

AR-based Remote Command and Control Service: Self-driving Vehicles Use Case

Oussama El Marai[1], Tarik Taleb[2,3], and JaeSeung Song[3]
[1]Department of Communications and Networking, School of Electrical Engineering, Aalto University, Finland
[2]Faculty of Information Technology and Electrical Engineering, Oulu University, Finland
[3]Department of Computer and Information Security, Sejong University, Seoul 05006, South Korea
Emails: oussama.elmarai@aalto.fi, tarik.taleb@oulu.fi, jssong@sejong.ac.kr

***Abstract*—The recent technological advances in many fields have significantly contributed to the development of the Advanced Driver Assistance System (ADAS), which in turn will greatly contribute to the flourishing of self-driving vehicles that can operate autonomously in all road scenarios. Until then, keeping the human input in the loop remains vital to either make decisions in unseen situations or approve vehicles' proposed decisions. In this paper, we leverage VR technology to provide remote assistance for self-driving in critical situations. Specifically, we study the delivery of a 360° live stream at high resolution (4K) to a remote operation center for supporting self-driving vehicles' decisions when, for example, merging onto the highway. The 360° video stream will be consumed by a human operator wearing a head-mounted display for increased flexibility, faster control, and an immersive experience. In addition, the 360° stream is augmented with relevant context data, such as the vehicle's speed and distance to other road objects, in order to increase the human operator's awareness of the vehicle and its surroundings. Depending on the human operator's proximity to the source, the video stream can either be viewed through the cloud or the edge, which further reduces the glass-to-glass latency. Experimental results demonstrate the effectiveness of employing VR technology to remotely command and control self-driving vehicles in critical situations. The results show that a 360° stream at 4K resolution can be delivered in sub-second glass-to-glass latency, which allows the operator to make timely decisions.**

***Index Terms*—360° Stream, Live Streaming, Virtual/Augmented Reality, Remote Assistance, and Self-driving Vehicles.**

I. Introduction

Nowadays, we are in the early stages of the launch of the Fifth Generation (5G) mobile network, and some mobile operators have already started its commercialization. In terms of mobile broadband (up to 10 Gbps), ultra-reliable and low-latency (nearly 1ms), and massive machine-type communications (1 million devices/km^2), this emerging technology promises to cater to diversified services with varying and stringent requirements. While this would certainly increase the revenues of mobile operators, it also places colossal strain on the underlying physical infrastructure, and introduces new technical challenges particularly in respect of Service Level Agreements (SLAs).

In most facets of our daily lives, including education, entertainment, and surveillance, video-on-demand (VoD) and live streaming have become an essential service of many developed systems. Video streaming applications rank as the most bandwidth-intensive services, especially when viewed at higher resolutions, such as FHD and 4K. As a matter of fact, people today are watching FHD videos as the standard, and due to the larger bandwidth promised by the 5G mobile networks, 4K produced videos are rapidly increasing. Consequently, the underlying infrastructure would be further stressed as well as new delivery challenges would be introduced, especially for real-time interactive services where user reactions are needed. The problem becomes worse when 360° streams are delivered in Augmented Reality (AR) and Mixed Reality (MR) applications. According to Cisco forecast[1], it is expected that HD VR and UHD VR will dominate the bandwidth of future homes by 2023.

Virtual Reality applications have a great deal of potential in various domains because of the immersive and engaging experience they offer to end-users. Microsoft reported that the use of Mixed Reality, using the Microsoft HoloLens Head-Mounted Display (HMD), has multifold benefits in the education field, including boosting student engagement by 35% and improving test scores by 22%[2]. The retail and shopping industries are also potential fields that would certainly benefit from this emerging technology. Many companies, such as IKEA, have already explored the potential of AR by developing their AR-based apps to allow customers to explore their products in various ways. In turn, the AR technology has been also used by the healthcare sector to perform surgeries [1] and explain complex concepts of human body at high efficiency. VR-based applications can also play a crucial role in protecting the environment. In fact, the immersive and engaging experience offered by VR applications, such as remote collaboration, remote assistance, and remote tourism, constitute an ideal alternative to the physical experience. Consequently, unnecessary trips and many expenses associated with transportation and office leases could be reduced to a greater degree, thereby enabling better mitigation of climate change from a greenhouse gas perspective.

The purpose of this paper is to leverage the AR technology for remote driving assistance of autonomous vehicles. We specifically investigate the efficiency of using a 360° stream, augmented with vehicle-related information, such as speed,

[1]https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html
[2]https://www.microsoft.com/en-us/education/mixed-reality

by a remote human operator (RHO) using an HMD wearable device to assist self-driving cars at some critical circumstances such as overtaking and merging into the highway. The use of the HMD device is motivated by the three degrees of freedom (3DOF) offered to the user movements, which boosts the visibility of roads and enables faster reaction times. Towards this end, we conducted several experiments comparing different streaming protocols and modes (push vs. pull) and evaluated the delivery of 4K 360° live streams in terms of glass-to-glass (G2G) latency. In contrast, when the video stream is automatically processed and decisions are made instantly by computers, end-to-end (E2E) latency might be a better measure of relevance. A previous study [2] reported that the average driver reaction brake time to be 2.3 seconds. Our numerical results show that 360° 4K streams can be delivered with sub-second G2G latency, giving the RHO enough time to react and send its decision to the autonomous vehicle via commands.

The rest of this paper is structured as follows. In Section II, we describe some selected work employing AR technology in real-world scenarios. Thereafter, we present relevant use cases where AR-based remote control would be of great importance. The proposed system architecture is described in Section IV. Next, we provide experimental results in Section V. Lastly, Section VI draws conclusions and outlines some future directions.

II. Related Work

Augmented Reality systems are increasingly getting momentum in recent years, especially with the noticeable advances in their enabling technologies such as the processing capabilities, the improved bandwidth to carry a large amount of generated data, and the advanced displaying devices (e.g., HMD). In Augmented Reality technology, real-world environments are extended or enhanced with virtual objects, such as colors, images, and texts, to provide viewers with an immersive experience. One of the main driving points behind the success of AR-based applications is the surging popularity of 360° video streaming in recent years. This is essentially owed to the richer and captivating experience offered compared to traditional streaming. In the first part of this section, we look at some previous work on delivering 360° streaming. Then, we review some recently conducted work exploring the potential of AR technology for various purposes and domains. For a comprehensive survey on both 360° and AR-based applications, we point the reader to [3] and [4], respectively.

A. 360° Streaming

Previous research on 360° video delivery has focused on different aspects at different stages, namely content creation and preparation (e.g., encoding at single or multi-bitrate for adaptive video streaming, stitching various viewports coming from an assortment of cameras), transmission over the network using different protocols, and eventually the display (e.g., active viewport and bitrate dynamic adaptation) on the end-user's devices (e.g., HMD and web-based players). In this subsection, we particularly introduce some previous work leveraging 360° videos in various applications.

Noronha et al. [5] propose a web-based application for Sight Surfing. The proposed system allows users to visualize and navigate through 360° hypervideos (i.e. 360° videos that hyperlink to other 360° videos). To this end, 360° videos should be associated with the trajectory geolocation where they have been captured.

In [6], the authors propose an online shopping system based on 360° recorded videos of different products to support "disadvantaged shoppers" in Japan. The idea is to avoid going through the classic search methods that consist of typing keywords of the desired products, which might not be appropriate especially when the name of the products are unknown. Accordingly, the authors propose a shopping system where users can navigate around the shop using 360° video and select the products they are interested in. The users can also display the selected products in 3D as if it is in their hands.

In [7], Ozcinar et al. proposed an adaptive 360° viewport-aware bitrate level selection approach based on MPEG-DASH paradigm. The main objective is to increase the streaming performance by enhancing the displayed viewport video quality while taking into account the bandwidth limitation. To do so, the authors propose to divide the projected 360° Field Of View (FOV) to 2D plane into tiles, where each tile is encoded at different bitrates. This would result in further splitting the viewport into separate and self-decodable tiles, which introduce the inside- and outside-viewport tiles. During the streaming session, the proposed DASH VR player requests all the tiles of a specific viewport from a video representation. However, only inside-viewport is requested from relatively high quality, and the bitrate of the outside-viewport tiles is gradually reduced to save bandwidth. Consequently, the video quality of the active viewport (i.e., the user is currently viewing) will improve. To support these changes, the authors also proposed an extended version of the standard MPD file proposed originally by DASH. Using 8K video quality, the authors corroborated the quality performance of their approach in a simulated environment and compared their promising results to commercial VR solutions.

B. Augmented Reality

In [8], the authors propose an end-to-end AR solution in the agriculture sector to support aquaculture farmers. The proposed system, named "AR + Cloud", aims to collect data about the aquaculture pond to live monitor and analyze in-situ water quality. To this end, they propose to reduce the big delay (up to 6 hours) between data collection and analysis, which is mainly caused by the manual process. As a remediation, Optical Character Recognition (OCR) and Wearable Data Collection Suit (WDCS) can be used for faster data acquisition and on-site decision-making, rather than transmitting the data to be processed elsewhere.

In [9], the authors propose the use of augmented reality and ontologies to help caregivers design and install smart homes for assisting frail people at home. To this end, the designer of

the smart home must provide the activities, as a set of tasks and actions. These activities consist of the personalized scenarios of the senior person. The defined scenarios are stored in an ontology that holds the links between objects, actions, and sensors. At the design phase, the caregiver uses a Virtual Advisor (VA) powered with an object recognition module that associates actions to physical elements. Using REST APIs, the VA extracts from the ontology domain-specific knowledge, maps the physical elements inside the ontology, and eventually returns directions for assisting seniors.

AR applications are also used in the tourism sector to provide immersive discovering experiences to worldwide tourists. It also serves as an exciting means for disseminating and advertising the heritage of a region, which motivates travelers to visit specific places. For instance, Museums' assets can be offered to tourists remotely through AR applications by visualizing heritage assets and sites while adding relevant historical information as annotations. In [10], the authors proposed *JejuView*, an AR/VR web application that provides consumers with an immersive experience to discover and explore the cultural heritage of the famous Jeju Island in South Korea using smartphone and HMD devices. Google also has its own AR Arts & Culture project where users can virtually visit over 2000 museums around the world and 10000 places as well.

III. VR-based Remote Control Use Cases

As introduced previously, AR/VR applications are increasingly embraced and integrated into many sectors due to their proven high capability in increasing end-user engagement. Most of the presented work in the previous section mainly focuses on leveraging AR/VR for non-real-time services. In this section, we showcase some use cases where real-time communication is required while exploiting AR/VR capabilities.

A. Remote Assistance for Self-driving Vehicles

Self-driving vehicles are one of the essential technologies envisaged for future smart cities. This technology is promising a variety of benefits among which improving human lifestyle and fostering greener environments by considerably reducing carbon emissions. These lofty goals, however, need extensive and heavy communication between all existing entities to increase situational awareness and require instant and automated, ideally local, decision making. Due to many challenges, these requirements might not be possible in the near future, therefore gradual transition, by keeping the human's decision in the loop for both monitoring and decision-making purposes, might be safer when dealing with unseen situations or lacking sufficient data by the self-driving vehicles. In this vein, the leverage of newly emerged technologies such as 360° video streams and VR applications would increase situational awareness, while introducing immersion, and conveniently enable faster decision making.

B. Drone Control

Drones, technically known as Unmanned Aerial Vehicles (UAVs), is a great and powerful technology that is penetrating many sectors nowadays such as agriculture, construction and mining, parcel delivery, among others. The continuous development and success of drones allowed to build many services and applications around this phenomenal technology. For instance, most drones are equipped with cameras to perform aerial photography or streaming in different domains such as tourism, cinematography, and live sports events, to deliver bird's eye view. If the onboard camera is 360°, this would enrich and provide a captivating user experience. They are also used for critical missions, such as inaccessible or dangerous locations (e.g. Firefighting), where sending a human being might be of high risk. In most of the aforementioned scenarios, and also using the on-board camera feed for controlling the drone itself notably behind the visual line-of-sight, sub-second G2G latency is a requirement for the operating human being to give him enough time for reaction.

C. Digital Twin-based Remote Surveillance

Digital twins are the clones of real-world assets in cyberspace. This technology consists of continuously gathering comprehensive real-time data from the physical entity and conveniently displaying it in a digital environment.The ultimate goal is to enable real-time monitoring and optimizing the functionalities of the physical asset. Leveraging VR technology in this field would greatly simplify the monitoring task while increasing the users' engagement. If we consider the remote surveillance based on digital twinning of the roads, for instance, where multiple 360° cameras are optimally deployed at strategic locations in the city, the remote live monitoring by security agents using HMD devices would be more efficient to identify crimes or any other misbehavior on the roads.

D. Telesurgery

Telesurgery is one of the most cutting-edge applications that enables highly qualified and experienced surgeons to perform critical surgeries remotely. This would bring great economic and environmental benefits by reducing medical staff travels, and allow for efficient surgeons time management, notably with the scarce experienced surgeons, which would increase the number of performed surgeries, and potentially the number of saved lives. This technology also offers a great way to remotely train or assist less experienced medical staff who are performing the surgery on-site, giving them more confidence and offering better practice. The introduction of both 360° and AR/VR technologies, which allow for overlaying patient-specific live health-related measurements onto 360° live streams, plays an essential role in facilitating the surgeon's remote operations and actions. However, this kind of application requires high reliability, as any disruption might endanger lives. Furthermore, sub-second G2G latency is an essential requirement in such an application due to the expected fast reaction of surgeons, especially when a patient's health deteriorates during the surgery.

IV. System Design

In this section, we describe the proposed system architecture to enable remote human assistance for autonomous vehicles in critical situations such as vehicles merging into the highway or merging multiple lanes.

Figure 1 depicts an overview of the proposed architecture for the self-driving vehicles' remote assistance system. In this figure, we illustrate the scenario of vehicles' merging into the highway, where we have the blue self-driving car that is coming from the sharp side road on the right and going to merge into the highway. This scenario has some challenges such as the blind spot, which is even difficult for a human driver. If no vehicle has been detected on the main road, the blue car can quickly merge into the highway at position 1 in the figure. On the other hand, when a vehicle (i.e., the red car) is detected on the target road, taking the decision whether to join or not and at which position (e.g., position 1 or 2) might be complicated and depends on many parameters such as the actual position and speed of the blue and red cars, among others. Making such critical decisions by the self-driving vehicle itself based merely on the vehicle's sensors might be difficult and not really safe, and any mistake or error (e.g., a deficient sensor reporting wrong measurements) would result in catastrophic crashes. In this context, human support might be highly needed to either approve or change the initial self-driving vehicle decision. To this end, a notification might be sent to RHO, along with the 360° live streaming augmented with the necessary information (e.g., speed and estimated distance to the closest vehicle) about the vehicle in question as well as the detected vehicles, asking for human's intervention and decision. From Figure 1, we identify the following entities:

A. Self-driving Vehicle

The self-driving vehicle is equipped with different sensors, such as LiDAR and cameras, to increase its situational awareness. These sensors continuously collect data about road participants (e.g., vehicles, pedestrians, and animals). The collected data is then typically processed locally to make instant decisions, such as speeding up, slowing down, and breaking. However, some decisions are too critical, notably on the move, which requires human assistance in unseen and critical situations, especially when the self-driving technology is not mature enough at early deployment stages.

In this work, we study the use of a 360° camera to send a panoramic view of the road to the RHO. The use of 360° video stream in this use case is mainly motivated by the 3DOF of movement given to the RHO to see the vehicle's surrounding environment, notably when using HMD devices, by simply rotating the head. In addition to the high movement flexibility, viewing the 360° video stream from the HMD device would result in faster reaction time compared to the traditional 360 web player. When the vehicle approaches the crossroad, it automatically triggers a remote assistance request to the first available and closest RHO, which contributes to further lowering the G2G latency. Based on the selected RHO's geolocation and its proximity to the vehicle in terms of network hops, the 360° live stream might be sent to either the edge or cloud server. This information is retrieved from the cloud servers, where all RHOs are registered in a database. After selecting and establishing communication with the nearest server (i.e., edge or cloud) to the closest available RHO, the vehicle starts sending the 360° stream to the selected edge server. It is worth noting that two different types of streaming can be employed. In the first, the camera acts as a streaming server that waits for clients to connect, while in the second mode, the camera keeps streaming to a provided sink. Considering the fact that remote driving is only required if the self-driving vehicle runs into problematic situations, which tends to consume bandwidth unnecessarily, we take the first mode into consideration in this work. Unlike previous work [11], which examines traditional streams, our study focuses on the delivery of 4K 360° live streaming for a more comprehensive view into the scene and the display of the 360° streams on VR HMD devices for smoother and faster reaction times. Although previous works [7], [12]–[14]have recommended delivering 360° streams at 8K resolution or higher, such as 12K and 16K, 4K resolution is also acceptable and was used by both previous works as well as YouTube, as indicated in [13]. We choose to use 4K resolution mainly because of the sub-second G2G latency, instead of 8K at nearly three seconds, which is totally incongruous with the critical use case of remote driving. The 8K resolution may be appropriate in less latency-critical use cases, or for 360° recorded streams where the encoding is not performed on-the-fly.

B. Edge Server

In case the edge server is the closest to the RHO, the incoming stream received at the edge server will be sent to two different sinks via two separate pipelines. While the first pipeline re-transmits the received stream to the cloud server for recording purposes and later use (e.g., possible investigation in case of wrong decision), the second pipeline converts the input stream to WebRTC technology to enable the RHO at the operation center to consume the 360° live stream using either the 360 web player or HMD device for a more immersive experience and presence.

C. Cloud Server

When the edge server is selected as the primary streaming server, in this case, the cloud server is used as a backup server where the live stream is recorded for later potential use. On the other hand, if the shortest path between the RHO and the vehicle passes through the cloud server (e.g., when no close RHO is available), the cloud server is then elected as the streaming server, and accordingly, the RHO will be pointed out to retrieve the live stream from it. Similarly, the cloud server will record and convert the incoming stream to a format that is readable from a 360 web player and HMD device. It would be expedient to mention that other metrics, such as the end-to-end bandwidth, could be considered for selecting the primary streaming server. However, the selection of the optimal path is out of the scope of this work.

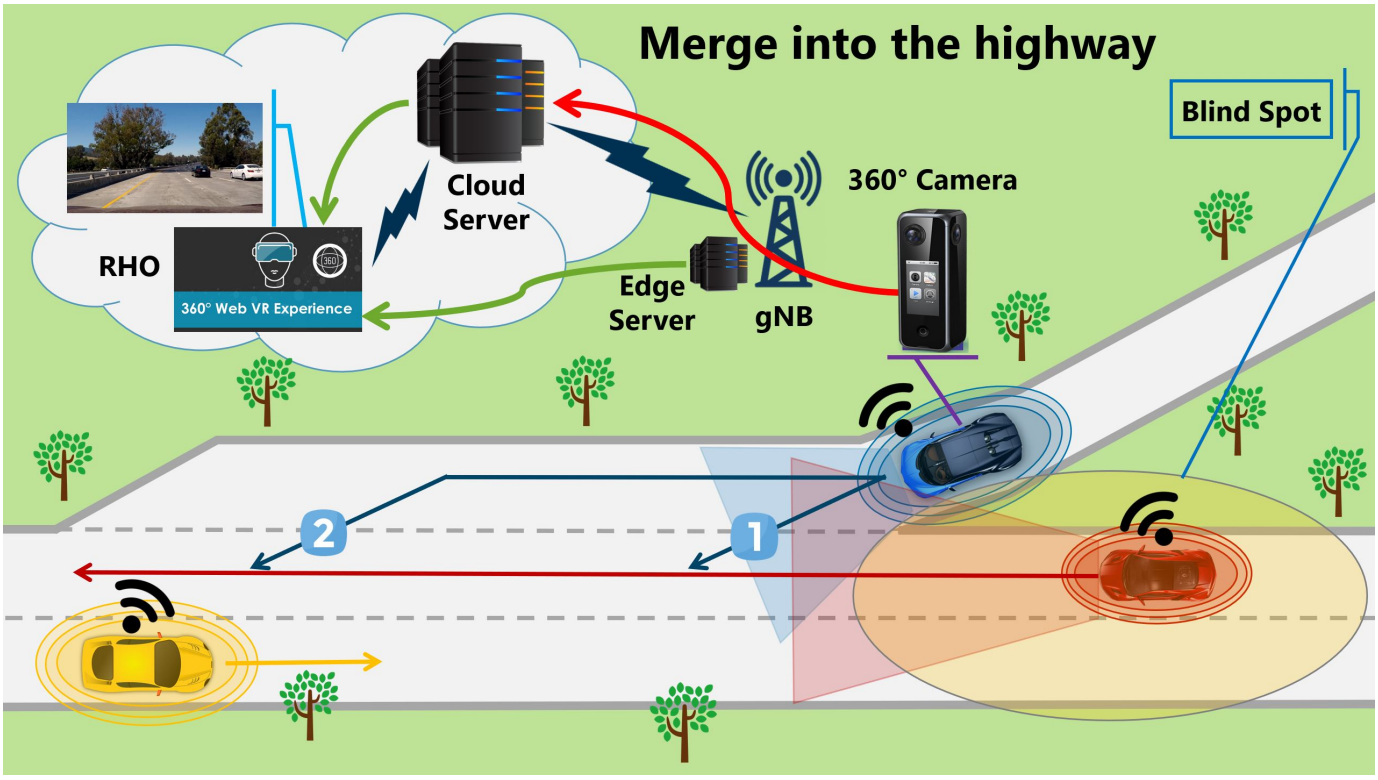


Fig. 1: AR-based self-driving vehicles' remote assistance system overview.

D. Remote Human Operator

Once the remote assistance request is triggered, the RHO promptly receives the 360° live stream in a separate window, where he can follow the vehicle's situation and change the viewport using the mouse or keyboard arrows. Instead, the end-user can watch the 360° live stream from the HMD device for a more flexible and captivating experience, which eventually results in a faster reaction time. Additionally, the 360° stream could be augmented with other relevant information such as the current speed of the blue vehicle as well as the estimated speed of the vehicle on the main road (the red car).

V. Performance Results

A. Experiment Settings

As illustrated in Figure 2, when remote driving assistance is triggered, the camera on the autonomous vehicle starts sending the stream to the closest available server to perform the conversion to a format playable on HTML5-based players. When the human operator receives a notification about the remote driving assistance, he can instantly render the received live stream in 360 panoramic view either from an HMD or desktop computer. The solid line in this figure represents the uplink, whereas the dotted ones show the downlink.

Regarding the hardware used for the evaluation, we used Labpano Pilot One 360° camera to capture, encode and send the 360° live stream to the server. The server's hardware configuration consists of an Intel® Xeon® Processor E3-1230 v5, 8M Cache, 3.40 GHz, 16GB RAM, nVidia GM107GL [Quadro K2200]. The receiver device is a Latitude 7490 - Core i5 8350U, 1.7 GHz, 16 GB DDR4, Intel UHD Graphics 620.

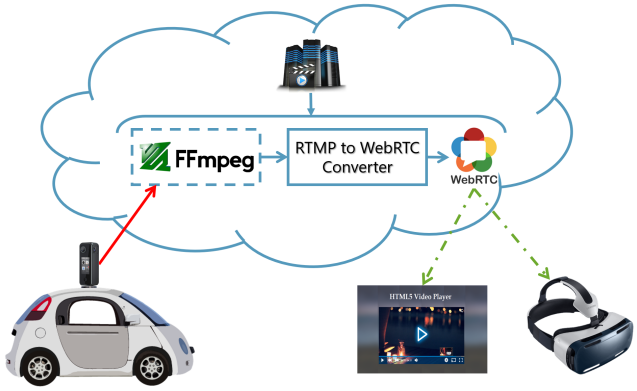


Fig. 2: Experimental setup.

In our experiments, we also evaluated the G2G latency under the three following streaming scenarios: a) *RTMP-PUSH:* where the camera pushes the stream towards a remote RTMP server, b) *RTMP-PULL:* where the remote server pulls the stream from the camera using RTMP protocol too, and c) *RTSP-PULL:* that is similar to the previous scenario but using RTSP protocol. These three streaming protocols were mainly selected due to their sub-second latency. To study the impact of the bitrate on the streaming performance, notably the latency, we evaluated each scenario at different bitrates, namely 6Mbps, 15Mbps, and 30Mbps. We also tried higher encoding rates (e.g., 40Mbps), but this resulted in poor user Quality of Experience (QoE), mainly due to frequent video

stalls. The video compression format used is H.264 at 4K (3840x2160) resolution.

On the server-side, a conversion from the received protocol (i.e., RTMP or RTSP) to an HTML5 friendly protocol (e.g., WebRTC) is necessary to allow the RHO to watch the 360° live stream on HMD or web-based applications. The conversion from the received RTMP and RTSP streams to WebRTC protocol is achieved through FFmpeg commands and OvenMediaEngine[3], which is an open-source streaming server. In this section, we evaluate the G2G (i.e., from the camera's lens to the RHO's display) latency because the consumer of the video stream is a human being, where some decisions, to be sent in the form of command messages, are expected from the RHO to either approve or alter the vehicle's decision. The G2G latency is measured at two different displays, namely the camera's display, and the end-user's display. First, we measure the G2G latency from the lens of the camera to the camera's display. This includes frame capturing, encoding, and rendering at the camera's display. Then, we measure the G2G latency between the camera's display and the end-user's display, which reflects the network and the rendering latency. Lastly, we measure the G2G latency from the source to the end user's display. It is worth noting that it may be more relevant to measure the end-to-end (E2E) latency [15], which is much lower than the G2G latency due to the display lag, when the video frames are automatically processed and analyzed (e.g., using ML algorithms) by a computer, and eventually, computer-based decisions are automatically made.

Our main focus in this paper is evaluating different video protocols and bitrates, as well as studying their impact on G2G latency for remote human-based vehicle control and decision aid. Therefore, we assume in our experiments that sufficient bandwidth is available for transmitting the video stream, regardless of the background traffic, notably during peak times. In this case, the use of the DASH technique might be a good alternative as it provides greater flexibility in adapting the bitrate of the stream, without compromising the exploitation of the live stream, on short time scales to changing network conditions, which ultimately enables as many vehicles as possible to stream. However, this technique exhibits a relatively high latency considering the remote driving use case. As a remediation, a stream selection approach, that consists of allowing/prohibiting vehicles to stream based on different parameters, such as the vehicle's location and streamed video quality, might be employed to avoid saturating the uplink, which negatively affects the overall system performance.

B. Experimental Results

This paper examines the 360° G2G latency under different scenarios and encoding rates (i.e., bitrates). In the first case, the experiments were done using RTMP push, where the camera is configured to push the 360° stream to a remote server. In the second scenario, the camera is configured as an RTMP server, then a client application connects to it to acquire the 360° stream and transmit them to the WebRTC converter module. For the third scenario, we evaluate 360° stream delivery over RTSP protocol, in the same manner as for the second scenario. The three scenarios are evaluated at bitrates of 6Mbps, 15Mbps, and 30Mbps, respectively. In general, the higher the bitrate, the larger the stream size, and the better the quality. Three evaluation stages take place: from *Source to Camera*, *Camera to RHO*, and *Source to RHO*. In the first stage, we measure G2G latency at the camera screen. This includes the time elapsed for stream acquisition, encoding, and display without any network transmission. The second stage measures the latency between the camera's display and the RHO's display, which basically measures the time it takes for network data to get from the camera to the RHO. The third stage represents the measurement of G2G latency between the source and the RHO's display.

[3]https://www.ovenmediaengine.com/ome

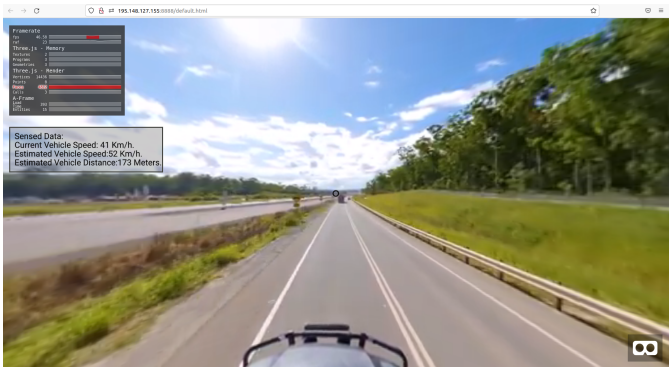
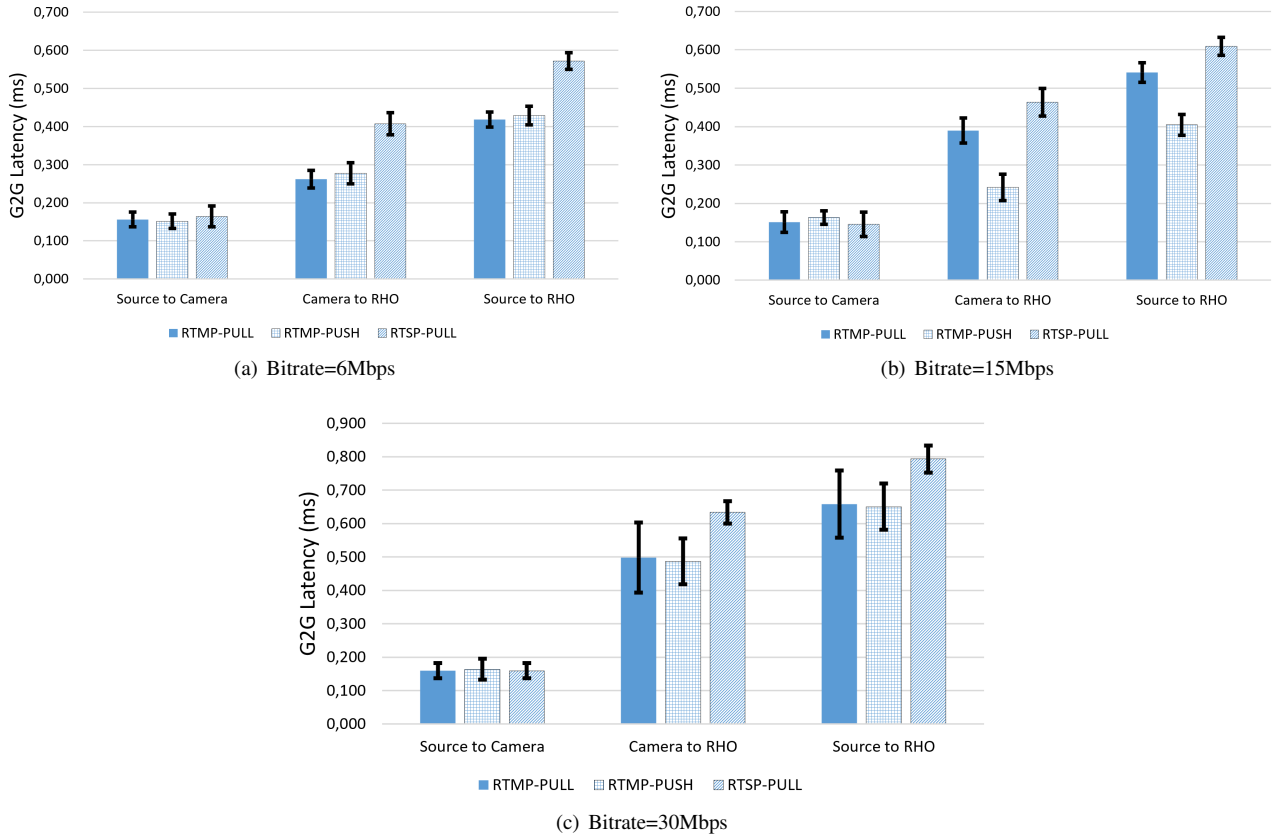


Fig. 3: 360° video stream augmented with sensed data about surrounding vehicles.

Figure 3 illustrates a screenshot of the proof-of-concept system. In order to make these services more flexible and immersive, RHO can view the 360° live stream through VR via desktop computers, using a 360 web-based player, or wear a HMD. In addition, the 360° stream is augmented with relevant information, displayed on the left-top side, regarding the vehicle in need of assistance, as well as its surroundings. Our VR experience was built using A-Frame, an open-source framework for developing virtual reality solutions. Using the stream feedback and the additional information provided, the RHO can send command messages to the vehicle indicating whether to engage the highway while speeding up, slowing down, or even stopping, allowing other vehicles on the main road to pass.

1) 6Mbps: Figure 4(a) shows the comparison of the average G2G latency, along with the standard deviation, between the three scenarios that correspond to the described streaming protocols with an encoding rate of 6Mbps. The acquisition, encoding, and display time, at the camera's display, are more or less the same (between 144ms and 164ms) for all three scenarios. However, the RTSP protocol has significantly higher network latency than both RTMP push and pull modes, while there is only a slight difference between the two RTMP modes.

2) 15Mbps: In Figure 4(b), we compare the G2G latency when the stream is encoded at 15Mbps. Acquisition and encoding latency appears to be roughly comparable to that



(a) Bitrate=6Mbps

(b) Bitrate=15Mbps

(c) Bitrate=30Mbps

Fig. 4: Glass-to-glass latency of different protocols when streaming at different bitrates.

of a 6Mbps encoded stream. In terms of network latency, it is remarkably higher than the stream encoded at 6Mbps for only RTSP and RTMP pull modes. However, for RTMP push mode, the network latency is roughly the same at both 6Mbps and 15Mbps, resulting in the best G2G latency between the source and the RHO's display.

3) 30Mbps: The comparison between the different streams encoded at 30Mbps depicted in Figure 4(c) shows a noticeable increase in the G2G latency for all streaming modes. Since the camera's hardware is capable of handling different bitrates encoding at nearly the same time, the G2G latency increase is largely caused by network delays.

VI. Conclusions

In autonomous vehicles, conventional cameras are one of the primary sensors that have a crucial role in various operations. Employing 360° cameras might bring new benefits and open up innovative applications.

In this work, we examine the use of VR technology, using 360° video stream, for controlling remote services. In particular, we focus on remote assistance for autonomous cars by RHO for better estimation and analysis in critical situations, such as merging onto the highway, where just the vehicle's sensors may not be enough to provide assistance to support autonomous operations. Upon receiving the assistance request, the RHO can view the 360° live stream either from a computer or through HMD for a more captivating experience, increased flexibility, and faster reaction time, by simply rotating the RHO's head. In addition, the 360° stream is augmented with other relevant information, such as the speed of the self-driving vehicle and the distance to the nearest object in the road. We mainly focused on evaluating the G2G latency of a 360° 4K live stream using different bitrates and protocols. Experimental results show that G2G latency can be lowered to near 400ms by pushing the RTMP stream towards the server, while RTSP exhibits the highest G2G latency of 800ms. In either case, the G2G latency is sub-second and allows enough time for the RHO to make timely decisions.

For future work, we plan to follow up our current work with real-time object detection and measure both the distance and speed of the detected vehicles on the road. Furthermore, to have a more complete experimental environment, we plan to incorporate this work with self-driving simulation environments, such as Carla, in order to gather data about road participants from the simulated environment.

Acknowledgment

This work was partially supported by the CHARITY project that received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016509. It is also partially funded by the Academy of Finland Project 6Genesis under grant agreements No. 318927. Prof. Song was supported by the Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government (MSIT) (2021-0-00188).

REFERENCES

[1] K. I. Adenuga, R. O. Adenuga, A. Ziraba, and P. E. Mbuh, "Healthcare augmentation: Social adoption of augmented reality glasses in medicine," in *Proceedings of the 2019 8th International Conference on Software and Information Engineering*, ser. ICSIE '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 71–74. [Online]. Available: https://doi.org/10.1145/3328833.3328840

[2] D. V. McGehee, E. N. Mazzae, and G. S. Baldwin, "Driver reaction time in crash avoidance research: Validation of a driving simulator study on a test track," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 20, pp. 3–320–3–323, 2000. [Online]. Available: https://doi.org/10.1177/154193120004402026

[3] C.-L. Fan, W.-C. Lo, Y.-T. Pai, and C.-H. Hsu, "A survey on 360° video streaming: Acquisition, transmission, and display," *ACM Comput. Surv.*, vol. 52, no. 4, Aug. 2019. [Online]. Available: https://doi.org/10.1145/3329119

[4] J. R. Rambach, G. Lilligreen, A. Schäfer, R. Bankanal, A. Wiebel, and D. Stricker, "A survey on applications of augmented, mixed and virtual reality for nature and environment," in *Proceedings of the 23rd HCI. International Conference on Human-Computer Interaction (HCII-2021), July 24-29, Washington, DC, United States*. Springer, 2021, virtual Conference.

[5] G. Noronha, C. Álvares, and T. Chambel, "Sharing and navigating 360° videos and maps in sight surfers," in *Proceeding of the 16th International Academic MindTrek Conference*, ser. MindTrek '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 255–262. [Online]. Available: https://doi.org/10.1145/2393132.2393189

[6] M. Ohta, S. Nagano, K. Nagata, and K. Yamashita, "Mixed-reality web shopping system using panoramic view inside real store," in *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*, ser. SA '15. New York, NY, USA: Association for Computing Machinery, 2015. [Online]. Available: https://doi.org/10.1145/2818427.2818456

[7] C. Ozcinar, A. De Abreu, and A. Smolic, "Viewport-aware adaptive 360° video streaming using tiles for virtual reality," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 2174–2178.

[8] M. Xi, M. Adcock, and J. Mcculloch, "An end-to-end augmented reality solution to support aquaculture farmers with data collection, storage, and analysis," in *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry*, ser. VRCAI '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3359997.3365721

[9] C. Haidon, H. K. Ngankam, S. Giroux, and H. Pigot, "Using augmented reality and ontologies to co-design assistive technologies in smart homes," in *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, ser. IUI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 126–127. [Online]. Available: https://doi.org/10.1145/3379336.3381492

[10] K. Jung, V. T. Nguyen, D. Piscarac, and S.-C. Yoo, "Meet the virtual jeju dol harubang—the mixed vr/ar application for cultural immersion in korea's main heritage," *ISPRS International Journal of Geo-Information*, vol. 9, no. 6, 2020. [Online]. Available: https://www.mdpi.com/2220-9964/9/6/367

[11] A. Alalewi, I. Dayoub, and S. Cherkaoui, "On 5g-v2x use cases and enabling technologies: A comprehensive survey," *IEEE Access*, vol. 9, pp. 107 710–107 737, 2021.

[12] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran, and M. D. Silva, "Vr is on the edge: How to deliver 360° videos in mobile networks," in *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network*, ser. VR/AR Network '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 30–35. [Online]. Available: https://doi.org/10.1145/3097895.3097901

[13] S. Aggarwal, S. Paul, P. Dash, N. S. Illa, Y. C. Hu, D. Koutsonikolas, and Z. Yan, "How to evaluate mobile 360° video streaming systems?" in *Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications*, ser. HotMobile '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 68–73. [Online]. Available: https://doi.org/10.1145/3376897.3377865

[14] A. Yaqoob, T. Bi, and G.-M. Muntean, "A survey on adaptive 360° video streaming: Solutions, challenges and opportunities," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2801–2838, 2020.

[15] O. E. Marai and T. Taleb, "Smooth and low latency video streaming for autonomous cars during handover," *IEEE Network*, vol. 34, no. 6, pp. 302–309, 2020.

Oussama El Marai received his engineering degree in computer science in 2005 from the University of Science and Technology Houari Boumediene (USTHB), Algiers, Algeria, the master degree from Ecole Nationale Superieure d'Informatique (ESI) in 2009, and he is pursuing his doctoral degree at the School of Electrical Engineering, Aalto University, Finland. His research interests include adaptive video delivery, user QoE optimization, digital twin, image processing, vehicular technology and intelligent transportation systems.

Tarik Taleb (Senior Member, IEEE) is currently a Professor at University of Oulu, Finland. He is the founder and director of the MOSA!C Lab (www.mosaic-lab.org). Between Oct. 2014 and Dec. 2021, he was a Professor at Aalto University. Prior to that, he was a senior researcher and 3GPP standards expert at NEC Europe Ltd., Germany. He also worked as assistant professor at Tohoku University, Japan. He holds a B.E. degree in information engineering, and M.Sc. Ph.D. degrees in information sciences from Tohoku University. His research interests lie in the field of beyond 5G and 6G, Autonomous Telco Cloud, and Network Softwarization.

JaeSeung Song (Senior Member, IEEE) is currently a professor at Sejong University. He received a Ph.D. from Imperial College London in the Department of Computing, United Kingdom. He holds B.S. and M.S. degrees in computer science from Sogang University. His research interests span the areas of beyond 5G and 6G, AI/ML enabled network systems, software engineering, networked systems and security, with focus on the design and engineering of reliable and intelligent IoT/M2M platforms.