Visual Novelty Detection for Inspection Tasks using Mobile Robots

Hugo Vieira Neto and Ulrich Nehmzow
Department of Computer Science
University of Essex
Wivenhoe Park
Colchester CO4 3SQ
United Kingdom
{*hvieir, udfn*}*@essex.ac.uk*

Abstract

Novelty detection – the ability to differentiate between common sensory stimuli and perceptions never experienced before – is a very useful competence for a mobile robot operating in a dynamic environment. Using such an ability, the robot can select which aspects of the environment are unusual and therefore deserve the attention from either a human operator – for instance, in supervised inspection or surveillance tasks – or its own computational resources for further processing.

Here we present a framework for novelty detection in which a mobile robot visually explores the environment and learns a model for it by means of the self-organisation of a neural network. After the learning process, the robot can be used to inspect the environment and highlight any perception that does not fit the acquired model of normality. We also present and discuss some experimental results from an inspection task involving the detection of arbitrary objects which were new to an environment that the robot had previously learnt.

1. Introduction

The limited computational resources available to an autonomous mobile robot often present challenges for applications that demand real-time processing of large amounts of sensory data, especially when artificial vision is involved.

A natural solution to cope with massive amounts of input stimuli is the use of a mechanism of attention to select aspects of interest and concentrate the available resources on those [1]. Selective attention is widely used in this manner, for instance in biological vision systems.

There is also little motivation to concentrate resources in concepts that are already well-known, but rather in new concepts that were never experienced before. In this sense, novelty detection is of fundamental importance to agents operating in a dynamic environment.

However, the ability to differentiate between common and unusual stimuli is a non-trivial task, as it is unclear beforehand which features of the environment are being looked for. Therefore, a method of model acquisition through robot learning is arguably the method of choice, rather than the explicit installation of *a priori* knowledge. Using this approach, a model of normality is learnt and used as means to separate novel from common perceptions.

Previous work using sonar readings as perceptual stimuli has successfully shown that novelty detection is possible without prior installation of models or any other kind of knowledge [2]. Nevertheless, the low resolution provided by sonar sensors poses serious limitations for real world surveillance and inspection tasks, where sensors with higher resolution are needed.

Therefore, we were interested to apply the previously developed novelty filter to visual information, instead of sonar sensors. As vision is fundamentally different from sonar sensing, processing of the visual information is necessary. Related work on visual novelty detection in constrained conditions (input images were close-ups of the walls which the robot was following) is reported in [3]. A completely different approach for novelty detection in video sequences using supervised learning can also be found in [4].

We describe in this paper a method to process colour visual information using a model of visual attention and local image statistics. A Grow-When-Required (GWR) neural network [5] is used as a novelty filter to highlight new, *arbitrary* features that may appear in the environment. Figure 1 shows a block diagram of the framework developed for our visual novelty detection mechanism.

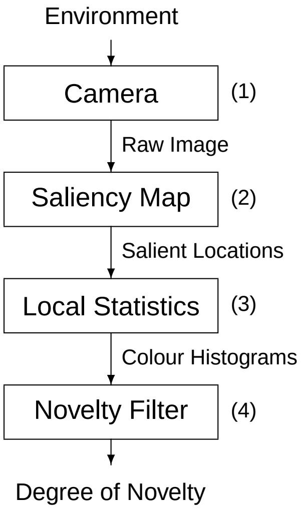

Figure 1. The visual novelty detection mechanism: local colour histograms are computed at salient locations of the image, whose degree of novelty is assigned by an artificial neural network.

Finally, we present some laboratory experiments involving a visual inspection task, in which the camera's field of view was not restricted to the walls of the environment. Results of these experiments show that our approach is promising for applications such as automated inspection and surveillance.

2. Image encoding

As mentioned in the previous section, the implementation of computer vision algorithms in mobile robots is a difficult issue: one normally desires to process a large amount of data with limited computational resources in real-time. Furthermore, the fact that images are acquired from a moving platform makes visual features subject to geometric transformations such as scaling, translation, rotation, changes in perspective and also occlusions.

In summary, a fast, compact image encoding technique is needed in order to generate robust feature vectors for higher levels of processing. A natural approach to overcome the speed difficulty is to limit image encoding to inexpensive techniques. Unfortunately, simplifying too much the encoding mechanism also limits its robustness to image transformations.

In previous work [3] some experiments in visual novelty detection were conducted using a wall-following robot. However, the robot's camera was positioned to solely acquire close-up images of the wall, restricting its field of view. This approach implicitly constrained the visual features almost only to texture and therefore also limited its usefulness for more general applications.

Here we present an image encoding method (blocks 2 and 3 in figure 1) that uses local colour statistics from salient locations within the image frame. This approach has proved to work efficiently with a wide, unrestricted field of view and demonstrated some robustness to image transformations.

2.1. Saliency map

In order to extract local features within the image it is necessary to select which regions of the image are "interesting" and deserve to be analysed in more detail. Further processing is therefore directed only to these regions, reducing computational cost.

In this work we have used the Saliency Map [6] as a model for selective visual attention. This model is inspired by the neural architecture of the early primate visual system and consists of multi- scale feature maps that allow the detection of local discontinuities in intensity, colour and orientation. Figure 2 presents a simplified block diagram for the Saliency Map.

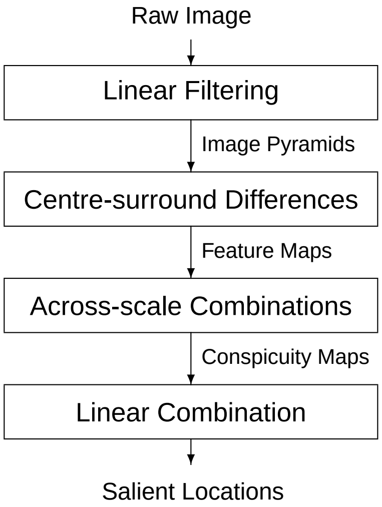

Figure 2. Simplified block diagram for the saliency map.

The feature maps are computed from a pyramidal structure similar to the one described in [7], obtained from the original input image (160×120 pixels in size). In our implementation, five Gaussian pyramids in four scales ($\sigma \in \{0, 1, 2, 3\}$) were built: $I(\sigma)$ for intensity and $R(\sigma)$ for red, $G(\sigma)$ for green, $B(\sigma)$ for blue and $Y(\sigma)$ for yellow, as described in [6]. Additionally, Gabor pyramids $O(\sigma, \theta)$ in four orientations ($\theta \in \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$) were also built using fast recursive Gabor filters [8].

Centre-surround differences were computed between a centre fine scale c and a surround coarse scale s from the pyramids to yield the feature maps. We have used $c \in \{0, 1\}$ and $s = c + 2$. The feature maps were combined in conspicuity maps for intensity, opponent colours and orientation, which were normalised and added to yield the final Saliency Map (the reader is referred to [6] for further implementation details).

The interesting property of salient points determined in this fashion is that they tend to be robust to geometric transformations, contributing to the desired general robustness of the image encoding mechanism.

We have used the ten highest values in the Saliency Map to indicate which locations of the image are likely to be the most “interesting” so that colour statistics could be calculated in their vicinity. Therefore, for each input image, we generated ten local histograms to feed the GWR-based novelty filter.

2.2. Colour histograms

Histograms are well-known statistical tools that, when applied to image features, show robustness against geometric transformations, changes in perspective and partial occlusion [9].

In this work we analyse the performance of local colour histograms, with no explicit encoding of any other image feature, such as shape or texture. To compute the colour histograms we first convert the images to the HSI (Hue-Saturation-Intensity) colour space from the RGB (Red-Green-Blue) colour space using (1), (2) and (3):

$$I = \frac{R + G + B}{3}, \tag{1}$$

$$S = 1 - \frac{\min(R, G, B)}{I}, \tag{2}$$

$$H = \arctan\left(\frac{\sqrt{3}(G - B)}{2R - G - B}\right). \tag{3}$$

Then we equally divide the hue interval $[-\pi, \pi]$ into M regions by defining the following membership functions f_m (4):

$$f_m = \begin{cases} 1 & \text{if} -\theta < H - (M - 2m)\,\theta \leq \theta \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $\theta = \frac{\pi}{M}$ and $m = 0, 1, ..., M - 1$.

A standard histogram can be computed by evaluating the responses of the membership functions q_m for each pixel in the image and adding them to the corresponding histogram bin (b_m), as shown in (5):

$$b_m = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} f_m(H_{x,y}), \tag{5}$$

where (x, y) are the pixel coordinates and $m = 0, 1, ..., M - 1$.

For the colour histograms used in the experiments reported here we have also included colour saturation information by weighting the response of the membership functions as given in (6):

$$b_m = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} f_m(H_{x,y})\, S_{x,y}. \tag{6}$$

Finally, we have normalised the histogram to satisfy the constraint $\sum_{m=0}^{M-1} b_m = 1$.

Our approach employs the above defined weighted histograms using $M = 32$ bins as input vectors for the GWR-based novelty filter. In order to exploit local information we have computed the colour histograms in sub-images of 32×32 pixels centred around the ten most salient points within the image frame.

3. Novelty filter

The novelty filter of our system (block 4 in figure 1) was based on the GWR neural network [5], which itself is derived from Kohonen’s Self Organising Feature Map (SOFM). Unlike the SOFM, however, the GWR network has the ability to add nodes to its structure in order to represent new input stimuli.

Training of the GWR network is done with an unsupervised winner-take-all approach, where the winner node and its topological neighbours have their weights adapted according to the learning rule given in (7), where $\mathbf{w}_i$ is the weight vector, ξ is the input vector and ϵ is the learning rate.

$$\Delta \mathbf{w}_i = \epsilon(\xi - \mathbf{w}_i). \tag{7}$$

The matching of the input is given by the corresponding activation value a_i of each node, as shown in (8).

$$a_i = \exp(- \parallel \xi - \mathbf{w}_i \parallel). \tag{8}$$

A model of habituation, which is a reduction in the behavioural response to stimuli that are repeatedly presented, is used as a measure of novelty. The habituation rule of a node is given in (9), where h_0 is the initial value of the habituation $h_i(t)$, $S(t)$ is the external stimulus, τ and α are time constants that control the habituation rate and the recovery rate, respectively.

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

Both activation and habituation values of the winner node for a given input are used to decide whether the stimulus is novel or not. Therefore, a new node is added every time that both activation and habituation values are below pre-defined thresholds a_T and h_T, respectively.

However, the algorithm used for the GWR network in this work is slightly different from the original presented in [2], as we have altered the learning and habituation rules for the topological neighbours of the winner node. The original approach used separate parameters ϵ_n and τ_n for the neighbours, which were just a constant fraction of ϵ_w and τ_w for the winner node. Therefore, ϵ_n and τ_n were completely independent of the distance between neighbour and winner nodes in input space. Our approach made the learning and habituation rates of the neighbour nodes proportional to their distance to the winner node in input space, as can be seen in (10) and (11), where a_w and a_n are respectively the activation of the winner and neighbour nodes and η is the proportionality factor ($0 < \eta < 1$).

$$\epsilon_n = \frac{\eta a_n}{a_w} \epsilon. \tag{10}$$

$$\tau_n = \frac{a_n}{\eta a_w} \tau. \tag{11}$$

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

For the experiments reported in this paper we have used the following parameters: $a_T = 0.9$, $h_T = 0.5$, $\eta = 0.1$, $\epsilon = 0.1$, $\tau = 3.33$, $\alpha = 1.05$, $h_0 = 1$ and $S(t) = 1$. 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. In addition, the low value assigned to the learning rate ϵ assures that nodes are not able to move too much from the location in input space where they were originally placed.

4. Experimental setup

The experiments discussed here were conducted using the colour vision system of a Magellan Pro mobile robot (figure 3), which is also equipped with a laser range scanner.

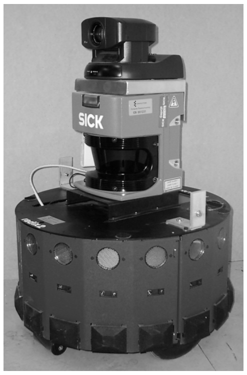

Figure 3. The Magellan Pro mobile robot used for the experiments.

The robot's navigation behaviour was determined exclusively by the data provided by the laser range scanner. We have employed the force field strategy, in which every distance measure covering the $180°$ in front of the robot acts like a virtual spring that pushes it towards the freest space in the environment. Basically, the robot translates very slowly (0.15m/s), until it finds an obstacle within a threshold distance of 0.5m, which causes it to stop and slowly rotate ($35°$/s maximum) towards free space again. In our experiments, this behaviour has shown to be extremely predictable and stable.

Figure 4 shows the top view of the environment used for the experiments, which consists of a closed arena surrounded by cardboard boxes and plastic cylinders.

The boxes and cylinders at the borders of the arena act as walls that limit the path of the robot and also its visual world. It can be noticed from figure 4 that the arena's floor has several marks, which contribute to add some visual heterogeneity to the environment.

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 ($-25°$). Therefore, the robot's field of view consisted of mostly of the floor and the walls of the arena. An example image of the environment from the robot's perspective is given in figure 5.

The images used in our experiments were acquired at one frame per second and without stopping the robot, resulting in a total of 45 image frames per loop around the arena.

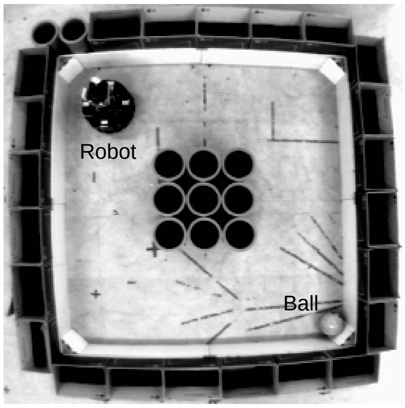

Figure 4. Top view of the arena used for the experiments: the robot is shown at its starting position and an orange football at the opposite corner.

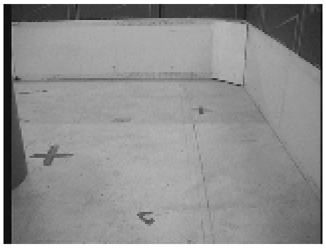

Figure 5. Robot's view of the environment from its starting position.

4.1. Task

Our experiments were designed to evaluate the ability of the devised mechanism to detect arbitrary novel visual features that may be inserted in the environment. Therefore, they were conducted in two stages: an exploration (learning) phase and an inspection (application) phase.

During the learning phase we acquired images while the robot was navigating around the empty arena. These images were used to generate local histogram-based feature vectors and train the GWR network.

During the application phase, some novel object was inserted inside the arena and again the robot was used to acquire images while navigating. This new sequence of images was then used to test the trained GWR network, using the habituation value of the winner node as a measure of novelty.

The expected outcome of these experiments was that the amount of novelty would progressively be reduced during the exploration phase, resulting from the self-organisation of the GWR network to represent the original environment. Additionally, it was expected during the inspection phase that peaks in the measure of novelty would appear where the novel object was inserted.

4.2. Results

The learning dataset was built with images acquired during five loops in the empty arena. They were used for off-line training of the GWR network and the amount of novelty in each frame was computed as the average of the habituation values of the winner nodes for each of the ten computed local histograms.

The amount of novelty measured during the learning phase is shown in figure 6. It can be noticed that the novelty values reduce as the robot explores the environment. Given the GWR network parameters we used, novelty values can range from a minimum of 0.05 and a maximum of 1.0. The four major peaks of novelty that appear in the graph for the first loop correspond to the corners of the arena, where the robot was turning.

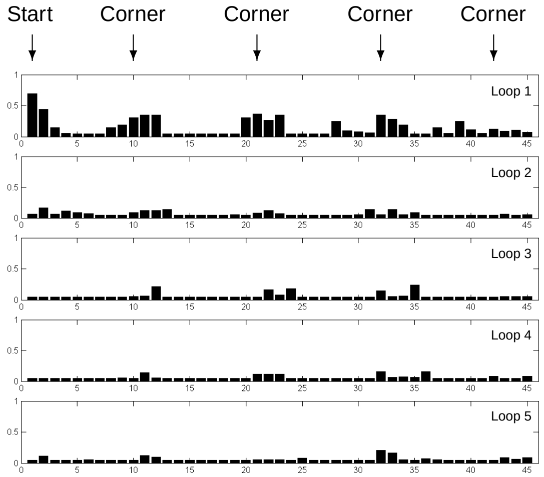

Figure 6. Original environment: the graphs depict the amount of novelty measured at every location in five consecutive loops around the empty arena. Learning of the GWR network was enabled.

For the application phase, an object was placed at one of the corners of the arena. Care was taken to select objects that did not interfere with the original path of the robot, i.e. objects that were not detected by the laser range scanner.

Figure 7 shows an example of the amount of novelty measured during the application phase when an orange football was placed in the arena (as shown in figure 4). The ball appeared within the field of view of the camera immediately after the robot turned the first corner, as indicated.

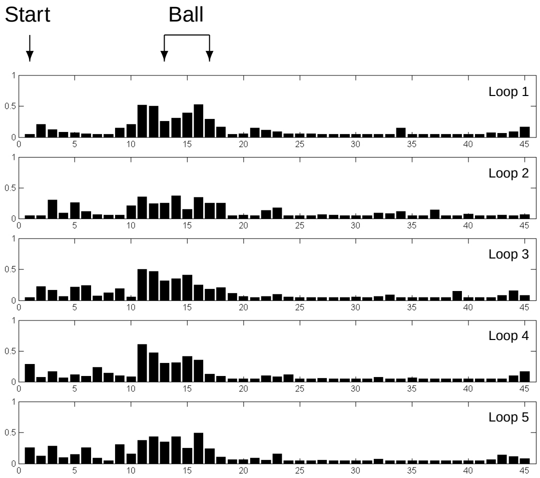

Figure 7. Altered environment: the graphs depict the amount of novelty measured at every location in five consecutive loops around the arena when an orange football was placed at one of its corners. Learning of the GWR network was disabled.

As can be seen in figure 7, the novel object is clearly detected and differentiated from the other visual stimuli observed in the arena.

5. Conclusion

In this paper, we presented a novelty-detection mechanism using vision, with potential applications in inspection tasks using mobile robots. The proposed approach takes into account local colour statistics from salient locations within the image frame to form a sub-symbolic representation of the environment without the installation of any *a priori* knowledge.

Colour histograms were computed in regions of the input image determined by a saliency-based mechanism of visual attention, which still has to be refined in terms of stability. Nevertheless, experiments conducted in a controlled scenario with a moving robot have shown that our approach has the ability to detect new, *arbitrary* objects as soon as they first appear in the camera's field of view.

The GWR network has quickly learnt a representation of the "normal" environment through our image encoding method, and was used in a straightforward manner to highlight "abnormal" features that were introduced later. Although the image encoding mechanism still needs further refinement, the results obtained are promising and have excellent potential to applications such as flexible automated inspection.

Future work includes the implementation of on-line training and the implementation of an improved image encoding mechanism, possibly including scene representation. As the computation of the Saliency Map is computationally expensive, we are interested in the possibility of using the available information from the image pyramids to compute local multidimensional receptive field histograms [9].

Acknowledgements

The first author is sponsored by the Brazilian Government through CAPES Foundation and the Federal Centre for Technological Education of Paraná (CEFET-PR), whose support is gratefully acknowledged.

References

[1] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[2] S. Marsland, U. Nehmzow, and J. Shapiro, "Environment-specific novelty detection," in *From Animals to Animats: Proceedings of the 7th International Conference on the Simulation of Adaptive Behaviour (SAB'02)*. Edinburgh, UK: MIT Press, 2002.

[3] S. Marsland, U. Nehmzow, and J. Shapiro, "Vision-based environmental novelty detection on a mobile robot," in *Proceedings of the International Conference on Neural Information Processing (ICONIP'01)*, Shangai, China, 2001.

[4] S. Singh and M. Markou, "An approach to novelty detection applied to the classification of image regions," *IEEE Transactions on Knowledge And Data Engineering*, vol. 16, no. 4, pp. 396–407, 2004.

[5] S. Marsland, J. Shapiro, and U. Nehmzow, "A self-organising network that grows when required," *Neural Networks*, vol. 15, no. 8-9, pp. 1041–1058, 2002.

[6] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[7] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C. H. Anderson, "Overcomplete steerable pyramid filters and rotation invariance," in *1994 Computer Vision and Pattern Recognition (CVPR'94)*, 1994, pp. 222–228.

[8] I. T. Young, L. J. van Vliet, and M. van Ginkel, "Recursive Gabor filtering," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2798–2805, 2000.

[9] B. Schiele and J. L. Crowley, "Object recognition without correspondence using multidimensional receptive field histograms," *International Journal on Computer Vision*, vol. 36, no. 1, pp. 31–50, 2000.