Ubiquitous Social Perception Abilities for Interaction Initiation in Human-Robot Interaction *

Amol Deshmukh
School of Mathematical and Computer Sciences
Heriot-Watt University, UK
a.deshmukh@hw.ac.uk

Ginevra Castellano
School of Electronic Engineering and Computer Science
Queen Mary University of London, UK
ginevra@dcs.qmul.ac.uk

Mei Yii Lim
School of Mathematical and Computer Sciences
Heriot-Watt University, UK
M.Lim@hw.ac.uk

Ruth Aylett
School of Mathematical and Computer Sciences
Heriot-Watt University, UK
ruth@macs.hw.ac.uk

Peter W. McOwan
School of Electronic Engineering and Computer Science
Queen Mary University of London, UK
pmco@dcs.qmul.ac.uk

ABSTRACT

Robots acting as assistants or companions in a social environment must be capable of sensing information about the location of the users and analysing and interpreting their social, affective signals in order to be able to plan and generate an appropriate response. Social perception abilities are thereby very important for the robot to evaluate whether it is appropriate or not to initiate an interaction with the user. In this paper we present the initial steps of the design of a ubiquitous social perception system for interaction initiation: users' social signals and expressive behaviour are analysed at different spatial locations and temporal instants. We propose an approach to evaluate whether it is appropriate for a robot to initiate an interaction with the user. We describe an autonomous algorithm to regulate the inter-entity distance between the robot and a person using visual face detection which can be used during interaction initiation and also discuss the role of memory abilities to remember what has happened throughout the interaction.

*This work is supported by European Community (EC) and is currently funded by the EU FP7 ICT-215554 project LIREC (LIving with Robots and IntEractive Companions). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

AFFINE'10, October 29, 2010, Florence, Italy.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; J.4 [**Computer Applications**]: Social and Behavioural Sciences

General Terms

Algorithms, Human Factors, Theory

Keywords

Interactive companions, human-robot interaction, socially intelligent behaviour, interaction initiation

1. INTRODUCTION

Building robots capable of interacting with humans as fluently as humans interact with each other has always been a major challenge in the research of robotic technology. Social perception abilities are necessary for a robot to interact with humans as a companion, a partner or an assistant, the lack of which often impedes human users from establishing a comfortable and productive interaction relationship. Robots acting as assistants or interaction partners in human social environments, such as robot companions, must be able to act in a socially competent way [1]. They must be capable of sensing information about the location of the users and analysing and interpreting their social, affective signals in order to be able to plan and generate an appropriate response [2]. From the perspective of automatic analysis of users' behaviour, most of the work has been done on analysis of face-to-face interaction cues [3] or monitoring of actions when users are far from the camera [4].

In this paper we present the initial steps of the design of a ubiquitous social perception system for interaction initiation: we believe that analysing the user's social signals when the user is at different spatial locations and distances from the robot can be useful for a robot to develop interaction initiation abilities. We investigate concepts related to interaction initiation and propose an approach to assess whether it is appropriate or not for a robot to start an interaction with the user, and to do so in a socially acceptable manner.

The paper is organised as follows. Section 2 gives a brief overview of the background of our work. Section 3 describes the scenario under investigation and the physical setup of the environment. Section 4 proposes an approach to assess the appropriateness of interaction initiation from the robot. In Section 5 we describe an autonomous mechanism to determine user proximity and robot approach using face detection. Section 6 discusses the role of memory abilities to allow the robot to memorise information about the user's location, signals, expressive behaviour, and in general all information about social perception that could be used for the purpose of evaluating whether to initiate an interaction. Finally, Section 7 summarises the proposed approach and discusses future work.

2. BACKGROUND

A socially intelligent companion must be able to assess the appropriateness of an interaction initiation condition with the user. In a mobile interaction scenario where robots and users are free to move in the environment, information about the position, movement and expressive behaviour of the users are of key importance for the robot to evaluate the user's level of engagement [5] and willingness to interact with the robot. How using this information to plan an appropriate interaction initiation is still a challenging issue, although some work was proposed in the literature. Michalowski et al. [6], for example, proposed an approach based on social space to categorise different stages of user engagement with a robot, such as *present*, *attending*, *engaged* and *interacting*. Peters proposed a perceptually-based theory of mind model for interaction initiation applied to virtual agents [7] and evaluated user perception of attention behaviours for interaction initiation in virtual environments [8].

After assessing the appropriateness for interaction initiation, the robot needs to approach the user or respond to the user's approach in a socially acceptable manner. Earlier studies have shown that successful human-robot interaction is impacted by comfortable approach distances between human and robot that respect the user's personal space [9]. Studies also indicate that the appearance of the robot influences the level of comfort in relation to approach distances [10]. Some researchers have assigned more precise numerical values to personal spaces in human to human interaction, e.g. Hall [11]. Hüttenrauch et al. [12] concluded that in human-robot interaction user trials most participants kept inter-personal distances from a PeopleBot^{TM} robot corresponding to Hall's personal spaces (0.45m to 1.2m). Other experiments and live human-robot interaction trials conducted [9], suggest the mean comfortable approach distances vary from 0.65m to 0.5m depending on appearance of the robot e.g. humanoid, mechanoid [10]. We anticipate that people will assume distances that correspond to social or personal zones (similar to distances people use having face-to-face conversation) while treating a robot as a *social being.*

3. SCENARIO

The work reported here is carried out as a part of the EU project LIREC (LIving with Robots and intEractive Companions, www.lirec.eu). The project aims to create interactive, emotionally intelligent companions which are capable of establishing long-term relationships with humans in social environments. The "Spirit of the Building" showcase at Heriot-Watt University, Edinburgh, aims to produce a social helper robot that can share a lab with human researchers and act as a "Team Buddy"- an assistant to facilitate long-term relationship with users. The "Team Buddy" that we call SARAH (Social Agent Robot to Aid Humans, see Figure 1) would act as a workplace buddy within a lab inhabited by a small group of people, performing tasks such as carrying the phone and printed material to the users, giving out reminders for important events such as meetings, paper deadlines, providing a lab tour to visitors, approaching and greeting the users, keeping track of who has entered/left the room, remembering people who are not there when they have gone, whilst maintaining a collective memory about user preferences such as lunch breaks, entry/exit time.

Figure 1: "Team Buddy" SARAH, Height 1.3m

This clearly requires social location, in which the robot is able to move within a human environment and interact with users who are busy with their normal work. In a lab like environment it is quite challenging to keep track of the user's location and perceive their interaction intentions. In order to engage in an interaction, the robot must have perception abilities such as detecting whether a user has entered the room or is sitting at their desk and user proximity control while approaching them. To perceive the user's interaction intentions from different spatial locations, it is necessary for the robot to be endowed with ubiquitous perception abilities.

3.1 Physical Setup

The robot operates in a lab with a group of 4-5 people who work there. The users in the lab work on their assigned desks, each of which has a desktop PC with a web camera facing the user. There is also a camera facing the room orthogonally to the door which can detect the user's entry/exits in the lab and a camera on the robot which provides information about lab surrounding when the robot moves. All the cameras in the lab can be interfaced together on a common system which can communicate with the robot. Figure 2 and Figure 3 provide an overview of the lab environment indicating the placement of cameras, desks, users, door and the mobile robot.

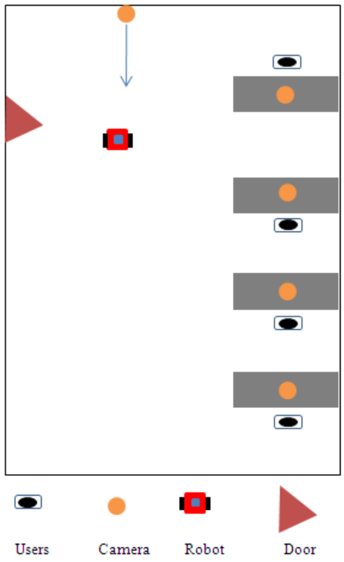

Figure 2: Lab environment map, Floor Area: 75 Sq. meters

Figure 3: Lab environment, users working in lab with web cam PCs on desks

4. ASSESSING APPROPRIATENESS FOR INTERACTION INIATIATION

Planning interaction initiation with a user requires the robot companion to have the ability to evaluate when this is appropriate. A socially intelligent companion acting as a personal assistant, in fact, must (1) be capable of assessing whether, based on the current situation, it can initiate an interaction with the user even when it is not requested (e.g., approaching the user for non urgent tasks such as greeting) and (2) be ready to start an interaction if the user shows willingness to engage with the robot.

In Sections 4.2 and 4.3 we analyse these two interaction conditions and propose an approach to evaluate the appropriateness of interaction initiation from the robot. Our approach is based on the analysis of the user's actions and social signals at different spatial locations and temporal instants. This implies analysing different types of signals and is achieved through the use of different types of sensors, including cameras at different locations and distances from the robot. In the following Section we list the social perception abilities that our companion is endowed with.

4.1 Social Perception Abilities

In the "Team Buddy" scenario, our robot companion is endowed with a set of vision abilities for social perception, which represent the starting point for the evaluation of the appropriateness of interaction initiation.

4.1.1 Face detection and tracking

Face detection is performed using the OpenCV method based on a Haar feature-based cascade classifier [13], while tracking is done using either the Camshift algorithm or an approach based on particle filters.

4.1.2 Motion direction detection

Our robot is able to detect the presence of motion within a specific spatial area and compute the overall motion direction of a person walking in an orthogonal direction to a camera. This can be achieved by the camera placed on the robot or at key locations in the lab (see Figure 2). The motion direction detection ability is based on the motion templates routines provided by OpenCV [13]. Using motion templates requires the automatic extraction of the user's silhouette. Given the user's silhouette, a motion template is built using a motion history image (mhi) and an indication of the overall motion can be derived by computing the gradient of the mhi (see Figure 4). Motion direction detection may also be used to perform some form of simple gesture or action recognition.

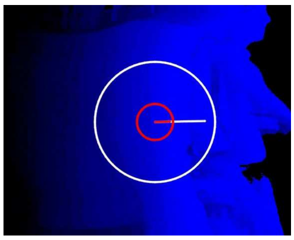

Figure 4: A measure of the overall motion direction detection of the user computed using motion templates

4.1.3 Affect sensitivity

Affect sensitivity refers to the ability, for a robot, to analyse and interpret expressive, affective behaviours and states displayed by the user. Currently a vision-based smile detector prototype based on Support Vector Machines is available to the robot to detect the probability that the user is smiling. This may give the robot some information about the user's mood. The user's level of activity, simple gestures, and physiological measurements are under investigation for possible integration into an affect recognition framework.

4.2 Evaluating interaction initiation when the user does not request it

Imagine a situation where the robot needs to engage with the user for non urgent tasks, e.g., greeting. A socially intelligent companion acting as a personal assistant would not start engaging with them if they are not in a good mood or are busy accomplishing some other tasks. Therefore, initiating interaction with a user when the latter does not take the initiative to interact with the robot requires evaluation of the current situation. Before reaching this level of analysis, though, the robot must know whether (1) a user is present in the room and (2) a user is sitting at a desk.

The motion direction detection ability is useful for the very first evaluation, that is, to detect whether a user has entered the room. A camera placed in an orthogonal direction to the door captures entries and exits of users from the room and a computation of the overall motion direction of the user allows the robot to infer whether the user has entered or left the room. The second step involves detecting whether a user is sitting at their desk. This is performed via face and motion detection, using a web camera placed on the desktop PC. Once the robot is aware of a user sitting at a desk, it can perform an overall analysis of the situation by activating the extraction of some simple indicators. An expression detection module performing smile detection using images captured with the web camera can help assess the mood of the user. Future work will consider the role of the task the user is involved in, as the non-verbal behaviors alone may not be representative if analysed out of context.

4.3 Inferring the user's willingness to interact with the robot

This scenario involves a user who is not necessarily sitting at a desk, but is walking in the office. The focus here is on the evaluation of the user's level of willingness to start an interaction with the robot, which may also give some information about the user's level of engagement with it [5]. As in the previous scenario, the very first thing that the robot needs to know is whether a user is present in the room. Depending on the robot's location, one or two cameras could be used to capture the user's motion and provide the input to the motion direction detection module. If motion is detected, the overall motion direction of the user can be used to evaluate whether the latter is walking towards the robot or entering a specific area under which the robot can respond. Motion direction computation at different spatial locations and temporal instants can then be used to help detect whether the user is willing to start an interaction with the robot. The user's face is an important cue in determining their interaction intentions. The human face can provide supplementary information to determine the user's interaction intentions, since a user is likely to face towards the robot if they are willing to engage in an interaction. Therefore face detection and tracking as mentioned in section 4.1.1 can be useful to determine if the user wants to initiate an interaction and is equally challenging to achieve using other sensors or methods.

5. USER PROXEMICS USING FACE DETECTION

In our scenario the "Team Buddy" interacts with members of the group through a robot (see Figure 1), a Pioneer^{TM} robot with enhanced superstructure and a built-in camera on the laptop PC placed on the superstructure. The robot's appearance can be perceived as mechanoid with a graphical face displayed on the laptop screen. In order to act as a companion robot to a small group of people, it is necessary for the robot to be able to approach users so that it can initiate interaction with them and maintain a comfortable distance from them. In our approach we propose the use of face detection for sensing user proximity taking into account user proxemics studies discussed in the background section.

5.1 Face Distance Calculation

We used the OpenCV library for detecting face in the environment [13]. We used the face detection further to estimate the position and distance of human face in relation to the camera placed on the robot (using the bounding box of a human face). Using a number of different human subjects facing towards the camera positioned at specific distances in a straight line, we recorded the difference between the image area and face bounding box detected by face detection algorithm. It is quite understandable that, the closer the face to the camera, the larger will be the face bounding box area and vice-versa. We used a camera with a resolution of 640 $\times$ 480 giving us a constant image area of 307200 (pixels). As the total image area is always constant, it is trivial to record the difference between image area and face bounding box area in pixels.

$$AreaDifference = ImageArea - FaceBoundingBox$$

We performed some experiments independently with 5 human subjects (4 Male, 1 Female), each were positioned at specified distance (0.3m, 0.5m, 1.0m, 1.5m, 2.0m) from the camera and recorded the area difference. Each reading was averaged over 20 samples taken at each position. The graph in Figure 5 illustrates the area difference readings (Y-axis, pixels) for each distance position (X-axis, meters) for 5 subjects P1 to P5 and the average area difference for the 5 subjects. We can observe from the curves that the area difference values are quite similar for all 5 human subjects (P1-P5) for all distance positions.

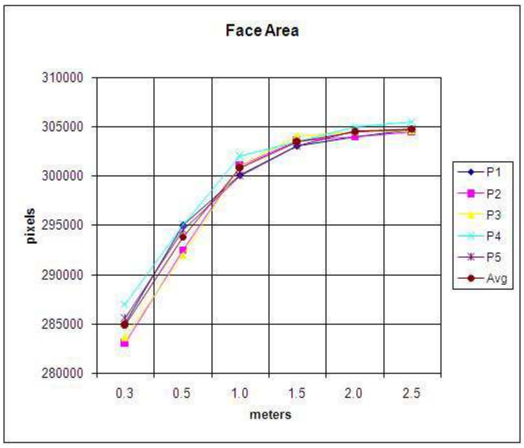

Figure 5: Face Area Graph

We thereby suppose that image area difference can be used effectively for user distance estimation using face area bounding box. Please note that the values may vary with different camera resolutions, but we anticipate that the nature of the curves will be similar. Another important factor to note here concerns face detections. We performed the experiments in well illuminated light conditions and recorded the average over 20 samples to test and improve the accuracy of the recordings. We also performed physical observations while recording the samples, to check if a face was present when we recorded the sample. Out of 20 $\times$ 5 (distance positions) $\times$ 5 (subjects) = 500 (total recordings), Total false detections = 56 (when no face was present, but detected), No Detections = 41 (when face was present, but not detected), overall face detection accuracy 80.6%. The results from the studies made earlier [11] and our average face area difference values are combined into Table 1.

Table 1: Face distance calculation and personal spaces

Face Distance	Spaces [11]	Area Difference
3m - 2m	Social Zone	304410
2m - 1m	Social Zone	303400
1m - 0.5m	Personal Zone	294000
0.5m - 0.2m	Intimate Zone	284620

5.1.1 *Face Position Estimation*

To further estimate the position of the detected face to the left or right from the camera's focal point, we calculated the difference in number of pixels from the face mid-point to center X-axis in the image. This pixel difference is further used to turn the robot (left/right) towards the detected face while approaching the user (see Figure 6).

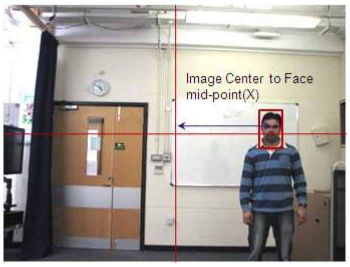

Figure 6: Face Position Estimation

5.1.2 *Automatic Distance Adjustment*

To enhance our approach we developed a mechanism for the robot to autonomously adjust the distance from the user. The robot moves backward (0.3m - 0.5m, 284620 - 294000 Area Difference in pixels) which corresponds to human intimate zone (Table 1) if it gets too close or the user chooses to approach it. When the robot detects a person stepping back, it approaches them to maintain its threshold (0.5m, 294000 Area Difference in pixels) which corresponds to the personal zone (Table 1). This mechanism provides an added advantage to user proxemic distance control when the subject is moving. Figure 7 illustrates the automatic distance adjustment. The face distance estimation algorithm was used to autonomously guide the robot towards the user and stop at a desired distance of 0.5m (Personal Zone) from the user. When multiple users are present in the environment, the robot approaches the closest person facing the robot.

5.1.3 *Preliminary studies*

We conducted some preliminary studies to test the effectiveness of our algorithm and to find out how people felt about the robot approaching them. The trials were carried out independently with 5 human subjects (4 male, 1 female). We conducted 3 trials per subject (total 15 trials), placed at 3 different positions in the lab facing the robot within range of 2 meters, before starting the program. At the end of 3 trials, each subject was given a short questionnaire and was asked to give a score between 0 to 5 (5 being the best).

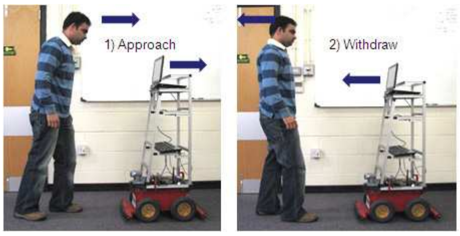

Figure 7: Automatic distance adjustment 1) Human approaches: The robot moves backward 2) Human withdraws: The robot approaches

1. Did you feel the robot found you and was actually approaching you? **Average score 4.4 after 15 trials**

2. How did you feel about the distance between you and the robot? Please rate your acceptance score

	Close	About Right	Far
No. of Votes	7	5	3
Average Score	3.4	3	3.6

3. Did you feel convenient that the robot moved backward when you tried to approach it? (a) It was acceptable: **4** Votes (b) It was discomforting: **1** Vote (c) can't say: **0**

The feedbacks to the questionnaire and average stopping distance measured after 15 trials (0.51m) indicate that people found the robot's approach reasonably acceptable (cumulative average acceptance score: **3.33**) and is in good agreement with previous experiments on robot to human approach distances [9, 10, 11, 12]. We would further enhance our approach by combining information from range of sensors such as passive infra-red (IR) sensor, laser range finder and sonar range in addition to the camera to estimate more accurately the proximity of human present. The autonomous user proxemics sensing and control can be used in interaction initiation for both cases mentioned in section 4.2 and 4.3.

6. MEMORY

Since our aim is to create a socially aware robot, it is important for the robot to be able to remember basic users' information (e.g. name, age, etc.) and preferences (e.g. when to an initiate interaction, entry/exit time), hence adapting its interaction accordingly [14]. Briefly, the memory of our robot is divided into 1) a higher-level symbolic memory consisting of a long-term memory (LTM), which stores the users' profiles and past episodic events, and a short term memory (STM), which stores information about current goals processing; 2) a lower-level sub-symbolic working memory (WM), which temporary holds the information perceived and is being processed by its different abilities [15].

In the following we explain how memory and social perception abilities are linked. When a face or motion are detected using the camera on the robot or the camera facing the room, the image is temporary captured onto the WM. This image is further processed by different abilities for different purposes. For example, the motion direction detection module will try to infer the overall motion direction of the user, and the user proxemics module will use this image to obtain the face position and distance. The resulting information can then be used by the robot to decide whether to initiate an interaction, and whether to approach or withdraw from the user.

Similarly, when a face is detected using the camera on a user's desk, information such as the captured image, the camera ID and the time is stored onto the WM. The image may be processed by other modules, such as the smile detection module. The camera ID can be used to identify a particular user by retrieving user-camera entry in its LTM. Once the user is identified, the user's preferences may be retrieved. This information can aid the robot's actions such as in deciding when to initiate an interaction (e.g. user A does not like to be disturbed when he is busy working). Additionally, information processed by the lower-level abilities is updated to the higher-level memory whenever necessary. For example, the detection event will be recorded onto the robot's episodic LTM that may be used in future to predict the user's availability. Thus, with prolonged interaction time, the robot will learn and get to know users better. This adaptive memory framework will enable the robot to remember the interactions over a sequence of events for a long period of time.

7. CONCLUSION AND FUTURE WORK

Robots acting as companions in a social environment must be capable of sensing information about the location and proximity of people in order to be able to plan and generate an appropriate response. The robot must be endowed with social perception abilities to initiate an interaction with the user in a socially acceptable manner. In this paper, we presented the initial steps of the design of a ubiquitous social perception system to sense interaction cues by analysing the user's signals at different spatial locations from the robot. We discussed the key steps to initiate an interaction with the user, such as detecting the presence of a user and their location (door entry/exit, face detection), assessing the appropriateness for interaction initiation in different circumstances and approaching the user in an acceptable manner. The role of memory and its link with social perception abilities was also been discussed. Future work will include the integration of the proposed social perception modules in a single computational framework in which they are interfaced with memory and decision-making. The definition of such a framework will benefit from further work on the design of a scenario-dependent affect recognition system and, in general, from the integration of contextual information in the social perception modules.

8. REFERENCES

[1] K. Dautenhahn. Socially intelligent robots: Dimensions of human- robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007.

[2] G. Castellano, Leite I., Pereira A., C. Martinho, A Paiva, and McOwan P. W. Affect recognition for interactive companions: Challenges and design in real world scenarios. *Journal on Multimodal User Interfaces, Springer*, 3(1):89–98, 2010.

[3] Z. Zeng, Roisman Pantic, M., G. I., and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.

[4] T. Xiang and S. Gong. Video behaviour profiling for anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):893–908, May 2008.

[5] C. Peters., G. Castellano, and S. de Freitas. An exploration of user engagement in hci. In *Proceedings of the International Workshop on Affect-Aware Virtual Agents and Social Robots (AFFINE'2009)*. Cambridge, MA. ACM New York, NY.

[6] M. Michalowski, S. Sabanovic, and R. Simmons. A spatial model of engagement for a social robot. In *Proceedings of the 9th International Workshop on Advanced Motion Control (AMC 2006)*, pages 762–767. IEEE, March.

[7] C. Peters. A perceptually-based theory of mind model for agent interaction initiation. *International Journal of Humanoid Robotics (IJHR), special issue Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids*, pages 321–340, 2006.

[8] C. Peters. Evaluating perception of interaction initiation in virtual environments using humanoid agents. In *Proceedings of the 17th European Conference on Artificial Intelligence*, pages 46–50. Riva Del Garda, Italy, August 2006.

[9] M. L. Walters, K. L. Koay, S. N. Woods, D. S. Syrdal, and K. Dautenhahn. Robot to human approaches: Comfortable distances and preferences. In *AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*. Stanford University, USA, 2007.

[10] M. L. Walters, K. Dautenhahn, S. N. Woods, K. L. Koay, R. te Boekhorst, and D. Lee. Exploratory studies on social spaces between humans and a mechanical-looking robot. *Journal of Connection Science, Special Issue on Android Science*, 18:429–442, 2006.

[11] Hall E.T. *The Hidden Dimension*. Doubleday, 1966.

[12] H. Hüttenrauch, K. Severinson Eklundh, A. Green, and E. A. Topp. Investigating spatial relationships in human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, pages 5052–5059. Bejing, China.

[13] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, 2008.

[14] M.Y. Lim, R. Aylett, W.C. Ho, Vargas P., and S Enz. A socially-aware memory for companion agents. In *9th International Conference on Intelligent Virtual Agents*, pages 51–62. Amsterdam, 2009.

[15] W.C. Ho, M.Y. Lim, P.A. Vargas, Dautenhahn K. Enz, S., and R. Aylett. An initial memory model for virtual and robot companions supporting migration and long-term interaction. In *IEEE International Symposium on Robot and Human Interactive Communication, ROMAN 2009*. Tokyo.