Real-time Automated Visual Inspection using Mobile Robots *

Hugo Vieira Neto
Department of Electronics
Federal University of Technology - Paraná
Avenida Sete de Setembro 3165
Curitiba-PR 80230-901
Brazil
hvieir@utfpr.edu.br

Ulrich Nehmzow
Department of Computer Science
University of Essex
Wivenhoe Park
Colchester CO4 3SQ
United Kingdom
udfn@essex.ac.uk

Abstract

We present a framework to perform novelty detection using visual input in which a mobile robot first learns a model of normality in its operating environment and later uses this to highlight uncommon visual features that may appear. This ability is of great importance for both robotic exploration and inspection tasks, because it enables the robot to allocate computational and attentional resources efficiently to those features which are novel.

At the heart of the proposed system is the image encoding mechanism which uses local colour statistics from regions selected by a biologically-inspired model of visual attention. Our approach works in real-time with a wide, unrestricted field of view and is robust to image transformations.

Experiments conducted in an engineered scenario demonstrate the efficiency and functionality of our method.

1   Introduction

The ability to distinguish between common and unusual sensory perceptions — novelty detection — is very useful for mobile robots operating in dynamic environments. Such a competence allows the robot to select those aspects of the surroundings that have an abnormal nature and therefore deserve closer attention from either a human operator, as it is the case for supervised surveillance or inspection tasks, or its own computational resources for further processing in other applications, such as environment exploration and mapping.

Novelty detection has been used in diverse problems, from medical diagnosis of masses in mammograms (Tarassenko et al., 1995) to fault monitoring and detection (Taylor and McIntyre, 1998). These problems share a common characteristic, which is the existence of large amounts of data in which the result of the test is negative (no disease diagnosed or no fault detected), whereas relatively few examples of the important features that are desired to be detected. It is therefore usually not possible to install or learn models of *abnormality*, because too little training data is available. Instead, a model of *normality* is acquired on-line and used to filter out *any* input stimuli that do not fit the learnt model (Marsland, 2003).

Either statistical approaches (Markou and Singh, 2003a) or neural networks (Markou and Singh, 2003b) can be used to implement novelty detection systems. In this work we use the self-organisation of a Grow-When-Required (GWR) neural network (Marsland et al., 2002b) to learn a model of the normal operating environment of a mobile robot and then use this model to highlight any anomalous perception that may appear.

The GWR network is capable of unsupervised and continuous (on-line) learning and has previously been successfully used in mobile robots having sonar readings as percep-

tual input (Marsland et al., 2002a). However, sonar sensors provide very low resolution sampling of the environment, posing serious limitations for real world applications. Therefore, we are interested in using the GWR network with visual information instead of sonar readings.

In this paper we describe a method to process colour vision using a model of visual attention and local colour statistics as means to generate feature vectors to the GWR-based novelty filter. Figure 1 shows a block diagram of the framework developed for our visual novelty detection mechanism.
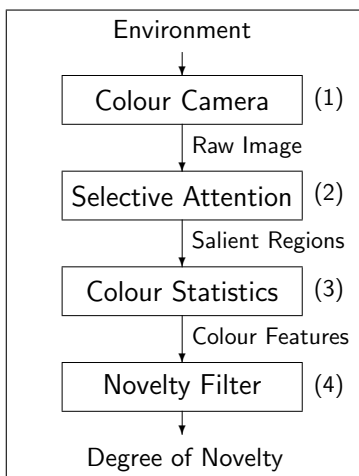

Figure 1: The proposed visual novelty detection mechanism: local colour features are computed at salient regions of the input image, whose degrees of novelty are determined by an artificial neural network.

The paper is structured as follows: first an account of related work in visual novelty detection is given, followed by a description of the image encoding procedure and the novelty filter we have used. Finally, we present and discuss laboratory experiments involving visual exploration and inspection of an engineered environment. We conclude that the system is able to discriminate new, *arbitrary* colour features and set the scene for future improvements using other visual features in addition to colour.

2   Related work

Previous work in novelty detection using visual input and mobile robots was conducted by Marsland *et al.* (Marsland et al., 2001). They have used a wall-following behaviour and a monochromatic camera positioned to acquire close-up images of the wall being followed.

This approach implicitly constrained the visual input to a very limited field of view, also limiting its usefulness to more general applications, although it was capable of on-line novelty detection through a GWR network.

Diehl and Hampshire have implemented a real-time novelty detection mechanism for video surveillance (Diehl and Hampshire II, 2002). Their method was based in the extraction of monochromatic spatial features in image sequences to represent moving objects. A classifier based on support vector learning was trained off-line with models of people and cars and was later used to reject different moving objects, such as bicycles, vans and trucks.

Recently, a new framework for novelty detection applied to region-segmented outdoor scenes in video sequences was proposed by Singh and Markou (Singh and Markou, 2004). Their approach made use of a feature-selection mechanism for various visual features to encode image regions and a feedforward neural network as a classifier. The classifier was used to reject any input not present in the training dataset so that the network could be retrained off-line.

In this paper we present an on-line novelty detection framework for visual input acquired from a moving platform, but limited to the use of colour statistics to encode visual features. Our approach contrasts with the aforementioned in the sense that it combines both unsupervised continuous learning and colour vision with no explicit image segmentation or restrictions in field of view.

3   Image encoding

The implementation of computer vision algorithms in mobile robots is a non-trivial issue, as one normally desires to process large amounts of data in real-time but with limited computational resources. Moreover, acquiring images from a moving platform makes visual features subject to numerous transformations, such as scaling, translation, rotation, changes in perspective and perceived illumination and also partial occlusions. Therefore, we are looking for a fast image encoding technique to provide compact and robust feature vectors for higher levels of processing.

A natural approach to overcome limitations posed by processing speed is to simplify the image encoding. Unfortunately, simplifying too much the image encoding mechanism also limits its robustness to the various transformations that images acquired from a mobile robot are susceptible to. Here we present an image encoding procedure (blocks 2 and 3 in Fig. 1) that uses local colour statistics from salient regions within the image frame. This approach has shown to work efficiently in real-time and to be robust to several image transformations.

3.1 Selective attention

In order to select which regions of the image contain useful information and hence deserve to be analysed in detail, we have used the saliency map model for visual attention (Itti et al., 1998), which is inspired by the neural architecture of the early primate visual system and is in agreement with the Feature-Integration Theory of Attention (Treisman and Gelade, 1980). Further processing to extract local visual features is therefore directed only to the selected regions.

Figure 2 presents the simplified architecture of the saliency map model, which consists in the construction and combination of multi-scale feature maps that allow the detection of local discontinuities in intensity, colour and orientation.
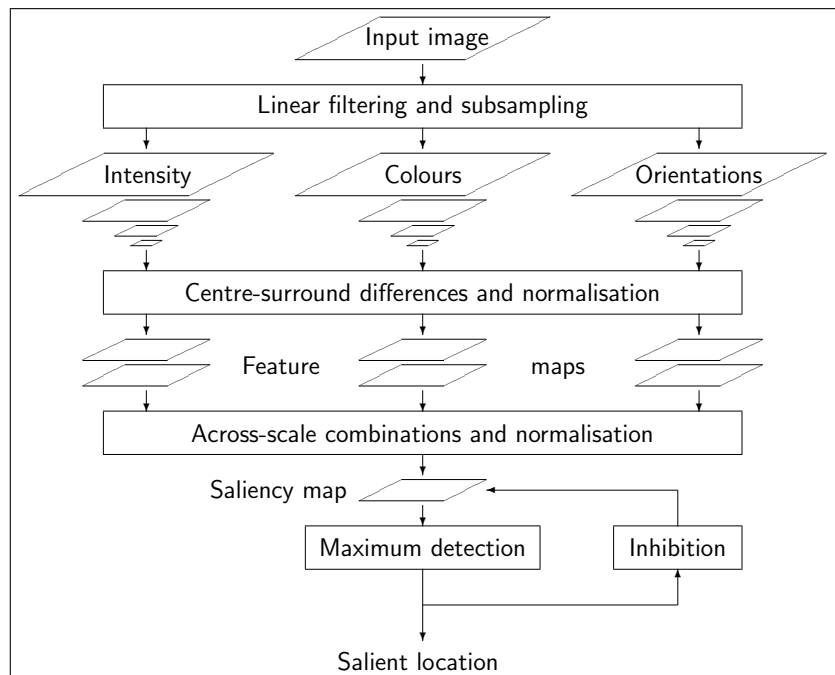

Figure 2: The saliency map model architecture.

The feature maps are computed from pyramidal structures (Greenspan et al., 1994), which are obtained by successive low-pass filtering and subsampling of the preprocessed input image (152×120 pixels in size). The resulting pyramids consist of five levels with reduction factors ranging from 1:1 (scale 0) to 1:16 (scale 4).

The first step in the extraction of early visual features is to obtain an intensity channel $I = (r + g + b)/3$ from the original red (r), green (g) and blue (b) channels of the input image. After that, intensity normalised channels $\hat{r}$, $\hat{g}$ and $\hat{b}$ 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 (in other locations $\hat{r}$, $\hat{g}$ and $\hat{b}$ are set to zero). Four broadly tuned colour channels are then computed: $R = \hat{r} - (\hat{g} + \hat{b})/2$ for red, $G = \hat{g} - (\hat{r} + \hat{b})/2$ for green, $B = \hat{b} - (\hat{r} + \hat{g})/2$ for blue and $Y = -2(B + |\hat{r} - \hat{g}|)$ for yellow, with negative values set to zero.

The intensity channel I and the broadly tuned colour channels R, G, B and Y are used to construct the Gaussian pyramids $I(\sigma)$, $R(\sigma)$, $G(\sigma)$, $B(\sigma)$ and $Y(\sigma)$, respectively. I is also used to construct oriented Gabor pyramids $O(\sigma, \theta)$ in the same manner as described in (Greenspan et al., 1994). We have used five scales ($\sigma \in \{0 \ldots 4\}$) and 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 the feature maps and are implemented as the difference between a fine centre scale and a coarser surround scale. Across-scale difference between scales (denoted by $\ominus$) is obtained by interpolation from the coarse scale to the finer scale and subsequent pixelwise subtraction. In our implementation, the centre consist of pixels at scale $c \in \{1, 2\}$ and the surround of the corresponding pixels at scale $s = c + 2$.

The first type of feature map is related to 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:

$$\mathcal{I}(c, s) = |I(c) \ominus I(s)|. \tag{1}$$

The second type of feature map is concerned with 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. Therefore colour feature maps are computed for red/green and blue/yellow double-opponent pairs as follows:

$$\mathcal{RG}(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))|, \tag{2}$$

$$\mathcal{BY}(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))|. \tag{3}$$

Finally, the third type of feature map accounts for local orientation contrast between centre and surround scales. They are computed separately for every orientation, as shown below:

$$\mathcal{O}(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)|. \tag{4}$$

In order to combine feature maps with different dynamic ranges into a single saliency map it is necessary to use a normalisation operator $\mathcal{N}(.)$, 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 original normalisation operator suggested in (Itti et al., 1998) involves searching for local maxima, but here we use a simpler way of normalising feature maps that yields comparable results: first we subtract the average value from the feature map F; second, we divide the resulting zero-mean feature map F_0 by its vector norm $\|F_0\|$; and finally we take the absolute value of the result.

The feature maps are combined in three conspicuity maps at scale $\sigma = 2$, one for each feature: intensity $\bar{\mathcal{I}}$, colour $\bar{\mathcal{C}}$ and orientation $\bar{\mathcal{O}}$. The conspicuity maps are obtained by computing across-scale addition (denoted by $\oplus$), which consists of resampling each map to scale 2 and subsequent pixelwise addition:

$$\bar{\mathcal{I}} = \bigoplus_{c=1}^{2} \mathcal{N}(\mathcal{I}(c, s)), \tag{5}$$

$$\bar{\mathcal{C}} = \bigoplus_{c=1}^{2} [\mathcal{N}(\mathcal{RG}(c,s)) + \mathcal{N}(\mathcal{BY}(c,s))], \tag{6}$$

$$\bar{\mathcal{O}} = \sum_{\theta \in \{0°,45°,90°,135°\}} \mathcal{N}\left(\bigoplus_{c=1}^{2} \mathcal{N}(\mathcal{O}(c,s,\theta))\right). \tag{7}$$

Finally, the three conspicuity maps are normalised and added to yield the final saliency map $\mathcal{S}$:

$$\mathcal{S} = \mathcal{N}(\bar{\mathcal{I}}) + \mathcal{N}(\bar{\mathcal{C}}) + \mathcal{N}(\bar{\mathcal{O}}). \tag{8}$$

The highest values in the saliency map correspond to the most salient locations within the input image and we use this information to select the regions where further analysis for feature extraction should be performed. After that, we inhibit a circular region of the saliency map centred around the current location and search for the next most salient position to be analysed. The process is repeated until a desired number of salient locations is reached.

Interest points determined using the saliency map are robust to geometric transformations, which contributes to the desired general robustness of the image encoding mechanism. We have used the nine highest salient values to indicate which locations of the image are likely to be the most interesting so that colour statistics could be calculated in their vicinity, as described next.

3.2 Colour statistics

Histograms are well-known statistical tools that can be successfully used to encode general image features and show robustness against geometric transformations, changes in perspective and partial occlusion. In fact, many approaches make use of histograms to characterise visual information, for example Swain and Ballard's colour indexing (Swain and Ballard, 1991) and Schiele and Crowley's multidimensional receptive field histograms (Schiele and Crowley, 2000).

In previous work (Vieira Neto and Nehmzow, 2004) we have studied the use of histograms using the HSI colour space to encode colour information, but found them to be very sensitive to changes in illumination. This shortcoming was also identified by Funt and Finlayson and lead them to improve Swain and Ballard's approach by developing a colour constant colour indexing scheme (Funt and Finlayson, 1995). Finlayson *et al.* have also proposed the colour angular indexing technique (Finlayson et al., 1996), a very compact colour representation that is invariant to illumination conditions.

Here we use the colour angular indexing approach to encode colour features present in the regions selected by the saliency map model (see subsection 3.1). This method encodes the characteristics of the colour distribution of an image as the angles between colour channel vectors in image space. For that, first we have to obtain zero-mean colour channels R_0, G_0 and B_0 by subtracting the average value from each original colour channel. The next step is to normalise the zero-mean colour channels to unitary length by dividing each one by their respective vector norm ($\|R_0\|$, $\|G_0\|$ and $\|G_0\|$). Finally, the three colour angles are computed as the inverse cosine of the dot product between the resulting normalised colour channels R_N, G_N and B_N: $\phi_{rg} = \arccos(\langle R_N, G_N \rangle)$, $\phi_{gb} = \arccos(\langle G_N, B_N \rangle)$ and $\phi_{rb} = \arccos(\langle R_N, B_N \rangle)$.

It should be noted that the angles of a colour distribution cannot discriminate unsaturated colours, i.e. shades of grey whose R, G and B components have the same value. To overcome this situation, we decided to also include intensity statistics in our representation. We have used the normalised standard deviation σ_i of the intensity values (average of the R, G and B components) of the image as a fourth element in our feature vector $\xi = \{\phi_{rg}, \phi_{gb}, \phi_{rb}, \sigma_i\}$.

We have computed the colour statistics described above for image patches of size 24×24 pixels around the nine most salient points identified by the saliency map in the input image. Therefore, for each input image we generated nine local feature vectors using colour statistics to feed into our novelty filter.

4 Novelty filter

The novelty filter used in our framework (block 4 in Fig. 1) is based on the GWR neural network (Marsland et al., 2002b), which is a clustering mechanism derived from Kohonen's self-organising feature map (SOFM) (Kohonen, 1984). However, unlike the SOFM, the GWR network has the ability to add nodes to its structure to represent new input stimuli using a model of habituation.

Figure 3 shows a schematic representation of the GWR network, which consists of a clustering layer of nodes and a single output node. The connecting synapses to the clustering layer are subject to habituation, which is a reduction in response to repeatedly presented stimuli. In other words, the more a node in the clustering layer fires, the less efficient its output synapse becomes.
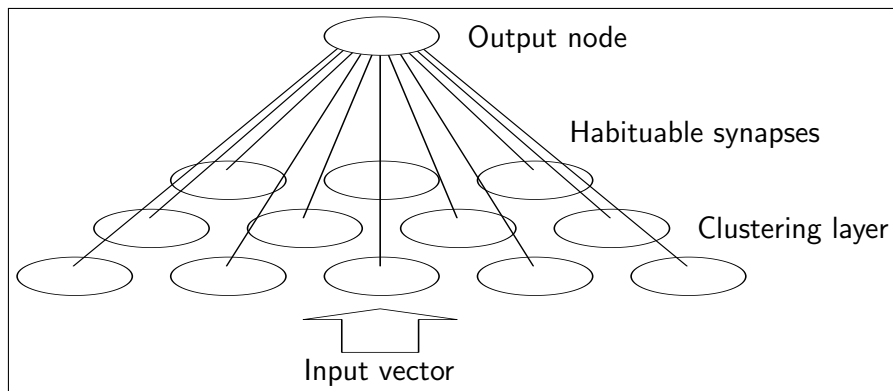

Figure 3: The GWR neural network architecture.

The habituation rule of a node's output synapse is given by the following first-order differential equation:

$$\tau_i \frac{dh_i(t)}{dt} = \alpha[h_0 - h_i(t)] - S(t), \tag{9}$$

where h_0 is the initial value of the habituation $h_i(t)$, $S(t)$ is the external stimulus, τ_i and α are time constants that control the habituation rate and the recovery rate, respectively. As the habituation value has a bounded output, it can be neatly used as a measure of the degree of novelty for any particular input.

The GWR network starts with two nodes in its map $M = \{c_1, c_2\}$, whose weights w_{c_1} and w_{c_2} are initialised randomly from the input distribution $p(\xi)$. There are no topological

connections between the two initial nodes and therefore the connection set is initialised to the empty set.

Training is done with an unsupervised winner-take-all approach, where the winner node (the best match to the input) and its topological neighbours have their weights adapted and their output synapses habituated. The distance from the input ξ to each node in the clustering layer is computed using the Euclidean distance $\|\xi - w_i\|$, where w_i is the weight vector of the node i, with i covering all the nodes in the current map M.

The best matching node is selected as $s = \arg\min_{i \in M} \|\xi - w_i\|$ and the second best as $t = \arg\min_{i \in M-\{s\}} \|\xi - w_i\|$. If there is already a connection between s and t, its age is set to zero, otherwise a new connection is created. The activation value of the winner node is computed as $a_s = \exp(-\|\xi - w_s\|^2)$. Both activation and habituation values of the winner node are used to decide whether an input is novel or not. Therefore, a new node r is added to the map every time that both activation and habituation values are below predefined thresholds a_T and h_T, respectively: $M = M \cup \{r\}$.

In our implementation the new weight vector w_r is set to the input ξ, whereas in the original GWR algorithm (Marsland et al., 2002a) it is set to the average between the winner node's weight vector w_s and the input. After inserting a new node in the map, it is also necessary to update its topological connections by removing the link between the nodes s and t and inserting links between r and s and between r and t. The positions of the winner node and all of its topological neighbours are adapted according to the learning rule $\Delta \mathbf{w}_i = \epsilon_i (\xi - \mathbf{w}_i)$, where ϵ_i is the particular learning rate for each of the nodes concerned (the winner and its neighbours only).

The implementation of the GWR training algorithm used in this work is slightly different from the original presented in (Marsland et al., 2002a), as we have altered the learning and habituation rules for the topological neighbours of the winner node. The original approach used parameters ϵ_i and τ_i for the neighbours, which were just a constant fraction of ϵ_s and τ_s for the winner node. Therefore, ϵ_i and τ_i were completely independent of the distance between neighbour and winner nodes in input space. In our approach, we made the learning and habituation rates of the neighbour nodes proportional to the ratio between the winner node and neighbour activations:

$$\epsilon_i = \frac{\eta a_i}{a_s}\epsilon_s, \tag{10}$$

$$\tau_i = \frac{a_s}{\eta a_i}\tau_s, \tag{11}$$

where a_s and a_i are the activation of the winner and neighbour nodes, respectively, and η is a proportionality factor $(0 < \eta < 1)$.

It can be noticed from Eq. 10 that the neighbour nodes will have their weights adapted to a lesser extent than the winner while Eq. 11 shows that neighbours will habituate in a slower rate than the winner node.

The final step of the GWR training iteration consists of checking for nodes and edges to be removed, i.e. nodes that no longer have any neighbours and edges whose age is greater than a predefined threshold age_{max}.

For the experiments reported here we have used the following parameters: $a_T = 0.9$, $h_T = 0.3$, $\eta = 0.1$, $\epsilon_s = 0.1$, $\tau_s = 3.33$, $\alpha = 1.05$, $h_0 = 1$, $S(t) = 1$ and $age_{max} = 20$ (see Eqs. 9, 10 and 11). The values for the node insertion thresholds a_T and h_T were selected to make sure that new nodes are added for every novel stimulus without the need of a large number of iterations. The low value assigned to the learning rate ϵ_s ensures that nodes are not able to move too much from the location in input space where they were originally placed. It is also important to mention that only habituation is modelled with the use of a constant positive value for $S(t)$, i.e. its reversal — dishabituation — is disabled in our implementation.

5 Experiments

We conducted a set of experiments to test our approach using the colour vision system of *Radix2*, the Magellan Pro robot shown in Fig. 4a. Besides the standard sonar, infra-red and bump sensors, our robot is also equipped with a laser range scanner whose readings were used for controlling the robot's navigation behaviour. Figure 4b also shows the top view of the engineered environment used in the experiments — a simple square arena built

of cardboard boxes in whose corners (numbered from 1 to 4) novel objects were gradually introduced.
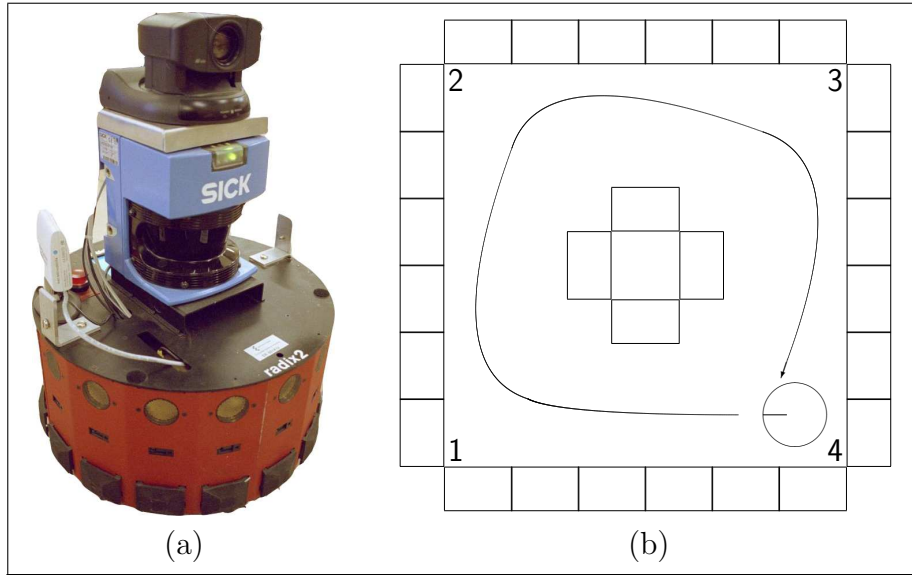

Figure 4: Experimental setup: (a) the mobile robot and (b) the laboratory environment.

The cardboard boxes at the borders of the arena act as walls that limit the robot's trajectory and also its visual world. With the sole intention of obtaining a completely controlled visual world for our experiments, the images were acquired with the robot's camera tilted down to its maximum ($-25°$), so that the field of view was constrained mostly to the arena's walls and floor. In spite of being able to process up to eight frames per second using the robot's onboard computer, the images used in the experiments reported here were acquired at one frame per second without stopping the robot, resulting in 50 frames per loop around the arena.

A simple obstacle-avoidance algorithm using the laser range scanner readings was used as the navigation behaviour for the robot. We have employed the force field strategy, in which every distance measure covering $90°$ in front of the robot is modulated to act as a virtual spring pushing it to the freest space in the environment. Basically, the robot slowly translated at 0.15m/s until it found an obstacle within a threshold distance of 0.5m, which made it stop and then rotate at a maximum speed of $35°/s$ towards free space again. In our experiments, this behaviour has shown to be extremely predictable and stable (the resulting approximate path followed by the robot is shown in Fig. 4b).

5.1 Robot task

Our experiments were designed in order to evaluate the ability of the proposed framework to detect arbitrary novel visual features that may appear in the robot's normal operating environment. Hence, they were divided in two stages: an exploration (learning) phase and an inspection (application) phase.

During the learning phase, images were acquired while the robot was navigating around the empty arena. These images were processed to generate local colour feature vectors as described in section 3 and to train the GWR network. After that, during the application phase, a novel object was introduced in the arena so that a new sequence of images could be acquired and used to test the trained GWR network as a novelty filter. The habituation value of the winner node's output synapse was used as a measure of the degree of novelty of the input.

The expected outcome of such experiments was that the amount of novelty measured would progressively be reduced during exploration as a result of the self-organisation and habituation of the GWR network to represent the original environment. In addition, it was expected that peaks in the novelty measure would appear in areas where the new object was inserted during the inspection phase.

5.2 Results and discussion

Both exploration and inspection phases involved processing images acquired during five loops around the arena. The amount of novelty in each frame was computed as the average of the winner node's habituation value for each of the nine local colour feature vectors generated per frame.

Figure 5 shows the average measure of novelty per frame during the learning phase in five consecutive loops around the empty arena. Notice that the average novelty value reduces as the robot explores the environment. Given the parameters used for the habituation rule in Eq. 9, novelty values range from a minimum of 0.05 and a maximum of 1.

Figure 5 also shows the output of the algorithm for the first image frame acquired
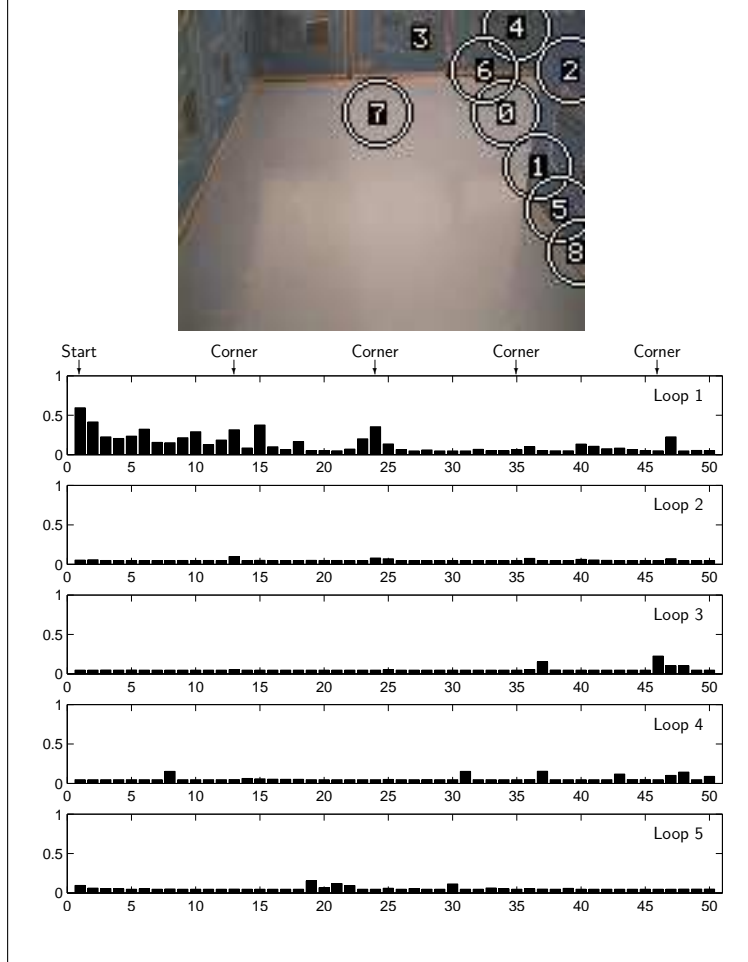
Figure 5: Exploration of the empty arena (learning of the GWR network enabled).

during exploration: the nine most salient locations are assigned with numbers (zero being the most salient), around which white circles are drawn if the surrounding region is considered novel (habituation value of the corresponding winner node above the habituation threshold h_T). As the GWR network has just initiated the learning process in this case, every salient region (with the exception of region 3) is considered as novel.

For the application phase, a new object — an orange ball — was placed in corner 3 of the arena. The amount of novelty measured in five consecutive loops around the arena with the ball is depicted in Fig. 6.

The new object appeared in the camera's field of view immediately after the robot has turned corner 2 and is clearly differentiated from other common visual stimuli observed in the arena (see Fig. 6). In this case, learning of the GWR network was disabled so that it would not habituate on the orange ball and hence detect it as many times as it appeared. Again, the visual output of the system is also shown for an image frame where the orange
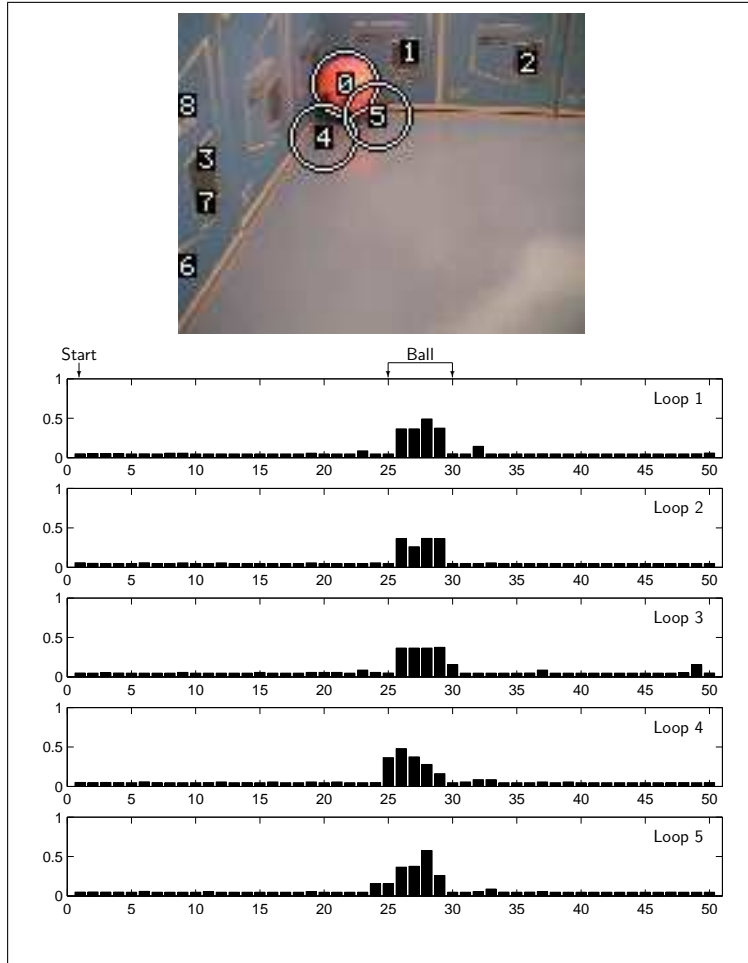
Figure 6: Inspection of the arena with the orange ball (new) in corner 3 (learning of the GWR network disabled).

ball is present, illustrating the ability of our algorithm to localise novel features.

As an extra experiment, learning of the GWR network was *enabled* and the arena with the ball in corner 3 was explored again in order to make the robot habituate on it. Another object was inserted in corner 2 of the arena, this time a grey box much less conspicuous than the orange ball. The arena was then inspected again and the results obtained are shown in Fig. 7.

Examination of Fig. 7 illustrates the ability of the system to localise the newly introduced object (salient locations 0 and 3) while rejecting very similar visual features in the same frame (salient locations 1, 2 and 4, corresponding to reflections of light on the arena's grey floor which share the same colour characteristics with the grey box).

We conducted statistical tests in oder to compare the performance of the system to the expected results. For these, ground truth was generated in the form of binary image
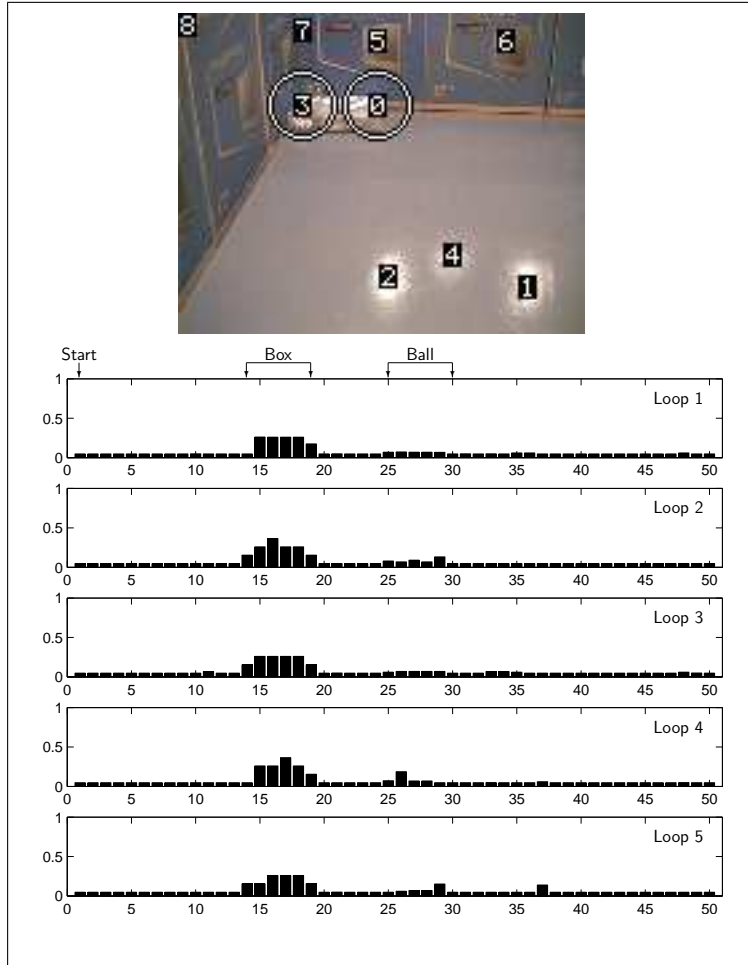
Figure 7: Inspection of the arena with the orange ball (habituated) in corner 3 and the grey box (new) in corner 2 (learning of the GWR network disabled).

masks with the location of the new object manually highlighted. The correlation between the system's response and the actual "novelty situation" was found to be significant (χ^2 test, $p < 0.05$).

6 Conclusion

Motivated by the great potential of a visual novelty-detection mechanism to applications in automated exploration and inspection using mobile robots, we have proposed a framework that uses local colour statistics from salient regions within input image frames to form a sub-symbolic representation of the operating environment without the installation of any *a priori* knowledge.

Colour angles (Finlayson et al., 1996) and the normalised standard deviation of intensity are computed in the vicinity of locations determined by a saliency-based mechanism

of visual attention (Itti et al., 1998) to form feature vectors. The image encoding procedure as a whole has shown to be very stable and robust to various image transformations, such as changes in perspective and perceived illumination. Experiments conducted in an engineered environment with a moving robot have shown that our approach has the ability to highlight new, *arbitrary* objects based on their colour characteristics as soon as they first appear in the camera's field of view.

The novelty filter based on the GWR neural network (Marsland et al., 2002b) was able to learn a representation of the normal environment quickly through our image encoding method, and later was used to detect any unusual features that were introduced after training. Our implementation is capable of on-line learning at eight frames per second when running on a Magellan Pro robot equipped with an 850MHz Pentium III processor. Furthermore, we have compared the performance of our mechanism with manually-generated ground truth and observed statistically significant correlation between system response and actual novelty status.

Ongoing work includes the improvement of the image encoding mechanism, which will incorporate other visual features in addition to colour. As the computation of the saliency map is expensive, we are interested in exploiting information already available from the image pyramids to encode local features. We also intend to implement automatic scale selection (Lindeberg, 1998) and investigate the possibility of using incremental PCA (Artač et al., 2002) to integrate image encoding and on-line novelty detection in a single method. Finally, the system's scalability to real-world environments is also under evaluation.

We consider to have achieved very good results in interfacing unrestricted visual stimuli to the GWR-based novelty filter. Our framework has shown to offer stability and robustness to general image transformations while providing consistent novelty detection based on local colour statistics.

Acknowledgements

References

Matej Artač, Matjaž Jogan, and Aleš Leonardis. Incremental PCA for on-line visual learning and recognition. In *Proc. of ICPR 2002*, 2002.

Christopher P. Diehl and John B. Hampshire II. Real-time object classification and novelty detection for collaborative video surveillance. In *Proc. of IJCNN'02*, pages 2620–2625, 2002.

Graham D. Finlayson, Subho S. Chatterjee, and Brian V. Funt. Color angular indexing. In *Proc. of ECCV'96*, pages 16–27, Cambridge, UK, 1996.

Brian V. Funt and Graham D. Finlayson. Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.

H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C. H. Anderson. Overcomplete steerable pyramid filters and rotation invariance. In *Proc. of CVPR'94*, pages 222–228, 1994.

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

Teuvo Kohonen. *Self-Organization and Associative Memory*. Springer-Verlag, New York, NY, 1984.

Tony Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):194–203, 1998.

Markos Markou and Sameer Singh. Novelty detection: A review - part 1: Statistical approaches. *Signal Processing*, 83:2481–2497, 2003a.

Markos Markou and Sameer Singh. Novelty detection: A review - part 2: Neural network based approaches. *Signal Processing*, 83:2499–2521, 2003b.

Stephen Marsland. Novelty detection in learning systems. *Neural Computing Surveys*, 3:157–195, 2003.

Stephen Marsland, Ulrich Nehmzow, and Jonathan Shapiro. Vision-based environmental novelty detection on a mobile robot. In *Proc. of ICONIP'01*, Shanghai, China, 2001.

Stephen Marsland, Ulrich Nehmzow, and Jonathan Shapiro. Environment-specific novelty detection. In *Proc. of SAB'02*, Edinburgh, UK, 2002a. MIT Press.

Stephen Marsland, Jonathan Shapiro, and Ulrich Nehmzow. A self-organising network that grows when required. *Neural Networks*, 15(8-9):1041–1058, 2002b.

Bernt Schiele and James L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1): 31–50, 2000.

Sameer Singh and Markos Markou. An approach to novelty detection applied to the classification of image regions. *IEEE Trans. on Knowledge and Data Engineering*, 16 (4):396–407, 2004.

M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

L. Tarassenko, P. Hayton, N. Cerneaz, and M. Brady. Novelty detection for the identification of masses in mammograms. In *Proc. of ICANN'95*, pages 442–447, 1995.

O. Taylor and J. McIntyre. Adaptive local fusion systems for novelty detection and diagnostics in condition monitoring. In *Proc. of the SPIE International Symposium on Aerospace/Defense Sensing*, 1998.

Anne M. Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

Hugo Vieira Neto and Ulrich Nehmzow. Visual novelty detection for inspection tasks using mobile robots. In *Proc. of SBRN 2004*, São Luís, Brazil, 2004.