International Journal of Advanced Science and Computer Applications
DOI: 10.47679/ijasca.v3i1.
https://ijasca.org/index.php/ijasca

RESEARCH ARTICLE

Abnormal event detection using intelligent video surveillance

M.Shaban [1], Marwa Elpeltagy [1], Ahmed Y Khedr [1], A.Al-Marakbey[1]

Abstract

Intelligent surveillance systems have a large number of cameras installed. Abnormal vehicle or human entry at a certain location or time may potentially result in monetary loss and/or fatalities. This study develops a multi-surveillance camera intelligent surveillance system that is new, adaptable, and fast. The user may choose the number of interest zones with any polygon shape for each camera. Furthermore, the sort of abnormal item and the direction of abnormal motion for each location separately. To identify items in a video frame, the Single Shot Multi-Box Detector (SSD_MobileNet_v3) deep neural network is utilized. After that, these items are tracked using a Kernelized Correlation Filters (KCF) tracker in order to identify the direction of aberrant motion. Also, a novelty is to determine the people's motion type, i.e., running or walking, by establishing a relationship between the real human dimension and the observed distances in the video. The system's performance is evaluated on both the Authentically Distorted Surveillance Videos dataset and the newly collected dataset. An accuracy of 88.22% has been scored for event detection and F1-score of 87%. for people's motion classification. The experimental results confirm the superiority of the suggested method over the current state-of-the-art methods.

Keyword: Intelligent Video Surveillance, Abnormal Behavior Detection, SSD, Human motion classification.

Systems and Computer Engineering Department, Faculty of Engineering, Al-Azhar University

*) *corresponding author*

M.Shaba

Email: mariamshaban@azhar.edu.eg

Received 18 October 2023 | Received in revised form 18 December 2024 | Accepted 11 February 2024
Available online 10 March 2024

Introduction

The video surveillance systems were designed for human operators to observe protected space remotely. Watching surveillance video is a tedious task and human observers can easily lose attention. Automation can help to reduce the manpower and increase the performance. It is very important to develop an intelligent video system that utilizes technology to automatically detect multiple abnormal behaviors.

There are various types of abnormal behaviors such as persons walking on the highway, baby climbing window or balcony fences, person intrusion, and Vehicles on a pedestrian walkway. Moreover, trucks on small cars ways, tamper with antiquities in Museums, and a car moving in the opposite direction are examples of abnormal behaviors. Also, they include a person entering the building in going out time and a human running inside a building actions.

Some of these abnormal behavior detection systems depend on the time of the action like a person in the bank at night. Also, some of them rely on the location of the event like a person climbing a fence. Other abnormal behavior detection systems depend on the motion type, the motion direction, and the motion speed.

These systems provide multiple options for the user to determine in each camera like surveillance time range. They allow the user to choose multi-regions of interest each with a specific abnormal type (persons, trucks,), different shape, and different direction. In addition, they can provide surveillance of three cameras at the same time each with multi-regions.

Kim et al. [1] introduced a method to detect intrusion in a restricted area. You Only Look Once (YOLO)detector is employed to identify the location of the pedestrian at a given interval.

Moreover, the vernalized correlation filter (KCF) [2] tracks the location in the rest of the video frames. The coordinate information of the intrusion and roaming area is obtained from a predefined Extensible Markup Language (XML) file. The ROI in the input image is determined according to the coordinates. The coordinate information of the tracking object is acquired to determine whether the object is inside or outside the ROI area.

Intrusion is determined by multiplying PersonN (x, y) × ROI (x, y). One of the drawbacks of this method is the difficulty of differentiating between partial and total intrusion because any intersection between person and ROI will give a value and count intrusion. Dohun Kim's method is evaluated using the KISA dataset which is clear except some videos have low-light environments.

In this paper, an abnormal behavior detection system was developed to cover most of abnormal situations with high performance. In the proposed the scheme, SSD_MobileNet_v3 [3] technique is employed for detection and KCF is utilized for Tracking with a detection rate of D/15F.

In the beginning, the algorithm decides to start surveilling or not depending on comparing the abnormal time with the actual time. If the system started working, all objects in the frame will be detected.

For each object, It checks if the object type is abnormal for the regions of interest or not. It also checks if this object is inside this region or not and if it is partially or totally included. After that, it tracks the object to determine its direction. Moreover, the abnormal human speed is detected. If the object is included in the region of interest, the system will give alarm. Also, if the direction is abnormal direction or the speed is abnormal, the system will give alarm.

In the proposed method, the user can draw any polygon shape and the system can extract the polygon vertices. In addition, it has the ability of differentiating between partial and total intrusion. It concerns four points in the BB. The decision depends on its positions corresponding to ROI. Also, the system provides multi-ROI each with different abnormal scenarios according to object type and direction.

Moreover, the proposed method performance is evaluated using the challenging Authentically Distorted Surveillance Videos dataset. The genuine distortions affecting these recorded videos are Defocus aberration (defocus), over-exposure, sub-exposure (exposure), and a combination of defocus and exposure. An average accuracy of 88.22% has been scored which differs according to intrusion type i.e. (prowl, walking, or running) and detection rate.

In addition, a novel human pose-based speed detection method is proposed employing key point information extracted from image. The required key points should include at least 6 points on the leg (left_ankle, left_knee, left_hip, right_hip, right_knee, right_ankle). Also, the static view from the side is necessary. The angle between the runner's legs can be obtained and utilized to determine the actual running frequency. Furthermore, the actual distance between the two legs can be easily obtained by ratio conversion. A relationship between the real distances (average human width in meters) and the observed distances in the video (BB width in pixels) is constructed. Cross multiplication is employed to calculate the real distance using the pixel's distance. In addition, the speed is computed and an F1_score of 87% has been scored. The proposed method is more general and works without any constraints compared to Zhao et al. [4] method. The proposed scheme is evaluated on a new large-scale dataset based on a combination of two popular datasets along with a newly collected dataset; the Authentically Distorted Surveillance Videos dataset, UCSD_Anomaly_Dataset. The newly collected part is gathered from YouTube outdoor videos, and Google images.

The proposed system accuracy depends on the detection rate for videos. According to the video, an accuracy rate of 88.22% is achieved for D/3F (one detection every three frames). Also, a rate of 84.67% is scored for D/15F, and finally, a rate of 78.59% is achieved for D/30F.

For images, an accuracy rate of 89.33% is achieved. Image accuracy is higher than videos because indoor videos are much distorted.The rest of the paper is organized as follows: Section 2 introduces a review of abnormal video detection methods and popular existing datasets. Section 3 proposes a new architecture for abnormal video detection. Section 4 is dedicated to the experimental results and analysis. Section 5 presents the conclusion and future work.

Related work

The progress of deep learning algorithms has raised the ease of abnormal action detection within video. Several methods have been elaborated to detect these actions relying on either traditional methods or the advanced deep learning-based approaches. Convolutional neural network (CNN), and Recurrent neural networks (RNN) are examples of powerful tools which can be employed for this purpose.

In [1], The abnormal intrusion is detected using YOLOv4. The 3D convolutional neural network (3D-CNN) is used for fall and violence detection. The suggested technique is evaluated using KISA Datasets and an average Recall Rate of 93.8% is achieved.

Additionally, a Deep Learning system [5] is developed for Suspicious Activity Detection within Surveillance Video. The framework consists of two Proceedings: convolutional neural networks CNN and Recurrent Neural Network (RNN). VGG-16 CNN is utilized for fracture extraction from video frames. LSTM is used for order dependence learning in the video frame sequence. An accuracy of 87.15% is achieved for CAVIAR and KTH datasets.

In Sun et al. [6], an adversarial 3D convolutional auto-encoder is introduced to capture robust normal spatio-temporal patterns. It consists of a 3D convolutional encoder and a 3D de-convolutional decoder that are used to encode the normal patterns in video sequences. An average accuracy rates of 90.6%, 92.7%, and 74.6% are achieved for UCSD, SUBWAY, and SHANGHAITECH, respectively.

Furthermore, Bouindour et al. [7] Proposed a 3D Convolutional Auto Encoder. a new loss function that combines MSE and compactness Mahalanobis loss is developed. This loss function helped in extracting robust spatiotemporal features while minimizing the hypersphere that encompasses the target class representations.

TensorFlow and Alex Net are used in [8] to recognize faces and detect intruders. Fires are detected using motion and color information. Balance point change, angles and movement distances of objects are used for Loitering detection. Finally, falls are detected using motion and acceleration features of the object. The system accuracy is 88.51% for intruder detection, 92.63% for fire detection, 80% for loitering detection, and 93.54% for fall detection.

In [9] anomalies within videos are detected. (CNN) is used for appearance encoding for each frame. Convolutional Long Short- Term Memory (ConvLSTM) is used for memorizing all previous frames to extract motion information. ConvNet and ConvLSTM are combined with Auto-Encoder to learn the regularity of appearance and motion for ordinary moments. ConvLSTM-AE accuracy rates are Avenue 77%, Ped1 75.5%, Ped2 88.1%, Subway Entrance 93.3%, and Exit 87%.

Moreover, Hasan et al. [10] detected irregular frames by learning regular patterns using autoencoders. The process began by learning a fully connected autoencoder using conventional spatiotemporal local features. Then, in a single learning framework, a fully convolutional autoencoder is constructed to learn both the local features and the classifiers. The accuracy is 70.2% for CUHK Avenue, 81.0% for UCSD Ped1, 90.0% for UCSD Ped2, 94.3% for Subway Entrance, and 80.7% for Subway Exit.

Adversarial event prediction (AEP) is presented in [11]. AEP derives the prediction model from normal event samples so that it can identify the relationship between the current and future course of events during the training phase. Adversarial learning for both the past and future of events is introduced to acquire the prediction model. The suggested adversarial learning constrains the learning for past events representation and compels AEP to learn the representation for forecasting future events.

Furthermore, Liu et al. [12] presented an Abnormal Human Activity Recognition system. The Bayes Classifier and VGG-16 are employed to differentiate between walking, running, punching and tripping. The extracted features are length, width ratio, entropy, and Hu invariant moment. KTH dataset is used for evaluation. The recognition accuracy based on Bayes reached 88%, 92%, 92% and 100% for each activity. The recognition accuracy based on CNN reached 92%, 96%, 100% and 100% for each activity.

Fence Climbing is detected Using Activity Recognition in Kolekar et al. [13]. Also, person motion is detected using background subtraction. The researcher concerned with two main features. These features are Centroid of the blob and centroid variations. Support Vector Machine classifier is used for classification of walking, climbing down, and climbing up activities.

Walking is a common human activity. Thus, if a human run suddenly, it may indicate that an abnormal event has been occurred. Also, speed detection is used to track physical fitness. Therefore, many algorithms were created to decide people's motion type i.e. (running or walking).

Lao et al. [14] presented a standard benchmark database with a diversity of scenes and ground truth for human running detection. The researcher developed a method based on the Farnebäck optical flow method to distinguish between running and walking. The system reached an average precision and recall rate of 67.8% and 90.1%, respectively.

Furthermore, a human pose-based system is designed in [4] to extract the image's key point information to detect the speed. The required key points should include at least 6 points on the leg (left_ankle, left_knee, left_hip, right_hip, right_knee, right_ankle). Also, the static view from the side is necessary to extract the key points.

The angle between the runner's legs can be obtained employing the key point information obtained. Based on the angle information, the actual running frequency is determined. After that, the actual distance between the two legs can be easily obtained by the ratio conversion. Experiments indicate that the detection accuracy of Simple Baselines can reach 89.3% in a certain situation using the MPII data set. Furthermore, the average relative error of speed sequences is 4.89%, which meets the practical detection requirements.

Seethi & Bharti [15], Wrist-worn Wearable Sensors is utilized for CNN-based Speed Detection i.e. (Walking and Running) with high precision. Data from 15 participants are collected while they are walking/running at different speeds on a treadmill. Accelerometer output and gyroscope sensory output from the wrist-worn device are the input for CNN individually. The max pooling outputs of Accelerometer and gyroscope sensory data are Concatenated and fed into dense layer then output layer for speed classification.

Moreover, Generative Adversarial Nets (GANs) are a wonderful way to solve classification issues because they can identify important features in the frames without the need for predefined anomaly categories. GANs are used in M.

Ravanbakhsh et al. [16] for abnormal event detection. An accuracy of 97.4% for UCSD Ped1, 93.5% for Ped2, and 99% for UMN has been reached. Ganokratanaa et al. [17] introduced a Deep Spatiotemporal Translation Network (DSTN) framework. This technique incorporates deep convolution neural network concepts from the GAN-based Edge Wrapping approach. This improves anomaly localization so that an accuracy of 98.5%, 95.5%, and 99.6% has been reached for UCSD Ped1, UCSD Ped2, and UMN, respectively.

Additionally, Ghedia & Vithalani [18] devised a surveillance system designed to detect outdoor objects against intricate and ever-changing backgrounds. They achieved this using a modified Gaussian Mixture Model (GMM) and Adaptive Thresholding, effectively addressing shadows and partial occlusions.

In a review study, researchers in [19] conducted a comparison of various research papers to illustrate different categories of suspicious events using both supervised and unsupervised machine learning techniques. Nandhini & Brindha, [20] developed a transfer learning model based on SSD to detect helmets and multiple riders from surveillance footage. Another work [21] introduced an intelligent vehicle tracking model based on queries to aid in the retrieval of stolen vehicles and identification of accident-involved vehicles. Also, in [22], a robust unsupervised approach utilizing deep autoencoders was presented for detecting anomalies in surveillance videos.

The proposed system

Figure 1 shows the proposed algorithm. At First, the user enters the required information through the GUI. These information's include working time, number of regions, object type, and motion direction. The event time is an important factor in deciding if the event is normal or abnormal. For example, Governmental institutions, banks, and markets work only at a certain time. If there is a motion in the closing time, it is categorized as abnormal motion. Consequently, the user must provide the start and the end time of detection. The camera works only at that time and switches off automatically otherwise. Then, the user can draw polygons for each region. Vertices for each polygon are extracted and SSD is employed to detect object types to determine the abnormal objects depending on their position.

KCF tracker is used to track these objects to determine their motion type and direction. These steps are explained in details in the following sections.

Single Shot Multi-Box Detector (SSD)

The Object type is the main essential factor which decide if the motion is normal or not. For example, in fence climbing, if it happens by a cat, it is normal but if it happens by humans, it is abnormal. Thus, we must know the object type. The system allows choosing the object type i.e. (persons, trucks, vehicles, persons& trucks, or all). There are two main categories of target detection methods: single-stage algorithms based on regression, like YOLO [23] and SSD [24], and two-stage algorithms based on proposed regions, such Faster R-CNN [25]. SSD employs a pyramid feature layer- based detection technique on feature maps of various sizes, performing both location regression and SoftMax classification.SSD also uses Faster R-CNN's anchor concept, with variable scales, aspect ratios, and previous frames; this will make it more accurate in identifying and localizing objects of varying sizes. the single-shot multi-box detector SSD is a robust object detection technique based on a feed-forward convolutional neural network (CNN). A set of bounding boxes and associated scores are generated, indicating the presence of object class instances in these boxes. The network in the frontend and an additional feature extraction layer at the backend compose the SSD network structure, due to the large computation and parameters of the VGG_16 network. Mobile_Net_V3 [26] replaces the original VGG_16 network. SSD_MobileNet_v3 is employed for this function because it is faster than YOLO while the detection performance is approximately the same.

Detection of abnormal events in a restricted region:

Abnormal event detection in a certain region in the camera view is needed in many cases as shown in Figure 2 The system allowsthe user to choose any shape for the region of interest.

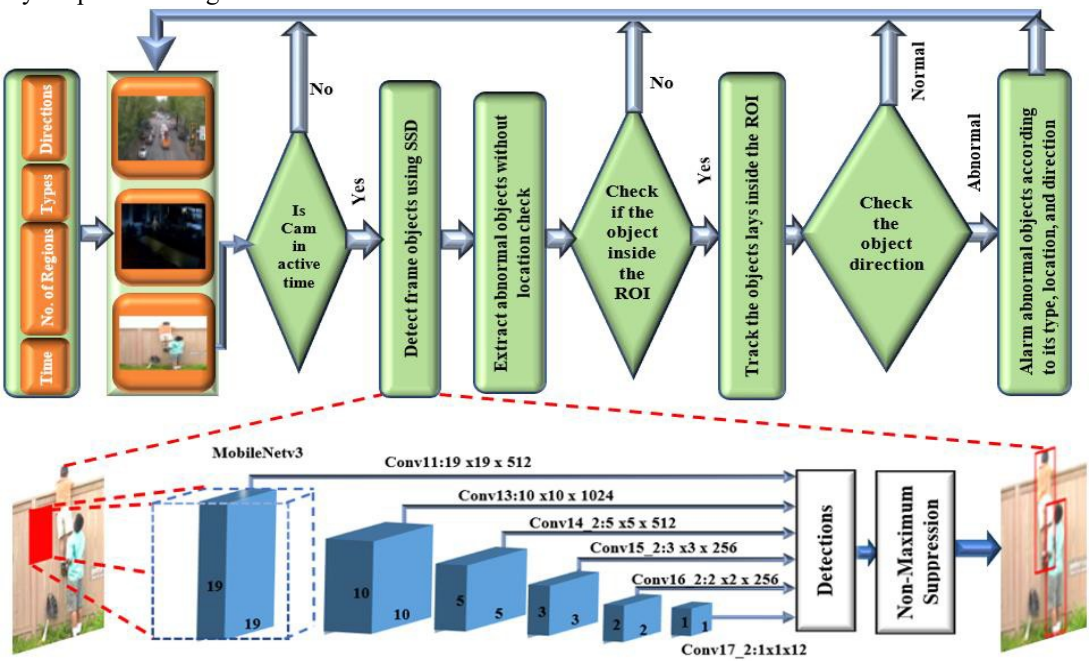


Fig. 1. system Flow chart with SSD_MobileNet_V3

Fig. 2. Detect certain abnormality in a restricted region

After the user draws the polygon, the system extracts the polygon Vertices and saves it in a three-dimensional list. This list contains three lists (list for each camera). Each camera list consists of several lists equal to the number of regions in this camera. Region list length equals to the number of polygon vertices in this region. From each polygon vertices list, we can determine if the object is partially or totally inside the region. To verify if the BB is inside the polygon, we need to carry out the following steps:

First all BB (bounding boxes) that lays above and under the polygon are excluded.

Second, for the points 1,2,3, and 4 of the BB investigate if this point lays inside or outside the polygon as Figure 3 a. Then check:

If all bounding box points (1, 2, 3 and 4) are inside the polygon, the object is inside (alarm) as BB_5 in Figure 3 b, Figure 3 c. If some points of the bounding box are outside, the object is partially inside (warning) as BB_6. As shown in Figure 3 c.

If all points of the bounding box are outside, the object is totally outside (do nothing).

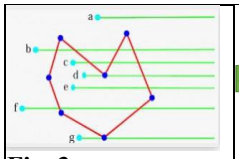


Fig. 3. a.

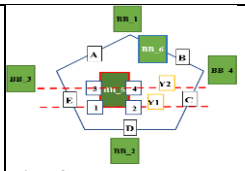


Fig. 3. b.

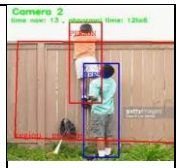


Fig. 3. c.

Human motion classification:

Walking is a common human activity. So, if a human runs suddenly, it is an indication that an abnormal event has been occurred.The Kernelized Correlation Filters (KCF) is employed for tracking objects due to its high speed and accuracy. KCF has gainedwidespread acceptance. It is suitable for raw pixel values in addition to the histogram of oriented gradient features. KCF is a variant of correlation filter in which the correlation between two samples is computed. The correlation scores the highest valuewhen these samples match.

The proposed algorithm has the ability to determine the people's motion type i.e. (running or walking). The main idea of the algorithm is to establish a relationship between the real distances and the observed distances in the video. Most adults have a pacing distance between 61:91.4 cm [27]. Figure 4 a, and average body breadth is between 58:65 cm [27]. as shown in Figure 4 b. Assuming the average human width is 0.65 meters for both the front view and the pacing distance for the side view. Consequently, we assume that the human width is one meter after wrapping BB around him. There are two BBs, BB1 for the first human location and BB2 for the second location. The average BB width is calculated in pixels: $D_{1\_pixels} = (width_1 + width_2)/2$ (1). Then, the distance between the centers of BB1 and BB2 is computed which represents the distance he ran in pixels: D2_pixels = $\sqrt{(X2 + Y2)}$ (2), as shown in Figure 4 c. If we assumed that D1_meter equals 1 meter, we can get

$$D2meter = \frac{D2pixels}{D1pixels} * D1meter$$

The moving-distance time is equal to the time between two consecutive frames. Finally, human

speed equal D2 meter/time. The average human walking speed is 1.4 m/s [28]. So, any speed above 3 m/s is considered running.

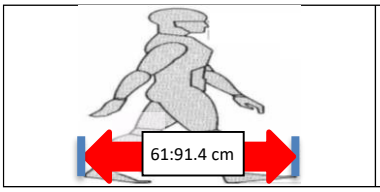


Fig. 4. a. Person pacing zone

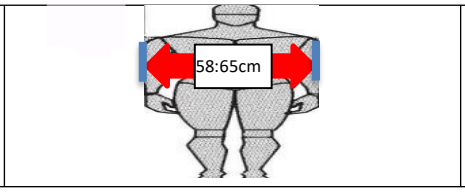


Fig. 4. b. Average body breadth

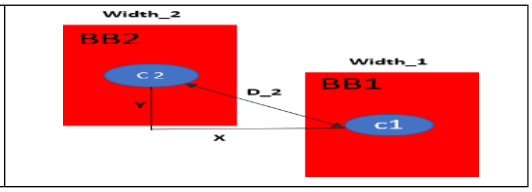


Fig. 4. c. distance calculation

Fig.5. a. Vehicle's wrong direction detection

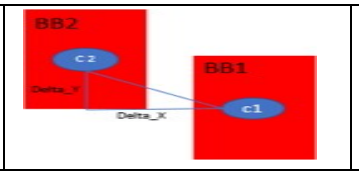


Fig.5. b. Distances calculations

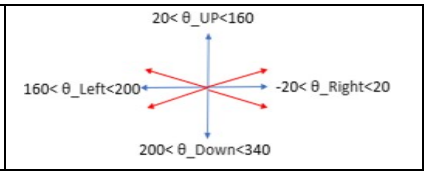


Fig.5. c. Direction calculation

Detection of the motion direction:

Abnormal event detection sometimes depends on the direction of the motion. For example, it is not allowed for cars to move in the opposite direction on the road as shown in Figure 5 a. For this situation, we let the user enter the abnormal direction (up, down, left, right, up&right,.. and stop) for each region as a list for each camera. After detecting an abnormal object in a certain region, the system tracks it to determine its direction and decides if it is normal or not. By comparing the new and old BB locations for each object. The differences in x_center, and y_center sites is calculated as shown in Figure 5 b. Also, the angle between BB1_center and BB2_center is computed. According to the angle value, the motion direction is defined as shown in Figure 5. c. due to shaking in the outdoor cameras sometimes. It may cause an inaccurate BB location. So, the direction is considered only if four consecutive frames give the same direction.

Anomalies detection for multiple regions and multiple cameras:

The ability to choose multi-regions in a single camera view each having any polygon shape and a certain abnormal event type is very useful. Instead of using multi-cameras each concerned with a certain region with different abnormal events, now only one camera makes the work of multi-cameras. For example, when using it in a highway region, the person's motion is abnormal,and in another region, the truck's motion is abnormal as shown in Figure 5 a. The user provides the number of regions he needsin a certain camera and a list of abnormal object types of each region for each camera. This list is used to alarm only the regionthat has the abnormal event. The proposed algorithm provides a multi-camera surveillance service. The system can cover 3 cameras each having all abilities we described above.

Experimental results

This section covers the dataset details, evaluation criteria and the dataset results as follow**.**

Dataset

The system performance is evaluated on both the Authentically Distorted Surveillance Videos dataset [29] and the newly collected dataset. A new small dataset for abnormal behavior detection is collected. It consists of 75 video clips for outdoors

which are gathered from YouTube. 45 videos of them are abnormal events and the rest are normal. In addition, there are 150 images which are collected from Google images and UCSD_Anomaly_Dataset for different events. 995 video clips from the Authentically Distorted Surveillance Videos dataset are included for indoor events. Defocus aberration (defocus), over- exposure, sub-exposure (exposure), and a combination of defocus and exposure are the genuine distortions affecting these recorded videos. After that, the merged dataset is divided into two parts; image dataset and video dataset.

The Video dataset: It is composed of 24 videos for Trucks detection and 51 video for Vehicle's wrong direction detection. In addition, there are 324 videos for Person Running intrusion (PR) and 321 videos for Person Walking intrusion (WL). Moreover, Person Prowl intrusion (PW) class has 350 videos in which the person makes suspicious movements.

The Images datasets:

This dataset includes different anomalous scenarios like Tamper with antiquities in Museums, Persons Walk on the highway, vehicles on pedestrian walkway, baby climbing window or balcony fence, and person intrusion.

System GUI (Graphical user interface):

To facilitate the interaction between the user and the proposed system a Graphical user interface (GUI) is designed as in Figure 6. The user can enter all abnormal conditions for each camera. The GUI screen is divided into five columns. The user can enter the surveillance time range in the first and second columns. In the third column, the user can provide the number of regions he needs to construct each camera along with a list containing abnormal object types for each region in the fourth column. Finally, in the fifth column, the user can provide a list containing the abnormal motion direction for each region.

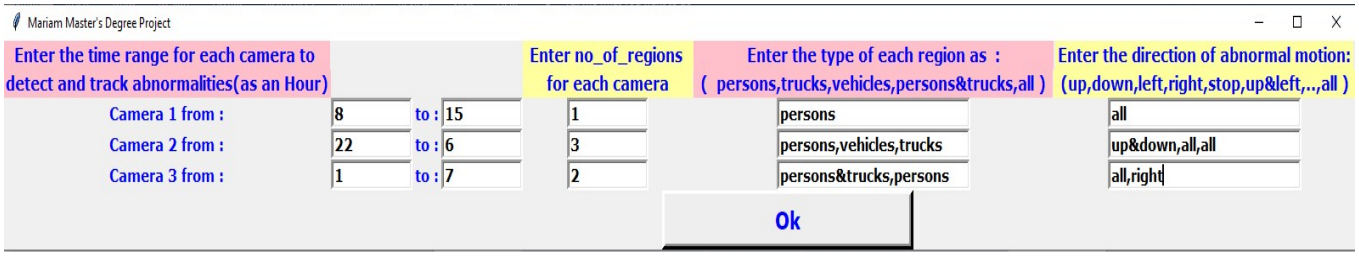


Fig 6. Abnormal behaviour detection GUI

Evaluation measures (metrics)

The accuracy, Recall, Precision and F1-Score are the popular evaluation metric which are utilized to assess the usefulness of the suggested abnormal event detection method. Speed is also an essential evaluation criteria in such applications. The equations of these evaluation measures are defined as follows

$$\text{accuracy} = \frac{\text{number of true negatives} + \text{number of true positives}}{\text{total number of samples}} \rightarrow (4)$$

$$\text{recall} = \frac{\text{number of true positives}}{\text{number of false negatives} + \text{number of true positives}} \rightarrow (5)$$

$$\text{precision} = \frac{\text{number of true positives}}{\text{number of false positives} + \text{number of true positives}} \rightarrow (6)$$

$$\text{F} - \text{measure} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \rightarrow (7)$$

Experimental results and discussion

The surveillance system efficiency depends on two factors, the system speed, and the abnormal event detection accuracy. Thesefactors will be discussed in the following sections.

System speed

The system speed may be affected by many factors such as the type of the applied detector and tracker. The number of working cameras, and the number of tracking objects are another factors. Also, the frame resolution affects the system speed. Accordingto the applied detector, Table 1. Shows the time efficiency for SSD_MobileNet_V3 and Yolo_v3 detectors. The recorded results approve that SSD is 2.5 times faster than YOLO, and more effected by frame resolution.

$$\text{Time decreased percentage} = \frac{REsolution\ 2\ time - resolution\ 1\ time}{resolution\ 1\ time} * 100 \quad (8)$$

Another study is carried out to detect the effect in the analysis time if there is an increase in the number of cameras, the video resolution, and finally the number of tracked objects. the study done using SSD for detection and KCF for tracking. The researchis conducted on two videos with different resolutions. The first resolution is twice the second. The first video includes only one-person, while the second video includes 8 persons. Results demonstrate that the number of frames analyzed per second decreased with an average of 2:3 frames when the number of cameras increased by one. The frame rate decreased by 5 F/S on average when the resolution was duplicated as shown in the Figure 7. It's decreased from 21 to 16 for one camera and from 19 to 14 fortwo cameras. Finally, when increasing the number of tracked objects the frame rate decreases as shown in the Figure 7. If one person is tracked for the same video and the same number of cameras, the frame rate reaches 21 F/S. If 4 people are tracked, itreaches 17 F/S and 13 F/S for 8 persons. The system frame rate depends on the detection rate for high-resolution videos (Authentically Distorted Surveillance Videos dataset). The System frame rate reaches 14 F/S for the detection rate D/3F, 16 F/Sfor the detection rate D/15F, and 17 F/S for the detection rate D/30F as shown in e Table 2.

Event detection accuracy rates

There are many types of intrusion motion, and each of them makes human shape differ in image. As we can see, there are three types of intrusion motion: walking, prowling, and running. Each of them has a different detection accuracy. Walking intrusion detection reaches the highest event detection rate with the distorted dataset compared with other intrusions due to the normal human shape in this case. Moreover, Prowl Intrusion has the lowest event detection rate compared to other intrusions due to the abnormal human shape in this case. Also, the video distortion causes loss in object detection in many frames. Run intrusion detection accuracy is between walking and prowling due to slightly abnormal human shape and speed. In addition, people may enter and exit without being detected in the small regions.

Roads have many abnormal events. Many roads accept only small vehicles and reject trucks either all the time or at certain times. A vehicle driving in the wrong direction is another

abnormal situation. The wrong direction detection depends on a specific area in the road because it is normal to drive up on half of the road while it is abnormal to drive down on the same half. Thus, multi regions each with different abnormality conditions are used. All these situations have been experimentally covered.

At the beginning, the surveillance system looks at the event time to decide if it is normal time or not. In the case of abnormal time, the system check if the object is inside the zone or not depending on the object type, shape, and direction.

In the case of image dataset, the conducted experimental results approved that the proposed algorithm scored an accuracy of 89.33%. According to the video dataset, Table 3, and Figure 8. illustrates the accuracy rates at different resolution levels. The recorded results show that the proposed system achieved an average accuracy of 88.22%.

TABLE 1. SSD V.S YOLO Time Efficiency

Model	Resolution :580*326	Resolution :1280*720	Average frame analysis time	Average no. of frame analysed/sec	Time decreased percentage
YOLO	0.174	0.211	0.1925	5.19	21.264%
SSD	0.0731	0.080	0.07655	13.06	9.986%

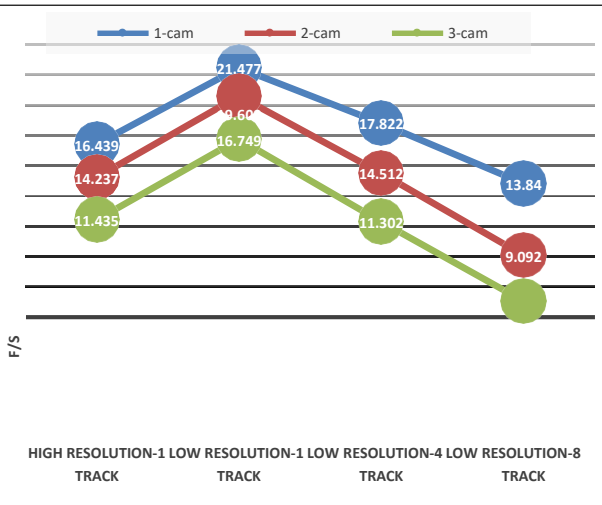


TABLE 2. System speed affected by detection rate for high resolution dataset.

Detection rate	Frame rate
D/3F	14
D/15F	16
D/30F	17

TABLE 3. Videos dataset Recall accuracy results.

Event \ Detection rate	D/3F	D/15F	D/30F
Trucks detection	95.83%	95.83%	95.83%
Wrong direction detection	92.15%	88.23%	88.23%
Prowl intrusion	83.14%	75.14%	64.57%
Walking intrusion	94.08%	92.52%	86.60%
Running intrusion	86.73%	85.80%	83.02%
Average accuracy	88.22%	84.67%	78.59%

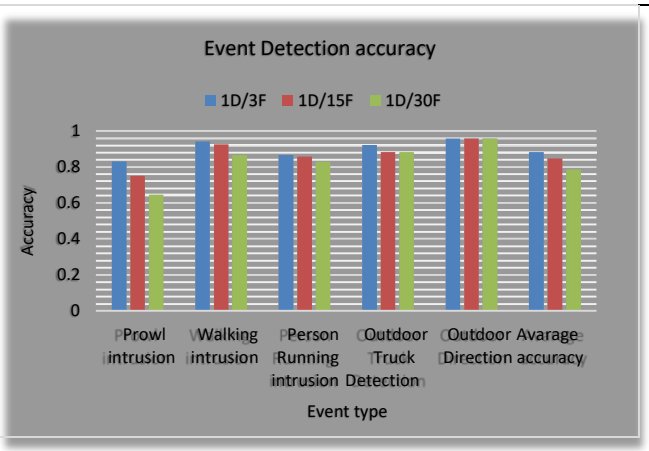


Fig. 8. Videos dataset Recall results.

Running detection accuracy rates

The run detection technique is applied to automatically classify the human motion type i.e. (run or walk). The technique results depend on the tracking algorithm. Many tracking methods have been applied as shown in Figure 9. and Figure 10. According to the experimental results, KCF is the best one compared to Median Flow, CSRT and Boosting. Boosting has the highest recall rate reaching 96%. However, it has a very low precision rate 54.7% which means half of the alarms are wrong. Consequently, we are concerned with the F1-score which is a combination of precision and recall rates. KCF has the highest F1-score reaching 87%. The Recall rate of applying KCF with a detection rate of 15 for run detection reaches approximately 82%. The precision rate reached 93.6% which means only 6.4% of alarms are wrong using KCF.

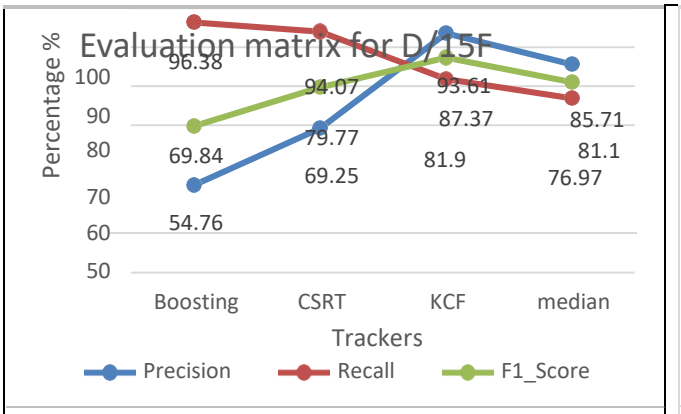


Fig. 9. Evaluation matrix for different trackers with D/15F

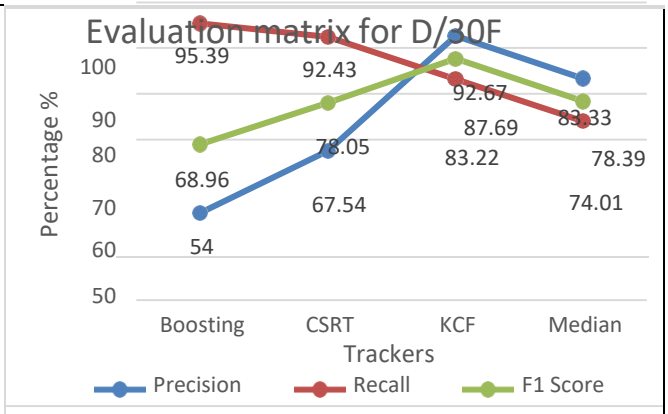


Fig. 10. Evaluation matrix for different trackers with D/30F

Figure 11. shows some images dataset results. and Figure 12. shows some videos dataset results.

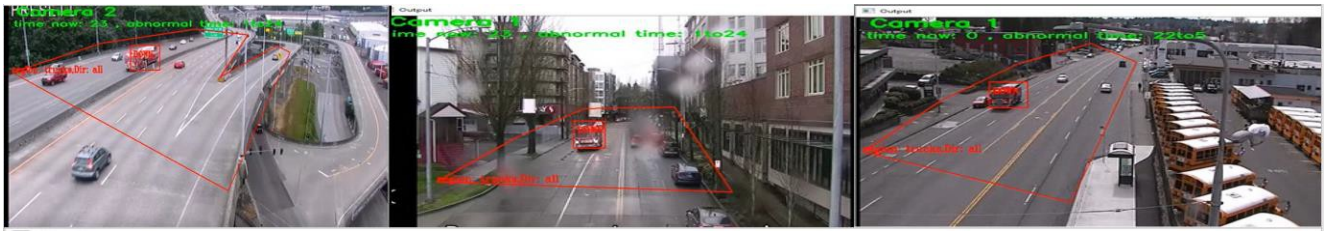
a.

b.

c.

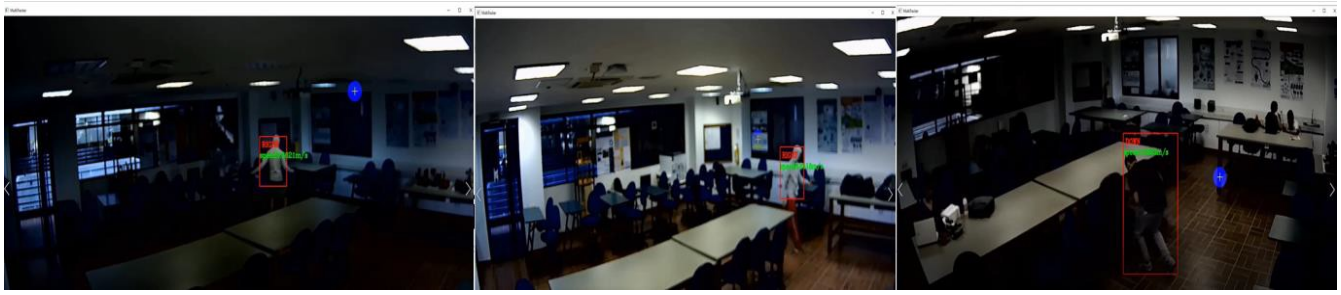
d.

Fig. 11 Image dataset results **a.** tamper with antiquities in Museums **b.** Persons Walk on the highway **c.** Vehicles on pedestrian walkway **d.** baby climbing window or balcony fence.

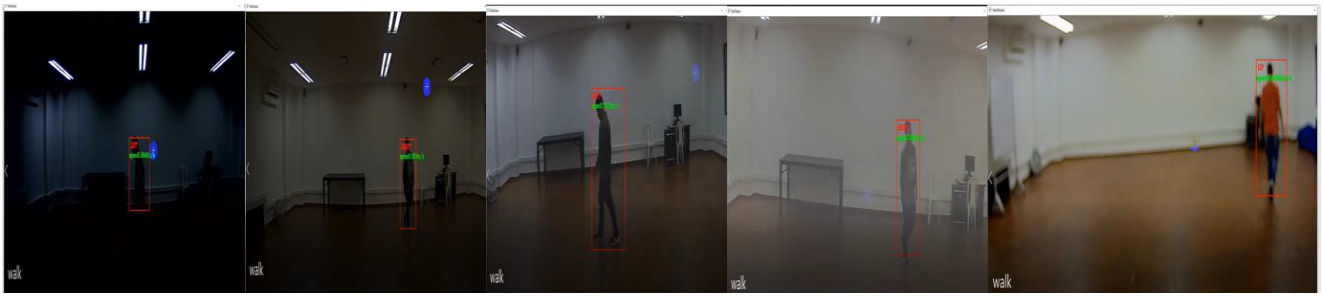


a.

b.



c.



d.

e.

Fig. 12. Videos dataset results **a.** Truck detection **b.** Wrong Direction detection **c.** Prowl intrusion detection **d.** Walking intrusion detection **e.** Run intrusion detection

System Setups

The experiments were conducted on Windows 10 and an HP laptop that has an Intel (R) Core (TM) i7-9750H CPU-16 GB and an RTX 2060 GPU-6 GB. The Python programming language, version 3.7.4, was used to implement the proposed model. Keras, Tensorflow, OpenCV, Sklearn, Xgboost, Numpy, Random, OS, and PIL are some of the libraries in Python which were employed for achieving the suggested model.

Conclusion

In this work, a new methodology for intelligent video surveillance based real-time anomalous behavior identification is demonstrated. This methodology is designed to detect many abnormal scenarios with multi-surveillance cameras. For each camera, the time range for abnormality detection and the number of interest regions with can be determined by the user. Moreover, the user can determine a specific abnormal object type and the direction of abnormal motion for each region individually. In the proposed method, the abnormal events depend on the event time, the location, the moving object type and the motion direction. In addition, the human speed is included. The suggested scheme is composed of two parts; a detector for object detection and a tracker for tracking the motion of abnormal objects to determine their speeds and directions. The SSD is used as a detector and KCF is used as a tracker. The human motion i.e. (running or walking) is classified by establishing a relationship between the real distance and the observed one in the video. The system is evaluated using the Authentically Distorted Surveillance Videos dataset along with the newly collected dataset. The experimental results approve that the proposed technique is extremely effective.

References

[1] Kim, D., Kim, H., Mok, Y., & Paik, J. (2021, July 2). Real-Time Surveillance System for Analyzing Abnormal Behavior of Pedestrians. Applied Sciences, 11(13), 6153. https://doi.org/10.3390/app11136153.

[2] George, M., Jose, B. R., & Mathew, J. (2018). Performance Evaluation of KCF based Trackers using VOT Dataset. ProcediaComputer Science, 125, 560–567. https://doi.org/10.1016/j.procs.2017.12.072

[3] Jian, Z., Yonghui, Z., Yan, Y., Ruonan, L., & Xueyao, W. (2020, November 28). MobileNet-SSD with adaptive expansionof receptive field. 2020 IEEE 3rd International Conference of Safe Production and Informatization(IICSPI). https://doi.org/10.1109/iicspi51290.2020.9332204

[4] Zhao, Z., Lan, S., & Zhang, S. (2020, October). Human Pose Estimation based Speed Detection System for Running on Treadmill. 2020 International Conference on Culture-Oriented Science & Technology (ICCST). https://doi.org/10.1109/iccst50977.2020.00108.

[5] Amrutha, C., Jyotsna, C., & Amudha, J. (2020, March). Deep Learning Approach for Suspicious Activity Detection from Surveillance Video. 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). https://doi.org/10.1109/icimia48430.2020.9074920.

[6] Sun, C., Jia, Y., Song, H., & Wu, Y. (2021). Adversarial 3D Convolutional Auto-Encoder for Abnormal Event Detection in Videos. IEEE Transactions on Multimedia, 23,3292–3305. https://doi.org/10.1109/tmm.2020.3023303.

[7] Bouindour, S., Hu, R., & Snoussi, H. (2019, June). Enhanced Convolutional Neural Network for Abnormal Event Detectionin Video Streams. 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). https://doi.org/10.1109/aike.2019.00039.

[8] Kim, J. S., Kim, M. G., & Pan, S. B. (2021, November 21). A study on implementation of real-time intelligent video surveillance system based on embedded module. EURASIP Journal on Image and Video Processing, 2021(1). https://doi.org/10.1186/s13640-021-00576-0.

[9] Luo, W., Liu, W., & Gao, S. (2017, July). Remembering history with convolutional LSTM for anomaly detection. 2017 IEEE International Conference on Multimedia and Expo(ICME). https://doi.org/10.1109/icme.2017.8019325.

[10] Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A. K., & Davis, L. S. (2016, June). Learning Temporal Regularity in Video Sequences. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2016.86.

[11] Yu, J., Lee, Y., Yow, K. C., Jeon, M., & Pedrycz, W. (2022, August). Abnormal Event Detection and Localization via Adversarial Event Prediction. IEEE Transactions on Neural Networks and Learning Systems, 33(8),3572–3586. https://doi.org/10.1109/tnnls.2021.3053563.

[12] Liu, C., Ying, J., Han, F., & Ruan, M. (2018, July). Abnormal Human Activity Recognition using Bayes Classifier and Convolutional Neural Network. 2018 IEEE 3rd International Conference on Signal and Image Processing(ICSIP). https://doi.org/10.1109/siprocess.2018.8600483

[13] Kolekar, M. H., Bharti, N., & Patil, P. N. (2016, November). Detection of fence climbing using activity recognition by Support Vector Machine classifier. 2016 IEEE Region 10 Conference (TENCON).https://doi.org/10.1109/tencon.2016.7848029

[14] Lao, S., Wang, D., li, F., & Zhang, H. (2016, December). Human running detection: Benchmark and baseline. Computer Vision and Image Understanding, 153, 143–150. https://doi.org/10.1016/j.cviu.2016.03.005

[15] Seethi, V. D. R., & Bharti, P. (2020, September). CNN-based Speed Detection Algorithm for Walking and Running usingWrist-worn Wearable Sensors. 2020 IEEE International Conference on Smart Computing (SMARTCOMP). https://doi.org/10.1109/smartcomp50058.2020.00064

[16] Ravanbakhsh, M., Nabi, M., Sangineto, E., Marcenaro, L., Regazzoni, C., & Sebe, N. (2017, September). Abnormal eventdetection in videos using generative adversarial nets. 2017 IEEE International Conference on Image Processing (ICIP). https://doi.org/10.1109/icip.2017.8296547.

[17] Ganokratanaa, T., Aramvith, S., & Sebe, N. (2020). Unsupervised Anomaly Detection and Localization Based on Deep Spatiotemporal Translation Network. IEEE Access,8,50312–50329. https://doi.org/10.1109/access.2020.2979869.

[18] Ghedia, N. S., & Vithalani, C. H. (2020, October 12). Outdoor object detection for surveillance based on modified GMM and Adaptive Thresholding. International Journal of Information Technology, 13(1), 185–193.https://doi.org/10.1007/s41870-020-00522-9

[19] Verma, K. K., Singh, B. M., & Dixit, A. (2019, September 20). A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. International Journal of Information Technology, 14(1), 397–410. https://doi.org/10.1007/s41870-019-00364-0

[20] Nandhini, C., & Brindha, M. (2022, August 23). Transfer learning based SSD model for helmet and multiple rider detection. International Journal of Information Technology,15(2),565–576. https://doi.org/10.1007/s41870-022-01058-w

[21] Sreedhar, S., Philip, A. O., & Sreeja, M. U. (2023, August 24). Autotrack: a framework for query-based vehicle tracking and retrieval from CCTV footages using machine learning at the edge. International Journal of Information Technology, 15(7), 3827–3837. https://doi.org/10.1007/s41870-023-01415-3

[22] Mishra, S., & Jabin, S. (2023, December 24). Anomaly detection in surveillance videos using deep autoencoder. International Journal of Information Technology,16(2),1111–1122. https://doi.org/10.1007/s41870-023-01659-z

[23] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. IEEE Conference on Computer Vision and Pattern Recognition, 89-95.

[24] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. Computer Vision – ECCV 2016, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

[25] Ren, S., He, K., Girshick, R., & Sun, J. (2017, June 1). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031.

[26] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, (2019) Searching for MobileNetV3. arXiv:1905.02244v5 [cs.CV]

[27] Panero, J., & Zelnik, M. (1979, November 1). Human Dimension and Interior Space. Watson-Guptill. http://books.google.ie/books?id=pQ1QAAAAMAAJ&q=)Human+Dimension+and+Interior+Space&dq=)Human+Dimension+and+Interior+Space&hl=&cd=2&source=gbs_api

[28] Browning, R. C., Baker, E. A., Herron, J. A. and Kram, R. (2006). Effects of obesity and sex on the energetic cost and preferred speed of walking. Journal of Applied Physiology. **100** (2): 390–398.

[29] Franco, C. A. (2022, May 18). Authentically Distorted Surveillance Videos Dataset. IEEE DataPort. https://ieee-dataport.org/open-access/authentically-distorted-surveillance-videos-dataset(Ren et al., 2017)