Smart Video Surveillance and Object Recognition System with novel RPCA-MFTSL and tracking with CNN

S. Naveen Kumar[1], R. Shoba Rani[2]

Abstract

In several application sectors, identifying human behaviour from surveillance footage is a challenging problem. The consequences of abnormal/unwanted actions in public areas, health-care organisations, and smart homes, together with other security concerns, necessitate a dependable and intelligent surveillance system for abnormal behaviour. Therefore, recognising and identifying objects and human movements in surveillance recordings, in addition to automated content encapsulation, has become a hot area of study. In this research, we will offer video surveillance abnormal activity detection. A rapidly evolving video analysis problem, aberrant event detection tries to distinguish between abnormal and normal occurrences in surveillance footage. Because normal and abnormal occurrences are equivalent in certain respects, it is necessary to examine additional distinguishing methods or motion data. Robust Principle Component Analysis was utilised for Background removal, whereas MFTSL was utilised for foreground identification. For Analytics, we employed a CNN-based tracking technique. Our suggested approach estimates moving objects with excellent precision.

Key Words: Background subtraction, RPCA, MFTSL, CNN, moving object

DOI Number: 10.14704/nq.2022.20.8.NQ44047 **NeuroQuantology 2022; 20(8):406-417**

Introduction

This is the most common method for distinguishing dynamic objects from their surroundings, which involves moving a camera. Human activities, object tracking, traffic dynamics, and PC vision are all examples of typical uses.. If you want to keep your data apart from the moving object, this technique refers to it as an additional backdrop area. In line with the final objective. It's more difficult to create background subtraction models than it is to create actual subtraction models. This motion detection method is now widely utilised [1]. A simple surveillance system, for instance, is incapable of detecting and preventing suspicious activities in public areas and essential structures. Due to a lack of physical fitness, CCTV cameras, and security personnel, the human operator was unable to avert a hazardous scenario by maintaining a watchful eye on the monitor. Police personnel may not be able to conduct continual monitoring. The question arises as to whether terrorist activity and suspicious incidents can be eliminated from public settings. Numerous scientists and academicians are committed to resolving this issue [2]. In this paper, we provided a technique and approaches used by researchers to conduct autonomous surveillance.

Visual surveillance experts have recently concentrated their efforts on action and behavior recognition. Not just in security situations where restricted areas are scrutinized, such as airports and shopping malls, but also in ordinary settings such as schools and workplaces, it is a massive and crucial obligation to comprehend what is occurring in one's surroundings. Contextually, a video surveillance sequence can be classified as either dynamic or static. [3].

Corresponding author: S. Naveen Kumar
Address: [1]Research Scholar, Department of Computer Science, Dr. M.G.R. Educational and Research Institute, Chennai, India, [2]Professor, Department of Computer Science and Engineering, , Dr. M.G.R. Educational and Research Institute, Chennai, India
E-mail: sk.naveenkumar@gmail.com[1], shobarani.cse@drmgrdu.ac.in[2]

Anomaly detection is the technique of finding patterns in data that don't match up with what would be expected based on previous observations of human behaviour. Anomalies, outliers, discordant observations, exceptions, aberrations, surprises, and uniqueness are some of the terms used to describe these phenomena. Many businesses can benefit greatly from the identification of anomalies, which is why anomaly detection is so important. A number of research groups have developed methods for detecting anomalies. The researchers designed numerous algorithms with various demanding circumstances in mind, but one constant remains: In order to describe an action, one must consider the amplitude and direction of the movement [5] [6].

The tremendous spatial and temporal continuity of video data is one of its most distinguishing characteristics, resulting in high dimension and excessive data redundancy. The resolution of taken photographs and videos is becoming increasingly high as electronic technology advances, resulting in a large increase in the amount of video data to be processed. When dealing with large amounts of data, treating them all the same can be impracticable, as it needs a large number of computer resources and a substantial amount of time. In addition, the duplicate information may impede video understanding and analysis. As a result, one of the most pressing challenges in video data processing is eliminating duplicate data while maintaining essential content for later processing. Saliency detection is a wonderful method for resolving this difficult task because of its ability to discover exciting material in photos or movies [7].

In a variety of fields, large datasets are becoming more common. To grasp such datasets, techniques must minimise their dimensionality while preserving the majority of their content. For this, principal component analysis is one of the earliest and most often used approaches among many others (PCA). If possible, it seeks to minimise the dataset's dimensionality while maintaining as much "variability" as possible (i.e., statistical information). Even though a multivariate normal (Gaussian) distribution is commonly assumed for inferential purposes, PCA as a descriptive tool requires no distributional assumptions and, as a result, is a highly adaptable exploratory method that may be applied to a wide range of numerical data types [8].

The explanation for this is the current state of global insecurity. As a result, an intelligent surveillance system is necessary, one that collects data in real time, communicates, analyses, and comprehends information about those being monitored. In this study, we will show how to detect unusual activities using video surveillance. Aberrant event identification is a video analysis task that is still in its early stages of development, with the goal of distinguishing between abnormal and normal occurrences in surveillance footage. Because normal and abnormal occurrences are comparable in some ways, better discriminating methods or motion data must be investigated. We used Robust Principle Component Analysis for Background subtraction and MFTSL for foreground detection. We used CNN based tracking algorithm for Analytics. Accuracy of estimation of moving objects is high in our proposed system.

Related Works

A Dependable and Intuitive Abnormal Conduct Surveillance systems are essential for averting the repercussions of unusual/unwanted behaviour in public areas, healthcare institutions, and besides additional security concerns, smart houses and this means that in surveillance recordings, if you can identify items and people, you've succeeded in addition to automated content encapsulation, has become a hot area of study. In this section, we show a statistical collection of selected articles on aberrant activity recognition that we reviewed in our study from each category.

Smriti .H et al. [9] We've presented a way for identifying unusual actions in the home. This is an attempt to develop a support system to help older persons who live alone. The proposed method finds Spatio-Temporal Interest Points for motion description in the frame by computing salient regions from frames. Language for activity descriptions is developed using the BoF method. To identify whether or not an activity is normal, SVM is employed as a classifier. We found the findings of our system encouraging because our goal is to accurately signal aberrant activity in senior persons. The system gives 100 percent accuracy for the UR Fall dataset; nonetheless, even if the system gives false alerts, the rate of appropriate categorization for aberrant behaviours is 96.83 percent for Dataset S by Le2i CNRS.

Yilin Wang et al. [10] suggested a unique, computationally efficient, and user-friendly approach for estimating spatiotemporal salience. The suggested method was motivated by the recent development of spectrum analysis-based visual saliency techniques, which utilised phase information to generate the image's saliency map. Recognizing that the estimated saliency map represented the human attention region for dynamic scenes, we proposed two approaches for two critical vision tasks that utilised this saliency map.

Mustafa Ayazoglu et al. [11] Structured nuclear norm minimization issues are easily solvable using an iterative Augmented Lagrangian Type (ALM) approach that requires just a mix of matrix thresholding and matrix inversion operations at each iteration. Multiple examples demonstrate that the suggested approach significantly reduced processing time and memory requirements as compared to interior point techniques.

Zhen Lei et al. [12] Using the simple AdaBoost algorithm, we demonstrated an enhanced Haar-like feature-based system for face detection. For face recognition, we additionally used hansdorff distance and an enhanced ICA. The system is built on a small AdaBoost and an upgraded Independent Component Analysis, both of which were independently proposed previously. The system is designed for video surveillance applications and can detect whether or not a particular person is present in a monitored region. Face detection and identification methods for video surveillance applications can be extended to other computer vision object detection and recognition challenges.

Khosro Rezaee et al. [13] Automatic study of crowd security becomes possible when crowd abnormalities or uncommon congestions occur. It is essential to automate the detection of abnormal crowd behaviour because activities such as terrorist operations, violence, and unusual and suspicious movements demand the supervision of operators and watchful staff. However, this is a significant obstacle that leads to excessive pricing and inaccurate decision-making. Consequently, constructing a fatigue- and error-free, real-time system employing the WoT platform will have a good effect on crowd behaviour control. This study investigates several ways for identifying crowd anomalies. Classification based on manual characteristics, classification using deep learning, and hybrid models are all considered.. Anomalies in crowds can be predicted and detected using deep learning and hybrid models, which have improved performance characteristics. However, computing complexity has been addressed in a few approaches for analysing crowd behaviour, where a reduction in processing time might decrease the response time to anomalous behaviour.

Yuncai Liu et al. [14] Deep network architecture for tracking a single target in the present mainstream application. A deep learning-based framework and algorithm for multi-camera video target tracking are offered in conjunction with the target identification network. The software has been able to track the target for an indefinite amount of time. As a result of this fusion technique, the target detection networks are trained using their own calibration data to improve their performance, resulting in enhanced network robustness and application specificity.

Hyochang Ahn et al. [15] Examining the current mainstream application's deep network architecture for a single target tracking task. In combination with the target identification network, a deep learning-based framework and algorithm for multicamera video target tracking are provided. The software has been able to track the target for an indefinite amount of time. The target tracking model is based on deep learning and a target detection network, and the corresponding fusion technique trains the target detection networks using its own calibration data, resulting in enhanced network robustness and application specificity.

Khot Harish S et al. [16] A smart video surveillance system can dramatically improve situational awareness. These systems transform video surveillance from a data collection tool to a system for obtaining information and intelligence. Smart surveillance systems can respond in real time thanks to real-time video analysis. Our system identifies the intruder and notifies authorised people, allowing them to take immediate action.

Osman Ibrahim et al. [17] Moving object detection is a well-known and demanding area of research... Detecting moving blobs provides a focus point for recognition, classification, and activity analysis, making these procedures more efficient because only "moving" pixels must be examined. The SDCS system is a software application designed primarily for traffic management. It has several advantages, including being a less expensive alternative to the conventional radar system. SDCS does not require the services of specialists due to its user-friendly interface and superb design.

Suseendran .G et al. [18] uses image processing and computer vision algorithms to give a pleasant and fashionable method for assessing road traffic density during the day. The video data collected is first weakened into frames, which are then preprocessed during a series of steps. Finally, the vehicles are detected and extracted from the then obtained because of the number of vehicles per unit area of the road section.

Summary

But still there is no method that avoids the crime happening by giving alert or increasing security especially in malls or other crowded area.
Deep Learning based tracking methods are better when compared to some Machine Learning methods.
This is due to the present situation of global insecurity. Consequently, a smart surveillance system for intelligent monitoring is required, one that gathers, transmits, analyses, and comprehends data pertaining to the persons being observed in real time.

Proposed Method

Human behaviour analysis using computer vision techniques has been a topic of interest for researchers. Video surveillance has been used in a pictures and Density estimated. The traffic density is

wide range of settings, including elderly care and home nursing, to mention a few. At a variety of spatial and temporal scales, smart video surveillance systems can improve situational awareness. In congested areas such as train stations, bus stops, and shopping malls, crime is on the rise. Aberrant event identification is a video analysis task that is still in its early stages of development, with the goal of distinguishing between abnormal and normal occurrences in surveillance footage. Additional distinguishing methods or motion information should be investigated because typical and abnormal events are similar in some aspects.
Because of the variety of activities and loudness in the scenes, abnormal event identification in crowd situations is more difficult. In this section, we used Robust Principle Component Analysis for Background subtraction and MFTSL for foreground detection. We used CNN based tracking algorithm for Analytics.

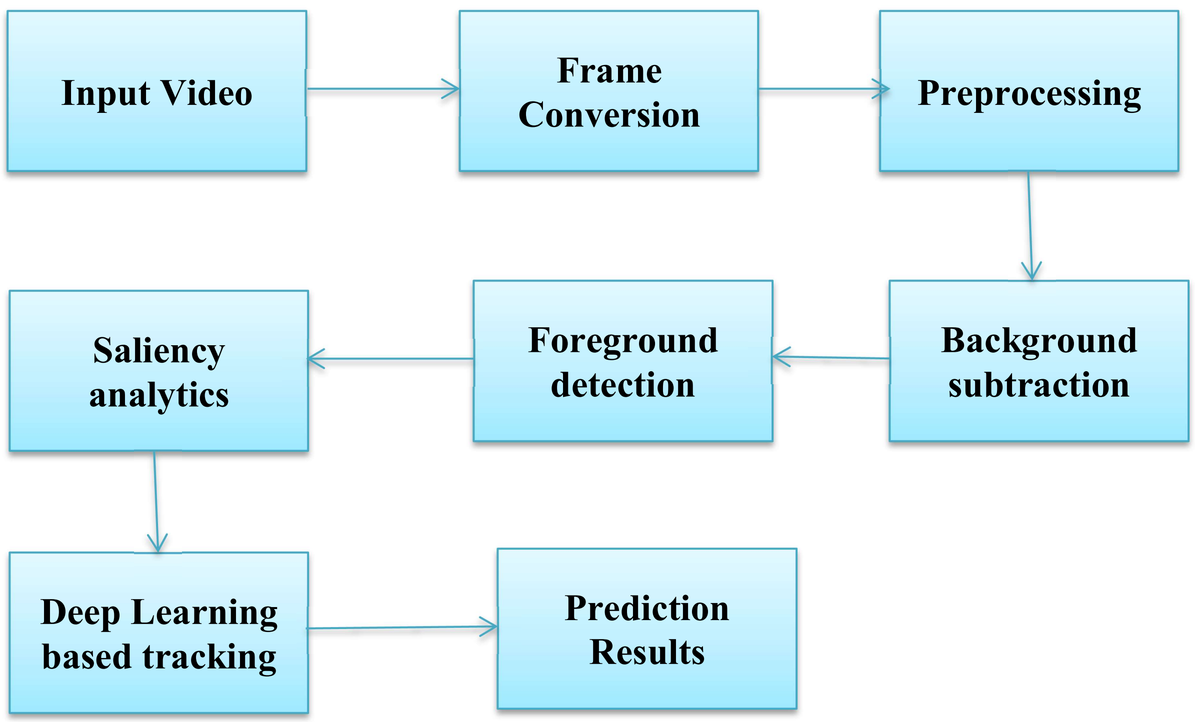


Figure 1. Block Diagram

Input Video

Even while video surveillance systems may be used to store video files, this is often a hardware device. The camera produces an image stream that acts as the input. Working with motion requires receiving a

video as an initial input for subsequent processing. It is vital to highlight that while a single image might teach us a lot of useful information, it is insufficient. CCTV video captures human behaviour in order to detect motion tracking.

Frame conversion or frame capturing

After taking the input video, this video must be transformed into frames in this phase. Framing is the process of dividing different human acts into discrete frames. It quickly determines how many motions are present.

Original video streams are modified frame-by-frame so that data may be sent along a specified path. Rather than transmitting individual ones and zeros, binary digits or bits are transmitted over an eight-bit channel, which is the size of a byte. Character streams, also known as the American Standard Code for Information Interchange (ASCII) standard format, are used to represent the majority of characters in a natural language alphabet. The following is a description of the procedure:

Set the coordinate frame for each pixel point (i, j)

$$FDn(i,j) = |fn(i,j) - fn-1(i,j)| \quad (1)$$

$$if\ FDn(i,j) < V \quad (2)$$

$$FDn(i,j) = 0 \quad (3)$$

Preprocessing

A dataset can be preprocessed to remove all of the data's noise. It collects all of the background noise and turns it into useful information. As an example, if two individuals are strolling in front of a tree, the tree is regarded to be their backdrop. When the backdrop is removed, calculations can be far more exact.

Background subtraction and fore ground extraction

In stable situations, the background subtraction method recognises the target, but it is sensitive to dynamic environments that alter due to illumination and extra events. A morphological procedure is necessary after raw subtraction and thresholding have been completed.

In the context of PCA, outliers and significant errors in the datasets are a given. Because of this, efforts have been made to construct resilient PCA versions, and the name RPCA has been used to identify a number of remedies to this issue. Recently, robust versions of PCA have been one of the most active research lines pertaining to Techniques linked to the use of PCA to deal with extremely large datasets in domains including image processing, machine learning, bioinformatics and web data analysis.

$$min_{L,S}||L|| * +\mu||S||1 \quad (4)$$

There are two types of data decomposition: low-rank (L) and sparse (S). Matrix components that minimise the linear combination of two different component norms were recognised as a convex optimization task. is the S matrix norm.

Foreground detection using MFTSL (Multi-Feature Tensor Subspace Learning). If the pixel is a present item or a background function, the maximum resolution must be provided. Up to the day of the test, the test runs the majority of the computations and then chooses one of the best outcomes. Numerous functional systems for identifying acceptable Entre Rios thresholds based on laws include extension approaches, procedures, and, at the very least, more extreme courses in square light.

$$T\,(m,n) = q(m,n) + k\sigma\,(m,n) \quad (5)$$

Where q (m, n) is the pixel's local mean (m, n) (m, n) = pixel standard deviation (m, n) The standard deviation influence is controlled by the weight k. Distance rejection, as well as global and local standards, are examples of variations. Let F0 (a, b) be the starting frame and Gref (a, b) be the reference frame, both of which have no foreground object. Both the reference backdrop frame and the FD separate non-static pixels from static pixels. A thresholding technique is used to perform the FD action. Alternatively, the image's pixels are compared to a predetermined threshold. The difference between the previous frame Ft1 (a, b) and the current frame Ft is used to identify the stationary set of pixels (a, b). Gref (a, b) is defined as the background reference frame;

$$G_t^{fd}\,(\mathrm{a,b}) = \begin{cases} \mathrm{Gref\,(a,b)\,,if\,|Ft\,(a,b)\,-\,Ft-1\,(a,b)|\,<\,\tau 1} \\ \mathrm{Gref\,(a,b)\,\times} \\ \mathrm{sgn\,(Ft\,(a,b)\,-\,Ft-1\,(a,b))\,,Otherwise} \end{cases} \quad (6)$$

Here, Gfd t (a, b) revealed as static pixels by frame difference method and τ1 denoted as the threshold value.

Algorithm

Input image frame

Low rank and sparse matrix decomposition $B_S = L_M + S_M$

L_M: Low rank matrix; S_M: Sparse Matrix

Energy minimization $Rank(L_M) + \tau||S_M||$

τ: *arbitrary balancing parameter.*

Sparse matrix B_{sm} another low rank matrix

$A_{sm} = e_i e_j^T$ and decomposition is $B_{s1m} = B_{sx} + e_i e_i^T, A_{s1m} = 0$.
Low rank Matrix A_{lm} and $B_{lm} = A_{lm}^1 e_1^T$. A_{lm}^1 : first column of A_{lm}. $A_{l1m} = A_{lm} + B_{lm}$ and $B_{l1m} = 0$.
Incremental SVD(φ_T, P^n, S^n)
If t=0 then
For i=1:n
$[X_t^n\ \varepsilon_t^n\ Y_t^n] \leftarrow SVD(\varphi_t^n, P^n, S^n)$
End
End for
Else
For i=1:n
$[X_t^n\ \varepsilon_t^n\ Y_t^n] \leftarrow iSVD(\varphi_t^n, P^n, S^n, X_{t-1}^n\ \varepsilon_{t-1}^n\ Y_{t-1}^n)$
End for
End

Saliency analytics

When navigating through visually busy situations, the human vision system is able to swiftly concentrate on critical regions for further processing while ignoring unnecessary information. It is essential to imitate these systems so that robots can interpret visual material in the same manner as humans. There are a number of uses for saliency detection that include summarising images and videos; human reidentification; video surveillance; object recognition; and image retrieval, among others.
A salient region is a part of an image that stands out from its surroundings and grabs the viewer's attention. The measure or quality with which a significant object or pixel in an image can be distinguished from its surrounds defines it. The moving foreground object in the video is the most important thing in the context of the application under discussion. The "image signature," a binary and holistic picture descriptor, was used to identify key elements in the image. The sign function of the Discrete Cosine Transform of an image (DCT). The salient image for further processing was computed using this method.
Assume the following deconstruction of grayscale images:
$y = e + a$ (7)
Where y, e, a ϵ Sm
The image signature is defined as:
$y' = DCT(y)$ (8)
Given an image that can be decomposed as in (7), the sign of the mixed signal y in the transformed domain can be taken as the support of e, and the reconstructed picture in spatial domain can be computed using inverse DCT.
$n = h * (y.y)$ (9)
Where h is the Gaussian kernel.

Deep learning based tracking algorithm

Machine learning has a subclass known as "deep learning" that is based on the way basic neural networks define depth. As a result of the interconnectedness of the deep neural network's layers (see Fig. for additional details), it is difficult to handle high-dimensional inputs. Consequently, if they do not have access to the proper dataset or processing resources, they may encounter problems like as over-fitting and be unable to build a model capable of making real-time judgments. In general, we used 3D filters to convolution the n-frame sequence in Figure 5. In order to recognise crowd behaviour, component filters are then utilised to concatenate k feature maps for classification categories such as anomalous event detection.
This study examined a variety of deep convolutional neural network architectures and configurations. This study studied the proper function for distinguishing security monitoring frames from natural frames recorded in the case of a person's aberrant state by adding luggage or other suspicious items. The network learns by the use of data disparities across successive frames. Their primary focus is on crowd surveillance and automated analysis of an uncontrolled environment, generally referred to as target detection or the individual effect of mobility in public places.
CNN is a form of artificial neural network developed to simulate human visual processing and suited for data processing. It is destined to become the next generation of fundamental classification algorithms, outperforming prior deep learning architectures in both video and audio signal processing.
The convolution procedure is performed on each phase's feature map by the convolution layer. Convolution utilises both the element-by-element multiplication of the feature map and the convolution mask. Using past knowledge, the learning samples are learnt in the right format, and the line masks are utilised differently for each layer. Each layer is assigned its own convolution masks. During the learning phase, the convolution process weight is determined. Equation 10 illustrates how convolution is calculated.
$x_{ab} = \sum_{j=0}^{i-1} \sum_{k=0}^{i-1} y_{a+j,b+k} * w_{jk}$ (10)
K, the convolution mask's size, is represented in Equation 10 by the letters a and b. Weighing and

input values are represented by letters w, y and x. Subsampling uses a mask to identify the largest or smallest value in the mask and estimates the mask's mean value.

Results and Discussion

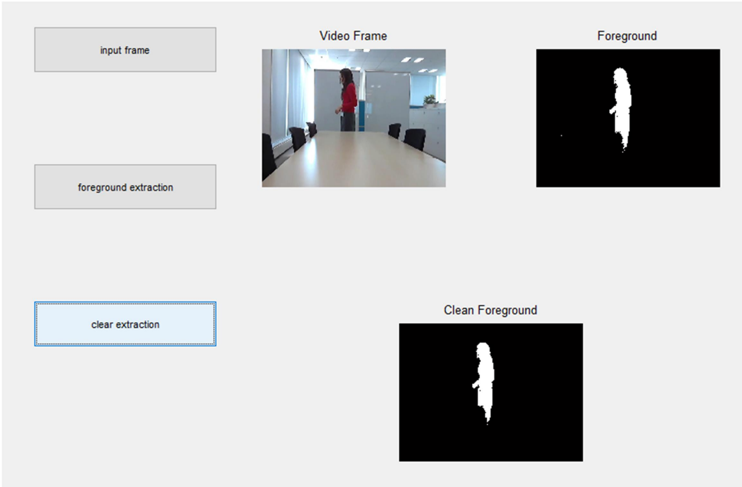


Figure 2. Moving object detected inn frames

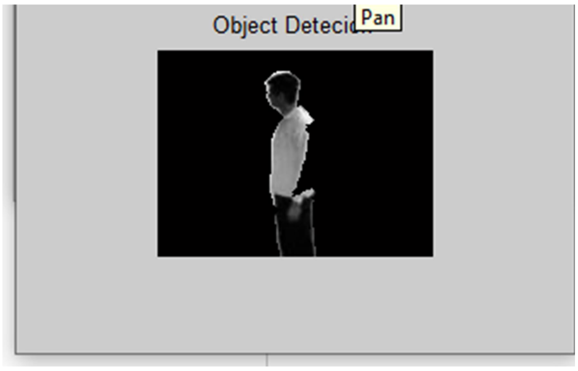


Figure 3. Moving object tracking with proposed methodology

The suggested numan besing moving object recognition and identification system is evaluated using accuracy, precision, and the f-measure. The parameters are broken down into sections in this section.

The ratio of specific positive samples to all positive classification samples should be kept in mind.

$$\text{Recall} = \frac{\text{number of detected target object in all frame}}{\text{difference in ground tooth and actual objects in frame}} \quad (11)$$

Precision: The fraction of correctly classified positive samples relative to the total number of samples per class.

$$\text{precision} = \frac{\text{Number of motion object detected in all frame}}{\text{Actual motion object in frame}} \quad (12)$$

The F-measure is the harmonic mean of accuracy and recall, as shown in the following equation.

$$F_{\text{Fmeasure}} = \frac{2 * \text{precision} * \text{recall}}{\text{precison} + \text{recall}} \quad (13)$$

Accuracy: The total number of valid samples equals the total number of samples categorised using the formulae below.

$$Accuracy = (TP + TN)/(T p + Tn + F p + f n) \quad (14)$$

Table 1. Confusion matrix parameters

Video	TP	TN	FP	FN
Video1	543	73	32	121
Video2	668	165	48	167

Video3	946	135	28	225
Video4	794	67	44	179
Video5	1047	98	17	264

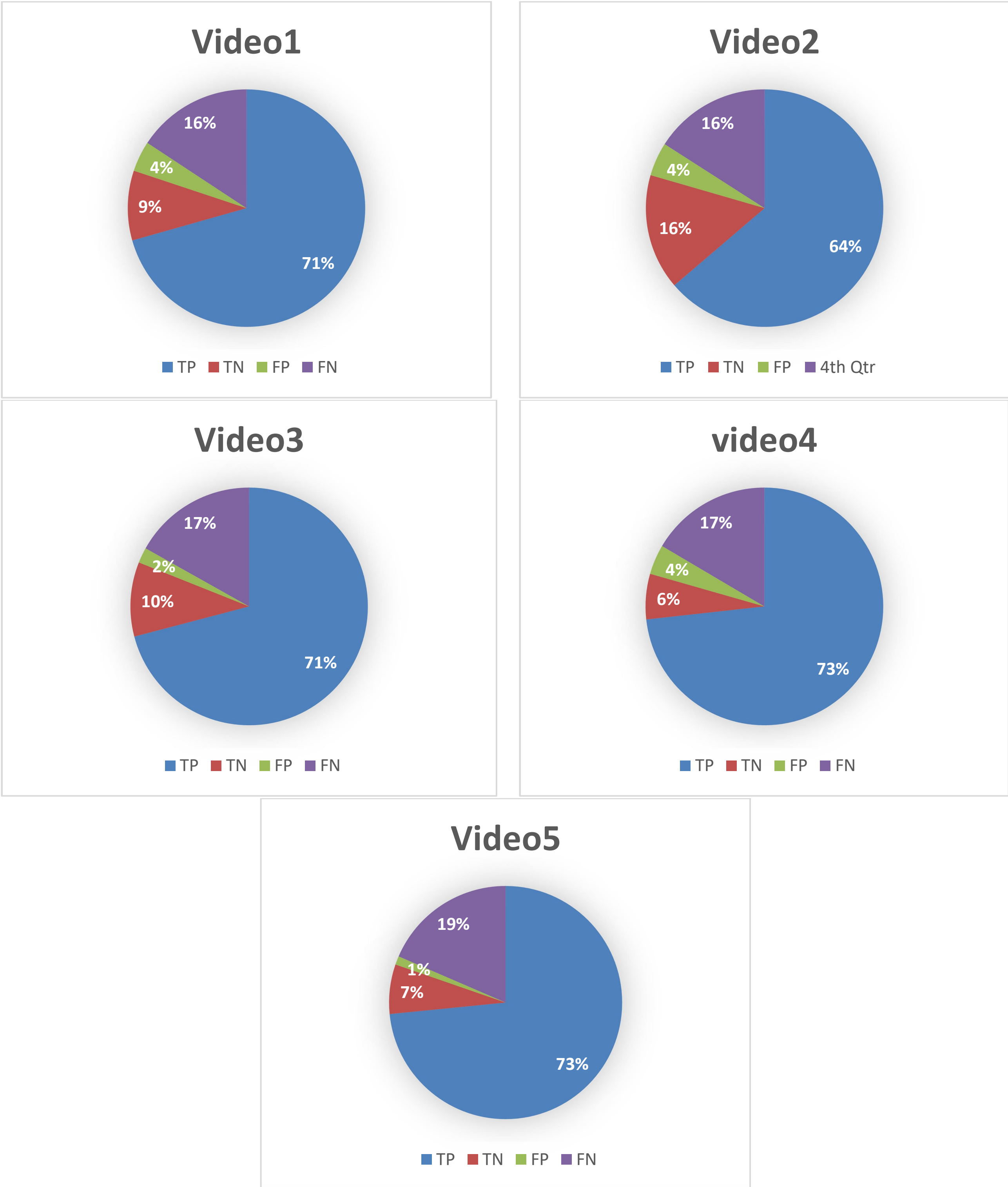


Figure 4. Confusion matrix parameters on Video1, Video2, Video3, Video4 and Video5

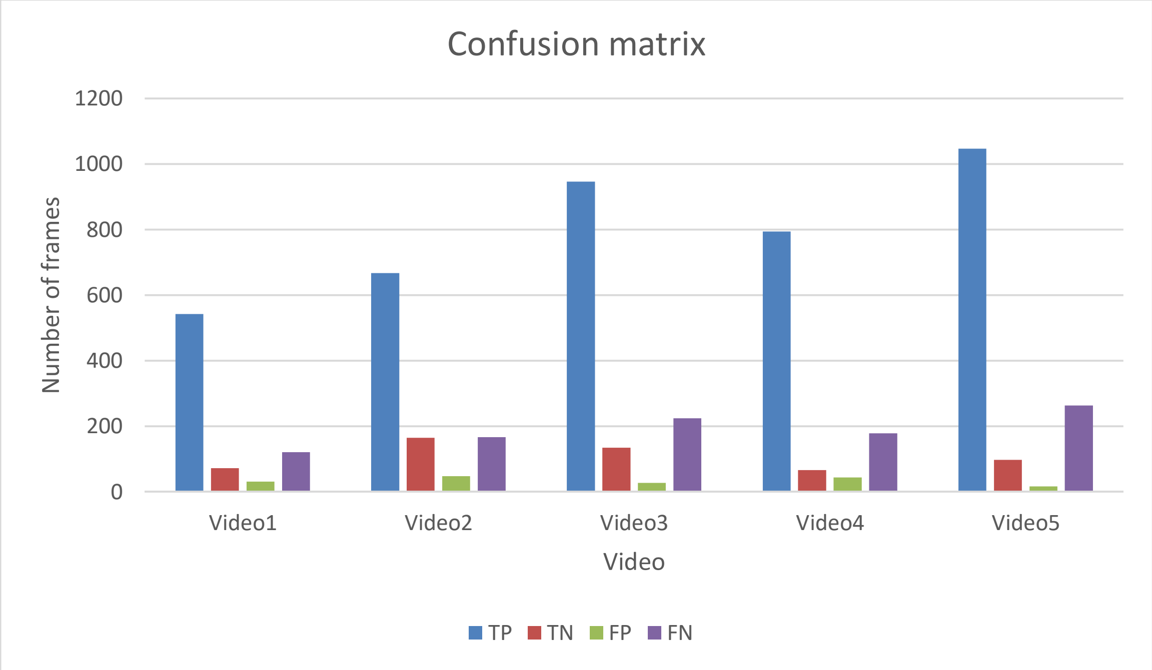
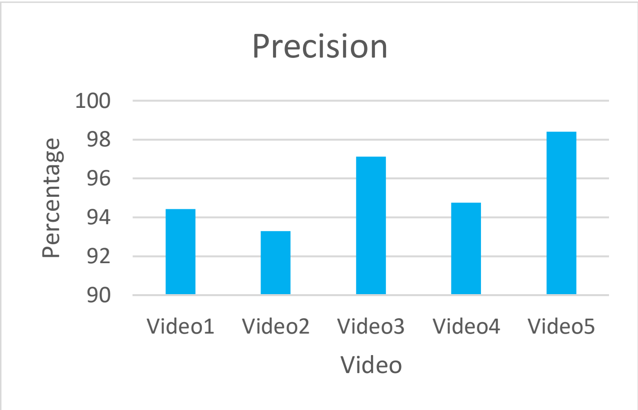
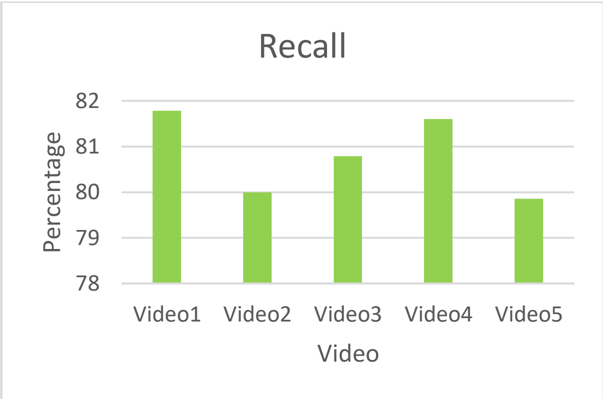


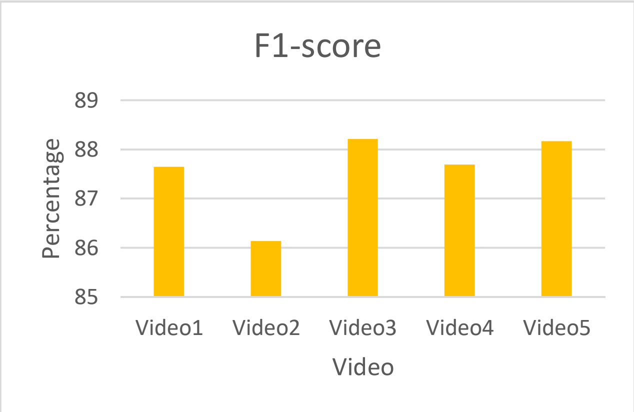
Figure 5. Confusion matrix parameters

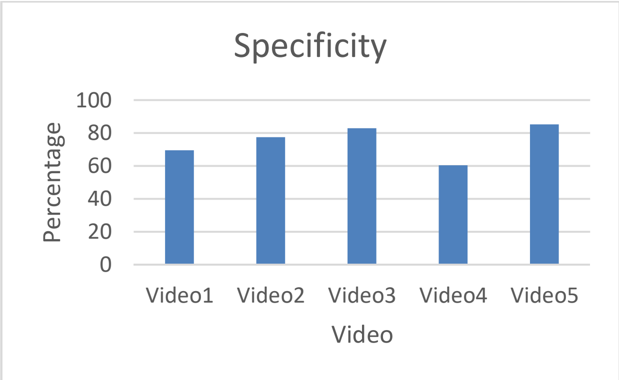
Table 2. Evaluation parameter of detected video

Video	Precision	Recall	F1-score	Accuracy	Specificity
Video1	94.43	81.78	87.65	80.10	69.52
Video2	93.30	80.00	86.14	79.48	77.46
Video3	97.13	80.79	88.21	81.03	82.82
Video4	94.75	81.60	87.69	79.43	60.36
Video5	98.40	79.86	88.17	80.29	85.22









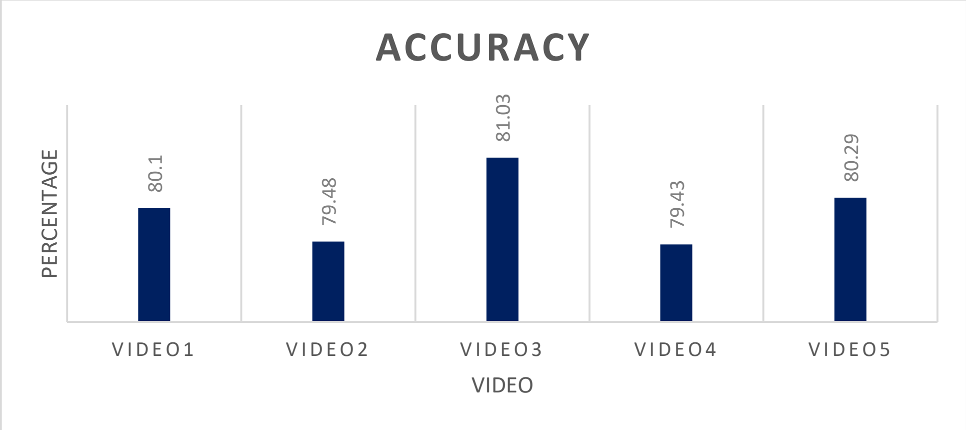


Figure 6. Precision, Recall, F1-Score, Specificity and sensitivity of different video images

Where TP = True Positive (total number of frames with actual moving object)) ;TN = True Negative (Total number of frames with Actual non moving objects);FP = False Positive (number of false moving object detection); FN = False negative (Number of flase non-moving object detected).

Table 3. Accuracy comparison of existing algorithms with proposed

Video	GMM	KNN	PCA	Proposed
Video1	68.78	72.65	78.2	80.10
Video2	65.43	69.53	72.6	79.48
Video3	48.67	73.33	74.87	81.03
Video4	71.1	65.24	71.11	79.43
Cideo5	69.38	76.66	75.43	80.29

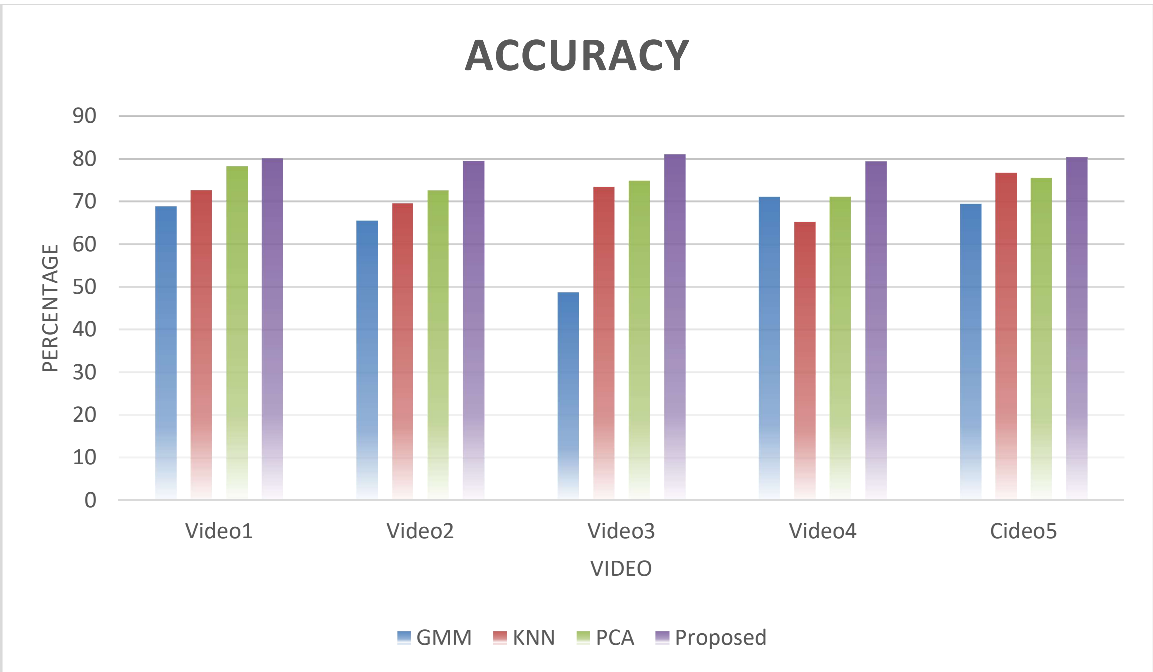


Figure 7. Accuracy comparison of Different algorithms

Videos are tested for moving object detection with some existing algorithms like GMM, KNN, and PCA and compared with the proposed algorithm. The above bar chart shown in figure6. Shows the performance of accuracy on different algorithm. Among those, our proposed method shows highest accuracy on object detection. Where GMM showing the least performance.

Conclusion

In the realm of computer vision, the automated recognition of ongoing actions and analysis of behaviour has become a popular area of study. We described a technique for detecting anomalous behaviour in surveillance footage. Deep learning has made significant advancements as a consequence of research and technological advances in deep learning. Deep learning is now being used with great success to image classification and recognition, speech recognition systems, and information retrieval systems, among other areas where it provides superior results and has greater growth potential than earlier approaches. We advocated using video monitoring to detect unusual behaviour. Aberrant event identification is a video analysis task that is still in its early stages of development, with the goal of distinguishing between abnormal and normal occurrences in surveillance footage. Additional distinguishing methods or motion information should be investigated because typical and abnormal events are similar in some aspects. We used Robust Principle Component Analysis for Background subtraction and MFTSL for foreground detection. We used CNN based tracking algorithm for Analytics. Accuracy of estimation of moving objects is high in our proposed system. Proposed methodology increases accuracy of average and high speed motion in surveillance video. The proposed method should be used to construct a surveillance system that can detect several objects and forecast movement in the future.

References

Thenmozhi, T., & Kalpana, A. M. (2020). Adaptive motion estimation and sequential outline separation based moving object detection in video surveillance system. Microprocessors and Microsystems, 76, 103084.

Vishwakarma, S., & Agrawal, A. (2013). A survey on activity recognition and behavior understanding in video surveillance. The Visual Computer, 29(10), 983-1009.

Rueda, H. F., Polania, L. F., & Barner, K. E. (2012, May). Robust tracking and anomaly detection in video surveillance sequences. In Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications IX (Vol. 8360, pp. 127-138). SPIE.

Kalli, S., Suresh, T., Prasanth, A., Muthumanickam, T., & Mohanram, K. (2021). An effective motion object detection using adaptive background modeling mechanism in video surveillance system. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-13.

Dhiman, C., & Vishwakarma, D. K. (2019). A review of state-of-the-art techniques for abnormal human activity recognition. Engineering Applications of Artificial Intelligence, 77, 21-45.

Ovhal, K. B., Patange, S. S., Shinde, R. S., Tarange, V. K., & Kotkar, V. A. (2017, December). Analysis of anomaly detection techniques in video surveillance. In 2017 International Conference on Intelligent Sustainable Systems (ICISS) (pp. 596-601). IEEE.

Shi, Y., Liu, Y., Zhang, Q., Yi, Y., & Li, W. (2016). Saliency-based abnormal event detection in crowded scenes. Journal of Electronic Imaging, 25(6), 061608.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.

Bhandari, S. H., & Babar, N. S. (2018, April). Abnormal Activity Recognition Using Saliency and Spatio-Temporal Interest Point Detector. In International Conference on Soft Computing Systems (pp. 291-301). Springer, Singapore.

Wang, Y., Zhang, Q., & Li, B. (2016, March). Efficient unsupervised abnormal crowd activity detection based on a spatiotemporal saliency detector. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1-9). IEEE.

Ayazoglu, M., Sznaier, M., & Camps, O. I. (2012, June). Fast algorithms for structured robust principal component analysis. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1704-1711). IEEE.

Lei, Z., Wang, C., Wang, Q., & Huang, Y. (2009, March). Real-time face detection and recognition for video surveillance applications. In 2009 WRI World Congress on Computer Science and Information Engineering (Vol. 5, pp. 168-172). IEEE.

Rezaee, K., Rezakhani, S. M., Khosravi, M. R., & Moghimi, M. K. (2021). A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. Personal and Ubiquitous Computing, 1-17.

Liu, Y., Wang, P., & Wang, H. (2018, November). Target tracking algorithm based on deep learning and multi-video monitoring. In 2018 5th International Conference on Systems and Informatics (ICSAI) (pp. 440-444). IEEE.

Ahn, H., & Cho, H. J. (2019). Research of multi-object detection and tracking using machine learning based on knowledge for video surveillance system. Personal and Ubiquitous Computing, 1-10.

Khot Harish, S., Gote Swati, R., & Khatal Sonali, B. Smart Video Surveillance.

Bonanno, G. (1987). Location choice, product proliferation and entry deterrence. The Review of Economic Studies, 54(1), 37-45.

Suseendran, G., Akila, D., Balaganesh, D., Elangovan, V. R., & Vijayalakshmi, V. (2021, January). Incremental Multi-Feature Tensor Subspace Learning Based Smart Traffic Control System and Traffic Density Calculation Using Image Processing. In 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 87-91). IEEE.