Kevin Yu*, Thomas Wegele, Daniel Ostler, Dirk Wilhelm and Hubertus Feußner

EyeRobot: enabling telemedicine using a robot arm and a head-mounted display

https://doi.org/10.1515/cdbme-2020-0019

Abstract: Telemedicine has become a valuable asset in emergency responses for assisting paramedics in decision making and first contact treatment. Paramedics in unfamiliar environments or time-critical situations often encounter complications for which they require external advice. Modern ambulance vehicles are equipped with microphones, cameras, and vital sensors, which allow experts to remotely join the local team. However, the visual channels are rarely used since the statically installed cameras only allow broad views at the patient. They neither allow a close-up view nor a dynamic viewpoint controlled by the remote expert. In this paper, we present *EyeRobot*, a concept which enables dynamic viewpoints for telepresence using the intuitive control of the user's head motion. In particular, *EyeRobot* utilizes the 6 degrees of freedom pose estimation capabilities of modern head-mounted displays and applies them in real-time to the pose of a robot arm. A stereo-camera, installed on the end-effector of the robot arm, serves as the eyes of the remote expert at the local site. We put forward an implementation of *EyeRobot* and present the results of our pilot study which indicates its intuitive control.

Keywords: head-mounted displays; telemanipulation.

Introduction

Communication is a key aspect of successful medical treatment. Due to the fast modernization of connected devices, establishing a communication lines between two parties often only require mere seconds. During the process of digitization of the health care section, telemedicine quickly gains importance during patient treatment. Telemedicine describes the practice of utilizing digital long-distance communication for connecting patients and doctors anywhere in the world and is a rapidly growing technology and beneficial for inpatient, outpatient, and remote care [1]. It naturally increases the availability of expert knowledge at a given time, reduces travel expenses, and synergizes with digital patient data acquisition.

During a mass-casualty incident, telemedicine can be life-saving. Such incidents involve a large number of injured people, but equally important, involve a large number of regional paramedics. If the impact of the incident overstrains the capacity of paramedical first response, even novices in medical practices are eligible to be deployed. In order to reduce the workload of paramedics and to aid less experienced health workers, remote experts from geographically distant locations connect to them via local telecommunication channels. Modern ambulance vehicles are equipped with microphones and cameras which are accessible for remote experts. Contrary to the intuition that the videos captured from inside the vehicle provides most of the relevant information for consultation, it is the audio channel that is most and frequently used. The cameras in those vehicles are installed statically on the rooftop. Naturally, from these viewpoints, the cameras are only able to capture a broad view of the patient. The remote experts are neither able to change their viewpoint nor able to have a close-up view of the patient. Instead, they have to instruct the paramedics to manually move a mobile camera to their desired view [2]. Additionally to these problems, the view of the camera can be occluded by local workers. We solve the presented problem by explaining our concept of *EyeRobot* in section 2. Furthermore, we show in section 3 that spatial exploration with the *EyeRobot* can be done faster with the head-mounted display compared to using keyboard and mouse.

1.1 Related works

There are several concepts published to digitally immerse distant persons into a remote location.

The first approach to enable telepresence with dynamic viewpoints uses human surrogates. Kasahara et al.'s [3] purposes their *JackIn Head* system, in which one user wears a 360° camera while the second person spectates the

***Corresponding author: Kevin Yu**, University Hospital Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany, E-mail: kevin.yu@tum.de
Thomas Wegele, Daniel Ostler, Dirk Wilhelm and Hubertus Feußner, University Hospital Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

video stream using a head-mounted display. They reported the sharing of the first-person view is a promising method. The human surrogate created dynamic viewpoints; however, they did not correspond to the movement of the second user. This confusion between the visual input and vestibular cues (i.e. the sense of balance) induced symptoms of cybersickness for the spectator [4, p. 49].

The second approach to enable telepresence utilizes RGB-D cameras, which are cameras able to capture the color and depth information of the scene. Escolano et al. [5] set up a dedicated area for the *HoloPortation* that is fully captured by multiple RGB-D cameras. They combined the data of all RGB-D camera for generating a fully 3D reconstructed scene that is transmitted to the Mixed Reality head-mounted display HoloLens worn by the remote user. This allowed the remote user to freely inspect the 3D reconstructed scene from arbitrary angles. Fuchs et al. [6] utilize bipolar projectors to project the 3D reconstructed scene on a flat surface and effectively creates the illusion of a window into the remote room. The local user can perceives the projection in 3D by wearing shutter glasses calibrated and synchronized beforehand to the projector. These approaches require high network bandwidth and strong computational power.

The third approach utilizes robotic components, coupled with a camera and a display. Higuchi et al. [7] use a drone in synchronization of the user's head motion. The *Flying Head* successfully translated three-dimensional head-motion to control the drone's movement and yielded faster navigation compared to using joystick controls. The *AESOP* surgical system is a robotic camera holder for laparoscopes during minimally invasive surgery, which motivated the *ZEUS* Robotic Surgical System and *da Vinci* Surgical System. The ZEUS and da Vinci systems share the concept of a remote console controlling several robot-arms [8]. Kristoffersson et al. [9] published a thorough review on mobile robotic telepresence on industrial and consumer levels. Those systems listed in the review article were perceived positively; however, they required unoccupied space on floor level to change their viewpoint and were limited in their degrees of freedom.

Method

EyeRobot emulates the paradigm of telepresence by using a robot arm as a surrogate for the user. This is accomplished by synchronizing the movement of a head-mounted display with the movement of the robot arm. A stereo camera, mounted onto the end-effector of the robot arm, provides immediate visual feedback from the perspective of the robot-arm. Lastly, by displaying the stereoscopic video stream onto the stereo displays inside the HMD, the user sees the local scene through the perspective of the robot arm.

Due to this setup, we expect the benefits of *EyeRobot* to be manifold: (a) The utilization of network bandwidth can be optimized rather easily due to existing well-optimized real-time video streaming libraries. (b) By capturing stereoscopic images with a baseline similar to the average interpupillary distance of a human, the 3D perception is intuitive to understand. (c) The control of the viewing angle and position of the robot deems to be intuitive since the end-effector of the robot arm mimics the movement of the user's head. (d) The remote expert gains an exceptional 3D spatial understanding through the two visual cues motion parallax and stereopsis.

A possible hardware combination of *EyeRobot* uses the soft robot by Panda Franka Emika GmbH, the ZED mini stereoscopic camera by Stereolabs Inc, and a Virtual-Reality Headset with inside-out tracking such as a Windows Mixed-Reality Headset by Acer (Figure 1). Alternative HMDs include optical see-through HMDs such as the Microsoft HoloLens, or VR HMDs such as the Oculus Rift series with outside-in tracking. *EyeRobot* is realizable with any HMD capable of real-time 6 DoF spatial tracking which includes translational and rotational information.

2.1 Implementation

EyeRobot involves several transformations. The notation ${}_{a}T^{b}$ refers to the 4-by-4 homogeneous transformation matrix from the coordinate space of a to b. An initiation phase is responsible for synchronizing the pose of the robot and the HMD. The robot takes in its default pose with the mounted camera pointing forwards. By pressing a virtual button, the user confirms the initial pose of the HMD ${}_{w}T^{c_{t_0}}$, with w being the world coordinate space and c_{t_0} being the camera coordinate space at the initial frame 0. Subsequent poses for frame i are computed as a relative transformation to ${}_{w}T^{c_{t_0}}$ by multiplying the matrix inverse with the current pose in world coordinate:

$$ {}_{c_{t_0}}T^{c_{t_i}} = \left({}_{w}T^{c_{t_0}}\right)^{-1} \cdot T^{c_{t_i}} \tag{1} $$

On the same note, the robot takes in the default pose and stores the transformation to its end-effector as ${}_{b}T^{e_{t_0}}$. The synchronization of both, robot and HMD, is fulfilled by setting the robot coordinate space of e_{t_0} equals to the camera coordinate space c_{t_0}. Following from (1), the equation

$$ {}_{b}T^{e_{t_i}} = {}_{b}T^{e_{t_0}} \cdot {}_{c_{t_0}}T^{c_{t_i}} \tag{2} $$

computes the new pose of the end-effector at frame i. The calculation and control of the final robot inverse kinematics are taken over by the robot interface. In order to allow the collaboration of *EyeRobot* in a relatively tight space on the local site, the control is designed as an active impedance controller inside a Cartesian space. This decision is crucial to the system since it allows human intervention on the robot and contributes to the safety of *EyeRobot* towards nearby people at the local site. The additional feather-damping system consisting of the final control signal U, and current position of the end effector e_{is} and its velocity $\dot{e}_{\text{is}}$.

$$ U(t_{\text{i}}) = K \cdot (e_{t_i} - e_{\text{is}}) + D \cdot \dot{e}_{\text{is}} \tag{3} $$

completes the control scheme of the robot arm.

For ensuring the *safety* of the local workers, *EyeRobot* contains several abort conditions that cause the robot arm to immediately cease manipulation and take on hold once met. The abort conditions

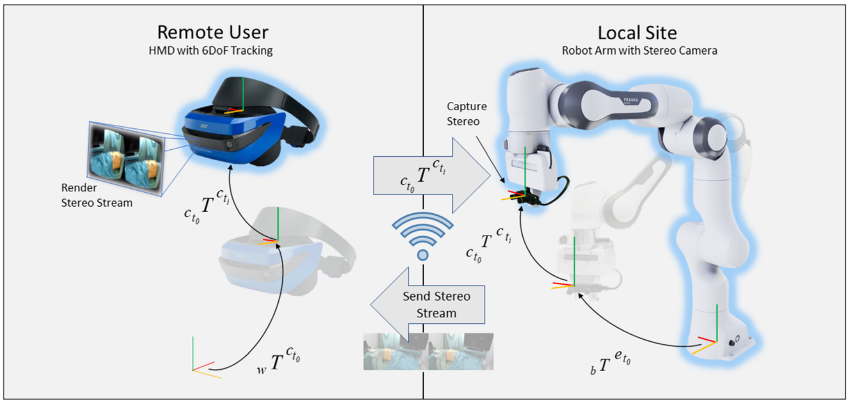

Figure 1: Depiction of the real-time update loop for building up the paradigm of telepresence in the *EyeRobot*. Both components, HMD and robot arm synchronize at an initiation step (the initial poses are depicted transparently). As a traditional master-slave system, the HMD controls the robot arm by sending the relative position from the initiated pose. In return, the robot arm sends the stereoscopic video, captured by a binocular camera mounted at the end-effector, back to the remote user as replacement of its sight.

are met if (a) the velocity of the joints or the end-effector surpasses a threshold value, (b) the desired end-position is outside the operating range of the robot, or (c) an external force is applied on the end-effector e.g. when colliding with an object or person.

The *network communication* between the remote user and the local site expands to the transmission of the relative transformation ${}_{c_{t_0}}T^{c_{t_i}}$ and the captured stereo video streams. We design the network architecture with the hindsight to minimize latency due to its major contribution to the phenomenon of cybersickness.

Due to the small encoded size of the transformation, the User Datagram Protocol is a suitable candidate. The transmission of the stereoscopic frame requires a more sophisticated approach due to the amount of the data. A suitable approach for transmitting the video stream utilizes GPU accelerated H.264 compression, followed by the transmission to the remote user via Real-Time Streaming Protocol.

Figure 2: Setup of the pilot study on comparing *EyeRobot* with Keyboard/Mouse navigation of the robot arm. The goal is to navigate the camera such that each letter is well visible.

3 Study on spatial navigation

In this section, we evaluate the *EyeRobot* in the aspects of intuitiveness and the precision of control. We prepared an environment containing five printed letters as seen in Figure 2. The task of the participants is to position the camera, and with it, the robot arm, to face every letter sequentially in a timely manner. First, the participants use the HMD to control the *EyeRobot* and the second time with a keyboard and mouse. WASD controlled the translation based on the camera's view direction, 'Q' and 'E' controlled up and down movement, and the mouse controlled pitch and yaw rotation. The time is measured for the entire procedure. We acquired

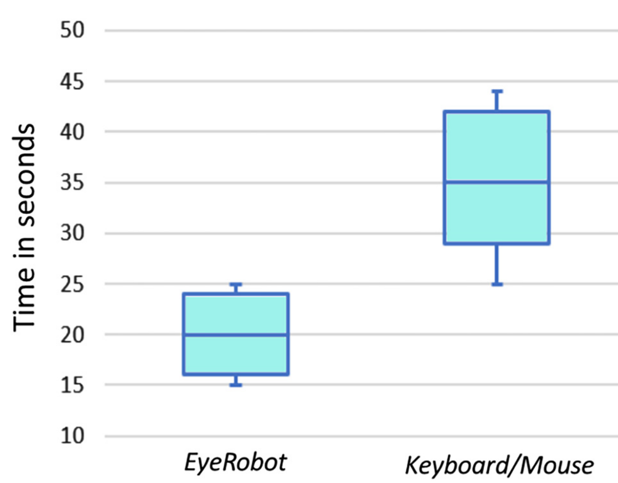

Figure 3: Box-plot time in seconds for exploring the test-scene via the robot arm: control with *EyeRobot* vs. Keyboard/Mouse control.

n=5 participants for our pilot study. None of them had experience with *EyeRobot* or a similar system.

3.1 Results and discussion

All of the participants were faster at finding the letters with *EyeRobot*. On average, they were about twice as fast, with 20 s compared to 35 s using keyboard and mouse as seen in Figure 3. One participant reported symptoms of cybersickness.

This result suggests users can quickly adapt to the control of *EyeRobot* since all participants were able to control the systems immediately without previous experience. Due to the physical properties of the robot arm, the working area is limited by its reach and self-collision. The area may be enlarged by mounting the robot arm on the ceiling and choosing a robot arm with longer links. A larger user study needs to be conducted in order to validate the added value of *EyeRobot* and the overall user experience for telemedicine.

4 Conclusion

We presented the concept of *EyeRobot*, a telepresence system with a manually controllable viewpoint and stereo vision in an AR/VR HMD. A remote expert using *EyeRobot* gains an in-depth spatial understanding using the two visual cues motion parallax and stereopsis. Furthermore, the direct translation of the movement of the HMD to the remote camera suggests an intuitive control with a low learning curve. We believe *EyeRobot* is a valuable asset to collaborative environments. For the future, we want to evaluate *EyeRobot* in a large collaborative setting with local paramedics and a remote expert in a realistic ambulant setting. In particular, we are curious about the feel of presence and acceptance of both local and remote users.

Research funding: This work has been developed in the project ArtekMed. ArtekMed (reference number: 16SV8088) is funded by the German Ministry of Education and Research (BMBF).

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: Authors state no conflict of interest.

Informed consent: Informed consent has been obtained from all individuals included in this study.

Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

1. Coombes CE, Gregory ME. The current and future use of telemedicine in infectious diseases practice. Curr Infect Dis Rep 2019;21:41.
2. Kurillo G, Yang AY, Shia V, Bair A, Bajcsy R. New emergency medicine paradigm via augmented telemedicine. In: International conference on virtual, augmented and mixed reality. Springer; 2016, pp. 502–11.
3. Kasahara S, Rekimoto J. JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In: Proceedings of the 21st ACM symposium on virtual reality software and technology; 2015, pp. 217–25.
4. Joseph J Jr, Kruijff E, McMahan RP, Bowman D, Poupyrev IP. 3D user interfaces: theory and practice. LaViola: Addison-Wesley Professional; 2017.
5. Orts-Escolano S, Rhemann C, Fanello S, Chang W, Kowdle A, Degtyarev Y, et al. Holoportation: virtual 3d teleportation in real-time. In: Proceedings of the 29th annual symposium on user interface software and technology; 2016, pp. 741–54.
6. Fuchs H, State A, Bazin J-C. Immersive 3D telepresence. Computer 2014;47:46–52.
7. Higuchi K, Rekimoto J. Flying head: a head motion synchronization mechanism for unmanned aerial vehicle control. In: CHI'13 extended abstracts on human factors in computing systems; 2013, pp. 2029–38.
8. Pugin F, Bucher P, Morel P. History of robotic surgery: from AESOP® and ZEUS® to da Vinci®. J Visceral Surg 2011;5:e3–8.
9. Kristoffersson A, Coradeschi S, Amy L. A review of mobile robotic telepresence. Adv Hum Comput Interact 2013;2013. https://doi.org/10.1155/2013/902316.