Enhancing Active Vision System Categorization Capability through Uniform Local Binary Pattern

Olalekan Lanihun, Bernie Tiddeman, Elio Tuci and Patricia Shaw

Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, United Kingdom.
http://www.aber.ac.uk

Abstract. Previous research in Neuro-evolution controlled Active vision systems has shown its potential to solve various shape categorization and discrimination problems. However, minimal investigation has been done in using this kind of evolved system in solving more complex vision problems. This partly due to variability in lighting conditions, reflection, shadowing etc, which may be inherent to these kind of problems. It could also be due to the fact that building an evolved system for these kind of problems may be too computationally expensive. We present an active vision system controlled neural network trained by a genetic algorithm that can autonomously scan through an image, pre-processed by Uniform Local Binary Pattern, [8] method. We demonstrate the ability of this system to categorize more complex images taken from the camera of a Humanoid (iCub) robot. Preliminary investigation results show that the proposed Uniform Local Binary Pattern [8] method performed better than the grayscale averaging method of [1] in the categorization tasks. This approach provides a framework that could be used for further research in using this kind of system for more complex image problems.

Keywords: Categorization, Neural Network, Genetic Algorithm, Uniform Local Binary Pattern

1 Introduction

Active vision is the process of exploring visual scene in order to obtain relevant features for subsequent meaningful and intelligent processing. This is very important and very useful in that visual systems usually have a form of control and are intelligently guided to only those areas of the image surface being processed that have relevant and valuable information to the task at hand. The control of the visual system can be done by various techniques, although, it is natural to use a neural network, because of its biological based inspiration and also their suitability for noisy data. However, developing an active vision system, particularly using the approach of evolving a neural network is still in its elementary stage [5]. In most cases, only simple vision problems have been solved using this approach, which could be attributed to inherent illumination conditions such

as reflection and shadowing in natural images, and also the computational cost that generally comes with using evolutionary techniques for more complex image problems. As a result, when the problem domain becomes more complex, the dimension of the feature vector input to the network increases, and therefore the benefits from this kind of system are soon outweighed by the computational cost. Consequently, categorization using active vision have being used for more simple vision problems and discrimination of very few stimuli. For instance in [1], an active vision system based genetic algorithm evolved neural network was used for categorizing five gray value italic letters. In [5], an active vision system of genetic algorithm evolved neural controller was used for basic 2D shape discrimination. In an attempt to overcome these problems , we have used Uniform Local Binary Pattern [8] for feature extraction and enhancement of more complex images taken from a Humanoid (iCub) robot camera. This can filter out to an appreciable degree impacts of image lighting conditions such as reflection and shadowing, and also reduces the feature vector size that is input into the network.

2 Related Works

The field of Evolved Active Vision Systems for categorization has been extensively studied. Mirolli and Nolfi [1] used an active vision system that is based on a genetic algorithm evolved neural network to categorize gray level italic alphabet letters in different scales (sizes). The movement of the artificial eye was controlled by motor neurons of the output units, which determine the eye location per time step, in-order to capture relevant input features for the neural controller. James and Tucker [5], developed an active vision system that is able to discriminate different 2D shapes by moving about in any direction with an ability to zoom and rotate. The system was able to discriminate different 2D shapes irrespective of their scales, location and rotation. An Active Vision System controlled by an evolved recurrent neural network was developed by Morimoto and Ikegami [6] that dynamically discriminates between rectangular and triangular objects. In this system when the agent moves through the environment, it develops neural states which are not just a symbolic representation of rectangles or triangles, but allow it to distinguish these objects. In the same vein Aditya and Nakul [2], used a neuro-evolution based active vision system to discriminate a target shape. The artificial retina used in their system has the ability to translate in co-ordinate X and Y directions, zoom-in and zoom-out, and ability to rotate as it scans over the image features. However in their work they introduced constraints to the environment of the active vision based system. The constraints to the environment are implemented in the form of force field in a certain direction. At each time step during the training and evaluation, a unit force is exerted on the artificial agent by the force field. This implies that at each step, the agent is forced to move a unit direction in the direction of the force field. Consequently, the actual movement of the agent per time step is determined by the vector sum of the change of location in X and Y direction as well as the force movement.

The constraints were added in order to make the system closer to the real world and also provides an opportunity to observe if the system is able to develop intelligent strategies for coping with them. In all the experiments, the system was able to perform better in the discriminating tasks, despite the constraints introduced. Floreano et al [3], also implemented an active vision based system that autonomously scans through gray scale images and was able to discriminate triangular shapes from square shapes. The images used in their experiment varied in scale and location. Finally, in relation to other research works listed above , the approach in this paper also uses an active vision system based on a genetic algorithm controlled neural network. We have adopted a similar approach used by Mirolli and Nolfi [1], but extended, with the enhancement of the images with Uniform Local Binary Pattern [8] to categorize more complex images from a Humanoid (iCub) robot camera.

3 Experimental Details

We have used a biologically inspired active vision system that combines sensorimotor information in order to determine the task done by an artificial agent. The artificial agent is provided with a movement eye that explores a visual scene (image) in other to extract relevant information in order to process the sensory stimuli. The vision system is controlled by a recurrent neural network evolved by a genetic algorithm, which is similar in approach to [1]. We have adopted the same fitness function used by Mirolli and Nolfi [1], but of a slightly different recurrent neural network architecture, of similar update equations in [7]. We have also adopted the periphery only architecture of [1], which also gave the best results in all the different architectures used in their experiments. Hereafter, we shall refer to the entire eye region as periphery in the remaining part of this paper. We have done three sets of experiments, which are: (i) the replication of the periphery only architecture of the original active vision system experiment presented in [1] for the categorization of five italic letters i.e , (l, u, n, o, j), which uses the grayscale averaging method of the pixel values of the periphery region (ii) our proposed method of pre-processing the periphery region with Uniform Local Binary Pattern [8] of the adopted periphery only architecture in [1], for the categorization of the objects on more complex images taken from Humanoid (iCub) robot camera, namely: soft toy, tv remote control set, microphone, board wiper and hammer, (iii) the periphery only architecture , using grayscale averaging of the pixel values, but in this case is used to categorize the same set of objects of images taken from Humanoid (iCub) robot camera. The neural network, evolutionary process and the fitness function are the same for the three experiments, only that the second experiment have a different input vector size as its the visual features are being processed by Uniform Local Binary Pattern [8] and that of the first and third experiments processed by grayscale averaging are the same. The number of trials and generations in the second and third experiment are also the same (10 and 5000 respectively), while that of the first experiment are (50 and 3000 respectively) . In each experiment we evaluated the

performance of the system based on its ability to correctly label the category of the letters or the objects. The three experiments were undertaken so as to do a quantitative and qualitative comparisons.

The Neural Network The Neural Network is a recurrent architecture that consists of one input layer of which vector size is determined by the type of visual features being processed, that is 243 in the case of Uniform Local Binary Pattern [8] and 32 for the grayscale methods. It also has 5 hidden recurrent neurons and 7 output neurons , 2 of which determines the movement of the eye per time step of maximal displacement of $[-12; 12]$ pixels in X and Y directions and the other 5 neurons for labelling of the category of the letters in case of experiment one and the category of the objects in the case of experiment two and three. The input layer consists of units which encodes the current state of activations of the neurons for the visual stimuli of the periphery region and the efferent copies of the 2 motor neurons and the 5 categorization units at previous time step $t-1$, (see experiment one and Fig. 1 for the grayscale letter categorization , experiment two and Fig. 2 for the proposed Uniform Local Binary Paterrn method [8] , and experiment three and Fig. 3 for graryscale method used for the same problem of object categorization on images taken from Humanoid (iCub) robot camera). The activations of the input neurons are normalized between 0 and 1 and a random value with a uniform distribution between the range of [-0.05;0.05] is added to those of the grayscale methods at each time step in order to take account the fact that gray level measured by the periphery is subject to noise. The outputs of the neurons in the hidden layer depend on the input received from the input neurons through the weighted connections and the activations of the hidden neurons at previous time step. The input activations scaled by the gained factor are represented by equation (1) below:

$$y_i = gI; i = 1, ., k; \tag{1}$$

Where k stands for the size of the input vector, parameter I in the equation represents the activation values of the input, y_i is the activation values of input scaled by the gain factor g. The update equation for the hidden nodes is as represented in the equation (2) below:

$$\tau_i \partial y_i = -y_i + \sum_{i=1}^{n} w_{ji} \sigma(y_j + \beta_j); i = 1, ., 5; \tag{2}$$

The update equation (2) for the hidden neurons is a differential equation. τ is the decay constant, y_i, is the outputs of hidden neurons at previous time step $t-1$, n is the total number of the input and the hidden neurons, w_{ji} is the weight of connections from input nodes to hidden nodes and $\sigma(y_j + \beta_j)$ is the firing rate, where β_j stands for the bias terms. i is the number of hidden neurons and j is the number of input neurons. Equation (3) below is used to compute

the output activations

$$y_i = \sum_{j=1}^{5} w_{ji}\sigma(y_j + \beta_j); i = 1, ., 7; \tag{3}$$

where y_i is the activations of the output neurons , w_{ji} is the connection weights from the hidden to output units, while i is the number of output neurons and j is the number of hidden neurons. σ is the sigmoid function used as shown in the equation (4)

$$\sigma(c) = \frac{1}{(1 + e^{-c})^{-1}} \tag{4}$$

The Evolutionary Task In each trial the eye is left to freely explore the image, however, a trial is terminated when the eye can no longer perceive any part of the letter or the object through the periphery vision for three consecutive times steps. The task of the agent is to correctly label the category of the current letter or object during second half of the trial, i.e when the agent has explored the image for enough time. The fitness is only computed in the second half of a trial with the fitness function shown below:

$$F = \frac{\sum_{t=1}^{nT}\sum_{c=sFC}^{nC}\left(0.5 * 2^{-rank} + 0.5 * \left(y_r^t * 0.5 + \sum_{y \in y_w^t}(1-y) * \frac{0.5}{nOL-1}\right)\right)}{nT * (nC - sFC)} \tag{5}$$

where 2^{-rank} rewards the agent's ability to activate the categorization unit corresponding to the current category compared to the other units and $\left(y_r^t * 0.5 + \sum_{y \in y_w^t}(1-y) * \frac{0.5}{nOL-1}\right)$ rewards the ability to maximize the activation of the correct unit while minimizing the activations of the wrong units, with the activation of the maximization of the correct unit weighing as much as the sum of the minimization of incorrect units. nT is the number of trials, nC is the number of steps in a trial, sFC is the time step in which we start to compute fitness, and rank is the ranking of the activation of the categorization corresponding to the correct letter or object, which is from 0, meaning the most activated and 4, meaning the least activated. y_i^t is the activation of the output corresponding to the right letter or object in trial t , y_w^t is the set of activations of the wrong letters or objects for trial t , and nOL is the number of letters or objects. The initial population consists of 100 randomly generated genotypes, each encoding the free parameters of the corresponding neural controller, which include all the connection weights, gain factors, biases and the decay constants of leaky hidden neurons. The parameters are encoded with 8 bits each. In order to generate the phenotypes, weights and biases linearly mapped in the range [-5;5], while time constants are mapped in [0;1].

3.1 Experiment One

Experiment was done in order to show the effectiveness of the grayscale in solving a simple image classification pproblem (i.e letter categorization). The experiment

consists of a moving eye located in front of a screen of 100 by 100 pixels and is used to display the letters to be categorized (one at a time). The artificial eye is a periphery only vision, which consists of a 5 by 5 photoreceptors uniformly distributed over a square and cover the entire retina of the eye. Each photoreceptor detect an average gray level of an area corresponding to 10 by 10 pixels of the image displayed in front of the screen. The activation of each photoreceptor ranges from 0 to 1, with 0 representing a fully white visual field while 1 representing a fully black. The screen is used to display five italic letters (l, u, n, o, j) of five different sizes each, with a variation of ± 10 and ± 20 percents to the intermediate size. (see Fig. 1 for the letter l). The letters are displayed in black and gray over a white background as shown in Fig. 1 for letter l

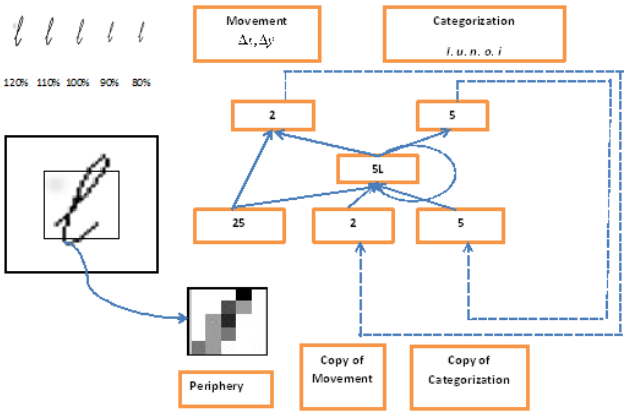

Fig. 1. The architecture of the network used in experiment one. It has 32 input neurons, 25 of which are for the periphery visual stimuli and 7 for efferent copies of the movement and categorization units. It also has 5 hidden neurons and 7 output neurons (i.e 2 for movement and 5 for categorization units). The left side of the figure, shows the different variations of the letter l and the periphery vision scanning part of the letter, with white image background

The agent is evaluated for 50 trials, lasting 100 time steps each. At the beginning of each trial: (i) one of the five letters in one of the 5 different sizes is displayed at the center of the image screen, with each letter of each size presented twice to an individual, (ii) the state of the internal neurons are initialized to 0.0, (iii) the eye is randomly initialized at the centre one third of the screen, (so that the agent can always perceive part of the letter with the periphery vision).

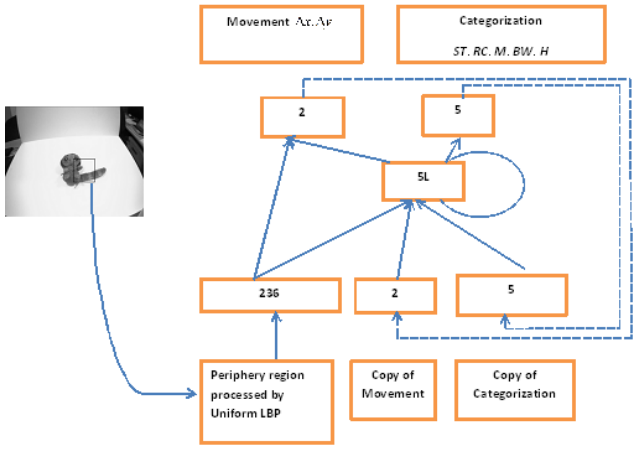

Fig. 2. The network of the experiment two with a total of 243 input neurons, of which 236 are for the visual stimuli and 7 are for the efferent copies from the output units activations. There are 5 hidden neurons of recurrence activations and 7 output neurons, of which 2 are the motor neurons and the other 5 for labelling categories of the objects. It also shows the periphery region as it scans the image

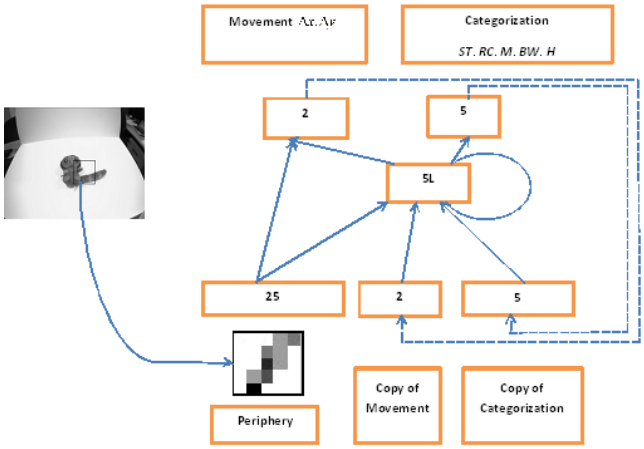

Fig. 3. The network architecture of the experiment three, of which the input features of the periphery region are processed by the grayscale method, the number of input neurons are 32. It also has 5 recurrence hidden neurons and 7 output neurons, of which 2 determines the displacement of the eye per time step and remaining 5 are for the categorization units. It also shows the periphery region as it scans the image.

3.2 Experiments Two and Three

In experiment two, we have adopted Uniform Local Binary Pattern method [8] for the pre-processing of the periphery region for the task of categorizing objects on images taken from humanoid (iCub) robot camera and in experiment three, we adopted the grayscale method for the same problem. The two systems are used to categorize coloured images of five different objects namely: soft toy, tv remote control set, microphone, board wiper and hammer. These are represented by letters ST, RC, M, BW, and H in the categorization units of the two networks (Fig. 2 and Fig. 3). The image sizes are 320 by 240 pixels. The original coloured images are first converted into gray images. The agents are evaluated for 10 trials lasting 100 time steps each. At the beginning of each trial: (i) each object on each image is presented twice to each individual, (ii) the state of the internal neurons are initialized to 0.0, (iii) the eye is initialized in a random position within the central one third of the object. Also in-order to make the images suited for the systems, in which trials are terminated when the eye (periphery region) loses visual contact with the object for three consecutive time steps, we used Canny Edge Detector to detect the edges on each image loaded per trial, and set rectangular masks on the objects in the images and set every other white pixels outside the boundary of these to black. Through this we are able to get images that consist of total outside boundary of black and the object of white and black. Fig. 4 show the gray images , Fig. 5 shows the images after being processed by the Canny Edge Detector and Fig. 6 shows the final look of the images after setting a rectangular mask on the Canny Edge Detector processed images. It should be noted that the above processing of the gray images by Canny Edge Detector and rectangular masking, which finally led to the images shown in Fig. 6 are only used to control the movement of the eye, so that every trial is terminated after the periphery vision loses total focus of the object for more than 3 time steps. It is the gray images that are processed by the Uniform Local Binary Pattern [8] (experiment two) and grayscale averaging (experiment three) and are used as input vector of the neural network along with efferent copies of the movement and categorization units (i.e activations at previous time step $t-1$).

Fig. 4. The above figure shows the gray images that are used in the categorization experiment

Experiment Two The experimental set up consist of a moving eye (artificial agent) , covering a total area of 50 by 50 pixels (periphery region) of the

Fig. 5. The above figure shows the images after being processed by the Canny Edge Detector

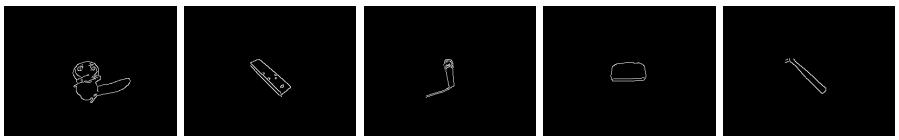

Fig. 6. The above figure shows the images after setting a rectangular mask on the Canny Edge Detector processed gray images

presented image per trial. The periphery image region is pre-processed with Uniform Local Binary Pattern [8], in order to enhance its quality and also reduce the feature vector size. In the experiment, we have divided the periphery region into 4 blocks of which histogram of uniform patterns are constructed for each block. Histograms of all the blocks are concatenated to form a feature vector, with each block giving a histogram of size 59. The feature vector is normalized between 0 and 1, with 0 representing a fully white visual scene and 1 representing a fully black, which forms the input vector of the neural network along with the efferent copies of the movement and categorization output units (i.e outputs at the previous time step), (see Fig. 2)

Experiment Three We performed a third experiment in-order to do a comparative analysis of the results with the results from our proposed method in experiment two. In this experiment, we adopted the grayscale averaging method in [1] for the processing of the periphery region of the images taken from the humanoid (iCub) robot camera (Fig. 3). The input into the neural network are: (i) activations of 5 by 5 photoreceptors, in which each one detects an average gray level of 10 by 10 pixels of the image displayed, and (ii) the efferent copies of the outputs of 2 motor units and 5 categorization units (i.e at previous time step). The activations ranges between 0 and 1, with 0 representing a fully white and 1 representing a fully black visual scene. The results of the experiment are described in section 4

4 Results

We show here the results of our experiments separately, as we have done three major experiments.

4.1 Experiment One

We have done 5 replication of the evolutionary run (Fig. 7 shows the graph of the best fitness), and also assessed the categorization capability of the system for the five letters (l, u, n, o, j) in the evaluation test (Fig. 8). The replicated grayscale experiment did very well in the task of labelling all the letters as demonstrated by the significance difference between the average activations of the current categories and the other categories in each labelling task. Also statistical data distributions of activations of the current categories are also very dense. The average performance accuracy of the system based on average activation and data distribution in all the categorization tasks is about 99 percents. (Fig. 8 shows the results of the performance evaluation test),

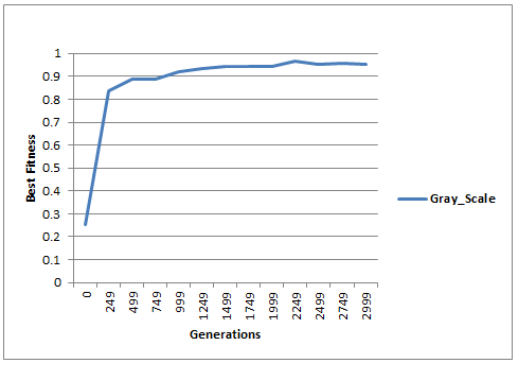

Fig. 7. The average of the best fitness in 5 replications of the evolutionary run for the experiment *one*

4.2 Experiment Two

We have also shown here the results of the performance evaluation test of our proposed Uniform Local Binary Patten method [8] for 5 replications of the evolutionary run (Fig. 9 shows the graph of the best fitness). The evaluation test was done for 5 set of images i.e soft toy, tv remote control, microphone, board wiper and hammer, taken from a humanoid robot camera . The set of images that were used in the evaluation test are static (i.e they are of the same orientations and scales). The assessment of the performance of the system was done using average of activations of each labelled category for about 400 trials. (Fig. 10). The results from our evaluation test show that the system was able to categorize the soft toy, microphone and the hammer and also did fairly well for ty remote control and hammer. The statistical data value difference between the average activations of the current categories(for the soft toy, microphone and hammer) and the other labelled categories in the case of correct categorization are quite close and the

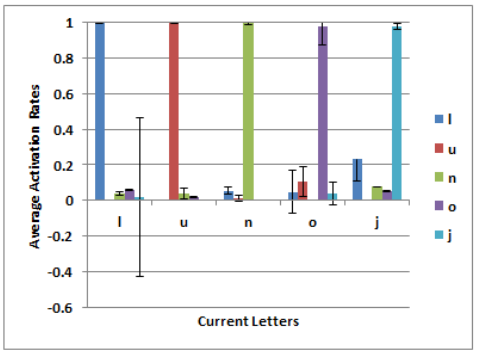

Fig. 8. The histogram of the results of performance evaluation test of experiment one for letter categorization using the average activations for labelled categories of (l, u, n, o, j) for 200 trials

data distribution is quite sparse. Likewise, for those ones of which the system was not fully recognising the objects, the difference between the average activations of the current categories (remote control and board wiper) in comparison with higher average activation in each categorization task are very insignificant and the data distribution of the categories with higher average activations than the current categories are also quite sparse, with similar pattern of distributions (Fig. 10). Overall, based on average activations of each categorization, statistical data difference and distribution the system has an average performance accuracy of about 55 percents in all the categorization tasks (Fig. 12 and Fig. 10).

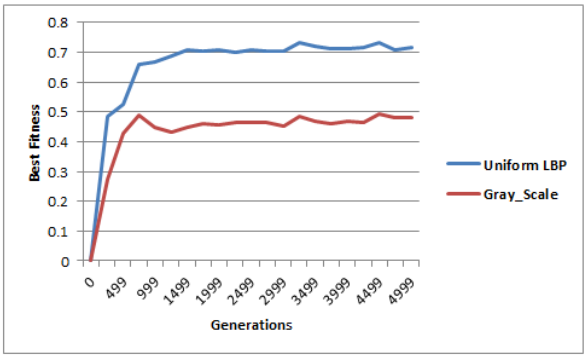

Fig. 9. shows the average of the best fitness of 5 replications of the evolutionary run for the Uniform Local Binary Pattern [8] experiment two and grayscale experiment three for categorizing the objects on the images taken from Humanoid (iCub) robot camera

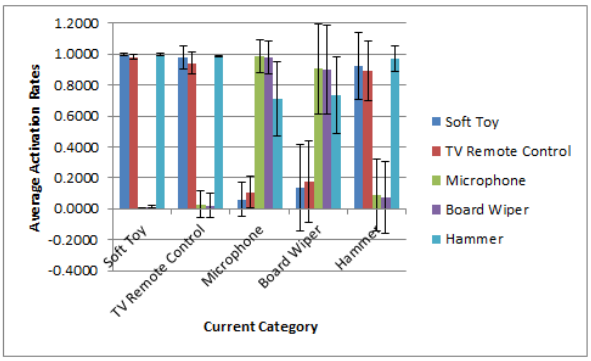

Fig. 10. The histogram of the results of the performance evaluation test for the Uniformed Local Binary Pattern [8] method using the average activations of the labelled categories of the objects on the images from Humanoid (iCub) robot camera for 400 trials

4.3 Experiment Three

We have done 5 replications of the evolutionary run, (Fig. 9 shows the best fitness graph), and the results of performance evaluation test in which performance were evaluated based on the average activation values of each labelled category of the objects in about 400 trials are shown in Fig. 11. The results show that the grayscale method was able to categorize only the tv remote control. The data distribution generally are also very sparse. However, it is also noticed that in all the categorization tasks the activation values of the tv remote control are the highest with almost the same value. Overall, the system has an average performance accuracy rate of about 20 percents (Fig. 12 and Fig. 11)

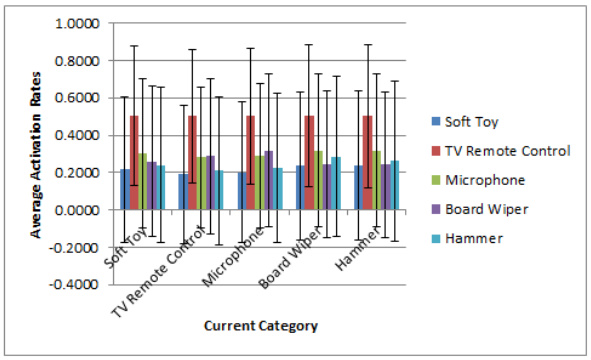

Fig. 11. The histogram of the results of the performance evaluation test for grayscale method in experiment three, using the average activations of the categories of the objects on the images from Humanoid (iCub) robot camera for 400 trials

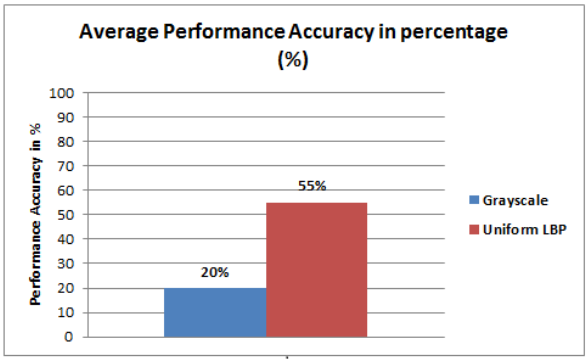

Fig. 12. The chart shows the average performance accuracy for the Grayscale and Uniform Local Binary Pattern method [8] in categorizing objects taken from Humanoid iCub robots camera in 400 trials

5 Discussion

This is a preliminary investigation and as such we have only used few dataset for the training and evaluation tests, and also not enough replications of the evolutionary run, for one to generalise the performance of the systems in the experiments. However, the approach of our proposed method of Uniform Local Binary Pattern [8] is a theoretical proof of concept of the kind of technique that can be use in solving more complex image problems. The first experiment using grayscale method was used to assess the capability of this method for ordinary letter categorization. The method did significantly well in all the letter categorization tasks in the performance evaluation test, with about 99 percents accuracy (Fig. 8). In the second experiment, the proposed pre-processing Uniform Local Binary Pattern Method [8], the system was able to categorize the soft toy, microphone, hammer, although not too statistically significant with relative average activations differences between the current categories and the second highest activation rates being very close. In the other two categorization tasks, i.e tv remote control and board wiper, the proposed system did fairly well, with the difference in average activation values of the current categores (tv remote control and board wiper) and the higher average activations rates very insignificant. Overall the system has an average accuracy rates of about 55 percents.(Fig. 10 and Fig. 12). The grayscale method in the third experiment was only able to categorize the tv remote control. The system has an aveage accuracy rate of about 20 percents for all the objects (Fig. 11 and Fig. 12). Also observation made from the histogram of the performance evaluation test show that the activation of the tv remote control was always consistently higher than the other categories in all the categorization tasks, and having a very similar values of about 0.5 (Fig. 11). However, the proposed Uniform Local Binary

Pattern [8] method look very promising for the following observed reasons: (i) The differences between the average activation rates of the current categories and those categories with higher average activation rates are quite insignificant, when the current categories are not correctly categorized, (ii) the statistical data distribution of the current category when correctly labelling a category, in the average are generally dense, (iii) the eye (periphery region) seems to have more focus on the objects in the grayscale method than in the proposed Uniform Local Binary Pattern [8] method. We therefore, have an intuition that if the eye could be controlled better, so that it could focus more on the objects, we may be able to improve further and achieved better results than grayscale method and so get the system also working very well in a variety of categorization tasks. Future work will be in this direction.

6 Conclusion

We have done a preliminary investigation using Uniform Local Binary Pattern [8] for pre-processing more complex images taken from Humanoid iCub robot camera for a neuro-evolution controlled active vision system. Our proposed method had about 55 percent accuracy as compared to grayscale method of about 20 percents in the categorization task. Future research we be done in-order to have a better control of the eye over visual scene probably using a form of zooming.

References

1. Mirolli. M., Ferrauto, T., Nolfi, S.: Categorisation through Evidence Accumulation in an Active Vision System. Connection Science, 22, 331-354 (2010)
2. Aditya, K., Nakul, I.: Evolving An Active Vision System for Constrained Environments. AI Project Report (2010)
3. Floreano, D., Kato, T., Marocco, D., Sauser, E.: Co-evolution of active vision and feature selection. Biological Cybernetics, 90, 218-228 (2004)
4. Stanley, K., Miikkulainen, R.: Evolving a roving eye for go. Genetic and Evolutionary Computation, pp 1226-1238 (2004)
5. James, D., Tucker, P.: Evolving a neural network active vision system for shape discrimination. Genetic and Evolutionary Computation Conference (2005)
6. Morimoto, G., Ikegami, T.: Evolution of plastic sensory-motor coupling and dynamic categorization. Artificial Life 9, 188-193 (2005)
7. Tuci, E.: Evolutionary Swarm Robotics: Genetic Diversity, Task-Allocation and Task-Switching Behaviour. Proceedings of the 9th International Conference on Swarm Intelligence, Brussels Belgium (2014)
8. Ojala, T., Pietikinen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions, 29, 51-59 (2002)