Int J of Soc Robotics (2017) 9:231–249
DOI 10.1007/s12369-016-0389-0

Teaching Robot's Proactive Behavior Using Human Assistance

A. Garrell[1] · M. Villamizar[1] · F. Moreno-Noguer[1] · A. Sanfeliu[1]

Accepted: 15 December 2016 / Published online: 5 January 2017

Abstract In recent years, there has been a growing interest in enabling autonomous social robots to interact with people. However, many questions remain unresolved regarding the social capabilities robots should have in order to perform this interaction in an ever more natural manner. In this paper, we tackle this problem through a comprehensive study of various topics involved in the interaction between a mobile robot and untrained human volunteers for a variety of tasks. In particular, this work presents a framework that enables the robot to proactively approach people and establish friendly interaction. To this end, we provided the robot with several perception and action skills, such as that of detecting people, planning an approach and communicating the intention to initiate a conversation while expressing an emotional status. We also introduce an interactive learning system that uses the person's volunteered assistance to incrementally improve the robot's perception skills. As a proof of concept, we focus on the particular task of online face learning and recognition. We conducted real-life experiments with our Tibi robot to validate the framework during the interaction process. Within this study, several surveys and user studies have been realized to reveal the social acceptability of the robot within the context of different tasks.

Keywords Human robot interaction · Social robotics · Human-assisted learning

This work was supported by the Spanish Ministry of Science and Innovation, project Rob-Int-Coop DPI2013-42458-P and AEROARMS European project H2020-ICT-2014-1-644271.

✉ A. Garrell
agarrell@iri.upc.edu

M. Villamizar
mvillami@iri.upc.edu

F. Moreno-Noguer
fmoreno@iri.upc.edu

A. Sanfeliu
sanfeliu@iri.upc.edu

[1] Institut de Robótica i informática industrial, CSIC-UPC, 08028 Barcelona, Spain

1 Introduction

Human-Robot Interaction (HRI) is an area of research that has received much attention in the recent years. There exists a wide range of applications in which HRI plays an important role, from the use of robots as companions for the eldery [23], to their ability to safely evacuate people in emergency situations [3].

One major topic within HRI research is that of giving robots the ability to initiate interaction with humans. It is commonly thought that social robots should engage in the same way as people do, using human-like physical signals and gestures [43]. In this spirit, recent studies have shown that while robots are able to encourage people to initiate interaction themselves [11,26], they consistently expect people to approach them instead of being the ones to initiate contact [37].

In this work, we go a step further and endow the mobile robot with proactive capabilities to seek out human interaction and to establish engagement with people, while revealing an expressive status through an emotional model, such that the person feels close to the robot and capable of forming a bond. Concretely, the presented approach is motivated by the appraisal models of humans emotions [40,46]. As stated in these models, a robot continuously appraises the situation is involved in, then, emotions can be triggered (e.g., the person is interacting with it or not). There exist a set of strategies that can be used to deal with a specific emotion, for example, by updating the agent's mental state (e.g. ., feeling happy if the volunteer is collaborating with the robot). Once this engagement has been established, we provide to the robot with cognitive and interactive capacities such that they may perform collaborative tasks wherein the human teaches the robot new skills in perception.

The contributions of this paper are as follows: first, we introduce a framework in which a mobile robot is able to initiate interaction with a person and develop an engagement proactively, focusing on the way the robot initiates the conversation in a manner perceived as natural by the person. Specifically, we examine the human communication model proposed by Clark [9], based on the notion that people in a conversation perceive the roles of other persons, such as a speaker, listener, and side participants. In order to develop this initial task, we gave our robot a visual module for detecting human faces in real time, with the caveat that faces must be non-occluded. To demonstrate the proper development of our model, we performed a user study wherein we discussed how the perceived acceptability of the robot is enhanced, as compared with two other simpler behaviors (base-line).

Our next contribution was to introduce a second robot-human communication framework, once the engagement had been initiated, wherein the human can naturally help the robot improve the performance of its facial recognition module. We used an online learning algorithm [56] that incorporates the human's assistance to enhance its performance. Following this interaction, the robot becomes able to detect faces in adverse conditions, such as when detection of visual targets is hindered by abrupt changes in light or partial occlusions. In addition, the robot learns the person's identity in order to engage in coherent dialogue with him/her in the future. In this online and real-time assisted algorithm, the human plays the role of teacher, guiding the robot through its learning process, and correcting the output of the facial recognition system in those difficult cases that require human assistance. The amount of human intervention lessens in intensity over time, and usually after a few seconds the robot's visual system becomes significantly more robust and reliable. Figure 1 shows different frames from a typical teaching process between a person and our mobile robot, Tibi.

The robot's demonstrated ability to approach people and learn to use human assistance leads to a number of possible applications. Among the most promising of these is the robot's capacity to independently look for people who can assist it, so as to progressively improve upon its skills throughout the interaction process. For instance, in urban spaces, if the robot loses its position, or it is looking for a special location in an unknown place, it can effectively ask for help from pedestrians. Moreover, in an scenario wherein an elderly person or a child is lost, the robot, rather than waiting for the lost individual to initiate contact, can move towards him/her proactively.

Finally, real-life experiments were conducted over the course of three weeks with our mobile service robot Tibi within different urban environments in Barcelona city, containing dynamic obstacles introduced to validate the framework during the interaction process. Furthermore, in this paper, questionnaires and user studies were carried out to explore the tolerance for the robot's different tasks. The results of these surveys are summarized and their most significant factors are discussed in detail.

The remainder of our study unfolds in the following manner. Section 2 introduces the related work in human-robot natural engagement, emotional models for social robots and human-assistance for recognition. Section 3 provides an overview of the contributions we describe in this article. Section 4 describes the robot navigation method employed to approach people in a friendly manner. In Sect. 5, we specify the details of the robot's proactive behavior capable of creating engagement. The emotional model used to achieve engagement with the person is presented in Sect. 6. The active learning for online face recognition is mentioned in Sect. 7. In Sect. 8, we present the setting of the experiments and our evaluation methodology, which is subsequently employed in the results. Finally, discussion and conclusions are given in Sects. 9 and 10, respectively.

2 Related Work

We next review related work and distinguish it according the following contributions of the present work: proactive methods for natural and friendly human-robot engagement, robots's emotions and visual recognition algorithms that progressively adapt their detections using both the observation of incoming data and the corrections provided by the human.

2.1 Initiation of Human Robot Interactions

Research in the field of social robots is still relatively new in comparison to what has been seen in traditional service robotics, e.g. robots serving food in hospitals or providing specific security-related services, applications which require minimal human-robot interaction. Therefore, prior research in this particular field is rather scarce [12,44].

Researchers are making efforts towards facilitating more natural approaches to human-robot interaction. A social robot should detect the human operator and carry out his/her commands [23]. In [11], researchers showed that in a fetch-and-carry task a seated person prefers to be approached by a robot from the left or right direction, rather than frontally or from behind. Further research showed that there are other important factors which can affect this preference, such as a person's prior experience interacting with robots [35], gen-

Fig. 1 Human-robot interaction and communication. *Left*: Tibi mobile robot approaches a person to initiate a conversation. *Right*: after the first contact, the person assists Tibi to improve its visual skills. A Wii's remote controller is used to help to validate and improve the visual face detector

der [11], or in which part of the room he/she is standing or sitting [57].

Recent efforts have focused on creating robots capable of starting conversations with humans in a friendly and natural manner [43]. In the present paper, we aim to give the robot the ability to approach a person and establish engagement with him/her. Some studies have focused on developing robots able to encourage people to initiate interaction [11,26]. The most common strategy has been to expect people to approach the robots to initiate a dialogue. In contrast to this, as shown in Fig. 1, our research introduces a mobile robot that is able to approach people in a safe and friendly manner in order to initiate contact.

In [4], the authors introduced a discussion of feature representations for analyzing human spatial behavior, proxemics, which can be applied to initiate an interaction between humans and robots.

Particularly, some studies have shown that robots may be capable of encouraging people to initiate interaction, instead of waiting for people to approach them to begin a dialogue [11]. Moreover, there has also been progress in the development of robots capable of initiating human interaction themselves. [49] proposed a model for robots to initiate interaction in a shopping mall. Another important topic that has been studied is that of computing the appropriate moment in which to begin the interaction or the human participation [51]. This should be the situation wherein both human and robot establish the mutual belief that they are sharing a conversation. However, in this work, the robot does not verbally indicate its intention to initiate a conversation, but rather expresses emotion to denote such readiness.

Michalowski [41] considered the characteristics of the spatial formation of people around a robot to select an individual and initiate interaction with him/her. Other approaches chose the person based on their motion trajectories or on their physical distance from the robot [31]. The social conventions rules and customs surrounding the beginning of a conversation have also been considered. For instance, [22] suggested that these social rules are essentially rituals that mutually confirm the start of a conversation. Here, we made use of these spatial formations in order to prevent the robot from invading the human's personal space.

In [54] researchers present an integrated motion synthesis framework designed for robots that interact with people. This model generates robot motions taking into account human's safety to socially interact with humans.

In this paper, we aim to go a step further, and propose that the robot proactively seeks the interaction with a human, with the purpose of convincing the human to contribute actively to improve its visual perception skills. The main problem in this context is that the person approached might not understand that the robot is trying to initiate a conversation with him/her and actually establish engagement.

Humans initiate conversation by eye gaze [42], but in the case of robots in real-life experiments, this task becomes difficult because robots do not easily recognize human gazes, as lighting changes is a common problem in outdoor environments. Instead, we opt using body orientation gestures, verbal, interaction and emotion expressions to signal intent to initiate conversation for. Assuming that the human has understood the robot's intentions, we have developed a communication protocol that allows the person and robot to work collaborating on the task of online face recognition.

2.2 Emotional Models for Social Robots

In this section, we present a brief overview of the current state of the development of emotional models for social robots. Expressions of affections of robots can provide benefits in many ways to human-robot interaction applications [1].

Emotions are essential for making robots able to interact with humans in order to resemble human interaction [7]. Researchers need different models to express emotions in different manners. Such modeling of emotions encourages social interaction. In [30] the authors defined the "Affective Loop" as an interactive process in which the user first expresses his/her emotions through non-verbal communication involving the body, and then, the system responds generating an affective expression, making use of colours, illuminations, or haptics, which affects the user making he/she respond and, consequently, feel more implicated with the system, in the present work, the robot.

In order to create the mentioned Affective Loop between users and robots, robots should detect if the user's feelings are positive or negative, and should reason and choose the emotional response to display at a cognitive level. The methodology by which robots are able to express their intended affective states should be effective, and the actions of the emotional robot would affect the user.

In the context of Human Robot Interaction, the affective interactions have different intentions. In the present work, we are interested, on the one hand, to increase the engagement in the social interaction context, emotions contribute to create engagements. Moreover, engagement is defined as "the process by which two (or more) participants establish, maintain and end their perceived connection" in the present context [53]. Besides, it has acquired more attention by the HRI community [45]. And, on the other hand, to increment the social presence in the long-term. The absence of emotional behaviour decreases the user's perception of social presence, concretely, during long-term interactions [38]. Social robots must not only transmit believable affective expressions, but also be able to do so in a personalized way, in order to be perceived as socially agents.

For instance, [20] described a long-term field study, which showed that robots' facial expression influenced the way and the time at which humans interacted with the robot. Moreover, using NAO robot, it has been demonstrated that emotional gestures can enhance participants' perception of a robot's expressibility [58]. Furthermore, [32] presented the emotion expressions capabilities of a robotic head. Or [36] introduced an emotional approach to proactively enhance the interaction between humans and robots.

Moreover, Yohanan et al. demonstrate in [59] that effective touch is a crucial element in human robot interaction. In their work, authors study how humans communicate their emotional state via touch communication using Haptic Creature and their expectations of its reactions. Their results can enhance the affective touch interactions.

Finally, according to [29], the emotional models can be classified into three different categories: emotion displays, virtual agents, and social robots. With regard to emotion, robots may have the ability to express emotions. Therefore, they can utilize facial expressions, gestures, or other non-verbal communicative means. Some examples of these systems are presented in [5,47]. Here, the robots are able to generate different facial expressions corresponding to the six basic emotions of anger, disgust, fear, happiness, sadness, and surprise. Most often, these robots have been used in hospitals for interacting with children. In this work, our robot, Tibi, is able to express its emotions through facial expressions (with led illumination) and gestures.

Secondly, virtual agents are simulated operators. The agents have no mechanical limitations to display their expressions, since their bodies are simulated. Nevertheless, they cannot physically interact with humans. One example of this kind of systems is the Roboreceptionist at Carnegie Mellon University [21]. This robot is equipped with a virtual face capable of expressing emotions.

And, thirdly, one of the most well-known social robots is Kismet [13] at the Massachusetts Institute of Technology. Its developed architecture consists of emotions, the robot's internal goals, and a behavior system. Furthermore, [48] described the interactive robot Maggie. Nevertheless, the purpose of the majority of these architectures is to work on only one robot and is optimized for working within only few environments context.

As stated above, the present work introduces a mobile robot able to express emotions using facial expressions, gestures, and speech in order to improve the relation between humans and robots.

2.3 Online Human-Assistance in Computer Vision

Object recognition is a very active topic in computer vision, with impressive results in spite of the difficulties inherent to this problem, such as lighting changes, partial occlusions, intra-class variations, and object's changes in appearance due to multiple views [25].

However, most of the methods are trained offline, either because they use large amounts of training data or because they require complex and time-consuming learning algorithms [14]. Nevertheless, there are some situations in which offline learning is not feasible, for example when the training data is obtained continuously, or when the size of the training set is very cumbersome. Another obstacle is presented when the learning is carried out with unknown objects. Such is the case in the present work, wherein the robot interacts with people so as to learn to recognize their faces from scratch, without any prior information or previous contact.

These kinds of scenarios have been addressed by online learning methods that use their own predictions to train and update the classifiers [55]. However, although approaches have shown great results using these adaptation capabilities, they are prone to suffer from drifting when updating the classifier with inaccurate predictions.

To increase the robustness of online learning algorithms, recent approaches have proposed using human assistance during the learning stage. In [56], a face classifier is computed on the fly, progressively updatting and improving the use of its own predictions and human corrections. Specifically, this method combines self learning and human assistance to avoid the drifting problem, and to teach the robot to discern accurate predictions from inaccurate ones.

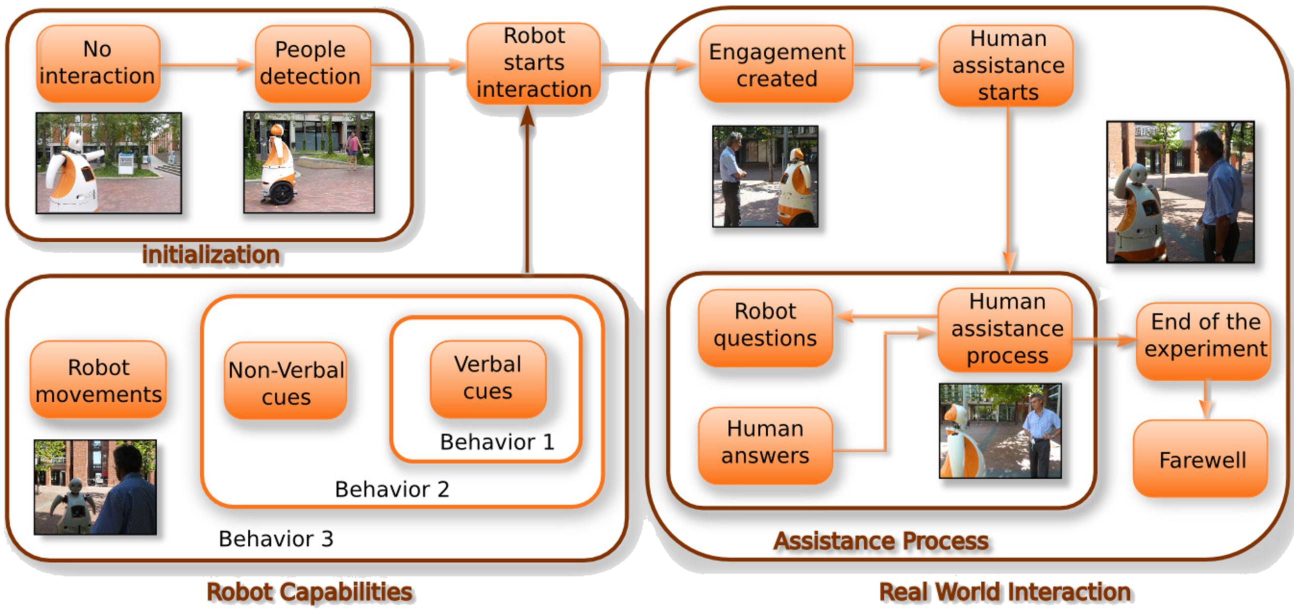

Fig. 2 Experiments overview. Sketch of the experiments performed to analyze different robot behaviors

Here, we go one step further, by introducing our social mobile robot to the pipeline of [56], and providing the robot with the ability to learn from a human, using a communication process that requires almost no-human effort. The robot is able to recognize the volunteer after following the interaction by saving the face descriptors. We believe that the integration of this kind of high-level learning algorithms in an autonomous robot, as well as the development of the engagement strategies, are important contributions for the HRI community, as they suggest that the robot will be able to learn from human interventions.

3 Overview of the Work

The next sections describe the architecture we developed to provide autonomous mobile robots with proactive behaviors. Our goal was to study previous approaches and take them one step further, by encouraging the robot to actively seek out human interaction and ask the person to help it improve its visual detection skills. The main obstacle in this scenario was the possibility that the person would not understand that the robot was trying to initiate a conversation with him/her. Humans typically initiate conversation by eye gaze [22], and in a real environment, it is very difficult for a robot to recognize this social gesture. Because of this, we relied more heavily on body position, gestures, and verbal cues. Once the human had effectively understood the robot's intentions, he/she could follow a specially-made, simple, and efficient communication protocol for teaching the robot. The protocol, developed specially for the purposes of this stage of the study, involved the following key components, as shown in Fig. 2:

- Robot's ability to proactively seek interaction: one of the main purposes of our research was to identify the optimal robot behavior for initiating interaction with a human. To do so, we analyzed three variations of this behavior, examining scenarios in which: (1) the robot uses only verbal cues to communicate with the participant; (2) the robot uses both verbal and non-verbal cues (e.g., gestures and eye gazes); and (3) the robot uses verbal and non-verbal cues and also moves towards the humans.
- Tibi's emotions: making the robot able to express its emotions, the created engagement is stronger, for that reason we synthesize Tibi's emotions of happiness, elation or surprise, among others, we use the model of the three dimensions of emotion [6], which characterizes emotions in terms of *valence*, *stance* and *arousal*. Rather than presenting emotions in terms of categories (happiness, sadness, frustration, etc.), some psychologists conceive of the dimensions that include the relationships between different emotions; this model has been developed in our robot in order to make it capable of expressing its internal emotions.
- Active learning for online face recognition: once the robot has engaged with a human, we propose an approach in which the robot is able to enhance its visual skills using the human's help. We will show that the robot's skills improved with each interaction.
- User study of robot's behavior: we also conducted a user study to determine whether the robot's behavior was perceived as socially appropriate by the experiment participants. We looked at various key aspects of the

interaction between a mobile robot and untrained human volunteers.

4 Social Robot Navigation

Before we review the contributions of our work, in this section we present the navigation method developed in order to allow the robot to be able to approach a person in an acceptable motion.

Here, we make use of the Social-Force Model (SFM), described in [27], to model robot navigation. SFM simulates pedestrian dynamics by using a set of interactive forces. It introduces a very general framework in which the details of human motion and behavior are expressed as a function of the pedestrians' relative and absolute positions and velocities. However, this model does not consider the interaction between a person and a robot, nor the interaction between obstacles. This was later considered in the so-called Extended Social-Force model (ESFM) [17]. We have applied this approach with some modifications to allow the robot to navigate in a friendly manner.

We consider the robot as a social agent, moving naturally in human environments according to the Extended Social-Force Model, and responding appropriately to the obstacles and people in its path. Furthermore, we believe that a more human-like navigation will increase the robot's acceptance among pedestrians, due primarily to the similarities between the robot's behavior and the anticipated movements of other pedestrians.

To this end, we describe robot navigation, understood as an instantaneous reaction to sensory information, driven by the social forces centered on the robot, as in the research conducted in [33], but focusing more on the social nature of the approach.

We first define an attractive force to the person with whom the robot is attempting to establish engagement. Assuming that robot tries to adapt its velocity within a *relaxation time* k_r^{-1},$\boldsymbol{f}_r^{goal}$ is given by:

$$\boldsymbol{f}_r^{goal} = k_r(\boldsymbol{v}_r^0 - \boldsymbol{v}_r) \tag{1}$$

where, v_r is the actual robot's velocity, and v_r^0 is the desired velocity. The relaxation time is the interval of time needed to reach the desired velocity and the desired direction.

Additionally, we define a set of repulsive forces due to the presence of other pedestrians:

$$\boldsymbol{F}_R^{per} = \sum_{p_j \in \mathcal{P}} \boldsymbol{f}_{R,j}^{int} \tag{2}$$

where forces $\boldsymbol{f}_{R,j}^{int}$ represent the repulsive interaction between the pedestrian p_j and robot R:

$$\boldsymbol{f}_{R,j}^{int} = A_{Rp} e^{(d_{Rp} - d_{R,j})/B_{Rp}} w(\varphi_{R,j}, \lambda_{Rp}) \tag{3}$$

The parameters $\{A_{pR}, B_{pR}, \lambda_{pR}, d_{pR}\}$ rule the kind of person-to-robot interaction, and depend on the specific robotic platform being used [15].

Regarding the interaction between the robot and obstacles, we consider the model:

$$\boldsymbol{F}_R^{obs} = \sum_{o \in \mathcal{O}} \boldsymbol{f}_{R,o}^{int} \tag{4}$$

where $\boldsymbol{f}_{R,o}^{int}$ is written as

$$\boldsymbol{f}_{R,o}^{int} = A_{Ro} e^{(d_{Ro} - d_{R,o})/B_{Ro}} w(\varphi_{R,o}, \lambda_{Ro}) \tag{5}$$

The parameters $\{A_{Ro}, B_{Ro}, \lambda_{Ro}, d_{Ro}\}$ rule the interaction person-obstacle.

Finally, the force governing the robot movement can be written as the weighted combination of all previous components:

$$\boldsymbol{F}^R = \alpha\, \boldsymbol{f}_{R,i}^{goal} + \beta\, \boldsymbol{F}_R^{per} + \gamma\, \boldsymbol{F}_R^{obs} \tag{6}$$

Once the reactive force action is obtained, the system responds duly to these stimuli, and linearly propagates its position and velocity according to this force value (Fig. 3). A detailed study of all these parameters values was presented in [16].

Additional constraints are also considered. All those robot propagations which result in a collision with an obstacle are discounted. Current robot maximum velocity is also a constraint, and depends on the robot's navigation state, which is a function of the proximity of persons:

$$v_R = \begin{cases} v_{safety} & \text{if } d_{R,i} \leq \mu_{safety} \\ v_{cruise} & \text{if } \mu_{safety} < d_{R,i} \leq \mu_{social} \\ v_{free} & \text{otherwise} \end{cases} \tag{7}$$

The v_{safety} is the maximum velocity the robot can achieve when at least one person is inside its inner safety zone. In contrast, v_{cruise} is the cruise velocity when someone is inside its social safety zone and v_{free} is the maximum robot velocity when there are no people inside its safety zone. The navigation states associated with these configurations are those of *social robot navigation* and *free robot navigation*, respectively.

The most interesting part of the system so far, resides in the fact that the proposed approach does not require static targets, the robot is able to move towards people. Moreover, it can approach those people who share a common destination.

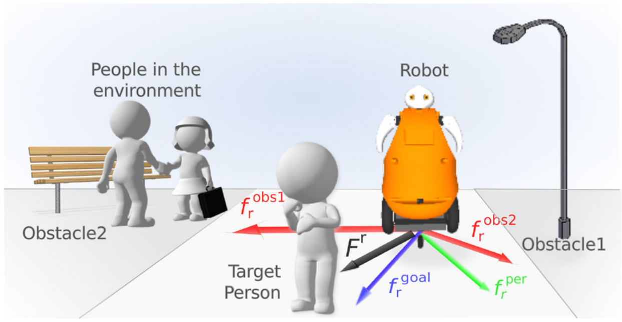

Fig. 3 Diagram the navigation forces. Internal elements of Tibi and a person during the experiments

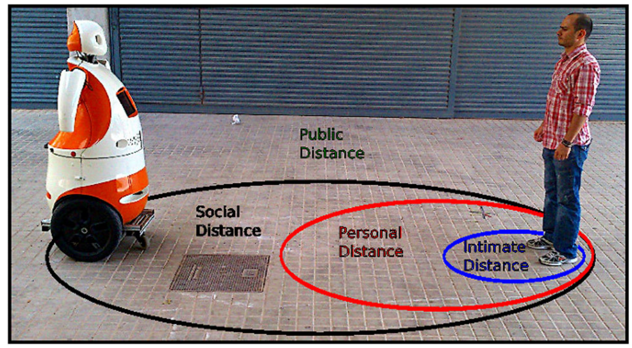

Fig. 4 Levels of engagement. Robot-to-person levels of distance, to distinguish levels of engagement while interacting

5 Robot's Proactively Seeking Interaction

The strategy for creating people-to-robot engagements is more proactive than those models which merely wait for the person to begin the interaction. In addition, the robot's ability to approach people opens up a wide range of possible applications. These include an invitation service, wherein, for example, a robot might approach people to offer city information and invite them on a tour; or the application proposed above, where proactive behavior is used to improve the robot's perception and visual skills by enabling it to learn from the human it engages with.

To allow the robot to independently initiate interaction with humans, we first used a laser range scanner to detect people in the space [2]. In the first part of the experiment, we make use only of the laser range scanner, since the person could be too far away, or there might be lighting changes, rendering the robot incapable of detecting pedestrians using only vision. After this initial localization phase, the robot approaches the person, always adhering to common conventions of what constitutes people's personal space. We also make the robot able to respond appropriately to human reactions. For example, if after the initial approach, the robot invites the selected person to come closer, and he/she does not notice, the robot will repeat the invitation. However, if the human simply declines to come closer, the robot will choose another volunteer. The robot will not begin the interaction process until the person visibly shows interest in the robot.

To define spatial bounds, we considered the conceptual framework known as "proxemics," proposed by Hall [24]. This research establishes the following taxonomy of distances between persons within a group of people:

– Intimate distance: the presence of another person is unmistakable, close friends or lovers (0–45 cm).
– Personal distance: comfortable spacing, friends (45 cm–1.22 m).
– Social distance: limited involvement, non-friends interaction (1.22–3 m).
– Public distance: outside circle of involvement, public speaking (>3 m).

Based on these proxemics, Michalowski et al. [41] classified the space around a robot in order to distinguish human levels of engagement while interacting with or in the presence of a robot. Figure 4 plots these four levels of distance and their corresponding engagements. In our framework we used the proxemics shown in the figure to try to maintain a "social distance" in the initial approach, assuming a "personal distance" only when the person had accepted the invitation to interact.

The robot's active behavior is implemented through the state machine shown in Fig. 5. Finite State Machines (FSMs) are widely used in many reactive systems to describe the dynamic behavior of an entity. The theoretical concepts of FSMs and an entity's specification, in terms of state transition diagrams, have been used for quite some time [19]. A deterministic finite state machine is a quintuple $(\mathcal{K}, \mathcal{H}, s_0, \varkappa, \mathcal{F})$, where: $\mathcal{K}$ is a finite, non-empty set of symbols; $\mathcal{H}$ is a finite, non-empty set of states; $s_0 \in \mathcal{H}$ is an initial state; $\varkappa$ is the state-transition function, $\varkappa\colon \mathcal{H} \times \mathcal{K} \rightarrow \mathcal{H}$; and $\mathcal{F}$ is the set of final states, a (possibly empty) subset of $\mathcal{H}$.

This state machine allows the robot to respond appropriately to people's behavior. The robot is able to determine if humans are interested in initiating interaction simply by tracking their positions.

One of the main objectives of our study was to determine the optimal mode of robot behavior for initiating interaction with a human. After reviewing the literature on empathy and pro-social behavior [10], we were able to identify three different modes of behavior: (1) the robot uses only verbal cues to communicate with the participants; (2) the robot uses both verbal cues and non-verbal cues (gestures and eye gazes); and (3) the robot performs verbal and non-verbal cues, and effectively approaches humans.

After the initial interaction has been established and the human has accepted it, the goal for the robot is then to

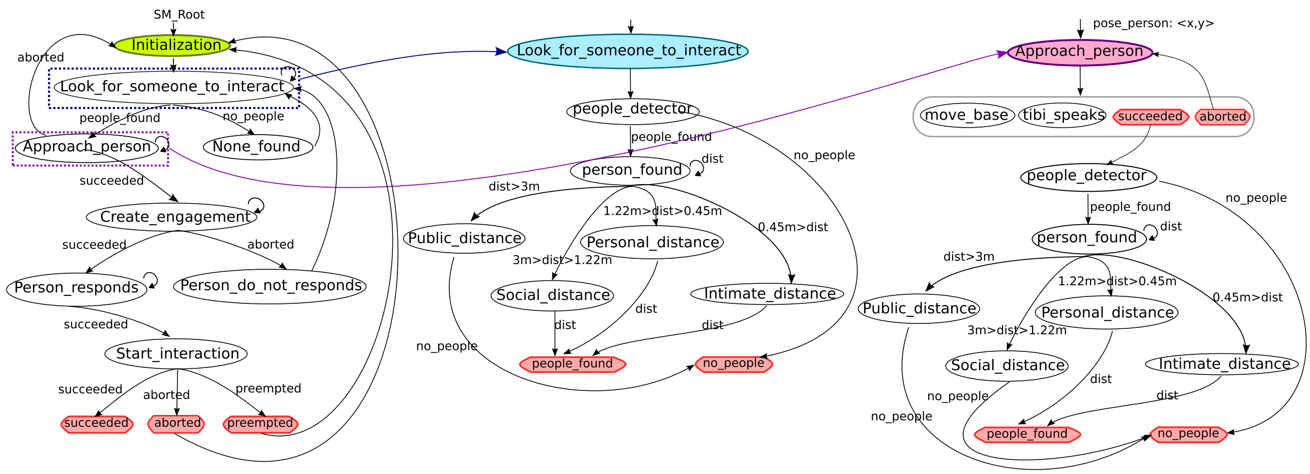

Fig. 5 Example of a state machine. The robot attempts to create an engagement with a person. Different components of the state machine

Table 1 Engagement expressions. Sample Tibi phrases to start interaction with a person

Assistance expressions	
Invitation to create an engagement	Hello, I am Tibi. I'm trying to learn to detect faces, could you help me?
	Hi, I am Tibi, I would like to learn to recognize different objects, will you be my teacher?
Invitation to continue the interaction	It will take just 2 min
	Please, don't go
	Let me explain you first the goal of this experiment, and then, you can decide if you want to stay

approach the person, moving from a "public distance" level to a "personal distance" level. In order to encourage the person to move even closer, the robot performs the following actions, depending on the aforementioned behaviors:

- Verbal communication: comments of encouragement, such as "Don't be afraid, I just want to talk with you." "Could you teach me to detect faces?"
- Non-verbal communication: gestures, arms and neck movements. A few samples are shown in Fig. 6.
- Robot motions: the robot approaches the person until reaching a "social distance."

Some phrases uttered by our robot are presented in Table 1.

It has been found that each of these strategies has a different impact on users. For that reason, we performed a set of experiments to analyze the relative acceptability of each behavior model.

6 Tibi's Emotional Model

Emotions play a significant role in human behavior, communication and interaction [6]. Accordingly, robot's emotions are important in our system. In order to bring the robot closer to humans, we gave the robot the ability to express its emotional status through speech and gestures.

To synthesize Tibi's emotions of happy, elated, surprised, relaxed, tired, bored, unhappy or angry , we used the model of the three dimensions of emotion suggested in [50]. This model characterizes emotions in terms of *stance* (open/close), *valence* (negative/positive) and *arousal* (low/high), thereby, it allows the robot to derive emotions from physiological variables. Our system relies on an open stance because Tibi is motivated to be openly involved in interaction with humans (see Fig. 7).

6.1 Arousal Factor

The arousal factor is determined by the human and the human's responses, and by factors such as whether Tibi finds the human, and whether the human responds. The intensity of the perceived stimuli is required for the implementation of the arousal factor. Furthermore, the perception system is able to rate the current state of engagement between the human and Tibi. In the current implementation, distance is used to measure intensity. Theses computations are based on the distance zones as described above (see Fig. 4).

The intensity of a human who stays in the public zone is rated at zero, whereas a person entering the intimate zone is assigned the maximum intensity value. The relative intensity of a person is more relevant than the absolute value. If a human enters the personal zone (from the social zone), intensity will increase, and arousal increases as well. Assuming that the volunteer remains in the personal zone, his/her inten-

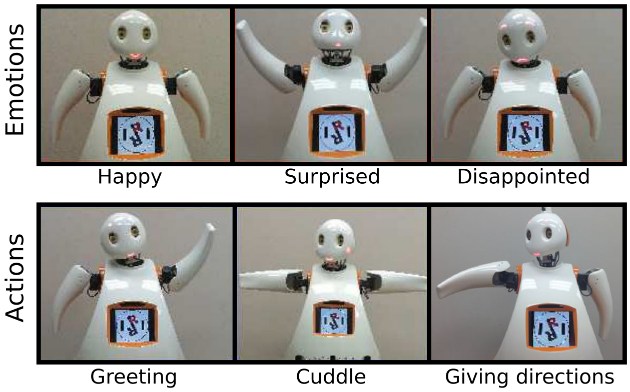

Fig. 6 Tibi gestures. Movements performed by Tibi during experiments. *Top*: three different emotional expressions. *Bottom*: three actions Tibi can perform

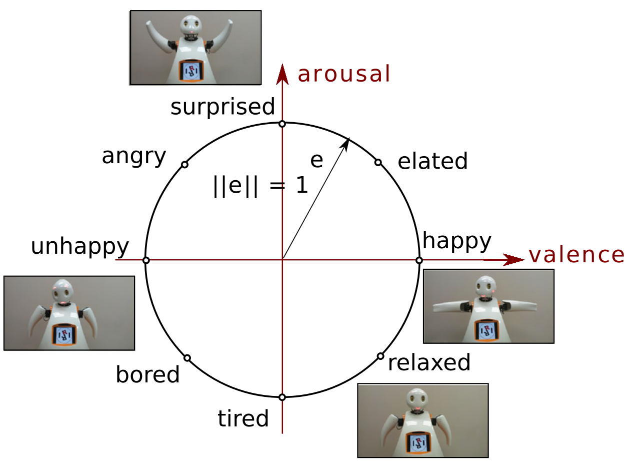

Fig. 7 Emotion space. This representation is used to define the actual emotional state of Tibi; every emotion can be described by the parameters arousal and valence

sity will remain the same, but the arousal level should not rise any further, as this would indicate a state of continuous fright.

To avoid these problems, the relative intensity is used for the social zone and for the public zone. Therefore, only changes in intensity are considered for calculating arousal. In contrast, for the intimate zone, absolute intensity is used. Based on these assumptions, the global intensity of the people currently detected can be calculated. The currently perceived people located within a specific distance zone are represented by $\Phi(t) = \{\phi_1, \ldots, \phi_n\}$, whereas the people previously located in this zone are described by recognized $\Phi(t-1)$. For each zone, a specific intensity level (zone level) is represented. In the current implementation these zone levels (ζ) are defined as follows: Public zone $\zeta = 0$, social zone $\zeta = 0.25$, personal zone $\zeta = 0.5$, and intimate zone $\zeta = 1$. This process is summarized in Algorithm 1.

To normalize the intensity to the range of [0, 1] the intensity value is divided by the number of currently perceived people. If a certain intensity has been detected, the previous

---

Algorithm 1: The intensity of perceived people is computed depending on their distance to Tibi.

Input: List of perceived people at time t: $\Phi(t)$
Input: Distance between each person $\phi_j(t)$ and the robot at time t: $\delta(t, j)$
Input: Size of $\Phi(t)$: N
Output: Intensity value: ι
for $r = 1, \ldots, N$ **do**
  Compute ζ
  Compute the intensity ι:

$$\iota(t) = \begin{cases} \iota + (\phi_j(t) - \phi_j(t-1))\zeta & \text{if } \zeta = 0.25, 0.5 \\ \iota + (\phi_j(t))\zeta & \text{if } \zeta = 1 \\ \iota & \text{otherwise} \end{cases}$$

Algorithm 2: Arousal: $A(t)$

Input: Intensity value at time t: ι
Input: List of perceived people at time t: $\Phi(t)$
Output: Arousal value at time t: $A(t)$
if $\iota > 0$ **then**

$$A(t) = A(t-1) + \omega \cdot \frac{\iota}{\#\Phi(t)}$$

else

$$A(t) = A(t-1) - \Delta \tag{8}$$

Limit $A(t)$ to the interval $[-1, 1]$

---

arousal value $A(t-1)$ is increased depending on the global intensity and a specific weight ω that indicates how fast the arousal value increases. In this work, the weight is set to 1. If no intensity is measured the arousal value is decreased. The value for decreasing is represented by Δ. Inspired in [28], the current Δ is set to 0.25. Finally the arousal value is limited to the range of [-1, 1], Algorithm 2 describes the process to compute the arousal value.

6.1.1 Valence Factor

Valence represents the robot's satisfaction with the current situation. For example, achieving a goal will cause an increase in valence. This depends on the current achievement of the internal goals of the robot. For instance, if the robot is currently pursuing one goal, the valence depends on the level of the achievement of the robot's internal goals. If the goal is almost achieved, the valence will be rather high; if the robot is far from achieving this goal the valence is low. If the robot is pursuing multiple goals, the valence is calculated based on the level of achievement associated with each goal.

Table 2 Assistance expression. Sample phrases uttered by the robot when updating the visual classifier

Assistance expressions	
Assistance	Can you tell me if your face is inside the rectangle?
	Is the detection correct?
No detection	I can't see you properly, move a little bit
	Can you stand in front of me?
Farewell	I'm so happy you helped me
	I hope to see you soon

In this study, valence is determined by whether the human responds appropriately to the robot's requests. As Tibi waits for a human response, indicated by pressing a "*yes*" or "*no*" button, if the human says something unexpected that Tibi cannot understand or if he/she fails to press either button, the negative response increases the emotion of anger; while a positive response leads to an increase in the emotion of happiness.

7 Active Learning for Online Face Recognition

We devised an approach to allow the human to improve the performance of the robot's visual skills, once the robot has initiated engagement.

The second objective of the present work is to allow our robot to benefit from the human's assistance. To this end, we equipped it with a screen depicting the results of the face detector. The robot was able to use verbal cues and gestures to be able to communicate with the human user. When the robot was not confident about the presence of a face in the input image, it requested the human's help, through a set of precise questions, which the human user could answer by pressing the "*yes*" or "*no*" button, using the Wii remote control. Table 2 shows some examples of these questions. The robot explains to the human how the Wii remote control functions within the context of the experiment.

Figure 8 illustrates entirety of the interaction between Tibi and a volunteer considering the internal elements.

Concretely, the goal of this section is to enhance the human-assisted facial recognition system based on the degree of human intervention and its effects on human-robot interaction. In particular, we focused on the duration of the established interactions and on the level of users' comfort therein.

The classifier used in the detection phase yields a score $\varsigma \in [0, 1]$, corresponding to the classifier confidence. Usually, when $\varsigma > 0.5$, the detection is assigned to a positive or object class (in this case, faces). Otherwise, the detection is considered as negative or belonging to the background class. However, there is a confidence interval ϑ around 0.5 in which the system is unable to assign the detection to a positive or negative class, due to the fact that the classifier is uncertain about the detection label (positive or negative) and the risk of misclassification is high. In these cases, we resort to the human's intervention to determine whether the detection belongs to a particular human face or falls in the background or on an incorrect person. This is then used to improve the classifier performance by updating the classifier only with correctly labeled detections. By conducting these experiments, we hoped to discover the range and degree of human assistance by which interaction becomes more effective.

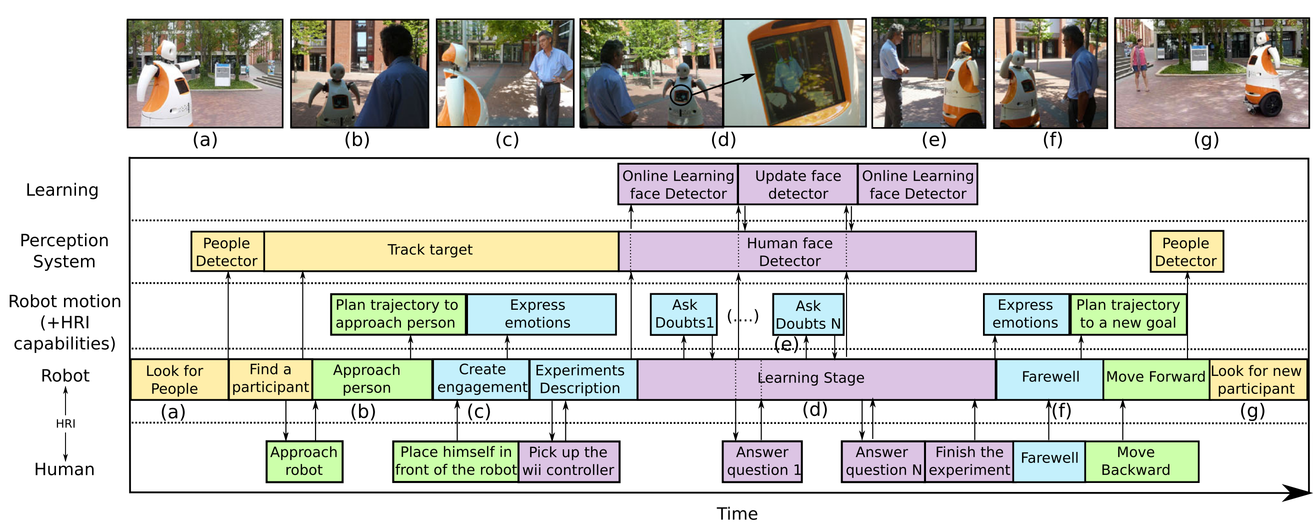

Fig. 8 Diagram of the interaction. Internal elements of Tibi and a person during the experiments

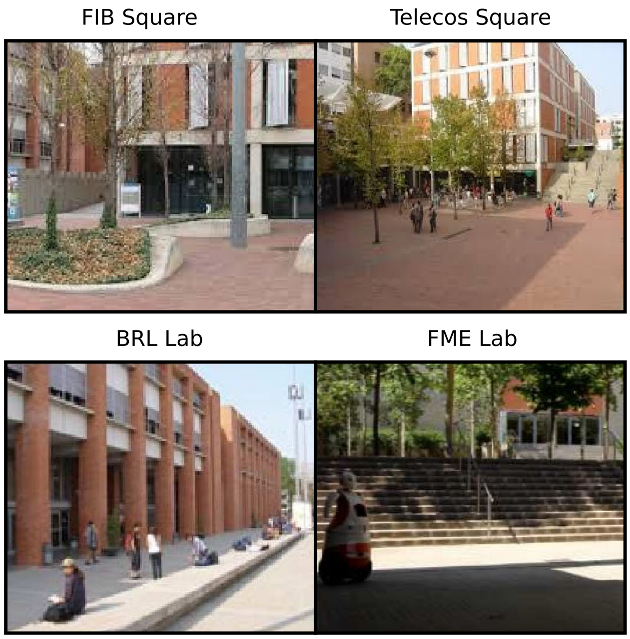

Fig. 9 Enviromental labs. Four different labs in UPC Campus where we performed the experiments with Tibi robot

8 Experimental Field of Study

In order to present the most realistic findings, we performed our analysis within real scenarios in the city of Barcelona. The tasks the robot had to accomplish were to: (1) approach a person to initiate interaction and stimulate interest in helping the robot; and (2) invite the person to help to enhance the robot's facial recognition system. In this section, we begin by introducing the robot and the environment domain; we move on to describe the experiment design and procedure; and finally we conclude with a review of the results obtained and the user study subsequently performed.

8.1 Robot and Environment Domain

Our working area consists of four different outdoor urban environments at the UPC Campus: the FIB square and the Telecos square in the North campus of UPC, the BRL (Barcelona Robot Lab), and the FME (Facultat de Matemàtiques i Estadística) lab in the South campus.

The North Campus is a large area outfitted as an experimental zone, covering over 20,000 m^2, and comprising different buildings squares, with multiple ramps, staircases, and typical obstacles such as bulletin boards, bicycle stands, trashcans and flower pots. The FME lab consists of a green space and a paved area, separated by stairs, see Fig. 9.

Tibi is a service robot, designed to operate in urban, pedestrian areas. It is based on a two-wheeled, self-balancing Segway RMP200 platform, and as such, is highly mobile, with a small footprint, a nominal speed of up to 4.4 m/s, and the ability to rotate on the spot (while stationary).

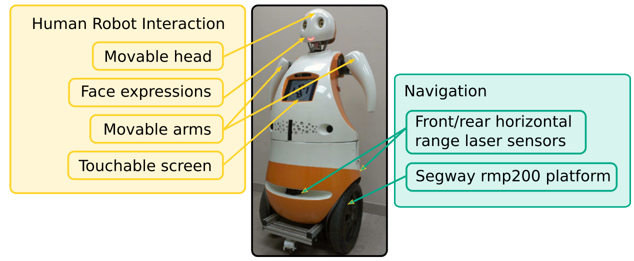

Fig. 10 Tibi robot. Mobile robot platform used in the experiments

The Tibi robot is 165 cm in height, occupies a clearance space of 80 cm, and weighs 110 kg. It is equipped with the following sensors, (see Fig. 10): two Hokuyo UTM-30LX 2D laser range sensors used to detect obstacles and people, giving scans over a local horizontal plane at 40 cm above ground, facing forward and backwards; a stereo Bumblebee camera located in the eyes, used for computer vision purposes; a touch screen to communicate with people; a speaker, movable arms and head to express emotion; two on-board computers (Intel Core 2 Quad CPU 2.66 and 3.00 GHz) which manage all the running processes and sensor signals; and finally, a laptop used for external monitoring.

The robot's communication is spoken out loud for the participants to hear and is also displayed on the touch-screen, but participants had to answer using a remote Wii control, as Tibi cannot understand speech in outdoor environments.

Moreover, Tibi was designed in order to interact with different people in open spaces. The robot is socially accepted, and humans take an interest in interacting with it, as its design is well-rendered, and its movements are smooth.

8.2 Experiment Design

To test our framework, we conducted the following experiments, wherein which the Tibi robot moved around the University Campus:

– Robot's Proactively Seeking Interaction: we compared the different robot behaviors described in Sects. 5–6 to initiate the interaction. In the initial phase, the robot only used voice instructions to attract people's attention. Following that, it was allowed to rotate to observe people's position. Finally, the robot had the capability to move towards people to interact with them. In all of these situations, the robot was able to effectively express its emotions.
– Active Learning for Online Face Recognition: we analyzed the effect of the human assistance on the robot's face recognition performance, and on the duration and ease of the human-robot interaction.

8.3 Experiment Procedure

8.3.1 Robot's Proactively Seeking Interaction

Our independent variables took into account whether the robot approached the person, or if it only used voice instructions. The main dependent variables involved participants' perceptions of the robot's persuasiveness, their compliance with the robot's suggestions, and their perceptions of the robot's social and intellectual characteristics. Each of these fields was evaluated by each participant through a questionnaire that was completed upon the conclusion of the experiment, based on [34]. The measurement was a rating on a Linkert-scale between 1 and 7, from "Not at all" to "Very much". For the evaluation score, an analysis of variance (ANOVA) measurement was conducted.

The information given to the volunteers about the robot was minimal at the start of the experiment, and hence, their behavior was not predefined at all. Participants were told to behave naturally, to listen to the robot's instructions, and to help it. Volunteers could decide whether to stay and perform the experiment or to skip the test at any time. Once the experiment was completed, participants answered the questionnaire.

8.3.2 Active Learning for Online Face Recognition

The face recognition system used in our experiments is based on the classifier proposed in [56]. This classifier, named Online Random Ferns, interactively computes a discriminative detector that allows the robot to recognize objects and human faces in real time. Although this classifier was shown to improve the recognition performance with higher rates of human assistance, previous studies did not explicitly evaluate the influence of human intervention on human-robot interaction.

In this work, we expanded upon [56] with empirical and quantitative evaluation of the human assistance from the perspective of HRI. The evaluation was carried out based on the interactions between the Tibi robot and several persons in a variety of environmental conditions. More specifically, we evaluated human-robot interaction for online face recognition in terms of the degree of human intervention. To arrive at this measure, we followed the criterion used in [56], wherein a confidence interval ϑ was established to determine when human intervention was required (human assistance interval). However, while in [56] this threshold was set at fixed value, in our study, we evaluated the face recognition module using different values of ϑ, and thus, different degrees of human intervention.

8.4 Participants

For the experiments, we selected 50 people (32 men, 18 women) on the University Campus. Participants ranged in age from 20 to 65 years (M = 35.72, SD = 14.13), and represented a variety of university majors and occupations including computer science, mathematics, biology, finance and chemistry. For each individual selected, we randomly activated one of the three robot behaviors to begin the interaction. Then, each participant assisted the robot to improve its visual skills (second experiment). It should be mentioned that none of the participants had previous experience working or interacting with robots.

8.5 Results in Real-Life Experiments

Before conducting the user study to determine whether different robot behaviors are socially appropriate for humans, we conducted real-life experiments to evaluate the robot's behavior over the course of two weeks. The approach proposed above was effectively tested at the BRL.

Real-world experimentation revealed unexpected obstacles that had not come up during the simulations. We observed severe limitations of the perception system, laser people detector, and tracker. People were not always properly detected, and data association was occasionally wrong. However, an in-depth discussion of the perception system falls outside the scope of the present work.

8.5.1 Robot's Proactively Seeking Interaction

We carried out our experiments with different untrained volunteers over the course of seven days. In each experiment, the robot was able to approach the participant and try to establish engagement. Figure 11 depicts some examples of the experiments performed with several volunteers in different urban environments. Figure 12 shows the paths taken by robots when approaching a person in the four different environments.

Once a significant number of real experiments with different volunteers was conducted, we concluded that the system worked, and that robot was able to approach humans and begin interactions with untrained people. We used these findings to proceed to conduct a user study, designed to determine whether the robot's behavior was socially acceptable to humans. This component is described in depth in Sect. 8.6.

8.5.2 Active Learning for Online Face Recognition

As mentioned above, the human-assisted facial recognition system was assessed based on the degree of human intervention and its effects on human-robot interaction. The classifier we used in the detection and recognition phase generates a

Fig. 11 Real-life experiments: Some examples of the real experiments conducted

score $\varsigma \in [0, 1]$, which corresponds to the classifier confidence. Nevertheless, this confidence interval ϑ is centered on 0.5, by which the system is not able to calculated if the detection is of a positive or negative class. Here, human intervention is required in order to determine if the detection belongs to a particular human face or falls in the background or on an incorrect person. Therefore, the interaction is used to enhance the performance of the classifier, updating it only with correctly labeled detections. In developing these experiments, we endeavor to discover the degree of human assistance at which the interaction becomes more effective. Figure 13 shows different volunteers assisting the robot in the task of face recognition.

Figure 14 shows the impact of human assistance on human-robot interaction. Figure 14-*Top-Left* depicts the average interaction and assistance times. As the degree of human assistance grows greater (and with it, interval size), the interaction time between robot and humans becomes shorter. It is also noteworthy that the interaction time with a smaller percentage of human intervention is relatively short. This is because when human participation is minimal (i.e., when human users seldom help the robot), people also lose interest in the task. Figure 14-*Top-Right* plots the percentage of human intervention for each interval. Again we see that the percentage of human assistance increases according to the uncertainty interval size. The graph on the left-bottom of Fig. 14 depicts the percentage of human acceptance of the robot's behavior. Finally, Fig. 14-*Bottom-Right* depicts the percentage of ignored requests. Note that as the number of times the robot asks for assistance increases, the number of ignored requests also increases. We found that a satisfactory compromise between the human's effort and interaction time was achieved for an assistance interval of $\vartheta = [0.4, 0.6]$. In other cases, people grew bored and thus the interaction failed as people declined to complete the experiments.

8.6 Measures

The results presented in the previous section demonstrate that the robot is able to approach people and initiate interaction, and that visual skills may be enhanced using human assistance. A user study was also conducted to determine whether the three strategies presented previously to initiate the interaction are perceived by people as socially appro-

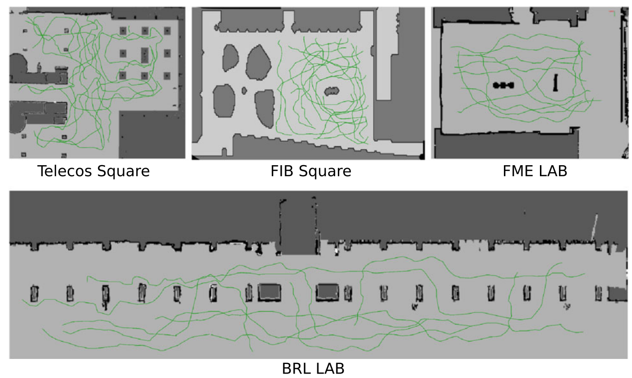

Fig. 12 Paths followed by Tibi to initiates an interaction. The robot approached different people and begins an interaction with them in the four different urban environments we performed the experiments

Fig. 13 Human assistance. *Top*: people assisting Tibi robot in outdoor scenarios. *Bottom*: Tibi's field of vision. The output of the recognition system is shown by *rectangles*. Correct detections are represented by *green boxes*; *blue boxes* indicate when the system is not confident and requires the help of a human

priate. Finally, we concluded this section by studying how our social navigation enhances a follower approach, wherein the robot only follows the person's trajectory, without considering any social conventions, and we should highlight that people perceived a difference between these two approaches.

The hypothesis we endeavored to test was as follows: "Participants will perceive a difference between the three robot behaviors and will assist at a greater rate when the robot is able to move and approach people according to accepted social conventions."

We compared the different robot behaviors for initiating interaction, as described in Sects. 5–6. At first, the robot used only verbal instructions to attract people's attention. Later, it

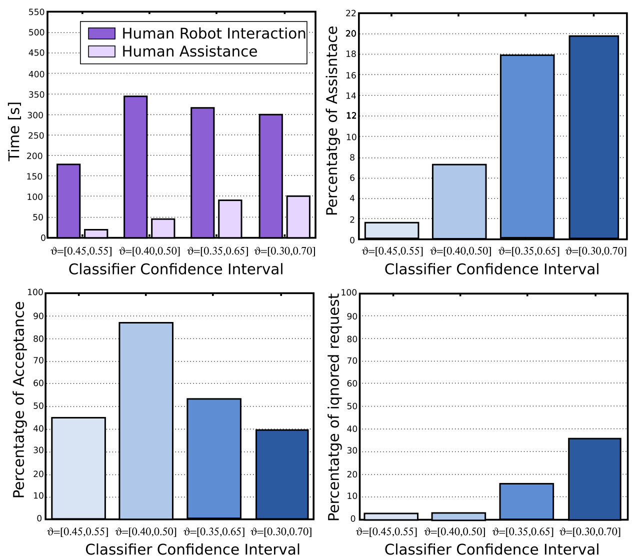

Fig. 14 Human assistance results. *Top:* Average times spent for human-robot interaction and human-assistance. And, percentage of human assistance in the face recognition system according to varying assistance intervals. *Bottom*: percentage of users' acceptance; percentage of ignored request

Table 3 Questionnaire. Survey questions asked of each participant. All questions were asked on a 7-point scale from "Not at all" to "Very much"

Survey's questions	
General robot behavior scale	Cronb. $\alpha = 0.74$
How comfortable did you feel near the robot?	
How safe did you feel around the robot?	
How human-like did the robot behave?	
Robot's sociability scale	Cronb. $\alpha = 0.82$
How social was the robot's behavior?	
How natural was the robot's behavior?	
How well did the robot's movements adhere to human	
Social norms?	
Robot's intelligence scale	Cronb. $\alpha = 0.79$
How intelligent did the robot behave?	
How well could the robot anticipate to your movements?	
How well could the robot understand your responses?	

was allowed to rotate so as to focus more closely on people's positions. Lastly, the robot was able to move towards the people to interact with them. Assistance could begin only once engagement had been initiated.

As we described in Sect. 8.4, we selected 50 people. For each participant, we randomly activated one of the three robot behaviors for initiating interaction. Then, each participant helped the robot to improve its visual skills. Again, none of the participants had previous experience working or interacting with robots.

Participants were asked to complete a variety of surveys. Our independent variables considered whether the robot approached the person or if it only used voice instructions. The main dependent variables involved participants' perceptions of the robot's persuasiveness, their compliance with the robot's suggestions, and their perceptions of the robot's social and intellectual characteristics. Each of these fields, was evaluated by every participant using a questionnaire to fill out after the experiment, based on [34]. Some questions are presented in Table 3.

8.6.1 Social Scales

Participants were asked to answer nine questions, as shown in Table 3, following their encounter with the robot in each mode of behavior. To analyze their responses, we grouped the survey questions into three scales: the first measured overall robot behavior, while the second and third evaluated more specific questions on the robot's movement. Both scales surpassed the commonly used 0.7 level of reliability (Cronbach's alpha).

Each scale response was computed by averaging the results of the survey questions comprising the scale. ANOVAs were run on each scale to highlight differences between the three robot behaviors.

Below, we provide the results of comparing the following three robot behaviors: *(B1)* the robot only uses verbal communication; *(B2)* the robot uses both verbal communication and gestures; and *(B3)* the robot uses verbal, nonverbal communication and may approach the person.

For the global evaluation score plotted in Fig. 15-*Left*, repeated ANOVA measures were computed. A significant main effect was found, $F(2, 47) = 41.52$, $p < 0.001$, $\eta^2 = 0.29$. Multiple comparisons with the Bonferroni method revealed that the score for *B3* is significantly higher than both behaviors *B1* ($p < 0.001$) and *B2* ($p < 0.001$). No significant difference was found between *B1* and *B2* ($p = 0.224$).

To analyze the source of the difference, additional scores were examined. For the sociability of the robot (Fig. 15-*Center*) a repeated-measures analysis of variance revealed a significant main effect, $F(2, 47) = 143.83$, $p < 0.001$, partial $\eta^2 = 0.14$. Pairwise comparison with Bonferroni showed a remarkable difference between the three strategies as well. *B1* vs. *B2*: $p < 0.01$; *B1* vs. *B3*: $p < 0.001$; *B2* vs. *B3*: $p < 0.001$.

Finally, for the robot's intelligence (Fig. 15-*Right*), a repeated-measures analysis of variance revealed a significant main effect, $F(2, 47) = 32.28$ $p < 0.001$, partial $\eta^2 = 0.31$. Pairwise comparison with Bonferroni revealed that the score

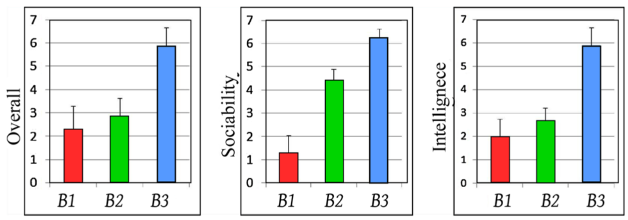

Fig. 15 HRI results I. Degree of acceptance of the three robot's behaviors. *Left*: global evaluation of the strategies. *Center*: robot's sociability. *Right*: robot's intelligence, as perceived by the humans

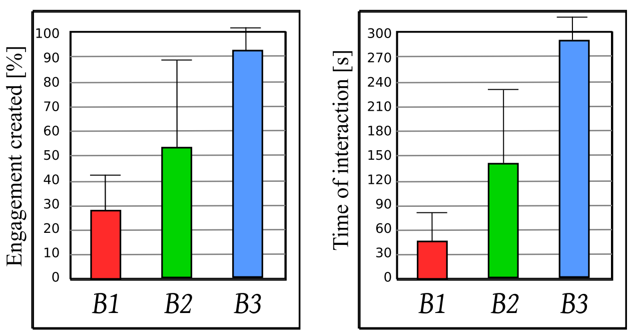

Fig. 16 HRI results II. *Left*: percentage of engagements created. *Right*: time of interaction in seconds

for *B3* is significantly higher than both *B1* ($p < 0.001$) and *B2* ($p = 0.0015$) strategies. No significant difference was found between *B1* and *B2* ($p = 0.42$).

In summary, from our analysis of the three different behaviors, we may conclude that when the robot uses verbal and non-verbal communication, and is able to approach the person, it has the largest rate of acceptance by humans. Under these circumstances, people generally perceived the robot to be more intelligent, seeing as it could detect and approach them; they also believed that it had better social skills.

Furthermore, we measure the percentage of successful goals, that is, the number of times the robot was able to create an engagement with the person, and we compared the three robot's behaviors. In Fig. 16-*left* the percentage for the three behaviors is plotted. And, finally, in Fig. 16-*right*, we show the duration of the interaction for the three behaviors. Note that when the robot is able to approach the person who is interacting with the duration and the interest of the volunteer is much larger.

In addition, we studied if the presented emotional model is well-perceived by participants. Here, to examine whether there are differences between the use of model of emotion or not, two scores were examined: "overall" and "robot's sociability", plotted in Fig. 17. Moreover, we compared the duration of the interactions. For the global evaluation and sociability, score plotted in Fig. 17-*Left, Center*, pairwise comparison with Bonferroni demonstrate a difference between the use of the emotional model, $p < 0.001$, in both cases. In terms of the duration of the interaction, it can be

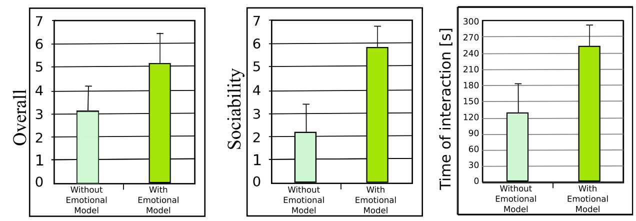

Fig. 17 HRI emotional model results. Degree of acceptance of the Emotional Model. *Left*: global evaluation if Tibi uses an emotional model or not. *Center*: robot's sociability. *Right*: duration of the interaction

seen, that if Tibi is able to express its emotions the duration of the experiments are longer.

Hence, once the three components has been analyzed, we can conclude that if our robot Tibi makes use of the emotional model, it has the largest acceptance. People perceived the robot to be more sociable, and the duration of the interactions were longer.

Finally, human perception has been studied in the navigation skill. To analyze the source of the difference, three scores were examined: "overall", "robot's sociability" and "robot's intelligence", plotted in Fig. 18. For the global evaluation score plotted in Fig. 18-*Left*, pairwise comparison with Bonferroni demonstrate a difference between the two kind of navigation approaches, $p < 0.001$. In terms of robot's sociability and intelligences the volunteers also perceived a difference between the two navigations, $p < 0.01$ in both cases.

Therefore, after analyzing these three components in navigation terms, we may conclude that if the robot has the ability to socially navigate and respect human conventions, it has the largest acceptance. People perceived the robot to be more intelligent more sociable.

8.6.2 Participants Comments

Each questionnaire included several blank lines underneath the social scales, where participants could include additional comments about the experiments. While we did not explicitly codify and analyze these comments, they do provide further insight into the effect of the three robot behaviors.

8.6.3 Comments When the Robot Uses Only Verbal Communication (B1)

Many of the participant comments reflect that the robot did not attract the attention. E.g.:

"I didn't think the robot was talking to me, because it wasn't moving."

"The only quality I can attribute to him is that he knew when I was walking around him."

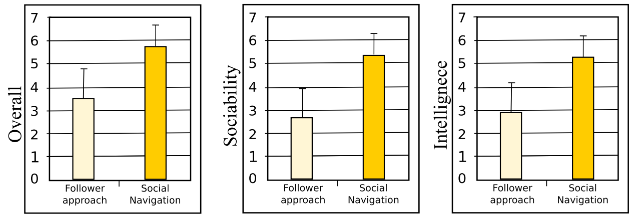

Fig. 18 HRI navigation results. Degree of acceptance of the robot navigation. *Left*: global evaluation of the two navigations. *Center*: robot's sociability. *Right*: robot's intelligence, as perceived by the humans

"The fact that the robot didn't move made it difficult for me to know whether it was interacting with me or not."

"The robot attracted my attention because it's cute, but not because of its behavior."

Note that the comments on this behavior indicate that participants felt that the robot did not try to initiate engagement with them.

8.6.4 Comments When the Robot Uses Both Verbal Communication and Gestures (B2)

Many of the comments reflect that the robot did not attract participants' attention to a satisfactory degree. Tibi was considered a social robot, but it was not perceived as intelligent:

"I like when she gestures, and attracts my attention, but I would have preferred that the robot also approached me, not just waited for me to act."

"I love when the robot greets me when I pass nearby, I find it very sociable."

"If Tibi was able to move, it would draw more attention and hold my interest, yet I find it very interesting that I could play the role of a teacher."

"I like that the robot comes do me and doesn't wait for me to approach it before speaking to me."

Note that the comments on this behavior generally indicate that although participants felt that the robot tried to initiate an engagement with them, it was not enough, and most participants wondered if Tibi was moving independently.

8.6.5 Comments When Robot Uses Verbal, Nonverbal Communication, and was Free to Approach the Person (B3)

Many of these comments indicated that participants felt that the robot tried to initiate engagement with them, and they were generally interested in the robot's skills:

"This is the first time I find myself around a robot who interrupts me in order to help me; it's very original."

"Tibi is very polite, and I find it charming that it follows me around until I pay attention."

"I felt that Tibi obeyed social conventions by approaching me and starting the interaction."

"Does it mean that Tibi will be here alone? That's original but may be dangerous for her."

"I feel that the robot is very intelligent because she knows when I'm nearby and approaches me in order to interact. I'd like to know what else she can do." (emphasis in original)

"It's funny that Tibi gets mad when I ignore her; it would be interesting to see if she remembered me next time she sees me."

Note that the comments on this behavior indicate that participants felt that the robot tried create an engagement with them. Moreover, Tibi behaved in a socially acceptable manner and generally understood if people wanted to interact with her or not.

9 Discussion

The findings presented in the previous section reinforce the notion that the robot's ability to initiate engagement is an important skill to master in order to achieve natural interaction with people. Overall, people were surprised to find a robot in a public space, and they were astonished when the robot caught their attention. Moreover, they enjoyed helping the robot to detect their faces and were surprised to see how the robot progressively improved its skills with their assistance.

The experiments we conducted yielded conclusive results. We found that people felt their interaction with the robot was more natural when the robot communicated through gestures, verbal cues, and motion. Detailed analysis showed that these capacities improved the human's perception of the robot's intelligence and sociability. We also found that the amount of speech and comments made by the robot seems to be appropriate for this type of scenario. Moreover, people felt comfortable using the Wii remote control to communicate with the robot. In order to study whether the emotional model improves the quality of the interaction between the robot and the human, we performed experiments to evaluate if the use of the emotional model enhanced the interaction. Volunteers percieved Tibi more sociable and closer when it expresses its emotional model, and the interactions were longer. Furthermore, the presented navigation has been perceived more sociable for users.

We were also able to effectively demonstrate that human assistance helped to enhance visual perception tasks such as online face recognition. The entire process was done with minimal human effort and great efficiency. The results show that the use of a social robot piques people's interest and encourages them to collaborate with the robot to enhance its visual skills.

We noticed that very few participants were capable of specifically naming the robot's disadvantages, and most of them provided helpful suggestions when asked about possible improvements for Tibi. People expressed an interest in communicating with the robot via voice commands, finding that kind of communication to be generally more comfortable. They also suggested that it would be interesting if they could teach the robot to identify new objects by pointing them at the robot's screen. Both of these remarks will be incorporated into our future research.

Finally, we must address some of the cultural limitations of our project. The parameters and definitions for human personal space, employed in the first set of experiments, are specific to European people and to the design of our own robot. Therefore, if this experiment were to be adapted in other cultures, its parameters would need to be adjusted accordingly through experimentation. In addition, the proposed model of interaction was tested in a specific scenario, and so its application in other situations is limited. It is possible that context and environment significantly affect humans' preference for a specific mode of robot behavior. For example, in a business environment, a mobile robot approaching people could be annoying, as its interruptions might disturb people. We believe that the University Campus is rather neutral, and can thus reflect general trends in interaction in many daily use scenarios. However, this question warrants further study.

9.1 When will This Capacity be Used?

We believe that robot's capacity to naturally establish engagement is a major function that should be implemented in future social robots. While other projects have assumed that people and robots can meet and initiate interaction, it has been observed that this is generally not the case in real world scenarios. In principle, robots might not need to initiate interaction themselves, because ideally people would be interested in the novelty and would approach them of their own volition. In these concrete cases, robots would not need to adjust their behavior to initiate interaction.

However, in most cases, humans will not initiate interaction with robots themselves, especially if the robots do not approach them and attract so much attention. Here, robots will often fail to initiate interaction [49].

There are many situations that involve a first meeting, such as a tour guide in cities or museum [8,18], nurse in hospitals [39] or a shopping assistant [52], which are actual and potential applications of social robots.

10 Conclusion and Future Work

We have introduced an autonomous mobile robot that seeks interaction for the purposes of human-assisted learning. The major contributions of this paper are two-fold. First, we have studied different robot behaviors for initiating interaction with humans. We showed that the robot was able to autonomously approach a person and establish an engagement with him/her.

Secondly, once engagement was established, people could assist the social robot to improve its visual skills. Following the assisted learning stage, the robot was able to detect people by using its visual skills even under challenging scenarios, such as when the objects were partially hidden.

Both contributions have been extensively and rigorously tested in a real environment in Barcelona city with non-trained volunteers. Our findings suggest that allowing the robot to take the initiative when communicating with people usually increased the number of human-to-robot interactions. This, in turn, allowed humans to assist robots in improving their visual skills, and engage in subsequent, and more predictable, interactions.

Finally, with respect to future work, humans routinely interact with other people and perform tasks individually and collectively on a daily basis. Robotic researchers are interested in designing robots that can interact with people in the same way as humans do. To be able to reach this goal, robots should learn from their interaction with humans and acquire the humans' skills which are used in our everyday life. The learned social behaviors could be used in a wide range of real-world scenarios, such as, domestic tasks, shopping, assistance, guidance, entertainment, surveillance, or rescue.

There are many examples where these interactions occur, but some of them are so basic that people might not realize the extreme difficulties that come with executing such tasks for a robot. Navigation in crowded environments, or the social engagement required to initiate a conversation, are some examples.

Continuing the work presented in this paper, we plan to develop new techniques to learn from interaction with humans using multi-modal interaction. The models can be learned offline or online, and humans can use information from inputs and outputs to train the system again in order to improve the models. We expect that with these new techniques, the multi-modal interactive system can improve the accuracy and robustness of the methods.

References

1. Anzalone SM, Boucenna S, Ivaldi S, Chetouani M (2015) Evaluating the engagement with social robots. Int J Soc Robot 7(4):465–478
2. Arras K, Mozos O, Burgard W (2007) Using boosted features for the detection of people in 2d range data. In: Proceedings of the IEEE international conference on robotics and automation, pp 3402–3407
3. Atkinson DJ, Clark MH (2014) Methodology for study of human-robot social interaction in dangerous situations. In: Proceedings of the second international conference on Human-agent interaction. ACM, pp 371–376
4. Bauer A, Klasing K, Lidoris G, Mühlbauer Q, Rohrmüller F, Sosnowski S, Xu T, Kühnlenz K, Wollherr D, Buss M (2009) The autonomous city explorer: towards natural human-robot interaction in urban environments. Int J Soc Robot 1(2):127–140
5. Bittermann A, Kuhnlenz K, Buss M (2007) On the evaluation of emotion expressing robots. In: IEEE international conference on robotics and automation, pp 2138–2143
6. Breazeal C (2003) Emotion and sociable humanoid robots. Int J Hum Comput Stud 59(1):119–155
7. Breazeal C (2009) Role of expressive behaviour for robots that learn from people. Philos Trans R Soc B Biol Sci 364(1535):3527–3538
8. Burgard W, Cremers A, Fox D, Hähnel D, Lakemeyer G, Schulz D, Steiner W, Thrun S (1998) The interactive museum tour-guide robot. In: Proceedings of the National Conference on Artificial Intelligence. John Wiley & Sons Ltd, pp 11–18
9. Clark H (1996) Using language, vol 4. Cambridge University Press, Cambridge
10. Cooper B, Brna P, Martins A (2000) Effective affective in intelligent systems–building on evidence of empathy in teaching and learning. In: Affective interactions. Springer, Berlin, pp 21–34
11. Dautenhahn K, Walters M, Woods S, Koay K, Nehaniv C, Sisbot A, Alami R, Siméon T (2006) How may i serve you?: A robot companion approaching a seated person in a helping context. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction. ACM, pp 172–179
12. Dautenhahn K, Woods S, Kaouri C, Walters M, Koay K, Werry I (2005) What is a robot companion-friend, assistant or butler? In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems, pp 1192–1197
13. Fellous JM, Arbib MA (2005) Who needs emotions?: The brain meets the robot. Oxford University Press, Oxford
14. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
15. Ferrer G, Garrell A, Sanfeliu A (2013) Robot companion: a social-force based approach with human awareness-navigation in crowded environments. In: Proceedings of IEEE/RSJ international conference on intelligent robots and systems, pp 1688–1694
16. Ferrer G, Garrell A, Villamizar M, Huerta I, Sanfeliu A (2013) Robot interactive learning through human assistance. In: Multimodal interaction in image and video applications. Springer, pp 185–203
17. Garrell A, Sanfeliu A (2012) Cooperative social robots to accompany groups of people. Int J Robot Res 31(13):1675–1701
18. Garrell A, Sanfeliu A, Moreno-Noguer F (2009) Discrete time motion model for guiding people in urban areas using multiple robots. In Proceedings of the IEEE/RSJ international conference on intelligent robots and systems, pp 486–491
19. Gill A (1962) Introduction to the theory of finite-state machines. McGraw-Hill, New York
20. Gockley R, Forlizzi J, Simmons R (2006) Interactions with a moody robot. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction. ACM, pp 186–193
21. Gockley R, Simmons R, Forlizzi J (2006) Modeling affect in socially interactive robots. In: The 15th IEEE international symposium on robot and human interactive communication. IEEE, pp 558–563
22. Goffman E (1966) Behavior in public places: notes on the social organization of gatherings. Free Press, New York
23. Haasch A, Hohenner S, Hüwel S, Kleinehagenbrock M, Lang S, Toptsis I, Fink G, Fritsch J, Wrede B, Sagerer G (2004) Biron–the bielefeld robot companion. In: Proceedings of the international workshop on advances in service robotics. Fraunhofer IRB Verlag, Stuttgart, Germany,pp 27–32
24. Hall ET (1969) The hidden dimension. Anchor Books, New York
25. Haralock RM, Shapiro LG (1991) Computer and robot vision. Addison-Wesley Longman Publishing Co., Inc., New York
26. Hayashi K, Sakamoto D, Kanda T, Shiomi M, Koizumi S, Ishiguro H, Ogasawara T, Hagita N (2007) Humanoid robots as a passive-social medium: a field experiment at a train station. In: Proceedings of the ACM/IEEE international conference on Human-robot interaction, pp 137–144
27. Helbing D, Molnar P (1995) Social force model for pedestrian dynamics. Phys Rev E 51(5):4282–4286
28. Hirth J (2012) Towards socially interactive robots-designing an emotion-based control architecture. Dr. Hut
29. Hirth J, Schmitz N, Berns K (2011) Towards social robots: designing an emotion-based architecture. Int J Soc Robot 3(3):273–290
30. Höök K (2009) Affective loop experiences: designing for interactional embodiment. Philos Trans R Soc London B Biol Sci 364(1535):3585–3595
31. Kanda T, Shiomi M, Miyashita Z, Ishiguro H, Hagita N (2009) An affective guide robot in a shopping mall. In: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, pp 173–180
32. Kędzierski J, Muszyński R, Zoll C, Oleksy A, Frontkiewicz M (2013) Emys-emotive head of a social robot. Int J Soc Robot 5(2):237–249
33. Khatib O (1985) Real-time obstacle avoidance for manipulators and mobile robots. Proceedings of the IEEE international conference on robotics and automation, vol 2, pp 500–505
34. Kirby R (2010) Social robot navigation. Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, USA
35. Koay K, Sisbot E, Syrdal D, Walters M, Dautenhahn K, Alami R (2007) Exploratory study of a robot approaching a person in the context of handing over an object. In: Proceedings of the AAAI-SS on multi-disciplinary collaboration for socially assistive robotics, pp 18–24
36. Kühnlenz B, Sosnowski S, Buß M, Wollherr D, Kühnlenz K, Buss M (2013) Increasing helpfulness towards a robot by emotional adaption to the user. Int J Soc Robot 5(4):457–476
37. Lee M, Makatchev M (2009) How do people talk with a robot?: An analysis of human-robot dialogues in the real world. In: Proceedings of the international conference extended abstracts on Human factors in computing systems. ACM, pp 3769–3774
38. Leite I, Martinho C, Pereira A, Paiva A (2009) As time goes by: long-term evaluation of social presence in robotic companions. In: RO-MAN 2009-The 18th IEEE international symposium on robot and human interactive communication, pp 669–674
39. Ljungblad S, Kotrbova J, Jacobsson M, Cramer H, Niechwiadowicz K (2012) Hospital robot at work: something alien or an intelligent colleague? In: Proceedings of the ACM 2012 conference on computer supported cooperative Work. ACM, pp 177–186

40. Marsella SC, Gratch J (2009) EMA: a process model of appraisal dynamics. Cogn Syst Res 10(1):70–90
41. Michalowski M, Sabanovic S, Simmons R (2006) A spatial model of engagement for a social robot. In: Proceedings of the 9th IEEE international workshop on advanced motion control, pp 762–767
42. Morales Saiki L, Satake S, Huq R, Glass D, Kanda T, Hagita N (2012) How do people walk side-by-side?: using a computational model of human behavior for a social robot. In: Proceedings of the 7th ACM/IEEE international conference on human-robot interaction, pp 301–308
43. Mutlu B, Shiwa T, Kanda T, Ishiguro H, Hagita N (2009) Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, pp 61–68
44. Quigley M, Conley K, Gerkey B, Faust J, Foote T, Leibs J, Ng A (2009) Ros: an open-source robot operating system. In: ICRA workshop on open source software, vol 3, no 3.2, p 5
45. Rich C, Ponsler B, Holroyd A, Sidner CL (2010) Recognizing engagement in human-robot interaction. In: Human-robot interaction (HRI), 2010 5th ACM/IEEE international conference on. IEEE, pp 375–382
46. Roseman IJ (1996) Appraisal determinants of emotions: constructing a more accurate and comprehensive theory. Cogn Emot 10(3):241–278
47. Saldien J, Goris K, Vanderborght B, Vanderfaeillie J, Lefeber D (2010) Expressing emotions with the social robot probo. Int J Soc Robot 2(4):377–389
48. Salichs MA, Malfaz M (2006) Using emotions on autonomous agents. the role of happiness, sadness and fear. Integr Approaches Mach Conscious Part AISB 6:157–164
49. Satake S, Kanda T, Glas D, Imai M, Ishiguro H, Hagita N (2009) How to approach humans?-strategies for social robots to initiate interaction. In: Proceedings of the ACM/IEEE international conference on human-robot interaction, pp 109–116
50. Schlosberg H (1954) Three dimensions of emotion. Psychol Rev 61(2):81
51. Shi C, Shimada M, Kanda T, Ishiguro H, Hagita N (2011) Spatial formation model for initiating conversation. In: Proceedings of robotics: science and systems, Los Angeles
52. Shiomi M, Kanda T, Glas D, Satake S, Ishiguro H, Hagita N (2009) Field trial of networked social robots in a shopping mall. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems, pp 2846–2853
53. Sidner CL, Kidd CD, Lee C, Lesh N (2004) Where to look: a study of human-robot engagement. In: Proceedings of the 9th international conference on Intelligent user interfaces. ACM, pp 78–84
54. Sisbot EA, Marin-Urias LF, Broquere X, Sidobre D, Alami R (2010) Synthesizing robot motions adapted to human presence. Int J Soc Robot 2(3):329–343
55. Timofte R, Zimmermann K, Van Gool L (2014) Multi-view traffic sign detection, recognition, and 3d localisation. Mach Vis Appl 25(3):633–647
56. Villamizar M, Garrell A, Sanfeliu A, Moreno-Noguer F (2012) Online human-assisted learning using random ferns. In: Proceedings of the 21st international conference on pattern recognition, pp 2821–2824
57. Walters M, Dautenhahn K, Woods S, Koay K (2007) Robotic etiquette: results from user studies involving a fetch and carry task. In: Proceedings of the ACM/IEEE international conference on human-robot interaction, pp 317–324
58. Xu J, Broekens J, Hindriks K, Neerincx MA (2014) Robot mood is contagious: effects of robot body language in the imitation game. In: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, pp 973–980
59. Yohanan S, MacLean KE (2012) The role of affective touch in human-robot interaction: human intent and expectations in touching the haptic creature. Int J Soc Robot 4(2):163–180

A. Garrell received the BSEE degree of Mathematics from the Universitat de Barcelona (UB), Spain, in 2006. She also obtained her PhD degree in Authomatic, control and visioin from the Universitat Politècnica de Catalunya, Barcelona, Spain, in 2013, and obtained the second prize of the best PhD Award in robotics in Spain. She has participated on diverse national and European projects and has worked with other research groups including the Robotics Institute of Carnigie Mellon University (US) . Her research interests are focused on cooperative robotics, social robotics and human robot interaction.

M. Villamizar received the BSEE degree of mechatronics engineering from the Universidad de San Buenaventura (USB), Colombia, in 2004. He also obtained his phD in computer vision and robotics from Universitat Politecnica de Catalunya (UPC), Barcelona, in 2012. Michael has participated in diverse national and European projects and has worked with other research groups. His research interests are focused on computer vision and pattern recognition. More specifically, he is interested in object recognition for real-time robotics applications.

F. Moreno-Noguer received the MSc degrees in industrial engineering and electronics from the Technical University of Catalonia (UPC) and the Universitat de Barcelona in 2001 and 2002, respectively, and the PhD degree from UPC in 2005. From 2006 to 2008, he was a postdoctoral fellow at the computer vision departments of Columbia University and the École Polytechnique Fédérale de Lausanne. In 2009, he joined the Institut de Robòtica i Informàtica Industrial as an associate researcher of the Spanish Scientific Research Council. His research interests include retrieving rigid and nonrigid shape, motion, and camera pose from single images and video sequences. He received UP's Doctoral Dissertation Extraordinary Award for his work.

A. Sanfeliu received the BSEE and PhD degrees from the Universitat Politècnica de Catalunya (UPC), Spain, in 1978 and 1982 respectively. He joined the faculty of UPC in 1981 and is full professor of Computational Sciences and Artificial Intelligence. He is the coordinator of the research Artificial Vision and Intelligent System Group (VIS) and head of the research line Mobile Robotics and Intelligent System of the (IRI) Institut de Robòtica i Informàtica Industrial (UPC-CSIC). He was the former director of the Institut de Robòtica i Informàtica Industrial, UPC-CSIC, former director of the UPC's Automatic Control department and past president of AERFAI. He has worked on various theoretical aspects on pattern recognition, computer vision and robotics and on applications on vision defect detection, tracking, object recognition, robot vision and SLAM. He has several patents on quality control based on computer vision. He has authored books in pattern recognition and SLAM, and published more than 260 papers in international journals and conferences. He has lead and participated in 45 R&D projects, 16 of them funded by the European Commission, and he was the coordinator of the European project URUS (Ubiquitous Networking Robotics in Urban Areas). He is (or has been) member of editorial boards of top scientific journals in computer vision and pattern recognition and has been the General Co-Chairmen of the top international congresses of Computer Vision (ICCV'2011) and Pattern Recognition (ICPR'2000) and of a number of international and national congresses. He received the prize to the Technology given by the Generalitat de Catalonia and is Fellow of the International Association for Pattern Recognition. He is member of IEEE.