Design and Implementation of Intuitive Human-robot Teleoperation Interfaces

Lei Wu
University of South Florida

Follow this and additional works at: https://scholarcommons.usf.edu/etd
Part of the Robotics Commons

Scholar Commons Citation
Wu, Lei, "Design and Implementation of Intuitive Human-robot Teleoperation Interfaces" (2020). Graduate Theses and Dissertations. https://scholarcommons.usf.edu/etd/8604

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Design and Implementation of Intuitive Human-robot Teleoperation Interfaces

by

Lei Wu

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
Department of Mechanical Engineering
College of Engineering
University of South Florida

Co-Major Professor: Redwan Alqasemi, Ph.D.
Co-Major Professor: Rajiv Dubey, Ph.D.
Sudeep Sarkar, Ph.D.
Stephanie Carey, Ph.D.
Kandethody Ramachandran, Ph.D.

Date of Approval:
November 24, 2020

Keywords: Robotic Arm, Interactions, DoF, Tele-robotics, Coordinate

Copyright © 2020, Lei Wu
Dedication

To my wife, Jia, for her precious support in my research life and for the sacrifices she made and the encouragement she gave. To my child, Rebekah, who gave me great joy every day when I return from my work.

To my advisor, Dr. Alqasemi, who gave me countless support from day one when I first arrived here. Giving me precious research ideas and leads me to find the innovations of my research and helping me revise papers days and nights.

To my advisor, Dr. Dubey, who gave me a solid foundation in robotics and kinematics and the guidance in my Ph.D. research.

To God, who brings me up from every difficult situation.
# Table of Contents

List of Tables iii

List of Figures iv

Abstract viii

Chapter 1: Introduction 1
  1.1 Motivation 1
  1.2 Dissertation Objectives 2
  1.3 Dissertation Outline 3

Chapter 2: Background 4
  2.1 Human-Robot Interface Input Solutions 4
    2.1.1 Joysticks 4
    2.1.2 RGBD Cameras 6
    2.1.3 Inertial Measurement Units (IMUs) 7
    2.1.4 Smartphones or Touchscreens 9
    2.1.5 Haptic Devices 10
    2.1.6 EMG Sensors 11
    2.1.7 Leap Motion Sensor 13
    2.1.8 Physical Human-Robot Interface (pHRI) 15
    2.1.9 Brain Computer Interface (BCI) 16
    2.1.10 Motion Tracking System 17
    2.1.11 Non Commercially Available Devices 18
  2.2 Commercial Robotic Arms 19
  2.3 Robotic Arm Control Reference Frames 22

Chapter 3: Development of Human-robot Interface Using Smartphones 24
  3.1 Introduction 24
  3.2 Control Interfaces and Theories 24
    3.2.1 One Button Interface 29
    3.2.2 Three Buttons Interface 31
    3.2.3 Tilt Interface 34
  3.3 Experimental Setup 35
  3.4 Evaluation and Results 39
<table>
<thead>
<tr>
<th>Chapter 4: Development of Intuitive Reference Frame and Hybrid Kinematics</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>45</td>
</tr>
<tr>
<td>4.2 System Modeling</td>
<td>46</td>
</tr>
<tr>
<td>4.2.1 Ground Reference Frame Control System</td>
<td>47</td>
</tr>
<tr>
<td>4.2.2 End-effector Reference Frame Control System</td>
<td>49</td>
</tr>
<tr>
<td>4.2.3 Intuitive Reference Frame Control System</td>
<td>50</td>
</tr>
<tr>
<td>4.3 Experiments</td>
<td>55</td>
</tr>
<tr>
<td>4.3.1 Introduction to the ADL Task</td>
<td>55</td>
</tr>
<tr>
<td>4.3.2 Human Subjects Recruitment and Testing</td>
<td>56</td>
</tr>
<tr>
<td>4.4 Results</td>
<td>57</td>
</tr>
<tr>
<td>4.5 Analysis</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5: Performance Test of the Hybrid System with Different Input Devices.</th>
<th>62</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>62</td>
</tr>
<tr>
<td>5.2 Experiment Setup</td>
<td>63</td>
</tr>
<tr>
<td>5.3 Experimental Results</td>
<td>66</td>
</tr>
<tr>
<td>5.4 Results Analysis</td>
<td>77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 6: Conclusions and Future Work</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Conclusions</td>
<td>80</td>
</tr>
<tr>
<td>6.2 Future Work</td>
<td>82</td>
</tr>
</tbody>
</table>

References

Appendix A: Copyright Permissions
**List of Tables**

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Comparison of the state of the art human-robot interfaces.</td>
<td>25</td>
</tr>
<tr>
<td>4.1</td>
<td>Technical requirements of building an intuitive robotic arm tele-</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>operation system summarized from users’ feedback.</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Comparison between the four user input devices and the three</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>reference frames.</td>
<td></td>
</tr>
</tbody>
</table>
List of Figures

| Figure 2.1 | Jaco robotic arm and joysticks. | 5 |
| Figure 2.2 | Xbox controller and the mapped control. | 5 |
| Figure 2.3 | Kinect sensor: 1. Depth sensors, 2. RGB camera, 3. Motorized base. | 6 |
| Figure 2.4 | Wearable IMU modules: MYO armband. | 7 |
| Figure 2.5 | Object position recognition using Kinect. | 7 |
| Figure 2.6 | IMUs mounted on the shoulders. | 8 |
| Figure 2.7 | Wearable hat interface. | 9 |
| Figure 2.8 | Smartphone based teleoperation interface. | 10 |
| Figure 2.9 | Parallel haptic force feedback device. | 11 |
| Figure 2.10 | sEMG sensor pads. | 12 |
| Figure 2.11 | sEMG arm band. | 12 |
| Figure 2.12 | Grasping based on sEMG signal. | 13 |
| Figure 2.13 | Leap motion with RGBD camera. | 14 |
| Figure 2.14 | Control two robotic arms using a Leap motion device. | 14 |
| Figure 2.15 | Using a robotic arm as an input device. | 15 |
| Figure 2.16 | Exoskeleton physical human-robot interaction. | 16 |
| Figure 2.17 | Brain-computer interface controlled manipulation. | 16 |
| Figure 2.18 | Brain-computer interface with vision assist for autonomous grasping. | 17 |
| Figure 2.19 | Small motion tracking system. | 18 |
| Figure 2.20 | Artificial skin interface. | 18 |
Figure 2.21 Robotic arms performing welding tasks.

Figure 2.22 Robotic arms performing painting tasks.

Figure 2.23 Robotic arms performing laboratory tasks.

Figure 2.24 The robot Baxter from Rethink Robotics.

Figure 2.25 UR5 robotic arm from Universal Robotics.

Figure 2.26 Kuka LBR iiwa robotic arm.

Figure 2.27 End-effector frame.

Figure 2.28 End-effector frame used in visual-servoing scenarios.

Figure 3.1 Smartphone coordinate system.

Figure 3.2 One Button: slide translation & tilt rotation (combine).

Figure 3.3 Three Buttons: slide translation & tilt rotation (separate).

Figure 3.4 Tilt: tilt translation & tilt rotation (separate).

Figure 3.5 First ADL task: Water pouring.

Figure 3.6 Second ADL task: Pick and place plate & bowl from the dish washer to the tabletop.

Figure 3.7 Third ADL task: Seasoning the food with Pepper and Salt.

Figure 3.8 Time required to complete each task (the lower the better).

Figure 3.9 Robot end-effector distance traveled (the lower the better).

Figure 3.10 How intuitive the interface was in each task (The higher the better).

Figure 4.1 The user input device is a spacemouse and the illustration of 6 degrees of freedom controls include FB: forward or backward, UD: up or down, LR: left or right and roll, pitch, yaw for rotation.

Figure 4.2 Illustration of the Ground Reference Frame and the mapped 6 DoF user control vectors.

Figure 4.3 Illustration of how the End-effector Reference Frame is related to the current pose of end-effector and the mapped 6 DoF user control vectors.
Figure 4.4  Illustration of the Intuitive Reference Frame and the mapped 6 DoF user control vectors.

Figure 4.5  The ADL task: food seasoning.

Figure 4.6  Time required for completing the ADL task using the three different reference frames.

Figure 4.7  Intuitiveness ratings of the three control reference frames.

Figure 4.8  A case study of how the Ground(G), End-effector(E) and Intuitive(I) reference frames behave when manipulating a robotic arm in a typical ADL task scenario.

Figure 5.1  One Button: slide translation & tilt rotation(combine).

Figure 5.2  The UI design of Xbox interface.

Figure 5.3  Omni user input device.

Figure 5.4  The ADL task: food seasoning.

Figure 5.5  Time required for completing the ADL task when using the smartphone with the three reference frames (the lower the better).

Figure 5.6  Robot end-effector distance traveled when using the smartphone with three reference frames (the lower the better).

Figure 5.7  Time required for completing the ADL task when using the Xbox with the three reference frames (the lower the better).

Figure 5.8  Robot end-effector distance traveled when using the Xbox with the three reference frames (the lower the better).

Figure 5.9  Time required for completing the ADL task when using the Omni with the three reference frames (the lower the better).

Figure 5.10  Robot end-effector distance traveled when using the Omni with the three reference frames (the lower the better).

Figure 5.11  Time required for completing the ADL task when using the Spacemouse with the three reference frames (the lower the better).

Figure 5.12  Robot end-effector distance traveled when using the Spacemouse with the three reference frames (the lower the better).

Figure 5.13  Time required for completing the ADL task, including data from all devices used with the three reference frames (the lower the better).
Figure 5.14  Robot end-effector distance traveled, including data from all devices used with the three reference frames (the lower the better).

Figure 5.15  Performances comparison between devices regardless of control method - completion time.

Figure 5.16  Performances comparison between devices regardless of control method - distance traveled.
Abstract

We have designed and implemented a novel human-robot teleoperation interface based on an intuitive reference frame and hybrid inverse kinematics to perform activities of daily living (ADL) using multiple input devices.

Persons with disabilities often rely on caregivers or family members to assist in their daily living activities. Providing robotic assistants with easy and intuitive user interfaces to assist with ADL can improve their quality of life and lift some of the burdens on caregivers and family members. Current human-robot interface solutions, such as joysticks, Kinect based gesture recognition, and touchscreen-based solutions, including smartphones, are still far from being able to operate in an intuitive way when used for complex activities of daily living.

In this dissertation, we review the current popular human-robot interfaces and discuss their advantages and disadvantages. When developing our new interface system, we try to maximize as many advantages as possible while minimizing the disadvantages. In this era of smartphones that are packed with sensors, such as accelerometers, gyroscopes, and a precise touch screen, teleoperation control can be interfaced with smartphones to capture the user’s intended operation of the robot assistant. We developed three novel human-robot smartphone-based interfaces to operate a robotic arm for assisting persons with disabilities in their ADL tasks. Useful smartphone data, including 3-dimensional orientation and 2-dimensional touchscreen positions, are used as control variables to the robot motion in Cartesian teleoperation. The developed interfaces provide intuitiveness, low cost, and environmental adaptability.
Not only the interface devices affect the intuitiveness of the robotic arm teleoperation, but also the inverse kinematics algorithm of the robotic arm is crucial to the whole system intuitiveness as well. The two commonly used reference frames are the Ground and the End-effector reference frames, which do not always provide intuitive control from the perspective of human operators. We conducted preliminary testing in Ground and End-effector reference frames separately to maneuver objects in 3D space. Based on their feedback, we found how users wanted the robot to move and developed a new Intuitive Reference Frame with a novel hybrid inverse kinematics solution (the Hybrid system). This system provides a more natural and easier to use human-robot interface. Two conventional control reference frames (the Ground and the End-effector) are compared with our novel Intuitive control reference frame. An activity of daily living (ADL) task was used to test the performances of the three control reference frames. A 6-D spacemouse, Xbox controller, Omni and a smartphone were used as input devices for the human-robot interfaces to control a Baxter robotic arm and perform the same ADL task using the three different control reference frames. We tested the three control reference frame systems with human subjects and collected qualitative and quantitative data. The results show that overall our Intuitive robotic arm control reference frame with the hybrid inverse kinematics greatly reduced the time and effort needed to manipulate the robotic arm. The results also show that our Hybrid system improved the performances and intuitiveness for all tested input devices.
Chapter 1: Introduction

1.1 Motivation

There are more than 28 million people in the US who have some form of physical disability and require assistance[1]. Providing robotic assistants with easy and intuitive user interfaces to assist with activities of daily living (ADL) can improve their quality of life and lift some of the burden on caregivers and family members.

An intuitive user interface for teleoperation of a robotic assistant is key to its effectiveness and utilization. One of the most widely used robotic arm teleoperation device is joysticks, since it provides low-cost, high precision, and very reliable characteristics[2]. The drawback of the joystick-based solution is it can only control 2 to 3 degrees of freedom at a time, so switching between modes is required in order to control all 6 degrees of freedom in the Cartesian space, which will cause extra cognitive load for the users when operating the robot. In recent years, research groups developed many new interfaces by utilizing up-to-date technologies, such as Kinect cameras, touchscreens, and haptic devices. However, new interfaces introduce many new shortcomings as well, such as high-cost, long installation time, training needed, and low reliability.

Robot kinematics is also an important part of the robot that can make it easier for the user to utilize if optimized effectively[3]. We identified potential improvements in the robotic arm control system that can make the control system work more efficiently and intuitively for the user. The two commonly used reference frames are the Ground and the End-effector reference frames, which do not always provide intuitive control from the perspective of human users.
When using the Ground Reference Frame or the End-effector Reference Frame to control a robotic arm, no matter what input device is adopted, people still find it difficult to maneuver the robotic arm to complete complex tasks that need translational motion and rotational motion throughout the process. Due to the poor intuitiveness of the two reference frames, human operators need to make a lot of effort to think from the robot’s perspective, which causes frustration and abandonment[3].

1.2 Dissertation Objectives

We have conducted user interviews with human subjects using the Baxter robot to perform ADL tasks and collected preliminary data to focus our work on improving the interface and usability of the robotic arm. On the basis of these interviews, we extracted objectives for improving their experience. The objectives of this dissertation are as follows:

1. Design of a smartphone-based human-robot input interface that achieves the following characteristics:
   - Easy interaction with the least amount of buttons.
   - Minimum or no training is needed.
   - Precise and responsive enough so that subjects can complete complex ADL tasks in an unstructured environment with minimum or no frustration.
   - No calibration is needed.
   - Low cost.
   - Lightweight and wireless.

2. Design and implementation of an intuitive human-robot teleoperation mapping method using a new reference frame and hybrid inverse kinematics, that fulfills all user’s preferences as follows:
• Maneuvering the robotic arm forward/backward translation and roll rotation in the End-effector Reference Frame.

• Maneuvering the robotic arm up/down translation and yaw rotation in the Ground Reference Frame.

• Maneuvering left/right translational motion along a special axis, which is perpendicular to the end-effector and parallel to the ground surface. This axis is neither described in the Ground Reference Frame nor in the End-effector Reference Frame.

• Maneuvering pitch rotational motion around that special axis.

3. Testing of the intuitive human-robot teleoperation user interface with multiple input devices.

1.3 Dissertation Outline

Chapter 2 will give a background on current existing popular human-robot interface devices and methods. Chapter 3 will focus on the requirements needed to build an easy to use teleoperation interface and how to use a smartphone only to achieve the goals. In chapter 4, we present a novel Intuitive Reference Frame with hybrid inverse kinematics, the corresponding experimental work and test results. Chapter 5 verifies the adaptability and performance of our developed theory applied with our developed Smartphone user interface as well as other popular user input devices. Chapter 6 gives conclusions of this research and the discussion of future work.
Chapter 2: Background

In human-robot teleoperation interfaces, there are three relevant research focuses. First, the user input devices hardware and the UI design, including the customized sensors and ergonomics. The goal is to make easy-to-use, precise, and responsive input interface solutions with high reliability and environmental adaptability. Second, the robot hardware design with a goal to increase the safety, accuracy, and affordability so that it can be used in more scenarios such as chemical industry, surgical, biomedical, and rehabilitation. Third, the robot kinematics, which affects how robotic arms behave, and the goal is to increase the environmental adaptiveness and intuitiveness of the whole system. In this chapter, we will introduce the state-of-the-art user input solutions, popular robotic arms, and commonly used reference frames in robotic kinematics.

2.1 Human-Robot Interface Input Solutions

2.1.1 Joysticks

V. Maheu used the Jaco arm controlled by the joystick to do services for persons with disabilities [4] as shown in Figure 2.1. The drawback of joystick based solutions is that the joystick can control 2 to 3 degrees of freedom at a time, so in order to control all 6 degrees of freedom, we need to switch between translational control or rotational control back and forth frequently.
Herlant et al. used a traditional joystick and two mode-switching buttons. They achieved many complex ADL tasks, such as phone dialing, water pouring, and unscrewing a jar of coffee [2].

Campeau-Lecours et al. utilized a multiple joysticks device, which is an Xbox controller, to control the robotic arm, as shown in Figure 2.2. Since it has more than one joystick, users do not need to switch between modes to achieve 6 DoF control [5].
2.1.2 RGBD Cameras

Figure 2.3 shows a typical RGBD camera, Microsoft Kinect Xbox 360, which is first released in 2010. This device was designed to track human body motions. From the SDK provided, we can have real-time gesture recognition, speech recognition and body skeletal detection for up to four persons at a time.

Reddivari et al. presented a control interface of the Baxter robotic arm using Kinect body motion tracking capability[6]. Their algorithm can find the position of the human body joints' position. Through mapping, the Baxter robot can mimic human body motion, but since the rotation of human body joints can not be detected, it only controls the position of the robotic arm. In order to include the rotational control, Li et al. utilized multiple IMU sensors(Figure 2.4) mounted on the elbow of the user to acquire rotation information, plus the Kinect camera for position information, they achieved all 6 DoF Cartesian space controls [7].
Instead of using the Kinect camera to recognize human body motion, Bousquest-Jette et al. used it to detect the objects as shown in Figure 2.5, then the user chooses which object he/she wants to grasp, the robot will pick the object and bring it to the user autonomously [8]. This method reduced the load needed to maneuver the robotic arm, but the system needs a structured environment and predefined objects.

2.1.3 Inertial Measurement Units (IMUs)

Jain et al. used IMU devices mounted on the shoulder of a person with disabilities to control the robotic arm, as shown in Figure 2.6, but the user was getting tired too quickly, and was only able to control a very limited number of degrees of freedom [9].
Baldi et al. utilized an IMU module and an EMG module to build a wearable body machine interface, as shown in Figure 2.7, where all sensors were mounted on a hat, including the battery. They were able to complete a relatively complex water pouring task, however, the mode switching is still needed in this interface [10].

Figure 2.6. IMUs mounted on the shoulders.
2.1.4 Smartphones or Touchscreens

Parga et al. designed a smartphone-based interface to control a robotic arm as shown in Figure 2.8 [11]. However, it needed many buttons to control all 6 DoF, and no task was performed.
Figure 2.8. Smartphone based teleoperation interface.

Instead of using buttons, Rodriguez et al. used gestures to control a mobile robot moving on the ground [12]. Mandlekar et al. used the IMU sensors inside of the smartphone to design an intuitive interface for robotic arm control [13], however, due to the integral calculations needed for the inertial sensors, they sacrificed the accuracy.

2.1.5 Haptic Devices

Vu and Na developed a 6-DoF parallel haptic device [14], as shown in figure 2.9. The advantage of this method is that the device is motorized for all joints so that position control can be achieved, and it provided force feedback, but due to the small working space, it was difficult to achieve 1 to 1 scaled control.
2.1.6 EMG Sensors

Fall et al. designed a wireless surface electromyography (sEMG) based body-machine interface for persons with spinal cord injuries that affect the arms, hands, and fingers control, thus making use of joysticks or keypads impossible [15]. The sEMG sensors are attached on the user’s chest and shoulders in order to sense the muscle activities then translated to robotic arm motion commands, as shown in Figure 2.10.
Figure 2.10. sEMG sensor pads.

Cote-Allard et al. utilized armband sEMG sensors to read human wrist muscle movement. This solution offered a tidy design and wireless control as shown in Figure 2.11, but it was not able to control all 6 degrees of freedom at the same time as well.

Figure 2.11. sEMG arm band.

Meattini et al. used sEMG pads mounted on the lower arm. It read the gesture of the human hand very well, and used the information to perform grasping for irregularly shaped objects, as shown in Figure 2.12. However, the position and orientation of the end-effector cannot be controlled [16].
2.1.7 Leap Motion Sensor

Lin et al. used a Leap motion device, an RGBD camera and multiple computer monitors to build an augmented reality interface as shown in figure 2.13. It can recognize hand gestures to control the robotic arm [17].
Du and Zhang presented a robotic dual-arm control using only one Leap motion device, as shown in figure 2.14. The leap motion device can detect two hands on top of it. The X, Y, Z position and Yaw rotation of hands can be read [18], but the accuracy of the hand pose required the user’s hand to be constantly suspended at a specific distance from the device, which caused fatigue and abandonment.
2.1.8 Physical Human-Robot Interface (pHRI)

Bitz et al. used a robotic arm as the control input device to read the position and orientation of the user’s hand, as shown in figure 2.15. It can read all 6 DoF at the same time, which makes this solution very intuitive and easy to use [19]. However, the robotic arm is very difficult to transport, and it needs an external power supply.

Figure 2.15. Using a robotic arm as an input device.

Li et al. developed a wearable exoskeleton as an input device to control a robotic arm, as shown in figure 2.16, the biggest advantage of this method is it can realize direct position control, which responds quicker than velocity control [20].
2.1.9 Brain Computer Interface (BCI)

Muelling et al. utilized the BCI device to achieve many complex ADL tasks (Figure 2.17), such as door opening, pouring soda, and picking up wood blocks [21]. However, all objects are pre-defined and have to be in a constructed environment.
Mounir et al. presented a BCI-Controlled hands-free wheelchair navigation system. The wheelchair’s motion can be controlled without joysticks [22]. Pathirage et al. achieved autonomous grasping using P300 BCI system and an RGB camera [23]. Figure 2.18 shows the user controlling the robotic arm to grasp a cup without moving any part of his body.

![Image of user controlling robotic arm](image)

Figure 2.18. Brain-computer interface with vision assist for autonomous grasping.

2.1.10 Motion Tracking System

Li et al. used the motion tracking system to track the markers on the user’s hand to read the pose of the hand, as shown in Figure 2.19, then map the human arm motion to the robotic arm motion [24]. This local position tracking system was relatively small and low cost compared to traditional indoor position tracking systems, but there are many drawbacks that made it impractical. First, it required the user to wear five markers on a hand before using this system; second, it needed a very long time to re-calibrate the system if someone accidentally hit the stand of the cameras.
2.1.11 Non Commercially Available Devices

Pugach et al. designed an artificial skin material as a user intention recognition device to control the robotic arm, as shown in Figure 2.20, which is quite interesting and new. However, the control was limited to only joint level operation [25].

There are different research groups developed many customized user input devices and control strategies to perform teleoperated robotic surgery [26, 27, 28].
2.2 Commercial Robotic Arms

Robotic arms are most widely used in the car manufacturing industries due to the repetitive, hazardous and heavy payload tasks needed\[29\]. Figure 2.21 shows automobile welding task performed by multiple robotic arms\[30\]. It frees humans from high voltage, high heat environments, and potential eyesight damages.

![Figure 2.21. Robotic arms performing welding tasks.](image)

Robotic arms are also very useful in painting tasks. Figure 2.22 shows multiple robotic arms work together to paint cars efficiently and safely, so humans do not need to work in a toxic environment.

![Figure 2.22. Robotic arms performing painting tasks.](image)

Yaskawa designed a precise dual-arm robot, Motoman, for assembly and handling tasks\[31\]. Figure 2.23 shows the robotic arms working in a biochemistry laboratory and preparing
the solution for the experiment. It saves time for technicians from these tedious and repetitive laboratory preparation tasks [32].

Nowadays, robotic arm manufacturing companies start releasing many light-weight, transportable, and low-cost robotic arms for small businesses. Many researchers also use these robotic arms for academic research purposes. One of the most widely used robotic arms in research is Baxter from Rethink Robotics, as shown in figure 2.24. It is a dual-arm robot, both arms have 7 DoF configuration, so researchers can apply redundant control on this robot [33]. The built-in dual-arm self-collision avoidance protection makes it best fit for dual-arm collaboration tasks compared to using two separate single-arm setup [33, 34]. Many researchers used the Baxter robotic arm in their user interfaces due to the flexibility and efficient programming capability [35, 36]. Compared to another widely used research robot, ABB Yumi dual-arm, Baxter has a larger working space but less accuracy and motion smoothness.[37].

Figure 2.23. Robotic arms performing laboratory tasks.
UR5, from Universal Robotics, is also a popular robotic arm used in research areas, as shown in figure 2.25. It has six precise joints and a small body design. Fang et al. utilized it to design a teleoperation system using a wearable EMG device [38]. Omarali et al. used the UR5 robotic arm to mimic the human upper limb motion [39].

Kuka LBR iiwa is a widely used single arm 7 DoF robot [40, 41, 42]. It provides high torque, precise, and compact design. Due to the high power density characteristics, many
research groups use it to build humanoid robots, which can stand by itself [43, 44, 45], as shown in figure 2.26.

![Kuka LBR iiwa robotic arm](image)

Figure 2.26. Kuka LBR iiwa robotic arm.

### 2.3 Robotic Arm Control Reference Frames

There are two commonly used reference frames in robotic arm kinematics, the Ground Reference Frame and the End-effector Reference Frame [46], which do not always provide intuitive control from the perspective of human users [3]. When using the Ground Reference Frame, the motion of the end-effector (robot hand) is always with respect to the ground or robotic arm base. The translation and rotation of the frame will not be affected by the current pose of the robot end-effector [47].

R.Featherstone presented the mapping equations between joint angles and the end-effector frame positions [48]. When using the End-effector Reference Frame, the motion of the end-effector is always with respect to the current pose of the end-effector, as shown in figure 2.27.
Different from the Ground Reference Frame, the control perspective changes when the pose of the end-effector changes. This technique is crucial when the robotic arm is manipulated based on streamed live video from the camera mounted on the end-effector [49], as shown in figure 2.28, since the only view that operators can see is from the camera mounted on the end-effector. In this case, control in the end-effector frame becomes much easier than in the ground frame.
Chapter 3: Development of Human-robot Interface Using Smartphones

3.1 Introduction

The reviewed human-robot interfaces are summarized in table 3.1. From the reviews, we can see that there are different shortcomings in each method that make it more difficult or impossible to complete complex ADL tasks.

In this chapter\(^1\), our focus is to build novel interfaces based on smartphones, which are readily available devices, eliminating as many shortcomings as possible in current existing human-robot interfaces to provide a simple and easy to use interface for users to control the robotic arm and perform their daily living activities. Our novel interfaces should achieve the following objectives: 1. intuitive interaction with the least amount of buttons such that minimum or no training is needed; 2. precise and responsive enough so that healthy people and persons with disabilities can complete complex ADL tasks in an unstructured environment with minimum or no frustration; 3. no calibration is needed; 4. low cost; 5. lightweight and wireless. These goals will also distinguish our work from existing methods.

3.2 Control Interfaces and Theories

In this work, three different control interfaces have been developed for controlling a robotic arm using a smartphone. One uses a single touch anywhere on the screen (One Button), and the other two use three buttons on the screen (Three Buttons and Tilt).

\(^{1}\)This chapter was published in 2020 IEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 5835-5841, Oct. 2020, doi: 10.1109/LRA.2020.3010453. Permission is included in Appendix A
<table>
<thead>
<tr>
<th>Interfaces</th>
<th>Interactions</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMG based [50, 51, 52, 53, 54]</td>
<td>User wears EMG device on their lower arm</td>
<td>Amputees are able to use it</td>
<td>User needs to be trained, needs calibration, cannot control translation and orientation at the same time, high cost.</td>
</tr>
<tr>
<td>Kinect based [55, 56, 57]</td>
<td>Computer will process Kinect 3D images, and provide many predefined task options, such as grasping, get near and drop.</td>
<td>More autonomous than other methods, good for users who have difficulties with a touch screen, low cost, commercially available.</td>
<td>Not precise, not responsive, cannot see transparent objects, needs careful calibration, camera view can be obstructed by the robot arm, can only recognize predefined objects, limited task options.</td>
</tr>
<tr>
<td>Leap motion based [58, 59, 60]</td>
<td>Move hands above the device. Device can track the hand position within a specific range.</td>
<td>Low cost, commercial available, can recognize both hands using a single device.</td>
<td>Accuracy of the hand pose requires the user’s hand to be constantly suspended at a specific distance from the device, which may cause fatigue and abandonment.</td>
</tr>
<tr>
<td>IMU based [61, 62, 63, 64]</td>
<td>User wears an IMU device on their hands or head. The robotic arm is controlled by tilting the device.</td>
<td>Precise</td>
<td>Not commercial available, needs separate battery pack, can only control 3 DoF at a time, needs to switch modes to control all 6 DoF, not very intuitive for controlling XYZ translation.</td>
</tr>
<tr>
<td>Brain computer interfaces [65, 66, 67, 68]</td>
<td>User wears a BCI input device, and reacts to a stimulation patterns displayed on a screen.</td>
<td>Works good for users whose motor functions are completely paralyzed, and their cognitive abilities are intact.</td>
<td>Needs lengthy calibration, needs training, needs focused attention, causes fatigue, high cost, low responsiveness.</td>
</tr>
<tr>
<td>Joysticks based [69, 70]</td>
<td>One up button one down button, one control stick for X and Y, one stick for pitch and roll, and one stick for yaw</td>
<td>Low cost, commercially available, very precise and responsive.</td>
<td>Complicated 6 DoF control, not intuitive, requires both hands working together, which can be difficult for a person with upper limb disability.</td>
</tr>
</tbody>
</table>
### Table 3.1. (Continued)

<table>
<thead>
<tr>
<th>Control Interface</th>
<th>Methodology</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone based</td>
<td>Touch buttons to open/close gripper, drag virtual stick to control X/Y, drag slider to control Z.</td>
<td>Commercially available, low cost, wireless, light weight.</td>
<td>Too many buttons and sliders on the screen, cannot control orientation of the gripper. Can be useful for very simple tasks.</td>
</tr>
</tbody>
</table>

| Voice control     | Computer will recognize voice commands, including up, down, left, right, open and close | No need to use hands.                                                      | Not responsive, cannot control orientation of the gripper, very difficult to achieve ADLs. |

| Physical human-robot interaction | Users use their hand to hold a master robotic arm end-effector. The system will mirror the motion to the slave robot arm. | Responsive, precise, intuitive, can control 6 DoF at the same time.          | High cost, heavy device, difficult to carry.                                  |

| Haptic device     | Use Phantom Omni Premium to read the position and orientation of human hand, then directly mapping the motion to the robot hand. | Precise, responsive, can control 6 DoF at the same time.                     | High cost, heavy device with a lot of wires, difficult to carry.              |

| Others            | Connot achieve complex ADLs.                                                    |                                                                            |                                                                               |

For user safety, all control interfaces require the users to keep their finger on the phone screen while they control the robot, and will stop moving the arm if the user releases their finger from the screen.

When operating the robotic arm in the 3D Cartesian space, the user needs to provide the robot controller with six user inputs, three translation velocity values and three rotation velocity values, to control all Cartesian degrees of freedom (DoF). A typical smartphone with accelerometers and gyroscope sensors can supply 3 degrees of freedom of its own pose, which are directly related to the *pitch*, *roll* and *yaw* of the phone. These three values can be used as user input values to provide three of the six DoF values required to control the robot.
end-effector in the Cartesian coordinates. Furthermore, when the user touches the screen, the position of the touch on the screen can provide x and y directions on the screen surface, which can be used to control two more of the six DoF values required for Cartesian space motion. This provides a total of five input values out of the six needed input values from the user as shown in Figure 3.1.

We created a smartphone application that has several user interfaces. The application will save the initial touch position and orientation when the user places their finger on the screen, then as the user slides that finger on the screen or tilts the smartphone, the application is continuously using the current reading minus the initial reading, and the result is transmitted to the PC over WIFI through TCP/IP protocol at around 90Hz. If the user releases the finger from the screen, it will send a stop signal to the robot computer. If the user starts touching the screen again, the initial position and orientation will be updated. So the output signal of the smartphone is not velocities, it is rather displacement of positions and angles.
This control pattern is similar to how the joystick controls a wheelchair, the more you push the joystick the higher the wheelchair speed is.

Equation 3.1 shows how the phone output command is calculated.

\[
\begin{bmatrix}
\Delta x_{\text{phone}} \\
\Delta y_{\text{phone}} \\
\Delta \text{pitch}_{\text{phone}} \\
\Delta \text{roll}_{\text{phone}} \\
\Delta \text{yaw}_{\text{phone}}
\end{bmatrix}
= \begin{bmatrix}
x \\
y \\
\text{pitch} \\
\text{roll} \\
\text{yaw}
\end{bmatrix}_{\text{current}}
- \begin{bmatrix}
x \\
y \\
\text{pitch} \\
\text{roll} \\
\text{yaw}
\end{bmatrix}_{\text{initial}}
\]  
\text{(3.1)}

where \text{current} is the current position and orientation value of the phone, and \text{initial} is the position and orientation data recorded when the user starts touching the screen. Using relative phone pose reading will give users the freedom of using it in any gesture, such as lying on the bed or sitting on the sofa.

In order to utilize these phone output commands in controlling end-effector velocities, we use a diagonal square matrix \( C \) that includes coefficients that represent motion sensitivity gains for each motion direction, and unit converters from position and orientation to linear and angular velocities. Equation 3.2 shows the diagonal coefficient matrix \( C \).

\[
C = \begin{bmatrix}
c_x & 0 & 0 & 0 & 0 & 0 \\
0 & c_y & 0 & 0 & 0 & 0 \\
0 & 0 & c_z & 0 & 0 & 0 \\
0 & 0 & 0 & c_{\omega_x} & 0 & 0 \\
0 & 0 & 0 & 0 & c_{\omega_z} & 0 \\
0 & 0 & 0 & 0 & 0 & c_{\omega_z}
\end{bmatrix}
\]  
\text{(3.2)}
where \( c_x, c_y, c_z, c_{\omega_x}, c_{\omega_y}, c_{\omega_z} \) are scaling coefficients that are used as gains for the smartphone control vector, and unit conversion factors that convert displacement to speed.

### 3.2.1 One Button Interface

We decided to design an intuitive interface that is different from any other existing touchscreen/smartphone based solutions which have too many buttons. In this control interface, there is only one button, which is placed throughout the whole screen, that is required to control all 6 DoF of the end-effector. The user may start touching anywhere on the screen, and then drag their finger on screen to control the X and Y translations of the robot end-effector. Additionally, the user can control the robot end-effector orientation simultaneously by tilting the phone in a similar fashion as they want the end-effector to be oriented in the Cartesian space.

![One Button Interface](image)

**Figure 3.2.** One Button: slide translation & tilt rotation (combine).
Figure 3.2 shows the phone application view of the "One Button" interface. Equation 3.3 represents the mapping between the smartphone application output vector and the velocity vector of the robot end-effector. For the translation part, the farther you slide your finger on the screen, the faster the end-effector’s linear velocity is. For the rotational part, the more you tilt the smartphone, the faster the end-effector's angular velocity is.

\[
v_e = \begin{bmatrix}
\dot{x} \\
\dot{y} \\
\dot{z} \\
\omega_x \\
\omega_y \\
\omega_z
\end{bmatrix}_e = C \cdot \begin{bmatrix}
\Delta y_{phone} \\
\Delta x_{phone} \\
0 \\
\Delta roll_{phone} \\
\Delta pitch_{phone} \\
\Delta yaw_{phone}
\end{bmatrix}
\]

(3.3)

where \(v_e\) is the robot end-effector velocity vector, \(\dot{x}, \dot{y}, \dot{z}\) are translational speed in \(mm/s\). \(\omega_x, \omega_y, \omega_z\) are rotational speed in \(rad/s\). \(\Delta x_{phone}, \Delta y_{phone}\) are the planar distances between the initial touch position to the current touch position on the screen. \(\Delta pitch_{phone}, \Delta roll_{phone}\) and \(\Delta yaw_{phone}\) are the 3D orientation angles’ differences between the phone pose at the initial touch and the phone pose at the current touch on the screen. The unit of \(\Delta x_{phone}, \Delta y_{phone}\) is pixels, and the unit of \(\Delta pitch_{phone}, \Delta roll_{phone}\) and \(\Delta yaw_{phone}\) is degrees. In this application, \(c_x=0.25mm/pixels* s, c_y=0.3mm/pixels* s, c_{\omega_x}=0.003rad/degrees* s, c_{\omega_y}=c_{\omega_z}=0.0025rad/degrees* s\). These gain values were found by assigning a rough estimation first, then adjusting it according to users’ preference. The reason why the gain for sliding in the \(x\) direction is larger than that for sliding in the \(y\) direction is because the width of a smartphone is typically smaller than the length, and choosing a different gain value will yield a better range of velocities for the end-effector.

Notice that the motion in the \(Z\) direction is set to zero as a default since there are only 5 user input values from the phone as shown in equation 3.1. In order to control the robot in
the Cartesian Z direction without adding more buttons and sacrificing One Button design intention, we created the Z control mode. If the user would like to move the robot end-effector in the Z direction, the user needs to pitch the phone up prior to touching the screen so that the Z control mode is activated. If Z control mode is activated, only Y(left and right) and Z(up and down) axes of the robot end-effector can be controlled using the finger’s planar drag on the touch screen as shown in equation 3.4.

\[
v_e = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}_e = C \cdot \begin{bmatrix} 0 \\ \Delta x_{phone} \\ \Delta y_{phone} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.4}
\]

In this case, \( c_y = 0.3 \text{mm/pixels} \ast s \), \( c_z = 0.25 \text{mm/pixels} \ast s \).

The user can quit Z control mode by releasing their finger from the touch screen, and pitching the phone down to the floor plane. This alternating mapping of the finger dragging on the touch screen can provide intuitive control of the 3D position of the end-effector using the planar touch screen.

To toggle the status of gripper between the ”open” and ”close” positions, double-touching anywhere on the touch screen will alternate between the two gripper positions.

3.2.2 Three Buttons Interface

Some people prefer maneuvering translational motion and rotational motion independently, so Three Buttons interface is designed. It uses three buttons on the screen instead of a single button. Each one of these buttons, when touched, controls a specific Cartesian
motion of the gripper. The first button only activates the Cartesian orientation control of the robot end-effector. The second button only activates the Cartesian position control of the end-effector. The third button toggles the "open" and "close" positions of the end-effector.

Figure 3.3. Three Buttons: slide translation & tilt rotation (separate).

Figure 3.3 shows the user interface of the developed application. As shown, when Pitch-Roll-Yaw button is touched, tilting the phone in any direction will provide three DoF Cartesian orientation values for the robot end-effector to move to a similar orientation. In this case, the position of the end-effector will not be changed, dragging the finger on the touch screen will not cause any motion. Equation 3.5 shows the Cartesian velocity of the robot
end-effector when Pitch-Roll-Yaw button is touched.

\[
v_e = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}_e = C \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Delta roll_{\text{phone}} \\ \Delta pitch_{\text{phone}} \\ \Delta yaw_{\text{phone}} \end{bmatrix}
\] (3.5)

In this case, \(c_{\omega_x} = 0.003 \text{ rad/degrees} \ast s\), \(c_{\omega_y} = c_{\omega_z} = 0.0025 \text{ rad/degrees} \ast s\).

When XYZ button is touched, the three DoF Cartesian position values of the robot end-effector are provided through planar finger dragging on the touch screen for the X and Y values, and tilting the phone up and down (pitch direction) for the Z value. In this case, the orientation of the robot end-effector will not be changed, as the phone’s roll and yaw motions will not cause any motion. Equation 3.6 shows the Cartesian velocity of the robot end-effector when XYZ button is touched.

\[
v_e = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}_e = C \cdot \begin{bmatrix} \Delta y_{\text{phone}} \\ \Delta x_{\text{phone}} \\ \Delta pitch_{\text{phone}} \\ 0 \\ 0 \\ 0 \end{bmatrix}
\] (3.6)

In this case, \(c_x = 0.25 \text{ mm/pixels} \ast s\), \(c_y = 0.3 \text{ mm/pixels} \ast s\), \(c_z = 0.7 \text{ mm/degrees} \ast s\).

When the Gripper button is touched, the user will be able to control the "open" and "close" positions of the robot end-effector. A single touch will toggle the status of the gripper.
3.2.3 Tilt Interface

Some persons with upper extremity disabilities may have the ability to move their hand, but have limited finger motion and dexterity that makes it difficult to move their thumb across a smartphone screen. For such cases, we developed the third control interface, Tilt. This interface will eliminate the need for finger motion on the screen and will use only tilting to control all 6 DoF of the robotic arm.

![Image of Tilt interface](image)

Figure 3.4. Tilt: tilt translation & tilt rotation (separate).

As shown in Figure 3.4, this Tilt control interface also shows 3 buttons on the screen, similar to the Three Button interface. The function and control logic of the first button and third button are the same as that of the Three Buttons interface. The first button (Pitch-Roll-Yaw Tilt) uses the same mapping for the Cartesian velocity of the robot end-effector as in Equation 3.5. However, the second button (XYZ Tilt) uses different control logic. Equation 3.7 shows the Cartesian velocity vector of the robot end-effector when XYZ
button is touched.

\[
v_e = \begin{bmatrix}
\dot{x} \\
\dot{y} \\
\dot{z} \\
\omega_x \\
\omega_y \\
\omega_z
\end{bmatrix} = C \cdot \begin{bmatrix}
\Delta \text{roll}_{phone} \\
\Delta \text{yaw}_{phone} \\
\Delta \text{pitch}_{phone} \\
0 \\
0 \\
0
\end{bmatrix}
\]

(3.7)

In this case, \( c_x = c_y = c_z = 0.7 \, \text{mm/degrees} \times \text{s} \).

Inspired by how a screwdriver drives a screw, we use roll rotation of the phone to control the forward and backward motions of the end-effector. Rotating clockwise refers to moving forward and rotating counterclockwise refers to moving backward.

### 3.3 Experimental Setup

To perform preliminary testing of our new interfaces with ADL tasks, we recruited subjects that are not familiar with the system and have no prior experience in controlling robotic arms for ADL tasks. It is important to mention here that these tests are not meant to provide statistical significance, it is rather meant to provide feedback for adjustment of gains and to collect preliminary data on the metrics mentioned later in this chapter. For that purpose, we recruited a small number of human subjects to perform three specific ADL tasks. These subjects are three healthy subjects, and one subject with spinal cord injury who has lower body and upper limbs disabilities. Future work will include recruiting a significant number of subjects and performing clinical testing that can provide statistically significant data, and we will perform power analysis that we will publish in a future publication. The study was approved by the Internal Review Board under IRB#Pro00040871. All of the subjects have no experience in using our interfaces.
Three ADL tasks were selected for this test, based on a prior survey we conducted for persons with physical disabilities to find the most common ADL tasks for which they can use robotic assistance to perform. The first task is a water pouring task, in which, users need to control the robot arm to grasp a bottle of water and maneuver the bottle to pour water into a cup without spilling. As shown in Figure 3.5, the user controls the robotic arm using our developed interfaces to grasp the water bottle and carefully pour water into a cup, then place the water bottle back on top of the desk.

![Figure 3.5. First ADL task: Water pouring.](image)

(a) Grasping the bottle  
(b) Maneuvering  
(c) Pouring water into the bottle  
(d) Placing the bottle on the table

The second task is a plate&bowl pick-and-place task. As shown in Figure 3.6, the user controls the robotic arm to grasp a plate from the dishwasher and place it on top of the desk, then go back to the dish washer, grasp a bowl and place it on top of the plate.
Figure 3.6. Second ADL task: Pick and place plate & bowl from the dish washer to the tabletop.

The third task is pepper & salt grasping task for food seasoning, which includes opening cabinet doors and adding contents to a dining plate. As shown in Figure 3.7, the user controls the robotic arm to open the door of a cabinet, grasp a can of pepper, shake it onto the dining plate, and place the can back into the cabinet. Then grasp a can of salt, shake it onto the dining plate, and place the can back into the cabinet. Finally, the user closes the door of the cabinet.
Figure 3.7. Third ADL task: Seasoning the food with Pepper and Salt.

The hardware used for the experimental evaluation are the Baxter robot (for it’s low cost as an affordable assistant robot), a PC (which is a standard unit available in most homes and workplaces), and an iPhone SE (which is a typical cost effective smartphone). Our interfaces can be applied on most robotic arms and smartphones. The smartphone
sends the user input values to the PC through WiFi TCP/IP communication. Kinematics and redundancy resolution methods to control the end-effector in the Cartesian space are discussed in a separate publication [47], which is not the focus of this dissertation.

3.4 Evaluation and Results

We collected both quantitative and qualitative data from human subjects’ tests with the three ADL tasks. For the quantitative data, we recorded the time required for the users to perform each task, the lower performance time means the task was easier to complete using that interface. We also measured the total distance traveled by the robot end-effector, the lower distance traveled means the lower unnecessary motion executed when performing the task using that interface. All three interfaces were tested in reference to the ground (G) coordinate frame (One Button_G, Three Buttons_G, Tilt_G) and in reference to the end-effector (E) coordinate frame (One Button_E, Three Buttons_E, Tilt_E). Each user performed each task three times for each user interface and for each control reference frame. Since we have 3 user interfaces, 2 control reference frames, and 3 ADL tasks, the total number of experiments performed by each user was 54. After completing each ADL task, we asked the user to answer one qualitative question, which is: “How intuitive the interface was to complete this task?” Users answered the question on a scale from 0-10, where 10 is “very intuitive”, and 0 is “very difficult to use”.

Figure 3.8 compares the time required to complete each of the three ADL tasks using each interface. Subscripts _G and _E refer to the control in Ground reference and End-effector reference, respectively. The “X” markers in red color represent averages of all recorded time in each interface and task, and the dots in blue color represent the each recorded time. In the water pouring task, One Button and Three Buttons control in the ground reference frame required the least amount of time, while Tilt control in the ground reference frame required the highest amount of time. In the plate&bowl pick-and-place task, One Button
control in the ground reference frame required the least amount of time, while Tilt control in the end-effector reference frame required the highest amount of time. In the food seasoning task, Three Buttons control in both the ground and the end-effector reference frames achieved the shortest time, while Tilt control in the end-effector reference frame required
the highest amount of time. Compared to the joystick based solution presented in [70], our methods required 25% to 35% of the time. The main reason for this significant time savings is that they used a joystick, which controls 2 DoF at a time, and they needed to switch control modes frequently, which required much more time and effort to perform the task. However, our one button interface for controlling all 6 DoF has eliminated all unnecessary mode switching time.

Figure 3.9 compares the distance traveled of the end-effector to complete each ADL task using each user interface, the lower distance traveled means that users had more precise motion in using that interface and can easily eliminate unnecessary motions. Subscripts \(_G\) and \(_E\) refer to the control in Ground reference and End-effector reference, respectively. The “X” makers in red color represent averages of all recorded distance in each interface and task, and the dots in blue color represent the each recorded distance. In water pouring task, Three Buttons control in the ground reference frame achieved the least travel distance among the other interfaces, while Tilt control in the ground reference frame achieved the highest traveled distance among the other interfaces. In the plate&bowl pick and place task, One Button control in the ground reference frame achieved the least travel distance among the other interfaces due to the fact that the plate&bowl pick and place task requires a lot of rotational maneuvering, while Tilt control in the end-effector reference frame achieved the highest traveled distance among the other interfaces. One Button is the best interface choice for this situation, since it can control all 6 DoF at the same time. In the pepper&salt grasping task, Three Buttons control in both the ground and the end-effector reference frames required the least travel distance, while One Button control in the end-effector reference frame traveled the highest distance among the other interfaces.

Figure 3.10 compares the users’ feedback of average intuitiveness for each interface in the three ADL tasks. Subscripts \(_G\) and \(_E\) refer to the control in Ground reference and End-effector reference, respectively. In the water pouring task, both One Button and
Three Buttons control in the ground reference frame have high intuitiveness ratings, while Tilt control in the end-effector reference frame has the least intuitiveness rating. In the plate\&bowl pick and place task, users rated One Button control in the ground reference frame the highest intuitiveness rating, and Tilt control in the end-effector reference frame
3.5 Discussions

When the subject with disabilities used our smartphone interface, without any training, he was able to maneuver the robotic arm to open the cabinet door and grasp the object at the first trial. He was impressed with how easy it was to use, and he expressed how he liked it very much.
From users’ feedback, there are many areas that can be improved. First, for maneuvering the gripper in forward/backward translational motion, and in pitch or roll rotational motions, users preferred to maneuver in the end-effector reference frame. For maneuvering the gripper in left/right, up/down translational motions and in yaw rotational motion, users preferred to maneuver in the ground reference frame. We will develop a novel hybrid control coordinate system and its corresponding kinematics equations of the robotic arm so that we can achieve hybrid reference frame control for improved intuitiveness and more natural motion mapping. Second, users preferred One Button interface because it has only one button, and it eliminated the need for visual feedback. Users suggested adding haptic feedback when touching the buttons in Three Buttons and Tilt interfaces in order to eliminate the need for looking at the screen when touching the buttons. Additionally, users suggested adding audio feedback to all three interfaces when toggling the gripper status to open and close.

In the next chapter, we are going to develop a hybrid control kinematics methodology to fulfill users’ control preferences that were collected through human subject testing and survey.
Chapter 4: Development of Intuitive Reference Frame and Hybrid Kinematics

4.1 Introduction

We have identified potential improvements in the robotic arm control system that can make the control system work more efficiently and intuitively for the user. These improvements are related to the reference frame on which the end-effector Cartesian motion is based. The two commonly used reference frames are the Ground and the End-effector reference frames, which do not always provide intuitive control from the perspective of human users. When using the Ground Reference Frame or the End-effector Reference Frame to control a robotic arm, no matter what input device is used, people still find it difficult to maneuver the robotic arm to complete complex tasks that need translational motion and rotational motion throughout the process. Due to the poor intuitiveness of the two reference frames, human operators need to put a lot of effort to think from the robot’s perspective.

We conducted preliminary testing with human subjects to use Ground and End-effector reference frames separately to maneuver objects in 3D space[3]. Surveys from these subjects indicated that when operating a robotic arm, most subjects preferred to maneuver the forward/backward translation and roll rotation in the End-effector Reference Frame, but maneuver up/down translation and yaw rotation in the Ground Reference Frame. However, subjects also want to maneuver left/right translational motion along an axis which is perpendicular to the end-effector and parallel to the ground surface, which is neither described in the Ground Reference Frame nor in the End-effector Reference Frame. Subjects also want to maneuver pitch rotational motion around this axis.
This chapter introduces a new intuitive method to control the robotic arm using a new reference frame and hybrid inverse kinematics. We will first explain the control theories using the two conventional reference frames, then we will present the development of our new reference frame, we named it “Intuitive” reference frame, and hybrid inverse kinematics that fulfills all user’s preferences mentioned above. We will then present the setup of a complex ADL task to test the performance using the three different reference frames. Results show the quantitative and qualitative data collected with human subjects testing, and we will analyze the data to highlight the difference between the two conventional control reference frames, and control using our ”Intuitive” reference frame.

4.2 System Modeling

A complete 6 DoF Cartesian space user input velocity vector for controlling robotic arms is described in equation 4.1. This information comes from an input device, such as joysticks, Kinect camera, or IMU module. Figure 4.1 shows an example of how the 6 DoF user input vector is related to the hardware of the input device.

\[
v_u = \begin{bmatrix} \dot{FB} \\ \dot{LR} \\ \dot{UD} \\ \dot{roll} \\ \dot{pitch} \\ \dot{yaw} \end{bmatrix}
\]  

(4.1)

where \(v_u\) is the user input Cartesian velocity command consisting of three translational motion speeds and three rotational motion speeds. \(\dot{FB}\) refers to forward and backward speed, \(\dot{LR}\) refers to left and right speed, \(\dot{UD}\) refers to up and down speed, \(\dot{roll}, \dot{pitch}, \)
yaw refer to the Cartesian angular velocities commands. Vector $v_u$ have a different control perspective based on the used reference frame. We added pre-superscripts $G$, $E$, or $I$ to the vector $v_u$ in the following context, referring the Ground, the End-effector and the Intuitive reference frames, respectively.

![Diagram of 6 degrees of freedom controls](image)

Figure 4.1. The user input device is a spacemouse and the illustration of 6 degrees of freedom controls include FB: forward or backward, UD: up or down, LR: left or right and roll, pitch, yaw for rotation.

4.2.1 Ground Reference Frame Control System

To control the robotic arm in reference to the Ground frame, we use equation 4.2 as follows:

$$
\dot{q} = \begin{bmatrix}
\dot{\theta}_1 \\
\dot{\theta}_2 \\
\vdots \\
\dot{\theta}_n 
\end{bmatrix} = \left(J_G^T \right)^+ \cdot 
\begin{bmatrix}
FB \\
LR \\
UD \\
roll \n\end{bmatrix}^{G} \\
\begin{bmatrix}
pitch \\
yaw
\end{bmatrix} 
$$

(4.2)
where $J_G$ is the Jacobian matrix of the robotic arm with respect to the Ground Reference Frame, $J_G$ can be calculated from the forward kinematics. $[\ldots]^+$ is the operation of pseudo-inverse. $\dot{q}$ is the joint velocity of the robotic arm, which will be sent to the robot controller directly in real-time.

Figure 4.2 shows how the user input vector is mapped to the Ground Reference Frame. As we can see in this figure, regardless of how the Cartesian position and orientation of the end-effector is, the control of the end-effector is always with respect to the Ground Reference Frame.

Figure 4.2. Illustration of the Ground Reference Frame and the mapped 6 DoF user control vectors.
4.2.2 End-effector Reference Frame Control System

In order to find the Jacobian matrix with respect to the End-effector Reference Frame, a frame transformation can be performed as shown in equation 4.3:

$$J_E = \begin{bmatrix} G_{ER} & 0 \\ 0 & G_{ER} \end{bmatrix}^T \cdot J_G$$  \hspace{1cm} (4.3)

where $G_{ER}$ is the rotation matrix that describes the End-effector Reference Frame relative to the Ground Reference Frame, which can be calculated from the forward kinematics. $J_E$ is the Jacobian matrix with respect to the End-effector Reference Frame.

To control the robotic arm in reference to the End-effector frame, we use equation 4.4 as follows:

$$\dot{q} = \hat{\theta}_1 \dot{} \hat{\theta}_2 \cdots \hat{\theta}_n = [J_E]^+ \cdot \begin{bmatrix} FB \\ LR \\ UD \\ \text{roll} \\ \text{pitch} \\ \text{yaw} \end{bmatrix}$$  \hspace{1cm} (4.4)

Figure 4.3 shows how the user input vector is mapped when using the End-effector Reference Frame. Different from the Ground Reference Frame, the End-effector Reference Frame is changing with the current pose of the end-effector.
4.2.3 Intuitive Reference Frame Control System

We conducted testing with 12 human subjects to find out the control preference when teleoperating the Baxter robot to perform ADL tasks. User preference parameters were collected and analyzed to develop requirements for our new Intuitive Reference Frame. Table 4.1 shows a summary of the requirements that we adopted for our new Intuitive Reference Frame.
Table 4.1. Technical requirements of building an intuitive robotic arm teleoperation system summarized from users’ feedback.

<table>
<thead>
<tr>
<th>#</th>
<th>Direction</th>
<th>User preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Forward/</td>
<td>Along the $X$-axis of the End-effector Reference Frame, which is the $\hat{X}_E$ axis.</td>
</tr>
<tr>
<td></td>
<td>Backward</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Left/Right</td>
<td>Along an axis which is perpendicular to the end-effector axis $\hat{X}_E$ and parallel to the ground surface.</td>
</tr>
<tr>
<td>3</td>
<td>Up/Down</td>
<td>Along the $Z$-axis of the Ground Reference Frame, which is the $\hat{Z}_G$ axis.</td>
</tr>
<tr>
<td>4</td>
<td>roll</td>
<td>Around the $\hat{X}_E$ axis of the End-effector Reference Frame.</td>
</tr>
<tr>
<td>5</td>
<td>pitch</td>
<td>Around an axis which is perpendicular to the end-effector axis $\hat{X}_E$ and parallel to the ground surface.</td>
</tr>
<tr>
<td>6</td>
<td>yaw</td>
<td>Around the $\hat{Z}_G$ axis of the Ground Reference Frame.</td>
</tr>
</tbody>
</table>

According to requirements #1 and #4, we assigned the $X$ axis of the Intuitive Reference Frame to be the same as the $X$ axis of the End-effector Reference Frame.

$$^I\hat{X}_I = ^E\hat{X}_E = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad (4.5)$$

where $^I\hat{X}_I$ is the unit vector of $X$ axis of the Intuitive Reference Frame described in its own frame, and $^E\hat{X}_E$ is the unit vector of $X$ axis of the End-effector Reference Frame described in its own frame.
To describe the vector $\hat{X}_I$ relative to the Ground Reference Frame, we used a rotation operation as shown in equation 4.6.

$$G\hat{X}_I = G \hat{E} R \cdot \hat{I}_E \hat{X}_I = G \hat{E} R \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$  \hspace{1cm} (4.6)$$

where $G\hat{X}_I$ is the unit vector $\hat{X}_I$ described in the Ground Reference Frame. This fulfills the requirements #1 and #4, allowing the user to control FB based on the end-effector’s $X$ axis.

In order to fulfill requirements #2 and #5, we constructed a unit vector that is perpendicular to the end-effector and parallel to the ground surface, by finding the cross product between the $Z$ axis of the Ground frame and the $X$ axis of the Intuitive frame as shown in equation 4.7.

$$G\hat{Y}_I = G\hat{Z}_G \times G\hat{X}_I = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \times G\hat{X}_I$$  \hspace{1cm} (4.7)$$

where $G\hat{Y}_I$ is a vector that is perpendicular to the end-effector and parallel to the ground surface, described in the Ground Reference Frame. Please notice that when the end-effector is perpendicular to the ground surface, $G\hat{X}_I$ is equal to $G\hat{Z}_G$, which results in algorithmic singularity. When the angle between $G\hat{X}_I$ and $G\hat{Z}_G$ is less than 10 degrees, we switch from the Intuitive frame to the Ground frame to avoid the algorithmic singularity.

To normalize $G\hat{Y}_I$, we use equation 4.8 as follows.

$$G\hat{Y}_I = \frac{G\hat{Y}_I}{|G\hat{Y}_I|}$$  \hspace{1cm} (4.8)$$

where $G\hat{Y}_I$ is a unit vector of $G\hat{Y}_I$. 

52
The Z axis of the Intuitive Reference Frame can be calculated using the cross product of unit vectors X and Y of the Intuitive Reference Frame as shown in equation 4.9.

\[ \hat{Z}_I = \hat{X}_I \times \hat{Y}_I \]  

(4.9)

where \( \hat{Z}_I \) is a unit vector that is perpendicular to \( \hat{X}_I \) and \( \hat{Y}_I \), described in the Ground Reference Frame.

Now the Intuitive Reference Frame is fully defined, which is represented by \( \hat{X}_I \), \( \hat{Y}_I \), and \( \hat{Z}_I \). Figure 4.4 shows our Intuitive Reference Frame, which fulfills requirements #1, #2, #4 and #5.

Figure 4.4. Illustration of the Intuitive Reference Frame and the mapped 6 DoF user control vectors.
In order to find the Jacobian matrix relative to the Intuitive Reference Frame, we need to find the rotation matrix that describes the Ground Reference Frame relative to the Intuitive Reference Frame, as shown in equation 4.10.

\[
{^I_G R} = \begin{bmatrix}
{^G \hat{X}_G} {^G \hat{Y}_G} {^G \hat{Z}_G} \\
{^G \hat{X}_G} {^G \hat{Y}_G} {^G \hat{Z}_G} \\
{^G \hat{X}_G} {^G \hat{Y}_G} {^G \hat{Z}_G}
\end{bmatrix}
\]  

(4.10)

where \( {^G \hat{X}_G} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \), \( {^G \hat{Y}_G} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \), \( {^G \hat{Z}_G} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \). \( {^I_G R} \) is the rotation matrix that describes the Ground Reference Frame relative to the Intuitive Reference Frame.

We can now find the Jacobian equation relative to the Intuitive Reference Frame as shown in equation 4.11.

\[
J_I = \begin{bmatrix}
{^I_G R} & 0 \\
0 & {^I_G R}
\end{bmatrix} \cdot J_G
\]  

(4.11)

where \( J_I \) is the Jacobian matrix relative to the Intuitive Reference Frame.

In order to fulfill requirements #3 and #6, we developed a hybrid inverse kinematics solution that can account for Cartesian motions of the end-effector that are relative to different reference frames, as shown in equation 4.12.

\[
\dot{\mathbf{q}} = \begin{bmatrix}
\dot{\theta}_1 \\
\dot{\theta}_2 \\
\vdots \\
\dot{\theta}_n
\end{bmatrix} = \left[ J_G \right]^+ \begin{bmatrix}
0 \\
0 \\
\dot{U}D \\
0 \\
yaw
\end{bmatrix} + \left[ J_I \right]^+ \begin{bmatrix}
\dot{F}B \\
LR \\
0 \\
\dot{roll} \\
\dot{pitch} \\
0
\end{bmatrix}
\]  

(4.12)
Equation 4.12 fulfills all the requirements we proposed in table 4.1. The up/down translational motion and yaw rotational motions are controlled in the Ground Reference Frame. However, the left/right, forward/backward translational motions, and pitch, roll rotational motions are controlled in the Intuitive Reference Frame, as shown in figure 4.4.

4.3 Experiments

In order to test the performance of our novel Intuitive Reference Frame control system and the hybrid inverse kinematics solution. We designed a complex ADL task that includes many subtasks that require interaction with cabinets, utensils, dishes and cans. Human subjects were recruited to perform this ADL task using the three different reference control systems. In order to eliminate the uncertainty and possible noise from the input device, such as IMU-based solutions or camera-based methods, we chose to use a 6 DoF joystick based input device (spacemouse) from 3D connexion, as shown in figure 4.1. The two side buttons are used as open and close of the gripper.

4.3.1 Introduction to the ADL Task

In order to test all 6 degrees of motion, we chose to perform a food seasoning task. Users need to use the spacemouse to control the Baxter robotic arm to grasp a bowl and add sea salt into the bowl contents. This task is divided into the following steps (subtasks), as shown in figure 4.5. Step 1: the user needs to maneuver the robotic arm to open the cabinet door. Step 2: grasp a bowl from inside the cabinet, then place the bowl onto the desk. Step 3: grasp a can of sea salt from inside the cabinet, shake some sea salt into the bowl, then place the sea salt can back into the cabinet. Step 4: close the cabinet door.
4.3.2 Human Subjects Recruitment and Testing

Twelve healthy human subjects were recruited to perform the ADL task, 10 of them are males, 2 of them are females. Their age range is from 24 to 42. The study was approved by the Internal Review Board under IRB#Pro00040871. All of the subjects have no prior experience in operating robotic arms.
We collected both quantitative and qualitative data from human subjects’ tests. For the quantitative data, we recorded the time required for the subjects to perform the task, the less time required means that the task was easier to complete using that reference frame system. Each user performed each task two times for each reference frame system. Since we have 12 users, 3 reference frames, and 2 trials for each reference frame system, the total number of experiments performed was 72. After completing the test, we asked the user to answer one qualitative question, which is: “How intuitive the reference frame system was to complete this task?”. Users answered the question on a scale from 0-10, where 10 is “very intuitive”, and 0 is “very difficult to use”. Before each experiment, we allowed the user to learn to control the robot for 2-5 minutes in the reference frame that they will use for the task. The following symbols are used to indicate the reference frame used during the experiments: G or **Ground** stands for the Ground Reference Frame control system, E or **End-effector** stands for the End-effector Reference Frame control system, H or **Hybrid** stands for the Intuitive Reference Frame system with hybrid inverse kinematics. The sequence of testing the three reference frame systems are intentionally randomized, such as E H G H E G or H E G G H E, in order to eliminate the possible biases introduced by the user learning effect.

### 4.4 Results

Figure 4.6 and Figure 4.7 include each recorded values and their averages for each categories. In figure 4.6, we can see that when controlling the robot using the Ground reference frame, the average time required to complete the task is 372 seconds, the minimum is 203 seconds, and the maximum is 629 seconds. When controlling the robot using the End-effector reference frame, the average time required is 389 seconds, the minimum is 221 seconds, and the maximum is 710 seconds. When controlling the robot using the Intuitive reference frame with the hybrid inverse kinematics (Hybrid system), the average time required is 237 seconds, the minimum is 130 seconds, and the maximum is 354 seconds.
average time required, using the Hybrid system reduced the time required to complete the task by 36% compared to the Ground, and 39% compared to the End-effector. Please notice that when using H, the maximum time required is 354 seconds which is larger than the minimum time required in G (203 seconds) or in E (221 seconds). This does not mean that any subjects using G or E performed faster than using H. Actually, each one of the 12 subjects performed the ADL task faster when using the H.
Figure 4.7 shows the qualitative data of intuitiveness rating results from all subjects. Please note that due to the intuitiveness rating are discrete data, so there are data points coincident with each other. For the **Ground** system, the average intuitiveness rating is 4.7, the lowest is 1 and the highest is 8. For the **End-effector** system, the average intuitiveness rating is 5.8, the lowest is 3 and the highest is 8. For the **Hybrid** system, the average intuitiveness rating is 8.7, the lowest is 7 and the highest is 10.

After each subject completed the experiment, we conducted a short interview with the user, asking “how you feel about the three systems including the advantages and disadvantages of these systems and improvement suggestions”. Two subjects mentioned that they prefer to control translation and rotation motions separately, using two joystick knobs. Eight subjects mentioned that controlling the robotic arm using the **Hybrid** system is the way they wanted the robot to move. However, when using G or E, they needed to think from the robot’s perspective, which was more prone to making mistakes in maneuvering the robotic arm and it diverted their attention away from the task, then caused frustration over time.
4.5 Analysis

In order to better explain why the subjects feel that G and E systems are not easy to use, we created figure 4.8, which represents a typical scenario of manipulating a robotic arm during an ADL task. The figure shows that the user is going to control the robotic arm to move the grasped bowl to the left side of the table which is indicated by the arrow in black color. The three reference frames are placed into the figure near the end-effector according to the current pose of the end-effector.

Figure 4.8. A case study of how the Ground(G), End-effector(E) and Intuitive(I) reference frames behave when manipulating a robotic arm in a typical ADL task scenario.
If the user wants to move the grasped bowl to the left side of the table, when using G, the user needs to push the spacemouse forward and left at the same time, additionally, at the same pace which will add extra cognitive load to the user. When using E, the user needs to push the spacemouse downward in order to move the robotic arm to the left side of the table which is very unintuitive for the user. As we can see in frame E, the Y-axis is facing up due to the rotation of the gripper, so when the user pushes the spacemouse left or right, the end-effector will move up and down. However, when using the Intuitive Reference Frame, the user just needs to push the spacemouse left which is the same as where he/she wants the bowl to be moved.
Chapter 5: Performance Test of the Hybrid System with Different Input Devices.

5.1 Introduction

From the experimental results presented in chapter 4, we can see that when using the Spacemouse as an input device, the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid system) significantly improved the intuitiveness of the user interface and decreased the time required to perform an ADL task (see figures 4.6 and 4.7). Theoretically, all other user interfaces should also benefit from our Hybrid system automatically without the need to change any other part of their interface since the user input interfaces and the robotic arm kinematics are independent systems. When any user input device is used (for example, joysticks, Omni, Kinect cameras, or touchscreens), the output of these devices is the same six degrees of freedom Cartesian space velocity vector, which is X, Y, Z, Roll, Pitch, and Yaw.

In this chapter, we will further verify the adaptability, significance, and usefulness of our intuitive user interface. We will choose four different user input devices and test with the two conventional reference frames and our Hybrid system, perform the same ADL task, and analyze the data to find out if our method improves the performance and intuitiveness of the whole human-robot teleoperation system.
5.2 Experiment Setup

The first user input device is the smartphone. We chose one of our most easy to use user interfaces (UI), which is "One Button" interface, as mentioned in chapter 3. In this control interface, there is only one button, as shown in figure 5.1, which is placed throughout the whole screen, that is required to control all 6 DoF of the end-effector.

![One Button Interface](image)

Figure 5.1. *One Button*: slide translation & tilt rotation(combine).

The second user input device is a Microsoft Xbox wireless controller. We adopted the UI design from Campeau-Lecours group’s work [5]. Figure 5.2 shows the UI design of this interface. The forward/backward(X) and left/right(Y) directions are controlled by the left-hand joystick, the up/down(Z) direction is controlled by left hand two push buttons. Right-hand joystick controls Pitch and Roll rotations, right-hand buttons "X" and "B" control Yaw rotation, and right-hand buttons "Y" and "A" control the gripper’s open and close.
The third user input device is an Omni, as shown in figure 5.3, which has six joints. From the supplied SDK, we can have Cartesian space position and orientation information of the endpoint, which is the stylus tip. Since the first three joints are motorized for force feedback, we programmed this device to keep the stylus at the center point by providing a force that is proportional to the distance from the center point. The front button on the stylus is used to initiate the rotational control. The rotational output Roll, Pitch, Yaw are proportional to the difference between the current orientation and the initial orientation (when the front button pressed) of the stylus.
The fourth user input device is a Spacemouse, which is the same one we used in chapter 4. The UI design in this test is also identical to what we used in chapter 4.

The ADL task we chose is the same task used in chapter 4. The user uses the four different input devices with three different reference frame systems to control the Baxter robotic arm to perform a food seasoning task. This task is divided into the following steps (subtasks), as shown in figure 5.4, Step 1: maneuver the robotic arm to open the cabinet door. Step 2: grasp a bowl from inside the cabinet. Step 3: place the bowl onto the desk. Step 4: grasp a can of sea salt from inside the cabinet. Step 5: shake some sea salt into the bowl, then place the sea salt can back into the cabinet. Step 6: close the cabinet door.

Figure 5.4. The ADL task: food seasoning.
We recorded the time required for the user to perform the task, where lower performance time means the task was easier to complete using that reference frame. We also measured the total distance traveled by the robot’s end-effector. Lower traveled distance means less unnecessary robot motion during task execution when using that reference frame. All four user input devices were tested in reference to the Ground(G) Reference Frame, the End-effector(E) Reference Frame, and the Intuitive Reference Frame with hybrid inverse kinematics(H). The user performed the task three times for each device and for each reference frame. Since we have 4 user input devices, 3 control reference frames, and 3 trails each, the total number of experiments performed was 36.

5.3 Experimental Results

![Figure 5.5. Time required for completing the ADL task when using the smartphone with the three reference frames (the lower the better).](image)

Figure 5.5 shows the time required to perform the ADL task when using the smartphone as the input device. When controlling the robotic arm using the Ground Reference Frame, the average time required was 262 seconds, the maximum was 317 seconds, and the minimum was 232 seconds. When controlling the robotic arm using the End-effector Reference Frame,
the average time required was 204 seconds, the maximum was 230 seconds, and the minimum was 191 seconds. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average time required was 144 seconds, the maximum was 174 seconds, and the minimum was 110 seconds. Comparing the average time required, using the Hybrid system reduced the time required to complete the task by 45% compared to the Ground Reference Frame, and 29.4% compared to the End-effector Reference Frame.

Figure 5.6 shows the end-effector travel distance when using the smartphone as the input device. When controlling the robotic arm using the Ground Reference Frame, the average end-effector travel distance was 6719 mm, the maximum was 8380 mm, and the minimum was 5880 mm. When controlling the robotic arm using the End-effector Reference Frame, the average travel distance was 5939 mm, the maximum was 6416 mm, and the minimum was 5694 mm. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average travel distance was 4671 mm, the maximum was 5531 mm, and the minimum was 3674 mm. Comparing the average travel distance, using the Hybrid system reduced the end-effector travel distance needed to complete the task by
30.4% compared to the Ground Reference Frame, and 21.4% compared to the End-effector Reference Frame.

![Graph showing ADL completion time - Xbox](image)

Figure 5.7. Time required for completing the ADL task when using the Xbox with the three reference frames (the lower the better).

Figure 5.7 shows the time required to perform the ADL task when using Microsoft Xbox controller as the input device. When controlling the robotic arm using the Ground Reference Frame, the average time required was 131 seconds, the maximum was 146 seconds, and the minimum was 112 seconds. When controlling the robotic arm using the End-effector Reference Frame, the average time required was 226 seconds, the maximum was 277 seconds, and the minimum was 186 seconds. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average time required was 128 seconds, the maximum was 147 seconds, and the minimum was 115 seconds. Comparing the average time required, using the Hybrid system reduced the time required to complete the task by 2.3% compared to the Ground Reference Frame, and 43.4% compared to the End-effector Reference Frame.
Figure 5.8. Robot end-effector distance traveled when using the Xbox with the three reference frames (the lower the better).

Figure 5.8 shows the end-effector travel distance when using Microsoft Xbox controller as the input device. When controlling the robotic arm using the Ground Reference Frame, the average end-effector travel distance was 6594 mm, the maximum was 7013 mm, and the minimum was 6370 mm. When controlling the robotic arm using the End-effector Reference Frame, the average travel distance was 10143 mm, the maximum was 11752 mm, and the minimum was 8508 mm. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average travel distance was 6484 mm, the maximum was 6915 mm, and the minimum was 6002 mm. Comparing the average travel distance, using the Hybrid system reduced the end-effector travel distance needed to complete the task by 1.7% compared to the Ground Reference Frame, and 36.1% compared to the End-effector Reference Frame.
Figure 5.9. Time required for completing the ADL task when using the Omni with the three reference frames (the lower the better).

Figure 5.9 shows the time required to perform the ADL task when using Omni as the input device. When controlling the robotic arm using the Ground Reference Frame, the average time required was 177 seconds, the maximum was 278 seconds, and the minimum was 125 seconds. When controlling the robotic arm using the End-effector Reference Frame, the average time required was 163 seconds, the maximum was 173 seconds, and the minimum was 149 seconds. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average time required was 119 seconds, the maximum was 134 seconds, and the minimum was 101 seconds. Comparing the average time required, using the Hybrid system reduced the time required to complete the task by 32.8% compared to the Ground Reference Frame, and 27% compared to the End-effector Reference Frame.
Figure 5.10. Robot end-effector distance traveled when using the Omni with the three reference frames (the lower the better).

Figure 5.10 shows the end-effector travel distance when using Omni as the input device. When controlling the robotic arm using the Ground Reference Frame, the average end-effector travel distance was 10365 mm, the maximum was 18525 mm, and the minimum was 6144 mm. When controlling the robotic arm using the End-effector Reference Frame, the average travel distance was 8052 mm, the maximum was 8409 mm, and the minimum was 7434 mm. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average travel distance was 6601 mm, the maximum was 7703 mm, and the minimum was 5800 mm. Comparing the average travel distance, using the Hybrid system reduced the end-effector travel distance needed to complete the task by 36.3% compared to the Ground Reference Frame, and 18% compared to the End-effector Reference Frame.
Figure 5.11. Time required for completing the ADL task when using the Spacemouse with the three reference frames (the lower the better).

Figure 5.11 shows the time required to perform the ADL task when using Spacemouse as the input device. When controlling the robotic arm using the Ground Reference Frame, the average time required was 223 seconds, the maximum was 238 seconds, and the minimum was 195 seconds. When controlling the robotic arm using the End-effector Reference Frame, the average time required was 208 seconds, the maximum was 270 seconds, and the minimum was 165 seconds. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average time required was 142 seconds, the maximum was 150 seconds, and the minimum was 135 seconds. Comparing the average time required, using the Hybrid system reduced the time required to complete the task by 36.3% compared to the Ground Reference Frame, and 31.7% compared to the End-effector Reference Frame.
Figure 5.12 shows the end-effector travel distance when using Spacemouse as the input device. When controlling the robotic arm using the Ground Reference Frame, the average end-effector travel distance was 7957 mm, the maximum was 8840 mm, and the minimum was 6805 mm. When controlling the robotic arm using the End-effector Reference Frame, the average travel distance was 8098 mm, the maximum was 11417 mm, and the minimum was 6138 mm. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average travel distance was 5317 mm, the maximum was 5589 mm, and the minimum was 5094 mm. Comparing the average travel distance, using the Hybrid system reduced the end-effector travel distance needed to complete the task by 33.2% compared to the Ground Reference Frame, and 34.3% compared to the End-effector Reference Frame.
Figure 5.13. Time required for completing the ADL task, including data from all devices used with the three reference frames (the lower the better).

Figure 5.13 shows the time required to perform the ADL task when all test data is included from all four user input devices. When controlling the robotic arm using the Ground Reference Frame, the average time required was 198 seconds, the maximum was 317 seconds, and the minimum was 112 seconds. When controlling the robotic arm using the End-effector Reference Frame, the average time required was 201 seconds, the maximum was 277 seconds, and the minimum was 149 seconds. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average time required was 133 seconds, the maximum was 174 seconds, and the minimum was 101 seconds. Comparing the average time required, using the Hybrid system reduced the time required to complete the task by 32.8% compared to the Ground Reference Frame, and 33.8% compared to the End-effector Reference Frame.
Figure 5.14 shows the end-effector travel distance when all test data is included from all four user input devices. When controlling the robotic arm using the Ground Reference Frame, the average end-effector travel distance was 7909 mm, the maximum was 18525 mm, and the minimum was 5880 mm. When controlling the robotic arm using the End-effector Reference Frame, the average travel distance was 8058 mm, the maximum was 11752 mm, and the minimum was 5694 mm. When controlling the robotic arm using the Intuitive Reference Frame with hybrid inverse kinematics (Hybrid), the average travel distance was 5768 mm, the maximum was 7703 mm, and the minimum was 3674 mm. Comparing the average travel distance, using the Hybrid system reduced the end-effector travel distance needed to complete the task by 27.1% compared to the Ground Reference Frame, and 28.4% compared to the End-effector Reference Frame.
Figure 5.15. Performances comparison between devices regardless of control method - completion time.

Figure 5.15 shows a comparison between four user input devices for task completion time. Averages show all data collected when using each user interface device, regardless of what control frames were used. As we can see, using the Omni user interface achieved the least amount of time in average.
Figure 5.16. Performances comparison between devices regardless of control method - distance traveled.

Figure 5.16 shows a comparison between four user input devices for the end-effector travel distance. Averages show all data collected when using each user interface device, regardless of what control frames were used. As we can see, using the smartphone user interface achieved the least travel distance in average.

5.4 Results Analysis

In general, the results show that no matter which input device was adopted, our novel Intuitive Reference Frame with hybrid inverse kinematics (the Hybrid system) improved the speed and the accuracy of the human-robot interface system. Based on the users' feedback, this is because the Hybrid system is much more intuitive than the other two reference frames. When using the Ground Reference Frame or the End-effector Reference Frame, users needed to concentrate their mind completely on the task in order to avoid mistakes. Even with
full concentration, users were still making mistakes, resulting in unnecessary motions and efforts were needed to correct these mistakes. However, when using the Hybrid system, users rarely made mistakes, and fewer corrections were needed to complete the task, so users spent less time, less effort, and they found it much easier to maneuver the robotic arm. This also explains why the performance deviations were much larger when they used the Ground Reference Frame and the End-effector Reference Frame compared to when the Hybrid method was used.

It is interesting to see that when using the Microsoft Xbox wireless controller as the input device, the Ground Reference Frame and the Hybrid system show similar time and distance. The reason for this is that controlling translational motions and rotational motions simultaneously is not only possible but also comfortable using this device. When any mistakes occur, users can correct them (usually orientation imperfections) with the right-hand joystick while moving the object with the left-hand joystick. However, using the Ground Reference Frame does require much more concentration and effort.

When comparing the performance of the four tested user input devices, results show that the Omni with the Hybrid control achieved the fastest speed, while the Smartphone with the Hybrid control achieved the highest accuracy (the lower distance traveled means fewer mistakes occurred). The reason is that the four devices have different gain settings when mapped to the 6 DoF Cartesian velocity vectors. The gain setting in the smartphone interface was much smaller than used in the Omni interface. The higher gain results in higher maximum speed, but also a higher chance of making mistakes or having less accuracy. Gain values are irrelevant to the relative effectiveness of user input devices. All devices can adopt high or low gains. Based on the discussion with our subjects, novel users usually prefer smaller gains, and more experienced users prefer higher gains. Table 5.1 is a summary of the average values from all collected data points. Numbers in red color represent lowest value (the lower the better).
Table 5.1. Comparison between the four user input devices and the three reference frames.

<table>
<thead>
<tr>
<th>Interfaces</th>
<th>Ground</th>
<th>End-effector</th>
<th>Hybrid</th>
<th>All frames average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time(s)</td>
<td>Distance (mm)</td>
<td>Time(s)</td>
<td>Distance (mm)</td>
</tr>
<tr>
<td>Smartphone</td>
<td>262</td>
<td>6719</td>
<td>204</td>
<td>5939</td>
</tr>
<tr>
<td>Xbox</td>
<td>131</td>
<td>6594</td>
<td>226</td>
<td>10143</td>
</tr>
<tr>
<td>Omni</td>
<td>177</td>
<td>10365</td>
<td>163</td>
<td>8052</td>
</tr>
<tr>
<td>Spacemouse</td>
<td>223</td>
<td>7957</td>
<td>208</td>
<td>8098</td>
</tr>
<tr>
<td>All devices average</td>
<td>198</td>
<td>7909</td>
<td>201</td>
<td>8058</td>
</tr>
</tbody>
</table>
Chapter 6: Conclusions and Future Work

6.1 Conclusions

The goal of this research is to provide intuitive and easy to use human-robot teleoperation interfaces for users. We reviewed state-of-the-art interfaces and found many advantages and shortcomings in each reviewed interface method. For example, the IMU-based solution is very intuitive in controlling rotational motions, but it needs extra buttons for mode switching to control the translational motions. The most widely used joystick-based solution is very efficient in controlling translational motions, but it can control only 2-3 DoF at a time. We developed smartphone based user interfaces with three different mapping strategies using the X, Y translation of the touchscreen information and the Roll, Pitch, Yaw rotations from accelerometers and gyroscopes inside the smartphone. Our novel smartphone-based human-robot control interfaces achieved intuitive and effortless control of the robotic arm in all 6 DoF. It accomplished relatively complex activities of daily living tasks that other methods failed or were very difficult to perform. All users, without any prior experience, were able to get used to our intuitive smartphone-based interfaces quickly and successfully complete the challenging ADL tasks without any training. Compared to most widely used joystick based interfaces, our interfaces required a single hand, and were very intuitive, especially when controlling the rotational motion of the robot arm. Compared to other popular interfaces reviewed in chapter 2, the newly designed smartphone user interface has many advantages: (1) it does not need calibration, (2) it does not need training and users can get used to these interfaces in few minutes, (3) cost is very low, (4) it is highly responsive and precise, (5) it is
effortless, wireless, lightweight and uses relative phone pose readings, (6) users can use this interface in various positions, such as sitting on the wheelchair or laying on bed, and (7) it is safe as the robotic arm stops moving immediately once the user’s finger is released from the phone screen.

Not only the interface devices affect the intuitiveness of the robotic arm teleoperation, but also the frame of reference and inverse kinematics algorithm of the robotic arm are also crucial to the whole system intuitiveness. We found that the two conventional reference frames, ground and end-effector, do not always provide intuitive control from the perspective of human operators. Users indicated that when operating a robotic arm, most users preferred to maneuver the forward/backward translation and roll rotation in the End-effector Reference Frame, but maneuver up/down translation and yaw rotation in the Ground Reference Frame. However, users also want to maneuver left/right translational motion along an axis which is perpendicular to the end-effector and parallel to the ground surface, which is neither described in the Ground Reference Frame nor in the End-effector Reference Frame. Users also want to maneuver pitch rotational motion around this axis. We successfully fulfilled all users’ expectations by developing a novel Intuitive Reference frame and a hybrid inverse kinematics solution, which made maneuvering the robotic arm to perform ADL tasks easier for human operators. The Intuitive reference frame uses an intuitive Y axis, which is perpendicular to the end-effector linkage and is parallel to the ground surface to control the left and right direction of the robotic arm. The hybrid inverse kinematics decouples the Jacobian matrix in the ground frame from the Jacobian matrix in the end-effector frame to achieve control using multiple reference frames. The performance of this novel system is verified by the implementation of a complex ADL task. Based on both quantitative and qualitative data, control using the Intuitive Reference Frame with the hybrid inverse kinematics (Hybrid system) dramatically decreased the time required to complete the ADL task. All users felt
that the Hybrid system was much more natural and easier for the robotic arm manipulation than the conventional methods.

The adaptability and usefulness of the developed theory were further tested with many other user input solutions. The results show that all tested user input solutions benefited from our newly designed human-oriented intuitive hybrid inverse kinematics without any UI modifications or hardware changes.

6.2 Future Work

Based on the feedback from the subjects in the smartphone interface test, users preferred One Button interface because it has only one button, and it eliminated the need to look at the screen. Users suggested adding haptic feedback when touching the buttons in Three Buttons and Tilt interfaces in order to eliminate the need for looking at the screen when touching the buttons. In the future, we will add haptic feedback by utilizing the vibration motor inside the smartphone.

The Hybrid system was designed solely based on the users’ feedback, but there are many other possibilities that may yield better results. For example, the Hybrid system is not always an orthogonal coordinate frame, and when the end-effector’s pose is near perpendicular to the ground, two degrees of freedom will be aligned with other DoF’s. In such a case, using pure Intuitive Reference Frame with conventional inverse Kinematics (not the hybrid inverse kinematics) may have better environmental adaptiveness, since the Intuitive Reference Frame is an orthogonal coordinate frame, and it can be used to control all six Cartesian DoF’s in any robotic arm configuration without compromising any DoFs due to vertical approach motion. However, in the future, we plan to find the vertical approach region that compromises the two DoF’s and create an automatic gradual switching mechanism between the hybrid inverse kinematics and the conventional inverse kinematics that uses only the Intuitive Reference Frame.
Our novel Intuitive Reference Frame with hybrid inverse kinematics can be useful for any velocity-controlled robot teleoperation system. It will be an interesting endeavor to use our theory with other user interfaces that were presented in chapter 2, when using traditional joystick based interfaces, Kinect camera based solutions, and many others. The performance and intuitiveness of these interfaces are expected to be improved when used with our Hybrid system. We plan to recruit more subjects with disabilities to test our novel interfaces and collect valuable feedback and suggestions from them. In addition, our Hybrid system may benefit broader applications such as telerobotic surgeries, space exploration robot manipulation, oil platform repairs, and nuclear facilities maintenance.
References


Appendix A: Copyright Permissions

Chapter 3 of this dissertation was published in IEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 5835-5841, Oct. 2020, doi: 10.1109/LRA.2020.3010453 [3]. According to IEEE author rights, the IEEE does not require individuals working on a thesis or dissertation to obtain a formal reuse license.