Ricardo André Martins Mendonça

Licenciado em Ciências de
Engenharia Electrotécnica e de Computadores

A Learning Approach to Swarm-based Path Detection and Tracking

Dissertação para obtenção do Grau de Mestre em
Engenharia Electrotécnica e de Computadores

Orientador: José António Barata de Oliveira,
Professor Auxiliar, FCT-UNL

Co-orientador: Pedro Figueiredo Santana,
Professor Auxiliar, ISCTE-IUL

Júri:

Presidente: Prof. Doutor Pedro Alexandre da Costa Sousa
Arguente: Prof. Doutor Lino José Forte Marques
Vogais: Prof. Doutor José António Barata de Oliveira
Prof. Doutor Pedro Figueiredo Santana

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março 2012

Copyright

A Learning Approach to Swarm-Based Path Detection and Tracking

Acknowledgements

I would like to express my gratitude to my dissertation supervisor Prof. José Barata for his invaluable support and for giving me the outstanding opportunity not only to contribute with a work on the robotics domain, but also to join a brilliant and fun research team: Gonçalo Cândido, Eduardo Pinto, Magno Guedes, Pedro Deusdado, Giovanni di Orio, Pedro Gomes, and more recently, Francisco and André.

A special acknowledgement to my co-supervisor Pedro Santana for his effort in helping me become a better engineer, sharing his experience and scientific knowledge. I am also grateful for his complete availability to the development of my dissertation. Working with him have been a very rewarding experience.

My next acknowledgements goes to all my colleagues and friends that, along these years, made my academic journey more fun and interesting. In particular, a word of thanks to João Santos, Tiago Xavier, Juleo, Carlos Carvalho, Miguel Marques, Fábio Alves, Flávio Dinis, Fábio Januário, Pedro Ferreira de Almeida, and many others.

A special word of thanks to my friend Pedro Gomes for his friendship, support and feedback. I wish him every success in his life.

My infinite gratitude to my parents for giving me the motivation, the never-ending support and the unconditional love since my early days on the fascinating journey that life is.

Last but not least, a special thanks to my paramount star and jewel, Nita, who always enlightens and enriches my life every day.

Abstract

This dissertation presents a set of top-down modulation mechanisms for the modulation of the swarm-based visual saliency computation process proposed by Santana et al. (2010) in context of path detection and tracking. In the original visual saliency computation process, two swarms of agents sensitive to bottom-up conspicuity information interact via pheromone-like signals so as to converge on the most likely location of the path being sought. The behaviours ruling the agents' motion are composed of a set of perception-action rules that embed top-down knowledge about the path's overall layout. This reduces ambiguity in the face of distractors. However, distractors with a shape similar to the one of the path being sought can still misguide the system. To mitigate this issue, this dissertation proposes the use of a contrast model to modulate the conspicuity computation and the use of an appearance model to modulate the pheromone deployment. Given the heterogeneity of the paths, these models are learnt online. Using in a modulation context and not in a direct image processing, the complexity of these models can be reduced without hampering robustness. The result is a system computationally parsimonious with a work frequency of 20 Hz. Experimental results obtained from a data set encompassing 39 diverse videos show the ability of the proposed model to localise the path in 98.67 % of the 29789 evaluated frames.

keywords: swarm cognition, monocular path detection, visual saliency, bio-inspired methods, off-road navigation.

Resumo

Esta dissertação apresenta um conjunto de mecanismos para modulação do processo de computação de saliência visual proposto por Santana et al. (2010) no contexto da detecção e seguimento de caminhos. No processo de computação de saliência visual original, dois enxames de agentes sensíveis à informação de conspicuidade visual interagem através de feromonas virtuais, de modo a convergirem para a localização do caminho procurado. Os comportamentos que regem o deslocamento destes agentes são especificados através de um conjunto de regras percepção-acção, que incorporam conhecimento de alto nível sobre a morfologia de um caminho típico. Este conhecimento reduz a ambiguidade face a regiões salientes no campo visual que não pertencem ao caminho. No entanto, se estas regiões forem semelhantes à morfologia de caminhos típicos, podem desviar a actividade do enxame para fora do caminho. Com o objectivo de resolver este problema, esta dissertação propõe a utilização de um modelo de contraste para modular a computação da conspicuidade e de um modelo de aparência para modular o depósito de feromona. Dada a heterogeneidade dos caminhos, estes modelos são aprendidos em tempo de execução. Ao serem usados num contexto de modulação, e não para um processamento directo da imagem, a complexidade dos modelos pode ser reduzida sem com isso limitar a robustez. O resultado é um sistema computacionalmente parsimonioso capaz de funcionar a uma frequência de 20 Hz. Resultados experimentais, obtidos a partir de um conjunto de 39 vídeos, mostram a capacidade do modelo a localizar o caminho em 98,67 % do total de 29789 frames avaliados.

palavras-chave: cognição de enxame, detecção de caminhos monocular, saliência visual, métodos bio-inspirados, navegação todo-o-terreno.

Contents

List of Figures

List of Tables

List of Notations

α_b	Contribution of a behaviour b
β_1	Learning rate of the top-down appearance-based model
β_2	Learning rate of the top-down contrast-based model
δ	Maximum intensity deviation
$\Delta\rho$	Incremental value of ρ
ϵ	Pheromone level baseline
η_1	Maximum allowed iterations for the motion of a virtual ant
η_2	Initialisation phase's number of frames
γ	Weight of a stochastic behaviour
γ_0	Initial value of γ
γ_τ	Constant factor for the γ exponential decay
λ	Ratio of the neural field used to initialize the pheromone maps
$\mathbf{A}$	Top-down appearance-based probability map
$\mathbf{C}_s$	Centre-surround feature map
$\mathbf{C}_w^C$	Top-down colour contrast-based map
$\mathbf{C}_w^I$	Top-down intensity contrast-based map
$\mathbf{C}_{bu}^C$	Bottom-up colour conspicuity map
$\mathbf{C}_{bu}^I$	Bottom-up intensity conspicuity map
$\mathbf{C}_{bu}^{CI}$	Bottom-up colour and intensity conspicuity map
$\mathbf{C}_{td}^C$	Top-down colour conspicuity map
$\mathbf{C}_{td}^I$	Top-down intensity conspicuity map
$\mathbf{C}_{temp}$	Auxiliary conspicuity map
$\mathbf{E}_w$	Excitation map
$\mathbf{F}$	Dynamic neural field

$\mathbf{H}$	Homography matrix
$\mathbf{I}$	Input frame
$\mathbf{I}'$	Previous input frame
$\mathbf{I}_w$	Inhibition map
$\mathbf{P}^C$	Colour pheromone map
$\mathbf{P}^C_*$	Auxiliary colour pheromone map
$\mathbf{P}^I$	Intensity pheromone map
$\mathbf{P}^I_*$	Auxiliary intensity pheromone map
$\mathbf{R}_{msr}$	Most salient region mask
$\mathbf{R}_{search}$	Search mask
$\mathbf{S}$	Saliency map
$\mathbf{u}^m$	Unidimensional vector of maximum pheromone levels related to visual feature m
$\mathbf{v}^m$	Unidimensional vector of average conspicuity levels related to visual feature m
$\mathbf{w}$	Top-down contrast-based model (weight vector)
$\mathbf{w}'$	Weight vector with the new feature map weights for the current frame
$\mathbf{h_{ref}}$	Top-down appearance-based model (normalised image histogram)
$\mathbf{h_{sample}}$	Normalised histogram with new information about the target path
$\mathbf{u}'_i$	Corner point at $\mathbf{I}'$
$\mathbf{u}_i$	Corner point at $\mathbf{I}$
$\mathcal{B}$	Blue feature map
$\mathcal{G}$	Green feature map
$\mathcal{I}$	Intensity feature map
$\mathcal{R}$	Red feature map
$\mathcal{Y}$	Yellow feature map
$\mathcal{BY}$	Blue-yellow double-opponency feature map
$\mathcal{RG}$	Red-green double-opponency feature map
$\Phi(p_m)$	Level of pheromone deployed by the virtual ant p_m
ρ	Pheromone influence factor
ρ_0	Initial value of ρ
σ	Pixel's intensity value
υ	Amount of $\Phi(p_m)$ deployed in the pheromone map of the opposite visual feature, m'

A	Virtual ants' action space
a	Virtual ants' action
a_{p_m}	Action selected by the virtual ant associated to visual feature m
B	Virtual ants' set of behaviours
b	Blue channel
b_n	Normalised blue channel
b_{p_m}	Virtual ants' behaviour
C	Colour feature
c_1	c_1 channel
c_2	c_2 channel
c_3	c_3 channel
g	Green channel
g_n	Normalised green channel
h	Image's height
I	Intensity feature
$K(.)$	Image normalising operator for conspicuity maps
k_h	Number of intensity intervals
m	Visual feature identifier
$M(.)$	Function that returns the global maximum intensity of a given image
m'	Opposite visual feature in relation to m
m_1	Number of non-zero pixels of a given image mask
m_2	Number of zero pixels of a given image mask
m_w	Weighted average intensity of a conspicuity map
n	Number of virtual ants deployed on conspicuity maps
n_p	Number of pixels in a given image
n_{cs}	Number of centre-surround feature maps
o_{p_m}	Location at which a virtual ant, associated to visual feature m, is deployed
p_m	Virtual ant associated to visual feature m
r	Red channel
r_n	Normalised red channel
T	Set of pixels that belong to the path trajectory

V_{p_m}	Set of pixels that belong to the trajectory performed by a virtual ant
w	Image's width
$w(x, y)$	Weight of the pixel at row y and column x
w_i	Weight at the i^{th} position of the vector $\mathbf{w}$
x	Image's column
x_s	Column of the highest intensity pixel
y	Image's row
y_n	Normalised yellow channel
y_s	Row of the highest intensity pixel
z	Offset from the bottom row of the image
z_1	Trigger value for the deployment of a virtual ant
z_2	Sample value obtained from a uniform distribution

Chapter 1

Introduction

Paths always played an important role on human civilisations. Since the early days of mankind, paths were built to create trade routes between small remote settlements in search of items not available in their own locality. In addition, paths usually provide safe passages in demanding environments, thus reducing the chances of the traveller getting lost or incurring in dead-lock situations. Following this observation, field robots should also benefit from exploiting these visual structures. Furthermore, the presence of these paths can also be regard as an indirect visual cue that can be used to help performing a direct free-space visual assessment. In particular, the latter inference is achieved by analysing the terrain's surface volumetric properties, using range information provided by stereo-vision systems or laser scanners (Batavia and Singh, 2001; Lacroix et al., 2002; Manduchi et al., 2005; Broggi et al., 2005; Seraji, 2006; Konolige et al., 2009; Kolter et al., 2009; Rusu et al., 2009; Santana et al., 2011). Hence, the motivation behind this dissertation resides in the fact that off-road robots benefit from having the perceptual capabilities required to exploit paths, saving computation time in obstacle detection and path planning. A practical application of path following can be environmental surveillance and protection.

Detecting paths can be rather complex given their wide variety, ranging from structured paved roads to nature trails with varying shape and tread materials. This high diversity difficults the task of path detection, due to a lack of a well defined path's geometric structure and appearance information. Concretely, this hampers a straightforward definition and learning of either path or background models. Furthermore, nature trails usually impose a defiant acquisition process of helpful three-dimensional information, as they not always have a recognisable volumetric signature in typical off-road environments. Therefore, model-free solutions (or as free as possible) are essential for the development of robust and general path detection systems.

Paths are usually conspicuous structures in the visual field of the robot. Following this observation to help the task of path detection, the use of visual salience is exploited in a swarm-based model proposed by Santana et al. (2010), which was shown to operate where previous models fail. The merit of this approach is to not impose any hard constraints on the appearance or shape of both path and background. This happens in part because visual salience and path location are positively correlated. This model assumes the paths' overall layout is more predictable than other visual features (e.g. colour). This a priori knowledge is embedded in a motion behaviours set of simple swarm

agents that inhabit the visual bottom-up conspicuity maps (products of visual salience computation). Each path created by an agent is taken as a hypothesis. Moreover, these agents interact with each other, materialising the metaphor of *collective intelligence* (Franks, 1989) exhibited by social insects. The swarm's goal is to ensure that the agents cooperatively build up, through these pheromone-like interactions, a robust approximation of the actual path's skeleton. To accumulate evidence across frames, the model relies on a dynamic neural field (Amari, 1977; Rougier and Vitay, 2006), extended with a mechanism to compensate robot motion.

The presence of too many distractors or a considerably heterogeneity of the path itself can lead to situations where the bottom-up conspicuity maps are not so well behaved, creating ambiguities on trail hypotheses and, on the worst case, temporarily misleading the swarm off the path. This tracking challenge can be reduced by applying the general concept of top-down knowledge, boosting the set of visual features (e.g., colour), known beforehand to better describe the object being sought (Frintrop et al., 2005; Navalpakkam and Itti, 2005). This dissertation exploits this observation to improve the original swarm-based model proposed by Santana et al. (2010). Moreover, visual features can be considerably unpredictable in the case of trails in natural environments. Therefore, top-down knowledge about paths must be learnt and updated on-line, increasing robustness to sudden changes in the path. To obtain this goal, this dissertation proposes a simple learning mechanism to learn path appearance and contrast models. The activity of each agent is top-down biased by the appearance and contrast-based models. Moreover, pheromone is deployed proportionally to the likelihood between the agent's trajectory and the path. As a result of numerous pheromone-like interactions, this mechanism allows the swarm activity to be spatially biased according to the expected path's appearance. This approach renders a cross-influence between the perception of appearance and contrast, and the perception of shape, which promotes robustness without hampering computational parsimony.

To validate the proposed model, experimental results were obtained from a data set of 39 diverse videos. The results showed that the model herein proposed is able to localise the path in the robot's visual field in $98.67\,\%$ of the 29789 evaluated frames at $20\,\text{Hz}$, whereas the original model attained $84.66\,\%$ at $20\,\text{Hz}$ on the same data set.

1.1 Dissertation Outline

This dissertation is organised as follows:

Chapter 2 reviews the state-of-the-art for road and trail detection;

Chapter 3 presents the supporting concepts of this work. In particular, an overview of the original path detector (Santana, 2011; Santana et al., 2010) is provided. Namely, the creation of bottom-up visual attention maps, the deployment and execution of swarm agents, and the temporal filter used to integrate evidence across time and to promote the swarm perceptual grouping, are outlined.

Chapter 4 describes the extensions proposed to the original model (Santana, 2011; Santana et al., 2010). Namely, the addition of top-down knowledge models about the appearance and contrast

of the path, and a learning mechanism to update these models;

Chapter 5 presents the experimental setup and the set of results obtained from a data set of 39 diverse videos;

Chapter 6 aggregates a set of conclusions, main contributions of this dissertation, and further research opportunities on the subject.

Chapter 2

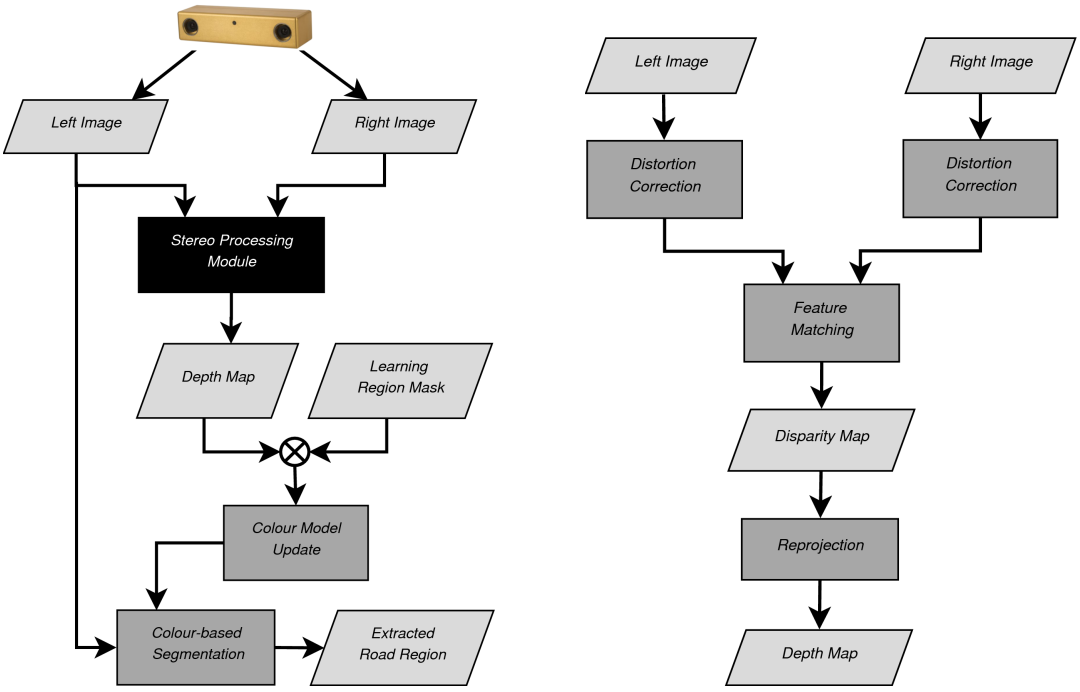
Related Work

There are different path detection methods that can be applied in diverse scenarios, given the wide variety of paths. Some methods are more suitable for structured paved roads, where others are better applied to nature trails. In particular, several road detection methods have been proposed and fielded successfully. These methods exploit the intrinsic characteristics of roads, such as: well defined boundaries; distinct appearance with respect to the surroundings; and they are somewhat monotonous structures, i.e., sudden changes in their shape seldom occurs. Conversely, nature trails are rather misbehaved structures. In some scenarios, there is a lack of strong edges delimiting them and the surroundings can be blended with the trail. As trail detection methods rely on work developed for the road domain, a survey is first done on road detection methods.

2.1 Road Detection Methods

The detection and tracking of paved roads is facilitated by the predictable appearance of the road's surface and by its delimiting strong edges. However, this is not the case of ill-structured unpaved rural roads. The typical solution on this latter case is to use a region-based approach to segment the road region from its surroundings. This segmentation can be achieved by a pixel-wise classification mechanism that can either be trained off-line or on-line. Off-line learning is done from a set of already labelled images (Chaturvedi and Malcolm, 2005; Alon et al., 2006), whereas on-line learning gives a more adaptive and robust operation, as it is done from a set of reference regions in the input image that the system was able to automatically label as road/non-road. The capability to discern autonomously between target and non-target regions is the challenge of the latter approach. A possible solution to this problem is to exploit short range volumetric information obtained from other sensors (e.g., laser or stereo) to discriminate the road plane from others (Thrun et al., 2006; Tue-Cuong et al., 2008). Fig. 2.1 depicts a stereo-based road detection system architecture.

Alternatively, one can assume that some regions can be logically labelled as road. For instance, if the road is wide and assuming that the robot is on it, then the near region in front of it can be labelled as road (Thorpe et al., 1988; Fernandez and Casals, 1997; Fernandez and Price, 2005; Song et al., 2007).



(a) Stereo-based road detection architecture

(b) Stereo processing module (black box in (a))

Figure 2.1: Overview of a stereo-based system for road region extraction. Adapted from Tue-Cuong et al. (2008). Shortly, a pair of images of the road are taken at the same time from two separate cameras (displaced horizontally from one another) and passed to a stereo processing module (b). The displacement of relative features among these images is measured to calculate a disparity map. Knowing the geometric arrangement of the cameras, the disparity map can be translated into a depth map. This depth map is classified into ground and non-ground patches. A learning region is then defined in front of the vehicle for colour data collection. At the intersection of this region and the ground (non-obstacle) patches, sample pixels are extracted to update the road colour model. Finally, the image is segmented according to the road colour model.

Once the road is segmented from the background, information regarding its appearance, geometry, and orientation can be extracted and used to build and update a road model. In general, a simplified model of the road (e.g., triangular) can be used to fit to the segmented image. To handle hard to model roads, region growing can be an interesting alternative to the model fitting process (Ghurchian et al., 2004; Fernandez and Price, 2005; Chaturvedi and Malcolm, 2005). In particular, Fernandez and Price (2005) presents a method for dirt road detection and tracking using colour vision and region growing technique. In order to segment the dirt road, this method assumes that a small rectangle at the centre-bottom of the image always contains a portion of the road that is suitable for analysis. Another assumption made is that the colour-space statistics of the road surface are different from the one of the surrounding regions. The analysis of the road area starts by computing the mean and standard deviation of the pixels' hue, saturation and intensity. The adaptation to road appearance changes is assured by a periodic update of these statistics. Then, a colour-based filter is parametrised according to the calculated means and standard deviations. The input image is processed using a recursive subdivision method. Shortly, the image is first divided into a small number of sub-regions that are processed coarsely. Upon finding a pixel that satisfies the filter equation (i.e. a pixel that belongs to road region), the current sub-region is further divided. This process proceeds until a minimal region is achieved. Afterwards, the image is divided into horizontal slices that will determine slices of the road by a region growing process. The outcome is a series of road

segments. In order to track the road, these slices' centres of mass are used to planning a trajectory along them, based on a variation of cubic spline fitting. Fig. 2.2 summarises the process. A limitation of this road detection method is its proneness to fail when the road's surface and surrounding regions share similar appearances. Another disadvantage that limits the application of this method is the assumption that the centre-bottom of the input image always contains part of the road, which cannot be guaranteed in the presence of narrow nature trails.

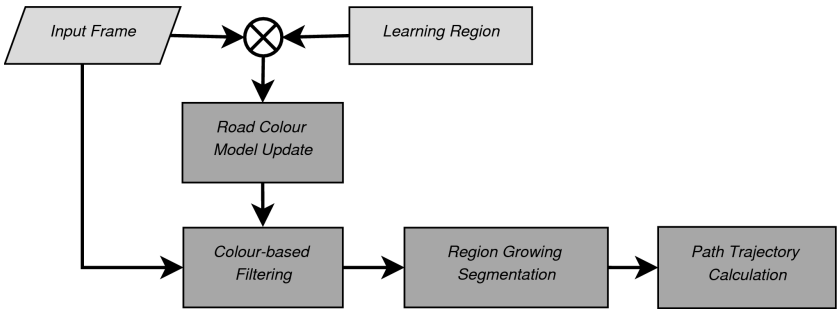


Figure 2.2: Overview of a region growing technique for road region extraction (adapted from Fernandez and Price (2005)). Briefly, an priori defined learning region is used to update a road colour model. A colour-based filter is then applied to the original image, removing the majority of non-road pixels. A region growing module segments the filtered image into road slices. The centers of mass of these slices are used to computed a motion trajectory.

An interesting alternative to the region growing process is to enforce a global shape constraint on the product of an unsupervised clustering mechanism (Crisman and Thorpe, 1991), which discards the need for a road/non-road pixel classification process. The trade-off on this method is the elimination of road appearance models at the cost of raising the number of possible ambiguities between regions with similar shape.

When the road and the background share the same appearance, the previously presented approaches may be inadequate. In this case, the dominant texture orientations, like road borders and wheel tracks, can be helpful to extract the road's vanishing point (Rasmussen, 2004, 2008; Kong et al., 2010). There are also hybrid architectures that integrate the orientations-based and regions-based approaches Alon et al. (2006); Song et al. (2007).

2.2 Trail Detection Methods

The road detection methods, described in the previous section, are the basis of most work done on trail detection. For instance, Bartel et al. (2007) use a region-based approach that relies on a priori knowledge about the colour distributions of both path and background. In this method, the path segmentation is done by detecting and extracting its borders, given the a priori knowledge that paths are grey and surrounded by grass or planted borders. Bartel et al. (2007) replace all green pixels of the input image by black ones, ensuring that the contrast between the path and boundaries is maximised. Next, a Gaussian convolution is applied to eliminate fine textures. The contrast of

the input image is then enhanced and a threshold filter is applied, leaving the path region as the brightest one. Afterwards, a gradient filter is applied to extract the edges of the detected pathway. To conclude the path segmentation, an object extraction algorithm is applied to remove shadows within the boundaries of the path. The outcome of these steps is shown in Fig. 2.3. This border extraction technique cannot be employed when there is no well defined path edges. Moreover, a priori colour knowledge about paths and their surroundings is of little use in less structured environments.

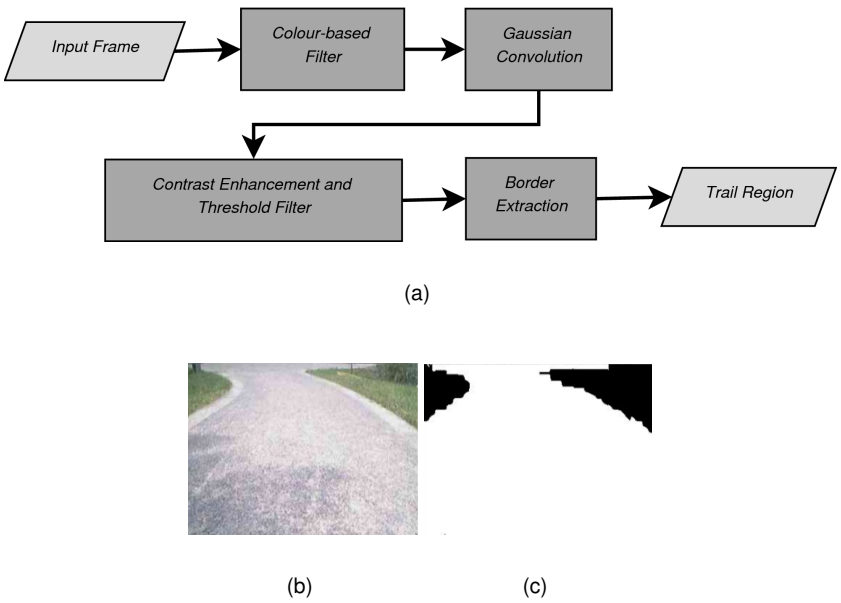


Figure 2.3: Processing steps for path border extraction using a priori knowledge. According to the model proposed by Bartel et al. (2007). (b) Input image. (c) Contrast enhancement and threshold filter result.

Substituting a priori models by self-supervised learning models helps in providing required adaptability in situations in which the trail's appearance is heterogeneous (Grudic and Mulligan, 2006; Rasmussen and Scott, 2008b). However, it is difficult to assure that the robot is on the trail when shape of the latter varies. From this observation, it follows that defining reference regions to supervise the learning process is not a straightforward task, as it is in the road domain. Moreover, the use of depth information to find the trail plane might be not so helpful if the trail and its surroundings exhibit the same height. Alternatively, as done in the road domain, the use of a global shape constraint, by first over-segmenting the image and then scoring a set of trail hypotheses against the global shape, have experienced some success (Rasmussen and Scott, 2008a; Rasmussen et al., 2009; Blas et al., 2008). In particular, Rasmussen et al. (2009) make the assumption that the trail's shape is approximately triangular under perspective and both left and right sides share the same appearance. (see Fig. 2.4). To track the trail, particle filtering is used. Each triangle hypothesis corresponds to a particle and its weight is the score given by a trail likelihood function. To measure appearance similarity between triangular regions, a technique based on histograms of *k*-means[1] (Lloyd, 1982) cluster labels is used. That is, a set of textons[2] describing colour features in CIE-Lab colour space is created at each pixel of the input image. *K*-means is performed only on textons with non-saturated pixels

[1]The K-means algorithm is a clustering method which aims to partition samples into k clusters. Each sample belongs to the cluster with the nearest mean. The result may depend on the initial clusters and there is no guarantee that it will converge to the global optimum.

[2]The term *textons* refers to fundamental micro-structures in generic natural images and the most basic elements in early (pre-attentive) visual perception. Please refer to (Zhu et al., 2005) for further information.

to create 8 texton labels. These are combined with the under- and over-saturated groups to yield a small set of final texton labels. The hypothesis region's colour distribution is modelled by a histogram of texton labels inside it, and the chi-squared metric is used to measure the appearance dissimilarity between the two histograms. The triangular shape-constraint presented by Rasmussen et al. (2009) is inadequate when the trail is considerably unstructured or interrupted. Moreover, false positives can emerge in situations where a distractor is on the trail. For instance, if the triangle hypothesis is over a bush at the centre of the trail, then the triangular neighbouring regions to its right and left are similar (trail regions), and the bush itself gives an high contrast with the surround. In this scenario, a misleading high score will be given to this trail hypothesis by the trail likelihood function.

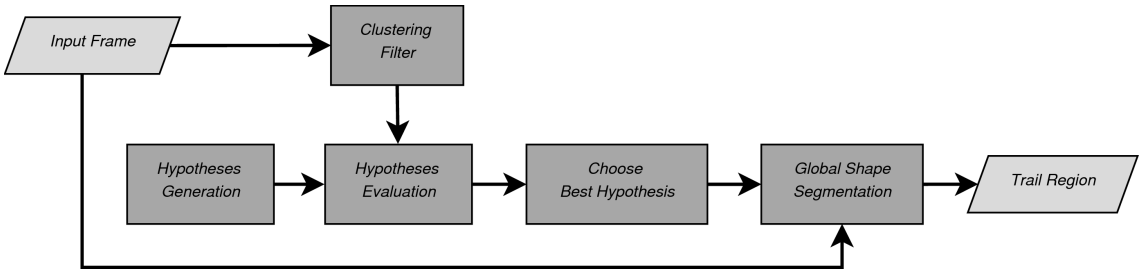


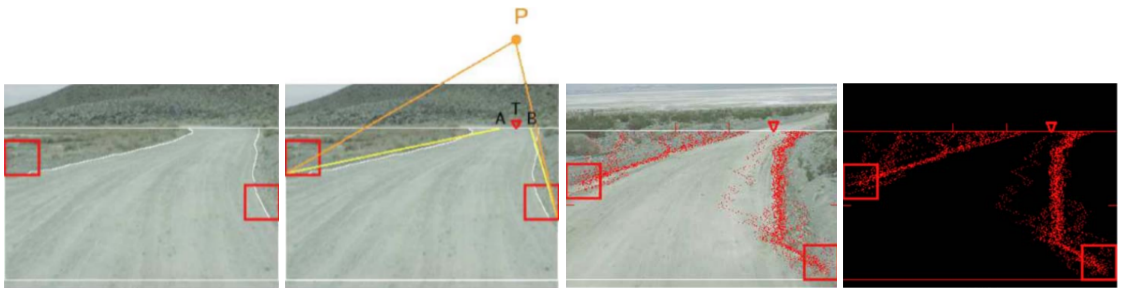
Figure 2.4: Model for trail detection and tracking based on shape-constraint (Rasmussen et al., 2009). The shape-based visual trail tracker assumes that the trail region is approximately triangular under perspective. It generates region hypotheses from a learnt distribution of expected trail width and curvature variation, which are scored according to the colour and brightness contrast with flanking regions.

In general, the use of the global shape constraint limits the type of trails that can be detected. Moreover, nature trails may have not clear edges segmenting them from the background, hampering the accuracy of the image over- segmentation process that precedes the global shape application.

Another concept with limited application in the trail domain is the vanishing point method. Despite of the good results in the road domain, the vanishing point method seldom applies in trail detection, as the global orientation of the trail is rarely indicated by dominant orientations.

An interesting line of research that can be applied in path detection is the use of the social insects metaphor (swarm intelligence). This concept has been applied in the design of several computer vision systems (Poli and Valli, 1993; Liu et al., 1997; 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; Santana, 2011). For instance, the ant foraging metaphor is exploited by Broggi and Cattani (2006), proposing a swarm-based system for trail border detection, in which two agent colonies are set to track each side of the trail. Agents move pixel by pixel, trying to find trail's borders. The motion rules are inspired by the behaviours of biological ants. Before executing the swarm algorithm, the starting regions and the height limit for the agents' motion in the input image must be defined. This is done by setting the starting areas in the periphery of the image, where a sufficient percentage of edges is present. Each agent is put randomly inside these areas. A point of attraction polarises the random moving component of the agents. Hence, the average moving direction of the colony is towards this point. Moreover, this point of attraction is computed frame by frame, using information on previously computed trail boundaries. The agents of a colony are divided into n different subsets, each one characterised by different moving rules parameters. The subsets are executed in sequence, and when all the agents of a certain subset have reached the upper limit pixels,

the pheromone trails are updated. The movements of the first subset are only based on heuristics, ignoring the pheromone deposit (edge-exploitation phase). As the execution proceeds, the other subsets of agents become increasingly sensitive to pheromone and less to heuristics (pheromone-exploitation). Finally, two agents, one per colony, only attracted by the pheromone trails are deployed. The final left and right road boundaries are defined by the trajectories executed by these two agents. Although the interesting results, this method is limited to well delimited paths. However, as aforementioned, trails rarely have strong edges in natural environments to help in trail border detection.



(a) (b) (c) (d)

Figure 2.5: Ant colony optimisation approach to detect off-road trail borders (Broggi and Cattani, 2006). The white curves on (a) represent road boundaries obtained from a previous frame. The horizontal lines are the bottom and upper limits of the agents' movement. The yellow straight lines in (b) are a linear approximation of the white curves, and the triangle T is the midpoint of the intersection between the upper limit line with the yellow ones. The new point of attraction is represented by point P. Two borders at two sides of the road are tracked by two agent colonies, where (c) presents the input image with the agents' paths superposed, and (d) shows only the ants' paths for an easier visual evaluation.

From a different perspective, Santana et al. (2010) propose the use of a conspicuity space together with swarm cognition concepts for the design of self-organising in visual attention, exploiting the observation that trails are conspicuous structures. The typical distributed and parallel design of swarm-based models is a major helpful feature, as the path hypothesis generation process demands an active selection of multiple pixels in order to approximate the skeleton of the path being sought. The swarm agents interact with each other indirectly by using a dynamical 2-D neural field that simulates the physical medium in which pheromone is deposited and propagated in time. In particular, these pheromone-like interactions are based on a phenomenon known as *stigmergy*[3]. A key advantage of the use of conspicuity space over the work proposed by Rasmussen et al. (2009), is that visual salience represents contrast information between trail and local surroundings, as well as between the path and the overall scene. Moreover, Santana (2011) showed that by modelling the cognitive process of visual attention as a self-organising process, the typical challenge of speed-accuracy trade-off in the face of context and task changes is more easily handled. The model proposed by Santana et al. (2010) discards hard constraints on the shape of both path and background, as well as the use of any a priori appearance and contrast knowledge.

The presence of too many visual distractors (e.g., salient non-path regions) can lead to situations in which the swarm-based model (Santana et al., 2010) does not converge to the path location. This issue can be diminished by top-down boosting the bottom-up visual features that describes the path being sought. The contribution of this dissertation lies in the integration of both contrast

[3]Stigmergy is a mechanism of indirect coordination between agents. This is achieved when an agent lefts a trace in the environment, stimulating the execution of a next action, by the same or a different agent. It was first observed in social insects and was introduced by Grassé (1959) (e.g., ants exchange information by laying down pheromones).

and appearance-based top-down knowledge into the original model and the addition of an adaptive pheromone deployment mechanism, based on the likelihood between the swarm agent's trajectory and the path. An on-line learning method is used to learn the appearance and contrast of the path, increasing robustness to sudden changes on the latter. Furthermore, the synergistic interaction between both bottom-up and top-down pathways reduces the dependency on accurate path models. Hence, the proposed model is potentially more suitable for path detection and tracking.

Chapter 3

Supporting Concepts

This chapter introduces the reader to the key aspects of the original model, proposed by Santana et al. (2010), which serves as basis for this dissertation. These key aspects are the use of visual attention for path detection and the use of swarm agents for its computation. Visual attention is introduced in Section 3.1, whereas the swarm paradigm is overviewed in Section 3.2. Finally, in Section 3.3, the details about the model itself are provided from (Santana, 2011; Santana et al., 2010), in which the interested reader may also refer for a detailed explanation.

3.1 Visual Attention

The ability of a visual system to detect salient regions on a given scene is known as visual attention. These regions can be used in complex vision tasks, such as detection, tracking or even recognition of objects. The active search for interesting regions, as done with eye movements in biological visual systems, is known as overt attention. The indirect perception of these regions is referred to as covert attention. For instance, without moving the eyes, humans can mentally acknowledge peripheral salient regions in their visual field. The belief that multiple covert attention processes co-exist in the human brain has been support by several studies (Pylyshyn and Storm, 1988; Doran et al., 2009).

Bottom-up and top-down factors can bias the focus of attention. Bottom-up attention is derived from instinctive and reflexive mechanisms that are triggered by the conspicuity of regions, like a source of light in a dark background. Top-down attention is a pro-active attention in the sense that, it is driven by expectations, motivations and goals of the subject, such as a priori knowledge about the object being sought. Therefore, bottom-up conspicuous regions obtained from the visual field can be analysed by a top-down cognitive process, which is based on knowledge of the object being sought. The outcome is a visual salience map which signals the regions of the visual field that are simultaneously conspicuous and share the general properties of the object of interest.

3.1.1 The Human Visual System

As the human perception englobes visual attention, a brief and simplified description about the human visual system is herein presented.

When the light achieves the human eye, it enters by the pupil, travels through the vitreous humour, and reaches the retina (see Fig. 3.1). The retina is formed by numerous photosensitive cells that transform the electromagnetic waves (light) into neural impulses. These photoreceptors cells are divided into two types: the rods and the cones (Kandel et al., 2000). The cones are colour sensitive, whereas the rods are sensitive to luminance. In particular, the cones are subdivided into three categories, each one sensitive to a specific colour frequency of the visible spectrum: red, green or blue. Rods and cones are connected to ganglion cells, via bipolar cells (Kandel et al., 2000). Ganglion cells transform the analog signal (graded potentials) to a discrete one by sending electrical discharges into the optic nerve. This stimulus travels from the optic nerve to the optic chiasm, where is divided in two pathways to each brain hemisphere (see Fig. 3.2). Visual cognitive processes in the brain are then fed from this stimulus.

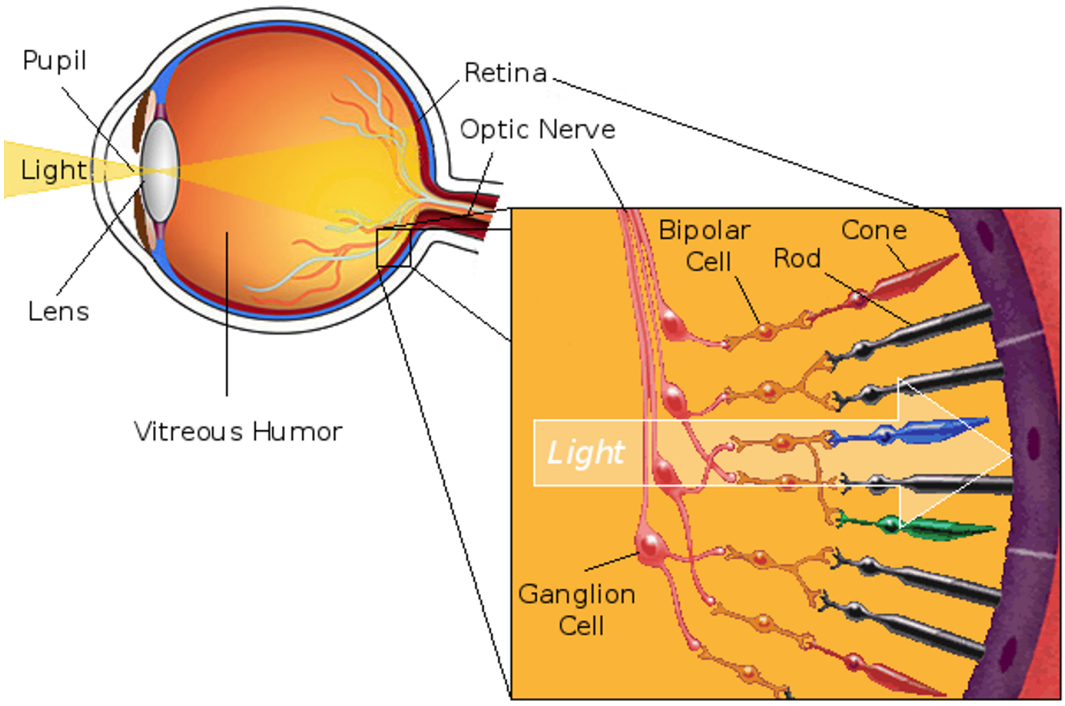


Figure 3.1: Simplified anatomy of the human eye. Adapted from (Reinhardt, 2010)) and from (Stroobandt, 1997).

The complex, but yet hierarchical, connections between the diverse neuronal cells composing the retina, are the key to form receptive fields with different complexity. Specifically, receptive fields of cells at one level of the visual system are formed from input by cells at a lower level. Thus, small and simple receptive fields are combined to form large and complex ones. For instance, the receptive field of a photoreceptor cell is a fictional cone-shaped volume in the visual field, that comprises all

the directions in which light activates that cell. On the other hand, the bipolar cells have a circular receptive field composed by a centre area and a surround area, both connected to numerous photoreceptors.

Bipolar cells are divided into two groups: on-centre cells and off-centre cells (Kandel et al., 2000). On-centre cells are excited by the activation of the photoreceptors that compose the centre area. However, if the activation occurs on the respective surrounding area, these cells are inhibited. An off-centre cell have the opposite response. In addition to centre and surround differences, colour opponency information is also generated with bipolar cells, as the latter could be connected to different cones. For instance, a bipolar cell can process the differences between the output of a red cone and a green cone, or the differences between blue cones and a combined signal from both red and green cones (blue-yellow opponency).

The receptive field of a ganglion cell encompasses all the photoreceptors connected to bipolar cells, which are in turn connected to this particular ganglion cell. Consequently, the organisation of ganglion cells' receptive fields provides a way of detecting not only light exposition through photoreceptors, but also the centre-surround differences, i.e., luminance and colour contrast information.

Finally, there are two major classes of ganglion cells: magnocellular and parvocellular (Kandel et al., 2000). Magnocellular cells are more sensitive to luminance (light-dark contrast) and can receive signals from both rods and cones. Parvocellular cells are sensitive to colour and, thus, only receive signals from cones. In particular, parvocellular are subdivided into two groups: one that receives red-green opponent signals, and one that receives blue-yellow opponent signals.

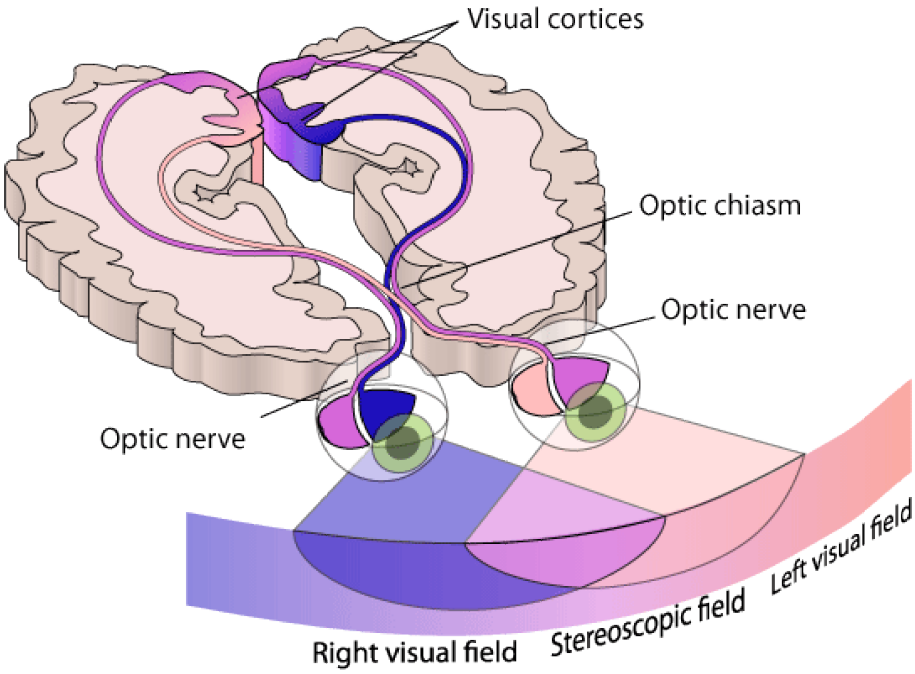


Figure 3.2: Human visual pathway (from ADInstruments (2009)).

3.1.2 Visual Attention Computational Model

In computer vision, the paradigm of visual attention has been widely investigated (Ahmed, 1991; Tsotsos et al., 1995; Koch and Ullman, 1985) and implemented in both software (Itti et al., 1998) and hardware (Ouerhani and Hugli, 2003b) domains. In particular, the biologically inspired saliency-based model of visual attention presented by Itti et al. (1998) has been used in several computer vision applications (Todt and Torras, 2000; Ouerhani and Hugli, 2003a). Moreover, the plausibility of this salience-based model has been assessed by Ouerhani et al. (2004) and encouraging results about correlation of human and computer attention were obtained. Under these circumstances, this visual attention model is properly adapted and used by Santana et al. (2010) and, therefore, herein described.

First, a set of visual feature maps is obtained from an input colour image. This set is composed by an intensity feature $\mathcal{I}$ and two double-opponency colour features, respectively for Red-Green, $\mathcal{RG}$, and for Blue-Yellow, $\mathcal{BY}$, opponency. The existence of this chromatic opponency in human visual cortex has been proved by Engel et al. (1997). Although only intensity and colour features are used for the sake of computational speed, additional features (e.g., orientations and depth) could be used for improved background-path segmentation. Formally, these visual features are computed in the following way:

$$\mathcal{I} = \frac{r+g+b}{3}, \tag{3.1}$$

$$r_n = \frac{r-(g+b)}{2}, \tag{3.2}$$

$$g_n = \frac{g-(r+b)}{2}, \tag{3.3}$$

$$b_n = \frac{b-(r+g)}{2}, \tag{3.4}$$

$$y_n = \frac{(r+g)}{2} - \frac{|r-g|}{2} - b, \tag{3.5}$$

$$\mathcal{RG} = \mathcal{R} - \mathcal{G}, \tag{3.6}$$

$$\mathcal{BY} = \mathcal{B} - \mathcal{Y}, \tag{3.7}$$

with r, g, and b being the red, green, and blue channels of the input colour image. The corresponding normalised channels are denoted by r_n, g_n, and b_n, respectively. The y_n is the normalised yellow channel. If y_n has negative values, they are set to zero. To decouple hue from intensity, these channels are normalised by $\mathcal{I}$, denoting $\mathcal{R}$, $\mathcal{G}$, $\mathcal{B}$, and $\mathcal{Y}$, respectively. As proposed by Itti et al. (1998), only pixels with $\mathcal{I}$ larger than 10% of its maximum value are submitted to this second normalisation, as hue variations are not perceivable at very low luminance. Other pixels yield a zero value.

Each feature map is transformed into its respective conspicuity map through a centre-surround mechanism (Itti et al., 1998), highlighting the regions of the input scene that strongly differ from their surroundings. This centre-surround mechanism mimics the behaviour of the retinal bipolar cells in the human eye. The centre-surround operator is illustrated in Fig. 3.3.

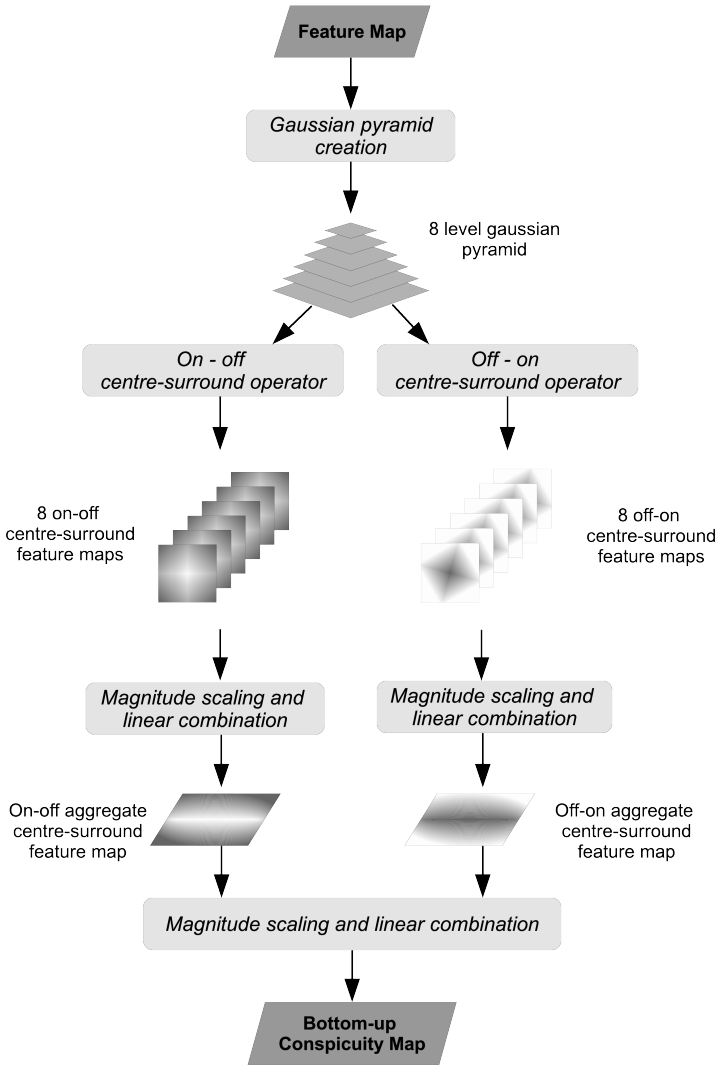


Figure 3.3: Bottom-up centre-surround operator.

Shortly, one dyadic Gaussian pyramid (Burt and Adelson, 1983), with eight levels (or scales), is computed from the intensity channel. Two additional pyramids, also with eight levels, are computed to account for the Red-Green and Blue-Yellow double-opponency color feature sub-channels. These various scales are used to perform centre-surround operations. The resulting set of on-off and off-on centre-surround maps per pyramid have higher intensity on those pixels whose corresponding feature differs the most from their surroundings. On-off centre-surround maps are built by across-scale point-by-point subtraction, between a level with a fine scale and a level with a coarser one. Off-on maps are computed the other way around, i.e., subtracting the coarser level from the finer one. Both on-off and off-on centre-surround maps are used separately, rather than considering the modulo of the difference, as done by Itti et al. (1998). This separation yields better results as shown in (Frintrop et al., 2005; Frintrop, 2006). All centre-surround maps built from the intensity pyramid are re-sized to a common size and independently scaled in magnitude according to a normalisation operator, and finally averaged together to produce the intensity conspicuity map $\mathbf{C}_{bu}^{I} \in [0, 1]$. The same process applies to create Red-Green and Blue-Yellow conspicuity maps, each one subsequently weighted and then averaged together to produce a single colour conspicuity map $\mathbf{C}_{bu}^{C} \in [0, 1]$. Fig. 3.4 depicts the bottom-up conspicuity computation process for a given input image, whereas Fig. 3.5 shows some more samples with the corresponding bottom-up conspicuity maps.

3.2 Swarm Intelligence

Swarm intelligence is the collective behaviour of decentralised and self-organizing systems that exhibit a collective intelligence (ben, 1989). From such systems, a coherent pattern emerges not from the influence of a central authority, but as the result of the local interactions among their distributed components. Notwithstanding the limited cognition capabilities of these individual processes, the system as a whole can solve complex problems more efficiently or solve those that go beyond the capability of a single individual process. Despite of external perturbations, self-organising systems can maintain its orderly behaviour and can be inherently robust to individual failures, if there is redundancy in their components. Moreover, only minimal complexity is required for each constituent parts of such systems. However, devising these individual processes can be challenging, as the connection between simple local rules and the desired complex global properties is indirect.

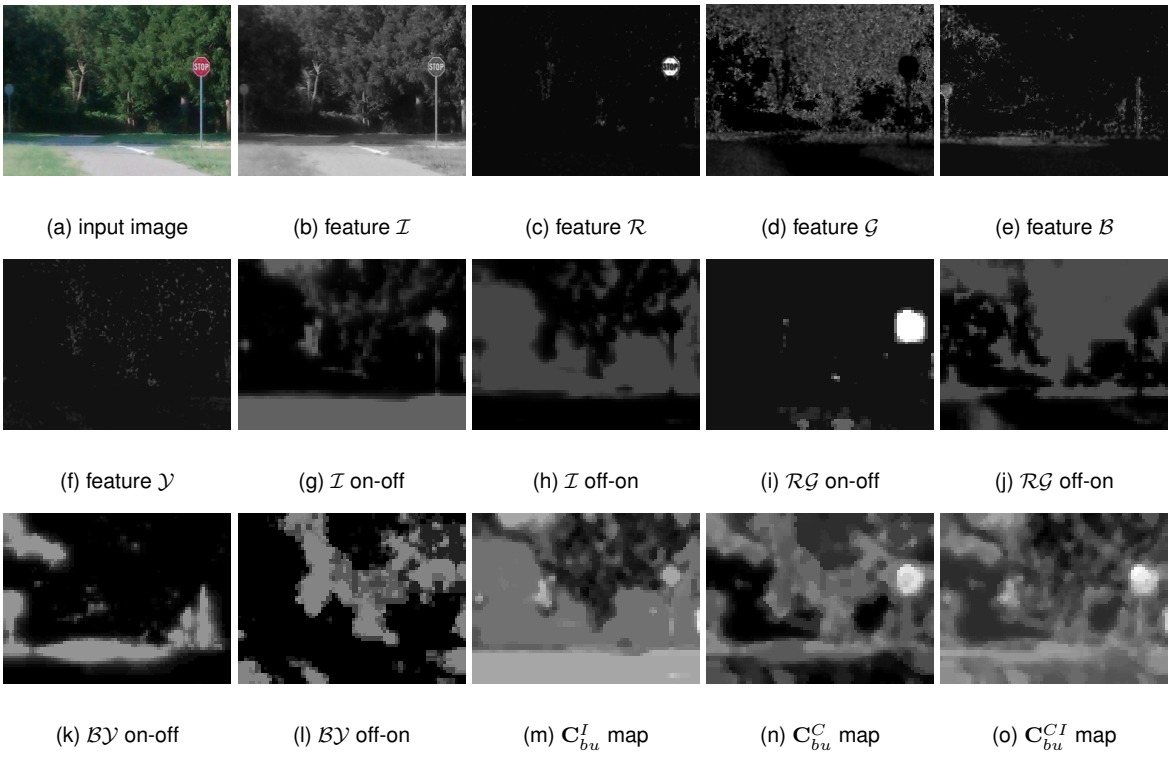
3.2.1 Biological Inspiration

Collective intelligence can be found in the animal kingdom, such as in bird flocking or exhibited by social insects. For instance, there is no centralised management in ant colonies. A coherent behaviour can be observed at the colony level, due to numerous interactions between individual ants, following simple rules and tasks. Some collect waste and perform maintenance duties, some search and collect food, others defend the colony, and so on. Another example is the search procedure done by bees when looking for a new hive location. When bees decide to move to a new hive and begin a new colony, scout bees fly out in all directions, searching for a suitable location. When one finds a interesting place, it flies back to the hive and communicates the new finding to the other scouts. The new hive location is chosen only when fifteen bees happens to arrive at the same location. Furthermore, a particular species of wasps organise themselves into different task-oriented groups, in which the size of each group is regulated according to the colony needs (Jeanne, 1996).

Finally, considering the example of the human body, one can see the emergence of a collective intelligence in this highly coordinated system of interacting swarms of cells, messenger molecules and bacterias.

3.2.2 Swarm-based Computational Models

From an engineering perspective, swarm-based computational models have desirable characteristics to solve complex problems. They are flexible in dynamic environments and implemented with simple and elementary rules to achieve a complex group behaviour. A certain randomness is also introduced to help the system exploring new and creative solutions. Furthermore, these models are robust to individual failures and inherently distributed and parallel with little or no supervision. Therefore, diverse swarm-based computational models have been applied to several domains with interesting results. For instance, to address the problem of reducing traffic jams, Oliveira and Bazzan (2006) proposes a swarm-based approach to coordinate and synchronise traffic lights in an efficient pattern. Each traffic light is an agent that interacts with other agents to perform adaptive signalling plans. Stimuli are provided in the form of produced pheromone according to the volume of traffic.



(a) input image (b) feature $\mathcal{I}$ (c) feature $\mathcal{R}$ (d) feature $\mathcal{G}$ (e) feature $\mathcal{B}$

(f) feature $\mathcal{Y}$ (g) $\mathcal{I}$ on-off (h) $\mathcal{I}$ off-on (i) $\mathcal{RG}$ on-off (j) $\mathcal{RG}$ off-on

(k) $\mathcal{BY}$ on-off (l) $\mathcal{BY}$ off-on (m) $\mathbf{C}_{bu}^{I}$ map (n) $\mathbf{C}_{bu}^{C}$ map (o) $\mathbf{C}_{bu}^{CI}$ map

Figure 3.4: Bottom-up conspicuity computation process. The set of bottom-up centre-surround maps are shown from (g) to (l). The combined bottom-up intensity and colour conspicuity map, obtained by a linear combination of (m) and (n), are depict by (o). The most salient regions are marked by white pixels on conspicuity maps. For instance, the stop signal are the most salient object in the scene.

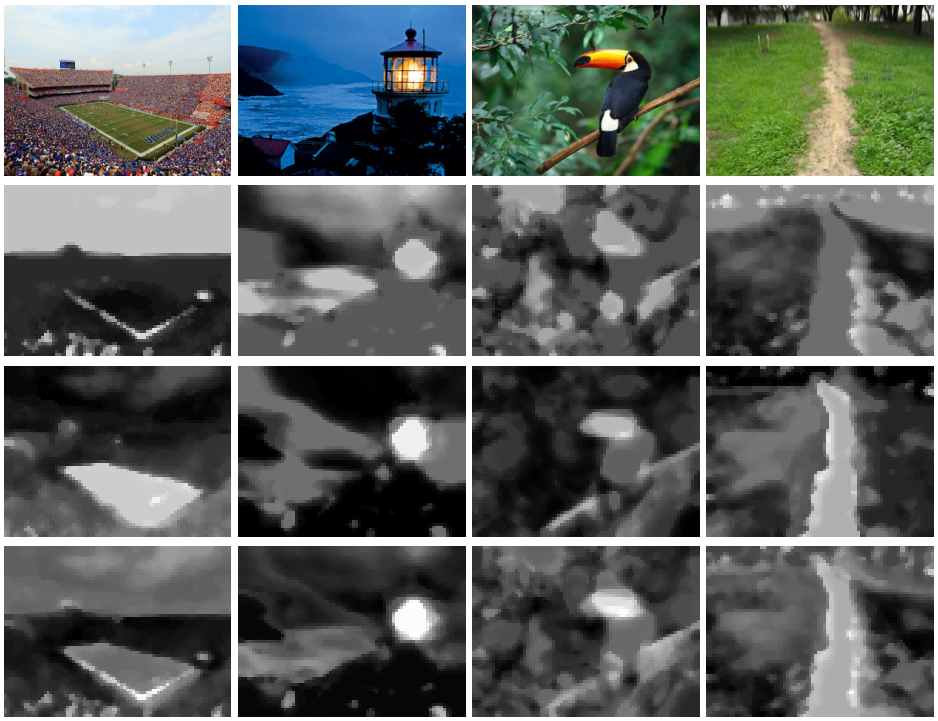


Figure 3.5: Examples of bottom-up conspicuity maps. Input image (top row), bottom-up intensity conspicuity map $\mathbf{C}_{bu}^{I}$ (top middle row), bottom-up colour conspicuity map $\mathbf{C}_{bu}^{C}$ (bottom middle row), and combined bottom-up colour and intensity conspicuity map $\mathbf{C}_{bu}^{CI}$ (bottom row). The field of the stadium, the light of the lighthouse, the toucan's beak, and the trail are all salient regions on the visual field.

Another interesting application of swarm intelligence is to model and simulate biological systems. Jacob et al. (2004) proposes a three-dimensional swarm-based model to simulate the human immune system reaction to viral antigen exposure. In particular, the production of antibodies in response to a viral population is modelled, as the reinforced memory response to a previously encountered pathogen. In the environmental domain, a swarm-based model of forest dynamics to simulate ecological disturbances is described in (Savage and Askenazi, 1998).

Finally, the original model (Santana et al., 2010) demonstrated that fast and relatively robust systems can be achieved with a simple and loosely coupled swarm-based design.

3.3 Swarm-based Path Detection

The original model (Santana et al., 2010) aims at detecting the path using bottom-up visual cues. Concretely, it computes visual conspicuity maps based on a set of bottom-up centre-surround feature maps, obtained from the input image at various spatial scales. These conspicuity maps are then shape-based filtered according to a priori knowledge about path's generic morphological properties. This is done implicitly by setting behavioural rules into a set of agents operating on the visual conspicuity maps. Their collective behaviour results in a final salience map that represents the path hypothesis. Fig. 3.6 depicts how the original model processes each frame.

3.3.1 Model Execution Overview

At each new frame $\mathbf{I}$, the two bottom-up conspicuity maps, $\mathbf{C}_{bu}^{C}$ for colour and $\mathbf{C}_{bu}^{I}$ for intensity information, are computed. The intensity of a pixel belonging to either one of these maps, signals how much that pixel detaches from the background at several scales. The bottom-up conspicuity maps, $\mathbf{C}_{bu}^{C}$ and $\mathbf{C}_{bu}^{I}$, are then shape-based filtered with the use of a priori top-down knowledge about typical path's shape. A set of n virtual ants (hereafter called p-ants, from perceptual-ants) is deployed on each bottom-up conspicuity map. The swarm activity builds two pheromone maps, $\mathbf{P}^{C} \in [0,1]$ for colour and $\mathbf{P}^{I} \in [0,1]$ for intensity information. Moreover, the p-ants' set of behaviours is designed to exploit a priori knowledge about typical paths approximate layout. Hence, the activation of pheromone maps is expected to match the path's location better than the activation of bottom-up conspicuity maps. Thus, rather than combining both bottom-up conspicuity maps to generate the final salience map $\mathbf{S}$, as typically done Itti et al. (1998); Frintrop et al. (2005), these map is obtained by combining both pheromone fields, $\mathbf{S} \leftarrow \frac{1}{2}\mathbf{P}^{I} + \frac{1}{2}\mathbf{P}^{C}$. Additionally, to create cross-modality influences, p-ants on a given pheromone map also affect the other pheromone map. This cross-modality increases robustness by allowing p-ants to exploit multiple cues indirectly.

The final salience map, $\mathbf{S}$, feeds a dynamic neural field (permanent pheromone map), $\mathbf{F} \in [0,1]$, which integrates pheromone (i.e., evidence) across frames, playing the role of a temporal filter. This allows self-organisation to occur at a longer time-scale and, as a consequence, to enable tracking. However, to avoid that the mentioned cross-modality influences propagate across frames and probably induce an undesirable neural field's activity build-up, two auxiliary pheromone maps, $\mathbf{P}_{*}^{I}$ and $\mathbf{P}_{*}^{C}$, are created free of these influences. Therefore, these auxiliary maps only encompasses the

pheromone deposited by the p-ants associated to the respective visual feature. These maps are then used to replace the pheromone maps, $\mathbf{P}^I \leftarrow \mathbf{P}^I_*, \mathbf{P}^C \leftarrow \mathbf{P}^C_*$, just before blending them for the purpose of creating $\mathbf{S}$. Moreover, in order to allow p-ants' creation and activity to be affected by history, at the onset of each frame, both instantaneous pheromone maps are initialised with a small ratio λ of the neural field after being motion compensated, $\mathbf{P}^I \leftarrow \lambda \cdot \mathbf{F}, \mathbf{P}^C \leftarrow \lambda \cdot \mathbf{F}$.

Motion compensation between current frame $\mathbf{I}$ and previous frame $\mathbf{I}'$ is also implemented so that the dynamics of the neural field can be decoupled from the dynamics of the robot. 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 path's pixel. Note that conspicuity maps, pheromone maps, final salience map, and neural field, all share the same width w and height h. These two values are selected bearing in mind real-time performance.

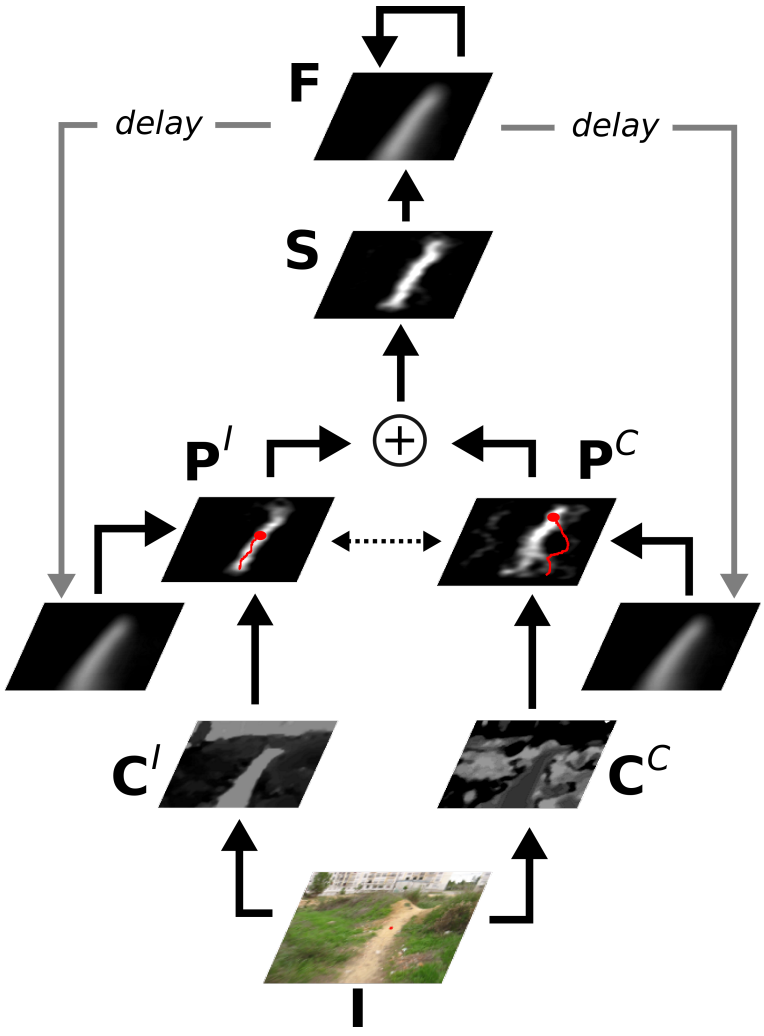


Figure 3.6: Operation overview of the model proposed by Santana et al. (2010). Two bottom-up conspicuity maps, $\mathbf{C}^C_{bu}$ and $\mathbf{C}^I_{bu}$ are computed from the input frame $\mathbf{I}$. A set of n virtual ants, embedding a priori top-down knowledge about typical path morphology, are deployed on these two bottom-up conspicuity maps. The result of the swarm activity are two pheromone maps, $\mathbf{P}^C$ and $\mathbf{P}^I$, that are combined to obtained a final salience map $\mathbf{S}$. To exploit multiple cues indirectly, p-ants on a given pheromone map also affect the other pheromone map. The final salience map, $\mathbf{S}$, feeds a dynamic neural field, $\mathbf{F}$, which integrates pheromone across frames. In order to allow p-ants' creation and activity to be affected by history, at the onset of each frame, both instantaneous pheromone maps are initialised with a small ratio λ of the neural field after being motion compensated.

3.3.2 Bottom-Up Conspicuity Maps Computation

The bottom-up conspicuity maps, $\mathbf{C}_{bu}^{C}$ and $\mathbf{C}_{bu}^{I}$, are computed according to the method proposed by Itti et al. (1998) and explained in Section 3.1. These bottom-up conspicuity maps and the centre-surround ones are magnitude scaled by recurring to a normalisation operator. The goal is to promote maps that have fewer conspicuous locations, avoiding that, when blending maps, strongly salient objects appearing in only a few maps are masked by noise or by other less-salient objects present in other maps. Therefore, the contribution of each pixel to the average is weighted according to its distance from the top row of the image. Formally, let $p(x, y)$ be the pixel in column x and row y of a given conspicuity map $\mathbf{C}$, with height h and, $M(.)$ a function that returns the global maximum intensity of $\mathbf{C}$. Let $w(x, y) = \sqrt{y/h}$ be the weight of pixel $p(x, y)$. The map's weighted average, m_w, is thus given by

$$m_w(C) = \frac{\sum\limits_{(x,y)\,\in\,C} p(x, y) \cdot w(x, y)}{\sum\limits_{(x,y)\,\in\,C} w(x, y)}, \tag{3.8}$$

and the normalising operator, $K(.)$, takes the form

$$K(\mathbf{C}) = \mathbf{C} \cdot \Big(M(\mathbf{C}) - m_w(\mathbf{C})\Big)^2. \tag{3.9}$$

All maps are 8-bit grayscale images, meaning that $M(\mathbf{C}) = 255$. Fig. 3.7 depicts normalised bottom-up conspicuity maps for a given input image. As stated by Santana et al. (2010), the $K(.)$ normalising operator shows a small quantitative improvement over other normalising operators, achieving better results in some key frames as it allocates higher levels of salience to path than to the background.

(a) Input image (b) Conspicuity map $\mathbf{C}_{bu}^{I}$ (c) Conspicuity map $\mathbf{C}_{bu}^{C}$

Figure 3.7: Bottom-up intensity (b) and colour (c) conspicuity maps obtained from image (a).

3.3.3 Pheromone Maps Computation

The bottom-up conspicuity maps, $\mathbf{C}_{bu}^{I}$ and $\mathbf{C}_{bu}^{C}$ are shape-based filtered by p-ants, so as to build two pheromone maps, $\mathbf{P}^{I}$ and $\mathbf{P}^{C}$, respectively. The pheromone maps computation begins with the creation and deployment of a p-ant, p_m, associated to a visual feature $m \in \{I, C\}$ (intensity or colour). The other visual feature is represented by m'.

The creation of a p-ant p_m on a given location o_{p_m} depends on the level of conspicuity and pheromone at that location on the corresponding conspicuity map $\mathbf{C}^m$ and pheromone map $\mathbf{P}^m$. Hence, to reduce sensitivity to any potentially noise at the boundaries of the conspicuity map $\mathbf{C}^m$, p-ants are deployed within a small randomly selected offset $z \in [0, 0.1 \cdot h]$ from the bottom of the map. In particular, a p-ant is deployed at row $r \in [h, h - z]$, where h is the height of the map[1]. In order to determine the respective deployment column, a uni-dimensional vector $\mathbf{v}^m = (v_0^m, \ldots, v_w^m)$ is computed, where 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 from the bottom row of the map, r. To compute the pheromone level, the same windowing process is applied to build the vector $\mathbf{u}^m = (u_0^m, \ldots, u_w^m)$, where the element u_k^m corresponds to the maximum pheromone level found in the window.

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, according with the following test:

$$z_1 < (\rho \cdot u_{z_2 \cdot w}^m + (1 - \rho) \cdot v_{z_2 \cdot w}^m), \tag{3.10}$$

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. Moreover, ρ operates as an adaptive process, changing the system from a conspicuity-driven behaviour (exploration) to a pheromone-driven behaviour (refinement/exploitation), by starting with a small ρ_0, and linearly increasing at each iteration by an amount $\Delta\rho$.

Observing carefully the deployment condition, Eq. 3.10, one can see that p-ants are progressively and probabilistically deployed on path-like locations, assuming that the path tend to be conspicuous and has been successfully detected in the previous frame (neural field dynamics), and that the pheromone accumulated by p-ants deployed in the current frame builds-up mostly around the actual path's location.

After being deployed, the p-ant must iterate (with a maximum of η_1 iterations) on the conspicuity map, $\mathbf{C}^m$, creating a trajectory that represents a path hypothesis. The p-ant's motion is ruled by a set of simple behaviours that make little assumptions regarding the path' structure. However, before specifying p-ants behaviours, it is necessary to specify their sensory and action spaces. The sensory space is defined by five receptive fields disposed around the p-ant's current position (see Fig. 3.8). An action $a \in A$ moves the p-ant to one of the five neighbour pixels not behind the current p-ant's position. The action space is thus defined by the set $A = \{1, 2, 3, 4, 5\}$.

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

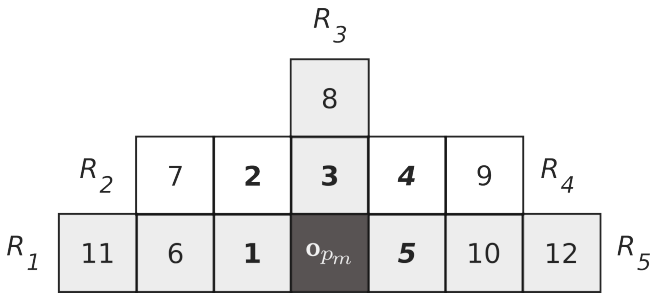


Figure 3.8: P-ant's set of receptive fields (Santana, 2011), namely, $R_1 = \{1, 6, 11\}$, $R_2 = \{2, 7\}$, $R_3 = \{3, 8\}$, $R_4 = \{4, 9\}$, and $R_5 = \{5, 10, 12\}$.

At each iteration, p-ant p_m executes a set of behaviours $B = \{greedy, track, centre, ahead, commit\}$, which independently vote on each possible action in A. Formally, behaviours are described as functions that return a vote in the interval $[0, 1]$ for each possible action $a \in A$. Following a typical approach of behaviour coordination (Rosenblatt, 1995), the most voted action is the one taken by the p-ant. Table 3.1 summarises, for each behaviour, which regions in the neighbourhood of the p-ant are associated to the most preferred action. Fig. 3.9 illustrates the pheromone trajectory created according to the set of p-ant's behaviours. As already stated, these behaviours embed top-down path shape information.

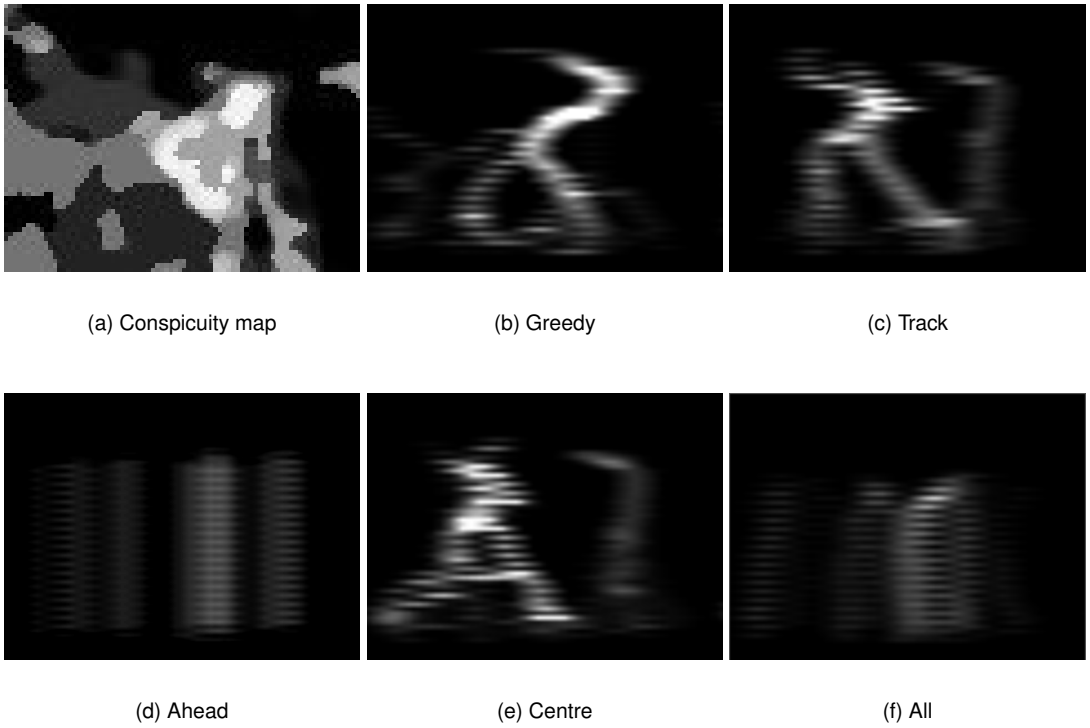
Table 3.1: P-Ant behaviours for path detection. Adapted from Santana (2011).

Behaviour	**Voting Preferences**
$greedy$	Regions of higher levels of conspicuity, under the assumption that trails are salient in the input image.
$track$	Regions whose average level of conspicuity is more similar to the average level of conspicuity of all the pixels visited by the p-ant, under the assumption that trails' appearance is homogeneous.
$centre$	Regions that maintain the p-ant equidistant to the boundaries of the trail hypothesis being pursued.
$ahead$	Upwards regions under the assumption that trails are often vertically elongated.
$commit$	Region targeted by the motor action at the previous iteration, under the assumption that trails' orientation tend to be monotonous.

The p-ant p_m selects its action a_{p_m} by maximising the following function, which incorporates behaviours' votes, pheromone-based interactions, and random fluctuations:

$$a_{p_m} = \arg\max_{a \in A} \left(\sum_{b_{p_m} \in B} \alpha_b f_b(p_m, a) + \mathbf{P}^m(R_a, o_{p_m}) + \gamma q \right), \tag{3.11}$$

where α_b is a user defined weight accounting for the contribution of behaviour $b_{p_m} \in B$ and γ is the weight accounting for stochastic behaviour, being $q \in [0, 1]$ a number sampled from a uniform



(a) Conspicuity map (b) Greedy (c) Track

(d) Ahead (e) Centre (f) All

Figure 3.9: P-ants' behaviours demonstration. The bottom-up conspicuity map in (a) was used to demonstrate the p-ants' behaviours. As can be seen in (b), the greedy behaviour chooses regions with maximum conspicuity, whereas the track behaviour (c) moves toward regions with conspicuity similar to the average conspicuity from all pixels visited by the p-ant, since the beginning of its motion. The ahead behaviour (d) bias the p-ants to move upward. A purely commit behaviour (not combined with other behaviours) has the same effect as the ahead behaviour in (d), due to replication of the p-ant's previous action. The centre behaviour (e) attracts the p-ants to the centre of a region, by detecting the boundaries created by regions with a different conspicuity level. The contribution of all behaviours, depicted in (f), shows a dominant pheromone trajectory with convergence to a region with high conspicuity.

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 of an immediate loop detection, namely, the p-ant moving recurrently from one pixel to another, then the action for the current iteration is randomly selected. The p-ant deploys pheromone in each corresponding position on $\mathbf{P}^m$ with a magnitude $\Phi(p_m)$, and a small portion of it, υ, in $\mathbf{P}^{m'}$:

$$\Phi(p_m) = \epsilon \tag{3.12}$$

where ϵ is an empirically defined pheromone level baseline. Finally, the p-ant's position o_{p_m} is updated according to the selected action. Another p-ant associated to the other visual feature, p'_m, is deployed and iterated following the same procedure. The modification of both pheromone maps (colour and intensity) by p-ants, enables a loosely coupled cross-modality influence, allows each p-ant to exploit multiple cues indirectly. This process is repeated n times, meaning that $2n$ p-ants are created and iterated. Algorithm 1 outlines the overall iteration process.

Once p-ants' activity has ceased, the computation of the two pheromone maps, $\mathbf{P}^I$ and $\mathbf{P}^C$, is

Algorithm 1 P-ant execution pseudo-code

Input: p-ant (p), conspicuity map ($\mathbf{C}_{bu}^{m}$), pheromone map ($\mathbf{P}^{m}$), pheromone map of other visual feature ($\mathbf{P}^{m'}$), pheromone map without cross-modality and neural field influences ($\mathbf{P}_{*}^{m}$))
Output: updated pheromone maps, $\mathbf{P}^{m}$, $\mathbf{P}^{m'}$, and $\mathbf{P}_{*}^{m}$
Data: $\eta_1, \epsilon, \upsilon$ are empirically defined constants.

// default previously selected action is forward motion
$a'_{p_m} \leftarrow 3$

// initialize list of scalars representing the conspicuity level at each p-ant's visited position
$V_{p_m} \leftarrow \emptyset$

// execute p-ant p for η_1 times
for η_1 *iterations* **do**

  use Equation 3.11 to obtain p-ant's action, a_{p_m}, based on $\mathbf{C}_{bu}^{m}$, $\mathbf{P}^{m}$, V_{p_m}, and a'_{p_m}

  // append conspicuity of the new p-ant's position
  $V_{p_m} \leftarrow V_{p_m} \bigcup \{\mathbf{C}^{m}(R_{a_{p_m}}, \mathbf{o}_{p_m})\}$

  // use obtained p-ant's action, a_{p_m}, to update p-ant's position, $\mathbf{o}_{p_m}$
  $\mathbf{P}^{m}(\mathbf{o}_{p_m}) \leftarrow \mathbf{P}^{m}(\mathbf{o}_{p_m}) + \Phi(p_m)$ *// update $\mathbf{P}^{m}$ at pixel $\mathbf{o}_{p_m}$*
  $\mathbf{P}_{*}^{m}(\mathbf{o}_{p_m}) \leftarrow \mathbf{P}_{*}^{m}(\mathbf{o}_{p_m}) + \Phi(p_m)$ *// update $\mathbf{P}_{*}^{m}$ at pixel $\mathbf{o}_{p_m}$*
  $\mathbf{P}^{m'}(\mathbf{o}_{p_m}) \leftarrow \mathbf{P}^{m'}(\mathbf{o}_{p_m}) + \upsilon \cdot \Phi(p_m)$ *// update $\mathbf{P}^{m'}$ at pixel $\mathbf{o}_{p_m}$*

  // store selected action
  $a'_{p_m} \leftarrow a_{p_m}$

end for

return $(\mathbf{P}^{m}, \mathbf{P}^{m'}, \mathbf{P}_{*}^{m})$

complete. Then, the two pheromone maps are merged into a final salience map, $\mathbf{S}$:

$$\mathbf{S} = \frac{1}{2} \cdot \mathbf{P}^{I} + \frac{1}{2} \cdot \mathbf{P}^{C}. \tag{3.13}$$

Fig. 3.10 illustrates typical pheromone maps. Moreover, it shows a competition between p-ants with, roughly, two pheromone trajectories, one at the centre and another at the right of the map. The centre trajectory has more pheromone, as more p-ants are attracted by the high conspicuity.

The instantaneous salience map, $\mathbf{S}$, feeds a two dimensional dynamic neural field $\mathbf{F}$ (Amari, 1977; Rougier and Vitay, 2006), which is a 2-D lattice of $w \times h$ neurons, each one corresponding to one pixel of the salience map. The activated neurons excite their neighbours and inhibit distant ones, promoting perceptual grouping and reducing ambiguities in the focus of attention. The goal of the neural field, $\mathbf{F}$, is to integrate evidence across time, to consider competition between multiple focus of attention, and to promote perceptual grouping.

The dynamical characteristic of the neural fields, displayed in the form of inertia, is the key element that enables information to be integrated across time. However, if not properly handled, this property causes the field to smear when the robot moves. A way of avoiding this undesirable effect is to shift

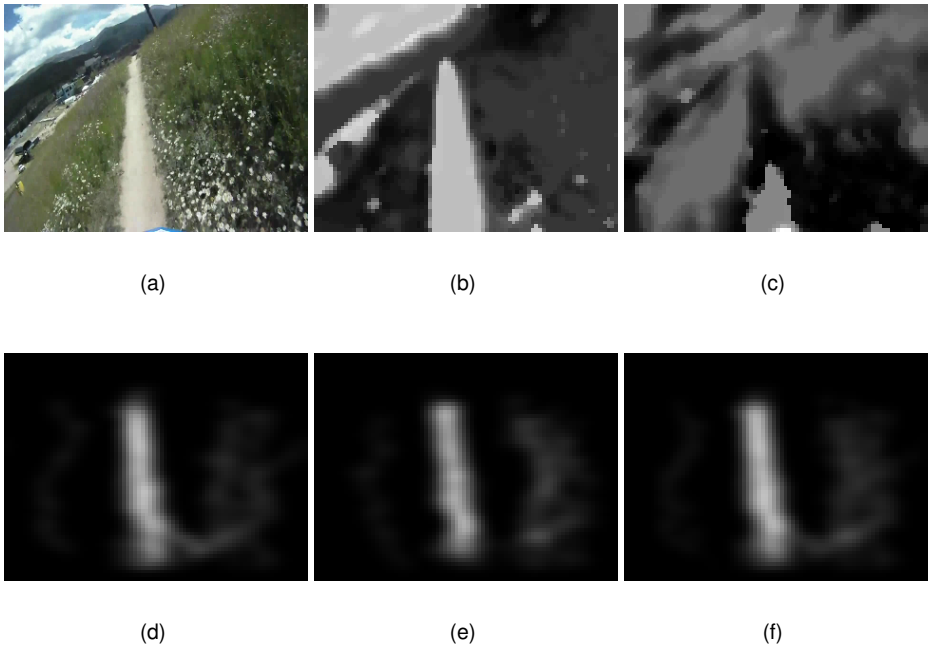


Figure 3.10: Pheromone maps computation. (a) Input image. (b) Bottom-up conspicuity map for intensity feature, $\mathbf{C}_{bu}^{I}$. (c) Bottom-up conspicuity map for colour feature, $\mathbf{C}_{bu}^{C}$. (d) Pheromone map for intensity feature, $\mathbf{P}^{I}$. (e) Pheromone map for colour feature, $\mathbf{P}^{C}$. (f) Salience map, $\mathbf{S}$, obtained by a linear combination of (d) and (e).

the neural field's activity according to the robot motion estimate by using asymmetrical kernels in the neurons (Zhang, 1996). The following three steps explicitly compensate the neural field for the camera motion engaged between the previous and current frames:

1. Estimate the homography matrix $\mathbf{H}$ that describes the projective transformation between the current frame, $\mathbf{I}$, and the previous one, $\mathbf{I}'$. To estimate the projective transformation $\mathbf{H}$, a set of corner points (Tomasi and Shi, 1994) is first detected in the previous frame, $\mathbf{I}'$. These points are then tracked in the current frame, $\mathbf{I}$, with a pyramidal implementation of the Lucas-Kanade feature tracker (Bouguet, 1999). The resulting sparse optical flow is then used to estimate the projective transformation relating both frames, i.e., the 3×3 homography matrix $\mathbf{H}$, such that,

$$\mathbf{u}_i' = \mathbf{H} \cdot \mathbf{u}_i, \tag{3.14}$$

   where $\mathbf{u}_i$ is a corner point found in $\mathbf{I}$ and $\mathbf{u}_i'$ its correspondence in $\mathbf{I}'$. If a minimum of four correspondences between corner points is not found, the homography matrix is set to the identity matrix, $\mathbf{H} = \text{diag}(1, 1, 1)$;

2. Obtain a motion compensated version of the previous neural field's state by using the estimated homography matrix, $\mathbf{F} \leftarrow \mathbf{HF}$;

3. Update $\mathbf{F}$ with the salience map $\mathbf{S}$ (Santana et al., 2010).

Chapter 4

Proposed Model

This chapter describes the work done in the framework of this thesis to improve the accuracy of the path detection and tracking. Concretely, a mechanism for on-line learning of top-down knowledge about the path being sought is added to the original model proposed by Santana et al. (2010). Section 4.1 presents the motivation behind this extension, whereas the proposed model is presented in Section 4.2.

4.1 Motivation

With the sudden presence of distractors in the bottom-up conspicuity maps, the sought path might not be as highly conspicuous as desired to help in its detection. Nevertheless, as stated by Santana et al. (2010), the original model is able to often detect the path's location in these scenarios, due to the key interaction between the neural field's inertia and the p-ants' sensorimotor coordination capabilities. However, if the path's conspicuity is low for several frames, the neural field's inertia cannot prevent the p-ants and, consequently, neither itself from migrating to the path's surrounding. The temporarily misleading of swarm activity to non-path regions can also be indirectly caused by an incorrect output in the optical flow, as this can severely affects the compensation of the neural field for the robot motion. Recovering from this situation can be difficult if the bottom-up conspicuity maps are populated with off-path distractors. Another issue is the presence of shadows. The latter tend to affect the bottom-up conspicuity maps, as they can break the segmented path's region, compromising its detection.

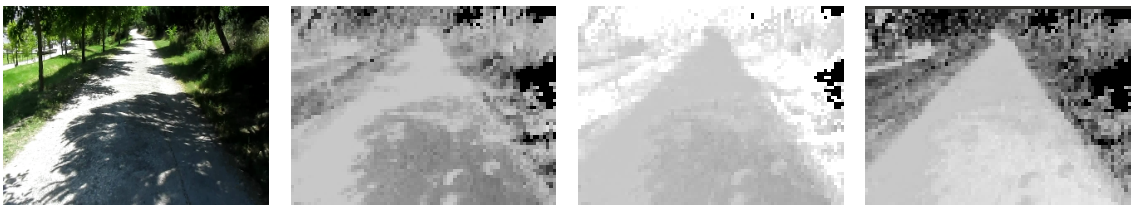
These are limitations that this dissertation mitigates by learning and using top-down knowledge to modulate the salience computation process and to influence the swarm activity, increasing robustness to sudden visual changes in the path. Fig. 4.1 illustrates the advantage of using top-down contrast-based modulation to path detection. The dependency on accurate path models is also reduced by the interactions between both bottom-up and top-down knowledge processes.

To minimise the effect of shadows, a shadow invariant colour space is used. A number of colour spaces such as HSI, normalised RGB, $c_1c_2c_3$, and $l_1l_2l_3$, have been tested and compared by Gevers

and Smeulders (1999). As the colour space $c_1c_2c_3$ showed to be the best shadow-invariant under indoor lightning (Gevers and Smeulders, 1999), as well as the best shadow and illumination invariant colour space for outdoor environments (Song et al., 2007), it is herein exploited. Fig. 4.2 shows an example in which this property is evident.

(a) (b) (c) (d)

Figure 4.1: Typical situation where top-down contrast-based modulation is key for a proper path detection. (a) Input image with most likely path location overlaid, using the top-down contrast-based modulated conspicuity maps depicted in (b). (c) Input image with most likely path location overlaid, using the bottom-up conspicuity map depicted in (d). The specific environmental configuration results in an inversion of the bottom-up conspicuity maps, attracting virtual ants to the sides of the path. By including top-down contrast knowledge, this problem is overcome.



(a) (b) (c) (d)

Figure 4.2: Shadow invariant colour space $c_1c_2c_3$. (a) RGB image. (b) Channel c_1. (c) Channel c_2. (d) Channel c_3. The shadow on path region, depicted in (a), is somewhat attenuated in the c_3 component (d).

Finally, the failure of the optical-flow motion compensation process remains as a limitation of the proposed model with no straightforward solution, as strong camera motion occurs in unstructured environments like nature trails. Moreover, the lack of well defined textures in poor lightning environments difficults the tracking of visual features, which are needed to compute the frame-wise translation and rotation matrices used to motion compensate the neural field. A hypothetical solution could be the use of inertial measurement units, as the latter does not rely on visual features to estimate the motion matrices.

4.2 Model Overview

The key concept of top-down knowledge is considered in the proposed model by several means. First, contrast knowledge is used to bias the computation of top-down contrast-based maps from

the bottom-up feature maps. Second, appearance knowledge plays a role in the computation of a probability map related to the path location and in an adaptive pheromone deployment process. Third, as in the original model, the swarm activity behaves like a shape-based filter, according to a priori knowledge about path's generic morphological properties. As results will show, the joint operation of these three sources of information is sufficient to ensure safe tracking of the path.

As seen on Chapter 3, a priori shape knowledge is implemented in the form of behavioural rules executed by the p-ants. Hence, this knowledge is taken as innate and, thus, not affected by learning. Conversely, top-down appearance and contrast models are learnt on-line, meaning that they need an initialisation phase. Therefore, the proposed model has two execution phases: the initialisation phase and the tracking phase. The initialisation phase lasts for η_2 frames, during which the system detects the path as in the original model, i.e., based upon bottom-up visual cues and without taking into account the top-down appearance and contrast models. In addition, the necessary information to build top-down knowledge about the target path is gathered. In the tracking phase, the model updates and uses the learnt top-down knowledge, so as to modulate the bottom-up conspicuity computation process. This way, the path location is more robustly tracked over time.

Similar to the original model, two bottom-up conspicuity maps, one encompassing intensity information, $\mathbf{C}_{bu}^{I}$, and another encompassing colour information, $\mathbf{C}_{bu}^{C}$, are computed from the input frame $\mathbf{I}$. In the tracking phase, these bottom-up conspicuity maps are biased by a top-down contrast-based model, $\mathbf{w}$, and fused with a top-down appearance-based probability map, $\mathbf{A}$. The top-down contrast model is implemented as an on-line learnt set of weights, each one representing the importance of a given visual feature to the detection of the path (see Section 4.3.2). The probability map, $\mathbf{A}$, is obtained by back projecting an on-line learnt top-down appearance-based model, which is implemented as a normalised histogram, $\mathbf{h_{ref}}$ (see Section 4.3.1).

The two top-down conspicuity maps, $\mathbf{C}_{td}^{I}$ and $\mathbf{C}_{td}^{C}$, are subsequently shape-based filtered into two pheromone maps, $\mathbf{P}^{I}$ and $\mathbf{P}^{C}$, by the activity of two swarm of p-ants, as in the original model. The pheromone maps are initialised with neural field's activity, as seen on Chapter 3. Moreover, each swarm operates over the respective pheromone map, using the information contained in the corresponding top-down conspicuity map. If system is in the initialisation phase, the bottom-up conspicuity map is used instead of the respective top-down map, as there is no top-down knowledge models created yet. Top-down appearance knowledge is used to modulate the level of pheromone deployed by the virtual ants. Rather than having p-ants deploying a constant level of pheromone along their paths, this approach compels p-ants deploying higher doses of pheromone on regions of the visual field whose appearance is similar to the one of the path.

After the shape-based filtering process, the resulting set of two pheromone maps are fused together, generating the salience map, $\mathbf{S}$. As in the original model, this salience map feeds the dynamical neural field, $\mathbf{F}$, which performs temporal filtering. Finally, the most salient region of the visual field is obtained as an image mask[1], which constrains the learning of information used to update both appearance and contrast models (see Section 4.4). Fig. 4.3 depicts the proposed model's pipeline, whereas Fig. 4.4 provides a system's snapshot on a typical situation. Algorithm 2 outlines the initialisation phase and Algorithm 3 the tracking one.

[1] An image mask is a binary image whose pixels' intensity belongs to {0, 1}.

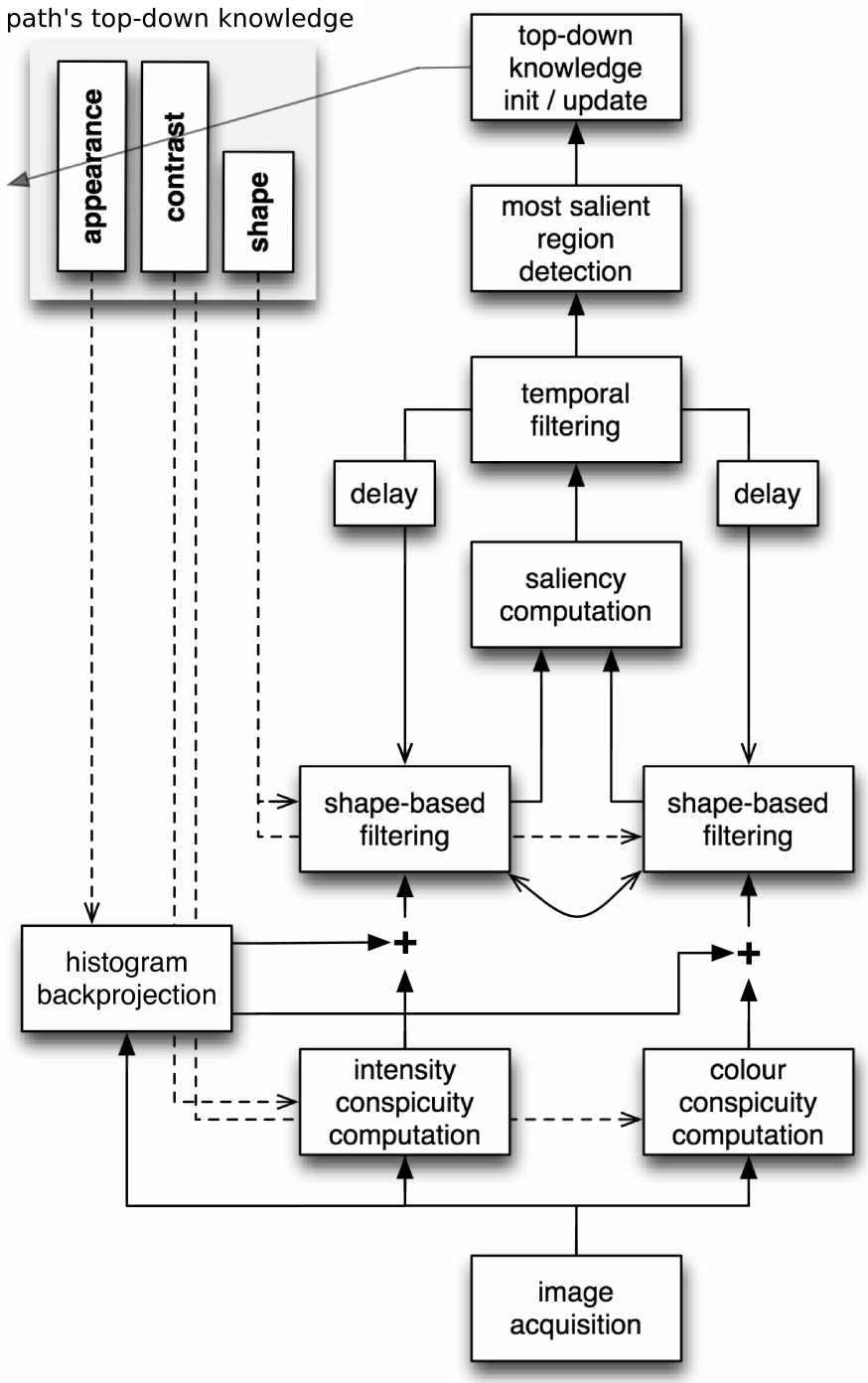


Figure 4.3: Proposed model's pipeline. The proposed model starts by computing two bottom-up conspicuity maps from the input image (intensity and colour information). After the initialisation phase, these maps are biased by the top-down contrast model and fused with an appearance-based probability map, computed from the top-down appearance model. The resulting top-down maps are shape-based filtered into two pheromone maps (one for colour and another for intensity information) by two swarms of virtual ants, respectively. The virtual ants use a set of behavioural rules that represent the top-down knowledge of the overall path's typical shape. These pheromone maps are initialised with the neural field's activity. Cross-modality, represented by the two-way arrow, are maintained by allowing swarms to share pheromone during their activity. Top-down appearance knowledge is used to modulate the level of pheromone deployed by the virtual ants. The two pheromone maps are fused together to generate a salience map. This latter map feeds the dynamical neural field, which performs temporal filtering. Finally, the most salient region is obtained as a mask to constrain the learning of both top-down appearance and contrast-based models.

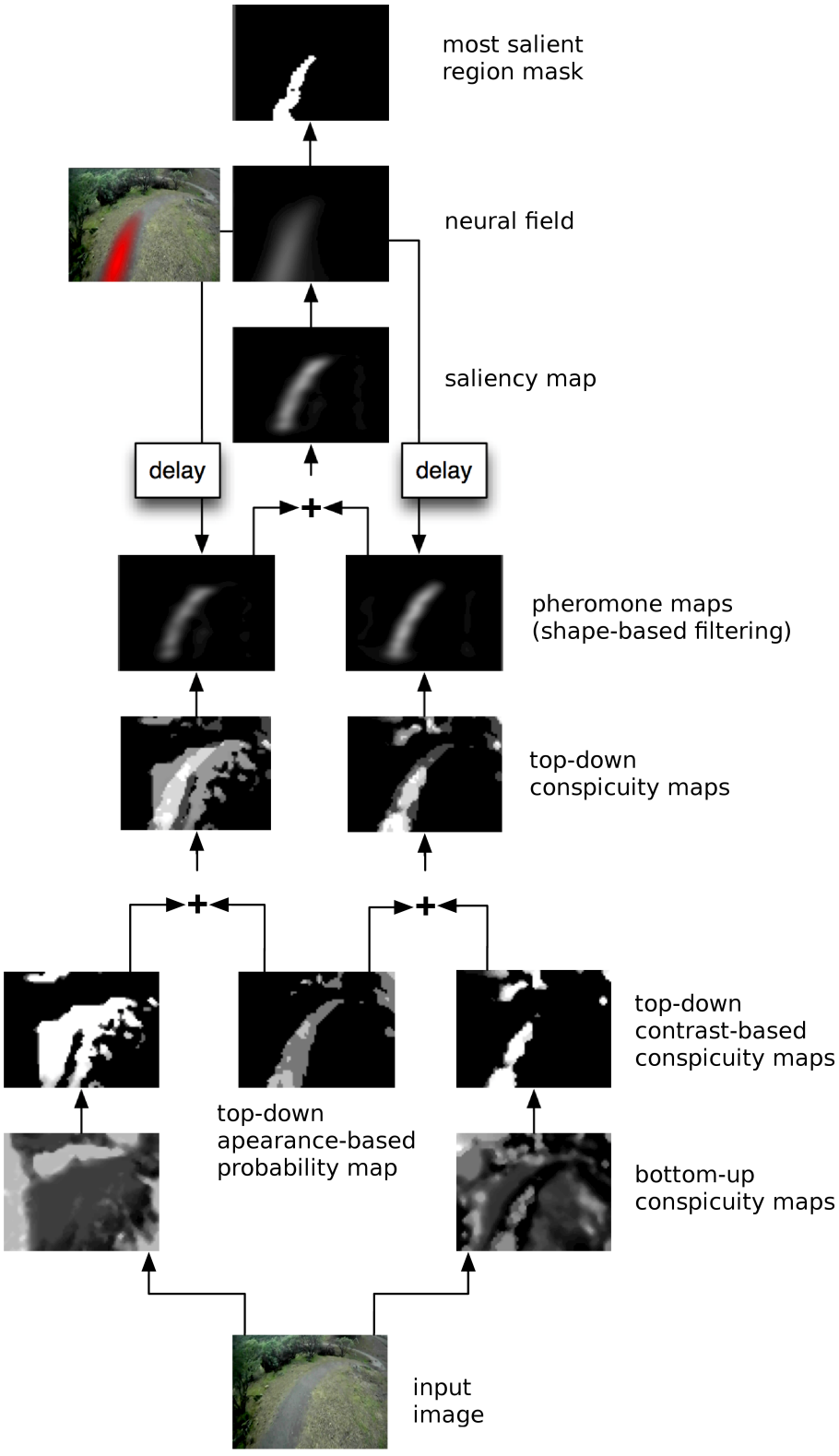


Figure 4.4: System's snapshot on a typical situation. Due to the environment's specific configuration, the bottom-up conspicuity maps, $\mathbf{C}_{bu}^{I}$ (left) and $\mathbf{C}_{bu}^{C}$ (right), are unable to unambiguously signal the presence of the path. By modulating the conspicuity process with the top-down contrast-based model learnt so far, path-background discrimination improves considerably. It is also possible to see that the back projection of the histogram (i.e., probability map, $\mathbf{A}$) used as top-down appearance-based model provides a discriminatory map that helps the swarms in performing the shape-based filtering. The good quality of the appearance-based model reflects the good correlation between the most salient region, $\mathbf{R}_{msr}$ of previous frames and the actual location of the path. The image on the left of the neural field, $\mathbf{F}$, corresponds to its overlay on the input image, $\mathbf{I}$.

Algorithm 2 Initialisation phase pseudo-code

Input: current frame $(\mathbf{I})$, previous frame $(\mathbf{I}')$, frame number (nr_frame)
Output: Neural field $(\mathbf{F})$
Data: Number of p-ants (n_ants) is an empirically defined constant. $\mathbf{C}_{temp}$ is an auxiliary image.

if $nr\_frame > 1$ **then**

use Equation 3.14 to estimate the homography matrix $\mathbf{H}$, from $\mathbf{I}$ and $\mathbf{I}'$
compensate neural field for robot motion, $\mathbf{F} \leftarrow \mathbf{H} \cdot \mathbf{F}$

end if

//initialise pheromone maps
$\mathbf{P}^I \leftarrow \lambda \cdot \mathbf{F}$;
$\mathbf{P}^I_* \leftarrow \emptyset$;
$\mathbf{P}^C \leftarrow \lambda \cdot \mathbf{F}$;
$\mathbf{P}^C_* \leftarrow \emptyset$;

compute bottom-up conspicuity maps, ($\mathbf{C}^I_{bu}$ and $\mathbf{C}^C_{bu}$), from $\mathbf{I}$ [see Section 3.3.2]

//update pheromone maps
for *n_ants* **do**

create p-ant p_I based on $\mathbf{C}^I_{bu}$ and $\mathbf{P}^I$ [see Section 3.3.3]
create p-ant p_C based on $\mathbf{C}^C_{bu}$ and $\mathbf{P}^C$

$(\mathbf{P^I}, \mathbf{P^C}, \mathbf{P^C_*}) \leftarrow$ **execute**$(p_I, \mathbf{C}^I_{bu}, \mathbf{P^C}, \mathbf{P^I}$ and $\mathbf{P^C_*})$ [see Algorithm 1]
$(\mathbf{P^C}, \mathbf{P^I}, \mathbf{P^I_*}) \leftarrow$ **execute**$(p_C, \mathbf{C}^C_{bu}, \mathbf{P^I}, \mathbf{P^C}$ and $\mathbf{P^I_*})$

remove(p_I, p_C)

end for

discard cross-modality and neural field influences, $(\mathbf{P}^I, \mathbf{P}^C) \leftarrow (\mathbf{P}^I_*, \mathbf{P}^C_*)$

compute salience map, $\mathbf{S} \leftarrow \frac{1}{2}\mathbf{P}^I + \frac{1}{2}\mathbf{P}^C$

update neural field $\mathbf{F}$ with $\mathbf{S}$ [see (Santana et al., 2010)]

$\mathbf{C}_{temp} \leftarrow \frac{1}{2} \cdot \mathbf{C}^I_{bu} + \frac{1}{2} \cdot \mathbf{C}^C_{bu}$

$(\mathbf{w}, \mathbf{h_{ref}}) \leftarrow$ **updateTopDownModels**$(\mathbf{I}, \mathbf{F}, \mathbf{w}, \mathbf{C}_{temp}, \mathbf{h_{ref}})$ [see Algorithm 5]

return $\mathbf{F}$

Algorithm 3 Tracking phase pseudo-code

Input: current frame $(\mathbf{I})$, previous frame $(\mathbf{I}')$
Output: Neural field $(\mathbf{F})$
Data: Number of p-ants (n_ants) is an empirically defined constant. $\mathbf{C}_{temp}$ is an auxiliary image.

use Equation 3.14 to estimate the homography matrix $\mathbf{H}$, from $\mathbf{I}$ and $\mathbf{I}'$

compensate neural field for robot motion, $\mathbf{F} \leftarrow \mathbf{H} \cdot \mathbf{F}$

//initialise pheromone maps
$\mathbf{P}^I \leftarrow \lambda \cdot \mathbf{F}$;
$\mathbf{P}^I_* \leftarrow \emptyset$;
$\mathbf{P}^C \leftarrow \lambda \cdot \mathbf{F}$;
$\mathbf{P}^C_* \leftarrow \emptyset$;

compute the probability map $\mathbf{A}$ from $\mathbf{I}$ and $\mathbf{h_{ref}}$ [see Section 4.5.2]

compute top-down contrast-based maps, $\mathbf{C}^I_w$ and $\mathbf{C}^C_w$, from $\mathbf{I}$ and $\mathbf{w}$ [see Section 4.5.1]

compute top-down conspicuity maps, $\mathbf{C}^I_{td}$ and $\mathbf{C}^C_{td}$, from $\mathbf{C}^I_w$,$\mathbf{C}^C_w$ and $\mathbf{A}$ [see Section 4.5.3]

//update pheromone maps
for *n_ants* **do**

  create p-ant p_I based on $\mathbf{C}^I_{td}$ and $\mathbf{P}^I$ [see Section 3.3.3]
  create p-ant p_C based on $\mathbf{C}^C_{td}$ and $\mathbf{P}^C$

  $(\mathbf{P^I}, \mathbf{P^C}, \mathbf{P^C_*}) \leftarrow$ **execute**$(p_I, \mathbf{C}^I_{td}, \mathbf{P^C}, \mathbf{P^I}, \mathbf{P^C_*})$ [see Algorithm 1]
  $(\mathbf{P^C}, \mathbf{P^I}, \mathbf{P^I_*}) \leftarrow$ **execute**$(p_C, \mathbf{C}^C_{td}, \mathbf{P^I}, \mathbf{P^C}, \mathbf{P^I_*})$

  remove(p_I, p_C)

end for

discard cross-modality and neural field influences, $(\mathbf{P}^I, \mathbf{P}^C) \leftarrow (\mathbf{P}^I_*, \mathbf{P}^C_*)$

compute salience map, $\mathbf{S} \leftarrow \frac{1}{2}\mathbf{P}^I + \frac{1}{2}\mathbf{P}^C$

update neural field $\mathbf{F}$ with $\mathbf{S}$ [see (Santana et al., 2010)]

$\mathbf{C}_{temp} \leftarrow \frac{1}{2} \cdot \mathbf{C}^I_w + \frac{1}{2} \cdot \mathbf{C}^C_w$

$(\mathbf{w}, \mathbf{h_{ref}}) \leftarrow$ **updateTopDownModels**$(\mathbf{I}, \mathbf{F}, \mathbf{w}, \mathbf{C}_{temp}, \mathbf{h_{ref}})$ [see Algorithm 5]

return $\mathbf{F}$

4.3 Top-down Knowledge Models

4.3.1 Appearance-based Model

The appearance model is used to promote the deployment of pheromone on the regions whose appearance is more similar to the one of the target path. To perform this image analysis, the appearance model is implemented as a histogram, as the latter classifies aspects of an image into discrete intervals to determine the correlation between images or features in an image. Shortly, a histogram is a function that returns the frequency of an intensity σ, i.e., the value of $\mathbf{h}(\sigma)$ is the number of pixels with intensity σ. Formally, being n_p the total number of pixels in a given image, and k_h the intensity intervals (or bins), the histogram $\mathbf{h}(.)$ meets the following condition:

$$n_p = \sum_{\sigma=0}^{k_h-1} \mathbf{h}(\sigma). \tag{4.1}$$

Furthermore, a normalised image histogram of a particular object, $\mathbf{h_n}(.)$, can be seen as a function giving the probability that a certain pixel belongs to this specific object. Formally, the probability of a pixel having an intensity of σ, $p(\sigma)$, in the 8-bit grayscale image $\mathbf{I}$, is defined as:

$$p(\sigma) = \mathbf{h_n}(\sigma) = \frac{\mathbf{h}(\sigma)}{n_p} \tag{4.2}$$

where $\mathbf{h}(.)$ is the non-normalised image histogram.

Concretely, the appearance model, $\mathbf{h_{ref}}$, is defined as a normalised three-dimensional 8-bit $16 \times 16 \times 16$ colour histogram. The use of 16 intervals for each dimension is empirically shown to better correlate a given image region with the information classified by $\mathbf{h_{ref}}$, i.e., the most likely path region. Being shadow invariant, the $c_1c_2c_3$ is used as the colour space that encodes the visual information, from which the model $\mathbf{h_{ref}}$ is updated (see Section 4.4).

4.3.2 Contrast-based Model

The top-down contrast-based model, $\mathbf{w}$, is defined as a vector of six elements, each representing the weight a given aggregate centre-surround feature map has to the detection of the path. Once the initialisation phase is over, these weights are updated and used at each frame to override the scaling function applied by default to the aggregate centre-surround maps (see Chapter 3). Table 4.1 lists the weights for the target path region delimited by a red rectangle, as depicted in Fig. 4.5 (a). This rectangular area is used herein only for the purpose of giving a simple example. As will be seen in the Section 4.4, the area used for updating $\mathbf{w}$, has a dynamic and complex shape. Fig. 4.5 also depicts the bottom-up conspicuity maps and the corresponding top-down contrast-based maps, computed using $\mathbf{w}$ (see Section 4.5.1).

Table 4.1: Computed weight vector $\mathbf{w}$ for the path depicted in Fig. 4.5 (a). The analysed path region corresponds to the area defined by the red rectangle in (a). These weights show that the learnt target's region is bright on a dark background (intensity) and more reddish with green background, whereas blue and yellow are less present than in the rest of the image and, thus, used for inhibition.

Centre-surround feature maps	Weights
Intensity On-Off	1.61601
Intensity Off-On	0.60327
Red-Green On-Off	2.27737
Red-Green Off-On	0.88842
Blue-Yellow On-Off	0.00000
Blue-Yellow Off-On	0.78312

(a) (b) (c) (d) (e)

Figure 4.5: Top-down contrast-based maps obtained from bottom-up feature maps, using the weight vector. (a) The input image with a hypothetical red rectangle marking the area used to compute the weight vector $\mathbf{w}$. (b) Bottom-up intensity conspicuity map, $\mathbf{C}_{bu}^{I}$. (c) Bottom-up colour conspicuity map, $\mathbf{C}_{bu}^{C}$. (d) Top-down intensity contrast map, $\mathbf{C}_{w}^{I}$. (e) Top-down colour contrast map, $\mathbf{C}_{w}^{C}$.

4.4 Learning Top-down Knowledge Models

To learn both top-down appearance and contrast models about the path being sought, it is necessary to specify, at each frame, the region of the input image that corresponds to the most likely location of the path.

As seen on Chapter 3, the high intensity region on the neural field represents a convergent behaviour that emerged from the p-ants' activity. Therefore, the highest intensity region of the neural field is more likely to belong to the path's location and, thus, it is used to find the most salient pixel on a temporary conspicuity map, $\mathbf{C}_{temp}$. During the initialisation phase, $\mathbf{C}_{temp}$ is computed as the average of both colour and intensity bottom-up conspicuity maps, $\mathbf{C}_{bu}^{I}$ and $\mathbf{C}_{bu}^{C}$, as the system does not acquired yet a top-down knowledge about the target path. In the tracking phase, $\mathbf{C}_{temp}$ is computed as the average of both colour and intensity top-down contrast-based maps, $\mathbf{C}_{w}^{C}$ and $\mathbf{C}_{w}^{I}$, respectively. Instead of searching in the whole $\mathbf{C}_{temp}$ map, the search is restricted to this region of interest defined

as $\mathbf{R}_{search}$ and, thus, computation time is saved.

Formally, the most salient pixel is found with the following expression:

$$(x_s, y_s) = \underset{(x,y)}{\arg\max}\{\mathbf{C}_{temp}(x, y)\}, \qquad \mathbf{R}_{search}(x, y) \neq 0 \tag{4.3}$$

where (x_s, y_s) is the location of the pixel with the highest intensity value found on $\mathbf{C}_{temp}$ and, simultaneously, belongs to a non-zero value of $\mathbf{R}_{search}$.

The location of the most salient pixel is used as a seed for a region growing process on $\mathbf{C}_{temp}$. The outcome of this process is a mask of the most salient region, $\mathbf{R}_{msr}$, which is provided to both appearance and contrast learning processes.

4.4.1 Most Salient Region Computation

The region growing process is herein implemented by a floodfill algorithm. Shortly, similarity between the seed and its eight immediate neighbours is verified. If a neighbour pixel meets the similarity criterion, it is labelled as belonging to the most salient region, $\mathbf{R}_{msr}$. After all eight neighbours have been processed, one of them, belonging to the seed's region, is chosen and its neighbours are checked against the similarity criterion. This process goes on until all pixels have been analysed or until all the neighbours of the seed's region do not meet the similarity criterion. Formally, this criterion is defined as:

$$\mathbf{C}_{temp}(x_s, y_s) - \delta \quad \leq \quad \mathbf{C}_{temp}(x, y) \quad \leq \quad \mathbf{C}_{temp}(x_s, y_s) + \delta, \tag{4.4}$$

where $\mathbf{C}_{temp}(x_s, y_s)$ is the seed's intensity, $\mathbf{C}_{temp}(x, y)$ is the intensity at location (x, y), and δ is the allowed intensity deviation. Fig. 4.6 depicts the obtained $\mathbf{R}_{msr}$ mask from a given input image. Algorithm 4 outlines the floodfill pseudo-code.

(a) (b) (c) (d) (e)

Figure 4.6: Search procedure used to find the most salient region $\mathbf{R}_{msr}$. (a) Temporary map $\mathbf{C}_{temp}$, computed in this example as the average of both colour and intensity bottom-up conspicuity maps. (b) Neural field $\mathbf{F}$. (c) Search mask $\mathbf{R}_{search}$. (d) Location of the highest intensity pixel p_s (green dot). (e) Most salient region $\mathbf{R}_{msr}$.

Algorithm 4 FloodFill pseudo-code

Input: Seed location (x_s, y_s), conspicuity map $(\mathbf{C}_{temp})$, Region mask $(\mathbf{R}_{msr})$
Output: Region mask, $\mathbf{R}_{msr}$
Data: Intensity threshold (δ), wait list $(\mathbf{Q})$

initialise list $\mathbf{Q}$ with the eight immediate neighbours of the seed pixel

while $\mathbf{Q} \neq \{\emptyset\}$ **do**

  get the next pixel's location (x_n, y_n) from the list $\mathbf{Q}$

  // comparison criterion
  if $(\mathbf{C}_{temp}(x_n, y_n) \leq \mathbf{C}_{temp}(x_s, y_s) + \delta)$ **and** $(\mathbf{C}_{temp}(x_n, y_n) \geq \mathbf{C}_{temp}(x_s, y_s) - \delta)$ **then**

    search to the left of (x_n, y_n) to find the location, (x_l, y_l), of a pixel that no longer matches the comparison criterion

    search to the right of (x_n, y_n) to find the location, (x_r, y_r), of a pixel that no longer matches the comparison criterion

    for all x_n *between* x_l *and* x_r **do**

      paint a white pixel at (x_n, y_n) on the mask $\mathbf{R}_{msr}$

      get the neighbour, (x_u, y_u), above (x_n, y_n)

      if $(\mathbf{C}_{temp}(x_u, y_u) \leq \mathbf{C}_{temp}(x_s, y_s) + \delta)$ **and** $(\mathbf{C}_{temp}(x_u, y_u) \geq \mathbf{C}_{temp}(x_s, y_s) - \delta)$ **then**

        add neighbour (x_u, y_u) to $\mathbf{Q}$

      end if

      get neighbour, (x_d, y_d), below (x_n, y_n)

      if $(\mathbf{C}_{temp}(x_d, y_d) \leq \mathbf{C}_{temp}(x_s, y_s) + \delta)$ **and** $(\mathbf{C}_{temp}(x_d, y_d) \geq \mathbf{C}_{temp}(x_s, y_s) - \delta)$ **then**

        add neighbour (x_d, y_d) to $\mathbf{Q}$

      end if

    end for

  end if

end while

return $\mathbf{R}_{msr}$

4.4.2 Appearance-based Model Update

The most salient region, $\mathbf{R}_{msr}$, is a key element in the update process of both appearance and contrast models, as non-path information can be filtered with the help of this image mask. Hence, to update the appearance model, the input RGB image is first converted to the $c_1c_2c_3$ colour space in the following way:

$$c_1 = \arctan\left(\frac{r}{\max(g,b)}\right), \tag{4.5}$$

$$c_2 = \arctan\left(\frac{g}{\max(r,b)}\right), \tag{4.6}$$

$$c_3 = \arctan\left(\frac{b}{\max(r,g)}\right). \tag{4.7}$$

The $c_1c_2c_3$ image is then pixel-wise multiplied by $\mathbf{R}_{msr}$. The resulting image region is used to build a histogram, $\mathbf{h_{sample}}$, that contains the new information about the path's appearance. Formally, this information is integrated in the appearance model $\mathbf{h_{ref}}$ according with the following expression:

$$\mathbf{h_{ref}} \leftarrow \mathbf{h_{ref}} \cdot (1-\beta_1) + \mathbf{h_{sample}} \cdot \beta_1, \tag{4.8}$$

where β_1 is the weight that the new information has in the $\mathbf{h_{ref}}$, i.e., the intrinsic adaptation speed of the appearance model. For instance, a higher β_1 value means a faster learning rate, but with the cost of a lower model's robustness to sudden misleading information.

4.4.3 Contrast-based Model Update

To update the contrast model $\mathbf{w}$, a weight vector $\mathbf{w}'$ is learnt for the current frame, based on the method proposed by Frintrop and Kessel (2009). Basically, this adapted method sets higher weights to bottom-up centre-surround maps that positively correlate with the most likely location of the object and lower weights to maps otherwise. This way, centre-surround maps are promoted according to their relevance to the object.

The weight of a centre-surround map is computed as the ratio between the average intensity of its pixels that simultaneously correspond to non-zero pixels in $\mathbf{R}_{msr}$, and the average intensity of the other pixels. Formally, the weight w_i of the centre-surround map, $\mathbf{C}_{s_i}$, is computed as

$$w_i = \frac{\frac{1}{m_1} \cdot \sum_{\mathbf{R}_{msr}(x,y)\neq 0} \mathbf{C}_{s_i}(x,y)}{\frac{1}{m_2} \cdot \sum_{\mathbf{R}_{msr}(x,y)=0} \mathbf{C}_{s_i}(x,y)} \qquad i \in \{1...6\}, \tag{4.9}$$

where m_1 and m_2 is the number of non-zero and zero pixels in $\mathbf{R}_{msr}$, respectively. However, it might happen that the most salient region mask, $\mathbf{R}_{msr}$, can partially cover path and non-path regions in the following two scenarios: (1) during the floodfill segmentation, the allowed intensity deviation value, δ, might be too loose; or (2) a misleading location of the seed, caused by high conspicuity of distractors. To increase robustness to sudden misleading information contained in $\mathbf{R}_{msr}$, the contrast knowledge model $\mathbf{w}$ has a smooth learning mechanism based on information learnt in the past. Hence, the adaptation of $\mathbf{w}$ is formulated as

$$\mathbf{w}(t) = (1-\beta_2)\mathbf{w}(t-1) + \beta_2 \mathbf{w}'(t), \tag{4.10}$$

where β_2 is the learning rate. Algorithm 5 outlines the appearance and contrast-based models update procedure.

Algorithm 5 Pseudo-code of the frame-wise top-down knowledge update process

1:
2: **Input:** Input image $(\mathbf{I})$, Neural field $(\mathbf{F})$, conspicuity map $(\mathbf{C}_{temp})$, weight vector $(\mathbf{w})$, histogram structure $(\mathbf{h_{ref}})$
3: **Output:** weight vector $(\mathbf{w})$, histogram structure $(\mathbf{h_{ref}})$
4:
5:
6: obtain neural field's $(\mathbf{F})$ highest activity region, $\mathbf{R}_{search}$
7:
8: find the coordinates (x_s, y_s) of the highest intensity pixel on $\mathbf{C}_{temp}$, using the mask $\mathbf{R}_{search}$
9:
10: $\mathbf{R}_{msr} \leftarrow$ **floodFill**$(x_s, y_s, \mathbf{C}_{temp})$ [see Algorithm 4]
11:
12: update weight vector $\mathbf{w}$ with $\mathbf{I}$ and $\mathbf{R}_{msr}$ [see Section 4.5.1]
13:
14: update top-down appearance-based model $\mathbf{h_{ref}}$ with $\mathbf{I}$ and $\mathbf{R}_{msr}$ [see Section 4.5.2]
15:
16: **return** $(\mathbf{w}, \mathbf{h_{ref}})$

4.5 Applying Top-down Knowledge Models

The weights of the vector $\mathbf{w}$ are used to excite or inhibit bottom-up visual features. The outcome of this procedure is two temporary top-down contrast-based maps, one for intensity information, $\mathbf{C}_w^I$, and another for colour information, $\mathbf{C}_w^C$. As stated before, each one of these maps are fused with the appearance-based probability map $\mathbf{A}$ to create two top-down intensity and colour conspicuity maps, $\mathbf{C}_{td}^I$ and $\mathbf{C}_{td}^C$, both based on contrast and appearance knowledge.

4.5.1 Contrast-based Maps Computation

Each top-down contrast-based map is obtained by subtracting an inhibition map from an excitation map, $\mathbf{E}_w - \mathbf{I}_w$. The excitation map $\mathbf{E}_w$ is the weighted sum of the bottom-up centre-surround maps that weight more than unity (important contribution for the target being sought), whereas the inhibition map $\mathbf{I}_w$ is the weighted sum of the maps that weight less than unity (not important). Centre-surround maps whose weight is equal to unity indicate that the mean salience of the target region is exactly the same as the mean salience of the background. Therefore, these maps have a zero contribution for the target salience and are discarded. Formally, the excitation and inhibition maps are computed from a set of n_{cs} centre-surround maps, according to the following two expressions:

$$\mathbf{E}_w = \sum_i (\omega_i \cdot \mathbf{C}_{s_i}) \qquad \forall i \in \{1..n_{cs}\} : \omega_i > 1; \tag{4.11}$$

$$\mathbf{I}_w = \sum_i \left(\frac{\mathbf{C}_{s_i}}{\omega_i} \right) \qquad \forall i \in \{1..n_{cs}\} : \omega_i < 1. \tag{4.12}$$

Specifically, the top-down intensity contrast-based map, $\mathbf{C}_w^I$, is computed using the weights of the on-off and off-on centre-surround intensity maps, whereas the weights of the four colour-opponency centre-surround maps are used to compute $\mathbf{C}_w^C$. The two contrast biased maps are then normalised with the operator $K(.)$, described in Chapter 3. Fig. 4.7 depicts the top-down contrast map computation process.

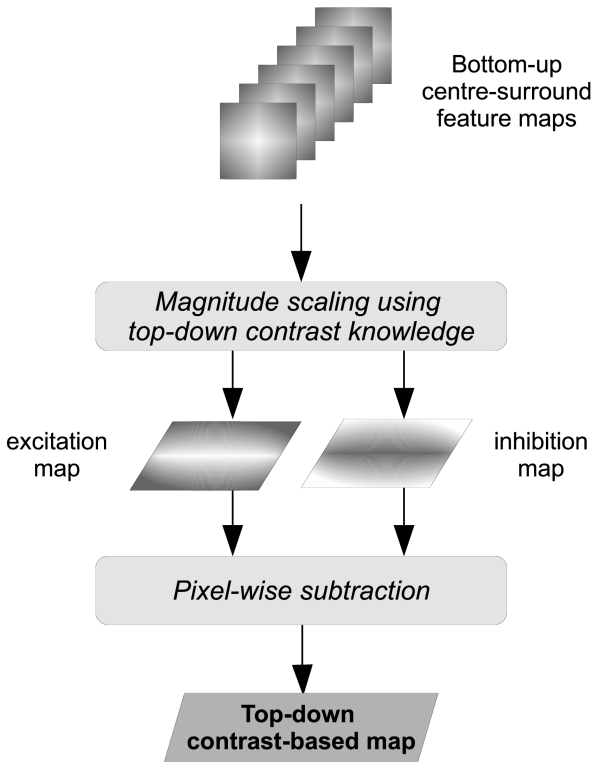


Figure 4.7: Top-down contrast-based conspicuity computation.

4.5.2 Appearance-based Probability Map Computation

The appearance-based probability map, $\mathbf{A}$, is obtained by performing the back projection of the normalised histogram, $\mathbf{h_{ref}}$. Concretely, this method computes a probability map by labelling each one of its pixels accordingly to the likelihood between the corresponding input image's pixels and the path region, which is classified by $\mathbf{h_{ref}}$. Hence, a white pixel in the appearance-based probability map, $\mathbf{A}$, means a full match whereas a black one means a full mismatch (see Fig. 4.8).

The appearance-based probability map helps the tracking of the path by the following two ways: (1) by its fusion with the top-down contrast-based maps to produce the final top-down conspicuity maps and, (2) by biasing the level of pheromone deployed by the virtual ants. In particular, the modulation of the pheromone deployment is implemented in the proposed model, as follows:

$$\Phi(p_m) = \epsilon + \beta \cdot p(T|V_{p_m}, \mathbf{A}) \tag{4.13}$$

where β is an empirically defined weighting factor, ϵ is an empirically defined pheromone level baseline, and $p(T|V_{p_m}, \mathbf{A})$ is the probability of the p-ant's path, V_{p_m}, to belong to the path T, given the information contained in the top-down appearance-based map $\mathbf{A}$. The probability $p(T|V_{p_m}, \mathbf{A})$ is approximated by the average probability computed by taking into account the probability of each pixel, visited by the ant, of belong to the path region. These pixels are represented by the set V_{p_m}, and their individual probabilities are obtained from the top-down appearance-based probability map, $\mathbf{A}$.

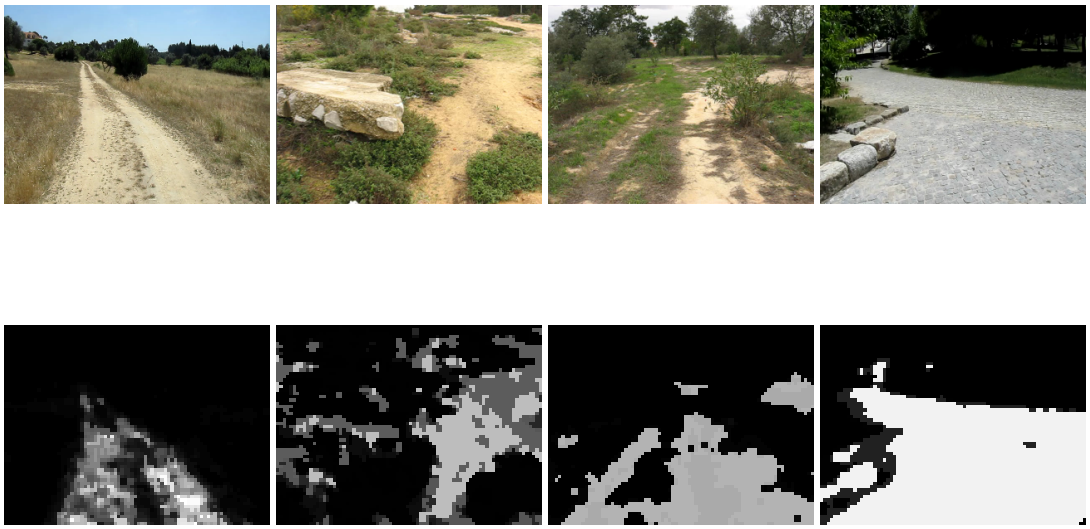


Figure 4.8: Appearance-based probability maps (bottom-row), and corresponding input images (top-row) for path detection. The probability maps are computed using the histogram back projection method, after the colour conversion process of the input RGB image to the $c_1c_2c_3$ colour space.

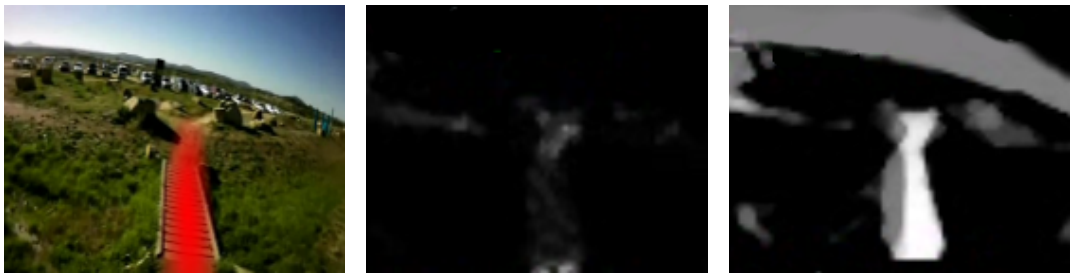
4.5.3 Top-down Conspicuity Maps Computation

Although useful, the top-down appearance-based model alone is unable to segment the path when its appearance suffers a sudden change, as in the situation depicted in Fig. 4.9. Hence, to overcome this failure case, the probability map, $\mathbf{A}$, is superposed with each top-down contrast-based map, $\mathbf{C}_w^I$ and $\mathbf{C}_w^C$, obtaining two top-down conspicuity maps, $\mathbf{C}_{td}^I$ and $\mathbf{C}_{td}^C$, as follows:

$$\mathbf{C}_{td}^I = \frac{1}{2} \cdot \mathbf{C}_w^I + \frac{1}{2} \cdot \mathbf{A}, \tag{4.14}$$

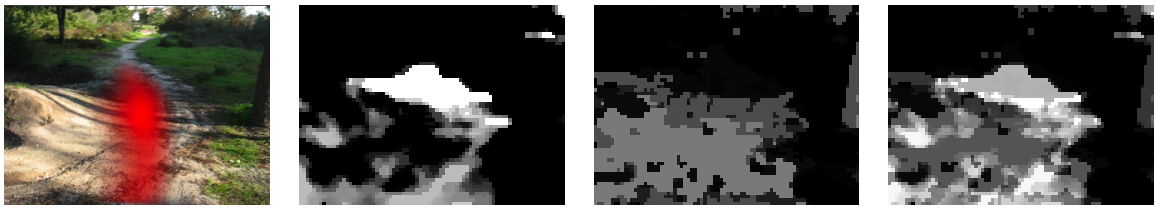
$$\mathbf{C}_{td}^C = \frac{1}{2} \cdot \mathbf{C}_w^C + \frac{1}{2} \cdot \mathbf{A}. \tag{4.15}$$

The outcome are maps with a reduced number of gaps which helps p-ants to track the path's skeleton in a more robust way (see Fig. 4.10).



(a) input image (b) back-projection map (c) top-down contrast map

Figure 4.9: Typical situation in which the back projection of the appearance model, depicted in (b), is insufficient to segment the path from the background. Conversely, the top-down colour contrast-based map, depicted in (c), is able to accurately localise the path in the input image. The failure of the top-down appearance model owes mostly to the sudden appearance of a bridge along the path.



(a) (b) (c) (d)

Figure 4.10: Typical situation in which the superposition of the top-down colour contrast-based map, $\mathbf{C}_w^C$, and the appearance model's probability map, $\mathbf{A}$, produces a top-down colour conspicuity map, $\mathbf{C}_{td}^C$, with a reduced number of gaps, thus facilitating the swarm operation. (a) Input image with system's output overlaid in red. (b) Top-down colour contrast-based map of (a), $\mathbf{C}_{td}^C$. Top-down appearance-based probability map. Superposition of (b) and (c).

Chapter 5

Experimental Results

This chapter presents the experimental setup and parametrisation used to assert the robustness and efficiency of the proposed model, as well as the obtained results. The failures cases of the proposed model and the discussion of the results, are described in Section 5.3.

5.1 Experimental Setup

The proposed model was implemented entirely in the C++ programming language and it was made fully compliant with the Robotics Operating System (ROS)[1] (Quigley et al., 2009). The system was tested in a Pentium(R) Dual-Core CPU T4300 2.10GHz with 4 Gb of RAM, running a 32-bit Linux distribution Ubuntu 10.10 (Maverick Meerkat), and using OpenCV 2.3 (Bradski and Kaehler, 2008) for low-level computer vision routines.

In order to measure the performance of the proposed model, an extensive data-set of 39 colour videos, encompassing a total of 29789 analysed frames with a resolution of 640×480, has been obtained with a hand-handled camera carried at an approximate height of 1.5 m and speed of 1 ms^{-1}. The dataset includes both natural and engineered paths in a wide variety of backgrounds (see Fig. 5.1). Experimental results were obtained running the model off-line.

5.2 Model Parametrisation

The proposed model's parametrisation follows practically the same values of the free parameters defined in the original model (Santana et al., 2010). Concretely, the number of p-ants deployed per map, n, has been empirically defined to 20. A smaller number may not ensure convergence, whereas a larger one did not exhibit considerable improvement in the tested data set. The same reasoning applies to the number of iterations applied to each p-ant, η_1, which has been set to 50. The

[1] ROS: `http://www.ros.org`

Figure 5.1: Data set representative frames. Each image corresponds to one video whose ID is given by increasing order from left to right and top to bottom. The redness of the blobs overlaid in the images correspond to the activity level of the neural field above $85\,\%$ of its maximum, representing the model's estimate of the path's location.

pheromone baseline deployed by a given p-ant on its associated pheromone map, ϵ, has been set to 2.0. The gain of the top-down appearance-based model's contribution to the deployed pheromone, β, is set to 2.0. The small portion of ϵ deployed in the other pheromone map, υ, has been set to 0.3. These values should not be set too high to avoid pheromone saturation, inhibiting the emergence of collective behaviour. The learning rates of both top-down appearance-based and contrast-based models, β_1 and β_2, are set to 0.1. The ratio of the robot motion compensated neural field used to initialise the pheromone maps at the onset of each frame has been set to $\lambda = 0.1$.

The p-ants are set to be more greedy in searching for high conspicuity regions or regions with similar average level of conspicuity, learnt from its past iterations. Beyond that, the p-ants behave with a little tendency to maintain equidistant to the boundaries of the path hypothesis. Hence, the contribution of each behaviour is $\alpha_{greedy} = 0.45$, $\alpha_{track} = 0.35$, $\alpha_{centre} = 0.10$, $\alpha_{ahead} = 0.05$, and $\alpha_{commit} = 0.05$. The goal of making $\alpha_{greedy} > \alpha_{centre} + \alpha_{ahead} + \alpha_{commit}$, is to ensure that p-ants exploit more strongly the conspicuity cue than the a priori knowledge on the expected path's shape. With a relatively high α_{track}, the swarm influences each individual p-ant to further reduce the problems associated with noise and distractors.

The width, δ_w, and the height, δ_h, of the window used to create p-ants (see Chapter 3) have been set to 9 and 5, respectively. The initial values of the random factor ρ (see Equation 3.10), ρ_0, and its increment at each iteration, $\Delta\rho$, have been set to 0.3 and 0.02, respectively. The initial values of the random factor γ (see Equation 3.11), γ_0, and the rate of its exponential decay at each iteration, γ_τ, have been set to 0.4 and 0.02, respectively. The number of input frames spent in the initialisation phase was defined empirically to $\eta_2 = 50$, which is empirically shown to be sufficient for the system to converge.

5.3 Results

This section presents the quantitative results obtained with the proposed model in the presented data set. Moreover, the following hypotheses are herein assessed:

1. Added value of the top-down appearance-based model;
2. Advantageous modulation of bottom-up visual feature cues into top-down contrast-based maps;
3. Increased path detection rate due to top-down knowledge;
4. Computational efficiency to ensure robust path following in off-road environments.

As seen on Chapter 4, the dynamic pheromone deployment process is biased by an appearance-based probability map that is learnt and updated on-line. Fig. 5.2 depicts several probability maps with a high probability value on path regions, confirming the first hypothesis.

As discussed in Chapter 3, in extreme situations where the path conspicuity is constantly low and noisy, the neural field inertia isn't enough, migrating slowly and together with the p-ants, to the

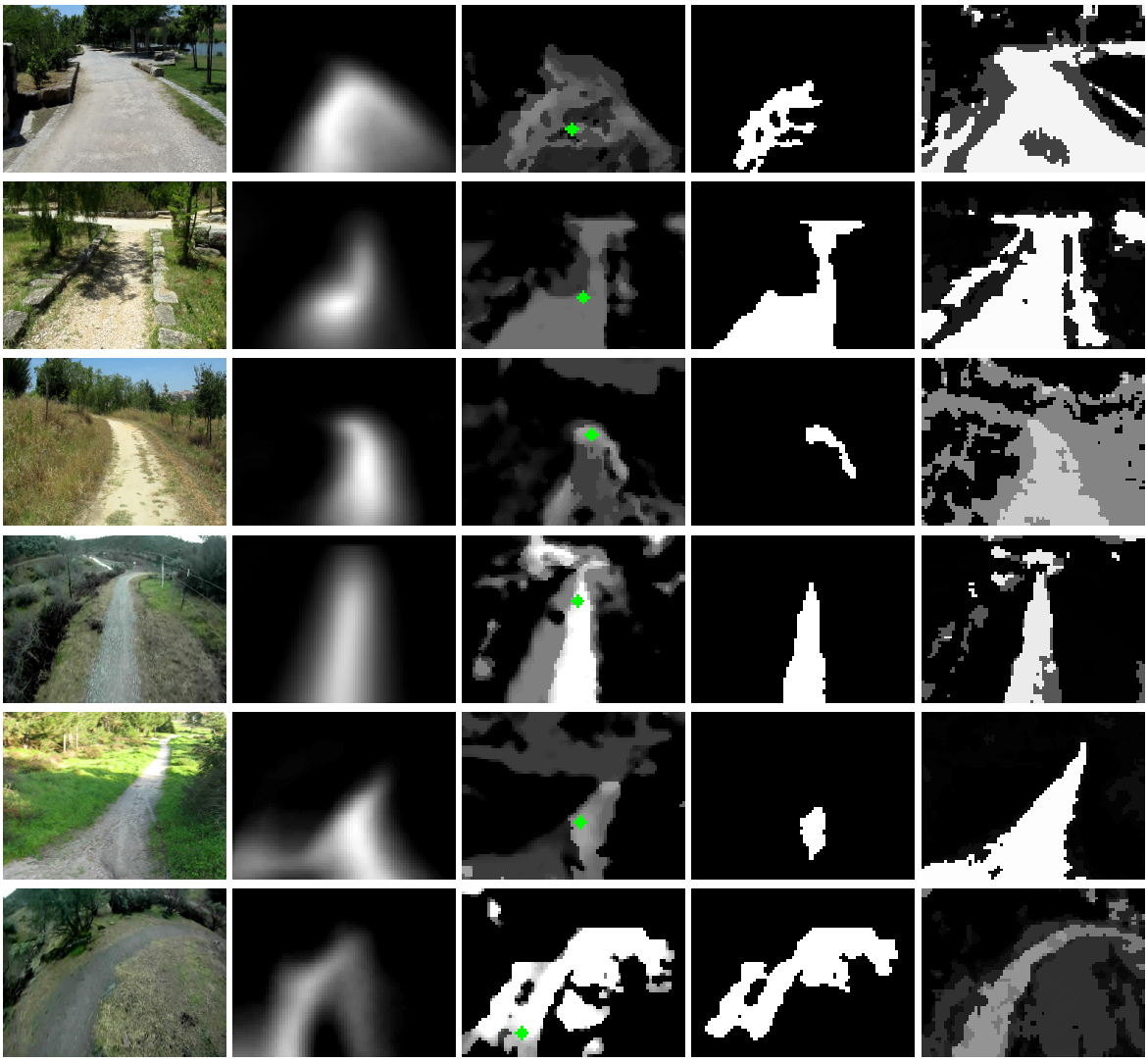


Figure 5.2: Evaluation of path segmentation based on the probability maps. Input image (first column), neural field $\mathbf{F}$ (second column), seed location marked by a green diamond on $\mathbf{C}_{temp}$ map (third column), most salient region mask $\mathbf{R}_{msr}$ (fourth column), and the corresponding probability map $\mathbf{I}_p$ (fifth column) . These figures show that path region is entirely (or partially) segmented on probability maps.

path' surroundings. These scenarios are not so prone to happen with new proposed model. This is achieved by the top-down contrast-based model, that allows the computation of top-down contrast-based maps less noisy than the corresponding bottom-up conspicuity maps. Moreover, as seen in Section 4.4.3, the contrast model has a smooth learning mechanism that gives importance to the past learnt information. Hence, it offers some robustness to sudden and misleading temporary changes on bottom-up conspicuity maps. Fig. 4.9 and Fig. 4.5 confirm the second hypothesis, as they illustrate the added value of the top-down contrast-based model.

The third hypothesis being tested in this dissertation is that the interaction between the bottom-up conspicuity maps and the top-down knowledge models are able to introduce the required added value to ensure robust path detection and tracking. Concretely, the path is considered correctly detected if the biggest blob of neural field activity (above $85\,\%$ of its maximum) is fully localised within the path's boundaries and roughly aligned with the path's orientation. Additionally, the whole neural field's activity can be taken as an approximation of the path/background segmentation. Such representation may be useful for detailed motion planning. Although a quantitative analysis is not

provided, the obtained qualitative results support this possibility - see the high correlation between neural field activity and path location in Fig. 5.2.

The results obtained from the data-set videos are shown in Table 5.1 to confirm the success of the fourth hypothesis. That is, in the tested data set, the proposed adaptive top-down swarm-based salience model predicts the path location 5 times more than a classical salience model based only on the bottom-up conspicuity maps. For the sake of fair comparison, the neural field $\mathbf{F}$, which is fed by $\mathbf{S}$, is used to generate the output on the different models. To assess the potential effects the probabilistic nature of p-ants behaviours might have in the model's repeatability, the results are obtained from averaging 5 runs per video. For each video run, the number of frames where the path is correctly detected is divided by the total number of frames. With a rather low standard deviation (see Table A.1), model's repeatability is demonstrated.

Table 5.1: Comparative detection results in the 39 data-set videos between several models: a classical salience model based only on the conspicuity maps, $\mathbf{S} \leftarrow \frac{1}{2}(\mathbf{C}^I) + \frac{1}{2}(\mathbf{C}^C)$; the original model (Santana et al., 2010) based on pheromone maps, $\mathbf{S} \leftarrow \frac{1}{2}(\mathbf{P}^I) + \frac{1}{2}(\mathbf{P}^C)$; and, the proposed model. Aggregate detection rate (mean $\pm$ standard deviation) computed as the average of the detection rates obtained per video. Refer to Table A.1 in Appendix A for details.

	Classical model (Itti et al., 1998)	Original model (Santana et al., 2010)	Proposed model
Detection rate [%]	19.60 ± 0.0	84.66 ± 0.14	98.67 ± 0.04

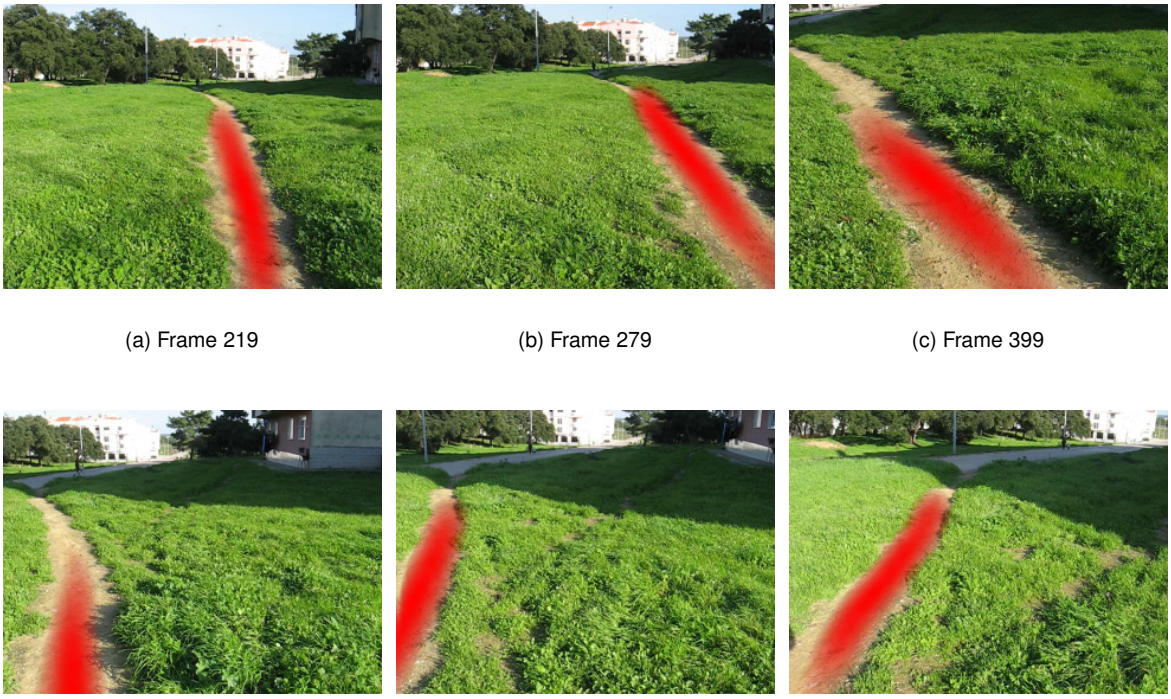
It is worth noting that in 28 of the 39 videos, the proposed model shows $100\,\%$ success rate, for all the 5 five runs. Videos 5, 27, 36, and 39 are accounted as long runs, above 4 minutes length, composed by more than 2500 frames. The $100\,\%$ success rate of the model in these videos (except for video 27 with a $99.47\,\%$ success rate) clearly shows its robustness in demanding situations.

To show that the model is capable of providing sufficient information to bring a robot back to the path after a forced exit (e.g., in the presence of an obstacle), the camera was frequently moved off path with a considerable level of oscillation. Fig. 5.3 depicts one of these situations.

The fourth hypothesis being assessed is whether the proposed model exhibits computational efficiency enough to ensure robust path following in off-road environments. Computational efficiency is attained with $20\,$Hz operation (see Table 5.2). Remarkably, the swarm-based pheromone maps computation, which is the only path-specific operation, takes only $2\,$ms on average per frame. With an average success rate of $98.67\,\%$ (see Table 5.1), the model shows itself capable of providing a great deal of, and possibly sufficient, information for a control system to guide a robot along most paths.

Table 5.2: Average computation times. The timing reported for the neural field update also includes optical flow computation, homography estimation and neural field wrapping.

	Neural field	Conspicuity maps computation	Pheromone maps computation	Total
Time [ms]	18	33	2	53



(a) Frame 219 (b) Frame 279 (c) Frame 399

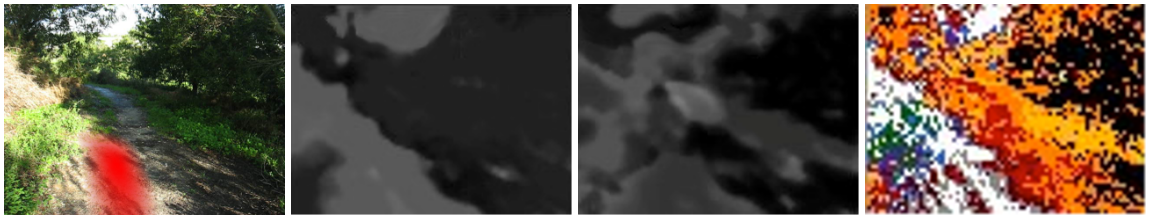
(d) Frame 447 (e) Frame 501 (f) Frame 546

Figure 5.3: System's output in a sequence of images from video 19 obtained with the camera moving on and off path. This shows the ability of the model to provide enough information about the path's location for a robot to return to the path after a forced exit. The redness of the pheromone cloud overlaid in the images correspond to the activity level of the neural field above $85\,\%$ of its maximum, representing the model's estimate of the path's location.

5.3.1 Failure Cases

Despite the overall good results, the model still presents some weaknesses that should be addressed in future work. As seen on Section 4.1, to reduce sensitivity to shadows, a different colour space than RGB was used to build the probability map, $\mathbf{A}$. However, strong shadows remove chromatic information, thus limiting the impact of the colour space. Fig. 5.4 illustrates one of these situations, along with a segmentation produced by a classic clustering-based approach (Rasmussen et al., 2009). The figure shows that the conspicuity maps produce segmentations similar to the clustering approach. This confirms the ability of salience maps to produce interesting segmentations of the input image. It also shows the difficulty classical segmentation approaches also have in handling strong shadows. Hence, this failure can be credited more to the poor signal-to-noise ratio of the input image than to any limitation of the proposed model. Given that shadows are cast mostly by tall objects, the fix to this problem may lie in the inclusion of 3-D information.

Sometimes a temporary discompensation of the neural field for the robot motion, caused by a failure in recovering the optic-flow, can temporarily mislead the swarm activity to non-path regions. In particular, this failure results in a concentration of neural field activity off the path and a strong competition between p-ants on and off the path, which can lasts for a few frames until symmetry is broken. For instance, p-ants responsible for the off path activity tend to rapidly converge to path regions due to "cleaner and accurate" top-down conspicuity maps that may suppress misleading



(a) Input image (b) Consp. map $\mathbf{C}_{bu}^{I}$ (c) Consp. map $\mathbf{C}_{bu}^{C}$ (d) Image segmentation

Figure 5.4: Failure caused by strong shadows. Situation in which the proposed model fails to track the path due to the presence of strong shadows. Note that conspicuity maps in (b)-(c) are themselves producing a segmentation of the input image similar to the one produced by a classical clustering-based segmentation approach (Rasmussen et al., 2009) in (d). Hence, failure is as likely in both cases.

non-path competition. But as one can infer, when the concentration of neural field activity is off the path, it can jeopardise the path's appearance and contrast learning due to a sudden misleading seed location, and consequently, threatening and compromising the integrity of top-down conspicuity maps. However, giving importance to past learnt contrast information, it offers some protection in these failure cases. Fig. 5.5 depicts a sequence of images from video 30, that shows the model's ability to recover from dramatic mismatches between the representation built so far and the sensory input.

5.3.2 Discussion

A key issue in the proposed model is the considerably high number of parameters that must be set. However, a single parametrisation is robust enough to cope with different situations. This ability can be verified by the success rate above $98\,\%$ in videos 26, 27, and 28, which include natural and engineered paths in both natural and man-made environments.

When the path is highly conspicuous in the environment, as most often occurs, ambiguity is rarely present in the system's output. When this assumption fails, and distractors are scattered, the model is still able to often perform correctly, as demonstrated by the quantitative results. This robustness owes to the synergistic operation between neural field inertia and p-ants' sensorimotor coordination capabilities, which allow an opportunistic exploitation of the path-background prioritised segmentation present in the conspicuity maps. Although the top-down appearance-based model is also responsible for this success, it is possible to depict in Fig. 5.6 that it is insufficient alone when the camera is compelled to change between paths of different appearance. In the same line, Fig 5.7 shows two additional situations in which the system successfully tracks the path despite its sudden appearance change.

As seen by the obtained results, the proposed model exhibited computational efficiency enough to ensure robust path following in off-road environments, as well as the capability to bring a robot back to the path after a forced exit.

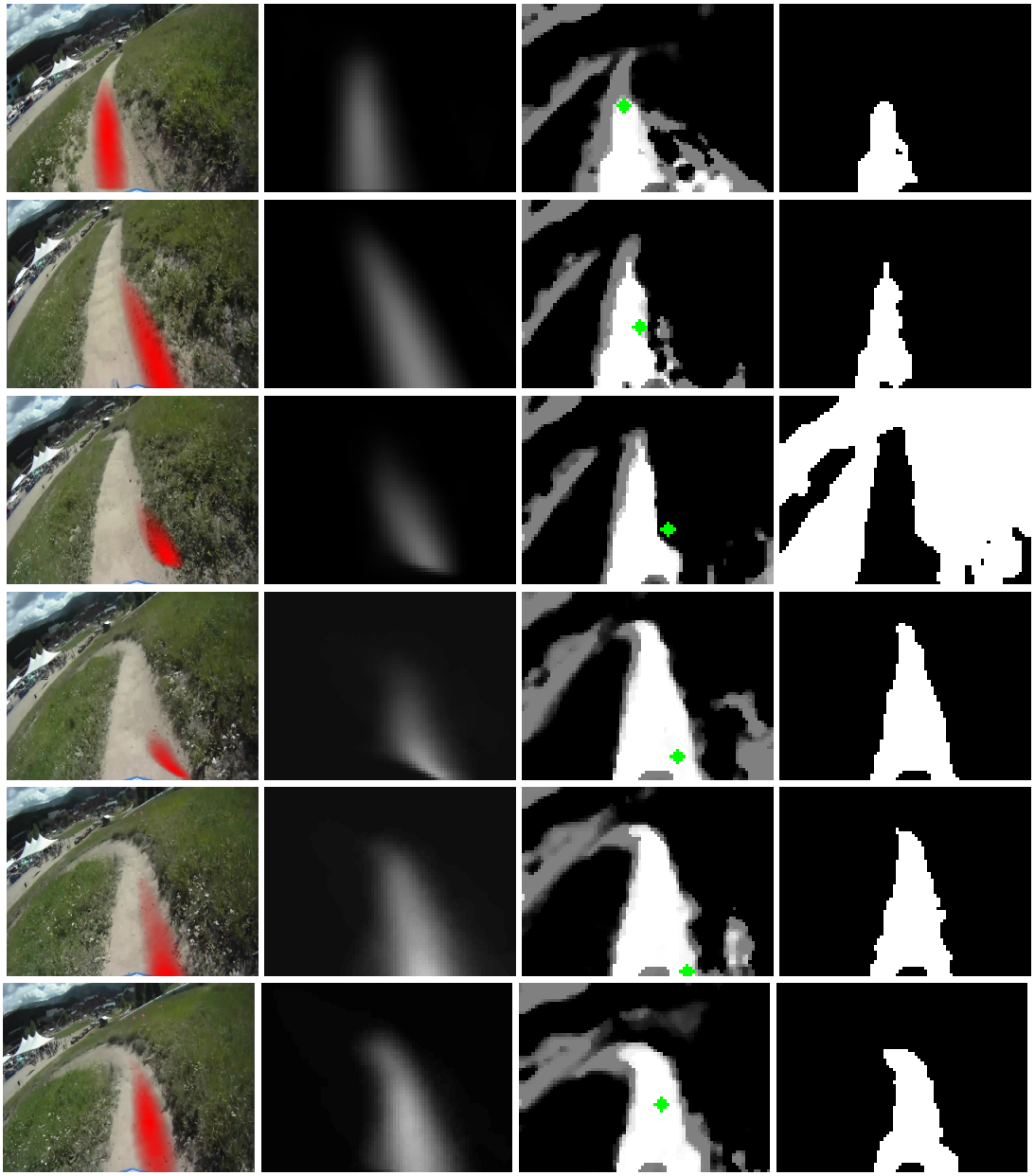


Figure 5.5: Misleading learning due to bad seed location caused by erroneous motion compensation. Situation in which the proposed model overcomes the failure of optic-flow and was able to track the path. From top to bottom, rows represent the model's state for key frames along the sequence. The figure shows that the smooth learning mechanism, gives importance to the past, boosting system robustness. First column shows the sequence of input frames. On the second, third and fourth columns, the obtained neural field, the most salient pixel and the most salient region mask are shown, respectively.

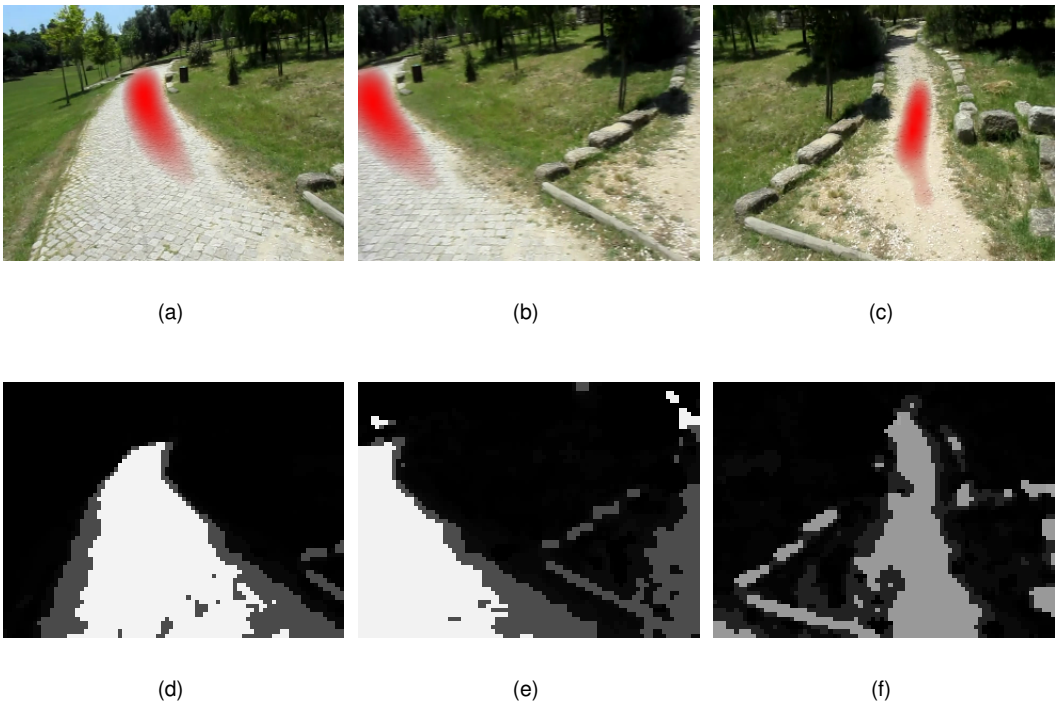


Figure 5.6: System's output in a sequence of images from video 27 obtained with the camera moving from the path depicted in (a) to the path depicted in (c), and the corresponding appearance model's back-projections (d) and (f). In (e), the pixel-wise path probability map of image (b) shows the inability of the learnt appearance model to indicate the presence of the new path. This is a result of both paths having different appearances. Nevertheless, the system is able to switch from one path to the other thanks to the contrast model. The appearance model eventually learns the new appearance of the path, as seen in (f).

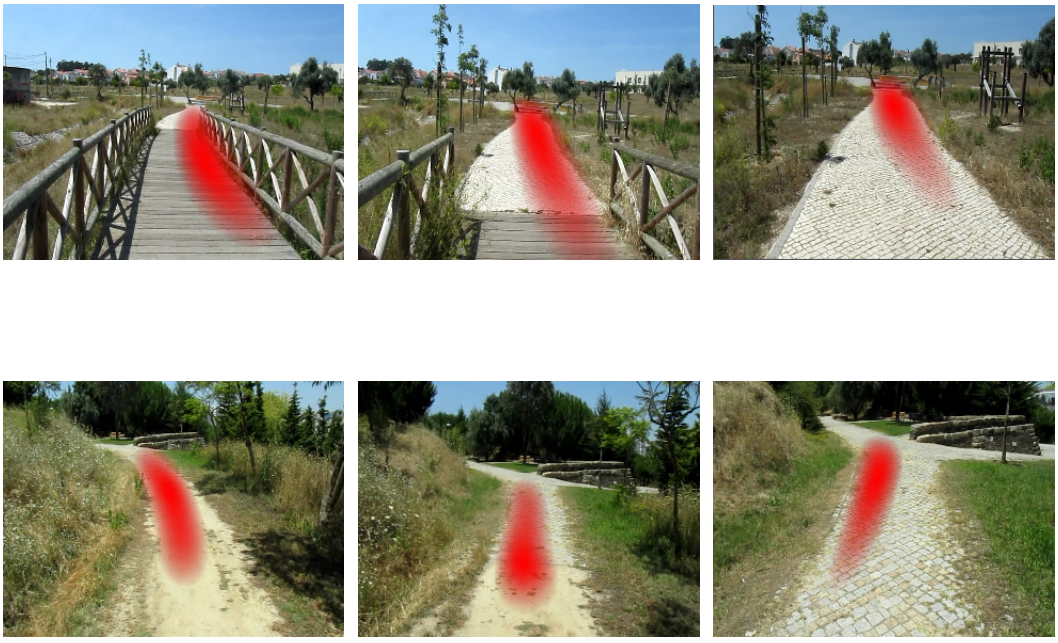


Figure 5.7: System's output in two sequences of images (top row from video 26 and bottom row from video 27) obtained with the camera moving along two paths whose appearance suddenly changes.

Chapter 6

Conclusions and Future Work

In this chapter, a summary of the results achieved in this dissertation is given, as well as some directions for future research.

6.1 Conclusions

This dissertation proposes the incorporation of an adaptive pheromone deployment and a top-down knowledge model into the swarm-based bottom-up visual attention model proposed by Santana et al. (2010). The goal of this extension is to improve the original model, as the swarm-based design developed by Santana et al. (2010) shows a fitted and adequate visual search, perceptual grouping, and multiple hypotheses tracking. In particular, as the original model uses visual salience both local and global cues on the path location are naturally exploited. That is, visual salience maps provide contrast information not only between the path and its surroundings, but also between the path and the overall scene. For instance, the appearance of the materials composing the path (e.g., soil) differ from the appearance of the materials composing its surroundings (e.g., grass, sidewalks) and the remainder elements present in overall scene (e.g., buildings, trees, sky). The complementarity of both local and global contrast information results in a robust handling of less structured trails.

As seen on the original model (Santana et al., 2010), two swarms of agents sensitive to bottom-up conspicuity information interact via pheromone-like signals so as to converge on the most likely location of the path being sought. The behaviours ruling the agents' motion are composed of a set of perception-action rules that embed top-down knowledge about the path's overall layout. This reduces ambiguity in the face of distractors. However, distractors with a shape similar to the one of the path being sought can still misguide the system. To mitigate this issue, the introduced top-down knowledge model consists of three components: contrast, appearance, and a priori shape knowledge. Using in a modulation context and not in a direct image processing, the complexity of these components can be reduced without hampering robustness. Hence, the appearance component was implemented by a simple $c_1c_2c_3$ colour space histogram that is used to build an appearance-based map about the path location and to modulate the pheromone deployment. The contrast component was implemented by a weight vector that modulates bottom-up visual conspicuity information into top-down contrast-

based maps. As already stated, the shape knowledge was implemented by the set of behaviours of each swarm agent.

The proposed model's architecture exploited synergistically contrast, appearance and shape information, which is essential to handle sudden path's appearance changes. In particular, it allows to further increase robustness and reliability of the original model by on-line learning of contrast and appearance top-down information of the path being followed.

As seen by the experimental results presented in Chapter 5, the model was successfully validated against a heterogeneous and challenging data-set, being capable of handling highly unstructured paths in natural environments, and exhibiting a low computational footprint. Moreover, low cost computation allows a faster robot motion. In particular, the model exhibited $98.67\,\%$ of success rate at 20 Hz, outperforming the previous non-adaptive model Santana et al. (2010), with $84.66\,\%$ success rate at 20 Hz. The proposed model was able to successfully track the path in situations which its sudden appearance changed (see Fig. 5.6). Furthermore, it was shown the ability of the model to provide enough information about the path's location for a robot to return to the path after a forced exit.

Another positive result that worth mention was that the selected model's parametrisation was not over-fit to a specific natural environment as can be seen by the high success rate across the diverse data-set, and thus highlighting its robustness.

Finally, proposed model contributes to the emerging swarm cognition field, which attempts to uncover the basic principles of cognition, i.e., adaptive behaviour, recurring to self-organising principles, mainly those exhibited by social insects.

6.2 Future Work

An interesting future development is the self-supervised adaptation of the weights of each p-ant's behaviour to achieve better results in different scenarios, providing an evolutive capability to the system. Another aspect to be taken into account is the creation of different expertises among the p-ants. For instance, one p-ant may be expert in finding path borders, whereas others may be expert in finding the horizon, keeping the p-ants away from exploring the sky. The addition of three-dimensional information, as an obstacle detection process, can helpfully bias the p-ants' motion away of obstacles.

To further increase the robustness of the proposed model, other visual features can be exploited, such as orientations. However, these features can increase significantly the processing time for each frame. Therefore, another interesting development would be the use of modern Graphics Processing Units (GPUs), as they are becoming more efficient and faster at performing calculations involving matrix and vector operations. Their highly parallel hardware capabilities makes them more effective than general-purpose CPUs for algorithms where processing of data is or can be done in parallel. Therefore, the parallel nature of the proposed model can be used to greatly improved computation time in future GPU-accelerated implementations, freeing more CPU resources for other tasks.

As the original model, this extended model does not handle bifurcations in the path. To overcome this limitation in the future, multiple focus of p-ants activity must be analysed and tracked in the neural field. The selection of which focus to track and, thus, which way the robot must proceed must be hinted by some external stimuli, like a directional signal on the path.

A open challenge in the proposed model is the autonomous assessment of the most propitious moment to automatically switch from initialisation to tracking phase.

6.3 Dissemination

Some of the concepts covered in this dissertation can be additionally viewed in the following publication, co-authored by the author:

- Santana, P., Mendonça, R., Correia, L., and Barata, J. (2011). Swarms for robot vision: The case of adaptive visual trail detection and tracking. *Advances in Artificial Life, ECAL 2011.*
- Santana, P., Mendonça, R., Alves, N., Correia, L., and Barata, J. (2011). Tracking natural trails with swarm-based visual saliency. *Journal of Field Robotics*.
- Santana, P., Mendonça, R., Correia, L., and Barata, J. (2012). Neural-Swarm Visual Saliency for Path Following. *Applied Soft Computing*.

The model herein proposed is being fielded on an all-terrain robot being developed with the SME Portuguese company IntRoSys, S.A. in the context of the QREN project Introsys Robot, project number 2008/002641.

Bibliography

(1989). *NATO Advanced Workshop on Robots and Biological Systems*. NATO.

ADInstruments (2009). Visual evoked potential (vep) (english) - education - adinstruments.

Ahmed, S. (1991). *VISIT: An Efficient Computational Model of Human Visual Attention*. PhD thesis.

Alon, Y., Ferencz, A., and Shashua, A. (2006). Off-road path following using region classification and geometric projection constraints. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 689–696. IEEE.

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

Antón-Canalís, L., Hernández-Tejera, M., and 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)*, pages 604–609. IEEE Computer Society, Washington, DC.

Bartel, A., Meyer, F., Sinke, C., Wiemann, T., Nchter, A., Lingemann, K., and Hertzberg, J. (2007). Real-time outdoor trail detection on a mobile robot. In *Proceedings of the 13th IASTED International Conference on Robotics, Applications and Telematics*, pages 477–482.

Batavia, P. H. and Singh, S. (2001). Obstacle detection using adaptive color segmentation and color stereo homography. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, volume 1, pages 705–710. IEEE Press, Piscataway.

Blas, M., Agrawal, M., Konolige, K., and Sundaresan, A. (2008). Fast color/texture segmentation for outdoor robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4078–4085. IEEE Press, Piscataway.

Bouguet, J. (1999). Pyramidal implementation of the lucas kanade feature tracker description of the algorithm. *Intel Corporation, Microprocessor Research Labs, OpenCV Documents*.

Broggi, A., Caraffi, C., Fedriga, R. I., and Grisleri, P. (2005). Obstacle detection with stereo vision for off-road vehicle navigation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) - Workshops*, pages 65–72. IEEE Computer Society, Washington, DC.

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

Burt, P. and Adelson, E. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540.

Chaturvedi, P. and Malcolm, A. (2005). Real-time road following in natural terrain. In *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems*, volume 2, pages 815–820. IEEE.

Crisman, J. and Thorpe, C. (1991). Unscarf-a color vision system for the detection of unstructured roads. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2496–2501. IEEE Press, Piscataway.

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.

Engel, S., Zhang, X., and Wandell, B. (1997). Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, 388(6637):68–71.

Fernandez, D. and Price, A. (2005). Visual detection and tracking of poorly structured dirt roads. In *Proceedings of the International Conference on Advanced Robotics (ICAR)*, pages 553–560. IEEE.

Fernandez, J. and Casals, A. (1997). Autonomous navigation in ill-structured outdoor environment. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 1, pages 395–400. IEEE Press, Piscataway.

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

Frintrop, S. (2006). *VOCUS: a visual attention system for object detection and goal-directed search*. PhD thesis, INAI, Vol. 3899, Germany.

Frintrop, S., Backer, G., and Rome, E. (2005). Goal-directed search with a top-down modulated computational attention system. In *Proceedings of the DAGM 2005, Lecture Notes on Computer Science, 3663*, pages 117–124. Springer-Verlag, Berlin, Germany.

Frintrop, S. and Kessel, M. (2009). Most salient region tracking. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1869–1874. IEEE Press, Piscataway.

Gevers, T. and Smeulders, A. (1999). Color-based object recognition. *Pattern Recognition*, (32):453–464.

Ghurchian, R., Hashino, S., and Nakano, E. (2004). A fast forest road segmentation for real-time robot self-navigation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 406–411 vol.1. IEEE Press, Piscataway.

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.

Grudic, G. and Mulligan, J. (2006). Outdoor path labeling using polynomial mahalanobis distance. In *Proceedings of Robotics: Science and Systems*, pages 16–19. MIT Press: Cambridge, MA.

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.

Jacob, C., Litorco, J., and Lee, L. (2004). Immunity through swarms: Agent-based simulations of the human immune system. In *Artificial Immune Systems*, volume 3239 of *Lecture Notes in Computer Science*, pages 400–412.

Jeanne, R. L. (1996). *Regulation of nest construction behaviour in Polybia occidentalis*, volume 52 of *Animal Behaviour*, pages 473–488. Elsevier.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science*. McGraw-Hill, fourth edition.

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

Kolter, J. Z., Kim, Y., and Ng, A. Y. (2009). Stereo vision and terrain modeling for quadruped robots. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 1557–1564. IEEE Press, Piscataway.

Kong, H., Audibert, J., and Ponce, J. (2010). General road detection from a single image. *IEEE Transactions on Image Processing*, 19(8):2211–2220.

Konolige, K., Agrawal, M., Blas, M. R., Bolles, R. C., Gerkey, B., Solà, J., and Sundaresan, A. (2009). Mapping, navigation, and learning for off-road traversal. *Journal of Field Robotics*, 26(1):88–113.

Lacroix, S., Mallet, A., Bonnafous, D., Bauzil, G., Fleury, S., Herrb, M., and Chatila, R. (2002). Autonomous rover navigation on unknown terrains: Functions and integration. *International Journal of Robotics Research*, 21(10-11):917–942.

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

Lloyd, S. P. (1982). *Least squares quantization in PCM*, volume 28.

Manduchi, R., Castano, A., Talukder, A., and Matthies, L. (2005). Obstacle detection and terrain classification for autonomous off-road navigation. *Autonomous Robots*, 18(1):81–102.

Mazouzi, S., Guessoum, Z., Michel, F., and Batouche, M. (2007). A multi-agent approach for range image segmentation. In *Proceedings of the 5th international Central and Eastern European conference on Multi-Agent Systems and Applications (CEEMAS), LNAI 4696*, volume 4696, pages 1–10. Springer-Verlag, Berlin, Germany.

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

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

Oliveira, D. and Bazzan, A. L. C. (2006). Traffic lights control with adaptive group formation based on swarm intelligence. In *Ant Colony Optimization and Swarm Intelligence*, volume 4150 of *Lecture Notes in Computer Science*, pages 520–521.

Ouerhani, N. and Hugli, H. (2003a). Maps: Multiscale attention-based presegmentation of color images. 2695:537–549.

Ouerhani, N. and Hugli, H. (2003b). Real-time visual attention on a massively parallel SIMD architecture. *International Journal of Real Time Imaging*, 9(3):189–196.

Ouerhani, N., Wartburg, R., Hügli, H., and Müri, R. (2004). Empirical validation of the saliency-based model of visual attention. *Electronic Letters on Computer Vision and Image Analysis*, pages 13–24.

Owechko, Y. and Medasani, S. (2005). A swarm-based volition/attention framework for object recognition. In *Proceedings of the IEEE Computer Vision and Pattern Recognition Workshop (CVPRW)*, pages 91–98. IEEE Computer Society, Washington, DC.

Poli, R. and Valli, G. (1993). Neural inhabitants of MR and echo images segment cardiac structures. In *Proceedings of the Computers in Cardiology*, pages 193–196. IEEE Computer Society, Washington, DC.

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

Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., Berger, E., Wheeler, R., and Ng, A. (2009). Ros: an open-source robot operating system. In *Proc. of the ICRA Open-Source Software Workshop*.

Ramos, V. and 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)*, pages 113–116, Belgium.

Rasmussen, C. (2004). Grouping dominant orientations for ill-structured road following. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1. IEEE Computer Society, Washington, DC.

Rasmussen, C. (2008). Roadcompass: following rural roads with vision+ ladar using vanishing point tracking. *Autonomous Robots*, 25(3):205–229.

Rasmussen, C., Lu, Y., and Kocamaz, M. (2009). Appearance contrast for fast, robust trail-following. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. IEEE Press, Piscataway.

Rasmussen, C. and Scott, D. (2008a). Shape-guided superpixel grouping for trail detection and tracking. In *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4092–4097. IEEE Press, Piscataway.

Rasmussen, C. and Scott, D. (2008b). Terrain-based sensor selection for autonomous trail following. In *Proceedings of the 2nd International Workshop on Robot Vision (Robvis 2008)*, pages 341–355.

Reinhardt, D. (2010). Human eye anatomy and basic eye facts of biology, chemistry and physics.

Rosenblatt, J. K. (1995). DAMN: a distributed architecture for mobile navigation. In *Proceedings of the AAAI Spring Symposium on Lessons Learned from Implemented Software Architectures for Physical Agents*, Stanford, CA.

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

Rusu, R., Sundaresan, A., Morisset, B., Hauser, K., Agrawal, M., Latombe, J., and Beetz, M. (2009). Leaving Flatland: Efficient real-time three-dimensional perception and motion planning. *Journal of Field Robotics*, 26(10):841–862.

Santana, P. (2011). *Visual Attention and Swarm Cognition for Off-Road Robots*. PhD thesis, Department of Computer Science, Faculty of Sciences, University of Lisbon.

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., Guedes, M., Correia, L., and Barata, J. (2011). Stereo-based all-terrain obstacle detection using visual saliency. *Journal of Field Robotics*, 28(2):241–263.

Savage, M. and Askenazi, M. (1998). Arborscapes: A swarm-based multi-agent ecological disturbance model. Working Papers 98-06-056, Santa Fe Institute.

Seraji, H. (2006). Safety measures for terrain classification and safest site selection. *Autonomous Robots*, 21(3):211–225.

Song, D., Lee, H., Yi, J., and Levandowski, A. (2007). Vision-based motion planning for an autonomous motorcycle on ill-structured roads. *Autonomous Robots*, 23(3):197–212.

Stroobandt, S. (1997). Analogy with the rods and cones of the eye's retina.

Thorpe, C., Hebert, M., Kanade, T., and Shafer, S. (1988). Vision and navigation for the Carnegie-Mellon Navlab. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(3):362–373.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., Lau, K., Oakley, C., Palatucci, M., Pratt, V., Stang, P., Strohband, S., Dupont, C., Jendrossek, L.-E., Koelen, C., Markey, C., Rummel, C., van Niekerk, J., Jensen, E., Alessandrini, P., Bradski, G., Davies, B., Ettinger, S., Kaehler, A., Nefian, A., and Mahoney, P. (2006). Stanley: The robot that won the darpa grand challenge. *Journal of Field Robotics*, 23(9):661–692.

Todt, E. and Torras, C. (2000). Detection of natural landmarks through multi-scale opponent features. *ICPR 2000*, 3:988–1001.

Tomasi, C. and Shi, J. (1994). Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 593–600. IEEE Computer Society, Washington, DC.

Tsotsos, J., Culhane, S., Wai, W., Lai, Y., Davis, N., and Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, 78:507–545.

Tue-Cuong, D.-S., Dong, G., Hwang, Y. C., and Heng, O. S. (2008). Extraction of shady roads using intrinsic colors on stereo camera. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 341–346. IEEE.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience*, 16(6):2112.

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.

Zhu, S.-C., Guo, C.-E., Wang, Y., and Xu, Z. (2005). *What are Textons?*, volume 62.

Appendix A

Results Detailed

Table A.1: Comparisation between different visual attention models for path detection. The aggregate path detection results were computed using a data-set composed by 39 videos.

Video ID	Nr of Frames	Classical model (Itti et al., 1998) detection rate [%]	Swarm-based model (Santana et al., 2010) detection rate [%]	Proposed model detection rate [%]
1	278	44.60	100.00 ± 0.00	100.00 ± 0.00
2	204	61.76	100.00 ± 0.00	100.00 ± 0.00
3	422	4.74	93.03 ± 0.21	100.00 ± 0.00
4	135	0.00	100.00 ± 0.00	100.00 ± 0.00
5	2854	32.48	93.90 ± 0.02	100.00 ± 0.00
6	186	27.96	97.53 ± 0.30	100.00 ± 0.00
7	121	0.00	100.00 ± 0.00	100.00 ± 0.00
8	124	0.00	88.06 ± 0.36	100.00 ± 0.00
9	301	18.77	98.38 ± 0.32	97.35 ± 0.13
10	147	49.66	92.11 ± 0.61	100.00 ± 0.00
11	386	0.00	100.00 ± 0.00	100.00 ± 0.00
12	158	0.00	88.48 ± 0.28	100.00 ± 0.00
13	134	40.30	87.31 ± 0.53	100.00 ± 0.00
14	676	44.23	99.14 ± 0.07	100.00 ± 0.00
15	683	26.50	91.22 ± 0.10	100.00 ± 0.00
16	770	4.55	82.96 ± 0.14	97.45 ± 0.06
17	403	34.99	93.90 ± 0.14	100.00 ± 0.00
18	335	97.01	86.21 ± 0.13	100.00 ± 0.00
19	230	84.78	76.43 ± 0.20	100.00 ± 0.00
20	439	6.38	82.92 ± 0.23	100.00 ± 0.00
21	490	3.67	93.31 ± 0.09	100.00 ± 0.00
22	230	10.87	100.00 ± 0.00	100.00 ± 0.00
23	600	6.00	90.10 ± 0.15	100.00 ± 0.00
24	802	0.00	95.06 ± 0.07	100.00 ± 0.00
25	907	0.00	94.42 ± 0.06	100.00 ± 0.08
26	1553	0.97	60.08 ± 0.00	100.00 ± 0.08
27	3011	0.70	24.50 ± 0.05	99.47 ± 0.04
28	1288	0.00	89.72 ± 0.14	100.00 ± 0.00
29	267	11.99	71.39 ± 0.18	96.25 ± 0.24
30	440	19.32	79.00 ± 0.31	84.82 ± 0.36
31	1027	18.70	68.96 ± 0.07	99.51 ± 0.06
32	1083	2.95	82.01 ± 0.04	89.33 ± 0.07
33	1649	2.91	90.42 ± 0.00	92.15 ± 0.10
34	591	5.25	50.19 ± 0.14	97.33 ± 0.20
35	388	17.27	73.97 ± 0.16	97.42 ± 0.23
36	2515	0.00	30.43 ± 0.04	100.00 ± 0.00
37	429	47.09	75.34 ± 0.09	100.00 ± 0.00
38	829	3.26	91.05 ± 0.05	96.61 ± 0.10
39	2696	34.85	90.22 ± 0.01	100.00 ± 0.00
	29789	19.60	84.66 ± 0.14	98.67 ± 0.04