International Journal of Computer Applications Technology and Research
Volume *– No.*, _____ 2012

1.1.1 Application of Cortical Learning Algorithms to Movement Classification

Abdullah Alshaikh
School of Computing and Digital Technologies
Staffordshire University
Stoke-on-Trent, United Kingdom
abdullah.alshaikh@research.staffs.ac.uk

Mohamed Sedky
School of Computing and Digital Technologies
Staffordshire University
Stoke-on-Trent, United Kingdom
m.h.sedky@staffs.ac.uk

Abstract: Classifying the objects' trajectories extracted from Closed-Circuit Television (CCTV) feeds is a key video analytic module to systematize or rather help to automate both the real-time monitoring and the video forensic process. Machine learning algorithms have been heavily proposed to solve the problem of movement classification. However, they still suffer from various limitations such as their limited ability to cope with multi-dimensional data streams or data with temporal behaviour. Recently, the Hierarchical Temporal Memory (HTM) and its implementation, the Cortical Learning Algorithms (CLA) have proven their success to detect temporal anomalies from a noisy data stream. In this paper, a novel CLA-based movement classification algorithm has been proposed and devised to detect abnormal movements in realistic video surveillance scenarios. Tests applied on twenty-three videos have been conducted and the proposed algorithm has been evaluated and compared against several state-of-the-art anomaly detection algorithms. Our algorithm has achieved 66.29% average F-measure, with an improvement of 15.5% compared to the k-Nearest Neighbour Global Anomaly Score (kNN-GAS) algorithm. The Independent Component Analysis-Local Outlier Probability (ICA-LoOP) scored 42.75%, the Singular Value Decomposition Influence Outlier (SVD-IO) achieved 34.82%, whilst the Connectivity Based Factor algorithm (CBOF) scored 8.72%. The proposed models have empirically portrayed positive potential and had exceeded in performance when compared to state-of-the-art algorithms.
Keywords: video analytic; movement classification; machine learning; video forensic; hierarchical temporal memory; cortical learning algorithms

2. INTRODUCTION

In recent years, the development of outdoor surveillance technologies has captured the interest of both researchers and practitioners across the globe. The objective of these technologies is to detect the presence of objects that are moving in the field of view of a CCTV camera(s) for national security, traffic monitoring in big cities, homes, banks and market safety applications or to automate the video forensic process [1]. The use of video analytic technologies has gained much attention in the research community and the global security around the world [2].

The purpose of intelligent visual surveillance in most cases is to detect, recognise, or learn interesting events that seem to constitute some challenge to the community or area of the target [3]. These challenges posed by defining and classifying events as unusual behaviour [4], abnormal behaviour [5], anomaly [6] or irregular behavior [7].

When considering a real environment and trying to relate the way objects interact in surveillance covered area, it is not so easy interpreting every activity correctly. Cluttered environments that contain so many moving objects pose a challenge for many anomaly-detection algorithms. However, in real life cases, these are the kind of scenarios we meet when considering movement classification in video surveillance [2].

There are many hurdles faced by outdoor surveillance system designers and implementers. The first step toward automated activity detection is detection, tracking and classification of moving objects in the field of view of CCTV cameras; another challenge is that sensor resolution is finite, and it is impractical for a single camera to observe the complete area of interest. Therefore, multiple cameras need to be deployed. Also, the detected objects are context-dependent, but for a general surveillance system any independently moving object such as a vehicle, animal or a person are deemed to be interesting, but detecting and classifying these objects is a difficult problem because of the dynamic nature of object appearances and viewing conditions in practical scenarios [8].

In general, to develop a video analytic system that can detect and classify the presence of objects moving in its field of view, that system must be able to: i) detect and classify objects into various categories, ii) track the detected objects over time and iii) classify their movements. Each of the above tasks poses its challenges in term of design and implementation. However, detecting, classifying and analysing the movements of objects were traditionally a manual job performed by humans in which the guaranty of absolute attention over time by a human on duty remains small, especially in practical scenarios [9].

A video analytic system consists of many modules; e.g. change/object detection, object classification, object tracking and movement classification. One key module is the movement classification module. In this module, the movements of detected objects are recorded and compared to infer anomaly. State-of-the-art movement classification rely mainly on rule-based classification techniques, i.e. applying a set of pre-determined spatio-temporal rules, often based on statistical learning techniques, which have been found to correlate to what humans, would interpret as situations of interest, corresponding to threats. Where the abnormalities in the video are traced and reported to the user [9]. Such techniques attempt to learn normal movements to identify abnormal movements.

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 data streams or data with temporal behaviour. In the attempt of mimicking the function of a human brain, learning models inspired by the neocortex has been proposed which offer better understating of how our brains function. Recently, new bio-inspired learning techniques have been proposed and have shown evidence of superior performance over traditional techniques. In this regard, Cortical Learning

Algorithms (CLA) inspired from the neocortex are more favored. The CLA processes streams of information, classify them, learning to spot the differences, and using time-based patterns to make predictions. In humans, these capabilities are largely performed by the neocortex. Hierarchical Temporal Memory (HTM theory attempts to computationally model how the neocortex performs these functions. HTM offers the promise of building machines that approach or exceed the human level performance for many cognitive tasks [11].

Considering the need for improved video analytic systems for the detection and classification of events in video feeds, various benchmark datasets are available in public domain [12, 13]. For example, the i-Lids, Imagery Library for Intelligent Detection Systems, datasets developed by the UK Home Office [14] and VIRAT dataset, a large-scale benchmark dataset for event recognition in surveillance video, developed by DARPA, Defence Advanced Research projects agency [15]. These datasets are captured from realistic surveillance scenarios.

This article introduces a novel CLA-based movement classification algorithm to classify the movements of moving objects in realistic video surveillance scenarios. The performance evaluation of how well the proposed algorithm can differentiate between an unusual movement and a normal movement was carried out based on the ground truth provided by the used dataset [15]. A comprehensive objective evaluation was adopted, which is targeted at comparing the output of the proposed algorithm to state-of-the-art movement classification algorithms.

3. RELATED WORK

A movement classification module attempts to understand the trajectories of tracked objects and the interactions between them. In this stage, the technique may attempt to classify the consistent and predictable object motion. Movements could be classified into two categories, stand-alone or interactive, where stand-alone movements refer to the action of an individual object, while interactive movements refer to the interaction between two or more objects. Statistical learning techniques are often utilised to classify between normal and abnormal activities, based on a priori information, and a user query. The overall aim is to produce a high level, compact, natural language description of the scene activities.

When considering movement classification, the question “what is going on in a scene” is considered [2]. In this sense, there must be a clear definition of what is considered normal/usual and abnormal/unusual. Abnormalities are defined as actions that are fundamentally different in appearance or action done at an unusual location, at an unusual time [16]. When considering anomaly detection algorithm, detecting the spot and where anomalies occur with little to no false alarm is of great emphasis.

4. CORTICAL LEARNING ALGORITHMS

In this study, the Cortical Learning Algorithm (CLA) is applied. The choice of classification algorithm depends on functionality and the design of such algorithm [1]. The CLA processes streams of information, classify them, learning to spot differences, and using time-based patterns to make predictions [19]. Fan et al. [21] critically analysed HTM theory and concluded that the CLAs enable the development of machines that approach or surpass performance level of human for numerous cognitive tasks. The neocortex is said to control virtually most of the important activities performed by mammals including touch, movement, vision, hearing, planning and language [11]. HTM models neurons which are arranged in columns, in layers, in regions, and in a hierarchy. HTM works on the basis of a user specifying the size of a hierarchy and what to train the system on, but how the information is stored is controlled by HTM. According to [9], the CLA processes streams of information, and also classifies the information, learning to identify variations, and using time-based patterns to make predictions. However, the place of time is significant in learning, inference and prediction. The temporal sequence is learned from HTM algorithm from the stream of input data; despite the difficulty in predicting the sequence of patterns. This HTM algorithm is very important since it captures the so-called building block of the neural organisation in the neocortex [22].

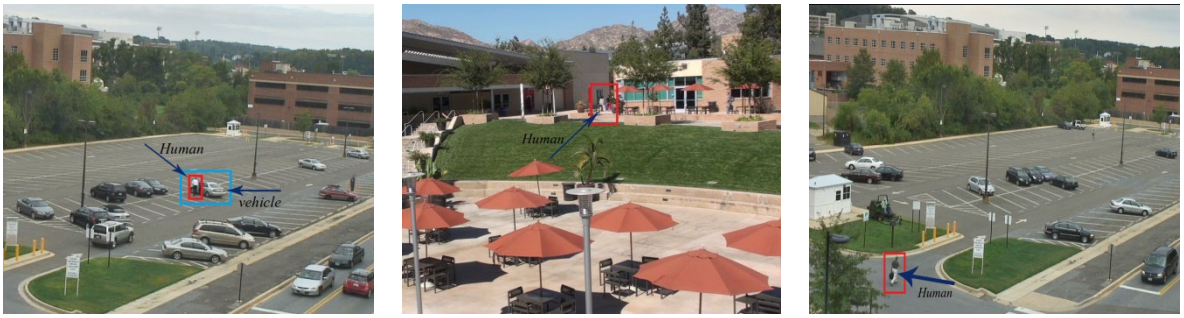
5. TEST AND EVALUATION

There are two modes of evaluation commonly used for testing datasets; they include scene-independent and scene-adapted learning recognitions. According to Wan et al.[27], scene-independent involves trained event detector on the scene which is not considered in the test. In this case, the test clips are used during the test process. Meanwhile, in the scene-adapted learning recognition, the used of clips are involved in training processes. Anjum et al. [17] stated that evaluation techniques consist of multi-object tracking and functional scene recognition that is ground-based annotation giving useful basis for large-scale performance evaluations and real-life performance measures. As a result, various metrics are devised for the evaluation of movement classification algorithms

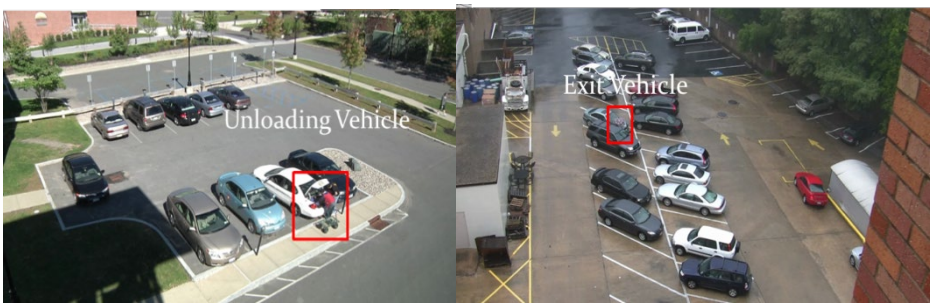
5.1 Datasets

Outdoor scenarios have been targeted in most post-incident analysis cases. Not all publicly available action recognition and movement classification datasets characterise realistic real-life surveillance scenes and/or events as they, mostly consist short clips that are not illustrative of each action performed [23]. Some of them provide limited annotations which comprise event examples and tracks for moving objects, and hence lack a solid basis for evaluations in large-scale. Moreover, according to Khanam and Deb [23], performance on current datasets has been flooded, and therefore requires a new more complex and large dataset to improve development.

large-scale dataset enables the evaluation of movement classification algorithms. The dataset was designed to challenge the video surveillance fields required to its background clutter, resolution, human activity categories and diversity in scenes than existing action recognition datasets. Therefore, VIRAT video datasets are distinguished by the following characteristics; diversity, quality, realistic, natural, ground, aerial, wider range of frame and resolution [24].



a. Human getting into vehicle b. Human entering a facility c. Human carrying a bag



d. Human accessing a trunk e. Human exiting car

f. Human entering a car g. Not known

Figure. 1 Snapshots from VIRAT video dataset

5.2 VIRAT video dataset

There are total of 11 scenes in VIRAT video that were captured by stationing high definition cameras and encoded in H.264. Individual scene consists of many video clips with various activities. 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. As shown in figure 16-1, the datasets snapshots show the VIRAT dataset in three sample activities. In this paper, the VIRAT video dataset is used to perform the evaluation for the proposed movement classification algorithm. There are two categories in which the video dataset is divided into testing and training datasets. The latter contains video scenes with several categories of human and vehicle activities recorded by stationary cameras, in a surveillance setting, in scenes considered realistic. Six object categories are included, unknown, person, car, other vehicle, other object and bike. Seven activities are presented, unknown, loading, unloading, opening-trunk, closing-trunk, getting-into-vehicle and getting-out-of-vehicle.

5.2.1 Annotation Standards

In annotation standards, 12 events are either fully annotated or partially annotated were present. The fully annotated videos have Thirteen (13) event types labelled from 0 to 12 while the partial annotation has Seven (7) event types labelled from 0 to 6. Event, activity, is represented as the set of objects involved with the temporal interval of interest e.g. “PERSON loading an OBJECT to a VEHICLE” and “PERSON unloading an OBJECT from a VEHICLE”. All this is clearly shown in the recorded videos. A person or object are annotated if they are within the vicinity of the camera and the dataset stops recording a few seconds after the object is out of the vicinity of the camera.

The training 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 1 and Table 2. These annotation files were generated manually and represent the ground truth. The training includes 66 videos representing three scenes.

5.3 Evaluation

This evaluation is basically based on the documents from VIRAT DATASET RELEASE 2.0 accessed from VIRAT DATASET . The VIRAT Video Dataset Release 2.0 is used in the analysis and evaluation of the data throughout this paper.

The results of the HTM anomaly detection algorithm is represented by an anomaly score for each field; a field represents a movement. The anomaly scores vary between Zero and One. Where Zero represents a normal movement

(ideally part of an event that has been learned) and One represents an abnormal movement. Values between Zero and One represent the anomaly score, where values close to Zero represent movements closer to normal ones and values closer to One represent movements that are closer to abnormal movements, i.e. suspicious.

First, the evaluation starts with the first scenario, for each record, the Precision, Recall and F-measure are calculated by comparing the resulted anomaly score with a threshold. If the anomaly score is less than the threshold, the detection is considered correct. In the case of an event that has not been shown in the training dataset, if the resulted anomaly score is greater than the threshold the result is considered correct. The True Positive, True Negative, False Positive and False Negative are considered as below:

- TP - the number of "true positives", positive Examples that have been correctly identified
- FP - the number of "false positives", negative Examples that have been incorrectly identified
- FN - the number of "false negatives", positive Examples that have been incorrectly identified
- TN - the number of "true negatives", negative Examples that have been correctly identified
- This process has been repeated for threshold values between 0.1 and 0.9 with a step of 0.05 to find the maximum accuracy and hence identify the optimum threshold.

5.4 Test Results

The table shown below explains the statistics of events presented in the seven experiments including the number of training and testing samples as well as the total number of samples for each hidden event.

Table 1. The hidden numbers of events

Hidden event	Training samples	Testing samples	Total samples
Event 0	59309	62726	
Event 1	60414	61621	
Event 2	61280	60755	
Event 3	57095	64940	122035
Event 4	58571	63464	
Event 5	47951	74084	
Event 6	32144	89891	

In this part of the experiment, an evaluation of the proposed HTM Cortical Learning Algorithm has been tested using the same dataset, Virat, to do a comparison of the performance metrics between each output of different machine learning technique.

Several anomaly detection algorithms are evaluated using RapidMiner Studio version 8.2. Each model's anomaly score is normalised to the range 0.0 to 1.0. The higher the value is, the higher the likelihood of an anomaly occurring.

6. CONCLUSION

Video analytic technologies have gained much attention especially in the context of the security of the community. The ultimate purpose of the intelligent visual surveillance is to handle different behaviours. Currently, the discovery of what is happening in a scene can be seen by automatic scrutiny of activities included in a video. Different algorithms that have been proposed to identify a solution to the movement classification problems. However, the required performance of such algorithms differs depending on the target scenario, and on the characteristics of the monitored scene.

Due to the diversity of video surveillance scenarios and the increasing development of movement classification algorithms, an automatic assessment procedure is desired to compare the results provided by different algorithms. This objective evaluation compares the output of the algorithm with the ground truth, obtained manually, and measures the differences using objective metrics. There are various datasets for activity and human action recognition, though older datasets provide limited ground truth classification to manual annotation at a simpler level, most of the modern datasets, in this case, VIRAT Video Dataset, gives high-quality ground truth.

In this paper, the proposed movement classification algorithm has been tested and its accuracy evaluated. Several experiments have been carried out to calculate the optimum anomaly threshold for each algorithm. the average achieved average F-measure for the proposed algorithm was 66.29%, with an improvement of 15.5% compared to the k-Nearest Neighbour Global Anomaly Score (kNN-GAS) algorithm. Our thanks to the experts who have contributed towards development of the template.

7. REFERENCES

[1] Adams, A.A. and Ferryman, J.M. 2015. The future of video analytics for surveillance and its ethical implications. Security Journal, 28(3), pp.272-289.

[2] Popoola, O.P. and Wang, K., 2012. Video-based abnormal human behavior recognition—A review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(6), pp.865-878.

[3] Lavee, G., Khan, L. and Thuraisingham, B., 2007. A framework for a video analysis tool for suspicious event detection. *Multimedia Tools and Applications*, *35*(1), pp.109-123.

[4] Hara, K., Omori, T. and Ueno, R., 2002, September. Detection of unusual human behavior in intelligent house. In *NNSP* (pp. 697-706).

[5] Lee, C.K., Ho, M.F., Wen, W.S. and Huang, C.L., 2006, December. Abnormal event detection in video using n-cut clustering. In *2006 International Conference on Intelligent Information Hiding and Multimedia* (pp. 407-410). IEEE.

[6] Pan, F. and Wang, W., 2006, January. Anomaly detection based-on the regularity of normal behaviors. In *2006 1st International Symposium on Systems and Control in Aerospace and Astronautics* (pp. 6-pp). IEEE.

[7] Zhang, Y. and Liu, Z.J., 2007, November. Irregular behavior recognition based on treading track. In *2007 International Conference on Wavelet Analysis and Pattern Recognition* (Vol. 3, pp. 1322-1326). IEEE.

[8] Alshaikh, A., & Sedky, M., 2016. Movement Classification Technique for Video Forensic Investigation. International Journal of Computer Applications, 135(12), pp. 1-7.

[9] Akintola, K. 2015. Real-time Object Detection and Tracking for Video Surveillance. VFAST Transactions on Software Engineering, 4(2), pp.9-20.

[10] Verma, B., Zhang, L. and Stockwell, D., 2017. *Roadside Video Data Analysis: Deep Learning* (Vol. 711). Springer.

[11] Hawkins, J., Ahmad, S. and Dubinsky, D. 2011. Hierarchical Temporal Memory Including HTM Cortical Learning Algorithms, 0.2. Techical report Numenta. Inc., September Palto Alto

[12] Li, C.-T, and IGI Global, 2013. Emerging digital forensics applications for crime detection, prevention, and security. Hershey, PA: IGI Global (701 E. Chocolate Avenue, Hershey, Pennsylvania, 17033, USA.

[13] Chris, D. and David, D., 2012. A New Approach of Digital Forensic Model for Digital Forensic Investigation. *International Journal of Advanced Computer Science and Application*s, 2(12). doi:10.14569/ijacsa.2011.021226.

[14] Branch, H.O.S.D., 2006, June. Imagery Library for Intelligent Detection Systems (i-LIDS). In Crime and Security, 2006. *The Institution of Engineering and Technology Conference on* (pp. 445-448). IET.

[15] Oh, S., Hoogs, A., Perera, A., Cuntoor, N., Chen, C.C., Lee, J.T., Mukherjee, S., Aggarwal, J.K., Lee, H., Davis, L. and Swears, E., 2011. A large-scale benchmark dataset for event recognition in surveillance video. *In Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on (pp. 3153-3160).

[16] Varadarajan, J. and Odobez, J.M., 2009. Topic models for scene analysis and abnormality detection. In Computer Vision Workshops (ICCV Workshops), *2009 IEEE 12th International Conference on* (pp. 1338-1345). IEEE.

[17] Anjum, A., Abdullah, T., Tariq, M., Baltaci, Y. and Antonopoulos, N. 2016. Video stream analysis in clouds: An object detection and classification framework for high performance video analytics. IEEE Transactions on Cloud Computing.

[18] Breunig, M., Kriegel, H., Raymond T., and Sander, J. 2000. LOF: identifying density-based local outliers. In SIGMOD Record volume 29, pages 93–104. ACM, 2000.

[19] Byrne, F. 2015. Encoding reality: Prediction-assisted cortical learning algorithm in hierarchical temporal memory. arXiv preprint arXiv:1509.08255.

[20] Comon, P. 1994. Independent component analysis, A new concept? Signal Processing 36 (1994) 287-314

[21] Fan, D. Sharad, M., Sengupta, A. and Roy, K. 2016. Hierarchical temporal memory based on spin-neurons and resistive memory for energy-efficient brain-inspired computing. IEEE transactions on neural networks and learning systems, 27(9), pp.1907-1919.

[22] Feris, R., Bobbitt, R., Pankanti, S. and Sun, M.T. 2015. August. Efficient 24/7 object detection in surveillance videos. In Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on (pp. 1-6)

[23] Khanam, T. and Deb, K. 2017. Baggage Recognition in Occluded Environment using Boosting Technique. KSII Transactions on Internet and Information Systems (TIIS), 11(11), pp.5436-5458.

[24] Moon, J., Kwon, Y., Kang, K. and Park, J. 2015. ActionNet-VE Dataset: A Dataset for Describing Visual Events by Extending VIRAT Ground 2.0. In Signal Processing, Image Processing and Pattern Recognition (SIP), 2015 8th International Conference on (pp. 1-4).

[25] Ramaswamy, s., Rajeev, R. and Kyuseok, S. 2000. Efficient algorithms for mining outliers from large data sets. In SIGMOD Record, volume 29, pages 427–438. ACM, 2000.

[26] Tang, Z., Mingxi, W. and Christopher J. 2002. Outlier detection by sampling with accuracy guarantees. In KDD , pages 767–772. ACM, 2002

[27] Wan, H., Wang, H., Guo, G. and Wei, X. 2018. Separability-Oriented Subclass Discriminant Analysis. IEEE transactions on pattern analysis and machine intelligence, 40(2), pp.409-422.