Intuitive Robot Teleoperation Through Multi-Sensor Informed Mixed Reality Visual Aids

SALVATORE LIVATINO[1], **DARIO C. GUASTELLA**[2], **(Associate Member, IEEE),**
GIOVANNI MUSCATO[2], **(Senior Member, IEEE), VINCENZO RINALDI**[3], **LUCIANO CANTELLI**[2],
CARMELO D. MELITA[2], **ALESSANDRO CANIGLIA**[4], **RICCARDO MAZZA**[5],
AND GIANLUCA PADULA[6]

[1]School of Physics, Engineering and Computer Science (SPECS), University of Hertfordshire, Hatfield AL10 9AB, U.K.
[2]Department of Electrical, Electronics and Computer Engineering (DIEEI), University of Catania, 95125 Catania, Italy
[3]Leverhulme Research Centre for Forensic Science, University of Dundee, Dundee DD1 4HN, U.K.
[4]Microsoft Italy, 20154 Milan, Italy
[5]Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland (SUPSI), 6928 Manno, Switzerland
[6]Academic Laboratory of Movement and Human Physical Performance, Dynamolab, Medical University of Łódź, 90-419 Łódź, Poland

Corresponding author: Salvatore Livatino (s.livatino@herts.ac.uk)

ABSTRACT Mobile robotic systems have evolved to include sensors capable of truthfully describing robot status and operating environment as accurately and reliably as never before. This possibility is challenged by effective sensor data exploitation, because of the cognitive load an operator is exposed to, due to the large amount of data and time-dependency constraints. This paper addresses this challenge in remote-vehicle teleoperation by proposing an intuitive way to present sensor data to users by means of using mixed reality and visual aids within the user interface. We propose a method for organizing information presentation and a set of visual aids to facilitate visual communication of data in teleoperation control panels. The resulting sensor-information presentation appears coherent and intuitive, making it easier for an operator to catch and comprehend information meaning. This increases situational awareness and speeds up decision-making. Our method is implemented on a real mobile robotic system operating outdoor equipped with on-board internal and external sensors, GPS, and a reconstructed 3D graphical model provided by an assistant drone. Experimentation verified feasibility while intuitive and comprehensive visual communication was confirmed through an assessment, which encourages further developments.

INDEX TERMS Virtual reality, augmented reality, user interfaces, graphical user interfaces, human-robot interaction, telerobotics, stereo vision.

I. INTRODUCTION

During the last two decades, robotic vehicles have been proposed for controlled environments such as depots and automated manufacture halls. Different systems have also been employed to help with dangerous tasks such as bomb disposal and mine discovery. Whereas mobile robots' autonomy has constantly increased, so has the awareness of the unreplaceable value of manual teleoperation, especially for challenging tasks and unknown environments.

The associate editor coordinating the review of this manuscript and approving it for publication was Pedro Neto.

Manually operated robotic vehicles, quite often behaving semi-autonomously, need to be commanded through their operator's interface. There has been an increased consciousness about the role of the interface in enhancing operator's situational awareness and its impact on operational performance.

Operator's situational awareness is a key aspect to operate effectively remote vehicles. This means among other things: understanding the surrounding environment, the robot location, the contextual robot-environment movements, and predicting future robot-environment behaviors [1], [2].

New literature contributions have proposed different ways to increase awareness, e.g. better representations of sensors

information [3], wider range of available commands and options [4], and more intuitive dashboards including the use of three-dimensional displays [5]. Despite the use of immersive displays and mixed reality (MR) representations has often been discussed [6]–[10], it is rare to find it on commercial products. There is nonetheless consensus about its potential and advantage in increasing operator's sense of presence in remote workspaces. This in turn means greater environment comprehension, which positively affects task performance and decision-making.

Teleoperation interfaces have recently included stereoscopic three-dimensional (S3D) monitors [11], [12], while in some fields such as telesurgery, S3D has become an established technology (see e.g. the spread of the *daVinci* system). The involvement of other human sensor modalities such as haptics is also being researched, but it does not appear on commercial products yet.

When remotely operating in outdoor natural environments, users may be challenged by: complexity of scenarios and objects, events dynamics, richness of the provided live-sensor data, limited display size, and also by the way available prior knowledge is communicated. Figure 1 illustrates main actors playing a role in operational behavior and decision process.
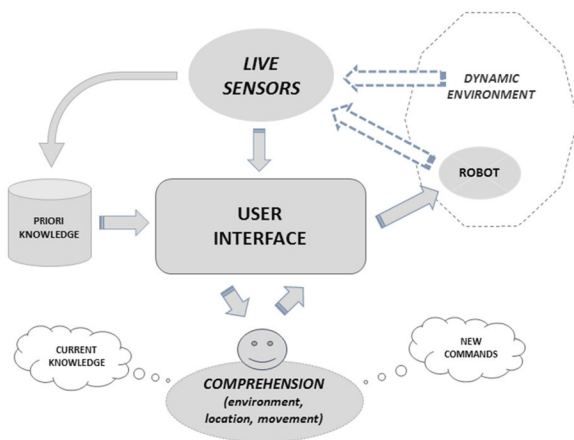
FIGURE 1. Conceptual illustration of key factors and user interface role during robot teleoperation.

This work proposes an intuitive way to present multiple sensor data to a remote-vehicles operator, which is expected to increase situational awareness. This is to be achieved: (1) by providing operators with as much knowledge as possible about the remote space and its current condition; and (2) by communicating information intuitively.

We have today many information-rich sensors at robot disposal [13], [14], e.g. 3D cameras, laser scanners, sonars, infra-red range finders and GPS, which can provide knowledge to greatly improve navigation and intervention performance. The challenge is then about how to communicate users a fairly large amount of sensor information effectively, therefore avoiding cognitive overload [1]. Our answer is to present sensor information visually and this should rely on:

- Coherent combination of different live sensors information. A lot can be achieved by applying Information Visualization theories to improve visual communication of data.
- MR representations and immersive displays. This combination can increase users' comprehension and sense of presence, and therefore remote-space awareness.

The operator will be observing an adaptive MR scenario consisting of three-dimensional streamed videos and graphical representations of sensor data. The latter will be designed as multi-sensor informed visual aids (VAs).

The focus is on outdoor natural scenarios, where robots are able to provide positional and attitude information together with live views of the surrounding environment. Previously or concurrently acquired environment maps can be considered, e.g. maps today achievable from drone aerial views.

The next section introduces the state of the art, whereas section III describes the proposed approach and specific choices. Section IV describes the implemented system, whereas section V analyzes results of experimentation trials. In section VI conclusions are drawn.

II. IMMERSIVE VISUALIZATION AND MR INTERFACES
A. VIRTUAL REALITY HEADSETS
Ivan Sutherland built the first HMD prototype in 1968 [15]. VR headsets had since then received continuous interest because of their potential in providing full visual immersion into artificially generated environments. However, despite the relevant improvements occurred in the last decades, with HMDs adopting optical tracking and OLED displays, while becoming smaller and lighter [16], some other issues remained. These were e.g. observed tunnel vision, tethering to PC, weight, portability and high cost.

They all limited HMD's adoption into the consumer market and HMD use was mainly confined within research labs. This happened up until this last decade. A notable step forward came then in 2012, with the first Oculus Rift system [17], which among other things, featured wide Field of View (FOV) and low-cost.

There has been great development since then with newer systems featuring wider displays and higher resolution, lower cost and wireless connection. The latter being one of the latest systems' focus, which saw first the making of "smartphone-based VR headsets", which rely on smartphones and their displays to operate [18], [19], then the standalone VR headsets, which rely on dedicated computing and display systems [20], [21]. They opened up to wide applications and a large audience [22]. Today's highest-specs VR headset have nonetheless remained those wired to a desktop PC. The reason being HMDs can exploit the greater PC processing and graphic power [23], [24]. Additional options are meantime being proposed, such as e.g. embedded eye-trackers [25], [26].

Compared to views of traditional 2D desktop monitors, but also to higher specs stereoscopic 3D monitors [27], [28], VR headsets feature high user's isolation from the
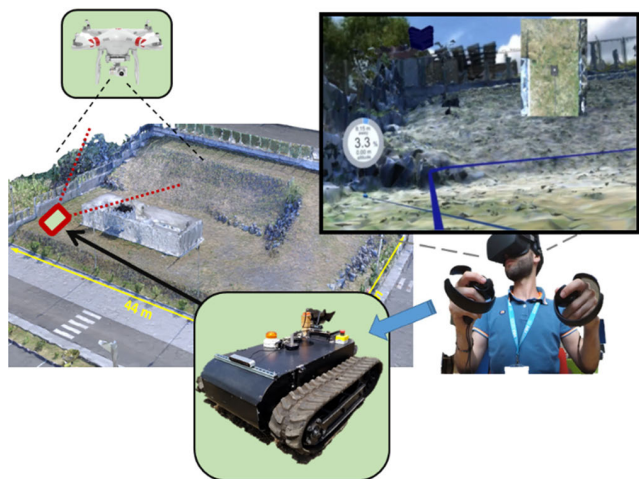
FIGURE 2. Illustrated overview of the proposed system and approach to outdoor robot teleoperation. The figure shows the robot-navigation workspace used in our experiments, which is surveyed by a drone that through its camera-system graphically reconstructs the below workspace area in 3D. An image of the ground robot used in our experiments is also included. On the right-hand side an image showing a robot operator wearing a VR headset, with on top an example of operator's view during navigation. The operator's view shows a MR scene depicted within a single-window representation. The view includes the video-stream image integrated with the reconstructed 3D model and four of the proposed visual aids (centerline, virtual pointer, top-view and guide arrow).

surrounding environment, while allowing for continuous omnidirectional / 360° viewing through natural head movements, rather than using a computer mouse or multi-dimensional joysticks. Figure 2 depicts our user wearing a VR headset.

B. MIXED REALITY REPRESENTATIONS

Virtual reality (VR) is the simulation of a world that can be real or invented. VR is often experienced through the use of immersive technologies, therefore involving human sensory inputs and especially vision. Computer graphics (CG) is typically used to visually represent VR environments, but also to represent signs, symbols, indicators, diagrams and numbers.

MR is the combination of real objects (live or recorded) with virtual objects, within consistent representations. Instances of MR are Augmented Reality (AR) and Augmented Virtuality (AV) [29] depending on which between reality and virtuality is the primary visual element. A main challenge for MR is alignment of real and virtual elements, which has been responsible for slower take up during last years compared to VR. Nonetheless, MR has recently got a new momentum because of the significant technology improvements of VR/AR headsets. These have become lighter, wireless, comfortable, sunlight insensitive and portraying sharp and well-aligned images.

MR representations are proposed in the literature to enable a more coherent combination of different information within the same visual context, which is more intuitive [3], [30]. The proposed interfaces focus on specific features and offer various elements to support operators while driving

remotely-located vehicles. MR representations have sometimes included the use of visual aids (VAs) to provide information on sensor data or on robot and environment status.

In case of sensor data, VAs mainly consist of graphic elements that exclusively display proprioceptive sensors data [2], [10]. Only in [3] some information about remote terrain (namely slope) is given to teleoperators. Other interfaces propose the use of stereo cameras to increase spatial awareness, but only recent works combine stereoscopic views with head-mounted displays (HMDs) [10], [12], [31]–[36].

In case of robot and environment status, VAs refer to the outcome of SLAM algorithms or, more generally, real-time 3D reconstructions, to obtain a 3D model of the region the robot is navigating into [31], [35]. Others do not take into account any information about the environment geometry [2], [12], [33], [34], [37] despite status information and 3D models have recently become more available and accessible.

Furthermore, most of the works exploiting MR have been designed for indoor scenarios or outdoor scenarios under certain conditions only [31]. A summary of related literature works, including interfaces' visualization characteristics and VAs is shown in Table 1.

III. METHOD: COMPREHENSIVE USER INTERFACE BASED ON VISUAL AIDS

This paper proposes a combination of elements within a user's visual interface for teleoperation of ground vehicles, which integrates live and pre-acquired sensors data, to represent robot and environment status within an immersive and adaptive MR view. Such a combination has never been proposed in the related literature. The interface combines:

a) **Immersive Visualization**. A natural approach to visual observation and interaction, which best suits the use of VR headsets. Interaction in the latter can take place through head movements and hand controllers.

b) **Three-Dimensional Mixed Reality**. An Augmented Virtuality (AV) visual representation [29], where real and virtual elements are integrated three-dimensionally on operator's display. These elements represent sensors data providing: positional information environment maps, graphical 3D reconstruction and live stereo images.

c) **Visual Aids**. A graphical representation of sensor data designed to provide specific types of aid in an intuitive way. They represent environment and direction information and aim to assist users during navigation.

The interface is designed according to following elements: *single-window representation*, *video-synthetic images*, *intuitive data viewing*, *regional and directional visual aids*.

A. SINGLE-WINDOW REPRESENTATION

The interface view shown on the VR headset represents a single window. This is the sole control panel users can rely on during robot teleoperation. The shown information may need to be rich (because of the many available sensors), while it may also need to be quickly comprehended, e.g.

TABLE 1. Summary of literature work's main characteristics we consider as reference. Ego/Exo stands for Egocentric/Exocentric.

	Viewing Frames	Display Type	Visual Context (video+graphics)		Visual Aids
2007 - Nielsen et al. [30]	Single	2D screen	AV (2D + 3D)	Ego/Exo	Robot's bonnet view, 3D map
2011 - Kelly et al. [3]	Single	2D monitor	AV (3D + 3D)	Ego/Exo	Vehicle's bonnet view, extended camera view, real-time video mapping, speed indicator, covered path trace, color-coded terrain slope
2012 - Ventura et al. [33]	Single	HMD	R (S3D)	Ego	-
2012 - Livatino et al. [7]	Single	S3D monitor	AR (S3D + 3D)	Ego	Proximity walls, laser rays, ground concentric circles, distance numerical values
2013 - Okura et al. [32]	Single	HMD	AV (S3D + 3D)	Ego/Exo	Robot's bonnet view, free-viewpoint generation, real-time video mapping, heading arrow
2015 - Krückel et al. [2]	Single	HMD	AR (S3D + 2D)	Ego	Artificial horizon and heading
2015 - Jankowski et al. [12]	Single	2D/S3D screen, or HMD	R (S3D)	Ego	-
2017 - Huy et al. [4]	Single	HMD	AR (S3D + 3D)	Ego	GUI for high-level task management, laser projected arrows and text
2017 - Tikanmäki et al. [35]	Single	HMD	AV (S3D + 3D)	Ego/Exo	Robot's bonnet view, real-time video mapping, extended camera view
2017 - Adamides et al. [37]	Single/Multi	2D screen, or HMD	R (3D)	Ego	-
2018 - Saputra et al. [34]	Single	HMD	R (S3D)	Ego	-
2018 - Kot et al. [10]	Multi	S3D screen, or HMD	AV (S3D + 3D) AR (S3D + 3D)	Ego	Robot's bonnet view, rear camera display, thermal camera display, robot status icons
2019 - Stotko et al. [31]	Single	HMD	AV (S3D + 3D)	Ego/Exo	Robot's bonnet view, real-time video mapping, extended camera view, interactive distance measurement
2019 - Roldán et al. [36]	Single/Multi	2D screen, or HMD	AV (S3D + 3D)	Exo	Robot's bonnet view, extended camera view, highlighted hints
Our work	Single	HMD	AV (S3D + 3D)	Ego/Exo	Robot's bonnet view, guide arrow, traversable area, virtual pointer, top view, centerline, extended camera view

when decisions are made while driving at speed. Sensors information presentation is therefore a relevant aspect. We see in literature control panels that only show subsets of available sensor data, others that show more sensor outputs but through different windows [38]–[40], and others that attempt grouping all data within a single window [3], [7], [30].

We want our interface to include all sensor data as well as any available prior knowledge, aiming at achieving a continuous and exhaustive monitoring of events. We follow recommendations from Gibson's ecological approach [41]. Therefore, we propose the use of a single window where we concurrently display live video-stream of the remote-scene and graphical representations of incoming sensor data combined with prior knowledge. The graphical elements are indications such as: distance to objects, traveled trajectories, environment features descriptors (slopes, obstacles), and a polygonal mesh of environment shape and objects resulted by the 3D reconstruction [42]. Video stream and graphics are three dimensionally integrated, a feature still quite uncommon in literature works (despite the use of 3D graphics). The single-window three-dimensional MR operator's view is a perfect design fit to the use of immersive displays, such as a VR headset. It also adapts well to 2D/3D desktop displays and
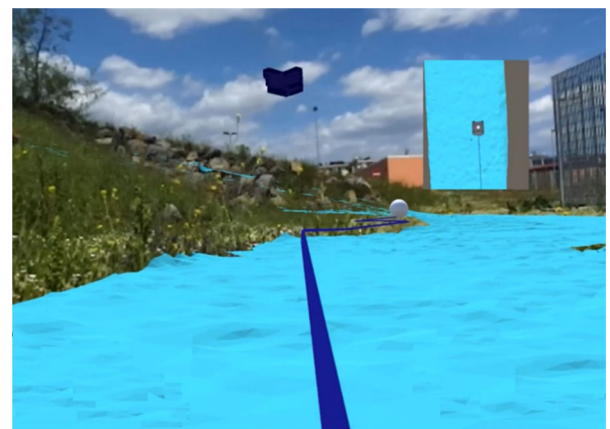
FIGURE 3. Example of operator's MR view with live video-stream in the background (showing sky, grass, rock-stones and buildings) and graphical representation of sensor data (suggested travel trajectory, guide arrow, top-view and traversable area in blue).

wall screens figure 2 top-right and figure 3 show examples of single-window representations.

B. VIDEO-SYNTHETIC IMAGES
A mixed reality visualization context containing both video and synthetic image textures is proposed, to support the

use of the single window representation and its consequent need to integrate data of different type and from different sources. In particular, we concurrently display a live video stream (reality) and graphical representations of other sensor data (virtuality).

The resulting MR view is an example of Augmented Virtuality [29]. It is in particular proposed to have the incoming video-stream graphically mapped on the reconstructed 3D graphical model. This allows the graphic engine to provide users with correct and coherent viewing perspective, and to correctly manage occlusions.

Vision being the dominant human sensory modality and the one we believe the most, we propose *reality* [29] to be always on screen whenever available (within the 3D MR context), and to portray the remote environment through the streaming of video-images (a rich information fast to be comprehended).

As for the *virtuality* representation, this consists of graphical objects representing trajectories, trends, slopes, walls, etc., which are three-dimensionally integrated in the MR context.

A careful design of functions and their appearance is therefore required. We apply useful guidelines to general interface design provided in Information Visualization literature [43], and we do it within the MANTRA context. Furthermore, we always have the option for users to adjust element views on demand.

Special attention is paid on spatial alignment in the 3D viewing space. We rely on a semi-automatic calibration process done at the start, which can also be repeated on demand during navigation. The proposed MR single-window view is visualized in S3D (for both video and graphics), which greatly helps users to better disambiguate among the different visualized elements based on the higher depth comprehension S3D provides. Figure 3 shows an example of our MR view.

The MR representation as the one proposed cannot be found in the literature. Table 1 collects related methods, which all have relevant differences with ours. We had also proposed one in recent years [7], but experiments ran indoors, on flat surfaces and with simple man-made scenarios. With this work we propose a new design framework, to address more challenging applications because of the uneven natural outdoor scenarios.

C. INTUITIVE DATA VISUALIZATION

When designing sensor data representation through the use of intuitive VAs, we propose to refer to Gestalt laws' most relevant indications and apply them to teleoperation interface's visual screen design. We focus on the eleven laws summarized by Chang *et al.* [44], which include:

- Smooth Continuation of lines and images [45]. We use it for path-indicators and mapped video streaming.
- Unambiguous and simple shapes. We use it for directional indicator and its design.
- Appropriate use of colors [46], [47]. We use it to design careful color pairing, limited hues, support expectations.

1) COLORS

We propose the use of standard color conventions, including those associated to danger–caution–safety [47], as they match users' expectations and recall the ordinary vehicle driving experience. We rely on the simplicity principle [44], use of small number of hues (but tuning up lightness and saturation), and colors pairing [46], [47]. The latter is especially relevant as we cannot control color-appearance of objects and landscapes in the incoming images, e.g. terrains and sky, but they act as background colors to contrast our VAs' color. We go for two dyad complementary colors as in MacDonald *et al.* [46].

2) TRANSPARENCIES

The need to manage multiple sensors data and their graphical representations, arising 3D occlusions, and the blending of video and graphics, make semi-transparency a key aspect. This plays a relevant role in supporting human attention and facilitating comprehension. The work of Harrison *et al.* [48] on transparency gives useful indications towards maintaining attention and fluency, which we apply to our MR views. We propose smooth blending and balancing between background video and the often sharper and less varying VAs. Furthermore, VAs' color and semi-transparency can all be adjusted on demand. Figure 4 shows an example of different level of VAs' transparency.


FIGURE 4. Adaptive Transparency VA with 3D reconstruction and traversable area: [left] both opaque, [center] only traversable area opaque, [right] full transparency of live streamed video.

3) ADAPTIVE VIEWS

To minimize cognitive overload, such as it could occur in dynamic situations requiring timely response [7], we follow the recommendations provided by Shneiderman [49] and Harrison *et al.* [48]. These concern with control panels' observations and the psychological problem of focused and divided attention. The answer is a screen view that dynamically changes depending on robot speed and user's preference. For instance, when moving it may show a qualitative overview that only shows relevant sensor information on path and close obstacles, graphically hiding object details, etc. This follows Shneiderman's Visual Information Seeking

Mantra (MANTRA) design approach [49]. Figure 5 shows examples of adaptive views using different color and transparency levels.
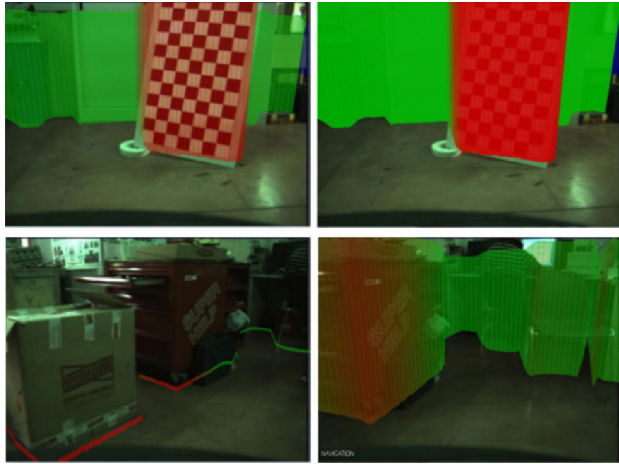
FIGURE 5. MR images showing qualitative overviews of an environment. Each row shows views of the same environment from the same viewpoint, with a video-image combined with a coloured graphical representation following objects' shape in 3D. The images to the right show instances of MR views occurring when the robot travels at higher speed. They show less objects details, which gives higher appreciation of free and occupied space. Such qualitative view is deemed relevant when driving at higher speed and therefore there is less time to look at details, but greater interest in avoiding collisions. The graphic mesh is coloured according to distance to objects, with red indicated higher proximity and risk of collision. Transparency changes according to vehicle's speed.

It is proposed that visibility and blending of predominant elements can change dynamically, adaptively, and on user's demand. We adopt the MANTRA approach for managing different data types [49]. We find very appropriate the suggested use of a small number of tasks, and we wish these to follow the order given by the MANTRA as it fits nicely to robot teleguide actions. We always Overview first, while Zooming, Filtering and Details are applied on demand [31], [35]. Automatic adaptation based on speed and distance to objects is an available option useful for specific environments and situations.

The Overview situation is empowered by the *exo-centric* viewing option. It can be combined with other VAs such as the *Traversable Area* and *Centerline* (introduced below). The Zooming can be enriched by measurements using *virtual pointers*. The Filter can be applied to colors and occlusions. We additionally consider within this approach the management of delays-driven latencies (based on timestamps). This to give consistency among visualized sensor data and to align them with the video-stream. Visible seams may occur because of image noise and sensor errors.

D. REGIONAL AND DIRECTIONAL VISUAL AIDS

The graphical elements, representing incoming sensor data, are specifically designed for teleoperation to visually aid understanding of: sensor information, remote-scene dynamics, vehicle's behavior and operator's commands. Eight types of VAs are proposed, which are grouped as Regional and Directional. The Regional VAs provide information on robot surrounding environment, while Directional VAs provide route following information (in our case this is based on traversing-cost path-planning [50]).

1) REGIONAL VISUAL AIDS

- ***Traversable Area***. It shows the crossable areas within current field of view. It can include expected difficulty and hide non-relevant details. The shown information is inserted and tuned either on-demand or adaptively (e.g. based on current speed). It is often an essential aid, especially for outdoor unstructured scenarios. Figures 3 and 4 include representations of the traversable area VA.

- ***Extended Camera View.*** It shows live camera images surrounded and integrated with graphics representing adjacent areas. This allows operators to perceive vehicle surrounding areas simultaneously to frontal view. It is helpful in many tasks to overcome narrow passages while providing greater sense of presence. It marries well with the proposed use of VR headset, which thanks to head-tracking adapts visualized images to user's head position. We propose live video-images to be inserted into a wider AV context through CG video-mapping. In particular, the streamed video is mapped onto a surface located in front of the robot according to robot heading. Our approach differs from literature works, such as Li and Fan [51] proposing sensor alignment through reconstructed 3D models (for outdoor applications), because we do not try to align real and virtual views. Rather, we map a live video feed into a surface inside the graphic representation. In this way to produce an extended camera view is straightforward through graphic rendering. We use image-processing to align video and graphical elements as in [7]. Our solution is feasible because we have knowledge of the robot surrounding environment provided by the robot sensors and reconstructed 3D model. Furthermore, our 3D model provides detailed graphical appearance because of the hi-resolution aerial cameras. This is very helpful to reduce the observed gap in quality between video and synthetic images. The main challenge for the proposed AV solution is alignment. We address it by careful horizon alignment [52], sensor data filtering and assessment, an initial and repeatable calibration [7], and graphic texture extrapolation in case of visible seam (due to lack of data). Figure 6 shows an extended camera view example.

- ***Exocentric View***. It shows the robot during navigation from a 3rd person view. We propose an exocentric 3D view generated from the available 3D model (updated whenever possible), GPS, IMU and odometric sensor data. This solution enables generation of any exocentric views, which is very useful in many occasions, such as parking and area overview. Operating with an exocentric view allows for faster space layout
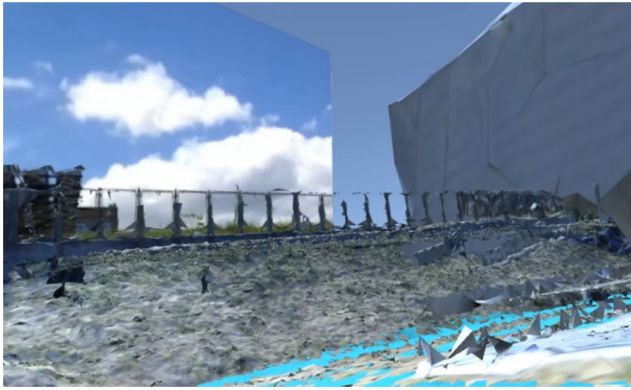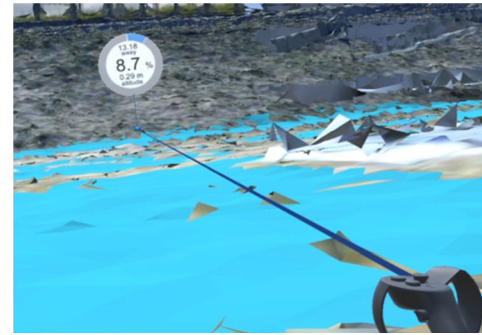
FIGURE 6. Extended Camera View VA: the image shows the graphically reconstructed 3D environment (from live sensor data), while the live video streamed from the onboard camera can be noted to the left in background.

FIGURE 7. Exocentric View VA with three vehicle's views captured during navigation. The left-image shows the Centerline VA on exocentric views.

FIGURE 8. Virtual Pointer VA with post raised on targeted point.

our pointer establishes a connection and distance-factor with the environment. A road-sign post like is shown at the position hit by the ray, displaying corresponding information in terms of altitude, slope, distance and traversability cost. Figure 4 shows an example. This VA supports the need one may have of reading specific measurements [48], which can be derived from the 3D reconstructed model or sensors data. Figure 8 shows an example.

2) DIRECTIONAL VISUAL AIDS

- ***Guide Arrow***. It shows a 3D arrow providing clear indication about the direction to follow. Figure 4 shows an example. Arrows are popular to indicate routes and very appreciated by inexperienced users [54]. They outperform alternatives such as compass and light sources [55], and are popular in latest video-games 54]. Our guide arrow is implemented through dynamic selection of most suitable centerline checkpoints and it is displayed in S3D making the indicated direction easy to catch visually. Figure 3 shows an example.
- ***Centerline***. It shows a graphic line that indicates the best vehicle position to have while navigating. A centerline represents a clear reference for vehicle's operators. Lines and segments are widely used on our roads to indicate lanes and specific behaviors. To use continuous lines to follow directions is in harmony with our brain sensitivity [45]. Lines provide that simplicity suggested in visual observations [44]. Differently from for our roads, our vehicle is expected to stand on top of the continuous line. This to ensure it keeps the best position while driving. The centerline is an intuitive and simple reference, which provides direction too. Figure 3 and figure 7 (right-hand side) show examples.
- ***Robot Bonnet***. It shows a graphical representation of robot front bonnet from ego-centric viewpoint. Therefore, only provided with this type of viewpoint. It includes an imaginary bonnet shape (with pointed end) to help operators comprehend vehicle-direction. Viewing vehicle's front bonnet effectively increases situational awareness 2], [31], and helps preventing collisions [5].

comprehension and effective vehicle maneuvering. Nielsen *et al.* [30] explain how an adjustable perspective can aid with all three levels of Endsley's situational awareness (perception, comprehension, and projection) [1]. Alternative solutions for exocentric view generation would be impractical as they could require for example large camera heads on top of vehicles or would help only on specific actions. We are aware that an exocentric view may slow down operations because it could divert operator's attention by uncovering details (the same way 3D views do compared to corresponding 2D views [5], [32]). This has nonetheless no impact on accuracy. Figure 7 shows two exocentric views iexamples.

- ***Top View***. It shows the environment from above. This type of view, popular on computer games to supports vehicles and person navigations, are considered of great help to catch vehicle attitude and plan future moves. Our top-view size and its position can be adjusted by users. Figure 3 shows an example.
- ***Virtual Pointer***. It shows a ray casted from the robot towards a location of interest in the environment. This type of VA already proposed in AR from early years [29] can be beneficial for human-computer interaction in virtual/augmented environments [53]. By casting a ray
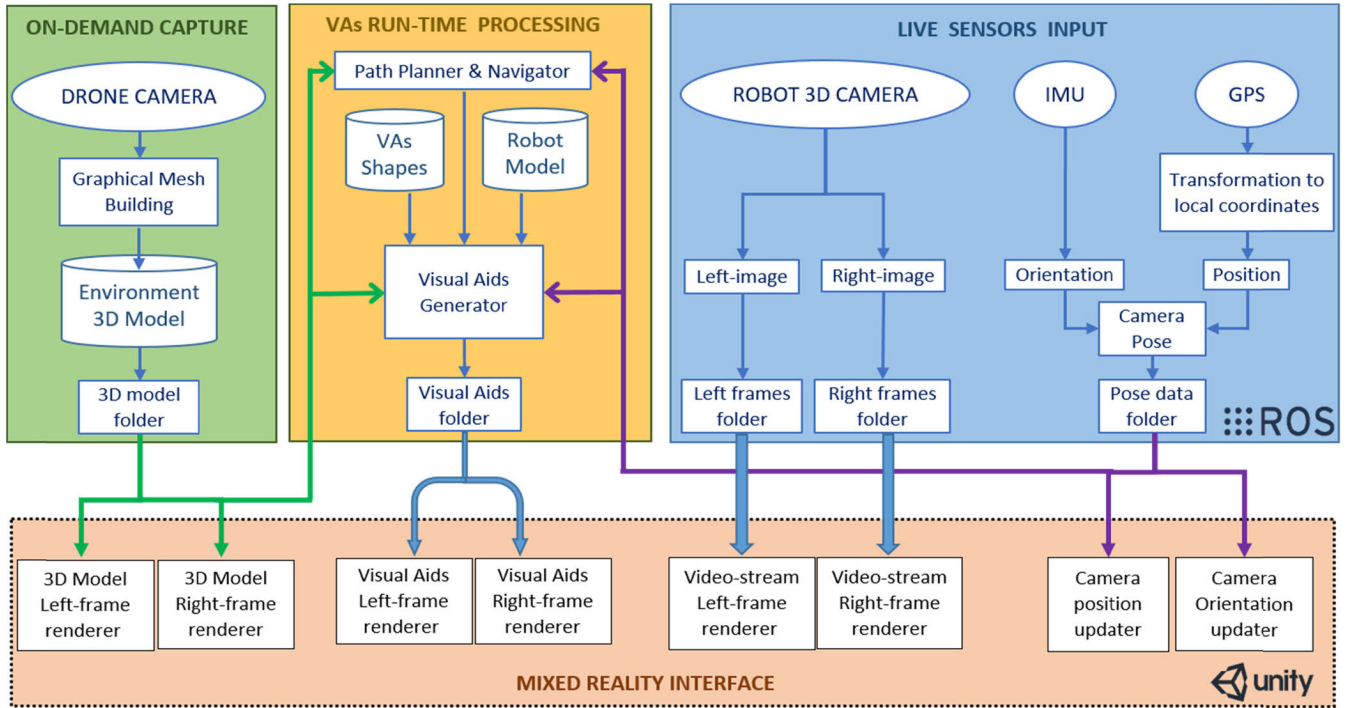
FIGURE 9. Main processing units with related blocks and data flow.

We cannot find in the literature an organized visual interface framework such the one proposed, which is aimed at enhancing user's comprehension through the described design elements and visual aids. The most relevant contribution is related to the intended application scenario, namely in uneven unstructured terrains, for which teleoperation interfaces have been scarcely investigated and developed.

IV. EXPERIMENTATION SYSTEM AND SETUP

Figure 1 gives an illustrated overview of the proposed system, while figure 9 shows the main processing units with the related blocks and data flow.

A. HARDWARE, SENSING AND VISUAL SCENES

In our experiments we use the *U-Go Robot* mobile platform, which is 75cm long, 88cm large, with 18cm wide rubbertracks. It is localized outdoor through the *Real-Time Kinematic (RTK) Differential GPS*. Its 3D attitude is acquired by the high-precision IMU *Xsense MTi*. A *ZED* stereo-camera connected to a *Jetson TX1* board sits to face the environment in front of the robot, while acquiring 15fps at 1280×720 px-res. High-level robot functions are managed by a *Raspberry Pi3* board, which runs the *Robot Operating System* (ROS) communication network. The ROS sets up distributed computing based on *TCP-IP* protocol and manages robot navigation. ROS drivers support the ZED cam. Figure 10 shows the system architecture.

The proposed teleoperation interface, built using the *Unity3D* software, contains environment textures mapped
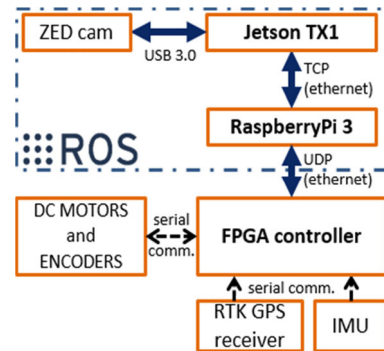

FIGURE 10. System architecture of the U-Go robot.

to an aerial photogrammetric reconstruction of the navigation area, obtained through the dedicated mapping software *Pix4D* [42]. A 2D map including traversing costs is derived from a terrain traversability analysis performed on the Digital Elevation Model (DEM), based on slope assessment and step detection [50]. This cost-map is matched against environment top-view. Our map pixel-size is 25cm. The Centerline VA is a polyline generated through multiple *checkpoint objects* laid over the 3D environment.

B. NAVIGATION, DATA COMMUNICATION AND CALIBRATION

The driver side of the teleoperation system consists of a PC remotely connected to the U-Go ROS network, which acts as robot-control station. The remote connection is

realized similarly to the work described in [56], thus letting the ROS middleware manage concurrencies and obtaining a reliable communication network. The remote operator wears an *Oculus Rift* HMD to get immersive MR views of robot and environment. The vehicle is guided through the *Oculus Touch* controllers. Alternatively, the user observes through a 27'' Acer Hn274h 2D/3D desktop monitor. *Unity3D* runs scripts and sends driving commands to robot [56]. During driving we store video frames, robot pose and associated timestamps. The operator's view is updated with latest sensor inputs and live-streamed environment images mapped on the reconstructed planes frontal to robot. The data-flow scheme (figure 9) also handles operator's run-time requests such as traversable areas (relying on the estimated 2D cost-map).

Camera images are aligned to environment 3D reconstruction automatically through the proposed image-stream mapping. Nonetheless, it is possible for an operator to make adjustments through the provided calibration tool. This helps overcome video-graphics mismatches. Calibration may take place at mission start and during teleoperation through the Touch controller. This feature is relevant because outdoor navigation causes vibrations and jumps, which may lead to camera-robot misalignments.

V. RESULTS AND DISCUSSION: SYSTEM FEASIBILITY AND INTERFACE EFFECTIVENESS

We wanted to gain an insight on overall system functionality and interface performance. We found it difficult to directly compare the proposed system with other state of the art proposals, such as e.g. those listed in Table 1. The reason being the different systems' setups, (such as the robot platform and the number of variables), the transmission delays and type of visual aids, which are difficult to equally recreate. We consequently aimed in our test trials:

- To confirm feasibility of the proposed systems and interface solutions, by implementing and running our interface on a real system setting.
- To check effectiveness of the proposed MR scenarios and visual aids, by asking users for their impression. Our qualitative assessment included observing sensors misalignments and visual aids usefulness. We also compared the use of VR headset with a more traditional desktop monitor.

Twenty-four trials took place on an outdoor field that included a number of navigation challenges: uneven ground, two different slopes, two uncrossable areas and limited ground visibility due to rich vegetation. A real robotic systems was used with maximum speed of 0.8 m/s. Figure 1 shows our test area through its 3D graphical reconstruction.

The test environment resembles two application scenarios: a terraced field, in the context of agricultural robotics, and a post-landslide scenario, where time-critical search and rescue operations have to be performed.

We asked 12 users to tele-operate the robot twice across the entire traversable area (after three minutes of practice), while either wearing a VR headset or observing the environment

in front of a 2D/3D desktop display. All users had some experience in using video-games, but not with robot teleoperation and VR. They followed a pre-determined schedule to counterbalance learning and fatigue effects. The linked video shows the system, all the VAs and a demo trial. Tests conformed to literature recommendations [57]–[59] and followed traditional approaches in terms of consent, forms and questionnaires [58]. Response to a questionnaire was provided based on the 7-point Likert scale $(-3,3)$, with '3' being the best score and '-3' the worst one.

There were 13 questions users were asked to answer with a score. The questions directly asked about the effectiveness of the 8 proposed VAs and of 4 viewing performance indicators (display suitability, used colours, image shaking and image misalignment). Users were carefully and equally explained the meaning of each question and scale values. We asked users to feel free to provide comments regarding our questions. In addition, we specifically asked to comment about sensor inconsistencies and video-graphic misalignments.

A. SENSOR DATA MISALIGNMENTS

Assessing communication challenges was outside our scope. Rather, we focused on evaluating the interface capability of delivering multiple inputs to operators (from GPS, IMU, Camera and 3D graphic model). We had an Internet connection with delays ranging from 1 to 4 secs, which is in line with several works in the literature [7], [31]. We exploited the data-associated timestamps in order to serve all inputs and response on first-in first-served schedule. The relevant outcome of open-handed questions and identified issues are discussed below:

- *Sensors Inconsistencies.* Differences among sensors in acquisition speed were of major interest in our test, as they would potentially affect image alignment and users' comfort. As for our choice sensor data were all transmitted as raw to save processing time. The GPS operated at 500ms, the IMU at 100ms, while live-camera images were sent at 15fps. The different sensor rates prevented us to opt for an encapsulation-based data transmission, as such approach would have bound us to the lowest data rate (i.e. the one of the GPS). The outcome indicated that discrepancies among sensors led to asynchronous inputs resulting in occasional images or VAs abrupt movements on display (image shaking). These were noted by 9 of our 12 users who judged them as minor and never critical (despite the uneven ground and lack of image-stabilization). The GPS lower update rate was one of the causes of inconsistencies despite mitigated through the use of a Kalman filter used as sensor fusion method and to combine acquired positions and odometry.
- *Video-Graphic Misalignment.* A consequence of sensor inconsistencies, but also errors and transmission delays, is misalignment in the MR image between video and graphical elements (responding to different sensors and data rates). During our tests we could occasionally

FIGURE 11. MR view during teleoperation toward the end of a trial. The superimposed red arrows indicate examples of misalignment between video and graphical elements.
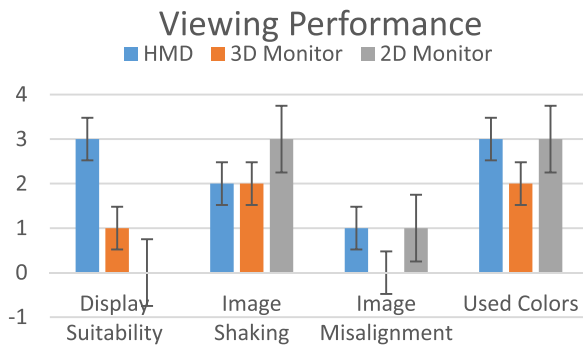
	Display Suitability	Image Shaking	Image Misalignment	Used Colors
HMD-3D	0.0001	0.0087	0.2261	0.1925
HMD-2D	0.0000	0.0001	0.0976	0.3799
3D-2D	0.0023	0.1070	0.0279	0.1251

FIGURE 12. Top-diagram: median scores and standard error of our qualitative user studies comparing user performance when tele-operating through a VR headset, 2D desktop, S3D desktop. The below-table shows T-Student *p*-values of coupled comparisons between the displays. The values in red indicate significant differences.

observe the actual vehicle attitude being different from the one communicated to operator's interface, which caused visual gaps between the texture-mapped graphical model and the camera streamed images. Figure 11 shows an example of image misalignment towards the end of a trial. The average score of the users for image misalignment was "good" (score 1) with 11 test-users judging that "incoming sensor and model information were received and displayed coherently". Occasionally image misalignments on attitude and position appeared (noted by all users), but they were not reported as relevant disturbance. We think this is a clear evidence that the robot hardware and the ROS network worked well and can be expected to do a good job under similar conditions and robot speed. The graphic model showed no delay in generating views; neither delays were perceived when mapping sensor data to VAs.

B. VIEWING PERFORMANCE

Figure 12 shows the outcome of our qualitative within subject user study. Users tele-operated the robot while either wearing the VR headset or observing through a less immersive

desktop screen. The latter included two viewing modalities: standard 2D and stereoscopic-3D (S3D). The diagrams show median values and standard error. The figures also include a table containing T-Student's *p*-values for coupled comparisons between displays.

- *Display Suitability*. The VR headset performed significantly better than any other screen, whereas the 3D monitor scored significantly better than its 2D version. This outcome clearly shows: (1) the contribution of having S3D viewing (HMD and 3D Desktop); (2) the contribution of having S3D viewing coupled to wide viewing angle (HMD). The higher sense of isolation is also a contributing factor.
- *Image Shaking*. This is generally commented as a minor issue. The significantly worse headset's performance compared to the monitors, and the similar scores of the two monitors, confirm the negative effect of the headset's greater involvement of user's peripheral vision.
- *Image Misalignment*. Visible gaps between image elements occurred mostly between streamed video and graphics, whereas graphical representations of sensor data and 3D model displayed coherently. The 2D monitor performed best. The reason being the misalignment were partially mitigated by the lack of depth awareness, which made them less noted or not perceived at all. The difference was significant only between the two monitors. The overall performance was lower than previous indicators, but still with a positive average.
- *Used Colors*. They were positively judged in terms of the chosen hues and their mapping to visual aids. They were, commented as well adapted to driving context (featuring green grass and brown terrain). Scores were among the highest, with no significant differences.
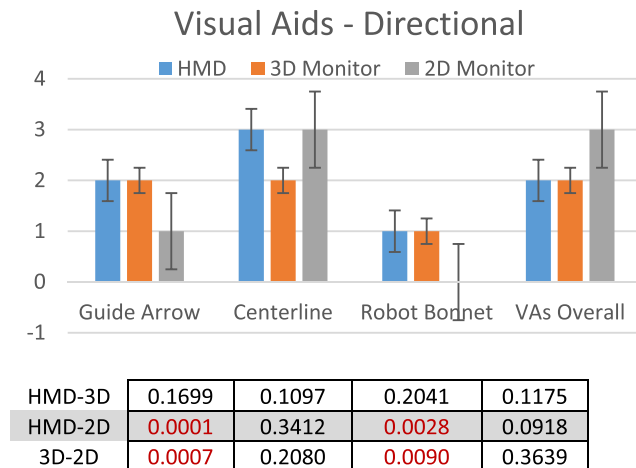
C. VISUAL AIDS

The figures 13 and 14 show the outcome of our qualitative analysis over VAs, analogously to figure 12.

- *Traversable Area*. It scored high on all displays with no significant difference among them. It was deemed needed for the specific type of environment, which featured reduced ground visibility because of the vegetation, and very helpful during planning and driving. Users stated this VA provided immediate comprehension facilitated by the use of colors and transparencies.
- *Extended Camera View*. The enhancement in terms of sense of presence was felt and positively judged. It was particularly appreciated on the headset because of the wide FOV. The graphic prevalence was judged excessive and suggested to be reduced. This VA scored significantly better on the VR headset than any other screen, whereas the 3D monitor scored significantly better than its 2D version.
- *Exocentric View*. It was commented as powerful and useful, but to require a couple of seconds to mentally adapt to the switch from egocentric to exocentric viewpoints. The graphic is deemed improvable to further increase

Visual Aids - Regional
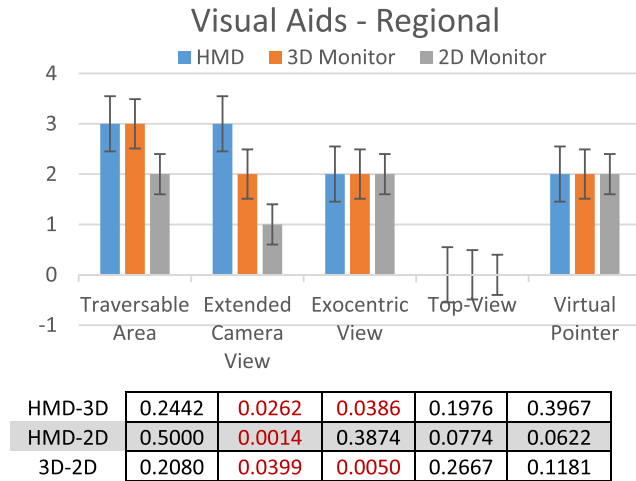
FIGURE 13. Top-diagram: median scores and standard error related to Regional Visual Aids. The below-table shows T-Student *p*-values of coupled comparisons between the displays (red values indicate significant differences).

HMD-3D	0.2442	0.0262	0.0386	0.1976	0.3967
HMD-2D	0.5000	0.0014	0.3874	0.0774	0.0622
3D-2D	0.2080	0.0399	0.0050	0.2667	0.1181

Visual Aids - Directional

FIGURE 14. Top-diagram: median scores and standard error related to Directional Visual Aids (and overall). The below-table shows T-Student *p*-values of coupled comparisons between the displays (red values indicate significant differences.

HMD-3D	0.1699	0.1097	0.2041	0.1175
HMD-2D	0.0001	0.3412	0.0028	0.0918
3D-2D	0.0007	0.2080	0.0090	0.3639

realism and presence, e.g. by including wheels' movement. The 3D monitor performed significantly better than both the other displays. 3D viewing was judged very helpful. It was less appreciated on the headset because of some occurring visual deformation.

- *Top-View*. This VA was judged occupying an excessive area of the overall view. Despite this, many users commented this aid was often ignored and lacked indication of the travelled path. A positive note was this VA being quicker to catch mentally than the exocentric view one. Scores were low with zero median value and no significant difference between screens.
- *Virtual Pointer*. It was commented as very useful to understand the surrounding environment during planning and under static conditions, whereas during navigation the generated occlusions sometime hindered

visibility. It scored high as median value on all screens, with no significant differences.

- *Guide Arrow*. This aid was judged to provide substantial help during navigation because it clearly indicated the driving direction. Users also commented its effectiveness was subject to hue choice, as the VA needs to stand out from current background. It was suggested its color should be adapted to vehicle's speed. The VA position on screen was judged suitable, and it was appreciated the option of changing it on demand. 3D viewing played a major role, as confirmed by the significantly better performance of headset and 3D monitor when compared to 2D monitor.
- *Centerline*. It was commented as the most needed help. It is intuitive and well indicated the position to hold during navigation and the path to follow. It also gave clear visibility to the underlying and surrounding environment. This VA achieved its highest scores on headset and 2D monitor, and slightly less on 3D monitor, with no significant differences between screens.
- *Robot Bonnet*. It was judged very helpful, but in need of some graphic improvements to its shape, to get maximum score. The 3D viewing made once again the greatest difference, especially in narrow passages, as it allowed users to clearly perceive the displacement between robot and closest obstacles. The headset and 3D monitors scored significantly better when compared to the 2D monitor.
- *VA Overall*. The results showed relevant variations among different visual aids and their effect on the tested displays. VAs were overall judged useful and providing real help for outdoor ground robot navigation. The VAs were judged to play a more relevant role in supporting navigation than the display. This we deem explained the high overall scores achieved by the 2D monitor and the non-significant difference between displays.

The proposed VAs, with the only exception of the *top-view*, were no doubt of great help to ground robot navigation. This was the case on any display. Their usefulness varied, with the regional VAs greatly appreciated for overview and planning, typically under static or low-motion conditions, whereas directional VAs were highly valued during motion.

The VR headset confirmed its great suitability over VAs that could exploit the wide FOV and head movement, therefore enhancing presence, and under static or nearly static conditions. This was particularly the case of the *extended camera view* and *exocentric view*. The VR headset also showed its advantage on both regional and directional VAs, in terms of 3D visualization. This was particularly the case for *guided arrow* and *robot bonnet*.

The 3D monitor was appreciated for enhancing depth perception, when compared to its equivalent 2D version. It was particularly appreciated for the *exocentric view*, *guide arrow* and *robot bonnet*. Testing on the 2D monitor was useful to see effectiveness of the VAs per se (regardless of the specific display). This was particularly shown by the *VA overall* high

scores and the non-significant difference between displays for *traversable area*, *virtual pointer, centerline* and *VA overall*.

VI. CONCLUSION

A new mixed reality visual context for robot teleoperation interfaces was proposed, aimed at improving performance by increasing operator's situational awareness. With the help of information visualization theories, related literature (and authors' experience), a way was devised to intuitively communicate available sensor information concurrently to streamed video input and environment knowledge.

The interface combined *immersive visualization*, *three-dimensional mixed reality* and *visual aids*; and its design included: *single-window representation*, *video-synthetic images*, *intuitive data viewing*, *regional and directional visual aids*. The use of specific visual aids was proposed, which were designed to best represent different sensor data within and around live video-images. Eight visual aids were proposed, classified as: Regional (*traversable area, extended camera, top view, exocentric view, and virtual pointer*) and Directional (*guide arrow, centerline, and robot bonnet*).

The proposed design was implemented on a real system that included: mobile platform (mobile robot with various sensors, a 3D-camera and processing unit), flying vehicle (drone with GPS and camera), and operator's unit (mixed reality interface, graphical processor and VR headset/3D-monitor). All were linked through a communication network.

The system was tested by twelve users through twenty-four practice trials on an uneven outdoor terrain presenting a few challenges. The outcome was very encouraging because of the positive feedback and acceptance given by all users towards the interface performance and visual aids.

All visual aids but one, were positively judged, deemed helpful and effective. Performance varied for different displays and viewing modalities, with the VR headset showing superior performance when either its wide FOV and head movement, or 3D viewing, could be exploited. This was the case with *extended camera view*, *exocentric view*, *guided arrow* and *robot bonnet*. The 3D monitor also showed good performance over its 2D version because of the enhanced depth-perception. Improvements were suggested on: *virtual pointer*, *guide arrow* and *robot bonnet*. While the *top view* was criticized in its current form.

Additionally, we plan to further mitigate sensor data mis-alignments and image shaking by respectively acting on the network, aiming at reducing communication delays, and on the robotic platform by introducing shock absorbers. As mentioned, communication related issues were not addressed in the present work, thus to be faced in future developments.

We think three-dimensional mixed reality is the future of teleoperation visual interfaces, which combined to visual aids has great potential in effectively conveying diverse sensor information visually. Three-dimensional mixed reality and visual aids also well marry with the use of a VR headset, which has now become mature technology.

REFERENCES

[1] M. R. Endsley, "Design and evaluation for situation awareness enhancement," in *Proc. Hum. Factors Soc. 32nd Annu. Meeting*, Los Angeles, CA, USA, 1988, pp. 97–101.

[2] K. Kruckel, F. Nolden, A. Ferrein, and I. Scholl, "Intuitive visual teleoperation for UGVs using free-look augmented reality displays," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Seattle, WA, USA, May 2015, pp. 4412–4417.

[3] A. Kelly, E. Capstick, D. Huber, H. Herman, P. Rander, and R. Warner, "Real-time photorealistic virtualized reality interface for remote mobile robot control," in *Robotics Research*, vol. 70. Berlin, Germany: Springer, 2011, pp. 211–226.

[4] D. Q. Huy, I. Vietcheslav, and G. S. G. Lee, "See-through and spatial augmented reality—A novel framework for human-robot interaction," in *Proc. 3rd Int. Conf. Control, Autom. Robot. (ICCAR)*, Nagoya, Japan, Apr. 2017, pp. 719–726.

[5] S. Livatino, G. Muscato, and F. Privitera, "Stereo viewing and virtual reality technologies in mobile robot teleguide," *IEEE Trans. Robot.*, vol. 25, no. 6, pp. 1343–1355, Dec. 2009.

[6] J. A. Frank, S. P. Krishnamoorthy, and V. Kapila, "Toward mobile mixed-reality interaction with multi-robot systems," *IEEE Robot. Autom. Lett.*, vol. 2, no. 4, pp. 1901–1908, Oct. 2017.

[7] S. Livatino, F. Banno, and G. Muscato, "3-D integration of robot vision and laser data with semiautomatic calibration in augmented reality stereoscopic visual interface," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 69–77, Feb. 2012.

[8] J. Xiao, P. Wang, H. Lu, and H. Zhang, "A three-dimensional mapping and virtual reality-based human–robot interaction for collaborative space exploration," *Int. J. Adv. Robotic Syst.*, vol. 17, no. 3, May 2020, Art. no. 172988142092529.

[9] N. Zaman, A. Tavakkoli, and, C. Papachristos, "Tele-robotics via an efficient immersive virtual reality architecture," in *Proc. 3rd Int. Workshop Virtual, Augmented, Mixed Reality HRI*, Cambridge, U.K., 2020.

[10] T. Kot, P. Novak, and J. Bajak, "Using HoloLens to create a virtual operator station for mobile robots," in *Proc. 19th Int. Carpathian Control Conf. (ICCC)*, Szilvasvarad, Hungary, May 2018, pp. 422–427.

[11] S. Livatino, L. T. De Paolis, M. D'Agostino, A. Zocco, A. Agrimi, A. De Santis, L. V. Bruno, and M. Lapresa, "Stereoscopic visualization and 3-D technologies in medical endoscopic teleoperation," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 525–534, Jul. 2015.

[12] J. Jankowski and A. Grabowski, "Usability evaluation of VR interface for mobile robot teleoperation," *Int. J. Hum.-Comput. Interact.*, vol. 31, no. 12, pp. 882–889, Dec. 2015.

[13] A. Nayyar, V. Puri, N. Nhu, and D. N. Le, "Smart surveillance robot for real-time monitoring and control system in environment and industrial applications," in *Information Systems Design and Intelligent Applications*. Singapore: Springer, 2018, pp. 229–243.

[14] R. S. Batth, A. Nayyar, and A. Nagpal, "Internet of robotic things: Driving intelligent robotics of future–concept, architecture, applications and technologies," in *Proc. 4th Int. Conf. Comput. Sci. (ICCS)*, Jalandhar, Punjab, Aug. 2018, pp. 151–160, doi: 10.1109/ICCS.2018.00033.

[15] I. E. Sutherland, "A head-mounted three dimensional display," in *Proc. Fall Joint Comput. Conf.*, vol. 1968, Dec. 1968, pp. 757–764.

[16] *Z800 3DVisor*. eMagin. Accessed: May 2, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Z800_3DVisor

[17] *Oculus Rift DK1*. Facebook. Accessed: Dec. 12, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Oculus_Rift#Development_Kit_1/

[18] Google. *Google Cardboard*. Accessed: Nov. 2020. [Online]. Available: https://arvr.google.com/cardboard/

[19] *Samsung/Oculus, Samsung Gear VR*. Accessed: Nov. 2020. [Online]. Available: https://www.oculus.com/gear-vr/

[20] *HTC Vive Focus Plus*. Accessed: Dec. 12, 2020. [Online]. Available: https://enterprise.vive.com/uk/product/focus-plus/

[21] *Oculus, Oculus QUEST*. Accessed: Nov. 2020. [Online]. Available: https://www.oculus.com/quest/

[22] A. Nayyar, B. Mahapatra, D. Le, and G. Suseendran, "Virtual reality (VR) & augmented reality (AR) technologies for tourism and hospitality industry," *Int. J. Eng. Technol.*, vol. 7, nos. 2–21, pp. 156–160, 2018.

[23] *Oculus Rift S*. Facebook. Accessed: Dec. 12, 2020. [Online]. Available: https://www.oculus.com/rift-s/

[24] *HTC Vive Pro*. Accessed: Dec. 12, 2020. [Online]. Available: https://www.vive.com

[25] *HTC Vive Pro Eye*. Accessed: Dec. 12, 2020. [Online]. Available: https://www.vive.com/uk/product/vive-pro-eye/overview/

[26] *FOVE VR Eye-Tracker*. Accessed: Dec. 12, 2020. [Online]. Available: https://fove-inc.com/

[27] *Best 3D Monitors is 2020*. Accessed: Dec. 12, 2020. [Online]. Available: https://www.youtube.com/watch?v=WBiJFGZ6XqA

[28] *BenQ PD3200U 32' 4K*. Accessed: Feb. 3, 2020. [Online]. Available: https://www.amazon.co.uk/dp/B01MY142C0?tag=sushee-21

[29] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Trans. Inf. Syst.*, vol. 77, no. 12, pp. 1321–1329, Dec. 1994.

[30] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks, "Ecological interfaces for improving mobile robot teleoperation," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 927–941, Oct. 2007.

[31] P. Stotko, S. Krumpen, M. Schwarz, C. Lenz, S. Behnke, R. Klein, and M. Weinmann, "A VR system for immersive teleoperation and live exploration with a mobile robot," *2019, arXiv:1908.02949*. [Online]. Available: http://arxiv.org/abs/1908.02949

[32] F. Okura, Y. Ueda, T. Sato, and N. Yokoya, "Teleoperation of mobile robots by generating augmented free-viewpoint images," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Tokyo, Japan, Nov. 2013, pp. 665–671.

[33] P. U. Lima, "Search and rescue robots: The civil protection teams of the future," in *Proc. 3rd Int. Conf. Emerg. Secur. Technol.*, Lisbon, Portugal, Sep. 2012, pp. 12–19.

[34] R. P. Saputra and P. Kormushev, "Resqbot: A mobile rescue robot with immersive teleperception for casualty extraction," in *Proc. 19th Int. Conf. Towards Auton. Robotic Syst.*, Bristol, U.K., 2018, pp. 209–220.

[35] A. Tikanmaki, T. Bedrnik, R. Raveendran, and J. Roning, "The remote operation and environment reconstruction of outdoor mobile robots using virtual reality," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Takamatsu, Japan, Aug. 2017, pp. 1526–1531.

[36] J. J. Roldan, E. Pena-Tapia, P. Garcia-Aunon, J. D. Cerro, and A. Barrientos, "Bringing adaptive and immersive interfaces to real-world multi-robot scenarios: Application to surveillance and intervention in infrastructures," *IEEE Access*, vol. 7, pp. 86319–86335, 2019.

[37] G. Adamides, C. Katsanos, Y. Parmet, G. Christou, M. Xenos, T. Hadzilacos, and Y. Edan, "HRI usability evaluation of interaction modes for a teleoperated agricultural robotic sprayer," *Appl. Ergonom.*, vol. 62, pp. 237–246, Jul. 2017.

[38] H. Yanco, J. Drury, and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *Hum.-Comput. Interact.*, vol. 19, no. 1, pp. 117–149, Jun. 2004.

[39] D. J. Bruemmer, D. A. Few, R. L. Boring, J. L. Marble, M. C. Walton, and C. W. Nielsen, "Shared understanding for collaborative control," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 35, no. 4, pp. 494–504, Jul. 2005.

[40] M. Baker, R. Casey, B. Keyes, and H. A. Yanco, "Improved interfaces for human-robot interaction in urban search and rescue," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Hauge, The Netherlands, Oct. 2004, pp. 2960–2965.

[41] J. J. Gibson, *The Ecological Approach to Visual Perception*. Boston, MA, USA: Houghton Mifflin, 1979.

[42] Pix4D. *Pix4Dmapper*. Accessed: Nov. 2020. [Online]. Available: https://pix4d.com/product/pix4dmapper-photogrammetry-software/

[43] R. Mazza, *Introduction to Information Visualization*. London, U.K.: Springer-Verlag, 2009.

[44] D. Chang, L. Dooley, and J. E. Tuovinen, "Gestalt theory in visual screen design—A new look at an old subject," in *Proc. World Conf. Comp. Ed.*, Copenhagen, Denmark, vol. 8, 2001, pp. 5–12.

[45] P. M. Lester, *Visual Communication: Images With Messages*. Belmont, CA, USA: Thomson Wadsworth, 2006.

[46] L. W. MacDonald, "Using color effectively in computer graphics," *IEEE Comput. Graph. Appl.*, vol. 19, no. 4, pp. 20–35, Jul/Aug 1999.

[47] R. Williams, *The Non-Designer's Design Book: Design and Typographic Principles for the Visual Novice*. London, U.K.: Pearson, 2015.

[48] B. L. Harrison, G. Kurtenbach, and K. J. Vicente, "An experimental evaluation of transparent user interface tools and information content," in *Proc. 8th Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, Pittsburgh, PA, USA, 1995, pp. 81–90.

[49] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proc. IEEE Symp. Vis. Lang.*, Boulder, CO, USA, Sep. 1996, pp. 336–343.

[50] D. C. Guastella, L. Cantelli, C. D. Melita, and G. Muscato, "A global path planning strategy for a UGV from aerial elevation maps for disaster response," in *Proc. 9th Int. Conf. Agents Artif. Intell.*, 2017, pp. 335–342.

[51] J. Li and X. Fan, "Outdoor augmented reality tracking using 3D city models and game engine," in *Proc. 7th Int. Congr. Image Signal Process.*, Dalian, China, Oct. 2014, pp. 104–108.

[52] M. A. Al-Mouhamed, O. Toker, and A. Al-Harthy, "A 3-D vision-based Man–Machine interface for hand-controlled telerobot," *IEEE Trans. Ind. Electron.*, vol. 52, no. 1, pp. 306–319, Feb. 2005.

[53] C. P. Quintero, R. T. Fomena, A. Shademan, N. Wolleb, T. Dick, and M. Jagersand, "SEPO: Selecting by pointing as an intuitive human-robot command interface," in *Proc. IEEE Int. Conf. Robot. Autom.*, Karlsruhe, Germany, May 2013, pp. 1166–1171.

[54] S. Burigat and L. Chittaro, "Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids," *Int. J. Hum.-Comput. Stud.*, vol. 65, no. 11, pp. 945–958, Nov. 2007.

[55] T. T. H. Nguyen, T. Duval, and C. Fleury, "Guiding techniques for collaborative exploration in multi-scale shared virtual environments," in *Proc. Int. Conf. Comp. Graph. Theory Appl.*, Barcelona, Spain, 2013, pp. 327–336.

[56] E. Rosen, D. Whitney, E. Phillips, D. Ullman, and S. Tellex, "Testing robot teleoperation using a VR interface with ROS reality," in *Proc. 1st Int. Workshop Virt. Augm. Mix. Real. (HRI)*, 2018, pp. 1–4.

[57] J. Rubin, *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests*. Hoboken, NJ, USA: Wiley, 1993.

[58] J. Nielson, *Usability Engineering*. San Mateo, CA, USA: Morgan Kaufmann, 1993.

[59] D. J. Kasik, J. J. Troy, S. R. Amorosi, M. O. Murray, and S. N. Swamy, "Evaluating graphics displays for complex 3D models," *IEEE Comput. Graph. Appl.*, vol. 22, no. 3, pp. 56–64, May 2002.

SALVATORE LIVATINO received the M.Sc. degree in computer science from the University of Pisa, Italy, in 1993, and the Ph.D. degree in computer science and engineering from Aalborg University, Denmark, in 2003. He was a Researcher with the Scuola Superiore Sant'Anna, Pisa, from 1993 to 1997, with the University of Leeds, U.K., in 1995, with INRIA Grenoble, France, in 1996, and the University of Edinburgh, U.K., in 2001. He worked for 12 years at Aalborg University, first as Research Fellow, then as an Assistant Professor, and finally as an Associate Professor. He is currently a Reader in Virtual Reality and Robotics with the University of Hertfordshire, Hatfield, U.K. His teaching experience has mostly been within problem-based learning and multidisciplinary education. He is the author of several journal and conference papers, and has contributed to many European and U.K., projects. His research interests include virtual and augmented reality user interfaces for tele-exploration and tele-operation, and focus on the use of stereocopic-3D visualization and immersive technology, computer vision and graphics algorithms, with applications in the field of telerobotics, telemedicine, and control panels and dashboards.

DARIO C. GUASTELLA (Associate Member, IEEE) received the master's degree in automation engineering and the Ph.D. degree in systems and computer engineering from the University of Catania, Italy, in 2015 and 2019, respectively. He is currently a Postdoctoral Research Fellow with the Robotic Systems Group, Department of Electrical, Electronics, and Computer Engineering, University of Catania. His research interests include cooperative mobile robots (both ground and aerial vehicles), terrain traversability analysis and Artificial Intelligence for autonomous navigation.

GIOVANNI MUSCATO (Senior Member, IEEE) received the degree in electrical engineering from the University of Catania, Catania, Italy, in 1988. After completing graduation, he was with the Centro di Studi sui Sistemi, Turin, Italy. In 1990, he joined the DIEEI University of Catania, where he is currently a Full-Time Professor of robotics and automatic control and since 2018, the Director of the Department. His current research interests include service robotics and the cooperation between ground and flying robots. He was the Coordinator of the EC project Robovolc and is the local coordinator of several national and European projects in robotics. He is the author of more than 300 papers in scientific journals and conference proceedings and three books in the fields of control and robotics. He is with the Board of Trustees of the Climbing and Walking Robots (CLAWAR) Association.

VINCENZO RINALDI received the B.S. degree and the M.S. degree in computer engineering from the University of Catania, Italy, in 2015 and 2018, respectively. From 2018 to 2019, he was a Research and Development Engineer with Arm23 SRL, Catania, Italy, focusing on the development of Mixed Reality applications. Since 2020, he has been a VR/AR Application Specialist with the Leverhulme Research Centre for Forensic Science, University of Dundee, Dundee, U.K. His work aims to aid research into the assessment of imaging techniques used for crime scene investigation, also developing novel methods of navigating and comparing 3D reconstructions in Virtual Reality.

LUCIANO CANTELLI received the M.S. degree in computer science engineering and the Ph.D. degree in electronic and automation engineering from the University of Catania, Italy, in 2003 and 2007, respectively. He is currently a Research Fellow in robotics and automatic control with the Dipartimento di Ingegneria Elettrica Elettronica e Informatica (DIEEI), University of Catania. Since 2003, he has been involved as Researcher in several national and European projects in robotics, such as the EC projects: ROBOVOLC (robot for volcano explorations), RAPOLAC (rapid production), and TIRAMISU (humanitarian demining). His research interests include industrial and service robotics, navigation and location systems for mobile robotics, mechatronics, multi-sensor data fusion, robotic assistive technologies, and bioengineering. He is a Co-Founder of the Etnamatica S.r.l, a Robotic Spin-Off company, in which he covers the role of the CEO and the Research and Development Manager.

CARMELO D. MELITA received the Ph.D. degree in electronics, automation, and complex systems control engineering from the University of Catania, thesis title: Unmanned Aerial Systems for Volcanic Sites Inspection'. From February 2009 to 2011, he was a Research Fellow with University of Catania to develop the subject Robots control methodologies and performances evaluation. From September 2005 to January 2006, he worked developing control algorithms and software for the Spiderbot robot, a climbing robot for the inspection of industrial tanks. He has been involved in several National and European projects (ROBO-VOLC, MOW-BY-SAT, and TIRAMISU). His recent research interests include the control and navigation topics of autonomous mobile robots (ground, flying, and underwater). In 2011, he founded Etnamatica srl: he is currently the CTO of the company.

ALESSANDRO CANIGLIA received the B.Sc. and M.Sc. degrees in computer engineering from the University of Catania, Italy, in 2008 and 2011, respectively. In 2010, he was a Visiting Researcher with the University of Hertfordshire, under the supervision of Dr. S. Livatino, and worked on a design and development of tele-guide systems for mobile robots using Augmented/Virtual Reality user interface. He is currently working with Microsoft as an Azure Technical Advisor and a Trainer.

RICCARDO MAZZA received the degree in computer sciences from the University of Pisa, Italy, in 1997, and the Ph.D. degree in communication sciences from the University of Lugano, Switzerland, in 2004, with a dissertation on the visual representation of students' data in Web-based distance education, pioneering the research in learning analytics already in 2001. From 1997 to 2020, he was a Researcher with the Institute of Communication Technologies, University of Lugano. Since 1999, he has been a Lecturer and a Researcher with the Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland (SUPSI). He has also been involved in a number of national and European research projects. He published the book *Introduction to Information Visualization* (Springer-Verlag, 2009), one of the few reference books in the discipline. His main research interests include information and data visualization, learning analytics, and distance education.

GIANLUCA PADULA received the B.Sc. degree in electrical engineering from the University of Catania, Italy, in 1996, the M.Sc. degree in biomedical engineering from the Polytechnic of Milan, Italy, in 2004, and the Ph.D. degree in biophysics and molecular biology from the Medical University of Lodz, Poland, in 2010. Since then, he has been working with the Medical University of Lodz, Poland, first as a Lecturer in physics, cardiology, and biophysics, and then as a Specialist in experimental and clinical physiology. In 2012, he became the Director of the Academic Laboratory of Movement and Human Physical Performance, Dynamolab, Medical University of Lodz. Since 2020, he has also a member of the Committee of Rehabilitation, Physical Education, and Social Integration of the Polish Academy of Science, Poland.

● ● ●