

Sensor Fusion for Social Robotics

B.R. Duffy*, C. Garcia[†], C.F.B. Rooney*, G.M.P. O'Hare*

* PRISM Laboratory, Dept. of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland

[†]ICS - Foundation for Research and Technology-Hellas (FORTH), Heraklion, Crete, Greece

Email: Brian.Duffy@ucd.ie, cgarcia@csi.forth.gr, Colm.Rooney@ucd.ie, Gregory.OHare@ucd.ie

ABSTRACT

This paper advocates the application of sensor fusion for the visualisation of social robotic behaviour. Experiments with the Virtual Reality Workbench integrate the key elements of Virtual Reality and robotics in a coherent and systematic manner. The deliberative focusing of attention and sensor fusion between vision systems and sonar sensors is implemented on autonomous mobile robots functioning in standard office environments.

Keywords: *mobile robots, vision, sensor fusion, multi-agent systems, social robotics.*

INTRODUCTION

Currently, rather than building complete world models, mobile roboticists tend towards reactive, embodied control strategies which focus more on sensing and acting [1] [2]. Motivation for this can be found in biological systems. For example, [3] introduces the concept of the *Umwelt* which describes the set of environmental triggers to which an animal responds. Likewise, robots may be configured to respond instinctively to a certain set of features. This forms the basis for Brooks' [1] subsumption architecture.

Such reactive behaviour not only applies to simple sensory modalities, e.g. tactile, infra-red, but also to rich media such as vision. *Animate vision* [4] is a body of research that promotes the use of fast, reactive routines, which intertwine sensing and action rather than converting visual signals into complex enduring world models.

This is exemplified in nature by insects, which despite having relatively simple brains can use visual clues such as landmarks to navigate. Such *snapshot* models of vision have been applied in the field of robotics [5].

The approach we advocate in this paper draws upon the above ideas. Modelling of the environment through the sonars alone can be inefficient. There are features which inhibit its use e.g. noise, reflections. Our other

alternative, vision, demands a high computational overhead to build up a detailed world model. However, if we abandon complex models and rather look for salient features, the advantages of both sonar and vision may be utilised.

The problems associated with obtaining robot position accurately without the use of tailored environments (by providing landmarks etc.) are rife [5] [6] [7] [8] [9]. Robot research has extensively documented problems associated with the cumulative positional error in the robots' own odometric sensor readings, rendering such sensors inadequate as exclusive sources of long-term positional information. Given the robots' internal position information the problem is to reduce the incremental error over time between this and the robots actual position in its environment.

Sonar provides a quick, simple way to determine approximate distances from obstacles. Experimentation shows that certain features possess characteristic *sonar footprints* e.g., open doorways, corners, walls. Similarly, such features may be recognised in a visual image. A robot traversing an environment searching for such features could be alerted by the presence of such a *sonar footprint* to the possible presence of a feature.

Effective identification of environmental cues is pivotal in circumventing cumulative positional error. Furthermore, it enables visualisation of virtual robots within a *Virtual Robotic Workbench* [10].

In the remainder of this paper we firstly introduce the robot, and then the Social Robot Architecture by which it is controlled. Subsequently, we describe the two primary strains of sensory stimulus those of sonar signatures and vision. Section 6 fuses these sources while section 7 describes their usage within the *Virtual Robotic Workbench*.

THE ROBOT

The Nomad Scout II robot is equipped with a sonar ring of 16 sensors, a Sony digital camera (Sony DCR-PC1 Digital Camcorder) using a Matrox Meteor PCI

framegrabber card, the Scout's odometry system and a bumper ring (fig. 1).



Figure 1: Nomad Scout II Robot with Vision System

Together these sensory systems can provide a large amount of sensory feedback. It would also be possible to add additional sensors to facilitate the development of a more complete model of the environment. However, high sensory feedback can lead to a high computational load, both in terms of converting sensory data into useful information and reasoning about this information.

Sections 4 and 5 respectively describe the sonar signatures and vision systems of the robot.

SOCIAL ROBOT ARCHITECTURE

Such robots are controlled by way of a *Social Robot Architecture* [11] [10]. This architecture aims at achieving team building and collaborative behaviour through the judicious synthesis of the reactive model with that of the deliberative model. The architecture (fig. 2) is comprised of four discrete layers: physical, reactive, deliberative developed using Agent Factory [12], and social.

Physical: Robots in terms of this research may take the form of either that of a physical entity, specifically the Nomad Scout II or a simulated entity (Nomadic Technologies XRDev robot).

Reactive: A series of fundamental reflex behaviours are

implemented at this level. The sensory information is processed resulting in clear *agent_events* and communicated to the deliberative level. *Agent_commands* are received from the deliberative layer.

Deliberative: This comprises of a Belief Desire Intention (BDI) [13] architecture developed through Agent Factory [12]. The perception process deals with converting *agent_events* into beliefs and adding them to the belief set providing the agent with an up to date model of its current perceived situation and results in the agents' commitments being updated accordingly. Pre-existing commitments are analysed and those pertaining to the current time frame honoured resulting in either a communicative act being sent to the social level or a physical act being passed to the actuators via the reactive level.

Social: Our agents interact via an Agent Communication Language (ACL), entitled Teanga.

More detailed information on *The Social Robot Architecture* and its applications are given in [14] [11] [10].

SONAR SIGNATURES

Figure 3 is a graph of the readings from three of the robot's sonar as it passes an open door. This sonar footprint is characteristic of an open door. First sonar 14¹ jumps as the free-space in the new room is detected. Next sonar 13 makes a similar jump, and finally sonar 12, the side sonar, jumps up too, indicating that the door is directly to the side of the robot. While the value from sonar 12 jumps, the values from sonar 13 and sonar 14 fall off again. This *sonar signature* reappears each time a door is passed. It differs from a corner sonar typical corner signature in that sonar 13 and sonar 14 decrease again in the door case. In addition to recognising the door through its sonar signature, we have successfully

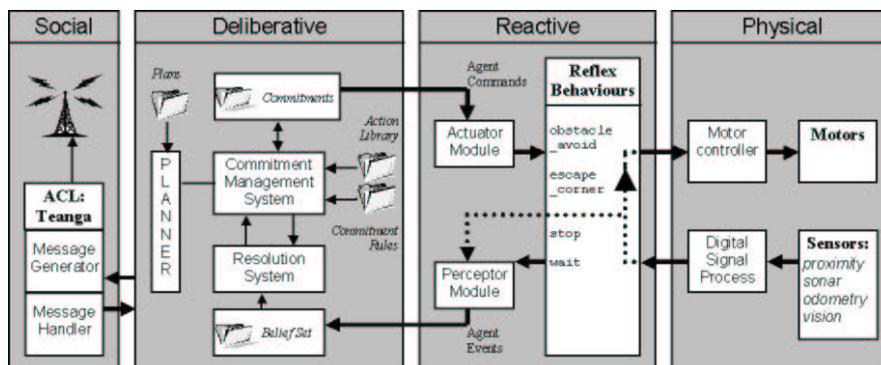


Figure 2. The Social Robot architecture: The Robot Agent

employed visual analysis for door recognition to provide corroborative evidence of the presence of a door. A library of sonar signatures can therefore be utilised for feature recognition.

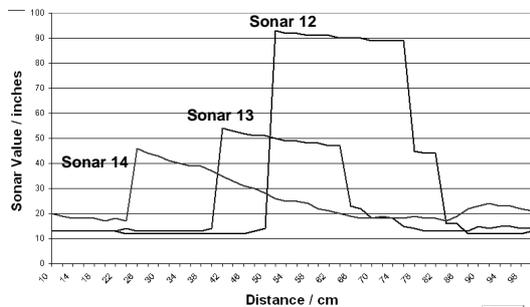


Figure 3: Sonar signature for a door

The footprint then triggers the vision system. A snapshot is taken and analysed in order to corroborate our initial feature hypothesis, within a given confidence interval. Such information would then be passed upward to the deliberative level. This *focusing of attention* combines the quick, real-time advantages of the sonar with the rich sensory feedback of vision.

THE VISION SYSTEM

Given the field of view of the mounted camera, a single feature is detected at a time. Let us consider detection of a doorframe feature. To do so, a scheme based on edge-detection, line fitting and shape analysis is performed. A doorframe is considered to be a “L” or “V”-shaped object in the image, as it may be seen in Fig.4a.

First, the Canny-Deriche edge detector [15] [16] is applied to the frame, in order to detect the outlines of the door. The Canny operator has been optimally designed by solving an optimisation problem with constraints related to sensitivity, localization and local unicity [15]. The method can be seen as a smoothing filtering performed with a linear combination of exponential functions, followed by derivative operations. Deriche developed a recursive implementation of this operator [16]. The high performance of the recursive version makes it more interesting for general applications.

This edge detection scheme makes explicitly use of a parameter α for controlling the granularity at which edges have to be detected. The parameter α directly defines the size of exponential filters and is chosen according to the expected size of grey level transition region, as well as to the noise level in the image.

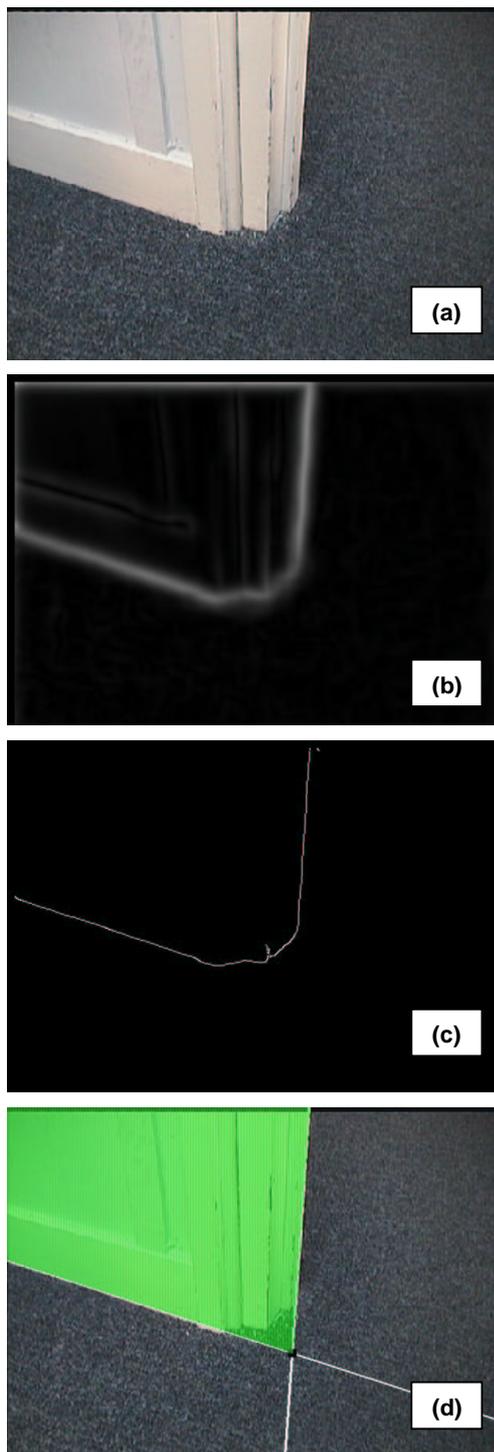


Figure 4: (a) Original image, (b) zero-crossing, (c) Selected zero-crossing and (d) area classification

The filter produces estimates of the horizontal and vertical components of the brightness gradient at every pixel. The intensity gradient at each pixel location can

then be estimated by taking the linear combination of these two directional values, providing estimated magnitude and direction. Fig.4a presents the original image and Fig.4b shows the magnitude image after filter application.

For all pixels, “non-maximum suppression” based on the gradient magnitude is performed by exploring in the direction of gradient. A pixel, called zero-crossing, is retained as a possible edge point only if it has a larger gradient than its neighbours located in the direction closest to that of the gradient, and than its neighbours located in the opposite direction. The remaining local maxima belong to one-pixel-wide edge segments. Thresholding based on gradient magnitude is then performed on these points. Any such point with magnitude above a threshold T_{high} is kept, as shown in Fig.4c.

One problem with all edge detection algorithms is that the edge points belonging to the same physical object borders are not all linked together, because of occlusion and lighting variations. There are several techniques to group points into lines in order to approximate the object borders, based on Least Squares criteria, which are very sensitive to noise or outlying edge points.

The most robust approach to link points into lines is to perform a Hough transform [17]. The Hough transform is a standard tool in image analysis that allows recognition of global patterns in an image space by recognition of local patterns (ideally a point) in a transformed parameter space. It is particularly useful when the patterns one is looking for are sparsely digitised, have “holes” and/or the pictures are noisy. The Hough Transform, doesn’t require connected or even nearby edge points. The idea is to accumulate evidence for different line equation parameters in “parameter space” for each edge point in “image space”.

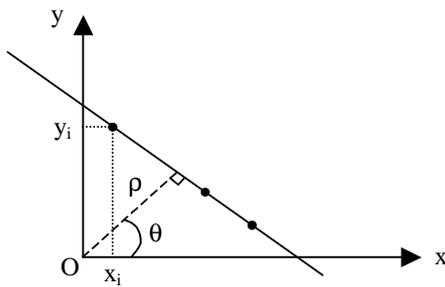


Figure 5: A line in parameter space

We assume lines to be parameterised in the form $\rho = x \cos(\theta) + y \sin(\theta)$ (1), where ρ is the perpendicular distance from the origin and θ is the angle with the normal, as it is shown in Fig.5. With this choice

of parameterisation, it is possible to describe all lines, even vertical ones.

Collinear points (x_i, y_i) , with $i=1..N$, are transformed into N sinusoidal curves of equation (1) in the (ρ, θ) plane, which intersect in the point (ρ_m, θ_m) .

The Hough transform works by letting each feature point vote in space for each possible line passing through it. These votes are totalled in an accumulator $Accu(\rho_m, \theta_m)$. For a certain range of quantised values of parameters ρ and θ , each (x_i, y_i) is mapped into the (ρ, θ) space and the points that map into the locations (ρ_m, θ_m) are accumulated in $Accu(\rho_m, \theta_m)$, i.e. $Accu(\rho_m, \theta_m) = Accu(\rho_m, \theta_m) + 1$.

The parameter space (ρ, θ) has to be quantised carefully. When the bins of the (ρ, θ) space are chosen too fine, each intersection of two sinusoidal curves can be in a different bin. When the quantisation is not fine enough, on the other hand, nearly parallel lines, which are close together, will lie in the same bin. For frame of size M lines by N columns, ρ varies from 0 to $\sqrt{M^2 + N^2}$ and θ from $-\pi/2$ to π . We chose to quantise the ρ and θ dimensions in n_ρ and n_θ cells.

Local maxima of the pixel intensity $Accu(\rho_m, \theta_m)$ identify straight-line segments in the original image space. The Hough domain has to be searched for maxima in an iterative way. The highest peak, which value is greater than the parameter min_{vote} , in the Hough domain is found and the corresponding line in the image space is identified. Then, the edge points corresponding to this line are discarded and the process is repeated.

Several tuning parameters are used in our implementation. The optimal values we found are : $\alpha = 0.2$, $T_{high}=80$, $n_\rho = 800$, $n_\theta = 500$ and $min_{vote}=50$, for images of size $M=480$ and $N=640$.

In Fig.4d, we show the lines that have been detected. Each pair of non-parallel detected lines give rise to an intersection point. Each intersection point and the associated two lines define four areas in the frame as shown in Fig.4d, which are to be analysed in order to detect the presence of a door. If present in the image, a doorframe has to be contained in upper areas 1 or 2, the others containing parts of wall and floor. The area containing the door frame is supposed to be lighter and homogeneous whereas the other areas are supposed to be darker and textured. Each area is classified according to the texture it contains.

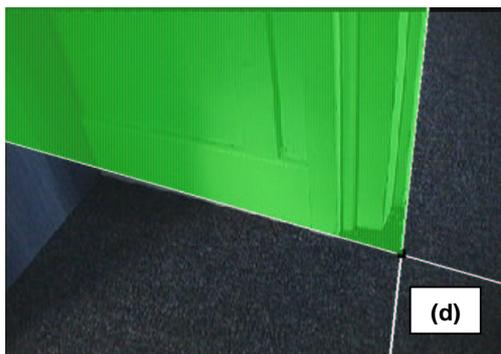
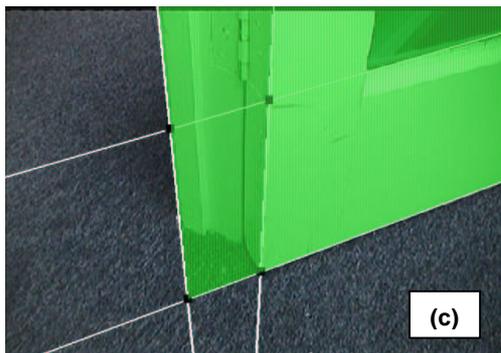
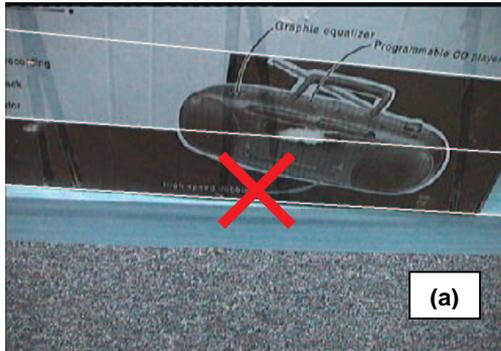


Figure 6: Sample results: (a),(b),(c) and (d)

A fast measure for classifying this texture is based on the variance of the intensity of the pixels contained in the area. We measure the density D of pixels whose intensity is greater than 0.7 times the average intensity

value among the areas. A doorframe is considered detected in one area if this area has a density D greater than 80% with an average intensity value greater than 150 and the three other areas have a density smaller than 50%. In the example of Fig.4, area 1 has been classified as containing a doorframe.

In Fig.6. we present some results of our scheme for door detection. Fig.6.a and Fig.6.b present some images, which do not contain doorframes (a superimposed cross indicates no door has been detected). Fig.6.c and Fig.6.d show the detected doorframes in two different configurations.

SENSOR FUSION

Sensor fusion in the mapping of unknown worlds is by no means new [18]. In terms of this research, the robots utilise sonar for a rough and rapid perception of the environment with vision used to reinforce the robots perception of particular features via a deliberative focusing of attention. This *attention focusing* restricts the computational loading imposed on the robot through the use of a vision system.

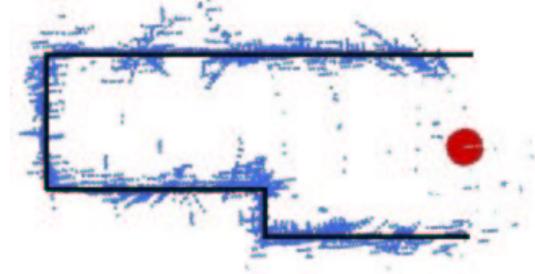


Figure 7: Wall and corner extraction from sonar information

Through the use of primary location cues (i.e. doors, walls, primary corners, desks etc.) and their corresponding sonar signatures (fig. 7), the robot activates its vision system to obtain snapshots providing sensor fusion between sonar and vision systems.

In the context of the door-location experiment the 3 front-right sonars are polled as the robot traverses its environment. When the values match the footprint an agent_event of the form possible_door is generated. Note that this footprint is not a static set of values for one time-frame. It reflects the changes in input that the robot receives as it approaches a door over time, in effect reflecting how the features *behaves* (in the robot's perspective) as the robot nears it.

This agent_event is then propagated to the deliberative level where it is adopted as a belief. This belief may then result in a commitment rule being fired and a

commitment being adopted. In this case the robot commits to activate the vision system. A couple of snapshots, from differing angles, are taken and analysed as described in the previous section. The result of this analysis is transformed into an `agent_event`, reflecting the presence (or not) of a door with an associated degree of certainty. This is then adopted at the deliberative level as a belief, which may in turn result in the firing of a commitment rule.

The use of `agent_events` from the physical level allows a degree of generality and abstraction between the modules. Different applications can define differing responses to `agent_events` through the commitment rules.

APPLICATION TO THE VIRTUAL ROBOTIC WORKBENCH

One of the key tenants of our research has been the provision of a *Virtual Robotic Workbench* [10], which supports the articulation of robot experiments and their subsequent visualisation across the medium of the Internet. This is achieved through the support of multiple views of multiple robot systems. The primary view is the physical perspective of the Nomad Scout IIs navigating the physical world. The secondary, more abstract view is a virtual reality perspective provided via the *Virtual Robotic Workbench*, which delivers a 3-D VRML world via the Internet (fig. 8) where robots are represented as robotic avatars and their associated movements as transformations within this world.

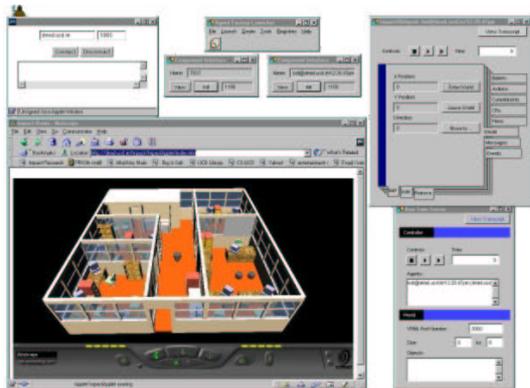


Figure 8: Virtual reality view of the IMPACT research environment

Herein we harness the advantages of using virtual environments by directly relating virtual reality and reality in a seamless manner. This permits multiple views, information hiding and abstraction, and behaviour scrutiny via snapshots and recordings.

The world is comprised of primary location cues. On passage between these locations, an event is sent to

Agent Factory. For instance, as the real robot passes through a door (recognized by both visual analysis and by its characteristic sonar signature) an event is sent such as `possible_door`. On the occurrence of such an event, we may update the avatar position accordingly, by zeroing both the real and the virtual position coordinates. Recalibration is thus achieved and when performed sufficiently frequently yields an accurate portrayal of the robots motion.

RESULTS AND CONCLUSIONS

The experimental results obtained illustrate the power derivable from a sensor fusion approach. The multiplexing of sonar footprints with that of the computer vision system produces a feature recognition system whose disambiguation abilities are considerably better than either of the constituent systems. The location of this sensory fusion subsystem within the social robotic architecture has offered a medium through which experiments can easily be performed.

Over a period of weeks detailed experiments have been conducted such that a robot ascribed with a wall following behaviour explored an environment comprised of a series of rooms. These rooms were not in any way sanitised and represent very real scenarios with the normal degree of clutter associated with a working office. In each test run frames were grabbed by the vision system resulting in a test set of some 400 images. These images are only acquired as a result of attention focusing in order to corroborate the existence of special features.

Experimental results have shown that in 96% of cases the robot is able to successfully identify a door. Experiments relating to other key features seem equally supportive although the number of test cases is considerably fewer. Corners could be successfully recognised in 92% of cases with convex corners proving more problematic than that of concave corners.

Improvements on this disambiguation threshold may be achieved in several ways. Currently we are investigating an idea inspired by work by O'Reagan [19] namely that a rich vein of information may be tapped through focusing on changes from scene to scene². One possible application involves describing how features behave in terms of changes within successive scenes.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of Enterprise Ireland through grant No.'s SC/96/621 COMEDY and

² analogous to the frame problem.

SP/97/08 IMPACT (Initiative for Mobile Programmable Agent-based Computing Technologies).

REFERENCES

- [1] R. Brooks, "A Robust Layered Control System for a Mobile Robot", *IEEE Jour. Rob. and Autom.*, 2(1) 1986
- [2] L. Steels, "Building Agents with Autonomous Behaviour Systems", *The 'artificial life' route to 'artificial intelligence'. Building situated embodied agents.* Lawrence Erlbaum Associates, New Haven. 1994
- [3] J. Von Uexkull, "A stroll through the worlds on animals and men." in *Instinctive Behaviour*, ed. K. Lashley. International University Press.
- [4] D. Ballard, "Animate vision" in *Artificial Intelligence* 48: pp57-86.
- [5] R. Moller, D. Lambrinos, R. Pfeifer, R. Wehner, "Insect strategies of visual homing in mobile robots", *Proc. Computer Vision and Mobile Robotics Workshop, CVMR'98*, pp37-45 FORTH, Heraklion, Greece, 1998.
- [6] A. Saffiotti, "Robot navigation under approximate self-localisation", *Robotics and Manufacturing Vol.6, Procs. 6th ISRAM Symposium*, pp589-594, 1996.
- [7] B. Scheile, and J. L. Crowley, "Certainty Grids: Perception and Localisation for a Mobile Robot", *IRS '93*, Zakopane, Poland, July 1993.
- [8] P.Ramos, F. Lobo Pereira, "Localisation system for an Autonomous Mobile Platform" *Proceedings of the ISIE'97 - IEEE International Symposium on Industrial Electronics*, Guimarães, Portugal, 7th-11th July, 1997.
- [9] K.D. Harris, and M. Recce, (1996) *Localisation and goal-directed behaviour for a mobile robot using place cells* in *Proc. Sintra Workshop on Spatiotemporal Models in Biological and Artificial Systems*, IOS Press, Amsterdam.
- [10] B.R. Duffy, G.M.P. O'Hare, R.P.S. O'Donoghue, C.F.B. Rooney, R.W. Collier, "Reality and virtual reality in mobile robotics", *1st International Workshop on Managing Interactions in Smart Environments MANSE'99*, Dublin December 1999
- [11] B.R. Duffy, R.W. Collier, G.M.P. O'Hare, C.F.B. Rooney, R.P.S. O'Donoghue, "SOCIAL ROBOTICS: Reality and Virtuality in Agent-Based Robotics", in *Bar-Ilan Symposium on the Foundations of Artificial Intelligence: Bridging Theory and Practice (BISFAD)*, 1999.
- [12] G.M.P. O'Hare, B.R. Duffy, R.W. Collier, C.F.B. Rooney, R.P.S. O'Donoghue, "Agent Factory: Towards Social Robots", *First International Workshop of Central and Eastern Europe on Multi - agent Systems (CEEMAS'99)*, 30st May-3rd June 1999, St. Petersburg, Russia
- [13] A.S. Rao, M.P. Georgeff, "Modelling Rational Agents within a BDI Architecture", *Prin. Of Knowl. Rep. & Reas.*, San Mateo, CA., 1991
- [14] B.R. Duffy, C.F.B. Rooney, G.M.P. O'Hare, R.P.S. O'Donoghue, "What is a Social Robot?", *10th Irish Conference on Artificial Intelligence & Cognitive Science*, 1-3 Sept., 1999 University College Cork, Ireland
- [15] J. F. Canny, "A computational approach to edge detection", *Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, November 1986, pp. 679-698.
- [16] R. Deriche, "Using canny's criteria to derive a recursively implemented optimal edge detector", *International Journal of Computer Vision*, vol.1, no. 2, pp. pp167—187, 1987.
- [17] P.V.C. Hough, "Machine Analysis of Bubble Chamber Pictures", *International Conference on High Energy Accelerators and Instrumentation*, CERN, 1959.
- [18] G. Dudek, P. Freedman, I.M. Rekleitis, "Just-in-time sensing: efficiently combining sonar and laser range data for exploring unknown worlds", *IEEE, International Conference in Robotics and Automation*, v1, pp667-671, April 1996.
- [19] J.K. O'Regan, A. Noë, "A sensorimotor theory of visual qualia, and its implications for experiments on change blindness", *Behavioural and Brain Sciences*, under revision, 1999.