Incremental PCA:
An Alternative Approach for Novelty Detection

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

Abstract

Exploration and inspection of dynamic environments using mobile robots are applications that benefit immensely from novelty detection algorithms. In this paper we propose the use of a new approach for on-line novelty detection based on incremental Principal Component Analysis and compare its performance and functionality with a previously studied technique based on a GWR neural network. We have conducted a series of experiments using visual input from a mobile robot interacting with a controlled laboratory environment that show advantages and disadvantages for each method.

1. Introduction

The ability to differentiate between normal and abnormal sensory perceptions is a desirable competence for mobile robots operating in dynamic environments – usually uncommon features carry the most useful information and therefore deserve to be analysed in more detail. Novelty detection is of interest to any agent aiming at true autonomy, continuous operation and adaptability to new situations through on-line unsupervised learning.

From an application point of view, reliable novelty detection systems would facilitate the implementation of automated inspection and surveillance. As for these tasks one commonly desires to detect any *previously unknown* feature – as opposed to recognition tasks in which features of interest are already known – the feasible approach to be followed is to learn a model of normality and use it to filter out abnormal perceptions.

Previous work has demonstrated that the approach of learning a model of normality instead of abnormality is very effective in mobile robots that use sonar readings as perceptual input (Marsland et al., 2002). This work resulted in the development of the Grow-When-Required (GWR) neural network, a self-organising learning mechanism able to determine whether an input is novel or not through the use of a model of habituation.

The GWR network has shown to work well with low-dimensional input data, such as a vector of sonar measurements, but in the past few years we have been interested in investigating the scalability of the GWR approach to another rather different sensor modality: vision. Our interest in using vision for novelty detection comes from the wide range of different information from the environment that it can provide the robot with.

A major difficulty that comes with vision is how to select which aspects of the data are important to be encoded, as it is undesirable to process high-dimensional data directly. So far, we have successfully employed a mechanism of attention to select salient locations of the input image frame and perform some image encoding in their vicinity (Vieira Neto and Nehmzow, 2004, Nehmzow and Vieira Neto, 2004). The purpose of the image encoding stage is to reduce dimensionality of input data to the novelty filter while trying to preserve discriminability between different classes of features as much as possible.

Designing the image encoding stage is a hard task as it is not always clear to the designer which aspects of the data are important to be encoded. It would be more desirable to select which aspects are the most relevant in a bottom-up approach and hence our interest in the Principal Component Analysis (PCA) algorithm. PCA consists in projecting the data onto principal axes – the axes in which variance is maximised – obtained from the data itself.

The standard PCA algorithm is usually computed in batch mode and needs all the input data to be available at once, making it unsuitable for on-line learning systems. Recently, however, an algorithm for incremental computation of PCA was introduced by Artač *et al.* (Artač et al., 2002) and we have identified its potential for novelty detection.

In this paper we implement an incremental PCA novelty filter and compare its performance with the already well-known GWR-based approach. Figure 1 shows the block diagram that illustrates the framework in which both methods were tested.

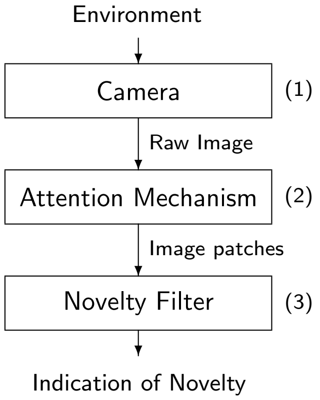


Figure 1: The visual novelty detection framework: salient image patches are selected within the images acquired from the environment and then either of the novelty filters is used to indicate the presence of novelty.

In order to generate the input for the novelty filters we have employed a visual attention mechanism to select smaller image patches within the input image frame. We briefly describe the attention mechanism in the next section, followed by both techniques for novelty detection. Finally, we present in detail the experimental setup in which we conducted our tests and discuss the results obtained.

2. Visual attention

The high dimensionality of visual data imposes serious constraints to real-time processing, especially when using mobile robots with restricted computational resources. It is therefore prohibitive to process an entire image frame, even when using low resolution – the use of some sort of dimensionality reduction scheme that preserves important and distinctive features of the visual data is needed before any higher level processing can be done.

Moreover, images acquired from a moving platform are subject to several geometrical transformations resulting from changes in perspective. Therefore, the naïve approach of comparing entire image frames directly would most definitely not work properly, leading to the misclassification of known features as novel due to simple image transformations.

In this paper we have used the saliency map (Itti et al., 1998) as a mechanism of visual attention to select distinctive regions in the input image frame (152×120 pixels in size). The saliency map consists in the combination of several multi-scale feature maps in intensity, colour and orientation of visual features, allowing the detection of conspicuous locations within the image frame that are generally robust to geometric transformations. Details of our implementation of the saliency map can be found in (Vieira Neto and Nehmzow, 2004).

We have used the nine highest values of the saliency map for each input frame to select the most "interesting" image regions so that image patches can be extracted from their vicinity. The image patches used in the experiments reported here have 24×24 pixels in size. As we have used RGB colour images, the resulting input vectors for the two alternative novelty filters being compared have $24 \times 24 \times 3 = 1728$ elements. Input vectors were normalised before being fed to the novelty filters to even out lighting conditions.

As in this work the similarity between image patches is computed in a pixel-by-pixel basis using the Euclidean distance, it is important to have the patches aligned as much as possible to minimise errors. The saliency map has an important role in this task as the location of salient points tend to be very stable and therefore minimises image patch misregistration.

At the moment our current implementation only handles translation effects, but we are currently investigating algorithms to achieve invariance to scale and rotation (Lowe, 2004), and also invariance to affine transformations (Mikolajczyk and Schmid, 2002). In this paper, generalisation with respect to scaling, rotation and affine transformations is achieved by means of the acquisition of several vectors for geometrically transformed versions of the same visual features.

3. Novelty filters

3.1 The GWR network

The Grow-When-Required (GWR) is an artificial neural network that was especially designed for the task of on-line novelty detection (Marsland et al., 2002). It combines a clustering mechanism and a model of habituation to decide if a determined input is novel and therefore needs to be incorporated to the current model. A summary of the operation of the GWR network is given in Algorithm 1.

The network is initialised with two unconnected nodes, whose weights are set to the two first input vectors. As input vectors are presented, the algorithm is able to decide which represent novel features based on how well they match existing habituated nodes, building and exploiting a topological map in input space.

We have previously used the GWR network for visual novelty detection (Vieira Neto and Nehmzow, 2004, Nehmzow and Vieira Neto, 2004), using local colour histograms as input vectors to achieve dimensionality reduction. In the experiments reported here, we used normalised raw image patches in order to fairly compare performance and functionality with an alternative novelty detection algorithm based on incremental PCA.

In this work, the parameters used for the GWR net-

Algorithm 1: GWR network novelty detection

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

Output: updated set of nodes A, updated set of connections C, novelty indication N.

1 Find the best and second best matching nodes s and t: $s = \arg\min_{i \in A^{(n)}} \| \mathbf{x} - \mathbf{w}_i \|$, $t = \arg\min_{i \in A^{(n)}/\{s\}} \| \mathbf{x} - \mathbf{w}_i \|$, where $\mathbf{w}_i$ is the weight vector of the node i.
2 If there is a connection between s and t, set its age to zero, otherwise create it: $C = C \cup \{(s,t)\}$).
3 Compute the activity of the best matching node: $a_s = \exp(- \| \mathbf{x} - \mathbf{w}_s \|)$.
4 Test if the activity and habituation values of the best matching node characterise novelty:
if $a_s < a_T$ *and* $h_s < h_T$ **then**
5   Add a new node: $A = A \cup \{r\}$.
6   Set the weight vector of the new node: $\mathbf{w}_r = (\mathbf{x} + \mathbf{w}_s)/2$.
7   Create connections between the new node and the best and second best matching nodes: $C = C \cup \{(r,s),(r,t)\}$.
8   Remove the connection between the best and second best matching nodes: $C = C/\{(s,t)\}$.
9   Indicate novelty detected: $N = 1$.
10 **end**
11 **else** Indicate no novelty detected: $N = 0$.
12 Compute the activity of the best matching node's neighbours (nodes with connections to the best matching node): $a_n = \exp(- \| \mathbf{x} - \mathbf{w}_n \|)$.
13 Adapt the positions of the best matching node and its neighbours: $\mathbf{w}_s = \mathbf{w}_s + \epsilon(\mathbf{x} - \mathbf{w}_s)$, $\mathbf{w}_n = \mathbf{w}_n + \frac{\eta a_n}{a_s}\epsilon(\mathbf{x} - \mathbf{w}_n)$.
14 Age connections to the best matching node: $age_{(s,n)} = age_{(s,n)} + 1$.
15 Habituate the best matching node and its neighbours: $\tau \frac{dh_s(t)}{dt} = \alpha[h_0 - h_s(t)] - S(t)$, $\frac{a_s}{\eta a_n}\tau \frac{dh_n(t)}{dt} = \alpha[h_0 - h_n(t)] - S(t)$.
16 Remove any nodes without any neighbours.
17 Remove any connections with age greater than age_{max}.

work were: $a_T = 0.9$, $h_T = 0.3$, $\eta = 0.1$, $\epsilon = 0.1$, $\tau = 3.33$, $\alpha = 1.05$, $h_0 = 1$, $S(t) = 1$ and $age_{max} = 20$.

3.2 *Incremental PCA*

Principal Component Analysis (PCA) is a very useful tool for dimensionality reduction that allows optimal reconstruction of the original data. It consists on projecting the input data onto its principal axes and is usually computed off-line, as the standard algorithm requires that all data samples are available *a priori*, making it unsuitable for applications that demand on-line learning.

However, a method for the incremental computation of PCA recently introduced by Artač *et al.* (Artač et al., 2002) makes simultaneous learning and recognition possible. Their technique allows the original data to be discarded immediately after the eigenspace is updated, storing only the (reduced dimension) data projected onto it.

In this work we employ the method proposed by Artač *et al.* to perform on-line novelty detection, using the magnitude of the residual vector, i.e. the RMS error between original data and the reconstruction of its projection onto the current eigenspace, to classify the input as novel or not. Algorithm 2 summarises how this approach is implemented.

Algorithm 2: Incremental PCA novelty detection

Input: current mean vector $\boldsymbol{\mu}^{(n)}$, current eigenvectors $\mathbf{U}^{(n)}$, current projected vectors $\mathbf{A}^{(n)}$, new input vector $\mathbf{x}$, residual threshold r_T.

Output: updated mean vector $\boldsymbol{\mu}^{(n+1)}$, updated eigenvectors $\mathbf{U}^{(n+1)}$, updated projected vectors $\mathbf{A}^{(n+1)}$, novelty indication N.

1 Compute the projection of the new input vector using the current basis: $\mathbf{a} = \mathbf{U}^{(n)\mathsf{T}}(\mathbf{x} - \boldsymbol{\mu}^{(n)})$.
2 Compute the reconstruction of the new input vector: $\mathbf{y} = \mathbf{U}^{(n)}\mathbf{a} + \boldsymbol{\mu}^{(n)}$.
3 Compute the residual vector (orthogonal to $\mathbf{U}^{(n)}$): $\mathbf{r} = \mathbf{x} - \mathbf{y}$.
4 Test if the magnitude of the residual vector is large enough to characterise novelty:
if $\| r \| > r_T$ **then**
5   Append residual vector as a new basis vector: $\mathbf{U}' = \begin{bmatrix} \mathbf{U}^{(n)} & \frac{\mathbf{r}}{\|\mathbf{r}\|} \end{bmatrix}$.
6   Append projected vector: $\mathbf{A}' = \begin{bmatrix} \mathbf{A}^{(n)} & \mathbf{a} \\ \mathbf{0} & \| \mathbf{r} \| \end{bmatrix}$.
7   Perform batch PCA on $\mathbf{A}'$, obtaining its mean vector $\boldsymbol{\mu}''$ and eigenvectors $\mathbf{U}''$.
8   Update projected vectors using the new basis: $\mathbf{A}^{(n+1)} = \mathbf{U}'(\mathbf{A}' - \boldsymbol{\mu}''\mathbf{1}_{1\times n+1})$.
9   Update eigenvectors: $\mathbf{U}^{(n+1)} = \mathbf{U}'\mathbf{U}''$.
10   Update mean vector: $\boldsymbol{\mu}^{(n+1)} = \boldsymbol{\mu}^{(n)} + \mathbf{U}'\boldsymbol{\mu}''$.
11   Indicate novelty detected: $N = 1$.
12 **end**
13 **else** Indicate no novelty detected: $N = 0$.

The algorithm is made completely incremental by initialising the eigenspace and projected vectors as follows: $\boldsymbol{\mu}^{(1)} = \mathbf{x}^{(1)}$, $\mathbf{U}^{(1)} = \mathbf{0}_{M\times 1}$ and $\mathbf{A}^{(1)} = 0$, where $\mathbf{x}^{(1)}$ is the first input vector and $\mathbf{0}_{M\times 1}$ denotes an $M \times 1$ matrix of zeros, M being the dimensionality of the input.

In this approach, dimensionality reduction is achieved by exploiting the fact that the number of eigenvectors in the model are likely to be less in number than the dimensionality of the input vectors ($24 \times 24 \times 3 = 1728$). If all eigenvectors are kept in the model, perfect reconstruction of the original data is achieved. This functionality allows the user to reconstruct the input image patches from the stored projected vectors and have a perfect notion of which aspects of the environment were learnt.

Further dimensionality reduction can be achieved by keeping only the eigenvectors corresponding to the $k < n$ largest eigenvalues in the model at the expense of losses in reconstruction (and possibly in the recognition rate of the system). The selection of eigenvectors can be done while computing the batch PCA during learning (step 7 in Algorithm 2). In this work we have set the threshold for the magnitude of the residual vector as $r_T = 0.2$ and we kept only eigenvectors whose corresponding eigenvalues were larger than 1% of the largest eigenvalue.

4. Experiments

In order to compare the performance of the aforementioned novelty detection mechanisms we have designed a series of experiments consisting of two different phases: an exploration phase in which the robot learns a visual model of normality for its operating environment; and an inspection phase in which the learnt model is used to highlight any abnormal visual feature that may appear in the environment.

We have built a bounded square arena with cardboard boxes and then used our Magellan Pro mobile robot to collect images while navigating using a simple obstacle avoidance behaviour. Although both novelty filters being compared in this paper are able to run in real-time, the images were collected for off-line processing in order to make fair comparisons by using the same dataset for both methods.

The performance of each method was evaluated by relating the actual presence of novelty and the response of the system in the form of a contingency table and later performing the χ^2 analysis, followed by the computation of Cramer's V and the uncertainty coefficient U. We have also compared the size of the models acquired by the two approaches.

Ground truth for the presence of novelty in each frame was generated manually in the form of binary images with the novel features highlighted. Any image patch selected by the mechanism of visual attention containing at least 10% of highlighted pixels was considered as novel for the purposes of performance analysis. Every round of the exploration or inspection phases consisted of five consecutive loops around the arena.

4.1 Results

The first round of experiments comprised the robot exploring the empty arena in order to build a model of normality for it. After that, the robot inspected the arena when two different objects were introduced in it: a very conspicuous orange football and a much less conspicuous grey box. The results obtained for both approaches are given in Table 1.

Table 1: Experiment 1 - exploration of the empty arena and inspection of the arena containing novel objects (orange football and a grey box).

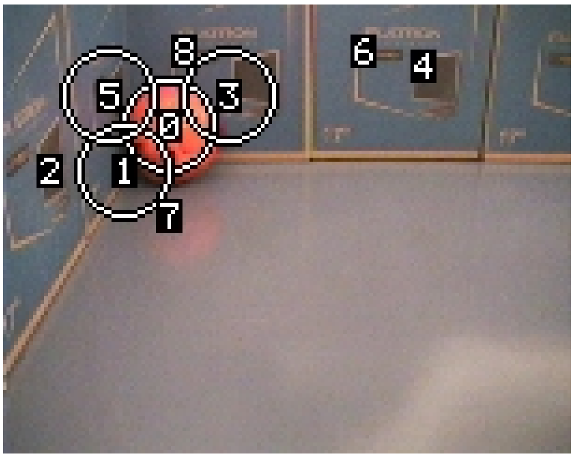
GWR Network		
	Novelty Detected	Novelty Not Detected
Novelty Present	102	14
Novelty Not Present	19	4365
Cramer's $V = 0.86$		
Uncertainty Coefficient $U = 0.73$		

Incremental PCA		
	Novelty Detected	Novelty Not Detected
Novelty Present	104	12
Novelty Not Present	15	4369
Cramer's $V = 0.88$		
Uncertainty Coefficient $U = 0.76$		

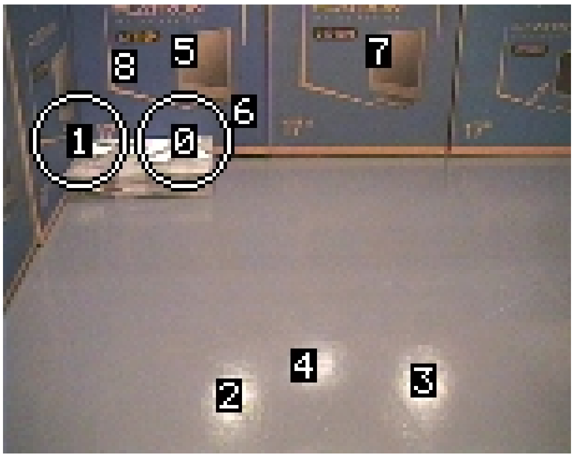
Both systems were able to highlight the novel objects, yielding statistically significant results according to the χ^2 analysis ($p < 0.05$). It can be noticed from Table 1 that results obtained for both systems were very similar and that the contingency table entries add up to 4500 samples, which correspond to 9 salient regions per image frame in a total of 500 frames (50 image frames per loop in a total of 10 loops around the arena).

Figure 2 visually depicts the results obtained for frames where the orange football and the grey box appeared, respectively. For the particular image frames shown in Figure 2, results were exactly the same for both approaches.

In order to establish whether or not the novelty detection systems were able to highlight novel but not very salient features, a second round of experiments was conducted. This time the robot has explored the arena containing the orange football and inspected it with the inclusion of the grey box in two different situations: first in a different location and then in the same location as the ball. Both systems were able to successfully detect the novel object in each situation, regardless of the presence



(a)



(b)

Figure 2: Results obtained with the visual novelty detection mechanism in experiment 1: (a) the orange football as new object; and (b) the grey box as new object. White circles designate image patches labelled as novel by the system, while numbers indicate salient locations within the image frame (lower numbers indicate an item of greater saliency).

or not of the much more salient orange ball in the same image frame, as shown in Figure 3.

The results obtained for this round of experiments are given in Table 2. Again statistically significant (χ^2 analysis, $p < 0.05$) for both systems, but slightly better for the incremental PCA.

Finally, we conducted a third round of experiments in order to evaluate the ability of the systems to differentiate between visually similar objects. This time the robot explored the arena containing the grey box and inspected it with another grey box, identical in colour and texture, but larger in size. Two instances were analysed: the larger box in the same location as the previous smaller box was and also in a different location in the arena.

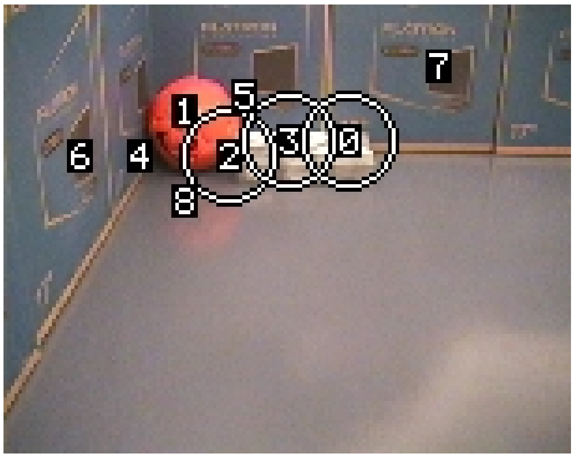


Figure 3: Results obtained with the visual novelty detection mechanism in experiment 2: the grey box is correctly detected as novel regardless of the presence of the much more salient but already known orange football. The output shown is the one provided by the incremental PCA approach, as the GWR-based approach only labelled region 2 as novel.

Table 2: Experiment 2 - exploration of the arena containing the orange football and inspection of the arena containing a novel object (a grey box in the same location as the ball and in a different location as the ball).

GWR Network		
	Novelty Detected	Novelty Not Detected
Novelty Present	77	90
Novelty Not Present	23	4310
Cramer's $V = 0.58$		
Uncertainty Coefficient $U = 0.31$		

Incremental PCA		
	Novelty Detected	Novelty Not Detected
Novelty Present	102	65
Novelty Not Present	22	4311
Cramer's $V = 0.70$		
Uncertainty Coefficient $U = 0.44$		

Both approaches were able to correctly detect the larger grey box as novel and the results obtained are given in Table 3. Once more, results were statistically significant (χ^2 analysis, $p < 0.05$) for both systems, but better for incremental PCA. Figure 4 shows an example of output obtained during the third round of experiments.

Table 3: Experiment 3 - exploration of the arena containing the grey box and inspection of the arena containing a novel object (a larger grey box visually similar to the original box in the same location and in a different location).

GWR Network		
	Novelty Detected	Novelty Not Detected
Novelty Present	49	213
Novelty Not Present	12	4226
Cramer's $V = 0.37$		
Uncertainty Coefficient $U = 0.11$		

Incremental PCA		
	Novelty Detected	Novelty Not Detected
Novelty Present	92	170
Novelty Not Present	17	4221
Cramer's $V = 0.53$		
Uncertainty Coefficient $U = 0.23$		

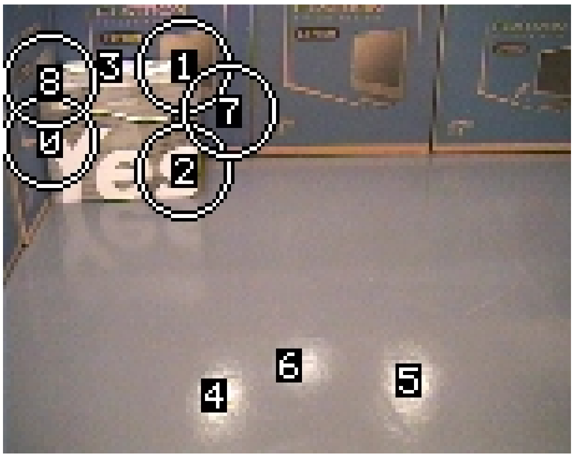


Figure 4: Results obtained with the visual novelty detection mechanism in experiment 3: the larger grey box is correctly detected as novel. The output shown is the one provided by the incremental PCA approach, as the GWR-based approach only labelled regions 0 and 8 as novel.

4.2 Discussion

The results given in Tables 1, 2 and 3 demonstrate the statistical significance in the association of the responses of both approaches and the actual novel features introduced in the robot's original environment. Although the strength of the association measured by Cramer's V and the uncertainty coefficient U can be thought as modest, we consider these values as very conservative for two reasons. First, if we take into account the consistency of novelties detected between successive image frames we are able to rule out most false positives (novelty detected but not present). And second, most false negatives (novelty present but not detected) can be eliminated by the fact that a single image patch within the new object labelled as new is enough to characterise the entire object as novel. Nevertheless, the values of V and U serve well to the purpose of comparing performances.

Results given by both approaches are similar in performance, although the size of the models acquired by each one are very different for the set of parameters used: in experiment 1 the GWR acquired only five vectors while the incremental PCA acquired 35; in experiment 2 the GWR acquired a total of 18 vectors and the incremental PCA 45; finally, in experiment 3, 11 vectors were acquired by the GWR and 47 by the incremental PCA.

The small amount of vectors learnt by the GWR had always the original input dimensionality (1728 elements), while the dimensionality of the vectors acquired by incremental PCA varied from 29 to 33 dimensions. However, every dimension of the projected vectors acquired by the incremental PCA approach corresponds to an eigenvector with 1728 elements.

Dimensionality issues become important when we consider that the Euclidean metric was used to determine similarity between vectors. When the Euclidean distance is used, a small difference between two high-dimensional vectors tend to be large in value, making it difficult to establish thresholds of similarity for high-dimensional spaces, as it is the case with the vectors acquired by the GWR network.

The incremental PCA approach offers a clear advantage as similarity between inputs is performed by the residual error in reconstruction from the projected space. Moreover, if a direct comparison of projected vectors is to be made, substitution of the Euclidean distance by the Mahalanobis distance can be easily implemented in the incremental PCA approach once the covariance matrix of the stored projected vectors is available as a sub-product of the method (step 7 in Algorithm 2). The Mahalanobis distance normalises the contribution of vector elements according to the covariance matrix of the data:

$$d_{\mathbf{xy}} = \sqrt{(\mathbf{x}-\mathbf{y})^{\mathsf{T}}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{y})} \quad (1)$$

where $d_{\mathbf{xy}}$ is the Mahalanobis distance between the column vectors $\mathbf{x}$ and $\mathbf{y}$ and $\mathbf{C}$ is the covariance matrix of the data. Euclidean distance corresponds to the special case where $\mathbf{C}$ is the identity matrix.

Another advantage of the incremental PCA approach is the ability to automatically reduce dimensionality allowing reconstruction of the original input image patch from the inverse transformation of the corresponding projected vector. Therefore, the user can evaluate which aspects of the environment were actually learnt by the system. Reconstruction of the stored vectors in the GWR network resulted in averaged image patches resulting from the learning procedure (step 13 in Algorithm 1).

The GWR network, however, has the advantage of building a topological map for the stored vectors, through connections between similar patterns. We have experimented increasing the number of stored vectors in the GWR approach by raising the activation threshold a_T in order to acquire a number of vectors as close as possible to the number of vectors acquired by the incremental PCA. This resulted in better reconstruction of the stored vectors, but has also sensibly decreased the overall performance of the GWR-based system. As one would expect, the number of false negatives has decreased, but on the other hand the number of false positives has increased immensely. We attribute this effect to the use of Euclidean distance in a high-dimensional space.

5. Conclusion

We have presented an alternative to perform on-line novelty detection using an incremental PCA approach and compared its performance against a previously studied approach based on the GWR network. The proposed incremental PCA approach provides slightly better performance, while offering advantages of embedded dimensionality reduction and good reconstruction ability, which is extremely useful to assess which aspects of the environment were actually learnt by the system.

The inability of the GWR network of evaluating similarity between inputs in reduced dimensions normally forces the system designer to use an additional preprocessing stage for dimensionality reduction, such as the use of colour statistics (Vieira Neto and Nehmzow, 2004). On the other hand, the GWR approach offers the functionality of constructing a topological relationship between inputs. Future investigations aim at combining the dimensionality reduction feature of the incremental PCA with the topological construction algorithm of the GWR network using the Mahalanobis distance as a measure of similarity between patterns.

Considering the overall system functionality, the attention mechanism plays an important role in generalisation by providing image patches that are robust to translations and therefore reducing the number of stored vectors. We are currently studying alternatives for the attention model that also offer invariance to scale and rotation (Lowe, 2004), and also affine transformations (Mikolajczyk and Schmid, 2002). Extensions to the incremental PCA algorithm to make it robust to occlusions (Skočaj and Leonardis, 2003) are also attractive for future work.

A final contribution of this paper is the introduction of a method to evaluate performance of novelty detection systems based on contingency table analysis. A quantitative assessment can be made by the computation of Cramer's V and the uncertainty coefficient U, while the statistical significance of the association between system response and actual novelty status can be made by χ^2 analysis.

Acknowledgements

Hugo Vieira Neto is sponsored by the Brazilian Government through CAPES Foundation and CEFET-PR, whose support is gratefully acknowledged.

References

Artač, M., Jogan, M., and Leonardis, A. (2002). Incremental PCA for on-line visual learning and recognition. In *Proceedings of the 16th International Conference on Pattern Recognition (ICPR'2002).*

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.

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

Marsland, S., Nehmzow, U., and Shapiro, J. (2002). 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.

Mikolajczyk, K. and Schmid, C. (2002). An affine invariant interest point detector. In *Computer Vision - ECCV 2002: 7th European Conference on Computer Vision (ECCV'2002)*, volume 1, pages 128–142, Copenhagen, Denmark.

Nehmzow, U. and Vieira Neto, H. (2004). Novelty-based visual inspection using mobile robots. In *Towards Autonomous Robotic Systems: Proceedings of the 5th British Conference on Mobile Robotics (TAROS'04)*, Colchester, UK.

Skočaj, D. and Leonardis, A. (2003). Weighted and robust incremental method for subspace learning. In *Proceedings of the 9th International Conference on Computer Vision (ICCV'03)*, pages 1494–1501.

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