The design and evaluation of an interface for teleoperation through eye gaze (TeleGaze)

Hemin Omer Latif, Student Member, IEEE, Nasser Sherkat, and Ahmad Lotfi, Senior Member, IEEE

ABSTRACT

In many teleoperation applications the operator is required to control the robot and monitor the status of the robot simultaneously from a remote location. Assuming a vision based feedback system is used for monitoring, the operator is required to closely monitor the images in real time while controlling the robot. This requires the eyes of the operator to be engaged in the monitoring task and the hands in the controlling task for the whole duration of the operation. Considering the fact that eyes are already involved in the monitoring task, inputs from their gaze can be used to aid in the controlling task to free operator’s hands from this task. However, the challenge here lies in distinguishing between the inputs that should be used for monitoring and those that should be used for controlling. To address this challenge, this paper presents a novel means of developing an interface for teleoperation through eye gaze or what is called TeleGaze hereafter. The interface developed in this work enables the operator to monitor as well as control a mobile robot remotely through the use of human eyes only. The TeleGaze interface was developed out of two different prototypes based on user satisfaction. Then, a task oriented evaluation for the interface was conducted in comparison with other interaction modalities for two different applications. Overall, it is found that TeleGaze provides efficient means of successfully teleoperating a mobile robot for a number of navigational tasks.

Index Terms - Eye Tracking, Human-Robot Interaction, Mobile Robots Motion Planning, Teleoperation
I. INTRODUCTION

Developing the required collaboration between human beings and robotic agents, which is mostly known as human-robot interaction, is one of the main remaining challenges in robotics [1]. Although, a fully autonomous robotic agent is the aim of a number of researchers in the field of robotics many other researchers, in contrast, are more interested in controllable agents [2]. This might be due to the belief that fully autonomous agents within real scenarios are not possible yet. It might also be due to the importance of the role of human beings in the cycle of human-robot interaction [3].

Whatever the reason might be, a wide range of these controllable agents require controlling from a remote location, which is mostly known as telerobotics or teleoperation [4]. Therefore, teleoperation as a mean of providing collaboration between human beings and robotic agents remains a widely addressed challenge in a variety of robotic applications up to the present day. Significant amount of efforts has been devoted to developing different teleoperation interfaces for different scenarios and applications [5].

Eye tracking, on the other hand, is entering its fourth era [6] and research is ongoing in developing faster, more reliable and easier to use eye tracking systems [7, 8]. Despite the current limitations of eye tracking systems, communication through gaze has been a goal for many years. With the ongoing research and the advance in technology, many usable systems can be seen for a range of applications [9-11]. Furthermore, real benefits are expected from gaze-based communication as systems become more able to make decisions about user intentions [12]. Therefore, the domain of eye tracking applications is expanding rapidly including both diagnostic and interactive applications [6]. As part of the development of interactive applications, inputs from the eyes have been used in developing a number of selective user interfaces for human-computer interaction [13]. Hence, a few attempts in using eye tracking for human-robot interaction can be seen as well [14].

Reducing the amount of body engagement in any human-robot interaction application, including teleoperation, would provide many other opportunities for the operator. Most teleoperation applications
require the operator to continuously monitor the status of the robot through a feedback system mechanism. This is regardless of the nature of the feedback system and the amount of information the operator needs to acknowledge in order to make the right decisions. Hence, the eyes of the operator need to be engaged in this monitoring task throughout the whole duration of the operation. Meanwhile, the hands of the operator need to be engaged, either partially or fully, in performing the commanding task using input devices. Since the eyes of the operator are engaged in the monitoring task, inputs from their gaze can be used to aid in the commanding task. This will free, again either fully or partially, the operator’s hands from this task.

To put this concept into test, navigating a mobile robot might be the most intuitive task to be implemented through human eyes because people mostly look where they want to go [15]. However, despite the correlation between the directions of the gaze and the intentions of movements, teleoperation through eye gaze is not free of challenges. The problem of modeling humans’ intentions and interpreting the inputs from the eyes into moving commands intuitively is quite challenging and the solution to this problem is not a trivial aim to achieve [16]. Therefore, the aim of this research is to develop a human-robot interaction interface that provides an intuitive teleoperation through eye gaze. In this research teleoperation through eye gaze is referred to as TeleGaze. Therefore, this acronym is used throughout this paper.

The motive behind TeleGaze, as it was mentioned earlier, is to reduce the amount of body engagement in performing any teleoperation task. Therefore, the interface that will be developed for TeleGaze, ideally, should work as naturally and as intuitively as possible. This constrain is so important to be taken into account that bypassing it might result in loosing the central concept of TeleGaze. TeleGaze should not increase the amount of workload on the eyes in order to decrease the amount of body engagement by for instance freeing the hands of the operator. However, this might not be achievable with a non-natural representation of the feedback information such as interfaces cluttered with maps, charts, and figures. Fortunately, the most widely used feedback system in current teleoperation applications is a stream of real-time images from cameras mounted on the robotic platform [17]. This natural representation of the
robot’s environment and status are the hope and motive behind the aim of creating natural interaction with the human operator.

To address the issues mentioned above this paper is organized as follows: in section 2 some related works are discussed followed by the introduction to the conceptual design of TeleGaze with the apparatus in section 3. Section 4 then presents the TeleGaze interface with a number of design considerations. In section 5 an observational study is described for two different prototypes leading to a refined prototype for the TeleGaze interface. Section 6 covers a task oriented evaluation for the interface and conclusions are drawn in section 7.

II. RELATED WORKS

The idea of using inputs from eyes in controlling robotic agents is relatively novel and only a few attempts can be found in the literature. Although the focus of most reported works is far from the development of an interaction interface for teleoperation through eye gaze, relevancy to this work can be found in the literature and a number of interesting attempts reported here.

In [18], a robotic arm was controlled using an experimental eye tracking algorithm. The interface provided the operator with feedback through images from a camera mounted in a fixed location. The interface was divided into a feedback region and a controlling region with a number of buttons. The focus of this work was more towards developing an eye tracking algorithm and testing it rather than developing a human-robot interaction interface using inputs from the eyes.

In [19], a robotic arm was controlled using an experimental eye tracking algorithm. The interface provided the operator with feedback through images from a camera mounted in a fixed location. The interface was divided into a feedback region and a controlling region with a number of buttons. The focus of this work was more towards developing an eye tracking algorithm and testing it rather than developing a human-robot interaction interface using inputs from the eyes.

An active vision system equipped with a pair of zoom cameras was presented in [19] and later in [20] to be controlled by an operator. The authors used skin color for face detection and ported a gaze tracking algorithm. The gaze tracking module however, had the priority over the face tracking in controlling the active head. They extended this work in [21] by implementing their active vision system in combination with a robotic arm to pick up an arbitrary object. They demonstrate the ability of the vision system to determine the object that the user is gazing at and then control the robotic arm to pick up the object.
In [22] various kinds of human-robot interface were developed to provide a range of services by a robotic arm for people with disabilities. As part of this work, the direction of the gaze was obtained through a developed eye tracking cap and used to interact with a menu-driven interface which was called the eye-mouse interface. The interface was equipped with a number of menus to control the robotic arm as well as controlling the direction of a pan/tilt camera. The user positioned an object in the centre of the scene and the interface calculated the 3D position of the object of interest which then was considered as the course 3D position of the object by the robot.

Recent research is reported in [23] into grasping an object using a robotic arm with the aid of inputs from the eyes. To guess the intention of the remote operator in grasping a particular object, a learning ability for a robotic arm was developed. Based on the data coming from the operator’s gaze and hand positions when looking at a monitor and contextually controlling the hand, the robotic arm was able to learn that the user’s intention is to grasp a particular object. The purpose of this work was to minimize bandwidth requirements for the teleoperation by transferring part of the processing required to the robotic arm after the learning process was done. No actual experiments, however, were reported in this work.

In an effort to develop smoother control of a remote-controlled humanoid robot, the authors of [24] presented the operator with a 3D sense of the remote environment by installing a set of two cameras on the robot. However, to minimize the amount of parallax for the object in the stereoscopic view, the operator was required to select the object of interest in a particular scene. A conventional way to do this is a mouse click, but to avoid the risk of manipulations, the mouse click required the operator to switch off the power of the manipulators, change to the mouse and select the object of interest, and then switch on the powers of the manipulators back. This would have increased the time that would be required to complete a task and increase the complexity of the manipulation. Therefore, eye-tracking was used to substitute the mouse click with an eye blink which saved the operator an amount of time and the effort of switching on and off the manipulators. Although, it was reported that the difference in the time of the task completion is not
significant, the authors mention the advantage of the work in reducing the amount of effort that is required by the operator.

To control a powered wheelchair through human eye gaze a controlling interface was developed in [25]. The interface, which was divided into four command regions and five non-command regions, allowed the operator to drive the wheelchair using inputs from the eyes. For the same purpose of controlling a powered wheelchair, earlier in [26] electrooculography was used to issue driving commands using an interface. In both of these works, as the operator was sitting on the powered wheelchair, no feedback information was presented to the user through the interface. Therefore, the interface was totally used as one way communication.

To experiment an eye tracking algorithm, a robotic toy was controlled in [14]. However, in this work, the location of the eye ball within the eye socket was used. Because the robot did not react based on the point of the gaze on the interface, the interface was used only to provide feedback to the operator.

The focus of the research work reported in this paper is the development of an interface for TeleGaze rather than an eye tracking mechanism. Furthermore, considering an intuitive interpretation of the inputs from human eyes to navigate a mobile robot from a remote location shapes the aim of this research. The operation of the TeleGaze is described in the next section.

III. TELEGAZE: TELEOPERATION THROUGH EYE GAZE

A. TeleGaze Conceptual Design:

A clear conceptual design for any interactive system is vital in delivering the required results [27]. The conceptual design of the TeleGaze system is shown in Fig. 1. The TeleGaze system consists of three sub-systems; namely robotic, eye tracking and interaction sub-systems. Robotic and eye tracking sub-systems are two easily separable sub-systems linked together with the TeleGaze interface and the software behind it.

Eye tracking sub-system provides the interaction sub-system with the coordinates of the operator’s
gaze on the interaction screen. This information then is interpreted into controlling commands and sent to the robotic sub-system according to a number of predefined interaction criteria that the interaction sub-system follows. Therefore, it can be said that the interaction sub-system works as a meeting point for the data flow. The interaction sub-system gets data coming from both ends of the whole TeleGaze system and feeds back both ends of the system with data. The data flow of the TeleGaze system is shown in Fig. 2.

The robotic sub-system consists of the locomotion controller and a vision based feedback component. The locomotion controller gets controlling commands from the interaction sub-system and implements them as robotic movements. The feedback component, which in this work is PTZ (Pan/Tilt/Zoom) cameras, sends real time images back to the interaction sub-system to be displayed to the operator using the TeleGaze interface.

**B. Apparatus:**

One of the main objectives of this work is to develop a platform-independent interface. A platform-independent interface would adapt to any robotic platform as long as it provides vision based feedback. It also adapts to any eye tracking equipment as long as it provides the same types of inputs from the eyes of the operator. In this section, however, the experimental platform used to test the TeleGaze interface is briefly described. The experimental platform, which is illustrated in Fig. 3., consists of a Wi-Fi enabled mobile robot as the robotic sub-system, an eye tracking equipment as the eye tracking sub-system and a teleoperation station which runs the interaction sub-system. The teleoperation station is where the operator is located and navigating the mobile robot remotely.

A commercial eye tracking equipment shown in Fig. 3a from the Applied Science Laboratories [11] was selected for the purpose of this experiment. This particular type of eye tracking equipment was selected because it uses techniques based on reflected light from the operator's eyes which is believed to be appropriate for this kind of interaction applications [13]. Although these systems are non-invasive and reasonably accurate the drawbacks are the requirements to keep the head still and also the difficulty to keep
a good contrast image. Therefore, the system is designed so that two people get involved in any experiment; one user and one supervisor. The supervisor needs to keep monitoring the status of the system and adjusting a number of thresholds in real time throughout the tracking period. Furthermore, a nine point calibration for each tracking session is required prior to commencing any experiments. This type of calibration is required in order to obtain more accuracy from the system [28].

The mobile robot shown in Fig. 3c is a modified wheelchair base equipped with a vision system consisting of two PTZ cameras from VIVOTECK [29]. Controlling a robot from a remote location requires awareness of any obstacles in the close surroundings of the robot. To gain this awareness, one of the cameras is mounted looking downward. This helps the operator to get an idea of the distance between the robot and any obstacles in the vicinity of the robot. The physical setup of the cameras is meant to provide the user with sufficient situational awareness about the remote location. A small view of the downward looking camera is displayed on the TeleGaze interface to the operator. The mobile robot is equipped with an onboard PC that maintains Wi-Fi connectivity with the teleoperation station. The teleoperation station is a normal PC running the interaction sub-system with which the operator interacts via the TeleGaze interface shown in Fig. 3b.

**IV. INTERFACE DESIGN CONSIDERATIONS**

One of the main objectives of the TeleGaze interface developed in this work is to build a two way communication channel. This is in order to enable monitoring and controlling at the same time through the use of the same space. Therefore, the concept of focus areas [27] is used to augment the images on the interface by creating, what are called in this work, *action regions*. The action regions are transparent regions on top of the images that are displayed to the operator rather than button-like shapes placed somewhere on the interface. The remaining of the interface area is called *no-action regions*. This is summarized in the following expression:
\{Interface (x,y)\} = \{AR(x,y)\} \cup \{NoAR(x,y)\}  \quad (1)

where:
\{Interface(x,y)\} is the set of pixels that makes the whole interface.
\{AR(x,y)\} is the set of pixels that makes the action regions.
\{NoAR(x,y)\} is the set of pixels that makes the no-action regions.

A relatively big size for the action regions was chosen in comparison with conventional command buttons on other graphical user interfaces. This is because pointing in small displays with high resolution is not possible given the status of the current eye tracking technology [13]. Depending on the requirements of the application, the size of the action regions can be chosen based on the performance of the eye tracking equipment for each individual user [22]. However, the transparency of the action regions in this work adds to their functionality in that it provides both monitoring and controlling. Hence, better use of space is achieved on the interface. Figs. 4 and 5 show the layout of the TeleGaze interface and an actual snapshot of it, respectively.

The nature of the inputs from the eyes differs from that of any other devices. Eyes are always engaged and therefore, a non-stopping stream of inputs needs to be dealt with during the course of the teleoperation. Picking up what actually should issue a command out of a continuous stream of inputs creates one of the biggest problems in any eye tracking application. This problem is well known as the -Midas Touch- problem [30]. Although a number of researchers recommend a multi modal interaction to overcome this problem [22], one of the proposed solutions in the research community is dwell time. Dwell time is keeping the gaze within a particular area continuously for a period of time to substitute for the act of clicking.

In general, the inputs from the eyes can be classified into eye movements and fixations. Limited information can be acquired from some types of eye movements due to the speed of the movements [31]. For the purpose of this work, however, it is fixations that are of major concern. Fixations can be defined as fixing the gaze on a particular point for viewing [32]. Also fixations that occur in the no-action regions are
of no interest in this application and will not be taken into any further considerations. It is only fixations that occur in the action regions which are considered for further processing. This is summarized in the following expression and illustrated in Fig. 1.

\[
\text{fix}(x,y) = \begin{cases} 
0 & \text{if } \text{fix}(x,y) \in \{\text{NoAR}(x,y)\} \\
\text{value} & \text{if } \text{fix}(x,y) \in \{\text{AR}(x,y)\}
\end{cases}
\]  

(2)

Whenever \(\text{fix}(x,y) = \text{value}\), then it is checked against expression (3) to determine whether it is a command or not.

\[
\text{fix}(x,y) = \text{cmd}(\text{AR}(x,y)) \quad \text{if} \quad \sum_{t=0}^{\text{dwell time}} \frac{\text{fix}(x,y)}{f} = \text{fix}(x,y)
\]

(3)

In this expression \(\text{cmd}(\text{AR}(x,y))\) is the command associated with the \(\text{AR}(x,y)\) and \(f\) is the frequency factor.

The frequency factor \((f)\) is determined based on the frequency of actions that the mobile robot can perform in one second. The frequency of actions of the mobile robot was set to 3Hz. Therefore, a dwell time of 1/3 of a second is used to consider a fixation as a command which makes \(t=330\) ms.

Research shows that independently controlling a camera mounted on a mobile robot helps in teleoperation [17]. Therefore, the operator was provided with action regions to control the onboard camera separately from the robot. This helps in avoiding to rotate the robot’s body around when a certain angle of view is required and can be achieved only by adjusting the view angle of the camera.

The eyes of the operator require rest from time to time. Furthermore, in some cases the operator might need to inspect the scene more deeply from time to time. Therefore the no-action regions on the interface provide the operator with rest for the eyes and with the opportunity to inspect parts of the scene. However, to provide the user with a greater opportunity to inspect the scene, a more radical solution is used in the TeleGaze interface.

The operator is provided with the option of using the interface either to interact with the system or to
inspect the scene only. The inspection mode of the TeleGaze allows the user to inspect the scene free from action regions. Switching between the interaction mode and the inspection mode is also performed using inputs from the eyes. In the latter mode, the only action region is one whose function is to switch back to the interaction mode and activate the rest of the action regions. A likely scenario where this functionality would help is driving the robot to a location and adjusting the direction of the camera using the interaction mode then switching to the inspection mode to read a poster using the whole scene without the fear of issuing any commands. Fig. 6. shows an actual snapshot of the inspection mode of the TeleGaze interface where only one action region can be seen.

V. OBSERVATIONAL STUDY

During the process of developing a prototype for the TeleGaze interface, it was realized that there are many alternative ways of designing the interface with more or less the same level of interaction. Making a final decision, as far as the design is concerned, turned out to be difficult both in terms of personal preference and system performance. Unlike the previous related attempts reported in [25, 33, 34] this research extended the design possibilities to actual users of the interface. Hence, to find out the most preferred design for the TeleGaze interface, prior to any task oriented evaluation, an observational study was conducted. Observational studies or what is referred to as the formative design by some authors [27] are recommended and implemented in the design of many interactive systems [35].

Two prototypes of the TeleGaze interface were initially used in this observational study with differences in the layout of the interface and in the provided level of control. However, despite the differences in the prototypes, the same design language was used in both prototypes. The same design language helps the users in learning less but doing more. Both prototypes are shown in Fig. 7. Fig. 7a shows the Edged-Interface (EI) and Fig. 7b shows the Centred-Interface (CI).
A. Design Constrains:

In the Edged-Interface (EI) the action regions are distributed alongside the edges of the interface, whereas, in the Centred-Interface (CI) the action regions are concentrated in the centre of the interface. The idea behind the design of EI was to keep the centre of the interface free from action regions so that the operator will be able to inspect the scene through the centre of the interface. However, the disadvantages of this design are thought to be the distance between the action regions and the non-intuitive positioning of some of the action regions such as looking upward to drive the robot forward. On the other hand in CI the operator can drive the robot forward and keep looking forward due to the intuitive positioning of the “forward” action region. However, the centre of CI is crowded and mostly occupied with action regions which was thought to distract the operator more than what would happen in EI.

Another point of difference between the two prototypes was that in EI the captions on the action regions were displayed continuously in a static way. In contrast, in CI the captions were not displayed and they were moving along with the point of the gaze. They were displayed in the form of tool tips rather than captions. Whichever region the user was gazing at, the caption was changing to the caption associated with that region. This was called dynamic caption functionality while the caption style used in EI was called static caption functionality. The dynamic caption functionality helped in freeing the action regions from writings and increased the amount of inspection that could be achieved through the action regions.

Regarding differences in the functionality of the interface in EI the operator did not have any control over the view from the downward camera which is located in the top-right corner of the interface. However, in CI the operator was able to enlarge and centre the view from the downward camera in order to have better view of the obstacles in the close surroundings of the robot body. An action region labeled “Swap Cam” provided the user with the capability of switching the main view between the forward camera and the downward camera. Fig. 8 shows a snapshot of the CI with enlarged and centered view from the downward camera. Comparing Fig. 7b and Fig.8 will give a clearer understanding of the concept to the reader.
Another point of difference in functionality is the automatic realignment of the camera once a moving command was issued to the robot. In EI the operator was presented with the capability of locking the camera at any angle and moving the robot, whereas, in CI the camera was realigned automatically whenever a moving command was issued to the robot. The operator had the option to set the realignment of the camera to either automatic or controlled through the use of an action region labeled “Lock Cam”. With these differences in the two prototypes, the observational study was carried to find out about user preferences for each difference.

B. Participants

A group of ten participants volunteered to participate in the TeleGaze observational study. The age of the participants ranged from very young students to some relatively old members of staff. The participants were both males and females. Among these participants were people with good familiarity with using computers and people with less familiarity. This group was chosen to cover a wide range of potential users.

At this stage and for the purpose of this observational study, the participants were not asked to perform any specific navigational tasks. Instead, they were left free to play around with the interface and try all the action regions that they might use for any navigational task. However, to perform a sensible comparison all participants were left to try both prototypes for an equal period of time.

All participants were given a brief verbal description of the idea of the study including a description of how the interface works. Then they were asked to investigate and try the interface freely using both prototypes. To avoid any bias towards any particular design, half of the participants were presented with EI first and then CI, whereas, this was turned around for the other half of the participants. However, a hard copy of both interfaces was delivered to the participant prior to commencing with the actual use of the prototypes to make them more familiar with the interfaces that they would be trying.

C. Results

The results of the observational study questionnaire are presented in Table I. The four major
differences in concern are listed with the number of the votes for each design. The results of the observational study are implemented in the refinement of the TeleGaze interface.

From the results shown in table I it can be seen that a higher number of votes went for EI as far as the general layout of the action regions is concerned. Most of the participants preferred the centre of the interface to be as free as possible to provide them with clearer view of the scene. However, they preferred the way that the action region “Forward” was located in CI as provided them with looking forward and moving forward.

Regarding displaying the captions on the action regions, a higher number of participants preferred static captions and not dynamic ones. They mentioned that the dynamic caption distracts them from the actual task of gazing and controlling. Therefore, they preferred the captions to be displayed continuously and statically on the interface. However, a number of the participants mentioned that with more practice, they will be able to discard all the captions because the number and the positioning of the action regions are easy to remember.

Similarly a higher number of votes went to EI regarding the presentation of the view from the downward camera without any further control over that view. Most of the participants believed that there is no need to enlarge and centre the view from the downward camera as long as it can be accessed whenever is required.

On the other hand, most of the votes went for CI when it came to the automatic realignment of the forward camera with the alignment of the robot. Despite the additional control that EI provided the operator with, most participants preferred automatic realignment. However, a few of the participants mentioned that they would like to have control over the alignment of the camera with an alignment indicator presented on the interface.

In general, it can be inferred from the results of this evaluation study that most participants preferred an easier, simpler more straightforward design rather than a hard, complex and twisted one despite the extra
level of control that the latter would provide.

D. **Refined Interface Design**

In the light of the results of the observational study covered earlier, a third prototype was developed for the TeleGaze interface that meets all the user preferences observed in the study. Fig. 9 shows a snapshot of the refined interface (RI). In RI the centre of the interface is nearly free of action regions as it was in EI. Also there is no control over the view of the downward camera, the cameras are automatically realigned and the captions are displayed statically.

Another point worth mentioning at this stage is that during the observational study, it was noticed that almost none of the participants moved the robot backward. In other words, none of the participants used the “Backward” action region. Therefore, the size of the “Backward” action region was reduced and moved out of the centre of the interface.

It can be argued that the design presented here is the best design for the purpose of the TeleGaze. The designing process is highly iterative, going forwards and backwards through the requirements of the system [27]. However, to obtain a reasonable balance between design and evaluation, it was decided to move forward to undertake a task oriented evaluation for the TeleGaze interface.

VI. **TASK ORIENTED EVALUATION**

To evaluate the refined design of the TeleGaze interface, a task oriented evaluation was carried out. The purpose of this task oriented evaluation was to compare the performance of the TeleGaze with other modalities of interaction. This section covers the task design and the evaluation metrics used in the evaluation together with a discussion of the results.

A. **Task Design**

Although a navigation task might differ from one application to another in many respects such as speed, required accuracy, complexity of the task and duration of the task, all these parameters are
application-dependent and liable to personal judgments. However, any task can be further divided into a number of subtasks. Furthermore, subtasks can be divided into a number of actions. In any navigational task, there are a number of subtasks that are most likely to be undertaken and therefore a number of actions. The most likely subtasks in any navigational task are moving along a straight line, turning to the right and to the left and finally stopping at a designated point. To include all these subtasks in the design of one navigational task, the track shown in Fig. 10. was planned.

To evaluate the TeleGaze interface in comparison with some other modes of interaction such as a conventional joystick, the participants were asked to drive the robot along the track using different modes of interaction including the TeleGaze interface. A brief explanation of each interaction mode was given to each participant with only one minute exercise prior to commencing the actual task. This was to get the participants familiar with each interaction mode and how the robot will respond to the commands. Then the time and accuracy of task completion were recorded for each participant and for each interaction mode.

One major character of teleoperation is that the control is from a remote location. This makes the task far more difficult to perform since the user is not interacting with the robot in real three dimensional spaces. Instead, the interaction will be in two dimensional spaces. In other words, the user is interacting only with feedback and different forms of data rather than looking at the actual robot itself in the same physical space. Therefore, with all modes of interaction, the participants were allowed to monitor the robot only through the interface shown in Fig. 9.

**B. Evaluation**

To the present day and due to the diversity of human-robot interaction applications there are no standard metrics to evaluate any newly developed interaction systems. However, there are a number of common metrics in any application domain that are most likely to be used to evaluate the developed system in that domain [36].

In many human-robot interaction applications addressing navigation and teleoperation, the most
commonly used metrics are **efficiency** and **effectiveness**. These two common metrics are used in the evaluation of the TeleGaze interface. Although, the definition of these metrics might vary from one application to another, for the purpose of this work efficiency was defined as *the time to complete the task* and effectiveness was defined as *the accuracy to keep on track*. Since the objective of this evaluation is to compare TeleGaze with other modes of interaction, the absolute time of task completion and accuracy are not a matter of concern as much as the relative ones are.

**B1. Joystick vs. TeleGaze Using Inputs from the Eyes**

Many commercially available robotic platforms, ranging from toy robots to robots for military applications, are controlled by some sort of joysticks [37]. Based on this information it can be assumed that a joystick is one of the most convenient ways of controlling a robot. Hence, the average time to complete the task and the accuracy of performing the task using a joystick is the target for the TeleGaze to meet. Based on this assumption, the operators were asked to perform the navigational task using the TeleGaze interface and a conventional joystick. The average time of task completion for both the TeleGaze interface and the joystick are illustrated in Fig. 11 with standard deviations of 45.16 seconds and 38.30 seconds, respectively.

From the results shown in Fig. 11 it can be concluded that the TeleGaze efficiency is 67% of the joystick efficiency based on the average time of task completion for ten participants. Based on a two tailed t-test, this difference in efficiency is considered statistically significant at \( p=0.05 \). This shows that the TeleGaze interface is significantly less efficient than the joystick. However, since this efficiency is calculated purely based on the average time of task completion, a very significant point should be considered.

The requirement of the task was to follow a painted track based on a tape laid down on the floor which the operator had to monitor from the view of the downward camera on the top-left corner (Fig. 9.). Using the joystick, the operator could look at that view and issue controlling commands for the whole
duration of the task. In contrast, with the TeleGaze commands could not be issued when the operator was looking at the view to monitor the position of the robot in relation to the tape on the floor. In other words, the operator did not issue controlling commands for the whole duration of the task which led the overall time of task completion to increase by 43% in comparison with the joystick.

This is less likely to happen in a different scenario such as moving from one point to another without checking the view from the downward camera. In this case, the operator would look through the main view and be able to issue commands as well. From the observations, it can be concluded that in such cases the efficiency of the TeleGaze is likely to increase by 20% and reach up to 80% of the efficiency of the joystick. However, the significant advantage of the TeleGaze should be mentioned here which is completing the navigational task with both hands absolutely free from the task.

To evaluate the effectiveness of the TeleGaze interface, the accuracy was defined as keeping the marked floor between the wheels of the robot. Despite moving out of an absolutely straight line, if no wheels of the robot crossed over the track then the accuracy considered as 100%. Overall, the task was repeated ten times for each interaction mode. Eight times out of these ten a 100% accuracy was achieved based on the criterion mentioned above with both the TeleGaze and the joystick. This means that the TeleGaze accuracy is 100% in comparison with the joystick accuracy. Whether the definition of the accuracy can be generalized for other navigational tasks or not is arguable but the accuracy of the TeleGaze is computed based on the accuracy of the joystick.

**B2. Joystick vs. TeleGaze using a computer mouse**

The TeleGaze interface is designed so that it can be used with a conventional mouse. Controlling the pointer on the screen can substitute for the inputs from the eyes and then, based on the same design language it can be used to teleoperate the mobile robot. An application of this functionality is using the TeleGaze interface with portable devices such as smart mobile phones or PDAs to navigate a mobile robot. With portable devices, due to the limitation in the space available for monitoring and due to the fact that
interacting with such devices is performed mostly using a pen, the TeleGaze interface might be an ideal solution for teleoperation.

From this point of view, the same task oriented evaluation was extended to include the TeleGaze interface using a conventional mouse. Surprisingly, in the case of using the mouse the efficiency is higher than that of the joystick as the average time to complete the task is less. The average time of task completion for both the TeleGaze interface using a mouse and the joystick are shown in Fig. 12 with standard deviations of 27.19 seconds and 38.30 seconds, respectively.

From Fig. 12 it can be observed that the efficiency of the TeleGaze interface using a mouse is 101.05% of the efficiency of the joystick. Based on a two tailed t-test with \( p=0.05 \) this increase in efficiency can not be considered statistically significant. However, using the mouse only required one hand of the operator to be engaged in the teleoperation while with the joystick both hands were engaged continuously.

As far as effectiveness is concerned, recalling the criteria designed earlier for the accuracy, the accuracy of the TeleGaze interface using a conventional mouse is 100% of the accuracy of a conventional joystick. Again it needs to be emphasized that this result is not for the absolute accuracy of the TeleGaze interface, rather it is the relative accuracy measured in comparison with the accuracy of the joystick.

**VII. CONCLUSIONS:**

In this paper, TeleGaze is introduced as a means of controlling mobile robots from a remote location. A TeleGaze interface is developed out of two different initial designs based on an observational study. Then task oriented evaluation was conducted to compare the efficiency and the effectiveness of the interface for two different applications in comparison to other modes of interaction.

From the work presented in this paper, it can be concluded that TeleGaze is likely to play a significant role in many teleoperation applications. An intuitive interpretation of the inputs from the eyes aids in the teleoperation and is likely to free the hands of the operator, either partially or fully, from the controlling
task. Although based on the evaluation presented earlier the TeleGaze efficiency has not beaten the efficiency of a conventional joystick but it has the significant advantage of allowing the operator to perform a navigational task with both hands completely free. Furthermore, continuous development in eye tracking technology will add significantly to the development of TeleGaze.

As part of ongoing research in further investigating TeleGaze, the next steps are likely to include diagnostic studies to find out the sources of limitations of TeleGaze. Although the authors believe that the limitations of the current eye tracking technology are the most likely sources of the limitations of TeleGaze, there are a number of other possibilities to be studied. The possibilities include but not limited to the design of the interface and the design language such as using alternatives for the dwell time to issue a command. Adding a layer of intelligence to the TeleGaze will be considered in the future studies as well.
REFERENCES


[12] H. Istance, "Communication through eye-gaze: Where we have been, where we are now and where we can go from here," in *Symposium on Eye Tracking Research and Applications*, 2006, pp. 9.


[22] Z. Bien, Myung-Jin Chung, Pyung-Hun Chang, Dong-Soo Kwon, Dae-Jin Kim, Jeong-Su Han, Jae-Hean Kim, Do-Hyung Kim, Hyung-Soon Park, Sang-Hoon Kang, K. Lee and Soo-Chul Lim,


LIST OF FIGURES

Fig. 1: The Conceptual Design of the TeleGaze System

Fig. 2: The Data Flow in the TeleGaze System

Fig. 3: The experimental platform used to experiment the TeleGaze interface; a) The teleoperation station, b) The TeleGaze interface, and c) The mobile robot platform.

Fig. 4: The Illustration of the layout of the TeleGaze interface.

Fig. 5: An actual snapshot of the TeleGaze interface.

Fig. 6: A snapshot of the inspection mode of the TeleGaze interface.

Fig. 7: Snapshots of two prototypes of the TeleGaze interface. a) The Edged-Interface (EI), and b) The Centred-Interface (CI)

Fig. 8: The TeleGaze interface with the enlarged view from the downward camera centered in the interface.

Fig. 9: A snapshot of the refined interface (RI) of the TeleGaze

Fig. 10: The illustration of the navigational task used in the evaluation.

Fig. 11. Figures of the evaluation results: average time to complete the task using a joystick and using TeleGaze with inputs from the eyes.

Fig. 12. Figures of the evaluation results: average time to complete the task using a joystick and using TeleGaze with a conventional computer mouse.

LIST OF TABLES

TABLE I: Results of the observational study questionnaire.
Fig. 1: The Conceptual Design of the TeleGaze System

Fig. 2: The Data Flow in the TeleGaze System
Fig. 3: The experimental platform used to experiment the TeleGaze interface; a) The teleoperation station, b) The TeleGaze interface, and c) The mobile robot platform.

Fig. 4: The Illustration of the layout of the TeleGaze interface.

Fig. 5: An actual snapshot of the TeleGaze interface.
Fig. 6: A snapshot of the inspection mode of the TeleGaze interface.

Fig. 7: Snapshots of two prototypes of the TeleGaze interface. a) The Edged-Interface (EI), and b) The Centred-Interface (CI)
Fig. 8: The TeleGaze interface with the enlarged view from the downward camera centered in the interface.

Fig. 9: A snapshot of the refined interface (RI) of the TeleGaze

Fig. 10: The illustration of the navigational task used in the evaluation.
Fig. 11. Figures of the evaluation results: average time to complete the task using a joystick and using TeleGaze with inputs from the eyes.

Fig. 12. Figures of the evaluation results: average time to complete the task using a joystick and using TeleGaze with a conventional computer mouse.
TABLE I: Results of the observational study questionnaire.

<table>
<thead>
<tr>
<th>Differences between the prototypes</th>
<th>No. of Votes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EI</td>
</tr>
<tr>
<td>General layout of the action regions (edged vs. centred)</td>
<td>6</td>
</tr>
<tr>
<td>Captions (static vs. dynamic)</td>
<td>7</td>
</tr>
<tr>
<td>Enlarging the view from the downward camera (no vs. yes)</td>
<td>7</td>
</tr>
<tr>
<td>Automatic camera realignment (yes vs. no)</td>
<td>1</td>
</tr>
</tbody>
</table>