

Improving Active Vision System Categorization Capability through Histogram of Oriented Gradients

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. In the previous work of Mirolli et al. [1], an active vision system controlled by a genetic algorithm evolved neural network was used in simple letter categorization system, using gray-scale average noise filtering of an artificial eye retina. Lanihun et al. [2] further extends on this work by using Uniform Local Binary Patterns (ULBP) [4] as a pre-processing technique, in order to enhance the robustness of the system in categorizing objects in more complex images taken from the camera of a Humanoid (iCub) robot. In this paper we extend on the work in [2], using Histogram of Oriented Gradients (HOG) [5] to improve the performance of this system for the same iCub image problem. We demonstrate this ability by performing comparative experiments among the three methods. Preliminary results show that the proposed HOG method performed better than the ULBP and the gray-scale averaging [1] methods. The approach of better pre-processing with HOG gives a representation that could translate to improve motor responses in enhancing categorization capability for robotic vision control systems.

Keywords: Categorization, Active Vision System, Neural Network, Genetic Algorithm, Histogram of Oriented Gradients

1 Introduction

An active vision system uses information from sensory-motor coordination in order to intelligently guide the vision mechanism to relevant salient features in a visual scene. The motor control of the visual system can be model by various techniques, but neural network seems to be very appropriate because of its biologically inspired background and suitability for noisy data. However, research in this area is still in its infancy [3]. Most of the work that has been done using this approach was for simple image problems. For Instance, James and Tucker [3] developed an Active Vision System for categorizing different 2D shapes. The system has the ability to move about in any direction, zoom and rotate; it was also able to categorize different 2D simple shapes irrespective of orientations, scales and locations. Mirolli et al. [1] developed an active vision system that is based on a genetic algorithm evolved neural network to categorize gray-scale

italic alphabet letters in different scales (sizes). The periphery region of the eye was processed by average gray-scale filtering. 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. However, Lanihun et al. [2] extended on the work in [1] by using ULBP to process the periphery region, so as to enhance its robustness in categorizing objects in more complex images taken from the camera of a Humanoid robot. They demonstrated the ability of the method to enhance the robustness of this kind of system with improved performance in their proposed system when compared to that of the gray-scale averaging [1] for the same Humanoid image problem. Lastly, in relation to other works listed above, the approach in this paper also involves a visuo-motor coordination of an active vision system based genetic algorithm evolved neural network. We have extended on the work done by Lanihun et al. [2] with the enhancement of the system with HOG to categorize objects in images taken from a Humanoid robot camera.

2 Experimental Details

We have performed three sets of experiments for the categorization of objects in the images taken from the Humanoid camera: (i) the gray-scale averaging experiment in [1], (ii) the ULBP pre-processing as used in [2], (iii) our proposed method that uses HOG for the processing of the retina region. However, we will not explain the details of the gray-scale averaging [1] and ULBP methods as these have been explained in our previous paper in [2]. The artificial agent is in form of a moving eye that explores a visual scene, in order to extract relevant information and 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 [2]. We have also adopted the periphery only architecture of [2] and [1](Fig. 2). The neural network, evolutionary process and the fitness function were the same for the three experiments. The images used in the experiments were coloured of size 320 x 240 pixels (Fig. 1), and were converted to gray-scale images for processing. The objects categorized were: soft toy, remote control set, microphone, board wiper and hammer, which are represented by (ST, RC, MC, BW, H) respectively in the output categorization units of the network (Fig. 2). We have divided our image datasets into two groups as follows:

1. A training set with each object of 3 different sizes with variation of ± 15 percent to the intermediate size, and each size with 3 different orientation in the range $[4, -4]$ degrees; and an evaluation set with objects of 2 different sizes, with a variation of 10 and 20 percent to the intermediate size, and each size of 5 different orientations in the range $[+3, -3]$ degrees.
2. A training set of objects of 5 different sizes varied between ± 20 percent to the intermediate size, and each size of 3 different orientation in the range $[5, -5]$ degrees; while the evaluation set has 5 varied sizes in the range ± 18 percent to the intermediate size and each size of 3 varied orientation within the range $[6, -6]$ degrees.



Fig. 1. The above figure shows the original coloured images taken from the Humanoid robot camera

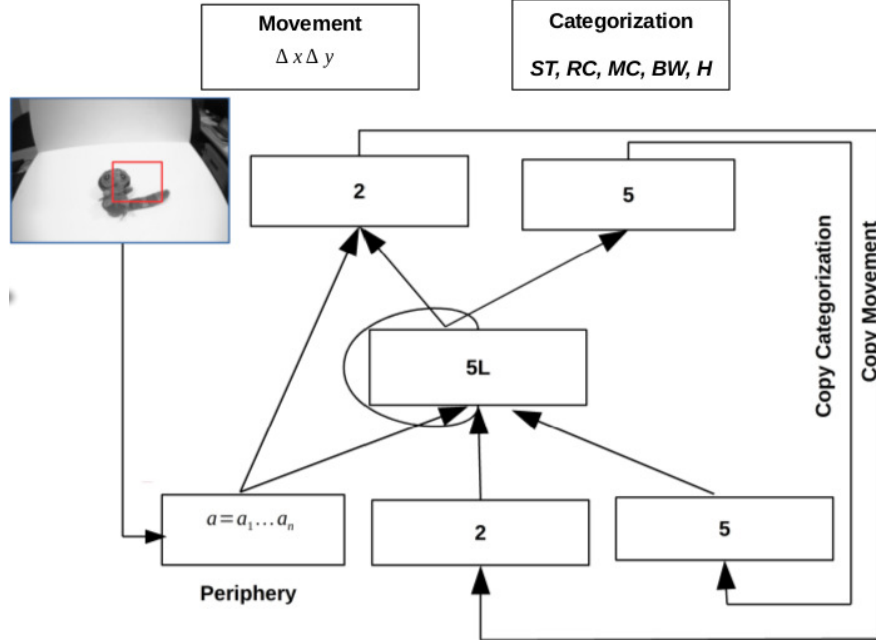


Fig. 2. The neural network architecture and the periphery region scanning the presented soft toy gray image in a trial for categorization

The neural network and the evolutionary task The recurrent neural network is a 3 layer architecture (Fig. 2). The input units activations were normalized between 0 and 1, and the vector size was determined by the method used in the processing of the periphery region. The input neurons encode the current states of the periphery region and the efferent copies of activations of 2 motor neurons and 5 categorization units (i.e. at previous time step $t - 1$). There are 5 hidden units of recurrent activations that depend on the input activations received from the input neurons through the weighted connections and its own activations at the previous time step. In the output layer, the 2 motor neurons determine the movement of the eye per time step (maximal displacement of $[-12, 12]$ pixels in X and Y directions), and the other 5 neurons are the categorization units that label the category. The agents were evaluated for 150 trials lasting 100 time steps each and for 5000 generations. In each trial the artificial

eye is left to freely explore the visual scene (image), however, a trial is ended when the eye can no longer perceive any part of the object in the image through the periphery vision for three consecutive time steps. The task of the agent is to correctly label the category of the current object during the second half of the trial, i.e, when the agent has explored the image for enough time. The agent was evaluated by the fitness function which has two components: the first one rewards the ability of the agent to rank the current category higher than the other categories; and the second rewards the ability of the agent to maximize the activation of the correct unit while minimizing the activations of the incorrect units, with the maximization of the correct unit contributing as much as the sum of the minimization of incorrect units.

2.1 The Proposed Histogram of Oriented Gradients Method

We have used HOG method to process the periphery region of the presented image per trial. However, we made some changes to the original HOG implementation by Dalal and Triggs [5], this was done in order to make a trade-off between performance and accuracy due to the specific nature of our problem and the evolved neural network method. This is because implementing it the way it was done in the original experiment in [5] that had a feature vector size of 3780 will be too computationally expensive for an evolved neural network. Moreover, we noted in our experiment for the iCub images that increasing the size of the feature descriptor did not result into reasonable improve performance. Our proposed system was implemented as follows: (i) a window size of 50 x 50 pixels which represents the periphery region and of 2 x 2 cells giving a total of 4 cells; (ii) gradient orientations were quantized into 9 bins (bin size was 20 degrees for each cell), and orientations were between 0 -180 degrees (i.e signed gradients were ignored), because including signed gradients with orientation of 0 to 360 degrees reduces the performance accuracy of our system for this particular image problem; (iii) gradient magnitudes were truncated into bins and; (iv) histograms were concatenated into feature descriptor of dimension 4 cells x 9 bins giving a feature vector of size 36. This feature vector size is substantially smaller than that of the original implementation by Dalal and Triggs [5] and therefore a more reasonable input vector size for the neural network.

3 Results

We have performed 10 replications of the evolutionary run for each of the 2 training groups for the three set of experiments , and evaluated the performance of the system using the genomes of the best evolved individual from each group. The evaluations were done for 10,000 trials. The system performance was assessed using the average results of the two evaluation groups based on: the percentage of times during the second half of each trial the categorization unit corresponding to the current object was the most activated, and the average activation rates of current objects in all categorization tasks. In terms of the percentage of times

Table 1. Experiment One (Gray-scale): Evaluation Test

	Average Activation Rates of Objects (Highest Activation Rates in Bold)				
Current Objects	Soft toy	Remote control	Microphone	Board wiper	Hammer
Soft toy	0.999940	0.499999	0.001440	0.500421	0.504063
Remote control	0.999940	0.499999	0.001441	0.500422	0.504058
Microphone	0.999940	0.499999	0.001441	0.500422	0.504059
Board wiper	0.999940	0.499999	0.001442	0.500422	0.504055
Hammer	0.999940	0.499999	0.001441	0.500422	0.504060

Table 2. Experiment Two (Uniform Local Binary Pattern): Evaluation Test

	Average Activation Rates of Objects (Highest Activation Rates in Bold)				
Current Objects	Soft toy	Remote control	Microphone	Board wiper	Hammer
Soft toy	0.927989	0.296624	0.430732	0.400178	0.382848
Remote control	0.303231	0.851935	0.222896	0.180930	0.649742
Microphone	0.349986	0.389921	0.771348	0.254532	0.529921
Board wiper	0.504821	0.286945	0.000448	0.770973	0.002129
Hammer	0.053441	0.894695	0.465550	0.114323	0.958321

in which current object was the most activated, our proposed HOG method had 81.98 percent accuracy as compared to that of ULBP of 72.98 percent and gray-scale averaging [1] of 20.1 percent (Fig. 3). However, in terms of average activation rates, the gray-scale averaging [1] method was only able to categorize the soft toy (table 1); while the ULBP (table 2) and our proposed HOG method (table 3) were successful for all of the categories. More also, our proposed HOG method had an average activation accuracy rates of 92.42 as compared to ULBP of 85.61 percent and gray-scale averaging [1] method of 50.41 percent of current objects in all categorization tasks (Fig. 4). In respect to the average activation in which our proposed HOG method and the ULBP were both successful in all categorization tasks, the higher differences between the current categories and the other categories in our proposed method as compared to that of the ULBP shows the statistical significant of our method over the ULBP.

Table 3. Experiment Three (Histogram of Oriented Gradients): Evaluation Test

	Average Activation Rates of Objects (Highest Activation Rates in Bold)				
Current Objects	Soft toy	Remote control	Microphone	Board wiper	Hammer
Soft toy	0.920484	0.399340	0.024266	0.555876	0.095414
Remote control	0.363643	0.819481	0.005429	0.221149	0.612242
Microphone	0.114605	0.066181	0.921041	0.072493	0.036939
Board wiper	0.508343	0.067864	0.190685	0.992130	0.571213
Hammer	0.038975	0.539647	0.002391	0.481335	0.967917

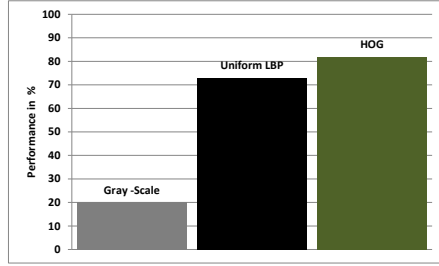


Fig. 3. Accuracy rates in terms of percentage of correct classification.

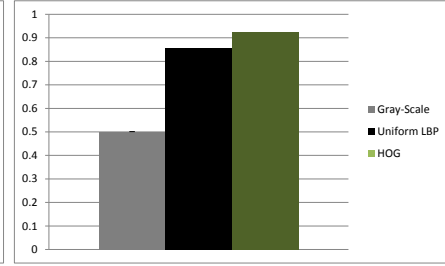


Fig. 4. Accuracy rates in terms of average activations.

4 Conclusion

We have done a preliminary investigation using HOG for pre-processing images taken from Humanoid (iCub) robot camera for an active vision system controlled by an evolved neural network. Our proposed method performed better than the ULBP and the gray-scale averaging [1] methods. The results of our investigation show that this kind of approach using pre-processing techniques could be used for solving more complex image analysis of control problems such as in robot vision. Future investigation will be in deep learning techniques as a form of pre-processing so as to give better motor responses in enhancing categorization capabilities.

References

1. Mirolli. M., Ferrauto, T., Nolfi, S.: Categorisation through Evidence Accumulation in an Active Vision System. *Connection Science*, 22, 331-354 (2010)
2. Lanahun. O., Tiddeman. B., Tuci E., Shaw, P.: Enhancing Active Vision Categorization Capability through Uniform Local Binary Patterns. *British Symposium on Artificial Life and Intelligent Agents, CCIS 519*, 1-13 (2015)
3. James, D., Tucker, P.: Evolving a Neural Network Active Vision System for Shape Discrimination. *Genetic and Evolutionary Computation Conference* (2005)
4. Ojala, T., Pietikinen, M., Maenpaa, T.: Multi-Resolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 29, 51-59 (2002)
5. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. *Computer Vision and Pattern Recognition*, 1, 886-893 (2005)