Journal of Intelligent & Robotic Systems (2022) 106:48
https://doi.org/10.1007/s10846-022-01749-3

SHORT PAPER

High Latency Unmanned Ground Vehicle Teleoperation Enhancement by Presentation of Estimated Future through Video Transformation

MD Moniruzzaman[1] · Alexander Rassau[1] · Douglas Chai[1] · Syed Mohammed Shamsul Islam[2]

Received: 25 May 2022 / Accepted: 26 September 2022 / Published online: 13 October 2022

Abstract

Long-distance, high latency teleoperation tasks are difficult, highly stressful for teleoperators, and prone to over-corrections, which can lead to loss of control. At higher latencies, or when teleoperating at higher vehicle speed, the situation becomes progressively worse. To explore potential solutions, this research work investigates two 2D visual feedback-based assistive interfaces (sliding-only and sliding-and-zooming windows) that apply simple but effective video transformations to enhance teleoperation. A teleoperation simulator that can replicate teleoperation scenarios affected by high and adjustable latency has been developed to explore the effectiveness of the proposed assistive interfaces. Three image comparison metrics have been used to fine-tune and optimise the proposed interfaces. An operator survey was conducted to evaluate and compare performance with and without the assistance. The survey has shown that a 900ms latency increases task completion time by up to 205% for an on-road and 147% for an off-road driving track. Further, the overcorrection-induced oscillations increase by up to 718% with this level of latency. The survey has shown the sliding-only video transformation reduces the task completion time by up to 25.53%, and the sliding-and-zooming transformation reduces the task completion time by up to 21.82%. The sliding-only interface reduces the oscillation count by up to 66.28%, and the sliding-and-zooming interface reduces it by up to 75.58%. The qualitative feedback from the participants also shows that both types of assistive interfaces offer better visual situational awareness, comfort, and controllability, and significantly reduce the impact of latency and intermittency on the teleoperation task.

Keywords Teleoperation · Robotic vehicle · Video transformation · Latency · Control · Enhancement techniques

1 Introduction

From the conceptual development to the current state, technology related to teleoperation has improved significantly over the last century [1]. Necessity and usability in surveillance, human exploration, safer transportation, mining, environmental observation, agriculture, medical surgery or even space exploration have inspired and boosted the motivation for robotic teleoperation technology [2]. The word teleoperation comes from the Greek term 'tele' which means "at a distance" or "far off". Therefore, teleoperation naturally implies the act of distant operation [3]. Teleoperation of robots, manipulators and robotic vehicles is a genre of teleoperation where an operator acts as a master, acquires information of the remote environment and establishes communication with the robotic entity over a communication channel, provides orders and supervisory suggestions, and the desired task is executed by the robotic entity according to the control and feedback from the human operator [4].

In the modern concept of teleoperation, the level of control of the human operator can differ according to the varying levels of artificial intelligence and autonomy associated with the robotic entity [5]. For facilitating varying levels of control through teleoperation, several techniques have been proposed namely, adjustable autonomy [6], collaborative

✉ MD Moniruzzaman
m.moniruzzaman@ecu.edu.au

Alexander Rassau
a.rassau@ecu.edu.au

Douglas Chai
d.chai@ecu.edu.au

Syed Mohammed Shamsul Islam
syed.islam@ecu.edu.au

[1] School of Engineering, Edith Cowan University, 270 Joondalup Drive, Joondalup WA 6027, Perth, Australia

[2] School of Science, Edith Cowan University, 270 Joondalup Drive, Joondalup WA 6027, Perth, Australia

control [7], mixed-initiative control [8], and sliding autonomy [9] amongst others. Although these techniques attempt to enhance teleoperation through ensuring appropriate control, they are all impacted by communication constraints and delay. In this study, we are focusing on exploring methods for enhancing teleoperation of unmanned ground vehicles (UGVs) by reducing the impact of delay on the teleoperation system.

1.1 Background

Delay in teleoperation, also known as latency, lag or command delay, occurs due to the data transmission time through the communication medium [10]. In robotic teleoperation, latency refers to the delay between the operators' control input and its impact on the visual feedback from the robotic environment [11]. Latency or delay in teleoperation, especially in teleoperation of ground robotic vehicles is not only difficult to handle, but also stressful, and overburdens the human operator with high cognitive workloads [12]. Latency decreases teleoperation performance [11] by reducing accuracy and increasing completion time of the teleoperation tasks [13]. A teleoperator's perception starts to be affected by latency from as little as 10-20 milliseconds (ms) [14]. Teleoperation reaction time increases by up to 64% if the latency increases to 225 ms [10]. Several studies such as [13] and [15] reported compromise in pursuit tracking performance of the teleoperator when the latency crosses 300 ms. Jitter or variable latency have an even more detrimental impact on teleoperation. While teleoperating field robotic vehicles such as unmanned ground vehicles (UGVs), unmanned aerial vehicles UAVs and remotely operated vehicles (ROVs) at a reasonable ground speed, latency causes the operator to execute repeated commands and overcorrect the steering. This overcorrection causes undesirable oscillation and potential loss of control [11, 16]. Oscillation and overcorrection make obstacle avoidance difficult and can cause damage both to the robot and the remote environment. Moreover, latency has been reported to cause motion sickness to the operator [11].

Not only is the inconvenience created by latency affecting teleoperation, but the teleoperation enhancement research itself is also impacted by the bottleneck created by logistical and data deficiency issues. Conventional teleoperation enhancement research requires expensive robots and robotic vehicles along with other control, visual and communication equipment. Moreover, outdoor experimentation is prone to hazards. Recent advancements in digital imaging technologies [2], high performance computation facilities [17], human-machine interfaces [18], and computer vision, and artificial intelligence technologies [19, 20] have opened up the possibilities of better teleoperation experiences. However, there is a significant shortage of proper visual and control input data sets required for teleoperation research with these new techniques. All of the above factors motivated us to design a new teleoperation simulator in order to explore possible methods to reduce the impact of delays in the control loop and thus enhance teleoperation.

1.2 Literature Review

Teleoperation is the perception of being present inside or within a simulated or physically distant environment [21]. Based on the mode of robot mobility and the type of remote environment, teleoperation can be categorised into four different types: stationary manipulator teleoperation, ground robotic vehicle teleoperation, aerial robotic vehicle teleoperation and underwater robotic vehicle teleoperation . All these types of teleoperation are influenced by camera viewpoint or field-of-view (FOV), depth perception, orientation, speed or motion of the vehicles, and quality of the transmitted video including frame rate and latency [11]. Considering the presence of all of the mentioned challenges in ground vehicle teleoperation, our initial research focus is on enhancing teleoperation of ground robotic vehicles. There could be four different teleoperation modes for ground robotic vehicles: direct control, multimodal, supervisory, and novel [22]. Our research is focused on the direct teleoperation control mode for a single ground vehicle teleoperation system where the human operators rely on video feedback from the robotic platform and provide control feedback through conventional controllers such as a steering wheel, and brake and acceleration pedals [23].

For enhancing teleoperation, a significant amount of research work has been conducted by the research community over more than half a century to solve fundamental control problems that arise over the communication link [24]. The enhancement techniques can be categorised as visual and non-visual enhancement techniques. Visual enhancement techniques can be further divided into 2D and 3D enhancement techniques. Exocentric view, automatic view adjustment, stereoscopic vision, virtual environment, vision-based object tracking and predictive systems are included in visual feedback enhancement techniques for teleoperation enhancement. In this paper, we will provide a background for 2D predictive feedback based enhancement techniques as we aim to reduce the impact of latency with an assistive predictive interface. Our core original idea is to design an interface to mimic a real-time visual feed for teleoperation based on predicting the anticipated future state. The state-of-the-art predictive techniques are summarised in Table 1 and discussed below.

1.2.1 Current 2D Predictive Display-Based Enhancement Techniques

Increased perceptional awareness of the environment enhances control over teleoperated mobile vehicles. However, communication delay between the teleoperator and

Table 1 Summary of 2D predictive feedback-based teleoperation enhancement techniques

Author, Year & Ref.	Technique	Medium	Robot type
2D predictive feedback			
Dybvik et al. (2021) [25]	Positional and scale transformation to the video display	Wired	Wheeled ROV
Wilde et al. (2020) [26]	Predictive flight path using velocity telemetry, camera feed and control inceptor deflection	WiFi	UAV
Ha et al. (2018) [27]	Future state and collision prediction for a multi mobile robot leader follower system	-	Wheeled ROV
Wang et al. (2016) [28]	Truncated prediction of states for nonlinear multi-agent teleoperation	Any medium	-
Matheson et al. (2013) [12]	Projected field of view by cropping and zooming	Wireless	Space Rover

the robotic platform is inevitable and degrades performance significantly. In challenging environments, longer delays or higher levels of jitter (time variance of delay) can even make teleoperation effectively impossible [29]. One prospective way to reduce the impact of teleoperation delay is by predicting the evolution of the state variable for the period of delay [30, 31]. Delays in the control loop motivated the development of predictive displays from the 1990s. The earlier approaches such as [32–35] tried to implement the concept of predictive displays that allows the operator to view the response of the system before it actually happens and hence avoid possible collisions. Witus et al. [36] implemented the state prediction of the UGV as a form of iconography for both AR and VR. However, these approaches were mostly prediction based on non-video-based feedback and do not provide an intuitive control interface, and are not suited for higher speed operation.

Some approaches have tried to solve problems that have arisen due to communication delays by predicting robot pose or states from 2D video feedback. Wang et al. [28] investigated input delays and nonlinearities for a teleoperation scenario and proposed a solution that relied on truncated prediction of Lipschitz nonlinear multi-agent systems. Their approach considered system state integral terms by tentatively applying the Krasovskii functional method. Ha et al. [27] used a propagation stage prediction technique for teleoperating a set of non-holonomic mobile robots. Their 2D predictive display showed the current and future poses of all the mobile robots. Their prediction horizon was up to two metres. However, the average speed of the mobile robots was only about 0.15m/s which was low.

To address the multi-second delay for space rovers Matheson et al. [12] described a simple, however somewhat effective technique. While the moving vehicle moves forward in a single direction, a zoom in to the images can give a future prediction to the trajectory. By cropping, zooming and projecting the image they were able to reduce the impact of the high latency of 3 s and halved the time of task completion. However, this technique is not applicable for parallax movement. A further improvement of the approach by Matheson et al. [12] has been implemented recently by Dybvik et al. [25]. In addition to the zooming method, positional and scale transformation was implemented for a better predictive display, but this was only applied to vehicles operating at low ground speeds.

Simulation of some predictive frameworks that do not include predictive displays as an operator aid, only algorithm-based predictions, includes the work by Zheng et al. [37] who described a model of predictor framework for UGVs and Zhang and Li [38] who attempted to design a predictor model based on the Clohessy-Wiltshire relative dynamic equation.

All of the above 2D predictive enhancement techniques provide either first-order state prediction of the robotic system or future pose estimation only. They offer a very limited horizon for prediction and can only be applied to stationary or slow-moving robotic vehicles. As these systems are designed as specific robot-in-the-loop systems, they are not easily transferable to other robot types. Moreover, these teleportation systems and enhancement techniques cannot be easily integrated into modern AI-based systems and techniques, nor have they been designed to collect data that can be used to train deep learning or AI-based enhancement tools. Therefore, in a time when AI and neural networks are being used to very effectively solve problems across a wide range of scientific research domains, a system that can simulate teleoperation with controllable latency, and can be used to generate synthetic images and control data for investigating AI-based enhancement techniques is of significant value. Further, an effective technique is required that can enhance teleoperation for long-distance and high-latency teleoperation scenarios. The research presented in this paper proposes methods to fill these identified research gaps.

1.3 Contribution of the Paper

In this paper, we describe a system that has the capability to resolve the data availability issue for AI-based teleoperation enhancement research, while also proposing a method to enhance teleoperation based on low cost equipment and easily implementable techniques. We have developed

a teleoperation simulation model that is capable of simulating teleoperation with controllable latency without the use of a real robotic vehicle. For teleoperation enhancement, to explore the effectiveness of video prediction based enhancement techniques, we have initially investigated straightforward video transformation techniques, which are easy to replicate and integrate into a real-time teleoperation system. To evaluate the performance of the model and enhancement approach, and to fine-tune it, we have used pixel matching based image quality measuring metrics including Peak Signal to Noise Ratio (PSNR) [39] and the structural difference matching metric Structural Similarity Index Measure (SSIM) [40]. However, to properly assess the level of enhancement achieved for a teleoperation scenario, the operator experience is what really matters. Therefore, we have conducted a human operator survey to determine whether our video transformation-based assistive interfaces are genuinely impacting the teleoperation experience in a positive way. The stages of the research and our contributions are highlighted using the flow chart in Fig. 1.

To the best of our knowledge, there have been no previous approaches that have implemented a low cost teleoperation simulator, that is capable of simulating controllable latency for ground vehicle teleoperation scenarios, and can be used to generate large amounts of synthetic image, video, and control input data for AI and deep learning based teleoperation enhancement research. Including this, the main contributions of this paper are as follows.

1. Designing a Simulink-based human-in-the-loop ground vehicle teleoperation simulation platform with controllable latency that can simulate remote ground vehicle operation at high speeds, while generating synthetic image, video, and control input data for teleoperation enhancement research.
2. Formulating an algorithm that accounts for the control input signals (acceleration, deceleration, steering), and simulated vehicle speed to enhance teleoperation through image/video transformation.
3. Applying pixel and structural similarity-based image analysis techniques for ground vehicle teleoperation simulator optimisation (image analysis techniques have rarely been used for teleoperation evaluation, and have never been used for UGV simulator performance measurement and optimisation).
4. Performing a human operator survey to investigate the impacts of high latency on ground vehicle teleoperation and evaluate the effectiveness of the video transformation-based enhancement technique to enhance teleoperation.
5. Providing an in-depth discussion of the qualitative and quantitative aspects of the survey outcomes and teleoperation performance evaluations and the implications of these for future teleoperation enhancement research.

The rest of the paper is structured as follows. Section 2 describes in detail the system we designed for teleoperation simulation and the interfaces we developed for teleoperation enhancement. Section 3 illustrates the model tuning techniques and the evaluation methods used in this research. Section 4 presents the results and Section 5 discusses these, including the operator survey feedback. Conclusion can be found in Section 6.

2 Teleoperation Task Simulator Design with Integrated Predictive Interface

One of the major components of this research is to eliminate the requirement for a real ground robotic vehicle in the teleoperation research chain. As illustrated in Fig. 2, one of the prime components of the simulator we designed is a virtual vehicle driving platform. Other major parts of the simulator are the control input equipment, visual feedback capture device and the latency control unit. These major components of the simulator are described below along with the detailed description of the algorithms of the predictive interface explored as an enhancement technique. Figure 2 presents the schematic diagram of

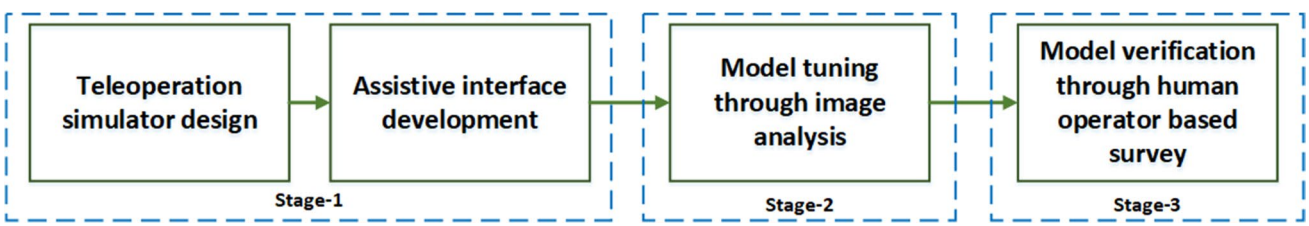

Fig. 1 The overall process flow of the research. Stage-1 of the research involved the development of a novel simulator with the proposed assistive interface (as described in Section-2); In stage-2, the model has been fine-tuned and optimised through image analysis (as described in Section-3.1); in the final stage a human operator based survey has been designed and carried out to verify the concept and quantify the effectiveness of the model (as described in Section-3.2)

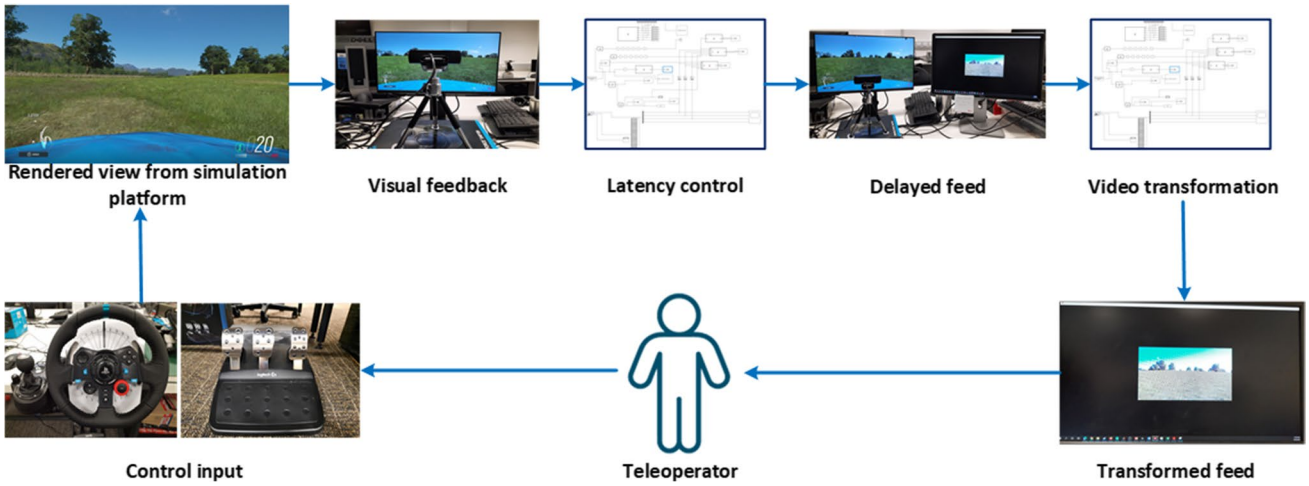

Fig. 2 System diagram of the proposed teleoperation simulator

the teleoperation simulator with incorporated assistive interface.

2.1 Simulation Platform

To create a teleoperation platform for a virtual vehicle that can be used as a teleoperated ground vehicle we have chosen a commercial racing game called 'Forza Horizon 4'. It was developed by Microsoft Studios and released in October 2018. This racing game offers a driving experience in fictionalised regions of Great Britain. It facilitates both on-road and off-road driving experiences. Moreover, its near-photorealistic visual representation of the environment coupled with reasonably realistic ground vehicle physics has made it as an excellent candidate to be used as a virtual environment to simulate our teleoperation task.

2.2 Control Input Equipment

To provide operator control inputs to the simulation platform, we have used a 'Logitech G29 Driving Force Racing Wheel' along with brake and acceleration pedals (Fig. 2(b)). The Logitech Wheel provides 900-degree lock-to-lock rotation similar to a real car steering wheel. The throttle, brake and clutch pedals are integrated into the unit's separate floor pedal. The brake pedal is nonlinear and is capable of mimicking a pressure-sensitive brake system.

2.3 Visual Feedback Capture Device

Controllers connected to a simulation platform only provide a real-time road driving experience. To simulate a teleoperated environment, we require a visual feedback capture device that will receive visual feedback from the simulation platform or game screen and feed it to a latency adding and controlling model. We have used a 'Logitech C922' pro stream webcam for this purpose. This webcam is capable of capturing videos in 1080p at 30 frames per second (fps) or 720p at 60 fps with a 78-degree field of view. It can accommodate flickering lights from light sources, which is vital for our use case as we use the camera to capture the computer screen. To represent the typical scenario of teleoperation of a ground vehicle robot in a remote environment via low bandwidth communication channels, the frame rate is limited to 10 fps for the video feed, which is adequate for real-time driving at reasonable ground speeds.

2.4 Latency Controller and Delayed Feed Display

It is inevitable for teleoperation to be affected by some degree of latency. For any long-distance teleoperation scenario, the impact is even more noticeable. Therefore, one of the major components of a teleoperation simulator is the capability of inducing and controlling remote visual feed delay. The simulator should be able to accept and compile control inputs (i.e., steering wheel rotation, brake, acceleration, etc.) along with displaying and saving them for later analytical purposes. It should also be able to receive and compile video feedback either from the simulation platform or from the visual feedback capture device, process the video signal, and save, and display the result. Most importantly, the simulator must be able to add latency to the visual feedback so that the operators feel the impact when the simulation is running.

To offer all of the above-mentioned characteristics to the simulator, we have developed a Simulink® model. To capture the video feed from the Logitech C922 camera and

bring the image frames of the video feed to the Simulink® model, we have used the 'From Video Device' block sets. This block is capable of queuing all the incoming frames in a first-in, first-out (FIFO) buffer and delivering one image frame for each simulation time step. We have configured the block to have only one output port that works as a gateway for the RGB frames to the rest of the model.

As mentioned earlier, for driving a vehicle in the Forza Horizon 4 gaming platform, a Logitech G29 controller unit is used. To capture and bring the induced signals such as wheel rotation, brake and acceleration pedal press and their intensity in real-time we have used a 'Joystick Input' block. This block provides interaction of control signals between the virtual world and the Simulink® model. The joystick input block feeds all the signals from the controller unit as a single axes signal. The signal is demultiplexed to split it into separate signals from the steering wheel, brake and acceleration pedals.

Our model is capable of controlling latency to the video feed at any desired level from a minimum of 300 ms (base Simulink® delay). For our experiment we have set the total latency to 900 ms. Any amount of latency can be added to the video feed using additional Simulink 'Unit Delay'blocks. This is a simple input-output block and equivalent to the z^{-1} discrete-time operator that holds its input by one iteration. We chose to keep the latency to below 1 second as the participants repeatedly failed to keep control of the ground vehicle during teleoperation sessions for latencies higher than one second during the human survey stage of this research, making comparisons between enhanced and unenhanced operation difficult. To visualise the delayed feed to the operator we have used the Video Display block from Simulink®. This block is capable of displaying high definition video.

Our model is simple, yet capable of accurately simulating the teleoperation of a ground vehicle. It is capable of easily inducing and controlling latency. Therefore, this model can be used to experiment with the impacts of variable latency on teleoperators. It can also be used to develop and test techniques to enhance teleoperation. This simple teleoperation platform is capable of saving control input signals and video feeds. These control input signals and videos can be used to develop and evaluate video transformation, AI, and deep neural network-based teleoperation enhancement techniques, which are planned future steps for this research. To the best of the authors' knowledge, there is no UGV teleoperation simulation platform that uses a virtual robotic ground vehicle, but provides a high quality representation of a real-world scenario, facilitates high-speed teleoperation, and saves the video feed along with the control signals for further experimentation. Note, an assumption has been made in the development of this teleoperation simulation platform that visual feedback delay and control input delay are effectively equivalent from the perspective of the operator, and so to maintain platform simplicity, only visual feedback delay is implemented. A validation of this assumption is provided in Appendix.

2.5 Predictive Interface Development

A teleoperator driving a ground vehicle through the delayed feed of our simulation platform will experience a similar impact from latency to a real-world UGV teleoperation scenario with the same amount of latency. We hypothesise that if the real-time effect of a teleoperator's control actions can be predicted and imposed on the delayed video feed, i.e. if the delayed feed can be transformed to show the impacts of the control input in real-time, the teleoperation experience can be enhanced, and the driveability of the vehicle significantly improved, even at relatively high ground speeds and for latencies as high as one second in a real-world on- or off-road environment. To test the hypothesis, we need to transform the delayed feed according to the change in position of the vehicle resulting from turning due to rotation of the steering wheel. We also need to transform the acceleration or deceleration of the vehicle resulting from inputs to the brake or accelerator pedals. In our simulator model, we have integrated assistive interfaces that achieve both of these goals through simple video transformation as discussed in this subsection.

2.5.1 Video Transformation Based on Steering Wheel Rotation

For driving and teleoperation of a ground vehicle, the whole of the peripheral vision of the driver (for our use case, the whole of the scene captured by the remote camera) is not strictly required. Therefore, as an initial video transformation approach, 30% of the whole image frames are cropped out. The width and height of the modified frames are determined as follows:

$$Modified\ frame\ width = Original\ frame\ width \times 0.70 \quad (1)$$

$$Modified\ frame\ height = Original\ frame\ height \times 0.70 \quad (2)$$

Figure 3 shows both the whole image frame and the cropped feed that is presented to the teleoperator. The red arrows show the objects visible in the whole scene that are absent in the cropped feed, but that do not meaningfully impact the situational awareness of the teleoperator or the control decisions they would make. This may not apply to objects in very close proximity to the vehicle, but in a high latency teleoperation environment it is likely to be too late for control action to avoid a collision if objects reach that close to the vehicle in any case.

$$x = (Original\ frame\ width - Modified\ frame\ width) \div \alpha \quad (3)$$

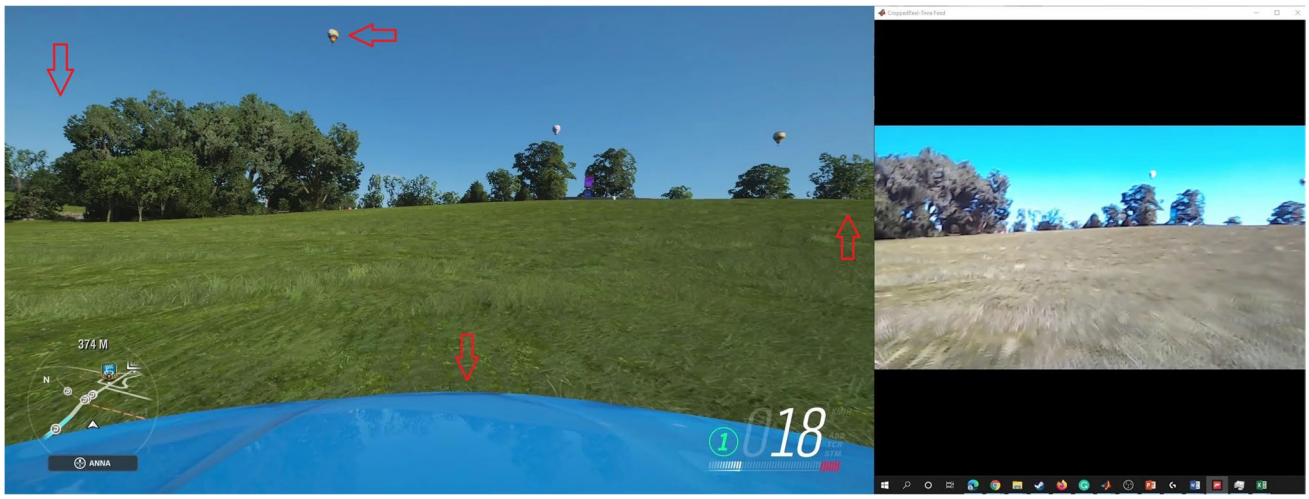

Fig. 3 The cropped feed (right) provides an operator with sufficient situational awareness to teleoperate a car in both on- and off-road scenarios

$$y = (Original\ frame\ height - Modified\ frame\ height) \div \beta \tag{4}$$

Due to latency, the teleoperator is experiencing a delay in the reflection of their control inputs in the visual feed. As the cropped feed is 70% of the full transmitted scene, we have the option to slide our cropped window anywhere in the full- frame based on the movement of the control input signals in real-time to create a perception of the real-time reflection of the operator's control input. To determine the exact location of the initial cropped 70% window, it is sufficient to know the top left pixel location of the window. We set the x and y coordinates of the top left pixel location according to Eqs. 3 and 4. We have kept the α value as 2 and the β value as 6 so that the initial cropped window remains in a central location along the width and the teleoperator can receive optimal situational information about the remote environment. To mimic the real-time direction change of the vehicle in the gaming platform through steering wheel rotation, the cropped window is moved either to the left or to the right in accordance with the direction of rotation of the steering wheel by the operator based on Eq. 5. In this equation, the wheel is the wheel rotation signal that ranges between $-1 \leq wheel \leq 1$. The value of the Speed depends on the UGV speed and c is a hyperparameter we used to fine-tune the predicted field-of-view of the cropped window by adjusting its movement. In our Simulink-based simulator, the assisted feed is generated and passed to the video display block by our custom simulator function block that runs Algorithm 1.

$$x_{assisted} = x + x \times c \times Wheel\ Rotation \times \mathbb{R}\left(log_{10}(1 + Speed)\right) \tag{5}$$

The assisted feed is a function of the delayed image feed, steering wheel rotation by the teleoperator and the speed of the simulated vehicle. As the simulation platform is a commercial video game and there is no way to directly access a signal representing the speed, we have acquired the speed of the vehicle from the captured feed by implementing an optical character recognition (OCR) [41] technique. After segmenting the portion of the screen that displays the speed of the vehicle, grayscale conversion and filtering have been applied prior to the OCR to increase the accuracy and reduce the error of the OCR function. We have named the window of the visual feed generated through this video transformation algorithm as the 'Sliding-only (SO)' window throughout the rest of the paper.

2.5.2 Video Transformation Based on Acceleration and Deceleration

In a standard forward-moving vehicle, from the driver's perspective, all the external objects approach the driver at varying speed according to the speed of the vehicle. However, in a latency impacted teleoperation scenario, the objects in a distant location in the delayed feed are always further away than the same object in the actual environment. To compensate for this disparity, if the delayed frames are scaled up or zoomed-in in accordance to the speed of the vehicle, it would offer the teleoperators' a predicted future view of the objects in front of the vehicle. Figure 4 shows how the simple transformation operation of zooming-in to the delayed feed achieves the desired outcome of effectively predicting the future location of objects in a teleoperation scenario.

To achieve the zoomed-in effect we have warped the delayed frames using the Simulink® 'Warp' block with a

Algorithm 1 Algorithm to transform the video feed as a predictive sliding-only window to reflect the operator control inputs in real-time

Require: $Assisted\ Feed = f(Image\ frame, Wheel\ Rotation, Speed)$

Ensure: $x = x$ coordinate of pixel;
Ensure: $y = y$ coordinate of pixel;
Ensure: $width =$ width of the $Image frame$;
Ensure: $height =$ height of the $Image frame$;
while The system runs **do**
  $x_{initial} = (Original\ frame\ width - Modified\ frame\ width) \div \alpha$;
  $y_{initial} = (Original\ frame\ height - Modified\ frame\ height) \div \beta$;
  $width_{final} = 70\% \times width$;
  $height_{final} = 70\% \times height$;
  $x_{assisted} = x_{initial} + x_{initial} \times c \times Wheel\ Rotation \times \mathbb{R}(log_{10}(1 + Speed)\alpha$;
  $Assisted\ Feed = \text{crop}\ [x_{assisted}\ y_{initial}\ width_{final}\ \ height_{final}]$;
end while
Result: Shifting the cropped window according to the wheel rotation and speed initialisation

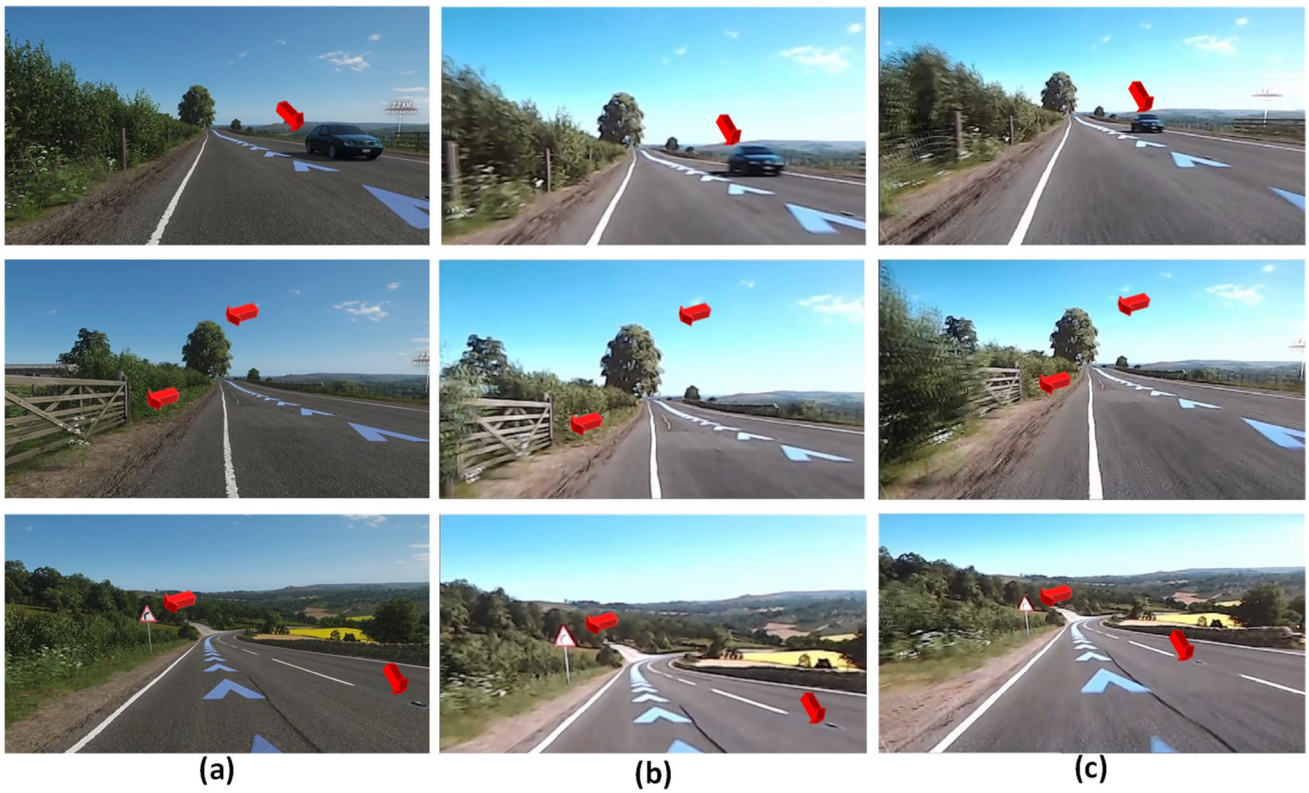

Fig. 4 A zooming-in transformation (**b**) of the delayed feed (**c**) provides the predicted future location of the objects (such as the car, tree, road sign, road driver mark, etc.) in the assisted feed that is closer to the ground truth (**a**)

zoomed transformation factor. The zoom factor was achieved using Eq. 6. In this equation, γ is an adjustable hyper-parameter we used to fine-tune the amount of zooming our model applies for better teleoperation enhancement.

$$zoom = \mathbb{R}(log_{10}(10 + Speed \times \gamma)) \tag{6}$$

As the logarithmic scale compresses the range, our zoom factor prevents the transformed video feed from being

zoomed-in too much for any speed of the vehicle and creates a parity between the teleoperator's expectation and the transformed feed. To prevent the zoom factor to ever be zero, a constant of 10 has been added to the speed factor. After a fair amount of initial testing, we set the value of γ to 0.4 to generate a video feed that offers a reasonably accurate level of future prediction relative to the ground truth while not making the field of view too narrow to lose the necessary situational awareness. The warp block is capable of applying either an affine or a projective transformation to an image frame. We have taken advantage of the projective transformation. For our zoom-in effect transformation, we have used bilinear interpolation while warping the delayed frames. In bilinear interpolation, the new pixel value after transformation is the weighted average of the four nearest pixel values. Along with the predicted future location of the environment, the zoom in and out effect based on the speed of the vehicle offers an immediate visual impact on the delayed feed based on the acceleration and deceleration by the teleoperator. As the teleoperator feels the real-time effect of pressing the acceleration and brake pedal in the delayed feed, it helps to enhance the teleoperation. We have incorporated the zooming-in effect on top of the sliding-only transformation and called the generated feed window, 'the sliding with zooming (SZ)' window throughout the rest of the paper. After incorporating the acceleration and deceleration based zooming effect our assisted feed transformation was generated using Algorithm 2.

Algorithm 2 Algorithm to transform the video feed as a predictive sliding-with-zooming window to reflect the operator control inputs in real-time

Require: $Assisted\ Feed = f(Image\ frame, Wheel\ Rotation, Speed)$

Ensure: $x = x$ coordinate of pixel;
Ensure: $y = y$ coordinate of pixel;
Ensure: $width$ = width of the $Image frame$;
Ensure: $height$ = height of the $Image frame$;
while The system runs **do**
  $x_{initial} = (Original\ frame\ width - Modified\ frame\ width) \div \alpha$;
  $y_{initial} = (Original\ frame\ height - Modified\ frame\ height) \div \beta$;
  $zoom = \mathbb{R}(log_{10}(10 + Speed \times \gamma))$;
  $x_{zoomed} = x_{initial} \times zoom$;
  $y_{zoomed} = y_{initial} \times zoom$;
  $width_{final} = 70\% \times width$;
  $height_{final} = 70\% \times height$;
  $x_{assisted} = x_{zoomed} + x_{initial} \times c \times Wheel\ Rotation \times \mathbb{R}(log_{10}(1+Speed)\,\alpha$;
  $Assisted\ Feed = \text{crop}\ [x_{assisted}\ y_{zoomed}\ width_{final}\ \ height_{final}]$;
end while
Result: Shifting the cropped window according to the wheel rotation, acceleration, deceleration, and speed initialisation

3 Evaluation Methodology

3.1 Model Tuning Through Image Analysis

Pixel based image analysis is not common in the literature for robotic teleoperation enhancement research. Although pixel analysis and comparison is not an ideal way to measure the accuracy of a simulator like the one presented in this work, such techniques may offer some insight into the capability of the simulator and its enhancement techniques, and allow quantitative assessment of the quality of the video frame prediction. Moreover, such quantitative assessment can be used for tuning the performance of the assistive windows prior to the human operator based evaluation. To perform pixel analysis and comparison we have experimented with peak signal to noise ratio (PSNR) [39], the structural similarity index measure (SSIM) [42], and multi-scale SSIM [43].

3.1.1 Image Processing for Pixel Analysis

Before pixel analysis and comparison, we have preprocessed the recorded video feed of the ground truth (gaming video feed from the simulating platform), the raw delayed feed, and the sliding transformation-based assisted delayed feed. All these visual windows were recorded simultaneously using open broadcaster software (OBS) as a single video recording to avoid any disparity among them. We have converted the recorded

video of a teleoperation event into individual frames. After converting them into frames, the three windows have been separated and converted to grayscale as the colour of the remote environment is less of a significant factor for a teleoperator. We have normalised the individual frames of the segregated windows to eliminate variations in brightness resulting from the recording of the screen by the capture device. During the normalisation process, we have used the Sobel-Feldman operator [44] to look for edges in the frames and return a numeric matrix that has been converted to grayscale normalised frames. The normalised non-assisted delayed and the assisted predicted frames are then compared with the respective ground truth frames using PSNR, SSIM, and multi-SSIM evaluation and comparison metrics. Figure 5 shows the different stages of the frame pre-processing prior to pixel analysis and comparison.

3.1.2 Comparison with PSNR

The peak signal to noise ratio (PSNR) [39] expresses the ratio of the maximum strength of a signal and the strength or power of the noise corrupting the signal. PSNR is widely used in image processing research, mainly to measure the image quality after a compression or transformation task relative to the original image. PSNR performs a pixel-by-pixel comparison using Eq. 7.

$$PSNR = 20 \times log_{10}(MAX) - 10 \times log_{10}(MSE) \tag{7}$$

Here, the MAX is the maximum value of the pixel and the MSE is the mean squared error. A higher PSNR value indicates the transformation is closer to the original image frame. For our experimentation, for a teleoperation event, we have compared the assisted (sliding-only) window and non-assisted (delayed window) respectively to the ground

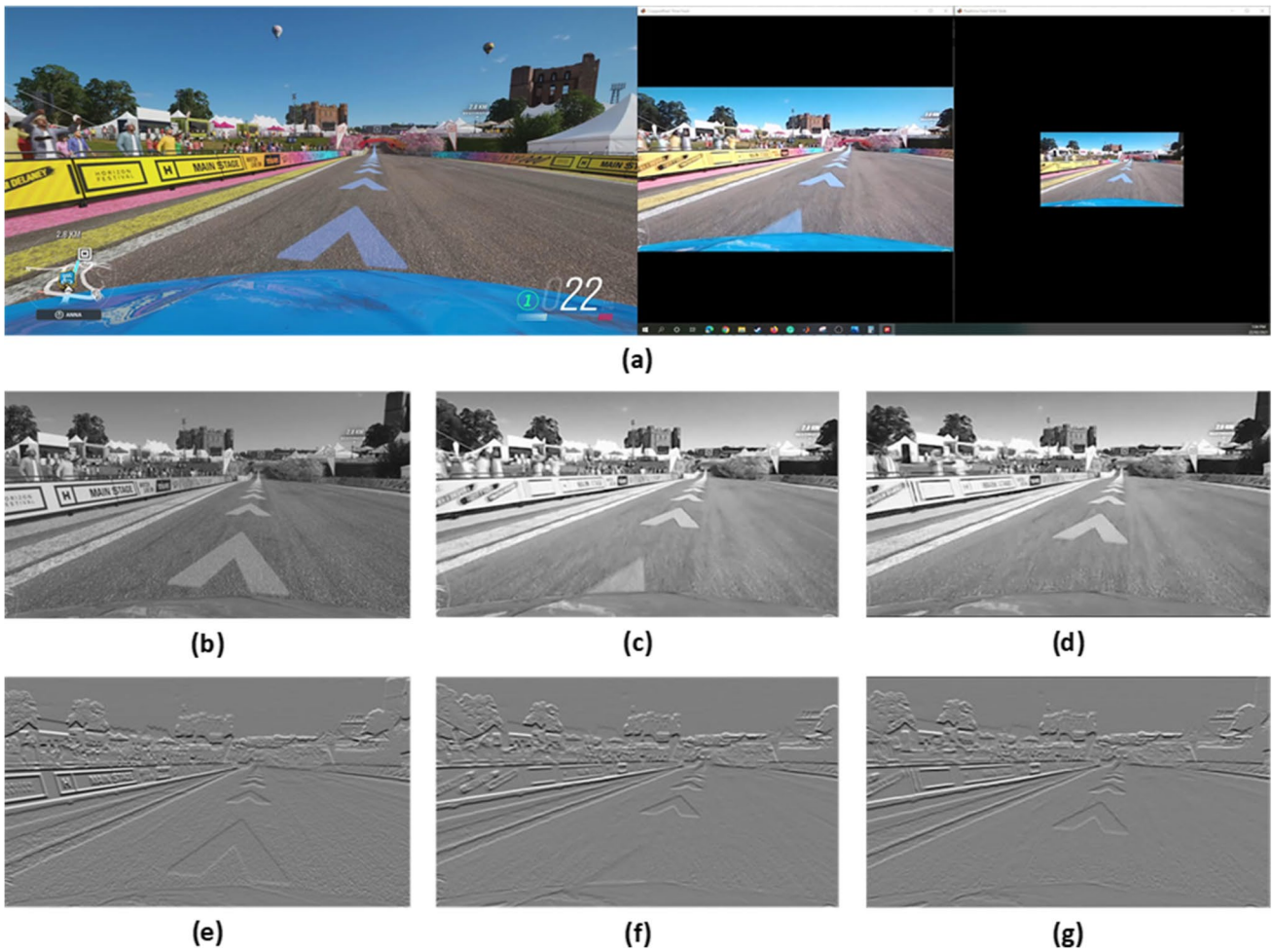

Fig. 5 Example of a teleoperation session: (**a**) Single frame of the recorded video using OBS, separated sections of the ground truth (**b**), non-assisted (**c**), and assisted (**d**) feed, and frames for ground truth (**e**), non-assisted (**f**), and assisted (**g**) feeds normalised with Sobel-Feldman operator

truth or the cropped gaming window to calculate the respective PSNR value and additionally plotted the difference of their PSNR values to demonstrate the comparison as shown in Fig. 7.

3.1.3 Comparison with SSIM

PSNR only considers the values of the pixels and estimates absolute errors. For our use case, the structural difference between the ground truth and delayed feed is more significant than the absolute pixel level differences. The assisted delayed feed is intended to provide a prediction of the future frames via video transformation. Therefore, it is expected that there will be more structural similarity between the assisted windows and the ground truth than that with the non-assisted window. To compare the structural differences we have experimented with the structural similarity index measure (SSIM) to compare the visual feed. SSIM considers the structural information changes along with the change of contrast and luminance to measure the image degradation. If the measure of two different windows x and y having the same size of $N * N$, The SSIM for these two windows would be [43],

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

Here, μ_x is the average of x, μ_y is the average of y, σ_x^2 is the variance of x, σ_y^2 is the variance of y, σ_{xy} is the covariance of x and y, and c_1-c_2 are the variables to stabilise the division. The variance and covariance part of the SSIM algorithm accounts for the structural change among images.

While comparing the SSIM of two different image frames with a ground truth image frame, the higher the SSIM, the closer the frame is to ground truth. For a teleoperation instance, we have measured and compared the SSIM values for our assisted and non assisted feeds. The SSIM comparison and their difference graph are shown in Fig. 8. To visually show the structural difference, maps can be produced comparing the assisted and non-assisted frames with the respective ground truth frames (see Fig. 9).

3.1.4 Comparison with Multi-scale SSIM

In addition to PSNR and SSIM, we have also experimented with the multiscale-SSIM (MS-SSIM). The literature [43, 45–47] suggests MS-SSIM is more robust and performs better for both images and video data. MS-SSIM uses the same SSIM algorithm, however, conducts the operation over multiple scales using a process of multiple sub-sampling stages. In the MS-SSIM process, the system downsamples the images by a factor of 2 before passing them through a low-pass filter. We have compared the same processed video frames using MS-SSIM that was fed to the SSIM and this comparison is shown in Fig. 10. This shows a clearer difference between the assisted and non-assisted frames and so was used for the rest of the work to quantify the quality of the prediction.

3.2 Model Verification Through Human Operator-Based Survey

We have conducted a human-operator-based survey to attain further evaluation and validation of the model. Appropriate approval has been granted by the Human Ethics Committee (HREC) of Edith Cowan University (ECU) prior to any operator data collection. In the survey, we have considered both the quantitative objective measures of the performance and the qualitative subjective measures of the operators' personal experience with the assisted and non-assisted delayed feeds. The detailed information regarding this survey is presented in this section.

3.2.1 Participants

For validation of our model with the assisted teleoperation interfaces, participants were invited to volunteer to participate. The majority of the participants are postgraduate students from different disciplines in the School of Science and the School of Engineering at Edith Cowan University. A total of 10 participants took part in the experiment. The median age of the participants is 32 years. All participants have a driver's license and are regular domestic drivers. We have also collected information on the participants' familiarity with racing games, prior experience of using a driving simulator or console, and any previous experience of operating a robotic vehicle such as driving or teleoperation of drones or remote control cars. Based on their previous experience and familiarity, participants can rate themselves to a maximum of 15 and a minimum of zero. Based on this familiarity index, we found that our participants are normally distributed where their mean familiarity index is 6.6. This implies that if the assistive visual interfaces enhance the teleoperation for these participants, they will enhance teleoperation for even novice teleoperators who have little previous experience of teleoperating ground vehicles. Figure 6 represents the data distribution of the participants based on their familiarity with the teleoperation task. Before taking part in the experiment, the participants were asked about whether they are under the effect of alcohol or any medication that could affect a standard teleoperation experience.

3.2.2 Experimental Setup and Task

For our experiment, all the participants were asked to attend the teleoperation session individually, not in a group. This has ensured the participants do not have any prior knowledge

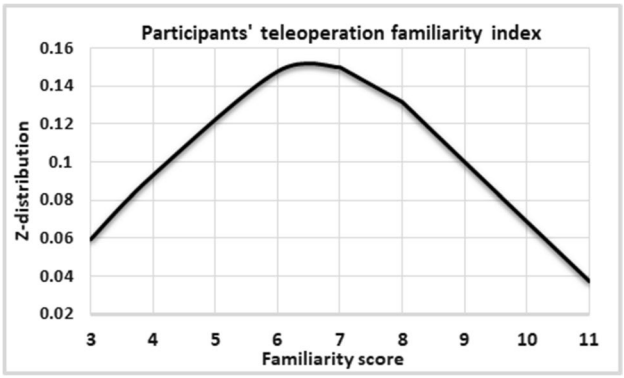

Fig. 6 Survey participants' familiarity index and their distribution

about the experimental setup or the procedure. Once the participants entered the experiment venue, they were shown the simulator and briefed so that they entirely understood the task they were to complete. The participants were provided with a chair in front of the monitors that work as the interfaces through which the visual feedback of the remote environment is received. Out of two, one of the monitors provides the real-time 60 fps full high definition (FHD) 1080p gaming feed that offers a near-photorealistic real-world visual impression and works as the ground truth, and the other one represents the delayed feeds from the simulating platform. Once the participants had been briefed, they were provided access to the control devices (steering wheel and acceleration and brake pedals).

All the participants were given the task to drive a car as a teleoperated ground vehicle through the Forza Horizon 4 gaming environment using the commercial steering wheel and acceleration and brake pedals shown in the system diagram of Fig. 2. Before starting to collect data, the participants were allowed to practise driving on the different tracks to get used to the system. The data collection started only when the participants felt comfortable and willing to start the experiment.

3.2.3 Manipulating Variables

While designing the human operators' performance based evaluation procedure for the video transformation based enhancement technique, the primary manipulating variables we have considered are the level of difficulty of the remote environment, the amount of latency added to the teleoperation loop, the quality of the situational awareness provided through the visual feed to the operator, and the additional assistance offered through our enhancement techniques. For our research, the participants were asked to drive the vehicle on two different tracks: one on-road and one off-road, four times each. The on-road track is approximately 2 km and the off-road track is approximately 1.5 km. The participants drove the vehicle in each track using 4 different visual feeds: the ground truth FHD feed, the non-assisted delayed feed, the assisted feed with sliding-only window, and the assisted feed with sliding and zooming effect window. However, except for the ground truth feed, the participants were not made aware of the capability and features of each of the video feed windows. Except for the ground truth, for all the sessions the amount of latency inserted into the delayed feeds was the same (900 ms). All the participants were directed to start and end the sessions in a track from the same starting and ending points to maintain the same direction through the same path to avoid the bias of opposite directional difficulty. Further, in the sessions where the participants drove based on the delayed feeds, the monitor displaying the real-time ground truth feed was not visible to them.

3.2.4 Quantitative Performance Measurement

We have prepared a quantitative objective performance measurement metric to measure the participants' performance during the teleoperation sessions. Measurement of the task completion time and teleoperation performance scores are common parameters when experimenting with different forms of teleoperation experience [37, 38, 48, 49]. As a scoring mechanism, we have considered the number of times the participants lose control and the teleoperated vehicle goes out of the defined track and the number of times the vehicle oscillates due to overcorrection on the track. The higher the counts, the poorer the performance for any session by the operator. We have also counted the time taken to complete a single session on a selected track, as well as calculating the average speed of the vehicle on a teleoperation session and using it as a performance measurement parameter. Increasing latency tends to reduce the driving speed of the vehicle for a given operator, therefore, the lower the average speed, the poorer the performance of the operation was considered for a particular session. The above parameters are considered to provide us with enough evidence to measure the objective performance of a teleoperator for both the assistive and non-assistive video feed-based teleoperation sessions.

3.2.5 Qualitative Performance Measurement

Additional to the quantitative performance measurement we have asked all the participants to rate their experience after every session to measure the qualitative aspect of the system and the effectiveness of the transformed visual feedback windows. The authors believe the outcome of the participants' survey would reinforce the outcome of the quantitative objective performance measurement and prove the robustness of the model based on the participants' experience. The participants were asked to rate six different aspects of their experience on a scale of 0 to 5, where 0 implies the

Table 2 Pixel based comparison of a random sample teleoperation simulation session

Parameter	Value
PSNR for assisted feed	20.30
PSNR for non-assisted feed	20.12
SSIM for assisted feed	0.49
SSIM for non-assisted feed	0.48
Multi-SSIM for assisted feed	0.62
Multi-SSIM for non-assisted feed	0.59

lowest, hardest, or negative experiences, and 5 implies easiest, most comfortable, or most positive experiences. In this survey, the participants were asked to provide feedback on the visibility of the remote environment through the user interface, the impact of latency and intermittency while teleoperating, controllability of the vehicle during the session, level of comfort in the speed they managed to drive and the environmental challenge of the track they were driving the vehicle on. The authors believe the answers to the survey questions provide insight into the impact of communication latency and intermittency on situational awareness while teleoperating a ground vehicle. The operator assessed experience of controllability can also be cross-examined relative to the overcorrection-induced oscillations and out-of-track incidents experienced by the participants. The survey also intends to draw a relation between the level of challenging environment and teleoperation experience.

4 Results

4.1 Pixel-Based Image Analysis Outcome

The principle evaluation of the performance of our simulator and video transformation-based enhancement technique has been performed through the human operator survey. However, we have used the pixel-based image analysis techniques such as PSNR, SSIM, and multi-SSIM for the purposes of fine-tuning our video transformation algorithm prior to the operator survey. Table 2 presents the PSNR, SSIM, and multi-SSIM values of the extracted frames of two windows from a recorded video of a random teleoperation session: one with the assisted feed, which is the delayed feed transformed using our video transformation algorithm, and another that is simply the delayed feed without any transformation to assist the operator. From table 2 we can see that for all three parameters, the mean values for the assisted feed are higher than the non-assisted feed. For the non-assisted feed, the mean PSNR, SSIM, and multi-SSIM values are 20.12, 0.48, and 1.59. These values improved to 20.30, 0.49, and 0.62 respectively for the assisted feed. The PSNR, SSIM, and multi-SSIM values for the assisted and non-assisted feeds are plotted in Figs. 7, 8, and 10. The mean values and the graphs indicate that according to the pixel and structural similarity indices the transformed assisted feed is closer to the ground truth than the delayed non-assisted feed. Figure 9 shows the SSIM difference maps of an assisted frame and a non-assisted frame with their ground truth frames. More black and grey areas in a map mean a higher difference in the frame compared to the ground truth.

4.2 Human Operator-Based Survey Results

4.2.1 Quantitative Performance Measurement

Table 3 presents a summary of the quantitative outcome of the human operator performances for both the on-road and off-road scenarios. The average time taken for participants to complete a 2 km long on-road track looking into the real-time high definition (HD) visual feedback was 171.1 s with an average speed of 46.25 km/h. For the 1.5 km long off-road track the participants spent an average of 125.8 s with an average speed of 44.20 km/h. This implies that the participants faced similar levels of challenge for both on-road and off-road tracks and were able to drive at almost the same speed when using the real-time (non-delayed) visual feedback. We have considered these values with HD real-time visual feed as the ground truth for our analysis.

During the driving sessions when the latency has been applied and no assistive technique is in place, the average time to complete a 2 km on-road track increased by 205.67% (i.e. 523 - 171.1 s). For the 1.5 km long off-road sessions it increased by 147.14% (310.9 s). Compared to the ground truth the average speed had a 69.71% drop (to 14 km/h) and for off-road, it had a 59.47% drop (to 17.91 km/h). For the off-road, the decrease is a little less, this may have been due to the fact that the on-road sessions included AI controlled traffic creating more unpredictability on the track. From the above numbers, we can make a general statement that for a 900 ms delay the task completion time increases by around 150-200% for a ground vehicle teleoperation scenario with reasonable ground speed. When the participants used the sliding-only window-based visual feedback, their performance increased by an appreciable amount. The average time to complete the 2 km on-road track was 389.5 s which is 25.53% less than the delayed-only feed performance. For the off-road track, the required time was reduced by 8.14% relative to the non-assisted feed. Similarly, the average speed increased by 36.37% and 8.75% for on-road and off-road tracks respectively. Based on the task completion time and average speed metrics, the assisted display with sliding and zooming combined also performed better than the delayed feed only and further enhanced the teleoperation experience.

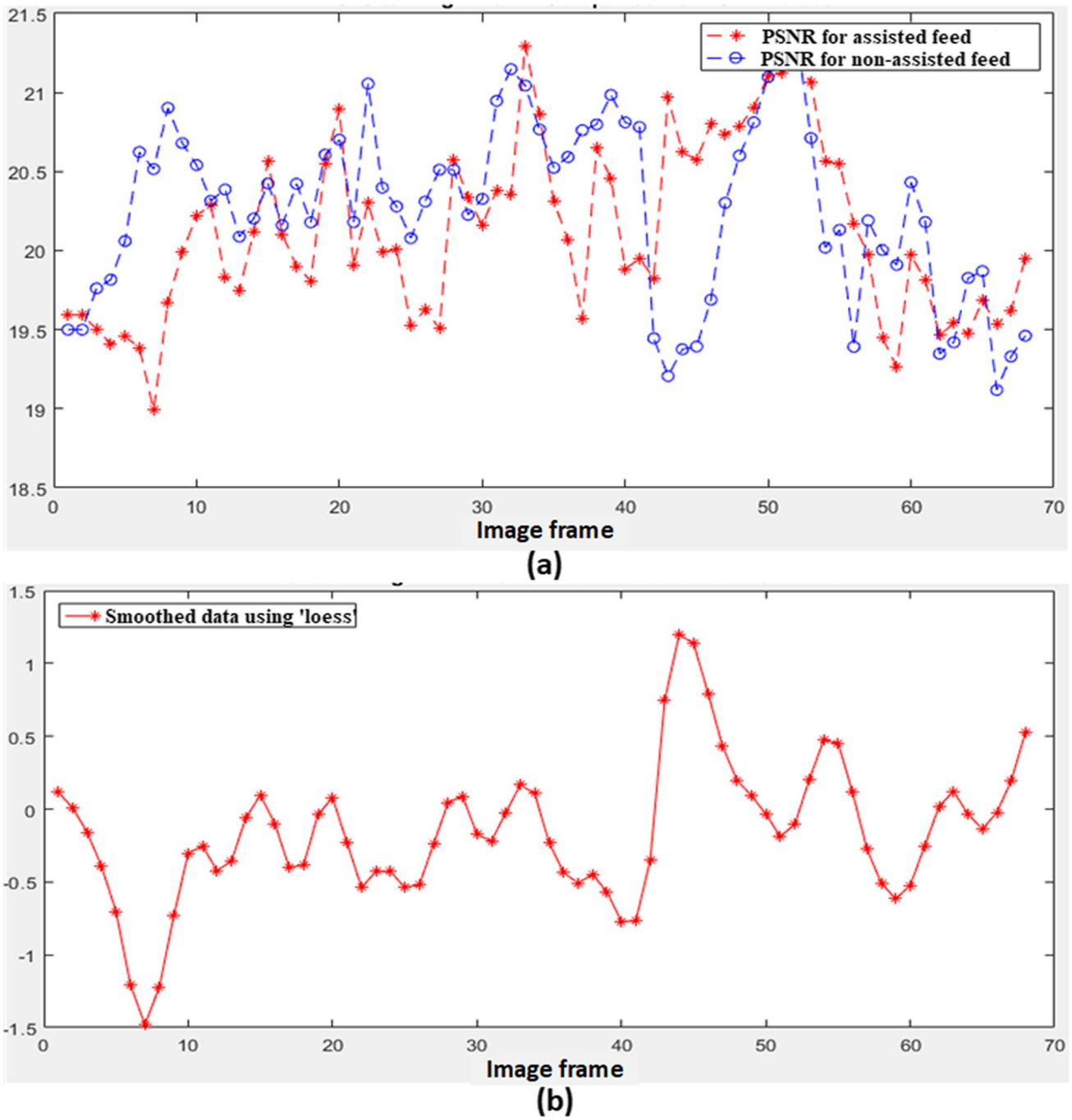

Fig. 7 Performance comparison of assisted and non-assisted feeds for a teleoperation session: (**a**) comparison of PSNR values for assisted (red) and non-assisted (blue) frames (**b**) curve showing the differences of PSNR values to visualise the comparison easily (best seen in colour)

With the sliding and zooming transformation, the on-road average time requirement was 408.9 s and the off-road time requirement was 285.6 s. These completion times were 21.82% and 11.74% less than the non-assisted delayed feed respectively. The average speed also increased by 28.75% and 13.53% for the on-road and off-road tracks respectively.

The other two crucial parameters of the performance evaluation are the count of oscillations and out-of-track incidents during the teleoperation sessions. As the count of oscillations can be a somewhat subjective judgement, we have considered the oscillations only when the vehicle crossed the road edges partially due to the impact of over-correction and ignored any oscillations where the impact

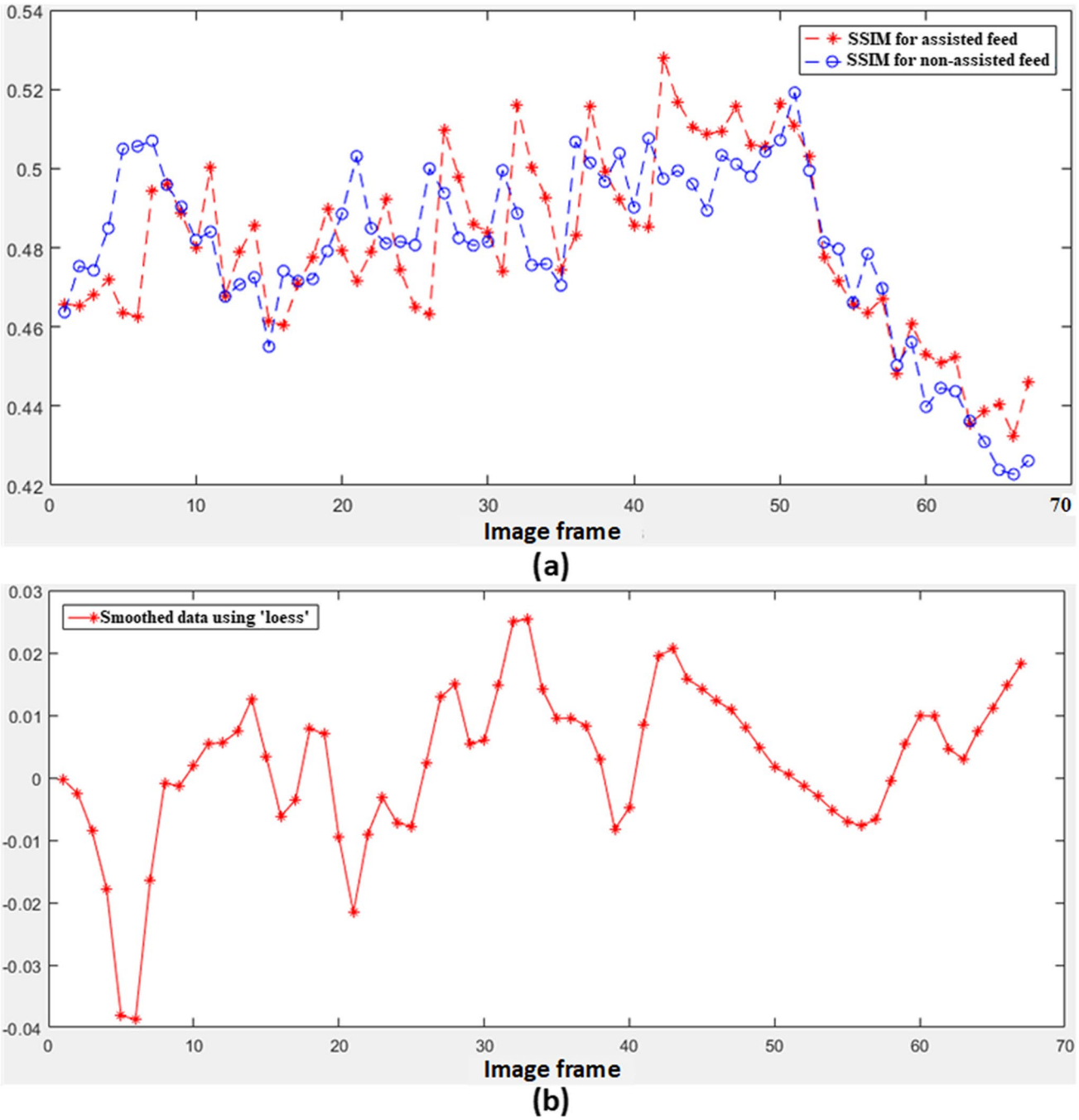

Fig. 8 For a teleoperation session: (**a**) comparison of SSIM values for assisted (red) and non-assisted (blue) frames, and (**b**) curve showing the differences of SSIM values to visualise the comparison easily (best seen in colour)

of overcorrection was minor and did not affect the vehicle enough for it to cross the road or terrain edge. From Table 3, the total number of oscillations that occurred during all the sessions that use the ground truth visual feedback was 17 which is an average of 1.7 times (count) for each participant for the on-road track. For the off-road, the total for the ground truth was 11 which is 1.1 times per participant. When latency was added, the number of oscillations increased to an average of 8.6 counts per person and a total of 86 counts for the on-road and a total of 90 counts for the off-road track. When the participants were aided with our sliding-only window feedback, the oscillation counts dropped significantly. For the on-road track the number of oscillations dropped 66.28% to 29 counts and for the off-road track, the number dropped 41.11% to 53 counts. For the sliding with zooming feed the outcome is even more encouraging. The number of

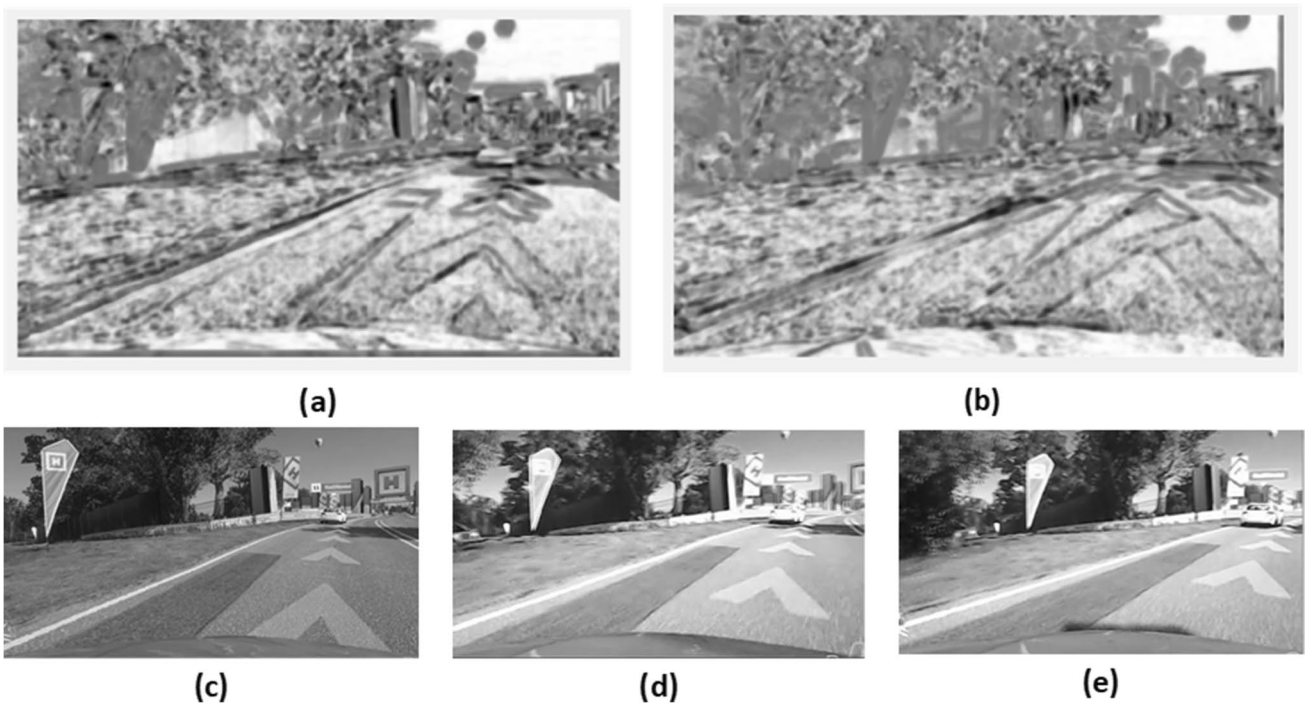

Fig. 9 SSIM maps ((**a**) difference map of assisted frame and ground truth frame; (**b**) difference map of non-assisted frame and ground truth frame) for a single assisted (**d**) and non-assisted (**e**) frames with the respective ground truth (**c**) frame. The more black the map is (**b**), more the difference compared to the ground truth

oscillations dropped further to a total of 21 counts for the on-road track, which is a 75.58% drop, and to 43 counts for the off-road track, which is a 52.2% drop.

A similar outcome has been observed for the out-of-track incidents as well. When the participants lost control and the vehicles fully crossed the edge to a point that the whole vehicle is out of the road or terrain, we have considered those incidents as out-of-track. For the FHD real-time visual feed, the total number of out-of-track incidents was expected to be lower. For the on-road track there was only one incident where the vehicle went out of the track. For the off-road track, there were no out-of-track incidents for the real-time FHD video feed. However, for the raw delayed feed the number climbed considerably for both the on-road and off-road tracks. For the on-road, the number of out-of-track incidents was 37 and for the off-road track, it was 35 in total. When driving with the assisted visual feeds the number dropped drastically. Using the sliding-only window the out-of-track incidents dropped to only 5 for the on-road and

Table 3 Quantitative survey outcome

Parameter	Ground truth (GT)	Without assistance (WA)	Sliding only (SO)	Sliding and zooming (SZ)	Relative change GT-WA	Relative change SO-WA	Relative change SZ-WA
On-Road							
Average time (s)	171.1	523	389.5	408.9	205.67%	-25.53%	-21.82%
Average speed (km/h)	46.25	14.01	19.10	18.04	-69.71%	36.37%	28.75%
Total oscillation (count)	17	86	29	21	405.88%	-66.28%	-75.58%
Total out of track (count)	1	37	5	8	3600%	-86.49%	-78.0%
Off-Road							
Average time (s)	125.8	310.9	285.6	274.4	147.14%	-8.14%	-11.74%
Average speed (km/h)	44.20	17.91	19.48	20.34	-59.47%	8.75%	13.53%
Total oscillation (count)	11	90	53	43	718.18%	-41.11%	-52.22%
Total out of track (count)	0	35	7	7	-	-80.0%	-80.0%

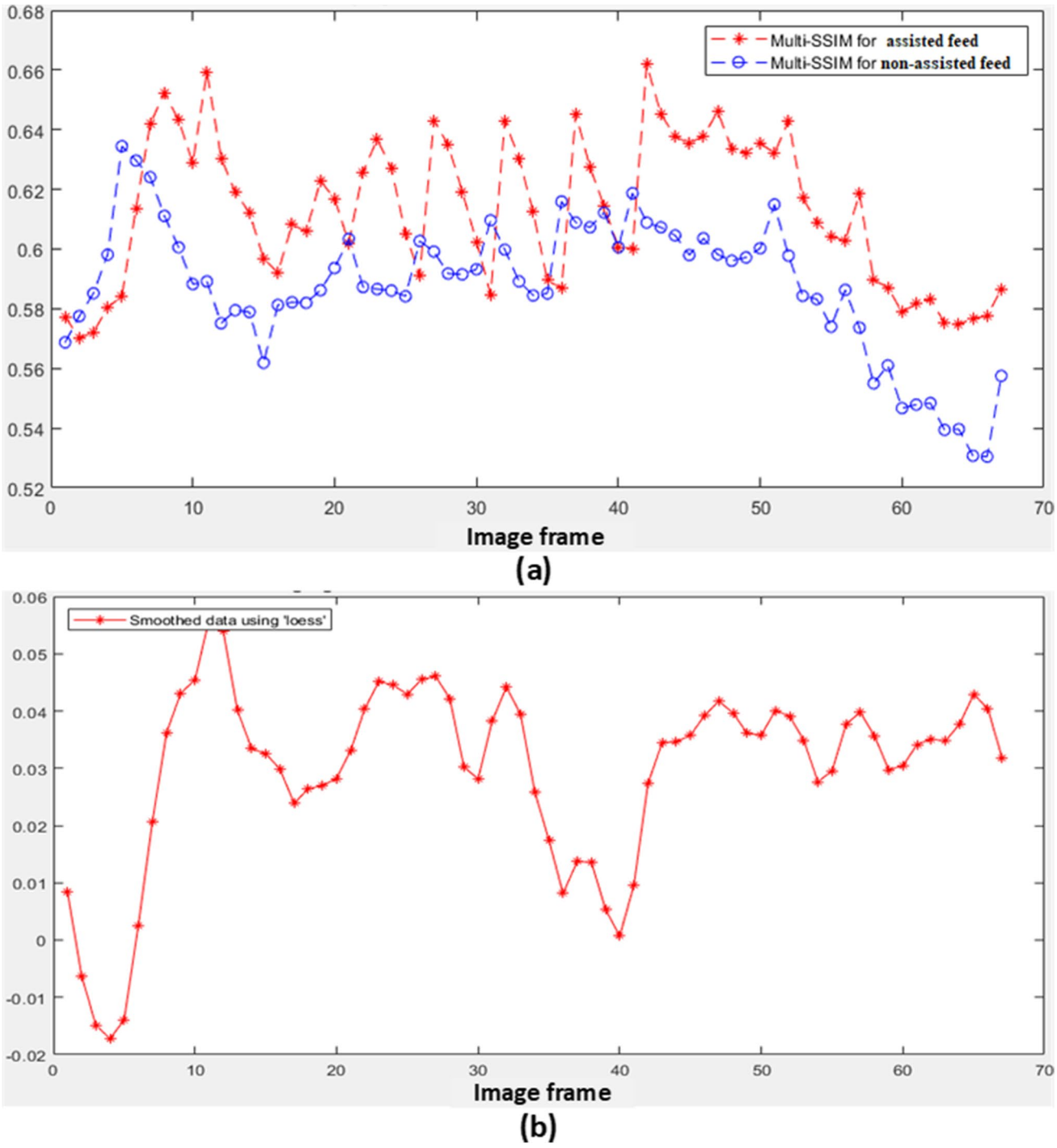

Fig. 10 For a teleoperation session (**a**) comparison of MS-SSIM values for assisted (red) and non-assisted (blue) frames (**b**) Curve showing the differences of MS-SSIM values to visualise the comparison easily (best seen in colour)

7 for the off-road track. Using the sliding and zooming feed, the decrease is similar, with only 8 incidents for the on-road and 7 for the off-road track. Thus, out-of-track incidents dropped around 80% for both the on-road and off-road tracks using both the sliding-only and sliding with zooming windows.

4.2.2 Qualitative Performance Measurement

To get feedback about the qualitative aspect of the different visual feeds and the participants' experience, all the

participants were asked to rate every session regarding the visual quality, speed and comfort, impact of latency, intermittency, and the controllability of the vehicle. During the experimental teleoperation sessions, the participants were unaware of the quantitative parameters that were used to measure the performance of their driving sessions. Therefore, their feedback is free from the biases of the quantitative parameters. Table 4 presents the summary of the participants' feedback on the different qualitative aspects. The participants rate the experience with the ground truth HD real-time feed as the best with a total score for the convenience of 20.2 for on-road and 24.2 for off-road out of a maximum possible of 30. This high score is expected for non-delayed teleoperation through the HD video feed. However, our point of interest is the comparison of the qualitative scores between the delayed feeds: without assistance, sliding-only, and sliding with zooming.

For the on-road track, according to the participants' feedback, the visibility of the driving track (2.6), and the controllability of the vehicle (3) are better for the sliding only assisted window. The participants found it easier to maneuver at a higher speed while using the sliding and zooming assisted window (2.7). The impact of latency was also deemed milder (2.3; a higher score means lower impact) while using this assisted window. The participants faced less cognitive challenge while teleoperating using the assisted windows (mean score of 2.70 for sliding and zooming and 2.60 for sliding only). According to the total qualitative score, the participants found both the sliding-only assisted window (15.3) and sliding and zooming assisted window (14.6) better than the non-assisted window (13.4).

The participants found it easier to perform teleoperation in the off-road track environment using the assisted windows similar to the on-road track. For all the aspects, visibility, controllability, cognitive challenge, impacts of latency, and intermittency, for both the sliding only, and sliding and zooming assisted windows are preferable to the operators compared to the non-assisted window. The total qualitative score given by the operators is 16.1 for the assisted windows, whereas, the score is 12.2 for the non-assisted window.

5 Discussion

5.1 Pixel-Based Analysis

The comparison of PSNR (Fig. 7(a)) is somewhat difficult to interpret as the PSNR values are very close for image frames from both the assisted and non-assisted windows. Figure 7 (b) helps to interpret and visualise the difference *(PSNR for assisted frames - PSNR for non-assisted frames)* in a clearer way. For most of the image frames, the difference is positive. This implies, for pixel-wise comparison, the assisted delayed feed is closer to the ground truth than the non-assisted delayed feed. While experimenting with PSNR, we found that when the speed of the vehicle increases the PSNR for the assisted feed increases in comparison to the non-assisted feed as the algorithm we developed used speed as a factor. We also found out that increasing the turning angle increases the PSNR for the assisted window. The assisted feed slides left or right in real-time according to the turning of the vehicle resulting from the turning of the steering

Table 4 Summary of Qualitative survey outcomes (In a scale of 0 to 5)

Parameter	Ground truth	Without assistance	Sliding only	Sliding and zooming
On-Road				
Mean feedback on visibility	4.5	2.5	2.6	2.3
Mean comfort on driving speed	3.6	2.1	2.2	2.7
Mean impact of latency (lower is high impact)	4.2	2	2.2	2.3
Mean impact of intermittency (lower is high impact)	4.2	2	2.7	2.2
Controllability	3.7	2.5	3	2.4
Mean challenge felt (lower is higher challenge)	3.90	2.30	2.60	2.70
Total qualitative score	20.2	13.4	15.3	14.6
Off-Road				
Mean feedback on visibility	4.5	2.39	3.01	2.9
Mean comfort on driving speed	3.8	1.7	2.6	2.5
Mean impact of latency (lower is high impact)	4.1	1.6	1.8	2.5
Mean impact of intermittency (lower is high impact)	4.5	1.9	2.9	2.1
Controllability	4	1.8	2.9	3
Mean challenge felt (lower is higher challenge)	3.3	2.7	2.8	2.9
Total qualitative score	24.2	12.2	16.1	16.1

wheel. This proves that the assisted feed works to provide a closer prediction to a future frame from a delayed feed.

While experimenting with the SSIM based analysis of the assisted and non-assisted frames, we found that the assisted feed provides slightly better SSIM values than the non-assisted feed. Therefore curves in Fig. 8(a) are more or less similar to that of the PSNR comparison graph. However, the difference curve Fig. 8(b) of assisted and non-assisted SSIM values shows a graph having most of the values positive. This implies that the assisted feed is closer to the ground truth. Figure 9 shows the difference map of an assisted and non-assisted frame compared with the ground truth frame. More black and grey regions are shown on the non-assisted difference map. That also implies that the assisted video feed is closer to the ground truth.

While applying the MS-SSIM on the same image data that were used for comparing with SSIM, we see a much clearer and more significant difference between the MS-SSIM values for assisted and non-assisted image frames (Fig. 10(a)). Figure 10 (b) shows the relative differences between the assisted and non-assisted feeds and the values are higher than that of SSIM, giving clearer differentiation. Therefore, we have used MS-SSIM to experiment with our video transformation algorithm and fine-tune the system to achieve the best performance for teleoperation enhancement. We have found that for Eq. 5, the value of $c = 5.5$ provides the maximum system performance in terms of MS-SSIM values. To the best of our knowledge, using pixel and structural similarity indices to evaluate and fine-tune a teleoperation simulator is a new approach in the teleoperation research domain.

5.2 Operator Survey-Based Quantitative Analysis

We assume that the performance of the participants during the teleoperation sessions are entirely dependent upon the 2D visual feedback they are provided with, as the operators have access to no other forms of feedback such as audio, text, signs, etc. Prior to starting data collection, the participants were offered sufficient practice sessions and time to get used to the system. While teleoperating through a delayed system, the operators may, over time, get more used to the latency and perform better in the later sessions. Also, for a longer set of teleoperation sessions, their performance may drop due to high cognitive workload, loss of human eye-hand coordination, fatigue, and tiredness. To minimise the impacts of these external factors, we have randomly changed the order of the sessions using the non-assisted visual feedback, the assisted with sliding-only window, and the assisted with sliding and zooming window as the visual feedback. The operators were not informed about which of the feeds they were going to use during each session.

Figure 11(a) and (b) plot the required time and speed for all of the participants during on-road track sessions and Fig. 12(a) and (b) plot the same graphs for the off-road track for all the different visual feeds. The graphs demonstrate that the participants performed better in terms of task completion time and speed for both on-road and off-road tracks using the assisted visual feedback. While comparing the performance graphs between the sliding-only window and the sliding and zooming window, in the on-road track, four participants performed better with the sliding and zooming window and six participants performed better with the sliding-only window. However, in the off-road track, the overwhelming majority of the participants performed better in terms of speed and time with the sliding and zooming window. Our system collects the vehicle speed information from the ground truth visual feedback and uses OCR to convert the digits that are displayed in a white colour on the screen. However, at times when the white speed digits overlap the white lane marks and borders of the road, the OCR generates irregular values that creates an irregular and unexpected zooming effect. Therefore, the sliding window with zooming feed had some irregular flickering incidents and this affected the completion time and speed. As the off-road track does not have any lane markings or road edge borders this issue did not occur and all participants performed better with the feed that included the zooming window.

Figure 11(c) and (d) plot counts of overcorrection related oscillations and out of track incidents for all the participants during the on-road track sessions and Fig. 12(c) and (d) plot the same graphs for the off-road track for all of the different visual feeds. From these graphs and Table 3 it can be claimed that the assisted feeds definitely enhanced the teleoperation by significantly reducing the effect of overcorrection and decreasing the number of oscillations and out-of-track incidents for both the on-road and off-road tracks. Further, based on the oscillation graphs and figures from the table, it can be seen that the sliding-and-zooming window has enhanced teleoperation to a greater extent and would make the teleoperation task safer in a real-life, relatively high-speed ground vehicle teleoperation scenario.

5.3 Operator Survey-Based Qualitative Analysis

From Table 4 we can see, the participants rated the sliding-only window highest in terms of visibility of the remote environment through the transformed feed. All the delayed feeds of our system have the same pixel dimension and frame rate. However, the sliding-only window gives the operator an impression of shifts in point of view according to the turning of the steering wheel and the expected future direction of the ground vehicle. Therefore, along with enhancing the teleoperation experience, the operators' expectation of the visual direction of the remote environment matched with the sliding window's point of view. Although, for the

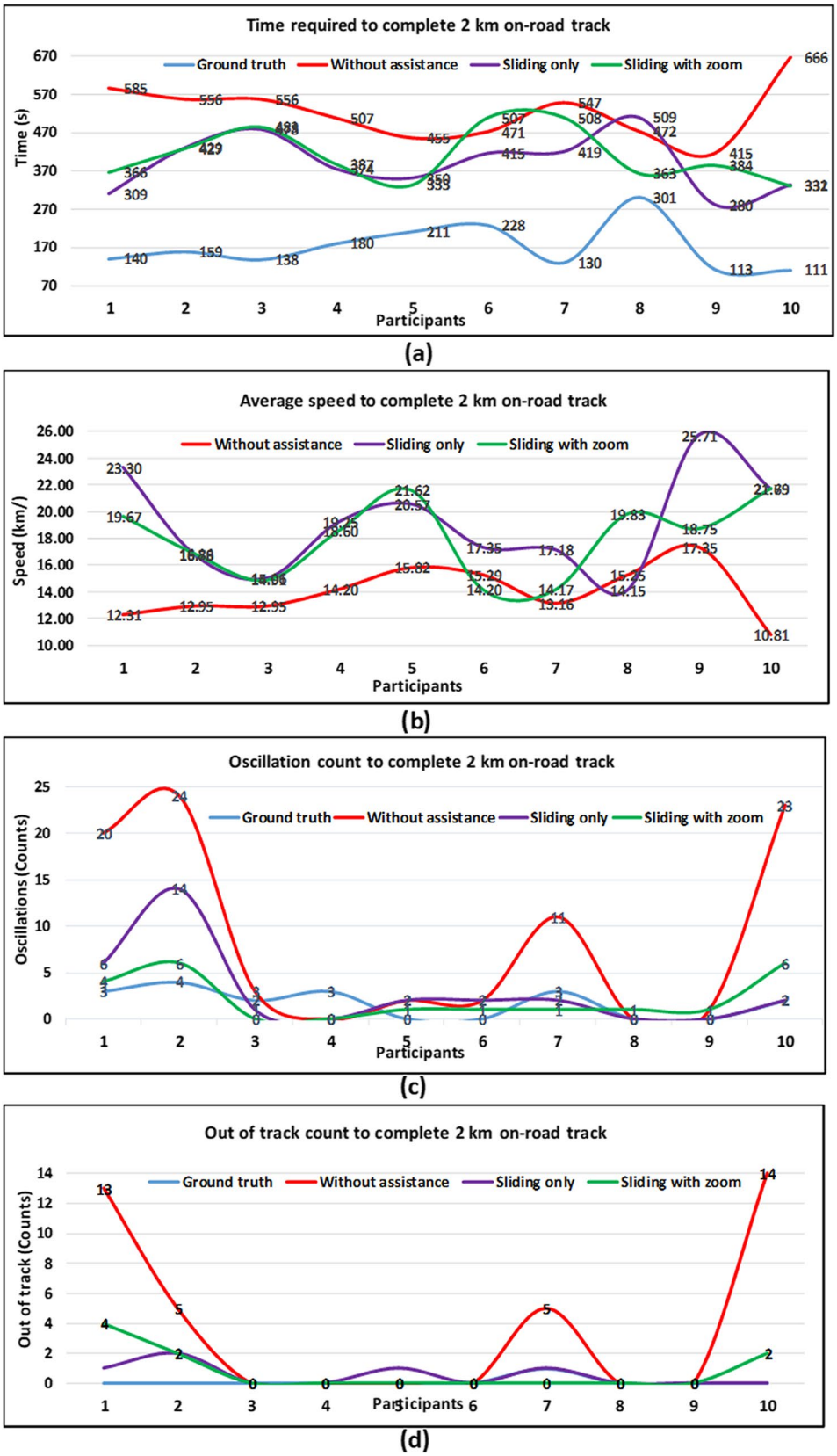

Fig. 11 Performance of each participant using different 2D visual interfaces on the 2 km on-road track: (**a**) time required to complete the track, (**b**) average speed of the vehicle during different sessions, (**c**) counts of oscillations due to overcorrection, (**d**) counts of out of track incident (best seen in colour)

on-road track, participants rated the sliding with zooming feed lower, for the off-road track the participants rated this window almost as high as the sliding-only feed. The likely reason has already been described in the previous subsection- the OCR produces anomalous values when the white digits of the speedometer overlaps with the white road markings, resulting in the transformed feed flickering irregularly at times for on-road tracks. As a result, the visibility feedback from the participants is lower for the on-road track with the sliding with zoom window. However, for off-road tracks, the rating is better.

From Table 4 we observe, that for both, the on-road and off-road tracks, the participants felt more comfortable with the speed they were driving while teleoperating when looking into the remote environment through the sliding with zooming feed. As the participants' average driving speed was higher with the sliding and zooming feed and also the participants were more comfortable using this feed, it can be claimed that the sliding and zooming transformation enhanced the teleoperation both in terms of the quantitative and the qualitative aspect of the driving experience. Moreover, according to the participants' feedback, the impact of latency was lowest for the sliding and zooming feed and worst for the raw delayed feed without any assistance. Furthermore, for both the on-road and the off-road tracks, the participants found the remote environment less challenging when teleoperating based on the sliding and zooming feed. There is no additional element of intermittency that was added to the video transformation. However, to replicate a real teleoperation video feed we reduced the frame rate and resolution, which created some intermittency effects equally for all the delayed feeds. However, the impact of intermittency was felt to be lowest for the sliding-only feed. Although the participants rated the intermittency as being higher (low score means higher impact) for the sliding with zooming feed than for the sliding-only feed, the score is still better than that of the raw delayed feed without any assistance. The OCR-induced speedometer-related flickering negatively impacted the participants' feedback on intermittency. According to the qualitative and quantitative aspects, especially for the off-road track teleoperation sessions, the participants overwhelmingly found the sliding with zooming feed makes the teleoperation easier than the sliding-only or the delayed non-assisted feed.

Besides the above discussions, none of the prior 2D predictive feedback based teleoperation enhancement approaches mentioned in Table 1 designed a universal simulator with controllable latency that can be used to simulate ground vehicle teleoperation (easily transferable to other robotic vehicle types) with a varying range of speed, latency, frame-rate (intermittency), and environmental difficulty. Moreover, the capability of collecting and saving synchronised image, video, and control signal data has made the simulator suitable for neural network and AI-based teleoperation enhancement research. Some of the previous approaches used forms of simple image transformation e.g. [12, 25], however, our video transformation algorithm is much more complex, it accommodates high-speed maneuvers, adjusts based on vehicle speed, incorporates inputs from control signals such as steering, acceleration, and deceleration, and offers a predictive 2D visual feed. The results indicate that our video transformation-based enhancement techniques are effective and significantly reduce task completion time, overcorrection-related oscillations, and out-of-track incidents. Therefore, we are confident that our approach represents a novel and effective teleoperation enhancement system.

6 Conclusion

This research work focuses on enhancing the experience and effectiveness of teleoperation, especially for long-distance, high-latency ground vehicles operating at a reasonable ground speed both in on-road and off-road terrains, using only 2D visual feedback to achieve the entire situational awareness of the remote environment. This type of teleoperation task is potentially dangerous, and collection of real-world control and visual data is difficult and not without some level of risk. To avoid this, for this research a system has been developed that can be used to simulate high-speed ground vehicle teleoperation tasks with configurable latency. This model can be used both for simulation and testing, and for data collection that can be used for future AI-based teleoperation enhancement research. In this research, we have also designed and evaluated two video transformation-based assistive visual interfaces (sliding-only and sliding with zooming) to enhance the teleoperation. This research implemented pixel-based analysis techniques such as PSNR, SSIM, and Multi-SSIM to evaluate and fine-tune the video transformation-based assisted interfaces. Overall performance evaluation of the model and comparative analysis of the assistive interfaces has been performed via a human operator-based survey. The survey results indicated that for a ground vehicle teleoperated at a reasonable ground speed with a latency of 900 ms, task completion time was increased by up to 200%. A delay of 1200 ms was found to impact the teleoperation to the extent that the overcorrection and oscillations make a full track completion almost impossible due to frequent losses of control. The survey results also showed that using our sliding-only visual transformation reduced task completion times by up to 25.53% and our sliding with zooming transformation-based techniques completion time was reduced by up to 21.82%. In terms of overcorrection-related oscillation reduction, the sliding with zooming transformation performed better and reduced the over-correction and oscillation by a large margin of up to 75.58%. Therefore, the sliding window with zooming transformation has been shown to effectively enhance

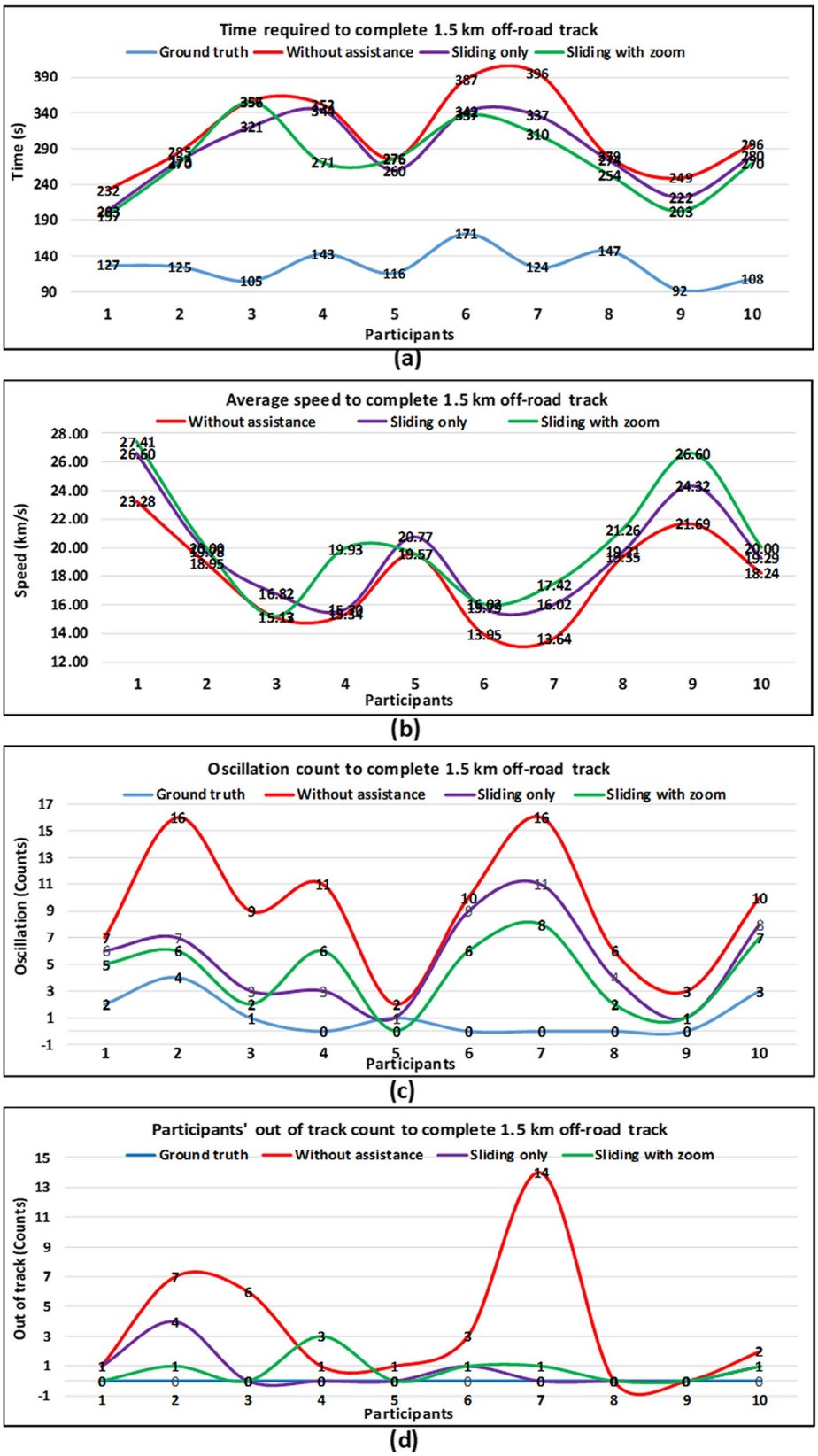

Fig. 12 Performance of each participant using different 2D visual interfaces on the 1.5 km off-road track: (**a**) time required to complete the track, (**b**) average speed of the vehicle during different sessions, (**c**) counts of oscillations due to overcorrection, (**d**) counts of out of track incidents (best seen in colour)

teleoperation by reducing both task completion time and overcorrection. This model is specifically suited for long-distance, high-latency, high-speed teleoperation tasks. However, it can be tuned and be adopted to any other teleoperation scenario. Furthermore, our designed simulator can be used to train operators to drive in higher latency situations. The simplified video transformation methods presented in this paper show significant potential to enhance teleoperation in high latency environments.To further validate the use of synthetically generated video feeds for latency mitigation, testing will need to be conducted with larger numbers of operators. For this research, the number of participants included for the human operator survey was limited due to the significant amount of time required for each participant to complete all of the required experimental driving sessions (around 2 hours for each participant). Many future refinements are possible to further improve the quality of the prediction and the overall standard of visual feedback presented to an operator and further research work will be carried out to explore these additional enhancements. In addition to conventional techniques, AI and deep learning-based future frame prediction and synthetic future video feed generation is an arena we plan to explore in our future research into teleoperation enhancement.

Appendix: Incorporating Control Input Delay with Visual Latency

Latency into the simulation platform can be achieved either by adding delay to the visual feed or by delaying the control input to the simulation platform. The authors of the paper are confident that the impacts of either of these delay factors, or a combination of them, are perceived as the same from the perspective of a teleoperator. This appendix discusses the methods through which the control input delay can be achieved and the impacts on a teleoperator.

In the main sections of the paper, latency was achieved purely by delaying the visual feed. To achieve the visual delay, a single computer unit was used to connect the simulating platform with controller devices, visual feed receiver camera, and the latency control and teleoperation enhancement Simulink® model. However, to incorporate control input delay along with the visual delay, we have used another personal computer (PC) unit, referred to as PC-2 in the rest of this section. The system diagram of the modified teleoperation simulation system has been provided in Fig. 13. In this modified system, the controller devices (steering wheel, brake and acceleration pedals) are connected to the PC-2. The vehicle simulation platform (game engine), the visual feed receiver, and teleoperation enhancement model is hosted by the previously used PC, named PC-1 hereafter. The two computers can be connected either by local area network (LAN), wide area network (WAN), or Internet connection. For our case, the computers were connected to the university LAN.

Aside from these new amendments, the rest of the system is as described in Section 3.

To add latency between the two computers we have used a third party software called Clumsy. Clumsy uses the Windows Packet Divert library to stop, capture, lag, drop, or tamper with packets on a living network. Any amount of system to system latency is achievable using this tool. An example of varied latency between two computers using Clumsy is shown in Fig. 14. To connect and receive control input from PC-2 to PC-1 we have used another third party software

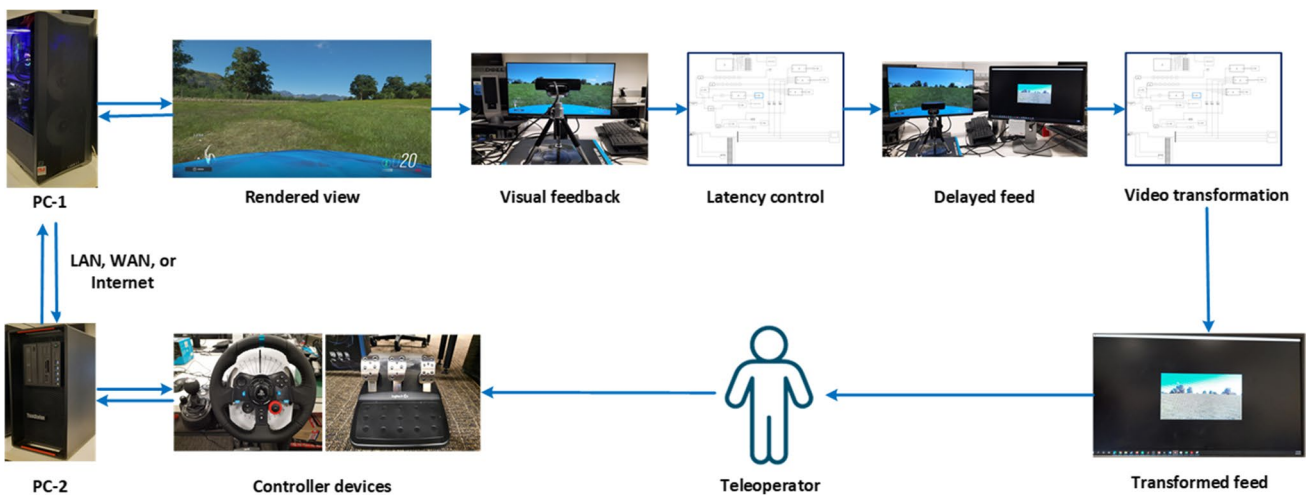

Fig. 13 System diagram to incorporate both control and visual latency to the system. Here PC-1 hosts the simulation platform and the enhancement model, and PC-2 connects to the controller devices

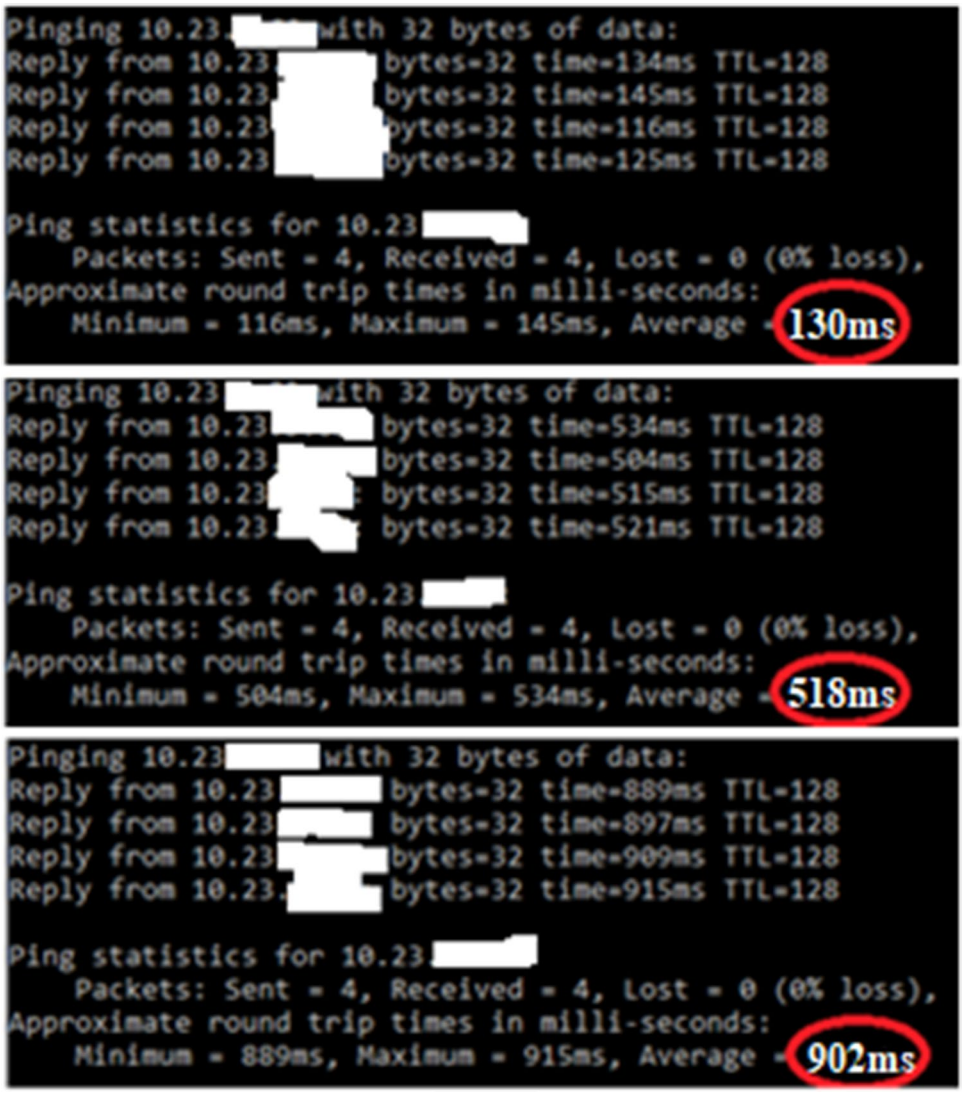

Fig. 14 PC-to-PC packet delay (any amount) facilitated by the WinDivert based Clumsy tool

Table 5 Comparison of Control delay only and visual feed delay only teleoperation

Delay type	Delay (ms)	Time (s)	Speed (km/h)	Oscillation (Count)	Out of Track (count)
On-Road					
Control Only	1032	499	14.42	18	3
Visual Only	900	523	14.01	10	4
Off-Road					
Control Only	1032	342	15.8	16	1
Visual Only	900	310.9	17.91	9	3.5

Table 6 Comparison of teleoperation performance affected by combined latency

Parameter	Delay (ms)	Without assistance	Sliding only	Sliding and zooming
On-Road				
Time (s)	1200	499	428	422
Speed (km/h)	1200	14.42	16.8	17.1
Oscillation	1200	10	0	0
Out of Track	1200	2	1	0
Off-Road				
Time (s)	1076	385	282	281
Speed (km/h)	1076	14.02	19.14	19.22
Oscillation	1076	11	0	0
Out of Track	1076	1	0	0

called VirtualHere. Although USB devices usually need to be directly connected to a computer to be used, VirtualHere facilitates the transmission of USB signals over a LAN, WAN or Internet connection to a remote machine, allowing for virtual connection of USB devices over a network.

To investigate the impact of control latency we have run teleoperation sessions on both on-road and off-road tracks. These are the same tracks used by the survey participants. We have compared the outcomes with the non-assisted delayed visual feed outcomes by the participants (From Table 3). The non-assisted delayed feed is affected by the visual delay only. The comparison is in Table 5 below. For both the on-road and off-road tracks, the average speed and time taken to complete one lap is very similar. A small amount of difference exists as the control only delay is a little higher than that of the visual delay.

Table 5 demonstrates that a certain amount of latency or delay in the teleoperation control loop, regardless of its source, will have the same impact on the teleoperator. To reinforce this statement and to prove the robustness of our teleoperation enhancement technique, we have conducted teleoperation runs on on-road and off-road tracks where the total delay in the loop consists of both control input delay and visual output delay. The outcome of the teleoperation sessions is provided in Table 6. For the on-road teleoperation session, the cumulative latency was 1200 ms where the control input latency was 560 ms and the visual feed latency was 640 ms. Without any video transformation based assistance it took 499 s to complete the 2 km track with an average speed of 14.42 km/h. Using our sliding window enhancement, the completion time reduced to 428 s with an average speed of 16.8 km/h. Using the sliding and zooming window both the time and speed improved further. Incidents of oscillation and out of track events also improved substantially using our assisted windows.

For the off-road track, the figures correspond with the previous numbers. In this case, the total latency is 1076 ms where 560 ms latency is induced from the controller to PC-2 and 516 ms latency is induced by our Simulink® visual transformation model. Using the sliding window the completion time reduced from 385 s to 282 s. For sliding and zooming the time is further reduced. The oscillation and out of track events reduced to zero for both the sliding only and sliding with zooming windows. Table 6 reflects the same outcome to that obtained via the participants' survey. It reconfirms the findings of our research that our video transformation-based assisted windows enhance an operator's performance for high latency ground-vehicle teleoperation. Based on these results, it can confidently be stated that whether the delay is present only in the visual feed, only in the control input, or a combination of the two (as would be the case in a real-world scenario) the impact is effectively the same from the operator perspective. This validates the platform that has been setup to evaluate the teleoperation enhancement techniques.

Acknowledgements The authors acknowledge the contribution of the survey participants who took part to evaluate the teleoperation system and provided valuable feedback. The authors also thank Mr. Hassan Mahmood and Mr. Shinsuke Matsubara for their suggestions and help with Matlab and Simulink as well as Dr Guy Gallasch and Dr Robert Hunjet from DSTG for their feedback on the teleoperation simulation and enhancement model.

Author Contributions The contributions of the authors are as follows.
- MD Moniruzzaman: Conceptualisation, model creation, data collection, experimentation, manuscript preparation.
- Alexander Rassau: Conceptualisation, supervision, manuscript preparation, and editing.
- Douglas Chai: Conceptualisation, supervision, manuscript preparation, and editing.
- Syed Mohammed Shamsul Islam: Conceptualisation, supervision, manuscript preparation, and editing.

Funding Information Open Access funding enabled and organized by CAUL and its Member Institutions This research has been funded by the ECU-DSTG Industry PhD scholarship.

Data availability If requested, the authors will share the data collected and used in this paper.

Code availability If requested, the authors will share the code and Simulink model prepared and used in this paper.

Declarations

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval Appropriate approval has been granted by the Human Ethics Committee (HREC) of Edith Cowan University (ECU) prior to any operator data collection. Approval number 2020-01643.

Consent to participate The authors declare that appropriate consent has been taken from the participants of the human survey described in this paper.

Consent for publication This paper does not contain materials, figures, tables, etc that have been published anywhere else. Therefore consent for publication is not applicable.

References

1. Moniruzzaman, M., Rassau, A., Chai, D., Islam, S.M.S.: Robotic teleoperation methods and enhancement techniques: A comprehensive survey. Robotics and Autonomous Systems **150**, 103973 (2022). https://doi.org/10.1016/j.robot.2021.103973

2. Little, C.L., Perry, E.E., Fefer, J.P., Brownlee, M.T., Sharp, R.L.: An interdisciplinary review of camera image collection and analysis techniques, with considerations for environmental conservation social science. Data **5**(2), 51 (2020). https://doi.org/10.3390/data5020051
3. Fong, T., Thorpe, C., Baur, C.: Advanced interfaces for vehicle teleoperation: Collaborative control, sensor fusion displays, and remote driving tools. Auton. Robots **11**(1), 77–85 (2001). https://doi.org/10.1023/A:1011212313630
4. Sheridan, T.B.: Telerobotics. Automatica **25**(4), 487–507 (1989). https://doi.org/10.1016/0005-1098(89)90093-9
5. Choi, J.J., Kim, Y., Kwak, S.S.: The autonomy levels and the human intervention levels of robots: The impact of robot types in human-robot interaction. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp. 1069–1074. IEEE, Edinburgh (2014). https://doi.org/10.1109/ROMAN.2014.6926394
6. Dorais, G., Bonasso, R.P., Kortenkamp, D., Pell, B., Schreckenghost, D.: Adjustable autonomy for human-centered autonomous systems. In: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence Workshop on Adjustable Autonomy Systems, pp. 16–35 (1999)
7. Fong, T., Thorpe, C., Baur, C.: Robot as partner: Vehicle teleoperation with collaborative control. In: Multi-Robot Systems: From Swarms to Intelligent Automata, pp. 195–202. Springer, Washington DC, USA (2002). https://doi.org/10.1007/978-94-017-2376-3_21
8. Bruemmer, D.J., Marble, J.L., Dudenhoeffer, D.D., Anderson, M., McKay, M.D.: Mixed-initiative control for remote characterization of hazardous environments. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS), pp. 9–17. IEEE, Big Island, USA (2003). https://doi.org/10.1109/HICSS.2003.1174289
9. Sellner, B., Heger, F.W., Hiatt, L.M., Simmons, R., Singh, S.: Coordinated multiagent teams and sliding autonomy for large-scale assembly. Proc. IEEE **94**(7), 1425–1444 (2006). https://doi.org/10.1109/JPROC.2006.876966
10. MacKenzie, I.S., Ware, C.: Lag as a determinant of human performance in interactive systems. In: Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems, Amsterdam, Netherlands, pp. 488–493 (1993). https://doi.org/10.1145/169059.169431
11. Chen, J.Y., Haas, E.C., Barnes, M.J.: Human performance issues and user interface design for teleoperated robots. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **37**(6), 1231–1245 (2007). https://doi.org/10.1109/TSMCC.2007.905819
12. Matheson, A., Donmez, B., Rehmatullah, F., Jasiobedzki, P., Ng, H.-K., Panwar, V., Li, M.: The effects of predictive displays on performance in driving tasks with multi-second latency: Aiding tele-operation of lunar rovers. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 57, pp. 21–25. SAGE Publications, Los Angeles, CA (2013). https://doi.org/10.1177/1541931213571007
13. Lane, J.C., Carignan, C.R., Sullivan, B.R., Akin, D.L., Hunt, T., Cohen, R.: Effects of time delay on telerobotic control of neutral buoyancy vehicles. In: Proceedings 2002 IEEE International Conference on Robotics and Automation, vol. 3, pp. 2874–2879. IEEE, Washington DC, USA (2002). https://doi.org/10.1109/ROBOT.2002.1013668
14. Ellis, S.R., Mania, K., Adelstein, B.D., Hill, M.I.: Generalizeability of latency detection in a variety of virtual environments. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 48, pp. 2632–2636. SAGE Publications, Los Angeles, CA (2004). https://doi.org/10.1177/154193120404802306
15. Held, R., Efstathiou, A., Greene, M.: Adaptation to displaced and delayed visual feedback from the hand. J. Exp. Psychol. **72**(6), 887 (1966). https://doi.org/10.1037/h0023868
16. Barnes, M., Cosenzo, K., Mitchell, D., Chen, J.: Human robot teams as soldier augmentation in future battlefields: An overview. In: Proceedings of the 11th International Conference of Hum (2005)
17. Sanzharov, V., Frolov, V., Galaktionov, V.: Survey of nvidia rtx technology. Program. Comput. Softw. **46**(4), 297–304 (2020). https://doi.org/10.1134/S0361768820030068
18. Mahmud, S., Lin, X., Kim, J.-H.: Interface for human machine interaction for assistant devices: a review. In: 10th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0768–0773. IEEE, Las Vegas, NV, USA (2020). https://doi.org/10.1109/CCWC47524.2020.9031244
19. Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S.: A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. **53**(8), 5455–5516 (2020). https://doi.org/10.1007/s10462-020-09825-6
20. Al-Garadi, M.A., Mohamed, A., Al-Ali, A.K., Du, X., Ali, I., Guizani, M.: A survey of machine and deep learning methods for internet of things (iot) security. IEEE Commun. Surv. Tutor. **22**(3), 1646–1685 (2020). https://doi.org/10.1109/COMST.2020.2988293
21. Draper, J.V., Kaber, D.B., Usher, J.M.: Telepresence. Hum. Factors **40**(3), 354–375 (1998). https://doi.org/10.1518/001872098779591386
22. Fong, T., Thorpe, C.: Vehicle teleoperation interfaces. Auton. Robots **11**(1), 9–18 (2001). https://doi.org/10.1023/A:1011295826834
23. Fong, T., Thorpe, C., Baur, C.: Advanced interfaces for vehicle teleoperation: Collaborative control, sensor fusion displays, and remote driving tools. Auton. Robots **11**(1), 77–85 (2001). https://doi.org/10.1023/A:1011212313630
24. Nof, S.Y.: Springer Handbook of Automation. Springer, Heidelberg, Germany (2009)
25. Dybvik, H., Løland, M., Gerstenberg, A., Slåttsveen, K.B., Steinert, M.: A low-cost predictive display for teleoperation: Investigating effects on human performance and workload. Int. J. Hum. Comput. Stud. **145**, 102536 (2021). https://doi.org/10.1016/j.ijhcs.2020.102536
26. Wilde, M., Chan, M., Kish, B.: Predictive human-machine interface for teleoperation of air and space vehicles over time delay. In: 2020 IEEE Aerospace Conference, pp. 1–14. IEEE, Big Sky, MT, USA (2020). https://doi.org/10.1109/AERO47225.2020.9172297
27. Ha, C., Yoon, J., Kim, C., Lee, Y., Kwon, S., Lee, D.: Teleoperation of a platoon of distributed wheeled mobile robots with predictive display. Auton. Robots **42**(8), 1819–1836 (2018)
28. Wang, C., Zuo, Z., Lin, Z., Ding, Z.: A truncated prediction approach to consensus control of lipschitz nonlinear multiagent systems with input delay. IEEE Trans. Control. Netw. Syst. **4**(4), 716–724 (2016). https://doi.org/10.1109/TCNS.2016.2545860
29. Richard, J.-P.: Time-delay systems: an overview of some recent advances and open problems. Automatica **39**(10), 1667–1694 (2003). https://doi.org/10.1016/S0005-1098(03)00167-5
30. Manitius, A., Olbrot, A.: Finite spectrum assignment problem for systems with delays. IEEE Trans. Autom. Control **24**(4), 541–552 (1979). https://doi.org/10.1109/TAC.1979.1102124
31. Artstein, Z.: Linear systems with delayed controls: A reduction. IEEE Trans. Autom. Control **27**(4), 869–879 (1982). https://doi.org/10.1109/TAC.1982.1103023
32. Bejczy, A.K., Kim, W.S.: Predictive displays and shared compliance control for time-delayed telemanipulation. In: IEEE International Workshop on Intelligent Robots and Systems, Towards a New Frontier of Applications, pp. 407–412. IEEE, Ibaraki, Japan (1990). https://doi.org/10.1109/IROS.1990.262418
33. Bejczy, A.K., Kim, W.S., Venema, S.C.: The phantom robot: predictive displays for teleoperation with time delay. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Cincinnati, OH, USA, pp. 546–551 (1990). https://doi.org/10.1109/ROBOT.1990.126037

34. Buzan, F.T., Sheridan, T.B.: A model-based predictive operator aid for telemanipulators with time delay. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Cambridge, MA, USA, pp. 138–143 (1989). https://doi.org/10.1109/ICSMC.1989.71268
35. Hirzinger, G., Heindl, J., Landzettel, K.: Predictive and knowledge-based telerobotic control concepts. In: IEEE International Conference on Robotics and Automation, Scottsdale, AZ, USA, pp. 1768–1769 (1989). https://doi.org/10.1109/ROBOT.1989.100231
36. Witus, G., Hunt, S., Janicki, P.: Methods for ugv teleoperation with high latency communications. In: Unmanned Systems Technology XIII, vol. 8045, p. 80450. SPIE, Orlando, Florida, United States (2011). https://doi.org/10.1117/12.886058
37. Zheng, Y., Brudnak, M.J., Jayakumar, P., Stein, J.L., Ersal, T.: An experimental evaluation of a model-free predictor framework in teleoperated vehicles. IFAC-PapersOnLine **49**(10), 157–164 (2016). https://doi.org/10.1016/j.ifacol.2016.07.513
38. Zhang, Y., Li, H.: Handling qualities evaluation of predictive display model for rendezvous and docking in lunar orbit with large time delay. In: 2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC), pp. 742–747. IEEE, Nanjing, China (2016). https://doi.org/10.1109/CGNCC.2016.7828878
39. Johnson, D.H.: Signal-to-noise ratio. Scholarpedia **1**(12), 2088 (2006). https://doi.org/10.4249/scholarpedia.2088
40. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
41. Eikvil, L.: Optical Character Recognition. Citeseer, Princeton, New Jersey, USA (1993)
42. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index. IEEE Trans. Image Process. **21**(4), 1488–1499 (2011). https://doi.org/10.1109/TIP.2011.2173206
43. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, vol. 2, pp. 1398–1402. IEEE, Pacific Grove, CA, USA (2003). https://doi.org/10.1109/ACSSC.2003.1292216
44. Sobel, I., Duda, R., Hart, P., Wiley, J.: Sobel-Feldman Operator. Preprint at https://www.researchgate.net/profile/Irwin-Sobel/publication/285159837. Accessed 20 Apr 2022
45. Rouse, D.M., Hemami, S.S.: Analyzing the role of visual structure in the recognition of natural image content with multi-scale ssim. In: Human Vision and Electronic Imaging XIII, vol. 6806, pp. 680615. SPIE, San Jose, California, United States (2008). https://doi.org/10.1117/12.768060
46. Søgaard, J., Krasula, L., Shahid, M., Temel, D., Brunnström, K., Razaak, M.: Applicability of existing objective metrics of perceptual quality for adaptive video streaming. Electronic Imaging **2016**(13), 1–7 (2016)
47. Dosselmann, R., Yang, X.D.: A comprehensive assessment of the structural similarity index. Signal Image Video Process. **5**(1), 81–91 (2011). https://doi.org/10.1007/s11760-009-0144-1
48. Lu, S., Zhang, M.Y., Ersal, T., Yang, X.J.: Effects of a delay compensation aid on teleoperation of unmanned ground vehicles. In: Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 179–180. Association for Computing Machinery, Chicago, USA (2018). https://doi.org/10.1145/3173386.3177064
49. Mathan, S., Hyndman, A., Fischer, K., Blatz, J., Brams, D.: Efficacy of a predictive display, steering device, and vehicle body representation in the operation of a lunar vehicle. In: Conference Companion on Human Factors in Computing Systems, pp. 71–72. Association for Computing Machinery, Vancouver, BC, Canada (1996)

MD Moniruzzaman is a current PhD candidate of the School of Engineering at Edith Cowan University, Western Australia. He completed his masters in Computer Science from Edith Cowan University and a bachelor in Electrical and Electronic Engineering from Khulna University of Engineering & Technology. He has several years of experience in telecommunication network deployment and casual tutoring for computer systems analysis and mathematics. He is a current student member for IEEE. His research interest includes robotic teleoperation, AI in robotics, and computer vision.

Alexander Rassau received a Bachelor of Science (Cybernetics and Control Engineering) and a PhD in microelectronics from the University of Reading in the United Kingdom in 1997 and 2000 respectively. He is currently the Associate Dean Teaching and Learning for the School of Engineering at Edith Cowan University. He has played an active role in the very rapid expansion of the School over the past decade, overseeing the introduction of a wide range of new and innovative engineering programs and courses. Since 1998, he has also been actively involved as a researcher and educator in the areas of embedded systems, intelligent control, machine learning, automation and robotics.

Douglas Chai completed the BE (Hons) and PhD degrees in Electrical and Electronic Engineering from the University of Western Australia, Australia, in 1994 and 1999, respectively. He is currently the Associate Dean (Academic) with the School of Engineering at Edith Cowan University, Australia. His research interests include image analysis, pattern recognition and computer vision. He has published over 90 technical papers, and received over 5,000 citations according to Google Scholar. He was an associate editor of the Australian Journal of Electrical and Electronics Engineering (AJEEE) in 2014-2017. Chai is a senior member of the Institute of Electrical and Electronics Engineers (IEEE). He has served in various IEEE committees for over 22 years, including secretary of IEEE Region 10 (in 2023-2024), chairmanship of IEEE Western Australia (WA) Section (in 2003-2004, 2007-2008), the IEEE Signal Processing WA Chapter (in 2005-2006, 2008, 2011-2013, 2018-2020), and the IEEE ComSoc WA Chapter (2021-2022). He received the Outstanding Volunteer Award from the IEEE WA Section in 2019.

Syed Mohammed Shamsul Islam completed his PhD with Distinction in Computer Engineering (CE) from the University of Western Australia (UWA) in 2011. He also received an MSc Comp Sci (KFUPM) and BSc EEE (IUT) in 2005 and 2000 respectively. He worked at UWA and Curtin University in the past, and currently working as a Senior Lecturer in Computing at Edith Cowan University (ECU). He is the Colead for the 3D sensing, visualisation and analytics lab and a founding member of the Centre for Artificial Intelligence and Machine Learning in the School of Science. He has secured the NHMRC Ideas Grant 2019 (A$ 467,980) and 12 other external research grants totalling over a million Australian dollars. He also attracted tens of public media releases including a TV news, and awards including the High Achieving Researcher 2021 award by ECU. He has published over 65 fully refereed scholarly articles including 22 journals, and three conference papers with best paper awards. He is an Associate Editor of IEEE Access, and reviewer/committee member of many highly ranked journals/conferences. He is a Senior member of both the IEEE and Australian Computer Society. His research interest includes Artificial Intelligence, Biomedical Engineering, and Computer Vision.