18

Semiotics and Human-Robot Interaction

João Sequeira and M.Isabel Ribeiro
Institute Superior Técnico / Institute for Systems and Robotics
Portugal

1. Introduction

Creating artificial creatures capable of interacting with human beings, following standard social conventions and breaking them as humans do, is part of the technological expression of mankind. Around the 17th century renowned craftsman started producing mechanical automata with behavioral capabilities that imitated basic human skills, mainly related to locomotion and manipulation. A multitude of fictional robots were developed by science fiction authors since the early 20th century, most of them exhibiting behavioral capabilities far ahead of what science and technology would allow. Since then robots populate collective imagination and technological societies established as unconscious goal developing robots aiming at obtaining a human alter ego.

In addition to strict intelligence, a key feature of human beings, robotics also targets human like physical interaction properties such as locomotion. After a century of scientific research it seems clear that achieving intelligence in robotics requires mastering cognition, learning, reasoning, and physical interaction techniques. The Turing test to assess the intelligence of a generic machine can also be used to assess the intelligence of a robot. If human can be deceived by a robot, in a dialogue and also in a physical interaction with the environment, through locomotion or manipulation, then it passes the test. As a complementary argument, the way the interaction is performed might influence whether or not the robot qualifies as intelligent. For example, a robot can avoid obstacles in different manners, according to the environment conditions, and induce different perceptions in a human watching the motion. In the end, an intelligent robot must interact with an ordinary human being, probably not experienced in what concerns robotics, as if it were a human.

The robotics research community only recently started to pay attention to human robot interaction (HRI) as an independent research area. Besides the pure research interest, mass applications, both socially and economically relevant, are being envisaged for robots, namely as home companions, personnal assistants, security agents, office assistants, and generic workers. State of the art humanoid robotics is already capable of simple tasks in factory environments but the interaction abilities are still not up to pass a Turing test.

Generic HRI must assume that humans are inexperienced in what concerns robot motion and hence the interaction techniques robots should use must clone those used by humans among themselves. This suggests that models of human interactions be used to support the research and development of HRI models. In addition, these models might support human like schemes for interaction among robots themselves hence avoiding having to consider separate competences for each type of interaction.

Source: Human-Robot Interaction, Book edited by Nilanjan Sarkar,
ISBN 978-3-902613-13-4, pp.522, September 2007, Itech Education and Publishing, Vienna, Austria

The chapter reviews a number of topics directly related to HRI and describes a research effort to develop a HRI model inspired in semiotics concepts developed in linguistics to model interactions among humans. A set of experiments is presented to illustrate the ideas developed.

2. A brief overview of HRI related research

HRI evolved from the human-machine and human-computer interaction. Often it has been reduced to the design of interfaces aiming at optimizing specific performance indices. Nowadays, typical human strategies used to convey information, such as expressing emotions and specifying intentions through motion, are being also addressed in HRI research.

As an example, the design of keyboards is often subject to studies to optimize usability measures such as the time required by a human to type a benchmark sequence of keys, (Carroll, 2003). Interface design techniques have also rely on the study of maps of human thought obtained by cognitive psychology, (Raskin, 2000). Interfacing tools are always present in a robot control architecture though its synthesis does not aim directly at simplifying the interaction between robots and humans. Usability has been studied in (Ryu and Lee, 2006) in the context of map based interfaces. An agent based architecture for HRI based on an adaptive graphical interface is described in (Kawamura et al., 2003). The robot agent provides the human with the necessary information on the robot and environment. A commander agent maintains a model of the user that is used to decide the message forwarding policy from the human to the appropriate robot. A paradigm in which robots and humans cooperate through the abiliy to recognize emotions is described in (Rani and Sarkar, 2004). Universal user friendly human-computer interfaces were addressed in (Savidis, A. and Stephanidis, C., 2004). Physical indicators used in HRI analysis criteria can also be used in decision making, (Dautenhahn and Werry, 2002).

Robot control architectures have always been a key subject in robotics, fostering research work in multiple enabling areas, e.g., sensors, kinematics and dynamics modeling and control, and in specific functionalities, e.g., path planning and following, obstacle avoidance, world mapping, and localisation. During the 80's the concept of behavior gained wide visibility in robotics. The semantic content associated with the behavior concept seemed to indicate that robot missions could be easily specified almost as if using natural language (or, more generically, a natural interface). Despite multiple efforts to create a formal support for this concept[1] this has been an elusive concept in what concerns simplifying HRI. Human-computer interaction models have been used in (Scholtz, 2003) to define HRI models based on a set of roles, such as supervisor, operator and bystander. In a sense, these roles can be identified with the linguistic notion of behavior though they yield only weak guidelines for synthesis.

The research in robot control architectures is huge. Most of the general purpose architectures can be classified somewhere in the span of behavioral and functional models. The first tend to be specified in terms of models of global performance whereas the later use functional blocks to describe goal behaviors. For example, (Ogasawara, 1991) identifies five components in control architectures, namely, percepts, decomposition, strategies, arbitration

---

[1] A framework including a formal definition of behavior for generic dynamical systems can be found in (Willems, 1991).

and actions. Despite the semantic content identified with the names of the components, at the implementation level there are functional blocks such as map builders, obstacle avoidance and world modeling strategies, and user interfaces. An arbitration block controls the action to be executed. Other examples of single and multiple robot architectures can easily be found in the literature using concepts from artificial intelligence, biology, semiotics and economic trade markets (see for instance (Parker, 1998; Sequeira and M.I. Ribeiro, 2006a; Bias et al., 2005)).

HRI is also a key area in active surveillance systems. The development of interaction strategies that can be used both by robots and humans, reasonably independent of their relative skills, is likely to improve the performance of the systems. The vast majority of the existing commercial surveillance systems rely on three main components, namely (i) networks of fixed sensors covering the perimeter under surveillance, (ii) visual and keyboard interfaces as interaction tools, and (iii) human supervisors to handle contingency situations. When moving to robotics, a critical issue is that the devices must be able to interact with humans unskilled in what concerns robot specific issues, such as kinematics and dynamics.

First generation commercial surveillance systems rely mainly on networks of fixed sensors, e.g., CCTV systems and motion detectors, to acquire and send data directly to human experts. The development of computer vision led to smart cameras able to process images and extract specific features. Image processing techniques for detection and identification of human activities is an area with huge influence in the ability of robotic systems to interact and even socialize with humans. The surveys in (Valera and Velastin, 2005; Hu et al., 2004) identify key issues in image processing related to the surveillance problem, e.g., human identification using biometric data, the use of multiple cameras, and 2D/3D target modelling. An example of a network of portable video sensors is presented in (Kogut et al., 2003) to detect, track and classify moving targets, and gathering information used to control unmanned ground vehicles.

Robots have been employed in commercial surveillance systems mainly as mobile platforms to carry sensors. The PatrolBot (www.mobilerobots.com) is used in the surveillance of buildings like the Victoria Secret's headquarters at Columbus, USA, and in the United Nations building at Geneva, Switzerland. Mostitech (www.mostitech.com), Personal Robots (www.personalrobots.com), and Fujitsu (www.fujitsu.com) currently sell robots for domestic intruder detection (off the shelf video cameras, eventually with pan-tilt capabilities, constitute also simple robots that can be configured to detect intruders).

In military and police scenarios the robots are, in general, teleoperated to gather information on the enemy positions and in explosives ordinance disposal, (Nguyen and Bott, 2000; Everett, 2003). The Robowatch robot (www.robowatch.de) is supervised by a human through a graphic interface and allowed limited autonomy through information from ultrasound, radar, video and infrared sensors. These robots do not aim at cooperative operation with other robots or humans. Upon detecting unexpected events they just signal a human supervisor through the interface. The interfacing techniques are often developed to accomodate technical constraints, e.g., small number of sensors allowed due to power supply constraints, which might limit their usability. In difficult applications, such as bomb disposal, there are learning curves of several months (robot non-holonomy tends to be an issue in such applications).

In some examples with large number of robots, the interfacing aspects become much more relevant as having humans supervising of each of the individual robots may not be feasible. For example, the Centibots project, (Konolidge et al., 2004), aims at deploying a large number of robots in unexplored areas for world mapping, target searching and surveillance tasks. Distributed map building and fault tolerant communications are just two of the functionalities in each robot. The robots are organized hierarchically, in groups, each with a team leader, and are able to exchange data within the limited range of the communications system. There are four types of interaction allowed, null interaction, hypothesis generation, hypothesis testing and coordinated exploration. Basically, this corresponds to a negotiation strategy that controls the exchange of sensor data. Another example is given by the team of miniature robots with onboard cameras in (Rybski et al., 2000) for reconnaissance and surveillance tasks. The limited computational capabilities of these robots require that image processing and the decision processes are done offboard. The control of large groups of robots using loose interaction protocols in a surveillance task is discussed in (Khrisna and Hexmoor, 2003). Besides the common localisation, navigation and collision avoidance modules, the proposed architecture includes functionalities such as social notions, values, cooperative and shared reasoning and intruder detection.

Including social skills in robots is likely to facilitate their integration in human environments. A robot might be instructed to approach groups of people as a mean to demonstrate that it needs to communicate explicitly or just to acquire information on the group. This sort of social behavior matches typically human social behaviors; in some circunstances a single human tends to approach groups of people in order to foster interaction with the other humans in the group. Standard techniques can be used to design behaviors that convey information on the intentions of a robot to the outside environment (see for instance (Nicolescu and Mataric, 2001)). Still, current strategies to describe in a unified way the synthesis and detection/recognition from sensor data of such behaviours, both for humans and robots, do not yield user friendly interfacing.

3. Abstract concepts in HRI modeling

In general, robots and humans work at very different levels of abstraction. Developing new forms of representing human-robot interactions close to those used among humans, e.g., natural interfaces, is likely to yield robotic systems able to perform complex missions that currently can only be accomplished by humans.

Most of the usual interactions among machines, and robots in particular, are supported on well defined protocols to wrap and transport data. This means that every intervenient knows exactly what to do when it receives data, how to transmit it and what to do with it. Interactions among humans follow different principles. Often, information is exchanged using loosely defined protocols and symbolic information is exchanged either conveying explicit data or wrapping a particular meaning that later it will be inferred by the receiver.

The difficulties in creating a model for linguistic interactions with the above characteristics are obviously immense. Despite the research efforts, dating back to the work of Chomsky, Harman and others, (Chomsky, 1968; Harman, 1969), current natural interfacing tools are still not powerful enough to mimic human natural language capabilities.

Mapping high level abstract concepts into low level concrete objects requires a roadmap, i.e., a set of organizing principles. Category theory (CT) provides a suitable framework to represent the objects and relations among them. Other than providing deep formal results, CT clarifies these organizing principles.

Diagram 1 represents a model of a hierachy of abstractions (the level of abstraction increases towards the righthand side of the diagram). Each level follows a classical sense-think-act pipeline. The H_i objects represent the data perceived by the robot at each abstraction level. The $abst_i$ functors define the data processing between levels. The beh_i functors represent the decision processes on the perceived data. The A_i objects contain the information that directly affects the motion of the robot. The act_i functors stand for the processes that transform high level information into actuator controls. The circle arrows indicate endofunctors in each of the categories involved.

$$\begin{tikzcd}
H_0^{\circlearrowright} \arrow[r, "abst_0"] \arrow[d, "beh_0"'] & H_1^{\circlearrowright} \arrow[r, "abst_1"] \arrow[d, "beh_1"'] & \dots \arrow[r, "abst_n"] & H_n^{\circlearrowright} \arrow[d, "beh_n"'] \\
A_0^{\circlearrowright} & A_1^{\circlearrowright} \arrow[l, "act_0"] & \dots \arrow[l, "act_1"] & A_n^{\circlearrowright} \arrow[l, "act_n"]
\end{tikzcd} \tag{1}$$

At the lowest level of abstraction, H_0 includes objects such as configuration spaces, the A_0 contains the control spaces, and the beh_0 account for low level control strategies, e.g., motor control feedback loops. Coordinate transformations are examples of endomaps in H_0. At the intermediate levels, the H_i can represent data, services, and functionalities such as path planning and world map building algorithms. At the highest level of abstraction, H_n stands for the objects used in natural interactions, that is, information units exchanged during a natural interactions such as those occurring when using natural language. The A_n stands for high level processing of such units. In an implementation perspective, the diagram suggests the use of concurrent beh_i maps, operating in different levels of abstraction and competing to deliver their outputs to actuators. In a sense, it generalizes the well known idea of subsumption architecture.

Abstractions defined through projection maps have been used in the context of dynamical systems to represent hierarchies of models that preserve good properties, e.g., controllability and observability; see for instance (Stankovic and Siljak, 2002) for examples with linear time invariant systems, (Asarin and Maler, 1994) for a definition on discrete event systems, or (Asarin and Dang, 2004) for nonlinear continuous systems. Simulation and bisimulation maps also express notions of equivalence (from an external perspective) between systems; see (van der Schaft, 2004) on linear time invariant systems.

Diagram 1, accounting for multiple decision levels through the beh_i maps, raises interesting robotics problems, namely (i) defining criteria for the number of abstraction levels, and (ii) optimally distributing the information processing through the abstraction levels, i.e., designing the $abst_i$ and act_i maps. The diagram can also be interpreted as relating different perspectives for modeling and controlling robots[2] and hence might represent a unifying structure for robot control architectures. In a sense, each of the abstraction levels represents

---

[2] This idea of different perspectives to system modeling is common in information systems architectures, see for instance the IEEE 1471 standard, (IEEE, 2000).

a different perspective from where to look to a problem and all the perspectives contribute to the global outcome.

4. A semiotics based HRI model

Semiotics is a branch of general philosophy addressing the linguistic interactions among humans. It has been used by several authors mainly to provide guidelines for control architectures synthesis (see for instance (Meystel and Albus, 2002)) and often related to the symbol grounding concept of artificial intelligence. Categorical approaches to semiotics have been proposed in human-computer interaction area. See for instance (Neumuller, 2000) for an application of hypertext theory to world wide web, (Codognet, 2002) for research on machine-machine and human-human interactions over electronic media, and (Malcolm and Goguen, 1998) for an algebraic formulation for semiotics and its use in interface design.

The underlying idea in this chapter is (i) to use the concept of *semiotic sign* as the basis information unit used by humans to interact among themselves, (ii) to establish principles for processing these semiotic signs, and (iii) to define adequate maps between sets of signs and spaces of control variables.

Interaction using natural language is probably the most complete example of interaction among heterogeneous agents using semiotic signs. Semiotic signs evolved mainly from the work of C. S. Pierce and F. Saussure (see for instance (Chandler, 2003; Bignell, 1997)). The concepts developed by Pierce and Saussure differ slightly, with Pierce's model being more flexible that Saussure's model. The signs and morphisms defined among them form *sign systems.*

Following Pierce's work, signs can be of three categories, (Codognet, 2002; Malcolm and Goguen, 1998), (i) symbols, expressing arbitrary relationships, such as conventions, (ii) icons, such as images, (iii) indices, as indicators of facts or conditions. Signs defined in these three categories can represent any of the abstract entities, of arbitrary complexity, that are used by humans in linguistic interactions, (Bignell, 1997).

A generic semiotic sign, in any of the three classes above, encapsulates three atomic objects, named *meaning, object,* and *label,* and the relations between them. Under reasonable assumptions on the existence of identity maps, map composition, and composition association, signs can be modeled as a category. Diagram (2) illustrates the sign category, hereafter named SIGNS. A similar representation, known as the "semiotic triangle", is often used in the literature on semiotics (see for instance (Chandler, 2003)). For the sake of completeness it is worth to point that Saussure's model of a sign included only two objects, a signifier, defining, the form the sign takes, and a signified, defining, the concept it represents.

$$\begin{tikzcd}
L \arrow[r, "semantics"] \arrow[d, "syntactics"'] & M \\
O \arrow[ur, "pragmatics"'] &
\end{tikzcd} \tag{2}$$

The *Labels,* (L), represent the vehicle through which the sign is used, *Meanings,* (M), stand for what the users understand when referring to the sign, and *Objects,* (O), stand for the real objects signs refer to. The morphisms are named *semantics,* standing for the map that extracts the meaning of an object, *syntactics,* standing for the map that constructs a sign from a set of syntactic rules, and *pragmatics,* standing for the maps that extract hidden meanings from signs, i.e., perform inference on the sign to extract the meaning.

In the mobile robotics context the objects in a concrete category resulting from SIGNS must be able to represent in a unified way (i) regular information exchanges, e.g., state data exchanged over regular media, and (ii) robot motion information. In addition, they should fit the capabilities of unskilled humans when interacting with the robots, much like a natural language.

A forgetful functor assigns to each object in SIGNS a set that is relevant for the above objectives. The co-domain category is denoted ACTIONS and is defined in Diagram (3),

$$\begin{tikzcd}
(q_0, a, B_a) \arrow[r, "semantics"] \arrow[d, "synctatics"'] & (q_0, B_a) \\
A \arrow[ur, "pragmatics"'] &
\end{tikzcd} \tag{3}$$

where A represents the practical implementation of a semiotic sign, q_0 stands for an initial condition that marks the creation of the semiotic sign, e.g., the configuration of a robot, a stands for a process or algorithm that implements a functionality associated with the semiotic sign, e.g., a procedure to compute an uncertainty measure at q_0, B_a stands for a set in the domain space of a, e.g., a compact region in the workspace. A practical way to read Diagram (3) is to consider the objects in A as having an internal structure of the form (q_0,a,B_a) of which (q_0, B_a) is of particular interest to represent a meaning for some classes of problems.

The *syntactics* morphism is the constructor of the object. It implements the syntactic rules that create an A object. The constructors in object oriented programming languages are typical examples of such morphisms.

The *semantics* morphism is just a projection operator. In this case the semantics of interest is chosen as the projection onto $\{ q_0 \times B_a\}$.

The *pragmatics* morphism implements the maps used to reason over signs. For instance, the meaning of a sign can in general be obtained directly from the behavior of the sign as described by the object A (instead of having it extracted from the label)[3].

Diagram 3 already imposes some structure on the component objects of a semiotic sign. This structure is tailored to deal with robot motion and alternative structures are of course possible. For instance, the objects in ACTIONS can be extended to include additional components expressing events of interest.

---

[3] This form of inference is named *hidden meaning* in semiotics.

5. From abstract to concrete objects

Humans interact among each others using a mixture of loosely defined concepts and precise concepts. In a mixed human-robot system, data exchanges usually refer to robot configurations, uncertainty or confidence levels, events and specific functions to be associated with the numeric and symbolic data. This means that the objects in ACTIONS must be flexible enough to cope with all these possibilities. In the robot motion context, looseness can be identified with an amount of uncertainty when specifying a path, i.e., instead of specifying a precise path only a region where the path is to be contained is specified, together with a motion trend or intention of motion. To an unskilled observer, this region bounding the path conveys a notion of equivalence between all the paths contained therein and hence it embedds some semantic content.

The objects in ACTIONS span the above caracteristics. For mobile robots the following are two examples.

- The action map *a* represents a trajectory generation algorithm that is applied at some initial configuration q_0 and constrained to stay inside a region B_a, or
- The action map *a* stands for an event detection strategy, from state and/or sensor data, when the robot is at configuration qo with an uncertainty given by B_a.

Definition 1 illustrates an action that is able to represent motion trends and intentions of motion.

Definition 1 (ACTIONS) *Let k be a time index,* q_0 *the configuration of a robot where the action starts to be applied and* $a(q_0)\,|_k$ *the configuration at time k of a path generated by action a.*

A free action is defined by a triple (q_0,a,B_a) *where* B_a *is a compact set and the* q_0 *the initial condition of the action, and verifies,*

$$q_0 \in B_a, \tag{4}$$

$$a(q_0)|_0 = q_0, \tag{4b}$$

$$\exists_{\epsilon > \epsilon_{\min}} \; : \; \mathcal{B}(q_0, \epsilon) \subseteq B_a, \tag{4c}$$

with $\mathcal{B}(q_0,\epsilon)$ *a ball of radius* ϵ *centered at* q_0*, and*

$$\forall_{k \geq 0} \;\; a(q_0)|_k \in B_a \tag{4d}$$

Definition 1 establishes a loose form of equivalence between paths generated by *a*, starting in a neighborhood of q_0, and evolving in the bounding region B_a. This equivalence can be fully characterized through the definition of an equality operator in ACTIONS. The resulting notion is more general than the classical notion of simulation (and bisimulation)[4] as the relation between trajectories is weaker. Objects as in Definition 1 are rather general. These objects can be associated with spaces other than configuration spaces and workspaces. Also, it is possible to define an algebraic framework in ACTIONS with a set of free operators that express the motion in the space of actions, (Sequeira and M.I. Ribeiro, 2006b). The

---

[4] Recall that a simulation is a relation between spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ such that trajectories in both spaces are similar independent of the disturbances in $\mathcal{X}_1$. A bisimulation extends the similarity also to disturbances in X2 (see (van der Schaft, 2004)).

interest of having such a framework is that it allows to determine conditions under which good properties such as controllability are preserved.

A free action verifying Definition 1 can be implemented as in the following proposition.

Proposition 1 (Action) *Let* $a(q_0)$ *be a free action. The paths generated by* $a(q_0)$ *are solutions of a system in the following form,*

$$\dot{q} \in F_a(q) \tag{5}$$

where F_a *is a Lipschitzian set-valued map with closed convex values verifying,*

$$F_a(q) \subseteq T_{B_a}(q) \tag{6}$$

where $T_{B_a}(q)$ *is the contingent cone to* B_a *at* q*.*

The demonstration of this proposition is just a restatement of Theorem 5.6 in (Smirnov, 2002) on the existence of invariant sets for the inclusion (5).

When $\{q\}$ is a configuration space of a robot, points in the interior of B_a have T_{B_a} equal to the whole space. When q is over the boundary of B_a the contingent cone is the tangent space to B_a at q. Therefore, when q is in the interior of B_a it is necessary to constrain $T_{B_a}(q)$ to obtain motion directions that can drive a robot (i) through a path in the interior or over the boundary of B_a, and (ii) towards a mission goal.

In general, bounding regions of interest are nonconvex. To comply with Proposition 1 triangulation procedures might be used to obtain a covering formed by convex elements (the simplicial complexes). Assuming that any bounding region can be described by the union of such simplicial complexes[5], the generation of paths by an action requires that (i) an additional adjustment of admissible motion directions such that the boundary of a simplicial complex can be crossed, and (ii) the detection of adequate events involved, e.g., approaching the boundary of a complex and crossing of the border between adjacent complexes. When in the interior of a simplicial complex, the path is generated by some law that verifies Proposition 1.

The transition between complexes is thus, in general, a nonsmooth process. Figure 1 shows 2D examples that strictly follow the conditions in Proposition 1 with very different behaviors.

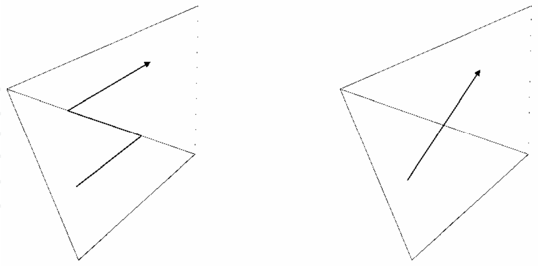


Figure 1. Examples for 2D simplicial complex crossing

---

[5] See for instance (Shewchuk, J. R., 1998) for conditions on the existence of constrained Delaunay triangulations.

In robotics it is important to ensure that transitions between complexes occur as smoothly as possible in addition to having the paths staying inside the overall bounding region. Proposition 2 states sufficient conditions for transitions avoiding moving over the boundary in adjacent complexes.

Proposition 2 (Crossing adjacent convex elements) *Let a path q be generated by motion directions verifying Proposition 1, and consider two adjacent (i.e., sharing part of their boundaries) simplicial complexes* B_1 *and* B_2 *assume that the desired crossing sequence is* B_1 *to* B_2*. Furthermore, let* $\mathcal{B}(q, \epsilon)$ *be a neighbourhood of q with radius e and define the points* q_{ci} *and sets* C_i*, i* = 1,2 *as,*

$$\begin{aligned} q_{c_i} &: \mathcal{B}(q_{c_i}, \epsilon_i) \subset B_i, \quad \textit{for some } \epsilon_i \\ C_i &= q_{c_i} + \lambda_i \left(\mathcal{B}_j - q_{c_i}\right), \quad \lambda_i \in [0, 1] \end{aligned}$$

and let the set of admissible motion directions be defined by

$$T_{B_1 \cup B_2} = \begin{cases} T_{B_1}(q) & \textit{if} \quad q \in B_1 \backslash C_1 \\ (\mathcal{B}_2 - q_{c_1}) & \textit{if} \quad q \in B_1 \cap C_1 \\ -(\mathcal{B}_1 - q_{c_2})) & \textit{if} \quad q \in B_2 \cap C_2 \\ T_{B_2}(q) & \textit{if} \quad q \in B_2 \backslash C_2 \end{cases} \tag{7}$$

Then the path q crosses the boundary of B_1 *and enters* B_2 *with minimal motion on the border* $B_1 \cap B_2$*.*

The demonstration follows by showing that when *q* is over the border between B_1 and B_2 the motion directions given by $(B_1 \cap B_2) - q$ are not admissible.

Expression (7) determines three transitions. The first transition occurs when the robot moves from a point in the interior of B_1, but outside C_1, to a point inside $B_1 \cap C_1$. The admissible motion directions are then those that drive the robot along paths inside B_1 as if no transition would have to occur. At this event the admissible motions directions drive the robot towards the border between B_1 and B_2 because C_1 is a viability domain as $\mathcal{B}_2 - q_{c_1} \subset T_{C_1}$. The second transition occurs when the robot crosses the border between B_1 and B_2. At this point the admissible motion directions $-(\mathcal{B}_1 - q_{c_2}) \subset T_{C_2}$ and hence the path moves away from the border $B_1 \cap B_2$ towards the interior of B_2. At the third transition the path enters $B_2 \backslash C_2$ and the admissible motion directions yield paths inside B_2.

While $q \in B_1 \cap B_2$ there are no admissible motion directions either in $T_{B1}(q)$ *or* $T_{B2}(q)$ and hence the overlapping between the trajectory and the border is minimal.

Examples of actions can be easily created. Following the procedure outline above, a generic bounding region in a 2D euclidean space, with boundary defined by a polygonal line, it can be (i) covered with convex elements obtained through the Delaunay triangulation on the vertices of the polygonal line (the simplicial complexes), and (ii) stripped out of the elements which have at least one point outside the region. The resulting covering defines a topological map for which a visibility graph can be easily computed.

Figure 2 shows an example of an action with a polygonal bounding region, defined in a 2D euclidean space with the bounding region covered with convex elements obtained with Delaunay triangulation. The convex elements in green form the covering. The o marks inside each element stand for the corresponding center of mass, used to define the nodes of the visibility graph. The edges of elements that are not completely contained inside the polygonal region are shown in blue. The red lines represent edges of the visibility graph of which the shortest path between the start and end positions are shown in blue.

Proposition 2 requires the computation of additional points, the q_{ci}. In this simple example it is enough to choose them as the centers of mass of the triangle elements. The neighbourhoods $B(q_{ci}, \varepsilon)$ can simply be chosen as the circles of maximal radius that can be inscribed in each triangle element. The second transition in (7) is not used in this example. The admissible motion directions are simply given by

$$\dot{q}(t) \in \mathcal{B} - q \tag{8}$$

where B stands for the neighborhood in the convex element where to cross, as described above.

The line in magenta represents the trajectory of a unicycle robot, starting with 0 rad orientation. The linear velocity is set to a constant value whereas the angular velocity is defined such as to project the velocity vector onto (8).

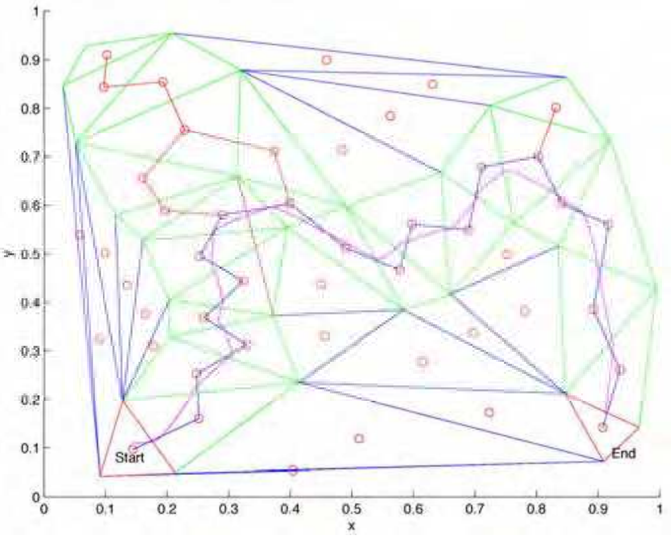


Figure 2. Robot moving inside a bounding region

6. Experiments

Conveying meanings through motion in the context of the framework described in this chapter requires sensing and actuation capabilities defined in that framework, i.e., that robots and humans have (i) adequate motion control, and (ii) the ability to extract a meaning from the motion being observed.

Motion control has been demonstrated in real experiments extensively described in the robotics literature. Most of that work is related to accurate path following. As aforementioned, in a wide range of situations this motion does not need to be completely specified, i.e., it is not necessary to specify an exact trajectory. Instead, defining a region where the robot is allowed to move and a goal region to reach might be enough. The tasks that demonstrate interactions involving robots controlled under the framework described in this chapter are not different for classical robotics tasks, e.g., reaching a location in the workspace.

The extraction of meanings is primarily related to sensing. The experiments in this area assess if specific strategies yield bounding regions that can be easily perceived by humans and can also be used by robots for motion control.

Two kinds of experiments are addressed in this chapter, (i) using simulated robots, and (ii) using real robots. The former allow the assessment of the ideas previously defined under controlled conditions, namely the assessment of the performance of the robots independently of the noise/uncertainties introduced by the sensing and actuation devices. The later illustrate the real performance.

6.1 Sensing bounding regions

Following Diagram 3, a meaning conveyed by motion lies in some bounding region. The extraction of meanings from motion by robots or humans thus amounts to obtain a region bounding their trajectories. In general, this is an ill-posed problem. A possible solution is given by

$$\begin{aligned} q_0 &= \hat{q}(t-h) \\ B_a(t) &= \cup_t \mathcal{B}(\hat{q}(t), \epsilon) \end{aligned} \quad (9)$$

where $\hat{q}(t)$ is the estimated robot configuration at time t, $\mathcal{B}(\hat{q}(t), \epsilon)$, is a ball of radius ϵ and centered at $\hat{q}(t)$, and h is a time window that marks the initial configuration of the action. This solution bears some inspiration in typically human charateristics. For instance, when looking at people moving there is a short term memory of the space spanned in the image plane. Reasoning on this spanned space might help extrapolating some motion features.

In practical terms, different techniques to compute bounding regions can be used depending on the type of data available. When data is composed of sparse information, e.g., a set of points, clustering techniques can be applied. This might involve (i) computing a dissimilarity matrix for these points, (ii) computing a set of clusters of similar points, (iii) map each of the clusters into adequate objects, e.g., the convex hull, (iv) define the relation between these objects, and (iv) remove any objects that might interfere with the workspace.

Imaging sensors are commonly used to acquire information on the environment. Under fair lighting conditions, computing bounding regions from image data can be done using image subtraction and contour extraction techniques[6]. Figure 3 illustrates examples of bounding regions extracted from the motion of a robot, sampled from visual data at irregular rate. A basic procedure consisting in image subtraction, transformation to grayscale and edge detection is used to obtain a cluster of points that are next transformed in a single object using the convex hull. These objects are successively joined, following (9), with a small time window. The effect of this time window can be seen between frames 3 and 4, where the first object detected was removed from the bounding region.

The height of the moving agent clearly influences the region captured. However, if a calibrated camera is used it is possible to estimate this height. High level criteria and a priori knowledge on the environment can be used to crop it to a suitable bounding region. Lower abstraction levels in control architectures might supsump high level motion commands

---

[6] Multiple techniques to extract contours in an image are widely available (see for instance (Qiu, L. and Li, L., 1998; Fan, X. and Qi, C. and Liang, D. and Huang, H., 2005)).

computed after such bounding regions that might not be entirely adequate. A typical example would be having a low level obstacle avoidance strategy that overcomes a motion command computed after a bounding region obtained without accounting for obstacles.

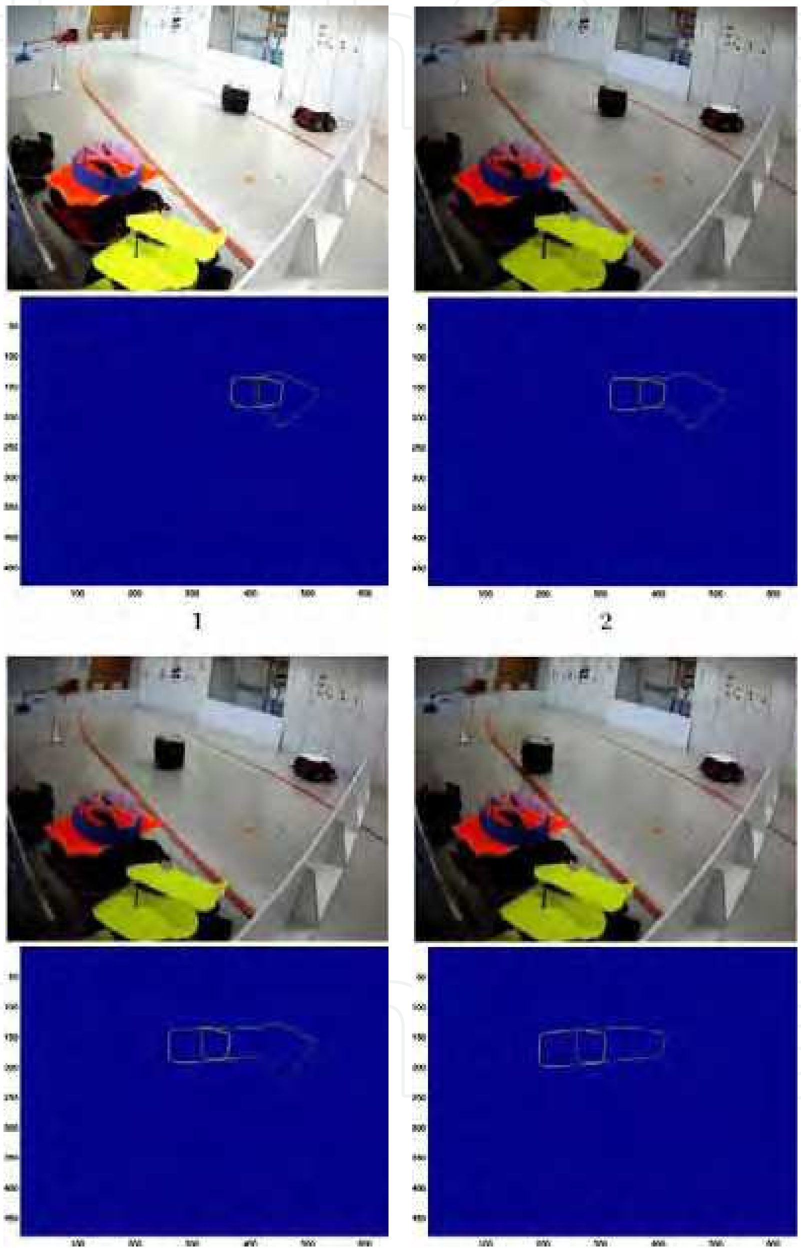


Figure 3. Bounding region extracted from the motion of a robot

6.2 Human interacting with a robot

Motion interactions between humans and robots often occur in feedback loops. These interactions are mainly due to commands that adjust the motion of the robot. While moving the robot spans a region in the free space that is left for the human to extract the respective meaning. A sort of error action is computed by the human which is used to define new goal actions to adjust the motion of the robot.

Map based graphical interfaces are common choices for an unskilled human to control a robot. Motion commands specifying that the robot is to reach a specific region in the workspace can be defined using the framework previously defined, forming a crude language. The interactions through these interfaces occur at sparse instants of time, meaning that the human is not constantly adjusting bounding regions. Therefore, for the purpose of illustrating the interaction under the framework described it suffices to demonstrate the motion when a human specifies a single bounding region.

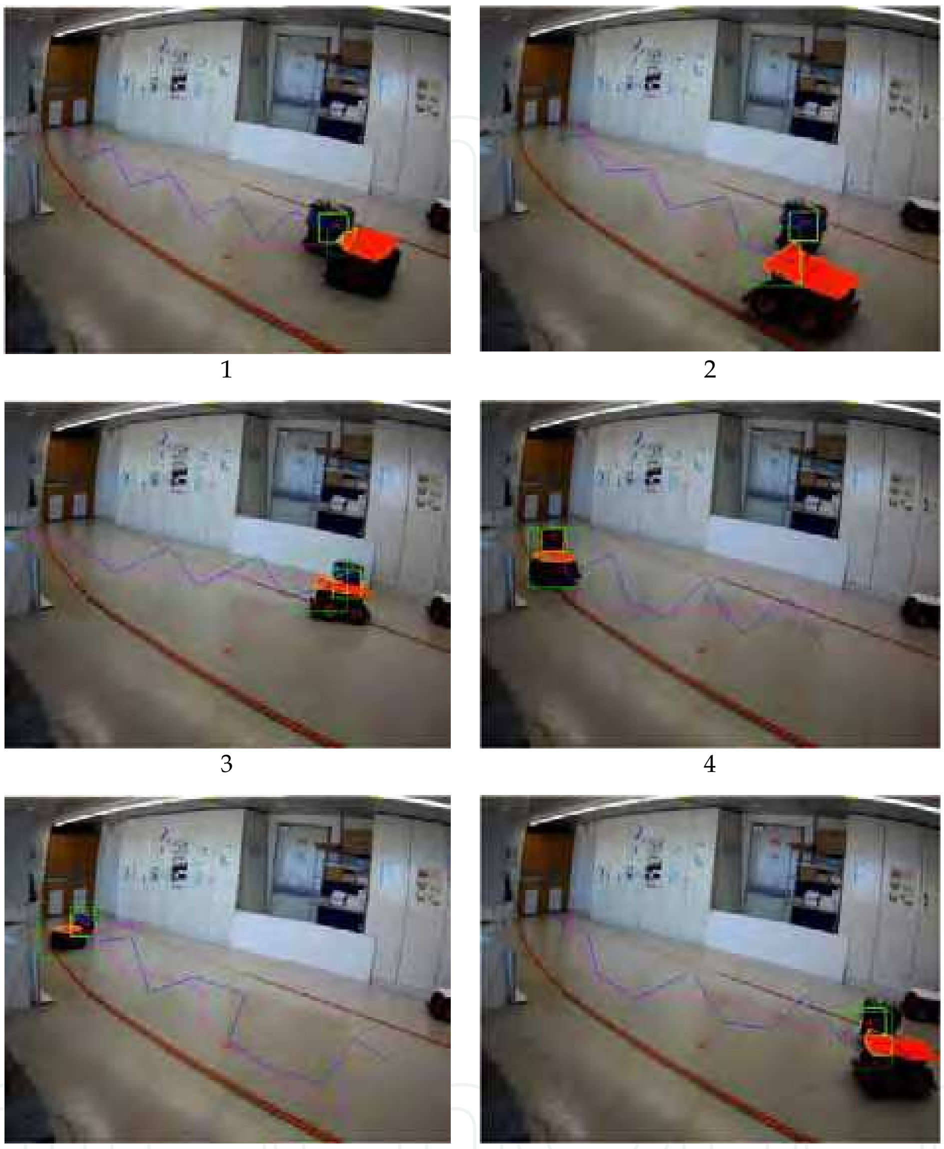
Figure 4 illustrates the motion of a unicycle robot in six typical indoor missions. For the purpose of this experiment, the robot extracts its own position and orientation from the image obtained by a fixed, uncalibrated, camera mounted on the test scenario[7]. Position is computed after a rough procedure based on color segmentation. Orientation is obtained through the timed position difference. A first order low pass filtering is used to smooth the resulting information. It is worth to point that sophisticated techniques for estimating the configuration of a robot from this kind of data, namely those using a priori knowledge on the robot motion model, are widely available. Naturally, the accuracy of such estimates is higher than the one provided by the method outline above. However, observing human interactions in real life suggests that only sub-optimal estimation strategies are used and hence for the sake of comparison it is of interest to use also a non-optimal strategy. Furthermore, this technique limits the complexity of the experiment.

A Pioneer robot (shown in a bright red cover) is commanded to go to the location of a Scout target robot (held static), using a bounding region defined directly over the same image that is used to estimate the position and orientation. Snapshots 1, 2, 3 and 6 show the Pioneer robot starting in the lefthand side of the image whereas the target robot is placed on the righthand side. In snapshots 4 and 5 the region of the starting and goal locations are reversed.

The blue line shows the edges of the visibility graph that corresponds to the bounding region defined (the actual bounding region was omitted to avoid cumbersome graphics). The line in magenta represents the trajectory executed. All the computations are done in image plane coordinates. Snapshot 5 shows the effect of a low level obstacle avoidance strategy running onboard the Pioneer robot. Near the target the ultrasound sensors perceive the target as an obstacle and force the robot to take an evasive action. Once the obstacle is no longer perceived the robot moves again towards the target, this time reaching a close neighborhood without the obstacle avoidance having to interfere.

---

[7] Localisation strategies have been tackled by multiple researchers (see for instance, (Betke and Gurvits, 1997; Fox, D. and Thrun, S. and Burgard, W. and Dellaert, F., 2001)). Current state of the art techniques involve the use of accurate sensors e.g., inertial sensors, data fusion and map building techniques.



1 2 3 4 5 6

Figure 4. Robot intercepting a static target

6.3 Interacting robots

Within the framework described in this chapter, having two robots interacting using the actions framework is basically the same as having a human and a robot as in the previous section. The main difference is that the bounding regions are processed automatically by each of the robots.

In this experiment a Scout robot is used as a static target while two Pionner 3AT robots interact with each other aiming at reaching the target robot. The communication between the two Pioneer robots is based on images from a single camera. Both robots have access to the same image, from which they must infer the actions the teammates are executing.

Each of the Pioneer robots uses a bounding region for its own mission defined after criteria similar to those used by typical humans, i.e., chooses its bounding regions to complement the one chosen by the teammate.

A bounding region spanned by the target is arbitrarily identified by each of the chasers (and in general they do not coincide). Denoting by B_1 and B_2 the regions being used by the chasing robots in the absence of target and by B_{g_1} and B_{g_2} regions where the target robot was identified by each of them the bounding region of each of the chasers is simple $B_i' = B_i \cup B_{g_i}$. If shortest routes between each of the chasers and the target are required then it suffices to make the $B_i = q_i + \lambda(B_{g_i} - q_i),\ \lambda \in [0,1]$.

The two bounding regions, B_1' and B_2', overlap around the target region.

The inclusion of the B_{g_i} aims at creating enough space around the target such that the chaser robots can approach the target without activating their obstacle avoidance strategies. Figure 5 shows two simulations of this problem with unicycle robots. The target location is marked with a yellow $*$. In the lefthand image the target is static whereas in the righthand side one uniform random noise was added both to the target position and to the B_{g_i} areas. In both experiments the goal was to reach the target within a 0.1 distance.

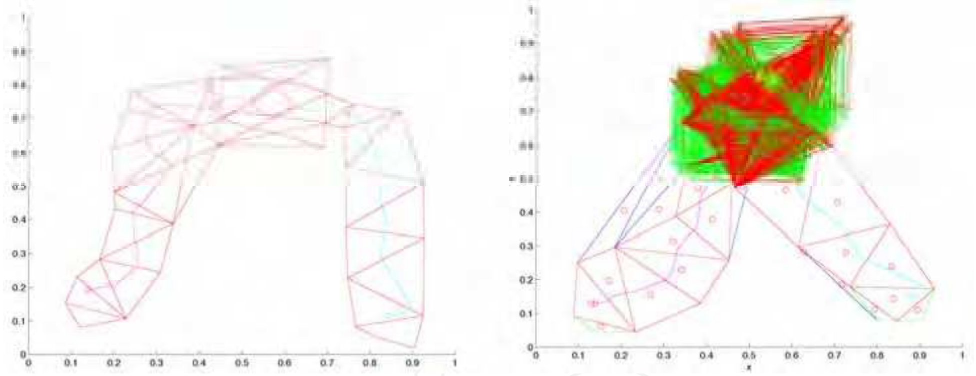


Figure 5. Intercepting an intruder

Figure 6 shows a sequence of snapshots obtained in three experiments with real robots and a static target. These snapshots were taken directly from the image data stream being used by the robots. The trajectories and bounding regions are shown superimposed.

It should be noted that the aspects related to robot dynamics might have a major influence in the results of these experiments. The framework presented can be easily adjusted to account for robot dynamics. However, a major motivation to develop this sort of framework is to be able to have robots with different functionalities, and often uncertain, dynamics interacting. Therefore, following the strategy outlined in Diagram 1, situations in which a robot violates the boundary of a bounding region due, for example, to dynamic constraints can be assumed to be taken care by lower levels of abstraction.

Figure 6. Intercepting an intruder

7. Conclusions

The social barriers that still constrain the use of robots in modern societies will tend to vanish with the sophistication increase of interaction strategies. Communication and interaction between people and robots occurring in a friendly manner and being accessible to everyone, independent of their skills in robotics issues, will certainly foster the breaking of barriers.

Socializing behaviors, such as following people, are relatively easy to obtain with current state of the art robotics. However, achieving human like interaction through motion (as people do) requires the development of alternative models to synthesize behaviors for robots. The framework outlined in this chapter shows how models of human interactions from social sciences, can be merged with robot control techniques to yield a set of objects that simplifies the development of robotics applications.

The experiments presented demonstrate interactions involving humans and robots similar to those arising in classical approaches. Even though these similarities, for example measured through the visual quality of trajectories generated by the robots, the effort to develop these experiments was only a fraction of the effort that a classical approach would have cost. Furthermore, the results show that robots can operate and interact both among themselves and with people, with significant quality, in poorly modeled environments. The experiments were designed for minimal technological requirements, hence avoiding shadowing the performance of the framework described.

The discussion on how to make a concrete object out of an abstract concept, such as meaning, might lead to other alternative frameworks. The one described in this chapter priviledges the locomotion features that characterizes a robot, namely by using as support space the configuration space. Still, multiple extensions can be made out of the ideas developed. A virtual agent might require additional components in the objects in SIGNS or even alternative support spaces, for instance to simplify reasoning processes.

As a final comment, though this work aims at approaching robots to people, as referred in (Scholtz, 2003), robot designers should also strieve to enhance human skills through robot technology in addition to trying to substituting robot skills by human ones.

8. References

Asarin, A. and Dang, T. (2004). Abstraction by projection and application to multi-affine systems. In *Procs. of 7th Int. Workshop on Hybrid Systems: Control and Computation, HSCC 04,* volume 2993 of *Lecture Notes in Computer Science,* pages 32-47. Springer. Philadelphia, USA, March 25-27.

Asarin, E. and Maler, O. (1994). On some relations between dynamical systems and transition systems. In Abiteboul, S. and Shamir, E., editors, *Procs. of 21st Int. Colloquium on Automata, Languages and Programming, 1C ALP 94,* volume 820 of *Lecture Notes in Computer Science,* pages 59-72. Springer. Jerusalem, Israel, July 11-14.

Betke, M. and Gurvits, L. (1997). Mobile Robot Localization Using Landmarks. *IEEE Transactions on Robotics and Automation,* 13(2).

Bignell, J. (1997). *Media Semiotics: An Introduction.* Manchester University Press, 2nd edition.

Carroll, J., editor (2003). HCI Models, Theories, and Frameworks - Towards a Multidisciplinary Science. Morgan Kaufmann Publishers.

Chandler, D. (2003). *Semiotics, The basics.* Rutledge.

Chomsky, N. (1968). *Language and Mind.* Harcourt Brace Jovanovich Inc.

Codognet, P. (2002). The semiotics of the web. In *Leonardo,* volume 35(1). The MIT Press.

Dautenhahn, K. and Werry, I. (2002). A Quantitative Technique for Analysing Robot-Human Interactions. In *Procs. 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems.* EPFL, Lausanne, Switzerland, October.

Bias, B., Zlot, R., Kalra, N., and A., S. (2005). Market-based multirobot coordination: A survey and analysis.

Everett, H. (2003). Robotic security systems. *IEEE Instrumentation & Measurement Magazine,* 6(4):30-34.

Fan, X. and Qi, C. and Liang, D. and Huang, H. (2005). Probabilistic Contour Extraction Using Hierarchical Shape Representation. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV05).*

Fox, D. and Thrun, S. and Burgard, W. and Dellaert, F. (2001). Particle Filters for Mobile Robot Localization. In A. Doucet, N. de Freitas, and N. Gordon, editor, *Sequential Monte Carlo Methods in Practice,* pages 499-516. Springer Verlag.

Harman (1969). Linguistic competence and empiricism. In *Language and Philosophy.* New York University Press.

Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviours. *IEEE Transactions on Systems, Man and Cybernetics, Part C,* 34(3):334-352.

IEEE (2000). IEEE Recommended Practice for Architecture Description of Software-Intensive Systems. E-ISBN 0-7381-2519-9, ISBN 0-7381-2518-0.

Kawamura, K., Nilas, P., Muguruma, K., Adams, J., and Zhou, C. (2003). An agent-based architecture for an adaptive human-robot interface. In *Procs. 36th Hawaii Int. Conf. on System Sciences.*

Khrisna, K. and H., H. (2003). Social control of a group of collaborating multi-robot multi-target tracking agents. In *Procs. 22nd Digital Avionics Systems Conference (DASC-22).*

Kogut, G., Blackburn, M., and Everett, H. (2003). Using video sensor networks to command and control unmanned ground vehicles. In *Procs Unmanned Systems in International Security 2003 (USIS 03).*

Konolidge, K., Fox, D., Ortiz, C., Agno, A., Eriksen, M., Limketkai, B., Ko, J., Morisset, B., Schulz, D., Stewart, B., and R., V. (2004). Centibots: Very large scale distributed robotic teams. In *Procs. Int. Symp. on Experimental Robotics (ISER-04).*

Malcolm, G. and Goguen, J. (1998). Signs and representations: Semiotics for user interface design. In *Procs. Workshop in Computing,* pages 163-172. Springer. Liverpool, UK.

Meystel, A. and Albus, J. (2002). *Intelligent Systems: Architecture, Design, and Control.* Wiley Series on Intelligent Systems. J. Wiley and Sons.

Neumiiller, M. (2000). Applying computer semiotics to hypertext theory and the world wide web. In Reich, S. and Anderson, K., editors, *Procs. of the 6th Int. Workshop on Open Hypertext Systems and Structural Computing,* Lecture Notes in Computer Science, pages 57-65. Springer-Verlag.

Nguyen, H. and Bott, J. (2000). Robotics for law enforcement: Applications beyond explosive ordnance disposal. Number 4232, pages 433-454.

Nicolescu, M. and Mataric, M. (2001). Learning and interacting in human-robot domains. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans,* 31(5):419-430.

Ogasawara, G. (1991). A distributed, decision-theoretic control system for a mobile robot. *ACM SIGART Bulletin,* 2(4):140-145.

Parker, L. (1998). ALLIANCE: An Architecture for Fault Tolerant Multirobot Cooperation. *IEEE Transactions on Robotics and Automation,* 14(2):220-240.

Qiu, L. and Li, L. (1998). Contour Extraction of Moving Objects. In *Procs. of 14th International Conference on Pattern Recognition (ICPR98).* Brisbane, Australia, August.

Rani, P. and Sarkar, N. (2004). Emotion-Sensitive Robots - A New Paradigm for Human-Robot Interaction. In *IEEE-RAS/RSJ International Conference on Hu-manoid Robots (Humanoids 2004).* Santa Monica, Los Angeles, CA, USA, November 10-12.

Raskin, J. (2000). The Human Interface - New Directions for Designing Interactive Systems. Addison-Wesley.

Rybski, P., Hougen, D., Stoeter, S., Gini, M., and Papanikolopoulos, N. (2000). Control of multiple small surveillance robots at AAAI 2000. In *Procs. AAAI 2000 Mobile Robot Competition and Exhibition Workshop.*

Ryu, H. and Lee, W. (2006). Where You Point is Where the Robot is. In *Procs. of 7th ACM SIGCHI New Zealand Chapter's Int. Conf. on Computer-Human Interaction: Design Centered HCI,* volume 158, pages 33-42. ACM International Conference Proceeding Series. Christchurch, New Zealand.

Savidis, A. and Stephanidis, C. (2004). Unified user interface design: designing universally accessible interactions. *Interacting with Computers,* 16(2):243—270.

Scholtz, J. (2003). Theory and evaluation of human robot interactions. In *Procs. of the 36th Hawaii Int. Conf. on System Sciences, HICSS'03.* Hawaii, January 5-8, USA,.

Sequeira, J. and M.I. Ribeiro, M. (2006a). Human-robot interaction and robot control. *Springer's Lecture Notes in Control and Information Sciences,* 335:375-390.

Sequeira, J. and M.I. Ribeiro, M. (2006b). A semiotic apporach to the control of semi-autonomous robots. *International Journal of Systems Science,* 37(6):361-376.

Shewchuk, J. R. (1998). A condition guaranteeing the existence of higher-dimensional constrained Delaunay triangulations. In *Procs. of 14th Annual Symposium on Computational Geometry,* pages 76—85. Minneapolis, Minnesota, USA.

Smirnov, G. (2002). Introduction to the Theory of Differential Inclusions, volume 41 of Graduate Studies in Mathematics. American Mathematical Society.

Stankovic, S. and Siljak, D. (2002). Model abstraction and inclusion principle: A comparison. *IEEE Transactions on Automatic Control,* 47(3):529-532.

Valera, M. and Velastin, S. (2005). Intelligent distributed surveillance systems: a review. *IEE Procs. On Vision, Image and Signal Processing,* 152(2):192-204.

van der Schaft, A. (2004). Equivalence of dynamical systems by bisimulation. *IEEE Transactions on Automatic Control,* 49(12).

Willems, J. (1991). Paradigms and puzzles in the theory of dynamical systems. *IEEE Transactions on Automatic Control,* 36(3).

Human Robot Interaction
Edited by Nilanjan Sarkar

ISBN 978-3-902613-13-4
Hard cover, 522 pages
Publisher I-Tech Education and Publishing

Published in print edition September, 2007

Human-robot interaction research is diverse and covers a wide range of topics. All aspects of human factors and robotics are within the purview of HRI research so far as they provide insight into how to improve our understanding in developing effective tools, protocols, and systems to enhance HRI. For example, a significant research effort is being devoted to designing human-robot interface that makes it easier for the people to interact with robots. HRI is an extremely active research field where new and important work is being published at a fast pace. It is neither possible nor is it our intention to cover every important work in this important research field in one volume. However, we believe that HRI as a research field has matured enough to merit a compilation of the outstanding work in the field in the form of a book. This book, which presents outstanding work from the leading HRI researchers covering a wide spectrum of topics, is an effort to capture and present some of the important contributions in HRI in one volume. We hope that this book will benefit both experts and novice and provide a thorough understanding of the exciting field of HRI.

Joao Sequeira and M.Isabel Ribeiro (2007). Semiotics and Human-Robot Interaction, Human Robot Interaction, Nilanjan Sarkar (Ed.), ISBN: 978-3-902613-13-4, InTech, Available from: http://www.intechopen.com/books/human_robot_interaction/semiotics_and_human-robot_interaction

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821