Received: October 17, 2024. Revised: December 12, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Detecting Anomalous Crowd Behaviour with Optical Flow and Energy-Based Methods

Nay Htet Linn[1]* **Zin Mar Win[2]**

[1]*University of Computer Studies, Mandalay, Myanmar*
*Corresponding author's Email: nayhtetlin@ucsm.edu.mm

Abstract: In the domain of intelligent surveillance for public safety, rapid anomaly detection in crowded environments is essential. This study presents an approach to crowd behaviour analysis by measuring crowd energy changes. Image pixels are modeled as particles, and optical flow techniques are used to extract velocity vectors and directions. To mitigate the noise, occlusions, and lighting challenges of optical flow, the system incorporates pixel motion estimation across frames, improving temporal coherence for smoother motion. Image grey entropy and Otsu's segmentation are employed to separate foreground from background, enabling detailed energy distribution analysis. Abnormal crowd activity is detected by observing sudden changes in motion intensity. Evaluation on the UMN dataset shows that the proposed method achieves an accuracy of 96.87% in anomaly detection, outperforming other conventional techniques. These results highlight the improved accuracy and efficiency of the method in detecting anomalous crowd behaviour in complex environments.

Keywords: Abnormal detection, Crowd behaviour analysis, Optical flow, Co-occurrence matrix.

1. Introduction

Recent advances in computer vision have improved the detection, tracking, and interpretation of crowd behaviour in surveillance footage. This capability is vital for identifying panic and escape behaviours during riots, chaotic incidents, natural disasters, and violent events. A significant challenge in intelligent video surveillance is developing an autonomous system that can detect anomalies in complex, crowded scenes.

Analysing crowd behaviour involves two main approaches: methods inspired by physical principles and machine learning techniques. Machine learning methods process and analyse data to extract crowd features. For example, A. Al-Dhamari et al. [1] proposed abnormal behaviour detection using sparse representations through sequential generalization of k-means. In contrast, physically inspired methods treat crowds as complex systems [2]. There are two primary strategies for simulating crowd dynamics. The microscopic approach treats the crowd as a collection of individuals, tracking each person's movements to infer overall behaviour [3]. This approach is particularly useful for smaller groups but becomes challenging in dense crowds due to occlusions. The macroscopic approach, on the other hand, views the crowd as a single entity, interpreting each pixel in an image as a particle and modelling these particles' collective characteristics [4]. Various global analysis techniques have been developed based on this perspective. The primary contributions of the proposed system are as follows:

- The approach leverages optical flow, specifically incorporating pixel motion estimation between frames, to address limitations like sensitivity to environmental factors. By smoothing motion estimates and improving temporal coherence, the method achieves robust detection even in noisy or occluded environments.
- Using image grey entropy and Otsu's segmentation, the system effectively separates foreground from background, allowing for precise energy distribution analysis that is critical for tracking movement

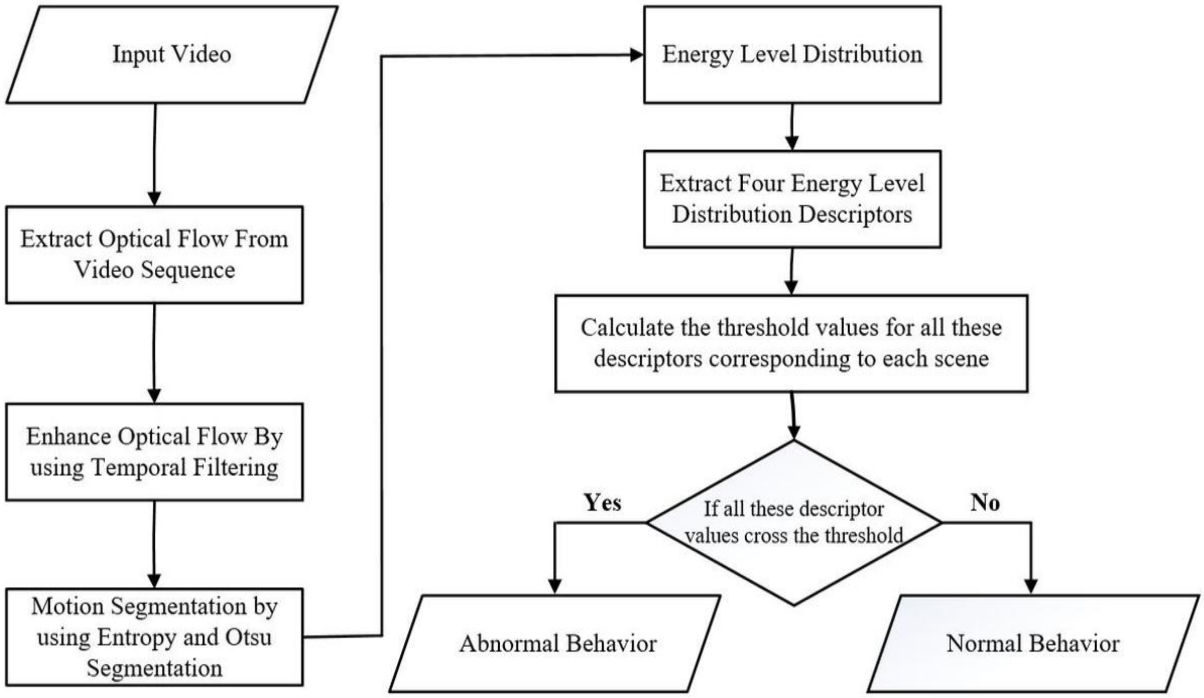

Figure. 1 The framework of anomaly crowd behavior detection system based on energy level distribution descriptors

across frames.

- For each segmented region, the system calculates energy level distribution, extracting four key descriptors—uniformity, entropy, contrast, and homogeneity—to characterize motion patterns within the scene. These descriptors offer a comprehensive basis for assessing normal versus abnormal activity.
- Thresholds for these descriptors are derived from observed normal behaviour, and any instance where all descriptors exceed their thresholds is flagged as abnormal, potentially indicating crowd disturbances or unusual behaviour.

The remainder of this paper is structured as follows: Section 2 describes related works, Section 3 discusses Optical Flow Extraction and Motion Segmentation, Section 4 covers Crowd Energy Distribution, Section 5 presents comparative experimental analysis, and finally, Section 6 includes the conclusion of the paper.

2. Related works

Abnormal crowd behaviour detection is vital for intelligent surveillance in crowded areas like airports and stadiums to prevent safety risks. Real-time detection is challenging due to factors like occlusion, environmental conditions, and complex crowd dynamics. Over the years, various methods, from traditional motion detection to advanced machine learning, have been developed, each with its strengths and limitations in real-world scenarios.

Aldhamari et al. [1] proposed an abnormal behaviour detection method using sparse representations with a sequential generalization of the k-means algorithm to capture unusual crowd dynamics. By employing sparse representation techniques, the model effectively reduces data complexity, enabling the identification of abnormal patterns across varying scenes. This approach demonstrates solid performance in detecting outlier behaviours while maintaining computational efficiency, making it suitable for moderate real-time applications. However, the reliance on k-means clustering limits the method's adaptability in highly complex or densely populated environments, where behaviour patterns are diverse and less distinct. Additionally, the sparse representation may struggle with scalability in large-scale settings, as it requires careful tuning to maintain detection accuracy amidst high variability in crowd behaviours.

Abdullah et al. [5] developed a multi-person tracking and crowd behaviour detection system utilizing a particles gradient motion descriptor combined with an improved entropy-based classifier. Their approach effectively captures crowd dynamics by leveraging motion gradients to track individuals and identify anomalous behaviours, yielding high accuracy across diverse crowd scenarios. This method demonstrated robust performance, particularly in small to mid-sized groups, where motion gradients and entropy measures efficiently distinguished normal and abnormal behaviours. However, the system’s reliance on particle motion descriptors limits its scalability in large, high-density crowds, where occlusions and complex interactions present tracking challenges. Overall, while the study advances multi-person tracking methods, its applicability to more extensive, densely populated settings remains a limitation for real-world crowd surveillance applications.

Luo et al. [6] introduced a crowd-level abnormal behaviour detection framework based on multi-scale motion consistency learning, aimed at capturing abnormal patterns across varying crowd densities. Their approach leverages multi-scale analysis to enhance detection accuracy by learning motion consistency across diverse regions, effectively identifying subtle anomalies within dense crowds. This technique demonstrated strong performance in both low- and high-density scenarios, showing adaptability to different crowd structures. However, the model’s multi-scale design increases computational load, which can limit its efficiency in real-time, large-scale surveillance applications.

Alafif et al. [7] leveraged generative adversarial networks (GANs) to address the complex task of detecting abnormal behaviors in massive crowds, using Hajj pilgrimage videos as a case study. Their approach combines optical flow with a GAN-based framework, enhancing the system’s capacity to distinguish subtle deviations in crowd behavior patterns. This method demonstrated high accuracy on benchmark datasets (UMN and UCSD) and achieved average detection accuracy on the HAJJ dataset, indicating its robustness in diverse, dense crowd settings. However, its performance in large-scale, occlusion-heavy environments like Hajj remains limited, with detection accuracy notably lower compared to simpler scenes due to challenges in managing distant camera views and extensive occlusions. This study contributes to the field by advancing GAN-based techniques for real-time surveillance but highlights ongoing challenges in scalability and accuracy under highly dynamic, large-scale crowd conditions.

Fan et al. [8] proposed a real-time abnormal behaviour detection system in videos, focusing on achieving both high accuracy and efficiency for practical applications. Their method utilizes a feature extraction approach that combines motion patterns and spatial information, enabling effective detection of anomalous activities in varied environments. The model demonstrated strong real-time performance and accuracy, making it suitable for dynamic monitoring scenarios. However, the reliance on handcrafted features limits the method's adaptability to highly complex or crowded scenes where behaviour patterns are unpredictable and vary significantly. While this study contributes valuable insights for responsive abnormal behaviour detection, its dependency on fixed feature designs constrains its scalability and robustness in more intricate video surveillance applications.

Alafif et al. [9] developed a hybrid classifier framework for real-time detection, tracking, and recognition of abnormal behaviours in large-scale Hajj crowds, integrating spatio-temporal features to address the complexities of dense crowd dynamics. This approach combines convolutional neural networks (CNNs) with traditional classifiers, enhancing detection accuracy by leveraging spatial and temporal information effectively. The method demonstrated robust performance on benchmark datasets and real-world Hajj data, showing its adaptability in high-density environments. However, the model's reliance on multiple classifiers and real-time processing requirements increases computational complexity, potentially limiting scalability in extensive surveillance systems.

Direkoglu [10] introduced an abnormal crowd behaviour detection method that combines motion information images with convolutional neural networks (CNNs) to identify anomalies. By converting motion data into image-like representations, this approach leverages CNNs’ spatial recognition capabilities to accurately detect unusual crowd behaviours, achieving competitive results across multiple crowd datasets. The model’s ability to extract meaningful features from motion information enhances detection accuracy and makes it adaptable to various crowd scenarios. However, the approach relies on hand-crafted motion features, which may not fully capture complex crowd dynamics. Moreover, the computational cost of CNNs poses challenges for real-time applications in large-scale surveillance.

Rajasekaran and Sekar [11] presented an abnormal crowd behaviour detection method leveraging an optimized Pyramidal Lucas-Kanade (PLK) technique to enhance motion tracking accuracy in dense crowds. By refining the PLK optical flow approach, the model improves detection precision, particularly in moderately crowded environments where individual motion patterns remain distinct. This optimization addresses some limitations of traditional optical flow methods, making the technique suitable for real-time applications. However, the approach is sensitive to noise and occlusions, which can reduce its accuracy in real-world environments. Additionally, the computational complexity of the pyramidal approach can limit its scalability for large-scale, real-time surveillance systems.

3. Optical flow extraction and motion segmentation

The optical flow method is utilized to derive the velocity field from a crowd video sequence. This study employs the Gunnar-Farneback optical flow algorithm, a popular technique for predicting dense optical flow fields in image sequences. The Farneback optical flow algorithm estimates the motion between consecutive frames in a video by analysing pixel displacements. It represents each frame as a quadratic polynomial, allowing the detection of motion in terms of pixel shifts. The algorithm computes pixel movement by comparing the intensity of neighbouring pixels across frames. Each frame is modeled as:

$$g(x) = xx^T B + xa^T + d \tag{1}$$

Where, x is the pixel position, B is a matrix representing the shape of the intensity variation, and a and d are coefficients. The motion is calculated as:

$$s = -\frac{1}{2} {B_1}^{-1}(a_2 - a_1) \tag{2}$$

Here, s is the displacement between consecutive frames, capturing pixel flow. This dense optical flow is then refined to enhance accuracy, making it useful for real-time applications like object tracking and motion detection.

In real-time applications like object tracking and motion analysis, the Farneback optical flow method is renowned for its speed and accuracy in estimating dense optical flow fields [12, 13]. To obtain more precise motion regions, the system applies temporal filtering after using the Farneback method. Fig. 2 shows the first two consecutive frames of the UMN dataset [14] and their corresponding optical flow.

3.1 Temporal filtering

In optical flow analysis, averaging serves as a straightforward yet powerful temporal filtering technique [15, 16] that mitigates noise and improves the accuracy of optical flow estimation over time. This method is particularly effective in reducing high-frequency noise within the optical flow field.

Figure. 2 Farneback Optical Flow of two consecutive frames

Accordingly, our system utilizes a moving average filter, which averages the optical flow vectors across a specified number of frames. Considering a sequence of N frames ($t_1, t_2, \ldots, t_N$), that is used to compute the filtered optical flow at time t_i. The filtered optical flow is denoted as follows:

$$Filter_f(x, y, t_i) = (\bar{u}, \bar{v}) \tag{3}$$

The temporal filter equation can be defined as:

$$\bar{u}(x, y, t_i) = \frac{1}{N} \sum_{j=1}^{N} u(x, y, t_i) \tag{4}$$

$$\bar{v}(x, y, t_i) = \frac{1}{N} \sum_{j=1}^{N} v(x, y, t_i) \tag{5}$$

Here, N denotes the number of frames used for filtering. The filtered flow vectors $(\bar{u}, \bar{v})$ at a specific pixel (x, y) and time t_i are obtained by averaging the flow vectors over the N frames. The choice of N depends on the video sequence characteristics and the desired balance between preserving motion details and achieving effective smoothing.

3.2 Motion region extraction

Motion segmentation is performed using flow field visualization with HSV colour mapping to distinguish moving targets by their velocity vectors, enhancing motion region visualization through colour assignment based on speed and direction. Motion region extraction involves evaluating uncertainty and determining the optimal threshold through pixel value distribution analysis. Our system integrates image gray entropy with the Otsu segmentation method for accurate foreground extraction [17, 18]. This approach ensures precise isolation of moving objects from the background, significantly improving motion segmentation

Received: October 17, 2024. Revised: December 12, 2024.

accuracy for applications in surveillance, object tracking, and video analysis.

3.2.1. Uncertainty estimation

Estimating uncertainty is crucial for understanding the confidence levels of a model's predictions, leading to more informed decision-making. In crowd motion analysis, the gray texture of the image highlights intensity distribution differences between motion and background regions. Entropy, a statistical measure of randomness, characterizes the texture of the input image. Higher entropy values denote more disordered intensity distributions.

The gray entropy [19], defined as:

$$E(y) = -\sum_{i=0}^{L-1} p(y_i) log_2 p(y_i) \tag{6}$$

is used to describe this texture information. L represents the total number of distinct gray levels, and $p\ (y_i)$ denotes the probability distribution.

In videos with moving crowds, motion regions have low entropy values (indicating structured patterns), while background regions have high entropy values (indicating greater disorder). This contrast in entropy values is critical for accurate segmentation and analysis of crowd movements. By utilizing entropy measurements, the system effectively distinguishes dynamic areas from static ones within video frames, enhancing detection accuracy and robustness in applications such as surveillance and video analysis.

3.2.2. Otsu segmentation

The Otsu method for image thresholding is a clustering-based approach designed to work effectively with bimodal histograms. It aims to minimize within-class variance while maximizing between-class variance.

At grey-level t, the image is divided into two classes: C0 and C1, or C0 = {0,1,2,……..,t} and C1 = {t+1,t+2,……,L-1}. L is the image's total number of gray levels.

The class probability estimates $w_0\ (t)$ and $w_1\ (t)$ are calculated using the subsequent formulas:

$$w_0(t) = \sum_{i=0}^{t-1} p(i) \tag{7}$$

$$w_1(t) = \sum_{i=t}^{L-1} p(i) \tag{8}$$

Where the probability that gray level i will occurs in the image is denoted by $p\ (i)$.

To optimize the threshold, the method focuses on maximizing the inter-class variance, which is equivalent to minimizing the intra-class variance. The between-class variance δ_b^2 (t) is defined as:

$$\delta_b{}^2(t) = \delta^2 - \delta_w{}^2(t) \tag{9}$$

$$= w_0\ (\mu_0 - \mu_T)^2 + w_1\ (\mu_1 - \mu_T)^2 \tag{10}$$

$$= w_0(t) w_1(t)\ [\mu_0(t) - \mu_1(t)]^2 \tag{11}$$

Which is expressed in terms of class probabilities w and class mean μ, where the class means $\mu_0\ (t)$, $\mu_1\ (t)$, and μ_T are:

$$\mu_0(t) = \frac{\sum_{i=0}^{t-1} i\, p(i)}{w_0(t)} \tag{12}$$

$$\mu_1(t) = \frac{\sum_{i=t}^{L-1} i\, p(i)}{w_1(t)} \tag{13}$$

$$\mu_T = \sum_{i=0}^{L-1} i\, p(i) \tag{14}$$

It is possible to compute the class probabilities and means iteratively. By maximizing the between-class variance$\delta_b^2\ (t)$, one can obtain the optimal threshold $T=ArgMax((\delta_b)^2\ (t))$. Threshold T allows for the segmentation of the motion and background zones. The input image $I_{in}\ (m,n)$ can be segmented from the output binary image $I_{out}\ (m,n)$ in the manner described below:

$$I_{out}(m,n) = \begin{cases} 1\ if\ I_{in}(m,n) \geq T \\ 0\ if\ I_{in}(m,n) < T \end{cases} \tag{15}$$

4. Crowd energy distribution

The quantitative assessment of crowd energy distribution reveals the energy levels of particles within the motion. This distribution is characterized using a co-occurrence matrix, which captures the spatial relationships between energy levels. The descriptors derived from this matrix are then utilized to accurately describe and analyze the crowd state, aiding in the detection of abnormal behaviours.

4.1 Particle energy model

The energy resulting from the movement of particles is constructed based on particle velocity. The energy of the mth frame optical flow with the coordinate of $(i,\ j)$ is defined as the following formula:

$$E(n) = \sum_{i=1}^{W} \sum_{j=1}^{H} \frac{1}{2} c_{i,j}(m) uv_{i,j}{}^2(m) \tag{16}$$

$E(n)$ is the energy of the optical flow in the motion region. W and H indicate the width and height of the motion area. The velocity of m^{th} frame image pixels $pix\ (i, j)$ is represented by the parameter $uv_{i,j}\ (m)$, and the coefficient $c_{i,j}\ (m)$ is obtained from the motion region of the current frame. The velocity of each pixel in the motion area is calculated by the following formula:

$$uv_{i,j}\ (\mathrm{m}) = \sqrt{\mathrm{u}_{i,j}{}^2 + \mathrm{v}_{i,j}{}^2} \tag{17}$$

4.2 Energy-level co-occurrence matrix

The Grey-Level Co-occurrence Matrix (GLCM), as outlined in [20], is a prevalent method for analyzing the distribution of grey values in an image. To compute GLCM, one must measure how often a pixel with intensity value i appears in a specific spatial relationship with a pixel of value j. Each element (i, j) in the GLCM represents the cumulative frequency of pixels with value i in relation to pixels with value j within the image. The size of matrix is based on the number of grey levels in the image, and typically, GLCM scales image intensity values down to eight levels.

This method generates a GLCM with a specific spatial relationship, using two horizontally adjacent pixels. However, a single GLCM cannot capture all textural characteristics, such as vertical textures. Therefore, multiple GLCMs with various offsets are needed to account for pixel relationships in different directions and distances. By defining an array of offsets in four directions (two diagonals, horizontal, and vertical) and four distances, the system produces 16 GLCMs per image, describing the distribution of energy levels and forming the energy-level co-occurrence matrix.

Consider an image of f with N potential energy levels, and let Q define the relative position of two pixels. The matrix G contains elements g_{ij}, each representing the frequency of pixel pairs with energy levels l_i and l_j at the offset Q, where $1 \leq \mathrm{i}, \mathrm{j} \leq N$. This matrix G is the energy-level co-occurrence matrix. By examining G with an appropriate position operator, one can identify the distribution of energy levels. A set of useful descriptors for characterizing the contents of G are listed below:

Uniformity: A uniformity metric inside the interval [0, 1]. For a constant energy-level, uniformity is 1.

$$Uniformity = \sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij}{}^2 \tag{18}$$

Entropy: Calculates how random each element of G is.

$$Entropy = -\sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij}\ log_2 p_{ij} \tag{19}$$

Contrast: The energy-level contrast between a particle and its neighbour through the whole image.

$$Contrast = \sum_{i=1}^{K} \sum_{j=1}^{K} (i-j)^2\, p_{ij} \tag{20}$$

Homogeneity: measures the similarity of intensity values in a local neighbourhood within an image. It quantifies how uniform or consistent the pixel intensities are within a region.

$$Homoteneity = \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{p_{ij}}{1+|i-j|} \tag{21}$$

Where, K is the row or column of the square matrix G, and p_{ij} is the estimation of the probability for a pair of points satisfying Q, which will have values $(\mathrm{l}_i, \mathrm{l}_j)$. It is defined as follows:

$$p_{ij} = \frac{g_{ij}}{num} \tag{22}$$

Where, num is the sum of the elements of G. The sum of these probability is one, and they fall between 0 and 1. The system uses four position operators with a distance of 1 and angles of 0, 45, 90, and 135 degrees to generate four energy-level co-occurrence matrices and calculate four descriptors for each image.

5. Experiment and discussion

This section presents the experimental results on abnormal crowd behaviour using the University of Minnesota (UMN) benchmark dataset [14]. The UMN dataset includes three scenes: a play court (scene 1), a museum (scene 2), and a ground (scene 3). The play court scene contains 2 videos, the museum contains 6 videos, and the ground contains 3 videos, totalling 11 videos. The video properties of the UMN dataset are 320x240 resolution, 30 frames per second, and 24 bits per pixel. Each sequence consists of varying train and test frames. In the UMN dataset, the sudden running of people is identified as abnormal behaviour as shown in Fig. 3.

Every video begins with scenes of people walking leisurely and ends with people rushing in a panicked state, providing a range of unusual test images. The proposed method can be applied to disaster prevention and safety monitoring. These applications guide the method's performance evaluation, emphasizing its ability to detect abnormal behaviour early.

Figure. 3 Normal and Abnormal frames of each scene of the UMN dataset

5.1 Threshold computation

The classification of a crowd as either normal or abnormal is based on comparing four specific parameters to their predetermined thresholds. Consequently, it is crucial to accurately estimate these thresholds. To achieve this, the first 300 frames from the initial video of each scene are used for parameter training. Subsequently, the values for these four parameters are computed for each frame in the video sequences. Using these computed values, the thresholds for each parameter in various scenes are determined according to the following formula.

$$[Threshold]_{P_s} = arg \max_{i=1\ldots300} [feature_n]_i + arg \min_{i=1\ldots300} \left[\frac{1}{(2\pi)^2} \sum_{j=0}^{\infty} \frac{(-1)^j (feature_n)^{2j+1}}{j!(2j+1)}\right]_i \quad (23)$$

Where P_s represent the sample video for various scenes (s=1, 2, 3), i denote the frame number in the video P_s $(i = 1,2,\ldots,300)$, and $feature_n$ be the value of the nth parameter (n=1, 2, 3, 4). To accurately estimate the thresholds, the system incorporates small Gaussian errors as a margin based on the maximum parameter value. Table 1 presents the threshold values for four parameters: uniformity, entropy, contrast, and homogeneity, across three different scenes from the UMN dataset.

5.2 Experimental results

The proposed system employs four feature descriptors derived from the energy-level co-occurrence matrix to assess crowd behaviour. To evaluate these descriptors, various videos from the UMN dataset were analyzed. The findings indicate that these descriptors effectively differentiate between normal and abnormal crowd behaviour. To minimize noise, the system triggers an alarm only if the value exceeds its threshold for 10 consecutive frames. The system successfully detected these anomalies in real time.

Fig. 4 illustrates the results of anomalous activity detection for an outdoor scene (Scene 1) from the UMN dataset. Initially, individuals are seen walking freely on the grass, but they suddenly start running and dispersing in various directions, signaling abnormal crowd behavior. The graphs of four feature descriptors—uniformity, entropy, contrast, and homogeneity—demonstrate their effectiveness in identifying this transition from normal to abnormal crowd activity. The feature descriptor graphs cross their respective thresholds when the abnormal behavior occurs, highlighting the variations that emerge as the crowd dynamics change and clearly delineating periods of normal and abnormal behavior.

In the detection of unusual behaviour in indoor Scene 2 from the UMN dataset, an anomaly is first indicated at the 532nd frame when the uniformity descriptor drops below its threshold for over 10 consecutive frames, triggering an alert. The entropy descriptor also exceeds its threshold at this point, reinforcing the abnormal behaviour detection. However, the contrast descriptor does not consistently breach its threshold until the 534th frame. Consequently, the system officially confirms the abnormal state starting from the 534th frame, based on uniformity, entropy, and homogeneity consistently exceeding their thresholds.

In the anomalous behaviour detection for outdoor Scene 3 from the UMN dataset, an alarm is triggered at the 561st frame when the uniformity descriptor drops below its threshold for 10 consecutive frames.

Table 1. Threshold values of Descriptors for the UMN dataset

Descriptors	**UMN Scene 1**	**UMN Scene 2**	**UMN Scene 3**
Uniformity	0.88499499	0.859084129	0.866036735
Entropy	0.770546159	0.912291634	0.874864375
Contrast	0.625572826	0.841601351	0.836982614
Homogeneity	0.970073083	0.963230224	0.966859871

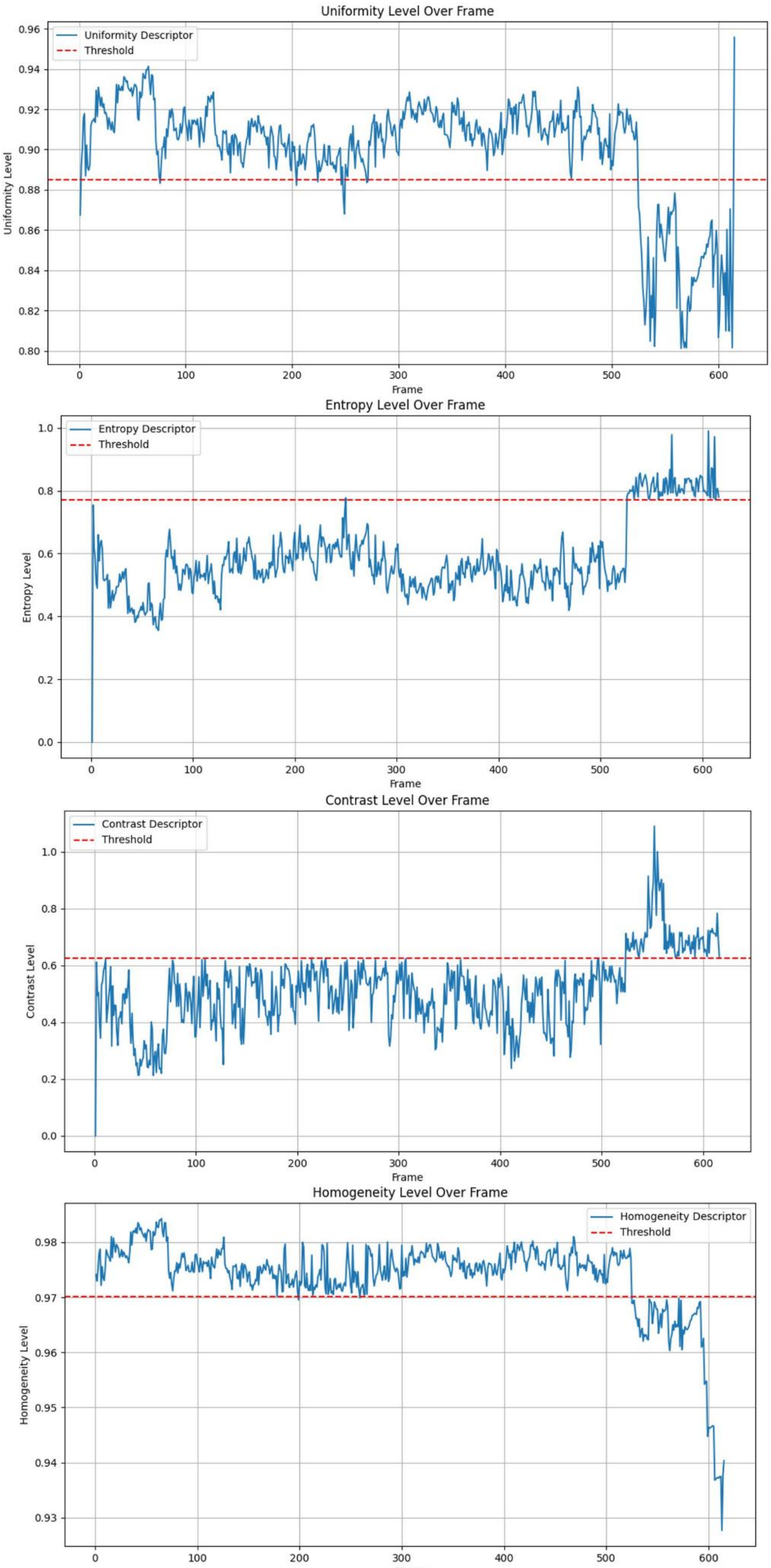

Figure. 4 Statuses of abnormal behaviour detection in the outdoor scene 1 from the UMN dataset

Received: October 17, 2024. Revised: December 12, 2024.

Table 2. Comparison of experimental results for each scene in the UMN dataset

Methods		Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
FGOFE + SGK [1]	Scene 1	93.910	N/A	N/A	N/A
	Scene 2	95.330	N/A	N/A	N/A
	Scene 3	93.667	N/A	N/A	N/A
	Average	94.302			
PGM + SURF + Improved Entropy [5]	Scene 1	87.430	N/A	N/A	N/A
	Scene 2	83.210	N/A	N/A	N/A
	Scene 3	90.630	N/A	N/A	N/A
	Average	86.060			
CNN + RFs [9]	Scene 1	88.850	99.340	87.350	92.960
	Scene 2	81.070	99.060	76.230	86.160
	Scene 3	93.330	99.400	93.320	96.260
	Average	87.750	99.267	85.633	91.793
OPLKTs [11]	Scene 1	92.500	97.414	90.400	93.776
	Scene 2	87.500	94.643	84.800	89.452
	Scene 3	95.000	95.276	96.800	96.032
	Average	91.667	95.778	90.667	93.087
Proposed Method	Scene 1	96.750	97.510	92.380	93.920
	Scene 2	96.050	94.110	92.220	94.630
	Scene 3	98.600	97.700	95.960	96.330
	Average	96.870	96.440	93.520	94.960

Although the uniformity dips at earlier frames, they do not sustain a consistent breach. The entropy descriptor surpasses its threshold at the 562nd frame, while the contrast descriptor consistently exceeds its threshold starting from the 558th frame. The homogeneity descriptor signals an anomaly at the 564th frame due to sustained values below the threshold. In summary, the system confirms an abnormal state when descriptors consistently breach thresholds over 10 consecutive frames, highlighting effective detection by uniformity, entropy, contrast, and homogeneity.

5.3 Comparison and analysis

In this section, the performance of the proposed method is compared with established approaches such as foreground optical flow energy (FGOFE) with sequential generalization of k-means (SGK) [1], particles gradient motion (PGM) with speeded up robust features (SURF) and improved entropy [5], CNN with random forests (RFs) [9], and optimized pyramidal Lucas-Kanade techniques (OPLKTs) [11] based on evaluation metrics such as Accuracy, Precision, Recall, and F-Measure.

Table 2 presents the comparative experimental results of different methods across all scenes of the UMN dataset, using metrics such as Accuracy, Precision, Recall, and F-Measure.

The proposed method outperforms all others, achieving the highest Accuracy (96.87%) and F-Measure (94.96%), indicating superior detection of anomalies. It also shows a strong balance between Precision (96.44%) and Recall (93.52%), reflecting its robustness in identifying abnormal crowd behaviour. Unlike CNN + RFs, which suffers from an imbalance between high precision and lower recall, the proposed method maintains a balanced and superior performance across all metrics. Furthermore, the robustness of the proposed approach is evident across different scenes, adapting well to complex crowd dynamics, whereas competing methods like FGOFE + SGK and PGM + SURF + Improved Entropy show significant performance drops. These results confirm that the proposed enhancements in motion estimation and feature extraction techniques effectively address the limitations of existing methods and provide a reliable solution for abnormal crowd behaviour detection. Figs. 5 and 6 show the comparison chart of the proposed method and the state-of-the-art methods.

In this work, enhanced pixel motion estimation through average temporal filtering significantly improved the performance of optical flow for abnormal crowd behaviour detection. Without filtering, pure optical flow yielded lower performance metrics, with accuracy dropping to 88.17%, precision to 86.26%, recall to 74.36%, and F-Measure to 79.89%. In contrast, the proposed enhancement resulted in substantially higher values across these metrics. This improvement highlights the importance of temporal filtering in achieving more accurate and reliable motion estimation.

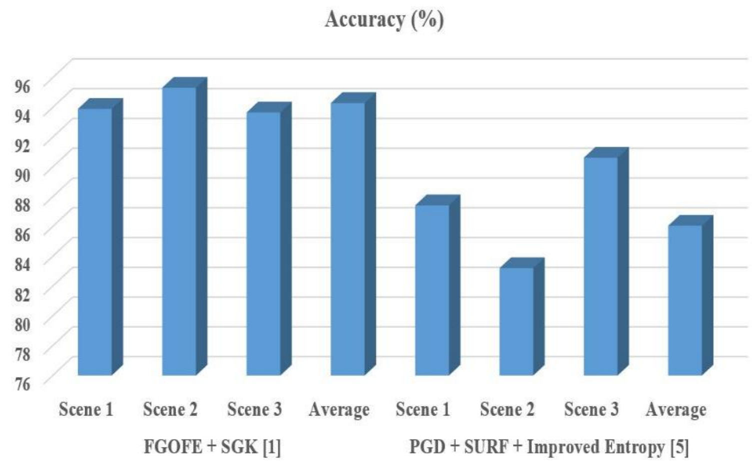

Figure. 5 Accuracy Comparison Chart

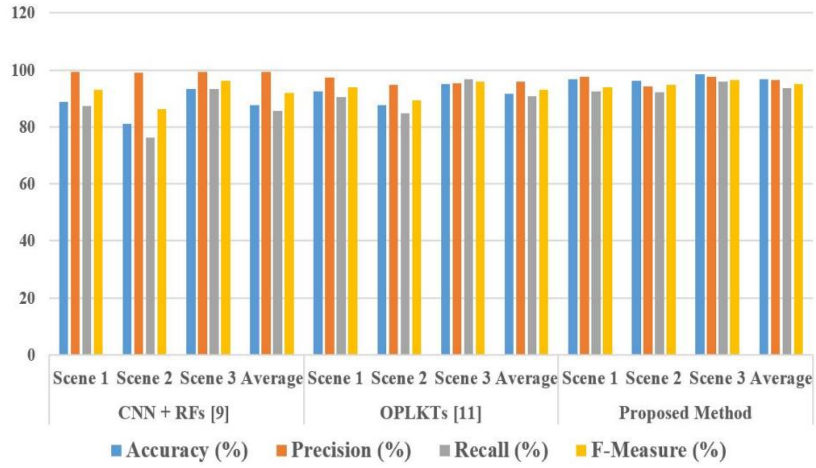

Figure. 6 Performance Comparison Chart

Table 3. AUC (%) comparison results on the UMN dataset

Methods		AUC (%)
FGOFE + SGK [1]	Scene 1	94.49
	Scene 2	92.11
	Scene 3	94.29
	Average	93.63
MSMC-Net [6]	Scene 1	N/A
	Scene 2	N/A
	Scene 3	N/A
	Average	94.4 ± 0.5
OF + GAN + U-Net & Flownet [7]	Scene 1	N/A
	Scene 2	N/A
	Scene 3	N/A
	Average	98.1
CNN + RFs [9]	Scene 1	97
	Scene 2	94.45
	Scene 3	97.38
	Average	96.28
Proposed Method	Scene 1	98.52
	Scene 2	97.68
	Scene 3	99.89
	Average	98.70

Table 3 presents the AUC (%) comparison results on each scene of the UMN dataset, highlighting the performance of the proposed method against existing approaches, including FGOFE with SGK [1], Multi-Scale Motion Consistency Network (MSMC-Net) [6], Optical Flow with Generative Adversarial Network (OF + GAN) [7], and CNN + RFs [9].

5.4 Discussion

The results of this study demonstrate that the proposed method for detecting abnormal crowd behaviour offers significant improvements in both accuracy and efficiency compared to existing techniques. By enhancing optical flow analysis and incorporating energy distribution methods, the system effectively captures motion dynamics, leading to more precise anomaly detection. The evaluation, conducted on the UMN dataset, shows that our approach not only outperforms traditional methods but also achieves a balance between computational complexity and detection accuracy.

Received: October 17, 2024. Revised: December 12, 2024.

6. Conclusion

In this paper, an unconventional method for detecting abnormal crowd behaviour using optical flow and energy-based techniques has been proposed and evaluated. The experimental results demonstrate that the proposed approach significantly improves accuracy, precision, recall, and F-Measure across multiple scenes of the UMN dataset, outperforming existing methods such as FGOFE with SGK, PGM + SURF with Improved Entropy, CNN with RFs, and OPLKTs. Specifically, the method achieved an overall accuracy of 96.87%, showing notable performance in both static and dynamic crowd scenarios.

The key contributions of this work include enhanced pixel motion estimation to address the limitations of optical flow, as well as the use of image grey entropy for effective background and foreground separation. These innovations have proven to be highly effective in improving anomaly detection in complex and crowded environments. Despite its success, future research could focus on further refining the method for more diverse and highly dynamic scenes, as well as reducing computational complexity to facilitate real-time implementation in large-scale surveillance systems. The results of this study provide a promising step towards more accurate and efficient crowd behaviour analysis in public safety applications.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization have been done by 1st author. The supervision and project administration have been done by 2nd author.

References

[1] A. Al-Dhamari, R. Sudirman, and N. H. Mahmood, “Abnormal behavior detection using sparse representations through sequential generalization of k-means”, *Turkish Journal of Electrical Engineering & Computer Sciences*, Vol. 29, No. 1, pp. 152-168, 2021.

[2] X. Zhang, X. Shu, and Z. He, “Crowd panic state detection using entropy of the distribution of enthalpy”, *Physica A: Statistical Mechanics and its Applications*, Vol. 525, pp. 935-945, 2019.

[3] M. Zahid, M. A. Khan, F. Azam, M. Sharif, S. Kadry, and J. R. Mohanty, “Pedestrian identification using motion-controlled deep neural network in real-time visual surveillance”, *Soft Computing*, Vol. 27, No. 1, pp. 453-469, 2023.

[4] B. Arbab-Zavar and Z. A. Sabeur, “Multi-scale crowd feature detection using vision sensing and statistical mechanics principles”, *Machine Vision and Applications*, Vol. 31, No. 4, p. 26, 2020.

[5] F. Abdullah, Y. Y. Ghadi, M. Gochoo, A. Jalal, and K. Kim, “Multi-person tracking and crowd behavior detection via particles gradient motion descriptor and improved entropy classifier”, *Entropy*, Vol. 23, No. 5, p. 628, 2021.

[6] L. Luo, Y. Li, H. Yin, S. Xie, R. Hu, and W. Cai, “Crowd-level abnormal behavior detection via multi-scale motion consistency learning”, In: *Proc. of the AAAI Conference on Artificial Intelligence*, Vol. 37, No. 7, pp. 8984-8992, 2023.

[7] T. Alafif, B. Alzahrani, Y. Cao, R. Alotaibi, A. Barnawi, and M. Chen, “Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study”, *Journal of Ambient Intelligence and Humanized Computing*, Vol. 13, No. 8, pp. 4077-4088, 2022.

[8] Z. Fan, J. Yin, Y. Song, and Z. Liu, “Real-time and accurate abnormal behavior detection in videos”, *Machine Vision and Applications*, Vol. 31, pp. 1-13, 2020.

[9] T. Alafif, A. Hadi, M. Allahyani, B. Alzahrani, A. Alhothali, R. Alotaibi, and A. Barnawi, “Hybrid classifiers for spatio-temporal real-time abnormal behaviors detection, tracking, and recognition in massive hajj crowds”, *Electronics*, Vol. 12, No. 5, p. 1165, 2023.

[10] C. Direkoglu, “Abnormal crowd behavior detection using motion information images and convolutional neural networks”, *IEEE Access*, Vol. 8, pp. 80408-80416, 2020.

[11] G. Rajasekaran and J. Raja Sekar, “Abnormal Crowd Behavior Detection Using Optimized Pyramidal Lucas-Kanade Technique”, *Intelligent Automation & Soft Computing*, Vol. 35, No. 2, pp. 2399-2412, 2023.

[12] N. Mukai, N. Nishimura, and Y. Chang, “Speed estimation using deep learning with optical flow”, In: *Proc. of International Workshop on Advanced Imaging Technology (IWAIT) 2024*, Langkawi, Malaysia: SPIE, pp. 6, May 2024.

[13] A. S. Lele and A. Raychowdhury, "Fusing frame and event vision for high-speed optical flow for edge application", In: *Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 804-808, 2022.

[14] "Monitoring Human Activity - Detection of Events", *Accessed: 14*, 2024. [Online]. Available: https://mha.cs.umn.edu/proj_events.shtml#crowd

[15] E. Xu, A. Wu, J. Li, H. Chen, X. Fan, and Q. Huang, "Infrared target detection based on joint spatio-temporal filtering and L1 norm regularization", *Sensors*, Vol. 22, No. 16, p. 6258, 2022.

[16] Y. Sun, J. Yang, and W. An, "Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 5, pp. 3737-3752, 2020.

[17] P. D. Sathya, R. Kalyani, and V. P. Sakthivel, "Color image segmentation using Kapur, Otsu and minimum cross entropy functions based on exchange market algorithm", *Expert Systems with Applications*, Vol. 172, p. 114636, 2021.

[18] J. Zheng, Y. Gao, H. Zhang, Y. Lei, and J. Zhang, "OTSU multi-threshold image segmentation based on improved particle swarm algorithm", *Applied Sciences*, Vol. 12, No. 22, p. 11514, 2022.

[19] A. Howedi, A. Lotfi, and A. Pourabdollah, "An Entropy-Based Approach for Anomaly Detection in Activities of Daily Living in the Presence of a Visitor", *Entropy*, Vol. 22, No. 8, p. 845, 2020.

[20] Z. Indra and Y. Jusman, "Performance of GLCM Algorithm for Extracting Features to Differentiate Normal and Abnormal Brain Images", *IOP Conference Series: Materials Science and Engineering*, Vol. 1082, No. 1, p. 012011, 2021.