Locating Objects Visually Using Opposing-Colour-Channel Coding

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

Abstract

The work presented in this paper shows how a simple algebraic manipulation of the RGB colour channels of an artificial vision system can significantly improve the performance of an object-locating mechanism based on unsupervised clustering.

The proposed image processing and clustering mechanism forms the first part of a novelty-detecting robot controller. Experiments with two different connectionist approaches (Self-Organising Feature Map and Principal Component Analysis Network) and an evaluation of their performances to the desired novelty-detection task are also presented.

1. Introduction

1.1 Motivation: Novelty Detection

Novelty detection — the ability to differentiate between common, previously perceived sensory stimuli and stimuli that have not been encountered before, within either a relative or a global timeframe — is of fundamental importance to all agents operating in and interacting with a dynamic environment. In living beings, the ability to detect novel stimuli reduces the threat from predators and other dangers, and enables the animal to focus on opportunities such as prey *etc.* In mobile robots, the main advantage of novelty detection from an operational point of view is the concentration of resources (e.g. computation, memory) onto relevant aspects, and from an application point of view the ability to carry out for instance surveillance and inspection tasks more efficiently and reliably.

We have shown that novelty detection without prior installation of models or any other kind of knowledge in mobile robots is possible (Marsland et al., 2000). That work used a Nomad 200 and its sonar sensors as perceptive stimuli. While sonar sensing is powerful, particularly with respect to operational aspects of mobile robots, its low resolution poses limitations for surveillance and inspection tasks. For such tasks, sensors with higher resolution are needed.

We were therefore interested to apply our previously developed novelty filter to visual stimuli, rather than sonar sensor signals. As visual images are fundamentally different to signals obtained from sonar sensors, processing of the visual stimuli is necessary.

This paper presents experiments for the development of image processing methods, based on self-organisation rather than pre-installed knowledge, that will ultimately be used as input stages to a novelty filter. In (Vieira Neto and Nehmzow, 2003) we presented an image processing mechanism based directly on the RGB colour channels. The results obtained were promising, but we are able to show in this paper that using opposing colour channels increases a robot's ability to locate target objects and to classify them considerably.

1.1.1 Task: Visual Location of Objects

As a first stage towards a novelty-detecting robot using visual stimuli, we were interested to develop a vision-based mechanism to identify (arbitrary) objects within an image frame. To achieve this, we divided the entire frame of 160×120 pixels into 25 sub-images of 32×24 pixels each, as shown in figure 1.

For the eventual novelty detection application, the robot's task would be to classify each of these sub-images as either novel or not novel. In the experiments reported here, the robot's task was to decide in which of the 25 sub-images a target

object shown before the experiment was located.

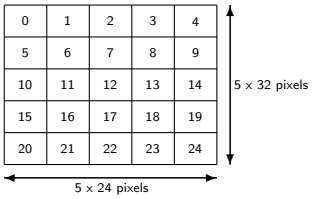

Figure 1: The 25 sub-images that result from partitioning the original image

1.1.2 Approach: Clustering Through Self-Organisation

We aimed to achieve the task of locating a target object in one of the 25 sub-images by encoding the image information according to colour and intensity (discussed in section 2.1), and clustering that information through self-organising, subsymbolic processes (discussed in section 3.3). This is depicted in figure 2.

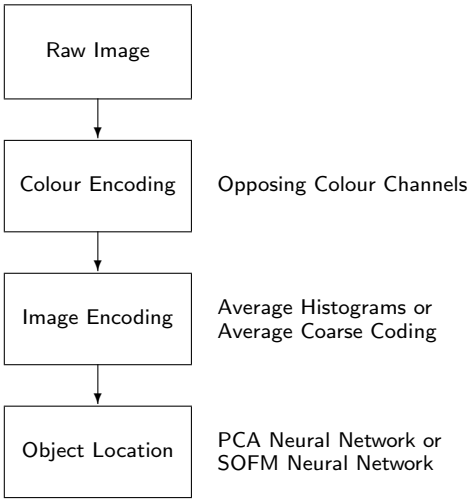

Figure 2: Functional blocks of the object-locating system

1.2 Related Work

The image processing mechanism presented here is based on the mechanism discussed in (Vieira Neto and Nehmzow, 2003), with the difference that the colour information is encoded differently.

The opposing-colour-channel encoding used here is based on the saliency map presented by (Itti et al., 1998). The artificial neural networks used were Kohonen's self-organising feature map (SOFM) (Kohonen, 1984) and the principal component analysis (PCA) network presented in (Ballard, 1997).

2. Image Encoding

2.1 Encoding Colour Information

Clearly, colour information will be relevant for locating target objects, and has therefore been used in previous experiments, as well as the experiments reported here.

We have used the information from the camera's three colour channels R (red), G (green) and B (blue) directly in previous experiments (Vieira Neto and Nehmzow, 2003). In contrast to this, the experiments reported here were based on the opponent-colour theory, which postulates that three processes mediate human vision — one achromatic and two chromatic.

In one of the chromatic processes, yellow and blue are combined in an opponent relationship, whereas in the other one, red and green form the antagonistic pair. Perception of black and white is mediated by the achromatic process (MacIlwain, 1996).

To obtain opposing colour channel information, we first compute the intensity of each pixel as $I = (R + G + B)/3$. In order to decouple hue from intensity, we normalise the R, G and B channels as indicated by equations 1, 2 and 3 (Itti et al., 1998), where I_{max} is the maximum pixel intensity.

$$r = \begin{cases} R/I & \text{if } I > I_{max}/10 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$g = \begin{cases} G/I & \text{if } I > I_{max}/10 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$b = \begin{cases} B/I & \text{if } I > I_{max}/10 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Using r, g and b, we then obtain four broadly tuned colour channels, as indicated by equations 4, 5, 6 and 7.

$$R' = r - (g + b)/2 \quad (4)$$

$$G' = g - (r + b)/2 \quad (5)$$

$$B' = b - (r + g)/2 \quad (6)$$

$$Y = (r + g)/2 - |r - g|/2 - b \quad (7)$$

The channels I, $R' - G'$ and $B' - Y$ are then used as input information to subsequent image

processing and clustering stages (described in sections 2.2, 2.3 and 3.3).

Our experiments had the purpose to determine whether encoding colour information in opposing channels has advantages over using RGB colour information directly.

2.2 Image Encoding Using Average Histograms

The first of the two image encoding techniques used consisted of computing the horizontal and vertical average histograms of the normalised intensity image I and the colour opponent channels $R'-G'$ and $B'-Y$. The histograms were stacked into single feature vectors with $3\times(32+24) = 168$ elements (figure 3).

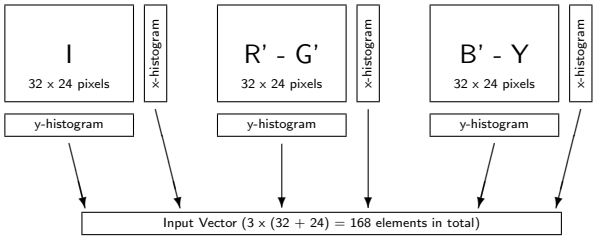

Figure 3: Image encoding using average histograms

2.3 Image Encoding Using Average Coarse Coding

The second image encoding mechanism used consisted of computing the average pixel value of the normalised intensity image I, and the colour opponent channels $R' - G'$ and $B'-Y$ within an 8×8 neighbourhood, resulting in feature vectors of $3\times(32/8\times 24/8) = 36$ elements (figure 4).

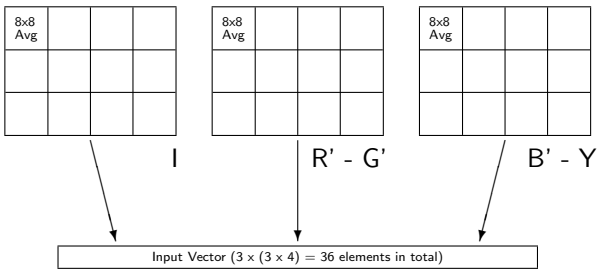

Figure 4: Image encoding using average coarse coding

3. Experimental Design

3.1 Data Used

Training images. For training, 50 different images were used. These images showed all three target objects (a blue cylinder, a green box and an orange football) together, against the unstructured background of the Brooker Laboratory at the University of Essex. The objects were distributed in random positions and orientations. Figure 5 (a) shows one example image from the training set.

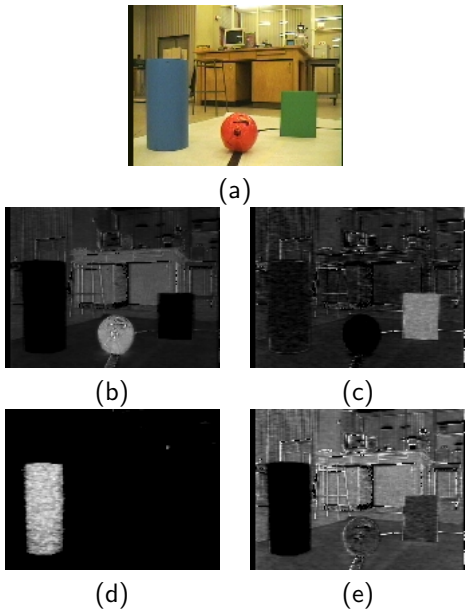

Figure 5: Images used. (a) original colour image showing the blue cylinder, the green box and the orange football; (b) R' channel emphasising the orange ball; (c) G' channel emphasising the green box; (d) B' channel emphasising the blue cylinder; (e) Y channel emphasising the yellowish background

Test image set 1. This set of test images comprised of 3×10 images, each containing exactly one of the three target objects in a randomly selected position within the frame. These test images were acquired in the same non-structured environment as the training images, although against a different background.

Test image set 2. A second testing dataset of 3×20 images, each containing all the objects included in the training dataset, was also acquired in order to evaluate the performance of the system when the target object is in the presence of distractors. This second test dataset was built to evaluate the performance of the neural networks in identifying a target object when multiple "known" objects (distractors) are present (see section 4.3).

Test image set 3. Finally, a third testing dataset was built in the same manner, but with objects that were not present in the training dataset, namely a black circle, a red cylinder and

a yellow box. This third test dataset was used in the experiments with "unknown" objects, described in section 4.4.

3.2 Robot

All experiments discussed in this paper were conducted using the colour camera system of our Magellan Pro mobile robot (figure 6).

Figure 6: The Magellan Pro mobile robot, whose vision system was used in the experiments

The raw images used were 160 × 120 pixels in size, encoded in the three colour channels R, G and B. The robot's ability to move was not used in the experiments discussed here, but will obviously be needed for future experiments, when the robot will be required to detect novel stimuli and to focus on them.

3.3 Object Classification Through Unsupervised Clustering

Having encoded the image information as described in sections 2.1, 2.2 and 2.3, we used two alternative subsymbolic methods for identifying and classifying objects: Kohonen's self-organising feature map (Kohonen, 1984) and the principal component analysis network (Ballard, 1997).

The task of both networks was to classify perceived visual stimuli. This was subsequently used for finding a target object within the image.

3.3.1 Self-Organising Feature Map

The SOFM used was a torus of 10x10 units (figure 7). It was trained according to the usual winner-takes-all approach (Kohonen, 1984), using the similarity matching given in equation 8, where $\vec{w_c}$ is the winner among all $\vec{w_i}$ units for a given input $\vec{x}$.

$$\| \vec{x}(t) - \vec{w_c}(t) \| = \min_i \{ \| \vec{x}(t) - \vec{w_i}(t) \| \} \quad (8)$$

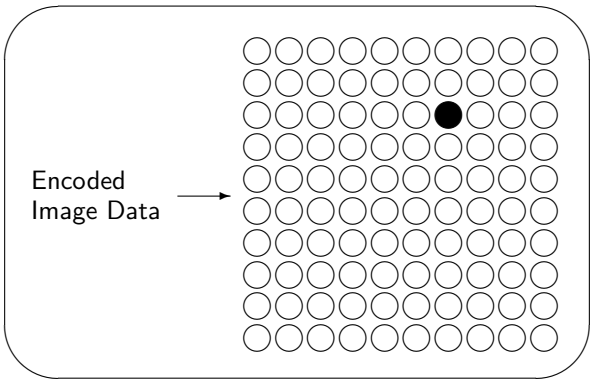

Figure 7: The self-organising feature map used

During the learning phase, the weight vectors of the winner unit and its neighbours were modified according to equation 9, where α is the learning rate ($0 < \alpha < 1$) and N_c is the topological neighbourhood of the winner.

$$\vec{w_i}(t+1) = \vec{w_i}(t) + \alpha[\vec{x}(t) - \vec{w_i}(t)],\ i \in N_c \quad (9)$$

In our experiments, both the learning rate and the topological neighbourhood size decreased with the training cycles, as shown in equations 10 and 11.

$$\alpha(m) = \exp(-10m/M) \quad (10)$$

$$N_c = \begin{cases} 3 & \text{if } 0 < m \leq 0.2M \\ 2 & \text{if } 0.2M < m \leq 0.5M \\ 1 & \text{if } 0.5M < m \leq M \end{cases} \quad (11)$$

M is the total number of training cycles (100 in all experiments with the SOFM).

3.3.2 Principal Component Analysis Network

As an alternative unsupervised clustering mechanism we used a single-layer feedforward network determining the 16 principal components of the input data. This network is discussed in (Ballard, 1997), and shown diagrammatically in figure 8.

The PCA network was trained with the Generalised Hebbian Algorithm (GHA) (Sanger, 1989). To compute the output vector $\vec{y}$ for a given input vector $\vec{x}$, equation 12 was used. Equation 13 describes how the weights w_{ij} of the network were adapted.

$$y_i = \sum_{j=1}^{J} w_{ij} x_j \quad (12)$$

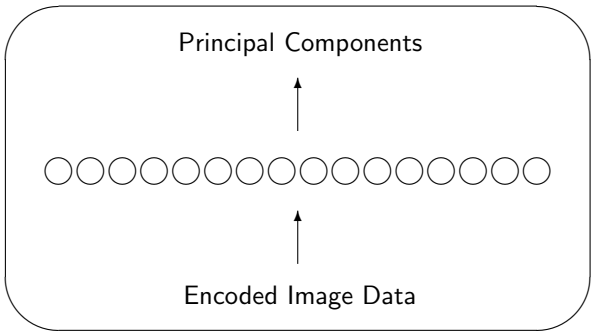

Figure 8: The PCA network used

$$\Delta w_{ij} = \alpha y_i [x_j - \sum_{k=1}^{i} y_k . w_{kj}]$$
$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij} \quad (13)$$

It was necessary to compute the average vector of the training data and then subtract it from all vectors in the set, in order to obtain zero-mean data. The learning rate for the PCA network was also made to decrease exponentially (equation 14).

$$\alpha(m) = 0.05 \exp(-m/M) \quad (14)$$

Again, M is the total number of training cycles (400 in all experiments with the PCA network).

4. Experiments and Results

4.1 Evaluation of Experimental Results

As stated before, the system's task was to identify that of the 25 subframes (figure 1) within which the target object was located. In order to assess the system's performance, we identified the "correct" answer for each test image manually. The criterion used for this decision was that any subimage that contained at least 25% of the target object was an acceptable answer[1].

We then performed a χ^2 analysis to determine whether there was a statistically significant correlation between the system's response and the correct answer. Because of the small amount of data available (which makes the χ^2 analysis unreliable), we grouped responses into the four quarters shown in figure 9.

The shape of the quadrants results from the requirement to group neighbouring subimages, in order to maintain the consistency of the χ^2 analisys. While this grouping into only four quadrants is coarse, it is still fine enough to make the camera's pan-tilt mount focus on the relevant stimulus, for instance.

[1] 25% was chosen on the grounds that in the worst case a target object could lie exactly at the center of four sub-images, so that one quarter of the object would lie within each subframe.

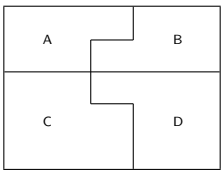

Figure 9: Sub-image groups used for χ^2 and Cramer's V analyses

A direct comparison between χ^2 analyses of different experiments is not possible, we therefore normalised the χ^2 results to the internal [0:1], using Cramer's V (Phi statistic (Sachs, 1982)). This allowed a direct quantitative comparison between different experiments. The closer the V value to 1, the stronger the correlation between system response and correct answer.

4.2 Experiment 1: Locating Individual, Known Objects

The first experiment concerned locating exactly one of three objects (blue cylinder, green box or orange ball) within the test image that had also been present in the training images.

Ten test images for each object, i.e. thirty test images in total were used to assess the performance. Success rates for the different image encoding and clustering schemes are shown in table 1.

Average Histograms			
	Blue Cylinder	Green Box	Orange Football
PCA Network	70%	80%	100%
SOFM Network	80%	100%	100%

Average Coarse Coding			
	Blue Cylinder	Green Box	Orange Football
PCA Network	80%	40%	70%
SOFM Network	50%	90%	100%

Table 1: Success rates for locating a single object that has been seen during training ("known" objects experiments)

There is a statistically significant correlation between identified and actual object location ($p = 0.05$).

4.2.1 Raw RGB encoding vs. Opposing-Channel-Encoding

In previous work we used straight RGB encoding (Vieira Neto and Nehmzow, 2003), rather than the opposing-colour-channel encoding described in section 2.1. We used Cramer's V to determine whether the former or the latter colour encoding scheme produces stronger correlations, the results are shown in tables 2 and 3.

	Average Coarse Coding	Average Histograms
PCA Network	0.63	0.65
SOFM Network	0.47	0.52

Table 2: Cramer's V for the "known" objects experiments using raw RGB channels (Vieira Neto and Nehmzow, 2003)

	Average Coarse Coding	Average Histograms
PCA Network	0.78	0.79
SOFM Network	0.90	0.96

Table 3: Cramer's V for the "known" objects experiments using the opposing-colour-channel pre-processing scheme (compare with table 2). Results are considerably better in all cases.

While both colour-encoding schemes resulted in statistically significant correlation, a comparison between tables 2 and 3 shows that the opposing-colour-channel encoding achieves superior results.

Figure 10 shows the results obtained for the test images containing the orange ball, using the histogram-based image coding and the SOFM network.

4.3 Experiment 2: Locating Known Objects With Distractors Present

The second of our experiments was conducted with the dataset which contain test images with all "known" objects. This experiment was designed to evaluate the occurence of misclassifications in the presence of distractors. However, it is in fact very similar to the first experiment and can be thought as a confirmation of the results obtained in the first experiments.

Table 4 shows the success rates obtained with the colour opponency pre-processing for the experiments with multiple "known" objects.

Once again, the χ^2 analisys for the experiments with multiple "known" objects was conducted

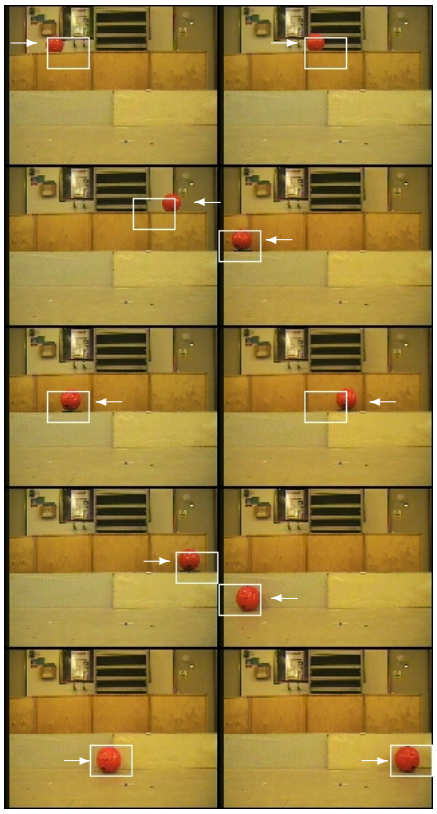

Figure 10: Sample results obtained using the histogram-based image coding and the SOFM network for an object present in the training dataset (the arrows indicate the target object and the rectangles indicate the sub-image in which the target was detected)

Average Histograms			
	Blue Cylinder	Green Box	Orange Football
PCA Network	40%	75%	60%
SOFM Network	70%	60%	60%

Average Coarse Coding			
	Blue Cylinder	Green Box	Orange Football
PCA Network	55%	75%	40%
SOFM Network	55%	65%	20%

Table 4: Success rates for the experiments with multiple "known" objects present in the test images (compare with table 1)

and have shown statistically significant correlation between actual and identified target positions for both PCA and SOFM networks ($p = 0.05$). This time, however, the test dataset was composed of 60 test images, all of which were used to

compute the χ^2 analisys.

Table 5 shows the results of the Cramer's V normalisation for the experiments with multiple "known" objects.

	Average Coarse Coding	Average Histograms
PCA Network	0.63	0.65
SOFM Network	0.61	0.78

Table 5: Cramer's V for the multiple "known" objects experiments using opposing-colour-channel (compare with table 3). It is evident that distraction affect performance adversely.

4.4 Experiment 3: Locating Unknown Objects

In the third set of experiments the system's task was to locate objects within the test image that had never been seen before, i.e. to locate objects that had not been part of the training images. An example of this is given in figure 11, which shows the results for the test images containing the red cylinder, using the histogram-based image coding and the SOFM network.

Table 6 shows the success rates obtained with the colour opponency pre-processing for the experiment with "unknown" objects, table 7 shows the results of the Cramer's V normalisation for the experiments with "unknown" objects.

Average Histograms			
	Black Circle	Red Cylinder	Yellow Box
PCA Network	60%	80%	60%
SOFM Network	20%	70%	30%

Average Coarse Coding			
	Black Circle	Red Cylinder	Yellow Box
PCA Network	30%	60%	50%
SOFM Network	10%	20%	60%

Table 6: Success rates for the experiments with "unknown" objects

Noticeable here is the "poor" performance of the SOFM, while the PCA network is well able to localise objects that were not present in the training data. Only for the PCA network there is a statistically significant correlation between identified and actual object location ($p = 0.05$), for the SOFM there is no significant correlation.

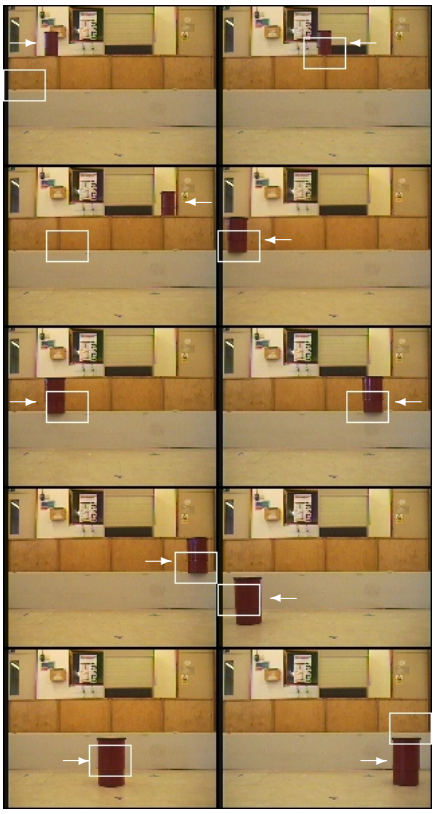

Figure 11: Sample results obtained using the histogram-based image coding and the SOFM network for an object not present in the training dataset (the arrows indicate the target object and the rectangles indicate the sub-image in which the target was detected)

	Average Coarse Coding	Average Histograms
PCA Network	0.52	0.83
SOFM Network	0.13	0.16

Table 7: Cramer's V for the "unknown" objects experiments using opposing-colour-channel (compare with tables 3 and 5)

Depending on the task, however, this result can be exploited: for novelty detection, for instance, it may be desirable to be "unable" to locate novel objects. The SOFM's disability to locate unknown objects might therefore be used to advantage.

5. Conclusions

There are good reasons to be interested in novelty detection in mobile robot control: focussing of computational and memory resources onto rel-

evant aspects of a task, the exploitation of opportunities, as well as the unsupervised execution of tasks such as surveillance or inspection.

In previous work (Marsland et al., 2000) we have presented a novelty filter that was able to detect novelty in sonar sensor signals. While this proved useful for all aspects of the robot's operability, the resolution of sonar sensors is too coarse to support tasks such as inspection or surveillance. We are therefore interested to use visual stimuli as inputs to a novelty filter.

In this paper, we have presented an image processing mechanism that is based on self-organisation, and does not use any pre-installed knowledge. Because novelty detection by definition deals with *a priori* unknown stimuli, processing through self-organisation is, arguably, the method of choice.

In contrast to previous work, where we used straight RGB colour channels as input (Vieira Neto and Nehmzow, 2003), we have used opposing colour channels (Itti et al., 1998) as input here. Our results demonstrate that the mechanism is able to locate target objects, and that it is robust: target objects were successfully located, even in the presence of background clutter and under differing lighting conditions.

We used two different image-encoding mechanisms — histograms and coarse coding — and two different clustering mechanisms — SOFM and PCA network. For locating "known" objects within the image frame, all combinations provided statistically significant correlation between estimated and true target location ($p = 0.05$), the same was true when distractors (known objects that were not targets) were present in the image. In both cases, the SOFM using histogram encoding provided the best results. In a third experiment, which involved locating objects that had not been seen during training, only the PCA net provided statistically significant correlations. Again, the histogram encoding provided better results than coarse coding.

In summary of the experimental results, histogram encoding provided the best results in all cases. The SOFM performs better than the PCA net in locating "known" objects, while the latter is better at locating "unknown" objects. We believe that this fact can later be exploited in novelty detection tasks.

Our future work will apply the image processing techniques presented here to novelty detection, incorporating information about shape and orientation as well.

Acknowledgement

Hugo Vieira Neto is sponsored by a CEFET-PR and CAPES Foundation fellowship, whose support is gratefully acknowledged.

References

Ballard, D. H. (1997). *An Introduction to Natural Computation*. MIT Press, Cambridge, MA.

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.

MacIlwain, J. T. (1996). *An Introduction to the Biology of Vision*. Cambridge University Press, Cambridge, UK.

Marsland, S., Nehmzow, U., and Shapiro, J. (2000). Detecting novel features of an environment using habituation. In *From Animals to Animats: Proceedings of the 6th International Conference on Simulation of Adaptive Behavior (SAB'2000)*, pages 189–198, Paris, France. MIT Press.

Sachs, L. (1982). *Applied Statistics*. Springer Verlag, New York, NY.

Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473.

Vieira Neto, H. and Nehmzow, U. (2003). Object localisation and tracking through subsymbolic classification. In *Proceedings of the AISB'03 Symposium on Biologically-Inspired Machine Vision, Theory and Application*, Aberystwyth, UK.