

CARNEGIE MELLON UNIVERSITY

School of Architecture
College of Fine Arts

Dissertation

Submitted in Partial Fulfillment of the requirements for the degree of

PhD in Computational Design

TITLE:

Space as Interface

A Computational Framework for Spatial Computing in Robotically-Assisted Construction

AUTHOR:

Jingyang (Leo) Liu

ACCEPTED BY ADVISORY COMMITTEE:

Daniel Cardoso Llach

Principal Advisor

June 10, 2024

DATE

Joshua Bard

Advisor

6.1.2024

DATE

Mayank Goel

Advisor

6/12/2024

DATE

Abstract

Since the early 1970s, many studies have explored integrating robotic systems into architecture, engineering, and construction (AEC) to address challenges like low productivity, worker shortages, and safety concerns. Efforts have aimed to automate construction. However, complexities such as job site environments, worker safety, and economic factors limit fully unmanned systems in AEC. The increasing trend of robotically-assisted practices underscores the need for an effective human-machine interface for robotic manipulation.

Recent advancements in spatial interfaces — immersive technologies such as augmented reality (AR) allowing users to interact with virtual information in space — offer a promising direction for Human-Robot Interaction (HRI) in construction. This dissertation introduces an AR for HRI framework tailored for robotically-assisted construction, empowering operators to leverage physical space as a shared canvas for information visualization, robot supervision, and manipulation. The framework optimizes components such as indoor localization and point cloud processing for AEC's unique demands. For instance, a multimodal localization approach combining visual simultaneous localization and mapping (vSLAM) with Ultra-Wideband (UWB) signals enhances robustness under complex lighting conditions.

Building on this framework, bespoke spatial interfaces are devised and tested in two use cases: proxemic-aware AR for HRI and robotically-assisted quality assessment and management. Case studies illustrate how AR interfaces facilitate robotic manipulation in construction contexts.

This dissertation contributes a novel AR-enabled HRI framework for robotically-assisted construction, initiating broader discussions on spatial interfaces' role in the built environment's future.

Acknowledgements

This work would not have been completed without the support and guidance I received during my PhD studies at Carnegie Mellon University. I would like to take this opportunity to express my deepest gratitude to my advisory committee members for their guidance and support. In particular, I want to extend my heartfelt appreciation to Daniel Cardoso Llach, my thesis advisor and committee chair. Your kind support, constructive feedback, and invaluable suggestions have significantly contributed to the depth and rigor of this work. I am also immensely thankful to Joshua Bard for the profound influence of your mentorship and encouragement throughout my Ph.D. journey. Your guidance helped me navigate the challenges of research and beyond. Additionally, my sincere gratitude goes to Mayank Goel for bringing interdisciplinary knowledge and inspiration to this work. Thank you for your constant support and invaluable contribution to the completion of the dissertation.

I would like to express my gratitude to my colleagues and friends from CodeLab. Thank you, MSCD alumni Yi-Chin Lee, Yaxin Hu, Wanlin Xiaoyi, Yumeng Zhuang, Hongyu Mao, and Yujeong Seo, for your valuable contribution to the research. I would also like to extend my gratitude to my Ph.D. peers - Emek Erdolu, Manuel Rodriguez Ladron de Guevara, Jinmo Rhee, Yuning Wu, Stella Shen, Zhenfang Chen, Myles Sampson, and Vina Wei. Thank you for all the joy you bring to this journey; working with you has been my pleasure. Thank you, Yunzhi Li, the research participants from the course Introduction to Architectural Robotics, and all collaborators and friends for your valuable input. Your feedback and suggestions significantly improved the work.

I would like to extend my gratitude to all faculty members in the School of Architecture. It is your contribution that made my time at CMU a wonderful experience. Specifically, I would like to thank Terry Hritz and all the assistants for your practical help and patient guidance in the DFab. I would like to thank Omar Kahn, Kai Gutschow, David Koltas for your trust and support for my PhD studies. Thank you, Alison Petrucci and Judy van Rheen, for your constant guidance on academic matters. I am honored to work with Mary-Lou Arscott, Dana Cupkova, Daragh Byrne, and Sarah Rafson on studio teaching, research, and exhibitions. Thank you for all the meaningful moments we have shared together.

With sincere appreciation, I would like to express my heartfelt thanks to Jenny E. Sabin. Thank you for introducing me to the world of academia and supporting me in every step of my academic pursuit.

Lastly, to my family, I am profoundly grateful for your unconditional support and encouragement, which made this endeavor possible.

Table of Contents

List of Figures	vii
List of Tables	xv
1 Introduction	1
1.1 Research Question	3
1.2 Vision	4
1.3 Contributions	6
1.4 Thesis Structure	9
2 Literature Review	11
2.1 Precedent works of AR for HRI	13
2.2 Domain Applications	14
2.3 Cultural and Historical Background of Spatial Interface	27
2.4 Relevance	28
2.5 Potential Areas for Future Research	31
3 Hypothesis	35
4 Method	39
4.1 Motion Planner for Site Reconstruction	43
4.2 Robust Localization Under Complex Environments	48
4.3 Geometry-informed Semantic Segmentation	59
5 Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction	67
5.1 Background	69

Table of Contents

5.2	Related Work	72
5.3	Method	75
5.4	Result	81
5.5	Conclusion	86
6	Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management	89
6.1	Background	91
6.2	Method	95
6.3	Results	102
6.4	Discussion	107
7	Conclusion and Discussion	113
7.1	Conclusion	114
7.2	Discussion	118
7.3	Limitations and Future Work	122
	Appendix A Appendix - Space and Computation	153

List of Figures

1.1	The technical framework introduced in this work can be used as a template for devising spatial interfaces for interaction in construction practices between human labor, robots and environments. Image credit: Author.	5
2.1	Miligram's Reality-Virtuality Continuum, adapted from (Milgram and Kishino 1994a).	12
2.2	AR systems support telepresent robot control (a) The telepresent control station of the Pioneer robot (Carnegie Mellon University and RedZone Robotics, Inc.) designed to inspect and assess the Chernobyl nuclear reactor (Fong and Thorpe 2001). (b) ARGOS system (Milgram, Zhai, et al. 1993a) introduced graphical elements such as virtual pointer and virtual tether to support AR-based visualization in HRI (c) AR systems designed for displaying predictive future robot actions based on the operator actions occurring (Ince, Bryant, and T. Brooks 1991).	13
2.3	AR systems integrated into a Da Vinci surgical robot for robotically-assisted surgery (RAS), AR in RAS can provide guidance to critical targets, reduce the surgeon's cognitive load, display preoperatively planned trajectories and increases the surgeon's spatial awareness (Qian, Deguet, and Kazanzides 2018a; Penza et al. 2017).	15

List of Figures

2.4	Research on HRI for creative practices (a) a tight human-machine coupling in the robotically-assisted weaving process, the robot weaves thread along a wooden frame through a series programmed motions, the user corrects the robot's mistakes on the fly, the human-machine interaction forms an overlooked type of relations termed as "assisted automation". (Cardoso, Bidgoli, and Darbari 2017)(b) A hybrid digital/physical workflow for combining human dexterity and machine repeatability in a robotic decorative plastering scenario, in which designers and craftspeople can modify the robotic trajectory through freehand sketch enabled by real time reality capturing techniques (Bard et al. 2016). (c) QUIPT, as a gesture-based control software, facilitates intuitive ways to program industrial robots using wearable markers and a motion capture system (Gannon 2014).	20
2.5	AR systems for HRI in design and making: (a) AR system enable the designer to view the predictive design feature on-the-fly in the physical environment. As soon as a design feature is completed, the robotic arm prints the new feature on-site (Peng et al. 2018).(b)AR system visualizes a spatial user interface (UI) with buttons for selecting, planning, and executing robotic instructions as well as overlays diagnostic feedback data on to the physical site (Kyjanek et al. 2019). (c)A projection-based AR system visualizes material properties such as thermodynamics for informed robot control (Johns 2014).	21
2.6	AR interface for robotic assembly and design prototyping with gesture input (Song, Koeck, and Luo 2021).	24
3.1	Research framework consisting of three research phases and their corresponding research elements. Image credit: Author.	36
4.1	The high-level framework design of a domain-specific Augmented Reality (AR) system in the context of Human-Robot Interaction (HRI), including essential components such as a motion planner, middleware, and spatial interface. The framework can be extended and implemented for various human-robot interaction tasks, particularly those relevant to construction-related scenarios. Image credit: Author.	41

List of Figures

4.2	Two phases of autonomous exploration —global exploration and local reconstruction. The global exploration phase involves mapping the surrounding area through (a) 3D exploration and (b) coverage path planning, while the local reconstruction phase captures the user-defined area at a fine resolution. In this study, I used an industrial robot platform (ABB IRB 6640) to evaluate the motion planning module for site mapping, industrial robots present reliable performance in accuracy and repeatability. However, the range and mobility of industrial robots can be confined due to factors such as singularity and a fixed number of Degree of Freedom (DoF). The integration of mobile platforms can potentially increase the work range with increased flexibility. Image credit: Author.	44
4.3	Challenging scenarios on job sites for visual SLAM including (a) a vertical white wall surface without rich textures (b) a roller door with repetitive line patterns (c) areas with artificial lighting that may change over time. The module for robust localization is exemplified and validated on an industrial robot (ABB IRB 6640). The result is compared with the built-in localization system estimated from the encoder of each joint. Image credit: Author.	49
4.4	System flowchart of the multimodal localization system. Image credit: Author.	52
4.5	Infrastructure layout of the UWB positioning system. Image credit: Author.	53
4.6	Trajectories obtained using UWB-only, VSLAM-only and the proposed multimodal localization system. Image credit: Author.	57
4.7	Performance comparison of VSLAM-only, UWB-only and the multimodal localization system under three conditions — (a) repetitive pattern, (b) poor textured surface and (c) inconsistent illumination. Image credit: Author.	57
4.8	The framework of the semantic-based system consisting of three components: 3D reconstruction, 3D segmentation and labelling. The 3D reconstruction component translates the acquired RGB and depth data from a Kinect sensor into point clouds. The 3D segmentation component segments point clouds into discrete surfaces with semantic label. The labelling component fuses geometric and semantic labels and creates anchor point for each segmented surface. Image credit: Author.	60

List of Figures

4.9	Simplified architecture of PointNet (Qi et al. 2017) (C): Concatenate (M): Max-pooling (S): Vertical stack. Image credit: Author.	63
4.10	Steps of 3D geometric segmentation : (a) Input point clouds (b) Estimate normal of each point (c) Cluster point clouds by normal and normal distance from the origin (d) Result of 3D geometric segmentation. Image credit: Author. . .	63
4.11	Qualitative model evaluation of the semantic-only segmentation approach and the label fusion segmentation approach the point cloud of scanned scenes. Image credit: Author.	65
5.1	A robot operator uses a mobile AR device to preview and edit the trajectory of an industrial robot for pick and place tasks, the AR interface overlays the waypoints, the task information, and a ghost preview of the robot target pose onto the physical environment for visual guidance and collision avoidance. Image credit: Author.	67
5.2	Proxemic-aware augmented reality (AR) may mitigate information overloading in AR-based human-robot interaction (HRI) (a) an AR interface displays all robot information at the same level of detail (b) virtual information is filtered based on human-robot proxemics, only the most relevant information is presented in the AR view. Image credit: Author.	70
5.3	System deployment diagram of the proxemic-aware AR for HRI. Image credit: Author.	70
5.4	The comparison between human-human proxemics and human-robot proxemics (a) human-human proxemics with discrete zoning (b) human-robot proxemic evaluation at sampled location (c) human-robot proxemic as a continuous function interpolated from the sampled location. Image credit: Author.	73
5.5	Compilation of co-design outputs, including sketches and prototypes, on AR visualization elements for two cases - robotic site scanning and pick & place. Image credit: Author.	78

List of Figures

5.6	Information visualization for various levels of human-robot proxemic - at the ambient engagement level, the AR system provides the overview information of the robot. At the task supervision level, more details of the ongoing task and robot will be revealed. At the third level, manipulation and control, the operator can view and edit the robot's trajectory. When the operator enters the danger zone, a safety aura and a warning sign will be displayed to alert the operator. Image credit: Author.	79
5.7	Experiment and baseline setup for surface inspection and pick & place scenario, the baseline setup displays all virtual information at the same level of details, the experiment setup selectively displays the relevant virtual information tailored to the current human-robot proxemic level. Image credit: Author.	81
5.8	Post-experiment questionnaire scores of the adaptive view management system. Error bars indicate the standard deviation. Image credit: Author.	82
5.9	Post-experiment questionnaire scores of the customized HRP model and the deterministic HRP model. Error bars indicate the standard deviation. Image credit: Author.	83
6.1	Elements of spatial annotation for robotically-assisted quality assessment and management. (1) an inspector with a mobile device, (2) a mock-up drywall for quality assessment, (3) the inspector annotates area of interest through an AR interface, (4) the robot trajectory presented via the AR interface, (5) IRB 6640 industrial robot equipped with sensors (6) linear track (7) Ultrawide-band station for indoor localization (8) a remote workstation for robot controlling and data processing. Image credit: Author.	90
6.2	An inspector with a mobile AR device chooses a target surface for robotic quality assessment. The captured scene in point clouds is partitioned into surface patches based on a point cloud segmentation pipeline, the inspector can choose a target surface by pointing the AR device to the surface and clicking on the surface within the AR view. Image credit: Author.	93
6.3	After the target surface is confirmed by the inspector, a traversal trajectory will be generated automatically for the robot (ABB IRB 6640) to cover the surface area for defect detection. Image credit: Author.	94

List of Figures

6.4	The defect detection result is registered to the physical environment and visualized through an AR interface. Image credit: Author.	95
6.5	Experiment setup (a) a 7 DoF industrial robot on a 5800 mm linear track and an interior wall mockup of 1500 mm (L) x 745 mm (W) x 1980 mm (H) (b) an end-effector sensor kit including RGB-D cameras (Kinect V2, Intel Realsense D415), laser range finders and a laser module (c) the robot highlights a detected surface defect with a laser cross-line. Image credit: Author.	96
6.6	The comparison between annotation with and without disambiguation mechanism: (a) without disambiguation mechanism, digital annotations drift while the operator moves around the target (b) with disambiguation mechanism, digital annotations anchored to the active surface invariant to the location of the AR device. Image credit: Author.	98
6.7	Motion planning for local scanning (a) the user-defined area of interest is synchronized from mobile device to the shared virtual model (b) the motion planner sample target over the area and calculate the shortest traversal path (c) the planner calculates joint value for industrial robot to reach the target. Image credit: Author.	99
6.8	Three types of defects on the dry wall component. Image credit: Author.	101
6.9	The four user study tasks: (a) surface selection (b) area of interest definition (c) robot trajectory preview (d) result preview and defect measurement. Image credit: Author.	103
6.10	The robotic quality assessment result of the interior wall mockup. Image credit: Author.	107

List of Figures

7.1	Spatial interface as a "technological mediation" augments human-world relations such as embodiment and hermeneutic relations. The diagram is inspired by Verbeek's descriptions on the mediating effect of technology (Rosenberger and Verbeek 2015; Hauser et al. 2018). Image credit: Author.	119
A.1	"Seek", 1970 by Nicholas Negroponte with the Architecture Machine Group.	157
A.2	Performative Ecologies, Ruairi Glynn, 2008	160
A.3	Left, PlaceLab Project at MIT (2003), Right, the Reconfigurable House (2008)	161
A.4	Left, the Colloquy of Mobile (1968), Right, the MusiColor Machine (1953) by Gordon Pask	163
A.5	Left, "Organisational Plan as Programme" (1965), Right, Plan and section drawings of Fun Palace (1960-1966)	165
A.6	Reconfigurable House built by Usman Haque and Adam Somlai-Fischer. . . .	171
A.7	Left, The 'CITY UNSEEN' project, Right, The Goldman projects	174

List of Tables

4.1	The translation error along each axis. Image credit: Author.	57
4.2	Quantitative model evaluation of the semantic-only segmentation approach and the revised label fusion segmentation approach on the point cloud of scanned scenes. Image credit: Author.	64
6.1	SUS questions	103
6.2	Additional question to evaluate the AR interface	104
6.3	Users (U_i) response to the SUS questions (Q_i)	105
6.4	Users (U_i) response to the supplemental questions (Q_i)	106

Introduction

With the development of spatial computing technologies, spatial interfaces such as AR/MR have been adopted in adjacent domains such as robotic surgery (Qian, Deguet, and Kazanzides 2018b, Qian, Deguet, Z. Wang, et al. 2019) and manufacturing (Soh-Khim Ong et al. 2020, Stadler et al. 2016). Within the domain of robotically-assisted construction, the study of spatial interfaces for human-robot interaction is still at an early stage. Several use cases demonstrate the application of AR in robotically-assisted design (Johns, Anderson, and Kilian 2019) and assembly (Kyjanek et al. 2019). However, a comprehensive framework for spatial interfaces in robotically-assisted construction still remains absent. Addressing the gap, this study aims to propose a computational framework to integrate spatial interfaces into robotic construction systems. My study identifies and explores the following potentials for spatial interfaces in the context of robotically-assisted construction: (1) contextualizing digital information such as robot status and trajectories, (2) bridging the mismatch between how robots and humans represent the surrounding environment, and (3) providing a shared spatial frame for information exchange. The aforementioned features may mitigate the tension between the existing human-machine interface directly imported from industrial automation and the unique domain challenges of the robot-assisted construction task. This dissertation also brings up a discussion on the inter-dependency between space and computation, providing an overview of the history, background, and intellectual lineage of computation in physical environments and tangibility in computing.

Chapter 1. Introduction

In terms of application area, the study is contextualized in robotically-assisted construction to investigate how the integration of context-aware AR systems may provide robot operators with embodied feedback from the surrounding environment. Taking the robotically-assisted quality assessment and management (QA&M) task as an example, robots equipped with sensors can assist inspectors by reconstructing the environment and identifying and quantifying defects such as deformation, discoloration, and surface anomalies. The integration of AR tools can aid inspectors by (1) supporting data visualization by overlaying the information of robots and building components on the physical environment: AR systems can synthesize the information collected by sensors on robots with contextual knowledge for more intuitive visual feedback. The physical environments in the background can help users to better understand and engage with the superimposed digital information. (2) facilitating intuitive robotic manipulation by enabling inspectors to program robots through well-entrenched interaction techniques: while existing robot interfaces often rely on online control via joysticks with a video interface or offline programming, AR systems can provide a high-level supervisory interface for robot manipulation. For instance, when programming a robot to inspect an area of interest on a wall surface, the robot operator can use hands to point to the target and annotate physical places and objects with digital information. By leveraging the inspector's intuitive spatial skills as interaction input, we can lower the entry barriers for supervising and manipulating robots in the robotically-assisted QA&M task. (3) enhancing collaboration between stakeholders by leveraging the physical environment as a shared workspace for information exchange: tangible objects in the physical environment become shared referents for virtual information, allowing both co-located and remote teams to use computationally mediated physical spaces to construct inter-referential awareness. The inspector can demonstrate the information associated with a set of objects by hand pointing to the real objects. Based on these collaborative referents, the anchored digital information can be communicated between both on-site and remote team members. From a technical perspective, this study introduces a novel framework for building a context-aware augmented reality (AR) system for robotically-assisted construction. The framework is composed of three key components: (1) a motion planner for autonomous mapping, (2) a middleware for robust localization and converting raw data into a compact

semantic¹ representation, (3) a spatial interface for users to intuitively interact with virtual information in the physical environments.

At the conceptual level, this study delves into the tight connection between physical space and computation, tracing the intellectual trajectories of physical space augmentation by computational systems and the spatial turn in the field of human-computer interaction.

This study proposes a technical framework for spatial computing in robotically-assisted construction practices, informed by both technical barriers and contextual constraints. Built upon the technical framework, I developed bespoke AR for HRI interfaces and documented two use cases. The use cases serve to demonstrate how spatial interface can provide an accessible interface for students² and practitioners (building inspectors) in the AEC domain to operate robots for construction-related tasks. Additionally, the discussion chapter reflects on the topic of the role of physical space in computation, providing a context that may pave the way for future research in spatial computing within the built environment.

1.1 Research Question

This research aims to explore the mutual constitution between the physical space and spatial interfaces and its application in human robot interaction (HRI) for robotically-assisted construction practices. Drawing from an enactive view³, the study proposes to leverage computa-

¹Semantics in machine learning and deep learning is a high-level data representation based on syntax (the rules of grammar in sentence structure) and deixis (common words that give context to a place, time, or person). Semantic technologies refer to associating the meaning in context to the input raw data, such as pixels and point clouds, through a statistical learning process. As a machine-processable and human-readable format, semantics can be an intermediate representation for building interpretable relationships among variables of the system. In statistical learning, the extent of correspondence between the real objects and their mathematical representations is bounded and dependent on the potency of the utilized idealization (simplifying properties of an object) and abstraction (representing an object with another object that is easier to handle) schemas (Tunç 2015), a human-understandable representation such as semantics can potentially inform us on the properties of real objects the model refers to.

²AR technologies provide a novel medium for robotic training with a potential enhancement in the learning experience. AR can provide several unique features in pedagogy, including (1) an interactive pre-training environment - students with a mobile device can interact with the virtual representation of a robot overlaid onto the physical environment before they actually operate the robot. (2) a step-by-step AR guidance - AR interface with superimposed visual cues, annotations, and interactive instructions can help students follow the procedure with reduced errors (3) a collaborative learning environment - Multiple students in a shared space can interact with the same virtual robot, which can foster teamwork and communication in the teaching session.

³Shared by both post-cognitivist and enactivist research frameworks, the evolution and development of the cognitive abilities of an organism are constructed based on the history of the interaction of that organism with

Chapter 1. Introduction

tionally augmented space as an interface for human-robot interaction and investigate the role of the physical space in HRI by asking a broad set of questions from three perspectives:

1. How might we use physical space to contextualize virtual information?
2. How might we leverage physical affordances for intuitive robot programming?
3. How might we use spatial relationships in physical space to mediate robot operators' interactions with virtual information?

1.2 Vision

My vision for AR-enabled construction is based on the assumption that humans can play a critical role in future construction in the domain of decision-making, design, and ethical considerations. Meanwhile, as a supplemental role, mechanical systems, including cranes, hoists, and variations of construction robots, can assist human labor in tasks such as heavy material handling and site monitoring with enhanced operational safety, efficiency, and precision. To facilitate human-robot interaction, this study explores how might we use physical environments as an interface for robot operators to supervise and control robots in robotically assisted construction practices.

With the development of spatial interfaces, such as augmented reality, digital information can be overlaid onto the physical world via wearable or handheld devices. Many studies have been conducted to explore the applications of spatial interfaces in diverse domains, including manufacturing, entertainment, and education (a more detailed review can be found in Section 2 - Literature Review). However, a comprehensive study on spatial interfaces in the domain of robotically-assisted construction remains absent. As an initial step, in this work, I attempt

its environment. Organic coordination within the inactive approach gives rise to the sensorimotor abilities of the organism. Enactivism and embodiment provide alternatives to cognitivism, computationalism, and Cartesian dualism by shifting the focus from an isolated cognitive process to a knowledge formation process dynamically unfolded in interaction with the environment (F. J. Varela, Thompson, and Rosch 2016; Jaegher and Rohde 2010; Hutto and Myin 2012). A discussion on the theories of enactivism and embodiment intersects with several broad fields of inquiry and analysis such as phenomenology and pragmatism. For instance, the writings of phenomenologist Maurice Merleau-Ponty in *The Phenomenology of Perception* (Merleau-Ponty 2013) and the American pragmatist John Dewey in *Art as Experience* (Dewey 2005) are highly influential for subsequent studies on both embodied cognition and multiple disciplines such as design (Kilian 2017), robotics (R. A. Brooks 1991b), artificial intelligence (Clancey 1997) as well as human-computer interaction (Dourish 2004a).

Chapter 1. Introduction

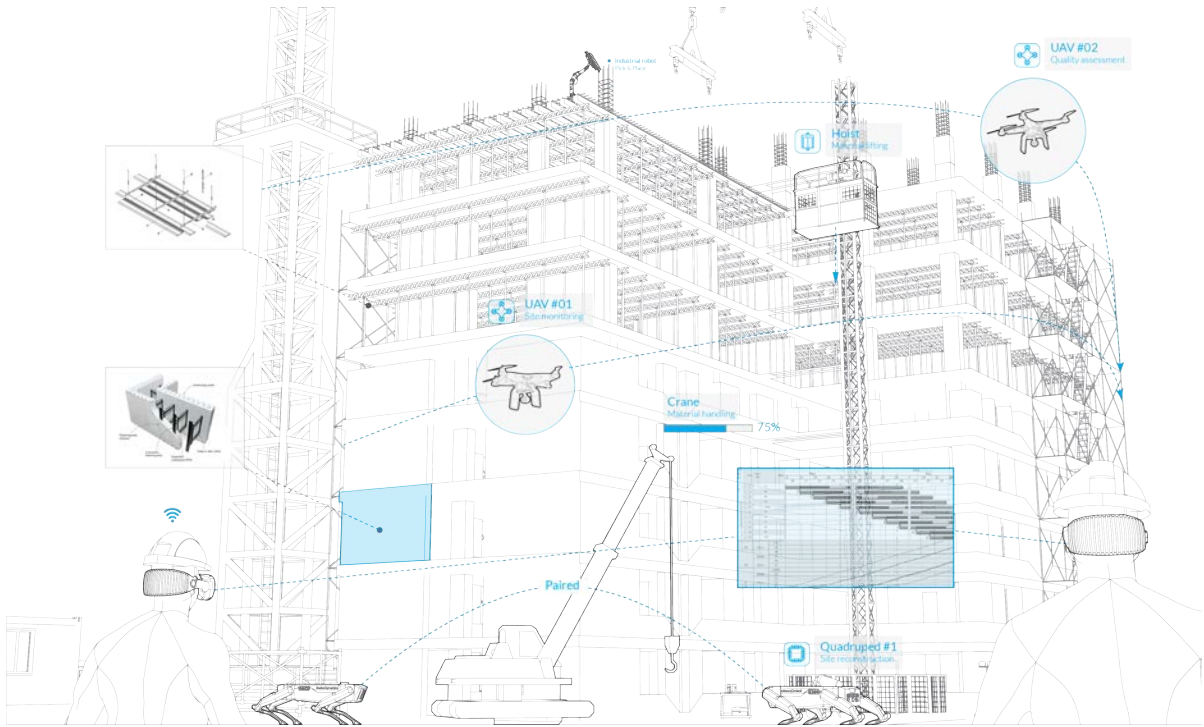


Figure 1.1: The technical framework introduced in this work can be used as a template for devising spatial interfaces for interaction in construction practices between human labor, robots and environments. Image credit: Author.

to provide a technical framework for AR-enabled HRI in construction robotics. In the case study, I found that the bespoke AR interfaces developed based on the framework provide robot operators with enhanced features in visualization and programming.

The modular structure of the AR for HRI system can be extended for the development of human-human and human-building interfaces within construction practices. For example, in construction, pipes, wirings, and structural elements can be hidden or occluded under floors. By registering the as-built data with the as-plan building information model, detailed schematics of the occluded building component can be juxtaposed with the associated physical structure, providing on-site workers with information on the internal layout of the building. In addition, as a human-human interface, AR allows all involved stakeholders on-site to collaboratively interact with virtual information in situ, the physical environment can serve as a shared spatial frame for all participants to anchor virtual information, digital annotations, and design variations in a real-time and interactive manner.

Shifting from the focus of automating construction practices, this work investigates the interface between humans and machines in construction. Leveraging recent developments in spatial interfaces, I designed a novel computational system for AR integration. The AR interface allows human workers to control and supervise robots through an immersive and embodied interface, fostering a synergy between human labor, machines, and the work environment within construction processes (Figure 1.1).

1.3 Contributions

This study contributes to the AEC domain by (1) presenting a novel computational framework for integrating spatial interfaces, such as augmented reality (AR), in robotically-assisted construction and (2) demonstrating how AR can serve as an effective human-machine interface in multiple scenarios such as surface inspection and material handling. The key components of the computational framework, including localization and 3D segmentation, are optimized to address challenges such as inconsistent lighting, cluttered environments, and unknown conditions typically encountered on construction sites. According to the technical evaluation, the components demonstrate improvements in accuracy and robustness compared to the state-of-the-art solution. Based on this framework, I documented two use cases illustrating how an AR interface can facilitate information visualization and in-situ robot programming in practices such as building inspection and material handling. The modular and reconfigurable framework introduced in this study can serve as a baseline and reference point for future research that focuses on the interaction between human operators and robotic systems in construction tasks.

Specifically, the contributions of this study lie in three perspectives.

- A computational framework for AR for HRI in robotically-assisted construction

With the goal of facilitating robotic programming and coherent collaboration, the study proposes to integrate context-aware AR systems into robotic manipulation. Users thus can leverage the inhabited environments for data visualization, task planning, and information exchange. Based on three core components, including a motion planner, a middleware, and a spatial interface, the framework provides robot operators with an embodied interface by leveraging physical environments to contextualize and share digital information.

Chapter 1. Introduction

Each core component of the framework is optimized for the specific limitations and requirements of the task. In robotically-assisted construction practices, the complete spatial information of the working environment may not always be accessible. A timely and accurate site information can be critical for robotic task planning and navigation. The back-end motion planning component generates near-optimal paths for robots to explore the surrounding environment and construct a multi-resolution map of the site. The component integrates frontier-based and information-based methods to maximize coverage. The detail of the implementation will be elaborated in the Chapter III. The middleware then maps the point cloud acquired by the laser scanner to high-level semantic data; this process is known as scene understanding in computer vision and robotics. In this study, I used surfels (planar surface patches (Pfister et al. 2000)) as the representation of the scene understanding model. In a comparison study with the state-of-the-art solution, the surfel representation achieves better performance in terms of accuracy and intersection over union (IoU). The third component, spatial interface, facilitates the interaction between robot operators and robots by integrating virtual elements into the physical environment. In the second case study, I collaborated with inspectors and learned how they usually annotate the area of defects and the component to be replaced with a spray gun or a pencil. Instead of replacing the sensorimotor experience of annotating, I exploited the experience of how inspectors interact with the surrounding environment through annotation and devised a task-oriented interaction modality.

- A novel demonstration of the proposed framework in robotically-assisted quality assessment and management application

Quality assessment and management (QA&M), as an integrated procedure in the post-construction phase, involves rigorous checks of fine details and rectification of defects. To assist inspectors in identifying and quantifying defects, multiple studies (Victores et al. 2011, Bandyopadhyay et al. 2018, La et al. 2019) have been conducted to adopt robotics systems equipped with sensors for in-progress monitoring and final quality inspection.

However, compared to the factory settings, the robotically-assisted QA&M in the AEC domain can be challenging due to complex environments. For example, the lighting conditions can be unpredictable which may significantly affect the reliability of machine vision systems, the environment may not be known a priori which can pose challenges to the robot navigation and task planning process. In the near future, the operator's judgment and tacit knowledge

Chapter 1. Introduction

can still play a significant role when in-situ decision-making is required. To facilitate in-situ robot control and supervision, this study proposes a novel workflow that integrates AR systems for effective robot programming. The inspector with a mobile device can annotate the area of interest for inspection through a spatial interface and use the inhabited environment as a shared medium to review the identified defects superimposed onto the physical environments.

- A conceptual framework to examine the role of physical environments in computation.

The investigation of space as an interface for human-robot interaction can lead to a broader set of questions in examining the tight linkage between the physical environment and computation. Tracing within the context of computation, the study outlines the changing role of space in computation in a shift from space as container, space as metaphor to space as interface.

The study draws from the intersection between the domain of the AEC (Architecture, Engineering, and Construction) and robotics, investigating a computational framework for AR-based human-machine interface in robotically-assisted construction practices. The combination of AR and robotics can offer a unique technical setup for collaborative robotics and interactive fabrication in computational design. The AR-based human-robot interface can potentially reconnect design exploration with materialization, allowing the designers to interactively explore the design space in the fabrication process. By leveraging the physical environment as a medium for visualizing and exchanging design intent, the study reframes the robot from a mere executor to a part of design ecosystems as a role that extends the malleability of computation to physical space (Johns, Anderson, and Kilian 2020). In a larger context, this work can contribute to computational design research by bridging the virtual design space with physical environments through an augmented reality system. As a computer-aided design and computer-aided manufacturing interface, the AR framework introduced in this study can bring new features and opportunities to assist computational designers, digital makers, and AEC practitioners by (1) integrating robotic fabrication into the design process, the novel AR interface allows designers to program robots and preview designs in physical environments through direct interaction, the feature can lower the entry barrier for robotic manipulation and enable an integrated design-to-fabrication workflow at an early stage. (2) facilitating in-situ design exploration, the AR system allows designers to explore design and fabrication strategies with the actual target environment, leveraging the physical interaction with the surroundings and on-site feedback for more informed decision-making (3) enhance the direct engagement with

the physical context, by connecting virtual design space with physical fabrication environment, the AR system allows the designer to directly engage with the materials, physical tools, and machines and contextualize the digital information in the work environment.

1.4 Thesis Structure

- Chapter 1 — introduction discusses the research background and outlines key research elements, including research questions, research methods, and novelties.
- Chapter 2 — literature review — entails a comprehensive examination of spatial interfaces in human-robot interaction. The chapter is organized into six sections. Section I - Precedent works of AR for HRI examines the intellectual origins of the field of AR for HRI. Section II outlines the landscape of how spatial interface may serve as a novel human-robot interface and transform state-of-the-art robotic practices in application domains such as manufacturing, design, and making. Sections III&IV discuss the potential challenges of integrating spatial interfaces in robotically-assisted construction from design to implementation. Section V discusses the mutual constitution between computation and space, outlining the historical and cultural background of spatial interface from the lens of architecture and human-computer interaction.
- Chapter 3 — hypothesis — introduces the hypothesis of the study derived from the background study, with a reflection on how AR for HRI might be applied for the context of robotically-assisted construction.
- Chapter 4 — method — introduces a computational framework for integrating spatial interface into robotically-assisted construction practices. The framework allows robot operators to preview virtual information of robots in the physical environment and manipulate robots through mobile AR devices. The framework can potentially provide systematic guidance for constructing and organizing components that constitute an AR for HRI systems. The chapter introduces three technical components in depth to demonstrate how the framework adapts to contextual constraints in robotically-assisted construction practices.

Chapter 1. Introduction

- In Chapter 5 and Chapter 6 are devoted to two case studies for testing the hypothesis. In the first case study — proxemic-aware AR for HRI, I investigate how the physical relationship between robot operators and robots can be leveraged to mitigate information overload in AR for HRI systems. In the second case — spatial interface for robotically-assisted quality assessment and management (Q&M), I documented how building inspectors may use the framework to annotate areas of interest for robotically-assisted inspection and preview the inspection result through an AR interface.
- Chapter 7 — conclusion and discussion— discusses the findings of case studies, limitations of the framework and potential directions for future research. With a review of the work, I revisit the definition of spatial interface and tentatively broaden the concept of spatial interface as an augmentation of human-world relationships. Following up in the appendix, I explore the tight connection between the physical space and computation from multiple perspectives, such as architecture, urban design, and human-computer interaction, to build a ground for future studies on the expanded domain of spatial computing and its specific applications within varied contexts.

Literature Review

Spatial interfaces¹ refer to a set of technologies focusing on overlaying computer-generated information on the real-world environment, commonly associated with immersive technologies such as virtual reality (VR), augmented reality (AR) and mixed reality (MR). As an umbrella term, spatial interfaces encompass user interface metaphors that superimpose 2D or 3D graphics on the user's view of the physical environment in real-time at different extents, or the *Reality-Virtuality Continuum* (Figure 2.1) (Milgram and Kishino 1994a). The continuum has two extrema, on the right side is the real environment and, the left side a fully virtual environment. Spatial interface technologies, such as augmented reality (AR)², is a subset of this continuum, combining virtual and real elements together and enhancing the user's view of the real world.

This chapter reviews precedents and state-of-the-art studies within the domain of augmented reality for human-robot interaction. It traces the intellectual lineage from pioneering works on tele-operated robots to recent developments in spatial interfaces, such as augmented reality for human-robot interaction, with applications in design, manufacturing, and medi-

¹In (Steinicke and Stuerzlinger 2015), Frank Steinicke et al. outline that spatial interfaces research focuses on "the user interface challenges that appear when humans interact in the space where the flat, 2D, digital world meets the volumetric, physical, 3D space we live in. This considers both spatial input and output, with an emphasis on the issues surrounding human and system interaction."

²Augmented reality is defined as an interface technology that fulfills three key requirements including: (1) It combines real and virtual content (2) It is interactive in real time (3) It is registered in 3D (Azuma 1997a). Namely, AR systems can display virtual content such as images and videos in real world, respond to user input through interactive graphics and find the position of other users viewpoint and enable the virtual image to appear fixed in the real world.

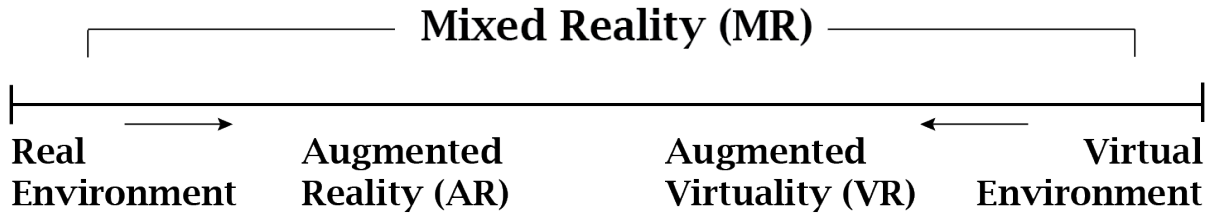


Figure 2.1: Miligram's Reality-Virtuality Continuum, adapted from (Milgram and Kishino 1994a).

cation. Within the context of robotically-assisted construction, I outline the challenges and gaps in integrating AR systems into the existing human-robot interaction workflow from multiple perspectives. Challenges from the work site environment, such as inconsistent lighting and clutterness, pose obstacles to the technical framework design. Organizational factors also impact how practitioners in the AEC sector integrate into the existing or future robotically-assisted construction workflow.

The study of the human-machine interface for industrial robots began in the 1940s. Early interactions were mostly unidirectional, relying on mechanical systems to encode and decode instructions. In the 1980s, high-level programming languages emerged to address the increasing complexity of robotic systems. In recent decades, human-robot interfaces have evolved beyond a single modality, incorporating speech, gestures, and body motion for more intuitive interaction.

In this work, I focus on the recent development of spatial interfaces, investigating how spatial interfaces may assist operators in robot control and supervision in robotically-assisted construction tasks.

This chapter first outlines the history and reviews recent developments in the AR for HRI domain. With a focus on the application of robotically-supported construction, I have identified three main research gaps in existing AR for HRI systems that could be potentially improved. These include (1) limited support for task-oriented interaction techniques, (2) the absence of high-level virtual-physical registration, and (3) limited cross-device collaboration. A detailed discussion can be found in section 2.5 potential areas for future research.

2.1 Precedent works of AR for HRI

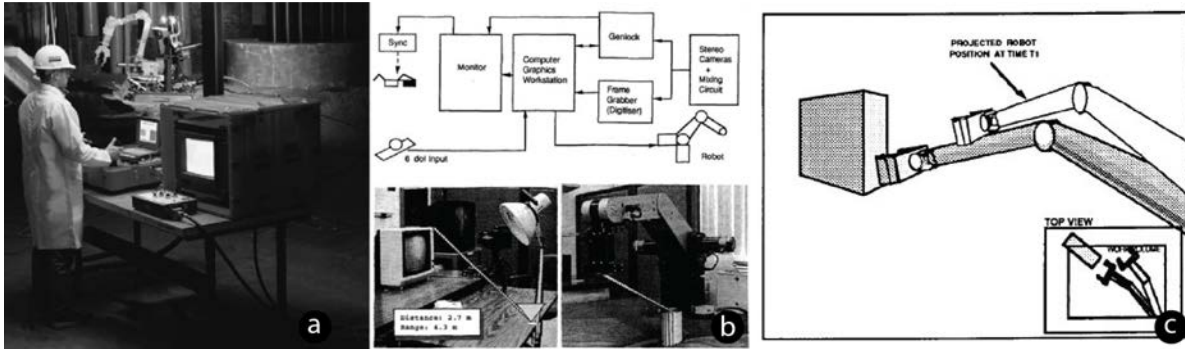


Figure 2.2: AR systems support telepresent robot control (a) The telepresent control station of the Pioneer robot (Carnegie Mellon University and RedZone Robotics, Inc.) designed to inspect and assess the Chernobyl nuclear reactor (Fong and Thorpe 2001). (b) ARGOS system (Milgram, Zhai, et al. 1993a) introduced graphical elements such as virtual pointer and virtual tether to support AR-based visualization in HRI (c) AR systems designed for displaying predictive future robot actions based on the operator actions occurring (Ince, Bryant, and T. Brooks 1991).

AR technology provides opportunities in the design and development of natural and intuitive mechanisms for human-robot interaction. The earliest major work leveraging AR for HRI appears to date back to the late 1980s with a focus on robot tele-operation system (Szaifir 2019) (Figure 2.2). The man-machine systems Lab at MIT contributed pioneering works to the development of tele-operation and supervisory control of robotic systems. To facilitate telerobotic system control, many studies explored effective subsystems for robot simulation (Browse and Little 1991), predictive graphics display (Bejczy, W. S. Kim, and Venema 1990a), and graphic operator aids (Ince, Bryant, and T. Brooks 1991 ; Zhai and Milgram 1992). One of the most developed prototype can be the Augmented Reality through Graphic Overlays on Stereovideo (ARGOS) system (Milgram, Zhai, et al. 1993a). ARGOS used overlaid virtual stereographical elements for information visualization to improve the communication between the operator and a teleoperated robot. Design concepts such as “stereographic pointers”, “virtual tape measure” and “virtual landmarks” were introduced to achieve effective “spatial human-robot communication.” After the 1990s, the research on spatial interfaces for HRI was conducted in a fragmented rather than cohesive manner (Szaifir 2019). For instance, Computer graphics communities and venues such as ISMAR (IEEE International Symposium on Mixed

Chapter 2. Literature Review

and Augmented Reality) and 3DUI (3D User Interfaces) focus on spatial display techniques in HRI. HCI and HRI communities and venues such as SIGCHI (The Special Interest Group on Computer–Human Interaction), UIST (The ACM Symposium on User Interface Software and Technology), IROS (International Conference on Intelligent Robots and Systems), ICRA (IEEE International Conference on Robotics and Automation) and etc. have published studies for evaluating the use of AR systems in HRI through user studies. The result showed that AR systems could significantly improve robotic control performance (Bowen et al. 2004 ; Maida, Bowen, and W Pace 2006). Domain-focused research in manufacturing, robotically-assisted surgery, and design illustrated how AR for HRI can be applied to specific user scenarios. As the research on AR for HRI gradually converges, a dedicated venue on the topic of Virtual, Augmented, and Mixed Reality for Human-Robot was held in conjunction with the IEEE/ACM HRI conference in 2018. Emerging studies on techniques, architectures and the broader impact of incorporating AR systems into HRI processes have been presented in an organized format.

2.2 Domain Applications

In this section, I will outline the state-of-art applications of AR for HRI in multiple domains, including industrial applications, design and making and construction robotics.

2.2.1 AR for HRI in Industrial Applications

For industrial applications, spatial interfaces for HRI are applied in domains such as surgery (Figure 2.3) to support features including but not limited to (1) task guidance (2) interactive planning (3) Motion supervision (4)tele-operation (5) sensory substitution (6) skill training and (7)safety.

- Task Guidance

Augmented reality systems can serve as a secondary visual source to overlay graphics on the main visual source for task guidance. Information that is hidden but useful to the robot operator can be registered and visualized to the main visual field to support cognition processes.

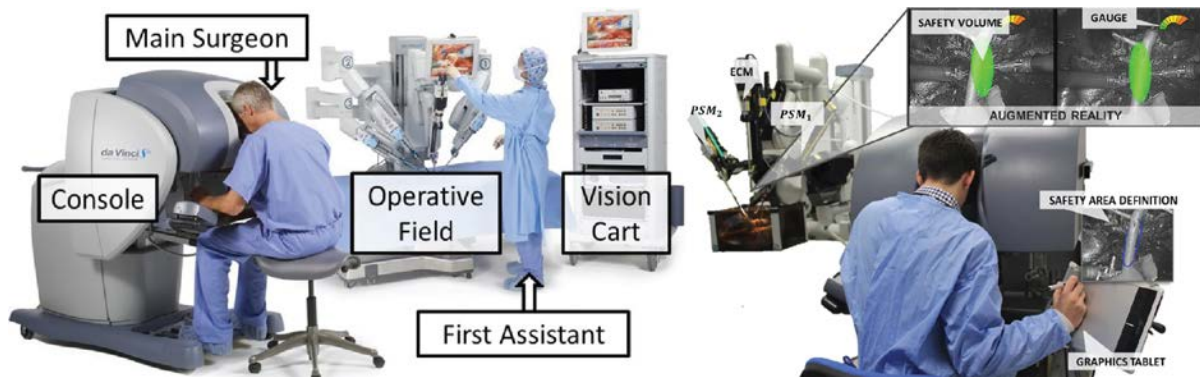


Figure 2.3: AR systems integrated into a Da Vinci surgical robot for robotically-assisted surgery (RAS), AR in RAS can provide guidance to critical targets, reduce the surgeon's cognitive load, display preoperatively planned trajectories and increases the surgeon's spatial awareness (Qian, Deguet, and Kazanzides 2018a; Penza et al. 2017).

The hidden information including critical structures that are difficult to distinguish, preoperative information about robots, and data from sensors can be visualized by AR systems. For instance, RPAR-II system (Fang, S. K. Ong, and A. Y. C. Nee 2012a) uses AR to enhance the human-virtual robot interaction in tasks such as pick-and-place, an AR system visualizes a virtual robot and its trajectory planning, and users can use a maker-cube to edit the virtual trajectory which will be further optimized into an executable robot path.

- Interactive Planning

Robot programming using AR allows operators to intuitively interact with robot programming progress using virtual robot models without having to interact physically with the real robot manipulators. The AR-based approach in robot planning can significantly reduce the setup and programming time and lower the learning curve for novice users. For instance, studies such as (Jonathan Wun Shiung Chong et al. 2009 ; S. K. Ong, J W S Chong, and A. Y. C. Nee 2010) presented methods to plan a collision-free path through guiding a virtual robot using a probe attached with a planar marker.

- Motion Supervision

In some cases, robots are designed to autonomously perform specific tasks during the period of autonomous operation; it is critical for the operator to closely monitor the robot's motion or action plan to ensure safety when the robot malfunctions. For instance, (Makhataeva,

Chapter 2. Literature Review

Zhakatayev, and Varol 2019a) presents the safety aura concept that uses AR to augment an operator's awareness of safety by overlaying a highlighted safety zone around an operating robot. Similarly, (Walker et al. 2018a) uses AR for robot motion intent visualization for an aerial robot during its autonomous flight over a predefined trajectory.

- Tele-operation

Robot tele-operation is a method of controlling a robot remotely by a human operator (Niemeyer et al. 2016 , Thomas B Sheridan 1992), which is widely adopted in domains including space exploration (Menchaca-Brandan et al. 2007), disaster response (F J Perez-Grau et al. 2017), and surgical robots (Okamura 2004). Tele-operation requires the constant attention of the operator which may bring cognitive burden and leave less time for other on-going tasks (Pitman and Cummings 2012). As support for robot piloting, two forms of interfaces are commonly used in robot tele-operation: a direct tele-operation interface allows operators to control robots using joysticks with a video display, or a high-level supervisory interface that represents robot motion by waypoints or other spatial-temporal indicators (Szafer, Mutlu, and Fong 2017). With the recent development in AR-supported hardware (e.g., AR head-mounted displays (ARHMD)), an increasing amount of study has been done to use AR to support natural spatial coordination in robotic tele-operation. For example, as a support for direct tele-operation, vision-based interfaces combine a live first-person view of robots with virtual information such as map data via display glasses. However, the missing surrounding environment outside of the robot's Field of View (FoV) can affect the operator's overall situational awareness (Hedayati, Walker, and Szafer 2018). Compared with direct tele-operation, a high-level supervisory interface can reduce the entry barrier without requiring additional familiarization and training (Ibrahimov et al. 2019). Operators can use well-entrenched skills such as hand gestures, arm movements, and gazes to interact with digital information projected onto the physical environment (Mistry and Maes 2009) or industrial manipulators (Ismail et al. 2019).

The integration of AR in tele-operation can (1) enable seamless and real-time remote collaboration and control. AR-based interface has been integrated in industrial robots to assist tele-operation as a replacement for conventional teachpendant, According to the usability test, AR interface can improve the speed of tele-operation tasks with enhanced ergonomics (García et al. 2022). (2) Remote visualization — In applications such as remote surgery, AR can provide the remote operator with immersive visual feedback including depth information, anatom-

Chapter 2. Literature Review

ical information, and surgical instrument information. The optical see-through head-mounted displays (OST-HMD) mounted at the end-effector of robots can stream views from different perspectives and visualize the key anatomies and instruments inside the patient's body (Z. Lin et al. 2022). (3) Predictive display — The latency from communication between the operator and the remote robot can be a technical challenge in robot tele-operation, long latency due to network limitations can accumulate to the second level when the operator is far from the telerobot (Lum et al. 2008). A predictive display, generally designed for space operations, can predict the potential result of an operation within a short amount of time (Richter et al. 2019). Supported by AR technology, the prediction can be rendered and overlaid onto raw video data as an approach to overcoming long latency in tele-operation. The predictive result based on high-fidelity simulation combined with real scenes can enhance the operator's telemanipulation task performance significantly in long-distance tele-operation tasks (Bejczy, W. S. Kim, and Venema 1990b).

- Sensory Substitution

AR can serve as an alternative visual solution for invisible sensory feedback such as haptics. In applications such as tele-operated robotically-assisted surgery, implementing direct haptic feedback remains impractical (Akinbiyi et al. 2006); thus, studies (Okamura, Verner, et al. 2011) explore how to transcode information from haptic modality to present it in a visual modality. Enabled by AR, force feedback can be visualized to the surgeon who operates surgical robots; the visual feedback receives a strongly positive acceptance according to the survey (I Aviles-Rivero et al. 2018).

- Skill Training

The main advantage of using AR for training is that the trainee can interact with real-world objects and simultaneously access the virtual information for guidance (Webel et al. 2013). Trainees can use intuition to create a mapping between the training and the real world. The tactile feedback during the training can be maintained, virtual information on the task delivered via AR supplements the trainee's task knowledge. Compared to the conventional training process in which the trainee read instructions on a separate medium such as user manuals, AR-based training platform allows for more efficient and embodied skill acquisition. Studies

Chapter 2. Literature Review

(Chowriappa et al. 2015 ; Dickey et al. 2016) use AR-enhanced training systems to improve the training process of robotically-assisted surgery.

- Safety

Physical human-robot interaction (pHRI) investigates how humans and robots physically interact with each other in a shared environment. The study focuses on the design, implementation, and evaluation of robotic systems that can work in proximity to humans. With the integration of robots, the issue of safety during human-robot collaborative tasks can become critical, which still remains an open problem in HRI.

Several solutions have been proposed to ensure physical safety during pHRI. In terms of control, studies have been conducted to avoid potential collisions through real-time motion planning. For instance, a soft motion trajectory planner can be used to ensure a collision-free trajectory that guarantees the safety and comfort of operators (Broquere, Sidobre, and Herrera-Aguilar 2008). By tracking the energy stored in the system, we can minimize the contact force between robots and operators in collaboration through online trajectory planning (Laffranchi, Tsagarakis, and Caldwell 2009). Adaptive control algorithms (Calanca, Muradore, and Fiorini 2015 , Vanderborght et al. 2013) can adjust robot control parameters such as speed and force when encountering unexpected obstacles or resistance from the environment. From the perspective of hardware design, the incorporation of force and torque sensors into joints can measure the actual forces and torques applied by robot's joints and end-effectors. Sudden changes may indicate potential contact or collisions with objects, emergent actions can then be taken to reduce the risk of harm. By limiting the inertia and weight of robots (Albu-Schäffer et al. 2007), the impact forces between operators and robots in accidental contact can be reduced, decreasing the risk of injury to humans.

Augmented reality as a novel interface for human-robot interaction can enhance information sharing and communication in the manufacturing process and improve machine control. Within the context of mitigating safety concerns, AR can play a role in communicating task-specific safety information to the operator and enhance the operator's awareness of potential risks. For instance, AR can serve as a virtual barrier to protect human operators in robotically-assisted manufacturing environments (Hoang et al. 2022). Together with depth sensors, the safety zones in HRI can be created and visualized through an AR headset (Hietanen et al. 2020). A safety metric based on distance and velocity can be visualized as a safety

Chapter 2. Literature Review

aura, an operator with a pair of AR goggles can identify the safe area while controlling the robot (Makhataeva, Zhakatayev, and Varol 2019b). In robotic surgery scenarios, errors such as trajectory misjudgment, time latency, and unexpected collision can result in serious safety issues for patients. AR display can serve as an assistive display tool to facilitate error-proof robotic manipulation and reduce the risks resulting from unintended consequences or errors. For instance, several authors proposed to utilize AR interfaces plan for the needle placement procedure in a surgeon-robot cooperative system, the safety and accuracy of a percutaneous ablation can be improved with the integration of AR interfaces based on a set of pre-clinical tests (Wen, Chui, et al. 2013, Wen, Tay, et al. 2014). However, the integration of AR interfaces can also bring new challenges to the existing HRI systems. When AR overlays virtual information over physical objects, the accuracy of registration is critical to the operator's situational awareness. In addition, the reliability of AR systems can affect the safety of the HRI system, there is still a significant gap between the advanced algorithms and clinical/industrial standards (Qian, J. Y. Wu, et al. 2019a).

As presented in the aforementioned review, in the domain of industrial applications, AR technologies can contribute to various industrial processes by improving efficiency, productivity, safety, and decision-making in human-robot interaction. AR technologies not only play a significant role in data visualization but also enhance the collaboration between workers in real physical environments. Especially with the ever-increasing networked devices in factory settings, AR can serve as a group-level human-machine interface for operators to access relevant information in real time and interact with virtual content superimposed on physical contexts. However, for industrial applications, several factors need to be taken into consideration. Firstly, AR technologies can bring extra privacy concerns in workplaces. Since the data can be monitored and presented to all stakeholders, the risk of issues such as unauthorized access, data breaches, or data misuse by third parties can increase in AR systems. The required supportive technologies for AR applications, such as location tracking, visual and audio recording, and biometric data processing, can rise privacy concerns. Secondly, the unique challenges and requirements of industrial environments can pose challenges to the design of interaction in AR systems. For example, workers in a factory settings often have their hands occupied with tools and machinery, hands-free interaction techniques allow workers to interact with AR systems without interrupting their on-going workflow. In addition, background noise

in factory environments can affect the voice-based interaction modality. Most importantly, safety issues need to be addressed in AR for HRI in industrial environments. Clear visibility of AR interface, warnings alerts for hazardous situations, and proximity sensors on devices for collision detection can be integrated into AR systems to prevent accidents and minimize distraction.

2.2.2 AR for HRI in Design and Making

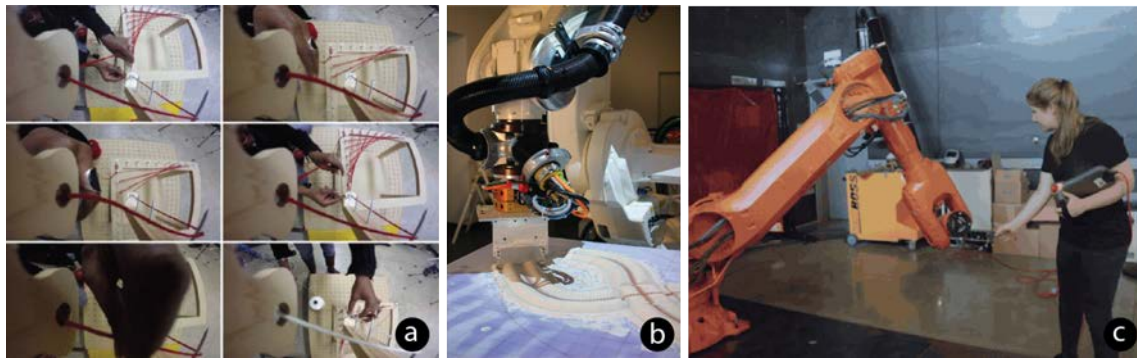


Figure 2.4: Research on HRI for creative practices (a) a tight human-machine coupling in the robotically-assisted weaving process, the robot weaves thread along a wooden frame through a series programmed motions, the user corrects the robot's mistakes on the fly, the human-machine interaction forms an overlooked type of relations termed as "assisted automation". (Cardoso, Bidgoli, and Darbari 2017)(b) A hybrid digital/physical workflow for combining human dexterity and machine repeatability in a robotic decorative plastering scenario, in which designers and craftspeople can modify the robotic trajectory through freehand sketch enabled by real time reality capturing techniques (Bard et al. 2016). (c) QUIPT, as a gesture-based control software, facilitates intuitive ways to program industrial robots using wearable markers and a motion capture system (Gannon 2014).

Human-robot interaction (HRI) can contribute to and enhance various creative practices by fostering collaboration between humans and robots. How machine fabrication can enhance human dexterity and creativity has long been a research topic in the domain of computational design. From a pedagogical view, assisted automation (Cardoso, Bidgoli, and Darbari 2017) offers a conceptual framework and concrete strategies to creatively incorporate robots into design education, drawing from a constructionist perspective. Alternative to the perspective that robots and numerically controlled machines can serve as "perfect slaves" for dirty work

Chapter 2. Literature Review

with materials, while designers can undertake the creative work. The term “assisted automation” and “robotic embodiment” highlighted “situated and embodied aspects of designing with robotic machines”. (Cardoso, Bidgoli, and Darbari 2017). Within the context of design and fabrication, reality is interface (Bard et al. 2016) explores how human dexterity and craft can be “the precision and repeatability of digital design and fabrication tools.” Sensing and visualization technologies, including live motion capture of gesture and projection mapping, are integrated into the robotic fabrication process to combine human skills with computational tools for creating a virtual-physical hybrid workflow in design and construction process such as decorative plastering and vault construction. Additionally, from a technological perspective, customized tools and software such as QUIPT (Gannon 2014) enable gesture-based human-robot interaction for creative applications, including fabrication, hybrid art, and tele-presence robots (Figure 2.4).

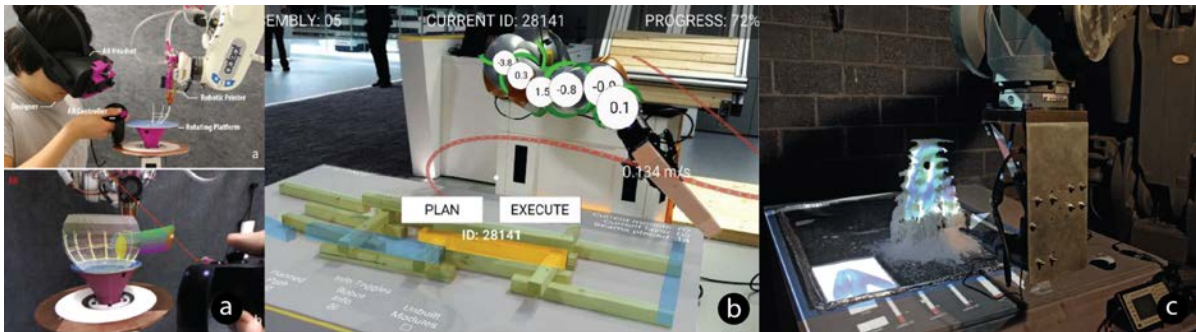


Figure 2.5: AR systems for HRI in design and making: (a) AR system enable the designer to view the predictive design feature on-the-fly in the physical environment. As soon as a design feature is completed, the robotic arm prints the new feature on-site (Peng et al. 2018). (b) AR system visualizes a spatial user interface (UI) with buttons for selecting, planning, and executing robotic instructions as well as overlays diagnostic feedback data on to the physical site (Kyjanek et al. 2019). (c) A projection-based AR system visualizes material properties such as thermodynamics for informed robot control (Johns 2014).

Since the 2010s, multiple studies have explored how spatial interfaces can facilitate effective collaboration between humans and robots and support in-situ design and fabrication tasks by providing faster and more intuitive robot programming than conventional techniques (Figure 2.5).

From a design and making perspective, spatial interfaces are applied to support a wide range of applications and creative practices by (1) contextual display of digital information

Chapter 2. Literature Review

and (2) intuitive interaction with virtual content (3) multi-stakeholder collaboration and information exchange

At the intersection of art, architecture, and robotics, several studies have explored using spatial interfaces for contextual display and intuitive interaction. For instance, studies such as (Weichel et al. 2014 ; Betti, Aziz, and Ron 2019 ; Kyjanek et al. 2019) explore how the designer can affect real-time robot-led fabrication processes. Spatial interfaces enable designers to modify toolpaths through intuitive approaches such as tapping and dragging. Alternative to a “file-to-factory” process, the operation is processed in a ‘byte to robot’ manner. The interruptible process allows the integration of human intuition and provides the potential to “bridge the gap between human sensibilities and material properties in the design process, thus ushering in a new kind of craft that is both materially responsive and “highly informed” (Johns 2014) (Figure 2.5). Studies including (Felbrich et al. 2018; Johns 2013a; Fazel and A. Izadi 2018a) utilize a combination of intuitive interaction techniques and spatial interfaces for design empowerment in form creation and materialization. Non-expert users can use bodily movement, such as gestural control, to shape and interact with forms. Created forms and associated fabrication information can be previewed and altered using an augmented reality system. Individuals can use intuitive means to explore design space and fabricate forms from table scale (Peng et al. 2018) to human scale (Johns 2013b).

The design process in AEC can be inherently complex involving multiple stakeholders from heterogeneous backgrounds and disciplines. Comprehensive communication between experts can facilitate knowledge sharing and skill transfer for integrated and holistic solutions. Collaborative Virtual Environments (CVEs), as a collaborative tool, explore the topic of transforming the network of computers into navigable 3D spaces that support collaborative work and social play (Benford et al. 2001). Early CVEs provide a digital landscape for multiple users to interact with each other, especially when they are geographically separated (Churchill and Snowden 1998). CVEs provide a digital environment for bi-directional data flow between offsite and onsite users. As a potential technology for CVEs implementation, augmented reality can bridge the gap between working in real and virtual environment. AR facilitate geographically distributed teams in collaborative tasks by providing common ground and spatial awareness in a virtual or mixed environment. The configurations of AR-enabled collaborative design can be categorized into four types (Calixte and Leclercq 2017). The group co-located

and remote revaluation configuration describes a collective design process with co-participants in co-presence and participants at a distant location. A mixed-reality immersive collaboration system can support bi-directional communication between on-site and off-site users (Mitterberger, Angelaki, et al. 2022) based on a shared digital twin in CVE. For on-site users, the fabrication instruction and the updates in the design model from the remote end can be visualized on-site via AR devices. For off-site users, site information can be accessed to inform design decisions. AR systems can improve performance time and mental effort in collaborative design tasks (X. Wang and Dunston 2011).In addition to telepresence collaboration, AR has the potential to enhance co-located face-to-face collaboration by providing asynchronous virtual information,for example, based on a projection-based AR, the designers on-site can preview the design result of instructions of robots (Mitterberger, Jenny, et al. 2022a). In the context of 3D artifact creation, AR can provide an effective solution for collaboration. By taking advantage of the flexibility of virtual content creation in AR, the collaborators can speed up the iteration process (Guo et al. 2019). In addition, collocated spatial interaction naturally supports deictic referencing and gestures to assist coordination (Ens et al. 2019). AR for collaborative design can afford natural coordination such as relative spatial position between collaborators and other non-verbal cues. These spatial cues are only meaningful when they are presented in a shared physical space (W Chastine et al. 2007).

2.2.3 AR for HRI in Construction Robotics

For construction and architectural robotics, AR systems have been introduced into the process of human-robot interaction, data sharing, and instruction display. The application domain includes interactive fabrication, on-site programming, and the tele-operation of construction robots.

- Interactive Fabrication

Interactive fabrication refers to the integration of digital technologies and interactive processes in the fabrication or manufacturing of physical form (Willis et al. 2010).Interactive fabrication usually involves digital modeling, computer-aided design, and real-time feedback for a dynamic and responsive fabrication process. The integration of AR systems can enhance

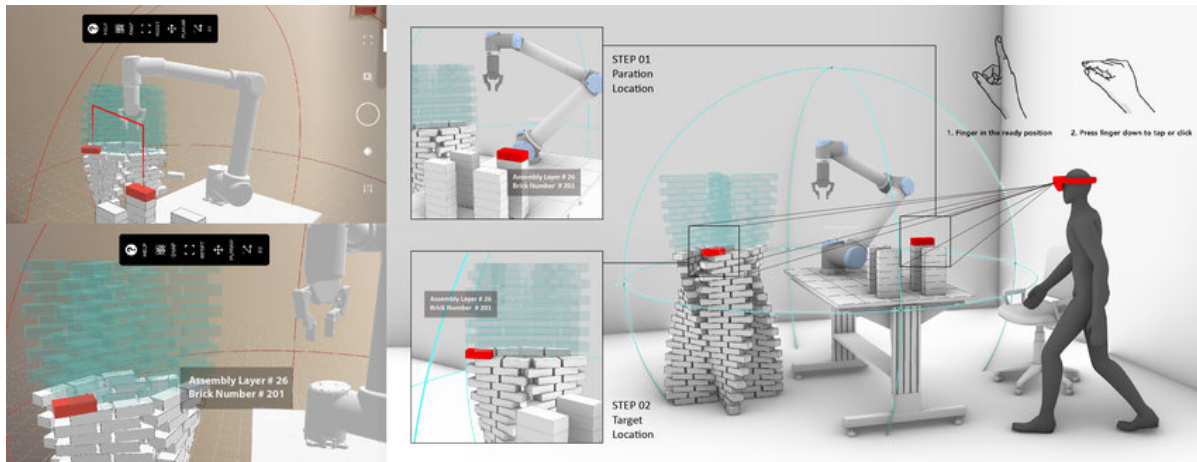


Figure 2.6: AR interface for robotic assembly and design prototyping with gesture input (Song, Koeck, and Luo 2021).

the information visualization of the existing digital fabrication or robotic systems. For example, a construction worker with an augmented reality head-mounted display can preview and plan the robotic trajectories, production sequencing, and diagnostic feedback in robotically-assisted timber construction (Ondrej Kyjanek et al. 2019). In addition, AR systems can augment physical objects with virtual information for on-site decision-making. The overlay of material properties over the inconsistent material in the dynamic construction process can enable manufacturers to assess the inaccuracies and mismatches between blueprints and physical entities in real-time (Jahn et al. 2020). In the interactive assembly process, AR tools can be used for visualizing the accurate location of components for assembly in real-world environments and assist multiple makers in collaboratively constructing complex forms (Fazel and A. Izadi 2018b, Song, Koeck, and Luo 2021) (Figure 2.6).

- On-site programming

Augmented reality as a spatial user interface can assist construction workers in controlling machines with multimodal interaction techniques. Gestures, voice, and many input modalities can be integrated into AR systems for high-level supervision and control. Especially for unskilled construction workers, the flexibility and intuitiveness of robot operation afforded by AR systems can lower the entry barriers and increase control efficiency. For example, in construction tasks such as assembly, AR-enabled interactive programming can assist operators in on-site path planning (Fang, S. K. Ong, and A. Y. C. Nee 2012b), virtual simulation

Chapter 2. Literature Review

(Ostanin, Yagfarov, and Klimchik 2019), and instructing robots via programming by demonstration (Soares, Petry, and Moreira 2021). For construction processes involving complex and soft materials such as concrete, plaster, and clay, a streamlined, well-defined automation process can be difficult. The material behavior of malleable materials can be difficult to simulate. A craft-specific user interface for robotic processes in construction can potentially leverage both machine precision and human tacit knowledge. For example, Interactive Robotic Plastering system (IRoP) systems enable the designers to interactively modify the robotic spray path through a controller-based interaction system and an augmented reality interface. The gestural input and an audio-visual AR interface demonstrated a high potential for enhancing interaction intuitiveness for construction workers and designers to perform “challenging-to-automate” processes on-site (Mitterberger, Jenny, et al. 2022b).

- Tele-operation

As part of construction robotics, tele-operated robots can assist construction workers in tasks performed in hazardous and inaccessible environments such as disaster sites and mines. A human operator can manipulate construction machines at a remote place through a supervisory control system with visual feedback based on a camera-display combination. Tele-operation shows a wide potential for tasks operated in dynamic and complex environments where human operators can be in the feedback loop and intervene in the decision-making process under uncertainties.

In robotic tele-operation tasks, AR systems can visualize non-graphical and invisible information to assist remote operators in construction machine operation (X. Wang 2008a). In maintenance tasks, the integration of AR into a robotic system can display the information of non-invasive sensors over the structure and guide the remote robot operator to the location of operation. For example, an AR-based equipment maintenance system can be used to present temperature variations and distributions for the maintenance of heating pipes (Klinker, Stricker, and Reiners 2001).

In tele-operated excavation tasks between the operator and the remote excavator, an interface is required for the operator to understand the environment where machines interact in the task (X. Wang 2008b). An intuitive interface can provide the operator with an onboarding view to enhance situational awareness (Zhou, Q. Zhu, and Du 2020). As a 3D visual system, AR and VR interfaces can transmit visual feedback and sensor information from the site to

Chapter 2. Literature Review

the remote operator. The data acquired from sensors on excavators, such as 3D point cloud data and camera views, can be visualized to assist remote decision-making. For example, the AR interface can synthesize geo-spatial data and construction informatics to improve urban excavation safety by visualizing and monitoring the proximity between invisible utilities and digging implements (X. Su et al. 2013). The efficacy of the vision-based interface can be circumscribed by the limitations of the remote video stream's field of view (FoV). The FoV defines the spatial range within which safe exploration can be conducted (X. Wang and Dunston 2006). The limited FoV can result in incomplete environmental information, such as soil, machines, and human workers, negatively impacting the manipulation efficiency (Masaru Ito et al. 2021).

In addition to vision, supplemental modalities can be integrated into AR-based tele-operation to improve the operator's understanding of the environment. Tactile feedback, such as high-frequency vibration, can be used to show the remote operator of the potential collisions. The tactile signals can enhance the operator's understanding of the field situation. For instance, tactile information can represent the material characteristics on-site for operators to perceive and recognize objects on job sites (Okamura, Cutkosky, and Dennerlein 2001). Auditory information can be supplementary to visual interfaces to present information such as alarms and warnings. When visual information is overloaded, the auditory can serve as an independent channel to notify the operator when there is a potential danger or emergency. The integration of auditory information into an AR-based tele-operation system needs to take the environmental noise into the consideration. The hearing protective equipment can potentially affect the efficiency of auditory interfaces (J. S. Lee et al. 2022).

In this section, I outlined how AR systems can facilitate HRI in the construction domain, from on-site application to teleoperation. At various stages of construction projects, AR interface can enable stakeholders to operate robots with enhanced efficiency, safety, and accuracy. However, several factors need to be considered for integrating AR technologies into construction practices. As AR is still in its infancy, experts and technicians may still be required to interface AR systems with existing construction software ecosystems, such as Building Information Modeling (BIM) software, CAD files, and project management platforms. The introduction of AR may bring changes to the existing well-established workflow across diverse stakeholders, including architects, engineers, and on-site workers. AR integration may encounter some resis-

tance due to the increased cost and the lack of familiarity of novel technology. Moreover, the working environments of jobsite can be more challenging than other industrial applications. Harsh conditions of job sites may require more consideration of durability in the hardware design of AR devices such as headsets and handheld devices. The battery duration can limit the AR application in outdoor job sites. In addition, the dynamic and complex construction sites can affect the robustness and accuracy of localization and virtual-physical registration. In Chapter 4, I described a multimodal localization system for AR systems operated in challenging lighting conditions. Finally, regulatory compliance, such as safety regulations and liability issues, may pose challenges to the deployment and use of AR for HRI systems in construction practices.

2.3 Cultural and Historical Background of Spatial Interface

In the section Appendix, I traced the intellectual lineage of the tight linkage between computation and space from the lens of architecture and computer science. The cultural and historical background provides fertile ground for reconnecting the concept of spatial interface with precedent academic inquiries. Within the domain of architecture, I draw from the early experiments in architectural cybernetics and discuss the concepts of homeostasis, autopoiesis, circularity, and conversation theory. Following this thread, I define spatial interfaces as an augmentation of the built environment, facilitating the interaction between humans and their inhabited space. Similarly, from the viewpoint of computer science, I reflect on the concept of spatial turn in human-computer interaction (HCI), a shift in focus from the experience with two-dimensional space towards considering spatial aspects of interaction. The spatial turn emphasizes the significance of physical space, location, movement, and embodiment in human-computer interaction. This shift expands the scope of HCI to include a broader range of contexts, such as immersive environments and tangible interaction. The concept of spatial interface can be considered as part of the spatial turn, with a special focus on the technology and experience of combining virtual objects with real-world surroundings. Explorations within the field of architecture and computer science directed me to expand the term spatial interface from a display technology to a technological mediation. This broadened definition

guided me to develop a new technical framework emphasizing virtual-physical registration and multimodal interaction, which I will elaborate on in Section 4 - Method.

2.4 Relevance

The examples discussed above shows how spatial interfaces could be applied to multiple domains where spatial awareness is necessary and important when interacting with machines, such as education (Quarles et al. 2008 ; Baykal et al. 2018; A. Wu et al. 2011), navigation (Boari and Fraser 2009), driving and orienting vehicles (Lajoie 2003), and leisure activities (e.g. solving puzzles (Antle 2013)).

In this thesis, I examine the potential of leveraging spatial interfaces for human-robot interaction in the architectural and, more specifically, construction domains.

The application domain of this research focuses on human-robot interaction (HRI) in robotically-supported construction. Specifically, I will use robotically-assisted quality assessment and management as a use case and evaluate how spatial interfaces such as augmented reality may assist inspectors in robotic manipulation and task allocation.

The motivation is rooted in the tension between the existing human-robot interface inherited from industrial automation and the unique challenges of HRI in architectural practices. The mismatch between the affordance of imported HRI interfaces and the sectoral context of design practices manifests itself from three aspects: (1) tasks, (2) environments, and (3) organizations.

- Tasks — Enactive actions

In the 1960s, early attempts were made to integrate robotics into the production of architectural components driven by safety, efficiency, and cost factors. In the early 2000s, architects and designers started to reappropriate robotics for design practices, shifting from a robot-centered paradigm to a human-centered paradigm (Bock 2007). Compared to the manufacturing tasks that are mostly predetermined, tasks in design practices rely on context-dependent constructs between humans, robots, and space, involving multifacetedness of human embodied experience. However, the existing interface for HRI mainly supports symbolic and iconic inter-

action³. The mutual construction between action and world can be limited as the handling of enactive knowledge still remains absent in the existing HRI interface. For example, the existing interface between human and robot often rely on predefined actions and responses programmed into the robot's behavior, Robots may not be able to dynamically adjust the actions based on evolving contexts or user input. The lack of sensory perception and unimodal interaction based on existing human-robot interfaces make it difficult to support enactive interactions. AR systems that support multimodal input and output can be applied to facilitate embodied HRI and align robotic technologies with the constitutive role in digitally mediated architectural practices.

- Environments — Spatial Awareness

In industrial automation, robots are designed and programmed for relatively static environments, running in work cells free from people and other interferences (Heyer 2010). Once programmed, robots are expected to execute predefined tasks repeatedly, isolated from external environments. However, environments for architectural practices are unstructured and undetermined. The physical constraints can dynamically change according to ongoing tasks, and these changes can in turn reconfigure the complex relation between humans and robots. Actions of robots or humans emerge from a hybrid mix of agencies that are embedded in coherent frames of action. Various physical connections are established through improvisation in response to in-situ conditions (Afsari et al. 2018). By integrating these contingent physical constraints into the interface, spatial interfaces can potentially shift HRI from a seamless batch-processing paradigm inherited from industrial automation to a situated, seamful “assisted automation” paradigm (Cardoso, Bidgoli, and Darbari 2017) that reveals the interdependence between humans, robot, and their co-inhabited space.

- Organizations — Communication and collaboration

From the perspective of organization, in contrast to serial products, construction processes can be considered as complex systems involving conditional decision-making between multiple

³(Bruner 1966) describes the three possible types of knowledge used when interacting with the world as symbolic, iconic and enactive. Symbolic knowledge involves conceptualization and abstract reasoning, iconic knowledge involves visual recognition and the ability to compare and contrast, and enactive knowledge is constructed on motor skills.

Chapter 2. Literature Review

stakeholders (Kohler, Gramazio, and Willmann 2014). Traditional robotic fabrication pipelines and interfaces assume a linear ‘file to factory’ process, which involves a predetermined batch of commands. The training and skills requirements for operating robots are relatively higher compared to conventional tasks (A. B. Moniz and Krings 2016a). The streamlined hierarchical structure and high skill requirements can lead to deskilling and communication barriers after integrating robotic technologies⁴. An intuitive human-robot interface can potentially enable users with limited prior knowledge of robotics to live program robots using well-entrenched spatial skills or natural language. The flexibility afforded by novel spatial interfaces could improve communication flows between different working groups and technical staff, which may support the adoption of robotics in established organizational configurations of architectural practices.

Within the broader context of computational design, the framework developed in this work can be extended to bridge the gap between digital design and physical prototyping, with a connection with research domains such as interactive fabrication and tangible computing. In Chapter 2.2.3, I briefly outlined the precedent works in the domain of interactive fabrication. This work may contribute to interactive fabrication by providing a novel interface for accessing digital information in the fabrication process. The AR system can provide an immersive design interface dynamically responding to hand gestures, material behaviors, and virtual simulations, creating a real-time feedback loop between the virtual design space and the physical world (Hahm et al. 2019 ; Song 2020). Moreover, the module for reality capturing in the AR system can integrate the physical properties of materials into the interactive fabrication process and augment the virtual 3D models with features acquired from the physical environments. Finally, The AR system can provide contextual and tangible descriptions of guiding information for on-site fabrication. For example, on-site builders with AR visual guidance can improve

⁴The speculation on the destabilization of the boundary between design and construction and seamless collaboration enabled by building information modelling (BIM) can be interfered with the conflicts between the structure of design organizations, established workflows and the affordance of technical systems, for instance, (Cardoso 2019) traces, visualizes and reflects on the socio-technical construction among “design ecologies” including technical systems, design collectives and their inter-relationships. Similarly, robotics as a tools for digital fabrication and construction have been increasingly adopted in architectural practices. However, according the author’s knowledge, few studies have been conducted to investigate how the introduction of a robotic system, a complex social-technical system, may reconfigure the existing work pattern, organization structure and professional roles in the AEC. Previous works (A. B. Moniz and Krings 2016b ; A. Moniz and Krings 2014; Blajer and Krawczyk 2002) examine the use of new robotic systems in adjacent domains such as manufacturing industry with social dimension.

the performance of manual operations and interactively build complex geometry with reduced errors. Within the domain of tangible computing, the AR system developed in this work can potentially enable natural interaction and collaboration in context-aware applications. For example, the semantic understanding feature of the AR interface can be used for dynamic context analysis and provide personalized visualization according to the situated environment and ongoing task. The virtual information overlaid onto the physical objects can enhance the sensory feedback in tangible interactions with additional visual and auditory information. The AR interface can provide a shared space where multiple participants can interact with both physical and virtual content in a collaborative manner.

2.5 Potential Areas for Future Research

Based on the aforementioned literature review, I identify several areas for future research within the domain of AR for HRI in construction.

- Task-oriented interaction techniques can be integrated for construction processes

Currently, the most prevalent input modalities for interactions in mixed reality are touchscreen and inertial measurement units (Z. Chen et al. 2017). Other modalities such as speech and gesture can provide users with increased bandwidth, efficiency, and robustness in human-robot interaction, especially in construction scenarios. However, according to the author's knowledge, no previous works have been presented on multi-modal spatial interfaces for human-robot interaction within the AEC domain. The technical challenges lie in the mapping between speech, digital content, and the physical domain of the interaction. To achieve the same level of situated natural language understanding and grounding still remains challenging for HRI in general and HRI in the AEC in particular. In this study, two potential solutions for spatial HRI interface in the AEC will be explored: (1) a domain-specific natural language understanding module that leverages specific domain knowledge in a specific construction task such as surface finishing to support more intuitive and efficient robot programming. (2) a context-aware module that leverages contextual information to disambiguate human operators' commands in real time.

- High-level virtual-physical registration can be established to enhance the context awareness of AR systems in robotically-assisted construction

Chapter 2. Literature Review

Due to a lack of connection between the digital and the physical environments, the existing spatial interface provides limited support for in-situ robotic tasks that require on-the-fly robot programming in adaptive to dynamic environments. Projects such as (Fang, S. K. Ong, and A. Y. C. Nee 2012a ; Quintero et al. 2018) can visualize predefined and pre-computed robot trajectories in a physical environment. However, the interdependency between digital information and the physical environment is absent in these systems. As a result, spatial interfaces decontextualize the robot control from real-world activities. The decontextualized setting can work for repetitive tasks in a static environment; however, in architectural practices, robots need to dynamically adapt to the changing environment, which may require in-situ authorization from the operator. For instance, in the surface finishing task, robots assist humans in applying malleable materials such as plaster to a physical surface with bespoke geometry. The robot task and trajectory planning are conducted in an on-the-fly mode adapting to the diverse and changing surface quality. The operator, thus, needs an intuitive HRI interface to visualize the real-time information of the changing environment and authorize in-situ robot path planning. The integration of spatial awareness into the existing spatial interface can facilitate situated information visualization and bridge the communication gap between users and robots.

- Multi-stakeholder collaboration can be further explored in AR for HRI research in the application domain with organizational complexity

Spatial interfaces such as augmented reality combine virtual and physical objects, specifically in architectural practices; the virtual design blueprint and real construction site can be combined and illustrated to designers, engineers, and workers in real time (Ahmed 2018). Being able to see both real-world and fellow operators during AR sessions makes AR an ideal medium to aid collaboration and facilitate more natural communication patterns (Kiyokawa et al. 2002). (Billinghurst and Kato 2002) provides a detailed explanation of how both the physical environment and tangible objects within it support collaboration during spatially referenced work activities, such as design and planning. Real objects used as semantic representations or spatial reference frames can support natural communication behaviors. However, most of the existing AR for HRI systems focus on the communication between robots and a single operator. The potential of AR as an interface to mediate face-to-face collaboration between co-located and remote users is still not fully explored yet. This study proposes to use

Chapter 2. Literature Review

physical space as a shared spatial reference for intuitive 3D computer-supported collaborative work. Users can fully leverage the physical environment augmented by a spatial interface for information exchange and communication.

Chapter 3

Hypothesis

My hypothesis is that the integration of AR systems can potentially provide robot operators with an embodied interface that may facilitate in-situ robot control and supervision in robotically-assisted construction practices. The first use case involved students without much prior experience in robotic manipulation. The system functioned as an instructional tool, enabling students to supervise and control the motion of the robot for construction-related tasks, such as picking and placing objects. The second use case documented an interface for building inspectors, exploring its capacity to enable accurate robot operation for the quantitative evaluation of building component quality.

Compared to existing two-dimensional interfaces, the integration of AR systems may support effective interaction by (1) improving information readability by contextualizing virtual content to its physical referents and places. For example, inspectors with mobile AR devices can assign tasks to robots on-site or on-factory and interactively view the quantified quality assessment result registered to a building component in the real environment. Compared to traditional HRI interfaces such as teachpendant, the in-situ visualization of geo-located information may enhance the interaction efficacy by increasing virtual information readability (2) increasing intuitiveness of human-robot interaction by supporting operator's task-specific spatial skills, for instance, in robotically-assisted quality assessment task, inspectors can use their well-entrenched skills such as annotations to define the area of interest for inspection, compared to the existing robotic control mechanisms through joysticks, the interaction technique adaptive to the inspector's internalized and embodied skills may enhance the intuitiveness of

Chapter 3. Hypothesis

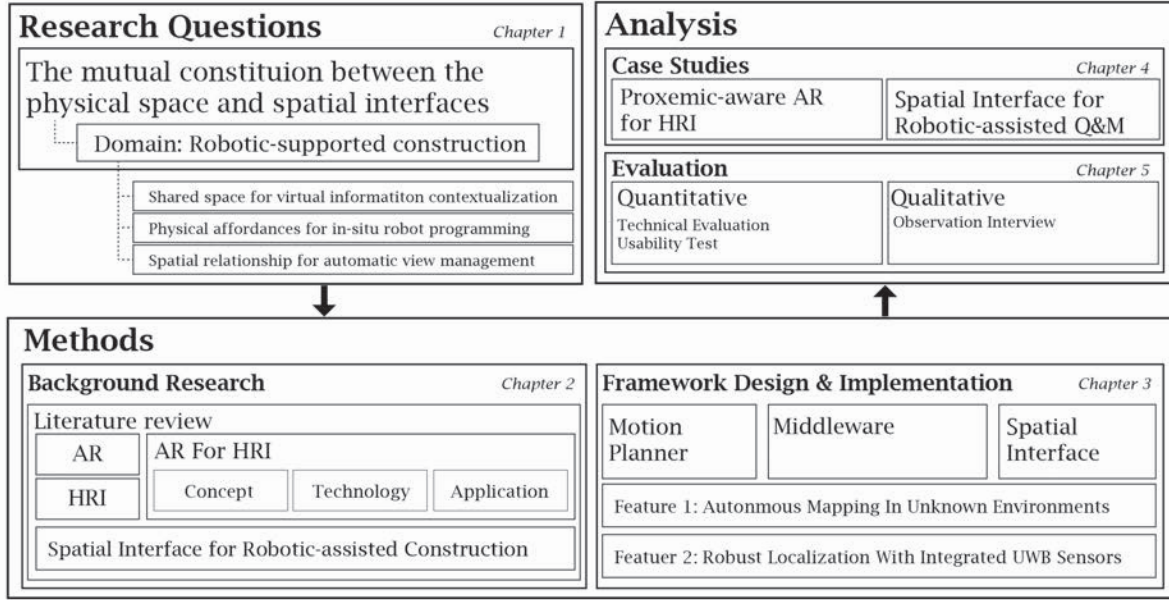


Figure 3.1: Research framework consisting of three research phases and their corresponding research elements. Image credit: Author.

the robotics manipulation process. (3) reducing the cognitive load by leveraging human-robot spatial relationships for adaptive view management. For example, in the first case, the AR system serves as an instruction tool for students operating robots for construction tasks. However, the work environment can be cluttered with obstacle obstructions in the interface view. I proposed a proxemic-aware view management system for displaying information at different levels of details adaptive to human-robot spatial relationships, with a hypothesis of reducing information overload in AR for robotically-assisted construction systems.

To validate the hypothesis, the study is conducted in three phases, including framework design, feature implementation, and case study & evaluation (Figure 3.1). The first phase examines the research background and contextualized the research questions in robotically-assisted construction practices. The domain challenges may uncover open opportunities in the design space.

The second phase, involving framework design and feature implementation, documents the process of designing and implementing a functional Augmented Reality for Human-Robot Interaction (AR for HRI) system guided by the framework. This framework acts as a high-

Chapter 3. Hypothesis

level conceptual structure to develop new AR for HRI systems, encapsulating core technical components. An industrial robot, ABB IRB-6640 mounted on a linear track, served as the experimental platform for feature implementation and tests.

The third phase, evaluation, documents two case studies focusing on the application of spatial interfaces in robotically-assisted construction practices. The study was conducted in both quantitative and qualitative manners to assess how the developed framework might support intuitive human-robot interaction. The evaluation method consists of technical evaluation, participant experiments, and qualitative interviews. The technical evaluation quantitatively benchmarks the developed models and algorithms in terms of accuracy, efficiency, and robustness. Participant experiments assess the usability of the developed system through A/B tests, while qualitative interviews gather contextual knowledge from the participants. The collection of experimental experiences may contribute to a broader understanding of expert knowledge that extends beyond the technical data. The implicit dimension of expert knowledge can shed light on the practical application of the developed framework in robotically-supported construction.

Chapter 4

Method

The Method chapter introduces the implementation of a domain-specific spatial computing framework for robotically-assisted construction. The framework is designed and implemented to address specific challenges within the context of robotically-assisted construction, including site information acquisition in unknown environments, robust device localization invariant to visual interference, user interface view management for cluttered surroundings, etc. The framework not only offers a robust solution to these challenges, but also provides an extensible template for integrating spatial interfaces into the domain of human-robot interaction for robotically-assisted construction tasks.

Compared to multiple general frameworks that laid the foundational groundwork for subsequent research in the domain of augmented reality (AR) within the context of HRI (Bejczy, W. S. Kim, and Venema 1990c, T B Sheridan 1987, Buzan 1987, Noyes 1984, Milgram, Zhai, et al. 1993b), this chapter documents a domain-specific framework tailored to robotically-assisted construction domain. Several distinct challenges can inform the design of the framework. Firstly, in most construction tasks, spatial information about the work environment can be limited or entirely inaccessible. Unlike domains such as manufacturing, medicine, or other indoor applications where augmented reality (AR) systems typically operate within well-defined and static environments, construction tasks are usually carried out in spatially unknown and dynamically changing settings. Consequently, a digital representation of the surroundings becomes crucial for both robots and AR devices.

Chapter 4. Method

The real-time representation of the surroundings is vital for robots to optimize paths for efficiency and simultaneously avoid obstacles in construction tasks. Additionally, based on a digital representation of the real environment, AR systems can ensure that the virtual information presented is coherent and consistent with the physical surroundings.

Secondly, the localization of spatial devices can be error-prone within a complex work environment. Environmental factors in construction tasks, such as inconsistent lighting and dynamic changes, can pose challenges to the localization of spatial devices. For example, construction sites operating during daylight hours can be affected by different lighting conditions throughout the day. Moving objects on jobsites can cast dynamic shadows and create inconsistencies in lighting, introducing inaccuracies into vision-based localization systems. The cumulative localization inaccuracy can result in errors in robotic navigation, misalignment in virtual-physical registration, and inconsistencies in human-robot interactions.

Thirdly, the cluttered construction environment can significantly impact the efficacy of AR systems. In densely packed or obstructed construction settings, alignment and registration between virtual information and physical referents can be affected, leading to disorientation and distraction for users. The superimposition of virtual information in such cluttered scenarios may add extra cognitive load to the operator, negatively impacting the operator's interaction with the AR for HRI system.

Informed by the challenges, I designed a AR for HRI framework tailored to construction robotics. The framework consists of (1) a motion planner module for robots to autonomously explore and digitize unknown environments, (2) a middleware for device localization and data conversion between raw point clouds and high-level semantics, and (3) a spatial interface for spatial interaction and automatic view management(Figure 4.1).

The motion planning module generates near-optimal traversal paths for robots to explore the surrounding environment while simultaneously constructing a comprehensive multi-resolution map of the site. The component integrates frontier-based and information-based methods to maximize coverage. The details of the implementation will be elaborated in the section 4.1.

The middleware module is designed for aligning virtual information with its corresponding physical objects. This alignment process is carried out through multi-level strategy. Firstly, at the level of location-aware registration, the module provides localization of spatial devices, maintaining the accurate alignment of virtual information within the operator's field of view.

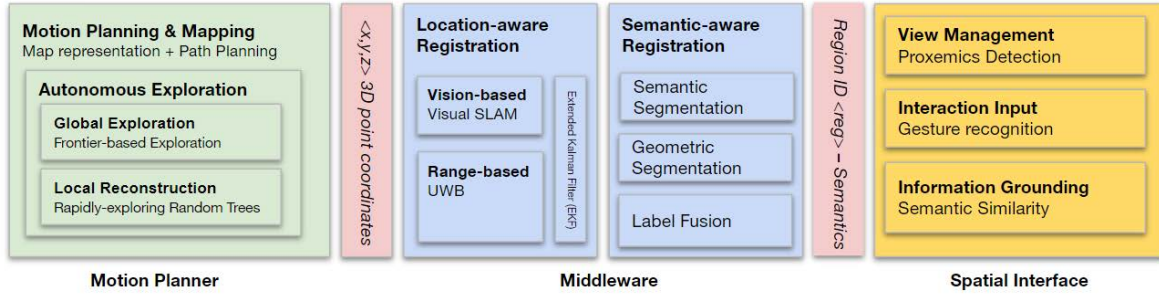


Figure 4.1: The high-level framework design of a domain-specific Augmented Reality (AR) system in the context of Human-Robot Interaction (HRI), including essential components such as a motion planner, middleware, and spatial interface. The framework can be extended and implemented for various human-robot interaction tasks, particularly those relevant to construction-related scenarios. Image credit: Author.

Secondly, at the semantic-aware registration level, the module maps the point cloud acquired by the laser scanner to high-level semantic data. This semantic knowledge empowers the system to meaningfully align virtual elements with the physical world. For instance, the module can intelligently categorize objects into distinct classes such as chairs, tables, floors, and doors. Subsequently, virtual information related to each class can be seamlessly juxtaposed with its corresponding physical referent, facilitating context-aware interaction.

The process of mapping point clouds to semantic data is commonly referred to as "scene understanding" within the domains of computer vision and robotics. A large number of techniques, primarily leveraging machine learning approaches, have been developed to automate the processing of 3D point cloud data. To enhance generalizability, most learning models focus on the end-to-end pipeline that takes raw point clouds as input and creates class labels in semantic format for point clouds as output. However, in construction applications, the environment can be cluttered which can result in noises and occlusions in the site scanning. This particular challenge is exacerbated when individual points are employed as the input for these techniques, potentially resulting in over-segmentation and inaccuracies. Informed by these challenges, this study introduces surfels (planar surface patches) as the foundational representation for the scene understanding model.

The selection of surfels is motivated by several considerations: (1) point clouds sharing the same planar surface patches in indoor environments likely belong to the same object (2) Planar surface patches can be used to filter out the noise and rectify over-segment point clouds (3) The

Chapter 4. Method

targets of surface inspection in architectural practices are usually large planar surface patches. The surfel-based representation demonstrates superior performance in terms of accuracy and intersection over union (IoU) when compared to alternative methods.

The third module, spatial interface module, is developed to superimpose virtual information onto the physical world. With a specialized focus on cluttered construction environments, this module incorporates a proxemic-aware view management feature. This feature has been engineered to automatically adjust the level of detail in the presentation of virtual information, contingent upon the spatial relationship between the operator and the robots. In alignment with the findings derived from the case study presented in Chapter 5, the implementation of proxemic-aware view management features yields a notable reduction in the operator's cognitive load and enhancing the efficiency of Augmented Reality (AR) for Human-Robot Interaction (HRI) systems.

With the aforementioned optimizations in localization, site mapping, and semantic segmentation, the framework can provide a virtual representation of physical assets and an interoperable ecosystem for heterogeneous hardware and software. The design of the technical framework is informed by the theoretical discussions presented in the section Appendix. Rather than solely focusing on enhancing the display performance of AR interfaces, this work emphasizes the mutual constitution between augmented reality and physical space. Physical environments are not abstract, nor is digital information placeless; they are integrated through a middleware that registers virtual content to physical referents and augments the physical world with digital information. The vision here is to leverage computation to augment the existing human-world relationship and protocols, thereby recombining the physical and virtual worlds into a hybrid continuum.

In this study, the framework is implemented using an industrial robot (ABB IRB 6640) as a testbed to illustrate and evaluate the accuracy, effectiveness, and robustness of each module. Given the repeatability and precision of industrial robots, it can be assumed that the errors resulting from the robot platform are controlled and consistent across the experiments and performance evaluation of each module. The framework is developed in a modular manner, leveraging semantics, a data format that is both human-readable and machine-processable, for information exchange between different systems and stakeholders. This framework has the potential to be extended to other mechanical systems in future construction applications, serv-

ing as a shared interface between human workers and machines. For example, an AR interface could visualize the operations of multiple equipment, such as hoists, cranes, and linear tracks, during material delivery tasks. By overlaying task progress and real-time site conditions onto the physical environment, operators can assign tasks to each equipment with enhanced situational awareness. However, when applying this framework to other robotic systems such as unmanned aerial vehicles and quadruped robots, additional testing and refinement are necessary to validate the framework's performance beyond the domain of industrial robotics.

4.1 Motion Planner for Site Reconstruction

The motion planner module generates a sequence of actions and a feasible path for the robot to explore and map an unknown environment autonomously. The module constructs the unknown site through three stages. At the first stage, the environment is unknown, the component returns the exploration path by searching and identifying the frontier points between free and unknown parts of the map. At the second stage, a robot operator can define a target volume within the partially known environment for detailed reconstruction, the component plans a next-best view based on the information gain of each sampled view candidate. At the third stage, a robot operator can define an area of interests within the known environment for detailed reconstruction. The component generates an efficient and collision-free path by minimizing the cost defined by the total traveling distance.

4.1.1 Global Exploration

At the global exploration phase, the robot explores the space by finding a set of poses along the frontier of the unknown environment and creates the occupied space map (a volumetric representation of space in a hierarchical structure). The robot first looks for a frontier located on the boundary between the explored and the unknown part of the environment, and then choose the one with the maximum information gain weighted by its cost to reach (C. Zhu et al. 2015a, Lu, Redondo, and Campoy 2020a), which can be defined as:

$$I_f = \frac{N_0}{N} \cdot \frac{1}{|P_f - P_0|}$$

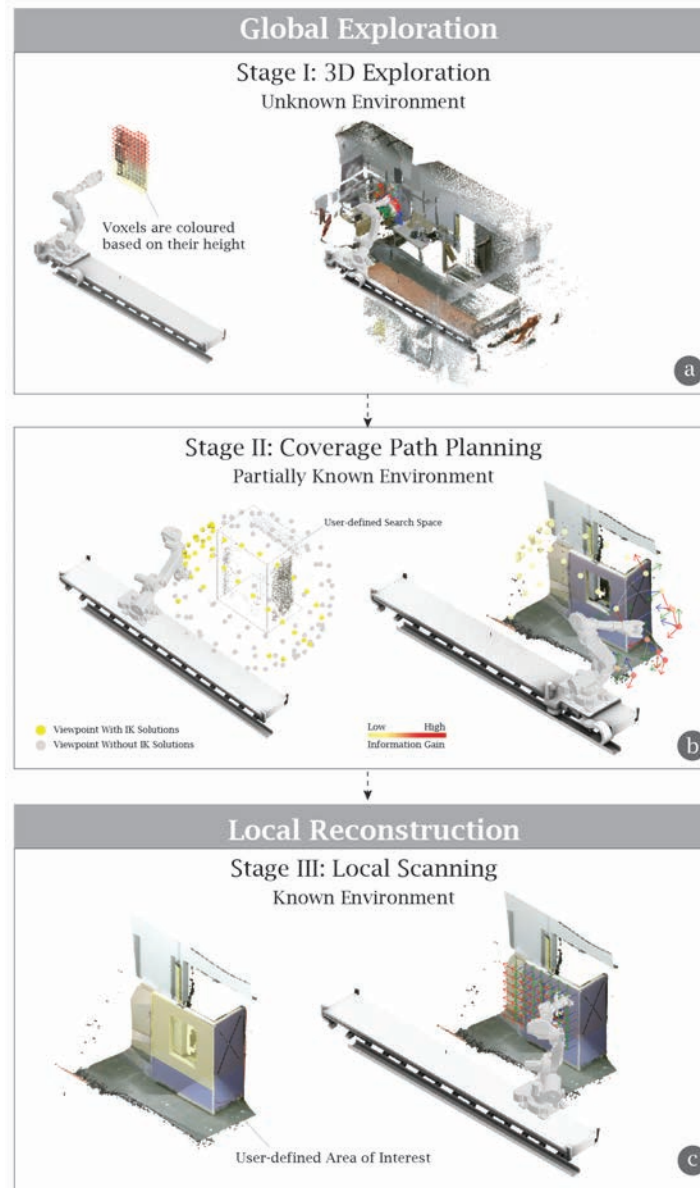


Figure 4.2: Two phases of autonomous exploration —global exploration and local reconstruction. The global exploration phase involves mapping the surrounding area through (a) 3D exploration and (b) coverage path planning, while the local reconstruction phase captures the user-defined area at a fine resolution. In this study, I used an industrial robot platform (ABB IRB 6640) to evaluate the motion planning module for site mapping, industrial robots present reliable performance in accuracy and repeatability. However, the range and mobility of industrial robots can be confined due to factors such as singularity and a fixed number of Degree of Freedom (DoF). The integration of mobile platforms can potentially increase the work range with increased flexibility. Image credit: Author.

Where I_f is the information gain of a candidate frontier cell. P_f and P_0 represent the coordinate of the candidate frontier cell and the current pose respectively. N_0 represents the unknown voxels within a sphere of the radius R around the candidate frontier that contains R voxels.

After the goal frontier is selected, the planner generates a path to move the robot 30 mm to the frontier and look towards the direction of the frontier through an Inverse Kinematic (IK) solver unless the planner can not find a valid IK solution. The exploration procedure terminates when the void voxels within the bounded space are classified as either free or occupied.

The 3D exploration phase reconstructs a partially known scene model of the site (93.67% coverage rate) at a relatively low resolution (50 points per m^3). Based on this model, the remote inspector can define a smaller search space for the second phase of coverage path planning. The coverage path planning aims at automatically generating collision-free robotic paths covering the required surface area within the target volume. I first combine the voxels in the volume into a binary status — “occupied” and “unoccupied” (Vasquez-Gomez et al. 2014a, Blaer and Allen 2007a). The system then uniformly samples a set of viewpoint candidates around a sphere to capture the most perspectives that can cover the target volume (Isler et al. 2016a). The quality of each view candidate is defined by measuring the mean entropy $e(x)$ over all voxels within the sensor frustum of a new pose, which can be defined as (Daudelin and Campbell 2017a):

$$e(x) = -\frac{1}{k} \sum_{i=1}^k p_i(x) \log p_i(x) + p_i(x^c) \log p_i(x^c)$$

Where $p_i(x)$ is the probability of voxel x being occupied, and $p_i(x^c)$ denotes the complement probability of $p_i(x)$, i.e. $p_i(x^c) = 1 - p_i(x)$. Based on this volumetric function, I select 10 candidates with the highest visible uncertainty for coverage path planning.

4.1.2 Local Reconstruction

After the global exploration phase, I use a region-growing approach based on smoothness constraints to cluster point clouds into surface patches. The robot operator can choose a specific surface patch for secondary scanning to build a high-fidelity reconstruction of the physical object.

Chapter 4. Method

Region growing methods use local features extracted from a neighborhood area around a selected point to aggregate nearby points with similar properties and segment a region of a point cloud (Paul J. Besl and Jain 1988). The goal is to identify distinct regions or segments within the point cloud that exhibit similar properties, such as color, intensity, or surface normals. This method is commonly used in computer vision, remote sensing, and 3D point cloud analysis.

Point normal is commonly used as a metric or criteria for region-growing methods. Specifically for the job site, the scanning of civil infrastructures contains both planar and curvilinear objects. The complex geometry of infrastructure, such as pipes and wires, can result in inaccurate estimation of point normals and further lead to errors in segmentation (Yoon et al. 2007). In this study, I built upon regression methods for normal estimation. Regression methods estimate normal with the whole neighborhood centered at the sampled point, which is commonly used in the normal estimation (Cazals and Pouget 2005, Gaël Guennebaud and Gross 2007). Algorithms such as Moving Least Squares (MLS) (Alexa et al. 2001), Local Kernel Regression (LKR) (Öztireli, Gael Guennebaud, and Gross 2009), and Voronoi PCA (Alliez et al. 2007) improved the performance of normal estimation by retrieving sharp features. Informed by the previous works, I implemented a novel normal estimation method for the automated segmentation of both planar and non-planar surfaces in point clouds (Khaloo and Lattanzi 2017a).

The method starts by classifying the point cloud into edge and non-edge sets. The point with sparse surrounding points in the neighborhood within smooth areas tends to be mislabeled as edge points, potentially negatively impacting the accuracy and efficiency of normal estimation algorithms. To improve the accuracy and robustness of edge point detection, I used the standard deviation σ_{ni} of point normal within the point's K nearest neighbors to identify edge points after calculating the initial normal of each point P_i through PCA.

$$\sigma_{ni} = \left| \sqrt{\frac{\sum_{j=0}^K n_j - \bar{n}_i}{K}} \right|$$

Where $\bar{n}_i$ is the average normal vector calculated from the normalized point normals within the P_i 's neighborhood, defined as

$$\bar{n}_i = \sum_{j=0}^K \frac{n_j}{K}$$

Each point P_i with a normal standard deviation larger than a threshold value defined by the user, is classified as an edge point. The edge points are removed from the whole point clouds before the surface segmentation. The non-edge points are used for the next step of region growth. The point P_i with the minimum curvature value was picked up as the initial start seed point.

Inspired by previous studies (Nurunnabi, Belton, and West 2012, Khaloo and Lattanzi 2017b), The similarity criteria include (1) Coplanar Distance (CD), (2) Eculiean Distance (ED) and (3) Angular Distance (θ) The threshold for CD and ED are locally adaptive. CD ensures the coplanarity of points. The threshold for CD combines the median and Q_n estimator (Rousseeuw and Croux 1993). ED is an indicator of the distance between a point and its neighbor. The threshold for ED is set to be the median of all pairs of ED between the points and the seed point. The θ the angle between two point normals, ensuring that the points around the sharp edges can be distinguished into separate surface patches. The value is defined by the user. In this study, I set the threshold to 3° .

Based on the similarity criteria, points are included in a region if the properties are similar to the seed point. The region iteratively adds neighboring points to the current region, when the growth process is complete, a new seed is selected for the next surface patch segment.

The robot operator can then choose a target surface or the area of interest on a surface for local scanning. To generate the local scanning path, I designed an offline path planning algorithm formulated as an optimization process of minimizing the cost function defined by the total length of robot trajectories covering the target area by adjusting decision variables such as candidate viewpoints and their traversing sequence. An online path-planning process is then integrated to address real-time adaptive motion planning to avoid collisions caused by dynamic objects or tolerances in reconstruction.

4.1.3 Post-processing

The post-processing module is designed for estimating the transformation matrix (scale, rotation, and translation) between two point clouds acquired from different sensors. In this study, the global exploration module built a site model with an Intel RealSense D435 sensor, the local reconstruction captured the area of interest with a Kinect V2 sensor. The registration between two point clouds from different sensors can be challenging due to scale differences, density differences, partial overlap, and large outliers. The challenge is known as cross-source point cloud (CSPC) registration in the domain of computer vision (Huang, Mei, and Zhang 2023).

The traditional registration method based on classical iterative closest point (ICP) may yield inaccurate pose estimation for the CSPC registration. To address the large discrepancies in densities, I implemented the clustering iterative closest point (CICP) approach (Tazir et al. 2018). CICP integrates a novel correspondence point selection process based on voxelization and clustering before the matching. The target point cloud was first subdivided into small voxels (spatial cubes). The points within one voxel are classified into subgroups, each subgroup represents a local surface with a shared normal vector. For the local surface, a representative point is selected as the one closest to the centroid of the local surface. Then the representative of both the target point cloud and source point cloud are included in a set for correspondence matching. Each selected point can represent a local surface in a voxel of the source and target point clouds. Based on the selected correspondence points, the matching process is invariant to the point cloud density and scanning pattern.

4.2 Robust Localization Under Complex Environments

For spatial devices, such as robots or augmented reality (AR) devices, accurate localization can be important in path-planning tasks and the alignment between virtual objects and the physical world. Visual simultaneous localization and mapping (VSLAM) is one of the most widely adopted techniques for indoor localization. By extracting the visual features of a sequence of images as descriptors, the position of the camera can be tracked in real time. However, several environmental factors in indoor construction sites can pose challenges to VSLAM (Figure 4.3).

4.2.1 Challenges

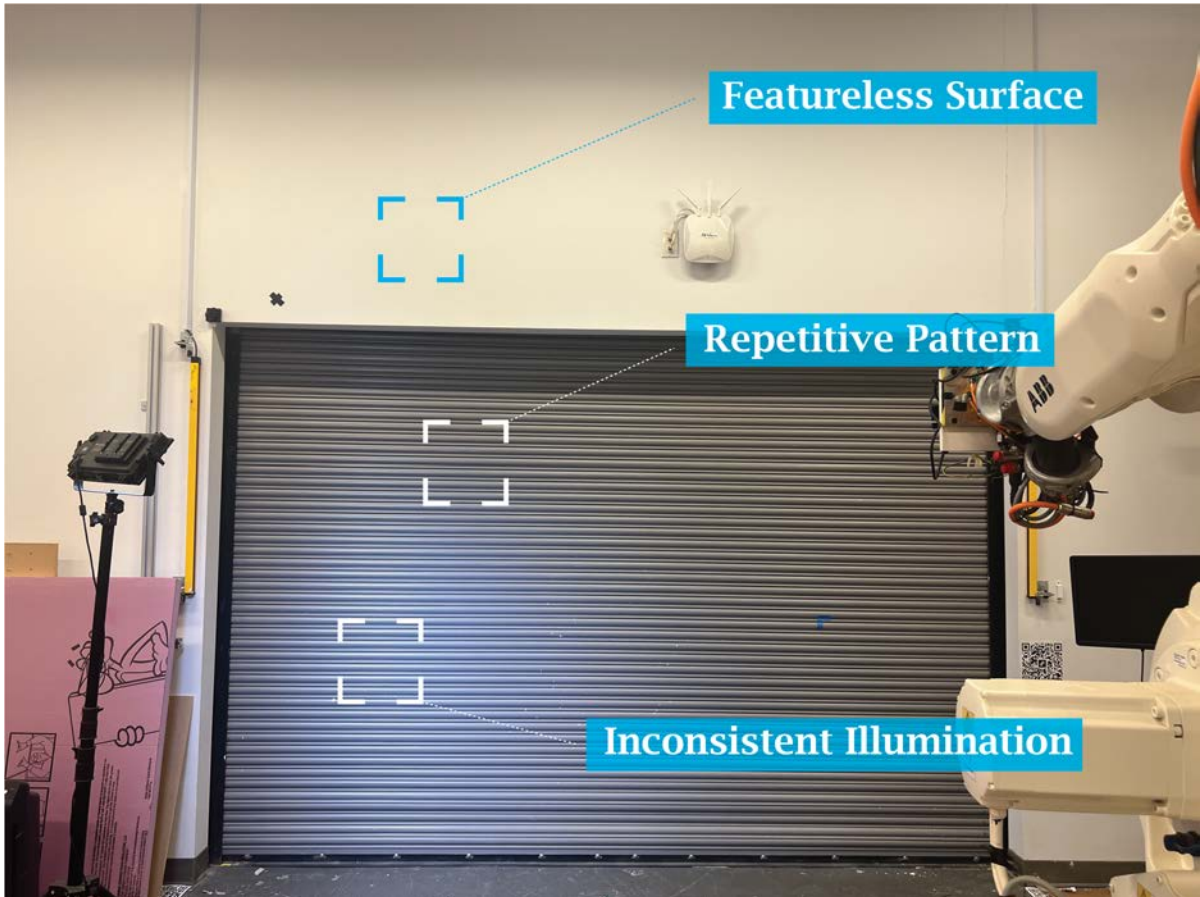


Figure 4.3: Challenging scenarios on job sites for visual SLAM including (a) a vertical white wall surface without rich textures (b) a roller door with repetitive line patterns (c) areas with artificial lighting that may change over time. The module for robust localization is exemplified and validated on an industrial robot (ABB IRB 6640). The result is compared with the built-in localization system estimated from the encoder of each joint. Image credit: Author.

- **Inconsistent illumination**

Construction sites can rely on temporary lighting or natural light. For areas with low light (shadows e.g.) and inconsistent illumination (turning on/off artificial lighting e.g.), VSLAM may fail in feature tracking leading to drift and errors in localization.

- **Poor textured surface**

VSLAM such as ORB-SLAM (Mur-Artal, Montiel, and Tardos 2015a) can yield impressive results in well-textured environments. However, construction job sites are not always texture rich. For instance, concrete buildings with flat walls can pose challenges to feature detection when the wall surface shares similar color and reflectivity.

- **Repetitive pattern**

Repetitive building components, such as scaffolding, are of insufficient variations in visual features. VSLAM is prone to fail due to the lack of distinctive feature points, visual ambiguity, and false matches.

In addition, global navigation satellite systems (GNSS) signals can be blocked or weakened by walls and other obstacles, leading to signal loss or attenuation and causing erroneous localization. To improve the robustness of VSLAM in such environments, this study proposes to integrate ultra-wideband (UWB) wireless localization and inertial measurement unit (IMU) into VSLAM systems to mitigate accumulative drift and enhance robustness. By leveraging the synergies between UWB, IMU, and VSLAM, our multi-modal localization can overcome the limitations of each sensing modality, providing robust localization in cluttered construction environments. When the VSLAM fails to provide reliable estimation, the complimentary sensing modalities can function to maintain localization consistency. Additionally, UWB as a global feature-independent localization system can correct the cumulative drift caused by VSLAM, leading to enhanced accuracy in long-term localization tasks.

4.2.2 Multimodal Localization

In this study, VSLAM system is built upon the ORB-SLAM framework (Mur-Artal, Montiel, and Tardos 2015b). UWB and IMU localization systems are combined with the VSLAM system through the extended Kalman filter (EKF). The proposed multimodal localization system is then tested under three scenarios — inconsistent illumination, repetitive pattern, and insufficient features. Through validating with the ground truth, I found that the integration of UWB and IMU can enhance robustness when VSLAM fails at feature extraction or correspondence

matching. Meanwhile, the cumulative drift caused by VSLAM can be reduced by UWB positioning. The real-world experiments demonstrated the feasibility and reliability of the system with applications in the domain of indoor localization for construction robots.

Ultra-wideband (UWB) as a wireless communication technique has been adopted for precise indoor localization in recent decades. Compared to other wireless communication technologies such as radio-frequency identification (RFID), Wi-Fi, and Bluetooth, UWB presents advantages including (1) accuracy: UWB can achieve high ranging accuracy (sub-centimeter) even in harsh environments due to its resistance to multipath (Dardari et al. 2009). (2) robustness: UWB is more robust to signal interference as the signal may be transmitted simultaneously over multiple frequency bands (H. Liu et al. 2007; F. Liu et al. 2020a). power consumption: UWB shows lower normalized energy consumption compared to other wireless communication protocols such as Wi-Fi and ZigBee (J.-S. Lee, Y.-W. Su, and Shen 2007). For long-term localization, UWB can be a suitable option in terms of energy efficiency. With the aforementioned aspects taken into account, many studies have been done to integrate UWB into VSLAM for mitigating scale ambiguity, improving accuracy, and enhancing robustness (T.-M. Nguyen, T. H. Nguyen, et al. 2019 , Xu et al. 2020). For example, UWB can effectively overcome scale ambiguity and scale drift when combined with monocular VSLAM systems (T. H. Nguyen, T.-M. Nguyen, and Xie 2020 , F. Liu et al. 2020b). In GPS-denied environments, UWB can serve as indoor GPS to supplement visual SLAM by providing global constraint (H.-Y. Lin and Yeh 2022) or local correction (Y. Cao and Beltrame 2021). For long-term localization, UWB can mitigate cumulative drift caused by visual SLAM (Francisco Javier Perez-Grau et al. 2017). For large-scale applications, UWB can be combined with LiDAR to enhance localization accuracy (Song, Guan, et al. 2019 , C. Wang et al. 2017). By tightly coupling multiple sensors such as IMU, monocular camera, and UWB, the optimization can be done in a joint manner for a more robust and accurate localization (T. H. Nguyen, T.-M. Nguyen, and Xie 2021 , T.-M. Nguyen, Yuan, et al. 2022). Extended from previous works, this study proposes a novel multimodal localization system for spatial devices such as construction robotics operated in unstructured environments, the integration of UWB can potentially address the VSLAM failures caused by inconsistent illumination, insufficient features, and repetitive patterns which are commonly seen in job sites.

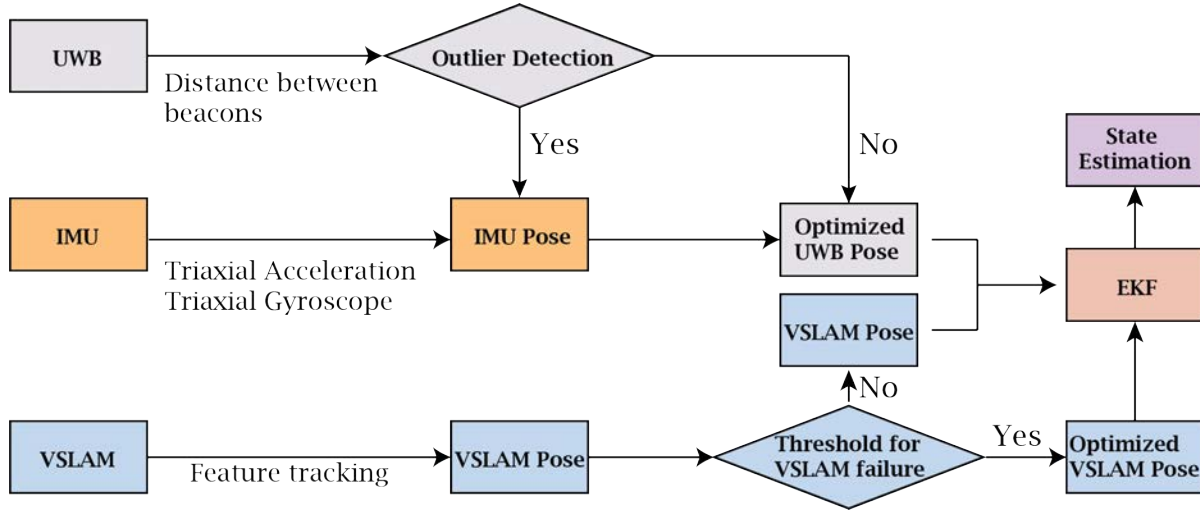


Figure 4.4: System flowchart of the multimodal localization system. Image credit: Author.

In this study, I used three measurements — IMU, the peer-to-peer ranges between UWB nodes, and the range between the robot and the 3D position of the visual features as the source for indoor localization. The IMU and UWB are first combined, and then I use EKF for the data fusion of VSLAM and the rectified UWB. I used a set of Marvelmind Super-MP-3D beacons for UWB localization and a Kinect sensor for VSLAM. The update rates for VSLAM and UWB are set to 30 Hz and 15 Hz respectively.

UWB localization uses ultra-wideband impulse radio to measure the distance between mobile and stationary base beacons. The localization of the target using UWB beacons requires (1) three known stationary beacons with known height or (2) four known stationary beacons (Fernando et al. 2019). In this work, I set up four stationary base beacons in a room of 5.0 m (W) $\times$ 15.0 m (L) $\times$ 2.8 m (H). The four base beacons are synchronized, and the position of each beacon is self-calibrated. The distance between the mobile beacon and the base beacon can be estimated by the Time-of-Arrival (ToA) of UWB signals. In each time slot, the mobile beacon transmits a request short-duration and low-power pulse signal at a unique channel with a time stamp, and the stationary beacon responds with a time stamp after a predefined time. Since the UWB sensors are synchronized at a picosecond level, we can precisely calculate the distance based on the difference between two time stamps and the speed of the signal. Then we can obtain the location of the mobile beacon using multilateration (Figure 4.5).

$$(x', y', z') = \sum_{i=1}^4 \left[\sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} - r_i \right]^2 \quad (1)$$

$$r_i = \frac{(T_{i,cycle} - T_{i,reply}) \times C}{2} \quad (2)$$

Where i is the index of the base beacons, (x_i, y_i, z_i) is the coordinate of i th stationary beacon which is known after self-calibration. (x', y', z') and (x, y, z) are the unknown coordinate of the mobile beacon before and after the error minimization of multilateration. r_i is the distance between each base beacon and the mobile beacon. $T_{i,cycle}$ is the time between the mobile beacon sending the signal and receiving the response. $T_{i,reply}$ is the time between a base beacon receiving the signal and sending the response. C is the speed of light (3×10^8 m/s).

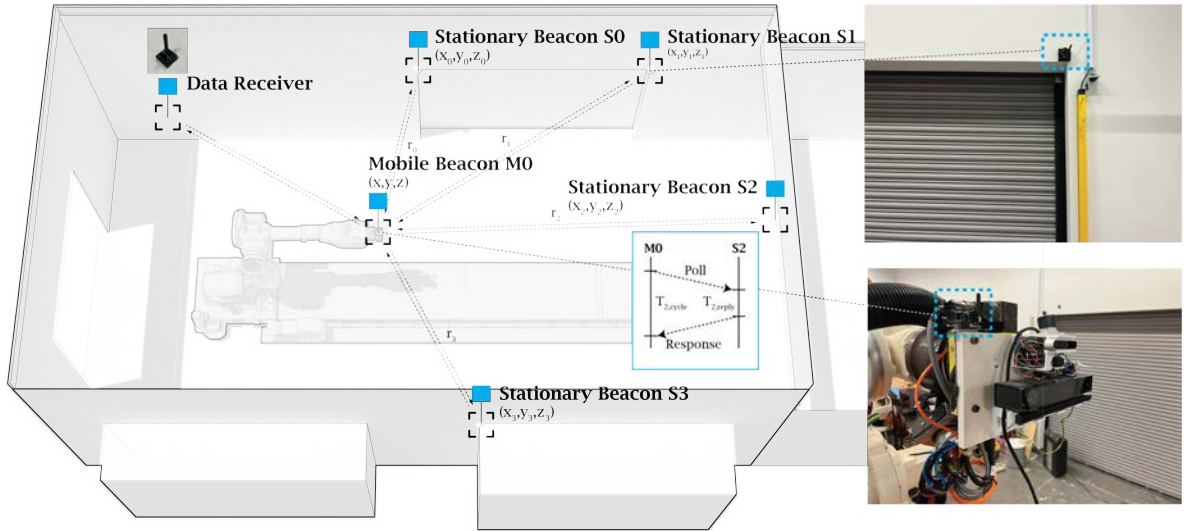


Figure 4.5: Infrastructure layout of the UWB positioning system. Image credit: Author.

The multipath interference caused by the reflections of UWB signals off walls and non-line-of-sight (NLOS) can affect signal accuracy which leads to inaccurate localization. Compared to UWB, IMU can provide more accurate short-term localization invariant to external factors. If the relative position between the UWB and IMU is larger than a threshold, e.g., 0.5 m, I use the predicted position of IMU to correct the UWB localization and remove the

Chapter 4. Method

outlier. As the update frequency rate of IMU is higher than UWB, the integration of IMU data is processed after two frames of UWB data are processed.

VSLAM tracks the features such as corners and edges in the environment and estimates the movement of the camera between consecutive frames. I use ORB-SLAM2 (Mur-Artal and Tardós 2017) for tracking the RGB-D camera's position and building the map of the environment. The RGB-D camera and UWB systems are synchronized to a common time reference, thus the localization information of the two systems can be aligned and paired for data fusion. As UWB localization is a general non-linear system, I use the extended Kalman filter (EKF) framework to couple UWB and VSLAM localization for more robust and accurate estimation.

The EKF is conducted in two steps — time updates and measurement updates. Once initialized, EKF predicts the system state at the next step and the uncertainty of the prediction. When the measurement is received, the system then updates the prediction and the uncertainty of the current state.

The state vector x consists of position and velocity (v_x, v_y, v_z) can be defined as:

$$x = [p_x, p_y, p_z, v_x, v_y, v_z] \in \mathbb{R}^6$$

The prediction of the next state can be defined as:

$$\bar{x}_{k+1,k} = f(\bar{x}_{k,k}, u_k) \quad (3)$$

$$P_{k+1,k} = F_{k,k} P_{k,k} F_{k,k}^T + Q_{k,k} \quad (4)$$

where F is the state transition matrix. $P_{k,k}$ and $Q_{k,k}$ are the covariances of $x_{k,k}$, $w_{k,k}$ respectively. u_k is the control input. $\bar{x}_{k+1,k}$ is a predicted system state vector at time step $k + 1$, $\bar{x}_{k,k}$ is an estimated system state vector at time step k .

In the update step, we can obtain the measurement z_k for calculating the Kalman gain G_k . I then update the state covariance matrix by

$$G_k = P_{k,k-1} H_k^T (H_k P_{k,k-1} H_k^T + R_k)^{-1} \quad (5)$$

where R_k is the covariance matrix of measurement uncertainty. H_k is the Jacobian matrix of the measurement, which can be defined as the partial derivatives of the measurement function

with respect to each state variable.

$$H_k = \partial h(i)/\partial x(j) \quad (6)$$

where i is the index of the measurement component. j is the index of the state variable. H_k needs to be computed at each time step.

We can then update the estimate with measurement by

$$\bar{x}_{k,k} = \bar{x}_{k,k-1} + G_k(z_k - H_k\bar{x}_{k,k-1}) \quad (7)$$

And update the uncertainty by

$$P_{k,k} = (I - G_k H_k) P_{k,k-1} \quad (8)$$

In this work, the implementation can be described as follows

- Initialization

I initialize the state vector by using an ArUco marker as a global reference frame for UWB and VSLAM (as robots can start from a known location), the covariance matrix was set based on prior empirical knowledge.

- Prediction

The measurement z_k at time stamp k contains the position and heading offset from the UWB and VSLAM localization, the predicted measurement $\bar{z}_{k+1,k}$ is defined as

$$\bar{z}_{k+1,k} = h(\bar{x}_{k+1,k}) + v \quad (9)$$

where $H \in \mathbb{R}^{8 \times 6}$ is the measurement model. $\bar{x}_{k+1,k} \in \mathbb{R}^{6 \times 1}$ is the predicted next position at time stamp k . v represents the uncertainty in the measurement. We can then obtain the innovation sequence i_k defined as

$$i_k = z_{k+1,k} - \bar{z}_{k+1,k} \quad (10)$$

The innovation sequence can be used to calculate the Kalman gain, which determines the weight given to the predicted state estimate and the actual measurement to update the state estimate.

- Estimation

The estimation can be done through Equation (7) The Kalman gain and state covariance are updated accordingly.

- Update

After estimation, we can update the state vector based on the measurements using the Kalman gain. The state vector contains information about the location and the velocity of the target.

VSLAM localization can fail due to environmental factors such as inconsistent illumination and insufficient features. I incorporate a data pre-processing step into the pipeline before the data fusion to identify VSLAM failures. I defined an empirical threshold e.g. 0.2 m . If the relative distance between the VSLAM and UWB localization is larger than the threshold, the VSLAM localization is considered as an outlier and replaced by a new position that shares the same heading offset with the UWB localization at an offset distance of the threshold value.

4.2.3 Evaluation

I evaluated the performance of the multimodal localization system in terms of accuracy and robustness. Accuracy is measured through absolute error evaluation along the x, y, and z-axis. The ground truth is obtained from the real-time joint reading from the built-in sensors of the industrial robot. Based on the joint angles of each joint, we can use forward kinematics to estimate the camera pose.

Table 4.1 gives the average absolute distances between the estimated and ground truth trajectory over time stamps. From the table we can see that the integration of VSLAM can provide a reliable short-term localization that can significantly reduce the impact of the UWB range measurement outliers caused by environmental factors such as noise, interference and occlusion (Figure 4.6) ¹.

¹Many indoor robot applications require centimeter-level accuracy to navigate safely and effectively within indoor environments, especially in crowded or dynamic spaces such as jobsite. The accuracy requirement can be varied depending on the tolerance of the task, for example, in foundation concrete casting practice, the maximum horizontal deviation from the plan dimension can be 1.27 cm (Committee 2006).

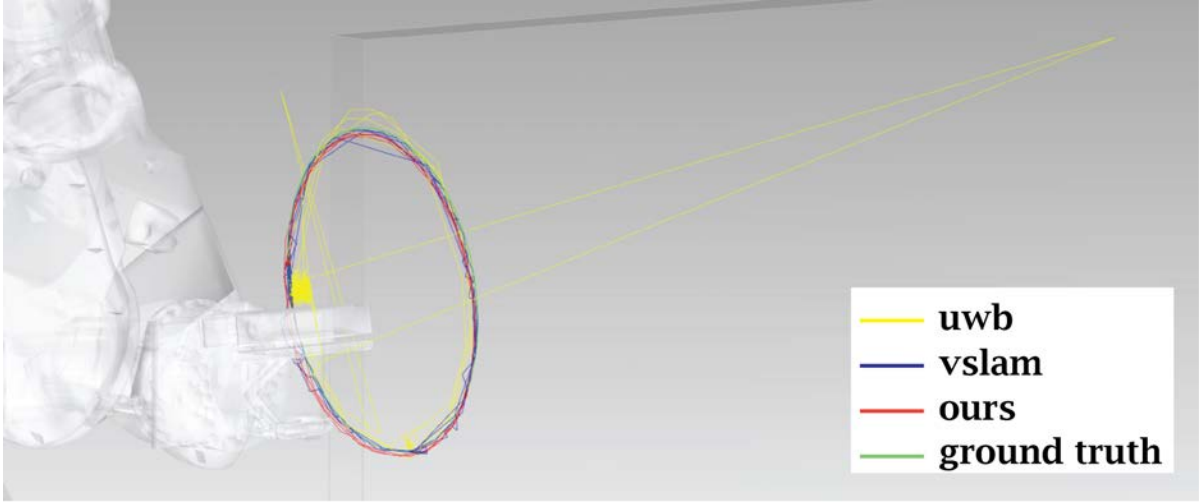


Figure 4.6: Trajectories obtained using UWB-only, VSLAM-only and the proposed multi-modal localization system. Image credit: Author.

Table 4.1: The translation error along each axis. Image credit: Author.

	VSLAM-Only	UWB-Only	Multimodal
$x(m)$	0.18	0.53	0.08
$y(m)$	0.12	0.13	0.05
$z(m)$	0.27	0.33	0.16

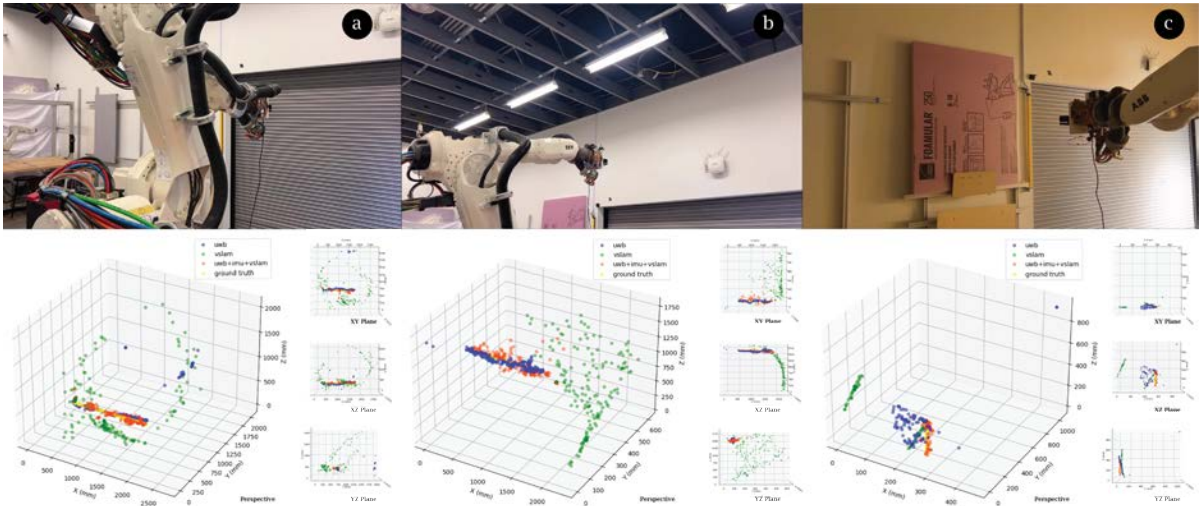


Figure 4.7: Performance comparison of VSLAM-only, UWB-only and the multimodal localization system under three conditions — (a) repetitive pattern, (b) poor textured surface and (c) inconsistent illumination. Image credit: Author.

To evaluate the robustness of the multimodal localization system, I used three environmental interference including inconsistent illumination, repetitive pattern, and insufficient features to evaluate the stability of different localization approaches.

- **Repetitive pattern**

Components such as roller doors contain repetitive visual features such as parallel lines which can pose challenges to the feature tracking in VSLAM localization. I use a roller door as a target object to test if the multimodal localization system can remain robust when symmetric or repetitive scene patterns bring ambiguities for feature correspondence.

- **Poor textured surface**

Some areas of perpendicular walls in the test environment are blank or poor textured. Insufficient visual features inliners can reduce the accuracy of VSLAM. I designed a trajectory passing through a blank wall to evaluate the multimodal localization system when distinct visual features are insufficient.

- **Inconsistent illumination**

To test the influence of illumination on the system's robustness, I set up two controllable illumination conditions — the dark environment and the normal environment. I define the artificial indoor light of 300 lux as a normal environment and 30 lux as a dark environment. We can test the impact of inconsistent illumination on the multimodal localization system by switching between two illumination conditions.

Under each condition, I tested three localization approaches including VSLAM-only, UWB-only, and the multimodal system (UWB, IMU and VSLAM). As can be seen in Figure 4.7 , when the lighting condition changes from normal to dark, VSLAM localization fails due to poor feature extraction and matching. When the lighting condition returns to normal, the cumulative errors lead to significant drift. Meanwhile, we can see that the UWB-only localization is affected by Non-Line-of-Sight (NLOS) when the UWB signal is obstructed. The attenuation of UWB signals leads to noises and inaccuracy in UWB-only localization. However, IMU sensors, as a complimentary modality, optimize the UWB localization when the current

UWB measurement is identified as an outlier. The localization trajectory obtained using the multimodal system remains robust when both VSLAM and UWB localization fail.

The repetitive patterns of the roller door and the poor textured blank wall both affect the accuracy of VSLAM. However, the UWB and IMU localization systems provide reliable and globally consistent range measurement when VSLAM fails. The outlier removal process before EKF fusion avoids the negative impact of VSLAM failure on the pose estimation. The result illustrates that our proposed multimodal localization system can enhance the system robustness and overcome the limitations of each localization system by its own.

This section presented a robust multimodal localization approach for indoor construction robotics. The approach integrates multiple sensor readings from RGB-D cameras, UWB and IMU to mitigate drift errors caused by each single sensor. I tested the multimodal localization system under three commonly seen conditions in job sites including inconsistent ambient lighting, surfaces with repetitive patterns, and poor textured surfaces. In the experiments, the multimodal localization system demonstrates robustness when VSLAM localization fails. When both UWB and VSLAM provide reliable localization, the UWB can reduce the drift error caused by VSLAM and VSLAM can enhance the accuracy when UWB is affected by noise in a local region. The limitation of the system lies in the coverage area of UWB sensors. In the experiment, I set up four UWB beacons to cover a room of 75 m^2 . The increased area may involve more stationary UWB beacons. In future work, an optimization approach for effective UWB beacon layout can be investigated to enhance the scalability of the localization system for large-scale construction sites such as tunnels. The strategically placed UWB beacons can increase the robustness of the localization system. UWB as a complementary global constraint can provide reliable localization, when sensors such as light detection and ranging (LiDAR) and GNSS fails due to signal blocking, insufficient geometric features, or object occlusion.

4.3 Geometry-informed Semantic Segmentation

To convert point cloud data into semantics, I built an end-to-end system based on a deep neural network. As depicted in Figure 4.8, the system first obtain RGB and depth data from a Kinect sensor. The system then automatically translates the low-level point cloud information into high-level semantic information through steps such as 3D reconstruction, 3D semantic

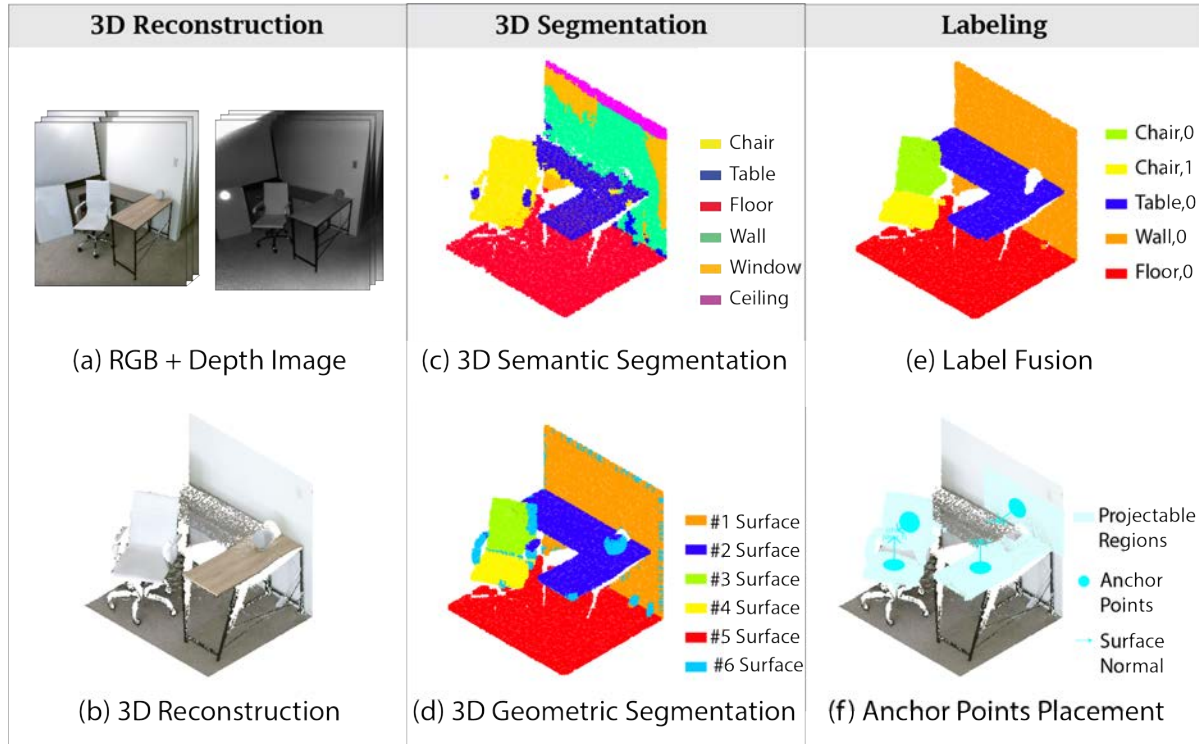


Figure 4.8: The framework of the semantic-based system consisting of three components: 3D reconstruction, 3D segmentation and labelling. The 3D reconstruction component translates the acquired RGB and depth data from a Kinect sensor into point clouds. The 3D segmentation component segments point clouds into discrete surfaces with semantic label. The labelling component fuses geometric and semantic labels and creates anchor point for each segmented surface. Image credit: Author.

segmentation, 3D geometric segmentation and labeling. Finally, the system places an anchor point at each segmented surface. Users can map virtual content onto a physical surface by referring to its semantic label. I use the Point Cloud Library in C++ for point cloud processing and TensorFlow Library in Python for building the deep learning architecture.

4.3.1 3D Reconstruction

The system obtains 3D reconstruction of the physical environment through dense simultaneous localization and mapping (SLAM). Following the KinectFusion framework (S. Izadi et al. 2011), I use a Kinect V2 sensor to reconstruct the scene in four steps:

1. I obtain raw depth information at each image pixel in the image domain. To reduce noise, I applied a bilateral filter to the raw depth map.
2. Each frame of depth images is transformed into 3D points and integrated into a 3D volumetric data structure.
3. Like live camera localization that involves estimating the current camera pose for each frame, I obtain the Kinect sensor pose by the full frame model ICP method (Paul J Besl and McKay 1992). I assume that only a small camera motion occurs from one frame to the next, thus I can use a fast projective data association algorithm to obtain correspondence points and the point-plane metric for Kinect sensor pose estimation.
4. The point cloud reconstructed contains noise and outliers inherent due to the errors of the depth camera, I use statistical outlier removal algorithms to remove outliers and prepare an effective 3D model for further processing.

4.3.2 3D Segmentation

The 3D segmentation component consists of two steps: 3D semantic segmentation and 3D geometric segmentation. The first step, 3D semantic segmentation, segments point clouds into clusters. Points in each cluster share the same semantic label. The second step, 3D geometric segmentation, is to further segment the point clouds into clusters based on their geometric properties. Each cluster contains the point clouds that are co-planar.

I used the PointNet architecture (Qi et al. 2017) for training a dataset of Stanford Large Scale 3D Indoor Scenes (Armeni, Sax, et al. 2017a; Armeni, Sener, et al. 2016). The dataset contains 6020 square meters of indoor areas from diversified building typologies such as offices, conference rooms and open spaces. 12 semantic elements cover most commonly seen objects indoor, such as structural elements (*ceiling, floor, wall, beam, column, window and door*) and furniture (*table, chair, sofa, bookcase and board*).

4.3.3 3D Labelling

After semantic segmentation and geometric segmentation, each point cloud is annotated with two labels — its semantic class and geometry ID. However, due to noise in the 3D reconstruc-

Chapter 4. Method

tion, the result of semantic segmentation is a combination of major correct point clouds and a small fraction of mislabeled point clouds.

I use a majority voting scheme to unify semantic labels of point clouds that shares the same plane. First, the result of 3D geometric segmentation is used to enclose a set of point clouds P as voters. Then we can assign a representative semantic label L_P . The representative semantic label assigned for all of the points in P is determined by choosing the semantic label with the highest probability. The final semantic label L_P is obtained by

$$L_P = \arg \max_{l \in \{1, \dots, \mathcal{L}\}} \frac{N_l}{N}$$

where l indexes through semantic labels and L is the number of semantic labels. N is the number of points in P , N_l is the number of points with the semantic label l .

I then determine the points that are visible from given locations of projectors based on their field of view using frustum culling algorithm (Assarsson and Moller 2000). By culling points visible from projector locations, we can determine regions in the scene that are projectable. For each region, I set an anchor point for virtual content at the centroid of all points within the region. An anchor point Pt_i is annotated with its Cartesian coordinates (x_i, y_i, z_i) , normal N_i , semantic label l_{S_i} and geometric label l_{G_i} , formatted as $((x_i, y_i, z_i), N_i, l_{S_i}, l_{G_i})$.

PointNet is a deep neural network that directly consumes point clouds and outputs the per point semantic class labels. To prepare the training data, I first split the captured point cloud with area 0.5 m by 0.5 m and randomly sample 2,048 points from each block. Each selected point is represented by its Cartesian coordinates, color information, and its normalized coordinates to the captured scene. The 9-dimensional vectors are mapped into high-dimensional space via Multi-Layer-Perceptrons (MLPs). The high-dimensional local features are then aggregated into the global feature via Max-pooling. The global feature and the local feature are then concatenated as the point feature. Finally, the point feature is mapped to the output class scores via MLPs (Figure 4.9).

After semantic segmentation, each point of the captured scene is classified into a given semantic class. I then segment the point cloud into parallel or shared planes based on the geometric properties of the point cloud (Figure 4.10).

Each point cloud is characterized by its estimated normal and normal distance from the origin (Berkmann and Caelli 1994). The normal vector of each point cloud can be estimated

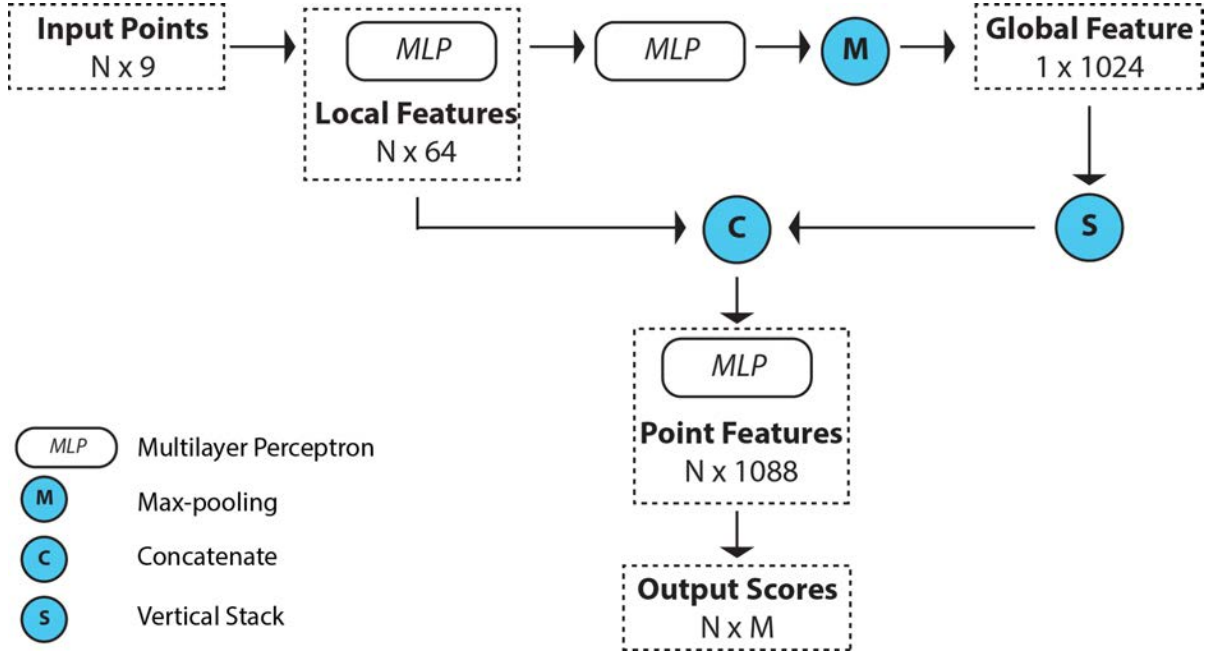


Figure 4.9: Simplified architecture of PointNet (Qi et al. 2017) (C): Concatenate (M): Max-pooling (S): Vertical stack. Image credit: Author.

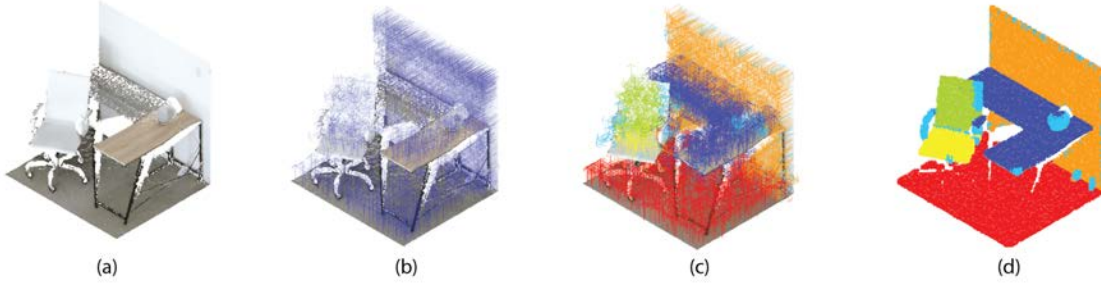


Figure 4.10: Steps of 3D geometric segmentation : (a) Input point clouds (b) Estimate normal of each point (c) Cluster point clouds by normal and normal distance from the origin (d) Result of 3D geometric segmentation. Image credit: Author.

based on its adjacent point clouds. For each point p_i , I pick k nearest neighbors N_p and compute the corresponding covariance matrix C , which is defined as:

$$C = \begin{bmatrix} p_{i_1} - \bar{p} \\ \dots \\ p_{i_k} - \bar{p} \end{bmatrix}^T \cdot \begin{bmatrix} p_{i_1} - \bar{p} \\ \dots \\ p_{i_k} - \bar{p} \end{bmatrix}, i_j \in N_P$$

$\bar{p}$ is the centroid of N_p . Then I estimate surface normal by finding the smallest eigenvalue λ_0 and its corresponding eigenvector v_0 of the covariance matrix C . Assuming $\lambda_0 < \lambda_1 < \lambda_2$, the surface variation σ_i can be estimated by:

$$\sigma_i = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}$$

σ_i is a feature for detecting edge points. When the point clouds are distributed in a plane, σ_i is small. If σ_i of a point is larger than a threshold σ_τ , the point can be categorized as a point on edges or borders.

After normal estimation, the point clouds are grouped into surfels based on the angle between normal vectors. The angle θ between vectors can be estimated by:

$$\theta = \cos^{-1}(u, v)$$

u and v are the normal of two points. If θ is within the defined angle threshold, two point clouds are grouped as parallel surfels. Finally, for each surfel, I use the normal distances of points from the origin to determine if the surfel shares a plane with other parallel surfels. Similarly, by setting a threshold d_t we can find clusters of co-planar surfels from parallel surfels.

Table 4.2: Quantitative model evaluation of the semantic-only segmentation approach and the revised label fusion segmentation approach on the point cloud of scanned scenes. Image credit: Author.

Scene	Method	Mean IoU	Overall Accuracy	Accu- racy	Avg. Class Ac- curacy
Scene 1	Semantic Only (PointNet)	66.3%	96.3%		95.6%
	Revised	71.0%	94.4%		76.0%
Scene 2	Semantic Only (PointNet)	71.3%	93.7%		79.1%
	Revised	86.4%	96.5%		98.1%
Scene 3	Semantic Only (PointNet)	46.2%	77.8%		56.8%
	Revised	65.7%	78.3%		69.8%

In order to obtain a robust and accurate segmentation, I used random sample consensus algorithm (RANSAC) to find inlier surfels and remove outliers. RANSAC algorithm first

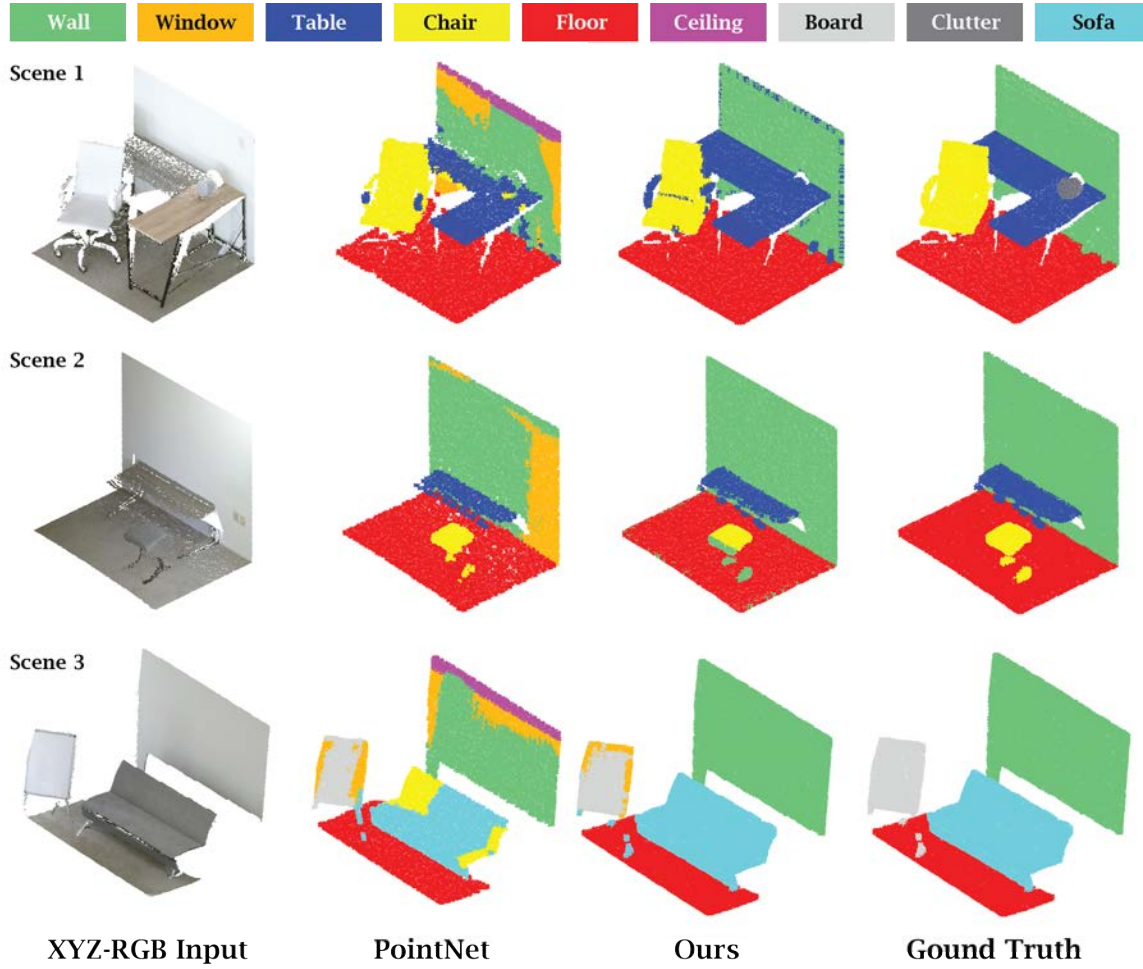


Figure 4.11: Qualitative model evaluation of the semantic-only segmentation approach and the label fusion segmentation approach the point cloud of scanned scenes. Image credit: Author.

estimates a hypothesis plane based on the randomly selected three points from coplanar surfels. Point clouds are categorized as inliers if the distance between points and the hypothesis plane is below a threshold. After iterative processing, we can find the plane that categorizes the maximum fraction of points as inliers. The outlier points are removed. The inlier points are labeled with the plane normal and each surfel is assigned a unique geometry ID.

4.3.4 Evaluation

To evaluate the usability of the system, I conducted both an accuracy study and user study. The accuracy study measures the accuracy of the 3D segmentation in both quantitative and qualitative manner.

I then evaluate our trained model on point clouds obtained from workplaces and homes using a Kinect Sensor. The ground truth for each scene is manually labeled. I compare semantic-only approach with the label fusion approach I adopted for the system, which uses the majority voting to unify the semantic label of points sharing a plane. As depicted in Figure 4.11, point clouds that were misclassified by the semantic-only approach are corrected by adding plane constraints, which means the system can produce more uniform and accurate result on environment point cloud I captured. Point clouds in this study are majorly composed of objects with simple geometry. I unified the semantic label of point clouds belonging to the same plane based on the assumption that co-planar point clouds share the same semantic class. Since the goal is to identify semantic labels of projectable surfaces, instead of recognizing semantic labels of every items, the geometry-prioritized approach I used can well serve for applications involving planes. This approach might not work for scenes that contain geometrically complex objects or various semantically different objects sharing similar geometry.

Chapter 5

Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

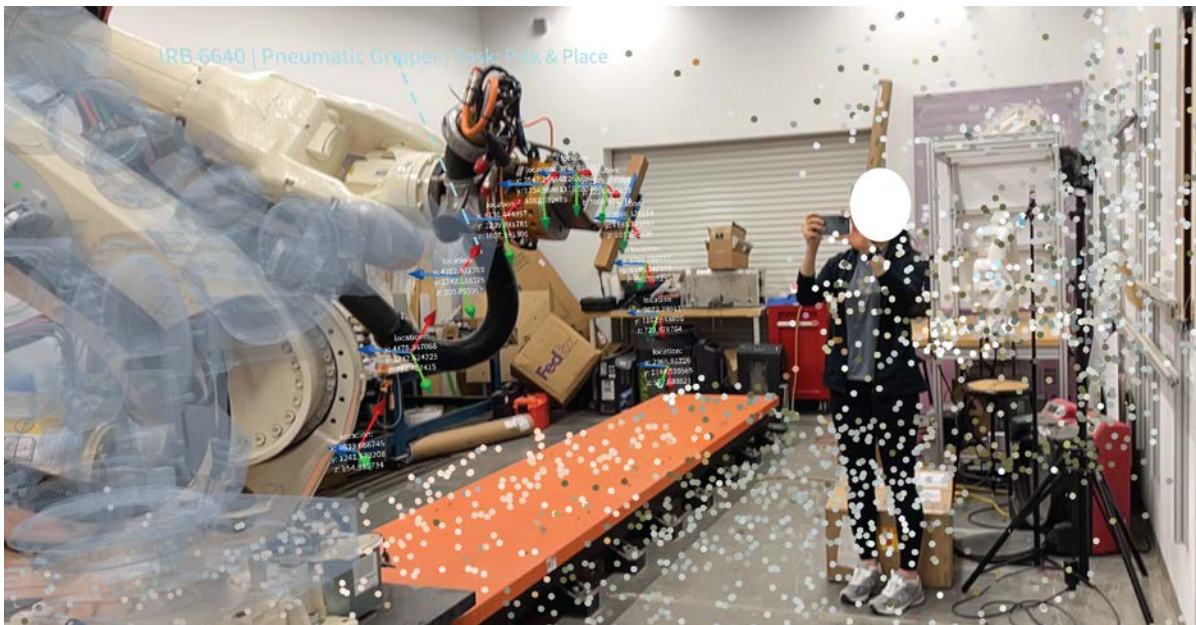


Figure 5.1: A robot operator uses a mobile AR device to preview and edit the trajectory of an industrial robot for pick and place tasks, the AR interface overlays the waypoints, the task information, and a ghost preview of the robot target pose onto the physical environment for visual guidance and collision avoidance. Image credit: Author.

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

This case study introduces a proxemic-aware augmented reality interface for (1) robotically-assisted pick & place and (2) site scanning. In both tasks, robotic systems can serve as a valuable construction tool for humans by providing potential benefits in safety, efficiency, and accuracy. For example, robots can assist humans in pick & place by automating labor-intensive material handling and reducing physically demanding tasks such as heavy object lifting. In site scanning, robotic systems can access hard-to-reach or hazardous areas to collect raw data for site monitoring, process tracking, and surveying. To facilitate online supervision and robot manipulation, this case study introduced a novel AR-based human-robot interface for overlaying the digital information of robots, such as trajectory, waypoints, and task specifics, over the physical environment (Figure 5.1). With consideration of the clutteredness of the work environment, this study introduces an AR system with adaptive view management to mitigate information overloading. The system leverages human-robot spatial relationship to automatically adjust what and how much visual content needs to be presented. Therefore, the robot operator can perceive the relevant data through AR interfaces without being overwhelmed by excessive information exposure. I defined a task-specific model for evaluating human-robot proxemic (HRP), where the system can identify HRP levels based on raw features, such as distance and orientation. Based on HRP levels, I design a set of visual elements for presenting robots' information at various levels of detail. To demonstrate the functionality of the system, I present a series of proof-of-concept applications showing that the AR system can assist the robot operator in a wide range of robotically-assisted construction tasks. Finally, I conducted a user study with students at Carnegie Mellon University, using the system as an instructional tool to operate the industrial robot at the Digital Fabrication Laboratory. The user study proves that the proxemic-aware AR system can reduce mental loading, increase visual clarity, and improve interaction efficiency in robotically-assisted tasks such as pick and place for novel users.

In this work, I contextualize the AR for HRI system in the manufacturing and construction domain. In construction and manufacturing, the working environment can be cluttered due to many factors. For example, the piles of construction materials, such as bricks and pipes, may not always be well organized and stored at the designated location. Temporary structures, such as scaffolding and formwork, and accumulated waste materials can add visual clutteredness. The AR application in cluttered construction environments may increase the cognitive load on

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

robot operators as they try to process both the physical environment and augmented digital information simultaneously. The cognitive overload can further lead to fatigue and errors in robotic manipulation. The cluttered AR interface makes it difficult for robot operators to focus on the most relevant AR content with the field of view. AR for HRI applications in cluttered environments such as construction sites needs to take information overlading and its impact on robot operators into consideration. In this case, I will introduce a proxemic-aware AR system and investigate how the system may mitigate information overloading issues in AR for robotically-assisted tasks.

The AR interface utilized in this case study builds upon the framework introduced in Section 4 - Method. The localization module provides raw data, including the location and orientation of handheld AR devices. This raw data is then used to estimate the proxemic level. The location of each spatial device is registered to a shared coordinate system that is constructed from a digital representation of the workspace. The motion planning module of the framework computes the near-optimal trajectory for site mapping and other tasks assigned by the operator. Additionally, I integrated a module for calculating human-robot proxemics (HRP), enabling the AR view to adapt to the current situation based on the HRP level.

5.1 Background

Augmented reality (AR) systems can serve as a human-robot interface, providing visual feedback to the operator by overlaying digital content onto the physical environment. However, the information-intensive environment can easily result in information overload that can negatively impact the user (Lipowski 1975) and cause safety issues in human-robot interaction (HRI). For instance, the annotations or labels in AR interfaces may occlude physical objects within the area point of interest. Too much information superimposed in AR views can cause ambiguity and make it difficult to filter out relevant information from trivial information(Keil et al. 2013). To mitigate the information overload in augmented reality, contextual factors are harnessed for optimizing user interfaces that can adapt to the user's cognitive load (Lindlbauer, Feit, and Hilliges 2019), ongoing tasks, and device characteristics (J. Zhu, Soh-Khim Ong, and A. Y. Nee 2015). This case study sets a first step to build a context-aware AR system

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

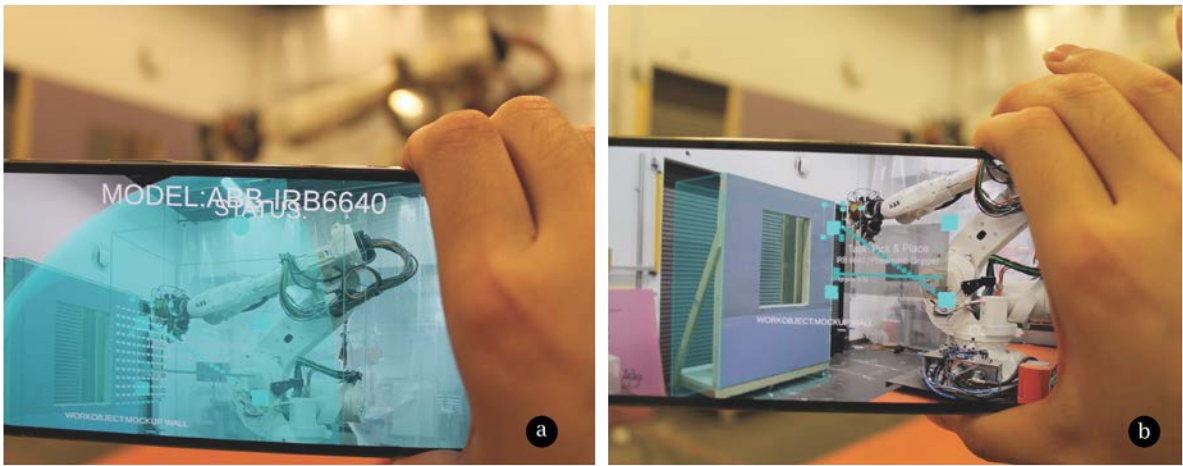


Figure 5.2: Proxemic-aware augmented reality (AR) may mitigate information overloading in AR-based human-robot interaction (HRI) (a) an AR interface displays all robot information at the same level of detail (b) virtual information is filtered based on human-robot proxemics, only the most relevant information is presented in the AR view. Image credit: Author.

that leverages human-robot proxemics (HRP) as an implicit cue for automatically adjusting the information presented in the user’s field of view (Figure 5.2).

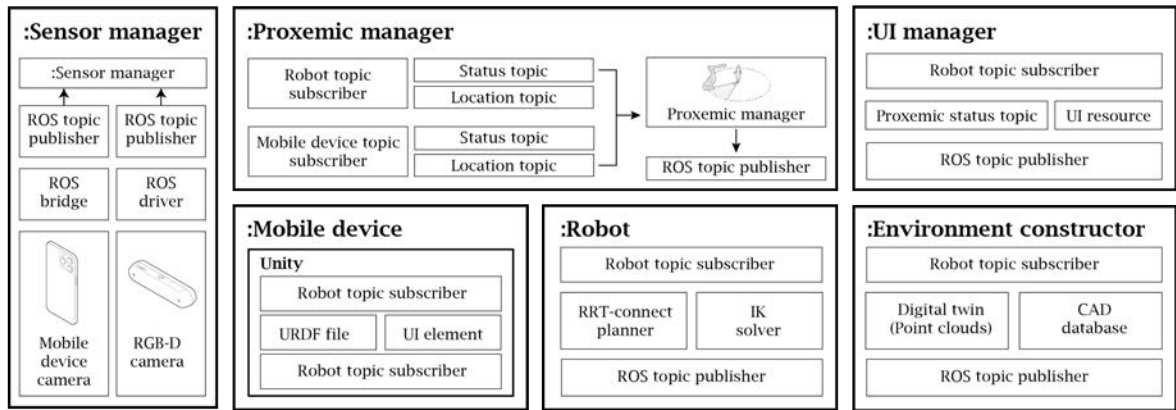


Figure 5.3: System deployment diagram of the proxemic-aware AR for HRI. Image credit: Author.

In this chapter, I will first discuss how I define human-robot proxemics. Proxemics was initially introduced by Edward T. Hall in the 1960s, defined as “the interrelated observations and theories of humans’ use of space as a specialized elaboration of culture”(Edmund T Hall

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

and Edward T Hall 1966). It has been widely used in human-machine interaction for the design of interfaces, devices, and systems that involve physical interaction with users. For example, fine-grained measurement of inter-entity relationships such as distance, orientation, identity, movement, and location can be exploited to design contextual-aware interfaces that mediate the simultaneous interaction of multiple people (Ballendat, Marquardt, and Greenberg 2010a). By understanding how humans use space and distance to communicate with each other, we can design robots with socially acceptable behaviors in scenarios such as navigation (Mead and Matarić 2017a) and object handling (Koay et al. 2014). Many studies have been conducted to quantitatively define human-robot proxemics (Mead and Mataric 2012, Mead and Matarić 2017b). Expanded from previous works, I introduced a novel approach to building a task-specific HRP model. The model is based on empirical data collected from robot operators consisting of three participants with prior experience in robot manipulation in the Digital Fabrication Laboratory at Carnegie Mellon University. Each participant was asked to stand at a set of predefined locations around robots and rate their perceived human-robot proxemic level through a mobile device. Different from previous proxemic models built upon discrete zones, I used a continuous function together with fuzzy inference to estimate human-robot proxemics. The non-crisp model better aligns with human's experience of proxemics, as we interpret spatial relationships in a gradual and progressive manner. The details of data collection and post-processing are elaborated in Chapter 5.3.1.

I then designed an adaptive visual system for presenting the operator with information about the robot at different granularities or levels of detail (LoD) (Lindlbauer, Feit, and Hilliges 2019). Each LoD reveals the information most relevant to the operator's current HRP levels. For example, when the operator shows minimal engagement with the robot, the AR interface will present overview information of the robot model and task in a brief textual format. When the operator moves toward the robot, the HRP level increases accordingly, indicating a higher possibility of interaction; more detailed information, such as trajectory and waypoints, will be revealed in the AR interface. For each HRP level, I designed a collection of widgets, such as panels and points, to help users understand and interact with the virtual information at the LoD tailored to the current context.

Finally, I will illustrate how the system can be used as an instructional tool to assist students in specific HRI scenarios such as 3D scanning and pickup & place. I implemented an

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

AR interface on a mobile phone to control and supervise a six-axis industrial mounted on a linear actuator. The industrial robot will perform two aforementioned tasks, pickplace and site scanning. In the pickplace task, the robot operator uses a mobile AR device to select one target object (e.g. wood bar post) and define the target location. The AR interface will present the operator with the pre-operative robotic trajectory and waypoints. After the operator confirms the command, the industrial robot will pick up the object and transport it to the target location. The AR interface can visualize the information about the robot toolpath before actions are performed. AR can highlight potential collision zones or obstacles in the robot's path, allowing workers and robot operators to mitigate potential collision risks. In the site scanning task, the industrial robot traverses the user-defined area and captures point cloud data of the site. The AR interface allows the robot operator to interactively edit the scanning area and overlay the supplemental information, such as blueprints and 3D models, for informed on-site decision making. The progress tracking bar in AR interfaces allows workers to estimate the remaining time that the robot will take to complete the scanning. Thus, workers can better allocate resources for different parallel tasks.

I compared the proposed system with the baseline - an AR interface for HRI without proxemic awareness, and evaluated the task performance by collecting responses from the participants through a post-experiment questionnaire. The results show that the system significantly outperformed the baseline in terms of reducing mental loading in the surface inspection and pick & place task (all $p < 0.01$). I believe the proxemic-aware AR system can serve as a novel human-robot interface that provides robot operators with relevant robot information when and where they need it.

5.2 Related Work

5.2.1 Augmented Reality For Human-Robot Interaction

The integration of robotics in complex scenarios such as construction, healthcare, and space exploration requires human intervention at varied levels. However, a robotic system's capacity to sense and interpret human signals which are accounted for in their programming in order to elicit suitable responses can be limited by the lack of information that robots can present (Walker et al. 2018b). Augmented reality (AR) interfaces can potentially address the

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

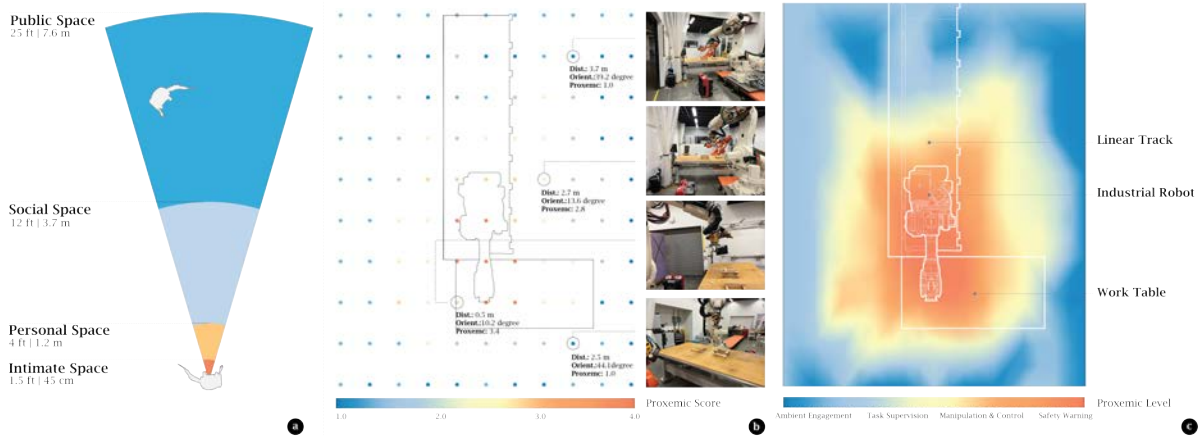


Figure 5.4: The comparison between human-human proxemics and human-robot proxemics (a) human-human proxemics with discrete zoning (b) human-robot proxemic evaluation at sampled location (c) human-robot proxemic as a continuous function interpolated from the sampled location. Image credit: Author.

constraint by overlaying supplemental virtual information onto robotic work cell. Suzuki, Ryo, et al. summarizes the categories of information presented in AR, which can be classified into robot information, environment information, pre-operative information, and supplemental information for expressive interaction (Suzuki, Karim, et al. 2022). Robot information describes the internal status of robots, such as poses, joint configurations, and robot position. Environmental information includes the robot's perception of the surroundings obtained from the equipped sensors. Pre-operative information is related to the robot's planned motion for the operator to perceive the future actions of the robot. Supplemental information can increase the robot's expressiveness by showing interactive content such as animation or dynamic data visualization (Suzuki, Zheng, et al. 2019). To visualize virtual information, visual elements such as panels, labels, interactive points, lines, and simulated robots are commonly used for presenting AR content at different levels of detail. Williams et al. outline how AR may enhance HRI in collaborative tasks, including navigation, object manipulation, prototyping, training, supervision, and social interaction (Williams et al. 2018). AR systems have been adopted in application domains such as robotically-assisted surgery (RAS), maintenance, design, and education (Makhataeva and Varol 2020). For example, AR interfaces can assist a surgeon in RAS by overlaying surgery plans, highlighting operational areas, and visualizing instrument status (Qian, J. Y. Wu, et al. 2019b). Within the context of manufacturing, AR for HRI can

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

be used for teleoperated maintenance (Yew, S. K. Ong, and A. Y. C. Nee 2017), assembly (Mourtzis, Zogopoulos, and Xanthi 2019), and intuitive robot programming (Soh-Khim Ong et al. 2020). In social robotics, AR interfaces can augment robot expression to enhance human-robot trust (Gao et al. 2019). When designing an AR for HRI system, an excessive amount of information that exceeds the user's cognitive capacity can lead to difficulties in perceiving, decision-making, and task performance in robotic manipulation. Visual clutteredness can make it challenging for users to focus on relevant information. Especially for construction practices, the site can be cluttered with piled materials, machinery, and tools. When abundant digital information is presented in an AR interface, users may find it difficult to engage with the surroundings and make effective decisions. Informed by the challenge of information overloading, this case study presented a novel proxemic-aware system that leverages the spatial relationship between humans and robots to adjust the amount and the level of details of digital information presented in AR views. According to the user study in section 5.4, the introduction of proxemics can enhance the robot manipulation performance measured by task completeness and task completion time.

5.2.2 Proxemic-aware Interaction

Proxemics describes how people use spatial relationships to mediate their relationships (Edmund T Hall and Edward T Hall 1966). People in interactions change distances and orientations to accommodate different social contexts. In ubiquitous computing, proxemic variables such as distance, orientation, identity, movement, and location (Ballendat, Marquardt, and Greenberg 2010b) are used to identify the spatial relationship between humans and surrounding devices. The interpersonal spatial relationship can be leveraged as an implicit cue for human-computer interaction, such as context-aware display (Ledo et al. 2015) and cross-device interaction (Grønbæk et al. 2020). For instance, by leveraging proxemic information, an interactive display surface can implicitly react to the surrounding people and their personal devices. When a person enters the room, the media player shows peripheral awareness content. As the person moves closer to the display, more detailed information is revealed (Ballendat, Marquardt, and Greenberg 2010b).

Within the domain of robotics, proxemics, as a socio-scientific theory, have been used as guidance for the design of socially-aware human-robot interaction. For instance, in au-

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

onomous navigation, proxemics are integrated with traditional navigational algorithms to develop socially acceptable navigation strategies for robots operated in crowded environments(Daza et al. 2021, Patompak et al. 2020a, Rios-Martinez, Spalanzani, and Laugier 2015). Compared to human-human proxemics, humans kept interpersonal distances from robots (0.45 *m* to 1.20 *m*) on the first encounter(Hüttenrauch et al. 2006a, Walters, Dautenhahn, Koay, et al. 2005). A number of HRI studies have been conducted to investigate the factors that can influence human-robot proxemics (HRP), such as robot appearance(Butler and Agah 2001), robot height(Dag Sverre Syrdal et al. 2007), robot behavior(Mumm and Mutlu 2011a), and personal preferences(Walters, Dag S Syrdal, et al. 2008). To define HRP, multiple studies have been done to build a rule-based system based on empirical experiments(Hüttenrauch et al. 2006b, Walters, Dautenhahn, Boekhorst, et al. 2009). To extend the application range of HRP frameworks, adaptation mechanisms such as policy gradient reinforcement learning (Mitsunaga et al. 2006) and fuzzy inference systems (Patompak et al. 2020b) are integrated for more personalized HRI.

Previous studies provide elegant and simple definitions of human-robot proxemics. However, for different tasks, factors such as work environments, operation area, and robot appearance can change fixed spatial human-robot relationships. In this work, I assume that the HRP is task-specific rather than deterministic. The case study documents an adaptive strategy to construct the HRP model based on empirical data collected from five robot operators, more details about data collection will be elaborated in Section 5.3.1 Human-Robot Proxemics Estimation. The model uses a continuous scale to measure human-robot proxemics, which can potentially avoid crisp transitions caused by discrete proxemic zoning.

5.3 Method

The proxemic-aware AR for the HRI system contains two parts: hardware and software (Figure 5.3). The software system is composed of four components, the proxemic manager, the view manager, the sensor manager, and the environment constructor. The sensor manager component publishes the location and pose of the robot and mobile devices obtained from sensor readings. The proxemic manager component subscribes to the sensor manager topic and outputs the estimated human-robot proxemic level. The UI manager components then adapt to the HRP level by automatically changing the UI content and the level of detail based on

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

a linear integer programming approach. The environment constructor component stores the CAD files of the robot and the reconstruction of the physical environment as a shared spatial coordinate system for both the robot and the mobile device. The communication between each component (node) is built upon the ROS framework.

For hardware setup, I use ABB IRB 6640 for experiments. I used the MoveIt framework and RRT-connect provided by the OMPL (The Open Motion Planning Library) for planning a collision-free path to reach a target position. The AR interface is implemented based on the ARKit framework on a mobile phone. The ROS-Unity is adopted for data streaming between the mobile phone and a remote laptop communicating with the robot via TCP/IP.

5.3.1 Human-Robot Proxemics Estimation

Traditional proxemics define the space within 0.45 m as intimate space. Space between intimate and personal distance is called personal space (ranging from 0.45 to 1.20 m). The space between social and personal distance is called the social space (ranging from 1.20 to 3.70 m). The space within public distance is called the public space (ranging from 3.70 m to 7.60 m)(Edmund T Hall and Edward T Hall 1966). Similarly to human-human proxemics, in human-robot interaction, spatial cues such as distance and orientation can affect how the operator perceives the robot. Like human-human proxemics, human-robot proxemics can be highly context-dependent and non-deterministic. Factors such as the robot's work range and appearance can have a significant impact on how the operator interacts with the robot. For example, users tend to stay a greater distance with a humanoid robot with mutual gaze with it(Mumm and Mutlu 2011b). A robot of a smaller size can be more approachable than a large one (Takayama and Pantofaru 2009). To build task-specific and user-centric HRP estimation systems, I collect the user's a priori information on HRP in a specific context. A user with a mobile device is asked to define the HRP level at a set of predefined locations around the robot. At each point, the user can assign a level of 1-4 (with 1 being the farthest and 4 being the nearest) to describe the perceived HRP (Figure 5.4). Meanwhile, raw features are recorded at each location, including the user's location, user orientation, human-robot relative distance, and relative orientation. The relative distance and orientation are based on a fiducial marker attached to a known location in the workspace. Both the robot and the mobile device use the fiducial marker as a shared landmark to build a coordinate system relative to real-world space.

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

We used the visual-inertial odometry of ARKit to track the real-time pose and location of the mobile device. The robot's location can be obtained through the status message published from the ROS-ABB driver node installed on the robot controller. Each sample point contains the raw features and user-defined proxemics in a set. Based on the collected sample data, we can create a continuous function to map the raw features to the HRP level. Since the collected points can be sparse, we chose to use an inverse distance weighting (IDW) approach.

Given a set of sample points x_i with known proxemic value p_i and $\{x_i, p_i \mid \text{for } x_i \in \mathbb{R}^n, p_i \in \mathbb{R}\}_{i=1}^N$, the IDW interpolation function $p(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as:

$$p(x) = \begin{cases} \frac{\sum_{i=1}^N w_i(x)p_i}{\sum_{i=1}^N w_i(x)} & \text{if } d(x, x_i) \neq 0 \text{ for all } i, \\ p_i & \text{if } d(x, x_i) = 0 \text{ for some } i, \end{cases}$$

where

$$w_i = d(x, x_i)^{-\beta}$$

x denotes an interpolated arbitrary point, x_i is a point with known proxemic value, d is a given distance from known x_i to the unknown point x , N is the total number of known points, β is a positive real number determining the degree to which the nearer point(s) are preferred over more distant points.

To ensure the safety of the operator, it is important for the user to keep a distance from the actual work area of the robot. I set all the points within the reachable area of the robot as level 4. The rest point will be evaluated and rounded to the actual proxemic level.

The output of the HRP model is converted into a fuzzy set through fuzzification. For each HRP level (from level 1 L_1 to level 4 L_4) we use the Gaussian membership function described as:

$$\mu(m, \sigma^2) = \begin{cases} \mu(m_{L_1}, \sigma^2) & \text{if } L_1 \\ \mu(m_{L_2}, \sigma^2) & \text{if } L_2 \\ \mu(m_{L_3}, \sigma^2) & \text{if } L_3 \\ \mu(m_{L_4}, \sigma^2) & \text{if } L_4 \end{cases}$$

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

The value of belongings determines the transparency of the visual elements at that level. When the user moves from one HRP level to another, the visual content of the previous level will gradually disappear when the visual content of the next level shows up, which can provide a seamless transition between HRP levels. To mitigate the interference between translucent visual elements, I created depth masks for visual elements with a larger value of belongings. The masks representing the shape of the visual elements were then used to occlude visual objects that may intersect within the user's view.

5.3.2 Adaptive view management

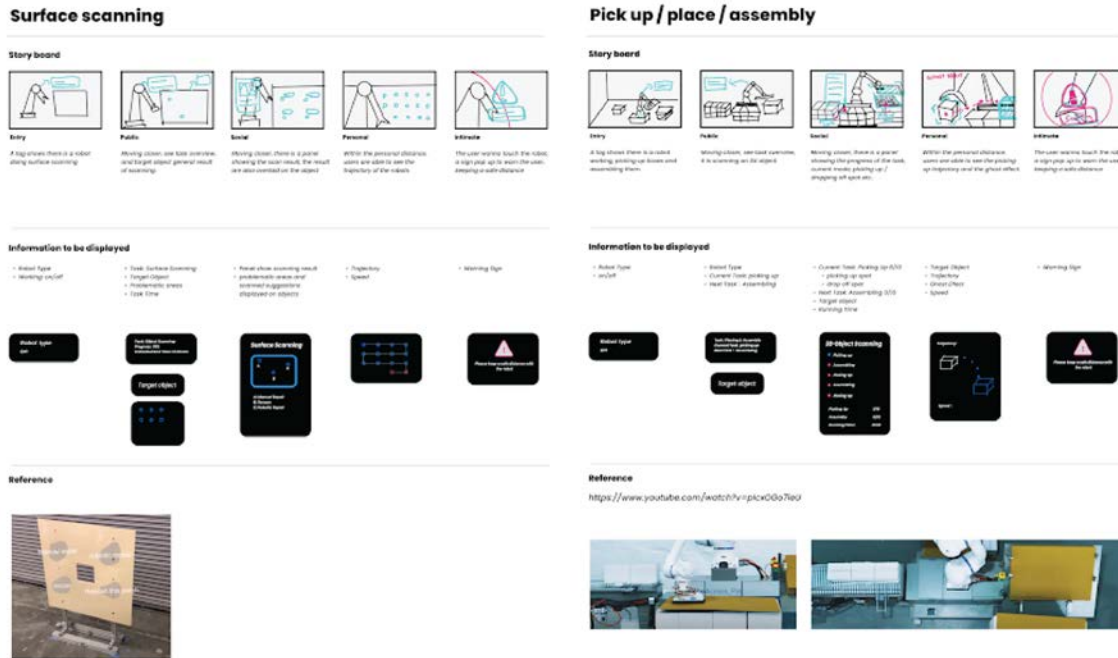


Figure 5.5: Compilation of co-design outputs, including sketches and prototypes, on AR visualization elements for two cases - robotic site scanning and pick & place. Image credit: Author.

A co-design pilot workshop was conducted with a cohort comprising four participants, including three robot operators and one designer, with a collective aim to develop an adaptive visualization system tailored for robotically-assisted construction. The co-design participants are graduate students at Carnegie Mellon University with prior experience in user interface design. Before the start of the workshop, I drafted a document describing the information to

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

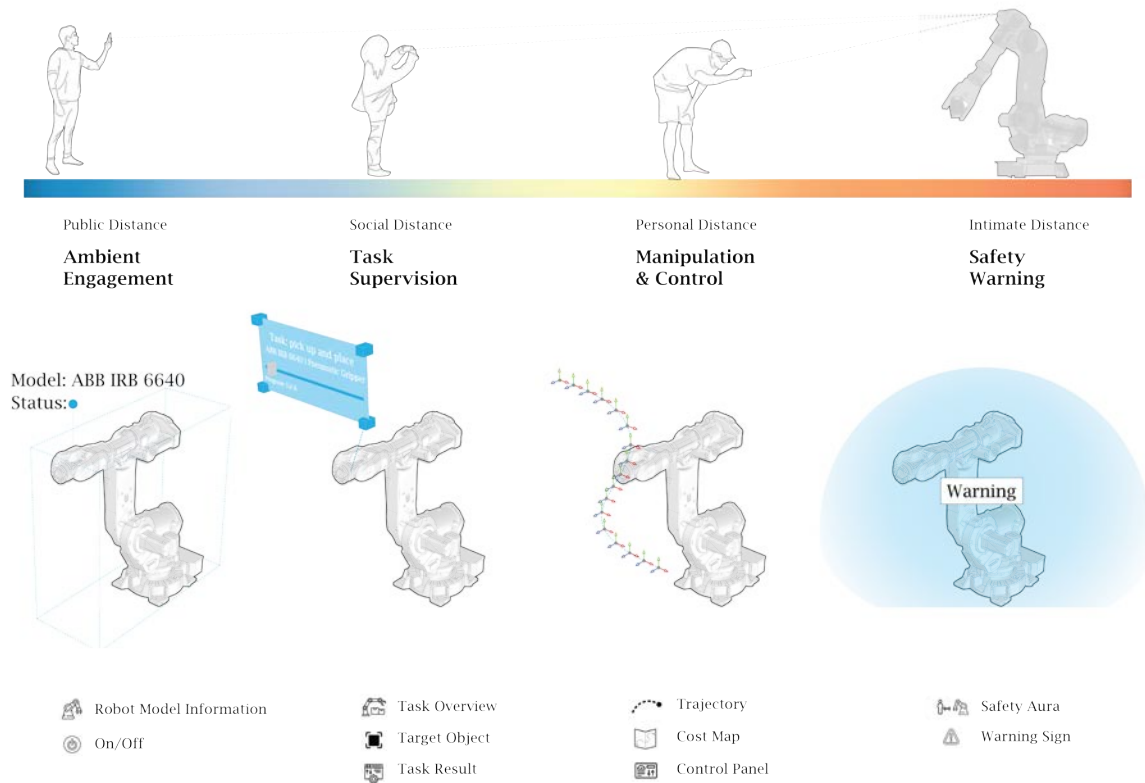


Figure 5.6: Information visualization for various levels of human-robot proxemic - at the ambient engagement level, the AR system provides the overview information of the robot. At the task supervision level, more details of the ongoing task and robot will be revealed. At the third level, manipulation and control, the operator can view and edit the robot's trajectory. When the operator enters the danger zone, a safety aura and a warning sign will be displayed to alert the operator. Image credit: Author.

be visualized in the process of robot manipulation, such as waypoints, safety zones, task descriptions, etc. In the first session, I presented the goal of the workshop - to collaboratively design an augmented reality (AR) interface for robot supervision and manipulation in construction. I then introduced key terms, including path planning, safety zones etc., and presented the precedent examples of user interface elements for tele-operated robotics (Milgram, Zhai, et al. 1993a). Transitioning into the second session, the workshop participants engaged in a brainstorming exercise coupled with sketching. We had initial ideation and then sketched out ideas for information visualization on digital devices. After discussion, for each type of informa-

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

tion, such as robotic trajectory, we chose one or two optimal visual representations. In the third session, based on the sketches, we used a collaborative web tool called "Figma" to create low-fidelity prototypes of their selected design concepts (Figure 5.5). The robot operators were then asked to rate the LoD of each piece of information of early prototypes and provide feedback on the visual element design. Through rounds of collaborative design sessions, the opinions of the designer and robot operators converge into a structured visual template, including a set of visual elements and their LoD. Designers then finalize the visual template by converting the sketches into digital assets for AR systems. There are four levels of HRP zones, each associated with a set of visual elements with a defined LoD (Figure 5.6).

At level 1 — ambient engagement, the system presents an overview of the task and the robot's current state information using labels and annotations. At this level, the detected human-robot proxemics indicates that the operator demonstrates an intention to engage with the robot. The AR system can quickly provide awareness of the robot, such as its model type and operation status, to facilitate the potential human-robot interaction by helping the operator briefly preview the task and the robot to interact with.

At level 2 — task supervision, the AR interface provides task-related information to assist the operator in better supervising the robot's task performance. As the operator moves closer to a robot, based on the human-robot proxemics definition, the operator shows increased engagement with the robot. The AR interface will reveal more details about the task through an information panel. For example, the task name and brief introduction will be presented, and a bar with a slider will be displayed to show the progress of the task.

At level 3 — robot manipulation and control: the AR interface provides information to enable direct manipulation and control of the robot. For example, the work area will be highlighted to show the robot's intent. The pre-operative paths will be displayed for real-time control, and the operator can edit, add, or delete the waypoints to program the robot on the fly.

At level 4 — safety warning: the AR interface presents information to ensure safe robot interaction and avoid collisions. As the operator enters the reachable region of the robot, a safety aura will be displayed to indicate the dangerous zones; the operator can thus avoid unexpected collisions while manipulating the robot.

5.4 Result



Figure 5.7: Experiment and baseline setup for surface inspection and pick & place scenario, the baseline setup displays all virtual information at the same level of details, the experiment setup selectively displays the relevant virtual information tailored to the current human-robot proxemic level. Image credit: Author.

5.4.1 Case Studies

The goal of case studies is to identify the impact of two independent variables - the adaptive view management system and the customized human-robot proxemics (HRP) model on AR-based robot manipulation. In collaboration with students at Carnegie Mellon University, I designed two use cases — 3D scanning and pick & place to investigate (1) the impact of the adaptive view management system on information overloading (2) the difference between the customized HRP model and the deterministic HRP model on task performance (Figure 5.7).

3D Scanning

In the 3D scanning scenario, the participant is asked to operate a seven-axis industrial robot to scan a designated area on a mock-up wall. In the baseline setup, virtual information, such as the inspection region, robot internal status, and guidance waypoints, is displayed at the same level of detail at the start of the task. In the experiment setup, I integrated the adaptive view management system. Based on the human-robot proxemics between the operator and the industrial robot, virtual information is displayed at four levels. At the first level, the operator

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

enters the ambient engagement zone, the AR interface displays the bounding box of the robot, and the model and operating status of the robot will be displayed above the box. As the operator moves closer to the robot and enters the task supervision zone, an information panel will show up with textual information about the task name, the instrument, and the sensors being used, and a sliding bar showing the completion rate of the inspection task. The target inspection region on the mock-up wall will be highlighted. At the third level - robot manipulation and control, the AR interface will display the guidance trajectory and waypoints of the robot. The operator can leverage the guidance to control the robot and inspect the target region. As the operator enters the safe robot interaction zone, the AR interface will display the safety aura with a warning sign to inform the operator about the potential collision.

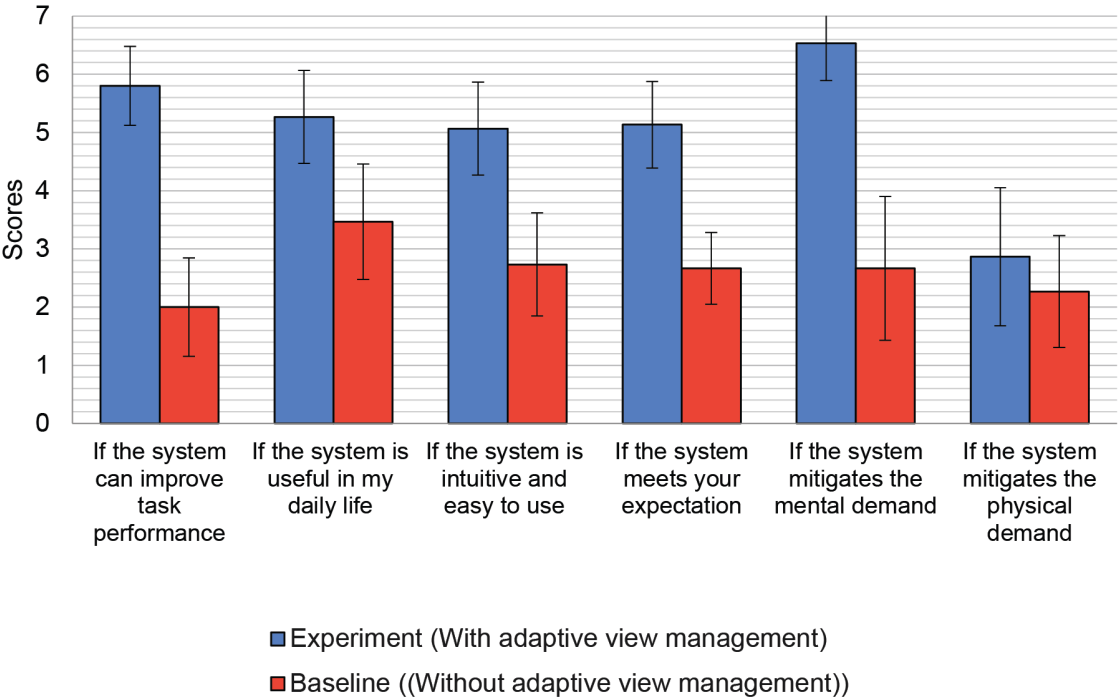


Figure 5.8: Post-experiment questionnaire scores of the adaptive view management system. Error bars indicate the standard deviation. Image credit: Author.

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

Pick and Place

In the pick and place scenario, the participants need to control a seven-axis industrial robot with a suction gripper to pick up a target wood block and place it at a specified location. In the experiment setup, we leverage human-robot proxemics for adaptive information visualization. Similar to the surface inspection case, when the student is in the safe robot interaction zone, the AR interface will display the reachable area of the robot to indicate potential collisions. As the student moves further, the trajectory of the pick and place action will be overlaid with the physical environment. As the student enters the task supervision zone, the AR interface will show detailed information about the ongoing task through an information panel. Finally, in the ambient engagement zone, the AR system will only display the operating status of the robot. Unlike the experiment setup, in the baseline setup, all of the aforementioned digital contents are displayed at the same level of detail.

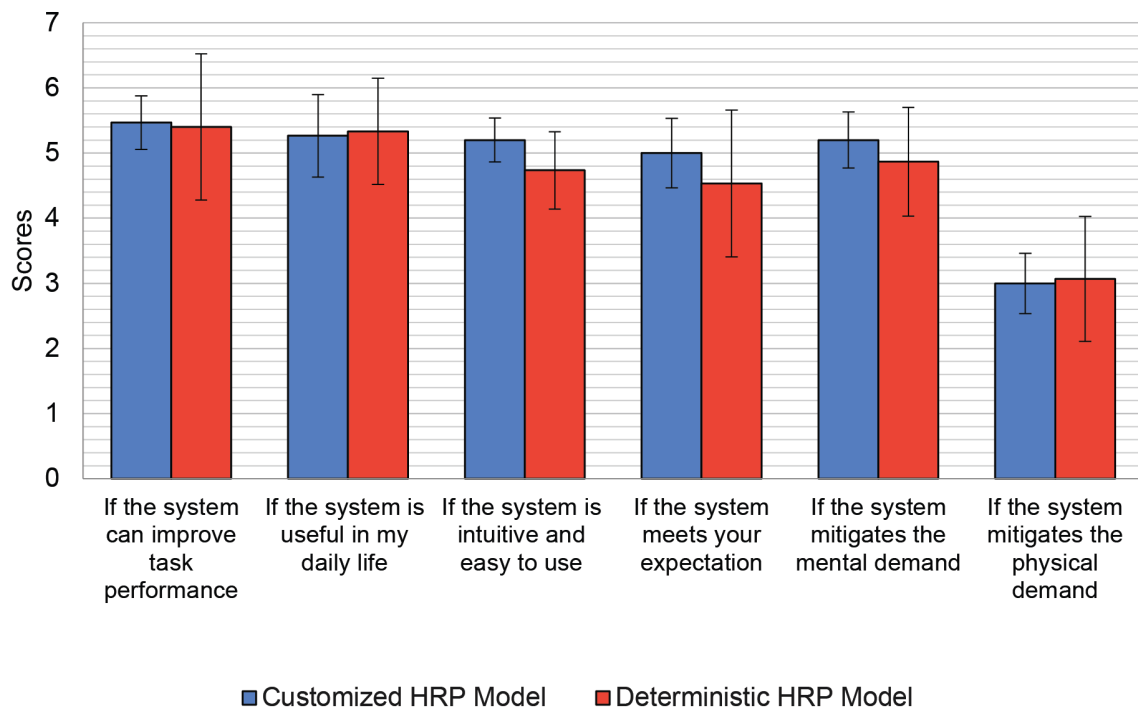


Figure 5.9: Post-experiment questionnaire scores of the customized HRP model and the deterministic HRP model. Error bars indicate the standard deviation. Image credit: Author.

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

Evaluation

To evaluate the efficacy of the proxemic-aware AR system, I used a post-experiment questionnaire to collect user's feedback on the system. In addition to the questionnaire, I logged the task performance data of each use case. In the first task — 3D scanning, the participants were asked to manually control a robot to cover a defined scanning area. The range of the area is visualized through a mobile device with an AR interface. The mobile device is attached to the teach pendant used for programming a seven-axis industrial robot. The participated students can program the robot's orientation and location using a joystick on the teach pendant. To evaluate the impact of the adaptive view management system, I conducted a repeated measures design. In the baseline setup, the guidance information is displayed at the same level of detail. In the experiment setup, the guidance information is visualized in different levels of detail according to the human-robot proxemics.

To compare the differences between the deterministic HRP model and the customized HRP model, I experimented with two models in human-robot proxemics measurement. The deterministic model defines four crisp HRP zones based on human-robot distance unaffected by tasks. The customized HRP model is based on empirical data collected from robot operators (Figure 5.4). The model uses a continuous function with a fuzzy inference to estimate the HRP level, and the definition of HRP zones is seamless and non-crisp. To minimize the order effect, the participants were split into two groups. The first group starts with the deterministic model, and the second group starts with the customized model. The task performance is measured by (1) task completeness: task completeness is defined by the coverage — the ratio between the actual inspected area and the target area (2) manipulation efficiency: this metric describes how many interactions the operator needs to make for an effective motion, which can be measured by the ratio between the number of interactions and the coverage rate.

There are 15 students (7 females and 8 males, After a pilot study, I determined effect size $|\rho|$ as 0.50 , error probability α as 0.20, and power as 0.80. Then we used a priori analysis in G * Power software to estimate the total sample size N , $N = 15$.) participated in the user study. All of the participants have prior experience with AR interfaces or similar interfaces, but none of them was trained for robot manipulation. We provided a training session on robot manipulation before we started the user study. The order of the experiment setup and the baseline setup was switched between tests to counterbalance the order effect that could impact

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

the study result. In each test, I removed the trials when participants claimed an error occurred (between 0 and 5 trials per test).

I found that in the 3D scanning task, the proposed proxemic-aware AR system outperforms the baseline setup in task completeness by 27.2% and motion efficiency by 31.3%. For the pick & place task, the proposed system shows a significant improvement in task completion time. The average time decreases from 51.27 sec ($SD = 11.27$ sec) to 22.18 sec ($SD = 7.11$ sec) with the integration of our proxemic-aware AR system (all $p < 0.01$). However, I did not find a significant effect of our system on the control accuracy. Based on the participants' feedback, the main reason could be (1) it can be difficult to perceive the location of virtual content in the real world, especially when there are occlusions between physical and virtual objects. (2) the participants mostly have limited experience in robot manipulation. It can be difficult for them to control the robot accurately with a joystick, which can be independent of differences in the AR setup.

I did not find a significant difference in task performance between the customized HRP model and the deterministic HRP model. The reason could be that the participants tend to remain static within the HRP zone suitable for robot manipulation until they complete the task. The differences in the definition of HRP zones did not significantly impact task performance.

I used a paired t-test to assess if the result demonstrated a statistically significant difference. The experiment setup yielded statistically significant improvements in the adaptive view management system test according to the post-experiment questionnaire (Figure 5.8). Participants rated the experiment setup ($M = 6.53$, $SD = 0.64$) higher in mitigating mental demand than the baseline setup ($M = 2.67$, $SD = 1.23$), $p < 0.01$. On the question "If the system can improve task performance?", participants rated the experiment setup ($M = 5.80$, $SD = 0.67$) higher than the baseline setup ($M = 2.00$, $SD = 0.84$), $p < 0.01$. On the question "If the system meets your expectation?", the participants rated higher on the experiment setup ($M = 5.13$, $SD = 0.74$) than the baseline setup ($M = 2.67$, $SD = 0.61$), $p < 0.01$. No significant differences were found between the experiment and baseline setup with regard to the physical demand.

I found that the proposed customized HRP model outperformed the deterministic HRP model with respect to intuitiveness (Figure 5.9). Participants rated the customized HRP model ($M = 5.20$, $SD = 0.67$) higher in "intuitive and easy to use" than the deterministic HRP

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

model ($M = 4.73$, $SD = 0.59$), $p = 0.014$. The context-sensitive customized HRP model may contribute to the intuitiveness of the system. For instance, in the pick & place task, the dimension of the active working area is $0.50m \times 0.80m \times 0.50m$. At a distance of $3.50m$, the operators already felt less engaged with the ongoing task and expected to overview high-level information, such as the robot's status. However, based on the deterministic HRP model, a $3.50m$ human-robot distance indicates a higher engagement level (social distance). The AR system presents a detailed introduction of the task with information panels. The panels may distract the robot operator when he or she needs to overview the status of the surrounding robots. The customized HRP model factors in the properties of the ongoing task, which can impact the robot operator's perception of human-robot proxemics. With regard to the question "if the system meets your expectation?", the customized HRP model ($M = 5.00$, $SD = 1.07$) reasonably outperformed the deterministic HRP model ($M = 4.53$, $SD = 1.13$), $p = 0.125$. According to the feedback from the participants, the customized HRP model is a continuous function, and the transition between HRP zones is seamless. The smooth transition is more aligned with the participant's perception of human-robot spatial relationships. However, the participants mentioned that juxtaposing multiple translucent virtual elements could reduce legibility. As an alternative to transparency, we can use the location of the visual elements as an indicator of LoD. For instance, as a robot operator enters a new HRP zone, the associated visual elements will be anchored to the main view, and the less relevant elements will be relocated to the peripheral area.

5.5 Conclusion

Extended from the framework introduced in Section 4 - Method, in this case study, I devised a bespoke AR system with proxemic-aware view management. The system leverages the spatial relationship between the operator and the robot to dynamically adjust the LoD of virtual information in AR interfaces. The case study demonstrates how the spatial relationship in the physical world can be used as an implicit cue for building a context-aware AR for human-robot interaction. The framework provides a basic template with three fundamental functions. Firstly, the site reconstruction module provides a digital representation of the workspace at different levels of resolution. Spatial and contextual data serve as a shared coordinate system

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

for spatial devices involved, including robots and handheld devices. In addition, the multiresolution representation provides an efficient data structure with reduced memory consumption. Especially in this case study, robots and AR devices are synchronized in a shared network. A multiresolution model can reduce the amount of data transmitted over the network. By sending lower-resolution models when high detail is not necessary, bandwidth usage is optimized, leading to faster loading times. The multiresolution representation reduces the latency from the second level to the millisecond level. In addition, the localization module provides important information on the location of each spatial device relative to the shared coordinate system. The spatial location is the critical feature for proxemic estimation, based on the distance between human and robot, the digital content presented within an AR view can be adjusted and adaptive to the current proxemic level. Especially, when a potential risk is identified due to the close proximity between humans and robots, the AR interface can provide a timely warning to the operator based on a precise localization module. As elaborated in Section 4.2 Robust Localization Under Complex Environments, with the integration of UWB and IMU measurement, when the vSLAM failed due to complex lighting environments, the localization module remains reliable. The robust localization module plays a critical role in proxemic estimation and high-level location-based AR experience. The spatial context built upon the framework serves as a foundation for the development of an adaptive AR system that leverages proxemics to mitigate information overload in human-robot interaction.

To build a customized HRP model, I collected empirical data from the participants on-site. The model transforms the conventional model of human-human proxemics which is composed of four discrete zones with fixed thresholds into a continuous function. In combination with a fuzzy inference framework, the proxemic estimation process demonstrated more intuitiveness according to the participants' feedback in the post-experiment survey. I designed a set of AR elements to visualize the robot's internal information and task information at different LoDs. To demonstrate the feasibility of the system, I presented how the proxemic-aware AR system could be applied in robotically-assisted construction tasks such as 3D scanning and pick & place. According to the user study, the integration of human-robot proxemics in AR interface view management can significantly alleviate the robot operator's cognitive load, increase the visualization clarity and enhance the HRI efficiency. I believe that proxemic-aware AR for HRI system can aid robot operators by rendering necessary virtual information that

Chapter 5. Case Study - Proxemic-aware Augmented Reality For Human-Robot Interaction

can not be directly perceived while mitigating distraction caused by information overloading. As a component of the human-robot interface within cluttered construction environments, an adaptive view management system serves to filter out redundant virtual information while presenting the most relevant data at an optimal level of detail, thereby mitigating the risk of information overload. Especially in dynamically changing job sites, extra attention to physical surroundings is required to avoid potential collisions. Instead of overwhelming the operator by presenting all the information up front, according to the participants' feedback in the survey, a gradually revealed and layered view management system is more aligned with the participant's expectations. Moreover, In both 3D scanning and pick & place, the proxemic-aware AR system demonstrates improved performance measured by task completion time. The AR systems underscore their potential in supervising and governing on-site construction robotics through real-time visualization. The visualization of robotic trajectory and task information in physical space can facilitate workers in robot manipulation by providing visual guidance, hazard detection, and building information in construction operations.

Chapter 6

Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

This case study investigates how augmented reality can assist building inspectors to operate robots on-site to identify and quantify surface defects on architectural components (Figure 6.1). Based on the framework introduced in Section 4 - Method, in this case study, I devised a bespoke AR interface to facilitate inspectors to program an industrial robot for surface inspection and preview the inspection result on-site. The development of the AR system is based on the three core modules extended from the framework. Firstly, the site reconstruction module provides a multiresolution digital representation of the site. In this case, I chose to use the low-resolution scene model for tasks such as robot motion planning and navigation, and the high-resolution model for robotically-assisted inspection tasks. Secondly, I tested the localization module, based on which I developed a task-specific interaction modality - spatial annotation. Spatial annotation can assist inspectors in operating robots for defect detection and in-situ visualization. With mobile AR devices, inspectors can define the area of interest over a physical target (Figure 6.2), the area will then be synchronized with a digital model of the real world, the coordinate systems of which are aligned. Based on the annotated area, the system will plan a set of actions and paths for a robot equipped with sensors. The pre-operative robotic trajectory will be visualized in the AR interface for the inspector to review

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

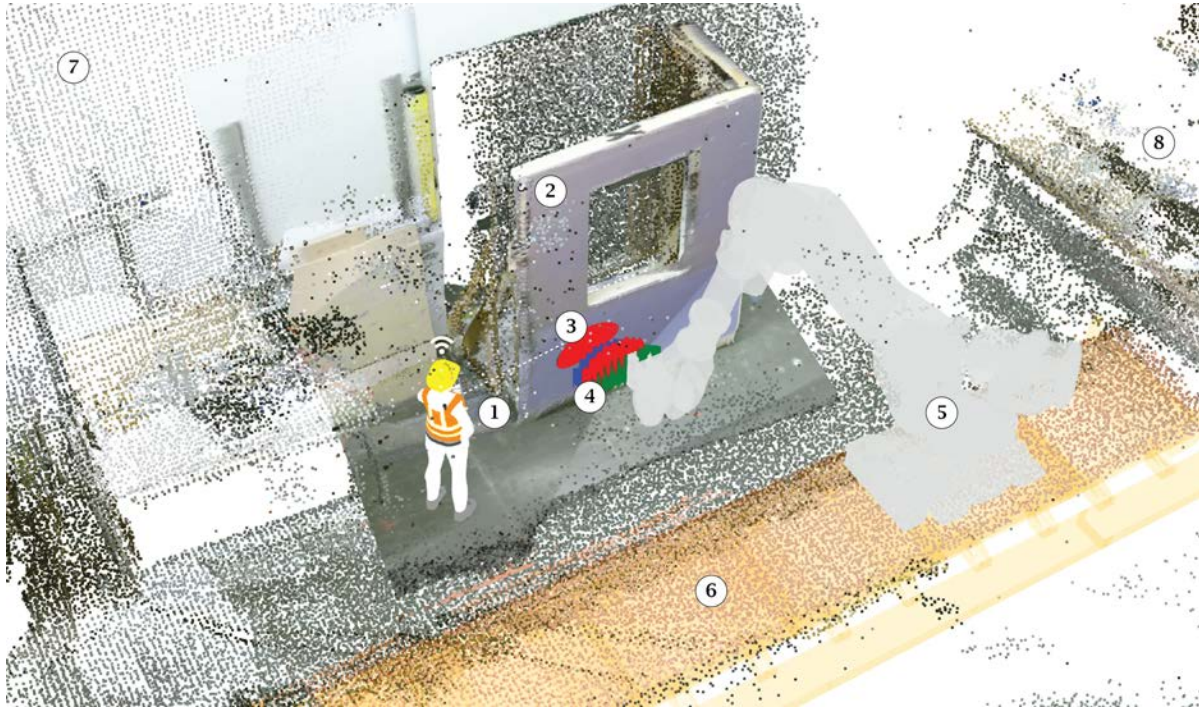


Figure 6.1: Elements of spatial annotation for robotically-assisted quality assessment and management. (1) an inspector with a mobile device, (2) a mock-up drywall for quality assessment, (3) the inspector annotates area of interest through an AR interface, (4) the robot trajectory presented via the AR interface, (5) IRB 6640 industrial robot equipped with sensors (6) linear track (7) Ultrawide-band station for indoor localization (8) a remote workstation for robot controlling and data processing. Image credit: Author.

(Figure 6.3). After the inspector approves the routine, the robot will perform the inspection over the defined area. The robotically-assisted inspection process can digitalize the detected defects for quantitative analysis and conduct repetitive inspection tasks. The spatial annotation interaction technique is dependent on a robust localization system that provides realtime and accurate location estimation of the handheld AR device. The drift and errors in localization can significantly impact the annotation accuracy, which directly affects the user experience. Finally, after the inspection process is complete, the inspection result registered to the real object can be viewed through the AR interface (Figure 6.4). To implement this feature, I used the semantic segmentation module from the framework to build a compact semantic representation over the raw point cloud data. The site model is then represented in multiple surfels, and each identified defect is registered to an associate surfel. When an inspector points to a surface, the

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

inspection result, such as surface anomalies, will be overlaid onto the physical surface. The inspector can measure the dimension via the AR tool and add more annotations. The semantic segmentation module plays an important role in surface selection and annotation placement, which is elaborated in Section 6.2.1 Annotation Placement.

As an information and communication tool, an AR interface can play an important role in future robotically-assisted construction practices by overlaying virtual information, such as controls, indicators, and guidance, over physical environments. The usability test underscores the efficacy of the augmented reality interface in facilitating robot manipulation during inspection tasks. A subsequent qualitative survey involving experts revealed the potential challenges associated with the integration of the augmented reality interface into practical construction practices.

6.1 Background

Robots equipped with sensors can be used for building inspection applications, including but not limited to maintenance inspections, quality control, as-built modeling, process tracking, and safety inspection (Halder and Afsari 2023). Robotically-assisted inspection can provide a safer alternative to manual inspection when the site is inaccessible, confined, or hazardous. Robots have been applied to tasks operating for inspection tasks in pipes (Inzartsev et al. 2019) and ducts (Tan et al. 2021), bridge decks (Khan et al. 2015), and sites after natural disasters (Torok, Golparvar-Fard, and Kochersberger 2014). Additionally, leveraging the recent developments in computer vision, robotically-assisted inspection can provide more accurate visual inspection results in identifying defects like structural cracks, wall cavities, surface finish defects, alignment errors, unevenness, and inclination defects (Prieto, Giakoumidis, and Soto 2021).

Fully automated inspection has been adopted in the manufacturing and automobile industry for decades (Edinbarough, Balderas, and Bose 2005; Fareh et al. 2014). However, the full automation of the inspection process in the AEC domain remains a challenge. Unlike the factory environment, comprehensive apriori site information of the construction environment is not always accessible. The site environment can be dynamic with unforeseen changes. For situations that require complex decision-making, human operators can strategically adjust

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

plans based on situational context, for instance, when unexpected obstacles block the robot, human operators intervene with problem-solving skills to ensure that ongoing inspection tasks proceed. In addition, the domain knowledge of the inspector can help with the interpretation of complex instructions, resolving ambiguities, and providing contextual information for the inspection task. Most importantly, when robotic systems are integrated into construction processes, the guidance and supervision of robots by a human operator who can oversee the environment are necessary for the safety of workers. Robots can contribute to multiple steps in building inspection. However, considering the aforementioned contextual factors, full automation of building inspection tasks remains challenging. Human supervision and control can play an important role in tasks that require human expertise, adaptability, and problem-solving skills.

Some studies have been conducted to investigate the potential of Augmented Reality & Virtual Reality as a human-robot interface in robotically-assisted inspection tasks. For instance, by combining the Building Information Modeling (BIM) with the inspection feed from robots, operators can evaluate the inspection result in a virtual environment and remotely monitor the construction site. On the reverse side, BIM information can be integrated into aerial inspection video for an unbiased and comprehensive condition assessment (D. Liu et al. 2021). AR interface facilitates the interaction between humans and robots by visualizing virtual information, such as task instructions and workflows, in a shared augmented environment.

In this case study, in collaboration with building inspectors from the city of Pittsburgh, I investigated how an AR interface can facilitate human-robot interaction in building component inspection. The user study demonstrates that the system provides the inspector and robot operator with an effective interface for robot manipulation, information visualization, and offsite-onsite collaboration. The case study was conducted with an interior wall mock-up with a 7 DoF industrial robot at the Carnegie Mellon University Digital Fabrication Lab. I used a Kinect V2 sensor for 3D exploration and a combination of an Intel RealSense D415 Sensor and laser range finders for the detection of surface anomalies (Figure 6.5). The proof of concept experiment setup is designed to validate the feasibility of an AR system for robot-assisted inspection. The setup in a lab environment may not encompass the full scope or complexity of on-site building inspection. For example, the existing robot platform, an industrial robot mounted on a linear track, is of limited working range. A mobile platform may be required

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

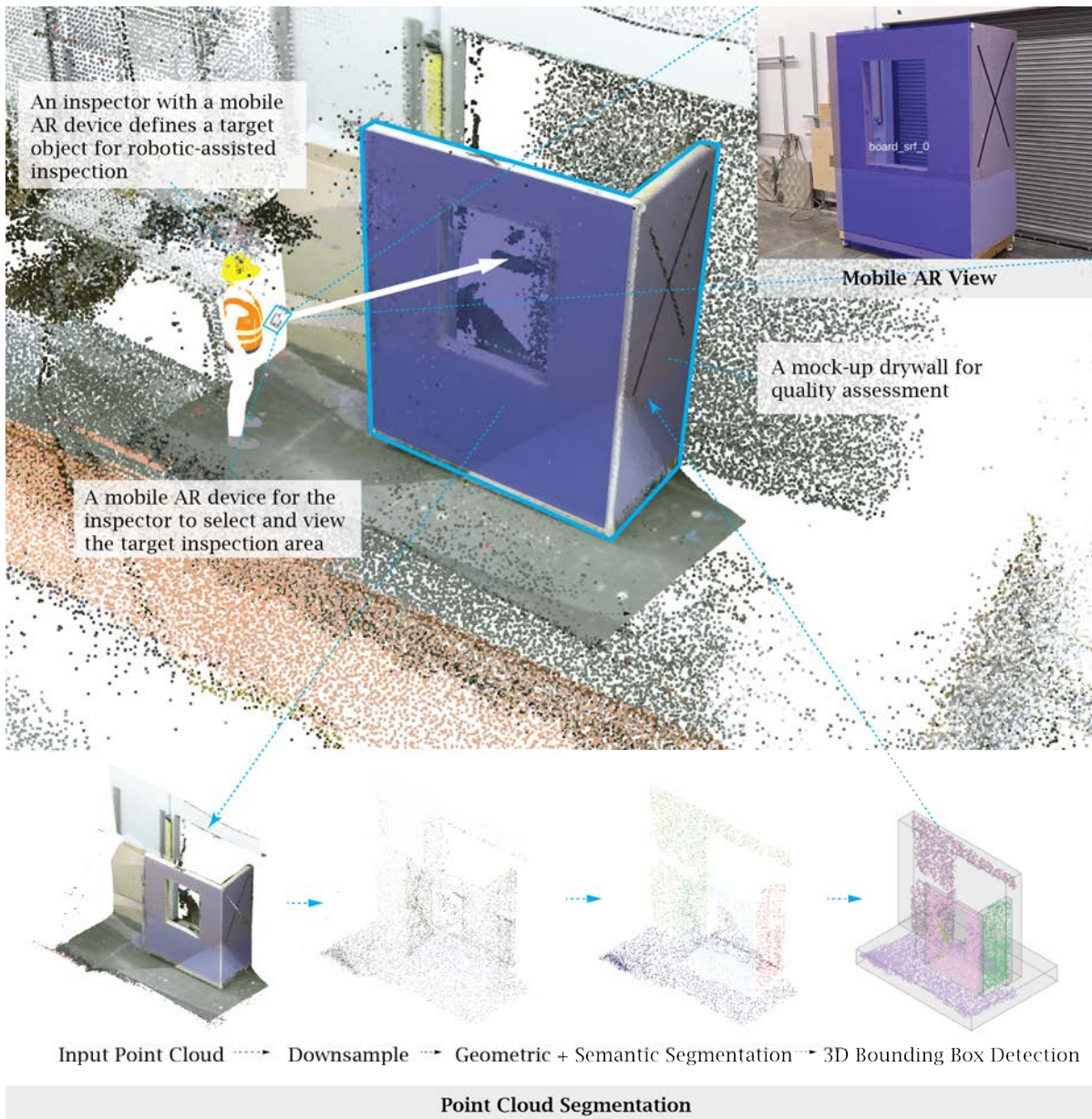


Figure 6.2: An inspector with a mobile AR device chooses a target surface for robotic quality assessment. The captured scene in point clouds is partitioned into surface patches based on a point cloud segmentation pipeline, the inspector can choose a target surface by pointing the AR device to the surface and clicking on the surface within the AR view. Image credit: Author.

with the increase in inspection area. The primary goal of this experiment setup is to build a platform for preliminary data collection on both the technical performance of the robotic in-

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

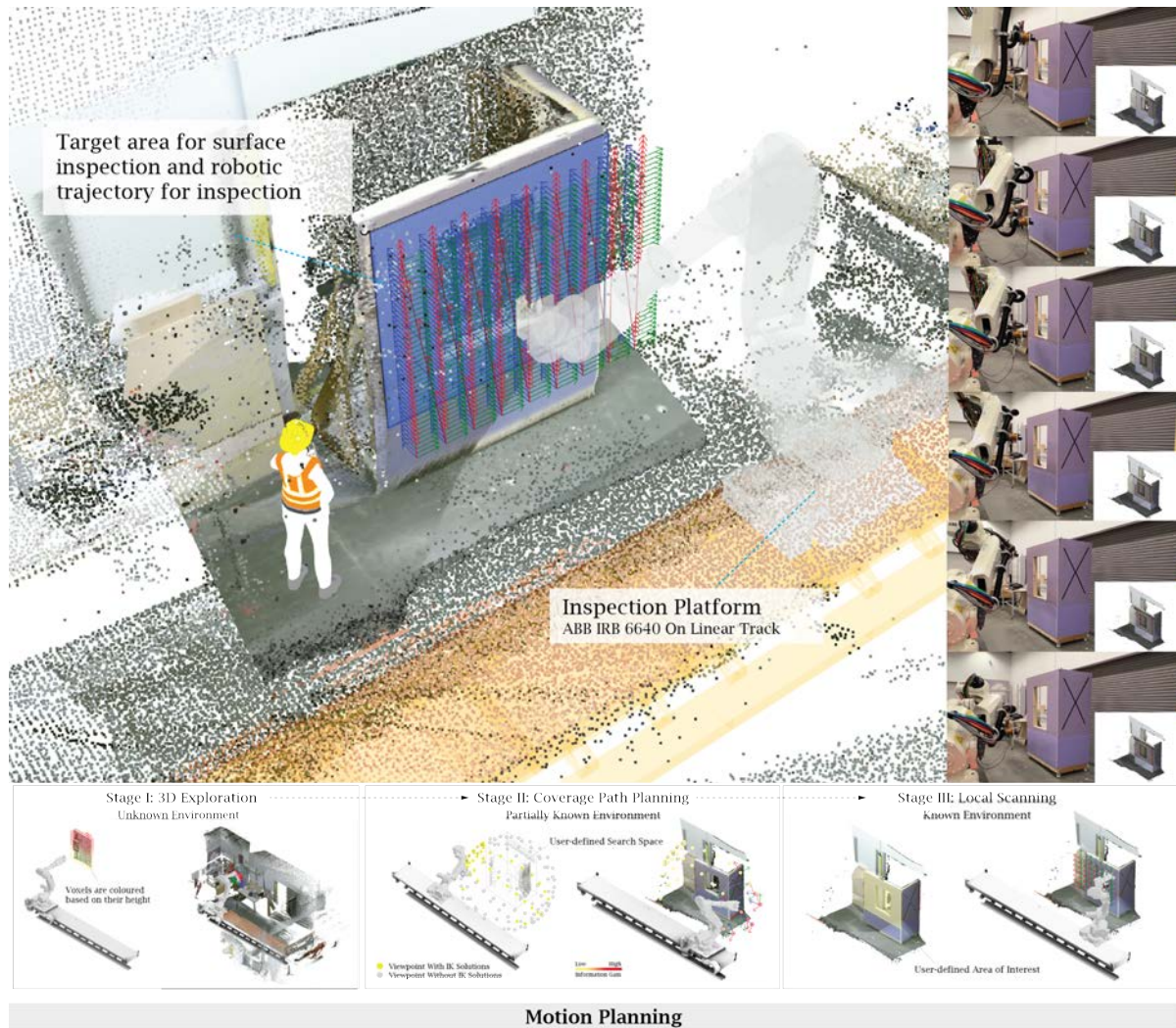


Figure 6.3: After the target surface is confirmed by the inspector, a traversal trajectory will be generated automatically for the robot (ABB IRB 6640) to cover the surface area for defect detection. Image credit: Author.

spection system and the usability of the AR interface. The setup paves the way for further refinement and the implementation of the system in the wild environment.

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management



Figure 6.4: The defect detection result is registered to the physical environment and visualized through an AR interface. Image credit: Author.

6.2 Method

The augmented reality (AR) system is structured based on the framework template outlined in Chapter 4, with extended features on interaction techniques, specifically a novel interactive modality — spatial annotation. Annotation serves as a crucial tool utilized by building inspectors to record, comment on, and highlight various aspects such as defects or areas of concern during manual inspections. Annotation is a widely used approach to documenting visual inspection findings, communicating with stakeholders such as property owners, and tracking inspection progress. Inspectors use paint or pens to mark the location of defects with arrows

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

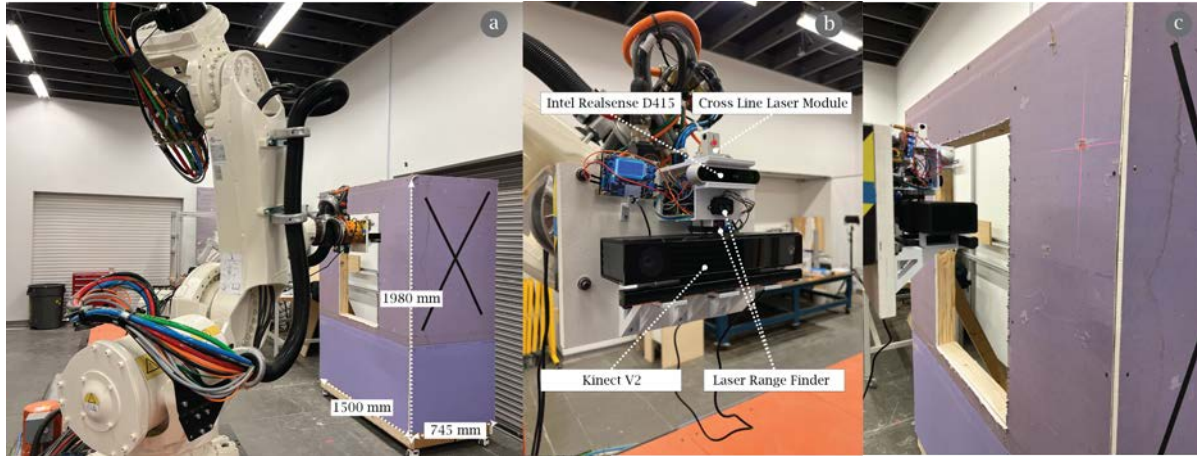


Figure 6.5: Experiment setup (a) a 7 DoF industrial robot on a 5800 mm linear track and an interior wall mockup of 1500 mm (L) x 745 mm (W) x 1980 mm (H) (b) an end-effector sensor kit including RGB-D cameras (Kinect V2, Intel Realsense D415), laser range finders and a laser module (c) the robot highlights a detected surface defect with a laser cross-line. Image credit: Author.

and circles and then transfer the annotation into digital format, for instance, taking a photo and uploading the images of the annotated defect to building assessment and management softwares. To emulate the paper-based inspection process, multiple building inspection software (e.g., Fieldwire by Hilti) and research works (García-Pereira et al. 2020) introduced annotations as an interaction modality for recording and documenting inspection items. Based on the observation of the paper-based inspection process, this study introduces a novel AR-based annotation system for inspectors to define and select inspection areas effectively. After the inspectors confirm the inspection area, the robotic system can autonomously generate a series of poses and near-optimal trajectories to facilitate the inspection process.

As introduced in Chapter 4, point clouds segmentation and indoor localization are two modules required for digital content selection and placement. This study extended the framework introduced in Chapter 4 with additional features for spatial annotation. In this section, I will elaborate on the supplemental modules for spatial annotation, such as annotation placement and the robotically-assisted inspection process, including motion planning, defect detection, and defect registration.

6.2.1 Annotation Placement

The system supports two types of annotation methods: the point-based method and the planar-based method. The point-based interaction enables the inspector to define a partial area of a surface as a target, while the planar-based method facilitates the selection of the entire target surface directly. Both interaction techniques are based on ray casting.

Ray casting is a technique widely used in computer graphics for identifying the intersection between rays extending from a device and a virtual object. We can assume a ray extends indefinitely like a laser pointer. The system selects one of the virtual objects that is closest to the user as the intersected object. Controlling a ray requires the origin point of the ray and its direction. These two parameters can be acquired from the indoor localization module.

The clutteredness of the job site environment can cause selection ambiguity and require repetitive motion to reach the desired accuracy. Ray-casting can be error-prone, especially when selecting small or highly occluded targets (Vanacken, Grossman, and Coninx 2007). Additionally, natural user behaviors, such as hand tremors, breathing-induced body sway, and task-related errors, such as fatigue, also impact the performance of the system. These innate behaviors contribute to increased jitter in the controller, potentially necessitating users to adjust their movements to precisely position a cursor during a 3D pointing task (Batmaz and Stuerzlinger 2019). In spatial annotation interaction, when user physically interact with the handheld device through tapping, the annotation result can be affected and unstable such as instant tilting and rotation. The measurement in localization can also add uncertainty to ray casting process cause errors in spatial annotation and object selection.

I added a disambiguation mechanism to optimize the ray-casting system. When a virtual ray is cast from the user's input device (e.g., a smartphone or AR headset) into the digital representation of the environment, the intersection is calculated based on the heuristic ray method (Haan, Koutek, and Post 2005, Ortega 2013). When the interaction session starts, each target surface will be evaluated based on a spatial temporal function.

$$s_t = s_{t-1}\lambda + (1 - \frac{\alpha(t)}{\beta})(1 - \lambda)$$

Where S_t denotes the score at time t , $\alpha(t)$ denotes the angle between the ray and the center of the target at time t , β is the threshold and λ is the time decay weight. The surface with the

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

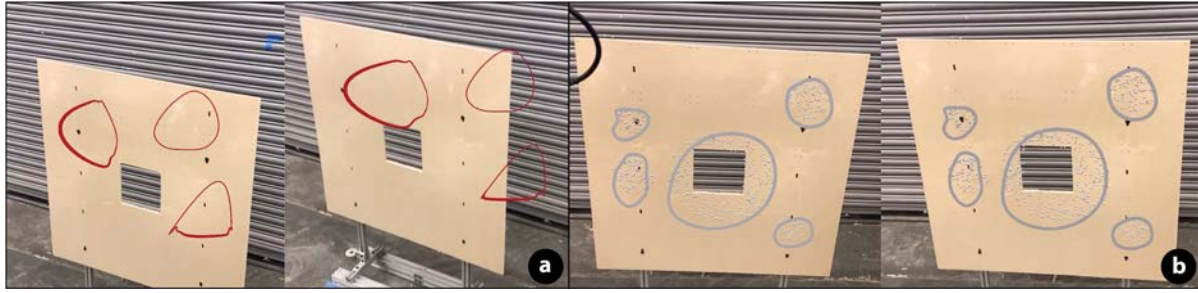


Figure 6.6: The comparison between annotation with and without disambiguation mechanism: (a) without disambiguation mechanism, digital annotations drift while the operator moves around the target (b) with disambiguation mechanism, digital annotations anchored to the active surface invariant to the location of the AR device. Image credit: Author.

highest score in the implicit scoring function will be set as the target surface, the target surface will be highlighted, and the operator can then press the button on the screen to confirm the selected surface. The selected surface will be defined as active, and the rest of the surface will be set as inactive. All the interaction processes and annotation would be attached to the active surface. As we can see in Figure 6.6, with the disambiguation mechanism, digital annotations can be associated with the physical object of the target, while the AR device changes location, digital annotations remain attached to the target board.

6.2.2 Motion Planning

Extended from the framework described in Chapter 4, after the inspector defined the area of inspection, the motion planning component automatically generates a collision-free path for a robot to reconstruct the site scene and inspect the area of interest. The motion planning problem can be formulated as a coverage path planning (CPP) problem or a variant of the well-known traveling salesman problem (TSP) — an agent needs to pass over a set of target points with minimum cost (usually time or path length). However, in contrast to the traditional TSP where the environment is stationary and known a priori, in site surface inspection operations, full prior knowledge of the environment might not always be accessible. For the case study, I use a three-phase approach including 3D exploration, coverage path planning, and local scanning for reconstructing a multi-resolution scene model of an unknown environment for robotic inspection.

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

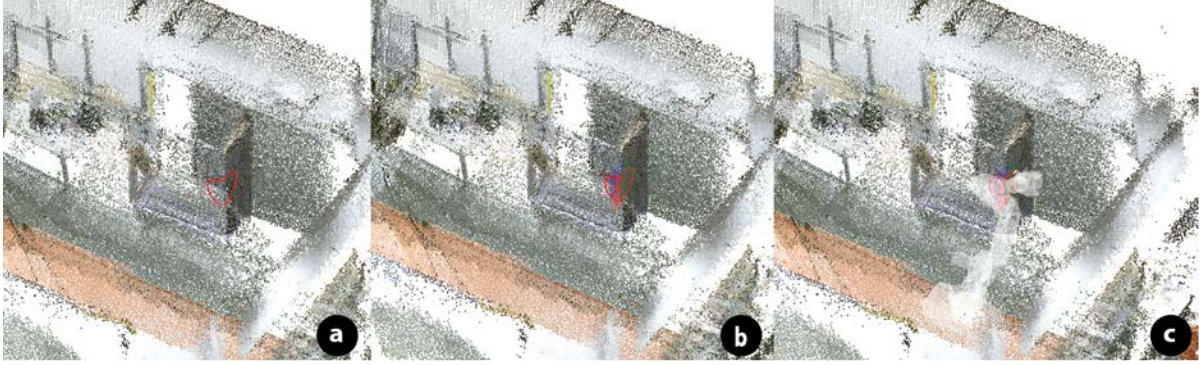


Figure 6.7: Motion planning for local scanning (a) the user-defined area of interest is synchronized from mobile device to the shared virtual model (b) the motion planner sample target over the area and calculate the shortest traversal path (c) the planner calculates joint value for industrial robot to reach the target. Image credit: Author.

In the initial exploration phase, the robot explores the space by finding a set of poses along the frontier of the unknown environment and creates the occupied space map (a volumetric representation of space in a hierarchical structure). The robot first looks for a frontier located on the boundary between the explored and the unknown part of the environment, and then chooses the one with the maximum information gain weighted by its cost to reach C. Zhu et al. 2015b Lu, Redondo, and Campoy 2020b, which can be defined as:

$$I_f = \frac{N_0}{N} \cdot \frac{1}{|P_f - P_0|},$$

where I_f is the information gain of a candidate frontier cell. P_f and P_0 represent the coordinates of the candidate frontier cell and the current pose respectively. N_0 represents the unknown voxels within a sphere of radius R around the candidate frontier that contains N voxels.

After the goal frontier is selected, the planner generates a path to move the robot 30 mm towards the frontier and look towards the direction of the frontier through an Inverse Kinematic (IK) solver unless the planner can not find a valid IK solution. The exploration procedure terminates when the void voxels within the bounded space are classified as either free or occupied.

The 3D exploration phase reconstructs a partially known scene model of the site (93.67% coverage rate) at a relatively low resolution (50 points per m^3). Based on this model, the

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

inspector can define a smaller search space for the second phase of coverage path planning. The coverage path planning aims at automatically generating collision-free robotic paths covering the required surface area within the target volume. I first combine the voxels in the volume into a binary status — “occupied” and “unoccupied” Vasquez-Gomez et al. 2014b Blaer and Allen 2007b. The system then uniformly samples a set of viewpoint candidates on a sphere to capture the most variations of perspectives that can cover the target volume Isler et al. 2016b. The quality of each viewpoint candidate is defined by measuring the mean entropy $e_{(x)}$ over all voxels within the sensor frustum of a potential next pose, which can be defined as Daudelin and Campbell 2017b :

$$e_{(x)} = -\frac{1}{k} \sum_{i=1}^k p_i(x) \log p_i(x) + p_i(x^c) \log p_i(x^c),$$

where $p_i(x)$ is the probability of voxel x being occupied, and $p_i(x^c)$ $p_i(x)$ denotes the complement probability of $p_i(x)$, i.e. $p_i(x^c) = 1 - p_i(x)$. Based on this volumetric function, we select 10 candidates with the highest visible uncertainties for coverage path planning.

After the exploration phase, I used a region-growing approach based on smoothness constraints to cluster point clouds into surface patches. The inspector can then choose a target surface or an area of interest on a surface for local scanning (Figure 6.7). To generate the local scanning path, I used an offline path planning algorithm formulated as an optimization process to minimize the cost function by adjusting decision variables such as candidate viewpoints and their traversing sequence. The function is defined by the total length of the robot trajectories that cover the target area. An online path planning process is then integrated to address adaptive motion planning in real time to avoid collisions caused by tolerances in reconstruction.

6.2.3 Defect Detection

There are several types of typical drywall defects from pre-installation to the maintenance stage, including cracks, warping, moisture damage, uneven joint finishing, and fire resistance rating. In this study, as a proof of concept demonstration, I focus on two types of defects — geometric deformation and surface anomalies — before surface finishing and painting and after gypsum board installation. Geometric deformation can be caused by irregularities in the underlying substructure and the failure of screwing onsite. The warping of the board due to

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management



Figure 6.8: Three types of defects on the dry wall component. Image credit: Author.

poor storage can also affect the drywall surface evenness (Gaião, Brito, and Silvestre 2011). Failures in screwing, such as loosening and recession, can cause surface discontinuities and result in bulges on the wall after surface finishing. Surface anomalies such as discoloration, cracks, and the detachment of the board covering can affect the aesthetics of the interior wall surfaces and cause the degradation of the structure, which needs to be repaired before the application of surface finishing materials (Figure 6.8).

To detect geometric deformation, I first use a principal component analysis-based approach to fit a plane within the selected point cloud. Then we can calculate the orthogonal distance between each point and the plane to evaluate surface flatness errors and calculate the connecting angle between two adjacent surfaces by computing the angle between their normal vectors.

Surface anomalies on gypsum boards can be visually detected by changes in color on the surface. Discoloration, scratches, and cracks often have clear boundaries and/or rougher texture than the normal areas. I apply Contrast Limited Adaptive Histogram Equalization (CLAHE) to blurred images to reduce the effect of lighting changes. Then I extract edges with a Canny edge detector and use morphological operations to cover the regions with dense edges, which correspond to the rougher texture of surface anomalies. I also use the depth information to discard the pixels with depth greater than 50 cm, since they are not on the wall. With the remaining regions, we filter out the regions with contour areas greater than one standard

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

deviation above average and classify them as surface anomalies. We also notice that surface anomaly detection and screw detection work optimally at different distances from the gypsum board.

As construction sites can be affected by complex and unpredictable lighting environments, to compensate for illumination variations and shadows, I combine both 2D and 3D data for defect detection. 2D image processing can extract the potential defects in images. 3D geometric information is used to detect volumetric defects defined by sharp changes in a depth reading. Depth reading is acquired by a fusion of a stereo camera and laser range finder to compensate for each individual sensor's deficiencies — stereo vision can produce dense output, but performs poorly on textureless surfaces or regions with repetitive patterns, and the data collected from the laser range finder is accurate but relatively sparse in nature Huber and Kanade 2011. The defects extracted from 2D and 3D data are fused at the decision-making level based on evidence theory — a general framework for modeling epistemic uncertainty for multisensor fusion.

6.3 Results

6.3.1 System Usability Test

The AR interface was tested by a total of seven participants. Five of them have reported to have very little or no experience using AR systems. Two of them have experience in the building inspection process as inspectors. They were asked to use the AR tool for a lab experiment using a six-DoF industrial robot mounted on a linear track for surface inspection of an L-shaped mock-up wall in the dimension of $1500mm(L) \times 745mm(W) \times 1980mm$. All the participants were asked to do the following tasks (Figure 6.9):

1. Define a surface of object for inspection through a mobile AR device.
2. On a defined surface, the participant was asked to annotate an area of interest for detailed inspection.
3. The participant was asked to use AR tool program robot navigate through tagged areas defined by an offsite user.

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

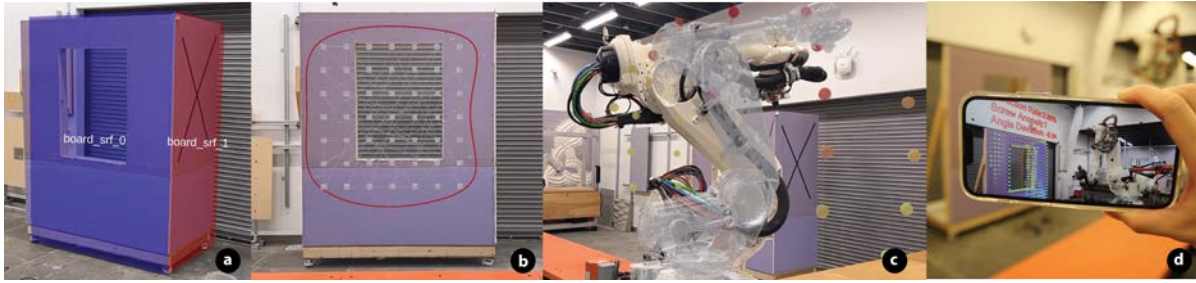


Figure 6.9: The four user study tasks: (a) surface selection (b) area of interest definition (c) robot trajectory preview (d) result preview and defect measurement. Image credit: Author.

4. After the robotic inspection, the participant was asked to choose a defect and quantitatively measure the dimension of the defect through the AR interface.

To validate the system, the participants were asked to complete a test of user satisfaction and provide qualitative comments on the performed tasks.

Table 6.1: SUS questions

Questions
Q1. I think that I would like to use this system frequently.
Q2. I found the system unnecessarily complex.
Q3. I thought the system was easy to use.
Q4. I think that I would need the support of a technical person to be able to use this system.
Q5. I found the various functions in this system were well integrated.
Q6. I thought there was too much inconsistency in this system.
Q7. I would imagine that the most people would learn to use this system very quickly.
Q8. I found the system very cumbersome to use.
Q9. I felt very confident using the system.
Q10. I needed to learn a lot of things before I could get going with this system.

A usability test is a method used to evaluate the ease of use and user-friendliness of a product, system, or service (Brooke 1996). The System Usability Scale (SUS) is a widely used tool for measuring the perceived usability of a system. It consists of a 10-item questionnaire that users complete after interacting with a system (Table 6.1). Each item is rated on a 5-point Likert scale, ranging from "Strongly Disagree" to "Strongly Agree." The scores from the SUS questionnaire are then calculated and normalized to produce a usability score ranging from 0

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

to 100. I chose the SUS as an evaluation tool due to several reasons, firstly, the SUS can provide meaningful usability insights with relatively small sample sizes (Klug 2017). Secondly, the SUS can provide customizability to the context, following the standardized SUS format, researchers can add questions to gather more detailed feedback on specific aspects of usability or user experience. In this study, to evaluate the spatial annotation feature, I added four more questions, which are depicted in Table 6.2. Finally, the SUS study has demonstrated good reliability and validity in studies across different domains for the past decades (coefficient alpha of 0.91) as a “highly robust and versatile tool for usability professionals” (Bangor, P. T. Kortum, and J. T. Miller 2008).

Table 6.2: Additional question to evaluate the AR interface

Questions
Q11.I believe that the AR interface can reduce the time in defining areas for robotic inspection.
Q12.The AR interface and robot allow me to perform the inspection task with more convenience.
Q13.I believe the system can be applied and suitable for my work in real environment.
Q14.I think that the spatial annotation provides an intuitive way for defining inspection areas.

The user satisfaction test is formulated in a way with varied tones. Benchmarks’ directionality differs for odd- and even-numbered items. Higher scores indicate better usability for odd-numbered items, while lower scores indicate a more positive user experience for even-numbered items. This strategy can help researchers identify invalid responses caused by participants who may not be fully engaged with the question statements.

The user satisfaction test is formulated in a way with varied tones. Benchmarks’ directionality differs for odd- and even-numbered items. Higher scores indicate better usability for odd-numbered items, while lower scores indicate a more positive user experience for even-numbered items. This strategy can help researchers identify invalid responses caused by participants who may not be fully engaged with the question statements (Lewis 2018). For odd-numbered questions (1, 3, 5, 7, 9), the score is determined by subtracting 1 from the selected rating. For even-numbered questions (2, 4, 6, 8, 10), the score is determined by subtracting the selected rating from 5. Once the scores for all 10 questions are calculated, they are added

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

together to obtain a total score. The last step is to normalize the score to 100-point scale, multiply the total score by 25, which can be represented by:

$$SUS\ Score = ((Q_1 - 1) + (5 - Q_2) + (Q_3 - 1) + (5 - Q_4) + (Q_5 - 1) + (5 - Q_6) + (Q_7 - 1) + (5 - Q_8) + (Q_9 - 1) + (5 - Q_{10})) \times 0.1 \times 25$$

Where $Q_1, Q_2, \dots, Q_{10}$ are the ratings for each of the 10 questions. The 100-point scale provides a standardized way to compare and assess the usability of a system.

Table 6.3: Users (U_i) response to the SUS questions (Q_i)

	U1	U2	U3	U4	U5	U6	U7	Min.	Max.	Average	S.D.
Q1	4	5	3	4	5	5	5	3	5	4.29	0.76
Q2	3	1	1	1	1	1	1	1	3	1.42	0.79
Q3	4	4	4	5	4	5	5	4	5	4.29	0.48
Q4	3	2	2	1	2	2	2	1	3	1.86	0.69
Q5	4	5	5	5	5	5	5	4	5	4.57	0.53
Q6	1	1	1	2	1	1	1	1	2	1.43	0.53
Q7	4	4	4	5	4	5	5	3	5	4.14	0.69
Q8	2	1	1	1	1	1	1	1	2	1.29	0.49
Q9	3	5	5	5	4	4	4	3	5	4.29	0.76
Q10	2	2	2	1	3	1	1	1	3	1.71	0.76
SUS	70.00	90.00	80.00	95.00	85.00	95.00	77.50	70.00	95.00	84.64	9.40

The 10 SUS statements which have been validated can be seen in Table x, the individual scores given by users and the calculated SUS score are shown in Table 6.3. With regard to the individual questions, Q7 — “*I found the various functions in this system were well integrated.*” received the highest value (Avg.= 4.57). Q4 — “*I think that I would need the support of a technical person to be able to use this system.*” (Avg. = (5.00 – 1.86) + 1.00 = 4.14) and “Q7 — *I would imagine that the most people would learn to use this system very quickly.*” (Avg. = 4.14) received the lowest score. Questions (Q1,Q2,Q4,Q9,Q10) all received a neutral score. The average score for all questions is above 4. The average SUS score for the system is 84.64, within the range from 0 to 100 (100 is the best result). Indicating by the score, the system can be considered as the level of “excellent” according to the SUS score

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

scale (Bangor, P. Kortum, and J. Miller 2009). In this SUS test, products that scored in the 90s were exceptional, products that scored in the 80s were good, and products that scored in the 70s were acceptable (Brooke 2013).

Table 6.4: Users (U_i) response to the supplemental questions (Q_i)

	U1	U2	U3	U4	U5	U6	U7	Min.	Max.	Average	S.D.
Q11	4	5	4	5	4	5	5	4	5	4.43	0.53
Q12	4	4	5	5	4	4	4	4	5	4.29	0.49
Q13	3	5	3	4	4	4	4	3	5	3.86	0.69
Q14	4	4	4	5	5	5	5	4	5	4.43	0.53
Average	3.75	4.50	4.00	4.75	4.25	4.50	4.50	3.75	4.75	4.25	0.35

The added supplemental questions can be seen in Table 6.4 . Ranging from strongly agree (5 points) to strongly disagree (1 point), the available response choices are aligned with the SUS questionnaire. However, all questions are formulated in a positive way, the highest achievable score for each question is 5 points. I analyze the correlation between two set of questions, the person correlation between the SUS score and the average score of additional questions is $r = 0.97$. U1 and U4 gave the lowest the highest score for both the SUS questions ($SUS = 70.00$, $SUS = 95.00$) and the additional questions ($Avg. = 3.75$, $Avg. = 4.75$), respectively. The individual scores given by users for additional questions can be seen in the Table 6.4.

6.3.2 Robotically-assisted Inspection Result

In addition to the user study, I document the result of robotic inspection and consult the result with the inspector to compare the difference between manual inspection and robotic inspection. According to the robotic surface inspection results (Figure 6.10), the mockup wall has 8.18% area identified as not flat (deviation from the fitted plane > 10 mm). The maximum displacement on the drywall is 28.95 mm. The angle between the two surfaces is -0.54 degrees deviating from the expected angle (90 degrees). The total area of surface anomalies is 85661.28 mm², accounting for 3.86% of the total surface area. 1 out of 33 of screws detected is identified as loose. The inspector reviews the result through the AR interface and approves the surface anomalies with an area of 82810.14 mm², the non-flat area approved by the inspector accounts for 8.03% of the total surface area, and the detected screw failure aligns with the

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

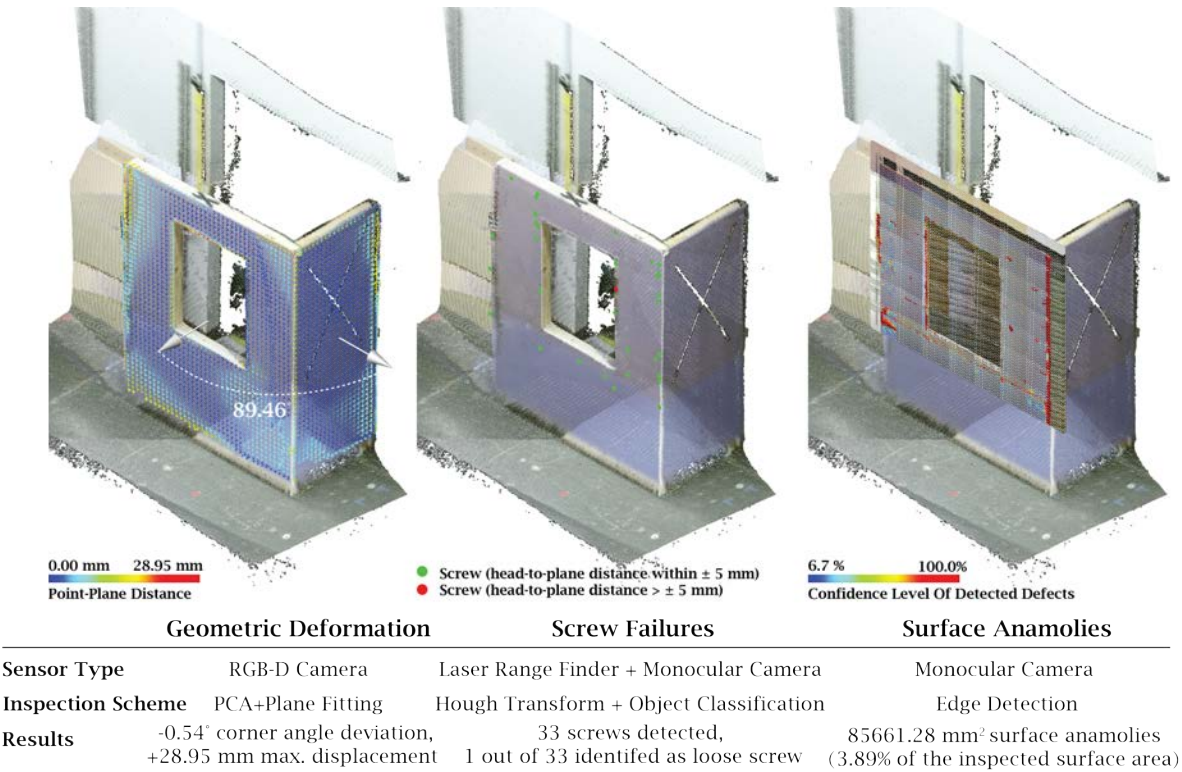


Figure 6.10: The robotic quality assessment result of the interior wall mockup. Image credit: Author.

result of manual inspection. In terms of the correlation between manual inspection and robotic inspection, I compared the performance of inspection results of two types of drywall defects - screw failure and surface anomalies, the overall correlation rate is 98.27%.

6.4 Discussion

After quantitative evaluation, I gathered qualitative feedback from participants regarding the augmented reality (AR) system and possible enhancements to the tool. In this subsection, I will provide a brief discussion of the system, from three perspectives: user experience, technical barriers, and contextual constraints.

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

6.4.1 User Experience

The AR for HRI content is presented through a handheld mobile device in this study. The participants report that they can feel fatigue after holding the mobile device for 5 minutes, when they need to put the phone around 30-40 cm away from the body, aligning with the eyes. U5 suggests that some the mobile device can be combined with wearables, such as exoskeletons to alleviate fatigue. The gesture and the weight of the device can affect the duration of spatial annotation. U3 mentioned that when annotating the object, one of hands need to stabilize the device at one side, and the other hand needs to apply pressure to the screen. The extra strength to avoid tilting can cause fatigue. The handheld device can slide and affect the accuracy of annotation. In the future iteration, additional hardware for fixing and balancing the handheld device can improve the user experience in interaction. For example, a case for gripping and handling the handheld device can mitigate the discomfort in spatial annotation. U6, with inspection experience, mentioned the need for the handheld device to attach to the inspector's body or clothing to avoid falling and interfering with ongoing tasks. In the future study, more intuitive interaction techniques can be further investigated. For example, tangible interfaces can be integrated to mitigate the reliance on screen-based interaction. Inspectors may use visual elements, such as post-it notes, as a cue for designating and annotating the target inspection area. The visual elements could serve to guide robotic systems in traversing the site and conducting inspections in regions identified by visual markers. By combining sticky notes with augmented reality (AR) interfaces, we can exploit the familiarity of physical objects within a digital context, thereby fostering a more intuitive and effective visualization and annotation experience within the physical environment.

Some participants mention about the screen size of the AR interface. In spatial annotation, the touchscreen is the primary interface to interact with AR content. The screen size can be a limit for the active working range. When performing the task of annotating area for local scanning, the area can extend out of the physical boundary of the screen. In the process of annotation, the user needs to put fingers constantly on the touchscreen, which may occlude some details of the physical view. When the site is cluttered, it can be challenging to highlight the small target. Precise interaction in spatial annotation has been reported as not easy by the participants. Especially when participants experience fatigue, it is difficult for them to stabilize the handheld device, U3 suggested it could be helpful if the AR interface could be locked when

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

the user needs to interact with small virtual objects precisely, U4 suggested that stylus can be an option for precise interaction. U6 reported the lighting conditions can affect the readability of the annotation, additional coatings can be applied to the screen to mitigate the visibility issues caused by ambient illumination.

Even though the majority of the participants had limited experience with AR technologies before, they reported that the system is easy and intuitive to use. The participants with inspection experience report that the interaction technique, spatial annotation, can lower the barriers to robotic manipulation. In terms of spatial annotation, the participants mentioned that compared to two-dimensional interaction, spatial interaction involves more degree of freedom. It increases challenges for precise interaction. A feature of recording the key frame can be helpful for users to record the location and edit the previous annotations efficiently. U4 mentions social acceptance of the spatial annotation, during the process of interaction, the user needs to constantly point the handheld device to the surrounding environment, which may cause concerns of privacy.

6.4.2 Technical barriers

Technical challenges in spatial annotation result from two perspectives, accuracy and efficiency. Accuracy depends largely on (1) the localization of robots and AR devices, as described in Chapter 4; environmental conditions, such as lighting conditions, can significantly impact the localization of devices. The addition of a UWB-based localization system significantly reduces drift errors in localization (2) the accuracy of the user interaction technique. In this study, inaccurate spatial annotations can result in deviations in robotic path planning, further affecting the inspection result. The efficiency of the system is dependent on network latency; the computation of the system is offloaded and stored to a workstation to overcome the restrictions of mobile devices. Virtual and real-world elements are synchronized through a local wireless network. The offloading latency can range from 250ms to 500ms, depending on the complexity of virtual visual elements. The system architecture and distribution of offloading can be further optimized to minimize latency within 100ms for a smooth interface. Lastly, the point cloud segmentation algorithm works mostly for physical environments composed of planar surfaces. In environments with complex geometry, the accuracy of the system can be affected by the inaccurate segmentation result.

6.4.3 Challenges and limitations

From the interview with the inspectors, the integration of AR tools and robots into inspection may bring several benefits and concerns. The result generated from robotic inspection and the AR interface can provide inspectors with quantitative information about the defects and assist them in their decision-making processes. The system can be potentially combined with building management systems for the full building life-cycle management. The robotically-assisted building inspection can supplement traditional inspection documentation such as anomaly repair forms and building pathology sheets with an accurate, time-efficient, and automated means of fault detection in buildings and other structures. However, several factors need to be considered for the future application. Inspectors mentioned the impact of the system on job opportunities. If the accuracy and efficiency of the AR tools and robots outperform human beings, inspection tasks can be conducted by people with less or no expertise. In terms of safety, holding a handheld AR tool can be a distraction for inspectors while walking on a job site. With regard to accuracy, the errors caused by the robotic system and AR tools, such as inaccurate or excessive information, may mislead the inspector. The robotic system errors may result from several perspectives, firstly, the deviation in sensing and perception may lead to inaccurate results. For example, in this case study, screw failure detection can be affected by the identification of screws process based on a data-driven pipeline. The lighting environment can affect the accuracy of screw detection, which may further impact the result of screw failure inspection. Additionally, the navigation and localization of robots may pose challenges to robotically-assisted inspection in the wild. Variations in terrain, the presence of obstacles, and weather conditions can all lead to errors in localization and failures for robots to traverse the designated inspection areas. For complex environments, human operator guidance and oversight are still needed for robot navigation. Finally, hardware failures can contribute to errors in data collection and analysis. In this study, vision-based sensors need maintenance in each set of experiments due to dust accumulation, and the low battery power of UWB-based stations can negatively affect localization accuracy. In addition, the participant mentioned potential issues in surveillance, the location tracking of the AR device without the user's consent can cause concerns in data privacy. Finally, inspectors also mentioned the cost-efficiency of the system, the hardware, the development costs, education costs, and the scalability of the system

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

needs to be taken into account, especially since the AR tool and robotics are still novel for inspection practices.

In conclusion, this case study illustrates how the technical framework can be extended for developing a human-robot interface in robotically-assisted surface inspection. The site reconstruction module provides a digital representation of the surroundings at two levels of resolution. The low-resolution global scene reconstruction provides a shared map for spatial devices such as robots and hand-held AR devices to synchronize and interact based on a shared coordinate system. AR devices and robots can be localized with reference to the same map. Spatial data can be anchored to this coordinate system and persist in the location over time. In this case study, the annotated area of interest by inspectors through an AR device can be shared with robots consistently, robots can localize those annotations in the map and navigate through the areas with planned trajectories and poses. Meanwhile, the result of the surface inspection can be registered to the shared map. Inspectors with AR devices can preview the result aligned with the area of interest when pointing to or approaching the location (Figure 6.4). The high-resolution local scene reconstruction was used for surface anomaly detection and analysis. The detailed surface reconstruction provides raw data for identifying defects such as geometric deformation, cracks, and discolorization, elaborated in Section 6.3.2 Robotically-assisted Inspection Result.

Additionally, the localization module based on vSLAM and UWB provides reliable location tracking for AR devices and robots. Especially under complexed lighting conditions when vSLAM fails due to inconsistent lighting and repetitive visual features, UWB and IMU provide accurate localization as a supplement. The robust localization module plays an important role in AR interaction. As introduced in Section 6.2.1 Annotation placement, the inspection area defined by an inspector is based on the ray casting algorithm. The accuracy of the annotation highly depends on the localization module. Based on feedback from participants in the user test, the integration of UWB and IMU in localization significantly reduced drifts and errors and led to a more consistent user experience in spatial annotation.

Finally, the case study validates the functionality of the semantic segmentation module in surface selection and inspection area annotation. The output of the site reconstruction module is a set of unordered point clouds, and the semantic segmentation module classifies point clouds into groups of surfaces with semantic labels. The compact representation provides the

Chapter 6. Case Study - Spatial Annotation for Robotically-assisted Quality Assessment and Management

inspector with a more efficient way to define a surface for inspection and preview defects associated with a specific physical surface. In Section 6.2.1 Annotation Placement, we can see how inspectors with AR devices choose a surface for inspection by clicking on it and how the spatial annotations can be placed onto and registered with a specific surface. The aforementioned features are based on the identification and classification of point clouds into meaningful and isolated surfaces, or surfels.

Conclusion and Discussion

The dissertation aims to provide a modular framework for AR-enabled robotically-assisted construction. The framework extends general AR frameworks with three optimized modules tailored for construction practices. The site reconstruction module serves for planning near-optimal paths for robots to explore and digitize jobsites which are mostly unknown a priori. The localization module is necessary for all AR devices and robots to position themselves with a reference to a shared map. The semantic segmentation module classifies point clouds into surfaces with semantic labels. In the following case studies, I developed two bespoke interfaces based on the modular framework and validated the framework in robotically-assisted tasks such as pick & place and surface inspection. In the case study spatial annotation for robotically-assisted quality assessment and management, I developed an AR interface for inspectors to define areas of interest for inspection, supervise robot actions, and preview inspection results. The case study exemplifies (1) how the site reconstruction module in the framework provides a shared digital representation of the work environment to register spatial data such as the robotic trajectory, and (2) how the semantic segmentation module can facilitate the inspector to define target surfaces or a local area of interest on a building component for robotic inspection. According to the system usability test, the AR interface developed based on the framework provides an enhanced user experience in robotic manipulation and supervision. The use case proxemic-aware AR for HRI investigates how the spatial relationship can be used to mitigate information overloading in AR for HRI systems operated under cluttered environments such as construction sites. In this case study, I demonstrated how the localiza-

Chapter 7. Conclusion and Discussion

tion module can provide reliable location information for estimating human-robot proxemics and then adjust the information presented in AR views. According to the user study, the AR system adaptive to human-robot proxemics contributes to operating efficiency and accuracy in tasks such as robotically-assisted material handling and site scanning. Beyond the discussion on technical implementation, both case studies demonstrate that the physical environment can play an important role in AR for HRI for information grounding, robot manipulation, and view management. In the rest of this chapter, I will further reflect on the objective, methods, and findings of the dissertation and then outline the limitations and discuss potential future directions for extending the study.

7.1 Conclusion

This thesis starts with the question of how spatial interface may assist human-robot interaction in robotically-assisted construction. Chapter 1, Introduction, I briefly discussed the background of the emerging intersection between spatial interface, specifically augmented reality (AR), and human-robot interaction (HRI), outlining its application in domains such as robotic surgery and manufacturing. Then, I described the potential benefits and challenges of AR for HRI in robotically-assisted construction. In Chapter 2, literature review, I traced early prototypes of spatial interface for robot teleoperation, which can be seen as the intellectual origin of AR for HRI. With a focus on recent developments, I described the state-of-the-art of how AR technology can contribute to effective human-robot interaction in adjacent domains. Then, I outline the challenges of introducing AR in robotically-assisted construction from the perspective of the environment and the practitioner. For example, the complex environments on job sites may negatively affect the crucial components of AR, such as mapping and localization. The introduction of a new interface may challenge the already established workflow in practice. Informed by the challenges, I hypothesize that physical space can be incorporated into AR for HRI to enhance user experience and interface usability. Following this hypothesis, I formalize three research questions of this research:

1. How might we use physical space to contextualize virtual information?
2. How might we leverage physical affordances for intuitive robot programming?

Chapter 7. Conclusion and Discussion

3. How might we use spatial relationships in physical space to mediate robot operators' interactions with virtual information?

To address these questions, I developed a novel computational framework for integrating AR in robotically-assisted construction practices. The framework contributes to the state-of-the-art AR for HRI by providing a more accurate and robust performance in digitalization, localization, and segmentation, which ultimately contributes to an enhanced user experience in AR-based robot manipulation. The key components of the framework are optimized for construction practices. The site reconstruction module provides a digital representation of the surroundings. Based on the context that the requirements of the level of details of a map can vary according to the task, I developed a global and local motion planner for building a multi-resolution site map, which can reduce memory consumption and the amount of data transmitted in networked applications. For example, in the case study robotically-assisted surface inspection, the low-resolution site map with a sparse point cloud is used for tasks such as navigation, and the high-resolution map with a dense point cloud is used for tasks that require high details such as surface anomaly analysis. The multi-resolution site map provides an efficient data format to represent the surroundings. Especially for construction practices, we can first use low-resolution scans to cover broad areas quickly, and then for critical areas, a detailed local scanning can provide precise measurement and sufficient visual information. In the development of the localization module, with the consideration of complex lighting environments and site conditions, I integrated UWB and IMU into a vSLAM system. Under the challenges of repetitive patterns, inconsistent illumination, and poor texture surfaces, which are commonly seen on jobsites, the localization module presents enhanced robustness and accuracy according to the experiment results in Section 4.2.3. The semantic segmentation module groups raw point clouds into surfaces with semantic labels. Partitioned surfaces are important for object selection and information registration in AR systems. Many techniques based on learning have been developed to automatically process 3D point clouds. To enhance generalizability, most learning models focus on the end-to-end pipeline that takes raw point clouds as input and creates class labels in semantic format for point clouds as output. However, in construction applications, the environment can be cluttered, which can result in noises and occlusions in the site scanning, the focus on individual points as input may lead to over-segmentation and inaccuracy. Considering the factors that (1) Point clouds sharing the

Chapter 7. Conclusion and Discussion

same planar surface patches in indoor environments likely belong to the same object, (2) Planar surface patches can be used to filter out the noise and rectify over-segment point clouds, (3) The targets of surface inspection in architectural practices are usually large planar surface patches. I used surfels as the representation of the scene understanding model. In a comparison study with the state-of-the-art solution in Section 4.3.4, the surfel representation achieves better performance in terms of accuracy and intersection over union (IoU).

Based on the framework described in Chapter 4, I investigated the hypothesis through two case studies, comprehensively documented in Chapter 5 and Chapter 6. Chapter 5 Proxemic-aware Augmented Reality for Human-robot Interaction, investigated the research question of how spatial relationships may mediate AR-based human-robot interaction. The study starts from a description of the information overloading issue found in AR for HRI. Excessing virtual information presented in the AR view may cause extra cognitive load and distraction to the robot operator, especially when on job sites where the environment can be cluttered. I proposed a novel proxemic-aware AR system to mitigate the information overload issue. Based on the human-robot spatial relationship, the AR view can adaptively adjust what and how much visual content needs to be presented. Proxemic is a concept describing how people use space around, which is widely adopted in spatial interaction. Different from the conventional definition of proxemics, in this thesis, I introduced two changes: (1) I used empirical data collected from students at the Digital Fabrication Laboratory at Carnegie Mellon University to build the human-robot proxemic (HRP) model, (2) Instead of adopting the crisp and discrete proxemic zoning, I chose to use a continuous function to define HRP. In accordance with HRP measurement, I conducted a co-design workshop with robot operators and designers to create a set of widgets and user interfaces with levels of detail (LoD). When the robot operator moves closer and further away from the robot, the AR view will adjust the LoD accordingly. Finally, I used the proxemic-aware AR for HRI system as an instructional tool and conducted two experiments with students. The students use AR interface to preview the virtual information in physical space and supervise the robot's motion in tasks such as 3D scanning and pick up & place. The result suggests that the AR for HRI system with adaptive view management can alleviate the cognitive load of the robot operator, increase visual clarity, and enhance interaction efficiency. This study made it clear that embodied spatial awareness can be incorporated into the AR interface to address information overloading issues in human-robot interaction.

Chapter 7. Conclusion and Discussion

Our natural understanding of the proximity and surrounding environment can impact our interaction with digital and virtual information. Learning how we interact with physical objects can inform the design of an effective interface that adapts to our preexisting skills. Chapter 6, Spatial Annotation for Robotically-assisted Quality Assessment and Management, introduces a task-specific AR interface for building inspectors to annotate areas on architecture components for robots to inspect and quantify defects. The study investigates how physical affordances can facilitate intuitive robot programming and how AR can contribute to in-situ visualization. In this study, I documented the implementation of an AR feature, spatial annotation, and its technical infrastructure extended from the AR for HRI framework described in Chapter 4. Inspired by how inspectors mark the area of inspection on site, I leverage this well-entrenched skill of inspector and the established workflow in building inspection practices for creating a similar experience in AR-mediated environments. I built a mock-up drywall and a set of digital inspection tools mounted onto a seven-axis industrial robot as an experiment site for validating the AR interface for robotically-assisted Q&M tasks. The user study shows that the AR interface with the spatial annotation feature meets the expectations in terms of usability. In the post-experiment survey, participants mentioned factors that were not revealed in the quantitative results. For example, the ergonomics of handheld devices and the interaction gestures can affect the accuracy of spatial annotation. The introduction of new technologies can cause concerns about job replacement, work surveillance, and privacy issues. Even though participants hold a positive perspective towards the AR-based human-robot interface and believe that the spatial annotation feature shows alignment with their working habits, they still remain neutral towards the novel robotically-assisted workflow, factors such as cost efficiency, safety, and accountability may require more evaluations before integrating robots into their existing practices. The case studies shed light on the importance of physical space in computationally mediated practices, especially its inseparable role in AR for HRI. Following this thread, in the next section, I tried to revisit the definition of spatial interface with a goal of setting a conceptual and theoretical ground for the future research in the domain of computation in built environments.

7.2 Discussion

In this study, I attempt to broaden the definition of spatial interface¹ by shifting the focus from the ontology of an interface to a constitutive impact of an interface, and tentatively defines spatial interfaces as an augmentation of human-world relations (Figure 7.1). Based on this idea, AR can be viewed as a technological mediation that re-mediate the already enmeshed relationship between humans and their inhabited space, rather than a sole visual interface that displays virtual content in the physical world².

The augmentation of relations induced by spatial interface includes embodiment and hermeneutics (Verbeek 2008). The augmentation of embodiment refers to the situation in which a user's experience is reshaped through the interface, with the interface itself in some ways taken into the user's bodily awareness. The interface serves as a human augmentation that mediates the bodily-perceptual relationship between the user and the world. In a hermeneutic relation, the interface is conceived as a part of reality augmentation. Reality is transformed by the interface to a hybrid virtual-physical continuum (Figure 2.1)³ that can be directly experienced by the user's senses. For instance, a spatial interface can convert invisible information of its associated physical objects into a readable format that can be "seen" by the user. By redirecting user's attention from technology itself to a relation of augmentation, a spatial interface provides new insights that address how interfaces can re-mediate one's engagement with the physical world instead of creating a dichotomy between the digital and the physical.

¹ Spatial interface refers to interfaces operating in a three-dimensional (3D) space, as a part of spatial interface, augmented reality (AR) focuses on overlaying digital content onto the physical environment to enhance user perception and experience with their surroundings. Unlike other variants of spatial interface, such as virtual reality, AR depends on the physical environment and supplements it with digital elements, instead of isolating the user from the real world.

² Craig 2013 characterizes AR as medium rather than a type of technology itself, similar definitions of AR can be seen in (Azuma 1997a ; Grubert, Langlotz, et al. 2016). The latter discuss conditions for pervasive Augmented Reality that would offer continuous experience of reality-registered presentations and be context-aware.

³ In (Milgram and Kishino 1994b) , Milgram and Kishino introduce the concept of "the reality-virtuality (RV) continuum". One end of the continuum scale is purely virtual environment, "consisting solely of virtual objects", the other end of the scale is purely real environment, "consisting solely of real objects". Any environment in between is mixed reality (MR), consisting of a blending of real and virtual objects. "Mixed reality environments where the real world is augmented with virtual content are called augmented reality (AR), while those where most of the content is virtual but there is some awareness or inclusion of real world objects are called augmented virtuality (AV)" (Skarbez, Smith, and Whitton 2021).

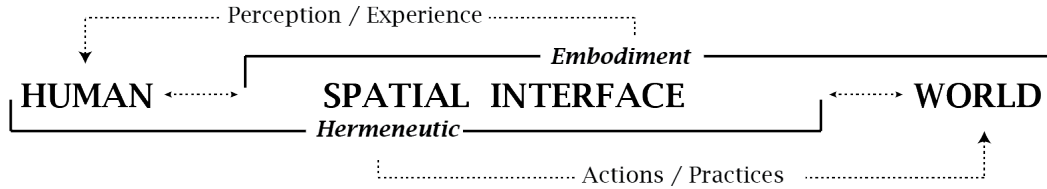


Figure 7.1: Spatial interface as a “technological mediation” augments human-world relations such as embodiment and hermeneutic relations. The diagram is inspired by Verbeek’s descriptions on the mediating effect of technology (Rosenberger and Verbeek 2015; Hauser et al. 2018). Image credit: Author.

The broadened definition of spatial interfaces may bring new potentials to AR systems by emphasizing the remediation of two aspects of human-world relations:

- Embodiment relations — human augmentation: Most AR systems rely on vision as a single channel for information exchange, or so-called unimodal interaction⁴. However, humans as “embodied beings” (Johnson 2015) understand the world through an immersive experience resulting from visual, auditory, and tactile senses. Our sensory, perceptual and cognitive abilities are inextricably tied to our physical being, situated in a dynamic world. Moving beyond a visual display technology, AR systems can incorporate multiple modalities for both input and output, allowing users to use sensorimotor knowledge, well-entrenched spatial skills and native language to interact with digital content naturally and intuitively in the physical world.
- Hermeneutic relations — environment augmentation: The disconnection between computer-generated static information and contingent reality in AR systems may lead to a gap of situatedness⁵. AR systems can reconstruct hermeneutic relations between spatial inter-

⁴The most prevalent input modalities for interactions in existing mixed reality systems are touchscreen and inertial measurement units (Z. Chen et al. 2017). As an alternative, previous pioneering works such as Put-That-There system (Bolt 1980) and (Kaiser et al. 2003) point out the potential of using spoken language as a supplemental I/O for vision to achieve natural 3D interaction.

⁵Augmented reality systems can embed virtual 3D objects convincingly, putting them into the context of the reality, but digital objects are mostly not connected to reality (Azuma 2016). The lack of meaningful virtual-

Chapter 7. Conclusion and Discussion

faces and reality by integrating contextual information, both physical and social. Instead of decontextualizing from or replacing reality, AR can be characterized as an interface for supplementing and augmenting physical environments.

In terms of AR for HRI⁶, the augmentation of embodiment relationship refers to the process of augmenting an operator's ability in robot supervision and control through on-body sensors and actuators. For instance, an operator with a head-mounted display can observe the preoperative robot trajectory overlayed onto the user's view. A surgeon using AR with a haptic feedback interface can teleoperate robot assistance module to perform minimally invasive surgery with improved hand-eye coordination. The feedback from the sensory devices on robots can be transferred to the surgeon via a haptic interface, and the spatial display can assist the surgeon to precisely and safely recognize important structures. The augmentation of hermeneutic relations refers to the visualization of the external information of the environment, which may improve the situation awareness and comprehension of the task of the robot operator in robot control and navigation. For instance, AR systems can superimpose registered information such as sensor stream data of the environment, cost map of search task, and the label of objects to enhance the operator's awareness of the surroundings.

From a technical perspective of view, to achieve the augmentation of embodiment and hermeneutic relations, several challenges still remains to be addressed:

- Data interoperability: multimodal input/output (I/O) can pose interoperability challenges in information exchange process since data encoding format for different interaction channels are varied. An architecture for providing interoperability among modality-specific components coming from heterogeneous I/O channels needs to be integrated to AR systems, for instance, the architecture needs to support data exchange between speech-based module using voice as input and vision-based module using images as input.

physical connections can result in decontextualization. For example, Pokemon Go For the majority of players, urban space has become a backdrop in the game, People are disconnected from those in their in their immediate environment (Munters 2017 ; Tsukayama and Guarino 2016).

⁶Human-robot interaction (HRI) in this study is defined as a process of human supervisory control robots in performance of routine tasks. Robots here is acutated mechanical systems that can take many forms, from industrial robot, mobile robots to robotic interface such as tangible interfaces, swarm robots and kinetic structures. In this study, I narrow down the research scope of robots to industrial robots for architectural practices.

Chapter 7. Conclusion and Discussion

- Virtual-physical registration : In order to present virtual objects over the real world accurately, an AR system needs to obtain (1) the current view of the real world to be augmented (2) the registration between the virtual object and the real world. The former problem is usually referred to as the tracking problem (Welch and Foxlin 2002), the latter one is often referred to as authoring problem (Shin et al. 2009). A meaningful authoring can enable AR systems to display necessary virtual information at a right location, thus avoiding issues including information overloading, de-contextualization and distraction. In order to create a meaningful authoring, an AR system need to require a good knowledge of both the virtual and real parts of a scene, which still remains an open question⁷. Recent developments in scene understanding enabled by deep learning techniques (Armeni, Sax, et al. 2017bKirillov et al. 2019) and semantic Simultaneous localization and mapping (SLAM) (Kostavelis and Gasteratos 2015) make it possible to automatically 3D reconstruct the physical environment and interpret 3D dynamic scene. This study will explore the potential of scene understanding and integrate semantic SLAM systems into AR for enhancing the virtual-physical registration.

The extended definition of spatial interface shares similarity with the term spatial computing defined as "human interaction with a machine in which the machine retains and manipulates referents to real objects and spaces" (Greenwold 2003). Both definitions depart from previous concepts on virtual placemaking and focus on the "addressability " (Dhaliwal 2022) between the physical world and computational systems. In this work, I demonstrated the implementation of the concept of virtual-physical registration, or hermeneutics, through the semantic understanding pipeline elaborated in Section 4.3. The proxemic-aware interaction introduced in Chapter 4 illustrated how the AR system can be adaptive to the embodiment of human-world relationship.

⁷Two of the most prevalent methods for virtual-physical registration are marker-based and markless. Marker-based AR uses markers (eg. 2D matrix code (Kato and Billinghurst 1999) , specific letters or characters (Wojciechowski et al. 2004) , pervasive point lights (Ahuja et al. 2019)) in a scene to anchor digital 3D content within the user's visual field. Markless AR is based on the natural characteristics of the scene captured by the camera for virtual content placement (Oufqir, Abderrahmani, and Satori 2020) . The prior knowledge of the environment is not required, a realtime simultaneous localization and mapping technology can scan the physical environment and create maps for virtual content placement. Markerless tracking has advantages such as not requiring prior knowledge of a user's locale, and not needing to place additional objects (e.g., fiducial markers) in an environment to create MAR experiences (J. Cao et al. 2021) .

7.3 Limitations and Future Work

This thesis provides a computational framework for integrating augmented reality as a human-robot interface for robotically-assisted construction and uses two case studies to demonstrate its potential applications in practice. While the framework shows great potential and generalizability, many technical limitations and practical challenges remain. From the perspective of technical constraints, the framework needs to further address three issues: precision, reliability, and responsiveness. In Chapter 4, I described how lighting conditions may affect localization in indoor environments and how the framework is enhanced with a multimodal localization method. Still, accurate indoor localization can be a challenging task in visually challenging and dynamic environments such as construction sites. The improvement of localization can significantly broaden the applicability of the framework in practice. Following the direction of the technical implementation in this thesis, I believe the combination of data-driven approaches and sensor fusion technology may bring some opportunities for robust localization solutions in AR for HRI applications, from indoor environments to the outdoor wild world. Secondly, the reliability of the technical infrastructure still needs to be improved. Due to the instability of signals caused by the environment, for example, occlusion and multipath issues, the juxtaposition of virtual content can experience temporary drift or failures. For safety-critical HRI applications in construction, we need to take into account the situation when the AR interface fails to provide accurate visual guidance. As the complexity of the system increases, we need to add more fail-proof mechanisms to ensure the safety of the operator when either robotic systems or AR systems malfunction. A more synergistic approach can be designed to improve the reliability from a user-interaction perspective.

Extended from the existing research, in the next step, several research opportunities in AR and robotics within the domain of ARC can be further explored. Firstly, extending from a human-machine interface, AR technology can serve as a shared and collaborative interface for multiple stakeholders to collaborate on job sites. AR interface can present a shared visualization as a support for communication and ensure that all involved participants have a clear view of construction plans and objectives. By overlaying real-time task assignment and status data onto the physical environment, AR interfaces can enable supervisors to monitor and manage a multi-robot team or multiple equipment on site by assigning tasks dynamically and optimizing

Chapter 7. Conclusion and Discussion

resource allocation for efficiency. In addition, the development of large language models may bring new opportunities in the novel natural input for AR for HRI on job site. In this thesis, I described a data-driven approach for semantic segmentation of point clouds, a process that can automatically map low-level point cloud data to a high-level semantic representation. The semantic information can be further combined with building information modeling to build a connection between as-built and as-design data. For robotics, the semantic information of the surroundings, together with natural language processing, can enable the operator to program the actions of robots with voice commands. Moreover, AR for HRI can be applied as a communication tool for teleoperating construction robots for space construction. Operators in mission centers, equipped with immersive visualization, guidance, and communication tools provided by AR interfaces, can operate robots, construction equipment, and mobile systems remotely and meanwhile communicate with astronauts in space stations for collaborative decision-making. Furthermore, from a broad view, reconfigurable actuators in physical environments can be considered robots. For example, shape-changing building components, wearable devices such as exoskeletons, and autonomous vehicles on job sites can be part of robotic systems with varied forms different than conventional industrial robots. AR for HRI system can be integrated for operators to interact with the equipment effectively. Finally, the first case study demonstrates the potential of introducing AR into personal training in construction robotics, enabling workers to develop practical skills through an enhanced learning experience. Interactive guidance enabled by AR interfaces can overlay instructions and tutorials onto the physical workspace, demonstrating the concepts, techniques, and skills of operating complex machines. Based on the feedback from the experiment participants, the system can provide a more personalized upskilling mentorship for workers to leverage digital tools and machines for enhanced safety in construction practices. The computational framework of AR for HRI with a focus on robotically-assisted construction set a first step for a future investigation in leveraging computationally enhanced physical environments for robot programming.

Bibliography

- Aarts, Emile and Reiner Wichert (2009). *Ambient intelligence*.
- Abowd, D et al. (1998). “Context-awareness in wearable and ubiquitous computing”. In: *Virtual Reality* 3.3, pp. 200–211.
- Afsari, Kereshmeh et al. (2018). “Applications of collaborative industrial robots in building construction”. In: *54th ASC Annual International Conference Proceedings*, pp. 472–479.
- Ahmed, Shakil (2018). “A review on using opportunities of augmented reality and virtual reality in construction project management”. In: *Organization, technology & management in construction: an international journal* 10.1, pp. 1839–1852.
- Ahuja, Karan et al. (2019). “Lightanchors: Appropriating point lights for spatially-anchored augmented reality interfaces”. In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 189–196.
- Akinbiyi, Takintope et al. (2006). “Dynamic augmented reality for sensory substitution in robot-assisted surgical systems”. In: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 567–570.
- Albu-Schäffer, Alin et al. (2007). “The DLR lightweight robot: design and control concepts for robots in human environments”. In: *Industrial Robot: an international journal* 34.5, pp. 376–385.
- Alexa, Marc et al. (2001). “Point set surfaces”. In: *Proceedings Visualization, 2001. VIS’01*. Pp. 21–29.
- Alliez, Pierre et al. (2007). “Voronoi-based variational reconstruction of unoriented point sets”. In: *Symposium on Geometry processing*. Vol. 7, pp. 39–48.

Bibliography

- Antle, Alissa N (2013). “Exploring how children use their hands to think: An embodied interactional analysis”. In: *Behaviour & Information Technology* 32.9, pp. 938–954.
- Armeni, Iro, Sasha Sax, et al. (2017a). “Joint 2d-3d-semantic data for indoor scene understanding”. In: *arXiv preprint arXiv:1702.01105*.
- (2017b). “Joint 2d-3d-semantic data for indoor scene understanding”. In: *arXiv preprint arXiv:1702.01105*.
- Armeni, Iro, Ozan Sener, et al. (2016). “3d semantic parsing of large-scale indoor spaces”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1534–1543.
- Ashby, W Ross and Jeffrey Goldstein (2011). “Variety, constraint, and the law of requisite variety”. In: *Emergence: Complexity and Organization* 13.1/2, p. 190.
- Assarsson, Ulf and Tomas Moller (2000). “Optimized view frustum culling algorithms for bounding boxes”. In: *Journal of graphics tools* 5.1, pp. 9–22.
- Azuma, Ronald T (1997a). “A survey of augmented reality”. In: *Presence: teleoperators & virtual environments* 6.4, pp. 355–385.
- (1997b). “A survey of augmented reality”. In: *Presence: Teleoperators & Virtual Environments* 6.4, pp. 355–385.
- (2016). “The most important challenge facing augmented reality”. In: *Presence: Teleoperators and Virtual Environments* 25.3, pp. 234–238.
- Ballendat, Till, Nicolai Marquardt, and Saul Greenberg (2010a). “Proxemic interaction: designing for a proximity and orientation-aware environment”. In: *ACM International Conference on Interactive Tabletops and Surfaces*, pp. 121–130.
- (2010b). “Proxemic interaction: designing for a proximity and orientation-aware environment”. In: *ACM International Conference on Interactive Tabletops and Surfaces*, pp. 121–130.
- Bandyopadhyay, Tirthankar et al. (2018). “Magneto: A versatile multi-limbed inspection robot”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2253–2260.
- Bangor, Aaron, Philip Kortum, and James Miller (2009). “Determining what individual SUS scores mean: Adding an adjective rating scale”. In: *Journal of usability studies* 4.3, pp. 114–123.

Bibliography

- Bangor, Aaron, Philip T Kortum, and James T Miller (2008). “An empirical evaluation of the system usability scale”. In: *Intl. Journal of Human-Computer Interaction* 24.6, pp. 574–594.
- Bard, Joshua David et al. (2016). “Reality is interface: Two motion capture case studies of human-machine collaboration in high-skill domains”. In: *International Journal of Architectural Computing* 14.4, pp. 398–408.
- Batmaz, Anil Ufuk and Wolfgang Stuerzlinger (2019). “Effects of 3D rotational jitter and selection methods on 3D pointing tasks”. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1687–1692.
- Baykal, Gökçe Elif et al. (2018). “A review on complementary natures of tangible user interfaces (TUIs) and early spatial learning”. In: *International journal of child-computer interaction* 16, pp. 104–113.
- Bejczy, Antal K, Won S Kim, and Steven C Venema (1990a). “The phantom robot: predictive displays for teleoperation with time delay”. In: *Proceedings., IEEE International Conference on Robotics and Automation*, pp. 546–551.
- (1990b). “The phantom robot: predictive displays for teleoperation with time delay”. In: *Proceedings., IEEE International Conference on Robotics and Automation*, pp. 546–551.
- (1990c). “The phantom robot: predictive displays for teleoperation with time delay”. In: *Proceedings., IEEE International Conference on Robotics and Automation*, pp. 546–551.
- Benford, Steve et al. (2001). “Collaborative virtual environments”. In: *Communications of the ACM* 44.7, pp. 79–85.
- Berkmann, Jens and Terry Caelli (1994). “Computation of surface geometry and segmentation using covariance techniques”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.11, pp. 1114–1116.
- Berri, Jawad, Rachid Benlamri, and Yacine Atif (2006). “Ontology-based framework for context-aware mobile learning”. In: *Proceedings of the 2006 international conference on Wireless communications and mobile computing*, pp. 1307–1310.
- Besl, Paul J and Neil D McKay (1992). “Method for registration of 3-D shapes”. In: *Sensor fusion IV: control paradigms and data structures*. Vol. 1611, pp. 586–606.
- Besl, Paul J. and Ramesh C Jain (1988). “Segmentation through variable-order surface fitting”. In: *IEEE Transactions on pattern analysis and machine intelligence* 10.2, pp. 167–192.

Bibliography

- Betti, Giovanni, Saqib Aziz, and Gili Ron (2019). “Pop Up Factory: Collaborative Design in Mixed Reality-Interactive live installation for the makeCity festival, 2018 Berlin”. In.
- Bich, Leonardo and Arantza Etxeberria (2013). “Systems, autopoietic”. In.
- Billinghurst, Mark and Hirokazu Kato (2002). “Collaborative augmented reality”. In: *Communications of the ACM* 45.7, pp. 64–70.
- Blaer, Paul S and Peter K Allen (2007a). “Data acquisition and view planning for 3-D modeling tasks”. In: *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 417–422.
- (2007b). “Data acquisition and view planning for 3-D modeling tasks”. In: *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 417–422.
- Blajer, Jacek A. Goszczyński Wojciech and Mariusz Krawczyk (2002). “The Inverse Simulation Study of Aircraft Flight Path Reconstruction”. In: *Transport* XVII.3, pp. 103–107.
- Boari, Douglas and Mike Fraser (2009). “Taking shortcuts: embedded physical interfaces for spatial navigation”. In: *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pp. 189–196.
- Bock, Thomas (2007). “Construction robotics”. In: *Autonomous Robots* 22, pp. 201–209.
- Bolt, Richard A (1980). ““Put-that-there” Voice and gesture at the graphics interface”. In: *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, pp. 262–270.
- Borgmann, Albert (2007). *Holding on to Reality: The Nature of Information at the Turn of the Millennium*. University of Chicago Press.
- Bowen, C et al. (2004). “Utilization of the Space Vision System as an Augmented Reality System for Mission Operations”. In: *Proceedings of AIAA Habitation Conference, Houston TX*.
- Brand, Stewart (1995). *How buildings learn: What happens after they’re built*. Penguin.
- Bremermann, H J (1977). “Complexity and transcomputability”. In: *The Encyclopaedia of Ignorance* 1.
- Brooke, John (1996). “Sus: a “quick and dirty” usability”. In: *Usability evaluation in industry* 189.3, pp. 189–194.
- (2013). “SUS: a retrospective”. In: *Journal of usability studies* 8.2, pp. 29–40.

Bibliography

- Brooks, Rodney A (1991a). “Intelligence without reason”. In: *Artificial intelligence: critical concepts* 3, pp. 107–63.
- (1991b). “New approaches to robotics”. In: *Science* 253.5025, pp. 1227–1232.
- Broquere, Xavier, Daniel Sidobre, and Ignacio Herrera-Aguilar (2008). “Soft motion trajectory planner for service manipulator robot”. In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2808–2813.
- Browse, R A and S A Little (1991). “The effectiveness of real-time graphic simulation in telerobotics”. In: *Conference Proceedings 1991 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 895–898.
- Bruner, Jerome Seymour (1966). *Toward a theory of instruction*. Vol. 59. Harvard University Press.
- Butler, John Travis and Arvin Agah (2001). “Psychological effects of behavior patterns of a mobile personal robot”. In: *Autonomous Robots* 10, pp. 185–202.
- Buzan, F T (1987). *Control of a teleoperator in the presence of large time delays*.
- Calanca, Andrea, Riccardo Muradore, and Paolo Fiorini (2015). “A review of algorithms for compliant control of stiff and fixed-compliance robots”. In: *IEEE/ASME transactions on mechatronics* 21.2, pp. 613–624.
- Calixte, Xaviéra and Pierre Leclercq (2017). “The interactive projection mapping as a spatial augmented reality to help collaborative design: Case study in architectural design”. In: *Co-operative Design, Visualization, and Engineering: 14th International Conference, CDVE 2017, Mallorca, Spain, September 17-20, 2017, Proceedings 14*, pp. 143–152.
- Cannon, Walter B (1929). “Organization for physiological homeostasis”. In: *Physiological reviews* 9.3, pp. 399–431.
- Cao, Jacky et al. (2021). “Mobile Augmented Reality: User Interfaces, Frameworks, and Intelligence”. In: *arXiv preprint arXiv:2106.08710*.
- Cao, Yanjun and Giovanni Beltrame (2021). “VIR-SLAM: Visual, inertial, and ranging SLAM for single and multi-robot systems”. In: *Autonomous Robots* 45, pp. 905–917.
- Cardoso, Daniel Llach (2019). “Tracing design ecologies: collecting and visualizing ephemeral data as a method in design and technology studies”. In: *digitalSTS*. Ed. by and others. Princeton University Press, pp. 451–471.

Bibliography

- Cardoso, Daniel Llach, Ardavan Bidgoli, and Shokofeh Darbari (2017). “Assisted automation: Three learning experiences in architectural robotics”. In: *International Journal of Architectural Computing* 15.1, pp. 87–102.
- Cazals, Frédéric and Marc Pouget (2005). “Estimating differential quantities using polynomial fitting of osculating jets”. In: *Computer Aided Geometric Design* 22.2, pp. 121–146.
- Chen, Harry et al. (2004). “Intelligent agents meet the semantic web in smart spaces”. In: *IEEE Internet computing* 8.6, pp. 69–79.
- Chen, Zhaorui et al. (2017). “Multimodal interaction in augmented reality”. In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 206–209.
- Chong, Jonathan Wun Shiung et al. (2009). “Robot programming using augmented reality: An interactive method for planning collision-free paths”. In: *Robotics and Computer-Integrated Manufacturing* 25.3, pp. 689–701.
- Chowriappa, Ashirwad et al. (2015). “Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: a multi-institutional randomised controlled trial”. In: *BJU Int* 115.2, pp. 336–345.
- Churchill, Elizabeth F and Dave Snowden (1998). “Collaborative virtual environments: an introductory review of issues and systems”. In: *virtual reality* 3, pp. 3–15.
- Clancey, William J (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge university press.
- Clark, Andy (1997). *Being there*. MIT Press Cambridge, MA.
- Committee, ACI (2006). “Specifications for tolerances for concrete construction and materials and commentary”. In: American Concrete Institute.
- Cook, Diane and Sajal Kumar Das (2004). *Smart environments: technology, protocols, and applications*. Vol. 43. John Wiley & Sons.
- Cook, Diane J, Michael Youngblood, and Sajal K Das (2006). *A multi-agent approach to controlling a smart environment*.
- Craig, Alan B (2013). *Understanding augmented reality: Concepts and applications*. Newnes.
- Dardari, Davide et al. (2009). “Ranging with ultrawide bandwidth signals in multipath environments”. In: *Proceedings of the IEEE* 97.2, pp. 404–426.

Bibliography

- Das, M et al. (2010). "Context aware e-learning system with dynamically composable learning objects". In: *International Journal on Computer Science and Engineering* 2.4, pp. 1245–1253.
- Daudelin, Jonathan and Mark Campbell (2017a). "An adaptable, probabilistic, next-best view algorithm for reconstruction of unknown 3-d objects". In: *IEEE Robotics and Automation Letters* 2.3, pp. 1540–1547.
- (2017b). "An adaptable, probabilistic, next-best view algorithm for reconstruction of unknown 3-d objects". In: *IEEE Robotics and Automation Letters* 2.3, pp. 1540–1547.
- Daza, Marcos et al. (2021). "An approach of social navigation based on proxemics for crowded environments of humans and robots". In: *Micromachines* 12.2, p. 193.
- Derrida, Jacques (2012). *Specters of Marx: The state of the debt, the work of mourning and the new international*. Routledge.
- Dewey, John (2005). *Art as experience*. Penguin.
- Dey, Anind K, Gregory D Abowd, and Daniel Salber (2001). "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications". In: *Human-Computer Interaction* 16.2-4, pp. 97–166.
- Dhaliwal, Ranjodh Singh (2022). "On Addressability, or What Even Is Computation?" In: *Critical Inquiry* 49.1, pp. 1–27.
- Dickey, Ryan M et al. (2016). "Augmented reality assisted surgery: a urologic training tool". In: *Asian journal of andrology* 18.5, p. 732.
- Dourish, Paul (2001). "Seeking a foundation for context-aware computing". In: *Human-Computer Interaction* 16.2-4, pp. 229–241.
- (2004a). *Where the action is: the foundations of embodied interaction*. MIT press.
- (2004b). *Where the action is: the foundations of embodied interaction*. MIT press.
- Eastman, Charles M (1972). *Adaptive conditional architecture*. Institute of Physical Planning, Carnegie-Mellon University.
- Edinbarough, Immanuel, Roberto Balderas, and Subhash Bose (2005). "A vision and robot based on-line inspection monitoring system for electronic manufacturing". In: *Computers in Industry* 56.8-9, pp. 986–996.

Bibliography

- Edwards, W Keith et al. (2002). “Using speakeasy for ad hoc peer-to-peer collaboration”. In: *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pp. 256–265.
- Ens, Barrett et al. (2019). “Revisiting collaboration through mixed reality: The evolution of groupware”. In: *International Journal of Human-Computer Studies* 131, pp. 81–98.
- Fang, H C, S K Ong, and A Y C Nee (2012a). “Interactive robot trajectory planning and simulation using augmented reality”. In: *Robotics and Computer-Integrated Manufacturing* 28.2, pp. 227–237.
- (2012b). “Interactive robot trajectory planning and simulation using augmented reality”. In: *Robotics and Computer-Integrated Manufacturing* 28.2, pp. 227–237.
- Fareh, Raouf et al. (2014). “An integrated vision-guided robotic system for rapid vehicle inspection”. In: *2014 IEEE International Systems Conference Proceedings*. IEEE, pp. 446–451.
- Fazel, Alireza and Abbasali Izadi (2018a). “An interactive augmented reality tool for constructing free-form modular surfaces”. In: *Automation in Construction* 85, pp. 135–145.
- (2018b). “An interactive augmented reality tool for constructing free-form modular surfaces”. In: *Automation in Construction* 85, pp. 135–145.
- Felbrich, Benjamin et al. (2018). “Self-organizing maps for intuitive gesture-based geometric modelling in augmented reality”. In: *2018 IEEE international conference on artificial intelligence and virtual reality (AIVR)*, pp. 61–67.
- Fernando, Eranga et al. (2019). “Observability analysis of position estimation for quadrotors with modified dynamics and range measurements”. In: *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 2783–2788.
- Foerster, Heinz Von (2003a). *Cybernetics of cybernetics*.
- (2003b). “Objects: tokens for (eigen-) behaviors”. In: *Understanding understanding: Essays on cybernetics and cognition*, pp. 261–271.
- Fong, Terrence and Charles Thorpe (2001). “Vehicle teleoperation interfaces”. In: *Autonomous robots* 11.1, pp. 9–18.
- Foth, Marcus et al. (2011). *From social butterfly to engaged citizen: Urban informatics, social media, ubiquitous computing, and mobile technology to support citizen engagement*. MIT Press.

Bibliography

- Foucault, Michel (2012). *Discipline and punish: The birth of the prison*. Vintage.
- Gaião, C, J de de Brito, and J Silvestre (2011). “Inspection and diagnosis of gypsum plaster-board walls”. In: *Journal of Performance of Constructed Facilities* 25.3, pp. 172–180.
- Gannon, Madeline (2014). *QUIPT*. URL: <http://www.madlab.cc/quipt>.
- Gao, Yuan et al. (2019). “Fast adaptation with meta-reinforcement learning for trust modelling in human-robot interaction”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 305–312.
- García, Alberto et al. (2022). “Augmented Reality-Based Interface for Bimanual Robot Teleoperation”. In: *Applied Sciences* 12.9, p. 4379.
- García-Pereira, Inma et al. (2020). “A collaborative augmented reality annotation tool for the inspection of prefabricated buildings”. In: *Multimedia Tools and Applications* 79.9, pp. 6483–6501.
- Gere, Charlie (2009). *Digital culture*. Reaktion Books.
- Giedion, Sigfried (1948). “Mechanization takes command a contribution to anonymous history”. In.
- Glynn, Ruairi (2007). “Performative Ecologies”. In: *Ruairi Glynn*.
- (2008). “Conversational environments revisited”. In.
- Gomez, Erick Sebastian (2019). “Placemaking & Augmented Reality: A Theoretical Exploration”. In.
- Gottfried, Interpretation-BMI B and H Aghajan (2009). “Learning about preferences and common behaviours of the user in an intelligent environment”. In: *Behaviour Monitoring and Interpretation-BMI: Smart Environments* 3, p. 289.
- Greenwold, Simon (2003). “Spatial computing”. In: *Massachusetts Institute of Technology, Master*.
- Grinter, Rebecca E et al. (2005). “The work to make a home network work”. In: *ECSCW 2005*, pp. 469–488.
- Grønbæk, Jens Emil et al. (2020). “Proxemics beyond proximity: Designing for flexible social interaction through cross-device interaction”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Bibliography

- Grubert, Jens, Tobias Langlotz, et al. (2016). “Towards pervasive augmented reality: Context-awareness in augmented reality”. In: *IEEE transactions on visualization and computer graphics* 23.6, pp. 1706–1724.
- Grubert, Jens and Dieter Schmalstieg (2013). “Playing it real again: a repeated evaluation of magic lens and static peephole interfaces in public space”. In: *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pp. 99–102.
- Gubbi, Jayavardhana et al. (2013). “Internet of Things (IoT): A vision, architectural elements, and future directions”. In: *Future generation computer systems* 29.7, pp. 1645–1660.
- Guennebaud, Gaël and Markus Gross (2007). *Algebraic point set surfaces*.
- Guo, Anhong et al. (2019). “Blocks: Collaborative and persistent augmented reality experiences”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3.3, pp. 1–24.
- Haan, Gerwin De, Michal Koutek, and Frits H Post (2005). “IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection.” In: *Ipt/egve*, pp. 201–209.
- Hahm, Soomeen et al. (2019). “FlowMorph-exploring the human-material interaction in digitally augmented craftsmanship”. In: *Intelligent & informed-proceedings of the 24th CAADRIA conference*. Vol. 1, pp. 553–62.
- Halder, Srijeet and Kereshmeh Afsari (2023). “Robots in inspection and monitoring of buildings and infrastructure: A systematic review”. In: *Applied Sciences* 13.4, p. 2304.
- Hall, Edmund T and Edward T Hall (1966). “The hidden dimension”. In: vol. 609. Anchor, p. 1.
- Haque, Usman (2007a). “The architectural relevance of Gordon Pask”. In: *Architectural Design* 77.4, pp. 54–61.
- (2007b). “The architectural relevance of Gordon Pask”. In: *Architectural Design* 77.4, pp. 54–61.
- Harper, Richard (2003). *Inside the smart home: Ideas, possibilities and methods*.
- Harrison, Steve and Paul Dourish (1996a). “Re-place-ing space: the roles of place and space in collaborative systems”. In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pp. 67–76.

Bibliography

- (1996b). “Re-place-ing space: the roles of place and space in collaborative systems”. In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pp. 67–76.
- Hauser, Sabrina et al. (2018). “An annotated portfolio on doing postphenomenology through research products”. In: *Proceedings of the 2018 designing interactive systems conference*, pp. 459–471.
- Hayles, N Katherine (1994). “Boundary disputes: Homeostasis, reflexivity, and the foundations of cybernetics”. In: *Configurations* 2.3, pp. 441–467.
- Hedayati, Hooman, Michael Walker, and Daniel Szafr (2018). “Improving collocated robot teleoperation with augmented reality”. In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 78–86.
- Heyer, Clint (2010). “Human-robot interaction and future industrial robotics applications”. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4749–4754.
- Hietanen, Antti et al. (2020). “AR-based interaction for human-robot collaborative manufacturing”. In: *Robotics and Computer-Integrated Manufacturing* 63, p. 101891.
- Hoang, Khoa C et al. (2022). “Virtual barriers in augmented reality for safe and effective human-robot cooperation in manufacturing”. In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1174–1180.
- Hong, Jong-yi, Eui-ho Suh, and Sung-Jin Kim (2009). “Context-aware systems: A literature review and classification”. In: *Expert Systems with applications* 36.4, pp. 8509–8522.
- Huang, Xiaoshui, Guofeng Mei, and Jian Zhang (2023). “Cross-source point cloud registration: Challenges, progress and prospects”. In: *Neurocomputing*, p. 126383.
- Huber, Daniel and Takeo Kanade (2011). “Integrating lidar into stereo for fast and improved disparity computation”. In: *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pp. 405–412.
- Hüttenrauch, Helge et al. (2006a). “Investigating spatial relationships in human-robot interaction”. In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5052–5059.
- (2006b). “Investigating spatial relationships in human-robot interaction”. In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5052–5059.

Bibliography

- Hutto, Daniel D and Erik Myin (2012). *Radicalizing enactivism: Basic minds without content*. MIT press.
- I Aviles-Rivero, Angelica et al. (2018). “Sensory substitution for force feedback recovery: a perception experimental study”. In: *ACM Transactions on Applied Perception (TAP)* 15.3, pp. 1–19.
- Ibrahimov, Roman et al. (2019). “Dronepick: Object picking and delivery teleoperation with the drone controlled by a wearable tactile display”. In: *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1–6.
- Ince, Ilhan, K Bryant, and T Brooks (1991). “Virtuality and reality: a video/graphics environment for teleoperation”. In: *Conference Proceedings 1991 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1083–1089.
- Intille, Stephen S et al. (2005). “The PlaceLab: A live-in laboratory for pervasive computing research (video)”. In: *Proceedings of PERVASIVE 2005 Video Program*.
- Inzartsev, Alexander et al. (2019). “Underwater pipeline inspection method for AUV based on laser line recognition: Simulation results”. In: *2019 IEEE Underwater Technology (UT)*. IEEE, pp. 1–8.
- Isler, Stefan et al. (2016a). “An information gain formulation for active volumetric 3D reconstruction”. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3477–3484.
- (2016b). “An information gain formulation for active volumetric 3D reconstruction”. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3477–3484.
- Ismail, Ajune Wanis et al. (2019). “Designing an Augmented Reality Multimodal Interface for 6DOF Manipulation Techniques: Multimodal Fusion Using Gesture and Speech Input for AR”. In: *Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 1*, pp. 309–322.
- Ito, Masaru et al. (2021). “Effects of machine instability feedback on safety during digging operation in teleoperated excavators”. In: *IEEE Access* 9, pp. 28987–28998.
- Ito, Mizuko (2005). *Mobile phones, Japanese youth, and the re-placement of social contact*.
- Izadi, Shahram et al. (2011). “KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera”. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pp. 559–568.

Bibliography

- Jaegher, Hanne De and Marieke Rohde (2010). *Enaction: Toward a new paradigm for cognitive science*. MIT Press.
- Jahn, Gwyllim et al. (2020). “Holographic construction”. In: *Impact: Design With All Senses: Proceedings of the Design Modelling Symposium, Berlin 2019*, pp. 314–324.
- JIN, Jiayi (2016). “New Poetics of Augmented Space: Re-conceptualise Augmented Space Based on Fundamentals of Augmented Reality”. In: *Lecture at the Conference of Museums and the Web (MW2016)*.
- Johns, Ryan Luke (2013a). *Augmented reality and the fabrication of gestural form*.
- (2013b). *Augmented reality and the fabrication of gestural form*.
- (2014). “Augmented materiality: modelling with material indeterminacy”. In: *Fabricate. gta Verlag, Zurich*, pp. 216–223.
- Johns, Ryan Luke, Jeffrey Anderson, and Axel Kilian (2019). “Robo-Stim: modes of human robot collaboration for design exploration”. In: *Design Modelling Symposium Berlin*, pp. 671–684.
- (2020). “Robo-Stim: Modes of human robot collaboration for design exploration”. In: *Impact: Design With All Senses: Proceedings of the Design Modelling Symposium, Berlin 2019*. Springer, pp. 671–684.
- Johnson, Mark (2015). “Embodied understanding”. In: *Frontiers in psychology* 6, p. 875.
- Kaiser, Ed et al. (2003). “Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality”. In: *Proceedings of the 5th international conference on Multimodal interfaces*, pp. 12–19.
- Kato, Hirokazu and Mark Billinghurst (1999). “Marker tracking and hmd calibration for a video-based augmented reality conferencing system”. In: *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR’99)*, pp. 85–94.
- Keil, Jens et al. (2013). “Controlling and filtering information density with spatial interaction techniques via handheld augmented reality”. In: *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments: 5th International Conference, VAMR 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I 5*, pp. 49–57.

Bibliography

- Khaloo, Ali and David Lattanzi (2017a). “Robust normal estimation and region growing segmentation of infrastructure 3D point cloud models”. In: *Advanced Engineering Informatics* 34, pp. 1–16.
- (2017b). “Robust normal estimation and region growing segmentation of infrastructure 3D point cloud models”. In: *Advanced Engineering Informatics* 34, pp. 1–16.
- Khan, F et al. (2015). “Investigation on bridge assessment using unmanned aerial systems”. In: *Structures Congress 2015*, pp. 404–413.
- Kientz, Julie A et al. (2008). *The georgia tech aware home*.
- Kilian, Axel (2017). *Embodied computation The changing relationship of physical form and geometry*.
- Kirillov, Alexander et al. (2019). “Panoptic segmentation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413.
- Kiyokawa, Kiyoshi et al. (2002). “Communication behaviors of co-located users in collaborative AR interfaces”. In: *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 139–148.
- Klinker, Gudrun, Didier Stricker, and Dirk Reiners (2001). “Augmented reality for exterior construction applications”. In: *Fundamentals of wearable computers and augmented reality*, pp. 379–427.
- Klug, Brandy (2017). “An overview of the system usability scale in library website and system usability testing”. In: *Weave: Journal of Library User Experience* 1.6.
- Koay, Kheng Lee et al. (2014). “Social roles and baseline proxemic preferences for a domestic service robot”. In: *International Journal of Social Robotics* 6, pp. 469–488.
- Kohler, Matthias, Fabio Gramazio, and Jan Willmann (2014). *The robotic touch: how robots change architecture*.
- Kostavelis, Ioannis and Antonios Gasteratos (2015). “Semantic mapping for mobile robotics tasks: A survey”. In: *Robotics and Autonomous Systems* 66, pp. 86–103.
- Kyjanek, O et al. (2019). “Implementation of an augmented reality ar workflow for human robot collaboration in timber prefabrication”. In: *Proceedings of the 36th International Symposium on Automation and Robotics in Construction*. Vol. 36.

Bibliography

- Kyjanek, Ondrej et al. (2019). “Implementation of an augmented reality AR workflow for human robot collaboration in timber prefabrication”. In: *Proceedings of the 36th international symposium on automation and robotics in construction, ISARC*, pp. 1223–1230.
- La, Hung Manh et al. (2019). “Automated robotic monitoring and inspection of steel structures and bridges”. In: *Robotica* 37.5, pp. 947–967.
- Laffranchi, Matteo, Nikolaos G Tsagarakis, and Darwin G Caldwell (2009). “Safe human robot interaction via energy regulation control”. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 35–41.
- Lajoie, Susanne P (2003). “Individual differences in spatial ability: Developing technologies to increase strategy awareness and skills”. In: *Educational Psychologist* 38.2, pp. 115–125.
- Langlotz, Tobias et al. (2014). “Next-generation augmented reality browsers: rich, seamless, and adaptive”. In: *Proceedings of the IEEE* 102.2, pp. 155–169.
- Law, John (2006). “Networks, relations, cyborgs: on the social study of technology”. In.
- Ledo, David et al. (2015). “Proxemic-aware controls: Designing remote controls for ubiquitous computing ecologies”. In: *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 187–198.
- Lee, Jin Sol et al. (2022). “Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation”. In: *Automation in Construction* 135, p. 104119.
- Lee, Jin-Shyan, Yu-Wei Su, and Chung-Chou Shen (2007). “A comparative study of wireless protocols: Bluetooth, UWB, ZigBee, and Wi-Fi”. In: *IECON 2007-33rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 46–51.
- Lewis, James R (2018). “The system usability scale: past, present, and future”. In: *International Journal of Human-Computer Interaction* 34.7, pp. 577–590.
- Lin, Huei-Yung and Ming-Chi Yeh (2022). “Drift-Free Visual SLAM for Mobile Robot Localization by Integrating UWB Technology”. In: *IEEE Access* 10, pp. 93636–93645.
- Lin, Zecai et al. (2022). “Robotic Telepresence Based on Augmented Reality and Human Motion Mapping for Interventional Medicine”. In: *IEEE Transactions on Medical Robotics and Bionics* 4.4, pp. 935–944.

Bibliography

- Lindlbauer, David, Anna Maria Feit, and Otmar Hilliges (2019). “Context-aware online adaptation of mixed reality interfaces”. In: *Proceedings of the 32nd annual ACM symposium on user interface software and technology*, pp. 147–160.
- Lipowski, Zbigniew J (1975). “Sensory and information inputs overload: behavioral effects.” In: *Comprehensive Psychiatry*.
- Liu, Donghai et al. (2021). “Integrating building information model and augmented reality for drone-based building inspection”. In: *Journal of Computing in Civil Engineering* 35.2, p. 04020073.
- Liu, Fei et al. (2020a). “An UWB/vision fusion scheme for determining pedestrians’ indoor location”. In: *Sensors* 20.4, p. 1139.
- (2020b). “An UWB/vision fusion scheme for determining pedestrians’ indoor location”. In: *Sensors* 20.4, p. 1139.
- Liu, Hui et al. (2007). “Survey of wireless indoor positioning techniques and systems”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37.6, pp. 1067–1080.
- Lu, Liang, Carlos Redondo, and Pascual Campoy (2020a). “Optimal frontier-based autonomous exploration in unconstructed environment using RGB-D sensor”. In: *Sensors* 20.22, p. 6507.
- (2020b). “Optimal frontier-based autonomous exploration in unconstructed environment using RGB-D sensor”. In: *Sensors* 20.22, p. 6507.
- Lucas, Chris (2004). “Cybernetics and Stochastic Systems”. In: *Complexity*.
- Lum, Mitchell JH et al. (2008). “Teleoperation of a surgical robot via airborne wireless radio and transatlantic internet links”. In: *Field and Service Robotics: Results of the 6th International Conference*, pp. 305–314.
- Maida, J, C Bowen, and W Pace (2006). “Enhanced lighting techniques and augmented reality to improve human task performance”. In: *NASA Tech Paper TP-2006-213724 July* 46.
- Makhataeva, Zhanat and Huseyin Atakan Varol (2020). “Augmented reality for robotics: A review”. In: *Robotics* 9.2, p. 21.
- Makhataeva, Zhanat, Altay Zhakatayev, and Huseyin Atakan Varol (2019a). “Safety Aura Visualization for Variable Impedance Actuated Robots”. In: *2019 IEEE/SICE International Symposium on System Integration (SII)*, pp. 805–810.

Bibliography

- (2019b). “Safety aura visualization for variable impedance actuated robots”. In: *2019 IEEE/SICE International Symposium on System Integration (SII)*, pp. 805–810.
- Mathews, Stanley (2007). *From agit-prop to free space: The architecture of Cedric Price*. Black Dog Pub. Limited.
- McCullough, Malcolm (2004). *Digital ground: Architecture, pervasive computing, and environmental knowing*. MIT press.
- Mead, Ross and Maja J Mataric (2012). “A probabilistic framework for autonomous proxemic control in situated and mobile human-robot interaction”. In: *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 193–194.
- Mead, Ross and Maja J Matarić (2017a). “Autonomous human-robot proxemics: socially aware navigation based on interaction potential”. In: *Autonomous Robots* 41.5, pp. 1189–1201.
- (2017b). “Autonomous human-robot proxemics: socially aware navigation based on interaction potential”. In: *Autonomous Robots* 41.5, pp. 1189–1201.
- Menchaca-Brandan, M Alejandra et al. (2007). “Influence of perspective-taking and mental rotation abilities in space teleoperation”. In: *Proceedings of the ACM/IEEE International Conference on Human-robot interaction*, pp. 271–278.
- Merleau-Ponty, Maurice (2013). *Phenomenology of perception*. Routledge.
- Merz, Hermann, Thomas Hansemann, and Christof Hübner (2009). *Building automation*. Springer.
- Milgram, Paul and Fumio Kishino (1994a). “A taxonomy of mixed reality visual displays”. In: *IEICE TRANSACTIONS on Information and Systems* 77.12, pp. 1321–1329.
- (1994b). “A taxonomy of mixed reality visual displays”. In: *IEICE TRANSACTIONS on Information and Systems* 77.12, pp. 1321–1329.
- Milgram, Paul, Shumin Zhai, et al. (1993a). “Applications of augmented reality for human-robot communication”. In: *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'93)*. Vol. 3, pp. 1467–1472.
- (1993b). “Applications of augmented reality for human-robot communication”. In: *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'93)*. Vol. 3, pp. 1467–1472.
- Mistry, Pranav and Pattie Maes (2009). *SixthSense: a wearable gestural interface*.

Bibliography

- Mitsunaga, Noriaki et al. (2006). “Robot behavior adaptation for human-robot interaction based on policy gradient reinforcement learning”. In: *Journal of the Robotics Society of Japan* 24.7, pp. 820–829.
- Mitterberger, Daniela, Evgenia-Makrina Angelaki, et al. (2022). “Extended reality collaboration: Virtual and mixed reality system for collaborative design and holographic-assisted on-site fabrication”. In: *Design Modelling Symposium Berlin*, pp. 283–295.
- Mitterberger, Daniela, Selen Ercan Jenny, et al. (2022a). “Interactive robotic plastering: Augmented interactive design and fabrication for on-site robotic plastering”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–18.
- (2022b). “Interactive robotic plastering: Augmented interactive design and fabrication for on-site robotic plastering”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–18.
- Moniz, António and Bettina-Johanna Krings (2014). “Technology assessment approach to human-robot interactions in work environments”. In: *2014 7th International Conference on Human System Interactions (HSI)*, pp. 282–289.
- Moniz, António B and Bettina-Johanna Krings (2016a). “Robots working with humans or humans working with robots? Searching for social dimensions in new human-robot interaction in industry”. In: *Societies* 6.3, p. 23.
- (2016b). “Robots working with humans or humans working with robots? Searching for social dimensions in new human-robot interaction in industry”. In: *Societies* 6.3, p. 23.
- Mourtzis, Dimitris, Vasilios Zogopoulos, and Fotini Xanthi (2019). “Augmented reality application to support the assembly of highly customized products and to adapt to production re-scheduling”. In: *The International Journal of Advanced Manufacturing Technology* 105, pp. 3899–3910.
- Mulloni, Alessandro et al. (2009). “Indoor positioning and navigation with camera phones”. In: *IEEE Pervasive Computing* 8.2, pp. 22–31.
- Mumford, Lewis (2010). *Technics and civilization*. University of Chicago Press.
- Mumm, Jonathan and Bilge Mutlu (2011a). “Human-robot proxemics: physical and psychological distancing in human-robot interaction”. In: *Proceedings of the 6th international conference on Human-robot interaction*, pp. 331–338.

Bibliography

- (2011b). “Human-robot proxemics: physical and psychological distancing in human-robot interaction”. In: *Proceedings of the 6th international conference on Human-robot interaction*, pp. 331–338.
- Munters, Gerald Martijn (2017). *Playful mediation of perception and engagement: urban space and location-based augmented reality applications through the lens of play*.
- Mur-Artal, Raul, Jose Maria Martinez Montiel, and Juan D Tardos (2015a). “ORB-SLAM: a versatile and accurate monocular SLAM system”. In: *IEEE transactions on robotics* 31.5, pp. 1147–1163.
- (2015b). “ORB-SLAM: a versatile and accurate monocular SLAM system”. In: *IEEE transactions on robotics* 31.5, pp. 1147–1163.
- Mur-Artal, Raul and Juan D Tardós (2017). “Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras”. In: *IEEE transactions on robotics* 33.5, pp. 1255–1262.
- Negroponte, Nicholas (1975). *Soft architecture machines*. MIT press Cambridge, MA.
- Nguyen, Thien Hoang, Thien-Minh Nguyen, and Lihua Xie (2020). “Tightly-coupled ultra-wideband-aided monocular visual SLAM with degenerate anchor configurations”. In: *Autonomous Robots* 44.8, pp. 1519–1534.
- (2021). “Range-focused fusion of camera-IMU-UWB for accurate and drift-reduced localization”. In: *IEEE Robotics and Automation Letters* 6.2, pp. 1678–1685.
- Nguyen, Thien-Minh, Thien Hoang Nguyen, et al. (2019). “Integrated uwb-vision approach for autonomous docking of uavs in gps-denied environments”. In: *2019 International Conference on Robotics and Automation (ICRA)*, pp. 9603–9609.
- Nguyen, Thien-Minh, Shenghai Yuan, et al. (2022). “Ntu viral: A visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint”. In: *The International Journal of Robotics Research* 41.3, pp. 270–280.
- Niemeyer, Günter et al. (2016). “Telerobotics”. In: *Springer handbook of robotics*, pp. 1085–1108.
- Norman, Donald A (1988). *The psychology of everyday things*. Basic books.
- (2016). *Living with complexity*. MIT press.
- Noyes, M V (1984). “A novel predictor for telemanipulation through a time delay”. In: *Proc. of Annual Conf. on Manual Control, Moffet Field, CA, NASA ARC*.

Bibliography

- Nurunnabi, Abdul, David Belton, and Geoff West (2012). “Robust segmentation in laser scanning 3D point cloud data”. In: *2012 International Conference on Digital Image Computing Techniques and Applications (DICTA)*, pp. 1–8.
- Okamura, Allison M (2004). “Methods for haptic feedback in teleoperated robot-assisted surgery”. In: *Industrial Robot: An International Journal* 31.6, pp. 499–508.
- Okamura, Allison M, Mark R Cutkosky, and Jack T Dennerlein (2001). “Reality-based models for vibration feedback in virtual environments”. In: *IEEE/ASME transactions on mechatronics* 6.3, pp. 245–252.
- Okamura, Allison M, Lawton N Verner, et al. (2011). “Force feedback and sensory substitution for robot-assisted surgery”. In: *Surgical robotics*, pp. 419–448.
- Ong, S K, J W S Chong, and A Y C Nee (2010). “A novel AR-based robot programming and path planning methodology”. In: *Robotics and Computer-Integrated Manufacturing* 26.3, pp. 240–249.
- Ong, Soh-Khim et al. (2020). “Augmented reality-assisted robot programming system for industrial applications”. In: *Robotics and Computer-Integrated Manufacturing* 61, p. 101820.
- Ong, Soh-Khim et al. (2020). “Augmented reality-assisted robot programming system for industrial applications”. In: *Robotics and Computer-Integrated Manufacturing* 61, p. 101820.
- Ortega, Michael (2013). “Hook: Heuristics for selecting 3D moving objects in dense target environments”. In: *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 119–122.
- Ostanin, Mikhail, Rauf Yagfarov, and Alexandr Klimchik (2019). “Interactive robots control using mixed reality”. In: *IFAC-PapersOnLine* 52.13, pp. 695–700.
- Oufqir, Zainab, Abdellatif El Abderrahmani, and Khalid Satori (2020). *From Marker to markerless in augmented reality*.
- Özkul, Didem (2015). “Location as a sense of place: Everyday life, mobile, and spatial practices in urban spaces”. In: *Mobility and locative media. Mobile communication in hybrid spaces*, pp. 101–116.
- Öztireli, A Cengiz, Gael Guennebaud, and Markus Gross (2009). “Feature preserving point set surfaces based on non-linear kernel regression”. In: *Computer graphics forum*. Vol. 28, pp. 493–501.
- Pask, Gordon (1968). “The colloquy of mobiles”. In: *Cybernetic Serenpidity, London, ICA; Institue of Contemporary Arts*.

Bibliography

- (1969). “The architectural relevance of cybernetics”. In: *Architectural Design* 39.9, pp. 494–496.
- (1971a). “A comment, a case history and a plan”. In: *Cybernetics, art and ideas*, pp. 76–99.
- (1971b). “A comment, a case history and a plan”. In: *Cybernetics, art and ideas*, pp. 76–99.
- Patompak, Pakpoom et al. (2020a). “Learning proxemics for personalized human-robot social interaction”. In: *International Journal of Social Robotics* 12, pp. 267–280.
- (2020b). “Learning proxemics for personalized human-robot social interaction”. In: *International Journal of Social Robotics* 12, pp. 267–280.
- Peng, Huaishu et al. (2018). “RoMA: Interactive fabrication with augmented reality and a robotic 3D printer”. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12.
- Penza, Veronica et al. (2017). “enVisors: enhanced Vision system for robotic surgery. a User-Defined safety Volume Tracking to Minimize the risk of intraoperative Bleeding”. In: *Frontiers in Robotics and AI* 4, p. 15.
- Perez-Grau, F J et al. (2017). “Semi-autonomous teleoperation of UAVs in search and rescue scenarios”. In: *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 1066–1074.
- Perez-Grau, Francisco Javier et al. (2017). “Multi-modal mapping and localization of unmanned aerial robots based on ultra-wideband and RGB-D sensing”. In: *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 3495–3502.
- Pfister, Hanspeter et al. (2000). “Surfels: Surface elements as rendering primitives”. In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pp. 335–342.
- Pickering, Andrew (2010a). *The cybernetic brain: Sketches of another future*. University of Chicago Press.
- (2010b). *The cybernetic brain: Sketches of another future*. University of Chicago Press.
- (2010c). *The mangle of practice: Time, agency, and science*. University of Chicago Press.
- Picon, Antoine and Alessandra Ponte (2003). *Architecture and the sciences: exchanging metaphors*. Vol. 4. Princeton Architectural Press.

Bibliography

- Pitman, David and Mary L Cummings (2012). “Collaborative exploration with a micro aerial vehicle: a novel interaction method for controlling a mav with a hand-held device”. In: *Advances in Human-Computer Interaction 2012*, pp. 18–18.
- Price, Cedric and Samantha Hardingham (1984). *Cedric Price*. Architectural Association.
- Prieto, SA, N Giakoumidis, and B García de Soto (2021). “AutoCIS: An Automated Construction Inspection System for Quality Inspection of Buildings”. In: *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. Vol. 38. IAARC Publications, pp. 669–676.
- Qi, Charles R et al. (2017). “Pointnet: Deep learning on point sets for 3d classification and segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660.
- Qian, Long, Anton Deguet, and Peter Kazanzides (2018a). “ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery”. In: *Healthcare technology letters* 5.5, pp. 194–200.
- (2018b). “ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery”. In: *Healthcare technology letters* 5.5, pp. 194–200.
- Qian, Long, Anton Deguet, Zerui Wang, et al. (2019). “Augmented reality assisted instrument insertion and tool manipulation for the first assistant in robotic surgery”. In: *2019 International Conference on Robotics and Automation (ICRA)*, pp. 5173–5179.
- Qian, Long, Jie Ying Wu, et al. (2019a). “A review of augmented reality in robotic-assisted surgery”. In: *IEEE Transactions on Medical Robotics and Bionics* 2.1, pp. 1–16.
- (2019b). “A review of augmented reality in robotic-assisted surgery”. In: *IEEE Transactions on Medical Robotics and Bionics* 2.1, pp. 1–16.
- Quarles, John et al. (2008). “Tangible user interfaces compensate for low spatial cognition”. In: *2008 IEEE Symposium on 3D User Interfaces*, pp. 11–18.
- Quintero, Camilo Perez et al. (2018). “Robot programming through augmented trajectories in augmented reality”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1838–1844.
- Richter, Florian et al. (2019). “Augmented reality predictive displays to help mitigate the effects of delayed telesurgery”. In: *2019 International Conference on Robotics and Automation (ICRA)*, pp. 444–450.

Bibliography

- Rios-Martinez, Jorge, Anne Spalanzani, and Christian Laugier (2015). “From proxemics theory to socially-aware navigation: A survey”. In: *International Journal of Social Robotics* 7, pp. 137–153.
- Rosenberger, Robert and Peter-Paul Verbeek (2015). “A field guide to postphenomenology”. In: *Postphenomenological investigations: Essays on human-technology relations*, pp. 9–41.
- Rousseeuw, Peter J and Christophe Croux (1993). “Alternatives to the median absolute deviation”. In: *Journal of the American Statistical association* 88.424, pp. 1273–1283.
- Satyanarayanan, Mahadev (2001). “Pervasive computing: Vision and challenges”. In: *IEEE Personal communications* 8.4, pp. 10–17.
- Schmalstieg, Dieter and Daniel Wagner (2007). “Experiences with handheld augmented reality”. In: *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 3–18.
- Shehan, Erika and W Keith Edwards (2007). “Home networking and HCI: What hath God wrought?” In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 547–556.
- Sheridan, T B (1987). “MIT research in telerobotics”. In: *Jet Propulsion Lab., California Inst. of Tech., Proceedings of the Workshop on Space Telerobotics, Volume 2*.
- Sheridan, Thomas B (1992). *Telerobotics, automation, and human supervisory control*. MIT press.
- Shin, Choonsung et al. (2009). “CAMAR 2.0: Future direction of context-aware mobile augmented reality”. In: *2009 International Symposium on Ubiquitous Virtual Reality*, pp. 21–24.
- Skarbez, Richard, Missie Smith, and Mary C Whitton (2021). “Revisiting Milgram and Kishino’s Reality-Virtuality Continuum”. In: *Frontiers in Virtual Reality* 2, p. 27.
- Soares, Inês, Marcelo Petry, and António Paulo Moreira (2021). “Programming robots by demonstration using augmented reality”. In: *Sensors* 21.17, p. 5976.
- Soliman, Moataz et al. (2013). “Smart home: Integrating internet of things with web services and cloud computing”. In: *2013 IEEE 5th international conference on cloud computing technology and science*. Vol. 2, pp. 317–320.

Bibliography

- Song, Yang (2020). “BloomShell-Augmented Reality for the assembly and real-time modification of complex curved structure”. In: *eCAADe Conference*. Vol. 9, pp. 345–354.
- Song, Yang, Mingyang Guan, et al. (2019). “Uwb/lidar fusion for cooperative range-only slam”. In: *2019 international conference on robotics and automation (ICRA)*, pp. 6568–6574.
- Song, Yang, Richard Koeck, and Shan Luo (2021). “AR Digi-Component - AR-assisted,real-time,immersive design and robotic fabrication workflow for parametric architectural structures”. In: *CAADRIA proceedings*. Hong Kong: CAADRIA.
- Stadler, Susanne et al. (2016). “Augmented reality for industrial robot programmers: Workload analysis for task-based, augmented reality-supported robot control”. In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 179–184.
- Stenson, Molly Wright (2022). *Architectural intelligence: How designers and architects created the digital landscape*. mit Press.
- Steinicke, Frank and Wolfgang Stuerzlinger (2015). “Spatial Interfaces”. In: *IEEE Computer Graphics and Applications* 35.4, pp. 6–7.
- Su, Xing et al. (2013). “Uncertainty-aware visualization and proximity monitoring in urban excavation: a geospatial augmented reality approach”. In: *Visualization in engineering* 1, pp. 1–13.
- Suchman, Lucy A (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.
- Sussman, Elisabeth (1997). *Keith Haring*. Whitney museum of American art.
- Sutherland, Ivan E (1968). “A head-mounted three dimensional display”. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pp. 757–764.
- Suzuki, Ryo, Adnan Karim, et al. (2022). “Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–33.
- Suzuki, Ryo, Clement Zheng, et al. (2019). “Shapebots: Shape-changing swarm robots”. In: *Proceedings of the 32nd annual ACM symposium on user interface software and technology*, pp. 493–505.

Bibliography

- Syrdal, Dag Sverre et al. (2007). “A personalized robot companion?-The role of individual differences on spatial preferences in HRI scenarios”. In: *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 1143–1148.
- Szafir, Daniel (2019). “Mediating human-robot interactions with virtual, augmented, and mixed reality”. In: *International Conference on Human-Computer Interaction*, pp. 124–149.
- Szafir, Daniel, Bilge Mutlu, and Terrence Fong (2017). “Designing planning and control interfaces to support user collaboration with flying robots”. In: *The International Journal of Robotics Research* 36.5-7, pp. 514–542.
- Takayama, Leila and Caroline Pantofaru (2009). “Influences on proxemic behaviors in human-robot interaction”. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5495–5502.
- Tan, Yi et al. (2021). “Automatic inspection data collection of building surface based on BIM and UAV”. In: *Automation in Construction* 131, p. 103881.
- Tazir, M Lamine et al. (2018). “CICP: Cluster Iterative Closest Point for sparse-dense point cloud registration”. In: *Robotics and Autonomous Systems* 108, pp. 66–86.
- Thomas, P CAUDELL and W M David (1992). “Augmented reality: An application of heads-up display technology to manual manufacturing processes”. In: *Hawaii international conference on system sciences*, pp. 659–669.
- Tomas, David (1995). “Feedback and Cybernetics: Reimagining the Body in the Age of the Cyborg”. In: *Body & Society* 1.3-4, pp. 21–43.
- Torok, Matthew M, Mani Golparvar-Fard, and Kevin B Kochersberger (2014). “Image-based automated 3D crack detection for post-disaster building assessment”. In: *Journal of Computing in Civil Engineering* 28.5, A4014004.
- Tsukayama, Hayley and Ben Guarino (2016). “What is really behind the Pokémon Go craze”. In: *The Washington Post*.
- Tunç, Birkan (2015). “Semantics of object representation in machine learning”. In: *Pattern Recognition Letters* 64, pp. 30–36.
- Vanacken, Lode, Tovi Grossman, and Karin Coninx (2007). “Exploring the effects of environment density and target visibility on object selection in 3D virtual environments”. In: *2007 IEEE symposium on 3D user interfaces*.

Bibliography

- Vanderborght, Bram et al. (2013). “Variable impedance actuators: A review”. In: *Robotics and autonomous systems* 61.12, pp. 1601–1614.
- Varela, Francisco G, Humberto R Maturana, and Ricardo Uribe (1974). “Autopoiesis: The organization of living systems, its characterization and a model”. In: *Biosystems* 5.4, pp. 187–196.
- Varela, Francisco J, Evan Thompson, and Eleanor Rosch (2016). *The embodied mind: Cognitive science and human experience*. MIT press.
- Vasquez-Gomez, J Irving et al. (2014a). “Volumetric next-best-view planning for 3D object reconstruction with positioning error”. In: *International Journal of Advanced Robotic Systems* 11.10, p. 159.
- (2014b). “Volumetric next-best-view planning for 3D object reconstruction with positioning error”. In: *International Journal of Advanced Robotic Systems* 11.10, p. 159.
- Verbeek, Peter-Paul (2002). “Devices of engagement: on Borgmann’s philosophy of information and technology”. In: *Techn{é}: Research in Philosophy and Technology* 6.1, pp. 48–63.
- (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press.
- (2008). “Cyborg intentionality: Rethinking the phenomenology of human-technology relations”. In: *Phenomenology and the Cognitive Sciences* 7.3, pp. 387–395.
- Victores, J G et al. (2011). “Robot-aided tunnel inspection and maintenance system by vision and proximity sensor integration”. In: *Automation in Construction* 20.5, pp. 629–636.
- W Chastine, Jeffrey et al. (2007). “Understanding the design space of referencing in collaborative augmented reality environments”. In: *Proceedings of graphics interface 2007*, pp. 207–214.
- W Fitzmaurice, George (1993). “Situated information spaces and spatially aware palmtop computers”. In: *Communications of the ACM* 36.7, pp. 39–49.
- W Krueger, Myron (1977). “Responsive environments”. In: *Proceedings of the June 13-16, 1977, national computer conference*, pp. 423–433.
- Walker, Michael et al. (2018a). “Communicating robot motion intent with augmented reality”. In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 316–324.

Bibliography

- (2018b). “Communicating robot motion intent with augmented reality”. In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 316–324.
- Walters, Michael L, Kerstin Dautenhahn, René Te Boekhorst, et al. (2009). “An empirical framework for human-robot proxemics”. In: *Procs of new frontiers in human-robot interaction*.
- Walters, Michael L, Kerstin Dautenhahn, Kheng Lee Koay, et al. (2005). “Close encounters: Spatial distances between people and a robot of mechanistic appearance”. In: *5th IEEE-RAS International Conference on Humanoid Robots, 2005*. Pp. 450–455.
- Walters, Michael L, Dag S Syrdal, et al. (2008). “Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion”. In: *Autonomous Robots* 24, pp. 159–178.
- Wang, Chen et al. (2017). “Ultra-wideband aided fast localization and mapping system”. In: *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 1602–1609.
- Wang, Peng et al. (2019). “A comprehensive survey of AR/MR-based co-design in manufacturing”. In: *Engineering with Computers*, pp. 1–24.
- Wang, Xiangyu (2008a). “Improving human-machine interfaces for construction equipment operations with mixed and augmented reality”. In: *Robotics and Automation in Construction*, pp. 349–362.
- (2008b). “Improving human-machine interfaces for construction equipment operations with mixed and augmented reality”. In: *Robotics and Automation in Construction*, pp. 349–362.
- Wang, Xiangyu and Phillip S Dunston (2006). *Mixed Reality—Enhanced Operator Interface for Teleoperation Systems in Unstructured Environment*.
- (2011). “Comparative effectiveness of mixed reality-based virtual environments in collaborative design”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41.3, pp. 284–296.
- Webel, Sabine et al. (2013). “An augmented reality training platform for assembly and maintenance skills”. In: *Robotics and autonomous systems* 61.4, pp. 398–403.
- Weichel, Christian et al. (2014). “MixFab: a mixed-reality environment for personal fabrication”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3855–3864.

Bibliography

- Welch, Greg and Eric Foxlin (2002). “Motion tracking: No silver bullet, but a respectable arsenal”. In: *IEEE Computer graphics and Applications* 22.6, pp. 24–38.
- Wen, Rong, Chee-Kong Chui, et al. (2013). “Projection-based visual guidance for robot-aided RF needle insertion”. In: *International journal of computer assisted radiology and surgery* 8, pp. 1015–1025.
- Wen, Rong, Wei-Liang Tay, et al. (2014). “Hand gesture guided robot-assisted surgery based on a direct augmented reality interface”. In: *Computer methods and programs in biomedicine* 116.2, pp. 68–80.
- Wiener, Norbert (1988). *The human use of human beings: Cybernetics and society*. 320. Da Capo Press.
- (2019). *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press.
- Williams, Tom et al. (2018). “Virtual, augmented, and mixed reality for human-robot interaction”. In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 403–404.
- Willis, Karl DD et al. (2010). “Interactive fabrication: new interfaces for digital fabrication”. In: *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pp. 69–72.
- Wojciechowski, Rafal et al. (2004). “Building virtual and augmented reality museum exhibitions”. In: *Proceedings of the ninth international conference on 3D Web technology*, pp. 135–144.
- Wu, Andy et al. (2011). “Kinesthetic pathways: a tabletop visualization to support discovery in systems biology”. In: *Proceedings of the 8th ACM conference on Creativity and cognition*, pp. 21–30.
- Xu, Hao et al. (2020). “Decentralized visual-inertial-UWB fusion for relative state estimation of aerial swarm”. In: *2020 IEEE international conference on robotics and automation (ICRA)*, pp. 8776–8782.
- Yew, A W W, S K Ong, and A Y C Nee (2017). “Immersive augmented reality environment for the teleoperation of maintenance robots”. In: *Procedia Cirp* 61, pp. 305–310.
- Yoon, Mincheol et al. (2007). “Surface and normal ensembles for surface reconstruction”. In: *Computer-Aided Design* 39.5, pp. 408–420.

Bibliography

- Zhai, Shumin and Paul Milgram (1992). “Telerobotic virtual control system”. In: *Cooperative Intelligent Robotics in Space II*. Vol. 1612, pp. 311–322.
- Zhou, Tianyu, Qi Zhu, and Jing Du (2020). “Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction”. In: *Advanced Engineering Informatics* 46, p. 101170.
- Zhu, Cheng et al. (2015a). “A 3d frontier-based exploration tool for mavs”. In: *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 348–352.
- (2015b). “A 3d frontier-based exploration tool for mavs”. In: *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 348–352.
- Zhu, J, Soh-Khim Ong, and Andrew YC Nee (2015). “A context-aware augmented reality assisted maintenance system”. In: *International Journal of Computer Integrated Manufacturing* 28.2, pp. 213–225.
- Zollmann, Stefanie et al. (2014). “Augmented reality for construction site monitoring and documentation”. In: *Proceedings of the IEEE* 102.2, pp. 137–154.

Appendix **A**

Appendix - Space and Computation

Technological innovation is constantly reshaping the materiality and mechanics of physical space. Especially since the postwar until today, with the miniaturization of computational devices and increased computing power, pervasive and unobtrusive computer systems together with modern information and communication technology (ICT) support humans in various applications in contexts such as home, workspace, and healthcare. The concept of architectural space with infiltrated computational systems have been discussed and represented in multiple disciplines. For instance, in the architectural domain, terms such as building automation and building management discuss automatic centralized control of computational systems in the built environment (Merz, Hanseemann, and Hübner 2009). Responsive environments and interactive architecture focus on the interaction between users and environments mediated by ICT (W Krueger 1977). Adaptive architecture is concerned with buildings that are designed to adapt to environments, inhabitants, and objects, as well as those buildings that are entirely driven by internal data (Brand 1995).

Meanwhile, computer scientists and engineers made several efforts to create systems in built environments. Since the early 1990s, ubiquitous computing and pervasive computing envisions the creation of environments saturated with computing and communication capability, yet gracefully integrated with human users (Satyanarayanan 2001). Following this thread, research initiatives such as ambient intelligence and intelligent environments employ wireless networks, user-friendly interfaces, and intelligent agents to implement the vision of ubiquitous computing.

Appendix A. Appendix - Space and Computation

Communities from architecture and computer science share interests in the integration of technology in the built environment. Many scholars have documented and traced the history of the connection between space and technology. For example, *Architectural Intelligence* (Steenson 2022) explores the work of architects from in the 1960s and 1970s who incorporated elements of interactivity into their works. The review of the works of architects such as Christopher Alexander, Richard Saul Wurman, Cedric Price, and Nicholas Negroponte and the MIT Architecture Machine Group laid the foundation of many contemporary interactive practices in built environments. Usman Haque, a British architect and researcher, draws specifically from the contribution of Gordon Pask on Cedric Price's Fun Palace project to illustrate the relationship between Pask's Conversation Theory and interactivity in architecture enabled by ubiquitous computing devices (Haque 2007a). In a similar vein, Malcolm McCullough offers an account of the intersections of architecture and interaction design, arguing that the ubiquitous technology does not obviate the human need for place (McCullough 2004).

From a broader scope, Andrew Pickering, a British sociologist, philosopher ,and historian of science, examines a largely forgotten group of British thinkers such as Grey Walter, Ross Ashby, and Gregory Bateson, and their singular work in a dazzling array of fields. Pickering contextualizes the rich history of cybernetics from psychiatry, architecture, engineering, and the sixties counterculture and contends cybernetics as a vision of the world as ultimately unknowable and a place where novelty is always emerging (Pickering 2010a). The essay collection *Architecture and the Sciences: Exchanging Metaphors* (Picon and Ponte 2003) illuminates how scientific paradigms migrated to architecture and what caused the migration to take place. Siegfried Giedion traces "mechanization" within the context of Western culture in a zig-zag and scattered manner. By contextualizing "machine" in architecture, especially domestic environments, the work of Giedion illustrates how a mechanized world impact on the human organism and on human feelings(Giedion 1948). In the work *Techics and Civilization*(Mumford 2010), Lewis Mumford presented a compelling history of the machine and a critical study of its effects on civilization. Mumford traced the origin of the machine age and its social results and critically argued that the development of modern technology had its roots in the Middle Ages rather than the Industrial Revolution. This viewpoint is thought-provoking to look to the Middle Ages, or even before it, to find human values as the impetus for what

Appendix A. Appendix - Space and Computation

would later become a wide-scale dependence on machines, rather than taking the advent of the steam engine as an autonomous causal force.

The aforementioned works provide a historical framework and create a space for opening up a discussion on the relationship between architecture, computation, and machines.

Specifically, in this section, I am drawing particularly from the experiment since the early 1960s, when cybernetics, together with the technology development originally for military applications, made inroads into architecture. Although the name of cybernetics gradually faded into obscurity, the concepts such as homeostasis, autopoiesis, circularity, and conversation theory are still profoundly embedded into contemporary practices in building automation, intelligent environment and interactive spaces. Or, as described by Charlie Gere, cybernetics is the direct predecessor of our current digital culture(Gere 2009). The scientific discipline Cybernetics suggested a certain way of thinking about the relationship between computation and the physical world. By retracing the connection between architecture and cybernetics, I aim to articulate how cybernetic thoughts still remain “not-present and not-absent”(Derrida 2012) to the architecture discipline, specifically in the research field within the intersection of computation and architecture. I tentatively define three types of capabilities of built environments augmented by embedded computational systems - 1) automation 2) adaptation 3) interaction and sketch their connections with early cybernetic thoughts.

A.0.1 Automation And Homeostasis

In 1948, Nobert Wiener, an American mathematician and philosopher, defined cybernetics as the “science of control and communication in the animal and the machine, organisms, and mechanisms”(Wiener 2019). He suggested an environmentally responsive machine that would put the principles of the living to work in the world of the non-living (Wiener 1988). The machine can bridge, ontologically and technically, the gap between the organic and the inorganic, the natural and the artificial, organisms, and technical objects.

Wiener’s standpoint arises from the fact that both entities could be regarded as self-regulating machines, able to control (or better reduce) entropy through feedback; that is, the capacity to regulate future behavior from previous operations and eventually learn from their experience. These machines had actuators to perform specific tasks, and sensors that allowed them to

Appendix A. Appendix - Space and Computation

record external conditions as much as their own operation, and return this information as input for their next action in a continuous self-regulatory circuit (Wiener 1988).

Self-regulation characterized the systems of the first cybernetic wave (1945-1960). The term “homeostasis”, coined by physiologist Walter Cannon, describes the ability of animal bodies to maintain a stable energy status, and to minimize energy loss in relation to environmental changes (Cannon 1929). Respectively, cyberneticists considered the homeostatic cybernetic model to be able to minimize information loss in the system, in favor of control and stability, through “negative feedback”(Hayles 1994).

A homeostatic machine would resist the tendency towards disorder, due to environmental disturbances, i.e. “noise,” eliminating possible unexpected behaviors and state changes(Tomas 1995). Such a system would be predictable, because it would use feedback loops to constantly follow its target, and maintain the stability of its state of operation, correcting all possible destabilizing activities.

The concept of homeostasis has then become a central position of cybernetics and further been introduced to architecture. In 1972, Charles Eastman developed the concept of “adaptive-conditional architecture,” which self-adjusts based on the feedback from the spaces and users, that of a boat constantly correcting its course against environmental disturbances (Eastman 1972). He used an analogy of a thermostat to describe the essential components: sensors that would register changes in the environment, control mechanisms (or algorithms) that would interpret sensor readings, actuators as devices that would produce changes in the environment, and a device (an interface) that would let users enter their preferences.

Similarly, roughly at the same time, Nicholas Negroponte was working on his “architecture machines” — a machine controlled environment. In one of his experiments, a greenhouse with opening-closing glass was regarded as a homeostatic system. The building is able to decode, encode and interpret the level of comfort of plants and functions under thermostat mechanisms (Negroponte 1975). Negroponte’s experimental work “SEEK” consisted of a Plexiglass encased, computer-controlled environment full of small blocks and inhabited by gerbils, who continuously changed the position of the blocks. Following instructions programmed by the authors the robotic arm automatically rearranged the blocks in a specific pattern. Once the arrangement was disrupted, a computer-controlled robotic arm rebuilt the block configurations in a manner its programmers believed followed the gerbil’s objectives (Figure A.1). The project

Appendix A. Appendix - Space and Computation



Figure A.1: "Seek", 1970 by Nicholas Negroponte with the Architecture Machine Group.

was an experiment in the symbiotic relationship between living organisms and machines. The model can be regarded as a model of a first-order cybernetics (Homeostatic) approach to architecture.

Computation embedded in architectural spaces serves as a central program, collecting environmental sensor input and sending an optimal instruction through regulatory feedback to actuators to change the state of the environment. Computation was used as either an inspirational tool or as literal technology to model self-regulation and eliminate system noise. Extending from the thought of computation as regulation, contemporary practices such as home automation aims to construct a feedback loop to maintain the built environment at a target status through a series of pre-programmed actions enacted by built-in actors. For example, in the 1950s, General Electric built All Electric House at Kansas, the house was equipped with automated functions such as turning on/off the lights when the time was right, watering the garden or making coffee at a specified time. Since then, many researchers, following this direction, have developed techniques to enhance and automate spatial services, by assessing situations and human needs to optimize control and performance in architectural space. Extending from this thread, recent works including adaptive house (D. Cook and Sajal Kumar Das 2004), Georgia Tech Aware Home (Kientz et al. 2008) , Orange (Harper 2003) equip

Appendix A. Appendix - Space and Computation

domestic space with networked sensors and actuators to automate and control multiple disparate systems. Focusing on "home automation", SpeakEasy (Edwards et al. 2002) improves device interoperability for efficient control. A related set of research has studied home networks (Grinter et al. 2005, Shehan and Edwards 2007), exploring how households configure, manage, live with their home networks and manage access control for sharing data and devices.

However, there are some limitations to the early concepts of computationally augmented space based on first-order cybernetic thinking. Firstly, the early concept of cybernetics was built upon a self-adjustable system aiming towards a predetermined and fixed goal. This means that the goal or next state of the system is fully specified by the combination of the system inputs (Lucas 2004). Laws of Requisite Variety states that, for any system to be controlled, or, to use one of the two other cybernetic synonyms, managed (the other is regulated), the variety or the number of states in the controller must exceed the variety in the system to be controlled (Ashby and Goldstein 2011). The mismatch between varieties of the physical world and computational space might turn the notion of control into control-as-restricting. Especially, compared to finite size of computation, problem definitions and solutions in physical environments can quickly become "transcomputable"(Bremermann 1977). The determinist framework that shared among many post-war cybernetists seem to be somehow inherently lacking to address undefinable real world states. Secondly, from the viewpoint of early cybernetics, users are detached from their inhabited computationally augmented space, which renders users as observers standing independently from the environment. This notion is in contradiction with the fact that users within the built environment are active as actors, and their actions are mostly unpredictable. Thus, the algorithm for regulating space might oversimplify the contexts in which decisions are already made without considering the presence of users. As a result, digitally-driven structures and computationally augmented spaces, like those discussed previously, only respond to stimuli or make decisions for proactive operations; more relational and interactive engagement with users are missing from the design. This deficiency was identified in the 1960s, by cyberneticists like Gordon Pask and cybernetics-inspired architects like Cedric Price. They proposed to rethink cybernetics with integration of users and focus on computationally mediated relations between users and architectural space. The shift in concept leads to a novel understanding of computationally augmented space as mutual rela-

Appendix A. Appendix - Space and Computation

tions, engagement and cooperation between users and the built environment. I will elaborate the concept in the next chapter - adaptation and autopoeisis.

A.0.2 Adaptation and Autopoeisis

In the previous chapter, I reviewed a variety of early experiments and projects under the theme of computationally augmented space influenced by post-war cybernetic thinking. Almost all projects shared an overall characteristic of embedding sensors and actuators and use computation to regulate space in adaptation to ever-changing context. However, like the early cybernetic thinking, early projects on computationally augmented space were conceptually bound to simplify computation as regulation - a series of pre-determined actions to external stimulus. As mentioned earlier, we might miss a more fertile conceptual ground by ignoring the indeterminacy of the physical world and the agency of users. As cybernetics moved into a system that include users rather than standing independent of them, or namely “the second order cybernetics”, computationally augmented space that followed this line of thinking gradually shifted from a self-regulatory system to a more interactive system.

In 1968, Mead’s paper (commissioned by Heinz von Foerster (Foerster 2003a)) in which she asked why cyberneticians did not treat their own systems as a cybernetic system: why not treat a cybernetics society through cybernetics itself. Hence, the title of the paper “The cybernetics of cybernetics” that also came to be called the new cybernetics and, more commonly, second-order cybernetics. The change of paradigm between first- and second-order cybernetics is summarized in the distinction between being part of and apart from the world. In a first-order account, one describes the world objectively as an independent observer; in a second-order account one is part of the world and so part of one’s own observation.

“Should one name one central concept, a first principle, of cybernetics, it would be circularity.”, stated by Heinz von Foerster. Compared to scientific systems that traditionally aim to be linear, cybernetic systems are mostly circular. Second-order cybernetics has developed several very particular circular systems. The most famous of these is the Autopoietic system of Varela, Maturana and Uribe (F. G. Varela, Maturana, and Uribe 1974). Autopoietic systems, also called initially autopoietic machines, explore the general relational scheme common to all living systems as the configuration of transformative processes whose result is the configuration itself, so that identity and activity, producer and product coincide. Unlike Turing

Appendix A. Appendix - Space and Computation

machines, set by external programmers (thus being heteropoietic) to compute problems referring to issues other than the system itself (thus being allopoietic), autopoietic machines realize a self-defined identity in a space of interactions (Bich and Etxeberria 2013). As noted by von Foerster, the self-referential autopoietic system provides a model for how, by a process of repeated action (such as observing) we can arrive at a stable and fixed outcome (Foerster 2003b). The concept of self-referential and self-producing autopoietic system challenges early cybernetic thinking based on a determinist framework by replacing external forces with internal interactions between agents.



Figure A.2: Performative Ecologies, Ruairi Glynn, 2008

Second-order cybernetics or autopoietic systems provides fertile conceptual ground for novel practices in computationally augmented space. For example, Glynn's work, *Performative Ecologies* (Glynn 2008), demonstrates the possibility that autopoietic systems in space can propose new unexpected behaviors. In this case the architect's role is to design and construct the framework within which the augmented environment will explore and develop its own behaviors and functions adaptively (Glynn 2007). The installation consists of three suspended robotic sculpture, each of them is programmed with an internal motive to attract visitor's attention. Instead of acting in a predetermined manner, three sculptures repetitively try to direct themselves and assess the levels of perceptual attention by analyzing detected face of visitors captured by an embedded camera. Through multiple recursions, the sculptures can determine which motions and behaviors are most successful in terms of attention, discarding unsuccessful actions (Figure A.2).

Appendix A. Appendix - Space and Computation

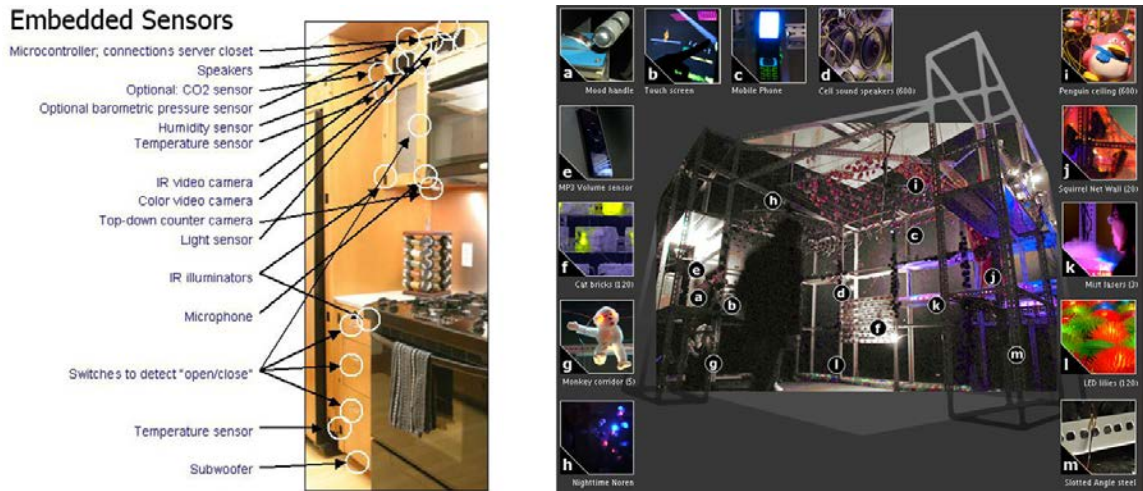


Figure A.3: Left, PlaceLab Project at MIT (2003), Right, the Re configurable House (2008)

This project demonstrates how internal motives of agents can generate indeterminate responses that are not pre-choreographed. It provides a novel framework to augment and animate physical space, shifting from self-regulation to self-production. Unlike a simple determinist reactive system, the system's behavior is emergent and unpredictable, operating according to the experience gained from interaction with humans.

In the context of autopoiesis, mental representations and cognitive processes are not symbolic and disembodied, as in a classical cognitivist view that favors a Cartesian dualist attitude to the mindbody relation, but rather embodied and situated, emerging out of the constraints of the body and its sensory-motor interactions with the environment. In adjacent fields such as robotics, the argument for embodied intelligence and control also emerges around the 1980s. Prior to the 1980s, the main trend was founded on a Cartesian model in Western epistemology, which emphasized centralized control, reflecting what Kevin Kelly considered to be the first type of machines that can be constructed, that is, mechanical predictable and hierarchical systems, such as the clock, following linear and sequential operations commanded by a central authority. This was featured in industrial robotic applications, which involved precise and stable operations guided by a particular albeit flexible software. Service robots, on the other hand, which would inhabit the social environment and cater for human needs, seemed incapable of coping with the uncertainty and fuzziness of everyday spaces. Rodney Brooks, in the 1980s, designed and constructed robots that would directly interact with their environment

Appendix A. Appendix - Space and Computation

through sensory-motor operations, without central control and pre-programmed representations. Inspired by the behavior of simple primitive organisms, these physical machines were equipped with a distributed system of sensors and actuators, and functioned by prioritizing actions according to a hierarchical set of in-built behavioral modes. Thus, intelligence was not an aspect of logical calculations on the basis of pre-given representations as in symbolic AI, but an emergent phenomenon of machines that could interact with local situations in the environment in which they were embedded, constantly adjusting their behavior by suppressing actions of lower priority. Intelligence would thus be, as Brooks stated, “in the eye of the observer.” (R. A. Brooks 1991a).

Extending from the concept of autopoiesis and non-representative AI, some contemporary research works on intelligent environments define computation more than ‘automation’ or ‘regulation’, properties like learning and proactivity become part of computationally augmented space. Together with the developments in deep learning and machine learning, projects, such as (D. J. Cook, Youngblood, and Sajal K Das 2006, Gottfried and Aghajan 2009), use environmental information as feedback to acquire knowledge and experience to optimize control and efficiency. The embedded computational system would be able to learn how to operate and respond efficiently despite the changes in user preferences or the layout of space (Aarts and Wichert 2009). For example, the House_N research (Figure A.3) conducted at the MIT Department of Architecture explores how new technologies, materials and strategies in design can lead to dynamic and evolving spaces that are responsive to the complexity of life (Intille et al. 2005). Researchers equipped a house with networked sensors including RF and motion detectors to identify user activities, the algorithm then statistically deduce users’ needs and control actuators to respond immediately and proactively. The computational system assists users by simplifying and anticipating their everyday activities and manage energy consumption economically by controlling lighting, heating and ventilation.

However, the issue of self-referential computational system is that the acceleration of technical dominance of the building’s functionality threatens to bypass co-development with the building inhabitants, potentially leading to non-democratic self-learning systems that make decisions without human involvement, contradicting the aims of sustainable development, which should provide for social sustainability, as well as economic and environmental sustainability. The absence of human in the computational systems may result in technocentrism – the be-

Appendix A. Appendix - Space and Computation

lief in technological innovation as the solution to social and environmental problems, wherein technology dominates human systems.

A.0.3 Interaction and Conversation

The constitution of machine and human can be traced back to Conversation Theory, a particularly coherent and potentially the most productive theory of interaction encompassing human-to-human, human-to-machine and machine-to-machine configurations in a common framework (Haque 2007b).

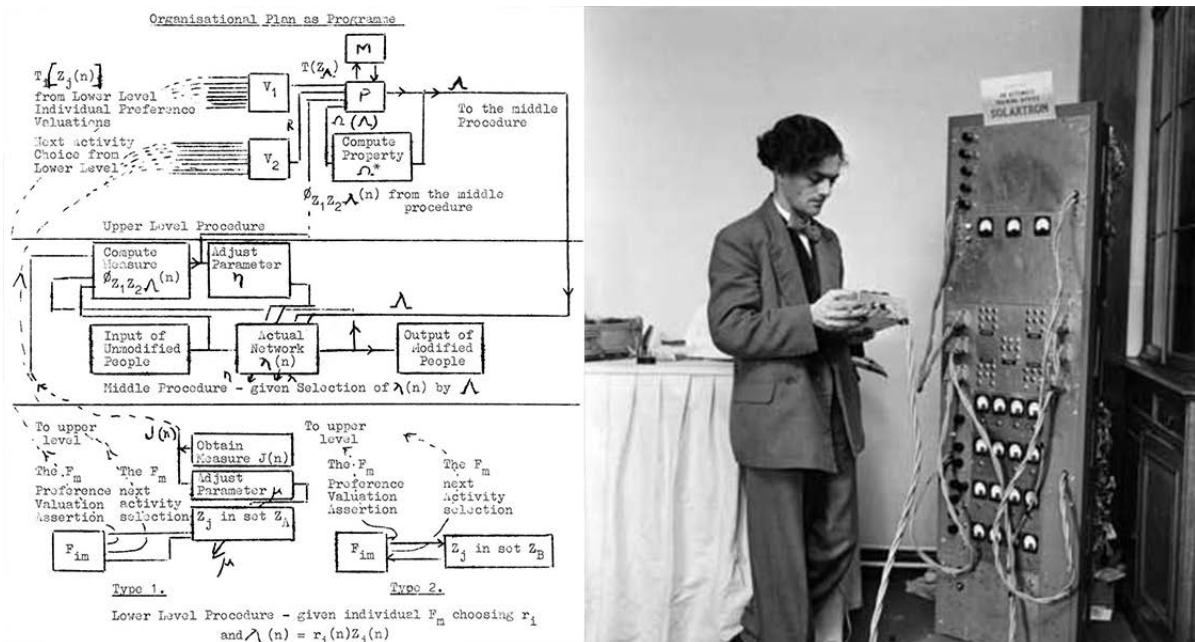


Figure A.4: Left, the Colloquy of Mobile (1968), Right, the MusiColor Machine (1953) by Gordon Pask

To perform the essential and distinguishing characteristics of conversation, Gordon Pask, English scientist, designer, researcher, academic, playwright, built a number of interactive systems, such as the Colloquy of Mobiles installation and the Musicolor machine (Figure A.4). These projects could respond to what Pask considered to be the most prevalent attribute of pleasurable interactive environments, that is, to “respond to a man, engage him in conversation and adapt its characteristics to the prevailing mode of discourse.” (Pask 1971a) In particular, what Pask considered to be a “socially oriented reactive and adaptive environment,” (Pask

Appendix A. Appendix - Space and Computation

1968) the Colloquy of Mobiles, exhibited at the Cybernetic Serendipity show at the ICA in London in 1968, consisted of five rotating suspended figures — the “mobiles” — endowed with the capacity to communicate through audiovisual signals. Three of the “mobiles” were designated as “female” and the other two as “male.” The signals were used in the process of satisfying the figures’ builtin competitive and incompatible goals, namely hierarchically placed “drives.” Represented by two states, an orange and a puce color light (attached on the bodies of the figures), the two “male” drives (which increased their intensity over time), would be satisfied once a light beam of the respective color hit a certain receptor on their bodies. This would be achieved only by “cooperation” with the female figures, which were equipped with reflector items able to send the beam back to the “male” that emitted it. Furthermore, motor-driven rotational and vertical motions of the entities and their reflectors would be engaged to negotiate cooperation and synchronize interaction between male and female, as well as induce “competitive” disturbances against “same sex” figures. The entities could emit specific sounds to signal goal achievement or state of cooperation, as well as learn from their experience and subsequently improve their performance.

This project did not take into account input from human visitors throughout — it was rather concerned mainly with internal interactions. Pask, however, concluded his description of the project wondering whether visitors, by interfering with the installation, would gradually find their own patterns of interaction with it and interpret its signals in a way that they felt would be pleasing. (Pask 1968) In fact, when the show was on, many visitors interacted with it for a long time while some used mirrors to redirect the beam lights (Pickering 2010b).

The Musicolor machine on the other hand, which Pask constructed along with his colleague Robin McKinnon-Wood in the early 1950s and used in various London theaters, was designed for improvisational performances, involving dynamic interactions between a human musician (usually a piano player) and a light display machine — in what Andrew Pickering called “a dance of agency.” (Pickering 2010c) Throughout the course of interaction, the system would detect the output properties (sound frequency, transience and rhythm) of the musician’s performance, and amplify them as electrical signals before filtering them, using a sound-input analog machine (consisting of uniselectors, capacitors and relays). Once the filtered sound signals exceeded certain threshold values, the machine would produce a variety of colored light patterns. To make the system more complex, these values were circuited to vary in time

Appendix A. Appendix - Space and Computation

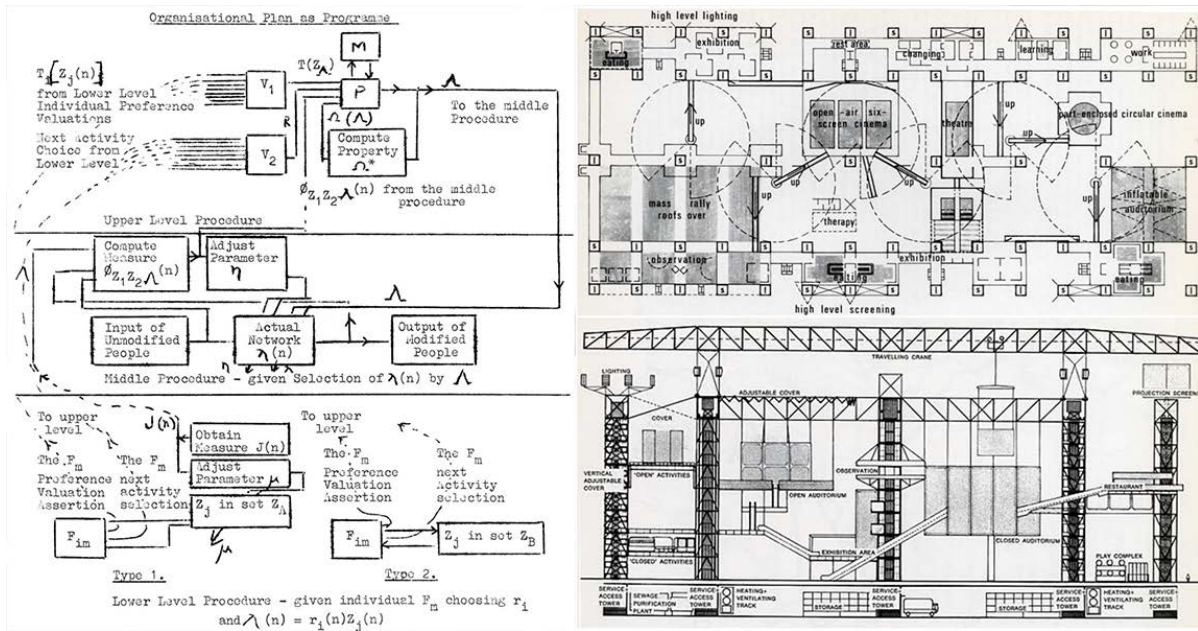


Figure A.5: Left, “Organisational Plan as Programme” (1965), Right, Plan and section drawings of Fun Palace (1960-1966)

depending on previous performance and machine behavior. In effect, as Pask noted, the machine was designed to get bored when repetitive actions were performed and shift towards novel activity (Pask 1971b). While Pask designed the Colloquy to interact internally within its machine parts (although visitors could engage with the installation with unpredictable outcomes), the Musicolor involved a feedback loop between human and machine, which entailed the musical score, the learning capacities of the machine and its visual display. Here the constant and dynamic interference of the user, i.e. the performer’s input (seen as second order cybernetics “disorder”), was an indeterminate factor that played its part in the formation of the overall performance. But not only did the human performance influence the machine output; the latter, with its variable liveliness, was fed back to influence the performer’s behavior and so on (Pickering 2010b). Unlike the Colloquy, the goals of the Musicolor were not built into the machine as fixed drives but were rather shared with the human performer, thus thematizing cooperation, suspending any possibility for control over the overall outcome.

There has also been a ground swell of interest in Pask’s work by architects, artists and designers, though his association with architects can be traced back to the 1960s. In the 1960s, Pask worked with the architect Cedric Price on his Fun Palace projects.

Appendix A. Appendix - Space and Computation

Using cybernetics and the latest computer technologies, Price hoped to create an improvisational architecture which would be capable of learning, anticipating, and adapting to the constantly evolving program. An array of sensors and inputs would provide real-time feedback on use and occupancy to computers which would allocate and alter spaces and resources according to projected needs (Price and Hardingham 1984). Gordon Pask was in charge of cybernetic system in the Fun Palace (Figure A.5), he described in his article, “The Architectural Relevance of Cybernetics”, how the cybernetic paradigm of design could link the aim or intention of users, with computationally augmented space that could start to learn, adapt, and respond to these aims and intentions (Pask 1969). The embedded computational programs could compute overall tendencies of visitors and tune parameters for transforming space by moving walls and corridors. The Fun Palace thus could “learn” users behavioral pattern and proactively plan future activities (Mathews 2007). The Fun Palace was a part of Cybernetic practices and discourses in Britain in the 1940s and 1950s when scientists constructed electromechanical machines that manifested conversational and adaptive behaviours. Similarly, we can see this concept noted by Pask in his text about the relevance of cybernetics to architecture. He proposed that the Fun Palace was a conceptualization of architecture as a partner that can converse with users. Pask mentioned the high point of functionalism is the concept of a house as a “machine for living in.” But the bias is towards a machine that acts as a tool serving the inhabitant. Pask noted that the notion would be refined into the concept of an environment with which the inhabitant cooperates and in which he can externalize his mental processes (Pask 1969). Although the mechanical system of the Fun Palace such as moveable hinged roof and trusses could point to a determinist and functionalist framework, the uncertainty in terms of its controlling programs brings a type of machine that could be different from Corbusier’s living machine. The open-endedness and indeterminacy implied in the Fun Palace depicted a possible future of computationally augmented space that goes beyond determinist limitations of mechanical flexibility.

From the perspective of HCI, space is more seen through the concept of place which is more than a location or a volume (Harrison and Dourish 1996a). Place organizes flows of people, resources, ideas, and confers identities (Foucault 2012). For example, places like markets, streets, and malls establish socio-cultural protocols, conventions, and values, as well as a means for shaping our shared conception of community and individual concepts of identity.

Appendix A. Appendix - Space and Computation

Place that pertains to concrete and situated activities, directly linked to the location and the functioning of intelligent systems. In this section, I will first draw from theories such as embodiment and phenomenology to illustrate the concept of mutual constitution between the physical world and technology and then use examples to demonstrate how the theoretical foundation provides fertile ground for the design of spatial interfaces.

A.0.4 Theoretical Foundation - From Phenomenology To Embodiment

In contrast to the Cartesian dualist view of classical cognitive science and AI, embodied cognitive science scholars argue that physical embodiment is a necessary condition for mental processes. The emphasis is on the role of sensory-motor engagement of organisms with their physical environment in cognitive processes. In a similar vein, HCI scholar, Paul Dourish, argues that the physical embeddedness of an active participatory subject in the real world is necessary for the creation, manipulation and sharing of meaning with the world. As Dourish argues: “embodiment is the property of our engagement with the world that allows us to make it meaningful.” (Dourish 2004b) The actions of subjects in embodied interaction theory are critical for understanding the world and its meanings. But instead of relying on a full non-representational stance towards cognition, (Clark 1997) Dourish, who mostly focuses on the design of software systems, questions but does not reject the role of representations in interaction design. To this end, he lays out a theory to propose general principles and directions for the design of interactive environments and artifacts, considering the centrality of embodiment in human–computer interaction. “What I am claiming for “embodied interaction” is not simply that it is a form of interaction that is embodied, but rather that it is an approach to the design and analysis of interaction that takes embodiment to be central to, even constitutive of, the whole phenomenon. This is certainly a departure from traditional approaches to HCI design.” (Dourish 2004b)

The theory of embodiment can be traced to its conceptual grounding - phenomenology. The phenomenological concept of embodiment refers to the common way by which we confront and experience the social and physical reality of the everyday world. Phenomenologists argue that our experiences depend on our interactions with the world, and the possibilities for actions that are provided by this world. Husserl’s emphasis on the phenomena of experience, Heidegger’s reconstruction of Husserl’s phenomenology through the concept of being as

Appendix A. Appendix - Space and Computation

being-in-world, Schutz's extension of phenomenology to problems of social interaction, and finally Merleau-Ponty's addressing of the role of the body in perception, explore the relations between embodied action and meaning, placing the source of meaning of things already in the world where humans act and which, in turn, acts on us.(Law 2006)

Peter Verbeek, Dutch philosopher of technology, challenges Heidegger's idea of tools as either "ready-to-hand" or "present-at-hand" from the perspective of post-phenomenology. Verbeek suggests that artifacts and technologies can be part of a continuum; they may promote neither full attention and involvement nor withdrawal. They can mediate action without fading into the background, and they may demand engagement by focal involvement or alterity relations to be useable. However, for Verbeek this mediation happens through the material utility of artifacts – their primary function – and not through their socio-cultural utility – their secondary function. In other words, the relations that are mediated take place on the basis of the product's functionality; things mediate relations, shaping human actions and experiences, through their functioning as material objects.(Verbeek 2005) Verbeek's post-phenomenological perspective, artifacts are considered to be mediators of human–world relations, pointing to a mutual constitution of humans and things, actively co-shaping the way humans act in and perceive the world. This is a useful conceptualization of the relations between human users and their technologies in domestic environments.

Similar to the idea that technological artifacts co-shape human actions and experience, Albert Borgmann, a German-born American philosopher, emphasizes the importance of involving users not only with the artifacts' functionality and meaning, such as through interfaces, buttons, signs, and symbols but also with their material machinery through sensory and bodily engagement. Borgmann focuses on specific aspects of information and communication technologies and the way that they help shape people's experiences with reality. In developing a philosophical analysis of information, Borgmann comes to view information as a relationship between humans and reality. Within a certain context, he argues, information consists of signs that inform people about things in reality (Borgmann 2007). A sign is hence not a thing but rather the promise of a thing, a relationship between humans and reality(Verbeek 2002).

The understanding of information as a sign or a promise of a thing echoes James Gibson's theory of affordances, brought into the field of design and HCI by Donald Norman in the late 1980s. In his book *The Design of Everyday Things*, Norman used the term to refer

Appendix A. Appendix - Space and Computation

to the fundamental properties of things that determine how they will be used. He made the distinction between real affordances and perceived affordances, namely what the user perceives the artifact can do.(Norman 1988) This concept is useful for designers who want to make the functionality and internal structure of artifacts, either physical or digital, visually perceivable through meaningful accessible representations and understandable conceptual models. Later, however, Norman dropped the concept to replace it with a more viable term, the signifier, and a more specific term, the social signifier. Social signifier is an indicator, a perceivable clue that provides meaningful information about the physical or social world, informing users about what an artifact can be used for.(Norman 2016)

A.0.5 Technical Development - From Tangible Interaction To Augmented Reality

Aforementioned theories and frameworks share similarities in emphasizing the mutual constitution between the physical world and technology. They lay foundations for technologies in HCI that focuses on the relational aspects of the physical world, to name a few, tangible interaction, context-aware applications and ambient intelligence.

Tangible Interaction

In the context of his embodied interaction theory, Paul Dourish defines tangible computing as an approach to computing that explores “how to get the computer ‘out of the way’ and provide people with a much more direct – tangible – interaction experience.”(Dourish 2004b) Tangible computing applications, which range from work in computer science, as well as art and design, assume the distribution of computational power across different devices or objects that can behave as active agents, reacting to the activities of people, while knowing their location and context. Instead of traditional computational devices, such as graphical computer interfaces and keyboards, tangible computing uses physical artifacts and their affordances in three-dimensional physical space.

There have been several examples and projects since the 1990s – from Durrell Bishop’s Marble Answering Machine and the work of the computer-related design department at the Royal College of Art directed by Gillian Crampton-Smith, to the work of the Tangible Media group at the MIT Media Lab, directed by Hiroshi Ishi. Designers have studied not only how computation moves into the world and out of the desktop PC, but also how computation may

Appendix A. Appendix - Space and Computation

manifest itself when it moves into the physical environment, and how physical manipulation of digitally enhanced tangible artifacts could help naturalize the ways we interact with computation, and thus make those interactions meaningful: “the essence of tangible computing lies in the way in which it allows computation to be manifest for us in the everyday world; a world that is available for our interpretation, and one that is meaningful for us in the ways in which we can understand and act in it.”

Tangible computing objects incorporate a “digital self” – hardware and software components – that represents the artifact, in the digital space, for other artifacts. Digital information is thus embedded in physical form while manifesting itself through physical changes in the domestic environment: for instance, lighting patterns, sound signals or movement of furniture and architectural elements. In tangible computing systems, there is no single point of interaction, a single device to interact with, as in traditional human computer interaction interfaces. The same activity can be distributed and shared between multiple devices or, more specifically, it can be achieved through the coordinated use of these artifacts.(Dourish 2004b) At the same time, interaction does not take place in a successive manner of information exchange, but in parallel, as there is no way to tell the next actions of users.

Thus, the role of designer in tangible computing changes from controller to facilitator. In traditional human–computer interaction, designers fully control the functionality, use and form of their artifacts. They start from pre-specified design models to anticipate the way these artifacts will interact with their users. (Dourish 2004b) But often users interact with and apply meaning onto augmented artifacts in surprising ways, which presents a possible alternative approach to design thinking for augmented artifacts. Meaning is constructed and communicated on many levels within a process that Dourish calls “coupling,” the intentional connection between users and artifacts that arises in the course of interaction. Instead of designing and anticipating the way people will use interactive objects, designers should explore methods to allow users to appropriate the objects, to act both on and through those objects. Users should be able not only to execute the normal functions of artifacts, but also to modify their interface, to break it and appropriate it by relocating its constituent elements. (Dourish 2004b)

This concept is demonstrated in projects, such as the Reconfigurable House (Figure A.6), an experimental proposal for an open source, user-driven interactive installation by Usman Haque and Adam Somlai-Fischer. It consisted of simple sound-, light- and touch-sensitive

Appendix A. Appendix - Space and Computation

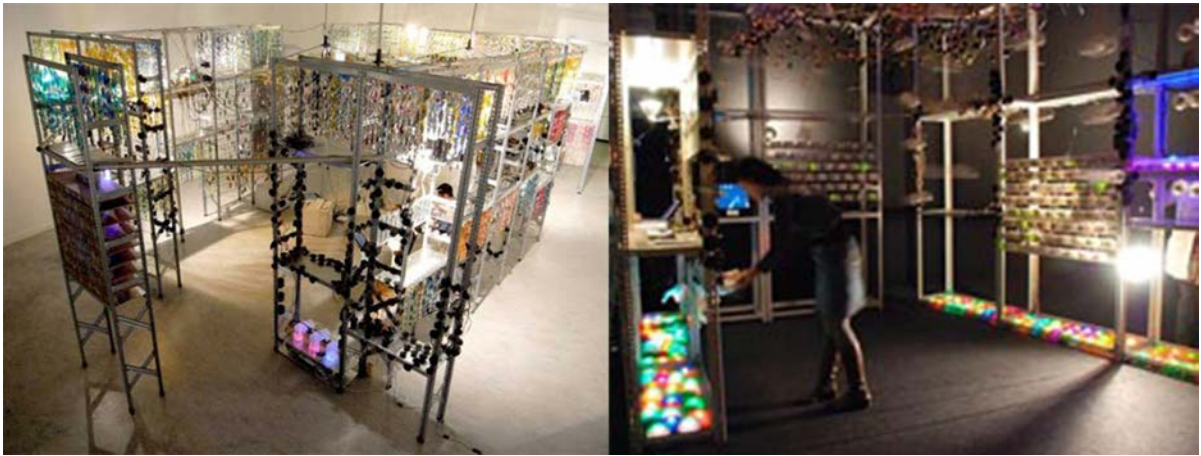


Figure A.6: Reconfigurable House built by Usman Haque and Adam Somlai-Fischer.

devices, such as low-tech toys and accessible everyday components, controlled by an Arduino open source microprocessor, which used the Processing open-source programming language. Users could interact either locally with the ambient environment, by changing the position of the different components, or remotely through the internet. The installation could be openly modified by its visitors: by rearranging the physical components, they could affect its software and thus its operation and behavior, its sound and light events, as well as noises and mist emission.

Context-aware Computing

Context-aware systems are systems that have the ability to sense and adapt to the environment. The context is not limited to spatial location such as user identity, the proximity of people and devices, but extended to a critical feature of sociologically-motivated exploration of interaction. As described in the book *Plans and Situated Actions*, Lucy Suchman drew on ethnomethodology, an analytic approach to the organization of social action, to provide a forceful critique of the then-dominant formal planning model in Artificial Intelligence. Suchman terms the “situated action” perspective, and stands in contrast to the traditional planning model in which the sequential organization of action is predetermined by an algorithmic exploration of the “search space” of goals and actions. Suchman does not reject the notion of “plans”; instead, she observes that plans, as prespecified formulations of future action, are merely one of a number of possible resources that people draw upon in answering the question, “what do I do next?”(Suchman 1987)

Appendix A. Appendix - Space and Computation

This perspective suggests a deeper role for context in interaction. It argues that the context in which actions take place is what allows people to find it meaningful. Context – the organizational and cultural context as much as the physical context – plays a critical role in shaping action, and also in providing people with the means to interpret and understand action.(Dourish 2001)

An overview of context-aware systems by Hong, Suh and Kim (Hong, Suh, and S.-J. Kim 2009) shows that context awareness involves acquiring, sensing or being aware of context, as well as adapting to it or using it. Correspondingly, according to Dey and Abowd (D. Abowd et al. 1998) context awareness falls into two categories, namely using context and adapting to context. To embed context into computational system, multiple middleware are designed to accomplish some particular objectives of individual applications. For example, Context Toolkit is one of the early research works in providing the framework to support developers by containing some necessary features and abstractions. It can be considered as the solution having the architecture based on loosely coupled major components that interact among each other. Three main abstractions are introduced including context widget for retrieving data from sensors, context interpreter for reasoning sensor data using different reasoning techniques, and context aggregator for creating higher level context.(Dey, G. D. Abowd, and Salber 2001) Co-BrA or Context Broker Architecture is an agent architecture called broker-centric architecture. It supports smart space application by providing knowledge sharing and context reasoning. CoBrA has component based architecture and uses Web Ontology Language (OWL) to define ontologies for context modeling and reasoning. A rule-based method is used to interpret and reason about context. A policy language and engine are used to control the sharing of user context. (H. Chen et al. 2004)

Context-aware computing is widely used in applications that can act between some degrees of rationally and personally with or without the intervention from the users. For example, in domestic environments, variety kinds of contexts such as motion, temperature, image, voice, video, etc. are used to cooperate with associated applications. Multi-sensors systems are employed for gathering contexts by accessing directly to the sensors or through the platforms which are designed to cooperate all sensors together. Together with IoT technology, the smart home applications can enhance the quality of life by introducing automated appliance control and assistive services.(Gubbi et al. 2013,Soliman et al. 2013) Context-aware applications can

Appendix A. Appendix - Space and Computation

be also deployed in education domain covering from the stand-alone application as automatic tutoring system to the web-based, mobile and ubiquitous learning environments. For example, the assessment of the learner's ability as commonly seen in the recommendation systems and personalized learning systems that can provide the appropriate feedback, recommendations or interventions to the individual learner. (M. Das et al. 2010, Berri, Benlamri, and Atif 2006)

Augmented Reality

AR encompasses user interfaces that allow for interaction with digital content embedded into the physical environment of users. For this purpose, AR interfaces superimpose digital information, such as 2D or 3D graphics on the user's view of the physical environment in real-time. Azuma

defined the main characteristics of AR interfaces as 1) the combination of virtual and physical elements, 2) being interactive in real-time and 3) being registered in 3D. (Azuma 1997b) First implementations appeared in the 1960s with Ivan Sutherland's "The sword of Damocles" (Sutherland 1968), first industrial applications with head-mounted display in the 1990s (Thomas and David 1992) as well as with handheld displays (W Fitzmaurice 1993). Today, mobile AR applications are used among others for browsing situated media with AR browsers (Langlotz et al. 2014), museums guides (Schmalstieg and Wagner 2007), mobile gaming (Grubert and Schmalstieg 2013), navigation tasks (Mulloni et al. 2009), product marketing as well as for industrial applications (Zollmann et al. 2014).

AR applications usually comprise three components: A tracking component, a rendering component, and an interaction component. All of these components can be considered as essential. The tracking component determines the device or user position in six degrees of freedom (DoF), which is required for visual registration of the digital content and its physical surroundings. Based on tracking data the scene is composed in the rendering component. Finally, the interaction component allows the user to interact with the physical and digital information.

Through experimenting place and creating objects into space, AR can build new ways for the 'physical world' (place, location, and space) to be understood and appropriated. However, the AR characteristics facilitate the transmutation of the physical spaces and real-time. In fact, it transforms 'space' to 'place' (JIN 2016). Similarly, sharing locational information, as a

Appendix A. Appendix - Space and Computation

crucial attribute of a place, influences ‘Placemaking’ and our understanding of places for both sides: those who share locational information and those who receive it (Özkul 2015).

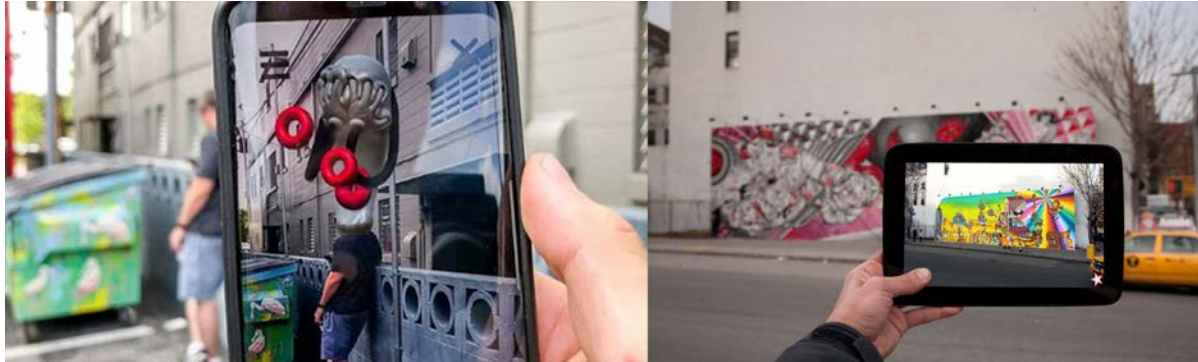


Figure A.7: Left, The ‘CITY UNSEEN’ project, Right, The Goldman projects

One of the key AR potentials is that its environments provide a perfect platform for Co-design, or Participatory design, or originally co-operative design which can have several alternative names, including co-located collaboration, remote collaboration, real-time collaboration, multiple-user collaboration, etc. (P. Wang et al. 2019). Hybrid spaces, which are comprised of both physical and virtual space (Harrison and Dourish 1996b), bring different forms of citizen engagement (Foth et al. 2011), and also novel ways of experiencing and navigating within cities (Mizuko Ito 2005). For example, The Goldman projects perfectly demonstrate how tactical placemaking and AR could be merged to present the brilliant historical part of street art. Keith Allen Haring, an American artist whose pop art and graffiti-like works grew out of the New York City street culture of the 80s (Sussman 1997), painted a mural on Houston and Bowery Street’s wall in 1982. Unfortunately, the original mural was covered by several ripped ads papers and tags. In 2008, however, the wall was restored by Tony Goldman, wall’s owner, and a group of public art experts and artists like curator Jeffrey Deitch. In June 2012, for mixing AR into the project, Heavy + PAC was utilizing AR to create a virtual history of the famous mural site (Figure A.7). So, the users can take a virtual journey of the wall and view each section as it originally resembled. Hence, it can be argued that this project is a popular re-engagement by transmitting historical memories of urban art through AR solutions. The ‘CITY UNSEEN’ is an ongoing public art project which removes artistic boundaries by launching a city into AR experiences, through fusions of works by internationally prominent artists and advanced AR technology. The contemporary art organization, Snap! Orlando, has

Appendix A. Appendix - Space and Computation

been launched the City Unseen project since 2018. This project covers a wide range of experiences and participation. The teleport brings the viewer into a virtual tour into an artist's studio, Projection mapping, and murals are eventually coming to life. (Figure A.7). This project is also used to rebuild and renovate some damaged and dilapidated murals (Gomez 2019) –the project launched by the CITY UNSEEN application on Android smartphones and Apple iOS devices. The application presents a map of Orlando city, which was tagged by the locations of art projects. The application presents a map of Orlando city, which was tagged by the locations of art projects. It can be said that this project brings a cutting-edge creative placemaking approach for Orlando's citizens.

Through the reflection on the history of the relationship between space and computation, I attempt to, on the one hand, sketch a detailed technical culture within the context of architecture, on the other, the story of space and computation is a part of larger narrative, from a historiographical and socioeconomic perspective. More importantly, if we look at the history of the close affinity between computation and space, we can find that a tendency of dissociation or division. As Mumford noted, “rested on the ability to dissociate lifting power from the arm and create a crane; to dissociate work from the action of men and animals and create the water-mill; to dissociate light from the combustion of wood and oil and create electric lamp”, similarly, the world of computation and the physical environment are separate with clear boundaries, the world of the concrete and the territory of abstractions. The world of machine tends to render the world of organism mechanic. The dualist notion can result in the understanding of computation as abstract, disembodied and placeless, which may further leads to reductionist, technology-determinism and misassumption of symmetry in human-machine interaction.

However, a range of novel technologies are fusing the physical, digital and biological worlds and impacting disciplines, economies and industries. We can see the virtual world in the physical world tends to recombine space, time and humans within a unified, networked and inseparable amalgamation. Conceptually, I am arguing that we need to consider the future computation in the built environment as an augmentation of space, or a spatially augmented interface pointing to the mutual constitution of human and things, where, artifacts, as mediators of human-world relations, actively co-shape the way humans act in and perceive the world. Similarly to Heidegger's early phenomenological insights discussed in the previous section,

Appendix A. Appendix - Space and Computation

spatially augmented interface invites engagement with their material function, rather than their symbolic and iconic representations in the form of entities and events. Thus, in alternative to detaching interface, we need to inhabit interface and provide meaningful experience of inhabitation. Technically, spatially augmented interface is a continuous and pervasive user interface that augments the physical world with digital information registered in 3D, while being aware of and responsive to the users context. Spatially augmented interface can be seen as a context-aware and responsive, continuous augmented experience within a physical environment, different from conventional augmentation technologies that are sporadically, ad-hoc and not in a continuous mode. Spatially augmented interface can potentially change our contemporary design and engineering practices.

For architectural design, spatially augmented interface can potentially challenge the traditional framework of design-construction-operation in placemaking and the role of architects as figures of authority. Rather to create a physical environment with predetermined function, we should aim at providing a protocol or interface for interaction and creating an open-ended environment guiding itself through the contingencies imposed on the system. Agency of space is shared among designers who create infrastructure, users with intentions and computational systems with “programs of actions”. Users can use mobile devices equipped with tangible, spatial and context-aware interface to define an open-ended, highly personalized environment without physical constraints and creatively engage with the functions of the ambient environment. From the viewpoint of computation, physical space is not indifferent to its embedded system, digital world is neither abstract, disembodied or placeless. Physical environments can serve as a ground that defines the way digital entities are distributed and function. Space, or more precisely termed as place, pertains to concrete and situated activities can directly guide the functioning of computational systems. Medium such as user-friendly interfaces and middleware will be able to manage and represent heterogeneous information in a common terms and provide exchange and interpretation of messages accordingly. In this sense, computation shifts from a means of division to a medium for recombination. Instead of turning our organic living environments into machines, we can envision a new architecture machine that is more organic and embodied, a dichotomy between the virtual and physical world will merge into a hybrid continuum unifying the concrete object and its representation.