BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF INFORMATION TECHNOLOGY
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

VEHICLE SPEED MEASUREMENT USING STEREO CAMERA PAIR

MĚŘENÍ RYCHLOSTI VOZIDEL POMOCÍ STEREO KAMERY

DOCTORAL THESIS
DISERTAČNÍ PRÁCE

AUTHOR   Ing. PAVEL NAJMAN
AUTOR PRÁCE

SUPERVISOR   prof. Dr. Ing. PAVEL ZEMČÍK
ŠKOLITEL

BRNO 2020

Abstract

This thesis aims to answer the question whether it is currently possible to autonomously measure the speed of vehicles using a stereoscopic method with the average error within ± 1 km/h, the maximum error within ± 3 km/h, and the standard deviation within ± 1 km/h. The error ranges are based on the requirements of the OIML whose Recommendations serve as templates for metrological legislations of many countries. To answer this question, a hypothesis is formulated and tested. A method that utilizes a stereo camera pair for vehicle speed measurement is proposed and experimentally evaluated. The experiments show that the technique overcomes state-of-the-art results with the mean error of approximately 0.05 km/h, the standard deviation of less than 0.20 km/h, and the maximum absolute error of less than 0.75 km/h. The results are within the required ranges, and therefore the formulated hypothesis holds.

Abstrakt

Tato práce se snaží najít odpověď na otázku, zda je v současnosti možné autonomně měřit rychlost vozidel pomocí stereoskopické měřící metody s průměrnou chybou v rozmezí ± 1 km/h, maximální chybou v rozmezí ± 3 km/h a směrodatnou odchylkou v rozmezí ± 1 km/h. Tyto rozsahy chyb jsou založené na požadavcích organizace OIML, jejichž doporučení jsou základem metrologických legislativ mnoha zemí. Pro zodpovězení této otázky je zformulována hypotéza, která je následně testována. Metoda, která využívá stereo kameru pro měření rychlosti vozidel je navržena a experimentálně vyhodnocena. Výsledky pokusů ukazují, že navržená metoda překonává výsledky dosavadních metod. Průměrná chyba měření je přibližně 0.05 km/h, směrodatná odchylka chyby je menší než 0.20 km/h a maximální absolutní hodnota chyby je menší než 0.75 km/h. Tyto výsledky jsou v požadovaném rozmezí a potvrzují tedy testovanou hypotézu.

Keywords

Vehicle speed measurement, traffic surveillance, stereo vision, stereo camera calibration, vehicle detection, vehicle tracking, sub-pixel registration.

Klíčová slova

Měření rychlosti vozidel, dopravní dohled, stereo vidění, kalibrace stereo kamery, detekce vozidel, sledování vozidel, sub-pixelová registrace

Reference

NAJMAN, Pavel. *Vehicle Speed Measurement Using Stereo Camera Pair*. Brno, 2020. Doctoral thesis. Brno University of Technology, Faculty of Information Technology. Supervisor prof. Dr. Ing. Pavel Zemčík

Vehicle Speed Measurement Using Stereo Camera Pair

Declaration

I hereby declare that this Ph.D. thesis was prepared as an original work by the author under the supervision of prof. Dr. Ing. Pavel Zemčík. I have listed all the literary sources, publications, and other sources which were used during the preparation of this thesis.

. . . .

Pavel Najman

August 31, 2020

Acknowledgements

Contents

Chapter 1

Introduction

The speed of a moving vehicle directly influences both the risk of a crash and the severity of its consequences. To minimize this risk and thus to increase the road traffic safety, the law imposes speed limits that should ensure that in the event of a crash, impact energies remain below the threshold likely to produce either death or serious injury.

The threshold value usually depends on the most probable crash scenario which varies with road location. In residential and high pedestrian traffic areas, it is often around 30 km/h. In cities or areas with a higher probability of side impact of vehicles and with a lower amount of pedestrian traffic, it is around 50 km/h. And for highways, where rear-end collisions are prevalent it is around 130 km/h [35, 55, 74].

In order for the speed limits to be effective in increasing road safety, they need to be enforced. The success of the enforcement depends on the ability of empowered authorities to accurately measure the speed of passing vehicles. While various measurement devices exist, only those that comply with metrological legislation implemented in a given country can be used for the enforcement.

The metrological legislation of individual countries defines the requirements in terms of working conditions, measurement range, precision, and accuracy which every speed measurement device has to meet. Before the device is approved and certified, it typically undergoes laboratory and field tests which verify that the device is able to provide reliable measurements within given tolerances. For countries that base their metrological legislation on recommendations of the International Organization of Legal Metrology, the maximum allowable errors for laboratory tests are ± 1 km/h for reference speeds up to 100 km/h and ± 1 % for greater speeds. For field tests, the maximum allowable errors are ± 3 km/h for reference speeds up to 100 km/h and ± 3 % for greater speeds. Additionally, the average error during the field test should be within ± 1 km/h. If the device is to work autonomously, it has to also meet the field test tolerances with 99.8 % probability. Therefore, if the distribution is Normal, the standard deviation should not exceed 1 km/h.

The speed measurement devices are based on various physical principles that influence their placement with respect to the road and their cost. Nowadays, the most attention is given to the camera-based technologies because they provide a rich array of data that can be quickly processed on contemporary hardware. Single-camera devices are very interesting from an application perspective and a lot of work has been done to improve the methods they use for calibration and speed estimation. But certification of these devices is problematic. Stereo camera devices, on the other hand, are based on more transparent calibration and speed estimation methods and their certification should be easier.

The main scientific goal of this thesis is the answer to the question of whether it is currently possible to autonomously measure the speed of vehicles using a stereoscopic method with the average error within $\pm$ 1 km/h, the maximum error within $\pm$ 3 km/h, and the standard deviation within $\pm$ 1 km/h. The error ranges are based on the requirements of the OIML whose Recommendations serve as templates for metrological legislations of many countries. The devices that would use the measurement method whose error is within the specified ranges could receive proper certification and be used for speed limits enforcement. This question is answered using the newly proposed stereoscopic vehicle speed measurement method that exploits novel stereo camera pair calibration approach and overcomes the current state-of-the-art techniques.

The structure of the thesis is as follows. The first three chapters are dedicated to the theoretical background. In Chapter 2, an overview of the image acquisition and processing techniques is presented. The chapter emphasizes the camera calibration, object detection, localization, and tracking methods as these form a basis on which the developed calibration and measurement methods are built. Chapter 3 describes the devices that are currently used for vehicle speed measurement and presents the requirements that are imposed on them by the metrological legislation. The usability of stereoscopic speed measurement is shown by comparison with the devices that are described in this chapter. State-of-the-art stereoscopic measurement and calibration methods are introduced in Chapter 4. The next chapters are dedicated to the main contribution of this thesis. Chapter 5 states the hypothesis of this work and proposes measurement and calibration methods on which the experiments are performed in order to test the presented hypothesis. The design and results of the experiments are described in Chapter 6. Chapter 7 focuses on possible applications and future work and Chapter 8 concludes this thesis.

Chapter 2

Image acquisition and feature detection

The representation of a digital image and its acquisition are described in the first three sections of this chapter. The last section is dedicated to the selected object detection and tracking techniques that are used in the proposed vehicle speed measurement method.

2.1 Digital image

An image may be defined as a two-dimensional continuous function, $f(x, y)$, where x and y are spatial coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity of the image at that point. In order to be handled by computers, the image must be digitized [32]. The digitization is a process that transforms the continuous image to its digital form. The digitization involves two independent processes:sampling and quantization. The spatial sampling digitizes the coordinate values and the quantization digitizes the amplitude values. The resulting digital image has finite amount of picture elements, each of which has an intensity value that is represented by a finite number of bits [22]. Although digital images could be acquired in numerous ways, this work focuses on images that are acquired by digital cameras.

2.2 Camera model

A camera is an optical instrument that provides a mapping between a 3D world and a 2D image. This mapping is represented by a matrix that models the camera projection. In this work, camera mapping is represented by a pinhole camera model [68].

Pinhole camera model

The pinhole camera model can be described using the camera calibration matrix

$$P = K[R|t], \tag{2.1}$$

where K is a 3x3 matrix of internal camera parameters and $[R|t]$ is a 3x4 matrix of external camera parameters which represents the position and orientation of the camera with respect to the world coordinate system and consists of a 3x3 rotation matrix R and a 3D translation
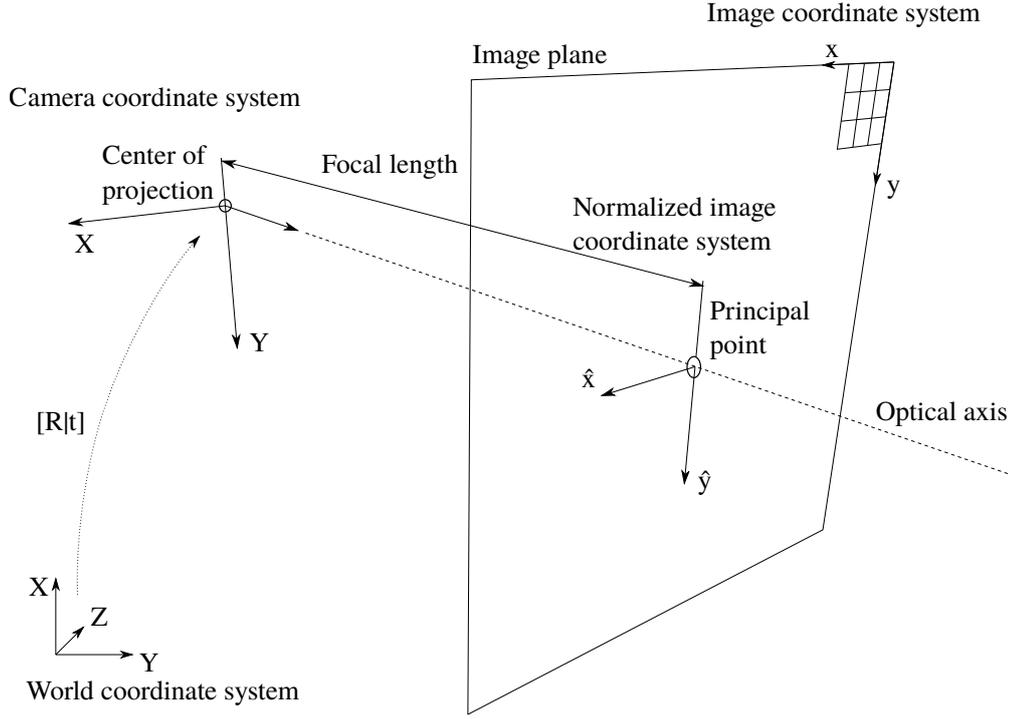
Figure 2.1: Pinhole camera model with geometrical representation of internal and external camera parameters and different coordinate systems.

vector t. The form of the matrix K is

$$K = \begin{bmatrix} f * m_x & s & c_x \\ 0 & f * m_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.2}$$

where f is a focal length of a camera, m_x and m_y are the numbers of pixels per distance unit in x and y direction, c_x and c_y are the image coordinates of a camera principal point, and s is a skew parameter. Overall the pinhole camera model has 11 degrees of freedom which need to be estimated in order to have a camera calibrated.

The geometrical representation of the internal and external camera parameters is shown in Fig. 2.1. The centre of projection is the origin of the camera coordinate system. It is the focal point for all rays that hit the camera. An image plane is a plane where the virtual image of the scene is formed. A distance between the centre of projection and the image plane is the focal length. The optical axis is a line that passes through the centre of projection and is orthogonal to the image plane. The point where the optical axis intersects the image plane is called the principal point. The principal point is the origin of the normalized image coordinate system, but his position is usually expressed in the image coordinates. The matrix of external parameters represents a transformation between the world coordinate system and the camera coordinate system.

The camera calibration process usually involves pointing the camera to a known calibration object or a pattern [78]. At least six 2D points (x) are selected in the captured
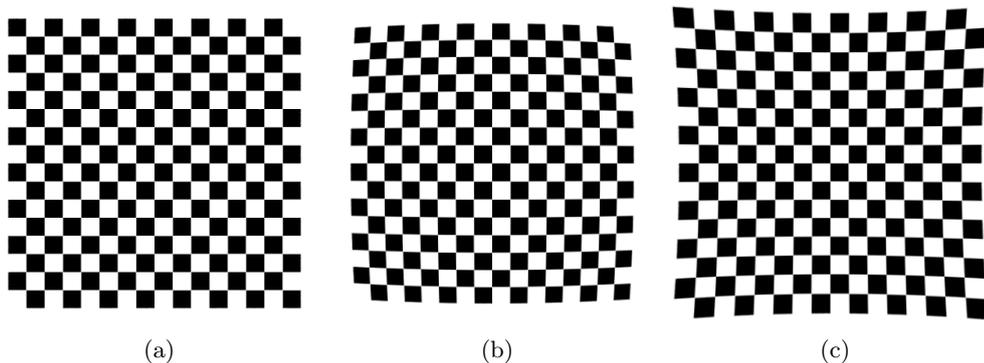
Figure 2.2: Most commonly encountered radial distortions. (a) Original undistorted image. (b) Barrel distortion. (c) Pincushion distortion.

image and related to their corresponding 3D points (X) using the formula

$$x_i = PX_i. \tag{2.3}$$

An initial estimate of the matrix P is obtained by a linear computation method and then further refined by minimizing geometric error

$$e = d(x_i, PX_i) \tag{2.4}$$

through a non-linear optimization.

Lens distortion

One important thing that is not covered by the pinhole camera model is a distortion introduced by a camera lens. The most common distortion is a radial distortion that is caused by an imperfect lens shape. The radial distortion distorts the ideal points radially from the distortion centre. The most commonly encountered types of radial distortions are shown in Fig. 2.2. Two widely used models for its description exist - Brown's [6] and Fitzgibbon's [21] (also known as the division model). Brown's model for the radial distortion is

$$\begin{aligned} x_u &= x_d + (x_d - x_c) * (K_1 * r^2 + K_2 * r^4 + ...) \\ y_u &= y_d + (y_d - y_c) * (K_1 * r^2 + K_2 * r^4 + ...) \end{aligned}, \tag{2.5}$$

where (x_u, y_u) are the undistorted point coordinates, (x_d, y_d) are the distorted point coordinates, (x_c, y_c) are the coordinates of a centre of distortion, K_x is the xth distortion coefficient, and r is a distance between the distorted point and the centre of distortion

$$r = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2}. \tag{2.6}$$

For lenses that exhibit severe radial distortions, the division model might be preferred as it is able to describe high distortions with fewer parameters [71]. The division model formula is

$$\begin{aligned} x_u &= x_c + \frac{x_d - x_c}{1 + K_1 * r^2 + K_2 * r^4 + ...} \\ y_u &= y_c + \frac{y_d - y_c}{1 + K_1 * r^2 + K_2 * r^4 + ...} \end{aligned}. \tag{2.7}$$

To compensate for the radial distortion, one needs to find out the values of the distortion coefficients and the distortion centre coordinates. These values can be computed, for example, by minimizing a cost based on the deviation from linear mapping (e.g. by keeping the images of straight scene lines straight) [24].

For the division model, a single parameter K is usually sufficient for the description of the radial distortion of most lenses [21]. Under this single-parameter division model, the images of the straight lines form circular arcs after distortion. The centre of the distortion and the single parameter K can then be computed by fitting circles to at least three arcs and solving a system of equations [7, 67].

Besides radial distortion, tangential and thin prism distortions exist, but their occurrence is less common [73]. It is likely that the distortion function is totally dominated by the radial component (especially dominated by the first term) with the other distortion components having a negligible effect [78].

2.3   Stereo camera pair

A stereo camera pair consists of two synchronized cameras with overlapping fields of view. If we are able to identify two points, one in each camera image, that are the projections of the same 3D object point, we can compute the position of the object point within a scene. Given enough corresponding points, we can reconstruct the whole scene. The mutual position of corresponding points is restrained by a relation between the two camera views that depend only on cameras' internal parameters and their relative pose and is independent of the scene structure. An epipolar geometry describes this relation [24].

Epipolar geometry

An epipolar geometry describes the relationship and constraints between two views. The relation between the two views of the calibrated cameras was described by Longuet-Higgins [42] as

$$\hat{x}'^{T} E \hat{x} = 0, \tag{2.8}$$

where E is a 3x3 essential matrix, $\hat{x}$ and $\hat{x}'$ are two points x and x' from the first and the second camera images respectively expressed in the normalized coordinates. The two points x and x' are the images of the same 3D point X, and therefore they correspond. Their normalized coordinates were obtained by removing the influence of the internal camera parameters

$$\begin{aligned} \hat{x} &= K^{-1}x \\ \hat{x}' &= K'^{-1}x' \end{aligned} \tag{2.9}$$

The essential matrix constrains the position of one of the points given the position of the other. For example, the position of point $\hat{x}'$ in the second camera image lies on a line $E\hat{x}$ and, vice versa, the position of the point $\hat{x}$ in the first camera image is constrained by a line $E^T \hat{x}'$. The constraining lines are called epipolar lines. All epipolar lines from one image plane intersect at the epipole which is also a point of intersection of baseline (the line that connects the two camera centres) with the image plane. The epipole also represents the image of the other camera centre. The elements of epipolar geometry are shown in Fig. 2.3.
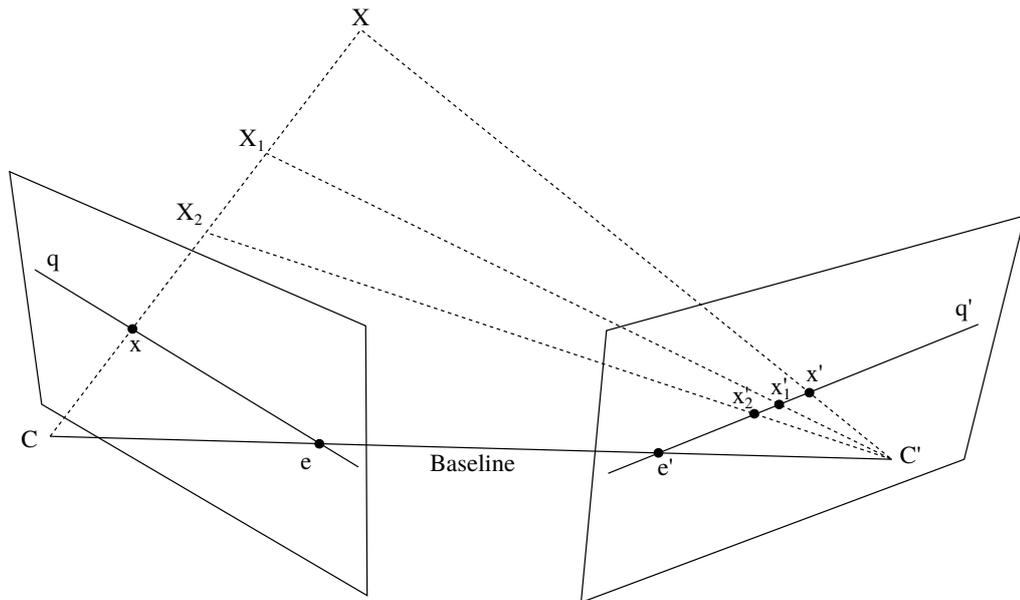
Figure 2.3: An epipolar geometry of a stereo camera pair. Points x and x' are two projections of 3D point X. The position of point x' in the second camera image is constrained by the epipolar line $q' = Fx$ and, vice versa, the position of point x is constrained by the epipolar line $q = F^T x'$. Points e and e' are the epipoles.

The concept of the essential matrix was later generalized for the uncalibrated cameras by Faugeras [19] and Hartley et al. [27]. They simultaneously introduced the fundamental matrix which does not require the normalization of point coordinates

$$x'Fx = 0. \tag{2.10}$$

The fundamental matrix can be computed from point correspondences alone. It has only seven degrees of freedom; thus, a minimum of seven correspondences are required for its computation. However, the use of only seven correspondences leads to a cubic polynomial equation which has one or three real solutions [26]. The seven-point algorithm can, therefore, yield one or three fundamental matrices. A unique solution can be obtained using eight or more correspondences. In order to achieve the best performance when using the eight-point algorithm, simple data normalization is required [29]. The methods for computation of fundamental matrices were reviewed by Zhang [79]. The same methods can be used to compute the essential matrix using the point correspondences expressed in normalized coordinates.

Scene reconstruction and triangulation

A position of a 3D point X within a scene can be reconstructed from its projections x and x'. The point X projects as

$$\begin{aligned} x &= PX \\ x' &= P'X \end{aligned}. \tag{2.11}$$

The reconstruction process is based on a triangulation which aims to find an intersection point of two rays in space. The two rays that are triangulated can be easily recovered

from the two projections and the camera calibration matrices. Although the task of rays intersection in space is trivial, in the presence of noise, the two rays are not guaranteed to intersect. In that case, the triangulation process seeks to find the best solution under an assumed noise model.

Hartley and Sturm [28] presented the optimal triangulation method under the assumption of Gaussian noise model. The method tries to correct the noisy correspondences which do not, in general, satisfy the epipolar constraint 2.10 by replacing them with the closest points u and u' that do satisfy the constraint. The closest points can be found by minimizing the function

$$d(x, u)^2 + d(x', u')^2. \tag{2.12}$$

Once the points u and u' have been identified, the corresponding rays will meet precisely, and their intersection can be found by any triangulation method.

Perhaps the most commonly used method for triangulation is the linear method which combines the equations 2.11 into a form $AX = 0$ where

$$A = \begin{bmatrix} x\mathbf{p}^{3T} - \mathbf{p}^{1T} \\ y\mathbf{p}^{3T} - \mathbf{p}^{2T} \\ x'\mathbf{p}'^{3T} - \mathbf{p}'^{1T} \\ y'\mathbf{p}'^{3T} - \mathbf{p}'^{2T} \end{bmatrix} \tag{2.13}$$

and $\mathbf{p}^{iT}$ is the ith row of P. This system of equations is linear in X and can be solved by, for example, singular value decomposition.

In order to perform triangulation and reconstruct the scene, we need to obtain the camera matrices P and P'. Different types of scene reconstructions exist, and they vary by the type of transformation (e. g. projective, affine, or similarity) that needs to be applied to them to get a true reconstruction. To perform measurements within a reconstructed scene, we need at least the similarity (also known as the metric) reconstruction. For this type of reconstruction the required camera matrices P and P' have forms

$$\begin{aligned} P &= K[I|0] \\ P' &= K'[R|t] \end{aligned} \tag{2.14}$$

where K and K' are the internal parameters for the first and the second camera, I is a 3x3 identity matrix and $[R|t]$ is a 3x4 matrix which specifies a position and an orientation of the second camera with respect to the first camera. Note that, for simplicity, the world coordinate frame coincides with the first camera. The camera matrices that allow metric reconstruction can be obtained, for example, by the progressive refinement of camera matrices retrieved from the fundamental matrix using the stratified method [17, 25] or directly by using five or more ground control points [24].

Correspondence problem

The correspondence problem is fundamental in computer vision. For the two views of a stereo camera pair, the correspondence problem can be formulated as follows. Given a projection of a 3D point in one view, find the corresponding projection of the same 3D point in the other view, if possible. The correspondence search area can be limited to a line if the essential or fundamental matrix is given. Two main classes of correspondence search algorithms exist - the correlation-based and the feature-based.

Inputs : A pair of images I_1, I_2, the width of the search window $2W + 1$, and a point $p_1 = (x_1, y_1) \in I_1$.

Output: A point $p_2 = (x_2, y_2) \in I_2$ that corresponds to the point p_1.

$c_{max} \leftarrow 0$;

$s_1 \leftarrow \sqrt{\sum_{i=-W}^{W} \sum_{j=-W}^{W} (I_1(x_1 + i, y_1 + j) - \overline{I_1(x_1, y_1)})^2}$;

foreach *pixel* $p = (x, y)$ *in image* I_2 **do**

$\qquad s_2 \leftarrow \sqrt{\sum_{i=-W}^{W} \sum_{j=-W}^{W} (I_2(x + i, y + j) - \overline{I_2(x, y)})^2}$;

$\qquad c \leftarrow \dfrac{\sum_{i=-W}^{W} \sum_{j=-W}^{W} (I_1(x_1 + i, y_1 + j) - \overline{I_1(x_1, y_1)})(I_2(x + i, y + j) - \overline{I_2(x_1, y_1)})}{s_1 s_2}$

$\qquad$ **if** $c > c_{max}$ **then**

$\qquad \qquad c_{max} \leftarrow c$;

$\qquad \qquad p_2 \leftarrow p$;

$\qquad$ **end**

end

Note $\overline{I_k(x, y)}, k = 1, 2$ is the average of the image k within the window of size $2W + 1$ centred around point (x, y).

Algorithm 1: Window-based correspondence search with zero-mean normalized cross-correlation similarity measure.

The correlation-based methods establish the correspondences by matching image intensities. These techniques can be broadly classified into two categories - the local and the global [61]. The local algorithms match intensity values within a finite image window. For each pixel in one image, a template window is constructed. The template window is then slid across the other image, and for every displacement, the similarity measure between the intensity values of the template window and the intensity values of the underlying area in the other image is computed. The match is established for the displacement where the similarity measure value is maximal. Widely used similarity measures are the sum of squared differences, the sum of absolute differences, and the normalized cross-correlation [18]. The Algorithm 1 shows a window-based search for a point p_2 that corresponds to a point p_1 with a zero-mean normalized cross-correlation similarity measure which is invariant to both shift and scale photometric distortion.

Unlike local methods, the global algorithms work with the whole image. Their goal is to estimate a displacement function that minimizes a global cost function that combines data and smoothness terms. This minimization can be done in various ways, such as graph cuts, simulated annealing, or probabilistic diffusion [61]. Both local and global methods produce dense correspondences; that is, they are able to find a correspondence for every image point.

Some methods that do not fit to the local and global categories exist. These methods include, for example, approaches based on wavelet transforms [3], such as [30, 40], or spiral search [49].

The feature-based approach to correspondence search is based on detecting and extracting the regions of interest from both images and matching them together using their descriptions. The image feature can be, for example, an edge, a corner, or a blob. Different types of features require different detectors to be used. Canny [9], Sobel [13] or Prewitt [60] detectors are the examples of the edge detectors. The corners can be detected using, for example, Laplacian of Gaussian, Difference of Gaussians, or Harris detector [23]. Maximally Stable Extremal Regions [48] or Principal curvature-based region detector [11] are the examples of the blob detectors.

Detected features are then submitted to a feature descriptor that attempts to describe the feature with a vector that is constructed from the feature's neighbourhood. The feature description strives to be invariant to scale, rotation, and shift. The SIFT [43], SURF [2], LBP [59], FREAK [1], and BRIEF [8] descriptors can be named as examples. Once the features of both images have been described, they can be matched together based on their distance in the feature space. The binary descriptors can be easily compared using the Hamming distance. The non-binary descriptors are usually compared using the Euclidean norm. Unlike the correlation-based approach, the feature-based methods produce only sparse correspondences, that is, they are able to find the correspondences only for the detected feature points and not for each image pixel.

2.4 Selected object detection and tracking methods

This section presents the selected object detection and tracking methods that are relevant to this work and that are used in the implementation. Object detection algorithms aim to locate the objects of interest within an image frame. The tracking algorithms estimate the positions of the detected objects in the following frames and help maintain the context in frames where the detector fails to locate the tracked object. Although many approaches to object detection and tracking exist [77, 34, 39], this section focuses only on those that are implemented as a part of the proposed vehicle speed measurement method, namely, WaldBoost and Kalman filter.

WaldBoost

The WaldBoost is an algorithm for computer vision two-class classification problems with near-optimal time and quality (error rate) trade-off. It was proposed by Šochman and Matas [63] and it integrates the AdaBoost algorithm and Wald's optimal sequential probability ratio test [70]. Hardware implemetations exhibit excellent accuracy and performance at very low power consumption [50].

A two-class classification is a task of classifying the object of a given set into two groups which can be decided by a sequential decision strategy. The sequential decision strategy S can be broken down into a series of steps. In each step t, a measurement x_t on the object x is taken and a decision function S_t decides whether the object belongs to the class $y \in \{+1, -1\}$ or whether the decision process should continue with the next step $t + 1$. The optimal strategy is the one that minimizes the time-to-decision and keeps the probabilities for both kinds of errors under the specified thresholds. Wald [70] showed that the sequential probability ratio test defined as

$$S_t^* = \begin{cases} +1, & R_t \leq B \\ -1, & R_t \geq A \\ continue, & B < R_t < A \end{cases}$$

is the optimal sequential decision strategy. The A and B are the error thresholds and the R_t is the likelihood ratio

$$R_t = \frac{p(x_1, ..., x_t | y = -1)}{p(x_1, ..., x_t | y = +1)}.$$

The optimal values for A and B are difficult to compute in practice and Wald suggests to set them to their upper and lower bounds

$$A' = \frac{1 - \beta}{\alpha}, \qquad B' = \frac{\beta}{1 - \alpha}, \tag{2.15}$$

where α is the probability of the error of the first kind and β is the probability of the error of the second kind.

To select and order the measurements and to estimate the joint probability density function, Šochman and Matas suggested using the AdaBoost algorithm. The AdaBoost algorithm produces a strong classifier of the form

$$H_T(x) = \sum_{t=1}^{T} h^{(t)}(x),$$

where $h^{(t)}$ are selected simple weak classifiers and T is the size of a labelled training set.

Using the trained classifier and desired error probabilities, the sequential probability ratio test becomes

$$S_t^* = \begin{cases} +1, & H_t \geq \theta_B^{(t)} \\ -1, & H_t \leq \theta_A^{(t)} \\ continue, & \theta_A^{(t)} < H_t < \theta_B^{(t)} \end{cases},$$

where $\theta_A^{(t)}$ and $\theta_B^{(t)}$ thresholds are set to the upper and lower bounds of A and B according to the Eq. 2.15. The WaldBoost algorithm uses the modified sequential probability ratio test for classification. In each step, the classifier output H_t is compared to the thresholds $\theta_A^{(t)}$ and $\theta_B^{(t)}$ and either a decision is made, or the next weak classifier is taken. If all the weak classifiers are executed, and a decision is not made, the H_T value is compared to a threshold of γ. If the H_T value is greater than the threshold, the object is classified as $+1$; otherwise, it is classified as -1.

Kalman Filter

The Kalman Filter [36] is a recursive algorithm that predicts the state of a linear system from past estimations and noisy measurements. Its applications are numerous, for example, object tracking, navigation, control systems, or signal processing. A comprehensive introduction to Kalman Filter can be found in [4, 72]. The algorithm utilizes five equations that are used in two alternating steps. The prediction step where the current state and the uncertainty of the current system state estimate are extrapolated and the update step in which the projected estimates are corrected.

The algorithm is initialized by initial system state and uncertainty (covariance) of the initial state. The initial values are passed to the prediction step as the current state $\hat{x}_k$ and current state uncertainty P_k estimates where they are extrapolated to the next system state $\hat{x}_{k+1}^-$ and the next state uncertainty P_{k+1}^- projections

$$\hat{x}_{k+1}^- = A_k \hat{x}_k + B_k u_k + w_k$$
$$P_{k+1}^- = A_k P_k A_k^T + Q_k \qquad ,$$

12
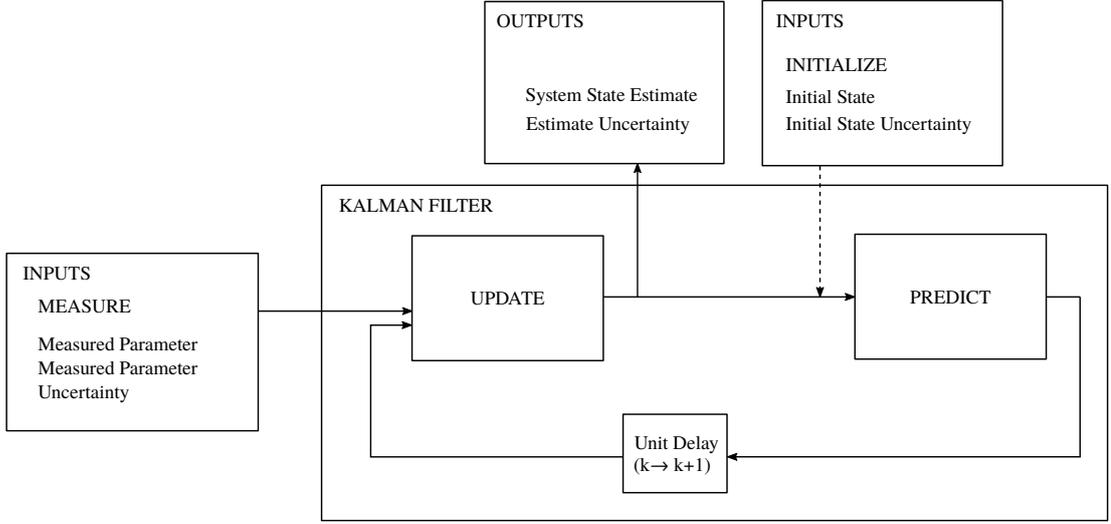
Figure 2.4: Schematic description of the Kalman Filter algorithm. Adapted from [4].

where the A_k is a state transition matrix, the B is a matrix that maps the control input u_k to a state, and the w_k is a zero-mean Gaussian process noise with process noise covariance Q_k.

Then a measurement is taken and the measured system state and measurement uncertainty are obtained. The measured and predicted values are submitted to the update step where the current system estimate $\hat{x}_k$ and the current state estimate uncertainty P_k are computed using equations

$$
\begin{aligned}
\hat{x}_k &= \hat{x}_k^- + K(z_k - H_k \hat{x}_k^-) \\
P_k &= (I - K_k H_k) P_k^-
\end{aligned}
,
$$

where $\hat{x}_k^-$ represents the predicted state, P_k^- is the predicted state uncertainty, z_k is the measured state, and I is the identity matrix. The H_k is a matrix that maps the true state x_k to the measured state z_k

$$
z_k = H_k x_k + v_k,
$$

where v_k is a measurement noise that is assumed to be zero-mean Gaussian with covariance R_k. The K is the Kalman gain

$$
K = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1},
$$

The output of the filter are the results of the update steps, that is, the current state estimate $\hat{x}_k$ and the current state uncertainty P_k. From these results, new predictions are made, and the new iteration starts.

Chapter 3

Vehicle speed measurement

The speed of a vehicle can be measured by various devices that can be divided into two classes according to the intrusiveness of their installation. The devices that belong to the first class have to be embedded into or placed onto the road, and because of that, they are classified as intrusive. The devices that belong to the second class are placed above or by the side of the road, and because of that, they are labelled as non-intrusive. The properties of the devices that belong to both classes are described in the first two sections of this chapter. The information about the devices described in these sections was taken from [5, 37, 45, 47]. Only the devices that pass the tests specified by local metrological legislation can be officially used for vehicle speed measurement. The test requirements that are common for many metrological legislations are presented in the final section of this chapter.

3.1 Intrusive technologies

Pneumatic tube detectors, inductive loops, magnetic and weigh-in-motion sensors belong to the intrusive category. Although the devices are accurate, and by themselves low cost, they need to be placed either on the road surface or embedded directly into it. Their installation and maintenance are therefore problematic and expensive because they usually require lane closure, which disrupts the traffic, and a pavement cut with a subsequent repair or resurfacing of the road. The devices detect the presence of the vehicles; therefore, at least two sensor pieces at a known distance apart are needed for the speed measurement. They are also insensitive to the weather as they are located in close proximity to passing vehicles.

Pneumatic tube detectors

Pneumatic tube detectors represent perhaps the oldest and the simplest of the vehicle detectors. The pneumatic tubes are placed across the road, perpendicularly to the traffic flow direction. As the vehicle's tires pass over them, a burst of air pressure along a rubber tube is sent. The air pressure pulse closes an air switch and sends an electric signal that marks the passage of a vehicle. The pneumatic tube detectors are low-cost, quick to install, portable, and power-efficient. They are able to measure speed on multi-lane and multi-direction roads but usually only for a single lane at a single time. They are unsuitable for high flow and high-speed roads because these conditions lower their detection accuracy and shorten even more their already short service life [5, 37, 45, 47].

Inductive loop detectors

Inductive loop detectors consist of three components: an oscillator, a lead-in cable, and one or more turns of insulated loop wire. The loop is usually embedded in the pavement or encased and placed on the road surface. The oscillator excites the loop with an electric signal. When a vehicle stops on or passes over the loop, the inductance of the loop is decreased, which increases the oscillation frequency. The change in frequency indicates the presence of a vehicle. Single loop or dual loop setups can be used [5, 37, 45, 47].

Single loop setups provide basic traffic parameters such as volume, presence, occupancy, classification, and gap. The speed can be measured under the assumption of known vehicle length as

$$s = \frac{l_v + l_d}{t_o},\tag{3.1}$$

where l_v and l_d are lengths of the vehicle and the detector, and t_o represents the detector occupancy time. Generally, the lengths of the individual vehicles are not known, and the average vehicle length is used instead. The speed data computed in such a way serve mostly for statistical purposes rather than for the speed limit enforcement.

The speed of individual vehicles can be measured using dual loop setups. These setups consist of two single loops placed at a known distance apart. When the first loop detects the vehicle, a timer is started and runs until the vehicle is detected by the second loop. The vehicle speed can be then computed as

$$s = \frac{d}{t_1 - t_0},\tag{3.2}$$

where d is a distance between the two loops and t_0 and t_1 are vehicle entry times at the first and the second loop.

The advantages of inductive loop detectors include low-cost, high accuracy, and applicability to a large variety of traffic surveillance tasks. On the other hand, the loops may fail in detecting the vehicles with relatively low metal content or unusual chassis configurations. One setup can provide measurements for one road lane only, so usually, multiple setups are needed to instrument a location. Other disadvantages include the need for pavement cut and line closure during the installation or the maintenance and vulnerability to damages caused by street maintenance activities, improper installation, heavy traffic, or temperature changes.

Magnetic sensors

Magnetic sensors detect the presence of a vehicle by monitoring the changes in the Earth's magnetic field created by the ferrous metal objects. Similarly to the induction loop detectors, two sensors are needed for vehicle speed and length measurement. Two types of magnetic field sensors exist.

Dual-axis flux-gate magnetometer consists of a primary and two secondary windings on a coil with high permeability soft magnetic core. When a change in the magnetic field occurs, the secondary windings generate a voltage that is compared to the sensitivity threshold. If the threshold is exceeded, the vehicle presence is recorded.

The induction or search coil magnetometer contains a single coil winding around a permeable magnetic rod. It is able to detect only vehicles in motion as it measures the voltage induced by the variations of the magnetic flux. When the measured voltage exceeds the sensitivity threshold, the vehicle is recorded.

The properties of the magnetic sensors are similar to the properties of the induction loop detectors. Unlike the induction loop detectors, their installation requires smaller pavement cut as they usually need only a small hole in the road surface. Due to their low profile, they are less susceptible to traffic stresses, but their detection zones are smaller [5, 37, 45, 47].

Weigh-in-Motion systems

Similarly to the pneumatic tube detectors, the weigh-in-motion sensors detect the vehicle force on the pavement. Bending plates and piezoelectric sensors are examples of the weigh-in-motion sensors. Two sensor pieces are required for vehicle speed measurement [5, 37, 45, 47].

The bending plate consists of a steel or a rubber plate with attached strain gauges. The strain gauges measure the deflection of the plate under the vehicle tires. The piezoelectric material is capable of converting the kinetic energy to electrical energy. When a sensor is subjected to the change in force created by a passing vehicle, it generates a voltage that is proportional to the weight of the vehicle.

Among the advantages of the weigh-in-motion systems are the low cost, low power consumption, high accuracy, and broad speed measurement range. The drawbacks are similar to the induction loop sensors. Their installation and maintenance require lane closure and usually a pavement cut. Multiple devices are necessary to cover a location fully, and they have a short service life.

3.2 Non-Intrusive technologies

Ultrasonic sensors, infrared sensors, laser detectors, radars, and camera-based technologies belong to the non-intrusive category of technologies. The non-intrusive devices represent an emergent field that expands rapidly with rising computational power and continuing advances in signal processing. Their main advantage is that they are placed either above the road or by its side, which makes their installation and maintenance easier and cheaper in comparison with the intrusive technologies. The low costs of their installation and maintenance at least partially offset the higher initial costs. The higher initial costs, weather susceptibility, and lower accuracy are their main disadvantages when compared to the intrusive devices.

Infrared sensors

The infrared sensors contain a light-sensitive element that converts the received light energy into the electrical signals. The vehicle presence is detected by analyzing the signal. The sensors are mounted either in the overhead or the side-looking configuration. The active and passive infrared sensors exist [5, 37, 45, 47].

The active infrared sensors use pulsed or continuous LED or laser diodes that operate in the near-infrared spectrum to illuminate the detection zones. The infrared light that is reflected from the vehicles passing through the detection zone is captured by the sensor. The infrared sensors measure the time for the reflected light to return. When the vehicle is present, the reflection time is lower, and on this basis, the vehicle is detected. The speed of a vehicle can be measured by splitting the transmitted light into two or more beams and recording the times at which the vehicle enters the detection area of each beam.

The passive sensors transmit no energy of their own. They detect the energy that is emitted from or reflected by vehicles, road surfaces, and other objects. When the vehicle enters the detection zone, its presence is registered by a change in the detected energy. Multiple detection zones need to be used for vehicle speed measurement.

The advantages of the infrared sensors include quick installation, reliability, high accuracy, and ability to provide measurements for multiple lanes in the side-looking configuration. On the other hand, their accuracy can be affected by poor weather conditions, and in the overhead configuration, multiple devices are usually needed to cover all road lanes.

Ultrasonic sensors

The ultrasonic sensors transmit sound waves at frequencies that are beyond the human audible range (above 20 kHz). The reflected acoustic waves are detected by the sensor and analyzed for the presence of a vehicle. The sensors are installed either in the overhead or the side-looking configuration. Side-looking configuration enables the measurement of multiple road lanes but is affected by an occlusion. Overhead configuration is unaffected by occlusion, but multiple devices are needed to cover multiple lanes. Pulsed and continuous ultrasonic devices exist [5, 37, 45, 47].

The pulsed devices transmit a series of ultrasonic pulses and measure the time it takes for the pulse to reflect and return to the sensor. Based on the measured time, the distance to an object in front of the sensor is measured. The vehicle is detected when the measured distance is shorter than the background distance. Two pulsed devices at a known distance apart are needed to measure vehicle speed.

The continuous detectors transmit a continuous ultrasonic wave and detect the frequency shift of returned waves. The moving vehicle is detected when the shift occurs. A single continuous ultrasonic detector is needed to measure the vehicle speed as it can be computed from the frequency shift using the Doppler principle (Eq. 3.3).

The advantages of ultrasonic sensors are quick and non-intrusive installation, high accuracy, and reliability. The disadvantages are high initial cost and performance influenced by the weather conditions.

Acoustic array sensors

Acoustic array sensors consist of a two-dimensional array of microphones that detect the sound of the approaching vehicles. The array of microphones can distinguish whether the detected sound is coming from the specified detection zones or not by monitoring the time of sound arrival at individual microphones. The sounds that originate in the detection zones are analyzed while the rest are attenuated. As the vehicle enters the detection zone, the change in sound energy is registered, and the vehicle is detected. The speed of a vehicle can be computed by tracking the incremental change in location at frequent intervals [38]. The acoustic array sensors are usually mounted on poles at the sides of the roads.

The advantages include quick and non-intrusive installation, passive operation, and multiple lane operation. The disadvantages include degraded performance during weather conditions that affect sound propagation and problematic detection of slow-moving or electric vehicles (i.e. vehicles that do not produce much noise) [5, 37, 45, 47].
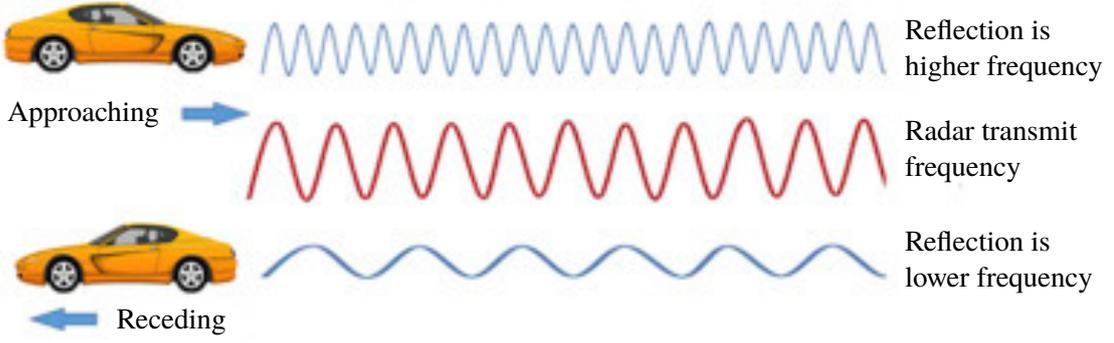
Figure 3.1: A frequency shift of a reflected signal for approaching and receding vehicles [33].

Microwave radar systems

The microwave radar emits electromagnetic signals with a frequency range of 1 GHz to 30 GHz. The reflected electromagnetic signals are detected by the device and analyzed for the presence of a vehicle. The radars are usually installed overhead on the poles, sign bridges, or overpasses. Side-looking configuration is also possible but less common. Continuous wave and frequency modulated continuous wave radars exist [5, 37, 45, 47].

The continuous wave radar transmits a signal that is constant in frequency with respect to time. When the signal reflects from a moving vehicle, its frequency changes due to the Doppler effect (see Fig. 3.1). The frequency shift is extracted from the received signal, and the presence of a vehicle is detected. If no vehicle is present, the frequency of the reflected signal is unchanged, and so the continuous wave radar cannot detect stationary vehicles. The speed of the vehicle can be computed directly from the shifted signal as

$$\Delta f = \frac{2v f_s cos(\theta)}{c} \tag{3.3}$$

where Δf is the Doppler shift, f_s is the transmitted frequency, θ is the angle between the direction of the transmitted signal and the direction of the vehicle passage, v is the speed of the vehicle, and c is the speed of the radar signal propagation.

The frequency modulated continuous wave radar transmits a signal with a frequency that is constantly changing with respect to time. In the reflected signal, the change in frequency is delayed (see Fig. 3.2), and from this delay, a distance to the reflector can be measured as

$$d = \frac{t_d c}{2}, \tag{3.4}$$

where t_d is the time delay, and c is the speed of the radar signal propagation. The vehicle is detected when the measured distance is smaller than the background distance. Because the vehicles are detected based on the measured distance, the stationary vehicles can also be detected. The speed of a vehicle is measured by dividing the radar field of view into several range bins. The radar then records times at which the vehicle entered each bin. From the recorded times and the known distance between bins, the speed can be computed. The disadvantage of this approach is that the radar operation is limited to a single lane.

The advantages of microwave radars include ease and non-intrusiveness of installation, high accuracy, long range of operation, ability to cover multiple lanes, and resistance to changing weather and lighting conditions. To perform correct measurements, the radar

18
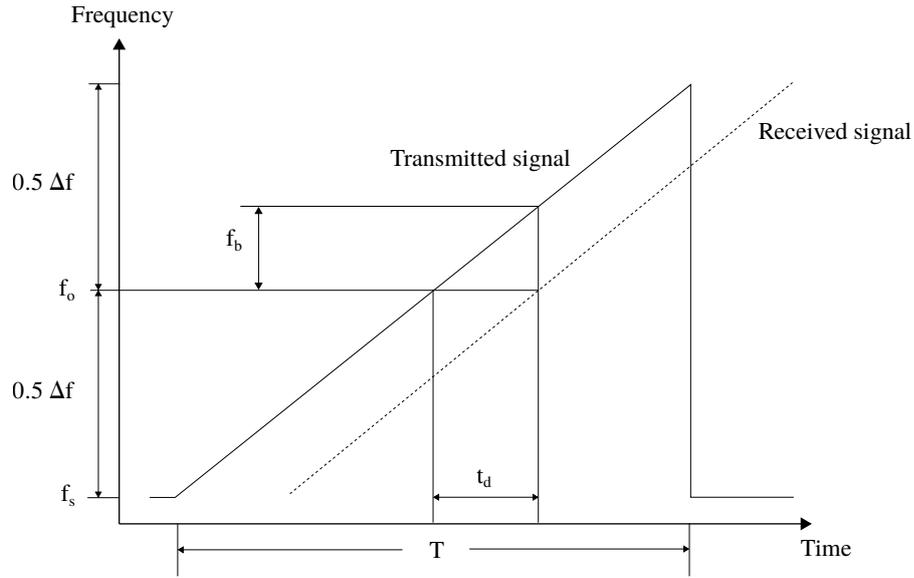
Figure 3.2: Reflected signal of frequency modulated continuous wave radar has the change in frequency delayed by t_d [10].

beam should cover only the target vehicle. This is not always possible since the radar beam widens as it gets further from the radar antenna. If there are more vehicles in the radar beam, the measurement that is obtained cannot be assigned to the targeted vehicle with certainty. Furthermore, the radar inclination angle with respect to the traffic flow has to be measured in order to compensate for cosine error [75]. The need for precise installation and careful aiming together with high initial cost belong among the disadvantages of the microwave radar systems.

Video image detection

Video image detection systems typically consist of one or more cameras that record passing vehicles. Recorded images are then analyzed using specialized software or on dedicated hardware that can be a part of the camera itself. The cameras are usually installed on the poles, sign bridges, or overpasses looking down at the upstream or downstream traffic. Three types of video image detection systems exist: tripline, closed-loop, and data association systems.

The tripline systems enable a user to define a number of detection zones in the field of view of the camera. When a vehicle enters the detection zone, a timer is started. The timer is stopped when the vehicle arrives at the end of the detection zone. The vehicle crossing is detected from the intensity changes in the tripline area. Given a known length of the detection zone, the vehicle speed can be computed. The speed measurement accuracy of the tripline systems is quite low [46] because the speed is measured over a fairly short distance (tens of meters) with a relatively high error of vehicle localization. Additionally, the length of the detection zone that is used for speed computation is longer than the actual distance that the vehicle travelled in the measured time (see Fig. 3.3). Therefore, even if the vehicle was located precisely, the measured speeds are always higher than the actual speeds.
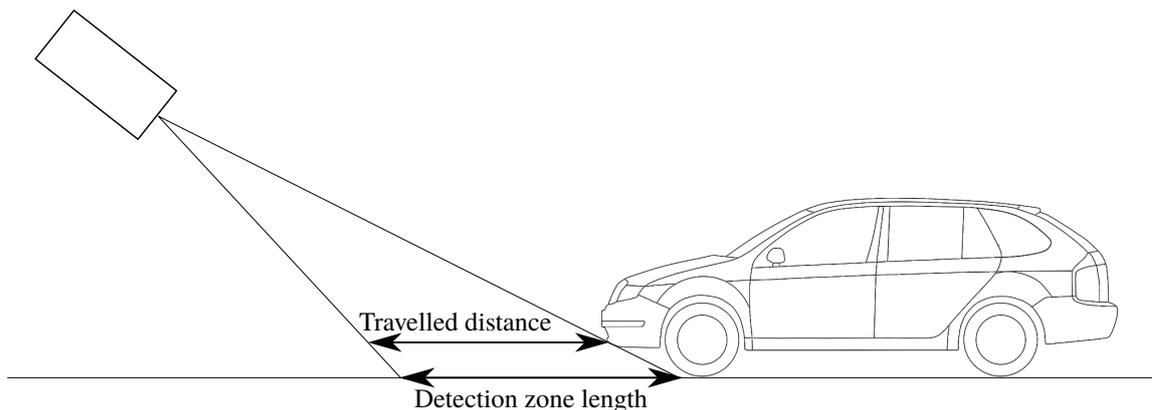
Figure 3.3: The discrepancy between the length of the detection zone and the travelled distance.

The closed-loop tracking systems detect and track passing vehicles continuously throughout the field of view of the camera. From multiple detections, the vehicle trajectory can be reconstructed and vehicle spot speed estimated. The state-of-the-art closed-loop tracking system measure vehicle speed by reprojecting the reconstructed trajectory to the road plane, measuring the length of the reprojected trajectory in pixels, converting the pixel length to metres by multiplying it by an appropriate scene scale factor, and dividing the metre length by time. In order to do this, the camera calibration parameters and scene scale factor need to be estimated.

Modern methods strive for the fully automatic estimation of the camera parameters from the scene vanishing points [12] and automatic scene scale factor estimation by constructing 3D boxes around vehicles and relating the pixel dimensions to the real-world dimensions of the recognized vehicle type [65, 66]. The fully automatic methods achieve a mean percentage error of 1.4 %, mean absolute error of 1.1 km/h, and 99 % of measurements have an absolute error below 3 km/h and percentage error below 4.1 %. The disadvantage of these methods is that they rely on the detection of the whole vehicle, which could lead to worse performance in, for example, low-light situations.

Alternatively, approaches that require manual input in the form of measured distances between several points on the road plane for each lane exist [44]. The points and their distances then provide a scene scale as well as a way to rectify the road plane through homography. To mitigate the discrepancy between the measured and the actual distance, the height of the tracked vehicle points above the road plane can be calculated [16] and used for correction. The manual methods achieve the mean error of -0.5 km/h with a standard deviation of 1.4 km/h and the worst errors of -4.7 km/h and 6 km/h. Although the speed measurement error significantly decreased since the first closed-loop tracking prototypes [46] thanks to the extensive research, the measurement error is still too high for these systems to be used for speed limit enforcement.

The data association tracking systems identify and track passing vehicles across multiple cameras. When a vehicle is detected in a camera view, a set of features that describe the vehicle is extracted. The set of features can be then used for the vehicle re-identification in other cameras' views. The data association tracking systems are typically used for section speed measurement as the cameras' views are usually non-overlapping, and the cameras are located at a greater distance from each other. However, approaches for spot

speed measurement also exist [41] although their accuracy is notably lower (mean absolute error of 1.44 km/h and the maximum error of 2.62 km/h). In the case of section speed measurement, the start and the end of the section are usually marked directly on the road surface to make the vehicle position estimation easier. Given the known distance between the two markers and the times at which the vehicle was detected at each marker, the vehicle speed can be computed. The greater the distance between the two cameras, the larger an error in vehicle position estimation can be while still preserving the ability of the system to measure the vehicle speed with the accuracy required for enforcement.

The video image detection systems provide a rich array of data that can be quickly processed on contemporary hardware. They have the ability to monitor multiple road lanes with multiple detection zones that are easy to set up and customize. Unlike other technologies, the video image detection systems are able to provide photographic evidence for enforcement of traffic law violations. Because the photographic evidence containing the identification of the offender (e. g. a vehicle with its license plate) is required as the proof of offence, the other systems are usually coupled with cameras when the enforcement is required which increases their costs. Other advantages include quick and non-intrusive installation, reliability and flexibility.

The performance of the system can be affected by low-visibility conditions, for example, due to bad weather (e. g. snow, rain, or fog) or insufficient lighting (e. g. during the night). Their performance can also be adversely affected by occlusion or camera motions due to vibrations. Other disadvantages include high initial cost and the need for precise calibration when used for speed or length measurement.

To better estimate the distance that vehicle travelled, the single cameras can be coupled with other devices, for example, IR or TOF sensors to create RGBD cameras [52]. However, more advantageous would be adding another camera in order to create a stereo camera pair. The distance data can be then extracted from the stereo image pairs using the techniques described in the previous chapter. Stereo-based camera devices are currently not utilized for live traffic surveillance operations but are in the focus of recent research. They have all the advantages of single-camera-based video detection systems and the potential to provide more precise and accurate measurements while being less affected by adverse weather and lighting conditions [31, 14, 76].

3.3 Metrological legislation

In order for the device to be recognized as a measurement device, it has to meet the requirements of the metrological legislation and pass the specified tests. The metrological legislations vary by country, and it is not possible to review them all within the scope of this thesis. Therefore the thesis focus on the metrological legislations that are based on the recommendations of the International Organization of Legal Metrology[1].

OIML is a worldwide, intergovernmental organization created in 1955 whose primary aim is to harmonize the regulations and metrological controls applied by the national metrological services, or related organizations, of its Member States [58]. At the time of writing this thesis, OIML has 61 Member States which participate in the work of OIML and that have ratified the OIML Convention [56], and 62 Corresponding Members who are countries that want to be informed of the OIML activities but do not want to actively participate in the decision-making process.

---

[1]https://www.oiml.org/ (OIML)

Table 3.1: Error requirements on the vehicle speed measuring devices according to the OIML Recommendation [57].

		Laboratory test	Field test
Max. error range	<= 100 km/h	± 1 km/h	<= 3 km/h
	> 100 km/h	± 1 %	<= 3 %
Mean error range	<= 100 km/h	-	± 1 km/h
	> 100 km/h	-	± 1 %
Stdev (autonomous)	<= 100 km/h	-	< 1 km/h
	> 100 km/h	-	< 1 %

The OIML publish several types of documents. The model regulations for a number of categories of measuring instruments are published as the Recommendations documents. The Member States are morally obliged to implement the model regulations as far as possible. The Recommendation that describes the requirements of the vehicle speed measurement devices is focused on the radar measuring equipment [57]. Nonetheless, the general requirements and principles of this Recommendation are usually utilized for other types of vehicle speed measuring equipment as well.

Apart from the construction and protection requirements, the documents describe the process of pattern approval. The pattern approval process consists of several tests that check the ability of the device to provide reliable measurements within an acceptable error range under different conditions. The device under test shall provide the measurement range that includes at least the range from 30 km/h to 150 km/h. The laboratory test is performed in a controlled condition environment. The measurement error for the laboratory test should be less than ± 1 km/h, or ± 1 % at speeds above 100 km/h. The tests of the effects of influence factors and disturbances test the mechanical and climatic resistance and the reliability of electronic and logical components.

The metrological field test is performed in actual traffic. During this test, 500 measurements are made, of which none should give a positive error larger than + 3 km/h, or + 3 % at speeds above 100 km/h. The average error of all results should be within ± 1 km/h. The device does not need to provide measurement for every passing car. If the measurement is recognized as faulty, it can be discarded. For the autonomous devices, the recognition and the discarding of the faulty measurements have to be done automatically by the device itself. For the manually operated devices, the decision can be made by the operators. Additionally, the autonomously operated devices shall provide a high level of confidence that the measurement error is within the permissible limits. That is, the autonomous devices have to meet the field test error tolerances with 99.8 % probability. Therefore, under the assumption of Normal distribution, the standard deviation of errors have to be less than 1 km/h. All measured speeds are compared to the ground truth measurements provided by a device that has uncertainty better than one-third of the device under the test; 99.8 % of the reference device results should have errors that are within ± 1 km/h, or ± 1% at speeds above 100 km/h. The laboratory and field test error requirements are summarized in Table 3.1.

The document [80] is a part of the metrological legislation of the Czech Republic. It can serve as an example of how the recommendations of the OIML might be incorporated to metrological legislation of its Member States. Some requirements in this documents are stricter, for example, the required measurement range is extended to at least 200 km/h and

the maximum negative error during field test is also specified and should not be larger than - 3 km/h, or - 3 % at speeds above 100 km/h. The scope of the document is also not limited to radar equipment only, but it is extended to all speed measuring devices. Although the metrological legislation of individual countries might be a bit stricter, the compliance with the OIML's recommendations is a good starting point for any vehicle speed measuring device.

Chapter 4

Stereoscopic measurement and calibration methods

This chapter presents the most relevant state-of-the-art methods that exploit stereo camera pairs for vehicle speed measurement and long-distance calibration.

4.1 Vehicle speed measurement methods

Stereovision-based methods for vehicle speed measurement usually assume synchronized cameras previously calibrated using already established methods. Their most important parts deal with feature point selection and correspondence search. Only three papers that use a stereo camera pair for vehicle speed measurement exist to the extent of my knowledge. All three works are from recent years which signifies a growing interest in this type of vehicle spot speed measurement.

4.2 Vehicle Speed Estimation Using Cascade Classifier and Sub-pixel Matching

Jalalat et al. [31] use a vertical stereo setup with 1.2 m baseline that was pre-calibrated using a chessboard pattern [78]. The algorithm that processes the stereo images consists of six steps: region of interest selection, background/foreground segmentation, vehicle detection, vehicle tracking, distance measurement, and speed measurement.

In the first step, the region of interest is manually specified by the user, and its actual dimensions are measured and stored for further processing. The region of interest delimits an area on the road in which the vehicles are processed, and it has, due to the perspective, a trapezoid shape. The bounding rectangle is constructed around the delimited area, and only the area within the bounding box is processed in the following steps.

The second step separates the background and the foreground objects in the area of interest. In order to fulfil real-time constraints, the background image is updated using a moving average. From each new frame, a background image is subtracted, and the result is thresholded. The threshold value is computed using an average background brightness which increases the robustness of the algorithm with respect to the changes in illumination.

The foreground mask is used in the third step to reduce the amount of falsely detected vehicles. The vehicles are detected in each frame using a Viola-Jones cascade classifier
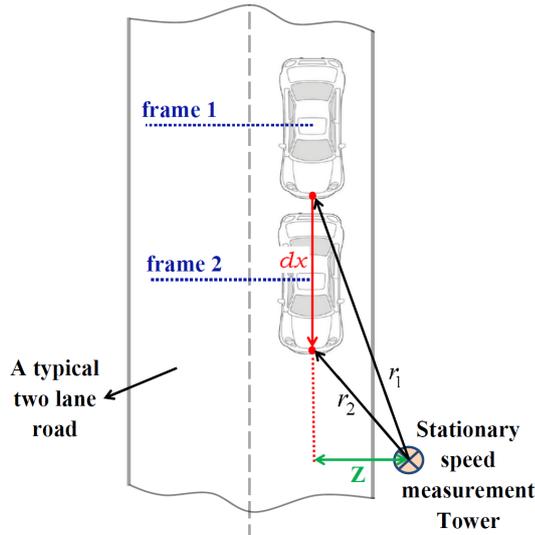
Figure 4.1: Geometry of distance measurement for a vehicle in two successive frames [31].

which has been trained on an extensive dataset that contains almost all types of vehicles on Iran highways.

The detected vehicles are tracked in the fourth step of the algorithm. Kalman filter that assumes linear velocity and constant acceleration of vehicles is utilized for tracking. Each tracked vehicle is assigned an identification number which is preserved until the vehicle leaves the area of interest.

In the fifth step, the vehicle distance is estimated. First, the stereo images are undistorted and rectified. Then, the correspondence search is performed, and the disparity map is computed. The disparity map is computed only for the detected vehicles and not for the whole region of interest. The correspondence search area is further limited to only the bottom third of the cascade detection bounding box. The area is then divided into three parts that are matched separately. The correspondence search is performed along the vertical epilines. The location of the corresponding point is initially estimated by cross-correlation in two times the upsampled resolution and then subsequently refined by exploiting the single-step DFT technique, which pinpoints the corresponding point location with sub-pixel precision.

In the final step, the vehicle speed is computed. First, the vehicle distance from the camera is triangulated using the disparity. Two distance in two successive frames are needed to compute the travelled distance using the Pythagorean theorem as

$$d = \sqrt{r_1^2 - Z^2} - \sqrt{r_2^2 - Z^2}, \tag{4.1}$$

where r_1 and r_2 are the measured distances from the camera in two frames and Z is the perpendicular distance from the camera to the direction of vehicle motion (see Fig. 4.1). The vehicle speed is then easily calculated by dividing the travelled distance by the time that passed between the two frames.

For the evaluation, they recorded five different video sequences that were used to compute the speed measurement error of their method with respect to the ground truth values provided by a laser speed measuring device. Overall the recorded dataset contains 441 vehicles for which the reference speeds are provided. They report the speed measurement

25

error of their approach as an arithmetic mean of the absolute error percentages compared to the reference measurements made by Fama Laser III (VHT-507/DVRM). The worst mean percentage error is 3.3 %, and the best mean percentage error is 1.8 %. The absolute error is not reported. Additionally, they report the percentages of measurements that are out of the ± 6 % range. The out of range percentages range from 1.4 % to 2.5 %.

4.3 Vehicle Speed Estimation Using Extracted SURF Features from Stereo Images

El Bouziady et al. [14] use a horizontal stereo setup pre-calibrated in a laboratory. The baseline of the stereo camera pair is not stated. The resolution of the captured images is 1400 x 1024 pixels. The video streams are sent via an Ethernet connection at a rate of 32 stereo frames per second to an FPGA where they are preprocessed and then forwarded to an industrial computer for further processing.

The FPGA performs background subtraction on the captured images. The foreground mask is obtained by first, subtracting the static background image from the current frame; then, the result is filtered and binarized using a fixed threshold. The binary mask is then subjected to the morphological operations that transform the white areas of the image to convex blobs. The blob contours are detected and their bounding rectangles constructed. The bounding boxes contain the detected vehicles. No tracking algorithms or algorithms that preserve the context between the two consecutive frames are described.

After the vehicle detection step, the sparse disparity map is constructed. The SURF feature detector and descriptor is used for the correspondence search. The searched area is limited to the feature-rich regions (e.g. license plate, manufacturer logo) within the vehicle bounding box. From the corresponding points, the disparities are computed, and the distances from the camera to the points on the vehicle surface are triangulated. Each distance r is then converted to the horizontal distance d using camera inclination angle α as

$$d = rcos(\alpha). \tag{4.2}$$

The horizontal distances from two consecutive frames and the camera refresh rate are used to compute the list of speeds from which the average speed is calculated.

For the evaluation, the dataset that consists of 12 passes of 6 vehicles in two different speed sessions was recorded. The normal speed session recorded vehicles that travelled with the speed in the range from 60 to 90 km/h. The high-speed session recorded vehicles that travelled with the speed in the range from 90 to 120 km/h. The reference speed values were measured using GPS. The mean squared error of 1.67 km/h is obtained for the normal speed session. For the high-speed session, the mean squared error of 2.33 km/h is obtained. The maximum absolute error was 2 km/h common for both sessions.

4.4 Vehicle Speed Measurement Based on Binocular Stereovision System

Yang et al. [76] use a horizontal stereo setup calibrated using Zhang's method [78]. Their cameras are equipped with a CCD sensor with a maximum resolution of 1288 x 964 pixels and connected to the computer through the USB 3.0 interface. The cameras captured ten frames per second. The length of the baseline is not specified. The processing algorithm

is divided into three steps: vehicle feature detection, vehicle tracking and stereo matching, and speed and trajectory measurement.

The inputs to vehicle feature detection are video streams from the two views, and the outputs are positions of detected license plates. The license plates are detected using an optimized Single Shot Detector network. The network structure was optimized to reflect the statistical distribution of the expected license plate sizes. The trained network achieves 97.8 % accuracy rate and is able to process 5.4 frames per second.

In the next step, the detected license plates are tracked, and the stereo matching is performed. For both tasks, the SURF features extracted from the license plates are utilized. As the features are invariant to changes in scale, they can be used to maintain the tracking context in consecutive frames of the monocular videos. And because they are also invariant to rotation and translation, they can be used for matching the license plates between both views. The extracted corresponding point pairs are submitted to the final step.

In order to increase the speed measurement efficiency, the amount of corresponding point pairs is reduced by filtering out the pairs whose both points do not lie in a circular area positioned in the middle of their license plates. The remaining points are triangulated, and their world position is obtained. Because the origin of the coordinate system coincides with the coordinate system of one of the cameras, the distance between two 3D points $p_i(x_i, y_i, z_i)$ and $p_j(x_j, y_j, z_j)$ can be easily computed as

$$d_{ij} = |p_j - p_i|. \tag{4.3}$$

For each point i, the distances to all other retained points are computed. The mean value μ and standard deviation σ of the distances are computed, and a Z-score for the point is computed as

$$Z_{ij} = \frac{d_{ij} - \mu}{\sigma}. \tag{4.4}$$

The points with an absolute Z-score greater than one are eliminated, and from the remaining points, the one that is closest to the centre of the license plate is selected as the exact spatial location of the target car in the current stereo frame pair. Given the known frame rate of the stereo camera pair, the speed of the vehicle can be then easily computed from its location in two frames.

For the evaluation, a dataset that contains four vehicle passes with a speed range of between 20 and 50 km/h was recorded. The reference speed values were obtained from a P-Gear P-510 professional satellite speed meter. The mean speed measurement error is 0.02 km/h, the mean squared error is 0.42 km/h, the maximum negative error is -1.6 km/h, the maximum positive error is 1.1 km/h, and the maximum percentage error of 3.8 %.

4.5   Long-distance calibration methods

In order to provide reliable measurements, the stereo camera pairs need to be calibrated. The calibration process of traffic surveillance cameras has specific challenges and complications. The traditional approaches that involve calibration patterns (e.g. [78, 51]) cannot be easily used after the cameras were installed because cameras are usually focused on a long shot and mounted on hardly accessible locations. The calibration in laboratory conditions prior to installation is also not suitable because the extrinsic parameters and focus are likely to change during the transportation or the installation. This section presents two existing methods for calibration of a stereo camera pair that is focused on a long distance.
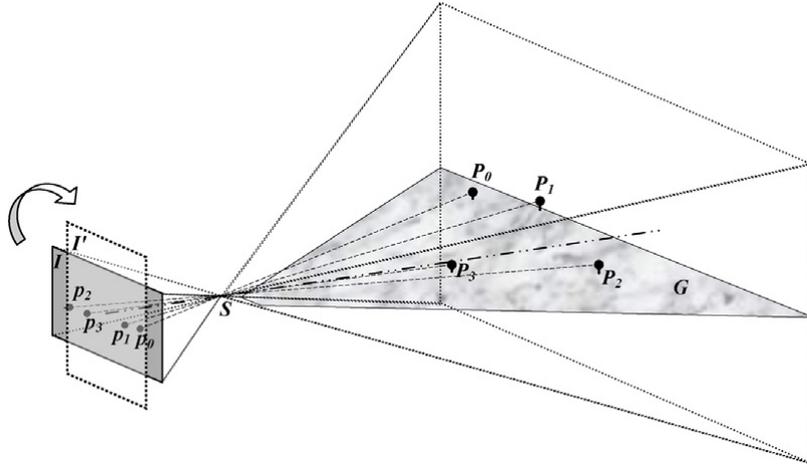
Figure 4.2: Two images captured by a camera in horizontal and vertical orientations containing four calibration points located at the same height as the camera are needed for a full calibration. [62]

4.6 A Camera Calibration Method for Large Field Optical Measurement

Shang et al. [62] presented a method that deals with a task of calibrating the cameras that are focused on a long distance with a large field of view. Their camera model consists of 15 parameters: 4 internal (position of the principal point (c_x, c_y) and the focal lengths f_x and f_y), 5 lens distortion, and 6 external (optical centre coordinates and rotation angles). The calibration method consists of two steps. First, the initial values for the camera parameters are obtained and then calculated precisely.

Initially, the principal point is assumed to be positioned in the centre of the image, and the values of the lens distortion parameters are all zeros. For the calculation of the initial values of focal lengths and rotation angles, the positions of at least four control points that lie on the same plane as the optical centre and the camera principal point have to be known. The positions of the control points, as well as the initial value for the position of the camera optical centre, are measured using a total station. Using two of the control points, the initial value for the focal length f_x can be computed. The initial value for the focal length f_y is set equal to the f_x. All four control points are needed to compute the initial values for the rotation angles.

The precise values of the camera parameters are calculated using bundle adjustment optimization, which minimizes the reprojection error of the control points. To control all parameters perfectly, two images have to be taken. One, when the camera is positioned normally, and the second one with the camera rotated 90 degrees along the optical axis (see Fig. 4.2). Alternatively, for the cameras with square pixels, negligible tangential distortion, and known vertical position of the principal point, only a single image would suffice.

The accuracy of the calibration method was evaluated by comparing the computed positions of six control points with the ground truth positions measured by a total station. The points were at the distances between 50 and 70 m from the stereo camera pair whose baseline was 30 m. The mean inter-point distance error was 0.008 m with a standard deviation of 0.04 m.
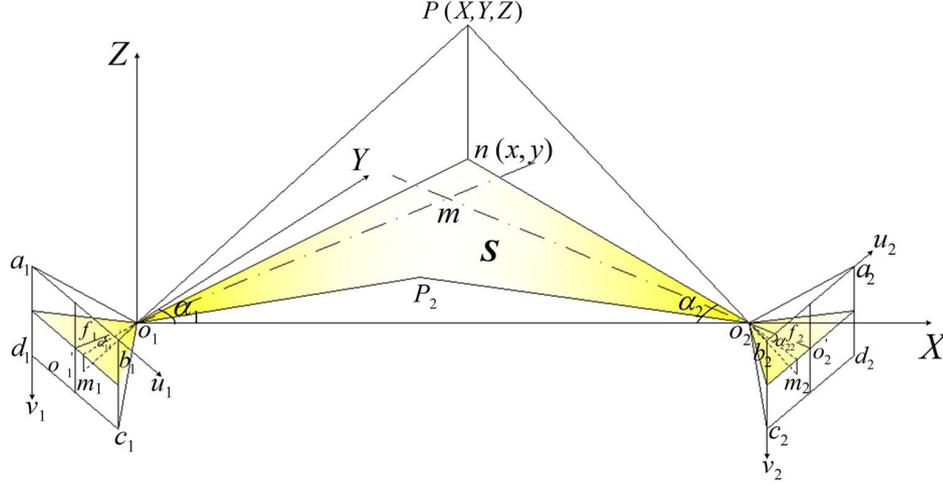
Figure 4.3: Simplified camera model requires a special positioning of the cameras where the cameras' epipolar planes are coincident (plane S) with the XY plane of the world coordinate system [69].

4.7 A Camera Calibration Method for Large Field Vision Metrology

Tian et al. [69] presented a two-step method with a similar approach as Shang et al. They used a simplified camera model which requires special positioning of the two cameras. The cameras have to be positioned in such a way that their epipolar planes are both coincident with the world coordinate system (see Fig. 4.3). The special positioning simplifies the triangulation process. The calibration process consists of two steps: camera adjustment and estimation of the parameters.

First, the cameras need to be adjusted, so their position and orientation suit the simplified camera model. In order to do this, two points at the same height as the camera optical centre are required. The orientation of the camera is then adjusted, so the projections of the two points appear in the longitudinal centre of the CCD sensor. After the adjustment, the extrinsic parameters are obtained using third-party tools.

For the estimation of the internal parameters, the third point is required. The additional point should be positioned at the same height as the other two points and thus lie in the same plane. From the geometrical relation between the projected points, the internal parameters are calculated.

For evaluation, the task of measuring the position of a UAV (unmanned aerial vehicle) was utilized. The world coordinates of the calibration points and the cameras were measured using a total station. The UAV flew at a distance of approximately 90 m from the stereo camera pair whose baseline was 79.88 m. Its position was tracked by a DGPS with a positional accuracy of 0.008 m. Eight reference positions were selected randomly and compared to the measured values. The mean inter-point distance error was 0.0062 m with the standard deviation of 0.0332 m.

Chapter 5

Proposed method for vehicle speed measurement

The current state-of-the-art vehicle speed measurement methods do not achieve the results that are within the ranges required for the autonomous measurement devices by the OIML Recommendation. This chapter presents a novel method for vehicle speed measurement using a stereo camera pair that is used in the following chapter to experimentally test the following hypothesis:

It is possible to measure the speed of vehicles using a stereoscopic method with the average error within ± 1 km/h, the maximum error within ± 3 km/h, and the standard deviation within ± 1 km/h.

The error ranges are based on the field test requirements of the OIML Recommendation and on the assumption of the Normal error distribution. Because the proposed measurement method requires a calibrated stereo camera pair and because the calibration of traffic surveillance cameras is a challenging task, this chapter also describes a novel method for stereo camera pair calibration that is suitable for traffic surveillance applications. Both methods were submitted to peer-reviewed journals [54, 53].

5.1 Vehicle speed measurement proposal

The equipment setup suitable for the proposed method consists of a synchronized and calibrated pair of two identical cameras with the same focal length. The proposed method exploits a stereo camera pair already calibrated with known calibration features (calibration error) and relies on existing algorithms of license plate detection. The performance of the license plate detection algorithm affects only the fact whether the speed measurement is performed at all, but does not affect its precision. First, the vehicles passing in front of the stereo pair in the series of frames are localized using their license plate co-ordinates. Consequently, the vehicle position is triangulated in the series of stereo images forming a trajectory using the information known about the stereo setup and the calibration information. Finally, once the trajectory and its individual points are known, the speed (and also acceleration along the trajectory) is computed. An overview of the proposed method is shown in Fig. 5.1.
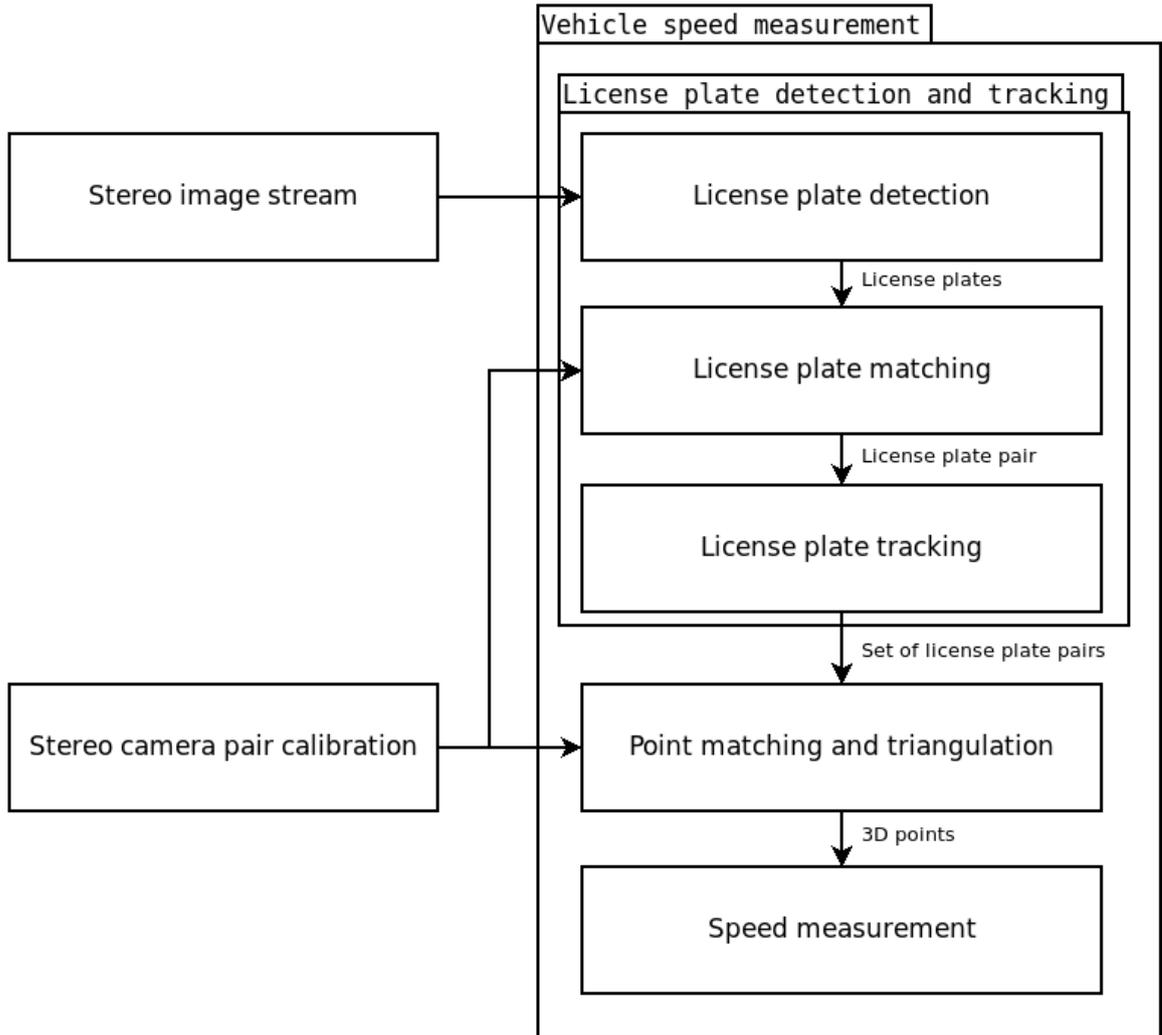
Figure 5.1: Overview of the proposed method. Vehicle trajectory is represented using a set of license plate pairs that are extracted from input stereo images. Several points are triangulated along the trajectory using known calibration parameters. Model of vehicle motion is fitted to the triangulated points in order to measure the vehicle speed.

License plates detection and tracking

The proposed method assumes that each vehicle has a license plate that is firmly attached to its body at a clearly visible place. This assumption reduces the task of vehicles detection to the task of the license plates detection, which is considered to be much easier due to the standardized appearance of the license plates of a given country. Although the appearances of the license plates differ among countries, they are usually similar enough, so that a detector trained on the license plates of one country is able to detect the license plates from other countries. In the first step of the method, the license plates of moving vehicles are detected and tracked.

The algorithm utilizes existing WaldBoost [63] detector with LBP features that was trained to detect license plates with a size of approximately 90x24 pixels and without or with very small rotation and perspective distortion. The detector works best for cameras

placed on a gate or a bridge above the road looking directly against or with the direction of traffic. Alternatively, the cameras can be placed on a pole on the side of the road, but the angle between the view direction and traffic flow should be kept small. The output of the detector is the top left coordinate and the size of a rectangle that contains the found license plate. License plates in the left and the right images are detected separately.

The detector finds the license plates in both images. Each detected license plate in the left camera image is assigned a corresponding license plate from the right camera image, if possible. The matching algorithm asserts that the matching license plates are approximately the same size, located at approximately the same position in both images and meet the epipolar constraint. The corresponding license plates then form a license plate pair. For the depiction of detected license plate pairs, see Fig. 5.2a.

For each license plate pair, the method needs to decide whether it belongs to a new vehicle or to a vehicle that was seen before. This can be done in numerous ways, for example, using license plate correlation, OCR, optical flow or Kalman filtering. In the proposed method, the latter is used, that is, license plate tracking using the Kalman filter based on the constant acceleration model. The constant acceleration model approximately corresponds to the motion of the license plates in the images. This approach was chosen because it provides a good trade-off between speed, accuracy, and complexity. The approximate predictions supplied by the Kalman filter are enough to perform the license plate re-identification based on its detected and predicted positions in the image and enough to maintain the tracking context in frames where the detection has failed. The tracking is implemented using re-detection and re-identification of license plates in each frame. The tracking algorithm maintains a Kalman filter for each vehicle that is currently passing in front of the cameras. Each license plate pair position is checked against all current Kalman filters' predictions. If its position is close enough to some predicted position, the algorithm assigns the checked license plate pair to a set of license plate pairs of the tracked vehicle and updates the filter accordingly. Otherwise, new vehicle tracking is initialized. When the vehicle passes out of view, the algorithm stops the tracking and starts processing vehicle's set of license plate pairs. A vehicle with its set of license plate pairs is shown in Fig. 5.2c.

Point matching and triangulation

After vehicle localization, the method triangulates its passage throughout the scene. First, from its set of license plate pairs, the matching algorithm selects the pair which contains the largest license plate images. Then, the algorithm takes the left license plate image of this pair and uniformly samples nine points on it. Around the sampled points, small rectangular regions of interest are constructed (see Fig. 5.3a).

Next, the matching algorithm takes another license plate pair from the set and matches the regions of interest to its left license plate image in order to obtain the points that match the sampled points. To achieve sub-pixel accurate matches, the regions of interest, as well as the template license plate image, are scaled and smoothed prior to template matching. For regions of interest, a scale factor of ten is used. Since the template license plate image is smaller than the sampled license plate image, the scale factor used to scale the template image should be greater. Because the exact value of this scale factor is unknown, the algorithm tries several scales within a reasonable range and chooses the one with the highest similarity score. The normalized cross-correlation is used as a similarity metric.

The matched points and the sampled points are used to compute homography transformation between the two left license plate images. This homography transformation is

further refined using enhanced correlation coefficient maximization [15]. Finally, the sampled points are transformed using the fine homography in order to obtain fine point matches. These fine point matches represent the sub-pixel accurate left image positions of sampled points in a different time. The algorithm repeats the matching process for the rest of the left license plate images from the set and essentially tracks the sampled points along the vehicle trajectory in left camera images with sub-pixel precision.

So far, the method extracted several points in left camera images and in order to perform triangulation, it needs to identify their correspondences in right camera images. A similar procedure as in the case of finding correspondences among left camera images is employed. First, the algorithm takes a license plate pair from a set of vehicle license plate pairs and selects the left license plate image. Then, it uniformly samples nine points in it, constructs the regions of interest, matches them to the right license plate image template, computes rough homography and refines it as it did before. Finally, the algorithm computes the fine point matches using sampled points from the left license plate image and fine homography transformation. The reference image and template images that are warped using the rough and fine homographies are shown in Fig. 5.3b, 5.3c, and 5.3d. The process is repeated for the rest of the license plate pairs from the set and, as a result, nine stereo point correspondences for each license plate pair are obtained.

Given stereo point correspondences and known internal and external stereo camera pair parameters, the method triangulates the 3D positions of points using existing algorithms. It uses the well known Linear-LS method [28].

Speed computation

To compute the average speed of a passing vehicle, the method utilizes the triangulated positions from the previous step and their timestamps. First, it corrects the triangulated positions by projecting them onto a common plane obtained as a least-square fit through the triangulated points with outliers removed using RANSAC [20]. The corrected points are then subdivided into nine sets in such a way that the same license plate points with different timestamps belong to the same set. These sets are processed separately. The points from a single set, together with their timestamps, serve as an input to a model describing the vehicle motion throughout the scene. The algorithm assumes that the vehicle is moving with constant or zero acceleration. This type of motion can be described by the following equation

$$p_i = p + v * \Delta t_i + \frac{1}{2} * a * \Delta t_i^2, \tag{5.1}$$

where p_i is a co-ordinate of the current position of a corrected triangulated license plate point in time i; p is a co-ordinate of initial license plate point position, v is a vector of initial speed, a is a vector of acceleration, and Δt_i is the time difference between the current and initial positions. The positional (p_i) and time (Δt_i) data are inserted to the model and a system of $N*3$ linear equations is constructed, where N is a number of triangulated positions available. This system is usually overdetermined ($N > 3$), and it can be formulated as

$$A * x = b \tag{5.2}$$

where A is a $N * 3$ matrix

$$A = \begin{bmatrix} 1 & \Delta t_0 & \frac{1}{2} * \Delta t_0^2 \\ 1 & \Delta t_1 & \frac{1}{2} * \Delta t_1^2 \\ ... & & \\ 1 & \Delta t_{N-1} & \frac{1}{2} * \Delta t_{N-1}^2 \end{bmatrix}; \tag{5.3}$$

b is a $N * 3$ matrix of triangulated positions

$$b = \begin{bmatrix} x_0 & y_0 & z_0 \\ x_1 & y_1 & z_1 \\ ... & & \\ x_{N-1} & y_{N-1} & z_{N-1} \end{bmatrix}; \tag{5.4}$$

and x is a $3 * 3$ matrix of unknown vectors p, v and a

$$x = \begin{bmatrix} p_x & p_y & p_z \\ v_x & v_y & v_z \\ a_x & a_y & a_z \end{bmatrix}. \tag{5.5}$$

This system can be solved using various methods for solving linear least square systems such as SVD decomposition. As a result, an initial point position, initial speed, and acceleration are obtained. From the obtained values, the average speed on the recorded track can be computed as:

$$v_{avg} = \frac{||A_{N-1} * x - p||}{\Delta t_{N-1}} \tag{5.6}$$

where A_{N-1} is the last row of matrix A. In order to make the computation of the average speed more robust to errors in position triangulation, the RANSAC [20] based approach to remove triangulated positions outliers is employed. The algorithm repeats the average speed computation process for each of the nine sets with triangulated license plate points and, in the end, nine average speeds are obtained. Finally, the median of the computed average speeds is selected as the average vehicle speed.

5.2 Stereo camera pair calibration proposal

The proposed method is suitable for the calibration of a synchronized stereo camera pair device that will look over a road section with passing vehicles. In order to calibrate the device, both camera matrices from Eq. 2.14 need to be estimated. The method splits the sought parameters to two groups. The first group consists of the parameters that can be estimated prior to the device installation. Their values are determined during an off-site calibration. The second group contains parameters that have to be estimated after the installation of the device. These are the external camera parameters and focal lengths. Their values are determined during an on-site calibration. An overview of the proposed calibration method is shown in Fig. 5.4.

Off-site calibration

The off-site calibration can be performed in a laboratory for each camera separately using the already established methods such as Zhang's [78]. Its goal is to find the values for the parameters that do not depend on the relative position of the two cameras or on the installation site. These parameters should stay the same for the whole time of installation

and should not change due to the external influences such as vibrations or heat. The parameters that are estimated during the off-site calibration include: the lens distortion parameters (x_c, y_c, K_1, K_2, ...), the position of the principal point (c_x, c_y), the skew (s), and the numbers of pixels per a distance unit (m_x, m_y). The precise values for m_x and m_y parameters can usually be found in a camera sensor datasheet if available.

On-site calibration

The on-site calibration can be performed after the cameras have been installed, and their optics have been focused on the incoming vehicles. Its goal is to find the values for the parameters that depend on the relative position of the two cameras or on the installation site. The parameters that are estimated during the on-site calibration include the position and the orientation of the second camera with respect to the first camera ($[R|t]$) and the focal lengths of both cameras (f, f'). The focal lengths depend on the distance at which the incoming traffic is recorded and should not change in time. The relative pose, on the other hand, will change in time due to the external forces and therefore, the device needs to be re-calibrated, or the change should be compensated in some way.

The on-site calibration requires two key conditions: the parameters estimated during the off-site calibration and the calibration vehicles. The calibration vehicles are recorded as they pass in front of the stereo camera pair and their initial speeds at the start of a measured road section and average accelerations on the measured road section are captured and paired with their record. In each record, the calibration vehicle is localized using existing algorithms of license plate detection and tracking. After that, several points are extracted from the images of the first camera along the trajectory of the localized vehicle, and their correspondences are identified in the images of the second camera. Finally, the set of corresponding points coupled with the speed and acceleration measurement for each calibration vehicle and the off-site calibration parameters are used for the estimation of on-site calibration parameters.

The goals of the first two steps of the calibration method are to localize the calibration vehicles and extract the corresponding points along their trajectories from their license plates. These steps are the same as the license plate detection and tracking and point matching steps from the vehicle speed measurement method described above. They are handled by almost the same algorithms. They differ in one thing. Because the information about the epipolar geometry is not yet known, its constraints cannot be exploited for limiting the image areas for searching or matching. The following section describes the last step of the on-site calibration process.

External parameters and focal lengths estimation

The external parameters and the focal lengths are estimated by minimization of error expressed as a difference between the distances that the calibration vehicles travelled and the distances that can be computed from their recordings. The error formula is

$$e = \sum_j^N \sum_i^{M-1} \sum_k^O (w_{i_k}^j (d_{i,i+1}^j - \|X_{i_k+1}^j - X_{i_k}^j\|))^2, \tag{5.7}$$

where j iterates over the number of calibration vehicles, i iterates over the number of license plate pairs detected during the vehicle passage, k iterates over the number of corresponding point pairs that are extracted from each license plate pair (nine in this case), $d_{i,i+1}$ is a distance that a vehicle travelled between two subsequent detections i and $i+1$, and X_{i_k}

are triangulated corresponding point pairs. The w is a weighting factor which helps us maintain the epipolar geometry.

Note that the algorithm expects that the vehicle license plates in two subsequent detections are close to parallel (i. e. the distances between the triangulated points on the first license plate and their corresponding points on the second license plate are the same). This expectation is reasonable because the measured road section is usually short, and the vehicle trajectory straight.

The value of the weighting factor w is based on the point distances to epilines. For a point from the first camera license plate, an epiline in the second camera image is computed using the Eq. 2.10. Then, a distance $l_{i_k}^j$ between its corresponding point and the epiline is computed. The same is performed for the point in the second image (i. e. an epiline is computed in the first image and then the distance $l'^j_{i_k}$ to its corresponding point in the first image). The value of w is then computed as

$$w_{i_k}^j = max(1.0, l_{i_k}^j, l'^j_{i_k}). \tag{5.8}$$

The weighting factor penalizes the estimates for which the distances to epipolar lines are high but does not benefits those estimates for which the distances to epipolar lines are low. That is, only the estimates that break epipolar geometry are penalized, and the quality of those that keep the geometry is decided only by the accuracy of the scene reconstruction that they provide.

The distance that the calibration vehicle travelled between two detections can be computed from its recorded initial speed and average acceleration as

$$d_{i,i+1} = v_i \Delta t_{i,i+1} + \frac{1}{2} a \Delta t_{i,i+1}^2, \tag{5.9}$$

where v_i is the vehicle initial speed at the time t_i, a is the average vehicle acceleration on the measured road section, and $\Delta t_{i,i+1} = t_{i+1} - t_i$ is the difference in times at which the frames i and $i+1$ were taken. The immediate speed v_i can be easily computed from vehicle initial speed v_0 at the start of the measured road section as

$$v_i = v_0 + \frac{1}{2} a(t_i - t_0). \tag{5.10}$$

The points $X_{i_k}^j$ and $X_{i_k+1}^j$ are triangulated using the well known Linear-LS method [28] that is based on Eq. 2.3. The minimization itself is done using the Levenberg-Marquardt algorithm.

In order to achieve optimal results, the tracks of calibration vehicles should cover the area of interest as uniformly as possible. For example, if the goal is to perform measurements on vehicles in two lanes, the tracks of the calibration vehicles should also cover both lanes.

(a)

(b)

(c)

Figure 5.2: License plates detection and tracking. (a) Detected license plates in the image from the left camera and their matching counterparts in the right camera image forming a license plate pairs. (b) The license plates search areas in both images were constrained by foreground masks constructed using background subtraction. The right camera image search area was further constrained using epipolar geometry. (c) A vehicle with its set of all license plate pairs.
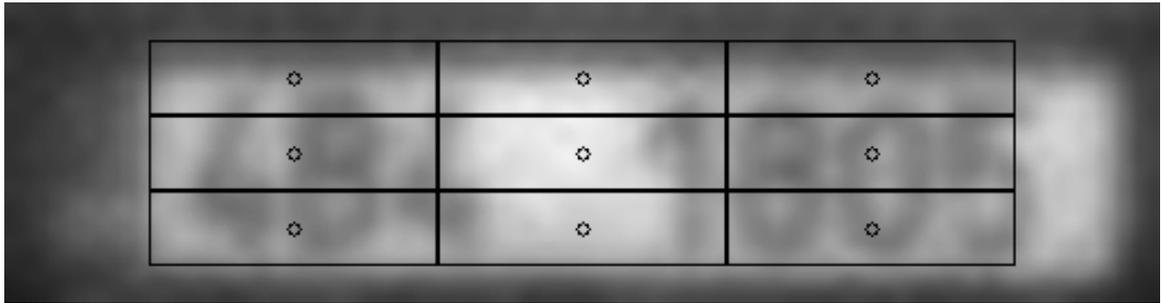
(a)
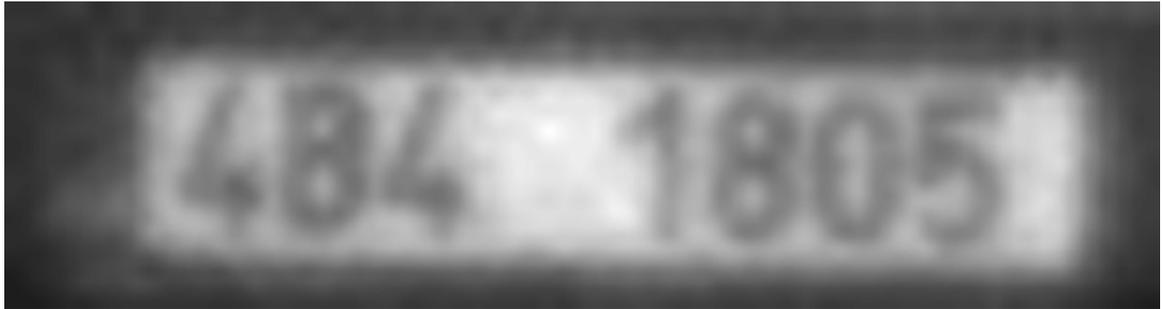
(b)
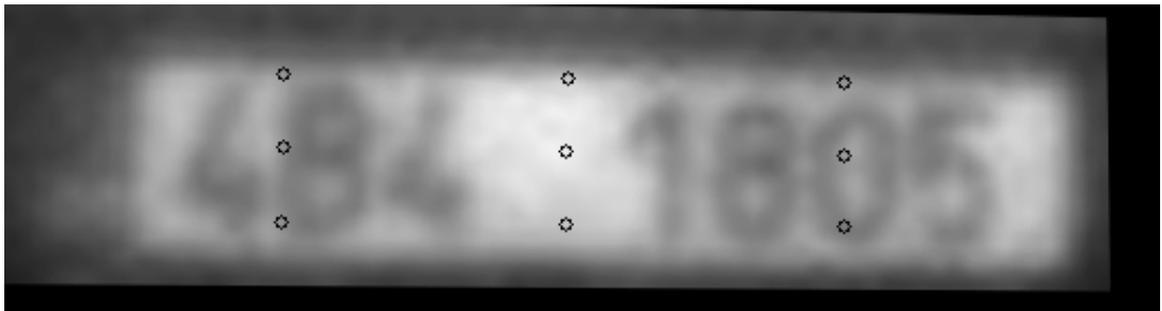
(c)
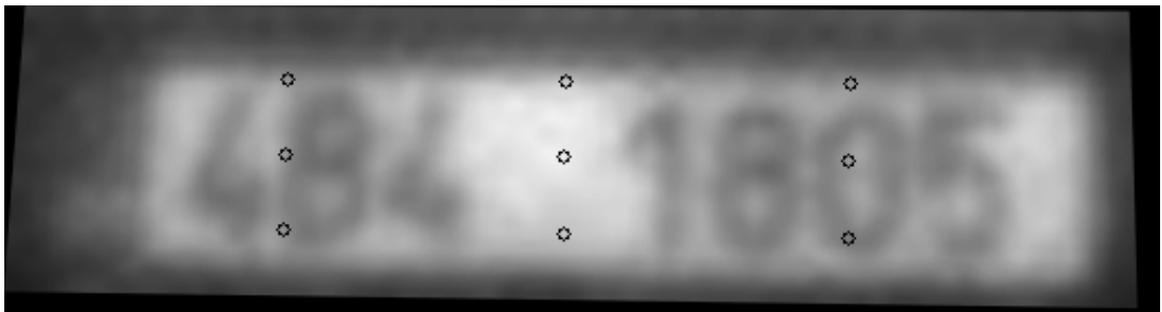
(d)

Figure 5.3: Point matching and triangulation. (a) Template left image with uniformly sampled points and regions of interests constructed around them. (b) Reference right image. (c) Template image with sampled points warped using rough homography computed from sampled points and their counterparts obtained by template matching the regions of interest to the reference image. (d) Template image with sampled points warped using fine homography computed using enhanced correlation coefficient maximization initialized with rough homography.
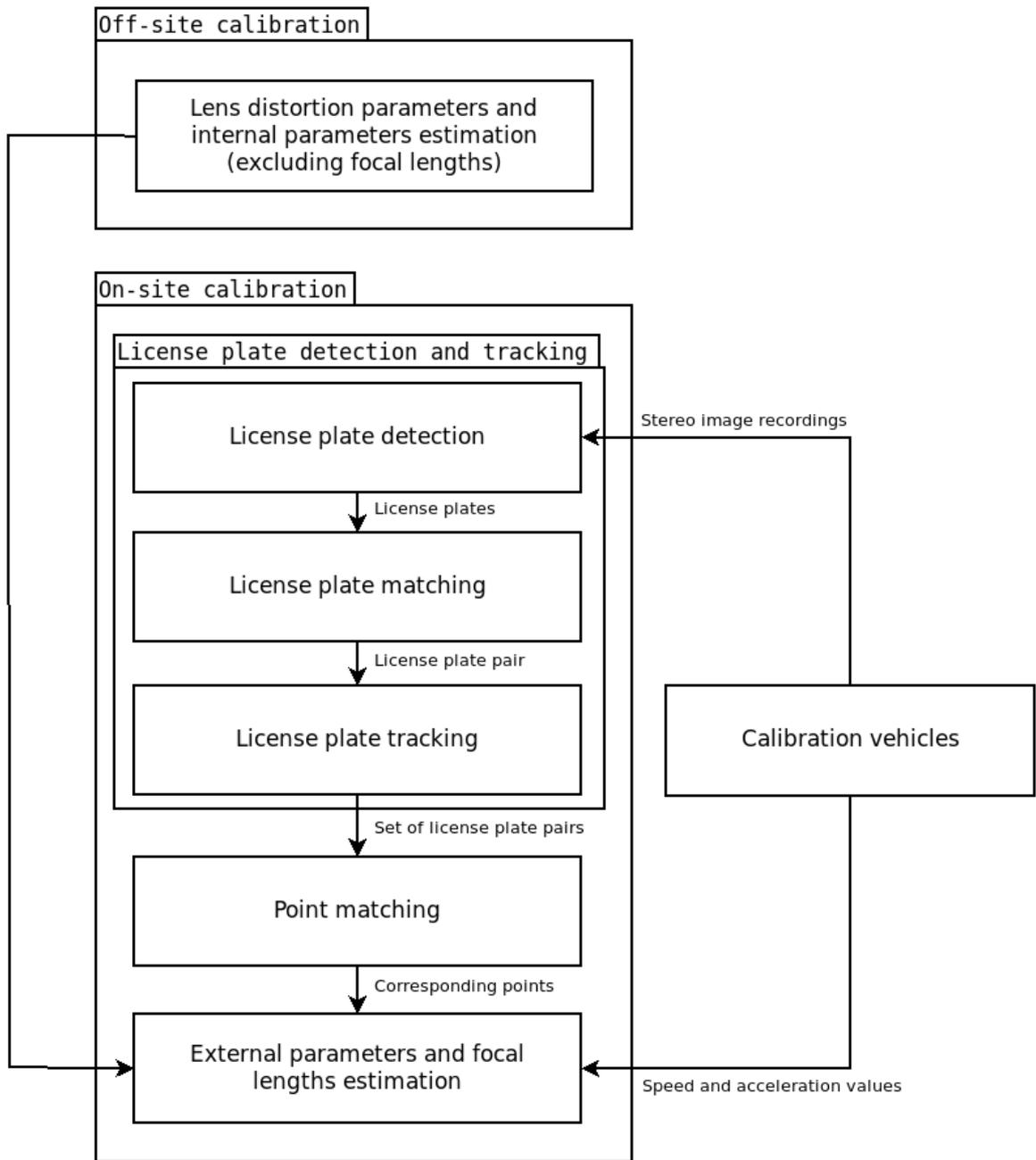
Figure 5.4: Overview of the proposed calibration method. Off-site calibration is performed first. For the on-site calibration, the calibration vehicles with known speed and acceleration are recorded. Their trajectories represented by a set of license plate pairs are extracted from the input stereo image stream. Several corresponding points are identified along the trajectories. The corresponding points, the known speed and acceleration of the calibration vehicles and the off-site calibration results are used to complete the on-site calibration.

Chapter 6

Experimental results of the proposed method

This chapter describes experiments that evaluate the novel measurement method with exploitation of the newly proposed calibration approach. The experiments utilize prototype hardware that is introduced in the first section of this chapter. Using the prototype hardware, a dataset that contains passing vehicles is recorded. The dataset and the reference data for the experiments are described in the second section.

The first experiment tests the presented hypothesis by evaluating the proposed method for vehicle speed measurement. The design of the experiment is similar to the design of the metrological field tests. In this experiment, the speed measurement error is computed from the speeds that are measured by the proposed method and the reference speeds. The results are compared to the existing stereo-based methods described in Chapter 4.1. The hypothesis holds if the maximum error, the mean error, and the standard deviation of error are within the ranges specified in Table 3.1.

The second experiment evaluates the proposed stereo camera pair calibration method. In this test, the stereo camera pair is calibrated by the proposed method and utilized for the measurement of distances that vehicles travelled between two consecutive frames. The measured distances are compared to the reference distances from the dataset, and the distance measurement error is reported. The results are compared to the existing stereo camera pair calibration methods described in Chapter 4.5.

6.1 Prototype hardware

The prototype hardware (see Fig. 6.1) consists of two custom made cameras mounted parallelly on a 1 m long aluminium profile placed on a sturdy tripod. The cameras are fitted with PYTHON 1300 global shutter CMOS image sensors and 35mm fixed focal length lens, which is positioned is a such a way that its principal axis is perpendicular to the sensor plane and intersects it at a sensor centre. The image sensors have 0.0048 mm x 0.0048 mm square pixels and provide monochrome 1280x1024 px images. Raw image data is streamed at a rate of 20 frames per second through a gigabit ethernet switch to a computer where the images are JPEG compressed and stored for further processing. The shutters of the cameras are synchronized using an external trigger with one camera being the master who sends the trigger signal to the second camera. The cameras and the switch are supplied power from two 6400mAh LiPo batteries attached to the profile.
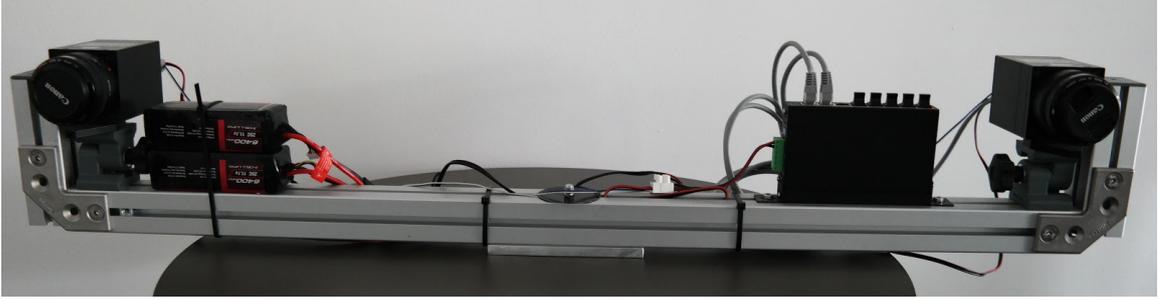
Figure 6.1: The stereo camera pair setup. Two cameras, ethernet switch, and two batteries mounted on an aluminium profile.
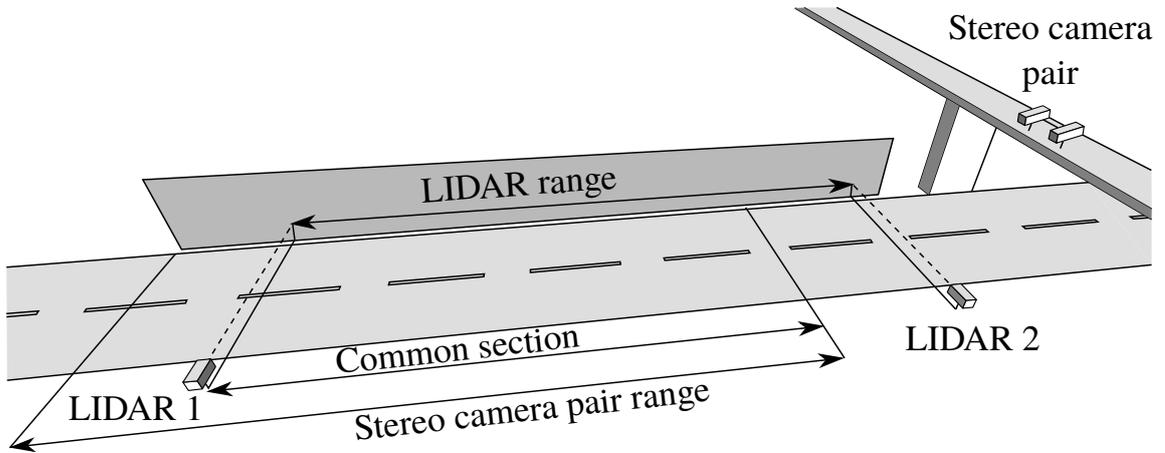
Figure 6.2: Schematic drawing of the relative positions of the sensors, their ranges, and their common area.

6.2 Dataset and reference data

For the purpose of evaluation, a dataset was recorded using the above-mentioned hardware. The whole dataset was recorded during a single session lasting approximately 40 minutes. During this session, 698 vehicle passes in two lanes were recorded. The left lane (from the point of view of cameras) is visible in full on both cameras while the right lane is only partially visible. The camera setup was placed on a footbridge across the road looking from above towards the incoming vehicles. Out of the 698 recorded vehicle passes, 44 were used for the on-site calibration of stereo camera pair and the remaining 654 were used for the testing.

To obtain the reference data, the same approach as Sochor et al. [64] was employed. Two LIDARs (LaserAce®IM HR 300) were placed at the same height parallelly to each other and perpendicularly to the street. The distance D between the LIDARs was 28.05 metres, and they were synchronized by the GPS time (Leadtek LR9540D). The distance and time data from both LIDARs were logged and processed separately. From the logged data, the processing algorithm calculated for each vehicle its immediate speed when entering the first and the second laser, its average speed on the distance D, its average acceleration, and
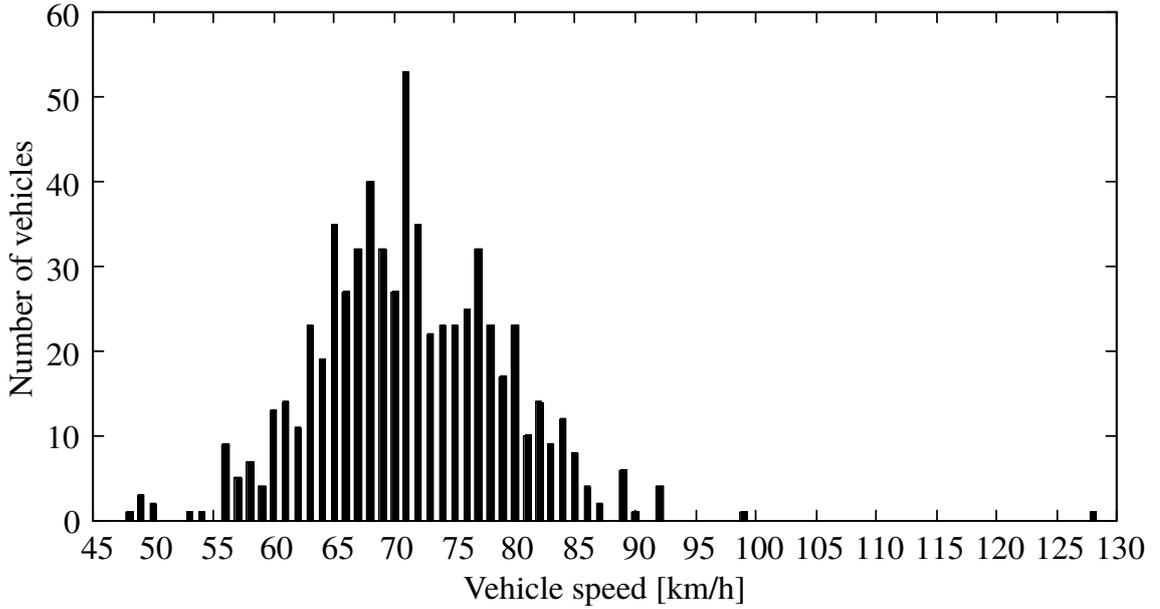
Figure 6.3: Histogram of reference average speeds.

its length. For more detail about the measurement process, reference data calculation, and the discussion of measurement error, see Sochor et al. [64].

Vehicle speed measurement

The reference data for the vehicle speed measurement experiment are the average speeds obtained from LIDARs. The reference average speeds are compared to the average speeds measured by the proposed method. However, these values are comparable if and only if they were both measured over the same section of the road. As the section of the road covered by the two LIDARs and the section of the road in the view of the stereo camera pair do not fully overlap, the reference and the measured average speeds are not directly comparable and need to be adjusted so that the road section, where the speed is measured, is common for both setups. The common section starts at the point where the vehicle enters the first LIDAR and ends at the point where the last vehicle license plate is recorded by both cameras (see Fig. 6.2). Because the cameras and LIDARs are time-synchronized, their timestamps can be utilized as a common ground for such an adjustment.

The adjusted reference average speed over the common section of the road is computed as:

$$v_r = v + \frac{1}{2} * a * (t_1 - t_0), \tag{6.1}$$

where v is the reference immediate speed when entering the first laser; a is the reference acceleration; t_0 is the reference time when the vehicle entered the first laser; and t_1 is the time when the last license plate of the vehicle was recorded.

The adjusted measured average speed over the common section of the road is computed as:

$$p_0 = p + v * (t_0 - t) + \frac{1}{2} * a * (t_0 - t)^2$$

$$p_1 = p + v * (t_1 - t) + \frac{1}{2} * a * (t_1 - t)^2 \qquad (6.2)$$

$$v_m = \frac{||p_1 - p_0||}{t_1 - t_0}$$

where p_0 and p_1 are co-ordinates of vehicle positions at times t_0 and t_1; t_0 is the reference time when the vehicle entered the first laser; t_1 is the time when the last license plate of the vehicle was recorded; p is a co-ordinate of the initial vehicle position at time t; v is the co-ordinate of immediate vehicle speed at time t; and a is the co-ordinate of vehicle acceleration. The values of p, v, and a are obtained from a vehicle motion model (Eq. 5.5). The histogram of reference average speeds is shown in Fig. 6.3.

Stereo camera pair calibration

The reference data for the stereo camera pair experiment are the distances that vehicles travelled between two consecutive frames. The reference distances were computed from data provided by the two LIDARs using formulas:

$$p_i = v * (t_i - t) + \frac{1}{2} * a * (t_i - t)^2$$

$$p_{i+1} = v * (t_{i+1} - t) + \frac{1}{2} * a * (t_{i+1} - t)^2 \qquad (6.3)$$

$$d_{i,i+1} = ||p_{i+1} - p_i||$$

where v is the reference immediate speed when entering the first laser at time t, a is the reference acceleration, and t_i and t_{i+1} are times at which the frames i and $i+1$ were taken. The histogram of reference distances is shown in Fig. 6.4. The mean distance is 0.982 m and the standard deviation is 0.108 m.The maximum distance is 1.783 m and the minimum is 0.665 m.

Timestamp assignment latency

One more thing should be considered, and that is the delay between the end of camera exposure and the timestamp assignment, which, in this case, takes place in the computer that stores the frames. The cameras, as soon as the exposure ends, pack the read-out lines into the UDP packets and send them to the computer where they are received by the software. The timestamp is assigned immediately after receiving the first UDP packet of a new frame, and it is the same for both images.

The timestamp assignment latency is measured by pointing the cameras on a series of LEDs which encode the millisecond part of current time in a binary format (see Fig. 6.5). The recorded frames were examined and compared the time encoded in LEDs to the millisecond portion of the frame timestamp. In this case, the timestamp assignment latency has a mean value of 2.5 ms and a standard deviation of 1 ms. The measured values are used to correct the recorded frame timestamps in Eq. 6.2 and Eq. 6.3.
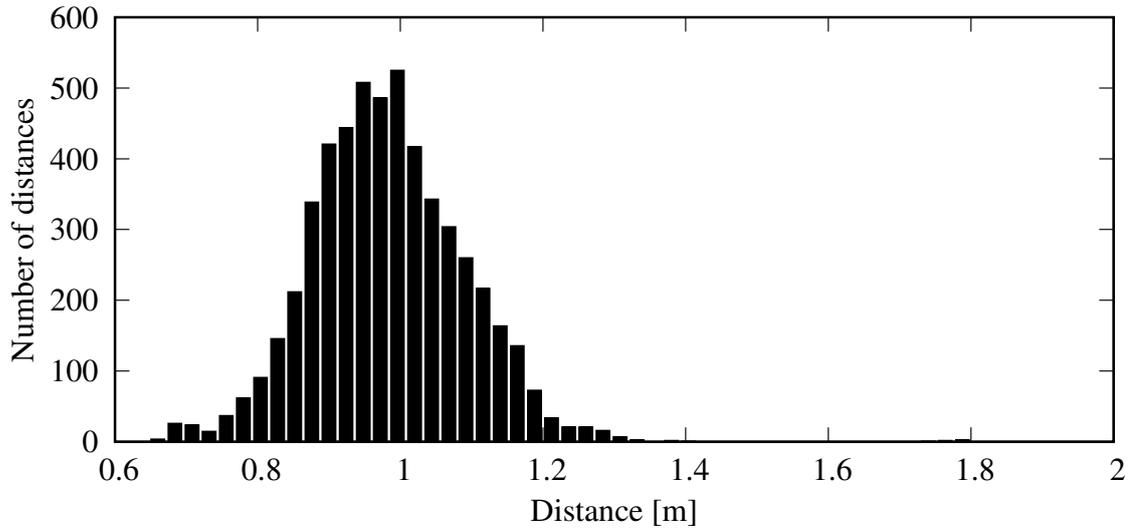
Figure 6.4: Histogram of reference distances measured by the two LIDARs.

6.3 Implementation

This section describes the implementations and execution times of both proposed methods.

Vehicle speed measurement

The proposed vehicle speed measurement method pipeline was broken into five smaller steps that were implemented as separate programs. These programs correspond to License plate detection and matching, License plate tracking, Point matching, Triangulation, and Speed measurement steps. The pipeline was implemented in C++ with the help of OpenCV and Boost libraries. The programs pass the data along the pipeline in JSON format through standard streams.

The implementation is focused more on the precision and accuracy of the measurement rather than on the optimality of the implementation or the speed of computation. Only a few optimizations were performed. They were aimed mostly at reducing the detection time of licence plates by constraining the image area in which the license plates are detected. To limit the license plate detection area, a background subtraction for each image of the stereo image pair is performed. The obtained foreground image masks are then morphologically dilatated to close the holes. After that, the license plates are detected in the masked left image. For each license plate found in the left image, an epipolar line in the right image is computed using a detected rectangle top left corner and fundamental matrix. The computed epipolar lines further limit the search area in the right image. The foreground masks are shown in Fig. 5.2b. Finally, the license plates in the masked right image are detected.

The average times of execution for each implemented step were evaluated separately. The evaluation took place on a Linux desktop computer with an Intel Core i5-6500 processor running at 3.2 GHz with 24 GB RAM. The data for the steps that are common to both methods, the Triangulation, and the Speed measurement steps is obtained using a sample vehicle recording that consists of 20 frames in which the vehicle license plate is successfully detected. The results are summarized in Table 6.1. The total processing time of the
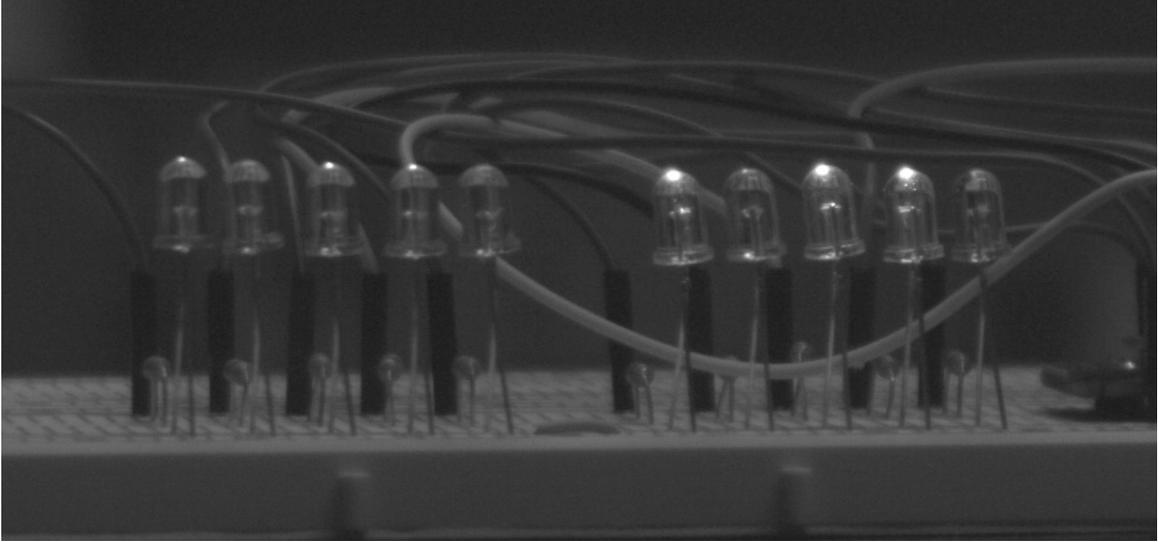
Figure 6.5: A camera image cutout that contains blinking LEDs. This image was assigned a timestamp with millisecond portion of 729. The LEDs in the image encode millisecond time just after the end of exposure which was 726 (whose binary image is 10110 10110 as seen above). The latency of timestamp assignment, in this case, is therefore 3 ms.

Table 6.1: Average per frame execution times of the individual pipeline steps of the vehicle speed measurement method.

Step	Time [ms]
License plate detection and matching	168.1
License plate tracking	3.6
Point matching	203.3
Triangulation	4.5
Speed measurement	10.0

vehicle speed measurement currently exceeds the 50 ms time limit for real-time processing of 20 frames per second supplied by the cameras. The most time-demanding steps are the License plate detection and matching and the Point matching steps, which makes them prime candidates for further optimization or implementation in hardware.

The total processing time per stereo frame with a single passing vehicle is approximately 0.4 seconds. The average number of stereo frames that were processed per vehicle in the dataset was 13, which means that a single vehicle can be processed in less than 6 seconds. If the average time distance between the two incoming vehicles is more than the processing time, all of the vehicles could be processed in near real-time. The mean time distance between the two incoming vehicles in the dataset acquired during „peak times" is approximately 3 seconds. Processing the traffic of this volume in near real-time would require cutting the processing time per vehicle to half or, for example, adding another computing unit. However, the traffic volume changes during the day, and, given enough storage, the low volume periods can be utilized to catch up with the processing of the stored frames. Because the real-time processing is usually not a requirement for traffic enforcement systems, the method should be able to compute the speed for all the passing vehicles as long

Table 6.2: Average per frame execution times of the individual pipeline steps of the stereo camera pair method.

Step	Time [ms]
License plate detection and matching	198.2
License plate tracking	3.7
Point matching	198.5
Calibration	85.0

as the traffic volume on a given location is less than approximately 14400 vehicles per day (600 per hour).

Stereo camera pair calibration

The same implementation approach was used for the proposed stereo camera pair calibration method. Its on-site calibration pipeline was divided into four smaller steps that were implemented as separate programs. These programs correspond to License plate detection and matching, License plate tracking, Point matching, and Calibration steps. Because the first three programs are the same for both pipelines, their implementation can be reused with just one small modification. The optimization that utilizes the epipolar geometry cannot be used as the information about the epipolar geometry is not yet known. The average execution times for each implemented step were obtained during the calibration process. During this process, the total of 748 stereo frames was processed, and 813 license plate pairs, which belonged to the 44 calibration vehicles, were detected. The evaluation took place on the same computer that was used for the evaluation of the vehicle speed measurement implementation. The results are summarized in Table 6.2.

The off-site calibration was performed in a laboratory using a chessboard pattern. After that, the prototype hardware was transported to a nearby footbridge across the road where it was properly set up and adjusted. The first 44 recorded vehicles, for which the ground truths were provided, were used for the on-site calibration. The road area covered by the calibration vehicles is shown in Fig. 6.6.

The estimation of external parameters and focal lengths is done by non-linear optimization process whose quality of results depends heavily on good initial estimates and proper limits. The initial estimates of focal lengths values and their upper and lower limits were chosen according to the used optics. The cameras used 35 mm fixed focal length lenses whose range of focus span from 0.25 m to infinity. The value of 35 mm was used as the initial estimates, and upper and lower limits were computed using the formula:

$$\frac{1}{l} = \frac{1}{f} + \frac{1}{s}, \tag{6.4}$$

where l is a lens focal length (35 mm in this case), f is a camera focal length, and s is a distance at which the fixed focal length lens is focused. Choosing 0.25 m and infinity for s the two f values that represent the upper and the lower limits for camera focal lengths estimation can be computed.

The external parameter t that describes the position of the second camera relative to the first camera is represented, for the minimization purposes, using spherical coordinates. This type of representation decouples vector magnitude from its direction, which enables specification of initial estimates and limits for each vector property separately. The highly

46

Figure 6.6: The area covered by the passes of the calibration vehicles through the measured road section is shown. Nine tracks that correspond to the extracted license plate points are drawn for each calibration vehicle.

accurate initial estimate of the vector magnitude can be easily provided by measuring the stereo camera pair baseline using, for example, laser distance meter. The magnitude limits, in this case, can be very narrow or the magnitude can be even excluded from the optimization and keep fixed. The measured baseline of the setup is 0.955 m. The initial estimates and limits for the spherical angles of the vector direction reflect the horizontal configuration of the stereo camera pair.

The external parameter R that describes the orientation of the second camera relative to the first is expressed using an angle-axis representation with the direction of its unit axis vector represented using spherical coordinates. The initial estimates and limits that were chosen for the three angles reflected the horizontal configuration of the stereo camera pair.

6.4 Results

This section presents and discusses the results of both experiments.
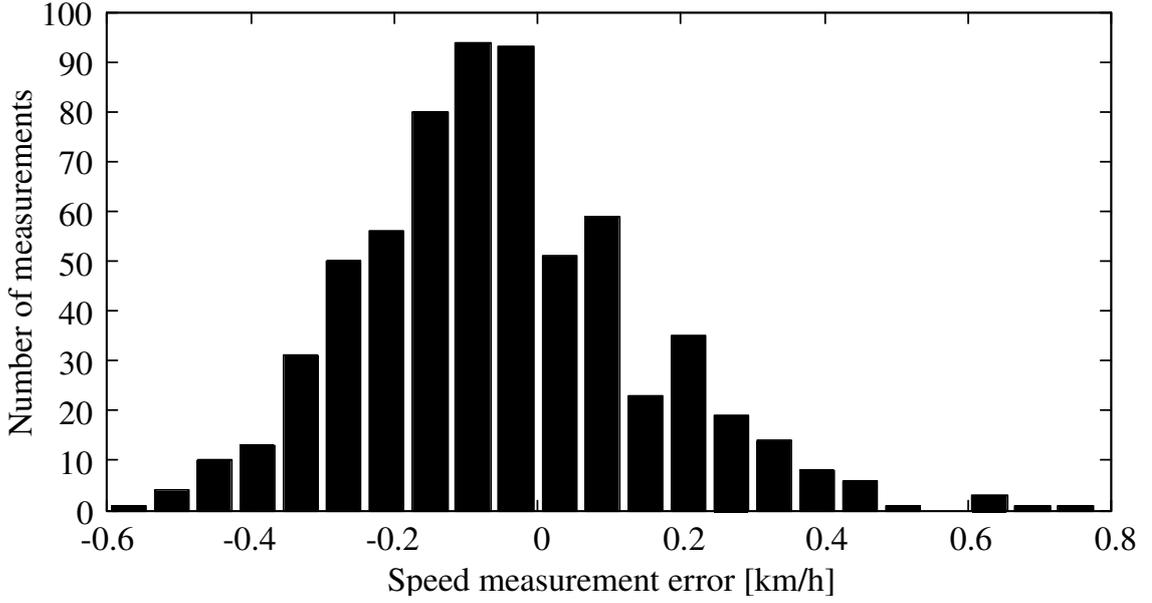
Figure 6.7: Histogram of speed measurement errors.

Vehicle speed measurement

The accuracy and precision of the speed measurement were evaluated using the above-mentioned dataset. The speed is measured for 653 of a total of 654 recorded vehicles. One measurement is missing because the license plate detector failed to detect the vehicle license plate. Detailed results for all vehicles in the dataset can be accessed online[1]. From the measured values, the measurement error was computed as:

$$e = v_m - v_d, \tag{6.5}$$

where v_m is average speed measured using the proposed method and v_d is reference average speed from the dataset. For the histogram of the speed measurement errors, see Fig. 6.7. Overall, the measured speed has a maximum negative error of -0.56 km/h and a maximum positive error of 0.72 km/h. The mean error is -0.05 km/h with a standard deviation of 0.20 km/h. The mean absolute percentage error is 0.23 % and maximum percentage error is 1.11 %. As all the error statistics are within the field test ranges recommended by the OIML (Table 3.1) for the autonomous measurement device, the described method should comply with the metrological legislations that are based on the OIML recommendations or those that have less strict error requirements. The presented hypothesis therefore holds.

The speed measurement errors are compared with three other stereo-based vehicle speed measurement methods mentioned in Chapter 4.1, namely, Jalalat et al.'s method [31], El Bouziady et al.'s method [14], and Yang et al.'s method [76]. The error comparison is shown in Table 6.4. For the datasets comparison see Table 6.3.

Table 6.3: Dataset comparison of stereo-based vehicle speed measurement methods.

	Dataset size	Number of different vehicles
Jalalat et al. [31]	441	441
El Bouziady et al. [14]	12	6
Yang et al. [76]	56	2
Proposed method	**653**	**653**

Table 6.4: Error comparison of stereo-based vehicle speed measurement methods.

	MSE [km/h]	STDEV [km/h]	Max abs. error [km/h]	Max perc. error [%]
Jalalat et al. [31]	NA	NA	NA	3.3
El Bouziady et al. [14]	2.33	NA	2.00	NA
Yang et al. [76]	0.42	NA	-1.60	3.80
Proposed method	**0.04**	**0.19**	**0.72**	**1.11**

Stereo camera pair calibration

The accuracy and precision of the presented calibration method were evaluated on a distance measurement task using the above-mentioned dataset. The distances that vehicles travelled between two consecutive frames are computed in several steps.

First, the License plate detection and matching and the License plate tracking steps from the on-site calibration part are applied to a vehicle record in order to extract a set of license plate pairs detected along the vehicle trajectory. Then, the algorithm transforms the set of license plate pairs to a set of license plate pair couples where each license plate pair couple represents a part of a vehicle trajectory whose length we want to compute. This transformation is done by coupling each license plate pair from the set with a license plate pair that was recorded in the consecutive frame, if possible. Next, the Point matching step is applied to each license plate pair couple from the set in order to obtain the nine corresponding point pairs for each license plate pair. Each of the nine point pairs from the first license plate pair of the couple is assigned its corresponding point pair from the second license plate pair of the couple. The point pair couples are then triangulated using the calibration method results, and the distance between the triangulated points of the couple is computed. This results in nine distances for each license plate pair couple, and the method takes their median distance as a distance that vehicle travelled between the two consecutive frames.

The distances are measured for 653 of a total of 654 recorded vehicles. Distances for one vehicle are missing because the license plate detector failed to detect the vehicle license plate. From the measured values, the measurement error is computed as:

$$e = d_m - d_d, \tag{6.6}$$

where d_m is the distance measured using the proposed method and d_d is reference distance speed from the dataset. For the histogram of the distance measurement errors, see Fig. 6.8. Overall, the 99 percentile absolute error is 0.05 m, and the mean error is -0.002 m with the standard deviation of 0.017 m.

---

[1] http://www.stud.fit.vutbr.cz/~xnajma00/results.json
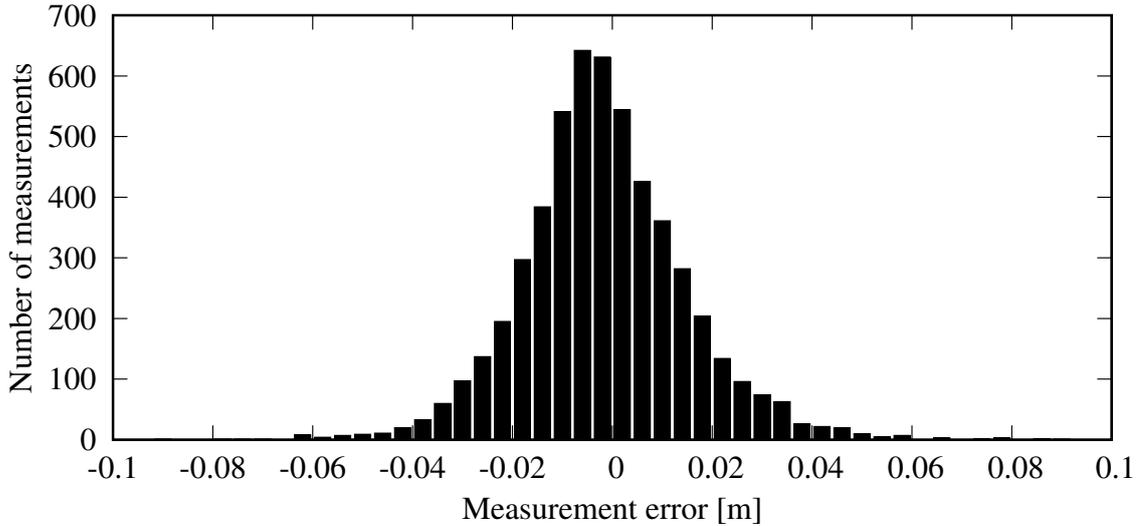
Figure 6.8: Histogram of distance measurement errors.

Table 6.5: Comparison of datasets and baselines of the stereo camera pair calibration methods.

	Distance from cameras [m]	Dataset size	Baseline [m]
Shang et al. [62]	50 - 70	15	30
Tian et al. [69]	49 - 85	28	79.88
Proposed method	**36 - 64**	**5365**	**0.955**

The distance measurement errors are compared with two other calibration methods that deal with cameras focused on a long shot mentioned in the chapter 4.5, namely, Shang et al.'s [62] method and Tian et al.'s [69] method. The comparison is shown in Table 6.6. Dataset properties and baselines are compared in Table 6.5. The proposed method achieves better results than the other methods in all compared statistics on a much bigger dataset with a much smaller baseline.

The setup that was calibrated using the proposed approach outperformed the existing methods that use the chessboard calibration (Jalalat et al.'s method [31], El Bouziady et al.'s method [14], and Yang et al.'s method [76]) in the precision and the accuracy of the vehicle speed measurement.

Table 6.6: Comparison of stereo camera pair calibration methods on a distance measurement task.

	MSE [m]	STDEV [m]	99 percentile abs. error [m]
Shang et al. [62]	0.0016	0.04	0.083
Tian et al. [69]	0.0011	0.033	0.063
Proposed method	**0.0003**	**0.017**	**0.05**

Discussion

The newly proposed vehicle speed measurement method achieves better results than the current state-of-the-art methods in all compared statistics on a much bigger and more diverse dataset. The speed measurement error of the measurement method is Normally distributed with the mean value of -0.05 km/h and the standard deviation of 0.20 km/h. The maximum absolute error is 0.72 km/h and the maximum percentage error is 1.11 %. These values were measured on a sample of 653 vehicles and are well within the ranges specified by the presented hypothesis. The hypothesis of this work, therefore, holds. Additionally, the proposed stereo camera pair calibration method also outperforms the current state-of-the-art long-distance calibration methods.

Chapter 7

Possible applications and future work

The first section of this chapter describes various possible applications of the proposed methods in the context of traffic surveillance. The ways in which the presented measurement and calibration approaches could be improved are presented in the second section.

7.1 Possible applications

The vehicle speed measurement method was designed mainly for the speed limit enforcement task. The devices that utilize this method can be permanently installed above the road and work autonomously. The CAMEA company showed interest in this type of application. The permanently installed devices could benefit from a wider baseline that allows for more precise measurement and extended multi-lane coverage. On the other hand, the small baseline can be exploited for portable setups that can be easily moved between locations or mounted on vehicles.

The speed enforcement task is not the only application for the stereo camera pair based devices in the area of traffic surveillance. The array of data that is provided by the two cameras is very rich and contains a lot of information that can be extracted and that is unique for this type of sensors. A simple modification of the presented method could be done in order to obtain information about vehicle count, traffic volume, and density. These are the basic traffic statistics that are often collected and used by civil engineers for further infrastructure development.

Another information that can be extracted from the stereo image pair is the vehicle class. The classification can be done in image space or by using a 3D reconstructed model of the vehicle. Unlike other devices that can usually provide only coarse classification, the stereo based devices could provide a very fine classification that could include vehicle manufacturer, vehicle type or even identification of non-standard vehicle customizations. The very fine classification could lead to easier vehicle re-identification using multiple setups on different locations that would not depend on license plate recognition. The re-identification can help to understand and monitor traffic flow patterns which are useful, for example, for smart cities development or air pollution control and improvement.

The 3D reconstruction does not need to be limited only to vehicles but could be extended to the whole scene. The reconstructed scene could enable better detection of dangerous

and undesired situations that include, for example, detection of pedestrians and animals on highways, incident detection, or traffic jams detection.

7.2   Future work

As of now, the presented system can be used only for short term monitoring. That is because it lacks the ability to maintain the correct calibration in time. The calibration correction is necessary for any long-term installed stereo-based device as the initial calibration is likely to change due to the external influences such as temperature changes or tremors.

The correct calibration can be maintained, for example, by tracking features of static objects in camera images and compensating for the changes in their positions. This approach would heavily rely on the precision of feature detection and tracking in both images and the presence of a static object with distinctive features in the scene. Alternatively, the use of the inertial measurement units (IMU) that would track the changes in the orientations of the cameras could be explored. This approach would not depend on the scene or the precision of feature tracking but would be limited only to the compensation of rotational changes because IMU cannot track the positional changes reliably due to the large drift. Another option is to add another camera that would observe the stereo camera pair and compute its external parameters. To make this task easier, the cameras of the stereo camera pair could be fitted with suitable markers.

The on-site part of the initial calibration process where the calibration vehicles need to be recorded, and their reference speed obtained is very time consuming and requires the involvement of multiple people. The calibration vehicles are used for the estimation of the external parameters and the focal lengths. If these values could be estimated with similar accuracy in a different way, the tedious process of recording the calibration vehicles could be removed. The re-calibration approaches mentioned above could help with the estimation of the external parameters. The focus of the auto-focus camera lenses can usually be controlled either manually or programmatically. The program can be utilized to estimate the current focal lengths of the cameras using the lens focal length, the range of focus and the number of focus steps taken.

The current implementation is not very efficient and can be optimized. Ideally, the total per-frame processing time should be less than the refresh rate of the cameras, which, in this case, is 50 ms. Reducing the processing time below that threshold would allow real-time processing of passing vehicles. This can be achieved in several ways. One way could be, for example, hardware acceleration of the most time-consuming parts of the pipeline which are the License plate detection and matching and the Point matching steps. Another way is to utilize multiple processing units and distribute the work among them.

Future work could also experiment with different license plate detection and point matching methods. A comparison of different approaches could yield interesting results, and it is probable that more suitable methods could be found for the given tasks.

The dataset that was used for the evaluation was recorded in the morning hours during favourable weather and lighting conditions. Expanding the limited scope of situations that are present in the dataset could also be the focus of future work. The situations that can expand the dataset are severe weather conditions, including the presence of rain, snow, or dust and otherwise varying weather and lighting conditions. The expanded dataset can then be used to evaluate the robustness and reliability of the presented approach in adverse conditions.

Chapter 8

Conclusion

This thesis answered the question of whether it is currently possible to measure the speed of vehicles using a stereoscopic method with the average error within ± 1 km/h, the maximum error within ± 3 km/h, and the standard deviation within ± 1 km/h. The error ranges were based on the requirements of the OIML whose Recommendations serve as templates for metrological legislations of many countries. Based on this question, a hypothesis was formulated and tested. A method that utilizes a stereo camera pair for vehicle speed measurement was proposed and experimentally evaluated.

The experimental evaluation was designed according to the design of the field test from the OIML Recommendation. Using the prototype hardware that consists of two synchronized cameras, ethernet switch, and two batteries mounted on one-meter long aluminium profile, a dataset was recorded. The dataset contains recordings of 698 vehicles for which the reference data is provided by a pair of LIDARs. The first 44 vehicles were used for stereo camera pair calibration. The remaining 654 vehicles were used for the evaluation. The reference average speeds and the average speeds measured by the method were used to compute the measurement errors. The results were compared to the existing stereo-based methods and to the OIML requirements.

The newly proposed method measures the speed of the passing vehicles more precisely than the other methods. The mean error, the standard deviation, and the maximum absolute error are -0.05 km/h, 0.2 km/h, and 0.72 km/h, respectively. These values are within the ranges specified in the hypothesis , which means that the presented hypothesis holds. The proposed stereoscopic vehicle speed measurement method should, therefore, comply with metrological legislations that are based on the OIML Recommendation.

Additionally, the stereo camera pair calibration method that is suitable for traffic surveillance applications was proposed and experimentally evaluated. The evaluation of the calibration method was based on a distance measurement task and utilized the same dataset as the vehicle speed measurement method. The results were compared to the results of the existing methods that deal with the calibration of stereo camera pairs that are focused on a long distance. The comparison showed that the newly proposed method is able to calibrate the stereo camera pair more precisely than the other methods.

The main focus of the future work should be on maintaining the correct calibration in time. Alternatively, one may also focus on decreasing the execution time, improving the resistance to adverse weather conditions, or experimenting with different detection, re-identification, or correspondence search methods.

Bibliography

[1] ALAHI, A., ORTIZ, R. and VANDERGHEYNST, P. Freak: Fast retina keypoint. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 2012, p. 510–517. ISSN 1063-6919.

[2] BAY, H., TUYTELAARS, T. and VAN GOOL, L. SURF: Speeded Up Robust Features. In: LEONARDIS, A., BISCHOF, H. and PINZ, A., ed. *Computer Vision – ECCV 2006.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, p. 404–417. ISBN 978-3-540-33833-8.

[3] BAŘINA, D., NAJMAN, P., KLEPÁRNÍK, P., KULA, M. and ZEMČÍK, P. The Parallel Algorithm for the 2D Discrete Wavelet Transform. In: *Ninth International Conference on Graphic and Image Processing (ICGIP 2017).* Qingdao: SPIE - the international society for optics and photonics, 2017, p. 1–6. ISBN 978-1-5106-1741-4.

[4] BECKER, A. *Kalman Filter Tutorial* [https://www.kalmanfilter.net]. 2018. Accessed: 2018-07-17.

[5] BEYER, P. Non-intrusive detection, the way forward. In:. Southern African Transport Conference, 2015.

[6] BROWN, D. C. Decentering distortion of lenses. *Photogrammetric Engineering and Remote Sensing.* 1966.

[7] BUKHARI, F. and DAILEY, M. N. Automatic radial distortion estimation from a single image. *Journal of mathematical imaging and vision.* Springer. 2013, vol. 45, no. 1, p. 31–45.

[8] CALONDER, M., LEPETIT, V., STRECHA, C. and FUA, P. BRIEF: Binary Robust Independent Elementary Features. In: DANIILIDIS, K., MARAGOS, P. and PARAGIOS, N., ed. *Computer Vision – ECCV 2010.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, p. 778–792. ISBN 978-3-642-15561-1.

[9] CANNY, J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence.* IEEE. 1986, no. 6, p. 679–698.

[10] CHAN, K.-L. and JUDAH, S. R. A beam scanning frequency modulated continuous wave radar. *IEEE Transactions on Instrumentation and Measurement.* IEEE. 1998, vol. 47, no. 5, p. 1223–1227.

[11] DENG, H., ZHANG, W., MORTENSEN, E., DIETTERICH, T. and SHAPIRO, L. Principal curvature-based region detector for object recognition. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 2007, p. 1–8. ISSN 1063-6919.

[12] DUBSKÁ, M., SOCHOR, J. and HEROUT, A. Automatic Camera Calibration for Traffic Understanding. In: *Proceedings of BMVC 2014*. 2014, p. 1–10.

[13] DUDA, R. O., HART, P. E. and STORK, D. G. *Pattern classification and scene analysis*. Wiley New York, 1973.

[14] EL BOUZIADY, A., THAMI, R. O. H., GHOGHO, M., BOURJA, O. and EL FKIHI, S. Vehicle speed estimation using extracted SURF features from stereo images. In: *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*. April 2018, p. 1–6. DOI: 10.1109/ISACV.2018.8354040.

[15] EVANGELIDIS, G. and PSARAKIS, E. Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Oct 2008, vol. 30, no. 10, p. 1858–1865. DOI: 10.1109/TPAMI.2008.113. ISSN 0162-8828. Available at: http://ieeexplore.ieee.org/document/4515873/.

[16] FAMOURI, M., AZIMIFAR, Z. and WONG, A. A Novel Motion Plane-Based Approach to Vehicle Speed Estimation. *IEEE Transactions on Intelligent Transportation Systems*. Apr 2019, vol. 20, no. 4, p. 1237–1246. DOI: 10.1109/TITS.2018.2847224. ISSN 1524-9050. Available at: https://ieeexplore.ieee.org/document/8401851/.

[17] FAUGERAS, O. Stratification of three-dimensional vision: projective, affine, and metric representations. *JOSA A*. Optical Society of America. 1995, vol. 12, no. 3, p. 465–484.

[18] FAUGERAS, O., HOTZ, B., MATHIEU, H., VIÉVILLE, T., ZHANG, Z. et al. *Real time correlation-based stereo: algorithm, implementations and applications*. Inria, 1993.

[19] FAUGERAS, O. D. What can be seen in three dimensions with an uncalibrated stereo rig? In: *European conference on computer vision*. Springer, 1992, p. 563–578.

[20] FISCHLER, M. A. and BOLLES, R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: FISCHLER, M. A. and FIRSCHEIN, O., ed. *Readings in Computer Vision*. San Francisco (CA): Morgan Kaufmann, 1987, p. 726 – 740. DOI: https://doi.org/10.1016/B978-0-08-051581-6.50070-2. ISBN 978-0-08-051581-6. Available at: http://www.sciencedirect.com/science/article/pii/B9780080515816500702.

[21] FITZGIBBON, A. W. Simultaneous linear estimation of multiple view geometry and lens distortion. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. IEEE, 2001, vol. 1, p. I–I.

[22] GONZALEZ, R. C. and WOODS, R. E. *Digital Image Processing*. 2ndth ed. USA: Addison-Wesley Longman Publishing Co., Inc., 2001. ISBN 0201180758.

[23] HARRIS, C. G., STEPHENS, M. et al. A combined corner and edge detector. In: *Alvey vision conference*. Citeseer, 1988, vol. 15, no. 50, p. 10–5244.

[24] HARTLEY, R. and ZISSERMAN, A. *Multiple view geometry in computer vision*. Cambridge, UK: Cambridge university press, 2003.

[25] HARTLEY, R. I. Euclidean reconstruction from uncalibrated views. In: *Joint European-US workshop on applications of invariance in computer vision.* Springer, 1993, p. 235–256.

[26] HARTLEY, R. I. Projective reconstruction and invariants from multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* IEEE. 1994, vol. 16, no. 10, p. 1036–1041.

[27] HARTLEY, R. I., GUPTA, R. and CHANG, T. Stereo from uncalibrated cameras. In: *CVPR.* 1992, vol. 92, p. 761–764.

[28] HARTLEY, R. I. and STURM, P. Triangulation. *Computer vision and image understanding.* Elsevier. 1997, vol. 68, no. 2, p. 146–157.

[29] HARTLEY, R. In defence of the 8-point algorithm. In: *Proceedings of the 5th International Conference on Computer Vision.* Cambridge: IEEE Computer Science Press, 1995, p. 1064–1070.

[30] HUANG, X. and DUBOIS, E. Dense disparity estimation based on the continuous wavelet transform [stereo image analysis]. In: *Canadian Conference on Electrical and Computer Engineering 2004 (IEEE Cat. No. 04CH37513).* IEEE, 2004, vol. 1, p. 465–468.

[31] JALALAT, M., NEJATI, M. and MAJIDI, A. Vehicle detection and speed estimation using cascade classifier and sub-pixel stereo matching. In: *2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS).* Dec 2016, p. 1–5. DOI: 10.1109/ICSPIS.2016.7869890.

[32] JAN, J. *Medical image processing, reconstruction and restoration: concepts and methods.* Boca Raton: CRC Press, 2005. ISBN 9781138310285.

[33] JIA, Y., GUO, L. and WANG, X. Real-time control systems. In: *Transportation Cyber-Physical Systems.* Elsevier, 2018, chap. 4, p. 81 – 113. DOI: https://doi.org/10.1016/B978-0-12-814295-0.00004-6. ISBN 978-0-12-814295-0. Available at: http://www.sciencedirect.com/science/article/pii/B9780128142950000046.

[34] JOSHI, K. A. and THAKORE, D. G. A survey on moving object detection and tracking in video surveillance system. *International Journal of Soft Computing and Engineering.* Citeseer. 2012, vol. 2, no. 3, p. 44–48.

[35] JUREWICZ, C., SOBHANI, A., WOOLLEY, J., DUTSCHKE, J. and CORBEN, B. Exploration of vehicle impact speed–injury severity relationships for application in safer road design. *Transportation research procedia.* Elsevier. 2016, vol. 14, p. 4247–4256.

[36] KALMAN, R. E. et al. A new approach to linear filtering and prediction problems [J]. *Journal of basic Engineering.* 1960, vol. 82, no. 1, p. 35–45.

[37] KLEIN, L. A., MILLS, M. K., GIBSON, D. R. et al. *Traffic detector handbook: Volume I.* Turner-Fairbank Highway Research Center, 2006.

[38] Lee, G. R. and Chen, Y.-M. *Passive acoustic traffic monitoring system.* 1999. US Patent 5,878,367.

[39] Li, X., Hu, W., Shen, C., Zhang, Z., Dick, A. et al. A survey of appearance models in visual object tracking. *ACM transactions on Intelligent Systems and Technology (TIST).* ACM New York, NY, USA. 2013, vol. 4, no. 4, p. 1–48.

[40] Liu, C., Pei, W., Niyokindi, S., Song, J. and Wang, L. Micro stereo matching based on wavelet transform and projective invariance. In:. IOP Publishing, 2006, vol. 17, no. 3, p. 565.

[41] Llorca, D. F., Salinas, C., Jiménez, M., Parra, I., Morcillo, A. G. et al. Two-camera based accurate vehicle speed measurement using average speed at a fixed point. In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC).* IEEE, Nov 2016, p. 2533–2538. DOI: 10.1109/ITSC.2016.7795963. ISBN 978-1-5090-1889-5. Available at: http://ieeexplore.ieee.org/document/7795963/.

[42] Longuet Higgins, H. C. A computer algorithm for reconstructing a scene from two projections. *Nature.* Springer. 1981, vol. 293, no. 5828, p. 133–135.

[43] Lowe, D. G. Object recognition from local scale-invariant features. In: *Proceedings of the seventh IEEE international conference on computer vision.* IEEE, 1999, vol. 2, p. 1150–1157. ISBN 0-7695-0164-8.

[44] Luvizon, D. C., Nassu, B. T. and Minetto, R. A Video-Based System for Vehicle Speed Measurement in Urban Roadways. *IEEE Transactions on Intelligent Transportation Systems.* June 2017, vol. 18, no. 6, p. 1393–1404. DOI: 10.1109/TITS.2016.2606369.

[45] Luz, E., Mimbela, Y. and Klein, L. A. Summary of vehicle detection and surveillance technologies used in intelligent transportation systems. *The Vehicle Detector Clearinghouse, Southwest Technology Development Institute (SWTDI) at New Mexico State University (NMSU), Fall.* 2007.

[46] MacCarley, C. A., Hockaday, S. L., Need, D. and Taff, S. S. Evaluation of video image processing systems for traffic detection (Abridgment). *Transportation Research Record.* 1992, no. 1360.

[47] Martin, P. T., Feng, Y., Wang, X. et al. *Detector technology evaluation.* Mountain-Plains Consortium Fargo, ND, 2003.

[48] Matas, J., Chum, O., Urban, M. and Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing.* 2004, vol. 22, no. 10, p. 761 – 767. DOI: https://doi.org/10.1016/j.imavis.2004.02.006. ISSN 0262-8856. British Machine Vision Computing 2002. Available at: http://www.sciencedirect.com/science/article/pii/S0262885604000435.

[49] Mondal, M. A. M. and Ali, M. H. On stereo correspondence estimation: a spiral search algorithm. In: *International Conference on Graphic and Image Processing (ICGIP 2011).* International Society for Optics and Photonics, 2011, vol. 8285, p. 82857H.

[50] MUSIL, P., JURÁNEK, R., MUSIL, M. and ZEMČÍK, P. Cascaded Stripe Memory Engines for Multi-Scale Object Detection in FPGA. *IEEE Transactions on Circuits and Systems for Video Technology.* 2019, vol. 30, no. 1, p. 267–280. DOI: 10.1109/TCSVT.2018.2886476. ISSN 1051-8215. Available at: https://www.fit.vut.cz/research/publication/11896.

[51] NAJMAN, P., ZAHRÁDKA, J. and ZEMČÍK, P. Projector-Leap Motion calibration for gestural interfaces. In: *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG).* Plzeň: Union Agency, 2015, p. 165–172. ISBN 978-80-86943-65-7.

[52] NAJMAN, P. and ZEMČÍK, P. *Summary Report of Contractual research - Tescan.* Brno: TESCAN Brno, s.r.o., 2018.

[53] NAJMAN, P. and ZEMČÍK, P. Stereo Camera Pair Calibration for Traffic Surveillance Applications. *IEEE Transactions on Intelligent Transportation Systems - UNDER REVIEW.* 2020.

[54] NAJMAN, P. and ZEMČÍK, P. Vehicle speed measurement using stereo camera pair. *IEEE Transactions on Intelligent Transportation Systems - MINOR REVISION.* 2020.

[55] OECD. *Towards Zero: Ambitious Road Safety Targets and the Safe System Approach.* Organisation for Economic Co-operation and Development, 2008.

[56] OIML. *OIML Convention* [https://www.oiml.org/en/files/pdf_b/b001-e68.pdf]. 1968. Accessed: 1999-12-13.

[57] OIML. *Radar equipment for the measurement of the speed of vehicles* [https://www.oiml.org/en/files/pdf_r/r091-e90.pdf]. 1990. Accessed: 2002-06-28.

[58] OIML. *OIML Strategy* [https://www.oiml.org/en/files/pdf_b/b015-e11.pdf]. 2011. Accessed: 2011-10-26.

[59] OJALA, T., PIETIKAINEN, M. and HARWOOD, D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In: *Proceedings of 12th International Conference on Pattern Recognition.* IEEE, 1994, vol. 1, p. 582–585. ISBN 0-8186-6265-4.

[60] PREWITT, J. M. Object enhancement and extraction. *Picture processing and Psychopictorics.* Academic Press New York. 1970, vol. 10, no. 1, p. 15–19.

[61] SCHARSTEIN, D. and SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision.* Springer. 2002, vol. 47, 1-3, p. 7–42.

[62] SHANG, Y., SUN, X., YANG, X., WANG, X. and YU, Q. A camera calibration method for large field optical measurement. *Optik.* Elsevier. 2013, vol. 124, no. 24, p. 6553–6558.

[63] SOCHMAN, J. and MATAS, J. Waldboost-learning for time constrained sequential detection. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05).* IEEE, 2005, vol. 2, p. 150–156.

[64] SOCHOR, J., JURÁNEK, R., ŠPAŇHEL, J., MARŠÍK, L., ŠIROKÝ, A. et al. Comprehensive Data Set for Automatic Single Camera Visual Speed Measurement. *IEEE Transactions on Intelligent Transportation Systems*. May 2019, vol. 20, no. 5, p. 1633–1643. DOI: 10.1109/TITS.2018.2825609.

[65] SOCHOR, J., JURÁNEK, R. and HEROUT, A. Traffic surveillance camera calibration by 3D model bounding box alignment for accurate vehicle speed measurement. *Computer Vision and Image Understanding*. Aug 2017, vol. 161, p. 87–98. DOI: 10.1016/j.cviu.2017.05.015. ISSN 10773142. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1077314217301108.

[66] SOCHOR, J., ŠPAŇHEL, J. and HEROUT, A. BoxCars: Improving Fine-Grained Recognition of Vehicles Using 3-D Bounding Boxes in Traffic Surveillance. *IEEE Transactions on Intelligent Transportation Systems*. Jan 2019, vol. 20, no. 1, p. 97–108. DOI: 10.1109/TITS.2018.2799228. ISSN 1524-9050. Available at: https://ieeexplore.ieee.org/document/8307405/.

[67] STRAND, R. and HAYMAN, E. Correcting Radial Distortion by Circle Fitting. In: *Proceedings of the British Machine Vision Conference*. BMVA Press, 2005, p. 9.1–9.10. DOI: 10.5244/C.19.9. ISBN 1-901725-29-4.

[68] STURM, P. Pinhole Camera Model. In: IKEUCHI, K., ed. *Computer Vision: A Reference Guide*. Boston, MA: Springer US, 2014, p. 610–613. DOI: 10.1007/978-0-387-31439-6_472. ISBN 978-0-387-31439-6. Available at: https://doi.org/10.1007/978-0-387-31439-6_472.

[69] TIAN, L., ZHU, W., LI, K. and YANG, Y. A camera calibration method for large field vision metrology. In: *2015 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE, 2015, p. 2632–2637.

[70] WALD, A. *Sequential Analysis*. New York: John Wiley, 1947.

[71] WANG, A., QIU, T. and SHAO, L. A simple method of radial distortion correction with centre of distortion estimation. *Journal of Mathematical Imaging and Vision*. Springer. 2009, vol. 35, no. 3, p. 165–172.

[72] WELCH, G. and BISHOP, G. *An Introduction to the Kalman Filter*. USA: University of North Carolina at Chapel Hill, 1995.

[73] WENG, J., COHEN, P. and HERNIOU, M. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. IEEE. 1992, no. 10, p. 965–980.

[74] WRAMBORG, P. A new approach to a safe and sustainable road structure and street design for urban areas. In: *Road safety on four continents conference*. Warsaw, Poland: Swedish National Road and Transport Research Institute (VTI), 2005, vol. 13, p. 12p–12p.

[75] WU, J., LIU, Z., LI, J., GU, C., SI, M. et al. An algorithm for automatic vehicle speed detection using video camera. In: *2009 4th International Conference on Computer Science & Education*. IEEE, 2009, p. 193–196.

[76] Yang, L., Li, M., Song, X., Xiong, Z., Hou, C. et al. Vehicle Speed Measurement Based on Binocular Stereovision System. *IEEE Access*. 2019, vol. 7, p. 106628–106641. DOI: 10.1109/ACCESS.2019.2932120. ISSN 2169-3536.

[77] Yilmaz, A., Javed, O. and Shah, M. Object tracking: A survey. *Acm computing surveys (CSUR)*. Acm New York, NY, USA. 2006, vol. 38, no. 4, p. 13–es.

[78] Zhang, Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Nov 2000, vol. 22, no. 11, p. 1330–1334. DOI: 10.1109/34.888718.

[79] Zhang, Z. Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*. Springer. 1998, vol. 27, no. 2, p. 161–195.

[80] ČMI. *Silniční rychloměry* [https://www.cmi.cz/sites/all/files/public/download/Uredni_deska/3405-ID-C_3405-ID-C.pdf]. 2010. Accessed: 2010-06-03.