Detecting Suspicious Behavior on Surveillance Videos: Dealing with Visual Behavior Similarity between Bystanders and Offenders

Guillermo A. Martínez-Mascorro, José C. Ortiz-Bayliss, Hugo Terashima-Marín
Tecnologico de Monterrey
School of Engineering and Sciences
A00824126@itesm.mx, {jcobayliss, terashima}@tec.mx

***Abstract*—Suspicious behavior is likely to threaten security, assets, life, or freedom. This behavior has no particular pattern, which complicates the tasks to detect it and define it. Even for human observers, it is complex to spot suspicious behavior in surveillance videos. Some proposals to tackle abnormal and suspicious behavior-related problems are available in the literature. However, they usually suffer from high false-positive rates due to different classes with high visual similarity. The Pre-Crime Behavior method removes information related to a crime commission to focus on suspicious behavior before the crime happens. The resulting samples from different types of crime have a high-visual similarity with normal-behavior samples. To address this problem, we implemented 3D Convolutional Neural Networks and trained them under different approaches. Also, we tested different values in the number-of-filter parameter to optimize computational resources. Finally, the comparison between the performance using different training approaches shows the best option to improve the suspicious behavior detection on surveillance videos.**

***Index Terms*—Suspicious behavior detection, Visual similarity, 3D Convolutional Neural Networks.**

I. Introduction

Suspicious behavior detection has become an essential and exciting topic due to its potential for several applications [1]–[3]. Recent years have witnessed a growing interest in developing automatic detection methods to improve security and surveillance systems [4]. However, defining "suspicious behavior" seems complicated since this behavior lacks a particular pattern or a fixed set of actions. Consequently, it is challenging to recognize such behavior accurately, even for human observers. People working in surveillance tasks suggest that suspicious behavior detection requires personal understanding and subjective interpretation [4], [5]. Human observers rely on their instincts, but it takes several years of practice and experience to develop this "gut feeling" [6].

Many behavioral specialists agree on describing several suspicious behaviors for better comprehension [7]–[9]. According to the location where the suspicious behavior takes place, it may have different interpretations, such as crime commission [10], terrorism [11], campus security [12], among others. As stated by the Hilliard Police Department [13], people are not suspicious, but their behavior is. It is the behavior that matters, not particular skin color, typical clothes, or facial expressions. Although it is hard to explain, most security and prevention associations try to involve the community to spot suspicious behaviors. The earlier those conducts are found, the less damage they can do.

Suspicious behavior detection systems are developed for particular situations and specific behaviors in mind. Usually, these systems are related to crime commission or prevention. For example, Tang and He looked for suspicious financial transactions to prevent money laundering [1], Pennington *et al*. examined storage-data access patterns to prevent intrusion and data theft [2], and Penmetsa *et al*. proposed aerial visual surveillance to detect violent actions, such as shooting, hitting, or choking [14].

Although many works have focused on crime scene detection [15] and abnormal behavior detection, such as shooting, robbery, or car accidents, only a few works focus on the behavior before the crime commission. More importantly, those few works provide no specific details about the type of offense. In this regard, it is important to mention the Pre-Crime Behavior (PCB) method [16], which has been applied to suspicious behavior detection in shoplifting cases. After using the PCB method and eliminating most of the crime-details information, normal and suspicious samples look very similar. Then, classifiers tend to produce high false-positive rates, given the visual similarity between different classes [17], [18]. A question arises from this situation: if suspicious behavior is generalized within the crime prevention context, should suspicious-behavior samples from different crimes be trained as one class or separately?

In this work, we present a comparison between three approaches that combine training and classification strategies for suspicious behavior detection. The first approach can only classify input samples as normal or suspicious since it relies on binary training / binary classification (the training examples have been labeled as normal or suspicious). The next two approaches rely on multi-class training (the training examples have been labeled as normal or suspicious, but now the labels include the type of crime that followed the suspicious behavior). For classifying, the second approach considers a multi-class classification (it discriminates among four crimes that originate the suspicious behavior as well as normal behavior). Finally, the third approach trains by using a multi-class setting, but the classification is binary. Then, the type of crime is not

important, as long as there is suspicious behavior. To test such approaches, we used 1278 samples videos from the complete UCF-Crime dataset [15], processed with the PCB method. These samples show normal-behavior and four types of crimes: shoplifting, stealing, arson, and abuse. The main contribution derived from this work is the finding that grouping suspicious-behavior samples from different types of crime while training improves the accuracy of the model.

The remainder of this document is organized as follows. The most relevant works related to this investigation are briefly introduced in Section II. In Section III, we present the methods developed as well as the experimental design. Section IV presents and discusses the results obtained. Finally, the conclusions and future work are presented in Section V.

II. Related work

Suspicious behavior is usually confused with abnormal behavior, but they have different meanings. Suspicious behavior regularly refers to unusual interactions between people or people and the objects around them [17]. Abnormal or anomalous behavior usually refers to everything outside the usual or expected behavior [3], [19]–[21]. Works on abnormal behavior detection focus on modeling normal conduct, and everything the model cannot classify as 'normal' is considered abnormal.

Abnormal behavior detection systems are the predecessors and foundation for suspicious behavior detection systems. Some examples of abnormal behavior detection applications include abnormal motion [22], abnormal trajectories [23], and abnormal crowd behavior [24]. However, some behaviors could be unusual or abnormal without being suspicious, such as a small group reunion in the street, or waiting for someone. Taking preventive measures against them could be considered discriminatory or even illegal [25]. Some studies about non-verbal behavior highlight the importance of the context of human behavior understanding tasks [5], [26]. In other words, an observer could consider a behavior suspicious in a particular context, but reasonable in a different situation.

Conversely, most of the available approaches for detecting suspicious behavior try to prevent specific scenarios, mostly related to crimes. For example, Rowe presented a suspicious-behavior detection system based on wireless sensors and changes in positions, velocities, and accelerations [25]. The system considered walking paths in a room a, looking or suspicious patterns and unusual spots to stop. Tang and He combined genetic algorithms and backpropagation neural networks to detect suspicious behavior on financial transactions [1]. They used a genetic algorithm to find the best initial weights for the network and tested their approach on a commercial bank dataset. Goodall *et al*. [27] introduced a tool to fight against cybercrime. This tool supports operators to discover unusual behaviors in streaming network data. The system parses several event streams and scores each of them. Then, it provides visual support for the analyst to explore and understand the context.

Also, many works aim to support CCTV surveillance tasks and prevent crimes like robbery, shoplift, and riots. For example, Mu *et al*. proposed a recognition algorithm for high definition videos based on motion vectors [17]. Their algorithm processes video samples of 1920×1080 pixels and uses a macroblock of size 4×4. They build a video dataset, which includes behaviors such as wandering, trailing, chasing, falling, and normal activity. Xia *et al*. presented a case-based reasoning approach, using a saliency-based visual attention model combined with time-attribute features [28]. They proposed to decompose the general behavior into sub-behaviors and estimate every body position. Using Hidden Markov models, they recognize each action and take start and end times. Martínez *et al*. presented a new approach to process criminal video samples [16]. They proposed the Pre-Crime Behavior method (PCB) to separate a crime-commission sample in three segments and extract a known-offender behavior before even acting suspiciously. Combining PCB and 3D Convolutional Neural Networks (3DCNN), the model detected suspicious behavior with 75% accuracy in shoplifting samples.

Suspicious and normal samples have a high visual similarity after applying the PCB method. Fig 1 shows three possible scenarios:

a) A hyperplane can easily split suspicious and normal samples.
b) A single hyperplane cannot separate the classes, so it is better to consider a multi-class approach.
c) The samples are very similar and cannot be separated.

Convolutional Neural Networks (CNN) have a high performance in computer vision and pattern recognition. Many approaches implemented them to tackle problems such as object detection [29], [30], identifying actions in images [31], and text recognition [32]. Ji *et al*. proposed a three-dimensional convolution on a CNN (3DCNN) architecture to analyze video data [33]. This new approach allowed them o extract spatial and temporal features and opened new analysis areas such as anomaly detection [34], gesture recognition [35] and Magnetic Resonance Imaging (IMR) analysis [36]–[38].

Convolutional Networks have shown a remarkable performance in object and action detection. Regarding related applications, some works have proposed approaches that rely on such networks to analyze a sequence of actions and look for specific behaviors or actions [34], [39], [40]. In these applications, 3DCNNs have proved capable of processing spatial-temporal information, such as video samples, and extract significant features.

Considering the previously described scenarios and types of samples, we propose to explore the visual similarity between behavior samples, after using the PCB method for extracting relevant segments from the samples. Section III presents the details about the dataset, the 3DCNN model, and the approaches to train it.

III. Experiments

In this investigation, we compared three models that combine training and classification approaches for detecting sus-

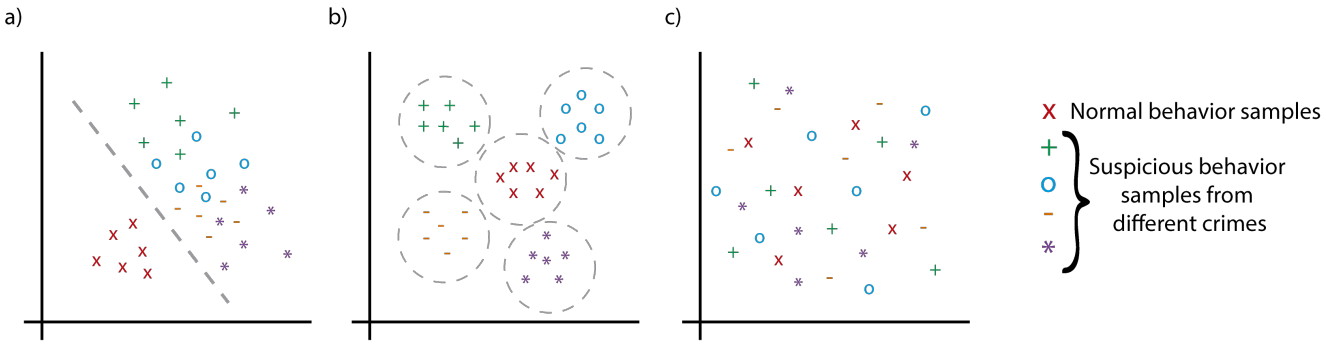

Fig. 1. Three possible sample-distribution scenarios: a), a binary classifier can easily discriminate between normal and suspicious behavior (it ignores the types of offense), b) a multi-class classifier discriminates between normal behavior as well as the different types of offense that give place to the suspicious behavior, and c) It is impossible to discriminate between normal and suspicious behavior since everything is mixed (samples from different classes are similar and different to split).

picious behavior in video samples (Fig. 3). We used crime video samples for the experiments and processed them with the PCB method. Then, we trained 540 different 3DCNNs using three combinations of training and classification approaches. We also modified the number-of-filters parameter for each convolutional layer to look for the necessary decomposition for this task. We considered six configurations for the number-of-filters in the convolutional layers to evaluate the necessary input decomposition.

In the first experiment, we compared the binary training approach against the multi-class training / multi-class classification one for classifying behavior. We assumed that suspicious-behavior samples from different crime videos are similar and easily separable from normal-behavior samples (Fig. 1a). The experiment includes the generation of 360 3DCNNs: 30 3DCNNs trained using the binary training approach (normal vs. suspicious behavior) for each of the six number-of-filters configurations (producing a total of 180 3DCNNs), and another 180 3DCNNs with a similar description but trained with the multi-class training / multi-class classification approach.

The second experiment evaluates the multi-class training approach's performance among two variants: binary classification and multi-class classification. In this case, we assumed that a single hyperplane could not separate the suspicious-behavior samples from the normal-behavior samples (Fig. 1b). We tested multi-class and binary classification from the multi-class trained 3DCNNs. Combinations between training and classification approaches are explained in Sect. III-D.

A. Dataset

The generated dataset includes 1278 video samples based on the UCF-Crime Dataset [15], from which 278 correspond to suspicious behavior and 1000 to normal conduct. Suspicious-behavior samples include shoplifting, stealing, abuse, and arson examples. The dataset contains a fraction of the videos from each class. The selection process included videos where at least one person's behavior is visible before the crime is committed. We processed all the videos with the PCB method and resized the video resolution to 80×60 pixels since this resolution showed a good performance in previous comparisons [16]. For the 3DCNN training, we used 198 suspicious-behavior samples and 700 normal-behavior ones.

In this work, we were interested in dealing with unbalanced datasets to resemble a more realistic scenario for crime prevention. We are interested in models that learn to detect suspicious behavior before a crime is committed, and balanced accuracy (bACC) estimates the model's performance by considering the unbalanced nature of such sets. That is the reason why we used it instead of the standard accuracy.

B. Pre-treatment

The PCB method allows processing video samples where an observer can see an offender's behavior before committing a crime. The method splits the video into three parts (Fig. 2): the crime evidence segment, the suspicious behavior segment, and the pre-crime behavior segment. To process the videos, an observer must watch the complete sample and detect specific moments. The crime evidence segment starts where the suspect in the video unquestionably commits an offense, denominated Strict Crime Moment (SCM). The suspicious behavior segment begins where the observer doubts the person in the video (Comprehensive Crime Moment, CCM). At this point, the suspect is behaving unusually. The third segment is what occurs from the first appearance of the suspect, and just before he/she starts to act suspicious.

PCB segments are like looking for regular clients in a store, to human sight. They only show people walking around the area and looking for products. These segments lack enough information to raise suspicion about a specific person. Therefore, they present a high visual similarity with normal-behavior samples.

We looked for a significant difference between suspicious-behavior and normal-behavior samples using the PCB segments to train 3DCNN models. If the difference is substantial,

the combination may work for suspicious behavior detection, even before a human observer can recognize it.

C. 3DCNN Architecture and Hardware

The 3DCNN general architecture consists of two convolutional layers with max-pooling and dropout stages, and two fully-connected layers. The group of continuous frames was set to ten as the input for the first convolutional layer. Kernels of size 3×3×3 process the information and generate as many output images as the number-of-filters configuration selected (for example, 32–64; 32 for the first layer and 64 for the second one). Each neuron uses a ReLU activation function. We used kernels of size 3×3×3 to perform the max-pooling operation and 25% for the dropout. Next, the output works as the input for the second convolutional layer. The configurations are kernels of size 3×3×3, ReLU activation function, the second value from the filters-pair as the number-of-filters, max-pooling of 3×3×3, and 25% of dropout.

After the second dropout, the information is flattened and used as the input for a fully connected layer of 512 neurons. Finally, the information goes to the output fully-connected layer with as many neurons as the number of trained classes (two for binary classification and five for multi-class). For multi-class training and binary classification, we obtained the confusion matrices and recalculated the hits and fails.

We performed the experiments on a Dell R840 server, which has 128GB in RAM, 57TB for storage, and a GPU NVIDIA Tesla V100 32GB.

D. Methodology

We choose binary and multi-class approaches for training and classification stages (Fig. 3). These approaches considered the sample's distribution from Fig. 1a and b. Binary training works with binary classification: we trained on samples labeled as normal or suspicious, and we classify them accordingly. However, in the case of multi-class training, we tested both binary classification and multi-class classification. Multi-class training / multi-class classification works as supposed; the examples are labeled as normal or as one of four types of crimes. Conversely, multi-class training / binary classification considers as a positive any crime sample classified as positive, regardless of the crime class. For example, if the model classifies a shoplifting sample in the arson class, it counts as a well-classified case.

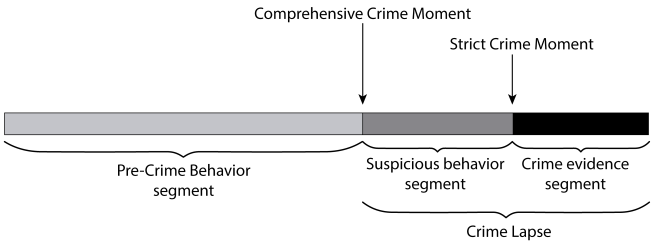

Fig. 2. PCB process representation.

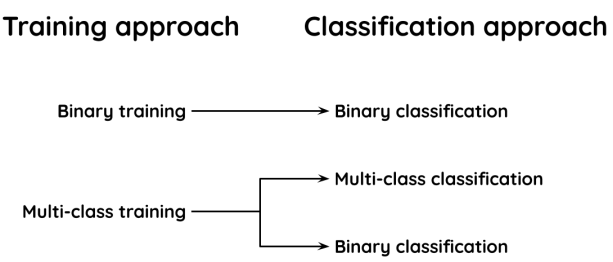

Fig. 3. Training and classification approaches considered for this work.

The experiment considers six number-of-filters configurations for each convolutional layer: 16–16, 32–32, 32–64, 64–64, 64–128, 128–32. The higher number-of-filters, the more details the network extracts, and the more computational resources it consumes. We looked for the best configuration of the number-of-filters to achieve a useful classification and optimizing resources.

We run each configuration 30 times to validate the results. Due to the unbalance in the dataset, we considered the balanced accuracy (bACC). It normalizes true positive (TP) and true negative (TN) predictions by the number of positive and negative samples. The balanced accuracy is defined as follows:

$$bACC = \frac{TPR + TNR}{2}$$

$$TPR = \frac{TP}{TP + FN}$$

$$TNR = \frac{TN}{TN + FP}$$

where TPR is the True Positive Rate, TNR is the True Negative Rate, TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative predictions of each classifier.

IV. Results

After getting the balanced accuracy of all the models, we compared the balanced accuracy of the binary classifiers against the ones of the multi-class ones. Although we aimed at suspicious-behavior detection, we tested the capability of the multi-class classifiers to distinguish suspicious behavior by the corresponding type of crime that aroused the behavior. For this purpose, we conducted a t-test on the means of the balanced accuracy of each method (binary training / binary classification and multi-class training / multi-class classification) for each configuration of the number-of-filters (Fig. 4). We contrasted two hypotheses: H_0, where the means of the two means are equal, and H_1, where means are not. Following the standard, we assumed a significance value of 5% ($\alpha = 0.05$). The results showed overwhelming statistical evidence in favor of binary training / binary classification. Table I shows the p-values of each test, which suggest that the mean of the balanced accuracy is not equal in any of the cases compared.

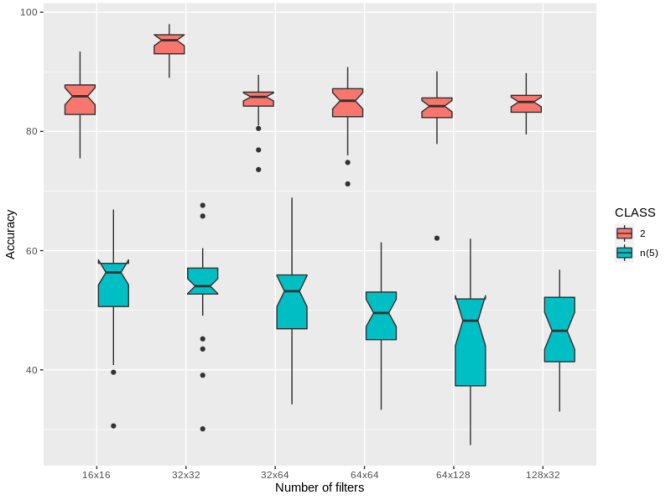

Fig. 4. Balanced accuracy comparison between binary training / binary classification approach and multi-class training / multi-class classification approach, using six number-of-filters configurations.

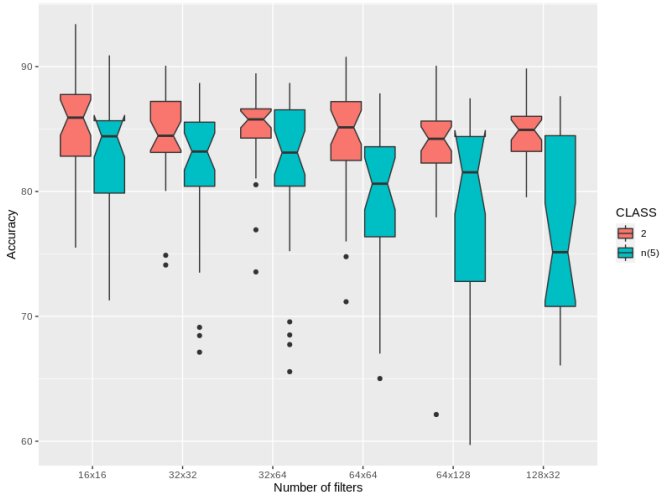

Fig. 5. Balanced accuracy comparison between binary training / binary classification approach and multi-class training / binary classification approach, using six number-of-filters configurations.

In a second comparison between binary classification models —one from binary training and other from multi-class training—, the multi-class training/binary classification approach did not exceed the binary training/binary classification approach even though the p-values increase. In all cases, as the p-values are lower than alpha, H_0 is rejected.

Table III presents the balanced accuracy of the tested approaches. The first approach — binary training and binary classification — shows a better performance than the other two. The difference against the second approach is overwhelming, while the one against the third approach is not as large as in the first comparison, but it is still statistically significant. The binary training approach has a better performance but cannot distinguish between the different types of crimes.

It is important to remember that this work aims to generalize suspicious behavior from different types of crimes. If the goal is to detect or prevent a specific kind of crime, it will be necessary to improve the multi-class detection approach. Also, the three methods present their best results when using 16 filters on both convolutional layers. This insight may support the development of a real-time detection app and a more agile neural network training process.

V. Conclusion

This work presented a comparison between three training approaches for suspicious behavior detection from different types of crime processed with the Pre-Crime Behavior method. As this method removes important information to detect crime in a video, the resulting samples have a high-visual similarity with normal-behavior samples. We looked for the best approach to avoid this similarity and classify the suspicious and normal behaviors correctly.

The multi-class training approach leads to multi-class and binary classification. The neural network trained each class by separate for the first classification approach. For the binary approach, we combined the suspicious behavior samples from different crimes into a single class. The first comparison was between binary training / binary classification and multi-class training/multi-class classification. The results showed an immense difference that favors binary training / binary classification, which achieved better performance (24.5% comparing the best models) than the second approach.

The second multi-class approach improves by reducing the classification from five to two classes. In this scenario, we consider as positive any suspicious-behavior sample classified in one of the four crime classes. After 30 runs of the model

TABLE I
P-values from t-tests comparing binary classification against multi-class classification, for each number-of-filters configuration.

Number-of-filters	p-value
16–16	$3.68e^{-19}$
32–32	$1.39e^{-23}$
32–64	$1.38e^{-20}$
64–64	$3.51e^{-21}$
64–128	$1.65e^{-17}$
128–32	$4.88e^{-23}$

TABLE II
P-values from t-tests comparing binary training / binary classification against multi-class training / binary classification, for each number-of-filters configuration.

Filter pair	p-value
16–16	$3.06e^{-2}$
32–32	$3.20e^{-2}$
32–64	$2.87e^{-2}$
64–64	$6.79e^{-4}$
64–128	$8.99e^{-3}$
128–32	$1.54e^{-5}$

TABLE III
BEST BALANCED ACCURACY (BACC) FOR EACH MODEL PER NUMBER-OF-FILTERS CONFIGURATION.

Number-of-filters configuration	Binary training / binary classification (%)	Multi-class training / multi-class classification (%)	Multi-class training / binary classification (%)
16–16	93.4	68.9	90.9
32–32	90.1	67.6	88.7
32–64	89.5	68.9	88.7
64–64	90.8	61.4	87.9
64–128	90.1	62.0	87.5
128–32	89.8	56.8	87.6

with both approaches and getting the balanced accuracy, we conducted a t-test on each pair of configurations. The results showed significance values to consider that the means are not equal and demonstrate that the first approach achieves a better performance again.

Since this approach does not search the detail of what type of crime does it belong to, it allows false positives between criminal classes. For example, even if the input is a 'shoplifting' sample and classified as 'stealing', the algorithm considers it a hit because it is within the crimes. We performed this adjustment because we considered it unfair that we were so strict with the multi-class training / multi-class classification model and detected each crime. With the binary, we considered only the type of behavior.

As mentioned, the binary training approach cannot identify a particular type of crime. This work aims to determine if suspicious behavior samples should be trained as a group, regardless of the nature of the offense. If the goal were to detect or prevent a particular type of crime, it would be necessary to improve the multi-class training approach or select a different one from the literature.

In a crime prevention scenario, as sooner the offender can be identified, the more time to react has the security team. PCB method and particularly Pre-Crime Behavior segments' training could improve the prevention by classifying a suspicious behavior from an early moment. A current limitation is that the algorithm returns a label from a complete sample, instead of a moment or a person committing something suspicious. As future work, we look for a generalization test by training with less criminal classes and testing with complete unknown ones. It will also be interesting to test these approaches to develop real-time detection for surveillance support by analyzing and labeling each group of frames separately. We consider it essential to collaborate with behavioral analysts from different areas, such as police officers, casino staff, behavioral experts, among others, to improve the project performance. We also look for implementing visualization techniques as Class Activation Maps or Saliency Maps to detect the regions where the offender is.

REFERENCES

[1] J. Tang and L. He, "Genetic optimization of BP neural network in the application of suspicious financial transactions pattern recognition," in *2012 International Conference on Management of e-Commerce and e-Government*, 2012, pp. 280–284.

[2] A. G. Pennington, J. L. Griffin, J. S. Bucy, J. D. Strunk, and G. R. Ganger, "Storage-based intrusion detection," *ACM Trans. Inf. Syst. Secur.*, vol. 13, no. 4, Dec. 2010. [Online]. Available: https://doi.org/10.1145/1880022.1880024

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, Jul. 2009.

[4] A. Wiliem, V. Madasu, W. Boles, and P. Yarlagadda, "A suspicious behaviour detection using a context space model for smart surveillance systems," *Computer Vision and Image Understanding*, vol. 116, no. 2, pp. 194–209, 2012.

[5] R. M. L. Pérez, F. G. León, and M. G. Olivares, *Comportamiento no verbal: más allá de la comunicación y el lenguaje*. Píramide, 2016.

[6] H. Wells, T. J. Allard, and P. Wilson, "Crime and CCTV in Australia: Understanding the relationship," in *Political Science*, 2006.

[7] Anti-Defamation League (ADL), "Recognizing and dealing with suspicious people," accessed May 4th 2020. [Online]. Available: https://www.adl.org/education/resources/tools-and-strategies/suspicious-people

[8] Berwyn Police Department, "What is suspicious activity?" 2012, accessed May 4th 2020. [Online]. Available: http://www.berwynpd.com/general_information/what_is_suspicious_activity

[9] D. Ross, "Defining 'suspicious behavior' without bias is harder than you think," 2018, accessed May 4th 2020. [Online]. Available: https://people.howstuffworks.com/defining-suspicious-behavior.htm

[10] Metropolitan Police Department, "Capital watch: What is suspicious behavior?" accessed May 4th 2020. [Online]. Available: https://mpdc.dc.gov/whatssuspicious

[11] Homeland Security, "What is suspicious activity?" accessed May 4th 2020. [Online]. Available: https://www.dhs.gov/see-something-say-something/what-suspicious-activity

[12] University of Michigan DPSS, "Reporte a crime or concern suspicious behavior," accessed May 4th 2020. [Online]. Available: https://dpss.umich.edu/content/services/report-a-crime/suspicious-behavior/

[13] Hilliard Police Department, "What is a suspicious person?" accessed May 4th 2020. [Online]. Available: https://hilliardohio.gov/wp-content/uploads/2018/07/suspicious-activity.pdf

[14] S. Penmetsa, F. Minhuj, A. Singh, and S. Omkar, "Autonomous UAV for suspicious action detection using pictorial human pose estimation and classification," *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, vol. 13, no. 1, pp. 18–32, 2014. [Online]. Available: https://elcvia.cvc.uab.es/article/view/v13-n1-penmetsa-minhuj-singh-omkar

[15] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6479–6488, 06 2018.

[16] G. A. Martínez-Mascorro, J. R. Abreu-Pederzini, J. C. Ortiz-Bayliss, and H. Terashima-Marín, "Suspicious behavior detection on shoplifting cases for crime prevention by using 3D convolutional neural networks." *arXiv:2005.02142v1 [cs.CV]*, 2020.

[17] C. Mu, J. Xie, W. Yan, T. Liu, and P. Li, "A fast recognition algorithm for suspicious behavior in high definition videos," *Multimedia Syst.*, vol. 22, no. 3, pp. 275–285, June 2016.

[18] X. Hu, J. Dai, Y. Huang, H. Yang, L. Zhang, W. Chen, G. Yang, and D. Zhang, "A weakly supervised framework for abnormal behavior detection and localization in crowded scenes," *Neurocomputing*, vol. 383, pp. 270–281, 2020.

[19] N. Vaswani, A. Roy Chowdhury, and R. Chellappa, "Activity recognition using the dynamics of the configuration of interacting objects," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, 2003, pp. II–633.

[20] F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Comput. Vis. Image Underst.*, vol. 115, no. 3, p. 323–333, Mar. 2011.

[21] M. Sabokrou, M. Fathy, and M. Hoseini, "Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder," *Electronics Letters*, vol. 52, no. 13, pp. 1122–1124, 2016.

[22] R. H. Huan, X. M. Tang, Z. H. Wang, and Q. Z. Chen, "Abnormal motion detection for intelligent video surveillance," in *Information Technology for Manufacturing Systems II*, ser. Applied Mechanics and Materials, vol. 58. Trans Tech Publications, 2011, pp. 2290–2295.

[23] X. Wang, H. Song, and H. Cui, "Pedestrian abnormal event detection based on multi-feature fusion in traffic video," *Optik*, vol. 154, pp. 22–32, 2018.

[24] G. Tripathi, K. Singh, and D. K. Vishwakarma, "Convolutional neural networks for crowd behaviour analysis: a survey," *The Visual Computer*, vol. 35, no. 5, pp. 753–776, 2019.

[25] *Detecting Suspicious Behavior From Only Positional Data With Distributed Sensor Networks*, ser. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. Volume 6: 5th International Conference on Multibody Systems, Nonlinear Dynamics, and Control, Parts A, B, and C, 09 2005.

[26] P. Bull, *Body Movement and Interpersonal Communication*. Wiley, 1983. [Online]. Available: https://books.google.com.mx/books?id=fY7BJQAACAAJ

[27] J. R. Goodall, E. D. Ragan, C. A. Steed, J. W. Reed, G. D. Richardson, K. M. T. Huffer, R. A. Bridges, and J. A. Laska, "Situ: Identifying and explaining suspicious behavior in networks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 204–214, 2019.

[28] L. Xia, B. Yang, and H. Tu, "Recognition of suspicious behavior using case-based reasoning," *Journal of Central South University*, vol. 22, no. 1, pp. 241–250, 2015.

[29] J. Lee, S. Lee, and S. Yang, "An ensemble method of CNN models for object detection," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 2018, pp. 898–901.

[30] S. Eum, H. Lee, H. Kwon, and D. Doermann, "IOD-CNN: Integrating object detection networks for event recognition," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 875–879.

[31] Y. Wang, W. Zhou, Q. Zhang, and H. Li, "Convolutional neural networks with generalized attentional pooling for action recognition," in *2018 IEEE Visual Communications and Image Processing (VCIP)*, 2018, pp. 1–4.

[32] R. R. Nair, N. Sankaran, B. U. Kota, S. Tulyakov, S. Setlur, and V. Govindaraju, "Knowledge transfer using neural network based approach for handwritten text recognition," in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, 2018, pp. 441–446.

[33] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.

[34] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, "Deep-cascade: Cascading 3D deep neural networks for fast anomaly detection and localization in crowded scenes," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1992–2004, 2017.

[35] L. Zhang, G. Zhu, P. Shen, and J. Song, "Learning spatiotemporal features using 3DCNN and convolutional LSTM for gesture recognition," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 3120–3128.

[36] M. Ueda, K. Ito, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, "An age estimation method using 3D-CNN from brain MRI images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 380–383.

[37] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI," *IEEE Access*, vol. 5, pp. 23 626–23 636, 2017.

[38] B. Khagi, C. G. Lee, and G. Kwon, "Alzheimer's disease classification from brain MRI based on transfer learning from CNN," in *2018 11th Biomedical Engineering International Conference (BMEiCON)*, 2018, pp. 1–4.

[39] J. Liu, J. Zhang, H. Zhang, X. Liang, and L. Zhuo, "Extracting deep video feature for mobile video classification with ELU-3DCNN," in *Internet Multimedia Computing and Service*, B. Huet, L. Nie, and R. Hong, Eds. Singapore: Springer Singapore, 2018, pp. 151–159.

[40] N. L. Hakim, T. K. Shih, S. P. Kasthuri Arachchi, W. Aditya, Y.-C. Chen, and C.-Y. Lin, "Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model," *Sensors*, vol. 19, no. 24, p. 5429, Dec 2019. [Online]. Available: http://dx.doi.org/10.3390/s19245429