Video-assisted Overtaking System Enabled by C-V2X Mode 4 Communications

Pedro Magalhães*‡, Pedro M. d'Orey†‡, Maria Teresa Andrade‡, and Francisco Castro*
*Bosch Car Multimédia Portugal, S.A., Braga, Portugal
†Instituto de Telecomunicações, Porto, Portugal
‡Universidade do Porto, Porto, Portugal
Email: {Pedro.Magalhaes2, Francisco.Castro2}@pt.bosch.com, pedro.dorey@dcc.fc.up.pt, mandrade@fe.up.pt

***Abstract*—The overtaking maneuver can be often dangerous and risky, especially when driving in rural roads, with a significant probability of leading to fatalities and serious injuries. Vehicle communication emerges as one of the leading technologies to improve the driving experience and road safety. This article presents a system designed to augment road safety in an Overtaking Maneuver, providing the ahead vehicle's Line-of-Sight by real-time video-streaming. It uses the C-V2X Transmission Mode 4 that, building upon cellular technology, has evolved as one of the main standards for vehicle communication. Results show an effective application of the system in real overtaking scenarios, enhancing the safety and proving its usefulness. Whilst having some impact on the latency, the solution incurred in low packet losses and provided good video quality.**

I. Introduction

Road traffic accidents have profound socio-economic impacts leading each year to millions of fatalities and serious injuries, according to the World Health Organization (WHO) [1]. Accidents are caused mostly by human error, namely driver inattention, lack of situation awareness, or misunderstanding the intentions of other road users. In the European Union, 54% of the fatalities occur in rural roads (non-motorway) according to the Annual Accident Report [2], where relative speeds between vehicles are higher, leading to frequent overtaking and other potentially dangerous maneuvers.

Several strategies have been adopted to improve road safety in rural environments, namely additional signing, modification of road layouts/infrastructure, installation of traffic lights, among others. Regarding technological solutions, several works [3] [4] [5] have been devoted to improving the safety of overtaking maneuvers. For instance, the *See-Through System* [3] assists drivers in the overtaking decision by providing road ahead conditions information (video feed of vehicle being overtaken) resorting to 802.11p V2V communications. However, previous works did not perform active *traffic scene understanding* (e.g. [6]) for improved safety, neither proposed an *automatic* overtaking system.

This work was supported in part by the Project MobiWise under grant POCI-01-0145-FEDER-016426, funded by the FEDER (European Regional Development Fund) through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) and by national funds through Fundação para a Ciência e Tecnologia, and grant UIDB/50008/2020 funded by FCT/MCTES through national funds and when applicable co-funded EU funds.

In the last decade, impressive developments have been made in 2D/3D object detection techniques resorting to a plethora of sensors, namely monocular or stereo cameras [7] or LIDAR [8], which allow improving scene understanding. In parallel, recent developments in vehicular networking have enabled the exchange of information between vehicles using 802.11p, namely of processed sensor data [e.g., cooperative awareness messages (CAM) [9]] or raw sensor information (e.g., video stream). More recently, LTE-V2X [10] has been presented as a strong contender for vehicular communications with the recent commercial availability of 3GPP release 14 devices. Specifically, C-V2X mode 4 allows vehicles to communicate directly without the support of cellular infrastructure. Nonetheless, few studies have *empirically* assessed the performance of C-V2X mode 4 networks and related applications.

In this paper, we propose an efficient and safe overtaking system enabled by video-streaming over C-V2X vehicular networks. Improved traffic scene understanding is achieved resorting to state-of-the-art *3D object detection techniques*, which assists in detecting unsafe overtaking maneuvers, thus reducing serious accidents due to lack of situational awareness. In addition, the proposed system *automatically* determines the start and end of (potential) overtaking maneuvers resorting to internal (i.e. GPS and CAN data) and external information (i.e. CAMs from other vehicles) prior to video streaming between vehicles. The system functioning is enabled by the exchange of status (i.e. CAM) and control messages (e.g. *play* message) through V2V communications. In addition, we present a preliminary study on the feasibility of delivering this application with varying demand over C-V2X networks that have been designed for handling periodic traffic.

To summarize, the main contributions of the paper are:

- definition of an efficient and safe overtaking system enabled by C-V2X networks and video-streaming in combination with object detection for improved safety;
- performance assessment of the system in a real-world scenario, including the empirical evaluation of the transmission of variable traffic over C-V2X mode 4 networks.

The remainder of this paper is organized as follows. The relevant state-of-the-art is presented in Section II. The proposed system is described in Section III. The system evaluation is given in Section IV. Section V provides the main conclusions.

II. Related Work

Advanced Driver Assistance Systems (ADAS). V2X communications allow improving the driving experience and safety by augmenting the information available to the driver. Several ADAS resort to video streaming over V2X wireless networks for their operation, namely teleoperated driving of vehicles from a remote control center [11]. Other works (e.g. [3], [5], [12]) use video streaming to assist drivers in overtaking maneuvers in challenging scenarios (e.g. rural roads). The See-Through System [3] [5] - an ADAS for assisting driver's on overtaking maneuvers - leverages on video streaming over V2V networks to transform vision-obstructing vehicles into transparent representations through the projection of received information into the windshield. Bohm et al. [4] proposed an overtaking assistance system solely based on adaptive cooperative awareness messaging. Pereira et al. [12] empirically evaluated the performance of an overtaking assistance system enabled by 802.11p networks in urban and highway scenarios.

Our work differentiates from other previous works in a number of ways. First, we propose a system that automatically detects overtaking maneuvers and triggers/terminates different system components resorting to CAMs and other sensor information (e.g. CAN bus). Second, we assess the feasibility of C-V2X mode 4 communications instead of 802.11p networks to enable efficient video streaming in an overtaking assistance system given the Semi-Persistent Scheduling scheme that has been specifically designed for the transmission of periodic traffic. In addition, we resort to object detection techniques combined with tracking to increase the safety of the system.

Video streaming over wireless networks. Video Streaming is a highly demanding service, being low tolerant to packet loss. Real-Time Streaming Protocol (RTSP), Real-Time Messaging Protocol (RTMP), and MPEG-Dash (Dynamic Adaptive Streaming) are among the most used video streaming protocols in wireless networks, which suffer from more challenging conditions when compared with wired networks. Aloman et al. [13] examined the performance of RTSP, RTMP, and MPEG-Dash in mobile networks. The results show a higher efficiency for RTSP, while RTMP presented higher Quality-of-Experience (QoE). Other studies evaluated different adaptions for improved video transmission. Vergados et al. [14] evaluated the performance of several DASH adaptations over LTE, which did not considerably increase the performance. Mammeri et al. [15] proposed the Erasure Coding-Real Time Protocol to address the high packet loss problem with the drawback of adding a small amount of communication delay.

Our work implements a *lightweight* streaming protocol based in RTSP but adapted to the vehicular communications. We propose a streaming protocol that is controlled accordingly to the overtaking flow. In addition, the video streaming is only started after the detection of an imminent overtaking maneuver, e.g. through the actions of the active vehicle driver that are captured in the Controller Area Network (CAN)[1] bus.

[1]The CAN bus interconnects - through a shared bus line - the Electronic Control Units (ECUs) and other devices that control subsystems in vehicles.

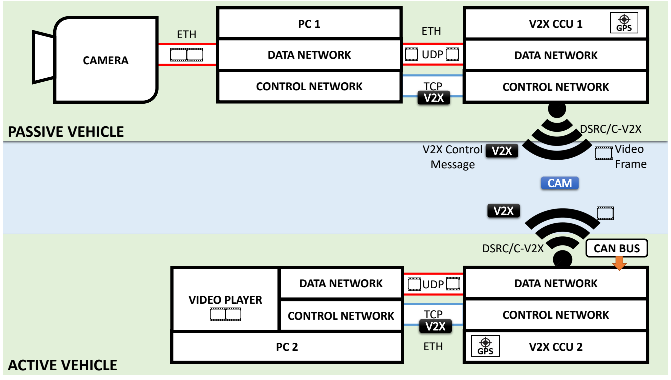


Fig. 1: Overall System Architecture

III. Video-assisted Overtaking System

We propose a system for safe overtaking maneuvers in challenging scenarios (e.g. rural roads) enabled by video streaming over C-V2X networks. The basic system functioning is based on the transmission of a compressed video feed acquired by one or multiple front-facing cameras installed in the vehicle being overtaken (*passive vehicle*) to the vehicle performing the maneuver (*active vehicle*), which is presented to the driver through a display or projected into the windshield after decompression. To enhance the system safety, the passive vehicle resorts to object detection techniques - combined with depth estimation and object tracking - to better understand the current traffic scene ahead, i.e to identify potentially dangerous situations arising from the lane change (e.g. head-on collision); the video streaming is only initiated if the passive vehicle considers that the overtaking maneuver can be performed safely.

A. Architecture

Fig. 1 depicts the modular system architecture. It consists in a distributed system, with stations installed in both vehicles, interacting through the exchange of standard and custom messages. Each vehicle is equipped with an Application Unit (AU) and a Communication Control Unit (CCU) that are connected over Ethernet using different transport protocols depending on the information type. Two message flows are considered for the transmission of 1) *video stream* using the UDP for reduced latency and 2) *application control information* (e.g. streaming startup) using the TCP for improved reliability.

Each unit runs dedicated software for enabling the system functioning. The main system components are herein described indicating in which vehicle these components are available:

- **Identification of safe overtaking conditions** (*passive*): This AU module resorts to the camera(s) of the passive vehicle to analyzes the traffic scene ahead using object detection techniques combined with object tracking [7] [16] and depth estimation to determine whether the overtaking maneuver is considered safe. The module outputs 3D object representations (reference object: passive vehicle) with the corresponding object identifiers which are constantly analyzed by the system to determine the safety level. Combining this information with the accurate position of the passive vehicle provided by the GPS sensor, and positioning information of other vehicles (e.g. active vehicle) contained in the received CAMs allows creating an *approximate* representation of the 3D world. Note that the CAM also contains the static vehicle's information (e.g. dimensions, type). This representation allows applying the time-to-collision [17] concept to identify potentially dangerous situations. The method projects to the future the trajectories of the two involved vehicles (overtaking vehicle and vehicle traveling in the opposite direction) and determines if at any given time interval the two dynamic objects are closer than a given safety distance. If the maneuver is deemed unsafe the system will prevent the transmission of the video stream. Note that vehicles traveling in the opposite direction might not be equipped with V2X communication equipment during initial deployment phases, which prevents the sole use of CAMs to create the world representation.

- **Video Streaming Algorithm** (*active & passive*): Prior to video streaming, a V2X Application (available in the CCU) predicts the overtaking maneuver by sensing the vehicle surroundings using local information (e.g. GPS) and information received from other vehicles via standard-compliant CAMs [18]. CAMs contain static and dynamic information of the ego vehicle (e.g heading, position, speed) that are used to detect overtaking maneuvers that meet certain conditions (e.g. vehicles with the same heading and relative distance smaller than a given value). When these conditions are met and the driver activates the left turn signal (recognized by reading information from the CAN bus), another stage of the overtaking system is triggered and the video streaming between vehicles starts. A detailed description of the VSA functioning is provided in Section III-C.
- **Frame Processing** (*passive*): After the video streaming has been triggered, this AU software module is responsible for the following tasks: i) *frame acquisition* from the selected camera at a given rate (e.g. 10 Hz), ii) *downscaling* to reduce the frame size, iii) (software-based)[2] *image compression* using a state-of-the-art algorithm (e.g. H.264/H.265), iv) *frame fragmentation* into smaller packets [termed Video Packet Messages (VPMs)] and v) *transmission* to the respective CCU using UDP for subsequent broadcast via C-V2X networks.
- **Frame Processing** (*active*): This AU software module performs the reserve operations of the passive vehicle, namely datagram reassembly, decompression, image upscaling to original resolution and display of video frame.

B. Information exchange between vehicles

C-V2X/LTE-V2X. 802.11p and C-V2X share the same frequency band. C-V2X networks operate in the 5.9 GHz band with 10 or 20 MHz channels. This technology employs a tailored SC-FDMA scheme for resource allocation. In the time domain, the frame is divided into ten 1 ms-long subframes; each subframe is composed of two time-slots each containing 7 OFDM symbols. In the frequency domain, a Resource Block (RB) of 180 kHz is composed by 12 subcarriers spaced by 15 Hz, which can be grouped into sub-channels. Resource allocation and interference management can be network-assisted in Mode 3 or done completely distributed in Mode 4.

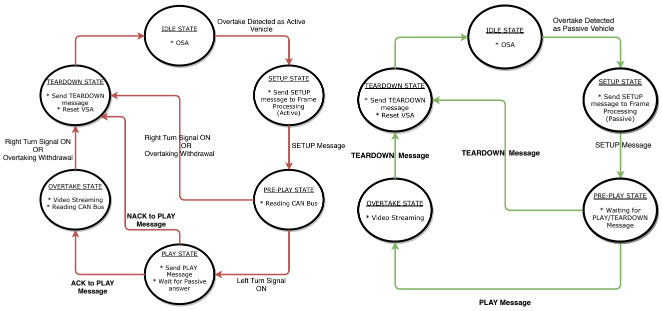
The communication between vehicles is done using C-V2X networks (release 14) in Mode 4, which does not require a base station for the resource channel allocation. In this mode, C-V2X uses the sensing-based Semi-Persistent Scheduling (SPS) scheme. SPS is a technique for resource channel allocation to support periodic packet transmissions with efficiency and to reduce packet collisions. Contrary to 802.11p which is based on CSMA-CA, C-V2X networks rely on the SPS contention-free scheme for optimized resource allocation. The selection of sub-channels is done in a distributed fashion by each network node considering a common set of parameters and a sensing mechanism. In the SPS scheme, nodes sense the different sub-channels and reserve the required resources that are not currently being used by other nodes. These reservations are communicated to other nodes through the Sidelink Control Information (SCI). Resources are kept during a random time interval (*Re-selection Counter*) within given bounds [5 to 15 Resource Reservation Intervals (RRI)]. When the re-selection counter expires, the node keeps the same configuration with probability p or selects new sub-channels with probability $1-p$. Additional sub-channels are also reserved if the message size increases or the latency requirement cannot be met.

At the MAC layer, C-V2X implements the HARQ technique for improved reliability and higher efficiency, wherein a given packet is sent twice within the packet delay budget with different coding schemes. In addition, channel coding in C-V2X is based on turbo codes (instead of convolutional codes for 802.11p), which are more robust in varying radio environments. More detailed information on C-V2X (Mode 4) and the impact of message variability on the C-V2X operation can be found in [19].

Protocol Messages. As detailed previously, the system resorts to standard-compliant CAMs for the VSA functioning. To reduce the system startup latency, we designed a custom streaming protocol based in RTSP but adapted to vehicular communications. To achieve this goal, two types of custom V2X Messages have been designed:

- **V2X Control Messages (CM)**: these messages are sent by the active vehicle to prepare image acquisition (*Setup* Message), start the video transmission (*Play* Message), and stop it (*Teardown) Message*. The generation of these messages depends on the prediction and behavior of vehicles on the road and the actions of the active vehicle driver that are read from the vehicle's CAN bus. The active vehicle driver must activate the left turn signal to start the video-streaming triggering the transmission of the *Play* Message, while the right turn signal stops the video-

[2]Hardware-based image compression considerably reduces the processing time but requires dedicated hardware.



(a) *Active Vehicle* (b) *Passive Vehicle*

Fig. 2: Video Streaming Algorithm (VSA).

streaming through the transmission of the *Teardown* message. Due to the unreliability of the wireless channel, we implemented a simple confirmation mechanism based on acknowledgments (ACK) for the CM messages. The *Play* and *Teardown* messages are exchanged between vehicles being forwarded to the AU at the receiving node, while the *Setup* message is only exchanged locally from the CCU to the AU.

- **Video Packet Messages (VPM)**: these video content messages are sent by the passive to the active vehicle.

C. Video Streaming Algorithm (VSA)

The system is controlled by the V2X application present in the CCUs. This software module lies in the Application Layer of the ITS Stack. The main functionalities of the V2X application are: 1) automatically identifying overtaking maneuvers and determining the vehicle role (active or passive); 2) start and stop the video-streaming between vehicles; 3) read sensor information from the CAN Bus and GPS sensor, and provide it to the V2X Stack; and 4) CMs generation and reception handling of the CMs, VPMs, and CAMs.

This software module implements the State Machine depicted in Fig. 2. The VSA execution differs depending on the vehicles' role, i.e. *active vehicle* or *passive vehicle*. The transition between states is triggered by two main events: i) reception of V2X control messages from the other vehicle (e.g. the reception of the *Play* message triggers the video streaming between vehicles) and ii) updated sensor information (e.g. the left turn signal makes the state change from the *PRE-PLAY* to the *PLAY* state at the active vehicle). The VSA is decomposed into two sub-modules: i) Overtaking Sensing Algorithm (OSA) to determine if an overtake maneuver is eminent and to determine the role of each vehicle and ii) Trigger Detection Algorithm (TDA) to control the video-streaming between vehicles. The OSA is incorporated into the *IDLE* state of the state machine, while the remaining states are devoted to the TDA functioning.

Overtaking Sensing Algorithm (OSA): The OSA operation comprises two phases. In the first phase, each vehicle identifies possible overtaking maneuvers by assessing whether the other vehicle travels in the same direction (i.e. similar headings) and is within a given relevance zone using its own location information provided by GPS and the received CAMs. In the second phase, the module assesses whether an approach is being made (i.e. the distance between the two vehicles decreases) using the positioning information of the two vehicles. The algorithm buffers five CAMs for improved reliability in the distance calculation. If the approach is materialized, the Trigger Detection Algorithm (TDA) is then executed. The algorithm also detects the vehicle's role in the overtaking (active or passive) by considering the relative positions of the vehicles in the road.

Trigger Detection Algorithm (TDA): After detecting an imminent overtake maneuver, the active vehicle sends a *Setup* message to its frame processing module. Similarly, the passive vehicle sends a *Setup* message to the frame processing module to setup image acquisition. In both cases, the system transitions to the *PRE-PLAY* state, in which it will wait for user actions (e.g. activation of turn left signal) reported by the CAN bus[3]. The TDA detects the activation of the turn light signals by reading information broadcasted on the CAN bus. After the activation of the left turn signal, the active vehicle transmits the *Play* Message to the passive vehicle (*PLAY* state). This vehicle verifies if the overtaking is feasible and replies providing a positive (ACK) [*OVERTAKE* state] or negative acknowledgment (NACK) [*TEARDOWN* state] in the following circumstances: i) another overtake maneuver is already being served or ii) the maneuver is not considered safe. The *Play* message is also used to synchronize the vehicles' state machines to the *OVERTAKE* state and trigger the video streaming between vehicles. The TDA also assesses a possible overtaking maneuver termination or withdrawal by 1) detecting a significant increase in distance between the two vehicles or 2) activation of the right turn signal [*TEARDOWN* state]. Upon this detection, the active vehicle sends the *Teardown* message to stop video transmission and reset the VSA to the *IDLE* state.

Parameter	Configuration
Channel [Bandwidth]	180 (5.9 GHz) [10 MHz]
CCU Tx Power	20 dBm
Tx & Rx Antenna Gain	5 dBi (Omnidirectional)
Antenna Placement	Vehicles's Roof
CAM Frequency	10 Hz
HARQ re-transmissions	enabled
Resource Selection Window (Ts)	100 ms

TABLE I: C-V2X configurations

IV. Results

A. Experimental platform & test site

We conducted several experiments to assess the performance of the proposed overtaking system. The measurement system was installed in two conventional vehicles (BMW i3 and 5 series) [see Fig.3]. The measurement system was composed

[3]The CAN bus allows acquiring real-time information on the status of different vehicle systems (e.g. the status of the engine module are reported by the engine control unit).

(a) Passive Vehicle (b) Active Vehicle

Fig. 3: Experimental platform.

of a laptop (AU), a GNSS receiver, and a communication system. A video camera was also installed in the passive vehicle. The communication system uses standard-compliant C-V2X radios operating in the 5.9 GHz frequency band. Omnidirectional antennas (MobileMark SMWG-313) with 5 dBi gain were installed in the vehicle's roof. The configuration of the communication system is given in Table I. To assess the system performance, each node captured the transmitted and received packets using the *tcpdump* tool, and simultaneously collected in statistics, position, and timing information.

The experiments were performed during May 2020 in a surface parking lot in the city of Braga, Portugal. Transmitter and receiver where predominantly in Line of Sight (LoS) conditions. The parking lot maximum length was approximately 300 m, which allowed to safely perform the overtaking maneuver. The initial distance between the vehicles was 95 m. During the measurements, both vehicles moved at a constant speed (active vehicle: 10 m/s and passive vehicle: 5 m/s), and when in close proximity the active vehicle overtook the passive vehicle. The experiments ended shortly after the overtaking maneuver was successfully executed. Four similar test runs were executed.

B. Evaluation Metrics

The following time-related metrics are considered:

- **E2E Latency**: time interval between the capture of a video frame at the transmitting station and the display of the same frame at the receiving station;
- **Trigger Latency**: time interval between the activation of the left turn signal and the initial video display;
- **Packet Inter-Reception Ratio (PIR)**: the time difference between the reception of two consecutive packets;

The following conventional video metrics are considered:

- **Mean Opinion Score (MOS)**: the average perceived quality evaluation with an audience of four people;
- **Peak Signal-to-Noise Ratio (PSNR)**: measures errors between original and reconstructed image. This metric typically ranges from 20 to 40 dB [20].
- **Structural Similarity (SSIM)**: measures images similarity (range: [0,1]) and is correlated to human quality. Values closer to 1 representing a higher resemblance between original and compressed images.

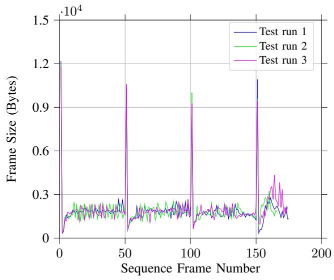


Fig. 4: Frame Size as a function of frame number.

TABLE II: Evaluation of time- and video-related metrics.

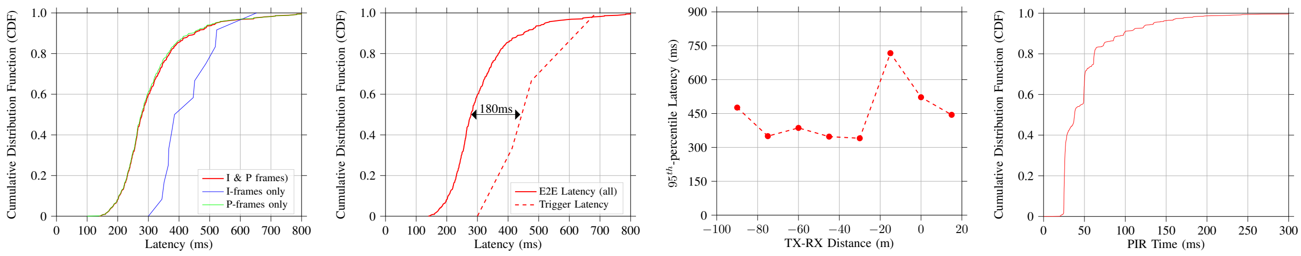
Type	Metric	Frame	Average	Std. Dev.	Unit
Time	E2E latency	all	362.528	113.697	ms
		I only	438.954	198.036	ms
	Trigger Latency	all	526.062	117.323	ms
	PIR latency	all	52.249	45.072	ms
Video	MOS	-	4.167	0.144	N/A
	SSIM	-	0.791	0.007	N/A
	PSNR	-	25.998	0.297	dB

C. Results

First, we present a brief overview of the generated traffic. Fig. 4 depicts the frame size as a function of time. We observe that the frame size increases considerably every 50 frames due to the transmission of *I-frames*, being the traffic closer to the target bit-rate for the remaining frames *P-frames*. Note that I-Frames represent a complete image (independent from other frames), while P-Frames exploit the differences from previous frames [21]. The size of P-frames is dependent on the dynamicity of the scene as can be seen for frame numbers larger than 150. Higher frame sizes require more fragmentation and consequently the transmission of a larger number of packets.

Time-related metrics. Table II presents a summary of the time-related and video-related results. Fig. 5 depicts the time-related metrics. The average E2E latency is acceptable (362 ms) and exhibits a reasonable variability (114 ms). We attribute the larger values of E2E latency to the transfer of the *I-frames* (see Fig. 5a) that lead to the transmission of a larger number of packets in a short time period. Although less frequent, the transmission of *P-frames* might also lead to high values of E2E latency; typically these high E2E latency values occur just the transmission of *I-frames*. Thus, we argue that due to its design the C-V2X scheduling algorithm introduces additional delay when handling (highly) variable traffic.

Analyzing now the results from an application point of view, we show in Fig. 5a that 60% and 94% of the video frames were captured and displayed in less than a requirement of 300 ms [22] or 500 ms [11], respectively. Fig. 5b compares the E2E and trigger latency. These results demonstrate that the average trigger latency is low (526 ms) and that the setup of the system (excluding E2E latency) is solely around 180 ms. The extreme trigger latency value is 687 ms, which corresponds to solely 17 m for a vehicle moving at 25 m/s. The overall system latency could be even further reduced with improved



(a) E2E Latency (b) E2E and Trigger Latency (c) E2E Latency (95^{th}-%ile) (d) PIR

Fig. 5: Time-related performance metrics

implementation. Fig. 5c depicts the 95^{th}-percentile of the E2E Latency as a function of the TX-RX distance. The results show that extreme latency values remain fairly constant as a function of distance; the differences between distance bins are explained by the transmission of *I-frames* that increase the extreme values of latency. Regarding the performance of the VSA, the algorithm was able to correctly detect all triggers in the test, specifically detecting the overtake maneuver, identifying the vehicle role and starting the video streaming. Fig. 5d presents the PIR CDF that helps to understand the occurrence of network problems in the reception of packets and perceiving the latency in capturing video frames. About 85% of the consecutive packets arrived with 100 ms of each other, meaning that the issues previously related did not impact the system most of the time, despite some higher values.

Video metrics. Regarding the image quality, the PSNR and SSIM metrics are within an acceptable range with low variability as shown in Table II, namely the SSIM is in the upper range of the [0, 1] interval and an average PSNR of 26 dB shows low errors in image reconstruction. For the perceived video quality, the audience rated the video with good quality (MOS = 4.167).

V. Conclusions

We proposed a safe and automatic overtaking system enabled by image processing techniques and video streaming over C-V2X networks. The evaluation demonstrated the system feasibility and its adequate performance. The system produced good video quality for the viewer but higher E2E latency values were introduced by C-V2X networks when handling variable traffic. As future work, we intend to understand in more detail the functioning of C-V2X networks under variable traffic. We also intend to make the system function more robust and adopt a probabilistic approach for processing sensor data.

References

[1] World Health Organization, *Global Status Report on Road Safety*. World Health Organization, 2018.

[2] European Road Safety Observatory, “Annual Accident Report 2018,” European Commission, Tech. Rep., jun 2018.

[3] C. Olaverri-Monreal, P. Gomes, R. Fernandes, F. Vieira, and M. Ferreira, “The See-Through System: A VANET-enabled Assistant for Overtaking Maneuvers,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2010.

[4] A. Bohm, M. Jonsson, and E. Uhlemann, “Adaptive cooperative awareness messaging for enhanced overtaking assistance on rural roads,” in *IEEE Vehicular Technology Conference*, 2011, pp. 1–5.

[5] P. Gomes, C. Olaverri-Monreal, and M. Ferreira, “Making Vehicles Transparent Through V2V Video Streaming,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 930–938, 2012.

[6] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, “3D Traffic Scene Understanding From Movable Platforms,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, 2014.

[7] P. Li, X. Chen, and S. Shen, “Stereo r-cnn based 3d object detection for autonomous driving,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7636–7644.

[8] Y. Zhou and O. Tuzel, “VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4490–4499.

[9] M. Boban and P. M. d’Orey, “Exploring the Practical Limits of Cooperative Awareness in Vehicular Communications,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3904–3916, 2016.

[10] R. Molina-Masegosa and J. Gozalvez, “Lte-v for sidelink 5g v2x vehicular communications,” *IEEE Vehicular Technology Magazine*, vol. 12, no. 4, pp. 30–39, 2017.

[11] P. M. d’Orey, A. Hosseini, J. Azevedo, F. Diermeyer, M. Ferreira, and M. Lienkamp, “Hail-a-Drone: Enabling Teleoperated Taxi Fleets,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2016, pp. 774–781.

[12] J. Pereira, M. Diaz-Cacho, S. Sargento, A. Zuquete, L. Guardalben, and M. Luis, “Vehicle-to-vehicle real-time video transmission through ieee 802.11p for assisted-driving,” in *IEEE Vehicular Technology Conference (VTC Spring)*, 2018, pp. 1–6.

[13] A. Aloman, A. I. Ispas, P. Ciotirnae, R. Sanchez-Iborra, and M. D. Cano, “Performance evaluation of video streaming using mpeg dash, rtsp, and rtmp in mobile networks,” in *IFIP Wireless and Mobile Networking Conference*, 2015, pp. 144–151.

[14] D. J. Vergados, K. Kralevska, A. Michalas, and D. D. Vergados, “Evaluation of http/dash adaptation algorithms on vehicular networks,” in *Global Information Infrastructure and Networking Symposium*, 2018.

[15] A. Mammeri, A. Boukerche, and Z. Fang, “Video streaming over vehicular ad hoc networks using erasure coding,” *IEEE Systems Journal*, vol. 10, no. 2, pp. 785–796, 2016.

[16] W. Luo, B. Yang, and R. Urtasun, “Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3569–3577.

[17] W. Wachenfeld, P. Junietz, R. Wenzel, and H. Winner, “The Worst-time-to-collision Metric for Situation Identification,” in *2016 IEEE Intelligent Vehicles Symposium (IV)*, 2016, pp. 729–734.

[18] ETSI, *ETSI EN 302 637-2: Specification of Cooperative Awareness Basic Service (V1.4.1)*.

[19] R. Molina-Masegosa, J. Gozalvez, and M. Sepulcre, “Comparison of IEEE 802.11p and LTE-V2X: An Evaluation With Periodic and Aperiodic Messages of Constant and Variable Size,” *IEEE Access*, vol. 8, pp. 121 526–121 548, 2020.

[20] H. Hosseini, B. Xiao, and R. Poovendran, “Google’s Cloud Vision API is Not Robust to Noise,” in *IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017, pp. 101–105.

[21] X.-W. Yao, W.-L. Wang, S.-H. Yang, Y.-F. Cen, X.-M. Yao, and T.-Q. Pan, “Ipb-frame adaptive mapping mechanism for video transmission over ieee 802.11e wlans,” *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 2, p. 5–12, Apr. 2014.

[22] ETSI, *ETSI TS 101 539-2: Intersection Collision Risk Warning (ICRW) application requirements specification (v1.1.1)*, June 2018.