18

Visual Novelty Detection for Mobile Inspection Robots

Hugo Vieira Neto[1] and Ulrich Nehmzow[2]
[1]*Federal University of Technology - Paraná*
[2]*University of Ulster*
[1]*Brazil*
[2]*UK*

1. Introduction

Novelty detection – the ability to identify perceptions that were never experienced before – and, more generally, selective attention are extremely important mechanisms to autonomous mobile robots with limited computational resources. From the operational point of view, the robot's resources can be used more efficiently by selecting those aspects of the surroundings which are relevant to the task in hand or uncommon aspects which deserve closer analysis.

In fact, identification of new concepts is central to any learning process, especially if knowledge is to be acquired incrementally and without supervision. In order to learn concepts, it is necessary to determine first if they are not already part of the current knowledge base of the agent. Simultaneous learning and recognition (Artač et al., 2002) is a fundamental ability to robots aiming at true autonomy, adaptability to new situations and continuous operation.

From the application point of view, reliable novelty detection mechanisms facilitate automated environment inspection and surveillance by highlighting unusual situations. Novelty detection and incremental learning are also vital in applications that demand unsupervised environment exploration and mapping.

In this chapter we are particularly interested in environment inspection using mobile robots to assist in the detection of abnormalities. An example of a practical application is the automatic identification of cracks, tree roots or any other kinds of faults in sewer pipes. Sewer inspection is currently performed manually by human operators watching logged video footage, a time-consuming and error-prone approach, due to human fatigue. This task would benefit immensely from the assistance of an inspection robot that is able to pinpoint just the unusual features in the sewer which are likely to correspond to potential faults.

Fault detection tasks are different from the usual pattern recognition problems in which the features of interest are usually determined beforehand – the aim in fault detection is to detect something that is unknown *a priori*. Therefore, it is argued that the most feasible approach to be followed is to learn a model of *normality* of the environment, and then use it to filter out *any abnormal* sensory perceptions (Tarassenko et. al., 1995). Abnormal perceptions are thus defined as anything that does not fit the acquired model of normality.

Previous work has demonstrated that the approach of learning models of normality and later on using them to highlight abnormalities is very effective for mobile robots that use low-resolution sonar readings as perceptual input (Marsland et al., 2002a).

The sensor modality used as perceptual input obviously plays an important role in the robot's performance for a given task or behaviour. If relevant features of the surroundings can not be properly sensed and discriminated, it will be impossible for the robot to respond appropriately. Mobile robots are usually equipped with tactile, distance (infrared, sonar and laser range finders) and vision sensors. Of all range of sensors, vision is the most versatile modality because it can detect colour, texture, shape, size and even distance to a physical object. Moreover, vision also has the advantage to be able to generate high-resolution readings in two dimensions, making the detection of small details of the environment more feasible.

Vision is therefore a primary source of information for a mobile robot operating in real world scenarios, such as in sewer fault inspection. The main reason for this is that the environment needs to be sensed in high resolution and preferentially using a two-dimensional field of view, so that the chances of missing important details are minimised. Furthermore, vision is a sense shared with humans and therefore provides common ground for collaboration between robots and operators while performing the inspection task.

Much of the previous research done in novelty detection applied to environment inspection using real mobile robots was made using exclusively sonar sensing (Crook et al., 2002; Marsland et al., 2002a) and little work was done using monochrome visual input in very restricted ways (Crook & Hayes, 2001; Marsland et al., 2001). There is also work related to novelty detection using sonar readings in simulated robots (Linåker & Niklasson, 2000; Linåker & Niklasson, 2001). The main idea behind these approaches was to use on-line unsupervised learning mechanisms in order to acquire models of normality for the environment.

Marsland et al. (2002a) have developed the Grow-When-Required (GWR) neural network and used it to highlight novel patterns in sonar scans, while Crook & Hayes (2001) used a novelty detector based on the Hopfield neural network (Hopfield, 1982). These approaches were qualitatively compared in (Crook et al., 2002) during a novelty detection task using sonar readings in a corridor. Linåker & Niklasson (2000) developed the Adaptive Resource Allocating Vector Quantisation (ARAVQ) network and used it in simulations. All of these mechanisms have shown to work very well with low-resolution sonar data according to qualitative assessment criteria. However, none of them was employed using high-resolution visual data in real world application scenarios. Also, quantitative tools to assess and compare the performance of novelty filters objectively were missing. Therefore, qualitative evaluation of novelty filters is one of the issues addressed in this work.

Here we are mainly interested in investigating novelty detection using colour visual input in real robots. However, a major difficulty that comes with vision is how to select which aspects of the visual data are important to be encoded and processed. It is undesirable to process raw high-dimensional visual data directly due to restrictions in computational resources in mobile robots. Hence, a possible solution to cope with massive amounts of visual input (tens of thousands of pixels per image frame) is the use of a mechanism of attention to select aspects of interest and concentrate the available resources on those (Itti & Koch, 2001).

A mechanism of visual attention selects interest points within the input image according to some criteria (for instance, edges or corners). Interest points selected by such attention mechanisms are usually locations containing very descriptive information – they are visually *salient* in the scope of the image frame. A small region in the vicinity of an interest point can then be encoded to represent local visual features. This process not only localises salient features within the image, but also concentrates computational resources where they are necessary. Local encoding of a small image region also has the advantage of reducing data dimensionality while preserving details.

A particularly interesting attention mechanism is the saliency map model (Itti et al., 1998), which combines different visual features (such as intensity, colour and orientation in multiple scales) to obtain a general indication of saliency for each image location – saliency can be thought as the property to stand out from the background. This approach is very convenient for novelty detection and, more specifically, inspection tasks in which the identification of uncommon features is precisely what is desired. Also, the use of a model of visual attention is essential to localise *where* the unusual features are within the image.

The approach we follow in this work is to use the saliency map as mechanism of attention to select a number of salient regions in the input frame, which are encoded into feature vectors and then fed to an unsupervised learning mechanism, either a GWR neural network (Marsland et al., 2002b) or an incremental Principal Component Analysis (PCA) algorithm (Artač et al., 2002). These learning mechanisms are used to build a model of normality for the perceptions acquired in the operating environment and, after the learning process, are used as novelty filters to highlight arbitrary novel features that may be encountered. This approach, as well as some tools for qualitative and quantitative performance assessment of novelty detection systems, is described further in the next section.

2. An Experimental Framework for Visual Novelty Detection

Although novelty detection using sonar sensing proved to be useful to detect open doors in corridors (Marsland et al., 2000) and even to identify in which corridor a mobile robot was operating (Marsland et al., 2002a), the very low resolution used – a robot's sonar ring is typically composed of a small number of sensors – and unreliable noisy readings pose serious limitations to more demanding real world applications. For example, it would be impossible to detect small cracks in a wall by using sonar sensors alone.

Vision, on the other hand, provides detailed information about the operating environment in high resolution. Of course, this comes at the expense of large amounts of data to be processed, which constitutes a serious difficulty when one desires real-time operation. Fortunately, the massive amount of information provided by a vision sensor is highly redundant and therefore can be compressed prior to higher levels of processing. Selecting which aspects of the visual data are the most relevant, however, is not a straightforward procedure and usually is dependant on the application.

Visual novelty depends on the multi-modal measures of the properties from the environment that the camera provides the robot with – some visual feature can be considered novel because of its colour, texture, shape, size, pose, motion or any combination of these and even other visual features – a much more complex scenario than the one of single mode sensors like sonars. Because multi-modal vision is very difficult to be accomplished in a mobile robot with limited computational resources, we had to decide which visual features were the most important to define novelty in our application domain.

In the context of an environment such as the inside of a sewer pipe, the visual novelties we are interested in are static, i.e. they do not move. Nevertheless, the sewer is a dynamic environment in the sense that new visual features that correspond to faults – cracks and tree roots, for instance – may appear at any time, hence the need of regular inspections. For this type of application, higher level visual interpretations – such as the concepts of size, pose or motion – are not as important as low-level features that characterise the essential appearance of visual objects. Therefore, we limited the visual features of interest in this work to colour, texture and shape.

Besides characterising novelty by visual appearance, spatial location of novel visual features in the environment is also important. Therefore, we are interested not only in detecting *which* features constitute potential faults but also *where* they are in the environment. Changes in location of visual features may also constitute novelty and are relevant in some application domains, such as automated surveillance. However, in the scope of this work we will not consider the location of a visual feature to be contextually important to determine novelty. In other words, it will not be possible to consider some visual feature as being novel based solely on its location in the environment.

Another difficulty related to visual novelty detection using a mobile robot concerns invariance to image transformations. Because the images are acquired from a moving platform, visual features are subject to several geometric transformations and it is undesirable that known visual features happen to be labelled as novel just because there were changes in appearance due to robot movement (e.g. changes in perspective). Hence, the image encoding procedure should be robust to small geometrical transformations that result from robot motion.

In order to localise novel features within the image frame, we decided to get inspiration from biological vision systems and use a mechanism of attention (Itti & Koch, 2001). Following this idea, smaller image regions selected by the visual attention mechanism from the input image can be encoded as feature vectors. Figure 1 depicts the block diagram of such an approach, in which the novelty filter is preceded by the attention mechanism. Instead of encoding the whole image frame in a single feature vector, several feature vectors are encoded per frame using the vicinity of salient image locations. Salient (or interest) points normally correspond to places with high information contents, i.e. strong peaks, edges or corners, depending on the criteria for their selection. In this work, we have used the saliency map (Itti et al., 1998) as our mechanism of selective attention.

By selecting interest regions in the form of image patches, we reduce the dimensionality of the data to be processed by the novelty filter and also gain robustness to some geometrical transformations, notably translations within the image frame due to robot movement. Furthermore, novel visual features can be immediately localised within the input image frame with the selection of local salient regions, as we will demonstrate experimentally.

The experiments that follow use raw image patches (24×24 pixels in size) extracted from salient locations within the input image (160×120 pixels in size) and compare performances of novelty filters based on the GWR neural network (Marsland et al., 2002b) and incremental PCA (Artač et al., 2002). There is hardly any other visual representation more specific than raw image patches and therefore generalisation in the experiments reported here is left to the learning mechanisms used as novelty filters. As a side-effect, the use of raw image patches allows visual feedback of the knowledge acquired during training (Vieira Neto & Nehmzow, 2005; Vieira Neto & Nehmzow, 2007).

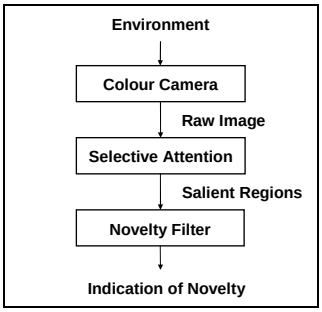


Figure 1. The framework for the investigation of visual novelty detection: an attention mechanism selects patches from the input image frame, which are then classified by the novelty filter as novel or non-novel.

2.1 Experimental setup

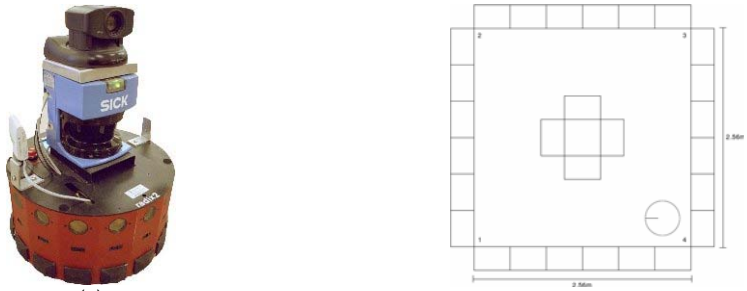
To evaluate the ability of our novelty detection framework to detect novel visual features that may appear in the robot's normal operating environment, we conducted experiments in controlled scenarios. Every experiment consisted of two stages: an exploration (learning) phase, in which the robot was used to acquire a model of normality of the environment, followed by an inspection (application) phase, in which the acquired model was then used to highlight any abnormal perception in the environment.

During the learning phase, images were acquired while the robot was navigating around a "baseline" environment (an empty arena or corridor, containing only "normal" features). These images were then processed by the attention mechanism to generate input feature vectors and train the novelty filter. After that, during the application phase, novel objects were placed in the environment so that a new sequence of images could be acquired and used to test the trained novelty filter.

The expected outcome of these experiments was that the amount of novelty detected would continuously decrease during exploration as a result of learning. At the beginning of the learning procedure everything is considered novel and, as the robot learns, less and less novelties should be found. During the inspection phase we expected that peaks in the novelty measure would appear only in areas where a new object had been placed. This hypothesis was tested using a real robot navigating in engineered (laboratory) and medium-scale real world environments. Figure 2 shows the experimental setup used for the laboratory experiments.

The colour vision system of *Radix*, the Magellan Pro robot shown in Figure 2a, was used to generate visual stimuli while navigating in the environment. The robot was equipped with standard sonar, infra-red and tactile sensors, and also with an additional laser range scanner whose readings were used for controlling the navigation behaviour. *Radix* operated completely autonomously in our experiments.

The robot's on-board computer was capable of processing on-line up to eight frames per second when running our control software, which was optimised for speed. Nevertheless, the images used in the experiments reported in this chapter were acquired at one frame per second (without stopping the robot) for off-line processing. This procedure was chosen in order to allow fair performance comparisons between different novelty detection mechanisms by using the same datasets.



(a) (b)

Figure 2. Experimental setup: (a) Magellan Pro mobile robot; (b) top view of a typical robot arena used as operating environment used in laboratory experiments. The robot arena was delimited by cardboard boxes, represented by rectangles, while the robot is represented by a circle with a line indicating its front, to where its camera is directed.

Figure 2b shows the top view of the engineered environment used in the laboratory experiments, a square arena delimited by cardboard boxes in whose corners (numbered from 1 to 4) novel objects were introduced. The cardboard boxes at the borders of the arena acted as walls (approximately 0.5m high) that limited the robot's trajectory and also its visual world. The images were acquired with the robot's camera tilted down to −25 degrees, so that the field of view was constrained to the arena's walls and floor, resulting in a completely controlled visual world for our experiments. A simple obstacle-avoidance algorithm using the robot's laser range scanner measurements was used as the navigation behaviour for the robot. In our experiments, this behaviour has shown to be very predictable and stable.

2.2 Assessment of results

Qualitative and quantitative assessment tools were devised to analyse our results. These assessment tools are very important to establish a reference for comparisons and therefore determine which of the studied methods perform better according to the desired application.

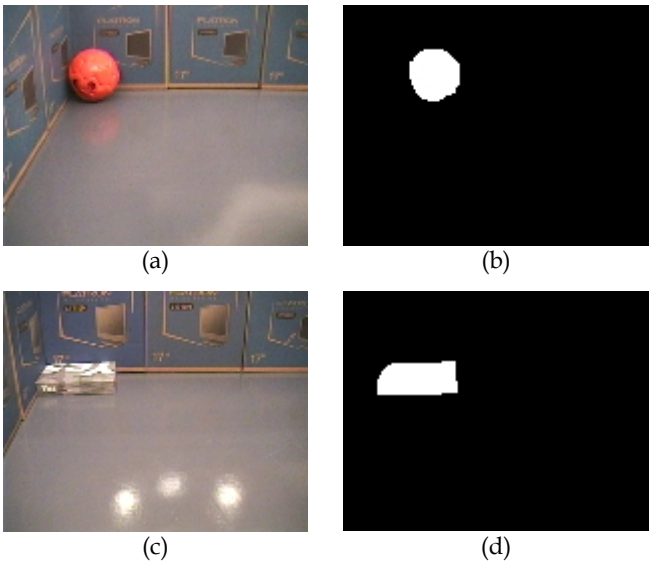
Qualitative assessment. In the following sections we use bar graphs in which novelty measurements provided by the novelty filter are plotted against time. They are used in order to obtain a qualitative indication of performance, in a similar fashion to (Marsland et al., 2002a). In these graphs, time essentially corresponds to a certain position and orientation of the robot in the environment because the navigation behaviour used in the experiments was highly repeatable.

The efficiency of learning during the exploration phase can be graphically assessed through inspection of the qualitative novelty graphs in multiple rounds. By looking at the novelty graphs for the exploration phase, one can determine how fast learning occurred and also assess if the amount of learning was adequate for the acquisition of an environmental model of normality.

In the inspection phase, a new object was introduced in the normal environment in order to test the system's ability to highlight abnormal perceptions. The measure of novelty was expected to be high only in places where the new object could be sensed, an expectation that should be reflected in the novelty graphs obtained. The inspection phase of experiments was

also carried out in multiple rounds with the learning mechanism disabled, so that unusual features in the environment were highlighted every time that they were perceived. Hence, the consistency of a novel feature being detected in a particular location of the environment but in different inspection rounds can also be evaluated using this qualitative assessment scheme.

Quantitative assessment. The off-line processing of image frames acquired with the robot in exploration and inspection phases also allows a quantitative assessment and direct performance comparison between different approaches through the use of identical datasets. For that, we manually generated ground truth in the form of a binary image for each input image where the novel object was present. In these binary images, the pixels corresponding to the novel object were highlighted (see examples in Figure 3).



(a) (b)

(c) (d)

Figure 3. Example of typical input images containing a novel object: an orange football in (a) and a grey box in (c), and their corresponding ground truth novelty templates in (b) and (d), respectively.

Using the manually generated ground truth information, contingency tables were built relating system response to actual novelty status, as shown in Table 1. If a given region of the input image had more than 10% of highlighted pixels in the corresponding ground truth template, then this region's novelty status was considered as "novelty present". In this case, if the system response was "novelty detected" this configured true novelty and therefore entry *A* in the contingency table shown in Table 1 was incremented; otherwise, if the system response was "novelty not detected", this configured a missed novelty with entry *B* being incremented. On the other hand, if the novelty status of a given image region was considered as "novelty not present" (less than 10% of highlighted pixels in the corresponding ground truth region) and nevertheless the system responded as "novelty detected", this configured a false novelty with entry *C* being incremented. Finally, if the system response agreed with the actual novelty status by attributing "novelty not detected"

to a region whose novelty status was "novelty not present", this represented true non-novelty and entry *D* was incremented.

	Novelty Detected	Novelty Not Detected
Novelty Present	*A*	*B*
Novelty Not Present	*C*	*D*

Table 1. Contingency table for the quantitative assessment of novelty filters.

An ideal association between system response and actual novelty status would have a contingency table in which values *B* and *C* in Table 1 are zero, while values *A* and *D* have non-zero values (in practice, *A* will be small in comparison to *D* as usually there are few examples of novel features in the inspected environment). The statistical significance of the association between the actual novelty status (ground truth) and the novelty filter response can be tested using χ^2 analysis (Nehmzow, 2003; Sachs, 2004). For the 2×2 contingency table shown in Table 1, the χ^2 statistic is computed using:

$$\chi^2 = \frac{N(AD - BC)^2}{(A+C)(C+D)(A+B)(B+D)}, \tag{1}$$

where $N = A + B + C + D$ is the total number of samples in the table.

If $\chi^2 > 3.84$ there is a significant correlation between novelty status and novelty filter response, with a probability $p \leq 0.05$ of this statement being wrong. If $\chi^2 > 6.64$ the significance level of the correlation is higher and the probability of being wrong decreases to $p \leq 0.01$. It is also important to mention that the χ^2 test is valid only for well-conditioned contingency tables – this entails the computation of a table of expected values, which must have no entries with expected values below 5 (Nehmzow, 2003).

The strength of the association was assessed by Cramer's *V*, which is directly based on the χ^2 statistic (Nehmzow, 2003):

$$V = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{(AD - BC)^2}{(A+C)(C+D)(A+B)(B+D)}}. \tag{2}$$

The uncertainty coefficient *U*, an entropy-based measure, was also used to estimate the strength of the association. Computation of the uncertainty coefficient relies on the fact that each sample in the contingency table shown in Table 1 has two attributes, the actual novelty status *S* and the novelty filter response *R*. The entropy of *S*, *H*(*S*), the entropy of *R*, *H*(*R*), and the mutual entropy of *S* and *R*, *H*(*S*, *R*), are given by the following equations:

$$H(S) = -\frac{A+B}{N}\ln\left(\frac{A+B}{N}\right) - \frac{C+D}{N}\ln\left(\frac{C+D}{N}\right), \tag{3}$$

$$H(R) = -\frac{A+C}{N}\ln\left(\frac{A+C}{N}\right) - \frac{B+D}{N}\ln\left(\frac{B+D}{N}\right), \tag{4}$$

$$H(S,R) = -\frac{A}{N}\ln\left(\frac{A}{N}\right) - \frac{B}{N}\ln\left(\frac{B}{N}\right) - \frac{C}{N}\ln\left(\frac{C}{N}\right) - \frac{D}{N}\ln\left(\frac{D}{N}\right). \quad (5)$$

When applying equations 3, 4 and 5, one must remember that $\lim_{p\to 0} p\ln p = 0$. The uncertainty coefficient U of S given R, $U(S|R)$, is finally computed using (Nehmzow, 2003):

$$U(S|R) = \frac{H(S) - H(S,R) + H(R)}{H(S)}. \quad (6)$$

Both V and U provide normalised measures of strength ranging from zero to one. Good associations result in V and U having values close to one, while poor associations result in values close to zero. Therefore, the values of V and U can be used to determine which among two or more novelty systems perform better in a given situation.
A further statistic that was used is the κ index of agreement, which is computed for 2×2 contingency tables as follows (Sachs, 2004):

$$\kappa = \frac{2(AD - BC)}{(A+C)(C+D) + (A+B)(B+D)}. \quad (7)$$

This statistic is used to assess the agreement between ground truth data and novelty filter response, in a similar way to what is done with V and U. However, it has the advantage of having an established semantic meaning associated with some intervals, as shown in Table 2 (Sachs, 2004).

Interval	Level of Agreement
$\kappa \leq 0.10$	No
$0.10 < \kappa \leq 0.40$	Weak
$0.40 < \kappa \leq 0.60$	Clear
$0.60 < \kappa \leq 0.80$	Strong
$0.80 < \kappa \leq 1.00$	Almost complete

Table 2. κ intervals and corresponding levels of agreement between ground truth and novelty filter response.

Unlike V and U, the κ statistic may yield negative values. If this happens, the level of *disagreement* between system response and manually generated ground truth can be assessed. Negative values result when the entries B and C in the resulting contingency table are larger than the entries A and D. In such a case, both U and V would still result in positive values because they are designed to measure the strength of the association (be it positive or negative) rather than the level of agreement (positive association) or disagreement (negative association).

3. Experiments in a Laboratory Environment

3.1 The GWR neural network as novelty filter

The GWR network (Marsland et al., 2002a; Marsland et al., 2002b) is a self-organising neural network based on the same principles as Kohonen's Self-Organising Map (Kohonen, 1984). Its structure is composed of nodes that represent the centres of clusters (model weight

vectors) in input space – every time that an input is presented, each network node will respond with higher or lower activity depending on how good its weight vector matches the input vector.

A novelty filter based on the GWR network basically consists of a clustering layer of nodes and a single output node. The connecting synapses to the output layer are subject to a model of habituation, which is a reduction in behavioural response to inputs that are repeatedly presented. In other words, the more a node in the clustering layer fires, the less efficient its output synapse becomes.

What makes the GWR network superior to the SOM is its ability to add nodes to its structure – hence the name Grow-When-Required – by identifying new input stimuli through the habituation model. Given an input vector, both the firing node's activity and habituation are used to determine if a new node should be allocated in order to represent the input space better.

The habituation rule of a clustering node's output synaptic efficacy is given by the following first-order differential equation (Marsland et al., 2002b):

$$\tau \frac{dh(t)}{dt} = \alpha[h_0 - h(t)] - S(t) \; . \tag{8}$$

where h_0 is the initial value of the efficacy $h(t)$, $S(t)$ is the external stimulus, τ and α are time constants that control the habituation rate and the recovery rate, respectively.

$S(t) = 1$ causes habituation (reduction in efficacy) and $S(t) = 0$ causes dishabituation (recovery of efficacy). It is important to mention that only habituation was modelled in our implementation – dishabituation was disabled by setting $S(t) = 1$. Using $\alpha = 1.05$ and $h_0 = 1$ results in efficacy values ranging from approximately 0.05 (meaning complete habituation) to 1 (meaning complete dishabituation). As synaptic efficacy has a bounded output, it can be used neatly as a measure of the degree of novelty for any particular input: higher efficacy values correspond to higher degrees of novelty. More detail is given in (Vieira Neto & Nehmzow, 2005).

Normality model acquisition. Exploration was conducted in five consecutive loops around the empty arena, with the robot being stopped and repositioned at the starting point in every loop. This procedure was used in order to ensure that the robot's trajectory would be as similar as possible for every loop, resulting in consistent novelty graphs for qualitative assessment. Images were acquired at the rate of one frame per second, resulting in a total of 50 images per loop around the arena. We used normalised raw image patches, selected by the saliency map from the acquired images, as input to the GWR network.

During the exploration phase, learning of the GWR network was enabled to allow the acquisition of a model of normality. As expected, the amount of novelty measured – the efficacy of the habituable synapse of the firing node – decreased as the network habituated on repeated stimuli. Only four nodes were acquired by the GWR network by the end of the fifth exploration loop.

Novelty detection. Having trained the GWR network during the exploration phase, we then used the acquired model to highlight any unusual visual features introduced in the empty arena: an orange football was placed as novel object in one of the corners and the robot was used to inspect the arena. The ball was selected not only because it contrasted well with the

arena's colour features, but also because it did not interfere with the robot's trajectory around the arena (it could not be sensed by the laser range scanner).
Learning of the GWR network was disabled during inspection, so that consistency in novelty indications could be verified over different loops around the arena. The novelty graphs obtained for the inspection phase of the arena containing the ball are shown in Figure 4.

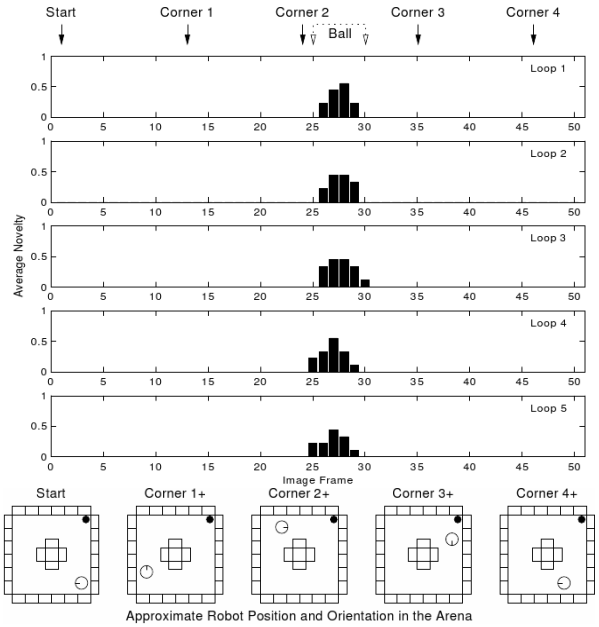


Figure 4. Inspection of the arena with the orange ball (novel stimulus) using the GWR network. The orange ball is clearly and consistently highlighted as novelty. The pictograms below the graphs indicate the approximate position and orientation of the robot in the arena while performing the inspection loops and also the location of the novel stimulus.

Because the images were acquired at the rate of one frame per second, the horizontal axis of the graphs in Figure 4 can also be interpreted as time in seconds. Pictograms indicating the approximate position and orientation of the robot in the arena are also shown (we use the notation "Corner 1+" to indicate position and orientation immediately after the robot has completely turned the first corner).
The set of frames where the orange football appeared in the camera's field of view are indicated by dotted arrows on the top of Figure 4. These frames correspond to locations where high values for the novelty measure were expected to happen (the ball appeared always in the same frames in every loop because the navigation behaviour was very stable). As one can notice in Figure 4, the ball was always correctly detected as the novel feature in the environment (see also figure 8a for the visual output of the system). Contingency table analysis through the χ^2 test revealed statistical significance between the novelty filter

response and actual novelty status ($p \leq 0.01$). The strength of the association revealed almost complete agreement between system response and actual novelty status.
A new inspection phase was then performed in the arena with another novel stimulus, a grey box instead of the orange ball. The grey box is much less conspicuous than the orange ball and the idea behind its use was to check the system's ability to detect a novel object similar in colour to the environmental background and therefore not very salient (the arena had predominantly grey and dark blue colours).
The frames in which the grey box appeared in the camera's field of view are indicated with dotted arrows in Figure 5, where the results of the new inspection round are also given.

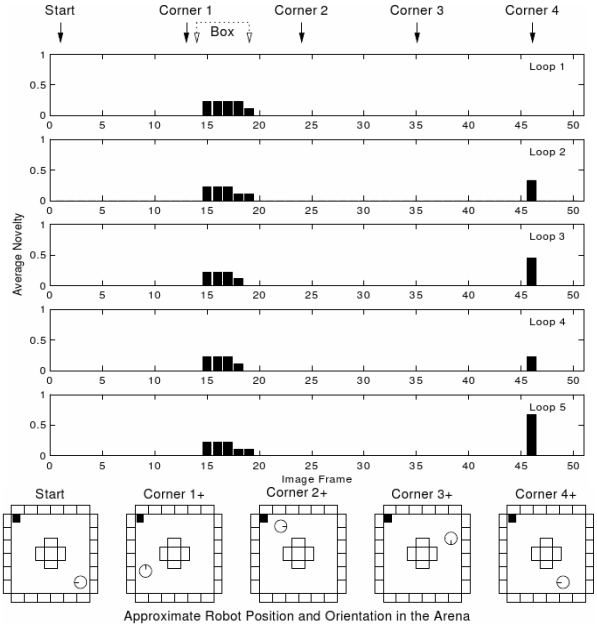


Figure 5. Inspection of the arena with the grey box (novel stimulus) using the GWR network. The grey box is clearly and consistently highlighted as novelty, but unexpected novelty indications also appeared consistently for image frame 46. The pictograms below the graphs indicate the approximate position and orientation of the robot in the arena while performing the inspection loops and also the location of the novel stimulus.

The GWR network correctly identified the grey box as novel, as shown in Figure 5 (see also figure 8b for the visual output of the system). However, unexpected novelty peaks also appeared consistently for image frame 46, when the robot turned a corner very close to the arena's wall. In this particular image frame, two thirds of the most salient regions correspond to a large edge between two of the cardboard boxes that constitute the wall. Although the robot was exposed before to edges between the cardboard boxes, it had never before been as close as happened in this case – this resulted in image patches containing

edges larger in scale than the GWR network was habituated to, resulting in their classification as novel.

Nevertheless, contingency table analysis using the χ^2 test revealed statistical significance between system response and ground truth ($p \leq 0.01$). The strength of the association was also measured and revealing strong agreement between novelty filter response and actual novelty status. As one can notice from these results, the false novelty indications due to the features present in image frame 46 depressed results slightly, but not statistically significantly.

3.2 Incremental PCA as novelty filter

PCA is a very useful tool for dimensionality reduction that allows optimal reconstruction of the original data, i.e. the squared reconstruction error is minimised. It consists of projecting the input data onto its principal axes – the axes along which variance is maximised – and is usually computed off-line because the standard algorithm requires that all data samples are available a priori, making it unsuitable for applications that demand on-line learning.

A method for the incremental computation of PCA recently introduced by Artač et al. (2002) makes simultaneous learning and recognition possible, which is an improvement to the algorithm originally proposed by Hall et al. (1998). Using this method, it is possible to discard the original input data immediately after the eigenspace is updated, storing only projected data with reduced dimensions.

We use incremental PCA as an alternative method to the GWR network to perform on-line novelty detection. The magnitude of the residual vector – the RMS error between the original input and the reconstruction of its projection onto the current eigenspace – is used to decide if a given input is novel and therefore should be added to the model. If the magnitude of the residual vector is above some threshold r_T, the corresponding input vector is not well represented by the current model and therefore must be a novel input. Complete implementation details of the incremental PCA algorithm used in this work can be found in (Vieira Neto & Nehmzow, 2005).

The previous laboratory experiments were repeated using the incremental PCA approach for comparison purposes. During the exploration phase it was verified that most of the eigenspace updates happened in the beginning of the first loop around the arena, becoming less frequent as the environment was explored. By the end of the fifth exploration loop, the incremental PCA algorithm acquired 35 model vectors.

As before, the model learnt during the exploration phase was used to highlight novel visual features in the arena during the inspection phase, while the learning mechanism was disabled. The results obtained for the inspection of the arena with the orange ball are given in Figure 6.

The orange ball was correctly identified as novel, as can be seen in Figure 6. Also, there were very few false indications of novelty. Inspection was repeated for the arena containing the grey box and the results obtained are given in Figure 7.

The grey box was also correctly highlighted by the incremental PCA approach with only a few spurious novelty indications. Incremental PCA coped better with the robot getting closer to the arena's walls (e.g. the large scale edge present in frame 46) because of our choice of parameters which influence generalisation – but this does not necessarily mean that incremental PCA always generalises better than the GWR network.

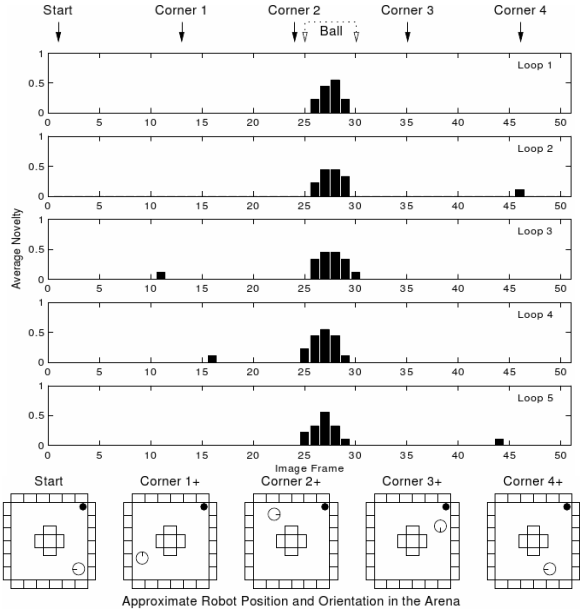


Figure 6. Inspection of the arena with the orange ball (novel stimulus) using incremental PCA. The orange ball is clearly and consistently highlighted with very few unexpected novelty indications.

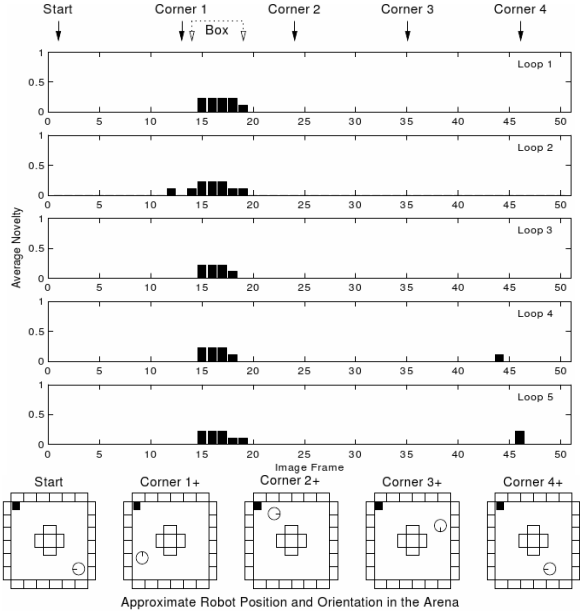


Figure 7. Inspection of the arena with the grey box (novel stimulus) using incremental PCA. The grey box is clearly and consistently highlighted with very few unexpected novelty indications.

3.3 Results

Table 3 shows a quantitative comparison of the results obtained with the GWR network and the incremental PCA approach. All cases presented statistically significant correlation between novelty filter response and actual novelty status according to the χ^2 analysis ($p \leq 0.01$).

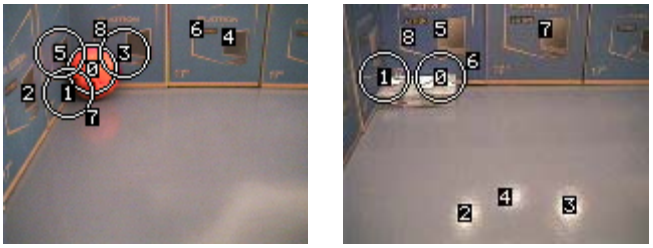
Overall performances (combined performances for the orange ball and the grey box) of both approaches are quantitatively very similar (almost complete agreement between novelty filter response and ground truth), although the incremental PCA algorithm yielded slightly better and more consistent overall results.

	GWR network	Incremental PCA
Orange ball	$V = 0.91$ $U = 0.74$ $\kappa = 0.91$	$V = 0.86$ $U = 0.68$ $\kappa = 0.86$
Grey box	$V = 0.70$ $U = 0.44$ $\kappa = 0.70$	$V = 0.83$ $U = 0.60$ $\kappa = 0.83$
Overall	$V = 0.82$ $U = 0.58$ $\kappa = 0.82$	$V = 0.85$ $U = 0.64$ $\kappa = 0.85$

Table 3. Performance comparison for the laboratory experiments: all results correspond to statistically significant correlation between system response and actual novelty status (χ^2 analysis, $p \leq 0.01$).

Figure 8 gives examples of output images in which the novel object was present in the field of view of the robot's camera. In these output images the numbers indicate the location of the interest points identified by the saliency map in order of relevance (0 corresponding to the most salient), while the presence of white circles indicate that the surrounding region was classified as novel by the filter.

One can notice in Figure 8 that in both cases, the system was able to indicate *where* the regions containing part of the novel object were within the image frame. For the particular output images shown, the results yielded by both novelty filters – GWR network and incremental PCA – were identical.



(a) (b)

Figure 8. Examples of output images for the laboratory experiments: (a) orange ball (GWR and PCA); (b) grey box (GWR and PCA). The numbers indicate the location of salient points

in order of importance (0 corresponds to the most salient) and the white circles indicate that the region corresponding to a particular salient point was labelled as novel.

4. Experiments in a Real World Environment

After the successful results obtained in experiments conducted in a laboratory environment, it was time to test the proposed visual novelty detection approach in a medium-scale real world environment.

The ideal scenario would be to send the robot down a sewer pipe to inspect for cracks, tree roots and other types of faults. However, our research robot was too large and also was not suitable to operate in such an environment. Furthermore, rigorous analysis and assessment of the system's behaviour in this kind of situation would be very difficult to perform due to the lack of knowledge and control of environmental characteristics – construction of the novelty ground truth, for instance, would constitute a difficult task.

Experimental setup. Hence, we decided to conduct experiments in one of the corridors at the Network Centre building at the University of Essex. The robot navigated along the corridor using the same navigation behaviour previously used in the laboratory experiments, acquiring one image frame per second, which resulted in the acquisition of 50 images per journey along the corridor. Differently from the previous laboratory experiments, the camera's pan-tilt unit was driven to its home position (facing straight towards the forward direction of the robot) for the experiments in the corridor.

Exploration was performed in the "empty" corridor to acquire a model of normality, as in the previous laboratory experiments, but limited to three journeys along the corridor. Finally, the learnt model of normality was used to inspect the corridor for unusual visual features that were manually inserted *a posteriori*.

We placed three different novel objects in the corridor at different times: a black rubbish bag, a dark brown bin and a yellow wooden board. These objects appeared in the robot's field of view immediately after the traversal of a door, which was present in the corridor.

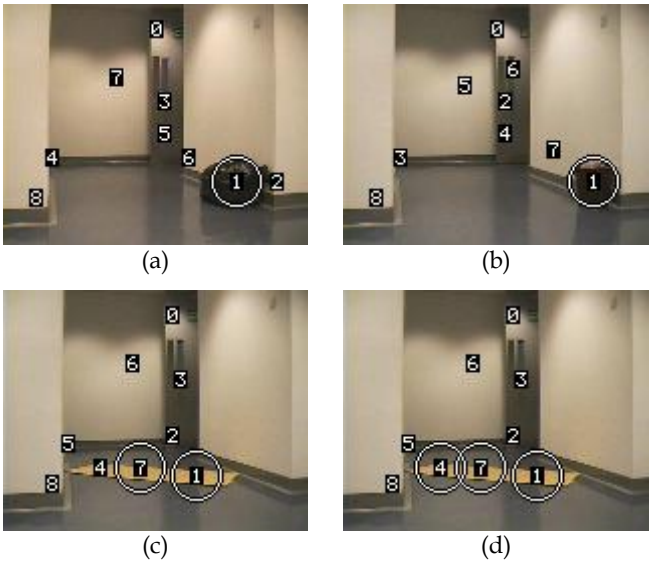
Results. After three exploration journeys along the empty corridor, the GWR network acquired 48 nodes, while the incremental PCA acquired 80 model vectors. Although the GWR network acquired fewer concepts, the incremental PCA algorithm has a more efficient and compact representation.

Apart from the wooden board, the chosen novel objects are dark and therefore to some extent similar to the dark areas of the normal environment. Contrast in the images acquired in the corridor was generally poor because no extra illumination was used, just the weak lighting already present. In spite of this fact, both GWR network and incremental PCA algorithm were able to correctly highlight the novel objects in the corridor during inspection, as shown in Figure 9.

However, both novelty filters also responded with false novelty indications for a pair of fire extinguishers that were present in the corridor. These novelty indications were unexpected because the fire extinguishers were already present during the exploration phase and therefore should have been part of the acquired model of normality. We attribute such false novelty indications to changes in the scale of the already known visual features of the fire extinguishers and believe that use of an image encoding method that is robust to changes in scale would contribute to reduce false novelty indications and enhance general performance

of the visual novelty filter. This hypothesis is currently under investigation (Vieira Neto & Nehmzow, 2007).

Table 4 shows a performance comparison in terms of Cramer's V, uncertainty coefficient U and κ index of agreement. All results showed statistically significant correlation between system response and actual novelty status (χ^2 analysis, $p \leq 0.01$). Overall performance (combined results for all three novel objects) indicates strong agreement between system response and actual novelty status. The GWR network presented the most consistent results.



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

Figure 9. Examples of output images for the real world experiments: (a) black rubbish bag (GWR and PCA); (b) dark brown bin (GWR and PCA); (c) yellow wooden board (GWR only); (d) yellow wooden board (PCA only). The white circles correctly indicate regions containing novel features.

	GWR network	Incremental PCA
Black bag	$V = 0.63$ $U = 0.35$ $\kappa = 0.63$	$V = 0.65$ $U = 0.37$ $\kappa = 0.65$
Brown bin	$V = 0.64$ $U = 0.38$ $\kappa = 0.64$	$V = 0.50$ $U = 0.23$ $\kappa = 0.50$
Yellow board	$V = 0.67$ $U = 0.37$ $\kappa = 0.67$	$V = 0.84$ $U = 0.69$ $\kappa = 0.84$
Overall	$V = 0.65$ $U = 0.36$ $\kappa = 0.65$	$V = 0.70$ $U = 0.44$ $\kappa = 0.70$

Table 4. Performance comparison for the real world experiments: all results correspond to statistically significant correlation between system response and actual novelty status (χ^2 analysis, $p \leq 0.01$).

5. Conclusion

To achieve novelty detection on visual images, for the purpose of automated inspection tasks, we used a mechanism of visual attention that selects candidate image patches in the input frame, combined with methods that classify image patches as novel or non-novel.

For real world inspection applications vision is the most appropriate sensor modality, as it provides colour, texture, shape, size, and distance information. All of this information is useful for robots operating in complex environments, but of course comes at the cost of processing large amounts of data, which is a particular challenge on autonomous mobile robots with limited computing power available.

We demonstrated that the use of the saliency map (Itti et al., 1998) as selective attention mechanism minimises the amount of data to be processed and at the same time makes it possible to localise *where* the novel features are in the image frame. The use of this attention mechanism avoids explicit segmentation of the input image (Singh & Markou, 2004) and makes the overall system more robust to translations due to robot motion.

Because novelty is of contextual nature and therefore can not be easily modelled, the approach that we follow is to first acquire a model of normality through robot learning and then use it as a means to highlight any abnormal features that are introduced in the environment. For this purpose, we have used unsupervised clustering mechanisms such as the GWR neural network (Marsland et al., 2002b) and the incremental PCA algorithm (Artač et al., 2002), which were both able to learn aspects of the environment incrementally and yielded very good results.

We proposed an experimental setup to evaluate performance and functionality of visual novelty filters, dividing the experimental procedure in two stages: an exploration phase, in which the learning mechanism was enabled to allow the robot to build a model of normality while experiencing the environment; and an inspection phase, in which the acquired model of normality was used as a novelty filter. Novel objects were inserted in the robot's environment during the inspection phase of experiments with the expected outcome that the visual novelty filter would produce indications of novelty and localise these new objects in the corresponding input image frame.

As the precise location and nature of the novelty introduced during the inspection phase is known by the experimenter, it is possible to generate ground truth data to be compared with the responses given by the novelty filter. In order to assess the performance of a novelty filter objectively, we used 2×2 contingency tables relating actual novelty status (ground truth) to system response, followed by the computation of statistical tests to quantify the association or agreement between them. Here we used the χ^2 test in order to check the statistical significance of the association between ground truth and novelty filter response, followed by the computation of Cramer's V, the uncertainty coefficient U and the κ index of agreement (Sachs, 2004).

Extensive experimental data was logged to evaluate and compare the efficiency of the visual novelty filter. The χ^2 analysis of the generated contingency tables revealed statistical significance in the associations between system response and actual novelty status in all of

the reported experiments. Typical quantitative analyses resulted in strong agreement with the ground truth data.
Qualitative assessment of the learning procedure during exploration, as well as consistent identification of novel features during inspection was made through the use of novelty bar graphs. In these graphs, a measure of the degree of novelty in each image frame is plotted against time/position. Novelty graphs are particularly useful to identify novelty indications in unexpected locations of the environment and investigate their reasons, leading to improvements in overall system robustness and ability to generalise.
We consider the results obtained to be very good and likely to succeed in real world applications that involve exploration and inspection of environments using vision. An example of such an application is the automated inspection of sewer pipes. However, more elaborate processing to become more robust to general affine transformations is likely to be necessary for applications in which the environment is not as structured as the arena and corridor that were used as operating environments in this work.

5.1 Contributions

One of our main contributions was to implement and experiment with visual novelty detection mechanisms for applications in automated inspection using autonomous mobile robots. Previous work done in novelty detection used only low-resolution sonar readings (Crook et al., 2002; Marsland et al., 2002a) or very restricted monochrome visual input (Crook & Hayes, 2001; Marsland et al., 2001). In contrast to this, the work presented here used colour visual stimuli with an unrestricted field of view. The selection of algorithms had emphasis on bottom-up and unsupervised learning approaches to allow exploitation of relevant characteristics of the acquired data from the ground-up.
Quantitative performance assessment tools based on contingency table analysis and statistical tests were developed in order to support objective comparisons between different visual novelty filters. For comparison purposes, novelty ground truth maps were generated in the form of binary images, in which novel visual features are highlighted manually. Because vision is a sensor modality shared between robots and humans, generation of novelty ground truth maps occurs in a natural and relatively easy way (although it demands time because of the volume of images involved).
Another main contribution was the demonstration that attention mechanisms extend the functionality of visual novelty filters, enabling them to localise where the novel regions are in the input frame and improving image encoding robustness to translations due to robot motion. Also, the use of an attention mechanism avoids explicit segmentation of the input image frame.

5.2 Future research

The results and conclusions drawn from the experiments in visual novelty detection reported in this work open a series of avenues for future investigations and improvements.
It would be interesting, for instance, to conduct more experiments using alternative attention mechanisms, especially those which can determine affine transformation parameters for the selected regions of interest. Possible options are the Harris-affine detector (Mikolajczyk & Schmid, 2002; Mikolajczyk & Schmid, 2004) and the interest point detector developed by Shi & Tomasi (1994). The use of such algorithms is expected to result in image encoding with extra robustness to affine transformations, improving the ability to generalise

and reducing the number of stored vectors or nodes by the novelty filter. Experiments are needed to compare performances with the attention mechanism already studied here to confirm or reject this hypothesis.

There are also some alternative methods of interest for the image encoding, which are likely to improve robustness to changes in scale and orientation of visual features. One possibility is the use of space-variant (log-polar) foveation (Bernardino et al., 2002) and the Fourier-Mellin transform (Derrode, 2001; Reddy & Chatterji, 1996) in order to encode visual features.

For applications that demand a systematic exploration of complex large-scale environments, such as a whole floor in a building, the integration of the proposed visual novelty detection framework with the environment exploration scheme developed by Prestes e Silva Jr. et al. (2002; 2004) is of particular interest. This approach uses potential fields to generate a dynamic exploration path that systematically covers the entire free area of the robot's environment, while generating a grid map of the obstacles that are present. Later on, the generated grid map can be used to produce arbitrary inspection paths or even paths towards specific goals.

If a novelty detection algorithm is used to learn and associate the local visual appearance of the environment to the grids of the environmental map, it is possible to determine novelty not only in terms of uncommon features that may appear in the environment, but also to establish if known features appear in unusual locations. A potential application of such ability is the automated organisation of a room, in which an autonomous mobile robot would be able to identify which objects are not in the places they were supposed to be and then take actions to correct the situation, "tidying up" the environment.

6. Acknowledgements

The research and experimental work reported in this chapter was conducted at the Robotics Research Laboratory of the University of Essex (UK) with the financial support of CAPES Foundation (Brazil), which is gratefully acknowledged.

7. References

Artač, M.; Jogan, M. & Leonardis, A. (2002). Incremental PCA for on-line visual learning and recognition. *Proceedings of the 16th International Conference on Pattern Recognition*, Vol. 3, pp. 781-784, ISBN 0-7695-1695-X, Quebec, Canada, August 2002.

Bernardino, A.; Santos-Victor, J. & Sandini, G. (2002). Model-based attention fixation using log-polar images, In: *Visual Attention Mechanisms*, Cantoni, V.; Petrosino, A. & Marinaro, M. (Ed.), pp. 79-92, Plenum Press, ISBN 978-0306474279, New York, USA.

Crook, P. & Hayes, G. (2001). A robot implementation of a biologically inspired method for novelty detection. *Proceedings of TIMR 2001 - Towards Intelligent Mobile Robots*, Manchester, UK, April 2001.

Crook, P.; Marsland, S.; Hayes, G. & Nehmzow, U. (2002). A tale of two filters – on-line novelty detection. *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*, Vol. 4, pp. 3894-3899, ISBN 0-7803-7272-7 Washington DC, May 2002.

Derrode, S. (2001). Robust and efficient Fourier-Mellin Transform approximations for gray-level image reconstruction and complete invariant description. *Computer Vision and Image Understanding*, Vol. 83, No. 1, pp. 57-78, ISSN 1077-3142.

Hall, P.; Marshall, D. & Martin, R. (1998). Incremental eigenanalysis for classification. *Proceedings of the 9th British Machine Vision Conference*, Vol. 1, pp. 286-295, ISBN 1-901725-04-9, Southampton, UK, September 1998.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences,* Vol. 79, No. 8, pp. 2554-2558, ISSN 0027-8424.

Itti, L. & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience,* Vol. 2, No. 3, pp. 194-203, ISSN 1471-0048.

Itti, L.; Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, pp. 1254-1259, ISSN 0162-8828.

Kohonen, T. (1984). *Self-organization and Associative Memory*, Springer-Verlag, ISBN 978-3540513872, New York, USA.

Linåker, F. & Niklasson, L. (2000). Time series segmentation using an adaptive resource allocating vector quantization network based on change detection, *Proceedings of the 2000 International Joint Conference on Neural Networks*, pp. 323-328, ISBN 978-0769506197, Como, Italy, July 2000, IEEE Computer Society Press.

Linåker, F. & Niklasson, L. (2001). Environment identification by alignment of abstract sensory flow representations, In: *Advances in Neural Networks and Applications,* Mastorakis, N.; Mladenov, V.; Suter, B. & Wang, L. (Ed.), 79-116, WSES Press, ISBN 960-8052-26-2.

Marsland, S.; Nehmzow, U. and Shapiro, J. (2000). Detecting novel features of an environment using habituation, *From Animals to Animats 6: Proceedings of the 6th International Conference on Simulation of Adaptive Behavior*, pp. 189-198, ISBN 978-0262632003, Paris, France, September 2000, MIT Press.

Marsland, S.; Nehmzow, U. & Shapiro, J. (2001). Vision-based environmental novelty detection on a mobile robot, *Proceedings of the International Conference on Neural Information Processing*, ISBN 978-7-309-03012-9, Shanghai, China, November 2001.

Marsland, S.; Nehmzow, U. & Shapiro, J. (2002a). Environment-specific novelty detection. *From Animals to Animats 7: Proceedings of the 7th International Conference on the Simulation of Adaptive Behaviour*, pp. 36-45, ISBN 0-262-58217-1, Edinburgh, UK, August 2002, MIT Press, Cambridge, USA.

Marsland, S.; Shapiro, J. & Nehmzow, U. (2002b). A self-organising network that grows when required. *Neural Networks*, Vol. 15, No. 8-9, pp. 1041-1058, ISSN 0893-6080.

Mikolajczyk, K. & Schmid, C. (2002). An affine invariant interest point detector, *Computer Vision - ECCV 2002: Proceedings of the 7th European Conference on Computer Vision*, pp. 128-142, ISBN 978-3540437444, Copenhagen, Denmark, May 2002, Springer-Verlag, Berlin, Germany.

Mikolajczyk, K. & Schmid, C. (2004). Scale and affine invariant interest point detectors. *International Journal of Computer Vision,* Vol. 60, No. 1, pp. 63-86, ISSN 0920-5691.

Nehmzow, U. (2003). *Mobile Robotics: A Practical Introduction*, 2nd ed., Springer-Verlag, ISBN 978-1852337261, London, UK.

Prestes e Silva Jr., E.; Engel, P. M.; Trevisan, M. & Idiart, M. A. P. (2002). Exploration method using harmonic functions. *Robotics and Autonomous Systems,* Vol. 40, No. 1, pp. 25-42, ISSN 0921-8890.

Prestes e Silva Jr., E.; Idiart, M. A. P.; Trevisan, M. & Engel, P. M. (2004). Autonomous learning architecture for environmental mapping. *Journal of Intelligent and Robotic Systems,* Vol. 39, No. 3, pp. 243-263, ISSN 0921-0296.

Reddy, B. S. & Chatterji, B. N. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing,* Vol. 5, No. 8, pp. 1266-1271, ISSN 1057-7149.

Sachs, L. (2004). *Angewandte Statistik: Anwendung statistischer Methoden,* Springer Verlag, ISBN 978-3540405559, Berlin, Germany.

Shi, J. & Tomasi, C. (1994). Good features to track, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600, ISBN 0-8186-5825-8, Seattle, USA, June 1994, IEEE Computer Society.

Singh, S. & Markou, M. (2004). An approach to novelty detection applied to the classification of image regions. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 4, pp. 396-407, ISSN 1041-4347.

Tarassenko, L; Hayton, P.; Cerneaz, N. & Brady, M. (1995). Novelty detection for the identification of masses in mammograms. *Proceedings of the 4th IEEE International Conference on Artificial Neural Networks*, pp. 442-447, ISBN 0-85296-641-5, Cambridge, UK, June 1995.

Vieira Neto, H. & Nehmzow, U. (2005). Automated exploration and inspection: comparing two visual novelty detectors. *International Journal of Advanced Robotic Systems*, Vol. 2, No. 4, pp. 355-362, ISSN 1729-8806.

Vieira Neto, H. & Nehmzow, U. (2007). Visual novelty detection with automatic scale selection. *Robotics and Autonomous Systems*, Vol. 55, No. 9, pp. 693-701, ISSN 0921-8890.