EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Abnormal motion pattern detection in video sequences by an unsupervised approach

Himal Acharya[1, 2] and Basanta Joshi*[1]

[1]Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Lalitpur, Nepal
[2]Department of Electrical and Electronics Engineering, Nepal College of Information Technology, Pokhara University, Lalitpur, Nepal

Received 28 June 2020
Revised 6 February 2021
Accepted 15 February 2021

Abstract

Identifying anomalous motion behavior in video sequences is a challenging task. Manual annotation of a large number of surveillance videos is time-consuming because of the limited human brain's visual attention. This work presents a new framework to detect abnormalities from unlabeled videos using motion patterns for the normal and abnormal event. This paper proposed an unsupervised hierarchical agglomerative clustering technique for finding the abnormal behavior motion patterns. Dense trajectories of feature points were extracted and grouped into feature points for different interval groups with characteristics of the feature points' motion speed. With results from partitioning interval groups by hierarchical clustering, anomalous motion patterns were localized in surveillance video sequences. We performed experiments on publicly available datasets containing different abnormal samples. The experimental results showed that the proposed framework achieved the highest frame-level accuracy of 96.68% for the UMN dataset. The experiment has achieved the highest rate of detection (up to 98.63%) for UCSD pedestrian datasets. The proposed framework has achieved outstanding performance in both pixel level and frame level evaluation.

Keywords: Anomaly detection, Dense trajectories, Hierarchical clustering, Motion pattern, Surveillance video

1. Introduction

In recent years, video surveillance systems are widely used for different security purposes [1] and traffic management. Detecting abnormal behavior has been active research in the computer vision and pattern recognition field for public safety [1]. Abnormal events are those activities that can threaten public safety and cause a hazard to human health and life. Abnormal behavior rarely occurs in video sequences. Humans can identify normal and abnormal events by manually inspecting short video frames. However, it is expensive to manually identify abnormal behavior from a large amount of data generated every day [2]. It is challenging to give meticulous attention over a long period. In such scenarios, machine vision algorithms can identify anomalous motion and behavior. Any activities that deviated from normal pedestrian activities are defined as abnormal (anomalous) activities [1]. In real-world scenarios, abnormal behavior is a complex behavior [3, 4] and contextual. Anomaly's definition may differ from one scene to another scene depending upon activities. A pedestrian walking on a pedestrian sideway is considered normal behavior but running, cycling is considered abnormal.

Abnormal event detection methods are categorized into two sub-groups, supervised and unsupervised methods. Existing supervised methods [5-10] required manual annotation of training videos that contain only normal events. However, it is not practical for a framework to have prior knowledge of every normal event. Any normal event absent during training videos may be identified as an abnormal event. So, this cannot be extended to new normal behavior. It is also time-consuming for human beings to specify normality in real-world scenarios. Existing unsupervised methods [11, 12] do not work well in detecting irregular motions and high dense abnormality. Test results show no significant change in accuracy and rate of detection by existing supervised and unsupervised methods.

This paper attempts to optimize the detection of abnormalities in surveillance video sequences by a hierarchical agglomerative clustering approach. The motion speed of feature points is calculated from dense trajectories of a temporal window of 15 frames. The homography is estimated to improve the accuracy of video sequences suffered from perspective distortion. Feature points are grouped into a certain number of intervals based on motion speed and dominant motion with no training phase. The hierarchical agglomerative clustering is then applied to cluster motion patterns to produce abnormal motion patterns and localize them in surveillance videos. The datasets used in the literature are surveillance videos of the pedestrian. So, this paper focuses on pedestrian activities. The significant contributions made by this paper based on UCSD Peds and UMN datasets as explained as,

i) The proposed unsupervised method extracts pixel-level low-level features obtained from dense trajectories that can detect sudden irregular motions. It can handle video sequences with high dense abnormality by taking the motion speed of non-stationary feature points.
ii) We propose a mechanism to detect and localize anomalous motion patterns from unlabeled videos by considering the pedestrian's dominant motion characteristics.

*Corresponding author. Tel.: +97 798 5119 0040
Email address: basanta@ioe.edu.np
doi: 10.14456/easr.2021.53

iii) We propose a hierarchical clustering to obtain normal and abnormal events. The strategy to choose a hierarchical clustering approach gives the flexibility to choose the number of motion pattern groups according to our proposed work.

2. Related works

For abnormal detection, two aspects are significant: feature extraction and model to estimate abnormality from features. Anomaly detection is classified under supervised and unsupervised techniques. Zhouyu et al. [13] and Basharat et al. [14] used high-level features obtained from object tracking to detect abnormalities. These papers did not address frequency object shading or blurring. Colque et al. [15] and Li et al. [16] used the samples' normal class labels to learn the normal patterns. Colque et al. [15] used entropy and nearest neighbor method for anomaly detection in UCSD Data Set, Subway Data Set. In the nearest neighbor method, anomalies are determined by proximity to their neighbors. Despite the success of supervised methods, their use in the real world is limited. There is a scarcity of representation of abnormal events in different datasets. All supervised methods require a training dataset with normal and abnormal events to obtain a normality model. Li and Mahadevan et al. [7] detected spatial and temporal abnormalities in the video using a robust low-level feature called Mixture of Dynamic Textures (MDT). This MDT jointly modeled the dynamics and appearance in crowded scenarios. Weixin et al. [7] improved the performance compared to Mahadevan et al. work [8] by performing a conditional random field filter on multi-scale anomaly maps, which have shown more effectiveness in modeling complex crowded cases than high-level features [13, 14]. Cong et al. [10] constructed a dictionary using normal events during the training phase and reconstruction error as the parameter for anomaly detection during the testing phase. A Multi-scale Histogram of Optical Flow (MHOF) was extracted to represent an event, and sparse reconstruction cost (SRC) determined anomalies.

Antic et al. [17] used a video parsing approach, which established a set of hypotheses using foreground detection with a support vector machine (SVM). Every pixel has an estimated probability of whether it belonged to the foreground. A probabilistic model was built using hypotheses from training videos. Abnormalities were detected from the test video using hypothesis tests that had achieved comparable detection results but required the entire dataset in advance. Wu et al. [6] detected abnormal events in crowd escape sequences using chaotic invariants. The Gaussian mixture model did the distribution of chaotic invariants. The trajectories were used to find if the crowd behavior is normal. That algorithm failed when escape behavior happened in the same direction (example-all escape in the same directions).

Similarly, Wu et al. [5] used the probabilistic model (Bayesian framework) in the training phase without escape motion patterns. Crowd behavior motion patterns were estimated by using optical flow fields. By using a Bayesian formulation, crowd escape motion was detected in low or medium-density crowded scenes. This Bayesian formulation cannot be applied to high-density crowded scenes. Pennisi et al. [18] proposed real-time crowd behavior detection based on background modeling using segmentation with the combination of Kanade-Lucas-Tomasi (KLT) feature extraction without the need for the training phase. It used an activity map to analyze image entropy trends (based on two consecutive frames) and temporal occupancy variation (TOV). Activity maps can identify only a few pixels because of the changing density of objects in a crowded scene. By using a threshold on entropy and TOV, an anomalous frame was identified. It fails when there is an abnormal event present from the beginning of the video sequence. Pixel level localization of anomalous events was not addressed in Pennisi et al. [18]. KLT feature tracker [19] based on sparse optical flow did not show sufficient quality and quantities of trajectories and was not robust to sudden irregular motions.

Lin et al. [11] used an online weighted clustering approach combined with a multi-target tracker (MTT) algorithm. Adaptive Multi-scale Histogram Optical Flow features (AMHOF) were obtained within the region of interest with optical flow vectors. This algorithm could track and detect only slow changes. Clusters with normal AMHOF were considered normal frames, and frames with the above threshold were considered abnormal frames. Kalman filter tracking was applied after online weighted clustering results to detect missing anomalies in frames. The Kalman filter method can track and detect only changes and fails to detect irregular motions. Abuolaim et al. [20] used k-means clustering at a coarse level to cluster data points to normal events and abnormal events. K-means required prior knowledge about the number of clusters. Such an approach does not provide the flexibility of choosing the number of clusters based on motion patterns.

Manjula [21] used background-subtraction to obtain high-level features. Generalized Autoregressive Conditional Heteroscedasticity (GARCH), a statistical model was applied to find events in the pedestrian pathway. Multilayer Perceptron (MLP) classified GARCH results as normal or abnormal events. The abnormal event was detected in video sequences at frame-level. Although these high-level feature-based approaches can find an abnormality, they are influenced by occlusions and noise. The high-level features-based approach fails to detect an abnormal event in both cluttered and crowded scenes. MLP is computationally expensive regarding time and resources [22]. Low-level motion features at pixel-level are extracted in our proposed framework to avoid the high-level features approach's detection problem in crowded scenes. The proposed framework can gain reliable performance without counting on large-set video sequences, long processing time [23, 24].

Both supervised and unsupervised methods discussed here have used sparse optical flow and KLT algorithm for tracking interest points. These trajectories obtained from KLT are not sufficient for action recognition [25]. As our work focuses on detecting anomalous behavior, understanding scenarios with a static background is not required. In this proposed approach, only objects that are changing their position in consecutive frames are considered. Training data are not needed in unsupervised methods, so this proposed framework can find anomalous motion patterns in test video clips.

3. Methodology

3.1 Proposed method

We propose a novel unsupervised abnormal event detection based on hierarchical agglomerative clustering of feature points. Our work focuses on pedestrian activities, so we consider pedestrian motion as normal motion (dominant motion). Skaters, cyclists, human panic behavior are considered abnormal events because they have different motion patterns than normal pedestrian motion. Our research goal is to detect abnormal behavior in pedestrian surveillance videos. Our proposed anomaly detection framework is illustrated in Figure 1. The proposed framework consists of three components- (i) feature extraction from dense trajectories (optical flow vector), (ii) grouping of feature points into n interval groups ranging from the minimum to the maximum speed of feature points, and (iii) clustering of feature points to obtain the abnormal motion pattern if the abnormal incident occurred. Such anomalous motion patterns are localized in the pedestrian surveillance videos.

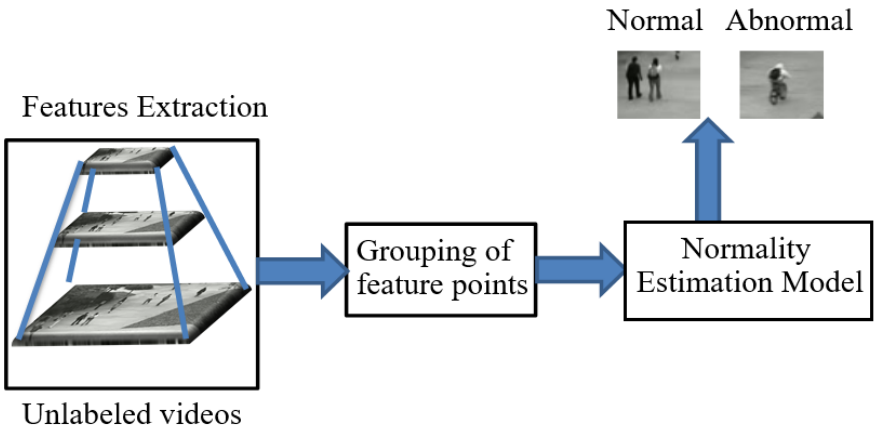

Figure 1 Flow graph of the proposed model.

3.1.1 Video processing-features extraction

Before the abnormal event detection, preprocessing of video sequences is needed. Video frames are first converted to gray-scale images and resized to multiple scales to cover image sequences with different sizes. Normal and abnormal events are distinguished by motion speed. Pedestrians walk motion is dominant in video sequences and is considered a normal motion pattern. Other activities (carts, skaters, cyclists) movement in pedestrian sideways are not dominant motion and considered anomalous behavior in the video sequences.

To extract features from video sequences, Wu et al. [6] used optical flow and particle advection to estimate the particle's position. Li et al. [11] used multi-target tracking with two clustering algorithms to detect regions of interest. This region tracking suffered from segmentation and more tracking errors in crowded scenes. The region-based features were extracted from regions partitioned in video frames, and features included histogram of optical flow (HOF) [5, 10, 11, 15], dynamic textures [7, 8], spatial regions [12]. This region-based feature representation focuses on the dynamics in regions [20] and does not account for anomalous object appearances. Li et al. [11] used a multi-tracking algorithm with two clustering methods that cannot be optimized in real-time. Those methods detected only large changes in a local context as abnormalities but cannot detect sudden irregular changes. In our proposed approach, videos are described by dense trajectories that sampled dense points from each frame. Such dense points are tracked based on displacement information from the dense optical flow field. Dense sampling showed improved results over sparse interest points obtained from the KLT tracker [25].

In unsupervised settings, anomaly detection is performed without knowing priorly of test videos. The clustering approach determines the abnormal event group from dense trajectories of feature points. Given the input video, distinctive feature points are extracted and tracked using dense trajectory features proposed by Wang et al. [19] and Farneback [26]. The proposed framework uses motion speed to obtain a feature vector in the video event. This framework tracks the feature points for 15 frames, and new feature points substitute for them.

In a dense optical flow field, each feature point f_i, i = 1, 2, ..,n has a sequence of locations $(P_t)_i$, $(P_{t+1})_i$, …., $(P_{t+L})_i$ in a particular time in trajectory length L.

$P_t = (x_t, y_t)$ at frame t is tracked to the next frame *t+1* with median filtering.

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * \omega)|_{(\overline{x_t}, \overline{y_t})} \quad (1)$$

where M is the median filtering kernel, (x_t, y_t) is a position of feature point at the frame at time t, $(\overline{x_t}, \overline{y_t})$ is a rounded position of (x_t, y_t), ω is dense optical flow field of the feature point.

Different speed activities are recognized by trajectories representation at multiple temporal scales. With classical physics, speed is calculated as the derivative of the position. All feature points of trajectory length L frames have x and y gradients. Speed is calculated by taking the magnitude of the x and y gradient [27]. For a temporal window of trajectory length L, the speed s_i of feature fi calculated as:

$$s_i = || (P_{t+L})_i - (P_t)_i || / L \quad (2)$$

where $(P_t)_i$, $(P_{t+L})_i$ are a position of the feature point f_i at frame t and t+L, respectively, L is the trajectory's length. A feature point's speed is calculated by taking a feature point at every L^{th} frame to avoid drifting. Normal motion pattern (dominant motion speed pattern) has a similar motion pattern in video sequences. Abnormal motion pattern deviates from normal motion pattern. For this approach, the clustering method is adopted.

Speed of feature points calculated suffered from motion parallax for test videos where a stationary camera faces the pedestrian movement. Such motion parallax does not occur in test videos for which the stationary camera is facing sideways to pedestrian movement. Homography estimation is employed for reducing motion parallax based on direct linear transformation (DLT) [28, 29]. Camera parameters are not available for the UCSD Peds dataset. The homography matrix is to be estimated without a camera's intrinsic and extrinsic parameters. It is estimated from a stationary frame by taking five points that form a square on the ground plane, as shown in Figure 2.

Then points in Figure 2 (left) are mapped into a square by assuming its coordinates as shown in Figure 2 (right). Based on DLT, applying the homograph matrix to UCSD Peds 1 image sequences, perspective distortion is removed.

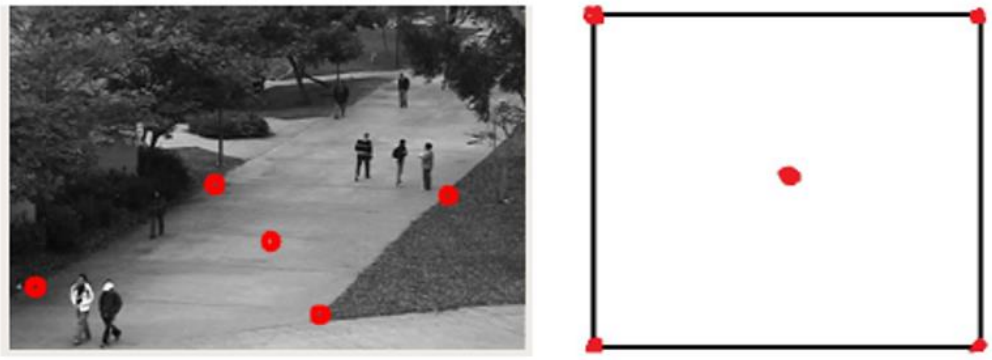

Figure 2 Five points extraction from the image (Left) and projected square (Right)

Figure 3 Projected image of UCSD Peds 1 test 01 at frame 1

Figure 3 shows the projected image after applying DLT on Figure 2 (Left) UCSD Peds 1 frame. The speed of feature points is calculated after perspective transformation for videos suffered from motion parallax.

3.1.2 Grouping of feature points

After finding the motion speed of pedestrian movement feature points in Section 3.1.1, such feature points fi are divided into n interval groups based on the method proposed by Abuolaim et al. [20]. Such interval groups contain the motion speed of feature points ranging from minimum to maximum motion speed. Each interval group is assumed as $D_j, j = 1,2\ldots., n.$ Each interval group *Dj* contains different size of feature points. The interval group containing the largest number of feature points is the dominant interval group. Abnormal activities occur in rare and not of dominant type. The last interval n^{th} group is of smaller size with a similar motion pattern. The average feature value $\overline{f_j}$ and the number of features in *Dj* characterize each interval group. Each interval group is normalized by dividing the interval group feature value and interval group size with the dominant interval group's average feature value and interval group size (number of feature points).

3.1.3 Normality estimation

From interval groups obtained from Section 3.1.2, an abnormal motion group is obtained by applying hierarchical agglomerative clustering. At first, the interval group D_j is assumed as a cluster by itself. The distance between clusters is calculated with a distance metric. Then, a new cluster is formed at each iteration by merging the nearest clusters [30]. This process continues until all the interval groups are gathered in one group.

Ward's method is used to ensure the intra-cluster difference in our approach. With the Euclidean distance square, Ward's criterion allows to minimize the total with-cluster variance or equivalently maximize the between-cluster variance. Hierarchical agglomerative clustering is executed to interval groups by Ward's method. Ward's criterion d_w is calculated as:

$$d^2{}_W(I_1, I_2) = \frac{nm}{n+m}\, d^2(G1, G2) \tag{3}$$

where G1, G2 are centroid of interval group 1 I_1 and interval group 2 I_2 and n, m be the number of feature points in two-interval groups respectively, d_w is the merging cost of combining interval group G1 and G2, $d^2(G1, G2)$ is the Euclidean distance between groups G1 and G2. The clusters with the lowest value of d_w are merged and form a single cluster.

Finding the number of clusters is an open problem in clustering. The dendrogram gives essential information to the identification of several groups. The hierarchical clustering approach provides flexibility to select many clusters during the partition of a dendrogram. Here, three clusters are chosen to partition n interval groups–normal, abnormal, and neutral. The neutral cluster widely separates the normal and abnormal clusters. Existing methods [5-7, 9, 11, 15, 16] used to classify events either as normal events or abnormal events depending upon the threshold conditions. Feature points near the threshold conditions are misclassified to a normal or abnormal group. While considering two groups- normal and abnormal, the motion speed of feature points slightly greater than the dominant interval

feature points is abnormal. Motion pattern is not accommodated to the dominant interval group and near to it cannot be considered as an abnormal event. Feature points having motion speed farthest away from the dominant interval group are considered as an abnormal event. In our experiments, pedestrian motion is a normal event group, but carts and cyclist motion are abnormal. But while extracting feature points and dividing them into interval groups, there are feature points near the dominant cluster. So, the neutral group contains trajectory features that cannot be accommodated to the normal and abnormal groups. Such abnormal feature points from abnormal cluster groups are localized in video sequences with f_i's corresponding trajectory positions.

4. Experiments

4.1 Datasets

This study uses UCSD Peds1, UCSD Peds2 [7], and UMN datasets [9]. These datasets are commonly used as a benchmark dataset for abnormal motion pattern detection in the literature. The proposed method's performance for abnormal motion pattern detection and localization was evaluated on uncrowded (UCSD Peds1, UCSD Peds2) and crowded (UMN) pedestrian surveillance scenes. Motion speed of feature points is extracted, and feature clustering is performed to detect anomalous motion patterns and localization of abnormal feature points in video sequences.

There are 34 training video sequences and 36 testing video sequences in the UCSD Peds1 dataset with 14000 frames with resolution 158× 238. In the UCSD Peds2 dataset, there are 16 training clips and 12 testing clips with 4560 total frames with resolution 240× 320. Testing videos of UCSD Peds1 and Peds2 datasets contain different anomalies like skaters, vehicles, and wheelchairs. Skaters, carts, cyclists, trucks move at a higher speed in comparison with pedestrians. Such activities are considered as abnormal motion in datasets. The UCSD Peds1 dataset is more challenging than UCSD Peds2 because the camera position creates motion parallax in UCSD Peds1.

The UMN dataset contains three different crowded scenes, and the resolution of each frame is 320×240 pixels. The total frames in the UMN dataset are 7,741 frames. UMN dataset contains 11 video sequences. In UMN test videos, anomalous events contain the panic escape behavior of the crowd.

4.2 Experimental setup and evaluation criteria

In this study, a temporal window for trajectory length L is set to 15, as proposed by Wang et al. [19]. Ward's method is used as a linkage criterion in hierarchical clustering. Experiments were conducted using PyCharm on the Linux platform with an Intel Core i5-5200 CPU with a 2.20 GHz processor and 8 GB of RAM. Two criterion pixel-level and frame-level criterion are common criteria to evaluate the performance of anomalous activity detection and localization [7]. At least one of the pixels in a frame is detected as abnormality compared with video sequences' frame-level ground truth annotations in frame-level criterion. Pixel-level criterion evaluates the localization by comparing the localization results with pixel-wise ground truth annotations. Pixel-wise ground truth annotations are available only for UCSD Peds1 and UCSD Peds2 datasets. A frame is true positive if more than 40% pixels of ground truth abnormal events are detected by the method. If it is negative and has abnormal behavior feature points' pixels localized, then a frame is a false positive. Two performance metrics - accuracy and rate of detection (RD) are used for evaluation as the same evaluation methodology used by Weixin et al. [7]. Rate of detection $RD= 1 – EER$, where EER stands for Equal Error Rate. An abnormal frame is correctly classified if at least one of its pixels is found abnormal compared to video sequences' frame-level ground truth annotations. This criterion is called the frame-level criterion. Pixel-level criterion evaluates the localization by comparing the localization results with pixel-wise ground truth annotations.

5. Results and discussion

5.1 Features labeling

We extracted dense trajectories from the test video sequences using the code publicly available by Wang et al. [25]. To illustrate feature labeling, we have taken a test video of the UCSD Peds1 Test 19 video. From dense trajectories of non-stationary objects, feature points are produced, feature points of the pedestrian walk, and cart movement are shown by green visualization in Figure 4. However, two persons standing near the cart's right side are not visualized with green color as we do not consider stationary objects. Anomalous behavior occurs with sudden changes. Such feature points are divided into ten interval groups, and hierarchical clustering is applied to interval groups. Figure 5 (a) shows that 10 ten interval groups are combined until all the data in one cluster, and a dendrogram tree is obtained. The dendrogram is partitioned into 3 clusters–normal cluster, abnormal cluster, and neutral cluster using agglomerative clustering. While partitioning into three clusters, interval group I1 makes one cluster, interval group I0, I2, I3, I4 make the second cluster group and the remaining interval groups make the third cluster group.

Figure 4 Dense trajectories of moving entities of UCSD Peds1 test 19.

The interval group with dominant motion (pedestrian walking) is considered a normal group, as shown by the green color in Figure 5 (b). The group farthest away from the normal group containing high-speed trajectories is an abnormal motion pattern represented by red color. Moreover, the third neutral group makes a good separation between the normal and abnormal groups represented by black color in Figure 5 (b). Figure 5 (b) contains the feature point speed in pixel /frame for a video clip with visualization of normal motion pattern, abnormal motion pattern, and neutral group.

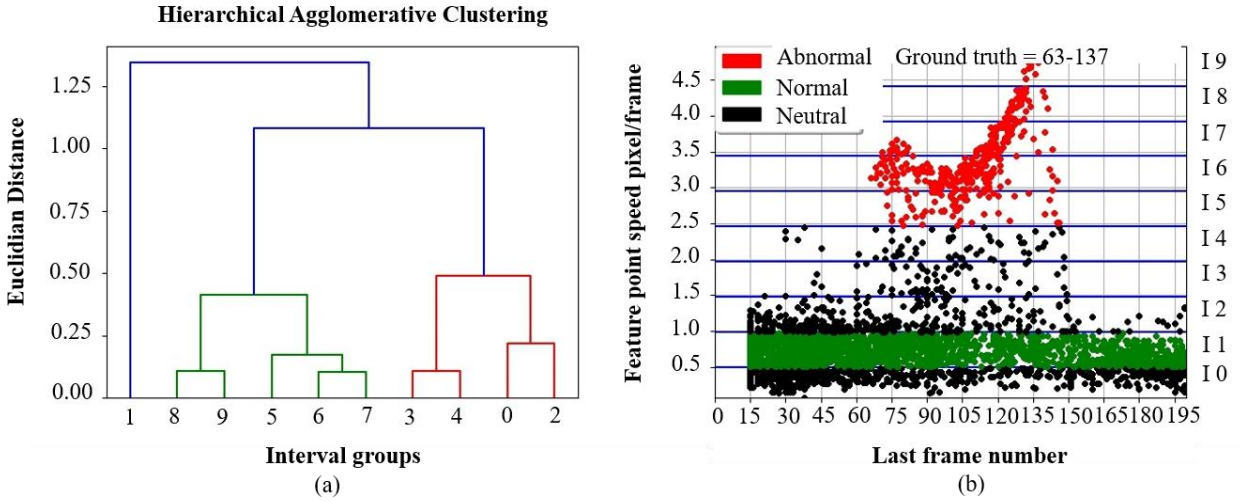

Figure 5 (a) Dendrogram of UCSD Peds1 test 19 (b) Plot of features point in pixel/frame with the normal event, abnormal event, neutral group of UCSD Peds1 test 19 video.

5.2 Effect of homography estimation

In UCSD Peds1 and UMN datasets, there arises perspective distortion. Stationary cameras facing towards the pedestrian walkway in UCSD Peds1, and crowds in UMN suffer distortion. Such a problem does not appear in the UCSD Peds2 dataset because the camera plane is parallel to the pedestrian movement. This distortion does not give the actual speed of feature points in video sequences. It is estimated from a stationary frame for homography by taking five points that form a square on the ground plane. The homography matrix is obtained by applying the direct linear transformation. Here, the effect of homography projection is illustrated with a test video example of the UCSD Peds1 Test19 video. In UCSD Peds1, the homography matrix is defined as:

$$H = \begin{bmatrix} 3.50252633e+00 & 6.12500188e+00 & -8.29249592e+02 \\ -3.85040199e-01 & 9.11910516e+00 & -7.02482457e+02 \\ 1.71963388e-03 & 2.03844857e-02 & 1.00000000e+00 \end{bmatrix} \quad (4)$$

In the test video, UCSD Peds1 test 19 vehicle appears in the 63rd frame to 137th frame, approaching the camera. In Figure 6 (a), an abnormal vehicle motion pattern is detected and localized where the vehicle just started to appear in frame sequences. However, when a vehicle reaches about mid of the pedestrian walkway, a vehicle's motion pattern is not detected, as shown in Figure 6 (b). In the top part of Figure 6 (a) and 6 (b), the test video frame sequence is shown on which the proposed framework is applied, and the detection and localization results are shown in the lower part of Figure 6 (a) and 6 (b). After estimating homography, such an undetected vehicle (considered abnormal motion pattern) in Figure 6 (b) lower part is detected and localized, as shown in Figure 7.

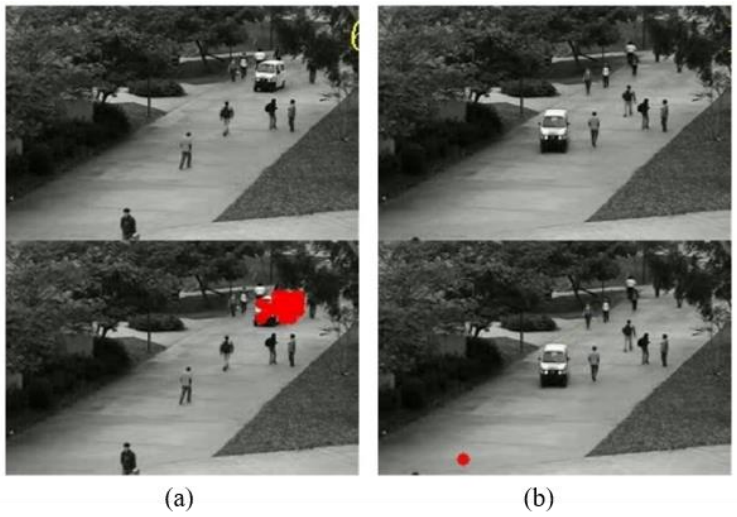

Figure 6 Visualization of anomalous feature points of UCSD Peds1 test 19 video (a) at frame 68 (b) at frame 101 before projection.

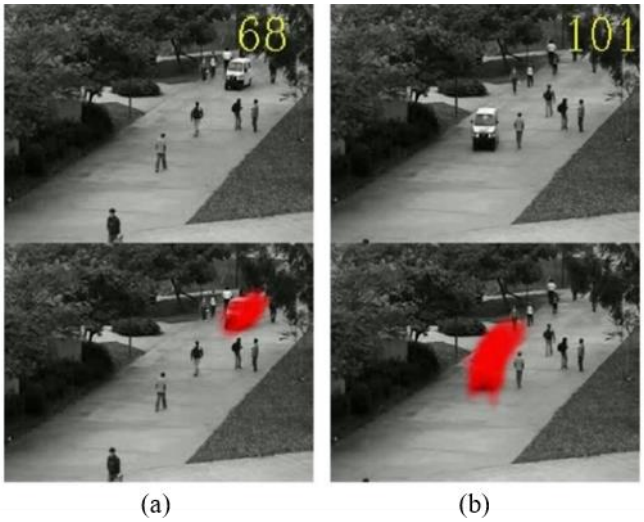

Figure 7 Visualization of anomalous feature points of UCSD Peds1 test 19 video at (a) frame 68 (b) frame 101 after projection

Figures 7 (a) and 7 (b) detect and localize the cart's movement as an anomalous motion pattern while the cart is in pedestrian sideways, which was not detected in all cases in Figure 6 (b).

Table 1 Effect of homography estimation in UCSD Peds1 test 19 video

Method	Frame level accuracy
Without projection	37.18%
With projection	82.41%

From Table 1, it is inferred that projecting feature points to the ground plane improves accuracy in the UCSD Peds1 dataset.

5.3 Evaluation of UCSD Ped dataset

Frame-level and pixel-level criteria are used to compare the proposed approach with other state-of-art approaches. Pixel-level criterion [7] is available for UCSD Peds1 and Peds2 datasets. Since the proposed method has a fixed parameter, only one point is obtained, and receiver operating characteristics (ROC) cannot be obtained. Hierarchical agglomerative clustering improves the detection results in both frame-level and pixel-level.

Table 2 Comparison of experimental results with state-of-art methods on UCSD Peds dataset

Method	Frame-level (%) RD		Pixel-level (%) RD	
	UCSD peds1	UCSD peds2	UCSD peds1	UCSD peds2
Proposed approach	94.03	98.63	94.43	90.86
Video parsing [17]	81.9	85.8	67.6	-
MDT-CRF [7]	82.2	81.5	64.9	70.1
Online weighted clustering- MTT [11]	82.9	86.1	76.7	78.9
HOFME [15]	66.9	80	-	-
HVOFA [16]	86.9	93.4	85.5	93.1

The proposed approach is compared with existing state-of-art methods on UCSD Peds1 and Peds2 data, including video parsing [17], MDT-CRF [7], online weighted clustering [11], HOFME [15], and HVOFA [16] with RD metric. Some existing methods did not calculate pixel-level RD. Therefore, dashes are used to represent them in the Table 2. Video parsing, MDT-CRF, online weighted clustering, HOFME, and HVOFA show comparable detection results but fail to detect sudden abnormal motion changes. Our proposed framework achieves the best rate of detection (RD) on frame-level and pixel-level annotation in the UCSD Peds1 dataset. In UCSD Peds2, the proposed framework obtained the comparable RD metrics in pixel-level and highest RD on frame-level achieving 98.63%. This result is because our proposed framework based on feature points of dense trajectories uses hierarchical clustering to detect abnormality from video sequences. The proposed framework gives more accurate anomaly detection and localization than the existing methods discussed above. It can be seen that the rate of detection (RD) of our proposed framework has been improved, which shows that our framework can detect more abnormal events than [7, 11, 15-17]. A neutral group between normal and abnormal event groups during the clustering stage reduces false positives.

5.4 Evaluation of the UMN dataset

For the UMN dataset, the only frame-level criterion is used to compare the performance with existing state-of-art methods. Pixel-level ground truth annotation is absent, so all other state-of-art methods perform a frame-level evaluation.

Table 3 Comparison of experimental results with state-of-art methods on UMN dataset

Method	Frame level accuracy%		
	Lawn	Plaza	Overall
Proposed Approach	95.01	98.35	96.68
Bayesian Approach [5]	95.36	96.63	95.9
Social Flow Model [9]	84.41	90.83	87.62
Chaotic Invariants [6]	90.62	91.58	91.1
Force Field [12]	88.69	77.92	83.3

The proposed method is compared to Bayesian, social flow, chaotic invariants, and force field methods for evaluating the model's performance. Table 3 shows a comparison of five methods in the lawn and plaza scene. The proposed method achieves higher accuracy of 98.35% in the plaza scene than in the lawn scene with an accuracy of 95.01%. However, in the lawn scene, the proposed methods show comparable detection results with the Bayesian model. However, the Bayesian approach is supervised and requires a training dataset but is limited. It is challenging to detect escape events by a force field, chaotic invariants, and social flow model methods because they model crowd behavior in non-escape cases. The proposed framework models crowd behavior in both escape-cases and non-escape cases. The proposed method's overall accuracy is 96.68%, which outperforms the sparse optical flow and social force model.

Ionescu et al. [31] used k-means clustering and one class SVM after features extraction from the last convolution layer of a pre-trained neural network. All the points are grouped into 30 clusters using k-means. One-class SVM is used to provide a tight boundary between clusters. The combination of k-means clustering and one-class SVM approach gives an overall accuracy of 95.1%, which is less than the proposed approach accuracy of 96.68%. Pixel-level ground truth annotation is not provided in the UMN dataset, but our proposed approach localizes abnormal motion patterns. This result shows the performance of our proposed framework is more accurate in all scenes. In these figures, even when multiple anomalous activities occur in the same frame sequence, different anomalies are detected.

Figure 8 shows some interesting qualitative results on the UMN dataset produced by the proposed framework. Normal motion patterns characterize people walking on the lawn and plaza. People running in all directions (panic behavior) are labeled as abnormal. Such panic behavior is localized in video sequences with anomalous motion patterns.

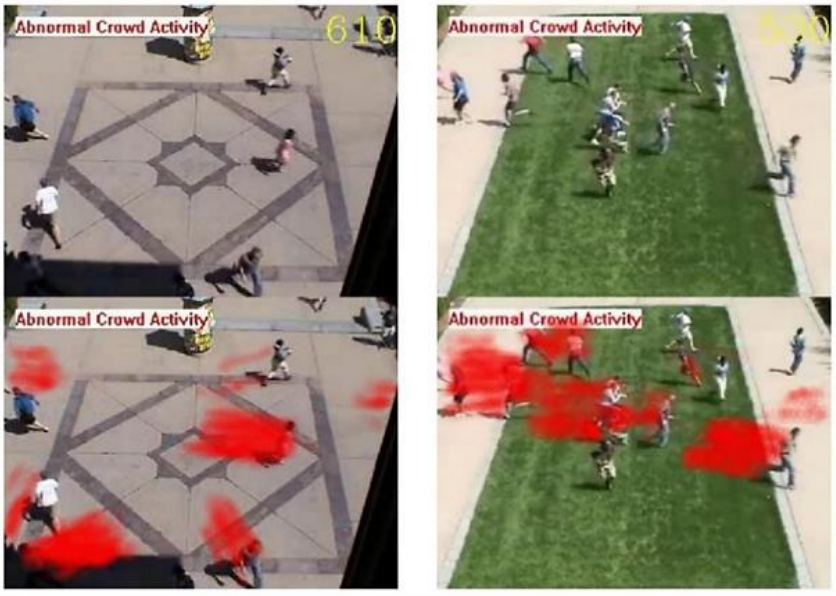

Figure 8 Crowd escape behavior detection and localization for different scenarios in the UMN dataset.

6. Conclusions

This paper presents a novel approach for detecting abnormal human behavior in surveillance video sequences. Motion speed obtained from dense trajectories of moving objects was grouped into intervals, and homography estimation was applied to video sequences that suffered from perspective distortion. Then, hierarchical clustering was used to cluster the intervals to detect abnormalities in surveillance video sequences. It has achieved the highest accuracy of 96.68% on the UMN dataset compared to existing methods. The proposed framework has achieved a frame-level rate of detection of 94.03% and 98.63% for the UCSD Peds1 and Peds2 datasets, respectively. For the pixel-level rate of detection, UCSD Peds1 achieved 94.43% and 90.86% for UCSD Peds2. The experimental evaluation shows that our framework outperforms existing state-of-the-art methods. In future work, the proposed framework can be extended to other real-world surveillance videos of road traffic, airport, indoor and outdoor scenes by considering dominant motion depending on the application context.

7. References

[1] Al-Dhamari A, Sudirman R, Mahmood NH. Abnormal behavior detection in automated surveillance videos: A review. J Theor Appl Inform Tech. 2017;95(19):5245-63.

[2] Brax C, Niklasson L, Smedberg M. Finding behavioural anomalies in public areas using video surveillance data. 2008 11th International conference on information fusion; 2008 Jun 30 - Jul 3; Cologne, Germany. New York: IEEE; 2008. p. 1-8.

[3] Mu C, Xie J, Yan W, Liu T, Li P. A fast recognition algorithm for suspicious behavior in high definition videos. Multimed Syst. 2016;22:275-85.

[4] Al-Dhamari A, Sudirman R, Mahmood NH, Khamis NH, Yahya A. Online video-based abnormal detection using highly motion techniques and statistical measures. Multimed Tool Appl. 2019;17:2039-47.

[5] Wu S, Wong HS, Yu Z. A Bayesian model for crowd escape behavior detection. IEEE Trans Circ Syst Video Tech. 2014;24: 85-98.

[6] Wu S, Moore BE, Shah M. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. 2010 IEEE computer society conference on computer vision and pattern recognition; 2010 Jun 13-18; San Francisco, USA. New York: IEEE; 2010. p. 2054-60.

[7] Weixin Li, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scenes. IEEE Trans Pattern Anal Mach Intell. 2014;36:18-32.

[8] Mahadevan V, Li W, Bhalodia V, Vasconcelos N. Anomaly detection in crowded scenes. 2010 IEEE computer society conference on computer vision and pattern recognition; 2010 Jun 13-18; San Francisco, USA. New York: IEEE; 2010. p. 1975-81.

[9] Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model. 2009 IEEE conference on computer vision and pattern recognition; 2009 Jun 20-25; Miami, USA. New York: IEEE; 2009. p. 935-42.

[10] Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection. 2011 IEEE Conference on computer vision and pattern recognition (CVPR); 2011 Jun 20-25; Colorado Springs, USA. New York: IEEE; 2011. p. 3449-56.

[11] Lin H, Deng JD, Woodford BJ, Shahi A. Online weighted clustering for real-time abnormal event detection in video surveillance. Proceedings of the 24th ACM international conference on multimedia; 2016 Oct 15-19; Amsterdam, Netherlands. New York: ACM Press; 2016. p. 536-40.

[12] Chen DY, Huang PC. Motion-based unusual event detection in human crowds. J Vis Comm Image Represent. 2011;22(2): 178-86.

[13] Fu Z, Hu W, Tan T. Similarity based vehicle trajectory clustering and anomaly detection. IEEE international conference on image processing; 2005 Sep 14; Genova, Italy. New York: IEEE; 2005. p. II-602.

[14] Basharat A, Gritai A, Shah M. Learning object motion patterns for anomaly detection and improved object detection. 2008 IEEE conference on computer vision and pattern recognition; 2008 Jun 23-28; Anchorage, USA. New York: IEEE; 2008. p. 1-8.

[15] Colque RM, Caetano C, de Andrade MT, Schwartz WR. Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos. IEEE Trans Circ Syst Video Tech. 2017;27:673-82.

[16] Li X, Li W, Liu B, Liu Q, Yu N. Object-oriented anomaly detection in surveillance videos. 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP); 2018 Apr 15-20; Calgary, Canada. New York: IEEE; 2018. p. 1907-11.

[17] Antic B, Ommer B. Video parsing for abnormality detection. 2011 International conference on computer vision; 2011 Nov 6-13; Barcelona, Spain. New York: IEEE; 2011. p. 2415-22.

[18] Pennisi A, Bloisi DD, Iocchi L. Online real-time crowd behavior detection in video sequences. Comput Vis Image Understand. 2016;144:166-76.

[19] Wang H, Klaser A, Schmid C, Liu CL. Action recognition by dense trajectories. CVPR 2011; 2011 Jun 20-25; Colorado Springs, USA. New York: IEEE; 2011. p. 3169-76.

[20] Abuolaim AA, Leow WK, Varadarajan J, Ahuja N. On the essence of unsupervised detection of anomalous event in surveillance videos. International conference on computer analysis of images and patterns; 2017 Aug 22-24; Ystad, Sweden. New York: Springer; 2017. p. 160-71.

[21] Pattnaik M. Abnormal event detection in pedestrian pathway using GARCH model and MLP classifier. Int J Signal Image Sci. 2019;5:15.

[22] Yu B, Liu Y, Sun Q. A content-adaptively sparse reconstruction method for abnormal events detection with low-rank property. IEEE Trans Syst Man Cybern Syst. 2017;47:704-16.

[23] Qasim T, Bhatti N. A low dimensional descriptor for detection of anomalies in crowd videos. Math Comput Simulat. 2019;166:245-52.

[24] Srinivasan A, Gnanavel VK. Multiple feature set with feature selection for anomaly search in videos using hybrid classification. Multimed Tool Appl. 2019;78:7713-25.

[25] Wang H, Schmid C. Action recognition with improved trajectories. IEEE international conference on computer vision; 2013 Dec 1-8; Sydney, Australia. New York: IEEE; 2013.

[26] Farneback G. Two-frame motion estimation based on polynomial expansion. 13th Scandinavian conference, SCIA; 2003 Jun 29 - Jul 2; Halmstad, Sweden. Berlin: Springer; 2003. p. 363-70.

[27] Rowland T. Velocity Vector [Internet]. Mathworld; 2019 [cited 2021 Feb 1]. Available from: http://mathworld.wolfram.com/VelocityVector.html.

[28] Dubrofsky E. Homography estimation. Vancouver: The University of British Columbia; 2009.

[29] LaRose D. A fast, affordable system for augmented reality. Pittsburgh: Carnegie Mellon University; 1998.

[30] Ongun C, Temizel A, Temizel TT. Local anomaly detection in crowded scenes using finite-time Lyapunov exponent based clustering. 2014 11th IEEE international conference on advanced video and signal based surveillance (AVSS); 2014 Aug 26-29; Seoul, Korea (South). New York: IEEE; 2014. p. 331-6.

[31] Ionescu RT, Smeureanu S, Alexe B, Popescu M. Unmasking the abnormal events in video. 2017 IEEE international conference on computer vision (ICCV); 2017 Oct 22-29; Venice, Italy. New York: IEEE; 2017.