Swarm Intell (2011) 5: 45–72
DOI 10.1007/s11721-010-0051-7

Swarm cognition on off-road autonomous robots

Pedro Santana · Luís Correia

Received: 13 May 2010 / Accepted: 3 December 2010 / Published online: 4 January 2011
© Springer Science + Business Media, LLC 2010

Abstract This paper contributes with the first validation of swarm cognition as a useful framework for the design of autonomous robots controllers. The proposed model is built upon the authors' previous work validated on a simulated robot performing local navigation on a 2-D deterministic world. Based on the ant foraging metaphor and motivated by the multiple covert attention hypothesis, the model consists of a set of simple virtual agents inhabiting the robot's visual input, searching in a collectively coordinated way for obstacles. Parsimonious and accurate visual attention, operating on a by-need basis, is attained by making the activity of these agents modulated by the robot's action selection process. A by-product of the system is the maintenance of active, parallel and sparse spatial working memories. In short, the model exhibits the self-organisation of a relevant set of features composing a cognitive system. To show its robustness, the model is extended in this paper to handle the challenges of physical off-road robots equipped with noisy stereoscopic vision sensors. Furthermore, an extensive aggregate of biological arguments sustaining the model is provided. Experimental results show the ability of the model to robustly control the robot on a local navigation task, with less than 1% of the robot's visual input being analysed. Hence, with this system the computational cost of perception is considerably reduced, thus fostering robot miniaturisation and energetic efficiency. This confirms the advantages of using a swarm-based system, operating in an intricate way with action selection, to judiciously

This work was partially supported by IntRoSys, S.A. and by FCT/MCTES grant No. SFRH/BD/27305/2006.

Electronic supplementary material The online version of this article (doi:10.1007/s11721-010-0051-7) contains supplementary material, which is available to authorized users.

P. Santana (✉) · L. Correia
LabMAg, University of Lisbon, Lisbon, Portugal
e-mail: Pedro.Santana@di.fc.ul.pt

L. Correia
e-mail: Luis.Correia@di.fc.ul.pt

P. Santana
UNINOVA, New University of Lisbon, Lisbon, Portugal

control visual attention and maintain sparse spatial memories, constituting a basic form of swarm cognition.

Keywords Swarm cognition · Visual attention · Spatial memory · Action selection · Autonomous robots · Biological inspiration

1 Introduction

Almost all embodied agents, either natural or artificial, deeply rely on perception for a proper interaction with their surrounding environment. Thus, understanding how perception self-organises is essential not only to deepen our understanding about the natural world, but also because it facilitates the synthesis of robust robotic systems. Although the function of perception is well understood, its self-organisation is not. That is, we understand what the perceptual system does, but not so much how it builds up from the interaction of distributed simple processing units. This is more striking if the interaction between perception, action selection, and memorisation is also taken into account in perception models. Uncovering these phenomena is key to enable, for instance, the synthesis of developmental mechanisms capable of allowing robotic systems to learn from their interaction with the environment. This is a daunting task, which could profit from a new modelling paradigm.

The most elementary unit model of robot control systems is the artificial neuron. In nature, neurons are the lower level units building up cognition. Being the physical substrate of cognition, neural activity and topological organisation have been the focus of several models (refer to Gerstner and Kistler 2002 for a thorough review on the field). With these models it is possible to emulate how neurons affect each other and how they behave collectively. This granularity is, however, often too fine when it comes to handle highly complex spatio-temporal problems, where several brain areas are involved. An example is perception which, as previously mentioned, must interact deeply with action selection and memory processes.

Therefore, a new modelling paradigm would be useful, in particular for the synthesis of embodied cognition systems, such as autonomous robots. A possibility that has been gaining momentum in the last few years is to model cognition in terms of the behaviour exhibited by social animals, with particular emphasis on insects. The swarm intelligence paradigm (Bonabeau et al. 1999) is particularly interesting to model perception. This is because insects are simple entities that need to move and interact so as to actively perceive their environment. Being simple, their abstraction is well suited as the atom of the self-organising cognitive system. The abstraction of their collective behaviour, in turn, is useful to model the interactions occurring between the atoms of the cognitive system. As will be shown, the active nature of the swarm elements can be easily exploited to model the active nature of embodied vision systems.

In previous work (Santana and Correia 2010), we have exploited this swarm-based paradigm to synthesise self-organising visual attention, operating in an intricate way with spatial memory and action selection. The result was an accurate and parsimonious perceptual system, performing on a by-need basis and capable of handling the speed-accuracy trade-off in a context-dependent way. However, the model was only validated on a simulated robot performing local navigation on a noise free planar environment.

To assess whether the model is able to exhibit the same interesting properties on a demanding embodied setup, this paper adapts the model to handle its realisation on an all-terrain physical robot equipped with stereoscopic vision. These adaptations are mostly related to the fact that the model must cope with a robot moving on a non-planar terrain, whose

sensory input is a noisy 3-D point cloud. These requirements demand the model to handle the projective nature of the vision sensor and to use evidence accumulation for noise sensitivity reduction. Furthermore, this work deepens the model's biological supporting arguments.

This paper is organised as follows. In Sect. 2, the rationale behind modelling visual attention under the swarm cognition framework is presented. Then, the problem definition is specified in Sect. 3. The proposed model is finally presented in Sect. 4. Next, the results of the validation of the method by a set of experiments in a physical robot is presented in Sect. 5. A discussion on the model's biological plausibility is given in Sect. 6. Finally, a set of conclusions and future work pointers are highlighted in Sect. 7.

2 Introducing swarm cognition for attention modelling

This section describes how this paper builds upon previous work on perception in both natural and applied sciences, as well as how it departs from it.

2.1 Active vision

The act of deploying attention, by means of eyes, head or body motion, is known to be an ubiquitous feature of animals endowed with rich vision (Land 1999). The ability to shift the focus of attention highlights the active nature of perception, rather than a passive one. This means that perception can only be properly studied when the agent is freely acting on its environment, as proposed by Gibson's ecological (ethological) approach to visual perception (Gibson 1979). In fact, body, nervous system and environment must be seen in a holistic way (Ashby 1952; Beer 1995; Thelen and Smith 1996).

In face of the limitations of passive set-ups (Fermuller and Aloimonos 1995), active vision (Bajcsy 1988; Aloimonos et al. 1988; Ballard 1991) has also emerged in the realm of artificial vision systems. Among the advantages of an active vision perspective there is the possibility for the agent to: (1) act in order to shape its sensory information flow (Sporns and Lungarella 2006) so as to make its processing tractable; (2) increase the sensory input signal-to-noise ratio; and (3) reduce ambiguities and consequently perceptual aliasing. These properties are a product of sensorimotor coordination, a concept that can be traced back to Dewey (1896), related to the enactive approach (Varela et al. 1991), and widely recognised as central to the sustainability of embodied adaptive behaviour (Brooks 1991; Ballard et al. 1997; Pfeifer and Scheier 1999; O'Regan and Noe 2001; Pfeifer and Bongard 2006; Mossio and Taraborelli 2008).

Sensorimotor coordination requires a tight coupling between the agent and the environment. Given that the involved timescales of this coupling require fast perceptual processing, the classical view that a global isomorphic representation of the environment is continuously maintained by the agent (Marr 1982) lacks support. Interacting perceptual pathways, dedicated to different purposes, are more likely to exist (Milner and Goodale 1995; Goodale 2008). For instance, pathways associated to motor action control, supposedly the dorsal visual stream, are prone to map directly to motor actions. This is congruent with the idea of the "world as its own best model" championed by Brooks (1991). Being perception so intricately linked to action, it is reasonable to assume that both unfold in parallel.

2.2 Parallel covert attention

Focussing perception on the most relevant aspects/regions of the environment, taking into account the current context, is advantageous to increase both performance and robustness

(Itti and Koch 2001; Hayhoe and Ballard 2005; Rothkopf et al. 2007). Furthermore, the intricate relationship between visual attention and other structures, such as bottom-up sensory pathways, categorical reasoning, and action selection, highlights its centrality and consequently its relevance to the understanding of the dynamics involved in an embodied cognitive system. Attention ultimately results in the motion of sense organs, e.g., eyes, towards the relevant stimulus source. This is called overt attention. A faster process is the one of mentally focussing on particular aspects of the sensory stimuli. This is called covert attention and it will be the main focus of the following discussion.

Computational models of visual attention typically assume the existence of a sensor-driven bottom-up pre-attentive component (Treisman and Gelade 1980; Koch and Ullman 1985; Itti et al. 1998; Palmer 1999; Corbetta and Shulman 2002; Hou and Zhang 2007), which drives attention to salient regions of the visual input. The more a region of the visual input detaches from the background, the higher its level of saliency. Top-down contextual knowledge on the visual search task, i.e., on the object being sought, is also known to play an important role on the modulation of attention (Yarbus 1967; Wolfe 1994; Tsotsos et al. 1995; Corbetta and Shulman 2002; Torralba et al. 2003; Frintrop et al. 2005; Navalpakkam and Itti 2005; Walther and Koch 2006; Neider and Zelinsky 2006; Rothkopf et al. 2007; Hwang et al. 2009), as has been shown by recent neurophysiological studies (Egner et al. 2008). The outcome of the interplay between these two processes is a saliency map whose activity level is higher in the regions of the visual input that are more likely to contain the object being sought. The saliency level is typically assumed to guide the motion of a "spotlight" from the most to the least relevant regions of the visual input. In the case of overt attention, this spotlight guides the motion of the sense organs to centre the visual input on the region of interest. In the covert case, a task-specific object recogniser is applied at each region where the spotlight falls (see Navalpakkam and Itti 2005). This approach highlights a dichotomy between the parallel nature of saliency map computation and the sequential nature of the processes operating on its basis. Due to the involvement of mechanical components, this sequencing is unavoidable in overt attention. Conversely, although constrained to process sensory input that is sequentially determined by the overt attention, the covert attention process is more likely to operate in a parallel fashion. Two main arguments can be put forth to support this hypothesis.

First, the human brain is a massively parallel structure, which renders unlikely that it would be instantaneously dedicated to a single covert attentional sequential process. That is, the effort of maintaining a unique purely sequential analysis is costly by itself, as it requires complex arbitration mechanisms. Second, studies with human subjects revealed the existence of a unitary spotlight of covert attention to be unlikely (Pylyshyn and Storm 1988; Doran et al. 2009). The tests involved subjects paying covert attention to objects that were moving so fast that the chances of effective attention switching were considerably low. The observed successful tracking of these objects supports the multiple spotlights hypothesis, which is also coherent with evidence on the existence of independent tracking mechanisms in the two cerebral hemispheres (Alvarez and Cavanagh 2005).

There is a considerable bulk of knowledge suggesting that both overt and covert shifts of attention share the same neural mechanisms (de Haan et al. 2008), supporting the premotor theory of attention (Rizzolatti et al. 1987). Impressively, this link between response preparation and spatial attention is not limited to the oculomotor system (Eimer et al. 2005). Thus, this theory predicts that common sensorimotor coordination mechanisms are involved in the control of attention and action. We argue that this opens the door for the application of sensorimotor coordination principles, studied mostly on overt control, to better understand and explore the dynamics of covert attention.

2.3 Agent abstraction of covert attention

In a more classical view, the control of the attention spotlight would be considered exogenous, in the sense that it is only function of sensory saliency and top-down modulation signals. However, under the assumption that covert attention is a process of sensorimotor coordination, it is more reasonable to consider the spotlight as a dynamical entity inhabiting the sensorimotor space of the embodied agent. This entity can then pro-actively move in this space to guide the focus of the embodied agent in a sensorimotor coordinated way, and thus exploiting all the aforementioned advantages of sensorimotor coordination to shape the sensory input. This entity, from now on called an agent, thus behaves as a locally sequential covert attentional process.

The feasibility of agent-based modelling for sensorimotor coordination has been extensively validated in the fields of embodied cognition and active vision (Scheier et al. 1998; Nolfi and Marocco 2002; Beer 2003; Fend et al. 2003; Balkenius et al. 2004; Floreano et al. 2004; Nolfi 2005; Suzuki and Floreano 2006; Pfeifer and Bongard 2006; Kim and Moeller 2006; Sporns and Lungarella 2006; de Croon and Postma 2007; Choe et al. 2008). A population of agents can generate coherent collective (parallel) behaviour. In perception, the agents' motion corresponds to an attention shift, whereas the collective spatio-temporal self-organised pattern implements a multi-focus attention process. Brain computational modelling with multiple agents is not a new idea (Minsky 1988; Chialvo and Millonas 1995). Although the first realisations for computer vision related problems are also not new (Poli and Valli 1993; Liu et al. 1997), only more recently it has received considerable attention (Ramos and Almeida 2000; Owechko and Medasani 2005; Antón-Canalís et al. 2006; Mobahi et al. 2006; Broggi and Cattani 2006; Mazouzi et al. 2007; Zhang et al. 2008). In general, these models exploit the metaphor of swarming behaviour on social insects.

These computational models are mostly stand-alone engineered parallel perceptual systems, lacking a body capable of purposive behaviour. This deficit undermines their explanation power regarding the mechanisms actually building up adaptive behaviour. Conversely, sensor planning, which is a relatively stable field in computer vision and robotics communities, is actually trying to bridge the gap between body motions and information gathering through the sensors (Dickmanns et al. 1990; Sukthankar et al. 1993; Kelly and Stentz 1998; Behringer and Muller 1998; Nabbe and Hebert 2003; Kwok and Fox 2004; Patel et al. 2005; Bagdanov et al. 2006; Urmson et al. 2006; Tessier et al. 2007; Hernandez et al. 2007; Sprague et al. 2007). However, none of these works models parallel covert attention operating in an intricate way with the action selection process.

2.4 Swarm cognition for attention modelling

The previous sections showed the relevance of a parallel model for covert attention, as well as the importance of its tight coupling with motor action preparation. It was also suggested that the multi-agent paradigm is appropriate for this modelling task.

A remarkable metaphor from the natural world encompassing these characteristics, and consequently exploited in this paper, is the foraging behaviour of army ants. By exploiting pheromone-based local interactions, these ants are able to forage in large areas around their nest in a parallel and robust way (Deneubourg et al. 1989). These foraging ants exhibit a sort of collective intelligence (Franks 1989), allowing the group to be seen as an individual decision making process.

Rather than static structures, like neurons, these agents are better viewed as active information particles that flow through the system. Hence, using agents, the design focus is on

the process and not so much on its supporting substrate. Being sensorimotor coordinated units, these information particles can actively shape their sensory input, use their sensorimotor history to induce long-range influences on other information particles, and in the limit improve their own behaviour. When together, these modular units can exploit the synergy of self-organisation and emergent properties. We argue that reaching the complexity of such a system with a connectionist model is possible but not likely tractable. Although not covered in this article, we foresee the usefulness of connectionist models to implement the agents themselves, as it would enable tractable neuro-evolution (Floreano et al. 2008) to build complex systems. Hence, we suggest that using an agent-based design the system modeller is able to reach higher levels of tractable complexity than when using connectionist models. Although the results reported in this paper show the benefits of an agent-based design, further phenomenological support is still required for a definitive comparison with connectionist models.

The use of the insect swarm metaphor to model vertebrate brain function has been also suggested by parallel and independent work (Passino et al. 2008; Couzin 2009; Marshall et al. 2009; Marshall and Franks 2009). Nest site selection in honey bees, for instance, exhibits some of the characteristics of decision making in brain, such as the statistical speed-accuracy trade-off predicted by diffusion models (Ratcliff and Smith 2004). Our work, instead, approaches the problem of studying cognition through social insects behaviour by building it on an embodied setup, following the synthetic approach to embodied cognition (Pfeifer and Scheier 1999) and Artificial Life (Bedau 2003). The advantages of using Artificial Life models for this purpose, though without the support of any practical realisation, were reviewed also in a parallel study (Trianni and Tuci 2010).

All these accounts can be framed in the emerging multidisciplinary field of Swarm Cognition, which attempts to uncover the basic principles of cognition exploiting self-organising principles, mainly those exhibited by social insects. However, as we only conceive cognition under the embodiment framework, we consider that a swarm-based model not involved in the embodied agent's sensorimotor coordination loop can hardly be considered as an instance of Swarm Cognition. Note that all these considerations are under the assumption that Swarm Cognition can model and help to understand the behaviour of a multi-cellular individual.

3 Problem definition

This section specifies the objective, hypotheses, and assumptions related to the proposed model. The problem to solve is primarily the one of minimising the perceptual resources required to support the proper execution of the action selection process controlling the embodied agent, hereafter simply called robot. Saving time in perception enables faster robot motion. The ultimate goal is to show that this complex cognitive problem can be easily managed by recurring to the swarm metaphor.

A perceptual resource is said to be expended when a given object detector is applied to a given region of the robot's visual input. Minimising perceptual resources expenditure is achieved by reducing the number of detector applications to the point from which the action selection process would no longer be sufficiently informed for a proper decision making.

Supported by the previous section, here we describe the three main hypothesis tested in this work. The first hypothesis is that the considered problem can be solved by allowing both perceptual and action selection processes to evolve in synergy, as it should promote the ap-

plication of the detector, i.e. focus attention, on pixels whose positive detection most impacts the action selection process, and consequently helps it in rapidly stabilising its output. If the robot's goal is to move forward, detecting an obstacle in its frontal region will compel a sudden change from forward motion to a turn. Conversely, an obstacle detected on the robot's side would effect no change in action selection. Hence, finding first the frontal obstacle has the highest impact on the action selection process and consequently helps the robot reaching the correct decision faster. The second hypothesis is that to ensure that the attentional process is robust, it should be parallel and provided with self-organising properties. Given the unpredictability of the environment, the complexity of taking into account the interaction with the action selection process, and the difficulty of defining a fixed speed-accuracy trade-off, turns the alternative of using an optimal covert visual attention policy into an infeasible solution. Finally, the third hypothesis is that an adequate formalism to model this parallel covert visual attention process is the one describing the collective behaviour exhibited by social insects.

In previous work (Santana and Correia 2010), we have already tested and confirmed these three hypotheses in the context of simulated robots performing local navigation in 2-D noise free environments. This article tests the three hypotheses on a more generalised setup, namely, on a physical robot performing local navigation in 3-D noisy environments. In turn, the added complexity of such an experimental apparatus demands some adaptations to our previous work (Santana and Correia 2010). Concretely, the model has been extended with an evidence accumulation mechanism (see Sect. 4.3.3) to better handle the noisy nature of the sensory input. This is a fundamental mechanism since, without it, noise impinging the robot's sensors would propagate through the system, hampering the maintenance of a coherent cognitive representation. We also needed to handle the projective nature of the employed vision sensor in a rough terrain.

Without loss of generalisation, the robot used to validate the hypotheses is a wheeled robot equipped with a stereoscopic vision sensor that must be able to perform local navigation in rough terrain. Hence, the object detector in this case is a 3-D obstacle detector, whereas the action selection process is a fast obstacle avoidance algorithm. The following sections detail these assumptions.

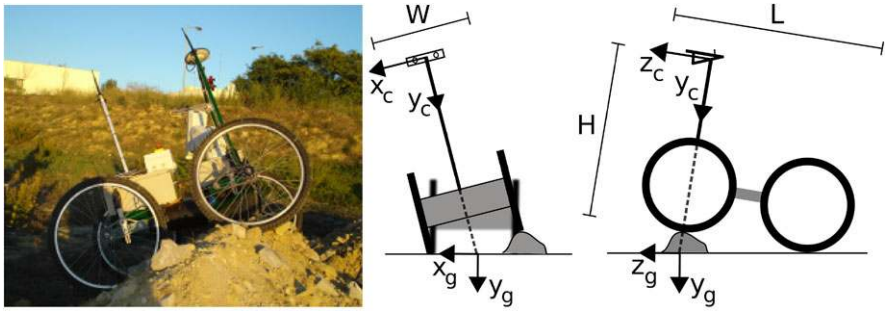
3.1 The Ares robot

The robot employed in this study, i.e., the Ares robot (Santana et al. 2008), is a vehicle with four independently steering wheels (see Fig. 1). Although it enables the use of several locomotion modes, here only the Double Ackerman mode is considered. In this mode, the robot moves in a car-like way, but with both front and rear wheels steering symmetrically. This gives the robot a considerable manoeuvrability.

Off-road terrains are uneven and densely populated with protuberances, such as small rocks (see Fig. 1). In these situations the robot exhibits considerable tilt levels, which makes the relative position and pose of the sensor with respect to the ground plane to vary as the robot moves.

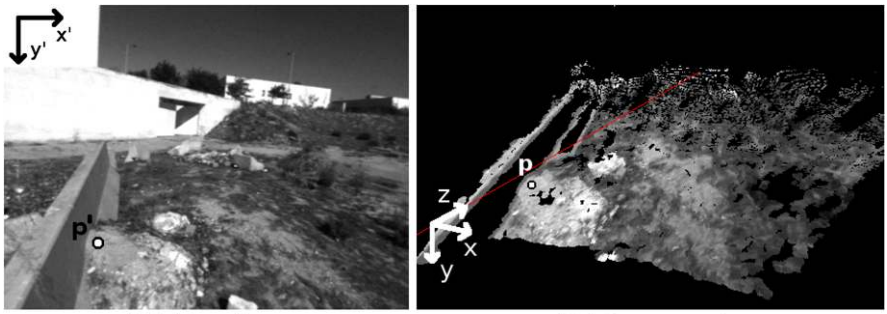
The robot is equipped with two cameras calibrated to perform stereoscopy. This means that each pixel from each camera is associated to a 3-D coordinate in the camera's frame of reference, provided that stereoscopy could be computed for the pixels in question (see below). The left input image, L, is taken as reference and hereafter simply called visual input or frame.

Given that stereoscopy can be computed for the pixel $\mathbf{p}'$ in L, this pixel can be back-projected so as to obtain the corresponding 3-D point $\mathbf{p}$. With a projection matrix P obtained



(a) physical robot (b) front view (c) side view

Fig. 1 The Ares robot. In the depicted situation the Ares robot is stepping an obstacle with its frontal left wheel, which results in non-zero pitch and roll angles with respect to the ground plane



(a) left input image (b) 3-D reconstruction

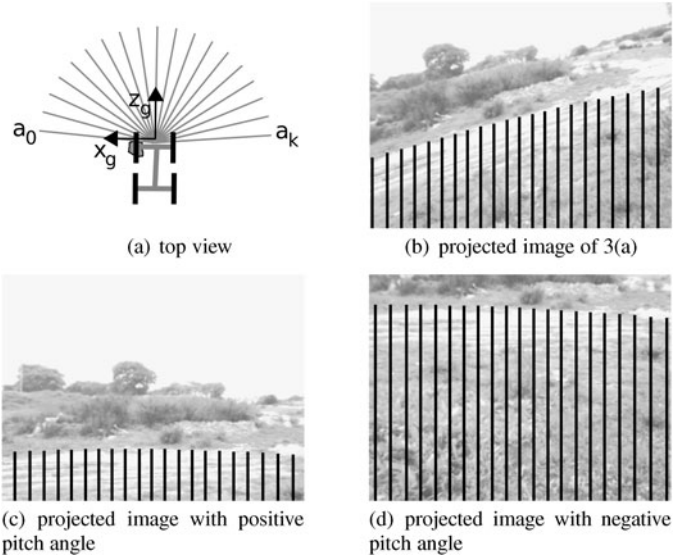
Fig. 2 (Colour online) 3-D reconstruction process. (**a**) Left input image obtained with stereoscopic vision, L. (**b**) Reconstructed 3-D model limited to a depth of 10 m. The 3-D point **p** is projected onto L in pixel $\mathbf{p}'$. The *red line* corresponds to the optical axis of the left camera. Note the absence of computed depth in many pixels, in particular along the homogeneous vertical structure on the left and frontal walls

from calibration, it is possible to do the inverse operation, i.e., to determine which pixel $\mathbf{p}'$ in L is the projection of the environment's 3-D point **p** (see Fig. 2).

The distance between both cameras is 30 cm. This enables sufficiently accurate depth computation between 2.5 m and 20 m. This means that obstacles close to the robot are not perceivable, making mapping an essential asset for this configuration. Two additional aspects render stereoscopic vision a challenging sensory modality. First, absence of 3-D information in low texture and under illuminated surfaces is rather common (see Fig. 2). This owes mostly to the difficulty of stereo processing to match both left and right camera images in those regions. Second, high levels of noise are common with poor lighting condition and depth computation error grows quadratically with range. Hence, in order to operate with a stereoscopic vision setup, a perceptual system must be highly robust.

3.2 Mapping between world and sensor coordinate frames

In this study, robot actions are considered to be linear trajectories radially overlaid on the ground in front of the robot, i.e., parallel to plane $z_g x_g$ and with $y_g = 0$ (see Fig. 1). To allow obstacle search to be affected by action selection, these linear trajectories must be projected onto L.



(a) top view (b) projected image of 3(a)

(c) projected image with positive pitch angle (d) projected image with negative pitch angle

Fig. 3 Action set parallel to the ground plane in (**a**), and respective projection on L in (**b**). In this situation, the robot is stepping an obstacle with its frontal left wheel (as depicted in Fig. 1). Note that the projected action set (*black lines*) is compensated for the robot's non-zero roll and pitch angles. The *central line* in all projected images correspond to the straight ahead motion. Note that even with considerably different pitch angles, as in (**c**) and (**d**), the projected lines terminate roughly in the same image regions. That is, they are clearly relative to the ground-plane coordinate system defined by $\{x_g, y_g, z_g\}$, and so invariant to the robot's pose

This projection requires the estimation of transformation Q, between the ground's frame of reference, $\{x_g, y_g, z_g\}$, and the camera's frame of reference, $\{x_c, y_c, z_c\}$. Due to terrain's roughness, the 4×4 homogeneous transformation matrix Q is computed at each frame by recurring to a robust ground-plane fitting process (Santana et al. 2009). After transforming with Q the linear trajectories to the camera's frame of reference, these can finally be projected onto L with projection matrix P (see Fig. 3).

In order to reduce the chances that obstacles in the environment may be confounded with the terrain itself during the ground-plane estimation process, the pixels of the obstacles that have been detected in the previous frame are discarded from the process (van der Mark et al. 2007).

3.3 Obstacle detection

In this study, obstacles are detected in a pixel-wise way over the visual input L. Concretely, a pixel is said to be an obstacle point whenever the pixel's associated 3-D point is at a height from the estimated ground plane that cannot be climbed by the robot (Santana et al. 2009). Formally, the detection result at pixel $\mathbf{p}'_\mathbf{a} = (x_a, y_a)$, with the corresponding 3-D point $\mathbf{p_a} = (X_a, Y_a, Z_a)$, is given by,

$$D(\mathbf{p}'_\mathbf{a}) = \begin{cases} 1 & \text{if } d(\mathbf{p_a}, (a, b, c, d)) > h_{\min} + \beta Z_a \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where, $d(\mathbf{p_a}, (a, b, c, d))$ is the orthogonal distance between point $\mathbf{p_a}$ and estimated ground plane $ax_c + by_c + cz_c + d = 0$, as computed in Sect. 3.2. Note that $\mathbf{p_a}$ is given relative to

$\{x_c, y_c, z_c\}$. The term βZ_a makes the detection threshold grow with the range, Z_a, of the point in question. This mechanism compensates for the range-dependent error growth in the 3-D reconstruction process. Thus, it reduces the potential for false positives in the far field. The lack of 3-D information in some pixels results in the impossibility of applying the detector in these cases.

Note that the simplicity of this detection method comes at the cost of only being applicable to moderately rough terrain. More complex detectors (see, e.g., Santana et al. 2010), which are required for more demanding environments, could be used in the model without any specific customisation. In fact, the more complex the detector, the more expensive is its computation, and consequently the more important is to focus the robot's attention.

3.3.1 Motion estimation

A key component of any complex robotic system is motion estimation. With it, the system will be able for instance to produce local maps of the environment.

Frame-to-frame robot motion, i.e., translation and rotation, is computed with a particular model (Santana and Correia 2008) of visual odometry (Matthies 1989; Agrawal and Konolige 2006). Visual odometry basically estimates the rigid body transformation matrix M that explains the change in position of notorious 3-D points in the environment across frames. The major advantage of this method over others, like dead-reckoning, is that it is fully synchronised with the sensor data feeding the perceptual process. This results in very accurate registration of information across frames.

4 Model description

This section starts by mapping the biological metaphor of foraging strategies in social insects to the concepts of parallel covert visual attention to be employed in the proposed model. In short, the metaphor is exploited to provide the model with the self-organising properties essential to maintain a robust parallel focus of attention. After presenting the biological inspiration, the proposed model is described. Since it is built upon authors' previous work, it is not described with full detail. For additional information the reader is referred to the previous publication (Santana and Correia 2010).

4.1 Biological inspiration

Covert visual attention is mostly about parallel search of objects in the robot's visual input. A remarkable metaphor for this process is the one of army ants engaging in foraging behaviour (see Sect. 2.4). Assuming that the environment where these ants inhabit corresponds to the robot's visual input, each ant can be seen as an individual covert visual attention process. As a consequence, their collective behaviour can be considered as a parallel covert visual attention process.

Following this metaphor from the natural world, the visual process in our model is composed of a swarm of simple homogeneous virtual agents that inhabit the visual input of the robot. These agents are probabilistically created (recruited) and therefore they receive hereafter the designation of p-ants. They search (forage) the robot's visual input along those regions where detected obstacles are more likely to stronger affect the action selection process. Hence, p-ants operate on a by-need basis being driven by the action selection process. In the

case of local navigation, the utility of perceiving a given region of the visual input is related to the utility of navigating on its corresponding region of the environment. In the proposed model, these regions are associated to the projected linear trajectories, as described in Sect. 3.2.

As in natural ants, p-ants do this search in a stochastic and coordinated way. By not behaving greedily, the system is more robust to unforeseen situations and faster in adapting to contextual changes. P-ants interact through *stigmergy*, i.e., they interact through perceptual shared mediums, for better coverage and tracking of detected obstacles. This allows the coexistence of positive and negative feedback loops that lead to robust collective behaviour. In conclusion, random fluctuations and both positive and negative feedback, which are necessary ingredients for self-organisation to occur (Bonabeau et al. 1999), are included in the model.

4.2 Proposed model

The proposed model is decomposed into two interacting processes, one for perception and another for action selection. The perceptual process includes the parallel covert visual attention aspects. Figure 4 illustrates the connectivity between both processes. Basically, after receiving a new frame, L, the perceptual and action selection processes interact (thicker arrows) for $i_{\max}$ iterations before a final motor action decision is reached and eventually engaged. These interactions occurring between both processes allow them to progressively unfold in parallel, and consequently, to enable accurate deployment of visual attention.

At each iteration, the action selection process sends a message to the perceptual process with an action utility vector, $\mathbf{u} = (u^1, u^2, \ldots, u^k)$, where $u^j \in [0, 1]$ is the utility of performing action j, and k is the cardinality of the robot's action repertoire. The action utility vector is computed according to a desired heading of motion, h, and constrained by information about free-space connectivity of the local environment, C, which has been sent by the perceptual process in the previous iteration. This set encompasses range information regarding those radial sectors of the local environment that contain free-space for robot motion. In this study, actions u^j are defined as linear trajectories centred on the robot and directed outwards in a radial pattern (see Sect. 3.2). Hence, the utility of each possible action is defined as the value of moving the robot along the corresponding linear trajectory given a goal location, $\mathbf{p}_{\text{goal}}$, and obstacles disposition in the environment. The more directed a given linear trajectory is to the goal location, the higher its utility. The closer an obstacle is to the robot along a given linear trajectory, the lower the trajectory's utility. The final utility is a combination of both factors.

At each iteration, the perceptual process uses the just received vector $\mathbf{u}$ to iterate its search for obstacles on a by-need basis (see Sect. 4.3). The positions of obstacles detected

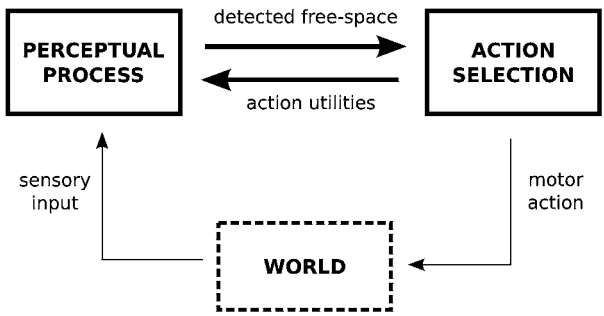


Fig. 4 Building blocks of the proposed model, adapted from Santana and Correia (2010)

since the arrival of the current frame are accumulated in set O. This set is then used to compute C. Finally, C is sent back as a message to the action selection process so as to induce its next move. The action with highest utility at the time the maximum number of iterations is reached is passed to the low-level motion controller through a low-pass filter. This way, sudden changes at the system's output are smoothed to avoid jitter at the actuators level.

4.3 Perceptual process

The perceptual process is composed of a set of p-ants, A, whose elements individually implement a covert attention process. The size of this set is zero at the first frame and varies along iterations.

At each iteration, the perceptual process starts by appending new p-ants to the set A (see Sect. 4.3.1). Then, each p-ant moves, according to its set of behaviours (see Sect. 4.3.2), updating its 2-D position in the visual input. Possibly, some of the p-ants will find an obstacle to track across iterations and even frames. In between frames, the position of every p-ant is updated to compensate for any robot motion that has occurred (see Sect. 4.3.4). This way, its position is kept relative to the environment and consequently invariant to robot motion. This update is done by applying to p-ant's 3-D position, obtained with stereoscopy from its 2-D position in the visual input, a transformation matrix representing the robot motion (see Sect. 3.3.1). The transformed 3-D point is then projected back onto the visual input with a projection matrix. In the process, due to robot motion, this projection may fall out of the visual input, which means that the p-ant has been moved out of the robot's visual field of view. If at that moment the p-ant was tracking an obstacle it is now said to constitute part of a body-centric local map. To maintain the local map updated (i.e., body centred), p-ants in that situation are also motion compensated in between frames. If a given part of the environment is revisited, the projected positions of some of these p-ants will be again within the visual field of view. These p-ants are then reactivated in order to track their associated obstacles in the visual input. At the end of each iteration, the 3-D positions of all p-ants that are tracking an obstacle, either in or out of the robot's visual field of view, may be used to extend the obstacles set O. Evidence accumulation (see Sect. 4.3.3) on the presence of an obstacle is used by each p-ant to determine whether or not to append its position to set O.

For this process to operate, every p-ant $a \in A$ needs to maintain state information across iterations and frames. This information includes p-ant's body-centred 3-D position, $\mathbf{p_a}$, and its 2-D projection onto a plane parallel to and centred on visual image L, $\mathbf{p'_a}$. With projection matrix P, $\mathbf{p'_a}$ can be computed given $\mathbf{p_a}$. Hence, $\mathbf{p'_a}$ is a 2-D vector defined in the reference frame of L. With stereoscopy, $\mathbf{p_a}$ can be computed given $\mathbf{p'_a}$, provided that $\mathbf{p'_a} \in FOV$, where FOV is the set of possible positions in L. If $\mathbf{p'_a} \notin FOV$ or stereoscopy failed to back-project $\mathbf{p'_a}$, then $\mathbf{p_a}$ can only be estimated, such as through compensating its previous state with the transformation matrix representing the robot motion estimate, M. Since with P it is possible to obtain $\mathbf{p'_a}$ from $\mathbf{p_a}$, one can then assess when the p-ant re-enters the visual input, i.e., when $\mathbf{p'_a} \in FOV$.

The following sections details how the perceptual process works in each iteration.

4.3.1 P-ants creation and removal

With the exception of the first iteration in each frame, the set of p-ants, A, is extended by creating new p-ants as a function of the incoming action utility vector, $\mathbf{u}$. This set is empty in the first frame. P-ants are not created in the first iteration of each frame because, at that time, vector $\mathbf{u}$ is not yet updated (set C not yet computed).

The higher the robot's speed, $s \in [0, 1]$, and the utility of a linear trajectory, j, the higher the chances of creating a corresponding p-ant, a_j, in L. The initial 2-D position of this p-ant, $\mathbf{p}'_\mathbf{a}$, is the point where the linear trajectory j projected onto L intersects the bottom row of L (see Sect. 3.2). This way p-ants start their search for obstacles in the close vicinity of the robot.

The newly created p-ant a_j is endowed with an initial energy level, ρ, which is reduced by one unit in each iteration and restored whenever the p-ant considers to be on an obstacle. With zero energy, the p-ant is removed from the system, $A \leftarrow A \backslash \{a_j\}$, to avoid memory and computation to grow unbounded.

4.3.2 P-ants behaviours

Figure 5 illustrates the finite state machine responsible for the switching of the behaviours ruling each p-ant's activity. The following describes these behaviours as well as their transitions.

After appending new p-ants to A, every p-ant $a \in A$ looks for obstacles at its position $\mathbf{p}'_\mathbf{a}$. To reduce computation, it first looks at a perceptual shared medium S, which represents obstacles detected previously in the current frame by any p-ant. If $\mathbf{p}'_\mathbf{a}$ is not represented in S and there is 3-D information associated to that point, the obstacle detector $D(\mathbf{p}'_\mathbf{a})$ (see (1)) is called and the result is stored in S. In the absence of 3-D information the point is considered without obstacle.

At creation time, every p-ant a starts in *search behaviour*. As depicted in Fig. 6(a), each iteration of this behaviour performs a simple stochastic motion step on L along the associated projected linear trajectory,

$$\mathbf{p}'_\mathbf{a} \leftarrow \mathbf{p}'_\mathbf{a} + \mathbf{v}'_\mathbf{a} \tag{2}$$

where $\mathbf{v}'_\mathbf{a}$ is a velocity vector with angular direction $\lambda_1 \cdot N(0, 1)$ and magnitude $\lambda_2 \cdot N(0, 1)$, with $N(0, 1)$ sampling a number from a Gaussian distribution with mean 0 and variance 1,

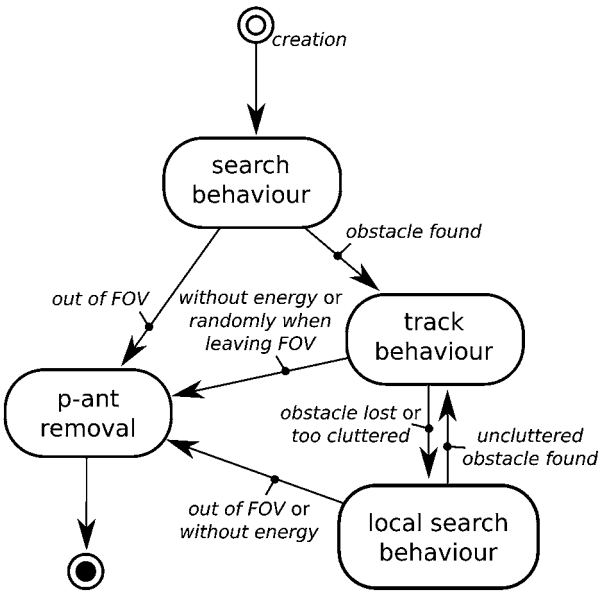


Fig. 5 Finite state machine representing each p-ant's activity. *Ovals* and *links* represent behavioural states and their transitions, respectively. The labels associated to each transition specify the conditions for the transitions to occur

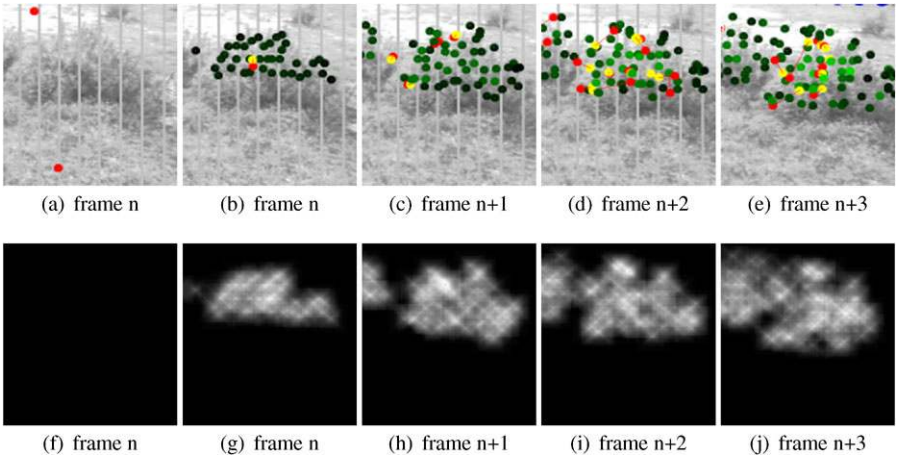


Fig. 6 (Colour online) Detection, diffusion and evidence accumulation example. The images correspond to a small region of the current frame. *Red dots* are p-ants in *search behaviour*. *Green dots* refer to those p-ants that are in *track behaviour*. The lighter the green the greater the evidence on the presence of the obstacle. The *vertical lines* correspond to a sub-set of projected linear trajectories with positive utility. Note that only at frame $n + 3$ these lines disappear, showing the progressive accumulation of evidence on the presence of the obstacle before reporting it, i.e., appending it to O. It is also possible to depict the diffusion process in (**b**), triggered by a single p-ant (*red dot* off the obstacle), where p-ants are cloned to better cover the obstacle. *Red dots* on the obstacle represent p-ants executing *local search behaviour* around a given anchor point, represented by the connected *yellow dots*. This occurs when the region is already too cluttered with other p-ants. The activity level in the perceptual shared medium, R, which is used by p-ants to assess the level of clutter and to help each other on finding the obstacle when in local search, is depicted in the *bottom row*. Videos with situations as the one depicted in this figure are available as online supplementary material

and λ_1 and λ_2 are empirically defined scalars. High λ_1 and λ_2 values facilitate fast detection of large obstacles, at the cost of missing smaller ones. Small values result in slower, though finer, detection. If while in this behaviour the p-ant moves out of the visual input boundaries, $\mathbf{p}'_\mathbf{a} \notin FOV$, the p-ant is removed from the system, i.e., $A \leftarrow A \backslash \{a\}$. This may happen due to p-ant's motion or due to the influence of the robot motion compensation process.

If an obstacle is found by a p-ant while in *search behaviour*, it switches to *track behaviour*. As illustrated in Fig. 6(b), in this new behaviour, p-ants are allowed to clone themselves for $r_{\max}$ times. A single cloning is attempted per iteration by randomly selecting a position in the neighbourhood (radius between 10 and 30 pixels) of the p-ant's current position. If the randomly selected position contains an obstacle and it is not too cluttered with other p-ants, then a clone is created and initiated on it. A clone inherits the number of replications of its ancestor so as to control the diffusion process, whose goal is to rapidly cover the detected obstacle.

A p-ant assesses the level of cluttering by inspecting the level of activity at its position on a perceptual shared medium, R. This 8-bit structure has the size of L. Therefore, p-ant's position there corresponds to p-ant's position on L. The activity level is a clutter measure because all p-ants in *track behaviour* increase the local activity of the perceptual shared medium R at the corresponding locations. Hence, R represents pheromone directly emitted by p-ants, and not a chemical they lay in the environment. For a p-ant a, this is done by adding to R a top-view pyramidal shape of top magnitude 20 and linear decay (0.9) outwards, centred on $\mathbf{p}'_\mathbf{a}$. If the activity at $\mathbf{p}'_\mathbf{a}$ in R is higher than an empirically defined scalar,

$\mathrm{R}(\mathbf{p_a}) > \eta$, it is said that the p-ant's position is too cluttered. Figure 6(f–j) traces the evolution of shared medium R for a given situation.

Besides cloning, p-ants in *track behaviour* have no additional activity. Nevertheless, robot motion compensation (see Sect. 4.3.4) will maintain their positions relative to the associated obstacles, independently of robot motion. As the robot moves, these p-ants will eventually leave the robot's visual input and contribute to the robot's spatial working memory with probability ψ. That is, not all p-ants leaving the visual input become part of the spatial memory. This helps on the regulation of the computational load by reducing the sampling of the peripheral environment.

Being inaccurate, the estimated robot motion is usually insufficient to completely cancel the effects of robot motion on the relative position of p-ants. The outcome is that p-ants eventually loose track of the obstacles. To cope with this, any p-ant in *track behaviour* that considers to have sufficiently accumulated evidence of being no longer on an obstacle switches to *local search behaviour*. As will be shown in Sect. 4.3.3, evidence grows and decays as the obstacle detector returns positive and negative, respectively. As a result, if evidence in obstacle presence is high, transient obstacle detection negatives may not be sufficient to engage behaviour switching. This is key to deal with noisy sensory data.

The switch to *local search behaviour* also occurs if the p-ant in position $\mathbf{p}'_\mathbf{a}$ determines that it is on an obstacle too cluttered with other p-ants, $\mathrm{R}(\mathbf{p}'_\mathbf{a}) > \zeta$. By making $\eta < \zeta$, a sort of hysteresis is implemented, and thus massive fluctuations of neighbouring p-ants entering and leaving *local search behaviour* are avoided.

As depicted in Fig. 6(c–e), when in *local search behaviour*, a given p-ant a performs a random walk around an anchor point, $\mathbf{z}'_\mathbf{a}$, whose initial position is defined as the p-ant's position at the time *local search behaviour* was initiated, $\mathbf{z}'_\mathbf{a} \leftarrow \mathbf{p}'_\mathbf{a}$. Hence, with the goal of re-detecting a lost obstacle or of finding a less cluttered region, $\mathrm{R}(\mathbf{p}'_\mathbf{a}) < \eta$, this behaviour randomly changes the position of the p-ant around the anchor point as follows,

$$\mathbf{p}'_\mathbf{a} \leftarrow \mathbf{z}'_\mathbf{a} + \sigma_a \big[N(0,1), N(0,1)\big]^T \tag{3}$$

where $\sigma_a \leftarrow \min(\sigma_a + 1, l)$ and $N(0,1)$ samples a number from a Gaussian distribution with mean 0 and variance 1. By increasing σ_a at each iteration, the local search spreads up to the upper-bound l, which constrains the search to avoid migration of p-ants between obstacles.

The anchor point changes to the p-ant's current position, $\mathbf{z}'_\mathbf{a} \leftarrow \mathbf{p}'_\mathbf{a}$, whenever the clutter level there is higher than the clutter level at the current anchor's location, yet not too high, $\eta > \mathrm{R}(\mathbf{p}'_\mathbf{a}) > \mathrm{R}(\mathbf{z}'_\mathbf{a})$. This directs the local search towards uncluttered regions where other p-ants reported the existence of an obstacle. If it is not possible to move the anchor point, p-ant's energy level is dramatically reduced by an amount υ. Since p-ants may be in *local search behaviour* across frames, the anchor points must be compensated for the robot's motion the same way p-ants positions are. If a p-ant in *local search behaviour* happens to detect a not too cluttered obstacle, $\mathrm{R}(\mathbf{p}'_\mathbf{a}) < \eta$, it switches to *track behaviour*. Finally, p-ants in *local search behaviour* that leave the robot's visual input are removed from the system. This may happen due to p-ant's motion or due to the influence of the robot motion compensation process.

4.3.3 Evidence accumulation

In the previous section nothing was said related to the update of the obstacles set O. This is done by having every p-ant a in *track behaviour*, whose *evidence* on the presence of an obstacle on its position is sufficiently high, appending its 3-D position, $\mathbf{p_a}$, to O. This

accumulation of evidence before reporting an obstacle is important to reduce sensitivity to noise in the 3-D point cloud. Otherwise, noise in the sensory input would propagate to the action selection process. The following describes how evidence is integrated by each p-ant.

At the first iteration of each frame, every p-ant a that (1) is in *track behaviour*, (2) is within visual input boundaries, $\mathbf{p}'_\mathbf{a} \in FOV$, and (3) its position has associated 3-D information, updates its evidence on obstacle presence, y_a, according to the following equation,

$$\tau \dot{y}_a = -y_a + \gamma D\left(\mathbf{p}'_\mathbf{a}\right)\mathrm{R}\left(\mathbf{p}'_\mathbf{a}\right) \tag{4}$$

where τ and γ are empirically defined scalars, $\mathrm{R}(\mathbf{p}'_\mathbf{a}) \in [0, 1]$ is the activity level of the perceptual shared medium R at p-ant's position, $\mathbf{p}'_\mathbf{a}$, and $D(\mathbf{p}'_\mathbf{a})$ returns the result of the pixel-wise obstacle detector (see above).

With this system, evidence decays to zero when an obstacle is not found for several frames, provided that 3-D information is available. This condition ensures that detected obstacles are not forgotten just because stereoscopy is no longer able to compute the obstacle's 3-D position. This allows for instance tracking an obstacle even when it is too close to the robot to be sensed by both left and right cameras.

If the obstacle is found for several frames, then the system converges towards the fixed point $y|_{\dot{y}_a \to 0} = \gamma \mathrm{R}(\mathbf{p}'_\mathbf{a})$. This means that evidence grows faster when the level of activity in the perceptual shared medium R is higher, i.e., when other p-ants also detect the presence of the obstacle. Finally, if $y_a > \alpha$, where α is a confidence threshold, then the p-ant is confident enough on the existence of an obstacle in its position, and so it can be reported to the action selection process, i.e., appended to O.

To append an obstacle to O, it is necessary to project the corresponding 3-D point, which is in the camera frame of reference, $\{x_c, y_c, z_c\}$, to the ground-plane frame of reference, $\{x_g, y_g, z_g\}$ (see Sect. 3.2). This way the obstacle is represented in the ground coordinate system, and so it affects the action selection process in a robot's pose invariant way.

Figure 6 illustrates a situation where: (a) p-ants are guided by the action selection process; (b) they find an obstacle and a diffusion process is initiated; (c), (d) evidence is accumulated for some frames; (e) and finally the obstacle is reported to the action selection process.

A probabilistic framework could also be considered for evidence accumulation, provided that a better noise description would be available. Nevertheless, the presented dynamical system implements a leaky-integrator and consequently a sound mechanism in terms of neural support. Moreover, allowing the collective to influence the individual decision is also sound as it implements a sort of lateral influence between neural structures. In fact, the use of stigmergy allows this to happen without the explicit modelling of a large set of connections.

4.3.4 Motion compensation

Motion compensation is performed in the first iteration of each frame by first transforming the 3-D position of every p-ant according to the robot motion estimate matrix M (see Sect. 3.3.1). Then, this transformed 3-D position is projected onto visual input L according to projection matrix P. The result is the p-ant's motion compensated position in L. See Fig. 7 and Fig. 8 for an illustration of the motion compensation used to update the spatial memory.

Allowing p-ants to be updated when out of the visual field of view to maintain spatial memories is important so as to avoid that the action selection process operates with a myopic view of the environment. Moreover, it allows the system to reduce perceptual requirements when revisiting a given environment (refer to our previous work on simulation (Santana

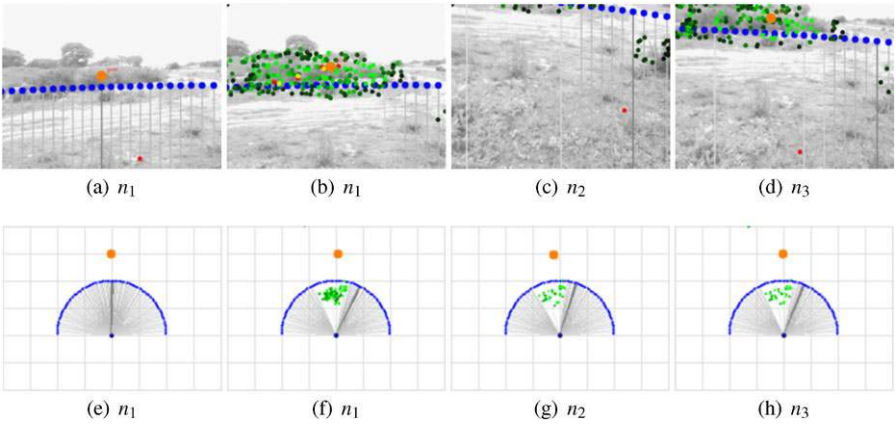


Fig. 7 Motion compensation example. (**a**) The robot moves toward the target. (**b**) A set of p-ants detect a large obstacle. (**c**) Some frames later the robot is tilted and these p-ants are obligated to track the obstacle outside the sensor's field of view. (**d**) When the tilt angle is zeroed these p-ants are once again in the sensor's field of view and consequently projected onto L. At that time, p-ants update their 3-D position and confirm the presence of the obstacle. The smaller number of p-ants, when compared to (**b**), is a consequence of the probabilistic removal of p-ants when they leave the field of view. The figure at the bottom of each image is a body-centred representation of the detected obstacles and linear trajectories. The darker the linear trajectory the higher its utility, as casted by the action selection process. It is possible to see the stability of the representation independently of the robot's current pose. Videos with situations as the one depicted in this figure are available as online supplementary material

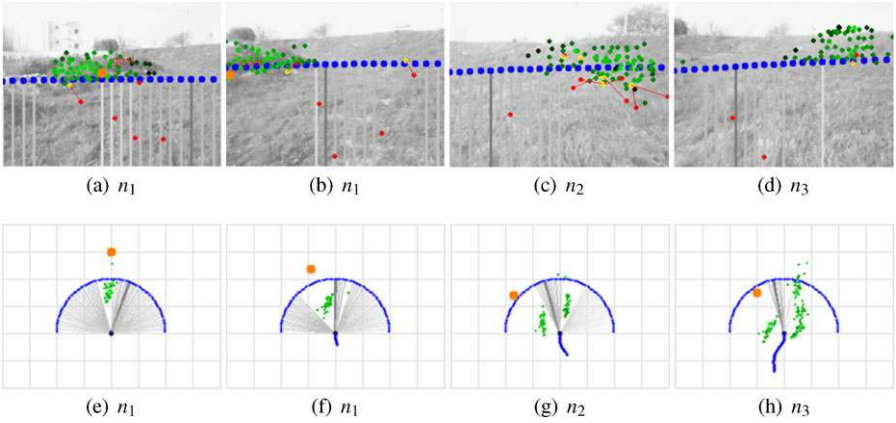


Fig. 8 Spatial memory operation example (same colour code as in Fig. 7). The example given refers to a situation where the robot has to move towards a goal location 30 m ahead of its start position. To reach its goal, the robot circumnavigates the large obstacle area visible in (**a**) and (**b**), which as a consequence of the robot's motion leaves the sensor's field of view in (**c**) and (**d**). Without the spatial memory, the action selection process would not be able to take into account the obstacles outside the sensor's field of view, and consequently it could lead the robot towards a collision. Videos with situations as the one depicted in this figure are available as online supplementary material

and Correia 2010) for experimental results showing this benefit). The fact that not all p-ants leaving the robot's visual field of view are kept in the system and that p-ants have a limited life span permits execution time to be kept low (see Sect. 5.2). In conclusion, the ability of

p-ants to operate as elements of the spatial working memory does not conflict with the idea of minimal perception.

5 Experimental results

5.1 Experimental setup

Small Vision System (SVS) (Konolige 1997; Konolige and Beymer 2007) and OpenCV (Bradski and Kaehler 2008) were used for stereo computation and other low-level computer vision routines on the 320×240 input images, respectively. Stereo computation uses an area-based L1 norm (absolute difference) correlation method, operating over Laplacian Of Gaussian (LOG) transformed images. The result is interpolated to a precision of $1/4$ pixel and the correlation window size is 11×11. To increase the amount of information available in the point cloud, the disparity calculation is carried out at the original resolution, and also on images reduced by $1/2$. With this multi-scale approach, the extra disparity information is used to fill in dropouts in the original disparity calculation.

SVS also provides a set of standard filters to reject 3-D points that are potentially erroneous at the cost of reducing too much the density of the 3-D point cloud in poorly textured environments. Briefly, according to a threshold f_c a confidence filter eliminates stereo matches that have a low probability of success due to lack of image texture. A uniqueness filter performs a consistency check to ensure that the minimum correlation value must be lower than all other match values by a threshold f_u. Finally, a speckle filter eliminates small disparity regions that are not correct by imposing a threshold f_s on the minimum region size. The three filters are used with $f_c = 12$, $f_u = 10$, $f_s = 400$.

The following describes the model's parametrisation used in the experiments. With $\gamma = 1, \tau = 0.8, \alpha = 0.6$, the dynamics of the confidence level in (4) is such that spurious noise could be effectively filtered out, without imposing a compromising latency to obstacles registration. Assuming a 10 Hz frame rate, with a maximum number of iterations per frame $i_{\max} = 10$, and a p-ant's top energy $\rho = 5000$, obstacles are maintained in the robot's spatial memory for about 50 s. This timing is sufficient for local obstacle avoidance. The strong energy decay occurring when local search is engaged, i.e. $\upsilon = 100$, guarantees that p-ants in that situation are removed from the system before a new frame is acquired. To produce an adequate speed-accuracy trade-off, the stochastic motion control parameters from (2), λ_1 and λ_2, have been empirically defined as 0.3 and 24, respectively. The number of times a p-ant is allowed to clone itself is $r_{\max} = 5$. The cardinality of the robot's action repertoire, k, is 80, which is sufficient to ensure navigation in cluttered environments. The control parameters to determine when an obstacle is too cluttered with p-ants, η and ζ, have been set to 40 and 250, respectively. This ensures a good coverage of the obstacle with p-ants without too much overlap. The control parameter to avoid migration of p-ants between obstacles when performing a local search, l, has been set to 20. Finally, the probability of removing a p-ant when it leaves the visual input is $\psi = 0.4$. This value establishes an adequate trade-off between computational efficiency and the density of the robot's spatial memory.

5.2 Experiments

To validate the proposed model the robot was asked to perform at a speed of $s = 0.5$ ms^{-1} a set of five runs on the off-road environment depicted in Fig. 9. In each run, the robot started from a different initial position and its goal was to reach a location 30 m ahead. Although

(a) (b)

Fig. 9 Experimental site. (**a**) The test site overlaid with approximate paths executed by the robot in two runs. In each run, the robot started from a different position and moved autonomously to reach the specified final position, represented by a circle. (**b**) Situation where the difficult conditions faced by the robot are evidenced: dense and tall vegetation, which often promotes the occurrence of false positives

Table 1 Results summary

Percentage of analysed pixels	0.9 ± 0.6 [%]
Control system computation time	112.1 ± 5.4 [ms]
Swarm computation time (t_s)	1.7 ± 0.6 [ms]
t_s/full analysis computation time	0.37 ± 0.17

mostly planar, the environment includes regions of variable slope. This, in addition to the considerable amount of tall vegetation, results in a considerable proneness for false positives occurrence.

Table 1 summarises the quantitative results obtained from the experiments. The robot reached the goal location in every run by negotiating all obstacles between itself and the goal. The executed paths were smooth and short, i.e., the robot avoided obstacles soon enough and smoothly followed their contour whenever necessary. This means that the system was capable of rapidly detecting obstacles at distance and kept tracking them in memory. Notice that whenever the robot followed the contour of obstacles outside its field of view we know that it is tracking them in memory.

Visual attention refers to the ability of accurately focussing perceptual resources where these are the most needed, according to the action selection process. The fact that the robot exhibits a near-optimal behaviour, i.e., it produces a highly directed motion towards the goal (see Fig. 9), with only 0.9% of analysed pixels on average per frame, is a clear demonstration that the model produces accurate visual attention. Note that accuracy here stands for being able to focus precisely on the regions of the current frame that are the most important for the action selection process.

For the following analysis we split the time spent by the swarm infrastructure from the one spent by p-ants when applying the obstacle detector to the image in their current position. Were the swarm's computation time larger than the one taken by a full image analysis, the benefits of the proposed model would be marginal. This is not the case, as it only requires 37%, on average, of the time spent by the brute force approach.

As the complexity of the obstacle detector grows, the less significant is the computation time of the swarm infrastructure to the overall cost. Note that the obstacle detector used by p-ants is actually the simplest one for a stereoscopic setup, as it only requires the computation

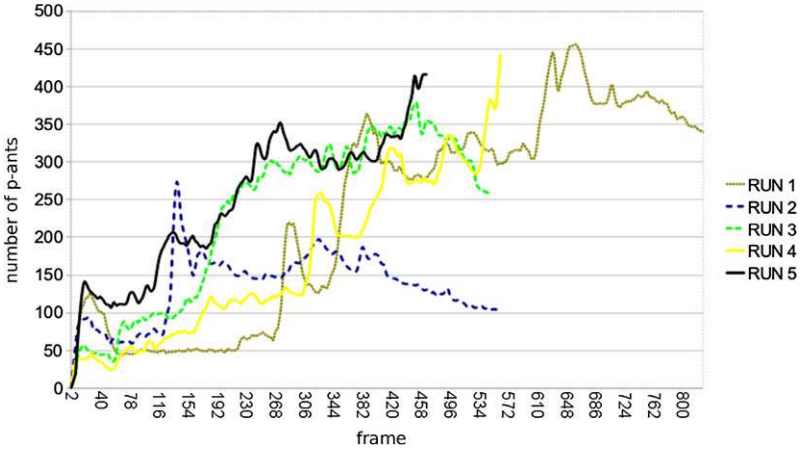


Fig. 10 Context-dependent load balancing results. Plots refer to the total number of p-ants during each run

of a distance to the ground plane. Hence, the reported results can be understood as the baseline performance of the proposed model. A final remark to say that the whole control system operates almost at 10 Hz, which showed to be sufficient to maintain a stable control of the robot at a speed of 0.5 ms^{-1}. This cost includes swarm update, visual odometry, ground-plane estimation and 3-D reconstruction.

One recognised characteristic of self-organised systems is their robustness to varying contexts. In this work, p-ants self-organise to build visual attention and spatial memories as the environment unfolds in the sensor's field of view. A by-product of this process is the adaptation to the context, which can be observed by the number of p-ants along time (see Fig. 10). A raise in the number of p-ants corresponds typically to situations where new obstacles are detected. When these are actually noise, their associated p-ants are eventually removed, resulting in a corresponding drop in their cardinality. A smoother decay in the quantity of p-ants usually refers to the removal of p-ants that have reached zero energy. Thus, the variability of the number of p-ants is a sign of system's adaptation to the environmental context and robot-environment interaction history.

Figure 11 shows the results of a comparison between the proposed model for visual attention and a random sampling of the visual input. From the comparison stands out that in 7% of the tested frames the random policy outperforms the swarm-based policy. However, this percentage corresponds to situations where no obstacles are present in sensor's field of view, meaning that the comparison is made solely on the basis of false positives and consequently is of little value.

One can then conclude that the results clearly show the benefits of the swarm-based solution guided by the action selection process over a baseline random one. With a larger field of view and a higher resolution, i.e., in a situation where the chances of randomly selecting an obstacle pixel are smaller, it is reasonable to assume that the comparison would further highlight the advantages of the proposed model.

Figure 12 shows that without the evidence accumulation mechanism, all false positives caused by the noisy nature of the sensory input are propagated to the action selection process and consequently to the actuators. Conversely, with evidence accumulation the actuators are barely affected by sensor noise. The inertia introduced by the mechanism to remove noise could cause latency in the sensory information processing. However, the proper behaviour

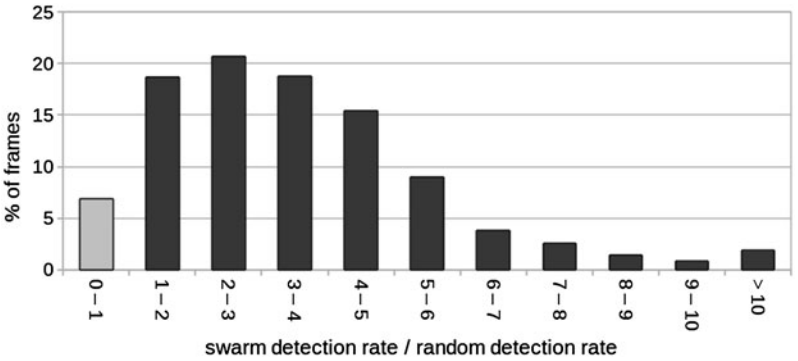


Fig. 11 Comparison between the proposed model and a random policy. The bins of the histogram refer to the ratio between the detection rate of the proposed method and the detection rate of a random policy, over the set of five runs. Detection rate is defined as the ratio of analysed pixels that are reported as obstacles. The higher the detection rate the more focused and effective is the attentive process. For each frame, the random policy operates by randomly sampling as much pixels as those that have been analysed by the swarm-based one in the same frame. That is, this test compares the effectiveness of each method for the same computational cost. The first bin (*light grey*) corresponds to the situations where the random policy outperforms the swarm-based one, 7% overall

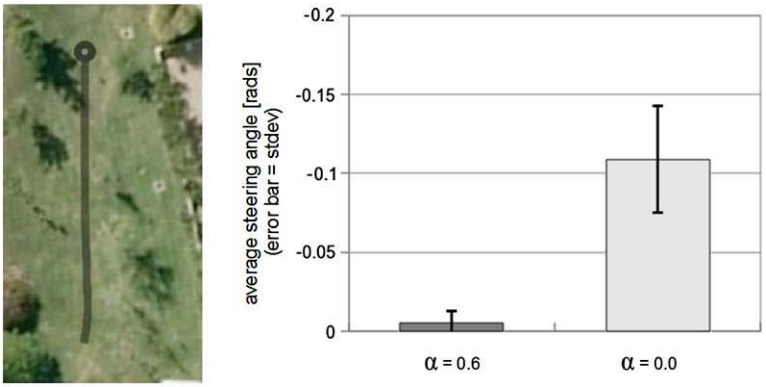


Fig. 12 Comparison between model with ($\alpha = 0.6$) and without ($\alpha = 0.0$) evidence accumulation. The comparison is based on statistics obtained from running each configuration five times on a pre-recorded video sequence. The video sequence was acquired by tele-operating the robot along the straight ahead ≈25 m motion overlaid in (**a**), whose final point is represented by a *circle*. To magnify the differences between both configurations, the obstacle detector's threshold is not set to grow with range, i.e., $\beta = 0$ in (1). The action selection's goal in both configurations is to move forward. No obstacles along the path inhibited the forward motion from being the one with highest utility. Therefore, in the absence of false positives influencing the action selection process, the latter would output a zero steering. The bars and error bars in (**b**) correspond to the average and standard deviation of the steering angle for each configuration

exhibited in the previous experiments allows us to conclude that this is not the case. Note that the evidence accumulation process is affected by the activity of the shared medium R, thus implementing a sort of implicit multi-agent consensus protocol. This is essential to help discriminating actual obstacles from spurious isolated noise, without imposing a slow evidence accumulation dynamics to every situation. Moreover, the use of this form of

implicit communication (*stigmergy*) to reach a sort of collective decision is a feed-forward mechanism without computational overhead.

A set of representative videos with the proposed model's output is available as online supplementary material. For complementary experimental evidence on simulation regarding the key role of some aspects of the model, such as stochastic vs. deterministic behaviour, memory-based vs. pure reflexive operation, parallel vs. a pure greedy deployment of p-ants, please refer to our previous work (Santana and Correia 2010).

6 Discussion

A p-ant represents a covert focus of attention to a given portion of the visual input, upon which a task-specific detector can be applied. Analogously, a natural ant can be seen as a mobile sensor that locally senses the environment for the presence of food items. Natural ants reinforce pheromone trails leading to fruitful locations so as to promote the rate at which food items are returned to the nest. P-ants also use a form of pheromone (activity in shared medium R) to recruit other p-ants to the location of detected obstacles. A negative feedback mechanism is included under the form of a too high pheromone concentration repelling p-ants. This guarantees an adequate number of p-ants to aggregate around a given obstacle.

Patterns in human brain may break when the visual input changes, but rapidly re-emerge in order to track the dynamics of the environment. Ant colonies are also known to be endowed with similar resilience in what regards tracking environmental changes. Thus, this ability of ant colonies can be used as metaphor to model tracking behaviour in parallel visual processes. Tracking is the action of updating the estimate of a given obstacle's position in the visual input, which can change, for instance, due to motion. In general this requires the tracking entity to perform a local search with the goal of re-detecting the obstacle. Accordingly, if a p-ant determines that it has lost the obstacle it has been focused on, it initiates a local search. As their natural counterparts, p-ants also exploit other p-ants pheromones to boost this search. In addition, p-ants are also sensitive to the robot's motion, allowing them to perform the respective compensation.

This ability of tracking and motion compensation endows the visual system with an intrinsically parallel and active spatial map of the environment. That is, each obstacle is represented by a set of p-ants that actively seek to maintain contact with it.

It is largely recognised that the amplification of random fluctuations is essential to allow the emergence of novel situations in foraging tasks, and in self-organised systems in general (Bonabeau et al. 1999). In this line, p-ants' behaviour also encompasses a probabilistic component. We add the following advantages of using randomness in search tasks. The goal of explicitly using random fluctuations allows p-ants to actively search the surroundings with complex search patterns, without their explicit coding. Moreover, the use of stochastic behaviour has the additional advantage of naturally handling noisy data. A deterministic program would require an extensive exception handling facility, which tends to be error-prone. In short, random fluctuations are essential to foster both cheap and robust design. In fact, the variability in neural response to identical stimulus has supporting evidence to be key in explaining the nearly optimal human brain performance, in a Bayesian sense (Ma et al. 2006). Refer to our previous work on simulation (Santana and Correia 2010) for experimental evidence supporting the importance of random fluctuations in the proposed model.

Neurophysiological studies reveal that when human subjects are requested to produce fast response times, in detriment of accuracy, an increase in the baseline activity is observed

in a network of brain areas related to decision, response preparation, and execution, in opposition to sensory related ones (van Veen et al. 2008). The raise in baseline activity allows reaching a decision threshold faster. In our model this is implemented by constraining the time available, i.e., maximum number of iterations per frame, for the action selection process to reach a conclusion. Until the time limit is reached, both action and perceptual processes act the same way for each possible speed-accuracy context. We speculate that this is only possible because the act of perceiving has the goal of pushing action selection forward, while being led by it. By using a parallel approach, our model allows perception to robustly track a good solution right from its onset and without the need of explicit context awareness. This allows both perceptual and action processes to progressively unfold by analysing best possible actions in decreasing order. The result is a good solution whatever the speed-accuracy trade-off. Thus, context modulation is emergent through the decision process, rather than from an explicit context-dependent modulatory signal operating on the perceptual process.

As for the perception-action coupling aspect, the parallel visual process herein proposed also maps seamlessly on the current understanding about the way covert attention operates. First, the use of a parallel system follows the idea of multiple loci of attention (Pylyshyn and Storm 1988; Doran et al. 2009). Second, being parallel, the approach robustly handles the speed-accuracy trade-off (van Veen et al. 2008). Finally, the proposed model includes a natural way of integrating noise, which is known to be very important to achieve near-optimal brain operation (Ma et al. 2006).

From an engineering standpoint, it is useful to compare the proposed model with particle filters as employed for the problem of simultaneous localisation and mapping (SLAM) (Thrun et al. 2005). In general, a particle in SLAM is a sample of the joint distribution robot/map, where a map is composed by a set of Extended Kalman Filters (EKF), each associated to a given landmark of the environment. Conversely, a p-ant is associated to a single landmark, and so it is more related to the role played by each EKF than with the particles themselves. Differently from EKFs in SLAM, p-ants perform an active search in addition to tracking, and do it guided by the action selection process. Moreover, while p-ants interact to improve their individual performance, EKFs in particles are independent.

P-ants do not contribute to the estimation of the localisation of the robot, as particles do. Local navigation being the focus of the proposed model, the localisation problem is handled solely with dead-reckoning. While SLAM methods could deliver both localisation and mapping, their high computational and memory cost makes them unsuited for parsimonious local navigation. For a more comprehensive discussion refer to our previous work (Santana and Correia 2010).

7 Conclusions

The first embodied realisation of swarm cognition was proposed and validated on the local navigation problem for vision-based off-road robots. Although simulations are recognised as important tools to validate minimalist cognition models (Slocum et al. 2000), we believe that only when facing the burdens of the real world, i.e., uncertainty, a model is capable of truly exhibiting its scalability. In this sense, this report is an important contribution to the demonstration of swarm cognition as a promising approach to the synthesis of cognitive systems.

The model considers that perceptual and action selection processes operate in an intricate way to focus attention and consequently reduce the processed image area. The underlying idea of the model is that perception should focus on the regions of the visual input where

detected objects most impact the action selection process. This helps action selection to rapidly stabilise its output. Motivated by the multiple covert attention hypothesis (Pylyshyn and Storm 1988; Doran et al. 2009), covert visual attention is modelled with swarms of simple virtual agents, named p-ants, based on the social insects foraging metaphor. Basically, p-ants perform local covert visual attention loops, whereas the self-organised collective behaviour maintains a robust global spatio-temporal coherence.

The benefits of these properties are confirmed by experimental results obtained with a physical all-terrain robot performing local navigation and equipped with a stereoscopic vision sensor. Concretely, the perceptual process is showed to be capable of providing the action selection process with sufficient information for a proper decision making with only 0.9% of visual input being analysed. This accuracy of the focus of attention is credited to: (1) the ability of action selection to guide perception on a by-need basis, as verified against a random policy; and (2) the self-organising principles exhibited by the parallel covert visual attention process, which provides the system with robustness and consequently the ability of maintaining an adequate focus of attention across environments. Spatial memory, another relevant asset of a cognitive system, was also shown to self-organise in the proposed model. The evidence accumulation mechanism added to the model revealed to be essential on the reduction of the sensitivity to false positives, which due to their high quantity would impair cognitive behaviour.

On the one hand, these results show the feasibility and advantages of relying on self-organisation to synthesise robust embodied cognitive systems, by recurring to the social insects metaphor. On the other hand, they show that with such a model, autonomous robots can considerably reduce the cost of perception, which in turn fosters real-time performance, computational parsimony, and energetic efficiency. This is an important asset to enable miniaturisation, long-lasting operation, and higher robot speed. The proposed model is a systematic method to exploit action selection information to maintain an accurate focus of perceptual attention. It allows the robot designer to exploit the most adequate detector for the task at hand without having to customise the method to cope with the computational requirements, and consequently to lose some of its detection capabilities.

As future work, we plan to devise a parallel implementation of the model in dedicated hardware, such as GPUs, thus fully exploiting the advantages of a multi-agent design. In addition, we intend to generalise the model so that it can be applied to other perception tasks.

Acknowledgements We would like to thank the fruitful discussions with Magno Guedes and Nelson Alves, as well as their support to the field experiments. We also thank IntRoSys, S.A. for the availability of the Ares robot.

References

Agrawal, M., & Konolige, K. (2006). Real-time localization in outdoor environments using stereo vision and inexpensive GPS. In *Proceedings of the 18th international conference on pattern recognition (ICPR)* (pp. 1063–1068). Los Alamitos: IEEE Comput. Soc.

Aloimonos, J., Weiss, I., & Bandyopadhyay, A. (1988). Active vision. *International Journal of Computer Vision*, *1*(4), 333–356.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.

Antón-Canalís, L., Hernández-Tejera, M., & Sánchez-Nielsen, E. (2006). Particle swarms as video sequence inhabitants for object tracking in computer vision. In *Proceedings of the sixth international conference on intelligent systems design and applications (ISDA)* (pp. 604–609). Los Alamitos: IEEE Comput. Soc.

Ashby, W. R. (1952). *Design for a brain*. London: Chapman and Hall.

Bagdanov, A. D., Bimbo, A. D., Nunziati, W., & Pernici, F. (2006). A reinforcement learning approach to active camera foveation. In *Proceedings of the 4th ACM international workshop on video surveillance and sensor networks* (pp. 179–186). New York: ACM.

Bajcsy, R. (1988). Active perception. *Proceedings of the IEEE*, *76*(8), 996–1005.

Balkenius, C., Eriksson, A. P., & Astrom, K. (2004). Learning in visual attention. In *Proceedings of the international conference on pattern recognition (ICPR), workshop on learning for adaptable visual systems (LAVS 2004)* (Vol. 4). Los Alamitos: IEEE Comput. Soc.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, *48*(1), 57–86.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723–767.

Bedau, M. A. (2003). Artificial life: organization, adaptation and complexity from the bottom up. *Trends in cognitive sciences*, *7*(11), 505–512.

Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, *72*(1–2), 173–215.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, *11*(4), 209–243.

Behringer, R., & Muller, N. (1998). Autonomous road vehicle guidance from autobahnen to narrow curves. *IEEE Transactions on Robotics and Automation*, *14*(5), 810–815.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems*. New York: Oxford University Press.

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: computer vision with the OpenCV library*. Sebastopol: O'Reilly Media, Inc.

Broggi, A., & Cattani, S. (2006). An agent based evolutionary approach to path detection for off-road vehicle guidance. *Pattern Recognition Letters*, *27*(11), 1164–1173.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*(1), 139–159.

Chialvo, D. R., & Millonas, M. M. (1995). How swarms build cognitive maps. In L. Steels (Ed.), *NATO ASI series: Vol. 144. The biology and technology of intelligent autonomous agents* (pp. 439–450).

Choe, Y., Yang, H. F., & Misra, N. (2008). Motor system's role in grounding, receptive field development, and shape recognition. In *Proceedings of the 7th international conference on development and learning (ICDL)* (pp. 67–72). Los Alamitos: IEEE Comput. Soc.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.

Couzin, I. (2009). Collective cognition in animal groups. *Trends in Cognitive Sciences*, *13*(1), 36–43.

de Croon, G., & Postma, E. O. (2007). Sensory-motor coordination in object detection. In *Proceedings of the IEEE symposium on artificial life (CI-ALIFE)* (pp. 147–154). Los Alamitos: IEEE CIS.

de Haan, B., Morgan, P. S., & Rorden, C. (2008). Covert orienting of attention and overt eye movements activate identical brain regions. *Brain research*, *1204*, 102–111.

Deneubourg, J.-L., Goss, S., Franks, N., & Pasteels, J. M. (1989). The blind leading the blind: modeling chemically mediated army ant raid patterns. *Journal of Insect Behavior*, *2*(5), 719–725.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, *3*, 357–370.

Dickmanns, E. D., Mysliwetz, B., & Christians, T. (1990). An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles. *IEEE Transactions on Systems, Man and Cybernetics*, *20*(6), 1273–1284.

Doran, M. M., Hoffman, J. E., & Scholl, B. J. (2009). The role of eye fixations in concentration and amplification effects during multiple object tracking. *Visual Cognition*, *17*(4), 574–597.

Egner, T., Monti, J. M. P., Trittschuh, E. H., Wieneke, C. A., Hirsch, J., & Mesulam, M. (2008). Neural integration of top-down spatial and feature-based information in visual search. *Journal of Neuroscience*, *28*(24), 6141.

Eimer, M., Forster, B., Velzen, J. V., & Prabhu, G. (2005). Covert manual response preparation triggers attentional shifts: Erp evidence for the premotor theory of attention. *Neuropsychologia*, *43*(6), 957–966.

Fend, M., Bovet, S., Yokoi, H., & Pfeifer, R. (2003). An active artificial whisker array for texture discrimination. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 1044–1049). New York: IEEE Press.

Fermuller, C., & Aloimonos, Y. (1995). Vision and action. *Image and Vision Computing*, *13*, 725–755.

Floreano, D., Toshifumi, K., Marocco, D., & Sauser, E. (2004). Coevolution of active vision and feature selection. *Biological Cybernetics*, *90*(3), 218–228.

Floreano, D., Durr, P., & Mattiussi, C. (2008). Neuroevolution: from architectures to learning. *Evolutionary Intelligence*, *1*(1), 47–62.

Franks, N. R. (1989). Army ants: a collective intelligence. *American Scientist*, *77*(2), 138–145.

Frintrop, S., Backer, G., & Rome, E. (2005). Goal-directed search with a top-down modulated computational attention system. In *Lecture notes on computer science: Vol. 3663. Proceedings of the DAGM 2005* (pp. 117–124). Berlin: Springer.

Gerstner, W., & Kistler, W. (2002). *Spiking neuron models: single neurons, populations, plasticity*. Cambridge: Cambridge University Press.

Gibson, J. (1979). *The ecological approach to visual perception*. Hillsdale: Erlbaum.

Goodale, M. A. (2008). Action without perception in human vision. *Cognitive Neuropsychology*, *25*(7), 891–919.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.

Hernandez, D., Cabrera, J., Naranjo, A., Dominguez, A., & Isern, J. (2007). Gaze control in a multiple-task active-vision system. In *Proceedings of the 5th international conference on computer vision systems (ICVS)*, Bielefeld, Germany, March 2007. Applied Computer Science Group.

Hou, X., & Zhang, L. (2007). Saliency detection: a spectral residual approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1–8). Los Alamitos: IEEE Comput. Soc.

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9*(5), 1–18.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews, Neuroscience*, *2*, 1–10.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 1254–1259.

Kelly, A., & Stentz, A. (1998). Rough terrain autonomous mobility—part 2: an active vision, predictive control approach. *Autonomous Robots*, *5*(2), 163–198.

Kim, D., & Moeller, R. (2006). Passive sensing and active sensing of a biomimetic whisker. In *Proceedings of the international conference on the simulation and synthesis of living systems (ALife X)* (pp. 282–288). Cambridge: The MIT Press.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227.

Konolige, K. (1997). Small vision systems: hardware and implementation. In *Proceedings of the international symposium on robotics research (ISRR)* (pp. 111–116). London: Springer.

Konolige, K., & Beymer, D. (2007). *SRI small vision system users manual*, May 2007.

Kwok, C., & Fox, D. (2004). Reinforcement learning for sensing strategies. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 3158–3163). New York: IEEE Press.

Land, M. (1999). Motion and vision: why animals move their eyes. *Journal of Computational Physiology A*, *185*, 341–352.

Liu, J., Tang, Y., & Cao, Y. (1997). An evolutionary autonomous agents approach to image feature extraction. *IEEE Transactions on Evolutionary Computation*, *1*(2), 141–158.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438.

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. New York: Henry Holt and Co., Inc.

Marshall, J. A. R., & Franks, N. R. (2009). Colony-level cognition. *Current Biology*, *19*(10), 395–396.

Marshall, J. A. R., Bogacz, R., Dornhaus, A., Planqué, R., Kovacs, T., & Franks, N. R. (2009). On optimal decision-making in brains and social insect colonies. *Journal of the Royal Society Interface*, *6*(40), 1065–1074.

Matthies, L. (1989). *Dynamic stereo vision*. Ph.D. thesis, School of Computer Science, Carnegie Mellon University.

Mazouzi, S., Guessoum, Z., Michel, F., & Batouche, M. (2007). A multi-agent approach for range image segmentation. In *LNAI: Vol. 4696*. *Proceedings of the 5th international Central and Eastern European conference on multi-agent systems and applications (CEEMAS)* (pp. 1–10). Berlin: Springer.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. London: Oxford University Press.

Minsky, M. (1988). *The society of mind*. New York: Simon & Schuster.

Mobahi, H., Ahmadabadi, M. N., & Araabi, B. N. (2006). Swarm contours: a fast self-organization approach for snake initialization. *Complexity*, *12*(1), 41–52.

Mossio, M., & Taraborelli, D. (2008). Action-dependent perceptual invariants: from ecological to sensorimotor approaches. *Consciousness and Cognition*, *17*(4), 1324–1340.

Nabbe, B., & Hebert, M. (2003). Where and when to look: how to extend the myopic planning horizon. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 920–927). New York: IEEE Press.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205–231.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621.

Nolfi, S. (2005). Categories formation in self-organizing embodied agents. In *Handbook of categorization in cognitive science* (pp. 869–889). Amsterdam: Elsevier.

Nolfi, S., & Marocco, D. (2002). Active perception: a sensorimotor account of object categorization. In *Proceedings of the 7th international conference on simulation of adaptive behavior (SAB)* (pp. 266–271), Edinburgh, August, 2002. Cambridge: MIT Press.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939–1031.

Owechko, Y., & Medasani, S. (2005). A swarm-based volition/attention framework for object recognition. In *Proceedings of the IEEE computer vision and pattern recognition workshop (CVPRW)* (pp. 91–98). Los Alamitos: IEEE Comput. Soc.

Palmer, S. E. (1999). *Vision science: photons to phenomenology*. Cambridge: MIT Press.

Passino, K. M., Seeley, T. D., & Visscher, P. K. (2008). Swarm cognition in honey bees. *Behavioral Ecology and Sociobiology*, *62*(3), 401–414.

Patel, K., Macklem, W., Thrun, S., & Montemerlo, M. (2005). Active sensing for high-speed offroad driving. In *Proceedings of the IEEE international conference robotics and automation (ICRA)* (pp. 3162–3168). New York: IEEE Press.

Pfeifer, R., & Bongard, J. C. (2006). *How the body shapes the way we think—a new view of intelligence*. Cambridge: MIT Press.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge: MIT Press.

Poli, R., & Valli, G. (1993). Neural inhabitants of MR and echo images segment cardiac structures. In *Proceedings of the computers in cardiology* (pp. 193–196). Los Alamitos: IEEE Comput. Soc.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179.

Ramos, V., & Almeida, F. (2000). Artificial ant colonies in digital image habitats—a mass behavior effect study on pattern recognition. In *Proceedings of the 2n international workshop on ant algorithms—from ant colonies to artificial ants (ANTS)* (pp. 113–116), Belgium.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two choice reaction time. *Psychological Review*, *111*, 333–367.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1A), 31–40.

Rothkopf, C., Ballard, D., & Hayhoe, M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14–16), 1–20.

Santana, P., & Correia, L. (2008). Improving visual odometry by removing outliers in optic flow. In *Proceedings of the 8th conference on autonomous robot systems and competitions*.

Santana, P., & Correia, L. (2010). A swarm cognition realization of attention, action selection and spatial memory. *Adaptive Behavior*, *18*(5), 428–447.

Santana, P., Cândido, C., Santos, P., Almeida, L., Correia, L., & Barata, J. (2008). The Ares robot: case study of an affordable service robot. In *Proceedings of the European robotics symposium, 2008 (EUROS)* (pp. 33–42). Berlin: Springer.

Santana, P., Guedes, M., Correia, L., & Barata, J. (2009). Saliency-based obstacle detection and ground-plane estimation for off-road vehicles. In *Proceedings of the international conference on computer vision systems (ICVS)* (pp. 275–284). Berlin: Springer.

Santana, P., Guedes, M., Correia, L., & Barata, J. (2010). A saliency-based solution for robust off-road obstacle detection. In *Proceedings of the international conference on robotics and automation (ICRA)* (pp. 3096–3101). New York: IEEE Press.

Scheier, C., Pfeifer, R., & Kuniyoshi, Y. (1998). Embedded neural networks: exploiting constraints. *Neural Networks*, *11*, 1551–1596.

Slocum, A., Downey, D., & Beer, R. (2000). Further experiments in the evolution of minimally cognitive behavior: from perceiving affordances to selective attention. In *Proceedings of the international conference on simulation of adaptive behavior (SAB)* (pp. 430–439). Cambridge: MIT Press.

Sporns, O., & Lungarella, M. (2006). Evolving coordinated behavior by maximizing information structure. In *Proceedings of ALife X* (pp. 3–7). Cambridge: MIT Press.

Sprague, N., Ballard, D., & Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Transactions on Applied Perception*, 4(2).

Sukthankar, R., Pomerleau, D., & Thorpe, C. (1993). Panacea: an active sensor controller for the alvinn autonomous driving system. In *Proceedings of international symposium on robotics research (ISRR)*. London: Springer.

Suzuki, M., & Floreano, D. (2006). Evolutionary active vision toward three dimensional landmark-navigation. In *Proceedings of the 9th international conference on the simulation of adaptive behavior (SAB)* (pp. 263–273). Cambridge: MIT Press.

Tessier, C., Berducat, M., Chapuis, R., Chausse, F., & Cemagref, A. (2007). A new landmark and sensor selection method for vehicle localization and guidance. In *Proceedings of the 2007 IEEE intelligent vehicles symposium* (pp. 123–129). New York: IEEE Press.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.

Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic robotics (intelligent robotics and autonomous agents)*. Cambridge: MIT Press.

Torralba, A., Murphy, K. P., Freeman, W. T., & Rubin, M. A. (2003). Context-based vision system for place and object recognition. In *Proceedings of the IEEE international conference on computer vision (ICCV)* (pp. 273–280). Los Alamitos: IEEE Comput. Soc.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.

Trianni, V., & Tuci, E. (2010). Swarm cognition and artificial life. In *LNCS/LNAI: Vols. 5777, 5778. Proceedings of the European conference on artificial life (ECAL)*. Berlin: Springer.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. Kei, Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial intelligence*, *78*(1–2), 507–545.

Urmson, C., Ragusa, C., Ray, D., Anhalt, J., Bartz, D., Galatali, T., Gutierrez, A., Johnston, J., Harbaugh, S., Kato, H., Messner, W., Miller, N., Peterson, K., Smith, B., Snider, J., Spiker, S., Ziglar, J., Whittaker, W., Clark, M., Koon, P., Mosher, A., & Struble, J. (2006). A robust approach to high-speed navigation for unrehearsed desert terrain. *Journal of Field Robotics*, *23*(8), 467–508.

van der Mark, W., Heuvel, J., & Groen, F. (2007). Stereo based obstacle detection with uncertainty in rough terrain. In *Proceedings of the IEEE intelligent vehicles symposium* (pp. 1005–1012). New York: IEEE Press.

van Veen, V., Krug, M. K., & Carter, C. S. (2008). The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *Journal of Cognitive Neuroscience*, *20*(11), 1952–1965.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge: MIT Press/Bradford Books.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*, 1395–1407.

Wolfe, J. M. (1994). Guided search 2. 0. a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.

Zhang, X., Hu, W., Maybank, S., Li, X., & Zhu, M. (2008). Sequential particle swarm optimization for visual tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1–8). Los Alamitos: IEEE Comput. Soc.