Ojala, Risto; Vepsalainen, Jari; Pirhonen, Jesse; Tammi, Kari

Infrastructure camera calibration with GNSS for vehicle localisation

Published in:
IET Intelligent Transport Systems

DOI:
10.1049/itr2.12262

Published: 01/02/2023

Document Version
Publisher's PDF, also known as Version of record

Received: 31 March 2022 | Revised: 20 July 2022 | Accepted: 22 August 2022

IET Intelligent Transport Systems

DOI: 10.1049/itr2.12262

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Infrastructure camera calibration with GNSS for vehicle localisation

Risto Ojala | **Jari Vepsäläinen** | **Jesse Pirhonen** | **Kari Tammi**

Department of Mechanical Engineering, Aalto University, Espoo, Finland

Correspondence
Risto Ojala, Department of Mechanical Engineering, Aalto University, Espoo, Finland.
Email: risto.j.ojala@aalto.fi

Funding information
Henry Ford Foundation Finland; Academy of Finland

Abstract

Intelligent transportation and smart city applications are currently on the rise. In many applications, diverse and accurate sensor perception of vehicles is crucial. Relevant information could be conveniently acquired with traffic cameras, as there is an abundance of cameras in cities. However, cameras have to be calibrated in order to acquire position data of vehicles. This paper proposes a novel automated calibration approach for partially connected vehicle environments. The approach utilises Global Navigation Satellite System positioning information shared by connected vehicles. Corresponding vehicle Global Navigation Satellite System locations and image coordinates are utilised to fit a direct transformation between image and ground plane coordinates. The proposed approach was validated with a research vehicle equipped with a Real-Time Kinematic-corrected Global Navigation Satellite System receiver driving past three different cameras. On average, the camera estimates contained errors ranging from 1.5 to 2.0 m, when compared to the Global Navigation Satellite System positions of the vehicle. Considering the vast lengths of the overlooked road sections, up to 140 m, the accuracy of the camera-based localisation should be adequate for a number of intelligent transportation applications. In future, the calibration approach should be evaluated with fusion of stand-alone Global Navigation Satellite System positioning and inertial measurements, to validate the calibration methodology with more common vehicle sensor equipment.

1 | INTRODUCTION

1.1 | Motivation and background

Transportation systems and city infrastructure are becoming increasingly intelligent, in an attempt to solve many problems of expanding urbanisation. Future technological advancements are relied on to increase the efficiency and safety of transportation. There lies great potential in the co-operation of connected vehicles and roadside sensors, as working in unison they can accurately monitor the overall surrounding traffic. Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) connections allow effortless information amongst vehicles and road users. Information transmitted in such connected systems can be utilised for controlling the traffic or assisting drivers with relevant information. To grasp the full benefit of these connected systems, road infrastructure should be updated with plethora of sensors capable of facilitating smart features. Smart features can range from safety and traffic signal control to urban planning. A potential approach for acquiring information on the infrastructure side would be to more efficiently utilise the traffic cameras already existing in many parts of road networks. Surveillance cameras are all over cities nowadays [1], with traffic being a popular area of monitoring. Cameras can be utilised for detecting the presence of different road users, such as vehicles, pedestrians or cyclists, as well localising them and evaluating their trajectories.

Applying detection and localisation algorithms on surveillance camera feeds in real-time, different connected safety systems could be implemented in the infrastructure. Vehicles and their drivers could be alerted of probable on-coming collisions, or alerts could be sent of hazardous behaviour of others. Such systems have been developed and discussed in the existing literature [2–4]. A smart safety system has also been developed

at our research group, with the goal of detecting road users in occluded intersection areas and warning drivers of probable collisions [5]. In order to use surveillance cameras to their maximum potential in different intelligent transportation systems, reliable localisation information should be acquired from them. This is a challenging issue, as acquiring three-dimensional real-world locations from the two-dimensional pixel information of image data is naturally not a straightforward process.

To improve the applicability of surveillance cameras in intelligent transportation infrastructure, this paper presents a convenient calibration approach based on connected vehicle technology. The goal of the calibration is to accurately transform the image coordinates of vehicles to coordinates on the road. Calibration is achieved via tracking a vehicle in the camera view with known geographic coordinates. Geographic coordinates of the vehicle are acquired from the Global Navigation Satellite System (GNSS), and shared with the infrastructure over V2I-connections. Although such vehicle GNSS positioning data is not currently available in the traffic infrastructure systems, the calibration approach is convenient for future adoption. Modern vehicles are already equipped with GNSS localisation capabilities, and connected vehicle technology will soon be introduced to the market. Connected vehicle technology will enable vehicles to share their GNSS positions via V2I-connections. The presented calibration technology is especially useful for infrastructure-based localisation systems in a future transition period when only a portion of vehicles are equipped with connected technology. This scenario is likely, considering the long lifespan of vehicles [6]. Calibration based on the GNSS coordinates can be considered reliable, as the coordinates offer a closed-loop calibration solution, in which the calibration accuracy of the automated process can be easily evaluated. Furthermore, cameras calibrated with GNSS coordinates are capable of localising vehicles in geographic coordinate systems, allowing seamless utilisation of the data in intelligent transportation systems (ITSs).

1.2 | Scientific contributions

This work presents novel methodology for automatically calibrating traffic cameras for vehicle localisation. The proposed calibration approach carries several benefits compared to previously presented calibration approaches [7–9]. Previous approaches have been limited by the challenge that they must make several assumptions regarding the structure of the visible scene, such as straight roads, known lane width, or certain visible vehicle keypoints. Additionally, these purely image-based approaches do not feature a feedback loop for automatically verifying the resulting calibration. Our novel contribution is a calibration method, that can be carried out in practically any outdoor environment. The proposed method is not reliant on assumptions of geometric clues visible in the structure of the scene, making the approach more generalisable. Furthermore, the GNSS-based calibration approach allows utilising the calibration data for validating the success of the calibration automatically, as the GNSS data inherently allow evaluation of the achieved localisation accuracy. GNSS-based calibration also enables camera-based extraction of vehicle coordinates in a global coordinate system, facilitating effortless usage of the camera localisation data in traffic monitoring systems. Previous approaches have been limited to a local camera-based coordinate system, as no global frame has been provided. Experimental results validating the reliability and accuracy of the proposed calibration approach are presented, highlighting the applicability of the method in actual traffic use cases.

2 | STATE OF THE ART

Calibration is an essential step in systems and devices that apply sensor data to assess their state and perceive the surrounding environment. Investigation of different calibration techniques has become a crucial field of research in robotics [10] and vehicular applications [11], where localisation and state-estimation must be performed accurately and robustly. Camera calibration has been a field of extensive research, as calibration is essential for acquiring reliable geometric information from images. In its simplest form, the calibration problem boils down to finding the camera matrix of a pinhole camera model, which linearly maps real-world 3D coordinates to the 2D camera coordinate system. Direct Linear Transformation (DLT), proposed by Abdel-Aziz and Karara [12], is commonly considered the most basic approach for this task. DLT calibration requires a set of known 3D world coordinate and 2D camera coordinate correspondences, and the camera matrix is found as the least squares solution to the linear problem. However, finding the camera matrix might not be sufficient for many applications. Weng *et al.* [13] have proposed a calibration method based on world coordinate and camera coordinate correspondences, which accounted for distortions in the camera model. In addition, their approach also solved the actual values of the camera intrinsic and extrinsic parameters, and not simply a solution to the linear projection problem. The approach was based on iterative least squares optimisation of the pinhole camera parameters and distortion parameters. In many practical scenarios, finding the 3D world coordinate and 2D camera coordinate correspondences for the calibration can be a laborious task, often requiring carefully constructed calibration setups. An alternative approach to calibrating a camera has been invented utilising vanishing points found in a camera view, as proposed by Caprile and Torre [14]. Vanishing points are theoretical points in the 2D image of a camera, where projected parallel 3D world lines appear to intersect. Utilising vanishing points derived from a camera view, the camera intrinsic parameters as well as the rotation matrix can be calibrated. This can be practical in many man-made environments, where parallel lines often exist. However, as the vanishing points contain no information of scale, the camera translation vector cannot be found via this method alone. Recently, the most popular camera calibration has been the one developed by Zhang [15]. This approach is capable of providing all camera parameters, including distortion parameters, based on multiple images of a planar checkerboard of known size. The approach first applies basic DLT to solve an

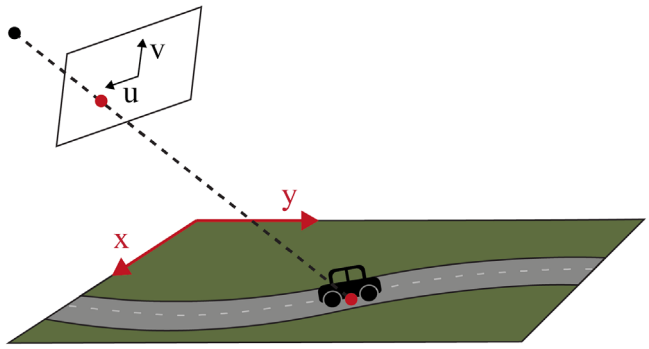

FIGURE 1 Simplified illustration of a traffic camera monitoring a traffic environment

analytical solution, followed by non-linear optimisation. Popularity of the approach is easy to understand, as no complex calibration setups are needed, only a planar pattern has to be shown in different orientation in front of the camera. However, in many real-world scenarios, especially with surveillance and traffic cameras, automated calibration procedures have been favoured, as the sheer number of surveillance cameras around the world is staggering.

Traffic camera calibration has been a widely studied problem among the ITS research community. Commonly the approaches have aimed for maximal automation, so that minimal manual work is required in the calibration procedure. The motive of the calibration has commonly been to acquire vehicle speeds and locations with the cameras, traditional applications including speed-limit violation monitoring and intelligent traffic control. As the ITS applications develop with advanced technology, need for convenient calibration increases for future applications as well. Generally, the calibration procedures and camera-based measurement techniques model the road environment as a plane on which the vehicles are located. This effectively reduces the positioning problem to a mapping from 2D world coordinates to 2D camera coordinates. Visual representation of the simplified problem is depicted in Figure 1. An extensive review presenting a multitude of traffic camera calibration approaches has been written by Sochor *et al.* [8]. Another review on the topic has been published by Kanhere and Birchfield [9], focusing especially on the commonly applied methodology based on vanishing points. As previously stated, vanishing point-based calibration methods are unable to generate a translation vector for the camera. Approaches based on this methodology typically utilise additional information of lane width, camera height, or distance between road and camera to compute a translation vector. This is necessary for acquiring distances or speeds in the camera view.

Vanishing point-based methods have been the most thoroughly investigated topic in traffic camera calibration. Bas and Crisman [16] were pioneers in the field, proposing an approach based on known camera height and tilt angle, as well as two manually marked points on both sides of the road. These roadside points were utilised for finding a vanishing point, demanding the road to be straight. Their measurements showed a relative accuracy of 1%–2% for points near the camera and 4%–6% for points further away on a single camera view. Similar yet more automated calibration methodology was presented by Schoepflin and Dailey [17]. Their approach first determined lanes based on vehicle motion maps, and extracted a vanishing point from the lanes, assuming that the lanes were straight. Another vanishing point, in an orthogonal direction, was found statistically analysing the vehicles with Hough transform. Assuming a known lane width they were able to carry out calibration of the camera. The accuracy of similar camera calibration has been studied by Zheng and Peng in a manual approach [18]. They also utilised two vanishing points, acquiring the first point along the ground from the lane markings and the second point in the vertical direction from lamp posts and other similar objects. Vanishing points were determined based on manual image annotation. Translation vector was estimated based on known dimensions of the lanes and the lane markings, as well as known heights of fences. Testing the accuracy of their approach with two 3.2 m and two 6 m known lengths in the image, they acquired an average relative accuracy of 1.5%. Traffic camera calibration mixing vanishing point detection and geometric road models has been proposed by Dawson and Birchfield [19]. Markov chain Monte Carlo search was utilised to fit the road geometry to the detected vanishing point. They assumed prior knowledge of the number of receding and oncoming lanes, and lane width. Analysing a segment with a known length on the ground plane, their approach had an average relative positioning accuracy of 10%. One of the most recent vanishing point-based methods has been presented by Dubska *et al.* [20]. Their goal was to fully automate the calibration process. Vehicles were assumed to move on a straight path in the camera view, and a vanishing point was acquired from tracking them. Vehicle appearance was utilised for finding an additional vanishing point, which was in the orthogonal direction compared to the other vanishing point, along the ground plane. Known statistics of vehicle sizes were utilised for estimating the translation vector of the camera. The work on the calibration method was continued by Sochor *et al.* [21]. They fine-tuned the method by improving the acquisition of vehicle 3D information, thus defining the scale of the scene more accurately. When measuring segments with known length in the road plane, their approach yielded an average relative error of 3.47%.

Calibration methods not based on vanishing points have been also presented. Manual traffic camera calibration has been explored by Ismail *et al.* [22], who utilised calibration data gathered from ortographic imagery and field-measurements carried out with a measuring wheel. Utilised calibration data included point correspondences, as well as lengths and angles of segments. They applied a novel multi-component loss function to reach calibration, reaching 6.9% relative error in estimating segment lengths. Another traffic camera calibration method based on known geometry in the images has been proposed by Do *et al.* [23]. They placed the corner markings of an equilateral triangle in the camera view, and used the known geometry for finding necessary camera parameters for distance measurements. A more recent automated calibration approach has been proposed by Bhardwaj *et al.* [7]. They applied deep learning for detecting vehicle keypoints, whose relative dimensions were

assumed known based on statistical data of the locally most common sedan models. They assumed known intrinsic camera parameters, as well as suitable point of view of vehicle rears. Necessary vehicle keypoints included vehicle tail lights and side-view mirrors.

Most approaches in traffic camera calibration, especially the automated methods, have been focused on utilising only the information available in the camera view. However, with the emerge of connected vehicles, additional sensor data may be available for the calibration procedure. In outdoor applications, GNSS offers an interesting solution for acquiring real-world coordinates in the camera view, enabling a multitude of calibration approaches. However, positions acquired with regular stand-alone GNSS receivers contain several metres of error, which can hinder the accuracy of the calibration [24–26]. GNSS positioning can be enhanced with different approaches, such as Real-Time Kinematic (RTK) corrections, which can reach centimetre-level accuracy in favourable conditions [27]. These approaches still suffer from some of the common limitations of GNSS navigation, such as signal reflection and non-line of sight issues in urban canyons [28–30]. GNSS-based calibration approaches are commonly applied in vehicles, fusing GNSS positioning information with the on-board camera imagery [31, 32]. Calibration of cameras installed in the infrastructure has been less commonly achieved with GNSS, although some related works exist. GNSS-based surveillance camera calibration has been implemented by Liao *et al.* [33], who performed calibration based on GNSS coordinates acquired from the mobile device of a pedestrian. Their system was designed for automated calibration, with an implementation of trajectory matching between the camera view and GNSS coordinates. Point correspondences of GNSS coordinates and image coordinates were utilised to carry out the well-established calibration method developed by Tsai [34]. Liu *et al.* [35] have also utilised GNSS coordinates, calibrating their stereovision surveillance system based on GNSS coordinates of a passing vehicle. GNSS-based calibration of camera systems has been expanded by Jiang and Sun [36]. With a small unmanned aerial vehicle equipped with a differential GNSS receiver, they visited different locations in the camera view. This yielded varied point correspondences in a notably wide range, ensuring a proper calibration.

As seen in previous literature, multiple calibration approaches for traffic cameras have been presented, both manual and automated calibration. There is an increasing need for automated traffic camera calibration methods, as a growing amount of localisation data are required from urban traffic environments. However, previously presented methods still require some preliminary knowledge or measurements of the scene in order to perform the calibration. This preliminary knowledge typically includes road or vehicle dimensions, or geometry. Due to their core assumptions, many methods can only function in specific environments, such as on straight roads. Furthermore, calibration methods based purely on the geometrical clues of the camera view miss an inherent feedback loop ensuring that the camera has been properly calibrated for the localisation task. We aim to solve the previously presented problems by proposing a novel calibration approach for localisation tasks. Our automated calibration approach utilises point correspondences of vehicle GNSS coordinates and image coordinates to calibrate traffic cameras. The proposed approach is designed for partially connected vehicle environments, where the shared GNSS coordinates of visible connected vehicles can be utilised. Proposed GNSS-based calibration can be carried out in any outdoor scene, not being reliant on assumptions regarding visible geometric clues in the structure of the scene. Furthermore, GNSS coordinates utilised in the calibration provide a feedback loop for testing the automatically calibrated camera, ensuring that the localisation is reliable. GNSS-based calibration also carries the benefit of the camera localisation results being automatically available in a relevant coordinate system. Localisation in the geographic coordinate system enables a multitude of ITS applications with the camera data.

3 | METHODS

3.1 | Calibration for localisation

The camera calibration approach presented here models the scene by placing vehicles on a flat ground plane. This reduces the problem to a 2D-to-2D mapping between world coordinate and pixel coordinate systems. The third dimension is completely omitted here, as it is irrelevant for the task of vehicle localisation in the two-dimensional ground plane. Therefore, a full camera calibration is not carried out and instead a homography [37, 38] is fitted between the image plane and the ground plane. This is achieved by matching multiple pairs of image plane coordinates of vehicles to their known ground plane coordinates. The image plane coordinates are acquired via object detection and tracking, and the ground plane coordinates are acquired with GNSS positioning. The location of a vehicle described by a 2D homogeneous coordinate $\mathbf{x}$, which is defined as

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{1}$$

The image plane coordinates $\mathbf{p}$ of the vehicle are defined in homogeneous coordinates as

$$\mathbf{p} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}. \tag{2}$$

In this paper, the centre point of the vehicle bounding box is chosen to represent $\mathbf{p}$. The described ground plane and image plane coordinate systems is presented in Figure 1. Any ith corresponding homogeneous coordinates in ground plane $\mathbf{x}_i$ and in image plane $\mathbf{p}_i$ are linked via homography as

$$\lambda \mathbf{x}_i = \mathbf{H}\mathbf{p}_i, \tag{3}$$

where the homography is defined by a 3×3 projection matrix $\mathbf{H}$. The scaling factor is denoted by λ.

Based on available point correspondences, the least squares optimal homography matrix can be solved via a homogeneous linear system representation as presented by Hartley and Zisserman [38]. For the corresponding ground plane $\mathbf{x}_i$ and image plane $\mathbf{p}_i$ coordinates, it holds that

$$\mathbf{x}_i \times \mathbf{H}\mathbf{p}_i = 0. \tag{4}$$

This can be manipulated to the form

$$\begin{bmatrix} 0 & -\mathbf{p}_i^T & y_i\mathbf{p}_i^T \\ \mathbf{p}_i^T & 0 & -x_i\mathbf{p}_i^T \\ -y_i\mathbf{p}_i^T & x_i\mathbf{p}_i^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{bmatrix} = 0, \tag{5}$$

where $\mathbf{h}$ denotes a row of the homography matrix. This form provides two linearly independent equations for each point correspondence. Requiring four or more unique point correspondences, a total of n point correspondences are bundled together as

$$\begin{bmatrix} 0 & -\mathbf{p}_1^T & y_1\mathbf{p}_1^T \\ \mathbf{p}_1^T & 0 & -x_1\mathbf{p}_1^T \\ 0 & -\mathbf{p}_2^T & y_2\mathbf{p}_2^T \\ \mathbf{p}_2^T & 0 & -x_2\mathbf{p}_2^T \\ & \vdots & \\ 0 & -\mathbf{p}_n^T & y_n\mathbf{p}_n^T \\ \mathbf{p}_n^T & 0 & -x_n\mathbf{p}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{bmatrix} = 0. \tag{6}$$

In the above homogeneous linear system, we denote the leftside matrix as $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 0 & -\mathbf{p}_1^T & y_1\mathbf{p}_1^T \\ \mathbf{p}_1^T & 0 & -x_1\mathbf{p}_1^T \\ 0 & -\mathbf{p}_2^T & y_2\mathbf{p}_2^T \\ \mathbf{p}_2^T & 0 & -x_2\mathbf{p}_2^T \\ & \vdots & \\ 0 & -\mathbf{p}_n^T & y_n\mathbf{p}_n^T \\ \mathbf{p}_n^T & 0 & -x_n\mathbf{p}_n^T \end{bmatrix}. \tag{7}$$

Constraining the scale of the homography matrix with its Frobenius norm $||\mathbf{H}||_F = 1$, the optimal homography matrix is represented by the eigenvector of $\mathbf{A}^T\mathbf{A}$ corresponding to the smallest eigenvalue. This eigenvector is the reshaped homography matrix which minimises the least squares error

$$\mathcal{L} = \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2, \tag{8}$$

where $\hat{x}$ and $\hat{y}$ denote a ground plane coordinate estimated with the homography matrix. With the calibrated homography matrix, image plane coordinates can be transformed into estimated ground plane coordinates, allowing acquisition of vehicle coordinates directly from images. The homography calibration can be effortlessly carried out once known pairs of ground plane coordinates and image coordinates are available. This calibration process was here experimentally validated on data of a research vehicle driving past traffic cameras.

3.2 | Random sample consensus for filtering the calibration data

Outlier coordinate pairs can be found in the calibration point correspondences due to problems in GNSS localisation or incorrect timestamps in the matched ground plane and image coordinates. Outliers in the calibration data can have a tremendous effect on the outcome of the homography calibration due to the least squares approach. In order to reduce the effects of outliers in the coordinate pairs utilised for calibration, Random Sample Consensus (RANSAC) [39] is additionally utilised in the calibration procedure presented here. The calibration process with RANSAC can be divided to the following separate steps.

1. A minimal random subset of four point correspondences are selected from the full set, called the initial inliers.
2. The homography is fitted to these initial inliers.
3. Ground plane estimates for all pixel coordinates are generated with the homography matrix, and those with an error beneath a chosen threshold are considered inliers.
4. The homography matrix is refitted to all of the inlier point correspondences.

In the experiments presented here, the RANSAC procedure was executed a total of thousand iterations, the chosen homography matrix being the one with the highest number of inliers. The inlier threshold was set at 3 m. The number of iterations and the inlier threshold value were chosen keeping in mind the total number of points correspondences and that GNSS localisation even with RTK corrections can occasionally include notable error [28–30]. The presented values should be generally applicable to any camera installations, although number of iterations could be increased if notably more point correspondences are available.

3.3 | Geodetic GNSS coordinates to ground plane

In order to transform geodetic vehicle GNSS coordinates to the two-dimensional ground plane, the following transformations were applied. First, the geodetic longitude, geodetic latitude, and ellipsoidal height coordinates were transformed to Earth Centred, Earth Fixed (ECEF) coordinates. This was achieved

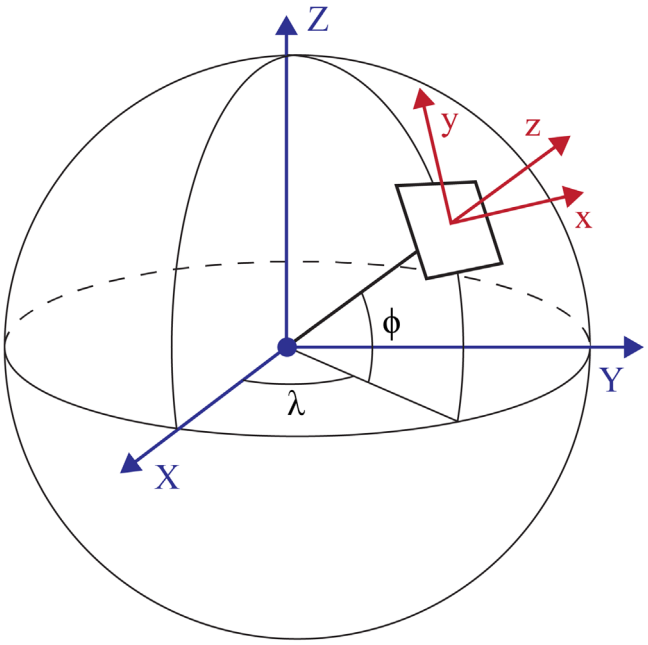

FIGURE 2 Geographic coordinate systems related to GNSS positioning

as

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} (N+h)\cos\phi\cos\lambda \\ (N+h)\cos\phi\sin\lambda \\ \left(\frac{b^2}{a^2}N+h\right)\sin\phi \end{bmatrix}, \tag{9}$$

where

$$N = \frac{a^2}{\sqrt{a^2\cos^2\phi + b^2\sin^2\phi}}. \tag{10}$$

Geodetic longitude and latitude are denoted as λ and ϕ, respectively. Ellipsoidal height is denoted by h, and a denotes equatorial radius, whereas b denotes polar radius. The acquired ECEF coordinates X, Y, and Z are further transformed to East, North, Up (ENU) coordinates as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\sin\lambda_o & \cos\lambda_o & 0 \\ -\sin\phi_o\cos\lambda_o & -\sin\phi_o\sin\lambda_o & \cos\phi_o \\ \cos\phi_o\cos\lambda_o & \cos\phi_o\sin\lambda_o & \sin\phi_o \end{bmatrix} \begin{bmatrix} X - X_o \\ Y - Y_o \\ Z - Z_o \end{bmatrix}, \tag{11}$$

where x, y, and z denoted the ENU coordinates. X_o, Y_o, and Z_o represent a chosen origin in the ECEF coordinate system, which was here chosen arbitrarily to be one of the measurement points in each experiment. The chosen origin defines the location of the ENU tangent plane, which here represents the ground plane. In order to place each calibration point on the ground plane, the z-coordinate is set to zero for all points. The different coordinate systems are presented graphically in Figure 2. It is noteworthy that the ground plane coordinates can be transformed back to ECEF and geodetic coordinates, depending on the coordinate system requirements of the particular ITS application.

3.4 | Experiment setup

To validate the calibration approach presented here, three roadside cameras were calibrated in the Helsinki metropolitan region by driving a research vehicle equipped with a GNSS receiver past them. The cameras are here referred to as Camera 1, Camera 2, and Camera 3. Samples of the camera views are provided in Figure 3. Camera 1 had a resolution of 1024×768 pixels, and monitored a high traffic road in central Helsinki. Camera 2 had a resolution of 1280×720 pixels, overlooking an urban road entering the Aalto University campus in Espoo. Camera 3 also had a resolution of 1280×720 pixels, and it monitored a section of a busy highway in Espoo. The camera views provided three distinct scenarios for validating the localisation results acquired with the calibration. Each camera view provided a unique combination of road geometry, speed of the research vehicle, camera viewpoint, as well as quality of GNSS positioning. GNSS positioning accuracy was mostly affected by the surrounding buildings, with the area of Camera 1 having the the most notable positioning problems due to the central location.

Validation of the calibration approach was practically carried out by driving a research vehicle past each camera view several times. In the view of Camera 1, each passing was done on a separate lane of the four-lane road. Camera 2 was passed a total of five times, driving back and forth the two-lane road. A total of six different lanes, three on both sides, were used for passing Camera 3 with the research vehicle. The research vehicle was equipped with an Indagon MTT130 RTK-positioning terminal, which was utilised for acquiring the GNSS location of the vehicle. In the urban surroundings, RTK fixed solutions were available for only a limited number of measurements, and vast majority of measurements were recorded with RTK float solutions. The corresponding image coordinates of the research vehicle were acquired with the well-established tracking combination of YOLOv4 [40] and DeepSORT [41]. The centre point of the vehicle bounding box was chosen to represent the vehicle image point coordinate. For Camera 1, Camera 2, and Camera 3, a total of 43, 38, and 40 corresponding GNSS locations and camera detections were recorded, respectively. The number of point correspondences was kept modest to highlight that the approach does not depend on a great number of data samples. Acquired GNSS locations as well as visualisations of tracking the research vehicle are presented for all three camera views in Figures 4–6. GNSS locations were plotted on samples extracted from OpenStreetMap [42] for clearer illustration of the points. When observing the map figures, one should note that OpenStreetMap is not a high-accuracy map and can contain notable errors [43]. This is why many of the camera localisations and GNSS positions can seemingly appear to be located outside of the road.

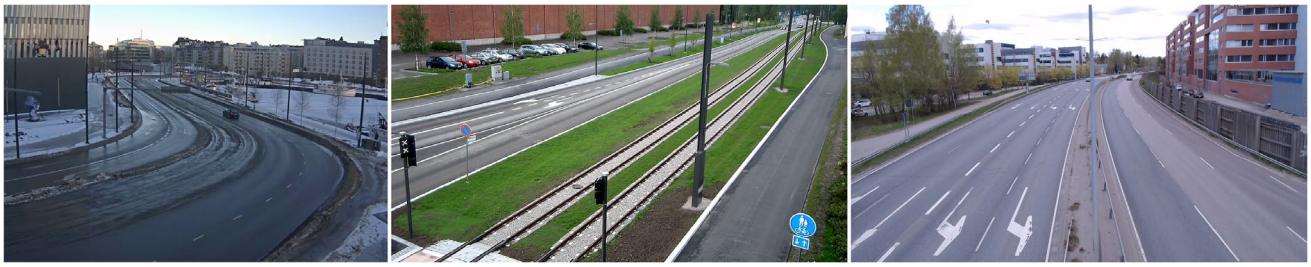

FIGURE 3 Camera views used for validating the calibration approach. Camera 1 on the left, Camera 2 in the middle, and Camera 3 on the right

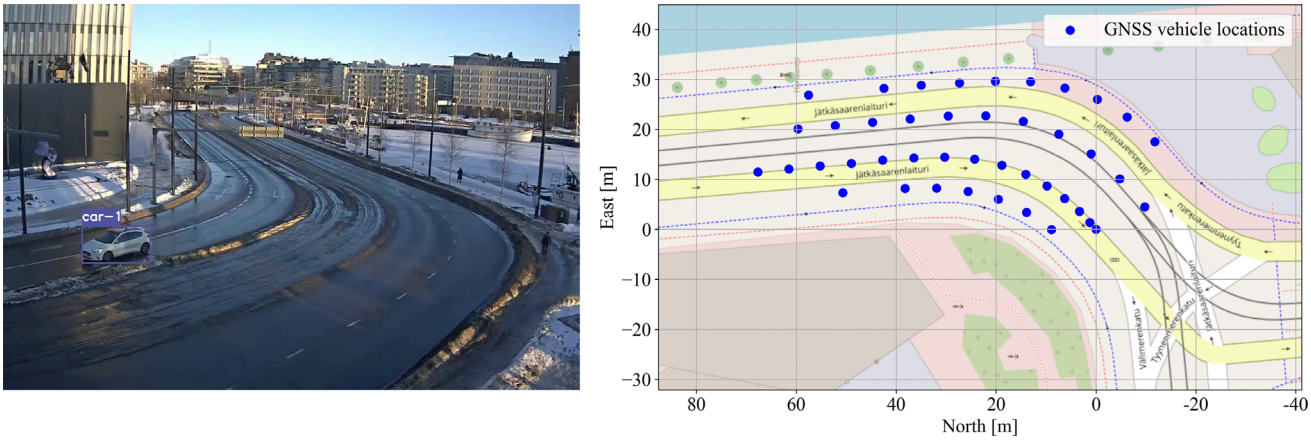

FIGURE 4 Visualisation of vehicle tracking in the view of Camera 1, and all the corresponding GNSS coordinates of multiple test runs

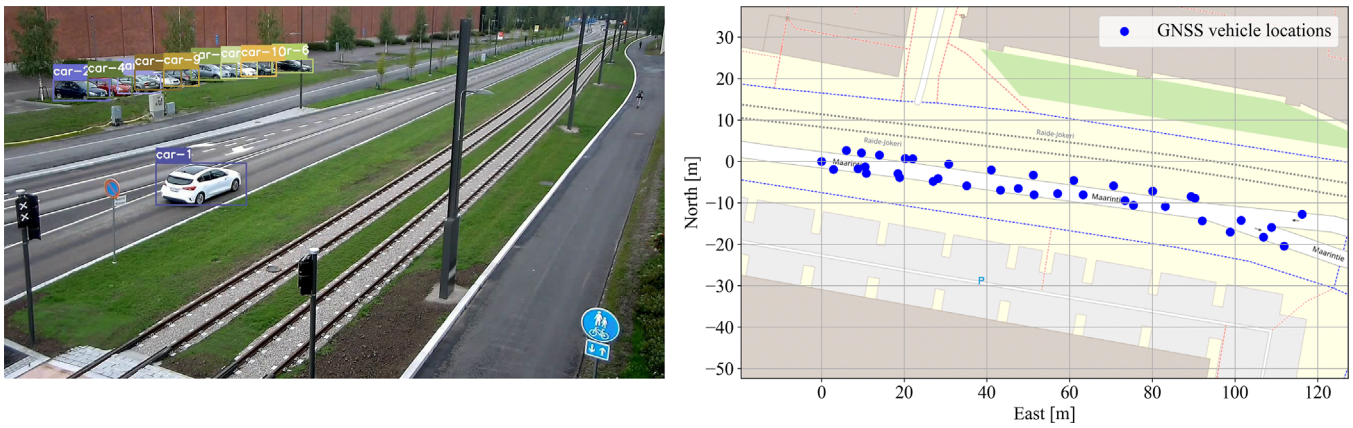

FIGURE 5 Visualisation of vehicle tracking in the view of Camera 2, and all the corresponding GNSS coordinates of multiple test runs

Since many existing traffic and surveillance cameras in cities capture video at lower resolution, the experiments were repeated with low-resolution video to ensure the calibration method can be applied on a wide variety of camera views. Low-resolution directly affects how accurately the vehicle position can be extracted from the images. Additionally, resolution impacts how well vehicles can be tracked in the images. The original video recordings were used, reducing the resolutions of Camera 1, Camera 2, and Camera 3 to 640×480, 852×480, and 852×480 pixels, respectively. Tracking the research vehicle in the low-resolution video footage of the cameras, a total of 45, 39, and 39 corresponding GNSS locations and camera detections were extracted, respectively.

3.5 | Error metric for comparison to other methods

Localisation errors are here mostly reported as the absolute distance between the camera localisation and the GNSS

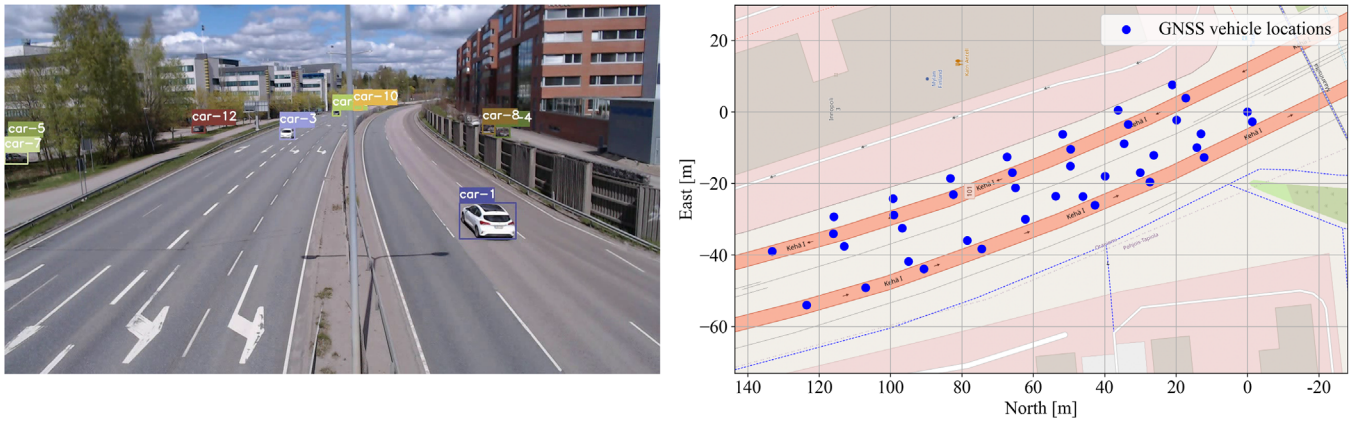

FIGURE 6 Visualisation of vehicle tracking in the view of Camera 3, and all the corresponding GNSS coordinates of multiple test runs

position. This offers a simplistic error metric for evaluating the applicability of the calibration approach to different traffic monitoring applications. However, this metric is inconvenient for comparing the achieved accuracy to previous approaches in the literature. This is due to the fact that the absolute measurement error typically grows as the measured distance grows. Expressing the errors as proportional to the measured distance yields results with better generalisability. To compare the proposed approach to previously published methods, the root-mean-square error (RMSE) metric proposed by Bhardwaj *et al.* [7] was adopted. The error metric analyses all possible pairs of validation points in the ground plane, and a normalised error is computed by comparing the real distance between the points and the distance reported by the camera. For any *i*th pair of validation points, the relative error ϵ_i^{norm} was computed as

$$\epsilon_i^{norm} = \frac{d_i^{reproj} - d_i^{real}}{d_i^{real}}, \tag{12}$$

where d_i^{reproj} is the distance between the camera localisations, and d_i^{real} is the distance between the ground truth points. Both distances lie on the ground plane. In this work, the GNSS positions were used as the ground truth points. Combining the relative errors of the point pairs of the validation sets, camera-specific RMSE values were computed as

$$RMSE = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(\epsilon_i^{norm})^2}, \tag{13}$$

where K denotes the total number of validation point pairs.

In this paper, the RMSE metric was used compare the accuracy achieved with the proposed method to the accuracy values reported in the work of Bhardwaj *et al.* [7]. Their work reported RMSE values measured for their novel AutoCalib approach, as well as the vanishing point approach proposed by Dubska *et al.* [20]. Their reported RMSE values were computed based on applying the calibration approaches in 10 different traffic camera views, and localising specific ground truth points in the ground plane.

4 | RESULTS

RTK-corrected GNSS data from the research vehicle and video data from the three cameras were utilised to validate the proposed calibration approach. Tenfold cross-validation was applied to meticulously evaluate the localisation capabilities of the calibrated cameras. The point correspondence data were divided into 10 separate subsets, of which nine subsets were utilised for calibration and the remaining subset was used for validation. All such combinations were exhaustively evaluated, using the validation sets for evaluating the error of the calibration. Each camera view was studied separately, and errors for the camera localisation were reported based on the distance between the location reported by the camera and the location acquired from the RTK-corrected GNSS receiver. Visualisations of the localisation results were generated for randomly selected subsamples of the validation processes. The visualisations were presented on OpenStreetMap, and one should again note that the maps are an approximation of the road environment. Furthermore, relative error was quantified for each camera view with the RMSE metric. These RMSE values were used to compare the proposed calibration approach to previous approaches, in addition to a general comparison of the features of the methods. Lastly, the efficacy of applying RANSAC for excluding GNSS mislocalisations was assessed, and the impact of timestamp errors on the calibration was analysed.

4.1 | Camera 1 localisation accuracy

A total of 43 corresponding GNSS locations and image points were gathered of the research vehicle with Camera 1. With 10-fold cross-validation, the localisation error was computed for each point when the respective point was in the subset used

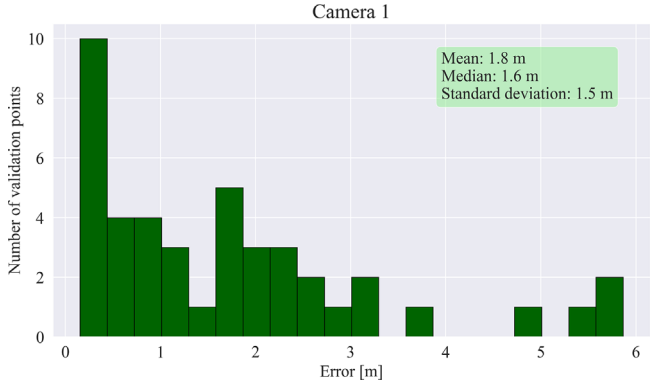

FIGURE 7 Camera localisation errors on the cross-validation data of Camera 1

for validation. A histogram and key statistics of these observed errors are presented in Figure 7.

Most of the errors can be seen in the range of 0–3 m, which is also reflected by the mean and standard deviation of the error distribution. The mode of errors can be found in the proximity of 0 m. However, the error distribution exhibits a heavy tail, with a number of outlier errors approaching the 6 m mark. These outlier errors were likely caused by GNSS inaccuracy, as the position estimates from the camera were directly compared to the GNSS locations.

Presented errors can also be witnessed observing localisation samples recorded at a random step of the cross-validation process, visualised in Figure 8. The GNSS locations and corresponding camera estimates are presented for the calibration and validation points separately. Most of the camera-based location estimates closely resemble the GNSS locations in the experiments, which depicts the capabilities of the camera-based localisation. The four lanes of the road are evidently distinguishable from the localisation results. However, some camera estimates can be seen clearly diverging from the GNSS locations near the tramway track, and these represent the outlier errors of Camera 1 previously presented in Figure 7.

Performing the same cross-validation on the low-resolution footage of Camera 1 did not drastically alter the results. With the reduced resolution, the localisation results showed a mean error of 2.0 m, median error of 1.5 m, and standard deviation of 1.6 m. Observing the original higher resolution validation results in Figure 7, the average error increased, yet the median error decreased. Therefore, the spread of the error slightly increased as the resolution was altered, which can also be witnessed by the increased standard deviation.

4.2 | Camera 2 localisation accuracy

Testing the calibration approach on Camera 2, a total of 38 corresponding GNSS locations and image points were utilised. Similar validation methodology was applied as with Camera 1, and the errors from the 10-fold cross-validation are presented in Figure 9.

Errors for Camera 2 also mostly ranged from 0 to 3 m. The error distribution resembles a bell curve, well represented by the mean and standard deviation. Mode of the distribution can be seen located at the mean. Compared to the error distribution of Camera 1, notably fewer errors were found in proximity of the 0 m mark. However, the distribution has barely any tail, except a single outlier at over 5 m of error. This outlier was likely again caused by GNSS measurement inconsistency. Localisations made by the camera system on a random cross-validation iteration are provided along the GNSS locations in Figure 10. Separate figures are presented for the calibration and validation points.

Camera estimates can be seen closely matching to the GNSS locations for both calibration and validation points. The two lanes of the road can be clearly distinguished from the localisation results. This indicates that the camera was well calibrated to the road environment, even though two-dimensional variation of the point data was minimal in the ground plane.

Cross-validation tests were repeated with low-resolution version of the video footage. On the low-resolution data, the camera localisation netted a mean error of 1.6 m, median error of 1.5 m, and standard deviation of 0.95 m. Change in resolution had therefore minimal impact on the calibration of Camera 2, when comparing to the results in Figure 9. Only the standard deviation of the errors shifted slightly.

4.3 | Camera 3 localisation accuracy

For Camera 3, a total of 40 corresponding GNSS positions and image points were utilised for assessing the accuracy of the calibration. Resulting errors from the 10-fold cross-validation are presented in Figure 11.

The achieved accuracy was distinctly similar to that achieved with Camera 1, with the error statistics showing nearly identical values. However, the mode of Camera 3 error distribution can be found at approximately 1.5 m, and there is notably more spread in the errors. Frequent errors are visible in the 3–5 m range. These errors were mostly caused due to localising the research vehicle at extreme distances. The road section for which the point correspondences were gathered was approximately 140 m long, and therefore the vehicle was nearly vanishing into the horizon during the farthest measurements. The difficulties in long-distance localisation can be witnessed in Figure 12, which depicts localisations performed by Camera 3 on a random cross-validation iteration along the GNSS positions. Greatest errors can be witnessed among localisations provided for distant points. Despite the errors, the lanes which the vehicle followed are again clearly visible from the localisation results.

The cross-validation was carried out with the low-resolution video of Camera 3 as well. In this cross-validation, the calibrated camera performed localisation with a mean error of 1.6 m, median error of 1.4 m, and standard deviation of 1.1 m. Surprisingly, the camera performed localisation more accurately with the low-resolution data, when comparing to the values in Figure 11. As the standard deviation decreased, some

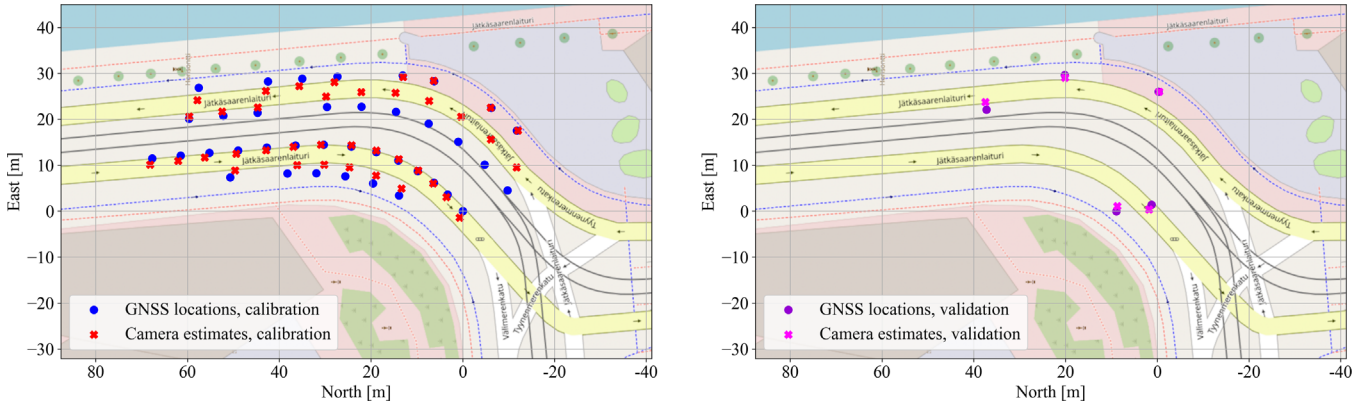

FIGURE 8 Sample from the Camera 1 cross-validation, highlighting camera location estimates on calibration as well as validation points

TABLE 1 The proposed GNSS-based calibration method compared to previously published methods

Calibration method	Mean RMSE	Min RMSE	Max RMSE	Generalisability	Prior info	Feedback loop	Coordinate system	Computational load	Connected traffic
AutoCalib [7]	8.98%	5.07%	12.3%	Fair	Required	No	Local	High	None
GNSS method	**12.0%**	**8.21%**	**18.0%**	**Excellent**	**None**	**Yes**	**Global**	**Low**	**Required**
VP method [7, 20]	21.6%	5.40%	56.4%	Good	Required	No	Local	Low	None

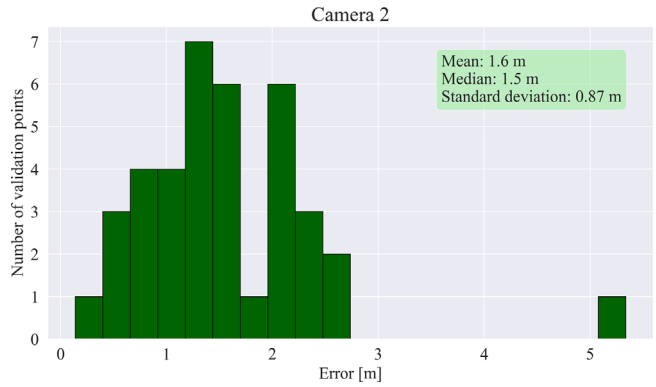

FIGURE 9 Camera localisation errors on the cross-validation data of Camera 2

of of the most extreme errors were absent. However, this was at least partially due to the fact that the low-resolution data had fewer long-range point correspondences. The research vehicle was not as successfully tracked at long ranges in the low-resolution video.

4.4 | Comparison to other calibration methods

Performing the previously described cross-validation on each of the cameras, the validation sets were utilised for evaluating relative errors comparable to the recent literature. RMSE values defined in Equation (13) were computed for each of the cameras. From the RMSE values of the cameras, the mean, minimum, and maximum are reported in Table 1, along with reference RMSE values reported for other automatic calibration methods in the literature [7]. In addition to the quantitative RMSE results, qualitative features of the calibration methods are compared in Table 1.

The proposed GNSS approach can be seen ranking slightly below the AutoCalib calibration approach in terms of the RMSE results. Compared to the commonly applied vanishing point-based calibration method (VP method), notably lower mean and maximum errors were achieved. These results highlight that the accuracy achieved with the proposed calibration approach is within the state of the art, yet not cutting-edge. However, the RMSE values do not offer an absolutely fair comparison, as the values found in the literature have been generated with a different dataset, featuring different characteristics, such as the accuracy of the ground truth measurements. Nevertheless, the provided RMSE values offer an indefinite metric for approximately comparing the accuracies of the calibration methods.

Accuracy is not the only factor that should be utilised for assessing the different methods. The qualitative features of the methods listed in Table 1 highlight the benefits of the proposed approach. The proposed GNSS-based method is generalisable to nearly any outdoor camera view, whereas AutoCalib requires a camera view with vehicles in certain poses. More specifically, AutoCalib assumes that vehicle sideview mirrors, taillights, and rear register plate are visible simultaneously in the images, as these keypoints are used for fitting the calibration. The VP method assumes that the vehicle trajectories follow straight lines. This assumption can break due to road geometry, or vehicle lane changes. Furthermore, the proposed method assumes

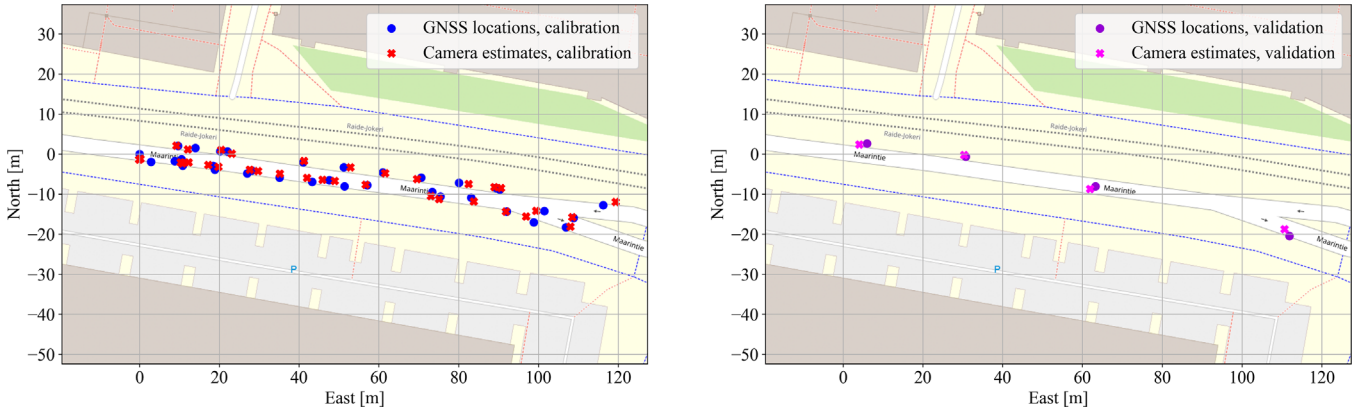

FIGURE 10 Sample from the Camera 2 cross-validation, highlighting camera location estimates on calibration as well as validation points

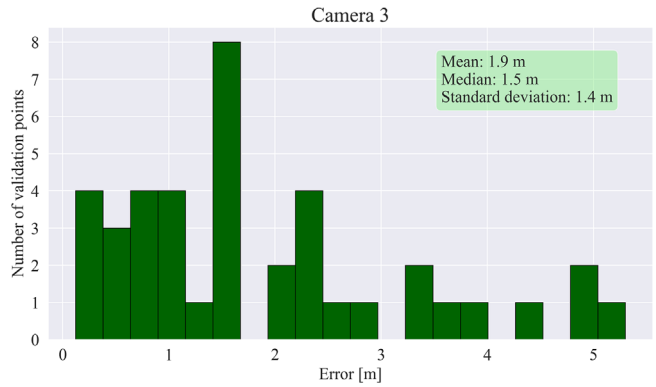

FIGURE 11 Camera localisation errors on the cross-validation data of Camera 3

no prior information other than flat ground plane, which is common to all methods. AutoCalib assumes known camera intrinsic parameters, as well as utilises a prior database of vehicle models common to the area. Similarly, the VP method utilises prior statistical knowledge of vehicle dimensions. A benefit of the GNSS-based method is also that it can utilise the GNSS coordinates as a feedback loop to assess the achieved calibration. The other methods are limited to the image data, and cannot reliably verify the result of the calibration. GNSS-based calibration additionally carries the benefit that the camera localisation is calibrated to the same global coordinate system used by other traffic systems. The other calibration methods provide localisation in an arbitrary local coordinate system, the orientation of which is unknown relative to the global frame. Lastly, the calibration process of the GNSS-based method is computationally lightweight, as only a handful of matrix operations on the point correspondence data are required each RANSAC iteration to compute the result. Similarly the VP method does not feature heavy computing, yet AutoCalib utilises a deep neural network for annotating vehicle keypoints, making its implementation computationally more demanding. The benefits of the proposed method stem from the utilisation of vehicle GNSS coordinates. However, the calibration method is reliant on connected vehicle and traffic infrastructure solutions, unlike the other calibration approaches. This limits the usage of the proposed approach in current traffic infrastructure.

4.5 | RANSAC for eliminating GNSS outliers

The ability of RANSAC to detect GNSS outliers and exclude them from the calibration was analysed with the point correspondences of Camera 1. Camera 1 was chosen as GNSS mislocalisations were clearly present, and the data points of adjacent lanes allowed for a convenient evaluation. Using all of the available point correspondences for calibration, the outliers excluded by RANSAC were labelled. These points are shown in Figure 13.

On the left in the figure, the outlier points can be seen located at a distance of roughly 10 m from the points of the adjacent lane. Considering an average lane width of 3–4 m, this indicates that the GNSS localisation has indeed performed extremely poorly in this particular area. The right side of the figure shows the camera location estimates acquired for the corresponding image points. The camera estimates can be seen ignoring the outlier GNSS points, forming the adjacent lanes with a fairly constant gap in between. Camera estimates of the outlier points are clearly shifted closer to the points of the adjacent lane. Distances between the outlier point estimates and estimates from the adjacent lane were notably more realistic, approximately 6 m. Compared to the distance of 10 m between the GNSS locations, this was a notable improvement. Therefore, it seems that RANSAC was able to correctly detect the outliers and exclude them from the calibration process. As a result, the camera localisation accuracy was not affected by the GNSS mispositioning, although the outlier points heavily affected the quantitative error results presented for Camera 1 in Figure 7.

4.6 | Sensitivity of calibration to errors in timestamps

In addition to GNSS inaccuracy, errors in image timestamps also hinder the outcome of the calibration. Matching the vehicle image coordinates to the GNSS position require the data

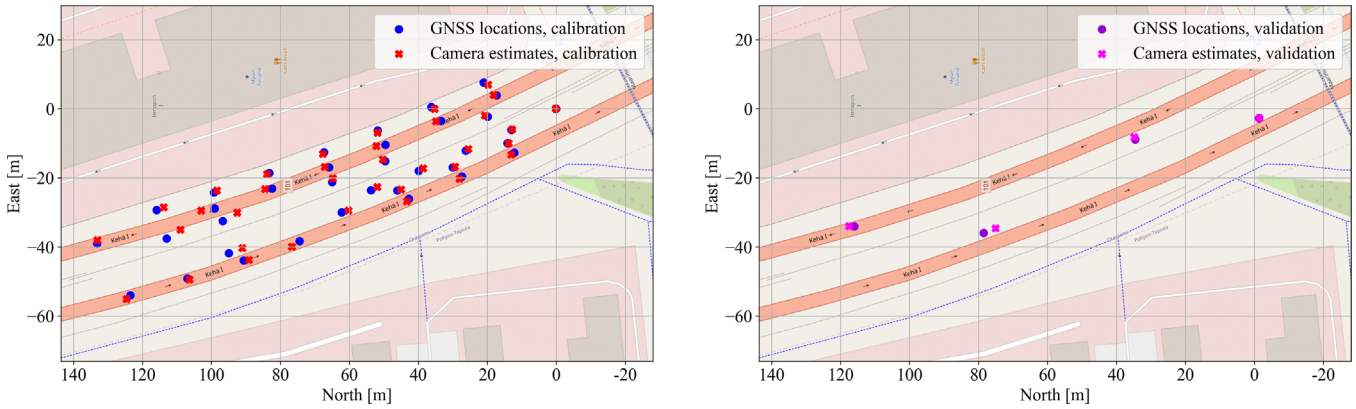

FIGURE 12 Sample from the Camera 3 cross-validation, highlighting camera location estimates on calibration as well as validation points

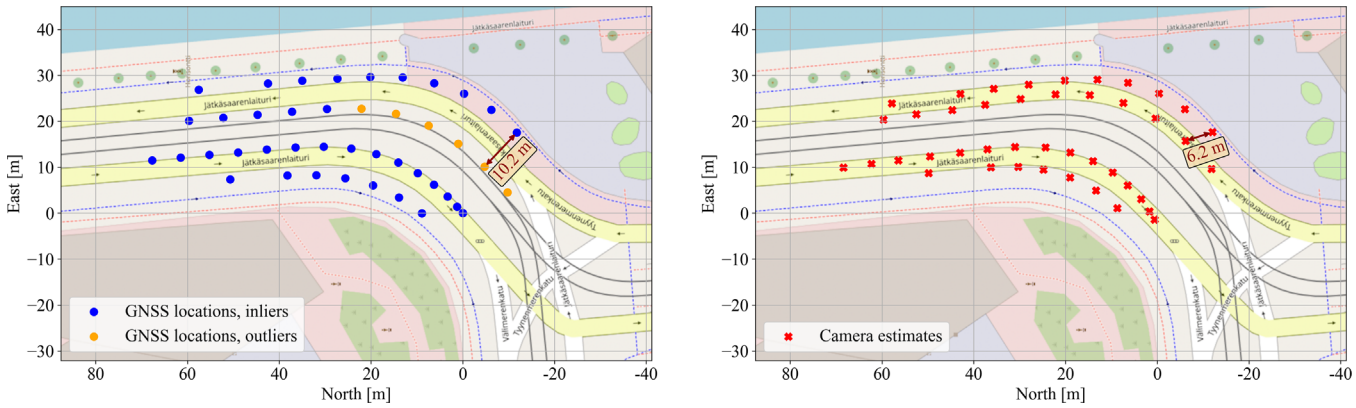

FIGURE 13 RANSAC effectively labelled unfit points as outliers. The outlier GNSS locations contained notable error, as they were at a distance of 10 m from the measurements of the adjacent lane. Due to excluding the outliers, the camera positioned the points closer to one another, roughly at a distance of 6 m

to be synced. GNSS receiver timestamps are highly accurate, yet image timestamps can include errors from multiple sources. The camera clock may not be properly synced to internet time services, or the image capturing and encoding process may be subject to delays which cause error in the timestamp. Discrepancy in the timestamps causes the vehicle to appear at a different location in the image than the one matching the position reported by the GNSS receiver. This leads to error in fitting the calibration, as the corresponding GNSS position and image coordinates of the data point do not match. Depending on the timestamp error and the vehicle speed, the vehicle may have moved to a drastically different position during the time between the acquisition of the corresponding image and GNSS position sample. In Table 2, distances moved by a vehicle during different quantities of timestamp error are presented.

Analysing the vehicle position errors presented in Table 2, error in timestamps is potentially a notable source of error in the overall accuracy of the calibration. Vehicles may move metres during the discrepancy between timestamps, which hinders the accuracy of the achieved calibration. Implementing the calibration in practical applications, timestamps provided by the cameras should be reasonably accurate. In the tests presented for the three cameras, errors in timestamps have likely

TABLE 2 Errors in observed vehicle position caused by timestamp errors at different vehicle speeds. The position errors have been colour-coded: green < 1 m, 1 m $\leq$ yellow ≤ 2 m, red > 2 m

		Vehicle speed			
		30 km/h	40 km/h	60 km/h	70 km/h
Timestamp error	30 ms	0.25 m	0.33 m	0.50 m	0.58 m
	50 ms	0.42 m	0.56 m	0.83 m	0.97 m
	100 ms	0.83 m	1.1 m	1.7 m	1.9 m
	200 ms	1.7 m	2.2 m	3.3 m	3.9 m

contributed to the overall localisation error. As the localisation errors on average ranged between 1.5 and 2.0 m, it seems probable that part of the error has been caused by slightly incorrect timestamps. However, on average the timestamps should have been fairly accurate in the tests, as the presented tests featured road sections with significantly different driving speeds, yet the average error remained relatively constant. Some outlier data correspondences in the tests could have been caused by occasional delay in image timestamps. In practical use, RANSAC should eliminate the outlier data correspondences caused by random timestamp errors, provided that the errors are not prevalent in the calibration data.

5 | DISCUSSION

5.1 | Accuracy of the calibration approach

Presented results demonstrate that the proposed calibration approach can be conveniently applied to accomplish reliable vehicle localisation from roadside surveillance cameras. The calibrated camera systems were capable of accurate localisation, considering the tests were carried out on roughly a 100 m long road sections. Despite monitoring long road sections, the average localisation errors of Camera 1, Camera 2, and Camera 3 were 1.8, 1.6, and 1.9 m, respectively. Evaluating the localisation accuracy on the relative RMSE metric presented in Equation (13), the error was similar to that achieved in the recent literature. Observing Table 1, the proposed approach can be seen reaching a mean RMSE of 12.0%, whereas the recently published AutoCalib was reported to reach RMSE of 8.98% in its respective paper [7]. Since the RMSE results have been extracted from different datasets, comparison between them is limited and only approximate. Nevertheless, their localisation errors can be considered rather similar.

The witnessed errors of the proposed calibration should be acceptable for many ITS applications, especially since the lane of the vehicle could be quite clearly determined from the localisation results. This is evident in the figures highlighting the camera localisation results on the map, especially in Figures 8 and 13. In these figures, the four lanes the vehicle has driven can be clearly defined from the camera localisations. The linear localisation procedure ensures that the error is more of a constant offset in certain regions, instead of highly varying random noise. This leads to the errors being highly predictable, and lanes should be conveniently extractable from the localisations with simple map-matching and pattern analysis algorithms. This type of position information combined with scene understanding is invaluable for many traffic control, monitoring, as well as safety applications. Such applications include traffic light control, road planning, and red-light violation warnings.

5.2 | Uncertainty in the GNSS positioning

In the presented results, there remains uncertainty regarding the exact accuracy of camera-based localisation, as the corresponding GNSS locations were used as reference for computing the errors. As mentioned earlier, even RTK-corrected GNSS can include outliers with metres of error when line of sight to the satellites is blocked. This naturally degrades the reliability of the accuracy measurements, as there is uncertainty regarding the ground truths.

Due to the occasional outliers in GNSS positioning, the RANSAC filtering was added to the calibration approach to negate the effects of anomalous positioning data. As seen in Figure 13, RANSAC effectively allowed the calibration process to ignore the outliers when fitting the homography. The camera can be seen localising the vehicle on clearly separate lanes, that are notably more realistic than the ones acquired from GNSS. However, ignoring these outlier points in the calibration caused the error histograms to include notable outlier errors, which in turn affected the error statistics. These outlier points ignored by RANSAC formed the tail of the error distribution of Camera 1 in Figure 7, which contained individual errors from 3 to 6 m. Camera 2 error distribution in Figure 9 had less outliers, only a single error of approximately 5 m. Similarly, Camera 3 did not apparently experience as many GNSS outliers. The highest errors of Camera 3 occurred at extreme distances from the camera, and therefore the distance was a more plausible cause for the errors. The GNSS positioning likely functioned more accurately in the views of Cameras 2 and 3 due to the better visibility of the sky, as well as the more predictable and straight trajectory of the vehicle. Camera 1 was located in central Helsinki, surrounded by tall buildings in immediate vicinity on multiple sides.

5.3 | Benefits and drawbacks of the calibration approach

The proposed GNSS-based automated calibration approach is applicable in a wide range of traffic environments, as minimal assumptions regarding the camera view are made. The only assumption in the proposed method is the model of the flat ground plane. In practice, the approach can be applied in nearly any outdoor environment. Previously proposed approaches in the literature have depended on different assumptions regarding the traffic scene, such as straight roads, vehicles moving in straight lines, known number of lanes, known lane width, known vehicle dimensions, or certain viewpoint of the camera. For example, the AutoCalib [7] approach used for comparison here assumed known intrinsic parameters of the camera, known statistics of vehicle dimensions, and such a camera viewpoint that the images contained vehicles in certain poses. This was necessary so that specific keypoints could be extracted from the vehicles. Such assumptions have naturally been necessary as the previous approaches have only utilised the information available in the images to perform the calibration. The proposed calibration approach takes advantage of an external data source, allowing for more reliable calibration. GNSS vehicle locations utilised for point correspondences in the images also allow for convenient validation of the calibration in automated operation. Specifically, the GNSS coordinates allow a feedback loop between camera estimates and actual locations. Purely image-based calibration approaches cannot verify the calibration from image data in any concrete way, reducing the reliability of the camera localisation.

Although widely applicable, GNSS-based calibration can have its drawbacks in some highly specific environments. Areas that are underground or inside naturally have extremely unreliable GNSS localisation. Therefore, cameras overlooking such areas cannot be calibrated with the proposed approach. Urban canyons can also hinder the GNSS positioning accuracy, as witnessed in the presented measurements. If a notable portion of outlier localisations are present in the calibration point correspondences, the calibration may be unreliable. However, in future traffic camera applications the appearance of connected vehicles in the camera view should be a somewhat regular occurrence. Continuous data acquisition from these connected

vehicles offers an increased number of point correspondences, which allows for increasingly reliable outlier elimination with RANSAC and a more stable optimal solution for the least-squares optimisation. This offers the possibility for updating the camera calibration, continuously enhancing the localisation accuracy and reliability of the camera system.

Another factor to note regarding applicability of the proposed approach is that the GNSS coordinates are transformed to the ENU ground plane. This method does not consider local changes in altitude, and localisation accuracy may consequently be limited in areas with significant local inclinations. This problem is common to all traffic camera localisation approaches, as a flat ground plane assumption is generally made in all single-camera localisation approaches. The proposed approach, however, also models the ground plane to be tangent to the earth ellipsoid. This should not pose any challenges in practice, as the inclination would have to be drastic to majorly affect the accuracy. Considering the steepness limitations for vehicle-operated roads, inclination should not cause notable errors in localisation.

5.4 | Future improvements to the calibration approach

In the presented tests, the acquired errors of the camera localisation were similar to the error in the GNSS positioning. This indicates that the accuracy of the achieved calibration is mostly limited by the accuracy of the GNSS. Calibration reliability and accuracy could further be improved by fusing the vehicle GNSS positioning with inertial measurements. Fusing inertial measurements with GNSS would effectively remove outliers caused by temporary problems with GNSS signal quality. Such fusion would likely enable usage of regular uncorrected GNSS positioning in the calibration without notable drawbacks in accuracy. RTK-corrected GNSS is a rare feature in vehicles, whereas regular GNSS fused with inertial measurements is common and cost-efficient technology available in most modern vehicles. Managing to carry out the calibration with existing on-board positioning equipment is naturally crucial for the adoption of the presented calibration technology.

Additionally, localisation and calibration reliability of the proposed approach could be improved with more accurate detection models of the vehicles in the images. Here, only two-dimensional bounding boxes were generated for the vehicles, and the centre point of the bounding box was used to reduce the vehicle to a point coordinate. This is clearly not an optimal approach, as the centre point of the two-dimensional vehicle bounding box depends on the angle in which the vehicle is observed, skewing the localisation results. More comprehensive algorithms capturing the vehicle three-dimensional bounding box should be applied to reach greater accuracy and improved reliability. The three-dimensional bounding box would provide more accurate pixel coordinates corresponding to the vehicle location on the ground plane. Nonrepresentative bounding boxes were undoubtedly a source of error in the presented results.

The results presented in this paper highlight that the camera localisation suffers from varying errors, yet the causes of these errors are not exhaustively analysed. These errors and the factors impacting them should be analysed more in-depth to further optimise the camera localisation capabilities. Such error factors include GNSS positioning, discrepancy in calibration data timestamps, vehicle bounding box, vehicle speed, camera pose, and camera resolution. Each of these factors has an effect on the accuracy of the point correspondence data used for the calibration process. Their impact on the calibration error should be analysed independently, finding the key attributes which might cause the calibration to fail. This detailed error analysis is left for future research on the topic.

Future research should also focus on fusing the proposed GNSS-based calibration approach with previously presented calibration methodologies. This would enable maximal exploitation of the benefits of different calibration methods. Since the other calibration methodologies apply vastly different processing techniques, their errors are likely not strongly correlated with the errors of the proposed approach. Fusion of the calibration approaches could be achieved in a number of ways, machine learning lately being a popular choice for fusion of perception technologies. Neural network-based approaches [44] or reinforcement learning-based approaches [45] could be applied to fuse the localisation results achieved with different automated calibration algorithms.

Another detail left for future work is the procedure of matching the specific vehicle in the images that has provided their GNSS coordinates. Here, the correct vehicle ID was selected by hand after recording the tracking results. In real automated calibration applications, the correct vehicle has to be automatically found in the images for which the known vehicle coordinates are acquired. This was left out of the scope of this paper, as the most convenient approach for this problem will depend on the information available in the connected vehicle message formats. Simple path matching and statistical methods should provide a suitable solution to the problem, yet additional information likely available in connected vehicle messages, such as vehicle colour and type, can notably improve the reliability of the matching.

5.5 | Camera localisation in connected vehicle environment

The presented calibration approach relies on vehicle GNSS coordinates being available on the infrastructure side, which is not the case in current traffic infrastructure. Therefore, implementation of the calibration approach can only be carried out once connected vehicles sharing their GNSS coordinates with the infrastructure are available on the roads. The GNSS timestamps must also be synced with the camera timestamps, yet this should not pose challenges in a connected vehicle environment with access to the internet. As for the future connectivity of vehicles, one could argue that vehicle communications will defeat the purpose of the roadside camera calibration and localisation technology presented here. Especially considering that the presented calibration approach is reliant on at least a partial

adoption of connected vehicle technology. In order to carry out the calibration in an automated manner, connected vehicles sharing their GNSS locations must drive past the cameras. If all vehicles shared their positions in real-time over a network, all external measurement approaches would naturally be redundant to some extent. However, reaching such scenario where all vehicles are connected will take a considerable amount of time.

Vehicle manufacturers and traffic authorities have had a long on-going debate regarding the implementation details of vehicle communications, and no definitive solution has yet been created. Meanwhile, an increasing number of modern vehicles include manufacturer-specific communication technologies. This indicates that while the industry is shifting towards connected vehicle technology, a great deal of effort is still required to reach unified vehicle communication networks. Furthermore, even if vehicle communications were unanimously standardised, adopted, and mandatory on modern vehicles, not all vehicles in traffic would be connected for a substantial time period. This is due to vehicles having a notably long lifespan, as most vehicles are in use for over a decade. Due to these reasons, for the near future a time period of partial connected vehicle environment seems probable. The calibration and localisation technology presented here is designed with this partial connectivity in mind, as the calibration can be performed in an automated manner if even some connected vehicles drive past the camera. As an example, if the maintenance vehicles of a city were equipped with connected technology, they would effectively allow convenient calibration of the cameras in the city while the vehicles were conducting their routine business. With the calibrated roadside cameras, a number of ITS applications requiring vehicle positions can be implemented without facing the uncertainty of unaccounted vehicles with no connected technology.

6 | CONCLUSION

An automated roadside camera calibration approach for localising vehicles was proposed in this paper. Future intelligent transportation systems rely on rich real-time information of traffic, and utilising existing traffic camera infrastructure offers a convenient way to acquire such information. The calibration approach is based on receiving GNSS coordinates from connected vehicles visible in the camera view. These GNSS coordinates are utilised to calibrate a homography, which can be utilised to transform image coordinates of vehicles to the respective ground plane locations. Distinct advantages of the presented approach are that it can be applied in practically any outdoor environment, and the GNSS coordinates of vehicles can be utilised as a feedback loop for the calibration, validating that the calibration has been successful. The measurements presented in this paper highlight that the presented calibration offers accurate vehicle localisation, although outliers in GNSS measurements complicated the exact quantification of the errors.

Future work on the calibration approach should aim to evaluate different methods for conveniently matching the vehicles in the camera view to the received GNSS-coordinates. This is crucial for adoption of the algorithm, as the calibration data points must be acquired reliably to prevent problems in the calibration. Furthermore, the calibration method should be validated with vehicle localisations performed via fusion of stand-alone GNSS and inertial measurements. This combination of sensors is already equipped on most modern vehicles, and the inertial measurements should fix most short-comings of GNSS in urban environments. Additional statistical methods could also be included to more reliably filter GNSS outliers.

Overall, the presented camera calibration approach can have a notable impact on how traffic is monitored and controlled. Adoption of real-time camera-based localisation technology in roadside cameras can generate an immense amount of data, which can solve traffic problems with more efficient control, design, and safety features. Infrastructure-based localisation offers advantages and robustness in partially connected vehicle environments, as fully connected vehicle fleets are still fairly distant future.

ACKNOWLEDGEMENTS

This work was supported by Henry Ford Foundation Finland and Academy of Finland. Code demonstrating the methodology of this work is publicly available in the following repository: https://github.com/ojalar/gnss-camera-calibration

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Risto Ojala https://orcid.org/0000-0003-0865-1775

REFERENCES

1. One billion surveillance cameras will be watching around the world in 2021, a new study says. https://www.cnbc.com/2019/12/06/one-billion-surveillance-cameras-will-be-watching-globally-in-2021.html (2019). Accessed 11 May 2021
2. Atev, S., Masoud, O., Janardan, R., Papanikolopoulos, N.: A collision prediction system for traffic intersections. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 169–174. IEEE, Piscataway, NJ (2005)
3. Zhou, Y., Liu, L., Shao, L., Mellor, M.: Dave: A unified framework for fast vehicle detection and annotation. In: European Conference on Computer Vision, pp. 278–293. Springer, Berlin (2016)
4. Cai, Y., Wang, H., Chen, X., Jiang, H.: Trajectory-based anomalous behaviour detection for intelligent traffic surveillance. IET Intel. Transport Syst. 9(8), 810–816 (2015)
5. Ojala, R., Vepsäläinen, J., Hanhirova, J., Hirvisalo, V., Tammi, K.: Novel convolutional neural network-based roadside unit for accurate pedestrian localisation. IEEE Trans. Intell. Transp. Syst. 21, 3756–3765 (2019)
6. Average age of the EU vehicle fleet, by country, The European Automobile Manufacturers' Association. https://www.acea.auto/figure/average-age-of-eu-vehicle-fleet-by-country/ (2021). Accessed 30 August 2021
7. Bhardwaj, R., Tummala, G.K., Ramalingam, G., Ramjee, R., Sinha, P.: Autocalib: Automatic traffic camera calibration at scale. ACM Trans. Sens. Netw. (TOSN) 14(3-4), 1–27 (2018)

8. Sochor, J., Juránek, R., Španhel, J., Maršık, L., Širokỳ, A., Herout, A., Zemcık, P.: Brnocompspeed: Review of traffic camera calibration and comprehensive dataset for monocular speed measurement. IEEE Trans. Intell. Transp. Syst. 3(5), 6 (2017)
9. Kanhere, N.K., Birchfield, S.T.: A taxonomy and analysis of camera calibration methods for traffic monitoring applications. IEEE Trans. Intell. Transp. Syst. 11(2), 441–452 (2010)
10. Li, Z., Li, S., Luo, X.: An overview of calibration technology of industrial robots. IEEE/CAA J. Autom. Sin. 8(1), 23–36 (2021)
11. Guindel, C., Beltrán, J., Martín, D., García, F.: Automatic extrinsic calibration for lidar-stereo vehicle sensor setups. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6. IEEE, Piscataway, NJ (2017)
12. Abdel-Aziz, Y., Karara, H., Hauck, M.: Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. Photogramm. Eng. Remote Sens. 81(2), 103–107 (2015)
13. Weng, J., Cohen, P., Herniou, M., et al.: Camera calibration with distortion models and accuracy evaluation. IEEE Trans. Pattern Anal. Mach. Intell. 14(10), 965–980 (1992)
14. Caprile, B., Torre, V.: Using vanishing points for camera calibration. Int. J. Comput. Vision 4(2), 127–139 (1990)
15. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22(11), 1330–1334 (2000)
16. E Bas, K., Crisman, J.D.: An easy to install camera calibration for traffic monitoring. In: Proceedings of Conference on Intelligent Transportation Systems, pp. 362–366. IEEE, Piscataway, NJ (1997)
17. Schoepflin, T.N., Dailey, D.J.: Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. IEEE Trans. Intell. Transp. Syst. 4(2), 90–98 (2003)
18. Zheng, Y., Peng, S.: A practical roadside camera calibration method based on least squares optimization. IEEE Trans. Intell. Transp. Syst. 15(2), 831–843 (2013)
19. Dawson, D.N., Birchfield, S.T.: An energy minimization approach to automatic traffic camera calibration. IEEE Trans. Intell. Transp. Syst. 14(3), 1095–1108 (2013)
20. Dubská, M., Herout, A., Juránek, R., Sochor, J.: Fully automatic roadside camera calibration for traffic surveillance. IEEE Trans. Intell. Transp. Syst. 16(3), 1162–1171 (2014)
21. Sochor, J., Juránek, R., Herout, A.: Traffic surveillance camera calibration by 3d model bounding box alignment for accurate vehicle speed measurement. Comput. Vision Image Understanding 161, 87–98 (2017)
22. Ismail, K., Sayed, T., Saunier, N.: A methodology for precise camera calibration for data collection applications in urban traffic scenes. Can. J. Civ. Eng. 40(1), 57–67 (2013)
23. Do, V.-H., Nghiem, L.-H., Thi, N.P., Ngoc, N.P.: A simple camera calibration method for vehicle velocity estimation. In: 2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 1–5. IEEE, Piscataway, NJ (2015)
24. Global Positioning System (GPS) Standard Positioning Service (SPS) Performance Analysis Report, Federal Aviation Administration. https://www.nstb.tc.faa.gov/reports/PAN96_0117.pdf#page=22 (2017). Accessed 30 July 2020
25. Merry, K., Bettinger, P.: Smartphone GPS accuracy study in an urban environment. PloS One 14(7), e0219890 (2019)
26. Wing, M.G., Eklund, A., Kellogg, L.D.: Consumer-grade global positioning system (GPS) accuracy and reliability. J. Forest. 103(4), 169–173 (2005)
27. Aponte, J., Meng, X., Hill, C., Moore, T., Burbidge, M., Dodson, A.: Quality assessment of a network-based RTK GPS service in the UK. J. Appl. Geo. 3(1), 25–34 (2009)
28. Lee, I.-S., Ge, L.: The performance of RTK-GPD for surveying under challenging environmental conditions. Earth Planets Space 58(5), 515–522 (2006)
29. Ong, R.B., Petovello, M.G., Lachapelle, G.: Assessment of GPS/glonass RTK under various operational conditions. In: Proceedings of the 22nd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2009), pp. 3297–3308. (2009)
30. Sheng, C., Fan, G., Yu, B., Gan, X., Zhang, J.: Optimized algorithm for bds/gps rtk suitable for urban canyons with a low-cost receiver. Meas. Sci. Technol. 31(11), 115007 (2020)
31. Chu, T., Guo, N., Backén, S., Akos, D.: Monocular camera/imu/gnss integration for ground vehicle navigation in challenging GNSS environments. Sensors 12(3), 3162–3185 (2012)
32. Tseng, W.-K., Schoellig, A.P., Barfoot, T.D.: Self-calibration of the offset between GPS and semantic map frames for robust localization. In: 2021 18th Conference on Robots and Vision (CRV), pp. 173–180. IEEE, Piscataway, NJ (2021)
33. Liao, H.-C., Wu, H.-J.: An automatic camera calibration method using gps-enabled mobile device. In: 2009 11th International Conference on Advanced Communication Technology, vol. 1, pp. 760–763. IEEE, Piscataway, NJ (2009)
34. Tsai, R.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. IEEE J. Robot. Autom. 3(4), 323–344 (1987)
35. Liu, H., Wang, C., Lu, H., Yang, W.: Outdoor camera calibration method for a GPS & camera based surveillance system. In: 2010 IEEE International Conference on Industrial Technology, pp. 263–267. IEEE, Piscataway, NJ (2010)
36. Jiang, H., Sun, X.: A camera calibration method based on differential GPS system for large field measurement. In: International Conference on Intelligent Robotics and Applications, pp. 508–519. Springer, Berlin (2017)
37. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer Science & Business Media, Berlin (2010)
38. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, New York, NY (2003)
39. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24(6), 381–395 (1981)
40. Bochkovskiy, A., Wang, C.-Y., H-Liao, Y.M.: Yolov4: Optimal speed and accuracy of object detection. Preprint, arXiv:2004.10934 (2020). https://doi.org/10.48550/arxiv.2004.10934
41. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 3645–3649. IEEE, Piscataway, NJ (2017)
42. Haklay, M., Weber, P.: Openstreetmap: User-generated street maps. IEEE Pervasive Comput. 7(4), 12–18 (2008)
43. Haklay, M.: How good is volunteered geographical information? A comparative study of openstreetmap and ordnance survey datasets. Environm. Planning B 37(4), 682–703 (2010)
44. Zhang, L., Han, H., Zhou, M., Al-Turki, Y., Abusorrah, A.: An improved discriminative model prediction approach to real-time tracking of objects with camera as sensors. IEEE Sens. J. 21(15), 17 308–17 317 (2021)
45. Zhou, T., Chen, M., Zou, J.: Reinforcement learning based data fusion method for multi-sensors. IEEE/CAA J. Autom. Sin. 7(6), 1489–1497 (2020)

How to cite this article: Ojala, R., Vepsäläinen, J., Pirhonen, J., Tammi, K.: Infrastructure camera calibration with GNSS for vehicle localisation. IET Intell. Transp. Syst. 17, 341–356 (2023). https://doi.org/10.1049/itr2.12262