A new interaction framework for human and robot

Dinh, Quang Huy

2018

Dinh, Q. H. (2018). A new interaction framework for human and robot. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/75865>

<https://doi.org/10.32657/10356/75865>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**A NEW INTERACTION
FRAMEWORK FOR HUMAN AND ROBOT**

DINH QUANG HUY

**INTERDISCIPLINARY GRADUATE SCHOOL
MEDIA INNOVATION INSTITUTE @ NTU (IMI)**

2017

**A NEW INTERACTION
FRAMEWORK FOR HUMAN AND ROBOT**

DINH QUANG HUY

Interdisciplinary Graduate School
Media Innovation Institute @ NTU (IMI)

A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for the
degree of
Doctor of Philosophy

2017

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

.....
Date

.....
Student Name

Acknowledgement

First and foremost, I would like to give my special thanks to Professor Seet Gim Lee, Gerald, my supervisor, for your generous support during all these years, for your mentoring, for the experienced guidance, academic supports, patience and valuable advices you have given during my research journey so far, especially for helping me find a way back at the time of being distracted from my research duty. I could not have imagined having a better advisor for my Ph.D study.

I want to deeply grateful to Dr. Viatcheslav V Iastrebov for his enthusiastic cooperation and suggestions over the last 4 years when we implemented our project together. My work would have not been possible without your knowledge and experience. Furthermore, you are always a close friend to me and share many funny stories with me during my stay at RRC. I also would like to give big thanks to all my colleagues in Robotics Research Center: Wee Ching, Burhan, Hendra, Choon Yue, William, and Ahmad. You guys are always so nice to me since the first day I came to NTU. I especially wish to thank Ms. Wee Ching for sharing her knowledge and expertise at the beginning phase of my PhD project.

I wish to show my great gratitude to my families. Without their unconditional supports and encouragements in moments of difficulties, I would not have been able to accomplish this result. Finally, I would love to specially thank my girlfriend, Ms. Duong Kim Cuc for always being by my side, supporting my every decision, and tolerating my shortcomings.

Abstract

Despite technological advancement in the robotic field, designing a user-friendly robot to interact with human remains one of the most technically challenging problems for researchers. One possible solution is to design a robot that can mimic human appearance and behavior to interact with a human. Although the idea is very promising, the current technologies still fail to create such a robot due to the complexity of human's body and brain. This makes the users feel frustrated and annoyed when interacting with these types of robots. In fact, users normally expect humanoids to be more human-like than it really is, ultimately leading to disappointment and uncomfortable when the robot fails to be as intelligent. As a result, the future of humanoid is still unforeseen. On the other hand, the indirect interaction approach provides another way for us to solve the problem. The method's framework utilizes a mediating object which is normally a standard projector such as a video or a LED projector to facilitate the interaction process. However, this framework also contains some limitations that make it hard to be accepted as a standardized framework for human-robot interaction to be applied to different contexts. Firstly, current indirect interaction interfaces provide only one way to communicate with the robot which is via the standard projector. This prohibits the robot to communicate with multiple users with different access levels. Also, the current framework setup cannot guarantee the safety of the system information since the information is projected on the floor and can be modified by any user. Next, the framework can only be applied to a limited number of applications as a normal projector cannot generate bright and high-contrast images in bright or outdoor environments. Finally, the current framework only presents a single-modal input device for the user to send input commands or to interact with the mediating system. This not only can limit the number of applications the interface can be applied to but also reduce the reliability of the interaction process because the user does not have any other input models to perform a validation process in case the robot answers with inappropriate responses.

This proposal aims to identify a framework for indirect interaction interface that recognizes the deficiencies of current systems. To overcome the challenges of the current system, the new framework must be able to support multiple users with different levels of interaction. The mediating channels must be reconfigured to increase the security of exchanging information while allowing the robot to send feedbacks to the surrounding environment to inform people sharing the same working environment of its operational status. Furthermore, the input device must support multimodal input modalities to enhance the robustness of transmitting commands from human to robot which, in turn, allows the system to be deployed in different environments. To do so, the system must combine two augmented reality techniques called: see-through augmented reality and spatial augmented reality. Furthermore, a laser writer and wearable handheld device are specially designed to be included in the interface to facilitate the interacting processes. Finally, a dialog framework between human and robot is also introduced to allow "the human to say less and the robot to do more" during the interaction process.

List of Publications

Book Chapter

1. Gerald Seet, Viatcheslav Iastrebov, Dinh Quang Huy and Pang Wee Ching.(2016). Laser Graphics in Augmented Reality Applications for Real-World Robot Deployment. In Guanghui Wang(Ed), Recent Advances in Robotic Systems(Chp. 8). Rijeka: InTech.

Conference Presentations and Proceedings

1. Huy Dinh Quang, I. Viatcheslav and G. Seet Gim Lee, "See-through and spatial augmented reality - a novel framework for human-robot interaction," 2017 3rd International Conference on Control, Automation and Robotics (ICCAR), Nagoya, 2017, pp. 719-726. doi: 10.1109/ICCAR.2017.7942791
2. Huy Dinh Quang Quilong Yuan, I. Viatcheslav, G. Seet, "Augmented Reality Interface for Taping Robot," 2017 18th International Conference on Advanced Robotics (ICAR), Hongkong.

Table of Contents

Abstract.....	i
List of Publications	iii
Table of Contents.....	iv
List of Table	vii
List of Figure	viii
Abbreviations	xi
Chapter 1.....	1
Introduction	1
1.1 Human-Robot Interaction– An Overview	2
1.2 Augmented Reality (AR) for Human-Robot Interaction in Industrial Contexts.....	6
1.3 Thesis Scope and Objectives	11
1.4 Applications and Contributions.....	13
1.5 Thesis Overview	15
Chapter 2.....	16
Literature Review.....	16
2.1 Overview	17
2.2 Augmented Reality.....	17
2.2.1 Augmented Reality-History	18
2.2.2 Augmented Reality – Application.....	19
2.2.3 See-through Augmented Reality	21
2.2.4 Spatial Augmented Reality	22
2.3 Wearable Handheld Device.....	23
2.4 Main trends in Human-Robot interaction.....	24
2.4.1 Anthropomorphic robots	25
2.4.2 Indirect Interaction Interface	27

2.5	Limitations of Current Indirect Interaction Interfaces	32
2.6	From Human-Human Interaction to Human-Robot Interaction	33
Chapter 3		36
A New Indirect Interaction Framework		36
3.1	Introduction of Methodology	37
3.1.1	The industrial contexts	37
3.1.2	The Interaction Framework.....	40
3.1.3	Modules in The New Framework	43
3.1.3.2	Multimodal Input Module.....	47
3.1.3.3	Spatial Task Sharing Module	50
3.3	Human-Robot Interaction Dialog.....	52
3.4	Framework Discussion	56
Chapter 4		63
Experimental Methodology		63
4.1	Framework Implementation	64
4.2	The Task Suggestion & Task Visualization Module with See-through Augmented Reality Glasses.....	65
4.2.1	Hardware Overview	65
4.2.2	Task Suggestion & Task Visualization Programming.....	66
4.3	The Spatial Task Sharing Module – The Laser Writer	73
4.3.1	Hardware Design.....	73
4.3.2	Image Distortion Compensation	82
4.4	The Multimodal Input Module – a Wearable Handheld Device.....	87
4.4.1	Hardware Design.....	87
Chapter 5		97
Case Studies and Discussion		97
5.1	Case Study 1: Human - Mobile Robot Interaction	98
5.1.1	The Mobile Navigation Scenario	98
5.1.2	Procedure.....	98
5.1.3	Objectives.....	100
5.1.3	Results	100
5.2	Case Study 2: Taping Robotic Platform	107
5.2.1	Taping Robot Interaction Problem.....	107

Table of Contents

5.2.2	Procedure and Objectives	108
5.2.3	Results	112
5.3	Discussion.....	117
Chapter 6		118
Contribution, Limitations, and Future Works		118
6.1	Contributions	119
6.1.1	A New Interaction Framework for Human-Robot.....	119
6.1.2	A Task Suggestion and Visualization Module for lowering Human Workload 121	
6.1.3	A New Multimodal Input Module for Industrial Contexts	122
6.1.4	A New Spatial Task Sharing Module with Laser Projection Technology	123
6.2	Limitations.....	124
6.3	Future Works.....	125

List of Table

Table 1. Summary of Interaction's Purpose for different groups of users.....	40
Table 2: Users with their access privileges to the interaction modules	42
Table 3. A Comparison between out proposed framework with other interface	60
Table 4: Usability test timings	133

List of Figure

Figure 1. Workers wearing safety equipment to perform industrial tasks [1].....	3
Figure 2. Cerevo's Tipron - Projector Robot Rolls Video Around the House [2].....	5
Figure 3. Augmented Reality Application with Meta smart glasses [3]	7
Figure 4. An example of projecting digital information onto the real object [8]	9
Figure 5. A person interacts with his friend in the same space virtually, while remaining miles apart using a Hololens from Microsoft [9]	17
Figure 6. The Reality–Virtuality continuum [10]	18
Figure 7. Printed magazine ads with AR content [26]	20
Figure 8. Microsoft Hololens Developer Edition [29]	21
Figure 9. Some typical wearable devices designed for the industrial contexts	24
Figure 10. Some current typical humanoid robot from industry	26
Figure 11. General framework for indirect interaction interface [11].....	28
Figure 12. Pre-visualization of a robot drawing program with individual projector drawings shown; robot executing program; result (top left - bottom right) [49].....	29
Figure 13. An overview on the current indirect interaction interface	31
Figure 14. An industrial scenario with human and robot sharing work space	37
Figure 15. Participants of Human-robot Interaction	38
Figure 16. The Proposed HRI Framework.	41
Figure 17. Interaction Framework for the operator with the robot	44
Figure 18. Architecture of the TSTV Module. Task Suggestion and Task Visualization are Augmented Reality Applications running on an Android Operating System.	46
Figure 19. An Overview of the Multimodal Input Module Block Diagram	48
Figure 20. Multimodal Input Module with four types of input modality supporting the interaction between the operator and the robot in our proposed framework.....	49
Figure 21. The Spatial Task Sharing Module with the ability to share the robot's operation information with the passersby. This kind of information must be very concise and easy to understand for everybody.....	51
Figure 22. A hierarchical task analysis for our human-robot interaction framework	54
Figure 23: Flow chart for operator-robot dialog	55
Figure 24: An overview of our proposed framework for human-robot interaction	57
Figure 25: Extended framework for interaction with multiple robots.....	61
Figure 26: Illustrations of framework components and applications	64
Figure 27: Moverio BT-200 augmented reality glasses	65
Figure 28: Augmented reality programming steps.....	66

Figure 29: Marker-based tracking block diagram	68
Figure 30: Overlaying virtual objects over real scene.....	69
Figure 31: Primary components of the augmented reality application running on the glasses	69
Figure 32: An example of augmented reality secondary components	70
Figure 33: Main components of the laser writer	74
Figure 34: The structure of a D'Arsonval meter and the design of the galvanometer.....	75
Figure 35: Images that is going to be generated by our laser writer	76
Figure 36: Our block diagram for the laser writer design	78
Figure 37: The laser writer hardware component block diagram	79
Figure 38: Laser Writer Prototype	79
Figure 39: Laser graphics projected to the floor in bright and low-light condition	80
Figure 40: Laser graphics generation process	81
Figure 41: The relationship between the laser writer coordinates, the physical surface coordinate and the camera coordinate.....	83
Figure 42: The image is pre-warped by multiplying with $W = P^{-1}S$ in the laser writer before projecting to the physical surface for obtaining a rectilinear image. S is just a scale rigid body transform to zoom the original image in the constraint of the laser writer.	84
Figure 43: Camera -Projector Calibration Using Homography methodology with the corresponding points extracted from a project square.	85
Figure 44: Laser Input Detection and Arrow Projection.....	86
Figure 45: Existing popular single-handed wearable devices	89
Figure 46: Our proposed design for the multimodal input module	90
Figure 47: The multimodal input device with its gesture specification	90
Figure 48: Handheld device prototype model	91
Figure 49: The first prototype of our handheld device.....	91
Figure 50: The handheld device hardware block diagram	93
Figure 51: Data collected from using buttons and gesture mapping input models	94
Figure 52: The Mobile Robot Navigation Scenario	99
Figure 53: The interaction dialog between the operator and the robot from selecting the robot to choosing the manual navigation mode.	101
Figure 54: Human-Robot interaction dialog for controlling the robot in manual mode	102
Figure 55: Using gesture mapping and buttons to control the robot movements. The input modalities are chosen to make the operator comfortable during the controlling process ..	103
Figure 56: Control the mobile robot with different input model from the MIM and the corresponding information projected by the laser writer for sharing with the passers-by .	104
Figure 57: The demonstration of semi-autonomous operation mode with laser pointer input: (a) Semi-autonomous mode instructions. (b) Use laser pointer to choose a goal point. (c) Robot moves to the goal point. (d) Robot stops at the goal point and project its “FINISH” status to the floor.	105
Figure 58: The human-robot interaction in the automatic mode where the operator can use laser pointer as the input for identify the goal point for the robot's navigation task	106
Figure 59: The taping robotic platform	107
Figure 60: Proposed Taping Robotic Platform	109
Figure 61: Our implementation for taping robot system.....	109
Figure 62: Workflow of the taping algorithm	111

Figure 63: Workflow of square taping	112
Figure 64: Workflow of spiral taping.....	112
Figure 65: Workflow of circular taping	112
Figure 66. The interaction between human-robot for the taping task from selecting the taping method to identifying the starting point of the taping area	114
Figure 67. The interaction between human-robot for the taping task from identifying the endpoint of the taping area to highlighting the taping area on the workpiece	115
Figure 68. The operator is monitoring the taping task performed by the robot and the result	116
Figure 69. A New Interaction Framework for Human-Robot with all the participants and the supporting modules	119
Figure 70. The TSTV module with its functionalities for helping the operator during the interaction.	121
Figure 71. The multimodal Input Module with multiple input modalities.....	122
Figure 72. The Spatial Task Sharing Module using laser source for generating very high-resolution images for the robot the share its actions with the passers-by	123
Figure 73: Using our framework to control the cleaning robot on a ship hull.....	125
Figure 74: Multimodal interface with augmented reality in a construction site	126
Figure 75: Mobile robot navigation pilot test	132
Figure 76: Joystick Controller.....	133
Figure 77: Times taken by participants to complete given task.....	134
Figure 78: Demographic of respondents	136
Figure 79: Comfort of the two types of controllers.....	137
Figure 80: Control of the handheld controller.....	138
Figure 81: Sustained use of the handheld controller	139
Figure 82: Fit and visibility.....	140
Figure 83: Peripheral visibility and comprehension with glasses	141
Figure 84: Control of the robot	142
Figure 85: Overall choice of controller.....	143

Abbreviations

STAR	See-through Augmented Reality
SAR	Spatial Augmented Reality
HRI	Human Robot Interaction
HCI	Human Computer Interaction
AR	Augmented Reality
ILDA	International Laser Display Association
HMD	Head Mounted Display
GUI	Graphical User Interface
DAC	Digital to Analog Converter

Chapter 1

Introduction

Robots have been used to increase the productivity and reduced risks and costs for many manufacturing processes in various fields so far. As we are moving to the era that is recently called “Industry 4.0”, the uses of robots are constantly improved to be able to adapt to the new trend which takes advantage of human-centered approaches, allows higher interaction level and is customizable. As a result, human-robot interfaces need to be newly designed or redesigned to satisfy these new requirements. Human-to-human communication is a multimodal process. People use different kinds of input modalities such as voice, body language, vision or even writing for conveying information. The main advantage of this communication model is to help human clarify any confusions happening during the conversation by using more than one kind of input. For example, sometimes it is difficult to verbally describe something. In that case, we use our hand to point to the real object or showing similar pictures as visual references. Inspired by this biological interaction model, the thesis is going to propose and implement a novel framework for human-to-robot communication which allows human and robot to combine different kinds of input modalities to exchange industrial task information.

1.1 Human-Robot Interaction– An Overview

Robots have been deployed in the industry for many years due to their accuracy, strength and repeatability features. In many industrial corporations, classical robots normally take the role of mindless operators used to perform repetitive, fixed sequence tasks. This type of robot operation is widely used in many automation manufacturing processes such as car assembling, smartphone assembling or cargo delivering. In these industrial scenarios, for the safety of human, manufacturers employing robots prohibit humans and robots to share the workspace at any time. However, with the advances in technology and human's standard of living, we are welcoming robots to our surrounding environments such as our home, our school or our workplace. This will help us to combine their strength, speed, repeatability, and persistence with our intelligence and dexterity to complete tasks more efficiently. Consequently, many robots are designed to not only share the workspace but also share goals in terms of task achievement with a human. In the industrial contexts, with the incoming of the "Industrial 4.0", the human-robot systems are required to adapt to new changes and requirements. Firstly, the cost and time for programming a robot becomes very significant in the scenario of "high-mix, low-volume" production where industrial tasks vary in application, lot size, and production processes. These high-variation and low-volume tasks are normally tailored to satisfy customer requirements, have a short time to market and can lower inventory requirement finished goods. Thus, tasks required of the robot, in such an environment, can be expected to be different and non-repetitive. As many manufacturers are currently allowing their products to be user-customizable due to their client demands, this trend is becoming more and more popular in many industrial fields such as cars or smartphones, especially with small and medium companies. Without a large number of task repetitions to offset programming effort, the deployment of the classical robotic interface becomes less attractive.

This has resulted in a new wave for another generation of manufacturing robots which are designed and targeted for small and medium enterprises especially when robots are not popular in these companies. Hence, the human-robot systems must be able to help these enterprises to deal with tasks that vary and can be changed frequently. On

the other hand, industrial environment usually requires workers to wear protective equipment. As a result, using mouse or keyboard to interact with the robot is very difficult or even impossible. This problem will become more serious in case of “high-mix, low-volume” tasks since the main operator is supposed to interact more with the robots to reconfigure robot parameters or redesign the task strategy. Consequently, using traditional input devices such as mice, keyboards or touch screens would become inappropriate in the upcoming era of industrial manufacturing. Figure 1 describes a worker working in an industrial environment with while wearing safety equipment.

Figure 1. Workers wearing safety equipment to perform industrial tasks [1]

As “high-mix, low-volume” product will increase the complexity of the manufacturing robot due to the growth in product depth and breadth driven by discriminating customers, the interaction interface between human and robot need to be improved as well. Specifically, robot programming method must be redesigned to rapidly reconfigure task parameters so that the robot can adapt to the variation in the task requirements which could cost a lot of time using traditional programming methods. Also, more robots are deployed in a semi-structured manufacturing environment or domestic environments. These industrial robots are becoming more collaborative, in their interactions with humans and are designed to work with people in the same environment without the provision of safety enclosures. In consequence,

the system interface must be able to increase the awareness of people sharing the same working environment of the robot operation. This is a very important feature for modern human robot interface implementation since industrial working contexts normally require the highest level of security for both workers and robots. Finally, as human is becoming the center factor for the designing process, it is important that the relationship between human and robot is changed to enhance productivity and encourage wider application. In the past, a human is supposed to be responsible for programming all the robot operations which require the users to master all the details of the industrial tasks while the robot just obeys the operator's commands. This kind of relationship requires human to take the role of the "master" while the robot just follows the orders as a "slave". This model undoubtedly imposes a significant load on the human. In this new era of manufacturing, human-robot interaction needs to evolve from one of that "master-slave" relationship to that of a "master-partner" one. The "master-partner" labels the human taking a decision responsibility assisted by a robot offering assistance. This proposed paradigm requires the robot to augment task description, from its appreciation of the human needs and the context of the situation. Under such a scenario, there would undoubtedly be many alternative options resulting in a different ranking of constraints and outcomes. To enable a robot to deliberate on a set of instructions and to result in a unique outcome would require all the constraints and desirables to be explicit, defined. Ultimately, this new and complex interaction process between human and robot has encouraged robotic scientists to develop methodologies to design and build robots that are user-friendly, smarter and safer to interact with. With that desired target, inspired by human-human communication models, many robotic researchers try to build robots that can mimic human appearance and behavior called "humanoids". These robots can resemble human body shape and contain human-like features such as hand gesture, voices, walking. This category of robots is called "anthropomorphism". Their idea is that if we can build robots that can communicate with human as human, it would be easy for us to communicate and understand them which, in turn, can encourage us to interact with them more. However, anthropomorphic robots can be very deceptive since a robot is not a human and trying to make them human can lead to many undesired consequences. In fact, resembling human's features is a very complex and costly

process. Furthermore, if the humanoid fails to respond as a person during the interaction process, people will feel very annoying and depressed. It will be a long journey for humanoid robotic researchers to achieve their final goal and the future of this field is still unforeseen. Researchers also attempted to solve the interaction problem using a different approach called “indirection interaction interface”. Indirect interaction interfaces are designed to allow human to communicate with the robot via a mediating object which is normally a projector that can project the graphical user interface onto a surface. An example of this interactive interface is shown in the below figure which introduces a robot that can go around and project information to the wall.

Figure 2. Cerevo's Tipron - Projector Robot Rolls Video Around the House [2]

This indirect approach can minimize the misleading problems that happen between human and humanoid robots since the interactive information is directly projected to the real environment by the mediator. The indirect interaction method mostly focuses on improving interaction process itself and pays attention to how task's information is presented to both human and robot. This kind of interface is straightforward and inexpensive to implement which is more appropriate for small and medium industrial entrepreneurs. However, the current framework for indirect interaction is not fully developed and contains some limitations that prevent it from being accepted popularly for different industrial contexts. The limitations are mainly due to the limited capability of the mediating object, the lack of communication channels

between human and robot, the lack of a sophisticated input device for conveying information from user to the robot. Current mediating object for indirect interface is normally a LED projector which can only work properly in indoor setting and becomes very blurry in an outdoor environment. Also, since there is only one single communication channel between human and robot, there is a problem with information security as well as extending use of the interface with multiple users. These gaps in the current framework has motivated us to develop a more comprehensive model for indirect interface which allow this kind of technology to not only overcome current difficulties but also be widely accepted in many industrial contexts. In the next section, we will discuss one emerging technology that can help us to solve these problems.

1.2 Augmented Reality (AR) for Human-Robot Interaction in Industrial Contexts

Research has shown that human cannot be completely replaced by robot in many specialized manufacturing processes. In fact, the human role should be elaborated and emphasized with the helps from the robots. As a result, the interfacing system between human-robot need to be improved for adapting to recent trends of industry. This evolution of Human-Robot interface requires the integration of different technologies from various fields and Augmented Reality will take one of the key technology. Augmented Reality is an emerging technology in computer vision field that allows virtual information to be added to the real environment. The core idea is to augment our real environment with selected computer-generated information to enhance the efficiency of a particular task or process. The AR system usually utilizes a camera to capture the real environment. This information stream is then processed to track the desired real target and add the corresponding virtual graphical contents to the system's display. Consequently, the user then can view both real scene and virtual contents at the same time via the same screen. Figure 6 introduces an example of using a smart glass for creating virtual objects and aligning them with the real scene.

Figure 3. Augmented Reality Application with Meta smart glasses [3]

AR technology can introduce many advantages to the human-robot interfaces in the fourth generation of industry because of its ability to provide the augmented information visualization at the right time and at the right place. In many manufacturing contexts, information or knowledge is always a key factor. The main operators need expert knowledge of the industrial systems or robots to manipulate them properly. Also, they are required to master the technical details of manufacturing systems to simulate the task, configure the technical parameters, monitor the systems in operation, repair and maintain the system, etc. Therefore, there is a heavy workload on human while performing a particular task in a classical industrial context. However, this situation can be changed with the next generation of human-robot interface where AR is a core factor. The core functionality of AR is to reduce the complexity of many manufacturing process by lowering the workload for main operators. Indeed, AR can participate and shift the workload from human to robotic system in many steps relating to an industrial process. The methodology behind AR technology is to provide expert knowledge to users by seeing it appropriately visualized. AR is actually welcomed in many industrial tasks and processes. Firstly, AR can joint in preoperational stages of human-robot interaction tasks such as robot programming, task planning, simulation, prototyping and design. For example, AR technology is implemented to construct a virtual model of the robot which is then manipulated for geometric path planning process [4]. This robotic programming

method allows users to view and manually edit the planned paths before any real execution from the robot. The method has shown great improvements in term of efficiency and programming time for many industrial tasks. Furthermore, a dynamic simulation of a robot is very useful for various robot tasks. By overlaying a correct kinematic and dynamic model of a robot over the real working environment, the users will be able to understand and predict the motion of a manipulator with a set of given initial conditions [5]. This feature can make the human-robot system become very intuitive and user-friendly. Secondly, AR can be applied to real-time manufacturing process as well. In a classical context, the main operator can only monitor the system status via a computer or a teach-pendant display with a lot of system parameters. It is also difficult to visualize the current status of the robot or its next movements. These limitations can be solved using a wearable device with Augmented Reality functionalities such as smart glasses. By overlaying virtual robot parameters as well as the trajectory of robot's operation over the real scene, the users can instantly understand the status of the robot and observe its next movements which, in turn, can improve the safety level of the system [6]. Additionally, the monitor process is more effective as the main operator can halt robot operation immediately if necessary by checking the virtual robot's information aligned with the real robot on the AR device or if there are some differences between the planned path and the robot's execution path. Finally, AR can successfully apply to post operational stages for industrial tasks such as repairing or maintenance. By aligning virtual information to the operator's view, these processes can be simplified and optimized. In the past, some industrial repairing and maintenance processes are so complex that the engineer must read manuals while performing the tasks. Furthermore, the maintenance engineer is distracted from the main task due to the eye and head movements for reading the user manuals. Consequently, it is reported that hard manuals delay maintenance operations in industry [7]. This problem can be solved using Augmented Reality features. The key idea is to remove the hard manuals and adding the instructional information to the engineer's perspective. The augmented information can be the text descriptions of the maintenance steps, the technical descriptions of required components or a virtual animation for performing the step. The maintainer can just follow the information and complete the task without remembering anything. Some AR systems

even provide the real-time feedback for user's operation and display warning messages to correct the procedure. Of course, the virtual guidance can be customized depending on the experience level of the operator. Even if the process is performed by an experienced user, this augmented reality guidance is still very useful as it can take the role as a reminder. The operator now can perform the repairing or maintenance task just by observing the computer-generated information from the AR system while looking at the task scene. This cannot happen in case of using hard manuals or a computer with mouse and keyboard to interface with the robot as the operator has turn away from the task's scene to read or to look at separate screen for instructional information.

Augmented Reality technology is also extended to another form called "spatial augmented reality (SAR)". This type of AR manipulates a projector to generate and add virtual information to the real scene. The virtual contents now can be directly projected to the working environment instead of being displayed by a device such as HMD or smart glasses. An example of SAR is shown in the below figure.

Figure 4. An example of projecting digital information onto the real object [8]

This method has the advantage of providing intuitive virtual contents to many users while removing the inconvenience of wearing an additional device. Therefore, there are many interfaces that integrate this SAR technology for facilitating the interaction between human and robot. One promising category of applications involves human-

robot interfaces with SAR technology which allow the user to directly create, add or edit the robot's path with intuitive feedback provided by the SAR system. Additionally, this kind of technology can fit nicely in the industrial working environment where there is a collaboration between users and robot. In fact, this feature is very important for human-robot interface especially in contexts which requires a high level of safety. By endowing a robotic platform with the ability to project interactive information on the working scene, robot's feedbacks can be continuously projected on the shared environment. As a result, robot information can be shared with users working in the same workplace in the form of the robot's path or upcoming intentions which, in turn, can increase the people's awareness of robot's operation as well as safety level of the workplace. However, this kind of technology contains some disadvantages as well. As the augmented information is generated by a projector, the system can only generate 2D information such as texts or images which is not appropriate for applications manipulating complex virtual information such as 3D robot model. Next, since computer-generated information is projected on a real background or surface, the environmental conditions such as lightning, the color of the background or occlusions will affect the information quality. Also, the calibration process to align projected virtual contents with real contents becomes more complicated compared to our traditional camera-tracking method. Current SAR interfaces [11][12] utilize projectors that can only provide low contrast images which may be good only for a presentation at a conference or a meeting. It is hard to see their projected graphics in a bright or outdoor environment which make it hard to adopt these interface to an industrial context. Therefore, a new projecting platform will be designed and implemented in our project to solve the current the limitations of traditional projectors and provide functionalities for spatial augmented reality purposes.

In short, AR technology is apparently welcomed to the industrial environment especially in the scenario of human-robot interaction because of its capability to enhance the efficiency, cost, user-friendliness, safety level, productivity for many human-robot systems. As a result, AR technology should be integrated and take an important role in the next generation of human-robot interaction in the upcoming Industrial 4.0.

1.3 Thesis Scope and Objectives

As mentioned in the previous sections, there is a need to design and implement a new interface for human-robot interaction to satisfy new trends in the robotic field especially for the upcoming “Industrie 4.0”. There are basically two schools of thought in robotic interface design to solve this problem including “anthropomorphism” (direct interaction interface) and “indirect interaction interface”. While the former idea seems to be very promising and ambitious, it will be a long and costly process until we can achieve our goal. The latter approach aims for a simpler and low-cost solution. As a result, we have decided to focus on solving current problems relating to this kind of interface in this thesis in order to extend their deployment in many industrial settings in the near future. To achieve this goal, our thesis will deliver the following objectives. A detailed literature review needs to be performed to list and analyze the related indirect interaction robotic interface to elaborate on their problems. From these problems and the original framework, we will propose a new framework for indirect interaction interface which integrates three new components that able to help the framework to eliminate current problems. A new theoretical interaction model is also introduced as a reference for researchers to design and implement indirect interface systems in the future. Due to the change in human-robot relation in the near future, the model is supposed to consider robot as a collaborator who can instruct human at any steps of the interaction task. This lead to the possibility of developing a formal interaction model which shifts the workload in the new framework from human to robot. In such a scenario, the robot should be equipped with the ability to provide suggestions to human and which allows the user to finish the task without understanding sophisticated technical details of the system. This type of interaction will be essential for simplifying the task complexity for new users and increasing the friendliness of the new human-robot interface. Another important target of the thesis is the implementation aspect of the new framework. By combining some emerging technology in the robotic field such as augmented reality, spatial augmented reality, and multimodal handheld device, an implementation of the proposed framework is also presented. Also, as current interface focuses more on problem-specific applications, the new framework should be more general and can

be applied to different industrial scenarios. Finally, it is essential for the framework to consider the requirements of industrial environments which are very different with an additional focus on robustness, reliability, safety, and ease-of-use. Let discuss some requirements of industrial settings in details.

1. Robustness and Reliability

These features are very important for determining the high-quality of the overall industrial system. In fact, industrial interaction systems are required to perform its designated duties for a long time and can repeat the task for thousands of times without any errors and can produce the same results. Also, it is essential for the framework to provide human and robot abilities, tools, data, visualizations to deal with unexpected situations in the context of robot-human co-existence. Furthermore, the framework can become more reliable if it can deal with abnormal situations with minimal human involvement.

2. Safety

Safety is always listed as one of the key requirements for industrial processes. In the industrial environment, faulty robot actions can lead to severe outcomes. As a result, the framework must include the users the ability to monitor on-going tasks and to intervene as necessary for the purpose of task optimization, accommodating new contingencies or mitigating contentions. On the other hand, people sharing the same environment with the robot should be aware or notified of the robot operations or intentions. The provision of such secure interface could keep users devoid of unnecessary ambiguities or dangerous collisions. Information security is also another concern for a new framework to be applied for an industrial setting. In particular, the system is supposed to authorize different information access levels to different kind of user with real-time response.

3. Ease-of-use

Designing a user-friendly interface for industrial workers is very different compared to other settings. There are some constraints in the manufacturing sites that people have to strictly follow to get accessed to their workplace. As a result, it is essential to take into account these factors during the interface design and implementation process. For example, people are supposed to

wear personal protective equipment such as full body suits, safety glasses, helmets or gloves. Thus, the traditional use of keyboards and mouse devices can be very difficult on the industrial shop floor. On the other hand, the deployment of a handheld device with gesture recognition functionalities would be more practical and convenient. In addition, to make interaction system more accessible, it should be able to operate by all kind of users ranging from newbies to experts. Therefore, there is a need to consider the issue of ease of programming or robot as an assistant.

In summary, the thesis will propose a framework that can overcome current limitations for indirect interaction method. Our framework also envisages the evolution of HRI in the upcoming future and provides solutions so that user can be more natural and allow the robot to provide a greater contribution to formulating the desired outcome. This approach, invariably, requires the robot and human towards greater interaction in terms of frequency and amount of relevant information exchange. In an environment shared between humans and robots, our solution will help the human to be able to recognize the intention of robots and vice versa. This is to avoid the possibility of conflicts and accidents.

1.4 Applications and Contributions

From the theoretical perspective:

- This work analyzed previous works in the indirect interaction field to figure out the current limitations of the general interaction framework. The thesis proposed a new framework for indirect interaction interface to solve all the limitations. The framework introduces the combination of two emerging technology in computer vision called “Augmented Reality”. In fact, this work is also the first project to combine see-through and spatial augmented reality into a unified framework for human-robot interaction.
- The thesis also extends the new framework for the case of multiple groups of users interact with multiple robots with the help of tracking module from augmented reality technology. Additionally, a new human-robot dialog model is proposed to be incorporated into the new framework. This conversation model

considers robot as a partner of a human for the task. Using augmented reality wearable handheld device as the supporting tools, the robot can suggest task options to the main user. As a result, the operator does not need to know all the details of the task. The idea can help to reduce the workload for human while leveraging the role of the robot in the human-robot relationship.

From the implementation perspective:

- Design and prototype a multimodal single-handed wearable device for a user to provide input commands to the interface. This device can be considered as a replacement for mouse and keyboard in programming robot tasks. The hand-controller is specially implemented so that the user can select the most appropriate input model for a task. It is also safe for the industrial environment since the worker can use it with only one hand while wearing protective equipment.
- As one of the main component of this framework, we introduce a novel laser writer for projecting robot operating information to the surrounding environment such as the floor or the wall. The primary function of this laser writer is as a mediating object between human and robot. Robots can manipulate this system to display its configuration, task information or incoming operations. This kind of projector uses a laser as its light source which can generate extreme bright and high-contrast images against different backgrounds. However, due to technology constraints, these images are only outlines without any interior details. Nonetheless, it is very useful for human-robot interaction since it can be deployed in both indoor and outdoor environments which are normally cannot be accomplished by using a normal projector. This laser writer will be our core component using spatial augmented reality technology in our framework.
- The proposed framework is applied two different industrial tasks for demonstration purposes. The first application is to integrate the interface to an industrial mobile robot. The main user will be able to select the task, test the manual control mode while observing visual feedback from the robot. This robot will project its movement intention to the floor for safety purposes. The other demonstration relates to applying the interface system in the scenario of a taping robot. In this context, we suggest using see-through technology configure robot

parameters while using laser writer to overlay taping information over the real object.

1.5 Thesis Overview

The rest of the thesis is organized as follow:

Chapters 2 provides the state-of-the-art of the related technologies including Human-Robot Interfaces, Augmented Reality and Wearable Handheld Devices with their possible applications.

Chapter 3 presents our novel human-robot framework for the indirect interaction field.

Chapter 4 describes the hardware design and software implementation of the framework components.

Chapter 5 elaborates on the implementations and results when applying the proposed framework to two different robotic platforms and compares our results with the related works.

Chapter 6 concludes our thesis with our contributions and proposes possible extensions of the framework as part of future work.

Chapter 2

Literature Review

The chapter provides an introduction as well as the applications for two emerging technologies including Augmented Reality and Wearable Handheld. These technologies will take important roles in our proposed framework in the next chapter. This chapter will also investigate how human-robot interaction interface was done previously. The chapter will end with a discussion on the limitations of the current human-robot interface.

2.1 Overview

This section provides an overview of relevant works that will be useful for our project. Detailed descriptions of Augmented Reality technology and the wearable handheld device will be given since they are going to be an important part of our proposed framework. For each of the above technologies, a brief history is also delivered to demonstrate how these fields emerged and made progress. In the final section, we introduce and discuss in detail the topic of human-robot interaction as this is the focus of the thesis.

2.2 Augmented Reality

To understand the concept of “Augmented Reality”, we need to mention the concept of “Mixed reality” first. Mixed reality refers to “the merging of real and virtual worlds to produce new environments and visualizations where physical and digital objects co-exist and interact in real time” [8]. An example of mixed reality is shown in the figure below.

Figure 5. A person interacts with his friend in the same space virtually, while remaining miles apart using a Hololens from Microsoft [9]

In figure 8, both real and digital elements are combined to improve user-experience for the user. Mixed Reality also refers to the entire spectrum of situations that span the continuum between virtual reality and actual reality which was first introduced by Milgram, 1994 [10].

Figure 6. The Reality–Virtuality continuum [10]

In this case, mixed reality can include augmented reality, virtual world, and other mixed configurations. From the continuum, we can see that Augmented Reality is closer to the real world, as opposed to the virtual world. The idea of Augmented Reality is to augment the real world with some aspects of the virtual world. This is because “augmented reality users remain in the real world while experiencing enhanced virtually created visuals, aural, and feelings. Augmented reality does this by layering virtual information and/or graphics on top of a user’s view of a real-world scene” [13]. Augmented Reality history was considered to begin in the 1960s with the motorcycle simulator called Sensorama designed by Morton Heilig [14]. The major development of Augmented Reality, however, likely starts in 1998 with the first Augmented Reality conference the IWAR (the International Workshop on Augmented Reality). As a result, people consider Augmented Reality to be an immature research field and there is a lot of work to be completed before this technology can successfully blend into many aspects of daily life. In the next sections, we will consider the history of Augmented Reality, their applications and two different types of Augmented Reality.

2.2.1 Augmented Reality-History

In 1962, Morton Heilig invented Sensorama system, a multi-modal interface enhancing the user’s movie experience by providing sound, visual, vibrate and scent effect. In 1975, Myron Krueger introduced Videoplace [15] which is a computer graphic system that allows the user to control computer-generated object. One

important application of AR was developed in 1990 when the Boeing R&D [16] team designed a system that facilitates the manufacturing and engineering process. This is also the first time that the term “Augmented Reality” is used. In 1994, Julie Martin [17] created “Dancing in the space”, an AR theater production, which also acts as an interactive interface between actors and virtual objects. Augmented Reality technology made a big step in 2000 with the foundation of the Total Immersion Company, a leading company in AR technology. Also, ARToolKit was released by Hirokazu Kato [18] in this year. This toolbox allows the users to overlay the designed computer graphics over the captured video. From this point of time, AR technology gained a large amount of attention from researchers and companies. In 2000, the first mobile AR game was developed by the Wearable Computer Lab at the University of South Australia called AR-Quake [19]. Mathias Mohring et al., 2004 [20] introduces a see-through AR system on a cell phone. This system can detect the marker on 2D video and render the 3D object. In 2006, AR developers began to show their interest in applying AR technology to information searching. One typical example is the Nokia Mobile Augmented Reality Application [21] aimed to create a phone application which overlays related virtual information over user’s input images.

In summary, there are a lot of efforts have been done in the history of AR to bridge the gap between the real and digital world. However, more developments need to be accomplished before AR becomes an essential technology in our lives. The next section will provide some recent impacts of AR technology in education, medical solutions, entertainments, and industry.

2.2.2 Augmented Reality – Application

The visual sensor is the most important sense of humanity. Most people gain understanding and are impressed by only what they really see with their naked eyes. As a result, the AR approach naturally becomes an essential part of many applications for many aspects of our lives. Hughes et al., 2005 [22] created program called “MR Sea Creatures” which create 3D animation effects for all the creature samples that are displayed in the museum. By using a special device called “MR portal” the visitor can see all the “static content of the museum come to life”. Audio and visual channels

are two most important effects to guarantee the success of this project. “MR MOUT: extreme reality” [23] is another interesting project for soldier training process. The project manipulates a decorated room in cooperation with computer-generated environment to create the process called “extend the realism” in which the real and digital environments are blended to create a rich MOUT site. The user will need a Head-Mounted Display (HMD) to participate in the battle. The trainees can practice their shooting skill with virtual character generated by the computer in an AR urban combat scenario. The gaming industry is a successful example of applying AR technology. The Wii gaming system [24] allows gamers to physically enjoy the game environment. The player can be the batter and try to hit the ball with his hand rather than to control the character by some buttons. Clearly, it is more fun to play the game this way. A map guidance system is also another successful story. Reitmayr and Schmalstieg, 2004 [25] design a map system for a mobile device with many appealing functions. The application adds the computer-generated path guidance with the target location to the real scene images. By using the camera of the phone, the user can intuitively find the correct direction wherever they are just by looking at the phone display. Advertisement industry is the one benefits the most from AR technology since AR technology allows everybody wants to see the real product before they buy it. For example, the figure below shows an advertisement campaign from Vespa Corporation. The reader can scan the ad through an app on their smartphone or tablet and build their own custom scooter from all the options available, including colors, styles, and accessories.

Figure 7. Printed magazine ads with AR content [26]

In 2009, with the release of Star Trek movies [27], the producers use AR along with the film poster as a marker to let the viewer interact with some of the movie related things such as the crow, the weapons, etc. Although it is believed to be an immature field, AR has shown a great potential for education, medical solutions, and entertainment. In the future, this technology will take a role of a leading element in combining the digital world with the daily life. In the next section, we will introduce two main type of augmented reality as the combination of them will be a core component of our proposed framework.

2.2.3 See-through Augmented Reality

See-through augmented reality is one the most popular and important trend in AR field as it provides a seamless way with the best possible experience for the operator. The technology requires the use of wearable computer glasses which have transparent displays that can generate graphical images and allows users to see through it. The technology can feed the bearer live information with high-resolution virtual contents during your activities. Some special glasses even allow the users to interact with those computer-generated objects. In 2013, Google started selling a prototype of their wearable device before it became available for commercial market on May 2014. It was developed by X with the mission of producing a ubiquitous computer. Google Glass displayed information in a smartphone-like hands-free format. Wearers communicated with the goggles via natural language voice commands. In 2015, Microsoft introduced their first AR glass called “HoloLens” [28] in one of their events. They started selling their first development edition in 2016. This product is the first self-contained, holographic computer that enables users to interact with digital contents and manipulate holograms in the surrounding world.

Figure 8. Microsoft Hololens Developer Edition [29]

Epson announced their Moverio BT-200 glass in 2014 at CES. Its second version called Moverio BT-300 was announced in 2016. Many other see-through glasses are introduced by different manufacturers such as Sony, Snap or Vuzix. These devices are able to generate very good quality virtual images for a wide range of applications. Various applications have been developed using this see-through technology including medical guidance [30][31], vision enhancement [32], game and training [33] or remote collaboration [34]. With the advances of the supported hardware and software, the future of see-through augmented reality technology is very promising.

2.2.4 Spatial Augmented Reality

Spatial augmented reality (SAR) is a branch of augmented reality that utilizes a beaming device to directly project graphical information on the real world. One important advantage of this technology is the elimination of smartphone, tablet or other wearable devices which reduces the complexity of the system setup. Additionally, a projector is included in the system setup to project information to the real world. The core difference between see-through augmented reality and spatial augmented reality is the way the augmented information is generated. In the case of SAR, digital information is projected directly to the real environment instead of being rendered by a smart device's display. Spatial augmented reality has proved its usefulness in many industrial scenarios where spatial information is the key factor. Some spatial augmented reality interfaces are implemented and apply to several automatic processes in shipping industry such as stud welding [35]. In these tasks, with the help of the projected information, the worker can intuitively and precisely identify the position or the area to perform the task. SAR becomes even more feasible when it is combined with some mobile robotic platforms. In 2011, a robot for guiding people by projecting instructions on the floor or the wall [36] is introduced. The setup includes a projector and a camera that are installed on top of the robot. The project illustrates the robot instructions on the floor while the camera monitors the targeted user. In material transportation field, an automatic navigation robot with spatial augmented reality function is built to safely operate in a shared industrial floor [37]. The ability to project the vehicle intention on the ground in front of the robot is very useful in term of increasing human's awareness of robot operation. For example,

Michael D. Coover et al., 2014 [38] have proved that SAR has a great potential in providing robot navigation information to users. Also, SAR has improved the interface safety level as users can totally focus on their working scene while still being able to follow the instructions generated by AR glasses.

2.3 Wearable Handheld Device

The idea of developing of wearable device can be dated back to the day of early computers which are so big and heavy that reinforcement is needed for protecting the floor below. As a result, designers always try to reduce the size and weight of a computer while increasing the memory capability and processing power. Furthermore, portability is always an intriguing feature for many products, especially in the electronic field. Therefore, many companies manage to design and build mobile portable device since 1950 such as the Sony Typecorder, the Xerox Notetaker or the MCM/70 Microcomputer [39]. True laptops appear in the 80s and 90s with the release of the first Apple's portable Macintosh, Pocket PC, Grid Compass and the ThinkPad. The next generation of handheld devices is organizer which was initially inspired by the handheld calculator. One famous example of organizers is the PalmPilot [40] which was the first wildly popular handheld computer. Also, during this time, handheld computers and mobile phones were studied and researched intensively by many corporations. From 1999, smartphones became popular around the world and connected people together and with the Internet. With the release of iPhone and other Android smartphones, people can easily get accessed to their email and online data, make online payments or stream video easily. In industrial and manufacturing field, handheld devices are designed and implemented to help human to interact naturally with computer since there are tasks that require data which is hard to capture using traditional tools such as mouse or keyboard. One typical example of a handheld device for the industry is the CyberGlove [41]. The device is integrated with flex sensor technology and provides clean, repeatable and accurate hand motion capture data. Another handheld device that was introduced in 2013 is the Reactive Grip Razer Hydra Prototype which maps motion-controlled data from user to 3D computer environment. The prototype also includes four sliders that move up and down in your

hand as you grip the unit. For certain situations, like swinging a sword or flail, the system creates an impressively convincing sensation that could bring us one step closer to immersive virtual reality. The device has great potential in applying to other fields such as medical robotic surgeries. There are some ideas relating creating a new computer mouse that can change the way people providing commands to the computers. The Loop Pointer [42] is a typical example. This special device from HillCrest Labs uses similar motion-sensing technology to let users control an on-screen cursor with the flick of their wrist.

Figure 9. Some typical wearable devices designed for the industrial contexts

In short, many wearable handheld devices are being developed and introduced to be applied to various fields to facilitate the interaction between human and computer. However, most mentioned devices only contain one or two input modalities which make them appropriate for a specific application. For our project, our idea is to design and implement a wearable device that allows the user to program and interact with the robot in any industrial tasks. Nevertheless, these works provided us the important and state-of-the-art technologies in the field of wearable electronics to be included in our future multimodal handheld device.

2.4 Main trends in Human-Robot interaction

In the field of human-robot interaction, there are two generalized schools of thought when it comes to human-robot interaction design. One method, called

anthropomorphism design, is inspired by the tendency of people to attribute human qualities to non-human entities. Consequently, anthropomorphic robots are created to resemble the human features such as body shape, gestures, voice, walking or even thinking models. The second method, called indirect interaction design, focuses on manipulating a mediating object to indirectly interact with the robot. In this section, an overview of the pros and cons of both methodologies is provided as well as the reasons why the second approach is selected for this PhD project.

2.4.1 Anthropomorphic robots

Human-robot interaction is a challenging issue in modern robotics. One main target of a human-robot interface is to allow human to define control input for the robot, monitor the state of the input and receive robot's feedback. As a result, a design of a robot that satisfies this requirement can greatly improve the quality of human-robot interaction. In order to achieve the target, the very first idea for many scientists is to design robots that mimic human appearance and behavior. This is due to a very natural argument. If we want to create a robot that can do as many things as we want as well as easy to communicate as we expect, we should make it as humanlike as possible. Furthermore, human is considered to be the product of millions of years of evolution and is perfect in term of intelligence and motion gesture. These reasons lead to the ultimate goal for many researchers in anthropomorphic robot field which is to create a humanoid robot that can replace human completely. This is, of course, a very long, challenged and costly journey for robotic researchers. It also requires the collaborations between scientists from different disciplines such as psychology, artificial intelligence, mechanical engineering, human-robot interaction or even human anatomy. Anyway, let take a look at two famous humanoid robots that our scientists are able to build so far which are Atlas from Boston Dynamic [43] and ASIMO from Honda [44]. Atlas is one of the most advanced humanoid robots that contain a control system which can coordinate motions of the arms, torso, and legs to achieve whole-body mobile manipulation. Therefore, Atlas robot can balance itself with a small footprint while performing tasks or walking on different types of terrain. On the other hand, ASIMO stands 130 cm tall and weighs 54 kg. The robot is famous for its ability to recognize human voices, faces and to perform human-like postures

and gestures. There are two cameras attached to the robot's head which allow it to capture the movements of multiple objects, determine distance and direction using visual information analysis. This information forms the camera can help ASIMO to follow or to face an approached human. An amazing function of the robot is to answer to questions by nodding or providing a verbal answer in different languages. With no doubt, the humanoid robotic design will be an intriguing field for many researchers and corporations to join and fulfill the ultimate dream of humanity.

Figure 10. Some current typical humanoid robot from industry

However, in a closer look, there are a lot of problems to consider for designing this kind of robot. One main argument for not using anthropomorphism in interface design is that anthropomorphic interfaces are deceptive and misleading. The idea is to build robots that can act like a human, however, robots are not human and in general, they are not very good at pretending to be human. Consequently, if we try to design a robot that pretends to be one of us and is not very good at it, people will get frustrated and annoying while interacting with the robot. This can be explained by the study [45] which found that users expect humanoids to be more human-like than they really are, ultimately leading to disappointment and discomfort when the robot fails to be as intelligent.

Another problem with anthropomorphism is relating to human factors that the robot designers are trying to resemble. Human behavior is very complex. For example, the

interaction model between human to human is a multimodal process. People use different kinds of communication inputs to convey information such as using voice and gesture to describe an object to other. In case the listener is unclear about the object, drawing a draft to illustrate the shape of the object is a possible solution. The example of human interaction model has shown that trying to mimic human features is very complicated and can become misleading easily. Even human gets confused sometimes talking to each other. The situation can get worse if the user cannot understand the humanoid robot's feedback and it is impossible for the robot to respond in a different manner due to technical limitations. To design a robot that is able to describe an answer using different communication modalities is a big problem. Finally, designing anthropomorphic robots is a very time-consuming and costly process which normally cannot be afforded by small and medium entrepreneurs. In fact, the history of humanoid robotic can be dated back to 1970 when Waseda University initiated the WABOT [46] - the world's first full-scale humanoid intelligent robot which was able to walk, communicate, measure distances and directions, grip and transport objects with hands. However, none of the existing anthropomorphic robotic platforms can achieve the human-level dexterity yet. While humanoid robots are developed in many research laboratories, their advent into the consumer market is still unforeseeable.

2.4.2 Indirect Interaction Interface

As the success of humanoid robots is not guaranteed, researchers must investigate another approach for interact with the robotic platform. This is when indirect interaction becomes a feasible method in HRI. The idea is to utilize a mediator between human and robot to exchange information. Such system can be implemented by mounting a projector on a robot. This projector can use the surrounding environment such as floors, walls or surfaces for displaying purposes. Figure 11 illustrates the general framework for an indirect interface for human-robot interaction.

Figure 11. General framework for indirect interaction interface [11]

While the mediating object can be attached to the robot, the user can be equipped with different type of input models to interact with robot including wearable handheld device or gesture recognition tools. The approach can help researchers to alleviate most core problems relating to human factors while focusing more on the user experience by enhancing the communication channel between human and robot. This interaction method is also very flexible in term of accommodating different kinds of users, tasks or interaction models. The idea of a robot with a projector was initially inspired by the movie “Star Wars” released in 1977 where robot R2D2 [47] can project information to the wall or the floor to communicate with a human. In 2006, Takafumi Matsumaru [12] introduced the first mobile robot with preliminary-announcement and display function for its incoming motion. The mobile robot was equipped with a projector that allows it to indicate its forthcoming operations to the people working in the surrounding environment. Basically, the robot communicates with human via the projected frame that is generated by the projector. A questionnaire evaluation for the robot was implemented with 200 visitors at an exhibition and the result showed that the robot gets 3.9 points and 4.5 points on the scale of 5 for how easy the participants understand the direction and speed of motion that are displayed. Overall, the evaluation showed that the robot is ineligible in general. In 2009, Jongkyeong Park and Gerard J. Kim [11] build a robot using the indirect interaction method that allows the operator to use an ultra-mobile PC or a laser pointer to play a chess game with the robot via an augmented board that is projected on the floor by the robotic system. Takuya Sasai et al, 2011 [36] using the proposed framework for indirect interface to design a guide robot for guiding visitors. The robot can project information anywhere in the surrounding environment while the user can provide commands to the robot by applying foot motion gesture to the dialog box generated by the robot and projected onto the floor. The system can detect the nearby users and

guide them by projecting moving directions onto the floor. In 2012, Ju-Hwan Seo et al [48] developed a similar robotic platform to interact with a human by using a system that consists of a beam projector and a laser scanner. While the projector generates images on the floor, the laser scanner data is processed to recognize user's foot motion to determine user input 's commands. The concept of indirect interaction is extended to other types of robots which is shown in the work by Florian Leutert et al [49] in 2013. They developed a projection system that consists of a fixed as well as a mobile projector mounted directly on the manipulator. The system enables the users to look at the pre-visualization of the robot drawings program by directly projecting the image of the drawing onto the canvas before the manipulator executes the drawing process. Finally, in 2015, Tomonobu Noguchi et al [50] studied a behavior model for information support with the ubiquitous display where they developed a system containing a mobile robot and a projector with pan-tilt mechanism. The robot can provide appropriate information for multiple using depending on user's interests and situations.

Figure 12. Pre-visualization of a robot drawing program with individual projector drawings shown; robot executing program; result (top left - bottom right) [49]

Also, in the same year, a new generation of an autonomous vehicle for transporting material [37] in industrial contexts is introduced. The robot can freely navigate in a shared working environment with workers so that it can provide flexible services such as loading and unloading of materials at a priori unknown positions. As navigating

freely is unpredictable and dangerous for people, the robot is equipped with the projector module for projecting its intentions on the ground plane in front of the robot. With this function, the robot will become more reliable to human which, in turn, increase the chance of accepting this kind of technology at the workplace. From the above examples, indirect interaction design method its great potential to be applied to a human-robot interface in various field. The design method is simple and easy to implement which is affordable for many small and medium entrepreneurs. There is also a study by Kwon and Gerald [51] showing that although people still felt that anthropomorphic robots are more user-friendly and emotional, the indirect interaction interface is more superior in term of understandability which is a very important criterion to evaluate the successfulness of designing a human-robot interface. Most subjects from the study also point out that the display content of the mediating objects in the indirect interfaces could be designed better to improve the level of friendliness and emotion for users while interacting with the robot.

Figure 13. An overview of the current indirect interaction interfaces

2.5 Limitations of Current Indirect Interaction Interfaces

All the related works mentioned follow strictly the idea of the indirect interaction framework where a projector takes the role of the mediator between human and robot. This has led to one major limitation for these interfaces since the quality of projected information depends greatly on the surrounding environment of the system. In fact, all the mentioned interfaces utilize standard LED projectors which are normally only good for meeting room or a cinema room or other indoor environments with low-light conditions. As a result, it is difficult or even impossible to extend the use of an indirect interface to other industrial tasks or another field. This has motivated us to implement a unique device that is able to replace the use of the standard projector in the current framework of indirect interaction interface.

Additionally, there is only one way in the framework for current indirect interaction systems to communicate with a human that is to project information onto the wall or the floor. This interaction model works well for tasks with only one main user. The model is not appropriate for the case with multiple users and some users have different interaction purposes with the robot. For example, there are some users who are directly working with the robot and they need to access the configuration of the robots while other users only share the same working environment and they just need to know the robot's intentions to avoid any potential collisions. Therefore, there is a need for improvement in term of communication model in the general framework so that the system can divide the robot information into different channels for different kind of users. Also, projecting all system information to the floor is not a good idea for industrial robots since unauthorized people can also get accessed and change the configuration of the systems. This explains why a new communication model also very essential for increasing the security level if we want to deploy this kind of robotic system in an industrial environment where security is a key factor.

Finally, most of the investigated projects use a single input model to control the robotic system such as foot's motion, pen, laser pointer or mobile personal computer. This is because these projects are implemented for a specific task and one single input model is enough to complete the task. However, for a general industrial task, one kind

of input is not enough to perform the task. The problem becomes even worse if we want to apply the system to different industrial tasks. For example, you might need to use the laser pointer to identify the object you want to robot to pick up while gesture recognition system is needed to confirm the picking process with the robot. Furthermore, multimodal communication is proved to be more reliable and less error-prone. For example, it would be frustrated for a human during the interaction process with a robot if the robot interprets wrongly the command from the user's gesture. It would be reasonable if there is a more reliable and tangible way for a human to confirm this command with the robot using another input model such as by pressing a button in this situation. Consequently, it is also important to design and implement a multimodal input device that allows human to send different kinds of input models to the robot via the interface to increase the flexibility and reliability of the whole system.

In this chapter, the history, as well as some typical examples of Augmented Reality and Wearable Handheld Device in the current literature, is provided. Two different approaches in the field of human-robot interaction are investigated in detail with some discussions and comparisons between the strengths and weaknesses of these design methodologies. The chapter ended with the current limitations of the indirect interaction method. In the next chapter, a new framework is proposed for indirect interaction interface. Also, the reasons for choosing Augmented Reality and Wearable Handheld Device as solutions for improving the current framework implementation will be discussed.

2.6 From Human-Human Interaction to Human-Robot Interaction

As robots are becoming popular in our environment such as our home, our school, our workplace, it is important to design a robotic system that not only is able to complete the task but also can communicate effectively with a human. In the last two decades, the main target for human-robot interaction field is to develop competent communication approaches. As a result, more intuitive user interfaces are designed that features advanced humanlike modalities such as face-to-face communication,

voice recognition or gesture. The interaction technology has developed far beyond the traditional limited human-robot interface for the industry. These advances in the robotic field are partially enabled by the studies from human-human interaction field. In fact, many studies, reports or designs for human-robot interaction are inspired or adapted from the corresponding models from the human-human communication theories. These studies [52][53] have shown that human-human interaction framework is very helpful and beneficial for human-computer and human-robot interaction. These works also conclude that many human-human communication models are actually repeated when humans interact with a robot or other intelligent agents. This can be explained that people communicate with the robot using their basic communication abilities and it is very difficult for them to develop a new form of communication. Though robots may communicate differently, human always tries to apply their developed communication approach to understand the interaction process with these artificial entities. These studies have established a foundation for us to consider the communication models of human-human interaction to apply and improve the current indirect interaction interface. From this foundation, using human-human interaction models to apply to human-computer or human-robot interaction is performed by many studies such as [53][54]. The procedure includes the following steps in sequence. First, a model for human-human interaction is generalized or selected. Next, the model is investigated and implemented to human-robot interaction. Finally, the model is tested by allowing the human to interact with the implemented robot. Although the approach can be applied to human-computer and human-robot interfaces, human-robot interfaces contain more intricate features than human-computer interface which we have to consider before performing any designing process. The study by Gerard Rigoll [55] showed that human-robot interface is a much more complicated problem which requires research in multidisciplinary areas such as Automation and Control, Artificial Intelligence, Mechatronics, Optimization, Psychology, Sociology, etc. This is because of the fact that robot is a complex mechanical embodiment. It consists of many components that can get involved in the interaction process which does not happen in the case of Human-Computer Interaction. This “embodiment effect” [56] may lead to many problems that need to be considered carefully if a human-robot interface is going to

be established. There are factors relating to the robot embodiment's attributes that may affect the quality of the exchanging information such as the distance between human and robot. For example, as the robot can move or walk, the distance or the angle of communication between human and robot will determine the quality of the information if speech is used as the communication signal which does not happen in the case of human-computer interaction. Another important point is the cooperation process between human and robot. Robots are much more complicated in term of mechanical functionalities as well as directional movements. Therefore, the robot not only can exchange information with a human but also can participate and help the user to complete a physical task. This implies that robot can form much complicated physical cooperation with a human in different applications and contexts. Other issues relating to human-robot interaction are the social acceptance as well as the psychological behaviors of the robot. In short, by transferring our findings and understandings of human-human interaction process to human-robot context, we can make some progress in the field. However, designing a framework or a system for human and robot interaction is a sophisticated process which requires many considerations from different disciplines and involves many different modalities. This background knowledge is very important for us during the design and implementation of our interface.

Chapter 3

A New Indirect Interaction Framework

In this chapter, a new framework is then proposed and presented for human and robot in industrial settings. The proposed framework is also extended to other interaction scenarios which may contain multiple robots and multiple users. A communication model for human and robot is also introduced to help the new framework adapt to new interaction contexts which consider human and robot to be partners.

3.1 Introduction of Methodology

3.1.1 The industrial contexts

Before describing in detail our proposed framework, let investigate the scenario of an industrial context where human and robot share the same working environment and collaborate to perform an industrial task. The following image illustrates a typical manufacturing context with different kind of users.

Figure 14. An industrial scenario with human and robot sharing workspace

In previous work, most human-robot interaction systems are designed for a specific application, in which, the main operators and the robot are the key factors. Our framework, on the other hand, takes this factory model into account and implements the idea of robots becoming human's partners and working together in a more general perspective. In this context, there are different kinds of robot deployed for different purposes. Some robots are fixed in limited locations while other might share the same floor space with human as shown in figure 26. Robots are controlled by human or they can operate autonomously. Finally, we have many users working in the same environment and may interact with these robots. Indeed, three different groups of people can be identified as shown in the figure below. They are grouped according to

their roles, their level of interaction relating to the robot actions, and on the nature of the information, they may require. The detailed information for each group is provided as follow.

Figure 15. Participants of Human-robot Interaction

Operator (Group I) – the human who is responsible for the control and supervision of the robot. Their interests are directly related to the deployment of the robotic platform such as to preprogram the robot, to simulate the industrial task for verification purposes, to be able to monitor robot's operation and to quickly halt the robot's operation if necessary. They can control or collaborate directly with the robot to perform or to complete a specific industrial task with the highest level of administration. Therefore, they are supposed to have expert knowledge of the robotic platform as well as the industrial process in order to manipulate the robot properly. They will be authorized to access robot system configuration to modify robot's parameters, task's specification for programming and controlling the robot. Also, the main operator is always required to fully aware of the robot's status when it is in operation mode to make sure it is performing as expected and satisfy all the safety procedures.

Bystanders (Group II) – the humans who are interested in monitoring the tasks being executed. Their interest might be to learn and understand the robot's operation or to discuss, cooperate with the main operator to solve a problem relating to the robotic system. Therefore, the bystanders would be able to access robot's configuration, specification or system's parameters during the collaboration process. However, they might not be able to change or to modify the system configurations. This feature is to differentiate the bystanders from the operators. Additionally, the bystander's work should be under the operator's supervision due to safety procedure. As a result, this kind of users has a medium level of information access to the system. A typical example for the bystanders is the new worker who wants to learn how to control and program a robot from an expert.

Passerby (Group III) – the people who are in the vicinity of interaction, but who are not directly related to, or interested in, the task being executed by the robot. The interest of this group of passersby arises from the sharing of common space and the need to accommodate the motion of the robot. Predominantly, the interest may be restricted to one of avoiding the robot and its workspace. Their interest is in the near-term actions of the robot as in the robot's current actions or in its next action. Therefore, this group of users has no interest in understanding the detailed operation, programming or controlling of the robotic system. Additionally, it is essential to prevent them from getting access to this kind of information because of safety reasons in a sharing working environment. On the other hand, because of the passersby's interest, they should be informed with the current operation status of the robot to understand and predict the robot's next movements or actions. This kind of information should be clear, simple and easy to understand as we need the action performed by the passersby in real time. Finally, it would be safe for the robot to detect and recognize the actions of the passersby from a distance to increase the safety of the interaction process. Due to their nature of the interaction, this group of users has the lowest level access to our system authorization.

In short, because each group of users has different ways and purposes to interact with the robot, the proposed interface framework is supposed to support these interaction demands by providing different interactive modules. Depending on their nature of the interaction, the contents of the information should be properly designed to maximize

their performance while sharing their work with the robot. Furthermore, while providing a user-friendly interface is an important objective, the new framework needs to assure that the requirements of the manufacturing environments are always fulfilled. The following table summarizes our framework's participants with their roles and objectives

Table 1. *Summary of Interaction's Purpose for different groups of users*

Agent	Operator	Bystander	Passerby	Robot
Objective	<ul style="list-style-type: none"> - Program the robot - Control and supervise the robot - Monitor robot's operation 	<ul style="list-style-type: none"> - Monitor robot's task - Learn robot's specification - Collaborate with the operators 	<ul style="list-style-type: none"> - Understand robot next action - Avoid the robot 	<ul style="list-style-type: none"> - Perform the task - Collaborate with the operator - Inform and avoid the passerby
Interaction level	High	Medium	Low	High

3.1.2 The Interaction Framework

Using the illustrated model and the limitations the current indirect interaction framework for human and robot in industrial contexts, we proposed our framework which is demonstrated in figure 28. In this framework, we introduce the integration of three communication models with different level of information security including the *multimodal input module (MIM)*, the *task suggestion & visualization module (TSVM)* and the *spatial task sharing module (STSM)*. The division among these modules is based on the targeted group of users and the security level of information that the module will support. These modules utilize various advanced technological tools to control and optimize the flow of information between human and robot. The main objective for this communication model is not only to facilitate the interaction process for all the participants in the framework but also to increase the user-friendliness as well as the security level for the interface within the industrial constraints. Depending on their purpose of interaction, each group of users can get access to the appropriate interactive modules. These modules with their corresponding components are illustrated in the below figure.

Figure 16. The Proposed HRI Framework. The framework consists of three interaction modules – multimodal Input, Task Suggestion and Visualization, and Spatial Task Sharing modules. The distinction between the three modules lies in the interaction purposes. The multimodal input module is responsible for receiving multimodal inputs from the operator. The interaction attributes in the TSTVM and STSM are more user-oriented. For more details, see section 3.1.3.

Firstly, only the *operator* is equipped with and get accessed to the *Task Suggestion & Visualization Module* and the *Multimodal Input Module*. This combination of interactive modules is to warranty that only the operator is permitted to program, control and modify robot action as they are the users with the highest authorization level to access the system. On the other hand, the *bystanders* could be provided with the *Task Suggestion & Visualization Module* for monitoring and learning purposes. They can monitor the processes of task configuration, task planning, task programming as well as collaborating with *the operator* via this channel. They would, however, not be permitted to control the actions of the robot or to modify any element of the main system, as control is only permitted through the *Multi-Modal Input Module*. Without the Multimodal Input Access, they would only be able to share in the notifications by the robot through the *Spatial Task Sharing Module*. This restriction in term of module access provides a clear differentiation between the two groups. Finally, as for the passersby in the vicinity of interaction scenario, but who are not directly interested in, the task being executed by the robot, they need to be provided with the ability to identify the robot's actions or its intention. The objective can be accomplished with the *Spatial Task Sharing Module*. By attaching this module to the robotic platform, the module can take the role of a communication channel for the robot to generate visual prompts for increasing the situational awareness for by this group. The following table summarizes the participants of the framework with the related interaction modules provided in the proposed interface.

Table 2: Users with their access privileges to the interaction modules within the framework.

	<i>Operator</i>	<i>Bystander</i>	<i>Passerby</i>
TSVM	X	X/O	O
MIM	X	O	O
STSM	X	X	X

In the next section, we will discuss each of the modules in the framework in details as well as their interaction with agents in our framework.

3.1.3 Modules in The New Framework

3.1.3.1 Task Suggestion & Task Visualization Module

In the past, communication between human and robot is not as important as today since classical robots only perform the task that is preprogrammed and barely interacts with a human. With the developments in technology and the new requirements of environment settings, it is very critical to create an effective communication modality between human and robot now. One of the most important objective for human-robot interaction is to reduce the workload for a human agent while leveraging the robot to take the role of a human's assistant. To achieve this goal in the manufacturing arena, an interface supporting a model of human-robot interaction with task/responsibility must be developed. The idea is to continue to utilize and refine the existing technology, incorporating new emerging technological solutions for user-defined tasks. One approach for solving this problem in the state-of-the-art is to encourage the development of collaborative solutions for human-robot interaction demands. This includes methodologies to enhance the capabilities, add advanced functionalities and to provide safety tools for adapting to the industrial context. In the scope of our framework, this problem will be solved using our *Task Suggestion & Visualization Module* which is specially designed for only the *operator* with the highest authorization level as well as most advanced functionality.

Firstly, the most important role of the module is to allow the operator to define the tasks for the robot. This task defining process will be performed in the form of a *task suggestion* process where the robot keeps providing the *operator* with suggestions to choose. The operator will then make a confirmation input and wait for feedback or other suggestions from the robot. The robot will keep this process until enough data is obtained to perform the task. This can reduce the workload for the human during the interaction with the robot as the operator does not require to remember or to master any step of the entire task as the detailed information will be provided by the robots. Additionally, by allowing the most secured information of the robot's operation to be transmitted via this module, the communication can be safe as only the operators and the robot can join this kind of interaction. The following figure

illustrates this functionality of the framework with detailed descriptions for the operator and the robot.

Figure 17. Interaction Framework for the *operator* with the *robot* using TSTV module and MI module. The *task procedure* will be generated by the robot's *Interaction Model* and sent to the *task suggestion* feature integrated in the TSTV module. The operator utilizes the TSTV and MIM module to make confirmations and define the task.

Secondly, *task visualization* is another important feature for supporting the communication process. This module is initially designed for displaying human-robot dialog after receiving the suggestion model from the task suggestion module. Also, as the framework will not require the *operator* to have expert knowledge of the interface system, the ability to instruct users to perform a task with the robot is very essential. By allowing the instructions to be visually displayed and overlaid properly over the working scene, the task will become simpler to understand and accomplish by the *operator*. Furthermore, *task visualization* is a great tool for performing task simulation or providing feedbacks from robot to human. It would be safer and more efficient if the *operator* can see and verify a simulation version of the task's trajectory or robot's actions before the real execution. This can be done using the *task visualization* function. Finally, the module is capable of feeding feedback from the robot to the operator. This function is very important as it was mentioned in the principle of closure [52] in human-human interaction field, people always try to generate the mutual belief that they have accomplished the goal of the current task after cooperating with other. This means that they are looking for the subtle feedbacks from others to confirm that the communication process is successful. In a similar manner, feedback is very important for human-robot [52][53] interaction as it is very

confusing for the user if there is no feedback from the robot after the user's action is performed. The role of robot's feedback becomes essential or even compulsory for many interaction schemes as the cooperation between human and robot is becoming more complex. For example, let's imagine a mobile robot navigating around an industrial factory, it is very important that this robot can indicate its feedback to the surrounding people to increase their awareness of the robot's operation. Without robot's feedbacks, robot's operation might become riskier in a shared working environment and a system failure can lead to catastrophic results. Indeed, one study even shows that feedback is one of the most important factors to evaluate the successfulness of a human-robot dialog [53]. Therefore, the ability for generating feedbacks via the *task visualization* module is very essential.

From the state of the art technologies, we recognize that see-through augmented reality is a great enabling solution for this module. With the ability to overlay computer graphics over an identified scene, this technology can provide our system a safer way to exchange information between human and robot since only the operators with the registered smart glass can view and interact with this information. The capability of adding computer-generated information to the real environment is very useful for task planning, task suggestion, and task simulation processes which are what we need for our TSTV module. Visual feedbacks from the robot to the human can become very intuitive and easy to understand for users without expert knowledge by using 3D rendering feature with computer vision engine. Furthermore, many studies [52] have shown that see-through augmented reality can be used to create a collaborative environment for human-robot interaction or for multiple users to share and interact with the same system which can be applied to our scenario where many *operators* and *bystanders* can discuss about the task without sharing this kind of information with the other group of users. Another advantage of see-through augmented reality technology is that it helps industrial operators to focus on the working scene while performing the task using the see-through feature. By wearing a transparent smart glass, the users can modify an ongoing task while keeping their eyes on the whole process instead of looking at the computer screen like what they did when using the traditional computer with mouse and keyboard. In fact, the technology not only provides an interactive and intuitive communication channel but

also increase the safety level for the operator which is very important for an industrial environment where safety is a big concern. The following image describes the overall software architecture of the augmented reality technology used in our proposed framework.

Figure 18. The architecture of the TSTV Module. *Task Suggestion* and *Task Visualization* are Augmented Reality Applications running on an Android Operating System.

A wearable see-through, transparent display is provided to allow the operator an unobstructed visual awareness of the environment. By taking advantage of the *computer vision engine* such as *2D tracking* as well as the *Augmented Reality API*, an *AR Application* can be created to perform the functionalities of the TSTV module. As a result, the task procedure received from the robot can be visualized and transformed into the human-robot interaction model. The model will be displayed using the functionality allowing for the projecting of high-resolution dialog actions, proposed by the partner-robot. In addition, relevant information required by the operator to allow for timely intervention would also be displayed such as task information, robot's operation status or user's commands.

In short, the task suggestion & visualization is one of the most important modules in our framework which acts as a secure communication channel for transferring information between the *operator* and the *robot*. This module can leverage the role of the robot to become a human's assistant while reducing the workload for the users.

3.1.3.2 Multimodal Input Module

As part of the interaction framework, we introduce the *Multimodal Input Module* which is only registered for the *operators* to interact with the robot. The motivations for this module come from the current difficulties of industrial workers who must wear protective equipment while programming or controlling the robot. In such a scenario, the use of traditional input devices like mouse, keyboard or teach pendant is very uncomfortable and not user-friendly. On the other hand, the use of a multimodal model with tangible inputs such as buttons or hand gesture could be more natural and favorable. However, a multimodal input model does not only mean the number of input modalities. It is the use of different input models to enhance the perceptual information received by the participants in the dialog. In fact, many studies [57][55] shown that multimodal input device provides many advantages for human in the interaction with the robot. With the ability to switch seamlessly across the different input modalities, the users can select the most fitting type of input to express their intention which, in turn, can reduce the ambiguity problem for human-robot communication. In the robotic field, this interaction model also can be applied for changing in the task configuration in an easy and fast manner without restarting the whole process. Furthermore, the combination of different modalities allows them to support each other in order to reduce the weaknesses and bring together their strengths to improve the overall functionality of the input device. The usefulness of multimodalities can be clearly observed in the case of communication failure. A replacement input model can take the role immediately to prevent any harmful consequences.

Figure 19. An Overview of the Multimodal Input Module Block Diagram

In our proposed framework, four types of input will be integrated into the *Multimodal Input Module* including human gesture, finger's pressing force, spatial pointing location, object's distance as shown in the figure below. These kinds of input modalities have been carefully considered before included in our wearable device. To choose the appropriate input models for our handheld interface, we must consider the task contexts and the user needs and the mental workload of each input model for the tasks [58]. Our focus will be the industrial contexts where many robot operators are required to wear protective equipment. First, there is a study [52] state that as human tends to react actively to the social cues which mean that if we can integrate social cues such as hand gesture mapping, speech recognition or pointing action into our device, the interaction process will become more natural. In this scenario, the use of speech should be eliminated as the industrial environment may contain a lot of surrounding noise which may lead to failure in communication. Also, the industrial environment is normally an open, large space where it is hard to speed to others. In this type of context, spatial information is becoming more crucial as it is intuitive and very robust compared to other input models. Finally, because of the protective gloves for safety purposes in an industrial environment, reliable and tangible buttons are included so that the industrial operator can feel the force of pressing the fingers while sending a command to the robot or to stop the robot in emergency situations. These buttons are very natural and simple to use while achieving extremely accurate results. The *operator* can choose the input modality or to combine them together to produce the most reliable output signal for the *operator* to control the robot depending on the industrial task and the working scenario. The following image illustrates the

deployment of the MIM in our proposed framework.

Figure 20. Multimodal Input Module with four types of input modality supporting the interaction between the *operator* and the *robot* in our proposed framework.

After getting the task information suggested by the robot via the TSTV module, the *operator* can select and generate the input signals and send them to the robot's perception system where these signals will be processed and confirmed with the robot's task model. The robot will then convert the information into robot's execution and return the feedback to the *operator* via the interaction model to the TSTV module. The *operator* can keep following the process until the robot finishes the required task. The multimodal input module is for improving the usability and the accessibility of the interface [60] for the *operator*. Many studies [60][54] report that the fusion of input modalities in the multimodal interface can optimize the human performance of the task while significantly reduce the overall workload for a human. Of course, providing multiple input modalities can introduce redundancies to our system. However, the redundant input signal not only improves the quality of the sending commands but also reduces the probability of sending an unreliable message to the robotic system. It is always more reliable to use different ways to explain and confirm your information with your robot in case of misunderstandings which is similar to the human-human communication method. The *multimodal input module* can also solve the many problems that we encountered for using monomodal input device for interfacing object. For example, it is very difficult for the single input model interface

to adapt to the frequent change in industrial tasks [55] such as defining the robot trajectory task. It is very costly, time-consuming and sometimes requires restarting the whole process with a single input device to fix the coordinate a small point in the trajectory. On the other hand, the multimodal input module can solve the problem by just simply switching between these different types of input for different task's requirement.

In short, there is a need to integrate different types of input into a unified input module. This module is required for to enhance the user performance for different tasks in different contexts. The module should contain multiple input modalities to make use of strengths of a multimodal input model. This module will help the *operator* to communicate with robot naturally and effectively. With the help of this device, the operator will be able to send multiple inputs to increase the reliability of the communication process in many different scenarios. Even if the task only needs one type of input from the user, the module is also very helpful since the user can send his commands and then confirm his actions using a different type of input with the robot in case the robot interprets wrongly the user's intentions.

3.1.3.3 Spatial Task Sharing Module

The motivation for the spatial task sharing module is because of widespread adoption of robots to various fields where robots may share the same working area with people. This raised the question of how to warranty the safety of people during the operation of the robot. In this scope of our framework, we introduce the *Spatial Task Sharing* module as a solution for this problem. This module can take the role of a communication channel supporting the communication between the robot and the surrounding people. The module will be attached to the robotic platform and is equipped with the ability to generate and project task information to the surrounding environment. Comparing to other modules, the STSM has the lowest level of authorization as it is designed for the robot to share its task information with people which means that everybody can get access to this kind of information if they are sharing the same working floor with the robot. The following figure describes the deployment of the STSM in our proposed framework.

Figure 21. The Spatial Task Sharing Module with the ability to share the robot's operation information with the passersby. This kind of information must be very concise and easy to understand for everybody.

Firstly, the *operators* and the *robot* will define the task's information. This information will be sent back to the *task model* and executed by the robot. The robot also sends the information to the *interaction model* for converting this amount of information to the form that most appropriate for increasing the awareness of the *passersby*. Next, the converted information will be transfer to the *Spatial Task Sharing module*. Before revealing robot's status to the surrounding people, the module must also consider other sources of information including the *environment information* and the *robot status*. The environment information can be the location, orientation of the surface where robot's information is going to be projected which is important for performing a distortion compensation. On the other hand, the *robot status* is the information of the relation of the robot with respect to people such as their distance. By combining the amount of information, a suitable message can be generated and sent to the *passerby*. Some studies [59] have suggested the requirements to evaluate the quality of information presented by an interface including as listed below. The number percentage in the bracket is the ranking for each feature.

- (a) Automatic presentation of contextually- appropriate information (18%)
- (b) Easy to use by the users (25%)

- (c) Graphical intuitive interface (16%)
- (d) Clear indication on the robot's display the next logical step (24%)

From the above criteria, the *Spatial Task Sharing module* must be able to act as an intuitive, clear and user-friendly communication channel to make the framework an effective interface. For this module, the emerging spatial augmented reality technology which has the ability to project computer-generated images directly to the real scene is a promising candidate. However, current implementations of this technology usually utilize the blur, low-contrast standard projector which can be seen in the original indirect interaction framework. In fact, most standard projectors cannot afford to provide bright and high-contrast images regardless of the conditions of the surrounding environment. From a review of the current technologies, we found that laser is the best source of light that can generate dazzling images against different kinds of environmental background. With computer assistance, animated laser graphics can generate very dazzling images even against a dark sky. Therefore, it is a good idea to use laser source to generate information for the interaction between human and robot in indirect interaction. Also, the images generated by laser source are very eye-catching which is very useful if we want to alert people of the ongoing operations of the robot. As a result, we decided to design and implement a special projector driving by a laser source and use it as our implementation for the spatial task sharing module which can replace the current standard projector in the current indirect interface system. As far as we know, our interface is the first system to design and implement this kind of projector in the field of human-robot interaction. The details of implementing this device will be carefully mentioned in the next chapter. In the following section, we will discuss in details human-robot interaction dialog between the *operators* and robots and how this communication model can improve the interaction between human and robot.

3.3 Human-Robot Interaction Dialog

In the scope of this project, we also consider the problem relating to the relationship between human and robot and how should the conversation between human and robot happen in the current “industry 4.0” settings. In the past, the human is normally the

boss who writes scripts and sends commands to a robot to control and verify the correctness of the robot's behavior. This communication model normally requires users to have expert knowledge of the entire system while the robot must take a role as a "slave" and obey all the commands from the main user. Consequently, it is very difficult for new users to control and interact properly with the robot for the first time. As robots are becoming more popular in various fields and interact more with human, it is very important to reduce the workload of human in the relationship with a robot. This leads to the trend in the robotic field that considers a robot to be a partner with humans. Also, robots are still very limited in term of intelligence, task execution, perceptual and cognitive abilities despite the recent technological advancements [61]. It is still a challenge for the robot to execute a complex task with high efficiency. Therefore, the cooperation between human and robot is a great solution [61][57] to improve the efficiency as well as the performance of the human-robot interface. If we look from a different angle, we should not consider the limitations of the robotic system is because of the functionality aspect of the robot. In contrast, we should recognize that a human and a robot are different entities with different qualitative features. Robot's strengths are its flexibility and repeatability while the human is more clever and intelligent. So, an interface which can create a collaborative environment for human and robot can help to balance the discrepancies and improve the performance of both human and robot. Additionally, the human theory of interaction which is called "Perspective taking" [52] has indicated that the ability to understand the thinking, feeling, motivations or taking other's perspective is the basis for successful communication. In the traditional robot scenarios, the main operator is usually the one who takes the robot's perspective and refine the communication messages for the robot to understanding properly. In such situation, many users must behave in an uncomfortable manner such as speak louder or slower with a repeated gesture. This kind of communication is much simpler compared to a normal human to human interaction due to the limitations of the robotic system compared to the complicated of the human body. However, as the human-centered approach is becoming the main focus of the robotic field, the requirement is to design a robotic system to serve human needs and the robot has to adapt to human's preference. For the incoming generation of robots, it should be the other way around where the robot

must take the human's perspective to form a successful communication between human and robot.

As a result, to enable this communication model, we introduce a suggestion system for human-robot cooperation. Building on this model, we propose our interaction model that can allow “human works less, robots do more” to complete a task completion or a process in the industrial context using the components in our proposed framework. The hierarchical analysis of the interaction model is illustrated in figure 22 where human and robot are partners and can help each other to complete the task.

Figure 22. A hierarchical task analysis for our human-robot interaction framework

The core idea for this interaction dialog is to allow the robot to suggest task's options to the operator using Augmented Reality technology. As shown in the task analysis model, the task can be divided into subtasks which can be further organized into a sequence of steps. For each step, the robotic system will render a suggestion for the user to select using the wearable input device. Finally, the robot will project a message to the environment for confirmation purposes. This interaction model will shift the workload of the task from the main operator to the robot. While the user still can determine the configuration of the entire task, he is not required to remember all the

details relating to the task since these elements will be suggested by the robot in the form of multiple options for a step. This kind of conversation will work beautifully with our proposed framework as we are using augmented reality for communication between human and robot and the strength of AR is for illustrations, visualizations, and demonstrations. The implementation of this interaction model is introduced in our case studies in chapter 4. The figure below is a flowchart that shows an example of how the flow of data input and decisions made during the operator-robot dialogue. During the dialogue, the robot assists the human in identifying task details and offering suitable alternatives for selection, by the human.

Figure 23: Flowchart for operator-robot dialog

With the augmented view provided by *task suggestion & visualization module*, the operator can provide the necessary support and commands to the robot. Selection and navigation through the menu options are executed using the *Multimodal Input* module. The operator can begin by addressing the chosen robot and indicating his choice of platform. The chosen robot affirms its selection through a visual indication or robot motion. Next is the identification of the required task. To reduce the number of interactions, the operator selects a task group and works towards identifying the

specific task that is required. With a prior knowledge or the ability to recognize the operator's intention, the robot would be able to better identify the required task and to reduce the number of steps required to complete the task definition and its required constraints. The operator subsequently initiates the task and continues to monitor task execution by direct visual observation and aided by updates of critical information. The operator may abort or pause (to modify) the progress of the task at any point during the execution.

3.4 Framework Discussion

In this work, we have proposed an interaction framework between the robot with a different group of users who also share the same working environment with it. The interaction process is supported via different interaction modules. The advantage of this framework is that can support multiple types of interaction between human and robot while making sure that the robot task information is secure in an industrial context. By combining different modules in the framework which have different features, the system can select the appropriate information and send to the corresponding modules for the target users. As a result, the new framework now can support different models of interactions such as *operator-robot*, *passerby-robot*. In the current indirect interaction interface, with one channel for communication, the system can only interact with one user at a time. This makes it difficult or even impossible to create a collaborative environment between multiple users with the robot which usually a requirement in many industrial tasks. Also, the multichannel configuration approach is very useful in increasing the security level of the system.

Figure 24: A summary of our proposed framework for human-robot interaction

This is because the framework can assign modules with different security levels to different groups of users depending on the administrator access level of users. The authorization privilege of users is determined by the interaction mediators they can access. As a result, only users with proper authorization privileges can setup, modify or stop the operation of the system. The *operator* group can get accessed to both TSTV channel and MIM channel which are secure channels while the *bystander* group can only access the STSM channel and probably the TSTV channel. The *passersby* group can only get robot information from the STSM channel. The ability to enhance the information security while maintaining different types of interaction is what differentiates our model from the previous indirect interaction framework. The feature is very important for many applications, especially in industrial context. Also, the multiple channel framework can make it easier for providing feedback between human and robot without interfering with other system information especially if the robot is operating in a shared environment with different people. Additionally, the introduction of the multimodal input module for the operator is another improvement our framework provided. First, it allows our system to be applied to different industrial scenarios with multiple ways to provide input to the robot for different tasks. The module is specially designed for industrial workers who

must wear protective equipment which limits their abilities to control traditional input devices to program the robot such as mouse and keyboard. The module can further improve the interaction process by reducing the workload for the user and make the input signal more reliable. Many works [62][60][59][63] have proved that communication with multiple inputs can improve the quality of human-robot communication. This is due to several reasons. First, human-human communication is a multimodal process [52] with different inputs transmitted via different channels. To convey information efficiently, people tend to use different modalities to describe the same source of information such as speaking about an object while using hand gestures to describe its shape. This human-human interaction model recommends that the similar multimodal input model should be applied to human-robot interaction. Secondly, the use of input model is very situational and each input model shines in a specific context [58][64]. For example, spatial information is better than speech in a construction site where there is a lot of environmental noise. By combining different input models, we would be able to widen the range of applications for our multimodal input module. Also, multiple modalities can increase the robustness of the system and reduce errors during the interaction process [61][65] because of the capability of ameliorating the number of errors between communication channels. Also, some studies [64] also prove that using multiple inputs can improve that quality of the individual input. For example, humans tend to speak more accurately while performing hand gesture [64]. Finally, a multimodal input device can be very effective in solving grounding problem between human and robot [57][65]. A common ground is very important for the interaction between human-human and human-robot since it helps us to coordinate, understand each other better while preventing misunderstanding and unwanted attempts. With the described advantages of the multimodal input model, the design and implementation of a multimodal input module are essential for the interaction in our framework. The combination of the TSTV module with see-through augmented reality technology and STSM with spatial augmented reality is an important contribution of our framework comparing to the original framework. In fact, our work is the first one to propose and implement this type of combined system in the field of human-robot interaction. The combination configuration can help our framework to take advantage of the strengths of each

technology and apply them to the human-robot interaction context. The following table will summarize the improvements of our framework comparing to the previous interface designed using the indirect interaction framework

Table 3. A Comparison between our proposed framework with other indirect interaction interface

Interface	Interaction modules	Ability to interact with different groups of users	Type of projector	Type of input device for the main operator	Level of Security
PMR-5R3, 2006 [12]	Projector, laser scanner	No	Standard	Monomodal	Low
Jongkyeong ark,2009 [11]	Projector, mobile PC	No	Standard	Monomodal	Low
TakuyaSasai et al.,2011 [36]	Projector, foot motion	No	Standard	Monomodal	Low
Ju-Hwan Seo et al.,2012[49]	Projector, laser scanner	No	Beam Projector	Monomodal	Low
UD, 2015 [50]	Projector, laser range finder	No	Standard	Monomodal	Low
AGV,2015[37]	Projector	No	Standard	None	Low
Our Interface	See-through glasses, laser writer, MIM	Yes	Laser	Multimodal	High

The framework can also be extended to the case of multiple robots working in the same working environment. The core idea is basically the same for this extended version. We will take advantage of our framework to design a solution for multiple robot interactions. The following figure illustrates our extended version of the proposed framework.

Figure 25: an Extended framework for interaction with multiple robots

The generalized version also includes three user groups with different levels of access to the interface. While each robot can be equipped with a STSM to help it to communicate with people and avoid any potential collisions, there is only a small change to our TSTV module. This is due to the fact that see-through augmented reality allows the tracking of multiple objects. As a result, if we consider one robotic system as a tracking object, we will be able to use one augmented reality application to interact with multiple robots which makes it very easy to implement this extended version of the framework. In fact, our implementation is to use multiple augmented reality markers attached to the robot so that the tracking module can differentiate the

robotic systems. Each robot is identified by an AR marker which can be tracked by the tracking system. By using the wearable device, we can select the robot to control and perform the desired task.

Chapter 4

Experimental Methodology

In this chapter, we proposed our implementation for the new framework. We will introduce the use of see-through augmented reality technology for the Task Suggestion & Task Visualization Module. The Multimodal Input Module is implemented in the form of a novel wearable handheld device which is presented both in term of hardware and software aspect. Finally, the Spatial Task Sharing Module with spatial augmented reality section is described in detail with the design and implementation of a special laser writer, a powerful device for projecting images to the background environment.

4.1 Framework Implementation

To assess the application feasibility of our interaction model in industrial contexts, we have designed and implemented three core interactive modules in our proposed framework and utilized the implemented devices for two different industrial tasks. The task suggestion & visualization module functionalities are recognized in the form of a see-through augmented reality glasses as see-through AR is the most suitable enabling technology for this module so far. As for the spatial task sharing module, a special laser writer will be prototyped that can overcome many limitations of the standard projectors. Finally, the multimodal input module is presented as a wearable handheld device which is specially designed for people working in the industrial environment. Our first application is related to the navigations of a mobile robot in an industrial environment as the mobile robot is becoming more popular in this context. The second scenario is for a taping robot system where our proposed interface will be used to improve the user experience during the interacting process. The user will be able to configure the taping task while monitoring the taping process directly on the workpiece. The following figure illustrates our implementations with the main components and their functions.

Figure 26: Illustrations of framework components and applications

For both implementations, three main components are utilized to support the interaction between human and robot as mentioned in the proposed framework. Each of the components will provide separate functions for the whole system. While the augmented reality glasses can overlay computer graphics over the real-world images, the wearable handheld device is used for sending input signal using different input models such as finger buttons, gesture mapping or laser pointer. The laser writer can provide visual feedback for the industrial task from the robot to the users or highlight the current task. The rest of the chapter will focus on describing the design and implementation of these components.

4.2 The Task Suggestion & Task Visualization Module with See-through Augmented Reality Glasses

4.2.1 Hardware Overview

In this project, we propose the use of Moverio BT-200 for performing the function of transparent smart glasses. The device which is designed to be used in various field including entertainment, manufacturing, medical science features a binocular, transparent glasses.

Figure 27: Moverio BT-200 augmented reality glasses

The BT-200 contains lens which has its own display, right in the field of vision. This feature allows the projection of virtual graphics into the surroundings. Furthermore, the lightweight device is integrated with Bluetooth and Wi-Fi technology. The most important feature of these smart glasses is the front-facing camera and motion tracker. This allows BT-200 to become a superior development platform for creating augmented reality applications and hands-free mixed reality experience. The device offers two viewing modes for developers including 2D or 3D mode. The 2D mode allows the user to view virtual contents overlaid over the video captured by the front-facing camera. On the other hand, the 3D mode basically provides the same functions except hiding the video stream from the camera. As the result, the user can monitor the computer graphical content while looking at the real environment via the see-through transparent glasses. Finally, the smart glasses run on Android 4.0.4 Operating System which allows us to take advantage of many features of Android platform for designing augmented reality contents. In the next section, we will provide details of how we will program and design our AR application for industrial scenarios.

4.2.2 Task Suggestion & Task Visualization Programming

To perform the task suggestion and task visualization features with augmented reality technology, the following steps are normally performed.

1. Create a task model or information for the robot system by representing it in the form of 2D or 3D graphical contents. This task information will be stored in the robotic system or in the memory of the smart glasses.
2. Perform the tracking process by using the integrated camera on the glasses to identify the designed targets for overlaying them with our stored task information.
3. Present the task information to the user via the smart glasses with a designed interface for the human-robot dialog to make it more user-friendly.

Figure 28: Augmented reality programming steps

4.2.2.1 Tracking

Tracking is the first process to achieve for implementing an augmented reality application. This is due to the fundamental concept of augmented reality that is to place virtual information to the real environment. As a result, one must be acknowledged about the positions and orientations of the locations in the real environment where the virtual contents are added. All this information is obtained via the tracking process. There are two main techniques for performing augmented reality tracking. The first technique is called marker tracking where a special AR marker is specially added to the working scene. The marker also contains a unique pattern that is very easy to detect using a normal camera. Consequently, this technique is very fast, simple and very stable. On the other hand, the marker-less tracking technique requires the use of natural features in the real scene to perform object tracking or position identification. The basic idea of this technique is to overlay virtual object into the scene without any knowledge of the user's environment. The method normally requires a deployment of an additional hardware to perform an environmental scanning process such as a 3D camera to detect natural features in the scene and use these features as the references for placing virtual objects into a real context. The approach is more complicated with lower tracking accuracy compared to the previous method. However, it is more flexible and more convenient to create mixed reality experience and can provide a new level of experiential immersion that marker tracking technique cannot offer. For this project, we decided to choose marker-based tracking method to be integrated into our Android application since our industrial application is applied in a limited working environment and require a high level of tracking accuracy. The details of programming the marker-based tracking process is illustrated in the figure below.

Marker-based Tracking

Figure 29: Marker-based tracking block diagram

The process requires a video stream from the front-facing camera from the smart glasses to be sent as the first input to the android app. In the next step, an image process method is used to identify the marker that is specially added to the working scene. The 3D position and orientation of this marker with respect to the camera are also obtained for overlaying virtual objects in the future. Finally, the identified marker is compared with a list of registered markers that are stored in the program’s database to match it with the correct one.

4.2.2.2 Overlaying Virtual Contents

The most fundamental problem of adding virtual information to the real scene is when and where to place those computer-generated objects. The question is partially answered as the marker-based tracking method provides us with the 3D position and orientation of the pattern with respect to the camera or our view perspective. The rest of the task depends on how we use the provided information to place, orient, and control the virtual objects. Specifically, the computer graphical object that we are going to overlay over the markers is a series of images that are specially designed to control or implement the flow of an industrial task. Depend on the requirements of the tasks and the working scenarios, we can have different ways to design the augmented reality images to satisfy the task performance. Basically, these images are already stored in the memory of the Android app. This app will control the “ON” or “OFF” state of these images using the camera video stream and the input signal received from the main user. As soon as the front camera of the smart glasses detects the marker in its field of view, it will signal the Android application running on the smart glasses and asking for user’s confirmation to start the image rendering process.

If the user input allows the rendering task to happen, the virtual images will be displayed at the marker position. The whole process is summarized in Figure 30.

Figure 30: Overlaying virtual objects over real scene

4.2.2.3 Designing the AR App

As we are designing an interface for human and robot in an industrial environment, the most important factor to consider is always user's experience. As a result, whatever the technology we are going to include in our interface, we are supposed to deliver a user-friendly product to our users. To achieve this goal, we divide our main app interface into two main categories including primary components and secondary components.

Primary components

The primary components of the Android application interface refer to the graphical elements that are fixed permanently to the application's background.

Figure 31: Primary components of the augmented reality application running on the glasses

These elements will maintain the user's awareness of the general configurations and parameters of the robot as well as the main user. Also, the primary components are designed to be small and placed at the corners of the main display so that they will not overlap with the field of view of the main user and other augmented reality contents. The above figure shows the primary components looking through the Moverio BT-200. The main interface contains the status of the robots that the system is able to control at the top left corner including M1, M2, M3 which represent Maven 1, Maven 2 and Maven 3 robot. The connect button at the top right corner is used to connect the robot system with the augmented reality glasses. At the bottom left, there is a map showing the position of the robot with respect to the main user. Finally, at the bottom right corner, there are icons and a menu showing the status of the handheld device which allows the user to see which button is pressed for confirmation purposes.

Secondary components

While the primary components are fixed to the application background, the secondary components will be more dynamic and focus on providing instructive information relating to the industrial tasks.

Figure 32: An example of augmented reality secondary components

During the interaction process between human and robot, these components will be generated, displayed and replaced depending on the task's contexts. Basically, as the user looks at the AR marker placed at the working scenario, the system will automatically identify the current stage of the task, generate and superimpose the corresponding augmented reality graphical information with respect to the orientation and location of the AR marker. Depending on the user inputs and selections, the secondary components will guide the main user go through the whole process to complete the entire task. Figure 32 is an example of the secondary components that will prompt the user to select the starting point of the object for the taping process to begin using the haptic button provided by the handheld device. This example is designed and applied to our taping system. Since we are dealing with an interface design problem, there is a rationale for adding and combining these elements together for the augmented reality framework. As the secondary components are the most important source of information for users to interact with the robot in our framework, it is essential to make it easy to understand for all kinds of users while maintaining a user-friendly and esthetical appearance to the users. Consequently, our secondary components contain three main elements including stage indicator, contextual information, and anchoring context. Firstly, our design provides the state indicator which allows the users to be informed of the current stage of the robot as well as the number of stages to be completed before the robot will carry out its task automatically. This kind of element is very important in user interface design since it gives the user a clear picture of the whole procedure that he is supposed to go through in order to complete the task while reducing the frustration of completing a long procedure. Also, it is also helpful for debugging purposes when the user accidentally recognize that he made a mistake during the calibration process. The stage indication will let the user remember and know exactly where to go back to fix it. Secondly, the contextual information component is generated to provide the user with task information, to instruct the user how to set up or configure a certain phase of the task or to offer options for the user to select in order to complete the current step. This element will vary depending on the status of the current industrial task. As was mentioned in the introduction chapter, one of our motivations for designing this user interface is to allow the user to receive graphical information from the system while

focusing on the working scene. As a result, this contextual information will be adjusted to make it unlikely to obstruct the user's vision. The visibility of this component is also very important since it is overlaid over the real scene which changes constantly and vulnerable to lightning conditions using the see-through glass. Therefore, the texts are customized in term of font style, size, and colors to guarantee that they are easy to read from a particular distance. For example, the text instructions use very high-contrast color including white and black. The hand-controller input instructions are designed with red, yellow and black so that the main user can easily interpret which button to press by looking at the hand graphics. Finally, an anchoring context is specially placed at the bottom of the secondary components that contains two different elements including a robot status and a controlling menu. The robot status enables users to quickly view the name of the robot as well as the current status of the task using a bar around the name of the robot. This bar will remain as green unless the robot encounters an error and requires troubleshooting which changes the color of the bar to red. Next, the menu's icons are designed to provide three main navigation functions for the augmented reality application which can be activated using the corresponding haptic buttons at the handheld device including emergency stop, reset the whole process and go back the previous stage. These functions are very important for the user while monitoring the operation of the robot since the user may need to intervene in emergency situations and these icons will serve as a great reminder to the user. In this section, we have investigated the augmented reality glass Moverio BT-200 which the first main component of our novel framework in terms of hardware specifications, software algorithm, and design. In the next section, we will introduce a very innovative device that can help our framework to overcome the contrast problem in many spatial augmented reality applications.

4.3 The Spatial Task Sharing Module – The Laser Writer

As was mentioned in chapter one, one of the main limitations of current spatial augmented reality technology is the low-contrast of the projector used. This kind of the projector requires specific environmental conditions to be effective in term of projecting augmented reality to the real scene such as an indoor environment with the low-light condition. As a result, it will be hard to apply the current type of spatial augmented reality technology to many industrial applications. On the other hand, industrial users who are interested in monitoring the tasks being executed by the robot require the robot information to be clearly visualized and presented in all different kind of environments. Therefore, designing a new type of projector for spatial augmented reality can improve the interaction between human and robot as well as extend the application of the indirect interaction framework to various fields.

To solve this problem, we recognize that laser is a very powerful source of light. Laser beams can generate animated laser graphics that are very eye-catching images even against a dark sky. This means that laser source could be a great solution for our current blurry projection problem in the current spatial augmented reality device. This has motivated us to design a laser-based projection system for projecting robot's information. This technology not only enables SAR technology to overcome current constraints but also is very user-friendly and affordable for low-budget projects. In the following sections, we will provide in detail our new laser writer device in both hardware and software aspect.

4.3.1 Hardware Design

4.3.1.4 Laser Projection Theory

Laser Graphic Projector can generate laser images by using the phenomena of the persistence of vision which says that “visual perception of an object does not cease for some time after the rays of light proceeding from it have ceased to enter the eye”. As a result, a laser image can be generated if there is a laser beam moving fast enough to create an optical illusion of images so that the viewer interprets them as continuous static or moving outlining shapes. Normally, human eyes and brain can only process

from 10 to 12 separate images per second and retain an image for up to a fifteenth of a second. If a subsequent image replaces it in this period, it will create the illusion of continuity. In order to implement such type of projector, two main systems are required including a laser source system and a mirror system which is shown in Figure 33. Firstly, a system with three laser diodes is needed to be able to emit beams of three basic colors (Red, Green & Blue). These beams will be mixed by a pair of dichroic mirrors into a single beam. Such three-color laser system with red, green and blue (RGB) theoretically is capable to generate all colors of the RGB-spectrum.

Figure 33: Main components of the laser writer

Next, this laser beam will be controlled by a system of two rotating mirrors. In fact, the rotation of these mirrors is generated by a system of two galvanometers that is attached to the mirrors. As shown in the figure above, the Galvo-Scanning Mirror X is placed perpendicular to the Galvo-Scanning Mirror Y. The rotation of these mirrors results in moving of the beam on a particular surface which, in turn, produces a laser image at a targeted surface. Now, let discuss in detail the physical structure as well as the fundamental operation of a galvanometer that is used in our rotating mirror system. The galvanometer is an electrical instrument that is normally used to detect small current. Such system consists of an electromagnet and a constant magnet mounted on the same axis with a mirror. When the current in the coil changes, a constant magnet interacts with the magnetic field of the coil and turns the shaft with the mirror to an angle proportional to the coil current. The combination of the two

Galvanometer scanners in an orthogonal plane allows us to control the position of the laser beam positioning on the surface. Historically, the galvanometer scanner evolved from the D'Arsonval meter movement which contains a DC moving coil – type movement in which an electromagnetic core is suspended within a strong permanent magnet field. A current directed through the coil will produce a magnetic field that opposes the permanent magnetic field and generates the rotation of the core. The core is restrained by springs so that its rotation is proportional to the current intensity. The core returns to its original position due to the restrained springs if there is no current in the moving coil. The galvanometer is also designed using the same principle. The galvanometer has a very thin rotor to minimize the rotor inertia in order to allow a fast movement. The moving coil is replaced with a high rigidly solid rotor, such as moving magnet or moving iron. This structure can be called a "Servo Motor".

Figure 34: The structure of a D'Arsonval meter and the design of the galvanometer

As shown in the figure above, the typical contemporary moving magnet galvanometer transducer is similar to a torque motor. Permanent magnets in the stator provide a fixed field that is augmented by the variable field developed by the control current through the stator coils. Seeking a new balanced field, the rotor (and mirror) executes a limited angular turning that is restrained by an elastic suspension system. The galvanometer is a broadband device that can be damped sufficiently to operate over a wide range of frequencies, from zero to an upper value depending on its mechanical resonance. Thus, it can provide the basic saw-tooth waveform for raster formation that is very helpful in creating laser graphical images.

4.3.1.2 Design Target

For the industrial contexts, the main objective for this device is to allow the robot to share its current operation status with people around. Therefore, our main target for the device focuses more on the ability to generate very bright images instead of paying attention to increase the image resolution as we are not dealing with images with many details. As a result, the system must be able to generate texts, images or shapes that are:

- Undistorted, high-contrast which can be observed clearly in both indoor and outdoor environment.
- Because of the limitations of the hardware, these images can be outlined. However, it should be understandable to people. Also, as this kind of information is targeted for the passersby, we will avoid generating complicated images or texts.

For our case study that will be presented in the next chapter which is relating to a navigation task performed by a mobile robot in a shared environment with human, the laser system should be able to render the following images projected on the floor.

Figure 35. Images that are going to be generated by our laser writer

The arrow shapes and the “STOP” word are for the mobile robot to share its movement direction while the rectangle and circle are for highlighting a specific area for masking in the taping process which will be presented in our case study section. As the information is projected on the floor, the color of the rendered image is also very important for increasing the visibility as well as the meaning of the shared

information. For example, the red color is normally mean “please be cautious” while green color could mean “safe” or “everything is okay”. Therefore, an RGB laser writer would be more helpful than a monochronic laser writer.

4.3.1.3 Design Specification

To achieve the desired images both in term of brightness and shape, the following components is included in the final design:

- Two laser galvanometer scanner sets with closed loop max 30000 points per second for vertical and horizontal direction.
- Class A red, green, and blue laser sources with maximum output power 5 milliwatts
- Driver boards for the galvanometers and the laser source
- A bridge board to convert input information from PC to the driver board (ILDA format is normally used for laser input)

For our system, the standard 24 frames per second are used to create the satisfactory fluidity for the eyes. As our galvanometer can move at maximum speed 30000 points per second, the maximum points we should use to draw our images are:

$$P = \frac{\text{Maximum Speed}}{\text{Maximum FPS}} = \frac{30000}{24} = 1250 \quad (4.1)$$

For our application, this number is enough to draw the mentioned images without any distortion or discomfort to the passersby’s eyes. Our projection distance will be about 1 meter which is approximately the height of our robot. A class-A (maximum power 5 milliwatts) laser source will provide enough power for creating high-contrast images and is accepted by the laboratory rule to be attached to the robotic system. The following block diagram describes the connection between components that will be included in the final design of the system.

Figure 36. Our block diagram for the laser writer design

4.3.1.4 Implementation Results

The completed hardware block diagram of the proposed laser writer is shown in figure 37. To convert an image stored in the computer into laser graphics, a laser controller board is used and interface with the computer using LAN network. This board receives the image and converts it into laser information format which is divided into two categories: RGB laser modulation and galvanometer orientation. The information will be wired directly to the drivers of the corresponding components including the RGB laser driver, X-Mirror galvanometer driver, and Y-Mirror galvanometer driver. As a result, we are not only able to control the intensity of the RGB laser but also to adjust the position of the laser beam with respect to a projecting surface. Color generation is reached via the intensity control of red, green and blue laser diodes using separate individual drivers for each source. These drivers control the intensity of each laser by increasing or decreasing the input voltage. All three beams are united in a special subtractive device that uses three dichroic semi-transparent mirrors to merge three different beams into a single powerful beam.

Figure 37: The laser writer hardware component block diagram

One main disadvantage of the laser writer is that there is a limit on the number of points the system can bounce to draw a laser shape due to the constraint of the galvanometer maximum speed. Therefore, if the system tries to render images containing more points than the laser writer can afford, there will be some distortion effects. Consequently, the laser images generated by the laser writer are normally unfilled drawings or outlines to reduce the bouncing points in a laser-generated image. However, the ability to project images in a bright outdoor environment is still a big advantage. Finally, we are also able to create animations with this laser system by projecting laser images that are in temporal sequencing synchronized in time. The following image shows our first laser writer prototype that is going to be attached to an industrial robot.

Figure 38: Laser Writer Prototype

Here are some results of laser graphical images created using our proposed laser writer system. For example, the following images show the laser writer projects the current movement status of the mobile robot in the bright and low-light environment.

Figure 39: Laser graphics projected to the floor in a bright and low-light condition

4.3.2 Image Generation

Laser graphics is accomplished by steering a single beam of laser light accurately through a series of specific points in space. Most graphics systems use frames of points for graphics storage and plotting sequential frame animation. The images are stored as a series of points and that animation frames are played back in sequence to create the illusion of motion like in a cartoon. The points that make up the image are converted by a digital to analog converters (DAC) or a display processor in the computer into analog voltages that drive the scanning system. The scanners use small mirrors mounted on galvanometers at right angles to each other to control the vertical and horizontal deflection of the laser beam. The points in the image are refreshed or redrawn by the scanners at such frequency, that human eye is tricked into seeing an image - typical refresh rates are 18-22 Hz. The rapid projection of a sequence of

slightly different images gives the illusion of movement just as in traditional animation cartoons. This section will provide in detail how a laser image is generated using the described hardware.

First, a computer with a laser editor program is required to create the desired laser image. For this project, we use Lasershow Designer 2000 editor to create and edit laser image. As the laser image is generated, the computer can save the content into an ILDA file. The file will later be transferred from the computer to the laser controller board using LAN network. ILDA format also allows the user to create and store multiple laser frames in one file. Each frame contains the positions in Cartesian coordinate and RGB values of each point in the image. As the controller board receives the ILDA file, it will read it and forward the corresponding information of each point in the laser image frame to the ILDA distributor using an ILDA cable. For each point in the laser frame, its information in the ILDA file will be converted into Red, Green, Blue intensity values and a galvanometer mirror angular orientation in both vertical and horizontal plane. The data is transferred to the corresponding driver modules as illustrated in figure 40. Finally, as the laser writer scans through all the points in the laser frame, the outlining beam will overlay a particular laser shape over the projected surface.

Figure 40: Laser graphics generation process

In the section, we have described how to generate laser graphics using the laser writer hardware. The next section will deal with performing laser image transformation to remove distortions while projecting the image onto a planar surface. Also, since there will be a camera attached to laser writer system, a laser writer – camera calibration is performed to transform information from the camera to the laser writer and vice versa.

4.3.3 Image Distortion Compensation

4.3.3.1 Laser Image Calibration

As the projector frame is not necessarily perpendicular to the projection surface, there will be visible distortions in the source image (Projector Frame) projected onto the surface (Projected Image Frame). This distortion needs to be corrected through a pre-warping process that is applied to the projector frame, before being projected to the surface which will allow the laser writer to generate undistorted images on the floor or on the wall. This process can be achieved using the idea of projector-camera homography. The assumptions for this problem are:

- The intrinsic and extrinsic parameters of both laser writer and camera are unknown and
- The projection surface is flat.

Let (x_p, y_p) is the pixel value of a point in the original image stored in the projector memory or the projector frame. This pixel will be projected by the laser writer to a projection surface at location (x_s, y_s) . Finally, the projected point is detected by the camera attached on the robot at location (x_c, y_c) on the camera frame. Thus, the point in the original image has been transformed twice before seen by the camera. This transformation is illustrated in the figure below where P is the perspective transformation matrix between the laser writer and the projection surface, C is the transformation matrix between the projection surface and the camera and H is the compound transformation matrix between the laser writer and the camera. To remove the perspective distortion from the projected image on the projection surface, the idea

is to figure out the value of P and then apply a pre-warped transform P^{-1} to the original image before projecting it to the physical surface. The relationship between P , C , and H is described in the figure where

$$P = C^{-1}H \quad (4.2)$$

Figure 41. The relationship between the laser writer coordinates, the physical surface coordinate, and the camera coordinate

The value of H can be obtained by applying the following homography transformation between the laser writer frame and the camera frame.

$$(x_p, y_p) = \left(\frac{h_1 x_c + h_2 y_c + h_3}{h_7 x_c + h_8 y_c + h_9}, \frac{h_4 x_c + h_5 y_c + h_6}{h_7 x_c + h_8 y_c + h_9} \right) \quad (4.3)$$

With

$$H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \quad (4.4)$$

H has eight degrees of freedom with the constraint $|(h_1 \dots h_9)^T| = 1$. The transform can be formulated in homogenous coordinate by the equation:

$$\begin{pmatrix} x_p w \\ y_p w \\ w \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix} \quad (4.5)$$

The value of H can be obtained from at least four pairs of corresponding points between the laser writer and the camera. In case of more than four correspondences, we use the least-squares method for n corresponding pairs $\{(x_{pi}, y_{pi}), (x_{ci}, y_{ci})\}$ with the following matrix:

$$T = \begin{pmatrix} x_{c1} & y_{c1} & 1 & 0 & 0 & 0 & -x_{c1}x_{p1} & -y_{c1}x_{p1} & -x_{p1} \\ 0 & 0 & 0 & x_{c1} & y_{c1} & 1 & -x_{c1}y_{p1} & -y_{c1}y_{p1} & -y_{p1} \\ x_{c2} & y_{c2} & 1 & 0 & 0 & 0 & -x_{c2}x_{p2} & -y_{c2}x_{p2} & -x_{p2} \\ 0 & 0 & 0 & x_{c2} & y_{c2} & 1 & -x_{c2}y_{p2} & -y_{c2}y_{p2} & -y_{p2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{cn} & y_{cn} & 1 & 0 & 0 & 0 & -x_{cn}x_{pn} & -y_{cn}x_{pn} & -x_{pn} \\ 0 & 0 & 0 & x_{cn} & y_{cn} & 1 & -x_{cn}y_{pn} & -y_{cn}y_{pn} & -y_{pn} \end{pmatrix} \quad (4.6)$$

The value of unit vector $\vec{h} = (h_1 \dots h_9)^T$ can be determined by finding the vector that minimize $|T\vec{h}|$ which can be obtained by calculating the eigenvector that corresponds to the smallest eigenvalue of $T^T T$. Similarly, the value of C can be calculated using the homography transformation between the projection surface and the camera. For demonstration purposes, we have fixed the laser writer to project directly to the floor and marking 4 different corners of the floor with a bright color to facilitate the camera's detection task. From these four corresponding points, we can calculate the value of matrix C using equation (4.6) and the value of matrix P using equation (4.2). The distortion compensation process now can be performed by multiplying the original image with P^{-1} to create the pre-warped image in the laser writer frame. The pre-warped image now can be projected to the floor by the laser writer. Although the projected image should appear undistorted, it should be scaled to become more readable for the surrounding people. Therefore, a scale transformation matrix S is also applied to the original image. This matrix values can be chosen manually for the laser writer depending on the size of the projected image.

Figure 42. The image is pre-warped by multiplying with $W = P^{-1}S$ in the laser writer before projecting to the physical surface for obtaining a rectilinear image. S is just a scale rigid body transform to zoom the original image in the constraint of the laser writer.

4.3.3.2 Laser Writer – Camera Calibration

As the multimodal handheld input is equipped with a laser pointer, the robotic system should be able to detect the laser dot generated by this pointer and highlight the location using the laser writer to show that it can understand the input information from the operator when the operator is indicating a reference goal location or choosing an item. This functionality is supported by the laser projector-camera system attached to the robot. The first step is to perform a calibration process between these devices to calculate the transformation matrix from the laser projector reference frame to the camera reference frame. Suppose (x_p, y_p) is the pixel value of a point in the image in the laser writer frame. This pixel will be projected and detected by the camera at location (x_c, y_c) . By projecting a square using the laser writer in the field of view of the camera, we would be able to identify the location of the square's corners in the frame captured by the camera. By applying equation (4.4), we would be able to calculate the values of the transformation matrix H .

Figure 43. Camera -Projector Calibration Using Homography methodology with the corresponding points extracted from a projected square.

The next step is to detect the laser dot in the camera frame which can be done by finding the maximum intensity value throughout the whole frame as the laser dot normally has the maximum value. Let $I(x_{co}, y_{co})$ is the intensity value of the laser dot in the image frame F_{xy} , the value of can be identified as:

$$I(x_{co}, y_{co}) = \max_{(a,b) \in F_{xy}} F(a, b) \quad (4.7)$$

The location of the laser dot in the camera frame can be calculated as:

$$\begin{pmatrix} x_{pow} \\ y_{pow} \\ w \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \begin{pmatrix} x_{co} \\ y_{co} \\ 1 \end{pmatrix} \quad (4.8)$$

By normalizing the result, we can get the location of the laser dot in the laser writer frame. Figure 44 demonstrates the use of the laser marker to indicate a position and the camera-projector system. The system confirms the position of the marker, by responding with the projection of an arrowhead that points to the marker location. In this scenario, the robot will project an arrow that follows the laser marker indicated by the user to show it is tracking the user's input.

Figure 44: Laser Input Detection and Arrow Projection

In this section, a real custom-designed laser outlining projection system has been implemented as part of Human-Robot interaction for generating graphics for augmented reality applications. Without the wearable displays, the visual interaction between the passer and Robot would only be possible to share in the notifications by the robot executed through the “Laser Writer”, which we have mounted on the robot. High-contrast real-time display of comments, intentions or visual feedback of the robot can be directly projected to the real environment. Whilst the unfilled line drawings may be considered as a disadvantage, the ability to project images in a bright outdoor environment is a powerful advantage of this component comparing to previous work in the field.

4.4 The Multimodal Input Module – a Wearable Handheld Device

This section will introduce the final component of our framework implementation which is a wearable handheld device for the main operator to provide a different type of input modalities to both the augmented reality system and the robotic platform. Our main motivation comes from the lack of a proper input device for an industrial worker to interact with the robotic system while wearing safety equipment. The traditional tools such as mouse and keyboard are inconvenient and difficult to use if, for example, the worker is wearing a pair of protective gloves. The interacting device must allow people working in an industrial environment to send commands signal to robot naturally and comfortably.

4.4.1 Hardware Design

4.4.1.1 Design Target

The core objective for this module is to provide the industrial workers a simple and comfortable device to send input commands to the robot during the interaction process. Therefore, the device must satisfy all the industrial requirements which are: safety, robustness, and easy-to-use to be popularly accepted by the related community.

Safety: One of the most important requirements for the industrial environment is safety, this motivated us to think about a wearable single-handed device as a single-handed tool would be considered to be safer. Furthermore, as an industrial worker must wear protective gloves, it would increase the safety level for the interaction if the device is attached firmly to the operator's hand.

Robustness: As mentioned in the previous section, a multimodal input model is proved to be more reliable and less error-prone comparing to the monomodal input framework.

Easy-to-use: All the chosen types of input are very natural and familiar to human input model which will make them easy to use for the new users.

Consequently, our device will incorporate the following input modalities:

- Hand gesture mapping.
- Haptic buttons.
- Laser pointer.

The reasons why the three mentioned input modalities are selected are as follow:

- Hand gesture mapping: using hand gesture is a natural way to interact with the robot as we use hand gestures to express our thinking every day. As a result, it will not take a lot of time or efforts for new users to get familiar with the device.
- Haptic buttons: to press a haptic button to select or confirm a step in a process is simple, reliable and intuitive for both novice and expert users. When human's fingers aspire to touch and feel things, they immediately attempt to press, shift, turn or pull them. This keeps the fingers busy and helps human to physically relax and thus be more concentrated mentally. Dexterous motions require antagonistic forces. The control should be accompanied by reasonable finger force applied to the spring-return force-feedback shafts of the displacement sensors with pleasant tactile buttons. Pushing force-returning pedals produce better sensations than, for instance, moving fingers in the control glove. Pushing buttons is like a fingers exercise for musicians. With the help of the force-returning pedals, the haptic button not only works as a single soft push-buttons but also as a proportional haptic input control affected by a single finger, or combination of fingers. Comparing with, for instance, joystick control or robotic glove, we have designed a single-handed naturally to use remote in which all 5 fingers of one hand are ergonomically interfaced to the buttons.
- Laser pointer: people get confused sometimes while receiving the target for a task or a process from others. As a result, they normally point directly to the target to make it more intuitive for others to understand their information. This explains why laser pointer will be extremely natural and useful for the interaction between human and mobile robot. With these functions integrated into one device, the system provides the users the option to choose and

combine these comfortable inputs to perform their industrial tasks depending on the industrial contexts and the requirements of a specific task.

For our demonstration with the mobile robot and the taping robot, the multimodal input device needs to provide the following functionalities:

- Use the gesture mapping and buttons features to control the robot manually
- Use the laser pointer input to define the goal location for the robot navigation in an automatic navigation task
- Use the laser pointer and the buttons to configure and define the masking area for the robot taping process

4.4.1.2 Design Specification

For designing and implementing our new handheld device, we have investigated existing and popular single-handed wearable devices for human to interact computer system as shown in the figure below.

Figure 45. Existing popular single-handed wearable devices

Most of them use haptic buttons and hand-motions as their inputs since they are considered to be natural and easy to perform for a human. As a result, these types of input will be added to our device. However, we will make some improvements relating to each input model to make it suitable for the industrial contexts. For the current prototype, we will consider the following design.

Figure 46. Our proposed design for the multimodal input module

The reason for us to come with this design is that we want a single-handed device that can attach firmly to our hand which not only makes the interaction process safer but also prevents the operator from feeling tired. Also, because of the need for input buttons, we need to find a proper number of buttons and arrange them properly to our device. Some input devices use one button which is not enough for many complicated tasks while others use too many buttons which may create confusion. Thus, we decide to add five buttons corresponding to our five fingers and arrange them next to our finger positions on the device. This arrangement will make the users feel very relax and they do not need to look at the position of these buttons to check if they are pressing the correct buttons which is very comfortable for working in the industrial contexts. As we are going to integrate three different input modalities to the module, the following are the specifications for each type of input:

Hand Gesture Mapping: As the device will be attached firmly to the operator's hand, this kind of input can be generated by installing an IMU sensor with 10 degrees of freedom to the body of the device to receive the orientation of the hand posture. The orientation of the human gesture should be recognized by collecting the roll-pitch-yaw data from the sensor. For our implementation, we expect to receive the following data

Figure 47. The multimodal input device with its gesture specification

Buttons: the buttons should be big enough to increase the pressing reliability. Furthermore, they are should be carefully chosen so that they can be pressed with a suitable force in a range of an adult's finger (1-3 Newtons) to warranty the tangibility and robustness of the system. A button which is too soft can create error and a hard button is not comfortable to press.

Laser pointer: A green laser pointer should be installed to our system as our eye is more sensitive to the green color region than other colors which make the green color is easier to detect. The maximum power for the laser pointer is also lower than 5 milliwatts to satisfy the safety requirement.

4.4.1.3 Implementation Results

Figure 48 also shows our first sketch of the handheld device to be added to our framework.

Figure 48: Handheld device prototype model

Figure 49 introduces our first prototype.

Figure 49: The first prototype of our handheld device

A lot of sensors are included to support these above interaction modalities. A green laser pointer is attached on top of the device for pointing action. A 10-degree of freedom inertial measurement unit (IMU) is embedded and programmed for mapping human gesture to robot operation. Five haptic buttons corresponding to five fingers are designed and arranged for controlling the system. The detailed technical specification of the device is as follow:

1. A small Arduino Pro Mini microprocessor has been selected as the heart for the handheld controller.
2. Five linear spring-return displacement sensors have been used for haptic input buttons. Five operational amplifiers are attached to these sensors to amplify the pressing and releasing signals of the springs to which they are connected.
3. A linear potentiometer has been installed to serve as an additional haptic sensor which can detect the slides by middle or ring finger that can be used for industrial tasks that require fine tuning control.
4. A 10 degree-of-freedom IMU is also added for the purpose of hand gestures detection.
5. The device supports wireless and secured communication using two modules: (a) Bluetooth Modem - BlueSMiRF Gold. (b) Wifi Transceiver ESP8266. These modules allow the communication between the handheld device, robotic platform, and HMD glasses.
6. Laser pointer module consists of TTL-controlled laser diode driver and the laser diode PLP520 for pointing purposes.
7. Distance towards the laser pointing is measured by LIDAR Lite rangefinder module, which is a compact high-performance optical distance measurement sensor.

Figure 50 shows the detailed hardware block diagram as well as the connections between those hardware components.

Figure 50: The handheld device hardware block diagram

All the desired input models are integrated into the multimodal device. We also add some additional type of input such as the laser scanner for distance measurement, however, it will not be used for our current demonstration. Finally, some measurement data from the input buttons and hand gesture mapping has been collected for checking the device performance. The data is presented on the next page. The buttons can support different level of pressing which output voltage range from 0 – 1050 millivolts for the thumb buttons and 0 – 900 millivolts for others. The gyroscope sensor can measure very accurately the human input angle which can be used later to map the operator gesture to robot's movement in the demonstration.

Figure 51. Data collected from using buttons and gesture mapping input models

In summary, we have implemented all the interaction modules in the new proposed framework human-robot interaction interface. We described in detail the design and implementation of the augmented reality system, the laser writer and a novel multimodal handheld for the interface. The devices are conceptualized and implemented to enable the transfer of different input modalities from human to robot. The modules can be applied to many applications in the human-robot interaction field.

Chapter 5

Case Studies and Discussion

In this chapter, we will apply our novel framework to two different interaction scenarios between human and robot in industrial environments. Each scenario has its own requirements and targets for both human and robot. We will introduce each case study in detail including the interaction scenario, proposed methodology, and results.

5.1 Case Study 1: Human - Mobile Robot Interaction

This case study will show how our framework can be applied to the interaction between human and mobile robots which are normally deployed in many industrial applications.

5.1.1 The Mobile Robot Navigation Scenario

Mobile robots have been deployed in various contexts such as material transportation, telepresence communication, sewer investigation or rescue tasks. With the developments of sensor technology and advanced algorithms recently, mobile robots not only able to perform the pre-programmed job but also can cooperate and help human to complete a task. In such a scenario, the mobile robot which is controlled by the main operator or be able to navigate automatically around the constraint work floor is supposed to interact with the main operator and with the people in the surrounding environment while performing the task. As a result, an interaction model must be applied to not only satisfy the interaction requirements but also to increase the safety of all the participants. The robot, as a partner to the human, must have the knowledge of the task it is going to perform so that it can provide options, suggestion or even corrections to the user. Only by fulfilling these requirements, robots can become the user's personal assistant and lower the cognitive workload for the operator. On the other hand, the operator must have the ability to configure, control, monitor, intervene, and stop the robot operation in emergency situations is essential to increase the safety level of the interaction process. Additionally, since mobile robots are sharing in the same environment with people, the robot must be aware of the people in the surrounding environment and the ability to communicate and inform people of its incoming movements is very important to avoid any possible collisions.

5.1.2 Procedure

In this case study, we will consider an interaction environment with three main agents including the operator, the robot, and the passers-by. While the operator is equipped with two main devices which are the see-through augmented reality glasses and a

multimodal input device, the MAVEN mobile robot is equipped a laser writer. The robotic system is already programmed with the function to navigate in a shared environment with people.

Figure 52. The Mobile Robot Navigation Scenario

The procedure for the interaction process is as follow:

1. The operator will wear the augmented reality glasses and the multimodal input device to interact with the robot. By looking at the robot's marker, the operator can receive the task's suggestion from the robot to start the navigation task configuration process.
2. The robot will start the interaction dialog by asking the operator the task he wants to perform and how to select it with the handheld device. After the operator chose *the navigation task*, the robot will suggest types of navigation it can perform including *a manual mode* and *an automatic mode*. If the operator chooses the *manual mode*, the robot will suggest instructions (task visualization) on how to control the robot with manual mode and. The process is similar in case the operator proceeds with the *automatic mode*.
3. With *the manual mode*, the operator will have the opportunity to use the gesture mapping feature of the multimodal input device to map his gesture to the movement of the robot. On the other hand, the laser pointer input with in

combination with the vision system of the robot is utilized to help the operator perform the navigation task in the *automatic mode*.

4. As the robot is performing the navigation task, the operator can receive the information of the robot current status via the see-through glasses. On the other hand, the passers-by will be continuously informed of the robot's operation via the laser writer.

5.1.3 Objectives

The objective of this case study is to demonstrate how our framework can be applied to the scenario of a mobile robot which can move automatically in an environment with surrounding people.

- To demonstrate the implementation feasibility of our proposed framework in a specific human-robot interaction problem
- To demonstrate the use of see-through augmented reality module with the ability to send task suggestion and visualization from the robot to the operators.
- To demonstrate the use of Multimodal Input Module with different types of input in supporting the operator to send input commands to the robot.
- To demonstrate the use of the laser writer which can generate very high-contrast images to allow the robot to share its operation status with the passers-by.

5.1.4 Results

5.1.4.1 The Operator-Robot dialog for manual control mode

At the beginning of the interaction process, the main operator must configure the navigation task for the mobile robot using see-through technology and the multimodal input device. While the robot takes the role of an assistant and suggests the task's options and requirements to the operator, the operator must use the multimodal input device to confirm his selections. The image below describes this type of human-robot dialog from the until the operator determines to choose the manual control mode.

<u>Operator Action with the MIM</u>		<u>Contents displayed AR Glasses</u>	<u>Mobile Robot Action</u>	<u>Step Description</u>
(2) Press the thumb button provided by the MIM to select the robot		(1) Suggests the operator to press thumb button to confirm his decision to take over this robot.	Step 1: A marker is attached to the robot as the reference position for displaying augmented reality contents. With the help of the augmented reality glass running on our augmented reality application, the main operator will be able to see a GUI which displays detailed instructions of how to select tasks and control the robot
	(4) Choose the Navigation task by pressing index button (which corresponds to letter "I" in the legend instructions)		(3) Suggests two tasks that can be performed including Navigation and Delivering	Step 2: The system suggests possible tasks that can be performed by the robot
	(6) Press the index finger to choose the manual mode		(5) Suggests two models that can be performed including Manual and Automatic	Step 3: The system asks the user to identify the controlling mode for the navigation task

Figure 53. The interaction dialog between the operator and the robot from selecting the robot to choosing the manual navigation mode.

As soon as the operator choose the manual mode, the robot will show the instructions to help the operator to control the robot in manual mode.

Operator Action with the MIM	Contents displayed by the AR Glasses	Mobile Robot Action
(8) Read the task instructions from the robot to figure out how to control it with gesture mapping and buttons		(7) Provide a menu with some instructions on how to control the robot via the AR glass
(10) Follow the instructions to manual control the robot		(9) Highlight the current turning left movement of the robot

Figure 54. Human-Robot interaction dialog for controlling the robot in manual mode

By using our designed modules, the task information is provided by the robot via the TSTV module which can shift the workload from the operator to the robot. Also, the interaction dialog between the operator and robot is very secure as all the information is displayed via the smart glass and the only way to modify the information is to use the multimodal input device which is equipped only to the main operator.

5.1.4.2 Multimodal input device for manual control mode

In the manual mode, the operator can use the gesture mapping in combination with the buttons to control the mobile robot. This shows the advantage of a multimodal input module where the user can choose the most comfortable input modality for the task requirements. For example, in this controlling scenario, gesture mapping would be suitable for turning the robot to the left and to the right and it is natural for a human to use our hand gesture for indicating the left or right direction. On the other hand, it is not user-friendly and unreliable to use gesture to indicate the robot to move forward and backward as point the MIM forward or backward is not natural. In this case, buttons can be used as a good substitution. The following image describes some results relating to control the robot with gesture mapping input and buttons from the multimodal input device.

<u>MIM Input</u>	<u>Robot Movement</u>
Tilt to the left	Turn left
Tilt to the right	Turn right
Press index button	Move forward
Press both index and middle Button	Move backward
Press all buttons except thumb	System Reset

Figure 55. Using gesture mapping and buttons to control the robot movements. The input modalities are chosen to make the operator comfortable during the controlling process

5.1.4.3 Task sharing module in manual control mode

While the operator interacts with the robot using the multimodal input device and the see-through augmented reality glasses, the robot with the attached laser writer will keep updating its status by projecting computer-generated texts or shapes on the floor. As a result, both main user and passersby are aware of the mobile robot operation.

<u>MIM Input</u>	<u>Robot Movement</u>	<u>Laser Writer Projection</u>
Tilt to the left	Turn right
Tilt to the right	Turn left
Press index button	Move forward
Press both index and middle Button	Move backward
No input	Stop

Figure 56. Control the mobile robot with different input model from the MIM and the corresponding information projected by the laser writer for sharing with the passers-by

The robot operation status with the corresponding projected images is described in the figure above where the robot can highlight its current with texts such as “READY”, “STOP”, “NAVIGATION” or its next movement with arrows. This also

shows the capability of our proposed interface to divide the system information into two different channels and direct them to different targets to improve the security level for the human-robot interaction process.

5.1.4.4 Automatic Navigation with Laser Pointer

In this second demo, the idea is to illustrate the use of laser pointer input from the multimodal input device to perform a task in an autonomous scenario. Basically, the operator uses the laser pointer to identify a goal point for MAVEN mobile robot. The robot will then project appropriate visual feedbacks and move to the destination. With see-through augmented reality system, the operator can select this mode of operation in the same manner as demonstrated in the first demo. However, the instruction on how to control the robot will be different. The AR glass shows that the robot is in autonomous control mode, and it suggests the user providing a goal point input using the laser pointer. The main user is providing a goal point by turning on the laser pointer, holding the index finger button in the pressed state and moving the laser dot for locating the desired goal target for the robot. A feedback arrow from the laser-writer is projected and follows the laser dot which is enabled by the camera sensor attached on top of the laser-writer. This is an intuitive way to guarantee that the mobile robot is correctly receiving destination information from the user. The user then pressed index and middle finger at the same time according to the instruction to confirm his goal point. As a result, the robot status changed to “MOVING” which was updated in both AR glass and on the floor. Finally, the robot stopped and projected “FINISH” word on the floor and waiting for the next instruction from the main user.

<u>Operator Action with the MIM</u>		<u>Contents displayed AR Glasses</u>	<u>Mobile Robot Action</u>	<u>Texts Projected by Laser Writer</u>
(2) Read the suggestions from the robot via the AR glasses		(1) Suggests the operator providing a goal point input using the laser pointer while projecting its current status to the floor
	(3) Turn on the laser pointer and press the index finger to set a goal point for the robot		(4) Tracking the operator input with the vision system and the laser writer
	(5) Press the index and middle finger at the same time to confirm the goal point location		(6) Detect the goal point and start the automatic navigation system, send task information to the laser writer for informing people around
	(7) Wait the robot to finish to reset the system or to select another task		(8) Stop and update its status with the laser writer

Figure 58. The human-robot interaction in the automatic mode where the operator can use a laser pointer as the input for identifying the goal point for the robot's navigation task

5.2 Case Study 2: Taping Robotic Platform

This case study will introduce an implementation of our framework in designing the interface for the current semi-automatic taping robotic system and the related calibration algorithms to enhance the user experience.

5.2.1 Taping Robot Interaction Problem

Applying masking tape to a particular area is a very important step for protecting an uninvolved surface in processes like mechanical part repairing or surface protection. In the past, the task was very time-consuming and requires a lot of manual works. In recent years, with some advances in the fields of automatic robotic system and computer vision, the task now can be completed with the help of an automatic taping system containing a 3D scanner for the 3D model reconstruction, a supporting platform, a taping robot and the robot taping tool. The supporting platform is used to mount the work-piece for taping. This platform can either be a simple fixed base or a rotating platform. The special design of the end-effector is required to meet the proper taping requirement. Meanwhile, the mechanism for cutting and holding the tape is also needed to accomplish the taping process. This automatic taping system, which has been proved to provide better quality and be at least twice as fast as the work done by a human operator, poses some limitations in term of the human-robot interface.

Figure 59: The taping robotic platform

Firstly, since the robot can only identify the area of the workpiece using a 3D scanned model stored in the computer software, there is no opportunity for the user to view the trajectory or the taping area the robot is going to cover on the real object. Secondly, although the main taping process is automatic by the robot, the user must define and configure the parameters such as the coordinates of the taping area, taping speed or taping pattern with the computer system using traditional mouse and keyboard. This process requires the expert knowledge from the user to make sure the system works correctly as expected. This not only reduces the user-friendliness of the system for new users but also not intuitive to interact with. This problem can be solved using our proposed framework for human-robot interaction. The Spatial Augmented Reality component (laser-writer) could be deployed to overlay the computer-generated information over the workpiece to make the taping process more visual. The information overlaying process is essential in a human-computer interface because it allows the user to intuitively monitor the intention of the robot and intervene if there is some misunderstanding during the masking process. The multimodal input device will allow the user to directly define the taping area on the workpiece's surface using a laser pointer output which is easy and very natural for all users. Finally, with the suggestions from the see-through augmented reality system, the new user can select the appropriate parameters for the taping process.

5.2.2 Procedure and Objectives

To demonstrate the application of our framework for the robot taping problem, the following set up will be used as shown in the below figures. The main operator again is equipped with the task suggestion & visualization module which is the Epson BT-200 see-through augmented reality glass and the multimodal input device. Our laser projection system is calibrated and added to the taping system reference frame. A marker is attached to the side of the taping platform for overlaying graphical information over the real environment. The main user will be able to control the whole system by looking at the instruction displayed at the marker using the see-through glass.

Figure 60: Proposed Taping Robotic Platform

Figure 61: Our implementation for taping robot system

To complete the interaction task, the following steps must be completed:

1. As the augmented reality glasses will render a digital user interface for the taping task, the operator is supposed to read the task information suggested by the robot via this device.
2. The taping robot can mask the work-piece with different patterns that varies in shape, size, and orientations. The operator needs to choose a taping method that is supported by our taping robot with the MIM device.
3. When asked by the augmented reality system, the user must use the laser pointer to define a certain area of the provided work-piece for the robot taping task. The laser system will track the operator's input and project the corresponding visual feedbacks.
4. The operator monitors the taping process with various taping information via the see-through glass and the laser projector.

The camera-laser writer calibration process that is mentioned before will be useful in this scenario to allow the laser writer to project correctly the digital taping shape onto the workpiece. The following steps describe the taping process in detail:

Step 1: User provides the taping starting and ending points using the laser pointer function.

Using the vision system, the starting and ending position for the taping segment provided by the user is identified using the following formula. Given the coordinate transformation from the workpiece system to the camera system, T_{CW} , the marked points $P_{w1}, P_{w2}, \dots, P_{wn}$ on the workpiece can be calculated from the following equation, where P_i is the location of the projected point with respect to the camera.

$$[P_{wi}; 1] = T_{CW}[P_i; 1]$$

The accuracy of the point estimation depends on the accuracy of the coordinate transformation, the vision measurement accuracy, and the size of the spot. After this step, the laser writer will highlight a taping area on the workpiece

Step 2: The video cameras take images and calculate the positions of the projected shape.

Using the same method, assume the vision system's outputs are a set of the coordinates of the projected points (sequentially provided) $P_1, P_2, \dots, P_n$. The marked points $P_{w1}, P_{w2}, \dots, P_{wn}$ on the workpiece can be calculated from the same equation,

$$[P_{wi}; 1] = T_{cw}[P_i; 1]$$

Step 3: The estimated points should be marching with the workpiece model.

The estimated points should be projected to be on the work-piece scanned model using the numerical method that is related to find the nearest mesh and interpolate the best estimation of it. Then, the marked area of the workpiece surface is known.

Step 4: Execute the taping process.

At this point, the system has enough information from the user to begin the masking task as usual. The diagram in Figure 62 summarizes the entire process of a taping task.

Figure 62: Workflow of the taping algorithm

For illustration purposes, the system supports three basic taping patterns namely, square, circular and spiral. The system will guide the user throughout the whole taping process which is arranged into stages. This step by step guide is simple to understand and therefore enable us to achieve our goal of allowing a novice user to use the robot efficiently. Figure 63, 64 and 65 provide a rough outline of the workflow of taping a spiral, square and circular pattern onto a workpiece respectively.

Figure 63: Workflow of square taping

Figure 64: Workflow of spiral taping

Figure 65: Workflow of circular taping

The objectives of this case study are as follow:

- To demonstrate the usefulness of the see-through augmented reality module in reducing the workload for the main user when interacting with the taping robotic platform and in simplifying the procedure for the new users.
- To demonstrate the use of the laser writer as a feedback system for the robot to return visual information of the task to the users.
- To demonstrate the usability of our multimodal input device in a different industrial context.

5.2.3 Results

This section provides some implementation results for applying the proposed framework to perform the square taping process. The user is equipped with our implemented interaction modules to complete the square taping task. With the help of the see-through glass, the graphical instructions are displayed at the marker position. This approach will allow the system to provide the operator instructive information to complete the taping process. The laser writer has shown great potential in providing visual feedback for confirmation purposes from robot to the operator while he can continuously monitor the taping area of the workpiece. The square taping task is described in detail in the figure below. The process begins with the user being asked to select the desired task for the robot to perform using the corresponding finger button on the handheld device. Here, he pressed an index finger to choose the taping

task. The system then suggests taping methods that are available to the robot. For demonstration purpose, we choose to tape a square area and pressed index finger button to confirm. Next, since the system normally requires a starting and an ending point to identify the square taping area, the system is asking the user to identify a starting point and the user selecting the starting point by using the laser pointer. The system confirms the user's starting point selection by projecting a small red circle at the user selected position using the laser writer. This provides intuitive and important feedbacks to the user to confirm his command. Similarly, the user can identify the endpoint for the taping process with the green laser pointer while the laser writer projects these points on the workpiece for confirmation. The system projects a laser image showing the area of the workpiece which is going to be taped while the see-through glass says that the taping process is starting. Finally, the robot arm proceeds the taping process by masking the white tape on the highlighted area.

<u>Operator Action with the MIM</u>		<u>Contents displayed by AR Glasses</u>	<u>Taping Robot Action</u>
<p>The main operator</p>	(2) Read the task's suggestions from the robot via the AR glasses		(1) Send taping's instructions to the operator and ask the operator for the desired task for the robot to perform using the corresponding input from the MIM.
	(3) Pressed the index finger to choose the taping task		(4) Suggests taping methods that are available to perform to the operator
	(5) Choose to tape a square area and pressed index finger button to confirm		(6) As defining a square area require a starting and an ending point, ask the user to identify a starting point first
	(7) Selecting the starting point by moving the laser pointer's output on the workpiece		(8) Tracking the operator's input with the vision system
			<p>The taping robot</p>

Figure 66. The interaction between human-robot for the taping task from selecting the taping method to identifying the starting point of the taping area

<u>Operator Action with the MIM</u>		<u>Contents displayed by AR Glasses</u>	<u>Taping Robot Action</u>
<p>The main operator</p>	(9) Read the task's suggestions from the robot via the AR glasses		(10) Confirms the user's starting point selection by projecting a small red circle at the selected position using the laser writer
	(12) Select the top-right corner with the laser pointer and confirm the location with index button		(11) Asking for the top right corner of the square area
	(14) Confirm with the robot if the position of the taping area is correct		(13) Confirm the user selection by projecting another small red circle at the selected position
	(16) Monitor the taping process from the robot and intervene if necessary		(15) Projecting a laser square image for highlighting the taping area while change the status of the robot in the augmented reality app.
			<p>The taping robot</p>

Figure 67. The interaction between human-robot for the taping task from identifying the endpoint of the taping area to highlighting the taping area on the workpiece

Figure 68. The operator is monitoring the taping task performed by the robot and the result

This demonstration has shown the capability of our proposed interface to enhance the user experience with robot taping system by allowing instruction information to be overlaid over the real working scenario which can greatly improve the user-friendliness of this human-robot interface.

5.3 Discussion

From the implementation of the two case studies, our framework has shown its ability to overcome the limitations of previous works so that we can achieve the targets mentioned in chapter 1. The multiple-channel configuration with the combination of see-through augmented reality and spatial augmented reality provide many advantages for this type of interface. It allows information to be divided into different channels for different target users. Also, with the help of the wearable device, we can generalize our users into three main groups with different needs. The setup also makes it easy to extend the framework for the case of multiple robots sharing the same task or same working environment. Augmented Reality is a great technology for improving the role of the robot in the relationship with human since it allows the robot to suggest task instructions and options to users. This can make the robot to be a partner or an assistant to a human in this scenario. Laser writer has shown its strength in generating bright and high-contrast digital graphics regardless of the environmental background. With the multimodal input device, the user can not only send different types of input command for different tasks but also can combine these input modalities together to communicate more effectively with the robots. Augmented Reality also has great advantages in interacting with multiple robot configurations. The ability to differentiate multiple robots with the help of the tracking module can really help to simplify the system configuration.

Chapter 6

Contribution, Limitations, and Future Works

This chapter summarizes some of the achievements we accomplished in designing and implementing the proposed framework in human-robot interaction field with respect to the proposed design targets. Also, some limitations of our novel framework are also carefully investigated to allow future research undertakings.

6.1 Contributions

The focus of this dissertation is to study and develop a new interaction framework for human-robot in the current industrial environment. By applying the indirect interaction approach, a new communication model is proposed which contains three main modules including the task suggestion & task visualization, multimodal input, spatial task sharing. The details of the framework with the contributions of each module are discussed as follows:

6.1.1 A New Interaction Framework for Human-Robot

Figure 69. A New Interaction Framework for Human-Robot with all the participants and the supporting modules

Our proposed framework has identified and categorized users who may interact with or share the working environment with the robot in the industrial contexts into three main group including the operators, the bystanders, and the passers-by. The operators communicate directly with the robot to collaborate or to control the robot for a specific task. The bystanders are users who are only interested in monitoring the operation of the robot or discussing with the main operator about the current robot's

task. Finally, the passersby who are not interested in the task but must share the same working area. Their concern is only the robot operation status or its next actions to avoid any possible collisions. Basing on the interaction interest of each group, we have been able to develop three modules to satisfy their demands. While the MIM is specially designed for the operator to providing commands to the robot, the TSTV module can be accessed by both the operator and the bystanders for their collaboration or discussion about the task or robot information. On the other hand, the passers-by can only be informed of the robot's operation via the STSM which is normally attached to the robot. The setup is not only for support the user's need but also help the user to feel more comfortable during the collaboration process as well as improving the information security of the robotic system.

Our framework can also be considered as an improvement to the current framework for "indirect interaction interface" in the field of human-robot interaction. This type human-robot interface only deploys a standard projector as a mediating object to facilitate the interaction between human and robot. The term "indirect" is used to refer to this mediator and to contrast with the direct approach where humanoid robots are specially designed to directly communicate with a human. Although the configuration of this framework has been shown to increase the understanding between human and robot, it contains many limitations that discourage its applications to many fields. Prior to the introduction of our proposed framework, the indirect interface supports only one channel of communication between human and robot. This only allows the framework to support the robotic system to interact with one user at a time. This prohibits the deployment of the interface in various applications, fields, and contexts where the capability of interacting with multiple users with different purposes at the same time is an important feature. A single mediating object model cannot guarantee the safe operation of the implemented interface. With one channel of communication, the interface cannot have a backup for emergency cases which means that the reliability of the framework is not ensured. Additionally, all the users can get accessed to the system configuration via this channel which means that the system security is not protected. These limitations have already been solved with our system. The detailed contributions of each module are provided in the sections below.

6.1.2 A Task Suggestion and Visualization Module for lowering Human Workload

This module concentrates on creating the most secure information communication channel between the operator and the robot as the robot configurations and parameters can only be seen via this channel. Furthermore, by taking advantage of the see-through augmented reality technology features, task configuration process becomes much simpler for the user using the augmented graphical information suggested by the robot. The interaction model provided by the module has shifted the role of human and robot compared to the classical model when the robot was a “slave” and must obey all the commands from the user. In this model, human and robot are partners where the robot suggests options and the user selects the most appropriate choice to configure or set up the task. Thus, users should be able to complete the task without mastering expert knowledge as the task procedure is provided by the robot via options and detailed instructions. This interaction paradigm between human-robot will become more popular in the incoming trend in the robotic field called “industrial 4.0” era.

Figure 70. The TSTV module with its functionalities for helping the operator during the interaction.

The computer graphics generated by the integrated augmented reality system also make the feedback information from the robot more interactive and easy to understand while keeping the information safe. The operator now can totally focus on the task scene while controlling the robot with this see-through technology supported by the module.

6.1.3 A New Multimodal Input Module for Industrial Contexts

Figure 71. The multimodal Input Module with multiple input modalities

To solve the problem of wearing protective equipment for industrial users, a new multimodal input module is developed and equipped for the operator. The module is utilized as a replacement for the traditional input devices such as computer mouse and keyboards which are not suitable for the industrial contexts. The multimodal input module introduces the integration of multiple user-friendly input modalities for sending user's commands to the interface including a laser pointer, haptic buttons, gesture mapping function, and a laser range sensor. This allows the operator to select the most appropriate inputs model for different tasks as well as to increase the

reliability of the input signal during the interaction with the robot. This improves the flexibility of the users in communication with the robot since they can choose different input models for different tasks. A task might require the user to utilize a combination of different input modalities to define the task's strategy which may be impossible to achieve with the single input device. The device also serves as another communication channel which can help us to differentiate the group of operators from the group of bystanders in the framework.

6.1.4 A New Spatial Task Sharing Module with Laser Projection Technology

Figure 72. The Spatial Task Sharing Module using laser source for generating very high-resolution images for the robot the share its actions with the passers-by

Prior to the implementation of our interface, the traditional indirect interfaces for human-robot interaction project interactive information to the real environment using standard LED projectors. This kind of projector can only render digital images with limited brightness and requires a low-light environmental condition to become effective. These projected images will become blurry and very difficult to see especially in a bright environment. As a result, it is difficult or even impossible to deploy the interface in outdoor tasks. The problem can be solved with our new spatial augmented reality module for task sharing purposes using laser projection

technology. The dazzling images generated by this module is essential to further increase the application range of our human-robot interaction framework. The combination of this spatial augmented reality task sharing module with the see-through augmented reality task suggestion & visualization module is the most important contribution of the frame for providing a secure, reliable and user-friendly communication model between human and robot. Also, from the previous work on evaluating a human-robot interface such as [61], the following features including (1) intuitive feedback, (2) high security, (3) robustness are essential for any interactive interface between human and robot which have also been satisfied by our module combination.

6.2 Limitations

In the current state, the framework still contains some avenues for future work. Although all the module is detailed described and implemented, a detailed experimental study on user heuristic evaluation is needed for a solid evaluation on the effectiveness of our framework. However, due to the limitations of manpower, the experiment cannot be performed at the time of writing the thesis. The study also needs to be extended to evaluating the current design of the handheld controller prototype as we are receiving some ideas regarding its size, weight, functionality so that the device can receive a popular acceptance in the industry.

As for the laser writer, although the ability to overlay high-contrast laser graphical images over the working environment is a strong advantage, the capability to perform automatic distortion correction is very important to consider and add in the future work. Now, the system can only perform distortion correction for a specific environment background. The calibration process is supposed to be performed in case there is a change in the projection surface. In the future, for example, when the mobile robot is smarter and can navigate automatically in a large industrial site, it may need to project its intention and task information to different places and for different people. The ability to perform image distortion correction by itself using the related surrounding environment information is an important feature.

6.3 Future Works

One important future work is to focus on designing and performing a heuristic evaluation of the proposed interface. For the study, about 20 people with different backgrounds are needed. They are supposed to interact with the interface and complete a task with the robot. The task will be performed via our new interface and one traditional interface such as the traditional controller for comparison purposes. The data collected will be used for analysis to demonstrate the advantage in term of mental workload (NASA-TLX score) or processing time. As for the handheld controller, it will be optimized in term of size, weight, shape, and functionalities to be accepted acceptance in the industrial environment. As for the laser writer, the automatic distortion correction capability will be implemented using a system including a camera, a laser scanner, and a laser writer. The system will support the robot to automatically identify the most suitable surface for projection with respect to the distance between the robot and the human to inform them of its next actions.

Another future work is to extend the application of the framework to other fields such as shipping industry, construction or manufacturing. For example, the below image illustrates a setup of the framework in the context of controlling a robot attached to a ship hull. In this scenario, the robot with the attached laser writer allows the user to have the ability to project, edit, customized the name of the ship before inscribing the real name to the ship hull using a robot. The full framework is shown in the figure below.

Figure 73: Using our framework to control the cleaning robot on a ship hull

Figure 74: Multimodal interface with augmented reality in a construction site

Additionally, the framework can be extended to the process of controlling the climbing robot to weld, clean or paint the ship hull. The operator can use the laser pointer to highlight the area of the hull while the laser writer confirms this command by projecting laser graphics to the corresponding position. The rest of the task is automatically performed by the robot. Next, this framework can also be applied in the context of a construction site as shown in the figure above. Normally, a big crane is controlled by an operator to lift and move a pile of material such as steel or iron to a predetermined location in a construction site. This process becomes very dangerous if the operator and the workers at the site do not pay enough attention. Our framework will be able to increase the safety level of the whole process by using the laser writer to project laser graphics to illustrate the path the crane is going to take while moving building materials. The framework can be extended to allow multiple workers to wear the handheld device to determine the target for moving the crane using the laser pointer.

References

- [1] MIG Welding – TIG Welding Retrieved Aug 24, 2017, from <http://machiningservicesinc.com/mig-welding-and-tig-welding-shop/>
- [2] Tom's Guide et Avram Piltch (2016, Jan 5). Projector Robot Rolls Video Around the House. *Tech Media Network*. Retrieved Aug 24, 2017, from <https://www.yahoo.com/tech/projector-robot-rolls-video-around-151330979.html>
- [3] Luke Hopewell (2013, Aug 15). Meta Smart Glasses: Meet The Australian Who Wants To Topple Google Glass. [Web log post] Retrieved Aug 24, 2017, from <https://www.gizmodo.com.au/2013/08/behind-meta-smart-glasses-how-a-silicon-valley-start-up-wants-to-topple-google-glass/>
- [4] H. C. Fang, S. K. Ong, and A. Y. C. Nee. 2012. Interactive robot trajectory planning and simulation using Augmented Reality. *Robot. Comput.-Integr. Manuf.* 28, 2 (April 2012), 227-237. DOI=<http://dx.doi.org/10.1016/j.rcim.2011.09.003>
- [5] Leutert, Florian & Schilling, Klaus. (2015). Augmented Reality for Telemaintenance and -inspection in Force-Sensitive Industrial Robot Applications. *IFAC-PapersOnLine*. 48. 153-158. 10.1016/j.ifacol.2015.08.124.
- [6] M. Gianni, G. Gonnelli, A. Sinha, M. Menna and F. Pirri, "An Augmented Reality approach for trajectory planning and control of tracked vehicles in rescue environments," 2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Linköping, 2013, pp. 1-6. doi: 10.1109/SSRR.2013.6719360
- [7] Palmarini, Riccardo & Roy, Rajkumar & Roy, Rajkumar & Torabmostaedi, Hosein. (2018). A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*. 49. 215-228. 10.1016/j.rcim.2017.06.002.
- [8] Virtual X-ray in the service-training. Retrieved Aug 24, 2017, from <http://www.tracking-challenge.de/content/tc/content/en/Object-Tracking/Applications.html>
- [9] Keith Curtin (2017, Jan). Mixed Reality will be most important tech of 2017. *Insider*. Retrieved Aug 24, 2017, from https://thenextweb.com/insider/2017/01/07/mixed-reality-will-be-most-important-tech-of-2017/#.tnw_vihFiACH
- [10] Milgram, Paul & Takemura, Haruo & Utsumi, Akira & Kishino, Fumio. (1994). Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies*. 2351. . 10.1117/12.197321.
- [11] J. Park and G. J. Kim, "Robots with projectors: An alternative to anthropomorphic HRI," 2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI), La Jolla, CA, 2009, pp. 221-222.
- [12] T. Matsumaru, "Mobile Robot with Preliminary-announcement and Display Function of Forthcoming Motion using Projection Equipment," ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, Hatfield, 2006, pp. 443-450.doi: 10.1109/ROMAN.2006.314368
- [13] The Ultimate Guide to Augmented Reality Technology. Retrieved Aug 24, 2017, from <http://www.realitytechnologies.com/mixed-reality>

References

- [14] The Father of Virtual Reality: Morton L. Heilig (1926-1997). Retrieved Aug 24, 2017, from <http://www.mortonheilig.com/>
- [15] Myron W. Krueger, Thomas Gionfriddo, Katrin Hinrichsen, "Videoplace—an artificial reality", Proceeding CHI '85 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Pages 35-40.
- [16] Tom Caudell. AR at Boeing (1990). [Web log post]. Retrieved Aug 24, 2017, from <http://www.idemployee.id.tue.nl/g.w.m.rauterberg/presentations/hci-history/tsld096.htm>
- [17] Cathy (2011, Aug 20). The History Of Augmented Reality. [Web log post]. Retrieved Aug 24, 2017, from <http://www.theopticalvisionsite.com/history-of-eyewear/the-history-of-augmented-reality/>
- [18] H. Kato, and M. Billinghurst, "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System", Proceedings of IWAR 99, pp. 85-94, 1999.
- [19] <http://wearables.unisa.edu.au/projects/arquake/>
- [20] Mathias Mohring, Christian Lessig, and Oliver Bimber. 2004. Video See-Through AR on Consumer Cell-Phones. In Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '04). IEEE Computer Society, Washington, DC, USA, 252-253. DOI=<http://dx.doi.org/10.1109/ISMAR.2004.63>
- [21] Tom Warren (2012, May 8). Nokia releases City Lens augmented reality beta app for Lumia Windows Phones. *The Verge*. Retrieved Aug 24, 2017, <https://www.theverge.com/2012/5/8/3006816/nokia-city-lens-augmented-reality-lumia-windows-phone>
- [22] Stapleton, C. B. & Hughes, C. E. (2005). Mixed Reality and Experiential Movie Trailers: Combining Emotions and Immersion to Innovate Entertainment Marketing. Proceedings of 2005 International Conference on Human-Computer Interface Advances in Modeling and Simulation (SIMCHI'05), New Orleans, January 23-27, 2005, 40-48.
- [23] MR MOUT: "Extreme MR: Going Beyond Reality to Create Extreme Multi-Modal Mixed Reality for Entertainment, Training and Education," Third IEEE and ACM International Symposium on Mixed and Augmented Reality, 2004, pp. 306-306. doi: 10.1109/ISMAR.2004.31
- [24] A. Steiger, B. Hein and H. Woern, "A Real-TimeWearable Projector-Wiimote-System for Augmented Reality Interaction Scenarios on Plane Objects," ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics), Munich, Germany, 2010, pp. 1-6.
- [25] Reitmayr, Gerhard and Dieter Schmalstieg. "Collaborative Augmented Reality for Outdoor Navigation and Information browsing." (2003).
- [26] Vespa-Augmented Reality. Retrieved Aug 24, 2017 from <http://vrmarketingdb.com/2017/03/16/vespa-augmented-reality-app/>
- [27] Caroline Siede (2014, Aug, 29). Beam me up: A beginner's guide to the Star Trek franchise. [Web log post]. Retrieved Aug 24, 2017 from <http://tv.avclub.com/beam-me-up-a-beginner-s-guide-to-the-star-trek-franchi-1798271640>
- [28] Adam Tuliper (2016, Nov). HoloLens - Introduction to the HoloLens. Retrieved Aug 24, 2017 from <https://www.microsoft.com/en-us/hololens>

References

- [29] Gary Sims (2015, Sep, 4). Epson Moverio BT-200 smart glasses review. [Web log post]. Retrieved Aug 24, 2017 from <http://www.androidauthority.com/epson-moverio-bt-200-review-636996/>
- [30] F. Sauer, Ali Khamene, B. Bascle, and G. J. Rubino. 2001. A Head-Mounted Display System for Augmented Reality Image Guidance: Towards Clinical Evaluation for iMRI-guided Neurosurgery. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, Wiro J. Niessen and Max A. Viergever (Eds.). Springer-Verlag, London, UK, UK, 707-716.
- [31] Kevin Kassil and A. James Stewart. 2009. Evaluation of a tool-mounted guidance display for computer-assisted surgery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1275-1278. DOI: <https://doi.org/10.1145/1518701.1518892>
- [32] Yuta Itoh and Gudrun Klinker. 2015. Vision enhancement: defocus correction via optical see-through head-mounted displays. In *Proceedings of the 6th Augmented Human International Conference (AH '15)*. ACM, New York, NY, USA, 1-8. DOI: <http://dx.doi.org/10.1145/2735711.2735787>
- [33] Hitesh Nidhi Sharma, Zachary O. Toups, Ajit Jain, and Andruid Kerne. 2015. Designing to Split Attention in a Mixed Reality Game. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)*. ACM, New York, NY, USA, 691-696. DOI: <https://doi.org/10.1145/2793107.2810289>
- [34] Daisuke Ochi, Kenta Niwa, Akio Kameda, Yutaka Kunita, and Akira Kojima. 2015. Dive into Remote Events: Omnidirectional Video Streaming with Acoustic Immersion. In *Proceedings of the 23rd ACM international conference on Multimedia (MM '15)*. ACM, New York, NY, USA, 737-738. DOI: <https://doi.org/10.1145/2733373.2807963>
- [35] R. S. Andersen, S. Bøgh, T. B. Moeslund and O. Madsen, "Intuitive task programming of stud welding robots for ship construction," 2015 IEEE International Conference on Industrial Technology (ICIT), Seville, 2015, pp. 3302-3307, doi:10.1109/ICIT.2015.7125587.
- [36] T. Sasai, Y. Takahashi, M. Kotani and A. Nakamura, "Development of a guide robot interacting with the user using information projection — Basic system," 2011 IEEE International Conference on Mechatronics and Automation, Beijing, 2011, pp. 1297-1302. doi: 10.1109/ICMA.2011.5985849
- [37] R. T. Chadalavada, H. Andreasson, R. Krug and A. J. Lilienthal, "That's on my mind! robot to human intention communication through on-board projection on shared floor space," 2015 European Conference on Mobile Robots (ECMR), Lincoln, 2015, pp. 1-6. doi: 10.1109/ECMR.2015.7403771
- [38] Michael D. Covert, Tiffany Lee, Ivan Shindeev, Yu Sun, "Spatial augmented reality as a method for a mobile robot to communicate intended movement," *Computers in Human Behavior*, vol. 34, May. 2014, pp. 241–248, doi:10.1016/j.chb.2014.02.001.
- [39] The First Wave of Portables. Retrieved Aug 24, 2017 from <http://www.computerhistory.org/revolution/story/316>
- [40] R.I.P. Palm: A History of the Smartphone/PDA Pioneer. Retrieved Aug 24, 2017 from <https://www.pcmag.com/feature/260439/r-i-p-palm-a-history-of-the-smartphone-pda-pioneer/1>
- [41] Cyberglove. Retrieved Aug 24, 2017 from <https://www.virtalis.com/blogs/products/cyberglove-systems/>

References

- [42] John Biggs (2009, Jun, 15). Video Review: Hillcrest Labs Loop Pointer. Retrieved Aug 24, 2017 from <https://techcrunch.com/2009/06/15/video-review-hillcrest-labs-loop-pointer/>
- [43] Atlas-The World's Most Dynamic Humanoid Retrieved Aug 24, 2017 from <https://www.bostondynamics.com/atlas>
- [44] Lee Ann Obringer & Jonathan Strickland. How ASIMO Works. Retrieved Aug 24, 2017 from <http://science.howstuffworks.com/asimo.htm>
- [45] Zlotowski, J., Proudfoot, D., Yogeewaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human-Robot Interaction. *International Journal of Social Robotics*, 7(3), 347-360.
- [46] Massimiliano ZECCA, Research on Emotional Humanoid Robotics: a good excuse to study yourself. Retrieved Aug 24, 2017 from http://www.yomiuri.co.jp/adv/wol/dy/campus/miscellany_1101.html
- [47] R2-D2. Retrieved Aug 24, 2017 from <http://www.starwars.com/databank/r2-d2>
- [48] J. H. Seo, J. Y. Yang and D. S. Kwon, "Laser scanner based foot motion detection for intuitive robot user interface system," 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, 2012, pp. 179-184. doi: 10.1109/ROMAN.2012.6343750
- [49] F. Leutert, C. Herrmann and K. Schilling, "A Spatial Augmented Reality system for intuitive display of robotic data," 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Tokyo, 2013, pp. 179-180. doi: 10.1109/HRI.2013.6483560
- [50] T. Noguchi, M. Kasetani, H. Yamazoe and J. H. Lee, "Proposal of a behavior model for information support with Ubiquitous Display," 2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Goyang, 2015, pp. 427-432. doi: 10.1109/URAI.2015.7358895
- [51] E. Kwon and G. J. Kim, "Humanoid robot vs. projector robot: Exploring an indirect approach to human robot interaction," 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Osaka, 2010, pp. 157-158. doi: 10.1109/HRI.2010.5453216
- [52] Krämer N.C., von der Pütten A., Eimler S. (2012) Human-Agent and Human-Robot Interaction Theory: Similarities to and Differences from Human-Human Interaction. In: Zacarias M., de Oliveira J.V. (eds) *Human-Computer Interaction: The Agency Perspective. Studies in Computational Intelligence*, vol 396. Springer, Berlin, Heidelberg
- [53] Weiss A., Mirnig N., Buchner R., Förster F., Tscheligi M. (2011) Transferring Human-Human Interaction Studies to HRI Scenarios in Public Space. In: Campos P., Graham N., Jorge J., Nunes N., Palanque P., Winckler M. (eds) *Human-Computer Interaction – INTERACT 2011. INTERACT 2011. Lecture Notes in Computer Science*, vol 6947. Springer, Berlin, Heidelberg
- [54] Torta E., van Heumen J., Cuijpers R.H., Juola J.F. (2012) How Can a Robot Attract the Attention of Its Human Partner? A Comparative Study over Different Modalities for Attracting Attention. In: Ge S.S., Khatib O., Cabibihan JJ., Simmons R., Williams MA. (eds) *Social Robotics. ICSR 2012. Lecture Notes in Computer Science*, vol 7621. Springer, Berlin, Heidelberg
- [55] Rigoll G. (2015) Multimodal Human-Robot Interaction from the Perspective of a Speech Scientist. In: Ronzhin A., Potapova R., Fakotakis N. (eds) *Speech and Computer. SPECOM 2015. Lecture Notes in Computer Science*, vol 9319. Springer, Cham

- [56] Kulms P., Kopp S. (2016) The Effect of Embodiment and Competence on Trust and Cooperation in Human-Agent Interaction. In: Traum D., Swartout W., Khooshabeh P., Kopp S., Scherer S., Leuski A. (eds) *Intelligent Virtual Agents. IVA 2016. Lecture Notes in Computer Science*, vol 10011. Springer, Cham
- [57] Randelli, G., Bonanni, T.M., Iocchi, L. et al. *Intel Serv Robotics* (2013) 6: 19. <https://doi.org/10.1007/s11370-012-0123-1>
- [58] Schneider M., Deml B. (2017) Analysis of a Multimodal Human-Robot-Interface in Terms of Mental Workload. In: Schlick C. et al. (eds) *Advances in Ergonomic Design of Systems, Products and Processes*. Springer, Berlin, Heidelberg
- [59] Mocan B., Fulea M., Brad S. (2016) Designing a Multimodal Human-Robot Interaction Interface for an Industrial Robot. In: Borangiu T. (eds) *Advances in Robot Design and Intelligent Control. Advances in Intelligent Systems and Computing*, vol 371. Springer, Cham
- [60] Trouvain B., Schlick C.M. (2008) Design and evaluation of a multimodal human-multirobot interface. In: Schmidt L., Grosche J., Schlick C.M. (eds) *Ergonomie und Mensch-Maschine-Systeme*. Springer, Berlin, Heidelberg
- [61] Lee K., Hwang JH., Kwon DS. (2008) A Comparative Study between Human-Human Interaction and Human-Robot Interaction. In: Nguyen N.T., Katarzyniak R. (eds) *New Challenges in Applied Intelligence Technologies. Studies in Computational Intelligence*, vol 134. Springer, Berlin, Heidelberg
- [62] Bannat A., Gast J., Rehrl T., Rösel W., Rigoll G., Wallhoff F. (2009) A Multimodal Human-Robot-Interaction Scenario: Working Together with an Industrial Robot. In: Jacko J.A. (eds) *Human-Computer Interaction. Novel Interaction Methods and Techniques. HCI 2009. Lecture Notes in Computer Science*, vol 5611. Springer, Berlin, Heidelberg
- [63] Elmogy, M., Habel, C. & Zhang, J. *Cogn Process* (2011) 12: 53. <https://doi.org/10.1007/s10339-010-0386-4>
- [64] Abich, J. & Barber, D.J. *J Multimodal User Interfaces* (2017) 11: 211. <https://doi.org/10.1007/s12193-016-0237-4>
- [65] Prodanov, P., Drygajlo, A., Richiardi, J. et al. *Intel Serv Robotics* (2008) 1: 3. <https://doi.org/10.1007/s11370-006-0001-9>

APPENDIX

Usability Testing

Whatever the technology you create, it is important to listen to feedbacks from the users in order to evaluate and improve your product. As a result, we have designed a usability test to find out whether the system is user-friendly, intuitive and is easy to learn. The usability test includes the following steps:

- Robot Navigation Test.
- Online Surveys.

A. Robot Navigation Test

The test is conducted with 20 participants that are currently undergraduate and graduate students from School of Mechanical/Aerospace Engineering. First, the pilot test requires each participant to navigate the mobile robot around traffic cones using our implemented interface as shown in the following figure.

Figure 75: Mobile robot navigation pilot test

For each participant, three time periods are recorded for later analysis. The first is the amount of time it took for them to trigger the robot's movements, to get through the screens necessary to select the task (Navigation/Delivery), the mode of control (Manual/Automatic) and the start of the control. This time indicates how user-friendliness the augmented reality interface is to first-time users and how much cognitive work they must do to be able to perform even the most basic task. The next period of time is relevant to the time to navigate the robot around the traffic cones. This measurement shows how easy the handheld controller is to control the robot and whether the AR glasses instructions are easy to understand. Finally, all participants were asked to perform the same navigation task using the joystick controller which

is a traditional device to control the robot instead of the handheld controller for comparison purposes.

Figure 76: Joystick Controller

Given below are the timings the 20 subjects registered during the test.

Table 4: Usability test timings

No.	Name	Demo	Start	End	Joystick
1	Huynh	36 sec	20 sec	2:24 min	55 sec
2	Hieu	34 sec	32 sec	2:03 min	35 sec
3	Tra	54 sec	1:18 min	2:46 min	54 sec
4	Tarshan	Expert user	13 sec	1:01 min	40 sec
5	Aditya	38 sec	52 sec	1:36 min	50 sec
6	Hendra Suratno Tju	20 sec	1:01 min	2:20 min	35 sec
7	Karthikeyan Gururaj	20 sec	1:50 min	2:10 min	53 sec
8	Kok Yuan Yik	50 sec	31 sec	1:40 min	28 sec
9	Li Xin	40 sec	5:00 min	4:02 min	1:20 min
10	Huang Yanpei	50 sec	3:30 min	3:50 min	1:40 min
11	Le Huu Minh	35 sec	1:54 min	3:30 min	1:00 min
12	Ehsan Asadi	40 sec	2:50 min	2:50 min	1:05 min
13	Ahmad	20 sec	50 sec	2:15 min	1:08 min
14	Wong Choon Yue	20 sec	1:00 min	2:06 min	34 sec
15	Vasanth Elangovan	40 sec	2:50 min	2:00 min	33 sec
16	Burhan	15 sec	47 sec	1:43 min	22 sec
17	Iastrebov Viatcheslav	20 sec	1:16 min	1:44 min	17 sec
18	Benjamin Chabanne	20 sec	22 sec	1:45 min	1:04 min
19	Nguyen Anh	25 sec	37 sec	3:10 min	1:00 min
20	Bui Xuan Hien	24 sec	1:09 min	1:34 min	1:24 min

From the table data, the following plot can be drawn for comparison purposes.

Figure 77: Times taken by participants to complete given task

As shown in the figure, the amount of time to navigate the mobile robot using the handheld controller is more than it is with the joystick controller. This is due to the small number of inputs present on the controller and the familiarity of the participants with the controller in comparison to the handheld device. On the one hand, there are no visual feedbacks the joystick to the users to read and remember to manipulate the robot to complete the task. On the other hand, the participants keep receiving feedbacks from the smart goggles while using the hand-controller to maneuver the robot since all the instructions and visual feedbacks are displayed at the AR glasses. This has shown one main feature of using augmented reality interface to perform industrial tasks is that although graphical instructions and feedback enable new users to complete the task with ease, it may slow the performance time due to the large number of input modalities the operator has to handle at the same time.

B. Online Survey

To understand more about the user experience with the designed interface, an online survey is conducted which following the procedure given by Stanton et al. The objective of this survey is to collect data to analyze our interface system with regards to human factors. The following questions are crafted into the online survey:

- What is your name? Nationality? Age? Educational background?
- Have you used remote control devices before? Where and how often?

- Are you comfortable with the joystick type controller?
- Did you like the fit of the hand controller?
- Could you reach all the buttons?
- Could you tell that you had to tilt to turn from the beginning? How come?
- How much do you like tilt to turn?
- How heavy did the device feel to you?
- Do you think you can use this system for long periods of time?
- How did the goggles fit?
- Could you see your surroundings while using the goggles?
- How clearly could you see the instruction screens?
- Were the instructions easy to read?
- Were the instructions easy to understand?
- Do you think the robot responded to commands with enough speed?
- Any area of improvements in the control of the robot?
- What possible applications do you see for this system?
- How easy was it to control the robot?
- Which controller is easier to use? The joystick type or the hand controller?
- Over a period of time and exposure to this device, which would be easier to use?
- What did you like in this system?
- What didn't you like in this system?
- What would you like to improve or change in this system?

The online survey has the advantage of allowing a larger number of participants to complete the test at the same time while reducing the time it takes for each participant to take the test which makes the survey more open and easy to be evaluated. The following section will provide in details results from the conducted survey.

Results regarding personal information of the subjects

Nationality

20 responses

Age

20 responses

Figure 78: Demographic of respondents

We have participants come from different nations and their age ranges between 25 and 35 years of age.

Results regarding the handheld controller

Are you comfortable with the joystick type controller?

19 responses

Did you like the fit of the hand controller?

20 responses

Figure 79: Comfort of the two types of controllers

Most of the participants feel comfortable using the joystick controller. However, most users have a problem figuring out that they have to press the button at the bottom right corner of the joystick to transmit the command inputs to the robot motor. Also, it is important to notice that some user cannot figure out how to turn the robot which requires the user to rotate the stick button in the middle of the device. Most users naturally try to figure out how to control the robot by pushing the stick forward, backward, left or right. Majority of the respondents report that they like the way the handheld controller fit their forearm.

Could you reach all the buttons comfortably?

20 responses

How comfortable are you with the tilt-to-turn function?

20 responses

Figure 80: Control of the handheld controller

Most users can reach all the buttons on the handheld device and only three users cannot reach the little finger (2 users) or the thumb buttons (1 user). 40% of the users feel positive with the tilt-to-turn function while only 25% did not like the function.

How heavy did the device feel to you?

20 responses

Do you think you can use this device for a long period of time?

20 responses

Figure 81: Sustained use of the handheld controller

The majority of the participants think that the device is light while only 15% of the user found it heavy. There are 25% of the users think that they can use the device for a long period of time while 40% of the answers are undecided about the question. This data has shown that the hand-controller need more improvements to match with the user expectations.

Results regarding the AR glasses

How well did the goggles fit?

20 responses

How clearly could you see the instruction screens?

20 responses

Figure 82: Fit and visibility

According to the user feedback, the fit of the glasses is approximately average. This problem since Moverio BT-200 is the first version of see-through AR glasses from Epson. At the time of writing this thesis, they already released a newer version which is lighter and more comfortable to wear. Most users can clearly see the instruction screens in the see-through mode which add a positive point to the product from Epson.

Could you see your surroundings while using the goggles?

20 responses

Were the instructions easy to read?

20 responses

Were the instructions easy to understand?

20 responses

Figure 83: Peripheral visibility and comprehension with glasses

The survey shows positive feedback regarding peripheral vision and comprehension with the goggles. Most users found that they were able to see the surrounding environment while performing the tasks. The instructions are easy to read and understand for the majority of the participant. This is a very promising result for the interface since the glasses is one of the core ideas of our framework that uses

computer-generated graphics to suggest task's options to new users so that the human's workload can be reduced during the interaction process.

Results regarding the robot

Do you think the robot responds to commands quickly?

20 responses

How easy was it to control the robot?

20 responses

Figure 84: Control of the robot

As shown in the chart, most participants believe the robot responds quickly to the user's command. Also, 80% of the users give positive feedback on how easy to control the robot. Some users found that the turning controls were a little sensitive with respect to their hand gesture. Finally, some were not comfortable with the usage of the word "tilt" as they think the word is used to indicate move device to the left or

right within the same plane instead of using the pronation-supination motion of the wrist.

Results regarding the overall system

Which controller is easier to use?

20 responses

Over a period of time and exposure to the system, which device would be easier?

20 responses

Figure 85: Overall choice of controller

As expected, the majority of participants would prefer the joystick over the handheld controller. This is due to the simplicity and familiarity of the joystick controller as opposed to a new, revolutionary technique of control that uses emerging technologies such as augmented reality. However, although all of the participants agree that the device would be very useful and applicable to a large variety of applications, they prefer not to switch to the handheld controller. There might be several reasons for this. First, the user tends to stick to technology that they are familiar with and they

may be afraid to adopt a totally new technology to their daily work. Also, navigation task is a simple task and can be completed quickly using the traditional joystick. It would be more desirable to perform a usability testing with a more complex industrial task to show the users the versatility of our interface. However, the goal of the study is not to convince users to accept the new interface but to give us an idea what we would possibly face if we want to introduce our new interface to the market.