Sapienza University of Rome

Ph.D. program in Computer Engineering

XXIII Cycle - 2011

Improving Human-Robot Awareness through Semantic-driven Tangible Interaction

Gabriele Randelli

SAPIENZA UNIVERSITY OF ROME

PH.D. PROGRAM IN COMPUTER ENGINEERING

XXIII CYCLE - 2011

Gabriele Randelli

Improving Human-Robot Awareness through Semantic-driven Tangible Interaction

Thesis Committee

Prof. Daniele Nardi (Advisor)
Prof. Marco Schaerf

Reviewers

Prof. Alessandro Saffiotti
Prof. Matthias Scheutz

Author's address:

Gabriele Randelli
Dipartimento di Informatica e Sistemistica
Sapienza Università di Roma
Via Ariosto 25, 00185 Roma, Italy
e-mail: randelli@dis.uniroma1.it
www: http://www.dis.uniroma1.it/∼randelli/

Contents

Chapter 1

Introduction

It is a well settled prediction that robots might be the next worldwide break-through, after the computer revolution in 1990s that gave rise to the *Information Age*. This passes through the achievement, by the robotic industry, of a critical consumer mass fostered by the massive deployment of robotic systems in our world. Robots for domestic services, health care, entertainment, work, or field applications, are just few examples of the manifold potential applications. All these scenarios share a common aspect: the interaction between humans and robots.

Effective human-robot interaction relies on attaining an appropriate level of awareness, in terms of the reciprocal understanding of the status of both the involved entities (humans and robots), their activities, environment and, when needed, their mission. From the human perspective, this is referred as *human-robot awareness*, described by Drury et al. (2003) as: *"the understanding that the humans have of the locations, identities, activities, status and surroundings of the robots"*. Conversely, from the robot perspective, this is typically denoted as *situation awareness*, defined by Endsley (1995) as: *"the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"*. In both cases, a lack of awareness significantly lowers the degree of human-robot interaction, which in turn weakens the overall task performance. Unfortunately, despite its relevance, this aspect is still an open research issue, since it involves non trivial knowledge abstraction processes.

From the standpoint of human-robot awareness, a relevant aspect that must be considered is that every interaction is realized via a communication mean. In particular, humans and robots communicate through a robot interface, which is often the Achilles' heel for a valuable awareness. Usually, robot

1

interfaces are composed of two distinct components: *(i) control* (or input) device, and *(ii) view* (or output) representation. As for the former aspect, the robotic community has often adopted common input devices, such as: keyboard, mouse, or joystick. While a common approach for human-computer interaction, these devices are inadequate when dealing with robots. First, robots are complex systems that exhibit a high degree of freedom with respect to the manipulation degree of common input interfaces. Second, robots are deployed, move and act in a real world, which is by far more complex and unpredictable than a virtual environment or a desktop representation, as in human-computer interaction. This requires comfortable interaction means, which guarantee fast response by leveraging unconscious and innate human skills. Finally, robots and humans are both situated in a real environment, where robots are typically perceived as active entities, more related to humans than to computers. Thereby, there is an instinctive expectation to communicate with them by everyday interaction means. When these aspects are not properly taken into account, robot control becomes unnatural. This burdens the operator cognitive effort with robot control, causing a degradation of the situation assessment process. The interaction is then characterized by a low-level, continuous and manual control, where the robot is not perceived as a standalone entity, rather as another type of computer. The result is that the operator has a reduced, if none at all, capability to understand the scenario and improve the overall human-robot awareness.

Concerning the view component of robot interfaces, the robotic community has, at least at the very beginning, designed interfaces as operation and debugging tools for robotic experts, that is, conceived as containers of huge amount of data, mostly represented in numerical form. Needless to say, this does not guarantee a proper awareness, as requires a significant effort to gather a high-level interpretation from numerical data. Such an approach cannot be applied to large scale consumers without any robotic expertise. Research has faced this problem with two different approaches: information visualization and novel view representations. Information visualization identifies design best practices for effective interface layouts, where data are organized in order to reduce the operator's cognitive effort. In particular, the trend is to collect data in a single window, to avoid the operator's continuous context switching among different parts of the interface. Concerning view representations, robot interfaces have been supported by novel visualization paradigms, such as: 3D viewers, virtual reality, immersive camera feedback, and so on, which enhance the perception capabilities of human operators. Despite this, neither information visualization nor enhanced visualizations are a definitive solution for non expert users to the human-robot awareness problem, since the assessment process is still demanded to the human operator. This requires to the operator a significant effort to collect data, still represented in numerical form,

aggregate them and extract a high-level interpretation of the scene. However, it is a concrete expectation for humans to be relieved by the robotic system in this process, for a faster human-robot awareness acquisition.

On the other hand, the latter aspect moves the problem to the robotic perspective: situation awareness. Situation awareness goes beyond low-level perceptions, as it relies on the achievement of a high-level knowledge meaning comprehension, which requires a grounding process. Nevertheless, robotic systems have still limited capabilities in symbol grounding. Furthermore, robots face the complexity of the real world where they are deployed. Perception errors, grounding ambiguities, knowledge inconsistencies and unexpected events are the most challenging factors that affect a proper situation assessment. The robotic community has dealt with these issues with several solutions: boosting the robot perception capabilities, adopting data-fusion criteria in order to solve ambiguities, designing methods to enhance robustness to unexpected conditions, and recovery procedures to deal with knowledge inconsistencies. All these approaches cope with the situation awareness problem only from the robot point of view. However, most of these issues are quite common for humans, which exhibit innate cognitive skills in grounding, and are used to manage explicit representations. Despite this, the role of humans has never been quite considered by the robotic community for this challenge, for example, through the design of *mixed* human-robot initiatives. Even when considered, the human contribution to high-level knowledge acquisition is typically constrained by computer-centered acquisition techniques, which is unnatural for humans. Considering that humans are situated in the real environment, and in that environment they effectively perceive relevant information and acquire it, the robotic community still needs to investigate novel approaches that may bring advantages to this process, and that can exploit the human natural attitude for grounding concepts in elements of the real world.

The main motivation of this thesis is the need to re-consider all the aforementioned research challenges, in terms of awareness in human-robot interaction, in light of novel technologies that are nowadays available and widespread, such as, portable tangible user interfaces. In fact, these devices exhibit interesting characteristics that are still unexplored in the robotic community. Concerning human-robot awareness, tangible user interfaces remove the sharp separation between input and output functionalities in robot interfaces, by enabling a direct interaction between humans and robots within real environments. As for situation awareness, these devices target conventional input controllers under a completely different light. Input devices are not only considered for robot operation (in the sense of a human operator controlling the robot locomotion), rather their role is augmented with novel functionalities useful for other activities, as in the case of human-centered knowledge grounding.

1.1   Scope

Human-robot interaction is a wide research area and it is applied to several contexts. In order to bind the scope of this thesis, we provide the following assumptions:

- we consider a subset of the possible scenarios, which is limited to **field applications**, that is, the use of mobile robots in environments, such as work sites or natural terrain, where the robots must safeguard themselves while performing non-repetitive tasks and objective sensing in dynamic environments; some examples are urban search and rescue and surveillance;

- among all the possible tasks in the considered applications, we focus on one specific activity: **robot operation**, that is, the human's control of a mobile robot, which is a fundamental activity orthogonal to any robotic application; on the other hand, we are not addressing interaction between humans and autonomous robots, nor the control of robot manipulators;

- anticipating one of the contributions of this thesis, we only deal with **portable tangible user interfaces**, hence not considering other types of tangible devices, such as interactive surfaces.

1.2   Contributions

The contributions of this thesis are based on a novel approach to the problem of awareness in human-robot interaction, that is, **improving awareness by enabling innovative interaction means**. A detailed description is reported in the following:

- concerning the achievement of human-robot awareness through effective input components, our solution is to design **tangible-based algorithms for robot operation**, since they embody high-level interaction paradigms that are comfortable for robot control. Lowering the cognitive effort for robot operation, tangible interfaces shift the operator engagement on the achievement of effective human-robot awareness, through an active involvement in the assessment process. We follow a research path that investigates from low-level to high-level tangible interaction paradigms: **grasping, tangible feedback, gesturing, pointing**. Each of them is in turn adopted as physical representation for the implementation of novel robot operation algorithms, which range from manual teleoperation to shared control;

- once relieved from the robot control, the human operator can be involved in the assessment process. We aim at defining a novel **assessment methodology** to boost up awareness in robotic systems. While previous research has mainly investigated how to improve the grounding capabilities of autonomous robots, our vision is to **put humans in the middle of the acquisition process**. Operators support robots by exploiting their innate skills in grounding and dealing with symbolic representations. Such a methodology satisfies the following requirements:

  - low impact on the operator cognitive effort;
  - acquisition leverages innate humans' skills performed in everyday activities;
  - acquisition is situated in the real world.

  Our solution is to define a twofold role for the input component of robot interfaces. From the one hand, as already mentioned, lowering the operator's effort in robot operation; from the other hand, supporting humans in the knowledge acquisition process, acting as **semantic-driven tangible interfaces**. TUIs exhibit novel interaction metaphors that humans adopt with little effort to acquire knowledge, such as gesturing or pointing. Moreover, tangible interfaces enable the **operator mobility** within the environment, which allows for fast and smart knowledge acquisition by a direct interaction with the elements in the scenario;

- once acquired, knowledge should be effectively represented to enhance awareness. Our solution is to adopt explicit formalisms to express the acquired information in terms of **semantic knowledge**. On the one hand, semantic knowledge is exploited by representation and reasoning systems to evince further knowledge, which improves robot autonomous skills and robustness. On the other hand, the same knowledge supports the human operator with a twofold benefit: semantic knowledge is based on symbolic representations, which are more comfortable for humans, and it is the result of a meaning extraction process which increases awareness.

1.3 Thesis Outline

This thesis is divided into four parts. Part I defines the research problem, introduces useful preliminaries, and provides a state of the art of the relevant topics covered by this work. Part II describes our contribution to robot operation through tangible user interfaces. In Part III we deal with tangible interfaces to enhance the assessment process. Finally, Part IV introduces

our last contribution, achieving an effective representation and management of the acquired knowledge through explicit representations, and exploiting it to enhance robot performance.

Part I

Chapter 2 formalizes the problem of awareness in human-robot interaction through two definitions: *human-robot awareness* and *situation awareness*. Furthermore, we introduce common metrics to estimate the level of awareness, which will be useful throughout the experimental evaluations presented in this thesis. Then, preliminaries are provided about tangible user interfaces. A formal analysis, based on three conceptual frameworks, allows us to highlight their main characteristics, which represents a common background retained throughout this thesis. Finally, since the scope of our thesis is to deal with robot operation, we introduce a formal definition of this activity, provide basic terminology, and a taxonomy of control styles.

Chapter 3 provides the state of the art about three relevant topics. First, common approaches to the problem of human-robot awareness are sketched out, to highlight their main limitations and provide a comparison with respect to our solution. Second, past implementations of TUIs in robotics are described, according to their interaction paradigms. Finally, since we deal with semantic representations, a survey of semantic knowledge in robotic systems is provided.

Part II

Following an increasing research path, Chapter 4 addresses low-level tangible interaction paradigms. More in depth, we present an interface for robot teleoperation, based on motion sensing, and we describe a tangible feedback system that leverages environment background information for safe teleoperation. Finally, we investigate intra-scenario operator mobility.

Chapter 5 moves the discussion towards high-level interaction paradigms. In particular, we deal with gesturing, and we present a gesture recognition system for a robot shared control policy. Such a system further lowers the operator cognitive effort by triggering commands and demanding their execution to the robot.

Part III

Chapter 6 presents another high-level tangible interaction paradigm, pointing, by defining the concept of *tangible pointing interface*, for the selection of environment elements on the ground. This also represents a first step towards the definition of a knowledge acquisition process.

Chapter 7 introduces the definition of *semantic-driven tangible interface*, and present a novel assessment methodology. It investigates how a tangible interface can be adopted to acquire knowledge and represent it through semantic formalisms. We provide a procedure for knowledge acquisition and implement a multimodal interface based on TUIs, speech and vision.

Part IV

Chapter 8 deals with the last contribution of this thesis: *semantic knowledge*. We present a *context-based architecture* that represents and manages semantic knowledge without hand-coding it into specific robotic modules, nor requiring a massive re-engineering of existent robotic systems. Our solution is then adopted to validate the effectiveness of semantic knowledge for situation and human-robot awareness.

Chapter 9 reports the thesis conclusions and future work.

1.4 Publications

Part of this thesis has been published in the following journal articles, conference and workshop proceedings.

Part II

Chapter 4: G. Randelli and M. Venanzi and D. Nardi. Tangible Interfaces for Robot Teleoperation. In *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, pages 231-232, Lausanne, 2011. [Randelli et al. (2011b)]

Chapter 4: G. Randelli and M. Venanzi and D. Nardi. Evaluating Tangible Paradigms for Ground Robot Teleoperation. In *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, Atlanta, GA, USA, 2011 (to appear). [Randelli et al. (2011c)]

Part III

Chapter 6: L. Marchetti, G. Randelli, and F. A. Marino. Multi-agent be-
haviour composition through adaptable software architectures and tan-
gible interfaces. In *RoboCup 2010: Robot Soccer World Cup XIV*, pages
278-290, Springer, 2010. [Randelli et al. (2011a)]

Chapter 7: L. Carlucci Aiello and D. Nardi and G. Randelli and C.M. Scalzo.
Suppose you have a robot. College Publications, 2011 (to appear).
[Aiello et al. (2011)]

Part IV

Chapter 8: G. Randelli and D. Nardi. Introducing ontology best practices
and design patterns into robotics: Usarenv. In *Proceedings of the 2010
conference on Modular Ontologies*, pages 67-80, IOS Press, The Nether-
lands, 2010. [Randelli and Nardi (2010)]

Chapter 8: D. Calisi, L. Iocchi, D. Nardi, G. Randelli, and V. Ziparo. Im-
proving search and rescue using contextual information. *Advanced Robot-
ics*, 23(9):1199-1216, 2009. [Calisi et al. (2009)]

Furthermore, Chapter 3 reports our contributions to the problem of human-
robot awareness through information visualization best practices and enhanced
view representations. This set of publications is tightly related to the motiva-
tions of this thesis, as it represents the starting point of our research investi-
gation about awareness in human-robot interaction:

- A. Valero and G. Randelli and F. Botta and M. Hernando and D.
  Rodriguez-Losada. Operator Performance in Exploration Robotics. In
  Journal of Intelligent & Robotic Systems, pages 1-21, Springer Nether-
  lands, 2011. [Valero et al. (2011)]

- A. Valero, C. Saracini, F. Botta, and G. Randelli. Spatial processes in
  mobile robot teleoperation. *Cognitive Processing*, 10(2):338-341, 2009.
  [Valero et al. (2009c)]

- A. Valero, G. Randelli, C. Saracini, F. Botta, and D. Nardi. Give me the
  control, i can see the robot! In *Proc. of IEEE Int. Workshop on Safety,
  Security, and Rescue Robotics (SSRR)*, 2009. [Valero et al. (2009b)]

- A. Valero, G. Randelli, C. Saracini, F. Botta, and M. Mecella. The advantage of mobility: Mobile tele-operation for mobile robots. In *Proceedings of the AISB 2009 Convention*, 2009. [Valero et al. (2009a)]

Part I

Preliminaries

Chapter 2

Background

2.1 Introduction

In this chapter we provide the reader with a theoretical background concerning our research problem and the main contributions of this thesis. Our aim is at introducing a common terminology and knowledge to support our research investigation throughout this work. On the other hand, a state of the art of relevant topics is presented in Chapter 3.

First, in Section 2.2 we formally introduce the research problem of this thesis, that is, awareness in human-robot interaction. More in depth, we present both the perspectives related to this topic: human-robot awareness and situation awareness. This points out the main research challenges.

Second, we move towards our approach to deal with the awareness problem: tangible user interfaces. Due to the novelty of these devices, we extensively overview them in Section 2.3. In particular, we are interested in evincing the relevant characteristics of these interfaces, to compare them with respect to traditional interfaces, and to identify which aspects of TUIs may effectively improve awareness.

As already mentioned in Section 1.1, the scope of our thesis is limited to robot operation. Robot operation is composed of a wide spectrum of approaches, whose exhibit different characteristics. Thereby, in Section 2.4 we overview these approaches, from low-level teleoperation to high-level operation paradigms. It is important to understand the main differences among these approaches, since most of them are considered in the rest of this work.

2.2   HRI and Situation Awareness

Awareness is quite an abstract concept. It is defined as: *"the state or ability to perceive, to feel, or to be conscious of events, objects or sensory patterns"*. Awareness is a relevant aspect, yet differently defined, in manifold fields: biology, neurosciences, computer science and, recently, robotics. Furthermore, there are different types of awareness: *self-awareness*, focused on our proprioceptive status, or *external awareness*, referred to the environment where we are situated, or its external elements (e.g. people, objects, situations, events). In this thesis we mainly deal with the latter. Finally, awareness should be contextualized according to the field we are considering. In our case, as already mentioned, we focus on field applications. Closely related to awareness is the concept of **assessment**, that is *the process that achieves the state of awareness.*

Throughout this thesis, we adopt the following terminology when referring to awareness: *human-robot interaction awareness* to identify the overall problem in HRI, *human-robot awareness* when addressing the human aspect, and *situation awareness* from the robot perspective. A state of the art of common robotic approaches to this problem is provided in Section 3.2, while in the rest of this thesis we propose a novel solution based on tangible user interfaces and semantic representations.

2.2.1   HRI Awareness

Our investigation about awareness in human-robot interactions moves from the analysis of the same concept in the human-computer interaction community. More in depth, we address *computer-supported cooperative work* (CSCW), since this field shares several characteristics with robotics. According to Wilson (1991), *"CSCW* is a generic term, which combines the understanding of the way people work in groups with the enabling technologies of computer networking, and associated hardware, software, services and techniques". Therefore, CSCW deals with both the technological challenges to support cooperative work among manifold users, and the psychological and cognitive processes to enhance it. When people cooperate, in particular in remote workgroups, awareness becomes a crucial aspect, since individuals need to gain some level of shared knowledge about each other's activities. Drury (2002) provides a formal definition of awareness in CSCW.

Definition 1 (Awareness)

Given two participants p_1 and p_2 who are collaborating via a synchronous collaborative application, awareness is the understanding that p_1 has of the identity and activities

of p_2. ∎

A question then arises: *is it possible to apply such a definition to human-robot interaction?* One the one hand, the answer is yes, since humans and robots cooperate for a joint goal, hence they need to understand each other. On the other hand, at least two aspects characterize the human-robot interaction setting as unique:

1. CSCW, and more in general human-computer interaction, is based on the interaction between humans and computers, the latter being passive entities; conversely, robots are perceived by humans as active entities, hence their active involvement is expected;

2. CSCW is based on the symmetrical interaction among humans, that is among entities with the same cognitive skills; human-robot interaction is asymmetrical and unbalanced, as robots do not have the same cognitive skills of humans.

Based on these considerations, a different characterization of awareness is needed in HRI. The first formal definition of **HRI awareness** is due to Drury et al. (2003):

Definition 2 (HRI Awareness - Base Case)

Given one human and one robot working on a task together, HRI awareness is the understanding that the human has of the location, activities, status, and surroundings of the robot; and the knowledge that the robot has of the human's commands necessary to direct its activities and the constraints under which it must operate. ∎

The same authors further extend this definition to the general case of multiple humans and robots.

Definition 3

Given n humans and m robots working together on a synchronous task, HRI awareness consists of five components:

1. **human-robot:** the understanding that the humans have of the locations, identities, activities, status and surroundings of the robots. Further, the understanding of the certainty with which humans know the aforementioned information.

2. **human-human:** the understanding that the humans have of the locations, identities and activities of their fellow human collaborators.

3. **robot-human:** the robots' knowledge of the humans' commands needed to direct activities and any human-delineated constraints that may require command noncompliance or a modified course of action.

4. **robot-robot:** the knowledge that the robots have of the commands given to them, if any, by other robots, the tactical plans of the other robots, and the robot-to-robot coordination necessary to dynamically reallocate tasks among robots if necessary.

5. **humans' overall mission awareness:** the humans' understanding of the overall goals of the joint human-robot activities and the measurement of the moment-by-moment progress obtained against the goals. ∎

In particular, throughout this thesis, we will focus on **human-robot awareness**, which has been considered by Drury et al. (2003) a key issue when operating robots.

HRI awareness affects the effectiveness of human-robot interaction, hence it is a crucial aspect to consider when designing a robotic system that involves the human presence. The lack of HRI awareness is defined as **HRI awareness violation**. The HRI community effort has focused on the identification of design best practices and guidelines, in order to boost up the assessment process. The most common approaches in this direction are described in Section 3.2. On the other hand, our contribution with respect to this aspect is mainly presented in Part II of this thesis.

2.2.2   Situation Awareness

Situation awareness (SA) is a concept widely spread in different applications: avionics, air traffic control, power plant operations, emergency activities, military strategic systems, all those scenarios characterized as intrinsically challenging. Despite nowadays manifold definitions exist, the most accepted conceptualization of situation awareness has been provided by Endsley (1995).

Definition 4 (Situation Awareness)

The perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future. ∎

Roughly speaking, acquiring an effective situation awareness is not only a matter of understanding what is happening around the person, but also to interpret this information according to the desired goals, and to project this information in the future. It is worth noting two aspects of this definition. First, it is intentionally a generic formalization, wide enough to be applied to manifold fields, and this has fostered several different definitions of situation awareness, either specialized to specific disciplines, or focused on different
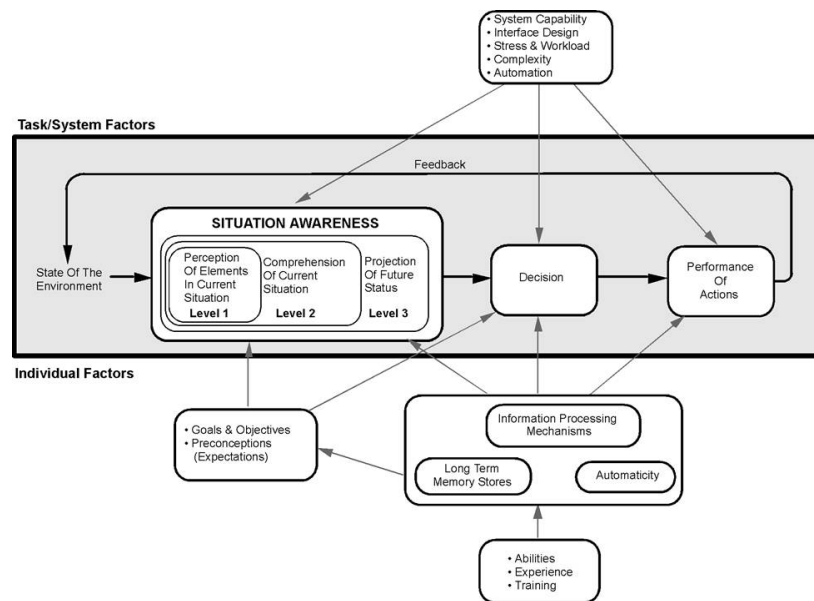
Figure 2.1: The situation awareness model proposed by Endsley [Adapted from Endsley (1995)].

points of view. Second, the definition of HRI awareness provided in the previous section can be considered as a specialization to robotics of the Endsley's definition.

Endsley also differentiates situation awareness, as a state of knowledge, from **situation assessment**, that is the processes involved to gain SA. The theoretical model of SA, as well as the three steps involved in SA assessment, are illustrated in Figure 2.1 and described below:

1. **Level 1 (Perception of the elements in the environment):** corresponds to the perception of the status, attributes, dynamics and relevant elements in the environment;

2. **Level 2 (Comprehension of the current situation):** the elements perceived at level 1 are aggregated in a comprehensive understanding of their overall significance, which in turn is matched to the operator goals. Two processes are involved in this step: the aggregation of isolated elements in a global context, and a high-level knowledge extraction from this integrated pattern;

3. **Level 3 (Projection of future status):** once aware of the status and dynamics of the single elements (level 1), and acquired the overall

situation (level 2), it is possible to project the future actions of the elements in the environment.

Our contribution to situation assessment is presented in Part III.

2.2.3   Measuring awareness in HRI

In order to evaluate whether a robotic system provides effective awareness, it is important to provide quantitative metrics to measure it. According to Hjelmfelt and Pokrant (1998), evaluation techniques can be roughly divided into three classes:

1. **explicit** techniques interrupt a participant to ask her questions about the situation and estimate SA from her answer; this approach has two main drawbacks: interrupting the operator degrades her SA, and participants tend to learn the type of questions;

2. **implicit** techniques never interrupt the user, and assess how well a task is accomplished, but the drawback is that this does not always imply to gain an effective SA;

3. **subjective** techniques ask subjects to self-rate their SA, but may be unreliable.

In this thesis all the experimental evaluations adopt implicit evaluation, by logging relevant metrics about the performance of participants. Furthermore, in Chapter 4 we also consider explicit techniques, through *think aloud*, and subjective techniques, providing post-run questionnaires.

Scholtz (2002) identifies some evaluation issues that have been successively adapted to robotics by Drury et al. (2004), and applied to urban search and rescue systems (USAR):

- *Is sufficient status and robot location information available so that the operator knows the robot is operating correctly and avoiding obstacles?*

- *Is the information coming from the robots presented in a manner that minimizes operator memory load, including the amount of information fusion that needs to be performed in the operators' heads?*

- *Are the means of interaction provided by the interface efficient and effective for the human and the robot (e.g., are shortcuts provided for the human)?*

- *Does the interface support the operator directing the actions of more than one robot simultaneously?*

- *Will the interface design allow for adding more sensors and more autonomy?*

It is interesting to note how common approaches to human-robot awareness, as we discuss in Section 3.2, address all these questions but the third, which is our concern in this thesis.

Different evaluation guidelines have been proposed to deal with these questions.

Information presence and presentation. This issue is closely related to achieving an effective situation awareness. It is possible to adopt a methodology known as *Situation Awareness Global Assessment Technique* (SAGAT), designed by Endsley (1988). It consists of goal-directed task decomposition, to evince a set of situational awareness requirements. A set of queries is then constructed to determine the operator's situational awareness step-by-step.

Interaction performance. This aspect relates to all types of interactions: human-human, human-robot, robot-human, and robot-robot. This can be accomplished through common HCI techniques, which typically are based on the definition and execution of a set of evaluation tasks. Adopted metrics will be in general different whether we are evaluating humans or robots.

Support for scalability and operator roles. Both these aspects can be analyzed using the aforementioned information presence and interaction performance techniques.

Olsen and Goodrich (2003) propose six performance metrics to evaluate human-robot interactions:

1. *task effectiveness*, that is how well a human-robot team accomplishes a task;

2. *neglect tolerance*, how much the robot's current task effectiveness declines over time when the operator is not controlling that robot;

3. *robot attention demand*, which measures the fraction of total time the operator controls a specific robot;

4. *free time*, defined as the converse of the robot attention demand;

5. *fan out*, the number of robots that a single operator can effectively control concurrently;

6. *interaction effort*, the time spent in interacting and its relative cognitive effort.

These metrics can be related to measuring HRI awareness, even if not specifically focused on this issue.

Finally, Drury et al. (2007) address the problem of evaluating situational awareness through a novel technique: *LASSO* (*L*ocation, *A*ctivities, *S*urroundings, *S*tatus, and *O*verall mission). This technique analyzes the users' recordings during "think aloud" intervals, meanwhile performing their task. Recorded utterances are classified with respect to the five categories, and estimated as positive or negative awareness. Under this point of view, it could be considered as a subjective methodology, but it is not performed at the end of the experiment, as most of the subjective techniques, rather on a moment-by-moment basis.

2.3   Tangible User Interfaces

Gesturing is a human innate communication skill in the real physical world where we live. Still, this skill is not exploited in digital systems, where most of the interaction is confined to graphical user interfaces and to ad-hoc interaction metaphors, known as *WIMP*, which stands for *Window, Icon, Menu, Pointing device.* Coupling gesturing with the digital world then becomes a need. A *tangible user interface* (TUI) is a user interface in which a person interacts with digital information through the physical environment. Unlike GUIs, tangible user interfaces use physical forms that fit seamlessly into the user's physical environment. As stated by Ishii (2008), "*TUIs make digital information directly manipulatable with our hands and perceptible through our peripheral senses through its physical embodiment*". In this section we provide an extensive background about TUIs and their characteristics (Sections 2.3.1, 2.3.2 and 2.3.3). We outline the main influences behind them, and we distinguish TUIs with respect to other tangible paradigms (Section 2.3.4). We relate tangible user interfaces to relevant cognitive science and human-computer interaction studies (Sections 2.3.5 and 2.3.7), which will be useful for the reader to evince how such systems could be adopted in robotics. Needless to say, since tangible interaction is a huge field, we selected those aspects that are relevant to the rest of this thesis. Concerning the use of tangible user interfaces in robotics, the reader may refer to the state of the art provided in Section 3.3. We deliberatively do not discuss in this section the different interaction metaphors exhibited by TUIs in robotics, since this part is a main issue of this thesis and will be discussed later.

2.3.1 From Ubiquitous Computing to TUIs

Every discovery is influenced by some preliminary works that partially unveil its breakthrough. Concerning tangible user interfaces, there are three main contributions worth mentioning.

History can be started in 1991, when Mark (1991) presents his theory of **ubiquitous computing**. Mark imagined the 21st century as an era where computer services would be so pervasive that no user would even notice them. Fully merged and deployed within the real environment, hundreds of computing processes, linked through wireless connections, working in the background without forcing humans to interact with them through classical computer interfaces. The interesting aspect of such an idea, which will be fruitful for the development of TUIs, is that digital systems are moved back into the real world, and are considered as background processes. In fact, traditional interfaces force the user to behave as the computer expects, not how we are used to do everyday.

Another influence for the development of TUIs has been **augmented reality** (AR). According to Wellner et al. (1993), augmented reality is a visual overlay of digital information onto a real-world imagery, hence it represents a way to couple digital and real world. Tangible interfaces are closely related to this coupling motif, but are more focused on exploiting graspable physical objects, rather than on vision techniques.

Finally, worth mentioning is the work by Fitzmaurice et al. (1995), who defined the term **graspable user interface** (which will be later replaced by TUI). A graspable user interface consists of virtual objects controlled through physical graspable handles called *bricks*. A brick can be attached to different virtual objects, and allows for space-multiplexed input and output. A Brick can be manipulated to select, move or rotate its corresponding virtual object. Groups of bricks can be linked to a single virtual object for complex operations: scaling, shape transformations, and so on.

Based on these influences, Ishii and Ullmer (1997) coined the term **tangible user interface**, defined as *a system that gives physical form to digital information, employing physical artifacts both as representations and controls for computational media*. Digital content is then represented through tangible objects, and it is manipulated via physical interaction with these objects. The core idea is literally to allow users to grasp data with their hands. In this section we will introduce three different theoretical models to formally characterize TUIs: the *(i)* **model-control-representation (physical and digital) framework** (Section 2.3.2), *(ii)* **foreground and background processes** (Section 2.3.3), and the *(iii)* **tangible interaction framework** (Section 2.3.4). The characteristics evinced by these theoretical models will be applied to the robotic context later on.

Graphical interfaces make a fundamental distinction between "input devices", such as keyboard and mouse, for control, and "output devices", like monitors, for representing information. On the other hand, TUIs eliminate this distinction, which motivates our interest in these devices for robotic applications. For example, a mouse exhibits control significance but little representational significance, that is, its associated graphic functionality (the cursor on the screen) could be realized by a trackball, joystick, or any other input device. When dealing with tangible interfaces, the physical shape is closely linked with the digital information they represent. These two aspects, control and representation, are essential to understand the difference between TUIs and other interfaces.

2.3.2   The MCRpd Interaction Framework

Starting from Ishii's definition, we aim at a formal assessment of the main characteristics of tangible user interfaces. Ullmer and Ishii (2000) introduce an interaction conceptual framework to characterize TUIs, known as "**model-control-representation (physical and digital)**" interaction framework (MCRpd). This model is based on the popular "model-view-controller" (MVC), defined for GUIs in combination with the Smalltalk-80 programming language.

As already mentioned, representation and control are two relevant aspects of tangible interfaces. In particular, MCRpd is based on the concept of *representation* as the *external* manifestation of information, that is, *perceivable by human sensing*. The MCRpd model assumes that external representations are divided into two classes: **physical representations**, physically embodied in a tangible form, and **digital representations**, which are computationally mediated by displays. As for the latter, even if manifested in the real world (for example, through a display or speakers), they are not physically embodied. *The unique trade-off between these two representations is the leading characteristic of TUIs with respect to conventional interfaces.* Needless to say, we consider this aspect as relevant for our investigation as well, since represents a different type of interaction with robots. As reported in Figure 2.2(a), the MVC model highlights the strong separation between the digital representation (**view**) and the **control** functionality. On the other hand, in the MCRpd model (see Figure 2.2(b)) the view element has been divided into *physical representation* ("rep-p") and *digital representation* ("rep-d"). With respect to the separation between view and control components, the MCRpd reveals a strong synergy between control and physical representation. Furthermore, the model highlights four characteristics of tangible interfaces:

1. physical representations (rep-p) are computationally coupled to digital information (model); for example, object shapes may represent digital

(a) Model-View-Controller Model

(b) Model-Control-Representation (physical and digital) Model
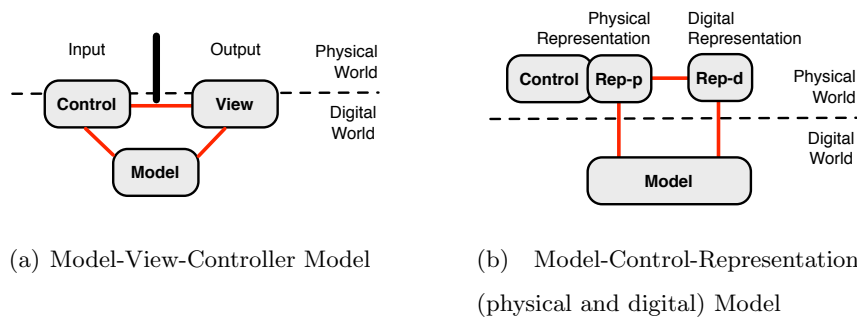
Figure 2.2: MVC and MCRpd models [Ullmer and Ishii (2000)].

geometric primitives;

2. physical representations (rep-p) embody interactive control mechanisms, since the interface movement is the primary mean for control;

3. physical representations (rep-p) are perceptually coupled with digital representation (rep-d); in fact, on the one hand, physical representations are coupled with the control component to trigger information processing, on the other hand, digital representations output the same information, once processed by the computational system;

4. the physical state of the interface partially embodies the digital state of the system.

The relationship between control and physical representation eases the robot control, since physical representations are natural for humans, and do not force to a continuous cognitive switch between the view component and the controller device, as it happens with classical graphic interfaces.

2.3.3   Foreground and Background Tangible Processes

Orthogonal to the architectural analysis of the MCRpd, a different theoretical model is focused on the perceptive level. Ishii and Ullmer (1997) characterize tangible interfaces in terms of *foreground* and *background* interactions:

1. *foreground* activities leverage humans capability to grasp and manipulate physical objects;

2. *background* interactions enable users to unconsciously perceive information from the environment background.
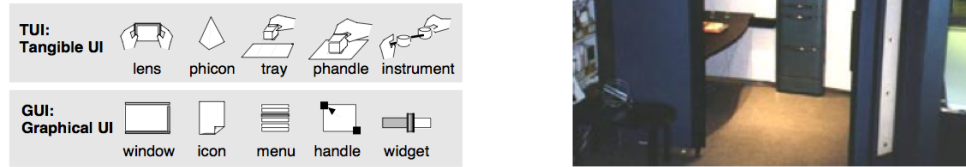
(a) metaDESK associates a graspable physical object to a corresponding virtual artifact

(b) ambientROOM provides background information (e.g. through surround speakers) and mechanisms to move such knowledge to the foreground.

Figure 2.3: Two research prototypes that highlight relevant elements, respectively in foreground and background cognitive processes [Ishii and Ullmer (1997)].

Two examples allow to clarify both foreground and background interactions. The metaDESK prototype attempts to instantiate physical objects as representations of typical GUI elements. For example, the concept of icon is represented by *phicons*, menus are embodied with *trays*, and windows as active *lens* (see Figure 2.3(a)). The ambientROOM projects aims at assessing how we can take advantage of natural background processing using ambient media, such as ambient light, shadow, sound, airflow, or water flow, to convey information. Furthermore, it is interesting to develop a seamless technique to elevate background processes as foreground tasks, as we often do in everyday life when we focus our attention on something previously unnoticed (see Figure 2.3(b)). The importance of background processes is notable in robotics. When controlling multiple robots, particularly in challenging environments, the flow of incoming perceptions is relevant and it is cognitively stressing for a single operator to manage such an amount of information. However, part of this knowledge could be significative for a proper awareness. Designing proactive processes that filter important background information and promote them as foreground perceptions supports an operator in controlling multiple

(a) MediaBlocks

(b) Embodied Games

(c) Breezeway

Figure 2.4: Examples of tangible, embodied and embedded interactions.

robots.

2.3.4 Tangible, Embodied and Embedded Interaction

The term tangible user interface, as well as its definition by Ishii and Ullmer, is not the only formal characterization proposed in literature. Furthermore, it can be seen as a subset of an ecology of approaches related to gesturing and tangibility. Some basic terminology allows to differentiate and contextualize TUIs from other tangible approaches. Jensen et al. (2005) adopt the umbrella term **tangible interaction** to identify a broader set of tangible approaches, which encompasses tangible user interfaces. The shift in phrasing from tangible interface to tangible interaction is intentional, since the qualities of the interaction are moved into the foreground of attention, and system designers are required to think about what people actually do with the system. All the tangible interaction approaches share common aspects such as: *tangibility, materiality, physical embodiment of data, embodied interaction,* and *embeddedness in real spaces.* Below is reported a taxonomy:

1. **Data-centered view.** This approach corresponds to the definition of TUI by Ishii and has been mainly adopted by the human-computer interaction community. It focuses on the idea of *physical objects* as representation and manipulation means of digital data, as already mentioned (see Figure 2.4(a)). We mainly deal with this class of tangibility;

2. **Expressive-Movement-centered view.** Adopted by the design and industrial community, it emphasizes bodily interaction with objects. This represents a shift of the focus from objects and their representations (as in the data-centered view), to the interaction process. According to Djajadiningrat et al. (2002), this consists of exploiting the "sensory richness and action potential of physical objects". Under this point of view, expressive-movement-centered view deals with analyzing what are the typical human actions, and designing systems in accordance to them, rather than dealing only with the design of graspable objects (see Figure 2.4(b)). This is closely related to the concept of **embodied interaction** by Dourish (2004), that will be discussed in Section 2.3.5;

3. **Space-centered view.** This approach has been discussed in interactive arts and architecture. It is based on the concept of **embedded interaction**, that is the reciprocal interaction between an object and the physical world, which in turn gives rise to cognitive processes. Roughly speaking, it concerns in considering the interface (an object, but even our body) as part of the space, the two combined and interacting as a single entity. For example, integrating tangible devices to "trigger display of digital content or reactive behaviors" (see Figure 2.4(c)).

To summarize, these three approaches cope with tangibility from different points of view: objects, object interactions, or spaces where objects are and interact.

Hornecker and Buur (2006) identify four core themes useful for tangible interaction design. We can outline their **tangible interaction framework**:

1. **Tangible Manipulation** (TM) refers to the material qualities and the manual manipulability realized through the bodily interaction with physical objects, that are coupled to computational resources;

2. **Spatial Interaction** (SI) refers to the space where the interaction is embedded. This involves moving physical objects in the space, but also moving someone own's body in the space, which exposes humans' body as the interface itself;

3. **Embodied Facilitation** (EF), that is, how the configuration of objects and space affects and directs actions and processes;

4. **Expressive Representation** (ER) concerns the adopted material and digital representation, which in turn affects the interface expressiveness. Some examples of representations will be presented in Section 2.3.6.

In particular, Chapter 4 focuses on the first and fourth category, while the second one is addressed in Section 5.

2.3.5 Related Work in Cognitive Sciences

Messing and Campbell (1999) show how gesturing, as well as speech, is one of the first learnt communication means (at around 10 months of age). Several studies addressed the cognitive processes involved in tangible interaction. Understanding these processes fosters the design of robotic systems that effectively exploits tangible interfaces.

A topic which has been extensively investigated in cognitive sciences is the concept of **embodiment**. According to the embodied cognition theory, cognitive artifacts (e.g. ideas, thoughts, concepts, or categories) are determined by humans' physical aspects, that is by our bodies. This theory, opposed to abstract cognitive theories (e.g. Cartesian dualism, or cognitivism), has been promoted by the artificial intelligence and robotic communities. Dourish (2004) states that tangible interaction relates on the fact that the ways in which we experience the world are through direct interaction with it, and that we act in the world by exploring the opportunities for action that it provides to us. Through this interaction with the physicality of real world, meaning is created, discovered, and shared.

Object manipulation and grasping is another important cognitive aspect. An object itself has a meaning for its size, dimension and, in particular, what actions we can perform through it. This latter aspect has been investigated by Gibson (1986), who coined the term **affordance**, that is *the possibilities for action that we perceive of an object in a situation*. For example, a button affords pressing, while a door affords opening, and so on. According to Norman (1999), there are two specific types of affordances: *perceived* and *physical* affordances. A button on a GUI is an example of perceived affordance, since users perceive that it can be pressed, but they do not physically interact with it. On the other hand, tangible user interfaces exhibit physical affordances, as humans act in the real world.

Another relevant research question is whether TUIs provide learning benefits through physical manipulation with respect to the traditional manipulation of graphical virtual objects. This conjecture has been supported by manifold cognitive researchers. According to Norman (1991), physical objects can be considered as *cognitive artifacts* designed to maintain, display, or operate upon information with a representational function, hence they can be considered as

additional knowledge sources and interaction means. Kirsh (1995) observes that people tend to modify their environment to support their cognitive processes for problem solving. Kirsh denotes this manipulation as **epistemic action**, whose benefits are to support humans' memory, lower their cognitive load, and to focus on specific affordances. Conversely, actions directly functional to the problem solution are called **pragmatic actions**. Patten and Ishii (2000) conducted a set of experiments to evaluate how exploiting spatial relations with TUIs can support cognitive processes with respect to traditional GUIs. Subjects had to read ten summaries of recent news articles. In the TUI interface these articles were bound to wooden blocks, while in the GUI version they were represented by icons on the screen. Subjects were then asked to indicate the location of each summary by pointing to the corresponding icon or wooden block. The authors observed that subjects who used the TUI localized better the summaries, for several reasons. First, they could move and organize blocks as they preferred, and the block layout mapped their cognitive representation. Second, moving blocks activates the so called **motor memory**, which is more effective when directly dealing with the content itself (because a block represents the content), instead of moving an intermediary tool (e.g. the mouse). To summarize, while the diffusion of computers forced us to unnatural cognitive processes, tangible interfaces re-activate our innate inclination towards gesturing, acting, and relating with the real environment.

2.3.6 Design of Tangible User Interfaces

According to the MCRpd framework, the design of effective TUIs requires an optimum matching between user requirements and the object-digital information coupling. We can further investigate the principal aspects of the *model*, *control* and *physical representation* components to accomplish this. Throughout this thesis we will focus on a subset of tangible user interfaces. In fact, we are interested in portable devices, equipped with localization and attitude sensors, interconnected with a larger system without any cable.

Model. A **digital binding** is the association established between digital information and physical representation. The question is: what kind of digital information can we bind? This is one of the main research issues in human-computer interaction. A TUI token can be associated to: *static digital media* (e.g. pictures, 3D models, and so on), *dynamic digital media* (e.g. movies, animations, and so on), *digital attributes* (e.g. colors, textures, and so on), *computational operations*, *data structures* (e.g. lists, trees, and so on), *people*, *places*, or *objects*.

Control. Associations are established through the control component, and can be divided into two classes: **static binding**, that is specified by the system designer a priori, and **dynamic binding**, established by the user of the interface, which can be changed at run-time.

Physical Representation. The physical representation is the most relevant aspect when designing a tangible interface. Nevertheless, not always design practices have been applied to TUIs. Most of times, the "found object" approach is used, that is, pre-existing objects are equipped with some sensors. Another approach is ruled by engineering electronic or mechanical components, without dealing with aesthetics. Finally, a last approach is to re-adapt physical artifacts adopted in workplace activities. The problem of an effective interface is the trade off between functionality, pragmatics, and aesthetics. Finally, physical representations can also involve more than one artifact, hence the physical-digital binding is established through the interaction among a group of objects. We can identify three interaction approaches:

1. **spatial approaches** map spatial configurations among the different artifacts, such as distances, differences, measures, or angles, onto the underlying digital information. For example, the position or orientation of small bricks can reflect the position of buildings on a map;

2. **relational approaches**, where logical relationships between artifacts enhance physical representations. Some relational examples are: ordering, sequences, proximity, stacking, or priorities;

3. **constructive approaches** allow to associate a meaning to the composition and aggregation of more artifacts, typically performed with modular objects.

In this thesis, we rely on spatial approaches, designing effective physical representations to enhance human-robot interaction.

2.3.7 Related Work in Human-Computer Interaction

Tangible user interfaces have been widely studied and applied in the human-computer interaction (HCI) community. Exploiting this expertise in robotics is only partially feasible. In fact, the human-robot interaction paradigm, as characterized by Drury et al. (2003), is unique. First of all, it involves two active entities: humans and robots. Furthermore, there is an asymmetric bidirectional situation assessment, since robot and human cognitive skills are not comparable. Still, some interesting results can be transferred to robotics. The need for a common characterization of the different tangible interfaces

available has fostered the definition of several taxonomies. Ullmer et al. (2005) identify three types of TUIs:

1. *Interactive Surfaces*: surfaces that interpret tangible objects placed on their top, as well as their relationships;

2. *Constructive Assembly*: modular objects composed in meaningful configurations, according to spatial or relational approaches;

3. *Token and Constraint (TAC)*: tokens (objects) constrained in the space by racks, stacks or slots.

In our work, we only consider portable devices belonging to the token and constraint class. Fishkin (2004) proposes a taxonomy based on two axes: **embodiment** and **metaphor**. The more an interface is far from the origin of the two axes, the more it can be considered as "tangible". The first axis, embodiment, represents *the degree of integration of the control (input) and view (output) components within the physical object that is manipulated*, and it is composed of four levels:

1. *full*, where the output device *is* the input device (e.g. an example could be a PDA-based interface);

2. *nearby*, where the output device is *near* to the input (e.g. in the Bricks prototype, the user moves bricks and can see the effect on the display below the bricks);

3. *environmental*, where the output device is *around* the user, like in the case of the ambientROOM prototype, where audio speakers are triggered by the input device (this is also referred as *non-graspable* tangible interaction);

4. *distant*, where the output is on another screen or another environment.

The second axis, metaphor, describes the type and strength of analogy between the interface and similar actions in the real world. This is particularly relevant in tangible interfaces, where grasping or manipulating an artifact is the metaphor of something else. This category includes five possible levels:

1. *none*, where the action performed on the input devices is not in analogy with the action effect (e.g. typing a command with a keyboard is not related to the command started);

2. *noun*, the look of an input object is closely tied to the look of some real world object, hence physical properties are more relevant than gesturing;

3. *verb*, the action performed with the input object is closely tied to the action performed on the corresponding real world object, hence gesturing is more relevant than physical properties;

4. *noun and verb*, combination on noun and verb metaphors, and the only difference is that input and real world objects are different;

5. *full*, in this case, there is not any metaphor, since even the two objects (interface and real world) are the same object (e.g. a digital pen is a real pen, and both can alter the same document).

As well as taxonomies, another aspect to understand the applications of tangible interfaces in robotics is to look at their application in human-computer interaction. TUIs in the HCI community have been mainly adopted and evaluated for learning, collaborative planning and entertainment applications. Fostering collaboration seems to be a relevant and peculiar aspect of TUIs, where significant benefits can be gained. This is especially true in those tasks which involve spatial cognition, such as graphic design. Kim and Maher (2005) compare design collaboration performed with GUI and TUI environments. In particular, different designers' behaviors were considered. For example, with the TUI environment, verbal interaction was less used in favor of moving 3D models. Moreover, with the GUI environment, people discussed but just one person operated the interface, while in a TUI environment everyone actively participated. This could be a potential benefit in robotics too, where human-robot dialogue systems still represent a challenging issue. The HCI community has also explored the combination of TUIs with augmented reality, which leads to *tangible augmented reality*. Since in human-computer interaction the user is not required to move in a real environment, virtual environments allow for spatial manipulation, while sitting in front of a calculator. An example is the work by Gallo et al. (2008). They adopt a Wiimote controller as a 3D interaction system for medical data manipulation in a virtual reality environment. Users can manipulate 3D objects reconstructed by 2D medical scans. Through the tangible interface they can rotate, translate, and zoom the object, crop and select object parts. An interesting aspect of their work is the virtual pointer mode, which allows for pointing inwards/outwards the scene through a *fishing reel* flavor of the ray casting technique. It is a challenging issue how to apply a similar technique in robotics, when dealing with a real environment, in order to have TUIs acting as tangible pointing devices. A study by Looser et al. (2007) compares three different selection techniques, still in a virtual environment, for object selection: tangible virtual lenses, virtual hand, and virtual pointer. Collaboration could lead to the design of multi-human TUI-based interfaces in robotics, but to the best of our knowledge it has never been applied yet. In particular, Marshall et al. (2007) performed an interesting study on

the potential benefits of TUIs as shareable interfaces, showing how non-native English speakers or shy people tend to interact more in collaborative tasks.

2.4   Robot Operation

Tangible user interfaces exhibit manifold interaction metaphors, that can be applied to different robotic applications for different tasks: social, entertainment, domestic, rescue, surveillance, and so on. Orthogonal to these domains there exists an activity that will be extensively treated in this thesis: the need for humans to operate one or more robots in the environment. This capability is necessary to perform any other high-level task, and achieving it in some *field applications*, is not trivial at all. The easier controlling a robot, the more an operator concentrates on her long-term goals, improving her awareness. Part of the robotic community addresses this problem under the label *teleoperation*. However, our feeling is that this term is nowadays overused and ambiguous. On the one hand, it has been used to represent the whole class of human-robot control approaches. On the other hand, it has been specialized to a specific type of control. Another common term is *robot control*. However, this is closely related to theory of control, which is something not related to the human-robot interaction. Finally, we do not adopt the generic term *human-robot interaction*, since it involves any type of interaction between humans and robots, while we are interested only in the control of the robot locomotion within an environment. Thereby, since we opt for the definition of a precise terminology, we will address this problem with the term **robot operation**, which embraces a whole set of approaches. Conversely, we refer to *robot teleoperation* as a particular approach that will be discussed later.

2.4.1   A Theoretical Framework for Robot Operation

Rephrasing a definition provided by Fong and Thorpe (2001), robot operation means simply: *operating a robot*. Here we propose a conceptual framework, composed of the following three components: **user**, **robot interface**, and **robot** (see Figure 2.5). The user is the human being who operates the robot. We adopt the generic term *user* since it includes all the different roles that a user can embody. Scholtz (2003) lists these roles (see Figure 2.6):

1. **operator**, who operates the robot through low-level *actions* for short-term goals;

2. **supervisor**, who monitors the overall situation with respect to long-term goals and can re-plan actions; they mainly deal with *perceptions* coming from the robot sensors for SA, and with planning;
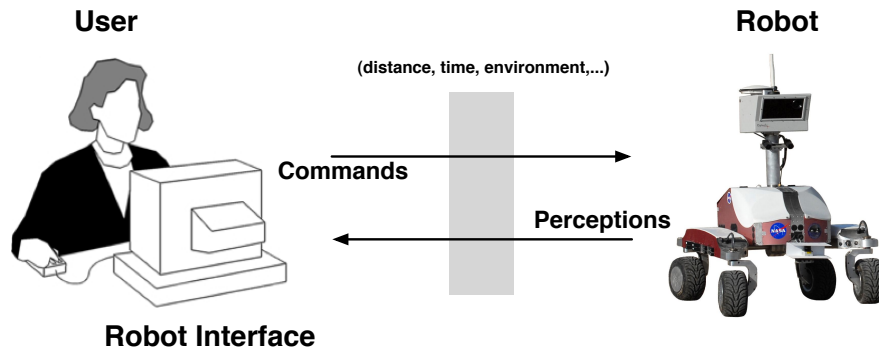
Figure 2.5: Our robot operation conceptual framework. The three main components are highlighted: *user*, *robot interface*, and *robot*.

3. **mechanic**, deals with hardware functionalities, in order to achieve the desired robot actions and behaviors;

4. **peer**, which is a robot teammate that can provide commands according to long-term goals, but cannot re-plan the mission as supervisors;

5. **bystander**, who typically is provided only with a subset of actions to control the robot, with a support role.

Our model is flexible enough to allow for one user, more than one user, or no user at all. The latter is a special case in the robot operation spectrum of capabilities, which corresponds to *full autonomy*. The fundamental component of this model is that robot operation requires a **robot interface**. Every robot interface provides at least two functionalities:

- *input functionalities*, to send commands to the robot;

- *output functionalities*, to visualize incoming perceptions from the robot sensors.

Due to the wide scope of this model, it encompasses manifold approaches to robot operation, which in turn can be classified according to different aspects: robot autonomy level, human/robot ratio, proximity, and so on. In the rest of this section, we briefly introduce these operation techniques.

2.4.2   Autonomy Level

Among all the different aspects of robot operation, the **level of autonomy** is one of the most relevant, and it has been considered in manifold taxonomies
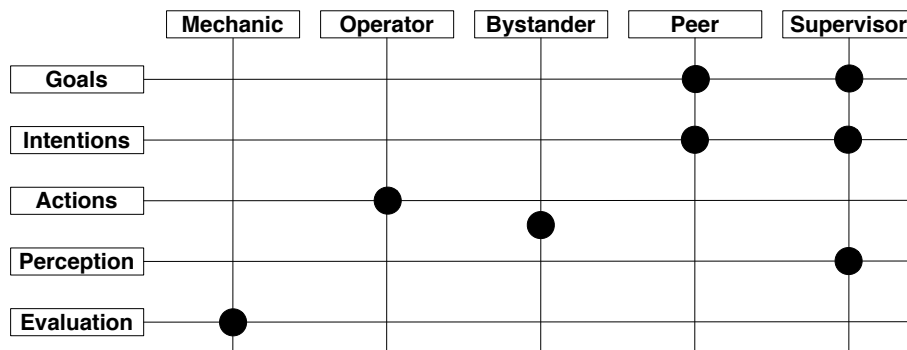
	Mechanic	Operator	Bystander	Peer	Supervisor
Goals				●	●
Intentions				●	●
Actions		●	●		
Perception					●
Evaluation	●				

Figure 2.6: The user roles proposed by Scholtz for HRI and inspired by the Norman's HCI model.

as a classification category. We attempt to find a trade-off among the different formalizations proposed in literature, adopting our own classification. We divide robot operation approaches into four classes: **teleoperation**, **safe teleoperation**, **supervisory control** and **full autonomy**.

Teleoperation

According to Murphy (2000), teleoperation means simply to operate a vehicle or a system over a distance (in fact, *tele* means "remote"). Indeed, despite the etymology of the term, for the rest of this thesis we remove the "remote" constraint, in order to include also those approaches characterized by same control characteristics, but in proximity of the robot. This is considered as the most basic operation paradigm. It involves a manual, continuous, and low-level control of one robot by a human, acting as an *operator*, through any type of *input device*: joystick, keyboard, wheel, and so on. Even in case of multiple robots, just one robot at a time can be operated.

In case of remote teleoperation, the operator is unable to access directly the environment; for example, because the robot is far away, as for space robots, or because it is in a protected area, as in a nuclear plant. Thereby, robots must be equipped with some kind of sensor to acquire data from the environment, and transmit them to any *output device* available to the operator. As we can guess, the bottleneck of this approach is then the communication bandwidth, which could be limited, and the sensors, which could be not enough to acquire a good situation awareness.

From the cognitive point of view, teleoperation is the most stressing operation type, for several reasons. First, the operator is focused only on contin-
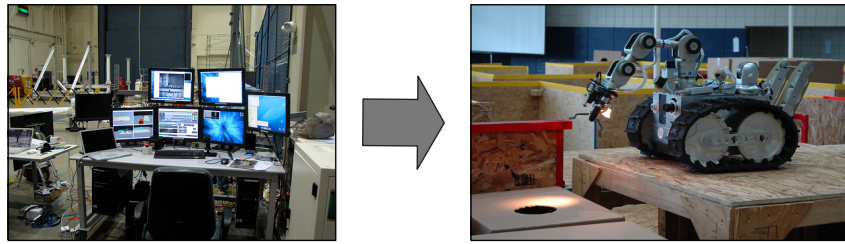
Figure 2.7: An example of interface for the teleoperation of a urban search and rescue robot.

uous robot operation, often forgetting about the real task. Second, situation assessment is constrained by the robot sensors and the output display, which is very fatiguing. Finally, in case of remote operation, a significant time delay can occur, which is frustrating. Still, nowadays teleoperation is the only realistic approach when dealing in very challenging situations, such as in field application. Thereby, we are interested in providing novel approaches to teleoperation, in order to lower the operator's cognitive effort, and we present a tangible-based interface in Chapter 4.

Safe Teleoperation

In case of scarce SA, teleoperation is not effective at all. The operator, who has full and direct access to the robot, could issue wrong commands, which in turn may cause safety risks for the robots and humans in the environment. Safe teleoperation is a particular type of teleoperation where autonomous preemptive techniques are used to prevent operators from wrong actions. This mode retains the direct control of the robot by the operator, but simple collision avoidance is performed. This allows the robot to avoid any part of the motion which would result in a collision, thus preventing the operator from accidentally causing damage by a wrong operation.

Supervisory Control

The term "supervisory" denotes a shift towards a higher level of autonomy, where the user acts as a *supervisor*. Her role is to command a robot for a specific task, or sub-goal, that the robot can accomplish autonomously. In particular, there are two types of supervisory control: **shared control** and **control trading**.
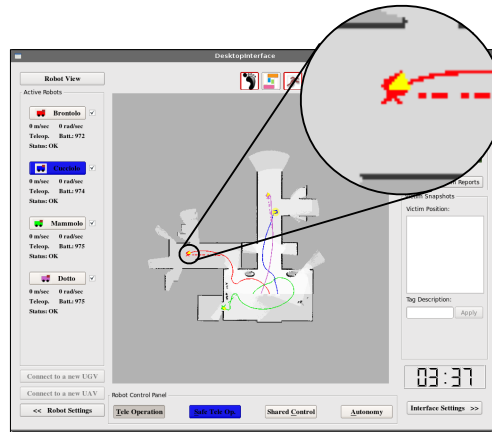
Figure 2.8: An example of shared control. The user clicks on a viewer to select the desired target pose for a robot, which moves towards the target autonomously.

Shared control is characterized by the possibility, for the supervisor, to delegate a sub-goal to the robot, which will accomplish it autonomously (see Figure 2.8). However, the supervisor has to control the robot execution and stop it whenever some problem arises (e.g. the robot gets stuck), or simply because she wants the control back. It is a useful approach to lower the cognitive stress in repetitive and boring tasks, but a continuous supervision is still required. We deal with this operation style in Chapter 5.

Control trading is even higher level. In this case, the supervisor only triggers the robot for a task, but she does not spend time in monitoring its execution, since the robot is considered fully capable of accomplishing it. It is a very effective approach, still humans tend not to trust completely robots, thereby shared control techniques are often preferred to control trading.

Full Autonomy

In most HRI taxonomies, full autonomy is not considered at all. However, we include it in the spectrum of robot operation approaches, considering it as a particular type, where the level of autonomy is full. Of course, in this case the role of the user is as a mere *observer*, and the human-robot interaction component is neglectable.
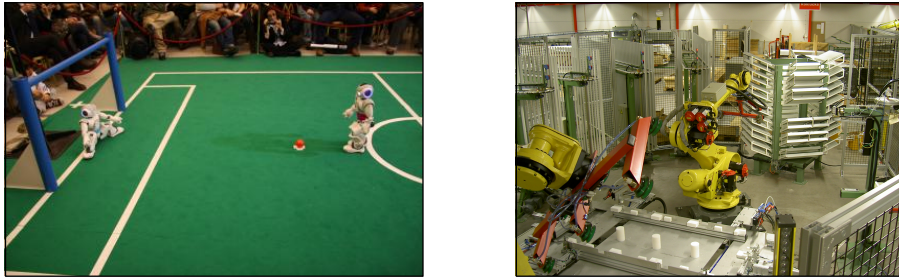
Figure 2.9: Two examples of robotic applications where it is possible to tend towards full autonomy: *(i)* soccer robotics and *(ii)* manufacturing.

2.4.3 Robot Operation Taxonomies

In the previous section we sketched out a taxonomy of robot operation approaches, classified according to their level of robot autonomy. However, different other aspects can be considered. Yanco and Drury (2005) propose a HRI taxonomy based on the following categories.

Task Type. The task type characterizes the design of the robot operation interface. For example, when acting in a urban search and rescue environment, a remote interface should be considered, and a teleoperation approach could be adopted, due to the challenging setting.

Ratio of people to robots. This ratio affects the human-robot interaction in a system. Here we are not considering a measure of the interaction between humans and robot, but simply the numbers of each. For example, Murphy (2004) characterizes USAR activities by a ratio of two, that is, at least two operators are required to control a single robot, because of the hard environmental conditions.

Level of shared interaction among teams. With respect to the ratio of people to robots, this aspect investigates the type of interaction among the team members. Yanco and Drury (2002) identify eight different relationships: *one human - one robot, one human - robot team, one human - multiple robots, human team - one robot, multiple humans - one robot, human team - robot team, human team - multiple robots, multiple humans - robot team.* The difference between interacting with a *group* or *multiple* is relevant. For example, in case of a robot team, the operator sends the command to the group, then the robots jointly nominate a robot as the receiver, hence every robot knows

what its teammates are doing. On the other hand, in case of multiple robot, the operator decides the receiver of an instruction. In this thesis, we present only single human - single robot interfaces.

Time/Space constraints.   There are four categories, according to time and space categories. As for time, humans and robots can operate at the same time (*synchronous*) or at different times (*asynchronous*). Concerning the space, they can be deployed in the same place (*co-located*) or in different places (*remote*). Some examples are reported in Table 2.1. We deal with synchronous approaches, either co-located or remote.

		Space	
		Same	**Different**
Time	**Same**	Robot Wheelchair	Urban Search and Rescue
	Different	Manufacturing Robots	Mars Rover

Table 2.1: Robotic applications classified according to space and time categories.

Chapter 3

Related Work

3.1 Introduction

In this chapter we provide an overview of relevant related work to situate our thesis contributions. Our analysis moves along three different research lines. First, we aim at describing how the problem of awareness, our research question, has been addressed by the robotic community (Section 3.2). In accordance with the main approaches present in the literature, we also discuss our past research investigation with respect to this problem. Our results highlight the limitations of current approaches and introduce the main motivations behind the contributions of this thesis. Second, in Section 3.3 we sketch out how tangible user interfaces have been adopted in robotic systems, focusing on robot operation. This highlights what may be the benefits of TUIs in robotic systems, and introduces some preliminary aspects further discussed in Part II and Part III of this thesis. Third, we address semantic knowledge in robotics (Section 3.4), in particular focusing on contextual knowledge, which represents our last goal of this thesis. A brief discussion concludes this chapter and summarizes the main aspects that motivate our contributions.

3.2 Approaches to Awareness in HRI

In Section 2.2 we introduced a formal definition of human-robot interaction awareness and situation awareness. Here we sketch out different approaches present in literature related to the problem of HRI awareness. Solutions to this problem can be roughly divided into two classes: enhancing the operator's cognitive capabilities by providing multilevel knowledge, and adopting

information visualization best practices for effective interface design.

Concerning the former category, we focus on **location awareness** and **surrounding awareness**, since these are fundamental for robot operation. In order to better understand how surrounding awareness enhances the operator's performance when she is operating a robot, we borrow two concepts from human spatial cognition: **route knowledge** and **survey knowledge**. The distinction between route and survey knowledge highlights which cognitive skills are needed by a human operator controlling a robot. Route perspective is closely linked to perceptual experience. It consists of an egocentric perspective in a "retinomorphous reference system", where the subject is able to perceive himself in the space. According to Herrmann (1996), this perspective emphasizes spatial relations between objects composing the surroundings in which the subject is situated. An example is an operator controlling a robot with a three-dimensional display that simulates the visual information that she would obtain by directly navigating in the environment. By contrast, survey perspective is characterized by Cohen (1989) as an external and allocentric perspective, such as an aerial or map-like display, hence it facilitates direct access to the global spatial layout. It recalls an operator with a device that enables a global, aerial view of the environment and the robot. Previous studies by Herrmann (1996) show that an operator having access to both perspectives exhibits more accurate performances.

We can therefore relate *location awareness* with *survey knowledge*, while *surrounding awareness* is correlated to *route knowledge*. For example, obstacle avoidance depends on the operator's surrounding awareness, that is an egocentric system of reference for deciding the robot direction. The problem, however, is that information about the overall environment remains rigid and relatively poor. By contrast, according to Werner et al. (1997), survey knowledge depends on the operator's location awareness, which is generally considered as a cognitive representation for fast, route-independent access to selected locations structured in an allocentric coordinate system. Most of the solutions embracing the multilevel knowledge approach improve human-robot awareness leveraging the operator's spatial cognitive abilities. This is accomplished in different ways: first, combining different perspectives of the environment (e.g. integrating location and surrounding awareness with a map-view and a first-person view), adding further sensors to the robotic platform, or introducing enhanced view representations, such as virtual reality. Such rich variety of information enables an operator looking at a graphical user interface to have access to more than one perspective at the same time. However, when considering operation conditions where the operator has intra-scenario mobility, these solutions still force the operator to act in a virtual environment and to rely on the interface itself, instead of moving her cognitive attention in the real environment, which can provide a better awareness state.

In the last several years the robotic community has also investigated a different category of approaches, that is, how to exploit spatial cognitive skills through effective human-robot interface design, based on information visualization best practices.

Adams (2002) analyzes how to fill the gap between human factors engineering and robotic research. Different best practices have been evinced through the design of robotic interfaces for demanding robotic applications, in particular those where the role of humans is fundamental, such as field applications. These techniques rely on lowering the operator's cognitive load by aggregating information and designing effective layouts. In particular, a requirement is to reduce the number of windows, and presenting information in a single widget to reduce the continuous switch between different parts of the interface. However, two considerations arise. First, these interfaces are still conceived as debugging applications for robotic experts. Second, information visualization does not affect the adopted representations; data are still represented in numerical form. The latter aspect can potentially improve human-robot awareness, but it is still not sufficiently explored, in particular with respect to field applications. To summarize, approaches based on multilevel knowledge tend to improve the awareness level of the operator, while solutions addressing information visualization attempt to lower her cognitive effort. As well as these two classes, there are also other approaches that allow to identify further guidelines:

- providing a granulation of the autonomy spectrum according to the operator's task and skills;

- supporting the choice of the appropriate robot autonomy level;

- preventing operator errors and predicting her intentions.

In the rest of this section we overview all these approaches, while also considering the intra-scenario operator mobility factor, since we deal with such aspect in our contribution about robot operation, presented in Chapter 4.

3.2.1 Desktop Interfaces

When an operator controls a robot through a desktop computer, often she cannot access the robot nor the navigating scenario. Under this assumption, the human's knowledge of the robot surroundings, location, activities and status is gathered solely through the interface. An insufficient or mistaken surrounding awareness of the robot might provoke a collision, and inadequate location awareness reveals that the explored area does not fit the requirements of the mission or that the task is not accomplished efficiently. When this

happens, the use of robots can be more of a detriment to the task than a benefit to it. It is worth noting that while these interfaces involve powerful graphic resources for output visualization, inputs are typically provided with standard devices, such as keyboards, mouses, or joysticks. While effective in human-computer interaction, these devices are not always the optimal solution to control a complex system like a robot.

Map-centric Interfaces

Map-centric interfaces are mainly focused on the representation of the robot inside the whole explored (or known a priori) environment map, and thereby seek to enhance the operator's location awareness. They provide a bird's eye view of the scenario. This allows an operator to follow the robot, meanwhile concentrating on the map where the robot is located. Map-centric interfaces are better suited for operating remote multi-robot systems than video-centric interfaces, given the inherent location awareness and global assessment that a map-centric interface can easily provide. In fact, the robots relationships, as well as their positions in the environment, can be seen easily understood. However, it is less clear whether map-centric interfaces are effective for single robot control. If the robot does not have any adequate sensing capability, creating the maps that these interfaces rely upon may not be possible. If the map is not properly produced (because of the interference of moving objects in the environment, the presence of open spaces without objects within the laser range, faulty sensors, software errors, and other factors), the user can get confused. Moreover, the emphasis on location awareness inhibits the effective mediation of good surrounding awareness. Prototypes of map-centric robotic interfaces have been implemented by Nielsen and Goodrich (2006), Nielsen et al. (2007), and Driewer et al. (2008).

Video-centric Interfaces

Studies reveal that operators heavily rely on the video feed coming from robots. Video-centric interfaces are thought to provide the most important information through the video, even if other information, including a map, is present. Video-centric interfaces are by far the most common type of interface used with remote robots, and they range from interfaces that consist only of the video image to more complex interfaces that incorporate other information and controls. However, the problem with video-centric interfaces is that whenever they include other information apart from the video, this information tends to be ignored, as demonstrated by Yanco and Drury (2004) and Nielsen and Goodrich (2006). Most existing interfaces are video-centric, one relevant ap-

proach in the human-robot interaction literature is that of UMass-Lowell GUI in Yanco and Drury (2004), Yanco et al. (2004), Yanco et al. (2007), and Drury et al. (2007). UMass-Lowell interface keeps the video centered, while the robot and the map rotate. On the other hand, map-centric interfaces typically move the robot, while keeping the map fixed. A joint work between Idaho National Laboratories GUI (map-centric) and UMass by Drury et al. (2007) compares their respective implemented interfaces. The authors demonstrate that this spatial reference difference does not influence the operator performance.

Integration between Map-centric and Video-centric Interfaces

In the last years, we spent a significant effort in investigating the problem of awareness in HRI according to the approaches introduced so far: information visualization and enhanced view representations. Our aim here is at evincing their drawbacks, which in turn represent the motivations of this thesis: shifting our quest towards innovative input and output interaction means.

More in depth, we present here a robot interface for single human - multi robot operation for structured and semi-structured environments, **DesktopInterface**, reported in Figure 3.1.
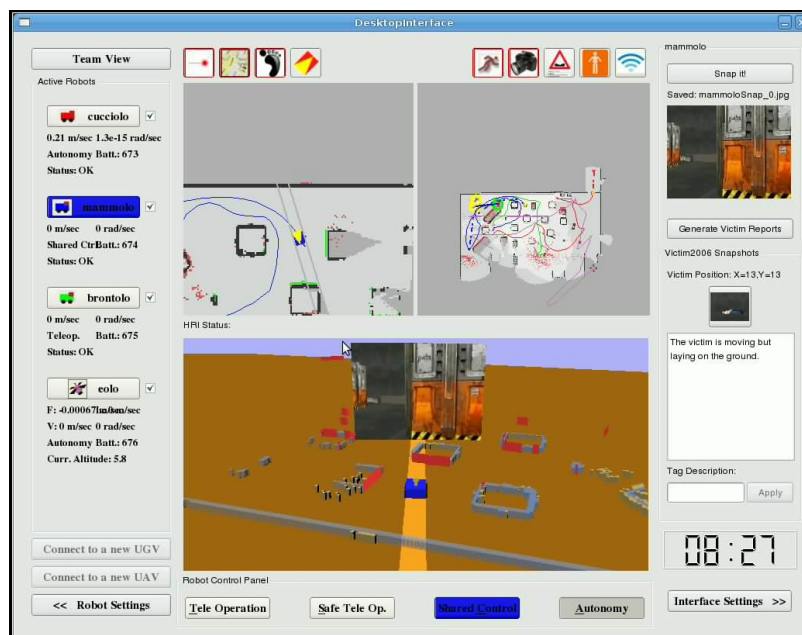

Figure 3.1: The main components of our robot interface *DesktopInterface*.

We sketch out the relevant characteristics of this interface:

1. with respect to the aforementioned approaches, *DesktopInterface* integrates map-centric and video-centric representations. Each representation is more effective under specific conditions, thereby their combination guarantees a proper level of awareness in different scenarios. More in depth, the interface provides three views. First, a *global map view* improves the location awareness of the whole robot team; second, a *local map view* boosts up surrounding awareness with an allocentric perspective when controlling a single robot. Finally, a *3D view* provides surrounding awareness from an egocentric point of view.

2. video feedback is always available and merged with the 3D view, hence the operator does not switch between the two representations, rather has a direct access to both of them, and perceives the video feedback as localized within the environment;

3. information is organized in a single widget, which reduces the operator's cognitive switch among the different interface views;

4. there is a preliminary focus on high-level representations, since the operator can take snapshots or tag relevant parts of the environment, and overlay this information on the map, which allows her to remember relevant aspects when observing explored areas.

The interface provides classical input means, such as keyboard and joystick.

Particularly relevant is the adopted development process, which is based on *user-centered* design. User-centered design is based on the continuous involvement of end users in the development process, in order to highlight the limitations of each prototype and refine it. In the end, the interface is the result of the evolution of several prototypes, continuously evaluated over two years through extensive experiments (Valero et al., 2009a,b,c, 2011). Figure 3.2 reports our first interface prototype compared with the last one. The first prototype has been refined through a set of experiments[1] conducted in our department for four days, involving 56 students of computer science. A second evaluation has been performed during the RoboCup Rescue Virtual Robots 2008, where this interface has been adopted to control our robotic system during the competitions[2], as described by Calisi et al. (2008d).

---

[1]Additional material (e.g. images, recordings, questionnaires) is provided at http://www.dis.uniroma1.it/ randelli/index.php?id=8.

[2]Further details about our robotic team, SPQR, and its involvement in the RoboCup competitions, are available at the RoboCup Rescue Virtual Robot Wiki Page.
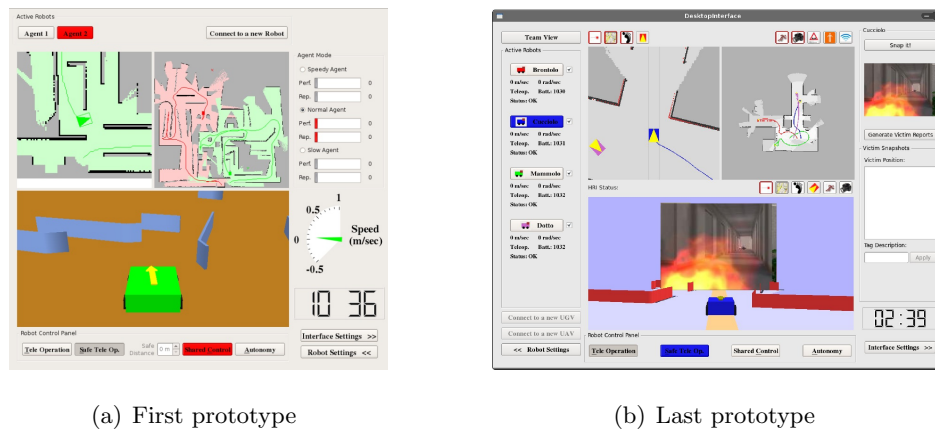
(a) First prototype

(b) Last prototype

Figure 3.2: The evolution of *DesktopInterface* after two extensive experimental evaluations.

The results of these experiments and the participants' feedback revealed how, despite our effort in improving the overall human-robot awareness through the aforementioned improvements, the are severe limitations, mainly caused by the lack of high-level information and the cognitive stress in operating robots. This suggests us to investigate alternative approaches to the problem of awareness. Under this point of view, the contributions of this thesis may be considered as a further evolution of our robot interface.

Augmented Reality

Augmented Reality (AR) is a technology that facilitates the overlay of computer graphics onto the real world. While in virtual reality (VR) a virtual environment replaces the entire physical world, AR "*augments*", rather than replacing, reality. Under this point of view, AR cannot be compared to the previous described map representations, both 2D or 3D, being these latter mere virtual reality simulations through computer graphics techniques. According to Green et al. (2008), AR supports the use of spatial dialogue and deictic gestures and allows the robot to visually communicate to its human collaborators its internal state through graphic overlays on the real worldview of the human, hence it strongly enhances human-robot awareness. In fact, AR allows a human partner to have an egocentric worldview, hence affording spatial understanding of the robot position relative to the surrounding environment. Here we only refer to AR output interfaces, that is reproducing an output

feedback to a human operator. However, it is worth mentioning that AR also fosters a tight integration with input devices, such as tangible interfaces, for a more natural human-robot interaction. Giesler et al. (2004) propose a system that allows a robot to interactively create a 3D model of an object on-the-fly. In this application, a laser scanner is used to acquire an unknown 3D object. The information from the laser scan is overlaid through AR onto the video feed of the real world, hence enhancing situation awareness through relevant object models. In a similar work, Giesler et al. (2005) implement an AR system that creates a path for a mobile robot. Fiducial markers are placed on the floor and used to calibrate the tracking coordinate system. A path is created node by node, by pointing the wand at the floor and giving voice commands for the meaning of a particular node. Map nodes can be interactively moved or deleted. As goal nodes are reached, the node depicted in the AR system changes color to keep the user informed of the robots progress. Finally, the benefits of AR have been confirmed by some experimental evaluations. Maida et al. (2006) show through user studies that AR significantly improves robot control performance. Drury et al. (2006) present a set of experiments showing that augmented real-time video with pre-loaded map terrain data result in a statistical difference in comprehension of 3D spatial relationships over using a simple 2D video for operators of unmanned aerial vehicles (UAVs). Koeda et al. (2005) design an annotation-based rescue assistance system for a teleoperated unmanned helicopter with a wearable augmented reality environment. In this system, an operator controls the helicopter remotely while watching an annotated view from the helicopter through a head mounted display (HMD). Virtual buildings and textual annotations assist the rescue operation indicating the position to search rapidly and intensively.

Haptic Interfaces

A haptic interface is a tactile feedback technology that takes advantage of a user's sense of touch by applying forces, vibrations, and/or motions to the user. This is particularly useful when operating "one-way" systems, that is, systems where the effects of the forces applied to the vehicle structure are not directly perceived by the operator. A common application is in modern airplanes, where a harsh command of the pilot can stall the vehicle, without noticing it. To replace this missing cues, haptic feedback can be adopted to provide the operator with a simulation of the effect of a specific command. Haptic interfaces have been mainly adopted in robotics for remote teleoperation. In particular, different works address the problem of robot manipulators (e.g. surgery robotics, manufacturing robotics). Under this point of view, the haptic feedback represents a force feedback generated by the contact of the

manipulator with the handled object. However, with respect to the scope of this thesis, we are more interested in haptic interfaces for mobile robot operation, that is, focused on the locomotion control of the platform. In fact, haptic technologies can support the operator, meanwhile controlling a robot, to prevent stall conditions, to be notified about unexpected obstacles, to leverage background information about the explored environment. A first relevant work is by Rosch et al. (2002), where a force feedback system for the teleoperation of mobile robots is presented. According to force measurements, detected in front of the vehicle, the operator's haptic joystick generates proportional forces. Every time the robot is pushing an object, the operator feels the related friction force on the interface. The major limitation of the proposed system is that the haptic feedback is continuous, which could frustrate the operator, and it is provided as a response to a bump, and not as a preemption mechanism.

Lee et al. (2002) present a different approach, which partially solves the aforementioned limitations. Their system simulates force feedback according to two different aspects: the position of the obstacles surrounding the robots (environmental force), and the relationship between the obstacles and the operator's commands, in terms of robot speed and jog (collision-preventing force). Thereby, force feedback is related to safe navigation, and not to object contact. A similar approach is proposed by Diolaiti and Melchiorri (2002).

All these solutions share a common interest in safe navigation and control of a mobile robot platform. Still, they do not adopt any particular user-centered design, which is common in the human-robot interaction community. For example, the burden of a continuous feedback on the operator's cognitive focus is never considered nor analyzed. Asynchronous approaches (e.g. discrete and temporally-limited notifications) have not been considered, while they will be a relevant part of this thesis, as described in Section 4.3.

3.2.2 PDA Interfaces

As we have seen in the previous considerations regarding spatial cognition, intra-scenario operator mobility is a great advantage in the context of acquiring situational awareness in robot teleoperation, as the operator has direct access to the environment and, in some situations, is co-located with the robot too. Even if remote operators, using powerful work stations, have access to a larger amount of data, responders carrying a PDA interface can boost the pervasiveness of robotic systems in mobile applications where operators cannot be pinned down in a particular place. Even if mobile devices are less powerful than desktop computers, they offer the operator the capacity to move, thus allowing him partially to view the actual scenario with the robot that

she is controlling. The disadvantages related to device limitations could be balanced by the advantage of mobility, and indeed novel smart-phones and tablets with much more computation power are emerging. Mobility could improve situational awareness and robot control. Operators could control, through innovative layout and input paradigms, a robot team with a PDA interface while having a partial view of the environment, and thus acquiring on-field information not retrievable by the robot sensors.

Some research groups have designed graphical user interfaces for PDAs. Fong et al. (2001) propose a PDA-based HRI interface well suited for unstructured or unknown terrain as well as for cluttered environments. A single window layout provides the robot relative position, rate, and waypoint (image and map-based) control modes. Another prototype by Fong et al. (2003) involves a teleoperation system model called *collaborative control*. In this interaction paradigm a robot asks questions to the human in order to obtain assistance with cognition and perception. Thus, human and robot work in a complementary manner, collaborating to solve problems. Collaboration is realized on a PDA device using written queries and replies. For exploration and navigation: Kaymaz-Keskinpala and Adams (2004) designed a PDA teleoperation interface that adopts only touch based interactions. Three screens were developed: an image-only screen, a sonar and laser range finder screen, and an image with sensory overlay screen. Skubic et al. (2003) propose a PDA sketching interface that can be used to direct a mobile robot along a specified path. In human-to-human communication, hand drawn route maps are often used to show the desired navigation path. Because sketched route maps are not drawn precisely or necessarily to scale, they do not attempt to analyze precise path information, but rather qualitative route information is extracted. The interface is developed on a PDA device, where the user can label relevant elements of the sketched map.

Intra-scenario operator mobility is a relevant topic of interest for us, since tangible user interfaces move human-robot interaction directly in the real world. However, as we will discuss in Chapter 4, because of intrinsic differences between PDAs and TUIs, the aforementioned results do not always apply to tangible devices.

3.3   Tangible User Interfaces in Robotics

Tangible user interfaces take advantage from two human fundamental skills: human presence in real environments and interaction with objects. Humans *act* in the space, *interact* with the space, *modify* the space. Generally speaking, two questions arise when introducing a novel interaction paradigm: (i) *which cognitive processes are involved when adopting such a paradigm?*, and (ii) *how*

can these processes be exploited and transferred to real applications?. We already addressed the former issue in Section 2.3. In this section we describe how TUIs have been applied to robotic systems and which interaction paradigms exhibit benefits within this context. However, as we will discuss later in Section 3.5, the feeling is that the two aforementioned research questions are still decoupled, and a comprehensive formal assessment of these devices in robotics is still missing. That is, despite several attempts to conduct empirical evaluations, or to propose a formal assessment, to quote Marshall et al. (2007): *"we really do not know why, how or whether"* tangible interface benefits can be substantiated.

3.3.1   Tangible Interfaces for Robot Operation

Teleoperation

Robotics includes a relevant set of tasks, labeled as *field applications*. Field applications are performed in unstructured or dynamic environments, such as: construction, forestry, agriculture, mining, subsea, intelligent highways, search and rescue, military, and space. According to Thorpe and Durrant-Whyte (2003), *field robotics* aims at the automation of platforms operating in environments where they need to safeguard themselves while performing non-repetitive tasks. Despite this, the more an environment is challenging, the more deploying fully autonomous robots is unrealistic, and the role of humans is fundamental. The simplest approach, under these conditions, is to adopt robot teleoperation. As already described in Section 2.4, this interaction paradigm is characterized by a low-level control that fully absorbs the operator with a high cognitive effort. Moreover, such an approach allows for controlling one robot per time.

Several studies address the problem of robot teleoperation with tangible interfaces through *motion detection*, that is the measurement of any speed change (represented in a vectorial form) in an object. Since several portable TUIs are equipped with accelerometers or gyroscopes, this is easy to accomplish. A formal definition of the motion detection problem is presented in Section 4.2. As well as detecting motion, a second relevant aspect is the definition of the physical representation that *maps* the interface motion onto robot commands (e.g. move forward, move backward, turn left, turn right). In fact, designing a natural and effective physical between the tangible interface and the corresponding robot commands is the most crucial aspect. Motion patterns should be easy to realize, as well as cognitive meaningful for their corresponding commands. Rouanet et al. (2009) associate the Wiimote forward/backward pitch to move a wheeled robot forward/backward,

and left/right roll to turn the robot left or right. A similar approach is proposed by Sharlin et al. (2004). Guo and Sharlin (2008a) define a higher level mapping to control a Sony AIBO robot. Using two Wiimotes, their pitch and roll are mapped onto eight commands (walk forward, stop, forward plus turning left, forward plus turning tight, strafe left, strafe right, rotate left, rotate right).

To the best of our knowledge, none of these approaches consider background information as a relevant aspect of robot operation. Conversely, in Section 4.3 we embed in our teleoperation interface two tactile feedback mechanisms to move relevant background information to the foreground, thus implementing a pre-attentive system that supports the operator to recognize unnoticed events, which in turn improves her awareness.

Gesture Recognition

Gesture recognition is the interpretation of human gestures via data processing techniques. Since gestures are body movements, motion detection is a first step towards gesture recognition, but the two concepts cannot be considered as equivalent. In fact, gesture recognition is a high-level process, which is based upon motion detection, but leverages on other processes as well. An interesting contribution of Part II of this thesis is to match each tangible interaction paradigm with a corresponding robot control style, in order to maximize the operation performance with respect to every paradigm. In particular, motion detection and gesturing address two different control styles: continuous control the former, and shared control the latter. With respect to camera-based gesture recognition, which is the most common approach, tangible interfaces allow for an easier gesture recognition in case of movement, while they exhibit problems in case of still poses. Varcholik et al. (2008) developed a hand and arm gestural control for a wheeled robot. They adopt a linear classifier based on 29 features that recognizes 3D gestures and requires few training samples per gestures (about 20), with a 95% classification accuracy. They recognize four gestures (move forward, stop, turn left, turn right), which are mapped onto four robot actions with a predefined speed. For example, after the command to move forward, the robot starts to move until the stop gesture is performed and recognized by the system. Mlích (2009) define a similar gesture recognition algorithm based on hidden Markov models to detect eight basic gestures: up, down, left, right, shake, shake to side, circle, square. Despite a reduced feature set (only ten), the overall system accuracy is rather low for some gestures (about 70% for shaking side, and about 78% for shaking). A critical aspect of gesture recognition for robot control is the error rate of the implemented systems. Errors in gesture recognition can prevent the robot

correct operation, which increases safety risks for humans.

In Chapter 5 we present our solution for a gesture recognition interface for robot shared control, which is based on a SVM classifier.

Tangible Pointing Interfaces

Moving further towards high level interaction metaphors, tangible interfaces can be used as pointing devices. A pointing device is an input interface that allows a user to input spatial data to a computer. For example, a mouse allows to move a graphical pointer and to select areas of the desktop. As already mentioned, the human-computer interaction community has developed different pointing techniques in virtual reality environments. Tangible interfaces can act as pointing devices in real environments, but this is a much more challenging problem. A formalization of the problem is provided in Section 6.2. Roughly speaking, there are two classes of problems: first, accelerometers and gyroscopes allow for proprio-perceptive localization and attitude of tangible devices, but they accumulate considerable error over time. Second, TUIs are not equipped with feedback sensors (e.g. infrared, ultrasonic, laser, and so on), hence measuring the distance between the device and the pointed location or retrieving the height of the selected object is not affordable. Without any other sensor, it is only possible to select ground areas, which is useless in 3D environments, in particular in case of harsh conditions. To the best of our knowledge, the only work facing this interaction metaphor in robotics is by Kemp et al. (2008). They present a novel interface for human-robot interaction that enables a human to intuitively select a 3D location and communicate to a mobile robot to move towards it. The user *clicks* the desired spot with a green laser pointer, while the robot detects the point with an omnidirectional camera, and moves towards it estimating its 3D location with a stereo pan/tilt camera. However, this approach does not involve any TUI. Selecting elements within a real scenario is a relevant aspect that we discuss in Part III of this work. In particular, we focus on two different issues: first, how to solve the selection problem and, second, how to exploit this interaction paradigm to acquire semantic knowledge, by grounding concepts represented with explicit formalisms into the selected perceptions. This establishes a tight synergy between tangible user interfaces and semantic knowledge. More in depth, we aim at considering a novel situation assessment methodology where the operator has an active role in the grounding process.

Such an interaction paradigm requires the operator and the robot to be, at least partially, co-located. When acting from a remote site, pointing in a real environment is not an option, but a virtual environment resembling a real one can be adopted. For example, a tablet surface can simulate a real

ground, where moving objects on its top corresponds to the movement of real robots. The more the object recalls the controlled robot, the more effective the interface is. In fact, objects would allow for the same affordances of the real robots. Taking inspiration from Ishii and Ullmer (1997)'s "phicons", Guo and Sharlin (2008b) adopted small physical objects resembling real robots, called *Ricons*, to control a team of Sony AIBO dogs.

Multi-Robot Systems

Tangible interfaces in robotics have not been so widely analyzed for collaboration as in human-computer interaction. Few approaches of single-human / multi-robot (SHMR), multi-human / single-robot (MHSR), and multi-human / multi-robot (MHMR) are nowadays available. When dealing with multi-robot systems, TUIs can also be used for task assignment, control switch, and robot cooperation. Lapides et al. (2008) developed a three dimensional tangible interface called 3D Tractus, based on the *drawing board* interaction metaphor. It consists of a table surface that slides up and down on four vertical tracks, simulating a vertical movement in the operation space (e.g. a building, the air, and so on). This interface has been designed for multi-robot control. The operator act as a team leader and can set target poses in the environment moving on the surface or sliding up and down the tablet itself. For example, to move a robot from one floor to another one (in case of a building), the operator would slide up/down the tablet. This interaction paradigm involves shared control rather than manual teleoperation.

3.3.2 Comparisons between TUIs and Other Techniques for Robot Teleoperation

Experimental evaluations about tangible interfaces as robot manual controllers (teleoperation) have mainly pursued two main goals: comparing tangible interfaces with more traditional operation approaches in HRI, and identifying effective *spatial mappings*. According to Sharlin et al. (2004), a spatial mapping is *the relationship between object's spatial characteristics and the way they are being used*. It is evident then that, under this point of view, HCI results cannot be transferred to HRI, since robot behaviors are pretty different. A big effort is committed to search effective spatial mappings, in order to associate natural gesture that reflect the physical state or functions of robots. Guo and Sharlin (2008a) evaluated the performance of a Wiimote with respect to a classic keypad through navigation and posture tasks, using a Sony AIBO robot. The main idea of their study is that robot teleoperation is a low-level cognitive fatiguing activity, which is performed to accomplish high-level tasks. Their re-

search question is whether tangible interfaces, being more natural than more conventional devices, lower the cognitive effort in robot teleoperation, shifting the operator attention towards high-level problem solving. Their experimental testbed is based on a navigation task for comparing the two interfaces in terms of speed, accuracy, and subjective preferences of the participants. The navigation task consists of two sections: an easy and a hard route. Participants have to complete both of them as fast as possible, and with as few errors as possible. Results show that the Wiimote device outperformed the keypad in terms of task completion time. In fact, participants do not need to focus on their hands and do not switch their attention between the robot and the interface. However, the differences between the two interfaces, although statistically significant, are underwhelming in their magnitude, and the authors do not claim that such a result can be extended to any teleoperation task. Analogous results are presented by Song et al. (2007), where they compare four types of devices: a pad-like device, a joystick, a driving device, and a Wiimote equipped with motion sensors, with the aim of measuring the user convenience. Participants control a Sentinel wheeled robot, and perform three tasks: rotation, moving, and speed. In each of these tasks the user performance in achieving specific angles, places, or speeds is evaluated. According to the data analysis, there is an overall performance homogeneity among the different devices. Buttons emerge as the worst interface, while joystick as the best one, and tangible interfaces are in the middle. However, results change noticeably according to different tasks, and to different conditions. Rouanet et al. (2009) present a comparison among three human-robot interfaces using handheld devices for teleoperation. To the best of our knowledge, this study represents the only attempt to compare handheld interfaces exhibiting novel interaction metaphors. In particular, the three proposed interfaces are: *(i)* a touch screen interface on a iPhone (where the user can define trajectories), *(ii)* a gesture interface on a Wiimote, and *(iii)* a virtual arrow interface, again on iPhone. Participants control a robot through two different obstacle courses: easy and hard, and they are co-located with the robots. Once again, it is interesting to see how there is no significant difference in the completion times among the three interfaces, with just a slight advantage with the tangible interface (probably because no context switching between the robot and the interface was required). Two important issues concerning the tangible interface arises. First, participants expect a better mapping between their movements with the Wiimote, and the corresponding movements of the robot. This is something that once again confirms how HCI results cannot easily be transferred in the robotic community. When dealing with TUIs, users expect a videogame-like behavior, which is not realistic nor affordable when controlling robots. Furthermore, participants feel frustrated for the lack of video feedback, as in the other two interfaces. The lack of video feedback is a well known issue with

tangible interfaces, and one of the main reasons why TUIs are often considered as a complementary tool of a multimodal interface, where video is provided as well.

Some considerations arise from these experimental evaluations. First, TUIs do not seem to overwhelm the other interfaces, rather a significant advantage arises when considering very specific environment and mobility conditions. However, a rigorous assessment of such conditions is still missing in literature. We address this aspect with an extensive experimental evaluation in Section 4.6. Second, asynchronous tactile feedback (for example, in the form of tangible cues through vibration) has never been evaluated in teleoperation for robot operation. Finally, there is the tendency to consider TUIs functionalities as part of multimodal interfaces and, in particular, when combined to vision or speech. As for the latter aspect, we present a multimodal interface for semantic knowledge acquisition in Chapter 7.

Finally, no comparisons have been conducted for high-level interaction paradigms with TUIs, such as gesture recognition or object selection.

3.4   Semantic Knowledge in Robotics

Hertzberg and Saffiotti (2008) characterize semantic knowledge in robotics by two properties: *(i)* the need for an *explicit representation* of knowledge inside the robot; and *(ii)* the need for *grounding the symbols* used in this representation in real physical objects, parameters, and events. Many robotic systems nowadays embody some sort of semantic knowledge, but it is often hard-coded in their implementation. This greatly reduces the possibility to reuse this knowledge, to reason on it, or to share it with other entities (e.g. robots, hardware devices, or humans). In the rest of this section we address works that provide an explicit representation to boost up robotic systems. At first, we focus on knowledge exploitation in human-robot interaction and then, we move towards applications in autonomous robots.

3.4.1   Semantic Knowledge in Human-Robot Interaction

Semantic knowledge not only improves autonomous robot performance, but is also an effective approach in human-robot interaction, particularly when the presence of humans is fundamental. There are several reasons supporting this claim:

- semantic knowledge is based on symbolic representations, which are more comfortable for humans. A robotic interface with numerical data is

cognitively stressing for an operator, while symbolic representations are closer to our communication attitude;

- semantic knowledge is the result of an intense process from raw perceptions (coming from the robot sensors), to data representation, aggregation, and meaning extraction. This resembles humans' assessment process but, when accomplished by robots, is free of cost;

- semantic knowledge is not a mere data mapping onto whatever symbolic representation, rather it is a process which adds further knowledge to the information provided by data on their own, mostly thanks to automated reasoning mechanisms;

- semantic knowledge is not only a benefit for human awareness, but it also allows for natural robot operation; for example, through query languages or natural language processing, humans can communicate with any robotic system comfortably.

Semantic Mapping and Human Augmented Mapping

A robotic task where semantic knowledge is a valuable tool for HRI is *semantic mapping*. In fact, several approaches attempt to model the environment beyond quantitative models, such as metric or topological maps. Concepts like rooms, corridors, surfaces are easier for the user to interact with a mobile robot. Think about the possibility for a human operator to command a robot *"go to the kitchen!"*, instead of *"go to x = 10, y = 200!"*. Nüchter et al. (2005) employ environmental knowledge by using geometric information to establish correspondences with data, providing for a more reliable and fast process. In the work by Galindo et al. (2005), environmental knowledge is represented by augmenting a topological map (extracted with fuzzy morphological operators) with semantic knowledge using anchoring. Environmental knowledge is also used in Diosi et al. (2005), where an interactive procedure and a watershed segmentation are employed to create a contextual topological map. Another work addressing the use of environmental knowledge is by Martínez Mozos and Burgard (2006), in which a context-based topological map is extracted from a metric one using AdaBoost. A different approach to semantic mapping and semantic knowledge is to put the human *in the middle* of the acquisition and meaning extraction process, cooperating with the robotic system in the grounding process. This approach is known as *human augmented mapping*, to indicate the active role that human-robot interaction plays in the robot acquisition of qualitative spatial knowledge. For example, Kruijff et al. (2006) introduce a system to improve the mapping process by interacting with the

robot (using natural language). However, when related to human-robot interaction, still few approaches in robotics adopt a symbolic representation. Related to human augmented mapping is our semantic-driven tangible interface, proposed in Section 7, whose output is a set of concepts that can be overlaid on a metric or topological map.

Semantic Knowledge for Robot Control

Closely related to mapping, is the problem of controlling robots with natural commands. Humans tend to explain how to reach places using *spatial relations*, which in turn can be codified with symbolic representations or through natural language. Marciniak and Vetulani (2002) design a natural language interface for a mobile robot, whose knowledge base contains spatial information describing static situations and actions, such as projective relations (e.g. *behind, in_front_of, left_of, right_of*) and relative distance (e.g. *far, close, in*). Skubic et al. (2004) investigate the use of spatial relations to ease human-robot communication. Such information is extracted by grid map, and adopted for human feedback and for human-robot commands. Semantic interpretations are stored and organized as *context predicates*, a logical form similar to propositional logic. Finally, Theobalt et al. (2002) propose a combined system where information flows in two directions. First, the navigation system supplies landmark information from the cognitive map used for the interpretation of the user's utterances in the dialogue system. Second, the semantic content of utterances analyzed by the dialogue system is used to adjust probabilities about the robot position in the navigation system. The knowledge representation in this case is *first-order logic*. Kruijff et al. (2007) present ontology-based approach to multi-layered conceptual spatial mapping that provides a common ground for human-robot dialogue. It is thus possible to establish references to spatial areas in a situated dialogue between a human and a robot about their environment. Their robotic system is based on an OWL ontology.

Semantic knowledge is applied to other HRI applications as well. Hasanuzzaman et al. (2007) define a frame-based knowledge model for person-centric gesture interpretation. The knowledge-based management system, SPAK, acquires knowledge from different software agents (e.g. gesture recognizer) and, through reasoning, determines the actions to be taken and submits the corresponding commands to the target robot control. Modayil and Kuipers (2004) propose an unsupervised learning method based on allocentric occupancy grids to identify and build an object ontology, but without any explicit symbolic representation. Loutfi et al. (2008) validate how a knowledge representation and reasoning system (KRR) can improve anchoring and human-robot interaction based tasks. Moreover, this is one of the few attempts in robotics to reuse

generic knowledge, since they adopt a general purpose upper level ontology, DOLCE (A Descriptive Ontology for Linguistic and Cognitive Engineering), instead of a content dependent one. This effort is significant, since it enhances the integration between robotic systems and other devices and fosters their deployment in everyday activities.

3.4.2 Context-based Knowledge in Autonomous Systems

Semantic knowledge enhances autonomous robot skills through reasoning mechanisms, which allow the robot for acquiring further knowledge and solving complex situations. In particular, semantic knowledge representing contextual information is useful for accomplishing manifold autonomous tasks. Turner (1998) defines contextual knowledge as: "any identifiable configuration of environmental, mission-related, and agent-related features that has predictive power for behavior".

Contextual knowledge can be used in robot mapping for selecting, possibly in a dynamic way, ad-hoc methods. As an example, Hahnel et al. (2002) propose a mapping technique for populated environments (e.g. public spots, such as stations, airports, and so on) in which a probabilistic method to track people is used to improve the mapping process. Tuning general techniques is another interesting option: for example, Montemerlo et al. (2002) improve SLAM techniques by using environmental knowledge to select the features. Grisetti et al. (2006) define three different phases in robot mapping algorithms, namely *exploration, localization* and *loop closure*. They propose to use introspective knowledge, by detecting those phases and by tuning the computation accordingly. Newman et al. (2006) exploit introspective and environmental knowledge by using two different algorithms for incremental mapping and loop closure: efficient incremental 3D scan matching is used when mapping open loop situations, while a vision based system detects possible loop closures. Wolf and Sukhatme (2008) develop techniques to build maps that represent activity and navigability of the environment. Their approach to semantic mapping is to combine machine learning techniques (e.g. HMM or SVM) with standard mapping algorithms. Still, the knowledge is embedded in the system without an explicit representation in most of these approaches.

In robot navigation and exploration, the use of contextual knowledge can help in the specialization process of general techniques to the problem at hand. For example, Coelho Jr et al. (1998) try to learn the most efficient navigation policies together by inferring environmental knowledge from system dynamics in response to robot motion actions. When phrasing search and exploration as a multi-objective task, mission related knowledge can change the relative importance of one kind of sub-goal with respect to the other ones. For example,

Calisi et al. (2007) highlight that search and exploration requires a choice among, often conflicting, sub-goals as exploration of unknown areas and search for features in known areas. Also, coordinated search and exploration can be improved using contextual knowledge. Stachniss et al. (2006) propose a coordination algorithm which takes into account environmental knowledge by using contextual knowledge (e.g., corridors, rooms). Sending at least one robot to explore the corridor allows to discover the structure of the environment and thus to enhance coordination with the other robots. However, this approach is focused on indoor structured scenarios and it does not apply to unstructured environmental elements, which typical of the field applications that we consider in this thesis.

The use of contextual knowledge has a long tradition in vision, both from a cognitive perspective and from an engineering perspective. Indeed, also robot perception can benefit significantly from contextual knowledge. Moreover, it is through the sensing capabilities of the robot that environmental knowledge can be acquired. In robot perception, normally, an iterative knowledge process occurs: a top-down analysis, in which the contribution given by the environmental and mission related knowledge helps the perception of features and objects in the scene; a bottom-up analysis, in which scene understanding increases the environmental knowledge. Torralba et al. (2003) use environmental knowledge, extracted from low-dimensional global images, to perform robust place recognition, categorization of novel places, and object priming. An example of the use of visual information to increase environmental knowledge is provided by Kamon et al. (1996), where a learning technique based on grasp trials is used to choose grasping points by considering the geometry of the object to grasp.

Finally, another relevant use of contextual knowledge is related to the design of basic behaviors, where it can be used for the fine tuning of the parameters. The use of contextual knowledge for behavior specialization is suggested by Beetz et al. (2001), where environmental and introspective knowledge is used to obtain smooth transitions between behaviors, by applying sampling-based inference methods. A more in rescue related task is the design of effective behaviors in rough terrains that has been pursued by exploiting terrain classification, for example the one by Triebel et al. (2006). Usually, in these cases, ad-hoc representations, such as behavior maps by Dornhege and Kleiner (2007), are used for representing features like the presence of ramps or open stairs. Nevertheless, this type of contextual knowledge can clearly be viewed as environmental knowledge and can be used to select or tune behaviors.

Despite the effort in leveraging the information level, most of these approaches cannot be considered as related to semantic knowledge, because they lack of an explicit representation. Even worse, these approaches are related to specific robotic components; for example, we outlined works related to

mapping, exploration, vision, behaviors, and so on. None of them considers semantic knowledge representation and management as an orthogonal aspect independent from its application, rather it is hard-coded in specific modules. Needless to say, this does not foster knowledge re-use and sharing across the whole robotic systems, or among different robots. This is why in Chapter 8 we address this issue presenting a generic solution to the problem of knowledge representation and management, independent from robotic modules and from the adopted explicit formalism.

3.5 Discussion

In this section, we extract relevant aspects coming out from the aforementioned related work that motivate our thesis contributions, which will be extensively discussed in the following chapters.

Tangible User Interfaces

Most of the previous research about human-robot awareness is focused on the design of effective human-robot interfaces, often based on information visualization best practices or enhanced view representations. However, as addressed in Section 3.2, robot interfaces are still conceived as applications for robotic experts and based on traditional interaction metaphors. On the other hand, nowadays innovative technologies are available, such as tangible interfaces, which exhibit novel interaction paradigms, whose benefits may be advantageous for different human-robot tasks, such as robot operation. Through the study of existent implementations and experimental evaluations addressing TUIs, we point out some interesting issues :

- a formal assessment of the role of tangible interfaces in robotic applications seems still missing, in particular there is the need to identify the most effective tangible interaction paradigm for each robot operation style;

- different experimental evaluations compare tangible interfaces with other interfaces for manual robot teleoperation, but they often do not apply any rigorous statistical analysis and do not address tactile feedback;

- tangible user interfaces are situated in the real world, hence intra-scenario operator mobility is an advantage that needs to be exploited and must be compared with respect to remote approaches;

- TUIs exhibit interesting functionalities through high-level interaction paradigms, beside acting as mere robot controllers, since they boost humans' innate attitude for gesturing or pointing;

- TUIs appear to be more effective if supported by other interaction means, such as a video feedback or speech recognition, in multimodal interfaces.

This motivates our research about tangible interfaces to move along a path from low to high level tangible interaction paradigms. As for low-level metaphors, we focus on an extensive and rigorous assessment of TUIs; concerning high-level interaction, we enable novel functionalities for TUIs devices: from acting as robot controllers to assuming the role of *semantic-driven tangible interfaces*.

Semantic knowledge

Concerning the use of semantic knowledge in human-robot systems, there has already been the tendency to put humans in the middle of the acquisition process. However, our feeling is that tangible interfaces acting as semantic knowledge acquisition tools have never been considered, and this lack motivates our aim to create a synergy between TUIs and semantic knowledge. In particular, we underline the effectiveness of an assessment methodology that allows users to acquire knowledge by a direct interaction with objects in the environment where they are situated, with a portable interface, and adopting an innate interaction style.

Despite several robotic systems attempts to produce some form of high-level knowledge, just few implementations represent this knowledge through any explicit representation. Furthermore, it is also important to design solutions that do not require the massive re-engineering of existent systems, and that fosters knowledge re-use and sharing among different robotic components. It is what we aim to address through context-based architectures. Finally, we already illustrated the manifold advantages of semantic knowledge: from the possibility to adopt reasoning mechanism, to knowledge reuse and robot performance boosting.

Part II

Tangible Interfaces for Robot Operation

Chapter 4

Teleoperation, Feedback, and Mobility

4.1 Introduction

Part II covers the first contribution of this work. We define here our research question for the next two chapters: *how can we exploit tangible interfaces to enhance human-robot awareness by lowering the operator's cognitive effort in robot operation?*

Our approach to this problem is twofold. On the one hand, we aim at defining tangible-based algorithms for robot control. On the other hand, we conduct extensive experiments to assess those characteristics of tangible user interfaces that can be effective for robot operation. As already mentioned in Section 2.3, TUIs exhibit several interaction paradigms that leverage different cognitive aspects. Throughout this part, we will follow a research path from low to high-level interaction metaphors, and this chapter covers the first steps of this conceptual framework: grasping and background tangible processes, which are respectively adopted for robot teleoperation control and environment tangible feedback. Furthermore, we also investigate a relevant aspect for robot control: intra-scenario operator mobility. Recalling the tangible interaction framework reported in Section 2.3.4, our interest in this chapter focuses on two characteristics of TUIs: *tangible manipulation* and *expressive representation.*

Robot teleoperation has been defined in Section 2.4 as a manual, continuous, and low-level control of one robot by a human, acting as an *operator*,

through any type of *input device*. Background tangible processes concern humans' unconscious perception of environmental background information, as reported in Section 2.3.3. Finally, operator mobility has been covered in Section 3.2, introducing PDA interfaces that enhance HRI awareness.

The discussion of related work about TUIs for robot teleoperation (see Section 3.5) highlighted some issues, here recalled for ease of presentation:

1. a formal assessment of the role of tangible interfaces in robotic applications seems still missing, in particular there is the need to identify the most effective tangible interaction paradigm for each robot operation style;

2. different experimental evaluations compare tangible interfaces with other interfaces for manual robot teleoperation, but they often do not apply any rigorous statistical analysis and do not address tactile feedback;

3. tangible user interfaces are situated in the real world, hence intra-scenario operator mobility is an advantage that needs to be exploited and must be compared with respect to remote approaches;

In Section 4.2, we propose a tangible interface for robot teleoperation that relies on accelerometer sensors for motion detection, defining a novel and comfortable human-robot interface. Section 4.3 introduces environmental background information relevant for robot teleoperation. Furthermore, we present a system that improves human-robot awareness through a tangible device equipped with a rumble motor. In Section 4.4, we characterize the operator mobility within the environment, its benefits and drawbacks with respect to remote approaches. Section 4.5 establishes our research questions and preliminary hypotheses, that are essential to compare tangible interfaces with respect to traditional teleoperation approaches. An extensive experimental evaluation, reported in Section 4.6, validates the effectiveness of these approaches for human-robot awareness, provides answers to the aforementioned research questions, and characterizes TUIs for robot teleoperation according to an *evaluation framework* here presented.

4.2   Robot Teleoperation and Motion Detection

When controlling a robot with a keyboard, the key pressure determines the robot movement, either linear or angular. A joystick achieves the same effect through the inclination of the stick. Every input device exhibits its own interaction style. How can a tangible interface perform low-level teleoperation?

We recall that we are focusing only on portable tangible interfaces. That is, we are not considering interactive surfaces, nor modular interfaces. We

deal with hand-held and graspable devices, which position and attitude in the environment can be altered by the operator through body gestures (typically, arms or hands). **Our solution exploits the device attitude as physical representation, and map relevant configurations onto digital information**. In particular, we map this information onto robot linear speed and jog. We rely on two classes of approaches:

- measuring changes in the device position and attitude, a problem known as **motion detection**;

- estimating the device position and attitude according to a reference frame, a problem known as **localization**.

Our system has been implemented using a particular tangible device: a Nintendo **Wiimote** controller. However, without loss of generality, the same principles can be adopted to other types of portable tangible interface. The Wiimote is a commercial-off-the-shelf (COTS) controller that is cheap, robust and with a comfortable design. Furthermore, it embeds different sensors: accelerometers, IR camera, rumble, leds, multiple buttons, internal speaker, and a cross-pad. However, its 3D spatial capabilities are error-prone and not comparable to professional equipment.

Attitude Estimation through Accelerometers

Given whatever portable tangible interface equipped with accelerometer sensors, we provide some preliminary definitions.

Definition 5 (Body-fixed Reference Frame)

Let B be a device equipped with accelerometer sensors. Then, F_B, denoted as body-fixed reference frame, is a right-handed Cartesian coordinate system with origin in B. ∎

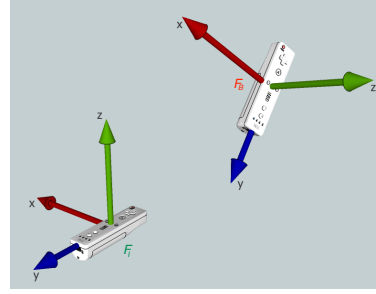In our case, the Wiimote-fixed reference frame is illustrated in Figure 4.1(a).

Definition 6 (Accelerometer Free Fall Reference Frame)

Let B be a device equipped with accelerometer sensors. Then, F_O is a local free fall reference frame relative to the accelerometers of B. ∎

Accelerometers actually measure the *proper linear acceleration* of the weight experienced by a test mass according to F_O, that is, as if the test mass were falling. The difference between F_B and F_O can be understood with an example. Imagine to lay the device face up on a table. Then, according to the free fall reference frame F_O, the acceleration reported is zero, since the test mass is
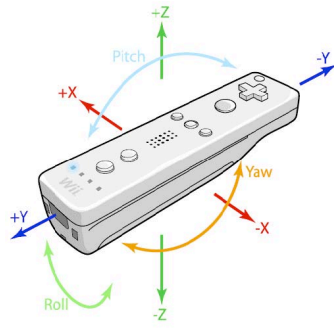
(a) The body-fixed reference frame of the Wiimote, F_B. It is worth noting that y axis is positive inwards, and x axis is positive when pointing left.

(b) 3D reconstruction of the body-fixed reference frame, F_B, with respect to the initial inertial reference frame, F_I.

Figure 4.1: Relevant reference frames for motion detection and localization.

falling in a free fall reference frame, hence no additional force is applied on it. On the other hand, according to the reference frame F_B, the acceleration reported along the z-axis is $|g|$, since it takes into account the force exhibited to counterbalance the gravity force on the test mass.

Definition 7 (Initial Inertial Reference Frame)

Let B be a device equipped with accelerometer sensors. Then, F_I is an inertial reference frame, which origin is the initial device position, and whose axes are congruent with the body-fixed axes when the device is still in its initial attitude.   ∎

The relationship between F_B and F_I is reported in Figure 4.1(b). Let $\vec{a}(t)$ be the acceleration reported at time t with respect to the body-fixed frame F_B. It is composed of two components: the actual device acceleration $\vec{a'}(t)$, and the gravity acceleration vector $\vec{g}$. Let $\theta(t)$ be the device pitch at time t. Then, $\theta(t)$ is given by (see Figure 4.2):

$$
\begin{aligned}
a_y(t) &= \cos(\theta(t) + 90)g_z \\
&= \arcsin(\theta(t))g_z
\end{aligned}
$$

from which

$$
\theta(t) = \arcsin\left(\frac{a_y(t)}{g_z}\right) \tag{4.1}
$$

Similarly, we can retrieve roll, defined as ψ in the following way:

$$\psi(t) = -\arctan\left(\frac{a_x(t)/g_z}{a_z(t)/g_z}\right) \qquad (4.2)$$

Unfortunately, accelerometers suffer from several limits: *(i)* accelerometers do not allow to retrieve the yaw angle, and *(ii)* meanwhile the device is moving, it is difficult to distinguish the gravity vector from the actual device acceleration. Thereby, they do not allow to retrieve a full spatial localization, in terms of position and attitude, of the tangible interface.
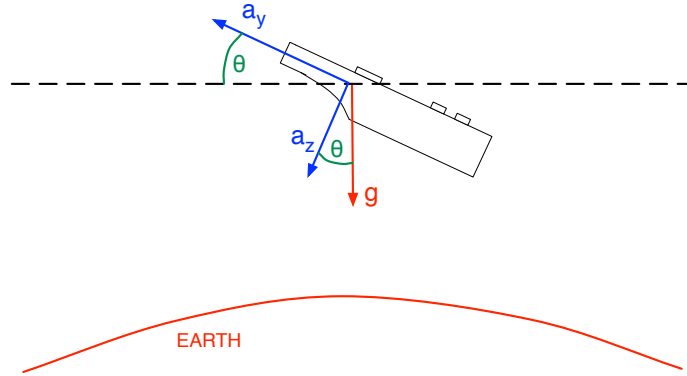
Figure 4.2: A sketch representing how to retrieve the device pitch from accelerometers.

Physical Representation for Robot Teleoperation

Once retrieved pitch and roll, we define a physical representation. At time t, the operator's arm affects the device status, $x_d(t)$, defined in terms of its roll and pitch:

$$x_d(t) \stackrel{def}{=} \langle\psi(t), \theta(t)\rangle$$

Analogously, let $x_r(t)$ be the robot status at time t, defined in terms of its linear speed ν and jog ϕ:

$$x_r(t) \stackrel{def}{=} \langle\nu(t), \phi(t)\rangle$$

Finally, we establish the following *representation function* f: *(i)* the device pitch is associated to the *robot linear speed*, and *(ii)* the device roll is mapped onto the *robot jog*:

$$x_r(t) = f(x_d(t)) = f(\psi(t), \theta(t))$$

We designed an intuitive mapping for the human operator. Rolling left (right) the device corresponds to a left (right) robot jog, as if we were driving with a wheel, while pitching down (up) the device increases (decreases) the robot linear speed, as acting on a stick. An example of a operator controlling a wheeled robot with our interface is shown in Figure 4.3. Our feeling is that such an interface reduces the operator's cognitive stress with respect to traditional approaches, and enhance human-robot awareness.

Figure 4.3: An operator controlling a wheeled robot with our tangible interface for teleoperation. Pitching down (up) the Wiimote device increases (decreases) the robot linear speed, while rolling left (right) the device corresponds to turning left (right) the robot.

4.3   Tactile Feedback

Robot teleoperation is a cognitively fatiguing activity, especially if conducted by a remote site. The operator is totally focused on controlling the robot, hence she does not consider manifold environmental details that are relevant for a proper assessment. Furthermore, the lack of perception due to the robot sensors or the interface limitations further degrades this process. Despite this, a tangible device equipped with a rumble feature can notify the operator about unexpected events by vibrating, **thus moving background information in the foreground** through tangible cues.

We embed in our teleoperation interface two different pre-attentive techniques, based on tactile feedback with a rumble motor:

- **Asynchronous Feedback**: vibration is activated in presence of unexpected events, to signal the operator a problem. An example is to activate the rumble if the robot is very close to an obstacle, or the platform

battery is running out. Unexpected events can be considered according to a broad spectrum of factors: *environment-dependent*, *mission-dependent*, or *proprioceptive.*

- **Synchronous Feedback**: vibration is attached to an environmental factor, which feedback is continuously provided to the operator. We clarify this through an example. When we drive a car, the wheel vibration reflects the asphalt roughness of the road. The more it is damaged, the more our wheels vibrate, and we are aware of this. Synchronous feedback provides the same functionality to the robot operator. Vibration can be modulated in intensity and frequency, in order to provide the roughness level of the ground where the robot is deployed. The drawback of synchronous feedback approaches is that they can frustrate the operator and they can loose effectiveness over time. Hence, we paid attention in the design of a non-intrusive, yet effective, rumble mechanism.

4.4 Operator Mobility

As already addressed in Section 3.2.2, intra-scenario operator mobility enables the direct acquisition of situational awareness in robot teleoperation, as the operator has direct access to the environment and, in some situations, to the robot. Needless to say, this is considered as an advantage, and tangible user interfaces, as already stated in Section 2.3.1, foster this approach, since they exhibit interaction paradigms that move humans back into the real world.

Effectiveness in operator mobility has been previously evaluated comparing PDA-based mobile interfaces with respect to remote desktop interfaces. However, these results can not be transferred to TUIs. First, PDA interfaces adopt different input technologies. They are mostly based on virtual keyboards, touch screens, or virtual arrows, projected on the PDA display. On the other hand, TUI exhibit a wide range of novel control paradigms, like our teleoperation interface proposed in Section 4.2. Second, both the approaches allow the operator to move within the environment and directly access the robot and its surroundings. However, PDA interfaces require a continuous cognitive switch between looking at the robot and looking at the PDA display. TUIs remove this constraint adopting a physical representation, gesturing, that does not distract the operator. Third, TUIs notify unexpected background information, through tactile feedback (see Section 4.3), which is not a relevant aspect in PDA-based interfaces.

Thereby, we aim at evaluating awareness when leveraging intra-scenario operator mobility with a TUI teleoperation interface. Even more, we investigate whether such a configuration can counterbalance the computational lim-

itations of TUIs with respect to the powerful graphic functionalities exhibited by desktop interfaces that are deployed on a remote site.

4.5   Preliminary Hypotheses

Our interface is a good representative of the whole class of portable tangible interfaces adopted for robot teleoperation. In fact, the same effect could be achieved with any other hand-held device equipped with accelerometers and a small rumble motor. Thereby, our aim is at analyzing its performance as a mere controller, with respect to traditional approaches (e.g. keyboards, joysticks, and so on). In this section, we formalize our research questions, while in Section 4.6 we describe an extensive experimental evaluation conducted to answer these questions. We also introduce an **evaluation framework**, represented as a four dimensional space, whose axes are: *mission-related performance*, *environment conditions*, *robot operation degree*, and *operator interaction comfort*. This framework takes inspiration from the different taxonomies proposed by the HRI community, and is specialized for robot teleoperation. It is generic enough to be applied in the evaluation of other input devices.

Our first research problem consists of *evaluating the effectiveness of tangible user interfaces in a remote robot teleoperation task, with respect to traditional interfaces, and under different robot mobility conditions within the environment.* We evaluate effectiveness through two metrics: navigation time and number of collisions. Consequently, we formulate four preliminary hypotheses:

PH1. *When adopting TUIs equipped with motion sensing, subjects may experience lower navigation times with respect to traditional interfaces, given any mobility condition,* because motion sensing guarantees a natural control, without any additional cognitive load due to the control mapping memorization;

PH2. *When adopting TUIs equipped with tactile feedback (through device vibration), subjects may experience a lower number of collisions with respect to traditional interfaces, given any terrain difficulty condition,* since in remote teleoperation tactile feedback boosts up the situation awareness provided by graphic interfaces;

PH3. *When increasing the terrain difficulty factor within the environment, TUIs may guarantee a lower navigation time increase with respect to traditional interfaces;*

PH4. *When increasing the terrain difficulty factor within the environment, TUIs may guarantee less collisions than traditional interfaces.*

Our second research question addresses the following issue: *evaluating the benefits of operator mobility and tactile feedback for real robot teleoperation, as well as estimating how much they can counterbalance the lack of a powerful graphic user interface.* The following preliminary hypotheses are then defined:

PH5. *When the operator directly accesses the environment, subjects may experience lower navigation times with respect to a remote control aided by powerful graphical interfaces*, since a better situation awareness is achieved.

PH6. *When the operator directly accesses the environment, and tactile feedback is provided, subjects may experience a lower number of collisions with respect to a remote control aided by powerful graphical interfaces*, since tactile feedback provides a pre-attentive mechanism to avoid collisions, even when unnoticed by the operator.

4.6 Experimental Evaluation

With the aim of validating our preliminary hypotheses, we organized a set of experiments for a comparative study of several interfaces for teleoperation[1]. In particular, we address the problem of robot teleoperation in urban search and rescue (USAR) scenarios. USAR consists of deploying a robot team within a disaster environment, to explore it and look for possible survivors. It is one of the most compelling robotic tasks, both for autonomous systems, and for the remote operator controlling the robotic team. The experimental analysis is divided into two parts: (*i*) a remote robot teleoperation in a virtual scenario, and (*ii*) a real robot teleoperation under different operator mobility conditions, performed in a real indoor scenario. There are two main distinctive features about these experiments: we conducted extensive runs, and paid attention to the realism of the setting. As for the former aspect, more than twenty hours of experiments (almost one hour per subject) have been performed, to measure the subject fatigue and relate this to a conceivable performance decay. Concerning the second aspect, the virtual scenario reproduces realistic physics, comprehensive of wireless communications constraints, robot turnovers, and sensor errors.

---

[1]Additional material (images, recordings and questionnaires) are provided at http://www.dis.uniroma1.it/ randelli/index.php?id=9.

Subjects

We tested 21 (3 females and 18 males) subjects taken from a class of students of Computer Science and AI, most of them with no experience of robot navigation. Ages ranged from 21 to 39 (mean $\approx$ 25.38, std $\approx$ 3.73). Three of them were left-handed. Being involved in computer science, all the participants are skilled in the use of computer, and are used to the keyboard interface. Nine of them have already participated to robotic projects, and out of these, only two have previously teleoperated a real robot. Finally, 14 participants declared not to have any previous experience with any Wiimote device, four to use it on a monthly basis, and just three weekly. Participants attended a preliminary training session to get basic confidence with each interface[2].

Interfaces and System Architecture

Robotic interfaces are different in the set of provided commands and, more in general, in the interaction paradigm they implement. We already introduced our tangible system. In our investigation, we also consider two representatives of traditional interfaces:

Joypad A joypad handled with one hand, which embeds a direction control cross-pad, accessible with the thumb (Figure 4.4(a)). Two extra buttons are used to reset the linear speed and the angular speed.

Keyboard A traditional four-arrows keypad with two extra buttons for stopping the linear speed and the angular speed (Figure 4.4(b)).

Concerning the joypad and the keyboard, the control technique is incremental, that is each command (button or key pressure) changes the desired robot speed or jog by one step. It is worth noting that while the keyboard is not portable, the tangible interface and the joypad allow for operator mobility. Due to the high number of runs required for each participant, we limited our comparison to these three control devices, which well represent established interaction styles. As a future work, it will be interesting to embrace other common interfaces, such as a joystick.

As well as the proposed controllers, the equipment adopted for the experiments consists of additional software and hardware support. The robotic architecture is based on our open source robotic framework OpenRDK[3], developed within the Ro.Co.Co. Laboratory[4] by Calisi et al. (2008a).

---

[2]As for training, subjects could try each interface for five minutes in a simulated open-space environment, before running the two experiments.
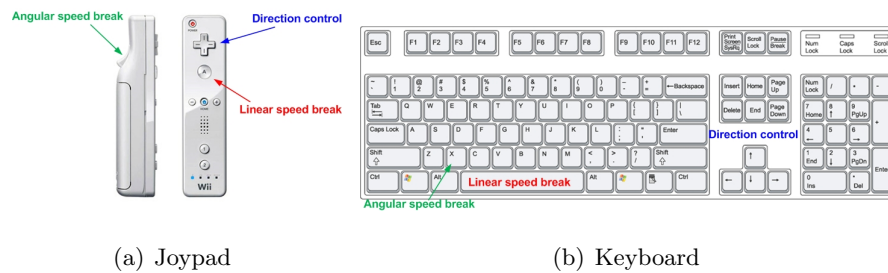
[3]http://openrdk.sf.net/

[4]http://labrococo.dis.uniroma1.it

(a) Joypad   (b) Keyboard

Figure 4.4: Overview of the implemented interfaces with their command configurations.

4.6.1 Remote Robot Teleoperation

Procedure and Task

The first case of study is a path-following task in a simulated rescue environment. The experiment scenario is implemented in the 3D robot simulator USARSim[5], developed by Balakirsky et al. (2006) at NIST. USARSim is currently adopted in the RoboCup Rescue Virtual Robots competitions, since it guarantees relevant characteristics, such as: a high degree of realism, through complex physic simulation, and full configurability, providing a wide spectrum of elements for rescue world models. The environment in the simulation is divided into three sections: *easy*, *medium* and *hard*, to simulate different fatigue conditions during the robot control. The pathway goes through a different degree of navigation difficulty, by including obstacles, cluttered areas and tricky passages the robot might encounter on its way. Subjects are required to guide the robot through the path. A run is considered failed when the robot crashes, flips over, or gets stuck. Participants must perform one run per each interface type: *Wiimote* with motion sensing and tactile feedback, *joypad*, and *keyboard*. The data gathered from this experiment are navigation time, number of collisions, robot pose and jog, and robot speed. Subjects fill a preliminary questionnaire and are interviewed after each run.

Design

We measure two dependent continuous quantitative variables: *navigation time* (measured in seconds) and *number of collisions*. As for the independent variables (IV), we consider *Interface Type* and *Terrain Difficulty*. Both the IVs

---

[5]http://usarsim.sf.net

are qualitative nominal variables and within-subjects factors, with three levels each: *Wiimote*, *Joypad*, *Keyboard* for the interface type and, *Easy*, *Medium*, *Hard* for the terrain difficulty factor. We adopt a 3x3 factorial design, controlling both the independent variables, for a total amount of nine possible treatments. We design a *repeated-measures* experiment. With this method, subjects test all the possible conditions, therefore the correlation between the different treatments is not affected by any difference due to subject skills. One drawback is the *crossover effect*: subjects tend to improve their knowledge about the experiment, since they perform it several times. This effect is counterbalanced through a *randomized* assignment of the treatment sequence.

Data Analysis

We perform three different analyses over the acquired data: evaluating the statistical difference of (*i*) the navigation time and (*ii*) the number of collisions with respect to the considered treatments, and (*iii*) a correlation study between navigation time and the number of collisions. As a result of robot turnovers or other unpredictable events (e.g. communication loss), which we deliberately chose to take into account for more realistic experiments, the dataset is affected by missing values and is unbalanced. To deal with these problems, we adopt a particular type of statistical analysis, known as *generalized linear mixed model*[6]. These models manage missing values, without dropping the whole case nor requiring any value imputation.

Navigation Time   To ensure that the normality assumption of GLMM is met, we apply a Shapiro-Wilk normality test against the collected data. Out of the nine treatments, four are not normally distributed. Dataset normalization was achieved through a logarithmic transformation and a further outlier elicitation (using both the z-score method for outliers and the Grubb's Test). We then apply the linear mixed model, addressing three fixed main effects: *Interface*, *Terrain Difficulty*, and *Interface x Terrain Difficulty*. The Type III Test of fixed effects shows that all the three main effects are statistically significant, as reported in Table 4.1. We further expand this analysis, through an analysis of the estimates of the fixed effects and their relative estimated marginal means:

- the *Interface* effect contrasts reveal how the Wiimote is significantly better than both the joypad ($p = 0.0, t = 5.922$) and the keyboard ($p = 0.0, t = 4.106$), and how the keyboard is better than the joypad ($p = 0.004$);

---

[6]An introduction to this statistical model is provided by West et al. (2006).

Fixed Effect	F	Sig.
Interface	19.208	0.0
Terrain D.	34.626	0.0
Interface x **Terrain D.**	3.696	0.01

Table 4.1: Type III test reveals that all the three effects are statistically significant.

- concerning the *Terrain Difficulty* effect, pairwise comparisons show that navigation times within an easy environment are significantly lower than those with medium terrain difficulty ($p = 0.025, t = 2.283$), which in turn are lower than times with hard mobility ($p = 0.001, t = 3.3401$);

- the interaction effect between *Interface* and *Terrain Difficulty* is significant mainly when comparing Wiimote and joypad performance for easy compared to medium mobility conditions ($p = 0.02, t = 2.576$)

In Figure 4.5 we report the interaction graph about the mean navigation times performed by the three interfaces throughout the different terrain difficulty conditions.

Collisions The outlier detection reduces the dataset to 19 samples. In this case, there is no preliminary data transformation, since the normality assumption for GLMM application is met. The same statistical method has been applied, again with three fixed main effects and no random effects. Table 4.2,

Fixed Effect	F	Sig.
Interface	8.298	0.001
Terrain D.	15.714	0.001
Interface x **Terrain D.**	0.098	0.907

Table 4.2: Type III test of fixed effects reports that only the **Interface** x **Terrain D.** is not statistically significant.
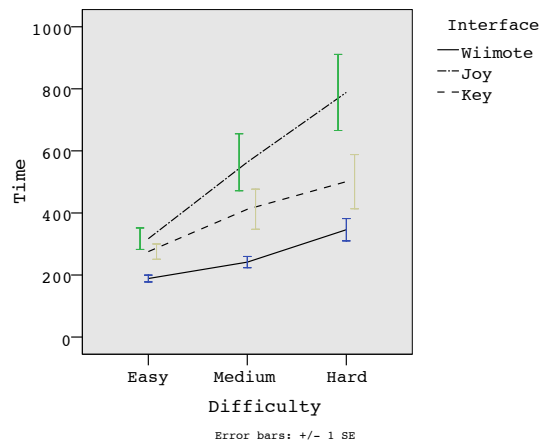
Figure 4.5: The interaction graph reporting the mean navigation time (in seconds).

reporting the fixed effects test, reveals how the *Interface* and *Terrain Difficulty* factors are statistically significant, whether *Interface x Terrain Difficulty* seems not to be. Moreover, we address the following considerations:

- the *Interface* effect contrasts show how both the Wiimote and the keyboard are significantly better, in terms of collisions, than the joypad ($p = 0.001$ and $p = 0.002$), while the Wiimote with tactile feedback is not statistically significant with respect to the keyboard ($p = 0.087$);

- comparisons among the levels of the *Terrain Difficulty* factor confirm that the number of collisions is significantly lower within easy environments, with respect to medium and hard ones ($p = 0.001$).

In Figure 4.6 we report the mean number of collisions for all the interfaces throughout the three mobility conditions.

Navigation Time - Collisions   Our last analysis clusters different teleoperation styles, to evince which are the most effective in terms of the operator's cognitive effort. For example, it can be interesting to assess whether subjects running the robot fast tend to accumulate a high number of collisions (or the converse), since this requires a significant level of intervention, which is stressing for the operator. As for our dataset, we cannot rely on any normality assumption. Therefore, we perform a *bivariate correlation* through the Spearman's correlation coefficient (see Figure 4.7). We conclude that there
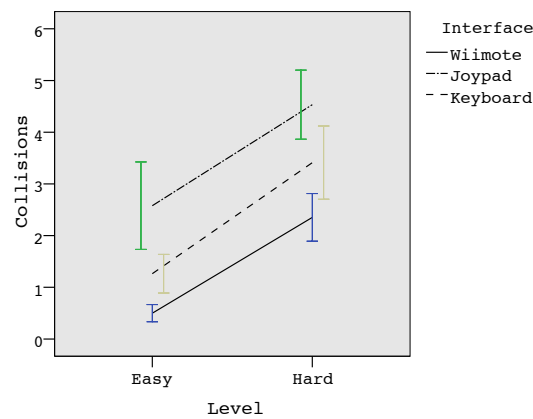
Figure 4.6: The interaction graph reporting the average number of robot collisions.

is a significant correlation between the two observed variables, and it is positive ($p = 0.002, r = 0.311$) That is, when navigation times increase, collisions increase as well.

Discussion

In this Section, we justify, from the cognitive point of view, the results provided by the data analysis. Furthermore, we also consider the feedback gathered through the questionnaires.

The tangible interface provides overall navigation times significantly lower than the other interfaces, as reported in Figure 4.8. It is also worth noting how it exhibits a tight standard deviation, with a small increase throughout all the terrain difficulty conditions. We argue this is due to two factors: a natural interaction and an intuitive control, due to the effectiveness of the physical representation adopted for the tangible interface. In fact, just few subjects had a previous experience with this interface. Nevertheless, even less skilled participants managed to limit the effort increase when facing hard conditions. This is further validated by the fact that 12 subjects selected the Wiimote as the best interface, seven preferred the keyboard, and only one the joypad.

Despite these results, **we do not claim that a tangible interface would always be the optimal solution in any robot teleoperation task**. In fact, as revealed by the analysis about the number of collisions, the Wiimote
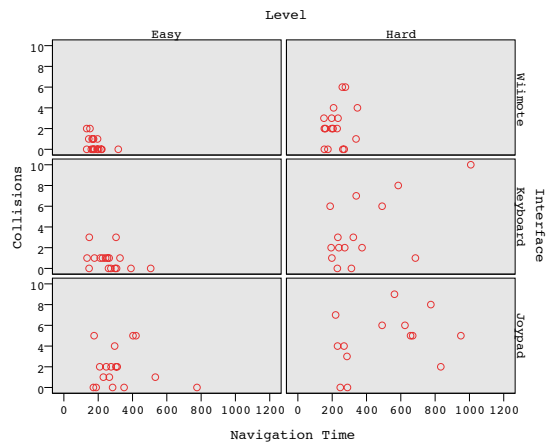
Figure 4.7: The scatter plot compares the navigation times (in seconds) with the number of robot collisions.
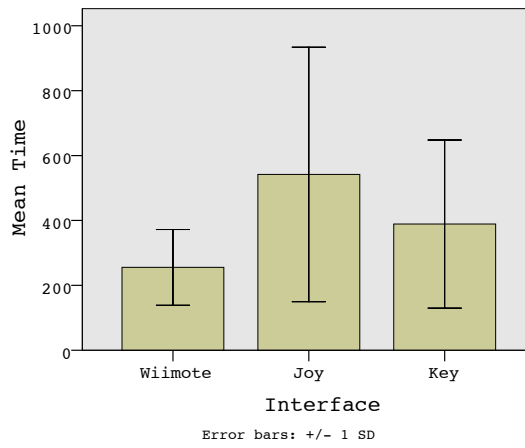
Figure 4.8: The bar chart reports the mean robot navigation time (in seconds).

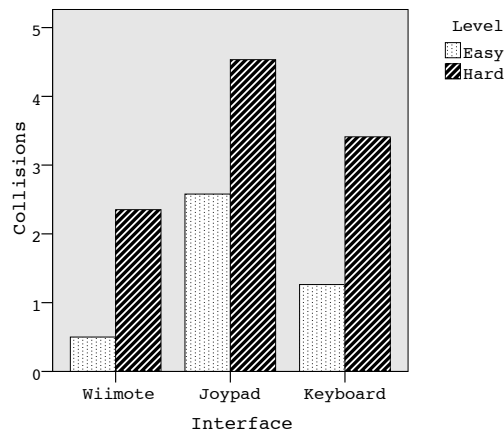versus keyboard difference is slightly significant. (see Figure 4.9). A justifi-

Figure 4.9: Bar chart showing the average number of robot collisions.

cation to this aspect seems to be the lower sensitivity of the keyboard with respect to the Wiimote interface. In fact, many subjects selected the keyboard as the best interface for controlling the robot in narrow spaces, whilst the Wiimote is too reactive for hard terrain conditions. This is further confirmed by the questionnaires, where subjects reported the following marks concerning the robot operation degree (ranged -3 to +3): 1.65 for the Wiimote, 1.5 for the keyboard, and 0.45 for the joypad. We argue this is due to the type of control. In fact, the keyboard implements an incremental control, resulting in more conservative and flexible movement in a constrained area. On the converse, the Wiimote provides continuous motion sensing[7]. Even more important, **using tactile feedback seems not to prevent collisions**. There is another drawback in the robot control with a tangible interfaces: subjects are not used to act within a 3D environment. Most of the common interfaces are somehow constrained just to a 2D manipulation space (e.g. moving a mouse on a surface), or 1D (e.g. pressing a button on the keyboard). While moving in a 3D space is very natural, yet precision is not easy to achieve. The Wiimote performs better in open spaces, where the error margin is greater, and the continuous control guarantees a more reactive control and lower navigation times.

As for the joypad, results showed it was the least effective interface, with a remarkable significant difference with respect to Wiimote and the keyboard.

---

[7]An incremental control would be unrealistic, since subjects should move their hand for every single increase or decrease.

Probably, this is due to two main reasons: the implemented control was incremental (like the keyboard), hence it was not so fast, and the learning rate was not so fast as for the Wiimote.

4.6.2   Operator Mobility

Procedure and Task

The second experiment is again a path-following task, like the first one, but this time it involves the teleoperation of a real robot. The scenario has been arranged within our department (see Figure 4.10), and it consists of two different parts, which are equivalent in terms of size and complexity. The choice to adopt different parts, even if comparable, minimizes the crossover effect in repeated-measures design. In the first section, subjects have a *full* access both to the environment and to the robot, that is, they can follow the robot during its path, even if not always with the same attitude of the robot. The second part is a remote teleoperation with *no visibility* nor access to the environment. Participants are supported by a powerful graphical desktop interface. Finally, we also considered a *partial* visibility section. In this case, the operator is still within the environment, but she cannot move. Hence, a proper view of the robot is not always guaranteed, because of possible occlusions. This partial visibility condition is helpful to assess *when* the operator mobility becomes no longer effective with respect to a powerful remote interface. However, due to network latency issues, data collected in this last part are biased, and will not be considered in our formal data analysis. Subjects are required to perform

Figure 4.10: A part of the real scenario set up for the real teleoperation experiment.

the full visibility section with a Wiimote, while the remote part with the keyboard. This assignment is due to several reasons. First, this experiment does not focus on any controller comparison, like the previous. Second, accessing the scenario with a keyboard would be unrealistic, and finally, due to the results in the first experiment, we discard the joypad interface. Data logging is the same of the previous experiment, and subjects have to fill a post-run questionnaire.

Design

We measure one dependent variable: *navigation time* (measured in seconds). As independent variable we consider the *operator visibility*. This is a qualitative nominal variable, which consists of three level: *Full Visibility*, *Partial Visibility*, *No Visibility*. We adopt a 3x1 *repeated-measures* design, that is, every participant performs all the visibility levels. Every run is set in a different environment (even if they are comparable for their extension), hence the *crossover effect* is minimized.

Data Analysis

Our analysis over the acquired data evaluates the statistical difference of navigation times with respect to the two considered visibility treatments. Since no robot stall nor flip has been reported, this time our dataset is balanced. We reduce the dataset to 18 samples through conventional outlier tests. Normality tests (we choose the Shapiro-Wilk test, because of the reduced set size) reveal that data collected during the remote runs are not normally distributed, hence we apply a non parametric test. Being a comparison between two related treatments of a within-subject factor, we select a Wilcoxon Signed-Rank Test. The analysis shows that navigation times performed with the Wiimote in the full visibility condition are significantly lower than those obtained in the remote run with the keyboard ($p = 0.006$, point probability $= 0.0$, $z = -2.678$). Mean navigation times for both the treatments are reported in Figure 4.11.

Discussion

As the data analysis of the real teleoperation experiment points out, **adopting tangible interfaces when the operator is co-located with the robot improves the overall performance**. From a cognitive point of view, this implies that the operator does not need any graphic interface, and the situation assessment acquired by her own perception is sufficient to accomplish the teleoperation task.
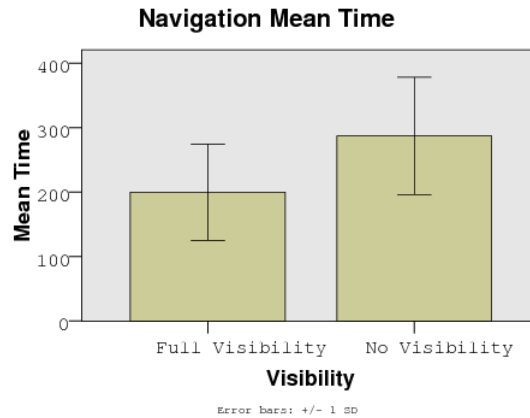
Figure 4.11: Bar chart reporting the mean navigation time (in seconds).

Nevertheless, it is interesting to correlate such a result with the visibility degree of the operator. In fact, there are at least two main factors that alter the operator visibility: (*i*) the distance between the operator and the robot, and (*ii*) occlusions which prevent a proper assessment of the area nearby the robot. *What is the maximum distance after which the operator access to the environment is not anymore an advantage? What is the operator behavior when she cannot have a complete assessment of the robot surroundings?* In order to answer these questions, it is worth taking into account the partial visibility condition, which allows us further considerations. Despite its validity from the point of view of a formal analysis, the partial visibility has been the worst section, in terms of navigation times. **Our feeling is that as soon as the robot is occluded or far, the operator performance degrades fast**. In fact, a partial view of the robot does not guarantee that the hidden part is free of obstacles. On the other hand, when the robot is too far, it is almost impossible to get a proper assessment of the trajectory of the path. Under these critical circumstances, the operator continuously switches her view between the robot and the graphical interface, which adds a relevant cognitive stress and leads to inattention. This slows down the task accomplishment and thrashes the overall performance.

4.7 Conclusions

We summarize the results evinced so far in our experimental analysis, according to the evaluation framework defined in Section 4.2, and focusing only on tangible interfaces:

1. *Mission-related performance.* Tangible interfaces are a valuable input approach for robot teleoperation, as addressed by their performance in terms of navigation times and number of collisions. However, tactile feedback does not significantly enhance the robot control in single-robot teleoperation. Nevertheless, it would be interesting to evaluate its potential benefits in a single-human / multi-robot (SHMR) paradigm.

2. *Environment conditions.* Best effectiveness is achieved in open spaces or semi-cluttered areas. The operator mobility and presence within the environment enhances the performance only under full visibility conditions, but is not robust with respect to the distance between the robot and the operator, nor to robot occlusions. Under these circumstances, a combination of direct visibility and remote interface is not fruitful and the best practice would be to adopt just a remote interface.

3. *Robot operation degree.* The main advantage of TUIs is their reactiveness, due to the continuous motion, which allows for a fast execution of complex movements with natural commands.

4. *Operator cognitive load and interaction comfort.* The learning rate is high, even for unskilled users, and the intuitive physical representation does not overload the operator cognitive effort, in particular when adopting manifold commands. This results in a lower effort increase when facing complex terrain conditions.

This experimental evaluation points out two main problems. First, TUIs are too sensitive for tight and cluttered areas. This problem can be solved through shared control paradigms. However, shared control cannot be well represented by simple grasping and motion detection, as in the case of teleoperation. Thereby, in the next chapter, we move towards high level tangible interaction metaphors, such as gesturing, and we introduce a novel approach to robot shared control. A second drawback is that acting in a 3D environment is somehow deceptive for precise control, as required by manual teleoperation. Thereby, 3D movements should be coupled with high-level commands, rather than low-level ones. Even this aspect is discussed in the next chapter.

Chapter 5

High-Level Tangible Interaction

5.1   Introduction

In Chapter 4 we addressed the problem of human-robot awareness by designing physical representations for tangible interfaces that lower the operator's cognitive effort. In particular, we focused on low-level interaction metaphors, mainly grasping and tactile feedback, through which we presented a robot teleoperation interface. An extensive experimental evaluation pointed out some drawbacks in this solution. First, robot teleoperation with tangible interfaces is not very effective when the robot is deployed in cluttered areas, since the interface is too reactive, due to its continuous control. Second, humans experience problems in performing precise movements in a 3D environment. This should not surprise the reader, as our world is full of examples that confirm this problem. In fact, every time a precise gesture is requested, humans tend to constraint their manipulation space to ease the process. For example, the wheel of a car acts on a 2D space. Under these circumstances, tangible interfaces do not provide significant benefits, which in turn does not allow the operator to focus on situation assessment.

Thereby, it is interesting to investigate high-level control styles, which do not require absolute precision, rather are based on discrete commands triggered by a user acting as a *supervisor*. As we will see, this in turn requires different tangible interaction paradigms. In fact, motion detection and grasping do not enable expressive representations for this type of control. This

chapter moves our research a step forward, covering high-level tangible interaction metaphors. Quoting Section 3.5, this time we address the following issue: "*TUIs exhibit interesting functionalities through high-level interaction paradigms, beside acting as mere robot controllers, since they boost humans' innate attitude for gesturing or pointing*". Moving away from grasping to gesturing we are addressing a relevant characteristic of TUIs, which is denoted in the tangible interaction framework (see Section 2.3.4) as *spatial interaction*. That is, we are focusing on the interaction itself, rather than on the object we manipulate.

Section 5.2 introduces the main limitations of low-level interaction paradigms when applied to shared control, and motivates the need to move towards high level metaphors, such as gesturing. In Section 5.3, we present a human-robot shared control interface based on gesture recognition. This approach exploits the humans' skill to move within the environment where they are situated. The result is that the operator triggers high-level commands that are autonomously executed by the robot. The operator interventions are only limited to robot failures, while most of the time she is dedicated to supervise the overall situation, which is a benefit for the situation assessment process. As in the case of motion detection, gestures are performed without any graphic interface, hence there is no cognitive attention switch between the interface and the robot. Finally, experimental evaluations and conclusions are provided at the end of this chapter.

5.2   Interaction Paradigms for Shared Control

For the sake of acquiring a proper awareness, the operator should not continuously switch between assessing the situation and controlling the robot. Thereby, the the operator is supported by a natural and comfortable control style, the more she will acquire a proper level of awareness. As already mentioned, grasping and object manipulation are not effective interaction paradigms when the robotic task gets more complex. This lowers the human performance in operating the robot, which in turn degrades human-robot awareness.

Let us motivate this with a case study. Imagine a dynamic structured indoor environment, like the floor of a building, composed of rooms, corridors, and doors. Our goal is to implement a robotic system that looks around and interacts with relevant landmarks, such as: objects, people, or environment spots. Examples of interactions are: moving towards a place in the environment, searching for an object, providing information to a person, and so on. In such a complex scenario, the operator cannot concurrently overview the scenario, control the robot, manage the different interactions, and deal with

unexpected events. Thereby, we specify two requirements:

- we aim at adopting a high-level robot operation policy, such as shared control, where a user specifies the target element, triggers a request (e.g. "go there"), and demands its execution to the robot, thus acting as *supervisor*;

- we provide the robot with some preliminary information, in terms of spatial hints.

Needless to say, a shared control technique lowers the cognitive effort in terms of operating the robot. However, this requires relevant autonomous skills, which is not always realistic. We believe that supporting a robot with spatial hints may boost up the system robustness, and lower the intervention rate of the supervisor in case of failures. For example, if you think of a robot in the middle of a corridor meanwhile searching for an object, then moving in the right direction can save a lot of time. This direction can be provided by the user as a spatial hint. Shared control relies on the fact that the robot is able to execute the supervisor's instructions, and to manage spatial hints as well. For example, we are assuming that the robot is provided with path planning functionalities to reach a place, and recognizes relevant concepts, such as: doors, crossings, and rooms. In this thesis, we only address the gesturing aspect, a not how the robot executes commands.

What would be an effective physical representation for shared control? We could reuse our previous approach based on motion detection. However, this paradigm is not enough expressive to deal with activities more complex than manual teleoperation. We provide an example that motivates such a claim. Let us suppose to instruct a robot to move towards a specific door. Only relying on the motion detection physical representation, that is, roll and pitch, we need to codify at least three types of information: identifying the desired door, asserting that it belongs to the concept of door, communicating the action "to move". Needless to say, it is unnatural and counter-intuitive to accomplish this by rolling or pitching the tangible interface. Thereby, as we raise the control style, we also need to leverage the interaction paradigm. We adopt **gesturing**, which is another natural tangible metaphor for humans. For example, gestures and speech are often integrated to explain how to reach a place (e.g. the utterance "go there" is typically associated with pointing a place with the arm).

This case study highlights one problem: *in order to acquire a proper awareness, it is required a physical representation that exploits a level of cognitive skills proportional to the task complexity*. **Our solution is a tangible gesture recognition system for robot shared control**. This passes through two steps:

1. implementing a gesture recognition system that relies on the perceptive capabilities of the tangible interface.

2. defining a meaningful physical representation through a **vocabulary** of gestures;

Once again, we adopt a Wiimote device equipped with accelerometers for our implementation and evaluation. As already mentioned in Section 4.2, this does not affect the generality of our solution.

5.3   A Gesture Recognition Robot Interface

Gesture Recognition through Accelerometers

We describe how to accomplish gesture recognition relying only on accelerometers. First, we provide our definition of *gesture*, that we will adopt throughout this section.

Definition 8 (Gesture)

Let $T \in \mathbb{Z}$ be a *sampling interval*. Then, a sample gesture G is a time-ordered sequence of accelerations $\vec{a}(kT)$, defined as:

$$((\vec{a}(kT)), 0 \leqslant k \leqslant n, n \in \mathbb{N}$$

According to this definition, the input of our recognition system is a gesture, represented as a sequence of timed triplets $(a_x, a_y, a_z)_T$, acquired with a sampling interval T. Accelerations greater in absolute value than the gravity acceleration are discarded, since they represent unreliable values. The acceleration components are pre-processed according to the calibration data and normalized with respect to the gravity acceleration value $|g|$ (see Figure 5.1. The first step of our system is to extract meaningful features from raw data. We identify eight features, collected in three groups:

- gesture displacement along the three axes (d_x, d_y, d_z), computed assuming a uniformly accelerated motion between two consecutive samples;

- gesture attitude in its final position n, in terms of pitch θ_n and roll ψ_n (computed as described in Section 4.2);

- gesture mean velocity along the three axes (v_x, v_y, v_z), computed assuming a uniformly accelerated motion between two consecutive samples.

Figure 5.1: Examples of accelerometer data representing some gestures.

Hence, the overall feature vector for a gesture G is $\Gamma_G = \langle d_x, d_y, d_z, \psi_n, \theta_n, v_x, v_y, v_z \rangle$. It is worth noting a couple of characteristics about this feature vector: *(i)* we tried to keep the vector size as small as possible to speed up the classification step, and *(ii)* we conducted a preliminary analysis to assess whether eliciting some features was possible, yet preserving the system performance. However, our results revealed that this set is minimal and cannot be further trimmed.

The feature vector represents the input of our recognition component, which is based on **support vector machines**. As described by Burges (1998), SVM is a partial case of kernel-based methods that maps feature vectors into higher-dimensional space using some kernel function, and then builds an optimal linear discriminating function in this space. The overall system is reported in Figure 5.2.
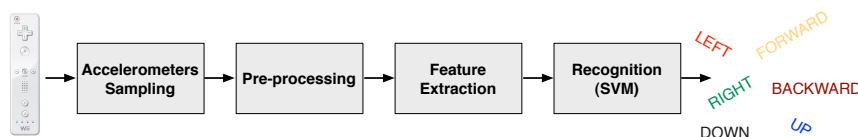
Figure 5.2: The architecture of our gesture recognition system. Once acquired a gesture, its accelerometer data are pre-processed and the feature vector is extracted. Then, a SVM machine learning algorithm classifies the gesture according to its feature vector.

Physical Representations for Shared Control

Under the hood of the physical representation adopted for our shared control system lay the same issues considered in the previous chapter for the robot teleoperation interface. That is, selecting a set of gestures as natural as possible, to allow for a seamless interaction. This time, we address arm gestures and, in particular, we define a set of six gestures, reported in Figure 5.3. This vocabulary has been applied to a specific aspect of our case study: providing spatial hints to our robot vision system, for landmark search and detection. This is quite an important aspect, since interacting with people or objects requires at first to detect them, and this process can block the system for a relevant amount of time. We report in Table 5.1 the six gestures, as well as their bound digital information. Since every machine learning technique is affected by a certain degree of error, the robot activates its speakers to play a recorded utterance corresponding to the detected gesture, hence the operator

Gesture	Digital Information	Behavior
Up	The landmark is higher	Tilt up the pan/tilt unit
Down	The landmark is lower	Tilt down the pan/tilt unit
Left	The landmark is on the left side	Pan left the pan/tilt unit or turn the robot left
Right	The landmark in on the right side	Pan right the pan/tilt unit or turn the robot right
Forward	Interact with the landmark	Move to the landmark, take a snapshot, and start interacting
Backward	Stop the robot	Stops all the ongoing activities and go back to its home position

Table 5.1: A description of the physical representation designed (gestures), the bound digital information, and the triggered robot behavior.
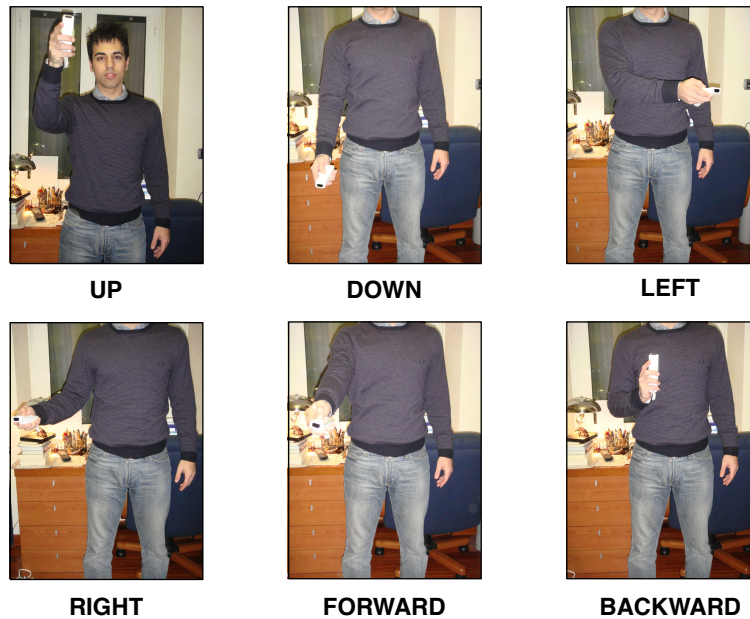
Figure 5.3: The gesture vocabulary defined for our interface: UP, $DOWN$, $LEFT$, $RIGHT$, $FORWARD$, $BACKWARD$.

can prevent any problem by stopping the robot. The evaluation of this system is reported in Section 5.4.

5.4   Experimental Evaluation

In order to evaluate our gesture recognition system, we organized an experimental evaluation with ten participants[1]. Ages range from 26 to 55 (mean $\approx$ 32.4, std $\approx$ 9.25). None of them had previous experience with a Wiimote, and they belong to different professional fields, hence it is an heterogeneous group. Each subject collects 30 gestures, five per each of the six gestures considered in our vocabulary. Gestures are acquired with random order through a particular application, then reducing the crossover effect. Before the acquisition step, each participant attended a five-minute training to get acquainted with the gestures.

Once acquired the accelerometer data, we extract for each gesture the feature vector $\Gamma_G = \langle d_x,\ d_y, d_z,\ \psi_n, \theta_n,\ v_x, v_y, v_z \rangle$. Figure 5.4 reports the

[1]Material about this experiment is provided at http://www.dis.uniroma1.it/ randelli/index.php?id=10.

extracted features on the training and validation set. We divide the total
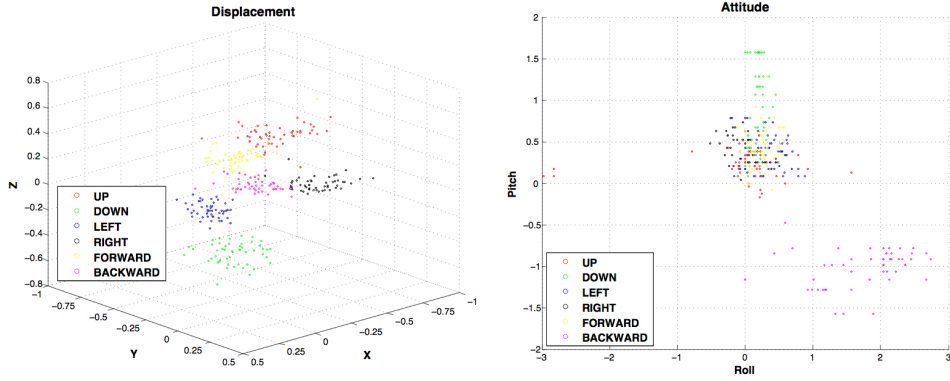
(a) Displacement along the three axes (d_x, d_y, d_z)

(b) Final attitude (ψ_n, θ_n)
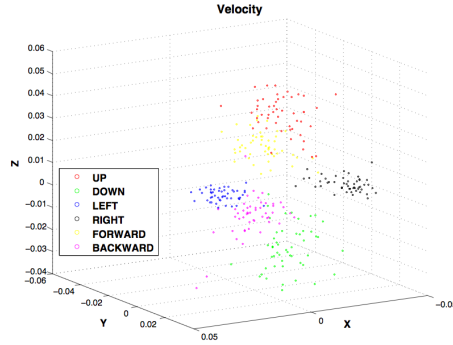
(c) Mean velocity along the three axes (v_x, v_y, v_z)

Figure 5.4: 3D plots with the feature vector extracted for gesture recognition.

amount of gestures into two sets with a 0.75 ratio: 225 samples as *training* set, and the remaining 75 samples as *validation* set. This split is performed randomly. In order to select an optimal classifier with respect to our feature set, we apply our training set to several common machine learning algorithms, and then validate each of them on the validation set. Upon this comparison, we select the approach which exhibits the best performance, which in our case is *support vector machines*. Table 5.2 reports the mean validation success rate. We implement, train and validate our SVM algorithm through the machine

KNN	Bayes	Decision Trees	Random Trees	SVM
97.725%	99.194%	96.66%	90.914%	99.335%

Table 5.2: The mean validation success rate of the applied machine learning algorithms reveals that SVM performs better than other approaches.

learning component of OpenCV library. Once trained, our classifier is tested on a *testing* set. The testing set has been acquired by the same ten participants, three gestures per type, for a total amount of 180 samples. Table 5.3 reports the test results with our SVM classifier.

UP	DOWN	LEFT	TOTAL
100% (0/30)	93.33% (2/30)	93.33% (2/30)	
RIGHT	**FORWARD**	**BACKWARD**	93.33% (12/180)
93.33% (2/30)	86.67% (4/30)	93.33% (2/30)	

Table 5.3: Success rate (and corresponding number of errors) for each gesture using a SVM classifier.

5.5   Conclusions

In this chapter we addressed high-level tangible interaction paradigms, in particular, focusing on gesturing. Another interesting paradigm will be introduced in Chapter 6, but it is related to the problem of knowledge acquisition, hence we postpone it to Part III. This chapter concludes our first contribution: adopting tangible user interfaces to lower the operator's cognitive effort for robot operation, thereby moving her cognitive skills on situation assessment.

Tangible interfaces are a recent technology still not widespread in the robotic community. Even more, their novel interaction metaphors have not been completely explored. This motivates our research path so far, which addressed a wide range of tangible paradigms: motion detection through grasping, background environment perception, operator mobility and gesturing.

This chapter concerns the design of a gesture-based interface for robot shared control. As soon as the task gets more complex, and robot operation becomes stressing for the human operator, awareness degrades drastically. Thereby, a shared control policy is required, where the operator acts as a *supervisor*, meanwhile the robot autonomously accomplishes the task. However, even state of the art robots may fail, and most of the robotic systems lack the required robustness to face these circumstances. Our approach is to support the robot with additional information, which reduces the probability of incurring in such a risk. We showed that achieving such a complex behavior with low-level interaction paradigms is not effective. Thereby, we move towards high level tangible interaction metaphors. In particular, we believe that gesturing is an optimal physical representation to trigger high-level commands to robots, thereby saving time and avoiding repetitive operations. Our gesture-based system allows to easily provide spatial information, and to avoid difficulties due to unexpected events. The success rate of our implementation is significantly high, and the support of robot speakers prevents from mistakes due to recognition failures.

As we expected at the beginning of this part, the most interesting aspects of TUIs concern high-level interaction paradigms, that leverage the seamless interaction of humans within real environments. On the other hand, traditional human-robot interaction suffer the same limitations of common GUI approaches adopted in human-computer interaction. In both cases, awareness is limited and impacts on the human cognitive processes. On the other hand, tangible user interfaces represent an opportunity to move back to a real interaction, where human-robot awareness can be acquired with the same means of human-human interaction. Throughout these two chapters, we provided the reader with novel means to teleoperate a robot, rely on pre-attentive notifications through background tangible processes, exploit intra-scenario mobility and, finally, adopt a shared control operation policy.

Part III

Knowledge Acquisition through Tangible Interfaces

Chapter 6

Tangible Pointing Interfaces

6.1 Introduction

Part III of this thesis addresses the problem of enhancing the situation assessment process. As already mentioned, this is often demanded to autonomous robotic systems. However, despite the recent progress in this field, the acquisition and grounding capabilities of autonomous robotic systems are still far from the expectations, mainly due to their lack of robustness and the complexity of real environments. On the other hand, knowledge grounding is quite a common activity for humans, because of their innate cognitive skills.

Thereby, our approach to this challenge is to define a novel methodology for situation assessment by establishing mixed human-robot initiatives, that is, moving humans actively in the knowledge grounding loop. Situation awareness is tightly coupled with human-robot awareness, since a robot system provided with significant knowledge is more robust to failures (and we will prove this in Chapter 8), and lowers the user intervention rate. Situation assessment outputs a high-level knowledge state that represents an improvement for the human's awareness, and it is more comfortable to deal with this knowledge, since it is represented in symbolic form.

Our solution to a human-centered situation assessment relies, once again, on tangible user interfaces. While in part II we proved TUIs effectiveness in terms of robot operation, they can also significantly contribute to the knowledge acquisition process, embodying interesting technologies that allows for the smart selection of environment elements. On the one hand, in this chapter we focus only on their role as selection tools. On the other hand, in Chapter 7 we will investigate another aspect of knowledge acquisition, that is, the relationship between tangible user interfaces and semantic knowledge.

In Section 6.2 we introduce the notion of *tangible pointing interface*, that is, the use of tangible user interfaces to select regions of interest in the environment. Section 6.3 provides a solution to a constrained version of this problem: selecting ground spots. This approach is evaluated in Section 6.4 for robot behavior composition. This methodology represents our best accomplishment, in terms of the synergy between humans, robots, and their environment, for a fast and seamless interaction and knowledge acquisition. Finally, experimental evaluations and conclusions are provided at the end of this chapter.

6.2   Tangible Pointing Interfaces

Our last contribution related to tangible user interfaces, here presented, is not related to robot operation, rather to the situation assessment process. Up to now, we have only considered tangible paradigms where the operator interacts *in* the environment, but not *with* the environment: perceiving, acting, but never *interacting*, or even *changing* the scenario. This is partially addressed with the concept of *tangible pointing interface* (TPI).

What is a tangible pointing interface? A tangible pointing interface is a tangible user interfaces that acts as a pointing device. There is no design nor hardware difference but the way its equipment is exploited, as reported in Figure 6.1. A tangible pointing interface is like a commercial pointer for presentations. It can be used by a human to point and select landmarks in the space: objects, areas, people, even events or situations. However, while a normal pointer is a mere tool, a TPI is an active device that interacts with robots. A tangible pointing interface is like a laser rangefinder, in the sense that can retrieve the distance from an object or its position, but embeds all the benefits of a tangible interface. Hence, it can also act as a controller for robot operation, as a device for behavior triggering, as a gesture-based system to add meaningful knowledge to the selected elements. In particular, in this thesis we assign to tangible user interfaces a twofold role: on the one hand, they act as robot controller, on the other hand, as knowledge acquisition tools.

How can tangible pointing interfaces enhance awareness? Situation assessment is a complex process, particularly when demanded to autonomous robots. First, there may be perception errors or ambiguities. Second, information may be not enough to guarantee a valuable assessment. Finally, autonomous systems are typically not enough robust for this process. Robots do not exhibit significant cognitive skills to accomplish disambiguation, grounding, and knowledge extraction. On the other hand, humans deal with this issue everyday. Thereby, putting them in the assessment loop with an active role may enhance the overall process. However, humans need smart solutions
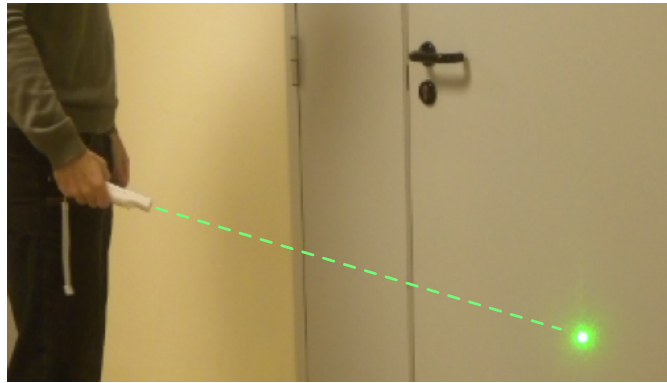
Figure 6.1: A simple example of a tangible pointing interface that selects a door (the green spot on the door is achieved through a commercial-off-the-shelf laser pointer and does not depend on the TPI).

to practically deal with this problem. The idea of a person sitting in front of a PC while tagging or introducing huge amount of data is not considerable. TPIs, integrating pointing and tangible characteristics, allow to quickly select relevant elements and ground them through gestures or other approaches. In particular, this last aspect characterizes the synergy between TUIs and semantic knowledge, as will be further discussed in Chapter 7.

In the rest of this section we focus on the pointing aspect, proposing our solution with TUIs, and present a testbed for smart robot behavior composition: a virtual coaching system for humanoid soccer robots.

6.3 Pointing with TUIs

In Section 3.3.1 we introduced some issues related to TUIs acting as pointing devices. In this section we address a specific case of the general problem: selecting elements on the ground, that is, not considering elevated elements. Again, we refer to a Wiimote device for the technical aspects.

As already stated in the previous chapter, accelerometers measure rough motion, and allow to retrieve the device roll and pitch, but they are not sufficient to retrieve the full attitude of the device within the space. A possible solution is to couple the tangible interface with a rangefinder sensor. For example, the Wiimote is equipped with an IR camera. However, this requires to cable the environment with IR leds, which is unrealistic and does not favor the operator mobility. Moreover, IR has a reduced range and field of view.

Our solution adopts gyroscopes to compute the interface attitude, which in turn is used to retrieve the position on the ground of the selected spot. This solution fosters the user mobility in the scenario, and does not require major changes in the environment.

From Gyroscopes to Device Attitude

To accomplish this, we extend the Wiimote with a small extension, called Nintendo Motion Plus, which provides two gyroscopes that report the three angular rates of the device. This allows to track the device attitude, even if a significant drifting error is accumulated over time.

Let F_B be a body-fixed reference frame. Then, $\omega^{F_B}(t) = [\omega_{pitch}(t), \omega_{roll}(t), \omega_{yaw}(t)]^T$ represents the roll, pitch and yaw rates measured at time t by gyros. It is worth mentioning that pitch and roll are inverted because, according to the Wiimote reference frame, roll is a rotation around the y-axis, while pitch represents a rotation around the x-axis. Now, let F_I be an initial inertial reference frame. Our goal is to track the device orientation with respect to F_I, hence we must solve an inverse differential kinematics problem. Let the quaternion vector $q(t) = [q_0(t), q_1(t), q_2(t), q_3(t)]^T$ be the attitude of the Wiimote at time t. We adopt quaternions to deal with singularities, since they are a more robust representation than Euler angles. Given the quaternion vector, $q(t)$, the following relationship represents the time derivative with respect to the angular rate given in the body frame $\omega^{F_B}(t) = [\omega_{pitch}(t), \omega_{roll}(t), \omega_{yaw}(t)]^T$.

$$\dot{q}(t) = \frac{1}{2}\Omega(t)q(t), \quad \Omega(t) = \begin{bmatrix} 0 & -\omega_{pitch}(t) & -\omega_{roll}(t) & -\omega_{yaw}(t) \\ \omega_{pitch}(t) & 0 & \omega_{yaw}(t) & -\omega_{roll}(t) \\ \omega_{roll}(t) & -\omega_{yaw}(t) & 0 & \omega_{pitch}(t) \\ \omega_{yaw}(t) & \omega_{roll}(t) & -\omega_{pitch}(t) & 0 \end{bmatrix}$$

$$(6.1)$$

The integration process has been implemented using the *classical Runge-Kutta method* (RK4). Finally, in order to have an easier interpretation of the attitude we convert back from the quaternion representation to Euler angles.

Let the Euler angle vector $\Theta^{F_I} = [\phi, \theta, \psi]^T$ be the attitude of the Wiimote at time t, with respect to the coordinate system F_I. Then, using the following equations, pitch, roll and yaw angles can be obtained in the absolute frame:

$$\begin{cases} \phi = \arctan\left(\frac{2(q_3q_3+q_1q_1)}{q_0^2-q_1^2-q_2^2+q_3^2}\right) \\ \theta = \arcsin(-2(q_1q_3 - q_1q_2)) \\ \psi = \arctan\left(\frac{2(q_2q_2+q_1q_3)}{q_0^2+q_1^2-q_2^2-q_3^2}\right) \end{cases} \qquad (6.2)$$

Since gyros are quite noisy, angular rates are preliminary filtered, before integrating them, using a *moving average filter*, as reported by Wei (2006). How-

ever, in case of orientation tracking over long time intervals, such a filtering method is not efficient, and it is necessary to adopt some control technique, such as a Kalman Filter, as proposed by Luinge and Veltink (2005).

From Device Attitude to Spot Position

Once retrieved the Euler angle vector, that is, the attitude of the Wiimote with respect to F_I, we retrieve the position of the selected spot on the ground. A sketch of the problem is reported in Figure 6.2. Unfortunately, gyroscopes do not allow to track the device position within the environment. Thereby, the operator cannot perform gestures while walking, but she must stand in a fixed spot of the space, which is known a priori. Let $P_0^{F_I} = (x_0, y_0, z0)$ be the operator position according to the absolute frame F_I. For the sake of simplicity, we assume x_0 and y_0 to be congruent with the origin of F_I, thereby $P_0^{F_I} = (0, 0, z_0)$, where z_0 is known a priori. Then, the position on the ground
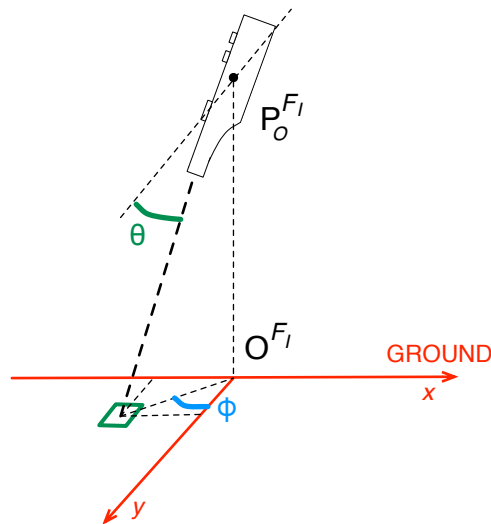
Figure 6.2: A sketch representing the problem of computing the ground spot position from the Wiimote attitude.

of the selected spot, $P_S^{F_I}$, is given by the following relation:

$$P_S^{F_I} = (x_S, y_S), \text{ where}$$

$$x_S = \tan\left(\frac{\pi}{2} - \theta\right)\cos\left(\frac{\pi}{2} - \phi\right),$$
$$y_S = \tan\left(\frac{\pi}{2} - \theta\right)\sin\left(\frac{\pi}{2} - \phi\right). \tag{6.3}$$

The implementation of this part is described in Section 6.5.

6.4   A System for Behavior Composition

We aim at exploiting the pointing system to create behaviors, by directly interacting with robots. The interaction comfort, and the proximity to the robots enhances the operator awareness of the situation. Furthermore, multimodal feedback, such as rumble or sounds, provides background information, lowering the cognitive fatigue.

In order to use the Wiimote and Motion Plus extension for behavior composition, we design the **Wiimote Attitude System**, with the following objectives:

- supporting multiple Wiimote devices;

- integrating the Wiimote with our robotic system;

- extracting and tracking high-level information from the raw data coming from the Wiimote sensors;

- computing the Wiimote attitude, given the angular rates measured by the gyros.

An interaction component diagram that describes the overall architecture is provided in Figure 6.3. To manage communications with the Wiimote, we
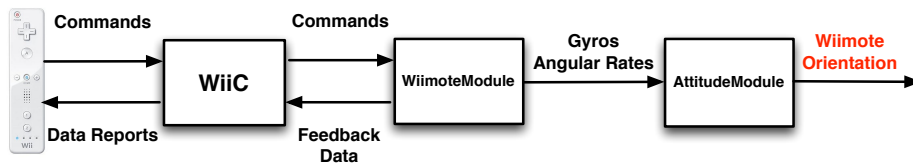
Figure 6.3: The Wiimote pointing system UML component diagram.

developed the **WiiC**[1] library. WiiC is a C++ layer that allows to connect to multiple Wiimote devices, as well as supporting the Motion Plus and Balance Board extensions, and all the other Wii devices. WiiC retrieves accelerometers, gyro, and IR data, and controls buttons, speakers, and rumble. Furthermore, it supports data logging, and is integrated with the machine learning algorithms provided by the OpenCV library. The **WiimoteModule** is an OpenRDK module that integrates the Wiimote with our robotic system. Finally, the **AttitudeModule** implements the attitude computation described

---

[1]WiiC is available at http://wiic.sf.net

in the previous section. The output of this system is the device attitude with respect to the initial inertial reference frame F_I.

6.5 Experimental Evaluation

To evaluate the effectiveness of our pointing system, we developed a prototype application for **virtual coaching**. The objective is to compose complex behaviors for multi-agent systems, such as game tactics, by using a user-friendly interface. The soccer field is divided into several zones, and the operator creates a tactic by assigning each robot to a specific area in the field. Our approach allows the operator to select areas with the tangible pointing interface, that is, directly in the field and with a natural interaction mean. Then, the tactic is composed by assigning a distinct role to each robot player. Since we are using real soccer rules, simple role behaviors are distinguished among: GoalKeeper, Defender, Supporter and Attacker. The overall system architecture is depicted in Figure 6.4. The input of our system is a Wiimote, equipped
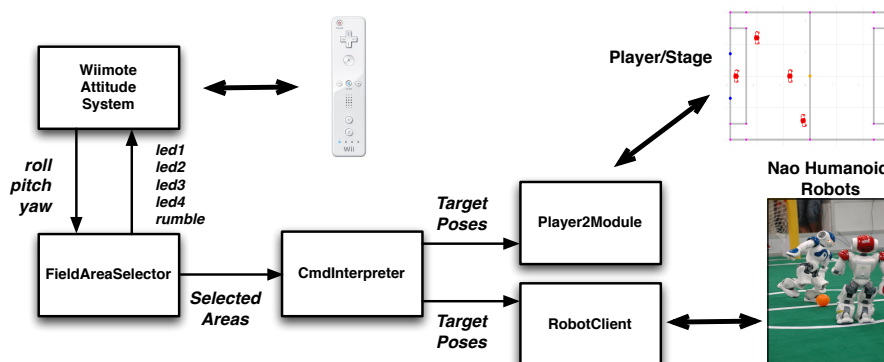
Figure 6.4: The proposed architecture for virtual coaching.

with the Motion Plus extension. First of all, the user calibrates the Wiimote. Then, she selects four areas of the soccer field. The complete finite state machine of the user interaction, implemented by the Wiimote Attitude System, is given in Figure 6.5(a). The **FieldAreaSelector** module takes as input the Wiimote attitude and computes the selected field area. Then the selected areas are translated into field coordinates and the coordination procedure is called. It is worth mentioning that the system also validates whether the selected target points correspond to free areas within the field, hence no outer parts of the field nor obstacles can be selected. At the end of this process,

the four area selections are codified in a string message sent to the **CmdInterpreter** module. CmdInterpreter accepts the string command, where the
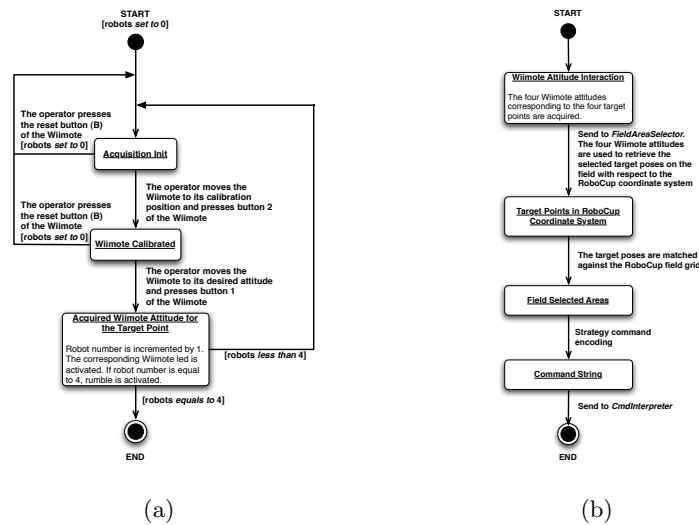
Figure 6.5: State diagrams about the spot selection with the Wiimote 6.5(a) and the overall interaction with the coaching system 6.5(b).

areas selected by the user are encoded, and outputs a target position for each robot player. The last step involves sending the four target positions to the corresponding robots, which move towards them. A rumble feedback is sent to the Wiimote to warn the user of the successful command.

We evaluated the precision of the pointing system for this task on real Nao humanoid robots, currently adopted in the RoboCup Standard Platform League (SPL), and on a Player/Stage simulated world, modeled upon the SPL rules. The complete data flow of the described application is shown in Figure 6.5(b).

Experimental results

Experiments have been conducted to establish the correctness of the proposed architecture. They are not intended to be exhaustive, since they lack a strong validation formalism, and should be considered as a pilot study.

Given the limitation of the sensors adopted, we define the following assumptions. In order to correctly determine selected areas, the device position is known a priori. Thereby, we assume that the user selects the points by

Scenario	User Interaction	Feedback Visual	Feedback Physical	Access to the Environment
TextOnly	Text	×	×	×
GUI	Point & Click	✓	×	×
TUI	Tangible	✓	✓	✓

Table 6.1: Qualitative results for Virtual Coaching experiments.

standing at the center sideline of the field. This information can be changed at runtime, and does not affect the accuracy of the system. Moreover, we discretized the field into nine distinct areas. We determined that the system has an accuracy of 0.4 meters. This means that the architecture is not able to distinguish areas whose size is less than $0.16m^2$. This is due to the noise in the pointing system, and it determines a lower bound for the accuracy that can be achieved. However, this value is still acceptable to be used in real applications, and we proved that by evaluating the virtual coaching example.

As well as accuracy measurements, a qualitative evaluation was conducted by using a real RoboCup SPL field. We compared the execution of the virtual coaching by using three different interfaces: text-based, GUI-based, and using a TUI. In the *TextOnly* scenario, the user had to insert the command by sending a string using a telnet-like program. In the *GUI* scenario, the operator interacted with the *RConsoleQT* program, provided by the OpenRDK framework. Finally, in the *TUI*, the operator used the aforementioned system.

Our goal was to evaluate the characteristics of the different approaches and the constraints the users faced with each input type. Table 6.1 resumes the expected features emerged by performing such qualitative experiments.

As future work, we will provide an extensive and quantitative evaluation of the whole system.

6.6 Conclusions

This chapter introduces the concept of *tangible pointing interface* (TPI), which allows humans to select landmarks in the environment, and to acquire them as relevant knowledge. On the one hand, we presented another high-level tangible interaction paradigm, focused on the user capability to interact *with* the environment, by pointing ground spots. Despite some technical limitations,

our implementation represents an interesting proof of concept. On the other hand, this represents a preliminary step to define a novel situation assessment methodology that relies on the active participation of the human operator. This aspect will be further explored in Chapter 7, discussing the integration between tangible interfaces and semantic representations.

To summarize, our solution has the following advantages with respect to the traditional approaches to knowledge acquisition: first, it fosters intra-scenario operator mobility. Second, the operator acquires relevant spots simply by selecting them, which is a fast and natural approach, common in human-human cooperation. Finally, the interface embodies a twofold role: robot controller and selection tool. These benefits enable the operator to support the robot in acquiring a significant amount of knowledge. This in turn is provided as input to knowledge representation and reasoning systems to infer further information, which contributes to a proper situation and human-robot awareness.

Chapter 7

Semantic-driven Tangible Interfaces

7.1 Introduction

In the previous chapter we have introduced a novel tangible interaction paradigm that allows the operator to acquire relevant knowledge, by selecting ground spots. Such an approach fosters intra-scenario mobility, thus the operator is free to move within the environment, even co-located with the robot. Knowledge acquisition is performed with a tangible pointing interface, that is, simply by pointing interesting spots. On the other hand, the same interface embeds all the controller functionalities presented so far. With respect to traditional systems, mainly based on computer graphic interfaces, this represents a more natural approach. However, two considerations arise. First, the interface presented in the previous chapter addresses a constrained version of the general pointing problem, that is, selecting only ground spots, and not elevated objects. Second, and even more important, we focused only on the selection aspect, without addressing how to ground this knowledge through explicit formalisms.

Thereby, we shift our focus from selection to *grounding*. Our approach to this challenge is to establish *mixed* human-robot initiatives, that is, moving humans actively in the knowledge grounding loop. This in turn significantly enhances human-robot awareness. In fact, semantic knowledge lessens the humans' burden in understanding the scene and interacting with mobile robots. Think, for example, about asking a robot to "go to the kitchen", as opposed to

"go to $x = 10, y = 200$". Conversely, let us consider the advantage of receiving symbolic feedback from robots, instead of interpreting dozens of numeric data.

In this chapter, we tackle this issue by *establishing a tight synergy between tangible user interfaces and semantic knowledge*. We refer to this as **semantic-driven tangible interfaces**. This synergy is supported by three aspects. First, we interpret user gestures to provide contextual information, as already described in Section 5.3. Second, in Section 7.3 we present a general solution to the problem of spot selection within the environment, this time addressing elevated elements too. Finally, acquired information is represented with a symbolic formalism. However, the problem of effectively represent, manage and exploit such knowledge will be discussed in Part IV of this thesis.

We present in Section 7.4 a multimodal interface, which combines gestures, vision and speech to allow for an easy interaction with humans, in order to specify both features of the operational environment and tasks to be accomplished. The result of this interaction is a symbol grounding process that allows for the acquisition of knowledge beyond the capabilities of current implementations. Section 7.5 reports a preliminary validation of such a system.

7.2   Study Case: Robot Surveillance

In order to motivate our work, let us briefly address a simple case study about indoor robotic surveillance. A number of research projects and robots in the market have been targeting it. We define a case study that resembles the scenario described in Section 5.2, but it is further expanded. We consider an indoor structured environment, composed of rooms and corridors, where the robot must be able to navigate and localize itself. The robot is expected to monitor the environment detecting anomalous situations, and, possibly, the presence of humans. How would a robotic system take advantage from contextual knowledge? Some examples are:

- differentiating *rooms* from other structural elements, makes it possible to trigger robots for specific patrolling paths;

- given knowledge about a specific *object*, the robot can search for it in specific locations and plan the search process accordingly;

- knowing about a *door* with respect to walls, a robot can check whether it is open;

- *windows* are critical entry points, hence robots should control more often these spots;

- when the building is closing, robots control whether all doors and windows are closed, turn off *lights*, and assess whether someone is in the corridors, using their speakers;

- knowing about *stairs* or *ramps* allows for a better motion and inspection of the corresponding areas;

- once selected specific places as *robot dock stations*, robots move there, when out of battery.

Many of the above examples are beyond the capabilities of current systems, unless they are tackled in ad-hoc ways, by embedding the knowledge about the environment and about the specific task into the system implementation. This makes existing solutions rather inflexible and incapable to deal with the level of abstraction that typical users would need. In our view, this is largely caused by the difficulty for the robot to have the relevant knowledge and to acquire it from the user. The key issue becomes then the definition of a situated world model for the robot, compact enough to acquire only those aspects that are relevant to accomplish the desired tasks. For example, to achieve a proper motion in the environment, we would like to model some structural elements such as: stairs, ramps, doors, floors, windows, and elevators; more generally, other elements of interest, could be: robot docking stations, crossings, people, RFIDs, artificial lights. Many of them have already been taken into consideration in several previous experiments involving real robots, but the implemented systems are very much dependent on the specific experimental scenario. Acquiring and grounding contextual information represented in any explicit form is still a challenging task.

More specifically, we can phrase our research questions as: *How is it possible to naturally refer to elements of the operational environment? How is it possible to ground semantic knowledge about contextual information effectively (and conduct this as an interactive and continuous process, involving humans)?* From these questions, we outline some requirements we expect to satisfy through our solution:

- semantic knowledge acquisition should involve humans as active participants;

- the adopted technique should be comfortable, to lower users' cognitive stress;

- acquisition and grounding should not be an isolated step, rather a continuous process, performed without interrupting humans' everyday activities;

- users' mobility should be fostered;

- acquisition and grounding overhead should be limited in time.

Our proposal is thus based on the integration of novel human-robot interaction metaphors with established artificial intelligence approaches, to design new human-centered semantic knowledge acquisition methodologies.

7.3   Semantic-driven Tangible Interfaces

In this section, we deal with the research questions phrased in Section 7.2, through a qualitative analysis of effective approaches to contextual knowledge acquisition and grounding. In particular, we focus on the role of tangible user interfaces, since we consider this interaction metaphor still unexplored. **Our solution is to establish a synergy between tangible user interfaces and semantic representations, designing a novel type of interface: semantic-driven tangible interfaces**. However, there exist different types of knowledge that can be grounded (e.g. spatial, temporal, objects, events, and so on), and it would be unrealistic to expect tangible interfaces as the optimal solution for all of them. Thereby, in order to identify the best solution for knowledge acquisition and grounding, it is important to couple tangible interfaces with other interaction means, such as speech and vision, designing a multimodal interface.

7.3.1   Spot Selection and Detection

Spot selection is related to our first research question: how to naturally refer to elements of the operational environment. We have already dealt with this problem in Section 3.3. In Section 6.2, we presented *tangible pointing interfaces* for a constrained version of the spot selection problem: pointing objects laying on the ground. In this section, we address the general problem, including the selection of elevated elements.

This problem is relatively easy to solve in virtual environments, where metric distances between objects are known a priori. For example, a common technique is *ray casting*. A virtual ray originating at the user's device shoots out in the direction she is pointing. Typically, the first object to be hit by the ray is selected. Gallo et al. (2008) adopt this method to manipulate 3D objects reconstructed in a virtual environment from medical data.

In real environments, the problem is much more challenging. In fact, most of the tangible interfaces are equipped with accelerometer sensors, and sometimes with gyroscopes and magnetometers too. Information coming from these sensors is used to detect the position and attitude of the tangible pointer, which is a preliminary step to point objects within the environment. Still, one
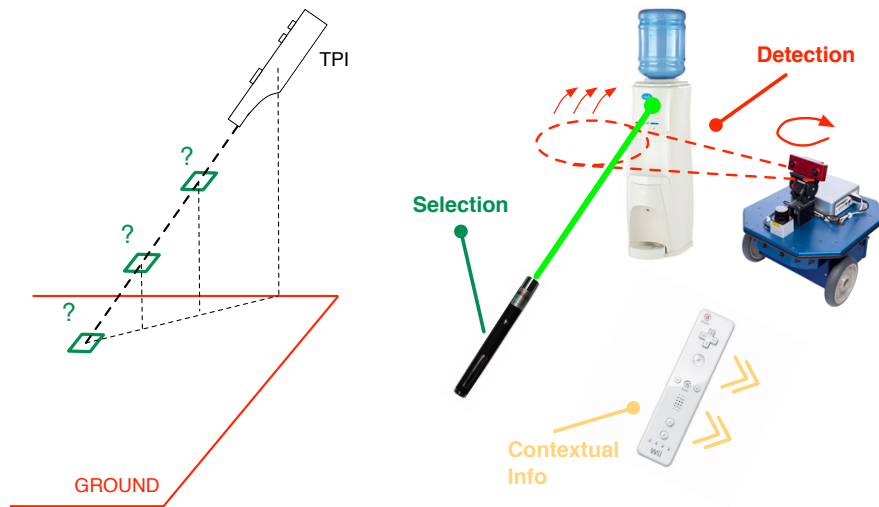
(a) Without any rangefinder sensor, it is impossible to measure the distance we are pointing at, hence to retrieve the z component of the selected spot.

(b) Decoupling spot selection from spot detection. Spot detection is enhanced by TUIs providing contextual information to speed up the process. In this case, the user performs a right gesture, which is codified as spatial hint for the pan/tilt unit that turns the stereo camera right.

Figure 7.1: Pointing landmarks in a 3D environment.

problem arises: these sensors do not provide any feedback about the distance between the interface and the selected objects, hence they are not valuable for localizing the object position in the 3D space, as reported in Figure 7.1(a). A first approach is to couple tangible interfaces with rangefinder sensors (e.g. sonars, lasers, and so on). However, this does not guarantee users' mobility, since these sensors are typically not portable. Another approach is to equip environments with sensors to detect the position of selected spots. For example, Sko and Gardner (2009) adopt multiple IR bars to enhance the Wiimote capabilities as controller in a virtual reality theatre. Yet, such a solution would require significant changes in the environment.

Our proposal takes the best of both the approaches, by decoupling the knowledge acquisition process into a *selection* and a *detection* component, and demanding this latter to the robotic platform (see Figure 7.1(b)). In particular, tangible interfaces are a valuable tool for spot selection, while spot de-

tection has been demanded to vision-based systems, which allow for relatively easy object recognition. The human operator is equipped with a commercial-off-the-shelf small pointer, which is used to highlight a relevant spot with a green dot. A vision subsystem, composed of a stereo camera and a pan/tilt unit, is delegated to recognize the green dot, hence detecting the selected spot. The human operator can thus point at spots in the environment, whose location can be determined by the vision system on the robot. At the same time, the user gestures are mapped onto contextual information for the pan/tilt unit, to narrow the search space for the green dot. Since we adopt the same system and vocabulary designed in Section 5.2, we will not describe again this component. The implemented system is described in Section 7.4.

7.3.2 Contextual Information Acquisition

Our second research question is concerned with symbol grounding: how we may effectively ground symbols into the representation. We have already mentioned that pure vision-based approaches suffer from several limitations. On the other hand, conventional approaches, such as tagging, or using graphical interfaces, are less comfortable for humans, in particular if symbol grounding is regarded as a continuous process (not restricted to system set-up). Relying only on gesturing is also not suited, since physical representations that map gestures with objects are not always effective, as already mentioned in Section 5.2. However, it has been proved that combining utterances and gestures is an effective interaction mean for humans, especially when dealing with spatial relations. Perzanowski et al. (2005) propose a multimodal interface where gestures disambiguates speech commands . For example, consider the following instruction "*go over there*". This would be useless without recognizing the user pointing to a specific place. Hence, we choose to deploy a speech recognition system that converts the operator's utterance into a symbolic representation, which is in turn grounded to the element previously detected by the vision subsystem. Once grounded, the symbol is stored in the knowledge base, and will be exploited to trigger robot behaviors. In particular, this last aspect is analyzed in Chapter 8, where we will discuss an effective solution to effectively manage semantic knowledge and enhance robot skills in accomplishing autonomous tasks.

7.4 Architecture for Multimodal Interaction

In order to validate the analysis in Section 7.3, we present a robotic architecture that allows to acquire contextual knowledge through an effective integra-
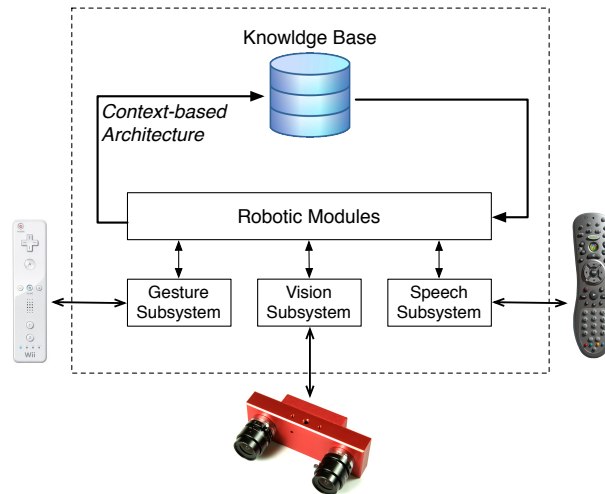
Figure 7.2: An overview of the system architecture. Here are reported the five components: *(i)* robotic subsystem, *(ii)* gesture subsystem, *(iii)* vision subsystem, *(iv)* speech subsystem, and *(v)* knowledge base. The overall system is tied by the context-based architecture.

tion of tangible, speech, and vision systems, in accordance with their aforementioned benefits. Our robotic architecture consists of five components: *(i)* robotic subsystem, *(ii)* gesture subsystem, *(iii)* vision subsystem, *(iv)* speech subsystem, and *(v)* knowledge base (see Figure 7.2).

Robotic Subsystem.   Includes common robot modules, all developed with the OpenRDK robotic framework: motion controller, path planner, localizer, and a mapper. They represent the low-level perceptive part of the system, based on numerical representations. Using the data acquired through the laser rangefinder the system performs SLAM and builds a map of the environment. The overall system runs on an Erratic wheeled robot.

Gesture Subsystem.   It manages communications between the tangible interface held by the user, and the robotic system. As usual, we adopted a Wiimote device. This system, based on our gesture recognition interface described in Section 5.3, is responsible for the following functionalities:

- manual robot teleoperation;

- recognizing gestures associated to contextual information hints for the vision-based detection of selected spots in the environment.

Table 7.1 reports the set of gestures. The interface manages two operation modes: *robot teleoperation* and *spot selection.* Button *One* on the Wiimote switches the desired operating mode. The Wiimote leds are used to remind users about the selected operating mode. To notify the operator about the correct recognition of gestures in spot selection the device rumble is activated.

Operation Mode	Command	Meaning
Robot Teleoperation	Wiimote Pitch	Move the robot forward/backward
	Wiimote Roll	Turn the robot left/right
	Move the arm towards the chest	Emergency stop of the robot
Spot Selection	Move the arm left (right)	The selected spot is on the left (right) with respect to the robot
	Move the arm upwards (downwards)	The selected spot is higher (lower) with respect to the robot
	Move the arm outwards (inwards)	The selected spot is farther (closer) with respect to the robot

Table 7.1: The command mapping on the Wiimote device for our tangible interface. Two operation modes are supported: *teleoperation* and *spot selection.*

Vision Subsystem.   It manages a pan/tilt unit and a stereo camera installed on top of it. In particular, this module is responsible for:

- detecting spots through perceptions coming from the stereo camera, according to Algorithm 1. Users can select spots through a commercial-off-the-shelf laser that emits a green dot, and can be tied to the tangible interface;

- controlling the pan/tilt unit according to contextual information hints, as described in Table 7.2;

- taking snapshots of each selected dot, which is used to tag a semantic map of the environment.

Speech Subsystem. The speech subsystem acquires human utterances to perform symbol grounding of the elements selected through the green laser and acquired by the vision subsystem. Furthermore, the robot speakers provide audio feedback to people in the environment (e.g. asking to open a door, communicating with people in corridors, alerting personnel when the department is closing, and so on). To implement this component, we adapted to our purposes the Speaky[1] speech recognition system. Speaky is a speech technology adopted in home automation, currently with a speech recognition engine for italian, that does not require any training. We defined a simple dictionary of words that are successfully recognized by the system and are sufficient for the purpose of symbol grounding. Integrating the speech component with a dialogue manager is in our research agenda.

Knowledge Base. It manages the acquired knowledge and the reasoning that triggers robotic behaviors. The considered world model is the same described in our case study in Section 7.2, and has been modeled through the ECL^{i}PS^e CPS library[2]. It is worth noting that we are not providing here the reader with any design choice, nor the adopted representation, as we postpone this discussion in Chapter 8, where an effective solution for semantic knowledge representation, management and exploitation will be presented. However, without going into details, at a given time, the knowledge base contains facts about structural elements of the environment (e.g. doors, rooms, stairs, and so on), mission-related relevant objects (e.g. robot docking stations, crossings), and people. Robotic behaviors are automatically triggered according to the knowledge defined in the knowledge base. For example, consider the case where the robot is in a room where a robot docking station is available, and its battery is running out. The desired behavior is then to reach the docking station for recharging. This might, for example, be expressed in the following

---

[1]http://www.mediavoice.it/

[2]http://eclipseclp.org/

Algorithm 1: Spot Detection

Input:

SpatialHintVocabulary: $U = \{UP, DOWN, LEFT, RIGHT, IN, OUT\}$

SpatialHint: $u_t \in U$ (spatial hint provided at time t)

CameraPan: ϕ_{t-1} (stereo camera pan at time $t-1$)

CameraTilt: θ_{t-1} (stereo camera tilt at time $t-1$)

CameraZoom: z_{t-1} (stereo camera zoom factor at time $t-1$)

PanStep: $\Delta\phi$ (stereo camera pan step)

TiltStep: $\Delta\theta$ (stereo camera tilt step)

ExposureStep: ΔE (exposure step), *Gain*: G (stereo camera gain)

CameraMaxPan: ϕ_{MAX} (stereo camera max pan)

CameraMaxTilt: θ_{MAX} (stereo camera max tilt)

CameraMaxZoom: z_{MAX} (stereo camera max zoom)

Output:

SpotPosition: $\boldsymbol{x}$ (spot position vector)

CameraPan: ϕ_t, *CameraTilt*: θ_t, *CameraZoom*: z_t

1 **for** *exposure* $\longleftarrow$ 0 **to** 100 **do**

2   GrabFrame(exposure,G)

3       $I_L \longleftarrow$ GetLeftFrame()

4       $I_R \longleftarrow$ GetRightFrame()

5       $I_d \longleftarrow$ ComputeDisparityMap(I_L, I_R)

6   **foreach** *pixel i in* I_d **do**

7   **if** $i > THRESHOLD$ **then**

8               $\boldsymbol{x} \longleftarrow$ ComputeSpotPosition(i)

9   **return** true

10   exposure $\longleftarrow$ exposure $+ \Delta E$

11 ComputePanTiltZoom ()

12 $\phi_t \longleftarrow$ CheckBounds $(\phi_t, -\phi_{MAX}, \phi_{MAX})$

13 $\theta_t \longleftarrow$ CheckBounds $(\theta_t, -\theta_{MAX}, \theta_{MAX})$

14 $z_t \longleftarrow$ CheckBounds $(z_t, 0, z_{MAX})$

15 **return** false

Gesture (u_t)	Behavior		
	Pan (ϕ_t)	**Tilt** (θ_t)	**Zoom** (z_t)
UP	$\phi_t \longleftarrow \phi_{t-1}$	$\theta_t \longleftarrow \theta_{t-1} - \Delta\theta$	$z_t = z_{t-1}$
DOWN	$\phi_t \longleftarrow \phi_{t-1}$	$\theta_t \longleftarrow \theta_{t-1} + \Delta\theta$	$z_t = z_{t-1}$
LEFT	$\phi_t \longleftarrow \phi_{t-1} - \Delta\phi$	$\theta_t \longleftarrow \theta_{t-1}$	$z_t = z_{t-1}$
RIGHT	$\phi_t \longleftarrow \phi_{t-1} + \Delta\phi$	$\theta_t \longleftarrow \theta_{t-1}$	$z_t = z_{t-1}$
IN	$\phi_t \longleftarrow \phi_{t-1}$	$\theta_t \longleftarrow \theta_{t-1}$	$z_t = z_{t-1} - 1$
OUT	$\phi_t \longleftarrow \phi_{t-1}$	$\theta_t \longleftarrow \theta_{t-1}$	$z_t = z_{t-1} + 1$

Table 7.2: The output of the *ComputePanTiltZoom()* step, which updates the pan/tilt/zoom position according to the provided spatial hint u_t at time t.

way:

$$
\begin{aligned}
\textbf{IF} \quad & robot(A) \land hasPose(A, PoseA) \\
& \land room(PoseA) \\
& \land hasDockingStation(PoseA, StationA) \\
& \land hasPose(StationA, PoseStationA) \\
& \land robotBatteryLevel(A) == low \\
\textbf{THEN} \quad & TARGET\_POSE = PoseStationA
\end{aligned}
$$

7.5 Experimental Evaluation

Our system has been deployed in a real environment for a preliminary validation of our case study: indoor robotic surveillance. Our aim is the assessment, from a quantitative point of view, of the effectiveness of the proposed solution with respect to our preliminary research questions. We report in Figure 7.3 the result of the knowledge acquisition and grounding process conducted during our validation. Such knowledge has been considered for the following scenario: performing an indoor robot patrolling, computing paths that prioritize offices and laboratories with respect to other rooms (e.g. toilettes), meanwhile checking that all the doors are closed, and adopting safe motion policies in presence of stairs and ramps. Robots should also be aware of the presence of docking station for recharging. However, it is worth remarking that in this section we are only validating the knowledge grounding step, and not the robot performance in accomplishing its task, that will be evaluated in the next chapter.

Figure 7.3: Output of the robot mapper through the semantic knowledge acquired by the operator for a surveillance scenario. Knowledge about structural elements and relevant elements is overlaid on the metric map built by the robot.

As for the tangible subsystem, gesture recognition relies on the SVM classifier introduced in Section 5.2, with a vocabulary of six arm gestures (upwards, downwards, left, right, outwards, inwards). Detected gestures provide contextual information for a shared control of the pan/tilt unit of the vision subsystem. They represent spatial hints that narrow the pan/tilt space meanwhile searching for a green laser dot.

As for the vision subsystem, the scene is acquired through a monochrome stereo camera. Our grabbing algorithm keeps exposure and gain low, in order to acquire only bright elements, such as laser dots. It then computes a disparity map and segments the image, to filter everything but the candidate green dot. Once retrieved the pixel coordinates of the dot in the disparity map, they are transformed with respect to a camera-fixed reference frame, which are in turn transformed according to an absolute reference frame. Finally, the mapper marks the acquired element on its metric map. During the system evaluation, it has been possible to detect spots up to ten meters far (with

a depth resolution of 0.18 meters), and as close as 1 meter (with a depth resolution of 0.002 meters), with a success rate greater than 90%. A snapshot of the acquired scene, the grabbed frame, and its relative disparity map, is reported in Figure 7.4.
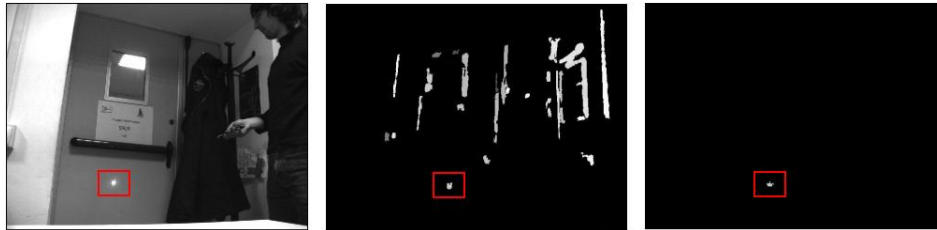
Figure 7.4: The scene acquired by the stereo camera, the grabbed frame with our vision algorithm with low exposure and gain, and the corresponding disparity map with the green dot selecting a door.

7.6 Conclusions

In this chapter we have presented a system that has been designed in order to achieve effective symbol grounding through human-robot interaction based on the integration of gestures, vision and speech. In our vision, tangible user interfaces are coupled with semantic knowledge representations, with a twofold role: providing high-level contextual information to speed up the grounding process, and easing knowledge selection in the environment. This latter aspect, in particular, enables the operator to naturally acquire relevant knowledge.

The capability to exploit the human operator knowledge about the operational environment and the task to be accomplished in a seamless way, can give substantial benefits in terms of leveraging the robot capabilities in the accomplishment of specific tasks, as we will discuss in Chapter 8. This in turn enhances the system robustness, and lessens the robot failure rate, hence the operator is more involved in acquiring a proper awareness, rather than continuously operating the robot. Even more, symbolic representations are a compact and natural knowledge representation, more comfortable for humans, which support them throughout the assessment process, without requiring a significant cognitive effort to interpret data.

The proposed approach has been validated, with several simplifications in order to achieve a running prototype, on a case study in indoor surveillance. It can nonetheless be deployed in a variety of domains and tasks, including

for example service robotics. To this end, it is worth mentioning that the experimental framework designed by Iocchi and van der Zant (2010) for the RoboCup@Home competition represents another very interesting test-bed for the proposed approach.

Tangible user interfaces may then improve the grounding process, providing smart acquisition methods, which are not comparable to an operator sitting in front of a computer performing the same task. Of course, as already mentioned, some activities of this process are better achieved using other means, such as vision or speech. This confirms our hypothesis (reported in Section 3.5) about the effectiveness of TUIs in combination with other interaction paradigms.

Part IV

Semantic Knowledge and Awareness

Chapter 8

Context-based Architecture

8.1 Introduction

Part IV of this thesis addresses the usage of semantic knowledge to enhance both human-robot awareness and situation awareness, and completes our research investigation. Our research problem is to assess how such knowledge, once grounded, may be effectively represented, managed, and exploited.

There are two major limitations of today systems in dealing with this problem. First, explicit representations are not always adopted, which does not allow to rely on reasoning mechanisms. Second, even when properly represented, knowledge is often hard-coded in manifold robotic modules, which requires an extensive re-engineering of the existent implementations to extract and integrate such a knowledge. This introduces the risk of inconsistency and does not foster any knowledge re-use and sharing.

In this chapter we look for a solution to properly represent and deal with semantic knowledge. First, knowledge must be represented according to an explicit representation. Second, we do not rely on a *specific* type of formal representation. That is, it is possible to adopt first-order logics, description logics, modal logics, and so on. Third, we aim at defining a system architecture that decouples the representation and reasoning system from the existent robotic modules, hence minimizing the re-engineering effort.

As a second major contribution, once defined such a system, we aim at validating that semantic knowledge provides benefits both to robots and humans. First, semantic knowledge is the main ingredient for representation and reasoning systems to enhance robot autonomous tasks. Second, it is represented through concepts, which are more comfortable for humans. A symbolic description is more tractable than dealing with a graphic display full of numerical

data that need to be interpreted by the operator. This latter aspect highlights how such a knowledge enhances human-robot awareness.

In particular, we focus on a fragment of knowledge, which is *contextual knowledge*, already defined in Section 3.4 as environmental, mission-related, and agent-related relevant information. The "contextual approach" has been shown to be an effective solution for robust autonomous systems. While it seems trivial that a system provided with such knowledge achieves a better performance with respect to a context-free system, this is not at all easy to accomplish, both in terms of acquiring this knowledge, and to properly represent and exploit it. It is also worth noting that the knowledge acquired in the previous chapter, thanks to semantic-driven tangible interfaces, is a subset of contextual knowledge.

In Section 8.2 we present our solution to the aforementioned issues, a **context-based architecture**, which extends conventional robotic systems with a KR&R component, without massive changes on the existent implementations. In Section 8.3 we apply our context-based architecture to a search and rescue (SAR) application, by adopting a specific knowledge representation and reasoning formalism. SAR operations involve the localization of victims trapped in confined spaces, using mixed teams of human operators and mobile robots. Finally, Section 8.4 provides an extensive evaluation of this system in a search and rescue environment. We show that, even if it is still unrealistic to define a robotic system capable to autonomously accomplish such a task, contextual knowledge can improve its performance in the search, hence lowering the operator's burden in controlling the platform during rescue operations, meanwhile increasing the operator's cognitive focus on the situation assessment.

It is worth noting that this is the result of a collaborative work, and the specific contributions presented in this thesis concern the definition of the context-based architecture, its extension to temporal and spatial constraints, and the whole experimental evaluation.

8.2　Context-based Architecture

A context-based architecture, introduced by Calisi et al. (2008b), resembles a feedback controller (see Figure 8.1). It can be formally defined as a quadruple $\langle S, T_{d/c}, R, T_{c/d} \rangle$, where:

- S is the context-free system, which represents any conventional robotic system, composed by modules such as motion controller, mapper, exploration module, localization module, and so on;

- $T_{d/c}$ is a finite set of data/context transduction modules, which process numerical output from S to extract contextual knowledge, represented in whatever symbolic language;

- R is a block of contextual reasoning modules, to infer new knowledge, useful for the tasks of the system;

- $T_{c/d}$ is a set of context/data transducers, which transform any symbolic representation into numerical data in order to control modules in S (closing the loop).
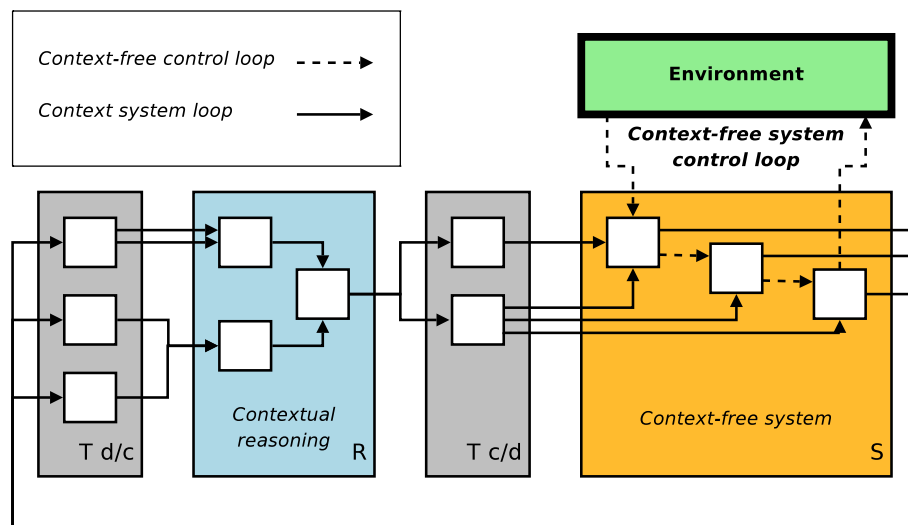
Figure 8.1: The main components according to the definition of context based architecture as a feedback controller.

Intuitively, based on the information extracted from the output of S, R can infer contextual knowledge which can be used to control the modules in S. The $T_{c/d}$ and $T_{d/c}$ modules are required to interface S and R by transforming data into symbols and symbols into data, respectively. The context system loop is executed at a much lower frequency then the context-free loop, thus not affecting the system reactivity. **The definition of R does not specify how the contextual reasoning module R should be implemented**. Indeed, this is a design choice which requires to tradeoff between expressive power and computational complexity.

A context-based architecture has three main advantages. First, it decouples contextual information and reasoning from the common robotic tasks (e.g.

navigation, mapping, and so on), thus it is generic and can be adapted to several scenarios. Second, the reasoner R functionalities are provided without modifying the S modules.

Our contextual architecture is an example of *heterogeneous layered architecture*. Thus, on the one hand, uncertainty is managed at a numerical level by S modules, based on specific methods to deal with uncertainty (e.g. SLAM techniques for localization and mapping uncertainty). On the other hand, information management and decision making is handled at a symbolic level through R modules. Concerning the robustness of the overall system, some considerations are in order. First of all, as already mentioned, the system reactivity is not affected because the contextual loop is performed at a low frequency. On the other hand, in certain circumstances, the use of contextual knowledge can improve reactivity, by allowing fast detection of environment changes (e.g. opening door), exploiting the information about the dynamics of objects. Furthermore, it should be noted that planning is done for few steps, thus avoiding long, failure prone plans. Finally, symbolic representations are more robust than numerical representations, as their discrete characterization partially overbears numerical data uncertainty and data noises.

8.3   A Context-based System for Rescue

In this section we present a robotic system, based on our context-based architecture, applied to urban search and rescue. Rescue operations require several tasks, such as navigation, exploration, mapping, localization, victim detection, and so on. Calisi et al. (2008c) show the effectiveness of a contextual architecture with respect to navigation and mapping tasks; here we focus on the improvement of the exploration and victim search, given some contextual knowledge. RoboCup Rescue Virtual Robots competitions aim at boosting robotic systems for urban search and rescue, by realistic simulations of rescue missions. In particular, competitors are provided with an initial assessment of the environment. Hence, context-based systems can exploit this knowledge to improve their performances in search and rescue activities.

In RoboCup Rescue, contextual knowledge, that is known before the mission, is coded into a georeferenced map, using a format standardized by the NIST[1], as reported by Jacoff et al. (2000). This "a priori" map contains coarse-grained knowledge about the difficulty levels concerning mobility and victim detection. A similar map can be assumed in realistic scenarios as well. In fact, it can be obtained from several sources:

---

[1]http://www.nist.gov

- first responder teams sending a first initial assessment to the control station;

- aerial views, possibly acquired by an UAV flying over the disaster area;

- cadastral maps of the environment;

- well known information about the disaster area (e.g. population density, presence of public spots).

In particular, focusing on first responder teams, they could effectively acquire an initial assessment through a responsive and fast system like our semantic-driven tangible interface, presented in Chapter 7. In fact, it would allow the rescue personnel to move within safe areas of the environment, meanwhile signaling relevant landmarks, dangerous areas, unstable structures, and so on. For example, a hard mobility region can contain slopes, ruined ground, holes, stairs, cluttered zones. A hard victim detection area can contain victims occluded by objects, moving victims, or situations where the victim detection subsystem can detect many false positives (nothing is stated about the probability to find victims). As we will see in the following, other contextual knowledge is inferred from S modules (through $T_{d/c}$ modules) during the mission. The main problems of the exploration in unstructured environments is that it requires the ability to avoid difficult areas, where the robot could stall (e.g., typical stall conditions are objects blocking the robot's motion, lack of reachable frontiers, rollovers, and so on). Even if those places could contain victims, a robot blocked in a hole cannot notify any human operator about the presence of any victim. Exploiting contextual knowledge allows to reduce risks, and to implement smarter heuristics to detect victims; for example, taking snapshots of hard zones, instead of moving the robot inside them.

Procedures for sensing victims based on artificial vision are prone to false positives, that need a further analysis by human operators in order to distinguish real victims from false alarms. This is accomplished by taking pictures of the areas where the victim sensor detects something interesting, to be reported to human operators. Moreover, in hard victim detection areas, it can be difficult also for a human operator to assess the presence of a victim, because of occluding objects or dust covering everything; in such cases, one photo could be not enough. Given that the robot is in this situation, it is possible to take photos from different points of view, or to take a panoramic photo from a distance. All these problems, as well as some heuristics to solve them, can be represented in a context-based system as the one proposed here, by defining the type of knowledge that is necessary to acquire and the rules to deal with it, in order to select the best parameters for S modules to improve exploration and victim detection.

Contextual Knowledge: R Module

In the following, the reasoner R is a rule-based engine (i.e., First Order Horn Clauses) implemented in PROLOG, that contains a set of facts concerning the environment and a set of rules. A rule is composed by a condition which, if verified, causes an effect:

$$\textbf{IF } \alpha \textbf{ THEN } PARAMETER = value$$

where α is a formula representing the condition of the rule.

In order to suitably model mission requirements, we need to represent spatial and temporal variables. Consequently, we allow a rule to include variables and function symbols. For example, for a given value of the spatial variable Pos, the function $mobility(Pos)$ returns the difficulty in navigating in Pos. Spatial and temporal variables predicate about conditions and events that happen in a certain place and at a certain time. Consider, for example, to use the contextual knowledge to assess the mobility hardness in frontier-based exploration. Frontiers, according to Yamauchi (1997), are regions on the boundary between open space and unexplored space. Given a frontier F located at $PosF$, the function $mobility(PosF)$ retrieves the frontier's mobility level. As for temporal variables: if we had a function $batteryLevel(T)$, which returns the battery level at a given time T, we could set a battery threshold beyond which the robot is forced to come back to its home position. It is also possible to imagine more complex situations where, using spatial or temporal variables, the robot can behave in different ways for different places or time intervals. Consider the following example which involves spatial variables: a robot is moving towards a target point in the environment, following a trajectory defined as a set of intermediate points. At a given time t, the robot is localized in the environment at $CurrentPose$, and it adapts its speed considering the mobility hardness of the next intermediate point to reach. If the next intermediate point has an easy mobility level, then the robot sets a high speed, otherwise it slows down, as described in the following rules:

IF $robotPose(CurrentPose, t) \wedge plan(CurrentPose, NextPose)$
           $\wedge\, mobility(NextPose) == hardMobility$
THEN $MAX\_SPEED = lowSpeed$

IF $robotPose(CurrentPose, t) \wedge plan(CurrentPose, NextPose)$
           $\wedge\, mobility(NextPose) == easyMobility$
THEN $MAX\_SPEED = highSpeed$

where $plan(CurrentPose, NextPose)$ tells whether the navigation path $CurrentPose$ and $NextPose$ is part of the current $plan$, given that at the current

time t the robot is in position $CurrentPose$. Similarly, it is possible to define rules with temporal variables. Consider a function $robotStalled(T)$, indicating a robot stall condition at time time $T - \Delta(T)$. If at time T, which is the current time, the robot is not stalled anymore, knowing that he just escaped from a stall, we can moderately increase the speed in order not to stall again:

IF $robotStalled(T - \Delta(T)) \ \wedge \ \neg robotStalled(T) \wedge now(T)$
THEN $MAX\_SPEED = mediumSpeed$

Contextual Function	Meaning	Detected by	Spat./Temp. Info
smallRamp	robot encounters a small ramp	Elevation Mapper	Nome
bigRamp	robot encounters a big ramp	Elevation Mapper	None
robotStalled	robot stalled	Motion Planner	None
mobility(Pos)	location *Pos* has a certain mobility difficulty	A Priori Map	Spatial
victims(Pos)	location *Pos* has a certain victims detection difficulty	A Priori Map	Spatial
batteryLevel(T)	robot battery level at time T	Battery Controller	Temporal
detectedVictim(Pos)	victim sensor detected a possible victim at position *Pos*	Victim Sensor	Spatial

Table 8.1: The predicates used for contextual reasoning, the S module involved in the detection of the corresponding perception, and their spatial and temporal relevance.

Table 8.1 reports the predicates and the functions used to assert facts extracted from S modules and turned into symbolic knowledge by $T_{d/c}$, additionally specifying whether the type of information is spatial or temporal. The facts acquired by S modules (reported in the third column of the same

table) characterize the self-diagnosis capability of our system, which is the feedback loop of the context-based architecture. Thus, the *elevation mapper* (which builds a representation of the ground surface topography using two differently tilted lasers) detects the presence of small or big ramps, which will be represented with the symbolic formulas *smallRamp* and *bigRamp*. The *victim sensor* detects possible victims and the *motion planner* notifies any robot stall condition. The contextual output of R subsystem (as reported in Table

Module	Parameter	Values
Navigator	MOTION_PLANNER	{fine,coarse}
	MAX_SPEED	{low,medium,high}
	MAX_JOG	{low, medium, high}
Mapper	MAP_ENABLED	{true,false}
	SCAN_MATCH	{on,off}
	ELEVATION_MAPPER	{on,off}
Exploration	EXPL_ENABLED	{true,false}
	MOB_WEIGHT	{low,high}
	VICT_WEIGHT	{null,low,medium,high}
	DIST_WEIGHT	{low,high}
	INVALIDATE_FRONTIER	{true,false}
Victim Manager	MULTI_PHOTO	{on,off}
	TAKE_SNAPSHOT	{true,false}
	INCREASE_SNAP_DISTANCE	{true,false}

Table 8.2: The parameters produced by the R modules. They will be transduced by the $T_{c/d}$ modules in numeric parameters for the S modules.

8.2) is then transduced into some numeric parameters (through $T_{c/d}$) for the modules of S.

Robot Functionalities: S Modules

The parameters MOB_WEIGHT, VICT_WEIGHT and DIST_WEIGHT are the weights that the reasoning modules R estimate using the robot's position and the *a priori* map of the environment. Then, the exploration module in

S will use them to select the "best" unvisited frontier within a set of candidates. The criteria to state which is the best frontier are the distance from the robot estimated position, the mobility hardness of the frontier location and the victim detection hardness. These three parameters are related each other, depending on the contextual knowledge about the robot's position, which determines the weights. For example, if the robot is located in an easy mobility area, then it is reasonable to give more importance to those areas where mobility remains easy, but where victims would be easily detectable, because there is no mobility problem in the present circumstances. On the other hand, if the robot is in a hard mobility zone, then it is very important to leave the area as soon as possible, hence mobility will have a greater weight. If the contextual knowledge identifies some zones (e.g. holes or stairs) as critical, then setting INVALIDATE_FRONTIER to true tells the exploration module to discard a given frontier. If the victim sensor measures a possible victim in a zone that the reasoning module recognizes as hard for detection, then MULTI_PHOTO parameter is switched on, to take photos from different points of view, in order to avoid possible occluding objects. Furthermore, the parameter IN-CREASE_SNAP_DISTANCE is also enabled, to take a panoramic snapshot from a distance, in case of elevated victims. Finally, if the robot encounters a hard mobility zone, while moving towards a possible victim, the it stops and take a snapshot from where it is (enabling TAKE_SNAPSHOT parameter), without stopping a failure-prone exploration of a difficult area. Some of the rules of the reasoning module that activate the heuristics introduced before are reported in Table 8.3. On the one hand, the context-based architecture allows for reasoning about contextual information. On the other hand, complex behaviors, which typically require a high interaction among several robotic modules, are controllable without the need to hardcode either the interaction, or the reasoning process. For example, the victim manager communicates the presence of a new potential victim, and the reasoner asserts that she is localized in a hard mobility area (the last row of Table 8.3). It then commands the navigator module to move towards the victim, meanwhile acting on the victim manager to increase the snapshot radius, to avoid the hard mobility zone. Nevertheless, the navigator communicates a robot stall condition, expressed by the function *robotStalled*. Again, the reasoner activates a more precise navigator to let the robot unstall, concurrently stopping the exploration module. The robot manages to unstall, but its battery level (*batteryLevel(t)*) is low, thus the reasoner module commands the navigator to go to the closest communication point, to send the acquired information. This example shows both how the reasoning system controls concurrently more then one robotic module, and how it manages the interaction and the information exchange among them. The last rule of the same table reports a typical example of how a spatial function, such as *mobility*, is used on two different locations (both

IF	*mobility(RobotPos) == easyMobility ∧ victims(RobotPos) == easyIdVictims*	
THEN	**Navigator:**	MAX_SPEED = highSpeed
		MAX_JOG = highJog
		MOTION_PLANNER = coarsePlanner
	Mapper:	SCAN_MATCH = on
		MAP_ENABLED = on
		ELEVATION_MAPPER = on
	Exploration:	DIST_WEIGHT = highDistWeight
		MOB_WEIGHT = lowMobWeight
		VICT_WEIGHT = highVictWeight
	Victim Manager:	MULTI_PHOTO = false
IF	*mobility(RobotPos) == hardMobility ∧ victims(RobotPos) == hardIdVictims*	
THEN	**Navigator:**	MAX_SPEED = lowSpeed
		MAX_JOG = lowJog
		MOTION_PLANNER = finePlanner
	Mapper:	SCAN_MATCH = off
		MAP_ENABLED = off
		ELEVATION_MAPPER = on
	Exploration:	DIST_WEIGHT = lowDistWeight
		MOB_WEIGHT = highMobWeight
		VICT_WEIGHT = nullVictWeight
	Victim Manager:	MULTI_PHOTO = true
IF	*robotStalled*	
THEN	**Navigator:**	MOTION_PLANNER = finePlanner
	Exploration:	EXPL_ENABLED = false
IF	*currentVictim(Victim) ∧ detectedVictimPos(Victim, VictimPos) ∧*	
	mobility(RobotPos) != hardMobility ∧ mobility(VictimPos) == hardMobility	
THEN	**Navigator:**	MOTION_PLANNER = finePlanner
	Victim Manager:	INCREASE_SNAP_RADIUS = true

Table 8.3: Some of the rules represented in the reasoning module *R*. In case a rule is matched, it is reported the set of modules and parameters involved by the triggered behavior.

the robot and the possible victim position), and applied to select a particular heuristic (in this case, to avoid to enter in the hard mobility area and take the photo from a distance).

8.4 Experimental Evaluation

8.4.1 Experiment Design

The proposed system has been tested with the USARSim simulator; the chosen environment is an indoor map used during RoboCup German Open 2008, that is reported in Figure 8.2. The contextual architecture (i.e., the inclusion of $T_{d/c}$, R and $T_{c/d}$ modules) has been added to a pre-existent system (hence taken as the S modules). Both the pre-existent S modules and the new R, $T_{d/c}$ and $T_{c/d}$ modules have been implemented using the OpenRDK framework. The experiments are performed in 20 runs, 10 with context-based

Figure 8.2: The environment adopted for the experimental evaluation, designed for the RoboCup German Open 2008.

architecture and 10 without, each one 15 minutes long[2]. The task is to explore the environment and look for possible victims. The robot used in the experiments is a P2AT wheeled robot, equipped with a horizontal laser scanner (used to build the 2D map of the environment and to localize the robot) and another laser scanner, that is inclined on the sagittal plane (this is used to build an elevation map in order to detect objects not lying on the horizontal laser plane). Moreover, the "victim sensor" simulates an image processing module, whose goal is to find human victims. Both the contextual-based system and the non-contextual-based one run without any human control; thus, in case of robot stall, the systems must solve the problem autonomously. Both systems are provided with the *a priori* map of the environment. It is beyond the scope of this chapter to evaluate how this map has been acquired. As already motivated in Section 8.3, it is realistic to assume that it has been built by the initial assessment of rescue teams equipped with our semantic-driven tangible interface, which has been presented in Chapter 7. It must be noted that the R module is totally decoupled from the S modules. The interaction between the reasoning and the robotic layer (R and S respectively) is only based on the feedback provided by S subsystem and the parameters set by the reasoning module, according to our definition of context-based architecture (see Section 8.2). Concerning the context-free system, its exploration exploits *a priori* information, but the mobility and victims weights are fixed (they are set to "highMobWeight" and "midVictWeight" respectively, as these values turned out to be a good trade-off after experimental evaluation). In fact, the robot does not consider any type of contextual knowledge to weight in different ways the reachable frontiers. Moreover, the parameter configuration used in the context-free system adopts a single snapshot policy, because the multi photo approach consists in taking the photo from different points of view, that could be difficult in hard mobility areas, and requires more time to be accomplished.

8.4.2   Experimental Results

The results of the experiments are shown in Table 8.4, where C denotes the context-based system, while NC the context-free one. It can be noticed that, after 15 minutes, the explored area with the context-based system is 20% greater than the area covered with the context-free system and, because of a low variance, it supports a systematic tendency, rather then an occasional result influenced by particular runs. Furthermore, it must be noted that the environment is so large that, in each run, the robot explored different areas,

---

[2]Experiment   material   is   available   at   http://www.dis.uniroma1.it/ randelli/index.php?id=11.

Run	Check1 (5 min.)	Check2 (10 min.)	Check3 (15 min.)	Victims
C_1	7146.27	10164.5	10276.7	3
C_2	7613.97	8538.87	8541.6	1
C_3	6918.8	7589 (stalled)	7589 (stalled)	2
C_4	4502.93	5686.27	5947.27	1
C_5	5657.4	8160.17	8766.03	1
C_6	6753.07	8109.8	8109.8	3
C_7	6663.03	9232.87	11071.57	3
C_8	6287.63	6943.8	7520.93	2
C_9	3908.43	4627.73	8094.13	1
C_{10}	7602.77	10395.73	11389.77	2
Avg.	**6305.43**	**7944.87**	**8730.68**	**1.9**
Std.	**1316.27**	**1386.54**	**973.06**	**0.72**
NC_1	5692.4	10030.2	10723.23 (stalled)	1
NC_2	6562.4	6641.47 (stalled)	6641.47 (stalled)	1
NC_3	3134.07	3721.5	4426.9 (stalled)	1
NC_4	7055.73	7342.63 (stalled)	7342.63 (stalled)	1
NC_5	6437.57	6853.93	9341.5	1
NC_6	7960.6	7968.27 (stalled)	7968.27 (stalled)	1
NC_7	5908.93	6103.77 (stalled)	6103.77 (stalled)	1
NC_8	4934.7	4935.9 (stalled)	4935.9 (stalled)	1
NC_9	6730.97	8267.7 (stalled)	8267.7 (stalled)	1
NC_{10}	6084.87	6088.57 (stalled)	6088.57 (stalled)	1
Avg.	**6050.22**	**6795.39**	**7183.99**	**1**
Std.	**906.16**	**1297.15**	**1544.67**	**0.0**

Table 8.4: Amount of explored area (in m^2) and number of detected victims.

Context-based System	Context-free System
14min 20sec	8min 12sec

Table 8.5: Mean Time To Live (TTL) for a robot, before an unsolvable stall condition.

and some of them were more difficult than others. Even in these cases, the robot using context-based system managed to explore a larger section of the environment. An example is the third run, when the robot entered in a very difficult room where, after entering, an obstacle fell down blocking the exit, thus blocking the access to a large part of the environment. In order to motivate these results, it is important to consider the evolution of the exploration, both with and without context. Without any contextual knowledge, the system selects frontiers using fixed weights, and the robot can easily enter in hard mobility areas, thus falling in a hole or rolling over. Conversely, the context-based system dynamically adapts to the environment nearby the robot. If it is in a hard mobility area, the robot moves towards easier zones (weighting more the mobility factor rather then the victim detection and distance factors); as a consequence this approach results in exploring first safer areas and to defer the hard parts of the environment, thus trading robustness with a greater chance to detect victims. In search and rescue activities, it is very important for a robot to communicate to the base station what it found, and therefore it is better to skip some zones and return in communication with the base station. If a robot stalled in a section without any communication link, collected data would be lost. Avoiding any type of stall condition is also essential to improve the amount of explored area. In fact, if we consider larger time intervals (e.g. 25 minutes instead of 15 minutes), the robot controlled by the context-based system could continue to explore, because there has been no stall, as reported in the same table. On the other hand, without any contextual knowledge the robot does not recognize any stall condition and does not react to solve it, so its explored area would not increase. This can be quantified as reported in Table 8.5, through the mean time to live (out of a total amount of 15 minutes) that is significantly greater for the context-based robot than for the context-free robot.

The table reports the number of correctly detected victims. However, it should be noted that in search and rescue activities and moreover, in RoboCup competitions, the only accurate method to assess whether a possible victim is a false positive or not, is to take a photo. In the ten runs without context, six times the photo correctly captured the victim. In all the other cases, the photo did not allow any detection, because the victim was occluded or she was elevated over the ground. Using the context-based system, and taking into consideration the hardness of victim detection in some areas, it has been possible to implement multi-photo heuristics from different points of view, thus improving victim detection. Furthermore, since the exploration tends to avoid hard mobility areas, if the victim sensor detects something in a critical part, the robot can move without entering in it and take a photo in the direction of the possible victim.

In order to assess the statistical significance of the collected data, we used

Lilliefors test, which fits well for small data samples and tests whether they come from a normally distributed population with unknown mean. The test showed us that both context-based data samples (*P-value = 0.3692*)[3] and context-free data samples (*P-value = 0.5*) can be assumed to follow a normal distribution.

The two data sets where further reduced in their size, applying Grubb's test for outliers: C_4 (*P-value = 0.4192, G = 1.6288, U = 0.6725*)[4] and NC_1 (*P-value = 0.2384, G = 1.8126, U = 0.5944*), where the two identified samples. Finally, we applied the *Welch's t-test* on the two obtained data sets. The result is that we can assess with at least 95% of confidence that the two sets belong to two different normal distributions (*P-value = 0.007*), hence they are statistically different (see Figure 8.3(b)). Roughly speaking, this allows to consider the difference of performances between the context-based system and the context-free one as statistically relevant, and not due to casual factors.

8.5   Conclusions

In this chapter we presented an approach based on contextual knowledge for designing autonomous robotic systems. A context-based architecture decouples the contextual reasoner from the other components of the system. This has the advantage of centralizing the collection and use of the contextual knowledge, and fostering a seamless integration with the existent robotic modules, without hand-coding it.

Even more important, we validated how semantic knowledge, through reasoning, leads to significant improvements of the system performance and robustness in a challenging environment, such as urban search and rescue. The reasoner makes use of a set of rules that are written using a compact formalism, that can describe spatial and temporal events. We performed several experiments that show how an effective representation and use of knowledge about the context can lead to significant improvements of the system performance, in changing operational conditions. Urban search and rescue is a challenging

---

[3]*P-value* is the probability of obtaining a result at least as extreme as the one that was actually observed, given that the hypothesis is true. In this case, the hypothesis was that data samples belonged to a normal distributed population. If the *P-value* is greater than a given significance level (usually set to 0.05, which corresponds to 95% of belief), then it is not possible to reject the hypothesis, hence we cannot refuse that data samples belong to a normal distribution.

[4]G represents the difference between outliers and the mean divided by the standard deviation, while U is the ratio of sample variances with or without suspect outliers.
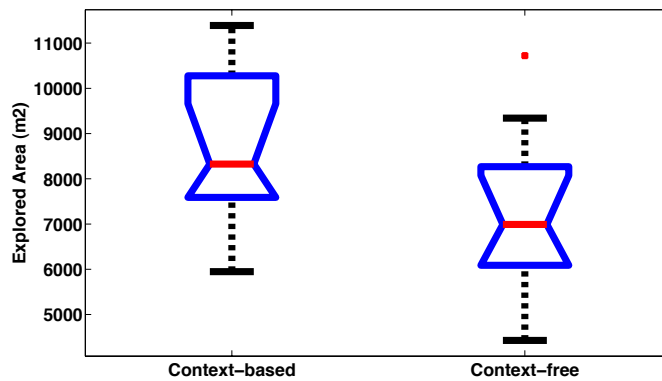
(a) *Context-based and context-free data samples. Red lines represent the medians of the sets, while red crosses are outliers.*
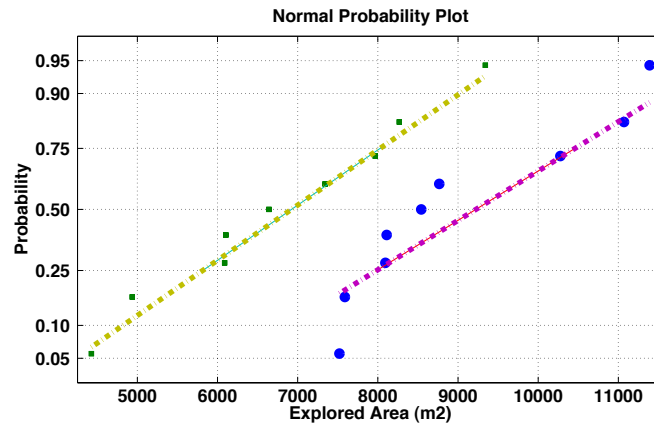
(b) *Data samples (without the two erased outliers) and their correspondent normal distributions. Circles are context-free samples, while squares are context-based samples. The two sets are sufficiently separable to assess that they belong to two different distributions.*

Figure 8.3: Context-based and context-free data samples statistics and their relative normal distributions.

robotic application, where the role of the human operator is mandatory, since it is still unrealistic to deploy completely autonomous robots to accomplish it. Nevertheless, by improving robot performance and robustness we are also significantly affecting human-robot awareness, in the sense that the human operator acquires a better assessment of the situation, she is less involved in low-level robot operation and, finally, she receives additional knowledge inferred by the reasoning component, represented in symbolic form, hence more comfortable to be interpreted.

Our architecture takes advantage of contextual information, in particular, acquired through a preliminary assessment involving rescue personnel. We already discussed that this step can be crucial for autonomous robots. This is where the human intervention, supported by tangible interfaces, leads to a significant improvement. This is possible through the tight synergy established between TUIs and semantic knowledge.

Chapter 9

Discussion

9.1 Conclusions

A key challenge in human-robot interaction for robot operation is the achievement of a proper level of awareness, in terms of human-robot awareness (from the human perspective), and situation awareness (from a robot point of view). Awareness is a multi-faceted aspect that depends on a crucial component of human-robot interaction: the robotic interface. A lack of awareness lessens the effectiveness of human-robot interaction, which in turn affects the performance of the robotic task to accomplish.

Up to now, awareness in human-robot interaction has been addressed with two main approaches: information visualization and enhanced output representations. Information visualization defines design best practices that optimize the data presentation layer in the interface, in order to lessen the cognitive effort of the operator, and highlighting unnoticed elements. Output representations provide a high-fidelity feedback of the robot surroundings and the overall environment. As well as conventional 2D maps, other approaches can improve this aspect, such as: 3D representations, camera feedback, virtual or augmented reality. Both these approaches are relevant for human-robot awareness, yet they are far from being definitive solutions, and they relate only to a subset of the aspects involving awareness.

In this thesis, we tackle the problem from a completely different perspective: investigating how awareness in HRI can be enhanced by adopting alternative interaction paradigms. In particular, we introduce tangible user interfaces as different input interaction, and semantic knowledge as output representation. This approach provides at least three significant contributions, analyzed in this thesis.

Our first contribution adopts TUIs for robot operation. We proved that tangible interfaces exhibit innovative interaction paradigms that, under specific circumstances, lower the operator cognitive effort in robot control. This shifts the operator focus on scene understanding, thereby improving human-robot awareness. In Chapter 4 we introduce low-level interaction paradigms: motion detection is considered for a robot teleoperation interface, while tactile feedback is adopted for a pre-attentive system that acquires relevant background information. We extensively evaluate both these metaphors to evince environment and mission-related conditions where they can bring real advantages in robotic systems. Finally, we consider intra-scenario operator mobility and assess whether, under this condition, a limited portable device can counterbalance a remote control through a powerful graphic interface. Chapter 5 moves our research towards high-level paradigms. In particular, we present a gesture-based interface for robot shared control, where the operator instructs the robot for a task, which executes it autonomously. We show how gesturing is a valuable paradigm for shared control, which further lowers the operator cognitive effort.

Our second contribution is a novel approach to the situation assessment process, which is based on the active human participation. By leveraging their cognitive skills in knowledge acquisition and grounding, humans overwhelm robot difficulties in acquiring knowledge autonomously. Thereby, the input component is not only used for robot control, but encompasses acquisition functionalities that enable human-robot cooperation. Chapter 6 presents this methodology by introducing tangible pointing interfaces, that is, tangible interfaces adopted for the natural and fast selection of relevant elements in a real environment. The main advantage is that the operator acquires information by directly interacting with real objects, and does not deal with any graphic display, which is unnatural. The pointing paradigm is further explored in Chapter 7, where we represent the acquired knowledge as semantic knowledge, through semantic-driven tangible interfaces. This is our ultimate contribution towards the acquisition of a high-level knowledge state.

Once acquired, our third and last contribution focuses on the exploitation of semantic knowledge, discussed in Chapter 8. Its benefit is twofold: first, it can be augmented by reasoning mechanisms that infer new knowledge. We proved how semantic knowledge, once acquired and exploited, improves robot situation awareness, which in turn lowers the intervention rate of the operator. Second, since semantic knowledge adopts symbolic representations, which are more comfortable for humans, it directly affects human-robot awareness, providing a higher level of information with respect to raw numerical data.

9.2 Future Work

Despite the results presented in this thesis, there is still room for different improvements to the problem of awareness in human-robot interaction.

Among the contributions of this thesis, the part dedicated to knowledge acquisition still requires an extensive and formal validation, such as in the case of the approaches presented in Chapter 6 and Chapter 7. Furthermore, it is worth noting that the scope of this thesis is limited to robot operation, thereby it would be interesting to extend this scope to other activities (e.g. manipulation). Finally, some assumptions in the knowledge acquisition techniques presented so far, mainly due to the noise of the sensors embedded in the adopted TUIs or to the lack of rangefinder sensors on-board, could be removed by adopting different interfaces or designing innovative ad-hoc devices.

Analyzing tangible user interfaces, we identified several interesting characteristics that match relevant issues of such a problem. However, there are still several aspects that need to be investigated to provide an exhaustive assessment of these devices within the context of awareness.

First, as future work, we would like to explore tangible devices as multi-human interfaces. In fact, in the human-computer community tangible interfaces have been effectively adopted for cooperative tasks, while this aspect has not been considered at all in robotics.

Once addressed multi-human robot interfaces, we would like to design effective cooperation strategies. We aim at two different approaches. On the on hand, analyzing the cooperation among operators equipped only with portable tangible interfaces and directly co-located with the robotic team. On the other hand, considering the relationship between operators with portable devices and remote personnel with desktop and powerful graphic interfaces. In the latter case, for example, intra-scenario operators acquire local awareness, meanwhile remote operators deal with global awareness.

Finally, we would like to implement a control transfer function based on the operator performance. Performance is assessed through the control feedback measured by tangible interfaces. In case of degrading performance, the control transfer function prevents dangerous events by transferring the control to another operator or to an autonomous system.

As for the usage of semantic knowledge, in this thesis we have mainly dealt with its acquisition, and we validated its effectiveness. However, as future work, we would like to introduce novel paradigms that enable human operators to easily access and query this knowledge. Our multimodal interface, proposed in Chapter 7, cues a possible solution in the integration of speech technologies and semantic knowledge. We believe that speech could be an effective and fast method to retrieve relevant knowledge, which contributes to acquiring a proper awareness.

List of Figures

147

List of Tables

Bibliography

J.A. Adams. Critical considerations for human-robot interface development. In *Proceedings of 2002 AAAI Fall Symposium*, pages 1–8, 2002. 41

L. Carlucci Aiello, D. Nardi, G. Randelli, and C.M. Scalzo. Suppose you have a robot. 2011. (to appear). 8

S. Balakirsky, C. Scrapper, S. Carpin, and M. Lewis. USARSim: providing a framework for multi-robot performance evaluation. In *Proceedings of the International Workshop on Performance Metrics for Intellingent Systems (PerMIS)*, Gaithersburg, MD, USA, 2006. 73

M. Beetz, T. Arbuckle, M. Bennewitz, W. Burgard, A. Cremers, D. Fox, H. Grosskreutz, D. Hahnel, and D. Schulz. Integrated plan-based control of autonomous service robots in human environments. *IEEE Intelligent Systems*, 16(5):56–65, 2001. URL citeseer.ist.psu.edu/beetz01integrated.html. 58

C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998. ISSN 1384-5810. 90

D. Calisi, A. Farinelli, L. Iocchi, and D. Nardi. Multi-objective exploration and search for autonomous rescue robots. *Journal of Field Robotics, Special Issue on Quantitative Performance Evaluation of Robotic and Intelligent Systems*, 24:763–777, August - September 2007. ISSN 1556-4959. doi: 10.1002/rob.v24:8/9. 58

D. Calisi, A. Censi, L. Iocchi, and D. Nardi. OpenRDK: a modular framework for robotic software development. In *Proc. of Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 1872–1877, September 2008a. ISBN 978-1-4244-2057-5. doi: 10.1109/IROS.2008.4651213. 72

D. Calisi, L. Iocchi, D. Nardi, C.M. Scalzo, and V. A. Ziparo. Context-based design of robotic systems. *Robotics and Autonomous Systems (RAS) - Special Issue on Semantic Knowledge in Robotics*, 56(11):992–1003, November 2008b. ISSN 0921-8890. doi: 10.1016/j.robot.2008.08.008. 126

D. Calisi, L. Iocchi, D. Nardi, C.M. Scalzo, and V. A. Ziparo. Contextual navigation and mapping for rescue robots. In *Proc. of IEEE Int. Workshop on Safety, Security & Rescue Robotics (SSRR)*, pages 19–24, Sendai, Japan, October 2008c. ISBN 978-1-4244-2031-5. doi: 10.1109/SSRR.2008.4745871. 128

D. Calisi, G. Randelli, A. Valero, L. Iocchi, and D. Nardi. RoboCupRescue - Virtual Robots League, Team SPQR-Virtual, Italy, 2008d. 44

D. Calisi, L. Iocchi, D. Nardi, G. Randelli, and V.A. Ziparo. Improving search and rescue using contextual information. *Advanced Robotics*, 23(9):1199–1216, 2009. ISSN 0169-1864. doi: 10.1163/156855309X452539. 8

J.A. Coelho Jr, E. Araujo, M. Huber, and R. Grupen. Contextual control policy selection. In CONALD'98 - Workshop on Robot Exploration and Learning, Pittsburgh, PA, June 1998., 1998. URL http://www.cs.cmu.edu/~conald/conald.shtml. 57

G. Cohen. *Memory in the real world*. Hove: Erlbaum, 1989. 40

N. Diolaiti and C. Melchiorri. Teleoperation of a mobile robot through haptic feedback. In *Haptic Virtual Environments and Their Applications, IEEE International Workshop 2002 HAVE*, pages 67–72. IEEE, 2002. 47

A. Diosi, G. Taylor, and L. Kleeman. Interactive slam using laser and advanced sonar. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1103–1108, Barcelona, Spain, 2005. 55

T. Djajadiningrat, K. Overbeeke, and S. Wensveen. But how, Donald, tell us how?: on the creation of meaning in interaction design through feedforward and inherent feedback. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 285–291. ACM, 2002. ISBN 1581135157. 26

C. Dornhege and A. Kleiner. Behavior maps for online planning of obstacle negotiation and climbing on rough terrain. Technical Report 233, University of Freiburg, 2007. 58

P. Dourish. *Where the action is: the foundations of embodied interaction*. The MIT Press, 2004. ISBN 0262541785. 26, 27

F. Driewer, M. Sauer, and K. Schilling. Design and evaluation of an user interface for the coordination of a group of mobile robots. In *17th International Symposium on Robot and Human Interactive Communication. RO-MAN 2008*, pages 237–242, August 2008. 42

J.L. Drury. *Extending usability inspection evaluation techniques for synchronous collaborative computing applications*. University of Massachusetts Lowell, 2002. ISBN 0493408002. 14

J.L. Drury, J. Scholtz, and H.A. Yanco. Awareness in human-robot interactions. In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, pages 111–119, 2003. 1, 15, 16, 29

J.L. Drury, J. Scholtz, and H.A. Yanco. Applying CSCW and HCI techniques to human-robot interaction. In *Proceedings CHI 2004 Workshop on Shaping Human-Robot Interaction*, pages 13–16, 2004. 18

J.L. Drury, J. Richer, N. Rackliffe, M.A. Goodrich, and MITRE CORP BEDFORD MA. Comparing situation awareness for two unmanned aerial vehicle human interface approaches. In *Proceedings of IEEE International Workshop on Safety, Security and Rescue Robotics (SSRR)*, Gaithersburg, MD, USA, 2006. 46

J.L. Drury, B. Keyes, and H.A. Yanco. LASSOing HRI: analyzing situation awareness in map-centric and video-centric interfaces. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 279–286. ACM, 2007. 20, 43

M.R. Endsley. Design and evaluation for situation awareness enhancement. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, volume 32, pages 97–101. Human Factors and Ergonomics Society, 1988. 19

M.R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995. ISSN 0018-7208. 1, 16, 17, 147

K.P. Fishkin. A taxonomy for and analysis of tangible interfaces. *Personal Ubiquitous Comput.*, 8(5):347–358, 2004. ISSN 1617-4909. doi: http://dx.doi.org/10.1007/s00779-004-0297-4. 30

G.W. Fitzmaurice, H. Ishii, and W.A.S. Buxton. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 442–449. ACM Press/Addison-Wesley Publishing Co., 1995. ISBN 0201847051. 21

T. Fong and C. Thorpe. Vehicle teleoperation interfaces. *Autonomous robots*, 11(1):9–18, 2001. ISSN 0929-5593. 32

T. Fong, C. E. Thorpe, and C. Baur. Advanced interfaces for vehicle teleoperation: Collaborative control, sensor fusion displays, and remote driving tools. *Autonomous Robots*, 11(1):77–85, 2001. 48

T. Fong, C.E. Thorpe, and C. Baur. Robot, asker of questions. *Robotics and Autonomous Systems*, 42(3-4):235–243, 2003. 48

C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernández-Madrigal, and J. González. Multi-hierarchical semantic maps for mobile robotics. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3492–3497, Edmonton, CA, 2005. Online at http://www.aass.oru.se/~asaffio/. 55

L. Gallo, G. De Pietro, and I. Marra. 3D interaction with volumetric medical data: experiencing the Wiimote. In *Proceedings of the 1st international conference on Ambient media and systems*, pages 1–6. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008. 31, 112

J.J. Gibson. *The ecological approach to visual perception*. Lawrence Erlbaum, 1986. ISBN 0898599598. 27

B. Giesler, P. Steinhaus, M. Walther, and R. Dillmann. Sharing skills: Using augmented reality for human-Robot collaboration. In *Proceedings of SPIE*, volume 5291, page 446, 2004. 46

B. Giesler, T. Salb, P. Steinhaus, and R. Dillmann. Using augmented reality to interact with an autonomous mobile platform. In *Proceedings of the IEEE International Conference on Robotics and Automation 2004*, volume 1, pages 1009–1014. IEEE, 2005. ISBN 0780382323. 46

S.A. Green, M. Billinghurst, X. Chen, and G.J. Chase. Human-robot collaboration: A literature review and augmented reality approach in design. *International Journal of Advanced Robotic Systems*, 5(1):1–18, 2008. 45

G. Grisetti, G.D. Tipaldi, C. Stachniss, W. Burgard, and D. Nardi. Speeding up rao blackwellized slam. In *ICRA*, pages 442–447, Orlando, FL, USA, 2006. 57

C. Guo and E. Sharlin. Exploring the use of tangible user interfaces for human-robot interaction: a comparative study. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing*

systems, pages 121–130, New York, NY, USA, 2008a. ACM. ISBN 978-1-60558-011-1. doi: http://doi.acm.org/10.1145/1357054.1357076. 50, 52

C. Guo and E. Sharlin. Utilizing Physical Objects and Metaphors for Human Robot Interaction. *Proceedings of Artificial Intelligence and Simulation of Behavior (AISB'08). Aberdeen, England (April 1–4, 2008). AISB Press*, 2008b. 52

D. Hahnel, D. Schulz, and W. Burgard. Map building with mobile robots in populated environments. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2002*, volume 1, pages 496–501. IEEE, 2002. ISBN 0780373987. 57

M. Hasanuzzaman, T. Zhang, V. Ampornaramveth, H. Gotoda, Y. Shirai, and H. Ueno. Adaptive visual gesture recognition for human-robot interaction using a knowledge-based software platform. *Robotics and Autonomous Systems*, 55(8):643–657, 2007. ISSN 0921-8890. 56

T. Herrmann. Blickpunkte und blickpunktsequenzen. In *Sprache & Kognition*, volume 15 of *217-233*. Huber, 1996. 40

J. Hertzberg and A. Saffiotti. Using semantic knowledge in robotics. *Robotics and Autonomous Systems*, 56(11):875–877, 2008. ISSN 0921-8890. doi: DOI:10.1016/j.robot.2008.08.002. URL http://www.sciencedirect.com/science/article/B6V16-4T72WWW-1/2/9edd0eb7357cb93ab0a7f0285979c469. Semantic Knowledge in Robotics. 54

A.T. Hjelmfelt and M.A. Pokrant. Coherent tactical picture. *CNA RM*, pages 97–129, 1998. 18

Eva Hornecker and Jacob Buur. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, pages 437–446, New York, NY, USA, 2006. ACM. ISBN 1-59593-372-7. doi: http://doi.acm.org/10.1145/1124772.1124838. URL http://doi.acm.org/10.1145/1124772.1124838. 26

L. Iocchi and T. van der Zant. Robocup@home: Adaptive benchmarking of robot bodies and minds. In *Proceedings of the International Conference on Simulation, Modeling and Programming for Autonomous Robots*, pages 171–182, 2010. 122

H. Ishii. The tangible user interface and its evolution. *Communications of the ACM*, 51(6):32–36, 2008. ISSN 0001-0782. 20

H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 234–241. ACM, 1997. ISBN 0897918029. 21, 23, 24, 52, 147

A. Jacoff, E. Messina, and J. Evans. A standard test course for urban search and rescue robots. In *Proceedings of the Performance Metrics for Intelligent Systems Workshop*, pages 499–503, Gaithersburg, Maryland, USA, 2000. 128

M.V. Jensen, J. Buur, and T. Djajadiningrat. Designing the user actions in tangible interaction. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*, pages 9–18. ACM, 2005. ISBN 1595932038. 25

I. Kamon, T. Flash, and S. Edelman. Learning to grasp using visual information. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, volume 3, pages 2470–2476, April 1996. ISBN 0-7803-2988-0. doi: 10.1109/ROBOT.1996.506534. 58

H. Kaymaz-Keskinpala and J.A. Adams. Objective data analysis for a pda-based human robotic interface. In *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics*, pages 2809–2814, 2004. 48

C.C. Kemp, C.D. Anderson, H. Nguyen, A.J. Trevor, and Z. Xu. A point-and-click interface for the real world: laser designation of objects for mobile manipulation. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 241–248. ACM, 2008. 51

M.I.J. Kim and M.L.O.U. Maher. Comparison of designers using a tangible user interface and a graphical user interface and the impact on spatial cognition. *Human Behaviour in Design 05*, page 81, 2005. 31

D. Kirsh. Complementary Strategies: Why we use our hands when we think. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, pages 212–217. Lawrence Erlbaum Associates, 1995. 28

M. Koeda, Y. Matsumoto, and T. Ogasawara. Annotation-based rescue assistance system for teleoperated unmanned helicopter with wearable augmented reality environment. In *Safety, Security and Rescue Robotics, Workshop*, pages 120–124, 2005. 46

G.J.M. Kruijff, H. Zender, P. Jensfelt, and H.I. Christensen. Clarification dialogues in human-augmented mapping. In *Proceedings of the 1st Annual Conference on Human-Robot Interaction (HRI'06)*, Salt Lake City, UT, March 2006. 55

G.J.M. Kruijff, H. Zender, P. Jensfelt, and H.I. Christensen. Situated dialogue and spatial organization: What, where... and why. *International Journal of Advanced Robotic Systems*, 4(2):125–138, 2007. 56

P. Lapides, E. Sharlin, and M. Costa Sousa. Three dimensional tangible user interface for controlling a robotic team. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 343–350. ACM, 2008. 52

S. Lee, G.S. Sukhatme, G.J. Kim, and C.M. Park. Haptic control of a mobile robot: A user study. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, volume 3, pages 2867–2874. IEEE, 2002. 47

J. Looser, M. Billinghurst, R. Grasset, and A. Cockburn. An evaluation of virtual lenses for object selection in augmented reality. In *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, pages 203–210. ACM, 2007. 31

A. Loutfi, S. Coradeschi, M. Daoutis, and J. Melchert. Using knowledge representation for perceptual anchoring in a robotic system. *International Journal on Artificial Intelligence Tools*, 17(5):925–944, 2008. 56

H.J. Luinge and P.H. Veltink. Measuring orientation of human body segments using miniature gyroscopes and accelerometers. *Medical and Biological Engineering and Computing*, 43(2):273–282, 2005. URL http://doc.utwente.nl/61405/. 103

J. Maida, C. Bowen, and W. Pace. Enhanced lighting techniques and augmented reality to improve human task performance. *NASA Tech Paper TP-2006-213724 July*, 2006. 46

J. Marciniak and Z. Vetulani. Ontology of Spatial Concepts in a Natural Language Interface for a Mobile Robot. *Applied Intelligence*, 17(3):271–274, 2002. ISSN 0924-669X. 56

W. Mark. The computer for the 21st century. *Scientific American*, 265(3):66–75, 1991. 21

P. Marshall, Y. Rogers, and E. Hornecker. Are tangible interfaces really any better than other kinds of interfaces? In *Workshop on Tangible User Interfaces in Context and Theory (CHI 2007)*, 2007. 31, 49

O. Martínez Mozos and W. Burgard. Supervised learning of topological maps using semantic information extracted from range data. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2772–2777, Beijing, China, 2006. 55

L. Messing and R. Campbell. *Gesture, speech, and sign.* Oxford University Press, 1999. 27

J. Mlích. Wiimote gesture recognition. In *Proceedings of the 15th Conference and Competition STUDENT EEICT 2009 Volume 4*, pages 344–349. Faculty of Electrical Engineering and Communication BUT, 2009. ISBN 978-80-214-3870-5. URL http://www.fit.vutbr.cz/research/view_pub.php?id=8933. 50

J. Modayil and B. Kuipers. Bootstrap learning for object discovery. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 742–747, 2004. 56

M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceedings of the International Conference of the American Association for Artificial Intelligence (AAAI)*, Edmonton, Canada, 2002. 57

R. Murphy. Human-robot interaction in rescue robotics. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34 (2):138–153, 2004. doi: 10.1109/TSMCC.2004.826267. URL http://dx.doi.org/10.1109/TSMCC.2004.826267. 37

R. Murphy. *Introduction to AI robotics.* The MIT Press, 2000. ISBN 0262133830. 34

P. Newman, D. Cole, and K. Ho. Outdoor SLAM using visual appearance and laser ranging. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1180–1187, Orlando, FL, USA, 2006. 57

C.W. Nielsen and M.A. Goodrich. Comparing the usefulness of video and map information in navigation tasks. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI 2006*, pages 95–101, 2006. 42

C.W. Nielsen, M.A. Goodrich, and R.W. Ricks. Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, 23(5): 927–941, 2007. 42

D.A. Norman. *Cognitive artifacts*, pages 17–38. Cambridge Univ Pr, 1991. 27

D.A. Norman. Affordance, conventions, and design. *interactions*, 6(3):38–43, 1999. ISSN 1072-5520. 27

A. Nüchter, O. Wulf, K. Lingemann, J. Hertzberg, B. Wagner, and H. Surmann. 3D Mapping with Semantic Knowledge. In *RoboCup 2005: Robot Soccer World Cup IX*, 2005. 55

D.R. Olsen and M.A. Goodrich. Metrics for evaluating human-robot interactions. In *Proceedings of the 2003 Performance Metrics for Intelligent Systems (PerMIS) Workshop*, 2003. 19

J. Patten and H. Ishii. A comparison of spatial organization strategies in graphical and tangible user interfaces. In *Proceedings of DARE 2000 on Designing augmented reality environments*, pages 41–50. ACM, 2000. 28

D. Perzanowski, A.C. Schultz, W. Adams, E. Marsh, and M. Bugajska. Building a multimodal human-robot interface. *Intelligent Systems, IEEE*, 16(1): 16–21, 2005. ISSN 1541-1672. 114

G. Randelli and D. Nardi. Introducing ontology best practices and design patterns into robotics: USAREnv. In *Proceeding of the 2010 conference on Modular Ontologies: Proceedings of the Fourth International Workshop (WoMO 2010)*, pages 67–80, Amsterdam, The Netherlands, The Netherlands, 2010. IOS Press. ISBN 978-1-60750-543-3. doi: 10.3233/978-1-60750-544-0-67. 8

G. Randelli, L. Marchetti, F. Marino, and L. Iocchi. Multi-agent behavior composition through adaptable software architectures and tangible interfaces. In Javier Ruiz-del Solar, Eric Chown, and Paul Ploger, editors, *RoboCup 2010: Robot Soccer World Cup XIV*, volume 6556 of *Lecture Notes in Computer Science*, pages 278–290. Springer Berlin / Heidelberg, 2011a. doi: http://dx.doi.org/10.1007/978-3-642-20217-9_24. URL http://dx.doi.org/10.1007/978-3-642-20217-9_24. 10.1007/978-3-642-20217-9_24. 8

G. Randelli, M. Venanzi, and D. Nardi. Tangible interfaces for robot teleoperation. In *Proceedings of the 6th international conference on Human-robot interaction*, HRI '11, pages 231–232, New York, NY, USA, 2011b. ACM. ISBN 978-1-4503-0561-7. doi: http://doi.acm.org/10.1145/1957656.1957746. URL http://doi.acm.org/10.1145/1957656.1957746. 7

G. Randelli, M. Venanzi, and D. Nardi. Evaluating tangible paradigms for ground robot teleoperation. In *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, 2011c. (to appear). 7

O.J. Rosch, K. Schilling, and H. Roth. Haptic interfaces for the remote control of mobile robots. *Control Engineering Practice*, 10(11):1309–1313, 2002. 47

P. Rouanet, J. Béchu, and P.Y. Oudeyer. A survey of interfaces using handheld devices to intuitively drive and show objects to a social robot. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Toyama, Japan, September 2009. 49, 53

J. Scholtz. Evaluation methods for human-system performance of intelligent systems. In *Proceedings of the 2002 Performance Metrics for Intelligent Systems (PerMIS) Workshop*, 2002. 18

J. Scholtz. Theory and evaluation of human robot interactions. In *Proceedings of HICSS '03 Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03)*, 2003. 32

E. Sharlin, B. Watson, Y. Kitamura, F. Kishino, and Y. Itoh. On tangible user interfaces, humans and spatiality. *Personal and Ubiquitous Computing*, 8(5):338–346, 2004. ISSN 1617-4909. 50, 52

T. Sko and H. Gardner. The Wiimote with multiple sensor bars: creating an affordable, virtual reality controller. In *CHINZ '09: Proceedings of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction*, pages 41–44, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-574-1. doi: http://doi.acm.org/10.1145/1577782.1577790. 113

M. Skubic, C. Bailey, and G Chronis. A Sketch Interface for Mobile Robots. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 918–924, 2003. 48

M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock. Spatial language for human-robot dialogs. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34(2):154–167, 2004. ISSN 1094-6977. 56

T.H. Song, J.H. Park, S.M. Chung, S.H. Hong, K.H. Kwon, S. Lee, and J.W. Jeon. A study on usability of human-robot interaction using a mobile computer and a human interface device. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pages 462–466, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-862-6. doi: http://doi.acm.org/10.1145/1377999.1378055. 53

C. Stachniss, O. Martínez-Mozos, and W. Burgard. Speeding-up multi-robot exploration by considering semantic place information. In *Proceedings of the IEEE International Conference on Robotics & Automation (ICRA)*, pages 1692–1697, Orlando, FL, USA, 2006. 58

C. Theobalt, J. Bos, T. Chapman, A. Espinosa-Romero, M. Fraser, G. Hayes, E. Klein, T. Oka, and R. Reeve. Talking to Godot: Dialogue with a mobile robot. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002)*, pages 1338–1343, 2002. 56

C. Thorpe and H. Durrant-Whyte. Field robots. *Robotics Research*, 6:329–340, 2003. URL http://dx.doi.org/10.1007/3-540-36460-9_22. 49

A. Torralba, K.P. Murphy, W.T. Freeman, and M.A. Rubin. Context-based vision system for place and object recognition. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 273–280. IEEE, 2003. ISBN 0769519504. 58

R. Triebel, P. Pfaff, and W. Burgard. Multi-level surface maps for outdoor terrain mapping and loop closing. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, 2006. 58

R.M. Turner. Context-mediated behavior for intelligent agents. *International Journal of Human-Computer Studies*, 48(3):307–330, 1998. URL citeseer. ist.psu.edu/turner97contextmediated.html. 57

B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. *IBM Systems Journal*, 39(3-4):915–931, 2000. ISSN 0018-8670. doi: http://dx.doi.org/10.1147/sj.393.0915. 22, 23, 147

B. Ullmer, H. Ishii, and R.J.K. Jacob. Token+ constraint systems for tangible interaction with digital information. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(1):81–118, 2005. ISSN 1073-0516. 30

A. Valero, G. Randelli, C. Saracini, F. Botta, and M. Mecella. The advantage of mobility: Mobile tele-operation for mobile robots. In *Proceedings of the AISB 2009 Convention*, 2009a. doi: 10.1.1.160.4359. 9, 44

A. Valero, G. Randelli, C. Saracini, F. Botta, and D. Nardi. Give me the control, i can see the robot! In *Proceedings of the IEEE International. Workshop on Safety, Security, and Rescue Robotics (SSRR 2009)*, pages 1–6, 2009b. doi: 10.1109/SSRR.2009.5424157. 8, 44

A. Valero, C. Saracini, F. Botta, and G. Randelli. Spatial processes in mobile robot teleoperation. *Cognitive Processing*, 10(2):338–341, 2009c. doi: 10. 1007/s10339-009-0299-2. URL http://www.springerlink.com/content/ 74q277101t38m6v2/. 8, 44

A. Valero, G. Randelli, F. Botta, M. Hernando, and D. RodrÃguez-Losada. Operator performance in exploration robotics. *Journal of Intelligent & Robotic Systems*, pages 1–21, 2011. ISSN 0921-0296. doi: http://dx. doi.org/10.1007/s10846-011-9539-7. URL http://dx.doi.org/10.1007/s10846-011-9539-7. 10.1007/s10846-011-9539-7. 8, 44

P. Varcholik, D. Barber, and D. Nicholson. Interactions and training with unmanned systems and the nintendo wiimote. In *Proceedings of the 2008 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 2008. 50

W.S. Wei. *Time Series Analysis*. Addison-Wesley, 2006. ISBN 0-321-32216-9. 102

P. Wellner, W. Mackay, and R. Gold. Back to the real world. *Communications of the ACM*, 36:24–26, July 1993. ISSN 0001-0782. doi: http://doi.acm. org/10.1145/159544.159555. URL http://doi.acm.org/10.1145/159544.159555. 21

S. Werner, B. Krieg-Brückner, H.A. Mallot, K. Schweizer, and C. Freksa. Spatial cognition: The role of landmark, route, and survey knowledge in human and robot navigation. In *Informatik Aktuell*, pages 41–50, 1997. 40

B. West, K. Welch, and A. Galecki. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman Hall / CRC Press, first edition, 2006. URL http://www-personal.umich.edu/~bwest/almmussp.html. 74

P. Wilson. *Computer supported cooperative work: an introduction*. Springer, 1991. ISBN 0792314468. 14

D.F. Wolf and G.S. Sukhatme. Semantic mapping using mobile robots. *Robotics, IEEE Transactions on*, 24(2):245–258, 2008. ISSN 1552-3098. 57

B. Yamauchi. A frontier based approach for autonomous exploration. In *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 1997. 130

H.A. Yanco and J. Drury. Classifying human-robot interaction: an updated taxonomy. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 3, pages 2841–2846. IEEE, 2005. ISBN 0780385667. 37

H.A. Yanco and J.L. Drury. "Where am I?" acquiring situation awareness using a remote robot platform. In *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics: The Hague, Netherlands, 10-13 October 2004*, pages 2835–2840, 2004. 42, 43

H.A. Yanco and J.L. Drury. A taxonomy for human-robot interaction. In *Proceedings of the AAAI Fall Symposium on Human-Robot Interaction*, pages 02–03, 2002. 37

H.A. Yanco, J.L. Drury, and J. Scholtz. Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. *Human–Computer Interaction*, 19(1):117–149, 2004. ISSN 0737-0024. 43

H.A. Yanco, B. Keyes, J.L. Drury, C.W. Nielsen, D.A. Few, and D.J. Bruemmer. Evolving interface design for robot search tasks. *Journal on Field Robotics*, 24(8-9):779–799, 2007. 43