International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Rare And Popular Event-Based Co-Located Pattern Recognition in Surveillance Videos Using Max-Min PPI-DBSCAN And GREVNN

Jayaram C V [1]*

[1]Assistant Professor, Department of Computer Science and Engineering,
Mysore College of Engineering and Management,
Mysuru, Karnataka, India.
Email: jayaramscv77@gmail.com

Dr. B K Raghavendra[2]
[2]Professor and HOD, Department of Information Science and Engineering,
Don Bosco Institute of Technology,
Bengaluru. India.
Email: bkraghavendra@dbit,.co.in

Abstract *:* Co-located pattern recognition is the process of identifying the sequence of patterns occurring in surveillance videos. In greater part of the existing works, the detection of rare and popular events for effective co-located pattern recognition is not concentrated. Therefore, this paper presents the automatic discovery of the co-located patterns based on rare and popular events in the video. First, the video is converted to frames, and the keyframes are preprocessed. Then, the foreground and background of the frames are estimated, and the rare and popular events are grouped using Maximum-Minimum Pixel-Per-Inch Density-Based Spatial Clustering of Applications with Noise (Max-MinPPI-DBSCAN). From the grouped image, the object detection and mapping are done, and the patch is extracted from it. Next, the edges are detected and from that, for the moving objects, motion is estimated by the Kullback-Leibler Kalman Filter (KLKF). Also, for non-moving objects, the objects/persons are tracked. From the motion estimated and tracked data, time series features are extracted. Then, the optimal features are selected using the Dung Beetle State Transition Probability Optimizer (DBSTPO). Finally, the co-located pattern is classified using a Generalized Recurrent Extreme Value Neural Network (GREVNN), and the alert message is given to the authorities. Hence, the proposed model selected the features in 53239.44ms and classified the event with 99.0723% accuracy and showed better performance than existing works.

Keywords-- Co-Location, Visual Pattern Mining, Canny Edge Detector (CED), Median Silhouette Filter (MSF), Faster Region-Convolutional Neural Network (FR-CNN), and Object Detection and Artificial Intelligence.

I. INTRODUCTION

The density of the crowd is determined based on the number of persons and the abandoned objects in a video frame [2]. Such objects are tracked through the stationery surveillance system or the dynamic cameras named Video Surveillance System (VSS) [5]. Based on color, motion, and shape, the abandoned objects are tracked by analyzing the video frames. The abnormal actions that did not happen in general on VSS are also known as a pattern [6],[14]. Further, the objects are recognized by acquiring the directional pattern features. Some of the popular models used for recognizing objects are the Convolution Neural Network (CNN), Recurrent CNN, and You Only Look Once (YOLO) algorithm [9],[17].

Nowadays, security is a major concern for people around the world, especially while traveling [11]. An object is declared as abandoned if its bounding box does not coincide with any person's bounding box [12]. However, object detection is efficient when it is beneficial in real-time. For recognizing the objects in real time, an edge-based detection strategy is introduced in the YOLO algorithm. Thus, the motion of the abandoned object is accurately recognized from the video [13].

In such moments, the persons used to place their baggage or luggage in transport zones, such as airports, railway stations, etc. To ensure security, the recognition of co-located objects and the respective person becomes predominant [16]. In order to detect moving objects, the background of the frame is subtracted using the Gaussian Mixture Model (GMM) to segment the foreground [19]. To specify the location of the object, the Region-based method is used in traditional CNN to analyze the object's position [24].

Recently, advanced VSSs have been developed for object detection using Artificial Intelligence techniques. Initially, the crowd behavior in public places is detected by extracting significant features [27]. Among the different methods, CNN cannot efficiently recognize objects from the video frames that contain variant illumination. So, the

Empirical mode decomposition approach is involved in CNN to analyze the orientation and phase of the object [28]. Hence, the temporal and spatial information of objects, including color, texture, and shape are identified [29].

However, none of the existing works focused on detecting abandoned objects in rare or unusual events. Therefore, in this work, the co-located patterns are recognized automatically in video streams using Max-MinPPI-DBSCAN and GREVNN techniques.

Some of the problems observed in existing works are mentioned as follows,

- The rare and popular events were not concentrated in existing works for identifying co-located patterns. Neglecting recognition of co-located objects in rare events results in inefficient pattern identification.
- The co-locating of abandoned objects is difficult among densely populated areas, such as railway stations, airports, etc [4].
- The tracking of abandoned baggage and individual behavior was not detected in time series in [21]. Thus, the respective person co-located with the object was not recognized.
- For the effective tracking of passengers and their baggage in [23] the presence of complex background, noise, occlusions, illumination variations, and ghost effects was not concentrated.
- Confusion and tracking loss occur between the passenger and their co-locating objects. This causes complexity in identifying co-locating patterns from the video.
- In surveillance videos, the features changing over time cause potential threats and challenges for recognizing co-located patterns.
- To overcome the limitations in existing works, the objectives focused by the proposed methodology are listed as follows,
- To identify co-located patterns efficiently, the rare and popular events are grouped by differencing frames using Max-MinPPI-DBSCAN.
- The moving and non-moving objects are detected using a canny edge detector for tracking the object and the corresponding co-located person over time.
- The extracted key frames are pre-processed using MSF for removing noise, regulating illumination, adjusting contrast, and suppressing the ghost effect.
- To prevent tracking loss and confusion, the detected objects are mapped by extracting patches among densely populated regions.
- The time-series features are extracted to minimize the potential threats in recognizing co-located patterns.

The rest of the paper is sorted as follows: Section 2 discusses the related works. Section 3 describes the proposed methodology and its performance is assessed in section 4. Finally, the paper is winded up with future suggestions in section 5.

II. Literature Review

Ahamad [1] suggested a hybrid video surveillance system for object detection. The different frames of video were processed using the Open-source Control Vision software. Then, the object or person was identified by the Vector Similarity Search algorithm. The performance was enhanced on account of accuracy and error rate. However, the pre-processing of video frames was not concentrated, which negatively impacted the object detection process.

Ayesha [3] developed a system to recognize objects by using Local Tri-conditional patterns in public places. The luggage was identified from the oriented gradient histogram feature of the image. Further, the deviations among every pixel along with its nearest pixels and central pixels were analyzed for detecting objects. The performance was enhanced in terms of detection accuracy and precision. But, the luggage detected using the histogram gradient of the image cannot recognize real-time objects efficiently.

Din [4] presented a model for tracking abandoned objects from surveillance video in real time. The background was modeled by using the initial video frame. Further, the motion was tracked to identify the moving objects by subtracting the background. The presented approach utilized a frame differencing technique for tracking the abandoned luggage. The performance was improved with respect to the object detection rate. However, the object was tracked simply based on its initial position, which was not efficient in tracking moving objects.

Gao [8] advanced a method for the detection of baggage in airports. For efficient detection, the baggage was detected by using the CenterNet-based hierarchical multi-object tracking framework. This method detected objects with higher mean average precision. But, the utilized CenterNet had computation complexity and overfitting risks, which degraded the object detection process.

Park [15] suggested a model to recognize abandoned objects from video surveillance. Initially, the position and area of the object were specified by the dual background system. Then, the objects were segmented using Mask R-CNN. Further, the segmented object and the specified objects were analyzed to detect the abandoned object. The model recognized objects with higher accuracy and efficiency. However, the R-CNN consumed more time and required large resources for object segmentation, which degraded the detection performance in certain cases.

Sathesh [21] presented the image-processing technique for detecting abandoned things in airports. The input video undergoes pre-processing, foreground, and background distinction using Kalman Filter (KF), and object recognition using a faster Region-based CNN (R-CNN). The suggested module recognized objects within less time. However, the utilized KF cannot recognize objects from videos containing non-Gaussian noise, which restricts the performance.

Siddique [23] explored an object detection approach for camera videos using a data augmentation method and CNN. The multiple detections of CNN were clustered by the Mean-Shift Algorithm (MSA). Further, a distance-based mechanism was introduced to co-locate the abandoned luggage with the person. The approach tracked objects with improved accuracy and

Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

precision. Yet, the adopted MSA could not cluster the high dimensional object efficiently, which degrades the detection accuracy.

Walia [25] presented a multi-cue model for tracking the object from video sequences. The significant particles were segmented from the unimportant objects by using the GMM approach. The object detection performance was analyzed both quantitatively and qualitatively from the video sequences. The presented model tracked objects with better precision. Since the GMM estimated various parameters to segment objects, it resulted in overfitting and convergence problems.

Wan [26] presented a method for detecting abnormal events in transportation systems. The redundant frames of videos were reduced and the long video was suppressed using Super frame segmentation.

.

redundant frames were not pre-processed, which affected the information retrieval about the abnormal event.

The object was represented in a time series using the selective traversal algorithm. The performance was improved in terms of precision and success rate. However, the utilized semantic network consumed more time to track objects due to the traversing of the entire video system

III. Materials and Methods

In this work, the rare and popular events in the video are grouped and the co-located pattern is classified using Max-MinPPI-DBSCAN and GREVNN, respectively. The architecture of the proposed work is given in Figure 1.

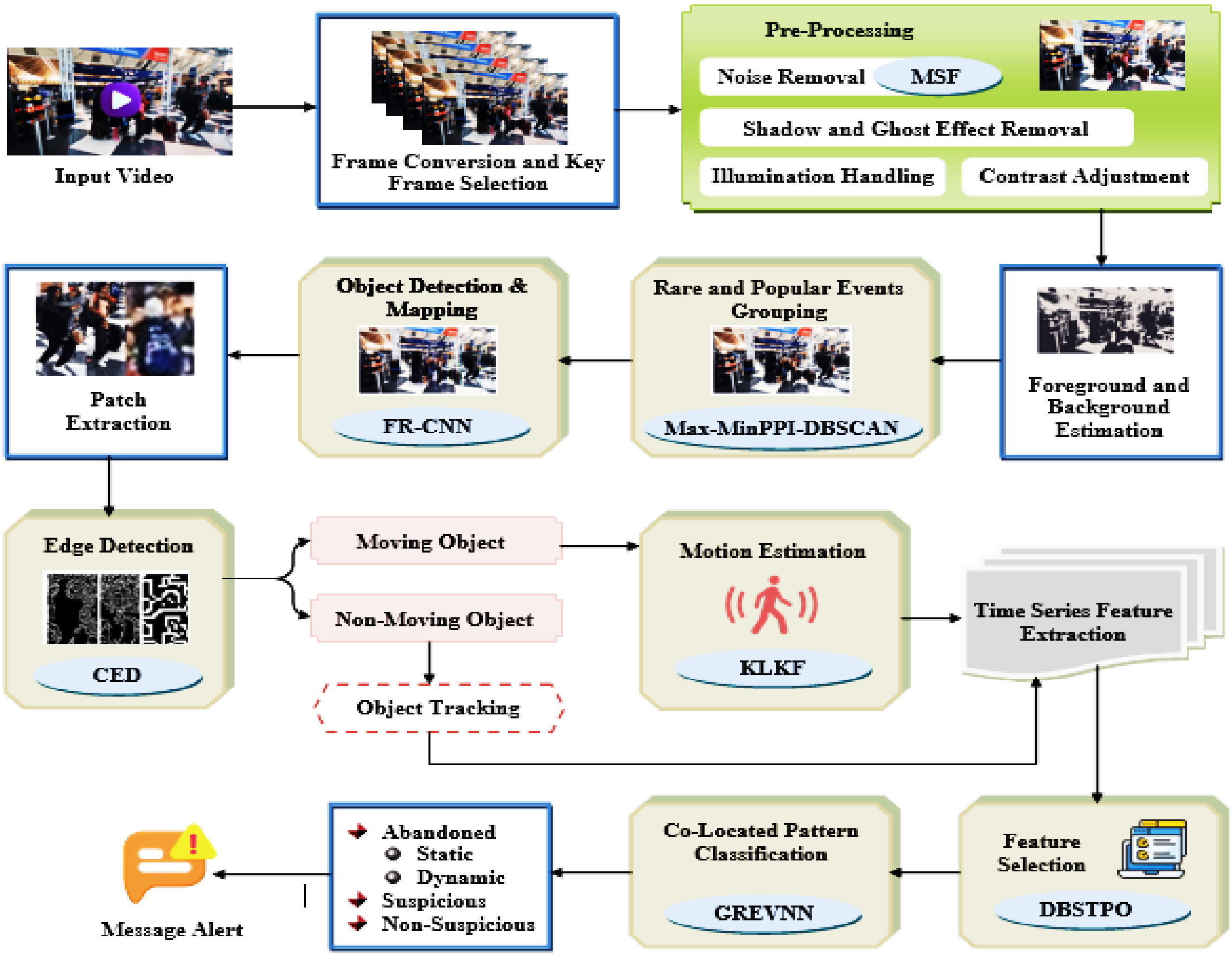


Figure 1: Framework of the Proposed Model

Zhang [30] recommended a model with multi-granular representation for object detection of videos. Using a semantic network, the object was tracked based on the semantic features. The abnormal event was then detected using a Deep convolution Network. The presented approach detected events with improved accuracy and processing time. But, the

A. Input Video, frame conversion, and keyframe selection

Initially, the surveillance videos from airports/railway stations (public places) are taken as input. Then, these videos are converted into frames/images(F), which are represented as,

$$F = [F_1, F_2, F_3, \dots, F_g] \quad (1)$$

Where, $(g)$is the number of frames extracted from the video. (F)has repeated frames, and it takes time to process the model. Therefore, keyframes are selected based on the variation between each frame and are given as,

$$F^* = \{F_1^*, F_2^*, F_3^*, \ldots, F_{g^*}^*\} \tag{2}$$

Where, (F^*) are the keyframes and (g^*) is the number of keyframes. Now, these keyframes are given for preprocessing.

B. Pre-processing

Here, (F^*) is preprocessed to remove the noise, shadow effect, and ghost effect and to adjust illumination and contrast.

Step1: Noise removal

The quality of the frame gets reduced due to the presence of white noise that has occurred by electrical interference. So, these noises are removed from (F^*) using a Median Filter (MF), which preserves the edges during noise removal. However, during filtering, the presence of minute Signal-to-Noise Ratio (SNR) breaks the edges of the objects, leading to false edges. Hence, to avoid this problem, silhouette extraction is done to preserve the image edges. The process of MSF is given below.

$$\alpha\,(k,l) = [k - l]/[\max(k,l)] \tag{3}$$

Where, $(k,l)$are the pixel values of (F^*). Now, the white noise is removed by passing (F^*)into the MSF as shown below,

$$\hat{F} = \left[\frac{k+l}{2}\right] \times \alpha(k,l) \tag{4}$$

Here, $(\hat{F})$is the noise-removed image, and this image is further preprocessed as follows,

Step2: Shadow and Ghost effect removal

The $(\hat{F})$ has shadows of objects/persons and a ghost effect due to the unwanted reflection of light. This causes misclassification during processing. So, these effects are removed from $(\hat{F})$ and the output obtained is given as $(\bar{F})$.

Step3: Illumination handling and Contrast Adjustment

Illumination is the amount of light present in the image, and it affects the information present in $(\bar{F})$. The contrast of the image can be enhanced effectively with correct illumination. So, firstly, the illumination handling is done for $(\bar{F})$, and the output is given as $(\ddot{F})$. Now, for $(\ddot{F})$, the contrast is adjusted to clearly display the information in the image. The contrast-adjusted image (F'') is represented by,

$$F'' = \langle F''_1, F''_2, F''_3, \ldots, F''_{g^*-1}, F''_{g^*}\rangle \tag{5}$$

Where, $(g^*)$are the number of preprocessed images, and the background and foreground are estimated for the preprocessed output.

C. Foreground and Background Estimation

The preprocessed image(F'')contains static and dynamic objects between each frame. These objects that are present in both background and foreground play a major role in rare and popular event recognition. So, the foreground and background are estimated to detect changes occurring in the image sequence. The foreground and background separated images are given as(Y). Now, from(Y), the rare and popular events are grouped as follows.

D. Rare and Popular Events Grouping

From(Y), the rare and popular events are grouped using the Max-MinPPI-DBSCAN method. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which is less affected by extreme values and separates the high-density region from the low-density region, is used for grouping. But, this model fixes the local density as a uniform value, which affects grouping with different densities. So, to overcome this issue, the epsilon (eps) value that locates the neighboring density is calculated using the Maximum-Minimum Pixel-Per-Inch (Max-MinPPI) formula. The process of the Max-MinPPI-DBSCAN is explained below.

The core points needed for grouping are based on two parameters, such as eps and minimum points. First, the eps (β) that determines the boundary to place the neighboring points is determined using the Max-MinPPI calculation.

$$\beta = \frac{\chi}{e * f} \tag{6}$$

Where, (χ) is the number of pixels present in (Y) and $(e * f)$ is the size of (Y). The maximum and minimum eps values $(\beta_{\max})$ and are used for grouping popular events and rare events, respectively. The min-point (δ) is the minimum number of points that are to be grouped to determine the core point(C).

Now, the grouping regarding$(\beta_{\max})$, $(\beta_{\min})$, and(δ) that shows the frame difference is done as given below.

$$C_1 = \beta_{\max} * \delta \tag{7}$$

$$C_2 = \beta_{\min} * \delta \tag{8}$$

Here, (C_1) and (C_2) are the core points for popular and rare events. The points (pixels) that do not come under the boundary of the core points are said to be noise points (γ), and clusters are not formed around them.

$$\gamma \rightarrow (< \delta) \;\; \forall(\beta_{\max}, \beta_{\min}) \tag{9}$$

These noise points (δ) are ignored and are not grouped. Hence, the rare and popular events are grouped to show the frame difference. The grouped image is represented as (M) and from this, the object is detected and mapped as described below.

E *Object Detection and Mapping*

The objects/persons are detected and mapped from the grouped image (M)using the FR-CNN method. Initially, this technique extracts important features and then calculates the anchor box. Subsequently, the object is classified and mapped. The process of FR-CNN is detailed as,

1. *Convolutional Layer*

Initially, the input image (M) is passed through the convolutional layer, which extracts important information from the image to create a feature map. The input is convolved and activated using the Rectifier Linear Unit (ReLU) (ϕ), which is equated as follows.

$$V = [(M \times w^{V}) + a] * \phi \tag{10}$$

Where, (V) is the output of the convolutional layer with weight value (w^{V}) and bias value (a).

2. *Region Proposal Network*

Then, $(V)$is passed through the Region Proposal Network (RPN), which places anchor boxes and creates a sliding window to adjust the anchor box. The RPN is activated by softmax (ε) and then max pooled regarding Region of Interest (ROI) as,

$$W = \max\{[(V \times w^{W}) + a] * \varepsilon\} \tag{11}$$

Where, (W) is the RPN output with the weight value (w^{W}).

3. *Fully-Connected Layer*

Here, the value (W) is fully connected to detect the object with the respective object score and then activated by (ε) as shown below.

$$U = [(\sum W \times w^{U}) + a] \times \varepsilon \tag{12}$$

Where, (U) is the output of the fully connected layer with the weight value (w^{U}). Hence, (U) is the object detected and mapped image, and the patch is extracted from this image as given below.

F. *Patch Extraction*

The images (U) have objects (o) and persons (p) detected in them. These objects and persons are made as a single patch to map the co-located objects/persons. It is equated as,

$$U^{*} = o\,(U) + p\,(U) \tag{13}$$

Where, (U^{*}) is the patch extracted image, and this image is given for edge detection as described below.

G. *Edge Detection*

The CED that removes the noise and false edges is used to detect the edges in (U^{*}). Edge detection is done to identify the moving and non-moving objects more effectively by analyzing the exact edges of the co-located objects/persons in (U^{*}). The description of CED is explained below.

The input is smoothed using the Gaussian filter to remove the noise in the input as shown below.

$$I = \frac{1}{2\pi\sigma^{2}} \times \exp^{\left[-\frac{(i^{2}+j^{2})}{2\sigma^{2}}\right]} \tag{14}$$

Where, (I) is the smoothened image with coordinates (i,j) and standard deviation (σ). The intensity gradients, such as Intensity Magnitude (Δ) and Intensity Direction (θ) for (I) is calculated as,

$$\Delta = \sqrt{\left(\frac{\partial I}{\partial i}\right)^{2} + \left(\frac{\partial I}{\partial j}\right)^{2}} \tag{15}$$

$$\theta = \tan^{-\left[\frac{(\partial I/\partial j)}{(\partial I/\partial i)}\right]} \tag{16}$$

Now, the non-maximum suppression is done to suppress all the weak edges and to obtain strong edges as given below.

$$P \to (\Delta > \theta) \tag{17}$$

When the magnitude is greater than the direction value, the edges in the image are considered as strong edges. The edge detected image is represented as (P). Next, from (P), the motion estimation and object tracking are carried out as detailed below.

H. *Moving and non-moving object separation*

For the purpose of estimating motion and object tracking, the moving and non-moving objects are separated from (P) as given below.

$$h\,(P) = P(h,\eta) - \eta\,(P) \tag{18}$$

Where, (h) is the moving object, and (η) is the non-moving object. Now, the motion is estimated for (h) as shown below.

1. *Motion Estimation*

In this phase, the motion of the moving object/person (h) is estimated using the KLKF method. The Kalman Filter (KF) that determines the varying quantities of moving objects regarding the law of motion is used for motion estimation. But, in KF, the error loss is calculated with respect to constant values and leads to premature convergence and improper motion estimation. Thus, to avoid this problem, the Kullback-Leibler (KL) divergence loss function that calculates the loss between the input and the prediction value is used as the error value in the KF. The KLKF algorithm is described below.

First, to find the information about the state of the object (h) with time (u), the state space (B) is calculated as,

$$B^{u+1} = [\Re * B^{u}(h)] + \mu \tag{19}$$

Where, (B^{u+1}) and $(\Re)$ are the current state and sequence of (h), (B^{u}) is the previous state of (h), and (μ) is the process disturbance. Next, the noisy observation (L^{u+1}) present in $(h)$is estimated as shown below.

$$L^{u+1} = [B^{u+1} \times (\Re)] + \Xi \tag{20}$$

Where, (Ξ) is the noise measured from (h). Now, the covariance error $\left(K^{u^*+1}\right)$ is calculated using the KL method, which measures the error by using the prior state$(B^{u+1})$and the posterior estimation$\left(B^{u^*+1}\right)$ for time $(u^* + 1)$ as derived below.

$$K^{u^*+1} = \sum B^{\quad u+1} * \log\left[B^{u+1}/_{B^{u^*+1}}\right] \tag{21}$$

Finally, the Kalman gain (J^{u+1}), which is the motion estimated value, is equated by,

$$J^{u+1} = K^{u^*+1} \times \left[L^{u+1} + K^{u^*+1}\right]^{-1} \tag{22}$$

After estimating motion, the non-moving objects are tracked as explained below.

2. Object Tracking

Now, the non-moving object/person (η) is tracked to effectively detect the abandoned, suspicious, and non-suspicious activity present in the surveillance video. By tracking, the path (location) of the object/person can be calculated. The object tracked value is represented as (N). Next, from the motion estimated and object-tracked output, time series features are extracted, which are given below.

I. Feature Extraction

The time series features, such as torso-based distance, spatiotemporal magnitude and direction angle, Histogram of Oriented Gradients-Depth Differential Silhouettes (HOG-DDS), trajectory, key joint-based distance, edges, ridges, and corners are extracted from both the motion estimated and object tracked output. The features (S) extracted are represented as,

$$S = \{S_1, S_2, S_3, .., S_q\} \tag{23}$$

Where, (q) is the number of features extracted. Next, from (S), optimal features are selected as mentioned below.

J. Feature Selection

Here, the optimal features required for the detection of co-located patterns from the surveillance video are selected from (S). The Dung Beetle Optimizer (DBO), which selects the value with high accuracy in a few iterations, is used for feature selection. However, the DBO selects the natural parameters to move the beetle in a random manner, affecting the optimal selection. To mitigate this issue, the State Transition Probability (STP) rule that transforms the movement of beetles based on its probability value is used. The method of DBSTPO is explained below.

1. .Initialization

The dung beetle (extracted features) population is first initialized to find the optimal position as shown below.

$$S_{d\times n} = \begin{bmatrix} S_{1,1} & \cdots & S_{1,y} & \cdots & S_{1,n} \\ \vdots & \ddots & \cdots & \iddots & \vdots \\ S_{z,1} & \cdots & S_{z,y} & \cdots & S_{z,n} \\ \vdots & \iddots & \cdots & \ddots & \vdots \\ S_{\ d,1} & \cdots & S_{d,y} & \cdots & S_{d,n} \end{bmatrix}_{d\times n} \tag{24}$$

Where, (d) is the number of dung beetles, and $\left(S_{z,y}\right)$ is the (z^{th}) beetle in the (y^{th}) dimension of the search space (n). The initial position of the dung beetle is calculated as follows.

$$S_{z,y} = [(ub^{y} - lb^{y}) \times r] + lb^{y} \tag{25}$$

Where, (ub^{y}, lb^{y}) are the upper bound and lower bound values regarding $\left(S_{z,y}\right)$ and $(r)$is the random number. The value $\left(S_{z,y}\right)$ is the current position of the dung beetle, and this position gets updated based on the fitness value and is given below.

2. Fitness

The fitness function (Ω), which determines the optimal features, is calculated regarding maximum classification accuracy (A) and is denoted as,

$$\Omega = \max(A) \tag{26}$$

By using (Ω), the dung beetle position gets updated regarding the search for a suitable location.

3. Position Updation

The dung beetle uses five behaviors, such as ball rolling, dancing, foraging, stealing, and reproduction to move to a suitable location. The position update of the dung beetle is described below.

4. *Rolling*

Here, the dung beetle rolls the dung ball in the direction regarding sunlight and wind, which are the natural parameters (E). Hence, the position updation of dung beetle $S_{z,y}(c+1)$ is equated as,

$$S_{z,y}(c+1) = S_{z,y}(c) + \left[E \times v \times S_{z,y}(c-1)\right] + (r * \varsigma) \tag{27}$$

$$\varsigma = \left[S_{z,y}(c) - S_{z,y}(\infty)\right] \tag{28}$$

Where, (c) and $(c-1)$ are the current and previous iterations, (v) is the deflection coefficient, and (ς) is the light intensity value regarding the present position $S_{z,y}(c)$ and worst position $S_{z,y}(\infty)$ of the dung beetle. The natural parameter (E) is obtained using the STP value, which generates rules and helps in the movement of the dung beetle, which is equated below.

$$E = \arg\max\,[\rho(c)\,]^{s} \times \left[S_{z,y}(c)\,\right]^{t} \tag{29}$$

Where, $[\rho(c)]$ is the pheromone concentration that determines the path of the dung beetle and (s, t) is the heuristic factor.

5. *Dancing*

Here, when the dung beetle faces obstacles, it dances and takes a new route. Thus, the position $S^{*}_{z,y}(c+1)$ is updated as follows.

$$S^{*}_{z,y}(c+1) = S_{z,y}(c) + \left\{\tan(\mathfrak{I}) \times \left[S_{z,y}(c) - S_{z,y}(c-1)\right]\,\right\} \tag{30}$$

Where, $(\mathfrak{I})$ is the deflection angle in the range off $(0-360°)$.

6. *Reproduction*

The female dung beetle finds the spawning area regarding the spawning upper bound and lower bound values (ub^{*}, lb^{*}) and lays the eggs for reproduction. The position of reproduction is given by,

$$S''_{z,y}(c+1) = \left[S_{z,y}(c) - lb^{*}\right] + \left[S_{z,y}(c) - ub^{*}\right] \tag{31}$$

7. *Foraging*

The dung beetles burrow into the ground in search of food based on the upper bound and lower bound values $(ub^{\mp}, lb^{\mp})$. The updated position $\hat{S}_{z,y}(c+1)$ is determined by,

$$\hat{S}_{z,y}(c+1) = S_{z,y}(c) + \left\{r * \left[S_{z,y}(c) - lb^{\mp}\right]\right\} + \left\{r * \left[S_{z,y}(c) - ub^{\mp}\right]\,\right\} \tag{32}$$

8. *Stealing*

Here, some dung beetle steals the food of other beetles and moves its position. The position $\dddot{S}_{z,y}(c+1)$ is equated by,

$$\dddot{S}_{z,y}(c+1) = \cancel{S}_{z,y}(c) + \left[r * \left(S_{z,y}(c) - \cancel{S}_{z,y}(c)\right)\,\right] \tag{33}$$

Where, $\cancel{S}_{z,y}(c)$ is the beetle with food.

Thus, from each behavior of the dung beetle, the optimal feature is obtained with respect to fitness (Ω), and it is represented as (S^{best}). The pseudocode for DBSTPO is detailed below.

Pseudocode for DBSTPO

```
Input: Extracted Features
Output: Optimal Feature ($S^{best}$)
Begin
Initialize population ($S_{d\times n}$), Iteration ($T, T^{\max}$)
Calculate current position of dung beetle
        $S_{z,y} = [(ub^{y} - lb^{y}) \times r] + lb^{y}$
Evaluate fitness function $\Omega = \max(A)$
While ($T \le T^{\max}$)
For ($\Omega$)
Update Position of dung beetle
/1/Rolling
Derive Natural Parameters
$E = \arg\max\,[\rho(c)\,]^{s} \times [S_{z,y}(c)\,]^{t}$
Dung Beetle position
    $S_{z,y}(c+1) = S_{z,y}(c) + [E \times v \times S_{z,y}(c-1)] + (r * \varsigma)$
/2/Dancing Position update $S^{*}_{z,y}(c+1)$
/3/Reproduction place
        $S''_{z,y}(c+1) = [S_{z,y}(c) - lb^{*}] + [S_{z,y}(c) - ub^{*}]$
/4/Foraging behavior in search of food $\hat{S}_{z,y}(c+1)$
/5/Stealing food
        $\dddot{S}_{z,y}(c+1) = \cancel{S}_{z,y}(c) + [r * (S_{z,y}(c) - \cancel{S}_{z,y}(c))\,]$
End for
End while
Return Optimal Feature ($S^{best}$)
End
```

Now, the optimal feature $(S^{best})$is given to the proposed classifier to detect the co-located pattern as mentioned below.

K. *Classification*

In this phase, the co-located patterns of objects from $(S^{best})$are detected using the GREVNN classifier. RNN remembers the information and classifies the time series data with various lengths accurately. But, the tanh activation function used in RNN restricts the longest video from processing. This leads to a reduction in classification efficiency. Thus, to avoid this issue,

Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Generalized Extreme Value (GEV), which processes the large videos efficiently, is used instead of the tanh activation function. The framework of GREVNN is given in Figure 2.

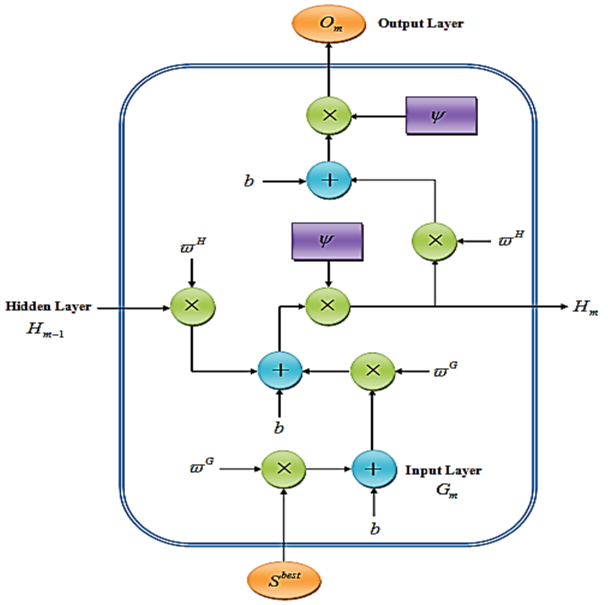


Figure 2: Architecture of GREVNN

The GREVNN has three layers, such as an Input Layer (G_m), a Hidden Layer (H_m), and an Output Layer (O_m). The process of GREVNN is described below.

1. *Input Layer*

The Input Layer (G_m) regarding time (m) consists of the optimal features (S^{best}), and it is converted to vector value as equated below.

$$G_m = (S^{best} \times \varpi^G) + b \tag{34}$$

Where, (ϖ^G) is the weight value of (G_m) and $(b)$is the bias value.

2. *Hidden Layer*

The Hidden Layer (H_m)collects the information from previous output (H_{m-1}) and processes it with present input (G_m). Then, the values are activated using the GEV activation function (ψ) to get the output of the Hidden Layer for time (m), which is given below.

$$H_m = \{\ [\,(G_m \times \varpi^G) + (H_{m-1} \times \varpi^H)\,] + b\} * \psi \tag{35}$$

Where, (ϖ^H) is the weight value of the hidden layer. The GEV activation function (ψ) tends to process the video with vast size, and it is calculated as follows.

$$\psi\,[S^{best}(\tau, x, \kappa)] =$$

$$\begin{cases} \exp\left(-\exp^{-\left(\frac{S^{best}-\tau}{x}\right)}\right) & \forall(\kappa = 0) \\ \exp\left[-\left(1 + \left(\kappa * \left(\frac{S^{best}-\tau}{x}\right)^{\frac{-1}{\kappa}}\right)\right)\right] & \forall(\kappa \neq 0) \end{cases} \tag{36}$$

Where, $(\tau, x, \kappa)$are the location parameter, scale parameter, and shape parameter of the input (S^{best}).

3. *Output Layer*

The hidden layer output (H_m) is processed and activated using $(\psi)$to get the final output, which is shown below.

$$O_m = \ [(H_m \times \varpi^H) + b] * \psi \tag{37}$$

Thus, (O_m) is the co-location pattern detected value and is represented as,

$$O_m = [O_m^1, O_m^2, O_m^3] \tag{38}$$

Where, (O_m^1) is the abandoned object (static/dynamic) in the video and (O_m^2) and (O_m^3) are the suspicious and non-suspicious activity detected in the video. The pseudocode for GREVNN is given below.

Pseudocode for GREVNN

Input: Optimal Feature (S^{best})
Output: Detected co-located pattern (O_m)

```
Begin
Initialize parameters$(\varpi^G)$, $(\varpi^H)$ , $(b)$
For $(S^{best})$
While time $(m)$
Derive Input Layer
        $G_m = (S^{best} \times \varpi^G) + b$
Evaluate activation function $(\psi)$
Calculate Hidden layer
        $H_m = \{\ [\,(G_m \times \varpi^G) + (H_{m-1} \times \varpi^H)\,] + b\} * \psi$
Find Output Layer
        $O_m = \ [(H_m \times \varpi^H) + b] * \Psi$
Detect co-located pattern
        $O_m = [O_m^1, O_m^2, O_m^3]$
End while
End for
Obtain detected co-located pattern $(O_m)$
End
```

Hence, after the detection of the co-located pattern in the surveillance video, an alert message is generated automatically and notified to the required authority or person about the event. Thus the proposed model effectively recognized the co-located patterns regarding rare and popular events in the surveillance video. The performance analysis of the proposed framework is explained in section 4

Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

IV. Result and Discussion

In this section, the performance of the proposed method is analyzed for the classification of co-located patterns, feature selection of objects, rare and popular event recognition, pre-processing of video frames, and estimation of moving objects. For assessing the performance, the proposed model is implemented using the PYTHON programming platform.

A. Dataset description

For analyzing the performance, the advanced video and signal-based surveillance dataset is utilized. This dataset provides Closed Circuit Television (CCTV) footage for event detection, including abandoned baggage in three different sequences.

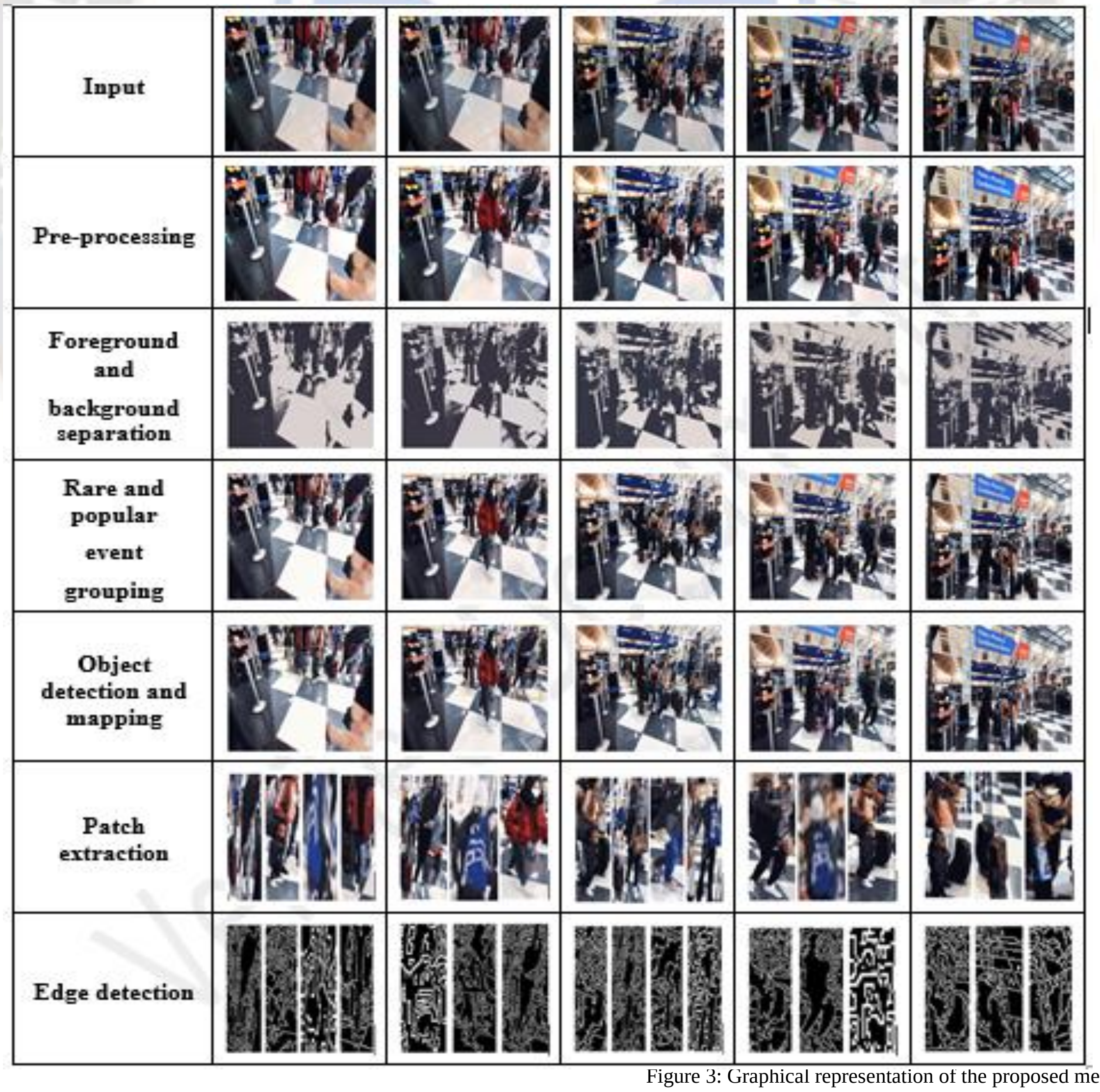
The video is sampled at the rate of 25 Hertz and the image size acquired from CCTV is in 720×576 pixels. From the dataset, 80% is used for training and 20% is used for testing the proposed model

B. Performance analysis

The performance of co-located patterns classification by the proposed method in terms of accuracy, precision, recall, and f-measure is evaluated by comparing the proposed method with existing techniques, such as RNN, Gated Recurrent Unit (GRU), Bi-directional Long Short Term Memory (Bi-LSTM), and LSTM.

The performance of the proposed technique in the classification of co-located patterns is represented in Figure 3. It is observed that the proposed approach classifies co-located patterns with 99.072% accuracy, 99.387% precision, 99.462% recall, and 99.454% f-measure. The existing methods attained an average of 94.904% accuracy, 94.740% precision, 94.132% recall, and 94.607% f-measure, which are lower than the proposed methods.

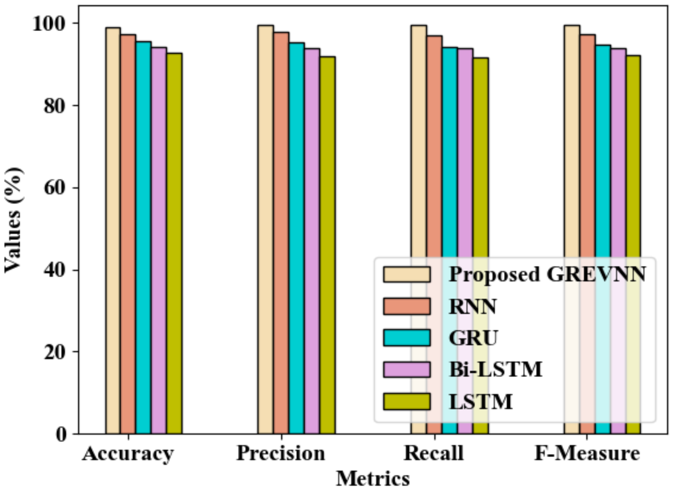
TABLE 1: IMAGE RESULTS OF THE PROPOSED METHOD


Input
Pre-processing
Foreground and background separation
Rare and popular event grouping
Object detection and mapping
Patch extraction
Edge detection

.

Figure 3: Graphical representation of the proposed method

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*



Since the object patches and their time series features are extracted, the proposed model efficiently classifies the co-located patterns.

TABLE 2: PERFORMANCE COMPARISON OF THE PROPOSED GREVNN METHOD

Methods	Specificity (%)	FPR	FNR
Proposed GREVNN	99.4823	0.4226	0.7414
RNN	97.5642	2.4614	2.2101
GRU	95.7246	5.3891	5.3023
Bi-LSTM	94.3975	7.3482	7.1286
LSTM	91.6487	9.5272	9.5348

The proposed GREVNN method on co-located pattern identification with respect to specificity, FPR, and FNR is depicted in Table 2. The proposed method achieves a specificity of 99.482%, which is higher than the existing methods. From Table 2, it is also noticed that the GREVNN recognizes co-located patterns with an FPR of 0.422 and an FNR of 0.741. In the meantime, the existing methods gained an average FPR of 6.181 and FNR of 6.043. As the GEV activation function is adopted in RNN by the proposed technique, the co-located objects are well categorized from the long video sequences.

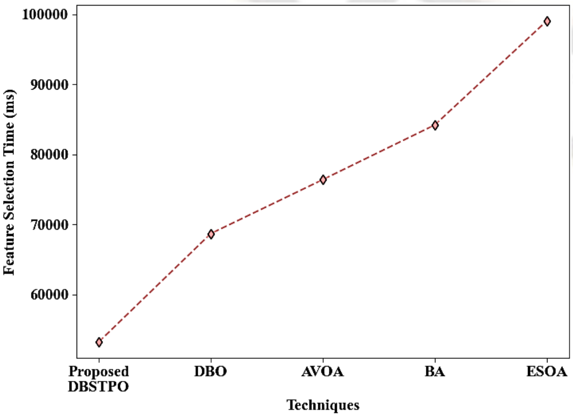


Figure 4: Performance illustration of the proposed algorithm

The performance of the proposed DBSTPO algorithm in selecting time series features of the object is illustrated in Figure 4. From Figure 4, the proposed DBSTPO selects objects' features within 53239.44 ms. For selecting features, the existing methods, such as Dung Beetle Optimizer (DBO) take 68688.66ms, African Vultures Optimization Algorithm (AVOA) takes 76390.88ms, Bees Algorithm (BA) takes 84190.91ms, and Egret Swarm Optimization Algorithm (ESOA) takes 99030.76ms. Since the time series features of the object are acquired using the proposed algorithm, the optimal features are selected within a short duration from the detected object.

TABLE 3: FITNESS ACHIEVEMENT OF THE PROPOSED ALGORITHM

Algorithms	Average Fitness
Proposed DBSTPO	95.2632
DBO	91.8919
AVOA	86.4706
BA	75.2941
ESOA	71.2857

The proposed DBSTPO algorithm performance in terms of average fitness in the feature selection of objects is indicated in Table 3. It is found that the proposed DBSTPO gained an average fitness value of 95.2632, which is higher than the existing algorithms' fitness. The state transition probability rule is adopted by the proposed algorithm. Thus, the average fitness is increased for DBPSTPO in selecting the optimal features of the object. The existing algorithms, such as DBO attained 91.891 of Average Fitness (AF), AVOA attained 86.470 AF, BA attained 75.294 AF, and ESOA attained 71.285 AF, which are lower than the proposed approach. Thus, the proposed DBSTPO algorithm efficiently selects object features over the traditional algorithms.

TABLE 4: PERFORMANCE OF RARE AND POPULAR EVENT RECOGNITION

Techniques	Grouping time (ms)
Proposed Max-MinPPI-DBSCAN	49592
DBSCAN	58561
PAM	75910
KMA	102025
FCM	119794

The performance of the proposed Max-MinPPI-DBSCAN method in grouping rare and popular events is compiled in Table 4. The proposed method is evaluated by comparing with existing techniques, such as DBSCAN, Partition Around Medoids (PAM), K-Means Algorithm (KMA), and Fuzzy C Means (FCM) clustering. From Table 4, it is observed that the proposed clustering method grouped rare and popular events within 49592 ms, whereas the existing approaches take an average of 89072 ms for grouping events.The foreground and background separation of video frames is carried out before event grouping for object detection. Further, the maximum and minimum pixels per inch are analyzed for grouping the rare and popular events by the proposed technique. Thus, the performance is enhanced for the proposed approach in clustering the rare and popular events over the prevailing methods.

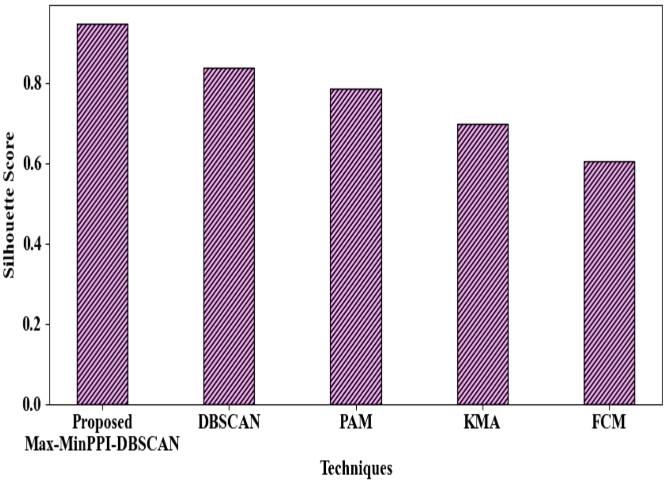


Figure 5: Performance comparison of the proposed event recognition technique

In terms of silhouette score, the proposed Max-MinPPI-DBSCAN in grouping events is represented in Figure 5. The proposed technique achieved a silhouette score of 0.947, which is higher than the existing techniques. Since preprocessing and foreground and background separation of video frames were carried out before event recognition, the proposed approach recognized rare and popular events more accurately. In the meantime, the existing techniques grouped events with an average lower silhouette score of 0.731. Thus, the proposed technique's performance is improved than the prevailing methods in grouping the rare and popular events.

TABLE 5: PRE-PROCESSING PERFORMANCE OF THE PROPOSED METHOD

Methods	PSNR (db)	MSE
Proposed MSF	39.11	0.9245
MF	32.65	1.1035
WF	30.83	1.9275
GF	21.58	2.7035
ADF	18.49	3.7125

The performance of the proposed MSF approach with respect to Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) in pre-processing video frames is compiled in Table 5. The pre-processing performance is analyzed by comparing it with the existing filters, such as Median Filter (MF), Wiener Filter (WF), Gaussian Filter (GF), and Anisotropic Diffusion Filter (ADF). From Table 5, it is found that the proposed filter is pre-processed with a PSNR of 39.11 db and MSE of 0.924, which are better outcomes than the existing pre-processing methods. The existing methods pre-processed video frames with an average PSNR of 25.88 db and MSE of 2.361. As the image edges are preserved while removing the noise of video frames, the PSNR and MSE performance are enhanced by the proposed MSF method over the existing methods.

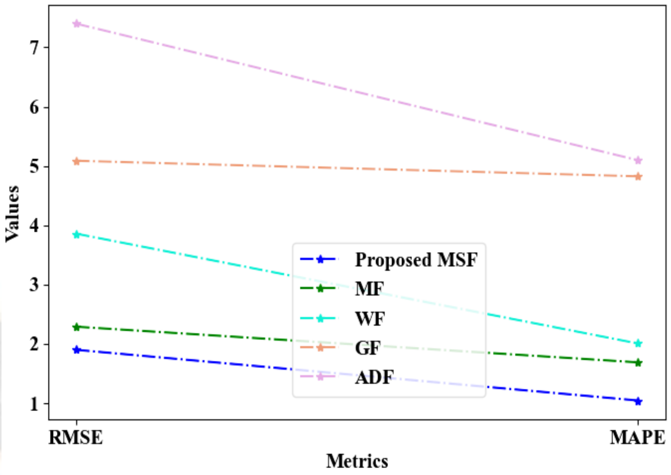


Figure 6: Performance evaluation of proposed MSF

The proposed MSF's performance for surveillance video pre-processing with respect to Root MSE (RMSE) and Mean Absolute Percentage Error (MAPE) is depicted in Figure 6. It is noticed from Figure 6 that the proposed method pre-processed video frames with an RMSE of 1.899 and MAPE of 1.046. In the meantime, the existing filters pre-processed video frames with an average RMSE of 4.658 and RMSE of 3.405, which are higher than the proposed filter. Due to the prevention of developing fake noisy edges, the proposed filter pre-processed frames with minimum error. Thus, the performance of the proposed method is better than the existing approaches in efficient pre-processing of video frames.

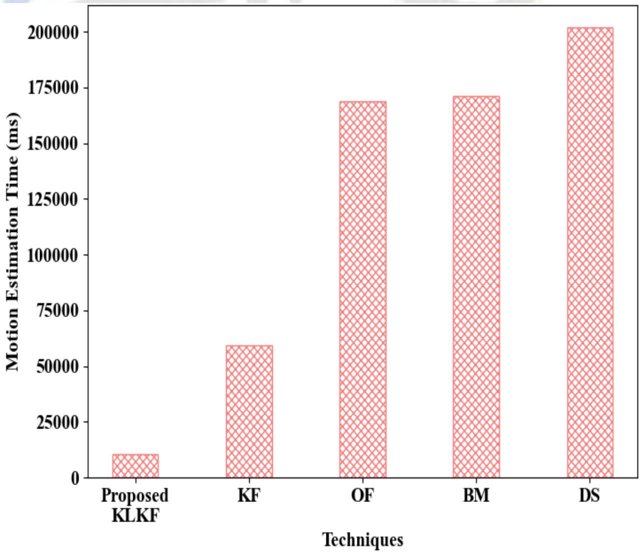


Figure 7: Performance representation of proposed KLKF

The performance of the proposed technique in estimating the motion of the moving object is represented in Figure 7. The improved performance of the proposed technique is validated by comparing it with existing techniques, such as KF, Optical Flow (OF), Block Matching (BM), and Diamond Search (DS). From Figure 7, it is observed that the proposed KLKF method estimates object motion in 10398ms, which is a lesser duration than the existing methods. In estimating object motion, the existing methods, such as KF take 59050ms, OF takes 16857ms, BM takes 171236ms, and DS takes 201833. The premature convergence that occurred during object estimation using KF is overcome by the proposed method. Thus, the proposed method takes minimal time for the estimation of moving objects over the existing methods.

TABLE 6: PERFORMANCE ANALYSIS OF KLKF

Methods	SAD
Proposed KLKF	2.6728
KF	3.2521
OF	3.7786
BM	4.0326
DS	6.2309

The proposed KLKF method's performance with respect to the Sum of Absolute Difference (SAD) for estimating object motion is shown in Table 6. In the proposed approach, the SAD in estimating moving objects is 2.672, which is lower than the existing methods. In the estimation of object motion, the existing methods attained an average SAD of 4.323. Since the object patches are extracted and edges are detected before estimating object motion, the proposed filter estimates object motion with minimum SAD. Thus, the performance of the proposed approach is improved in analyzing the object movement than the prevailing methods.

TABLE 7: PERFORMANCE COMPARISON WITH THE RELATED WORKS

References	Techniques used	Precision (%)	Recall (%)	F-measure (%)
Proposed	GREVNN	99.387	99.465	99.454
Raju[18]	CNN	90.00	98.00	93.00
Shahbano [22]	Joint scale LBP	98.89	95.11	-
Sarker [20]	T-C3D	95.33	96.33	95.67
Huang [10]	VTD-FastICA	81.45	77.02	73.32
Elhoseny[7]	MODT	63.08	64.16	92.42

The performance comparison of the proposed GREVNN model in recognizing co-located objects is weighing against some existing works. From Table 7, it is observed that the existing methods, such as CNN, detected objects with a precision of 90.00%, recall of 98%, and f-measure of 93%. Further, the techniques, such as joint scale Local Binary Pattern (LBP) and Temporal Convolutional 3-Dimensional (T-C3D) neural network recognized objects with an average of 97.15% precision, 95.72% recall, and 95.67% f-measure, which are lower than the proposed technique. Also, the existing approaches, including Video in Time Domain-Fast and Independent Component Analysis (VTD-FastICA) and Multi-Object Detection and Tracking (MODT), attained lower performance in co-located object detection. The existing works detected objects without recognizing the rare and popular events. However, the proposed model detects objects by grouping rare and popular events, which enhances the performance to a precision of 99.387% and recall of 99.465%. Thus, the proposed technique recognizes abandoned objects more efficiently than the traditional methods.

5. CONCLUSION

The co-located objects of the surveillance video streams are recognized using Max-MinPPI-DBSCAN and GREVNN in this paper. The input video was converted into frames and then pre-processed, followed by separating the background from the foreground. Subsequently, rare and popular event grouping, patch extraction, edge detection, movement-based object separation, and time series feature extraction are carried out for the efficient recognition of co-located patterns. From the performance analysis, it is proven that the video frames are pre-processed with a PSNR of 39.11db and MSE of 0.924; in addition, rare and popular events are grouped within 49592ms using the proposed approach. Furthermore, the proposed approach classifies abandoned, suspicious, and non-suspicious objects with an accuracy of 99.072% and a precision of 99.387%. Thus, the co-located objects are identified effectively based on rare and popular events by using the proposed technique.

FUTURE SUGGESTION

Although the co-located objects are recognized by extracting their time-series features, the periodic changes occurring in the co-located object are not concentrated. Therefore, in the future, object recognition will be further enhanced by monitoring the co-located object in a time series.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Author Contributions

The J.C.V is responsible for the paper's background research, ideation,methodology,dataset collecting,implementation, , result analysis, and comparison, as well as the preparation and editing of the draft and the visualization. The B.K.R has overseen, reviewed, edited and supervised the work.

REFERENCES

[1] Ahamad, R., & Mishra, K. N., "Hybrid approach for suspicious object surveillance using video clips and UAV images in cloud-IoT-based computing environment", Cluster Computing, 1–25, 2023.

[2] Akella, S., Abhang, P., Agrharkar, V., & Sonkusare, R, "Crowd Density Analysis and Suspicious Activity Detection", IEEE International Conference for Innovation in Technology, 2020, 6–9.

[3] Ayesha Khan, J. A., Ahmad, W., Nadeem, M., Zahra, S. W., Arshad, A., Riaz, S., & Shahid, U, "Baggage Detection and Recognition Using Local Tri-Directional Pattern", International Journal of Mobile Computing Technology, 1(1), 8–17,2023

[4] Din, M., Bashir, A., Basit, A., & Lakho, S., "Abandoned Object Detection using Frame Differencing and Background

Subtraction", International Journal of Advanced Computer Science and Applications, 11(7), 676–681,2020.

[5] Dogariu, M., Stefan, L. D., Constantin, M. G., & Ionescu, B,"Human-Object Interaction: Application to Abandoned Luggage Detection in Video Surveillance Scenarios",13th International Conference on Communications, 2020,157–160.

[6] Dwivedi, N., Singh, D. K., & Kushwaha, D. S, "An Approach for Unattended Object Detection through Contour Formation using Background Subtraction", Procedia Computer Science, 171, 1979–1988,2020

[7] Elhoseny, M. (2019). "Multi-object Detection and Tracking (MODT) Machine Learning Model for Real-Time Video Surveillance Systems", Circuits, Systems, and Signal Processing, 39, 611–630, 2019.

[8] Gao, Q., & Liang, P,"Airline Baggage Appearance Transportability Detection Based on A Novel Dataset and Sequential Hierarchical Sampling CNN Model", IEEE Access, 9, 41833–41843. 2021.

[9] Gomathy Nayagam, M., & Ramar, K," Reliable object recognition system for cloud video data based on LDP features", Computer Communications, 149, 343–349,2020.

[10] Huang, Y., Jiang, Q., & Qian, Y, "A Novel Method for Video Moving Object Detection Using Improved Independent Component Analysis", IEEE Transactions on Circuit and Systems for Video Technology, 31(6), pp 2217–2230,2021.

[11] Iqbal, M. J., Iqbal, M. M., Ahmad, I., Alassafi, M. O., Alfakeeh, A. S., & Alhomoud, A," Real-Time Surveillance Using Deep Learning", Security and Communication Networks, 2021, 1–17,2021.

[12] Jagad, C., Chokshi, I., Chokshi, I., Jain, C., Katre, N., Narvekar, M., & Mukhopadhyay, D, "A Study on Video Analytics and Their Performance Analysis for Various Object Detection Algorithms", IEEE IAS Global Conference on Emerging Technologies, 2022, 1095–1100.

[13] Lyu, Z., & Luo, J., "A Surveillance Video Real-Time Object Detection System Based on Edge-Cloud Cooperation in Airport Apron", Applied Science, 12(19), 1–17, (2022).

[14] Nayak, R., Pati, U. C., & Das, S. K,"A comprehensive review on deep learning-based methods for video anomaly detection", Image and Vision Computing, 106, 1–64, 2021.

[15] Park, H., Park, S., & Joo, Y, "Detection of Abandoned and Stolen Objects Based on Dual Background Model and Mask R-CNN", IEEE Access, 8, 80010–80019,2020.

[16] Pathan, S., Pardeshi, S., Uke, N., Jha, A., & Satpute, A, "Abandoned object detection for intelligent video surveillance", International Journal of Advances in Engineering Research, 23(3), 22–31,2022.

[17] Pudasaini, D., & Abhari, A, "Scalable Object Detection, Tracking and Pattern Recognition Model Using Edge Computing", Proceedings of the 2020 Spring Simulation Conference,2020, 1–11.

[18] Raju, D., & Preetha, K. G. (2019). "Efficient Abandoned Luggage Detection in Complex Surveillance Videos", In Innovative Data Communication Technologies and Application ,pp. 181–187,2019.

[19] Saluky, S., Supangkat, S. H., & Nugraha, I. B, "Abandoned Object Detection Method Using Convolutional Neural Network", 7th International Conference on ICT for Smart Society: AIoT for Smart Society,2020, 20–23.

[20] Sarker, M. I., Losada-Gutiérrez, C., Marrón-Romera, M., Fuentes-Jiménez, D., & Luengo-Sánchez, S,"Semi-Supervised Anomaly Detection in Video-Surveillance Scenes in the Wild",Sensors, 21(12), 1–20,2021.

[21] Sathesh, A., & Hamdan, Y. B, "Speedy Detection Module for Abandoned Belongings in Airport Using Improved Image Processing Technique", Journal of Trends in Computer Science and Smart Technology, 3(4), 251–262,2022.

[22] Shahbano, Ahmad, W., Shah, S. M. A., & Ashfaq, M, "Carried Baggage Detection and Classification using Joint Scale LBP", 5th International Electrical Engineering Conference,2020, 4–8.

[23] Siddique, A., & Medeiros, H, "Tracking Passengers and Baggage Items using Multi-camera Systems at Security Checkpoints", ArXiv, 1–14,2022.

[24] Srinath, R., Vrinidavaniam, J., Vasudev, V. P., Supreeth, S., Raj, H., & Kesarwani, A, "A Machine Learning Approach for Localization of Suspicious Objects using Multiple Cameras", IEEE International Conference for Innovation in Technology,2020, 1–6.

[25] Walia, G. S., Kumar, A., Saxena, A., Sharma, K., & Singh, K, "Robust object tracking with crow search optimized multi - cue particle filter", Pattern Analysis and Applications, 23, 1439–1455,2019.

[26] Wan, S., Xu, X., Wang, T., & Gu, Z,"An intelligent video analysis method for abnormal event detection in intelligent transportation systems", IEEE Transactions on Intelligent Transportation Systems, 22(7), 4487-4495,2020.

[27] Wang, B., & Yang, C," Video Anomaly Detection Based on Convolutional Recurrent AutoEncoder", Sensors, 22(12), 1–16, 2022.

[28] Yaseen, M. U., Anjum, A., Fortino, G., Liotta, A., & Hussain, A, "Cloud based Scalable Object Recognition from Video Streams using Orientation Fusion and Convolutional Neural Networks", Pattern Recognition, 121, 1–24,2022.

[29] Zhang, C., Wu, X., & Gao, X. (2020). "An improved Gaussian mixture modeling algorithm combining foreground matching and short-term stability measure for motion detection", Multimedia Tools and Applications, 79, 7049–7071,2020.

[30] Zhang, Z., Zhang, Y., Cheng, X., & Lu, G"Siamese network for object tracking with multi-granularity appearance representations", Pattern Recognition, 118, 1–13,2021.