Vehicle Tracking through Vision-Millimeter Wave Doppler Shift Fusion

Martín Trullenque Ortiz, Herbert Groll, Erich Zöchmann, Christoph F. Mecklenbräucker

Institute of Telecommunications, TU Wien, Vienna, Austria, *trullenqueortiz@gmail.com*

***Abstract*—Automated cars benefit greatly from millimeter wave broadband communication links. Also, computer vision is becoming more and more used in automotive applications. In this scenario, we propose a system capable to track the position of a given car on a road by fusing data from a camera system and a wireless radio link at 60 GHz. A Gaussian mixture model and epipolar geometry have been used for computer vision. From a V-band wireless link a Doppler shift estimate has been obtained. The effectiveness of our method is shown on measurement data.**

Keywords—millimeter wave, computer vision, data fusion

I. INTRODUCTION

Automated cars will be a short-term reality. Past research has mainly focused on computer vision, radar [1] and on the the vehicular communications standard IEEE 802.11p.

The time-critical nature of road safety applications has imposed the need to accurately design the wireless vehicular communication systems [2]. This need has shown that vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication should make use of millimeter waves (mm-wave) [3-5]. In [4], a location of a car relative to a base station (BS) helps to infer angle-of-arrival (AOA) and angle-of-departure (AOD) estimation and improve beamforming. In [6], analysis and characterization of all elements taking part in a communications link at 5.9 GHz in V2X environment are performed. For the 60 GHz band, measurements of cars driving at constant speed and a fixed distance have been performed in [7]. In [8] and [9] 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], [10]. In [1], a radar system is combined with a camera applying the Ada-Boost algorithm to perform car detection. Other ranging methods such as lidar are used for identifying the region of interest (ROI). For instance, in [11], a lidar structure is used to define the ROI and then, as well as in [1], Ada-Boost algorithm is used. In [12], 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 [10]; in that case, the radar itself is not used to estimate the velocity of the car.

In this paper, using a 360º video camera and wireless receiver data coming from a radio link at 60 GHz, a moving car is tracked during the measurement campaign.

This paper is structed as follows. In Section II we describe the measurements performed and the signal processing. In Section III we explain the different processing tools used to achieve our results, which will be detailed in Section IV.

II. MEASUREMENT CAMPAIGN

The measurements were performed in Vienna downtown on 25th September 2018, for details see [13]. The wireless communications link consisted of two TX antennas placed on top of a moving car and a receiver placed close to a street crossing on traffic light height. One TX had 0º beam elevation while the other one had 15º beam elevation. The parameters of the wireless system are summarized in Table 1.

TABLE I. WIRELESS LINK COMMUNICATION PARAMETERS

Parameter	***Value***	***Parameter***	***Value***
centre frequency	$f_{\mathrm{TX}} = 60.15$ GHz	TX antennas	20 dBi conical horn
number of TX antennas	2	RX antenna	omni-directional
bandwidth	100 MHz	snapshot rate	$T_{\mathrm{snap}} = 0.83$ms

From the communication link an estimate of the received power from antenna *i* at all snapshots *0,..S-1* is obtained $\underline{P_{R,i}} = [P_{R,i,0}, \dots, P_{R,i,S-1}]$. Equally, an estimate of the noise power $\underline{P_n} = [P_{n,0}, \dots, P_{n,S-1}]$ is obtained. Both parameter estimates lead us to define the SNR for each transmitter as follows:

$$\underline{SNR_i} = \underline{P_{R,i}} / \underline{P_n}. \quad (1)$$

Moreover, also using the techniques described in [13] a Doppler shift estimate is obtained for every snapshot $\underline{\hat{f}_{\mathrm{Dopp,i}}} = [\hat{f}_{\mathrm{Dopp,i,0}}, \dots, \hat{f}_{\mathrm{Dopp,i,S-1}}]$, which allows to estimate the radial velocity of the car.

On the vision side we have a 360º camera placed next to the receiver. The camera has been the GoXtreme Full Dome 360º. In order to align both data sources a light barrier has been placed 24 m away from the receiver. When the car goes through the light barrier the recording of the wireless channel starts and a LED array flashes up to align recorded video frames with the radio signal. A schematic of the measured scenario can be seen in Fig. 1 and in Fig. 2 the alignment strategy is illustrated.

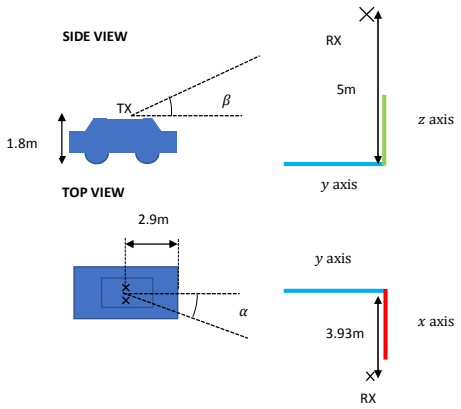

Fig. 1: side and top view as schematic of the measurement scenario with defined coordinate system

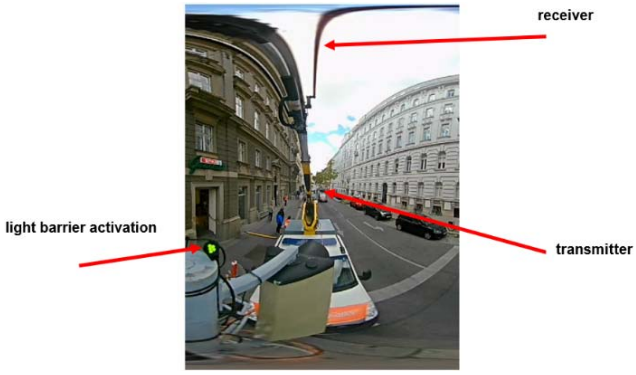

Fig. 1: synchronization of vision and radio data by means of a light barrier and optical trigger

III. Signal Processing

A. Image processing techniques

In order to get a position estimate from the video camera, several image processing tools have been used. Firstly, to isolate the car form the rest of the image, a Gaussian mixture model is used for background estimation. The implementation chosen has been the one provided by Matlab's (2017b) computer vision system toolbox. The chosen parameters are shown in Table II.

TABLE II. BACKGROUND ESTIMATION DESIGN PARAMETERS

Parameter	*Value*
number of Gaussians	3
minimum background ratio	0.7
learning rate	0.005
initial covariance	30
number of training frames	20

The number of Gaussians and the number of training frames have been adjusted empirically. All others are default values. The result after background estimation can be observed in Fig. 3.

From the background estimated frame, a position estimate is performed. For the position estimation depth information is needed. For that, calibration pictures with reference points were taken. Looking for these points' correspondences on the video frames we were able to compute the fundamental matrix. The correspondences are shown in Fig. 4.

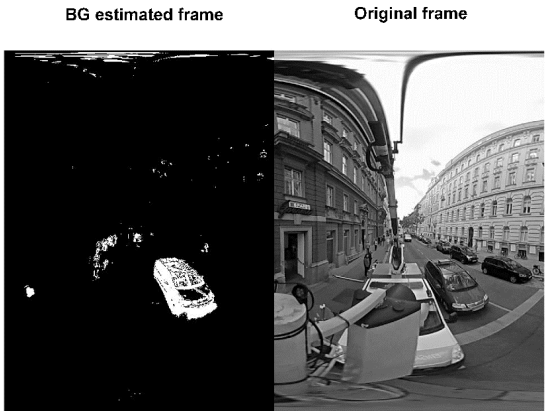

Fig. 2: background estimated frame

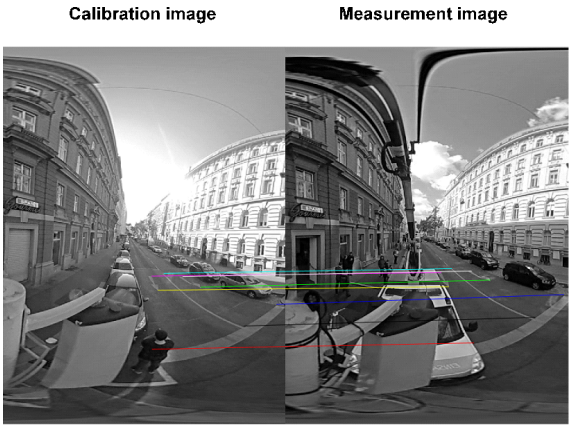

Fig. 3: correspondences of points in calibration (left-hand side) and measurement (right-hand side) images

Fig. 4: coarse position estimate

For the calibration images we took reference points at known distance. This allowed us, through the fundamental matrix, to compute the correspondences on the video frames and then to estimate the depth. The position estimate is updated once the car is at the reference position shown in Fig. 4. The position values in between are interpolated.

B. Wireless channel data processing

As described in Section II, from the wireless link an estimate of the received power contribution for each transmitter, noise contribution and a Doppler shift estimate are obtained. This last estimate gives us an estimate of the radial

velocity of the car. However, our target is to estimate the velocity along the y axis, for which we use the following equation.

$$f_{\text{Dopp}} = \frac{f_{\text{TX}}}{c} v_y \cos\alpha \cos\beta. \quad (2)$$

From (2) we have to know the angles α and β defined in Fig.1. For computing these angles, we use the position estimate obtained from the video frames.

C. Data fusion

In this contribution data fusion is implemented via the extended Kalman filter. The input to the filter are $\hat{x}_{\text{pos}}$ and $\hat{y}_{\text{pos}}$, the estimated coordinates from the coarse position estimate, as well as the cosine of angles α and β estimated applying trigonometric ratios. Finally, the last inputs are the Doppler frequency shift estimates from each transmitter. It is necessary to remark that the time between radio data segments is not the time between video frames. The snapshot rate of the wireless link is $T_{\text{snap}} = 0.83$ ms and between video frames it is $1/30$s.

The parameters to estimate in the filter have been x and y position ($\underline{x}_{\text{pos}}$ and $\underline{y}_{\text{pos}}$ respectively), the velocity along y axis v_y, and $\cos\alpha$ and $\cos\beta$.

Our notation for the Kalman filter is as follows. The transition equation is defined by

$$\underline{x}_{k+1} = \underline{\underline{A}}\,\underline{x}_k + \underline{w}, \quad (3)$$

where $\underline{\underline{A}}$ is the parameter evolution matrix modeling the dynamics of the system. 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 measurements equation is given by

$$\underline{z}_k = g(\underline{x}_k) + \underline{n}, \quad (4)$$

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

Since $g(.)$ is a non-linear function, it is linearized by

$$\underline{\underline{G}}_k = \left[\frac{\partial g(\underline{x}_k)}{\partial x_{k,1}}, \frac{\partial g(\underline{x}_k)}{\partial x_{k,2}}, \frac{\partial g(\underline{x}_k)}{\partial x_{k,4}} \frac{\partial g(\underline{x}_k)}{\partial x_{k,4}}, \frac{\partial g(\underline{x}_k)}{\partial x_{k,5}}, \frac{\partial g(\underline{x}_k)}{\partial x_{k,6}}\right] \quad (5)$$

The details for the extended Kalman filter implementation are shown in in Table III.

TABLE III. EXTENDED KALMAN FILTER PARAMETERS

Value
$\underline{x}_k = [x_{\text{pos},k} \quad y_{\text{pos},k} \quad z_{\text{pos},k} \quad v_{y,k} \quad \cos\alpha_k \quad \cos\beta_k]^T$
$\underline{\underline{A}}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & a(\underline{x}_k) & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
$\underline{\underline{G}}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial f_{Dopp,1,k}}{\partial x_{k,4}} & \frac{\partial f_{Dopp,1,k}}{\partial x_{k,5}} & \frac{\partial f_{Dopp,1,k}}{\partial x_{k,6}} \\ 0 & 0 & 0 & \frac{\partial f_{Dopp,2,k}}{\partial x_{k,4}} & \frac{\partial f_{Dopp,2,k}}{\partial x_{k,5}} & \frac{\partial f_{Dopp,2,k}}{\partial x_{k,6}} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
$\underline{z}_k = [\hat{x}_{\text{pos},k} \quad \hat{y}_{\text{pos},k} \quad \hat{z}_{\text{pos},k} \quad f_{\text{Dopp},1,k} \quad f_{\text{Dopp},2,k} \quad \cos(\hat{\alpha}_k) \quad \cos(\hat{\beta}_k)]^T$
$\underline{\underline{W}} = E[\underline{w}\underline{w}^H]$ $\quad$ $a(\underline{x}_k) = \frac{T_{\text{seg}}}{\underline{x}_{5,k}\underline{x}_{6,k}}$
$\underline{\underline{C}} = E[\underline{n}\underline{n}^H]$ $\quad$ $\frac{\partial f_{\text{Dopp}}}{\partial x_{k,4}} = \frac{f_{Tx}}{c}\underline{x}_{k,5}\underline{x}_{k,6}$
$\underline{\underline{P}}_k = (\underline{\underline{I}}_2 - \underline{\underline{K}}\underline{\underline{G}})\,\underline{\underline{\hat{P}}}_k$ $\quad$ $\frac{\partial f_{\text{Dopp}}}{\partial x_{k,5}} = \frac{f_{Tx}}{c}\underline{x}_{k,4}\underline{x}_{k,6}$
$\underline{\underline{K}}_k = \underline{\underline{P}}_k\,\underline{\underline{G}}^H(\underline{\underline{C}} + \underline{\underline{G}}\,\underline{\underline{P}}_k\,\underline{\underline{G}}^H)^{-1}$ $\quad$ $\frac{\partial f_{\text{Dopp}}}{\partial x_{k,6}} = \frac{f_{Tx}}{c}\underline{x}_{k,5}\underline{x}_{k,4}$

IV. RESULTS

With the implementation of the extended Kalman filter detailed above, we have evaluated the tracking of the car during the measurements campaign. The measured scenario and the measurement setup are as in [13]. The only difference from [13] is the additional usage of the video data. As explained in Section II, to synchronize both the camera and the wireless link data a light barrier has been set 24 m away from the receiver position. The RX and the 360º camera are placed close to a traffic light, 5 meters above the road level. The recording length of the wireless link has been set to 3.6 s, which let us acquire 4298 snapshots. Within that time, 108 video frames are capture by the camera.

For performing the tracking of the car, many parameters of the extended Kalman filter have been adjusted. These parameters have been set empirically, and are detailed in Table IV.

TABLE IV. EXTENDED KALMAN FILTER PARAMETERS SET UP

Parameter	***Value***
x axis state variance (${\sigma_{w1}}^2$)	0.01
y axis state variance (${\sigma_{w2}}^2$)	0.2
Velocity state variance (${\sigma_{w4}}^2$)	$1\text{e}-3$
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 (${\sigma_{n6,7}}^2$)	$1e-4$

Fig. 6 shows the estimated velocity obtained by the vision system alone and compares it to the obtained estimate by the proposed sensor fusion. The driver of the car kept the velocity as constant as possible, which is also confirmed by the obtained sensor fusion result. The obtained results from the video data is very unreliable due to the usage of only 8 reference points. The reason why only 8 reference points have been considered is the lack of reliability of other references in the calibration images. Then, in order to estimate the position, we have checked for each frame whether the car had gone through any of the 8 reference points or not.

To check the accuracy of the system, the average velocity is computed. It is $6.1414\ \frac{\mathrm{m}}{\mathrm{s}}$, which means that in 3.6s the car has covered $22.1\ m$. To evaluate how good this performance is, we must notice that taking the distance information from the video frames, in these 108 frames the car has covered $22\ m$, which makes our estimate very accurate.

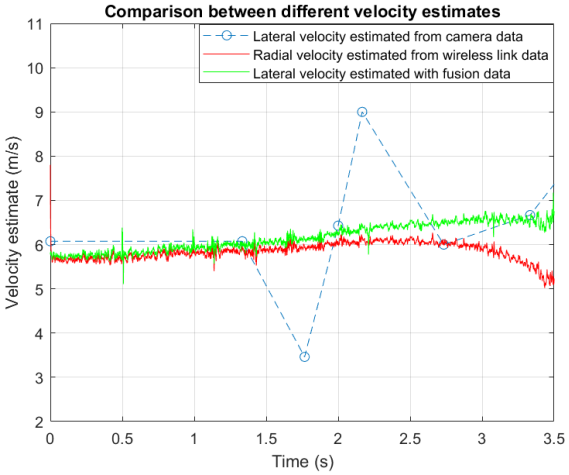

Fig. 6: velocity estimate using extended Kalman filter

This position estimate now allows us to plot the SNR from each transmitter over distance, as it can be seen in Fig. 8. Fig. 8 shows a coverage map of the road. The results differ for both transmitters. Firstly, one reason for having these different results is that the different beam elevation causes that destructive interference took place at different positions for each transmitter. 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 must be noticed that, as only 108 frames were taken, 41 segments are averaged for SNR per frame to reach alignment.

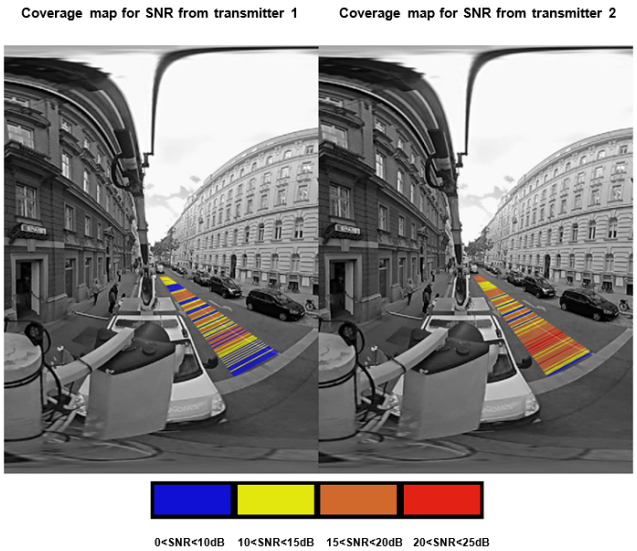

Fig. 8: Coverage map of the received signal.

V. CONCLUSIONS

In this paper we have described the performance of a system able to track the position and velocity of a moving car in an urban environment using two data sources: a video camera and a wireless communications link at 60 GHz. For developing this system both data sources have been fused using an extended Kalman filter. Using the data provided by [13] the received SNR has been shown along the road.

REFERENCES

[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] 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.

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

[4] 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, 2006.

[5] 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.

[6] 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

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

[8] 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.

[9] 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

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

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

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

[13] H. Groll et al. «Sparsity in the delay-Doppler domain for measured 60 GHz vehicle-to-infrastructure communication channels» *In Proceedings of IEEE International Conference on Communications (ICC), 2019.*