Novelty-Based Visual Inspection Using Mobile Robots

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

Abstract

In this paper we present a novelty detection mechanism in which a mobile robot learns a model for its environment through visual exploration. Once the learning process is finished, the robot can be used to inspect the environment and highlight stimuli that do not fit the acquired model of normality. Experimental results from a visual inspection task involving the detection of arbitrary objects are also presented in this paper.

1. Introduction

The ability to differentiate between common and unusual perceptions, also known as novelty detection, can be very useful for mobile robots that operate in dynamic environments. A robot with such a competence can select which aspects of the environment are abnormal 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.

However, the implementation of a novelty detection mechanism is not trivial. As it is unclear beforehand which features of the environment need to be searched for, it is not feasible to install explicit models *a priori*. Instead, models have to be acquired – furthermore, as novelty detection entails the identification of *any* novel stimuli, these models need to be models of normality, rather than abnormality.

Following this approach, a model of normality is learnt through the self- organisation of a neural network and used as means to separate novel from common perceptions.

Previous work using sonar readings as the source for perceptual stimuli has shown that novelty detection is possible without prior installation of any kind of knowledge (Marsland et al., 2002a). Nevertheless, the low sensory resolution provided by sonars poses serious limitations for *real world* surveillance and inspection tasks, where sensors with higher resolution are needed.

In this work we investigate the possibility of applying the novelty filter used in (Marsland et al., 2002a) to visual information, instead of sonar readings. We describe a method to process colour visual information using image statistics, generating feature vectors for a Grow-When-Required (GWR) neural network (Marsland et al., 2002b). The GWR network is then used to highlight new, *arbitrary* features that contrast with the acquired model of normality. A block diagram of our visual novelty detection mechanism is shown in figure 1.

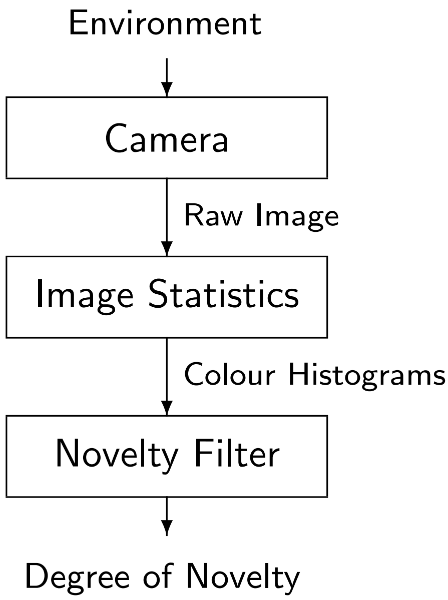

Figure 1: The visual novelty detection mechanism: colour histograms are computed from the images acquired from the environment and then a novelty filter is used to measure their degree of novelty.

Finally, we present some laboratory experiments which involve a visual inspection task to detect novel features in the environment. We discuss the use of global and local colour histograms – centred around visually salient locations within the image – as image descriptors.

2. Novelty Filter

As the concept of novelty is ill-defined, the most feasible approach for the implementation of a novelty filter consists in learning a model of *normality* for the environment and then using it as means to highlight abnormal (i.e. novel) perceptions.

Several techniques available in the literature – for instance the ones reported in (Ypma and Duin, 1997, Taylor and MacIntyre, 1998) – rely on Kohonen's Self-Organising Feature Maps (SOFM) (Kohonen, 1984) to build the model of normality. The main advantage of this approach is that there is no need of any *a priori* knowledge installation.

In this work we used a Grow-When-Required (GWR) neural network, which itself is derived from the SOFM, but has the additional capability to add nodes to its structure in order to represent new perceptions. The ability to grow is particularly interesting to avoid the network becoming "saturated" with the amount of information (Marsland et al., 2002b).

The GWR uses a model of habituation, which is a reversible reduction in responses to repeatedly presented stimuli. With the use of habituation, novelty can be defined more specifically as stimuli which have not been perceived in the current context. Moreover, habituation also allows an objective measure of the degree of novelty of a given stimulus over time (Marsland et al., 2002a).

The model of habituation is given by the following first-order differential equation:

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

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

In this work we have used $\tau = 3.33$, $\alpha = 1.05$, $h_0 = 1$ and $S(t) = 1$. These parameter values constrain the synaptic efficacy $h(i)$ of each network node to the range $[0.05, 1]$, the minimum value meaning complete habituation of the node to the input and the maximum value meaning complete novelty. As $S(t)$ is a constant positive value, there is no recovery (i.e. dishabituation) in our case, which allows us to use the synaptic efficacy of the winner node $h_w(t)$ as a direct measure of the degree of novelty for any given input stimulus.

The network is trained with a traditional winner-take-all approach, in which the weights of the winner node and its topological neighbours are adapted according to the learning rule given below:

$$\Delta \mathbf{w}_i = \epsilon(\xi - \mathbf{w}_i), \tag{2}$$

where $\mathbf{w}_i$ is the weight vector, ξ is the input vector and ϵ is the learning rate.

The activation value a_i of a given input vector for each node is determined by the equation that follows:

$$a_i = \exp(- \parallel \xi - \mathbf{w}_i \parallel). \tag{3}$$

If both habituation and activation values of the winner node are below pre-defined thresholds h_T and a_T, respectively, a new node is added to the network. We have used $h_T = 0.5$ and $a_T = 0.9$ to make sure that new nodes are added for every novel stimulus without the need of a large number of iterations. In addition, we have used a learning rate ϵ of 0.1, which assures that nodes are not able to move too much from the location in input space where they were originally placed.

Concerning the topological neighbours of the winner node, the training algorithm used in this work is slightly different from the original one (Marsland et al., 2002a), which used separate parameters ϵ_n and τ_n for the neighbours. These values were just a constant fraction of the parameters ϵ and τ. Our approach made the learning and habituation rates of the neighbour nodes proportional to their distance to the winner node in input space:

$$\epsilon_n = \frac{\eta a_n}{a_w} \epsilon, \tag{4}$$

$$\tau_n = \frac{a_n}{\eta a_w} \tau, \tag{5}$$

where a_w and a_n are respectively the activation of the winner and neighbour nodes and η is the proportionality factor ($0 < \eta < 1$).

It can be noticed from the equations above that neighbour nodes will have their weights adapted to a lesser extent and habituate in a slower rate than the winner node (in this work we have used $\eta = 0.1$). For further details about the implementation of the GWR network algorithm, please see (Marsland et al., 2002a, Marsland et al., 2002b).

3. Image Encoding

The use of computer vision algorithms in mobile robots is a challenging task: large amounts of data need to be processed in real-time with limited computational resources. Moreover, acquiring images from a moving platform make visual features subject to a number of geometric transformations such as scaling, translation, rotation, changes in perspective and occlusions.

In order to make efficient use of the GWR-based novelty filter it is necessary to generate input vectors that are at the same time compact, robust to image transformations and fast to compute. In this paper we present an image encoding method that uses colour statistics to represent image features.

3.1 Colour Histograms

Image histograms are well-known statistical tools that when properly used can present robustness against image transformations and partial occlusions (Schiele and Crowley, 2000). Here we are concerned with the performance of colour histograms, with no explicit encoding of any other image feature.

To compute the colour histograms we first convert the images to the HSI colour space from the RGB colour space as follows:

$$I = \frac{R+G+B}{3}, \tag{6}$$

$$S = 1 - \frac{\min(R,G,B)}{I}, \tag{7}$$

$$H = \arctan\left(\frac{\sqrt{3}(G-B)}{2R-G-B}\right), \tag{8}$$

where R, G and B are the red, green and blue channels and H, S and I are the hue, saturation and intensity for each pixel, respectively. The HSI colour space was selected with the purpose of separating explicitly chrominance (hue and saturation) from luminance (intensity).

We divided the hue interval $[-\pi, \pi]$ equally into M regions by defining the following membership functions f_m:

$$f_m = \begin{cases} 1 & \text{if} -\theta < H - (M-2m)\,\theta \leq \theta \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

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

The standard hue histogram is computed by adding the responses of the membership functions f_m for each pixel in the image to the corresponding histogram bin (b_m), as shown below:

$$b_m = \sum_{x=0}^{X-1}\sum_{y=0}^{Y-1} f_m(H_{x,y}), \tag{10}$$

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

We have also included saturation in the colour histograms used in our experiments by weighting the response of the membership functions as follows:

$$b_m = \sum_{x=0}^{X-1}\sum_{y=0}^{Y-1} f_m(H_{x,y})\,S_{x,y}. \tag{11}$$

Finally, we have normalised the histogram to satisfy the constraint $\sum b_m = 1$. Our approach employs the above defined histograms using $M = 32$ bins as input vectors for the GWR-based novelty filter.

3.2 Global and Local Histograms

In our first experiments we have used colour histograms in a global fashion, i.e. the histograms were computed for the whole image frame. This approach has shown to be able only to detect large visual alterations in the environment (see section 4.), but lacked the ability to point where exactly the alteration was in the image frame.

To provide the novelty detection mechanism with the ability to localise where in the image frame are the novel visual features, local statistics, rather than global, are needed. Therefore, a method to determine which regions of the image are the most "interesting" and deserve further analysis has to be employed.

In this work we have used the Saliency Map (Itti et al., 1998) as a model for selective visual attention. This model is inspired by the neural architecture of the early primate visual system and consists of multiscale feature maps that allow the detection of local discontinuities in intensity, colour and orientation. Further details of our implementation of the Saliency Map are given in (Vieira Neto and Nehmzow, 2004).

The interesting property of salient points determined in this fashion is that they tend to be robust to geometric transformations and contribute to the general desired robustness of the whole image encoding mechanism. We have selected the ten highest values in the Saliency Map to indicate which locations of the image are the most "interesting" so that colour histograms could be calculated in their vicinity. The salient points tend to correspond to edges, corners and similar discontinuities.

Therefore, for each input image, ten local histograms were generated to feed the GWR-based novelty filter. The histograms were computed from patches of 32×32 pixels centred around the salient points.

4. Experimental Setup

The experiments discussed here were conducted using the colour vision system of the Magellan Pro mobile robot *Radix 2*, which is shown in figure 2.

Figure 3 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, limiting the path of the robot and also its visual world. It can be noticed from figure 3 that the arena's floor has several marks, which contribute to add some visual heterogeneity to the environment.

With the 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 angle ($-25°$). Therefore, the robot's field of view consisted of mostly of the floor and the walls of the arena.

The images used in our experiments were acquired at one frame per second and without stopping the robot,

Figure 2: The Magellan Pro mobile robot used for the experiments. The laser range scanner visible in the photograph was used only for motion control, not for novelty detection.

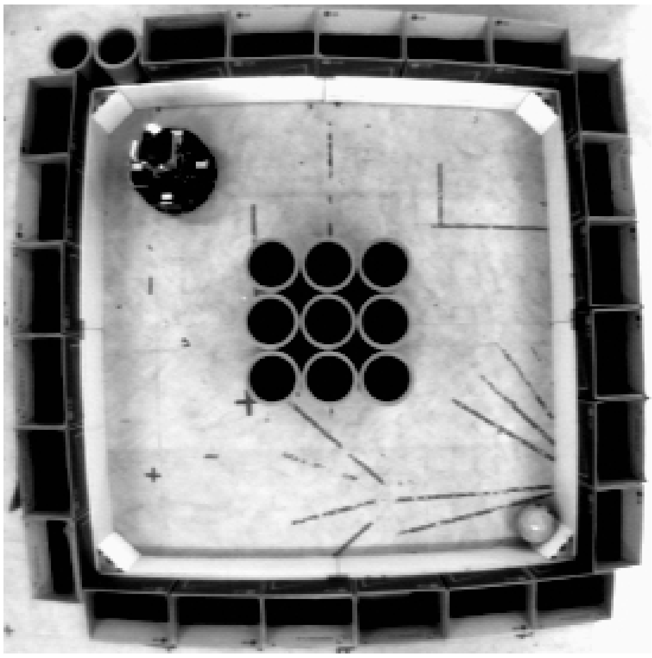

Figure 3: Top view of the arena used for the experiments: the robot is shown at its starting position and an orange football at the opposite corner.

resulting in a total of 45 image frames per loop around the arena.

4.1 Robot Behaviour

The navigation behaviour of the robot was exclusively determined by the distance information provided by the laser range scanner. We have used the force field strategy, in which every distance measure covering 90° in front of the robot acts like a virtual spring, pushing the robot towards free space in the environment.

The robot moves forward 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. This behaviour has shown to be extremely predictable and stable in our experiments.

4.2 Robot 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. They were conducted in two stages: an exploration (learning) phase and an inspection (application) phase.

During the exploration phase we acquired images while the robot was navigating around the empty arena. These images were used to generate the histogram-based feature vectors and train the GWR network. For the inspection phase, some novel object was inserted inside the arena and once more the robot was used to acquire images while navigating. The new sequence of images was then used to test the trained GWR network, using the synaptic efficacy of the winner node h_w 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. Additionally, it was expected during the inspection phase that peaks in the measure of novelty would appear mostly near to where the novel object was inserted.

4.3 Results and Discussion

The learning dataset was built with images acquired during five loops in the *empty* arena. Figure 4 shows the degree of novelty measured during the exploration phase when using global histograms (i.e. a single colour histogram for the whole image frame). Given the used GWR network parameters, novelty values can range from a minimum of 0.05 and a maximum of 1.0. It can be seen that the novelty values tend to slowly reduce as the robot explores the environment, but there are several unexpected peaks of novelty which indicate problems with the use of global histograms for a proper image encoding.

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

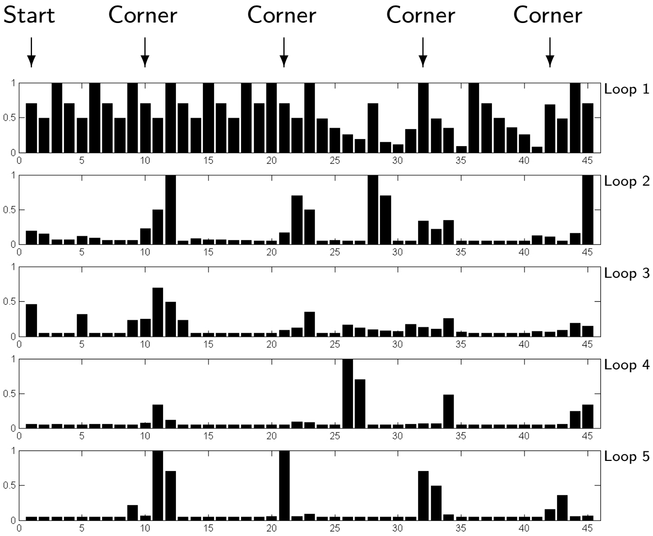

Figure 4: Original environment explored with the use of global histograms: the graphs depict the amount of novelty measured at every location in five consecutive loops around the empty arena. Learning of the GWR network was enabled.

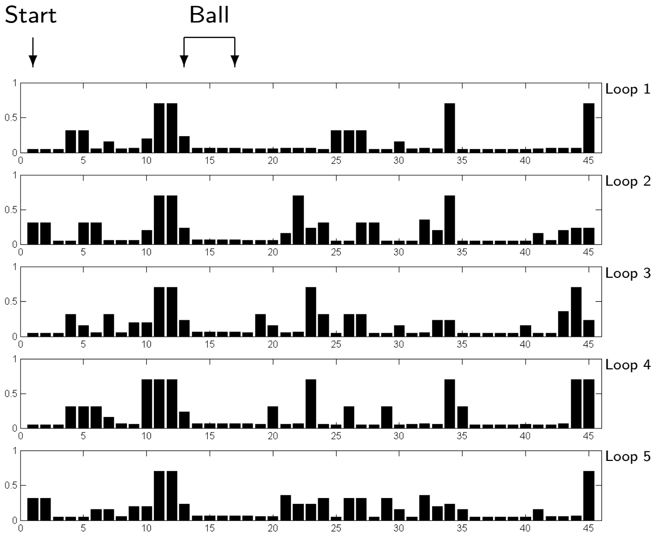

Figure 5: Altered environment inspected with the use of global histograms: the graphs depict the amount of novelty measured at every location in five consecutive loops around the arena when an orange football was placed at one of its corners. Learning of the GWR network was disabled.

range scanner.

Figure 5 shows an example of the amount of novelty measured during the inspection phase when an orange football was placed in the arena (as shown in figure 3), once again using global histograms as image descriptors. The ball appeared within the field of view of the camera immediately after the robot turned the first corner, as indicated. Unfortunately, in these graphs there is no clear indication that the novel object is detected.

As the approach using global histograms did not work as desired, we decided to use a mechanism of selective attention and compute local histograms in the vicinity of salient image locations, as described in the previous section. The local approach has shown to perform as desired, with the additional advantage of pointing out where the new visual features are located within the image frame.

Figure 6 shows the degree of novelty measured during the learning phase when using the mechanism of attention and local histograms. It can be noticed that the novelty values consistently reduce as the robot explores the environment. The amount of novelty in each frame was computed as the average of the synaptic efficacies of the winner nodes for each of the ten computed local histograms.

The four major peaks of novelty that repeatedly appear in the first loop correspond to the corners of the arena, where the robot was turning. These peaks are persistent over the next loops, indicating that our approach still needs to be further refined to cope with the major geometric transformations that occur in the images when the robot is turning a corner.

Finally, the degree of novelty measured during the inspection phase when using the mechanism of attention and local histograms is shown in figure 7.

As can be observed in figure 7, the novel object is clearly detected and differentiated from the other visual stimuli observed in the arena. Figure 8 shows an example of how the mechanism is able to locate the novel visual features within the image frame.

5. Conclusion

The ability to detect stimuli that contrast with stimuli commonly perceived in a robot's environment – novelty detection – is of immense importance to autonomous mobile robots, as it allows the precise commitment of sensory and computational resources, decreases reaction time and improves the reliability and security of robot operation. Besides, novelty detection is a core component of tasks such as surveillance, inspection and exploration.

In this paper we have presented experiments with a fully autonomous Magellan Pro mobile robot (*Radix 2*), that uses the mechanism of habituation and growing artificial neural networks to acquire models of normality, using visual input. These models can subsequently be used to detect anything not conforming with them, i.e. abnormality.

We have discussed two experiments in this paper, one using image encoding with global histograms, the other using local histograms. Both approaches are able to detect a foreign object introduced into the experimental

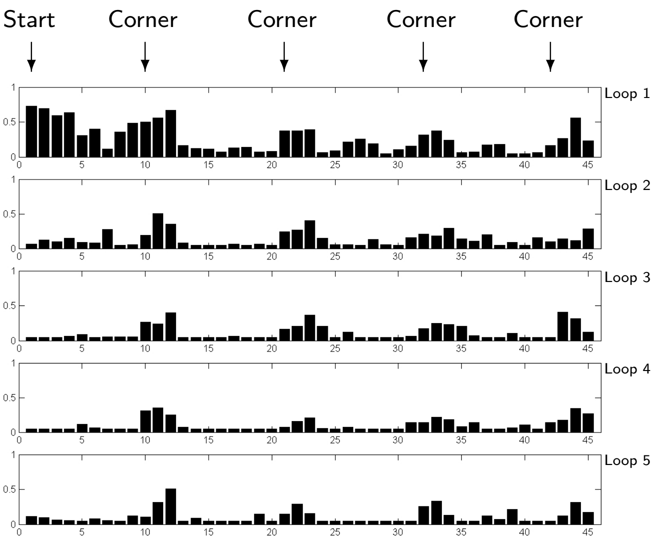

Figure 6: Original environment explored with the use of local histograms (to be contrasted with figure 4): the graphs depict the amount of novelty (average of the novelty values for the ten most salient regions) measured at every location in five consecutive loops around the empty arena. Learning of the GWR network was enabled.

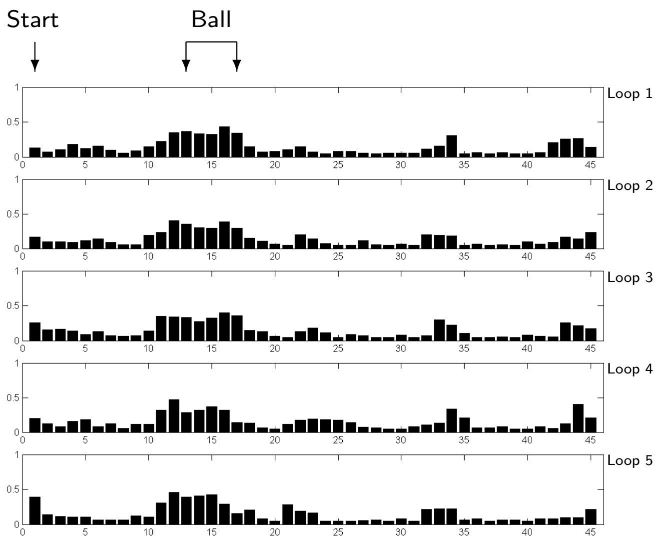

Figure 7: Altered environment inspected with the use of local histograms (to be contrasted with figure 5): the graphs depict the amount of novelty (average of the novelty values for the ten most salient regions) measured at every location in five consecutive loops around the arena when an orange football was placed at one of its corners. Learning of the GWR network was disabled.

arena, but the method using local histograms has the additional advantage of detecting the locus of the novel stimulus.

One important issue concerning industrial applications is the question of whether this approach will scale up. Ongoing work at Essex therefore investigates the application of these mechanisms to larger real world environments, with the objective to detect faults such as cracks and foreign objects in real-world environments.

Acknowledgements

The second 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

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

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

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

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

Schiele, B. and Crowley, J. L. (2000). Object recognition without correspondence using multidimensional receptive field histograms. *International Journal on Computer Vision*, 36(1):31–50.

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

Vieira Neto, H. and Nehmzow, U. (2004). Visual novelty detection for inspection tasks using mobile robots. In *Proc. 8th Brazilian Symposium on Neural Networks (SBRN 2004)*.

Ypma, A. and Duin, R. P. W. (1997). Novelty detection using self-organizing maps. In *Proceedings of the International Conference on Neural Information Processing and Intelligent Information Systems*, Dunedin, New-Zealand.

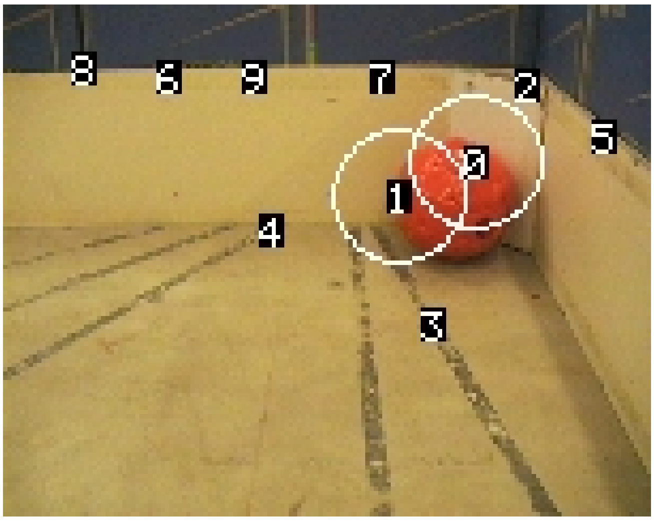

Figure 8: Localisation of novel features during the inspection phase: the numbers indicate the salient points in the image – lower numbers indicating an item of greater saliency – and the white circles indicate novelty values above the habituation threshold h_T.