

Research Article

Anomaly Prediction over Human Crowded Scenes via Associate-Based Data Mining and K-Ary Tree Hashing

Affan Yasin ¹, Sheikh Badar ud din Tahir ², Jaroslav Frnda ^{3,4}, Rubia Fatima ⁵,
Javed Ali Khan ⁶ and Muhammad Shahid Anwar ⁷

¹School of Software Engineering, Northwestern Polytechnical University (NPU), Xian 710072, Shaanxi, China

²Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad 44000, Pakistan

³Department of Quantitative Methods and Economic Informatics, Faculty of Operation and Economics of Transport and Communication, University of Zilina, Zilina 01026, Slovakia

⁴Department of Telecommunications, Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava, Ostrava 70800, Czech Republic

⁵School of Software Engineering, Tsinghua University (THU), Beijing, China

⁶Department of Software Engineering, University of Science and Technology, Bannu 28100, Pakistan

⁷Department of AI and Software, Gachon University, Seongnam-si 13120, Republic of Korea

Correspondence should be addressed to Rubia Fatima; rubiafatima91@hotmail.com and Muhammad Shahid Anwar; shahidanwar786@gachon.ac.kr

Received 18 November 2022; Revised 24 March 2023; Accepted 18 July 2023; Published 25 July 2023

Academic Editor: Lianyong Qi

Copyright © 2023 Affan Yasin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Anomaly detection and behavioral recognition are key research areas widely used to improve human safety. However, in recent times, with the extensive use of surveillance systems and the substantial increase in the volume of recorded scenes, the conventional analysis of categorizing anomalous events has proven to be a difficult task. As a result, machine learning researchers require a smart surveillance system to detect anomalies. This research introduces a robust system for predicting pedestrian anomalies. First, we acquired the crowd data as input from two benchmark datasets (including Avenue and ADOC). Then, different denoising techniques (such as frame conversion, background subtraction, and RGB-to-binary image conversion) for unfiltered data are carried out. Second, texton segmentation is performed to identify human subjects from acquired denoised data. Third, we used Gaussian smoothing and crowd clustering to analyze the multiple subjects from the acquired data for further estimations. The next step is to perform feature extraction to multiple abstract cues from the data. These bag of features include periodic motion, shape autocorrelation, and motion direction flow. Then, the abstracted features are mapped into a single vector in order to apply data optimization and mining techniques. Next, we apply the associate-based mining approach for optimized feature selection. Finally, the resultant vector is served to the k-ary tree hashing classifier to track normal and abnormal activities in pedestrian crowded scenes.

1. Introduction

Regarding human video-based surveillance systems, the visual interpretation of abnormal behaviors is a field of crucial significance. Anomalies in camera footage comprise illegalities or potentially harmful conditions that pose a risk to the public. Throughout outdoor spaces around the world, billions of video surveillance equipment are being installed.

However, the majority of the devices are actively gathering videos without any functionalities. Considering the huge amounts of data created by the camcorders every second, it is unbearable for humans to comprehend this massive database of video sequences. How to enhance the visual surveillance method for abnormal event prediction, estimation and identification of pedestrians, road accidents, crimes, robberies, violent conflicts, unauthorized intrusions, and

other anomalous activities has become a major research area in the various domains such as smart home and state security. In addition, researchers require computer vision techniques to detect irregularities in videos automatically.

Due to human cognitive limits, such as decreasing attention after a specific length and constraints in spatial cognition modules, evaluating huge datasets can be challenging [1]. In order to overcome these limitations, visual surveillance technology has gained importance in various fields, including security, healthcare, and education [2]. These technologies enable the efficient extraction of meaningful data and information from enormous volumes of video sequences and the automatic monitoring of aberrant circumstances, hence lowering the amount of manual labor necessary to do these activities [3]. In the real world, various domains seem to be fit for this technology such as smart management systems, the educational sector, smart healthcare and monitoring, data and information encryption methods [4], and surveillance and public places. Although there are a variety of approaches for designing intelligent video surveillance, the fundamental strategy is to characterize regular events and recognize those that do not fit the given framework [5]. The researcher prefers the preceding smart methodology because the anomalous characterization differs depending on the content, i.e., scenarios regarded anomalous for one sequence may be considered positive for another moment, and due to the difficulty of locating anomalous behavior variations.

Throughout the previous systems, a significant amount of research has been put into creating sure “hand-crafted” characteristics that effectively reflect actual activities. By using traditional artistic methods to characterize an object’s motion, item sequences of typical activities were retrieved [6]. Thereafter, such items associated with tracks that diverged from the patterns that were previously learned were regarded as an anomaly. The use of global rectangular shapes to simulate the path track has taken the role of trend line approaches, which were usually observed to be impracticable for evaluating problematic situations, which include moderate aspects that are obtained from two-dimensional video frames or nearby 3D multimedia frames, such as temporal and spatial shading [7], histograms of optical flow mixtures of dynamic patterns, and motion characteristics [8]. The approaches are based on classifiers which have a key drawback in that they are hard to adjust toward the wide range of detection and segmentation present in various scenarios.

Previously, efforts were made to practice supervised learning for the activity of anomalous prediction in movie-based media, which responds to the remarkable results of supervised learning on video processing tasks such as classification, identification, object detection, and motion detection [9]. Furthermore, researchers [10] initiated the use of convolution neural networks for the abnormal activity identified due to its acceptance. Most of these techniques start by extracting the features using machine learning and optimization algorithms, followed by training classification models, like a one-class classifier. However, researchers were not developed and customized for the entire issue, and such feature maps are not at their optimal [11–13].

Our research proposed a robust and smart approach to the pedestrian anomaly prediction (PAP) framework. We utilized the human-crowded video-based data as input for the proposed method. Initially, we perform the different preprocessing steps including denoising, frame conversion, background subtraction, and RGB image to binary image conversion. After this, we identify the human from the given input data, using texton-based segmentation. Then, we applied Gaussian smoothing and crowd clustering while analyzing the crowd for further calculations and results. Once we find the human and settle with crowd analysis, the next step is to find various features. Therefore, we extracted a bag of features involving three types: periodic motion, motion direction flow, and shape autocorrelation. After extracting features, we map them in a single vector and apply data optimization and mining techniques. For this, we adopt the associate-based mining technique, which gives us the optimized vector for further classification. Finally, we apply the K-ary tree hashing algorithm to find normal and abnormal activities in pedestrian crowd-based videos. Our main contribution is as follows:

- (1) To predict indoor and outdoor pedestrian behavior in crowd-based scenes, we proposed a robust approach.
- (2) We introduced a bag of features in which we extract various features such as periodic motion, motion direction flow, and shape autocorrelation to predict normal and abnormal activities.
- (3) To reduce the data, optimization, and mining, we apply an associate-based mining technique, and for classification, we apply a K-ary tree hashing algorithm over two benchmark datasets.
- (4) We compared the performance of our proposed PAP system with other state-of-the-art classifiers (such as XGBoost and SVM) on two benchmark datasets (including Avenue and ADOC). The results have shown that the proposed PAP system significantly outperformed other classifiers.

The remainder of our work is structured as follows: Section 2 presents the background study of the proposed work. Section 3 introduces the primary approaches applied in PAP methodology, covering denoising, human and crowd analysis, feature extraction (bag of features), data optimization, and classification. The experimental evaluation with results is presented in Section 4. Finally, the research concludes with a discussion of the PAP system in Section 5 and summarizes the conclusion drawn with the insight gained in Section 6.

2. Related Work

We briefly discuss the essential requirements that culminate in dramatically varied pedestrian video abnormal action identification methodologies. The variety of potential anomalous occurrences is the primary obstacle to the anomaly estimation problem. Numerous researchers improve this challenge by explicitly describing abnormalities

and essential characteristics that can be used efficiently for image classification, with gesture trajectory being the most popular. These investigations seek to establish sequences of item trajectories based on regular occurrences [14]. The process consists of four major phases: object recognition, monitoring, trajectory-based semantic segmentation, and identification [15]. The benefits of procedures in this segment include their straightforward adoption and rapid implementation. Several saliency approaches have attained state-of-the-art results [16]. Rather than specifying the exact and predicting distinct anomaly properties, other researchers view a given series of images as a sequence of smaller 3D regions. Meaningfully, a collection of subsequent frames are integrated with the spatiotemporal direction and then separated into same-size 3D areas, including a window sliding on the current structure [17]. During the inference phase, every 3D patch obtained from untrustworthy sources is represented as a complex combination of a training dataset's episodes. The reconstruction error will serve as the criterion for validating the conclusion. In addition to space partitioning, gating factor analysts are also used to construct 3D sections [18], while other scholars begin to understand the association among learning regions according to their frequency or pattern-matching identification [19].

Sparse coding presupposes that all instances may be estimated as a linear transformation of several elements of learned dictionaries and has been frequently employed in anomaly detection [4]. In particular, sparse coding-based anomaly detection generates a vocabulary under the feature space requirement during training. It employs the restoration decline to identify unusual frames (i.e., abnormal activities) during detection. The other primary benefits of instance segmentation are its operational effectiveness, which complicates practical examples like surveillance tape evaluation. For instance, the authors of [4] suggest an interactive sparse coding strategy to combat such a disadvantage. Representation [20] means abandoning the feature space requirement and learning many tiny translations to encapsulate texture features at various scales. In addition to computational effectiveness, these conventional works have also been plagued by the constraints of deeply embedded algorithms.

Recently, specific approaches have been utilizing LISTA [21] for anomaly identification. Although these techniques could benefit from swift implication frequency and concurrently learn dictionaries and sparse illustration due to their machine learning execution, researchers have sustained from the first limited sources because they are comparable to the conventional solutions [22]. Generally, researchers use a nonadaptive generating technique and do not consider existing data while optimizing. Sometimes, the proposed approach may be inadequate and result in lower productivity [23]. For example, sparse/large data typically necessitate a per-dimension patching strategy to conserve sufficient processing resources. In addition, optimal control social evaluations [24] have demonstrated that adding available information improves the generalization capability of optimization techniques.

In [25], researchers developed a movement filter that uses temporal as well as spatial data among successive frames of video to recover erroneous detection methods. Deep convolution neural network (DCNN) is the foundation of our architecture for estimating crowds in low population videos. Visual Geometry Group (VGG16), Zeiler and Fergus (ZF), and VGGM are utilized in the context of a region-based DCNN for object tracking. In [26], researchers describe a two-stage head identification approach that uses a fully convolutional network (FCN) to produce scale-aware suggestions, accompanied by a convolutional neural network (CNN) that separates each recommendation into two groups, i.e., head and backdrop. Experimental findings indicated that the utilisation of spectrum recommendations generated by FCN leads to improvements in the recall frequency as well as the mean and median precision (mAP) of the item". In [27], researchers propose a solution that utilizes a sophisticated model that functions as a head detection and takes into account the scale fluctuations of heads in movies. The strategy is based on the premise that the head is the most conspicuous body feature in sporting venues where vast crowds congregate. In order to address the issue of varying scales, they produce spectrum head ideas based on even a density. The convolutional neural network (CNN) is then fed spectrum recommendations and returns a reaction matrix comprising the presence percentages of humans seen across picture dimensions. The authors [28] present an end-to-end methodology for spectrum head identification that can accommodate a wide assortment of scales. By simulating a collection of magnitude convolutional neural networks with distinct receptive fields, we establish that scale changes may be accounted for. Several spectrum monitors are incorporated into a single communications system whose settings are optimized from the beginning to the end.

Multiscale histogram of optical flow (MHOF) is obtained to characterize activity in [29], and anomalies are identified and predicated on a sparse reconstruction cost (SRC). The authors in [30] revised the SRC classification algorithm to include a sparse collection of learning and 3D contour characteristics to characterize an activity. Furthermore, Del Giorno et al. [31] proposed identifying variations in video footage by distinguishing images from prior images. As an enhancement of the bag of video words (BOV) technique, Javan Roshtkhari and Levine [32] proposed a probability distribution function for encoding spatiotemporal arrangements of video dimensions using gradient characteristics. They identified the anomalous performance of the suggested statistical tests by merging statistical properties, including HOG and HOF, to indicate the activities. Colque et al. [33] introduced a new dynamical feature selection method, histograms of optical flow orientation and amplitude and probability, via adding mobility and stochastic variables to HOF (HOFME). Spatial cuboid-based approaches may try to perceive long-term actions, including hovering, despite their robustness when contending with complicated environments. It is because hovering is tied to a person's long-term mobility instead than their relatively local cuboid mobility.

Most anomaly prediction-based anomaly detection methods utilize a few prior frames to anticipate the objective session. Compared to probability estimation methodologies, researchers are more logical and sensible in learning, and the systems are straightforward to configure. In contrast to the frame modelling techniques [34], the frame estimation techniques evaluate the anomaly's presentation and orientation and activity. Furthermore, the benefits of frame projection strategies outweigh those of frame creation systems. P-GAN, a representative effort, predicted future images using the U-Net as the GAN generation [35]. The obtained frame estimated the intensities loss, gradients decline, optical flow destruction, and integrity verification decline of the GAN classification model [36]. The net loss characteristic of the entire network was developed using the aforementioned different indicators. Lastly, the anomaly detection utilized the discrepancy image and its estimate given by the learning algorithm [37]. The model's structure is uncomplicated, but its effectiveness is exceptional. Moreover, the anomaly prediction model maintains accurate results in anomaly identification at an adequate level.

All the above methods have some limitations in dealing with this; we proposed a robust approach. Initially, we incorporated data and converted it into images and frames, denoised; after this, we applied texton-based segmentation and crowd clustering. The next step is to extract features from image sequences, periodic motion, motion direction flow, and shape autocorrelation for data optimization and data mining via associate-based data mining and data identification of normal and abnormal prediction and classification using the k-ary tree hashing algorithm. We apply this methodology over two publicly available datasets and achieve significant improvements. Finally, we compare it with the existing state-of-the-art method.

3. Materials and Methods

This section focuses mostly on our suggested PAP approach, which comprises multiple subparts. Initially, we collect RGB data via a surveillance camera and multifunctional camera in video and image format. After receiving the input, we apply it to preprocesses such as video-to-frame conversion, frame scaling, denoising, and RGB-to-binary conversion. The next phase separates humans from subhumans, clusters the population, and analyzes data. The bag of the features extraction technique is then applied, and three robust features are extracted: periodic motion, motion direction flow, and shape autocorrelation, for data optimization and data mining via associate-based data mining and data identification of normal and abnormal prediction and classification utilizing the k-ary tree hashing algorithm. Figure 1 presents a comprehensive graphical summary of the proposed strategy.

3.1. Preprocessing. In this section, we discuss the preprocessing for the proposed method; initially, background subtraction is performed using change detection and connected components-based method and approaches. We

connected a component labeling method for segmenting the human silhouette and identifying skin pixels as salient sections. Upon obtaining the skin components, we segmented the human contour using histogram-oriented thresholding. Using Otsu's technique, various threshold values of Δ (equation (1)) were modified, and the maximum color intensity of stochastic histogram $hto(x, y)$ is expressed as follows:

$$hto(x, y) = \begin{cases} IR\left(\frac{TOi + Tehi_{max}}{4}\right), & \text{if } Tehi_{max} \leq 5, \\ IR\left(\frac{TOi + Tehi_{max}}{2}\right), & \text{if } Tehi_{max} > 5, \end{cases} \quad (1)$$

where IR is the overweight, TOi is a threshold which is proposed by Otsu's method, and $Tehi_{max}$ is the largest location of skin frequency over extracted histogram index. This method is functional for every grey scale sector of given image, which is expressed as follows:

$$Imge(x, y) = \begin{cases} (0, 0, 0), & \text{if } mgr(x, y) = 0, \\ (Imgr(x, y), Imgr(x, y), Ib), & \text{if } img(x, y) = 1. \end{cases} \quad (2)$$

Figure 2 presents the example results' background subtraction over given set of images from Avenue state-of-the-art dataset.

3.2. Human Detection and Crowd Analysis. In this section, human recognition and crowd analysis are discussed. After extracting the human contour, a mask was placed on the human figure to identify the human body shape. The frame encompasses all regions of humans, with red representing humans. Once human validation is accomplished in crowd-based datasets, the next step is to detect normal and deviant human behavior through crowd grouping and evaluation. Initially, Gaussian smoothing [38] is used to recognize humans. Gaussian smoothing functions similarly to thresholding because it softens an image. The Gaussian average difference measures the degree of reduction. The Gaussian gives an "evaluation model" of a given pixel's surroundings, with the estimate increasing toward the central pixel's illumination. Figure 3 shows the results of human detection and crowd analysis.

3.3. Bag of Features. In this step, we discuss the bag of features and the extraction approaches to finding these features over state-of-the-art datasets. We utilized three robust feature extraction frameworks: periodic motion, motion direction flow, and shape autocorrelation features. Algorithm 1 defines the complete picture of the feature extraction procedure.

3.4. Periodic Motion. This spatial feature detects human motion over body segments. The region of attention is the component of the human body that triggers repeating

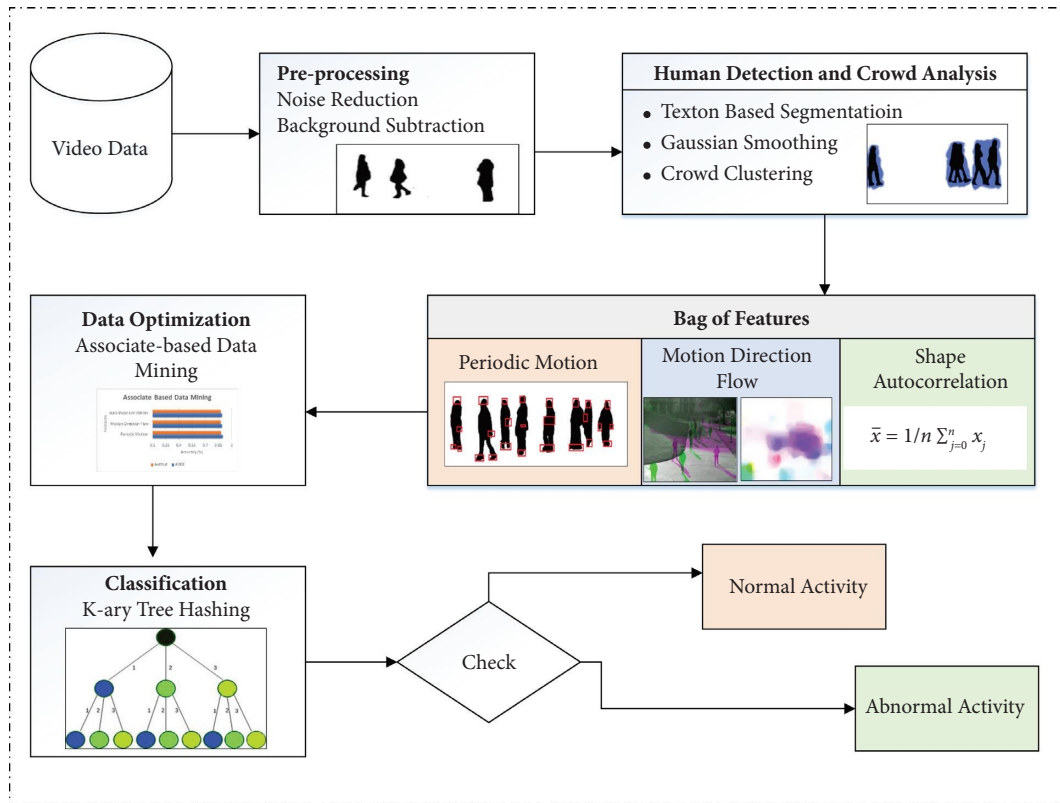


FIGURE 1: The architectural diagram of the proposed framework of pedestrian anomaly prediction.

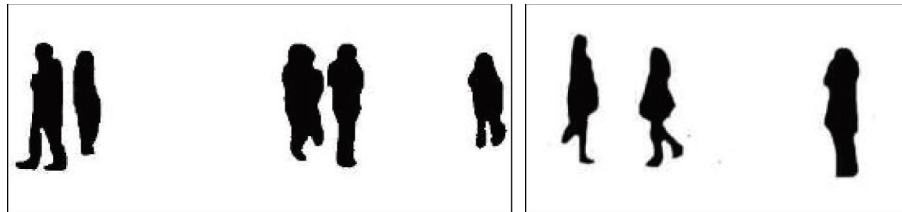


FIGURE 2: Background subtraction results on human subjects.

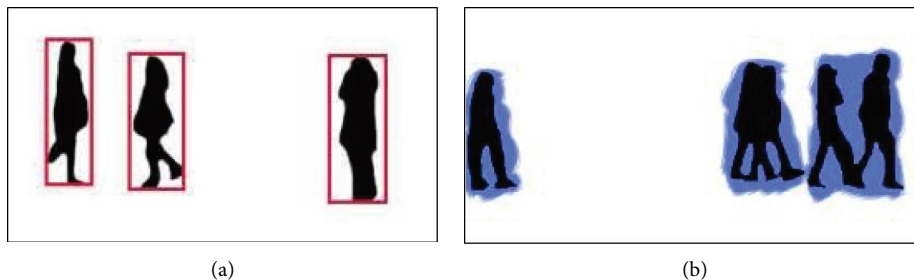


FIGURE 3: (a) Human detection and (b) crowd analysis via Gaussian clustering.

patterns. Such a region is identified using a fundamental examination of the human anatomy. A bounding section displays the area of focus. Equation (3) illustrates the mathematical formula for periodic motion.

$$PM(k) = \alpha \sin(\omega k + t), \tag{3}$$

where $PM(k)$ represents periodic motion and $\alpha \sin(\omega k + t)$ illustrates the human motion that is recurring in any specified order of images (see Figure 4).

```

Input: Frame_data
Output: Feature_vect ( $fe_1, fe_2, fe_3, \dots, fe_n$ )
Extracted_features  $\leftarrow []$ 
Data  $\leftarrow$  GetData_F()
Data_size_F  $\leftarrow$  GetData_F_size()
Procedure PAP (Video, Images)
Features_Vect  $\leftarrow []$ 
Denoise_Input_Data  $\leftarrow$  Preprocessing (Win, Median)
Sampled_Data (Denoise Data)
While exit invalid state do
[PM, MDF, SAC]  $\leftarrow$  ExtractFeatures (sample data)
Feature_Vect  $\leftarrow$  [PM, MDF, SAC]
Return MainfeaturesVector

```

ALGORITHM 1: Bag of features.

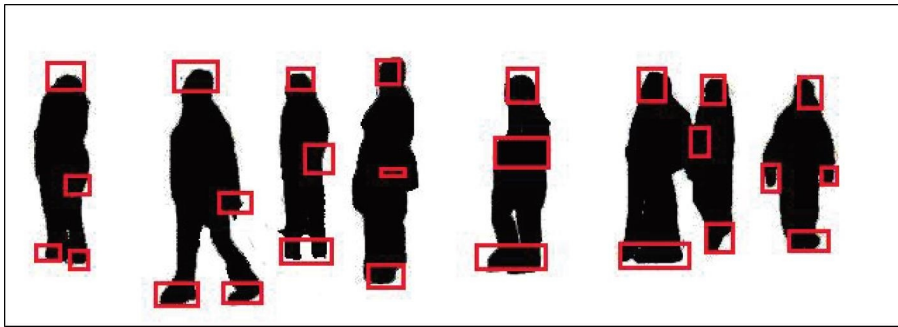


FIGURE 4: The example results of periodic motion over Avenue dataset.

3.5. Motion Direction Flow. In the motion flow features, a color scheme separation model was implemented for the provided data images. With it, we identify the motion direction and colorize it. We can now obtain such indicators and convert them to feature vector for use in future estimation and computations. These motion flow properties are formulated as follows:

$$M_{vf} = \sum_0^n \text{if } v(\text{piv}), \quad (4)$$

where M_{vf} defines the motion flow vector, piv is the detected RGB attributes, and if v are the input image and frame data. Figure 5 illustrates the result of direction flow features.

3.6. Shape Autocorrelation. We implemented the virtualization methodology to the detected human body as part of the shape-based autocorrelation feature. We considered particular positions to be centered and constructed a 5-by-5-pixel window from the middle of image 1 to image n . Then, we replicated this procedure for every discovered human body. After getting a 5×5 frame of all critical human body areas, autocorrelation was calculated. The median of x_j, \dots, x_n is determined using the given method described in the following equation:

$$\bar{x} = \frac{1}{n \sum_{j=0}^n x_j}, \quad (5)$$

where \bar{x} is the median and x_j is the given image sequence while time is represented as

$$R_p = \frac{s_p}{s_k}, \quad (6)$$

where s_k is the total variance of given data and R_p is the correlogram index values.

3.7. Data Optimization: Associate-Based Data Mining. The association rule-based features mining approach enables us to select the most distinctive characteristics by eliminating redundant and discordant elements from the obtained sample, which typically can reduce motion activity identification and pedestrian anomaly prediction accuracy and reliability. This is an underside method that begins with empty features and functionality nof and gradually adds new features depending on the selection of an optimization technique. This could reduce the variance, resulting in greater precision. Various sectors, including surveillance equipment, smart healthcare systems, and computer vision-dependent intelligent systems, frequently employ the association rule-based features mining methodology.

In the current infrastructure, the Bhattacharyya location estimation features minimization method is applied for numerous event-based classifications. It can calculate the separation rating $\text{nof}((x, y))$ between x and b regions and then validate it. This methodology enables the reduction of the significant aspects of domain material, but the appropriate outcome for pedestrian anomaly prediction is reliant



FIGURE 5: (a) The mapped direction flow and (b) the graphical representation of motion direction flow features.

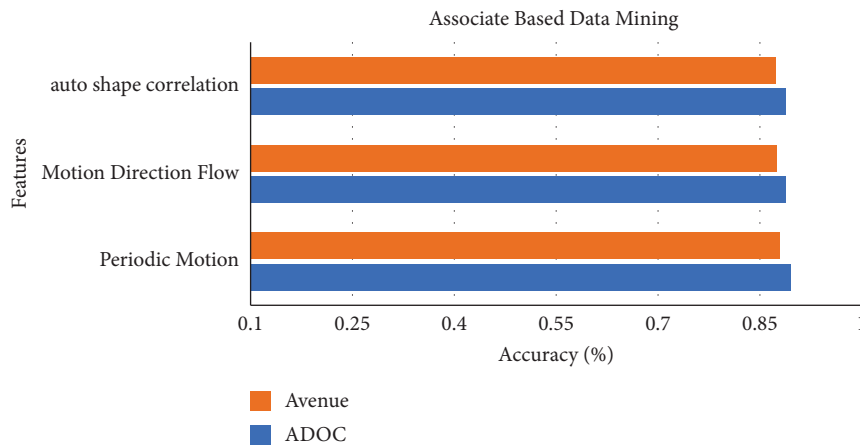


FIGURE 6: Rule-based features mining method over the Avenue and ADOC datasets.

on the optimization method of the features mining techniques.

$$\text{nof}_{(x,y)} = (k_x - k_y) \left(\frac{\Sigma_x - \Sigma_y}{2} \right) (k_x - k_y)^t, \quad (7)$$

where $\text{nof}_{(x,y)}$ is the optimized features vector, k_x are the mean index values, Σ_x are the class index covariance, x and k_y are the value average, and Σ_y are the class covariance of y for M statistics of normal and abnormal classes. The optimized predicted solution is calculated as follows:

$$\text{OFv} = \frac{1}{N^2} \sum_{u=1}^M \sum_{v=1}^M \text{nf}_{(u,v)}. \quad (8)$$

An identification assessment methodology is presented for predicting pedestrian behavior for the given dataset to gather feature extraction that is anticipated to decrease prediction errors and increase interclass accurateness across feature information. For the Avenue dataset, periodic motion, motion direction flow, and auto-shape correlation-based features are extracted. For the ADOC dataset, periodic motion, motion direction flow, and auto-shape

correlation-based features are extracted. Figure 6 represents the features optimization results of Avenue and ADOC datasets.

3.8. Classification: K-Ary Tree Hashing. The K-ary tree hashing technique utilized the procedures of the rooted tree which defines the node having maximum K children. For the recognition and classification approach, the lowest hashing approach has been applied which is considered as the prestep of K-ary tree hashing. This approach is based upon the conditional property utilizing the similarity check between two subsets B_i and B_j . B set of A hashing value estimator $|k_b|A/b = 1$ for B is used; furthermore, the min hashing operator for B is $k_b(B)$.

To produce a permutation index, we have

$$k_b(j) = \text{mod}((L_{d_j} + M_d), N_d). \quad (9)$$

The N_d, L_{d_j}, M_d are the available random values which are extracted from the dataset. The K-ary tree hashing technique uses two strategies to discover the optimal result (see Figure 7): a naive approach for determining the

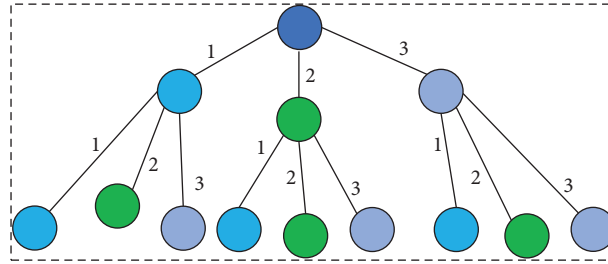


FIGURE 7: The architectural diagram of K-ary tree hashing.

Require: L, K_i
Ensure: $T(v)$
 % N is neighbor, L is Data, and T is size fixing approach%
 (1) $Temp \leftarrow \text{sort}(L(K_i))$
 (2) $j \leftarrow \min(j, |K_i|)$
 (3) $t(i) \leftarrow [i, \text{index}(temp(1:j))]$

ALGORITHM 2: Naïve approach for K-ary tree hashing.

frequency of neighboring nodes and min hashing to define the size of whatever statistic. The naive method is described in Algorithm 2.

4. Experimental Results and Analysis

4.1. Validation Methods. Numerous validation approaches have been studied in existing literature in order to assess the performance of a proposed system. These approaches include train-test holdout validation, leave-one-out, and k-fold cross-validation, respectively. However, subject overlap occurs when these methods are applied to datasets with several instances per subject. Therefore, in reviewing the literature and to prevent the challenges of subject overlap, leave-one-subject-out (LOSO) cross-validation is the best applicable approach. In this context, we adopted the LOSO cross-validation method for our proposed system. In LOSO, one subject is separated for testing while the system is trained on the provided data of the remaining subjects. The identical method is repeated for all subjects. Finally, the mean recognition rate of all subjects is determined.

4.2. Evaluation Metrics. Different evaluation metrics, such as accuracy, precision, recall, and F1 score, have been employed to assess the performance of the recommended systems. In our context, LOSO CV accurately recognized pedestrian anomalies in the benchmark datasets. Precision measures how well a proposed system can predict a certain class. Recall indicates the number of times a proposed system was able to recognize a particular category while f -measure delivers an individual score that addresses both precision and recall covers in a single number.

4.3. Dataset Descriptions. The Avenue database [30] contains 16 activities and 21 testing films. The videos were recorded on the central avenue of the Chinese University of Hong Kong (CUHK), with a range of 30652 pictures. The exercise videos are an example of formal conditions. Challenging clips involve both conventional and abnormal actions. The database consists of few more difficulties, including low camera calibration, anomalies, and median shapes, along with several samples of unusual entities, such as strange, inappropriate, and strange.

The second dataset is our PAP framework architecture design, ADOC [39], a 24 hrs. The second dataset in our study is the PAP framework architecture design, specifically the ADOC dataset [38]. This dataset consists of a 24-hour workout videos that includes 25 distinct sequence activities and a total of 721 instances. This is the largest gathering available with unique entity identifiers for security monitoring. A surveillance video camera installed on a large university campus has acquired the statistics. It exposes a walkway that incorporates different facilities and depicts the hectic everyday activity of youngsters, instructors, and staff.

4.4. Hardware Platform. MATLAB (R2022a) and Google Colab are used for all analysis and experimentation while Intel (R) Core (TM) i7-8665U CPU 1.90 GHz with 64-bit Windows 11 Pro was used as the main portable device. Additionally, the notebook involves 16 GB of internal memory (RAM).

4.5. Experimental Evaluation and Results. All the experiments are performed on two challenging benchmark datasets including Avenue and ADOC, respectively. The

TABLE 1: Confusion matrix of proposed PAP over Avenue database.

Scenes	ADR (%)	Error rate (ER) (%)
Scene 1	86	14
Scene 2	88	12
Scene 3	91	9
Mean (ADR, FPR, ER)	88.3	11.6

TABLE 2: Confusion matrix of proposed PAP over ADOC database.

Scenes	ADR (%)	Error rate (ER) (%)
Scene 1	87	13
Scene 2	89	11
Scene 3	92	8
Mean (ADR, FPR, ER)	89.3	10.6

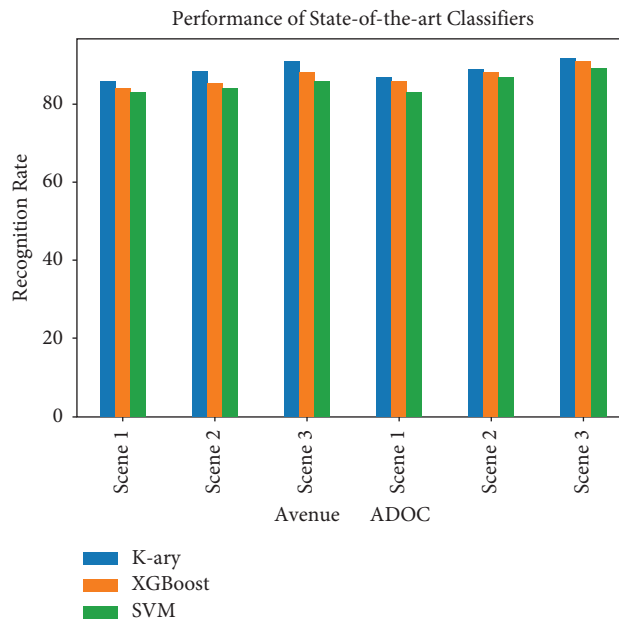


FIGURE 8: The comparison result of the recognition rate of k-ary with XGBoost and SVM classifiers.

TABLE 3: Comparison of the proposed PAP system using Avenue and ADOC databases.

Method	Avenue (%)	Method	ADOC
Handcrafted features [40]	74.5	Ano-graph [41]	81.4
Discriminative framework [20]	78.3	Leverage proxy task [42]	74.9
Binary classification (unmasking) [43]	80.6	Spatiotemporal auto-encoders [44]	84.60
Stacked RNN framework [45]	81.7		
Predicted future frame [46]	85.1		
Proposed PAP	88.3		89.3

next step is to distinguish between normal and abnormal activities in pedestrian crowd scenes with the support of the K-ary tree hashing algorithms. Table 1 presents the confusion matrix for the avenue database, highlighting an anomaly detection rate (ADR) of 88.3%, a false positive rate of 11.6%, whereas Table 2 represents the confusion matrix of ADOC database with ADR of 89.3% and ER of 10.6%.

In the next step (see Figure 8), we compare the performance of our proposed PAP system with other state-of-the-art classifiers on two benchmark datasets. The results have revealed that the proposed PAP system significantly outperformed other classifiers.

Table 3 presents the comparison of the proposed PAP method performance with other state-of-the-art systems using the Avenue and ADOC databases, respectively.

5. Discussion

The proposed PAP system is designed to predict the extraneous behavior of pedestrians using a k-ary tree hashing and associate-based feature mining. The research study is based on the denoising, human detection, bag of features, cues optimization, and anomaly detection steps. The proposed system is more accurate than the current state-of-the-art method. However, working on a real-time application presents possible challenges as it may involve processing overhead. Object occlusion is another issue that might make it harder to recognize pedestrians in crowded areas. This occurs when persons are obstructed from view by other individuals or objects, and it might make it more difficult to spot abnormal behavior.

Because human behavior is very diverse and complicated, it is challenging to construct anomaly prediction algorithms that precisely capture all conceivable actions. A considerable amount of training data is typically required for anomaly prediction algorithms to detect anomalies accurately. Yet, it can be difficult to acquire enough data to train the system in crowded environments, particularly for rare or unexpected behavior.

6. Conclusions

Humans are able to optimally find what they are searching for as the pedestrian vision system is adjusted to distinctive visual descriptors that make subjects of interest prominent. We introduce a method to replicate the human sensory system via bag of features such as periodic motion, shape autocorrelation, and motion direction flow. In addition, we optimized them to acknowledge the presence of pedestrians. Therefore, the resulting pedestrian anomaly predictor is an efficient and robust top-down saliency system.

The experimental results indicate that our PAP detector attains state-of-the-art performance on two benchmark datasets, Avenue and ADOC. Moreover, in both benchmark datasets, we determined that it outperformed all other current state-of-the-art methods evaluated. Based on these findings, it is promising to continue to explore more features impacted by human visual systems to detect more complex scenarios in various settings, including security, transportation, and emergency care.

Data Availability

The data are publicly available as described in the main text.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The research was cofunded by the European Union within the REFRESH project-Research Excellence for Region Sustainability and High-tech Industries ID No. CZ.10.03.01/00/22_003/0000048 of the European Just Transition Fund and by the Ministry of Education, Youth and Sports of the

Czech Republic (MEYS CZ) through the e-INFRA CZ project (ID: 90254) and also by the MEYS CZ within the project SGS ID No. SP 7/2023 conducted by VSB-Technical University of Ostrava.

References

- [1] C. Wang, H. Yao, and X. Sun, "Anomaly detection based on spatio-temporal sparse representation and visual attention analysis," *Multimedia Tools and Applications*, vol. 76, pp. 6263–6279, 2017.
- [2] E. S. Sezer and A. B. Can, "Anomaly detection in crowded scenes using log-euclidean covariance matrix," in *Proceedings of the VISIGRAPP (4: VISAPP)*, pp. 279–286, Funchal–Madeira, Portugal, January 2018.
- [3] C. Dhiman and D. K. Vishwakarma, "A review of state-of-the-art techniques for abnormal human activity recognition," *Engineering Applications of Artificial Intelligence*, vol. 77, pp. 21–45, 2019.
- [4] M. A. Ur Rehman, H. Raza, and I. Akhter, "Security enhancement of hill cipher by using NON-square matrix approach," in *Proceedings of the Proceedings of the 4th international conference on knowledge and innovation in Engineering Science and Technology*, Acavent, Hanoi, Vietnam, November 2018.
- [5] T. Chen, C. Hou, Z. Wang, and H. Chen, "Anomaly detection in crowded scenes using motion energy model," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 14137–14152, 2018.
- [6] K. Zhao, B. Liu, W. Li, N. Yu, and Z. Liu, "Anomaly detection and localization: a novel two-phase framework based on trajectory-level characteristics," in *Proceedings of the 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1–6, San Diego, California, USA, July 2018.
- [7] T. Ganokratanaa, S. Aramvith, and N. Sebe, "Anomaly event detection using generative adversarial network for surveillance videos," in *Proceedings of the 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1395–1399, Lanzhou, China, November 2019.
- [8] H. Rabiee, J. Haddadnia, H. Mousavi, M. Kalantarzadeh, M. Nabi, and V. Murino, "Novel dataset for fine-grained abnormal behavior understanding in crowd," in *Proceedings of the 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 95–101, Colorado Springs, CO, USA, August 2016.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [10] S. Gholami and S. Khashe, "Flight delay prediction using deep learning and conversational voice-based agents," *Am. Acad. Sci. Res. J. Eng. Technol. Sci.*, vol. 89, pp. 60–72, 2022.
- [11] Z. Chen and X. Huang, "End-to-end learning for lane keeping of self-driving cars," in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1856–1860, Los Angeles, CA, USA, June 2017.
- [12] S. Gholami and M. N. You, "Don't need labeled data for open-book question answering," *Applied Sciences*, vol. 111, 2022.
- [13] D. Bhargavi, E. P. Coyotl, S. Gholami, and K. Knock, "Who's there?--Identifying football player Jersey numbers with synthetic data," 2022, <https://arxiv.org/abs/2203.00734>.
- [14] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proceedings of the*

- Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6479–6488, Seattle, WA, USA, June 2018.
- [15] M. Xu, M. Gao, Y.-T. Chen, L. S. Davis, and D. J. Crandall, “Temporal recurrent networks for online action detection,” in *Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5532–5541, Montreal, BC, Canada, October 2019.
- [16] Y. Cong, J. Yuan, and J. Liu, “Sparse reconstruction cost for abnormal event detection,” in *Proceedings of the CVPR*, pp. 3449–3456, Colorado Springs, CO, USA, June 2011.
- [17] B. Zhao, L. Fei-Fei, and E. P. Xing, “Online detection of unusual events in videos via dynamic sparse coding,” in *Proceedings of the CVPR*, pp. 3313–3320, Colorado Springs, CO, USA, June 2011.
- [18] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, “Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression,” in *Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2909–2917, Seattle, WA, USA, June 2015.
- [19] J. Kim and K. Grauman, “Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates,” in *Proceedings of the 2009 IEEE conference on computer vision and pattern recognition*, pp. 2921–2928, Miami, FL, USA, June 2009.
- [20] A. Del Giorno, J. A. Bagnell, and M. Hebert, “A discriminative framework for anomaly detection in large videos,” in *Proceedings of the European Conference on Computer Vision*, pp. 334–349, Amsterdam, The Netherlands, October 2016.
- [21] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in *Proceedings of the 27th international conference on international conference on machine learning*, pp. 399–406, Haifa Israel, June 2010.
- [22] W. Luo, W. Liu, and S. Gao, “A revisit of sparse coding based anomaly detection in stacked rnn framework,” in *Proceedings of the Proceedings of the IEEE international conference on computer vision*, pp. 341–349, Montreal, BC, Canada, June 2017.
- [23] M. Genzel, J. Macdonald, and M. Marz, “Solving inverse problems with deep neural networks-robustness included,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, 2022.
- [24] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” *Journal of Machine Learning Research*, vol. 12, 2011.
- [25] M. Saqib, S. D. Khan, N. Sharma, and M. Blumenstein, “Crowd counting in low-resolution crowded scenes using region-based deep convolutional neural networks,” *IEEE Access*, vol. 7, 2019.
- [26] S. D. Khan, H. Ullah, M. Ullah, N. Conci, F. A. Cheikh, and A. Beghdadi, “Person head detection based deep model for people counting in Sports videos,” in *Proceedings of the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS, Taipei, Taiwan, September 2019*.
- [27] S. D. Khan and S. Basalamah, “Scale and density invariant head detection deep model for crowd counting in pedestrian crowds,” *The Visual Computer*, vol. 37, 2021.
- [28] S. D. Khan, Y. Ali, B. Zafar, and A. Noorwali, “Robust head detection in complex videos using two-stage deep convolution framework,” *IEEE Access*, vol. 8, 2020.
- [29] B. Zong, Q. Song, M. R. Min et al., “Deep autoencoding Gaussian mixture model for unsupervised anomaly detection,” in *Proceedings of the International conference on learning representations*, Vancouver, BC, Canada, April 2018.
- [30] C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in matlab,” in *Proceedings of the Proceedings of the IEEE international conference on computer vision*, pp. 2720–2727, Sydney, Australia, December 2013.
- [31] A. Del Giorno, J. A. Bagnell, and M. Hebert, “A discriminative framework for anomaly detection in large videos,” in *Proceedings of the European conference on computer vision*, pp. 334–349, Amsterdam, the Netherlands, October 2016.
- [32] M. Javan Roshtkhari and M. D. Levine, “Online dominant and anomalous behavior detection in videos,” in *Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2611–2618, Portland, Oregon, USA, June 2013.
- [33] R. V. H. M. Colque, C. Caetano, M. T. L. de Andrade, and W. R. Schwartz, “Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, pp. 673–682, 2016.
- [34] Y. S. Chong and Y. H. Tay, “Abnormal event detection in videos using spatiotemporal autoencoder,” in *Proceedings of the International symposium on neural networks*, pp. 189–196, Anchorage, Alaska, USA, May 2017.
- [35] W. Liu, W. Luo, D. Lian, and S. Gao, “Future frame prediction for anomaly detection--a new baseline,” in *Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6536–6545, Salt Lake City, UT, USA, June 2018.
- [36] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, “Abnormal event detection in videos using generative adversarial nets,” in *Proceedings of the 2017 IEEE international conference on image processing (ICIP)*, pp. 1577–1581, Beijing, China, September 2017.
- [37] S. Akcay, A. Atapour-Abarghouei, and T. P. G. Breckon, “Semi-supervised anomaly detection via adversarial training,” in *Proceedings of the Asian conference on computer vision*, pp. 622–637, Perth, Australia, May 2018.
- [38] B. Garg and G. K. Sharma, “A quality-aware Energy-scalable Gaussian Smoothing Filter for image processing applications,” *Microprocessors and Microsystems*, vol. 45, pp. 1–9, 2016.
- [39] M. Pranav and L. Zhenggang, “A day on campus-an anomaly detection dataset for events in a single camera,” in *Proceedings of the Proceedings of the Asian Conference on Computer Vision*, Kyoto, Japan, December 2020.
- [40] W. Liu, W. Luo, D. Lian, and S. Gao, “Future frame prediction for anomaly detection—a new baseline,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6536–6545, Salt Lake City, UT, USA, June 2018.
- [41] M. Pourreza, M. Salehi, and M. Sabokrou, “Ano-graph: learning normal scene contextual graphs to detect video anomalies,” 2021, <https://arxiv.org/abs/2103.10502>.
- [42] H. Park, J. Noh, and B. Ham, “Learning memory-guided normality for anomaly detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June 2020.
- [43] R. Tudor Ionescu, S. Smeureanu, B. Alexe, and M. Popescu, “Unmasking the abnormal events in video,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2895–2903, Venice, Italy, October 2017.
- [44] Y. S. Chong and Y. H. Tay, “Abnormal event detection in videos using spatiotemporal Autoencoder,” in *Proceedings of*

- the International Symposium on Neural Networks*, pp. 189–196, Sapporo, Japan, June 2017.
- [45] W. Luo, W. Liu, and S. Gao, “A revisit of sparse coding based anomaly detection in stacked rnn framework,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 341–349, Venice, Italy, October 2017.
- [46] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, “Learning temporal regularity in video sequences,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 733–742, Las Vegas, NV, USA, July 2016.