al., 2015), without accounting for distortion effects. For a set of n point correspondences $\{(\mathbf{X}_w^i, \mathbf{x}^i)\}_{i=1}^n$, we formulate the homogeneous system:

$$A\mathbf{p} = \mathbf{0} \tag{8}$$

where $A \in \mathbb{R}^{2n \times 12}$ is constructed from the point correspondences, and $\mathbf{p} \in \mathbb{R}^{12}$ contains the elements of the projection matrix $P = K[R|t]$. This system is solved via Singular Value Decomposition (SVD) to obtain the minimum norm solution.

Stage 2: Parameter Decomposition

The estimated projection matrix P is decomposed into its constituent intrinsic and extrinsic components using QR decomposition:

$$P = K[R|t] \tag{9}$$

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-1199-2026 |

This decomposition ensures that K is upper-triangular with positive diagonal entries and R satisfies the orthogonality constraint $R^T R = I$.

Stage 3: Distortion Parameter Estimation

Using the initial estimates of intrinsic and extrinsic parameters, we compute the expected undistorted normalized coordinates $\mathbf{x}_n^i$ for each GCP and subsequently estimate the distortion coefficients by minimizing the functional:

$$\min_{\mathbf{d}} \sum_{i=1}^{n} \|\mathbf{x}^i - \hat{\mathbf{x}}^i(\mathbf{d})\|^2 \tag{10}$$

where $\hat{\mathbf{x}}^i(\mathbf{d})$ represents the predicted image point after applying the distortion model with coefficients $\mathbf{d}$.

Stage 4: Global Optimization

All parameters are simultaneously refined using the Levenberg-Marquardt algorithm to minimize the reprojection error:

$$E(K, R, t, \mathbf{d}) = \sum_{i=1}^{n} \|\mathbf{x}^i - \Pi(K, R, t, \mathbf{d}, \mathbf{X}_w^i)\|^2 \tag{11}$$

This non-linear optimization is initialized with the parameters estimated in the preceding stages and incorporates robust loss functions to mitigate the influence of potential outliers.

2.4 Rigorous Validation and Uncertainty Analysis

We implement a comprehensive validation protocol to assess the calibration accuracy and quantify parameter uncertainty:

2.4.1 Reprojection Error Analysis For each GCP, we compute the Euclidean reprojection error:

$$e_i = \|\mathbf{x}^i - \Pi(K, R, t, \mathbf{d}, \mathbf{X}_w^i)\| \tag{12}$$

We analyze both the mean reprojection error:

$$\bar{e} = \frac{1}{n} \sum_{i=1}^{n} e_i \tag{13}$$

and the Root Mean Square Error (RMSE):

$$e_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2} \tag{14}$$

Additionally, we compute the spatial distribution of reprojection errors to identify potential systematic biases in the calibration.

2.4.2 Geometric Accuracy Assessment and Pose Validation We evaluate the geometric accuracy of the estimated camera pose by computing the relative error between actual world distances and their corresponding image-derived values. For pairs of points $(\mathbf{x}^i, \mathbf{x}^j)$ with corresponding world coordinates $(\mathbf{X}_w^i, \mathbf{X}_w^j)$, we quantify:

$$\epsilon_{ij} = \left| \frac{d_{ij}^{world} - d_{ij}^{image}}{d_{ij}^{world}} \right| \times 100\% \tag{15}$$

where d_{ij}^{world} represents the Euclidean distance between world points and d_{ij}^{image} is the distance between the back-projected image points.

Furthermore, we validate the estimated camera pose by comparing the computed position and orientation with independently measured ground truth values when available, or through consistency checks across multiple calibration runs with varying GCP configurations.

3. Data Collection

This section details the multiple data sources and collection methodologies employed in this study to ensure comprehensive spatial coverage and multi-perspective analysis capabilities.

3.1 Video Data Collection

Traffic monitoring video data was collected using an iPhone 14 smartphone positioned at a strategic location overlooking the study area on Robert-Bourassa Highway near Université Laval, Quebec City, Canada. The video was recorded at 4K resolution (3840 × 2160 pixels) at 30 frames per second. The smartphone was mounted on a stable tripod to minimize vibrations and ensure consistent framing throughout the recording period. (Figure 2)

3.2 Orthophoto and Point Cloud Data

High-resolution orthophoto and 3D point cloud data of the study area were acquired using a DJI Phantom 4 UAV and processed using Pix4D photogrammetric software. The orthophoto provides planimetric (XY) coordinates, while the point cloud provides corresponding elevation (Z) values for GCP extraction.

All coordinates are referenced to the NAD83 MTM Zone 7 coordinate system, which is the standard projected coordinate system for the Quebec City region. The study area covers approximately a 15 m × 11 m section of the roadway, as determined by the spatial extent of the selected ground control points (Figure 3).

4. Camera Calibration Results and Analysis

4.1 Intrinsic Parameters Analysis

The camera calibration procedure successfully determined the intrinsic parameters through point cloud-based correspondence matching. The derived camera matrix reveals several important characteristics about the imaging system. The focal length values of 848 pixels in both horizontal and vertical directions indicate a camera with square pixels and no significant skew, which is typical of modern digital cameras. This focal length,

Figure 1. A sample view of a major roadway near the Laval University campus.

combined with the principal point location at (424, 232) pixels, suggests an image resolution likely around 848×464 pixels or similar, as the principal point is positioned near the theoretical center. In practical photogrammetric applications, this simplifies the projection model significantly, as no distortion correction is required during subsequent image processing operations.

4.2 Extrinsic Parameters and Spatial Orientation

The extrinsic parameters (Table 1) reveal significant insights about the camera's position and orientation within the surveying coordinate system. The camera location at coordinates (245027.47, 5182626.64, 97.24) meters indicates operation within a projection coordinate system, likely a regional mapping grid, here a MTM (Provincial or State Plane coordinates system). The translation vector (3494305.47, 485298.62, 3804467.37) represents the camera center in the rotated coordinate system, and its magnitude indicates significant rotation between the camera and world coordinate frames. The Euler angles provide crucial information about camera orientation: the X-rotation of 98.42° suggests the camera is tilted significantly from the horizontal plane, the Y-rotation of 39.43° shows substantial pitch, and the Z-rotation of 175.48° indicates the camera is nearly inverted relative to the standard orientation. This complex orientation suggests either an unusual mounting configuration or a specialized surveying setup, often used for oblique aerial photography or ground-based terrestrial photogrammetry with unconventional viewing angles.

4.3 Control Point Distribution and Reprojection Accuracy Assessment

The spatial distribution of control points and their corresponding reprojection errors provide critical insights into calibration quality and reliability. The seven 3D control points span approximately 11 meters in the X-direction and 15 meters in the Y-direction, covering a localized area suitable for detailed photogrammetric analysis (Table 2). If we do not considerd the elevation of Point 7, which seems to have an erroneous Z value (8,03 m too high), the other 6 points elevations range from 90,15 m to 90,38m. The reprojection error analysis reveals concerning variations in measurement quality. The mean reprojection error of 5.80 pixels, while within acceptable ranges for many applications, indicates room for improvement. Point 1's excessive error of 13.13 pixels strongly suggests this point should be classified as an outlier and potentially removed from the calibration dataset. This large error could result from incorrect point identification in either the point cloud or image, measurement imprecision, or systematic issues with the correspondence. Conversely, Point 6's good accuracy of 0.70 pixels demonstrates that high-precision correspondence is achievable with proper measurement techniques. The standard deviation of reprojection errors across all points is approximately 4.2 pixels, indicating significant variability in measurement quality that warrants investigation of the point selection and measurement procedures.

The camera calibration process yielded promising results, as evidenced by the analysis of reprojection errors between the originally selected ground control points and their estimated positions after parameter optimization. Figure 3 illustrates the spatial distribution of the seven ground control points (depicted in red) alongside their reprojected counterparts (shown in green) following the calibration procedure.

The quantitative assessment of reprojection errors revealed a mean error of 2.94 pixels across all points, indicating a reasonable calibration quality considering the challenging nature of the urban environment and the relatively low number of control points utilized. Point-specific analysis showed varying degrees of accuracy, with errors ranging from approximately 1.30 pixels (Point 3) to 5.71 pixels (Point 2). The discrepancy in reprojection accuracy across points can be attributed to factors such as the geometric configuration of the points, potential occlusions, and inherent challenges in feature identification.

Through this calibration procedure, we successfully estimated the camera's geographic position and orientation within the world coordinate system. This geo-orientation establishes a robust mathematical relationship between the image coordinate system and real-world spatial references, enabling accurate geor-

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-1199-2026 | © Author(s) 2026. CC BY 4.0 License.

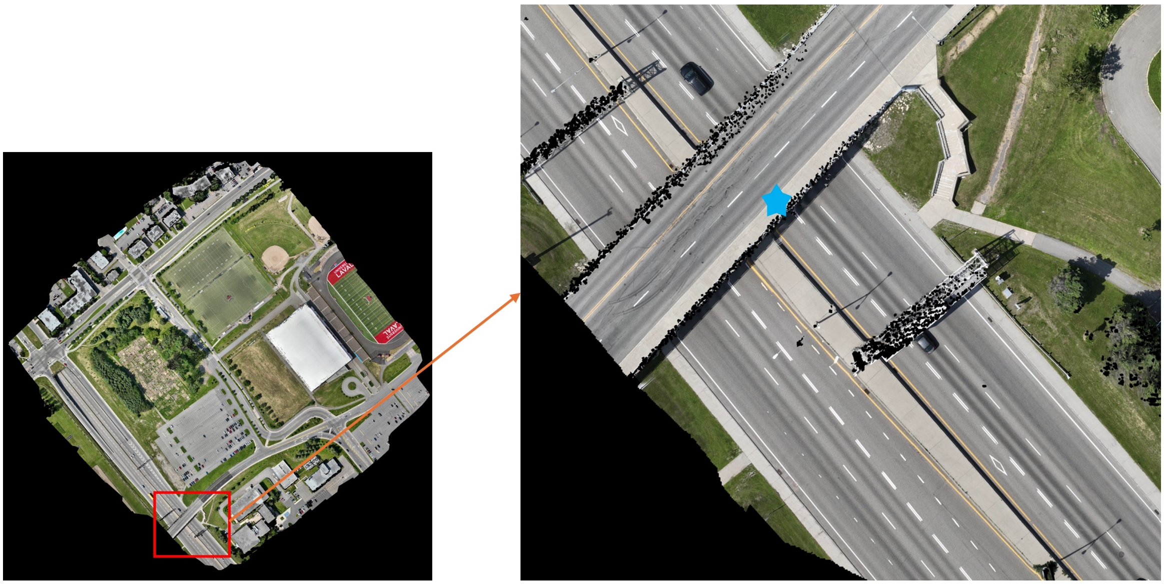

Figure 2. At left: Orthophoto of the study area. At right: The camera location (blue star on the orthophoto) of the recorded video frames.

Table 1. Camera Extrinsic Parameters and Coordinate Transformations

Parameter	Value	Description
Camera X Position	245027.47 m	World coordinate X
Camera Y Position	5182626.64 m	World coordinate Y
Camera Z Position	97.24 m	Elevation above datum
Translation X	3494305.47 m	Rotated frame translation
Translation Y	485298.62 m	Rotated frame translation
Translation Z	3804467.37 m	Rotated frame translation
Euler Angle X (Roll)	98.42°	Rotation about X-axis
Euler Angle Y (Pitch)	39.43°	Rotation about Y-axis
Euler Angle Z (Yaw)	175.48°	Rotation about Z-axis

Figure 3. Visualization of the seven selected ground control points (red) and their corresponding reprojected positions (green) after camera calibration and parameter estimation.

eferencing of detected objects within the video frames. The calibrated camera parameters provide the foundation for subsequent analysis of traffic patterns and vehicle movements within a spatially referenced framework, facilitating integration with other geospatial datasets and supporting quantitative spatial analysis of the observed phenomena.

4.4 Calibration Quality Assessment and Recommendations

The overall calibration shows mixed quality indicators, which require careful interpretation for practical applications. The geometric configuration of control points provides adequate spatial distribution for the surveyed area. However, the concentra-

Table 2. Control Point Spatial Distribution and Error Analysis

Point	X (m)	Y (m)	Z (m)	u (px)	v (px)	Error (px)	Quality
1	245030.06	5182608.21	90.28	372	255	13.13	Poor (Outlier)
2	245033.02	5182608.47	90.38	241	252	3.91	Good
3	245029.12	5182613.27	90.25	336	263	3.34	Good
4	245024.61	5182618.08	90.19	328	283	7.73	Moderate
5	245035.55	5182611.36	90.30	292	253	8.26	Moderate
6	245029.07	5182618.28	90.15	272	274	0.70	Excellent
7	245031.21	5182623.60	98.18	56	331	3.51	Good
Spatial Extent: X Range: 10.94 m, Y Range: 15.39 m, Z Range: 8.03 m **Error Statistics:** Mean: 5.80 px, Std Dev: 4.20 px, Max: 13.13 px, Min: 0.70 px							

tion within a relatively small 11×15 meter footprint may limit the extrapolation accuracy beyond this region. The reprojection error statistics reveal a bimodal quality distribution: four points (2, 3, 6, 7) exhibit good to excellent accuracy below 4 pixels, while three points (1, 4, 5) show moderate to poor performance above 7 pixels. This pattern suggests systematic issues in either the measurement methodology or the point selection criteria. For operational use, several improvements are recommended: (1) removal or remeasurement of Point 1 due to its outlier status, (2) verification of Points 4 and 7 correspondences to reduce their elevated errors, (3) addition of control points at greater distances from the current cluster to improve calibration robustness, and (4) implementation of more rigorous point identification procedures in both point cloud and image domains. The achieved accuracy level is suitable for applications requiring meter-level precision but may be insufficient for centimeter-level surveying tasks. Future calibration iterations should target a mean reprojection error below 2 pixels with no individual point exceeding 5 pixels for optimal photogrammetric performance.

4.5 Vehicle Detection and Spatial Transformation

Following successful camera calibration and geo-orientation, we applied the previously described vehicle detection algorithm (Shokri et al., 2023) to the highway video footage. To facilitate lane-specific traffic analysis, we first delineated individual road lanes by drawing reference lines across the roadway in the image space. These lines served as virtual detection zones for counting and classifying vehicles traveling in each lane.

The core innovation in our approach lies in transforming these image-space detection elements into geographic coordinate space. By applying the camera parameters estimated during calibration, we established a direct mapping between pixel coordinates in video frames and real-world geographic positions. This transformation enables several critical capabilities:

1. Accurate positioning of detected vehicles within the geographic reference system;

2. Precise measurements of inter-vehicle distances and lane widths in metric units;

3. Integration of detection results with other georeferenced datasets, such as road network databases;

4. Calculation of true vehicle speeds based on displacement in geographic space rather than image.

Figure 4 illustrates the implementation of our detection system, showing vehicles identified in the video frame alongside the lane delineation lines used for classification and counting. The detection algorithm effectively identifies vehicles across multiple lanes under challenging conditions, including partial occlusions, varying vehicle sizes, and perspective distortion.

The transformation from image to geographic coordinates was implemented by applying the projection model described in Section 2.2.3, utilizing the intrinsic and extrinsic camera parameters obtained during calibration. For each detected vehicle and lane boundary, we performed an inverse mapping from image coordinates (u, v) to world coordinates (X_w, Y_w, Z_w), constraining the solution to the known road surface elevation derived from the point cloud data.

This approach significantly enhances the utility of traffic monitoring data by anchoring observations to real-world coordinates, thereby facilitating more sophisticated analysis of traffic patterns and enabling integration with geographic information systems for broader transportation management applications.

5. CONCLUSION

This study has demonstrated a comprehensive methodology for integrating multiple data sources to achieve georeferenced traffic monitoring through camera calibration and vehicle detection. By leveraging drone-derived orthophotos and point cloud data for camera calibration, we established a robust mathematical relationship between image space and geographic coordinates, enabling spatially accurate analysis of traffic patterns captured in conventional video footage.

The camera calibration approach achieved a mean reprojection error of 2.94 pixels across five ground control points, providing sufficient accuracy for subsequent traffic analysis applications. The successful transformation of image-space vehicle detections and lane boundaries to geographic coordinates demonstrates the practical utility of the calibration, enabling physical measurements and spatial analyses that enhance the value of video-based traffic monitoring.

Key contributions of this work include: (1) A methodological framework for camera calibration using multiple geospatial data sources, (2) Practical strategies for ground control point selection and validation, (3) A complete workflow from data collection through calibration to georeferenced vehicle detection, and (4) Quantitative assessment of calibration accuracy and its implications for traffic monitoring applications.

In contrast to the classical stereophotogrammetric approach, the proposed procedure used a single image to transform 2D image coordinates into 3D world coordinates. The proposed method may currently present accuracy limitations, but is better suited

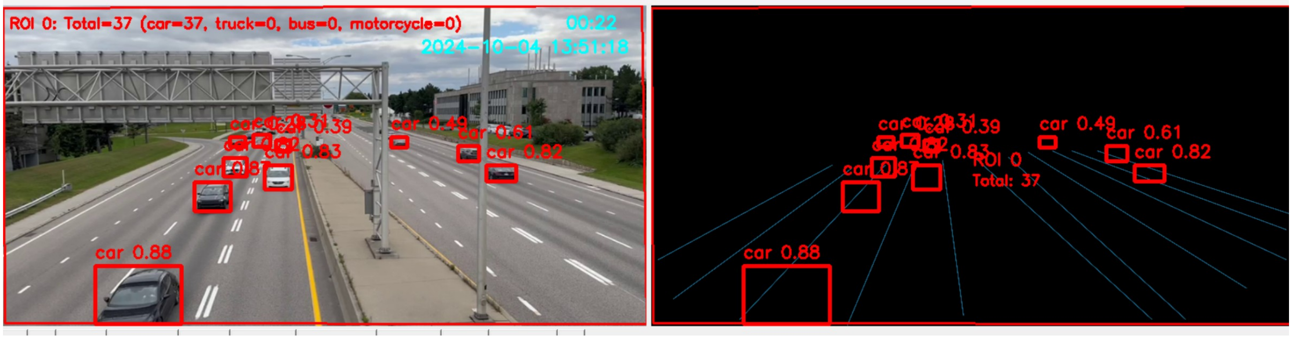

Figure 4. Vehicle detection results with lane delineation lines. Detected vehicles are highlighted with bounding boxes, while colored lines indicate the lane boundaries used for vehicle counting and classification.

for processing traffic images coming from a growing number of single surveillance cameras distributed over city roads and highways.

ACKNOWLEDGEMENTS

The authors would like to thank prof. Willian Ney Cassol and his team for providing high-resolution orthophoto and 3D point cloud data of the study area.

References

Abdel-Aziz, Y. I., Karara, H. M., Hauck, M., 2015. Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry*. *Photogrammetric Engineering and Remote Sensing*, 81(2), 103-107.

de França, J., Stemmer, M., de M. França, M., Piai, J., 2012. A new robust algorithmic for multi-camera calibration with a 1D object under general motions without prior knowledge of any camera intrinsic parameter. *Pattern Recognition*, 45(10), 3636–3647.

Devarshi, P., Yash, C., Patel, M., Raval, D., 2025. A Comprehensive Deep Learning Model for Improved Person Reidentification Using Multi-Camera Streaming Pipeline. *Procedia Computer Science*, 258, 455–466.

Divasson-J., A., Macarulla, A., Garcia, J., Borges, C., 2025. Agent-based modeling in urban human mobility: A systematic review. *Cities*, 158, 105697.

Dubská, M., Herout, A., Juránek, R., Sochor, J., 2014. Automatic camera calibration for traffic understanding. *Proceedings of the British Machine Vision Conference (BMVC)*.

Harrou, F., Zeroual, A., Kadri, F., Sun, Y., 2024. Enhancing road traffic flow prediction with improved deep learning using wavelet transforms. *Results in Engineering*, 23, 102342.

Mei, P., Karimi, H., Ou, L., Xie, H., Zhan, C., Li, G., Yang, S., 2025. Driving style classification and recognition methods for connected vehicle control in intelligent transportation systems: A review. *ISA Transactions*, 158, 167–183.

Milosavljević, A., Rančić, D., Dimitrijević, A., Predić, B., Kovačević, V., 2017. A Method for Estimating Surveillance Video Georeferences. *ISPRS International Journal of Geo-Information*, 6(7), 211.

Nijhoff, F., 1986. The direct linearizing transform for three-dimensional lattice equations. *Physica D: Nonlinear Phenomena*, 18(1), 380–381.

Nocua M., F., Pérez-Holguín, W.-J., Pardo-Beainy, C., 2025. Urban traffic monitoring based on deep learning on an embedded GPU. *Expert Systems with Applications*, 273, 126847.

Oluwatobi, A., Tayo, A., Oladele, A., Adesina, G., 2021. The design of a vehicle detector and counter system using inductive loop technology. *Procedia Computer Science*, 183, 493–503.

Pereira, H. d. M., Bessa Júnior, J., de Albuquerque Nóbrega, R., 2024. Geospatial-based decision support system for prioritizing road segments for maintenance and rehabilitation. *Case Studies on Transport Policy*, 16, 101170.

Shokri, D., Larouche, C., Homayouni, S., 2023. A Comparative Analysis of Multi-Label Deep Learning Classifiers for Real-Time Vehicle Detection to Support Intelligent Transportation Systems. *Smart Cities*, 6(5), 2982–3004. https://www.mdpi.com/2624-6511/6/5/134.

Shokri, D., Larouche, C., Homayouni, S., 2024. Proposing an Efficient Deep Learning Algorithm Based on Segment Anything Model for Detection and Tracking of Vehicles through Uncalibrated Urban Traffic Surveillance Cameras. *Electronics*, 13(14). https://www.mdpi.com/2079-9292/13/14/2883.

Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.