RES MILITARIS Social Science Journal

Surveillance Video Anomaly Detection and Recognition through Kernel Local Component Analysis and Deep Learning Classification

Priti Singh, Hari Om Sharan and C.S. Raghuvanshi

Faculty of Engineering and Technology, Rama University, Mandhana, Kanpur, Uttar Pradesh , India

preetirama05@gmail.com

Abstract

As global populations increase, the analysis of crowds garners significant attention from both social and technological perspectives. The diverse nature of crowd behavior poses challenges in assessment, particularly in complex and unpredictable environments where manual monitoring falls short, posing risks to public safety and security. Various methods for detecting abnormal behaviors aim to address issues such as execution time, computational complexity, efficiency, robustness against occlusion, and generalizability. This study introduces an innovative approach to human activity-based anomaly detection and recognition in surveillance videos, leveraging deep learning (DL) techniques. Video input undergoes preprocessing for noise reduction and smoothing, followed by feature extraction using kernel local component analysis to monitor human activities. Subsequently, Bayesian network-based spatiotemporal neural networks classify the extracted features, revealing anomalies within the surveillance video dataset. Simulation results across different crowd datasets demonstrate performance metrics including mean average error, mean square error, training accuracy, validation accuracy, specificity, and F-measure. The proposed methodology achieves an MAE of 58%, MSE of 63%, training accuracy of 92%, validation accuracy of 96%, specificity of 89%, and F-measure of 68%.

Keywords: Human Crowd, Behaviour Monitoring, Anomaly Detection, Local Component Analysis, Bayesian Network**.**

1. Introduction:

Over the past few years, there has been a noticeable rise in both crime and terrorism. Video surveillance has become a crucial weapon in the fight against crime and violence, particularly in crowded places and at events. In public and private spaces, video cameras have been deployed more frequently in recent years. Conversely, video quality has never been better due to technological advancements, but at the expense of more computational work. It is hard to analyze such large amounts of data manually. Hence automation processing becomes crucial. In this setting, automatic video surveillance becomes a research area that draws much interest. Numerous studies have been conducted in areas like face identification and face recognition for harmful objects. Automatic video monitoring, which includes studying crowd behaviour, is getting more and more attention [1].

Due to the extensive use of video surveillance techniques, it is now difficult, time-consuming, and ineffective to manually evaluate the massive amounts of video data collected from crowd-monitoring CCTV cameras. It requires a workforce and constant attention to determine if the captured behaviours are normal or aberrant. In order to accurately identify and detect anomalies in crowd scenes, surveillance systems must have an automatic anomaly detection functionality [2]. Rapid and automatic detection of anomalous behaviors in crowded contexts is crucial for enhancing safety, reducing hazards, and ensuring speedy reactions [3]. Anomaly detection aims to quickly and automatically find anomalies. Intelligent monitoring systems are becoming essential for efficient crowd control. It is impossible to describe abnormal behavior precisely because it depends so heavily on the standards established in the environment under consideration. For instance, turning around the Kaaba in a clockwise direction is not normal behavior. In general, congestion is categorized as abnormal. Sometimes it's a security challenge, which means that something unexpected has happened to cause an action. The development of WoT is linked to Web-based technologies for computer vision and machine learning that facilitate quick hybrid decision-making.

Additionally, WoT is gaining more control over our way of life, making it easier to get things done. The W3C's WoT criteria are standards for identifying interoperability issues with various Internet of Things (IoT) application domains and guiding principles [4]. Additionally, WoT Description is crucial to the game's building components. According to this concept, a Thing Description aids in realizing WoT as a tangible or intangible object. As a result, the model's information is deemed a Thing based on semantic language and a serialization based on JavaScript Object Notation (JSON). The literature has numerous computer vision methods for tracking, identifying, and analysing crowd behaviour. Researchers have presented several methodologies and ways to study crowd dynamics to provide a safe and secure environment and prevent crowding, riots, terror attacks, etc. Due to significant occlusion and complicated background scenarios, typical computer vision techniques are not usable in crowded environments [5]. The contribution of this research is as follows:

1. To propose a novel technique in human activity-based anomaly detection and recognition by surveillance video using deep learning techniques.
2. Video features are extracted for human activity monitoring by kernel local component analysis.

3. Extracted features are classified using Bayesian network-based spatiotemporal neural networks, and this classified output gives the detected anomaly based on monitoring and recognition.

2. Related works:

In the past two decades, numerous algorithms and methods for enhancing the accuracy and speed of operations have been created. The three most prevalent methods are individual feature-based, flow field-oriented, and spatiotemporal feature-driven approaches. Crowd behaviour in the first type is frequently viewed as an amalgamation of distinct actions each crowd entity takes [7]. For instance, a group of individuals travelling in the same direction can be used to identify crowd flows along a crowded route. These methods mainly focused on identifying, isolating, and analyzing each crowd member to describe the crowd properties. In [8], they used track-lets to acquire comprehensive trajectories for activity detection or tracking. The analysis and clustering of semantic regions in unstructured spaces are proposed using a variety of trackless-based approaches [9]. Unique to find "abnormal events in crowded scenes "online dictionary learning" technique to attain an "optimized dictionary based on the optimal dictionaries." It is obtained using the "Scale Invariant Feature Transforms (SIFT)" technique [10]. An improved process, the SIFT technique lowers the count of optical flow points [11], requiring less computation for computing HMOFP. Work [12] has developed a unique method for detecting anomalous occurrences by modeling the spatiotemporal distribution of crowd movements. Authors in [13] have treated anomaly detection as a "Maximum A Posteriori (MAP)" issue, in which the knowledge is derived via background subtraction with the help of "Robust Principal Component Analysis (RPCA)," and the probability is calculated by a "trained global maximum grid template" [14]. It's worth noting that prior understanding isn't restricted to only the "foreground binary map"; it may be replaced with other successful techniques. For anomaly identification, a simple motion feature is utilized, which is similar to the "Histogram of Optical Flow (HOF)" [15]. In recent decades, crowd surveillance has received the most attention; numerous talks and research projects have been conducted [16,17]. The most recent literature focuses on machine learning applications in crowd surveillance, such as abnormal event detection, object tracking, person tagging, etc. An agent-based model [18], a flow-based model [19], and a particle-based model are the standard models for crowd behavior. [20] has presented partial differential equations to describe the dynamics of a crowd in order to portray pedestrians as a continuous density field in space.

3. System model:

In this section, the novel technique for human activity detection with recognition and monitoring for anomaly detection utilizing DL methods. Input has been collected as video and processed for noise removal and smoothening. Then these video features are extracted for human activity monitoring by kernel local component analysis. Then the extracted features are classified using Bayesian network-based spatiotemporal NN, and the classified output shows detected anomaly. The proposed architecture is represented in figure-1.

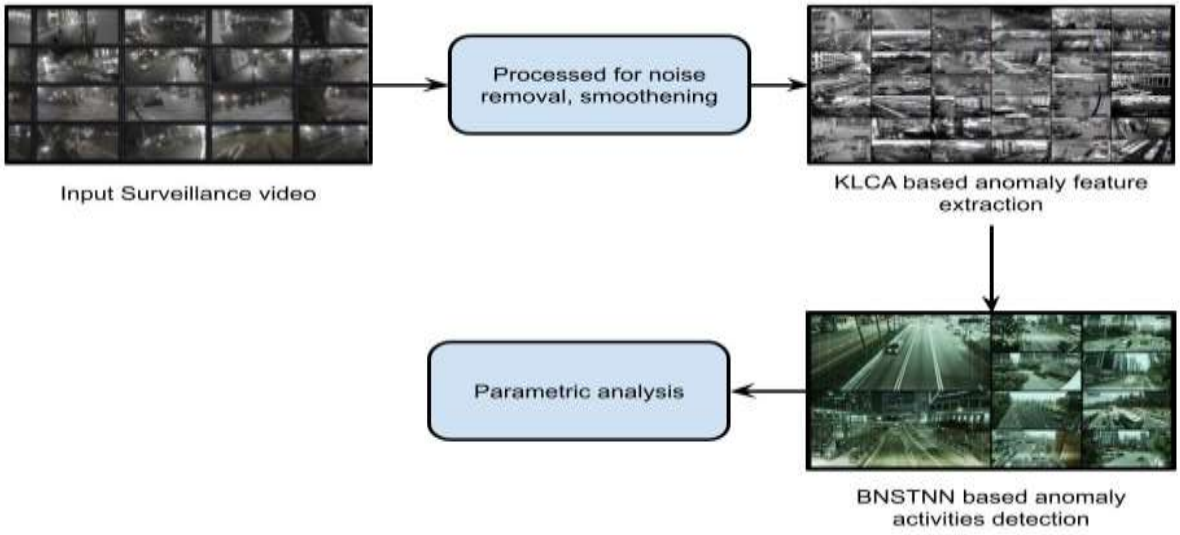

Figure 1: Proposed architecture

Crowd surveillance's primary goal is to find and report any unusual activity or occurrence in the region being watched. The occurrences of unexpected events in irregular behavior are known as abnormal activities or events in a film. An effective video surveillance system should be able to locate and recognize an object. In our situation, the object is a human being in the video, which is used to recognize the object's action. Even the difficult ones, like bikes, automobiles, cats, dogs, packages, etc., can be considered objects. An object can also be a collection of people. These discovered objects are tracked and analysed in further detail in the following frame. To follow the items across various camera feeds starting at the chosen time. Motion segmentation aims to categorize pedestrian activities in a crowd clearly and consistently. We may also employ our trajectory-level behavior features to segment motion patterns. Typically, spatial regions on an image or video are divided up using motion pattern segmentation techniques based on how similarly the pedestrians move.

We employ the K-means data-clustering technique to organize the behavior characteristics of the trajectory features obtained inside a specific time window. We exclude data collected before a specific threshold time or earlier frames because we focus on temporal local behavior analysis. These characteristics are grouped into K groups of flows that we refer to as behavior clusters. Equation (1) is used to compute a set of behavior clusters B = B1, B2,..., BK:

$$argmin_B \sum_{k=1}^{K} \sum_{b_i \in B_k} dist\,(b_i, \mu_k) \quad (1)$$

μk is the centroid of each cluster, **bi** is a behavior feature vector, and dist(bi, k) is the distance between arguments. In our example, eq(2) is used to calculate the distance between 2 pedestrian feature vectors

$$dist\,(b_i, b_j) = c_1 \parallel p_i - p_j \parallel + c_2 \parallel (p_i - v_i^{avg} wdt) - (p_j - v_j^{avg} wdt) \parallel + c_3 \parallel g_i - g_j \parallel \quad (2)$$

This is a weighted average of separation between 3 points: the current position, the prior position, and the future position. Weights c1, c2, and c3 are given. Based on the direction of centroid's velocity components, eight colors represent each behavior cluster. We demonstrate that the segmentation derived from the sparse data corresponds to the general crowd behavior patterns. Our algorithm's clustering computation throughout each frame only requires a few tens of milliseconds of processing time.

Kernel Local Component Analysis-Based Feature Extraction:

Here, we adapt these findings to our data-based Kernel Principal Component Analysis (KPCA). Traditional linear PCA can be produced in an HD space by utilizing PCA with a kernel function.

Self-adjoin operator for trace class G is implemented in Hilbert space H.

T ∈ B (H) is said to be tumor trace class if and only if series$\sum_{set} \langle \psi_i, |T| \psi_i \rangle, with\ |\underline{T}| = \sqrt{T^* T}$, is convergent for some ONBψ_i. In this case from eq.(3),

$$tr\,(T) := \sum \langle \psi_i, T \psi_i \rangle$$

$$c_1 \parallel f \parallel^2 \leq \sum_{\alpha \in A} |\langle h_\alpha, f \rangle|^2 \leq c_2 \parallel f \parallel^2 \quad for\ all\ f \in H \quad (3)$$

Let $(h_\alpha)_{\alpha\in A}$ be a frame in H. $Set\ L: H \to l^2$, then by eq.(4)

$$L: f \leftrightarrow \langle h_\alpha, f\rangle_{\alpha\in A} \tag{4}$$

Then $L*: l^2 \to H$ given by eq.(5),

$$L^*\big((c_\alpha)\big) = \sum_{\alpha\in A} c_\alpha h_\alpha \tag{5}$$

where $(c_\alpha) \in l^2$; and by eq.(6)

$$L^*L = \sum_{\alpha\in A} |h_\alpha\rangle\langle h_\alpha| \tag{6}$$

A function K : S × S → C is a positive definite (p.d.) kernel on S if, from eq (7)

$$\sum_{i,j=1}^{N} c^{-}_i c_j K\big(v_i, v_j\big) \geq 0 \tag{7}$$

for all | $\{x_i\}_{i=1}^N \subset S, \{c_i\}_{i=1}^N \subset C$, and $N \in N$

Given a p.d. kernel, a mapping : S H (K) such that from eqn (8), a reproducing kernel Hilbert space Φ : S → H (K) and a p.d (8)

$$K(x, y) = \langle \Phi(x), \Phi(y)\rangle_{H(K)} \tag{8}$$

The function Φ in the issue is referred to as a feature map. The replicating quality listed in equation (9) also includes:

$$f(x) = \langle K_x, f\rangle_{H(K)} \tag{9}$$

$$span\ \{K_x \coloneqq K(.,x) \tag{10}$$

H (K)-inner product is given by eq.(11)

$$\langle \sum c_i K_{x_i}, \sum d_j K_{x_j}\rangle_{H(K)} := \sum \underline{c_i} d_j K\big(x_i, x_j\big) \tag{11}$$

Consider a dataset xi where each xi is a D-dimensional vector and i = 1, 2, , N. The data will now be projected into an M-dimensional subspace, where M is the data's dimension. Consider the projection, which is represented by the equation y = Ax, where A is represented by [u T 1, u T , u T M], and u T k uk = 1 for k = 1, 2, u T , M. We're looking for eq(12,13) as a result

$$A^* = arg_A = \frac{1}{N}\sum_{i=1}^{N} \left(y_i - \underline{y}\right)\left(y_i - \underline{y}\right)^T \quad (12)$$

$$\underline{y} = \frac{1}{N}\sum_{i=1}^{N} x_y) \quad (13)$$

Let S_x be covariance matrix of {xi}. Since $tr(S_y) = tr(AS_xA^T)$, we get eq by utilizing the lagrangian multiplier and taking the derivative. (14)

$$S_y u_k = \lambda_k u_k \quad (14)$$

Which means that u_k in an eigenvector of S_x. Now x_i is given as eq. (15)

$$x_i = \sum_{k=1}^{D} (x_i^T u_k)u_k \quad (15)$$

x_i is approximated by eq. (16)

$$x_i = \sum_{k=1}^{M} (x_i^T u_k)u_k \quad (16)$$

Where u_k is eigenvector of S_x based on kth largest eigenvector.

The kernel principle components that arise can then be computed using eq. (17)

$$y_k(x) = \phi(x)^T v_k = \sum_{i=1}^{N} a_{ki} k(x_i, x_j) \quad (17)$$

We know that kernel takes the shape according to Mercer's hypothesis for kernels by eq. (18).

$$k(x_i, x_j) = \phi(x)^T \phi(x_i) \quad (18)$$

Multiply both sides of the equation by φ(xl), eq. (19)

$$\frac{1}{N}\sum_{i=1}^{N} \phi(xl)\phi(x_i) \sum_{i=1}^{N} a_{kj}\phi(x_i)\phi(x_j) = \lambda_k a_{ki}\phi(xl)\phi(x_i) \quad (19)$$

Which we can re-written as eq. (20)

$$\frac{1}{N}\kappa(x_l, x_i) \sum_{i=1}^{N} a_{kj}\kappa(x_i, x_j) = \lambda_k \sum_{i=1}^{N} \kappa(x_l, x_i) \quad (20)$$

$$K = \kappa(x_i, x_j)$$

AK is an N-dimensional column vector and is aki's eigenvector. By using eq(21), ak is solved

$$Ka_k = \lambda_k = Na_k \quad (21)$$

eq(22) provides the kernel principal component transformation that results

$$\hat{x} = \phi(x)^T u_k = \sum_{i=1}^{N} a_{ki} k(x_i, x_j) \quad (22)$$

In light of an uncentered kernel matrix, we calculate the kernel's zero means by eq (23)

$$\phi(x_i) = \phi(x_i) - \frac{1}{N}\sum_{i=1}^{N} \phi(x_j)$$

$$k = \|\phi(x_i) - \tfrac{1}{N}\textstyle\sum_{i=1}^{N} \phi(x_j)\|_2 - (\phi(x_i) - \tfrac{1}{N}\textstyle\sum_{i=1}^{N} \phi(x_j))^T(\phi(x_i) - \tfrac{1}{N}\textstyle\sum_{i=1}^{N} \phi(x_j)) \quad (23)$$

After expansion, we have that by eq. (24)

$$= k_{ij} - \phi(x_i) - \tfrac{1}{N}\textstyle\sum_{i=1}^{N} \phi(x_j)\tfrac{1}{N} - \tfrac{1}{N}\phi(x_i) - \tfrac{1}{N}\textstyle\sum_{i=1}^{N} \phi(x_j) - \tfrac{1}{N}\phi(x_i) - \tfrac{1}{N}\textstyle\sum_{i=1}^{N} \phi(x_j) \quad (24)$$

It can be rewritten in short as eq. (25)

$$K = K - 1, J - K1 + 1K1 \quad (25)$$

K is referred to as the Gram matrix where K=K ij (normalized kernel matrix). Despite numerous advancements in human activity analysis, it is still difficult to comprehend and keep track of human activity. Due to its individualized nature, abnormal behaviour is classified in several different ways. There has been much confusion caused by it in the literature. Some studies use frequency to describe the anomaly. "Abnormal" or "rare" refers to an infrequency incident. A controlled crowd is simple to evaluate, but an unstructured crowd is much riskier because of its erratic movement. When analyzing behavior, we primarily pay attention to flow rates, irregular occurrences like running and fighting, and velocity. Again, validating human activity is a difficult issue since it is difficult to determine and obtain ground truth video material showing certain anomalous crowd behaviours.

The qualities of a scene and the underlying application are often considered while determining the temporal window's length. Small time windows are useful for collecting details in dynamically changing situations with many rapid velocity changes brought on by some agents moving swiftly. Scenes with less varied pedestrian movement benefit from

larger periods, which tend to smooth out abrupt changes in motion. In our case, we keep the time frame in our applications between 0.5 and 1.0 seconds.

Bayesian Network-Based Spatiotemporal Neural Networks-Based Classification:

Anomaly detection is an important issue that has received attention from researchers in many different fields and applications. It relates to detecting pedestrians, events, or observations in a crowd dataset that don't fit an anticipated pattern or correspond to other pedestrians. Typically, aberrant objects or agents can be found, which can result in better automatic monitoring. Our trajectory-based behavior features are mostly used for local anomaly identification. In other words, we notice a few uncommon actions only seen in the movie at specific times. These intervals can range in length from several hundred frames to the entire length of the video. Put another way, and we define an anomaly as temporally unusual behaviour. For instance, if many other people are heading in the same direction, a person's behaviour that goes against the flow of the crowds may be picked up as an abnormality at one moment but not at another time in the frame.

Bayes factor can be understood as a summary indicator of the strength of evidence data gives us supporting hypothesis H1 in comparison to hypothesis H2, which contrasts. The Bayes factor and posterior odds are the same if H1's prior probability and H2's prior probability are equal (P(H1) = P(H2) = 0.5). Let's define their conditional probability density functions as fX|Y (x|y) and fY |X(y|x). Next, by eq (26)

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} \qquad (26)$$

so that eq(27). are utilized to express Bayes' theorem for continuous variables

$$f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} \qquad (27)$$

where $f_Y(y) = \int_X f_{Y|X}(y \mid x) f_X(x) dx = \int_{-\infty}^{+\infty} f_{XY}(x,y) dx$ because of the total probability theorem. The simple naive Bayes classifier classifies an instance based on these probabilities. We arrive at this via eq(28) by applying Bayes' theorem and somewhat compressing the notation

$$P(y_j \mid x_i) = \frac{P(x_i|y_j)P(y_j)}{P(x_i)} \quad (28)$$

$$P(x \mid y_j)P(y_j) = P(x, y_j) = P(x_1, x_2, \dots, x_p, y_j) = P(x_1 \mid x_2, x_3, \dots, x_p, y_j)P(x_2, x_3, \dots, x_p, y_j) \; because \; P(a, b) = P(a \mid b)P(b) = P(x_1 \mid x_2, x_3, \dots, x_p, y_j)P(x_2 \mid x_3, x_4, \dots, x_p, y_j)P(x_3, x_4, \dots, x_p, y_j) = P(x_1 \mid x_2, x_3, \dots, x_p, y_j)P(x_2 \mid x_3, x_4, \dots, x_p, y_j) \cdots P(x_p \mid y_j)P(y_j) \quad (29)$$

This is a significant presumption broken in most real-world situations, therefore the moniker "naive." According to this presumption, by eq (30)

$$P(x_1 \mid y_j) \cdot P(x_2 \mid y_j) \cdots P(x_p \mid y_j)P(y_j)$$

$$P(x_1 \mid y_j) \cdot P(x_2 \mid y_j) \cdots P(x_p \mid y_j)P(y_j) = \prod_{k=1}^{p} P(x_k \mid y_j)P(y_j) \quad (30)$$

which we can plug into Eq. 27, and we obtain by eq. (31)

$$P(y_j \mid x) = \frac{\prod_{k=1}^{p} P(x_k|y_j)P(y_j)}{P(x)} \quad (31)$$

We can easily determine the value of the numerator for every class as well as choose one for which it is maximal. The maximal posterior rule is the name given to this rule eq. (32):

$$\hat{y} = argmax_{y_j} \prod_{k=1}^{p} P(x_k \mid y_j)P(y_j) \quad (32)$$

An NB classifier is a method that applies Eq. 32. However, a clear categorization is not always preferred. For instance, we are frequently more interested in how well a method rates examples of one class in comparison to cases of other classes when ranking tasks involve positive and negative classes. Using the total probability theory once more, we can rewrite equation (33)

$$P(y_j \mid x) = \frac{\prod_{k=1}^{p} P(x_k|y_j)P(y_j)}{\prod_{k=1}^{p} P(x_k|y_j)P(y_j) + \prod_{k=1}^{p} P(x_k|y_j^c)P(y_j^c)} \quad (33)$$

We will give an example of how to determine the general form using eq(34) approximate answer using our method

$$x^m y''(x) = F(x, y(x), y'(x)) \quad (34)$$

where the domain is denoted by $m \in Z$, $x \in R$, $D \subset R$, and y(x) represents the answer that needs to be computed. Eq. (35) converts the issue to a discretized form if (x, p) signifies a trial solution with movable parameters p:

$$Min_P \sum_{x_i \in D} F\left(x_i, y_t(x_i, p), y_t'(x_i, p)\right) \quad (35)$$

By comparing discovered STT features with the normality human activity model, the system should assign either "normality" or "abnormality" to video samples to detect crowd anomalies. The choice is made for each buffered video clip for internet use as the video is playing. STT features are taken from STV slices found at S places by an average flow field, like learning progress is made. STT feature for detecting crowd anomalies is denoted as F j = [f jk], where j = 1, 2,..., S and k = 1, 2,..., N. In this study, the 3-sigma rule of Gaussian distribution, which is determined by eq(36), is used to make a simple binary decision for every STT element $d_{jk} = \{1(\,positive)\ \tilde{f}_{jk} \in [\mu_{jk} - 3\sigma_{jk}, \mu_{jk} + 3\sigma_{jk}]\ 0\ (negative)\ \ otherwise$ (36)

For one STV slice, this operation contains N subdivisions. Equation (37) has provided a "vote" technique for reaching a "final" conclusion.

$$D_j = \{1\ \frac{1}{N}\sum_k d_{jk} > T\ 0\ \ otherwise\ . \quad (37)$$

The positive rate is then evaluated with a threshold, T, before making a final choice for the STV slice. The threshold is understood as a pass rate for the choice in the voting process. A higher pass rate indicates that more positive voters for a slice must vote for the slice to be declared ultimately positive (djk = 1). Each DJ can indicate whether a crowded event in their location is typical or abnormal. The decision-making system in the designed prototype is efficient. The algorithm takes substantially less time than the time required to buffer and play videos. The abnormal human activity can be identified using a parallel programming method before the current video segment is cleared from the buffer, ensuring that decisions are made in real-time while the video is playing. We analyze the distance between each pedestrian's local and global attributes for anomaly detection. Anomaly features often tend to be isolated in a cluster they are a part of when they first appear in a scene. The classification of anomaly detection is shown in figure-2.

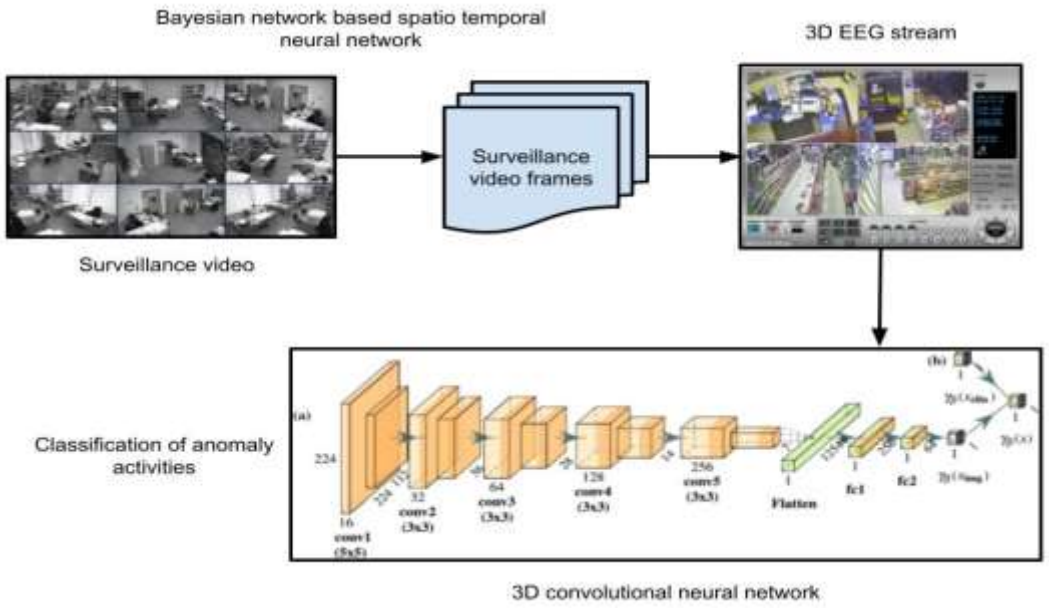

Figure-2 Classification Of Bayesian Network-Based Spatiotemporal Neural Network

In other words, the motion of the pedestrian will be distinct from that of the people around him. We categorize a feature as an anomaly by eq(38) if the Euclidean distance between the global and local features is greater than a threshold value.

$$dist\ (b^l, b^g) > \ Threshold \qquad (38)$$

The user can adjust this threshold. The sensitivity of the anomaly detection will rise if this threshold is set low, and vice versa.

Performance analysis

A PC running Windows 7 64-bit with an Intel Xeon E5-1650 [(3.60 GHz) six-core CPU, NVidia Titan XP 6 GB GPU (3840 CUDA cores), and 16 GB of memory] served as the training platform. The software tools included Microsoft Visual Studio 12.0, Python 2.7, CUDNN 7.5, and CUDA 8.1.

Dataset description

A stationary video camera with a 640 × 360 pixels resolution was used to collect the CUHK Avenue dataset while filming street activity near the City University of Hong Kong. This dataset includes 21 test video examples showing odd occurrences and human behaviour and 16 train video samples showing typical human behaviour. Persons walking on the sidewalk alone or in groups are considered regular behaviour, but people throwing things

away, loitering, walking in front of cameras, walking on grass, and abandoning things are considered aberrant behaviours. A stationary video camera with a resolution of 238 × 158 pixels focused on two pedestrian paths was used to record UCSD pedestrian Dataset [20]. UCSD ped 1 and UCSD ped 2 are two datasets that document various crowd scenarios, ranging from sparse to packed.34 train video samples and 36 test video samples make up Ped 1 dataset, whereas 16 train video samples and 12 test video samples make up Ped 2 dataset. 26 frames per second (FPS)frame rates were used to capture both of the chosen datasets.

Table-1 Analysis Of Various Human Crowd Behaviour Datasets

Dataset	Techniques	MAE	MSE	Training accuracy	Validation accuracy	Specificity	F-measure
CUHK Avenue dataset	**HMOFP**	45	51	77	81	75	55
	RPCA	48	53	79	83	79	59
	HCBM_KLCA_DLT	51	55	81	85	81	62
UCSD ped 1	**HMOFP**	47	53	79	82	77	59
	RPCA	52	56	83	84	82	62
	HCBM_KLCA_DLT	53	58	85	86	83	63
UCSD ped 2	**HMOFP**	49	55	85	85	79	61
	RPCA	55	59	88	89	85	65
	HCBM_KLCA_DLT	58	63	92	96	89	68

Table-1 shows analysis based on various human crowd behavior analyses. The CUHK Avenue dataset, UCSD ped 1, and UCSD ped 2 datasets are compared. The parameters analyzed are mean average error, mean square error, training accuracy, validation accuracy, specificity, and F-measure.

Pre-process inputs by contracting recovered edges to 224 x 224 pixels and normalizing pixel values by scaling between 0 and 1 since video tests have changed aspects. We pick profundity of the worldly cuboid, T=8, which generally compares to a surmised time of 33% of a second because of the casing pace of picked preparing data. Due to the huge profundity

of information cuboids, the decision of T should consider both boosting movement to be gotten inside sequential edges and decreasing union of DL strategy. In this examination, we prepared learning techniques with 1500 preparation ages at a learning pace of 0.01. The spatiotemporal autoencoder model is improved by utilizing a stochastic slope drop approach, and the cost capability for working out reproduction misfortune is a mean squared blunder. We utilized an early halting regularization procedure, where preparing stops when misfortune quits improving, to keep the model from becoming overfit. By isolating the informational index into 60% for the first and 20% for the second and third emphasis, the preparation was done in more than three consistent cycles. The fluffy measure in the dynamic learning stage was reproduction error.

The tracking algorithm controls the program's efficiency and precision, which can be erratic, sparse, or lose traces. Additionally, the real-time techniques used today might not be effective in movies of very dense crowds, such as those that contain thousands of agents in a single frame. While offline learning techniques may calculate numerous global properties, our online learning strategy is primarily useful for obtaining local pedestrian trajectory parameters. For instance, our motion segmentation and anomaly detection algorithms will only record unusual or uncommon behaviors seen in temporally close frames.

4. Conclusion:

The inability to manually monitor intricate and unpredictable human activity scenarios significantly impacts public space security and safety. This study introduces a technique for anomaly detection rooted in human activity analysis from video surveillance datasets. Preprocessing steps are applied to the input video, followed by feature extraction via kernel local component analysis and classification using Bayesian network-based spatiotemporal neural networks. The classified output enables the detection of anomalies within the video data. Addressing computational complexity, certain methods for human activity analysis aim to reduce processing time, facilitating faster responses to anomalous behavior. The proposed technique achieves an MAE of 58%, MSE of 63%, training accuracy of 92%, validation accuracy of 96%, specificity of 89%, and F-measure of 68%. Future prospects include integrating pedestrian dynamics features with other methodologies to capture complex crowd behaviors beyond the scope of this study.

Reference:

1) 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.

2) Sánchez, F. L., Hupont, I., Tabik, S., & Herrera, F. (2020). Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects. *Information Fusion*, *64*, 318-335.

3) Bahamid, A., &Mohd Ibrahim, A. (2022). A review on crowd analysis of evacuation and abnormality detection based on machine learning systems. *Neural Computing and Applications*, *34*(24), 21641-21655.

4) Aldayri, A., &Albattah, W. (2022). Taxonomy of Anomaly Detection Techniques in Crowd Scenes. *Sensors*, *22*(16), 6080.

5) Chaudhary, D., Kumar, S., & Dhaka, V. S. (2022). Video based human crowd analysis using machine learning: a survey. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *10*(2), 113-131.

6) Mishra, P. K., Mihailidis, A., & Khan, S. S. (2022). Skeletal Video Anomaly Detection using Deep Learning: Survey, Challenges and Future Directions. *arXiv preprint arXiv:2301.00114*.

7) Patrikar, D. R., &Parate, M. R. (2022). Anomaly detection using edge computing in video surveillance system. *International Journal of Multimedia Information Retrieval*, *11*(2), 85-110.

8) Bhuiyan, M. R., Abdullah, J., Hashim, N., & Al Farid, F. (2022). Video analytics using deep learning for crowd analysis: a review. *Multimedia Tools and Applications*, *81*(19), 27895-27922.

9) Zhang, S., Gong, M., Xie, Y., Qin, A. K., Li, H., Gao, Y., & Ong, Y. S. (2022). Influence-aware attention networks for anomaly detection in surveillance videos. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(8), 5427-5437.

10) Bamaqa, A., Sedky, M., Bosakowski, T., Bastaki, B. B., &Alshammari, N. O. (2022). SIMCD: SIMulated crowd data for anomaly detection and prediction. *Expert Systems with Applications*, *203*, 117475.

11) Shin, H., Na, K. I., Chang, J., & Uhm, T. (2022). Multimodal layer surveillance map based on anomaly detection using multi-agents for smart city security. *ETRI Journal*, *44*(2), 183-193.

12) Mohamed, A. A., Alqahtani, F., Shalaby, A., &Tolba, A. (2022). Texture classification-based feature processing for violence-based anomaly detection in crowded environments. *Image and vision computing*, *124*, 104488.

13) Chakole, P. D., Satpute, V. R., &Cheggoju, N. (2022, May). Crowd behavior anomaly detection using correlation of optical flow magnitude. In *Journal of Physics: Conference Series* (Vol. 2273, No. 1, p. 012023). IOP Publishing.

14) Khaire, P., & Kumar, P. (2022). A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments. *Forensic Science International: Digital Investigation*, *40*, 301346.

15) Ekanayake, E. M. C. L., Lei, Y., & Li, C. (2022). Crowd Density Level Estimation and Anomaly Detection Using Multicolumn Multistage Bilinear Convolution Attention Network (MCMS-BCNN-Attention). *Applied Sciences*, *13*(1), 248.

16) Bamaqa, A., Sedky, M., &Bastaki, B. (2022). Reactive and proactive anomaly detection in crowd management using hierarchical temporal memory. *International Journal of Machine Learning and Computing (IJMLC)*, *12*(1), 7-16.

17) Samani, H., Yang, C. Y., Li, C., Chung, C. L., & Li, S. (2022). Anomaly detection with vision-based deep learning for epidemic prevention and control. *Journal of Computational Design and Engineering*, *9*(1), 187-200.

18) Khan, A. A., Nauman, M. A., Shoaib, M., Jahangir, R., Alroobaea, R., Alsafyani, M., ... &Wechtaisong, C. (2022). Crowd Anomaly Detection in Video Frames Using Fine-Tuned AlexNet Model. *Electronics*, *11*(19), 3105.

19) MAJJI, V., KAKOLLU, V., KUMAR, M. S., SOUJANYA, K. N., KANTHAMMA, B., & RAO, G. B. (2022). VIDEOBEHAVIOR POSSIBLE IDENTIFICATION AND RECOGNITION OF ABNORMALITIES AND NORMAL BEHAVIOR PROFILING FOR ANOMALY DETECTION USING THE CNN MODEL. *Journal of Theoretical and Applied Information Technology*, *100*(14).

20) Halim, N. (2022). Intelligent Human Anomaly Identification and Classification in Crowded Scenes via Multi-fused Features and Restricted Boltzmann Machines.