



Mobile Mixed-Reality Interfaces That Enhance Human–Robot Interaction in Shared Spaces

Jared A. Frank, Matthew Moorhead and Vikram Kapila*

Mechatronics, Controls, and Robotics Laboratory, Mechanical and Aerospace Engineering Department, NYU Tandon School of Engineering, Brooklyn, NY, USA

OPEN ACCESS

Edited by:

Trung Dung Ngo,
University of Prince Edward
Island, Canada

Reviewed by:

Fady Alhajjar,
RIKEN Brain Science Institute
(BSI), Japan
Yongsheng Ou,
Shenzhen Institutes of Advanced
Technology (CAS), China

*Correspondence:

Vikram Kapila
vkapila@nyu.edu

Specialty section:

This article was submitted to
Vision Systems Theory,
Tools and Applications,
a section of the journal
Frontiers in Robotics and AI

Received: 21 October 2016

Accepted: 10 May 2017

Published: 09 June 2017

Citation:

Frank JA, Moorhead M and Kapila V
(2017) Mobile Mixed-Reality
Interfaces That Enhance Human–
Robot Interaction in Shared Spaces.
Front. Robot. AI 4:20.
doi: 10.3389/frobt.2017.00020

Although user interfaces with gesture-based input and augmented graphics have promoted intuitive human–robot interactions (HRI), they are often implemented in remote applications on research-grade platforms requiring significant training and limiting operator mobility. This paper proposes a mobile mixed-reality interface approach to enhance HRI in shared spaces. As a user points a mobile device at the robot’s workspace, a mixed-reality environment is rendered providing a common frame of reference for the user and robot to effectively communicate spatial information for performing object manipulation tasks, improving the user’s situational awareness while interacting with augmented graphics to intuitively command the robot. An evaluation with participants is conducted to examine task performance and user experience associated with the proposed interface strategy in comparison to conventional approaches that utilize egocentric or exocentric views from cameras mounted on the robot or in the environment, respectively. Results indicate that, despite the suitability of the conventional approaches in remote applications, the proposed interface approach provides comparable task performance and user experiences in shared spaces without the need to install operator stations or vision systems on or around the robot. Moreover, the proposed interface approach provides users the flexibility to direct robots from their own visual perspective (at the expense of some physical workload) and leverages the sensing capabilities of the tablet to expand the robot’s perceptual range.

Keywords: interaction, interface, robotics, manipulation, mixed-reality, tablet, vision, workspace

1. INTRODUCTION

While robotics technologies and applications have experienced accelerating advances, with robots vacuuming homes, assembling automobiles, exploring planets and oceans, and performing surgeries, their potential remains limited by their ability to effectively interact with people. Recent efforts have explored scenarios in which service robots function alongside people in shared workspaces (Shah and Breazeal, 2010). To ensure the physical safety and psychological comfort of users in these interaction scenarios, a high level of mutual attention and awareness are required of the human and robot (Drury et al., 2003). Efforts to maintain these conditions have typically been addressed in the design of the robot, with respect to its sensing (Fritzsche et al., 2011), reasoning (McGhan et al., 2015), and control (Lew et al., 2000), as well as in the modification of the environment (Lenz et al., 2012; Morato et al., 2014) for direct physical HRI applications. However, graphical interfaces

have begun to provide elegant solutions for interacting with home service robots (Lee et al., 2007; Sakamoto et al., 2016), drawing inspiration from interfaces used by trained engineers to remotely operate robotic vehicles and manipulators. Recent efforts have also sought new interface strategies that permit non-technical users to intuitively interact with robots by using non-verbal gestures captured through vision (Waldherr et al., 2000), touch (Micire et al., 2009), and inertial sensing (Kao and Li, 2010). However, current implementations often confine users to computer stations that can be costly, limited in mobility, or contain hardware and software that are unfamiliar or inconvenient for lay users. For seamless interaction with robots in shared spaces, readily accessible mobile solutions offer a compelling opportunity.

Recent advances in mobile technologies have allowed state-of-the-art features like image processing, multitouch gesture detection, device attitude estimation, and 3D graphics rendering to be integrated on mobile devices such as smartphones and tablet computers to provide portable interfaces for enhanced HRI. Moreover, with their familiarity and ease of use, mobile devices can support intuitive applications to operate robotic systems with comparable performance and usability *vis-a-vis* conventional research-grade interfaces, for a fraction of the cost and training (Su et al., 2015). Although the principles of efficient HRI have begun to be implemented in the design of mobile interfaces, such as those that enable shared or adjustable autonomy in interactions with service robots (Muszynski et al., 2012; Birkenkamp et al., 2014), they have principally been investigated in remote operations rather than in shared human–robot spaces. This presents a problem, since interaction techniques utilized in teleoperation are not necessarily directly applicable to interactions in close quarters.

In teleoperation scenarios, perception of the remote environment often relies on visual feedback provided from either egocentric perspectives using cameras mounted on the robot (Menchaca-Brandan et al., 2007) or exocentric perspectives using cameras mounted in the environment (Hashimoto et al., 2011). Several usability issues have been associated with these perspectives that can compromise the operator's situational awareness. For example, the limited field of view of cameras can cause a keyhole effect, in which operators miss important events that occur offscreen (Woods et al., 2004), requiring them to consult additional sensor information and store mental models of the remote environment in short-term memory (both of which introduce significant cognitive workload) to maintain situational awareness (Goodrich and Olsen, 2003). Moreover, latency and low image quality as a result of limited bandwidth, as well as orientation and frame of reference issues that stem from unnatural camera viewpoints, can disrupt situational awareness and sensations of telepresence (Chen et al., 2007). Proposed solutions to these challenges have included training operators to utilize multiple, multimodal, or ecological displays and controls (Hughes and Lewis, 2005; Marín et al., 2005; Nielsen et al., 2007; Green et al., 2008a). Moreover, virtual and augmented reality have played a fundamental role in recent efforts, providing environments that can serve as effective tools for spatial communication (Green et al., 2008b) and visualizations that promote situational awareness despite limited access to on-site cameras, suboptimal positioning of cameras, poor viewing angles, occlusion, etc.

(Ziaei et al., 2011). Specifically, Milgram et al. (1993) have examined the use of gesture-based interactions with virtual objects as a means to communicate spatial information to robots to command and control their performance of tasks with physical objects.

Many of the usability issues encountered with conventional egocentric and exocentric visual perspectives in teleoperation scenarios can be compounded when used in shared spaces. For example, using robot-mounted cameras to render egocentric perspectives can cause a keyhole effect, while exocentric views require the installation of cameras in the environment. With either perspective, users can experience orientation and frame of reference issues as they attempt to adapt to a different visual perspective from their own. When users are collocated with the robot, they may tend to switch their attention between the interface screen, which presents one visual perspective, and their direct perspective of the robot, which can lead to increased mental workload (Chen et al., 2007). Moreover, computational overhead associated with encoding and streaming video to the mobile device will introduce latency that can obstruct performance and user comfort. By leveraging the mobility, computational power, and embedded sensors of the interface device, many of these usability issues can be avoided.

This paper aims to investigate how a mobile mixed-reality interface approach implemented on mobile devices can enhance the interactive communication of spatial information with robots for object manipulation in shared spaces. Rather than limiting the user experience with unfamiliar, uncomfortable, and expensive research-grade equipment, the mobility of smartphones and tablets affords operators the flexibility of using an interface while in close proximity to the robot. By utilizing the video of the shared workspace from the device's rear-facing camera, a number of benefits emerge over conventional approaches. Orientation and frame of reference issues are averted since visual feedback is presented from the perspective of the user. Thus, the device screen can act as an interactive window that allows users to command the robot by interacting directly with the world, thus reducing the cognitive load associated with attention switching. Moreover, latency arising from encoding and streaming video from robot-mounted or environment-mounted cameras is eliminated, increasing responsiveness of the interface.

By implementing user interaction as well as robot perception, reasoning, and control in the user's frame of reference, a common ground is established between the user and robot that facilitates mutual awareness through information sharing. This information sharing can be used to improve situational awareness by granting users intuitive visual access to information, such as the state of the robot or status of a task, through augmented graphics. Since the shared visual perspective is mobile, it is able to capture regions of the workspace that are undetectable by conventional egocentric and exocentric views. Thus, the proposed mobile mixed-reality interface approach can be used to provide the robot with spatial information that can enhance and in some cases replace the information from the robot while planning and controlling its motion, due to sensorial and mechanical limitations of the robot (constraints on the field of view of its cameras or on the reachable space of its limbs).

The paper is organized as follows. In Sections 2 and 3, the materials and methods used in this study are described, respectively.

Section 2 introduces the robot, the workspace, and interface device technology used in the design of the study. Then, Section 3 presents the three mobile human–robot interface design strategies investigated in the study. Section 4 outlines the user evaluation conducted to compare the three interface strategies under investigation, including the experimental procedure, the task performed by participants, and the assessment and analysis methods utilized. Section 5 discusses the results of the evaluation, while Section 6 provides concluding remarks and future directions of the research.

2. SYSTEM OVERVIEW

The system used in this study consists of a humanoid robotic platform with two 6 degree-of-freedom (DOF) arms, a table with an assortment of blocks of different colors, a computer station for video recording and streaming, and a tablet device that is held by the user and provides a mixed-reality interface for interacting with the robot, as shown in **Figure 1**. The hardware and software architectures employed in this study are adapted from the study of Frank et al. (2016), in which the feasibility, precision, and user experience associated with mobile mixed-reality interfaces for HRI were recently investigated.

2.1. Robot

The robotic platform used in this study is a humanoid with two 6-DOF arms that can manipulate objects on a table. The robot is programmed to receive, over Wi-Fi, commands that contain desired poses for the robot's tool to pick up and place objects in its workspace. First, an algorithm determines which of the robot's arms to use to pick up and place the object. Then, an inverse kinematic model computes the angles required to orient the arm's joints such that its tool is brought to the desired pose. By constraining the tool to be oriented downward in a vertical plane, the robot's 6-DOF inverse kinematic model is decoupled into two 3-DOF solutions known as the inverse position kinematics and the inverse orientation kinematics of the arm (Spong et al., 2006). The inverse position kinematics give the joint angles necessary to position the wrist center of the arm, and the inverse orientation

kinematics give the joint angles necessary to orient the tool. An algorithm plans a path to complete the object manipulation task by generating a sequence of intermediate poses for the tool. Additional information regarding the robot's kinematics and path planning is provided in the study of Frank et al. (2016). The robot also has a stereoscopic vision system that consists of two small webcams mounted rigidly to a pan-and-tilt platform. In this study, the video captured by this system is utilized for HRI in only one of the three interface designs, *viz.*, the one using an egocentric perspective.

2.2. Workspace

The robot is surrounded by a symmetrical workspace composed of two rectangular sections, one on each side of the robot and one semicircular section in front of the robot (**Figure 2**). Directly in front of the robot on the semicircular section are three containers, a red, a green, and a blue. On each of the three sections of the workspace is a colored square block that must be picked up and placed by the robot into the container of corresponding color. On the semicircular section is a green block, on the rectangular section on the robot's left side is a red block, and on the rectangular region on the robot's right side is a blue block. Affixed to the surfaces of the workspace are twelve visual markers, which are arranged in a known pattern at predefined locations (**Figure 2**). These markers are used by the interface to establish a shared reference frame for the exchange of spatial information between the operator and the robot, as well as for the realistic display of augmented graphics. To facilitate the localization of the blocks with respect to this reference frame, markers are also affixed to each of the blocks. Each of the fifteen markers used in this study contains a distinct pattern that allows the marker to be uniquely identified. Mounted to the ceiling directly above the workspace is a camera, which is used to provide an overhead view. The video captured by this camera is recorded for all experimental trials performed in the study, but is utilized for HRI in only one of the three interface designs, *viz.*, the one using an exocentric perspective.

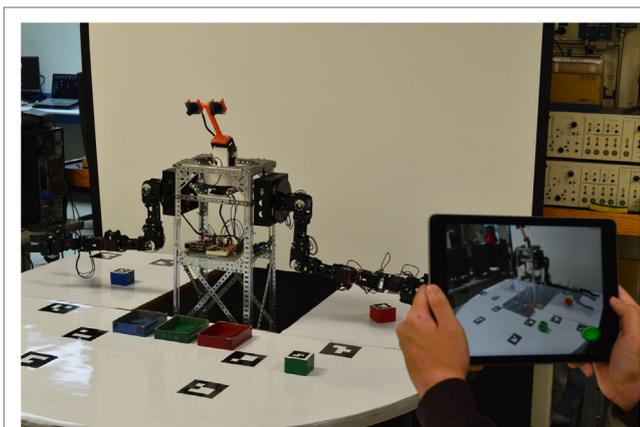


FIGURE 1 | The environment used to conduct the human–robot interaction study.

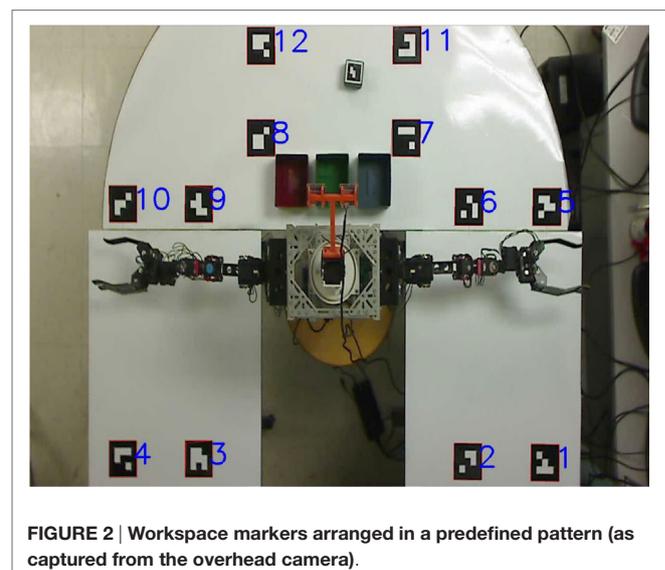


FIGURE 2 | Workspace markers arranged in a predefined pattern (as captured from the overhead camera).

2.3. Computer Station and Interface Device

To record and stream video from the overhead and robot-mounted cameras, a computer station is installed beside the robot's workspace. The interface device accesses these video streams in two of the three designs (i.e., the ones that use the egocentric and exocentric perspectives) by connecting to the station with a client–server architecture. The interface device used in conducting this study is an Apple iPad Pro. Released in 2016, this tablet computer has a 9.7" (250 mm) screen with a $2,048 \times 1,536$ pixel multitouch display (264 pixels per inch), a 12-megapixel 4 K resolution rear-facing camera, 256 GB of flash memory, a 2.16 GHz dual-core processor, and 4 GB of RAM. These specifications make the iPad Pro, and the generations of mobile devices to come, uniquely suited to provide mobile mixed-reality human–robot interfaces. This is so because these devices are capable of capturing and processing live video sequences, rendering 3D augmented graphics, recognizing multitouch gestures, estimating device attitude, and communicating over wireless networks, synchronously and in real time. Thus, these devices can provide more than intuitive user interfaces; running in the background, their mobile applications have the ability to offload some of the sensing, storage, and computation of the system. This ability is exemplified in one of the three interface designs explored in this study, *viz.*, the one using the rear-facing camera of the tablet to provide a mobile perspective for HRI.

3. INTERFACES

To explore mixed-reality as an enabler of HRI and investigate ways in which mobile technologies may enhance HRI in shared

spaces, three mobile interfaces are developed that utilize distinct design strategies. In the front end, each interface provides a mixed-reality view of the shared space that consists of a live video of the space augmented with computer-generated graphics, such as reference planes and axes registered in the view and virtual objects linked to corresponding physical objects in the space. The principal difference between the three alternative interface designs lies in the visual perspective used to render the mixed-reality environment. In the first interface, the perspective comes from a camera mounted to the robot. In the second interface, the perspective comes from a camera mounted on the ceiling above the workspace. In the third interface, the perspective comes from the rear-facing camera on board the interface device. All three interfaces provide the same augmented graphics responsible for enhanced visual feedback to promote situational awareness and intuitive commanding of the robot through touch interactions with the graphics. For example, interactive virtual blocks are projected on top of actual blocks detected in the video. Through taps, drags, and rotations of the user's fingers on the touchscreen, the virtual blocks can be intuitively manipulated to desired poses on the screen. A virtual grid is projected onto the surface of the table to provide users with a visual aid for precise placement of the virtual blocks in the mixed-reality workspace. However, due to the limited lengths of the robot's arms and the constraints on their configurations imposed by the grasping approach taken by the robot, not all locations in the workspace can be reached. To inform themselves of the allowable and prohibited regions in the workspace, users can double tap on the screen to toggle on and off the display of a semitransparent cross-sectional plot of the robot's reachable space that is projected with the grid (see **Figure 3**). In this plot, red indicates prohibited regions

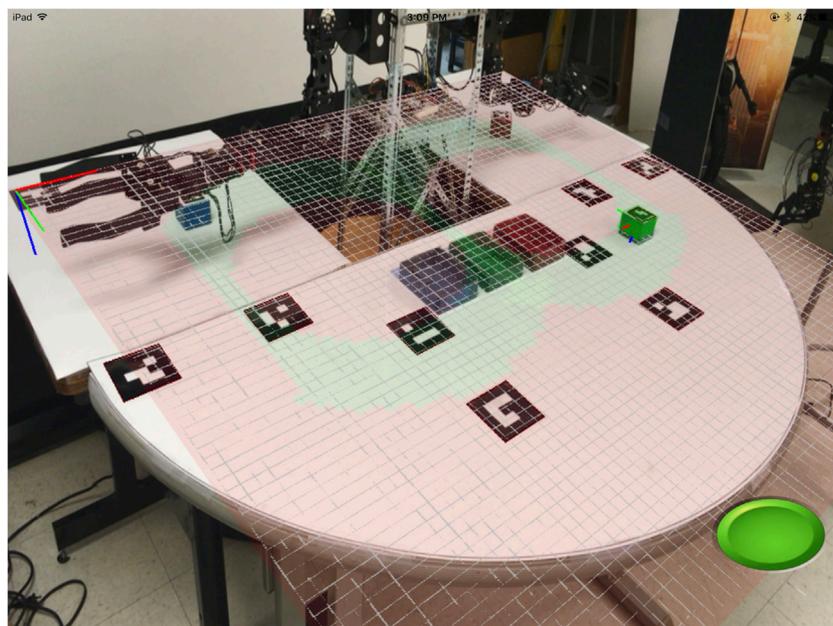


FIGURE 3 | Screenshot of the mixed-reality environment with the plot of the robot's reachable space enabled.

while green indicates allowable regions. Thus, the augmented graphics rendered by the interfaces can support human–robot collaboration for object manipulation tasks in shared spaces by forming mixed-reality graphical environments in which the user and robot exchange relevant spatial and task-specific information.

In the background of each interface, the mobile application must perform several tasks. First, each video frame is processed using an adaptive thresholding technique to detect the image locations of the centers of the markers affixed to the workspace and objects of interest. Since the twelve workspace markers are coplanar, if any four or more workspace markers have been detected in the video frame then an iterative approach is used to estimate the relative pose of the plane of the workspace through the solution of a perspective n -point problem (Oberkampff et al., 1996). To estimate the relative pose of each object of interest, the same approach is applied to the solution of a perspective 4-point problem for each object marker, where the points used are the four corners of the marker. Next, the relative poses are transformed so that the poses of all objects of interest are represented with respect to the fixed coordinate frame established by the workspace markers (see **Figure 4**). These transformations allow both the interface to augment the video with virtual elements that enhance the user’s situational awareness and provided the pose of the robot is known with

respect to the workspace frame, the performance of object manipulation tasks by the robot using the vision-based measurements obtained by the interface.

To determine which object users would like the robot to manipulate and at what pose they would like the robot to place the object, the interface first captures touch gestures (taps, drags, rotations, and releases) of the users’ fingers on the screen. Then, the relative transformation to the workspace frame is used to map users’ touch gestures on the screen to locations and orientations in the plane of the workspace with respect to this reference frame. As users interact with the virtual objects on the screen, the interface uses the mapped poses to display these objects underneath the users’ fingers, creating the sensation that they are moving the block along the surface of the table. Once users have completed their interactions with the virtual blocks and are satisfied with their interactions, a press of a button communicates the associated spatial commands to the robot that enable it to manipulate the corresponding physical object to the appropriate desired pose. Open source libraries are used to perform these tasks, such as the Open Source Computer Vision (OpenCV) library to process video frames, Open Source Graphics Library for Embedded Systems (OpenGL ES) to render augmented graphics, and the CocoaAsyncSocket library to communicate with the robot over Wi-Fi using the TCP/IP protocol.

3.1. Conventional Egocentric Interface

With the first interface, users interact with the robot while being provided visual feedback from the perspective of one of the robot’s cameras, which is mounted on a pan and tilt system. The limited field of view of this camera is only capable of capturing one of the three sections of interest in the workspace at any given time (see **Figure 5**). Thus, users are forced to pan and tilt the camera between these three sections if objects are to be moved from one section of the workspace to another. To give users the most natural control of the pan and tilt of the robot’s camera system, while reserving touchscreen gestures for interactions with virtual elements in the mixed reality environment, the interface uses the device’s accelerometer and gyroscope to estimate the device’s attitude. These estimates, which are obtained at a rate of 30 Hz, are represented as Euler angles with respect to a fixed reference frame whose Z axis is vertical along the downward-facing gravity vector and whose X axis points along the centroidal axis in the longitudinal direction of the tablet from when the interaction was initiated. Readings from

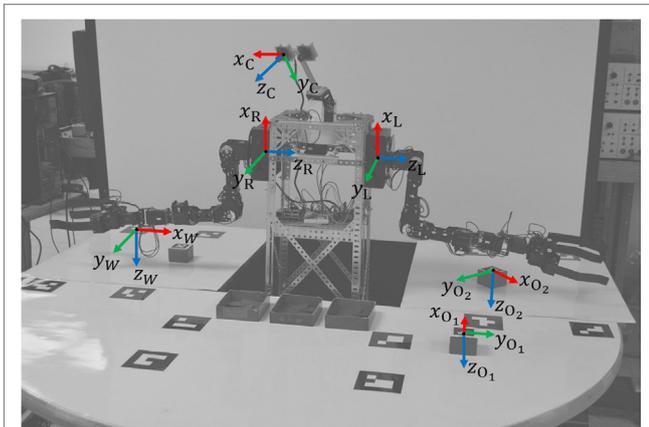


FIGURE 4 | Coordinate frames used by both the interface and robot to communicate spatial information.

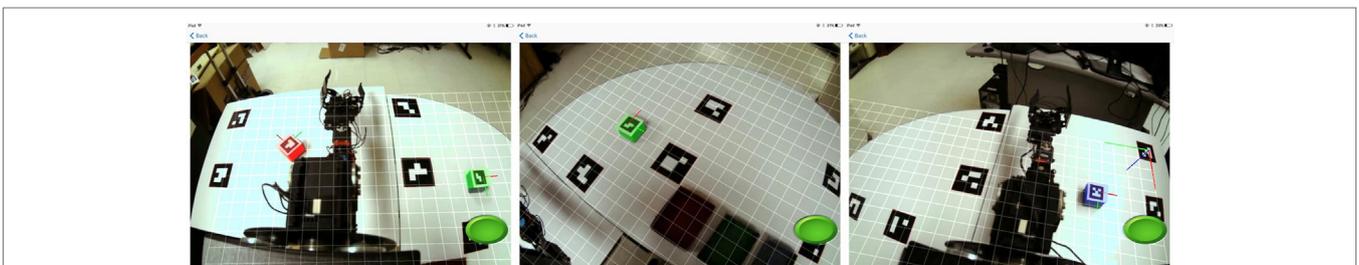


FIGURE 5 | Egocentric views of the workspace as captured from the robot-mounted camera.

the magnetometer are used to correct the direction of the X axis to maintain long-term accuracy. **Figure 6** shows the yaw rotation θ that transforms the XYZ frame to an intermediate frame $x'y'z'$ and the roll rotation ϕ that transforms the intermediate frame $x'y'z'$ to the frame xyz attached to the device. Since the yaw and roll of the device are directly analogous to the pan and tilt, respectively, of the robot’s camera system, a one-to-one mapping is used to command the attitude of the robot’s camera system from the estimated attitude of the device. However, the camera system can pan only 270° and tilt only $\pm 45^\circ$. Due to these limitations in the mobility of robot’s camera system and the limited field of view of its cameras, there are locations in the

workspace that are unobservable using this interface (i.e., close to the robot as well as behind the robot).

3.2. Conventional Exocentric Interface

To improve upon the limitations encountered with the first interface in observing the workspace, a second interface provides users with an expanded and more natural view from a camera mounted to the ceiling above the workspace (see **Figure 7**). Since the camera is fixed, no device motion is captured by the interface. Users simply interact with the mixed-reality environment generated from an overhead perspective of the workspace. However, there remain locations in the workspace that cannot be observed from the perspective presented by this interface (i.e., underneath the robot’s arms). Due to the inability to adjust the perspective provided by this interface, users still cannot interact with objects everywhere in the robot’s reachable space without the installation of additional cameras.

3.3. Proposed Mobile Mixed-Reality Interface

With the third interface, users hold the tablet such that its rear-facing camera is pointed at the robot and its workspace from an arbitrary perspective. In this interface design, the tablet’s camera captures the video used to render the mixed-reality environment on the screen for interacting with the robot (see **Figure 8**). The mobility afforded with this interface approach offers a few distinct advantages over the other two approaches, particularly the ability to access every location in the robot’s reachable space without the need for installed sensing or computation beyond the mobile device. These features allow for more seamless and economic implementation of instrumented human–robot interactions in a wider range of applications.

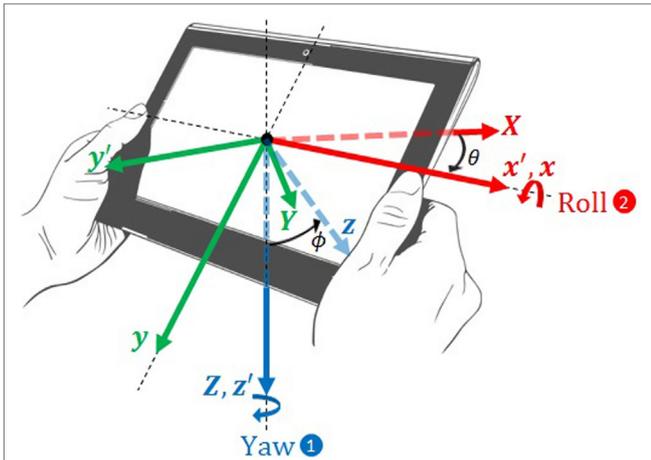


FIGURE 6 | Rotations considered between coordinate frames in estimating the roll and yaw of the device.

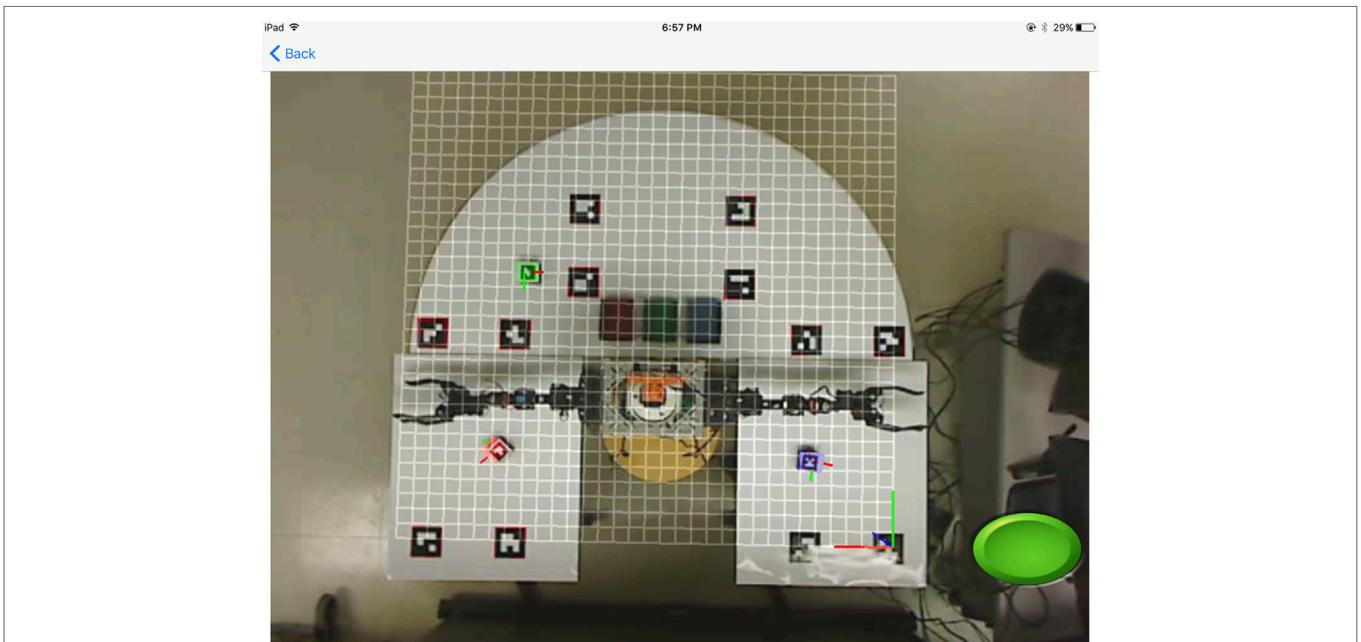


FIGURE 7 | Exocentric view of the workspace as captured from an overhead camera.

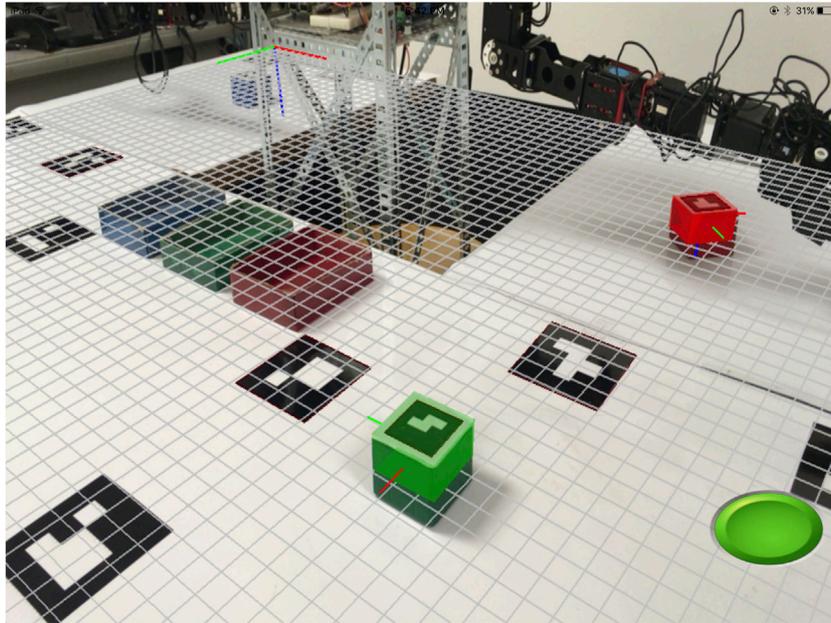


FIGURE 8 | Mobile view of the workspace as captured from the tablet's rear-facing camera.

4. EVALUATION

To assess the user experience and performance in interacting with the humanoid robot using the developed interfaces, an experimental study was conducted with participants, who were each asked to command the robot to complete an object manipulation task using one of the interfaces. The objective of the task was to pick up each of the three blocks in the workspace and to place it in the container of matching color in as little time as possible. See <http://engineering.nyu.edu/mechatronics/videos/mmrmanipulation.html> for a video illustrating user interaction for task completion with each of the three interfaces. The interface assignment for each participant was generated randomly to prevent any potentially biased data. The proposed mobile mixed-reality interface has a distinct advantage over the other two interfaces since, if one of the blocks were placed directly underneath either of the robot's arms, that block will be out of the field of vision of both the egocentric and exocentric interfaces. This would cause participants to lose visible and interactive access to the block from either of these interfaces. This advantage of the mobile mixed-reality interface exists in general when only a single camera is mounted to either the environment or the robot, since the robot and obstacles in the environment will always occlude a portion of the workspace. Thus, to level the playing field and ensure consistent and unbiased experimental data, the three blocks started in locations that were visible with each of the three interfaces.

First, each participant completed a preliminary assessment, which was used to gage the participant's familiarity with mobile devices, AR, and robots. Then, participants were given a one-minute introduction to the task to be performed and the interface to be used to perform the task. Participants were instructed about

the objective of the task, i.e., to place each block in the container of the same color. Participants using the mobile interface were instructed to keep as many workspace markers on the screen as possible to achieve the most accurate performance. Each participant was given only one chance to place each block.

Prior to performing the task, with the containers removed from the workspace, each participant was given one pretrial to practice commanding the robot to pick up and place the green block to an arbitrary pose in the workspace. For half of the participants (10 in each group of 20 participants), the augmented plot of the robot's reachable space was enabled, and for the other half the plot was disabled. Due to the lengths and configurations of the robot's arms, and the grasping approach implemented in the study, approximately 80.1% of the workspace is unreachable, 18.7% can be reached by one of the robot's arms, and 1.2% can be reached by both arms. Thus, the pretrials were conducted to evaluate the extent to which the augmented plot improves participants' situational awareness such that they can successfully complete an object manipulation to a reachable location in the workspace.

As participants performed the task, they stood facing the robot from across the semicircular section of the workspace. With the third interface, participants were encouraged to move around as needed to see blocks that might be hidden behind the robot. To assess the participants' performance of the task using each of the interfaces, the time elapsed and the success of each trial were manually recorded by the evaluator observing the trial. Since participants were asked to pick up and place three blocks, their success was graded as the percentage (from 0 to 100%) of these six operations that they completed successfully.

After participants performed the task, they were asked to respond to an evaluation that assessed their experiences in two

parts. In the first part of the evaluation, the participants were asked to indicate their level of agreement with nine positive and negative statements on a 5-point scale. In the second part of the evaluation, the NASA Task Load index (NASA-TLX) was used to assess the workload associated with using each interface (Hart, 2006). This index, which is calculated on a 0–100 scale, indicates the overall demand of the task on the participant and is composed of six categories: mental workload, physical workload, temporal workload, performance, effort, and frustration. To obtain the NASA-TLX, participants complete a workload evaluation sheet in which, for each of the six categories, they mark an “X” on the 0–100 scale that is separated into 20 equally spaced segments of five points. In the Raw TLX (RTLX) variation of the assessment, the ratings reported for each category are uniformly weighed and used to calculate an average overall workload for the task. To aid them in carefully responding to the workload evaluation, participants were provided with a guide that described each of the categories in detail. Finally, at the end of the evaluation, participants were provided a blank sheet of paper where they were encouraged to leave comments and feedback about their experience while performing the task.

5. EVALUATION RESULTS AND DISCUSSION

To evaluate aspects of the performance and user experience associated with interacting with the humanoid robot using each of the three user interfaces, the study was conducted with 60 undergraduate engineering student participants. All 60 participants owned a smartphone and 61.7% owned a tablet. Since participants were students in engineering, roughly half (50.6%) had interacted with, worked on, or built a robot before, while the rest (49.6%) had neither seen nor interacted with a robot up close prior to their participation in the study. Due to recent releases of popular games and utilities with augmented reality, 38.1% of the participants reported having at least one recent mobile augmented reality experience. **Figure 9** summarizes participants’ self-reported levels of familiarity with each of the emerging technologies relevant to the study.

5.1. Performance

As participants interacted with the robot using each of the interfaces, their performance of the task was evaluated by recording the number of successful pickups and placements of each block on the table and the amount of time taken to complete the task.

5.1.1. Success Rate

Of the 60 participants in the study, 60% performed the pretrial successfully. Of these 36 participants, 72.2% had the augmented plot of the robot’s reachable space enabled. Of the 24 that failed the pretrial, 83.3% did not have the augmented plot enabled. **Table 1** shows the number of participants who had the augmented plot enabled and passed the pretrial, those who had the augmented plot enabled but failed the pretrial, those who did not have the plot enabled but passed the pretrial, and those who did not have the plot enabled and failed the pretrial. The

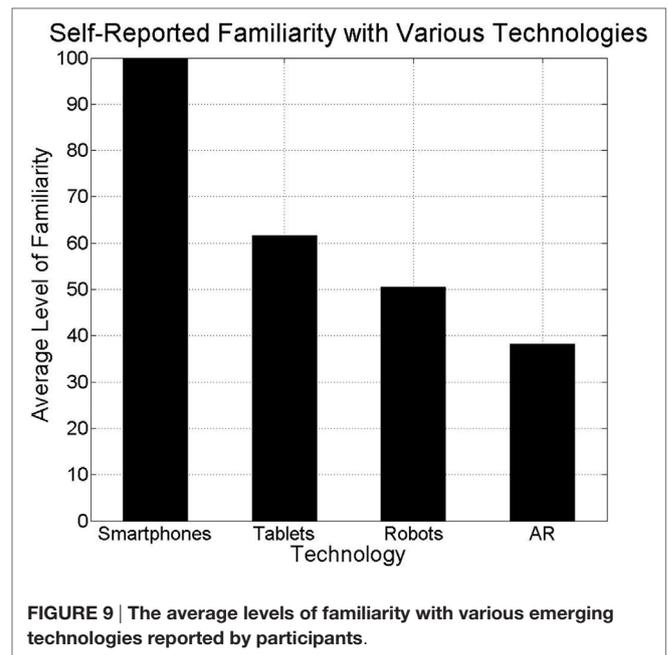


TABLE 1 | The performance of participants in the pretrial, with and without access to the augmented plot.

| | With AR plot | Without AR plot | Total |
|-----------------|--------------|-----------------|-------|
| Passed pretrial | 26 | 10 | 36 |
| Failed pretrial | 4 | 20 | 24 |
| Total | 30 | 30 | 60 |

majority of participants who were provided the augmented plot succeeded in accomplishing the test (86.67%) and those who were not provided the augmented plot failed in accomplishing the test (66.67%). To determine whether access to the augmented plot had a significant effect on participants’ success in the pretrial, the *N-1* Two-Proportion Test proposed by Campbell (2007) was used and resulted in rejection of the null hypothesis that participants perform equally well with and without the augmented plot ($z = 4.181, p < 0.0001$). These results indicate the significant benefit that the augmented plot has in raising participants’ awareness of the robot’s spatial limitations. Note that, although usability issues prevented a small group of participants from passing the pretrials with the augmented plot enabled, 33.33% of the participants were able to accomplish the pretrial successfully without access to the plot. Since 18.7% of the workspace is reachable, the participants without the augmented plot who guessed a reachable location is nearly double of the amount that would be expected as a chance occurrence. This is due to several participants’ preferences to command the blocks to locations directly in front of the robot, where most of the reachable space is concentrated. Note that no statistically significant differences were found between the mean performances of the pretrial achieved with each interface.

Overall, participants performed the object manipulation task very well with each of the interfaces. Of each group of 20 participants, 16, 18, and 17 participants succeeded in picking and

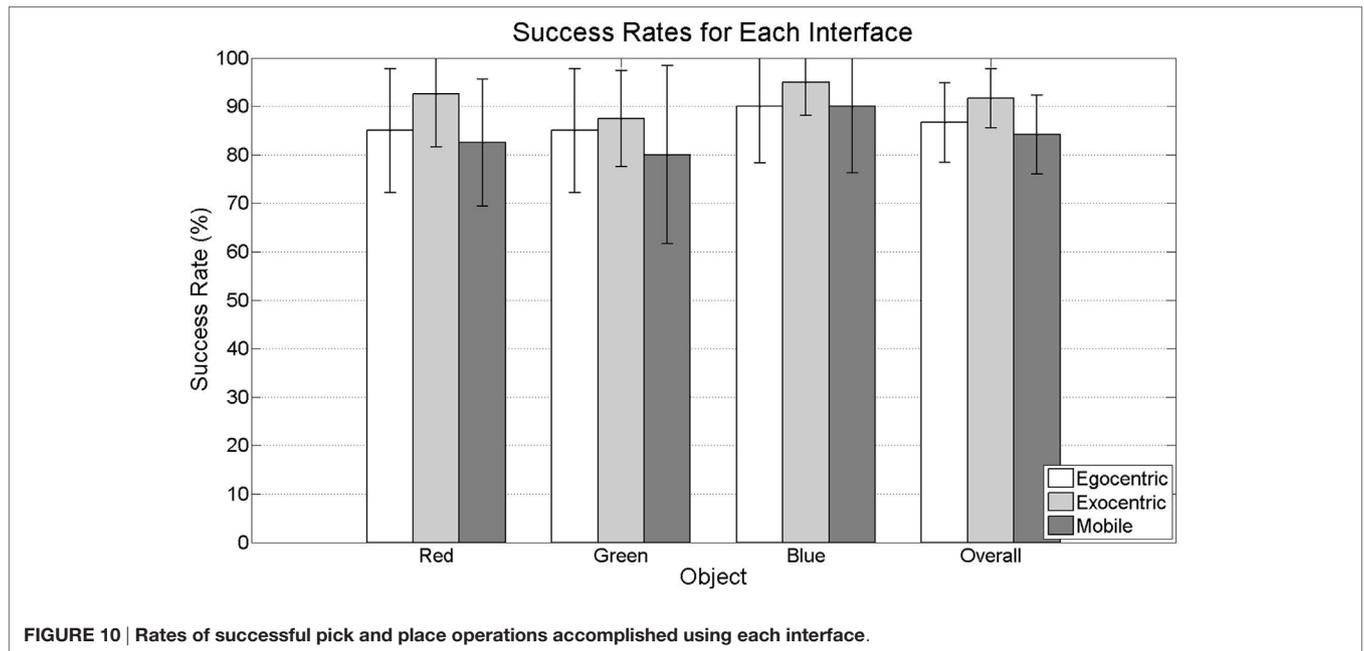


FIGURE 10 | Rates of successful pick and place operations accomplished using each interface.

placing at least 2 of the 3 blocks using the egocentric, exocentric, and mobile mixed-reality interfaces, respectively. Participants performed an average of 86.7%, 84.2%, and 91.7% of their interactions successfully with the red, green, and blue blocks, respectively. **Figure 10** shows the average percentage of successful pick and place operations performed by participants using each interface, both for each of the blocks and overall during a trial. Although **Figure 10** shows that, on average, participants were more successful with the green block and performed better using exocentric interface than the other two, no statistically significant differences were found, neither between interactions with the different blocks nor between the different interfaces. In other words, participants perform the task just as well with the mobile mixed-reality interface (which communicates directly with the robot without the need for robot-mounted or environmentally mounted sensors) as with the egocentric or exocentric interfaces (which require cameras and their associated processing to be installed in the environment or onto the robot).

5.1.2. Time

Although participants completed the task at the same rate using each of the interfaces, the amount of time taken to complete the task differed depending on which interface was used. **Figure 11** presents the sample medians of the times elapsed while performing the task with each of the interfaces (169, 132.5, and 139 s with the egocentric, exocentric, and mobile mixed-reality interfaces, respectively), along with their 95% confidence intervals, calculated according to Sauro and Lewis (2012). With each interface, the same amount of the trial time is spent waiting for the robot to complete its operations on the blocks. Thus, differences in total trial times are largely due to differences in interaction time of the participants, which can arise from variations between participants or between the efforts required to perform the task

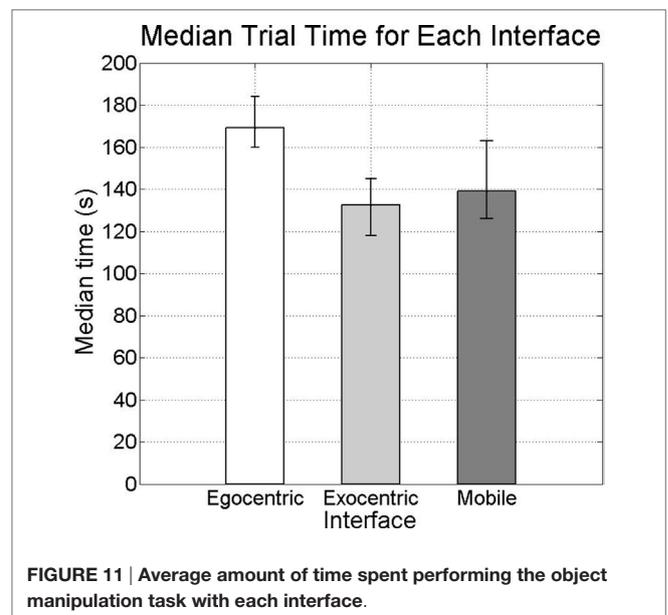


FIGURE 11 | Average amount of time spent performing the object manipulation task with each interface.

using different interfaces. The results of independent two-tailed *t*-tests indicate no statistically significant differences between the egocentric and mobile interfaces or between the exocentric and mobile interfaces. However, a statistically significant difference is found between the egocentric and exocentric interfaces ($t(38) = 7.0636, p = 2.0174 \times 10^{-8} < 0.05$). This difference in time can be explained by the distinct amounts of mental and physical efforts demanded by each interface. Specifically, participants using the exocentric interface do not have to change the visual perspective provided by the interface. Thus, participants only need to become comfortable with the visual perspective once

and can spend most of the interaction time manipulating the augmented graphics. Meanwhile, using the egocentric interface required participants to physically turn their bodies to adjust the perspective of the robot-mounted camera. Use of the mobile interface requires participants to physically move around the environment to adjust the interface's visual perspective. However, since participants with mobile interface choose very different vantage points from which to observe the workspace and take largely varying amounts of time to get settled at these vantage points, there is much larger variation in the time they take to complete the task than participants with the other two interfaces. In fact, the standard deviation of the times taken with the mobile interface (42.1 s) is more than twice the amount with the egocentric interface (16.8 s) and with the exocentric interface (19.9 s). Thus, to calculate the degrees of freedom for the *t*-tests involving the mobile interface, the Welch–Satterthwaite procedure is used to account for these large differences in variances (Sauro and Lewis, 2012). Although the egocentric interface required less physical motion than the mobile interface, it will be shown that participants found the egocentric perspective unnatural and uncomfortable. Thus, participants using the egocentric interface may have spent a large amount of interaction time readjusting to changes in visual perspective.

5.2. User Experience

In addition to the performance achieved, participants' perceptions of their experience form a fundamental part of evaluating the mixed-reality interfaces. The user experience associated with each interface was assessed using three mechanisms: a usability questionnaire, a NASA RTLX self-assessment, and by reviewing participant comments and feedback.

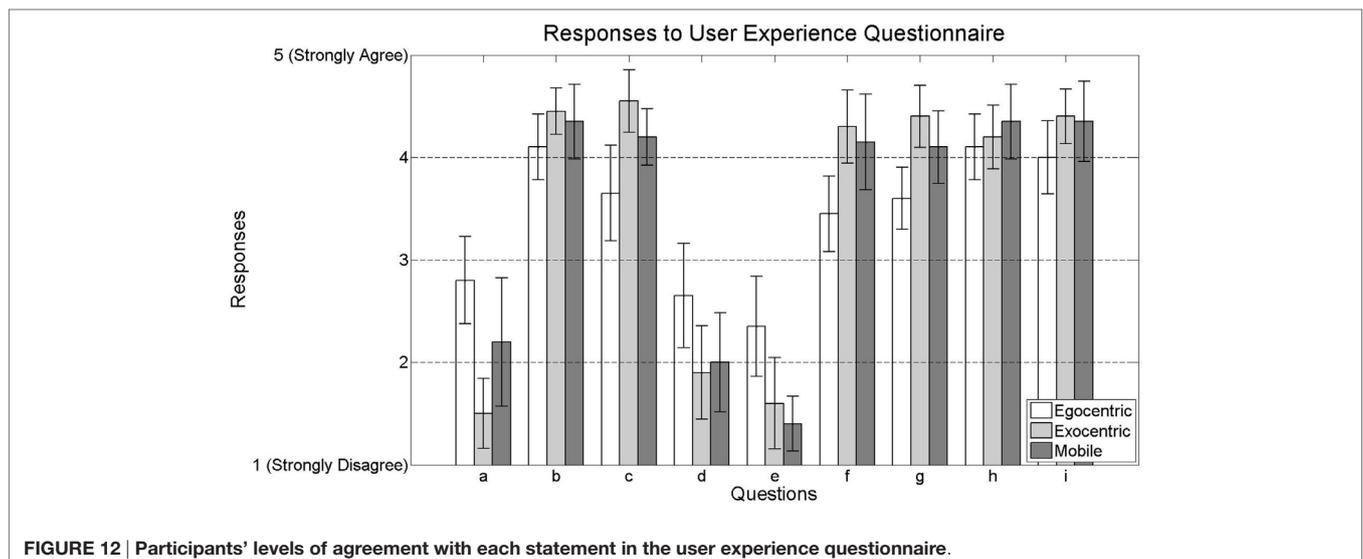
5.2.1. Usability Questionnaire Results

To assess aspects of participants' user experience, a questionnaire inspired by the Post-Study System Usability Questionnaire (Lewis, 2002) was designed and administered to participants. This questionnaire asked participants to indicate their level of agreement

with the following nine positive and negative statements on a 5-point scale (1: strong disagreement and 5: strong agreement).

- It was difficult to interact with the virtual elements on the screen.
- The virtual graphics on the screen were useful visual aids.
- Overall, the application made it easy and fun to interact with the robot.
- I required assistance to interact with the robot.
- It took a long time for me to become comfortable using the application.
- It was easy to place and orient blocks on the table using this application.
- Overall, I felt that I was able to use the application to accurately communicate my intentions to the robot.
- Overall, I would recommend this application to people who work with robots at home or work.
- I would like to see more applications like this for people who may one day have robots at home or work.

Figure 12 shows the participant responses for each statement in the questionnaire. These responses look promising, indicating relatively low amounts of perceived difficulty and relatively moderate to high amounts of satisfaction and perceived performance with the interfaces. By conducting independent two-tailed *t*-tests on the response data, no statistically significant differences were found between the responses given by participants using the exocentric interface and those using the mobile interface. However, several statistically significant differences were found between the responses given by participants using the egocentric interface and those using the exocentric and mobile interfaces. For example, although many participants either disagreed or strongly disagreed that any of the interfaces made it difficult to interact with the virtual elements on the screen, participants using the egocentric interface found that it was more difficult than those using exocentric interface ($t(38) = 4.7721, p = 2.6985 \times 10^{-5} < 0.05$).



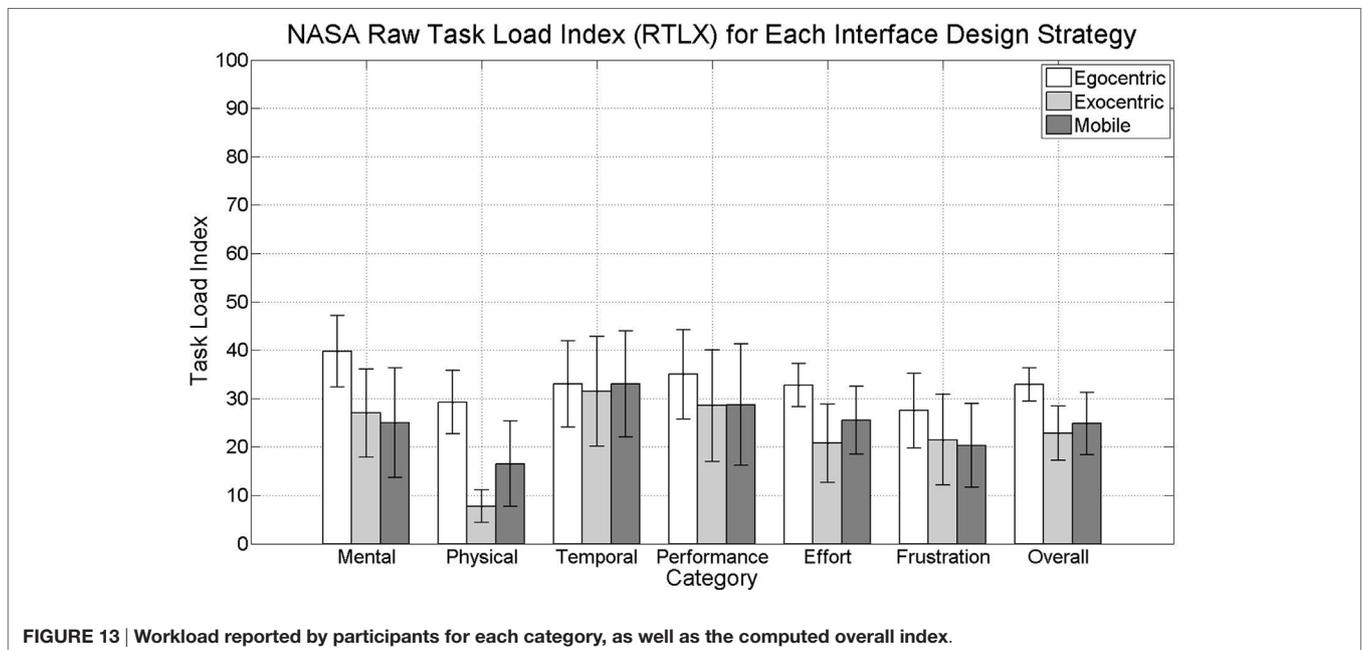
Moreover, participants using the egocentric interface found that it was not as easy to place and orient blocks on the table as those who used either the exocentric interface ($t(38) = 3.3040$, $p = 0.0021 < 0.05$) or mobile interface ($t(38) = 2.3576$, $p = 0.0236 < 0.05$). Overall, participants using the egocentric interface consistently agreed less with the statement that: (1) it was easy and fun to interact with the robot than those who used the exocentric interface ($t(38) = 3.2302$, $p = 0.0026 < 0.05$) or mobile interface ($t(38) = 2.0353$, $p = 0.0488 < 0.05$) interfaces and (2) they were able to use the application to accurately communicate their intentions to the robot than those who used the exocentric interface ($t(38) = 3.7173$, $p = 6.4693 \times 10^{-4} < 0.05$) or mobile interface ($t(38) = 2.1475$, $p = 0.0382 < 0.05$). These differences in perceived difficulty and ability could be due to the fact that using the egocentric interface requires participants to first turn the device to find the object of interest, tap on the object to select it, then turn again to find a view containing the goal location for the object and finally tap and drag on the goal location, whereas the exocentric interface contains both the start and goal locations of each object in the view at all times and does not require alternating sequences of device movements and touchscreen gestures to interact with the robot. However, since the mobile interface sometimes may require some movement to achieve a convenient perspective of the workspace, participants find its level of difficulty somewhere between that of the two conventional interface designs on average.

The difficulty and discomfort associated with needing to move the device to find objects and goal locations in the perspective of the robot explain why the egocentric interface required more time to complete the task than the other two interfaces. This argument is supported by the finding that participants using the egocentric interface responded that they required more time to become comfortable using the interface than participants using

the exocentric interface ($t(38) = 2.2736$, $p = 0.0287 < 0.05$) or mobile interface ($t(38) = 3.4183$, $p = 0.0015 < 0.05$). Moreover, participants using the egocentric interface responded that they required more assistance from the laboratory assistant in order to use the interface than participants using the exocentric interface ($t(38) = 2.1954$, $p = 0.0343 < 0.05$).

5.2.2. Workload

Further insight into the user experience associated with performing the task with each interface is obtained by reviewing the responses to the workload evaluation. Using the RTLX approach, the ratings reported for each category are used to compute the average overall workload index. **Figure 13** shows the mean values of the workload reported by the participants for each category, as well as the mean values of overall workload computed from these reported values. These results are promising since they show that participants using each interface reported a relatively low workload in each of the categories. As was the case with the user experience results, no statistically significant differences were found between the exocentric and mobile interfaces. However, statistically significant differences were found between the overall index computed for the egocentric interface and both the exocentric interface ($t(38) = 3.0592$, $p = 0.0041 < 0.05$) and the mobile interface ($t(38) = 2.2041$, $p = 0.0336 < 0.05$). These differences are due to several statistically significant differences that were found between participants' perceived mental workload, physical workload, and effort with the egocentric interface and the workload amounts perceived by participants using the other two interfaces. Specifically, participants using the egocentric interface reported experiencing both significantly higher mental and physical workload than the participants using either the exocentric interface (mental: $t(38) = 2.1804$, $p = 0.0355 < 0.05$,



physical: $t(38) = 5.8473, p = 0.0343 < 0.05$) or mobile interface (mental: $t(38) = 2.1941, p = 9.2227 \times 10^{-7} < 0.05$, physical: $t(38) = 2.3158, p = 0.0261 < 0.05$). These results are consistent with comments made earlier in the paper regarding the disadvantages of using the egocentric perspective. Note that although participants using the mobile interface, on average, report experiencing more than twice the amount of physical workload than those using the exocentric interface, a large variance in the physical workload perceived by these participants prevents this difference from being statistically significant. This large variance could be due to the fact that there were several options for how participants needed to move to interact with a block that was hidden behind the robot, resulting in different amounts of movement for different participants. Moreover, varying amounts of comfort and engagement with the mobile interface may have influenced participants' perceptions of the physical demands of the task. Despite significant differences in recorded trial times, significant differences were not found between the temporal demand experienced with different interfaces. Furthermore, despite significant differences between the level of difficulty experienced by participants, significant differences were not found in the workload evaluation between the frustration experienced by participants using different interfaces. However, a significant difference was found in the reported effort required to accomplish the task between participants using the egocentric interface and those using the exocentric interface. These results are consistent with the significant differences found in mental and physical workload.

5.2.3. Participant Comments

Further insights regarding the differences in user experiences and perceived workload between the interfaces can be obtained by reviewing the participants' comments and feedback. Of the 60 participants in the study, 44 provided descriptive comments and suggestions for future improvement to the design of the interfaces. Comments included praises, positively biased statements that do not offer recommendations for improvement, such as

The interface definitely made it fun and easy to interact with the robot.—Participant 22 who used the proposed mobile mixed-reality interface.

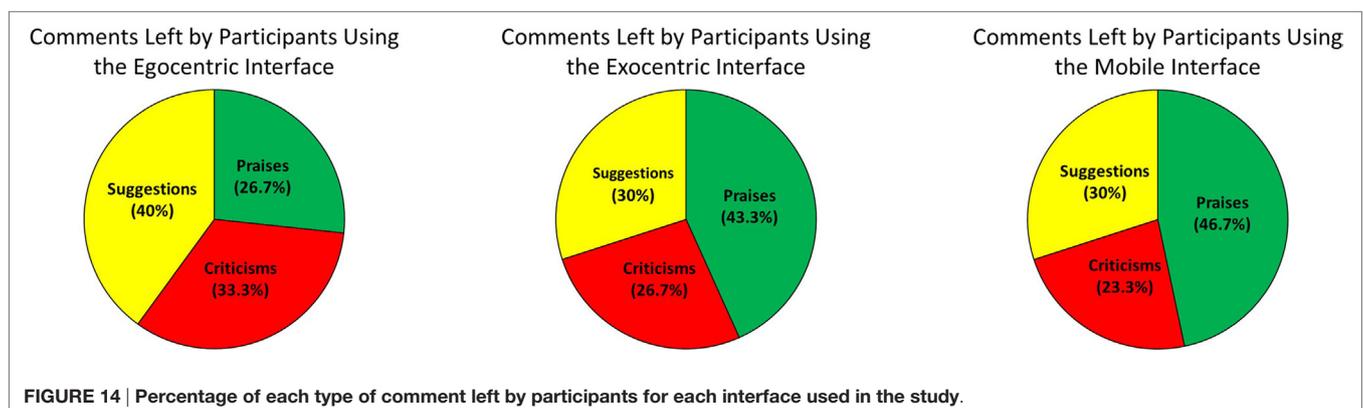
There were also a few criticisms, characterized as negatively biased statements that do not offer recommendations for improvement, such as

The video had significant lag.—Participant 52 who used the conventional exocentric interface.

However, most participants left useful suggestions for improvement that were neither positively or negatively biased, such as

Perhaps make the augmented grid lines less visible so they are more user-friendly.—Participant 18 who used the conventional egocentric interface.

Figure 14 presents the percentages of each type of comment left by participants using each interface in the study. These results show that a majority of the participants were affected enough by their participation in the study to leave meaningful comments for developers and that a sizable portion of participants were satisfied enough to praise their experience in writing. Praises were taken as validation of our arguments for integrating mobile hardware and software in the development of mixed-reality interfaces for interacting with robots, such as the ones proposed in this paper. Many of the praises touched on the satisfaction associated with being able to successfully control the actions of the humanoid robot. Participants were also impressed with the ability to naturally interact with the augmented blocks in the mixed-reality environment and the intimate connection observed between their actions with the augmented blocks and the resulting operations performed by the robot on the physical blocks. Criticisms served as a reminder that this class of mobile mixed-reality human–robot interfaces is still in its infancy, and much work remains to be done to improve the visual and interactive aspects of the interfaces. These criticisms included the observation that a small amount of delay was present in the video provided by the egocentric and exocentric interfaces. Moreover, although the overhead camera was placed at the minimal distance needed to capture the entire workspace, participants felt the perspective used by the exocentric interface was too far above the table, making the blocks appear too small and difficult to interact with. Furthermore, several participants who used the egocentric interface found it uncomfortable to see themselves



performing the task from the perspective of the robot. Criticisms of the mobile interface were not related to the perspective provided by the camera or any video delay, but instead were focused on usability issues associated with the visual and interactive aspects of the interface. Suggestions included alternative ways that the augmented grid and blocks ought to be displayed in the environment (e.g., changing the appearance and lowering the transparency of the visuals, providing augmented reality containers, etc.), as well as alternative ways in which users ought to interact with these augmented elements (e.g., allowing virtual elements to “snap” to locations in the environment, allowing users to perform pinch gestures on the screen to zoom in and out on the view, etc.). These suggestions will be considered in the development and testing of future prototypes before the next user study with the interfaces is conducted.

6. CONCLUSION

This paper proposed a novel mobile mixed-reality interface approach to interact with robots in shared spaces using the mobile devices that users already own and are familiar with. The proposed approach appropriately leverages the capabilities of mobile devices to render mixed-reality environments from views provided by their rear-facing cameras. This allows for the development of interfaces with more intuitive visual perspectives, reduced latency, improved responsiveness, and expanded perceptual range in comparison to implementing conventional egocentric and exocentric interfaces on mobile devices.

To evaluate aspects of the performance and user experience associated with these interfaces, a user study was conducted in which participants interacted with a humanoid robot to perform object manipulation on a tabletop. The study revealed the following benefits of the proposed approach. First, by integrating touchscreen interaction with augmented graphics, the interfaces allowed users to naturally command the robot to manipulate physical objects. Second, the interactions enabled by the approach were as successful and yielded comparable or better user experiences as the conventional interface approaches. Third, the mobility associated with the approach provided the sensation of directly interacting with objects in the robot’s workspace through a visually engaging mixed-reality window. Fourth, by allowing users to move about while pointing their device from different perspectives, the approach resolved usability issues commonly exhibited by conventional interfaces, such as the keyhole effect and occlusion. The practical significance of the proposed approach is evidenced from the following. First, the approach provides intuitive and natural interactive mixed-reality interaction with robots without the need for large operator stations or sophisticated vision systems mounted in the environment or on the robot. Second, the approach can be generalized to disparate robots and workspaces, provided that an accurate kinematic model of the robot and an accurate geometric model of the visual features in the environment are known. Third, the approach enables robotic platforms to be operated outside the traditional laboratory environment, which may preclude the installation of vision sensors. Fourth, the approach can significantly reduce the cost and complexity of implementing human–robot interaction

systems. Finally, the proposed approach can be adapted to a broad range of applications, such as assembly, machining, packaging, and handling tasks commonly encountered in manufacturing and construction.

The following are the specific conditions under which the proposed approach works well. First, the approach addresses applications in which the user and robot occupy a shared space. Second, the approach requires $n \geq 4$ visual markers in the workspace of the robot that are coplanar and detectable by the mobile device’s camera from various perspectives. Third, the approach is deemed feasible with mobile devices with comparable or superior capability to an iPad Pro *vis-à-vis* its sensing and processing capabilities. Despite its benefits, the proposed approach is limited as revealed by the aforementioned specific conditions for it to function well. For example, instead of being limited to shared spaces, conventional approaches can be implemented in remote operation scenarios also. However, the list of applications in which people interact with robots in shared spaces is steadily growing (e.g., in education, medicine, recreation, and domestic applications). Next, as the robot and objects in the workspace obstruct the view of one or more visual markers from the mobile device camera, additional visual markers must be introduced into the workspace (e.g., twelve visual markers were affixed to the tabletop in this study). Thus, to relieve constraints on the robot’s workspace, future work will explore alternative methods of plane estimation by exploiting (i) the visual features on the robot rather than the workspace, (ii) the concept of vanishing points, and (iii) the inertial measurement unit of mobile devices. Finally, the proposed approach has only been tested with the robot performing relatively simple pick-and-place tasks on rigid blocks that have simple shapes and are affixed with visual markers. However, a variety of vision techniques exist for detecting and estimating the pose of objects without the use of visual markers. In an effort to generalize the proposed approach to practical applications of robotic manipulation, future efforts will integrate these techniques to the treatment of irregular-shaped or deformable objects. Furthermore, complex tasks with obstacles will be addressed by the fusion of sensor data captured by the mobile device and by the robot to construct a map of the robot’s configuration space to be used for path planning.

AUTHOR CONTRIBUTIONS

JF is the main contributor, both in the development and research with the mobile mixed-reality interfaces as well as with the humanoid robot. MM was involved in the design and construction of the robot and in the programming of the robotic kinematics and path planning algorithms. VK supervised all aspects of research, including checking the accuracy of results and helping with the writing of the paper.

FUNDING

This work was supported in part by the National Science Foundation awards RET Site EEC-1132482 and EEC-1542286, DRK-12 DRL: 1417769, ITEST DRL: 1614085, and GK-12 Fellows DGE: 0741714, and NY Space Grant Consortium grant 76156-10488.

REFERENCES

- Birkenkamp, P., Leidner, D., and Borst, C. (2014). “A knowledge-driven shared autonomy human-robot interface for tablet computers,” in *Proc. IEEE-RAS Int. Conf. Humanoid Robots* (Madrid: IEEE), 152–159.
- Campbell, I. (2007). Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations. *Stat. Med.* 26, 3661–3675. doi:10.1002/sim.2832
- Chen, J. Y., Haas, E. C., and Barnes, M. J. (2007). Human performance issues and user interface design for teleoperated robots. *IEEE Trans. Syst. Man. Cybern. C Appl. Rev.* 37, 1231–1245. doi:10.1109/TSMCC.2007.905819
- Drury, J. L., Scholtz, J., and Yanco, H. A. (2003). “Awareness in human-robot interactions,” in *IEEE Int. Conf. Systems, Man and Cybernetics*, Vol. 1 (Washington, DC: IEEE), 912–918.
- Frank, J. A., Moorhead, M., and Kapila, V. (2016). “Realizing mixed-reality environments with tablets for intuitive human-robot collaboration for object manipulation tasks,” in *Proc. Int. Symp. Robot-Human Interactive Communication* (New York, NY: IEEE), 302–307.
- Fritzsche, M., Elkmann, N., and Schulenburg, E. (2011). “Tactile sensing: a key technology for safe physical human robot interaction,” in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction* (Lausanne: IEEE), 139–140.
- Goodrich, M., and Olsen, D. Jr. (2003). “Seven principles of efficient human robot interaction,” in *IEEE Int. Conf. Systems, Man, and Cybernetics*, Vol. 4 (Washington, DC: IEEE), 3942–3948.
- Green, S. A., Chase, J. G., Chen, X., and Billingham, M. (2008a). “Evaluating the augmented reality human-robot collaboration system,” in *Int. Conf. Mechatronics and Machine Vision in Practice* (Auckland: IEEE), 575–580.
- Green, S., Billingham, M., Chen, X., and Chase, J. (2008b). Human-robot collaboration: a literature review and augmented reality approach in design. *Int. J. Adv. Rob. Syst.* 5, 1–18. doi:10.5772/5664
- Hart, S. G. (2006). “Nasa-task load index (NASA-TLX); 20 years later,” in *Proc. Human Factors and Ergonomics Society Annual Meeting*, Vol. 50 (San Francisco, CA: Sage Publications), 904–908.
- Hashimoto, S., Ishida, A., Inami, M., and Igarashi, T. (2011). “Touchme: an augmented reality based remote robot manipulation,” in *Proc. Int. Conf. Artificial Reality and Telexistence* (Osaka: CiteSeerX).
- Hughes, S. B., and Lewis, M. (2005). Task-driven camera operations for robotic exploration. *IEEE Trans. Syst. Man. Cybern. A Syst. Hum.* 35, 513–522. doi:10.1109/TSMCA.2005.850602
- Kao, M.-C., and Li, T.-H. (2010). “Design and implementation of interaction system between humanoid robot and human hand gesture,” in *Proc. of SICE Annual Conf* (Taipei: IEEE), 1616–1621.
- Lee, W., Ryu, H., Yang, G., Kim, H., Park, Y., and Bang, S. (2007). Design guidelines for map-based human–robot interfaces: a collocated workspace perspective. *Int. J. Indus. Ergon.* 37, 589–604. doi:10.1016/j.ergon.2007.03.004
- Lenz, C., Grimm, M., Röder, T., and Knoll, A. (2012). “Fusing multiple kinects to survey shared human-robot-workspaces,” in *Technische Universität München, Munich, Germany, Tech. Rep. TUM-I1214*.
- Lew, J. Y., Jou, Y.-T., and Pasic, H. (2000). “Interactive control of human/robot sharing same workspace,” in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Vol. 1 (Takamatsu: IEEE), 535–540.
- Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *Int. J. Hum. Comput. Interact.* 14, 463–488. doi:10.1080/10447318.2002.9669130
- Marin, R., Sanz, P. J., Nebot, P., and Wirz, R. (2005). A multimodal interface to control a robot arm via the web: a case study on remote programming. *IEEE Trans. Indus. Electron.* 52, 1506–1520. doi:10.1109/TIE.2005.858733
- McGhan, C. L., Nasir, A., and Atkins, E. M. (2015). Human intent prediction using Markov decision processes. *J. Aerosp. Inf. Syst.* 12, 393–397. doi:10.2514/1.I010090
- Menchaca-Brandan, M. A., Liu, A. M., Oman, C. M., and Natapoff, A. (2007). “Influence of perspective-taking and mental rotation abilities in space teleoperation,” in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction* (Washington, DC: IEEE), 271–278.
- Micire, M., Desai, M., Courtemanche, A., Tsui, K., and Yanco, H. (2009). “Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces,” in *Proc. ACM Int. Conf. Interactive Tabletops and Surfaces* (Banff: ACM), 41–48.
- Milgram, P., Zhai, S., and Drascic, D. (1993). “Applications of augmented reality for human-robot communication,” in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Vol. 3 (Yokohama: IEEE), 1467–1472.
- Morato, C., Kaipa, K. N., Zhao, B., and Gupta, S. K. (2014). Toward safe human robot collaboration by using multiple kinects based real-time human tracking. *J. Comput. Inf. Sci. Eng.* 14, 011006. doi:10.1115/1.4025810
- Muszynski, S., Stücker, J., and Behnke, S. (2012). “Adjustable autonomy for mobile teleoperation of personal service robots,” in *Proc. IEEE Int. Symp. Robot and Human Interactive Communication* (Paris: IEEE), 933–940.
- Nielsen, C. W., Goodrich, M. A., and Ricks, R. W. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE Trans. Robot.* 23, 927–941. doi:10.1109/TRO.2007.907479
- Oberkamp, D., DeMenthon, D. F., and Davis, L. S. (1996). Iterative pose estimation using coplanar feature points. *Comput. Vision Image Understand.* 63, 495–511. doi:10.1006/cviu.1996.0037
- Sakamoto, D., Sugiura, Y., Inami, M., and Igarashi, T. (2016). Graphical instruction for home robots. *Computer* 49, 20–25. doi:10.1109/MC.2016.195
- Sauro, J., and Lewis, J. R. (2012). *Quantifying the User Experience: Practical Statistics for User Research*. San Francisco, CA: Morgan Kaufmann.
- Shah, J., and Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Hum. Factors* 52, 234–245. doi:10.1177/0018720809350882
- Spong, M., Hutchinson, S., and Vidyasagar, M. (2006). *Robot Modeling and Control*. New York, NY: John Wiley and Sons.
- Su, Y.-H., Hsiao, C.-C., and Young, K.-Y. (2015). “Manipulation system design for industrial robot manipulators based on tablet PC,” in *Intelligent Robotics and Applications* (Portsmouth, UK: Springer), 27–36.
- Waldherr, S., Romero, R., and Thrun, S. (2000). A gesture based interface for human-robot interaction. *Auton. Robots* 9, 151–173. doi:10.1023/A:1008918401478
- Woods, D. D., Tittle, J., Feil, M., and Roesler, A. (2004). Envisioning human-robot coordination in future operations. *IEEE Trans. Syst. Man. Cybern. C Appl. Rev.* 34, 210–218. doi:10.1109/TSMCC.2004.826272
- Ziaei, Z., Hahto, A., Mattila, J., Siuko, M., and Semeraro, L. (2011). Real-time markerless augmented reality for remote handling system in bad viewing conditions. *Fusion Eng. Design* 86, 2033–2038. doi:10.1016/j.fusengdes.2010.12.082

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Frank, Moorhead and Kapila. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.