# Guided 360-Degree Visual Perception for Mobile Telepresence Robots

Kishan Chandan, Xiaohan Zhang, Jack Albertson, Xiaoyang Zhang, Yao Liu, Shiqi Zhang
SUNY Binghamton, Binghamton, NY 13902 USA
{kchanda2;xzhan244;jalbert5;xzhan211;yaoliu;zhangs}@binghamton.edu

***Abstract*—Omnidirectional (360-degree) cameras have provided robots a field of view that covers approximately the entire sphere, whereas people's visual field spans only about $120^\circ$ horizontally. The difference in visual field coverage places a challenge for people to directly process the rich visual information (in spherical or planar forms) from the robots' 360-degree vision system. In this paper, we focus on mobile telepresence robot platforms equipped with 360-degree vision, where the human operator perceives and interacts with a remote environment through a mobile robot. We develop a framework that, for the first time, leverages the mobile robot's 360-degree scene analysis capability to guide human attention to the remote areas with potentially rich visual information. We have evaluated our framework using a remote target search task, where a human operator needs to use a semiautonomous Segway-based mobile robot to locate a target object in a remote environment. We have compared our approach to a baseline from the literature that supports 360-degree visual perception, but does not provide guidance to human attention. From the results, we observed significant improvements in the efficiency of the human-robot system conducting remote target search tasks.**

(a) (b)

Fig. 1: Segway-based mobile robot platform (RMP-110) with a 360-degree camera (Ricoh Theta V), mounted on top of the robot, and laser ranger finder (SICK TiM571 2D) that has been used for prototyping and evaluation purposes in this research.

## I. Introduction

Telepresence is an illusion of spatial presence, at a place other than the true location [17]. In particular, mobile telepresence robots help the human operator to extend their perception capabilities along with the ability of moving and actuating in a remote environment [14]. The rich literature of mobile telepresence robotics has demonstrated applications in domains such as offices [8, 27], academic conferences [18, 22], elderly care [7, 26], and education [3, 23].

Recently, researchers have equipped mobile telepresence robots with a 360-degree camera to perceive the entire sphere of the remote environment [31, 10]. In comparison, traditional cameras, including the pan-tilt ones, can only capture a limited visual field of the remote environment. On the one hand, such 360-degree cameras have equipped the robots with the omnidirectional scene analysis capability using computer vision algorithms. On the other hand, the human vision system by nature is not developed for, and hence not good at analyzing 360-degree visual information. For instance, computer vision algorithms can be readily applied to visual inputs of different field coverages, whereas the binocular visual field of the human eye spans only about $120^\circ$ of arc [25]. The field that is effective to complex visual processing is even more limited, e.g., only $7^\circ$ of visual angle for facial information [20].

One can project 360-degree, spherical views onto equirectangular video frames (e.g., panorama frames). However, people are not used to viewing such 360-degree equirectangular videos directly [9]. For instance, Google Street View provides only a relatively small field of view to the users, even though the 360-degree frames are readily available. Within the telepresence robotics context, this research is motivated by the gap between the off-the-shelf technology for 360-video capturing and the difficulty of the human to process the all-degree view of the remote environment directly.

In this paper, we propose a framework that, for the first time, equips mobile telepresence robots with simultaneous capabilities of 360-degree scene analysis and guiding human attention to areas of interest. The all-degree scene analysis capability of our framework, at the same time, passively keeps track of the remote environment and actively guides human attention to information-rich areas. In existing systems, the human operator has to *"look around"* to find such areas. Using our framework, once the information-rich areas are identified in the remote environment, the human operator's attention is guided towards such areas through indicators such as arrows, bounding boxes, and pop-ups. Such 360-degree scene analysis coupled with the visual guidance helps the human operator perceive an all-degree view of the remote environment.

We have evaluated our framework with experiments in a target search scenario. From the results, we observed that our framework significantly improves the efficiency of the human-robot system in finding the target object, in comparison with a competitive baseline that is equipped with 360-degree vision but does not provide guidance to human perception.

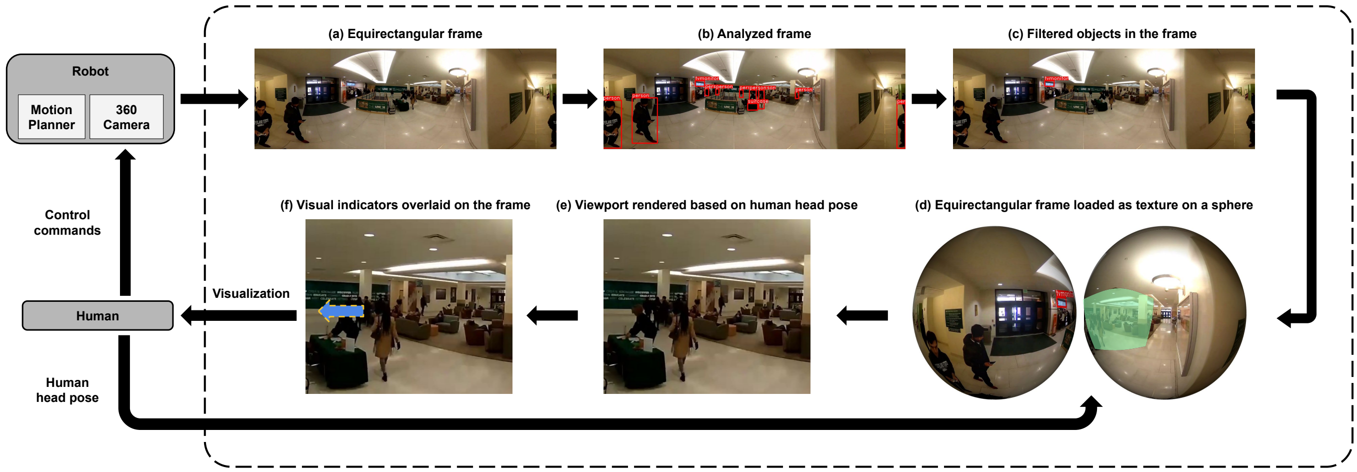


Fig. 2: Key components of our developed framework for guided 360-degree visual perception for mobile telepresence robots. The framework aims to enable simultaneous capabilities of visually perceiving the remote environment and guiding human attention to information-rich areas, toward improving shared autonomy in mobile telepresence robotics.

A recent article on how robotics can help in combating COVID-19 has pointed out that robotic avatars and controls can be used by attendees of exhibitions and conferences, which will result in reduction of both the infection rate of COVID-19, and also the carbon footprint [30]. During COVID-19 times, our framework can help the human operator to be virtually present at a remote location using a mobile telepresence robot with an enhanced perception of the remote environment.

## II. Related Work

There is rich literature of research on mobile telepresence robotics. Early systems include a personal telepresence robot, called PEBBLES [28], and a telepresence robot for enabling remote mentoring, called Telementoring [2]. One limitation of those systems is that their fixed view cameras result in a narrow field of view. As a consequence, the human operator must control the robot to manually change the field of view, losing the capability of all-degree scene analysis. In contrast, our framework uses a 360-degree vision system for all-degree perception and scene analysis of the remote environment.

To overcome the barrier of limited field of view in robotics, researchers have used multiple cameras embodied on the robot [12, 13], wide-angle cameras [29, 16], and pan-tilt cameras [21, 5, 15]. Although the pan-tilt cameras allow the human operator to move the camera to look around in the remote environment, there is the issue that the operator has the additional cognitive load to control the pan-tilt camera in addition to controlling the robot platform. To alleviate this issue, researchers have combined head motion to automatically control the motion of pan-tilt cameras using head-mounted displays (HMDs) [6, 4]. While such systems eliminated the need for manual control of a pan-tilt camera, they suffer from the limited visual field, and can only perceive the remote environment partially. Our framework overcomes this limitation through a 360-degree camera that enables the robot to perceive the entire sphere of the environment, while at the same time freeing the human operator from manually controlling the camera's angle.

Very recently, researchers have developed 360-degree camera-based systems for mobile telepresence robots [11, 10]. For instance, Zhang et al. developed a prototype system for redirected walking with a telepresence mobile robot mounted with a 360-degree camera [31]; and Oh et al. used a 360-degree camera for mobile telepresence robots in virtual tour scenarios [19]. In comparison to their work, our framework equips the robot with the scene analysis capability in 360 degrees, and further enables the robot to use the outputs of scene analysis to guide human attention to improve the human operator's situational awareness.

## III. Guided Telepresence Robotics Framework

In this section, we present our guided telepresence robotics framework that leverages the mobile robot's 360-degree scene analysis capability to guide human attention to the remote areas with potentially rich visual information.

Fig. 2 shows an overview of our framework. The input to the framework includes live 360-degree video frames of the remote environment obtained from the 360-degree camera mounted on the robot (Fig. 2 (a)). These frames encode the 360-degree information using the equirectangular projection, and we refer to them as equirectangular frames. These frames are processed using an object detection system to output an **analyzed frame** (Fig. 2 (b)). The detected objects are then filtered to highlight only the objects of interest (Fig. 2 (c)). Depending on the robot's current tasks, the objects of interest can be different.

Equirectangular frames are encoded and transmitted to the human operator's device. Once received, we can construct a sphere and use these equirectangular frames as texture of the sphere as shown in Fig. 2 (d). A user may only be viewing one portion of the sphere at a time, with limited field of view, e.g., as indicated in the green-shaded portion in Fig. 2 (d).

The portion viewed by the user is called the **viewport**. At runtime, based on the head pose of the human operator, views in the viewport are rendered accordingly as shown in Fig. 2 (e). Finally, the framework overlays visual indicators in real-time to guide the human attention towards the object of interest while providing an immersive 360-degree view of the remote environment (Fig. 2 (f)).

Next, we describe the framework in Algorithm 1, and then we describe a full instantiation of the framework.

### A. Algorithm

Algorithm 1 delineates the procedure of our guided telepresence robotics framework and explains the different stages as well as the data flow at each stage.

The input of Algorithm 1 includes the following:

- $\Theta$: horizontal field of view of human eyes (typically 120-degree). $\Theta$ can be tuned according to the device being used for visualization
- $\kappa$: name of the object of interest as a string
- $ObjDet$: real-time object detection system

Firstly, we enter a while-loop in Line 6, which is the main control loop of our algorithm. Then in Line 7, we check if new frames are available from the robot. Next, in Line 8, the current equirectangular frame ($F$) is obtained from the robot. Then, $ObjDet$ processes each frame and returns a set of detected objects with their corresponding locations in the frame as $P$ (Line 9). $P$ is stored in the form of a dictionary with the name of the detected objects ($Obj_{name}$) as keys and their locations ($OLoc$) as values. In Line 10, $\kappa$ is used to query the location of the object of interest, and then the bounding boxes are drawn around it (Line 11). Next, in Line 12, $F$ is used as texture information of a sphere. Based on human operator's head pose obtained in Line 13, the viewport is rendered from the sphere.

In Line 15, we calculate the relative location of the object of interest with respect to the viewport. Depending on the relative location of the object, $F^{'}$ is overlayed with a "left indicator", or a "right indicator", to guide the human attention in the direction of the object, whereas, if the object is in the viewport, then the frame is not modified (Lines 16-21). $F^{'}$ is then presented to the user via the interface, and the feedback is collected in the form of control commands (Line 23). Then, we enter a for-loop in Line 24 that converts every control command to a *Teleop message* (Line 25). Based on the control command the robot is remotely actuated and the 2D visualizable map is updated accordingly to show the new robot pose.

### B. Framework Instantiation

We design a web-based **telepresence interface** using the guided telepresence robotics framework, as shown in Fig. 3. The telepresence interface enables the human operator to perceive the remote environment, track the robot position in the remote environment, and also control the robot remotely.

For object detection, we use state-of-the-art convolutional neural network (CNN)-based framework **YOLOv2** [24] in our implementation. We use **A-Frame**, an open-source web framework for building virtual reality experiences [1] and for rendering the equirectangular frames to visualize an immersive 360-degree view of the remote environment. User can use "click" and "drag" operations to look around in the remote environment. The telepresence interface provides a scroll bar to adjust the field of view of the human operator. Our telepresence interface also allows the users to easily change the web-based visualization to an immersive virtual reality 360-degree experience via head-mounted displays (HMDs) via a **VR** button at the bottom right of the interface. In the VR mode, human users can use their head motions to adjust the

**Algorithm 1** Guided Telepresence Robotics Framework

1: **Require**: $\Theta$, $\kappa$, and a real-time object detection system ($ObjDet$)
2: Initialize an empty list $C$ to store the control commands as $\emptyset$
3: Initialize a quaternion $H$ (Human Pose) as (0,0,0,0)
4: Initialize a 2D image $F$ with all pixel values set as 0
5: Initialize a dictionary $P$ = {}
6: **while** True **do**
7:     **if** new frame is available **then**
8:         Obtain current frame $F$ of the remote environment
9:         $P \leftarrow ObjDet(F)$ ▷ Detect objects in the frame
10:         $OLoc = P[\kappa]$ ▷ Get the location of the object of interest
11:         Overlay bounding box for object of interest over $F$
12:         Construct a sphere using equirectangular frame as its texture
13:         Obtain current human head pose $H$
14:         Render a *viewport* $F^{'}$ to match the human operator's head pose
15:         Calculate the relative location of the object based on viewport
16:         **if** Object is in the viewport **then**
17:             No modifications are made to $F^{'}$
18:         **else if** Object is closer to the left of the viewport **then**
19:             Overlay "left indicator" over $F^{'}$
20:         **else if** Object is closer to the right of the viewport **then**
21:             Overlay "right indicator" over $F^{'}$
22:         Present $F^{'}$ to the user via the interface
23:         Obtain feedback from the human operator and store it in $C$
24:         **for each** $c \in C$ **do**
25:             $\psi \leftarrow \tau(c)$ ▷ Generate a teleop message
26:             $\psi$ is sent to the robot to actuate it
27:             Updated map with the robot pose is presented via the interface

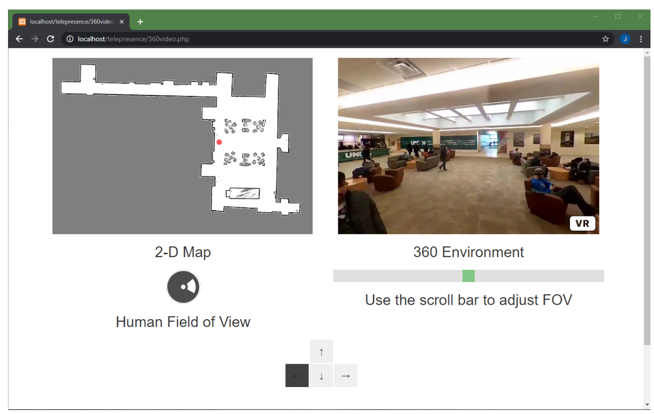


Fig. 3: *Telepresence Interface* showing the 2D map, the frame of the remote environment, a 2D icon to show human head orientation, a scrollbar to adjust field of view, and a set of icons to indicate the current teleoperation commands given by the human operator.

view of the remote environment.

The telepresence interface also shows the 2D map of the remote environment. Based on the robot pose, a red circle is overlaid on the 2D map to indicate the live robot location in the remote environment. Additionally, we also show a 2D icon to convey the head orientation of the human operator. The 2D icon also shows the part of the 360-degree frame which is in the field of view of the human operator (Fig 3). We use the arrow keys on the keyboard as an input mode to enable the human operator to give feedback to the robot. The control commands are converted to a ROS "teleop" message and passed on to the robot to remotely actuate the robot. The telepresence interface highlights the keys that are pressed by the human operator.

## IV. Experiments

We conducted experiments evaluating the guided telepresence robotics framework in a target-search scenario. In this scenario, a human operator was assigned a task of finding a target object in the remote environment using a mobile telepresence robot. For evaluation purposes, we designed a **hybrid evaluation platform** to simulate the robot and human behaviors. Next, we present details of the hybrid evaluation platform.

### A. Hybrid Evaluation Platform

We used 360-degree videos captured from the real world and simulated human and robot behaviors to build our hybrid evaluation platform. To capture 360-degree videos, we used a Segway-based platform (RMP-110) as the mobile robot platform with an on-board 360-degree camera (Ricoh Theta V). The robot was teleoperated in a public space, while the robot's live locations were estimated offline using a SLAM algorithm.

We model a virtual human in the hybrid simulator that replicates the behavior of the human operator. The virtual human can both track the location of the simulated robot through the 2D map and perceive the 360-degree view of the remote environment via the Telepresence interface. The virtual human can also look around in the remote environment and send control signals to teleoperate the simulated robot using the interface. In our hybrid evaluation platform, the control signal can be either *forward*, or *backward*, because we use a simulated robot with a pre-recorded video of the remote environment. Based on the control signal, the location of the robot is changed, and the interface is updated accordingly.

### B. Baseline vs. Guided Telepresence Robotics Framework

We compared our framework using a target-search scenario with a baseline from the literature that supports 360-degree visual perception, but does not guide the human attention [11, 10]. The location of the target object (in our case, a **television**) was constant, whereas, there were nine different start positions for the robot. The orientation of the robot was randomly selected, and a set of one thousand paired trials were carried out for every two-meter distance from the target object until 18 meters. The virtual human carried out one random action every two seconds out of "*move forward*", "*move backward*", "*look left*", and "*look right*". Using random actions, the simulated user could explore the remote environment. Once the target object is in the field of view of the human operator, the trial is terminated. Compared to baseline trials where only 360-degree view of the remote environment was visualized, in trials of our proposed framework, we performed all-degree scene analysis and overlaid visual indicator over the frames to guide human attention.

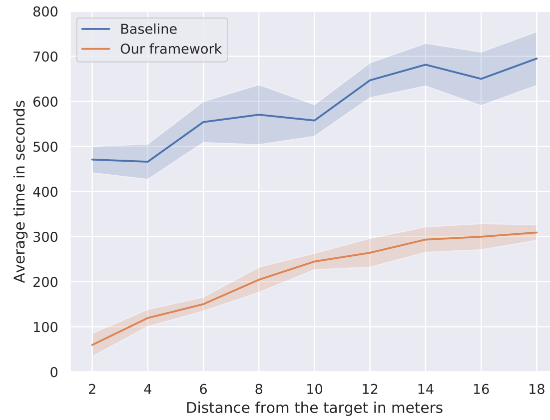

Fig. 4: Comparisons are made between our framework and the baseline in terms of average time (y-axis) taken to find the target object from nine different distances (x-axis).

### C. Results

Fig. 4 shows the overall performance of our framework compared to the baseline. Each data point represents the average task completion time (y-axis) for nine different distances (x-axis) (i.e., randomly sampled positions in each of the nine distances). One thousand trials were divided over four batches to plot each data point. The shaded regions denote the batched standard deviations for each of the nine distances.

It can be observed that the task completion time of the human operator in finding the target object was significantly reduced in our framework compared to the baseline with a $p$-value $< 0.001$ for all the distances. The results suggest that both the 360-degree scene analysis and the human guidance contribute to the improvement of the human operator's efficiency in locating target objects in the remote environment.

## V. Conclusions and Discussions

In this paper, we develop a guided telepresence robotics framework that enables a mobile telepresence robot to passively monitor the remote environment with all-degree scene analysis and actively guide the human attention to the key areas at the same time. We conducted experiments to evaluate our framework in a target search scenario. From the results, we observed significant improvements in the efficiency of the human-robot system in comparison to a baseline from the literature that does not provide guidance to human attention.

## Acknowledgments

Part of this work has taken place in the Autonomous Intelligent Robotics (AIR) Group at SUNY Binghamton. AIR research is supported in part by grants from NSF (NRI-1925044), Ford (URP Award), OPPO (Faculty Research Award), and SUNY Research Foundation. Jack Albertson is supported by NSF under Research Experiences for Undergraduates (REU) grant NRI-2032626. The work of Xiaoyang Zhang and Yao Liu is partially supported by NSF under grant CNS-1618931.

## References

[1] A-frame-a web framework for building virtual reality experiences. https://aframe.io/.

[2] Rahul Agarwal, Adam W Levinson, Mohamad Allaf, Danil V Makarov, Alex Nason, and Li-Ming Su. The roboconsultant: telementoring and remote presence in the operating room during minimally invasive urologic surgeries using a novel mobile robotic interface. *Urology*, 70(5):970–974, 2007.

[3] Veronica Ahumada-Newhart and Judith S Olson. Going to school on a robot: Robot and user interface design features that matter. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(4):25, 2019.

[4] Tamay Aykut, Christoph Burgmair, Mojtaba Karimi, Jingyi Xu, and Eckehard Steinbach. Delay compensation for actuated stereoscopic 360 degree telepresence systems with probabilistic head motion prediction. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2010–2018. IEEE, 2018.

[5] Michael Baker, Robert Casey, Brenden Keyes, and Holly A Yanco. Improved interfaces for human-robot interaction in urban search and rescue. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 3, pages 2960–2965. IEEE, 2004.

[6] Mark T Bolas and Scott S Fisher. Head-coupled remote stereoscopic camera system for telepresence applications. In *Stereoscopic Displays and Applications*, volume 1256, pages 113–123. International Society for Optics and Photonics, 1990.

[7] Amedeo Cesta, Silvia Coradeschi, Gabriella Cortellessa, Javier Gonzalez, Lorenza Tiberio, and Stephen Von Rump. Enabling social interaction through embodiment in excite. In *ForItAAL: Second Italian Forum on Ambient Assisted Living*, pages 1–7, 2010.

[8] Munjal Desai, Katherine M Tsui, Holly A Yanco, and Chris Uhlik. Essential features of telepresence robots. In *2011 IEEE Conference on Technologies for Practical Robot Applications*, pages 15–20. IEEE, 2011.

[9] Vamsidhar Reddy Gaddam, Michael Riegler, Ragnhild Eg, Carsten Griwodz, and Pål Halvorsen. Tiling in interactive panoramic video: Approaches and evaluation. *IEEE Transactions on Multimedia*, 18(9):1819–1831, 2016.

[10] John Paulin Hansen, Alexandre Alapetite, Martin Thomsen, Zhongyu Wang, Katsumi Minakata, and Guangtao Zhang. Head and gaze control of a telepresence robot with an hmd. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, page 82. ACM, 2018.

[11] Yasamin Heshmat, Brennan Jones, Xiaoxuan Xiong, Carman Neustaedter, Anthony Tang, Bernhard E Riecke, and Lillian Yang. Geocaching with a beam: Shared outdoor activities through a telepresence robot with 360 degree viewing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018.

[12] Sei Ikeda, Tomokazu Sato, and Naokazu Yokoya. High-resolution panoramic movie generation from video streams acquired by an omnidirectional multi-camera system. In *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI2003.*, pages 155–160. IEEE, 2003.

[13] Mojtaba Karimi, Tamay Aykut, and Eckehard Steinbach. Mavi: A research platform for telepresence and teleoperation. *arXiv preprint arXiv:1805.09447*, 2018.

[14] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction*, 2013:3, 2013.

[15] Hideaki Kuzuoka, Shinya Oyama, Keiichi Yamazaki, Kenji Suzuki, and Mamoru Mitsuishi. Gestureman: a mobile robot that embodies a remote instructor's actions. In *Proceedings of the ACM conference on Computer supported cooperative work*, pages 155–162, 2000.

[16] Gordon M Mair. Telepresence-the technology and its economic and social implications. In *1997 International Symposium on Technology and Society Technology and Society at a Time of Sweeping Change. Proceedings*, pages 118–124. IEEE, 1997.

[17] M. Minsky. Telepresence. *OMNI Magazine*, 1980.

[18] Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. To beam or not to beam: A study of remote telepresence attendance at an academic conference. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*, pages 418–431. ACM, 2016.

[19] Yeonju Oh, Ramviyas Parasuraman, Tim McGraw, and Byung-Cheol Min. 360 vr based robot teleoperation interface for virtual tour. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[20] Michael Papinutto, Junpeng Lao, Meike Ramon, Roberto Caldara, and Sébastien Miellet. The facespan—the perceptual span for face recognition. *Journal of vision*, 17 (5):16–16, 2017.

[21] Eric Paulos and John Canny. Social tele-embodiment: Understanding presence. *Autonomous Robots*, 11(1):87–95, 2001.

[22] Irene Rae and Carman Neustaedter. Robotic telepresence at scale. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 313–324. ACM, 2017.

[23] Irene Rae, Gina Venolia, John C Tang, and David Molnar. A framework for understanding and designing telepres-

ence. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 1552–1566. ACM, 2015.

[24] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger. *arXiv preprint arXiv:1612.08242*, 2016.

[25] Theodore C Ruch and John F Fulton. Medical physiology and biophysics. *Academic Medicine*, 35(11):1067, 1960.

[26] Alessandra Maria Sabelli, Takayuki Kanda, and Norihiro Hagita. A conversational robot in an elderly care center: an ethnographic study. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 37–44. IEEE, 2011.

[27] Leila Takayama and Janet Go. Mixing metaphors in mobile remote presence. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 495–504. ACM, 2012.

[28] Patrice L Tamar Weiss, Carolynn P Whiteley, Jutta Treviranus, and Deborah I Fels. Pebbles: A personal technology for meeting educational, social and emotional needs of hospitalised children. *Personal and Ubiquitous Computing*, 5(3):157–168, 2001.

[29] Kimiya Yamaashi, Jeremy R. Cooperstock, Tracy Narine, and William Buxton. Beating the limitations of camera-monitor mediated telepresence with extra eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 50–57, 1996.

[30] Guang-Zhong Yang, Bradley J Nelson, Robin R Murphy, Howie Choset, Henrik Christensen, Steven H Collins, Paolo Dario, Ken Goldberg, Koji Ikuta, Neil Jacobstein, et al. Combating covid-19—the role of robotics in managing public health and infectious diseases, 2020.

[31] Jingxin Zhang, Eike Langbehn, Dennis Krupke, Nicholas Katzakis, and Frank Steinicke. A 360 video-based robot platform for telepresent redirected walking. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interactions (VAM-HRI)*, pages 58–62, 2018.