Journal of Unmanned System Technology

A Novel Study on Data Rate by the Video Transmission for Teleoperated Road Vehicles

Tito Tang, Frederic Chucholowski, Min Yan, and Markus Lienkamp
Institute of Automotive Technology, Technische Universität München, München, Germany

Abstract—This paper presents three approaches to reduce the data rate required for video image transmission of teleoperated road vehicles and compares their efficacy to that of an ordinary H.264/AVC video compression. In addition, the paper suggests a setup of the open source H.264/AVC encoder x264 that provides a robust video stream with minimized encoding latency. While the two investigated edge detection algorithms, Canny-Algorithm and Laplace-Algorithm, tends to increase data rate, the Foveated-Imaging-Technique can reduce data rate for less important parts of the image. All investigated algorithms, including video compression, run in real-time since their computation time for one frame is less than forty milliseconds.

Keywords—Teleoperation, road vehicles, data reduction, video streaming.

I. INTRODUCTION

RESEARCH on the subject of “autonomous vehicles” has been increasing significantly since the 1980s. The fields of autonomous vehicle navigation [1] and control [2] are just some of the areas of interest. In addition to the early successes of the PROMOTHEUS EU-Project from the 1990s, the competitions sponsored by the U.S. Department of Defense (Darpa Challenges) [3],[4] have influenced public perception. The Institute devoted itself to the field of autonomous vehicles as part of a research topic from the “Deutsche Forschungsgemeinschaft (DFG; English: German Research Foundation)” Collaborative Research Center [5]. In summary, it can be stated that automatic driving is already possible, but only in certain less complex situations. Some highway, congestion or stop-and-go traffic scenarios partly fulfill these conditions. For other scenarios, especially in the area of urban traffic, there will probably be no autonomous vehicles driving on public roads in the next few decades. The main reason is that the environment is too complex because of the different road users and confusing road topologies.

Considering the abovementioned reasons, research at the Institute of Automotive Technology focuses on “teleoperated vehicles” and “semi-autonomous vehicles”. This approach, already known in the field of robotics [[6] & [7]], consists of bringing a human driver into the control loop in order to deal with the complexities of the road.

Since humans are good at interpreting complex scenarios [8], an external operator drives the vehicle remotely [9] using a live streaming video. Here, the operator is connected to the vehicle via wireless data transmission, such as a cellular network. A server handles the communication and can additionally perform computation-intensive tasks and provide information from the Internet to the driver, such as upcoming road conditions. The communication consists of a bidirectional system, where control signals are sent from the operator workstation to the vehicle while relevant vehicle information, such as vehicle speed and position, are simultaneously sent back to the operator at the workstation. One important piece of information is the video stream from the on-site environment.

Current technologies such as Wireless Local Area Networks (WLAN or Wi-Fi), Universal Mobile Telecommunications Systems (UMTS) and Long Term Evolution (LTE) provide the possibility of establishing a wireless connection between the vehicle and the workstation. Although the recent standard in WLAN technology, 802.11n, provides sufficient bandwidth for the transmission of video streaming, the operating range is very limited [10]. In comparison, cellular networks have a much higher operating range but a more limited bandwidth. The nominal uplink rate for a release 8 3G network is 11 Mbit/s[1] and for current 4G networks it is 75 Mbit/s. However, the actual achievable data rate is often restricted by the provider and highly depends on the number of simultaneous users in one cell and signal strength [11].

A raw uncompressed digital video image normally needs a large video rate. For example, using current camera settings a bandwidth of 58.6 Mbit per second is necessary, for a gray-scale video sequence with a resolution of 640 x 480 pixels and a bit rate of 25 frames per second.

It is clear from this calculation that video data cannot be transmitted in raw uncompressed format even with the current 4G cellular networks. It is therefore necessary to compress the video images of the robot’s surroundings to be able to send them to the operator using a cellular connection.

This paper presents and evaluates the three approaches: A) Canny Algorithm, B) Laplace Algorithm and C) Foveated-Imaging-Technique, to reduce the required data rate. In addition, a configuration of the open source H.264/AVC video encoder x264 is presented to generate efficient and robust video streams with minimized encoding latency.

Corresponding author: Tito Tang Chan (e-mail: tang@ftm.mw.tum.de).
This paper was submitted on November 07, 2013; revised on December 16, 2013; and accepted on March 13, 2014.

---

[1] 3rd Generation Partnership Project (3GPP), “3GPP – HSPA”, Version 2013, July 30. Available: http://www.3gpp.org/HSPA

II. STATE OF THE ART

As mentioned earlier, the term “teleoperation” refers to the remote operation of a vehicle or a system [9]. Teleoperation can be classified into different categories according to their type of data communication or power supply. Here the systems can be classified into “wired” and “wireless” operation. In the case of wired teleoperation, the robot or remote-controlled system is connected to a control station through a cable that provides the necessary power supply and the interface for data communication. On the other hand, in the case of wireless teleoperation, there is no cable between the control station and the remote-controlled system. The system must therefore possess an independent energy source. The data communication between the system and the operator is achieved through a cellular network.

For video data transmission, it is common to see the use of video CODECs (COder – DECoder) to compress the data volume transmitted. FFMPEG, a complete solution for gathering and converting audio and video streams, is a commonly used free library for multimedia processing. Many commonly known programs make use of the FFMPEG project for their development. FFMPEG’s libavcodec library supports more than 100 different codecs, including H.264/AVC, H.263, MPEG-2, MPEG-4 and MPEG-TS [12]. For the compression of H.264/AVC video it makes use of the open source x264 library.

III. PROBLEM STATEMENT

Teleoperated systems use video transmission systems in order to help the operator understand the environment surrounding the robot. A high bandwidth is required for a smooth display of the video streams. However, current available bandwidth through cellular networks is limited. The current paper presents a novel study on methods to reduce the data rate for the video data transmission used in teleoperated vehicles.

IV. EXPERIMENTAL SETUP

The Institute is devoted to research on “teleoperated vehicles” as an alternative to Unmanned Ground Vehicles in scenarios that cannot yet be fully interpreted by autonomous vehicles, such as urban traffic. An experimental vehicle is used at the Institute for the corresponding research activities. It consists of an Audi Q7 with corresponding sensors and actuators. The Audi Q7 is equipped with 8 cameras to provide a full view of the surroundings. Five GigE Ethernet cameras provide a front view of approximately 240^0. Two cameras are installed at each of the side mirrors and an extra camera is installed on the back of the vehicle to provide a view of the surroundings behind the vehicle [13].

Figure 1 shows a diagram of the current setup at the Institute of Automotive Technology. The Audi Q7 vehicle is connected to the Internet through cellular networks. Similarly, the operator workstation is connected to the internet via a LAN connection. Data coming from the cameras on the vehicle are transmitted to the workstation through the above-mentioned system; at the same time, control signals coming from the operator are sent to the vehicle through the same communication channel.

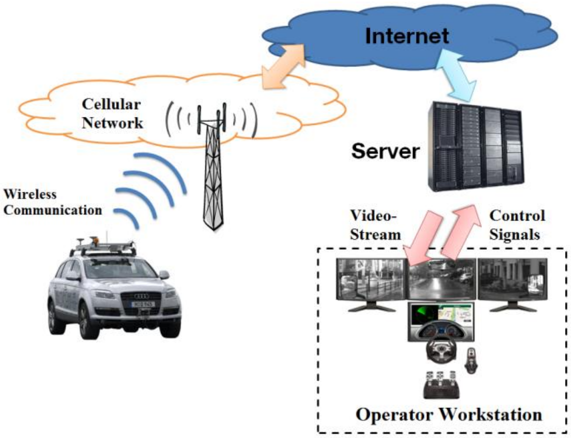

Figure 1 Diagram of the teleoperated vehicle at the Institute

The experimental vehicle and its overall design are described in [5] and the teleoperation interface is further described in [14].

Currently each video from the front GigE cameras is set to transmit video streams with a resolution of 640 x 480 pixels, at a frame rate of 25 fps. The experiments are conducted with two previously recorded scenarios consisting of videos taken during test drives. Scenario 1 consists of a homogeneous scenario without other vehicles. Scenario 2 is a more complex test drive, where different vehicles and pedestrians appear in the video sequence. **Figure 2** and **Figure 3** show sample frames of the two scenarios.

Studies and comparisons of the results are made using the two above-mentioned scenarios. Scenario 1 has a length of 1 minute 42 seconds at 25 fps bitrate, and Scenario 2 has a length of 2 minutes 50 seconds at 25 fps bitrate.

Figure 2 Sample frame from Scenario 1

Figure 3 Sample frame from Scenario 2

V. APPROACH

In the following we will demonstrate the Canny-Algorithm, the Laplace-Algorithm and the Foveated-Imaging-Technique, which might be suitable for data rate reduction. We also provide an x264 H.264/AVC encoder configuration to generate efficient and robust video streams with minimum encoding latency.

A. Canny Algorithm

The Canny-Algorithm for edge detection was developed in 1986. This method was named after its developer, John Canny [1]. The Canny operator is known as one of the best edge-detection algorithms and is still capable of delivering very competitive results 20 years after its development. Three main goals were defined for the Canny operator:

- Good detection
- Good localization
- Minimal number of incorrectly detected edges

Figure 4 shows an example of edge detection using the Canny-Algorithm for Scenario 1.

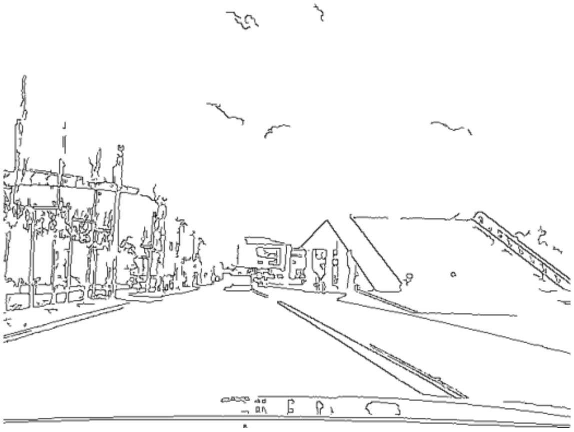

Figure 4 Edge-detection sample using Canny-Algorithm

The optimal function of the Canny-Algorithm consists of the sum of a constant and a set of four exponentials. This set of exponentials can be approximated by the first derivative of a Gaussian. The points that represent the local maximum values in the second derivative are the candidates for the number of edges [16]. Gaussian smoothing is first implemented in order to reduce the noise. The maximum values in the four directions are then used as candidates for the edges. Large slopes in intensity tend to belong to edges.

B. Laplace-Algorithm

The operator developed in 1979 by Marr and Hildreth, which represents the second derivative of a Gaussian-function, is also known as “Laplacian of Gaussian” [17]. The basic principle involves computing the second derivative by a function and using the zero-crossing as the criteria for edge detection. A Gaussian noise reduction filter should be implemented first because of the strong influence of noise. The Laplace operator can be represented using (1) and (2) [16], where I(x,y)represents the brightness value of a pixel at the position (x,y). **Figure 5** shows an example of edge detection using the Laplace-Algorithm for Scenario 1.

$$L(x,y) = \nabla^2 I(x,y) \tag{1}$$

$$\nabla^2 I(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \tag{2}$$

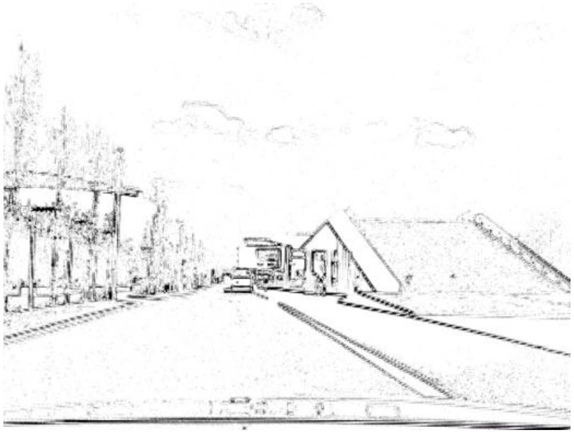

Figure 5 Edge-detection sample using LaPlace-Algorithm

C. Foveated Imaging Technique

There is a limitation in human perception in which the spatial resolution of the human visual system decreases dramatically from the point of fixation. A perspective of 2.5 degrees against the focus point has the highest resolution. Outside this range and up to a 10-degree perspective, the resolution is reduced by a factor of 2. From a perspective of 10 degrees on the human eye only perceives one-tenth of the central resolution [18].

The application of video transmission in teleoperated vehicles consists of a constant resolution throughout the entire image. It is possible to reduce the data rate by using the limits of human perception. In other words, bandwidth is being used in order to send information that the human eye cannot perceive. A Foveated-Imaging-Technique system normally consists of two or three different resolutions; a high resolution range will only be displayed at the point of interest (or point of focus). One method for reducing the resolution in the video frames is using the Gaussian-Image-Pyramid.

Figure 6 Sample frame of a Foveated-Image with an area of interested focused on the vehicle

Figure 6 shows an example of a frame where the Gaussian-Image-Pyramid method is implemented to reduce the resolution of the whole video frame. An area of interest with a focal point is defined for the vehicle on the right. Only the area of interest is presented in high resolution.

D. H.264/AVC Video Compression

We used the state of the art H.264/AVC video codec for video compression. The open source project x264 provides an easy to use compression library with good encoding efficiency for the H.264/AVC video codec. Modern video compression algorithms normally utilize I-frames or I-slices, P-frames/slices and B-frames/slices. Each I-frame contains all the information required to reconstruct the full picture without dependencies on other images. P-frames reference one previous image and contain movement predictions of specific areas of the image. B-frames can reference previous and additionally also following frames [19]. To be able to reference following frames, the encoder has to process several video images before the first encoded frames can be output. This results in an additional time delay during image processing, which is critical for real time video streaming with minimum time delay. We therefore prevent the use of B-frames.

Since the data size of P- and B-frames is usually significantly smaller than that of I-frames, data rate can be minimized by using as few I-frames as possible. However, in case of packet loss during the transmission, the P- and B-frames cannot be decoded correctly since they might reference lost frames. It is therefore important to find a tradeoff between data rate and robustness. For our system this is one I-frame at least every two seconds.

The bigger size of I-frames also has the negative effect of data rate bursts when these are generated. To avoid these we make use of the "intra refresh" feature provided by the x264 encoder. Instead of encoding a full image as I-frame, smaller parts of the image alternately contain all necessary information to reconstruct the part of the image without references. These are called I-slices. This distributes the required data rate of I-frames to a longer time span.

The H.264/AVC standard defines several profiles for different purposes that determine which compression algorithms and features may be used. The "baseline profile" for example is often used in embedded systems and only supports I- and P-slices while the "high profile" is often used for high definition disc storage and broadcasting and also supports B-slices. Since the omission of B-slices decreases encoding latency we use the "baseline profile". In addition to profile settings, the x264 encoder provides parameter presets which preconfigure the encoder for specific tasks. Since the reduction of encoding time is more important than image quality for our system, we use the "veryfast" preset in conjunction with the "zerolatency" tuning option[2]. The "zerolatency" tuning option disables all features which would result in additional encoding delays, such as rate control lookahead, multiple threads or B-frames.

These settings lead to a robust video stream with minimum encoding latency. The required data rate depends on the specified image quality, which can be adjusted dynamically depending on the available transmission bandwidth.

VI. RESULTS AND DISCUSSION

TABLE I shows the average time needed for processing each video frame using the respective algorithm. The computer used is a CPU Intel© Core (TM) i5, 2.4 GHz M520 running on Windows XP operating system with 2.0 GB RAM.

At the Institute of Automotive Technology, real-time implementation is an important aspect for the application of teleoperated vehicles. In order to achieve real-time implementation, a bit rate of 25 frames per second is necessary, which corresponds to a maximum processing time of 40 milliseconds per frame. The average time for data processing does not exceed the limit of 40 milliseconds in any of the cases, thus guaranteeing a real-time application.

TABLE I AVERAGE TIME FOR PROCESSING EACH FRAME

	Scenario 1	**Scenario 2**
H.264/AVC	10.17 [ms]	11.46 [ms]
Canny-Algorithm	27.27 [ms]	27.06 [ms]
LaPlace-Algorithm	14.08 [ms]	14.20 [ms]
Foveated-Imaging	09.00 [ms]	10.37 [ms]

A comparison in data size for the two edge-detection algorithms Canny-Algorithm and Laplace-Algorithm, the Foveated-Imaging-Technique and the H.264/AVC encoded video can be seen in **Figure 7** for both scenarios.

Edge-detection algorithms show an increase in data rate. In addition, they greatly reduce the amount of information that the video images provide to the operator. These are therefore not

---

[2] MeWiki, "X264 Settings," Version 2012, December 9. Available: http://mewiki.project357.com/index.php?title=X264_Settings&oldid=5207

suitable to reduce data rates for the teleoperated driving. One of the main reasons for this increase in data rate is that H.264/AVC uses the “Tree structured macroblock partition and motion compensation” [20] technique. The quantity of bits needed to represent the motion vectors and macroblocks allocation can vary depending on the selection of the macroblocks and sub-macroblocks. The selection of the macroblock allocation plays an important role in determining the compression rate. In the case of a “sharp” frame, a higher number of smaller macroblocks is needed to completely represent its details. The more detailed the frame is the smaller the compression rate achieved by the CODEC. For a frame represented with edges, a higher number of smaller macroblocks is needed, thus increasing the data rate.

The Foveated-Image-Technique, on the other hand, shows a slight decrease in data rate, which would produce a considerable amount of data reduction. In the first scenario, a reduction of about 19% is achieved and in the second scenario, a reduction of about 17%. However, this comes with the negative effect of reduced image resolution.

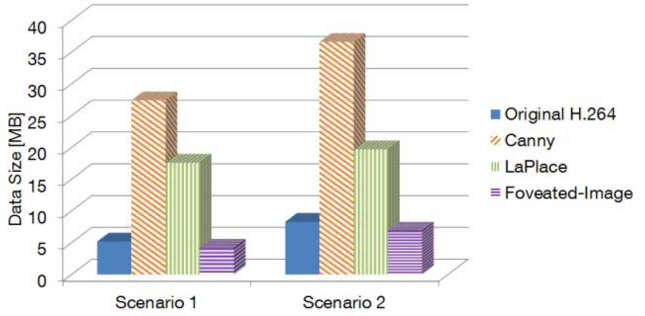

Figure 7 Comparison between original video with H.264/AVC CODEC and edge-detection algorithms and with Foveated-Imaging Technique

VII. CONCLUSIONS

This paper presents a novel study for data reduction in video transmission for teleoperated vehicles. The study consists of implementing different approaches to reduce the data rate needed for data transmission from the teleoperated vehicle to the operator. These approaches are the edge-detection algorithms Canny-Algorithm and Laplace-Algorithm and the Foveated-Imaging-Technique. Additionally it presents the setup of the x264 H.264/AVC video encoder to get an efficient and robust video stream with minimum encoding latency and compares data rate to that of the presented approaches.

While all provided approaches satisfy the requirement of real-time capability, results show that edge-detection algorithms do not provide a reduction in video data rate. However, taking advantage of the fact that humans can only perceive an area of the entire picture in high resolution, the Foveated-Imaging-Technique could be implemented to transmit the entire frame at a lower resolution, with only the area of interest in higher resolution. Results showed that the data rate can be reduced compared to original H.264/AVC encoding with the drawback of lower image resolution for most parts of the image.

VIII. FURTHER RESEARCH

In order to implement the Foveated-Imaging-Technique in real-time teleoperated vehicles applications, the area of interest has to be known. Further research concentrates on head-tracking and eye-tracking devices to determine the operator’s angle and point of focus. Results from the head- and eye-tracking devices will be merged to determine the area of the video that needs to be transmitted in high resolution.

The required data rate can probably be further reduced using the upcoming High Efficiency Video Codec (HEVC or H.265). Once the standard is released and efficient encoders are available, this codec should further be investigated.

Besides the problems with bandwidth for video stream transmission, another important problem that needs to be solved is the time delays in transmission. These time delays influence the overall performance of the driving task. The Institute of Automotive Technology addresses the above mentioned problem through the development of different approaches. Among these are indirect control strategies as shown in [21] or predictive displays as shown in [22] and [23].

REFERENCES

[1] C. K. Ngai and N. H. C. Yung, “DAQL-Enabled Autonomous Vehicle Navigation in Dynamically Changing Environment,” in in *Advances in Reinforcement Learning*, A. Mellouk, Ed. 2011, pp. 385–410.

[2] M. B. Trabia, L. Z. Shi, and N. E. Hodge, “A Fuzzy Logic Controller for Autonomous Wheeled Vehicles,” in in *Mobile Robotics, Moving Intelligence*, J. Buchli, Ed. 2006, pp. 175–200.

[3] T. Jungmann, “Grand Challenge 2005, die Rallye der Rechner,” *ATZ Online*, 2005.

[4] R. Schedel, “Darpa Urban Challenge 2007,” *ATZ - Automob. Zeitschrift*, vol. 110, no. 1, pp. 28–30, 2008.

[5] M. Goebl, M. Althoff, M. Buss, G. Faerber, F. Hecker, B. Heissing, S. Kraus, R. Nagel, F. P. Leon, F. Rattei, M. Russ, M. Schweitzer, M. Thuy, C. Wang, and H. J. Wuensche, “Design and capabilities of the Munich Cognitive Automobile,” in Intelligent Vehicles Symposium, 2008 IEEE, 2008, pp. 1101–1107.

[6] M. Nohmi and T. Bock, “Contact Task by Force Feedback Teleoperation Under Communication Time Delay,” in *Robotics and Biomimetics, 2006. ROBIO ’06. IEEE International Conference on*, 2006, pp. 55–60.

[7] M. U. Keerio, W. Zhang, and A. R. Jafri, “Cooperation of Humanoid Robots using Teleoperation for Transferring an Object,” *Int. J. Adv. Robot. Syst.*, vol. 5, no. 1, pp. 31–40, 2008.

[8] H. Winner, S. Hakuli, and G. Wolf, Eds., *Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort*. Vieweg + Teubner, 2009.

[9] T. Fong and C. E. Thorpe, “Vehicle teleoperation interfaces,” *Auton. Robots*, vol. 11, no. 1, pp. 9–18, 2001. CrossRef

[10] Kioskea, “Introduction to Wi-Fi (802.11 or WiFi),” 2013. [Online]. [Accessed: 31-Oct-2013] VIEW ITEM

[11] S. Tenorio, K. Exadaktylos, B. McWilliams, and Y. Le Pezennec, “Mobile broadband field network performance with HSPA+,” in *Wireless Conference (EW), 2010 European*, 2010, pp. 269–273.

[12] FFmpeg Team, “FFmpeg General Documentation.” 2013.

[13] S. Gnatzig, F. Chucholowski, T. Tang, and M. Lienkamp, “A System Design for Teleoperated Road Vehicles,” in *ICINCO2013, 10th International Conference on Informatics in Control, Automation and Robotics*, 2013, pp. 231–238.

[14] F. Diermeyer, S. Gnatzig, F. Chucholowski, T. Tang, and M. Lienkamp, “Der Mensch als Sensor - Der Weg zum teleoperierten Fahren,” in *AAET*, 2011.

[15] J. Canny, “A Computational Approach to Edge Detection,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. PAMI-8, no. 6, pp. 679–698, 1986. CrossRef

[16] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed. O’Reilly Media, 2008, p. 555.

[17] D. Marr and E. Hildreth, “Theory of Edge Detection,” *Proc. R. Soc. London. Ser. B. Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, 1980.

[18] W. S. Geisler and J. S. Perry, “A real-time foveated multiresolution system for low-bandwidth video communication,” in *Human Vision and Electronic Imaging, SPIE Proceedings*, 1998, pp. 294–305.

[19] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *Circuits Syst. Video Technol. IEEE Trans.*, vol. 13, no. 7, pp. 560–576, 2003. CrossRef

[20] A. Puri, X. Chen, and A. Luthra, “Video coding using the H.264/MPEG-4 AVC compression standard,” *Signal Process. Image Commun.*, vol. 19, no. 9, pp. 793–849, 2004. CrossRef

[21] S. Gnatzig, F. Schuller, and M. Lienkamp, “Human-machine interaction as key technology for driverless driving - A trajectory-based shared autonomy control approach,” in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012, pp. 913–918.

[22] F. Chucholowski, S. Buechner, J. Reicheneder, and M. Lienkamp, “Prediction Methods for Teleoperated Road Vehicles,” in *Conference on Future Automotive Technology*, 2013.

[23] F. Chucholowski, M. Sauer, and M. Lienkamp, “Evaluation of Display Methods for the Teleoperation of Road Vehicles,” in *9th International Conference on Intelligent Unmanned Systems*, 2013.