University of Kentucky
UKnowledge

Theses and Dissertations--Computer Science | Computer Science

2015

Towards Intelligent Telerobotics: Visualization and Control of Remote Robot

Bo Fu
University of Kentucky, ukyfubo@gmail.com

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Fu, Bo, "Towards Intelligent Telerobotics: Visualization and Control of Remote Robot" (2015). *Theses and Dissertations--Computer Science*. 40.
https://uknowledge.uky.edu/cs_etds/40

This Doctoral Dissertation is brought to you for free and open access by the Computer Science at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Computer Science by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Bo Fu, Student

Dr. Ruigang Yang, Major Professor

Dr. Miroslaw Truszczynski, Director of Graduate Studies

Towards Intelligent Telerobotics: Visualization and Control of Remote Robot

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy in the
College of Engineering at the
University of Kentucky

By
Bo Fu
Lexington, Kentucky

Director: Dr. Ruigang Yang, Professor of Computer Science
Lexington, Kentucky 2015

Copyright © Bo Fu 2015

ABSTRACT OF DISSERTATION

Towards Intelligent Telerobotics: Visualization and Control of Remote Robot

Human-machine cooperative or co-robotics has been recognized as the next generation of robotics. In contrast to current systems that use limited-reasoning strategies or address problems in narrow contexts, new co-robot systems will be characterized by their flexibility, resourcefulness, varied modeling or reasoning approaches, and use of real-world data in real time, demonstrating a level of intelligence and adaptability seen in humans and animals. The research I focused is in the two sub-field of co-robotics: teleoperation and telepresence.

We firstly explore the ways of teleoperation using mixed reality techniques. I proposed a new type of display: hybrid-reality display (HRD) system, which utilizes commodity projection device to project captured video frame onto 3D replica of the actual target surface. It provides a direct alignment between the frame of reference for the human subject and that of the displayed image. The advantage of this approach lies in the fact that no wearing device needed for the users, providing minimal intrusiveness and accommodating users eyes during focusing. The field-of-view is also significantly increased. From a user-centered design standpoint, the HRD is motivated by teleoperation accidents, incidents, and user research in military reconnaissance etc. Teleoperation in these environments is compromised by the Keyhole Effect, which results from the limited field of view of reference. The technique contribution of the proposed HRD system is the multi-system calibration which mainly involves motion sensor, projector, cameras and robotic arm. Due to the purpose of the system, the accuracy of calibration should also be restricted within millimeter level. The followed up research of HRD is focused on high accuracy 3D reconstruction of the replica via commodity devices for better alignment of video frame. Conventional 3D scanner lacks either depth resolution or be very expensive. We proposed a structured light scanning based 3D sensing system with accuracy within 1 millimeter while robust to global illumination and surface reflection. Extensive user study prove the performance of our proposed algorithm. In order to compensate the unsynchronization between the local station and remote station due to latency introduced during data sensing and communication, 1-step-ahead predictive control algorithm is presented. The latency between human control and robot movement can be formulated as a linear equation group with a smooth coefficient ranging from 0 to 1. This predictive control algorithm can be further formulated by optimizing a cost function.

We then explore the aspect of telepresence. Many hardware designs have been developed to allow a camera to be placed optically directly behind the screen. The purpose of such setups is to enable two-way video teleconferencing that maintains eye-contact. However, the image from the see-through camera usually exhibits a number of imaging artifacts such as low signal to noise ratio, incorrect color balance, and lost of details. Thus we develop a novel image enhancement framework that utilizes an auxiliary color+depth camera that is mounted on the side of the screen. By fusing the information from both cameras, we are able to significantly improve the quality of the see-through image. Experimental results have demonstrated that our fusion method compares favorably against traditional image enhancement/warping methods that uses only a single image.

KEYWORDS: Telerobotics, 3D Reconstruction, Multi-ystem calibration, Predictive Control, Robotic arm control

Author's signature: Bo Fu

Date: 05/06/2015

Towards Intelligent Telerobotics: Visualization and Control of Remote Robot

By
Bo Fu

Director of Dissertation: Ruigang Yang

Director of Graduate Studies: Miroslaw Truszczynski

Date: 05/06/2015

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Ruigang Yang. He has guided me into this field, inspired me to pursue my PhD on this topic and helped me on this thesis in too many ways to enumerate. I could not have reached this point without him.

I also want to give many thanks to my committee members, Dr. Fuhua Cheng, Dr. Yuming Zhang, Dr. Laurence Hassebrook, Dr. Melody Carswell and Dr. Nathan Jacobs, for the time they spent on my thesis review. I appreciate all of the suggestions and comments.

My deepest gratitude goes to the people who have had such a significant impact on my life. My parents encouraged me to enter a Ph.D. program in the United States. They have always given me unconditional support and love.

Dedicated to my beloved parents and my girlfriend.

TABLE OF CONTENTS

Acknowledgments iii

Table of Contents iv

List of Figures vi

List of Tables ix

Chapter 1 Introduction 1
  1.1 Motivation and Goals 1
  1.2 A brief note on the research 5
  1.3 Innovation 8
  1.4 Dissertation Outline 9

Chapter 2 Background and Related Work 10
  2.1 Structured Light 3D Scanning 10
  2.2 Visual Display Technique 15
  2.3 Hand-eye calibration for Robotic Vision 18
  2.4 Human-Robot Interaction (HRI) 21
  2.5 Predictive Control and Robot Control 21
  2.6 Telepresence Enhancement 22

Chapter 3 Visualization and Monitoring of Remote Welding 24
  3.1 System Platform 27
  3.2 Multi-sensor Calibration 29
  3.3 Experiment and Usability Evaluation 42
  3.4 Conclusion 49

Chapter 4 Teleoperation of Remote Welding 50
  4.1 System Platform 51
  4.2 System Calibration 55
  4.3 Predictive Control of Robot Speed 56
  4.4 Experiment 61
  4.5 Conclusion 68

Chapter 5 Enhancement of Telepresence 70
  5.1 Overview 71
  5.2 Our Approach 72
  5.3 Experiments 79
  5.4 Conclusion 82

Chapter 6 Conclusion and Future Work 83

6.1 Contribution 83
6.2 Future Work 85

Chapter 7 Appendix 86
7.1 Camera Calibration 86
7.2 Motion Sensor Calibration 87
7.3 Work piece-projector Calibration 88

Bibliography 90

Vita 108

LIST OF FIGURES

1.1 The concept of co-robotics. 1
1.2 Model of operations and the associated research challenges. 4

2.1 The classification of existing structured light techniques. 12
2.2 The characteristics of display techniques. 18
2.3 The Concept of hand-eye calibration. 19

3.1 Virtualized Welding: Above: an illustration of our virtualized welding operation, in which existing welding robots are augmented with a video cameras to capture the working environment. The operator can monitor the welding process in an augmented display setup from differenct angles, in which welding images are projected on a mock-up 3D surface, as if he/she was right next to the actual welding. Below: our current implementation of virtualized welding. . 25
3.2 General structure of the virtualized welding system. 27
3.3 Overview of real welding workstation. The global view camera is facing towards the robotic arm and work piece. 28
3.4 Detailed view of the virtual welding workstation. Major components are a mock up, an auxiliary camera and a projector. The mockup is generated by 3D printer. 29
3.5 Calibration of the virtualized welding system. Projector, local view camera and robotic arm are calibrated together. 30
3.6 Triangulation via line-plane intersection between camera frame and projector frame. 33
3.7 Gray Code structured light sequences. The image represents the sequence of bit planes displayed during data acquisition. Image rows correspond to the bit planes encoding the projector columns, assuming a projector resolution of 1024 by 768, ordered from most to least significant bit (from top to bottom). . . 33
3.8 Decode result based on gray code structured light scan. Left: gray code illumination sequence. Right: decoded result. Color represent coordinate in projector cooridnate system. 34
3.9 Strong Interrefelction and sub-surface scattering result unfavourable errors in the recontructed model. 34
3.10 Visualization of different binary coding. Bottom two are logical XOR04 and logical XOR02 code respectively. 35
3.11 a) Object illumiated by a single pixel stripe pattern. b) Close up image on the marked region. 36
3.12 The pipeline of High Resolution Structured Light Scanning algorithm. 37
3.13 Overview of real welding workstation. The global view camera is facing towards the robotic arm and work piece. 38

3.14 Visualization of the work piece. (a) global view of the work piece; (b-d) local view of the work piece; (e-h) blank HRD and HRD of b-d; (i-l) observation of HRD from different view point; (m-o) global view of the working environment corresponding to b-d. The frames are cpatured without electric arc for visual purpose. 44
3.15 Screen shot of experiment for user study. Left: local view; Right: Global view; During study, the global view and local view are displayed on screen side by side. These two views are cropped and image-enhanced for viusal purpose in this paper. 44
3.16 Left: Mental Demand subscale score from the NASAS-RTLX rating scale. Mental Demand was rated on a 0 (low) - 100 (high) scale. Error bars are constructed using 1 standard error from the mean. Right: Graph of reported confidence of the users in their first response. Confidence rating were reported on a 1 (low confidence) 10 (high confidence) scale. 46
3.17 User estimate time allocation between displays for use in a monitoring task. . . 48

4.1 Virtualized Welding: Above: an illustration of our virtualized welding operation, in which existing welding robots are augmented with 3D sensors and video cameras to capture the working environment. The operator can monitor and control the welding process in an augmented display setup in which welding images are projected on a mock-up 3D surface. The operators motion is tracked. 50
4.2 General structure of the virtualized welding system. 51
4.3 Overview of the real workstation. 52
4.4 Detailed view of the virtual welding workstation. Major components are a mock up, a motion sensor, an auxiliary camera and a projector. The mockup is generated by 3D printer. 53
4.5 System work flow. 54
4.6 Calibration of the virtualized welding system. Projector, local view camera, motion sensor and robotic arm are calibrated together. 55
4.7 Sample human movement. 57
4.8 5-step-ahead prediction error versus smoothing coefficient. 58
4.9 5-step-ahead prediction coordinates. 58
4.10 5-step-ahead prediction errors. 58
4.11 weld pathes on the pipe. 62
4.12 Error of 3D scaned model, the difference between reconstructed model and ground truth is less than 1 mm. 63
4.13 Several patterns for tracking experiment. We designed four different type of pattern on the surface: straight line, quadratic curve, sine wave and broken line. Note that when applying these pattern on work piece, all of them turn to 3D curve. 63
4.14 Robot tracking performance of a sine wave using Leap Sensor in 3D space. . . 64
4.15 X-Y plane projection for Figure 4.14. 64
4.16 Tracking errors for Figure 4.14. 65

4.17 Three different control method. (a): user controls the robotic arm via motion sensor while viewing the HRD; (b): close up look of a; (c): user controls the robotic arm via 3D mouse while viewing the flat display device; (d): user controls the robotic arm via 3D mouse whiel viewing the HRD. 66
4.18 Tracking performance: straight line. 66
4.19 X-Y plane projection for Figure 4.18. 67
4.20 Tracking performance: sine wave. 67
4.21 X-Y plane projection for Figure 4.20. 67
4.22 Tracking performance comparison between different implementation. The result proves our implementation: 3D motion sensor combined with HRD could handle more complicated senario, 3D curve in our experiment. 68
4.23 Visualization on HRD during control welding. We use a laser dot in real welding workstaion to indicate where the weld gun is pointing at. User should use the laser dot in video feedback rendered on HRD as a spatial and 3D cue. (a,e,i): viewing blank HRD from different angle; (b-d): welding on straight line; (f-h): welding on curve; (j-l): welding on sine wave; (m-o): welding on line segments. 69

5.1 Sideview of of a typical teleconferencing system. 71
5.2 Our hybrid setup 73
5.3 An example of the *side-view image* I_s, *eye-view image* I_e. Note the large difference both in field of view and view angles between I_s and I_e, as well as the extreme low quality of I_e due to under-exposure and noises. 74
5.4 (a A warped sparse image I_w and (b the image after JBF. Note besides the large holes in (b that are due to view point variation, there are also some small holes after filtering. These are mainly due to inaccuracy of filtered depth map. 76
5.5 (a Enhanced *eye-view image* through histogram matching; (b Hole-filled image. 77
5.6 Detected face region and skin colo region, which is used for noise variance estimation. 78
5.7 Out processing pipeline. The green blocks are our inputs and the red block is our output; while all the rest are our intermediate processing modules 79
5.8 Comparison of noise variance estimations. 80
5.9 Comparison of our result (third column) with bilateral filtering [1](fourth column) and wavelet domain denoising [2] (fifth column). The first and second columns are images from *side-view camera*, i.e. Kinect, and *eye-view camera*, respectively. 81

7.1 Camera undistortion. a)-d) sample input images for undistortion toolbox; e) camera view before undistortion; f) camera view after undistortion. 87
7.2 Conept of motion sensor calibration. Left, coordinate system defined on motion sensor and mock up; Right, procedure of triangulation. 88
7.3 Conept of lookup table based calibration. At each control point, a 2D-3D projection is calculated based on cooresponding feature points. The interpolation is introduced between adjacent control points. 89

LIST OF TABLES

2.1 Equipment requirement comparison between 3 invisible structured light scanning techniques 15

3.1 Cameras and projector configuration. 41

4.1 Cameras and projector configuration. 54
4.2 Error Comparison Between Leap Sensor and 3D Mouse 68

Chapter 1 Introduction

1.1 Motivation and Goals

Human-machine cooperative, or co-robotics (Fig. 1.1 [3]) has been recognized as the next generation of robotics. In contrast to current systems that use limited-reasoning strategies or address problems in narrow contexts, new co-robot systems will be characterized by their flexibility, resourcefulness, varied modeling or reasoning approaches, and use of real-world data in real time, demonstrating a level of intelligence and adaptability seen in humans and animals. Research on relevant aspects of human cognition, perception, and action has the potential to be especially useful in this regard. This type of research may enhance the design of robotic systems by mimicking human learning, reasoning and action planning. This approach may also be helpful for designing co-robotic systems that will be able to fruitfully collaborate with humans. Thus, the research program is necessarily cross-disciplinary engaging basic research in the behavioral and social sciences, education, as well as computer science and engineering.

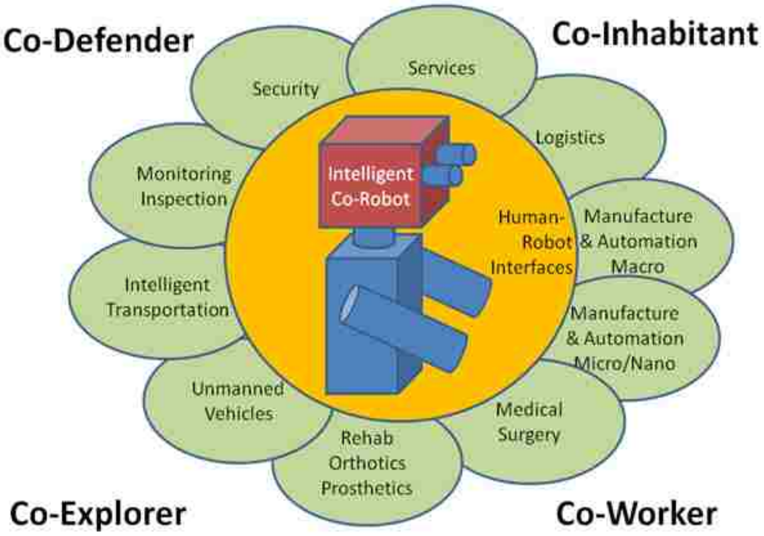


Figure 1.1: The concept of co-robotics.

Based on the promising outcome and expectation, national robotics initiative (NRI) is organized to support the relative research on co-robot systems. Multiple federal government including the National Science Foundation (NSF), NASA, NIH and USDA are participating in funding these researches.

We envision a new human-machine interactive paradigm that can transfer human intelligence to robots. An existing dumb robot will be augmented with sensors to observe the work piece, as well as its surroundings. These sensors are able to record and reconstruct the process in 3D, which includes the working environment, the pose of the torch, etc. The reconstructed data are transmitted to a control room and visualized with novel augmented reality techniques: A skilled controller can look at the process from different angles, as if he/she was right next to the actual work piece. Parameters can be adjusted by the human (with intelligence) and executed by the robot (with precision).

The objective of this proposal is to study novel ways of telerobotics by using mixed reality techniques. This effort can be fullfilled by establishing the software/hardware and robotic control algorithms foundation to allow a robot to combine its accurate motion control and physical strengths with the intelligence of a human controller through real-time human robot cooperation. Toward this goal of intelligent welding robots with autonomy, I propose an innovative robotic control platform that is capable of monitoring and control the remote robot activity using 3D imaging techniques and visualize the welding process remotely using augmented reality techniques, which is inspired by Takeo Kanada, who coined the term Virtualized Reality [4]. A traditional robot is augmented with a number of sensor s to acquire all the relevant information during movement, including torch position, work piece geometry, etc. These parameters are transferred to a visualization workstation to re-enact the process. The workstation uses an articulated arm with a video projector, which is referred to as the torch surrogated, to simulate the movement of the torch and the visual changes under it. Note that the projection surface, constructed from template pieces, is a proxy for the actual welding pieces. Furthermore, the surrogate can also be used as an input device to control the position of the robot.

One application of our proposed algorithm with significant social impact is tele-operational welding, which is a widely used in manufacturing process that is labor intensive and sometime hazardous. While industrial welding robots have been in use for several decades, they are pre-programmed actuators with limited, if any, intelligence. As a result welding robots are primarily used in well-controlled environments, such as assembly lines for mass production, in which the work pieces may be accurately prepared and positioned at reasonable costs. Given that manufacturing is moving towards more customized productions, the next generation of welding robots that can intelligently adjust to various welding tasks is urgently needed. Unfortunately, equipping robots with intelligence is challenging. Current welding robots are basically articulated arms with a pre-programmed set of movement. Although some robots are equipped with seam tracking capabilities, they all lack the intelligence skilled human welders possess and their adaptation to different welding conditions is limited. They require precision prepared work pieces with little variation in geometry and material properties. Therefore their applications are mostly limited to assembly lines for mass-produced products, such as automobiles. They are thus unable to produce a consistent weld bead in these situations. Rather than explicitly modeling the physical welding process in a parametric way, a data-driven approach that analyzes a rich set of welding process data will be an effective way to mimic the human intelligence in reaction to various welding tasks. In order to collect and utilize the human intelligence of professional welding experience, and in the meanwhile, enable direct human-robot interaction for better user experience, developing a virtual welding platform which integrates data collection, virtual reality and human-machine interaction is therefore necessity of that demanded.

The proposed Virtualized Teleoperation can be operated at three different levels: (1) Remote-controlling (teleportation). Similar to the well-known da Vinci surgical robot, now a welder can remotely operate the torch by controlling the surrogate in the visualization station, with the benefit of looking at the process from different viewpoints, as if the welder is right in front of the work piece. This would enable working in hazardous areas for human operators. (2) Co-operative (supervisory control). The human welder is mostly monitoring the progress of the welding process carried out by the pre-programmed welding robot. The welder can occasionally make adjustment by taking control of the surrogate. Different from teleportation, the adjustment is usually small and occasional. In this way, the precision of the welding robot and be combined with the intelligence of an experienced welder. (3) Autonomous. The robot is running in a fully automatic way under close-loop control. The control decision is based on the sensor input, such as 3D geometry of the weld pool and the work piece. This type of adaptive control is fundamentally more advanced than the current welding robots. At all three-levels (Fig. 1.2), a critical distinction of the proposed virtualized welding is the 3D reconstruction of the welding environment and recording of all welding parameters. This offers the opportunities for our robot to initially learn from human welders, and then eventually achieve full autonomy with capabilities exceeding experienced welders.

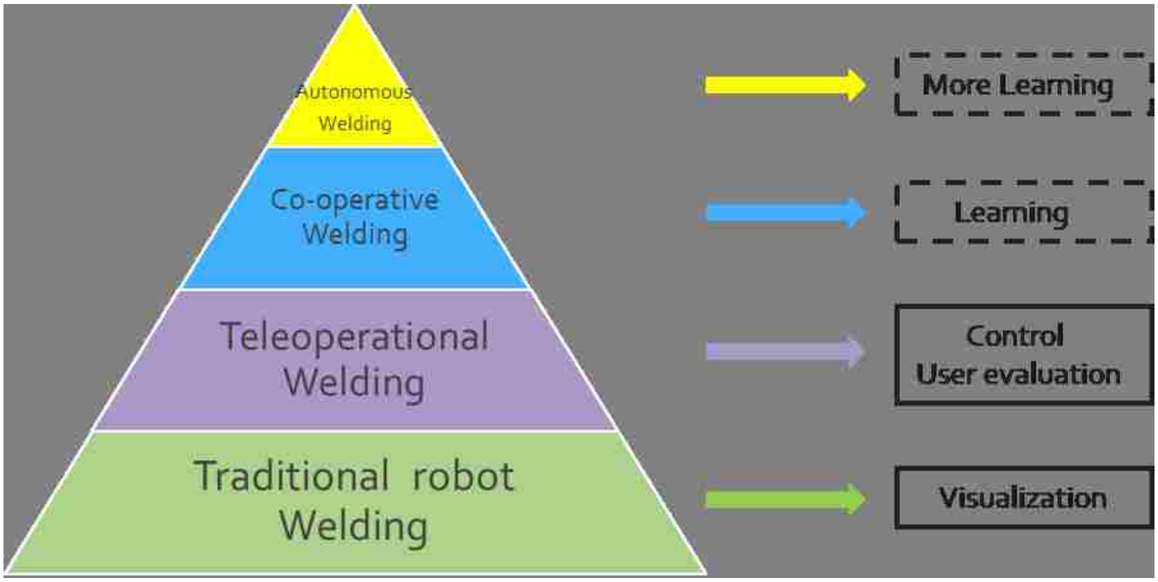


Figure 1.2: Model of operations and the associated research challenges.

1.2 A brief note on the research

Telerobotics is the area of robotics concerned with the control of semi-autonomous robots from a distance, chiefly using Wireless network (like Wi-Fi, Bluetooth, the Deep Space Network, and similar) or tethered connections It is a combination of two major subfields, teleoperation and telepresence. Teleoperation indicates operation of a machine at a distance. It is similar in meaning to the phrase ”remote control” but is usually encountered in research, academic and technical environments. It is most commonly associated with robotics and mobile robots but can be applied to a whole range of circumstances in which a device or machine is operated by a person from a distance.

Teleoperation is standard term in use both in research and technical communities and is by far the most standard term for referring to operation at a distance. This is opposed to ”telepresence” that is a less standard term and might refer to a whole range of existence or interaction that include a remote connotation. A telemanipulator (or teleoperator) is a device that is controlled remotely by a human operator. If such a device has the ability to perform autonomous work, it is called a telerobot. If the device is completely autonomous, it is called a robot.

In simple cases the controlling operator’s command actions correspond directly to actions in the device controlled, as for example in a radio controlled model aircraft or a tethered deep submergence vehicle. Where communications delays make direct control impractical (such as a remote planetary rover), or it is desired to reduce operator workload (as in a remotely controlled spy or attack aircraft), the device will not be controlled directly, instead being commanded to follow a specified path. At increasing levels of sophistication the device may operate somewhat independently in matters such as obstacle avoidance, also commonly employed in planetary rovers.

Devices designed to allow the operator to control a robot at a distance is sometimes called telecheric robotics.

Two major components of Telerobotics and Telepresence are the visual and control applications. A remote camera provides a visual representation of the view from the robot. Placing the robotic camera in a perspective that allows intuitive control is a recent technique that although based in Science Fiction (Robert A. Heinlein's Waldo 1942) has not been fruitful as the speed, resolution and bandwidth have only recently been adequate to the task of being able to control the robot camera in a meaningful way. Using a head mounted display, the control of the camera can be facilitated by tracking the head as shown in the figure below.

This only works if the user feels comfortable with the latency of the system, the lag in the response to movements, and the visual representation. Any issues such as, inadequate resolution, latency of the video image, lag in the mechanical and computer processing of the movement and response, and optical distortion due to camera lens and head mounted display lenses, can cause the user 'simulator sickness' that is exacerbated by the lack of vestibular stimulation with visual representation of motion.

Mismatch between the users motions such as registration errors, lag in movement response due to overfiltering, inadequate resolution for small movements, and slow speed can contribute to these problems.

The same technology can control the robot, but then the eyehand coordination issues become even more pervasive through the system, and user tension or frustration can make the system difficult to use.

The tendency to build robots has been to minimize the degrees of freedom because that reduces the control problems. Recent improvements in computers has shifted the emphasis to more degrees of freedom, allowing robotic devices that seem more intelligent and more human in their motions. This also allows more direct teleoperation as the user can control the robot with their own motions.

A telerobotic interface can be as simple as a common MMK (monitor-mouse-keyboard) interface. While this is not immersive, it is inexpensive. Telerobotics driven by internet connections are often of this type. A valuable modification to MMK is a joystick, which provides a more intuitive navigation scheme for planar robot movement.

Dedicated telepresence setups utilize a head mounted display with either single or dual eye display, and an ergonomically matched interface with joystick and related button, slider, trigger controls.

Future interfaces will merge fully immersive virtual reality interfaces and port real-time video instead of computer-generated images. Another example would be to use an omnidirectional treadmill with an immersive display system so that the robot is driven by the person walking or running. Additional modifications may include merged data displays such as Infrared thermal imaging, real-time threat assessment, or device schematics.

The prevalence of high quality video conferencing using mobile devices, tablets and portable computers has enabled a drastic growth in Telepresence Robots to help give a better sense of remote physical presence for communication and collaboration in the office, home, school, etc. when one cannot be there in person. The robot avatar can move or look around at the command of the remote person.

For over 20 years, telepresence robots, also sometimes referred to as remote-presence devices have been a vision of the tech industry. Until recently, engineers did not have the processors, the miniature microphones, cameras and sensors, or the cheap, fast broadband necessary to support them. But in the last five years, a number of companies have been introducing functional devices. As the value of skilled labor rises, these companies are beginning to see a way to eliminate the barrier of geography between offices. Traditional videoconferencing systems and telepresence rooms generally offer Pan / Tilt / Zoom cameras with far end control. The ability for the remote user to turn the devices head and look around naturally during a meeting is often seen as the strongest feature of a telepresence robot. For this reason, the developers have emerged in the new category of desktop telepresence robots that concentrate on this strongest feature to create a much lower cost robot. The Desktop Telepresence Robots, also called Head and Neck Robots allow users to look around during a meeting and are small enough to be carried from location to location, eliminating the need for remote navigation.

1.3 Innovation

In this dissertation, we propose the algorithms and frameworks for the improvement of performance of telerobotics. These algorithm enhance the overall performance from the aspects of teleoperation and telepresence.

The first framework which related to teleoperation is based on our novel hybrid-reality display (HRD) assisted teleoperational welding system, which we refer to as virtualized welding. It will allow a controller to monitor and remote control welding process with proper 3D and spatial cues in real time. On the visualization aspect, it contains a hybrid reality display (HRD) system, which utilizes projectors to project a captured video image onto a 3D replica of the actual weld surface. It provides a direct alignment between the frame of reference for the operator and that of the displayed image. The algorithm presented focuses on how to robustly calibrate multi-sensor in the framework, enable a fully immersed operation environment. On the human-robot interaction aspect, the latency of communication and transmission introduces unreality of remote scene. The algorithm proposed utilize the concept of predictive control and is capable of minimize the incoherence between remote video feedback and visualization on display device.

The second framework related to the domain of telepresence. We present a novel image enhancement framework to significantly improve the image quality captured by a camera from a see-through screen. Rather than performing traditional image enhancement, which are often under constrained, we employ an additional color+depth camera mounted on the side of the screen to make the problem better constrained. A novel sensor fusion algorithm is developed to allow the recovery of a low-noise, high-fidelity image with correct color reproduction and enhanced details.

1.4 Dissertation Outline

The reminder of this dissertation is structured as follows. Chapter 2 provides discussion of existing works mainly on these categories: structured light scanning technique, display techniques in augmented reality, research progress in robot vision, techniques in reconstruction of specular highlight surface and discussion on telepresence. Our approach on multi-sensor calibration in visualization and monitoring on proposed novel hybrid reality display(HRD) is detailed in Chapter 3. Since it is the first time the pipeline of hybrid reality display assist teleoperation on welding has ever been proposed, we first introduce the concept and setup. The algorithm that calibrate and combine all essential element in the system is then presented. The performance is quantitatively evaluated base on extensive user study. In Chapter 4, we move to the improved version of the pipeline discussed in Chapter 3 by adding human factors via remote control. Due to this extra element in the system, multiple delay may be introduced into the teleoperation system. We proposed an algorithm based on predictive control to compensate the in-synchronization issue. In Chapter 5, we address our effort on improve the performance of telepresence. Chapter 6 concludes with outlook of several future research possibilities.

Copyright © Bo Fu, 2015.

Chapter 2 Background and Related Work

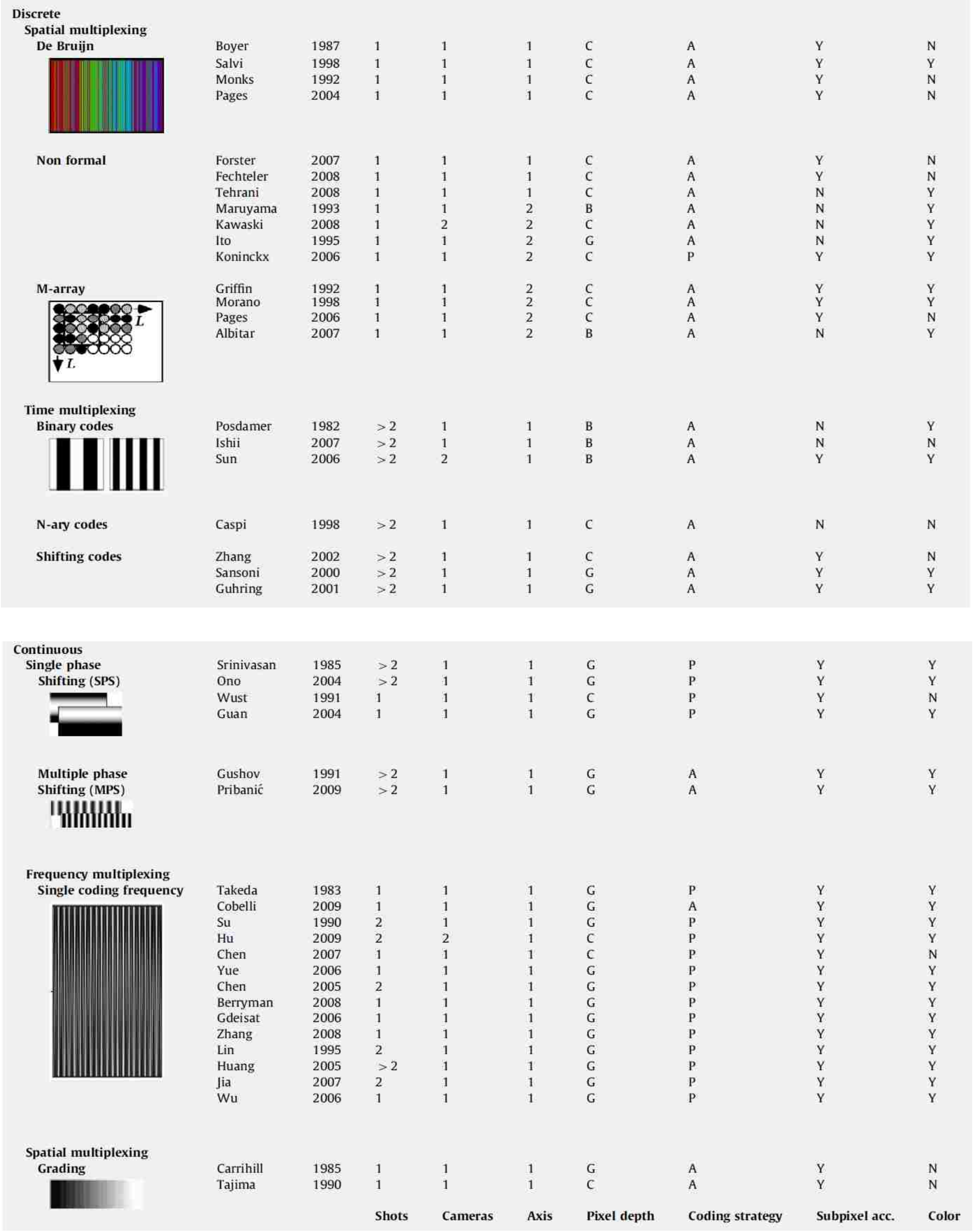
2.1 Structured Light 3D Scanning

3D scanning using structured light is one of the oldest computer vision techniques. Since the first paper [5–7] , a lot of progress has been made in terms of reconstruction speed, accuracy and resolution. Broadly, these techniques are divided into discrete [8] and continuous [9] coding schemes. For an exhaustive survey on structured light techniques, reader is referred to the survey by Salvi et al [10]. In addition, hybrid techniques that combine structured light with photometric stereo based techniques have been proposed as well [11, 12].

The seminal work of Nayar et al. [13] presented an iterative approach for reconstructing shape of Lambertian objects in the presence of interreflections. Liu et al. [14] proposed a method to estimate the geometry of a Lambertian scene by using the second bounce light transport matrix. Gupta et al. [15] presented methods for recovering depths using projector defocus [16] under indirect illumination effects. Chandraker et al. [17] use interreflections to resolve the basrelief ambiguity inherent in shape-from-shading techniques. Holroyd et al. [18] proposed an active multiview stereo technique where high-frequency illumination is used as scene texture that is invariant to indirect illumination. Park et al. [19, 20] move the camera or the scene to mitigate the errors due to indirect illumination in a structured light setup. Hermans et al. [21] use a moving projector in a variant of structured light triangulation. The depth measure used in this technique (frequency of the intensity profile at each pixel) is invariant to indirect illumination. In this paper, our focus is on designing structured light systems that are applicable for a wide range of scenes, and which require a single camera and a projector, without any moving parts. Nayar et al. showed that the direct and indirect components of scene radiance could be efficiently separated [22] using high-frequency illumination patterns. This has led to several attempts to perform structured light scanning under indirect illumination [23–25]. All these techniques rely on subtracting or reducing the indirect component and apply conventional approaches on the residual direct component.While these approaches have shown promise, there are three issues that prevent them from being applicable broadly: (a) the di- rect component estimation may fail due to strong interreflections (as with shiny metallic parts), (b) the residual direct component may be too low and noisy (as with translucent surfaces, milk and murky water), and (c) they require significantly higher number of images than traditional approaches, or rely on weak cues like polarization. Recently, Couture et al. [26] proposed using band-pass unstructured patterns to handle interreflections. Their approach involves capturing a large number (200) of images with random high-frequency patterns projected on the scene. In contrast, [27] explicitly design ensembles of illumination patterns that are resilient to a broader range of indirect illumination effects (interreflections, subsurface scattering, defocus, diffusion, and combinations of multiple effects), while using significantly fewer images.

Active illumination has also been used to measure density distribution of volumetric media [28, 29] and reconstruct transparent objects [30, 31]. For a detailed survey on techniques for reconstructing transparent and specular surfaces, please refer to the state of the art report by Ihrke et al. [32]. There have also been techniques for performing 3D scanning in the presence of volumetric media using light striping [33, 34]. Our techniques can not handle volumetric scattering. The focus of this work is on reconstructing opaque and translucent surfaces with complex shapes.

Fig. 2.1 [10] shows a classification of the existing pattern projection techniques.The main distinction has been done regarding the discrete or continuous nature of the pattern, rather than the codification process. Discrete patterns present a digital profile having the same value for the region represented by the same code word. The size of this region largely determines the density of the reconstructed object. Besides, continuous patterns present a smooth profile where every pixel has a unique code word within the non-periodicity region, assuring dense reconstruction. A posterior sub-classification is done regarding spatial, time and frequency multiplexing. Columns on the right indicate the value of some intrinsic attributes common to all the patterns.



			Shots	Cameras	Axis	Pixel depth	Coding strategy	Subpixel acc.	Color
Discrete									
Spatial multiplexing									
De Bruijn	Boyer	1987	1	1	1	C	A	Y	N
	Salvi	1998	1	1	1	C	A	Y	Y
	Monks	1992	1	1	1	C	A	Y	N
	Pages	2004	1	1	1	C	A	Y	N
Non formal	Forster	2007	1	1	1	C	A	Y	N
	Fechteler	2008	1	1	1	C	A	Y	N
	Tehrani	2008	1	1	1	C	A	N	Y
	Maruyama	1993	1	1	2	B	A	N	Y
	Kawaski	2008	1	2	2	C	A	N	Y
	Ito	1995	1	1	2	G	A	N	Y
	Koninckx	2006	1	1	2	C	P	Y	Y
M-array	Griffin	1992	1	1	2	C	A	Y	Y
	Morano	1998	1	1	2	C	A	Y	Y
	Pages	2006	1	1	2	C	A	Y	N
	Albitar	2007	1	1	2	B	A	N	Y
Time multiplexing									
Binary codes	Posdamer	1982	> 2	1	1	B	A	N	Y
	Ishii	2007	> 2	1	1	B	A	N	N
	Sun	2006	> 2	2	1	B	A	Y	Y
N-ary codes	Caspi	1998	> 2	1	1	C	A	N	N
Shifting codes	Zhang	2002	> 2	1	1	C	A	Y	N
	Sansoni	2000	> 2	1	1	G	A	Y	Y
	Guhring	2001	> 2	1	1	G	A	Y	Y
Continuous									
Single phase Shifting (SPS)	Srinivasan	1985	> 2	1	1	G	P	Y	Y
	Ono	2004	> 2	1	1	G	P	Y	Y
	Wust	1991	1	1	1	C	P	Y	N
	Guan	2004	1	1	1	G	P	Y	Y
Multiple phase Shifting (MPS)	Gushov	1991	> 2	1	1	G	A	Y	Y
	Pribanić	2009	> 2	1	1	G	A	Y	Y
Frequency multiplexing									
Single coding frequency	Takeda	1983	1	1	1	G	P	Y	Y
	Cobelli	2009	1	1	1	G	A	Y	Y
	Su	1990	2	1	1	G	P	Y	Y
	Hu	2009	2	2	1	C	P	Y	Y
	Chen	2007	1	1	1	C	P	Y	N
	Yue	2006	1	1	1	G	P	Y	Y
	Chen	2005	2	1	1	G	P	Y	Y
	Berryman	2008	1	1	1	G	P	Y	Y
	Gdeisat	2006	1	1	1	G	P	Y	Y
	Zhang	2008	1	1	1	G	P	Y	Y
	Lin	1995	2	1	1	G	P	Y	Y
	Huang	2005	> 2	1	1	G	P	Y	Y
	Jia	2007	2	1	1	G	P	Y	Y
	Wu	2006	1	1	1	G	P	Y	Y
Spatial multiplexing									
Grading	Carrihill	1985	1	1	1	G	A	Y	N
	Tajima	1990	1	1	1	C	A	Y	N

Figure 2.1: The classification of existing structured light techniques.

These attributes are: 1) Number of projected patterns: determines whether the method is valid or not for measuring moving objects. 2) Number of cameras: the method uses stereo-vision(two or more cameras)coupled to an on-calibrated pattern used only to get texture on the surface pattern,or an unique camera coupled to a calibrated projector. 3) Axis codification: the pattern is coded along one or two axis. 4) Pixel depth: refers to the color and luminance level of the projected pattern(B,G and C stands for binary, gray scale and color, respectively). 5) Coding strategy: refers to the periodicity of the set of patterns projected on the surface(A stands for absolute and P stands for periodic). 6) Sub-pixel accuracy: determines whether the features are found considering sub-pixel precision,thus providing better reconstruction results (yes or no). 7) Color: determines whether the technique can cope with colored objects (yes or no).

The visible light based structured light scanning technique, however put up with a major disadvantage: the light pattern projection forms an invasive signal that can be objectionable in some cases. It yields to the loss or corruption of colorimetrical and textural information of the lighted surfaces, to the inconsistence of the optical flow and, moreover, to the offensive, indeed dangerous aspect of the illumination (think about potential danger of LASER sources, faces measurements even with slide light projector, etc.) In addition, some inspection systems, working in outdoor or partial outdoor environment, have to be discreet and without risks. Among these, systems of sensitive zone surveillance, systems of collision detection on some vehicles, systems of environment recognition used in robotics and many others can be mentioned. These systems have to acquire geometric measurements on objects that cross their detection field, without disrupting, modifying, or putting in danger the environment. In order to benefit from the advantages of structured light vision while avoiding these drawbacks, we laid down the objective to design a sensor with light pattern projection in the nonvisible spectrum. With this aim, several options occur, each one based on a different type of light: InfraRed Structured Light (IRSL), Imperceptible Structured Light (ISL) and Filtered Structured Light (FSL).

For Infrared Structured Light scanning, An infrared laser beam is used to generate invisible patterns that can be single dot, single line or bi-dimensional patterns. The light is usually projected in the near-infrared, i.e. from 640nm to 2500nm or from 4000 to 15600 cm-1. The scene is observed by a CCD camera, because of the spectral sensibility of CCD, from 300nm to 1100nm (an infrared camera is not necessary). This technique is widely used for 3D scanning, robot navigation and so on. Boverie, Devy and Lerasle recently used infrared structured light for 3D perception applied to airbag generation [35]. They also compared structured light and stereovision [36]: they concluded that stereoscopy gives very high reliability but remains heavy in terms of computational time, whereas structured light proposes a good compromise between accuracy and speed. They also recall one of the most interesting characteristic of structured light: its capability to retrieve 3D information from a non-textured surface. Infrared structured light has also been used in omnidirectional vision [37]. A rotative sensor composed of two CCD cameras and a diffracted laser beam observed the scene along 360. By filtering the images, this sensor is used for the 3D reconstruction thanks to structured light and color acquisition thanks to RGB information. A Cold filter is used to grab infrared structured light image.

For imperceptible structured light , sensors are composed by a unique light source and two cameras. The light source projects a light pattern followed by its complement (inverse pattern) onto the scene at high frequency, so that the resulting pattern is uniform. The first camera is synchronized with the projection of the first pattern and permits to reconstruct the scene thanks to the capabilities of structured light vision; the second camera has a longer integration time and observes the scene under uniform light (as a result of pattern and complement projection) which permits to get a classical gray-level or colored image and processes it. This technique is known as imperceptible structured light [38,39]. The aim is to combine the advantages of structured lighting (easy correspondence and reconstruction of homogeneous surfaces) with the advantages of classical vision (color or texture analysis, etc.) in order to achieve a 3D reconstruction of the scene with the mapping of surface colors and textures.

Whether the frequency of projection reaches the critical Ilicker Irequency (from 75Hz

Table 2.1: Equipment requirement comparison between 3 invisible structured light scanning techniques

	Light Source	Camera	Additional Requirement
IRSL	Only one light source is needed: laser beam, diffracted laser beam.	One CCD camera or, eventually, one infrared camera.	None
ISL	One video-projector is needed.	Two CCD cameras are needed.	None
FSL	Only one light source is needed: laser beam, diffracted laser beam or video-projector.	One CCD camera or, eventually, one infrared camera.	One IR filter.

according to Watson [40]), the pattern and the inverse pattern are visually integrated over time, so that the result is the appearance of flat field (”white” light). Critical flicker frequency is defined as the highest frequency at which a person can detect the flicker in a flickering light source.

For filtered structured light , The light source is filtered so that only infrared structured light passes. The light pattern can be projected through a laser source or a video-projector. An IR filter, set in front of the light source, permits to ”cancel” light below 750nm, 800nm, 850nm, etc. by acting as an high-pass filter.

Based on the required equipment, the 3 invisible structured light scanning techniques can be compared in (Table. 2.1 [41]).

2.2 Visual Display Technique

Mixed reality [42, 43] have been recognized for welder training [44] and adapted in the test generation of training equipment. Some sophisticated systems for training with HMD have been available recently, such as ARC+ [45], the Fronius Virtual Welding system [46], VRTEX 360 [47], and EWI AdvanceTrainerTM [48]. These systems do not employ see-through method, instead, they apply fully simulated environment on the display, thus are categorized into virtual reality. Among these methods, VRTEX 360 is probably one of the most sophisticated. It is co-developed by a company specializing in visual training and simulation. A mock-up welding torch is equipped with sensors so that it can be fully tracked. A welders helmet is fitted with a head mounted display to provide simulated images. A graphics simulation program is developed to simulate the welding process, including the sparks and the weld pool. It is not clear how the simulation is achieved. As a training tool, the images shown to the trainee are entirely simulated. While this may be adequate for the purpose of training, it is unlikely to be able to simulate the complexity and possible variations in a real welding environment. Another drawback is that the focus distance is fixed in most display types, resulting poor eye accommodation. Therefore we choose to use augmented reality (AR) techniques for the visualization aspect of virtualized welding. AR allows a user to see the real world, with virtual objects superimposed upon or composited with the real world [49]. AR has been used in many application areas, such as education, health care, the military, and entertainment. However, its application to welding has not been reported.

Display techniques of AR [50, 51] can be divided into three categories: head-mounted displays, projection-based displays and handheld displays.

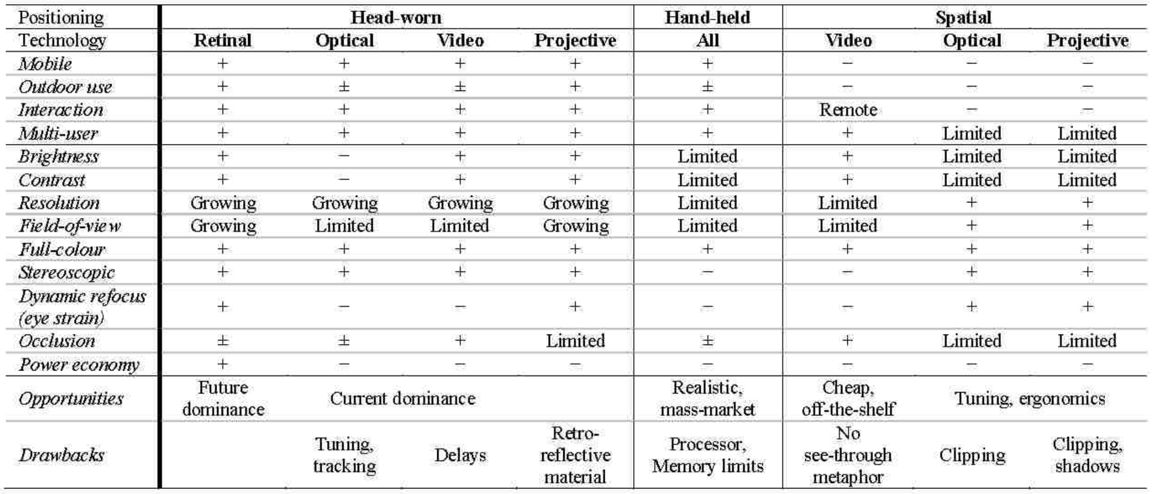
HMD provides user a virtualized or partially virtualized immersive environment by combining reality information and virtual information. A see-through HMD can thereafter help the user to see more information than just the reality by superimposing virtual objects in the reality. Differenced by the mean of visual presence, it can be divided into three categories: optical see-through [52], video see-through and head-mounted projective displays. Tightly related to virtual reality is video see-through, where the virtual environment is replaced by video of reality and images are overlaid by the AR. In [53], an AR system integrated in welding helmet is presented. The scene is acquired by a stereoscopic high dynamic ranged CMOS camera that enabling simultaneous observation of the welding arc and environment. Then the scene is displayed on a video see-through HMD. Disadvantages of these methods include a low resolution of reality, a limited field-of-view and user disorientation due to a parallax caused by the cameras positioning at a distance from the viewers true eye location, resulting significant adjustment effort. In [54], however, a video see-through HMD was built with zero eye offset from commercial components and a mount fabricated via rapid prototyping, but this method still suffers from other drawbacks other than the parallax issue.

Another method is optical see-through which abandons the real world perception but only displays the AR overlay by means of transparent mirrors and lenses. A popular product that falls into this category is Google glass. In [52] an autostereoscopic optical see-through system for AR is presented. It uses a transparent holographic optical element (HOE) to separate the views produced by two, or more, digital projectors. It is a minimally intrusive AR system that does not require the user to wear special glasses or other equipment. The main challenge is the generation of correct occlusion effects between virtual and real objects. [55] solves this problem by proposing a display which is capable of mutual occlusions. Advantages of optical see-through method are that they are cheaper and parallax-free. The major limitation that makes optical see-through inapplicable to welding application is other input devices such as cameras are required for registration. Also, by combining the virtual objects holographically through transparent mirrors, the brightness and contrast of both the images and reality perception will be reduced, making this method less suitable for environment with spectral high light caused by the electric arc during welding.

As an alternative to HMD, head-mounted projective displays [56] use a pair of portable projectors mounted on the headset that project images onto retro-reflective material which is then reflected back into users eye. Advantage of this method is that it supports a large field of view and is able to display on curved surface. The limitation is that the user must be precisely tracked to provide a stable and convincing image.

Different from head-mounted projective displays, projection-based displays project AR overlay onto real objects to result in a projective display. In [57], a method named Shader Lamp is proposed. The idea is to use projectors to graphically animate physical objects in the real world. This method is a good option for applications that do not require users to wear anything, providing minimal intrusiveness and accommodating users eyes during focusing. They can cover large surfaces for a whole field of view. Projection surface can varying from flat to complex models. We therefore utilize this method for visualization aspect of our system.

Generally, the advantages and disadvantages of these techniques is presented in Fig. 2.2 [58].



Positioning	Head-worn				Hand-held	Spatial		
Technology	**Retinal**	**Optical**	**Video**	**Projective**	**All**	**Video**	**Optical**	**Projective**
Mobile	+	+	+	+	+	−	−	−
Outdoor use	+	±	±	+	±	−	−	−
Interaction	+	+	+	+	+	Remote	−	−
Multi-user	+	+	+	+	+	+	Limited	Limited
Brightness	+	−	+	+	Limited	+	Limited	Limited
Contrast	+	−	+	+	Limited	+	Limited	Limited
Resolution	Growing	Growing	Growing	Growing	Limited	Limited	+	+
Field-of-view	Growing	Limited	Limited	Growing	Limited	Limited	+	+
Full-colour	+	+	+	+	+	+	+	+
Stereoscopic	+	+	+	+	−	−	+	+
Dynamic refocus (eye strain)	+	−	−	+	−	−	+	+
Occlusion	±	±	+	Limited	±	+	Limited	Limited
Power economy	+	−	−	−	−	−	−	−
Opportunities	Future dominance	Current dominance			Realistic, mass-market	Cheap, off-the-shelf	Tuning, ergonomics	
Drawbacks		Tuning, tracking	Delays	Retro-reflective material	Processor, Memory limits	No see-through metaphor	Clipping	Clipping, shadows

Figure 2.2: The characteristics of display techniques.

2.3 Hand-eye calibration for Robotic Vision

There is a strong need for an accurate hand-eye calibration (Fig. 2.3 [59]). The reasons are twofold: i) to map sensor-centered measurements into the robot/world frame and ii) to allow for an accurate prediction of the pose of the sensor on the basis of the arm motion in fact these are often complementary aspects of the same problem.

When performing hand-eye calibration on the basis of both the pose of tool with respect to the robot base frame, and pose of camera with respect to the world frame, there are two main approaches in order to estimate the hand-eye transformation:

1) Move the hand and observe/perceive the movement of the eye.

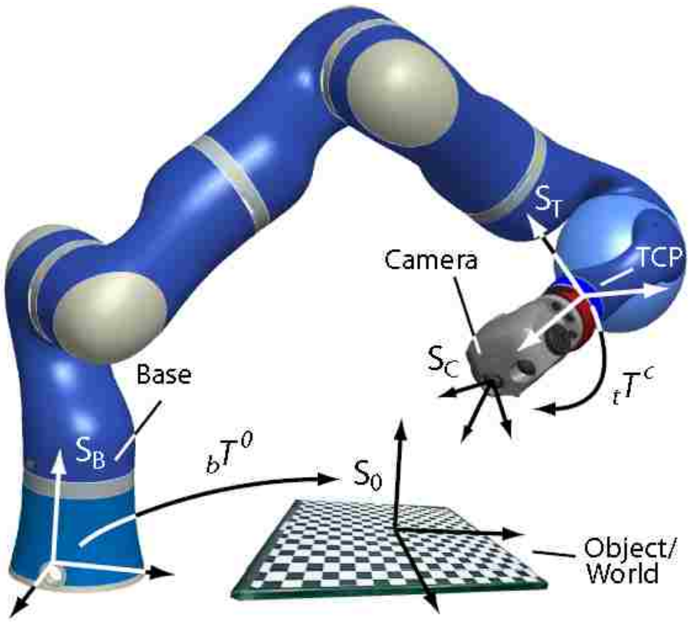


Figure 2.3: The Concept of hand-eye calibration.

This is the classical approach. Early solutions regard the rotational part of this equation decoupled from the translational one, yielding uncomplex, fast, but error-prone formulations, since rotation estimation errors propagate to the translational part. Seminal articles are Shiu and Ahmad 1989 [60] (least squares fitting of rotation, then translation, using angle-axis representation) and Tsai and Lenz 1989 [61] (similar to [60] with closedform solution). Zhuang and Roth 1991 [62] simplified the formulation introducing quaternions for the estimation of the rotational part, in the same way as Chou and Kamel 1991 [63], who make use of the singular value decomposition (SVD). Chen 1991 [64] for the first time does not decouple rotational and translational terms by using the screw theory. Wang 1992 in [65] compares [60] and [61] resulting in a slight advantage for the latter. Zhuang and Shiu 1993 [66] apply nonlinear optimization for both parts, minimizing a similar expression to Frobenius norms of homogeneous matrices of transformation errors. They additionally offer the possibility to disregard the camera orientation for the estimation. A similar approach was presented by Fassi and Legnani 2005 [67]. Park and Martin 1994 [68] perform nonlinear optimization in the same way, but again in the detached formulation. Lu and Chou 1995 [69] introduce the eight-space formulation based on quaternions, linearly optimizing both parts at the same time using the SVD. Horaud and Dornaika 1995 [70] nonlinearly optimize both the rotational (formulated with quaternions) and the translational parts one-to-one. Wei, Arbter, and Hirzinger 1998 [71] nonlinearly minimize algebraic distances performing simultaneous hand-eye and camera calibration. Daniilidis 1999 [72] introduces the dual quaternions an algebraic representation of the screw theory to describe motions. This enables the author to find a fast SVD-based joint solution for rotation and translation within linear formulation. Bayro-Corrochano et al. 2000 [73] in the same way produce a SVD-based linear solution of the coupled problem by the use of motors within the geometric algebra framework. Andreff et al. 2001 [74] do the job properly, employing this particular formulation for X-from-motion applications. They get rid of the nonlinear orthogonality constraint by increasing the dimensionality of the rotational part and manage to formulate the problem as a single homogeneous linear system.

2) Simultaneous estimation of the hand-eye transformation and the pose of the robot in the world.

To the best of our knowledge it was Wang in 1992 [65] who first submitted this formulation explicitly for hand-eye calibration. Surprisingly, none of the further approaches refer to him in this context. Zhuang et al. 1994 [75] apply quaternions in order to get a simple linear solution of the rotational part by the use of the SVD. Remy et al. 1997 [76] nonlinearly optimize both parts by minimizing reprojected 3D Euclidean error distances in S0. Dornaika and Horaud 1998 [77] solve the rotational problem linearly with quaternions and also nonlinearly optimize both parts by one-to-one minimizing of Frobenius norms and two penalty functions. Other approaches integrate the hand-eye calibration with the intrinsic camera calibration and minimize the Root Mean Square (RMS) of the image frame errors. The optimization criteria for both approaches are often suboptimal and no attention is paid to proper parametrizations. Since the purpose of model-based3 calibration is the accurate parametrization of the system model, maximum accuracy optimal calibration is achieved when minimizing model fitting errors with regard to the actually erroneous elements. Here we propose a metric on the group of rigid transformations SE(3) for this purpose. Moreover, with the exception of [76], a thorough comparison of these very different approaches is missing.

2.4 Human-Robot Interaction (HRI)

From a user-cantered design standpoint, our proposed display system is motivated by teleoperation accidents, incidents, and user research in military reconnaissance [78], surgery (e.g., [79]), urban search and rescue (e.g., [80]), and space exploration (e.g., [81]). Teleoperation in these environments is compromised by the Keyhole Effect, limited depth cues, and misalignments of robot, display, and human frames of reference. The Keyhole Effect results from the limited field of view that is usually provided by robot-mounted cameras. This keyhole view disrupts the operators normal attention control, limits situation awareness, disrupts spatial comprehension, and makes object identification more difficult [82]. Luckily, for welding applications, welders are trained to work with a limited field of view, and the area of interest is small. Frame of Reference problems occur when the operators control axes are not spatially aligned with the axes of motion of the robot. Misalignment requires the operator to mentally rotate and transform displayed axes, increasing cognitive workload, response times, and control errors [83]. Finally, limitations of depth information during teleoperation require the human operator to use less efficient control strategies and results in greater mental workload (e.g., [79]).

2.5 Predictive Control and Robot Control

Robot control has been an active research area since early 1980s. Different control methods have been proposed, ranging from passivity, compliance, predictive and adaptive control, and variable structures [84]. Wave variables method, as a modification to the passivity theory, is considered to be a robust approach to solve arbitrary time delay problem [85–87]. However, wave variables are not physically measurable and thus may not be as intuitive as velocity and force data. Garcia [88] proposed a hybrid control method for controlling a telerobotic system. It was designed to modify the references sent from the local station to the remote station when force and position thresholds were overcome or when communication was interrupted. Nonlinear adaptive control was also adopted by various researchers [89–91]. The fuzzy control method and neuro-fuzzy technology have been demonstrated to have advantages of robustness and ability to model and control complex nonlinear systems [92, 93].

Predictive control of linear systems has received considerable attention in past decades due to its robustness with respect to model uncertainty [94–96]. Recently nonlinear predictive control method has been extensively studied to control the robot arm. Makarov [97] presented a model-based predictive approach for trajectory tracking of an anthropomorphic robot arm. Wang [98] proposed a multivariable predictive-repetitive controller. Closed-loop performance of the proposed control system in terms of reference trajectory following, disturbance rejection, and measurement noise attenuation was also demonstrated. In this paper control of the robotic arm movement speed is formulated as a predictive control problem, and an analytical solution is derived to control the robot speed in real-time.

2.6 Telepresence Enhancement

The issue of image or video *denosing* has been an active research topic for decades. Effective approaches include as non-local means [99], bilateral filters [100], etc. Unfortunately, high-frequency details are usually lost after denosing, since it is impossible to distinguish high-frequency contents with random noise in a single image.

The use of additional optical component can also alter the color balance of the captured image. For example, the ConnectBoard system uses an wavelength dependent diffuser to interleave the projected image and see-through image [101]. The color transfer of the diffuser is approximated as a piece-wise linear affine transformation. Color transfer from different images can also be achieved by looking at some image statistics (such as mean and image histogram) [102].

Recent imaging techniques combine two or more images in the gradient domain (e.g., [103–105]). These algorithm usually deal with a stack of images taken from the same perspective, for which the pixel correspondences across images are accurate and given.

Chapter 3 Visualization and Monitoring of Remote Welding

Historically, many technological systems that require skilled operator control eventually transition to semi-automated or fully automated systems. This is certainly the case for welding, where automated welding systems have existed for several decades. These automated systems reside primarily in manufacturing, where weld parameters may be tightly controlled (e.g. weld types, work piece position, environmental conditions, etc.). Situations where variation occurs have required the skill of expert welders. Driven by increasing demands in manufacturing to produce more customized products in small batches [106], semi-automated processes for more complex welding tasks are likely to occur in the near future. This will allow a welding robot to intelligently adapt to various welding tasks, while requiring the expert welder to monitor progress in real time and make changes when necessary. This transition from manual operation to monitoring, is likely to bring about a new set of cognitive and physical task demands for the welder [107, 108]. One of the best methods to counter future workload issues may be through the use of efficient, user-centered design of new displays [109].

In this chapter we present a new type of hybrid-reality display (HRD) system, which we refer to as *virtualized welding*, that will allow a welder to monitor a remote welding process with proper 3D and spatial cues in real time (Fig. 3.1). It is assembled out of readily available sensing and visualization hardware. In particular, we present an augmented display that utilized projectors to project a captured video image onto a 3D replica of the actual weld surface. We have chosen this approach due to a wide range of human factors, ergonomics, and usability research that has identified the limitations of traditional planar displays for supporting navigation and teleoperation tasks [110]. User performance often suffers because of misalignments between the frame of reference of the operator and that of the displayed image, reductions in visual context, and limited depth cues. Skilled operators,

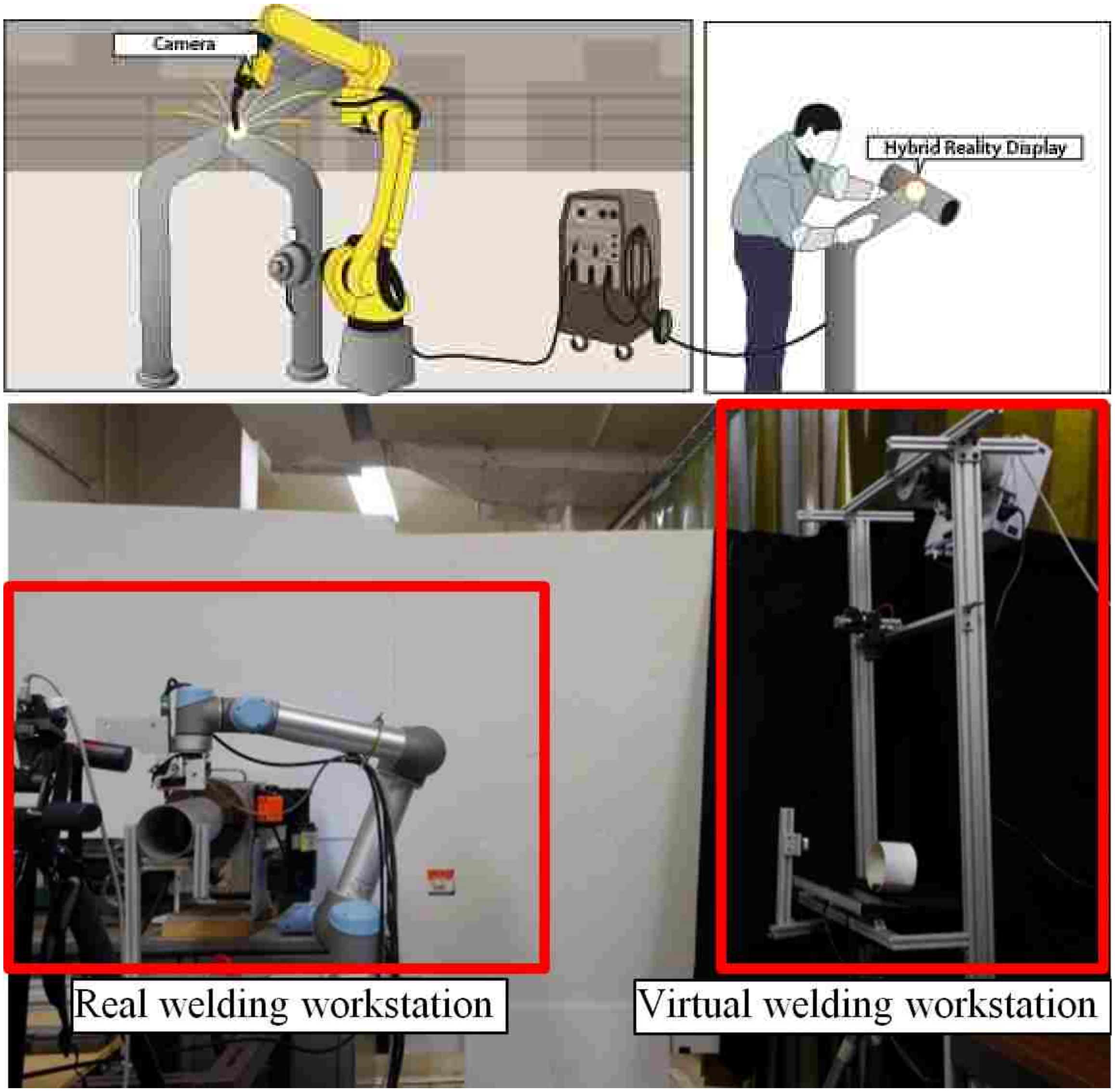


Figure 3.1: Virtualized Welding: Above: an illustration of our virtualized welding operation, in which existing welding robots are augmented with a video cameras to capture the working environment. The operator can monitor the welding process in an augmented display setup from differenct angles, in which welding images are projected on a mock-up 3D surface, as if he/she was right next to the actual welding. Below: our current implementation of virtualized welding.

such as seasoned surgeons performing laparoscopic surgery, can often compensate for these limitations, but doing so increases their cognitive load, perceived stress, and fatigue [111]. Thus, we argue that a fundamental goal of visual workstation design must focus not only on immediate operator performance but also on the enhancement of cognitive metrics that have implications for long-term operator well-being and proficiency.

Some sophisticated systems for training with head-mounted displays(HMD) have been available recently, such as ARC+ [45], the Fronius Virtual Welding system [46], VR-TEX360 [47], and EWI AdvanceTrainerTM [48]. HMDs provide the user with a virtualized or partially virtualized immersive environment by combining reality information and virtual information. While improvements have been made since their inception, disadvantages to HMDs include: field of view restrictions, low resolution, parallax issues, and user disorientation. As such, these problems are still the focus of extensive study [112–115].

Our focus on both experiential and performance outcomes has led us to bypass some current VR workstation options, such as HMDs, in favor of a workstation that makes use of surrogate objects as projection surfaces. Projection-based displays project an AR overlay onto real objects to result in a projective display. For example, in [57], a method named *ShaderLamp* is proposed. The idea is to use projectors to graphically animate physical objects in the real world. This method allows for applications that do not require users to wear anything, thereby limiting intrusiveness, allowing natural eye focusing. In addition, the image may encompass large surfaces across a users whole field of view, and the projection surfaces may vary from flat to complex models. To be clear, the goal of our initial development and evaluation efforts are not to directly compete with alternative, evolving VR display modalities, but rather to see if a surrogate-base projection workstation can yield immediate improvements in user outcomes over those associated with the direct video feeds of the weld pool provided by current robotic welding systems.

The current work presents in detail our system design and methods for calibration of the system. Furthermore, a user performance study was conducted aimed at assessing any possible performance and cognitive benefits of the current system, specifically in regards to monitoring of the weld process in real-time. Results demonstrate the hybrid-reality system yields immediate advantages in user performance and workload over more traditional planar displays presenting the work piece with common camera viewpoints.

3.1 System Platform

Our virtualized welding system (Fig. 3.1) consists of two workstations: real welding workstation and virtual welding workstation. Illustrated in Fig. 3.2, The real welding workstation is responsible for completing the welding task on the work piece and acquiring visual information of the work piece. It contains the robot control system and the visual information acquisition system. The virtual welding workstation (Fig. 3.4), on the other hand, focuses on visualizing the work piece. Data communication between the two workstation is bridged by network.

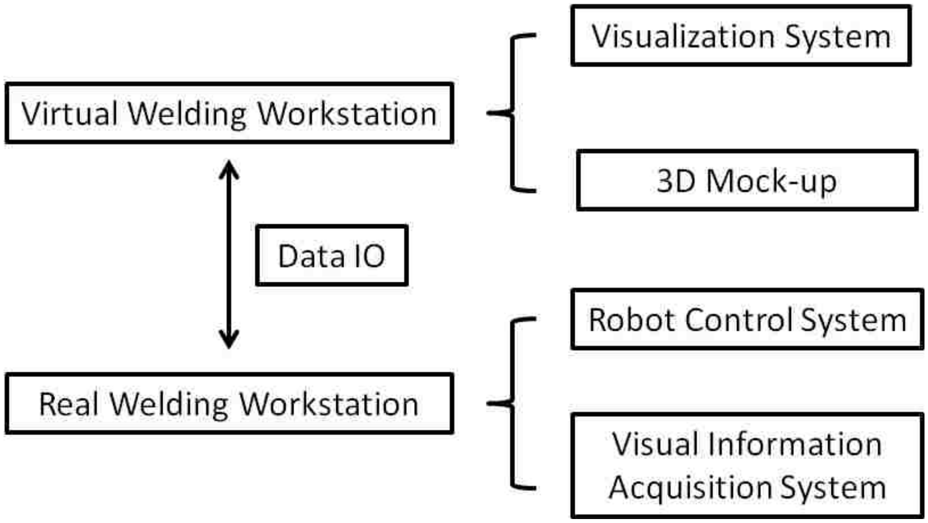


Figure 3.2: General structure of the virtualized welding system.

Real Welding Workstation

Illustrated in Fig. 3.3, the weld gun and a video camera are rigidly mounted to the end effector of a robotic arm. The camera, which is referred to as the *local view* camera, observes the work piece while welding is in progress. The view angle of the camera is adjusted to be similar to that of human welders, providing a more realistic visualization experience when the visual information taken by the local view camera is rendered in the virtual welding workstation. Another camera is mounted in a fixed location. This camera, which is referred to as the *global view* camera, captures a wide view of the working environment. It provides more reference in the surrounding area. The videos captured by the global view camera is mainly used for usability evaluation. The robotic arm follows the command from the virtual welding workstation and drives the end effector in real time. In addition, a 3D scanner based on structured light is used to scan the work pieces in high resolution (less than 1mm in depth accuracy).

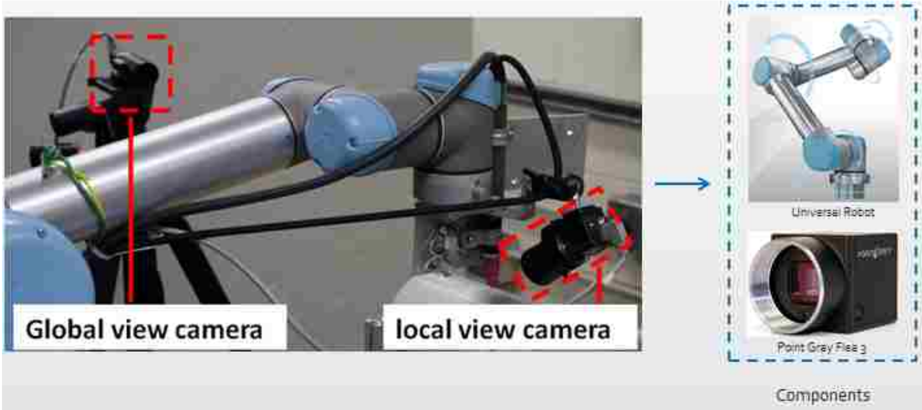


Figure 3.3: Overview of real welding workstation. The global view camera is facing towards the robotic arm and work piece.

Virtual Welding Workstation

A mockup of the work pieces is reconstructed in the virtual welding station. The mockup can be assembled from identical work pieces, or in our case, 3D printed. A video projector is used to project imagery from the real-welding station onto the mockup. Since the mockup and the actual work pieces have (almost) identical surface geometry, the resulting HRD is *autostereoscipic* and provides the same spatial cue as in the real welding. In order to achieve that, the projector must be calibrated with respect to the mockup. This requires the use of an auxiliary camera to observe the projected images for calibration purpose. In addition, the projector-camera pair can be used as a 3D scanner. We usually re-scan the mockup to accommodate the errors introduced in 3D printing or assembly.

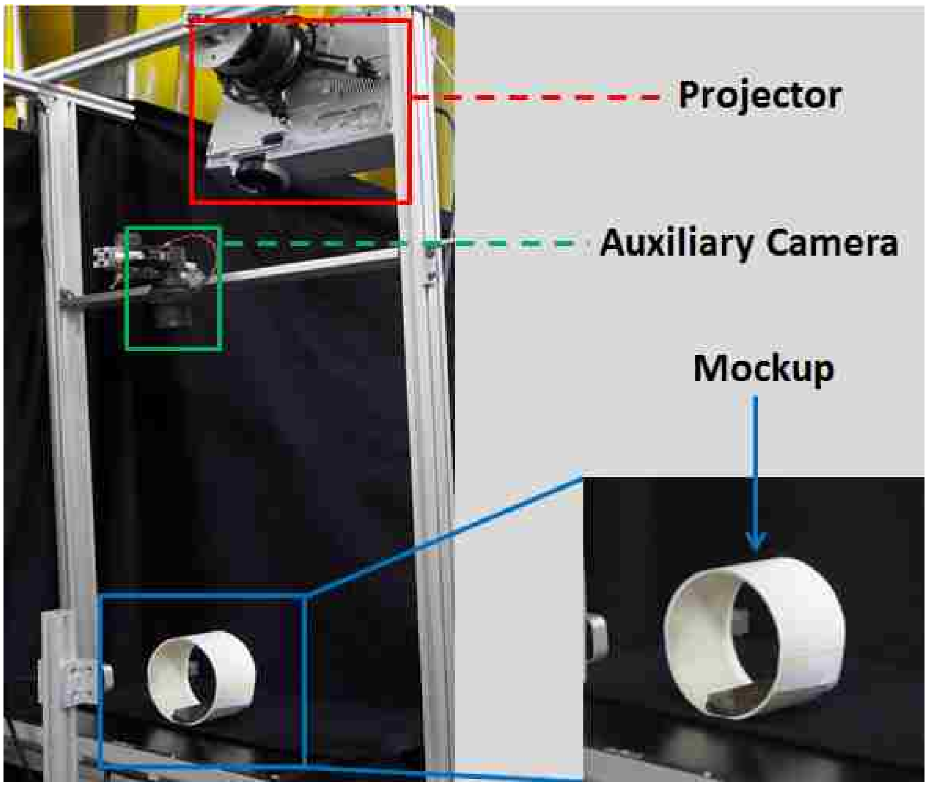


Figure 3.4: Detailed view of the virtual welding workstation. Major components are a mock up, an auxiliary camera and a projector. The mockup is generated by 3D printer.

3.2 Multi-sensor Calibration

Extensive calibrations between components of the proposed system are conducted before the system performs the teleoperation and visualization task. Fig. 3.5 illustrates these calibrations in our system.

The calibration work concerns (a) the pose of the local-view camera at the end-effector of robotic arms and (b) linking the coordinate frame on the real welding station with that from the virtual welding station. Since we have geometrically identical objects in both stations, we use them as the common coordinate frame (e.g., S_0). Then the task of calibration is to find the intrinsic and extrinsic parameters of the cameras and projectors. More specially, for the local-view camera, hand-eye calibration consists of calculating the unknown position (translation) and orientation (rotation) of camera frame S_C w.r.t the robot end-effector frame S_H when the camera is mounted on the robotic arm rigidly. The other pair

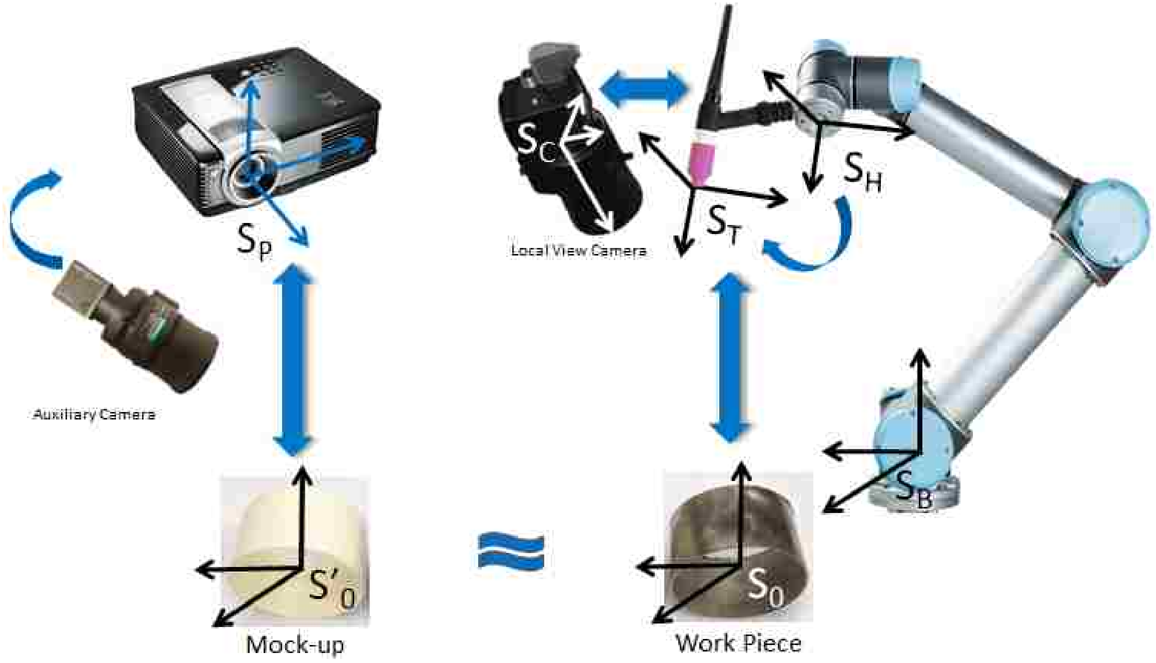


Figure 3.5: Calibration of the virtualized welding system. Projector, local view camera and robotic arm are calibrated together.

of coordinate frames, S_H and robot base frame S_B are easily linked by forward kinematic.

Hand-Eye Calibration

Regarding (S_C, S_H) calibration, we can simultaneously estimate the hand-eye transformation and pose of the robot in the world: $AX = ZB$, where A is the homogeneous transformation relating pose of S_C to the pose of world frame S_0 ${}_0T^c$, B is the homogeneous transformation linking the pose of S_H and the pose of robot base frame S_B ${}_bT^h$, and X and Z are the eye-hand and world-base transformation [59]. The estimation can be further formulated to the predictive parametric model, which can directly reproduces the rigid transformations in a loop way: $camera - hand - base - world - camera$:

$$ {}_0T^c\,{}_cT^h = {}_0T^b\,{}_bT^h \rightleftharpoons \begin{tikzcd} S_C \arrow[r, "{}_cT^h"] & S_T \\ S_0 \arrow[u, "{}_0T^c"] \arrow[r, "{}_0T^b"] \arrow[ur] & S_B \arrow[u, "{}_bT^h"'] \end{tikzcd} \tag{3.1} $$

Since the transformation ${}_0T^b$ is not our concern in this specific case of hand-eye calibration, we can further eliminate this term by replacing it with two different instants i and j in Eq.3.1:

$$
{}_{c_i}T^{c_j}{}_cT^h = {}_cT^h{}_{h_i}T^{h_j} \rightleftharpoons
\begin{tikzcd}
{}_0T^{c_j} \arrow[r, "{}_cT^h,({}_0T^b)"] & {}_bT^{h_j} \\
{}_0T^{c_i} \arrow[u, "{}_{c_j}T^{c_j}"] \arrow[ur] \arrow[r, "{}_cT^h,({}_0T^b)"] & {}_bT^{h_i} \arrow[u, "{}_{h_i}T^{h_j}"']
\end{tikzcd}
\tag{3.2}
$$

The equation can be further decomposed into rotation and translation:

$$
\begin{cases}
{}_{c_i}R^{c_j}{}_cR^h = {}_cR^h{}_{h_i}R^{h_j} \\
{}_{c_i}R^{c_j}{}_ct^h + {}_{c_i}t^{c_j} = {}_cR^h{}_{h_i}t^{h_j} + {}_ct^h
\end{cases}
\tag{3.3}
$$

with error metric [59]:

$$
\{{}_tT^c, {}_bT^0\}^* = \arg\min_{{}_tT^c,{}_bT^0}\left(\sum_{i=1}^{n} \frac{(\mathrm{O}_i^{rot})^2}{{}^*\sigma_{rot}^2} + \frac{(\mathrm{O}_i^{tra})^2}{{}^*\sigma_{tra}^2}\right)
\tag{3.4}
$$

where ${}^*\sigma_{rot}^2$ and ${}^*\sigma_{tra}^2$ are the 2^{nd} moments of the independent Gaussian probability density function in rotation and translation error.

by solving Eq.3.3, we can calibrate frame S_C and frame S_H.

Projector-Mockup Calibration

In the virtualized working station, the visualization of real working environment is performed via projector and the mockup onto which the rendered video frame will be projected. Without prior knowledge of the transformation between projector frame S_P and common frame S_0', it is necessary to discuss the calibration procedure.

In order to discover the 3D transformation between S_P and S_0', it is naturally to adapt the technique of structured light scanning since only an extra auxiliary camera is needed to fulfill the SL's requirment, and the accuracy can be as less than 1 mm. Other scanning method, such as Line scanning systems, e.g. laser scanners, are capable of acquiring accurate depth data within 1 mm, but needs a relative long acquisition time, and extra effort is required to further calibrate projector with existing mock up. Once we get the 3D geometry of the mockup via the scanner, S'_0 is automatically registered in S_P. Another purpose of the structured light scanner is, given a work piece, we need to get its 3D model to print out the mockup for displaying via 3D printing technique. Given the complexity in geometry and material of the work piece, that is, in practice, the work piece is not necessarily being a plate or pipe, the presence of interreflections, subsurface scatting and defocus, as well as unfavorable surface color, such as black, in these scenario, the performance of scanner will be greatly affected. Here we adapt a multi-gray code pattern based visible/Near Infrared structured light scanning with subpixel refinement.

The gray-code based structured light scanning, once the correspondence between projector's coordinate system and camera's coordinate system has been discovered, is all about solving triangulation (illustrated in Fig. 3.6 [116]) between projector and camera coordinates. The major technical difficulty comes from decoding the captured gray-code illumination sequence (Fig. 3.7 [117]). Fig. 3.8 illustrates a standard gray code based structured light scan and the decoded reuslt. The challenges can be divided into the following categories: 1) non-ideal illumination 2) surface reflection 3) limited resolution.

The critical part of decoding primarily relies on high contrast gray-code pattern sequence. Under visible light condition, certain type of object, such as dark colored work piece can not reflect gray code patterns with enough contrast ratio. Although varying painting material on the work piece provides different reflective behavior from visible or near ifrared light (NIR) [118], the dark colored painting absorbs most of visible light spectrum while being more reflective in NIR lighting condition. Base on this observation, we adapt projector with both visible and NIR light source. Some experiment result can be referred in section 5.3.

Another issue related to SL is interreflections, subsurface scatting and defocus over

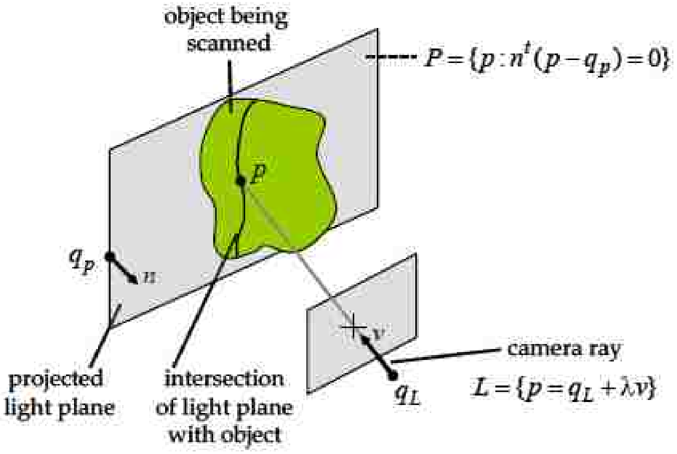


Figure 3.6: Triangulation via line-plane intersection between camera frame and projector frame.

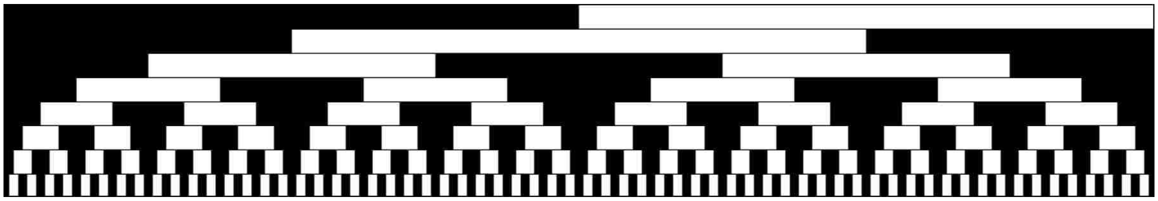


Figure 3.7: Gray Code structured light sequences. The image represents the sequence of bit planes displayed during data acquisition. Image rows correspond to the bit planes encoding the projector columns, assuming a projector resolution of 1024 by 768, ordered from most to least significant bit (from top to bottom).

surface of work piece, as illustrated in Fig. 3.9 [27]. As the work piece in our application are mainly made of metal with/without painting, the interreflections is inevitable under illumination of structured light scanning. Thus we introduced four different types of Gray Code, each corresponding to conquer certain type of the surface. Illustrated in Fig 3.10 [27].

The third issue is that with standard structured light decoding schemes one is limited by the resolution of the projector. That is, while one can decode a corresponding projector pixel coordinate for every pixel in the image frame, the quantization of the projector ultimately limits the accuracy of the reconstruction. We adapt a method [119] based on exploiting the blur induced by the optics of the projector to achieve subpixel resolution of the recovered projector coordinates. By this mean, we can localize scene points more precisely

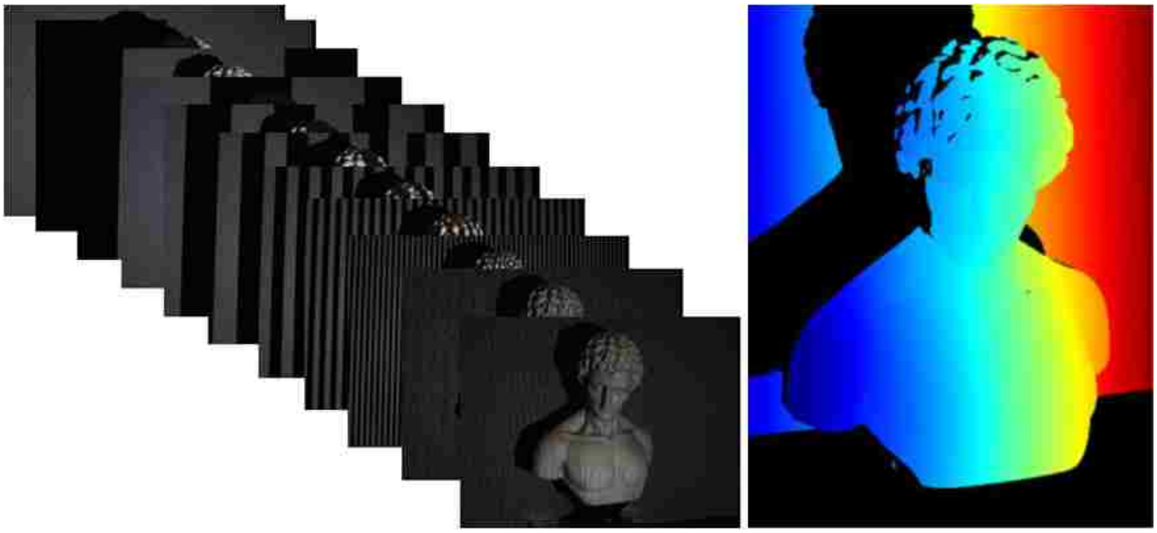


Figure 3.8: Decode result based on gray code structured light scan. Left: gray code illumination sequence. Right: decoded result. Color represent coordinate in projector cooridnate system.

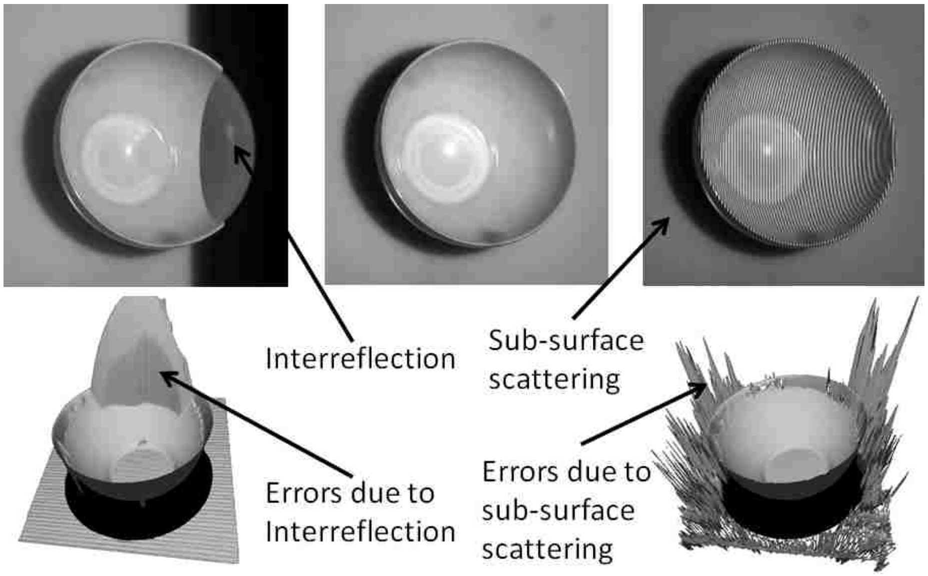

Interreflection
Sub-surface scattering
Errors due to Interreflection
Errors due to sub-surface scattering

Figure 3.9: Strong Interrefelction and sub-surface scattering result unfavourable errors in the recontructed model.

in the projector frame and thus improve the accuracy of the resulting 3D reconstruction. Eight single stripe patterns are introduced to assist the projector blur estimation, Fig. 3.11 shows one of the images where the object is being illuminated with one of the single pixel thick stripe patterns.

The Eq. 3.5 models how the observed intensity of the pixel $I(k)$varies as the stripe is

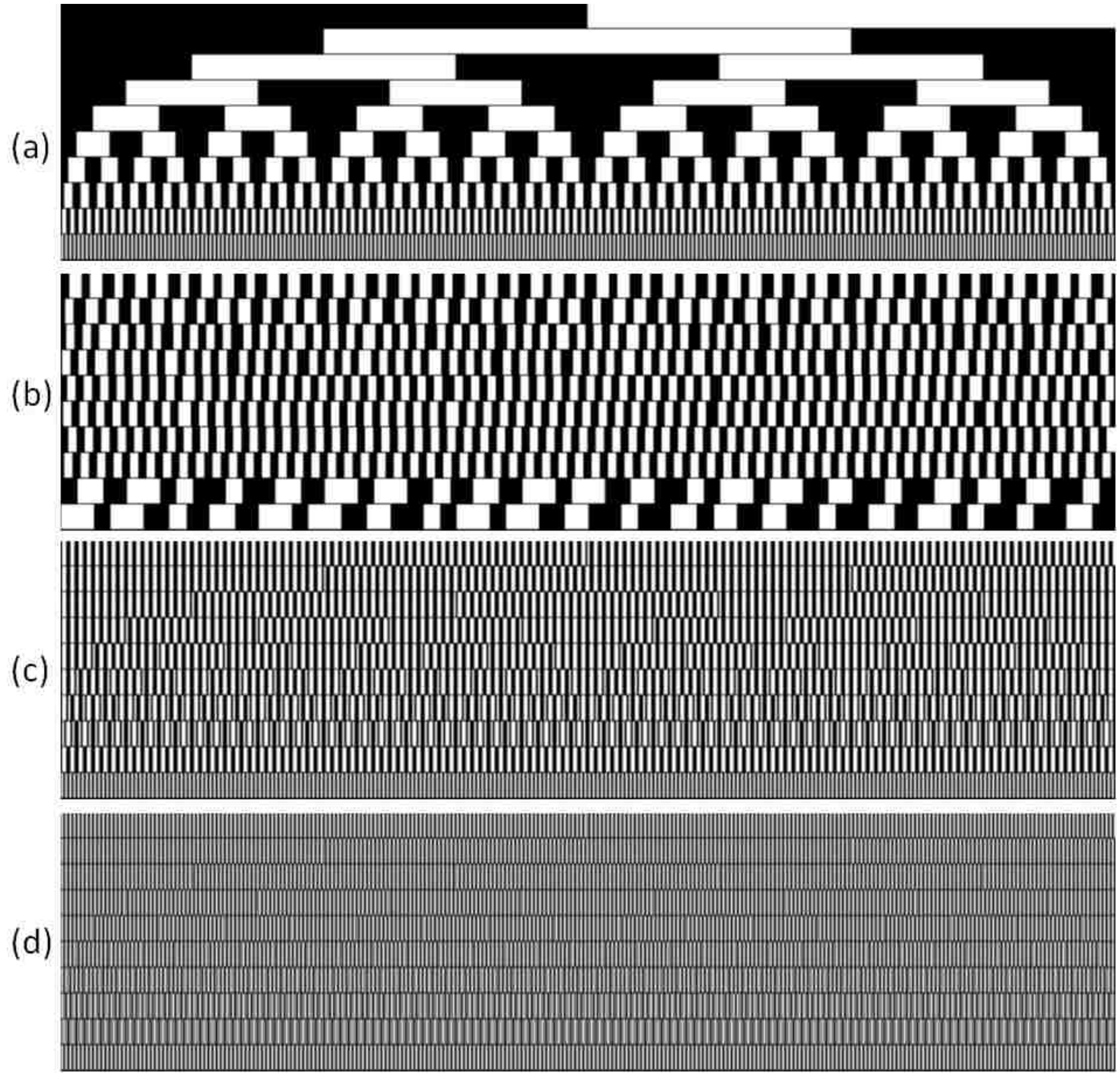


Figure 3.10: Visualization of different binary coding. Bottom two are logical XOR04 and logical XOR02 code respectively.

marched across the scene.

$$I(k) = I_1 exp(\frac{-(k-\delta)^2}{\sigma}) + I_0 \tag{3.5}$$

Where $I_1 = f(\theta_0, \theta_i) E_0 \cos\theta_i$, $f(\theta_0, \theta_i)$ represent the BRDF at the scene point, I_0 represent the scene irradiance due to ambient illumination, k denotes the stripe index from 0 to 7, δ is the projection of scene point in the projector frame, θ models the width of the

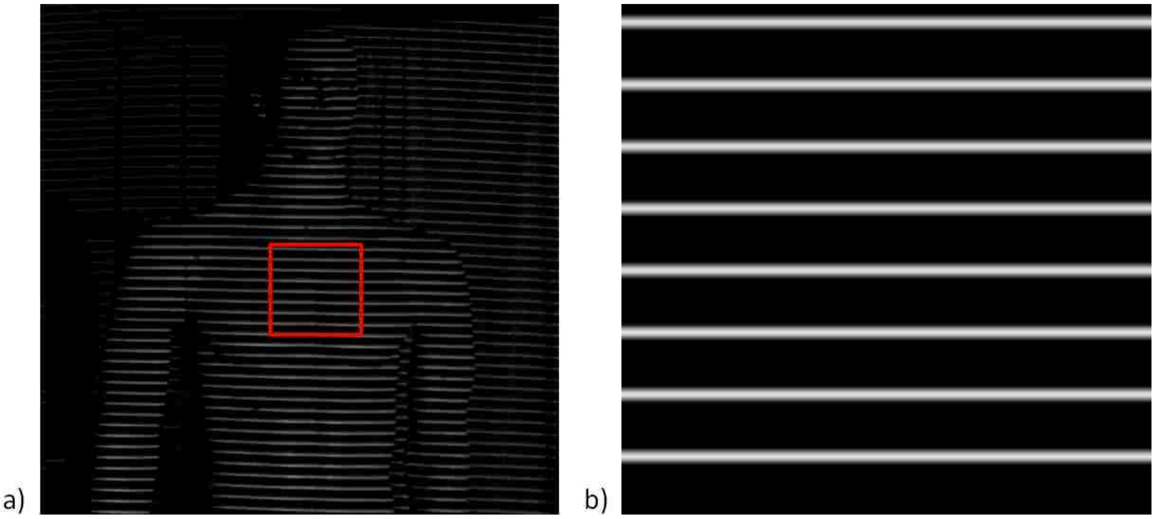


Figure 3.11: a) Object illumiated by a single pixel stripe pattern. b) Close up image on the marked region.

blur kernel at certain point in the scene. Converting Eq. 3.5 into the following equation:

$$log(I(k) - I_0) = (logI_1 - (\frac{\delta^2}{\sigma})) + (\frac{2\delta}{\sigma})k - (\frac{1}{\sigma})k^2 \tag{3.6}$$

Eq. 3.6 demonstrates the relationship between the parameters of interest (δ) and the coefficients of the local quadratic fit. We can thus recovers a floating point offset between -0.5 to 7.5 at each scene point which effectively corresponds to the lower bits of the projection of the scene point in the projector frame.

Combining the three methods for different scenario of the work piece, we present the following algorithm Alg. 2 in pseudo-code. The visualized pipeline is illustrated in Fig. 3.12.

Tool-Hand Calibration

In order to enable the robotic arm sense the 3D environment, local view camera serves as the observation device of the robotic arm, in section 3.2, we discussed the calibration between the camera and the end effector. However, Tool (or any device) attached to the end-effector of the robotic arm is its ultimate interactive subject with the working environment.

Algorithm 1 High Resolution Structured Light Scanning in the presence of global illumination and varying colored painting

```
1: Project patterns and capture images for the 5 codes - two Gray codes ( Conventional Gray $GC$ and Gray codes with maximum min-SW $GMM$), the two logical codes (XOR02 and XOR04), and the single pixel stripe pattern $GS$.
2: Compute the depth values under the $GC$ and $GMM$ using conventional decoding and XOR02 and XOR04 using the logical decoding.
3: Compare the depth values. If any two codes agree, return that value as the correct depth. If the two Gray codes $GC$ and $GMM$ agree, return the value computed by $GMM$.
4: Compute the depth values under $GS$ using subpixel refinement method and improve the resolution of the result in step 3.
5: Mark the camera pixels where no two codes agree as error pixel.
6: Mask the pattern so that only the scene points corresponding to the error pixels are lit [120]. Repeat steps 1-6 to progressively reduce the residual errors.
```

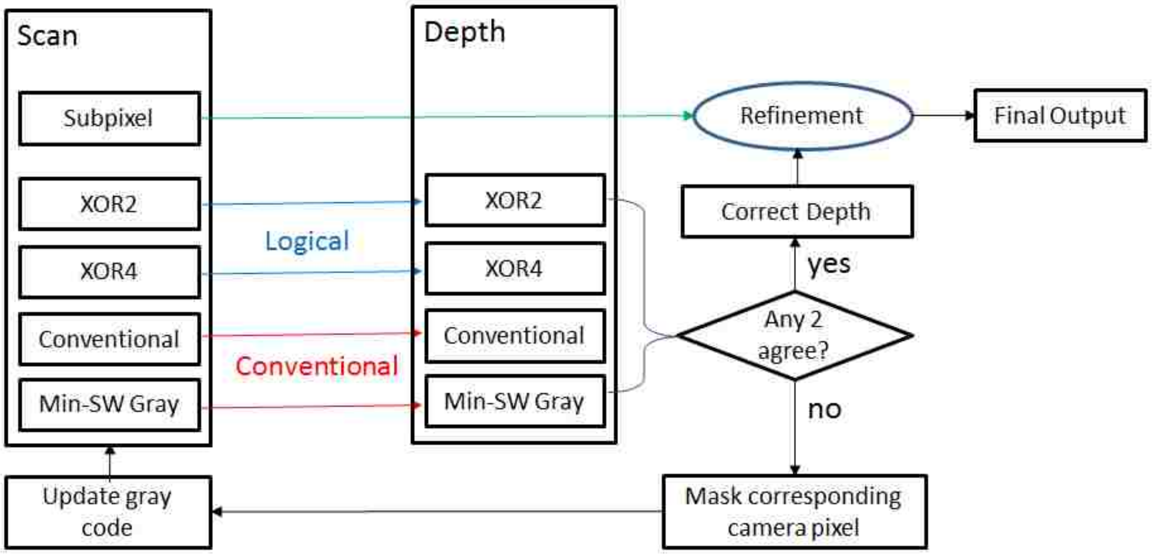


Figure 3.12: The pipeline of High Resolution Structured Light Scanning algorithm.

Thus, it is necessary to calibrate the frame of tool S_T and frame of end-effector S_H.

In real setup, geometric relationship between S_T and S_H is unknown due to non-standardized mounting device of local view camera. A trivial method is provided by Microsoft Kinect Fusion. However, due to limited GPU memory and depth sensor resolution [121], this method is not suitable on our application which requires millimeter level of accuracy. Based on the algorithm proposed in subsection 3.2, we therefore proposed a method which combines the structured light scanning, iterative closet point(ICP) and principle component analysis (PCA) to discover the transformation between S_T and S_H.

The general pipeline of the proposed algorithm is illustrated in Fig. 3.13.

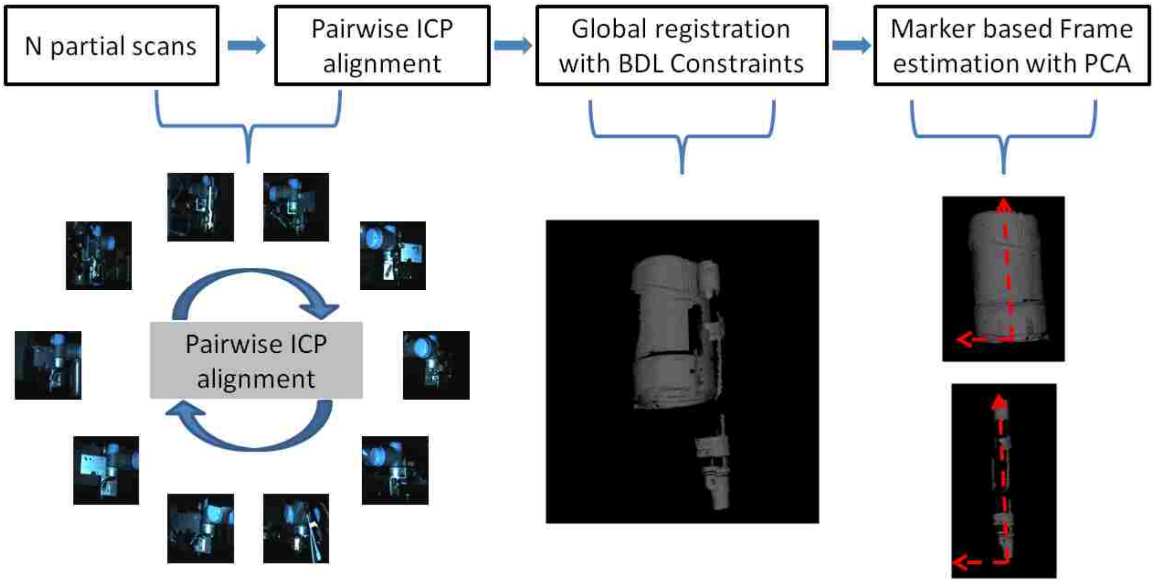

N partial scans
Pairwise ICP alignment
Global registration with BDL Constraints
Marker based Frame estimation with PCA
Pairwise ICP alignment

Figure 3.13: Overview of real welding workstation. The global view camera is facing towards the robotic arm and work piece.

Now we have a loop of n scans $M_i, i = 1, \ldots, n$, the graph T_1 is aligned with M_1 correctly and we use it as the embedded graph to register M_1 to M_2. After the registration, M_1, T_1 are deformed as $M_{1,2}, T_{1,2}$ and transformations are denoted as $\left\{\left(\mathbf{R}_1^k, \mathbf{t}_1^k\right)\right\}$. Using the weight and node indices of T_2 but the node positions of $T_{1,2}$, we register M_2 to M_3 and get $M_{2,3}, T_{2,3}$. The process continues until registering T_n back to T_1 having transformations $\left\{\left(\mathbf{R}_n^k, \mathbf{t}_n^k\right)\right\}$ and we call this process the pairwise registration. For a globally correct registration, we have $T_{n,1} = T_1$, that is for each node, $\mathbf{t}_1^k + \mathbf{t}_2^k + \cdots + \mathbf{t}_n^k = 0$, and the deformed mesh $M_{n,1}$ is consistent with M_1. Given the deformation multiplication property when the deformation is highly rigid, the total deformation should be an identity and we have the rotation consistency constraint, $\mathbf{R}_n^k \mathbf{R}_{n-1}^k \cdots \mathbf{R}_1^k = \mathbf{I}$.

Due to error accumulation, the pairwise registration will drift to violate such constraints. Similar to [122], we distribute the accumulated rotational and translational error individually and choose a weight $w_i = 1/Dist(M_{i,i+1}, M_{i+1})$ to transformations $\left\{\left(\mathbf{R}_i^k, \mathbf{t}_i^k\right)\right\}$, where $Dist(M_{i,i+1}, M_{i+1})$ is the average fitting error of E_{fit}, for all $i = 1, \ldots, n$. ($n+1$ we refer to 1.) Since the error distribution of each node is performed in the same way, we ignore the superscript k in the following for simplicity.

The translational error is distributed by solving the following optimization,

$$\min \sum_{i=1}^{n} w_i^2 \left\|\hat{\mathbf{t}}_i - \mathbf{t}_i\right\|^2, \quad s.t., \sum_{i=1}^{n} \mathbf{t}_i = 0, \tag{3.7}$$

and the solution is found using Lagrange multipliers, $\hat{\mathbf{t}}_i = \mathbf{t}_i - \alpha_i \sum_{j=1}^{n} \mathbf{t}_j$, with the scalar α_i as

$$\alpha_i = \frac{1}{w_i^2} \Bigg/ \sum_{j=1}^{n} \frac{1}{w_j^2} \tag{3.8}$$

The rotational error distribution is to minimize the total rotational deviation:

$$\min \sum_{i=1}^{n} w_i \angle(\hat{\mathbf{R}}_i, \mathbf{R}_i), \quad s.t., \mathbf{R}_n^k \mathbf{R}_{n-1}^k \cdots \mathbf{R}_1^k = \mathbf{I}, \tag{3.9}$$

where the angle between two rotations is defined as $\angle(\mathbf{A}, \mathbf{B}) = \cos^{-1}\left(\frac{tr(\mathbf{A}^{-1}\mathbf{B})-1}{2}\right)$. Analyzed in [123], the optimal $\hat{\mathbf{R}}_i$ is computed as

$$\begin{aligned} &\hat{\mathbf{R}}_i = \mathbf{E}_i^{<\alpha_i>} \mathbf{R}_i, \\ &\mathbf{E}_i = (\mathbf{R}_k \mathbf{R}_{k-1} \cdots \mathbf{R}_1 \mathbf{R}_n \mathbf{R}_{n-1} \cdots \mathbf{R}_{k+1})^{-1}, \end{aligned} \tag{3.10}$$

where α_i is referred to equation 3.8, and $E_i^{<\alpha_i>}$ is defined to be the rotation matrix that shares the same axis of rotation as E_i but the angle of rotation has been scaled by α_i.

Once all the optimal $\left\{\left(\hat{\mathbf{R}}_i^k, \hat{\mathbf{t}}_i^k\right)\right\}$ are obtained, we use the total transformation $\left\{\left(\left(\hat{\mathbf{R}}_1^k \ldots \hat{\mathbf{R}}_{i-1}^k \hat{\mathbf{R}}_i^k\right)^{-1}, -\hat{\mathbf{t}}_i^k - \hat{\mathbf{t}}_{i-1}^k - \cdots - \hat{\mathbf{t}}_1^k\right)\right\}$ to deform the mesh M_i with $T_{i-1,i}$ back to M_1. After all the meshes M_i are updated, we can repeat the pairwise registration step from M_1 and T_1. The graphs $T_1, T_{1,2}, \ldots, T_{n,1}$ will finally converge to a constant graph and $\left\{\left(\hat{\mathbf{R}}_i^k, \hat{\mathbf{t}}_i^k\right)\right\}$ converges to the globally optimal solution.

In the sense that the effect of error distribution step can be considered to prevent the graph drifting and pull it towards the optimal position, we do not only use high rigidity and regularization weights, but also perform an interleaved bi-directional error distribution to make it more robust to large errors. The basic idea is to perform an inverted iteration using the order of $M_1, M_n, M_{n-1}, \ldots, M_3, M_2, M_1$ after a forward directional iteration. The directional scheme is in essential the same to the multiple cycle blending technique described in [123] and the total time complexity to convergence is the same because they traverse both direction in one iteration and we perform each direction once but need two iterations.

To summarize the algorithm:

Algorithm 2 Robust Transformation estimation of rigid objects

```
1: Scan and reconstruct multiple partial patches of the objects, $M_i, i = 1, \ldots, n$.
2: Compute the pairwise rigid ICP between partial scans of adjacent view, $\{M_i, M_i + 1 | i = 1, \ldots, n-1\}$, the resulting transformation are denoted as $\left\{\left(\mathbf{R}_1^k, \mathbf{t}_1^k\right)\right\}$.
3: Process the pairwise transformation in step 2 by applying bi-directional loop constraint (BDL), the optimal $\left\{\left(\hat{\mathbf{R}}_i^k, \hat{\mathbf{t}}_i^k\right)\right\}$ are obtained. Repeat step 2 and step 3, until the $\left\{\left(\hat{\mathbf{R}}_i^k, \hat{\mathbf{t}}_i^k\right)\right\}$ converges to the globally optimal solution.
4: Once all partial scans $M_i, i = 1, \ldots, n$ are registered to the target scan $M_1$, the final surface $S$ is extracted by using Screened Poisson Surface method [124].
5: Isolate the surfaces of two rigid object $S_1$ and $S_2$. Run PCA on $S_1$ and $S_2$ respectively.
6: With the 3D coordinates of the predefined marker on the surface of $S_1$ and $S_2$, together with the main axis calculated in step 5, $Z_1$ and $Z_2$, estimate the frame of $S_1$ and $S_2$, and then calculate the transformation between these two rigid objects $\{(R, t)\}$.
```

Projector-Eye Calibration

Regarding the calibration of the projector and the local-view camera (S_C, S_P), it involves two steps. In the first step we use the auxiliary camera to calibrate the projector-camera using standard calibration techniques [125]. In the second step, we scan the 3D mockup. We assume that S_C and S_P are sharing the same coordinate frame S_0 since the camera and the projector are aiming at two identical objects. From the scanned mockup we can estimate transformation ${}_0T^p$. Let ${}_0T^c$ be the homogeneous transformation bridging S_0 and S_C. ${}_0T^c$ is the extrinsic parameters of the local view camera. By defining N markers on both the work piece and the mock-up, we can get a set of coordinate pairs:

$$S = \left\{ \left(p_i, p_i'\right) \middle| \, p_i \in S_P, p_i' \in S_C, i = 1, \cdots, N \right\} \tag{3.11}$$

The extrinsic ${}_cT^p$ between S_C and S_P can then be solved:

$$E = \begin{bmatrix} E_1 \\ \vdots \\ E_N \end{bmatrix} = 0 \tag{3.12}$$

where $E_i = \|p''_i - Hp_i\|^2 - \left\|p_i - H^{-1}p''_i\right\|^2$ is symmetric transfer error.

Table 3.1: Cameras and projector configuration.

terms	${}_0T^c$	${}_0T^p$	${}_cT^p$	${}_bT^c$
error	0.632	1.332	2.382	$\theta = 5.88 \times 10^{-3\circ}$ $t = 0.1mm$

Let the coordinate system of the screen be represented as a Cartesian coordinate system, $\{O_s, X_s, Y_s, Z_s\}$, using unit of pixel. Let the coordinate system of the robotic arm be represented as a Cartesian coordinate system, $\{O_r, X_r, Y_r, Z_r\}$, using unit of mm. The System defines a certain number (denote as N) of marker points in the screen coordinate system, $P_i'(x,y), i = 1, \cdots, N$. By pointing the weld gun at corresponding marker point on the work piece, system can record the coordinate of end effector of robotic arm P_i. we can get following set of coordinate pairs:

$$S = \{(P_i', P_i) | i = 1, \cdots, N\} \tag{3.13}$$

In order to calculate the projective matrix M between the two coordinate systems, the unit should also be unified. Let the pixel density of the screen to be α, then we can get a new set of coordinate pairs:

$$S_1 = \{(P''_i, P_i) | i = 1, \cdots, N\} \tag{3.14}$$

where $P''_i = \frac{P'_i}{\alpha}, i = 0, \cdots, N$.

We can get these two coordinate systems calibrated by solving the homography in non linear method:

$$E = \begin{bmatrix} E_1 \\ \vdots \\ E_N \end{bmatrix} = 0 \tag{3.15}$$

where $E_i = \|p''_i - Hp_i\|^2 - \left\|p_i - H^{-1}p''_i\right\|^2$ is symmetric transfer error.

To summarize, given a pose (based on the feedback from the robotic controller) in S_B, with known ${}_bT^t, {}_tT^c$, our system can estimate the pose of the local view camera, then, by applying ${}_cT^p$,the visual information captured by the local view camera can be correctly rendered and projected onto the mock-up. The details of the projection correction algorithm can found in [57]. We have verified that our reprojection error is less than one millimeter on the mockup throughout the movement range of the robotic arm.

3.3 Experiment and Usability Evaluation

An initial user study was conducted in order to determine any possible impacts on performance due to the VR Pipe display. The selected usability evaluation methods were based on the premise that our virtualized welding workstation will eventually be used to both directly control and monitor welding robots. Rather than studying both functions, however, we chose to focus on the monitoring function of the end user for two reasons. First, existing research has more frequently focused on direct teleoperation [126, 127] and thus there is a greater need for a better understanding of the design parameters that enhance monitoring (supervisory control) tasks. Second, as training of the robots progresses, the monitoring role will become a larger and larger portion of the human operators task.

Method

We here define successful monitoring as the operators ability to create and maintain accurate mental models of the weld pool and weld site in order to accurately predict welding outcomes, thus facilitating timely human intervention when needed. Because monitoring tasks in related domains (e.g., UAV control) are known to become more cognitively demanding as physical engagement is reduced, we believe cognitive outcomes such as reduced mental workload, enhanced confidence, and user preference are as critical as objective performance in predicting long-term success of the system, and for identifying opportunities for improvement.

Participants

12 students participated in the current study. 5 of the participants were second-year welding students (3 male, $\bar{x}$ age = 36) from a local community and technical college and 1 participant was a certified welding educator (male, 43 years age). The remaining 6 participants were graduate students from a university (2 male, $\bar{x}$ age = 24). Participants were paid for their efforts. To assure that there were no significant differences between groups, population was included as a fixed factor in all RM-ANOVAs. No main effect of population was observed for any of the dependent measures $F(1,10) > 2.4, p > 0.14$.

Stimuli and apparatus

15 welds running parallel across the surface of a steel pipe were created (see Fig. 3.14(a)). Visual appearances of the welds were altered by manipulating direction, length, amperage, speed, and end point of the weld. Video of each weld being performed was captured for use in the three display conditions L, G+L (Fig. 3.15) and HRD (Fig. 3.14(e-l)). To ensure

that all the welds were visually distinct, 6 sets of 3 welds each were chosen from 9 of the welds (3 training sets, 3 experimental sets).

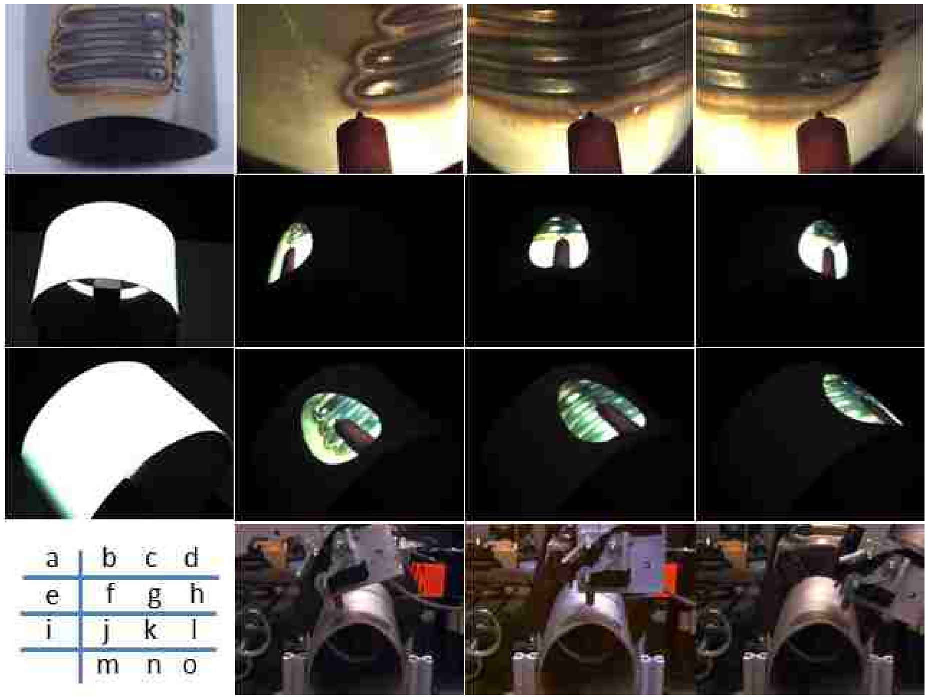


Figure 3.14: Visualization of the work piece. (a) global view of the work piece; (b-d) local view of the work piece; (e-h) blank HRD and HRD of b-d; (i-l) observation of HRD from different view point; (m-o) global view of the working environment corresponding to b-d. The frames are cpatured without electric arc for visual purpose.

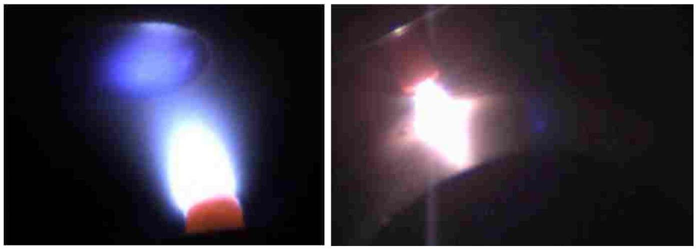


Figure 3.15: Screen shot of experiment for user study. Left: local view; Right: Global view; During study, the global view and local view are displayed on screen side by side. These two views are cropped and image-enhanced for viusal purpose in this paper.

Procedure

Participants were greeted, given a brief over view of the study, and asked to sign informed consent forms. An overview of the welding apparatus was provided including a simple explanation of the welding hardware, camera positions and the created welds. This overview included the factors that were manipulated in order to achieve visual differences between the different welds. The participants whom were unfamiliar with welding were given a slightly more in-depth overview if there they were unsure about any of the information. Participants then performed the matching-to-sample task. First, participants were handed the metal pipe and asked to inspect a set of three weld samples and specifically informed to remember differences between them. Participants then watched the weld being performed on one of the three display types (target weld), and were not allowed to look at the pipe for reference while the videos were being presented. After the video finished, participants were instructed to describe which of the three weld samples matched the target weld, and their confidence in this response on a 1 (low confidence) to 10 (high confidence) scale. Feedback was provided on the accuracy of their response, and if they were incorrect a second opportunity was given. Feedback was again provided on the second attempt. A block of three trials, one for each display, was used as training prior to the experimental procedure. For the experimental block, each display condition was encountered 3 times for a total of 9 trials. Display and target weld order were randomized for both the training and experimental blocks. After each non-training trial of the matching to sample task (i.e. after a correct responses or 2 incorrect responses), a NASA-RTLX questionnaire was administered. Upon completion of all 9 trials, participant completed a time allocation questionnaire and an open response questionnaire where they reported positive and negative aspects of each display.

Results

All dependent measures were submitted to a repeated-measures analysis of variance (RM-ANOVA). If further investigation was warranted, two-tail pairwise comparisons were conducted using the HRD as the reference condition.

Comparison task performance and confidence

The matching-to-sample comparison task was used to assess the ability of the user to accurately predict the final weld given information provided by the displays. In this regard, the measure may be considered a test of the users situational awareness afforded by the current displays [128]. Performance was high across all conditions, with participants on average requiring less than 2 attempts to identify the welds (HRD $\bar{x} = 1.5$, L $\bar{x} = 1.47$, G+L $\bar{x} = 1.63$), and differences did not achieve significance $F(2,22) = 0.905, p > 0.4$. However, differences were observed in participants confidence in their answers after viewing a particular display $F(2,22) = 9.82, p < 0.001$, see Figure 3.16(right) for results. On trials where the HRD was viewed, participants reported higher confidence in their selection ($\bar{x} = 8.25$) than with the Local display ($\bar{x} = 7.833$), $t(11) = 3.29, p < 0.01$, or the Global+Local display ($\bar{x} = 7.08$), $t(11) = 2.32, p = 0.04$.

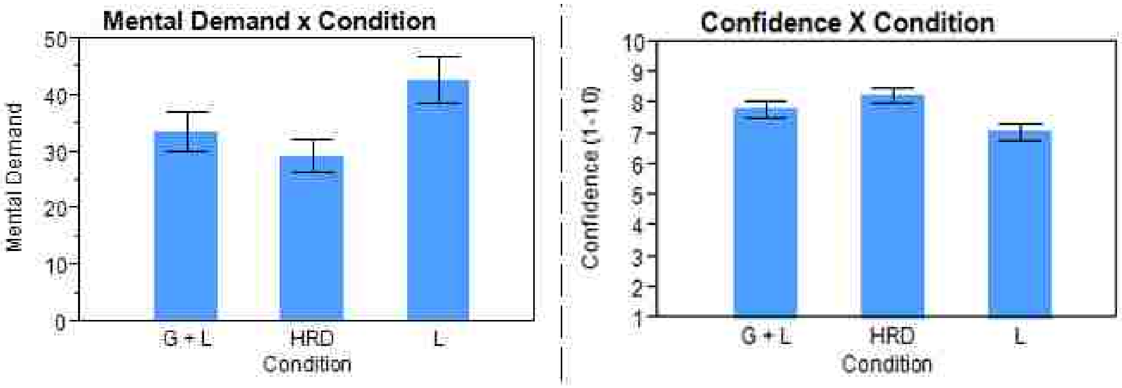


Figure 3.16: Left: Mental Demand subscale score from the NASAS-RTLX rating scale. Mental Demand was rated on a 0 (low) - 100 (high) scale. Error bars are constructed using 1 standard error from the mean. Right: Graph of reported confidence of the users in their first response. Confidence rating were reported on a 1 (low confidence) 10 (high confidence) scale.

Workload analysis:NASA-RTLX

The NASA Raw Task Load Index (NASA-RTLX) version of the NASA-TLX [129] was administered to assess subjective workload demands between displays. The NASA-RTLX is a multidimensional rating scale that consists of 6 separate subscales (i.e. mental demand, physical demand, temporal demand, performance, effort, and frustration). Previous studies have demonstrated high correlations between the weighted means of the TLX and unweighted means of the RTLX [130]. All 6 subscales were averaged to create an overall measure of workload. No significant workload differences were observed between conditions, $F(2,22) = 2.258, p = 0.12$. It was hypothesized that the cognitive demands dimension of the RTLX would be more sensitive to differences in the display, and this turned out to be correct, $F(2,22) = 6.41, p < 0.01$. Participants reported greater cognitive workload requirements (Fig. 3.16(left)) in the Local display condition $(\bar{x} = 42.83)$ than the HRD condition $(\bar{x} = 29.30), t(11) = -3.29, p < 0.01$. The difference between the HRD and Global+Local display did not achieve significance, $t(11) = -1.61, p = 0.134$.

Time allocation

For the time allocation procedure, participants were presented with a hypothetical situation in which they could choose any of the displays for a similar monitoring task. However, the displays were located in different areas, and only one display could be viewed at a time. Participants were given 100 hours of work time to allocate between the three displays.

As the data violate the assumption of independence, a standard RM-ANOVA could not be conducted. Instead, the data was submitted to a Related-Samples Freidmans Two-Way ANOVA. Although a significant results were achieved for the total study population, analysis was conducted selecting only welding students as an assumption was made that they would be better able to estimate the utility of each individual display in a more traditional welding environment. See Figure 3.17 for results. A significant effect was found for how welders would allocate their time across the displays, $X_r^2(2) = 7.71, p < 0.025$. Participants estimated that they would spend a majority of their time monitoring the HRD $(\bar{x} = 67)$, a significantly greater amount than with either the Local display $(\bar{x} = 20)$, or Global+Local display $(\bar{x} = 13)$.

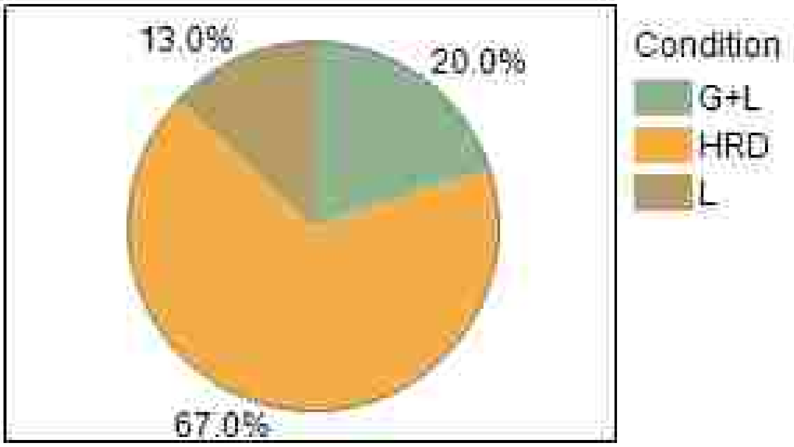


Figure 3.17: User estimate time allocation between displays for use in a monitoring task.

User study discussion

The HRD demonstrated several advantages in the user study. First, although overall workload was consistent across displays, mental workload was significantly decreased when monitoring with the HRD as opposed to the Local display. We speculate that this may be due to participants not needing to perform high-demand mental spatial transformations to account for weld pool travel over the geometry of the pipe. This would also explain why differences in mental workload did not achieve significance between the HRD and the Global+Local display, as the Global display mitigates some of the necessary transformations. Future research may further investigate these claims by increasing the variation in what differentiates weld types and investigating strengths and weaknesses of the other displays.

When welding students were asked about how they would opt to allocate their time between displays in a monitoring task on the job, they chose to allocate a majority of their time to HRD. While we only have anecdotal evidence to suggest an explanation for this, many of the welders commented that HRD allowed for monitoring of the weld pool and the spatial location of the weld, while the global display would be utilized to monitor the performance of the welding hardware, a task not afforded by HRD or Local display. Participants allocating some time to the Local display often commented that the higher resolution of the display allowed for increased detection of subtle weld pool features. This opens the future possibility of dual display systems (e.g.HRD + Global display) and a need for higher fidelity imaging in the HRD.

A question remains as to why the participants did not actually perform better in the matching-to-sample task with any particular display. The reason for this may lie in the simplicity of the welds for testing, which allowed students not familiar with the welding process to perform at a high level regardless of display type. Specifically, all welds followed a relatively simple and consistent path across the weld surface, allowing users to focus on specific differences in determining the target weld (viz. direction and end point). While the performance measure did not prove sensitive enough to demonstrate a difference between displays, participants viewing the HRD reported higher confidence in their responses than with the other display types. This may suggest the HRD allowed for more robust mental representations of the weld, but further research is necessary. Increasing the complexity of the weld path, especially though minute lateral deviations, may better demonstrate the benefit of the HRD.

3.4 Conclusion

In this chapter, we have proposed a novel hybrid reality system for tele-operated weld monitoring. A high accuracy 3D scanning technique is utilized to create digital models of work pieces to be welded. Based on the type of welding job, a mock-up is constructed from a set of templates. The welding process is captured by the camera mounted on the robotic arm and visualized on the mock-up using projectors. The welder can see the welding process as if she/he were next to the actual welding. User studies show that our HRD has reduced the mental workload and is preferred by welders.

Chapter 4 Teleoperation of Remote Welding

Welding is a widely used manufacturing process that is labor intensive and sometime hazardous. While industrial welding robots have been in use for several decades, they are pre-programmed actuators with limited, if any, intelligence. As a result welding robots are primarily used in well-controlled environments, such as assembly lines for mass production, in which the work pieces may be accurately prepared and positioned at reasonable costs. Given that manufacturing is moving towards more customized productions, the next generation of welding robots that can intelligently adjust to various welding tasks is urgently needed. Unfortunately, equipping robots with intelligence is challenging.

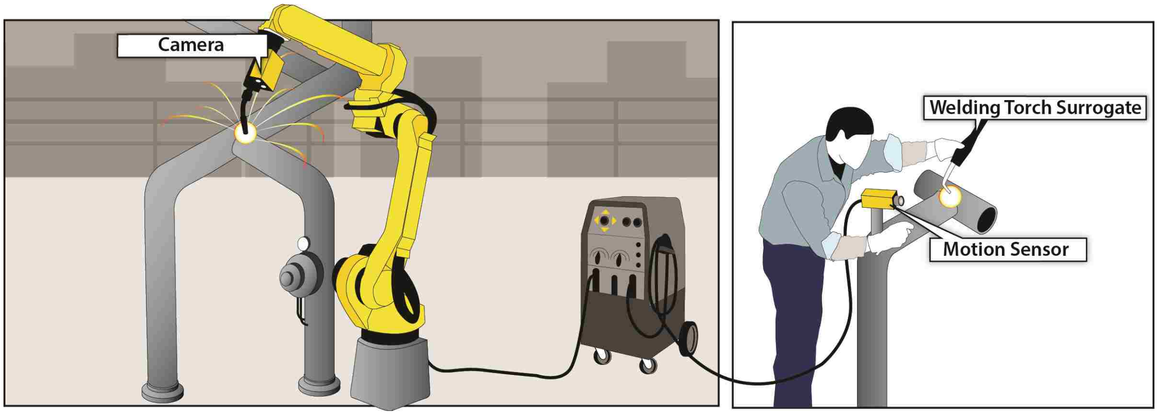


Figure 4.1: Virtualized Welding: Above: an illustration of our virtualized welding operation, in which existing welding robots are augmented with 3D sensors and video cameras to capture the working environment. The operator can monitor and control the welding process in an augmented display setup in which welding images are projected on a mock-up 3D surface. The operators motion is tracked.

In this chapter we present a prototype of our virtualized welding system (Fig. 4.1), which is an improved version of our previous prototype that had no robot control [131]. It is developed using commodity sensing and display components. On the visualization aspect, it contains a hybrid-reality display (HRD) system, which utilizes projectors to project a captured video image onto a 3D replica of the actual weld surface. It provides a direct alignment between the frame of reference for the operator and that of the displayed image [131]. On the human-robot interaction aspect, we developed and tested several interaction means, including the use of hand tracking and traditional control with a 3D mouse. The details of various system components and methods to calibrate and control them are presented. Results show that our system can provide better control accuracy, in particular when the welding process is carried on complex surfaces. We contribute is success to our surrogate-based projection workstation, which provides a natural means to support navigation and teleoperation tasks.

4.1 System Platform

Similar to the system proposed in 3.1, There are two workstations in our virtualized welding system (Fig. 3.1): real welding workstation and virtual welding workstation. The real welding workstation (illustrated in Fig. 4.2) is primarily for conducting the welding task on the work piece while acquiring visual information of the work piece simultaneously. It contains the visual information acquisition system and the robot control system. The virtual welding workstation, is responsible for visualizing the work piece while tracking tool's motion. Data communication between there two workstation is linked by network.

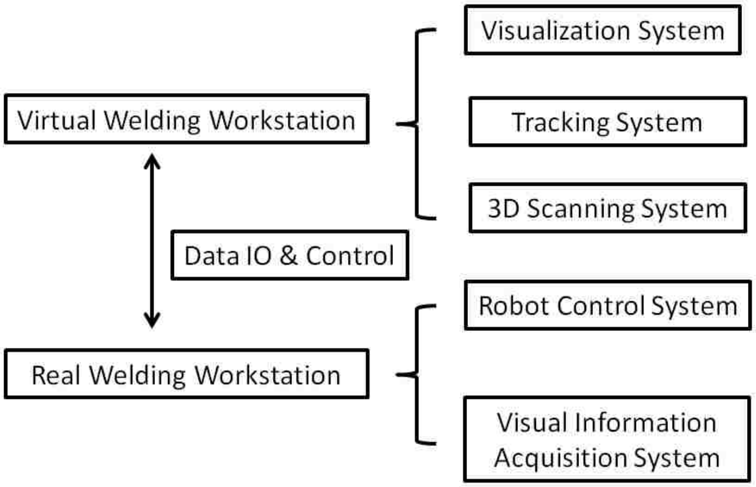


Figure 4.2: General structure of the virtualized welding system.

Real Welding Workstation

Illustrated in Fig. 4.3, The major component of real welding workstation are (1) HD camera (Point Grey Flea 1394a CCD camera, 1024×768), which is referred to as the local view camera; (2) robotic arm (Universal Robots UR5 6 axis robot arm, 1/10 mm accuracy). The real weld gun and a video camera as a group is rigidly mounted to the end effector of the robotic arm. The local view camera, locally observes the work piece. The view angle of the camera is adjusted to mimic that of human welders'. Thus a more realistic visualization experience could be achieved when the visual information taken by the local view camera is rendered on HRD. The robotic arm follows the command from the virtual welding workstation and drives the end effector in real time. In addition, a 3D scanner based on structured light is utilized to scan the work pieces in high resolution (depth accuracy is less than 1mm, refer to Fig.4.12).

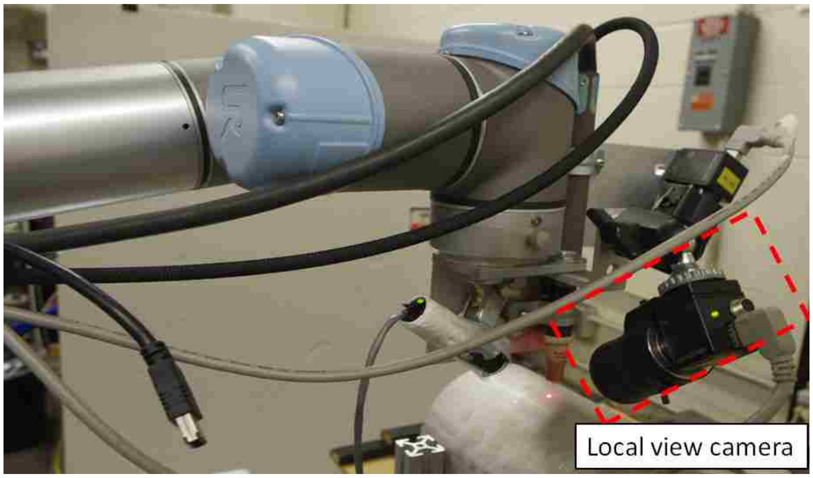


Figure 4.3: Overview of the real workstation.

Virutal Welding Workstation

A HD projector (DLP, 1920×1080) is used to project image from the real-welding station onto the mockup. Since the mockup and the actual work pieces have the (almost) identical surface geometry, or differs only in scale, the resulting HRD is autostereoscipic and provides the same spatial cue as in the real welding. In order to achieve that, the projector must be calibrated with respect to the mockup. This requires the use of an auxiliary camera (Point Grey Flea3, 1600×1200) to observe the projected images for calibration purpose. In addition, the projector-camera pair can be used as a 3D scanner. We usually rescan the mockup to accommodate the errors introduced in 3D printing or assembly. The virtual welding workstation is illustrated in Fig. 4.4. A 3D mockup of the work pieces is reconstructed in the virtual welding station. The mockup can be assembled from identical work pieces, or in our case, 3D printed. By utilizing the structured light scan technique [116] with subpixel refinement [119] for reconstructing geometry of 3D objects, this system can provide a high accuracy point cloud of the mock up. Visualization system can utilize this accurate measurement of the mock up while rendering video on hybrid reality display. Since the mock-up has the same 3D geometry as the real work piece and the projector is calibrated w.r.t the mock-up, the display provides accurate spatial context and 3D cues. A motion tracking sensor (Leap Motion sensor, 1/100 mm accuracy)is employed to track

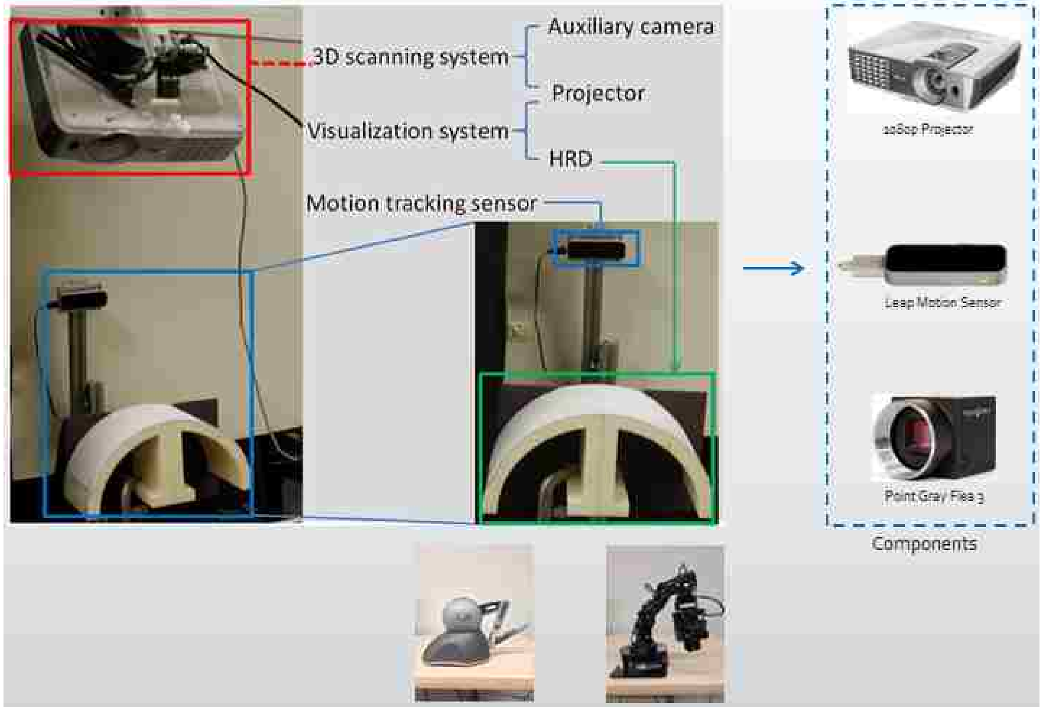


Figure 4.4: Detailed view of the virtual welding workstation. Major components are a mock up, a motion sensor, an auxiliary camera and a projector. The mockup is generated by 3D printer.

tool's motion in the virtual welding workstation.

Table 4.1: Cameras and projector configuration.

config	Camera(aux)	Camera(local)	projector
resolution	1600×1200	1024×768	1920×1080
focal length	3697.3	2728.3	2023.8
frame rate	15 FPS	30 FPS	60 FPS

System Workflow

As illustrated in Fig 4.5, the motion sensor in virtual welding workstation monitors possible motion of the tool of specific shape. Valid motion will trigger a motion command from motion sensor to robotic arm. Local view camera keeps sending visual information to virtual workstation for rendering on HRD, and video rendering is self-running. Since the camera is rigidly mounted on robotic arm, A motion of robotic arm will cause a view change, which consequentially results a different observation on HRD. User will adjust their movement based on the observation.

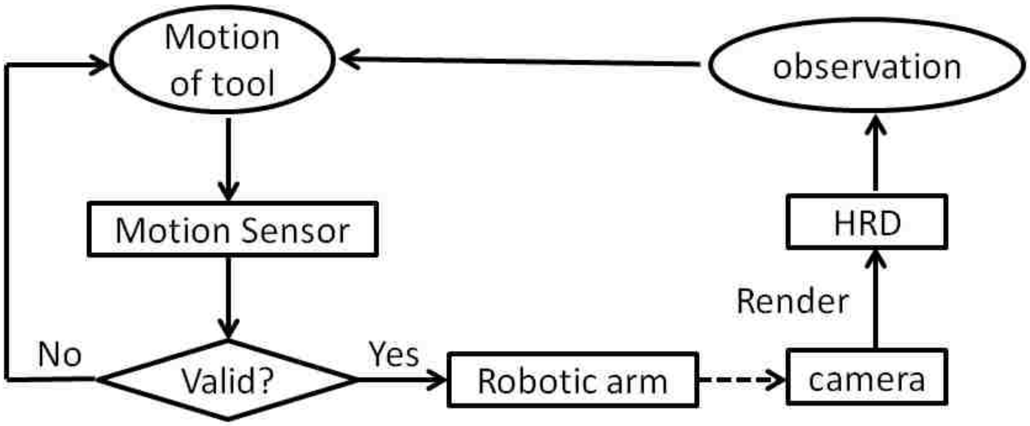


Figure 4.5: System work flow.

4.2 System Calibration

Extensive calibrations between components of the proposed system are conducted before the system performs the welding and visualization task. Fig. 4.6 illustrates these calibrations in our system.

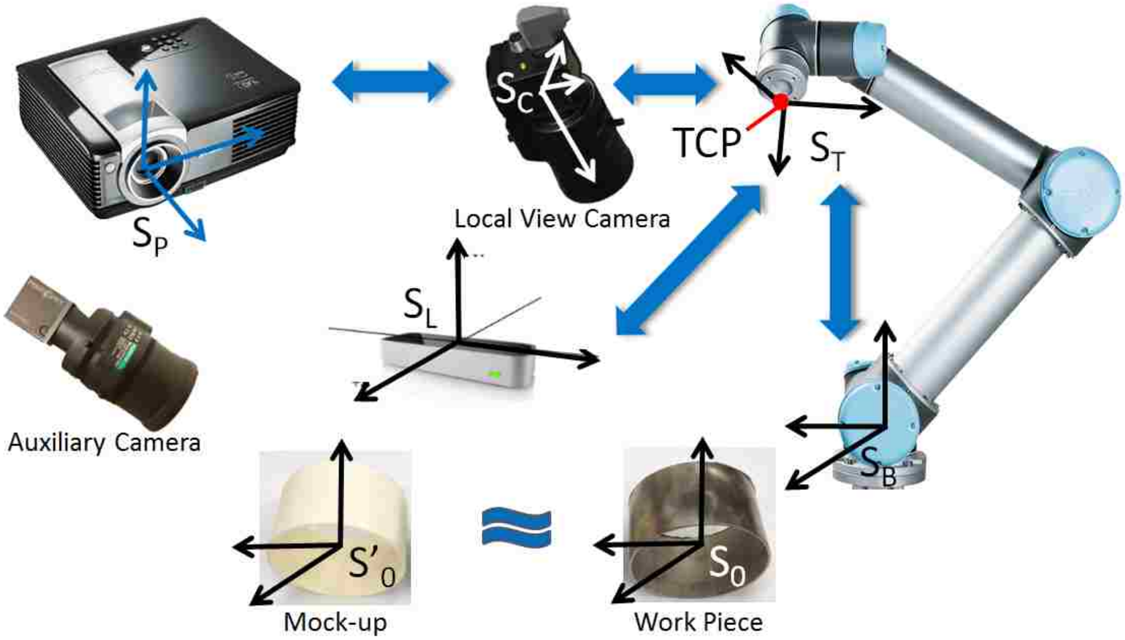


Figure 4.6: Calibration of the virtualized welding system. Projector, local view camera, motion sensor and robotic arm are calibrated together.

In Section 3.2, the calibrations of most elements in the system have been discussed. Thus, in this section, we only cover the calibration of motion sensor and the TCP frame.

calibration between motion sensor and TCP frame (S_L, S_T) can be conducted based on Eq. 3.13 and Eq. 3.15. Different from (S_C, S_P) calibration, where the predefined point can be observed, motion sensor does not detect the 3D coordinate of predefined point on HRD surface, especially when occlusion presents. Thus a method of triangulation should be applied in order to get correct corresponding point pair. Denote a marker on surface of HRD as P in motion sensor coordinate system, and as R in S_T. By aiming the weld gun toward P at different orientation, we can define two lines as:

$$p_1 = q_1 + \lambda_1 v_1, p_2 = q_2 + \lambda_2 v_2 \tag{4.1}$$

By solving a cost function for triangulation, of line-line intersection.

$$\begin{aligned} \Phi(p, \lambda_1, \lambda_2) = \|q_1 + \lambda_1 p_1 - p_1\|^2 - \|q_2 + \lambda_2 p v_2 - p_2\|^2 & \\ w.r.t \frac{\partial \Phi}{\partial p} = (p - p_1) + (p - p_2) = 0 \end{aligned} \tag{4.2}$$

we can get the 3D coordinate of P in motion sensor space S_L. In order to get R in S_T, robotic arm should be driven so that TCP can touch the marker on surface of work piece. Now, with given point pair R and P, we can calculate ${}_lT^t$.

4.3 Predictive Control of Robot Speed

In this section a predictive control algorithm is derived. By setting a relatively large robot speed, tracking performance is guaranteed. However, for large robot speed, the robot is suffered from large vibration with consistent accelerating and decelerating. Thus, it is preferred to track the command movement signal (human hand movement speed) with minimum robot speed. In this case, a pre-defined robot speed may not be sufficient for tracking human hand movement which has inevitably varying speed. In this paper, a systematic way to determine the robot speed is proposed.

Prediction of the Human Movement

In predictive control [95] a reference signal is needed to compute the control actions. In our study, the reference signal is the human hand movement. At instant k, the controller needs to determine the speed $u(k) = \sqrt{u_x^2(k) + u_y^2(k) + u_z^2(k)}$based on the robot tip position feedback $\phi(k) = [x(k), y(k), z(k)]$ to drive the robot to track human hand movement $\phi_r(k) =$

$[x_r(k), y_r(k), z_r(k)]$.

The prediction range N should be large enough to achieve a robust control. However, the regulation speed decreases as N increases. It is found that $N = 5$ can achieve the satisfactory regulation speed and good robustness. In our application, the desired trajectory $\phi_r(k+j)$ is defined as:

$$\begin{cases} u_{i,f}(k+j) = \alpha u_{i,f}(k+j-1) + (1-\alpha)u_i(k+j) \\ \phi_r(k+j) = \phi(k) + u_{i,f}(k) jT_s, j = 1, ..., N \end{cases} \tag{4.3}$$

where $i = x, y, z$, $u_{i,f}$ is the filtered speed in x, y, z axis, T_s is the sampling time, and α is the smoothing coefficient. As α becomes larger, the system will track the set point with slower speed but better robustness and smoothness. To choose an appropriate α, prediction errors are evaluated. Figure 4.8 illustrates 5-step-ahead prediction error in x axis for a sample human movement specified in Figure 4.7, with respect to smoothing coefficient from 0 to 1. It is observed that $\alpha = 0.9$ can achieve a good trade-off between response speed and robustness, and the 5-step-ahead prediction error reaches its minimum when $\alpha = 0.9$.

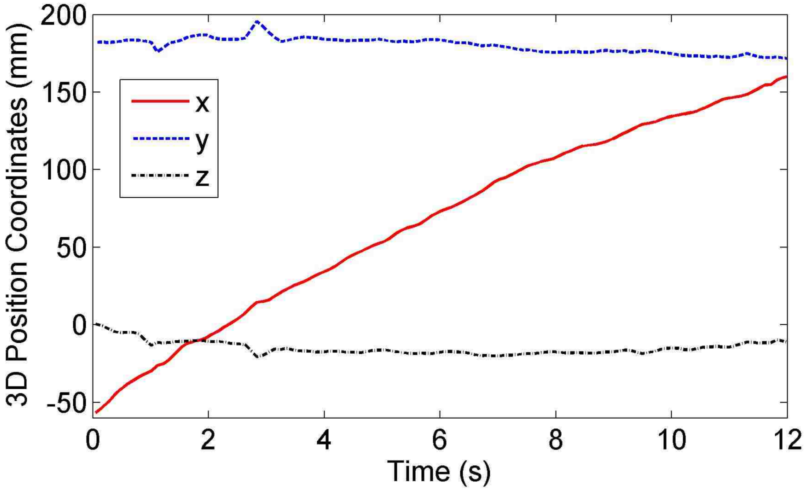


Figure 4.7: Sample human movement.

Because the smoothness of the human hand movement varies from person to person, it is evident that different operators should have different smoothing coefficients. To obtain

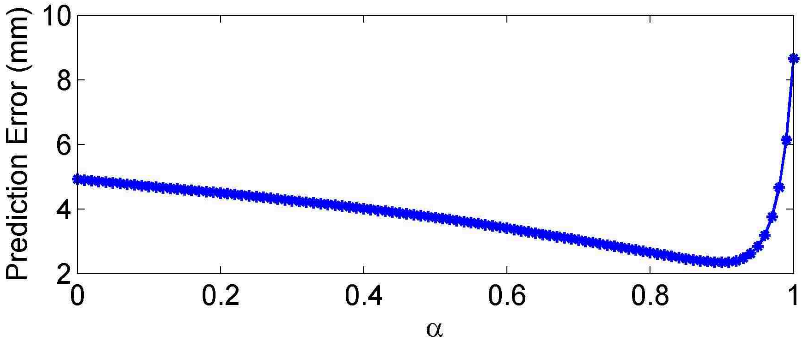


Figure 4.8: 5-step-ahead prediction error versus smoothing coefficient.

the smoothing coefficient for a specific operator, a training period can be conducted and process described in this section can be applied accordingly.

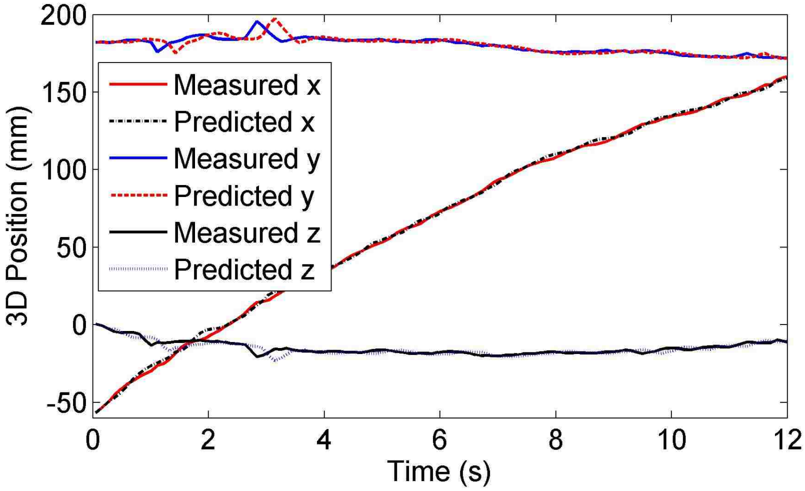


Figure 4.9: 5-step-ahead prediction coordinates.

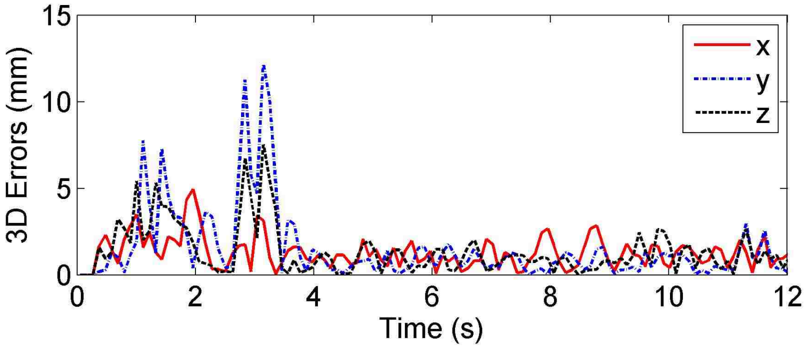


Figure 4.10: 5-step-ahead prediction errors.

Figure 4.9 and Figure 4.10 depicts the 5-step-ahead prediction performance using Equation 4.3. It is observed that the prediction errors remain less than 2 mm for most of the time. For a relatively long range predictive period (5-step or 2.5 s in this study), this prediction result is considered acceptable. In the following subsection, this prediction will be utilized in the predictive control algorithm.

Predictive Control Algorithm

In this section predictive control of robot motion for teleoperation is derived. Given the sampling time T_s, the following 1-step-ahead prediction equations can be obtained:

$$\begin{cases} \hat{x}(k+1) = x(k) + (u_x(k-1) + \Delta u_x(k))T_s \\ \hat{y}(k+1) = y(k) + (u_y(k-1) + \Delta u_y(k))T_s \\ \hat{z}(k+1) = z(k) + (u_z(k-1) + \Delta u_z(k))T_s \end{cases} \tag{4.4}$$

Suppose the future control action is constant, i.e. $\Delta u_x(k+j) = \Delta u_y(k+j) = \Delta u_z(k+j) = 0, j = 1, ..., N$, j-step-ahead prediction yields:

$$\begin{cases} \hat{x}(k+j) = x(k) + (u_x(k-1) + \Delta u_x(k))jT_s \\ \hat{y}(k+j) = y(k) + (u_y(k-1) + \Delta u_y(k))jT_s \\ \hat{z}(k+j) = z(k) + (u_z(k-1) + \Delta u_z(k))jT_s \end{cases} \tag{4.5}$$

The prediction equation can be further expressed in matrix form:

$$\begin{cases} [\hat{X}]_{Nx1} = [X(k)]_{Nx1} + [F_x]_{Nx1} u_x(k-1) + [F_x]_{Nx3} \Delta u_x(k) \\ [\hat{Y}]_{Nx1} = [Y(k)]_{Nx1} + [F_y]_{Nx1} u_y(k-1) + [F_y]_{Nx3} \Delta u_y(k) \\ [\hat{Z}]_{Nx1} = [Z(k)]_{Nx1} + [F_z]_{Nx1} u_z(k-1) + [F_z]_{Nx3} \Delta u_z(k) \end{cases} \tag{4.6}$$

where $\hat{X} = \begin{pmatrix} \hat{x}(k+1) \\ \hat{x}(k+2) \\ \vdots \\ \hat{x}(k+N) \end{pmatrix}$, $X(k) = \begin{pmatrix} x(k) \\ x(k) \\ \vdots \\ x(k) \end{pmatrix}$, $F_x = \begin{pmatrix} T_s & 0 & 0 \\ 2T_s & 0 & 0 \\ \vdots & \vdots & \vdots \\ NT_s & 0 & 0 \end{pmatrix}$.

Or

$$[\hat{\Phi}]_{3Nx1} = [\Phi(k)]_{3Nx1} + [F]_{3Nx3}[U]_{3x1} + [F]_{3Nx3}[\Delta U(k)]_{3x1} \tag{4.7}$$

where $\hat{\Phi} = \begin{pmatrix} \hat{X} \\ \hat{Y} \\ \hat{Z} \end{pmatrix}$, $\Phi(k) = \begin{pmatrix} X(k) \\ Y(k) \\ Z(k) \end{pmatrix}$, $U = \begin{pmatrix} u_x(k-1) \\ u_y(k-1) \\ u_z(k-1) \end{pmatrix}$, and $\Delta U(k) = \begin{pmatrix} \Delta u_x(k) \\ \Delta u_y(k) \\ \Delta u_z(k) \end{pmatrix}$.

The predictive control algorithm can be formulated by optimizing a cost function. In our application the tracking accuracy of the robot arm is important to perform the teleoperation. The following cost function is first proposed:

$$\begin{aligned} &J(\Delta U(k)) = \textstyle\sum_{j=1}^{N}[(x(k+j) - x_r(k+j))^2 + (y(k+j) - y_r(k+j))^2 \\ &+(z(k+j) - z_r(k+j))^2] \end{aligned} \tag{4.8}$$

Large robot movement speed may generate non-smooth robot movement and shaking, which is not preferable in our application. Thus, the control objective should minimize the tracking error as well as the robot speed. The following cost function is used:

$$\begin{aligned} &J(\Delta U(k)) = \textstyle\sum_{j=1}^{N}[(x(k+j) - x_r(k+j))^2 + (y(k+j) - y_r(k+j))^2 \\ &+(z(k+j) - z_r(k+j))^2] + \lambda(u_x^2(k) + u_y^2(k) + u_z^2(k)) \\ &= [\Phi(k) + FU + F\Delta U(k) - \Phi_r(k)]^T[\Phi(k) + FU + F\Delta U(k) - \Phi_r(k)] \\ &+[U + \Delta U(k)]^T\Lambda[U + \Delta U(k)] \end{aligned} \tag{4.9}$$

where $[\Lambda]_{3x3} = diag(\lambda)$. The value of the weight λ can be determined based on their physical meaning. For applications where smooth robot movement is more preferable, λ should be relatively large. While in applications where tracking performance is more important, smaller should be chosen. In this study $\lambda = 1$ $(mm/1mm/s)^2$ is chosen, which implies that an error of $1mm$ in the position has the same contribution to the cost function as robot speed change of $1mm/s$.

The control law is calculated such that:

$$\frac{\partial J(\Delta U(k))}{\partial \Delta U(k)} = 0 \tag{4.10}$$

Equation 4.10 can be further expressed as:

$$\begin{aligned}
&\frac{\partial J(\Delta U(k))}{\partial \Delta U(k)} = \frac{\partial [U+\Delta U(k)]^T \Lambda [U+\Delta U(k)]}{\partial \Delta U(k)} \\
&+\frac{\partial [\Phi(k)+FU+F\Delta U(k)-\Phi_r(k)]^T [\Phi(k)+FU+F\Delta U(k)-\Phi_r(k)]}{\partial \Delta U(k)} \\
&= 2F^T F[\Phi(k)+FU-\Phi_r(k)]+2\Lambda[U+\Delta U(k)] \\
&= 2(F^T F+\Lambda)\Delta U(k)+2F^T[\Phi(k)+FU-\Phi_r(k)] = 0
\end{aligned} \tag{4.11}$$

The predictive control law is finally expressed as:

$$\Delta U(k) = -(F^T F+\Lambda)^{-1}[F^T(\Phi(k)+FU-\Phi_r(k))+\Lambda U] \tag{4.12}$$

The robot movement speed can thus be calculated by $u(k)=\sqrt{(u_x(k)+\Delta u_x(k))^2+(u_y(k)+\Delta u_y(k))^2+(u_z(k)+\Delta u_z(k))^2}$ and sent to the robot together with the 3D coordinates of the next pose.

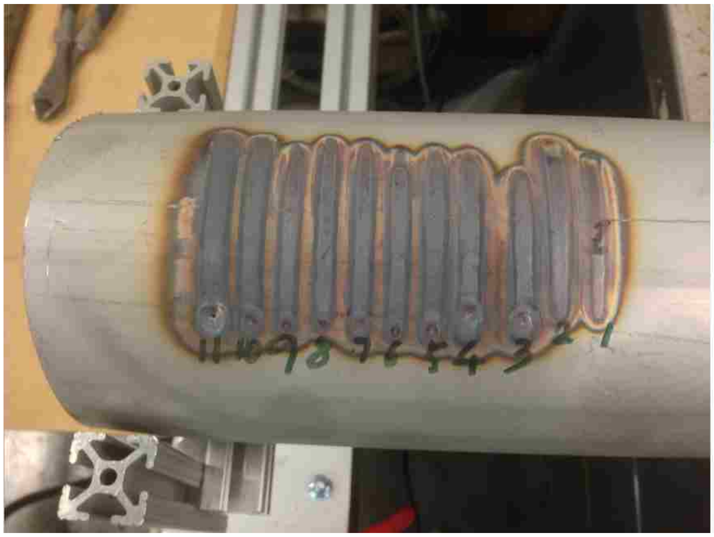
Equation 4.12 is an analytical solution to the optimization of the cost function specified in Equation 4.9, and can thus be implemented in real-time.

4.4 Experiment

In order to demonstrate the performance of our virtualized welding system, we conduct the following experiments. First, the visualization of work piece on HRD is discussed. Second, three different control and tracking experiments are conducted: (a) Control robotic arm with 3D mouse (3DConnexion SpaceMouse Pro 3DX-700040) and visualize video from local view on flat computer screen; (b) 3D mouse control and visualizing on HRD; (c) Control robotic arm with motion sensor while rendering video on HRD.

Visualization on HRD

Each weld path on the work piece is generated by our welding system, with a dimension of 60 80 millimeters in length and about 5 millimeter in width. By measuring the actual dimension of each weld path projected on the HRD, a qualitative assessment can be conducted. 15 weld paths (see Fig. 4.11,there are 4 paths on the other side) were tested, and result can be found in Fig. 4.11(Due to space limit, we only show average result of every 5 paths in a group).



Path Id	Length error (mm)	Width error (mm)
1-5	1.5	0.5
6-10	1.5	0.5
11-15	1.8	0.5

Figure 4.11: weld pathes on the pipe.

The mock up used for HRD is 3D scanned based on real work piece. Our visualization assumes the mock up is geometrically identical to the real work piece, or with a scale factor. However, 3D scanner can introduce potential errors. In Fig. 4.12 , we demonstrates that,when compared with ground truth, the model generated by our 3D scanner has fairly small error.

Robot Tracking Performance

To assess the performance of the human motion tracking and robot control in our system, 5 human welders is tasked to follow two types of trajectory on the HRD: a straight line, and a sine wave (Fig. 4.13).

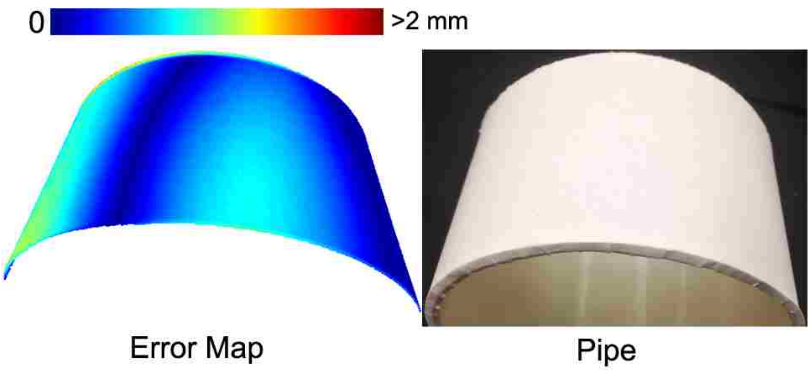


Figure 4.12: Error of 3D scaned model, the difference between reconstructed model and ground truth is less than 1 mm.

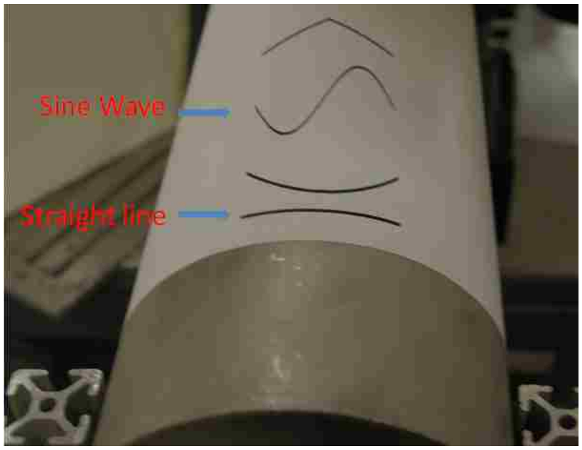


Figure 4.13: Several patterns for tracking experiment. We designed four different type of pattern on the surface: straight line, quadratic curve, sine wave and broken line. Note that when applying these pattern on work piece, all of them turn to 3D curve.

In this subsection, the robot tracking performance is evaluated. It is focus on how accurate the robotic arm when sending it certain coordinate. Figure 4.14 and Figure 4.15 illustrates the performance of the motion sensor tracking a sine wave in 3D space. Figure 4.16 plots the tracking errors. It is seen that the sent 3D coordinates and measured robot coordinates are matched well. The tracking errors are maintained smaller than 2 *mm*, which is considered acceptable in our application. For tracking a straight line and motion tracked by 3D mouse, similar tracking performances are observed.

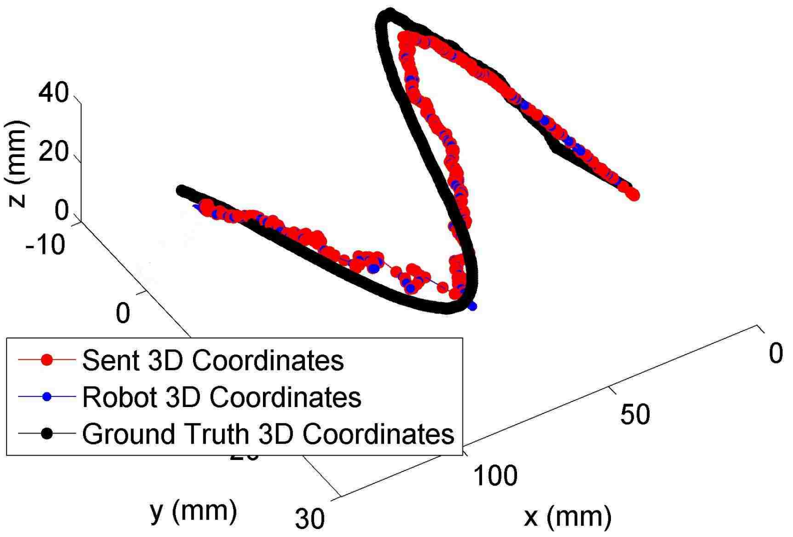


Figure 4.14: Robot tracking performance of a sine wave using Leap Sensor in 3D space.

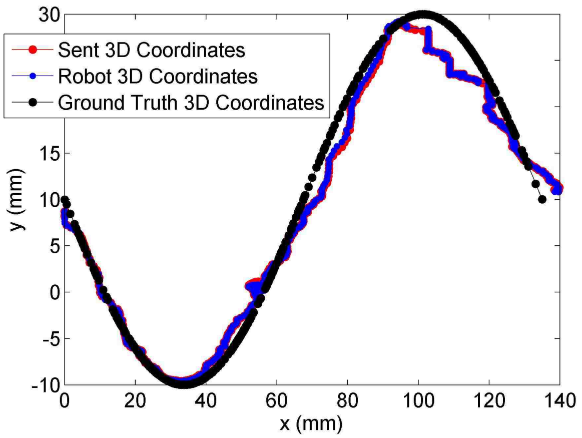


Figure 4.15: X-Y plane projection for Figure 4.14.

Comparison Between 3 Algorithms

In this subsection, we compare the reading of motion sensor and the ground truth 3D coordinate. Since the user adjust the motion of the tool based on the video feedback of local view camera, a good matching proves the overall system accuracy.

The 3 tracking methods are compared and the experiment results are analyzed. Fig. 4.17 describes how these 3 type of experiments are conducted. Figure 4.18 and Figure 4.19 illustrate the tracking performance of a straight line in 3D space. It is shown that the

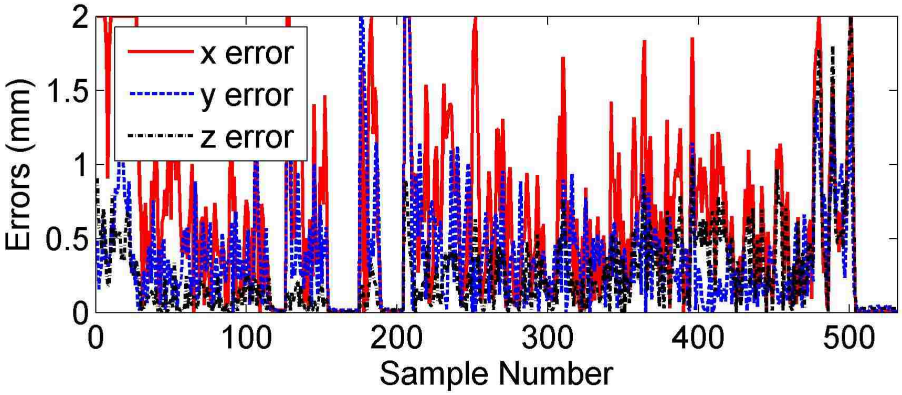


Figure 4.16: Tracking errors for Figure 4.14.

tracking performance compared to ground truth is excellent for 3D mouse tracking, either with 2D screen display or mockup display. For Leap Sensor tracking, the performance is slightly deteriorated by the human hand movement. However, most of the errors are kept within 2 *mm*, which is considered acceptable in our application.

Then a sine wave is tracked by the proposed 3 methods, and the experimental results are plotted in Figure 4.20 and Figure 4.21. It is shown that compared to 3D mouse tracking, Leap Sensor can track this complex trajectory with better accuracy.

To quantitatively evaluate the performances of the proposed 3 tracking methods, the following two criteria are defined. The model average error is;

$$E_{ave} = \frac{1}{n}\sum_{k=1}^{n} |\hat{y}(k) - y(k)|, (k = 1, ..., n) \tag{4.13}$$

where n is the number of data points, $y(k)$ is the ground truth position at instant k, and $\hat{y}(k)$ is the measured robot position. The root mean squares error (RMSE) is calculated by:

$$RMSE = \sqrt{\sum_{k=1}^{n} (\hat{y}(k) - y(k))^2/n} \tag{4.14}$$

Table 4.2 depicts the errors associated with 3 tracking methods. It is seen that for simple tracking task (e.g., a straight line), tracking with 3D mouse either with 2D screen or mockup display performs better than Leap Sensor. This is expected because for such

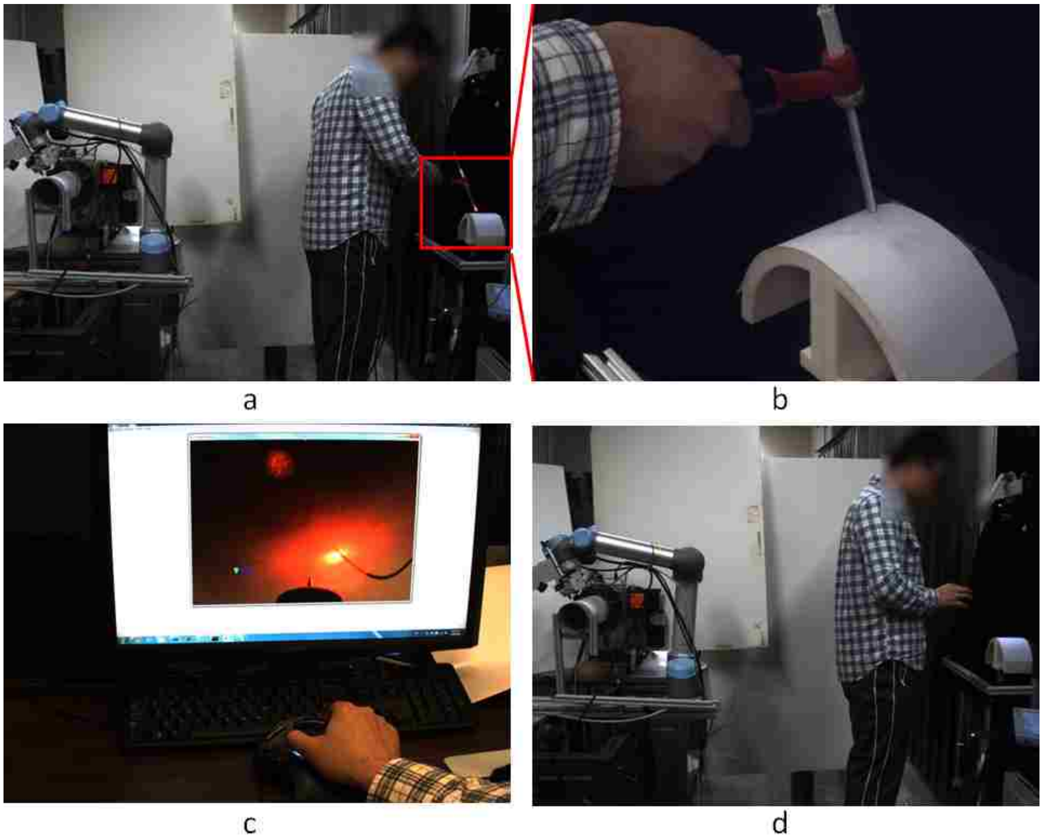


Figure 4.17: Three different control method. (a): user controls the robotic arm via motion sensor while viewing the HRD; (b): close up look of a; (c): user controls the robotic arm via 3D mouse while viewing the flat display device; (d): user controls the robotic arm via 3D mouse whiel viewing the HRD.

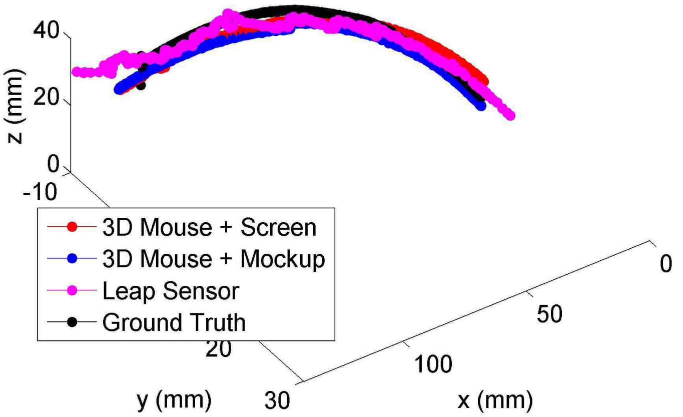


Figure 4.18: Tracking performance: straight line.

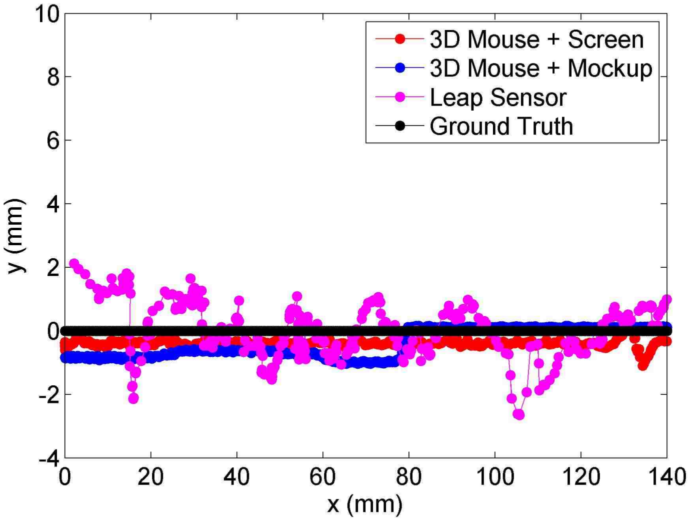


Figure 4.19: X-Y plane projection for Figure 4.18.

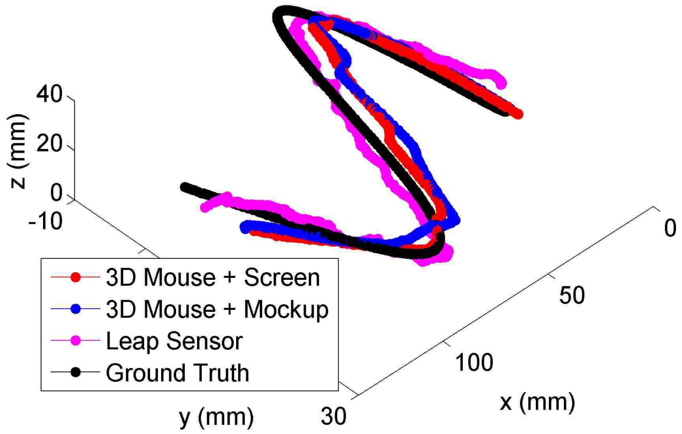


Figure 4.20: Tracking performance: sine wave.

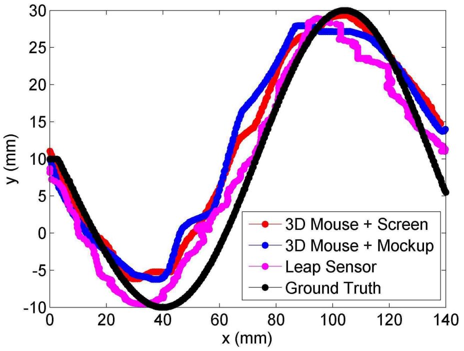


Figure 4.21: X-Y plane projection for Figure 4.20.

type of motion, adjusting 3D mouse in one direction is convenient and robust. For Leap Sensor, however, since human hand motion is intrinsically shaking, the performance might be deteriorated. Yet, for complex 3D tracking (e.g., sine wave in 3D space), leap sensor outperform other two tracking methods since the major advantage of the proposed tracking system is that it has the flexibility like the human hand.

Table 4.2: Error Comparison Between Leap Sensor and 3D Mouse

	Line		Sine Wave	
	RMSE	E_{ave}	RMSE	E_{ave}
3D Mouse + Screen	0.7439	0.7458	4.3342	5.4551
3D Mouse + Mockup	0.4798	0.5889	4.0978	4.9812
Leap Sensor	1.1215	1.3266	2.9106	3.4699

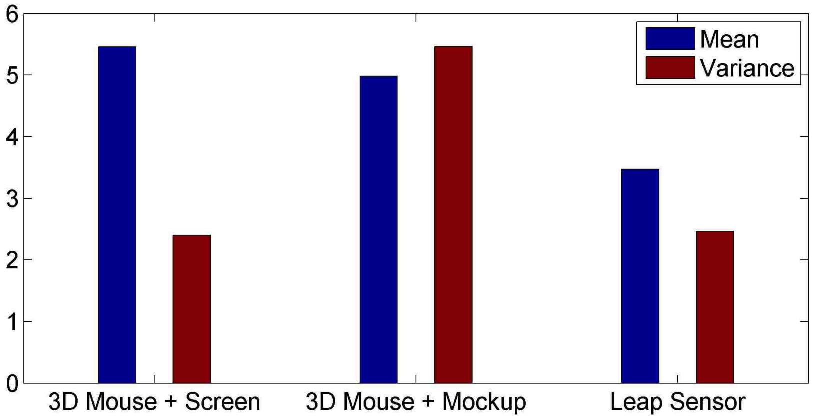


Figure 4.22: Tracking performance comparison between different implementation. The result proves our implementation: 3D motion sensor combined with HRD could handle more complicated senario, 3D curve in our experiment.

More visualization result for our motion control is shown in Fig. 4.23. Even more visualization result is in supplementary material.

4.5 Conclusion

In this chapter, we have proposed a novel mixed reality system for tele-operated welding. A high accuracy 3D scanning technique is utilized to create digital models of work pieces to be welded. Based on the type of welding job, a mock-up is constructed from a set of

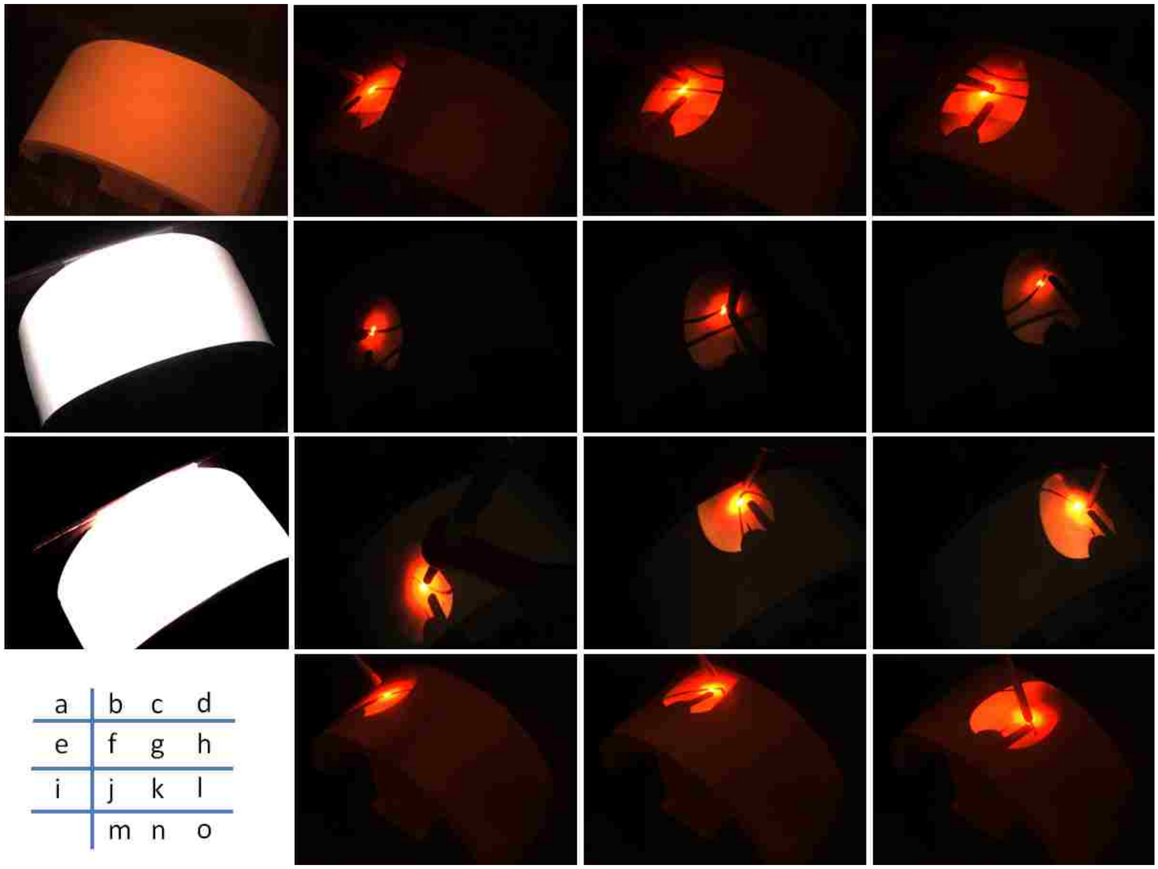


Figure 4.23: Visualization on HRD during control welding. We use a laser dot in real welding workstaion to indicate where the weld gun is pointing at. User should use the laser dot in video feedback rendered on HRD as a spatial and 3D cue. (a,e,i): viewing blank HRD from different angle; (b-d): welding on straight line; (f-h): welding on curve; (j-l): welding on sine wave; (m-o): welding on line segments.

templates. The welding process is captured by cameras mount on the robotic arm and visualized on the mock-up using projectors. The welder can see the welding process as if she/he were next to the actual welding. Experiment shows that our VR display outperforms the conventional 2D tele-operated welding.

Chapter 5 Enhancement of Telepresence

Video has risen to be a dominant force on the Internet. With sufficient computational power and increasing network bandwidth, a new generation of teleconferencing systems is appearing. Typically marketed as telepresence systems, these devices employ carefully designed visual and audio environments that address human factors for the participants. By using large displays, a dedicated network, and high-quality codecs, telepresence systems show life-size participants with accurate flesh tones and fluid motion, simulating the experience of face-to-face meetings.

While commercial telepresence systems - such as those from CISCO, Polycom, or Hewlett-Packard (HP) - offer significant improvement over traditional video teleconferencing systems, one important human factor is still missing: eye contact. As Simmel remarked [132], eye contact ”represents the most perfect reciprocity in the entire field of human relationship.” Because eye gaze is vital in the flow of natural communication, the lack of eye contact is one of the first things participants in video communications notice. The observation is caused by the fact that the display screen and the video camera cannot be co-located or co-linear. In a typical teleconferencing setup (shown in Fig. 5.1), the user often looks at the remote party displayed on the screen, while the local camera often captures the user from above the screen, creating gaze disparity. To preserve eye contact, the ideal camera shall be placed behind the screen.

The correction of eye gaze has been studied for decades. Both software solutions and hardware ones have been introduced. Software methods (e.g., [133–135]) typically use one more or cameras mounted around the screen to capture the user and then apply image warping techniques to create a synthesize view from the ideal position (e.g., behind the screen). This view synthesis problem is very challenging and existing solutions are quite fragile in practice, therefore none has been commercialized. In the hardware track, various tech-

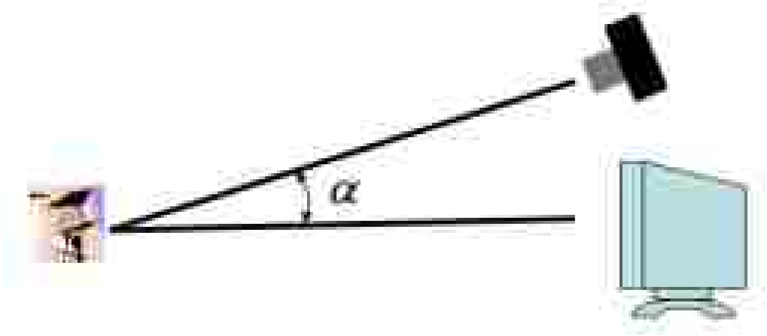


Figure 5.1: Sideview of of a typical teleconferencing system.

niques have been developed to create a see-through screen so that a camera can be placed behind the screen which at the same time serves the main display area. They range from the use of half-silver mirror [136], switchable liquid crystal diffusers [137], anisotropic diffuser [138], to weave fabric [139]. A common problem of these systems is the reduced image quality due to the additional optical components in the camera's imaging path. For example half-silver mirror reduces the amount of incident light by 50%. Therefore the captured images behind these screens are usually under-exposed, noisy, and of poor color fidelity.

In this chapter we present a novel image enhancement framework to significantly improve the image quality captured by a camera from a see-through screen. Rather then performing traditional image enhancement, which are often under constrained, we employ an additional color+depth camera mounted on the side of the screen to make the problem better constrained. A novel sensor fusion algorithm is developed to allow the recovery of a low-noise, high-fidelity image with correct color reproduction and enhanced details.

5.1 Overview

Our desire to put a camera behind a display screen is motivated by the need for maintaining eye gaze during teleconferencing. However, all current see-through screens will significantly reduce the amount of light that can be captured by the camera, because of either the optical design or the need for fast switching. The resulting image therefore exhibits a number of artifacts. The most common ones are high noise level, incorrect color balance, and lack of details (as if seeing through a fog).

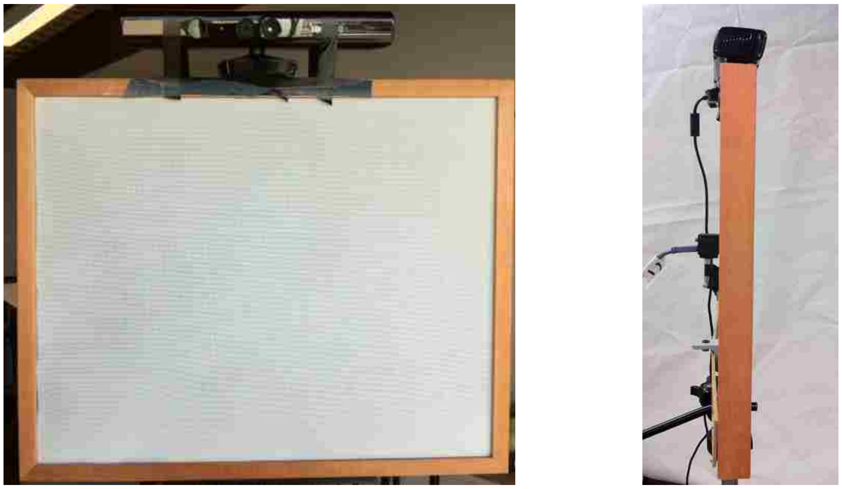
The issue of image or video *denosing* has been an active research topic for decades. Effective approaches include as non-local means [99], bilateral filters [100], etc. Unfortunately, high-frequency details are usually lost after denosing, since it is impossible to distinguish high-frequency contents with random noise in a single image. In order to address this, we incorporate an additional image to provide more constraints for denoising.

The use of additional optical component can also alter the color balance of the captured image. For example, the ConnectBoard system uses an wavelength dependent diffuser to interleave the projected image and see-through image [101]. The color transfer of the diffuser is approximated as a piece-wise linear affine transformation. Color transfer from different images can also be achieved by looking at some image statistics (such as mean and image histogram) [102]. In this paper rather than explicitly modeling the color transfer between two devices/images, we directly warp pixels from the reference view to the see-through view to directly colorize the see-through image.

Our proposed algorithm is related to recent imaging techniques that combine two or more images in the gradient domain (e.g., [103–105]). These algorithm usually deal with a stack of images taken from the same perspective, for which the pixel correspondences across images are accurate and given. In our setup we have two images taken from different perspectives and (effective) illuminations. Our formulation is designed to be robust against erroneous and spare correspondences.

5.2 Our Approach

We assume a hybrid setup which includes two cameras. One is mounted around the edge of the display, which we refer to as the *side-view camera*. In our current setup we choose to use the Kinect camera from Microsoft since it can produce a RGB+depth image. The image from the side view is denoted as I_s. The other is mounted behind a screen, which we refer to as the *eye-view camera*. In order to achieve see-through capability we choose to use the approach that uses a weave fabric for its simplicity and low-cost construction [139]. The image from the eye-view camera is denoted as I_e. Figure 5.2 shows our setup and Figure 5.3 are one sample pair of images from these two cameras. The task is to use the information in I_s to enhance I_e. Our approach consists of two phases, namely guided image warping and denoising. We would explain each step in the following sections.

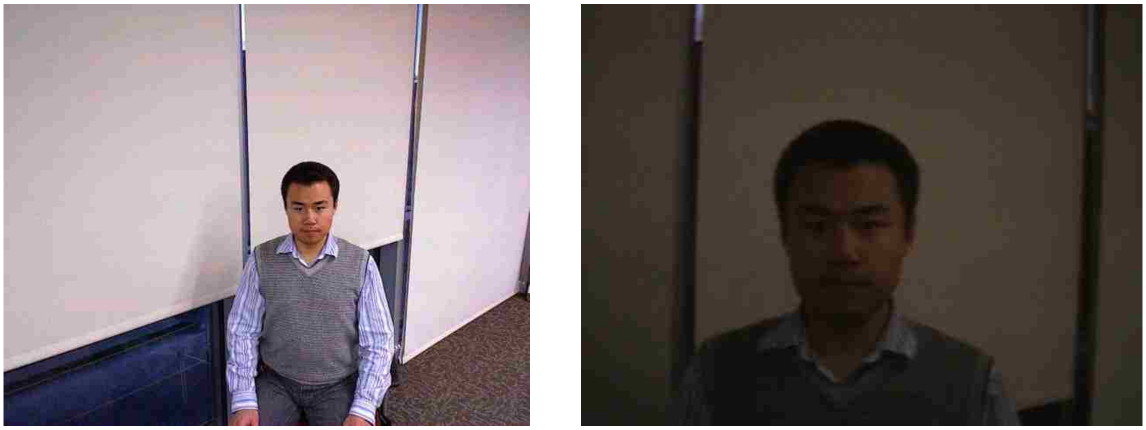


(a) Frontal view (b) Side view

Figure 5.2: Our hybrid setup

Guided Image Warping

The two cameras are pre-calibrated in a common coordinate system. From the calibration information and the depth map contained in I_s, we can back project pixels in I_s to find their corresponding pixels in I_e. This mapping is denoted $f : I_e(u,v) = I_s(p,q)$. Note that f is entirely based on the geometry of the camera and the scene depth map. Due to error in calibration, changes in visibility, and inaccuracy in the depth map, the mapping can be erroneous and sparse, that is, some pixels in I_e have the wrong or even no correspondences in I_s. Therefore simply replacing the pixel values at $I_e(u,v)$ with those from $I_s(p,q)$ will yield pool result. We thereby need a better way to enhance I_e.



(a) Side-view image from Kinect camera

(b) Eye-view image

Figure 5.3: An example of the *side-view image* I_s, *eye-view image* I_e. Note the large difference both in field of view and view angles between I_s and I_e, as well as the extreme low quality of I_e due to under-exposure and noises.

We first check the validity of f by assuming color constancy between the two views. More specifically, we obtain a $N_e \times M_e$ patch of pixels around $I_e(u,v)$ and match it within a slightly larger area of size $N_s \times M_s$, where $N_s > N_e$ and $M_s > M_e$ around $I_s(p,q)$. To deal with the difference in camera gain and exposure, we use normalized cross-correlation (NCC) as the matching metric. If a matching score is above a certain threshold T, the mapping is updated as $I_e(u,v) = I_s(p',q')$ where p',q' is the coordinate in I_s that leads to a high-enough matching score. For the sake of simplicity in notation, we denote the updated mapping as f as well in the following sections. Now we can warp the pixels in I_s according to f to the same space as I_e; and the resulting image is denoted as I_w. Note that I_w is sparse due to pruning process with NCC. Figure 5.4(a)shows the warped image I_w corresponding to images in Figure 5.3.

In order to obtain a dense image from I_w, we first apply Joint Bilateral Filtering (JBF) [140] technique using the *eye-view image* I_e as guidance image. The filtering process can be expressed as

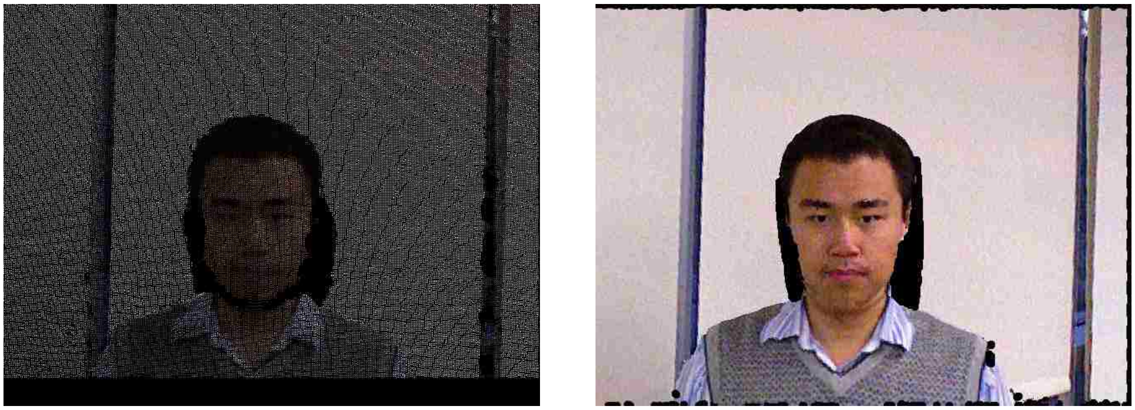
$$I_w^J(p,q) = \frac{1}{W(p,q)} \sum_{(i,j)\in\mathscr{N}_{p,q}} I_w(p,q)w(i,j,p,q); \tag{5.1}$$

where $\mathscr{N}_{p,q}$ denotes the neighborhood of pixel $I_w(p,q)$ and

$$\begin{aligned} w(i,j,p,q) &= e^{\frac{\|I_S(p,q)-I_S(i,j)\|^2}{2\sigma_r^2}} \cdot e^{\frac{(i-p)^2+(j-q)^2}{2\sigma_s^2}}, & (5.2) \\ W(p,q) &= \sum_{(i,j)\in\mathscr{N}_{p,q}} w(i,j,p,q)\delta(I_w(p,q) \neq 0) & (5.3) \end{aligned}$$

where $\delta(x)$ is an indicator function, with value being 1 if and only if x holds. The two parameters σ_r^2 and σ_s^2 controls the shape of the range and spatial gaussian kernel respectively and therefore the degree of filtering. Here we have two alternatives that JBF can be applied to, namely the warped color image I_w and its corresponding sparse depth map. It is well known that filtering will cause blurry effect. Therefore, directly apply JBF on the warped color image I_w would result in lost of details. While the depth map after filtering would also be blurred and accuracy could be compromised, the color image obtained by back-projecting the depth map to *side-view camera* space still preserves the details. There is a trade-off between these two options. We choose to apply JBF on depth map and then obtain color image by back-projection, instead of filtering the warped color image.

Figure 5.4(b) shows the image obtained after applying the above filtering procedure to image I_w in Figure 5.5. As we can see, JBF is capable of filling in small holes. However, due to large view differences between these two cameras, there exist several relatively large regions in I_w that are not visible in I_s, where JBF cannot improve. Since in these regions, no information exists in the *side-view image* I_s, we can only utilize the *eye-view image* I_e as guidance. Obviously simply copy and paste pixel values from I_e to I_w would result in visible artifacts due to large image differences. We address this problem by solving the



(a) Warped image

(b) Filtered image using JBF

Figure 5.4: (a A warped sparse image I_w and (b the image after JBF. Note besides the large holes in (b that are due to view point variation, there are also some small holes after filtering. These are mainly due to inaccuracy of filtered depth map.

following Poisson equation [141]:

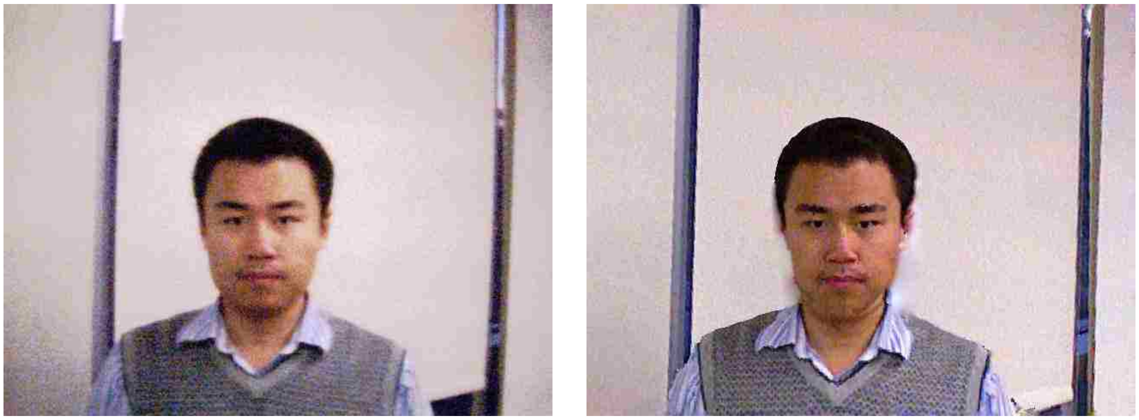
$$I_w^P = argmin_I \sum_{(i,j)\in\mathscr{H}} (\nabla I(i,j) - \mathbf{v}(i,j))^2; \tag{5.4}$$

subject to the boundary conditions

$$I_w^P(p,q) = I_w^J(p,q), \forall (p,q) \notin \mathscr{H}; \tag{5.5}$$

The $\mathscr{H}$ above denotes the hole regions in I_w^J, i.e. after applying JBF. ∇I is the gradient field of image I and $\mathbf{v}$ is the gradient field of a guidance image, which could be I_e. However it is observed that due to the low dynamic range of the *eye-view camera* in our particular setup, the gradient field of the original *eye-view image* I_e does not provide much useful information. We therefore perform histogram matching on I_e with I_w^J to obtain a contrast enhanced image I_e^M. We then set $\mathbf{v} = \nabla I_W^J$ and plug it into Eq. 5.4 to solve for image I_w^p. Figure 5.5 shows the enhanced *eye-view image* and the hole-filled image by solving the Poisson equation.

Clearly, the hole-filled image I_w^P possesses more high-frequency detail than both the



(a) Enhanced eye-view image

(b) Hole-filled image

Figure 5.5: (a Enhanced *eye-view image* through histogram matching; (b Hole-filled image.

original *eye-view image* I_e and the enhanced image I_e^M. Since the image is warped from the *side-view image* I_s, it does not suffer from the low contrast and color washout issues as in I_e. However, the hole-filled image I_w^P still has large amount of visible noises and requires denoising, which is explained in the next section.

Denoising

There has been numerous research papers on image and video denoising. In particular, wavelet-based techniques have been shown to be effective on single image denoising [142–144]. The general wavelet-based denoising procedure select appropriate threshold limit at coefficients at each scale to best remove the noises in image, and then perform inverse wavelet transform of the processed wavelet coefficients to get denoised image. Here we employ Bayes Least-Squares using Gaussian Scale Mixtures (BLS_GSM) algorithm [142] to remove noise from the image I_w^P. Instead of using threshold on coefficients, BLS_GSM removes noise based on a statistical model of the coefficients of an over-complete multi-scale oriented basis.

This technique models the wavelet coefficients, denoted as y, within a local patch in

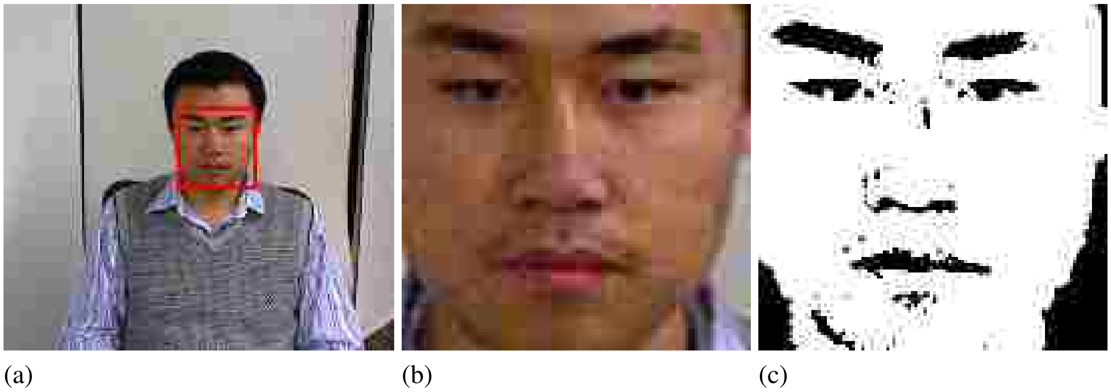

(a) (b) (c)

Figure 5.6: Detected face region and skin colo region, which is used for noise variance estimation.

each scale for an observed image as a Gaussian Scale Mixtures (GSM), as in Eq. 5.6

$$y = x + w = \sqrt{z}u + w \tag{5.6}$$

where z (scale), u (underlying signal) and w (noise) are all zero-mean Gaussians. One assumption made in [142] about the noise behavior is knowledge of noise variance, denoted as σ^2. In real case, the assumption usually does not hold, though pre-calibration can be conducted to estimate the noise variance and use it as an approximation afterwards. However, noise behavior, and thereby the noise variance σ^2, might vary under different lighting conditions and changes of other environmental factors. Due to our specific application, namely video teleconferencing, we can assume existence of human faces in the image. Therefore, we propose to estimate the noise variance σ^2 from face regions in the image. The basic idea is to perform face detection [145] and then skin color detection [146] on the face region to select a set of candidate pixels, which we denote as $\mathscr{S}$. In figure 5.6 we show an example of detected face and skin color region on the face.

The underlying assumption of our approach is uniformity of skin color in human face, which is valid in most cases. Nonetheless, observed skin color in images are not necessarily uniform because of light conditions. Therefore, we use a local method to estimate the noise variance σ^2. A set of noises in small local neighborhood of the pixels in $\mathscr{S}$ are estimated

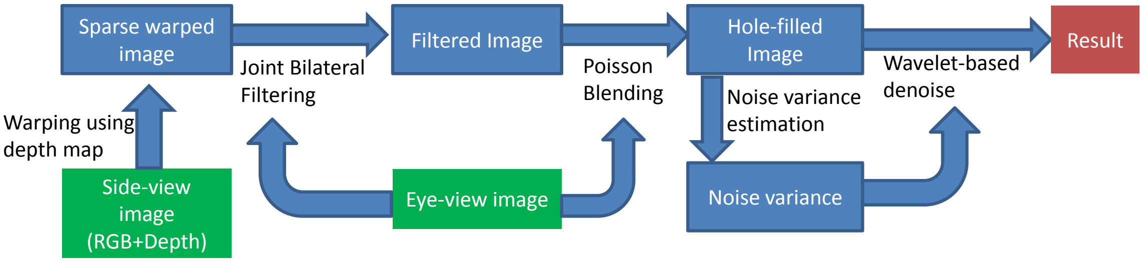


Figure 5.7: Out processing pipeline. The green blocks are our inputs and the red block is our output; while all the rest are our intermediate processing modules

by subtracting local mean:

$$\begin{aligned}\Phi(i,j) = \{ &|I_w^P(p,q) - \frac{1}{|\mathcal{N}_{i,j}|} \sum_{(k,l)\in\mathcal{N}_{i,j}} I_w^P(k,l)|; \\ &\forall(p,q)\in\mathcal{N}_{i,j}\}, \forall(i,j)\in\mathcal{S};\end{aligned} \tag{5.7}$$

Then variance of all these noises are calculated and treated estimation of global noise variance.

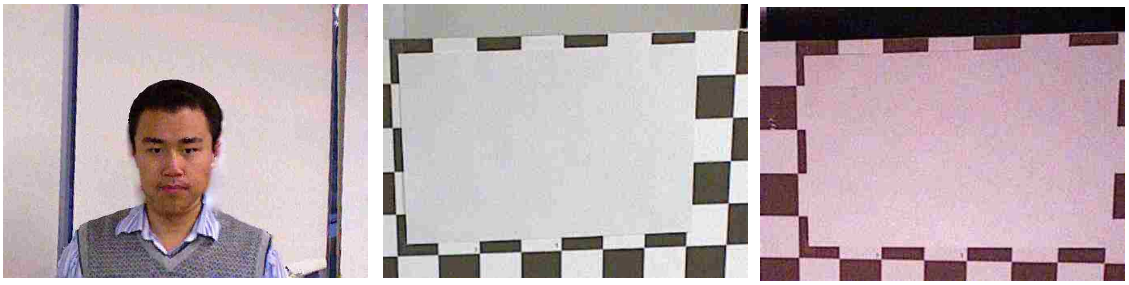
$$\sigma^2 = Var\{\bigcup_{(i,j)\in\mathcal{S}} \Phi(i,j)\}; \tag{5.8}$$

where $Var\{\mathcal{X}\}$ means the variance of all the elements in the set $\mathcal{X}$. We then plug in the estimated σ^2 into the BLS_GSM algorithm for noise reduction. After this step, the final image can be ready for display. The results are shown in section 5.3. Out entire pipeline is summed up in Fig. 5.7.

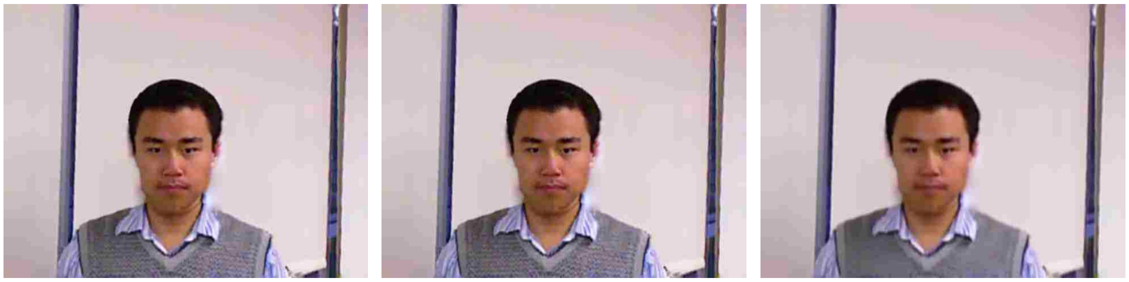
5.3 Experiments

In this section we first demonstrate the advantages of employing each components in our approach and then qualitatively evaluate the performance of our approach. While the effectiveness of JBF and Poisson Blending has been shown in Sec. 5.2, here we show that the noise variance estimation presented in Sec. 5.2 offers more robustness to lighting changes. We use a uniform color board to pre-calculate the noise variance under two lighting conditions, one similar to the one under which we capture the real scene while another is fairly different. Figure 5.8 shows the comparisons. In the case where lighting conditions are similar, pre-calibration is almost equivalent to our skin-color-based estimation, as show in the left and right column. However, with lighting changed, the pre-calibrated noise behavior is usually not consistent with the real scene; and therefore denoising would result in over-smoothed or under-smoothed. Here in the right column of Figure 5.8, we show a case of over-smoothed.

In the third column of Figure 5.9, we show results applying our approach to several real scenes. Compared to the original *eye-view images* in second column, our approach achieves significantly higher image qualities, both in terms of high-frequency details and level of noises. For the sake of comparing overall performance, we also apply bilateral filters [1] and wavelet domain filters [2] with histogram matching on the same set of eye-view images. We keep the parameters for both comparison methods fixed for all these



(a) Source image from which noise variance are estimated. Left: real scene image. Middle: with lighting similar to real scene. Right: with lighting different to real scene.



(b) Denoising results with the noise variances estimated from the images in (a), from left to right respectively.

Figure 5.8: Comparison of noise variance estimations.

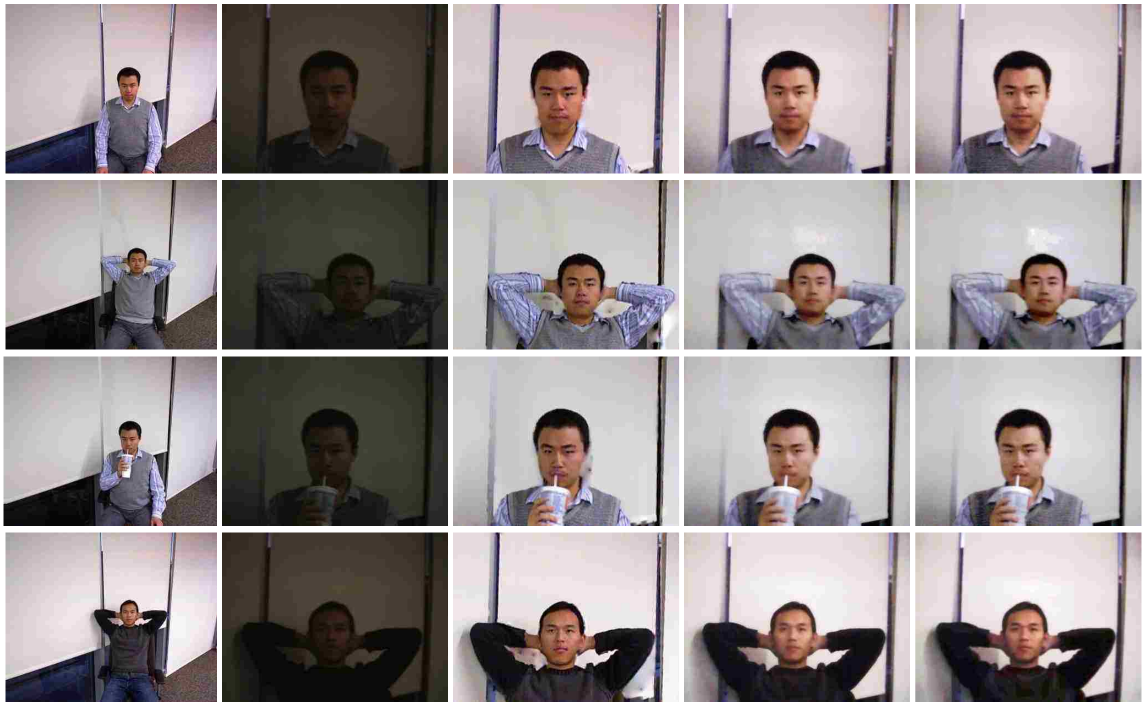


Figure 5.9: Comparison of our result (third column) with bilateral filtering [1](fourth column) and wavelet domain denoising [2] (fifth column). The first and second columns are images from *side-view camera*, i.e. Kinect, and *eye-view camera*, respectively.

experiments. For bilateral filters, we use 5×5 neighborhood and $[3, 0.1]$ as spatial and range kernel variance; while the noise variance σ^2 for wavelet-based methods is the one estimated from the image in the middle column in Figure 5.8. Figure 5.9 shows results of four real cases, with both our results and the others' for comparison. It is clear that our approach preserves more details. By contrast, both bilateral filter and wavelet filters suffer from more severe lost of details.

There are still rooms for improvement in our results. Some of the occluded areas are overly smoothed. Two sources of errors could contribute to these artifacts: misalignment due to inaccuracy of depth map and imperfect inlier selection with NCC. With such small errors, Poisson Blending would propagate the error to some extent. However, since we only apply Poisson Blending on the hole regions due to view-point difference while other small holes are filled using JBF, the artifacts are limited to those small regions.

5.4 Conclusion

In this chapter we present a novel image enhancement method to effectively improve the frame visual quality captured by camera behind a see-through screen. Our framework differs from present image enhancement by adding additional color and depth information captured by Kinect camera. This unique setup makes our algorithm outperform traditional image enhancement method in recovering nature colored image with less noise and more detail information. We develop a novel pipeline that adopt state-of-the-art image warping, filtering, and fusing techniques to enhance the underexposed and blurry see-through image. By comparing our approach with bilateral filter and typical denoising algorithm, we demonstrate our algorithms ability of better preserving detail image information while reducing noise. Looking into the future we plan to use graphics hardware to make the processing in real time. Since almost all of our operations are local, they can be easily accelerated on the GPU. In addition we want to explore ways to directly estimate an image noise model without explicitly correspondences so we can use a regular side-view camera. This might be possible if the baseline between these two camera is small.

Chapter 6 Conclusion and Future Work

In this dissertation, we have explored several aspects in the area of telerobotics. In the tele-operation, we present novel ways by using mixed reality techniques. The proposed algorithms benefit the multi-sensor calibration which is critical to tele-operation. Based on the proposed algorithms, we further developed a platform for monitoring, visualization and remote control of a teleoperational system. We apply our techniques in application such as remote welding and the promising result proves the performance of our algorithms. In the tele-presence, present a novel image enhancement method to effectively improve the frame visual quality captured by camera behind a see-through screen. By comparing our approach with bilateral filter and typical denoising algorithm, we demonstrate our algorithms ability of better preserving detail image information while reducing noise.

6.1 Contribution

Teleoperation with proper visual information assistant is still a challenging problem. Due to wide range of human factors, ergonomics, and usability research that has identified the limitations of traditional planar displays for supporting navigation and teleoperation tasks, I proposed a new type of display: hybrid-reality display (HRD) system [147], which utilizes commodity projection device to project captured video frame onto 3D replica of the actual target surface. It provides a direct alignment between the frame of reference for the human subject and that of the displayed image. The advantage of this approach lies in the fact that no wearing device needed for the users, providing minimal intrusiveness and accommodating users eyes during focusing. The field-of-view is also significantly increased. From a user-centered design standpoint, the HRD is motivated by teleoperation accidents, incidents, and user research in military reconnaissance etc. Teleoperation in these environments is compromised by the Keyhole Effect, which results from the limited field of view of reference. The followed up research of HRD is focused on high accuracy 3D reconstruction of the replica via commodity devices for better alignment of video frame. There are three conventional approaches: Time of Flight (TOF) sensor based, Kinect Fusion Based, and structured light scanning based method. The first two methods suffer from relatively low depth resolution due to the limitation of the depth sensor. The third one can provide sub-millimeter accuracy while sensitive to spectrum nature of object. In [148], I improved the performance of structured light scanning by utilizing a high speed near infrared projector, which is robust to the color of object.

Robot control has been an active research area since early 1980s. Different control methods have been proposed, ranging from passivity, compliance, predictive and adaptive control, and variable structures. Predictive control of linear systems has received considerable attention in past decades due to its robustness with respect to model uncertainty. In [149], I proposed 1-step-ahead predictive control algorithm. The latency between human control and robot movement can be formulated as a linear equation group with a smooth coefficient ranging from 0 to 1. This predictive control algorithm can be further formulated by optimizing a cost function. Another aspect of research regarding to improve the performance of robot control relies on high accuracy inter-devices calibration, which is also a hot topic in robotic vision. I mainly focused on calibration between auxiliary device (mounted camera, etc.) and robot. In [150], I addressed a rigid/non-rigid model completion based on iterative closest point (ICP) and global optimization of error distribution. This method adapts the fact that the alignment error introduced during bounding of each pair of adjacent patches of 3D scan is inevitable. With global optimization, the generated complete model suffers from fewer artifacts.

In the scope of tele-presence, all current see-through screens will significantly reduce the amount of light that can be captured by the camera, because of either the optical design or the need for fast switching. The resulting image therefore exhibits a number of artifacts. The most common ones are high noise level, incorrect color balance, and lack of details (as if seeing through a fog). In this paper rather than explicitly modeling the color transfer between two devices/images, we directly warp pixels from the reference view to the see-through view to directly colorize the see-through image. Our proposed algorithm is related to recent imaging techniques that combine two or more images in the gradient domain. These algorithm usually deal with a stack of images taken from the same perspective, for which the pixel correspondences across images are accurate and given. In our setup we have two images taken from different perspectives and (effective) illuminations. Our formulation is designed to be robust against erroneous and spare correspondences

6.2 Future Work

Shard the point with science fiction masterwork, the ultimate goal of robot is highly intelligent, independent device with the merit of robustness and accuracy. The potential research areas which also interested me are 1) automation: parameterize the rich set of human control data via data-driven approach. The ultimate goal is to enable the development of future generation of multi-task robot that can sense and adapt to different jobs with little or no human intervention. 2) Robust sensing: sensor is the main environment information interface of robot. The accuracy and robustness of sensor and related algorithm greatly affect the performance of task driven robot. The perception of the scene in three dimensions, especially with existence of object with specular highlight is still an open question. Thus, my next research topic is real-time reconstruction of 3D model of object with specular highlight.

In the scope of tele-presence, looking into the future we plan to use graphics hardware to make the processing in real time. Since almost all of our operations are local, they can be easily accelerated on the GPU. In addition we want to explore ways to directly estimate an image noise model without explicitly correspondences so we can use a regular side-view camera. This might be possible if the baseline between these two camera is small.

Chapter 7 Appendix

Due to the complexity of the experiment environment setup, this chapter statements the general procedure of system calibration for the remote welding in Chapter 3 and Chapter 4.

7.1 Camera Calibration

In our experiment, there are two video cameras involved. One is mounted on the robotic arm (denoted as Cam_a) and another is the auxiliary camera paired with the projector in virtual workstation (denoted as Cam_b).

Camera Calibration of Cam_a

Due to the nature of welding process, the distance between the tool holding by the robotic arm and the work piece should be less than 5 mm. Consequently, the Cam_a, in order to gain relatively high resolution of region of interest (ROI) while keep the FOV for environment awareness, should be mounted with a wide FOV lens while keep the camera as close as to the work piece. The wide FOV lens, inevitably, could introduce barrel shape optical distortion. An matured solution for undistortion is by using the Matlab based camera self-calibration toolbox [125]. Once the parameter of distortion is calculated by the toolbox, there are plenty of real-time image undistortion toolboxes available. The experiment data can be viewed in Fig. 7.1.

Camera Calibration of Cam_b

The calibration of the auxiliary camera Cam_b is essentially following the calibration of structured light scanner, which is a joint calibration of camera and the paired projector. The procedure can be refereed to [116].

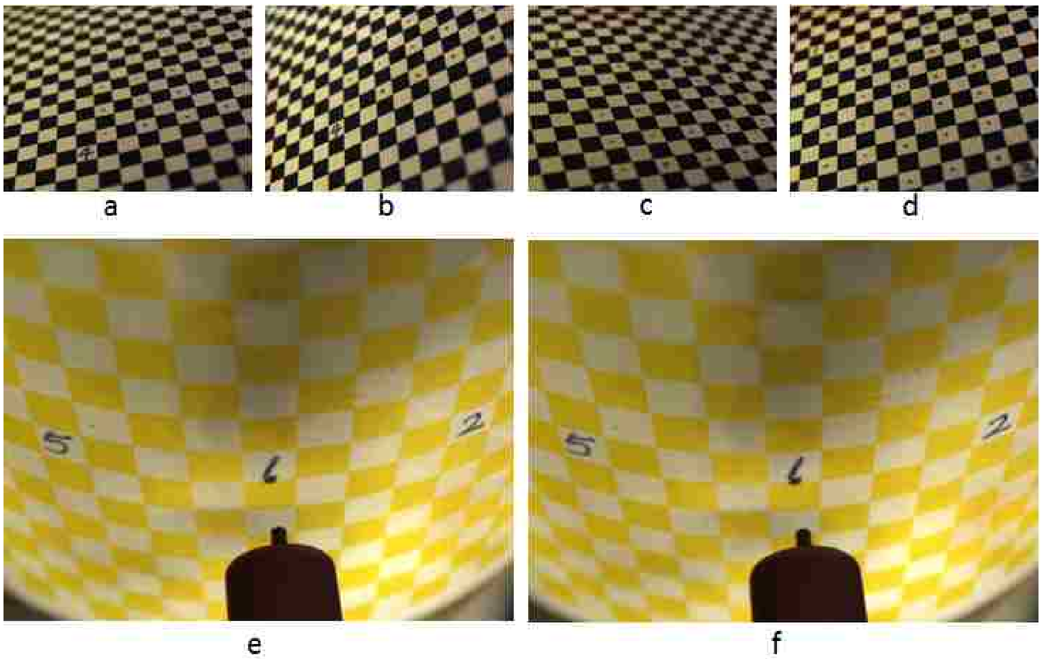


Figure 7.1: Camera undistortion. a)-d) sample input images for undistortion toolbox; e) camera view before undistortion; f) camera view after undistortion.

7.2 Motion Sensor Calibration

The motion sensor applied in our experiment setup is primarily for hand detection and hand motion tracking. Consequently, only the 3D coordinate of tip of bar-shaped object and its 3D orientation can be extracted from the sensor's IO interface. In order to calibrate the motion sensor with the mock up, the conventional method failed since the sensor itself is not a fully functional 3D scanner. Inspired by the concept of triangulation, we introduced a unique way to fulfill the requirement of calibration. The concept is illustrated in Fig. 7.2.

Step1. randomly assign and mark a few dots on the surface of the mock up.

Step2. record the 3D coordinate of the tip location and 3D pose of the bar when it is pointing at certain marker from at least 3 different directions. Based on triangulation, calculate the 3D coordinate of the marker in the coordinate system of motion sensor.

Step3. Based on the 3D coordinate of the markers, employ RANSAC to assist the calculation of the transform matrix between the coordinate systems defined on the mock up and the motion sensor.

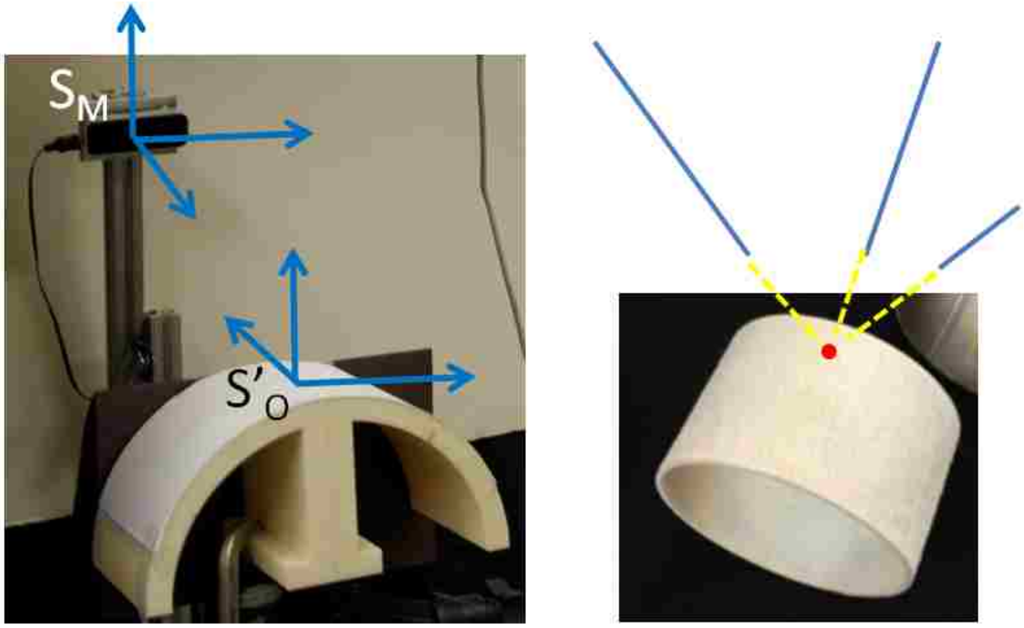


Figure 7.2: Conept of motion sensor calibration. Left, coordinate system defined on motion sensor and mock up; Right, procedure of triangulation.

7.3 Work piece-projector Calibration

Finally, the work piece and projector should be calibrated such that the video feedback from the Cam_a could be properly rendered and projected onto the surface of the corresponding mock up. Based on the motion trajectory of the tool holding by the robotic arm and the actual effective rendering area of the mock up, we can pre-define several control point on the work piece. By pointing the tool towards the control points, several images of the surface of work piece with chessboard pattern will be captured via the Cam_a. Based on the calibration in Section 3.2, the 3D model of the mock up with texture in the coordinate system of projector can be calculated. The lookup table based calibration can therefore be processed. Each image (2D) can be registered with certain portion of the textured 3D model (3D) via 2D to 3D projection. The missing data between adjacent control point will be interpolated. The concept is illustrated in Fig. 7.3.

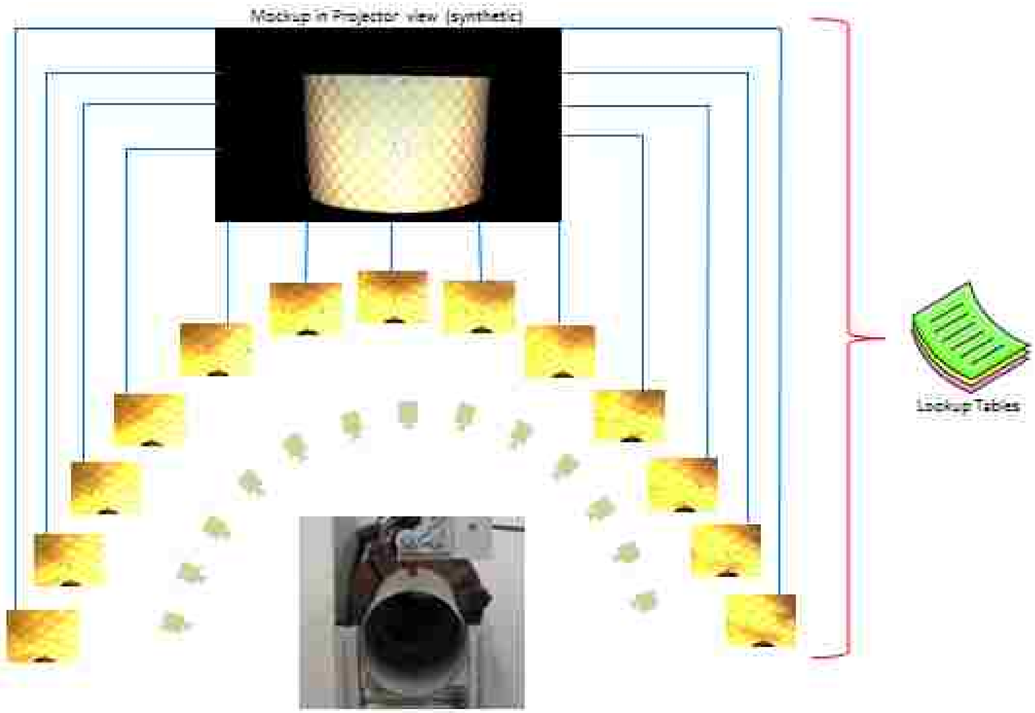


Figure 7.3: Conept of lookup table based calibration. At each control point, a 2D-3D projection is calculated based on cooresponding feature points. The interpolation is introduced between adjacent control points.

Bibliography

[1] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.

[2] Ivan W Selesnick and Ke Yong Li. Video denoising using 2d and 3d dual-tree complex wavelet transforms. In *Optical Science and Technology, SPIE's 48th Annual Meeting*, pages 607–618. International Society for Optics and Photonics, 2003.

[3] National robotics initiative report. `http://www.nsf.gov/pubs/2014/nsf14500/nsf14500.htm`, 2014.

[4] Takeo Kanade, Peter Rander, and PJ Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE multimedia*, 4(1):34–47, 1997.

[5] Peter M Will and Keith S Pennington. Grid coding: A preprocessing technique for robot and machine vision. *Artificial Intelligence*, 2(3):319–329, 1972.

[6] Michihiko MIMOU, Takeo Kanade, and Toshiyuki SAKAI. A method of time-coded parallel planes of light for depth measurement. *IEICE TRANSACTIONS (1976-1990)*, 64(8):521–528, 1981.

[7] JL Posdamer and MD Altschuler. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1):1–17, 1982.

[8] Eli Horn and Nahum Kiryati. Toward optimal structured light patterns. *Image and Vision Computing*, 17(2):87–97, 1999.

[9] Song Zhang and Peisen S Huang. High-resolution, real-time three-dimensional shape measurement. *Optical Engineering*, 45(12):123601–123601, 2006.

[10] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. *Pattern recognition*, 43(8):2666–2680, 2010.

[11] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM transactions on graphics (TOG)*, 24(3):536–543, 2005.

[12] Daniel G Aliaga and Yi Xu. Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3d modeling. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[13] Shree K Nayar, Katsushi Ikeuchi, and Takeo Kanade. Shape from interreflections. *International Journal of Computer Vision*, 6(3):173–195, 1991.

[14] Siying Liu, Tian-Tsong Ng, and Yasuyuki Matsushita. Shape from second-bounce of light transport. In *Computer Vision–ECCV 2010*, pages 280–293. Springer, 2010.

[15] Mohit Gupta, Yuandong Tian, Srinivasa G Narasimhan, and Li Zhang. (de) focusing on global light transport for active scene recovery. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2969–2976. IEEE, 2009.

[16] Li Zhang and Shree Nayar. Projection defocus analysis for scene capture and image display. *ACM Transactions on Graphics (TOG)*, 25(3):907–915, 2006.

[17] Manmohan Krishna Chandraker, Fredrik Kahl, and David J Kriegman. Reflections on the generalized bas-relief ambiguity. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 788–795. IEEE, 2005.

[18] Michael Holroyd, Jason Lawrence, and Todd Zickler. A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance. *ACM Transactions on Graphics (TOG)*, 29(4):99, 2010.

[19] Johnny Park and Avinash C Kak. 3d modeling of optically challenging objects. *Visualization and Computer Graphics, IEEE Transactions on*, 14(2):246–262, 2008.

[20] John Park and Avinash C Kak. Multi-peak range imaging for accurate 3d reconstruction of specular objects. In *6th Asian Conference on Computer Vision*, pages 1–6, 2004.

[21] Chris Hermans, Yannick Francken, Tom Cuypers, and Philippe Bekaert. Depth from sliding projections. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1865–1872. IEEE, 2009.

[22] Shree K Nayar, Gurunandan Krishnan, Michael D Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics (TOG)*, 25(3):935–944, 2006.

[23] Tongbo Chen, Hendrik Lensch, Christian Fuchs, and H-P Seidel. Polarization and phase-shifting for 3d scanning of translucent objects. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[24] Tongbo Chen, H-P Seidel, and Hendrik Lensch. Modulated phase-shifting for 3d scanning. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[25] Jinwei Gu, Toshihiro Kobayashi, Mohit Gupta, and Shree K Nayar. Multiplexed illumination for scene recovery in the presence of global illumination. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 691–698. IEEE, 2011.

[26] Vincent Couture, Nicolas Martin, and Sebastien Roy. Unstructured light scanning to overcome interreflections. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1895–1902. IEEE, 2011.

[27] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G Narasimhan. A practical approach to 3d scanning in the presence of interreflections, subsurface scattering and defocus. *International journal of computer vision*, 102(1-3):33–55, 2013.

[28] Bradley Atcheson, Ivo Ihrke, Wolfgang Heidrich, Art Tevs, Derek Bradley, Marcus Magnor, and Hans-Peter Seidel. Time-resolved 3d capture of non-stationary gas flows. In *ACM transactions on graphics (TOG)*, volume 27, page 132. ACM, 2008.

[29] Jinwei Gu, Shree K Nayar, Eitan Grinspun, Peter N Belhumeur, and Ravi Ramamoorthi. Compressive structured light for recovering inhomogeneous participating media. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(3):1–1, 2013.

[30] Kiriakos N Kutulakos and Eron Steger. A theory of refractive and specular 3d shape by light-path triangulation. *International Journal of Computer Vision*, 76(1):13–29, 2008.

[31] Nigel JW Morris and Kiriakos N Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[32] Ivo Ihrke, Kiriakos N Kutulakos, Hendrik PA Lensch, Marcus Magnor, and Wolfgang Heidrich. State of the art in transparent and specular object reconstruction. In *EUROGRAPHICS 2008 STAR–STATE OF THE ART REPORT*. Citeseer, 2008.

[33] Srinivasa G Narasimhan, Shree K Nayar, Bo Sun, and Sanjeev J Koppal. Structured light in scattering media. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 420–427. IEEE, 2005.

[34] Mohit Gupta, Srinivasa G Narasimhan, and Yoav Y Schechner. On controlling light transport in poor visibility environments. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[35] S Boverie, M Devy, and F Lerasle. 3d perception for new airbag generations . In *15th IFAC World Congress on Automatic Control, Barcelona, Spain*, 2002.

[36] S Boverie, M Devy, and F Lerasle. Comparison of structured light and stereovision sensors for new airbag generations. *Control Engineering Practice*, 11(12):1413–1421, 2003.

[37] Olivier Romain, T Ea, Claude Gastaud, and Patrick Garda. Un capteur multispectral de vision panoramique 3d. *ORASIS*, 1:359–366, 2001.

[38] Mark Alan Livingston. *Vision-based tracking with dynamic structured light for video see-through augmented reality*. PhD thesis, The University of North Carolina at Chapel Hill, 1998.

[39] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM, 1998.

[40] Andrew B Watson et al. Temporal sensitivity. *Handbook of perception and human performance*, 1:6–1, 1986.

[41] David Fofi, Tadeusz Sliwa, and Yvon Voisin. A comparative survey on invisible structured light. In *Electronic Imaging 2004*, pages 90–98. International Society for Optics and Photonics, 2004.

[42] Fumio Kishino Paul Milgram. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, E77-D(12):1321–1329, 1994.

[43] Ivan E Sutherland. The ultimate display. *Multimedia: From Wagner to virtual reality*, 1965.

[44] Kenneth Fast, Timothy Gifford, and Robert Yancey. Virtual training for welding. In *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*, pages 298–299, 2004.

[45] Claude Choquet. Arc+: Todays virtual reality solution for welders. In *International Conference of Safety and Reliability of Welded Components in Energy and Processing Industry*, 2008.

[46] Fronius International. Fronius virtual welding. `http://www.fronius.com/cps/rde/xchg/SID-A0A61DDC-33C5ACA6/fronius_international/hs.xsl/79_15490_ENG_HTML.htm`.

[47] The Lincoln Electric Company. Vrex 360-virtual reality arc welding (vraw) training trainer. `http://www.Lincolnelectric.com/en-us/equipment/training-equipment/Pages/vrtex360.aspx`.

[48] EWI. Advancedtrainer. `http://www.ewi.org/ewi-advancetrainer%E2%84%A2-innovation-in-welder-training`.

[49] Ronald T Azuma et al. A survey of augmented reality. *Presence*, 6(4):355–385, 1997.

[50] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 193–202. IEEE Computer Society, 2008.

[51] DWF Van Krevelen and R Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2):1, 2010.

[52] Alex Olwal, Christoffer Lindfors, Jonny Gustafsson, Torsten Kjellberg, and Lars Mattsson. Astor: An autostereoscopic optical see-through augmented reality system. In *Mixed and Augmented Reality, 2005. Proceedings. Fourth IEEE and ACM International Symposium on*, pages 24–27. IEEE, 2005.

[53] Bernd Hillers, Dorin Aiteanu, and Axel Gräser. Augmented realityhelmet for the manual welding process. In *Virtual and Augmented Reality Applications in Manufacturing*, pages 361–381. Springer, 2004.

[54] Kurtis P Keller, Henry Fuchs, et al. Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 28–31. IEEE Computer Society, 2005.

[55] Ozan Cakmakci, Yonggang Ha, and Jannick P Rolland. A compact optical see-through head-worn display with occlusion support. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 16–25. IEEE Computer Society, 2004.

[56] Hong Hua, Chunyu Gao, Leonard D Brown, Narendra Ahuja, and Jannick P Rolland. Using a head-mounted projective display in interactive augmented environments. In *Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on*, pages 217–223. IEEE, 2001.

[57] Ramesh Raskar, Greg Welch, Kok-Lim Low, and Deepak Bandyopadhyay. Shader lamps: Animating real objects with image-based illumination. In *Rendering Techniques 2001*, pages 89–102. Springer, 2001.

[58] DWF Van Krevelen and R Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2):1, 2010.

[59] Klaus H Strobl and Gerd Hirzinger. Optimal hand-eye calibration. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 4647–4653, 2006.

[60] Yiu Cheung Shiu and Shaheen Ahmad. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax= xb. *Robotics and Automation, IEEE Transactions on*, 5(1):16–29, 1989.

[61] Roger Y Tsai and Reimar K Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *Robotics and Automation, IEEE Transactions on*, 5(3):345–358, 1989.

[62] Hanqi Zhuang, Zvi S Roth, Yiu Cheung Shiu, and Shaheen Ahmad. Comments on" calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax= xb"[with reply]. *Robotics and Automation, IEEE Transactions on*, 7(6):877–878, 1991.

[63] Jack CK Chou and M Kamel. Finding the position and orientation of a sensor on a robot manipulator using quaternions. *The international journal of robotics research*, 10(3):240–254, 1991.

[64] Homer H Chen. A screw motion approach to uniqueness analysis of head-eye geometry. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 145–151. IEEE, 1991.

[65] C-C Wang. Extrinsic calibration of a vision sensor mounted on a robot. *Robotics and Automation, IEEE Transactions on*, 8(2):161–175, 1992.

[66] Hanqi Zuang and Yiu Cheung Shiu. A noise-tolerant algorithm for robotic hand-eye calibration with or without sensor orientation measurement. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(4):1168–1175, 1993.

[67] Irene Fassi and Giovanni Legnani. Hand to sensor calibration: A geometrical interpretation of the matrix equation ax= xb. *Journal of Robotic Systems*, 22(9):497–506, 2005.

[68] Frank C Park and Bryan J Martin. Robot sensor calibration: solving ax= xb on the euclidean group. *IEEE Transactions on Robotics and Automation (Institute of Electrical and Electronics Engineers);(United States)*, 10(5), 1994.

[69] Ying-Cherng Lu and Jack CK Chou. Eight-space quaternion approach for robotic hand-eye calibration. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, volume 4, pages 3316–3321. IEEE, 1995.

[70] Radu Horaud and Fadi Dornaika. Hand-eye calibration. *The international journal of robotics research*, 14(3):195–210, 1995.

[71] Guo-Qing Wei, Klaus Arbter, and Gerd Hirzinger. Active self-calibration of robotic eyes and hand-eye relationships with model identification. *Robotics and Automation, IEEE Transactions on*, 14(1):158–166, 1998.

[72] Konstantinos Daniilidis. Hand-eye calibration using dual quaternions. *The International Journal of Robotics Research*, 18(3):286–298, 1999.

[73] Eduardo Bayro-Corrochano, Kostas Daniilidis, and Gerald Sommer. Motor algebra for 3d kinematics: The case of the hand-eye calibration. *Journal of Mathematical Imaging and Vision*, 13(2):79–100, 2000.

[74] Nicolas Andreff, Radu Horaud, and Bernard Espiau. Robot hand-eye calibration using structure-from-motion. *The International Journal of Robotics Research*, 20(3):228–248, 2001.

[75] Hanqi Zhuang, Zvi S Roth, and Raghavan Sudhakar. Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form ax= yb. *Robotics and Automation, IEEE Transactions on*, 10(4):549–554, 1994.

[76] Sandrine Remy, Michel Dhome, J-M Lavest, and Nadine Daucher. Hand-eye calibration. In *Intelligent Robots and Systems, 1997. IROS'97., Proceedings of the 1997 IEEE/RSJ International Conference on*, volume 2, pages 1057–1065. IEEE, 1997.

[77] Fadi Dornaika and Radu Horaud. Simultaneous robot-world and hand-eye calibration. *Robotics and Automation, IEEE Transactions on*, 14(4):617–622, 1998.

[78] Jessie YC Chen, Ellen C Haas, and Michael J Barnes. Human performance issues and user interface design for teleoperated robots. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1231–1245, 2007.

[79] Martina I Klein, Cindy H Lio, Russel Grant, C Meldoy Carswell, and Stephen Strup. A mental workload study on the 2d and 3d viewing conditions of the da vinci surgical robot. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 53, pages 1186–1190. SAGE Publications, 2009.

[80] Jennifer Casper and Robin R. Murphy. Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 33(3):367–385, 2003.

[81] Stephen R Ellis. Collision in space. *Ergonomics in design: the magazine of human factors applications*, 8(1):4–9, 1999.

[82] Martin Voshell, David D Woods, and Flip Phillips. Overcoming the keyhole in human-robot coordination: simulation and evaluation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 442–446. SAGE Publications, 2005.

[83] Christopher D Wickens, John W Keller, and Ronald L Small. Left. no, right! development of the frame of reference transformation tool (fort). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 54, pages 1022–1026. SAGE Publications, 2010.

[84] T.B. Sheridan. Space teleoperation through time delay: review and prognosis. *IEEE Transactions on Robotics and Automation*, 9(5):592–606, 1993.

[85] G. Niemeyer and J.E. Slotine. Internet-based teleoperation using wave variables with prediction. *IEEE/ASME Transactions on Mechatronics*, 7(2):124–133, 2002.

[86] P. Arcara and C. Melchiorri. Control schemes for teleoperation with time delay: a comparative study. *Robotics and Autonomous Systems*, 38:49–64, 2002.

[87] S. Munir and W.J. Book. Internet-based teleoperation using wave variables with prediction. *IEEE/ASME Transactions on Mechatronics*, 7(2):124–133, 2002.

[88] C. Garcia, J. Posto, and C. Soria. Supervisory control for a telerobotics system: a hybrid control approach. *Control Engineering Practice*, 11(7):805–817, 2003.

[89] H.K. Lee and M.J. Chung. Adaptive controller of a master-slave system for transparent teleoperation. *Journal of Robotic Systems*, 15(8):465–475, 1998.

[90] W. H. Zhu and S.E. Salcudeam. Stability guaranteed teleoperation: an apdative motion / force control approach. *IEEE Transactions on Automatic Control*, 45(11):1951–1969, 2000.

[91] N. Hung, T. Narikiyo, and H. Tuan. Nonlinear adaptive control of master-slave system in teleoperation. *Control Engineering Practice*, 11:1–10, 2003.

[92] S.Y. Yi and M.J. Chung. Robustness of fuzzy logic control for an uncertain dynamic system. *IEEE Transactions on Fuzzy System*, 6(2):216–225, 1998.

[93] F. Cuesta, A. Ollero, B.C. Arrue, and R. Braunstingl. Intelligent control of nonholonomic mobile robots with fuzzy perception. *Fuzzy Sets and Systems*, 134:47–64, 2003.

[94] S. J. Qin and T.A.Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, 11(7), 2003.

[95] Y.K. Liu and Y.M. Zhang. Model-based predictive control of weld penetration in gas tungsten arc welding. *IEEE Transactions on Control Systems Technology*, PP:1–12, 2013.

[96] Y.K. Liu and Y.M. Zhang. Control of 3d weld pool surface. *Control Engineering Practice*, 21(11), 2013.

[97] M. Makarov, M. Grossard, P. Rodriguez-Ayerbe, and D. Dumur. Generalized predictive control of an anthropomorphic robot arm for trajectory tracking. In *2011 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, July 2011.

[98] L. Wang, C.T. Freeman, S. Chai, and E. Rogers. Experimentally validated repetitive-predictive control of a robot arm with constraints. In *2012 American Control Conference*, June 2012.

[99] A. Buades, B. Coll, and J.-M. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, 2008.

[100] M. Zhang and B. Gunturk. Multiresolution bilateral filtering for image denoising. *IEEE Trans. on Image Processing*, 17(12):2324–2333, 2008.

[101] K. Tan, I. Robinson, R. Samadani, B. Lee, D. Gelb, A. Vorbau, B. Culbertson, and J. Apostolopoulos. Connectboard: a remote collaboration system that supports gaze-aware interaction and sharing. In *MMSP*, 2009.

[102] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, pages 34–41, 2001.

[103] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249256, 2002.

[104] P. Perez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Trans. Graph.*, 22:313–318, August 2003.

[105] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph.*, 23:664–672, August 2004.

[106] A roadmap for us robotic: From internet to robotics. `http://www.us-robotics.us/reports/CCC%20Report.pdf`, May 2009.

[107] Natasha Merat, A Hamish Jamson, Frank CH Lai, and Oliver Carsten. Highly automated driving, secondary task performance, and driver state. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(5):762–771, 2012.

[108] Raja Parasuraman, Mustapha Mouloua, and Robert Molloy. Effects of adaptive task allocation on monitoring of automated systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(4):665–679, 1996.

[109] PA Hancock. In search of vigilance: The problem of iatrogenically created psychological phenomena. *American Psychologist*, 68(2):97, 2013.

[110] Jessie YC Chen, Ellen C Haas, and Michael J Barnes. Human performance issues and user interface design for teleoperated robots. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1231–1245, 2007.

[111] C. Melody Carswell, Duncan Clarke, and W. Brent Seales. Assessing mental workload during laparoscopic surgery. *Surgical Innovation*, 12(1):80–90, 2005.

[112] Beatriz Sousa Santos, Paulo Dias, Angela Pimentel, Jan-Willem Baggerman, Carlos Ferreira, Samuel Silva, and Joaquim Madeira. Head-mounted display versus desktop for 3d navigation in virtual reality: a user study. *Multimedia Tools and Applications*, 41(1):161–181, 2009.

[113] Claudius Pfendler, Jürgen Thun, Thomas Alexander, and Christopher Schlick. The influence of different electronic maps and displays on performance and operator state in a geographic orientation task. *Behaviour & Information Technology*, 30(6):833–844, 2011.

[114] Robert Patterson, Marc D. Winterbottom, and Byron J. Pierce. Perceptual issues in the use of head-mounted visual displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(3):555–573, 2006.

[115] Tal Oron-Gilad, Elizabeth S. Redden, and Yaniv Minkov. Robotic displays for dismounted warfighters: A field study. *Journal of Cognitive Engineering and Decision Making*, 5(1):29–54, 2011.

[116] Douglas Lanman and Gabriel Taubin. Build your own 3d scanner: optical triangulation for beginners. In *ACM SIGGRAPH ASIA 2009 Courses*, page 2. ACM, 2009.

[117] Douglas Lanman and Gabriel Taubin. Build your own 3d scanner: 3d photography for beginners. In *ACM SIGGRAPH 2009 Courses*, page 8. ACM, 2009.

[118] MD Blue and S Perkowitz. Reflectivity of common materials in the submillimeter region. *IEEE Transactions on Microwave Theory Techniques*, 25:491–493, 1977.

[119] Camillo J Taylor. Implementing high resolution structured light by exploiting projector blur. In *Applications of Computer Vision (WACV),IEEE Workshop on*, pages 9–16. IEEE, 2012.

[120] Yi Xu and Daniel G Aliaga. An adaptive correspondence algorithm for modeling scenes with strong interreflections. *Visualization and Computer Graphics, IEEE Transactions on*, 15(3):465–480, 2009.

[121] Kinect fusion. `https://msdn.microsoft.com/en-us/library/dn188670.aspx`, 2015.

[122] Jing Tong, Jin Zhou, Ligang Liu, Zhigeng Pan, and Hao Yan. Scanning 3d full human bodies using kinects. *Visualization and Computer Graphics, IEEE Transactions on*, 18(4):643–650, 2012.

[123] Gregory C Sharp, Sang W Lee, and David K Wehe. Multiview registration of 3d scenes by minimizing error between coordinate frames. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(8):1037–1050, 2004.

[124] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (TOG)*, 32(3):29, 2013.

[125] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000.

[126] Sung Ha Park and Jeffrey C Woldstad. Multiple two-dimensional displays as an alternative to three-dimensional displays in telerobotic tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(4):592–603, 2000.

[127] Carole Ferrel, Jean-Pierre Orliaguet, Daniel Leifflen, Chantal Bard, and Michelle Fleury. Visual context and the control of movements through video display. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1):56–65, 2001.

[128] Justin G Hollands and Matthew Lamb. Viewpoint tethering for remotely operated vehicles effects on complex terrain navigation and spatial awareness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(2):154–167, 2011.

[129] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Human mental workload*, 1(3):139–183, 1988.

[130] C Carswell, Cindy H Lio, Russell Grant, Martina I Klein, Duncan Clarke, W Brent Seales, and Stephen Strup. Hands-free administration of subjective workload scales: Acceptability in a surgical training environment. *Applied ergonomics*, 42(1):138–145, 2010.

[131] B. Fu, W. Seidelman, Y. Liu, T. Kent, and R.G. Yang. Towards virtualized welding: Visualization and monitoring of remote welding. *ICME to appear*, 2014.

[132] G. Simmel. Sociology of the senses: Visual interaction. In R. Park and E. Burgess, editors, *Introduction to the Science of Sociology*. University of Chicago Press, 1921.

[133] M. Ott, J. Lewis, and I. Cox. Teleconferencing eye contact using a virtual camera. In *INTERCHI*, pages 119–130, 1993.

[134] Tat-Jen Cham, S. Krishnamoorthy, and M. Jones. Analogous view transfer for gaze correction in video sequences. In *7th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 1415–1420, 2002.

[135] R. Yang and Z. Zhang. Eye gaze correction with stereovision for video teleconferencing. In *Proc. Europ. Conf. Computer Vision*, volume 2, pages 479–494, 2002.

[136] J. Oppenheimer. Prompting apparatus, 1959. US Patent 2883902.

[137] S. Shiwa and M. Ishibashi. A large-screen visual telecommunication device enabling eye contact. In *SID Digest*, volume 22, pages 327–328, 1991.

[138] A. Wilson. Touchlight: An imaging touch screen and display for gesture-based interaction. In *Proceedings International Conference on Multimodal Interfaces (ICMI)*, 2004.

[139] C. Zhang, R. Yang, T. Large, and Z. Zhang. A novel see-through screen based on weave fabrics. In *IEEE International Conference onMultimedia and Expo (ICME)*, pages 1–6, 2011.

[140] Q. Yang, R. Yang, J. Davis, and D. Nister. Spatial-depth super resolution for range images. In *CVPR*, june 2007.

[141] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH ’03, pages 313–318. ACM, 2003.

[142] S.G. Chang, Bin Yu, and M. Vetterli. Spatially adaptive wavelet thresholding with context modeling for image denoising. *Image Processing, IEEE Transactions on*, 9(9):1522 –1531, sep 2000.

[143] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *Image Processing, IEEE Transactions on*, 12(11):1338 – 1351, nov. 2003.

[144] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4:490–530, 2005.

[145] M. Nilsson, J. Nordberg, and I. Claesson. Face detection using local smqt features and split up snow classifier. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 2, pages II–589 –II–592, april 2007.

[146] Vladimir Vezhnevets, Vassili Sazonov, and Alla Andreeva. A survey on pixel-based skin color detection techniques. In *IN PROC. GRAPHICON-2003*, pages 85–92, 2003.

[147] Bo Fu, Will Seidelman, Yukang Liu, Travis Kent, Melody Carswell, Yuming Zhang, and Ruigang Yang. Towards virtualized welding: Visualization and monitoring of remote welding. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.

[148] Bo Fu and Ruigang Yang. Robust near-infrared structured light scanning for 3d human model reconstruction. In *SPIE MOEMS-MEMS*, pages 89790A–89790A. International Society for Optics and Photonics, 2014.

[149] Bo Fu, Yukang Liu, Yuming Zhang, and Ruigang Yang. Virtualized welding: a new paradigm for tele-operated welding. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pages 241–242. ACM, 2014.

[150] Qing Zhang, Bo Fu, Mao Ye, and Ruigang Yang. Quality dynamic human body modeling using a single low-cost depth camera. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 676–683. IEEE, 2014.

Vita

NAME

Bo Fu

BIRTH PLACE and YEAR

1983, Xi’An, Shaanxi Province, China

EDUCATION

July 2005: B.S. in Information and Computing Science, Shaanxi Normal University, Xi’An, Shaanxi, China

July 2007: M.E. in Signal and Information Processing, Shaanxi Normal University, Xi’An, Shaanxi, China

PUBLICATIONS

B. Fu, S. Will, Y. Liu, T. Kent, M. Carswell, Y. Zhang, R. Yang. Towards Virtualized Welding: Visualization and Monitoring of Remote Welding. In IEEE International Conference on Multimedia & Expo (ICME), 2014.

B. Fu, Y. Liu, R. Yang. Virtualized Welding: A New Paradigm for Tele-Operated Welding. the 20th ACM symposium on Virtual Reality Software and Technology (VRST), 2014.

B. Fu, R. Yang. Robust near-infrared structured light scanning for 3D human model reconstruction. In SPIE MOEMS-MEMS. International Society for Optics and Photonics, 2014.

B. Fu, M. Ye, R. Yang, C. Zhang. See-Through Image Enhancement Through Sensor Fusion. In IEEE International Conference on Multimedia & Expo (ICME), 2012.