International Journal of Computer Applications Technology and Research
Volume 6–Issue 12, 501-510, 2017, ISSN:-2319–8656

A Cortical Learning Movement Classification Algorithm for Video Surveillance

Abdullah Alshaikh
Staffordshire University
School of Computing and
Digital Technologies
Stoke-on-Trent, UK

Mohamed Sedky
Staffordshire University
School of Computing and
Digital Technologies
Stoke-on-Trent, UK

Abstract: Classifying the movements of objects detected from a video feed is a key module to achieve a cognitive surveillance system. Machine learning techniques have been heavily proposed to solve the problem of movement classification. However, they still suffer from various limitations such as their limited ability to learn from streamed data. Recently, Hierarchical Temporal Memory (HTM) theory has introduced a new computational learning model, Cortical Learning Algorithm (CLA), inspired from the neocortex, which offers a better understanding of how our brains process temporal information. This paper proposes a novel biologically-inspired movement classification algorithm based on the HTM theory for video surveillance applications. The proposed algorithm has been tested using twenty-three videos, from VIRAT dataset, and an average accuracy of 85% was achieved.

Keywords: hierarchical temporal memory, cortical learning algorithms, movement classification, video forensic, post incident analysis

1. INTRODUCTION

Movement classification algorithms aim at learning motion patterns of objects of interest in a surveillance scenario to classify a new movement. They also attempt to understand the trajectories of tracked objects and the interactions between them. The application domain where these algorithms are engaged is a cognitive surveillance system 1].

Several movement classification methods have been proposed in the literature, which has several merits and demerits, but the advances in these methods have continued and are recently gaining importance and attention of many researchers due to the need for flexible, adaptable ways of solving movement classification problems [2].

Numerous computation techniques have been introduced to enhance computation beyond the physical limits of computers for solving complex problems. One such approach is called biologically inspired computing, also known as a bio-Inspired approach. 'Learning from experience' is a basic task of the human brain which is not yet fulfilled satisfactorily by computers. Moreover, recently cope with this issue, several researchers have been involved in bio-inspired approaches, where a learning method is proposed based on a model derived from neurophysiological observations of the generation of the sense of self which is connected to the memorisation of the interaction with external entities. Therefore, bio-inspired algorithms are based on the structure and functioning of complex natural systems and tend to solve problems in an adaptable and distributed fashion.

New bio-inspired machine learning techniques have been proposed in the attempt of mimicking the function of a human brain. Hierarchical Temporal Memory (HTM) theory has proposed new computational learning models, Cortical Learning Algorithms (CLA), inspired from the neocortex, which offer a better understanding of how our brains function. HTM gives an adaptable and naturally precise system for settling expectation, grouping, and oddity location issues for a wide scope of information sorts [3, **Error! Reference source not found.**].

This paper proposes a novel bio-inspired movement classification algorithm based on the CLA. The proposed algorithm can be used to automate video analytics and video forensic systems.

The remaining of the paper is structured as follows: Section 2 presents a review of related works in movement classification, it includes literature that studies the CLA. The proposed movement classification technique is presented in section 3. The used dataset and the evaluation criteria are presented in section 4. The results are analysed in section 5, and discussed. The paper concludes with section 7 by drawing insights from the proposed techniques, experimentation and results.

2. PREVIOUS WORK

The use of video analytic technologies has gained wide attention in the research community and the global security around the world [5]. The purpose of intelligent visual surveillance in most cases is to learn, detect and recognise interesting events that seem to constitute challenges to the

community or area of the target [6]. These challenges posed by defining and classifying events as unusual behaviour [7], abnormal behaviour [8], anomaly [9] or irregular behaviour [10]. Activity recognition techniques are reviewed, in [1].

Biologically inspired algorithms or bio-inspired algorithms for classification are a class of algorithms that imitate specific phenomena from nature. Bio-inspired algorithms are usually bottom-up, decentralized approaches which specify a basic set of conditions and rules that attempt to solve a complex problem by iteratively applying them. Such algorithms aim to be adaptive, reactive and distributed fashion [14]

2.1 Cortical Learning Algorithms

Cortical Learning Algorithms (CLAs) comprises an effort by Numenta Incorporation [25] to design a model that can perceptually and computationally analyse neocortex learning in the brain. The cortical learning algorithms are utilized as a part of the second implementation of a designed framework for perceptual learning called Hierarchical Temporal Memory (HTM). The algorithm, CLA, functions on a set of data structure, and the two of them together accomplish some level of spatial and temporal pattern recognition. The data structure utilised is a gathering of segments of cells, called a locale. A cell in a section is a neuron-like substance, which makes associations with different cells, and totals their action to decide its state of initiation.

It is biologically proven that neocortex is the seat of intelligent thought in the human or mammalian brain. Intelligent properties such as vision, movement, hearing, touching, etc. are all performed by this intelligent seat, this cognitive tasks that are primarily performed by the neocortex of humans are challenging to design in real life scenarios.

2.1.1 Types of CLA Components

The CLA consists of four main components: Encoder, Spatial Pooler, Temporal Memory and a classifier

Encoder: The initial step of utilising an HTM framework is to change from an information source into a Sparse Distributive Representations (SDRs) using an encoder. The encoder changes over the local configuration of the information into an SDR that can be bolstered into an HTM framework. The encoder is in charge of figuring out which bits ought to be ones, and which ought to be zeros, for a given information esteem in such a route as to catch the essential semantic qualities of the information. Comparative information qualities ought to deliver overlapping SDRs [**Error! Reference source not found.**17]. HTM frameworks require information contribution to the type of SDRs [**Error! Reference source not found.**]. An SDR comprises of a vast exhibit of bits of which most are zeros. The encoder aims to generate a code where every piece conveys some semantic meaning so if two SDRs have more than a couple overlapping one-bits, then those two SDRs have comparable implications.

Spatial pooling: The open field of every section is a settled number of information sources that are arbitrarily chosen from a much bigger number of hub data sources. Considering the info design, a few segments will get more dynamic information values. Spatial pooling chooses a consistent number of the most dynamic sections and inactivates (represses) different segments in the region of the dynamic ones. Comparable information designs tend to actuate a steady arrangement of sections.

Temporal Memory: Temporal memory has been a dynamic region of research for HTMs. The significance of temporal memory and the general objectives of temporal pooling have been to a great extent predictable [19]. Be that as it may, the expression "temporal memory" has been utilised for various diverse executions and looking through the code, and past documentation can be to some degree confounding. The first CLA Whitepaper utilised the term temporal pooler to depict a specific usage. This usage was unpredictably tied in with succession memory. Like this the succession memory and transient pooling were both alluded to as "temporal pooling", and the two capacities were perplexed [19].

Classifier: HTM-CLA plans to learn and speak to structures and groupings in light of memory predictions. In any case, the classifier used to interpret the arrangement yield from HTM-CLA are a long way from palatable. Classifiers utilised as a part of the NuPIC structure are KNN, CLA and SDR Classifiers [20]. Two new classifiers are also proposed by [20] given various similitude assessment strategies. The principal technique is H-DS Classifier given Dot Similarity and the second strategy is H-MSC Classifier given Mean-Shift Clustering, in an attempt to make the classifiers in HTM-CLA more productive and powerful.

2.1.3 The Choice of CLA

CLA being an online learning algorithm and needs no pre-processing and requires less training time. For example, CLA has been applied to solve the problem of classifying Electrocardiogram (ECG) samples into sick and healthy groups discriminating subsequence eliminated in the signal after supervision which could otherwise be done by the human supervisor [29].

This paper proposes a bio-inspired Movement Classification technique that tends to achieve an efficient and effective performance.

3. PROPOSED MOVEMENT CLASSIFICATION TECHNIQUE

The proposed bio-inspired movement classification is based on the CLA that learns to predict a sequence of movements. A slightly erroneous copy of the learned sequences will be presented to the algorithm, which will recover quickly after any unexpected or suspicious movement patterns.

However, going to predict the rest of the sequence, this would be a desirable property since real-world data is likely to be noisy and dynamic. The proposed Cortical learning movement classification algorithm presents a unique and novel way of approaching this problem.

3.2 Movement Classification Datasets

Most post-incident analysis cases target outdoor scenarios. Not all publicly available movement classification and action recognition datasets represent realistic real-world surveillance scenes and scenarios as they contain short clips that are not representative of expected actions in these scenarios. Some of them provide limited annotations which comprise event examples and trajectories for moving objects, and hence lack a solid basis for evaluations in large-scale.

VIRAT video dataset is a large-scale dataset that facilitates the assessing of movement classification algorithms. The dataset used for this study was designed to be natural, realistic, and challenging for video surveillance domains stipulated to its background clutter, resolution, human event/activity categories and diversity in scenes than existing action recognition datasets [28].

According to [28] the dataset distinguishing characteristics are as the following:

- **Realism and natural scenes:** VIRAT's data is collected in natural scenes by showing people in standard contexts performing normal actions, with cluttered backgrounds in an uncontrolled environment.
- **Diversity:** VIRAT's data is collected from multiple sites through a variety of camera resolutions and viewpoints, while many different people perform actions.
- **Quantity:** Various types of human-vehicle and human actions interaction are included with a large number of examples (>30) per action class.
- **A wide range of frame rates and resolution:** Many applications operate across a wide range of temporal and spatial resolutions such as video surveillance. Therefore, the dataset is designed purposely to capture the ranges, (with 2–30Hz) frame rates and 10–200 pixels in person-height.

4. Dataset and Evaluation Criteria

VIRAT dataset includes a total of eleven scenes that were recorded in the videos captured by

stationing high definition cameras. Due to the wind, the videos reordered might experience a little clutched and encoded in H.264 as highlighted from VIRAT Dataset. Each scene contains many video clips, and each clip has zero or many instances. The file name format is unique which makes it easier for the identification of videos that are from the same scene using the last four digits that indicate collection group ID and scene ID.

4.1 Annotation Standard

There is a total of twelve different types of events which are either fully annotated or partially annotated. The event is represented as the set of activities objects are involved within a time interval, e.g. "PERSON loading an OBJECT into a VEHICLE" and "PERSON unloading an OBJECT from a VEHICLE". Objects are annotated as long as they are within the vicinity of the camera and stop recording a few seconds after the object is out of the vicinity of the camera, all this and much more are considered in terms of analysis and evaluation purposes. MATLAB software is used for this evaluation.

VIRAT dataset includes two sets of annotation files that describe (a) the objects and (b) the events depicted in the videos. Samples of the event annotation files and the object annotation files are shown in Table 4-1 and Table 4-2 these annotation files were generated manually and represent the ground truth used for evaluation. The training includes 66 videos representing three scenes.

The events included in VIRAT training dataset are:

- *unknown=0,*
- *loading=1,*
- *unloading=2,*
- *opening_trunk=3,*
- *closing_trunk=4,*
- *getting_into_vehicle=5,*
- *getting_out_of_vehicle = 6.*

Table 4-1 Sample of VIRAT's object annotation file

Object ID	Duration of object	Frame number	bbox X_lt	bbox Y_lt	bbox Width	bbox Height	Object Type
1	385	3495	157	659	76	132	1
1	385	3496	162	658	76	132	1
. .	. .	. .	. .	. .	. .	. .	. .
1	385	3838	747	498	73	97	1
1	385	3839	747	498	73	97	1
3	4732	0	613	469	254	189	2
3	4732	1	612	468	255	190	2
. .	. .	. .	. .	. .	. .	. .	. .

Object Type: type of object (Unknown=0, person=1, car=2, other vehicle=3, other object=4, bike=5)

4.2 Combining the two files

A Matlab script has been developed to generate a file that combines information from VIRAT object annotation files with corresponding information from VIRAT events annotation files for each video file. Table 4-3 shows the combination of the events and objects annotation file obtained from the sample VIRAT dataset.

Table 4-2 Sample of VIRAT training dataset object annotation file

Reset	Event-ID	Frame No.	Event Type	Object Type	Object ID	bbox X_lt	bbox Y_lt	bbox	bbox
1	**0**	**0**	**6**	**2**	**2**	**648**	**497**	**154**	**66**
0	**0**	**1**	**6**	**1**	**1**	**720**	**490**	**26**	**22**
0	**0**	**1**	**6**	**2**	**2**	**648**	**497**	**154**	**66**
..	..	..	..	..	..	..	..	..	..
1	**1**	**0**	**3**	**2**	**1**	**457**	**432**	**93**	**58**
0	**1**	**0**	**3**	**1**	**2**	**533**	**479**	**21**	**48**
0	**1**	**0**	**3**	**2**	**3**	**205**	**371**	**71**	**44**

4.3 Performance Evaluation

Scene-independent and scene-adapted learning recognitions are the two evaluation modes that are used for testing datasets. Scene-independent has a trained event detector on the scene which is not included in the test, while scene-adapted recognition applied to the clips that may be used for training processes, but the test clips are not used during the process.

This evaluation is based on the documents from VIRAT dataset release 2.0 [27] [28]. This document from the VIRAT dataset website has the following contents that are described below

4.4 EXPERIMENTAL SET-UP

Table 4-3 Sample of the combined generated data

E-ID	E-Ty	E-Lg	E-S-Fr	E-E-Fr	bbox X_lt	bbox Y_lt	bbox Width	bbox Height	NOO
1	5	172	3670	3841	670	454	267	228	2
2	5	217	10413	10629	985	406	209	204	2
3	2	66	10068	10133	891	357	202	128	3
4	6	131	9614	9744	983	399	226	211	2
5	6	112	15812	15923	1220	378	241	126	2
6	5	151	17522	17672	1253	380	198	126	2
2	5	217	10413	10629	985	406	209	204	2

The test egins by isolating the training part of VIRAT vdeo dataset into two sections. The initial segment is utilised for training purposes and the second part is utilised for testing, 60% of the data has been utilised for training and the remaining has been utilised for testing. Each try begins by moving one event or two events, from the training dataset to the testing dataset. The record name demonstrates the shrouded event which has been moved to the testing dataset and is not shown to the algorithm in the training phase e.g. Event0, Event1, Event2 .. Event6. The point is to conceal those occasions in the preparation and to present them in the testing to discover how the proposed algorithm can identify a new event as an anomaly.

4.5 Data Preparation

The data preparation starts by moving one event or two events to the end of the file and the purpose of that to hide those events during the training phase and present it in the testing phase to find out how the system has learnt and understood from previous events.

The results of the CLA anomaly detection algorithm is represented by an anomaly score for each field. This score varies between Zero and One. Where values close to Zero represent movements closer to normal ones and values closer to one represent movements that are abnormal.
First, the evaluation starts from the first test field until a first record that represents an event, which has been hidden in training, appears. The accuracy is calculated by comparing the resulted anomaly score with a threshold. If the anomaly score is less than the threshold the movement is considered normal.

The second step starts when the first record of a hidden event appears. In this case if the resulted anomaly score is greater than the threshold, the result is considered correct. This process has been repeated for threshold values between 0.1 and 0.9 with a step of 0.1 to find the maximum accuracy and hence to identify the optimum threshold.

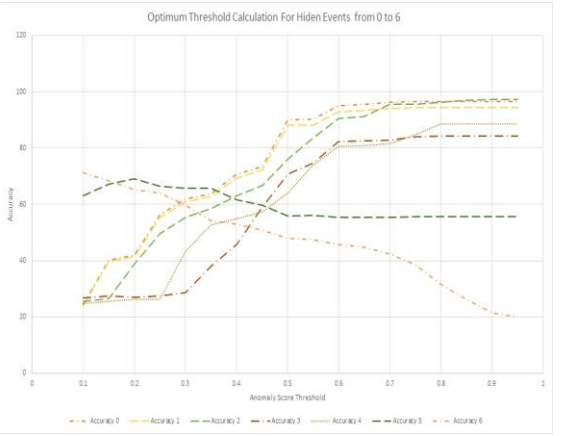
The mathematical equation below defines the calculated accuracy.

$$Accuracy = \frac{Total\ number\ of\ correct\ detection}{The\ total\ numbers\ of\ tested\ events}$$

5. DATASETS ANALYSIS RESULT (CLA ANOMALY ALGORITHM)

Figure 1 highlights the various accuracy trends of the proposed algorithm. The figure shows the change of those anomalies are with different threshold scores.

Figure 1: Optimum Threshold Calculation for Hidden Events from 0 to 6



predicted events where the blue bars refers to the ground truth or the VIRAT dataset, and the orange bars to the predicted ones.
The second machine learning is Decision Tree Learning [25], which is one of the predictive modelling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

6. Conclusion

Artificial Intelligence (AI) and Neural Networks (NN) have been widely used to solve the movement classification problem. However, they still have various limitations for instance, their limited scope of operations. In the attempt of mimicking the function of a human brain, learning models inspired from the neocortex has been proposed which provide better understating of how our brains function. Recently, new bio-inspired learning techniques have been proposed and their results have shown evidence of superior performance over traditional techniques. The CLA processes streams of information, classify them, learning to spot the differences, and using time-based patterns in order to predict. In humans, these capabilities are mainly performed

by the neocortex. Hierarchical Temporal Memory (HTM) is a technology modelled on how the neocortex performs these functions. HTM provides the promise of building machines that approach or exceed the human level performance for many cognitive tasks.

The proposed Bio-Inspired movement classification technique is based on HTM and is biologically used to solve many problems looking at the set of requirement that bio-inspired movement classification technique uses.

The conclusions, from this study, which is drawn are stated below: -

1. The neocortex inspired learning techniques was suitable for correctly learning and predicting a sequence of movement and can then be presented with a slightly erroneous copy of the sequence, which will cover quickly after any unexpected or suspicious movement patterns.

2. As it is also going to predict the rest of the sequence, this would be a desirable property since real world data is likely to be noisy and dynamic in nature.

This study has given indications that neocotex inspired learning techniques are applicable for activities in movement classification aspect of the analysis of video forensic evidence.

7. REFERENCES

1. Vishwakarma, S. and Agrawal, A., 2013. A survey on activity recognition and behavior understanding in video surveillance. The Visual Computer, 29(10), pp.983-1009.

2. Forbes, N., 2000. Biologically inspired computing. Computing in Science & Engineering, 2(6), pp.83-87.

3. Hawkins, J., & Ahmad, S. (2015). Why Neurons Have Thousands of Synapses, A Theory of Sequence Memory in Neocortex. arXiv preprint arXiv:1511.00083.

4. Ahmad, S., & Hawkins, J. (2016). How do neurons operate on sparse distributed representations? A mathematical theory of sparsity, neurons and active dendrites. arXiv preprint arXiv:1601.00720.

5. Popoola, P. and Wang, J. (2012) 'Video-Based Abnormal Human Behaviour Recognition-A Review' IEEE Transaction on Systems, MAN, and Cybernetics-Part C: Applications and Review, vol. 42, no. 6, Nov., 2012.

6. Lavee, G. Khan, L. and Thuraisingham, B (2007) 'A framework for a video analysis tool for suspicious event detection' Multimedia Tools Appl., vol. 35, pp. 109–123, 2007.

7. Hara, K. Omori, T. and Ueno, R (2002) 'Detection of unusual human behavior in intelligent house' in Proc. 2002 12th IEEE Workshop Neural Netw. Signal Process., 2002, pp. 697–706.

8. Lee, C. K. Ho, M. F. Wen, W. S. and Huang, C.L (2006) 'Abnormal event detection in video using N cut clustering' in Proc. Int. Conf. Intell. Inf. Hiding Multimedia Signal Process., 2006, pp. 407–410.

9. Feng P. and Weinong, W (2006) 'Anomaly detection based on the regularity of normal behaviors' in Proc. 1st Int. Symp. Syst. Control Aerosp. Astronautics, Jan.19–21, 2006, pp. 1041–1046.

10. Zhang Y. And Liu, Z (2007) 'Irregular behavior recognition based on trading track' in Proc. Int. Conf. Wavelet Anal. Pattern Recog., 2007, pp. 1322–1326.

11. Kobayashi, M., Okabe, T. and Sato, Y. 2010. Detecting forgery from static-scene video based on inconsistency in noise level functions, Information Forensics and Security, IEEE Transactions on 5 (4) (2010) 883{892.

12. Jing Zhang , Yuting Su , Mingyu Zhang, 2009. Exposing digital video forgery by

ghost shadow artifact, Proceedings of the First ACM workshop on Multimedia in forensics, October 23-23, 2009, Beijing, China

13. Wang, W. and Farid, H. 2009. Exposing digital forgeries in the video by detecting double quantization, in: Proceedings of the 11th ACM workshop on Multimedia and security, ACM, 2009, pp. 39-48.

14. Ding, S., Li, H, Su, C., and Yu, J. 'Evolutionary artificial neural networks: a review' Artif Intel Rev, 2011, vol 39, pp 251-260.

15. Costello, C.J. and Wang, I., 2005, December. Surveillance camera coordination through distributed scheduling. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on* (pp. 1485-1490). IEEE.

16. Kuehne, H., Jhuang, H., Garrote, E., Poggio, T. and Serre, T., 2011, November. HMDB: a large video database for human motion recognition. In*Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 2556-2563). IEEE.

17. Purdy, S., 2016. Encoding data for HTM systems. arXiv preprint arXiv:1602.05925.

18. Bosch, A., Zisserman, A. and Muoz, X. (2008) Scene classification using a hybrid generative/discriminative approach. IEEE Trans. Pattern Analysis and Machine Intell., 30(04):712–727, 2008.

19. Melis, W.J., Chizuwa, S. and Kameyama, M., 2009, May. Evaluation of the hierarchical temporal memory as a soft computing platform and its VLSI architecture. In Multiple-Valued Logic, 2009. ISMVL'09. 39th International Symposium on (pp. 233-238). IEEE.

20. Zhituo, X., Hao, R. and Hao, W., 2012, October. A Content-Based Image Retrieval System Using Multiple Hierarchical Temporal Memory Classifiers. In Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on (Vol. 2, pp. 438-441). IEEE.

21. Balasubramaniam, J., Krishnaa, C. G., & Zhu, F. (2015). Enhancement of Classifiers in HTM-CLA Using Similarity Evaluation Methods. Procedia Computer Science, 60, 1516-1523.

22. Ermoliev, Y., 1983. Stochastic quasigradient methods and their application to system optimization. Stochastics: An International Journal of Probability and Stochastic Processes, 9(1-2), pp.1-36.

23. Bottou, L., 2010. Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010 (pp. 177-186). Physica-Verlag HD.

24. Schapire, R.E., 2003. The boosting approach to machine learning: An overview. In Nonlinear estimation and classification (pp. 149-171). Springer New York.

25. Rodriguez, M., Orrite, C., Medrano, C. and Makris, D., 2016. A time flexible kernel framework for video-based activity recognition. Image and Vision Computing, 48, pp.26-36.

26. George, D. and Hawkins, J., 2009. Towards a mathematical theory of cortical micro-circuits. PLoS computational biology, 5(10), p.e1000532.

27. Moon, J., Kwon, Y., and Kang, K., 2015. ActionNet-VE Dataset: A Dataset for Describing Visual Events by Extending VIRAT Ground 2.0. *2015 8th International Conference on Signal Processing, Image Processing and Pattern Recognition (SIP)*

28. Oh, S., Hoogs, A., Perera, A., Cuntoor, N., Chen, C., Lee, J. T., Mukherjee, S., Aggarwal, J. K., Lee, H., Davis, L., Swears, E., Wang, X., Ji, Q., Reddy, K., Shah, M., Vondrick, C., Pirsiavash, H., Ramanan, D., Yuen, J., Torralba, A., Song, B., Fong, A., Roy-Chowdhury, A., and Desai, M. 2011. A Large-scale Benchmark Dataset for Event Recognition in Surveillance Video. *In Proceedings of IEEE Comptuer Vision and Pattern Recognition (CVPR), 2011.*

29. Akrami, A., Akrami, A., Solhjoo, S. & Nasrabad, A. (2005). EEG-Based Mental Task Classification: Linear and Nonlinear Classification of Movement Imagery. *In Engineering in Medicine and Biology 27th Annual Conference. Shanghai, China, pp. 4626–4629.*