Analyzing Image Quality Metrics for Situational Awareness Estimation

Scientific work to obtain the degree
Bachelor of Science (B.Sc.)
at the Department of Mobility Systems Engineering
of the TUM School of Engineering and Design
at the Technical University of Munich

Supervised by	Prof. Dr.-Ing. Markus Lienkamp
	Nils Jannik Gehrke, M.Sc.
	Chair of Automotive Technology
Submitted by	Miguel Navarro Regodón, B.Sc.
	Lochhamer 7
	82152 Planeg
Submitted on	16.08.2024

Chair of Automotive Technology
TUM School of Engineering and Design
Technical University of Munich

Project description

Analyzing Image Quality Metrics for Situational Awareness Estimation

The future of our roads belongs to automated driving. In the near future, however, automated vehicles (AVs) will fail in different situations in our everyday traffic, making fallback solutions relevant for successful mission completion. In such an AV fail case, a remote operator will replace the failed AV module from afar and assist the AV in the specific situation. For this assistance, the operator must have high situational awareness and understanding. The situational awareness is built primarily using the available video stream that has undergone compression and possible data loss during transmission.

The goal of this bachelor thesis is to research and implement different image quality metrics and evaluate and discuss their correlation with situational awareness. To achieve this goal, in the first step, you will perform a literature review with regard to image quality metrics, and already identified correlations with the human to understand and identify the content. You will then select a subset of the identified metrics to implement them in either C++ or Python to analyze a ROS2 video stream on a frame basis. The implemented metrics will then be used to evaluate via a study the correlation between the metric scores and the capability of humans to identify objects in the image and build up their situational awareness. This includes deriving a study design, conducting the study, and evaluating the study results to obtain the correlation values. Further, the results are discussed and summarized.

The thesis comprises the following work packages:

- Literature review regarding image quality metrics and their correlation with human perception
- Recording suitable sample video scenes with the research vehicle EDGAR that include different environments, light conditions and weather conditions as well as visually challenging scenarios
- Selection of a promising set of image quality metrics that are part of an open source framework
- Designing and Conducting a study to determine the correlation between metric scores and the human situational awareness / human perception of objects in the video stream
- Evaluation of study results, discussion and summary

Experience with Python is required. Previous work with respective Python libraries (e.g. cv2) and experience with study design is of help.

The thesis should document the individual work steps in a clear form. The candidate undertakes to complete the Bachelor's thesis independently and to indicate the scientific aids used.

Ausgabe: 16.08.2024

Abgabe: 16.08.2024

Prof. Dr.-Ing. M. Lienkamp

Betreuer: Nils Jannik Gehrke, M. Sc.

Chair of Automotive Technology
TUM School of Engineering and Design
Technical University of Munich

Geheimhaltungsverpflichtung

Herr/Frau: **Navarro Regodón, Miguel**

Gegenstand der Geheimhaltungsverpflichtung sind alle mündlichen, schriftlichen und digitalen Informationen und Materialien, die der Unterzeichner vom Lehrstuhl oder von Dritten im Rahmen seiner Tätigkeit am Lehrstuhl erhält. Dazu zählen vor allem Daten, Simulationswerkzeuge und Programmcode sowie Informationen zu Projekten, Prototypen und Produkten.

Der Unterzeichner verpflichtet sich, alle derartigen Informationen und Unterlagen, die ihm während seiner Tätigkeit am Lehrstuhl für Fahrzeugtechnik zugänglich werden, strikt vertraulich zu behandeln.

Er verpflichtet sich insbesondere:

- derartige Informationen betriebsintern zum Zwecke der Diskussion nur dann zu verwenden, wenn ein ihm erteilter Auftrag dies erfordert,
- keine derartigen Informationen ohne die vorherige schriftliche Zustimmung des Betreuers an Dritte weiterzuleiten,
- ohne Zustimmung eines Mitarbeiters keine Fotografien, Zeichnungen oder sonstige Darstellungen von Prototypen oder technischen Unterlagen hierzu anzufertigen,
- auf Anforderung des Lehrstuhls für Fahrzeugtechnik oder unaufgefordert spätestens bei seinem Ausscheiden aus dem Lehrstuhl für Fahrzeugtechnik alle Dokumente und Datenträger, die derartige Informationen enthalten, an den Lehrstuhl für Fahrzeugtechnik zurückzugeben.

Besondere Sorgfalt gilt im Umgang mit digitalen Daten:

- Für den Dateiaustausch dürfen keine Dienste verwendet werden, bei denen die Daten über einen Server im Ausland geleitet oder gespeichert werden (Es dürfen nur Dienste des LRZ genutzt werden (Lehrstuhllaufwerke, Sync&Share, GigaMove).
- Vertrauliche Informationen dürfen nur in verschlüsselter Form per E-Mail versendet werden.
- Nachrichten des geschäftlichen E-Mail Kontos, die vertrauliche Informationen enthalten, dürfen nicht an einen externen E-Mail Anbieter weitergeleitet werden.
- Die Kommunikation sollte nach Möglichkeit über die (my)TUM-Mailadresse erfolgen.

Die Verpflichtung zur Geheimhaltung endet nicht mit dem Ausscheiden aus dem Lehrstuhl für Fahrzeugtechnik, sondern bleibt 5 Jahre nach dem Zeitpunkt des Ausscheidens in vollem Umfang bestehen. Die eingereichte schriftliche Ausarbeitung darf der Unterzeichner nach Bekanntgabe der Note frei veröffentlichen.

Der Unterzeichner willigt ein, dass die Inhalte seiner Studienarbeit in darauf aufbauenden Studienarbeiten und Dissertationen mit der nötigen Kennzeichnung verwendet werden dürfen.

Datum: 16.08.2024

Unterschrift: _______

Erklärung

Ich versichere hiermit, dass ich die von mir eingereichte Abschlussarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Garching, den 16.08.2024

_____

Miguel Navarro Regodón, B. Sc.

Declaration of Consent, Open Source

Hereby I, Navarro Regodón, Miguel, born on November 13, 2002, make the software I developed during my Bachelor thesis available to the Institute of Automotive Technology under the terms of the license below.

Garching, 16.08.2024

_____

Miguel Navarro Regodón, B. Sc.

Table of contents

List of abbreviations

AD	Automated Driving
AI	Artificial Intelligence
SA	Situational Awareness
AV	Automated Vehicle
VQM	Video Quality Metric
ROS	Robot Operating System
SPIDER	Scanning, Predicting, Identifying, Deciding and Executing appropriate Responses
SAGAT	Situation Awareness Global Assessment Technique
6LM	6-Layer Model
V2X	Vehicle-to-Everything
RGB	Red Green Blue
MID	Morion-in-Depth
MOS	Mean Opinion Score
PSNR	Peak Signal-to-Noise Ratio
dB	Decibel
MSE	Mean Squared Error
SSIM	Structural Similarity Index Measure
VIF	Visual Information Fidelity
DLM	Detail Loss Metric
VMAF	Video Multi-Method Assessment Fusion
WebRTC	Web Real-Time Communication
HD	High Definition
VR	Virtual Reality
HS	Horn and Schunck
EPE	End-Point Error
HSV	Hue Saturation Value

Formula symbols

Formula symbols	Unit	Description
x	-	Vector from the reference image
y	-	Vector from processed image
$f(x,y)$	-	Input variable (color variable of the original pixel)
$g(x,y)$	-	Output variable (color variable of the processed pixel)
m	-	Number of pixels horizontally
n	-	Number of pixels vertically
I	-	Maximum pixel luminance value
$l(x,y)$	-	Local luminance comparison function
$C(x,y)$	-	Local contrast comparison function
$S(x,y)$	-	Local structure comparison function
μ_x	-	Sample mean of x
μ_y	-	Sample mean of y
σ_x	-	Sample standard deviation of x
σ_y	-	Sample standard deviation of y
σ_{xy}	-	Correlation coefficient between x and y
C_1	-	Constant for luminance function
C_2	-	Constant for contrast function
C_3	-	Constant for structure function
α	-	Relative importance of luminance
β	-	Relative importance of contrast
γ	-	Relative importance of structure
n	-	Total number of questions
i	-	Question number
ω_i	-	Assigned weight for question i

r_i	-	Rating for question i
$\sum$	-	Sum

1 Introduction

1.1 Motivation

Automated Driving (AD) has become a very popular technology in recent years. Different companies are actively developing this technology by making large investments in the sector; potential to improve safety and efficiency is also part of the actual research [1]. For example, in Europe, Easymile works to achieve sustainable public transport by using automated minibuses [2]. In America, Waymo develops this technology to avoid injuries and fatalities in the road as well as to improve mobility [3].

However, technology is only sometimes perfect; this means that AD will sometimes fail, and a solution is needed. Sinha et al. [4] summarized some of the accidents that occurred in San Francisco between 2014 and 2019 for the future deployment of the vehicles. The statistics on the type of accident are shown in Figure 1. Crash rate report [4]. This thesis aims to move forward to achieve that objective.

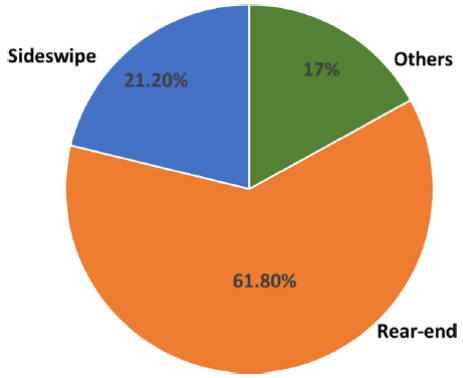

Figure 1. Crash rate report [4].

The most important condition to safe driving is perfectly perceiving and interpreting the vehicle's surroundings and the rest of the traffic. Being able to plan and control the situation according to them is crucial for that objective [5]. For that reason, for an operator controlling the vehicle from an external room, perfectly understanding the environment is even more crucial. It is essential to notice that the operator will typically intervene when the AD fails, which means that the situation the controller will be in may be a potential hazard and not easy to solve. Therefore, providing the operator with precise and comprehensive information is necessary to control such scenarios effectively.

Fully autonomous driving vehicles are already driving in the streets. In California, the legislation allows driverless taxis and typical vehicles to share the roads, as can be seen in Figure 2. Waymo

driverless taxi in the streets of San Francisco [3]. Therefore, this thesis is a good opportunity to explore and initialize in AD and teleoperation.

When an AD vehicle fails, a human teleoperator takes over control of the vehicle from a remote control center. For this assistance and for an effective intervention, the teleoperator must have good knowledge of the situation and the vehicle's surroundings to know how to act or what to do. This is called Situational Awareness (SA), and in the context of this thesis, it will mainly be achieved through live video stream sent by the Automated Vehicle (AV).

Figure 2. Waymo driverless taxi in the streets of San Francisco [3].

1.2 Research purpose

While driving, lots of different scenarios and events can occur. Therefore, it is crucial to identify the SA and the environment of the roads at the moment when the teleoperation is happening so that every important factor is communicated to the operator. Also, some things can be displayed differently in the teleoperation station than in reality. That could happen due to data loss during the video transmission, among other problems.

Live video stream is a critical source of real time information that allows teleoperators to stay informed during critical moments of intervention [6]. To ensure the safety and efficacy of the AV, video quality is essential to improve the SA of the operator. Therefore, the improvement and development of SA for the teleoperator is helpful for the field of AD.

During this thesis, multiple videos will be chosen to evaluate some important aspects of the SA and the environment of different driving situations. The evaluation will be done based on [7]. In this paper, many areas of the environment, divided into six layers of driving, are included. For the work, six videos were planned and filmed with the research vehicle EDGAR. Alternatively, if any video cannot be filmed due to weather conditions or problems, the simulator CARLA [8] can be used.

The main objective of the thesis is to improve and analyze the SA of the remote teleoperator. For that reason, some questions open up:

Can the SA of a teleoperator be measured?

How can the SA be measured?

Many options for this might be possible. SA is an abstract thing that cannot be directly evaluated from zero to ten. The SA of a scenario not only depends on the scenario but also on the driver or, in this case, the teleoperator. Experience, confidence and ability are some factors that influence the environment perception. Also, people do not perceive danger in the same way, which is also essential to SA assessment.

How can video quality be evaluated?

Many people think that resolution is equal to quality, but that is not true. The quality of a video is the result from the evaluation of multiple factors. For that reason, the existence of Video Quality Metrics (VQMs) is beneficial for this quality assessment. The different VQMs do not work all the same and do not focus on the same aspect of video quality. A deeper explanation of VQMs will be done in the next chapters.

This work will focus on estimating the SA by correlating it with the quality of the live video stream. Figure 3. Research methodology of the thesis depicts the methodology followed to achieve the objectives.

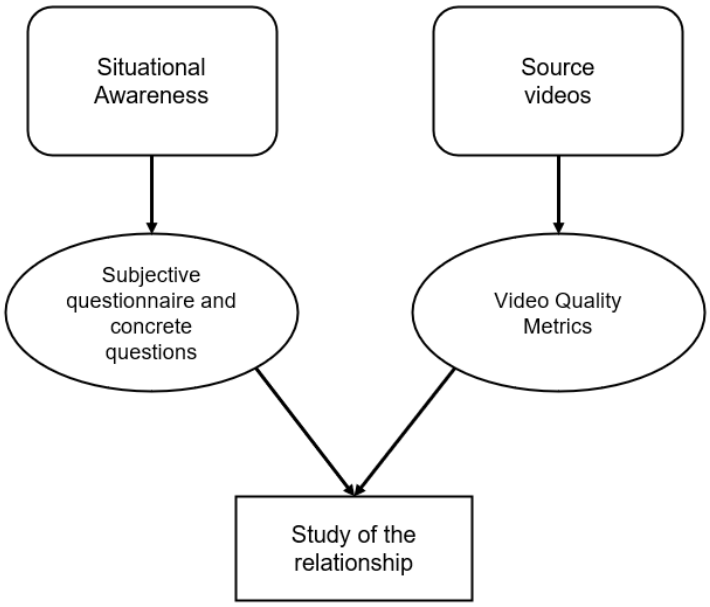

Figure 3. Research methodology of the thesis.

On one hand, the idea is to see if the perceived video quality is related to environment perception. SA evaluation is possible through subjective questionnaires and concrete observations. Also, some formal and detailed questions are essential to determine the environment's comprehension. For example, distance or time reactions can be useful in analyzing that perception.

On the other hand, video quality can be easily evaluated as it is data and can be processed. The actual state of the art provides different VQMs that are able to evaluate a source video frame by frame. Determining what VQM is suitable for this analysis is also part of this work.

1.3 Structure of the thesis

logical order of arguments is presented as part of the objective to achieve the thesis goals. In Chapter 2, the most important definitions and the state of the art development are done. In the end, the research gap of the thesis is explained according to the previous documentation.

Chapter 3 introduces the methodology that will be followed in the thesis. The discussion and election of six different video samples according to the current state of the art and SA definition is explained in this section. In this chapter, some VQMs are selected for later implementation in different environments. A subjective study will be developed, and that includes the creation of an online survey. The justification for the survey questions will be discussed in this chapter as well. The execution of the VQMs, the online survey, and the data process management method are also included.

The results obtained in the study and with the VQM assessment are displayed in Chapter 4. The analysis of this data is a great part of the chapter as it allows to achieve conclusions in the following chapters. Graphics and tables are part of the statistical results obtained in this part.

Chapter 5 is responsible for the discussion of the results previously obtained. Conclusive and consolidated statistical results will be shown to explain how good the objectives of the thesis were achieved. In the last section, chapter 6, the summary of the work is done, and conclusions are obtained. An overview of the work is also done.

2 State of the art

This section describes the state of the art in the relevant areas for automated driving. It includes the definition of teleoperation for AD and some teleoperation models, as well as the definition of SA and the suggested ways of evaluating it. On the other hand, some relevant qualities of human vision for the work are explained, and the current state of the art of the most important VQMs and some derivates are summarized. Also, the correlation between SA and VQMs is studied.

2.1 Teleoperation for automated driving

As AVs remain under continuous development, there are still multiple situations that cannot be solved by the vehicle itself. So-called disengagements, however, must be studied to maintain the potential of the business. There are two types of disengagements: failure detection and safety operations. Sinha et al. [9] classified and studied multiple AD disengagements from different companies and concluded that failure detection ones were more present with lower cumulative miles. Even though the author stated that previous studies are premature and cumulative disengagement studies in relation to cumulative miles are likely to be unreliable. A potential fallback solution to continue the vehicle operation is teleoperation [10].

Teleoperation has the potential to substitute certain tasks of the automation by a human. It is important to note that human and machines are good in complementary tasks. Humans thrive in cognitive tasks, which include situational analysis, decision making, and planning, due to their capability of coping, and this makes them valuable. On the other hand, machines can be more precise and have shorter reaction time [11]. This is further depicted in Figure 4.

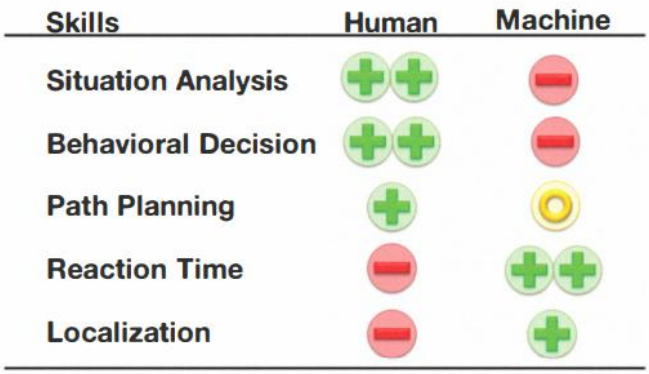

Figure 4. Relevant skills between humans and machines according to [10].

To replace different subsets of the automated driving software, multiple teleoperations are presented in the literature [12–14]. The further work will focus on six concepts described in [10].

Direct control

Direct control is the most fundamental teleoperation concept (Figure 5. Conceptual description of direct control concept according to [1].). It replaces the entire automated system and requires the vehicle to obtain and collect information of the surroundings to execute basic control commands [15].

In direct control setup, the input data are video streams from the vehicle. These are collected with cameras on the vehicle and sent to the operator through a private network. This requires compression of the images, and sound stimulus can also be sent from the vehicle. Schimpe et al. [6] proposed a software for carrying out various teleoperation concepts research. In the study, direct control was included as one of the control modes and explained how primary controls are directly transmitted to the vehicle.

Based on the video stream and other information proportioned by the vehicle, the operator controls the vehicle with commands similar to manual driving inside a vehicle. This includes steering wheel angle, throttle pressure, brake pressure, revolution management, and many other commands. Also, sound stimulus can be used. Then, all these commands are transmitted to the vehicle and executed [10].

Different studies have been conducted in relation to direct control. For example, [16] provided an open source framework for teleoperated driving. In the study, low-cost off-the-shelf components were used, and Robot Operating System (ROS) was the bridge to connect CARLA and the operator's station. A flexible video streaming framework was also presented [17] and achieved promising results on bitrate handling and optimizing video resolution scaling factors.

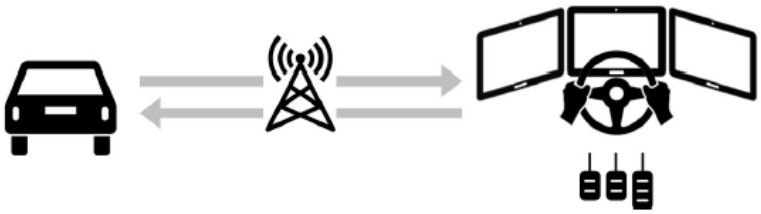

Figure 5. Conceptual description of direct control concept according to [1].

Shared control

Direct control relies on the operator to safely drive the vehicle. Humans, however, can also make mistakes. To encounter this issue, shared control is proposed. This concept is based on the collaboration between humans and autonomous software to accomplish a task. The type of interaction, the task type, and the forms of user feedback are aspects that the software must take into account. Experience has shown that full autonomy still faces some problems under challenging tasks and that shared control improves performance by not relying entirely on humans, even though lots of reliability problems are yet to be solved [18, 19].

Majstorović et al. [10] described shared control teleoperation as very similar to direct control in terms of command making. The main difference is that commands are accepted by the shared controller but not necessarily executed. The shared controller is responsible for determining the safety of the commands and deciding whether to execute them or intervene. In a critical situation, the shared controller may not consider the input commands safe and will reject them, overriding the control of the vehicle to avoid an accident [20].

Trajectory guidance

As an alternative to direct and shared control, trajectory guidance is created. In this teleoperation control mode, the operator provides trajectory commands and a velocity profile. The operator must perceive the environment and plan the commands. The autonomous system makes only low-level control decisions. Such decisions include stabilization of the vehicle and acceleration, among others. Some experiments have proven that trajectory guidance is a good option for teleoperation, but they have encountered some problems. Trajectory guidance requires the operator to look several meters ahead, and this is sometimes only possible in predictable environments [10].

Video streaming latency can also be a major problem in this teleoperation mode. Gnatzig et al. explained the possibility of safe and reliable teleoperated system based on trajectory guidance. In the same paper, a teleoperation system was designed, including path management, lateral control, and longitudinal control as key elements to trajectory guidance teleoperation. The system included some safety features, for example, the vehicle was set to always stop at the end of the path. The results showed that the teleoperator needs to view several meters ahead, even with the path already being defined and this is possible in scenarios with few unpredictable events [11].

Waypoint guidance

Waypoint guidance further reduces the responsibility of the operator compared to trajectory guidance. Via waypoint input from the operator inside a drivable area, the autonomous system is then responsible for planning a trajectory between these waypoints. This also includes setting a target velocity for the different trajectory sections by the vehicle and avoiding potential hazards along the trajectory. The vehicle takes over the same low-level control decisions from the trajectory guidance. Solely the decision making process remains at the human operator [10].

This concept is capable of functioning effectively in higher latencies. The operator's input is primarily based on the static environment surrounding the vehicle, and both latency and varying video quality do not present a safety concern during operation. In the event of obstacles within the driving path, the vehicle will come to a safe stop without the need for human intervention, which might result in time-inefficient stop-and-go driving behavior [10].

Interactive path planning

In the context of interactive path planning, the operator bears the responsibility for decision-making. The operator is presented with a variety of potential paths, generated by the automated software. It is crucial to note that the software does not propose trajectories that lead to collisions or otherwise fail to achieve the desired outcome. The majority of the developed methods can be

classified into two categories: heuristic search algorithms and numerical optimization approaches. The initial method prioritizes computational concerns and real-time control, whereas numerical optimization emphasizes dynamic behavior and trajectory performance [10, 21].

This method avoids the stop-and-go driving behavior that appears on the last two teleoperation modes. It is possible because the machine provides the operator with different path options while driving, so the effect is avoided. Additionally, this method aims to present a limited yet sufficient number of path options to avoid overwhelming the driver and simplifying decision-making processes [21].

2.2 Situational awareness

Driving is a difficult task that requires a high level of focus and attention. While driving, lots of stimuli are perceived, and the driver must be able to avoid them in order to drive safely. The act of texting while driving, whether it is sending or receiving of messages, can result in a potential threat on the road. These threats are closely associated with a reduction in the driver's perception of their surroundings [22].

Endsley defines Situational Awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" on page 97 in [23]. Many things can affect the level of SA of a driver. For example, workload and stress can negatively affect conduction, but alternatively, training and experience can increase the level of focus and attention [23].

Fisher and Strayer [24] reached a model of the relation between SA and crash risk (Figure 6. Relationship between SPIDER processes, situational awareness and crash risk). This model is called SPIDER, and it includes Scanning, Predicting, Identifying, Deciding, and Executing appropriate Responses. In the same paper, they made a study where this relation was statistically analyzed. Even though many simplifications were done, they concluded that the probability of completing an action depends on the prior state's SA. They demonstrate a large relative crash risk increase with the interference of a secondary task in any SPIDER event.

Figure 6. Relationship between SPIDER processes, situational awareness and crash risk according to [24].

Endsley [23] developed the Situation Awareness Global Assessment Technique (SAGAT) to evaluate SA. In a proposed example study, the SA of multiple pilots was evaluated. SAGAT

consisted in stopping in the middle of a simulation and asking SA questions. The questions depend on the environment surrounding, as more elements in the surroundings, more challenging the questions are. Various stops were done, and random questions were selected to allow statistical validity and consistency. SAGAT scores from multiple pilots are stratified into three types: immediate, intermediate, and long-range SA. The only inconvenience this method presented is the constant interruptions of the simulation. However, it could not be done during the simulation because it would interfere with environment perception.

Uhrmeister conducted a study for automated driving using the SAGAT. For different scenarios, multiples questions were formulated. Such as, Information of other vehicles when the simulation stopped, the use of lighting by other vehicles or questions about participant opinion in accident avoidance [25].

SA as an overall depends on the awareness of multiple different objects and elements in the surrounding of the vehicle. Describing these elements is a complex task. Classifying and organizing the environmental factors that influence SA is needed to make a full description of the AV surroundings.

Layer number	Layer description	Examples
Layer 1	Road network and traffic guidance objects	Roads, sidewalks, parking spaces
Layer 2	Roadside structures	Buildings, bridges, street illumination, publicity signs, nature
Layer 3	Temporary modifications of Layer1 and Layer 2	Roadwork signs, emergency warnings, covered markings, auxiliar roads
Layer 4	Dynamic objects	Other vehicles (non moving as well), pedestrians, animals, moving objects
Layer 5	Environmental conditions	Precipitation, illumination, wind, pavement condition due to weather
Layer 6	Digital information	Changing traffic lights, information panels, switchable signs

Table 1. Layer description of 6LM.

Scholtes [7] describes the 6-Layer Model (6LM) for environmental evaluation. This model aims to formulate an environmental description independent from the system; for that reason, it should not contain any goals, values, or norms. A description of the layers from the 6LM will be done. Table 1 generally describes the different layers and Figure 7 shows a real example layer representation.

Layer one specifies where and how the driver should drive in the street, which means that this layer includes everything necessary for that to happen, excluding exceptional scenarios mentioned in other layers. This layer also includes traffic lights, driving signals, and everything related to pavement conditions or materials. In general, layer one describes the base of the road environment and makes its own driving possible in a normal driving scenario [7].

In layer two, everything that surrounds the road is included. It contains every static object that can be found not on the road. Layer two is the same as layer one but with the difference that layer two includes everything static not on the road and layer one everything static in the road. This layer increases the complexity of the environment and for that reason is very important to define it [7].

The next layer is a layer on its own, but in reality, it is the evolution of layers one and two. It comprises the temporary modifications of these two last layers. Actually, layer three does not include any different types of objects or signals. A clarifying example of this layer will be a construction area; temporary signals might impede the normal circulation of vehicles or the reconstruction of a bridge, disrupting the usual environment [7].

In layer four, dynamic objects are included, this means that now, the environment includes other moving vehicles and sudden movements. In reality, layer four considers all the objects that potentially move but do not necessarily have to be in movement. Parked cars are, according to that description, included in this layer, as well as garbage containers. With this layer, every object that can appear in a driving environment is classified [7].

Layer five explains the environmental conditions that can occur in driving scenarios. It includes from weather to light conditions. This layer is crucial for the SA because it can generate risky driving conditions. Reflexes because wet pavement, fog or icy roads are mentioned in this section. Lighting conditions are critical in environmental perception, whether it is artificial or natural lightning. The lack of light or the excess of it is also part of this layer [7].

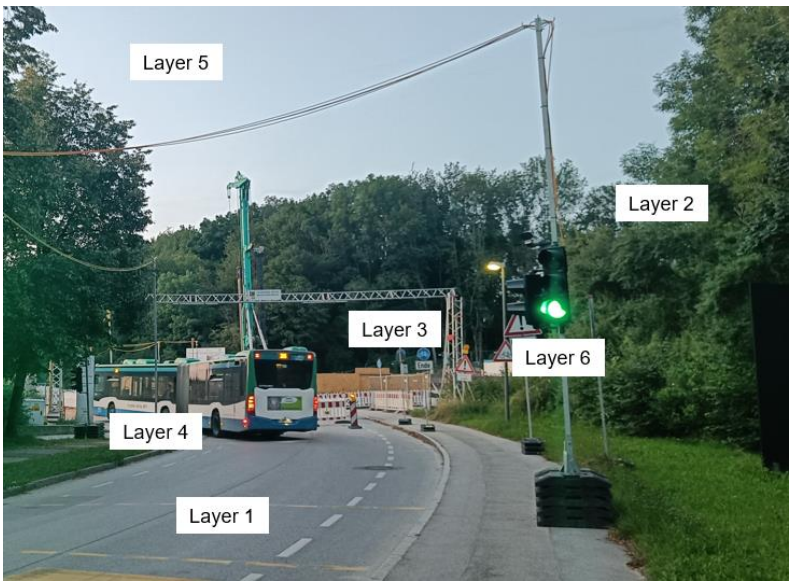

Figure 7. Representation of layers in a real scenario.

The last layer, layer six, focuses on communication or data exchange. This section also includes communication between different vehicles or infrastructure. Traffic management systems, including digital information signals, are mentioned in this section of the 6LM. Vehicle-to-Everything (V2X) also is part of this layer as it is a technology that enables communication of

information that might be occluded to the driver. For example, information about an incoming traffic jam or an intersection can be transmitted to drivers through V2X, and therefore, it will be represented in this layer [7].

For a future teleoperation control and SA perception, the operator will make use of a teleoperation system. The current teleoperation setup available in the Chair of Automotive Technology at the Technical University of Munich is a good example of a workplace and can be seen in Figure 8. The workplace has three monitors of 31.5" with 2K resolution and a framerate of 144 Hz. The setup is provided with pedals and steering wheel and the Vehicle Software runs on the EDGAR vehicle x86 PC, the videos are displayed on a projected sphere so that the teleoperator better perceives the surroundings of the vehicle. The computer used for the simulator has two graphic processors, 4090 RTX, an Intel i9 14900K CPU, and 192 GB of RAM to simulate all six cameras from the real vehicle at once.

Figure 8. Current teleoperation workplace at the Chair of Automotive Technology at Technical University of Munich.

2.3 Human vision and perception

Human vision is one of the most critical parts of our senses, and it involves different psychological and physiological components. That is why comprehending some aspects of it is useful to understand how the environment is perceived. Also, it is important to know the visual system's capabilities in order to simulate some of them during teleoperation control [26].

Peripheral vision and acuity

Peripheral vision is defined as the capability to see what someone is directly looking at and some part of the rest of the environment. When a gaze is made directly to an object, it can be clearly seen, the surroundings of the gaze can also be seen but not as clearly as the main object [27].

This encompasses the ability to see objects or detect movement outside the direct line of vision. This ability is essential in detecting movement as it makes threats and sudden movement detection easy to predict. However, peripheral vision implies lower acuity and resolution, so it is not as easy to detect fine details and specific shapes [27].

Color perception

Humans have two types of photoreceptors, this includes rods and cones. Our interest focuses on cones responsible for photopic vision at high light levels. There are three types of cones that can be classified according to the spatial sensitivity. L-cones, M-cones, and S-cones are named according to their sensitivity to long, medium, or short wavelengths. The cones have different peak sensitivities that coincide approximately with Red Blue Green (RGB) wavelengths. For that reason, monitor technology uses this same RGB combination to achieve a wide color display [28].

Foster [29] describes color constancy as the effect where the perceived or apparent color of a surface does not change even the intensity or spectral composition of the illumination. This effect allows humans to achieve a good perception of surfaces with independence of illumination variations, leading to better interaction with the real world.

Depth perception

Depth perception is crucial in relation to SA as it informs the observer about some key environmental information. Visual features that are perceived with both eyes can be used to determine the distance based on the shift in the position where the features are seen for each eye individually. This concept is also applied in the stereo vision using two cameras. But also monocular cues paired with the human past experience about certain object sizes or the area of overlap between different objects can be used to extract depth information with only one eye [30].

Thompson [31] analyzed different moving objects in a survey. For this purpose, Motion-in-Depth (MID) is required. The experiment concluded that binocular MID varies across eccentricity-matched locations, and monocular MID showed different results depending on the eye. Additionally, the study determined that sensitivity in experimental conditions is relatively poor compared to experiences in the world; this indicates that sensory signals that contribute to natural perception are not necessarily used in an experiment.

Light adaptation and contrast sensitivity

The visual system is able to adapt to a huge range of light intensities. This adaptation allows for the discrimination of luminance variation at every level. This is important to be able to achieve a good environment perception in every condition. Pupillary aperture, photoreceptors, and neural adaptation are responsible for the light adaptation process [28].

Contrast sensibility is the effect that occurs when the visual system acknowledges the difference in luminance between objects so they can be distinguished. For example, imagine an object in a uniform background; contrast sensitivity is the relative difference in luminance between the object and the background [32].

2.4 Image quality metrics

Image quality assessment has become a significant technological activity, and a lot of research is being done. The most important subjective quality metric is the Mean Opinion Score (MOS).

It is a subjective metric because it relies on human assessment, and apart from video, audio, and many different multimedia content can be evaluated [33].

VQMs are algorithms designed to evaluate the quality of a video. The goal is to predict the MOS of viewers. It is very useful when evaluating some video data that has been compressed or modified. With these metrics, loss of quality during any process can be evaluated. Metrics can be classified by the way they work. One of those classifications is the reference they used to assess the video quality [34, 35]**:**

- Full reference. The metrics included in this section perform a frame-by-frame comparison between the video and a reference video. Therefore, they need the full original video, and they are suitable for quality analysis after compression and decompression or after video transmission. Full reference implies spatial and time alignment, which can be an issue [34, 35].
- No reference. These metrics analyze only the test video with no need of any other input. This makes them more flexible and easier to use as they do not need spatial and temporal alignment [34, 35].
- Reduced reference. This section includes the metrics between full reference and no reference. These metrics only use some parts of the reference or test video and use them to execute the comparison. This method allows for avoiding assumptions that no reference metrics make while managing enough information [34, 35].

2.4.1 Peak Signal to Noise Ratio (PSNR)

This metric is a comparison between the highest power of a signal and the highest corrupting noise. PSNR is expressed in a logarithmic scale with a maximum value of 100 decibel (dB) and typical values in image processing from 30 dB to 40 dB. As said, the metric compares two signals; therefore, it is a full reference metric. The most common use is to measure the quality of images after compression coders. PSNR uses for the calculation the Mean Squared Error (MSE) [36, 37].

The following equations explain how PSNR works [36]:

$$MSE = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}[f(x,y) - g(x,y)]^2 \quad (1)$$

$$PSNR = 10 \cdot \log_{10}\left(\frac{I^2}{MSE}\right)\ [dB] \quad (2)$$

where I is the maximum pixel luminance value that is usually in 8-bit representation 255.

PSNR is very used due to the simplicity and velocity of execution; it is, at the same time, easy to understand. Nowadays, the development of technology is exposing some limitations of the metric. Some studies are starting to prove that PSNR has a low correlation with subjective opinion [38–40].

Even though PSNR is very used, and, for example, standardization of video codecs depends on this metric, the current state of the art continuously argues about the adequation of PSNR for video quality evaluation [41].

2.4.2 Structural Similarity Index Measure (SSIM)

SSIM is a metric that analyses structural similarities between two images. It is a perception-based metric that considers the change of structural information to be image degradation. This metric uses some important perception concepts like contrast and luminance masking. Similarly to SSIM, a metric is able to compare feature similarities; it is called the Feature Similarity Index Measure (FSIM). Both metrics are full reference metrics, and the index assessment can vary from zero to one, being one when the compared videos are the same [42, 43].

The mathematical explanation of the metric is as follows [44]:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (3)$$

$$C(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (4)$$

$$S(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (5)$$

$$SSIM = l(x,y)^{\alpha} \cdot C(x,y)^{\beta} \cdot S(x,y)^{\gamma} \quad (6)$$

When $\alpha = \beta = \gamma = 1$ and $C_1 = C_2$: (7)

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (8)$$

SSIM needs grayscale images as input, but Hassan and Bhagvati [45] developed a variation based on SSIM to evaluate color images directly. By doing this, more meaningful information is analyzed. They demonstrated that this metric outperformed SSIM and some derivatives from it.

2.4.3 Visual Information Fidelity (VIF)

This metric is a full reference metric that uses both information sources. One is the mutual information between the input and output of the visual system and the other is the mutual information between the input of distorted image and the output of the visual system. It is important to explain that this metric uses a human vision model to weigh and evaluate the perceptual importance of the image as it takes into consideration the ability of humans to perceive visual data. The mathematical description of the metric is quite difficult but can be seen in [46, 47]. The metric result is comprehended between zero and one, being one the best score possible.

Y. Han et al. [48] developed a metric based on VIF that demonstrated better predictive performance and had lower computational complexity. In this metric, VIF is used to get visual information from the sources to obtain 'effective visual information', and then to get the assessment result, all the information is put together in a fusion metric.

2.4.4 Detail Loss Metric (DLM)

DLM is a full reference metric that focuses on luminance only to evaluate video quality. This metric is based on separating the measurement of detail loss that affects content visibility and the impairment that may distract the viewer. The inputs of the metric are converted to a grayscale for the analysis. The metric consists of decoupling a restored image's additive impairments to apply human visual system processing later. This model uses only a low-level vision system model to implement two characteristics: contrast sensitivity and contrast masking. Then, a calculation and adaptative comparison of the two quality measures is executed [49, 50].

2.4.5 Video Multi-Method Assessment Fusion (VMAF)

With the objective of improving viewers experience, Netflix [50] developed VMAF. This metric uses multiple other metrics to predict subjective quality; it combines the strengths and weaknesses of some of the metrics used to provide a 0 to 100 result. At the moment this paper was published, VMAF used Visual Information Fidelity (VIF) and Detail Loss Metric (DLM) to assess quality. Simplifications of these metrics are used in VMAF; for example, for VIF, the loss of fidelity is represented as an elementary metric, and in DLM, it is also taken as an elementary metric. Netflix conducted a study in this same paper where VMAF outperformed other metrics excepting when the information source was live video streaming.

García [51] tried to analyze Web Real-Time Communication (WebRTC) with VMAF by conducting an experiment where different metrics were tested against MOS, and the results obtained were that VMAF and VIF achieved better correlation than SSIM and PSNR, including some derivatives. Rassool [52] did a similar experiment resulting in a 0.948 for VMAF score correlation with MOS proving the good performance of the metric.

Orduna [53] has also proven the practicality of VMAF by trying different applications for the metric. Originally, the metric was designed to work with Full High Definition (HD), but different studies have proved the capabilities of this metric to work with different types of content. This paper focused on 360 Virtual Reality (VR) content with no adaptation or training.

It is important to say that VMAF is currently an open source metric [54] where developers can improve and propose new code modifications.

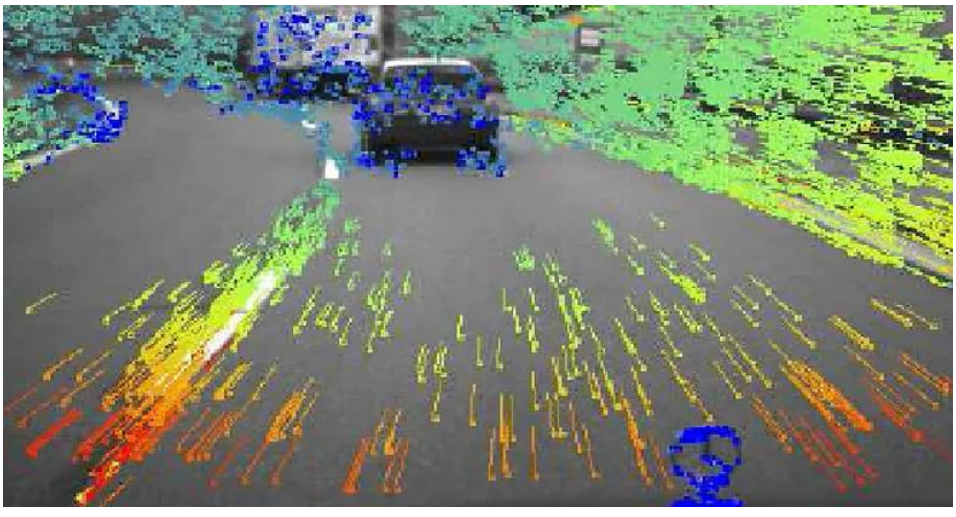

Figure 9. Optical flow movement detection on a moving vehicle.

2.4.6 Optical flow

Optical flow arises from the necessity to evaluate the motion of objects. Motion is crucial for the visual experience as it supports various visual tasks, including three-dimensional shapes and oculomotor control, perceptual organization, object recognition, and scene comprehension [55].

Specifically, optical flow is defined as the distribution of apparent velocities of objects, surfaces or edges in an image. Optical flow arises from the relative motion between objects and the observer [55–57]. Figure 9. Optical flow movement detection on a moving vehicle shows a real evaluation of motion using the optical flow concept.

When taking optical flow to a real test, some problems are presented. Objects can move in three dimensions, but these images are typically displayed in a two-dimensional display. This means that the velocity of a moving object perceived on a screen is not necessarily the same as the real velocity due to this problem. Another problem with the relation between motion and image sequence is that some movements can be hidden from optical flow algorithms. For example, in a sphere rotating by its diameter line, optical flow cannot detect any apparent movement due to geometrical characteristics and because anything is changing in the image source, assuming there are no shade changes in the video [58].

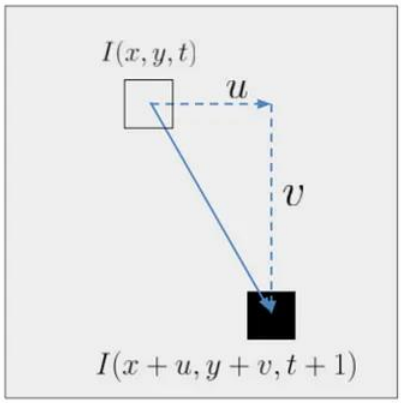

Figure 10. Pixel position change according to [59].

The mathematical description of optical flow can be tough and needs some assumptions. Figure 10. Pixel position change shows the concept of a moving pixel from a point to another in a specified time. It is easy to calculate the velocity with this information, but the main problem is detecting the pixels' correspondence. In this model, the motion is assumed to be small, so the same pixel must be in the vicinity of the first one. Also, the appearance of the pixel is assumed to not change in that small period of time. An approach to this is done [59, 60]:

$$I(x,y,t) = I(x+u,y+v,t+1) \quad (9)$$

$$I(x+u,y+v) \approx I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v \quad (10)$$

The equality (9) is known as the brightness constancy constraint. If (9) and (10) are combined, the following is obtained:

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v = 0 \quad (11)$$

$$I_t + I_x u + I_y v = 0 \quad (12)$$

This result explains that the spatial motion of the camera movement explains any appearance change of a pixel.

The mathematical explanation of optical flow extends much further, but usually, when the classical method is used, three different data and spatial penalty functions need to be implemented. These penalty functions are the quadratic Horn and Schunck (HS), equation 13, the Charbonnier equation 14, and the Lorentzian equation 15 [61].

$$\rho(x) = x^2 \quad (13)$$

$$\rho(x) = \sqrt{x^2 + \epsilon^2} \quad (14)$$

$$\rho(x) = \log\left(1 + \frac{x^2}{2\sigma^2}\right) \quad (15)$$

Sun [62] evaluated these three penalty functions and many more variations on the Middlebury optical flow benchmark. The results in Table 2. Average rank and EPE on the Middlebury test set with the different penalty functions [62] show the average ranking and end-point error (EPE) on the Middlebury test set. The Charbonnier showed good results despite being the simplest, and Lorentzian and HS performed well.

Penalty function	Average Rank	Average EPE
Charbonnier	34.8	0.408
HS	49.0	0.501
Lorentzian	42.7	0.530

Table 2. Average rank and EPE on the Middlebury test set with the different penalty functions [62].

2.5 Situational awareness evaluation

SA estimation is not a simple task, as it is known to represent perception and comprehension of the environment. Therefore, subjective opinions usually estimate SA but a method to estimate SA objectively is needed and that is where VQMs take action.

Hayashi et al. [63] intended to evaluate SA with a standard glance model. The work focused on unscheduled takeover situations that they considered the most dangerous because the driver's attention is not always focused on related driving tasks. The created model analyzed how and where a driver looks at while driving. The results concluded that this model predicted the SA of the drivers well as most subjects focus on the right parts of the environment and parts of the car when executing a maneuver.

In [64] SA is described as a tool to dynamic decision making as can be seen in Figure 11. Situational awareness and dynamic decision-making. Munir et al. described how to evaluate SA for better performance and the challenges that it represents. The paper is oriented toward the

military and Air Force perspectives but can be applied to every field. Many SA measurement techniques are proposed in this paper to achieve an SA score, and many of them utilize different types of probes. These measurement methods include self-rating techniques, observer-rating techniques, evaluation techniques during the activity, or performance-based techniques. The results showed that the metrics focused on some area of SA rather than in the totality for the assessment. However, the results obtained from the work recognize some techniques and technologies for improvement in SA.

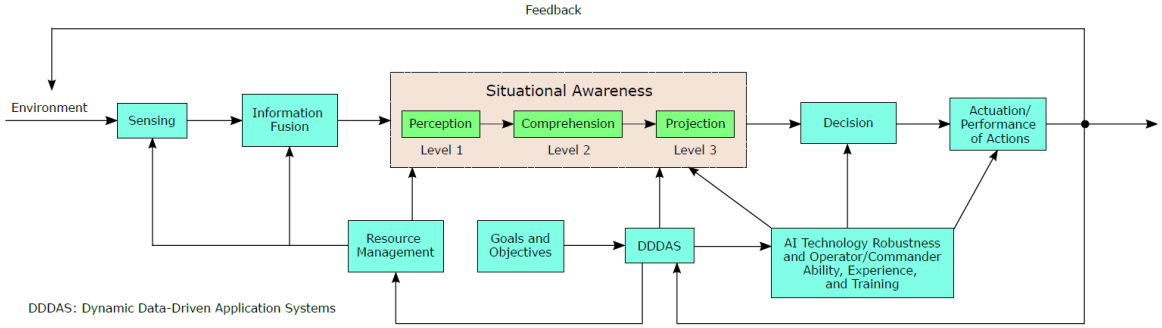

Figure 11. Situational awareness and dynamic decision-making.

Engelke et al. [65] propose two models to complement different VQMs because most of them do not take into account the spatial regions of videos where saliency variations occur and the effect on the viewer's attention. These models showed improvement in quality prediction compared to VQMs, which had none of these models implemented. This study allows for future approaches in spatial and temporal visual degradation comprehension to evaluate saliency information.

When evaluating SA, risk is sometimes forgotten in the different models, and it is actually one of the main objectives of the application of good SA knowledge. BowTie is a software that enables risk assessment at a tactical and strategic level. This tool enables teleoperators to comprehend the underlying logic of the environment and enhances the SA of the individuals for example to, operators that manage extensive streams of events [66].

The quality assessment in teleoperated driving is a challenging task, and the criteria for determining the minimum acceptable quality for remote driving still need to be determined. A study was conducted to determine the requisite quality for teleoperated driving. It employed an online survey and a variety of video quality metrics. The results demonstrated that video quality is dependent on environmental conditions, such as poor lighting or rainy weather. Conversely, the study concluded that a bitrate ranging from 299.20 Kbit/s to 831.92 Kbit/s is sufficient for a single camera under typical conditions [67].

2.5.1 Correlation between SA and VQMs

In [41] the results obtained showed the correlation of PSNR with subjective opinion. The conclusion was that the metric should not be used for direct comparison to subjective opinion as it can vary from linear to non-linear content. PSNR is a specific metric that should only be used when different optimizations are done and for specific use. Table 3. Correlation between PSNR and subjective quality shows the performance of the mentioned metric compared with subjective opinion with different video contents and when all content is jointly evaluated.

Content	SRC1	SRC2	SRC3	SRC4	SRC5	SRC6
Correlation	0.98	0.99	0.99	0.98	0.99	0.98

Content	SRC7	SRC8	SRC9	SRC10	ALL
Correlation	0.99	0.98	0.99	0.98	0.71

Table 3. Correlation between PSNR and subjective quality.

García et al. [51] showed the correlation of the principal VQMs when compared to MOS using the Pearson correlation. In this study, every metric obtained good results, but according to the results in Table 4. Pearson correlation between MOS and objective metrics, the better predictive metrics were VMAF and VIF.

	VMAF	VIFp	SSIM	MS-SSIM	PSNR	PSNR-HVS	PSNR-HVS-M
MOS	0.915	0.938	0.885	0.809	0.878	0.879	0.871

Table 4. Pearson correlation between MOS and objective metrics.

As a summary of the state of the art, many papers conclude that PSNR is not a very adaptative metric and cannot be used with every video content [41]. Also, VMAF is considered as a very adaptative metric for all kinds of content and has proven to perform well [53, 68]. On the other hand, VIF has proven to be a good metric, better than many traditional human vision system metrics or even SSIM [46, 47].

Menon et al. [69] analyzed the Pearson correlation between different metrics for a thousand video sequences. The results in **¡Error! No se encuentra el origen de la referencia.** show the positive similarities in quality prediction between VMAF and SSIM.

Content	PSNR	SSIM	VMAF
PSNR	1.00	0.70	0.83
SSIM	0.70	1.00	0.88
VMAF	0.83	0.88	1.00

Table 5. Pearson correlation between VQMs.

2.6 Research gap

Multiple scientific publications emphasize the fact that the objective evaluation of SA is a significant development. At the same time, the state of the art also agrees that teleoperation SA is critical [5, 23, 24].

VQMs are well consolidated, and new variations and improvements of previous metrics are constantly being developed. A profound evaluation of individual metric performances has been conducted in the literature among different video contents [34, 37, 42]. A few studies that simulate real-world driving conditions and study the robustness of VQMs in these conditions exist [1, 67]. These metrics are well tested and usually are compared to each other with subjective studies to verify quality assessment. This verifies the continuous work to obtain a metric that best evaluates reality as humans can.

However, there needs to be more focused research on how VQMs can relate with subjective opinion to obtain a SA estimation, and even less existing work is focused on teleoperation environments. Actually, existing studies do not sufficiently address the requirements and challenges of teleoperation scenarios, such as latency, real-time processing, or data loss during transmission. By applying VQMs on selected driving scenes and at the same time assessing the SA via state of the art study methods, a deeper correlation between the individual VQM score and the teleoperator SA is investigated. The potential results serve as a first step in closing the research gap.

3 Method

This chapter explains how the work has been developed, starting with the proposition of three hypotheses that will be answered in the results chapter. Following, the video source selection and the content of it, taking into account multiple environment options, is presented. Further, the objective quality evaluation is explained, discussing which metrics were implemented and how they were achieved. In the last part, the subjective evaluation through a survey is detailed, and the survey execution and the objectives of it are included.

3.1 Hypothesis

To assess the relationship between video quality and situational awareness, three hypotheses are formulated. The first hypothesis comprehends the quality evaluation change depending on the evolving complexity of the videos measured by the involvement of the different layers that describe SA. The objective is to evaluate if the constant video quality is less suited for the scene when the complexity is increased.

In order to respond to the second hypothesis, a layer classification needs to be done, where the apparition of different layers determines a video complexity. The goal of the hypotheses is to answer if the VQM show different quality assessment depending on the layers involved.

The last hypothesis is to assess VQM importance and performance. This hypothesis aims to determine whether VQMs show similar results in assessing quality as human do. Ranking the sample videos based of objective quality evaluation and the individual metrics is part of this task, to show statistically how similar to human opinion and reliable the metrics are.

Figure 12. Front center camera of the vehicle in CARLA simulator.

3.2 Sample videos selection

For this thesis, multiple videos were filmed to evaluate the correlation of SA and objective quality. The videos are planned to describe urban driving situations. They must be able to resemble real situations and scenarios and for doing that, the current literature is of help. As explained in State of the art, with the help of some scientific papers [7] SA was fully covered in the sample content.

Some of the sample videos were filmed using the open-source simulator CARLA with the EDGAR vehicle on it. On the other hand, some of them were filmed with the real EDGAR vehicle in the streets of Munich. The videos are always filmed with the front camera of the vehicle, both in the simulator and in reality. Also, no sound is recorded. This is showed in Figure 12 and Figure 13. Among all the video samples, the selection is primarily decided for the content and coverage of the different layers to represent SA. On the other hand, the second criterion is the framerate, aiming to achieve an approximate average of 20 frames per second. Resolution is also accounted for, and a minimum of 720p is the objective.

CARLA simulator is a powerful tool that provides images of different characteristics. The sample videos were filmed with a framework based on the Carla Python API, and then the ROS2 interface of EDGAR was simulated. This includes ROS2 messages for the vehicle commands that are subscribed and executed in the simulator. Multiple functions for this purpose can be found in the following repository [70]. Then, the sensor topics are published by subscribing to the camera sensors and transforming them into messages to the respective topics. Further, the EDGAR data is published in different messages that include gear selection, velocity, indicators, honk and more elements.

The videos filmed with the research vehicle are filmed and then compressed. The camera provides a video resolution of 1920 x 1200 therefore it meets the minimum requirement. As later indicated in Table 6. Description of source video material, the videos on average have a framerate radio between 18.4 to 19 frames per second. Before the compression, the videos were anonymized with a Meta tool [71] that blurs out license plates and faces. Then, a compression at 5000 Kbit is performed.

Figure 13. Front center camera of the research vehicle EDGAR.

The filmed videos last approximately from 8 to 12 seconds. The estimated duration is due in order to isolate a determinate situation and not a full driving experience. At the same time, the duration of the videos is conditioned to the subjective experiments. Videos should be long

enough to evaluate determinate characteristics but should not take too long for the subject to lose concentration or focus. For better resemblance with a real teleoperation situation, the chosen videos are all from the research vehicle EDGAR. However, some videos from the vehicle were not good enough to participate in the analysis. For example, videos inside a tunnel had very low framerate, and therefore, the quality was not good enough. That created some difficulties on the analysis of Layer 5.

The content of the videos has been chosen based on the 6-layer model environment description explained in 2.2. Therefore, the videos include some of the different layers already mentioned. A brief description of the video samples is provided in Table 6.

Video sample	Source	Frame rate average	Description
Video 1	EDGAR	18.7226	Straight line driving with few side vehicles or objects; medium speed with detention upcoming
Video 2	EDGAR	18.5292	Straight line driving with multiple vehicles to the side; heavy speed in a highway
Video 3	EDGAR	18.4779	Upcoming detection with multiple vehicles and emergency vehicle at high speed; medium speed
Video 4	EDGAR	18.7761	Under construction area with vehicles in both directions; light speed with light steering
Video 5	EDGAR	18.8615	Interurban road leading to a tunnel with low lighting; High speed
Video 6	EDGAR	18.8305	Intersection with high number of vehicles and pedestrians; low speed with braking

Table 6. Description of source video material.

Video 1

The first video is a low complexity sample and represents an easy situation with no evident distractions. This video aims to be as an example of the methodology for the subjective survey. The video consists of moderated straight speed with no side vehicles or objects, just a normal driving situation with low traffic intensity in the street surrounded by many buildings.

In the video, Layer 1, Layer 2 and Layer 4 are included as it is the only and simplest way to describe an environment with other vehicles and pedestrians on it.

Video 2

Video 2 represents another typical situation on an interurban environment. It has been filmed in a highway. The vehicle has a high speed according to the road requirements. The video aims to evaluate heavy speed perception as well as side information perception. At the end of the video, the vehicle approximates to a bridge and that alters lighting perceived by the camera. Throughout the entire video, multiple vehicles and road signs appear.

With this video Layer 1, Layer 2, Layer 4 and Layer 5 are included. It is crucial to evaluate how dynamic objects affect perception and SA.

Video 3

In the third video, a road leading to an intersection is presented. The vehicle is travelling at a medium speed. Multiple vehicles appear, and they are closely situated to the vehicle, affecting safety perception and multiple aspects of SA. At the same time, an emergency vehicle with its corresponding blue lighting appears at high speed in the opposite direction making that an exciting event to evaluate.

The video covers Layer 1, Layer 2, Layer 4 and Layer 6. It is crucial to evaluate videos with high speed difference to analyze the distraction that creates. To be able to detect a unique vehicle like this in a teleoperation drive is essential to guarantee safety. Layer 6 in present in the exchange of digital information and communication with the lighting. This video increases in complexity as dynamic objects are part of it and vehicles are close to the observer.

Video 4

The fourth video depicts a rout traversing a road under construction. The vehicle is driving at light speed through different curves signaled with temporary modifications on the street. Detained and moving vehicles can be seen in the other direction as the observer moves. The lane is confusing as many symbols appear, and white and yellow lines appear.

Layer 3 is the principal focus of the video, as temporary objects on the road heavily modify the road and, therefore, its comprehension. At the same time, Layer 1, Layer 2, Layer 4 and Layer 6 are part of the video represented by the road, surrounding buildings, moving vehicles and traffic lights respectively.

Video 5

The fifth video runs on a high speed road where the vehicle is about to enter a tunnel. To be able to correctly look into a darker area and detect potential hazards is a mandatory capability for teleoperation. In the video, other vehicles appear at high speed.

With this content, Layer 1, Layer 2, Layer 4 and Layer 5 are covered. Light changes are crucial to evaluate and can occur suddenly. Even though light changes can be present in multiple and various conditions, this is a good example and subject of analysis.

Video 6

In video 6, a high quantity of moving objects is presented. This includes pedestrians, cyclists and motorized vehicles. The video sample shows the EDGAR vehicle approximating and detaining behind another vehicle. At the intersection, vehicles are observed moving in multiple directions, some even overtaking stationary vehicles. Additionally, cyclists are observed traversing the roadway from the sidewalk.

This video includes the highest quantity of objects with the capability of movement which implies a high definition of Layer 4. Layers 1 and 2 are also present in the description of non-moving objects in the environment. The last layer covered is Layer 6, represented in the traffic lights that communicate information about preference and direction.

3.3 Objective quality assessment

Objective quality assessment includes multiple algorithms and mathematical models to evaluate video quality without involving subjective human opinion. These methods are transformed into software and will be executed and tested. Results will be collected and later studied.

The quality assessment is crucial to achieve a reliable method on quality estimation for the SA on teleoperation driving scenarios. Therefore, the data collected must be consistent with the original sources and, of course, with subjective opinions.

In this chapter, the implementation of different algorithms will be discussed. PSNR, SSIM, VMAF, and optical flow implementation are treated in this chapter. These metrics help quantify video streams' fidelity and motion characteristics, which are essential for teleoperation applications.

3.3.1 VQMs election

For the implementation of the metrics a previous selection of which will be implemented needs to be done. The VQMs were initially planned to be implemented and executed through ROS2, which is a communication software between machines. The purpose of using this communication method is for future direct implementation with the research vehicle EDGAR.

In reality, ROS2 implementation revealed many technical problems and ended up not being a priority for the work. So, for the objective evaluation, some metrics were executed through ROS2, and others were locally executed.

In order to make a full study about quality assessment in the field of AD, the principal metrics needed to be tested. PSNR, SSIM, VMAF, and some extra metrics were elected to assess quality through their mathematical algorithms.

PSNR proved to not be a suitable metric in many environments and that has been proven in Section 2.4.1 but more research needs to be done with driving situational samples. For that reason, PSNR was one of the metrics to be evaluated to determine how well it performs and to study if it should be part of objective quality assessment in this field.

The next metric to be implemented and later evaluated is SSIM. This metric has so many different variations and improvements and thus it must be studied. In this research, just the normal metric is used to see the performance of SA quality assessment.

As literature has proven, VMAF is one if not the best metric for the majority of content. Especially, VMAF has proven to be adaptative to many different contents and even formats. Therefore, it is essential to see how well it performs with SA quality assessment.

Optical flow is the final part of objective evaluation. The motion is essential in this kind of content so optical flow evaluation is key to analyze part of the SA of the teleoperator.

3.3.2 ROS2 implementation

For the implementation of PSNR, ROS2 environment and Publisher-Subscriber communication is used. This communication enables the exchange of information between a publisher and a subscriber by the use of a topic, an accessible node where the publisher publishes information and the subscriber nodes subscribe to it. When the node publisher publishes a message, in this case, the video frame by frame, a topic is created with this content. Then the topic needs to be accessed using the topic name by the node subscriber. Topics can have multiple publishers and multiple subscribers. In the last step, the subscriber executes his program with the topic information. This communication is the actual communication between the EDGAR research vehicle being the publisher of the live video stream and the programmer being the subscriber, and vice versa. In the future, when live video quality assessment is available, EDGAR vehicle will be a publisher of the video stream, and the quality evaluation station will be the subscriber to evaluate the video source [72].

The process starts with the execution of mp4converternode.py that can be found in the repository associated to this work. This Python code creates two different publishers that create two different topics with the original video and the compressed video frame by frame. Further, the PSNR script is executed subscribing to both topics and delivering the result of the algorithm for every frame.

As stated in 2.4.1, PSNR is mathematically a straightforward algorithm and can be almost directly integrated into Python language programming. For this program, multiple external libraries need to be used. Numpy is used to calculate MSE of the different video frames, Math library is also needed to execute the logarithmic function as it is not included in base Python language. Finally, to allow ROS2 communication, the library rclpy [72] must be included.

The function was implemented from an open-source developer [73]. he Python script analyzes the video and evaluates it with a rating in the logarithmic scale from 0 to 100. It is essential to note that for the later evaluation and comparison between metrics.

3.3.3 FFmpeg implementation

FFmpeg is a multimedia framework that can interact with almost anything humans and machines have created. This includes decode, encode, transcode, mux, demux, stream, filter and play. This framework is characterized by supporting many different formats. The open-source framework aims to provide the best technical solution for developers and users of applications or programs [74].

For the analysis of some quality metrics such as SSIM and VMAF, FFmpeg is the most suitable framework for testing them easily. FFmpeg includes many libraries, Libvmaf being one of them. This library allows the user to calculate the VMAF score for original and compressed videos. As explained in in State of the art, the VMAF method includes many other metrics, and that is why

they can be tested by following the same instructions. These are the so-called additional features of the library. Therefore, a quality assessment of VMAF and SSIM will be performed using this method.

The metrics are tested locally on a computer. Using a Windows terminal, the following command results on the creation of a text file with the information of VMAF evaluation for each frame.

ffmpeg -i distorted.mp4 -i reference.mpg -lavfi libvmaf=log_path=output.xml -f null -

Additionally, Libvmaf has many more options. Correctly configuring the additional features, SSIM and PSNR evaluation can be easily achieved. For organizational and efficiency reasons, PSNR will also be executed via FFmpeg even being implemented in a ROS2 environment. The following commands result in two text files with PSNR and SSIM evaluation frame by frame.

ffmpeg -i distorted.mp4 -i reference.mp4 -lavfi " [0:v][1:v]ssim=stats_file=ssim.log; [0:v][1:v]psnr=stats_file=psnr.log" -f null -

3.3.4 Optical flow

Høirup Nielsen [75] developed an open-source Python software regarding optical flow. The code is able to calculate the average variation of the motion in the video frame by frame. It also provides image source of the quantity of motion in the video, that can be seen in Figure 14 and Figure 15. With this content, the observer can have an idea of the average motion of the video.

Figure 14. Frame of a studied video with the flow displayed with arrows meaning direction and quantity of movement.

The algorithm calculates movement vectors for each pixel and compares them with the same pixel in the next frame. The difference of these vectors is what we call motion. This is possible because of the OpenCV library included in the program. For motion assessment, the function calcOpticalFlowFarneback calculates the dense optical flow using the Farnebäck [76] algorithm, that can be found in Bigun and Gustavsson's book [77]. Besides the OpenCV library, Numpy is also needed to average the motion of each pixel to present a signal optical flow rating estimation per frame.

Figure 15. Frame of a studied video with the Hue Saturation Value (HSV) meaning quantity of motion and direction.

The metric was implemented locally in a Windows environment and tested with the source videos. The Python file is directly executed in a Windows terminal with the correct features and desired videos. OpenCv is responsible for obtaining and supplying the image to the rest of the code. However, the first and the more accessible videos don't show to much relevant information in the motion assessment as no sudden movements or high velocities appear. Very little variance can be found between the optical flow average values from the different pixels in the first videos. The data obtained by the metric is then saved in a different file for later statistical evaluation.

3.4 Subjective study

Subjective quality assessment involves the participation of humans to evaluate the quality of the video content. Unlike objective rating, this methodology relies on human personal perception and judgment. This is very useful to understand the end-user perception of video quality. This subchapter includes the design of the survey, the motivation of the questionnaire, and the execution methodology.

3.4.1 Survey design

The objective of the survey is to estimate a subjective evaluation for the different video samples, focusing on SA estimation.

The survey is aimed for a participant with some driving experience. Therefore, the age, whether the subject has a driver's license, how many years of experience the subject has, and the approximate weekly usage of a vehicle are important to the survey. Participants will mostly be in an age range from 18 to 35, preferably with driving experience. Experience with driving simulators could be useful.

A significant factor that must be considered is where the participants will participate in the survey. It is clear that quality and perception of a video is not the same in a HD computer display than in a smartphone. This information is essential to filter the results and to distinguish quality assessment from quality visualization limitations.

The survey will consist of a short introduction of the video, followed by the actual video, and after that, a specific questionnaire. Table 7 provides a summary of the questions common to every

video that will be asked following the viewing of the video. These questions are evaluated on a scale from 1 to 10, with 1 being the worst rate and 10 being the best rate. The ratings obtained in this section will be then compared and classified according to different criterions. The initial questions will be decisive to classify the information obtained in this section as maybe different driving experience result in different quality or conduction evaluation.

On the other side, some specific questions for each video will be formulated. The reason for this is the increasing complexity of the video content and the increasing distractions and scenarios. Some questions will be asked to evaluate some specific factors of SA and concrete events. This depends on the content of the videos already explained in 3.2. For example, the comprehension of static or dynamic objects is evaluated. Additionally, distance safety perception or perception of a special vehicle is asked.

Question number	Description
1	How clear was the overall scene in the video?
2	How would you rate the quality of the video stream?
3	Do you think the vehicle was well situated in the street?
4	Do you think the velocity was according to the situation?
5	How safe will you feel being in that vehicle?
6	How good would you rate the conduction of the video?
7	How good could you perceive the environment?
8	How good would you rate visibility?
9	Do you think the video provided sufficient information for a teleoperator to take control if needed?

Table 7. Subjective study questions common to all videos and rated from 1 to 10.

3.4.2 Subjective experiment execution

An online LimeSurvey survey was created to display the video selection and evaluation of quality assessment and SA comprehension. In the first part of the survey, participants are informed about the field of the study and some basic knowledge about the technology of teleoperated driving. Then, some demographic questions about participants were asked, as well as the device used for the survey execution.

The survey is structured in groups with questions. For each video, two groups were used to avoid the possibility of the participant going back and watching the video multiple times. In the first group, the video is displayed, and a verification question is asked. Then, on the second group, some questions are proposed. The evaluation from the main questions is a scale from 1 to 10, where 1 is the lowest rate and 10 is the highest. After this, participants are asked about the perceived velocity of the vehicle and then some specific questions for each video. These

questions are evaluated from 1 to 5 being 1 the lowest rating and 5 the highest or with yes/no type questions. At the end of each question group, a blank space is left for any extra comments.

For the execution of the study, the use of a computer or a portable computer is recommended. The participants were told to watch the videos in full-screen mode and to avoid watching the video multiple times. The survey was designed to be completed in 8 to 12 minutes and was not limited to any specific software or device. In order to guarantee variation, multiple participants close to the author and supervisor of this work were contacted and sent the invitation for the study.

For the analysis of the results, some previous determinations need to be done. The normality of the population of the survey is studied. For that purpose, the Shapiro-Wilk test [78, 79] is implemented. This method helps to determine and explain how the distribution of the participants is. Participants age, driving experience, weekly driving and the used device are studied with this test. For another part of the data analysis, the Friedman Test [80] will be used. This test is used to determine statistically significant difference between means from various groups which show the same subjects. Further, for the significant different groups, a correlation study is done. The Pearson-Correlation-Method will determine the correlation between metrics, layers and the survey questions. The last statistical method to be used will be the T-test. By this method we manage to study the significant difference between two groups, in this case the difference for the layers with low and high presence. We do this for each layer and for each metric to know the influence that the layer has on the evaluation of the metric.

In order to obtain results that are comparable with objective quality assessment, the questions will be compared independently and converted to a comparable scale. This includes the calculation of the averages from all participants that finished the survey, avoiding any corrupted data, the variance, the maximum, and the minimum. Based on these statistical key values, the correlation between the rating and a VQM is determined using the Pearson-Correlation-Method. The other questions will be studied independently to evaluate the comprehension of some concrete parts of the videos.

4 Results

This chapter summarizes the results obtained for numerous videos. In the first place, a great volume of videos is objectively evaluated, and the metrics results can be found. The videos can be differentiated between being CARLA simulator videos, real videos recorded in Munich or the six previously selected videos.

Secondly, the results from the subjective experiment execute with an online survey are presented. The distribution of the participants is studied, and the overall result are presented. The obtained results are used to validate the veracity of the proposed hypothesis. This is also included in the discussion of the results.

For the most parts, in the following sections, the results will be color-coded by using blue for PSNR, orange for SSIM and green for VMAF. The results will be presented in different formats, including boxplots, histograms, linear graphic plots and data tables.

4.1 Objective analysis

During the execution of this work, many video samples have been studied. In this section of the chapter, the objective analysis of some videos filmed in the simulator CARLA and a great quantity of videos filmed in the streets of Munich is done.

The videos are always analyzed with the three VQMs previously mentioned, and the processing of the results is executed in Excel. The graphic presentation and results are also obtained with this tool. The entire results are attached in the digital complements of this thesis, and usually the average, standard deviation, maximum and minimum are calculated. The color-coded results are constant during the entire documents.

For a correct objective analysis, it is important to notice the ranges of the metric evaluation. PSNR for example is a metric which ranges can vary between 0 and 100, but in this kind of content, as mentioned in the state of the art, a rating from 30 dB to 40 dB is a common result [36]. On the other hand, SSIM rating can vary from 0 to 1 with results achieving close to 1 value when quality is good. Finally, VMAF can rate video quality with a scale from 0 to 100. For these reasons, the metrics cannot be compared directly as the ratings scales are not equal or constant.

4.1.1 CARLA simulator videos

Six different scenarios were tested with the VQMs. The different samples differ on traffic intensity and speed. For example, scenarios 1, 2, 3 and 5 describe a vehicle driving at 20 kph starting with no traffic and ending with high density traffic respectively. On the other side, scenario 4

shows a construction area and scenario 6 depicts an urban high traffic situation with a speed of 25 kph.

The results show similar performance between the metrics comparing the different scenarios. Quality is also well evaluated by the metric, which does not mean to be enough or sufficient for teleoperation. This can be better appreciated in Figure 16.

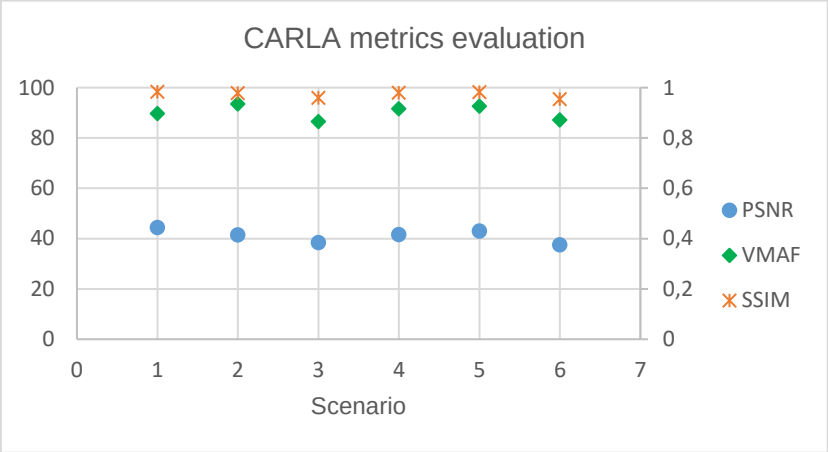

Figure 16. VQMs results for six different scenarios filmed in the simulator CARLA. PSNR and VMAF ranking from 0 to 100 and SSIM ranking from 0 to 1.

Even though these results, a ranking for the different scenarios with each metric was done. In this study, the metrics did not show the same performance indicating different orders. Actually, PSNR and SSIM showed the same order on quality assessment, but not VMAF. On the other hand, every metric agrees by their methodology that scenario 5 should be ranked as the second best video in terms of quality. The metrics also agree on ranking scenario 4 as the third best quality video. The rest of the scenarios are not in the same order among the different metric but in reality, the evaluation of the worst to videos is closely matched. Table 8 shows the full ranking of the videos in term of quality for the different scenarios.

The evolution of the quality during the video has been conducted for these samples. Figure 17 shows an example of this representation for scenario 3, where PSNR and VMAF refer to the scale from 0 to 100 and SSIM to the scale from 0 to 1.

Order	PSNR	SSIM	VMAF
First	Scenario 1	Scenario 1	Scenario 2
Second	Scenario 5	Scenario 5	Scenario 5
Third	Scenario 4	Scenario 4	Scenario 4
Fourth	Scenario 2	Scenario 2	Scenario 1
Fifth	Scenario 3	Scenario 3	Scenario 6
Sixth	Scenario 6	Scenario 6	Scenario 3

Table 8. Simulator CARLA videos order by quality assessed by three different metrics.

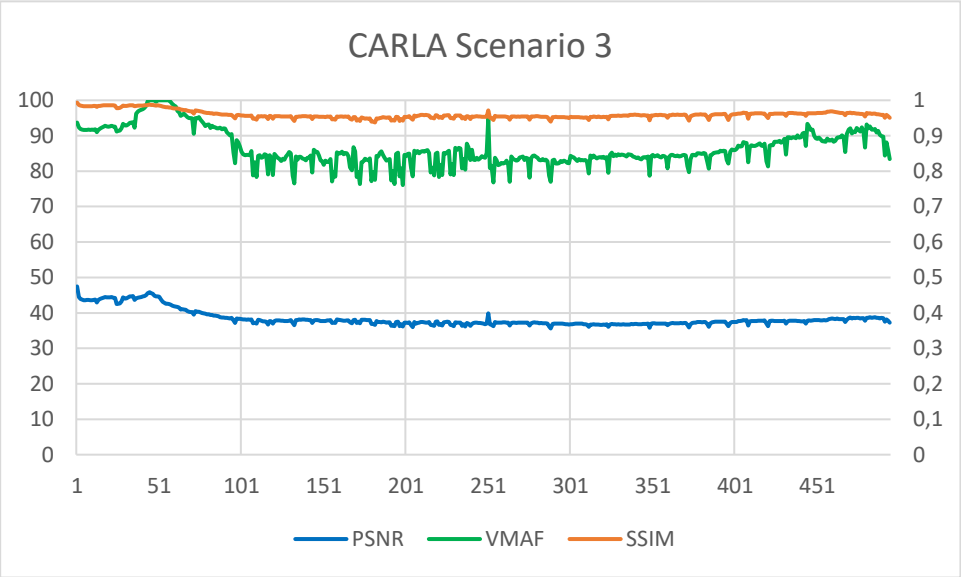

Figure 17. Evolution of the evaluation of three video quality metrics for scenario 3 of CARLA videos.

4.1.2 Munich recording videos

During the research vehicle EDGAR filming journey, many videos were filmed. This work includes 205 videos as an initial source for analyzing quality. These videos are differentiated by areas and streets, being the 205 videos organized in nine different categories. The different categories do not have the same number of videos and not all categories define and include at the same level of layers for SA assessment.

The videos are tested with the implemented metrics, achieving results about average, median, standard deviation, maximum, minimum and quartile information. This information is then used to show in general terms the performance of the metrics. The results considerate all the videos

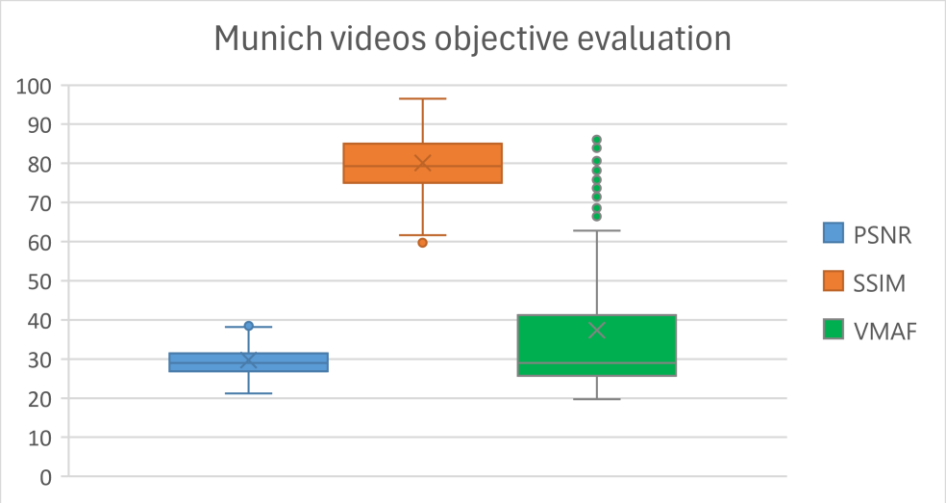

Figure 18. Video quality metrics for all the videos filmed in Munich. The scale ranges from 0 for a poor evaluation to 100 for good quality assessment.

at the same time so they should be accounted as that. These results are then an indicator for quality among big sources of video and therefore should not be interpreted for specific quality results. Figure 18 depicts the previous commented results. For this polt, SSIM was normalized by multiplying by a factor of one hundreg to display correctly with the other metrics.

For further results in following sections, the layers of each video were clasified in different categories. Table 9 shows the classification and the criteria followed. A part from the layers, the light burnouts were also accounted for. Overall this indicates what defines the situation on the video and helps to achieve a conclusion for hypothesis 2.

Layer	None	Low	High
Layer 1	The layer is not present in the video	One direction street	Multiple directions or an intersection
Layer 2	The layer is not present in the video	Little to no traffic infrastructure	More than two traffic signs and narrow surrounding
Layer 3	The layer is not present in the video	Minor elements (for example, cones)	Affects the ego lane and highly modifies the scene
Layer 4	The layer is not present in the video	Objects do not show imminent interaction with the vehicle	Objects might result in action of the ego vehicle
Layer 5	Evaluation id one depending on the weather, for example, sunny, cloudy, foggy…		
Layer 6	This layer is not present in any videos as exchange of information is hard to capture through a video.		
Burnout	There is no burnout in the video	-	A burnout appears during some of the duration of the video

Table 9. Layer rating explanation for SA assessment for Munich videos. Burnout of the image also included.

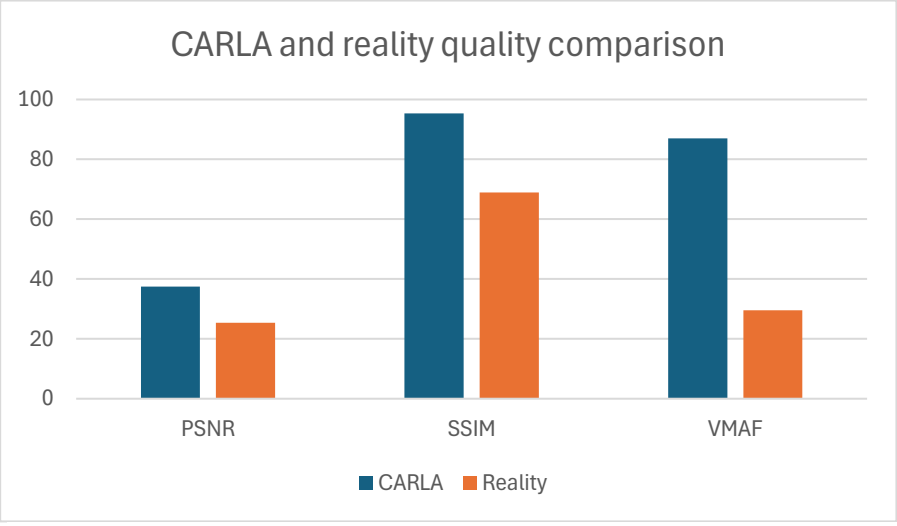

Figure 19. Comparison of two similar video content from CARLA simulator and reality. SSIM was normalized so it can be compared to the rest of the metrics.

Finally, a comparison between similar videos of the CARLA simulator and real videos is done. The videos recorded in the simulator are longer than the real ones. In the study case, the simulator video lasts one minute and twenty seconds and the real Munich recording lasts nine seconds. The velocity of the ego vehicle is also quite different, being slower in the simulator than in reality but overall, the videos present various environmental similarities. Both of the videos represent a moving vehicle crossing an intersection with other vehicles on it and then a straight driving through a wide avenue with two traffic directions. Figure 19 shows the comparison between both videos analyzed by the three metrics.

4.1.3 Video selection analysis

In this section, the selected videos in 3.2 are evaluated. The results obtained include metric evaluation frame per frame for each video, optical flow analysis and layer intensity assessment.

The videos that are used in this section were selected with different criteria and therefore show different performance. The evolution of the metrics is essential to study these differences in quality for each of the samples. Figure 20 to Figure 25 show the evolution of the PSNR, SSIM and VMAF during time for every video. PSNR and VMAF are ranged in a scale from 0 to 100 and SSIM in a scale from 0 to 1.

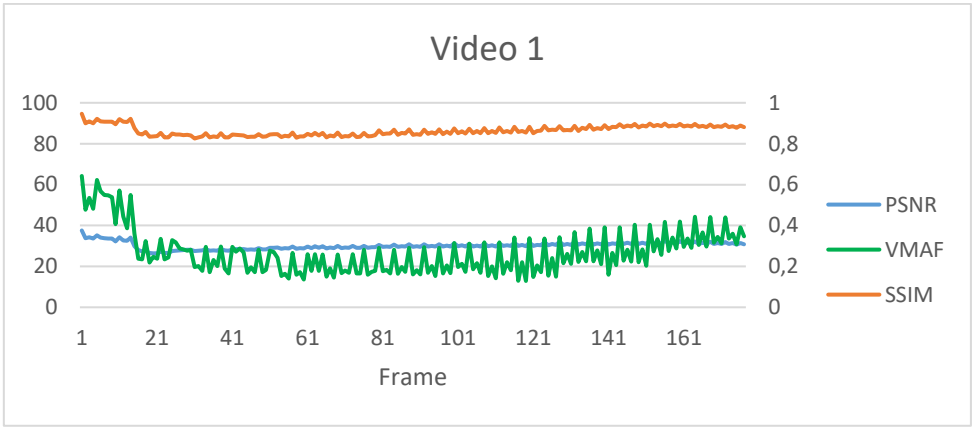

Figure 20. Evolution of the VQMs for video 1.

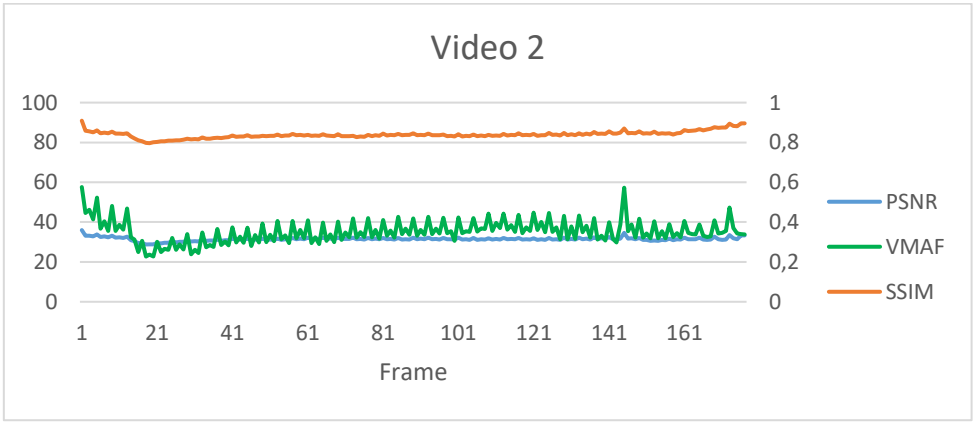

Figure 21. Evolution of the VQMs for video 2.

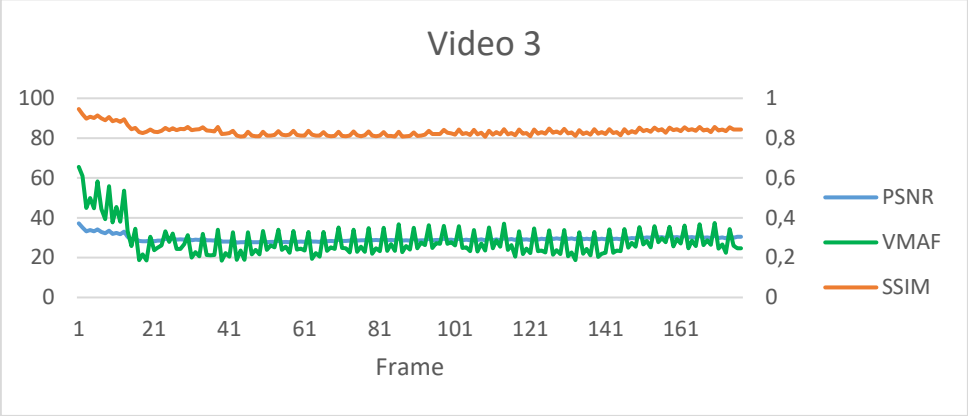

Figure 22. Evolution of the VQMs for video 3.

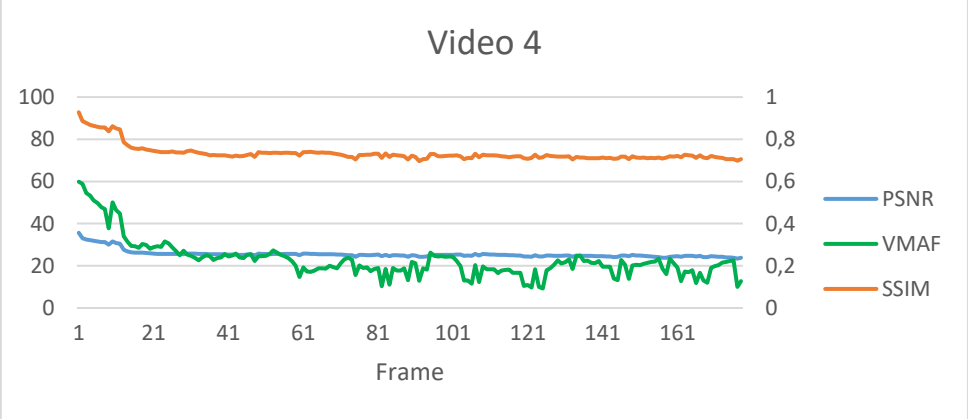

Figure 23. Evolution of the VQMs for video 4.

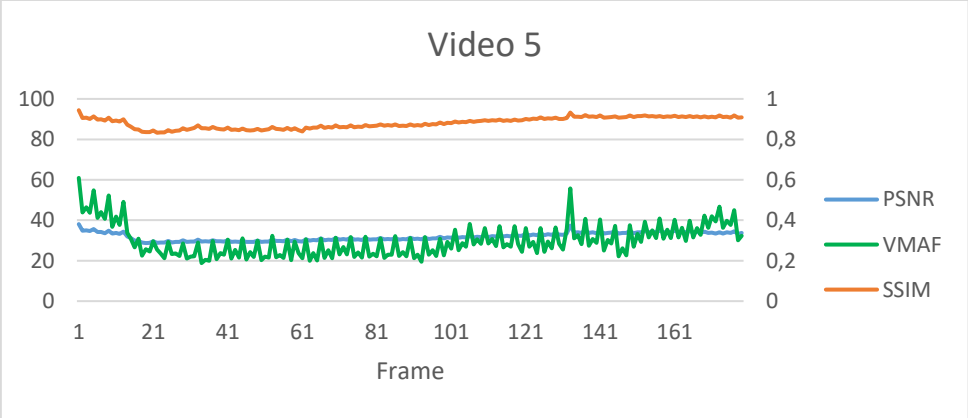

Figure 24. Evolution of the VQMs for video 5.

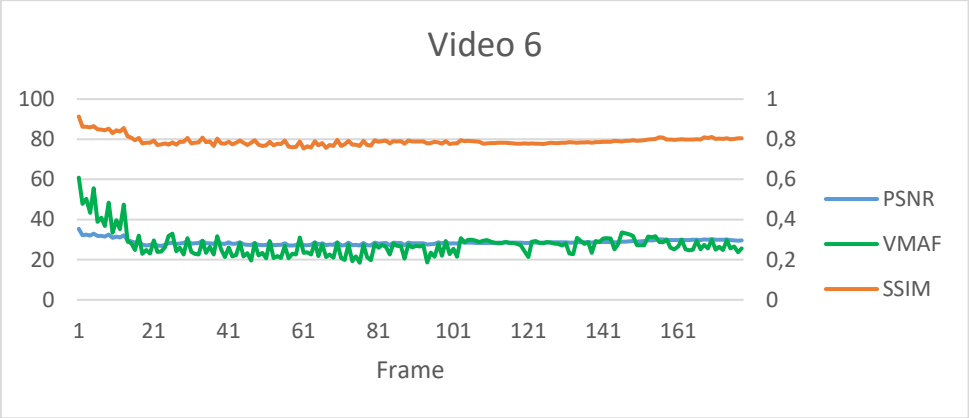

Figure 25. Evolution of the VQMs for video 6.

Optical flow is a big part when analyzing a video content. In this section, the analysis over time of optical flow is done. Figure 26 shows the evolution of the optical flow rating for video 4. The rating represents the average of the motion of each pixel in every frame of the video, this was further explained in 3.3.4. This is an example of the rest of the evaluations for the other videos.

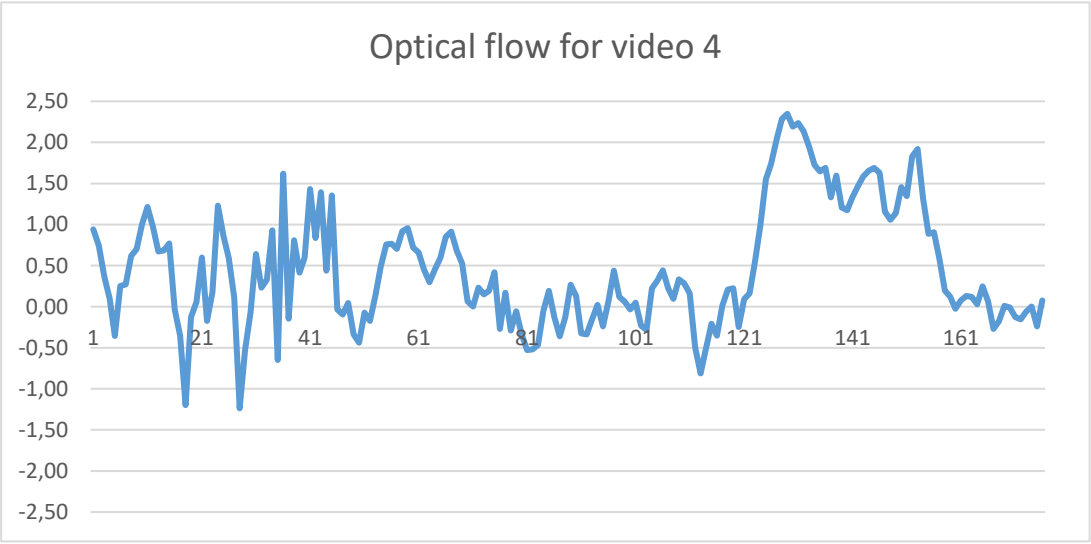

Figure 26. Optical flow evolution during time for video 4.

In this section, an analysis of the involvement of the layers is also done. This is useful for the discussion of hypothesis two in further sections. The assessment was carried based on the same criteria as the entire Munich sample videos explained in Table 9. In the same way, layer 6 is not present throughout the videos and no burnouts of the image are considered. Therefore, the rest of the layers are evaluated according to the mentioned table. Furthermore, a weighted average of the different metrics for each layer is done considering the value none equal to zero, low equal to one and high equal to two. Table 10 shows the averages for each layer.

Layer	PSNR	SSIM	VMAF
Layer 1	29.684	0.822	29.135
Layer 2	28.820	0.812	27.303
Layer 3	25.515	0.733	22.719
Layer 4	29.245	0.814	28.410
Layer 5	29.948	0.834	28.997
Layer 6	0	0	0

Table 10. Weighted average considering no existence, low or high presence of the layer for every video. Layer 6 is not present and therefore shows a zero value.

4.2 Survey results

The survey period covered 5 days starting on August 2nd of 2024. The survey took approximately from 10 to 15 minutes to complete. A total of 79 participants took part. Frome these participants, 70 (97.5%) completed at least one page of the survey and 52 (66%) fully complete the survey until last question. From these 52 final participants, 51 stated to have driving license. The participants chose their age range being the majority in a range from 15 to 30 years old (31 participants, 59,6%). The Shapiro-Wilk Expanded test demonstrated that the age distribution over the different age ranges differing 15 years was not normally distributed. The same happened with years of experience and weekly driving average, obtaining average results for driving experience of 17,5 years and for weekly driving of 3,2 days. The instructions of the study recommended the use of a computer or laptop to better display the contents of the survey, however, it was not a compulsory requirement. The device election was also not normally distributed being the smartphone the most used device with 24 participants using it. This information is further described in A General participant information.

During the survey, after the main questions, the estimated velocity was asked. For some videos, traffic signals indicated it, but the results show that the distribution of the estimated velocity for every video does not follow a normal distribution except for video 5. Figure 27 shows the normal distribution of the data for video 5. On the other hand, the average of the ratings indicates an adequate velocity for the majority of the videos. This will be further discussed in Discussion.

Some participants encountered some doubts about the videos and questions. In relation with the perception of the safety, it was not clear if the question meant safety while driving or being driven by the AV. On the videos, the vehicle was controlled by a driver but in a future implementation, AD technology will develop this task or teleoperation if needed. Some participants also encountered serious problems when big light changes appeared, for example on tunnels or approximating to a bridge. One of the participants suggested to increase the safety distance because at some moments, vehicles in front could not be seen. On the other hand, a few participants commented about the quality being too low for high speeds but at the same time other subjects were able to identify traffic signals or even exit indication. A great number of

participants noted that in the construction zone displayed on video 4, the quality was not good enough to perceive the situation and it was very difficult to see the yellow lane lines.

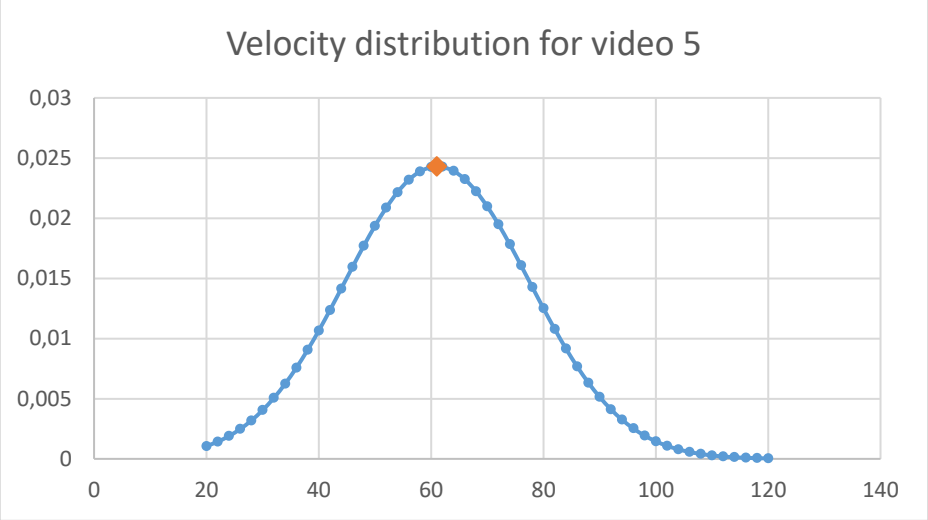

Figure 27. Normal velocity distribution assessed in the subjective opinion about video 5.

4.3 Hypothesis 1: adaptation of constant quality

To demonstrate the veracity of the hypothesis, the evaluation of the constant quality was done based on complexity evolution. The main data of the survey for every video was classified into nine topics including overall clearance, quality, situation on the street, velocity accordance, safety feeling, conduction rating, environmental perception, visibility and teleoperation control. These topics were rated on a scale from 1 to 10. After the data collection, results were organized depending on the topic, resulting then in a comparison for all the videos. The Friedman test was executed to determine the existence of significant difference between the videos. The results show that four out of nine questions had no significant difference between the rating of the six videos. The four topics that indicated no difference were the situation on the street, velocity accordance, environmental perception and teleoperation control. In further chapters, these results are discussed, and logical answers are proposed.

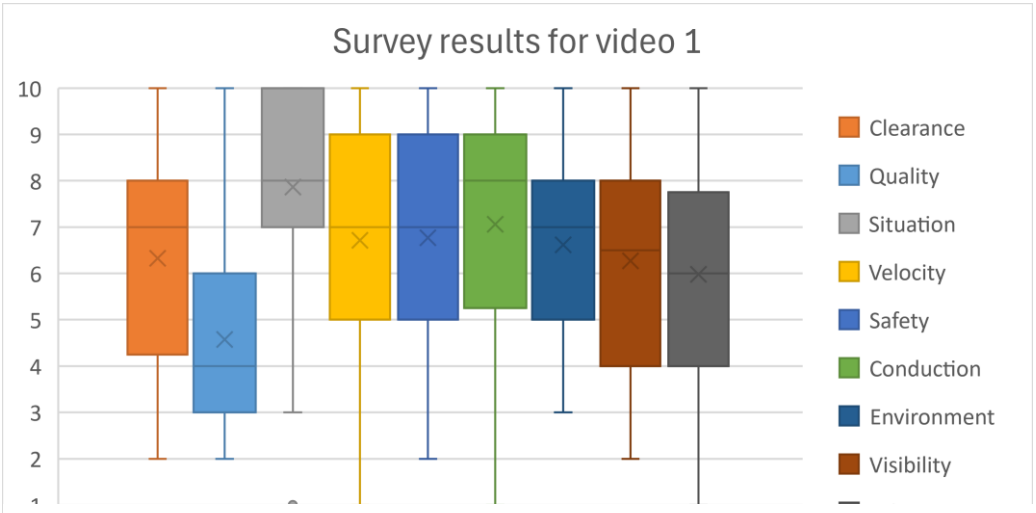

Figure 28. Statistical results for the nine topics in the subjective survey for video 1.

Figure 28 depicts as an example, the evaluation of the nine different topics for the video 1. This boxplot was also obtained for the rest of the videos.

4.4 Hypothesis 2: involvement of layers

Hypothesis two implies the demonstration of a correlation between layer involvement and video quality. The hypothesis states that when more layers are involved, the quality should be worse than when few layers are present. For this purpose, two similar studies were done. First the correlation between the layers and the metrics for the six video selection based on the Pearson-Correlation-Method is done. The second study consists of the analysis of the significant difference for the layers using a source of 110 real videos and then the Pearson correlation study.

In the study of the six selected videos, the Pearson method revealed very different results. Some layers are clearly correlated to the metric and other layers are clearly not. Table 11 shows the results of the method. When the value is higher than 0.5, it is considered to be correlated with the metric. A positive result indicates a direct correlation, and a negative result indicates an inverse correlation.

Layer	PSNR	SSIM	VMAF
Layer 1	0.186	-0.126	0.567
Layer 2	-0.617	-0.446	-0.522
Layer 3	-0.858	-0.822	-0.681
Layer 4	-0.505	-0.700	0.020
Layer 5	0.235	0.257	0.200
Layer 6	0	0	0

Table 11. Pearson correlation between layers and metrics for the six selected video samples.

When the subject of the analysis are 110 videos, results change. For the significance difference between the metrics, the Friedman method was substituted by the T-test, even though both aim to achieve the same results. The test shows significant difference between the low and high presence of the layers. So, for each quality metric, the layer involvement is evaluated, and quality assessments determines if the layer is or not significant. Table 12 depicts the results of this test. When the T-test value is below 0.05, it is considered that data has significant differences with a 95% confidence, on the other case, data is considered not to have significant differences. For the results of the layers that presented significant differences, the Pearson correlation was carried out. The method presents no correlation between the layers and the metrics. It is important to note that this method studies a linear correlation between the data.

Layer	PSNR	SSIM	VMAF
Layer 1	0.466	0.408	0.051
Layer 2	0.405	0.412	0.171
Layer 3	0.045	0.012	0.215
Layer 4	0.068	0.019	0.290
Layer 5	1.9E-21	0.010	7.4E-24
Burnout	0.037	0.034	0.385

Table 12. T-test results for low and high presence of layers for each metric with a source of 110 videos.

4.5 Hypothesis 3: video ranking comparison

Regarding this hypothesis, some of the data used for hypothesis one is also used here. In this case, the focus falls on the general analysis of quality. For that, the subjective and objective data for the six selected videos is compared. Figure 29 shows the subjective quality assessment for the safety feeling, which has a significant difference among the six videos. The topics of the survey were analyzed with the Friedman test and the significant differences were presented in hypothesis one.

Figure 29. Subjective ratings about safety feeling by 52 participants in online survey.

Further, for the significant different topics, a correlation study between them and the metrics is done. The method used is again the Pearson-Correlation-Method. The correlation results for the significant questions can be seen in Table 13. The values show promising results that will be further discussed. Negative values indicate inverse correlation and positive values indicate direct correlation.

Question	PSNR	SSIM	VMAF
Overall clearance	0.754	0.848	0.875
Quality of the stream	0.560	0.808	0.731
Safety feeling	0.833	0.927	0.962
Conduction rating	0.793	0.887	0.920
Visibility rating	0.762	0.881	0.886

Table 13. Correlation between the metrics and the significant questions from the survey. Done with the Pearson-Correlation-Method.

5 Discussion

In this chapter, the results previously presented are discussed. First, the video election is explained, and its limitations are presented. Secondly, the VQMs results for simulation and real videos are compared and explained. This includes the analysis of the advantages or disadvantages the different frameworks. Also in this section, the six selected videos comparison between objective and subjective opinion is elaborated. This includes general ratings from the metrics and from the participants in the survey as well as a complexity study with the help of layer involvement. Finally, the demonstration of the hypothesis closes the chapter.

5.1 Video election

Initially, video election caused some inconveniences as real EDGAR videos and ROS2 framework complicated the selection process. For the development of the work, the more videos available the better. That is why videos from both the CARLA simulator and reality are included. The different format of the videos between the simulator and the EDGAR research vehicle presented problems when evaluating using VQMs. The videos were mostly transformed to mp4 format to be studied by the metrics. The video recordings conducted in Munich were subject to certain limitations. All videos were captured on the same day, resulting in no major differences in lighting conditions.

In addition, the survey further complicated the choice because only a limited number of videos could be shown. These videos had to be carefully selected to represent a driver's environment in different situations. The selection criteria of the survey videos is clearly marked by the subjectivity of the author and that is why the analysis of a larger video source is also carried out. Another limitation faced by the survey was the correct viewing of the videos. Although the participants were told to view the content in full screen mode, the type of device may not have permitted this. On the other hand, in the survey, the integrated player, Youtube, was in charge of displaying the videos and, as commented by one participant, the quality of these could be worse than the maximum possible by the player.

5.2 Objective and subjective evaluation

The video contents in the simulator and in reality were very different. That is why the simulator CARLA results show a very good quality estimation compared to the real videos. Because they are already a digital content, the video compression is smoother and therefore when comparing original videos with compressed ones we get excellent results. In addition, the quality ranking of

each metric is identical in PSNR and SSIM for the different scenarios and in some cases coincides with VMAF as well.

For the study of the real videos, the opportunity to have 205 videos is very helpful. With this analysis we can get a measure of how different metrics interpret quality for many different scenarios and situations. The results show not very high but sufficient values in most cases. For example, for PSNR, knowing that between 30 dB and 40 dB are good values for this type of content, the results are good. SSIM also shows good values on average. However, VMAF does not show very promising values as well as having a lot of dispersion in some data.

Results clearly indicate that the video quality is better in the simulator than in the real videos after compression. Figure 19 clearly shows this statement.

Further, the six videos chosen for their representation of SA were evaluated. The video quality evolution graphs for each metric show very stable values in most cases. However, all videos show an initial drop in quality for all VQMs. On the other hand, the optical flow of the sequences was evaluated as a fairly good parameter to identify large movements in the video. For example, in Figure 26 there are several peaks representing oncoming vehicles crossing the ego vehicle. In the most pronounced peak and contrasting with the original video, two bulky vehicles are clearly identified in the image. This analysis can be beneficial in identifying the number of moving objects in a video and how conspicuous they are to the driver.

For the same videos, a weighted average was obtained depending on the presence of the layers describing the environment. Despite having few videos for this analysis, the results show very similar values for each metric in each layer apart from layer three which obtains lower results than the rest. This indicates that with high presence of this layer, temporary objects such roadwork signs or work fences, the quality is worse than the rest.

The survey execution had some limitations. Among them is the number of participants, despite being an acceptable number, the more opinions, the better the general results. In addition, the age distribution was quite concentrated and that is why, together with the rest of the initial information of the participant, normal distributions of the information were not obtained. The positive and beneficial part for the experiment is that both drivers with great experience and drivers with few years of experience participated. Finally, the survey included a question about the perception of speed in each video. The results do not show a normal distribution in the speed data except for video 5. This is explained because this video takes place on a highway, and it is easier to estimate the speed. On the other hand, video 2 also takes place on a road of the same type with more vehicles and the same speed results are not obtained.

5.3 Hypothesis achievement

To prove the first hypothesis, the Friedman test determined that four out of nine questions did not show significant differences between the videos, which means that there is a general opinion on this aspect. This indicates that the increase in complexity in four out of nine videos is not reflected in the quality assessment. It can be due to multiple factors depending on the question, for example, situation on the street or velocity accordance. Regarding these questions, the perception of the vehicle driving correctly between the lines or driving with an according velocity is explained because of the driving rules the vehicle followed during the recording of the videos. The hypothesis statement then is false for these four questions of the survey. On the other hand, five questions presented significant differences between the videos, and thus it means that the

complexity changes on the content affects the view evaluation. That is easy to see on the safety feeling for example, where perception of security depends on multiple factors including the different scenarios or personal experiences and not entirely on video information.

Hypothesis two poses a serious limitation: subjectivity in assigning the intensity of the layers. In addition, there is also the limitation that this assignment must be done manually by a human and that reduces the possible volume of videos. In this work, 110 videos were used to study the influence of the layers with the evaluations of the video metrics. The T-test method provided the significant difference in the quality of each layer between its evaluation as low or high presence. This was done depending on the video metric. The results of this analysis were shown in Table 12. After this, the Pearson correlation was run and surprisingly it did not provide any relationship. However, it is important to know that the Pearson method measures a linear correlation, and this data may be correlated but not linearly.

Further, to demonstrate hypothesis three, the Pearson correlation was performed to the questions and metrics, obtaining positive and very high values for the questions without significant differences. Therefore, this means that for these questions, the measurement made by the metrics is very good compared to human opinion. Overall, VMAF presented very good results, including high positive correlations with question assessment. However, PSNR performed even better, being the metric that best correlates with the questions in four of the five topics. In addition to this analysis, the correlation was evaluated using the same method with the mean order of all participants. In this case, VMAF stands out in four out of five topics and PSNR in the remaining one. Taking into account all these results, the hypothesis concludes that PSNR and VMAF are quite appropriate for analyzing teleoperation content and that SSIM does not show good results.

6 Conclusion

This chapter provides an overview of the current state of the art in teleoperation and video metrics. In addition, the study carried out and the hypotheses raised are presented. Some of the most important results are then highlighted. Finally, windows for improvement found during the development of the thesis are identified with the aim of motivating future work and the improvement of video quality in teleoperated driving.

6.1 Summary

Teleoperation is a relatively modern technology and presents great opportunities for improving autonomous driving. This thesis aims to help determine the video quality perceived by humans using video metrics. For this purpose, from among those that the state of the art has developed, we take advantage of PSNR, SSIM and VMAF. These metrics have been shown in previous studies to be functional and useful and have therefore been chosen to analyze this content.

To achieve the objective of determining the quality of an image perceived by a human, an online experiment with 52 final participants was conducted. The subjective study shared some similarities with SAGAT. This experiment included the viewing of six videos pre-selected to describe a driver's environment. In addition to the viewing, participants were asked for ratings for different topics about the video including the overall clearance, quality, situation on the street, velocity accordance, safety feeling, conduction rating, environmental perception, visibility and teleoperation control.

As part of the quality assessment, 215 real videos recorded in the city of Munich were available as well as numerous videos recorded in the CARLA simulator. These videos were recorded in different parts of the city in order to obtain a wide variety of scenarios and situations. The content generally consisted of ten-second videos for the real and some simulator content, although other simulator videos lasted between one and two minutes. Subsequently, these videos are compressed to 5000 Kbit simulating the compression necessary for teleoperated driving. The results of the simulator videos showed that the compression slightly affects these contents, presenting very good quality for all metrics compared to real contents.

During the development of the work, three hypotheses are raised. The first aims to evaluate the increase in the complexity of the videos according to the subjective opinion of the survey participants. The second aims to determine the significant differences between the video metrics depending on the incorporation of the different layers that constitute a complete description of the environment. Finally, the third hypothesis studies the general correlation that exists between the video metrics and human opinion to determine which ones are similar to subjective opinion and therefore are more useful in the sector.

The results obtained previously demonstrate for the first hypothesis that some aspects of driving are subjective to the quality of the image provided, but other aspects are independent of this because they are influenced by external factors. Therefore, only for some of the topics covered in the survey, the increase in complexity of the situation affects the assessment of these.

The results of the second hypothesis obtained by means of the T-test indicate that the most basic attributes that constitute a driver's environment, layers 1 and 2, are not significantly different in any of the metrics used. However, for the rest of the attributes that describe the environment, there is some variability depending on the metric. This means that for PSNR and SSIM, the layers mostly show significant differences, whereas in VMAF they do not. Exceptionally, layer 5, responsible for describing the lighting conditions, shows significant differences in all the metrics. Despite these results, none of the significantly different layers present a linear correlation with the quality evaluated by the metrics.

The last hypothesis shows that not all metrics have the same performance for analyzing content of this type. SSIM does not present promising results while PSNR and VMAF obtain very positive values. For the survey questions that showed significant differences, the Pearson correlation was performed. This method indicated that PSNR is the metric that most closely approximates the mean rating of the online survey, VMAF also obtained good values in this study. However, when the quality order of the six videos for these significant questions is analyzed, it is VMAF that comes closest followed by PSNR. It is therefore concluded that both metrics have a promising use and that the simplicity of PSNR provides performance advantages for the execution of analysis.

6.2 Outlook

The hypotheses raised have been satisfactorily resolved, although some details can be improved. During the writing of the thesis, some new questions were generated. These questions range from improving the subjective experiment to new tests with different content.

The online experiment carried out yielded good results, however, a presential experiment that better represents the situation of a teleoperator would be beneficial to improve the veracity of the data. This includes the visualization of the environment with more cameras in the vehicle available and even with several screens as available in the chair of Automotive Technology at Technical University of Munich (Figure 8). In addition, the selective decision of participants could generate new results. The question therefore is whether with improvements in the subjective survey, the results obtained are verified or if otherwise there are differences.

This thesis has primarily focused on video content from urban environments and that, therefore raises the last question. Unlike urban roads, rural roads typically contain fewer objects but often involve a higher frequency of curves and intersections, typically with higher speeds. Furthermore, temporary or unexpected objects that may appear on rural roads are often less predictable compared to urban environments. Teleoperation is studied in this case to improve AD and that, therefore, involves all types of spaces including all types of roads and available paths. This raises the question of whether VQMs are equally valid and effective in both urban and rural environments.

List of illustrations

Table directory

Bibliography

[1] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick und F. Diermeyer, "Survey on Scenario-Based Safety Assessment of Automated Vehicles", *IEEE Access*, Jg. 8, S. 87456–87477, 2020, doi: 10.1109/ACCESS.2020.2993730.

[2] L. Barthelmes, M. E. Görgülü und M. Kagerbauer, "Wissenschaftliche Begleitung der Easy-Mile-Busse in Monheim am Rhein - Ergebnisbericht", 2024.

[3] M. Schwall, T. Daniel, T. Victor, F. Favarò und H. Hohnhold, "Waymo-Public-Road-Safety-Performance-Data"

[4] A. Sinha, S. Chand, V. Vu, H. Chen und V. Dixit, "Crash and disengagement data of autonomous vehicles on public roads in California" (eng), *Scientific data*, Jg. 8, Nr. 1, S. 298, 2021, doi: 10.1038/s41597-021-01083-7.

[5] M. Baumann and J. F. Krems, "Situation Awareness and Driving"

[6] A. Schimpe, J. Feiler, S. Hoffmann, D. Majstorovic und F. Diermeyer, "Open Source Software for Teleoperated Driving" in *2022 International Conference on Connected Vehicle and Expo (ICCVE)*, Lakeland, FL, USA, 2022, S. 1–6, doi: 10.1109/ICCVE52871.2022.9742859.

[7] M. Scholtes *et al.,* "6-Layer Model for a Structured Description and Categorization of Urban Traffic and Environment", *IEEE Access*, Jg. 9, S. 59131–59147, 2021, doi: 10.1109/ACCESS.2021.3072739.

[8] A. Dosovitskiy, "CARLA an open driving simulator", *Proceedings of the 1st Annual Conference on Robot Learning*, S. 1–16, 2017.

[9] A. Sinha, V. Vu, S. Chand, K. Wijayaratna und V. Dixit, "A Crash Injury Model Involving Autonomous Vehicle: Investigating of Crash and Disengagement Reports", *Sustainability*, Jg. 13, Nr. 14, S. 7938, 2021, doi: 10.3390/su13147938.

[10] D. Majstorovic, S. Hoffmann, F. Pfab, A. Schimpe, M.-M. Wolf und F. Diermeyer, "Survey on Teleoperation Concepts for Automated Vehicles", 18. Aug. 2022. [Online]. Verfügbar unter: http://arxiv.org/pdf/2208.08876v1.

[11] S. Gnatzig, F. Schuller und M. Lienkamp, "Human-machine interaction as key technology for driverless driving - A trajectory-based shared autonomy control approach" in *2012 RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, Paris, France, 2012, S. 913–918, doi: 10.1109/ROMAN.2012.6343867.

[12] M. P. Stéphane Bensoussan, "Computer-Aided Teleoperation of an Urban Vehicle - Advanced Robotics, 1997. ICAR '97. Proceedings., 8th International Conference on", 1997.

[13] G. Niemeyer und J.-J. Slotine, "Stable adaptive teleoperation", *IEEE J. Oceanic Eng.*, Jg. 16, Nr. 1, S. 152–162, 1991, doi: 10.1109/48.64895.

[14] L. Basañez und R. Suárez, "Teleoperation" in *Springer Handbook of Automation*, S. Y. Nof, Hg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, S. 449–468, doi: 10.1007/978-3-540-78831-7_27.

[15] T. Fong und C. Thorpe, "Vehicle Teleoperation Interfaces"

[16] M. Hofbauer, C. B. Kuhn, G. Petrovic und E. Steinbach, "TELECARLA: An Open Source Extension of the CARLA Simulator for Teleoperated Driving Research Using Off-the-Shelf Components" in *2020 IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, USA, 2020, S. 335–340, doi: 10.1109/IV47402.2020.9304676.

[17] A. Schimpe, S. Hoffmann und F. Diermeyer, "Adaptive Video Configuration and Bitrate Allocation for Teleoperated Vehicles" in *2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*, Nagoya, Japan, 2021, S. 148–153, doi: 10.1109/IVWorkshops54471.2021.9669258.

[18] F. Stroppa *et al.,* "Shared-Control Teleoperation Paradigms on a Soft-Growing Robot Manipulator", *J Intell Robot Syst*, Jg. 109, Nr. 2, 22. Sep. 2023, doi: 10.1007/s10846-023-01919-x.

[19] Y. Li, A. Takagi und K. P. Tee, "Editorial: Shared Control for Tele-Operation Systems" (eng), *Frontiers in robotics and AI*, Jg. 9, 2022, doi: 10.3389/frobt.2022.915187.

[20] M. Andreas Julius Schimpe, "Uncoupled Shared Control Designs for Teleoperation of Highly-Automated Vehicles"

[21] A. Hosseini, T. Wiedemann und M. Lienkamp, "Interactive path planning for teleoperated road vehicles in urban environments" in *2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, China, 2014, S. 400–405, doi: 10.1109/ITSC.2014.6957723.

[22] T. M. Allen, H. Lunenfeld und G. J. Alexander, "Driver information needs", *Comittee on Motorist Information System 50th annual meeting*.

[23] M. R. Endsley, "Design and Evaluation for Situation Awareness Enhancement", *Proceedings of the Human Factors Annual Meeting*, Jg. 32, S. 97–101, 1988.

[24] Donald L. Fisher, "Modeling situation awareness and crash risk", 2014.

[25] J. Uhrmeister, "The validity of the SAGAT-questionnaire:: An empirical study using simulated driving situations", 2013.

[26] H. R. Wu und K. R. Rao, "Digital Video Image Quality and Perceptual Coding"

[27] H. Strasburger, I. Rentschler und M. Jüttner, "Peripheral vision and pattern recognition: a review" (eng), *Journal of vision*, Jg. 11, Nr. 5, S. 13, 2011, doi: 10.1167/11.5.13.

[28] S. Winkler, "Digital Video Quality: Vision Models and Metrics"

[29] D. H. Foster, "Color constancy" (eng), *Vision research*, Jg. 51, Nr. 7, S. 674–700, 2011, doi: 10.1016/j.visres.2010.09.006.

[30] Adriana Fiorentini and Lamberto Maffei, "Binocular depth perception without geometrical cues", 1971.

[31] L. Thompson, M. Ji, B. Rokers und A. Rosenberg, "Contributions of binocular and monocular cues to motion-in-depth perception" (eng), *Journal of vision*, Jg. 19, Nr. 3, S. 2, 2019, doi: 10.1167/19.3.2.

[32] D. G. Pelli und P. Bex, "Measuring contrast sensitivity" (eng), *Vision research*, Jg. 90, S. 10–14, 2013, doi: 10.1016/j.visres.2013.04.015.

[33] Y. Gao, X. Min, Y. Zhu, J. Li, X.-P. Zhang und G. Zhai, "Image Quality Assessment: From Mean Opinion Score to Opinion Score Distribution" in *MM '22: The 30th ACM International Conference on Multimedia*, Lisboa Portugal, 2022, S. 997–1005, doi: 10.1145/3503161.3547872.

[34] S. Winkler und P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics", *IEEE Trans. on Broadcast.*, Jg. 54, Nr. 3, S. 660–668, 2008, doi: 10.1109/TBC.2008.2000733.

[35] S. Li, L. Ma und K. N. Ngan, "Full-Reference Video Quality Assessment by Decoupling Detail Losses and Additive Impairments", *IEEE Trans. Circuits Syst. Video Technol.*, Jg. 22, Nr. 7, S. 1100–1112, 2012, doi: 10.1109/TCSVT.2012.2190473.

[36] Z. Kotevski und P. Mitrevski, "Experimental Comparison of PSNR and SSIM Metrics for Video Quality Estimation"

[37] Q. Fan, W. Luo, Y. Xia, G. Li und D. He, "Metrics and methods of video quality assessment: a brief review", *Multimed Tools Appl*, Jg. 78, Nr. 22, S. 31019–31033, 2019, doi: 10.1007/s11042-017-4848-x.

[38] B. Girod, "What's wrong with mean-squared error? Digital images and human vision", *MIT Press*, S. 207–220, 1993.

[39] Z. Wang, A. C. Bovik und L. Lu, "Why is image quality assessment so difficult?" in *Proceedings of ICASSP '02*, Orlando, FL, USA, 2002, IV-3313-IV-3316, doi: 10.1109/ICASSP.2002.5745362.

[40] P. C. Teo und D. J. Heeger, "Perceptual image distortion" in *1st International Conference on Image Processing*, Austin, TX, USA, 1994, S. 982–986, doi: 10.1109/ICIP.1994.413502.

[41] Q. Huynh-Thu und M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates", *Telecommun Syst*, Jg. 49, Nr. 1, S. 35–48, 2012, doi: 10.1007/s11235-010-9351-x.

[42] U. Sara, M. Akter und M. S. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study", *JCC*, Jg. 07, Nr. 03, S. 8–18, 2019, doi: 10.4236/jcc.2019.73002.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh und E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity" (eng), *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, Jg. 13, Nr. 4, S. 600–612, 2004, doi: 10.1109/tip.2003.819861.

[44] S. S. Channappayya, A. C. Bovik und R. W. Heath, "Rate bounds on SSIM index of quantized images" (eng), *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, Jg. 17, Nr. 9, S. 1624–1639, 2008, doi: 10.1109/TIP.2008.2001400.

[45] M. Hassan, "Structural Similarity Measure for Color Images"

[46] H. R. Sheikh und A. C. Bovik, "Image information and visual quality" (eng), *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, Jg. 15, Nr. 2, S. 430–444, 2006, doi: 10.1109/tip.2005.859378.

[47] S. Rezazadeh und S. Coulombe, "Low-complexity computation of visual information fidelity in the discrete wavelet domain" in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2010*, Dallas, TX, 2010, S. 2438–2441, doi: 10.1109/ICASSP.2010.5496298.

[48] Y. Han, Y. Cai, Y. Cao und X. Xu, "A new image fusion performance metric based on visual information fidelity", *Information Fusion*, Jg. 14, Nr. 2, S. 127–135, 2013, doi: 10.1016/j.inffus.2011.08.002.

[49] S. Li, F. Zhang, L. Ma und K. N. Ngan, "Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments", *IEEE Trans. Multimedia*, Jg. 13, Nr. 5, S. 935–949, 2011, doi: 10.1109/TMM.2011.2152382.

[50] Z. Li und A. Aaron, "Toward A Practical Perceptual Video Quality Metric", *Netflix TechBlog*, Juni 2016.

[51] B. García, L. López-Fernández, F. Gortázar und M. Gallego, "Practical Evaluation of VMAF Perceptual Video Quality for WebRTC Applications", *Electronics*, Jg. 8, Nr. 8, S. 854, 2019, doi: 10.3390/electronics8080854.

[52] R. Rassool, "VMAF reproducibility: Validating a perceptual practical video quality metric" in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Cagliari, Italy, 2017, S. 1–2, doi: 10.1109/BMSB.2017.7986143.

[53] M. Orduna, C. Diaz, L. Munoz, P. Perez, I. Benito und N. Garcia, "Video Multimethod Assessment Fusion (VMAF) on 360VR Contents", *IEEE Trans. Consumer Electron.*, Jg. 66, Nr. 1, S. 22–31, 2020, doi: 10.1109/TCE.2019.2957987.

[54] *VMAF GitHub repository*. Netflix. [Online]. Verfügbar unter: https://github.com/Netflix/vmaf

[55] D. Fleet and Y. Weiss, "Optical flow estimation"

[56] B. K. Horn und B. G. Scgunck, "Determining Optical Flow", *Artificial Inteligence*, Jg. 17, S. 185–203, 1981.

[57] Arush, *Part 1 - Visual Feature Detection for Autonomous Vehicle Video Streams.* [Online]. Verfügbar unter: https://medium.com/building-autonomous-flight-software/using-opencv-to-detect-features-in-autonomous-driving-4d7c5348ee4 (Zugriff am: 17. Juli 2024).

[58] S. Beauchemin und J. L. Barron, "The computation of optical flow", *ACM Computing Surveys*, Jg. 27, Nr. 3, 3. Sep. 1995.

[59] Arush, *Part 2 - The Math Behind Optical Flow.* [Online]. Verfügbar unter: https://medium.com/building-autonomous-flight-software/math-behind-optical-flow-1c38a25b1fe8 (Zugriff am: 17. Juli 2024).

[60] Arush, *Part 3 - Lucas-Kanade Optical Flow.* [Online]. Verfügbar unter: https://medium.com/building-autonomous-flight-software/lucas-kanade-optical-flow-942d6bc5a078 (Zugriff am: 17. Juli 2024).

[61] M. Otte und H.-H. Nagel, "Optical Flow Estimation: Advances and Comparisons"

[62] D. Sun, S. Roth und M. J. Black, "A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them", *Int J Comput Vis*, Jg. 106, Nr. 2, S. 115–137, 2014, doi: 10.1007/s11263-013-0644-x.

[63] Hiroaki Hayashi *et al., A Driver Situational Awareness Estimation System Based on Standard Glance Model for Unscheduled Takeover Situations*. Piscataway, New Jersey: IEEE, 2019. [Online]. Verfügbar unter: https://ieeexplore.ieee.org/servlet/opac?punumber=8792328

[64] A. Munir, A. Aved und E. Blasch, "Situational Awareness: Techniques, Challenges, and Prospects", *AI*, Jg. 3, Nr. 1, S. 55–77, 2022, doi: 10.3390/ai3010005.

[65] U. Engelke, M. Barkowsky, P. Le Callet und H.-J. Zepernick, "Modelling saliency awareness for objective video quality assessment" in *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX 2010)*, Trondheim, Norway, 2010, S. 212–217, doi: 10.1109/QOMEX.2010.5516159.

[66] E. Blaauwgeers, L. Dubois und L. Ryckaert, "Real-time risk estimation for better situational awareness", *IFAC Proceedings Volumes*, Jg. 46, Nr. 15, S. 232–239, 2013, doi: 10.3182/20130811-5-US-2037.00036.

[67] S. Neumeier, S. Stapf und C. Facchi, "The Visual Quality of Teleoperated Driving Scenarios How good is good enough?" in *2020 International Symposium on Networks, Computers and Communications (ISNCC)*, Montreal, QC, Canada, 2020, S. 1–8, doi: 10.1109/ISNCC49221.2020.9297343.

[68] A. V. Katsenou, F. Zhang, K. Swanson, M. Afonso, J. Sole und D. R. Bull, "VMAF-based Bitrate Ladder Estimation for Adaptive Streaming" in *2021 Picture Coding Symposium (PCS)*, Bristol, United Kingdom, 2021, S. 1–5, doi: 10.1109/PCS50896.2021.9477469.

[69] V. V. Menon, P. T. Rajendran, R. Farahani, K. Schoeffmann und C. Timmerer, "Video Quality Assessment with Texture Information Fusion for Streaming Applications" in *MHV '24: Mile-High Video Conference*, Denver CO USA, 2024, S. 1–6, doi: 10.1145/3638036.3640798.

[70] *Separate autoware_auto_msgs into several packages.* [Online]. Verfügbar unter: https://gitlab.com/autowarefoundation/autoware.auto/autoware_auto_msgs/-/blob/master/autoware_auto_control_msgs/msg/AckermannControlCommand.idl?ref_type=heads

[71] N. Raina *et al.,* "EgoBlur: Responsible Innovation in Aria", 24. Aug. 2023. [Online]. Verfügbar unter: http://arxiv.org/pdf/2308.13093v2.

[72] S. Macenski, T. Foote, B. Gerkey, C. Lalancette und W. Woodall, *Robot Operating System 2: Design, architecture, and uses in the wild.* [Online]. Verfügbar unter: https://docs.ros.org/en/humble/Concepts/Basic/About-Topics.html.

[73] Imavijit, *Python | Spitzen-Signal-Rausch-Verhältnis (PSNR).* [Online]. Verfügbar unter: https://www.geeksforgeeks.org/python-peak-signal-to-noise-ratio-psnr/ (Zugriff am: 15. Mai 2024).

[74] J. Newmarch, "FFmpeg/Libav" in *Linux Sound Programming*, J. Newmarch, Hg., Berkeley, CA: Apress, 2017, S. 227–234, doi: 10.1007/978-1-4842-2496-0_12.

[75] *Computer Vision: Optical Flow*. GitHub, 2021. [Online]. Verfügbar unter: https://github.com/niconielsen32/ComputerVision/blob/master/opticalFlow/denseOpticalFlow.py

[76] Gunnar Farnebäck, "Two-Frame Motion Estimation Based on Polynomial Expansion"

[77] J. Bigün und T. Gustavsson, *Image analysis: 13th Scandinavian conference, SCIA 2003, Halmstad, Sweden, June 29-July 2, 2003 proceedings*. Berlin, New York: Springer, 2003.

[78] P. Royston, "Approximating the Shapiro-Wilk W-test for non-normality", Jg. 2, S. 117–119, 1992.

[79] C. Zaiontz, *Shapiro-Wilk Expand Test.* [Online]. Verfügbar unter: https://real-statistics.com/tests-normality-and-symmetry/statistical-tests-normality-symmetry/shapiro-wilk-expanded-test/ (Zugriff am: 10. August 2024).

[80] Z. Bobbitt, *How to Perform the Friedman Test in Excel* (Zugriff am: 10. August 2024).

Appendix

A General participant information

The following data describes the general information of the participants in the online survey. Some important values from this were used in the different chapters of the thesis.

Number of records in this query:	52
Total records in survey:	52
Percentage of total:	100,00%

Summary for E01Q01
Wie alt bist du?

Answer	Count	Percentage
Von 0 bis 15 (AO01)	1	1,92%
Von 15 bis 30 (AO02)	25	48,08%
Von 30 bis 45 (AO03)	8	15,38%
Von 45 bis 60 (AO04)	14	26,92%
+60 (AO05)	4	7,69%
No answer	0	0,00%
Not displayed	0	0,00%

Summary for E01Q02
Haben Sie einen Führerschein?

Answer	Count	Percentage
Yes (Y)	51	98,08%
No (N)	1	1,92%
No answer	0	0,00%
Not displayed	0	0,00%

Summary for E01Q03
Wie viele Jahre fahren Sie schon?

Calculation	Result
Count	52
Sum	910,5
Standard deviation	14,6
Average	17,51
Minimum	0
1st quartile (Q1)	3,25
2nd quartile (Median)	12
3rd quartile (Q3)	30
Maximum	46

Null values are ignored in calculations

Q1 and Q3 calculated using minitab method

Summary for E01Q04
Wie viele Tage fahren Sie durchschnittlich pro Woche?

Calculation	Result
Count	52
Sum	168
Standard deviation	2,06
Average	3,23
Minimum	0
1st quartile (Q1)	1
2nd quartile (Median)	3
3rd quartile (Q3)	5
Maximum	7

Null values are ignored in calculations
Q1 and Q3 calculated using minitab method

Summary for E01Q05
Wo machen Sie das Experiment?

Answer	Count	Percentage
Computer (1)	8	15,38%
Laptop (2)	18	34,62%
Tablet (3)	2	3,85%
Smartphone (4)	24	46,15%
Other	0	0,00%
No answer	0	0,00%
Not displayed	0	0,00%