Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2023.027205
Article

Tech Science Press

Automatic Anomaly Monitoring in Public Surveillance Areas

Mohammed Alarfaj[1], Mahwish Pervaiz[2], Yazeed Yasin Ghadi[3], Tamara al Shloul[4], Suliman A. Alsuhibany[5], Ahmad Jalal[6] and Jeongmin Park[7,*]

[1]Department of Electrical Engineering, College of Engineering, King Faisal University, Al-Ahsa, 31982, Saudi Arabia
[2]Department of Computer Science, Bahria University, Islamabad, Pakistan
[3]Department of Computer Science and Software Engineering, Al Ain University, Al Ain, 15551, UAE
[4]Department of Humanities and Social Science, Al Ain University, Al Ain, 15551, UAE
[5]Department of Computer Science, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia
[6]Department of Computer Science, Air University, Islamabad, Pakistan
[7]Department of Computer Engineering, Korea Polytechnic University, 237 Sangidaehak-ro Siheung-si, Gyeonggi-do, 15073, Korea
*Corresponding Author: Jeongmin Park. Email: jmpark@kpu.ac.kr
Received: 13 January 2022; Accepted: 08 June 2022

Abstract: With the dramatic increase in video surveillance applications and public safety measures, the need for an accurate and effective system for abnormal/suspicious activity classification also increases. Although it has multiple applications, the problem is very challenging. In this paper, a novel approach for detecting normal/abnormal activity has been proposed. We used the Gaussian Mixture Model (GMM) and Kalman filter to detect and track the objects, respectively. After that, we performed shadow removal to segment an object and its shadow. After object segmentation we performed occlusion detection method to detect occlusion between multiple human silhouettes and we implemented a novel method for region shrinking to isolate occluded humans. Fuzzy c-mean is utilized to verify human silhouettes and motion based features including velocity and optical flow are extracted for each identified silhouettes. Gray Wolf Optimizer (GWO) is used to optimize feature set followed by abnormal event classification that is performed using the XG-Boost classifier. This system is applicable in any surveillance application used for event detection or anomaly detection. Performance of proposed system is evaluated using University of Minnesota (UMN) dataset and UBI (University of Beira Interior)-Fight dataset, each having different type of anomaly. The mean accuracy for the UMN and UBI-Fight datasets is 90.14% and 76.9% respectively. These results are more accurate as compared to other existing methods.

Keywords: Abnormal event classification; gray wolf optimizer; region shrinking; xg-boost classifier

1 Introduction

Due to increased security concerns and need for safety applications [1–4], development, in the video. The major use of such applications is to identify the abnormal events/threats occurred in public areas [5–8]. Abnormal event detection is considered a major research topic in computer vision. Video surveillance and public safety awareness are considered to be the main reasons to invest time in the research area of abnormal event detection. Multiple applications have been built that reduce the cost and increase the significance of public safety. Due to the huge increase in population and to avoid any unwanted condition, the need for accurate and significant applications has been increased. Surveillance cameras are used in public places such as markets, stadiums, airports, museums, train stations to identify any abnormal event that can cause inconvenience for people [9–11]. There exist many challenges that make the detection of abnormal events difficult.

The main challenges we found in the literature are the nature of ground truth available publically, size of datasets, type of dataset either it is real-time or acted, background variance, and, the absence of a universal definition of anomalies existing datasets. Each dataset contains a different type of anomaly and has multiple videos in different environments. We considered the datasets with different environments and anomalies to evaluate this system. The main target of this research is to build an effective, robust, and adaptive system that can perform well in different environments and detect multiple types of anomalies using a single application [12].

The proposed system can be applicable in multiple potential settings, such as security applications, crowd estimation, application used to take care of vulnerable and elder people, inspection and analysis, military applications, traffic monitoring of pedestrian, robotic and video surveillance. Evaluation has been performed using two public benchmark and results presents supremacy of proposed system over existing applications. Following are the main contribution of this research works.

a) We introduce an accurate method for the detection of moving objects using Gaussian Mixture Model (GMM) and Kalman filter. The modified system detectsed occluded people and objects accurately.
b) We designed a region shrinking model that is used to isolate the person who is occluded for a long period. This helps us to enhance the accuracy of human silhouette detection in complex and occluded environments.
c) To enhance the accuracy of the proposed model, we extracted features that include velocity, speed, and optical flow in the temporal domain. Fuzzy c-means with Hough transform has been used to verify human silhouettes.
d) For feature optimization, we applied gray wolf optimizer to optimize data and reduce computation cost. An XG-Boost classifier with a local descriptor is applied to classify normal and abnormal events.

The rest of this paper is organized as follows: Section 2 contains a detailed overview of the related works. In Section 3, the methodology of abnormal event detection is discussed. Section 4 describes the complete description of the experimental setup and a comprehensive comparison of the proposed system with existing state-of-the-art systems. In Section 5, future directions and conclusions are defined.

2 Literature Review

With advances in crowd ana and surveillance applications and technologies, more effective and adaptive systems have also been developed. The research community has developed many robust, novel, and advanced methods for video and crowd analysis [13]. In the past decade, many researchers have proposed versatile approaches for human detection, event detection, and anomaly/abnormal event detection in surveillance videos. Summary of research work performed by different researchers have been displayed in Tab. 1.

In this paper we considered the medium density crowd to detect the abnormal events with different environments, we considered the data set with in-door and out-door scenarios and consider the problems of occlusion, arbitrary movement, and anomaly detection.

Table 1: Summary of existing work done in field of activity recognition and event recognition

Authors	Approach	Application	Reference
Franco et al.	Approach for abnormal event detection using dynamic sparse coding based on the reconstruction of query signal sparsely using special-temporal information	Abnormal event detection	[14]
Jalal et al.	A learning framework of sparse combinations is implemented that reduces major processing cost and improves the effectiveness	Action recognition	[15]
Ullah et al.	Proposed a system using sparse reconstruction cost (SRC) and reduced the computational complexity using sparsely formulation	Action and activity recognition	[16]
Amft et al.	Proposed an approach of a learning framework and builds an effective method to develop an adaptive cascade dictionary to build a learning framework to detect an anomaly	Anomaly detection	[17]
Iang et al. Chen et al.	Proposed a model for action recognition using body parts estimation and their movement estimation	Action recognition	[18] [19]
Li et al.	Merged 2-D and 3-D approach using explicit shape model to track people behavior	Behavior estimation of pedestrian	[20]
Einfalt et al, Yu et al, Franklin et al.	Proposed system with new combination of features to predict mishap using neural network	Pedestrian protection system and anomaly detection	[21] [22] [23]
Lohithashva et al.	Proposed system using gray level co-occurrence metrics and local binary pattern texture for violent videos	Event detection in violent videos	[24]
Feng et al.	Utilized attention guided LSTM and detected spatio temporal events like fall detection	Abnormal event detection	[25]
Khan et al.	Proposed a system to estimate body pats movement using maker based approach	Body parts analysis	[26]
Golestani et al.	Applied neural network on induction based signal and their motion analysis, reduce computational time.	Activity recognition	[27]
Dalal et al.	Used histogram of gradient features and build system to detect humans. Performance of object detection improved	Pedestrian detection and tracking	[28]
Zhu et al., Lowe et al.	Proposed system using local scale invariant features to improve object detection to add on the performance of activities performed by the humans.	Human detection and human activities detection	[29] [30]
Wang et al., Niebles et al.	Proposed system for activity detection using segmentation and features extracted for motion segments	Activity detection	[31] [32]

3 Material and Methods

Proposed system comprised of six main modules preprocessing, human silhouette extraction, human verification, feature extraction and optimization, and event classification. The complete process is given in Algorithm 1.

Algorithm 1: Abnormal Event Detection

Input: F: Extracted set of video frames

Output:E: Event class

Variables: P = processed image, s = speed, o = optical flow, T = Time, N = number of frames in set F, Ob = Number of objects detected after preprocessing, T_i = size of average human silhouette and C = position of semi-circle in object

```
Repeat
    For i = 1 to N do
        img = GaussianMixtureMode(i)
        detectedObject = KalmanFilter(img)
        K = KalmanFilter(D)
        S ← ShadowRemoval(K)
    Repeat
     For j = 1 to Ob do
        If(size(j) > Ti)
                C = FindCircl(i)
                if(true)
                        Sh = ShrinkRegion(Si)
                Else
                        Obj = i
        End If
    End For
   f← ExtractFeature(S, O)
   Opt←GrayWolfOptimizer(f)
   Behavior←XGBoost(Opt)
until complete set of video frames traversed.
E ← Behavior
Return E
End
```

Preprocessing is the first step, we performed on extracted video frames from 2 different datasets. GMM is applied for background subtraction and identification of foreground objects. Then we used the Kalman filter to detect and track moving objects across the video. There were some objects having shadows that can reduce the efficiency of object detection. To deal with them and to enhance the appearance of the objects, we converted the images into Hue, Saturation, Value(HSV) domain to find the clear contrast between an object and its shadow and removed the shadow using a threshold value. The next step we followed is human silhouette extraction that is performed using region shrinking or object separation that helps to isolate multiple occluded people. Human/non-human silhouette verification is done using fuzzy c-means and local descriptor. We applied a gray-wolf optimizer to optimize the features and an XG boost classifier to classify the behavior of the crowd. Fig. 1 presents the complete architecture of the proposed system. We evaluated the performance of the proposed approach on publicly available benchmarks: University of Minnesota (UMN) [33] and UBI (University of Beira Interior)-Fight [34] datasets, and the proposed method was fully validated for efficacy and proved superior to the other state-of-the-art methods.

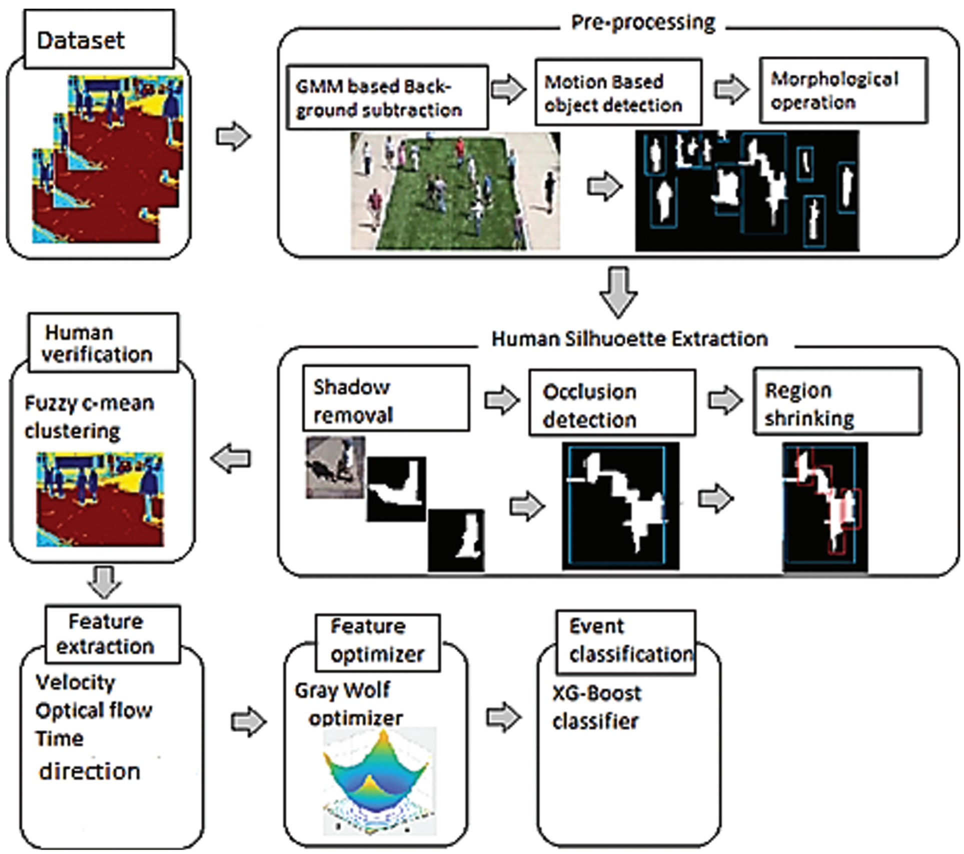


Figure 1: The architecture flow diagram of the proposed anomaly detection system

3.1 Pre-processing

The proposed system consists of sub phases to pre-process the data. GMM is used to remove the background and moving objects are detected using kalman filter in sequences of videos. Further, multiple morphological operations [35] are performed to filter the objects.

The most important part of preprocessing is the declaration of foreground and background objects in an image. Correct detection of an object in any image analysis system leads to an effective and accurate performance of the system. In the proposed system we used GMM to differentiate foreground and background objects and remove the background. GMM picks every pixel in an image frame and models it into Gaussian distribution. Each pixel is divided into the foreground (FG)/ background (BG) pixel using the intensity of pixel [36]. To declare pixels as foreground or background pixels, probability of each pixel is calculated using Eq. (1) and decides whether to include it in FG or BG.

$$P(X_t) = \sum\nolimits_{i=1}^{K} (w.n\left(X_t,\ U_t,\ \sum i,\ t\right) \tag{1}$$

where X_t is the value of a current pixel in frame t, K is the number of distributions in the mixture, $\mu_{i,t}$ is the value of the mean of the k^{th} distribution in frame t, $\Sigma_{i,t}$ is the standard deviation of the k^{th} distribution in frame t and $\eta(X_t, \mu_t, \Sigma\ i, t)$ is the probability density function (pdf) [37] given in Eq. (2).

$$\eta(X_t, \mu, \sum) = \frac{1}{(2\pi)^{\frac{n}{2}}|\sum|^{\frac{1}{2}}} exp^{\frac{-1}{2}(X_t-\mu)\sum^{-1}(X_t-\mu)} \tag{2}$$

According to Stauffer and Grimson [38], every color space in an image is uncorrelated with others. So, a difference in intensity can result in a uniform standard deviation [39]. The covariance matrix can be computed using Eq. (3).

$$\sum_{i,t} = \sigma_{i,t}^2 I \tag{3}$$

Every Gaussian that have a bigger value than the chosen threshold, has been classified as background as given in Eq. (4) and Gaussian with lesser values has been included in foreground part.

$$B = argmin_b\left(\sum\nolimits_{i=1}^{b} \omega_{i,t} > T\right) \tag{4}$$

Value of ω, μ, and σ have been updated using Eqs. (5)–(7) respectively if pixel matches any of the K Gaussian [40].

$$\omega_{i,t+1} = (1-\alpha)\omega_{i,t} + \alpha \tag{5}$$

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho.X_{t+1} \tag{6}$$

$$\sigma_{i,t+1}^2 = (1-\rho)\sigma_{i,t}^2 + \quad \rho(X_{t+1} - \mu_{i,t+1}).(X_{t+1} - \mu_{i,t+1})^T \tag{7}$$

where

$$\rho = \alpha \times \eta\left(X_{t+1},\ \mu_{i,} \sum\nolimits_i \quad\right) \tag{8}$$

Meanwhile, if in case, no Gaussian value [41] matches, then only the value of ω is updated using Eq. (9).

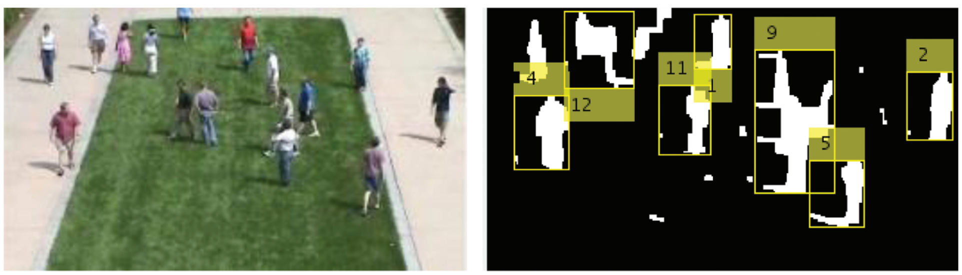
$$\omega_{j,t+1} = (1-\alpha)\omega_{j,t} \tag{9}$$

If values of all parameters are found, foreground pixels can be estimated easily. GMM results on UMN datasets are given in Fig. 2.

After background removal Kalman filter [42] has been utilized to detect moving objects across the videos. Processed images of Gaussian filter [43] is used as input and moving objects are extracted. Results of Kalman filters are displayed in Fig. 3.

(a) (b)

Figure 2: Object detection using GMM. (a) original image and (b) objects detected using GMM



(a) (b)

Figure 3: Object tracking using Kalman Filter. (a) original image and (b) object tracked using Kalman filter where yellow boxes with labels show the number of objects being tracked

3.2 Human Silhouette Extraction

We observed that some human silhouettes in the datasets have shadows that can mislead the extraction of human silhouettes and mix them with non–human objects. To cope with this, we performed shadow removal before human silhouette extraction. Then we applied region shrinking to isolate multiple occluded human silhouettes.

We used objects after shadow removal as input and we detected that there are still some objects that are occluded and have a larger size than the average size of a human. We used a semi-circle as an input pattern and found that the regions that have a larger size than average human silhouette size resembled semicircles. Coordination of detected points have been extracted and the top point is considered as head. After head point has been expected, left most, right most, and bottom points are calculated using a heuristic based average size and height of human silhouette. Detected parts are isolated as independent human silhouettes and the region is subtracted from the originally detected object (see Algorithm 2). The process performed recurrent steps until detected object size lies within the limit of human size [44]. Fig. 4 presents the results of region shrinking phases.

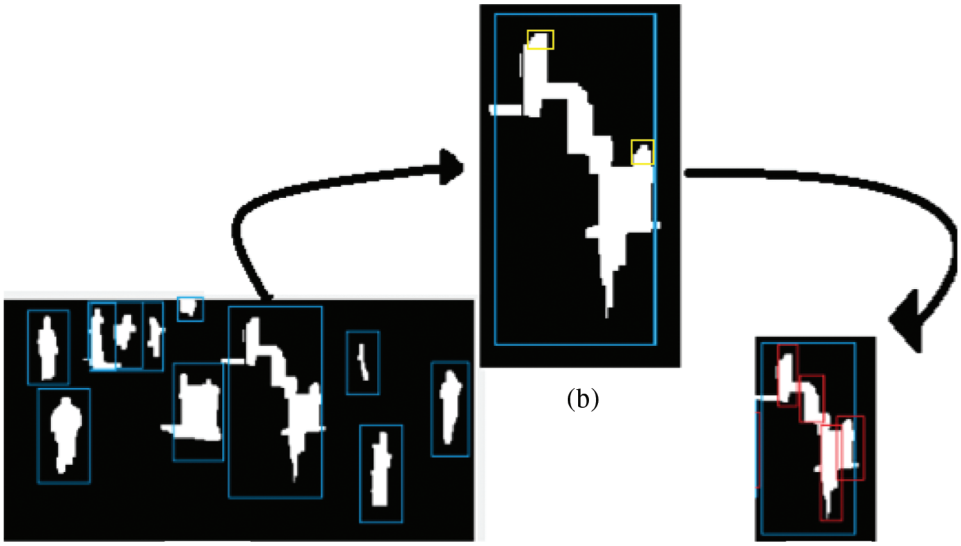


Figure 4: Region shrinking process. (a) objects filtered after pre-processing, (2) region selected and head points detected at multiple places, and (c) region shrinked and multiple persons isolated

---

Algorithm 2: Region Shrinking Algorithm

Input: O: Object having size larger than human silhouette, t: target image (semi circle), H_s : avergae size

Output: H: set of huma silhoette extracted from object O

Variable: (x, y) = x and y coodinates of head/circle found, tl = lower value of human silhouette, th upper value of human silhouette, tle = Estimated left point , tre = Estimated right point, (N, v) = y coodinates of estimated end points

Begin

S = Size(O)

If(O >AverageSize(H_s))

(x, y) = MatchCircle(t)

tl = EstimateLowerPoint

th = EstimateUpperPoint

tle = EstimateleftPoint

tre = EstimaterightPoint

(x, y) = (x + tle, x + tre)

(N, v) = (y + tl, y + th)

H = calculateBox((x, y), (N,V))

S = subtract(O-H)

Repeat untill S < = H_s

End If

End

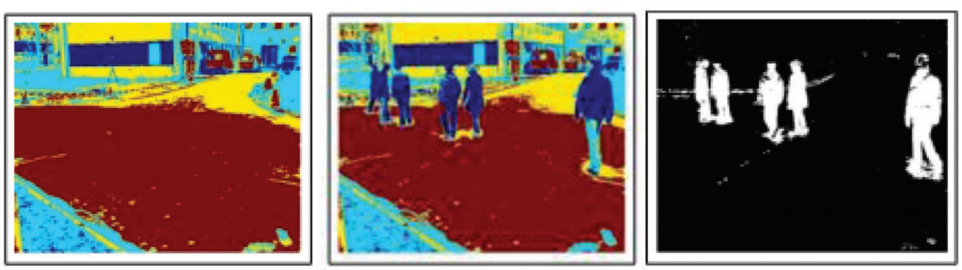
Performance of region shrinking algorithms has been evaluated using the UMN dataset and the UBI-Fight dataset. Results are displayed in Tab. 2. It is clearly shown that region shrinking algorithm boost the performance of object detection module. There were multiple occluded objects that were counted as a single object are separated and divided into multiple objects after applying region shrinking.

Table 2: Results of region shrinking algorithm with occluded objects

Datasets	Ground truth	Number of objects using object detection	Number of objects after region shrinking
UMN	6	3	5
UMN	7	4	7
UBI-Fight	4	2	4
UBI-Fight	3	1	2

3.3 Human Silhouette Verification

We used fuzzy c-means to verify the human silhouettes. Fuzzy c-means is a clustering technique used to cluster the human silhouettes. Fuzzy c-means is based on an un-supervised clustering technique that connects the similar data points together and isolates data points that are dissimilar and far apart from each other. As compared to other methods of clustering, fuzzy c-means depicts higher accuracy. We used fuzzy c-means with soft clustering approach to cluster human and non-human objects (see Fig. 5) for human silhouettes verification [45].



(a) (b) (c)

Figure 5: Human silhouette verification results of UMN dataset. (a) background image, (b) clustered objects, and (c) verified humans

3.4 Feature Extraction

In this section, we discuss the features we used to classify the events. This list includes velocity of a human silhouette, optical flow of different points of human silhouettes and direction of movement. Features of each human silhouette are extracted independently and discussed below in detail.

3.4.1 Velocity

Distance of objects between two points has been computed using value of their centroid [46]. Euclidean distance has been used to calculate distance. It is calculated by using the Euclidean distance formula given by Eq. (10). Position of pixels are used as initial and final stage of objects.

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{10}$$

where x_1 is position of previous pixel, x_2 is position of present pixel in terms of width, y_1 is the position of previous pixel and y_2 is position present pixel in terms of height. Considering the distance velocity of all moving objects have been calculated using distance travelled per unit time with respect to frame rate as Eq. (11). The velocity of the object is presented in Fig. 6.

$$Velocity = DistanceTravelled/FrameRate \tag{11}$$

(a) (b)

Figure 6: The velocity of moving objects. (a) original image and (b) moving objects velocity

3.4.2 Optical Flow

To estimate the motion pattern optical flow has been extracted for the selected pixels belonging to moving objects using Horn-Schunck [47] optical flow algorithm. Speed of a pixels relates with its neighboring pixels. For every point in optical flow change of speed is smooth with no sudden changes. Smoothing constraint has been described using Eqs. (12) and (13). Fig. 7 presents the optical flow of multiple points selected from the human silhouette.

$$\nabla^2 u + \nabla^2 v \tag{12}$$

where

$$\nabla^2 u = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2, \ \nabla^2 v = \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2, \tag{13}$$

(a) (b)

Figure 7: Optical flow of moving objects (a) original image and (b) yellow arrow presents optical flow for moving objects with head pointing their direction

3.5 Abnormal Classification Using XG-Boost Classifier

In this section, we describe the machine learning algorithm XG-boost classifier [48] that is used for abnormal event classification. XG Boost classifier supervised learning model that is highly efficient, portable and flexible gradient boosting library that can be used as classifier and for regression also. We used XG-Boost classifier over three datasets: UMN and UBI-Fight. XG-Boost is one of the most popular machine learning algorithm these days. As compared to other machine learning algorithms, XG-Boost classifier gives better performance regardless of data types [49]. It provides parallel computation using extreme gradient boosting Eqs. (14) and (15) computation on a single machine.

$$F_o(x) = \operatorname{argmin} \sum\nolimits_{i-1}^{n} L(y_i,\ \gamma) \tag{14}$$

$$F_m(x) = F_{m-1}(x) + \operatorname{argmin}\left[\ \sum\nolimits_{i=1}^{n} L(y_i,\ F_{m-1}(x_i) + h_m(x_i))\right] \tag{15}$$

where $\{x_{1,}\ x_{2,}\ x_{3,\ldots}\ x_n\}$ is training set, F_x is approximation function with low cost and h_m is learner function. We performed a comparative analysis of Decision Tree [50] and XG-Boost classifier to ensure the effectiveness of the classifier used in this study. Tab. 3 shows the results of two classifiers used in the comparative analysis and it is clearly shown that XG-Boost performs very well as compared to the other classifiers.

Table 3: Comparison table for well-known classifiers over three datasets, considering precision, recall and f-score

Classifier	Decision tree			XG-boost		
Datasets	Precision	Recall	F-score	Precision	Recall	F-score
UMN dataset	0.84	0.87	0.84	0.90	0.97	1.00
UBI-fight	0.71	0.78	0.77	0.76	0.86	0.89

4 Experiments and Results

Experiments were performed using Intel Core i5 2.80 GHz CPU having 16 GB RAM and 64-bit operating system. System has been implemented in Matlab 2013 b and image processing toolbox. For each video, frames are extracted and used to process the objects and classify the events.

4.1 Datasets Description

This section provides details of the different datasets used in the evaluation of the proposed system. We used two datasets having different type of anomalies in different settings.

4.1.1 UBI Fight Dataset

UBI-Fight dataset is publically available dataset specifically designed for anomaly detection provides multiple fighting scenarios (see Fig. 8), the UBI-Fights dataset contains dataset 1000 videos comprising 80 h of different scenarios. 216 videos belongs to fight events and remaining, 784 are normal daily-life activities. All videos are annotated at frame level.

4.1.2 UMN Dataset

This dataset was prepared by the University of Minnesota and publically available. It containing 11 scenarios from 3 different places includes both indoors and outdoors environment. The events are belongs to local, global, temporal and spatial categories. Fig. 9 presents different normal and abnormal views of the UMN dataset.

Figure 8: Some examples of the UBI fight dataset

Figure 9: Normal and abnormal views of the UMN dataset

4.2 Performance Measurement and Result Analysis

For system evaluation, precision has been taken as the performance measure to evaluate the performance of our system. Precision [51] has been calculated using Eq. (16).

$$\text{Precision} = \frac{t_c}{(t_c + f_c)} \tag{16}$$

where t_c is the number of true anomalies detected and fc is the number of false anomalies detected. Tab. 4 displays the results over the UMN dataset, and Tab. 5 presents the results of the UBI-Fights dataset.

Table 4: Results of anomaly detection with UMN dataset

Sequence	Detected anomalies count	Ground truth	Truly detected	False detected	Precision
100	2	2	2	0	1.00
100	3	3	3	0	1.00
150	3	3	3	0	1.00
100	2	2	2	0	1.00
180	4	3	2	1	0.66
120	2	3	2	1	0.66
120	9	9	9	0	1.00
Average anomaly detection accuracy rate = **90.4%**					

Table 5: Results of anomaly detection with UBI fight datasets

Sequence	Detected anomalies count	Ground truth	Truly detected	False detected	Precision
100	2	2	1	1	0.50
100	2	3	2	1	0.66
150	4	3	3	1	0.75
100	3	3	2	1	0.66
180	4	3	3	1	0.75
120	2	2	2	0	1.00
120	1	1	1	0	1.00
Average anomaly detection accuracy rate = **76.19%**					

Table 6: Performance measures of all datasets used to evaluate the system

Datasets	AUC	Decidability	EER
UMN Dataset	0.906	1.386	0.16
UBI-Fight	0.769	0.323	0.427

We used the area under the ROC Curve (AUC) [52], decidability [53], and Error Equivalence Rate (EER) [54] to further evaluate the performance of our system over the three datasets as shown in Tab. 6. It consider the level of uncertainty. The cumulative performance of both datasets of the proposed model are depicted in Tab. 7.

Table 7: Comparison table of proposed system with the state of the art methods

Dataset	Methods	Accuracy	Methods	Accuracy	Methods	Accuracy
UMN	Mehran et al. [55]	87	Cui et al. [56]	85	Proposed method	90.6
UBI-fight	Degardin et al. [57]	75	Zhu et al. [58]	72	Proposed method	76.9

To evaluate the effectiveness of proposed system comparison was performed with state of the art methods with two datasets used in system (See Tab. 7). It is clearly visible that the proposed system performed very well as compared to other methods.

Results show that the proposed system performs well when evaluated with both dataset as compare to other methods used in literature.

5 Discussion and Analysis

With dramatic increase in world population, the number of surveillance application has also increased. Abnormal event detection is one of the widely research topics for video surveillance applications. The main goal is to detect any anomalous activity at public places like shopping malls, public rallies, outdoor events, or in educational institutes. A number of applications have been made, still abnormal event detection is a challenging task due to variance of background or environment. Also, there is no standard definition of anomaly different events are considered anomalous in different environments.

Proposed system has been tested using two different datasets having different set of anomaly and perform well as compare to other existing approach for abnormal event detection. There are some limitation of our system that can effect upon the performance of this system, theses includes

• **Nature of the crowd:** In heavy density crowd where person are highly occluded and occlusion continues for long time, in such environment our object detection module may not detect the human silhouettes as accurate as it can detect in low to medium density crowd environment.

• **Type of objects:** This system identifies the anomaly created by human silhouettes, For example falling of a human, Running in panic, Fighting between peoples, but did not cover anomalies caused by objects like entrance of motor car in pedestrian area, car accident or any other that does not involve human. We have considered only human silhouettes as of object of interest.

We will work on these limitation in future and will try to enhance performance of system in variance environment and all type of anomalies specially we will work on the high-density crowds.

6 Conclusion

This paper proposes an effective method for the detection of abnormal events in public places for remote sensing data. To enhance the performance of our system and to handle the occluded situation we designed a new region shrinking algorithm to extract human silhouettes in the occluded situations. Fuzzy c-means clustering is used to verify objects as human or non-human silhouettes. We extracted motion-related features which include velocity and optical flow to classify the abnormal events. To reduce computational cost and optimize data, the Gray Wolf optimization model is implemented. The proposed abnormal event detection system works in surveillance applications with different environments for remote sensing data. The system also has theoretical implications in identifying the suspected or abnormal activities in shopping places, railway stations, airports, shopping malls.

In the future, we will work on the high-density crowds with moving background objects in a complex environment. We will work for places with high levels of an occluded occlusion, like at holy events, rallies, shopping festivals.

Funding Statement: The authors acknowledge the Deanship of Scientific Research at King Faisal University for the financial support under Nasher Track (Grant No. NA000239). Also, this research was supported by a Grant (2021R1F1A1063634) of the Basic Science Research Program through the National Research Foundation (NRF) funded by the Ministry of Education, Republic of Korea.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. S. Khanal, P. Prasad, A. Alsadoon and A. Maag, “A systematic review: Machine learning based recommendation systems for e-learning,” *Education and Infomation Technolology*, vol. 25, pp. 2635–2664, 2020.

[2] M. Gochoo, T. H. Tan, S. H. Liu, F. R. Jean, F. S. Alnajjar *et al.,* “Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and DCNN,” *IEEE Jounal of Biomedical and Health. Informatics*, vol. 23, pp. 693–702, 2018.

[3] J. K. Aggarwal and Q. Cai, “Human motion analysis: A review,” *Computer Vision and Image Understanding*, vol. 73, pp. 428–440, 1999.

[4] M. W. Lee and R. Nevatia, “Body part detection for human pose estimation and tracking,” in *Proc. of the IEEE Workshop on Motion and Video Computing*, Austin, USA, pp. 23–23, 2007.

[5] M. Gochoo, T. Tan, S. Huang, T. Batjargal, J. Hsieh *et al.,* "Novel IoT-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning," *IEEE Internet of Things Journal*, vol. 6, pp. 7192–7200, 2019.

[6] L. W. Hu and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, pp. 585–601, 2003.

[7] A. Jalal, N. Sarif, J. T. Kim and T. S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home," *Indoor Built on Environment*, vol. 22, pp. 271–279, 2013.

[8] J. Liu, J. Luo and M. Shah, "Recognizing realistic actions from videos in the wild," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops*, Miami, USA, pp. 1996–2003, 2009.

[9] Y. Wang and G. Mori, "Multiple tree models for occlusion and spatial constraints in human pose estimation," in *Proc. of the Lecture Notes in Computer Science*, Glasgow, UK, pp. 710–724, 2008.

[10] Y. Zou, Y. Shi, D. Shi, Y. Wang, Y. Liang *et al.,* "Adaptation-oriented feature projection for one-shot action recognition," *IEEE Transaction and Multimedia*, vol. 22, pp. 3166–3179, 2020.

[11] A. Jalal, M. Mahmood and A. S. Hasan, "Multi-features descriptors for human activity tracking and recognition in indoor-outdoor environments," in *Proc. of the Proc. of 2019 16th Int. Bhurban Conf. on Applied Sciences and Technology*, Rawalpindi, Pakistan, pp. 371–376, 2019.

[12] M. A. Khan, K. Javed, S. A. Khan, T. Saba, U. Habib *et al.,* "Human action recognition using fusion of multiview and deep features: An application to video surveillance," *Multimedia Tools and Application*, vol. 79, pp. 1–27, 2020.

[13] Y. Wang, B. Du, Y. Shen, K. Wu, G. Zhao *et al.,* "EV-Gait: Event-based robust gait recognition using dynamic vision sensors," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 6358–6367, 2019.

[14] A. Franco, A. Magnani and D. Maio, "A multimodal approach for human activity recognition based on skeleton and RGB data," *Pattern Recognition Letters*, vol. 131, pp. 293–299, 2020.

[15] A. Jalal, M. Uddin and T. S. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home," *IEEE Transaction on Consumer Electronics*, vol. 3, pp. 863–871, 2012.

[16] A. Ullah, K. Muhammad, I. U. Haq and S. W. Baik, "Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments," *Future Generation Computer System*, vol. 96, pp. 386–397, 2019.

[17] O. Amft and G. Tröster, "Recognition of dietary activity events using on-body sensors," *Artificial Intelligence in Medicine*, vol. 42, pp. 121–136, 2008.

[18] Y. G. Iang, Q. Dai, T. Mei, Y. Rui and S. F. Chang, "Super fast event recognition in internet videos," *IEEE Transaction on Multimedia*, vol. 17, pp. 1174–1186, 2015.

[19] Y. Chen, X. Zhao and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, pp. 2381–2392, 2015.

[20] A. Li, Z. Miao, Y. Cen, X. Zhang, L. Zhang *et al.,* "Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning," *Pattern Recognition*, vol. 108, pp. 107355, 2020.

[21] M. Einfalt, C. Dampeyrou, D. Zecha and R. Lienhart, " Frame-level event detection in athletics videos with pose-based convolutional sequence networks," in *Proc. of the 2nd Int. Workshop on Multimedia Content Analysis in Sports*, New York, USA, pp. 42–50, 2019.

[22] J. Yu, A. Lei and Y. Hu, "Soccer video event detection based on deep learning," in *Proc. of the Lecture Notes in Computer Science*, San Jose, USA, pp. 377–389, 2019.

[23] R. J. Franklin and D. V. Mohana, "Anomaly detection in videos for video surveillance applications using neural networks," in *Proc. of the 4th Int. Conf. on Inventive Systems and Control*, Coimbatore, India, pp. 632–637, 2020.

[24] B. H. Lohithashva, V. N. M. radhya and D. S. Guru, "Violent video event detection based on integrated LBP and GLCM texture features," *Rev. D'Intelligence Artif.*, vol. 34, pp. 179–187, 2020.

[25] Q. Feng, C. Gao, L. Wang, Y. Zhao, T. Song *et al.,* "Spatio-temporal fall event detection in complex scenes using attention guided LSTM, " *Pattern Recognition Letters*, vol. 130, pp. 242–249, 2020.

[26] M. H. Khan, M. Zöller, M. S. Farid and M. Grzegorzek, "Marker-based movement analysis of human body parts in therapeutic procedure," *Sensors*, vol. 20, pp. 3312, 2020.

[27] N. Golestani and M. Moghaddam, "Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks," *Nature Communication*, vol. 11, pp. 1–11, 2020.

[28] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, San Diego, USA, pp. 886–893, 2005.

[29] X. Zhu, J. Pang, C. Yang, J. Shi and D. Lin, "Adapting object detectors via selective cross-domain alignment," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 687–696, 2019.

[30] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. of the Seventh IEEE Int. Conf. on Computer Vision*, Kerkyra, Greece, vol. 2, pp. 1150–1157, 1999.

[31] X. Wang, H. Gong, H. Zhang, B. Li and Z. Zhuang, "Palmprint identification using boosting local binary pattern," in *18th Int. Conf. on Pattern Recognition*, Hong Kong, China, pp. 503–506, 2006.

[32] J. C. Niebles, C. W. Chen and L. Fei-Fei, "Modeling temporal structure of decomposable motion segments for activity classification," in *Proc. of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Crete, Greece, pp. 392–405, 2010.

[33] S. Sun, Z. Kuang, L. Sheng, W. Ouyang and W. Zhang, "Optical flow guided feature: A fast and robust motion representation for video action recognition," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 1390–1399, 2018.

[34] Y. Zhao, L. Zhou, K. Fu and J. Yang, "Abnormal event detection using spatio-temporal feature and nonnegative locality-constrained linear coding," in *IEEE Int. Conf. on Image Processing*, Phoenix, USA, pp. 3354–3358, 2016.

[35] C. Lu, J. Shi and J. Jia, "Abnormal event detection at 150 fps in matlab," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Sydney, Australia, pp. 2720–2727, 2013.

[36] Y. Cong, J. Yuan and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Colorado Springs, USA, pp. 3449–3456, 2011.

[37] W. Wen, W. Su, H. Tang, W. Le, X. Zhang *et al.,* "Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing," *Cell Discovery*, vol. 6, pp. 1–18, 2020.

[38] L. Zhang and X. Xiang, "Video event classification based on two-stage neural network," *Multimedia Tools and Application*, vol. 79, pp. 1–16, 2020.

[39] C. Rodriguez and B. L. Fernando, "Action anticipation by predicting future dynamic images," in *Proc. of the European Conf. on Computer Vision*, Germany, pp. 1–15, 2018.

[40] Y. Ghadi, W. Manahil, S. Tamara, A. Suliman, A. Jalal *et al.,* "Automated parts-based model for recognizing human-object interactions from aerial imagery with fully convolutional network," *Remote Sensing*, vol. 14, pp. 1–15, 2022.

[41] M. Jain, H. Jegou and P. Bouthemy, "Better exploiting motion for better action recognition," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Portland, USA, pp. 2555–2562, 2013.

[42] Y. Li and S. G. Liu, "Temporal-coherency-aware human pose estimation in video via pre-trained res-net and flow-CNN," in *Proc. of the Int. Conf. on Computer Animation and Social Agents*, Seoul, South Korea, pp. 150–159, 2017.

[43] H. Wang, D. Oneata, J. Verbeek and C. Schmid, "A robust and efficient video representation for action recognition," *International Journal of Computer Vision*, vol. 119, pp. 219–238, 2016.

[44] A. Jalal, S. Kamal and D. Kim, "Depth silhouettes context: A new robust feature for human tracking and activity recognition based on embedded HMMs," in *Proc. of the 12th Int. Conf. on Ubiquitous Robots and Ambient Intelligence*, Goyangi, Korea (South), pp. 294–299, 2015.

[45] K. Kim, A. Jalal and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *Journal of Electrical Engineering & Technology*, vol. 14, pp. 2567–2573, 2019.

[46] N. Amir, A. Jalal and K. Kim, "Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model," *Multimedia Tools and Applications*, vol. 80, pp. 21465–21498, 2021.

[47] I. Akhter, A. Jalal and K. Kim, "Pose estimation and detection for event recognition using sense-aware features and adaboost classifier," in *Proc. of Conf. on Applied Sciences and Technologies*, Islamabad, Pakistan, pp. 500–505, 2021.

[48] A. Kaveh and S. Talatahari, "A novel heuristic optimization method: Charged system search," *Acta Mechanic*, vol. 213, pp. 267–289, 2010.

[49] E. Park, X. Han, T. L. Berg and A. C. Berg, "Combining multiple sources of knowledge in deep CNNs for action recognition," in *Proc. of the 2016 IEEE Winter Conf. on Applications of Computer Vision*, Lake Placid, NY, USA, pp. 1–8, 2016.

[50] A. Jalal, I. Akhtar and K. Kim, "Human posture estimation and sustainable events classification via pseudo-2D stick model and k-ary tree hashing," *Sustainability*, vol. 12, pp. 9814, 2020.

[51] S. Gomathi and T. Santhanam, "Application of rectangular feature for detection of parts of human body," *Advance Computational Science and Technology*, vol. 11, pp. 43–55, 2018.

[52] Y. Kong, W. Liang, Z. Dong and Y. Jia, "Recognising human interaction from videos by a discriminative model," *IET Computer Vision*, vol. 8, pp. 277–286, 2014.

[53] L. Zhang and X. Xiang, "Video event classification based on two-stage neural network," *Multimedia Tools and Application*, vol. 79, pp. 21471–21486, 2020.

[54] S. Sun, Z. Kuang, L. Sheng, W. Ouyang and W. Zhang, "Optical flow guided feature: A fast and robust motion representation for video action recognition," in *Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 1390–1399, 2018.

[55] R. Mehran, A. Oyama and M. Shah, "Abnormal crowd behavior detection using social force model," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, USA, pp. 935–942, 2009.

[56] X. Cui, Q. Liu and M. Gao, "Abnormal detection using interaction energy potentials," in *Proc. on Computer Vision and Pattern Recognition*, Colorado Springs, USA, pp. 3161–3167, 2011.

[57] B. Degardin and H. Proença, "Iterative weak/self-supervised classification framework for abnormal events detection," *Pattern Recognition Letters*, vol. 145, pp. 50–57, 2021.

[58] Y. Zhu, K. Zhou, M. Wang, Y. Zhao and Z. Zhao, "A comprehensive solution for detecting events in complex surveillance videos," *Multimedia Tools and Application*, vol. 78, pp. 817–838, 2018.