**IAES International Journal of Artificial Intelligence (IJ-AI)**
Vol. 13, No. 4, December 2024, pp. 3900~3908
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp3900-3908


# Framework for abnormal event detection and tracking based on effective sparse factorization strategy

**Divyaprabha[1], Guruprasad Seebaiah[2]**
[1]Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology, Tumkur, India
[2]Department of Biomedical Engineering, Sri Siddhartha Institute of Technology, Tumkur, India

**Article Info**

***Article history:***

Received Dec 13, 2023
Revised Mar 28, 2024
Accepted Apr 17, 2024

***Keywords:***

Abnormal events
Computer vision
Mobile object dynamics
Sparse matrix factorization
Video tracking

**ABSTRACT**

The idea of tracking video objects has evolved to facilitate the area of surveillance systems. However, most current research efforts lie in speedy abnormal event detection and tracking of objects of interest tracking. However, the primary challenge is dealing with complex video structures' inherent redundancy. The existing research models for video tracking are more inclined towards improving accuracy. In contrast, the consideration of a more significant proportion of mobile object dynamics, e.g. abnormal events, in motion over the crowded video frame sequence is mainly overlooked, which is essential to study a specific movement pattern of the object of interest appearing in the frame sequence concerning the cost of computation factors. The study thereby introduces a unique strategy of speedy abnormal event detection and tracking, which facilitates video tracking to assess a specific pattern of object of interest movement over complex and crowded video scenes, considering a unique learning-based approach. The extensive simulation outcome further shows that the proposed tracking model accomplishes better tracking accuracy yet retains an optimized computation cost compared to the baseline studies. The computation of video tracking also accomplishes higher detection rates even in the challenging constraints of partial/complete occlusion, illumination variation and background clutter.



***Corresponding Author:***

Divyaprabha
Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology
Tumkur, India
Email: divyasy74@gmail.com

## 1. INTRODUCTION

With the growing security demands, surveillance systems are commonly deployed. The background idea of abnormal event detection also includes an object of interest tracking (OIT) when it comes to analysing the specific movement patterns of objects [1]. The mechanism of identification and tracking of video objects (ITVO) is associated with computer vision with a broader range of applications [1]. The prime notion of ITVO is to extract logical information from streams of video feeds to facilitate better understanding and practical interpretation of dynamic scenes [2]. ITVO is considered one of the most critical applications contributing to video surveillance systems. There has been a growing interest among computer vision researchers towards analysis of densely crowded environment in a video sequence for precise detection of OIT which could appear in the form of an abnormal event. The abnormal event detection and tracking problem from crowded video scenes have been motivated by the ubiquity of surveillance cameras, challenges in crowd modeling and the importance of crowd monitoring for various applications. Here the challenge is not to analyse the normal crowd behaviour but to detect the deviations among OIT movement patterns which differs from the other objects mobility dynamics within the video scenes these are referred as anomalous or abnormal events. However, ITVO has many other wide range of applications, including security applications, autonomous vehicles, human-computer interaction, sports analysis, healthcare, marketing, and retail [3], [4]. Some of the complex problems in visual tracking is people tracking and traffic density estimation [5], [6]. ITVO has also been extensively studied under motion estimation techniques. Various schemes are being evolved to deal with the ITVO aspects and fulfil the requirement of video surveillance applications. Bombardelli *et al.* [7] stated the conventional algorithms developed by researchers are able to tackle only some of the issues in object tracking, however, they still do not provide a foolproof solution. The scope of the study attempts to explore the abnormal events that could take place in the captured video data, and in that aspect, the design objective aims to study and track the behaviour of the object of interest that could be either abnormal movement of non-pedestrian entities in the walkways, or it could be anomalous pedestrian movement patterns. A few instances of abnormalities could occur due to different objects of interest and their anomalous movement patterns, including bikers, skaters, and small carts. However, several challenges are associated with the conventional ITVO study models, which further derive the scope of this research [8].

The extensive review of literature considered analysing a set of related baseline strategies which have mostly considered machine learning and statistical computing-based modeling for video object tracking considering abnormal event detection and tracking and also worked on the similar line of research. The exploration of baseline strategies mostly narrowed down to the abnormal event detection strategies which have been inspired from the sparse dictionary-basis vector learning strategy and utilized non-negative matrix factorization (NMF) methods to improvised the tracking of unusual events and object of interests from the video sequence. The study also explored the related design methodology which have also followed the similar processes of analytical modeling consisting of data acquisition, pre-processing, feature engineering, and many other essential standard steps and customized their own strategic methods for feature engineering and feature learning prospects [9]–[12]. The baseline strategy of Ren *et al.* [13] extensively illustrated the significant and potential advantages of NMF in robust extraction of intrinsic feature attribute structures from the video data and also derive its scope of enhancement towards robust and cost-effective OIT from abnormal events-oriented video scenes. It also shown how useful the NMF is to deal with the decomposition of multivariate data [14]. At present, there exist various forms of related studies which have also focused on developing ITVO models for precise object detection and tracking purpose. Aghili [15] introduced an ITVO framework which adopts the adaptive scheme of kalman filter (KF) for detection and tracking of moving objects over the given video scenes. The design strategy is found suitable for fault tolerant operations even in the presence of challenging constraints. Banerjee *et al.* [16] considered developing an adaptive optimization strategy based ITVO framework towards multiple objects tracking where the strategy considers the involvement of Viterbi algorithm and KF. A specific learning based ITVO framework has been designed considering convolutional neural network (CNN) in [17], [18]. In their study addressed the occlusion problem and introduced a gradient model of multi-channel ITVO framework. The experimental results claim to attain better consistency and classification performance. The adoption of CNN is also found in the work of Kim and Ha [19] where the strategy mostly covers ITVO of an object considering stable foreground. However, the computational aspect of the model is mostly overlooked. Li *et al.* [20] addresses the problem of saliency in ITVO and further introduces a semantic based approach to overcome the challenges during object tracking scenario. Su *et al.* [21] utilized CNN for key point detection on ITVO. A similar approach of deep-learning based approach for selection and extraction of local features also could be seen in [22].

The study explores the research trend in ITVO and identifies the gap that restrict the evolution of this research track. The identified research problems in the existing mechanism of object detection and tracking viz: i) majority of the studies are mostly inclined towards single moving object detection and tracking where its associated dynamics were not much studied under motion estimation techniques; ii) it can also be seen that even though significant research effort had been laid towards designing an efficient ITVO system but very lesser existing studies have addressed the challenge of discriminating a specific movement pattern of moving objects from other moving objects considering dynamic crowd oriented video scenes analysis; iii) even though majority of the research approaches of ITVO explores different learning-based strategies but their focus is mostly on improving the tracking accuracy rather considering the aspects of cost of computation. It is also observed that very few strategies have explored the compelling ideas of NMF and least-square approach to explore the sparsity factors of features and learns from the features to improve the performance of feature extraction modeling, whereas these methods are suitable for dimension reduction and appropriate feature computation considering lexicon vector based dictionary construction modeling unlike other computationally expensive feature extraction methods; and iv) also in the existing studies maximum utilization of learning-based methodologies are found which demands for higher computing resources during processing and storage utilization which affects the tracking speed from the perspective of computational performance which has never been judged in majority of the existing system. Therefore, a scope remains in learning-based solution where smart amendment and strategic modeling of feature computation could retain effective tracking performance in ITVO yet balance the computational factors in the presence of constraints such as partial or full occlusion, background clutter problem, and illumination variation.

The contribution of the study is as follows: i) in the proposed system context the analytical strategy for speedy abnormal event detection and tracking (SAEDT) follows simplified computational steps of execution which are not at all similar to the existing baselines of deep learning and other NMF based tracking models; ii) the customized NMF in the proposed systems learning aspects not only reduces the dimensionality of the complex video data but also retain appropriate feature computation considering feature engine structure reposit (FESR) modeling which forms a dictionary for the purpose of effective feature learning; iii) the design strategy here employs an efficient feature extraction mechanism with sparse combination learning strategy that not only contributes towards enhancing tracking accuracy for target mobile objects but also towards faster tracking with reduced computation cost; iv) it has to be noted that the proposed strategic model has tactfully dealt with the features and with limited features only the model is trained which reduces the computational burden to the system resources which is lesser explored in the existing baseline models; v) the novelty of this approach is that it can differentiate specific movements of mobile object of interest from other mobile objects that appear over crowded scenes; and vi) the experimental results further claim its effectiveness in terms of accuracy and cost of computation.The next section discusses about adopted method of proposed study.

## 2. METHOD

The study introduces a novel computational framework design and modelling for abnormal event detection and tracking in the research context of ITVO. Here, the study aims to analyse the object of interest movement through SAEDT by processing different instances of surveillance video sequences. The study assumes that to design and develop an effective video tracking system; the primary criteria is to meet the requirements to deal with the inherent redundancy of video structures. Considering this aspect, the strategic execution workflow modelling of the SAEDT framework is proposed. The study constructed model for FESR for appropriate feature computation. Here the study formulated a customized and strategic feature extraction model execution workflow namely FESR which is influenced from the concept of dictionary learning [23], [24]. The formulation of the feature extraction considers block-size modeling to update the feature vector with respect to block entities which also represents the essential pixel attributes. It also implements a column-oriented structure modeling for the representation of frame block entities and finally updates the FESR for the purpose of dictionary based learning paradigm where an efficient sparse combination modeling also utilized [25]. Here the system also utilizes statistical functionalities to evaluate the proposed feature extraction process for sparse dictionary-basis vector learning strategy.

### 2.1. Design challenges in object of interest tracking from abnormal events

The traditional practices of ITVO, precise detection and tracking of abnormal events is a challenging task based on what installed surveillance cameras capture. The traditional practices of OIT are pretty labor-intensive and require non-stop human attention, making the process never-ending and boring. The possibilities also arise where the chances of abnormal events become significantly less, making 99% of the effort of watching the videos go wasted. The current research practices are more inclined towards automatic detection and tracking of abnormal events, which meets the requirements for present or emerging computer vision applications [26], [11]. It has to be noted that OIT in the form of abnormal event detection and tracking is not a typical classification problem, as difficulties arise in listing out all the possible negative samples [25].

### 2.2. Data acquisition for UCSD dataset

The data acquisition of the considered dataset [27] of UCSD anomaly detection took place with the stationary cameras installed at an adjacent elevation of walkways. The mounted cameras were subjected to overlook the movement patterns of pedestrians over the walkways. The dataset is considered to model the SAEDT workflow strategy where walkways in the video data frames have variable crowd density, which ranges from sparse to extremely crowded scenes. In the normal settings, video frames appear with pedestrians walking over the walkways as recorded by the mounted cameras.

### 2.3. Dataset description

The study in the design and modeling of SAEDT framework considers the UCSD dataset [27] for the purpose of experimental design and analytical simulation modeling. It aims to construct a basis for automatic OIT from abnormal events. Here the standard video dataset $v_d$=[$v_1$ , $v_2 \cdots v_i$] is splitted into two subsets which could be analytically represented as (1):

$$v_d\colon \forall v_d \leftarrow \{v_t \cup v_{ts}\} \tag{1}$$

Here $v_d$ represents the original video dataset which is splitted into training video data $(v_t)$ and testing video data$(v_{ts})$. Here the video footage recorded from each scene was splitted into video clips $v_i$ of $N$ number of frame sequence $(F_s)$ where $N$=200. Here the entire dataset consists of $\{v_d(i)\}_{i=1}^{M}$ such that $v_d$=$[v_1, v_2 \cdots v_M]$ which is further splitted into $\{v_t(i)\}_{i=1}^{P}$ and $\{v_{ts}(i)\}_{i=1}^{Q}$ such that $\{v_t(i)\}_{i=1}^{P} \subseteq v_d$ and $\{v_{ts}(i)\}_{i=1}^{Q} \subseteq v_d$. The total frame sequence $TF_t$ for training dataset of $\{v_t(i)\}_{i=1}^{P}$ is computed with (2).

$$TF_t = N \times v_t(i) \tag{2}$$

Similarly the computation of total frame sequence $(TF_{ts})$for testing dataset $\{v_{ts}(i)\}_{i=1}^{Q}$ is also computed with (3).

$$TF_{ts} = N \times v_{ts}(i) \tag{3}$$

Here dataset of $v_d$ is splitted into a ratio of $TF_t$:$TF_{ts}$ which appears to be 63% $TF_t$ for training data out of $v_d$ and remaining 36.99% for$TF_{ts}$. The research study also explores both training and testing frame sequence such as $F_s \in \{v_t, v_{ts}\}$ and its corresponding data description to understand its structure of representation from the perspective of computational analysis. The next stage of operational process considers data exploration for SAEDT framework.

### 2.4. SAEDT: video data exploration

The video data exploration process in SAEDT framework initially locates the selected training data of $v_t(i) \subseteq v_d$ and locates its corresponding directory tree structure vector $(dS)$ for each $v_t(i)$. Here the numerical computational approach further applies and explicit function of $f_1(x) \leftarrow [dS, F_t]$ to compute the attribute of location information $(dS_{loc})$ for the respective frame $F_j$ of specific type $t$. The study further computes the frame location structure $F_{loc}$[struc] for individual $\{F(j)\}_{j=1}^{w}$. Here w indicates the upper bound correspond to $F_j$. Here the computing process of $F_{loc}$[struc] generates a structure array with several fields such as $\{Fname_1, dt_2, sF_3, Flag_4, sdt_5\}$. The field $\mathrm{Fname}_1$ indicates name of the particular file/folder, whereas $\mathrm{dt}_2$ refers to modification date and time stamp associated with that particular$F(j)$. $\mathrm{sF}_3$ refers to the size of $F(j)$ and $\mathrm{Flag}_4$ indicates whether the file correspond to directory or not. Finally $\mathrm{sdt}_5$ indicates the serial date number for the particular file.If the system of $f_3(x)$ performs concatenation on the two different sets such as $s_1, s_2$ , here $s_1$ represent the partial location information structure and $s_2$ represents the attributes of the field of name for $\mathrm{Fname}_i$ in the form of string for all the respective $Fname_{i=1}^{w}$: $Fname_{i=1}^{w} \forall V_t(i)$ and then it can be represented as (4):

$$s_1 s_2 = \{e_1 e_2 : e_1 \in s_1, e_2 \in s_2(i)\} \ where\ 1 \le i \le w \tag{4}$$

The process flow further enables another explicit function of $f_4(x)$ which basically reads the frames $\{F_j\}_{j=1}^{w}$ considering $s_1, s_2$ and perform digitization of the individual $F_j$. The computational approach further applies the same workflow model of execution for $v_{ts}(i) \subseteq v_d$ where each $v_{ts}(i)$ also consists of $\{F_j\}_{j=1}^{w}$.

### 2.5. SAEDT: construct feature engine structure reposit for feature extraction

The computation process strategically models this process of FESR for the purpose of effective feature computation to facilitate OIT. Here the computation considers the attributes from the previous phase which are $s_1, s_2$ and the system also enables reading of the $\{F_j\}_{j=1}^{w}$ using $f_4(x)$ from UCSD anomaly detection dataset. The computation generates the numerically computable representation of the $\{F_j\}_{j=1}^{w}$ in the form of $F_j$. Here the system also applies a novel block-size $(Bs^3)$ modeling considering $(m, n)$ for $F_j \in v_t(i) \subseteq v_d$which is further used for FESR modeling. Here the system also considers the number of frames for training$(\mathrm{nF_j})$.The strategy of feature computation and exploration initially divide the block size into m-by-m which is further passed to an implicit function of $f_5(x)$ along with $\mathrm{F_j}$. Here the computation considers m-by-m blocks of $B_j(m^2) \in F_j$ into a column structure of $S_{col}$. Finally, $S_{col}$ comprises of the concatenated columns in the form of matrix. The construction of FESR model basically creates bucket $b_{FESR}$ which considers the size of $(m^3 \times \chi \times \varepsilon)$. Here the computation of ε can be performed as (5):

$$\varepsilon = \frac{nF_j}{m} \tag{5}$$

The computing process further enables the block-based FESR modeling where initially it considers the $\kappa$ number of frames and also creates another $\mathrm{F_{j2}}$ which is a null high dimensional matrix of size$row, col, m$. The computation of $F_{j2}$ can be represented as (6). The algorithm of FESR feature extraction modeling is provided as (6):

$$F_{j2} \leftarrow F_j[row, col, m_{i+1}] \text{ where } 1 \le j \le \kappa \text{ , } 0 \le i \le 4 \tag{6}$$

Here the value of $\kappa$ is computed as (7):

$$\kappa = nF_j - m \tag{7}$$

SAEDT: Algorithm for design of FESR in feature extraction process
Input: s_1, s_2, 〖{F_j}〗_(j=1)^w∈v_t (i)⊆v_d
Output: ρ_FESR
Begin
1. Init s_1 ,s_2 , 〖{F_j}〗_(j=1)^w
2. Apply: 𝑓_4 (x)
3. Process: 〖{F_j}〗_(j=1)^w in the form of F_j
4. IF ∂(F_j )<E^3
5. Reduce: ∂(F_j )=E^1
6. Else
7. Consider F_j of E^1
8. Set 〖Bs〗^3 , nF_j
9. Divide block_size(m-by-m) for F_j.
10. Apply: 𝑓_5 (x)
11. 〖S_col (m^2×χ) ←B〗_j (m^2) // organizes the frame blocks (pixels) into a column structure
12. S_col=[B_1 ,B_2,B_3…B_j]
13. b_FESR of (m^3×χ×ε)
14. For j←1 toκ
15. For i←0 tom-1
16. F_j2←F_j [row,col,m_(i+1) ]
17. End
18. End
19. Update b_FESR (m^3×χ) // block-oriented FESR
20. Mean computation using Eq. (10)
21. Feature Vector: fv=〖b_FESR (m^3×χ)〗_j where 1≤ j≤χ
22. Normalization of fv1
23. Update b_FESR (m^3×χ×ε)← 〖b_FESR (m^3×χ)〗_fv1
End

The computing process further also initializes the cell_blocks in FESR and for each $m_i$ it updates the $F_{j2}$ $[row, col, i]$. The computing process further also divides the $F_{j2}$ into $(m \times m)$ blocks and column-wise arrangement considering $f_5$(x) and update $S_{col} = [B_1, B_2, B_3 \dots B_j]$. The feature vector computation takes place with (8).

$$fv = b_{FESR}(m^3 \times \chi)_j \text{ where } 1 \le j \le \chi \tag{8}$$

The further computation perform subtraction of $\mu$ from $fv$ as (9):

$$fv1 = (fv - \mu) \tag{9}$$

So that the summation of $fv1$ remains 1. Further computation also normalize the feature vector $fv1$ and store the values of $fv1$ to $b_{FESR}(m^3 \times \chi)$.

### 2.6. SAEDT: feature learning strategy based on effective sparse factorization

The proposed study has explored the potential advantages of NMF with sparseness in dimensionality reduction of features where it can effectively deal with the inherently nonnegative data. The study formulates a customized approach for non-negative matrix approximation which even though follows the fundamental steps of NMF. The idea of sparse NMF has evolved for the purpose of computation of parts-based, linear representation of non-negative data. Ren *et al.* [13] claims that even though standard factorization of data matrix utilizes singular value decomposition (SVD) which is incorporated over principles component analysis (PCA) but these methods lack effectiveness when the dataset consists of sequence of frames and textural attributes. However, for many dataset of image sequence and text, the original data matrices are found non-negative. If there is a non-negative video representation data $\in \mathbb{R}^{N\times T}$, then the NMF factorization basically factorize the $N \times T$ dimensional data of $V$ into non-negative factors of $W \in \mathbb{R}^{N\times k}$ and $H \in \mathbb{R}^{k\times T}$ such (10) can be satisfied.

$$V = W \times H \tag{10}$$

Here $k$ represents the basis components for NMF factorization into $W \in \mathbb{R}^{N\times k}$ with the lower dimensionality factor which also implies the number of clusters Figure 1.

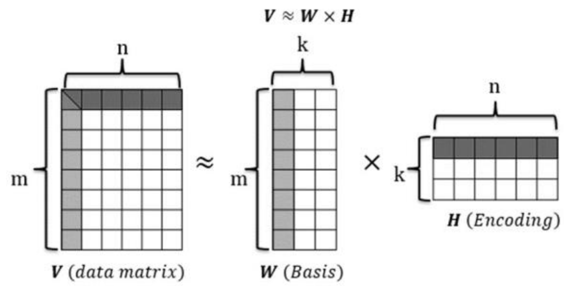



Figure 1. Decomposing V data matrix into W, H with NMF [28]

The NMF makes use of these three metrics to describe the decomposition of features with potential physical meaning which helps in appropriate feature computation and learning for object of interests. Here $k$ dimension also presents the features of the data and it is selected in such a way to satisfy the constraints $(N + T)k \leq NT$. Considering the data matrix $V$, the decomposition of $W * H$ takes place which is shown in the Figure 1. Here, $H$ implies which feature is related to which $F_j$ and with what intensity factor$(I_\xi)$. The objective of the method NMF is to minimize the distance between $V, W \times H$ while preserving the non-negativity of $W, H$. The problem formulation of distance minimization in the form of cost can be shown with (11).

$$MIN \rightarrow\ E = \left||V - W \times H|\right|^2\ w.r.tW\&Hs.tW, H > 0 \tag{11}$$

Here $E$ implies the error which is minimized with respect to the desired sparseness of $w_t \rightarrow S_w$ and$h_t \rightarrow S_h$. The NMF algorithm in the proposed workflow of feature learning strategy aims to obtain $W, H$ for the given sparsity constraints which is measured by $l_1$ norm of the vector.

### 2.7. SAEDT: object of interest detection and tracking

One the feature matrix is generated along with the learning of the model, the next strategic module of SAEDT performs the abnormal event detection and tracking in the form of OIT. Here initially it considers the test video sequence of $\{v_{ts}(i)\}_{i=1}^{Q} \subseteq v_d$ and compute the ground truth $GT$. Here initially the computing process again execute the SAEDT: Algorithm 2 design and updates the metrices of $X(m^3, \varepsilon)$ for decomposed matrix form of $F_{set}\epsilon V$. Further the system computes the distance factor in the form of probabilistic measure $P \leftarrow prob(X(m^3, \varepsilon)\colon \forall F_{set}\epsilon V$ in the measure of squared outcome and further compute the summation of the overall probabilistic measure with (12).

$$E_P = \sqrt[2]{\sum X(m^3, \varepsilon)} \tag{12}$$

The computation further obtain the ground truth annotated $\{v_{ts}(i)\}_{i=1}^{Q} \subseteq v_d$ data and check the dimensionality of individual $F_j$ and the strategic execution of the proposed tracking model SAEDT further employs another strategic function of $f_7(x)$ which basically evaluates the testing and validation considering the ground truth value $GT, E_{min}, \psi(E_{min})$. The computation of appropriate block-representation also forms feature vector with essential feature entities which helps in matrix factorization and sparse coding. The analysis of block-coordinate descent factors also helps converging towards global optimum features which is essential for learning. Here the feature extraction process basically helps the learning strategy for achieving two goals where the first goal is effective representation of features and the second is to find normalized optimum feature combinations from the redundant surveillance video information.

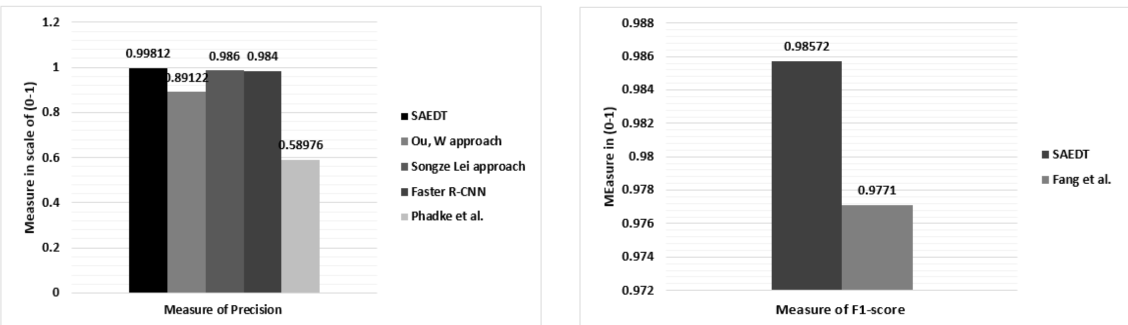
## 3. RESULTS AND DISCUSSION

The study considers MATLAB 2015a mathematical computing software along with the system configuration of processor: Intel(R) Core (TM) i5-8250U CPU @ 1.60 GHz 1.80 GHz, installed RAM: 12.0 GB and system type: 64-bit operating system, x64-based processor to realise the formulated analytical algorithms of SAEDT framework. SAEDT evaluation criteria considers a set of performance metrics such as computation of specificity, precision, recall, and F1_score along with cost of computation in the form of processing time to justify the SAEDT system outcome. The visualization of the $\mathrm{F_{loc}[struc]}$ and its corresponding computed attributes from the training video clip 1 can be shown in Table 1.

Table 1. Visualization of $\mathrm{F_{loc}[struc]}$ and its corresponding attribute

| Fields | Fname1 | dt2 | sF3 | Flag4 | sdt5 |
|---|---|---|---|---|---|
| 1. | F(1) | 17-Oct-2012 06:16:26' | 38056 | 0 | 7.351592614120371e+05 |
| 2. | F(2) | '17-Oct-2012 06:16:26' | 38026 | 0 | 7.351592614120371e+05 |
| 3. | F(3) | '17-Oct-2012 06:16:26' | 38008 | 0 | 7.351592614120371e+05 |
| 4. | F(4) | '17-Oct-2012 06:16:26' | 38044 | 0 | 7.351592614120371e+05 |
| - | - | - | - | - | - |
| - | - | - | - | - | - |
| - | - | - | - | - | - |
| 200. | F(200) | '17-Oct-2012 06:16:26' | 38052 | | 7.351592614120371e+05 |

The study evaluated the proposed SAEDT tracking model over ten training video clips and estimated the performance metrics of recall, precision, specificity, F1-score, and feature extraction time for different test instances. The study considers comparative analysis by comparing the precision measure of SAEDT with the existing baseline studies by [29]–[32]. The outcome in Figure 2(a) clearly shows that the proposed SAEDT attains comparable outcome for the measure of precision which is approximately 0.99812. On the other hand, faster R-CNN also attains precision score of 0.986 and 0.984 respectively. Several research [29], [32] approach do not ensure better precision outcome. It is quite clear that the SAEDT not only accomplishes better tracking accuracy but also its strategic modeling has significantly reduced computational burden to the training model. Figure 2(b) shows that comparison outcome of F1-score. The comparative analysis of F1-score also shows that the proposed SAEDT outperforms the approach of Fang *et al.* [33], the study explored many related baseline approaches and their design approaches which follows slightly different evaluation strategy considering sparsity based abnormal event detection. The study further evaluated the complexities associated with the existing works and normalized or approximated their outcome with mean computation and further synthetically generated the comparable data with respect to precision, F1-score, specificity, and sensitivity and computing time to show the effectiveness of the proposed system. That means in the proposed system the precision score is found approximately 0.99812 which is found on average 0.89122 in the case by [29] even in the presence of occlusion and illumination variation constraints.




(a) (b)

Figure 2. Comparison of (a) precision and (b) F1-score measure with popular baseline studies

The insights into the research, comparisons and analysis shows that the accuracy outcome of the proposed study also found comparable with the existing baseline solutions with significant improvement over the processing speed. It has to be also noted that the proposed strategic modeling accurately tracks object of interest from the abnormal event even in the presence of various constraints such as partial or full occlusion, illumination variation, and background clutter. The comparable feature extraction time is also found on an average of 0.128124 ms which also ensures its applicability over wide range of realtime survellaince applications. It has to be noted that this proposed research study inclines with the thoughts on the gap extracted by the same authors in [34], [35].

## 4. CONCLUSION

The study in this manuscript introduces a proposal for the conceptual modeling of effective ITVO model which balances the trade-off that exist between tracking accuracy and the cost of computation factors in the measure of algorithm execution time. The entire strategic modeling is simplified with optimized flow of functional execution where the methods of NMF makes the features appropriate for learning from both appropriateness and computation point of view. Here the feature learning strategic model basically improvises the learning from unlabelled data. Here the selected features also undergoes for extraction of unique combinations where column-wise unit representation unit norm to prevent over-fitting. The extensive simulation results show that the proposed strategy of SAEDT outperforms the existing techniques and also addresses the complexity problems of deep learning models in tracking of significant events where abnormal objects movement patterns are tracked among other mobile objects presence in a dynamic and complex video scene of pedestrians. The tracking model performance is evaluated under a specific dataset. However, its integration can be tested for other video clips as well. The novelty of this approach is it attains faster and accurate tracking performance even in the presence of occlusion, clutters, and other constraints. The future research work also focus towards building another cost-effective predictive modeling of SAEDT considering mode complex dataset and parameterized constraints.

## REFERENCES


[1] V. Sharma, M. Gupta, A. Kumar, and D. Mishra, "Video processing using deep learning techniques: a systematic literature review," *IEEE Access*, vol. 9, pp. 139489–139507, 2021, doi: 10.1109/ACCESS.2021.3118541.

[2] J. Kaur and W. Singh, "Tools, techniques, datasets, and application areas for object detection in an image: a review," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 38297–38351, 2022, doi: 10.1007/s11042-022-13153-y.

[3] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: datasets, metrics, and methods," *Applied Sciences*, vol. 10, no. 21, pp. 1–24, 2020, doi: 10.3390/app10217834.

[4] B. G. -Garcia, T. Bouwmans, and A. J. R. Silva, "Background subtraction in real applications: challenges, current models, and future directions," *Computer Science Review*, vol. 35, 2020, doi: 10.1016/j.cosrev.2019.100204.

[5] S. Bao, X. Zhong, R. Zhu, X. Zhang, Z. Li, and M. Li, "Single shot anchor refinement network for oriented object detection in optical remote sensing imagery," *IEEE Access*, vol. 7, pp. 87150–87161, 2019, doi: 10.1109/ACCESS.2019.2924643.

[6] S. A. Qureshi *et al.*, "Kalman filtering and bipartite matching based super-vhained tracker model for online multi object tracking in video sequences," *Applied Science*, vol. 12, no. 19, 2022, doi: 10.3390/app12199538.

[7] F. Bombardelli, S. Gul, D. Becker, M. Schmidt, and C. Hellge, "Efficient object tracking in compressed video streams with graph cuts," in *2018 IEEE 20th International Workshop on Multimedia Signal Processing, MMSP 2018*, 2018, pp. 1–6, doi: 10.1109/MMSP.2018.8547120.

[8] R. Pereira, G. Carvalho, L. Garrote, and U. J. Nunes, "sort and deep-sort based multi-object tracking for mobile robotics: evaluation with new data association metrics," *Applied Sciences*, vol. 12, no. 3, 2022, doi: 10.3390/app12031319.

[9] N. Gillis, "Sparse and unique nonnegative matrix factorization through data preprocessing," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 3349–3386, 2012.

[10] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, 2014, doi: 10.1109/TPAMI.2013.111.

[11] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1975–1981, doi: 10.1109/CVPR.2010.5539872.

[12] G. Zhou, A. Cichocki, and S. Xie, "Fast nonnegative matrix/tensor factorization based on low-rank approximation," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2928–2940, 2012, doi: 10.1109/TSP.2012.2190410.

[13] B. Ren, L. Pueyo, G. Ben Zhu, J. Debes, and G. Duchêne, "Non-negative matrix factorization: robust extraction of extended structures," *The Astrophysical Journal*, vol. 852, no. 2, 2018, doi: 10.3847/1538-4357/aaa1f2.

[14] J. Wang, M. Zhang, X. Hu, and T. Ni, "Incremental learning algorithm based on graph regularized non-negative matrix factorization with sparseness constraints," in *2021 4th International Conference on Artificial Intelligence and Big Data, ICAIBD 2021*, 2021, pp. 125–128, doi: 10.1109/ICAIBD51990.2021.9459040.

[15] F. Aghili, "Fault-tolerant and adaptive visual servoing for capturing moving objects," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 3, pp. 1773–1783, 2022, doi: 10.1109/TMECH.2021.3087729.

[16] S. Banerjee, H. H. Chopp, J. G. Serra, H. T. Yang, O. Cossairt, and A. K. Katsaggelos, "An adaptive video acquisition scheme for object tracking and its performance optimization," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17227–17243, 2021, doi: 10.1109/JSEN.2021.3081351.

[17] X. Chen, H. Li, Q. Wu, K. N. Ngan, and L. Xu, "High-quality R-CNN object detection using multipath detection calibration network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 715–727, 2021, doi: 10.1109/TCSVT.2020.2987465.

[18] J. Chen, Z. Xi, C. Wei, J. Lu, Y. Niu, and Z. Li, "Multiple object tracking using edge multi-channel gradient model with ORB feature," *IEEE Access*, vol. 9, pp. 2294–2309, 2021, doi: 10.1109/ACCESS.2020.3046763.
[19] J. Y. Kim and J. E. Ha, "Foreground objects detection using a fully convolutional network with a background model image and multiple original images," *IEEE Access*, vol. 8, pp. 159864–159878, 2020, doi: 10.1109/ACCESS.2020.3020818.
[20] X. Li, D. Song, and Y. Dong, "Hierarchical feature fusion network for salient object detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 9165–9175, 2020, doi: 10.1109/TIP.2020.3023774.
[21] J. Su, J. J. Liao, D. Gu, Z. Wang, and G. Cai, "Object detection in aerial images using a multiscale keypoint detection network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1389–1398, 2021, doi: 10.1109/JSTARS.2020.3044733.
[22] H. Xu, X. Lv, X. Wang, Z. Ren, N. Bodla, and R. Chellappa, "Deep regionlets: blended representation and deep learning for generic object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1914–1927, 2021, doi: 10.1109/TPAMI.2019.2957780.
[23] L. Fanghua, R. Ruolin, N. Hao, and W. Aixia, "Face super-resolution via robust online dictionary learning," in *2016 8th International Conference on Wireless Communications and Signal Processing, WCSP 2016*, 2016, pp. 1–5, doi: 10.1109/WCSP.2016.7752454.
[24] J. Shi, X. Ren, G. Dai, J. Wang, and Z. Zhang, "A non-convex relaxation approach to sparse dictionary learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1809–1816, doi: 10.1109/CVPR.2011.5995592.
[25] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2720–2727, doi: 10.1109/ICCV.2013.338.
[26] F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Computer Vision and Image Understanding*, vol. 115, no. 3, pp. 323–333, 2011, doi: 10.1016/j.cviu.2010.10.008.
[27] L. Hu and F. Hu, "anomaly detection in crowded scenes via SA-MHOF and sparse combination," in *Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017*, 2017, pp. 421–424, doi: 10.1109/ISCID.2017.130.
[28] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007, doi: 10.1016/j.csda.2006.11.006.
[29] W. Ou, D. Yuan, Q. Liu, and Y. Cao, "Object tracking based on online representative sample selection via non-negative least square," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10569–10587, 2018, doi: 10.1007/s11042-017-4672-3.
[30] S. Lei, B. Zhang, Y. Wang, B. Dong, X. Li, and F. Xiao, "Object recognition using non-negative matrix factorization with sparseness constraint and neural network," *Information*, vol. 10, no. 2, 2019, doi: 10.3390/info10020037.
[31] Y. Zhang, J. Wang, and X. Yang, "Real-time vehicle detection and tracking in video based on faster R-CNN," *Journal of Physics: Conference Series*, vol. 887, Aug. 2017, doi: 10.1088/1742-6596/887/1/012068.
[32] G. Phadke, "Robust multiple target tracking under occlusion using fragmented mean shift and Kalman filter," in *ICCSP 2011 - 2011 International Conference on Communications and Signal Processing*, 2011, pp. 517–521, doi: 10.1109/ICCSP.2011.5739376.
[33] Z. Fang *et al.*, "Abnormal event detection in crowded scenes based on deep learning," *Multimedia Tools and Applications*, vol. 75, no. 22, pp. 14617–14639, 2016, doi: 10.1007/s11042-016-3316-3.
[34] Divyaprabha and S. Guruprasad, "Design strategy for identification and tracking of video objects over crowded video scenes using a novel feature-learning algorithm," in *3rd IEEE International Conference on Mobile Networks and Wireless Communications, ICMNWC 2023*, 2023, pp. 1–8, doi: 10.1109/ICMNWC60182.2023.10435734.
[35] Divyaprabha and M. Z. Kurian, "Methodological insights towards leveraging performance in video object tracking and detection," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, pp. 460–474, 2023, doi: 10.14569/IJACSA.2023.0140851.


## BIOGRAPHIES OF AUTHORS

**Divyaprabha** is Associate Professor at Sri Siddhartha Institute of Technology, Tumkur, India. She received her Bachelor's degree from Bangalore University and Master's degree from BITS, Pilani, Rajasthan. She is member of ISTE. Her field of interests are image processing and machine learning. She has published papers in conferences and journals. She can be contacted at email: divyaprabha@ssit.edu.in.

**Guruprasad Seebaiah** is Associate Professor at Sri Siddhartha Institute of Technology, Tumkur, India. He received his Bachelor's, Master's and doctoral degree from Visvesvaraya technological university (VTU), Belgaum, India. He is member of ISTE. His field of interests are image processing and biomedical instrumentation. He has published the research papers in reputed conferences and international journals. He can be contacted at email: guruprasads@ssit.edu.in.