On-line Visual Novelty Detection in Autonomous Mobile Robots

Hugo Vieira Neto
Graduate School of Electrical Engineering and Applied Computer Science,
Federal University of Technology – Paraná, Brazil

1 Introduction

Autonomous mobile robots face great challenges in order to operate in complex dynamic environments. On one hand, it is necessary to sense the environment in such a way that relevant information is gathered for the execution of the desired robotic task – in this sense, some of the most useful and interesting applications require the use of relatively high resolution vision sensors, which are potentially able to provide multi-modal environmental information ranging from low-level colour and texture to high-level depth and motion. But on the other hand, the limited computational resources available to an autonomous mobile robot are often not enough to process all the massive amounts of collected visual data in real-time.

A possible solution to this problem is to enable the robot to learn continuously about its operating environment, using selective attention to filter out aspects from the environment which are likely to be relevant for the desired task and novelty detection to highlight unusual stimuli that should be subject to higher levels of processing and analysis. Novelty detection – the competence to identify perceptions that were never experienced before – is a fundamental ability for autonomous mobile robots that aim at learning details from their operating environment continuously. This is the case for tasks that involve unsupervised environment exploration and mapping, in which knowledge is to be acquired incrementally and without supervision. Moreover, novelty detection mechanisms also make automated inspection and surveillance tasks possible by providing the robot with the ability to pinpoint unusual situations in its environment.

In this chapter we present variations of an on-line visual novelty detection framework that can be used in environment exploration and inspection by autonomous mobile robots. Novelty detection tasks are fundamentally different from usual pattern recognition problems in which the main features of interest are known beforehand. For example, in a face recognition task one can determine beforehand that a face is roughly composed of two eyes, a nose and a mouth in a particular geometrical arrangement – *i.e.* there is a model, which can be more or less refined, to be searched for. In contrast, in novelty detection tasks there are no *a priori* models to be searched for, because novelty can be *anything* that deviates from the usual perceptions from the environment. Arguably, the most feasible approach to be followed to solve the novelty detection problem is to learn a model of *normality* of the robot's environment, and then use it to filter out *any abnormal* sensory perceptions (Tarassenko et al., 1995). Following this approach, abnormal perceptions are conveniently defined as anything that does not fit the acquired model of normality.

The first successful use of novelty detection mechanisms in physical autonomous mobile robots performing exploration and inspections tasks was initially done using low-resolution sonar data (Crook et al., 2002; Marsland et al., 2002a), following some trials using very rough monochromatic vision (Crook & Hayes, 2001; Marsland et al., 2001). More recently, a more comprehensive framework for novelty detection using more refined high-resolution colour vision in autonomous mobile robots, which is discussed in this chapter, was proposed in (Vieira Neto, 2006). The main idea behind all of these approaches is to use an on-line unsupervised learning mechanism to acquire a model of normality of the environment and then use it to highlight any unusual sensory perceptions. In this work we initially focus on the use of the Grow-When-Required (GWR) neural network (Marsland et al., 2002b) (section 3) as an on-line novelty filter for visual features. More details about this and other available novelty detection approaches can be found in specialised surveys (Hodge & Austin, 2004; Markou & Singh, 2003a; Markou & Singh, 2003b; Marsland, 2003).

In order to use high-resolution visual information as the main perceptual input for learning mechanisms in general, especially if real-time operation is desired, it is almost always imperative to perform some sort of dimensionality reduction that preserves essential features and throws away unnecessary details of the raw visual data. One way of achieving dimensionality reduction is to model the probability distribution of the features of interest in the image – a popular choice for this task is the use of image descriptors based on histograms (Bay et al., 2008; Lowe, 2004; Swain & Ballard, 1991) or colour angular indexing (Finlayson et al., 1996), which can successfully be used to represent features from the entire image frame in a global fashion.

Figure 1 shows a block diagram of a generic visual novelty detection framework that uses a global image encoding approach in order to achieve dimensionality reduction.

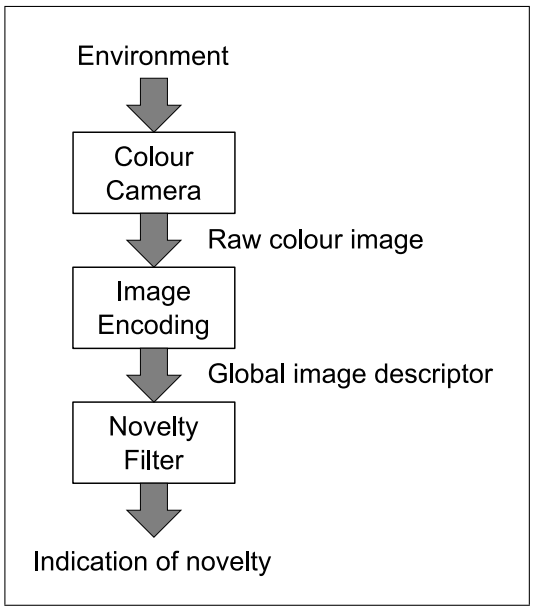


Figure 1: Global visual novelty detection block diagram: a colour camera is used to acquire an image from the environment, which is then encoded in a global feature descriptor with lower dimensionality, which is fed to a novelty filter that determines *when* novel features are present in the camera's field of view.

The global image description approach shown in figure 1 allows the novelty detection system to determine *when* novel features enter the field of view of the robot's camera, as will be shown in section 4. However, a much

more useful approach should not only determine *when* but also *where* novel features are localised within the field of view of the robot's camera, as will be shown in section 6. In this new approach, a visual attention mechanism is used in order to select relevant image regions that are represented by local feature descriptors. Local description of small image regions in the vicinity of interest points selected by an attention mechanism not only results in data dimensionality reduction and allows to determine the location of important features, but also represents details that would otherwise be lost in a global description.

A block diagram of a generic visual novelty detection framework that uses an attention-based local image encoding approach is shown in figure 2.

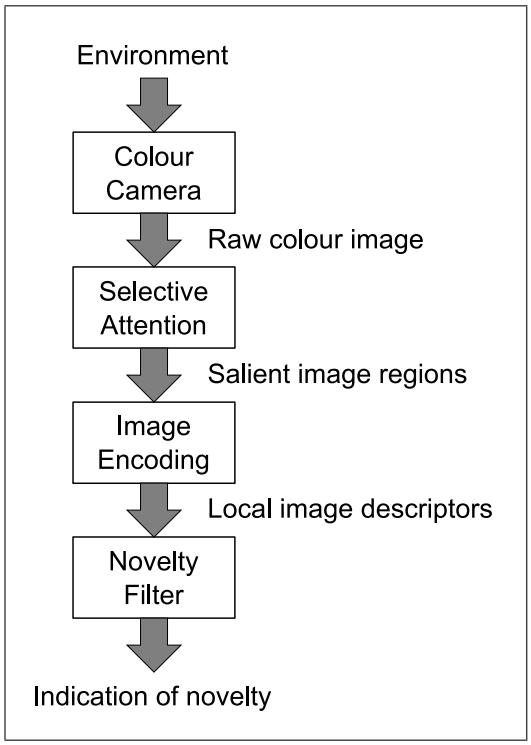


Figure 2: Local visual novelty detection block diagram: the colour image acquired from the environment is subject to an attention mechanism, which selects salient regions to be encoded in local feature descriptors, which are then fed to a novelty filter that not only determines *when* but also *where* novel features are present within the camera's field of view.

There are several choices for the selective attention mechanism (Bay et al., 2008; Kadir & Brady, 2003; Lowe, 2004; Mikolajczyk & Schmid, 2004; Itti et al., 1998), which aims at selecting interest points around which the local descriptive information content is maximised according to some criteria. However, a particularly interesting selective attention mechanism in the context of the local visual novelty detection framework presented in figure 2 is the saliency map model (Itti et al., 1998) (section 5), which combines different visual features in multiple scales in order to obtain a general indication of visual saliency at each image location. The concept of assessing saliency – the property to stand out from the background – is very convenient for the identification of uncommon features, which is precisely what is desired in novelty detection tasks. In this local image description approach, the saliency map model acts as the first selection stage of candidate unusual regions within single image frames that will be subject to further analysis by the novelty filter itself in a more general sense.

One final aspect, investigated in section 8, is the possibility of achieving both image description and novelty detection within a single algorithm. As explained earlier, some sort of dimensionality reduction of the raw visual data is often necessary to allow real-time operation of the learning mechanisms involved in novelty detection. The low-dimensionality image description to be used should ideally preserve essential features and discard unnecessary details of the visual data. However, "essential features" and "unnecessary details" are not always clear to the designer of such a system and therefore a bottom-up approach that *autonomously* extracts relevant information from the raw data itself seems to be much more convenient. In this sense, a very interesting mechanism that allows both bottom-up dimensionality reduction of raw visual data and novelty detection is the incremental Principal Component Analysis algorithm for on-line visual learning and recognition (Artač et al., 2002) (section 7). This approach yields a bottom-up description that allows image reconstruction, which provides the user with *visual* information of which aspects were acquired in the model of normality of the environment, something that was not possible when using the previously discussed top-down image description approaches.

The remaining sections of this chapter describe details of the algorithms used in all variations of the on-line visual novelty detection framework in discussion and also show practical results of their use in the experimental setup described in section 2. Section 3 describes the GWR neural network and how it can be used as a novelty filter. A demonstration of the performance of a visual novelty detection approach using the GWR neural network and *global* image descriptors in a physical mobile robot follows in section 4. Next, the description of the saliency map model of visual attention is given in section 5, followed by a demonstration in section 6 of the performance of the same visual novelty detection approach, but now using *local* image descriptors. Section 7 describes how incremental PCA can be used alternatively as a visual novelty filter with *autonomous* image representation, whose performance is then demonstrated in section 8. Finally, conclusions are drawn in section 9.

2 Experimental Setup

In order to demonstrate the performance of the visual novelty detection framework in discussion, experiments were devised and conducted in a controlled laboratory scenario. Every experimental round consists of two stages: an exploration (learning) phase, in which a physical mobile robot is used to acquire a model of normality of the operating environment, followed by an inspection (application) phase, in which the acquired model is then used to highlight any abnormal perceptions that may be encountered by the robot in the same operating environment.

During the learning phase, images are acquired while the robot navigates around a "baseline" environment – an empty arena delimited by cardboard boxes. These images are encoded in order to generate input feature vectors and train the learning mechanism (GWR neural network or incremental PCA algorithm) that later will be used as novelty filter. The expected outcome is that the amount of novelty detected during the exploration phase continuously decreases as a result of learning. During the application phase, novel objects are deliberately placed within the robot's operating environment and new images are acquired during navigation in order to test the performance of the trained novelty filter. It is expected that peaks in the novelty measure now appear only when a new object appears in the field of view of the robot's camera.

Figure 3 depicts the experimental setup used for the laboratory experiments that follow in this chapter. The colour vision system of the Magellan Pro mobile robot shown in figure 3a was used to generate visual stimuli while navigating in the square arena delimited by cardboard boxes shown in figure 3b. The cardboard boxes act as walls that limit the robot's trajectory and visual world – only the walls of the arena and the floor are within the camera's field of view. Figure 3c also shows the top view of the trajectory followed by the robot inside the arena with its corners numbered from one to four. Novel objects were placed in the corners of the arena so that the performance of the novelty filter could be assessed.

(a) (b) (c)

Figure 3: Experimental setup: (a) Magellan Pro mobile robot; (b) robot arena made of cardboard boxes; (c) top view of the trajectory followed by the robot (represented by a circle with a line indicating its front) inside the arena delimited by the cardboard boxes (represented by rectangles).

3 The GWR Neural Network

The GWR neural network (Marsland et al., 2002a; Marsland et al., 2002b) is a self-organising neural network that consists of a clustering layer of nodes and a single output node. Clustering nodes consist of model weight vectors that represent centres of clusters in input space, associated to radial basis activation functions that determine the receptive field of each cluster. The synapses which connect nodes in the clustering layer to the output node are subject to a model of habituation, which is a reduction in response to inputs that are repeatedly presented – the more a clustering node fires, the weaker its output synapse becomes.

The habituation rule of the synaptic efficacy of clustering nodes to the output node is given by the following first-order differential equation:

$$\tau \frac{dh(t)}{dt} = \alpha[h_0 - h(t)] - S(t), \tag{1}$$

where h_0 is the initial value of the synaptic efficacy $h(t)$, $S(t)$ is the stimulus, and τ and α are time constants that control the habituation rate and recovery rate, respectively. $S(t) = 1$ causes habituation (reduction in synaptic efficacy) and $S(t) = 0$ causes dishabituation (recovery of synaptic efficacy).

Typically, $\tau = 3.33$, $\alpha = 1.05$, $h_0 = 1$ and $S = 1$, which results in synaptic efficacy ranging from approximately 0.05 (meaning complete habituation) to 1 (meaning complete dishabituation). As synaptic efficacy is bounded, it can be used neatly as a measurement of the degree of novelty for any particular input – higher synaptic efficacies correspond to higher degrees of novelty.

Learning of the GWR network is performed using a winner-take-all approach. Every time that an input vector is presented to the network, each node in the clustering layer will have higher or lower activation depending on how well its weight vector matches the input. However, only the best matching node fires in response to a given input, inhibiting all other clustering nodes. Then, the weights of the winner clustering node are adapted and so are the weights of each of its topological neighbours, although to a lesser extent than the winner node.

The GWR neural network has the ability to add nodes to its clustering structure by identifying new input stimuli through the habituation model. Given an input vector, both the winner node activation and habituation are used to a determine if a new clustering node should be allocated in order to represent the input space better. The network is first initialised with two completely dishabituated nodes (c_1 and c_2) in its clustering map M:

$$M = \{c_1, c_2\}. \tag{2}$$

The weight vectors for these two initial clustering nodes can be initialised with the first two input vectors presented to the network. At first there are no topological connections between the clustering nodes and therefore the connection set C is initialised to the empty set:

$$C = \emptyset. \tag{3}$$

From the third input vector onwards, the best matching node s (winner node) and second best matching node t are found by computing the Euclidean distance from the input vector $\mathbf{x}$ to each clustering node:

$$s = \arg\min_{i \in M} \| \mathbf{x} - \mathbf{w}_i \|, \tag{4}$$

$$t = \arg\min_{i \in M/\{s\}} \| \mathbf{x} - \mathbf{w}_i \|, \tag{5}$$

where $\mathbf{w}_i$ is the weight vector of node i, with i covering all the existing nodes in the current map M.

If there is an existing connection between clustering nodes s and t, its age is set to zero (the age of a connection corresponds to how many iterations of the algorithm have elapsed since the connection was created), otherwise a new connection between clustering nodes s and t is created with age zero:

$$C = C \cup \{(s,t)\}. \tag{6}$$

Activation of the clustering nodes is computed using the following radial basis function:

$$a_i = \exp(-\|\mathbf{x} - \mathbf{w}_i\|^2). \tag{7}$$

Both activation and habituation values of the winner node are used to decide whether a given input is considered novel or not. Every time that both activation and habituation values are below predefined thresholds a_T and h_T, respectively, a new node r is added to the clustering layer:

$$M = M \cup \{r\}, \tag{8}$$

whose weight vector $\mathbf{w}_r$ is set to the average between the winner weight vector $\mathbf{w}_s$ and the input vector $\mathbf{x}$.

After inserting a new node, it is necessary to update the network topology by removing the connection between nodes s and t:

$$C = C/\{(s,t)\} \tag{9}$$

and by inserting connections between nodes r and s, and between nodes r and t:

$$C = C \cup \{(r,s),(r,t)\}. \tag{10}$$

Then, the winner node and all of its topological neighbours have their output synapses habituated according to equation 1, and their cluster centres are adapted according to the following learning rule:

$$\Delta \mathbf{w}_i = \varepsilon(\mathbf{x} - \mathbf{w}_i), \tag{11}$$

where ε is the learning rate ($0 < \varepsilon < 1$).

The learning and habituation rates of the neighbour nodes – ε_j and τ_j, respectively – are made proportional to the ratio between winner and neighbour node activations (Vieira Neto & Nehmzow, 2004), so that neighbour nodes have their weights adapted to a lesser extent than the winner node and also habituate in a slower rate.

The final step of the GWR network learning iteration consists of incrementing the age of every existing connection, and checking for nodes that no longer have any neighbours or connections whose age is greater than a predefined threshold age_{max} to be removed.

A summary of the use of the GWR network as a novelty filter is given in algorithm 1.

Algorithm 1: GWR neural network novelty detection. Typical parameters: $a_T = 0.9$, $h_T = 0.3$, $\eta = 0.1$, $\varepsilon = 0.1$, $\tau = 3.33$, $\alpha = 1.05$, $h_0 = 1$, $S = 1$ and $age_{max} = 20$.

Input: current set of nodes A, current set of connections C, new input vector $\mathbf{x}$.
Output: updated set of nodes A, updated set of connections C, novelty indication N.

1 Find the best and second best matching nodes s and t to the new input vector $\mathbf{x}$: $s = \arg\min_{i \in A} \| \mathbf{x} - \mathbf{w}_i \|$, $t = \arg\min_{i \in A/\{s\}} \| \mathbf{x} - \mathbf{w}_i \|$, where $\mathbf{w}_i$ is the weight vector of node i.
2 **if** there is a connection between nodes s and t **then** set the connection's age to zero: $age_{(s,t)} = 0$ **else** create a new connection between nodes s and t: $C \Leftarrow C \cup \{(s,t)\}$, $age_{(s,t)} = 0$.
3 Compute the activation of the best matching node: $a_s = \exp(- \| \mathbf{x} - \mathbf{w}_s \|)$.
4 Test if the activation and habituation values of the best matching node characterise novelty:
if $a_s < a_T$ and $h_s < h_T$ **then**
5 Add a new node r: $A \Leftarrow A \cup \{r\}$.
6 Set the weight vector of the new node r: $\mathbf{w}_r = \frac{1}{2}(\mathbf{x} + \mathbf{w}_s)$.
7 Remove the connection between nodes s and t: $C \Leftarrow C/\{(s,t)\}$.
8 Create connections between nodes r and s and between nodes r and t: $C \Leftarrow C \cup \{(r,s),(r,t)\}$, $age_{(r,s)} = 0$, $age_{(r,t)} = 0$.
9 Indicate novelty detected: $N = 1$.
10 **else** indicate no novelty detected: $N = 0$.
11 Compute the activation of the best matching node's neighbour nodes, *i.e.* nodes with connections to the best matching node: $a_j = \exp(- \| \mathbf{x} - \mathbf{w}_j \|)$.
12 Adapt the weight vector of the best matching node: $\mathbf{w}_s \Leftarrow \mathbf{w}_s + \varepsilon(\mathbf{x} - \mathbf{w}_s)$
13 Adapt the weight vectors of the neighbour nodes: $\mathbf{w}_j \Leftarrow \mathbf{w}_j + \frac{\eta a_j}{a_s}[\varepsilon(\mathbf{x} - \mathbf{w}_j)]$.
14 Habituate the best matching node: $h_s \Leftarrow h_s + \frac{\alpha(h_0 - h_s) - S}{\tau}$.
15 Habituate the neighbour nodes: $h_j \Leftarrow h_j + \frac{\eta a j}{a_s}\left[\frac{\alpha(h_0 - h_j) - S]}{\tau}\right]$.
16 Increment the age of connections to the best matching node.
17 Remove any connections with $age_{(s,j)} > age_{max}$.
18 Remove any nodes without neighbours.

4 Global Image Descriptors and Visual Novelty Detection

The first experiment to be shown using visual input to a GWR neural network involves the use of global image descriptors based on colour statistics. The idea is to use a simple and fast image encoding technique to reduce the dimensionality of the input visual data (160 × 120 pixels), so that it can be efficiently processed by the novelty filter. In this initial experiment, the performance of colour angles as global image descriptors is analysed – colour angular indexing (Finlayson et al., 1996) provides a very compact colour constant representation based on the characteristics of the colour distribution within the image frame, in the form of angles between colour vectors **r**, **g** and **b** that contain all pixel values of the red, green and blue channels, respectively, of the input image in scanning order.

The first step in order to compute the colour angles is to obtain zero-mean colour vectors $\mathbf{r}_0$, $\mathbf{g}_0$ and $\mathbf{b}_0$ by subtracting the corresponding average pixel value of each original colour vector:

$$\mathbf{r}_0 = \mathbf{r} - \bar{r}, \tag{12}$$

$$\mathbf{g}_0 = \mathbf{g} - \bar{g}, \tag{13}$$

$$\mathbf{b}_0 = \mathbf{b} - \bar{b}, \tag{14}$$

where $\bar{r}$, $\bar{g}$ and $\bar{b}$ are the average pixel values of the original colour vectors **r**, **g** and **b**, respectively.

The next step is to normalise the zero-mean colour vectors to unitary length by dividing each one by their respective norm:

$$\mathbf{r}_N = \frac{\mathbf{r}_0}{\|\mathbf{r}_0\|}, \tag{15}$$

$$\mathbf{g}_N = \frac{\mathbf{g}_0}{\|\mathbf{g}_0\|}, \tag{16}$$

$$\mathbf{b}_N = \frac{\mathbf{b}_0}{\|\mathbf{b}_0\|}, \tag{17}$$

where $\mathbf{r}_N$, $\mathbf{g}_N$ and $\mathbf{b}_N$ are the normalised zero-mean colour vectors.

Colour channel covariances are equivalent to dot products, which in this case geometrically correspond to the cosine of the angles between the corresponding unitary length colour vectors. These angles are invariant to changes in illumination and can be computed by the inverse cosine of colour vector dot products:

$$\phi_{rg} = \arccos(\langle \mathbf{r}_N, \mathbf{g}_N \rangle), \tag{18}$$

$$\phi_{gb} = \arccos(\langle \mathbf{g}_N, \mathbf{b}_N \rangle), \tag{19}$$

$$\phi_{rb} = \arccos(\langle \mathbf{r}_N, \mathbf{b}_N \rangle), \tag{20}$$

where $\langle \mathbf{i}, \mathbf{j} \rangle$ denotes the dot product of vectors **i** and **j**.

The interesting aspect of this colour representation is that changes in sampled illumination due to robot motion result in a rotation of the colour channel vectors of the image. As the angles between these vectors remain the same in spite of any rotation, the resulting descriptors are robust to illumination changes. However, colour angles cannot discriminate shades of grey because in this particular case colour angles always result in $\phi_{rg} = \phi_{gb} = \phi_{rb} = 0$. In order to solve this problem, intensity information was included in the image descriptor as an additional element, resulting in feature vectors of four dimensions. The normalised intensity standard deviation σ_I is a relative measurement that was used with success in (Vieira Neto, 2006):

$$\sigma_I = \frac{2}{I_{max} N} \sqrt{\sum_{n=1}^{N} (I(n) - \bar{I})^2}, \tag{21}$$

where $I(n)$ are the image intensity values, $\bar{I}$ is the mean intensity value, I_{max} is the maximum intensity value and N is the total number of pixels in the image.

The experiment starts with an exploration phase, in which the robot is used to acquire a model of normality of the environment. In this work, exploration was always 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 data for qualitative comparisons between loops.

Images were acquired at one frame per second, resulting in a total of 50 images per loop around the arena. A global descriptor using colour angles and normalised intensity standard deviation was computed for every acquired image frame and fed as input vector to a GWR neural network with the typical parameter values

discussed in section 3. During exploration, learning was enabled to allow the GWR network to acquire a model of normality of the operating environment.

Figure 4 shows the novelty graphs for each of the five consecutive exploration loops around the empty arena. Novelty graphs are used for qualitative performance assessment, in which the novelty measurements provided by the novelty filter are plotted against the image frame number, in a similar fashion to what was done in (Marsland et al., 2002a) using sonar scans. In these graphs, each frame number essentially corresponds to a certain position and orientation of the robot in the environment because the navigation behaviour that was used is highly repeatable (Vieira Neto, 2006).

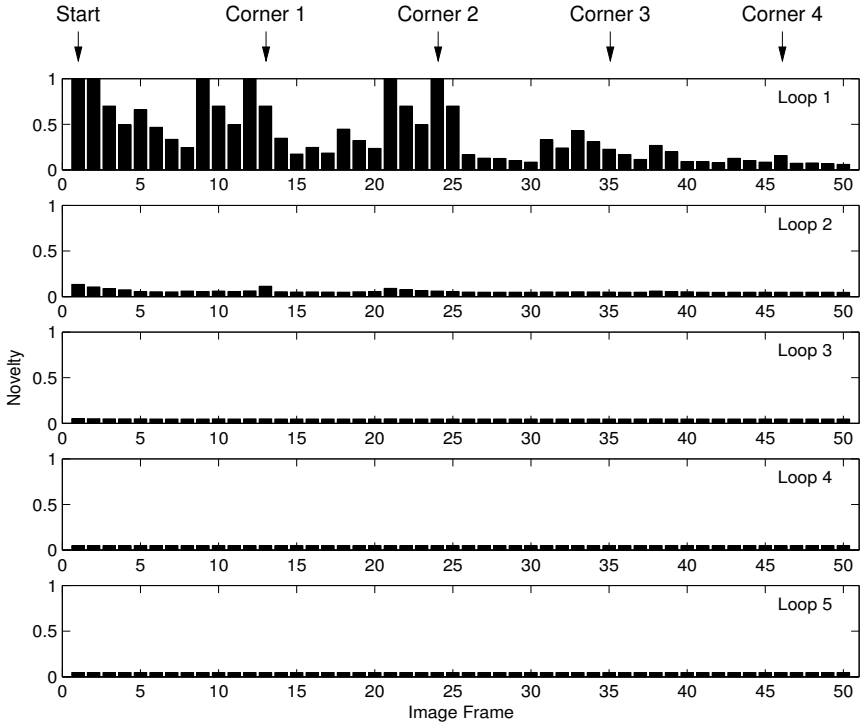


Figure 4: Exploration of the empty arena as normal environment, using global colour angles as image encoding scheme. The graphs show that the amount of novelty gradually decreases as the GWR neural network habituates on repeated stimuli. Complete habituation is achieved by the end of the second exploration loop.

Because the images were acquired at one frame per second, the horizontal axis of the novelty graphs in figure 4 can also be interpreted as time in seconds. It can be noticed that the amount of novelty measured during the exploration phase declined over time as the robot repeatedly explored its environment and progressively habituated to it, as expected. The efficiency of learning during the exploration phase can be graphically assessed through inspection of the novelty graphs in multiple rounds of training – in this case, the GWR neural network was completely habituated to the environment by the end of the second loop. After training, the model of normality acquired by the GWR network had six nodes, each containing a prototype colour descriptor learnt from the explored environment.

Once the GWR neural network network is trained, the acquired environmental model of normality can be used in the inspection phase to highlight any unusual visual features in the arena. During the inspection phase, a new object is introduced in the normal environment in order to test the system's ability to highlight abnormal

perceptions – the measurement of novelty is expected to be high only in places where the new object can be sensed by the robot. The inspection phase of the experiment was also carried out in five loops around the arena, but with the learning mechanism disabled so that unusual features in the environment could be highlighted every time that they were sensed.

In this experiment, an orange ball was deliberately placed as the novel object in one of the corners of the the arena and the robot was used to inspect it. The ball was selected not only because it had a good contrast to the normal environmental colour features, but also because it did not interfere with the robot's trajectory around the arena. 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 results obtained for the inspection phase of the arena containing the ball as novel object are shown in figure 5.

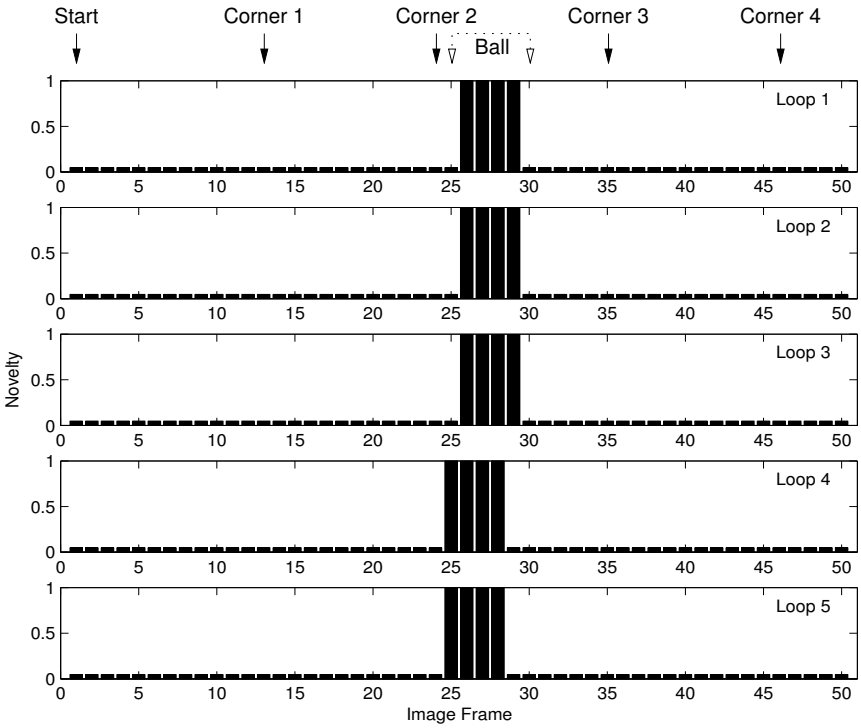


Figure 5: Inspection of the arena with an orange ball as novel object, using global colour angles as image encoding scheme. Locations where the ball was in the camera's field of view are indicated by dotted arrows on the top of the figure. The graphs show that the novel stimulus is correctly and consistently identified by the system in every inspection loop.

The set of frames where the orange ball appeared in the camera's field of view are indicated by dotted arrows on the top of figure 5 – these frames correspond to locations where high values for the novelty measure were expected to happen, as the ball appeared nearly always in the same image frames in each loop. It can be noticed clearly from the inspection novelty graphs shown that the ball was correctly and consistently highlighted as the novel stimulus in every inspection loop.

With the experiment using global image descriptors just shown, one can notice that a novelty filter is able to detect *when* novel visual features enter the field of view of the robot's camera. However, in order to be able also to determine *where* the novel features are localised within the image frame, local image descriptors should be used together with a selective visual attention mechanism, as will be discussed in the following sections.

5 The Saliency Map Model of Visual Attention

A simplified architecture for the computation of the saliency map model of visual attention (Itti et al., 1998), which consists in the construction of multi-scale feature maps, is presented in figure 6. Gaussian and oriented Gabor pyramids are constructed by successive filtering and subsampling of the input image, and then combined into feature maps that enable the detection of local variations in intensity, colour and orientation.

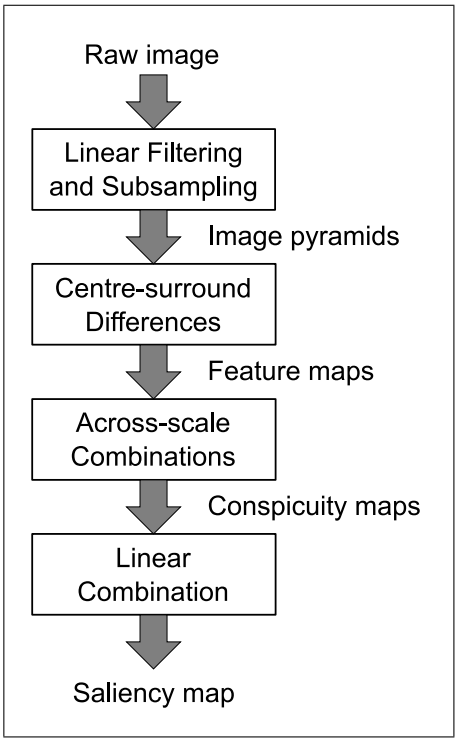


Figure 6: Simplified saliency map block diagram: multi-scale Gaussian and oriented Gabor pyramids are constructed from the input image and then combined into feature maps that yield a single saliency map, which indicates the location of unusual features within the image.

The first step in the extraction of early visual features is to obtain an intensity channel (**I**) from the original red (**R**), green (**G**) and blue (**B**) channels of the input image:

$$\mathbf{I} = \frac{1}{3}(\mathbf{R} + \mathbf{G} + \mathbf{B}). \tag{22}$$

After that, intensity normalised channels **r**, **g** and **b** are computed in order to decouple hue from intensity, but only at those locations where I is larger than $1/10$ of the maximum intensity I_{max}:

$$r = \begin{cases} R/I, & \text{if } I > I_{max}/10; \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

$$g = \begin{cases} G/I, & \text{if } I > I_{max}/10; \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

$$b = \begin{cases} B/I, & \text{if } I > I_{max}/10. \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

Four broadly tuned colour channels for red (**R**), green (**G**), blue (**B**) and yellow (**Y**) are then computed using the following equations:

$$R = \max\{0, r-(g+b)/2\}, \tag{26}$$

$$G = \max\{0, g-(r+b)/2\}, \tag{27}$$

$$B = \max\{0, b-(r+g)/2\}, \tag{28}$$

$$Y = \max\{0, -2B-|r-g|\}. \tag{29}$$

The intensity channel **I** and the broadly tuned colour channels **R**, **G**, **B** and **Y** are used to construct Gaussian pyramids (Burt & Adelson, 1983) $\mathbf{I}(\sigma)$, $\mathbf{R}(\sigma)$, $\mathbf{G}(\sigma)$, $\mathbf{B}(\sigma)$ and $\mathbf{Y}(\sigma)$, respectively ($\sigma \in [0, \sigma_{max}]$). The intensity channel **I** is also used to construct oriented Gabor pyramids (Greenspan et al., 1994) $\mathbf{O}(\sigma, \theta)$ in four orientations ($\theta \in \{0°, 45°, 90°, 135°\}$).

Centre-surround linear operations similar to receptive fields found in neurons along the visual pathway of mammals are used to obtain feature maps, which are implemented as the difference between a fine centre scale $c \in [c_{min}, c_{max}]$ and a coarser surround scale $s = c + \delta$, with $\delta \in [\delta_{min}, \delta_{max}]$. Across-scale difference (denoted by $\ominus$) is obtained by bilinear interpolation from the coarse scale to the fine scale and subsequent pixelwise subtraction.

The first type of feature map is related to local intensity contrast, detected by neurons sensitive to bright centres and dark surrounds or vice-versa. Both types of sensitivity are simultaneously obtained by the use of rectification:

$$I(c,s) = |\mathbf{I}(c) \ominus \mathbf{I}(s)|. \tag{30}$$

The second type of feature map accounts for colour double-opponency, which is detected by neurons whose centres are excited by one colour and inhibited by another, while the opposite excitation relationship holds for the surrounds. Colour feature maps are computed for red/green and blue/yellow double-opponent pairs:

$$\mathcal{RG}(c,s) = |(\mathbf{R}(c) - \mathbf{G}(c)) \ominus (\mathbf{R}(s) - \mathbf{G}(s))|, \tag{31}$$

$$\mathcal{BY}(c,s) = |(\mathbf{B}(c) - \mathbf{Y}(c)) \ominus (\mathbf{B}(s) - \mathbf{Y}(s))|. \tag{32}$$

The third type of feature map is concerned with local orientation contrast between centre and surround scales. Orientation feature maps are computed separately for every orientation:

$$O(c,s,\theta) = |\mathbf{O}(c,\theta) \ominus \mathbf{O}(s,\theta)|. \tag{33}$$

In order to combine feature maps with different dynamic ranges into a single saliency map it is necessary to use a normalisation operator $\mathcal{N}(.)$ (Itti et al., 1998; Vieira Neto, 2006). Otherwise, salient features that are strongly present in a few maps may be masked by noise or less salient features that appear more frequently. The use of the normalisation operator ultimately results in giving more weight to unusual features in the scope of the input image frame and therefore makes the saliency map an excellent choice for the task of selecting candidate regions to be processed by a novelty filter.

Feature maps for each feature – intensity, colour and orientation – are then combined in three conspicuity maps at scale c_{min}. The conspicuity maps are obtained by computing across-scale addition (denoted by $\oplus$), which consists of resampling each feature map to scale c_{min} and subsequent pixelwise addition:

$$\bar{I} = \bigoplus_{c=c_{min}}^{c_{max}} \bigoplus_{s=c+\delta_{min}}^{c+\delta_{max}} \mathcal{N}(I(c,s)), \tag{34}$$

$$\bar{C} = \bigoplus_{c=c_{min}}^{c_{max}} \bigoplus_{s=c+\delta_{min}}^{c+\delta_{max}} [\mathcal{N}(\mathcal{RG}(c,s)) + \mathcal{N}(\mathcal{BY}(c,s))], \tag{35}$$

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} \mathcal{N}\left(\bigoplus_{c=c_{min}}^{c_{max}} \bigoplus_{s=c+\delta_{min}}^{c+\delta_{max}} \mathcal{N}(O(c,s,\theta)) \right). \tag{36}$$

Finally, the three conspicuity maps are normalised and averaged to yield the final saliency map:

$$\mathcal{S} = \frac{1}{3}(\mathcal{N}(\bar{I}) + \mathcal{N}(\bar{C}) + \mathcal{N}(\bar{O})) \tag{37}$$

Figure 7 shows an example of saliency map obtained for a typical image acquired from the environment used for the experiments discussed in this chapter. It can be noticed that the most salient region of the image corresponds to an orange ball, followed by dark objects printed on the cardboard boxes that compose the walls of the environment.

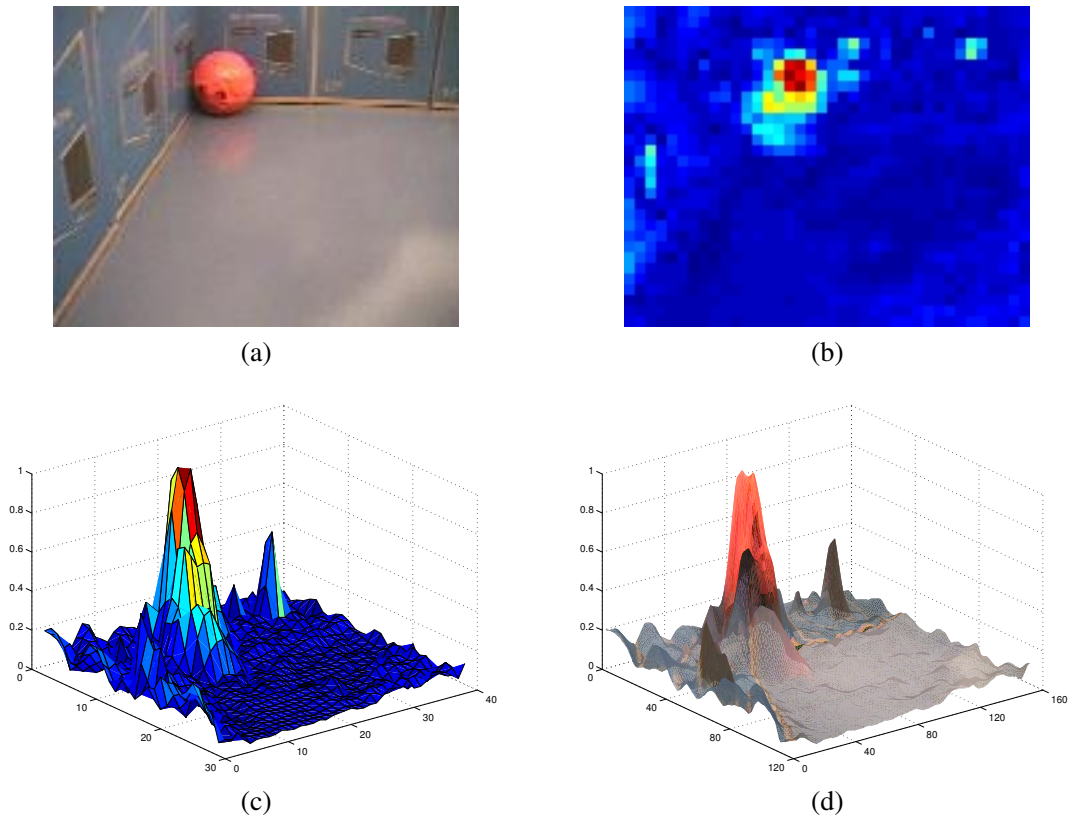


Figure 7: Saliency map example: (a) input image; (b) saliency map highlighting the most salient features in the input image; (c) 3D saliency surface visualisation; (d) 3D saliency warped input image.

The highest values in the saliency map correspond to the most salient locations within the input image, which can also be ranked according to their magnitudes. More precise localisation of local maxima in the saliency map may be obtained by interpolation to subpixel accuracy using a second order Taylor expansion (Vieira Neto, 2006).

A summary of the operations performed to compute the saliency map is given in algorithm 2.

Algorithm 2: Saliency map model of visual attention. Typical parameters: $c_{min} = 2$, $c_{max} = 4$, $\delta_{min} = 3$, $\delta_{max} = 4$, $\sigma_{max} = 8$ and $I_{max} = 255$.

Input: input image colour channels **R**, **G** and **B**.
Output: saliency map **S**.

1 Compute an intensity image from the input image colour channels: $\mathbf{I} = \frac{1}{3}(\mathbf{R}+\mathbf{G}+\mathbf{B})$.

2 Normalise **R**, **G** and **B** colour channels by **I** in order to decouple hue from intensity: $r = \frac{R}{I}$, $g = \frac{G}{I}$, $b = \frac{B}{I}$. Normalisation is applied only at locations where $I > \frac{I_{max}}{10}$.

3 Compute four broadly-tuned colour channels: $\mathbf{R} = \mathbf{r} - \frac{1}{2}(\mathbf{g}+\mathbf{b})$ for red, $\mathbf{G} = \mathbf{g} - \frac{1}{2}(\mathbf{r}+\mathbf{b})$ for green, $\mathbf{B} = \mathbf{b} - \frac{1}{2}(\mathbf{r}+\mathbf{g})$ for blue and $\mathbf{Y} = -2\mathbf{B} - |\mathbf{r}-\mathbf{g}|$ for yellow. Negative values are set to zero.

4 Construct a Gaussian pyramid $\mathbf{I}(\sigma)$ from **I**, with scales $\sigma \in [0, \sigma_{max}]$.

5 Construct four Gaussian pyramids $\mathbf{R}(\sigma)$, $\mathbf{G}(\sigma)$, $\mathbf{B}(\sigma)$ and $\mathbf{Y}(\sigma)$ from **R**, **G**, **B** and **Y**.

6 Construct four oriented Gabor pyramids $\mathbf{O}(\sigma,\theta)$ from $\mathbf{I}(\sigma)$, with preferred orientations $\theta \in \{0°, 45°, 90°, 135°\}$.

7 Compute a set of centre-surround difference feature maps for the intensity channel: $I(c,s) = |\mathbf{I}(c) \ominus \mathbf{I}(s)|$, with $c \in [c_{min}, c_{max}]$ and $s = c + \delta$, $\delta \in [\delta_{min}, \delta_{max}]$.

8 Compute two sets of centre-surround difference feature maps for double-opponent colour channels: $\mathcal{RG}(c,s) = |(\mathbf{R}(c) - \mathbf{G}(c)) \ominus (\mathbf{R}(s) - \mathbf{G}(s))|$ and $\mathcal{BY}(c,s) = |(\mathbf{B}(c) - \mathbf{Y}(c)) \ominus (\mathbf{B}(s) - \mathbf{Y}(s))|$.

9 Compute four sets of centre-surround difference feature maps for the orientation channels: $O(c,s,\theta) = |\mathbf{O}(c,\theta) \ominus \mathbf{O}(s,\theta)|$.

10 Compute the conspicuity map for intensity from across-scale combinations of the set of intensity feature maps: $\bar{I} = \bigoplus_{c=c_{min}}^{c_{max}} \bigoplus_{s=c+\delta_{min}}^{c+\delta_{max}} \mathcal{N}(I(c,s))$, where $\mathcal{N}(.)$ is a normalisation operator.

11 Compute the conspicuity map for colour from across-scale combinations of the two sets of double-opponent colour feature maps: $\bar{C} = \bigoplus_{c=c_{min}}^{c_{max}} \bigoplus_{s=c+\delta_{min}}^{c+\delta_{max}} [\mathcal{N}(\mathcal{RG}(c,s)) + \mathcal{N}(\mathcal{BY}(c,s))]$.

12 Compute the conspicuity map for orientation from across-scale combinations of the four sets of orientation feature maps: $\bar{O} = \sum_{\theta \in \{0°,45°,90°,135°\}} \mathcal{N}\left(\bigoplus_{c=c_{min}}^{c_{max}} \bigoplus_{s=c+\delta_{min}}^{c+\delta_{max}} \mathcal{N}(O(c,s,\theta)) \right)$.

13 Compute the saliency map from the conspicuity maps for intensity, colour and orientation: $\mathcal{S} = \frac{1}{3}(\mathcal{N}(\bar{I}) + \mathcal{N}(\bar{C}) + \mathcal{N}(\bar{O}))$.

14 Search the saliency map for maxima (salient locations).

It should be noted that the final saliency map can be biased by giving a higher weight for any particular feature in equation 37. By doing so, one can easily render colour features more salient than intensity or orientation features, for example. More weight can also be given to a particular scale or orientation in equations 34, 35 or 36. This is an important detail because it makes possible to use top-down biasing if some *a priori* information is available about the features of interest for a given application. For instance, if it is known beforehand that blue vertical lines are important to be detected in a certain exploration or inspection task, the saliency map can be easily biased and give more weight to the relevant feature maps (blue/yellow double-opponent colour channel and 90° orientation channel). The saliency map architecture also offers flexibility to be extended in order to include other visual features, such as flicker and motion (Dhavale & Itti, 2003), if the application requires so.

6 Local Image Descriptors and Visual Novelty Detection

Although statistical descriptors of visual features can be powerful in many applications, their use in a global fashion weakens the ability to capture and represent small details present in the visual field. Statistical representations in general tend to dilute the contribution of features that appear less frequently in the sea of more common features. Visual features that occupy small areas relative to the size of the entire image will have small contributions to a global statistical descriptor – such small features would be probably disregarded by higher levels of processing as if they were noise.

As small details are often relevant for novelty detection tasks, global representations in general seem not to be a good solution for the image description problem. Furthermore, a novelty filter that highlights *when* an image frame contains some novel visual features is not as useful as a novelty filter that also locates *where* these novel features are within the image frame, something that is plausible when local image descriptors are used.

The experiment to be presented now uses colour statistics at selected locations rather than global colour statistics, which means obtaining several image descriptors from different regions of the image. The approach followed here is to use the saliency map model to select a fixed number of salient regions in order to encode multiple local feature vectors per image frame, using the same image descriptors based on colour angles and the novelty filter based on the GWR neural network, as was done in the experiment involving global image descriptors.

A saliency map was computed for each input frame and the nine most salient points found were used to establish the centre of image patches of 24×24 pixels in size. Each of the selected image patches had corresponding image descriptors based on colour angles and normalised intensity standard deviation computed, which were individually fed to a GWR neural network with the same parameters as before. The same two-stage experimental procedure involving exploration and inspection phases that was adopted in the earlier experiment was followed once more.

Using this approach, the GWR network completely habituated on the environment after the fourth exploration loop, acquiring 21 nodes to represent its model of normality. From this information, it can be noticed that the environment was represented in much more detail when using local descriptors – the size of the trained network is more than three times larger than previously and the time necessary for proper training was doubled (see section 4).

The novelty graphs referring to the exploration phase using local image descriptors are not shown here, but rather the ones referring to two inspection trials, one of them involving the same orange ball as before (figure 8) and also another one involving an inconspicuous grey box (figure 9). Because nine feature vectors were generated and classified for each input frame, the novelty graphs for qualitative assessment of results had to be adapted – for experiments using local image descriptors, the novelty graphs depict the average measurement of novelty given by the GWR network in each frame.

As can be visualised in figure 8, the use of local descriptors allows correct detection of the image frames in which the orange ball was present, as in the previous experiment using global image descriptors. Very few false novelties were detected, while the image frames in which the ball appears were clearly identified as having high levels of novelty – this is indicated by the novelty graphs at the left. Moreover, because a local image encoding procedure is now in use with a selective attention mechanism, it is also possible to generate output images indicating which features of the environment are considered salient and, among those, which are considered novel. The use of local descriptors made the localisation of novel features possible, as illustrated with the four output image frames at the right, in which the orange ball was largely present within the camera's field of view during the first inspection loop. In the image frames shown in figure 8, the numbers indicate the most salient locations in ascending order (0 corresponds to the most salient location), whose corresponding image patches were highlighted with white circles when considered to be encompassing novel visual features.

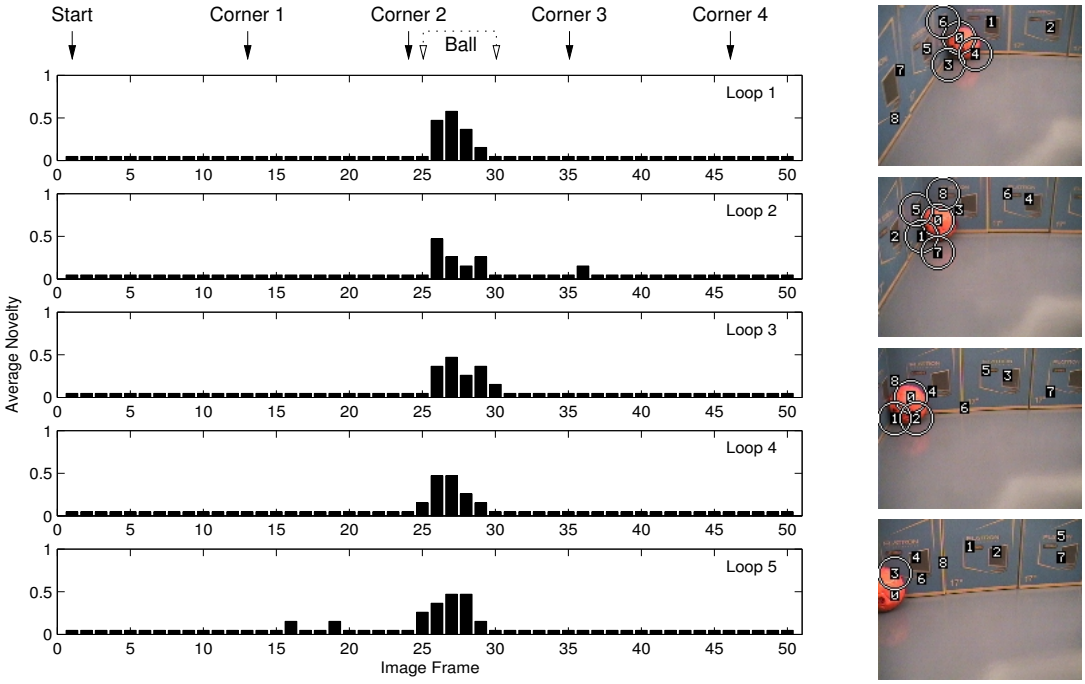


Figure 8: Inspection of the arena with an orange ball as novel object, using local colour angles as image encoding scheme. Locations where the ball was in the camera's field of view are indicated by dotted arrows on the top of the figure. The graphs at the left show that the novel stimulus is correctly and consistently identified by the system in every inspection loop. The images at the right show the corresponding frames in the first loop with circles highlighting the regions considered as novel.

A second inspection trial involved a grey box which is much less conspicuous than the orange ball used previously. In this second trial, the orange ball was removed from the arena and then the grey box was placed in a different corner of the arena than the one where the ball used to be.

The novelty graphs at the left of figure 9 show that the novelty filter was also able to determine consistently which image frames contained the novel features introduced in the environment for this new inspection round, *i.e.* the grey box. The frames in which the grey box appeared in the camera's field of view (after the robot turned the first corner of the arena) are indicated with dotted arrows.

The grey box was clearly identified, as shown by the four output image frames at the right of figure 9, which were generated during the first inspection loop. Very few unexpected (*i.e.* false) indications of novelty occurred in other places of the environment.

The impact of using a local approach for image encoding was clearly positive from a qualitative perspective, as the mechanism of attention offers a clear contribution to the efficient representation of visual data by splitting a relatively large image frame into several small image patches with high information contents (salient regions). In the next sections, an alternative approach to perform bottom-up local image representation and novelty detection simultaneously in a single algorithm based on incremental PCA is considered, discarding the need of specialised top-down image descriptors like the one based in colour angular indexing that was used until this moment.

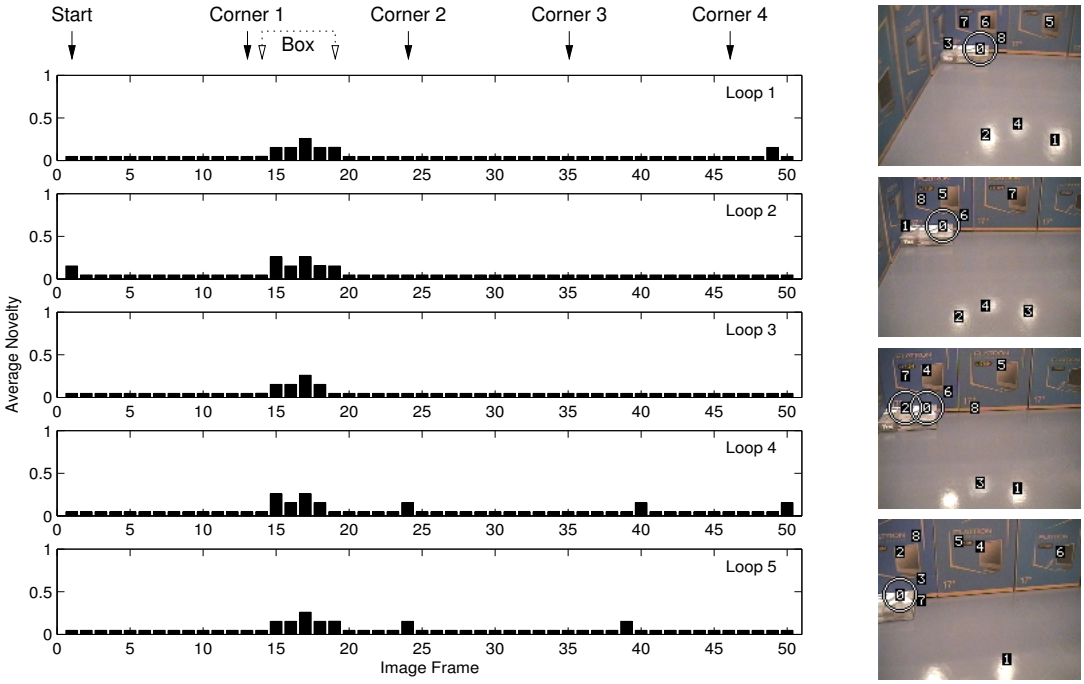


Figure 9: Inspection of the arena with a grey box as novel object, using local colour angles as image encoding scheme. Locations where the box was in the camera's field of view are indicated by dotted arrows on the top of the figure. The graphs at the left show that the novel stimulus is correctly and consistently identified by the system in every inspection loop. The images at the right show the corresponding frames in the first loop with circles highlighting the regions considered as novel.

7 Incremental Principal Component Analysis

Principal Component Analysis can be used as a method for dimensionality reduction in which the input data is projected onto principal axes – the axes along which the variance of the input distribution is maximised. This operation is reversible and allows optimal reconstruction of the original input data.

Standard PCA consists in solving an eigensystem for the covariance matrix $\mathbf{C}$ of normalised input vectors:

$$\mathbf{C}\mathbf{U} = \mathbf{U}\Lambda, \tag{38}$$

$$\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^{\mathsf{T}}, \tag{39}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \tag{40}$$

where n is the number of input vectors $\mathbf{x}_i$, μ is the mean vector, $\mathbf{U}$ contains the eigenvectors and Λ contains their corresponding eigenvalues.

The eigenvectors that correspond to non-zero eigenvalues span a subspace of up to the original m dimensions of the input vectors, but in order to achieve dimensionality reduction of the input data, only eigenvectors corresponding to the arbitrarily chosen $k < m$ largest eigenvalues are selected to be included in a reduced eigenmodel. The original m-dimensional input vectors can therefore be projected onto the k-dimensional subspace spanned by this reduced eigenmodel:

$$\mathbf{a}_i = \mathbf{U}^\mathsf{T}(\mathbf{x}_i - \mu), \tag{41}$$

where $\mathbf{a}_i$ are the projected vectors.

The process can be reversed, reconstructing the input vectors from the projected vectors with minimal squared error:

$$\mathbf{y}_i = \mathbf{U}\mathbf{a}_i + \mu, \tag{42}$$

where $\mathbf{y}_i$ are the reconstructed input vectors.

Standard PCA requires that all data samples are available *a priori* for batch processing, but incremental computation of PCA is also possible (Skočaj & Leonardis, 2003; Artač et al., 2002; Hall et al., 1998), making it suitable for applications that demand on-line learning. This work is based on the incremental PCA algorithm originally proposed in (Hall et al., 1998), which assumes that an initial eigenmodel – composed by a mean vector μ and an eigenvector set $\mathbf{U}$ – is already available. The algorithm discussed here also includes a set of projected vectors $\mathbf{A}$ in the eigenmodel, in order to enable simultaneous learning and recognition (Artač et al., 2002).

When a new input vector $\mathbf{x}$ is available, the set of eigenvectors is updated by appending a new orthogonal basis vector and then applying a rotational transformation (Hall et al., 1998). The new basis vector is obtained by projecting the new input vector onto the current eigenspace (equation 41) and using its reconstruction (equation 42) to compute the residual vector $\mathbf{r} = \mathbf{x} - \mathbf{y}$.

The normalised residual vector is orthogonal to the current eigenspace and therefore is a natural choice for the new basis vector:

$$\mathbf{U}' = \begin{bmatrix} \mathbf{U} & \frac{\mathbf{r}}{\|\mathbf{r}\|} \end{bmatrix}. \tag{43}$$

A major contribution made in (Artač et al., 2002) was to allow projected vectors to be stored and updated, so that the original input vectors can be discarded. Adding a new eigenvector and a new projected vector to the eigenmodel results in increasing the dimensionality of the appended set of projected vectors:

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A} & \mathbf{a} \\ \mathbf{z} & \|\mathbf{r}\| \end{bmatrix}, \tag{44}$$

where $\mathbf{z}$ is a row vector of zeroes.

Performing batch PCA on the appended set of projected vectors $\mathbf{A}'$ yields a mean vector η and a rotation matrix $\mathbf{R}$ that will be used to update the eigenspace (Skočaj & Leonardis, 2003):

$$\mathbf{U} = \mathbf{U}'\mathbf{R}, \tag{45}$$

$$\mu \Leftarrow \mu + \mathbf{U}'\eta, \tag{46}$$

$$\mathbf{A} = \mathbf{R}^\mathsf{T}(\mathbf{A}' - \eta\mathbf{o}), \tag{47}$$

where $\mathbf{o}$ is a row vector of ones.

It is possible to discard eigenvectors whose corresponding eigenvalues are below some threshold, which yields dimensionality reduction of the projected data at the cost of some loss of information. A small percentage of the largest eigenvalue is normally used as the threshold to determine which eigenvectors to keep.

The algorithm is made completely incremental by initialising the eigenspace and projected vectors as follows: $\mu = \mathbf{x}_1$, $\mathbf{U} = \mathbf{z}$ and $\mathbf{A} = 0$, where $\mathbf{x}_1$ is the first input vector and $\mathbf{z}$ denotes a column vector of zeroes with the dimensionality of the input.

In section 8, incremental PCA is used as an alternative method to perform on-line novelty detection. The magnitude of the residual vector – effectively the RMS error between original data and the reconstruction of its

projection onto the current eigenspace – is used to decide if a given input is novel and should be added to the eigenmodel. In practice, if the magnitude of the residual vector is above some threshold r_T, the corresponding input vector is considered not to be well represented by the current eigenmodel and therefore must constitute novelty (Vieira Neto & Nehmzow, 2005).

A summary of the use of the incremental PCA algorithm as a novelty filter is given in algorithm 3.

Algorithm 3: Incremental PCA novelty detection.

Input: current mean vector μ, current eigenvectors $\mathbf{U}$, current projected vectors $\mathbf{A}$, new input vector $\mathbf{x}$, residual threshold r_T.

Output: updated mean vector μ, updated eigenvectors $\mathbf{U}$, updated projected vectors $\mathbf{A}$, novelty indication N.

Compute the projection of the new input vector using the current basis: $\mathbf{a} = \mathbf{U}^\top(\mathbf{x} - \mu)$.
Compute the reconstruction of the new input vector from its projection: $\mathbf{y} = \mathbf{U}\mathbf{a} + \mu$.
Compute the residual vector (orthogonal to $\mathbf{U}$): $\mathbf{r} = \mathbf{x} - \mathbf{y}$.
Test if the magnitude of the residual vector is large enough to characterise novelty:
if $\| \mathbf{r} \| > r_T$ **then**
    Append normalised residual vector: $\mathbf{U}' = \begin{bmatrix} \mathbf{U} & \frac{\mathbf{r}}{\|\mathbf{r}\|} \end{bmatrix}$.
    Append projected new input vector: $\mathbf{A}' = \begin{bmatrix} \mathbf{A} & \mathbf{a} \\ \mathbf{0} & \| \mathbf{r} \| \end{bmatrix}$.
    Perform batch PCA on $\mathbf{A}'$, obtaining its mean vector η and eigenvectors $\mathbf{R}$.
    Update eigenvectors: $\mathbf{U} = \mathbf{U}'\mathbf{R}$.
    Update mean vector: $\mu \Leftarrow \mu + \mathbf{U}'\eta$.
    Update projected vectors: $\mathbf{A} = \mathbf{R}^\top(\mathbf{A}' - \eta\mathbf{o})$, where $\mathbf{o}$ is a row vector of ones.
    Indicate novelty detected: $N = 1$.
else indicate no novelty detected: $N = 0$.

8 Autonomous Image Representation and Visual Novelty Detection

The experiments using colour statistics as image descriptors within the discussed visual novelty detection framework have yielded successful results so far. However, image encoding based solely on colour statistics does not hold enough information to reconstruct the original image. If one examines the weights of the GWR network nodes to analyse which visual aspects of the environment were acquired during learning, the best information that can be retrieved is the relative amount of different colours present in a given region of the environment.

In the experiments to be presented now, an image encoding procedure that allows image reconstruction is exploited, so that examination of the acquired normality model provides valuable *visual* information about which aspects of the environment were actually learnt. In order to achieve this, raw image patches selected by the visual attention mechanism are normalised and used directly as input vectors to the incremental PCA algorithm described in section 7.

Using RGB image patches with 24×24 pixels in size results in input vectors with $24 \times 24 \times 3 = 1728$ dimensions, which were normalised to unit length in order to even out lighting conditions. An input space of such relatively high dimensionality may in fact not be entirely necessary to represent the acquired visual data appropriately. Incremental PCA autonomously provides dimensionality reduction by exploiting the fact that the number of eigenvectors in the acquired model is likely to be less than the number of dimensions of the input vectors. Further reduction in dimensionality can be achieved by keeping only the eigenvectors corresponding

to the largest eigenvalues in the model, at the expense of losses in reconstruction and possibly in the overall recognition rate of the system. If all eigenvectors are kept in the model, perfect reconstruction of the original visual data is achieved.

The experiments of this section use the same two-stage experimental procedure adopted in the earlier experiments, which involves five exploration loops around the environment for the acquisition of the model of normality by the incremental PCA algorithm and five inspection loops for testing the obtained novelty filter. During exploration, the residual error threshold of the incremental PCA algorithm was set to $r_T = 0.25$ and only eigenvectors whose corresponding eigenvalues were larger than 1% of the largest eigenvalue in the model were kept.

Most of the eigenspace updates (indications of novelty) during the exploration phase happened in the beginning of the first loop around the arena, becoming less frequent as the environment was explored. At the end of the learning process, the incremental PCA algorithm acquired 35 model vectors with 33 dimensions, representing a data compression factor of more than 98%. However, a mean vector and 33 eigenvectors with the original 1728 dimensions were also stored in the model.

Figure 10 shows the faithful reconstruction of the 35 model vectors acquired during the exploration of the environment, which provide *visual* feedback of which aspects of the environment were captured.

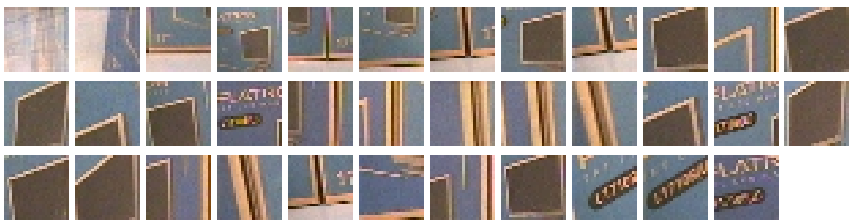


Figure 10: Environmental model acquired by the incremental PCA algorithm for the empty arena. A total of 35 image patches acquired from the environment were faithfully reconstructed.

From the reconstructions in figure 10, it is clear that the model captured a detailed representation of the environment, which mostly consists of inscriptions and objects printed on the cardboard boxes that constitute the walls of the arena, and the edges between these boxes.

The acquired model of normality was then used as novelty filter during the inspection phase of the experiment, which was conducted in two trials, each having a different novel object deliberately placed in the arena – the same orange ball and grey box used in previous experiments were used once again. Learning of the incremental PCA algorithm was disabled during the inspection phase so that the novel objects could be repeatedly detected in different inspection loops, following the same procedure as in previous experiments.

The first trial of five inspection loops around the arena containing the orange ball as the novel object yielded the results shown in figure 11. The novelty graphs at the left show that incremental PCA was able to detect the orange ball consistently in every inspection loop with very few false indications of novelty. The four output image frames at the right show the situations during the first inspection loop in which regions containing parts of the ball were highlighted as having novel features.

A second inspection trial was conducted, now with the orange ball being removed and the orange box being placed in a different corner of the arena – results are shown in figure 12. Once more, the incremental PCA approach was able to detect the novel object correctly and consistently, as shown by the novelty graphs at the left. There were very few spurious novelty indications, particularly at the end of the fourth and fifth inspection loops. At the right, output image frames show four of the five situations during the first loop in which regions partially containing the novel features were correctly highlighted with white circles.

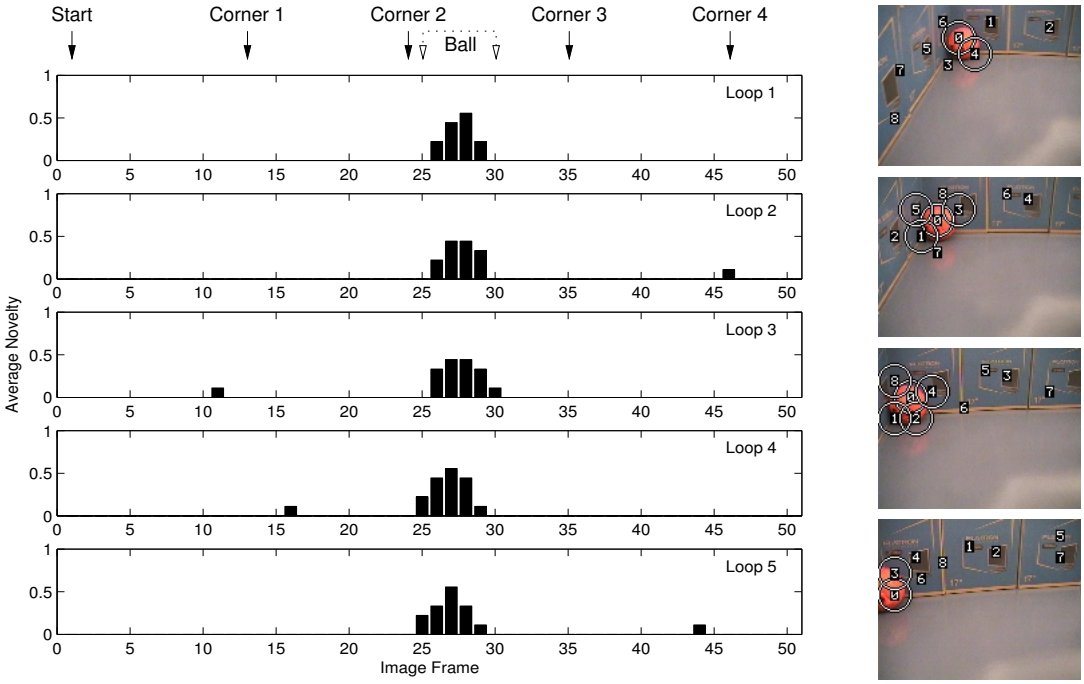


Figure 11: Inspection of the arena with an orange ball as novel object, using incremental PCA as image encoding and novelty detection scheme. Locations where the ball was in the camera's field of view are indicated by dotted arrows on the top of the figure. The graphs at the left show that the novel stimulus is correctly and consistently identified by the system in every inspection loop. The images at the right show the corresponding frames in the first loop with circles highlighting the regions considered as novel.

A final experiment was conducted, in which the robot once again explored the arena, this time with the conspicuous orange ball already present in one of its corners, and inspected it afterwards with the inclusion of the inconspicuous grey box next to the ball, *i.e.* in the same corner of the arena. The incremental PCA algorithm acquired 45 model vectors with 32 dimensions after this new exploration phase. The image reconstructions obtained are shown in figure 13, clearly illustrating that features from the orange ball were included in the new model of normality.

The fact that the grey box was placed next to the ball during the inspection phase obviously affects the response of the attention mechanism because the two objects of interest are present at the same time in some of the image frames, competing for saliency. However, the grey box was correctly identified as the novel object even in the presence of the more conspicuous but already known orange ball, which was ignored by the novelty filter. Figure 14 shows two examples of output image frames that illustrate the response of the novelty filter, one of them with salient region 0 containing novel features that were actually missed by the system.

The experiments in this section show that novelty filters based on incremental PCA are an interesting alternative to novelty filters based on the GWR neural network. Also, the use of raw image patches as input to the novelty filter allows faithful image reconstruction from the vector models acquired, adding the extra functionality of *visual* feedback of which aspects of the environment were actually learnt during the exploration phase.

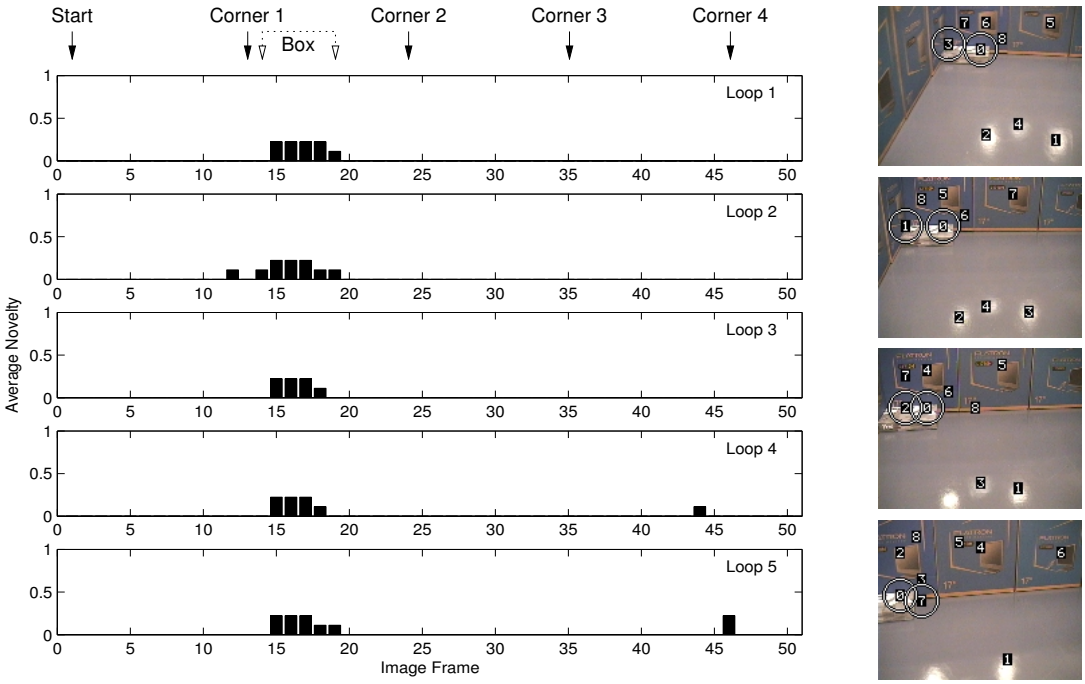


Figure 12: Inspection of the arena with a grey box as novel object, using incremental PCA as image encoding and novelty detection scheme. Locations where the box was in the camera's field of view are indicated by dotted arrows on the top of the figure. The graphs at the left show that the novel stimulus is correctly and consistently identified by the system in every inspection loop. The images at the right show the corresponding frames in the first loop with circles highlighting the regions considered as novel.

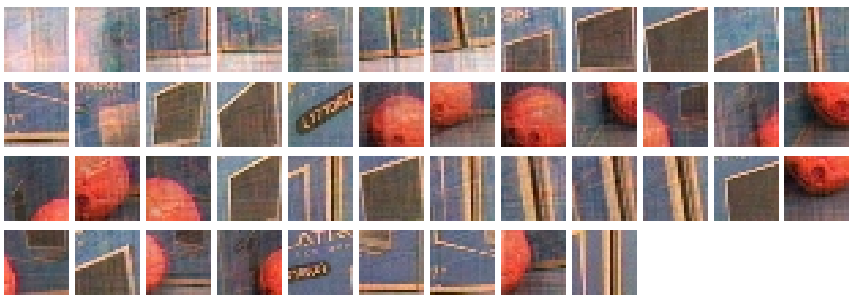
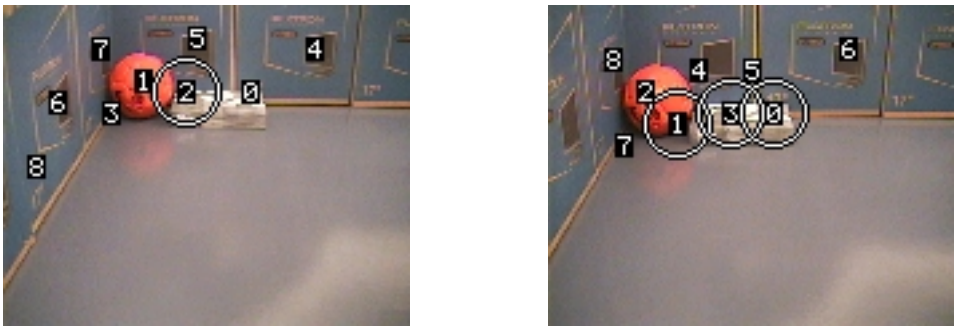


Figure 13: Environmental model acquired by the incremental PCA algorithm for the arena containing the orange ball. A total of 45 image patches acquired from the environment were reconstructed with some deterioration due to the inclusion of the orange ball.



(a) (b)

Figure 14: Examples of output images of the new inspection trial. In (a) and (b) the grey box is correctly identified as being the novel object in spite of the presence of the much more conspicuous but already known orange ball.

9 Conclusion

The ability to differentiate between common and uncommon stimuli is essential to robots operating in dynamic environments and is at the core of applications involving automated exploration and inspection. Because novelty is of contextual nature, and therefore can not be easily modelled, the most feasible approach to be followed is to first acquire a model of *normality* through robot learning via unsupervised clustering mechanisms and then use it as a means to highlight *any abnormal* features that may appear in the operating environment. This approach was used previously in mobile robots using low-resolution sensor modalities such as sonar sensing.

A comprehensive on-line visual novelty detection framework that can be used in autonomous mobile robots for exploration and inspection tasks was presented in this chapter. The use of novelty filters and global image descriptors to detect *when* novel features enter the field of view of the robot was described and discussed, as well as the use of a visual attention mechanism and local image descriptors to detect *where* novel features are within the image frame. The unsupervised clustering mechanisms used as novelty filters were a GWR neural network operating on image descriptors, or an incremental PCA algorithm operating directly on raw image patches, which had their performances evaluated in several experimental examples.

Experiments conducted in engineered environments with a physical mobile robot have demonstrated that the proposed visual novelty detection framework has the ability to highlight and locate new, *arbitrary* objects as soon as they first appear in the field of view of the robot's camera. The on-line unsupervised clustering mechanisms used were able to learn adequate representations of the robot's normal operating environment very efficiently. When using raw image patches as input to a novelty filter based on the incremental PCA algorithm, it was also possible to have *visual* feedback of the aspects learnt from the environment through image reconstruction.

For further experimental results obtained in different environmental conditions, including analysis and discussion of the influence of the robot's trajectory on the performance of the visual novelty detection framework presented in this chapter, the reader is referred to (Vieira Neto, 2006).

Acknowledgement

This chapter is dedicated to the memory of Prof. Ulrich Nehmzow.

References

Artač, M., Jogan, M., & Leonardis, A. (2002). Incremental PCA for on-line visual learning and recognition. In *Proceedings of the 16th International Conference on Pattern Recognition*, volume 3 (pp. 781–784). Quebec, Canada.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346–359.

Burt, P. J. & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transact ons on Communications*, 31(4), 532–540.

Crook, P. A. & Hayes, G. (2001). A robot implementation of a biologically inspired method for novelty detection. In *Towards Intelligent Mobile Robots*.

Crook, P. A., Marsland, S., Hayes, G., & Nehmzow, U. (2002). A tale of two filters — on-line novelty detection. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation* (pp. 3894–3899). Washington, DC.

Dhavale, N. & Itti, L. (2003). Saliency-based multi-foveated MPEG compression. In *Proceedings of the 7th IEEE International Symposium on Signal Processing and its Applications* Paris, France.

Finlayson, G. D., Chatterjee, S. S., & Funt, B. V. (1996). Color angular indexing. In *Proceedings of the 4th European Conference in Computer Vision* (pp. 16–27). Cambridge, UK.

Greenspan, H., Belongie, S., Goodman, R., Perona, P., Rakshit, S., & Anderson, C. H. (1994). Overcomplete steerable pyramid filters and rotation invariance. In *Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 222–228).

Hall, P. M., Marshall, D., & Martin, R. R. (1998). Incremental eigenanalysis for classification. In *Proceedings of the 9th British Machine Vision Conference* (pp. 286–295).

Hodge, V. J. & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2), 85–126.

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

Kadir, T. & Brady, M. (2003). Scale saliency: A novel approach to salient feature and scale selection. In *Proceedings of the International Conference on Visual Information Engineering* (pp. 25–28).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.

Markou, M. & Singh, S. (2003a). Novelty detection: A review - part 1: Statistical approaches. *Signal Processing*, 83, 2481–2497.

Markou, M. & Singh, S. (2003b). Novelty detection: A review - part 2: Neural network based approaches. *Signal Processing*, 83, 2499–2521.

Marsland, S. (2003). Novelty detection in learning systems. *Neural Computing Surveys*, 3, 157–195.

Marsland, S., Nehmzow, U., & Shapiro, J. (2001). Vision-based environmental novelty detection on a mobile robot. In *Proceedings of the International Conference on Neural Information Processing* Shanghai, China.

Marsland, S., Nehmzow, U., & Shapiro, J. (2002a). Environment-specific novelty detection. In *From Animals to Animats: Proceedings of the 7th International Conference on the Simulation of Adaptive Behaviour* Edinburgh, UK: MIT Press.

Marsland, S., Shapiro, J., & Nehmzow, U. (2002b). A self-organising network that grows when required. *Neural Networks*, 15(8-9), 1041–1058.

Mikolajczyk, K. & Schmid, C. (2004). Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1), 63–86.

Skočaj, D. & Leonardis, A. (2003). Weighted and robust incremental method for subspace learning. In *Proceedings of the 9th International Conference on Computer Vision* (pp. 1494–1501).

Swain, M. J. & Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, 7(11), 11–32.

Tarassenko, L., Hayton, P., Cerneaz, N., & Brady, M. (1995). Novelty detection for the identification of masses in mammograms. In *Proceedings of the 4th IEE International Conference on Artificial Neural Networks* (pp. 442–447).

Vieira Neto, H. (2006). *Visual Novelty Detection for Autonomous Inspection Robots*. PhD thesis, University of Essex, Colchester, UK.

Vieira Neto, H. & Nehmzow, U. (2004). Visual novelty detection for inspection tasks using mobile robots. In *Proceedings of the 8th Brazilian Symposium on Neural Networks* São Luís, Brazil.

Vieira Neto, H. & Nehmzow, U. (2005). Automated exploration and inspection: Comparing two visual novelty detectors. *International Journal of Advanced Robotic Systems*, 2(4), 355–362.