Vehicle Tracking Through Vision-Millimeter Wave Doppler Shift Fusion

diploma thesis

Martín Trullenque Ortiz

Advisors: Univ. Prof. Christoph F. Mecklenbräuker, Erich Zöchmann and Herbert Groll

Vienna, February 2019

Abstract

Vehicular communications is a topic of interest in the race towards autonomous cars. Parallel, there is a future need to move towards higher frequencies in vehicular communications. In the same way, computer vision is becoming every time more used in automotive applications. In this thesis, I propose a system capable to track the position of a given car on a road with a camera and a wireless radio link in the V-band. For this I have developed some computer vision and image segmentation approaches with Gaussian mixture model for background estimation, erosion and epipolar geometry. Also, from a V-band wireless link and a signal analyser, a Doppler shift estimate has been obtained from a measurement campaign. Finally, a Kalman filter has been implemented for tracking the car's position with a highly accurate performance.

Kurzfassung

Fahrzeugkommunikation ist ein wichtiges Thema im Kontext von autonomen Fahrzeugen. Parallel dazu besteht in der Fahrzeugkommunikation ein zukünftiger Bedarf nach der Verwendung von höheren Frequenzen. Ebenso wird Computervision in Automobilanwendungen immer häufiger eingesetzt. In dieser Diplomarbeit schlage ich ein System vor, das die Position eines Autos auf einer Straße mit einer Kamera und einer drahtlosen Funkverbindung im V-Band verfolgen kann. Hierfür habe ich einige Ansätze zur Bildverarbeitung und Bildsegmentierung mit einer Gaußschen Mischverteilung für die Hintergrundschätzung, Erosion und epipolare Geometrie entwickelt. Außerdem wurde von einer V-Band-Funkverbindung und einem Signalanalysator die Doppler-Verschiebung geschätzt. Schließlich wurde ein Kalman-Filter implementiert, um die Position des Fahrzeugs mit hoher Genauigkeit zu verfolgen.

Dedication

I dedicate this work to the first person that taught me a lesson at university, that instilled me his passion for calculus and willingness for learning. To the person that knew how to make me laugh and advised me in the crucial and hard times in my first years at university. Because if he hadn't passed away so early, it would have been a pleasure to keep learning from him.

To the best professor I have ever had.

I dedicate this thesis to José Gómez Martí.

Acknowledgements

I would like to acknowledge my advisor Christoph F. Mecklenbräuker for giving me the opportunity to carry out my diploma thesis in Vienna.

To my advisors Erich Zöchmann and Herbert Groll for their help every time I needed it with an everlasting patience.

To my professors at UPC Javier Ruiz and Phillipe Salembier for easing me the material I have needed to perform some part of the thesis.

Table of contents

List of Figures

List of Tables

1. Introduction

Enhancing driving experience has been a topic in which many stakeholders have been working on. Making a quick review to the recent history of cars evolution in the last decades, at one point power steering was added for driving. Then cameras and distance sensors appeared for parking assistance. More research brought us very accurate depth measurements with cameras and different computer vision approaches such as Ada-Boost algorithms [1] [2]. Parallel to that and thanks to a huge research on the topic, the usage of unlicensed 5.9 GHz band for vehicular communications was allowed through the standard IEEE 802.11p [3]. These huge steps brought the well-known Tesla cars, which are nowadays the most autonomous cars.

However, there are two reasons that indicate that moving towards higher frequencies when referring to vehicular communications could be convenient. Firstly, there is a constant increase of data to be transmitted in vehicular communications. This leads to a higher throughput to transmit it in the same time [4]. Secondly, vehicular applications are growing nowadays. This means that the 5.9 GHz band in a couple of years will have to be shared among an exponentially increasing number of vehicles [5]. This can make us run out of radio resources. Moving towards higher frequencies would provide us of more throughput to share among the increasing number of users we expect to have.

There is another reason why radio applications are important in vehicular communications. Even though computer vision has improved its performance exponentially when referring to image segmentation and recognition (as we can see in the increase of biometrics applications), it still cannot provide the benefits radio applications can. Going deeper into a real case, Tesla cars perform their long-distance car detection through radar, so cameras do not seem to be a feasible option for cases where a long braking distance is needed.

In the framework of moving towards higher frequencies, V-band has become an interesting option. The unlicensed 60 GHz band is available worldwide and provides very high throughput. Moreover, it has the standard IEEE 802.11ad that regulates it.

In this latest framework the need arises to perform wide research towards autonomous car. One necessary application for autonomous cars, which is the main goal of this thesis, is vehicle tracking. Here, this goal is tackled using video frames and wireless radio data at V-band.

For developing this system, the measurements taken place on 25th September 2018 in Vienna downtown are used. The goal is approached in two independent ways, with video frames and with wireless radio data.

For processing video data, different image processing tools have been used. To isolate the car from the image, a background estimation using a Gaussian mixture model is implemented. Then, an erosion stage is used to remove artifacts and unwanted objects in the foreground detected image. After that, (as it is needed to measure some distances from the road to the camera), epipolar geometry is used to relate calibration images with the video ones. Finally, a first position estimation with the eroded image and some distance measured points has been made.

Parallel to that, from the wireless radio data, a frequency shift estimate is obtained using the techniques described in [6]. With this estimate and the position estimation previously done, a Kalman filter is implemented for fusing both data and provide us position and velocity tracking.

1.1. Background research

Going more in detail about the race of the last decades towards autonomous car it can be pointed out that vehicular communications and its architecture has been a popular topic among research for autonomous driving [7] [8] [9] [10] [11] [12]. In the past, research on the influence of the environment to wireless communications at higher frequencies has also been a popular focus [13].

The time-critical nature of road safety applications has imposed the need to accurately design the operation and performance of wireless vehicular communication systems [7]. This need has showed that vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication should go towards the usage of mm-wave [14] [15] [16]. In [16], an exhaustive study of the angle-of-arrival (AOA), angle-of-departure (AOD) and RMS delay spread is realised to evaluate the feasibility of the usage of radar in 28 GHz and 38 GHz band. In [15], a location of a car relative to a base station (BS) is performed through an AOA and AOD estimation. In [17], a characterization of 5.2 GHz channel in highway is performed. In [18], analysis and characterization of all elements taking part in a communications link at 5.9 GHz in V2X environment are performed.

Going in depth on the 60 GHz band, measurements of cars driving at constant speed and a fixed distance have been performed in [19]. In [20] and [21] the effect of overtaking cars in an urban environment in the 60 GHz band is shown.

The usage of cameras to detect moving cars has been a topic of research for the last decades [1], [22]. A radar system is combined with a camera applying Ada-Boost algorithm to perform car detection in [1]. However, this is not a joint estimation. In this case, the radar selects a region of interest (ROI) and afterwards Ada-Boost algorithm is performed. Other ranging methods such as lidar are used for identifying the ROI. For instance, in [2], a lidar structure is used to define the ROI and then, as well as in [1], Ada-Boost algorithm is used. In [23] a joint estimation is performed with lidar and a 3D camera that uses convolutional neural networks (CNN). An independent 76 GHz radar and camera car detection is performed in [22]. In that case, the radar itself is not used to estimate the velocity of the car and the computer vision (CV) technique used is support vector machine (SVM). In [24], using stereo images, different CV approaches are evaluated where the histogram of oriented gradients (HOG) is selected for the best recall/precision trade-off. Also using stereo images, a lane detector is implemented in [25], using a Hough transform.

1.2. Outline

This work is structured as follows: in the second chapter, the main tools used to achieve the goal are presented. In that section, starting from a main scheme of the system designed, the main characteristics of each stage of the system are explained. Also, in some of them preliminary results are provided, which are considered afterwards.

In the third chapter, results are provided starting from a brief description of a coarse position estimate and a radial velocity estimate. After that, data fusion of these first results with a Kalman filter is described. Furthermore, Kalman filter performance is compared to an alternative extended Kalman filter.

Finally, some conclusions about the results obtained are provided as well as mentioning some possible research paths.

2. Methodology

The aim of this section is to explain the employed tools to get a position and velocity estimate of a moving car, which is the main objective of this thesis.

For achieving this goal, two data sources are used. One of them is a 360 degrees video camera. The other one is a wireless radio communication system consisting of a transmitter array placed on top of a car and a receiver at a fixed position.

The camera is the GoXtreme Full Dome 360º, which has the following main features [26].

Double lens (220º each)
1920x960 pixels @30 fps video
3008x1504 px still image resolution
4MP sensor 2x
360º angle
WiFi function

Table 1: main parameters of Full Dome 360º

The transmitter of the communication system is an array placed on top of a car transmitting data continuously in V-Band. This transmitter array is formed by two transmitters, transmitter 1 with 0º beam elevation in z axis and transmitter 2 with 15º beam elevation in z axis. Axis are defined in Figure 12. On the receive site, the known transmit signal is captured with a R&S FSW67 signal analyser (SA). The parameters of the wireless communications system are summarized in the following table.

Parameter	**Value**
Centre frequency	$f_{\mathrm{TX}} = 60.15\ \mathrm{GHz}$
Number of TX antennas	2
Bandwidth	$100\ \mathrm{MHz}$
Transmit antennas	$20\ \mathrm{dBi}$ conical horn
Receive antenna	omnidirectional
Recording time	$T_{\mathrm{rec}} = 3.6\ \mathrm{s}$
Segment time length	$T_{\mathrm{seg}} = 0.83\mathrm{ms}$

Table 2: main parameters of the wireless channel

To synchronise both data sources, a light barrier on the road has been placed 24 m far away from the receiver. When the car is on the road it continuously transmits the transmission signal. Once it goes through the light barrier, a trigger activates the recording of the SA and also flashes a green light that it is seen from the camera to align recorded video frames with the radio signal. It must be highlighted that the transmitter array is located 1.8 m above the road and that the receiver is located 5 m above the road. It is also noticeable that the transmitter is 2.9 m behind the car hood. This means that when the car goes through the light barrier, the distance between the transmitter and the receiver is 26.9 m in y axis.

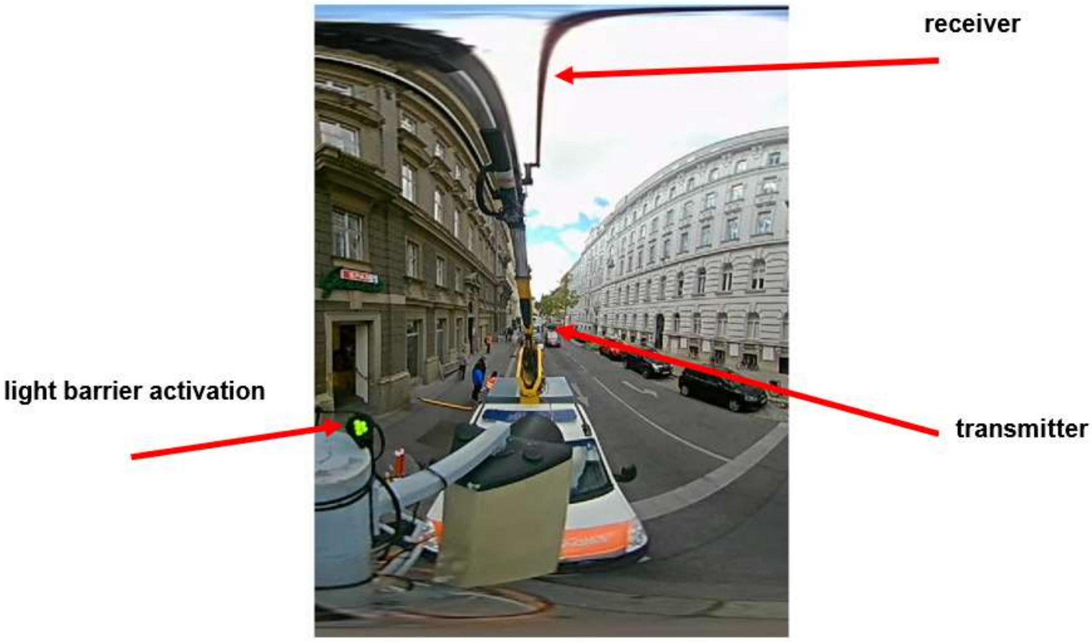

Figure 1: aligned recording frame

Figure 1 shows the trigger moment where the signal analyser receives the activation signal and LEDs are flashing for a visual marker in the recorded video frames.

The approach chosen for implementing the tracker has been to fuse a coarse position estimate coming from the video frames with a processed Doppler shift estimate obtained in [6], as it is shown in Figure 2.

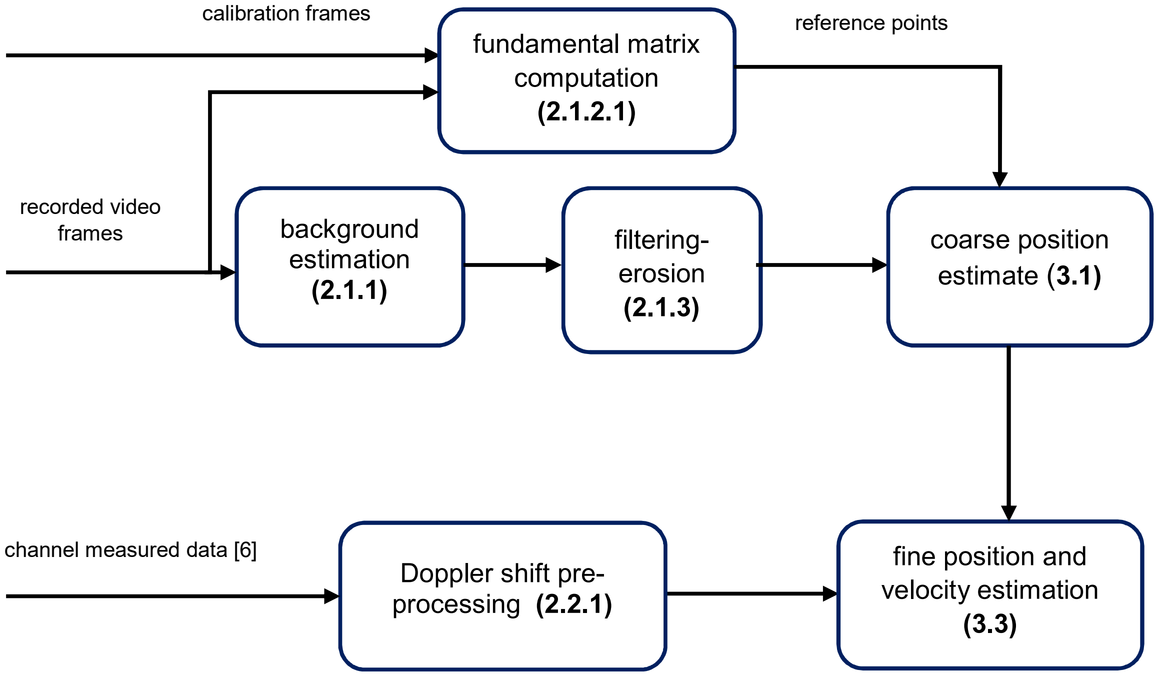

Figure 2: proposed system block scheme

Firstly, a background estimation is made from the recorded video, which helps to split the image into foreground and background. Secondly, different artifacts and unwanted objects are filtered through an erosion stage. Additionally, epipolar geometry has been applied due to the fact that, at the day of the measurements, it was not possible to take distances from the road to the SA. This arose the need of having to take some calibration images and relate video frames and calibration images with the fundamental matrix. Thirdly, an evaluation for each frame has been performed to know if the car had gone through the points of known distances.

Regardless of the position estimation through camera data, wireless communication data in the course of channel estimation described as in [6] has been obtained. This data consists on a received power estimate contribution for each transmitter, a noise power estimate and a Doppler shift estimate for each transmitter. The Doppler shift estimates are used afterwards to obtain an estimate of the lateral velocity of the car.

A Kalman filter and an extended Kalman filter are proposed to fuse the coarse position estimate obtained from camera data and the wireless radio data.

2.1. Camera data

The aim of this part is to explain the employed tools to extract the position of the car from the camera data.

Prior to this explanation, it is interesting to give some insights about some concepts that are assumed to be known later on.

Pinhole camera model: describes the mathematical relationship between the coordinates of a point in three-dimensional space and its projection onto the image plane in an ideal pinhole camera [27].

Stereoscopy: is a technique for creating or enhancing the illusion of depth in an image by means of stereopsis for binocular vision [28].

Epipolar geometry: is the geometry of stereovision. Assuming that two cameras view a 3D scene from two distinct positions, there are a number of geometric relations between the 3D points and their projections onto the 2D images that lead to constraints between the image points [28]. These relations are derived based on the assumption that the cameras can be approximated by the pinhole camera model [28]

Epipole: is the point of intersection of the line joining the camera centres with the image plane.

Epipolar plane: contains the baseline. There is a one-parameter family of epipolar planes [28].

Epipolar line: is the intersection of an epipolar plane with the image plane [28].

8-points algorithm: is an algorithm used in computer vision to estimate the essential matrix or the fundamental matrix related to a stereo camera pair from a set of corresponding image points [28]. The algorithm's name derives from the fact that it estimates the fundamental matrix form a set of eight corresponding image points.

It is relevant to note that, since the target is to know where the car is and the car is the only movement object of interest, frames can be decomposed into background and foreground, where the car is.

2.1.1. Background estimation

One practical way to get data from the video sequences is to isolate the car. This is feasible as it is the only moving object to differentiate between foreground (FG) and background (BG). For doing this, a background estimator has been selected, which can also be seen as a foreground detector.

To get a detection of the foreground, an estimation of the background is performed to then subtract it from the image of interest following the method presented in [29] and in [30].

Let $\underline{x}_t$ be the value of a pixel at a given time in a defined coordinate system. To do background estimation, it is necessary to decide whether this pixel belongs to FG or BG.

Given a pixel $\underline{x}_t$ we can model the probability of being included either in FG or BG. Making the Bayesian estimate we get to

$$R = \frac{\mathrm{p}(BG|\underline{x}_t)}{\mathrm{p}(FG|\underline{x}_t)} = \frac{\mathrm{p}(\underline{x}_t|BG)\mathrm{p}(BG)}{\mathrm{p}(\underline{x}_t|FG)\mathrm{p}(FG)}. \quad (1)$$

Where R is defined as the ratio between the probability of a known pixel to be background and the probability of a known pixel to be foreground.

Following the reasoning described in [31], assuming no prior knowledge about FG and BG (i.e. $\mathrm{p}(BG) = \mathrm{p}(FG)$), and also assuming that $\mathrm{p}(\underline{x}_t|FG)$ has a uniform distribution (i.e. $\mathrm{p}(\underline{x}_t|FG) = \gamma$), then it is decided that $\underline{x}_t$ belongs to BG if

$$\mathrm{p}(\underline{x}_t|BG) > \gamma, \quad (2)$$

Which can also be expressed as $\mathrm{p}(\underline{x}_t|BG) > \mathrm{p}(\underline{x}_t|FG)$.

As explained in [31] it is necessary to create a model of $\mathrm{p}(\underline{x}_t|BG)$. This model is estimated from a training set considering only the previous *L* frames. The choice of *L* has been done through an educated guess. However, it is interesting to note that large *L* in static scenarios will provide a exact model estimation, but in dynamic background scenarios (e.g. with many changes of light or image planes from one frame to the following one) a large *L* may lead to an interpretation of many pixels as false FG.

The model created from $\mathrm{p}(\underline{x}_t|BG)$ is called background model. But, as there are only *L* training frames, this background model is just a background estimation model, as it depends on the previous *L* frames.

After having done several research, the model selected to estimate the background has been a Gaussian mixture model. This model consists on a weighted sum of *K* variables with Gaussian distribution, mean $\underline{m_i}$ and covariance $\underline{\underline{\Sigma}}_i$.

For the implementation of the model, the foreground detection function provided by Matlab 2017b computer vision system toolbox [32] has been selected. This method consists on creating a foreground mask from subtracting a previously estimated background.

For the model execution, several parameters have to be fixed. In this work all the parameters except *K* and *L* have been kept to the default value Matlab 2017b computer vision system toolbox had [32]. After an educated guess, the number of Gaussians has been set to 3 as it is the lowest number that provides a good performance when the car is far away. Number of training frames *L* has been set to 20, as it was the maximum number frames where the car can be estimated from a long distance and still being able to estimate it in the "*test"* frames. Parameters used are summarized in Table 3.

Parameter	**Value**
Number of Gaussians (*K*)	3
Minimum background ratio	0.7
Learning rate	0.005
Initial covariance	30
Number of training frames (*L*)	20

Table 3: parameters used for foreground detection

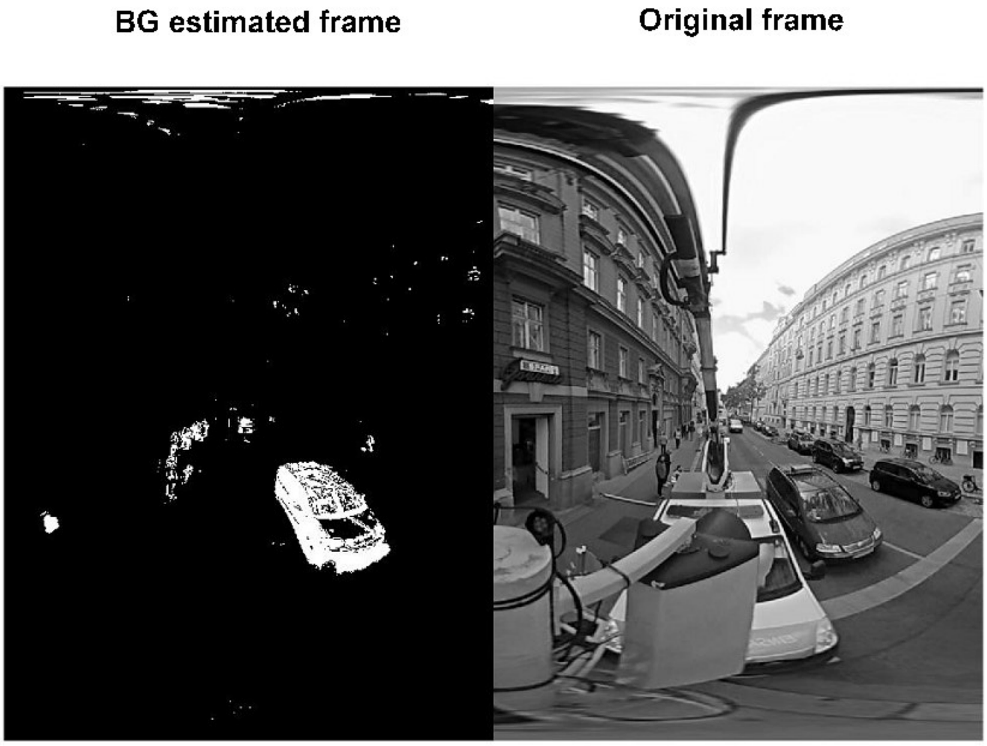

Figure 3: in the left-hand side, background estimated frame. In the right-hand side, original frame

In Figure 3, background estimation for the measurements site is shown. The right-hand side shows a selected frame for background estimation. The left-hand side shows the result of background estimation. The car is well estimated. Some unwanted objects, like the man walking through, can still be seen. This can be further improved by defining a ROI.

2.1.2. Epipolar geometry

Epipolar geometry is the intrinsic projective geometry between two views [28]. In this case, the need for applying epipolar geometry arises from the fact that the at day of the measurements it was not possible to measure distances between the SA and the road. Then, another day distances were measured and videos were recorded with the same camera. However, even trying to set the camera taking the calibration frames in the same position as the measurement day, some offsets could not have been avoided, so both measurement and calibration images have been treated as two different views. To correct these offsets, epipolar geometry has been used.

From the left-hand side scheme in Figure 4 it is possible to state the reasoning detailed in [28]: let a point P in 3-dimensional space be projected in x in one image and x' in another. To see the relation between them, it should be noted that x, x', space point P, and camera centres C and C' are coplanar [28]. Let me define this plane as π. The rays projected from x and x' intersect at P, and the rays are coplanar lying in π. This property is the most important for looking for a correspondence between the two views [28].

A correspondence connects two points from two different views, which refer to the same point in real coordinates.

Moving now to the right-hand side scheme of Figure 4, assuming that only x is known. If we look for x', we should consider that the plane π is defined by the baseline b and the ray defined by x [28]. We also know that the ray corresponding to the unknown point x' lies in π, then x' lies on the line of intersection l' of π with the second image plane [28]. This line l' is the image in the second view of the ray projected from x. It is the epipolar line corresponding to x [28].

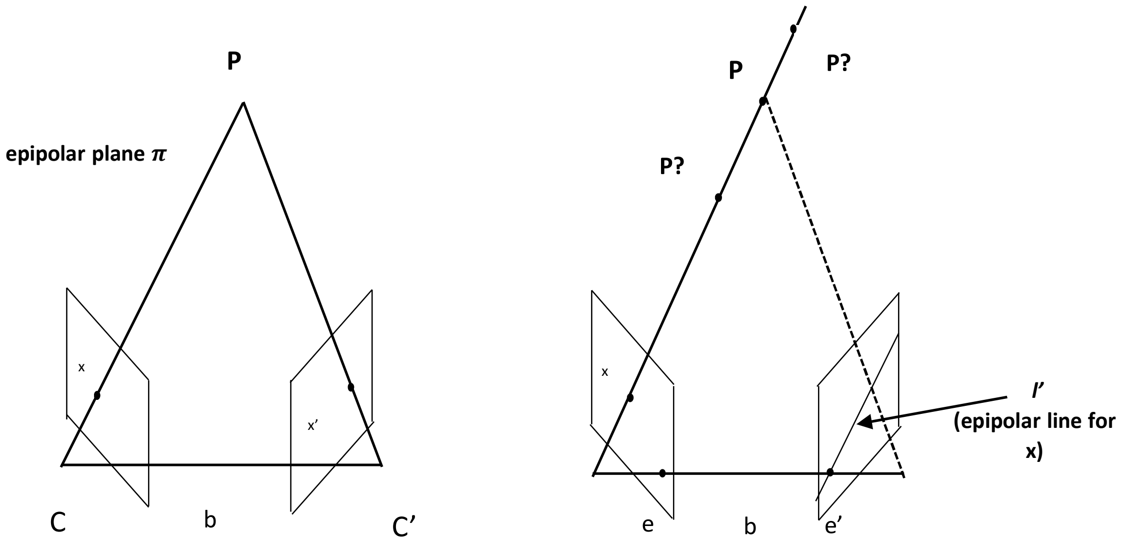

Figure 4: epipolar geometry schemes.

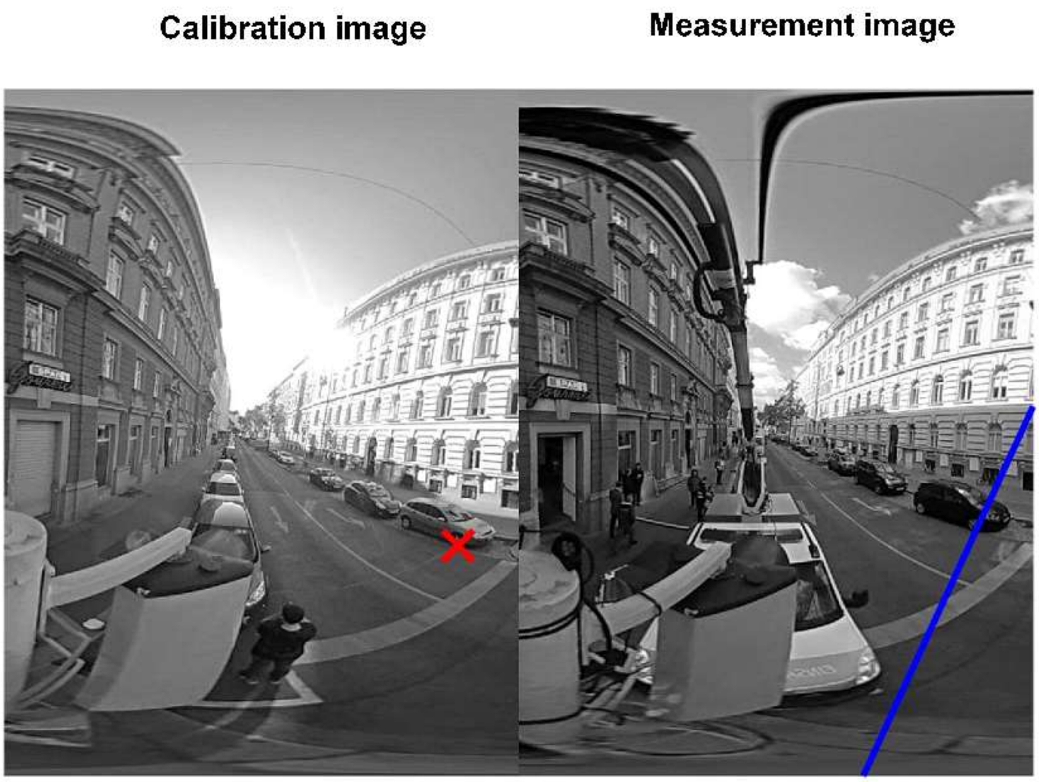

Figure 5: example of an epipolar line computation

In Figure 5 a real example of the theoretical development above can be seen.

In the left-hand side, a point P in world coordinates is projected into the red marker in the calibration image. The distance from P to the centre of the calibration image is known. Then, the goal is to find P in the measurement image. For this the epipolar line of the red marker is computed. This line is drawn in the measurement image (right-hand side of Figure 5).

2.1.2.1. Fundamental matrix

To compute epipolar lines from one image to another it is needed to compute the fundamental matrix.

The fundamental matrix F encapsulates the epipolar geometry. It is a 3x3 matrix of rank 2. If a point in 3D space (and homogeneous coordinates) P is imaged as $x = [u \quad v \quad 1]^T$ in the first view and $x' = [u' \quad v' \quad 1]^T$ in the second, then the image points satisfy the relation $x'^T F x = 0$, which is decomposed in Equation (3) [28].

$$[u' \; v' \; 1]\begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0 \tag{3}$$

Equation (3) is reformulated to

$$[uu' \; vu' \; u' \; uv' \; vv' \; v' \; u \; v \; 1]\begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{bmatrix} = 0. \tag{4}$$

Using this property and in order to estimate this matrix from two images, we use the 8-points algorithm. This algorithm consists on introducing 8 pairs of different and known x and x'. Then, Equation (4) allows to reformulate

$$\begin{bmatrix} u_1u_1' & v_1u_1' & u_1' & u_1v_1' & v_1v_1' & v_1' & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_8u_8' & v_8u_8' & u_8' & u_8v_8' & v_8v_8' & v_8' & u_8 & v_8 & 1 \end{bmatrix}\begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{bmatrix} = 0. \tag{5}$$

On the day of the wireless radio measurements it was not possible to compute some distances from the ground to the SA. This meant that there was a zone of interest on the road (from the direction arrows to the stop line) where no reference could be taken. To create these references, calibration images were taken separately. Afterwards the fundamental matrix was computed to relate the 3D points (ground) with two 2D points (the one of the calibration image and the video frame). The ends of the lines in Figure 6 are the correspondences that were taken.

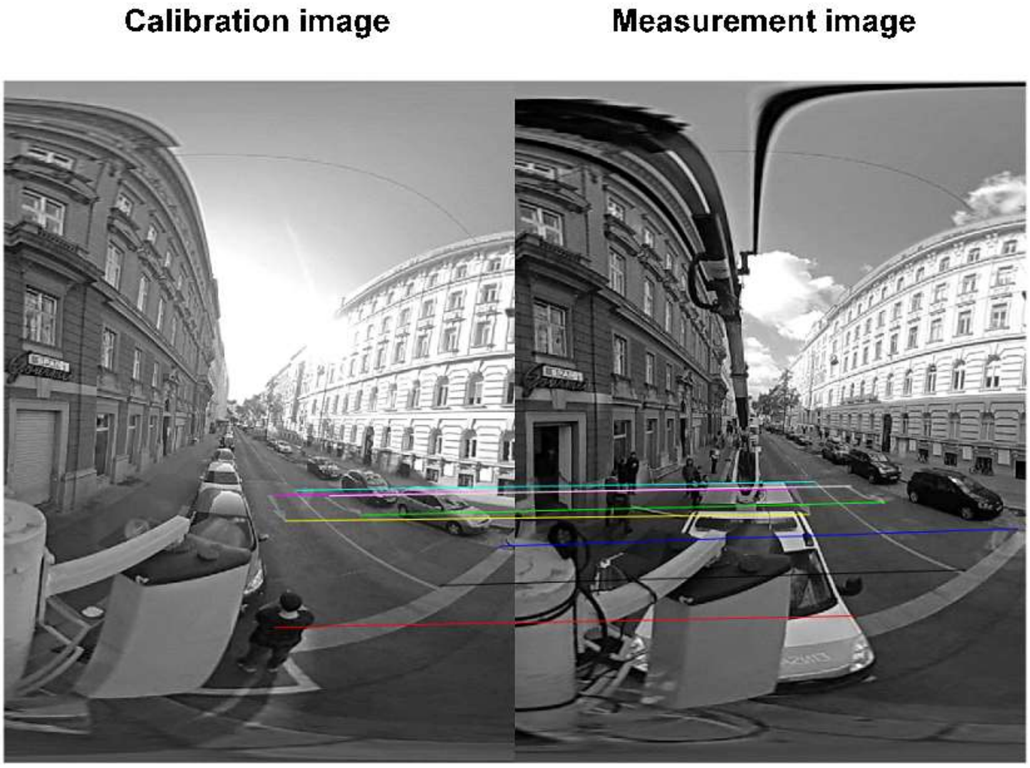

Figure 6: correspondences of points in calibration (left-hand side) and measurement (right-hand side) images

Once the fundamental matrix was estimated, points of the road with known distances to the calibration camera were selected (left-hand side of Figure 7). Using the fundamental matrix, I computed the epipolar lines of every of these points in the measurement image (right-hand side in Figure 7). As the selected points were intentionally at the boundary of the road, this let us conclude that the equivalent points will be the intersection of the boundary of the road with the epipolar lines, as it can be seen in right-hand side in Figure 7.

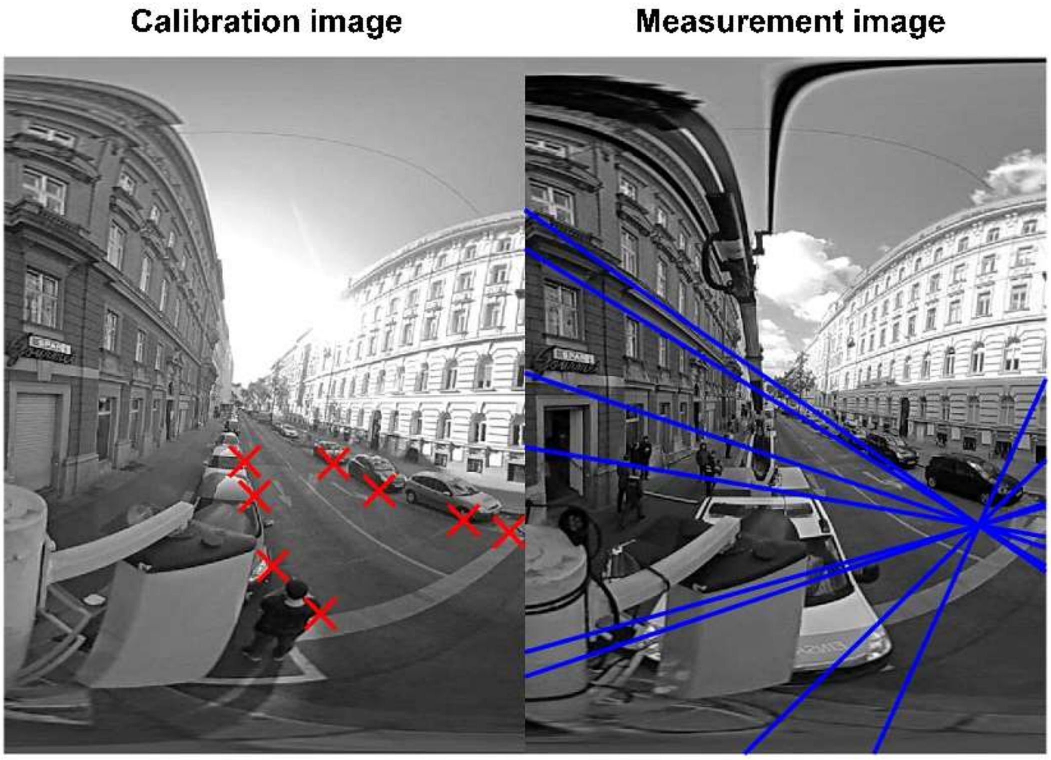

Figure 7: measured points in the calibration image (left-hand side) and their epipolar lines in the measurement image (right-hand side)

In the left-hand side of Figure 7, measured points in the calibration image are shown. For each point, its epipolar line in the image taken from the measurement's day has been computed, and shown in the right-hand side.

The distance between more points on the road and the traffic light was measured to have more distance marks without having to compute the equivalent point in the video frames.

2.1.3. Filtering - Erosion

To remove the different artifacts and unwanted objects, a filtering stage has to be implemented. In this thesis, an erosion stage has been implemented. Erosion is a filtering in which the output image will have in its centred pixel the infimum of the points of a filtered area of the input image. The element that defines this area is called structuring element (SE).

In a formal expression, the erosion of a binary image A given by a binary structuring element B can be expressed as.

$$A \ominus B = \bigcap_{b \in B} A_{-b} \tag{6}$$

Where A_{-b} denotes the translation of A by $-b$ and $\ominus$ is the erosion operator.

To illustrate it with a dummy example, let A be a 7x7 matrix such as

$$A = \begin{bmatrix} \underline{1} & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{7}$$

Where $\underline{1}$ expresses the $(0{,}0)$ position.

Let B be a structuring element such as

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \underline{1} & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{8}$$

Applying zero padding to A, after eroding A by B, the result is as follows

$$C = \begin{bmatrix} \underline{0} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{9}$$

From (9) it can be seen that the frame of 0s is due to zero padding. Also, the 0 in (7) in the middle when doing the erosion ends up being a 3x3 square of 0s in (9), as 3x3 is the size of the structuring element.

In our case, 3 different structuring elements were proposed and investigated further.

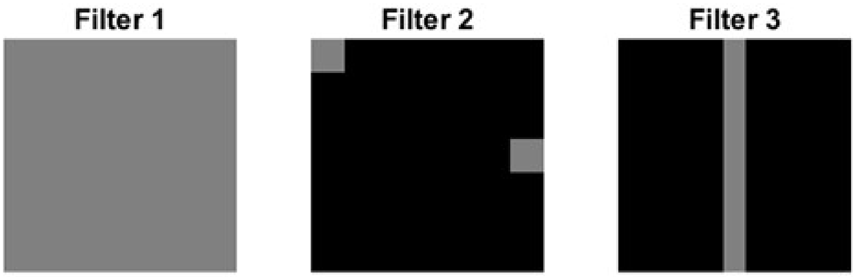

Figure 8: different structuring elements analysed

From Figure 8 it can be seen that structuring element 1 is a 7x7 squared matrix. Then the eroded image will have in each pixel the infimum of the given pixels and the 3 closest neighbours in each direction. In the second structuring element the resulting image will have in each pixel the infimum of the pixel located 3 pixels left-hand side and 3 up and the pixel located 3 pixels away at right-hand side. Finally, in the third structuring element, output image has in each pixel the minimum of the pixels of the 3 closest pixels in the input image pixel's columns.

The result of applying each structuring element to a given foreground detection image is shown in Figure 9.

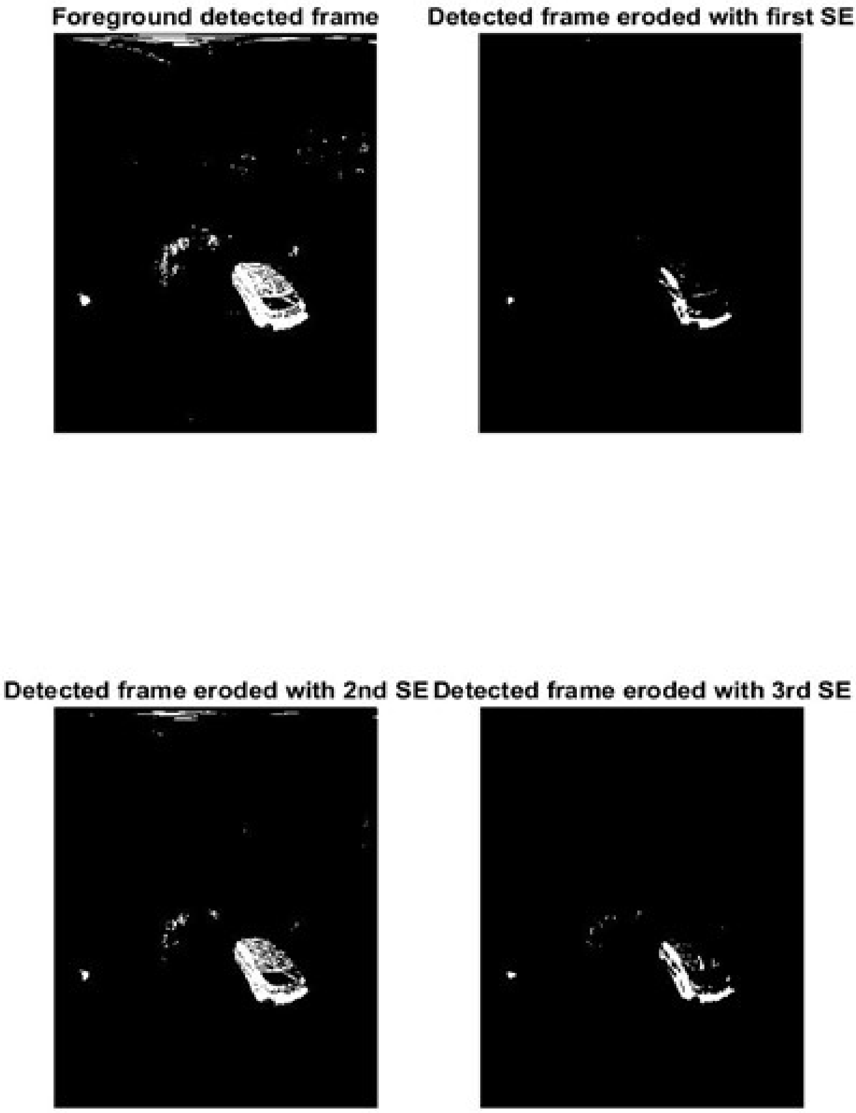

Figure 9: from top to bottom and left to right: background estimated frame. BG estimated frame afterwards being eroded with the first SE. BG estimated frame afterwards being eroded with the second SE. BG estimated frame afterwards being eroded with the third SE.

The second structuring element removes few artifacts as long as the image is being eroded with just one pixel. However, it keeps very well the accuracy of the front car hood. Then, both first and third SE provide a similar performance removing the artifacts, even none of them can remove accurately the unwanted elements. However, there are still some noticeable differences because of the fact that the first SE also erodes in the horizontal component, removing more artifacts and unwanted objects.

Deciding which SE should be chosen leads to a trade-off. Removing as many artifacts as possible lead to a cleaner image. However, this procedure leads to a loss of accuracy when estimating the position of the image. As a high accuracy is needed, the second SE has been chosen. Furthermore, as it is only checked if the car goes through the line that divide the lanes, many of these artifacts are not crucial.

2.2. Wireless radio data

The aim of this section is to explain the techniques used to estimate the velocity from the recorded radio data. For this, a short introduction of the acquired data is done before providing the notation of the extended Kalman filtering.

The acquired wireless radio data comes from an urban scenario (as it can be seen in Figure 10), in which the transmitter is an antenna array located on top of a moving car. The receiver is positioned at the crossroads, at 5m height. The goal is to estimate the velocity of the moving vehicle during its approach to the crossroads. The receiver is a Rohde and Schwarz signal analyser (SA) located next to a traffic light. The centre frequency of both transmitter and receiver is at 60.15 GHz. More details of the setup can be seen in [6].

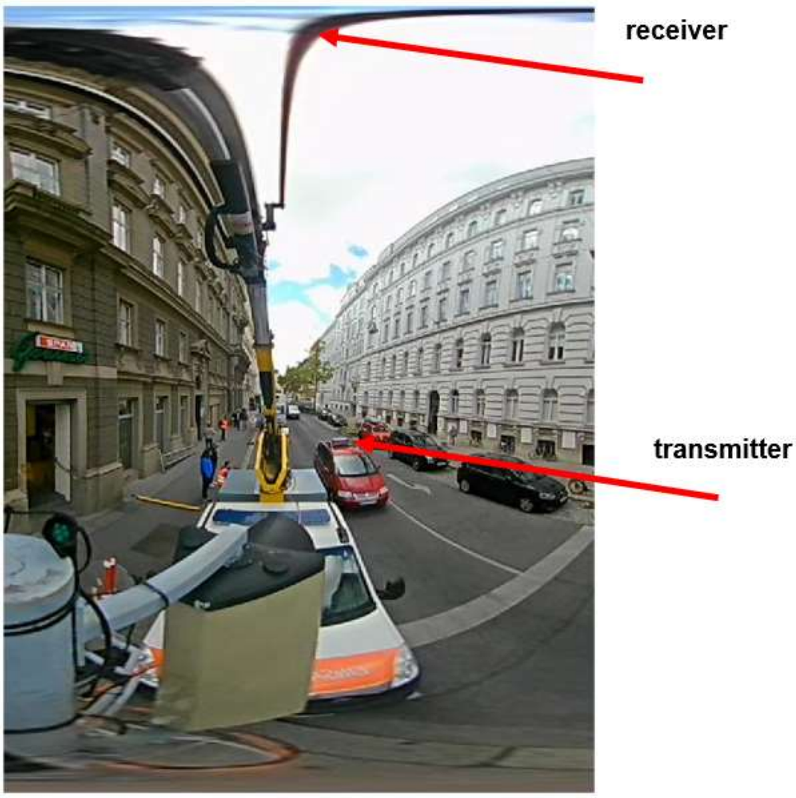

Figure 10: wireless channel setup

From the measurement evaluation performed in [6] an estimation of the received power contribution of each of the transmitters $\underline{P_{R,i}} = [P_{R,i,0}, \ldots, P_{R,i,S-1}]$ is obtained, as well as an estimate of the noise power $\underline{P_n} = [P_{n,0}, \ldots, P_{n,S-1}]$. It is important to note that index *i* stands for the transmitter and each component of the vectors stand for the segment of time the sample is referring to. The number of segments is *S*, which translates to a recording time of 3.6 s. Having the estimate of the received power and the noise, leads to define the signal-to-noise ratio (SNR) for each transmitter as follows

$$\underline{SNR_i} = \underline{P_{R,i}} / \underline{P_n}. \tag{10}$$

Furthermore, a Doppler frequency shift estimate for each transmitter is obtained in every segment $\underline{\hat{f}_{\mathrm{Dopp,i}}} = [\hat{f}_{\mathrm{Dopp,i,0}}, \ldots, \hat{f}_{\mathrm{Dopp,i},S-1}]$, which allows to estimate the radial velocity of the car, due to Doppler shift.

2.2.1. Doppler shift pre-processing

From the acquired wireless radio data, a Doppler shift frequency estimate for each transmitter is obtained. This provides an estimate of the radial velocity of the car with respect to the receiver according to

$$f_{\mathrm{Dopp}} = \frac{f_{\mathrm{TX}}}{c} v_r. \tag{11}$$

From (11), f_{Dopp} is the Doppler frequency shift, f_{TX} the transmission frequency, c the speed of the light in vacuum and v_r the radial velocity estimate.

In this case, and defining the angles α and β as in Figure 11, v_r can be expressed as:

$$v_r = v_y \cos\alpha \cos\beta. \tag{12}$$

Where in (12) v_y, is the velocity of the car along y axis, defining the axis as in Figure 12.

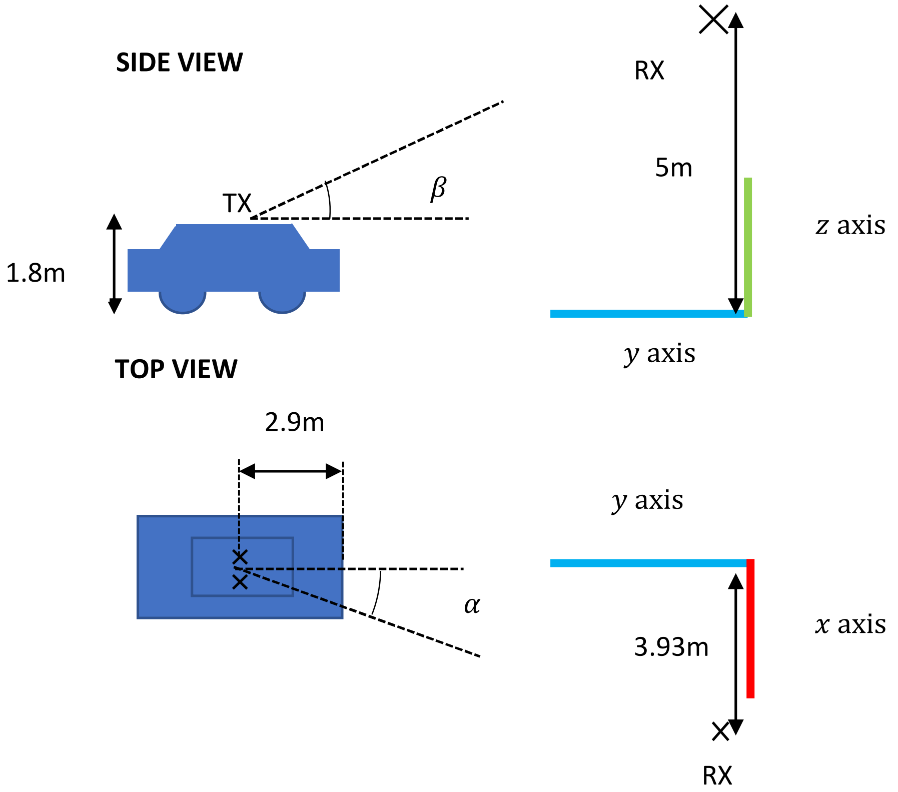

Figure 11: side and top view scheme of the measurement scenario

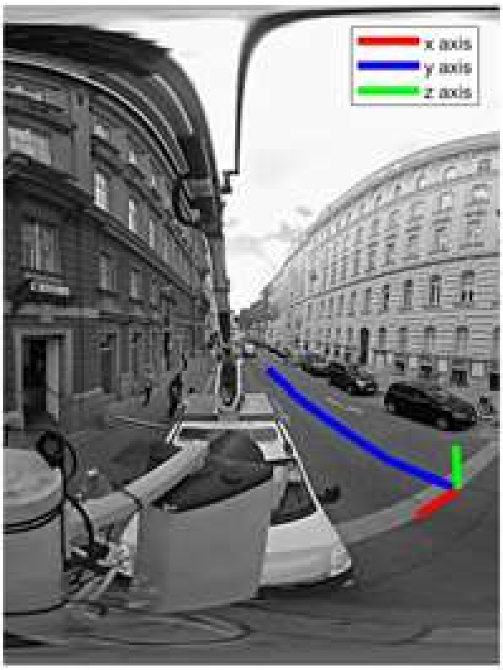

Figure 12: axis definition with street view coordinates

The angle α in Figure 11 is the angle between y axis and the visual line of sight (LOS) between the transmitter and the receiver projected onto the xy plane. Then β in Figure 11 is the angle between y axis at the transmitter height, and the LOS between the transmitter and the receiver projected onto the yz plane.

The estimate of the Doppler shift and the coarse position estimate from the camera data are used as inputs for the Kalman tracker, which is used for fine position and velocity estimation.

2.2.2. Extended Kalman filter

Here the notation for our Kalman filter is introduced.

The transition equation is given by

$$\underline{x}_{k+1} = \underline{\underline{A}}\,\underline{x}_k + \underline{w}. \tag{13}$$

Where $\underline{\underline{A}}$ is a matrix that expresses the evolution of the parameters to be estimated. $\underline{\underline{A}}$ is a model of the dynamics of the system parameters. The vector $\underline{x}$ represents the parameters to be estimated and the index *k* expresses the evolving time.

In addition, to allow a model mismatch, a noise term $\underline{w}$ is added. This vector contains as many components as parameters to be estimated and each component follows a Gaussian distribution with 0 mean and variance ${\sigma_{wi}}^2$. This variance is a free parameter which must be adjusted. This captures the uncertainty of the model. The covariance matrix of $\underline{w}$ is

$$\underline{\underline{W}} = E\left[\underline{w}\underline{w}^H\right], \tag{14}$$

which is chosen empirically, considering how trustable the parameter's evolution in time is. On the one side, setting a low variance means that it won't adapt to sudden changes, but on the other side, setting a high variance means that it might not follow the model accurately.

The measurements equation is given by

$$\underline{z}_k = \; g(\underline{x}_k) + \underline{n}. \tag{15}$$

Where $\underline{z}_k$ is the vector of measurements obtained, $g(.)$ expresses the relationship of the measurement and the parameters vector, and $\underline{n}$ is a noise vector.

In case of having linear relationship, $g(.)$ is expressed as a matrix $\underline{\underline{G}}$, which is the case of Kalman filtering.

When $g(.)$ is not a linear function, then it's linearized by

$$\underline{\underline{G}}_k = \left[\frac{\partial g(\underline{x}_k)}{\partial x_{k,1}}, \frac{\partial g(\underline{x}_k)}{\partial x_{k,2}}, \frac{\partial g(\underline{x}_k)}{\partial x_{k,3}}\right], \tag{16}$$

where the second index in the denominator shows the i^{th} component of the parameter's vector.

The noise vector $\underline{n}$ in (15) expresses the uncertainty of the measurements. From the estimated parameters, a linear combination of them is performed to get the measurements with a certain noise term $\underline{n}$. This vector contains as many components as observations we have, and each component is a Gaussian noise of 0 mean and variance ${\sigma_{ni}}^2$.

The covariance matrix of $\underline{n}$ is again

$$\underline{\underline{C}} = E\left[\underline{nn}^H\right], \tag{17}$$

which is chosen intentionally to reflect the noise of the measurements taken.

Furthermore, to update $\underline{x}$ in time considering the estimation error, the error covariance matrix $\underline{\underline{P}}$ and the gain matrix $\underline{\underline{K}}$ are defined respectively.

In this thesis, a Kalman filter and an extended Kalman filter have been implemented. Both of them have as input $\underline{\hat{x}}_{\text{pos}}$ and $\underline{\hat{y}}_{\text{pos}}$ estimated from the coarse position estimate stage. Angles α and β (described in Figure 11) have also been estimated applying trigonometric ratios. It is necessary to remark that the time between radio data segments is not the time between video frames. The interval between segments is $T_{\text{seg}} = 0.83\ \text{ms}$ .and between frames is $1/_{30}$ seconds.

In Kalman filtering x and y position ($\underline{x}_{\text{pos}}$and $\underline{y}_{\text{pos}}$ respectively) as well as the velocity along y axis $\underline{v}_y$ are estimated. It must be noted that the velocity estimate is the one along y axis and not the radial velocity. This is due to the inputs $\underline{q}_1$ and $\underline{q}_2$ where $f_{\text{Dopp,i,k}}$ is already divided by the product of the cosines of the two angles as Kalman filter only admits linear relations. However, in extended Kalman filtering the input is the estimated parameters from the coarse position estimation and $\underline{f}_{\text{Dopp,i}}$ as defined in (11), which is linearized in $\underline{\underline{G_k}}$

Parameters are summarized in Table 4.

Kalman filter	**Extended Kalman filter**
$\underline{x}_k = [x_{\text{pos},k} \quad y_{\text{pos},k} \quad v_{y,k}]^T$	$\underline{x}_k = [x_{\text{pos},k} \quad y_{\text{pos},k} \quad z_{\text{pos},k} \quad v_{y,k} \quad \cos\alpha_k \quad \cos\beta_k]^T$
$\underline{\underline{A}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & T_{\text{seg}} \\ 0 & 0 & 1 \end{bmatrix}$	$\underline{\underline{A}}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & a(\underline{x}_k) & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

$\underline{\underline{G}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{f_{Tx}}{c} \\ 0 & 0 & \frac{f_{Tx}}{c} \end{bmatrix}$	$\underline{\underline{G_k}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial f_{Dopp,1,k}}{\partial x_{k,4}} & \frac{\partial f_{Dopp,1,k}}{\partial x_{k,5}} & \frac{\partial f_{Dopp,1,k}}{\partial x_{k,6}} \\ 0 & 0 & 0 & \frac{\partial f_{Dopp,2,k}}{\partial x_{k,4}} & \frac{\partial f_{Dopp,2,k}}{\partial x_{k,5}} & \frac{\partial f_{Dopp,2,k}}{\partial x_{k,6}} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
$\underline{z_k} = [\hat{x}_{\mathrm{pos},k} \quad \hat{y}_{\mathrm{pos},k} \quad q_{1,k}(\alpha_k, \beta_k) \quad q_{2,k}(\alpha_k, \beta_k)]^T$	$\underline{z_k} = [\hat{x}_{\mathrm{pos},k} \quad \hat{y}_{\mathrm{pos},k} \quad \hat{z}_{\mathrm{pos},k} \quad f_{\mathrm{Dopp},1,k} \quad f_{\mathrm{Dopp},2,k} \quad \cos(\hat{\alpha}_k) \quad \cos(\hat{\beta}_k)]^T$
$\underline{\underline{K_k}} = \underline{\underline{P_k}}\, \underline{\underline{G}}^H (\underline{\underline{C}} + \underline{\underline{G}}\, \underline{\underline{P_k}}\, \underline{\underline{G}}^H)^{-1}$	
$\underline{\underline{P_k}} = (\underline{\underline{I_2}} - \underline{\underline{KG}})\, \underline{\underline{\hat{P}_k}}$	
	$a(\underline{x_k}) = \frac{T_{\mathrm{seg}}}{\underline{x}_{5,k}\underline{x}_{6,k}}$
$q_{i,k}(\alpha_k, \beta_k) = f_{\mathrm{Dopp},i,k} / \cos(\hat{\alpha}_k)\cos(\hat{\beta}_k)$	
	$\frac{\partial f_{\mathrm{Dopp}}}{\partial x_{k,4}} = \frac{f_{Tx}}{c}\underline{x}_{k,5}\underline{x}_{k,6}$
	$\frac{\partial f_{\mathrm{Dopp}}}{\partial x_{k,5}} = \frac{f_{Tx}}{c}\underline{x}_{k,4}\underline{x}_{k,6}$
	$\frac{\partial f_{\mathrm{Dopp}}}{\partial x_{k,6}} = \frac{f_{Tx}}{c}\underline{x}_{k,5}\underline{x}_{k,4}$

Table 4: Kalman filter and EKF equations definition

3. Results

The aim of this section is to present and evaluate the different results obtained from the measurements taken for the thesis. The measured scenario and the measurement setup are similar as in [6]. The only difference from [6] is that in this work additionally video data is being used.

As explained in Chapter 2, to synchronise both the camera and the SA, a light barrier has been set 24 m away from the SA position. When the car goes through this light barrier, a trigger activates the recording of the SA and also flashes a green light that it is seen from the camera to align recorded frames. This allows to align both data sources.

The SA and the 360º camera are placed on top of a traffic light, 5 meters above the road level.

The recording time from the SA has been set to 3.6s, which let us acquire 4298 segments of data. Within that time 108 frames are captured by the camera, as the frame rate is 30 fps. The software used for all post-processing and evaluation stage has been Matlab version 2017b.

3.1. Tracking results

3.1.1. Coarse position estimation

From the images taken in the calibration stage, reference points are obtained. These references can be seen in Figure 13. Blue points are the selected afterwards in Figure 15.

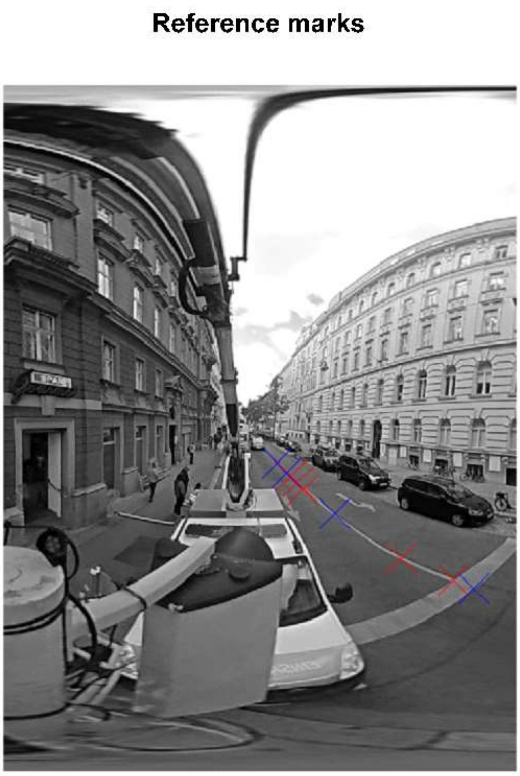

Figure 13: reference marks

To make the position estimate, an evaluation for each of the eroded frames is done to check which references the car, as foreground, has passed. For points between two marks, the trajectory of the car is linearly interpolated. The results of an exemplary measurement run are shown in Figure 14.

Coarse trajectory estimate

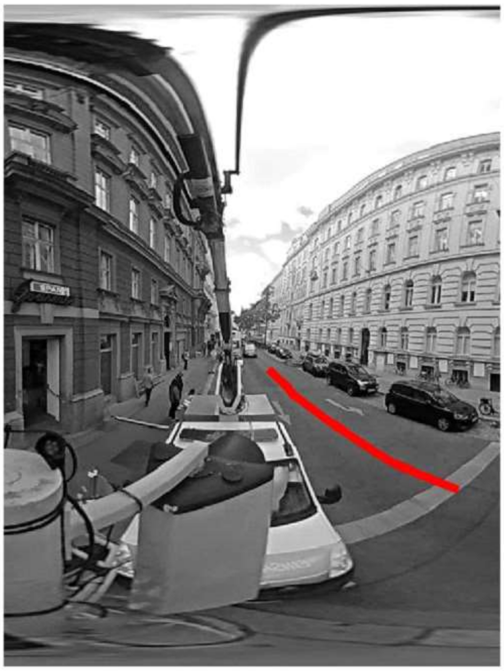

Figure 14: coarse trajectory estimate

The red line shows the trajectory of the car position during the measurement. The positions are relative to the previously defined coordinate system. For real world coordinates, further corrections are necessary. The real-world coordinates of the blue marked points of Figure 13 have been obtained using ViennaGIS, and the result is shown in Figure 15. The trajectory is exported to Google Maps.

Exported trajectory estimate

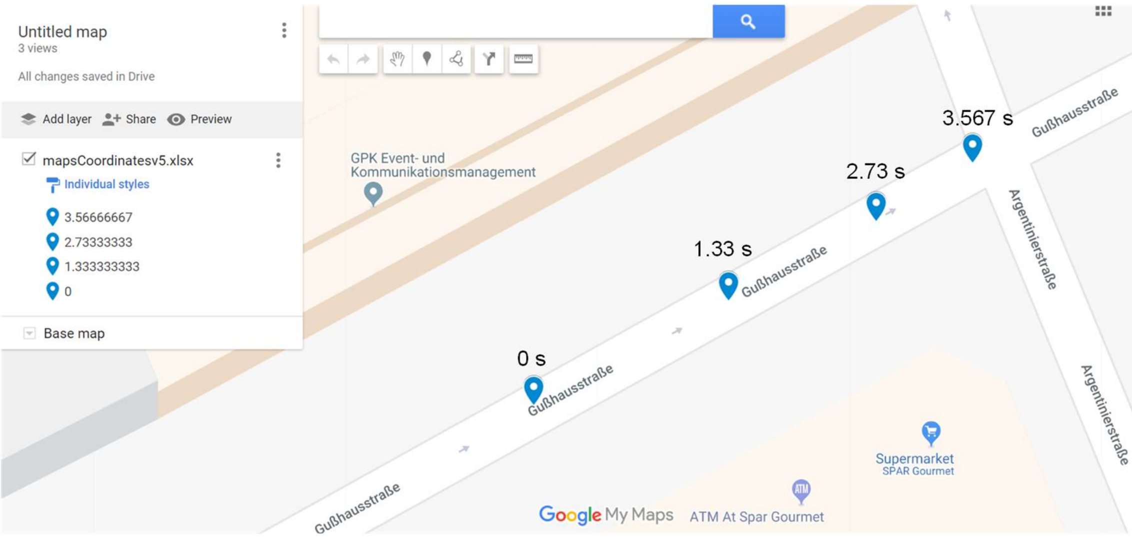

Figure 15: map pointing the coordinates associated to some of our points of interest

To understand the image, it should be noted that the blue marked locations match with selected locations placed in Figure 13. The title of the locations, which is also below each location, is the elapsed time for passing the light barrier in seconds when the car went through the checkpoints.

3.1.2. Velocity estimate through Doppler shift estimate

As explained in Section 2.2, from the radio data acquired using [6], a Doppler shift frequency estimate $\underline{f}_{\mathrm{Dopp,i}}$ is obtained for the received signal from each transmitter. From this estimate a Kalman filter has been developed having both $\underline{f}_{\mathrm{Dopp,i}}$ as inputs and v_r as estimate. As in this case just v_r is estimated, matrix $\underline{\underline{A}}$ turns into a scalar value a in (18). For the same reason, $\underline{\underline{G}}$ turns into a vector of two components $\underline{g}$ in (19), where each of the components stands for each $\underline{f}_{\mathrm{Dop}\ ,\mathrm{i}}$.

This estimation is done for both transmitters.

The Kalman filter matrices are as follows

$$a = 1, \tag{18}$$

$$\underline{g} = \begin{bmatrix} f_{TX}/c \\ f_{TX}/c \end{bmatrix}. \tag{19}$$

As the velocity evolution of the car cannot be predicted, no time evolution model is proposed and no previous knowledge can be applied. In the measurement's equation the conversion from the Doppler frequency shift to the radial velocity is made applying (11). The central transmitted frequency f_{TX} has been assumed the same both transmitters. Selected variance and initialization parameters are summarized in Table 5.

Value	Parameter
Velocity variance ($\sigma_w{}^2$)	0.1
Measurements variance for transmitter 1 ($\sigma_{n1}{}^2$)	700
Measurements variance for transmitter 2 ($\sigma_{n2}{}^2$)	200
Velocity initialization (m/s)	10

Table 5: first Kalman radial velocity estimate

A high velocity initialization has been chosen to check how fast the Kalman filter could adapt to sudden changes. Also, as it is known that the time between samples is lower than a millisecond, car velocity cannot change a lot. This is why the velocity variance chosen has been low.

Results are shown in Figure 16 and Figure 17.

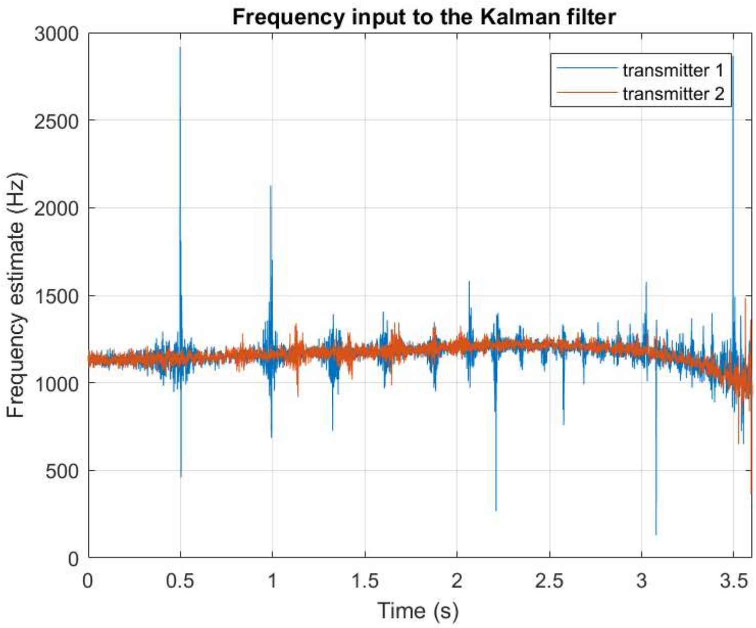

Figure 16: frequency input of the Kalman filter for both transmitters

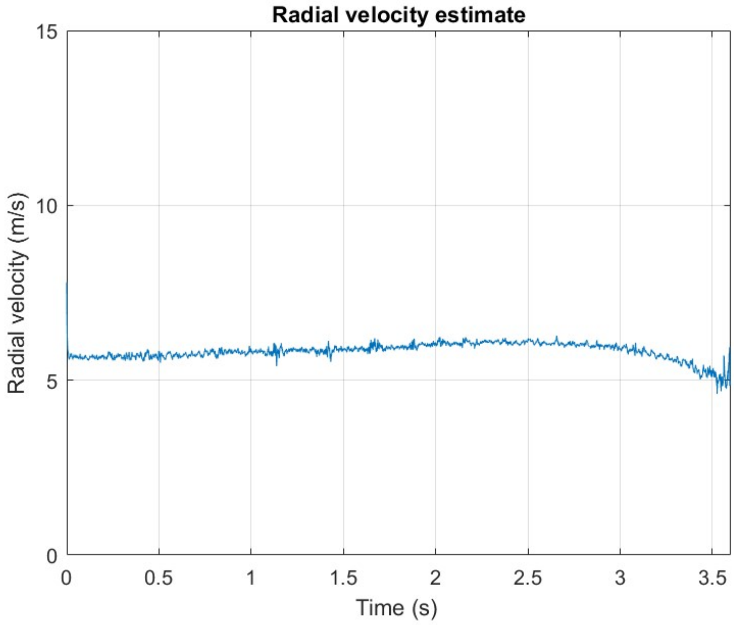

Figure 17: result of the radial velocity estimate for both transmitters

The behaviour of the Kalman filter parameter can be seen. Peaks at erroneous frequency shift estimate due to fading holes of the wireless channel have been smoothed. This has happened because of having set the variance of the measurements far higher than the initial state. This means that if the input sample is very different from the one according to (13), the model should trust only a variance of the order of the initial state. Then, sudden variations are being removed. Another point to remark is that the measurement variance of the Doppler frequency estimate from transmitter 2 is lower than transmitter 1 as the signal is smoother with less peaks. This can be explained with the 15º beam-elevation being less susceptible to fading due to spatial filtering of the road environment.

3.1.3. Joint estimation

Finally, after getting the results of the coarse position estimate, two algorithms have been developed to make a joint estimate of position and velocity: a Kalman filter and an extended Kalman filter.

3.1.3.1. Kalman filter estimate

Having the model described in Table 4, inputs have been the coarse position estimate and the frequency shift estimate. The parameters have been adjusted as follows in Table 6.

The choice of the ${\sigma_{w3}}^2$ variance comes from taking into account the time between frequency shift segments. It must be noted that ${\sigma_{w3}}^2$ is lower than in Section 3.1.2. This is done taking into consideration that Kalman filtering reduces the general weighted estimate error. After adding $x_{\mathrm{pos,k}}$ and $y_{\mathrm{pos,k}}$ to the parameter's vector, the estimation error increases. As the radial velocity estimate in Section 3.1.2 was quite steady, v_y has been assumed to be quite steady too and for this ${\sigma_{w3}}^2$ has been chosen far lower than ${\sigma_{w1}}^2$ and ${\sigma_{w2}}^2$.

As in Section 3.1.2, ${\sigma_{n4}}^2$ is lower than ${\sigma_{n3}}^2$ as the frequency shift estimate is more reliable for transmitter 2.

Parameter	Value
$x\ axis$ state variance (${\sigma_{w1}}^2$)	0.01
$y\ axis$ state variance (${\sigma_{w2}}^2$)	0.2
Velocity state variance (${\sigma_{w3}}^2$)	$1e-3$
x axis measurement variance (${\sigma_{n1}}^2$)	0.5
y axis measurement variance (${\sigma_{n2}}^2$)	0.4
Frequency measurement variance (${\sigma_{n3}}^2$)	700
Frequency measurement variance (${\sigma_{n4}}^2$)	200

Table 6: joint position and velocity estimate Kalman filter parameters

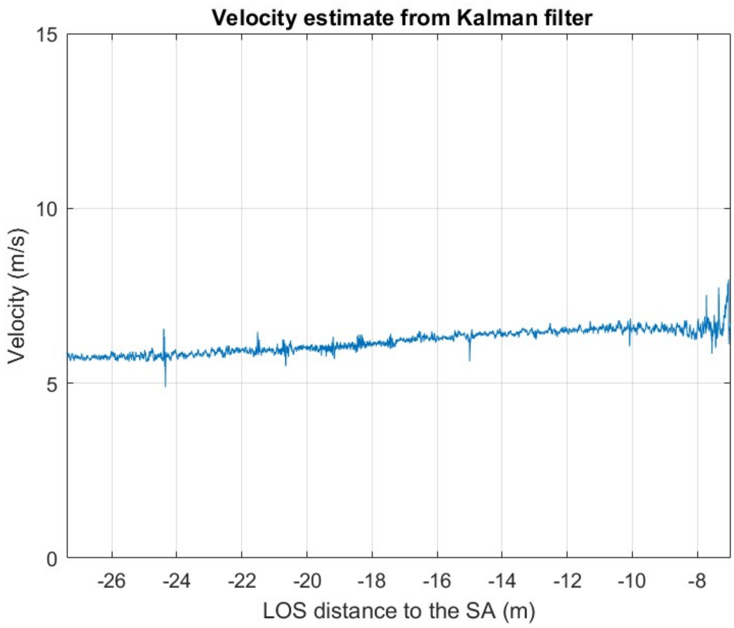

Figure 18: velocity estimate with Kalman filter

An increasing velocity estimate v_y can be observed in comparison to radial velocity v_r in Figure 18. Using (12) an explanation can be obtained. In the right-hand side image of Figure 16 it can be seen how the velocity decreases as the car goes towards the SA. However, considering angles α and β, velocity increases. However, having introduced position estimate leads to a less accurate velocity estimate, as Kalman filtering reduces general estimation error. This means that the estimation error is now weighted for every transmitter.

Even though no direct comparison between radial velocity from the frequency shift estimate and velocity along y axis from camera data is possible, this last estimate can be compared with the velocity estimate from the coarse position estimate. The results of the comparison can be found in Figure 19 and in Figure 20.

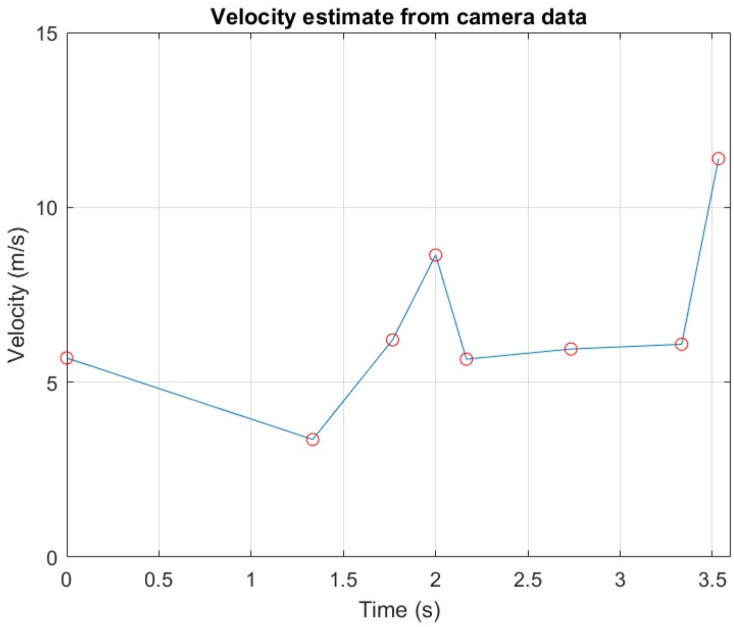

Figure 19: velocity estimate from camera data

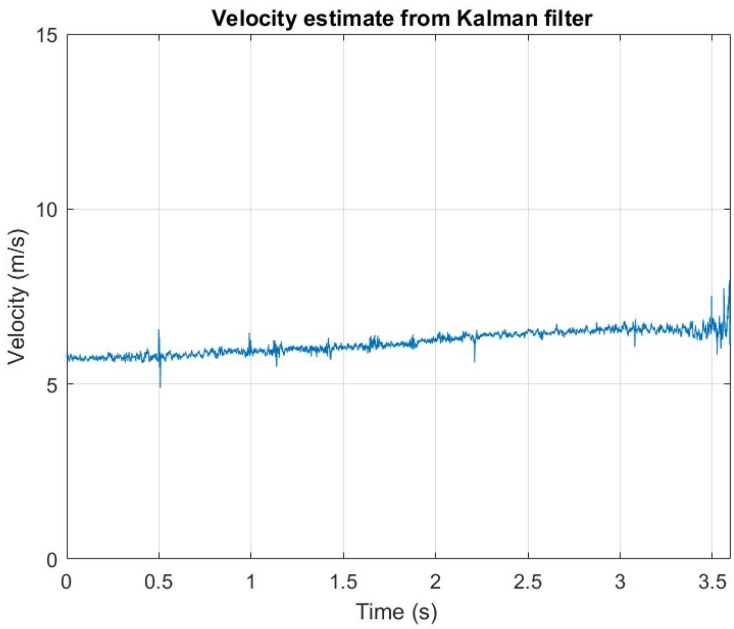

Figure 20: Velocity estimate over time with Kalman filtering

In Figure 19 it can be seen how the coarse position estimate leads to a very abrupted velocity estimate, which can't match with the real velocity of the car. Therefore, the usage of the wireless radio data is necessary to get a plausible estimate.

To check the accuracy of the car velocity estimate, the average estimated velocity from Figure 20 has been computed, which has turned out to be $6.1414\ \frac{\mathrm{m}}{\mathrm{s}}$. This means that in $3.6\ \mathrm{s}$ the distance covered is $22.1\ \mathrm{m}$. Now, the distance from the light barrier until the last frame (which is placed at 2 m distance), has turned out to be $22\ \mathrm{m}$ which is quite accurate.

3.1.3.2. Estimate with extended Kalman filter

Once seen that Kalman filtering has a promising performance, some improvements have been introduced.

One of the techniques used in the last approach has been the usage of the coarse position estimate to compute the angles. For improving this, an extended Kalman filter has been developed. The objective of this technique is to estimate in every iteration α and β defined in Figure 11 from the finely estimated position estimate. The extended Kalman filter is defined by the model expressed in Table 4.

As we did in Section 3.1.3.1 parameters must be adjusted, and it has been done as summarized in Table 7. Parameters from Table 6 have been kept with the same value, but new parameters have been added.

Parameter	Value
$x\ axis$ state variance (${\sigma_{w1}}^2$)	0.01
$y\ axis$ state variance (${\sigma_{w2}}^2$)	0.2
Velocity state variance (${\sigma_{w4}}^2$)	$1e-3$
Angles state variance (${\sigma_{w5,6}}^2$)	$1e-4$
$x\ axis$ measurement variance (${\sigma_{n1}}^2$)	0.5
$y\ axis$ measurement variance (${\sigma_{n2}}^2$)	0.4
Frequency measurement variance (${\sigma_{n4}}^2$)	700
Frequency measurement variance (${\sigma_{n5}}^2$)	200
Angles measurement variance (${\sigma_{n6,7}}^2$)	$1e-4$

Table 7: extended Kalman filter parameters

It must be remarked from the parameter's selection that the variances of both angles are the same for both equations. This has been done to assume the same degree of ignorance for both coarse position estimate and transition equation.

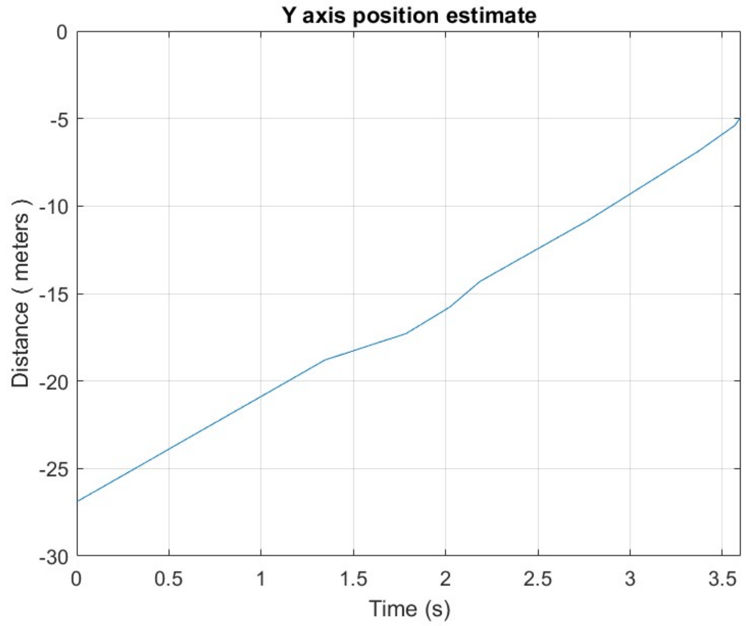

Figure 21: distance estimate along y axis from transmitter to receiver

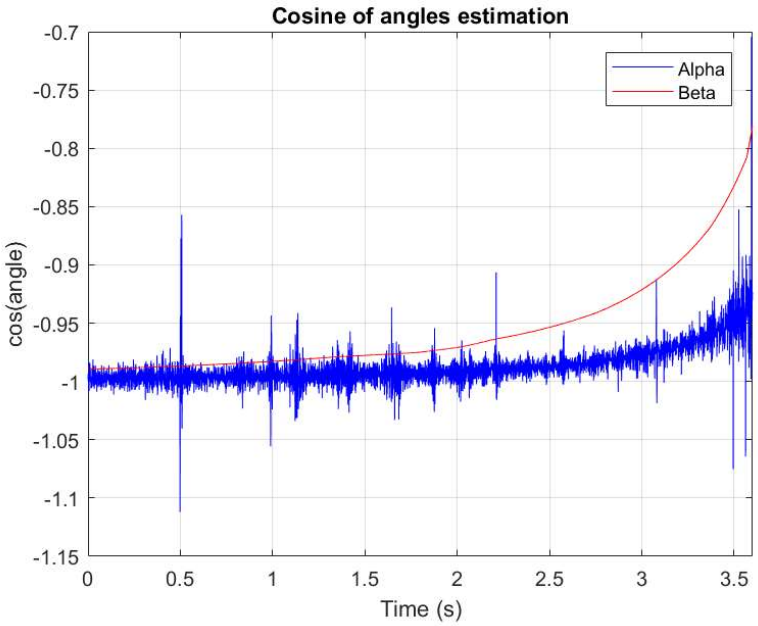

Figure 22: estimate of the cosines of α and β over time using extended Kalman filter

Figure 21 and Figure 22 show the estimate of the y axis position estimate and α and β angles respectively. It can be extracted from Figure 21 that the y axis position estimate is quite linear. This was expectable as long as in Figure 20 the velocity estimate was quite steady. In Figure 22 the estimate of α and β over time is plotted. Angle β is well estimated. However, angle α is noisy. This happens because the movement along x axis is assumed to change during time due to a not completely straight-line driving. The frequency measurements error is translated to an estimate error of α.

A comparison of the performance of Kalman filtering and extended Kalman filtering is done in Figure 23, where both velocities estimates are plotted to appreciate better the differences.

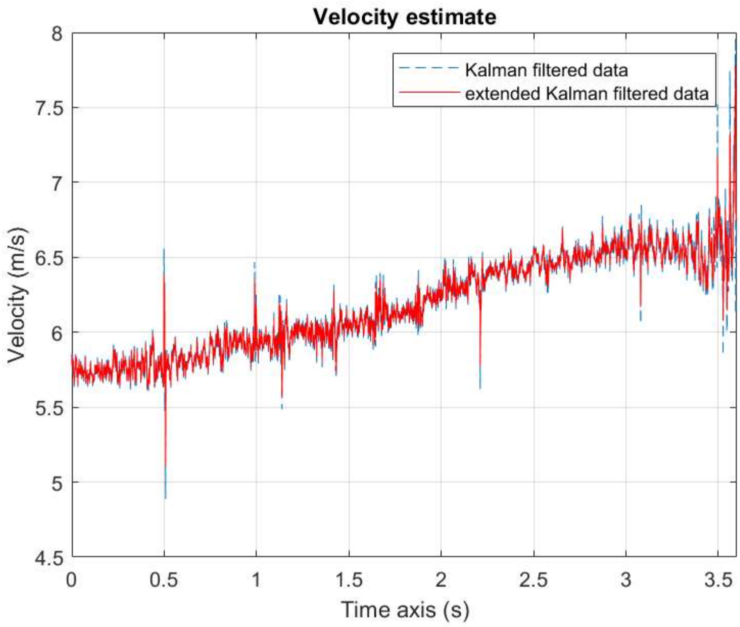

Figure 23: velocity estimate with Kalman filter and with EKF

As it can be seen in Figure 23 both estimates are similar. However, as EKF estimates also angles α and β, it can correct the error introduced by the computation of the angles and it is less sensitive to sudden changes. To get more insights about how much do the estimates differ, the mean square root difference has been computed. This metric has been defined as follows

$$\mathrm{diff}_{\mathrm{RMS}} = \sqrt{\frac{\sum_{k=0}^{S-1}(v_{y,KF,k} - v_{y,EKF,k})^2}{S}}. \tag{20}$$

Where $v_{y,KF,k}$ stands for every segment of the estimated velocity using Kalman filtering and $v_{y,EKF,k}$ stands for every segment of the estimated velocity using extended Kalman filtering.

The result of computing (20) is $0.05\ \mathrm{m/s}$, which is due mainly to a better correction in the peaks using extended Kalman filtering.

3.2. Applications of vehicle tracking

Position estimate from Kalman filtering allows to transform the time index into a position index which was not possible so far in [6]. This allows for instance the plot of the received signal-to-noise ratio (SNR) over the estimated distance, as it is done in Figure 24.

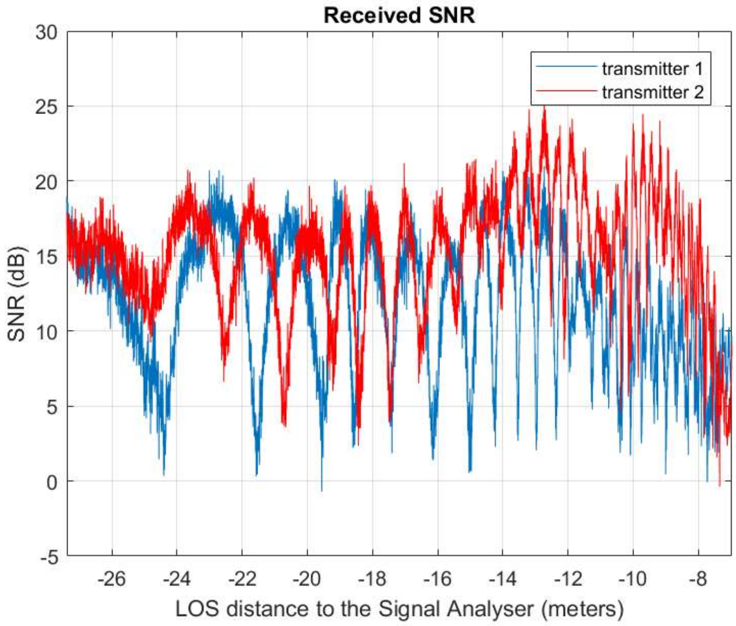

Figure 24: plot of the SNR over the line-of-sight estimated distance

The SNR in Figure 24 is shown for each of the two transmitters over their estimated visual line-of-sight (LOS) distance to the receiver. Results differ for both transmitters. Firstly, one reason for having these different results is that the different beam elevations of the antennas caused that destructive interference took place at different positions for each transmitter. It is interesting to see how this effect has influence also in the frequency estimate in Figure 16. Secondly, it can also be observed that the 15º elevation beam of transmitter 2 provides on average higher SNR when the car approaches the receiver. Spatial filtering near the receiver is more relevant for transmitter 1 and the corresponding received signal gets weaker.

It is also interesting to get more insights about the signal-power-to-noise ratio (SNR) depending on the estimated distance. For this, in Figure 25 a coverage map has of the road has been built. It must be noticed that, as only 108 frames were taken, 41 segments are averaged for SNR per frame to reach alignment.

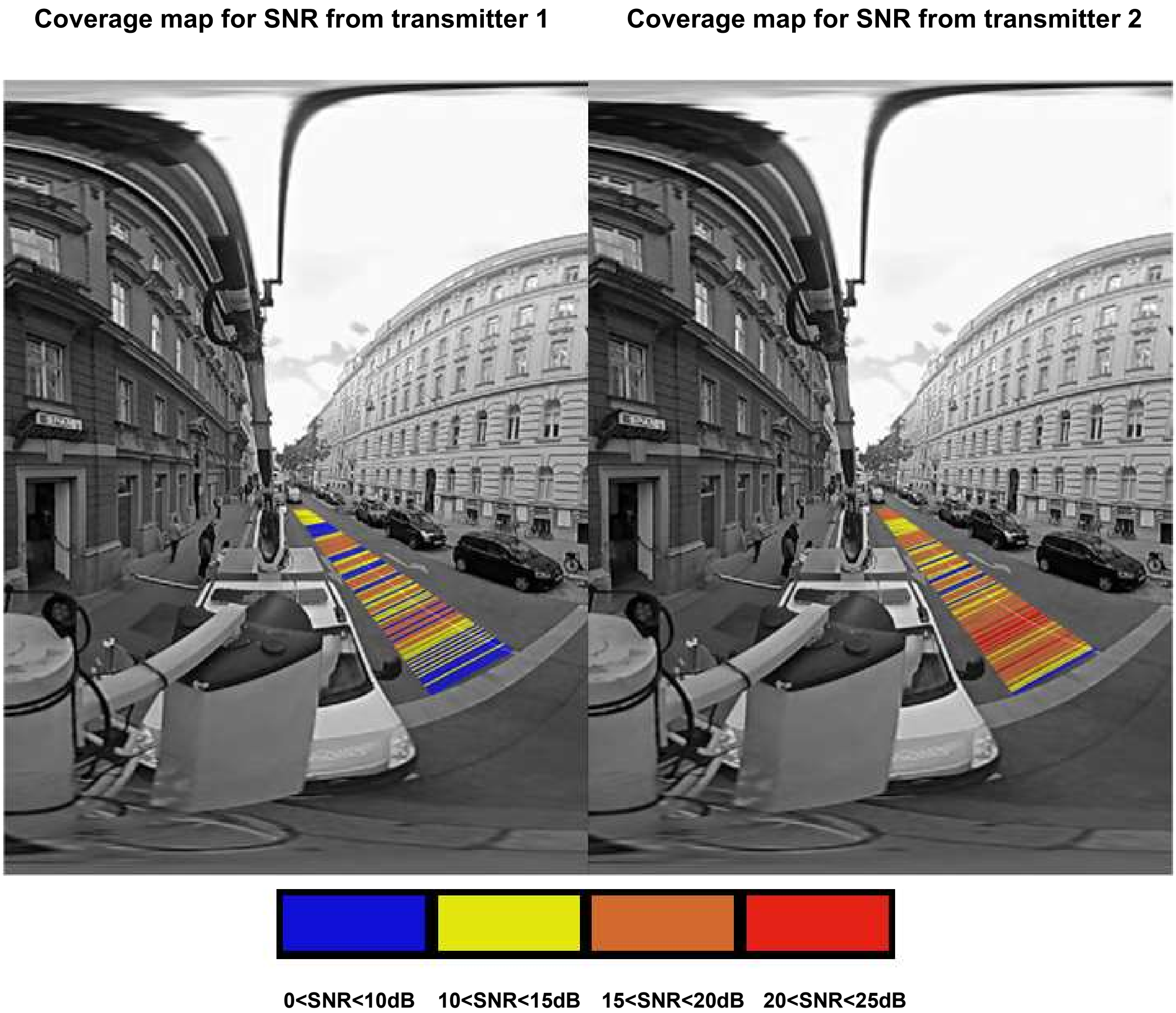

Figure 25: in the left-hand side, SNR coverage map for transmitter 1. In the right-hand side coverage map for transmitter 2

Figure 25 shows how results in SNR differ between transmitters as it was explained in Figure 24.

Another interesting result to compute is the SNR with selection diversity, which is taking the sample of the signal with most power of both transmitters.

To illustrate this, Figure 26 shows the selection diversity plotted together with the SNR for each of the transmitters.

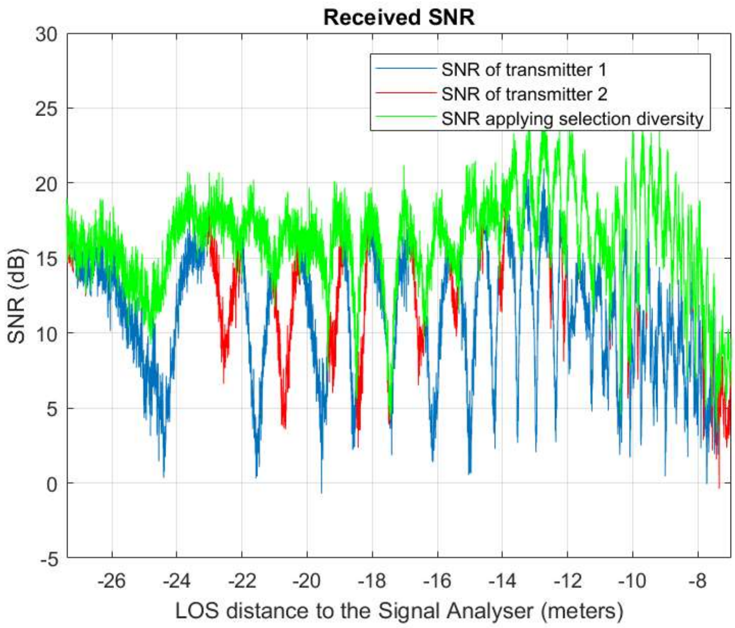

Figure 26: selection diversity plot with both SNR estimates from both transmitters

It can be seen in Figure 26, that the 15º beam elevation of transmitter 2 is less susceptible to fading due to spatial filtering, which means that the signal is better on average. Therefore, as destructive interferences took place in different positions for each transmitter, selection diversity avoids these interferences, having always a good SNR.

Finally, the same overlay method as shown in Figure 25, is now done for SNR with selection diversity in Figure 27.

Coverage map for SNR with selection diversity

0<SNR<10dB 10dB<SNR<15dB 15dB<SNR<20dB 20dB<SNR<25dB

Figure 27: coverage map of selection diversity plot

4. Conclusions and future development

The aim of this thesis was to develop a system able to track the position and velocity of a moving car in an urban environment with two data sources: a camera and a wireless communication link formed by a transmitter array on top of a car and a signal analyser (SA) acting as receiver. The system has been evaluated with measurements taken place in Vienna downtown the 25th of September 2018. For this, a 360º camera has been chosen and the communication link employed on the transmitter side an antenna array of 2 directional elements and on the receiver side an omnidirectional antenna. The communication link operated in the V-frequency band.

For developing this system, an estimate from each source has been done in an independent way. From camera data side, two different stages have been used. For isolating the car, a Gaussian mixture model for background extraction has been implemented followed by an erosion. These techniques have been successful as I was able to isolate the car in the region of interest.

Parallel to this approach, calibration images have been taken to get references of the distance between points on the road and the receiver. Also, these images have been used to get the distance of some points to the SA through epipolar geometry. Both techniques have allowed to make a position estimate of the car of interest.

From the communications link side, using radio data from [6], a Doppler shift estimation has been obtained for 2 transmitters. Both estimates have been fed as inputs for a Kalman filter to track position and velocity of the car.

As shown in the results chapter, the usage of this system has provided an accurate estimate of the velocity as well as the car position. Furthermore, using the data provided by [6], the received SNR has been printed along the road. We have also seen that EKF gives a slightly better performance than Kalman filtering.

As further development of this thesis there are many issues that could be interesting to face in the long term. However, I think that the most crucial ones are a different image processing approach and a study of the antenna orientation relating it with the received power would be really interesting. For camera data, it would be intriguing to observe the position estimation using for example CNN or Ada-Boost algorithm, that where mentioned in the state of the art. From fine position estimate, taking into account that coarse position estimation may not have a Gaussian noise, it would be interesting to evaluate the performance of other kind of filters such as particle filters.

As a final opinion of this thesis, I think that the goal was ambitious in the sense of having had to work directly with measurements and also having to face computer vision and image segmentation issues what were completely new and we had to learn from scratch. Nonetheless, a complete solution has been set out and some paths have been considered to keep researching.

5. Bibliography

[1] A. Haselhoff et al. «Radar-vision fusion with an application to car-following using an improved adaboost detection algorithm» *In Proceedings of Intelligent Transportation Systems Conference*, 2007.

[2] L. Huang and M. Barth, «Tightly-coupled lidar and computer vision integration for vehicle detection» *In Intelligent Vehicles Symposium*, 2009.

[3] D. Jiang and L. Delgrossi, «IEEE 802.11p: Towards an International Standard for Wireless Access in Vehicular Environments» *In Proceedings of VTC Spring 2008-IEEE*, Singapore, 2008.

[4] Choi et al. «Millimeter-wave vehicular communication to support massive automotive sensing» *IEEE Communications Magazine, 54, 12*.

[5] T. Litman, «Autonomous vehicle implementation predictions. Implications for transport planning» *Victoria Transport Policy Institute,* 2018.

[6] H. Groll et al. «Sparsity in the delay-Doppler domain for measured 60 GHz vehicle-to-infrastructure communication channels» arXiv: 1901.10817, Jan 2019.

[7] M. Sepulcre and J. Gozalvez, «On the importance of radio channel modeling for the dimensioning of wireless vehicular communication systems» *In Proceedings of 7th International Conference on ITS*, 2007.

[8] P. Alexander et al. «Cooperative intelligent transport systems: 5.9-GHz field trials» *Proceedings of the IEEE,* vol. 99, nº 7, pp. 1213-1235, 2011.

[9] P. Papadimitratos et al. «Architecture for secure and private vehicular communications» *In Proceedings of 7th International Conference on ITS*, 2007.

[10] G. Liu et al. «A routing strategy for metropolis vehicular communications» *In Proceedings of International Conference on Information Networking*, Berlin, 2004.

[11] P. Papadimitratos et al. «Vehicular communication systems: enabling technologies, applications, and future outlook on intelligent transportation» *IEEE Communications Magazine, 47,* pp. 74-85, 2009.

[12] R. Verdone «Outage probability analysis for short-range communication systems at 60 GHz in ATT urban environments» *IEEE Transactions on Vehicular Technology,* vol. 46, nº 4, pp. 1027-1039, 1997.

[13] P. F. Smulders and L. M. Correia, «Characterisation of propagation in 60 GHz radio channels» *Electronics & Communication Engineering Journal,* vol. 9, nº 2, pp. 73-80, 1997.

[14] J. Choi et al. «Millimeter-wave vehicular communication to support massive automotive sensing» *IEEE Communications Magazine,* vol. 54, nº 12, pp. 160-167, 2016.

[15] N. Garcia et al. «Location-aided mm-wave channel estimation for vehicular communication» *In Proceedings of 17th Signal Processing Advances in Wireless Communications,* pp. 1-5, 206.

[16] T. S. Rappaport et al. «Millimeter wave mobile communications for 5G cellular: It will work!» *IEEE Access,* vol. 1, nº 1, pp. 335-349, 2013.

[17] A. Paier et al. «Characterization of vehicle-to-vehicle radio channels from measurements at 5.2 GHz» *Wireless Personal Communications,* vol. 50, nº 1, pp. 19-32, 2009.

[18] C. F. Mecklenbräuker et al. «Vehicular channel characterization and its implications for wireless system design and performance» *Proceedings of the IEEE,* vol. 99, nº 7, pp. 1189-1212, 2011.

[19] M. G. Sánchez et al. «Millimeter wave radio channel characterization for 5G vehicle-to-vehicle communications» *Measurement,* nº 95, pp. 223-229, 2017.

[20] E. Zöchmann et al. «Measured delay and Doppler profiles of overtaking vehicles at 60 GHz» *In Proceedings of the IEEE 20th APS Topical Conference on Antennas and Propagation in Wireless Communications* 2018.

[21] E. Zöchmann et al. «Statistical evaluation of delay and Doppler spread in 60 GHz vehicle-to-vehicle channels during overtaking» *In Proceedings of IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications*, 2018.

[22] X. Liu et al. «On-road vehicle detection fusing radar and vision» *In Proceedings of Vehicular Electronics and Safety (ICVES)*, 2011.

[23] A. Asvadi et al. «Multimodal vehicle detection: fusing 3d-lidar and color camera data» *Pattern Recognition Letters,* nº 115, pp. 20-29, 2018.

[24] D. Neumann et al. «Online vehicle detection using Haar-like, LBP and HOG feature based image classifiers with stereo vision preselection» *In Intelligent Vehicles Symposium*, Redondo Beach, CA, USA, 2017.

[25] W. Song et al., «Real-time lane detection and forward collision warning system based on stereo vision» *Intelligent Vehicles Symposium*, Redondo Beach, CA, USA, 2017.

[26] G. FullDome, «GoXtreme FullDome 360º» [online]. Available: http://www.goxtreme-action-cams.com/en/goxtreme-full-dome-360/. [Last access: 10 January 2019].

[27] D. A. Forsyth and J. Ponce, « A modern approach» Computer vision: a modern approach, pp. 21-48, 2003.

[28] R. Hartley and A. Zisserman, «Multiple view geometry in computer vision» Cambridge: Cambridge University Press, 2004.

[29] A. Nurhadiyatna et al. «Background subtraction using Gaussian mixture model enhanced by hole filling algorithm (gmmhf)» *In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2013.

[30] C. Stauffer and W. E. L. Grimson, «Adaptive background mixteure models for rea-time tracking» *In Proceedings of CVPR*, Massachusets, 1999.

[31] Z. Zivkovic, «Improved adaptive Gaussian mixture model for background subtraction» *In Proceedings of ICPR*, Cambridge, 2004.

[32] MathWorks, «Computer vision system toolbox (2017b)».