

Participatory telerobotics

Alexander D. Wissner-Gross^{*a,b,c}, Timothy M. Sullivan^c

^aInstitute for Applied Computational Science, Harvard University, 29 Oxford St., Cambridge, MA, USA 02138-2933; ^bThe Media Laboratory, Massachusetts Institute of Technology, 20 Ames St., Cambridge, MA, USA 02142-1308; ^cGemedi, Inc., 485 Massachusetts Ave., Cambridge, MA, USA 02139-4018

ABSTRACT

We present a novel “participatory telerobotics” system that generalizes the existing concept of participatory sensing to include real-time teleoperation and telepresence by treating humans with mobile devices as ad-hoc telerobots. In our approach, operators or analysts first choose a desired location for remote surveillance or activity from a live geographic map and are then automatically connected via a coordination server to the nearest available trusted human. That human’s device is then activated and begins recording and streaming back to the operator a live audiovisual feed for telepresence, while allowing the operator in turn to request complex teleoperative motions or actions from the human. Supported action requests currently include walking, running, leaning, and turning, all with controllable magnitudes and directions. Compliance with requests is automatically measured and scored in real time by fusing information received from the device’s onboard sensors, including its accelerometers, gyroscope, magnetometer, GPS receiver, and cameras. Streams of action requests are visually presented by each device to its human in the form of an augmented reality game that rewards prompt physical compliance while remaining tolerant of network latency. Because of its ability to interactively elicit physical knowledge and operations through ad-hoc collaboration, we anticipate that our participatory telerobotics system will have immediate applications in the intelligence, retail, healthcare, security, and travel industries.

Keywords: Telepresence, telerobotics, crowdsourcing, participatory sensing, mobile devices, street view

1. INTRODUCTION

As of 2011, there are nearly 6 billion mobile-cellular subscriptions in the world, with 87% global penetration and 79% penetration in the developing world¹. With nearly 1.2 billion mobile-broadband subscriptions, we have reached a stage of sufficient high-bandwidth interconnection that it is now possible to do more than instantaneously speak or correspond with more people than ever before¹. We may now send and receive images and live video streams in real time. Moreover, as we will argue here, our mobile devices and broadband connections can also now enable our civilization to have the sort of immersive long-distance visual experiences currently permitted only through the use of telepresence robots, which have become accessible in the past few years only at a high fixed cost inaccessible to most individual consumers. At present, however, mobile devices do not permit us to access such experiences at will, or to command and instruct end users to navigate their spaces as desired or as one might expect using a telepresence robot that will move exactly as instructed by the end user.

We present here a novel “participatory telerobotics” system that represents a solution to the above-cited problem and that is the culmination of progress at the intersection of telerobotics and mobile devices, as well as crowdsourcing. In particular, it promises to reduce the relatively high fixed cost of telerobotics to near zero, within the reach of the billions of mobile phone users worldwide. Effectively, we leverage developments in cloud technology, mobile video cameras, and platforms to crowdsource the functionality of a telepresence robot without the expense or learning curve of an expensive telerobot. To understand the trajectory of this achievement, and why it is only now possible, we must examine the technological developments each of these fields over the past 10-15 years.

* alexwg@post.harvard.edu; <http://www.alexwg.org>

2. BACKGROUND

Street view services such as Google Street View represented a marked advance in the global use of mobile devices for remotely viewing geography. Launched in May 2007, Google's solution captures images of, among others, city streets, museum and retail interiors, using specially designed mobile cameras mounted to automobiles, backpacks, trikes, etc. Having grown to encompass significant portions of the globe, the Street View model has allowed a small number of people to capture still images that can later be interacted with by users of Google's Maps and Earth applications. One can "walk down" a street or "through a store" on a predefined path, and zoom in or out of the still image. While some images may now be viewed in three dimensions by those using red/cyan glasses, the experience has significant limitations. Above all, it does not permit real-time viewing or full interactivity with the environment. One can not pick up or examine objects in the store view, if those objects even still exist within the store by the time one views the image. Finally, camera resolutions are often quite limited, rendering it impossible to get high quality detail views even with the zoom functionality.

To interact with the environment, the field of telepresence robotics has enjoyed increasing popularity in recent years. Telepresence robots enable individuals at a distance to manipulate a robot while viewing and even verbally interacting with a distant environment. They may also provide a video screen enabling those distant to view the operator of the robot. Such immersive robotic experiences are opening up new ways for remote workers to interact at distant locations. For example, InTouch Health robots have provided surgeons with advice² from a distance of over 8,000 kilometers. In a personal robotics market of US \$1.16 billion, such robots may become important in industrial applications like security, surveillance, and eldercare². However, with a representative purchase price of US \$15,000 (e.g., for a QB telepresence robot from Anybots, Inc.), such devices will continue to be well out of the reach of the average individual for the years to come³.

While roboticists have been laboring over the creation of improved telepresence robots to extend our reach, another technology trend has arisen: crowdsourcing. While the practice of crowdsourcing has historical precedent, the term itself was only first introduced by Jeff Howe and Mark in June 2006 in Wired Magazine, as new platform technologies like the Internet and mobile devices enabled individuals to seek connections with vast numbers of people hitherto difficult to access at once⁴. According to their definition, "crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers."⁴

For most of the past decade, crowdsourcing has been widely employed for both skilled and unskilled labor by individuals and organizations. Amazon's Mechanical Turk, launched in 2005, has provided micro-job employers with labor to conduct social science experiments, artistic and educational research, and even missing-person searches. The website iStockphoto.com connects individuals with thousands of photographers. A variant on crowdsourcing known as crowdfunding permits individuals to broadcast requests for financial support and investment from large groups. It has seen rapid success in the website Kickstarter.com.

Related to crowdsourcing and crowdfunding is the notion of "participatory sensing," which involves groups of connected or disconnected people utilizing sensing devices – such as those built into mobile phones – to amass data to obtain new knowledge. While such sensing may now seem obvious and ubiquitous – used, for example, to detect traffic jams in smartphone mapping programs – this was not the case until quite recently. It was not until 2006 that J. Burke, et al. introduced the concept, noting the ubiquity of global mobile phone usage and its potential to "gather, analyze and share local knowledge."⁵

3. IMPLEMENTATION

Our proposed general-purpose participatory telerobotics architecture, as depicted in Figure 1, consists of six distinct stages:

1. selection of a human to serve as an "avatar" for remote control by an "operator";

2. initiation of one or more network connections to a mobile device near the human;
3. submission of one or more requested actions from the operator to the device;
4. presentation of some or all of the requested actions by the mobile device to the human avatar;
5. scoring of compliance of human with requested actions; and
6. reporting of data on its local environment by the mobile device back over the network.

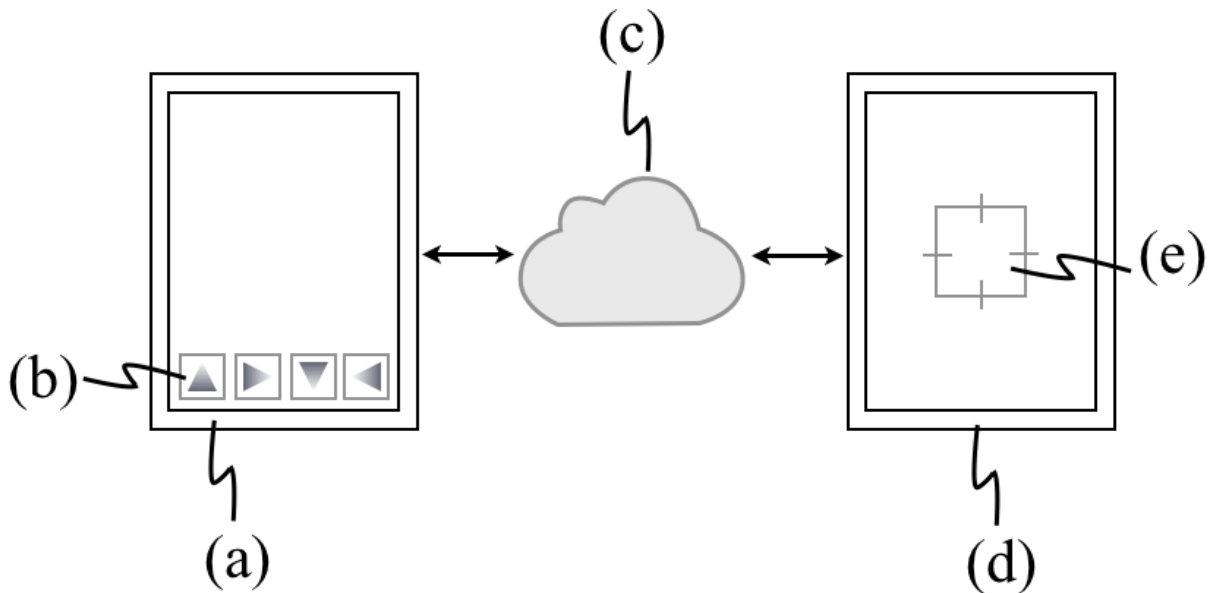


Figure 1. Schematic illustration of our participatory telerobotic implementation. (a) Operator's delegate device. (b) Icons to request actions. (c) Server and network connections. (d) Avatar's delegate device. (e) Representative augmented reality overlay for scoring tolerance region.

The selection stage can encompass more than one human avatar to receive the requested actions in parallel, and can in principle be accomplished by any of the following methods:

- by hyperlink (e.g., for a pre-arranged connection between the operator and the avatar);
- by map view (i.e., by viewing the locations of all available avatars on a map, and choosing a specific avatar);
- by self-reported status search (e.g., by available avatars who self-report anticipating being in the vicinity of a place of interest);
- by specification of a spatial region (e.g., a geographic polygon within which any available avatar would be acceptable); and
- by specification of target coordinates (the default method in our implementation, as shown in Figure 2(a)).

In our proposed general architecture, after target geographic coordinates have been specified, potential human avatars can then in principle be sorted according to an arbitrary function of their present and historical states and then serially notified of an incoming session request. That sorting function could take into account each human avatar's:

- current distance from the specified coordinates, penalizing greater distances;
- historical latencies in responding to session requests, rewarding prompt responses;
- historical compliance with the requested actions, rewarding higher scores;

- recency of activity, taking advantage of the increased likelihood of availability of avatars who have engaged in sessions recently;
- self-reported availability status, allowing potential avatars to opt in or out of receiving session requests; and
- social network connections, leveraging trusted relationships derived from knowledge of the social graph.

Selection and initiation of the network connection are mediated by an intermediate server, positioned on the network in order to minimize end-to-end latency⁶. If the connection involves only a single avatar, the audiovisual connections is peer-to-peer between the operator's and avatar's delegate devices, whereas multi-avatar connections route audiovisual data through intermediate servers. From a prospective avatar's perspective, a session initiation request could in principle include an arbitrary amount of information about the prospective operator and be delivered via a notification, and approved, declined, or ignored by the prospective avatar.

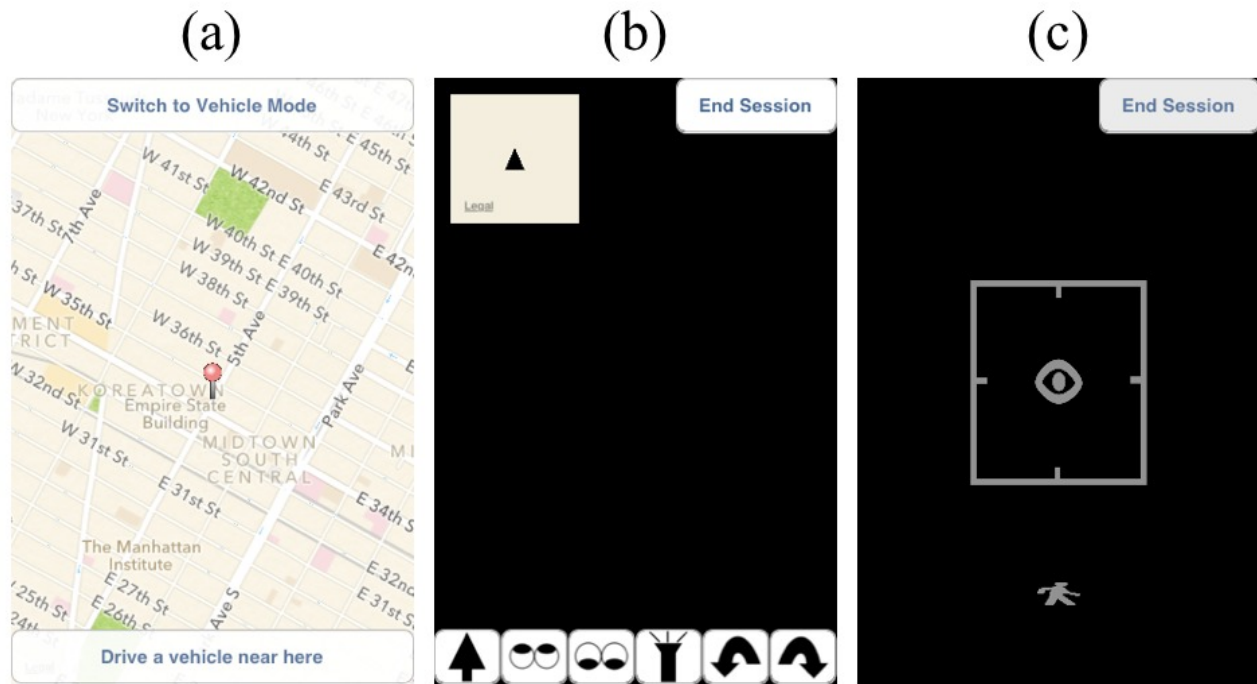


Figure 2. Screenshots of our participatory telerobotic implementation. (a) Map view on operator's delegate device. (b) Driving view on operator's device. (c) Vehicle view on avatar's device.

Supported avatar action requests in our implementation, as partially shown in Figure 2(b), include asking the human avatar to:

- translate the device delegate toward or away from the direction in which a preassigned part of it (the rear face, in our implementation) is facing;
- rotate the device clockwise or counterclockwise relative to the local axis of gravity;
- tilt the device forward or backward;
- translate the device left or right; and
- rotate the device clockwise or counterclockwise about the axis in which a preassigned part of it (again, the rear face in our implementation).

Additionally, in our implementation, actions that do not require human avatar intervention can also be requested, such as turning on and off the LED torch component of the device delegate.

The requests that do require human action are then presented to the human via the device delegate's display in the form of a scored augmented reality "game." In the game, icons for target ground speed and device orientation are displayed as dynamic overlays over a copy of the local video stream, as shown in Figure 2(c), with the objective being to keep both icons as close as possible to the center of a target box. Points are graphically awarded in real-time as a function of the distance of both icons from the center of the target box. In our implementation, the incremental score increase at each timing step was given by

$$\Delta s = \begin{cases} 100 \left(1 - \frac{|h|}{h_{\text{tol}}} \right) \left(1 - \frac{|p|}{p_{\text{tol}}} \right) \left(1 - \frac{|v|}{v_{\text{tol}}} \right) & \text{for } |h| \leq h_{\text{tol}}, |p| \leq p_{\text{tol}}, |v| \leq v_{\text{tol}} \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where h denotes the device heading displacement from the target value, p denotes the device pitch displacement from the target value, v denotes the device ground velocity displacement from the target value, and h_{tol} , p_{tol} , and v_{tol} denote the respective tolerances defining the visible target box size.

As a result, the desired magnitude and direction of both the desired ground velocity and device orientation are seamlessly visualized in real time, enabling, for example, the operator to request running versus walking. Our scoring implementation was designed to leverage the modern sensor suite built into most modern mobile devices, and uses as input streams:

- the GPS receiver (determining the device's location and speed); and
- the accelerometer, gyroscope, and magnetometer (determining the device's 3-D orientation and angular velocity).

In future implementations, we also anticipate using the back-facing camera to recognize barcodes as a method for supporting augmented reality overlays registered to specific environmental objects⁷.

4. DISCUSSION

We expect that our "participatory telerobotics" system, when fully deployed, will have many uses, from intelligence and defense to public health, tourism and retail.

Using a one-to-many anonymized interface, intelligence analysts could rapidly mobilize and "activate" up to thousands of individuals to survey a region, look for suspicious activity, or even silently lead individuals away from danger. Currently, ubiquitous monitoring has high labor and technical costs, as well as substantial infrastructural hurdles. With a participatory telerobotics system, the costs would be substantially reduced, and might be expanded and contracted at will, leveraging civilians to provide micro-aid to the intelligence agencies.

Doctors and public health officials could use such a system to rapidly diagnose health-related problems from remote locations, and instruct individuals in how to administer topical treatments, take medications, etc. Using the increasingly high-resolution cameras on consumer smartphones, doctors could remotely direct patients to move their bodies and zoom in appropriately in order to diagnose and triage visible conditions. Health practitioners could also check-in on infirm patients living alone or in groups at a distance.

For consumers, novel, brief tourism experiences might become available, and perhaps even gamified, permitting individuals to briefly activate and direct others so that they can experience new places they would be unlikely or unable to visit. In addition to virtual tourism, new opportunities open up for remotely inspecting hotels, restaurants, and tourist attractions during trip planning.

Similarly rich experiences also promise to open up for retailers and their customers, enabling the latter to examine products available online or at remote locations in a manner far richer than a 2D image would provide. Similarly, online auction bidders would be able to view products not professionally photographed in three dimensions, before placing a bid.

Of course, emerging mobile device form factors, such as heads-up displays, will offer major advancements to our participatory telerobotic experience, enabling even more immersive telepresence from a truer visual perspective for all of the above scenarios. With full immersion possible, one might be able to have a tour guide with capabilities of fine remote manipulation of the environment not possible with the present generation of telepresence robots, at a fraction of the fixed price.

5. CONCLUSION

In conclusion, we have presented a novel "participatory telerobotics" system that generalizes the existing concept of participatory sensing to include real-time teleoperation and telepresence by treating humans with mobile devices as ad-hoc telerobots. Because of its ability to interactively elicit physical knowledge and operations through ad-hoc collaboration, we anticipate that our participatory telerobotics system will have immediate applications in the intelligence, retail, healthcare, security, and travel industries.

REFERENCES

- [1] International Telecommunications Union, "The World in 2011: ICT Facts and Figures," <<http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2011.pdf>> (6 April 2013).
- [2] Corley, A.-M., "The reality of robot surrogates," IEEE Spectrum, 23 September 2009, <<http://spectrum.ieee.org/robotics/humanoids/the-reality-of-robot-surrogates/0>> (6 April 2013).
- [3] Guizzo, E., "When my avatar went to work," IEEE Spectrum, September 2010, <<http://spectrum.ieee.org/robotics/industrial-robots/when-my-avatar-went-to-work>> (6 April 2013).
- [4] Brabham, D. C., "Crowdsourcing as a model for problem solving," *Convergence: The International Journal of Research into New Media Technologies* 14(1), 75-90 (2008).
- [5] Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M. B., "Participatory sensing," In *Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications*, 117-134 (2006).
- [6] Wissner-Gross, A. D. and Freer, C. E., "Relativistic statistical arbitrage," *Phys. Rev. E* 82, 056104 (2010).
- [7] Wissner-Gross, A. D. and Sullivan, T. M., "Multicolor symbology for remotely scannable 2D barcodes," *Proc. SPIE* 6623, 662304 (2008).