International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Intelligent Video Analytics for Abnormal Event Detection

Dr.Shweta Marigoudar[1] Dean FCIT,GM University, Davanagere, Karnataka, India
Ms.Shamina Attar[2]**,**Assistant Professor, GM University, Davanagere, Karnataka, India

Abstract

Surveillance cameras have become a primary tool for monitoring human movement and preventing unwanted or unintended activities. In today's world, security management professionals increasingly rely on video surveillance to combat crime and mitigate incidents that could adversely affect society. The number of surveillance camera installations has dramatically increased in both the private and public sectors, serving to monitor public activities effectively. Video surveillance is considered one of the most efficient methods for ensuring security. Installing a surveillance camera allows the captured footage to be transmitted to security personnel; however, merely having video footage is insufficient for identifying abnormal activities. To effectively analyze this footage, it is essential to integrate an intelligent system.This paper aims to design and implement an Intelligent Video Analytics Model (IVAM), also known as the Human Object Detection (HOD) method, for analyzing and detecting video-based anomalies and abnormal human activities. IVAM can be deployed alongside surveillance cameras in various public locations such as Institutions, airports, hospitals, shopping malls, and railway stations, allowing for the automatic identification of unusual events. The IVAM was experimented with using MATLAB software, and the results were thoroughly verified. The performance of IVAM was assessed by comparing the obtained results with those from existing approaches, demonstrating that IVAM outperforms contemporary methods concerning accurate anomaly detection, lower error rates, and higher classification accuracy.

1.Introduction

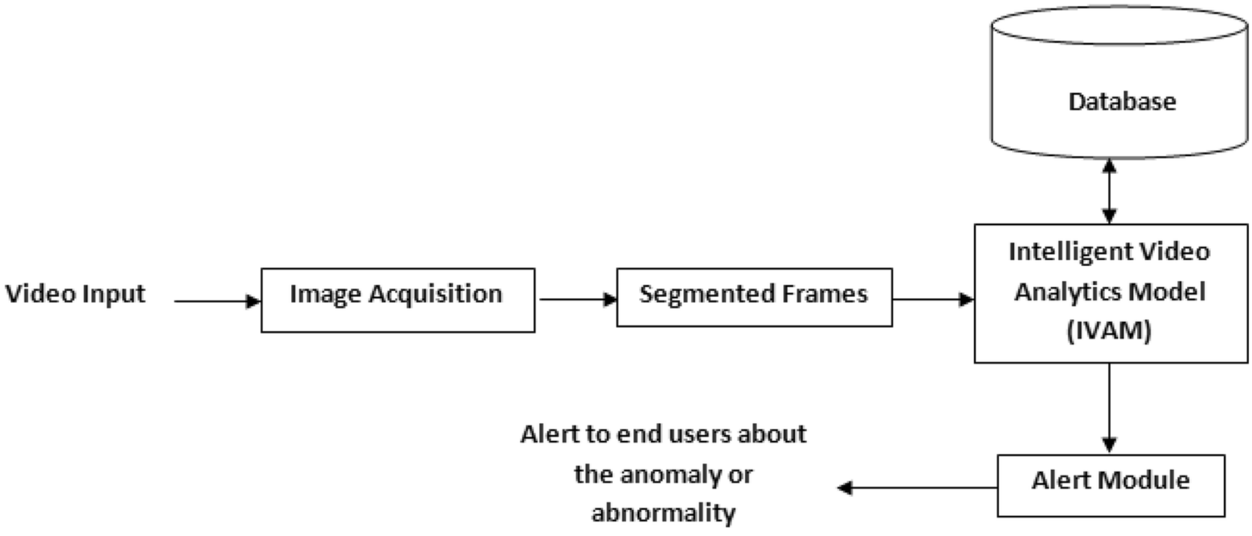
In recent days an Artificial Intelligent (AI) system makes a person work more efficiently by simplifying their job with more precision and accurate results. Personal and organization security cannot be brought by just installation of a surveillance camera. AI-based surveillance practices along with intelligent video analytics calculates and provides security measures that have a significant impact in preventing anomalies. Video analytics deploy several cutting edge technologies that a camera alone cannot accomplish. This paper is motivated towards integrating video analytics into a surveillance system to detect abnormal activities which pose a serious threat towards the security of a person or an organization. Surveillance systems that deploy vision-based video analytics analyze the video recordings in real-time environment and detect abnormal activities that could pose a serious threat to the security of an individual or an organization. Technologies based on video analytics aid security systems to understand and "learn" how to distinguish between normal and abnormal activities so that it can easily detect unusual and potentially unsafe activities that may even get unnoticed by security personnel. This can be accomplished in two ways; first the objects in the environment are monitored closely to identify the human and vehicle movements in the surrounding area and second by collecting the feedback from the personnel about the occurrence and accuracy of various events and having this assimilated knowledge incorporated as an intelligence system itself, thereby improving the overall functional aspect of the system through which a "teachable" system is developed. Artificial Intelligence when properly deployed and integrated with any system provides ultimate security, which may not even be matched by eminent human security systems. The primary factor that drives the need to develop and deploy an AI-based system for security is to nullify and suppress the occurrence of human errors while monitoring the applications. Several studies indicate that any person who is involved in monitoring task have full attention and perform to their best for a maximum time of 20 min beyond which the attention level tends to decrease gradually based on the nature of the security person. This issue is compounded by the fact that the security person has to keep in track and monitor multiple tasks and events occurring in parallel which might further deteriorate the attention levels. Thus, video analytics based AI systems have become the need of the hour to promote efficient surveillance. To detect the abnormal events various innovative ideas in video analytics and AI-based surveillance has been used quite frequently in several areas such as temporal object tracking [1], people detection [2], traffic monitoring [3] and host detection [4]. In addition, different methods were used to detect moving objects in online video surveillance systems.

A camera is installed with AI and video analytics is designed to capture a snap of a person who manages to sneak into a secured place even under poor lighting conditions. Another instance where these intelligent systems can outsmart human surveillance is to spot and track persons loitering around the perimeters of secured buildings such as banks, schools, ATMs, government organizations, etc. and these systems in turn alert the security personnel to follow-up and track the person suspected to be anomalous. This way the AI systems remove the complete dependency of human monitoring and instead provide notifications to the security team well in advance even before the occurrence of an anomaly or at the time of occurrence of anomalous behavior. The turnaround time to detect issues is thus reduced. The objective of this paper is to design and implement an automatic intelligent video surveillance system to identify, detect and alert the abnormal object and human exists in the video. The overall flow of the work is given below.

- A video is recorded using a surveillance camera and this video is in turn transmitted to a server (or Database) in a real-time environment.
- The video is segmented and each segment is converted into frames.
- All the frames are preprocessed, background subtracted and morphological operation based objects are applied to detect objects.
- The detected objects are then compared with the ground truth objects to find out the abnormal objects (ODD) as defined by the user.
- Detect multiple objects which include moving objects, people, vehicle, etc., and the corresponding behavior is analyzed at the same time.
- Whenever an anomaly object is detected, it is labeled and alerted back to the server or to the end user who in general is the security personnel.

The block diagram of Intelligent Video Analytics Model (IVAM) is depicted in Fig. 1. The camera is fitted inside a college, street and other important places where it needs security highly. In this paper, the two different videos are taken for experimenting the proposed approach. One is a benchmark video described in the next section and the other one is custom video (user-defined) taken by a higher-end digital camera inside a campus. The file format of the video data should be in a codec (container) format where this format describes the blueprint of the file while coding (encoding–decoding) the video data. Several containers format holds the data encoded with various codecs. To read the video file, the file should be in a recognized format such as in AVI. In addition, it is able to access the codec format where it helps to decode the easily and hastily.

Fig. 1



The primary objective of this model is to ensure that the IVAM model quickly recognizes the anomalies present in the video sequence by utilizing very minimal resources and quick processing time. To understand the problem statement of this study, it is necessary to learn the previous research works focused towards event detection and classification tasks in videos which is carried out in the subsequent section.

2.Related works

Most of the time, the objective is to identify, detect and diagnose the objects, events and human activities, which is factually defined as a normal or abnormal behavior [5, 6]. Few of the authors in [7,8,9,10,11,12,13,14] have carried out an overview on abnormal detection in automatic surveillance systems. Brezeale and Cook in [7], have performed a review about categorizing activities of a specific video produced in 2008, while the author in [8] has spoken in detail about understanding the activities or events recorded in video data. The author in [9] has discussed about various earlier research works based on understanding the dynamic events occur in the video. Dee and Hogg, and Haering et al. in [11] presented a detailed survey where the methods proposed for anomaly detection in real-time surveillance systems. Then Raty [12] delivered an overview of the same topics. In addition, there were no earlier reviews that can sufficiently depict the process of detecting anomaly concurrently including both the fields such as, surveillance and anomaly detection. It is noticed that most of the works carried out towards anomaly detection have used the same method to undertake the project of anomaly detection. Vapnik and Lerner [15] formerly anticipated the support vector machine for taxonomy or regression. Xinyi Cui et al. [16] in 2012 have suggested a new method, where they make use of an interaction energy potential function that represents the current behavioral state of objects and according to that SVM classified is used to classify the normal and abnormal events in the video. In addition to these, the approaches which have been discussed as part of [17,18,19,20,21,22,23] have been studied to understand the working of best-fit algorithms that will aid in anomaly detection.

Additionally, these methods do not rely on object detection and object tracking, but outperforms with better efficiency than social force [24] and optical flow [25]. The efficiency of the optical flow method can be used for abnormal detection. The author in [26] said that, when the people crowd increased then the surveillance tracking method becomes unreliable. Various researches have confirmed that the histogram value of the optical flow alignment descriptor is fit for reciting the movement information under global scenario such as video frames and foreground frames [27,28,29,30], but does not perform that well in sparse scenes. Pretty recently, Tian Wang et al. [31] proposed a framework for representing and analyzing the objects based on the optical flow method based visual features. Author in [32] used 1-class SVM method and PCA [33, 34] for abnormal behavior detection. Methods like SVM need to learn the entire data for accurate classification. Similarly, various existing approaches depend on various internal and external processes. The proposed approach is developed with an eye to bridge the gaps in these proposed approaches and to improve the overall efficiency of the surveillance system.

3.Problem statement and motivation

One of the fast-growing challenges in recent days is protecting people and their valuable belongings in public places. To address this issue, video surveillance-based object analysis methods are used. Hence, it is essential to deploy surveillance cameras across many places where security is needed. Thus, video processing and understanding of the activities involved in the video becomes a popular research work nowadays. These video analysis techniques enable the camera to identify, detect and classify the targets of objects and their relevant activities. One of the critical tasks is to identify and detect odd events or unexpected activities occurring in a video sequence. However, this problem is not solved accurately and with higher efficiency levels. In this case, a video surveillance system which can understand the scene and automatically distinguish abnormal activities can play a pivotal role. The system would then alert the operators or users accordingly. This problem has motivated us to develop an intelligent video surveillance system for identifying, detecting the abnormal events accurately and alerting the concerned users of the abnormal events.

4 Limitations pertaining to existing research works

From the literature survey, the limitations pertaining to the existing research works are: the odd events are obtained by comparing new images with the next or to the previous images in the sequence. The memory utilization to store all the images of the video is high and comparison time between frames is too long since lots of redundant and unnecessary comparisons are involved. With this in mind, some of the earlier research works stored only some novel images (like key frames) are kept and used for comparison. Few of the research works used a bag of video words to perform anomaly detection, whereas it also takes longer time for comparison and more memory for storing the bag of video words as distributed.

5 Need for the proposed work

The necessity of this proposed work is to increase the security levels in any kind of video surveillance applications by improving the accuracy and efficiency of classifying anomalous events. Placing video cameras are envisioned to record crimes or criminals. This system acts as a security guard, which is predominantly used to investigate the unexciting droning sequence images. Expecting a security person at all the times to remain vigilant is very demanding. Hence this work eliminates the onus of detecting unusual scenarios by the security personnel to a little extent, dwellings it on the video surveillance system. In addition, it is fully automatic and it can be executed and monitored the odd situations in the environment by the admin or the user at anytime from anywhere in the network.

6.Dataset used

There are two different datasets that are used to verify and evaluate the performance of the IVAM approach namely UCSD-AD dataset and a custom dataset. Both datasets were acquired by a normal canonical camera mounted at the height of a 35 feet long from the ground, overlooking pedestrian walkways and a college campus, respectively. The density of the people crowd varies from 5 to 20 numbers. To increase the accuracy of the object classifications, some of the points are considered as assumptions and it is given below.

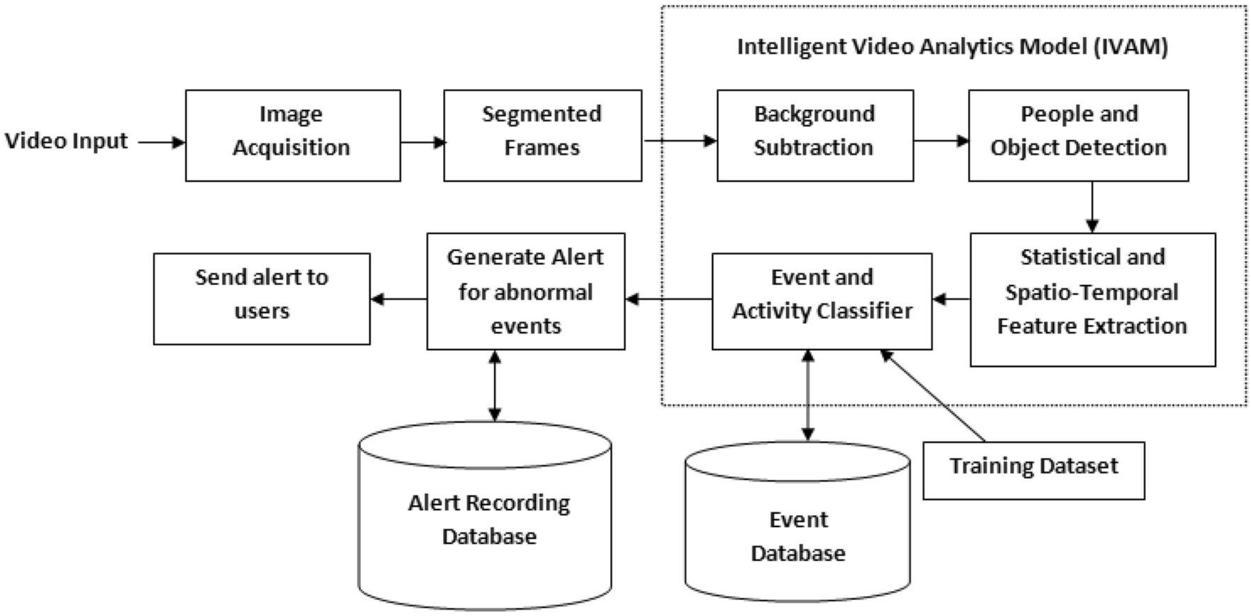
- The experimental videos contain people walking over a pedestrian lane and some moving cars.
- The anomalies considered in these videos are riding bicycle, skating, riding a mini truck, etc., in a pedestrian lane where people normally walk.
- Also, these moving objects in the video such as cars, trucks etc., are of different sizes and colors.
- Also, it is assumed that the abnormalities are unsurprisingly occurring.
- Each video segment comprises below 200 frames, due to the reduction in the complexity of the video processing.

There are 50 video samples used for training and 48 video samples used for testing which are taken from UCSD-AD dataset which is publicly available at "http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm". To verify the working of the proposed approach, the classified result is compared with the ground truth annotation which is included in the dataset. Since this is a benchmark dataset, each frame has a true-annotation with a binary-flag showing that particular frame has abnormal event generated by a person or an object. A set of video segments has their own binary mask showing the region of the abnormal human/objects. It also helps to increase performance evaluation with respect to the capability of the proposed approach in detecting abnormalities.

7. Intelligent video analytics model

The IVA model utilizes a Human Object Detection (HOD) method to detect the abnormal images or objects in a video sequence. It is well known that a video is a collection of sequential images called as frames. Each frame is considered as an image and any kinds of image processing methods can be applied to the frames. Since the surveillance camera records the environmental video round the clock, the size of a video is not fixed and can be segmented for easy process. Each video considered for the experiment is segmented first. Then the segments are converted into a set of frames. The built-in functionalities of computer vision in MATLAB is used for read the video, segment the video and convert videos into frames. Frames are individual images and various image processing methods are applied for pre processing, background subtraction, and efficient segmentation of objects. To segment the objects, morphological operations are used here. Then a classifier is applied to analyze the behavior of the people and objects and determine the behavior as normal or abnormal. Here the abnormality is defined as dimensions where their likelihood is not matched with the determined threshold pertaining to a normal object. Unusual and rare behavior, not observed in the training process is considered as abnormal. The unusual object is detected by comparing the new objects with the background objects. All the video processing related applications commonly use moving object detection. The regions of the moving objects like people and vehicles in a video is the primary stage of the vision system, because it gives a focus of attention and streamlines the processing on successive analysis stages. Motion detection will not be a reliable process when a dynamic change happens due to weather changes or any other nature imposed change. Some of the methods used frequently for object detection are preprocessing, background subtraction, frame differentiation, statistical methods, feature extraction and mapping, and optical flow. Figure 2 shows the complete architecture of IVAM.

Fig. 2



A video V is divided into K number of segments $\{S1,2,\ldots,SK\}$where the value of K depends on the video length L. Length L represents the running time T of the video. Time taken for each video segment Si is $=\{t1,t2,\ldots,tk\}$. In this paper, N number of videos $\{V1,2,\ldots,Vi,\ldots,VN\}$ are used for the training process and the segments are given in the form of matrix as

S=[$SS11SS12$:$SS1kkSS21SS22$:$SS2kk$::::$SSNN1SSNN2$:$SSNNkk$]

where each segment Sij comprises N number of frames, and assumed that the number of frames in each segment is lesser than 200, due to the reduction in the computational and comparison complexity. The set of all events in Video Vi is

Vi=$\{ei\}i$=1N

having the set of measures M where the abnormal event ea is defined as

ea∈Vi

Ma, i=(ea, ei),ei∈V−{ea}

(1)

ea is abnormaliff$f$∀i,, i⩽γ

where γ is a threshold. It means that the event eq is not matched with the other observed events. The new events are always compared with the valid events constructed in training class. It can be obtained by computing the likelihood ratio that identifies the abnormal patterns which are inferred by choosing video arrangements having lesser likelihood values. Initially, the redundant data is removed from the video frames by computing the similarity of the spatio-temporal values of the frames. It is assumed that in a 1-min video, the number of events occurring is roughly calculated in 5–10. In case of lengthy videos which are captured for a longer duration, the video volumes are approximated. To create a probabilistic model, only a small number of video frames having valid events is taken for the process of training. This leads to constructing the problem formulation for abnormal detection. To make sure that, the input objects are classified as abnormal event in a short amount of time, information about the objects' region of interest should be stored effectively in a database. The proposed approach is aimed to reduce the time, and effective reconstruction process in abnormal detection. In addition, the frame information is reduced to increase the memory utilization and restrict the searching time to increase the accuracy of the event classification. Focusing on the frame information maintenance, leads to classifying the normal and abnormal events and since it is used in the training process effectively.

7.1 Background subtraction

The concept of background subtraction is particularly employed in the field of image processing to automate computer vision. This process is mainly used for object recognition like human, car or any objects visible in the foreground of the frame. In our method, this is mainly used to detect the abnormal behavior of moving objects. To detect the objects in the frame, background subtraction is applied in a particular frame to identify an object. There will be a static background or dynamic background according to the nature of the scene. But in our case, the scene is considered as dynamic as there will be more objects moving across in a given scene. Here, we need to detect the objects by subtracting the current frame from a reference frame. In this work, the reference frame is considered to be a fixed frame. Hence we consider the background as static and the foreground as dynamic. By subtracting the current frame which have different objects (human or object) with the referenced frame which have the static background. Background subtraction method is straightforward and easy to understand, and exactly extracts the individuality of target object, but it is aware to the transform of external environment, therefore, it is valid to the condition that the background is well known. After fixing the background frame, subtraction of the current frame may be done to detect the objects. Subtracting both frames can be done pixel-by-pixel. Current frame pixel is marked as a foreground pixel that is $\mathrm{fr}(x, y)$ and as well as the corresponding background frame pixel marked as $\mathrm{bg}(x, y)$. The predefined threshold value called γ is computed using (2).
$\gamma<\mathrm{fr}(x, y)-\mathrm{bg}(x, y)$.

(2)

7.2 Morphologically object detection

After subtracting the frames the resultant frames are subjected to segmentation. The difference value is compared with the predefined threshold value. If the difference in pixel values is larger than the predefined threshold γ, then it can be assumed that the pixels appearing correspond to that of a moving object, otherwise, they can be assumed to be the background pixels. The target object will be identified after applying the threshold operation. The expression for this is shown in (3) as follows:
$d(x, y)=\{1, \text{ if}(|\gamma<\mathrm{fr}(x,y)-\mathrm{bg}(x,y)0\text{otherwise}$

(3)

The resultant frame which got out of this expression provides the subtracted value. Each pixel of this subtracted frame is compared with the threshold value. If the threshold value is lesser than the subtracted pixel value, the above-mentioned expression returns the value 1 and will return the value 0 otherwise. In general, when it comes to image processing, the value of 0 is considered as black and a value of 1 is considered as white. Thus, we can easily identify the target objects in the current frame where the target objects will be white in color while the background becomes black.

The segmentation on binary images leads to limitless issues. Making a binary image with simple threshold contains more noises. Morphology in image processing alters the object depending upon the desired shape. It is used across various applications like noise elimination, texture analysis, boundary extraction etc. The issues in the binary images are eliminated by morphological techniques. Instead of their numerical values, morphological operations are convinced only on the related ordering of pixel values. Hence, there is more attention on binary images. Morphological operations can also be applied to gray-scale images. In this case, the light transfer functions corresponding to gray images are unknown so that their absolute pixel values are not considered. A small template image is selected and applied to all the locations of the input image and the same sized output images are generated. This will give the new image that have non zero pixels unlike the input image. We may consider the templates in the shape of square, diamond or may be of any shape. The basic operations are dilation and erosion which are based on logical AND, OR notation and are described by set analysis. Dilation includes pixels at boundaries of objects while erosion is the process where the pixels at boundaries of the objects are eliminated. This including or excluding of pixels depends on the structuring element that is used for image processing. In morphology, the basic operation is dilation and it can be applied for both binary image and gray-scale image. This will cause the growth of object. It increases the boundary of the foreground and hence it leads to making the holes smaller. Dilation happens in two parts. First one is dilation and the second is structuring. With the help of structuring, we can determine how much the input image should be dilated. The mathematical definition of dilation can be as follows:

Consider Si is a set of coordinates pertaining to input images and SS~ is a set of structuring element coordinates and Six is a translation of SS so that its origin is at x. Thus, dilation of Si by SS is set of all points of x such that intersection of Six with Si is not null. In terms of set operations, dilation of S_i denoted by SS is defined in (4).
$Si\Theta Sj=\{x \,|(Sj^\wedge)x\cap\neq\emptyset\}$

(4)

The boundaries of the objects are segmented by the dilation operation after which the region is filled. In binary images, the background pixels are changed to foreground pixels until the object boundaries are received as a gray-scale value. This is followed by the filling operation which fills the various holes in the acquired input images automatically. Erosion is the mathematical morphology where it erodes the foreground objects in accordance to their region. In addition, this operation can shrink or enlarge by itself according to the objects size. The mathematical definition of erosion function is: Let Si is the input image coordinates, Sj is the coordinates of the structuring element, then Sjx is the translation of Sj, so that its origin is at x. Thus, dilation of Si by Sj is the set of all points of x such that Sjx is one of the subset of Si. The erosion of Si by Sj is written in (5).
$Si\Theta Sj=\{x|(Sj)x\cap Sic\neq\emptyset$

(5)

The set of all objects are considered as the events ei and they are classified as normal or abnormal by comparing their space–time oriented information.

8.Conclusion

The main objective of this paper is to design and implement a novel intelligent video analytical model as a human object detection method for surveillance video. From the surveillance video, the activity in a particular region is recorded, then the video is segmented, segments are converted into frames. Then frames are preprocessed by removing the noise and background is subtracted. After background subtraction, the frames automatically becomes binary images and the objects are detected using morphological operations. Since the objects are moving this paper uses the space–time related features for classifying the detected objects as normal or abnormal. To evaluate and verify the efficiency of the proposed IVA model, it is programmed in MATLAB software and the results are verified. From the experimental results, it is found and concluded that the IVA model outperforms most of the existing approaches in terms of object detection and classification with less error. From the results, it is observed that the IVA model has achieved 99.77% and 98.19% accuracy in correctly classifying the normal and abnormal frames from benchmark dataset and custom dataset, respectively. Similarly, 0.23% and 2.03% of normal and abnormal frames are classified incorrectly from benchmark dataset and custom dataset, respectively. Hence it can be concluded that the IVA model is better and suitable for online video processing when compared to other contemporary methods.

Future work will be to take this approach and experiment and verify the results with more benchmark dataset and custom dataset and to integrate this and model along with a sophisticated framework that can be deployed for smart surveillance.

References

1. Chaturvedi, P.P., Rajput, A.S., Jain, A.: Video object tracking based on automatic background segmentation and updating using RBF neural network. Int. J. Adv. Comput. Res. **3**, 866 (2013)

   Google Scholar

2. Kim, J., Grauman, K., Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June (2009). pp. 2921–2928

3. Zhan B, Monekosso DN, Remagnino P, Velastin SA, Xi IQ, Crowd analysis: a survey. Mach. Vis. Appl. 19:345–357 (2008)

   Article Google Scholar

4. De Almeida IR, Jung, CR, Change detection in human crowds. In: Proceedings Conference on Graphics, Patterns and Images, August (2013), pp. 63–69

5. Oluwatoyin P, Popoola, Wang K, Video-based abnormal human behavior recognition—a review. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 42, 6, 2012, 865–878

   Article Google Scholar

6. Saykol, E., Gudukbay, U., Ulusoy, O.: Scenario-based query processing for video-surveillance archives. Eng. Appl. Artif. Intell. **23**(3), 331–345 (2010)

   Article Google Scholar

7. Brezeale, D., Cook, D.J.: Automatic video classification: a survey of the literature, IEEE Transactions on System, Man and Cybernetics-Part C: Applications and Reviews, 38, 3, 416–430, 2008

8. Lavee, G., Rifling, E., Rudzsky, M.: Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video. Syst. Man Cybern. Part C Appl. Rev. 39, 5, 489–504, 2009

   Article Google Scholar

9. Buxton, H.: Learning and understanding dynamic scene activity: a review. Image Vis. Comput. 21, 1, 2003, 125–136

   Article Google Scholar

10. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 34, 3, 2004, 334–352

Article **Google Scholar**

11. Haering, N., Venetianer, P.L., Lipton, A.: The evolution of video surveillance: an overview. Mach. Vis. Appl. 19, 5–6, 2008, 279–290