**TELKOMNIKA Telecommunication, Computing, Electronics and Control**
Vol. 18, No. 5, October 2020, pp. 2447~2453
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i5.16634 

# Abnormal activity detection in surveillance video scenes

**Jwan Jamal Ali, Narjis Mezaal Shati, Methaq Talib Gaata**
College of Science, Department of Computer Science, Al-Mustansiriya University, Iraq

**Article Info**

***Article history:***

Received Feb 7, 2020
Revised Apr 25, 2020
Accepted May 8, 2020

***Keywords:***

Anomaly detection
Motion object detection
Real-time processing
Tracking
Video surveillance

**ABSTRACT**

Automated detection of abnormal activity assumes a significant task in surveillance applications. This paper presents an intelligent framework video surveillance to detect abnormal human activity in an academic environment that takes into account the security and emergency aspects by focusing on three abnormal activities (falling, boxing and waving). This framework designed to consist of the two essential processes: the first one is a tracking system that can follow targets with identify sets of features to understand human activity and measure descriptive information of each target. The second one is a decision system that can realize if the activity of the target track is "normal" or "abnormal" then energizing alarm when recognized abnormal activities.



***Corresponding Author:***

Jwan Jamal Ali,
College of Science, Department of Computer Science,
Al-Mustansiriya University,
Baghdad, Iraq.
Email: jwan.jamal484@gmail.com

## 1. INTRODUCTION

The Anomaly defined as a behavior that deviates based on what viewed as normal or standard according to the domain. The analysis of the abnormal activity in the video sequence has attracted remarkable and rapid attention in the field of video surveillance due to its efficiency by filtering a large number of unnecessary information and the availability of many human and material resources [1]. Video surveillance systems have become one of the most prominent research in security and emergency applications involving the identification and classification of human activities into normal or abnormal activities. In spite of the fact that there are many proposed algorithms and techniques in this motivating topic, research in this domain still insufficiency to two significant things: datasets that public accessible and general comparative assessment. Also, little changes in illumination, shadows, number of objects that appear in the scene can turn the look of the scene causing a lot of false alarms ratio, loss of object tracking, and other related problems that may not respond in real-time to any suspicious event. therefore, such systems have a wide area of research to promote some of their capabilities [1].

The term "surveillance" is the activity to look out for. Surveillance is the demonstration of observing tireless and transient objects inside a certain environment. It targets to formulate useful data from a large of "videos" overall by observation cameras using: detection, segmentation, interest objects tracking, automatic recognize and understand their behavior. They are significant tools that support humans by expanding their cognitive abilities and reasoning about different interesting situations [2].

– Research AIM

The goal of this paper is to design a system that has the ability to detect humans, tracking behavior, and classification as a normal activity (walking) or abnormal activities (falling, boxing and waving) from videos of monitoring students in an academic environment that takes into account the security and emergency aspects.

This paper proposes ways to complete the tracking and classification of human activity by preprocessing steps using the Gaussian mixture model (GMM) for detecting the moving objects. Then used the fuzzy C-means clustering (FCM) technique to segment the images for further customize the object and distinguish it from adjacent objects. Combined (HARRIS-SIFT) algorithms together to extract features and the Kalman filter to tracking targets. Finally, K-nearest neighbor (KNN) used to classify the activities "normal and abnormal”. When recognizing one of the abnormal activities the system generates sound alarms to identify this activity with a red label around the person or persons involved. Whilst the yellow label identified around the persons that the system recognized as normal activity.

– Literature review

In [3] Proposed a system to detect unusual activity in real-time video surveillance. This paper proposed a method to consist of three essential processes. First, background subtraction used to detect the motion. Second, skeletonization algorithms applied. Finally, unusual event detection by matching the Skeleton image frame with a reference image frame in the database and setting a red box on the frame.

In [4] Proposed of ISS to detect human behavior in a universal environment by using temporal-differencing for moving object detection and Gaussian function to locate motions region. This method used a filter, its shape model based on equation named (OMEGA) to ignore non-human objects and a support vector machine (SVM) to classify objects into normal and abnormal behavior, and use model for retrieving the object detected from the dataset to identification by used of content-based image retrieval (CBIR).

In [5] Proposed a security system to the detection of an anomaly motion by classifying different motions using the support vector machine (SVM) classifier. This was done by using the background subtraction to detect the moving objects and a Kalman filter for tracking. In [6] Proposed a system to detect object aware abnormal activity based on block foreground segmentation to restrict the analysis of moving objects from the sparse matrix. The objects are then represented using the trajectories and then the histogram is built.

## 2. DATASET

The video surveillance dataset includes four classes of activities recorded by a camera. Then separated into normal activities and abnormal activities. The normal activities include one class (walking) and the abnormal activities include three classes (falling, boxing and waving) from various scenarios at different times (morning, afternoon), sunny and cloudy weather, and different distance between the camera and persons. Figure 1 shown frames as examples of these activities.

The activities performed in our indoor academic department with a static camera (Logitech HD Pro Webcam C920), which contains123 videos (80 training, 43 test). All with AVI file format. The videos that used for training and testing have the following metrics:

- Frames per Second (fps): 30
- Resolution frame: 480x360
- Total Training Frames: 20453 frames
- Total Testing Frames: 10793 frames

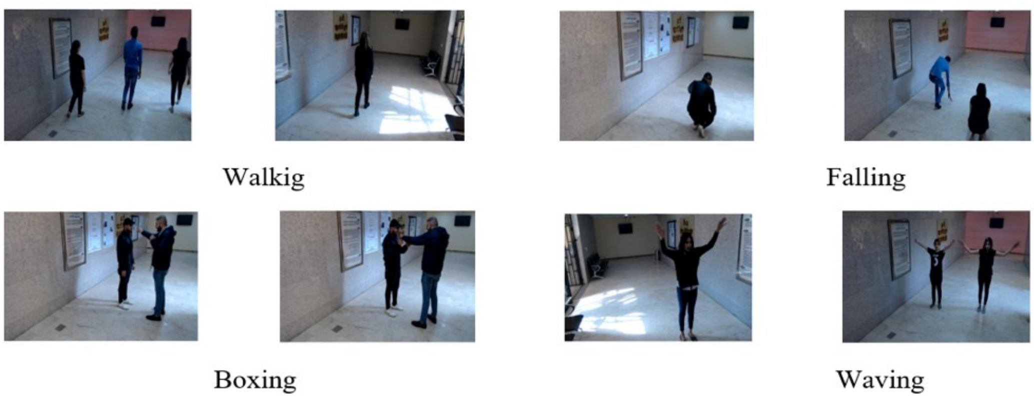


Figure 1. Examples of some the normal and abnormal of the human activities

## 3. RESEARCH METHOD

The proposed system focuses on monitoring human in an academic environment that has the ability to detect the moving objects from video by separating the foreground from the background, extract features of motion, tracking, and recognition of normal activities: (walking) and abnormal activities: (falling, boxing and waving). Figure 2 illustrated the proposed system. The abnormal activity detection system step can be summarized as the following steps:

Step1: Input video (raw data) from the camera sensor.

Step2: Pre-Processing with two stages:

- Convert video to frames
- Motion detection

In the first stage, video clips are converted into sequential frames and the target detection by the GMM algorithm in the second stage.

Step3: Use fuzzy C-means (FCM) clustering to more accurate analysis.

Step4: The features of targets for each sequential frame are extracted using (Harris- SIFT) features.

Step5: Use Kalman filter tracking to track targets.

Step6: Recognition of the activities and classification into normal or abnormal using the KNN algorithm.

Step7: Generate alarm when recognizing one of the abnormal activities.

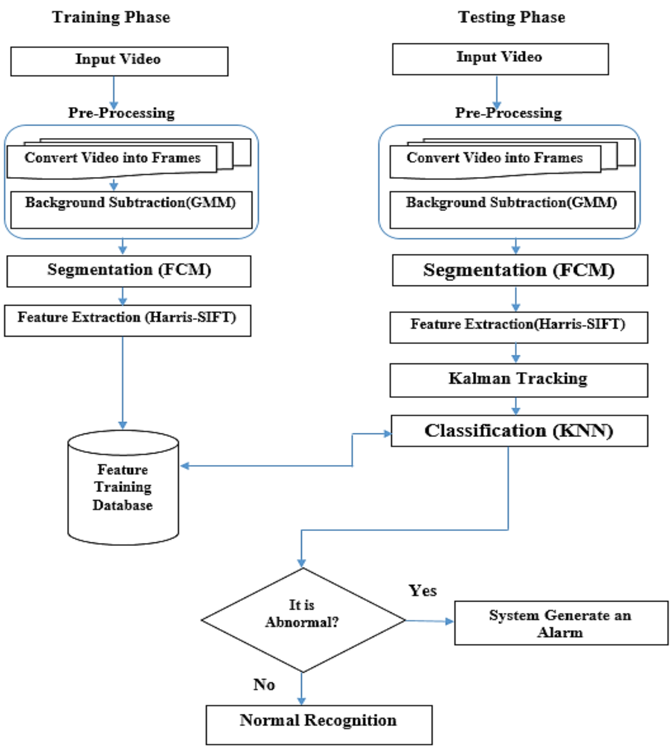


Figure 2. The proposed system diagram

### 3.1. Preprocessing

The primary preprocessing basis is to enhance image quality that prevents unwanted distortion. Making it appropriate for human interpretation and machine perception that improves the extraction of the important object features for further processing. Therefore, considered to be one of the essential and necessary phases of the video surveillance systems [7]. Videos are sequences of images every one of these images called a frame, showed in quick enough frequency for these reasons the eyes of humans can percept the congruity of its content. All techniques of image processing can be able to apply individual frames that manipulation and analysis, in particular, to enhance its quality within the limits of the work and purpose of the processing [8]. Figure 3 shows the video frames of different scenarios, and the frame sequence of videos dataset are visible.

Figure 3. Frames sequence of videos dataset

#### 3.1.1. Noise and images enhancement

Noise is an undesirable effect on pixel values and will present rather different intensity values within digital images such as background noise and blurred objects that always present through image acquisition, transmission, encoding and processing steps [9].

– Contrast enhancement

Contrast is a significant factor in each subjective assessment of image quality, whereas contrast is the variation in visual properties that make an object identifiable from other objects and background. Color image contrast enhancement is done by converting an image to another color space (L*a*b*) on L* (luminosity layer) the contrast adjustment accomplished after that transformed the image back to the color space (RGB). The pixels intensity impacts the luminosity, whereas real colors preserved [10].

– Spatial filtering

Spatial filtering modifies the image by exchanging the value of each by pixel with the function of the values of that pixel and it's nearby [11]. By using the following syntax:

h = fspecial ('disk’, radius)

#### 3.1.2. Background subtraction

A technique used to detect motion in the video scenes by using the video frames to subtract the current frame from the background model that previously obtained and classified through two main steps: background modeling and foreground extraction [12]. Gaussian mixture model (GMM) is one of the popular and robust techniques for construct a background model to segment moving objects from the background that assigns to a probability density function represented as a sum of Gaussian densities. Based on variance and persistence, Gaussians categorized as "foreground" and "background". The value of pixels if not represent the background distributions then take into the foreground this will be done until it is Gaussian with consistent and sufficient evidence supported [13].

Accurate foreground detection is a difficult task in real-time because the real-world of the video frames sequences include many critical situations. The main challenge to change detection algorithms is the casting shadows accompanying foreground objects. For those reasons, the foreground mask was processed using a 3×3 medium filter with a morphological operation (opening and closing). It is often used since it is enough to eliminate noise and at the same time preserve the edges. See Figure 4 the experiment processing frames in indoor activities as showed in: (a) the frame with a moving of the object, (b) the contrast enhancement, (c) the foreground detected with the shadow, and (d) showed the foreground mask after the processing.


Figure 4. The experiment processing frames in indoor activities: (a) frame with a moving object, (b) contrast enhancement, (c) foreground detected with shadow, and (d) foreground mask after processed

### 3.2. Image segmentation

Digital image partitioning into multiple parts (groups of pixels) referred to as image segmentation. The objective is to facilitate (and/or) modifying the regions of interest in the image to be simpler analyzed and more beneficial processed especially after foreground extraction when an inaccurate merger occurs between people and the opacity of the object. fuzzy C-means (FCM) is a soft clustering algorithm where each element can have a place with more than one gathering, thus (FCM) can be extremely fast because of the number of iterations demand to obtain a specific clustering practice identity to the demanding accuracy [14].

### 3.3. Motion tracking and feature extraction

For tracking any object, feature extraction plays a significant role. Combined (Harris and Scale Invariant Feature Transform (SIFT)) which is proposed in this system to extract features of objects that reduction of dimensionality and reduce the number of resources needed to describe a large range of data [15-17]. Tracking objects detected in frames sequence and matching them is a critical stage of intelligent security systems because of the ability to extract the objects and analyze their behavior [18]. Kalman filter is a mathematical and recursive filter. The word "filter" is utilized because it is the process of returns the best estimate of noisy data a "filter out" the noise. It is an estimate acquired by integrated both "prediction" and "correction" [19, 20]. The major objectives to use a Kalman filter are as follows:

- Predict the future location of objects.
- Reduce noise due to inaccurate detection.
- Associate multiple objects for their tracks.

### 3.4. Action recognition and classification

After extraction features, the classification algorithm applied to images or videos [20, 21]. Classification is the data analysis process by a set of training data to find a model that distinguishes and describes data classes and thus recognition of activities that used for the objective of abnormality detection [1]. Low complexity and simple yet efficient in many cases classifier algorithm named "K-Nearest Neighbors (KNN)". This algorithm classified by matching the anonymous data with groups of similar trained data using the "Euclidean Distance" as a similarity measure. to prevent the larger range attributes from overriding of the smaller range attributes, attribute values are set [22, 23]. The classification of (KNN), meant that the most anonymous pattern assigned the most prevalent classes among the nearest neighbors' classes. In [24, 25] if there are two classes linked, the link of the lowest average distance assigned to the anonymous pattern [26, 27].

## 4. EXPERIMENTAL RESULTS AND DISCUSSION

The random nature and high complexity of human movements, in addition to, the issue of inadequacy datasets that publicly accessible leads to difficulty and tricky of classifying suspicious human activities of interest and makes the detection of suspicious activities is challenging. These issues leading to inconsistencies in experimental results in several papers in literature. To validate the proposed system performance, we applied the approach to a video sequence in real-time that recorded at our academic department in an indoor environment with the static camera (Logitech HD Pro Webcam C920). For this purpose, MATLAB (R2019a) tool used and implemented on a PC with a 2.80 GHz processor core i7 and 16.0 GB RAM.

In fact, that in most of the tracking scenarios, there is not much change in the object's articulation between the successive frames. So, after the experiments, we have adopted in our system the basis of processing every three frames instead of the high computational of frame-to-frame feature matching that requires extensive processing to run in real-time. The experimental results of the system showed achieved promising results in most test conditions. To evaluate the system performs the following metrics used:

Accuracy (ACC)

$$ACC = \frac{No.of\ objects\ correctly\ classified}{Total\ no.of\ occurrence\ of\ the\ activity} \quad (1)$$

Detection Rate (DR)

$$DR = \frac{\text{True Positive}}{\text{True Positive + FalseNegative}} \quad (2)$$

False Alarm Rate (FAR)

$$FAR = \frac{\text{False Positive}}{\text{True Negative + False Positive}} \quad (3)$$

The experimental result of action classification is as shown in Figure 5. Table 1 Shows samples of the experimental results that were listing to show the quantitative results associated with activity interest behavior.

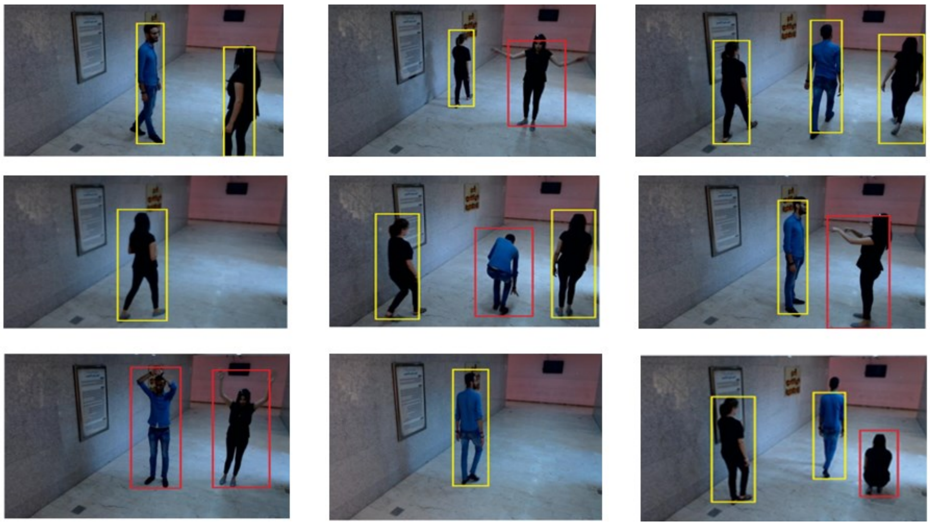

Figure 5. Result of action classification in a different scene

Table 1. Experimental result

| Number of people | Behavior | Accuracy% | Detection Rate% | False Alarm Rate% |
|---|---|---|---|---|
| 1 | normal | 99.15% | 99.28% | 2.05% |
| 1 | abnormal | 97.14% | 98.07% | 3.05% |
| 2 | Normal and abnormal | 96.50% | 98.18% | 4.12% |
| 2 | normal | 97.05% | 96.04% | 5.84% |
| 2 | abnormal | 98.80% | 98.78% | 4.88% |
| multiple | normal | 94.56% | 96.57% | 8.54% |
| multiple | abnormal | 94.79% | 95.59% | 9.59% |
| multiple | Normal and abnormal | 93.96% | 95.56% | 9.45% |

## 5. CONCLUSION AND FUTURE SCOPE

This work proposes an automated real-time video surveillance system to track targets and recognize normal and abnormal human activity in an academic field that takes into account the security and emergency aspects. The test of results of the real-time video sequences demonstrate the effectiveness of design to recognize and evaluate human activity due to the high values of accuracy, detection rate while a low false alarm rate. As future work classified more suspicious activities that can identify accurate actions. More advanced algorithms designed for real-time video surveillance.

## ACKNOWLEDGMENT

The author thanks the Department of Computer Science, College of Science, Mustansiriyah University, for supporting this work.

## REFERENCES

[1] Al-Dhamari A. A., Sudirman R., Mahmood N. H., "Abnormal Behavior Detection in Automated Surveillance Videos: A Review," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 19, pp. 5245-5263, 2017.

[2] Sahasri M., C. Gireesh, "Object Motion Detection and Tracking for Video Surveillance," *International Journal of Engineering Trends and Technology*, pp. 161-164, 2017.

[3] Adhvaryu A., Kalpesh J., "Real Time Unusual Event Detection in Video Sequences," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 4, no. 3, pp. 565-572, 2015.

[4] Al-Nawashi M., Al-Hazaimeh O. M., Saraee M., "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments," *Neural Computing and Applications,* vol. 28, pp. 565-572, 2017.

[5] Zaidi S., et al., "Video Anomaly Detection and Classification for Human Activity Recognition," *International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC),* pp. 544-548, 2017.

[6] Gogawale R. C., Kakade K. A., Kale P. N., Yadav A. S., Hanwate P. S., "Detecting Abnormal Activities from Input Videos and Reporting to Authorities," *International Journal of Advance Research and Innovative Ideas in Education,* vol. 5, no. 1, pp. 301-307, 2019.

[7] Chen Y., Chen A., Jiang Z., Zhong J., "Surveillance Video Portrait Recognition Preprocessing Technology for Police Actual Combat," *Open Journal of Applied Sciences*, vol. 09, no. 05, pp. 394-402, 2019.

[8] Student P. G., Engineering C., College B. V. M. E., and Nagar V. V., "A Survey on Object Detection and Tracking," *International Journal of Advance Engineering and Research Development*, vol. 3, no.01, pp. 2970-2978, 2018.

[9] Alazawi S. A., Shati N. M., Abbas A. H., "Texture features extraction based on GLCM for face retrieval system," *Periodicals of Engineering and Natural Sciences,* vol. 7, no. 3, pp. 1459-1467, 2019.

[10] Mariena A. A. & Sathiaseelan J. G. R., "Contrast enhancement of grayscale and color images using adaptive techniques," *International Journal of Engineering and Technology*, vol. 7, no. 2, pp. 1-4, 2018.

[11] Pérez-Benito C., Morillas S., Jordán C., Conejero J. A., "Smoothing vs. sharpening of colour images: Together or separated," *Applied Mathematics and Nonlinear Sciences*, vol. 2, no. 1, pp. 299-316, 2017.

[12] Shati N. M., Alazawi S. A., Abdulbaqi H. A., "Background Subtraction (BS) using Instant Pixel Histogram," *Journal of Southwest Jiaotong University,* vol. 54, no. 5, pp. 1-6, 2019.

[13] Goyal K., Singhai J., "Review of background subtraction methods using Gaussian mixture model for video surveillance systems," *Artificial Intelligence Review*, vol. 50, no. 2, pp. 241-259, 2018.

[14] Manikandan S. S. S., "Outdoor Scene Image Segmentation Based on Kernelized Fuzzy C Means Clustering and Pom," *International Journal of Current Engineering and Scientific Research,* vol. 6, no. 3, pp. 20-26, 2019.

[15] Alshibani D. R., Shati N. M., & Ahmed N. T., "DNA Genetic Recombination based Image Encryption using Chaotic Maps," *Indian Journal of Public Health Research & Development*, vol.10, no. 6, 2019.

[16] Awad A. I., Hassaballah M., "Image Feature Detectors and Descriptors: Foundations and Applications," *Studies in Computational Intelligence*, 2020.

[17] Dong C., Liu J., Xu F., Liu C., "Ship detection from optical remote sensing images using multi-scale analysis and fourier HOG descriptor," *Remote Sensing*, vol. 11, no. 13, pp. 1-19, 2019.

[18] Wen L., et al., "Multi-Object Detection and Tracking," *International Journal for Research in Applied Science Engineering & Technology,* vol. 6, pp. 809-813, 2018.

[19] Tirupathamma M., "Object Detection and Tracking in video using Kalman Filter," *International Journal of Recent Advances in Engineering & Technology,* vol. 5, no. 5, pp. 102-106, 2016.

[20] Kumar A., Kushwaha S., "A Framework for Human Activity Recognition using Pose Feature for Video Surveillance System," *International Journal of Computer Applications*, vol. 1, pp. 1-4, 2016.

[21] Hota S., Pathak S., "KNN classifier-based approach for multi-class sentiment analysis of twitter data," *International Journal of Engineering & Technology*, vol. 7, no. 3, pp. 1372-1375, 2018.

[22] Isa N. E. Z. Md, Amir A., Ilyas M. Z., Razalli M. S., "The Performance Analysis of K-Nearest Neighbors (K-NN) Algorithm for Motor Imagery Classification Based on EEG Signal," *Matec Web Conf.*, vol. 140, 2017.

[23] N. D. Zaki, N. Y. Hashim, Y. M. Mohialden, M. A. Mohammed, T. Sutikno, and A. H. Ali, "A real-time big data sentiment analysis for iraqi tweets using spark streaming," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 6, pp. 1411-1419, 2020.

[24] O. A. Hammood, M. N. M. Kahar, W. A. Hammood, R. A. Hasan, M. A. Mohammed, A. A. Yoob, et al., "An effective transmit packet coding with trust-based relay nodes in VANETs," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 2, pp. 685-697, 2020.

[25] A. F. M. Raffei, T. Sutikno, H. Asmuni, R. Hassan, R. M. Othman, S. Kasim, et al., "Fusion Iris and Periocular Recognitions in Non-Cooperative Environment," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 7, no. 3, pp. 543-554, 2019.

[26] Z. F. Hussain, H. R. Ibraheem, M. Alsajri, A. Hussein Ali, M. A. Ismail, S. Kasim, et al., "A new model for iris data set classification based on linear support vector machine parameter's optimization," *International Journal of Electrical & Computer Engineering*, vol. 10, no. 1, pp. 1079-1084, 2020.

[27] T. Zhang, Z. Yang, W. Jia, B. Yang, J. Yang, and X. He, "A new method for violence detection in surveillance scenes," *Multimedia Tools and Applications,* vol. 75, no. 12, pp. 7327-7349, 2016.