DOI: 10.36297/vw.jei.v6i2.178

VW Engineering International, Volume: 6, Issue: 2, 148-156

Enhanced People Detection and Tracking in Smart Surveillance Using Probabilistic Modeling

Aravind B. B[*1], Sanjeev Kumar Shah[2], D. Abdul Jaleel[3], G. Praburam[4], Premalatha V[5]

[*1]Department of Computer Science and Engineering (Data Science), Malla Reddy Engineering College, Hyderabad
[2]Department of Computer Science and Engineering, Uttaranchal Institute of Technology, Uttaranchal University, Dehradun
[3]Department of Computer Science & Technology Madanapalle Institute of Technology & Science, Madanapalle
[4]Department of Computer Science and Engineering, Viswam Engineering College, Madanapalle
[5]Department of Computer Science and Engineering, SASI Institute of Technology and Engineering,

Received: March 28, 2025
Accepted: April 18, 2025
Published online: April 26, 2025

Abstract: People detection and tracking play a crucial role in smart surveillance systems, enabling various applications such as behavior analysis, identity recognition, and anomaly detection. Accurately detecting and tracking individuals in dynamic environments remains challenging due to variations in lighting, occlusions, and background noise, presenting a new and innovative approach for robust people detection and tracking, incorporating multiple processing stages: foreground extraction, noise removal, people detection, and tracking with re-identification.

The proposed method begins by extracting the foreground region from the background image to isolate potential moving objects. Subsequently, a noise removal technique is applied to enhance the precision of detection by eliminating unwanted artifacts. Once the foreground is refined, individuals are detected and tracked across consecutive frames. A re-identification mechanism ensures the correct association of individuals, even in cases of occlusion or temporary disappearance. To evaluate the effectiveness of the approach, extensive experiments were conducted on the PETS2009 dataset, a widely recognized benchmark for multi-object tracking in surveillance applications, suggesting the achievement of higher accuracy in detecting and tracking multiple individuals by the proposed method, even in complex scenarios. Furthermore, the method enhances people's counting capabilities, reducing errors commonly observed in traditional tracking techniques. The findings suggest that the proposed approach can be effectively integrated into real-time surveillance systems, improving automated monitoring and security applications.

Keywords: people detection, people tracking, conditional probability, people counting

1. Introduction

In recent years, identifying individuals in surveillance video footage has garnered considerable interest due to its diverse applications. These include identifying unusual activities, analyzing human gait, counting individuals in crowded areas, personal identification, gender classification, and monitoring falls among elderly individuals [1-2]. Surveillance videos typically have low resolution, and when captured by stationary cameras, they often feature minimal background changes. In outdoor environments like industrial zones, railway stations, marketplaces, and educational institutions, tasks such as detecting unusual activities, managing crowds, and regulating traffic pose significant challenges [3-5]. Smart Surveillance Systems (SSS) have gained significant attention from researchers for overcoming challenges in surveillance. These challenges involve detecting stationary and moving objects, tracking multiple individuals, classifying objects, analyzing intricate motion patterns, and recognizing human activities [6-8]. Usually, the objects are frequently recorded at a considerable distance, and it still depends on human observers to identify activities in real-time video feeds [9-10]. However, human attention has limitations when it comes to simultaneously monitoring multiple events on surveillance screens. As a result, automated human motion analysis has gained prominence as a research area in computer vision and pattern recognition.

Intelligent systems can detect and track motion information from moving objects, aiding in precise classification and further analysis. This study focuses specifically on human detection rather than recognizing complex actions. Identifying humans in video footage is challenging due to body posture, clothing, and lighting conditions, and the detection process generally involves two primary stages: object identification and classification. Object detection can be accomplished through various techniques, including background subtraction, optical flow analysis, and spatio-temporal filtering. Among these, background subtraction is a widely adopted approach that isolates moving objects by analyzing the disparity between the current frame and a reference background frame, either at the pixel level or in segmented blocks [12]. Several methods are available for background subtraction, such as adaptive Gaussian mixture models, non-parametric background models, temporal differencing, warping-based background models, and hierarchical background models. On the other hand, optical flow-based detection relies on tracking motion vectors of objects over time to identify movement within a sequence of images. However, this technique is often affected by noise, variations in color, and inconsistent lighting. Additionally, most optical flow methods require substantial computational resources and are sensitive to abrupt motion changes [13].

Detecting and tracking multiple individuals in surveillance systems pose significant challenges due to issues such as occlusion, illumination variations, and overlapping objects. Conventional methods, including histogram-based representation and probability estimation of appearance, have been employed for people tracking using color feature vectors. However, these approaches often fail in occlusion scenarios [14]. Bounding box-based detection methods have also been utilized to track individuals moving in different directions. While effective under controlled conditions, these methods struggle in complex environments with severe occlusions and illumination variations [15]. Overhead cameras offer an alternative solution for people counting by minimizing occlusion issues. However, their field of view is limited, making them unsuitable for wide-area monitoring. To address these challenges, feature point clustering techniques have been introduced, enabling identifying moving entities based on independent motion. This approach clusters feature points in frame sequences, estimates object trajectories, and counts individuals accordingly [16]. Despite its advantages, this method is highly sensitive to camera perspectives, leading to instability in feature point tracking. Multi-object tracking and event detection have been extensively explored over the last twenty years [17]. Motion estimation techniques, such as Kalman filtering and Bayesian modality fusion, have been developed to track objects across multiple camera views.

2. Literature Review

Over the years, researchers have developed various methods to improve people's detection and tracking accuracy and efficiency in surveillance systems. Mahalingam *et al.* introduced a three-step method involving pedestrian people recognition, tracking, and analysis, where an adaptive Gaussian mixture of was used for segmentation, followed by a fuzzy morphological filter to remove foreground noise [18]. Similarly, Tian *et al.* proposed an algorithm for multi-object tracking in online videos, which utilized object identification and tracking processes to classify tracked entities into single and multiple-object categories, ensuring effective management across different environments [19]. Hamuda *et al.* employed Kalman filtering to estimate an object's position in a frame sequence to enhance tracking accuracy. A data affiliation technique was further integrated to refine cropping operations, significantly reducing location failure rates [20]. Another approach, proposed by Anindaputri *et al.*, combined mean shift tracking with a particle Kalman filter, where mean shift served as the primary tracker under normal conditions, but when occlusion occurred, the particle Kalman filter improved tracking accuracy [21]. Also, Chen *et al.* introduced a lower-grade object detection technique utilizing a stable relationship between outcomes across two scales, reducing processing time and computational expense [22].

For handling gradually moving objects, Sahoo *et al.* developed a segmentation algorithm to differentiate between actual movement and incorrect motion correlations. This was achieved through statistical analysis of frame means and valley-based thresholding to segment motion regions effectively [23]. Meanwhile, Ewees *et al.* applied an opposition-based learning method to optimize engineering problems, outperforming traditional algorithms across multiple benchmark tests [24]. In real-time tracking scenarios, Masaki *et al.* developed a 2-D color histogram-based method to detect human regions using gradient orientation in localized image sections. Since multiple human regions are often detected around a single individual, clustering techniques based on Euclidean and Bhattacharyya distances were used to refine the tracking process, ensuring a single representative region per person [25]. Similarly, Loris *et al.* explored structural component analysis to identify human body parts in crowded areas, utilizing super-pixel segmentation to group visually similar features such as forehead and cheeks. However, variations in lighting conditions posed challenges to consistency [26]. Further refinements in tracking methods were introduced by Ahsan Shehzad *et al.* who applied the Jaccard similarity index to compare features of detected individuals across consecutive frames. By determining the degree of similarity between feature sets, the method could identify and track individuals with high accuracy [27]. However, occlusions and merged detections sometimes led to missing observations, affecting precision. Despite these challenges, feature extraction, segmentation, and tracking algorithm advancements continue to drive improvements in people detection systems, making them increasingly robust for real-world applications.

3. The Proposed CPM Method

This paper presents a Conditional Probability Model (CPM) for multi-person detection and tracking in a Single Static Surveillance (SSS) system using a monocular camera. A verification process distinguishes human subjects from other objects to ensure accurate detection. The identified individuals are then assigned unique labels for effective tracking. The proposed approach utilizes color-based conditional probability representation combined with pairwise distance similarity measures to enhance tracking accuracy. Experimental results demonstrate that the model effectively adapts to varying lighting conditions, making it suitable for real-world applications.

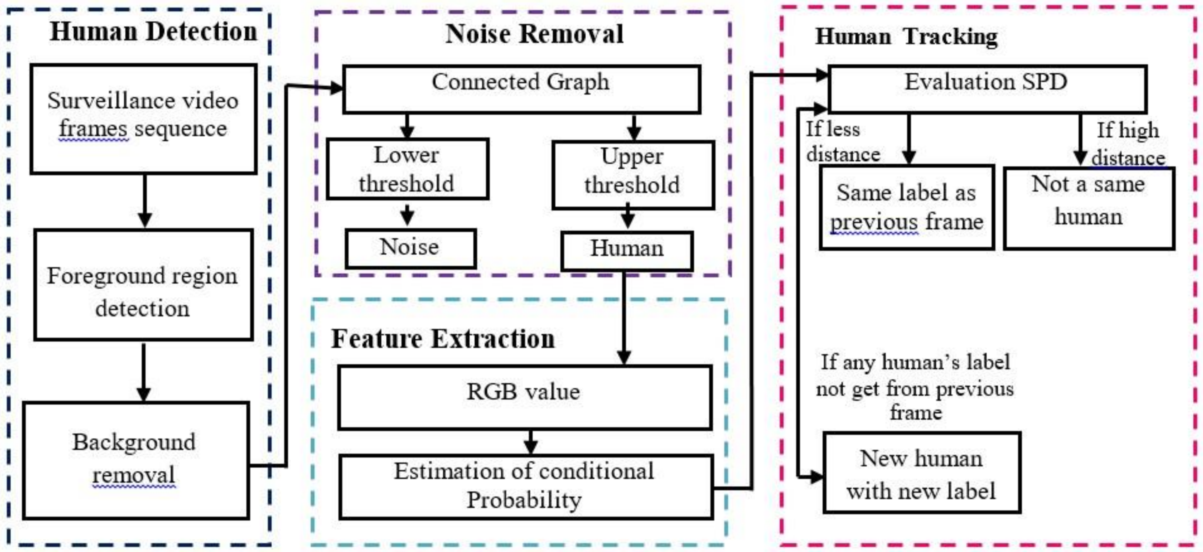

Figure 1: Workflow of CPM Model

The objective of the CPM is to track individuals across successive frames. The flowchart of the CPM is illustrated in Figure 1. The proposed CPM takes a sequence of frames from surveillance footage as input. Various processing steps are applied, including foreground extraction, background elimination, human detection, tracking, and similarity evaluation. These tasks are performed using techniques such as background modelling, connected component analysis, and conditional probability methods.

3.1 Formulation of the Problem

The next step is to control the sequence of frames, the surveillance footage, Si = $\{S_1, S_2, S_3, ..., S_n\}$, where i = 1 to n, detecting the (h) foreground region in the sequence and gets the next M frame by subtracting from a basic frame H, relies on the investigation functionality of the h frame in settings, noise must be prevented in the foreground region since the foreground region contains static or noise events as static objects now. To minimize the noise, background subtraction of the output from the sequence of frames from the foreground region is carried out and the remaining are represented as Kj = $\{K_1, K_2, K_3, ..., K_a\}$ in S_i where j = 1 to a. In S_i and M_l =$\{M_1, M_2, M_3, ..., M_b\}$ in S_{i+1} where i = 1 to b. Given Kj, Ml, the conditional probability, used for tracking the object in the subsequent frame.

3.2 Foreground Region Detection

The color frame sequence Si from the security footage serves as the CPM's input. The foreground region that appears in the Si is detected using the RGB Background Model (BM) [28], where H is the frame of reference and each S_i frame is subtracted from h using equation 1.

$$B\_S_i = \lfloor h_{xy} \rfloor - \lfloor s_{xy} \rfloor \qquad (1)$$

A frame sequence with a foreground region where i = 1 to n is denoted by B_Si. The x-axis and y-axis corresponding location pixels in the reference and input frames are indicated by the coordinate values hxy and Sxy, respectively.

3.3 Background Removal

In Background where B_Si ={B_Si1,B_Si2...B_Sin_,...,B_Si_ni} is a group of frames that i = 1...n in a background region; the coordinate values hxy and sxy represent the pixel location coordinates in the reference frame and input frame, respectively; additionally, moving foreground regions provide additional attention with additional information, such as their shadows; consequently, the unwanted foreground region is extracted when the pixel value of the corresponding frame is subtracted from its corresponding pixels of the RGB background model; to get the best foreground region, the parameter is fixed to modify the input frame's pixel values [29]. The value of the parameter ^ will be adjusted based on the surveillance environment. The RGB color frame sequence has a set threshold T value. To get the equivalent pixel value in the reference frame, the ^ value is added or removed with the sxy if the color value reaches a predetermined threshold. The RGB backdrop model's initial intensity value serves as the threshold value.

$$s_{xy} = \begin{cases} s_{xy} & if \;\; h_{xy} > T \\ s_{xy} \pm \Omega & Otherwise \end{cases} \qquad (2)$$

3.4 People Detection

There may be noise and people in the foreground region frames. An undirected graph is first used to temporarily name the foreground region of the Kj individuals [30]. Equation 3 mathematically combines the Connected Components Nodes (CN).

$$CN = \{ AR_{(h+1,r)}, AR_{(h-1,r)}, AR_{(h,r+1)}, AR_{(h,r-1)} \}, \qquad (3)$$

In this case, $AR_{(h+1,r)}$ is at the right, $AR_{(h-1,r)}$ is at the left, $AR_{(h,r+1)}$ is at the top, and $AR_{(h,r-1)}$ is at the bottom of the current pixel. There may be noise or people in the connected area. The average human size is fixed to eliminate the noise in the foreground area. When a person's size falls below a predetermined threshold, a specific associated foreground region is deemed to be noise. It will be eliminated from subsequent procedures. The remaining Connected Graph (CG) is regarded as the people once the noise has been eliminated. The equation 4 represents it.

$$CG = \begin{cases} K_j & if \; size of \; (K_j) > \Delta \\ noise & otherwise \end{cases} \qquad (4)$$

4.1

$$K_j \rightarrow \lfloor K_{1\,CP1}, K_{1\,CP2}, \ldots, K_{1\,CP10} \rfloor, \lfloor K_{2\,CP1}, K_{2\,CP2}, \ldots, K_{2\,CP10} \rfloor, \ldots\ldots\ldots \lfloor K_{a\,CP1}, K_{a\,CP2}, \ldots, K_{a\,CP10} \rfloor \qquad (5)$$

5.1

$$M_l \rightarrow \left[M_{1_{CP1}}, M_{1_{CP2}}, ..., M_{1_{CP10}}\right], \left[M_{2_{CP1}}, M_{2_{CP2}}, ..., M_{2_{CP10}}\right],$$

6.1 $$........, \left[M_{b_{CP1}}, M_{b_{CP2}}, ..., M_{b_{CP10}}\right] \quad (6)$$

Where, CP_1, CP_2, ...,....CP_{10} are the feature points for each person. Though a single value is extracted from the strips, it reflects the data distribution center using the average value.

3.5 People Tracking

This section describes the process of tracking individuals labeled as Kj, beginning with its initialization as the input. To ensure accuracy, the tracking mechanism relies on extracting and comparing key features across consecutive frames. To enhance precision, the space is divided into ten equally spaced vertical strips, and Conditional Probability (CPY) is computed based on the color distribution within each strip. The process starts with strip-wise color normalization, where pixel values are scaled between 0 and 1. CPY then represents the likelihood of a color value occurring simultaneously between two individuals, Kj and Ml. After this, relevant features are extracted from both Kj and M_l, and a condition is formulated to retrieve features from Ml selectively. Finally, each strip's average feature value is refined, ensuring accurate tracking. The mathematical details of this approach are provided in Equations (5) and (6).

3.6 *Similarity Analysis*

To find a respective human with its label, the sum of Pair Distance (SPD) is used to find the distance between K_j and M_l. SPD sums the distance of the corresponding pair of points [32]. Equation 7 can be used to compute this

7.1 $$SPD(K_j, M_l) = \sum_{j}^{a} CP_w(k_j), CP_w(m_l) \quad (7)$$

Where $j \neq k$ and let w be the corresponding positional CPY where $w = w$. The tracking is done between j^{th} CPY in K_j and k^{th} CPY in M_l.

When the SPD value is high, the algorithm maintains the same human label as in the previous frame. A new label is assigned if the value is low, identifying it as a different individual. The tracking process is outlined in Algorithm 1.

Algorithm 1 Tracking the People

Input: Collection of CPY (CPY1, CPY2, ..., CPY10) of K_j &

Collection of CPY (CPY1, CPY2, ..., CPY10) of M_l

Output: M_l

1. START
2. For each K_j
3. Calculate CPY (CPY1, CPY2, ..., CPY10)
4. For each M_l
5. Calculate CPY(CPY1,CPY2, ...CPY10)
6. END FOR
7. FOR each M_l and K_j
8. Calculate SPD
9. END FOR
10. Find the most relevant information related to equation 7.

11. Start Tracking the labels
12. Calculate SPD (there is no similarity for any K_i)
13. Assigning the new label is done
14. END

4. Result and discussion

The proposed CPM was developed and executed using MATLAB 2013a on a system with 4.00 GB RAM and an Intel(R) i5 3337U processor running at 1.80 GHz. The evaluation was conducted using the PETS09 dataset, which includes multiple video sequences captured from different perspectives. The experiments were performed under varying parameters, denoted as $\wedge$ and $\otimes$. To assess CPM's effectiveness, its performance was compared with different methods, including holistic property-based techniques for people detection and statistical as well as pattern-matching approaches for people tracking.

A. Dataset Description

The PETS2009 dataset was taken from https://cs.binghamton.edu/~mrldata/pets2009 [34] in which PETS09-S2L1 was only considered for analyzing the results by the proposed CPM model. The nature of the PETS09-S2L1 is the sparse crowd, captured in different surveillance areas and weather conditions.

B. Performance Analysis

The effectiveness of CPM is assessed by calculating detection and tracking accuracy using Equations 8 and 9. These metrics are evaluated against the ground truth for people detection and tracking. The results are presented in Tables I and II, where Table I details the detection accuracy for the first 30 frames in view 1, yielding a mean accuracy of 93%. Similarly, Table II outlines the tracking accuracy over the same 30-frame sequence. The mean accuracy is 90%. The detection and tracking accuracy is computed using correctly detected people and total people in a frame [27], [28]. The following equations explain it. (8) and (9)

$$Detection\,Accuracy = \frac{Correctly\,Detected\,People}{Total\,People\,in\,a\,Frame}, \quad (8)$$

$$Tracking\,Accuracy = \frac{Correctly\,Tracked\,People}{Total\,People\,in\,a\,Frame}, \quad (9)$$

TABLE I: People Detection ACCURACY (View 1)

Step No.	Projected count	Recorded count	Accuracy
10	6	6	99.99%
15	6	6	99.9%
20	6	5	80%
25	4	5	80%
30	5	5	100%
35	4	5	80%
Average people detection accuracy			90%

TABLE II: People Tracking Accuracy (View 1)

Sequence No.	Successful	Failure	Accuracy
10	4	1	80%
15	3	-	100%
20	4	-	80%
25	4	1	75%
30	5	-	100%
35	4	1	75%
Mean Human Tracking Accuracy			85%

C. Correlative Examination

Tables III and IV provide a comparative analysis of the proposed CMP method against other approaches. The holistic approach and pattern matching were used for people counting, where CMP achieved higher accuracy than existing techniques. Likewise, statistical methods and pattern matching were applied for people tracking, with CPM delivering improved performance over other methods.

TABLE III: ComparATIVE ANALYSIS OF HUMAN Detection Methods

Method	People detection accuracy
Holistic properties	85.21%
Comparative Mapping	86.40%
Recommended Strategy	90.00%

TABLE IV: Comparison with People Tracking Methods

Method	People counting accuracy
Statistical method [25]	89.00%
Pattern matching method [28]	87.01%
Proposed Approach	90.00%

The proposed approach is structured into four stages. Initially, the foreground region is identified by separating it from the background frame. In the second stage, smaller connected components are eliminated based on the average size of individuals to reduce noise in the foreground. The third stage involves extracting the color feature CP from the detected region to track the assigned label across consecutive frames. In the final stage, SPD estimates feature values between individuals. While this method performs effectively in standard frame sequences without occlusion, it does not account for occlusion scenarios. This is illustrated in Figure 2.

V. Conclusion

This study presents the CPM method for various surveillance applications, including people counting, detection, and tracking. A novel approach is introduced for multi-person detection and tracking, encompassing foreground extraction, noise filtering, individual identification, movement tracking, and similarity assessment. Experimental results demonstrated high accuracy, achieving 93.00% for people counting and 90.00% for tracking. However, the method does not account for occlusion scenarios. Future research can extend this work to include unusual event detection, person recognition, and other advanced applications.

References

1) K. Buys, C. Cagniart, A. Baksheev, T.-D. Laet, J. D. Schutter and C. Pantofaru, "An adaptable system for RGB-D based people body detection and pose estimation," Journal of visual communication and image representation, vol. 25, pp. 39-52, Jan 2014.
2) Jalal and S. Kamal, "Improved Behavior Monitoring and Classification Using Cues Parameters Extraction from Camera Array Images," IJMI, vol. 5, pp. 71-78, 2019
3) Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal and D. Kim, "Robust people activity recognition from depth video using spatiotemporal multi-fused feature," Pattern recognition, vol. 61, pp. 295-308, 2017.
4) Enyedi, L. Konyha and K. Fazekas, "Threshold procedures and image segmentation," in proc. of the IEEE International symposium ELMAR, pp. 119-124, 2005.
5) F. Farooq, A. Jalal, L. Zheng, "Facial expression recognition using hybrid features and self-organizing maps", in proc. of ICME, pp. 409-414, July 2017.
6) X. Yang and Y. Tian, "Eigenjoints-based action recognition using naivebayes-neartest-neighbor," in proc. of the CVPR, pp.14-19, June 2012.
7) Jalal, and S. Kamal, "Real-time life logging via a depth silhouettebased people activity recognition system for smart home services," in Proceedings of AVSS, Korea, pp. 74-80, Aug 2014.
8) Jalal, M. Maria and A. Hasan, "Multi-features descriptors for people activity tracking and recognition in Indoor-outdoor environments," in proc. on IBCAST, 2019.
9) Jalal and S. Kim, "The mechanism of edge detection using the block matching criteria for the motion estimation," in Proceedings of HCI Conference, Korea, pp. 484-489, Jan 2005.
10) Sony, K. Ajith, K. Thomas, T. Thomas, and P. L. Deepa, "Video summarization by clustering using euclidean distance," in proc. of the SCCNT, 2011.
11) S. Kamal and A. Jalal, "A hybrid feature extraction approach for people detection, tracking and activity recognition using depth sensors," Arabian Journal for science and engineering, 2016.
12) T. H. Chen, T.-Y. Chen and Z.-X. Chen, "A Intelligent People-Flow Counting Method for Passing Through a Gate," in proc. of the CIS, pp. 573-578, 2006.
13) J.-W. Kim, B.-D. Choi and S.-J. Ko, "Real-time vision-based people counting system for the security door," in proc. of the Inter. conf. On CSCC, pp. 1-4, 2002.
14) V. Rabaud and S. Belongie, "Counting crowded moving objects," in proc. of the conf. on CVPR, 2006.
15) G. Antonini and J.P Thiran, "Counting pedestrians in video sequences using trajectory clustering," IEEE Trans. on circuits and systems for video technology, vol. 16(8), 2006.

16) T.-H. Chang and S. Gand, “Tracking multiple people with a multi-camera system,” in proc. of the IEEE workshop on multi-object tracking, 2001.
17) T. Ding, M. Sznaier and O. I. Camps, “Fast track matching and event detection,” in proc. of the conf. on CVPR, 2008.
18) T. Mahalingam, M. Subramoniam, “A robust single and multiple moving object detection, tracking and classification,” Applied Comput. Inf., 2018.
19) P.K. Sahoo, P. Kanungo, S. Mishra, “A fast valley-based segmentation for detection of slowly moving objects,” Signal Image Video Process. **7**, 2018.
20) E. Hamuda, B.Mc Ginley, M. Glavin, E. Jones, “Improved image processing-based crop detection using Kalman filtering and the Hungarian algorithm,” Comput. Electron. Agric. 148, 2018.
21) Iswanto, B. Li, “Visual object tracking based on mean-shift and particle-Kalman filter,” Proc. Comput. Sci. 116, 2017.
22) K. Chen, J. Wang, S. Yang, X. Zhang, Y. Xiong, C.C. Loy, D. Lin, “Optimizing video object detection via a scale-time lattice,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
23) P.K. Sahoo, P. Kanungo, S. Mishra, “A fast valley-based segmentation for detection of slowly moving objects,” Signal Image Video Process. 2018.
24) A.A. Ewees, M.A. Elaziz, E.H. Houssein, “Improved grasshopper optimization algorithm using opposition-based learning,” Expert Syst. Appl. **112**, 2018.
25) Takahashi, M., Fujii, M., Shibata, M. et al. Robust Recognition of Specific Human Behaviors in Crowded Surveillance Video Sequences. EURASIP J. Adv. Signal Process. pp. 1- 14, 2010.
26) Loris Bazzani, Marco Cristani, Vittorio Murino, “Symmetry-driven accumulation of local features for human characterization and re-identification,” Elsevier, Computer Vision and Image Understanding, pp. 130 – 144, 2013.
27) A. Shehzed, A. Jalal and K. Kim, "Multi-Person Tracking in Smart Surveillance System for Crowd Counting and Normal/Abnormal Events Detection," 2019 International Conference on Applied and Engineering Mathematics (ICAEM), 2019, pp. 163-168.
28) M. Elhoseny, “Multi-object Detection and Tracking (MODT) Machine Learning Model for Real-Time Video Surveillance Systems,” Circuits, Syst. Signal Process., vol. 39, no. 2, pp. 611–630, 2020.
29) J. S. Kim, D. H. Yeom and Y. H. Joo, "Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems," in IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1165-1170, August 2011.
30) Shehzed, A. Jalal and K. Kim, "Multi-Person Tracking in Smart Surveillance System for Crowd Counting and Normal/Abnormal Events Detection," 2019 International Conference on Applied and Engineering Mathematics (ICAEM), 2019.
31) Kim, D.; Kim, H.; Mok, Y.; Paik, J. “Real-Time Surveillance System for Analyzing Abnormal Behavior of Pedestrians,” Appl. Sci. **2021**.
32) Huang and H. Wang, "A Novel Key-Frames Selection Framework for Comprehensive Video Summarization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 2, pp. 577-589, Feb. 2020.
33) Chan, M. Morrow and V. Nuno, “Analysis of crowded scenes using holistic properties,” in proc. of the IEEE workshop on CVPR, pp. 101- 108. 2009.