GALAXY INTERNATIONAL INTERDISCIPLINARY RESEARCH JOURNAL (GIIRJ)
ISSN (E): 2347-6915
Vol. 10, Issue 2, Feb. (2022)

ANOMALOUS HUMAN ACTIVITY RECOGNITION FROM VIDEO SEQUENCES USING BRISK FEATURES AND CONVOLUTIONAL NEURAL NETWORKS

Vishnu Priya P.
Research Scholar, Department of Computer Applications,
Bharathiar University, Coimbatore

Rajeswari R.
Associate Professor, Department of Computer Applications,
Bharathiar University, Coimbatore

ABSTRACT

In the present world, smart video surveillance system is essential for people to utilize distinctive sorts of security system to keep their property safe from unapproved person's entry. A security system helps individuals to feel safe while they have to travel or go out of their home for work purposes and others. The current security systems against robbery are costly as a lot of money must be paid to the administration supplier to store the recorded video despite the very fact that there's no human movement recognized. The solution to this problem is an intelligent video surveillance system that can detect anomalous human activities automatically. This eventually minimizes the specified space for storing and makes the system cost-effective. A typical smart video surveillance system consists of two steps namely, feature extraction and classification. In the proposed research work the Binary Robust Invariant Scalable Points (BRISK) features are extracted from the given videos during the feature extraction step. The obtained features are given as input to 3D Convolutional Neural Network (CNN) which classifies the videos as anomalous and normal based on the features. The present work uses University of California (UCF) dataset which contains thirteen video sets with one normal and twelve abnormal activity videos to evaluate the proposed method. The proposed BRISK and CNN based method provides a 100% training and 87.5% testing accuracy which is high compared to the existing methods that are based on shallow neural networks in detecting anomalous activities.

Keywords: video surveillance, anomalous, extraction, network.

INTRODUCTION

Intelligent video surveillance has become more popular among homeowners, businesses, neighbourhoods, communities, towns, and major cities because the system is effective in handling the time taken to watch the traditional video. It improves security, simplifies the monitoring of videos, and helps to stop criminal activity or suspicious acts. Video surveillance systems have a number of applications including elderly care, home nursing, security, traffic monitoring, fire detection, and human action understanding [Ramachandra et al., 2020]. The ability of intelligent video systems allow homeowners, businesses, and other municipalities to cut back the quantity of data gathered, highly reduces the quantity of manpower required and hours required to manage the video system.

The aim of Human activity recognition (HAR) is to recognize activities from a video surveillance and observe the actions of the subject and the environmental conditions. The detection and localization of anomaly behaviour in the crowded scenes are elaborated [Li et al., 2013]. A new framework is presented in [Mahadevan et al., 2017] for anomaly detection in crowded scenes. A study on human action recognition in video surveillance system is given in [Hemangee et al., 2019]. They describe the methods for extraction of different moving objects from surveillance video and identify the face and action.

In this paper, a video surveillance system is proposed to detect the human activities as abnormal and normal. In the proposed system BRISK feature extraction method is used to extract the features. 3D CNN are more accurate when compared to other CNN techniques. Hence, in the proposed method 3D CNN takes the 3D BRISK features of videos as input to categorize the videos as anomalous and normal videos. A benchmark dataset viz., University of California (UCF) dataset is used so that null frames, blur or any other problems in the videos can be avoided. The rest of the paper is organized as follows. Section 2 gives a brief review of the related work. Section 3 describes the involved in the proposed video surveillance system. Section 4 presents the experimental results obtained using the proposed method for the UCF dataset. Section 5 gives the conclusion for this chapter

LITERETURE REVIEW

Benzeth et al. [2012] summarize a survey on the study of the background subtraction algorithm. The main aim of their survey paper is to provide a solid analytical ground to underscore the strengths and weaknesses of the most widely implemented motion detection methods. They make use of 29 video sequences 15 real-time videos, 10 semi-synthetic containing up-to 3000 frames.

Sultani et al. [2019] explain the variety of realistic anomalies in surveillance videos that includes both normal and anomalous videos. In their paper, they have divided the normal and abnormal videos into bags and video segments, where a deep anomaly ranking model is used so that it predicts high anomaly scores for anomalous video segments. Here they have used a new large-scale dataset of 128 hours of videos, with 1900 long untrimmed real-world surveillance videos with 13 realistic anomalies such as fighting, abuse, robbery and shooting.

Miao et al. [2016] proposed an intelligent video surveillance system based on moving object detection and tracking. The operation of this system still relies on artificially found abnormal events and incidents. This intelligent video processing technology has strong functions, low labor cost, flexible applications, and effectively detects the target.

Kaliraj et al. [2015] explained the moving object detection in the surveillance systems. A novel approach for object detection in video surveillance is presented. This system consists of various steps including video compression, object detection, and object localization.

Shafie et al. [2014] developed a traffic signal optimization using vehicle flow statistics. The traffic signal light changes its color based on the number of vehicles in the signal to avoid traffic jams. A background model technique is being proposed in this system to successfully detect target objects such as motorbike, car and bus. The surveillance test result that has been conducted early in the morning of the day is discussed.

Tavagad et al. [2016] developed a smart video surveillance system using Raspberry Pi. This is a new technology which is less expensive in their work, and used as a standalone platform for image processing. Using web applications the information is captured and transmitted via a 3G dongle to a smart phone by the Raspberry Pi. A high-speed video monitoring sub-system, using 802.11 is used and realized. Video surveillance is used for real-time applications and multimedia surveillance.

Zablocki et al. [2014] provide a survey about the intelligent video surveillance system installed in public spaces. Smart video surveillance system is developed due to the advancements in the algorithms of video content analysis. The variety of solutions in a system based on object detection, tracking, movement analysis, vehicle detection and traffic or parking are discussed.

Muhammad et al. [2018] developed a fire detection and localization in video surveillance applications based on deep CNN. This system is mainly developed for fire detection so that it can avoid any fire disasters. CNN is used because of its less memory and computational requirement compared to other deep neural network. CNN approaches are mainly used throughout their work.

PROPOSED METHODOLOGY AND IMPLEMENTATION

The proposed system is based on BRISK features and CNN for detecting anomalous activities in video sequences. Figure 1 represents the steps involved in proposed method. The first step is the feature extraction. The proposed system mainly focuses on video surveillance system with BRISK features and 3D CNN. The proposed work uses 3D CNN for classifying or recognizing various normal and abnormal actions. The 3D CNN is considered as very effective method due to their feed-forward nature while compared to other classification techniques. 3D CNN has huge computation and memory efficiency compared to other deep learning networks.

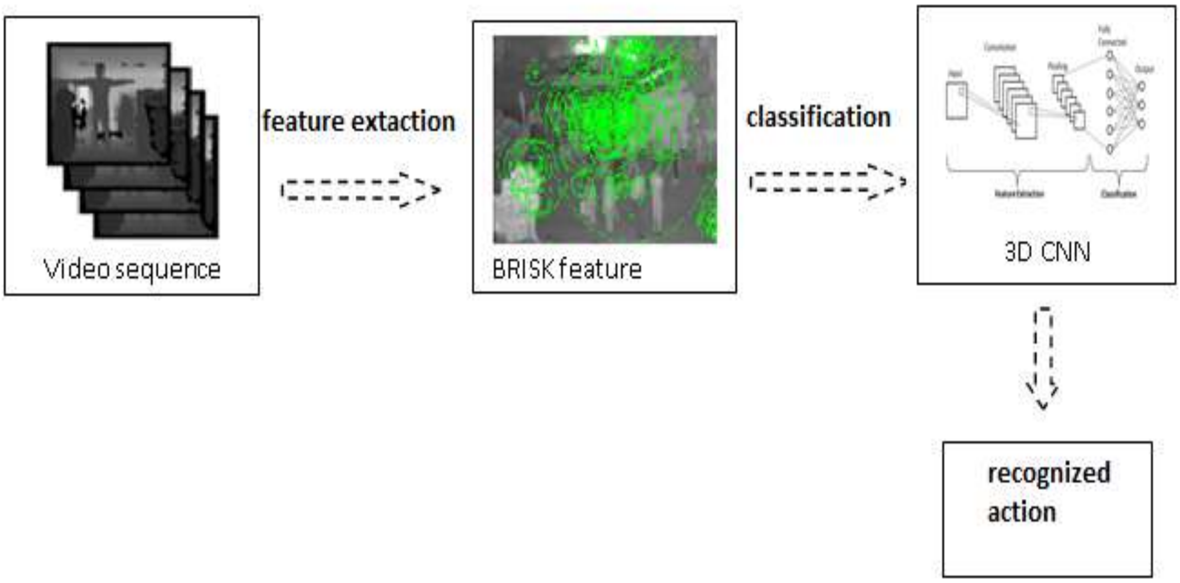


Figure 1: Proposed smart video surveillance system

3.1 Brisk Feature Extraction

BRISK algorithm is employed for feature extraction. This BRISK algorithm is more suitable than SIFT and SURF algorithm. It constructs the feature descriptor of the local image through the greyscale relationship of random point pairs within the neighborhood of the local image and obtains the binary feature descriptor. BRISK features are comparatively more accurate than other feature descriptors. The BRISK feature takes more time to detect but the advantage is that it is rotation and scale invariant. Other features like FAST takes less time but it's not scale invariant. The figure 2 shows the normal video frame with the extracted BRISK features.

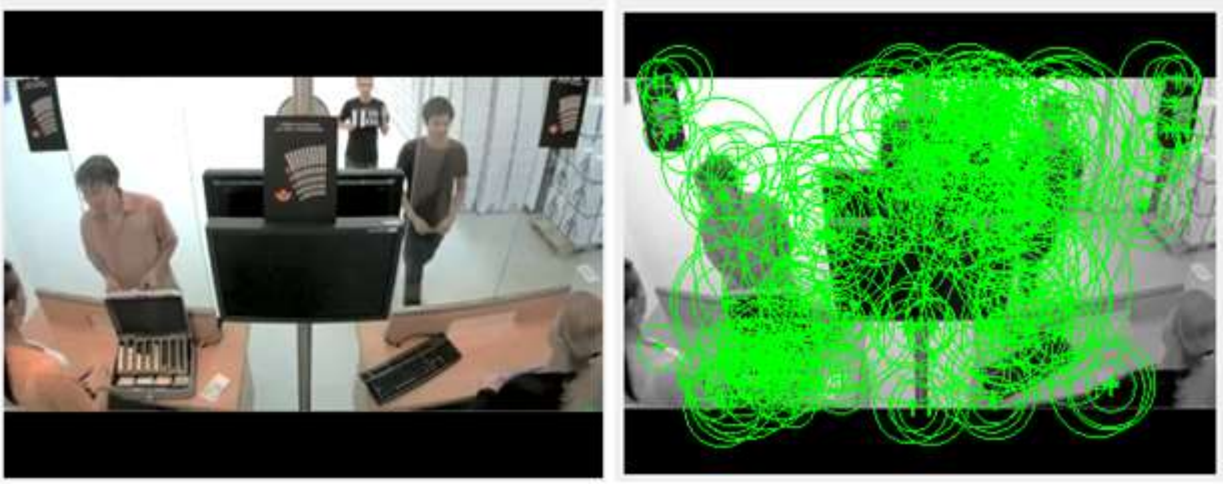


Figure 2 Frame in a normal video with its extracted BRISK features

3.2 Classification

Convolutional neural networks (CNNs) are a kind of deep model that will act directly on the raw inputs. CNN architecture is inspired by the organization and functionality of the cortical area and designed to mimic the connectivity pattern of neurons. The neurons within a CNN are split into a three-dimensional structure, with each set of neurons analyzing a little region or feature of the image. The 3D CNNs are being created to enhance the identification of moving and 3D images, like video from security cameras a time-consuming process that currently requires expert analysis. In the proposed work 3D-CNN models are used to recognize human actions abnormal and abnormal from the extracted 3D BRISK features. The 3DCNN consists of two convolutional layers interspersed with 2 max-pooling layers followed by 2 fully connected layers. A 2×2×2 max-pooling_3d is applied to the output of every convolutional layer. This model takes the 3D BRISK features extracted from the feature extraction step as input. Since it is a 3D CNN it is able to extract features from both the spatial and temporal dimensions by performing 3D convolutions.

3.3 Implementation

The features extracted videos are given as the input to the CNN. The proposed method is implemented using MATLAB and Python. To detect multiscale corner features detect BRISK() features function uses a binary robust invariant scalable keypoints (BRISK) algorithm. The MATLAB function returns extracted feature vectors, also referred to as descriptors and BRISK points. Then machine learning based CNN models are generated using keras package in Python. The 3D matrices of BRISK features for the 140 video data are classified. The actions recognized are categorised as 0's and 1's for normal and anomalous activities respectively. 70% of the data is used for training and 30% for testing. The trained Conv3D model has two input layers and one output layer and all parameters are trained in Conv3D model. The training accuracy is 100% and test accuracy is up to 87.5%. The model sequence with two input layers and other model with two input and max polling is obtained.

EXPERIMENTAL RESULTS

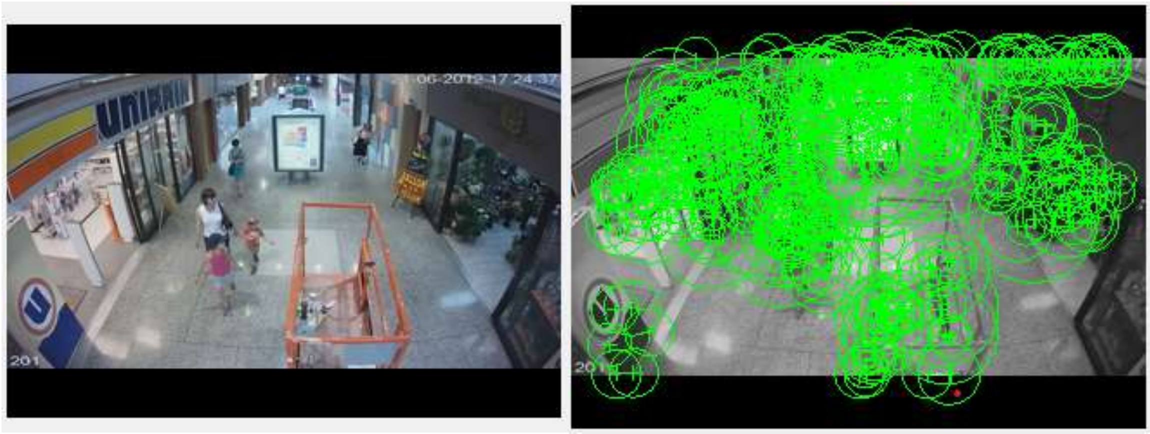
The proposed method uses BRISK features for action recognition. One challenge for the supervised machine learning method is that they require sufficient labeled data for the model. The proposed method is based on convolutional neural network which is a supervised learning method. The developed model is evaluated using UCF dataset.

4.1 Dataset

The most common dataset that is utilized in anomaly detection in surveillance videos is the University of Central Florida (UCF) Crime dataset [Chadha et al., 2017, Thakur et al., 2019, and Sultani et al., 2019]. In order to evaluate the proposed method, UCF dataset is used. It consists of long untrimmed surveillance videos that cover 13 planet anomalies, including Abuse, Arrest, Arson Assault, Road Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism. This dataset is employed because the data are mostly associated with public safety. There's a necessity for real-life datasets to determine the effectiveness of anomaly detection techniques.

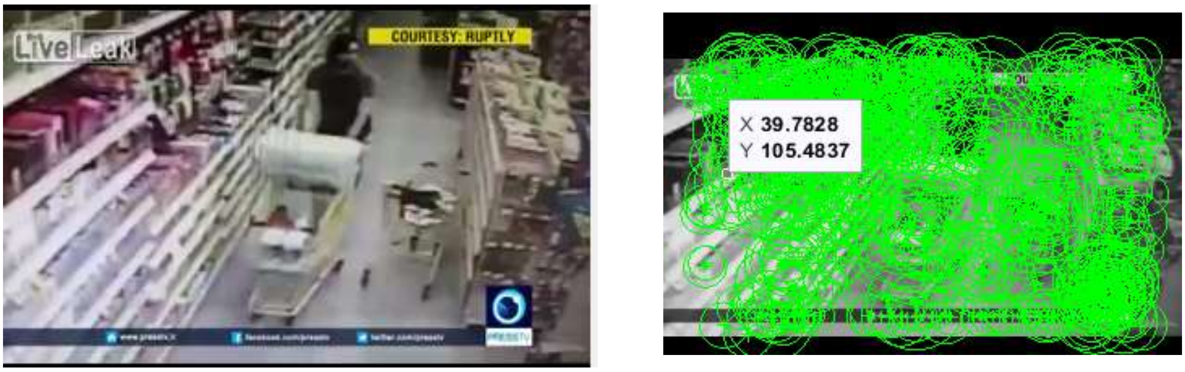
4.2 Feature Extraction

There are 1290 average frames in the given video data and for each frame the features are extracted. In the proposed method BRISK features are used for detecting the features and descriptors. A frame from the normal video with its BRISK features is shown in figure 3. A frame from abuse, arrest, arson, assault, road accident, burglary, explosion, fighting, robbery, shooting, stealing, shop lifting and vandalism and the extracted BRISK features are given in figures 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 and 16 respectively.



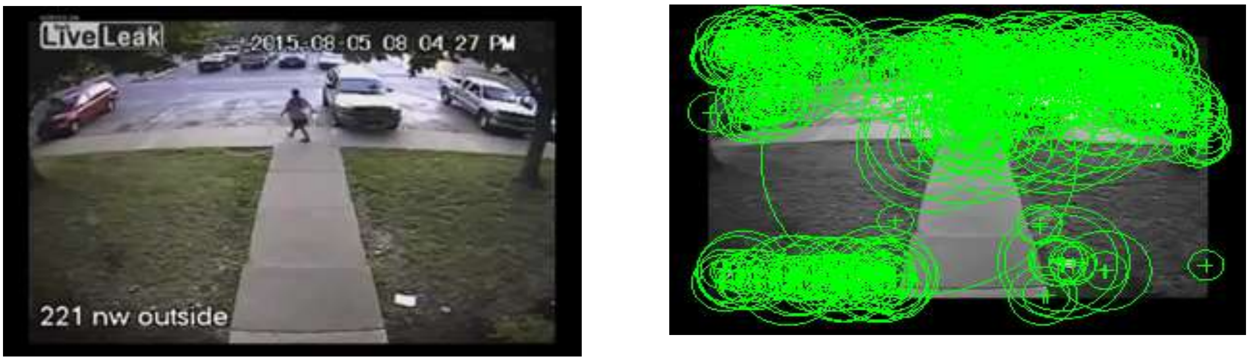
(a) (b)

Figure 3: normal video frame and features (a) original frame and (b) extracted BRISK features



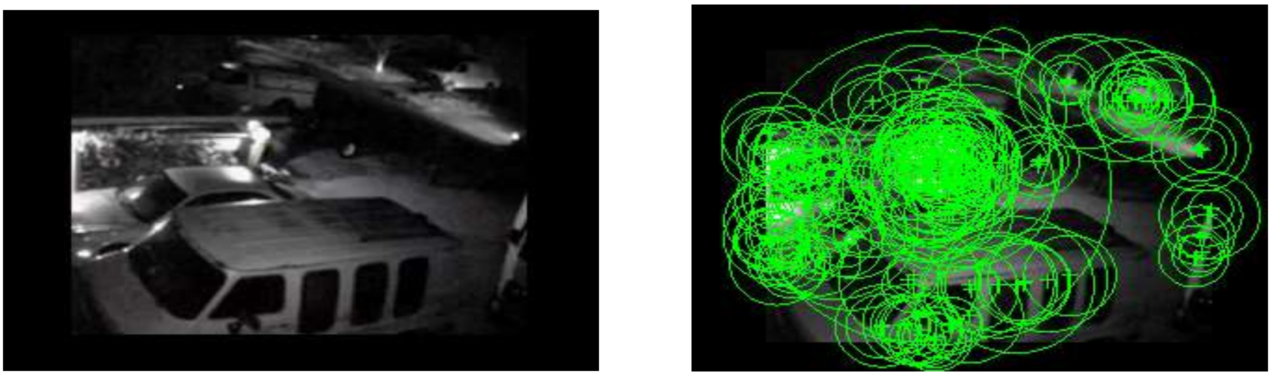
(a) (b)

Figure 4: abuse video frame and features (a) original frame and (b) extracted BRISK features



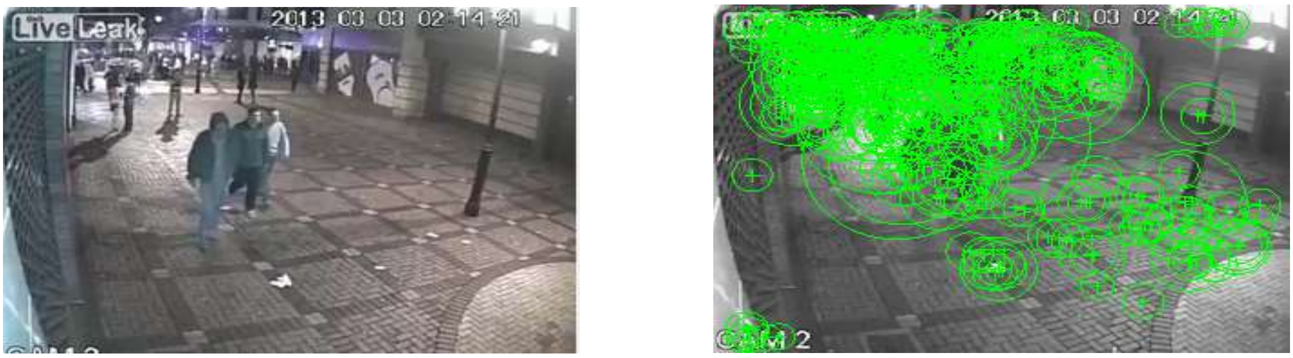
(a) (b)

Figure 5: arrest video frame and features (a) original frame and (b) extracted BRISK features



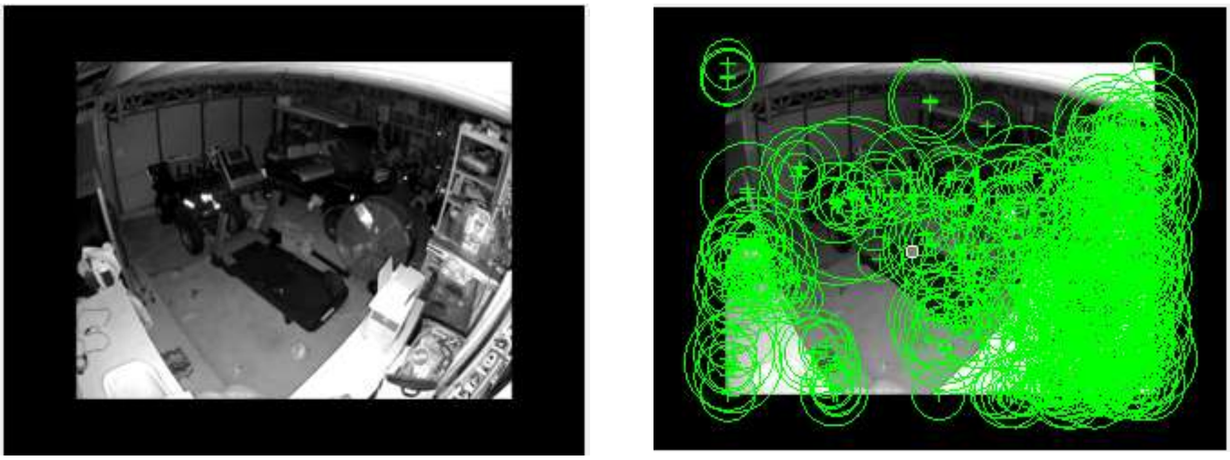
(a) (b)

Figure 6: arson video frame and features (a) original frame and (b) extracted BRISK features



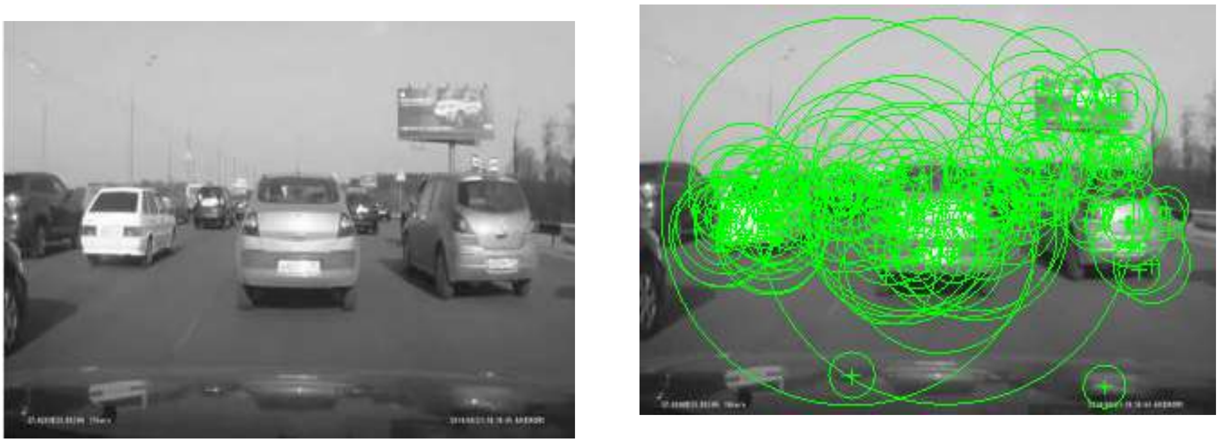
(a) (b)

Figure 7: assault video frame and features (a) original frame and (b) extracted BRISK features



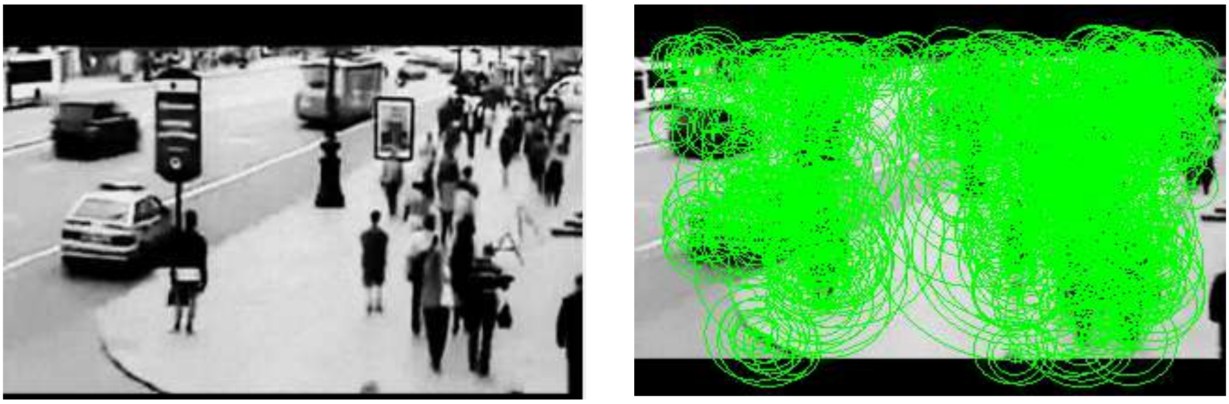
(a) (b)

Figure 8: burglary video frame and features (a) original frame and (b) extracted BRISK features



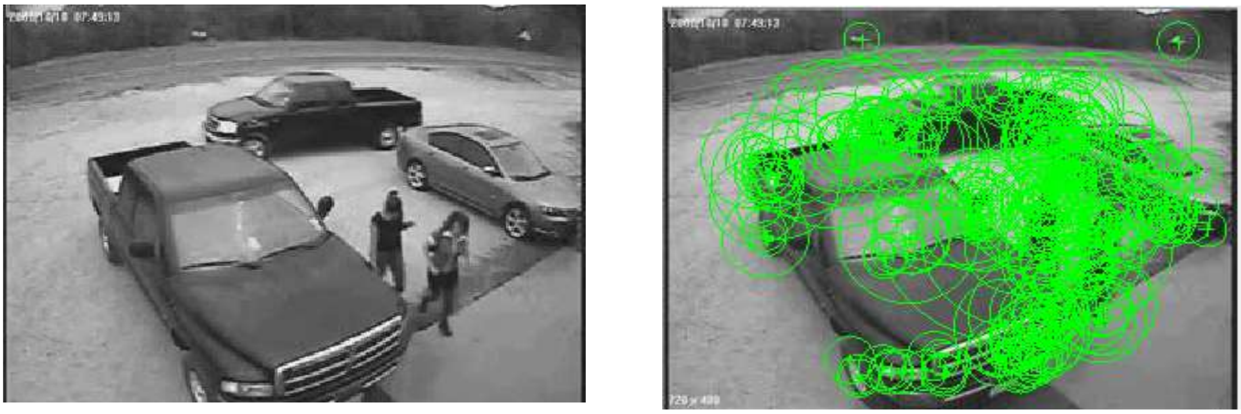
(a) (b)

Figure 9: explosion video frame and features (a) original frame and (b) extracted BRISK features



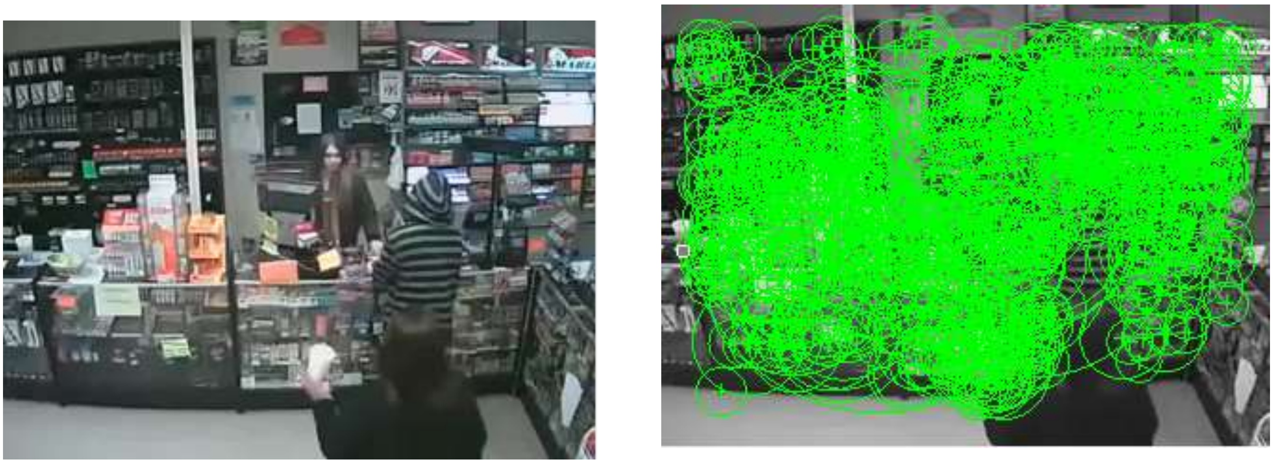
(a) (b)

Figure 10: fighting video frame and features (a) original frame and (b) extracted BRISK features



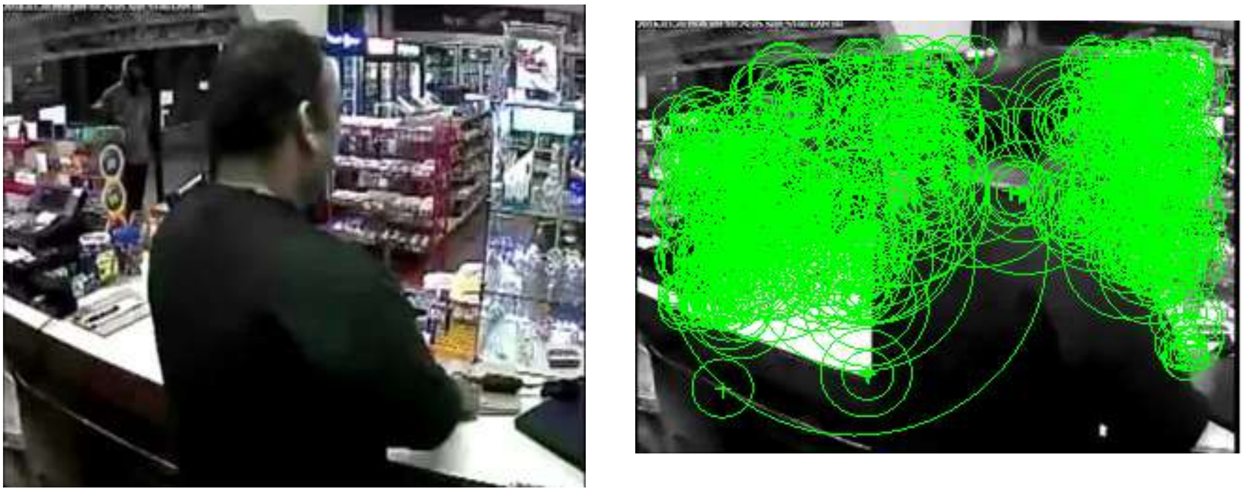
(a) (b)

Figure 11: accident video frame and features (a) original frame and (b) extracted BRISK features



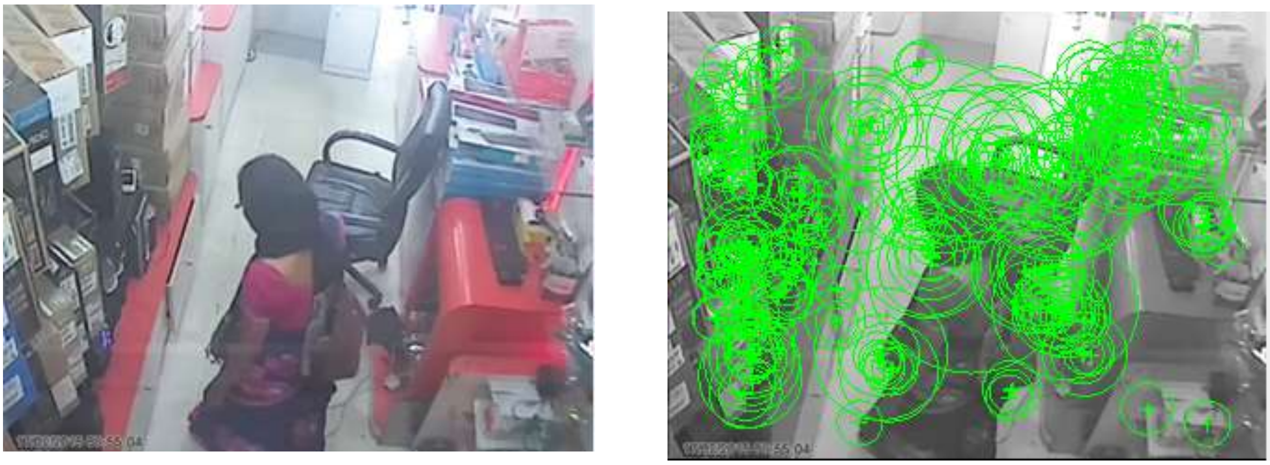
(a) (b)

Figure 12: robbery video frame and features (a) original frame and (b) extracted BRISK features



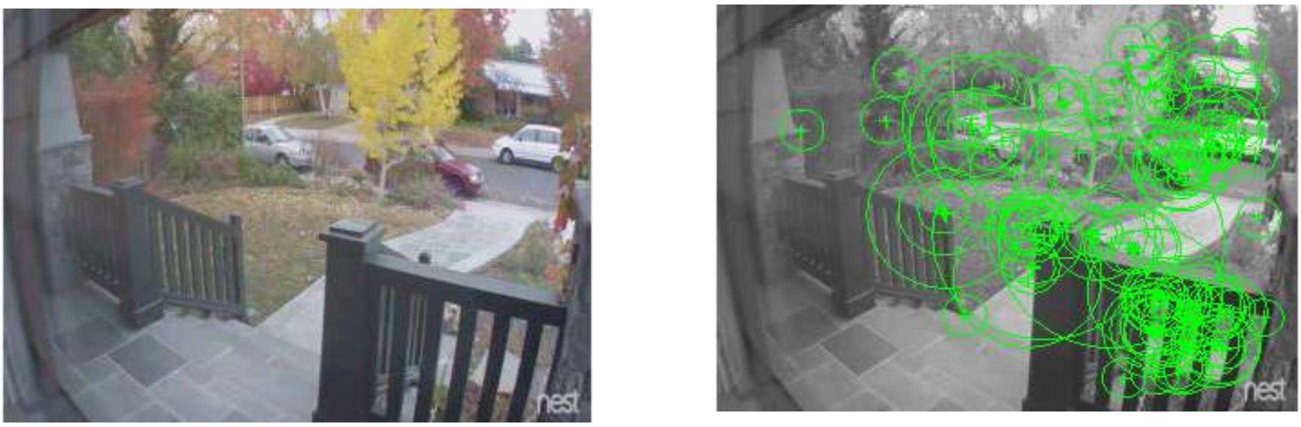
(a) (b)

Figure 13: shooting video frame and features (a) original frame and (b) extracted BRISK features



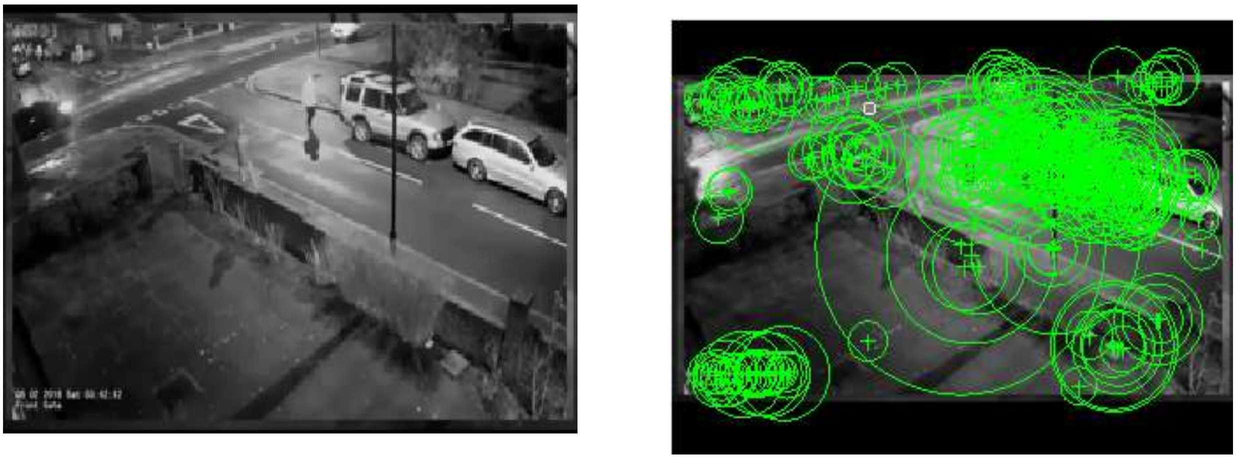
(a) (b)

Figure 14: shoplifting video frame and features (a) original frame and (b) extracted BRISK features



(a) (b)

Figure 15: stealing video frame and features (a) original frame and (b) extracted BRISK features

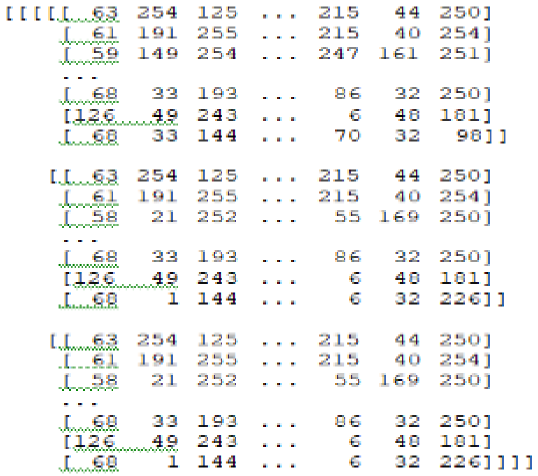


(a) (b)

Figure 16: vandalism video frame and features (a) original frame and (b) extracted BRISK features

4.3 Classification

In the proposed method 3D CNN is used for classifying the features as normal and abnormal. Open CV and Keras environments are used for video dataset classification. The extracted features are fetched by using python for classification by the model. At first the output of brisk feature is converted to the 3D arrays of BRISK feature as shown in figure 17. A sample of 140 videos including 25 normal and 115 abnormal video data are used to form a 3D array and the dimension of the resulting array is [140, 209, 11, 64]. The 0's and 1's represents the categories of action.



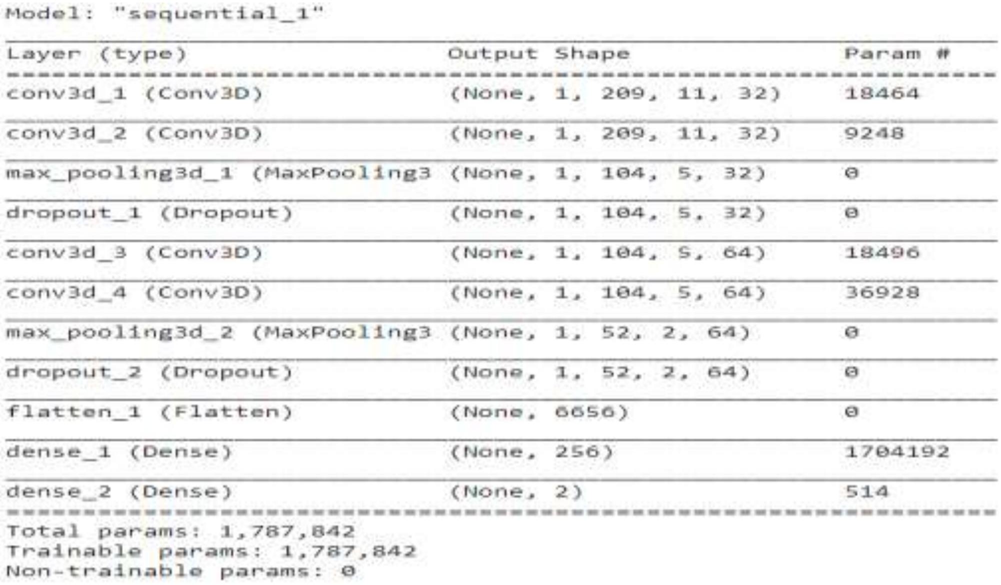
```
[[[[[ 63 254 125 ... 215  44 250]
    [ 61 191 255 ... 215  40 254]
    [ 59 149 254 ... 247 161 251]
    ...
    [ 68  33 193 ...  86  32 250]
    [126  49 243 ...   6  48 181]
    [ 68  33 144 ...  70  32  98]]

   [[ 63 254 125 ... 215  44 250]
    [ 61 191 255 ... 215  40 254]
    [ 58  21 252 ...  55 169 250]
    ...
    [ 68  33 193 ...  86  32 250]
    [126  49 243 ...   6  48 181]
    [ 68   1 144 ...   6  32 226]]

   [[ 63 254 125 ... 215  44 250]
    [ 61 191 255 ... 215  40 254]
    [ 58  21 252 ...  55 169 250]
    ...
    [ 68  33 193 ...  86  32 250]
    [126  49 243 ...   6  48 181]
    [ 68   1 144 ...   6  32 226]]]]
```

Figure 17: 3D array of BRISK features

The Conv3D model trained is summarized in figure 18. There are two hidden layers in the proposed methods with the 3d max pooling. The total output value of the defined model is (1, 209, 11, 32) for first hidden layer and (1, 104, 5, 64) for the second hidden layer. The total parameters used in this model is 1,787,842 and all the parameters are used for training in this work. 32 samples were used for train and 8 samples for testing. The test score for the created model is 87.5%. The performance measure values of the proposed 3D CNN model obtained during training is shown in table I.



```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv3d_1 (Conv3D)            (None, 1, 209, 11, 32)    18464
_________________________________________________________________
conv3d_2 (Conv3D)            (None, 1, 209, 11, 32)    9248
_________________________________________________________________
max_pooling3d_1 (MaxPooling3 (None, 1, 104, 5, 32)     0
_________________________________________________________________
dropout_1 (Dropout)          (None, 1, 104, 5, 32)     0
_________________________________________________________________
conv3d_3 (Conv3D)            (None, 1, 104, 5, 64)     18496
_________________________________________________________________
conv3d_4 (Conv3D)            (None, 1, 104, 5, 64)     36928
_________________________________________________________________
max_pooling3d_2 (MaxPooling3 (None, 1, 52, 2, 64)      0
_________________________________________________________________
dropout_2 (Dropout)          (None, 1, 52, 2, 64)      0
_________________________________________________________________
flatten_1 (Flatten)          (None, 6656)              0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               1704192
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 514
=================================================================
Total params: 1,787,842
Trainable params: 1,787,842
Non-trainable params: 0
```

Figure 18: trained Conv3D model

Table I: Measures obtained during training of proposed 3D CNN model

Measure	Value
Training loss	0.0
Training accuracy	100%
Validation loss	0.21
Validation accuracy	91.7%
Testing loss	0.15
Testing accuracy	87.5%

A primary method of 3D CNN is used in proposed method and the testing accuracy obtained for the proposed method is 87.5%. The accuracy rates of UCF crime dataset by using the proposed 3D CNN method is compared with those obtained in existing methods and are given in Table II.

Table II comparison with existing methods

Method	Test Accuracy
Sultani et al., 2019	68.5%
Landi et al., 2017	74.7%
Proposed method	87.5%

CONCLUSION

The video surveillance system involves observing a scene or scenes and searching for specific behaviors that are inappropriate or might indicate the emergence or presence of inappropriate behavior. Intelligent video surveillance system improves it by automatically detecting the anomalous behavior of humans from surveillance videos. In the proposed method binary feature extraction algorithm (BRISK) is used to extract features from videos. The 3D CNN is used in the proposed method as it is very effective method due to its feed-forward nature while compared to other classification techniques. 3D CNN has huge computation and memory efficiency. The University of California (UCF) dataset set is used to evaluate the proposed system. The proposed method produced testing accuracy of 87.5% which is high compared to the results generated by shallow neural networks.

ACKNOWLEDGEMENTS

The authors acknowledge the support of Rashtriya Uchchatar Shiksha Abhiyan (RUSA) 2.0 - Bharathiar University Entrepreneurship Innovation and Career Hub (BEICH), Department of Higher Education, Government of India. The authors would also like to thank Bharathiar University for providing the necessary support.

REFERENCES

1. Benezeth.Y, Jodoin.P.M, Emile, Laurent, Rosenberger, Comparative study of background subtraction algorithms, Journal of electronic imaging, Vol. 19(3), pp 1-31, 2012.
2. Chadha.A, Abbas.A, Andreopoulos.Y, Video classification with CNNs: using the codec as a spatio-temporal activity sensor, IEEE transactions on circuits and systems for video technology, Vol. 29, pp 475-485, 2017.
3. Hemangee.D, Bhattacharyya.A, Agarwal.R, Chowdhury.S.R, A review on human action recognition in surveillance videos, Reserved by international journal of advanced science and engineering, Vol. 6, pp 60-63, 2019.
4. Kalirajan.K, Sudha.M, Moving object detection for video surveillance, Scientific world journal, Vol. 2015, pp 1-10, 2015.
5. Landi.F, Snoek.C, Cucchiara.R, Anomaly locality in video surveillance, Cornell University Library, arXiv:1901.10364, 2019.
6. Li.Y, Xia.R, Huang.Q, Xie.W, Li.X, Survey of spatio-temporal interest point detection algorithms in video, IEEE translations and content mining are permitted for academic research, Vol. 5, pp 10323-10323, 2017.
7. Mahadevan.V, Bhalodia.V, Vasconcelos.N, Anomaly detection and localization in crowded scenes, IEEE transactions on pattern analysis and machine intelligence, Vol. 36, pp 18-32, 2017.
8. Miao.**Z,** Zou.**S,** Li.**Y,** Zhang.**X,** Wang.**J,** He.**M**, Intelligent video surveillance system based on moving object detection and tracking, in Proc. Int. Conf. Inf. Eng. Commun. Technol., pp. 1-4, 2016.
9. Muhammad.K, Ahmad.J, Lv.Z, Bellavista.P, Yang.P, Baik.S.W, Efficient deep CNN based fire detection and localization in video surveillance applications, IEEE transactions on system, man, and cybernetics: systems, Vol. 49, pp 1419-1434, 2019.
10. Ramachandra.B, Jones.M.J, Street scene: a new dataset and evaluation protocol for video anomaly detection, IEEE winter conference on application of computer vision, pp 2558-2567, 2020.
11. Shafie A.A, Ali M.H, Hafiz.F, Roslizar, Ali, Smart video surveillance system for vehicle detection and traffic flow control, Journal of engineering science and technology, Vol. 6(4), pp 469-480, 2011.
12. Sultani.W, Chen.C, Shah.M, Real-world anomaly detection in surveillance videos, Cornell University Library, arXiv: 1801.04264, 2019.
13. Tavagad.S, Bhosale.S, Prakash.A, Kumar.D, Survey paper on smart surveillance system, International research journal of engineering and technology, Vol. 3, pp 315-318, 2016.
14. Thakur.D, Kaur.R, An optimized CNN based real world8nomaly detection in surveillance videos, International journal of innovative technology and exploring engineering, Vol. 8, pp 465-473, 2019.
15. Zabłocki.M, Gosciewska.K, Frejlichowski.D, Hofman. R, Video surveillance systems for public spaces – a survey, Journal of theoretical and applied computer science, Vol. 8, pp. 13-27, 2014.