Nelson Miguel Rosa Alves

Vision Based Trail Detection for All-Terrain Robots

Lisboa
2010

UNIVERSIDADE NOVA DE LISBOA

Faculdade de Ciências e Tecnologia

Departamento de Engenharia Electrotécnica

Vision Based Trail Detection for All-Terrain Robots

Nelson Miguel Rosa Alves

Dissertação apresentada na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa para obtenção do grau de Mestre em Engenharia Electrotécnica e de Computadores.

Orientador: Prof. José António Barata de Oliveira

Lisboa
2010

Acknowledgements

I would like to begin by thanking my supervisor Prof. José Barata, for motivation and for giving me the opportunity to work on the subject of this dissertation. I would also like to thank Pedro Santana for all his constant support and guidance throughout the whole work.

Next, I want to thank all my colleagues and friends, in particular David and Magno, for all the help, support and friendship along the years we spent together in this university. Samuel, Pedro, André, Zé, and all those who I might not have mentioned, you also have thanks.

My parents also deserve a word of appreciation, for all the continuous efforts in my education, all the support and belief. No words are enough to express my thanks to both of you, and to my brother as well.

Last, but not least, a very, very special thank you to all the members of the Comissão A.G.A., you are the best, and I'm glad to have so many friends like you.

Não tenhamos pressa,

mas não percamos tempo.

José Saramago

Resumo

Esta dissertação propõe um modelo para detecção de trilhos baseado na observação de que estes são estruturas salientes no campo visual do robô. Devido à complexidade dos ambientes naturais, uma aplicação directa dos modelos tradiconais de saliência visual não é suficientemente robusta para prever a localização dos trilhos. Tal como noutras tarefas de detecção, a robustez pode ser aumentada através da modulação da computação da saliência com conhecimento implicito acerca das características visuais (e.g. cor) que permitem uma melhor representação do objecto a encontrar. Esta dissertação propõe o uso da estrutura global do objecto, sendo esta uma característica mais estável e previsivel para o caso de trilhos naturais. Esta nova componente de conhecimento implicito é especificada em termos de regras de percepção activa, que controlam o comportamento de agentes simples que se comportam em conjunto para computar o mapa de saliência da imagem de entrada. Para o propósito de acumulação de informação histórica acerca da localização do trilho é utilizado um campo neuronal dinâmico com compensação de movimento. Resultados experimentais num conjunto de dados vasto revelam a habilidade do modelo de produzir uma taxa de sucesso de 91% a $20\,$Hz. O modelo demonstra ser robusto em situações onde outros detectores falhariam, tal como quando o trilho não emerge da parte de baixo da imagem, ou quando se encontra consideravelmente interrompido.

Abstract

This dissertation proposes a model for trail detection that builds upon the observation that trails are salient structures in the robot's visual field. Due to the complexity of natural environments, the straightforward application of bottom-up visual saliency models is not sufficiently robust to predict the location of trails. As for other detection tasks, robustness can be increased by modulating the saliency computation with top-down knowledge about which pixel-wise visual features (e.g., colour) are the most representative of the object being sought. This dissertation proposes the use of the object's overall layout instead, as it is a more stable and predictable feature in the case of natural trails. This novel component of top-down knowledge is specified in terms of perception-action rules, which control the behaviour of simple agents performing as a swarm to compute the saliency map of the input image. For the purpose of multi-frame evidence accumulation about the trail location, a motion compensated dynamic neural field is used. Experimental results on a large data-set reveal the ability of the model to produce a success rate of 91% at $20\,$Hz. The model shows to be robust in situations where previous trail detectors would fail, such as when the trail does not emerge from the lower part of the image or when it is considerably interrupted.

List of Symbols and Notations

Symbol	Description
ROC	Receiver Operating Characteristic
TPR	True Positive Rate
FPR	False Positive Rate
IOR	Inhibition-Of-Return
$N(.)$	original model normalisation operator [Itti et al., 1998]
$W(.)$	normalisation operator proposed by [Frintrop, 2006]
$K(.)$	proposed normalisation operator
$M(X)$	global maximum of a given map X
$m(X)$	local maxima of a given map X
$\mathbf{I}(t)$	input image
$\mathbf{C}^{\mathbf{C}}(t)$	colour conspicuity map
$\mathbf{C}^{\mathbf{I}}(t)$	intensity conspicuity map
$\mathbf{S}(t)$	saliency map
$\mathbf{P}^{\mathbf{C}}(t)$	pheromone map associated to the colour conspicuity map
$\mathbf{P}^{\mathbf{I}}(t)$	pheromone map associated to the intensity conspicuity map
$\mathbf{F}(t)$	dynamic neural field
$\mathbf{H}(t)$	perspective transformation homography matrix

Symbol	**Description**
E_m	set of agents
B	set of agent behaviours
O	set of possible motor actions
$o(n)$	agent position at iteration n
$a^+(n)$	most voted motor action at iteration n
s	score of an agent
$f_b(m, a, n)$	motor action evaluation function for stochastic behaviour
$g_b(p, a, n)$	motor action evaluation function for pheromone contribution

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Autonomous robotics has been, over the last 30 years, an increasing source of inspiration for research and development. Ever since the first unmanned vehicles many efforts have been made towards solving the problem of autonomous navigation, from complete autonomy in rural, off-road, aggressive environments to driving assistance in urban scenarios.

In outdoor environments the exploitation of any sort of structure is essential for safe robot navigation. An example is the ability to detect and follow trails, thus reducing the chances of collision with obstacles, in addition to lowering the cognitive load associated to path and trajectory planning.

On the account of path (roads and trails, paved or dirt) following, its importance is easy to understand, for paths are usually deprived of obstacles, thus providing safe passageway for both humans and vehicles. Besides the mentioned need to navigate through clear areas (i.e., with the least number of obstacles), trails in natural environments often offer some structure, which can be exploited and used for navigation purposes. This dissertation contributes to this line of research by proposing a computationally fast trail detector, with a good success rate, and a good level of robustness, for outdoor environments.

Most of the challenges of trail detection relate to their lack of a well defined morphology or appearance. This hampers a straightforward learning of trail models. In addition, they exist in environments that are unstructured themselves. This in turn complicates the learning of background models. Moreover, the problem of supervising the learning process remains an open issue. This is aggravated by the fact that trails change over time, thus rendering hand-labelling unsuited for the task at hand.

The majority of the models proposed in an attempt to solve the trail detection problem in mobile robots rely on hard assumptions concerning the shape of the trail and its surroundings, their appearance, or the relative position of the robot. In this line of thought, a common solution is to assume that the robot is already inside the trails' boundaries and oriented along it and take a sample patch of the area in front of the robot to build a colour model for the trail [Fernandez and Price, 2005] or background [Rasmussen and Scott, 2008b]. Similarly, there are some approaches that make use of 3-D information obtained from a LADAR to ascertain the drivable area before building the model [Dahlkamp et al., 2006]. Although LADAR has been widely used with success for robot navigation, namely in the DARPA Grand Challenges [Cremean et al., 2006], [Urmson et al., 2006], [Thrun et al., 2006], for low cost service robots operating in natural environments this is not the best approach. Besides the previous assumption, other methods also consider that the trail is surrounded by vegetation and strong edges, thus classifying areas mostly green as non-trail [Bartel et al., 2007] or using evolutionary algorithms to explore the borderlines [Broggi and Cattani, 2006]. When considering natural trails the mentioned assumptions might be too harsh, for these often appear somewhat homogeneous with the surroundings, thus reducing the effectiveness of the referred approaches.

An alternative is to make use of traditional segmentation methods [Zhang and Nagel, 1994], [Felzenszwalb and Huttenlocher, 2004], [Unser, 1995], [Jain and Farrokhnia, 1991] to discriminate the path in the visual field of the robot [Nabbe et al., 2006], [Kim et al., 2007]. The segmented image can then be analysed by searching for clusters with geometric properties identical to a trail [Blas et al., 2008], or by grouping segments according to its approximate shape and performing tests on their appearance [Rasmussen and Scott, 2008a]. However, good segmentation methods tend to be computationally intensive. Furthermore, natural trails usually present themselves with great unpredictability in shape and appearance. Still, geometric considerations can be used to generate trail hypotheses [Rasmussen et al., 2009], thus improving the efficiency but not solving all the problems that avert from natural environments.

A careful observation of natural images highlights the fact that trails are typically conspicuous in the visual field of the robot, i.e., are structures that easily pop-out. This observation has alerted to the possibility of using visual saliency as a means to focus the attention of an accurate trail detector in an unbiased way, thus not imposing any hard constraints on the appearance or shape of both trail and background. Hence, this dissertation contributes to this line of research by proposing a model-free solution for robust, reliable and computationally efficient trail detection in natural environments. Additionally, this dissertation extends considerably this concept by recurring to the swarm-based collective behaviour metaphor and by exploiting evidence accumulation across frames for improved robustness.

1.1 Problem Statement

This dissertation covers the problem of vision-based trail detection for mobile robots. The need to operate in unstructured environments in a sufficiently fast and robust way imposes two main requirements:

R1 - The proposed solution must be model-free, thus avoiding the necessity to rely on hard assumptions on the appearance and morphology of the trail, which, in turn, allows it to be used as a focus to guide a specialized detector. Such a solution also discards the need for straightforward learning of trail models, which tend to become outdated. Furthermore, not using given or learned models also results in increased robustness, necessary to deal with the highly unpredictable characteristics of trails in natural environments.

R2 - The trail detector should be computationally efficient. It is also desirable that the model lends itself to parallel implementation, thus allowing its application in parallel emergent and distributed systems.

1.2 Solution Prospect

This dissertation proposes the following solutions to comply with the specified requirements:

- The model makes use of visual saliency under the observation that trails are typically conspicuous in the visual field of the robot. However, saliency maps tend to be noisy due to the ubiquity of distractors and the heterogeneity of trails and therefore additional top-down knowledge on them is required. Hence, the trails' overall layout is used in order to deal with their lack of a well defined morphology and appearance, as it is a more stable and predictable feature in natural environments. This approach does not impose any hard constraints on the appearance or shape of both trail and surroundings. To isolate the proposed model's characteristics and more easily assess the major contributions, the work presented is divided into two parts.

  - In the first part, simple agents operating on the saliency maps generate trail skeleton hypotheses, whose behaviour embodies implicit general knowledge about trails' overall layout. Being simple, the agents are fast to compute and therefore compliant with the requirement R2. This part of the work validates the positive correlation between visual saliency and the trail location, as well as the application of the agent-based method to trail detection.

  - In the second part, the agent-based method is extended by allowing the agents to perform as a swarm. Being self-organised, the agents' collective exhibits accuracy and robustness without hampering computation efficiency. Temporal evidence accumulating the trail location is exploited by recurring to a fast to compute dynamic neural field, further increasing the mentioned properties.

1.3 Dissertation Outline

This dissertation is organised as follows:

Chapter 2 presents a brief overview of the state of the art in the context of trail detection;

Chapter 3 describes the first part of the work, a saliency-based model using simple agents to explore the attention based maps for the detection of the trail;

Chapter 4 describes the second part, a swarm-based model for trail detection, extending the previous one by allowing the agents to exhibit collective behaviour and accumulating evidence across frames;

Chapter 5 draws some conclusions concerning the work presented, as well as possible future improvements;

1.4 Further Readings

The work on trail detection using visual saliency and collective behaviour presented in this dissertation has already been published:

[Santana et al., 2010a] Santana, P., Alves, N., Correia, L., and Barata, J. (2010). A saliency-based approach to boost trail detection. *In Proc. of the 2010 IEEE Intl. Conf. on Robotics and Automation* (ICRA 2010), pages 1426-1431, May 3-8, 2010, Anchorage, Alaska.

[Santana et al., 2010b] Santana, P., Alves, N., Correia, L., and Barata, J. (2010). Swarm-based visual saliency for trail detection. *To appear in Proc. of the 2010 IEEE Intl. Conf. on Intelligent Robots and Systems* (IROS 2010), Taipei, Taiwan.

Chapter 2

State of the Art

This chapter surveys the state of the art in trail detection algorithms. Trails are usually safe pathways and also free of dead-lock situations. A robot following a trail is thus able to traverse large distances in off-road environments in an effortless, more secure way. On the one hand, computation for obstacle detection and trajectory or path planning is saved, thus allowing the allocation of resources to other tasks. On the other hand, fewer are the chances of getting lost or incurring into collisions, therefore contributing to the preservation of the mechanical structure and any payload the robot may be carrying. The importance of trail and road detection, suitable for real-time application in all-terrain service robots, has promoted the research on this subject over the past ten years. Several approaches have been proposed in this continuous search for newer, faster and more robust detection techniques.

The most successful and interesting methods to solve the problem of trail detection are presented next in this chapter.

Typical solutions often tend to rely on assumptions concerning the position of the robot and general characteristics of the trails, like their appearance and structure. Some approaches assume that the robot is already on the trail and oriented along it (Section 2.1), and make use of this information to build a colour model of the trail or background, thus separating one from the other. Besides this assumption, other approaches also assume that strong edges segment the trail from its surroundings (Section 2.2). The process of finding the boundaries can be very diverse, while some techniques rely on the use of colour information, others go a little further and recur to evolutionary, biologically inspired algorithms. However, these two assumptions often fail to occur on realistic situations. An alternative is to segment the image, group some of the segments to build hypotheses, and then score these hypotheses against a model of the trail (Section 2.3). Good segmentation techniques tend to be computationally intensive, thus rendering the use of this method unsuitable for real-time applications. In an attempt to reduce this computation time problem, and assuming a trail viewed under perspective is well approximated by a triangular shape, hypotheses may be generated directly, and then scored using appearance contrast between a trail hypothesis and its surrounding regions (Section 2.4).

The model presented in this dissertation differs from the previous approaches by making use of visual attention techniques (Section 2.5). Due to the computationally intensive nature of visual search algorithms, the ability to highlight features and places of interest in a context-dependent way might prove useful. Although visual saliency mechanisms are not common in trail detection, in other applications including attentional systems for humanoids and obstacle detection for mobile robots several approaches have risen.

2.1 On-trail Approaches

When considering autonomous navigation for mobile robots, most of the time it is reasonable to assume that the robot is already following a road or trail, and therefore inside its boundaries and oriented along it. This assumption has given birth to a number of methods, being most of them colour-based algorithms.

On the account of road detection, a method based on self-supervised learning has been proposed by Dahlkamp et al. [Dahlkamp et al., 2006]. This model relies on a laser range finder to scan for flat, drivable surface area in the vicinity of the vehicle, which is assumed to be road. The colour information associated with this area is then used to construct appearance models to classify the entire field of view of the camera. Additionally, GPS information is used to guarantee that the robot is on the road and therefore the drivable surface identified by the laser

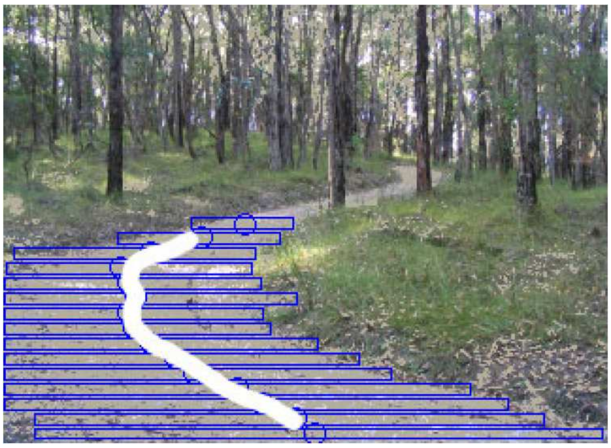


Figure 2.1: Trail detection using colour-based clustering as proposed in [Fernandez and Price, 2005]. Blue rectangles represent the culled segments. The determined trajectory is overlaid in white.

range finder is correct. In natural unstructured off-road environments, where paths commonly appear as trails, thus being more narrow and unpredictable than roads and surrounded by dense vegetation or trees, GPS information might not be so reliable.

The model proposed by Fernandez and Price [Fernandez and Price, 2005] relies on colour vision for detection and tracking of poorly structured dirt roads in natural environments. The prime assumption for this method is that road surface displays colour-space statistics different from the surrounding regions. In this model, the task of road detecting and tracking is accomplished in three steps: characterisation of the road, clustering of road regions, and modelling of its trajectory. For the first step, it is yet assumed that a small rectangle in the centre-bottom of the image always contains a portion of the road, which will be used to characterise it. Analysis of this region in a Hue, Saturation, Intensity (HSI) variant colour-space is the basis for the creation of a colour-based filter, which is then used to find pixels that may belong to the road area. The next step consists of aggregating these pixels into regions representing segments of the road, by assuming it is generally presented in the image flowing from bottom to top. This is accomplished by first dividing the image in horizontal slices, then performing a series of region-growing segmentation operations using the slices as borderlines, and finally merging and culling the segments, retaining only the largest segment per slice. In the third and final step the centres of mass of the segments mentioned in the previous step are computed and used

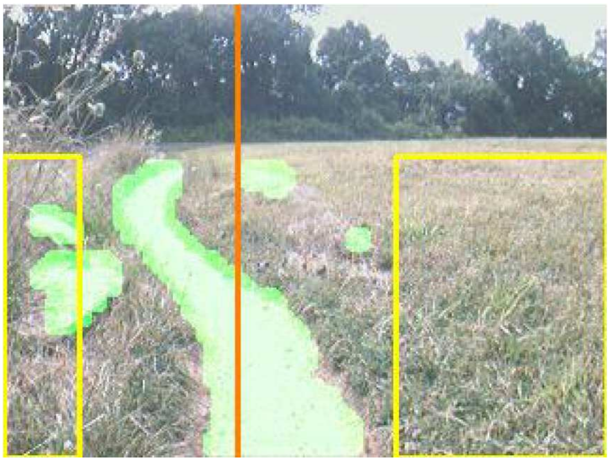


Figure 2.2: Trail detection using histogram colour classification as proposed in [Rasmussen and Scott, 2008b]. Yellow lines delimit the reference areas. Pixels classified as on-trail represented in green.

to determine the trajectory, by means of a spline curve using the weighted centres of mass as control points, as depicted in Fig. 2.1.

Another approach based on similar assumptions, applied to unstructured trails in natural environments, is the model proposed by Rasmussen and Scott [Rasmussen and Scott, 2008b]. In this method the terrain is first classified as flat, thick, or forested by analysis of a *ladar* scan. In the case of thick or forested terrain, *ladar* information alone is used to guide the robot by finding the empty space between the vegetation. In flat terrain though, this is not sufficient for selecting the region of the image corresponding to the trail and therefore an image-based trail segmentation is performed. In this case, the first step of the method consists of constructing a 3-D histogram corresponding to a colour model for the background, based on the RGB values of the pixels contained in two narrow rectangular areas in the left and right sides of the image, extending from its bottom to a horizon line. This background model is then used to classify all the pixels below the horizon line as on or off-trail. Finally, the reference left and right areas are adapted according to an estimate of the trail width. An example of the result of this model for flat terrain is shown in Fig. 2.2.

As mentioned, the above models work nicely in situations following the main assumption that the robot is already on the trail and oriented along it. Another requirement for the success of these algorithms is that the trails present colour statistics that are different from the background. Bearing this in mind, these algorithms are prone to fail in cases where: (1) variations that are sufficient to cause the robot to lose track occur in the structure of the trail; (2) the robot is not on nor aligned with the trail; and (3) the colour statistics of the trail and the background are identical. Unpredictability in the structure, appearance, and orientation of natural trails makes these situations of possible failure occur more frequently than expected, thus highlighting the need for algorithms not sensitive to these factors.

2.2 Edge Detection based Approaches

Trails and roads often present visual cues that distinguish them from the background. The idea of exploring the characteristics of the road-scene has given birth to a number of approaches based on the detection of the boundaries of the trail or road, thus finding and classifying the internal subsequent area as navigable. These methods are most commonly applied to well structured roads and lanes, as can be found in urban environments.

Well known work in vision-based road following includes the models proposed by Southall and Taylor [Southall and Taylor, 2001], which consists of extracting the lane markings exploring the contrast on colour images and then estimate the vehicle's position using a particle filter, and Apostoloff and Zelinsky [Apostoloff and Zelinsky, 2003], which relies on particle filtering and cue fusion technologies to build a multiple-cue visual lane tracking system. Both methods are designed to work on paved or painted roads with sharp edges. Conversely, Rasmussen [Rasmussen, 2004] proposed a model for following ill-structured roads using the dominant texture orientations of every pixel in the image to estimate a vanishing point.

In rural and off-road areas, where roads appear with trail-like characteristics, i.e., narrow and unstructured, this kind of approach is not so common but used in the work described below.

A visual method for outdoor trail localization relying on edge detection is proposed by Bartel et al. [Bartel et al., 2007]. In this model the visual classification algorithm is processed in three steps: trail border detection, object extraction and direction control. For detecting the

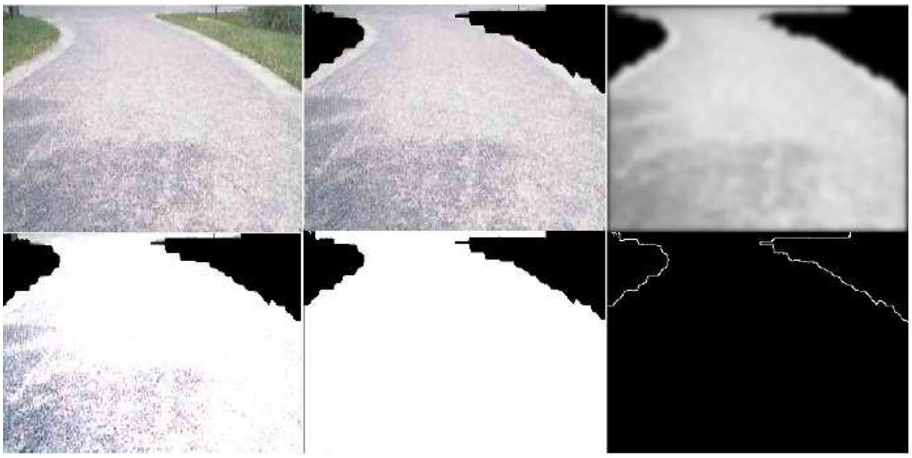


Figure 2.3: Border extraction process, as proposed in [Bartel et al., 2007]. Processing steps, from left to right, top to bottom, are: original image, green to black, gaussian blur, contrast enhancement, thresholding and border extraction

trail border, it is assumed that most trails have grass or planted borders and therefore all green pixels of the image are painted black. Next, a gaussian blur is applied, followed by a contrast enhancement, thus revealing the pathway as the brightest area in the image. Finally, the image is thresholded, separating the path from non-trail areas, and its edges are extracted by means of a gradient filter. See Fig. 2.3 for an illustrative example of this process. In the second step, an object extraction algorithm is used to select the biggest contour surrounding smaller ones, which allows the rejection of wrongly classified structures lying within the boundaries of the trail. For the last step, the centre of the extracted boundaries on several horizontal lines is used to generate a control signal.

An alternative to the above method is the use of simple agents as proposed by Broggi and Cattani [Broggi and Cattani, 2006]. The implemented algorithm is based on the Ant Colony Optimization (ACO) [Dorigo and Stützle, 2004], which consists in a parallel meta-heuristic for combinatorial optimization problem inspired by the foraging behaviour of biological ants. The first step is to localize the optimal starting states, which are placed in peripheral areas where a sufficient percentage of edges is present. By computing the vectorial euclidean distance between a RGB Normalized transformation over the input image and the temporal average of Red, Green and Blue values of road pixels over all the images, and then applying a gradient operator,

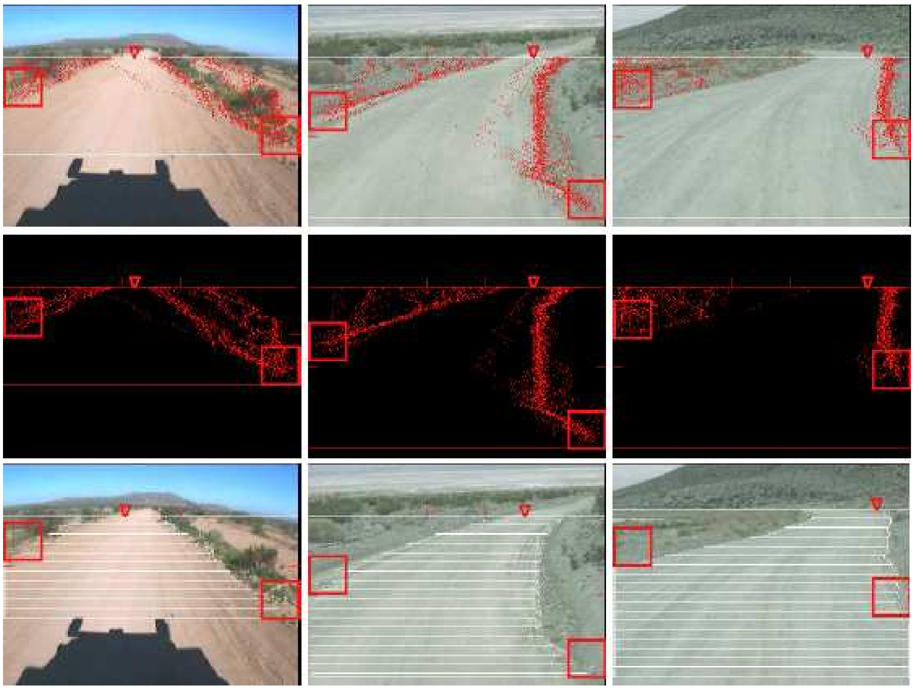


Figure 2.4: Evolutionary approach to path detection proposed in [Broggi and Cattani, 2006]. Top row presents the original images with the agents' paths overlaid. Middle row shows only the ants' paths, for clearer visualization. Bottom row reveals the path detection results.

a monochromatic edge image is obtained. This represents the only a priori information given to the agents. The obtained image is then used to define the local heuristic function, under the rule that the attractiveness of a pixel is proportional to the brightness of its correspondent in the edge image, and the cost function, under the rule that the cost of movement towards a pixel is inversely proportional to the brightness of its correspondent in the edge image. Next, the deployed agents move according to their random-proportional and pseudo-random-proportional rules, with edge-exploitation and pheromone-exploitation behaviours, and random movement polarized by a point of attraction. Agents are divided in subsets with different parameters for the moving rules, meaning that as the execution proceeds they become more sensitive to pheromone, and less to heuristics. When every agent of a subset has reached the final pixels the pheromone trails are updated according to an evaporation ratio and the ants contributions, in order to enhance paths formed by bright pixels and recurrently visited. To extract the final solution a single agent is created in each colony, which move pixel by pixel attracted only to pheromone until a final pixel is reached, thus building a representation for the road boundaries.

The agents paths and detection results can be seen in Fig. 2.4.

These models work nicely in well delimited trails surrounded by dense vegetation or sidewalk, and in roads with lanes well demarcated from the background. In situations where the trail is somewhat mixed with its surroundings, like most often occurs in natural off-road environments, edge detection may become a difficult task, therefore making these kind of algorithms achieve a low success rate.

2.3 Segmentation-based Approaches

Finding the edges of a road in order to segment it from the background might be a difficult challenge in natural terrain. In this line of thought, other approaches based on image segmentation techniques have emerged. These methods should be accurate enough to allow distinction between background and trail. Therefore, the detection algorithm focuses on selecting the segment or group of segments that best fit a trail model.

The work proposed by Soquet et al. [Soquet et al., 2007] makes use of stereovision to estimate free space in the image, and then applies colour segmentation to extract road segments. Anisotropic texture features of roads are explored by Zhang and Nagel [Zhang and Nagel, 1994] for the segmentation purpose. These algorithms focus in paved and fairly structured roads, which possess characteristics not present in natural trails like hiking and biking paths.

In this context, Rasmussen and Scott [Rasmussen and Scott, 2008a] proposed a model for the trail detection problem. In this method it is assumed that there is only one trail region, and that it follows the shape of a triangle with its base aligned with the bottom edge of the image. The detection algorithm begins by generating a set of trail hypotheses based on the grouping of superpixels generated by an over-segmentation algorithm. These hypotheses are then scored according to several shape and appearance criteria, and finally the one with the highest score is picked as the representation of the trail region. An iterative, agglomerative process is used to generate the hypotheses. First, a pixel is selected randomly from the bottom of the image and used as the seed. Each iteration of the agglomeration process consists of adding a new

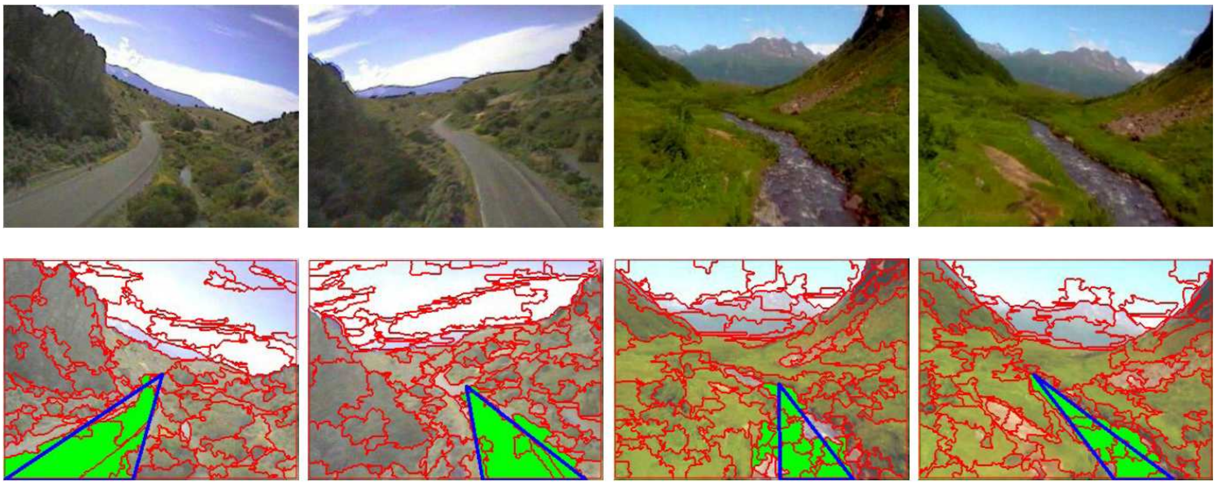


Figure 2.5: Superpixel segmentation based trail detection as proposed in [Rasmussen and Scott, 2008a]. Top row represents the input images. Bottom row shows the output of the detector, overlaid in the input image. Red delimits segments obtained from the segmentation process. The best-scoring grouping is represented in green, and the respective fitted triangle in blue.

member to the superpixel grouping, chosen from its set of neighbours and with probability given by the Euclidean distance in RGB space between the neighbouring superpixel and the current group. The appearance variation and the overall size of the agglomeration are used to determine the final number of superpixels in it. To assess the trail likelihood of each hypothesis, the grouping is scored using a triangle as the trail shape template and according to three terms: shape, appearance, and deformation. This is done similarly to the work of Sclaroff and Liu [Sclaroff and Liu, 2001], which uses a deformable model to guide the grouping of regions in search for relatively simple shapes like bananas and street signs in fairly uncluttered images. The first term, shape likelihood, consists in approximating the grouping with a triangle fitted with its highest, leftmost and rightmost points, and then measuring the similarity between both shapes. The second, appearance likelihood, measures the difference in appearance between a grouping and its neighbouring superpixels, and the variation within. Lastly, deformation likelihood measures how different the approximated triangle is from a learned model of the trail. The triangle fitted to the best-scoring grouping is propagated to the next frame, and used to evaluate the appearance likelihoods of the newly generated hypotheses. Fig. 2.5 presents some illustrative results of the detector.

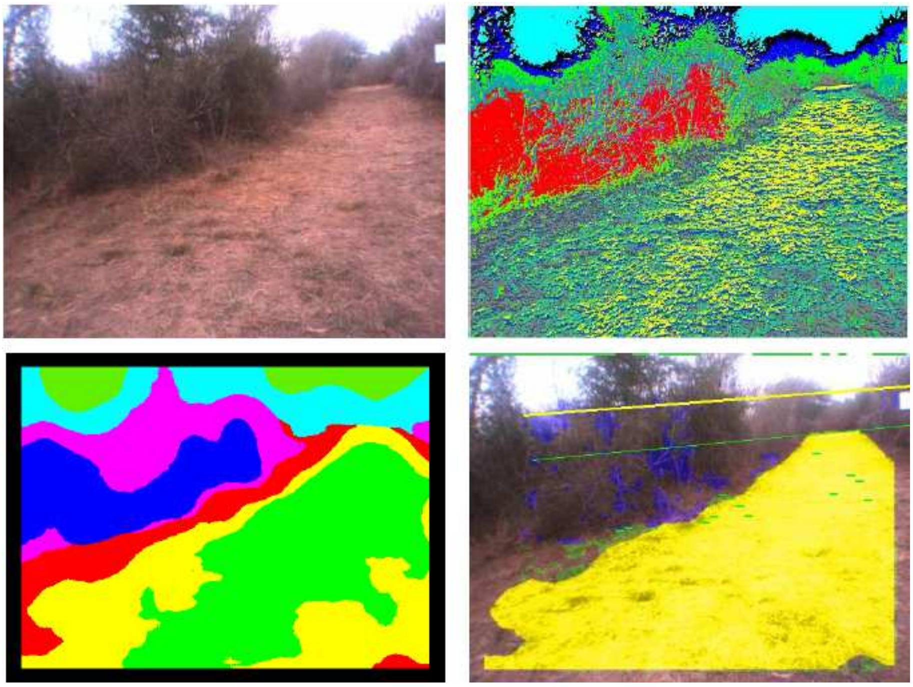


Figure 2.6: Segmentation algorithm for path detection as proposed in [Blas et al., 2008]. Top-left shows the input image. Top-right represents the assignment of each pixel to a texton. Bottom-left is the final segmentation, and recognized path is shown in bottom-right.

Another segmentation based method applied to trail detection is the one proposed by Blas et al. [Blas et al., 2008]. This appearance-based segmentation algorithm makes use of colour and texture in conjunction with 3-D information provided by a stereo camera. For colour and texture representation compact descriptors composed by the colour information of the centre pixel and the relative change in intensity in a local neighbourhood are used. The computed descriptors are grouped using a k-means algorithm [Jain and Dubes, 1988], [Duda and Hart, 1973]. Alternative clustering methods include graph-cut-based approaches [Martin et al., 2004], Self-Organizing Maps [Martin-Herrero et al., 2004], or level-sets [Liapis et al., 2004]. The grouped descriptors are then assigned to basis vectors, i.e., textons [Leung and Malik, 2001], which are vocabularies for tiny surface patches with associated local geometric and photometric properties. Histograms of these textons are clustered again using k-means to find similar regions in the image, which are merged to provide the final segmentation. The segmented image contains only a small number of regions, which are then analysed in search for the ones presenting geometric attributes more similar to a path. These steps can be visualized in Fig. 2.6.

Although a good segmentation of the terrain may provide a valuable assistance in trail detection, contemporary models for robust image segmentation and subsequent grouping are computationally intensive and consequently unsuitable for real-time requirements. Moreover, grouping tends to fail in the presence of interrupted trails.

2.4 Contrast-based Approaches

In a study parallel to the one presented in this dissertation, and extending the superpixel model described in the previous section, Rasmussen et al. [Rasmussen et al., 2009] proposes the use of appearance contrast for trail detection. The main idea behind this method is to look for a triangular region which contrasts with the surroundings. The basic framework for trail finding is to generate trail hypotheses and score each of them with a likelihood function, assuming that a trail viewed under perspective may be associated with a triangle shape starting from the bottom of the image. Trail hypotheses are generated from a learned distribution of expected trail width and curvature variation. For each hypothesis, two additional triangles are defined in its left and right neighbouring regions. Histograms of k-means cluster labels in a CIE-Lab colour space are computed for the three triangles. The trail likelihood is captured by measuring the dissimilarity between the trail region and the surrounding ones, as well as the symmetry of the flanking regions, and the highest scored hypothesis is chosen for trail representation, as can be seen in Fig. 2.7. Several alternatives to measure similarity between image regions include colour and texture histogram measures such as Bhattacharyya or χ^2 [Dunlop et al., 2007], [Varma and Zisserman, 2005], brightness in grayscale images [Ren and Malik, 2003], Euclidean colour distance [Martin et al., 2004], and the Earth Mover's Distance [Mori, 2005].

In the referred method it is assumed that trails are imaged as perfect triangles and both their left and right sides share the same appearance. Although these assumptions comply with a large set of situations, natural trails not always possess these properties. Additionally, the extensive use of 3-D information to bias the detection process in the model complicates the assessment of the role played by the appearance-based component in the results.

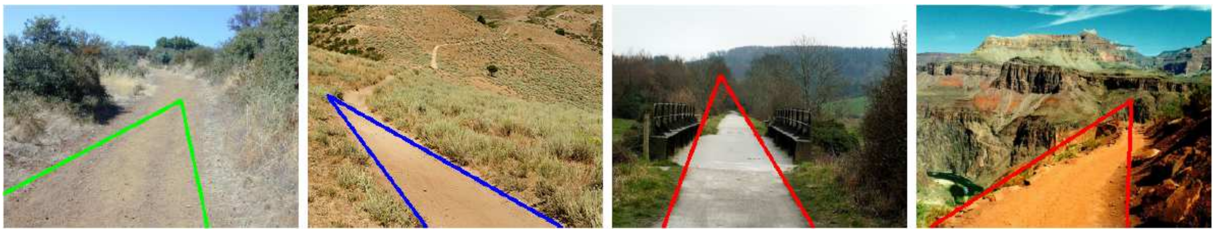


Figure 2.7: Appearance contrast method for trail detection as proposed in [Rasmussen et al., 2009]. The outputs of the detector are depicted as coloured triangles overlayed in the input image.

2.5 Visual Attention Models for Autonomous Robots

Cognitively rich robots make use of visual perception for interaction with humans and their surroundings in a context-dependent way. For this purpose, the ability to highlight features that have a high probability of being relevant, thus allowing the system to filter unimportant information, is a great advantage.

Vision-based attentional systems for humanoid robots have been proposed by Moren et al. [Moren et al., 2008] and Ruesch et al. [Ruesch et al., 2008]. Both models make use of visual saliency to control the gaze of a humanoid head. In the first, specific features of the objects being sought according to a task are used to apply top-down modulation to bottom-up saliency maps. This results in an increase of the saliency of these features, thus highlighting objects in context-dependent way. The second model integrates multi-modal saliency information (visual and auditory) into a unified spacial representation. The points with the highest overall saliency value are the ones considered interesting and used to focus the attention of the robot.

Visual attention mechanisms have also been applied to mobile robotics with the purpose of reducing the computational cost in expensive tasks like object detection and characterisation. Concerning vision-based navigation for all-terrain ground robots, visual saliency has been used for successful guidance of an obstacle detector by Santana et al. [Santana et al., 2009], [Santana et al., 2010c]. In these models, visual saliency is used to focus the attention of the detector by selecting areas of the image corresponding to regions that contain obstacles, thus narrowing the analysed data, which results in reduced computation times and lower sensitivity to noise. Although these models make use of visual saliency in the robot navigation context, this dissertation reports its first time application to the task of trail detection.

A different attentional mechanism is proposed by Hong et al. [Hong et al., 2002] to focus a colour-based detector of puddles and road signs. This model makes use of laser and colour information to build a world model. The data gathered is then used to predict which regions of future images should be analysed.

Since it is rather common the use of saliency for other tasks in cognitively rich robots, the overhead of its computation is diluted over all modules using it. Bearing this in mind, the use of a bottom-up saliency mechanism to guide the focus of selective attention in context-dependent tasks, like object or trail detection, is by itself a means of decreasing computation time and therefore an important process for real-time applications. However, in unstructured off-road environments, although bottom-up attention provides a means for constraining the focus of attention, a top-down mechanism is needed in order to find and keep the correct focus of interest on the object being sought (obstacles, trails) in spite of unrelated salient features.

Specifically in the trail detection problem, although visual saliency can be used as a means to segment the input image by determining which regions of the visual field detach more from the background, the saliency maps generated may not be accurate enough to allow an immediate and correct detection, namely in the presence of distractors or when the trail is considerably heterogeneous. A method for diminishing this problem is to use top-down boosting of visual features that are known to describe the object being sought. However, these features are considerably unpredictable in the case of trails in natural environments. To overcome these difficulties this dissertation proposes a novel use of top-down knowledge in the form of behaviours ruling the motion of simple agents inhabiting the saliency and its intermediate conspicuity maps.

Chapter 3

A Saliency-Based Approach to Boost Trail Detection

This chapter presents a saliency-based solution to boost trail detection. A careful observation of natural images highlights the fact that trails are structures that easily pop-out. Bearing this some quantitative support, and visual saliency could then be applied to focus the attention of an accurate trail detector in an unbiased way. Experimental results herein presented support this assumption and furthermore show that, with proper analysis, saliency information alone provides enough cues to reduce the ambiguity regarding both trail's position and approximate skeleton to three hypotheses, in the vast and diverse used dataset. This analysis is performed by a set of agents inhabiting the saliency and feature specific intermediate maps. These agents' behaviours exploit implicit, top-down knowledge about the object being sought in an active way. With the proposed model, computationally demanding accurate trail detectors are able to focus their activity to a fraction of the input image, thus promoting robustness and real-time performance. Notably, this robustness is revealed with the model's ability to detect what we humans would select as the most navigable area, in images where trails are almost indistinguishable or not even present. See Fig. 3.1 for some representative examples.

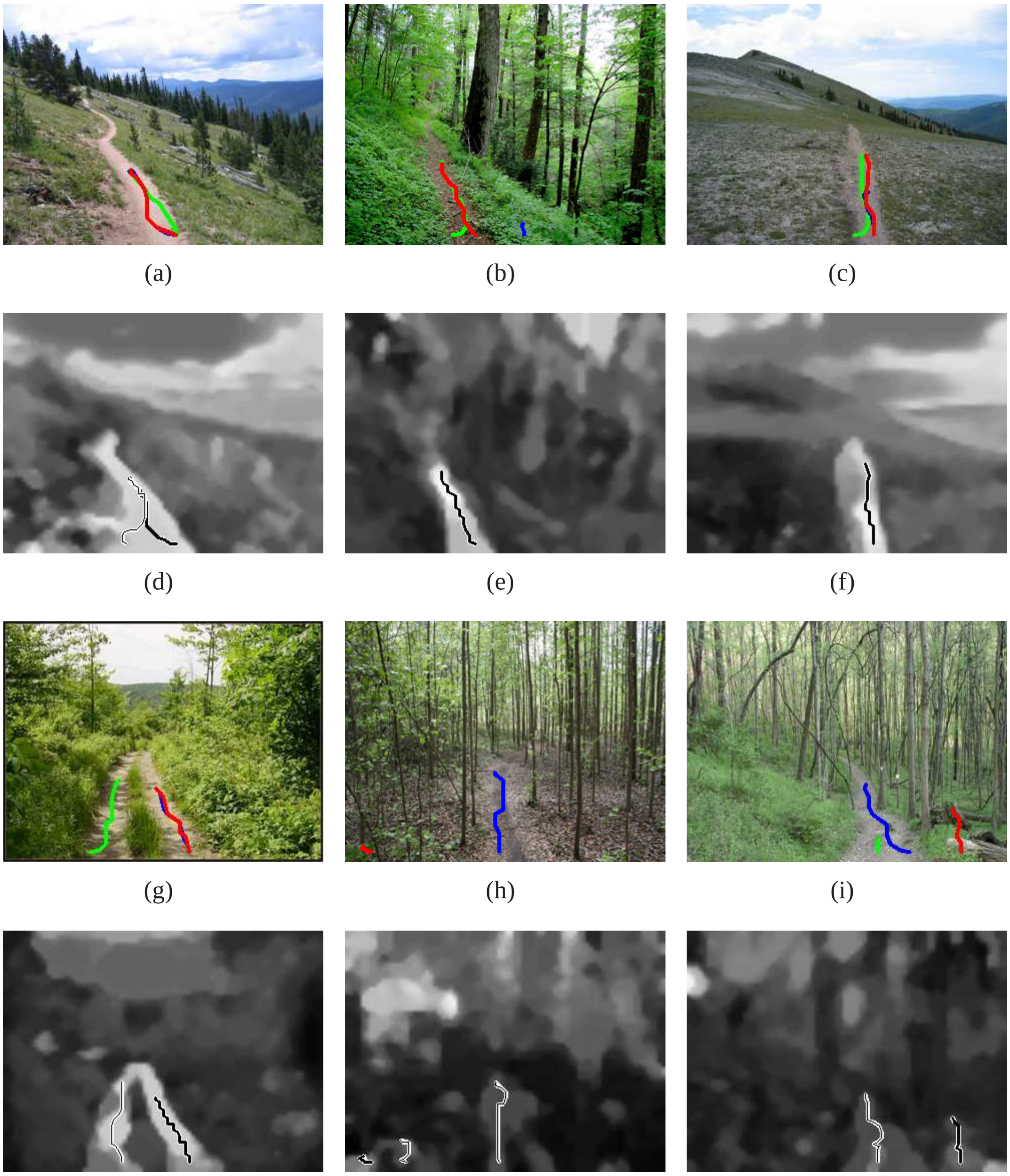


Figure 3.1: Input images (above) and respective saliency maps (below), where saliency is represented in grey level. These maps are the superposition of two conspicuity maps, one for colour and another for intensity channels. Each of these maps is searched for trails by agents (see Section 3.2), whose paths are described by the overlaid lines. Thicker lines refer to the most probable trail candidate, which appears in the input image in red. The best agent found on the intensity and colour conspicuity maps is represented in blue and green, respectively.

3.1 Saliency Computation

Saliency computation is about determining which regions of the input image are more conspicuous, i.e. detach from the background, at several scales and feature channels. In this model only intensity and colour channels are used, and saliency is computed according to the biologically inspired model proposed by Itti et al. [Itti et al., 1998], properly adapted to the task at hand.

Shortly, one dyadic Gaussian pyramid, with eight levels, 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 colour feature channels. The various scales are then used to perform centre-surround operations [Itti et al., 1998]. The resulting centre-surround maps have higher intensity on those pixels whose corresponding feature differs the most from their surroundings. An example is a dark patch on a bright background (off-on), as well as the other way around (on-off). On-off centre-surround operations are performed 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. Rather than considering the modulo of the difference, as in the original model [Itti et al., 1998], both on-off and off-on centre-surround maps are considered separately, which has been shown to yield better results [Frintrop et al., 2005, Frintrop, 2006]. Then, the centre-surround maps are blended to produce two conspicuity maps $\mathbf{C}^{\mathbf{C}}(t) \in [0,1]$ and $\mathbf{C}^{\mathbf{I}}(t) \in [0,1]$, one aggregating colour and another aggregating intensity information, respectively. Finally, these two maps are blended in a final saliency map $\mathbf{S}(t) \in [0,1]$ [Itti et al., 1998].

When blending maps, the most discriminant ones, i.e. those that highlight a smaller number of objects, are typically promoted by recurring to a normalisation operator. In the original model [Itti et al., 1998], this is done by scaling a given map X according to the normalisation operator $N(.)$. This operator is defined by the square of the difference between its global maximum,

$M(X)$, and the average of all its other local maxima, $\bar{m}(X)$, i.e.

$$N(X) = X \cdot (M(X) - \bar{m}(X))^2 \tag{3.1}$$

A similar normalisation operator has been proposed by Frintrop et al. [Frintrop et al., 2005, Frintrop, 2006]. In this case, the uniqueness operator,

$$W(X) = X / \sqrt{m(X)} \tag{3.2}$$

scales the map X according to the number of its local maxima above a given threshold, $m(X)$. In this work the threshold is set to its default value, i.e. 50% of the map's global maximum [Frintrop, 2006]. This method allows, among other things, to account for the proportion of objects competing for attention when determining their saliency.

Common to both methods is the use of local maxima information, which though appealing not always embodies the information intended to capture. Large homogeneous structures for instance, such as the sky, generally encompass only a few local maxima. In this situation, the sky would be undesirably considered highly conspicuous, despite its large foot-print in the whole image. A second aspect is that the two analysed saliency models consider that all pixels contribute equally to the saliency computation. However, excepting for extreme tilt/roll angles, the upper region of the image has little relevant information for trail detection. As a consequence, without a space-variant contribution to the final saliency map, feature maps that are only discriminative in the lower part of the image, and consequently interesting for trail detection, would not be adequately promoted.

In face of these limitations a new normalisation operator is herein proposed. Rather than considering only the map's local maxima when averaging, as it is done in $N(.)$, it is proposed to use all pixels. Furthermore, the contribution of each pixel to the average is weighted according to its distance from the top row. Formally, let $Int(X, c, r)$ return the intensity of the pixel in column c and row r of a given map X, with height $h(X)$. Let

$$w(X, c, r) = \sqrt{r/h(X)} \tag{3.3}$$

be the weight of pixel at position (c, r). The map's weighted average, m_w, is thus given by

$$m_w(X) = \frac{\sum_{(c,r)\in X} Int(X, c, r) \cdot w(X, c, r)}{\sum_{(c,r)\in X} w(X, c, r)} \tag{3.4}$$

and similarly to the operator $N(.)$, the proposed normalising operator, $K(.)$, takes the form

$$K(X) = X \cdot (M(X) - m_w(X))^2 \tag{3.5}$$

To reduce computational cost, the proposed system uses image operators over 8-bit images, whose magnitude is clamped to $[0, 255]$ by thresholding. In addition, prior to normalisation, maps are scaled to cover the interval $[0, 255]$, meaning that $M(X) = 255$ for all cases.

The Receiver Operating Characteristic (ROC) curves depicted in Fig. 3.2 show that, for the tested dataset (see Appendix A, Figs. A.1-A.6), the proposed procedure produces consistently a better trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) than the other two methods. The small difference between the ROC curves could suggest that only a small quantitative improvement was obtained with the proposed model. However, the averaging procedure used to build the curves hide the fact that none of the other methods was able to consistently allocate higher levels of saliency to trail regions than to the background as often as the proposed one.

Fig. 3.2 also shows that saliency is considerably correlated with trail location, which is an important contribution by itself. This correlation can also be observed for typical images in Fig. 3.1. However, it is still lower that the one required for accurate trail detection. That is, there is no single threshold on the saliency map that clearly segments the trail for all images in the dataset. It is thus important to devise a mechanism able to overcome this limitation. As it will be shown in the next section, an agent-based design is the adequate tool for the purpose.

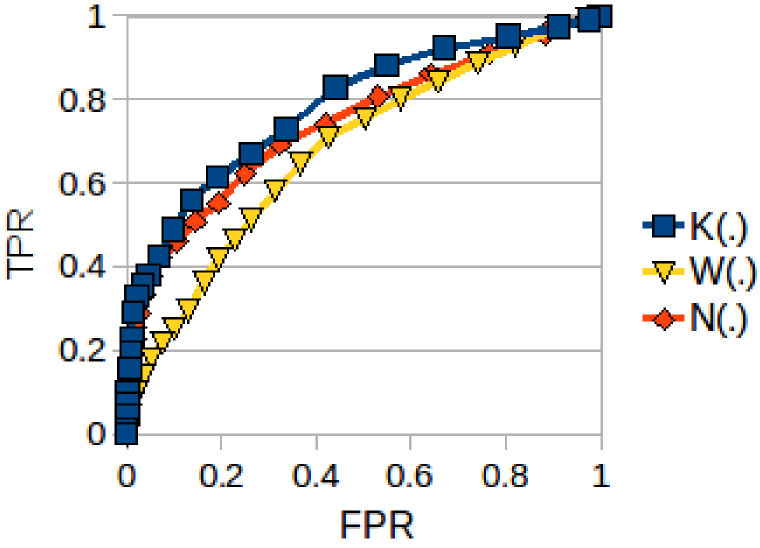


Figure 3.2: Normalisation operators comparison. Each plot is the average ROC curve over all images in the dataset, for a given normalisation operator. ROC curves were built by thresholding the final saliency map and comparing the resulting binarised image against the hand-labelled ground-truth of the dataset. All operators result in curves above the line of no-discrimination, $y = x$, thus showing the positive correlation between visual saliency and trail presence. Moreover, the higher area under the curve for the proposed model, $K(.)$, demonstrates that it is the most adequate for the task at hand.

3.2 Trail Detection Agents

Rather than considering image analysis as information processing, this work follows the idea of considering it as the result of a sensori-motor coordination process. Under this paradigm, the agent-based approach to image analysis, in particular for object recognition, is showing promising results [Floreano et al., 2004, Owechko and Medasani, 2005, de Croon and Postma, 2007, Choe et al., 2008]. This success story can be in part understood by the fact that agents realise active vision local loops, and thus exploiting all the known advantages of considering perception as an active process [Ballard, 1991]. Being this work in line with this novel way of developing robust perceptual systems, its potential success contributes to the body of evidence on the relevance of an agent-based design for perceptual systems.

In a context different from the one considered in this dissertation, i.e. road detection, the agent-based design has already and successfully been used [Broggi and Cattani, 2006]. Despite the fact that the work herein proposed focuses on trails instead of roads, some additional differences between this model and the one of Broggi & Cattani [Broggi and Cattani, 2006] can be observed. As it will be described, in this method agents inhabit conspicuity and saliency maps, rather than the image space itself. The focus is set on the structure being sought, i.e. the trail, and not on its boundaries. In addition, the hard assumption that the robot is on the trail or road is herein disregarded.

The system is composed of a set of agents, E_m, deployed in each conspicuity and saliency map $m \in \{\mathbf{C}^{\mathbf{C}}(t), \mathbf{C}^{\mathbf{I}}(t), \mathbf{S}(t)\}$, with width $w(m) = 320$ and height $h(m) = 240$. Each agent moves on one of these maps, according to a set of rules, in an attempt of following a given trail hypothesis.

3.2.1 Agent Recruitment

Let us first describe how agents are deployed in the three maps, which occurs according to the maps intensity level, i.e. the level of conspicuity or saliency, depending on the map in question. In order to avoid any noise potentially present at the map's boundaries, agents are deployed with a small offset of the bottom of the map in question, i.e. at row $r = h - 15$, where h is the height of the maps.

To determine the column where each agent is deployed, the unidimensional vector

$$\mathbf{v^m} = (v_0^m, \ldots, v_w^m) \tag{3.6}$$

is first computed, where w is the width of the maps. The element v_k^m of $\mathbf{v^m}$ refers to the average intensity of the pixels in column k, contained between row r and row $r - \delta$, where $\delta = 10$ to avoid deploying agents in columns with spurious high intensity pixels. Formally,

$$v_k^m = \sum_{l \in [r, r-\delta]} m(k,l)/\delta \tag{3.7}$$

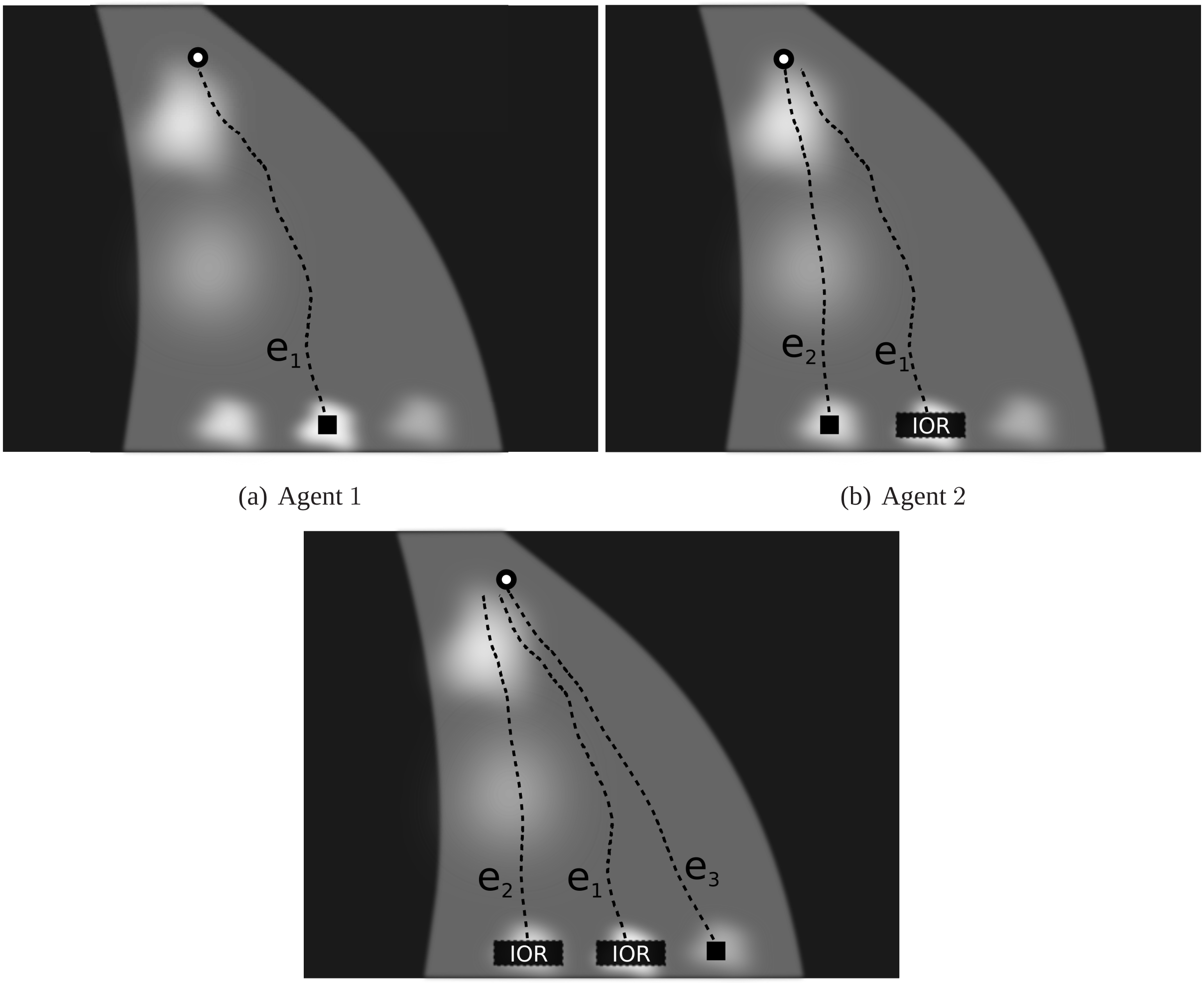
where $m(k,l)$ is the intensity or saliency level, depending on the map in question, at pixel in column k and row l.

Finally, the agent $e \in E_m$ is deployed in column

$$c(e) = \arg\max_k v_k^m \tag{3.8}$$

thus compelling it to be initiated in the most salient region, according to $\mathbf{v^m}$.

To analyse the second most salient region, an Inhibition-Of-Return (IOR) mechanism is used. This is implemented by zeroing the elements of $\mathbf{v^m}$ that are connected to $v_{c(e)}^m$ through elements with values similar to it. This agent deployment sequence is repeated until one of the following holds: (1) a maximum number of agents, z_{max}, has been deployed in the map or (2) the current highest value of $\mathbf{v^m}$, $\max(\mathbf{v^m})$, is below a fraction η of its initial value, i.e. before the first agent was deployed. In this work $\eta = 0.7$, which avoids the deployment of agents in low intensity (conspicuous/salient) regions. An illustration of the recruitment procedure is presented in Fig. 3.3.



(a) Agent 1

(b) Agent 2

(c) Agent 3

Figure 3.3: Agent deployment process. Illustrative example for 3 agents deployed in a map. (a) The first agent, e_1, is deployed in the region of the map with the highest intensity (conspicuity/saliency level), according to the vector $\mathbf{v^m}$. (b) Inhibition-Of-Return (IOR) is applied to $\mathbf{v^m}$, and the second agent, e_2, is deployed in the next highest intensity region. (c) Same procedure for the third agent, e_3. IOR is applied and e_3 is deployed. The dotted lines represent the agents motions since their onset (square), $o(0)$, until the current iteration (circle), $o(n)$, embedding the behaviours described in Section 3.2.2. The map depicted is hand-made and purely illustrative.

3.2.2 Agent Behaviours

Let us now describe the behaviour of each deployed agent. For simplicity, agents and maps indexes will be discarded in the remainder of this section. That is to say that the following applies to a single agent e allocated to a specific map m.

The set $O = \{1, 2, 3, 4, 5\}$ defines agent motor actions in terms of an index to the nearest neighbour pixels whereto the agent can move from its current position, $o(n)$, at iteration n (see Fig. 3.4). To reduce both sensitivity to noise and computational cost, the agent's surroundings are segmented into regions $R_1 \ldots R_5$ (see Fig. 3.4). The average intensity of a region containing pixel p is given by $A(p)$. For instance, both $A(1)$ and $A(6)$ correspond to the average intensity of the pixels contained within region R_1, as this region is composed of pixels 1, 6 and 11. Thus, regions are indirectly indexed by their encompassed pixels. The straight intensity of a pixel p is simply given by $Int(p)$.

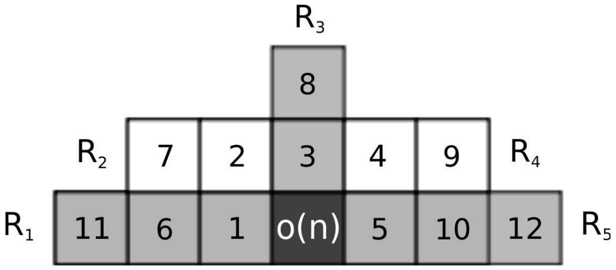


Figure 3.4: Agent neighbourhood relative pixel indexes. Numbers correspond to the pixels' index relative to the current position of the agent, $o(n)$. Regions surrounding $o(n)$ are segmented in $R_1 = \{1, 6, 11\}, R_2 = \{2, 7\}, R_3 = \{3, 8\}, R_4 = \{4, 9\}, R_5 = \{5, 10, 12\}$.

To account for top-down knowledge on the structure of the object being sought, a set of five perception-action rules, i.e. behaviours,

$$B = \{greedy, track, centre, ahead, commit\} \tag{3.9}$$

vote for each possible action, $a \in O$, according to the behaviour-based voting command fusion approach [Rosenblatt, 1997]. The most voted action, $a^+(n)$, is selected by the agent as the next motion, which is then used to update its position, $o(n)$,

$$\dot{o}(n) = \Gamma\big(a^{+}(n)\big), \;\; a^{+}(n) = \arg\max_{a \in O} \sum_{b \in B} w_b \cdot f_b(m, a, n) \tag{3.10}$$

where $\Gamma(.)$ transforms a motor action, $a \in O$, onto pixel coordinates centred on the current agent's position, w_b is the weight accounting for the contribution of behaviour $b \in B$, described by the evaluation function $f_b(m, a, n)$ as follows,

$$f_{track}(a, n) = 1 - \frac{\left|A(a) - \sum_{q \in Q} \frac{q}{n-1}\right|}{255} \tag{3.11}$$

$$f_{ahead}(a, n) = 1 - \frac{|3 - a|}{2} \tag{3.12}$$

$$f_{commit}(a, n) = 1 - \frac{|a^{+}(n-1) - a|}{4} \tag{3.13}$$

$$f_{centre}(a, n) = \left| d_x(n) \cdot \left(\frac{6 \cdot \mathcal{H}\big(-d_x(n)\big) - a}{5} \right) \right| \tag{3.14}$$

$$f_{greedy}(a, n) = \frac{A(a)}{255} \tag{3.15}$$

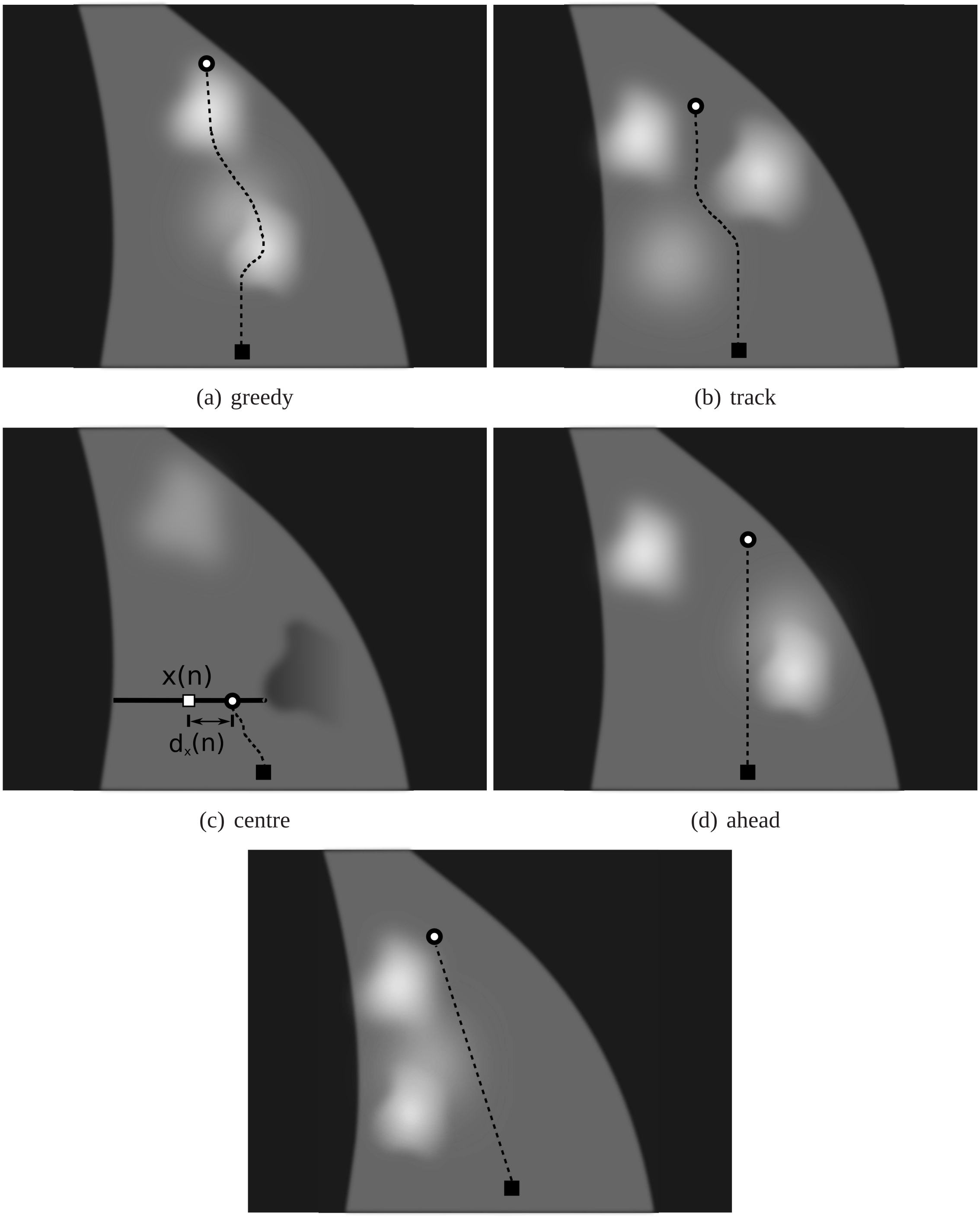
where $d(n)$ is computed as described in Fig. 3.5(a) and $\mathcal{H}(.)$ is the Heaviside function. Q is a set whose elements are scalars with the intensity of the pixels crossed by the agent along its path. Formally, $Int\big(o(n)\big)$, is inserted to Q as follows, $Q(n) \leftarrow Q(n-1) \bigcup \{Int\big(o(n)\big)\}$. Refer to Table 3.1 for further details on each behaviour and Fig. 3.1 for examples of agent typical motions. The best performance has been empirically obtained with the following trade-off, $w_{greedy} = 0.45$, $w_{track} = 0.35$, $w_{centre} = 0.10$, $w_{ahead} = 0.05$, $w_{commit} = 0.05$.

The agent is allowed to move until one of the following stopping conditions is met: (1) a maximum number of α_1 iterations is performed; (2) the agent reaches row α_2 (row zero at image's top); (3) the average intensity of regions $R_1 \ldots R_5$ is below a given proportion $\beta < 1$ of the average intensity of the pixels visited by the agent.

$$\beta \cdot \sum_{q \in Q} \frac{q}{n-1} > \sum_{j=1}^{5} \frac{A(j)}{5} \tag{3.16}$$

Behaviour	**Voting Preferences**
$greedy$	Regions of highest intensity, under the assumption that trails are the most salient structures in the map. See Fig. 3.5 (a) for an illustration of this process.
$track$	Regions whose average intensity is more similar to the average intensity of the pixels visited by the agent, under the assumption that trails' conspicuity is somewhat homogeneous. See Fig. 3.5 (b) for an illustration of this process.
$centre$	Regions closer to the centroid, $x(n)$, of the set of pixels, $S(n)$, that: (1) share the row with $o(n)$; (2) display intensities similar (i.e. within a given margin γ) to the one of $o(n)$; and (3) are connected to $o(n)$ through a set of pixels complying with the first two conditions. The goal is to maintain the agent equidistant to the trail's boundaries, where the deviation to the centroid is given by $d_x(n) = \frac{c(o(n))-c(x(n))}{D(n)}$ with $D(n) = \|S(n)\|$. Remember that $c(p)$ returns the column of pixel p. See Fig. 3.5 (c) for an illustration of this process.
$ahead$	Upwards regions under the assumption that trails appear as vertical elongated structures. See Fig. 3.5 (d) for an illustration of this process.
$commit$	Previously selected region, to reduce sensibility to local noise, under the assumption that trails' outline is somewhat monotonous. See Fig. 3.5 (e) for an illustration of this process.

Table 3.1: Behaviours ruling the trail detection agents. Illustrations of each behaviour are shown in Fig. 3.5.



(a) greedy

(b) track

(c) centre

(d) ahead

(e) commit

Figure 3.5: Behaviours ruling the trail detection agents. The dotted lines represent the agents motions since their onset (square), $o(0)$, until the current iteration (circle), $o(n)$. The map depicted is hand-made and purely illustrative. (a) Greedy behaviour. The agent will follow higher intensity regions. (b) Track behaviour. The agent will prefer regions similar to the ones already visited by it. (c) Centre Behaviour. The pixels composing the thicker horizontal line define the set $S(n)$. The agent will try to approach this line's centroid $x(n)$, represented by the white square, which is deviated from the current agent's position, $o(n)$, by $|d_x(n)|$ pixels. (d) Ahead behaviour. The agent will prioritize upwards regions. (e) Commit behaviour. The agent will follow the previously selected region.

where, $\alpha_1 = 50$, $\alpha_2 = 160$, and $\beta = 0.7$ are empirically defined scalars.

The set of agents deployed in a given map must be ranked in order to select the one that better represents the trail. Consequently, as soon as one of the previously mentioned stopping conditions is met, the score of the agent is computed,

$$\begin{aligned} s = & \sum_{i=0}^{n} \frac{\mu_1 D'(i) - \mu_2 D''(i)}{n} + \\ & \sum_{q \in Q} \frac{\mu_3 q}{n-1} + \mu_4 d\Big(o(n), o(0)\Big) \end{aligned} \tag{3.17}$$

where $\mu_1 = 0.01$, $\mu_2 = 0.01$, $\mu_3 = 0.5$, $\mu_4 = 0.5$ are empirically defined scalars and $d(o(n), o(0))$ is the Euclidean distance between the two points. The first parcel of the score function accumulates the first, D', and second, D'', derivatives of D along the agent's path. This parcel favours paths where D progressively shrinks towards a vanishing point. The second parcel promotes agents whose path contains highly salient pixels. Finally, the third parcel disfavours short paths.

3.3 Experimental Results

This section presents a set of experimental results obtained with a dataset composed of 50 colour images, with resolution 640×480, obtained from Google (see Appendix A, Figs. A.1-A.6). The dataset only encompasses images obtained with cameras roughly located at the eyes height, and thus providing a vantage point that would be plausible for a medium-size robot. The trail detector has been implemented without thorough code optimisation, and tested in a Centrino Dual Core 2 GHz, running Linux, and OpenCV for computer vision low-level routines.

Since the output generated by the trail detector is the set of the agents' paths, and not the trail's outline, it is difficult to find a way of comparing the results against some sort of ground truth. The following describes the assumptions taken to assess whether a given agent has been able to represent the trail. Trails are considered correctly detected if the agent is deployed inside the trail and finishes its run also inside, or very close to, the trail. In addition, curves and zigzags described by the agent are considered valid as long as they also stay inside the trail, or very close to its borders.

Table 3.2 summarises the results obtained as a function of the maximum number of allowed agents per map $z_{max} \in \{1, 3, 5, 7\}$. In a first analysis, success rate is calculated per map. This allows to determine the proneness of each map alone to provide enough cues for its highest score agent to properly represent the trail. In a second combined analysis, success is obtained when at least one of three map's best agent, succeeds. In this case, the ambiguity regarding both trail's position and approximate skeleton is of up to three hypotheses, i.e. one per map, in 98% of the tested images. This clearly shows that the proposed method accurately focus agents on the most promising regions of the image.

The obtained results also confirm the positive correlation between saliency and the presence of trails (see Fig. 3.2). Would this correlation be nonexistent and the trail detection results would be linearly affected by the number of agents. Instead, with a single agent per map, the trails in 90% of the images were properly detected, whereas an increment of only 6% is observed if two additional agents per map are deployed. An even smaller differential is obtained when we go from three to five agents, namely 2%. Adding more agents reflects in a null gain.

Figs. A.7, A.8 show the trail of the best agent per map in the images composing the dataset. These images are very diverse and in some cases no trail can be found altogether, not even by the human eye. The system still produces a correct answer, that is, selects the open region through which the robot would be able to traverse. This is a sign of generalisation capability, which was only possible due to the use of a non-specific detector, as it is the case of saliency.

Hence, even in the most difficult situations, saliency and conspicuity maps were able to maintain a globally coherent description of the environment. However, the existence of local intensity variations requires the system to have a considerable level of robustness in order to be unaffected by those local artifacts. The agent-based approach showed to be that robust, mostly due to the fusion of several behaviours. Moreover, being a purely bottom-up and feed-forward approach, the method is exceptionally fast, taking an average of $1\,\mathrm{ms}$ per map. This includes

Nr. of Agents	**Colour Map**	**Intensity Map**	**Saliency Map**	**Combined**
$z_{max} = 1$	$44\,\%$	$64\,\%$	$74\,\%$	$90\,\%$
$z_{max} = 3$	$54\,\%$	$78\,\%$	$82\,\%$	$96\,\%$
$z_{max} = 5$	$58\,\%$	$80\,\%$	$82\,\%$	$98\,\%$
$z_{max} = 7$	$58\,\%$	$80\,\%$	$82\,\%$	$98\,\%$

Table 3.2: Trail detection results.

finding all the potential trails, finding their length, and choosing the correct one. An additional cost must be considered, which refers to the computation of the three maps, which takes on average 30 ms. These maps have two remarkable embedded properties: (1) they segment the input image in a very efficient way, and (2) they naturally prioritise the segments according to their conspicuity.

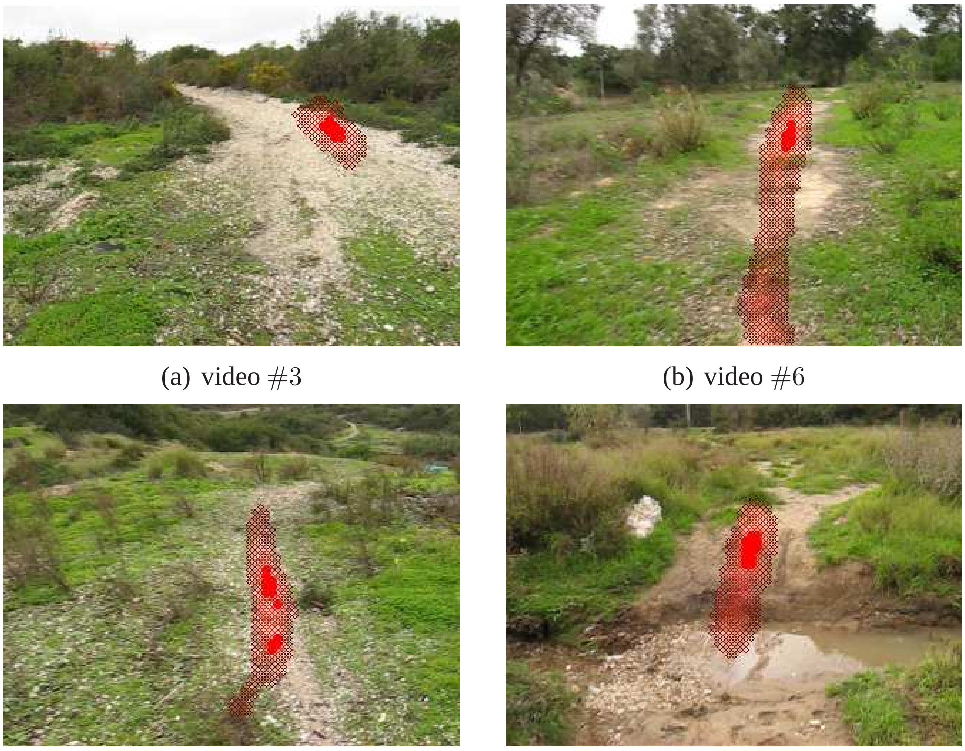
Chapter 4

Swarm-Based Visual Saliency for Trail Detection

In the previous chapter was shown that the saliency map of a given image corresponds itself to an efficiently computed segmentation of the latter. That is, the segmentation of the input image, which can be a computationally intensive task, can be obtained as a by-product of determining which regions of the visual field detach more from the background. Furthermore, the obtained segments are already prioritised by their conspicuity level. It was also shown that visual saliency and trail location in the input image are positively correlated.

From these findings it should follow that the highest priority segment in the saliency map matches the location of the trail in the input image. In practise, this is a brittle assumption in the face of not so well behaved saliency maps, which may occur in the presence of distractors or when the trail is considerably heterogeneous. This difficulty can be diminished with top-down boosting of visual features (e.g., colour) that are known to describe the object being sought [Frintrop et al., 2005, Navalpakkam and Itti, 2005]. However, these visual features are considerably unpredictable in the case of trails in natural environments. In opposition, trails' overall layout is a much more predictable feature. For example, the projection of trails onto the input image typically converges towards a vanishing point. This novel use of top-down knowledge was embedded in the previous model in the form of behaviours ruling the motion of simple agents inhabiting the saliency and its intermediate conspicuity maps. The motion paths of these agents were then taken as the skeleton of a set of trail hypotheses, which were then scored, and three of them selected as the output of the system.

Despite its overall good results, the previous model was unable to reduce the ambiguity of three trail hypotheses, it was brittle in the presence of interrupted trails, and it was unable to exploit historical information to improve its robustness. Fig. 4.2 depicts the model proposed in this chapter, which extends the previous one to overcome its limitations: (1) by allowing the agents to exhibit collective behaviour through pheromone-based interactions, and (2) by allowing the system to accumulate evidence about the most likely trail location across multiple frames through the use of a dynamic neural field. See Fig. 4.1 for typical results obtained with the extended model.



(a) video #3

(b) video #6

(c) video #24

(d) video #25

Figure 4.1: Typical trail detection results (red overlay) obtained with the swarm-based model. These results show that model is able to localise the trail even when it is highly interrupted, blends itself with the background, or does not start from the bottom of the image.

4.1 System Overview

In short, two conspicuity maps, $\mathbf{C}^{\mathbf{C}}(t) \in [0,1]$ for colour and $\mathbf{C}^{\mathbf{I}}(t) \in [0,1]$ for intensity information, are computed from the input image $\mathbf{I}(t)$, as described in Chapter 3, Section 3.1. A set of agents is then deployed on each map. These agents interact with the corresponding conspicuity map according to their perception-action rules, which embed the trail-specific top-down modulation process, as defined in Chapter 3, Section 3.2. During the process, pheromone is deployed and sensed by the agents in two pheromone fields, $\mathbf{P}^{\mathbf{C}}(t) \in [0,1]$ and $\mathbf{P}^{\mathbf{I}}(t) \in [0,1]$, according to the ant foraging metaphor. An additional perception-action rule is introduced to make the agents' behaviour sensible to the pheromone deployed by the swarm, and thus enabling coherent collective behaviour to emerge. This way, agents help each other on the task of perceptual completion, resulting in a global behaviour that is robust to the local variations inherent to trails.

Being the deployed pheromone a function of agents' sensations across their trajectories on the corresponding conspicuity maps, it is influenced by the activity occurring in distant regions of the map. This long-range spatial connectivity allows handling the potentially large size of trails in a robust and parsimonious way.

Rather than blending both conspicuity maps, $\mathbf{C}^{\mathbf{C}}(t)$ and $\mathbf{C}^{\mathbf{I}}(t)$, to generate the final saliency map $\mathbf{S}(t) \in [0,1]$, as typically done [Itti et al., 1998, Frintrop et al., 2005], in this work $\mathbf{S}(t)$ is obtained by blending both pheromone fields. The final saliency map $\mathbf{S}(t)$ feeds a dynamic neural field [Amari, 1977, Rougier and Vitay, 2006], $\mathbf{F}(t) \in [0,1]$, which integrates pheromone (i.e., evidence) across frames and also implements both lateral excitation and long-range inhibition. This neural field allows the system to maintain a coherent focus of attention across time [Rougier and Vitay, 2006]. Motion compensation is also implemented so that the dynamics of the neural field can be decoupled from the dynamics of the robot. The neural field's state feeds back both pheromone fields so that history influences agents' activity. The output of the system is given by the current state of the neural field, where the higher the activation of a given neuron the higher its chances of being associated to a trail pixel.

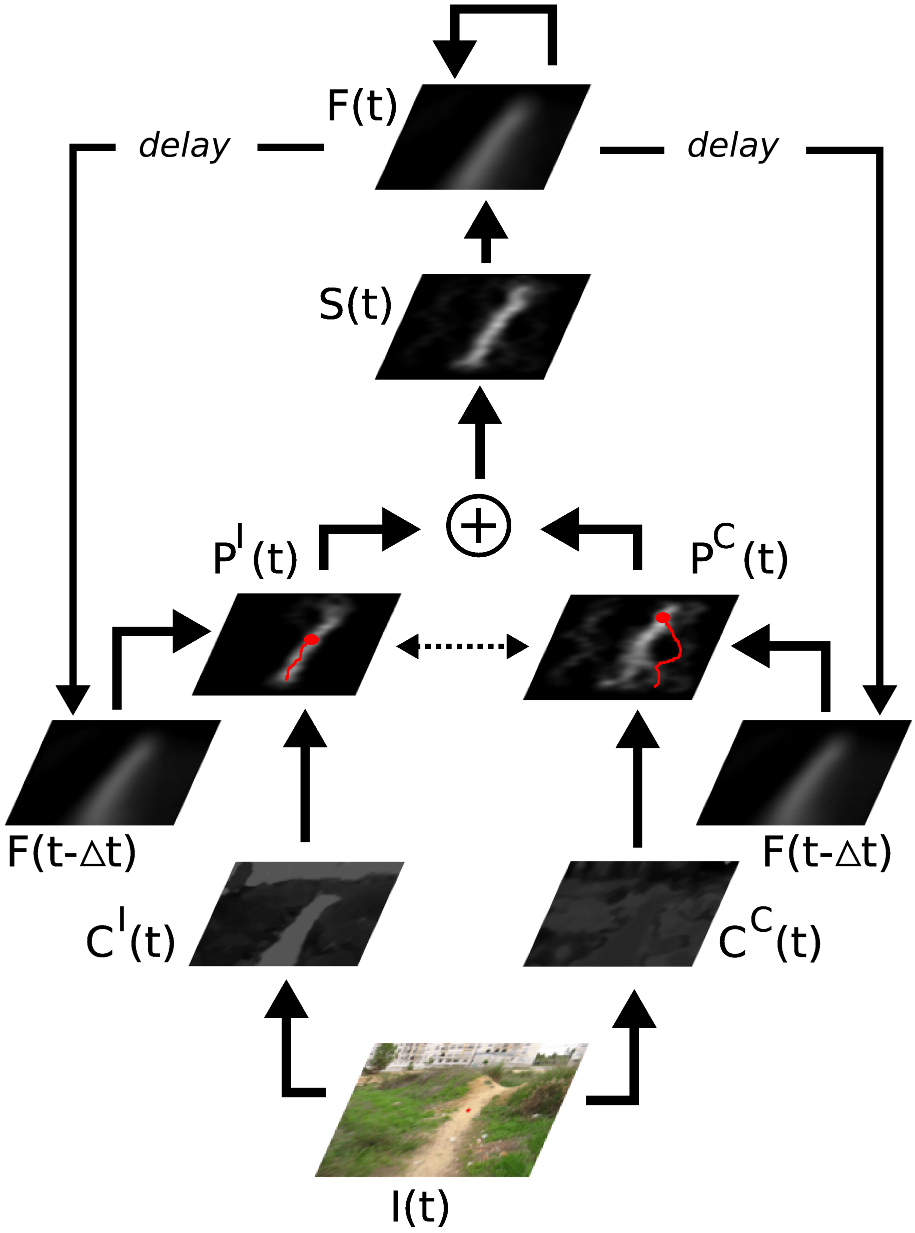


Figure 4.2: System overview. The red overlays in both pheromone fields, $\mathbf{P}^{\mathbf{C}}(t)$ and $\mathbf{P}^{\mathbf{I}}(t)$, are two illustrative agent paths. For the sake of clarity, motion compensation aspects are not represented.

4.2 Conspicuity Maps Computation

Conspicuousness computation is about determining which regions of the input image detach from the background at several scales and feature channels. In this model, as in the previous one, only intensity and colour channels are used.

Shortly, one dyadic Gaussian pyramid with eight levels 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 colour feature channels. The various scales are then used to perform centre-surround operations [Itti et al., 1998]. The resulting centre-surround maps have higher intensity on those pixels whose corresponding feature differs the most from their surroundings. An example is a bright patch on a dark background (on-off), as well as the other way around (off-on). On-off centre-surround operations are performed 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. Then, the centre-surround maps are blended to produce a colour conspicuity map, $\mathbf{C}^{\mathbf{C}}(t) \in [0, 1]$, and an intensity conspicuity map, $\mathbf{C}^{\mathbf{I}}(t) \in [0, 1]$. The width, w, and height, h, of both maps is 80 and 60, respectively.

When blending maps, the most discriminant ones are promoted by recurring to a normalisation operator. Here the normalisation operator described in the previous model is followed, which was shown to outperform other known models [Itti et al., 1998, Frintrop et al., 2005] in trail detection. Please refer to Chapter 3, Section 3.1 for further details and to Fig. 4.2 for examples of conspicuity maps.

4.3 Collective Behaviour

This section describes how an agent deployed on a conspicuity map $m \in \{\mathbf{C}^{\mathbf{C}}(t), \mathbf{C}^{\mathbf{I}}(t)\}$ behaves in order to generate a pheromone field $p \in \{\mathbf{P}^{\mathbf{C}}(t), \mathbf{P}^{\mathbf{I}}(t)\}$, in cooperation with other agents, whose activity level is correlated with the localisation of the trail. If the agent is allocated to the colour conspicuity map, $\mathbf{C^C}(t)$, then it contributes to the colour pheromone field, $\mathbf{P^C}(t)$. Analogously, if allocated to the intensity conspicuity map, $\mathbf{C^I}(t)$, the agent contributes to the intensity pheromone field, $\mathbf{P^I}(t)$. The process through which agents are deployed in the maps is also explained in this section.

4.3.1 Agent Behaviours

At the onset of each frame, both pheromone fields are zeroed and subsequently affected by a small ratio λ of the robot motion compensated neural field's previous state, $\mathbf{F}'(t-\Delta t)$,

$$\mathbf{P^C}(t) = \mathbf{P^I}(t) = \lambda \mathbf{F}'(t-\Delta t) \tag{4.1}$$

In this study $\lambda = 0.1$. Refer to Section 4.4 for details on the computation of $\mathbf{F}'(t-\Delta t)$. This pheromone level offset allows agents' activity to be affected by history, which induces stability, robustness to noise and across-frames progressive improvement.

For a given number $n_{max} = 50$ of iterations, whose index is represented by n, the agent builds up a trail hypothesis by updating its position, $o(n)$, according to a set of behaviours B, which are sensible to the level of conspicuity in the agent's surroundings. These behaviours embed top-down information on the object being sought, such as its approximate shape. The agent's motion is also affected by other agents' activity according to the ant foraging metaphor, i.e., via *stigmergy*. That is, agents interact with each other through a pheromone field built by them while moving. Conspicuity-based behaviours and pheromone influence contribute to the agent's motion according to the following voting mechanism,

$$a^+(n) = \arg\max_{a \in O} \left(\sum_{b \in B} \alpha_b f_b(m,a,n) + \beta g(p,a,n) + \gamma q \right) \tag{4.2}$$

$$\dot{o}(n) = \Gamma\Big(a^+(n)\Big) \tag{4.3}$$

where: O is the set of possible agent motor actions (e.g., “move to the right”); $\Gamma(.)$ transforms a motor action, $a \in O$, onto pixel coordinates centred on the current agent’s position; β is the weight accounting for the contribution of pheromone, which is described by the motor action evaluation function $g(p, a, n) \in [0, 1]$; α_b is the weight accounting for the contribution of behaviour $b \in B$, which is described by the motor action evaluation function $f_b(m, a, n) \in [0, 1]$; and γ is the weight accounting for stochastic behaviour, being $q \in [0, 1]$ a number sampled from a uniform distribution each time the action is evaluated.

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

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

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

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

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

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

The newly proposed evaluation function $g(p, a, n)$ greedily provides higher score to the motor actions that take the agent to regions of higher level of pheromone. By making the score proportional to the level of pheromone, this evaluation function guides the agent towards regions recurrently visited by other agents. The outcome is coordinated collective behaviour. By the end of each iteration, the agent contributes to pheromone field p by deploying an amount of pheromone ϵ in its current position, $o(n)$, and to the other pheromone field p' a small portion of ϵ, υ. That is, if $p = \mathbf{P^C}(t)$ then $p' = \mathbf{P^I}(t)$, and the other way around. This process enables loosely coupled cross-modality influence, thus allowing each agent to exploit multiple cues indirectly, and therefore to maintain their simplicity. In this study $\epsilon = 0.008$ and $\upsilon = 0.3$.

The ratio used to control the importance of the collective over the individual experience,

$$\beta/(\sum_{b \in B} \alpha_b + \gamma) \tag{4.4}$$

has, in this study, $\beta = 1.0$ and $\gamma = 0.8$. Please refer to Chapter 3, Section 3.2 for further details on the agent motor actions set O, on the behaviour set B, on its associated weights α_b, and on how the agent's local surroundings is segmented into regions.

4.3.2 Agent Recruitment

A set of agents, E_m, is deployed at each conspicuity map $m \in \{\mathbf{C^C}(t), \mathbf{C^I}(t)\}$. The chance of deploying an agent on a given location of map m depends on the level of conspicuity at that location and on the level of pheromone at the same position of the corresponding pheromone field p. The following describes in detail the deployment process.

To avoid any noise potentially present at the map's boundaries, agents are deployed with a small offset of the bottom of the conspicuity map in question, i.e., at row $r = h - 5$, where h is the height of the conspicuity maps.

To determine the column where each agent is deployed, the unidimensional vector

$$\mathbf{v^m} = (v_0^m, \ldots, v_w^m) \tag{4.5}$$

is first computed, where w is the width of the conspicuity maps. The element v_k^m of $\mathbf{v^m}$ refers to the average conspicuity level of the pixels in column k, contained between row r and row $r - \delta$, where $\delta = 5$ to avoid deploying agents in columns with spurious highly conspicuous pixels. Formally,

$$v_k^m = \sum_{l \in [r, r-\delta]} m(k,l)/\delta \tag{4.6}$$

where $m(k,l)$ is the conspicuity level at pixel in column k and row l.

The same process is repeated to build a vector for the pheromone field in question,

$$\mathbf{v^p} = (v_0^p, \ldots, v_w^p) \tag{4.7}$$

where $p(k,l)$ is the pheromone level at pixel in column k and row l. In this case,

$$v_k^p = \sum_{l \in [r, r-\delta]} p(k,l)/\delta \tag{4.8}$$

Then, the test

$$z < (v_{j \cdot w}^m + \max(v_{j \cdot w-4}^p, v_{j \cdot w+4}^p)) \tag{4.9}$$

is repeated until it succeeds, where $z \in [0,1]$ and $j \in [0,1]$ are numbers sampled from a uniform distribution each time the test is performed. At that time, the agent is deployed in column $j \cdot w$. With this test, the chance of deploying an agent in a randomly selected column $j \cdot w$ is as high as the conspicuity and pheromone levels at the deployment region. This sampling process is repeated until $|E_m| = 20$ agents are deployed per map m.

4.4 Evidence Accumulation

To integrate evidence across time, to consider competition between multiple focus of attention, and to promote perceptual grouping, the fusion of both pheromone fields,

$$\mathbf{S}(t) = \frac{1}{2}\mathbf{P^C}(t) + \frac{1}{2}\mathbf{P^I}(t) \tag{4.10}$$

feeds a 2-D dynamic neural field $\mathbf{F}(t)$. Note that this process only occurs after the agents' activity has ceased, and therefore the pheromone fields have been fully updated.

The dynamical characteristic of the neural fields [Amari, 1977, Rougier and Vitay, 2006] is what enables their ability to integrate information across time. To avoid the blurring of the neural field when the robot moves, 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 $\mathrm{H}(t)$ that describes the perspective transformation between the current frame, $\mathbf{I}(t)$, and the previous one, $\mathbf{I}(t-\Delta t)$. This step is further detailed in Section 4.4.1.

2. Obtain a perspective compensated version of the previous neural field's state by using the estimated homography matrix,

$$\mathbf{F}'(t-\Delta t) = \mathbf{H}(t)\mathbf{F}(t-\Delta t) \tag{4.11}$$

3. Obtain $\mathbf{F}(t)$ by updating the perspective compensated neural field $\mathbf{F}'(t-\Delta t)$ with the pheromone field $\mathbf{S}(t)$. This step is further detailed in Section 4.4.2.

4.4.1 Homography Matrix Estimation

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

$$\mathbf{u}'_i = \mathrm{H}(t)\mathbf{u_i} \tag{4.12}$$

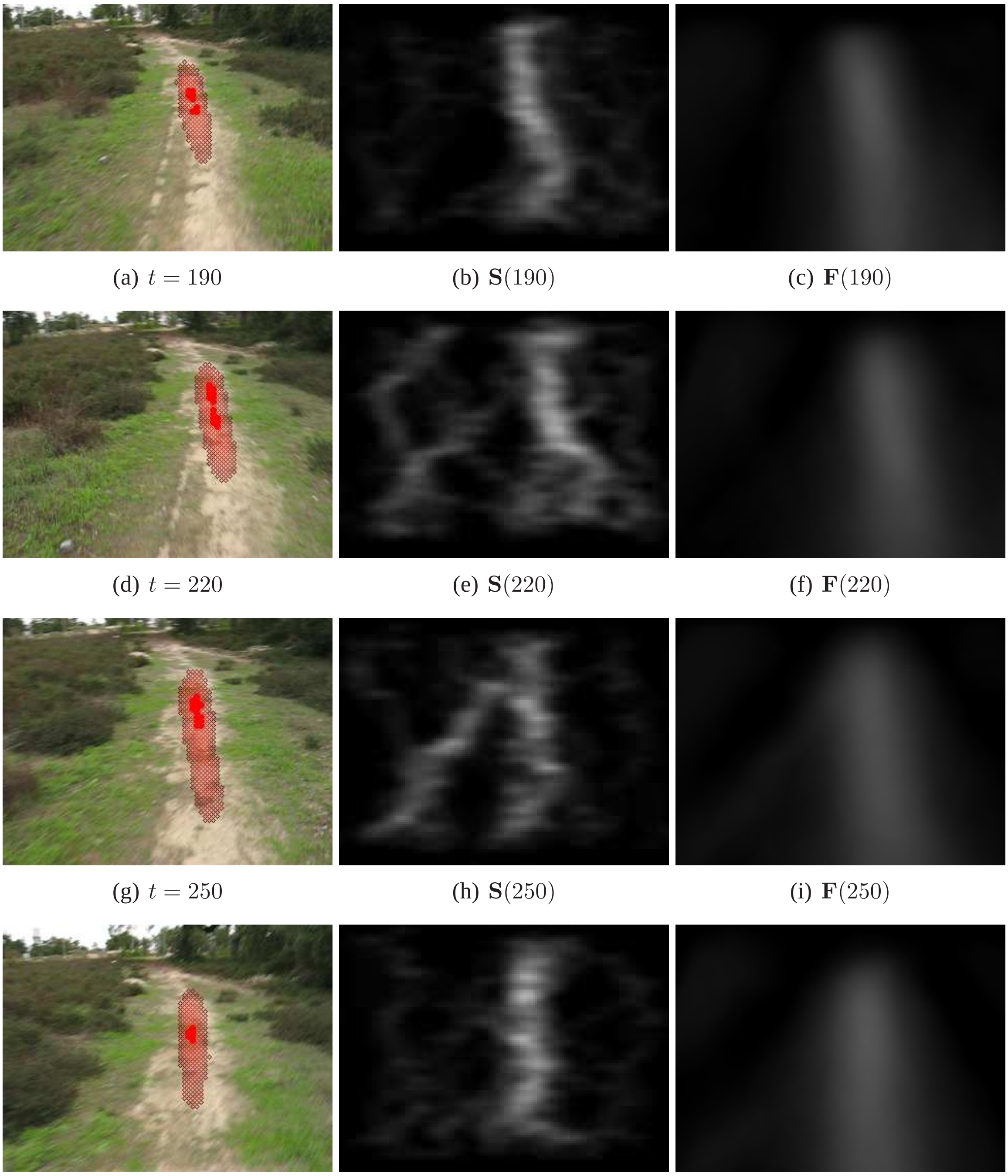
where $\mathbf{u_i}$ is a local feature found in $\mathbf{I}(t-\Delta t)$ and $\mathbf{u'_i}$ its correspondence in $\mathbf{I}(t)$. Due to noise in the tracking process, the homography matrix is calculated as the least-squares solution that minimises the back-projection error [Bradski and Kaehler, 2008]. This process assumes that distortion introduced by the camera lens into the input images has been corrected. It also assumes that either: (1) the terrain in front of the robot is planar or (2) the camera was only rotated, and not displaced, between frames. None of these two constraints can be strictly ensured in off-road environments. Still, in most situations the terrain is somewhat planar and the attitude of the camera changes more significantly than its position. Experiments have shown that the co-occurrence of these two relaxed constraints is sufficient to maintain a robust operation. If a minimum of four correspondences is not found, the homography matrix is set to the identity matrix, $\mathrm{H}(t) = diag(1, 1, 1)$.

4.4.2 Neural Field Update

The neural field $\mathbf{F}(t)$ is a 2D lattice of $w \times h$ neurons with "Mexican-hat"-shaped lateral coupling. This pattern of connectivity helps in the formation of a coherent focus of attention [Rougier and Vitay, 2006]. On the one hand, activated neurons excite their neighbours, thus promoting perceptual grouping. On the other hand, activated neurons tend to inhibit distant ones, thus reducing ambiguities in the focus of attention. Formally, the connection's weight between a neuron in position **x** and a neuron in position **x**′ is given by a Difference of Gaussians (DoG), function of the Euclidean distance between both, $w(\mathbf{x}, \mathbf{x}')$.

In addition to lateral connectivity, the neural field also has afferent interactions with pheromone field $\mathbf{S}(t)$. The weight of a connection between an element of $\mathbf{S}(t)$ in position **y** and a neuron of $\mathbf{F}(t)$ in position **x** is given by a Gaussian function of the Euclidean distance between both, $d(\mathbf{x}, \mathbf{y})$. This operation enlarges neurons' receptive field to reduce sensitivity to noise.

The average membrane potential of a given neuron at position **x** can now be expressed by the following nonlinear integro-differential equation,



(a) $t = 190$ (b) $\mathbf{S}(190)$ (c) $\mathbf{F}(190)$

(d) $t = 220$ (e) $\mathbf{S}(220)$ (f) $\mathbf{F}(220)$

(g) $t = 250$ (h) $\mathbf{S}(250)$ (i) $\mathbf{F}(250)$

(j) $t = 280$ (k) $\mathbf{S}(280)$ (l) $\mathbf{F}(280)$

Figure 4.3: Example of neural field competition in a situation represented by four ordered frames obtained from video $\#11$ of the tested dataset. The trail is present in the input image for several frames prior to $t = 220$, thus eliciting high level of activity in the neural field, $\mathbf{F}(190)$. Although the transient appearance of a trail-like grass segment in the bottom-left region of the image is felt in the pheromone field, $\mathbf{S}(220)$ and $\mathbf{S}(250)$, this distractor is actively inhibited in the neural field, $\mathbf{F}(220)$ and $\mathbf{F}(250)$.

$$\begin{aligned}\tau\frac{\partial\mathbf{F}(\mathbf{x},t)}{\partial t} = & -\mathbf{F}(\mathbf{x},t)+\\ & \int w(\mathbf{x},\mathbf{x}')\,f\left(\mathbf{F}(\mathbf{x}',t)\right)\mathbf{dx}'+\\ & \int d(\mathbf{x},\mathbf{y})\mathbf{S}(\mathbf{y},t)\mathbf{dy}+h\end{aligned} \tag{4.13}$$

where $f(x) = x$ in this work, τ is a time constant and $h = 0$ is the neuron threshold. For numerical integration, the Euler forward method is used to obtain an approximation of the neural field, which in matrix form results in the following rearranged expression,

$$\begin{aligned}\mathbf{F}(t) = \mathbf{F}'(t-\Delta t) + \frac{\Delta t}{\tau}\Big(& -a\cdot\big(\mathbf{F}'(t-\Delta t)\big)+ \\ & b\cdot\big(DoG^{k_1,k_2}_{\sigma_1,\sigma_2} * \mathbf{F}'(t-\Delta t)\big)+\\ & c\cdot\big(G^{k_3}_{\sigma_3} * \mathbf{S}(t)\big)+h\Big)\end{aligned} \tag{4.14}$$

where $*$ is the convolution operator, a, b and c are weights defining the contribution of each term, $DoG^{k_1,k_2}_{\sigma_1,\sigma_2} = G^{k_1}_{\sigma_1} - G^{k_2}_{\sigma_2}$, G^{k}_{σ} is a Gaussian kernel of size $k \times k$ and width σ. Note that the neural field's previous state, $\mathbf{F}(t-\Delta t)$, is substituted by its motion compensated counter part, $\mathbf{F}'(t-\Delta t)$. The neural field free parameters have been empirically defined, $\sigma_1 = 4.25$, $\sigma_2 = 14.15$, $\sigma_3 = 2.15$, $k_1 = 25$, $k_2 = 91$, $k_3 = 11$, $a = 2$, $b = 2.5$, $c = 8$, and $\frac{\Delta t}{\tau} = 0.03$. The system showed robustness to small variations around these values as long as the proportions are roughly maintained.

To enable fast computation, the model is synchronously evaluated, meaning that at time t neurons are updated based on the network state at time $t-\Delta t$. Due to robot motion, any potential symmetry at the sensory input does not prevail, making neural field oscillations unlikely to occur over relevant periods of time.

The dynamical characteristic of the model in conjunction with the long-range lateral inhibition results in the following property. The higher the number of frames with the same spot with high activity the more difficult it is, due to lateral connectivity, for other regions to become activated. Hence, transient distractors are actively inhibited once a large evidence on the trail location is accumulated (see Fig. 4.3).

4.5 Experimental Results

An extensive dataset of 25 colour videos encompassing a total of 12023 frames with a resolution of 640×480 has been obtained with a hand-held camera (see Appendix B). The camera was carried at an approximate height of $1.5\,\mathrm{m}$ and at an approximate speed of $1\,\mathrm{ms}^{-1}$. The trail detector was evaluated on a Core2 Duo 2.8 GHz running Linux. OpenCV was used for low-level routines. Table 4.1 shows that the model runs on average at $20\,\mathrm{Hz}$, where only 4% refers to the swarm-based activity. The timing reported for the neural field update also includes optical flow computation, homography estimation, and neural field wrapping.

The experimental results are twofold. First it is shown that the proposed swarm-based saliency model is more robust than a classical one [Itti et al., 1998, Frintrop et al., 2005], where conspicuity maps are blended,

$$\mathbf{S}(t) = \frac{1}{2}\mathbf{C}^{\mathbf{C}}(t) + \frac{1}{2}\mathbf{C}^{\mathbf{I}}(t) \tag{4.15}$$

rather than their corresponding pheromone fields,

$$\mathbf{S}(t) = \frac{1}{2}\mathbf{P}^{\mathbf{C}}(t) + \frac{1}{2}\mathbf{P}^{\mathbf{I}}(t) \tag{4.16}$$

For the sake of fair comparison, the neural field $\mathbf{F}(t)$, which is fed by $\mathbf{S}(t)$, is used to generate the output in both cases. Then, a qualitative comparison with related trail detectors highlights the advantages of the proposed model. To handle the probabilistic nature of the agents behaviours, a set of 5 runs was performed per video.

The trail is considered correctly localised if the biggest blob of neural field activity above 0.85 (from a maximum of 1) is fully within the trail's boundaries. In cases of ambiguity caused

	Neural Field	**Conspicuity Maps Computation**	**Swarm Computation**	**Total**
time (ms)	12	36	2	$\mathbf{50}$

Table 4.1: Average computation times.

by co-occurrence of two similar blobs, the pheromone field $\mathbf{S}(t)$ is used to assess which blob is being reinforced and consequently should be taken as the output.

A comparative analysis between Table 4.2 and Table 4.3 reveals that the proposed swarm-based saliency model clearly outperforms the classical one. That is, a higher average success rate is obtained along with a smaller standard deviation. It follows from the success rate of $91\% \pm 12\%$ that the proposed model is well suited for off-road autonomous robots. This result is more stringent if the difficulty of the tested dataset is taken into account. To our knowledge no previous work has been tested against a dataset with trails as narrow, unstructured and discontinuous as the ones herein considered. Moreover, differently from previous works [Rasmussen and Scott, 2008a], [Fernandez and Price, 2005], [Blas et al., 2008], and [Rasmussen et al., 2009], the model succeeds in situations where the trail is not starting from the bottom of the image (see Fig. 4.1(a)).

It is also worth noting that in 7 of the 25 videos, the proposed model shows 100% success rate for all the 5 five runs. Video 5 is accounted as a long run with almost 5 minutes length. Besides being often interrupted and highly unstructured, the trail in this video also exhibits a variable width. Moreover, the terrain surrounding the trail is heterogeneous and highly populated with potential distractors, such as trees and bushes. The 85% success rate of the model in this video clearly shows its robustness in demanding situations. About 5% of the fail cases refer to situations where the trail is nevertheless noticeable in the neural field. In this case, as in other lower performance videos, ambiguity between trail and surroundings could be reduced by considering additional perceptual modalities, such as texture and depth.

When the trail is highly conspicuous in the environment, as most often occurs, ambiguity is rarely present. When this assumption fails and distractors are scattered, the model is still able to perform correctly. This robustness owes to the agents' sensori-motor coordination capabilities, which allow an opportunistic exploitation of the trail-background segmentation present in the conspicuity maps.

Video ID	Nr. of Frames	Nr. of Correct Frames	% of Correct Frames
1	278	124	44.60
2	204	126	61.76
3	422	20	4.74
4	135	0	00.00
5	2854	927	32.48
6	186	52	27.96
7	121	0	00.00
8	124	0	00.00
9	309	58	18.77
10	147	73	49.66
11	386	0	00.00
12	158	0	00.00
13	134	54	40.30
14	676	299	44.23
15	683	181	26.50
16	770	35	4.55
17	403	141	34.99
18	335	325	97.01
19	230	195	84.78
20	439	28	6.38
21	490	18	3.67
22	230	25	10.87
23	600	36	6.00
24	802	0	00.00
25	907	0	00.00
	$\sum = 12023$	$\sum = 2717$	$(\mu \pm \sigma) = (\mathbf{23.97} \pm \mathbf{27.73})$

Table 4.2: Trail detection results - Classic saliency computation: $\mathbf{S}(t) = \frac{1}{2}\mathbf{C}^{\mathbf{C}}(t) + \frac{1}{2}\mathbf{C}^{\mathbf{I}}(t)$.

Video ID	Nr. of Frames	Average Nr. of Correct Frames	Average % of Correct Frames
1	278	124	44.60
2	204	126	61.76
3	422	20	4.74
4	135	0	00.00
5	2854	927	32.48
6	186	52	27.96
7	121	0	00.00
8	124	0	00.00
9	309	58	18.77
10	147	73	49.66
11	386	0	00.00
12	158	0	00.00
13	134	54	40.30
14	676	299	44.23
15	683	181	26.50
16	770	35	4.55
17	403	141	34.99
18	335	325	97.01
19	230	195	84.78
20	439	28	6.38
21	490	18	3.67
22	230	25	10.87
23	600	36	6.00
24	802	0	00.00
25	907	0	00.00
	$\sum = 12023$	$\sum = (\mathbf{10577.60 \pm 109.80})$	$(\mu \pm \sigma) = (\mathbf{91.32 \pm 1.01})$

Table 4.3: Trail detection results - Proposed saliency computation: $\mathbf{S}(t) = \frac{1}{2}\mathbf{P}^{\mathbf{C}}(t) + \frac{1}{2}\mathbf{P}^{\mathbf{I}}(t)$.

Chapter 5

Conclusions, Contributions and Future Work

This chapter summarises the work presented in this dissertation, providing a set of conclusions and contributions concerning the proposed models and the results obtained, as well as some aspects for future work.

5.1 Conclusions

This dissertation reported for the first time the use visual saliency to the trail detection problem. The model showed to be a computationally efficient solution with overall good results (91% success rate at $20\,$Hz), performing in situations where previous detectors tend to fail, such as when the trail does not emerge from the lower part of the image or when it is considerably interrupted. These results are mostly due to the effective segmentation obtained through the visual saliency method, and to the swarm-based design used to exploit this information. Furthermore, no hard assumptions on the appearance and morphology of the trails are done, conversely to most of the solutions proposed so far, which makes this model-free approach suitable for diverse and demanding natural environments. To our knowledge, this work is the most complex application of the agent-based sensori-motor coordination approach to object detection. The work was presented in two parts.

The first part focused on a saliency-based method using simple agents to exploit the saliency maps, with the purpose of validating the application of visual saliency to agent-based trail detection. A positive correlation was shown to exist between visual saliency and trail location. This preattentive mechanism also revealed as a promising method for fast prioritised (according to saliency) segmentation of the input image. Furthermore, seeing trails as conspicuous parts of the scene allowed the system to generalise. That is to say that in situations where trails could be hardly identified, even by the human eye, the system reported as trail open regions of the environment. A newly proposed normalisation operator for saliency computation played an important role in this achievement. The good prioritised segmentation properties exhibited by the visual saliency method, though not sufficient for accurate trail detection, present a good basis for boosting a focused detector.

To rapidly extract the trail skeleton from the prioritised saliency maps, an agent-based solution was proposed. This approach showed to be adequate, with experimental results showing that up to three trail hypotheses are generated by the method, being at least one of them correct in 98% of the cases. These results contribute to the growing evidence of agent-based approaches for the development of robust perceptual systems. This model is also innovative on the way top-down knowledge of the object being sought is considered. Typically, visual features are boosted according to the expected object's scale, colour and intensity [Navalpakkam and Itti, 2005]. Instead, in this work the object's (trail) approximate shape is implicitly considered, by means of feed-forward and consequently fast perception-action rules dictating the behaviour of each agent. Despite its overall good results, the model showed an inability to overcome some difficulties, namely: (1) it was unable to reduce the ambiguity of three trail hypotheses; (2) it was brittle in the presence of interrupted trails; and (3) it was unable to exploit historical information to improve its robustness.

In the second part of the work, the model was extended to overcome these difficulties by: (1) allowing the agents to exhibit collective behaviour; and (2) allowing the system to accumulate and make use of historical information. Hence, this part presented a swarm-based solution for trail detection, in which agents influence each other through shared mediums, i.e., via *stigmergy*. Evidence of trail location is accumulated across frames in a motion compensated dynamic neural field.

This solution has been successfully validated against a highly demanding and diverse dataset composed by video sequences, exhibiting 91% success rate at $20\,$Hz. These results due to large extent to the swarm-based design, which enabled a robust self-organisation of visual search, perceptual grouping, and multiple hypotheses tracking. The dynamic neural field showed to be a fast and efficient means for the integration of evidence across frames, implementing both lateral excitation and long-range inhibition, which increased the resilience of the system in the presence of distractors, namely rocks and trail-like grass. The motion compensation allowed the dynamics of the neural field to be decoupled from the dynamics of the robot, thus contributing to the stability of the system in outdoor environments. Finally, the high success rate across the diverse dataset shows that the selected parametrisation is not over-fit to a specific environment, thus highlighting its robustness.

5.2 Future Work

A more extensive sensitivity analysis of the model still needs to be addressed in future work. In this context, a mechanism for the self-parametrisation of the system can be considered. Other perceptual modalities, such as texture and depth, can be further analysed as alternative or complement to the used colour and intensity conspicuity maps, and might be included to further increase the robustness of the model. Testing the swarm-based saliency model to other visual search tasks is also object of future prospect. Lastly, the implementation of the detector in a physical robot and its testing in the mentioned environments will allow the assessment of the applicability of the proposed model.

Bibliography

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

[Apostoloff and Zelinsky, 2003] Apostoloff, N. and Zelinsky, A. (2003). Robust vision based lane tracking using multiple cues and particle filtering. In *Proc. of the IEEE Intelligent Vehicles Symposium*, pages 558–563.

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

[Bartel et al., 2007] 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 *Proc. of the IASTED International Conference on Robotics, Applications and Telematics*, pages 477–482.

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

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

[Bradski and Kaehler, 2008] Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.

[Broggi and Cattani, 2006] 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.

[Choe et al., 2008] Choe, Y., Yang, H. F., and Misra, N. (2008). Motor system's role in grounding, receptive field development, and shape recognition. In *Proc. of the International Conference on Development and Learning*, pages 67–72.

[Cremean et al., 2006] Cremean, L., Foote, T., Gillula, J., Hines, G., Kogan, D., Kriechbaum, K., Lamb, J., Leibs, J., Lindzey, L., Rasmussen, C., Stewart, A., Burdick, J., and Murray, R. (2006). Alice: an information-rich autonomous vehicle for high-speed desert navigation. *Journal of Field Robotics*, 23(9):777–810.

[Dahlkamp et al., 2006] Dahlkamp, H., Kaehler, A., Stavens, D., Thrun, S., and Bradski, G. (2006). Self-supervised monocular road detection in desert terrain. In *Proc. of Robotics: Science and Systems*.

[de Croon and Postma, 2007] de Croon, G. and Postma, E. O. (2007). Sensory-motor coordination in object detection. In *Proc. of the IEEE Symposium on Artificial Life*, pages 147–154.

[Dorigo and Stützle, 2004] Dorigo, M. and Stützle, T. (2004). *Ant Colony Optimization*. MIT Press.

[Duda and Hart, 1973] Duda, R. and Hart, P. (1973). *Pattern recognition and scene analysis*. Wiley, New York.

[Dunlop et al., 2007] Dunlop, H., Thompson, D., and Wettergreen, D. (2007). Multi-scale features for detection and segmentation of rocks in mars images. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7.

[Felzenszwalb and Huttenlocher, 2004] Felzenszwalb, P. and Huttenlocher, D. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181.

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

[Floreano et al., 2004] Floreano, D., Toshifumi, K., Marocco, D., and Sauser, E. (2004). Co-evolution of active vision and feature selection. *Biological Cybernetics*, 90(3):218–228.

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

[Frintrop et al., 2005] Frintrop, S., Backer, G., and Rome, E. (2005). Goal-directed search with a top-down modulated computational attention system. *Lecture Notes In Computer Science*, LNCS 3663:117–124.

[Hong et al., 2002] Hong, T. H., Rasmussen, C., Chang, T., and Shneier, M. (2002). Fusing ladar and color image information for mobile robot feature detection and tracking. In *Proc. of the International Conference on Intelligent Autonomous Systems*, pages 124–133.

[Itti et al., 1998] 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.

[Jain and Dubes, 1988] Jain, A. and Dubes, R. (1988). *Algorithms for clustering data*. Prentice Hall Englewood Cliffs, NJ.

[Jain and Farrokhnia, 1991] Jain, A. and Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern recognition*, 24(12):1167–1186.

[Kim et al., 2007] Kim, D., Oh, S., and Rehg, J. (2007). Traversability classification for ugv navigation: a comparison of patch and superpixel representations. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3166–3173.

[Leung and Malik, 2001] Leung, T. and Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44.

[Liapis et al., 2004] Liapis, S., Sifakis, E., and Tziritas, G. (2004). Colour and texture segmentation using wavelet frame analysis, deterministic relaxation, and fast marching algorithms. *Journal of Visual Communication and Image Representation*, 15(1):1–26.

[Martin et al., 2004] Martin, D., Fowlkes, C., and Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549.

[Martin-Herrero et al., 2004] Martin-Herrero, J., Ferreiro-Arman, M., and Alba-Castro, J. (2004). Grading textured surfaces with automated soft clustering in a supervised som. *Lecture Notes In Computer Science*, LNCS 3212(5):323–330.

[Moren et al., 2008] Moren, J., Ude, A., Koene, A., and Cheng, G. (2008). Biologically based top-down attention modulation for humanoid interactions. *International Journal of Humanoid Robotics*, 5(1):3–24.

[Mori, 2005] Mori, G. (2005). Guiding model search using segmentation. In *Proc. of the IEEE International Conference on Computer Vision*, pages 1417–1423.

[Nabbe et al., 2006] Nabbe, B., Hoiem, D., Efros, A., and Hebert, M. (2006). Opportunistic use of vision to push back the path-planning horizon. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2388–2393.

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

[Owechko and Medasani, 2005] Owechko, Y. and Medasani, S. (2005). A swarm-based volition/attention framework for object recognition. In *Proc. of the IEEE Computer Vision and Pattern Recognition - Workshops*, pages 91–99.

[Rasmussen, 2004] Rasmussen, C. (2004). Texture-based vanishing point voting for road shape estimation. In *Proc. of the British Machine Vision Conference*, pages 470–477.

[Rasmussen et al., 2009] Rasmussen, C., Lu, Y., and Kocamaz, M. (2009). Appearance contrast for fast, robust trail-following. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3505–3512.

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

[Rasmussen and Scott, 2008b] Rasmussen, C. and Scott, D. (2008b). Terrain-based sensor selection for autonomous trail following. In *Proc. of the International Conference on Robot Vision*, pages 341–355.

[Ren and Malik, 2003] Ren, X. and Malik, J. (2003). Learning a classification model for segmentation. In *Proc. of the IEEE International Conference on Computer Vision*, pages 10–17.

[Rosenblatt, 1997] Rosenblatt, J. (1997). DAMN: A distributed architecture for mobile navigation. *Journal of Experimental & Theoretical Artificial Intelligence*, 9(2):339–360.

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

[Ruesch et al., 2008] Ruesch, J., Lopes, M., Bernardino, A., Hornstein, J., Santos-Victor, J., and Pfeifer, R. (2008). Multimodal saliency-based bottom-up attention a framework for the humanoid robot icub. In *Proc. of the IEEE International Conference on Robotics and Automation*, pages 962–967.

[Santana et al., 2010a] Santana, P., Alves, N., Correia, L., and Barata, J. (2010a). A saliency-based approach to boost trail detection. In *Proc. of the IEEE International Conference on Robotics and Automation*.

[Santana et al., 2010b] Santana, P., Alves, N., Correia, L., and Barata, J. (2010b). Swarm-based visual saliency for trail detection. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[Santana et al., 2009] Santana, P., Guedes, M., Correia, L., and Barata, J. (2009). Saliency-based obstacle detection and ground-plane estimation for off-road vehicles. In *Proc. of the International Conference on Computer Vision Systems*, pages 275–284.

[Santana et al., 2010c] Santana, P., Guedes, M., Correia, L., and Barata, J. (2010c). A saliency-based solution for robust off-road obstacle detection. In *Proc. of the IEEE International Conference on Robotics and Automation*, pages 3096–3101.

[Sclaroff and Liu, 2001] Sclaroff, S. and Liu, L. (2001). Deformable Shape Detection and Description via Model-Based Region Grouping. *Proc of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(5):475–489.

[Soquet et al., 2007] Soquet, N., Aubert, D., and Hautiere, N. (2007). Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation. In *Proc. of the IEEE Intelligent Vehicles Symposium*, pages 160–165.

[Southall and Taylor, 2001] Southall, B. and Taylor, C. (2001). Stochastic road shape estimation. In *Proc. of the IEEE International Conference on Computer Vision*, pages 205–212.

[Thrun et al., 2006] Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., et al. (2006). Stanley: The robot that won the DARPA Grand Challenge. *Journal of field Robotics*, 23(9):661–692.

[Tomasi and Shi, 1994] Tomasi, C. and Shi, J. (1994). Good features to track. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600.

[Unser, 1995] Unser, M. (1995). Texture classification and segmentation using wavelet frames. *IEEE Transactions on image processing*, 4(11):1549–1560.

[Urmson et al., 2006] Urmson, C., Ragusa, C., Ray, D., Anhalt, J., Bartz, D., Galatali, T., Gutierrez, A., Johnston, J., Harbaugh, S., ” Yu” Kato, H., et al. (2006). A robust approach to high-speed navigation for unrehearsed desert terrain. *Journal of Field Robotics*, 23(8):467–508.

[Varma and Zisserman, 2005] Varma, M. and Zisserman, A. (2005). A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1):61–81.

[Zhang and Nagel, 1994] Zhang, J. and Nagel, H.-H. (1994). Texture-based segmentation of road images. In *Proc. of the IEEE Intelligent Vehicles Symposium*, pages 260–265.

Appendix A

A Saliency-Based Approach to Boost Trail Detection - Dataset and Image Results

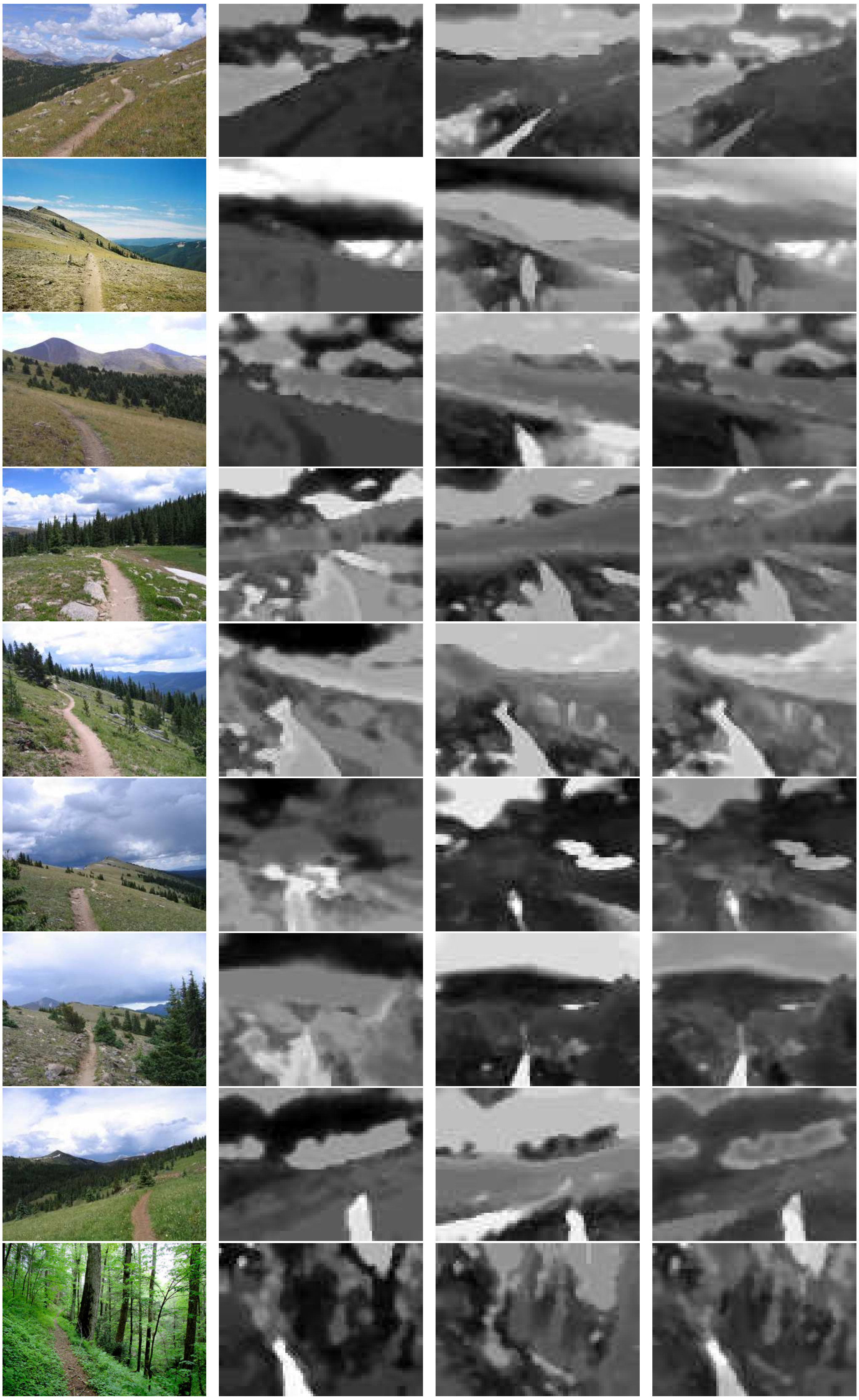


Figure A.1: Dataset used in the saliency-based model, composed by images #01 to #09. The first column presents the input image. The second, third and fourth columns show the colour conspicuity, the intensity conspicuity, and the saliency maps, respectively.

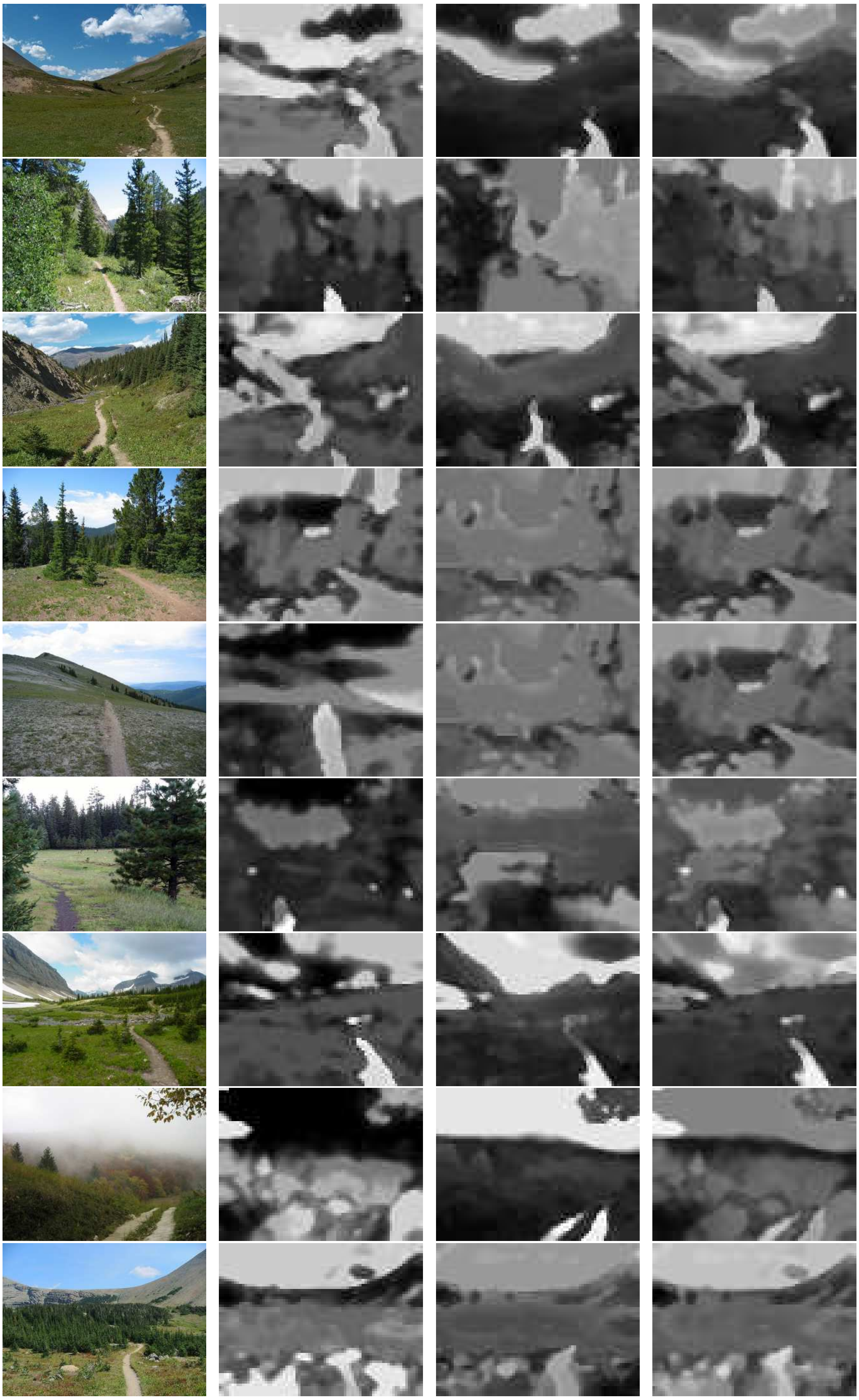


Figure A.2: Dataset used in the saliency-based model, composed by images #10 to #18. The first column presents the input image. The second, third and fourth columns show the colour conspicuity, the intensity conspicuity, and the saliency maps, respectively.

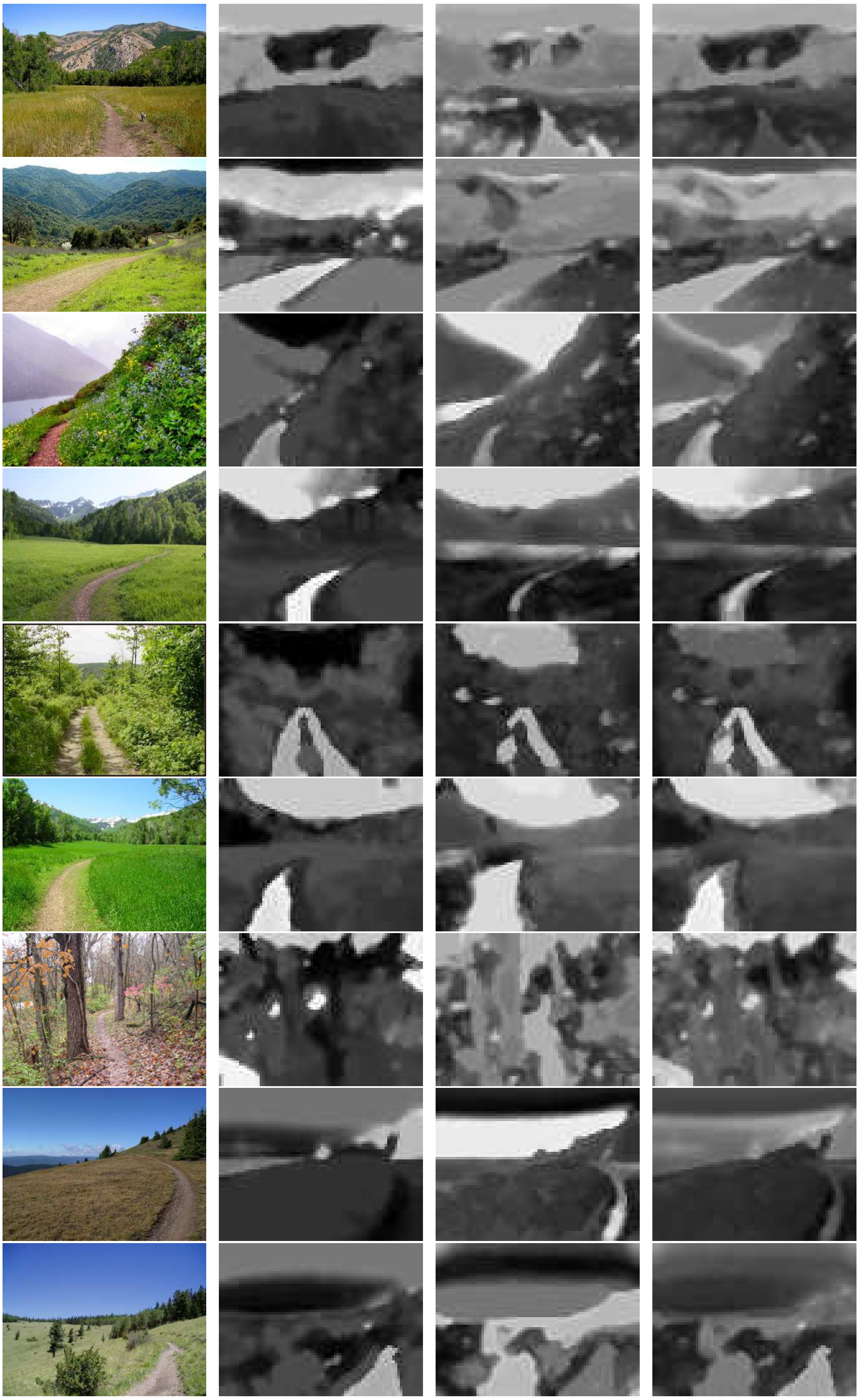


Figure A.3: Dataset used in the saliency-based model, composed by images #19 to #27. The first column presents the input image. The second, third and fourth columns show the colour conspicuity, the intensity conspicuity, and the saliency maps, respectively.

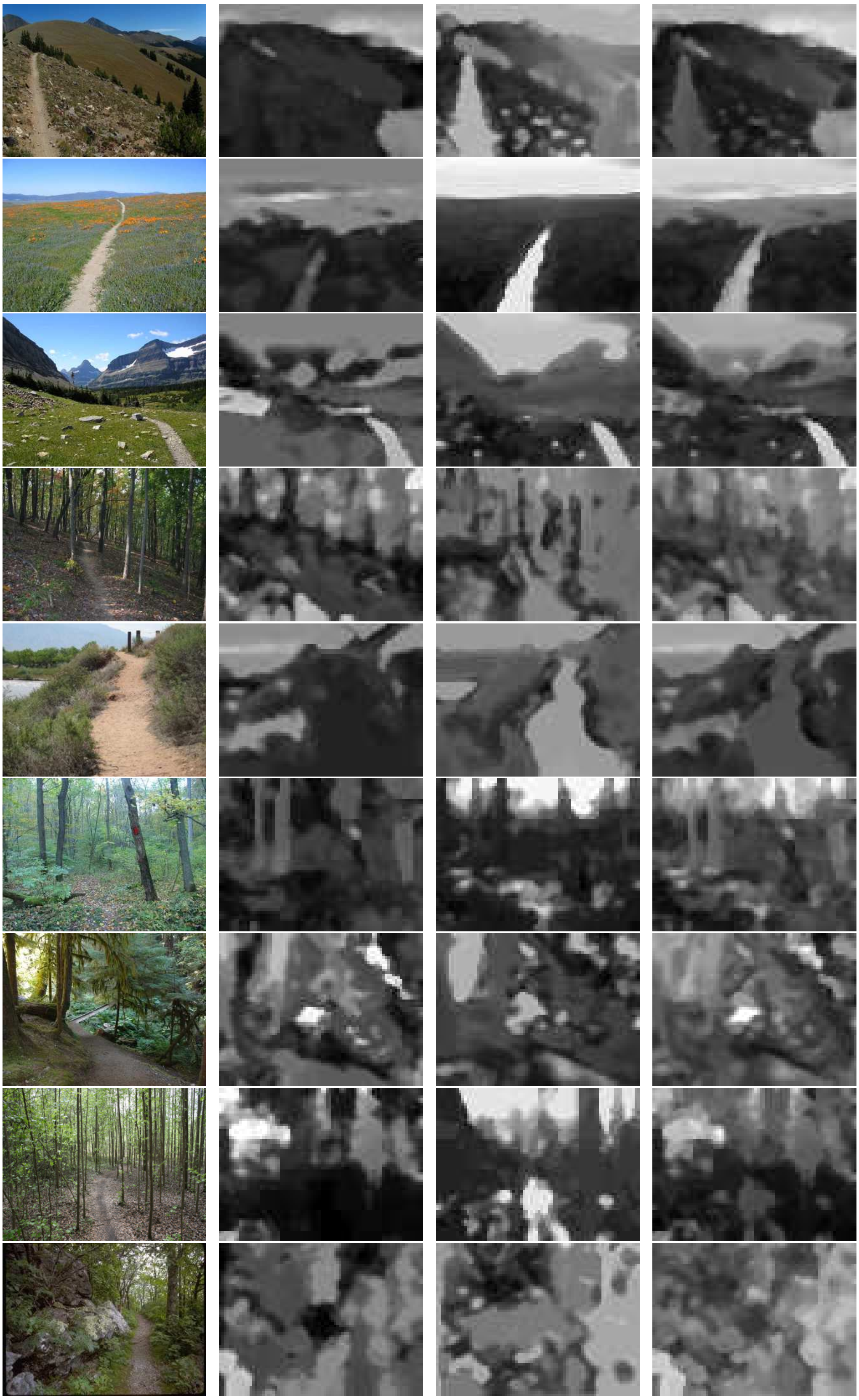


Figure A.4: Dataset used in the saliency-based model, composed by images #28 to #36. The first column presents the input image. The second, third and fourth columns show the colour conspicuity, the intensity conspicuity, and the saliency maps, respectively.

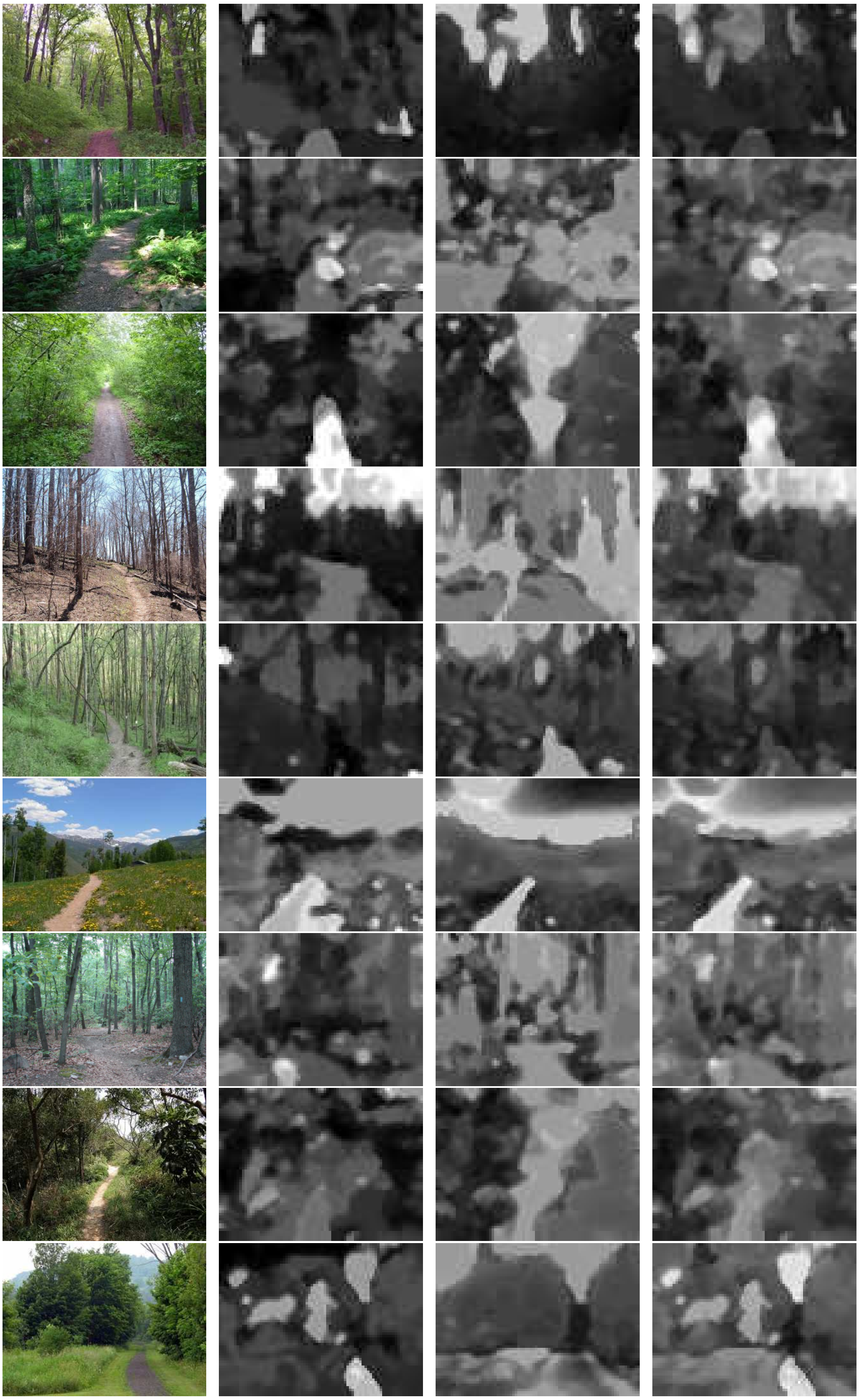


Figure A.5: Dataset used in the saliency-based model, composed by images #37 to #45. The first column presents the input image. The second, third and fourth columns show the colour conspicuity, the intensity conspicuity, and the saliency maps, respectively.

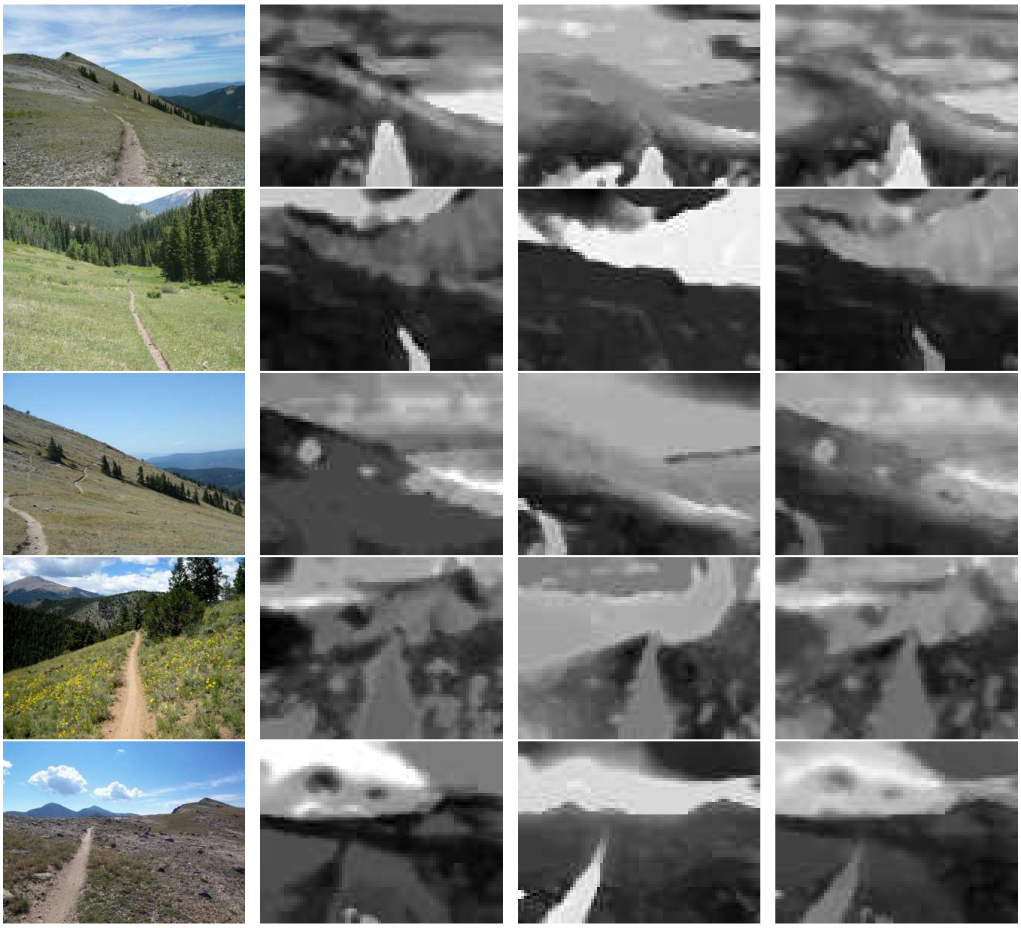


Figure A.6: Dataset used in the saliency-based model, composed by images #46 to #50. The first column presents the input image. The second, third and fourth columns show the colour conspicuity, the intensity conspicuity, and the saliency maps, respectively.

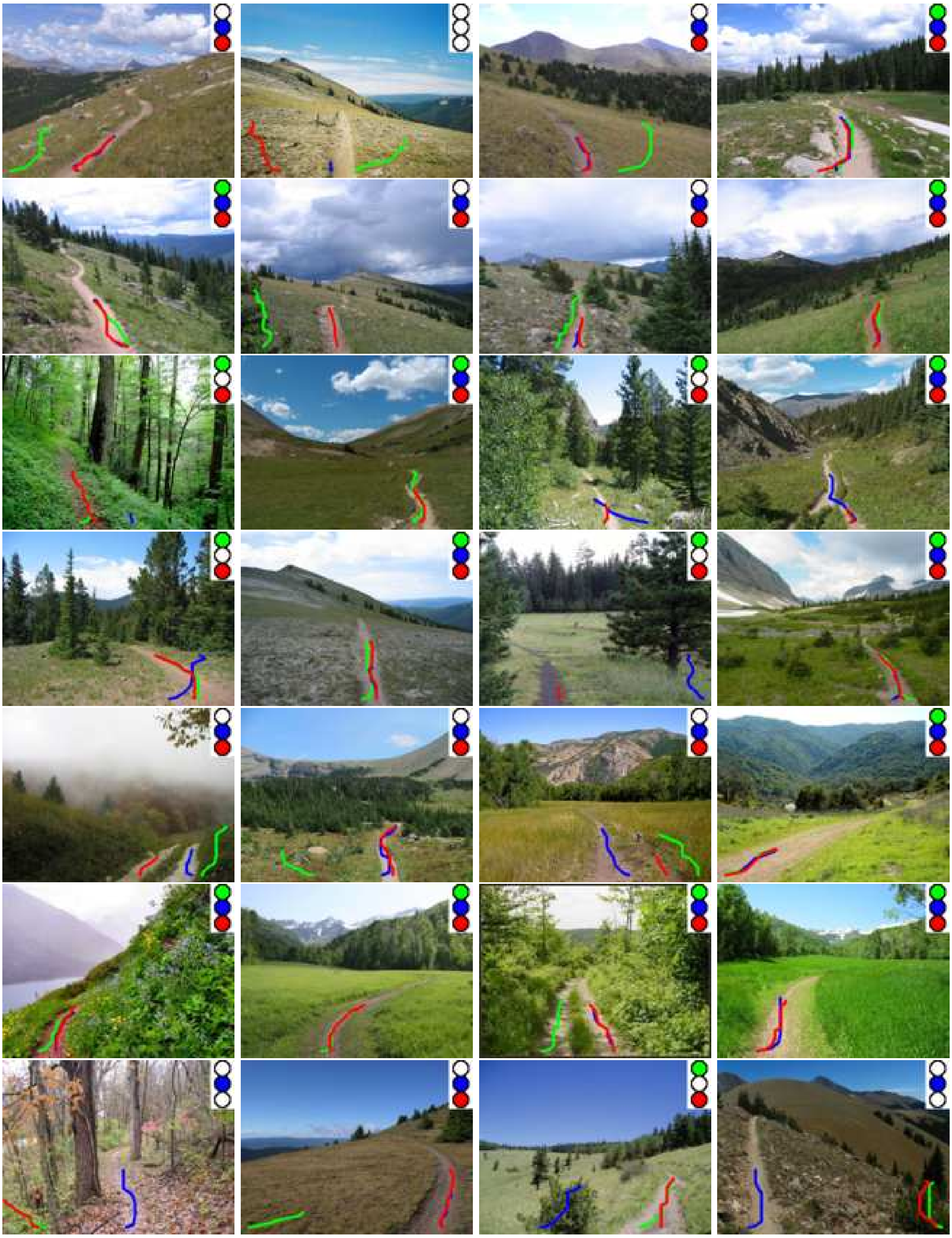


Figure A.7: Trail detection results (#01 to #28). The best agent's path in each map is superposed on the corresponding input image. Path colour is green, blue and red for the colour, intensity and saliency maps, respectively. In the top-right corner of each image, the presence of a filled circle with a given maps' colour, indicates that the best agent's path of the corresponding map correctly represents the trail.

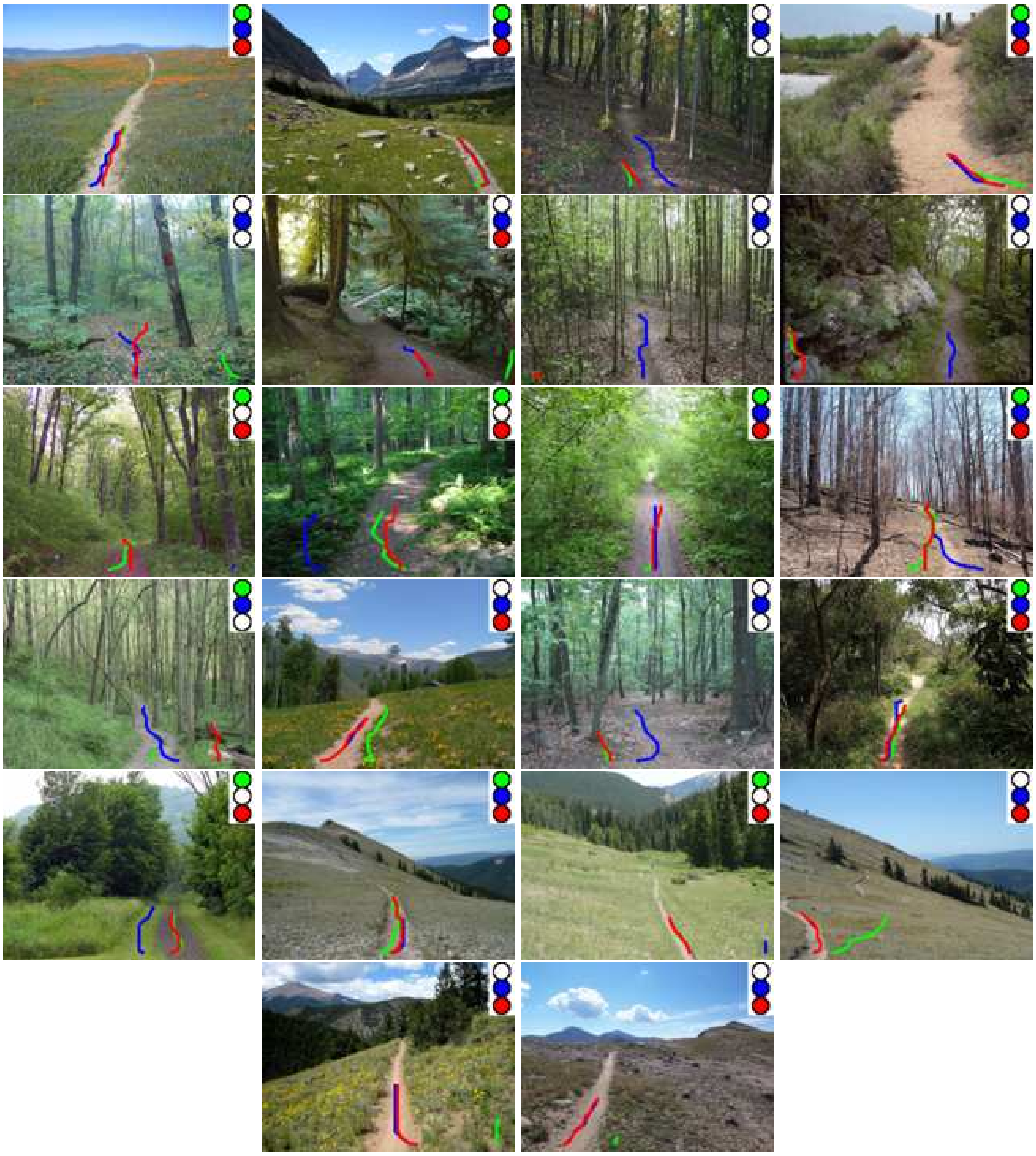
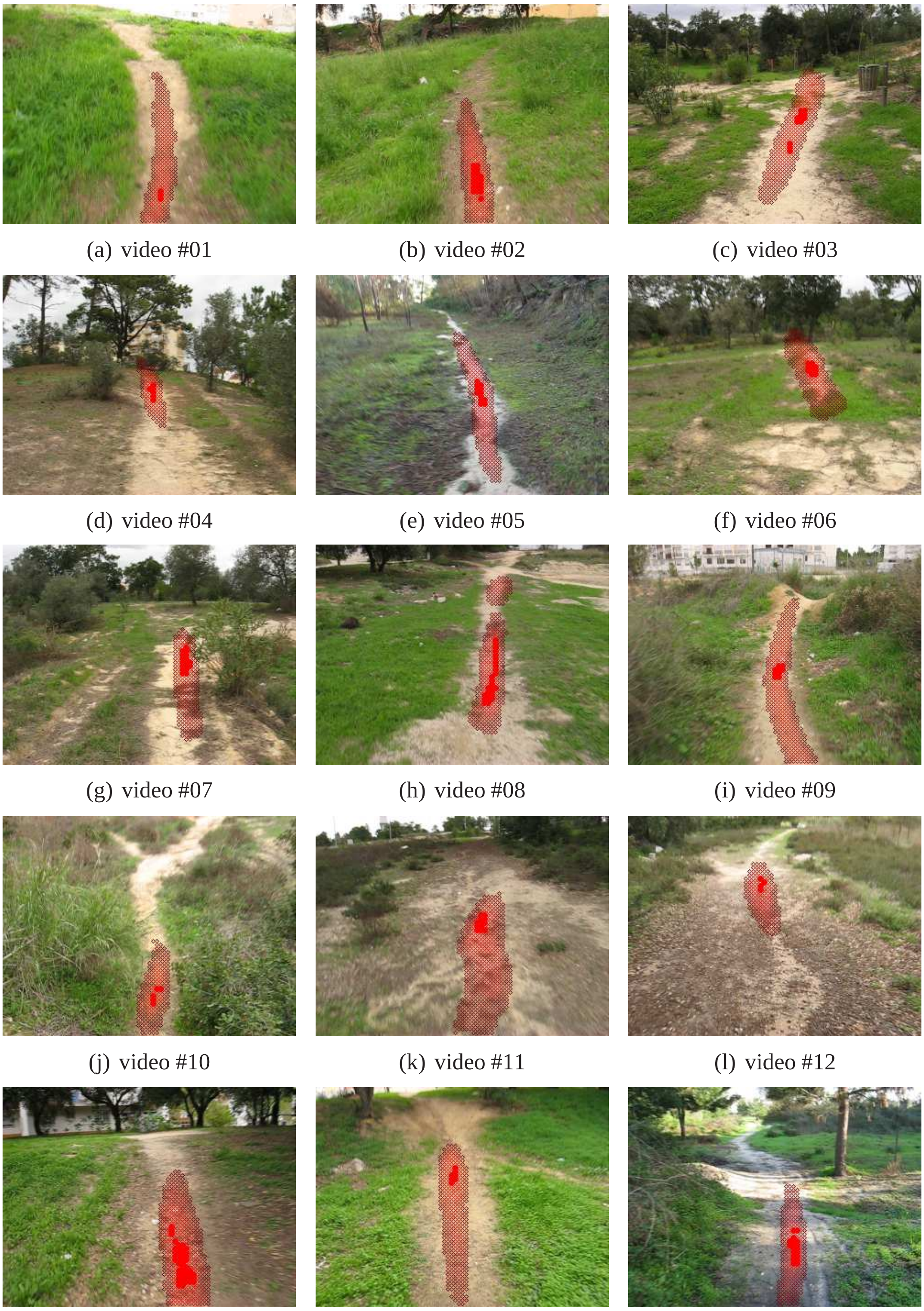


Figure A.8: Trail detection results (#29 to #50). The best agent's path in each map is superposed on the corresponding input image. Path colour is green, blue and red for the colour, intensity and saliency maps, respectively. In the top-right corner of each image, the presence of a filled circle with a given maps' colour, indicates that the best agent's path of the corresponding map correctly represents the trail.

Appendix B

Swarm-Based Visual Saliency for Trail Detection - Dataset and Image Results



(a) video #01 (b) video #02 (c) video #03

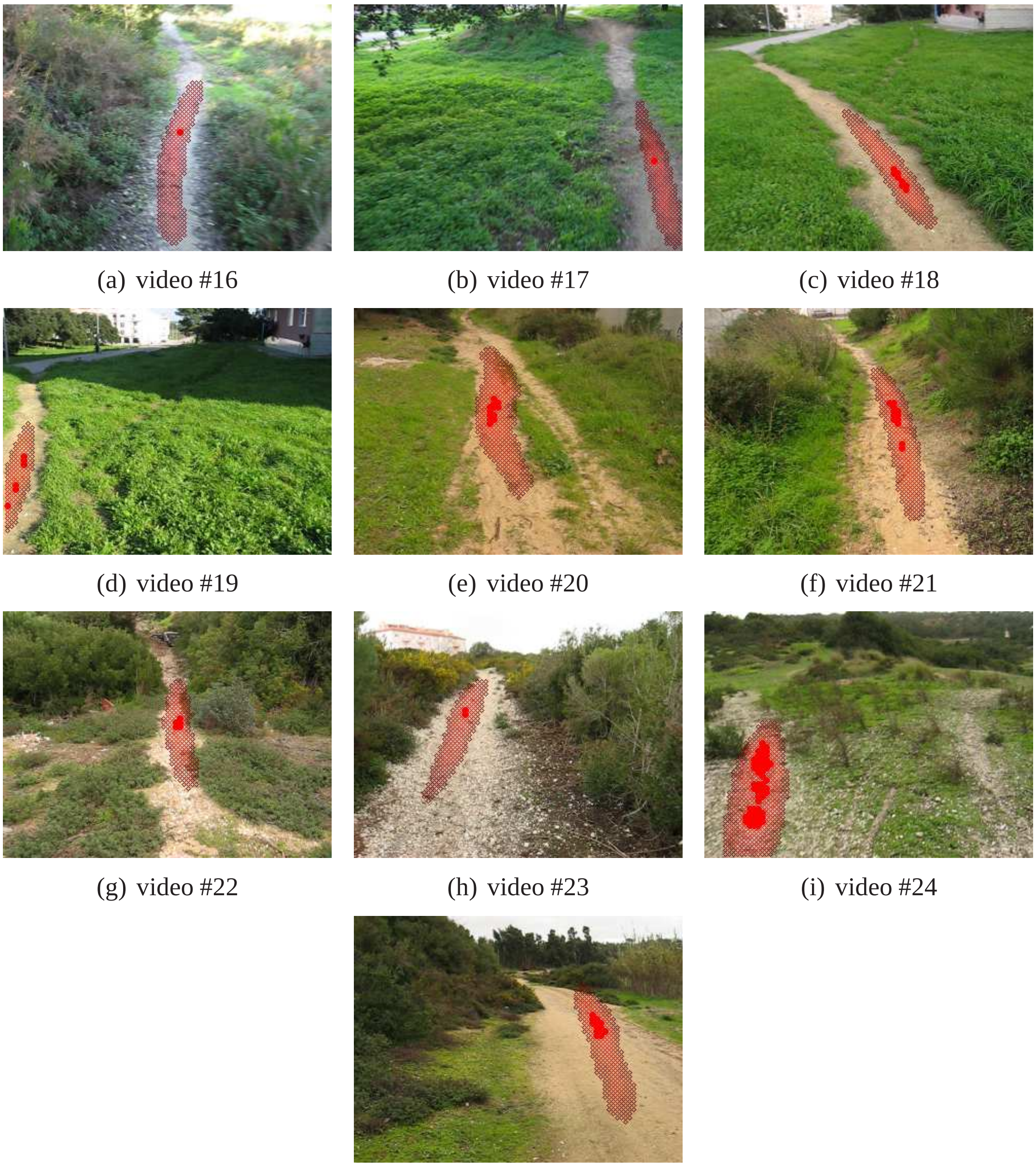
(d) video #04 (e) video #05 (f) video #06

(g) video #07 (h) video #08 (i) video #09

(j) video #10 (k) video #11 (l) video #12

(m) video #13 (n) video #14 (o) video #15

Figure B.1: Dataset used in the swarm model, representative frames from videos #01 to #15. Each image corresponds to one video whose ID is given by increasing order from left to right and top to bottom. The overlaid red blobs represent the model's estimate of the trail location, which corresponds to an activity of the neural field above 0.85.



(a) video #16 (b) video #17 (c) video #18

(d) video #19 (e) video #20 (f) video #21

(g) video #22 (h) video #23 (i) video #24

(j) video #25

Figure B.2: Dataset used in the swarm model, representative frames from videos #16 to #25. Each image corresponds to one video whose ID is given by increasing order from left to right and top to bottom. The overlaid red blobs represent the model's estimate of the trail location, which corresponds to an activity of the neural field above 0.85.