Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777658

TBC-K-MEANS BASED CO-LOCATED OBJECT RECOGNITION WITH CO-LOCATED OBJECT STATUS IDENTIFICATION FRAMEWORK USING MAX-GRU

Jayaram C V
Assistant professor, Dept. of cse, Mysore college of engineering, and management
Mysore-28, Jayaramscv77@gmail.com

Dr B K Raghavendra
Ph.D, Professor and HOD, Dept. of Information science and engineering, Don Bosco Institute of technology, Bengaluru, raghavendrabk69@gmail.com

Abstract: In the application of detached object recognition in public places like railway terminals, the recognition of the co-located objects in the video is a more vital process. Nevertheless, owing to the occurrence of multiple co-located object instances, the analysis of the status of the co-located object in the video is a challenging process. Hence, for solving this issue, this paper proposes the Min-Max Distance based K-Means (MMD-K-Means)-centric co-located object recognition with object status identification. Primarily, the input video from the railway is converted to frames. Subsequently, it was improved using Contrast Limited Adaptive Histogram Equalization (CLAHE). Next, Tukey's Bi-weight Correlation-based Byte Tacking (TBC-BT) and MMD-K-Means clustering are done for the detection and tracking of moving and non-moving objects. Subsequently, the Cyclic Neighbor-based Connected Component Analysis (CN-CCA) process was done from the static and moving object-detected frames for the main and co-located object labeling. Next, it executed the patch extraction for the separate analysis of each instance. At last, the Maxout-based Gated Recurrent Unit (Max-GRU) determined the object status in CN-CCA processed frame with the estimated distance between objects and extracted features from the static objects. The proposed system's performance is experimentally proved with several performance metrics.

Keywords: *Co-located object recognition, video stream, Min-Max Distance based MMD-K-Means, Tukey's Bi-weight Correlation-based Byte Tacking (TBC-BT), Cyclic Neighbor-based Connected Component Analysis (CN-CCA), Maxout-based Gated Recurrent Unit (GRU).*

1. INTRODUCTION

For visual applications, Pattern Recognition (PR) is a necessary objective. Here, solving numerous high-level intelligent issues depends heavily on the success of automatic and accurate PR (Zhang et al., 2020). In addition, the exploration of new boundaries like studies on image and video, and co-located object recognition is enabled by PR (Wu et al., 2022). In spatial data and spatio-temporal data like videos, the co-located PR can be done. In the application of surveillance, namely airways, railways, etc for the detached co-located object recognition, the co-located PR in the video is utilized. As a co-located object (i.e. bag) can move along with its owner for a long time, it can be left without movement at some point (Popov et al., 2021). Hence, for accurate co-located object recognition and tracking, video content analysis approaches are utilized.

In spontaneously evaluating video to observe and regulate spatial and temporal events, Video Content Analytics (VCA) is the proficiency (Jayaram & Bhajantri, 2019). Centered on the human co-located-object Interaction, the co-located object recognition in VCA is done. The human-based object detection and tracking task, which intends for recognizing betwixt persons and objects in a video, is named Human–co-located object interaction (Sun et al., 2021)(T. Wang et al., 2022). A process that locates the position of one or more objects in the image or video with the help of a bounding box is termed object detection (Kaur & Singh, 2022). Throughout the years, object detection approaches have enhanced to the point in which their error rate is less than human beings when monitoring surveillance (Alzaabi et al., 2020). Mask Regions with CNN (RCNN) (Dogariu et al., 2020), You Only Look Once (YOLO) (Santad et al., 2018), et cetera are a few prevailing approaches for the identification of abandoned Co-located object recognition. But, they are usually designed for a single target domain (W. Wang et al., 2022) and are designed for static objects only. Hence, for co-located object status recognition, the Multiple Object Tracking (MOT) (X. Yang et al., 2020) is considered as an effective approach for the dynamic moving co-located objects.

The MOT task is largely divided into locating multiple objects, maintaining their identities, and yielding their individual trajectories from an input video (Luo et al., 2021). Long Short Term Memory (LSTM) (Tsai et al., 2020), Kalman filtering (Elhoseny, 2020), and so on are the prevailing works developed for the MOT. Nevertheless, in the baseline studies, when the video is investigated as a whole, the status of the entire co-located object can't be identified. Therefore, for solving this issue, this paper proposes patch-based co-located object status detection with MMD-K-means-based co-located object recognition

1.1 Problem statements

The existing studies for co-located object recognition in the video stream include the succeeding setbacks.

- There was a higher probability that some of the co-located object statuses could be missed when the multiple co-located object instances were tracked in the video for the co-located object status determination.

- In existing works, the tracking of the co-located objects became unreliable, while the connections betwixt the connected and co-located objects are neglected.

- In most of the existing works, the detached co-located object is recognized grounded on its long-time stationary behavior, which is unreliable for object status determination.

- When the pattern of the main and co-located object is not studied, the recognition of the detached co-located object is difficult.

By considering these issues, developing an effective co-located objects recognition system for determining the status of all the co-located objects in the video is the goal of the proposed scheme.

- The patch-centric distance evaluation of objects is done to determine the status of all the co-located objects in the video.

- The CN-CCA process is executed to determine the connected main object of a co-located object.
- The Max-GRU recognition based on the distance evaluation and features of static objects is proposed to determine the co-located object status efficiently.

The remaining part of the paper is arranged as follows: the related works of the proposed system are elucidated in Section 2, the proposed approach is elaborated in Section 3, the experimental outcomes of the proposed study are discussed in Section 4, and at last, the paper is concluded with the future recommendation in Section 5.

2. RELATED WORKS

(Siddique & Medeiros, 2022) recommended Self-Supervised Learning (SSL) system for tracking passengers and baggage items at security checkpoints. For object detection, the Convolutional Neural Network (CNN) was utilized, and for the recognition of objects, the temporal identifiers were utilized. As per the outcomes, the multi-object tracking accuracy was enhanced by self-supervision. However, the SSL system was prevented from reaching the Supervised-Learning plan by the challenges in the differentiation of baggage from the other structures.

(Ramadan et al., 2022) established the detection and classification plan for human-carrying baggage. The Densely-connected-convolutional Network (DenseNet-161) and the Fit One cycle executed the detection and classification. The system's efficacy was proved by the binary and multi-class classification accuracy. Nevertheless, during training with more model parameters, the system could have caused computation overhead.

(Russel & Selvaraj, 2021) explored a system for the carried objects' recognition in the gait energy image. For improving the single CNN's performance, a parallel deep CNN architecture with customized filters was developed. According to the outcomes, for the recognition of real-time carried objects, the established system attained superior performance. But, owing to objects' shape and motion pattern, the system reliability was limited.

(P. Yang et al., 2021) established a plan for the discovery of co-location patterns from massive spatial datasets with or without rare features. For detecting the co-locations, an interesting measure named Weighted Participation Index (WPI) was established. The demonstrated scheme's effectiveness and scalability were proved by the experimental outcomes. Nevertheless, in the system, the prevalence index's determination was a difficult process.

(Bao et al., 2021) propounded a mining technique for the recognition of Super Participation Index-closed (SPI-closed) co-location patterns. For discovering the SPI-closed Co-location Patterns (SCPs), a hash table formed with the correct neighboring cliques was utilized. As per the experiments, the SCP technique was more flexible compared to similar techniques. But, the time efficiency of the Maximal Co-located Patterns (MCPs) was not satisfactory.

(Tran et al., 2021) presented a spatial co-location pattern mining technique centered on overlapping cliques. For the recognition of spatial co-location pattern mining, this framework introduces a two-level filter mechanism, where the first level is a feature type filter and the second level is a neighboring instance filter. According to the experimental study, more than any other algorithm, the established approach responded to user requirements quickly.

However, when storing the instances in the hash map structure, the system could suffer from high storage overhead.

(Mehta & Kaur, 2020) implemented an automated technique for detecting and labeling abandoned objects from videos. In the implemented technique, for object detection, the Point-Tracker algorithm with the generalized Region Of Interest was utilized. As per the outcomes, the detection system's accuracy was enhanced by the generalized ROI. Yet, this approach wasn't reliable for effectively abandoned object detection without the scene analysis from the key-frames of the video.

(Anwar et al., 2020) recommended a pattern mining technique utilizing the Location-Based online Social Networks (LBSN) data for inferring types of diverse locations. The frequent co-located users and user components were mined. Subsequently, for categorizing the locations, temporal pattern analysis was done with Frequent Pattern (FP) growth algorithm. The model's efficacy regarding mean reciprocal rank was proved by the experimental results. However, the tree formation in the FP-growth was a difficult process, which would elevate the time for the pattern analysis.

(Zhou et al., 2021) suggested the Maximal Instance Algorithm (MIA) for the spatial co-location patterns' fast mining. For finding the maximum instances from a spatial dataset, a row instance tree was performed. Next, for identifying the co-location patterns, MIA with no minimal join operation was established. As per the experimental evaluation, the MIA achieved superior performance on running time. However, as the MIA model was applicable only to the static co-located pattern, it was unreliable.

(Thenmozhi & Kalpana, 2020) employed moving object detection in a video surveillance system centered on Adaptive Motion Estimation (AME) and Sequential Outline Separation (SOS). The approaches utilized for identifying moving items and classifying moving items are the AME and SOS. During the experimental evaluation, the SOS and AME approaches attained superior outcomes for accuracy. But, the time taken for the computation was more than the analogized approach for the video data.

3. PROPOSED METHODOLOGY FOR THE CO-LOCATED OBJECT RECOGNITION WITH STATUS ESTIMATION

In the application of determining the status of the object, the recognition of co-located objects in the video helps. But, owing to the lack of detailed analysis of all co-located instances, existing works are limited in the determination of co-located object status recognition. Hence, this study proposes a patch-centric Max-GRU scheme for the recognition of the co-located object and its status in the video. The proposed system's architecture is provided below:

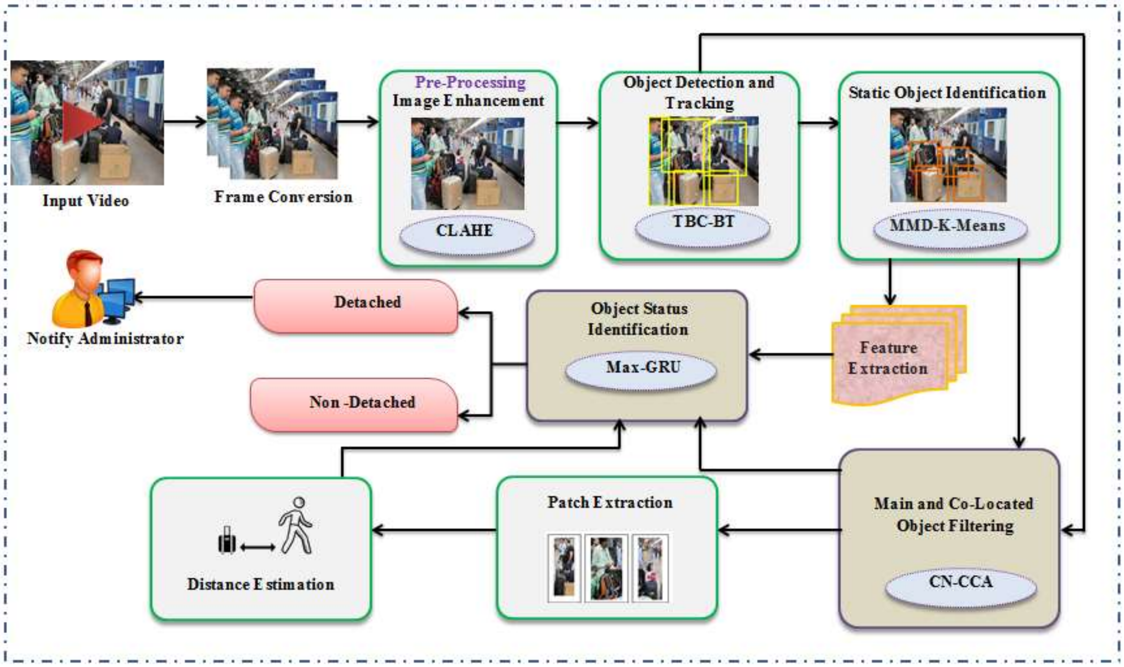


Figure 1: Architecture of the proposed scheme

3.1 Input data

Primarily, the proposed system takes input surveillance video from the railway terminals and converts it into frames for the recognition of co-located objects. In this, the frames $\left(S_n\right)$ acquired from the input video are signified as,

$$S_n = \{S_1, S_2, ..., S_z\} \tag{1}$$

Where, S_z depicts the z^{th} frame obtained from the video.

3.2 Pre-processing

The CLAHE approach is utilized for improving the quality of the video frames $\left(S_n\right)$ and enhancing the recognition rate of the objects in the frames. Utilizing CLAHE has the main advantage of reducing the over-amplification of contrast in the homogeneous regions of an image. Rather than the entire image, it operates on the smaller portion of an image named tiles. Next, to eliminate the artificial boundaries and improve the image contrast, the nearest regions are merged utilizing bilinear interpolation.

Primarily, the frames $\left(S_n\right)$are partitioned into tiles of size $u \times v$ and the histogram of each region is calculated centered on its gray level. Subsequently, to make the image appear more natural, the intensity of each region is enhanced based on the Rayleigh transform. The density of each intensity value $(d_x \in u, v)$ grounded on Rayleigh transform is termed as,

$$d_x = d_{\min} + \sqrt{2\lambda^2\left[-\ln(-\Theta(x))\right]} \tag{2}$$

Here, the lower bound of the pixel (u, v) value is $d_{\min} \in u, v$, Rayleigh scaling parameter is depicted as λ, and the transfer function is signified as $\Theta()$. The contrast of the image will be enhanced by a higher λ value. The enhanced frame thus obtained is denoted as e_n.

3.3 Object detection and tracking

Next, the frames are improved; the objects in the frames are detected and tracked with the TBC-BT technique, which relies on the motion of objects between successive frames. Owing to its advantage in object detection with tracking behavior even in occluded scenarios, Byte Tracking (BT) is chosen. However, to detect the objects with only the feature distance, BT takes more time. Hence, for solving this issue, Tukey's Biweight Correlation (TBC) technique is included in BT. The steps in TBC-BT are explained below:

Object detection: Utilizing the bounding boxes, the TBC-BT uses YOLO for the detection of the objects in the frames e_n. The YOLO v8 predicts multiple bounding boxes per grid cell in which the head predicts the bounding box of an object and give the center coordinates (p_c, q_c), width (W), and height (H). Hence, the final predicted bounding box (ρ) is estimated as,

$$\rho_t = \varsigma(\Re_m) + B_t \tag{3}$$

$$\rho_l = \varsigma(\Re_n) + B_l \tag{4}$$

$$\rho_W = \Re_W . \exp(A_W) \tag{5}$$

$$\rho_H = \Re_H . \exp(A_H) \tag{6}$$

Here, the predicted offsets from the anchor box are depicted as $\Re$, the coordinates of the top-left corner of the bounding box are signified as B_t, B_l, the sigmoid activation is indicated as ς, and the anchor box's height and width are depicted as A_W, A_H.

After determining the bounding boxes for an object, the overlapped bounding boxes are detected using the Intersection of the union (U) as,

$$U = \frac{\beta_\rho \cap T_\rho}{\beta_\rho \cup T_\rho} \tag{7}$$

Here, the predicted and target bounding boxes are indicated as β_ρ, T_ρ. In this, the bounding box with the highest U value is chosen as the detected objects' bounding box. Hence, the bounding boxes with diverse scores are signified as,

$$b_v = \{b_1, b_2, ..., b_k\} \tag{8}$$

Here, the k^{th} detected objects' bounding box is signified as b_k.

Object tracking: In TBC-BT, after the object detection, the objects are tracked based on the association process, which is a matching process by estimating the similarity with tracklets. In TBC-BT, the bounding box with a high score and low score are separated by the TBC (α) and Euclidean distance (dis) as,

$$\alpha = b_{high}(1 - b_v). \tag{9}$$

$$dis = \sqrt{\sum_{v=1}^{k} \left| b_{high} - b_v{}^2 \right|} \tag{10}$$

Here, the bounding box with the highest score is signified as $b_{high} \in b_n$, which is determined in the TBC-BT. The boxes with high α value and the minimum dis value are considered as high-scored tracklets and vice-versa. Next, using the Kalman filtering, the bounding boxes with all the score values are associated and tracked centered on the area difference $(\delta(\ ))$ as,

$$\delta(v, v+1) = \frac{\left|\varepsilon_n^v - \varepsilon_{n+1}^v\right|}{\max\left|\varepsilon_n^{v+1} - \varepsilon_{n+1}^{v+1}\right|} \tag{11}$$

Here, the tracking window's area is depicted as ε, which indicates the degree of the window's deformation. The smaller ε value signifies a closer description of the two objects' shapes. The $v^{th}, v+1^{th}$ objects in the frames $n, n+1$ are depicted as $(v, v+1)$, $(n, n+1)$, correspondingly. Hence, the objects tracked frames are acquired, which is represented mathematically as,

$$O_n = \{O_1, O_2, \ldots\ldots, O_z\} \tag{12}$$

Here, the z^{th} object detected and tracked frame is given as O_z. Based on motion, the TBC-BT detects and tracks the objects; hence, only the co-located objects in motion are detected, while static objects were not identified. MMD-K-Means is used here to detect static objects.

3.4 Static object detection

The static object detection (i.e. separation of co-located objects) is performed using the MMD-K-Means algorithm to determine all the co-located objects in the video frame. In this, due to the efficient classification of the stationary and nan-stationary pixels, the K-Means algorithm is chosen. However, the identification efficiency decreases without the proper centroid initialization. Hence, for solving this issue, the Min-Max Distance (MMD) between the histogram of pixels is estimated for the centroid selection, and is explained below:

Initialization: The O_n frame is fed as input to the MMD-K-Means clustering for which the κ clusters are also initialized and are depicted as,

$$\aleph_\kappa = \{\aleph_1, \aleph_2\} \tag{13}$$

Here, the initial clusters generated are signified as $\aleph_1, \aleph_2$.

Centroid estimation: The clusters' centroids (E_κ) are calculated with MMD of pixel histogram as,

$$E_\kappa = |\gamma - \gamma'| \tag{14}$$

Here, the cluster's center is signified as γ', and the pixel in the object detected frame is depicted as $\gamma \in O_n$, which ranges from 0 to 255 pixel values. Obtaining the biggest pixel value, γ' is determined. Hence, the final cluster centroid C_κ is determined by estimating the MMD value and obtaining the minimum E_κ value.

Assigning clusters: Subsequent to the centroid (C_κ) calculation, the pixel points are assigned to the clusters utilizing the distance (ζ) formula,

$$\zeta = \left(\sum_{\gamma=1}^{b} C_\kappa - p_\gamma\right) \tag{15}$$

Here, the pixel data value in the frame is depicted as p_γ, and the total number of pixels in the frame is signified as b. The data points are assigned to the respective clusters with the minimum distance based on this distance calculation. Next, the final static $(\aleph_{static})$ and dynamic $(\aleph_{dynamic})$ pixel objects in the frame are clustered as,

$$\aleph_{\kappa} = \{\aleph_{static}, \aleph_{dynamic}\} \tag{16}$$

~~**Pseudocode of proposed MMD-K-Means**~~

```
Input: Objects tracked frames $\{O_1, O_2, ....., O_z\}$
Output: Clusters $\aleph_{\kappa}$
Begin
    Initialize frames , pixels $\gamma, \gamma'$
    For frame $O_n$ do
        Initialize clusters $\{\aleph_1, \aleph_2\}$
        Initialize centroids $C_{\kappa}$ with MMD
            For each pixel points $\gamma$ in the frame $O_n$
                Calculate distance using $C_{\kappa} - p_{\gamma}$
            End for
            Assign pixel to $\aleph_{\kappa}$ with minimum distance
    End for
    Return clusters $(\aleph_{\kappa})$
End
```

3.5 Main and co-located object filtering

Subsequently, the CN-CCA process determines the main object corresponding to the co-located object in the frame O_n with the $\aleph_{static}$ as a reference. In this, the CCA is utilized for the advantage of evaluating the connected class with the main class. However, the CCA has the random usage of midpoints that degrades the interconnection recognition of objects. Hence, for overcoming this, the Cyclic Neighborhood (CN) technique is utilized in CCA.

The CN-CCA takes the input pixels (η_j) in $\aleph_{static}$, and checks whether the η in O_n connects with any other pixels in the main object pixel (M_j), which is present inside the bounding box. This process is performed by the CN approach as,

$$\mu_{dist}(\eta_j, M_j) = \left[\min\left(\left|\eta_j - M_j\right|, 1 - \left|\eta_j - M_j\right|\right)\right]^2 \tag{17}$$

Here, the distance between the current and the neighboring pixel value is depicted as μ_{dist}, and when the minimum μ_{dist} is acquired, the object corresponding to it is labeled as the main object $(lab(M_j))$. The assignment of labeling is given as,

$$Lab(M_{j+1}) = \begin{cases} 0, & V(M_{j+1}) = 0 \\ lab(M_j), & V(M_{j+1}) = V(M_j) \\ lab(M_j) + 1 & V(M_{j+1}) \neq 0, V(M_j) \end{cases} \tag{18}$$

Where, the label of the objects is depicted as lab, and the value of the pixels is signified as $V(\ )$. In the CN-CCA approach, when calculating the distance of the pixel values to the neighbor pixels, if the non-zero pixels in the co-located object are connected to another pixel of the main connected object, it is filtered and labeled. Otherwise, the non-zero pixels are assigned with another label $lab(M_j)+1$ during the first pass.

Next, after all the objects in the frame O_n are labeled, the second pass is done for checking whether the $Lab(M_{j+1})$ has equivalent labels as other objects' pixels and solve them. Subsequently, the final filtered main-co-located pair is acquired based on this process. The filtered main and co-located pair (P_i) is given as,

$$P_i = \{P_1, P_2, \ldots., P_q\} \tag{19}$$

Here, the q^{th} filtered main- and co-located object is signified as P_q and D_n is the acquired frame.

3.6 Patch Extraction

To analyze the trajectory of the co-located objects individually after filtering the main and co-located object in the frame, each main- and co-located object is converted to patches. After the patch conversion, to avoid missing the track of a single co-located object, the trajectory of the co-located object is observed in each patch. The converted patches are indicated as,

$$\Im_i = \{\Im_1, \Im_2, \ldots, \Im_q\} \tag{20}$$

Here, $\Im_q$ denotes the patch for the q^{th} main-colocated object pair.

3.7 Distance Estimation

After the patches are obtained, based on the radius, the distance between the main and the collocated objects in the frames is determined. The distance is estimated as,

$$\wp_n^i = \sqrt{\left(o_{co}^i - o_{main}^i\right)} \tag{21}$$

Here, the distance calculation in the n^{th} frame for i^{th} patch is depicted as $\wp_n$, and the main and co-located object in the i^{th} patch is signified as o_{co}^i, o_{main}^i. Hence, the object is considered as detached object when the $\wp_n$ exceeds the threshold Ω.

3.8 Feature extraction

The SURF and HOG patterns of the co-located objects are also given to the object status identifier to differentiate the main and co-located object during the co-located object status identification and to improve the accuracy of the status identification, which are given below:

Speeded Up Robust Features (SURF): For the feature extraction in the image, SURF uses square-shaped filters. Using SURF descriptors (s), the SURF feature is extracted and is given as,

$$s = \varpi(xx, yy) * G \tag{22}$$

Here, the isotropic and separable Gaussian kernel is signified as G, and the pixel co-ordinates in the frame O_n are depicted as xx, yy. A feature set is obtained with this descriptor, and it is denoted as $\{N_{feat}\}$.

Histogram Of Gradients (HOG): A feature descriptor to determine the co-occurrence counts of gradient orientation in the portion of a frame is named HOG.

In this, by combining the angle and magnitude obtained from the image, the gradients are obtained. The angle of gradients on the x-axis $(xx_G)$and y-axis $(yy_G)$is calculated as,

$$\phi = Sin^{-1}\left(\frac{\left(\frac{yy_G}{xx_G}\right)}{\sqrt{1+\left(\frac{yy_G}{xx_G}\right)^2}} \right) \tag{23}$$

Here, the gradient angle is signified as ϕ . The magnitude of the gradient $|G|$ is found by the given equation,

$$|G| = \left((xx_G)^2 + (xx_G)^2\right)^{1/2} \tag{24}$$

At last, the obtained features of the co-located objects are indicated as,

$$\Gamma_r = \{\Gamma_1, \Gamma_2, ..., \Gamma_\vartheta\} \tag{25}$$

Here, the ϑ^{th} feature is depicted as Γ_ϑ .

3.9 Co-located object status identification

After the features are extracted, the features $\Gamma_r$along with the estimated distance$(\wp_n)$, the filtered object frame(D_n), and the status of the co-located object are identified. In this, the Max-GRU classifier is utilized for status identification. Due to its efficiency in video applications, the GRU is considered. However, the GRU has the drawback of increased time-consuming with increased model complexity. Hence, for overcoming this drawback, the Maxout activation is utilized in GRU. Figure 2 gives the architecture of the proposed Max-GRU as,

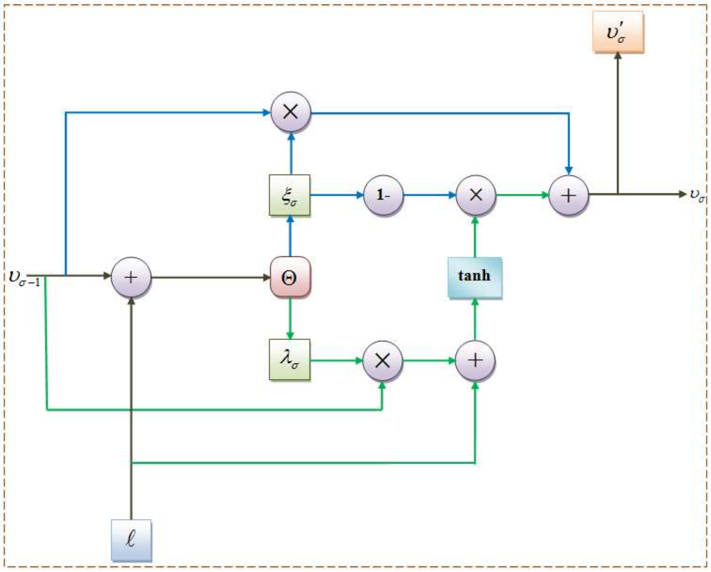


Figure 2: Structure of proposed Max-GRU

Input: The combination of Γ_r distance$(\wp_n)$ and the filtered object frame(D_n) is the input of the Max-GRU, which is denoted as,

$$\ell = \langle \Gamma_r, \wp_n, D_n \rangle \tag{26}$$

Two inputs that include the current input $(\ell_{\tau-1})$and the previous hidden state as vectors are taken by the reset gate and Update gate in the Max-GRU at each time. By performing element-wise multiplication between the obtained vector and the respective weights (ω) for each gate, the output of each gate is obtained. The operations of the Reset gate and Update gate are given as,

Reset Gate: In the reset gate, the linear sum betwixt the freshly computed state and the prevailing state with the bias parameter is calculated, and is mathematically indicated as,

$$\lambda_\sigma = \Theta(\omega_\lambda * \ell + \omega_\lambda \upsilon_{\sigma-1} + v_\lambda) \tag{27}$$

Here, λ_σ depicts the reset gate determined by the input vector and the information at the previous memory gate $(\upsilon_{\sigma-1})$, the bias value is depicted v, and the Maxout activation function is signified as $\Theta(\ )$. The Maxout activation of gates is given as,

$$\Theta(\lambda_\sigma) = \max(\omega_\lambda^T . \ell + v_\lambda) \tag{28}$$

Here, the transpose function is signified as $(\ )^T$.

Update Gate: The update gate determines how much of the earlier information from previous time $(\sigma - 1)$ steps are needed to be updated. The update gate can be calculated as,

$$\xi_\sigma = \Theta(\omega_\xi * \ell + \omega_\xi \upsilon_{\sigma-1} + v_\xi) \tag{29}$$

$$\Theta(\xi_\sigma) = \max(\omega_\xi^T . \ell + v_\xi) \tag{30}$$

Here, the update gate computed by the newly computed state is depicted as ξ_σ .

To pass the relevant information, the current memory content $\upsilon_{\sigma-1}$ needs the reset gate (λ_σ) , while the final memory unit (υ_σ)holds the information for the current unit and passes it to the network utilizing the update gate. The memory in the Max-GRU is calculated as,

$$\upsilon_\sigma = (1 - \xi_\sigma)\upsilon_{\sigma-1} + \xi_\sigma \upsilon_{\sigma-1} \tag{31}$$

$$\upsilon'_\sigma = \tanh(\omega_\upsilon \cdot \ell + \omega_{\upsilon,\upsilon'}(\lambda_\sigma * \upsilon_{\sigma-1}) + v_\upsilon) \tag{32}$$

Here, the hyperbolic tangent activation function is signified as tanh. Hence, from the final memory unit, the final output class (υ'_σ) in the Max-GRU is estimated.

Pseudocode of proposed Max-GRU

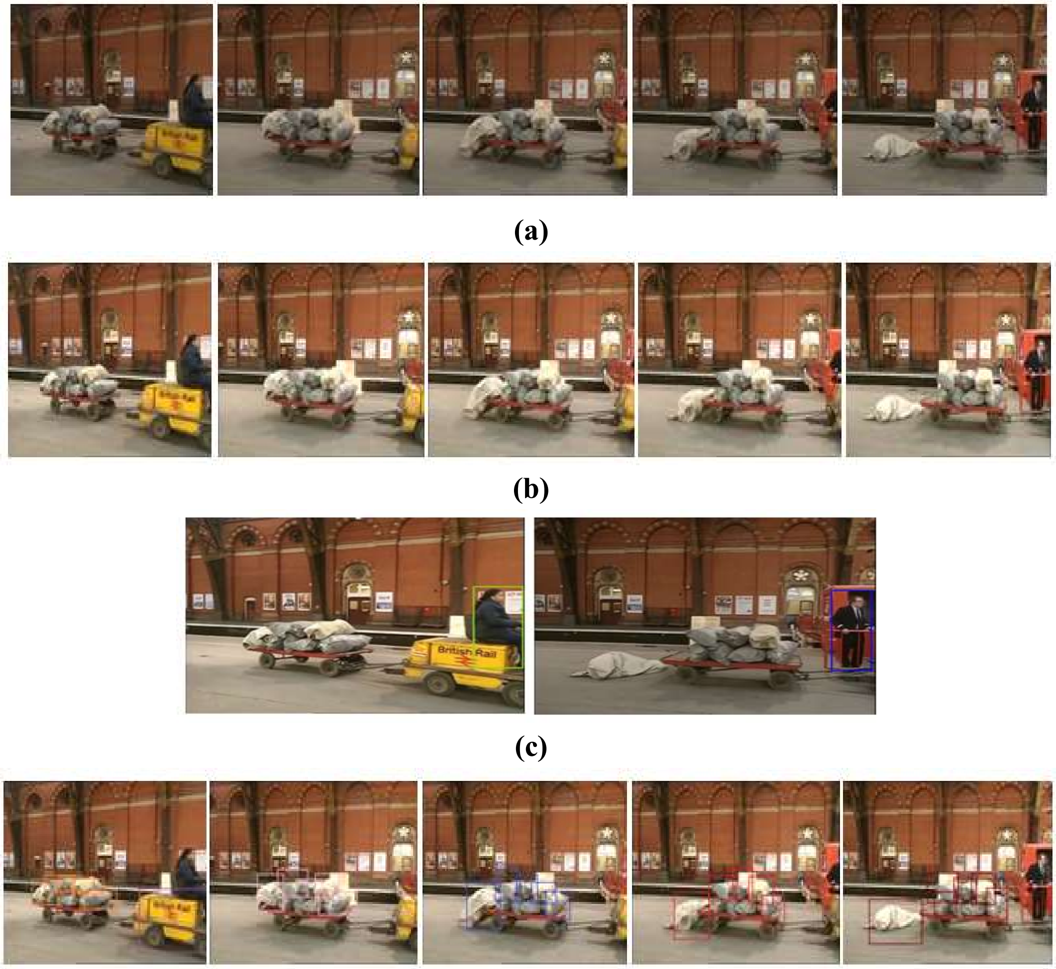
```
Input: $\ell = \langle \Gamma_r, \wp_n, D_n \rangle$
Output: Output class
Begin
    Initialize input data $(\ell)$, parameters $(v, \omega)$,number of layers
    Set initial memory content $\upsilon_\sigma = \upsilon_0$
    For time step $(e)$do
        Estimate update gatewith $\Theta(\omega_\xi * \ell + \omega_\xi \upsilon_{\sigma-1} + v_\xi)$
        Activate update gate using maxoutactvation $(\Theta)$
        Calculate reset gate using $(\omega_\xi, \ell)$ and $(\omega_\xi, \upsilon_{\sigma-1})$
```

```
            Compute final memory content $(\upsilon'_{\sigma})$
        End for
        Return output $\upsilon'_{\sigma}$
End
```

In this, the output classes are the status (i.e. detached or not-detached) of the object. After obtaining the status of the Max-GRU, if the co-located object is found to be detached, the information is notified to the surveillance administrator. Through a public announcement, the administrator can notify the users.

4. RESULTS AND DISCUSSION

This section evaluates the proposed approach's performance centered on the performance metrics and analogizes the outcomes with the prevailing approaches. Utilizing publically gathered surveillance video data in railway stations, the proposed approach is deployed in the working platform of PYTHON. Figure 3 gives the sample image outcomes.

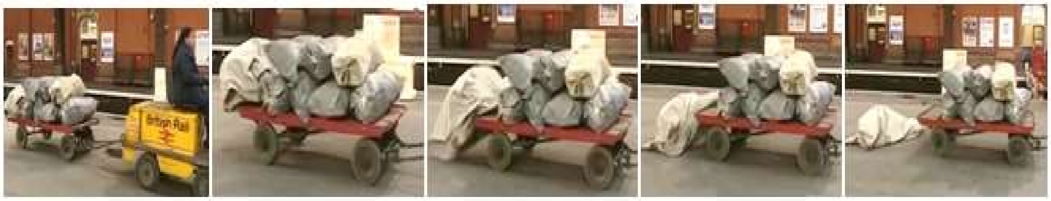


(a)

(b)

(c)

(d)



(e)

Figure 3: Sample image results for (a) frame conversion (b) Image enhancement (c) detected and tracked objects (d) static object recognition and € patch extraction

4.1 Performance Analysis of Co-located Object Recognition

In this subsection, the proposed MMD-K-Means approach’s performance for identifying the co-located objects is analogized with the existing algorithm like K-Means, K-Medoid, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) and Clustering LARge Applications (CLARA).

Table 1: Recognition Rate Analysis

Techniques	**Recognition Rate (%)**
Proposed MMD-K-Means	97.92
K-Means	95.2
K-Medoid	93.23
BIRCH	91.04
CLARA	86.49

The recognition rate achieved by several approaches is displayed in Table 1. Here, when analogized to the prevailing K-means system having a 95.2% recognition rate, the proposed MMD-K-Means system’s recognition rate enhanced by about 2.72%. Similarly, the proposed algorithm surpassed other prevailing systems also. This result shows that the clustering approach’s performance has been enhanced by the proposed MMD for centroid selection, hence leading to co-located objects’ better recognition rate.

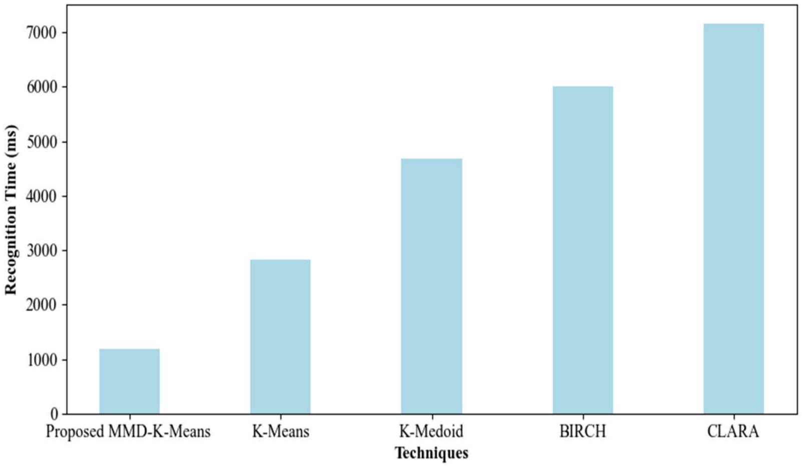


Figure 4: Performance Analysis of proposed MMD-K-means

Figure 4 illustrates the time taken by the proposed and baseline systems to recognize the objects. The static and co-located objects were recognized by the proposed algorithm in 1184ms; however, for recognizing the objects, the prevailing K-Means needed 2834ms, K-Medoid needed 4691ms, BIRCH needed 6015 ms, and CLARA needed 7153 ms. Thus, the MMD-centric measurement usage aided in effective clustering with less time.

4.2 Performance Analysis of object status Identification

The proposed Max-GRU classifier's performance for identifying the object status is examined with the prevailing approaches like GRU, LSTM, Bidirectional LSTM (Bi-LSTM), and Recurrent Neural Networks (RNN).

Table 2: Performance Measure of Max-GRU

Techniques	Accuracy (%)	Precision (%)
Proposed Max-GRU	98.13	97.34
GRU	96.98	95.55
LSTM	94.34	93.14
Bi-LSTM	91.84	91.77
RNN	89.26	89.9

The proposed classifier's performance centered on accuracy and precision is shown in Table 2. For effective object status identification, accuracy and precision values should be higher. Consequently, the objects' status was identified by the proposed classifier with 98.13% accuracy and 97.34% precision, while the prevailing approaches render relatively lower performance. This exhibits the proposed system's superiority in recognizing the objects' status. Generally, when compared to prevailing approaches, the proposed classifier provided superior performance.

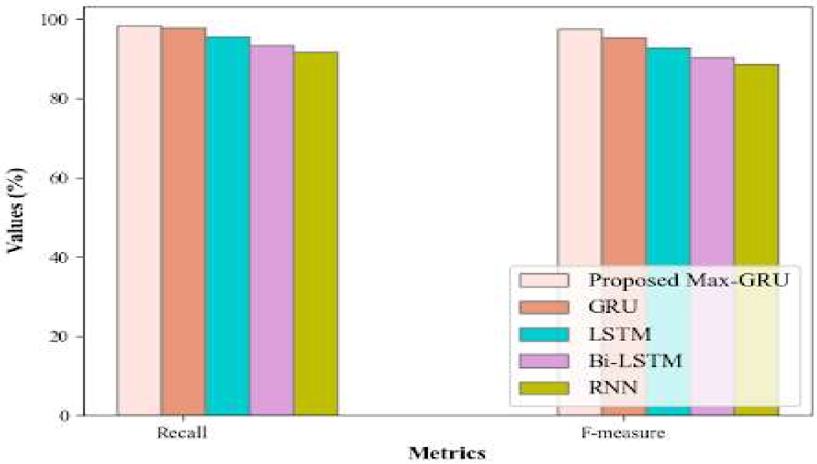


Figure 5: Performance Validation based on Recall and F-measure

The proposed and prevailing approaches' performance are investigated regarding Recall and F-measure in Figure 5. Recall and F-measure attained by the proposed technique are 98.37% and 97.64%, correspondingly, while the prevailing GRU, LSTM, Bi-LSTM, and RNN approaches proffered lower values. The proposed classifier's performance has been improved to a greater extent with the inclusion of the maxout activation function in GRU.

Table 3: Comparison of Proposed and Existing Classifiers

Techniques	Training Time (ms)	Specificity (%)
Proposed Max-GRU	9920	96.94
GRU	11354	94.11
LSTM	12943	92.34
Bi-LSTM	14647	89.99
RNN	15761	87.42

The training time and specificity of the proposed Max-GRU algorithm and the prevailing classifiers are exposed in Table 3. The proposed system consumed 9920 ms for effectively training the system with minimum error, while the prevailing systems like GRU, LSTM, Bi-LSTM, and RNN needed 11354 ms, 12943 ms, 14647 ms, and 15761 ms, correspondingly. Likewise, when analogized to the prevailing GRU, the proposed object status identifier's specificity exhibited a 2.83% improvement. The proposed system's efficacy is exposed clearly by this analysis.

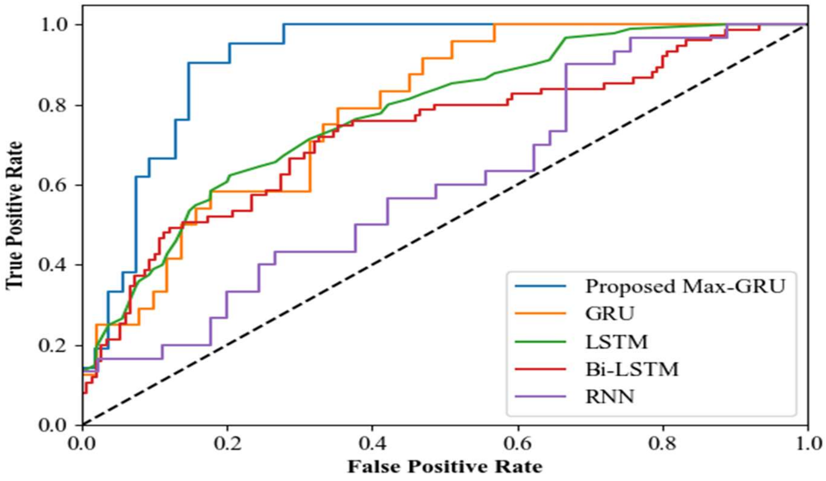


Figure 6: ROC Analysis

For discriminating between detached and non-detached objects, Receiver Operating Characteristic (ROC) curve determines the accuracy level of the object status identifier (Max-GRU). The ROC curve is the curve generated by the plot between True Positive Rate (TPR) and False Positive Rate (FPR) across varying cut-offs. Better performance is signified by a ROC curve lying above the diagonal level and a monotonically increasing curve. Hence, according to Figure 6, a high saturation (0.98) was attained by the curve for the proposed Max-GRU, while other prevailing classification systems exhibit an increasing curve below the proposed level (GRU-0.96, LSTM-0.94, Bi-LSTM-0.91, RNN-0.89). Hence, when compared to the existing ones, the higher value achieved by the proposed system displays significant improvement in performance.

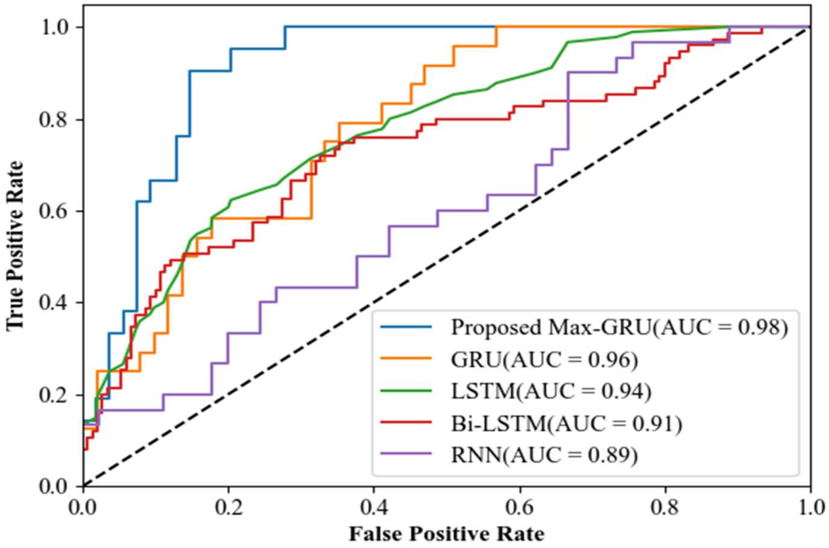


Figure 7: AUC Analysis

The measure of sensitivity and specificity that evaluates the classification outcome is named the Area Under the ROC Curve (AUC). The correct classification of all the inputs with their exact target is described by the highest AUC. With this point from Figure 7, the proposed Max-GRU achieves an AUC value of 0.98, which signifies that it classifies the detached and non-attached objects correctly with maximum probability. On the other hand, when analogized to the proposed system, other conventional classification approaches achieve less AUC value. This shows the proposed classifier's effectiveness for object status identification.

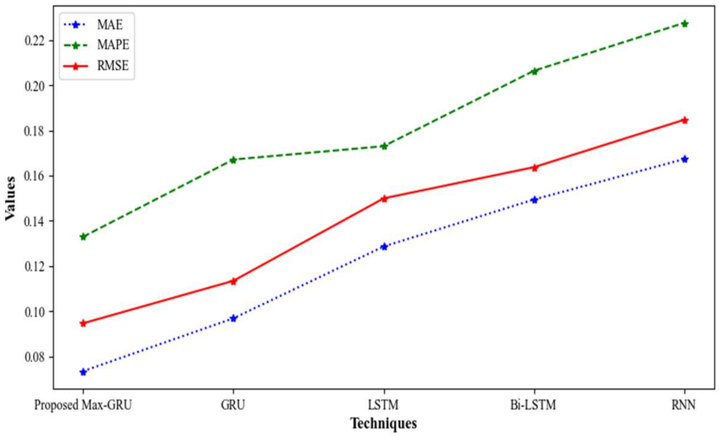


Figure 8: Performance Analysis of Proposed Max-GRU

The proposed and existing classifiers' error metrics Mean Absolute Error, Mean Absolute Percentage Error, and Root Mean Square Error (MAE, MAPE, and RMSE) are illustrated in the above figure 8. Superior experimental outcomes are demonstrated by the lower error values. Centered on the magnitude of error produced by the system, the aforementioned error metrics evaluate the prediction rate. Thus, the graph shows that the proposed classifier attained 0.0734, 0.1329, and 0.0946 for MAE, MAPE, and RMSE, respectively. Similarly, the MAE, MAPE, and RMSE attained by the prevailing GRU, LSTM, Bi-LSTM, and RNN classifiers are (0.0968, 0.1671, 0.1134), (0.1286, 0.173, 0.1499), (0.1494, 0.2064, 0.1637), and (0.1673, 0.2276, 0.1847), correspondingly. On analogizing these values, the proposed system has lower values compared to other prevailing approaches. Hereafter, it is concluded that the proposed study achieves superior accuracy with lower errors and proves to be more proficient.

4.3 Performance Analysis of Object Filtering

Here, the proposed co-located objects filtering system is contrasted with the baseline approaches like CCA, Aggressive Relabeling (AR), Rosenfeld Quick Union (RQU), and Contour Tracing (CT).

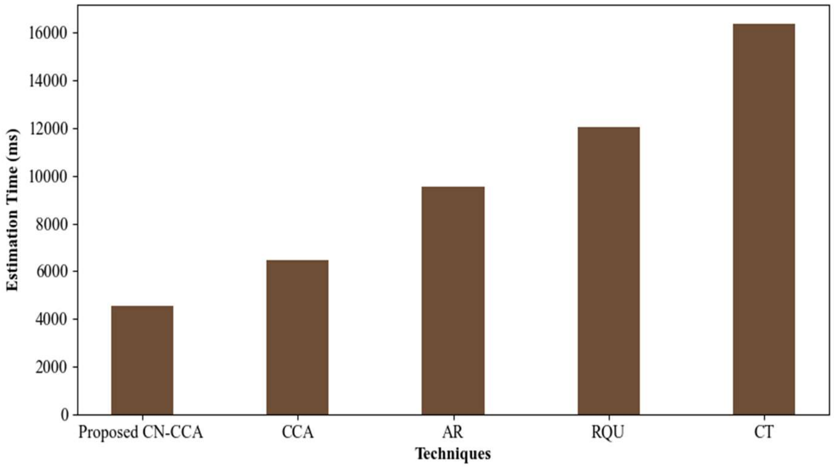


Figure 9: Estimation Time

The performance analysis of the proposed and existing systems grounded on the time consumed for estimating the connected component is indicated in Figure 9. The proposed CN-CCA approach's estimation time is 4535ms. Conversely, the prevailing approaches, such as CCA, AR, RQU, and CT consumed 6487ms, 9543ms, 12043ms, and 16349ms, correspondingly. Better performance in filtering objects with minimum time has been displayed by the CN inclusion in the proposed approach.

4.4 Performance Analysis of Object Tracking

Grounded on Multiple Object Tracking Accuracy (MOTA), the performance of the proposed object tracking algorithm TBC-BT is analogized with the baseline techniques, namely BT, deep Simple Online Real-time Tracking (SORT), SORT, and Kalman Filter (KF).

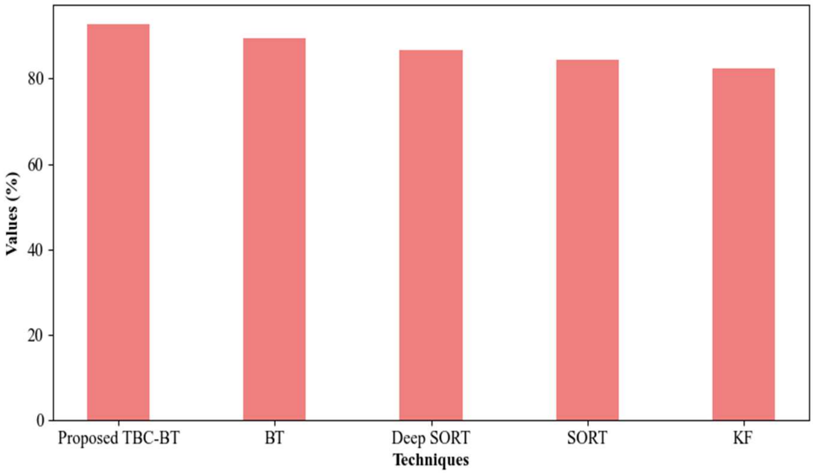


Figure 10: MOTA analysis

Figure 10 depicts the MOTA acquired by the proposed TBC-BT algorithm and prevailing techniques. The multi-object tracking process' efficiency is determined by MOTA analysis. Figure 10 illustrates that the MOTA acquired by the proposed system is 92.64. However, the MOTA values acquired by the prevailing systems are lower (BT-89.43%, Deep SORT-86.75%, SORT-84.33%, and KF-82.46%). Hence, TBC in the conventional BT algorithm has exhibited effective performance in object tracking.

4.5 Comparative analysis

Here, for proving the proposed system's efficiency, the efficiency of the proposed and the existing studies of (Siddique & Medeiros, 2022), (Mehta & Kaur, 2020), and (Thenmozhi & Kalpana, 2020) are contrasted.

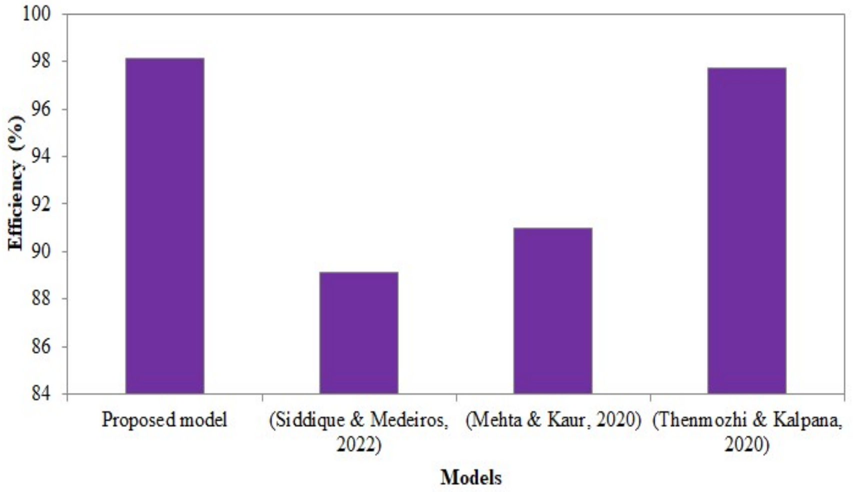


Figure 11: Comparative analysis

A comparative analysis of the proposed and the existing systems regarding efficiency is given in Figure 11. In this, for the co-located object status identification, the efficiency is evaluated with the recognition of co-located objects. In this, the (Siddique & Medeiros, 2022), (Mehta & Kaur, 2020), and (Thenmozhi & Kalpana, 2020) systems provide 9.03%, 7.13%, and 0.41% less efficiency compared to the proposed system only for the detection of objects. However, the proposed system attained higher efficiency compared to the other systems even with the status identification. Hence, higher efficiency is due to the higher recognition rate and patch-wise analysis in the proposed model. This makes the proposed system more appropriate for co-located object recognition and status identification.

5. CONCLUSION

The co-located objects in railway terminal videos for the recognition of object status are identified in this paper. In this, MMD-K-Means and TBC-BT are used for the co-located object recognition, and the Max-GRU model was introduced for the co-located object status recognition. Next, patch-wise analysis is performed for covering all the co-located object instances in the video. Subsequently, the proposed system was experimentally analyzed and verified in contrast to the prevailing approaches. In 1184ms, the proposed MMD-K-Means attained a 97.92% co-located object recognition rate. Next, the Max-GRU acquired 98.13% identification accuracy, and for other performance metrics also, it acquired superior outcomes. Subsequently, the proposed TBC-BT and CN-CCA also attained superior results than the analogized approaches. Lastly, the efficacy of the proposed system over the other baseline systems was proved by the comparative analysis. But, the proposed system is executed with a video acquired with a single surveillance video, where when the co-located object moved to the next location, it can’t be detected. Hence, for better-co-located object status recognition, multiple successive surveillance videos for a single railway terminal can be utilized in the future.

REFERENCES

Alzaabi, A., Talib, M. A., Nassif, A. B., Sajwani, A., & Einea, O. (2020). A Systematic Literature Review on Machine Learning in Object Detection Security. *2020 IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020*, *November*, 136–139. https://doi.org/10.1109/ICCCA49541.2020.9250836

Anwar, T., Liao, K., Goyal, A., Sellis, T., Kayes, A. S. M., & Shen, H. (2020). Inferring Location Types with Geo-Socialoral Pattern Mining. *IEEE Access*, *8*, 154789–154799. https://doi.org/10.1109/ACCESS.2020.3018997

Bao, X., Lu, J., Gu, T., Chang, L., Xu, Z., & Wang, L. (2021). *Mining Non-Redundant Co-Location Patterns*. 1–14.

Dogariu, M., Stefan, L. D., Constantin, M. G., & Ionescu, B. (2020). Human-Object Interaction: Application to Abandoned Luggage Detection in Video Surveillance Scenarios. *2020 13th International Conference on Communications, COMM 2020 - Proceedings*, 157–160. https://doi.org/10.1109/COMM48946.2020.9141973

Elhoseny, M. (2020). Multi-object Detection and Tracking (MODT) Machine Learning Model for Real-Time Video Surveillance Systems. *Circuits, Systems, and Signal Processing*, *39*(2), 611–630. https://doi.org/10.1007/s00034-019-01234-7

Jayaram, C. V., & Bhajantri, N. U. (2019). A Review Study: Recognition of Co-located Pattern in Video Stream. *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, 353–356. https://doi.org/10.1109/ICACCS.2019.8728340

Kaur, J., & Singh, W. (2022). Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, *81*(27), 38297–38351. https://doi.org/10.1007/s11042-022-13153-y

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T. K. (2021). Multiple object tracking: A literature review. *Artificial Intelligence*, *293*, 103448. https://doi.org/10.1016/j.artint.2020.103448

Mehta, M., & Kaur, K. (2020). An Automated Approach to Detect and Label Abandoned Objects from Videos using Generalized ROIs. *2020 IEEE 17th India Council International Conference, INDICON 2020*. https://doi.org/10.1109/INDICON49873.2020.9342599

Popov, A. Y., Ibragimov, S. V., Malyshev, S. A., & Abdurakhmanova, R. A. (2021). Multiple Objects Association System for the Smart City. *Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2021*, 2211–2216. https://doi.org/10.1109/ElConRus51938.2021.9396415

Ramadan, M. K., Youssif, A. A. A., & El-Behaidy, W. H. (2022). Detection and Classification of Human-Carrying Baggage Using DenseNet-161 and Fit One Cycle. *Big Data and Cognitive Computing*, *6*(4). https://doi.org/10.3390/bdcc6040108

Russel, N. S., & Selvaraj, A. (2021). Gender discrimination, age group classification and carried object recognition from gait energy image using fusion of parallel convolutional neural network. *IET Image Processing*, *15*(1), 239–251. https://doi.org/10.1049/ipr2.12024

Santad, T., Silapasupphakornwong, P., Choensawat, W., & Sookhanaphibarn, K. (2018). Application of YOLO Deep Learning Model for Real Time Abandoned Baggage Detection. *2018 IEEE 7th Global Conference on Consumer Electronics, GCCE 2018*, 114–115. https://doi.org/10.1109/GCCE.2018.8574819

Siddique, A., & Medeiros, H. (2022). Tracking Passengers and Baggage Items Using Multiple Overhead Cameras at Security Checkpoints. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 1–13. https://doi.org/10.1109/tsmc.2022.3225252

Sun, X., He, Y., Ren, T., & Wu, G. (2021). Spatial temporal human object interaction detection. *2021 IEEE International Conference on Multimedia and Expo (ICME)*.

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*. https://doi.org/10.1016/j.micpro.2020.103084

Tran, V., Wang, L., & Zhou, L. (2021). A spatial co-location pattern mining framework insensitive to prevalence thresholds based on overlapping cliques. In *Distributed and Parallel Databases* (Issue 0123456789). Springer US. https://doi.org/10.1007/s10619-021-07333-2

Tsai, W.-J., Huang, Z.-J., & Chen-En Chung. (2020). Joint detection, re-identification , and LSTM in muti-object tracking. *2020 IEEE International Conference on Multimedia and Expo (ICME)*.

Wang, T., Lu, T., Fang, W., & Zhang, Y. (2022). Human- object interaction detection with ratio transformer. *Symmetry*, 1–10.

Wang, W., Zhang, J., Zhai, W., Cao, Y., & Tao, D. (2022). Robust Object Detection via Adversarial Novel Style Exploration. *IEEE Transactions on Image Processing*, *31*, 1949–1962. https://doi.org/10.1109/TIP.2022.3146017

Wu, Y., Wang, D. H., Lu, X. T., Yang, F., Yao, M., Dong, W. S., Shi, J. B., & Li, G. Q. (2022). Efficient Visual Recognition: A Survey on Recent Advances and Brain-inspired Methodologies. *Machine Intelligence Research*, *19*(5), 366–411. https://doi.org/10.1007/s11633-022-1340-5

Yang, P., Wang, L., Wang, X., & Zhou, L. (2021). Efficient discovery of co-location patterns from massive spatial datasets with or without rare features. *Knowledge and Information Systems*, *63*(6), 1365–1395. https://doi.org/10.1007/s10115-021-01559-3

Yang, X., Wei, Z., Wang, N., Song, B., & Gao, X. (2020). A novel deformable body partition model for MMW suspicious object detection and dynamic tracking. *Signal Processing*, *174*, 107627. https://doi.org/10.1016/j.sigpro.2020.107627

Zhang, X. Y., Liu, C. L., & Suen, C. Y. (2020). Towards Robust Pattern Recognition: A Review. *Proceedings of the IEEE*, *108*(6), 894–922. https://doi.org/10.1109/JPROC.2020.2989782

Zhou, G., Li, Q., & Deng, G. (2021). Maximal instance algorithm for fast mining of spatial co-location patterns. *Remote Sensing*, *13*(5), 1–20. https://doi.org/10.3390/rs13050960