Swarms for Robot Vision: The Case of Adaptive Visual Trail Detection and Tracking

Pedro Santana[1,2], Ricardo Mendonça[1], Luís Correia[2] and José Barata[1]

[1]UNINOVA, Universidade Nova de Lisboa, Portugal
[2]LabMAg, Universidade de Lisboa, Portugal
pfs@uninova.pt

Abstract

Previous work has shown that a pheromone-based visual saliency map can be computed by a swarm of simple agents inhabiting the robot's input image. It was also shown that, with a proper set of behaviours controlling the agents, the saliency map can be used to localise trails present in the robot's visual field. Under the assumption that the robot starts its autonomous operation already on the trail, this paper extends that work by enabling the agents to learn online an appearance model of the trail. The learned model is then used to increase the level of pheromone deployed in the regions of the input image that are more probable of belonging to the trail. This is motivated by the well-known importance that a priori object knowledge has to improve visual search. The outcome of this extension is a self-organising behaviour capable of detecting trails in 98% of the evaluated situations, outperforming the original work. The agents being simple their computation is fast, resulting in a $12\,\mathrm{Hz}$ performance. Thus, by introducing a parsimonious learning mechanism, this paper contributes to increase robustness of swarm-based robot vision systems.

1. Introduction

An important sensory modality for autonomous robots is vision. However, the richness of vision comes with the price of complex processing. The complexity inherent to vision calls for fine and contextualised focus of computational resources on the most relevant stimuli obtained from the environment. This process is called visual attention, which has been extensively studied in humans (Oliva and Torralba, 2007; Wolfe et al., ress). By focusing perception: (1) computation, and consequently, energy are more efficiently used; (2) the robot becomes less sensitive to noise and perceptual aliasing; and as a consequence of the previous two, (3) faster robot motion, lower cost, and reduced robot size are enabled.

Models of visual attention typically assume the existence of a sensory-driven bottom-up pre-attentive component (Treisman and Gelade, 1980; Itti et al., 1998), which is modulated by top-down context aware pathways (Tsotsos et al., 1995; Neider and Zelinsky, 2006). The use of top-down modulation is important when bottom-up saliency information is insufficient to focus attention in the presence of distractors. These distractors are other objects or perceptual aliasing in the environment that happen to detach from the background at least as much as the object being sought. However, top-down information (e.g., expected colour and morphology of the object) is quite dependent on the environmental context. As a result, adapting this knowledge is key when facing unstructured environments.

Visual attention ultimately drives 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 focusing particular aspects of the sensory stimuli. This is called covert attention and its modelling is the focus of this work. Studies on human subjects support the hypothesis that multiple covert attention processes co-exist in the brain (Doran et al., 2009).

In previous work (Santana and Correia, 2010, 2011), we have explored the idea of existing multiple covert attention processes to model visual attention on autonomous robots as the product of a self-organising process supported by a set of virtual agents inhabiting the sensorimotor space of the robot. In particular, we have devised a model where the action selection process is used as top-down context knowledge to guide visual obstacle detection. In that work, agents perform local covert visual attention loops, whereas the self-organising collective behaviour maintains global spatio-temporal coherence. In a related research line (Santana et al., 2010), we have shown that a swarm of agents is able to create saliency maps using implicit knowledge about the object being sought. The model was shown to be able to detect and track trails in natural environments. This top-down knowledge was defined in terms of the behaviours controlling the agents. Focusing on the shape of trails, rather than in their photometric appearance, is advantageous given trails variability. However, photometric appearance may be useful to compensate for situations where shape information is not reliable. Due to its variability under different contexts, photometric appearance must be considered under an adaptive framework, capable of being tuned to the specificities of the environment. The current paper addresses this problem by including an adaptive mechanism into the agents compos-

ing the swarms responsible for the localisation and tracking of the trail. Concretely, the output generated by the swarm in previous frames is used to supervise the learning process of a trail's appearance model. In turn, this model is used to modulate the pheromone deployed by the agents, thus helping them concentrate their activity on the image regions whose appearance is more similar to the one of the trail being tracked.

2. System Overview

Typically, object-related a priori knowledge is used by top-down boosting of the set of features (e.g., colour) known beforehand to be more representative of the object being sought. Instead, the object's overall layout, which is a more stable and predictable feature in the case of natural trails, whose local appearance often blends with the background, is used in this work. This type of a priori knowledge is specified indirectly in the proposed model as perception-action rules controlling the behaviour of simple agents inhabiting the robot's visual input. These agents are called p-ants (from perceptual-ants) and represent local covert attention processes. Their self-organising collective behaviour results in a saliency map of the input image, and thus, in a global covert attention process.

Fig. 1 depicts the base model (Santana et al., 2010) of this work. In short, at each new frame $\mathbf{I}$, two conspicuity maps, $\mathbf{C}^C \in [0,1]$ for colour and $\mathbf{C}^I \in [0,1]$ for intensity information, are computed (Santana et al., 2010). The intensity of a pixel in a given conspicuity map signals how much the pixel detaches from the background at several scales (i.e., resolutions), in the scope of a given visual feature. A set of n p-ants is then deployed on each map. These p-ants interact based on the ant-foraging metaphor for several iterations in order to build two pheromone maps, $\mathbf{P}^C \in [0,1]$ and $\mathbf{P}^I \in [0,1]$. The behaviour of these p-ants is designed to exploit some a priori knowledge about typical trails approximate layout. The activation of the pheromone maps is therefore expected to match the trail's location better than the activation of the conspicuity maps, which are only sensory driven. Additionally, by allowing p-ants on a given pheromone map to also affect the other pheromone map, cross-modality influences are implicitly, i.e., through stigmergy (Grassé, 1959), maintained in the system. This increases robustness by allowing p-ants to exploit multiple cues indirectly, in a simple and fast to compute way.

Rather than blending both conspicuity maps to generate the final saliency map $\mathbf{S} \leftarrow \frac{1}{2}\mathbf{C}^I + \frac{1}{2}\mathbf{C}^C$, as typically done (Itti et al., 1998), in this work $\mathbf{S}$ is obtained by blending both pheromone fields, $\mathbf{S} \leftarrow \frac{1}{2}\mathbf{P}^I + \frac{1}{2}\mathbf{P}^C$. This way the saliency map is no longer a result of purely bottom-up sensory-driven process; instead, the bottom-up information is exploited under the context of some a priori knowledge about typical trails approximate layout. The result is a more robust and accurate focus of attention at the cost of a residual computational overhead.

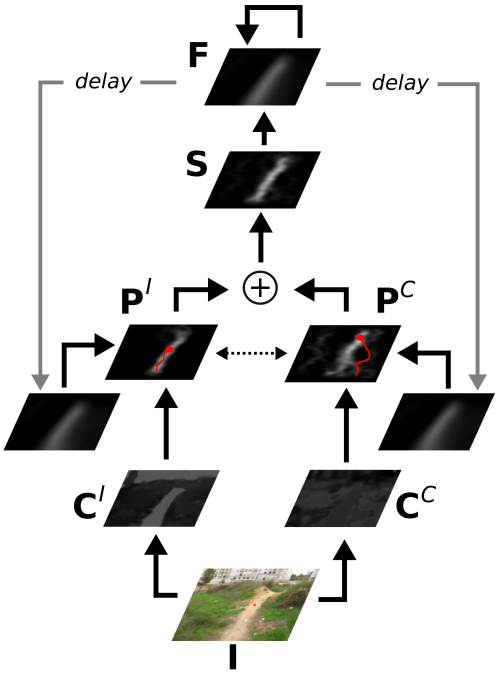


Figure 1: System's operation overview (Santana et al., 2010). The red overlays in both pheromone fields, $\mathbf{P}^C$ and $\mathbf{P}^I$, are two illustrative p-ant paths. Motion compensation aspects are not represented. Note that the brightest region in the neural field, $\mathbf{F}$, correctly corresponds to the trail location in the input image, $\mathbf{I}$.

For across-frames integration of trail location evidence, the final saliency map $\mathbf{S}$ feeds a dynamic neural field, $\mathbf{F} \in [0,1]$, that is, a 2-D lattice of dynamical neurons with Mexican-hat shaped lateral coupling (Amari, 1977). This coupling implements inter-neuron local lateral excitation and long-range inhibition, which helps the neural field on the production of a single focus of attention (Rougier and Vitay, 2006). In order to decouple the dynamics of the neural field from the dynamics of the robot, the projective transformation estimated between frames is applied to the neural field. Finally, the output of the system is given by the current state of the neural field, in which the higher the activation of a given neuron the higher its chances of being associated to a trail's pixel (refer to (Santana et al., 2010) for details on dynamical field processing).

In order to allow p-ants' creation and activity to be affected by history, at the onset of each frame, both pheromone maps are initialised with a small ratio λ of the neural field after being motion compensated, $\mathbf{P}^I \leftarrow \lambda\mathbf{F}, \mathbf{P}^C \leftarrow \lambda\mathbf{F}$. This induces stability and robustness to noise and temporarily mis-behaved conspicuity maps (i.e., unable to properly discern between trail and background in the presence of distractors), as well as it enables across-frames progressive im-

provement.

With the purpose of reducing the effects of strong distractors when tracking the trail, this paper includes into the swarm-based system an adaptive mechanism. The goal is to learn and update in each frame a simple appearance model of the trail, so that p-ants can strengthen the deployment of pheromone on regions whose appearance match the learned one. The result is a stronger stigmergic behaviour around the true location of the trail. Learning occurs by sampling the region of the visual input corresponding to the region of the neural field with highest activity. That is, the model is updated under the assumption that the trail location estimated in the previous frame is correct.

3. Pheromone Maps Computation

This section describes how the two pheromone maps, $\mathbf{P}^I$ and $\mathbf{P}^C$, are built from the two conspicuity maps, $\mathbf{C}^I$ and $\mathbf{C}^C$. For this purpose, a given p-ant, p_m, is created and associated to a given visual feature $m \in \{I, C\}$. The other visual feature is represented by m'. While being iterated for η times, p_m will move on $\mathbf{C}^m$, influenced by the pheromone present in $\mathbf{P}^m$. In the non-adaptive model, while moving, this p-ant deploys pheromone in each position visited in $\mathbf{P}^m$ with a magnitude ϵ_0, and a small portion of ϵ_0, υ, in $\mathbf{P}^{m'}$.

After the iterations for this p-ant, a p-ant associated to the other visual feature, $p_{m'}$, is created and iterated following the same procedure. Afterwards, the two p-ants are removed from the system and the process is repeated n times, meaning that $2n$ p-ants are created and iterated. As it will be shown, the deployed pheromone is a function of p-ants' sensations across their trajectories on their associated conspicuity maps. Hence, it is influenced by the activity occurring in distant regions of the map. This long-range spatial connectivity allows handling the potentially large size of trails in a robust and parsimonious way.

3.1. P-Ant's Creation

The chances of creating a p-ant p_m on a given location $\mathbf{o}_{p_m}$ of the conspicuity map $\mathbf{C}^m$ depends on the level of conspicuity at that location and on the level of pheromone at the same location in the corresponding pheromone map, $\mathbf{P}^m$. Hence, p-ants are progressively and probabilistically deployed where there are more chances of being a trail, under the assumptions that: (1) trails tend to be conspicuous; (2) the trail has been successfully detected in the previous frame (represented by the feedback provided by the delayed neural field state); and (3) that the pheromone accumulated by p-ants deployed in the current frame builds-up mostly around the actual trail's location.

By assuming that trails often start from the bottom of the image, p-ants are deployed with a small randomly selected offset $z \in [0, 0.1 \cdot h]$ of the bottom of the conspicuity map in question, i.e., at row $r \in [h - z, h]$, where h is the height of the map[1]. This random small offset reduces sensitivity to any noise potentially present at the map's boundaries.

In order to determine the column where p_m is deployed, a unidimensional vector $\mathbf{v}^m = (v_0^m, \ldots, v_w^m)$ is first computed. The element v_k^m of $\mathbf{v}^m$ refers to the average conspicuity level of the pixels in a small window centred on column k and with a randomly selected offset with respect to the bottom row of the map, r,

$$v_k^m = \sum_{l,j} \frac{\mathbf{C}^m(l,j)}{\delta_w \cdot \delta_h} \tag{1}$$

where $l \in [k - \delta_w/2, k + \delta_w/2]$, $j \in [r - \delta_h, r]$, $\mathbf{C}^m(l,j)$ returns the conspicuity level in position (l,j), and δ_w and δ_h are the width and the height of the window, respectively. The same windowing process is applied to build a vector for the pheromone field in question, $\mathbf{u}^m = (u_0^m, \ldots, u_w^m)$. Element u_k^m corresponds to the maximum pheromone level found in the window:

$$u_k^m = \max\{\mathbf{P}^m(l,j)\}_{l,j} \tag{2}$$

where $\mathbf{P}^m(l,j) \in [0,1]$ returns the pheromone level in position (l,j). The max operator is employed to benefit those regions where the paths of p-ants overlap more often and consequently where there is a higher consensus on the trail's skeleton position.

Using these two vectors in the following test, which is repeated until it succeeds, the chances of deploying a p-ant in a randomly selected column $z_2 \cdot w$ is as high as the conspicuity and pheromone levels at the deployment region,

$$z_1 < \left(\rho \cdot u_{z_2 \cdot w}^m + (1 - \rho) \cdot v_{z_2 \cdot w}^m\right) \tag{3}$$

where $z_1 \in [0,1]$ and $z_2 \in [0,1]$ are numbers sampled from a uniform distribution each time the test is performed and ρ is a weight factor used to trade-off the influence of both pheromone and conspicuity information. By starting with a small value, ρ_0, and by linearly growing at each iteration by an amount $\Delta\rho$, ρ operates as an adaptive process, compelling the system to move from a conspicuity-driven operation (exploration) to a pheromone-driven operation (refinement/exploitation).

3.2. P-Ant's Execution

Before specifying p-ants behaviours, it is necessary to specify their sensory and action spaces. To reduce both sensitivity to noise and computational cost, the sensory input is defined by 5 coarse receptive fields disposed around the p-ant's current position, $R_1 \ldots R_5$ (see Fig. 2). For a given visual feature m and p-ant's position $\mathbf{o}_{p_m}$,

[1] Rows are indexed in increasing order from the top to the bottom of the map.

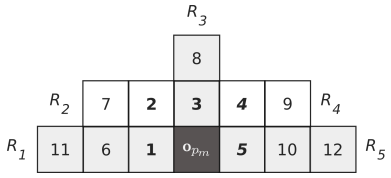


Figure 2: P-ants' sensory and action spaces. Regions surrounding current p-ant's position, $\mathbf{o}_{p_m}$, are segmented into a set of receptive fields, $R_1 = \{1, 6, 11\}, R_2 = \{2, 7\}, R_3 = \{3, 8\}, R_4 = \{4, 9\}, R_5 = \{5, 10, 12\}$, whose composing pixels are numbered as in the figure. If a given action $a \in A$ is selected, then the next p-ant's position will be the closest pixel to the p-ant, represented by the pixels in bold.

$\mathbf{C}^m(R_k, \mathbf{o}_{p_m})$ and $\mathbf{P}^m(R_k, \mathbf{o}_{p_m})$ return the average conspicuity and pheromone levels of the pixels constituting receptive field R_k, respectively. Parameter $\mathbf{o}_{p_m}$ is used to transform the p-ant's centred receptive field onto the map's frame of reference. To refer directly to the pixel-wise conspicuity and pheromone levels at the p-ant's position, $\mathbf{C}^m(\mathbf{o}_{p_m})$ and $\mathbf{P}^m(\mathbf{o}_{p_m})$ are used, respectively. An action $a \in A$ moves the p-ant to one of the 5 neighbour pixels not behind the current p-ant's position. The action space is thus defined by the set $A = \{1, 2, 3, 4, 5\}$ (see Fig. 2).

At each of η iterations, p-ant p_m executes a set of behaviours $B = \{greedy, track, centre, ahead, commit\}$, which independently vote on each possible action in A. Then, the most voted action is the one taken by the p-ant.

In order to allow the system to operate with unstructured trails, these behaviours are simple and make little assumptions regarding the trail's structure. Each behaviour exploits a priori knowledge of trail's shape or appearance so as to make p-ants producing trajectories that approximate the trail's skeleton. For instance, under the assumption that trails are somewhat monotonous structures, p-ants should move under the influence of some inertia. This is implemented by having the $commit$ behaviour voting more strongly on the action that is most similar to the one selected in the previous iteration.

The following describes which regions in the local neighbourhood of the current agent position are selected as its next position by each of the five behaviours, and thus embody top-down knowledge about trails:

1. Greedy: Regions of higher levels of conspicuity, under the assumption that trails are salient in the input image;

2. Track: Regions whose average level of conspicuity is more similar to the average level of conspicuity of the pixels visited by the agent, under the assumption that trails' appearance is somewhat homogeneous;

3. Centre: Regions that maintain the agent equidistant to the boundaries of the trail hypothesis being pursued;

4. Ahead: Upwards regions under the assumption that trails are often vertically elongated;

5. Commit: Region targeted by the motor action at the previous iteration, under the assumption that trails' outline is somewhat monotonous.

Formally, for a given p-ant p_m, behaviours are described as functions that return a vote in the interval $[0, 1]$ for each possible action $a \in A$. As an example consider the *greedy* behaviour (refer to (Santana et al., 2010) for the other behaviours),

$$f_{greedy}(p_m, a) = \mathbf{C}^m(R_a, \mathbf{o}_{p_m}). \quad (4)$$

As it will be shown, all these behaviours contribute to p-ants trajectories that closely represent the trail's skeleton. The absence of an explicit scoring function, which would require a model-based imposition of constraints on the trail's shape, hampers a post-ranking of all deployed p-ants to determine the "best trajectory". Moreover, not all p-ants will be deployed on the trail and so not all are able to follow the actual trail. To overcome these challenges two ingredients of the system are determinant.

The first ingredient comes in the form of positive feedback raising from the amplification of random fluctuations. With additive random fluctuations at p-ants actuation level, those that are deployed off the trail will diverge, whereas p-ants deployed on the trail will converge towards its vanishing point, thanks to the *centre* behaviour. Hence, there will be higher concentrations of pheromone on trail regions. This happens because the presence of the trail tends to be a global constraint which is only felt by the p-ants deployed on it. In a sense, the trail operates as an attractor for the self-organising system.

The second ingredient is the use of stigmergy in the form of pheromone-based interactions. By making p-ants attracted to high pheromone concentration regions, we positively reinforce the difference between diverging and converging p-ants (symmetry breaking). Hence, this second ingredient ensures that, along time, the structure imposed by the presence of the trail on the *centre* behaviour is stronger than the effects of random fluctuations. This effect is magnified by the fact that p-ants are deployed according to the level of pheromone already present in the pheromone maps. Moreover, the fact that robot forward motion tends to make the neural field skew towards the bottom of the image makes regions of higher activity in deep visual field more likely to invoke p-ants. The use of pheromone-based interactions has the additional advantage of overcoming the brittleness of controlling p-ants based on myopic behaviours. The local interruption of a trail, that could inhibit the *centre* behaviour from properly leading the p-ant along the trail, is overcome by having p-ants progressively building a pheromone "bridge" over the interruption thanks to *commit* and *ahead* behaviours.

In order to take these considerations into account, in each

iteration a p-ant p_m selects its action by maximising the following utility function, which incorporates behaviours' votes, pheromone-based interactions, and random fluctuations,

$$a_{p_m} = \arg\max_{a \in A} \left(\sum_{b \in B} \alpha_b f_b(p_m, a) + \mathbf{P}^m(R_a, \mathbf{o}_{p_m}) + \gamma q \right)$$

where: α_b is a user defined weight accounting for the contribution of behaviour $b \in B$; and γ is the weight accounting for stochastic behaviour, being $q \in [0, 1]$ a number sampled from a uniform distribution each time the action is evaluated. To match the randomness magnitude with the scale of the image, which is typically smaller for pixels in upper regions of the image, the weight γ starts with an initial value γ_0 and exponentially decays by a constant factor γ_τ at each iteration.

In case an immediate loop is detected, namely, the p-ant moving recurrently from one pixel to another, then the action for the current iteration is randomly selected. Finally, the p-ant's position $\mathbf{o}_{p_m}$ is updated according to the selected action[2].

4. Adaptive Process

This section describes how (see Section 4.1) and when (see Section 4.2) the appearance model of the trail is learned and updated. To help p-ants disambiguate in situations where the conspicuity information is not sufficient by itself, the learned model is used to promote the deployment of pheromone on regions of the image whose appearance is more likely to belong to the one of the trail (see Section 4.3). To allow learning the model from scratch, some assumptions regarding the initial position of the trail with respect to the robot are made (see Section 4.4).

4.1. Appearance Model

The trail's appearance model of the current frame is a simple colour histogram, $\mathbf{h}$, of the pixels in the region of higher neural field's activity. To reduce sensitivity to illumination effects, the HSV colour space is used. To further reduce this sensitivity, the H(ue) component is described by 12 bins, the S(aturation) component by only 8 bins, and the V(alue) component is discarded altogether.

This frame-wise appearance model is used to update an across-frames appearance reference model,

$$\mathbf{h_{ref}} \leftarrow \Theta(\mathbf{F})\mathbf{h_{ref}} + \big(1 - \Theta(\mathbf{F})\big)\mathbf{h} \tag{5}$$

where $\Theta(\mathbf{F}) = \kappa \cdot \max(\mathbf{F})$ makes the speed the reference model adapts to changes in the trail's appearance proportional to the neural field's maximum activity. This weighted approach allows the appearance model to be updated more strongly when the system is more sure of its output being a correct segmentation of the trail from the background. This assumption follows from the fact that the more stable the pheromone maps' activity across-frames the higher the neural field's maximum. Hence, the presence of distractors is less prone to affect the reference appearance model.

[2]For the sake of completeness, the pseudo-code of the models here described can be found at: http://www.uninova.pt/∼pfs/ecal2011trail.html

4.2. When to learn

To further reduce the chances of learning erroneous appearance models due to the presence of distractors, the appearance reference model, $\mathbf{h_{ref}}$, is only updated with Eq. 5 if the neural field in the current frame reports the trail as being roughly located ($\pm\, 10\%$ of the map's width) at the centre of the image. This is a reasonable heuristic under the assumption that the robot is actively centring itself along the trail in order to follow it.

This learning gating process allows the reference model not to learn the appearance of transient distractors appearing in the sides of the trail. Furthermore, it allows the system to delay the learning phase when the robot does not start centred on the trail.

4.3. Adaptive pheromone deployment

In Santana et al. (2010), p-ants deploy a constant level of pheromone along their paths, ϵ_0 (see Section 3). In this work, instead, a given p-ant p_m deployed in map m deposits a non-fixed level of pheromone,

$$\epsilon = \epsilon_0 + \beta \cdot p(T|V_{p_m}) \tag{6}$$

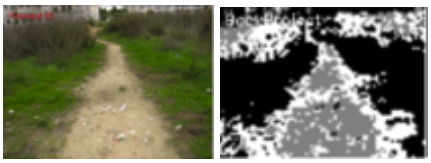
where β is an empirically defined weighting factor and $p(T|V_{p_m})$ is the probability of the p-ant's path, V_{p_m}, to belong to the trail (T).

The probability $p(T|V_{p_m})$ is approximated by the average probability of pixels visited by the p-ant of belonging to the trail. These pixels are represented by the set V_{p_m}, and their individual probabilities are obtained directly from the normalised histogram $\mathbf{h_{ref}}$, according to a technique known as histogram back-projection (see Fig. 3 for typical results). As the experimental results will show, this simple approach suffices to help p-ants tracking the trail.

4.4. The First Frames

The advantages of using learning comes at the price of solving the bootstrapping problem. That is, in the absence of a learned model, the detector has a reduced chance of generating a good output to supervise the learning, which in turn hampers the learning of the model altogether. To solve this problem we start from the assumption that the detector is turned on when the robot is already roughly located on the trail. Therefore, we can assume that in the first frames the trail is centred on the robot's input image.

(a) input image (b) trail probability



(c) input image (d) trail probability

Figure 3: Pixel-wise trail probability (brightness level) for two typical images.

Bearing this in mind, in the first frames, instead of considering the maps' entire width, w, when selecting the deployment column of a newly created p-ant (see Eq. 1 and Eq. 2), the adaptive model assumes that the deployment region is constrained by a band centred on the map and with a frame-wise upper-bounded growing width. Concretely, in the first frame, the width of the band is 10% of w. Then, this width is increased by 0.5% at each new frame until the upper-bound w is reached. From then on, it remains static. At this moment the learned model is sufficiently mature to help the detector tracking the trail.

5. Experimental Results

This section quantifies the improvement the adaptive mechanism brings to the overall method and how well it suits the fast computation requirements imposed by physical robots. In order to measure the performance of the adaptive model, we relied on the same data-set used to evaluate the non-adaptive model (Santana et al., 2010). This data-set consists of 25 colour videos, encompassing a total of 12023 frames with 640×480 resolution, which have been obtained with a hand-held camera[3]. This camera was carried at an approximate speed of $1\,\mathrm{ms}^{-1}$. The trail detector was evaluated on an Intel T4300 2.1 GHz dual core, running Linux. OpenCV was used for low-level routines. To handle the probabilistic nature of the agents behaviour, a set of 5 runs was performed per video. In some of these videos the robot does not start on the trail, which is important to validate the ability of the detector to delay the learning phase.

Performance is measured as the percentage of frames in which the biggest blob of neural field activity above 0.85 (from a maximum of 1) is fully within the trail boundaries. The system parameters related to the adaptive mechanism, κ, β, and ϵ_0, have been empirically set to 0.001, 0.01, and 0.008, respectively. The remainder of the free parameters have been set as in the original model (Santana et al., 2010).

With a success rate of $92.98\,\% \pm 0.16\,\%$ over the 25 videos, the base model already attains an impressive result, operating ≈ 4 times better than a classical saliency model and in situations where previous detectors fail (details in (Santana et al., 2010)). However, a single failure in an embodied setup may result in dramatic consequences. Therefore, full success must be pursued. With the adaptive mechanism, the model reaches a success rate of $97.94\,\% \pm 0.17\,\%$ over the 25 videos, and a $100\,\%$ success rate in 12 of the 25 videos (see Table 1). Conversely, the non-adaptive model obtains a $100\,\%$ success rate only in 6 of the 25 videos. Fig. 4 shows frames from some videos belonging to the 25 video data-set where the non-adaptive model fails to detect the trail, whereas the adaptive one succeeds. Although typically transient, these failures could drive the robot off trail. They usually occur when the assumption that trails are conspicuous structures fails due to the overall scene configuration. Sometimes it also happens that a sudden camera motion is not captured by the motion detection method, resulting in a mismatch between the neural field and the environment.

In terms of computation time, the non-adaptive model runs at 13 Hz whereas the adaptive one runs at 12 Hz. Note that only roughly $8\,\%$ of the computation time refers to swarm-based activity - the remainder includes robot motion estimation, neural field update, and conspicuity maps computation. The conclusion is that the adaptive mechanism, which improves the method's accuracy, adds little computational overhead.

It is important to point out that the dependency of the overall process on an appearance model makes the learning process a critical one. This is reflected on the need for a learning bootstrapping process and for trail's appearance transitions to be smooth. That is, the improvement in performance is obtained at the cost of introducing assumptions, which are, nevertheless, acceptable under a trail tracking framework.

6. Discussion

Rather than static structures, like neurons, agents are better viewed as active information particles that flow and change in the system. Hence, using agents, the design focus is on the process and not so much on its supporting substrate. Additionally, agents being sensorimotor coordinated units can exploit the benefits of active vision (Bajcsy, 1988; Ballard, 1991) at the information processing level. These include the ability of agents to actively select and shape their sensory input so as to increase noise-to-signal ratio and increase their discriminatory power, to augment rotation and scale invariance, and also to exploit sensorimotor history with the purpose of inducing long-range influences and in the limit of improving their own behaviour (Scheier et al.,

[3]The model's output overlaid on these videos is available at: http://www.uninova.pt/∼pfs/ecal2011trail.html

(a) non-adaptive (b) adaptive

(c) non-adaptive (d) adaptive

(e) non-adaptive (f) adaptive

Figure 4: Examples of situations where the adaptive method outperforms the non-adaptive method. The red blobs represent the estimated trail location, which corresponds to the neural field activity above 85% of its maximum. In the adaptive case, besides localising the trail, the red blob is well aligned with its orientation. This means that the system is able to output both position and orientation of the trail.

Video ID	Nr. of frames	Non-adaptive model correct frames [%]	Adaptive model correct frames [%]
1	278	100.00 ± 0.00	100.00 ± 0.00
2	204	100.00 ± 0.00	100.00 ± 0.00
3	422	93.03 ± 0.21	99.15 ± 0.13
4	135	100.00 ± 0.00	100.00 ± 0.00
5	2854	93.90 ± 0.02	97.79 ± 0.03
6	186	97.53 ± 0.29	95.91 ± 0.48
7	121	100.00 ± 0.00	100.00 ± 0.00
8	124	88.06 ± 0.36	100.00 ± 0.00
9	309	98.38 ± 0.32	95.79 ± 0.32
10	147	92.11 ± 0.61	97.41 ± 1.12
11	386	100.00 ± 0.00	100.00 ± 0.00
12	158	88.48 ± 0.28	100.00 ± 0.00
13	134	87.31 ± 0.53	100.00 ± 0.00
14	676	99.14 ± 0.07	98.46 ± 0.17
15	683	91.22 ± 0.10	91.51 ± 0.10
16	770	82.96 ± 0.14	86.83 ± 0.30
17	403	93.90 ± 0.14	94.14 ± 0.83
18	335	86.21 ± 0.13	98.81 ± 0.30
19	230	76.43 ± 0.19	100.00 ± 0.00
20	439	82.92 ± 0.23	95.54 ± 0.38
21	490	93.31 ± 0.09	100.00 ± 0.00
22	230	100.00 ± 0.00	100.00 ± 0.00
23	600	90.10 ± 0.15	100.00 ± 0.00
24	802	95.06 ± 0.07	99.10 ± 0.10
25	907	94.42 ± 0.06	98.10 ± 0.09
Total	12023	$\mathbf{92.98 \pm 0.16}$	$\mathbf{97.94 \pm 0.17}$

Table 1: Comparative results summary.

1998; Nolfi and Marocco, 2002; Beer, 2003; Floreano et al., 2004; Sporns and Lungarella, 2006; Mirolli et al., 2010). Furthermore, the use of multiple agents in the task of modelling cognitive behaviour exploits biological knowledge obtained from similar processes that can be found in Nature.

In our line of research, we have used a model inspired by *swarm cognition* of social insects, whose considerable similarities with brain cognitive function are becoming widely recognised (Passino et al., 2008; Couzin, 2009; Marshall and Franks, 2009; Trianni and Tuci, 2011; Santana and Correia, 2010; Turner, 2011; Trianni et al., 2011). In this work, the ant foraging metaphor previously used was extended with learning capabilities, resulting in a system that can better adapt to different environmental contexts.

The use of learned appearance models to swarm-based object tracking has already been explored in the context of PSO-based models (Zhang et al., 2008). However, our work is the first applying learning to the problem of swarm-based trail detection and tracking. This is an important difference as the appearance of trails change more drastically than the one of typical objects. Furthermore, our model uses the appearance model to modulate pheromone deployment, a concept inexistent in PSO models.

7. Conclusions

This article proposes a model to incorporate an adaptive mechanism into a swarm-based trail detector previously published. The goal of this mechanism is to allow the detector to learn and exploit appearance models of the trail being followed. Experimental results confirmed the ability of the adaptive model to outperform the non-adaptive one, under the assumption that the robot starts its operation already on the trail.

The learned trail's appearance model is used to modulate the swarm operation, rather than, to directly classify the input image as in a convolution-like typical computer vision operation. First, this approach allows the system to exploit synergistically both appearance and shape information, which is pivotal to handle sudden trail's appearance changes. Second, this multi-modal approach allows the use of simple appearance models, i.e., histograms. Third, the appearance model and the behaviours controlling the agents being simple enable a fast to compute system.

With a bottom-up self-organising approach, the model is capable of handling highly unstructured trails without exhibiting a high computational load. In fact, we have shown in previous work (Santana et al., 2010) that the non-adaptive model performs in situations where previous detectors employing classical computer vision techniques would fail. In this work, we have improved the previous model by introducing elementary learning of the photometric appearance of the trail. All this leads us to conclude that swarm-based models are an interesting alternative to classical computer vision techniques. This means that besides contributing with

a useful model to improve off-road robot navigation, this work intends to encourage the artificial life community to employ their bulk of knowledge at the service of the high impact problem of synthesising robust and fast computer vision systems.

An interesting future development would be to expand the learning capabilities to other aspects of the model. An example is the adaptation of the weights controlling how much each agent's behaviour contributes to the overall behaviour. It would also be interesting to learn the behaviours themselves. An additional aspect that might be considered is the emergence of hierarchical organisation among the agents. Finally, the method's ability to deal with strong camera motion must be evaluated on a physical robot embodiment.

Acknowledgements

This work was partially supported by FCT/MCTES grant No. SFRH/BD/27305/2006 and CTS multi-annual funding, through the PIDDAC Program funds.

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):77–87.

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

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

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

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

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

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

Grassé, P.-P. (1959). La reconstruction du nid et les coordinations inter-individuelles chez bellicositermes et cubitermes sp. la théorie de la stigmergie: Essai d'interprétationdu comportement des termites con- structeurs. *Insectes Sociaux*, 6:41–80.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.

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

Mirolli, M., Ferrauto, T., and Nolfi, S. (2010). Categorisation through evidence accumulation in an active vision system. *Connection Science*, 22:331–354.

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

Nolfi, S. and Marocco, D. (2002). Active perception: a sensorimotor account of object categorization. In *Proceedings of the 7th International Conference on Simulation of Adaptive Behavior (SAB)*, pages 266–271. MIT Press.

Oliva, A. and Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12):520–527.

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

Rougier, N. and Vitay, J. (2006). Emergence of attention within a neural population. *Neural Networks*, 19(5):573–581.

Santana, P., Alves, N., Correia, L., and Barata, J. (2010). Swarm-based visual saliency for trail detection. In *Proceedings of the IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems (IROS)*, pages 759–765. IEEE Press, Piscataway.

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

Santana, P. and Correia, L. (2011). Swarm cognition on off-road autonomous robots. *Swarm Intelligence*, 5(1):45–72.

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

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

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

Trianni, V. and Tuci, E. (2011). Swarm cognition and artificial life. In *Proceedings of the 10th European Conference on Artificial Life (ECAL 2009)*, volume LNCS/LNAI 5777, 5778. Springer-Verlag, Berlin, Germany.

Trianni, V., Tuci, E., Passino, K., and Marshall, J. (2011). Swarm cognition: an interdisciplinary approach to the study of self-organising biological collectives. *Swarm Intelligence*, 5(1):3–18.

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

Turner, J. (2011). Termites as models of swarm cognition. *Swarm Intelligence*, 5(1):19–43.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., and Greene, M. R. (In Press). Visual search in scenes involves selective and nonselective pathways. *Trends in cognitive sciences*, doi:10.1016/j.tics.2010.12.001.

Zhang, X., Hu, W., Maybank, S., Li, X., and Zhu, M. (2008). Sequential particle swarm optimization for visual tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE Computer Socienty, Washington, DC.