entropy

MDPI

Article

Multi-Person Tracking and Crowd Behavior Detection via Particles Gradient Motion Descriptor and Improved Entropy Classifier

Faisal Abdullah [1], Yazeed Yasin Ghadi [2], Munkhjargal Gochoo [3], Ahmad Jalal [1] and Kibum Kim [4,*]

[1] Department of Computer Science, Air University, Islamabad 44000, Pakistan; 191633@students.au.edu.pk (F.A.); ahmadjalal@mail.au.edu.pk (A.J.)
[2] Department of Computer Science and Software Engineering, Al Ain University, Abu Dhabi 122612, United Arab Emirates; Yazeed.ghadi@aau.ac.ae
[3] Department of Computer Science and Software Engineering, United Arab Emirates University, Al Ain 15551, United Arab Emirates; mgochoo@uaeu.ac.ae
[4] Department of Human-Computer Interaction, Hanyang University, Ansan 15588, Korea
* Correspondence: kibum@hanyang.ac.kr

Abstract: To prevent disasters and to control and supervise crowds, automated video surveillance has become indispensable. In today's complex and crowded environments, manual surveillance and monitoring systems are inefficient, labor intensive, and unwieldy. Automated video surveillance systems offer promising solutions, but challenges remain. One of the major challenges is the extraction of true foregrounds of pixels representing humans only. Furthermore, to accurately understand and interpret crowd behavior, human crowd behavior (HCB) systems require robust feature extraction methods, along with powerful and reliable decision-making classifiers. In this paper, we describe our approach to these issues by presenting a novel Particles Force Model for multi-person tracking, a vigorous fusion of global and local descriptors, along with a robust improved entropy classifier for detecting and interpreting crowd behavior. In the proposed model, necessary preprocessing steps are followed by the application of a first distance algorithm for the removal of background clutter; true-foreground elements are then extracted via a Particles Force Model. The detected human forms are then counted by labeling and performing cluster estimation, using a K-nearest neighbors search algorithm. After that, the location of all the human silhouettes is fixed and, using the Jaccard similarity index and normalized cross-correlation as a cost function, multi-person tracking is performed. For HCB detection, we introduced human crowd contour extraction as a global feature and a particles gradient motion (PGD) descriptor, along with geometrical and speeded up robust features (SURF) for local features. After features were extracted, we applied bat optimization for optimal features, which also works as a pre-classifier. Finally, we introduced a robust improved entropy classifier for decision making and automated crowd behavior detection in smart surveillance systems. We evaluated the performance of our proposed system on a publicly available benchmark PETS2009 and UMN dataset. Experimental results show that our system performed better compared to existing well-known state-of-the-art methods by achieving higher accuracy rates. The proposed system can be deployed to great benefit in numerous public places, such as airports, shopping malls, city centers, and train stations to control, supervise, and protect crowds.

Keywords: bat optimization; human crowd behavior (HCB); improved entropy (IE); Jaccard similarity; multi-person counting; particles gradient motion (PGM); speeded up robust features (SURF)

Citation: Abdullah, F.; Ghadi, Y.Y.; Gochoo, M.; Jalal, A.; Kim, K. Multi-Person Tracking and Crowd Behavior Detection via Particles Gradient Motion Descriptor and Improved Entropy Classifier. *Entropy* **2021**, *23*, 628. https://doi.org/10.3390/e23050628

Academic Editor: Amelia Carolina Sparavigna

Received: 1 April 2021
Accepted: 12 May 2021
Published: 18 May 2021

1. Introduction

Multi-person tracking is currently one of the most essential and challenging research topics in the computer vision community [1–9]. Because of the common availability of high-quality low-cost video cameras and considering the inefficiency of manual surveillance and monitoring systems, automated video surveillance is now essential for today's crowded and complex environments. To monitor, control, and protect crowds, accurate information about numbers plays a vital role in operational and security efficiencies [10–16]. The counting and tracking of many persons is a challenging problem [17–25] due to occlusions, the constant displacement of people, different perspectives and behaviors, varying illumination levels, and because, as the crowd gets bigger, the allocation of pixels per person decreases.

A primary concern in surveillance and monitoring systems is to identify human crowd behaviors and supervise the crowd to prevent disasters and unforeseen events [26–34]. The analysis of human behavior in crowded scenes is one of the most important and challenging areas in current research [35–43]. Traditional visual surveillance systems that depend purely on manpower to analyze videos is inefficient because of the enormous number of cameras and screens that require monitoring, human fatigue due to time spent on lengthy monitoring periods, paucity of essential fore-knowledge and training in what to look for, and also because of the colossal amount of video data that is generated per day. Such issues necessitate an automated visual surveillance system that can reliably detect, isolate, analyze, identify, and alert responders to unusual events in real time. Automated surveillance systems seek to detect human behaviors automatically in crowded scenes, and it has many potential applications, such as security, care of the elderly and infirm, traffic monitoring, inspection tasks, military applications, robotic vision, sports analysis, video surveillance, and pedestrian traffic monitoring [44–52].

In this research article, we propose a robust new particles-based approach for multi-person counting and tracking, which addresses the problematic fact that, as the density of a crowd increases, the number of pixels allocated per human decreases. By using our particles-based approach, we were able to count and track multiple persons in crowded scenes and efficiently deal with occlusions, arbitrary movements, and overlaps. We also propose a new approach for crowd behavior detection using an improved entropy classifier based on the fusion of global and local descriptors extraction. First of all, we applied pre-processing steps on extracted video frames for noise removal, edge detection, and contrast adjustment, then human/non-human detection was performed using multi-level thresholding and morphological operations. We applied a distance algorithm for human silhouette extraction. After that, our work involved two facets: (i) multi-people tracking and (ii) crowd behavior detection. In the multi-person tracking phase, we first verified the extracted silhouettes by a particles force model, then we converted extracted foreground objects into particles, and, using physics phenomena of the mutually interacting particles force model, non-human objects were discarded. As every extracted human silhouette is a collection of particles, by treating groups of particles that make one silhouette as a cluster, we performed labeling and cluster estimation using a K-nearest neighbors search algorithm to count the persons. We then fixed the human silhouettes with a unique integer ID, and, using normalized cross correlation as a cost function and the Jaccard similarity index, multi-person tracking was performed. However, for crowd behavior detection, we used a fusion of global and local descriptors, that is, after foreground extraction, we extracted a human crowd contour as a global descriptor and a particles gradient motion (PGM) descriptor, along with geometric and speeded up robust features (SURF) as local descriptors. Using this fusion of global and local descriptors, bat optimization was then applied for optimal descriptors. Finally, by using Shannon's information entropy theory [53], we introduced an improved entropy classifier to detect crowd behavior.

Experimental results show that our proposed system performed better compared to existing well-known state-of-the-art methods. The proposed system has huge potential applications, such as crowd density estimation, security, care of the elderly and vulnerable, sports analysis, inspection tasks, military applications, robotic vision, video surveillance, and pedestrian traffic monitoring. The major contributions of this paper can be highlighted as follows:

1. We propose a new particles force model for human silhouettes verification, which is a necessary step for accurate counting and tracking of multiple persons in crowded scenes.
2. We developed a novel particles gradient motion local descriptor and human crowd contour as a global descriptor, while the fusion of global and local features was used for crowd behavior detection.
3. We designed an improved entropy classifier to analyze contextual information and classify crowd behavior in a more efficient manner.
4. We evaluated the performance of our proposed multi-person tracking approach on a publicly available benchmark PETS2009 dataset while crowd behavior detection performance was evaluated on the publicly available benchmark UMN dataset and the proposed method was fully validated for efficacy, surpassing other state-of-the-art methods, including deep learning.

The remaining structure of this paper was arranged as follows: Section 2 describes related work. A detailed overview of the proposed model for multi-person tracking and crowd behavior detection is mentioned in Section 3, which includes preprocessing, human silhouettes extraction, the particles force model, multi-person counting, multi-person tracking, global and local features extraction, bat optimization, and an improved entropy classifier. In Section 4, we evaluate the performance of our proposed approach on a publicly available benchmark dataset and give a detailed comparison of our proposed approach with other state-of-the-art methods. Lastly, in Section 5, we sum up the paper and outline future directions.

2. Related Work

During the last few years, several algorithms and systems have been developed by different researchers for crowd counting, tracking, and human behavior detection [54–62]. Here, we divide the related work into two parts, namely, human crowd behavior detection systems and multi-person counting and tracking systems.

2.1. Crowd Behavior Detection Systems

Many contributions have been proposed to describe crowd behavior using various models [63–69]. Crowd behavior detection is a challenging problem due to the arbitrary movements of individuals and groups, partial or full occlusions, different outlooks and behaviors, posture changes, and composite backgrounds [70–76]. To detect human behaviors automatically in crowded areas, S. Wu et al. in [77] constructed a density function of optical flow based on class-conditional probability and described the motion of crowds using divergent centers and potential destinations so that anomalies can be detected on the basis of a Bayesian framework. However, the system is not effective for arbitrary movements or overlaps. S. Choudhary et al. in [78] proposed a SIFT feature extraction technique, along with a Genetic Algorithm for optimal feature extraction; anomalies were detected by checking feature set movement behaviors. Their proposed system has a very high computational processing demand. Direkoglu et al. in [79] used a one-class SVM, along with features based on optical flow to detect crowd behavior; their system is limited by the accuracy limitations of optical flow estimation. W. G. Aguilar et al. in [80] introduced a moved-pixels density-based statistical modeling approach for detecting abnormal crowd behavior. This system has low computational cost, but the efficiency decreases with increasing complexity of the situation being monitored, e.g., serious occlusions. A. Shehzed et al. in [81] first detected humans and then the gaussian smoothing technique was used to detect anomalous behavior; however, the accuracy of the system decreases with illumination changes and occlusions because thresholding is used for detection. W. Ren et al. in [82] introduced a behavior entropy model for detecting abnormal crowd behavior using spatio-temporal information, along with behavior certainty of pixels, but the system is vulnerable to certain misclassifications due to interclass similarities. G. Wang et al. in [83] addressed the crowd behavior detection problem by using the pyramid Lucas-

Kanade optical flow [84] method based on location estimation of adjacent flow; however, the proposed method is not effective for an unstructured crowd. R Mehran et al. in [85] placed a grid of particles on the image and introduced a social force model for detecting crowd behavior. Bellomo, N. et al. in [86] pursued two specific objectives: the derivation of a general mathematical structure based on appropriate developments of the kinetic theory suitable for capturing the main features of crowd dynamics and the derivation of macroscopic equations from the underlying mesoscopic description. Colombo, R.M. et al. in [87] dealt with macroscopic modelling of crowd movements, particularly how non-local interactions are influenced by walls, obstacles, and exits. An ad hoc numerical algorithm, along with heuristic evaluation of its convergence, was also provided. Khan, S.D. et al. in [88] proposed scale estimation network SENet and head detection network. The SENet takes the input image and predicts the distribution of scales (in terms of histogram) of all heads in the input image, which are later on classified by a detection network.

2.2. Multi-Person Counting and Tracking Systems

True foreground extraction, i.e., human pixels, is only one of the primary steps for accurate counting and tracking of humans in crowded scenes [89–93]. Several approaches and systems have been introduced by many researchers for multi-person counting and tracking. In [94], S. Choudri et al. proposed a pixels-based people counting model using the fusion of a pixel map-based algorithm along with human detection to count only human classified pixels. They applied a depth map, image segmentation, and a human presence map that was updated with a human mask for the purpose of counting people; however, the system has misclassification problems due to interclass similarities. H. Chen et al. in [95] proposed a new color and intensity patch segmentation approach for tracking and detection of human body parts and for the full body. They applied fusion of color space segmentations for the detection of body parts and for the full body. For tracking, based on the velocity of a target, they adaptively selected the track gate size. A target's likely forward position was predicted based on the target's previous velocity and direction. The proposed algorithm achieved satisfactory results only when the count of peoples was limited in the view, i.e., efficiency decreases as the crowd increases. In [96], J. Garcia et al. introduced a head tracking-based directional people counter. Using several circular patterns and preprocessing steps, people's heads were detected. For the tracking application, a Kalman filter was used, and counting was achieved on the bases of head detection and tracking. The effectiveness of the proposed algorithm decreases during serious occlusions, arbitrary movements, and overlaps. M. Vinod et al. in [97] introduced object tracking and counting using new morphological techniques. The frame-difference technique, followed by morphological processing and region growing, was used for counting people. Moving objects were extracted by determining their movements, and then tracking was performed using color features. As the illumination of the scene changed, the efficiency of the proposed algorithm decreased. G. Liu et al. in [98] proposed a tracker based on a correlation filter. Kalman filter applications were used for tracking. They designed a tracker that detects numerous positions and alternate templates. However, the system was not efficient in dealing with complex situations, such as occlusions and random movements. E. Ristani et al. in [99] used deep learning to track multi-persons. Using CNN, they extracted features and then introduced a weighted triple loss strategy to assign weights during training. Their system was computationally complex, and a huge dataset was essential for training. H. Xu et al. in [100] located humans by their shoulders and heads, and, for tracking, they used trajectory analysis and the Kalman filter, but the system was not effective for arbitrary movements or overlaps.

3. Proposed System Methodology

This section elaborates our proposed methodology for multi-person tracking and crowd behavior detection. We propose a robust multi-person tracking system based on a particles force model and human crowd behavior detection system using an improved entropy classifier with spatio-temporal and particles gradient motion descriptors. In the proposed system, the first step is the preprocessing of extracted video frames from a static camera. Secondly, object detection is transacted using multi-level thresholding, morphological operations, and labeling. Thirdly, for human silhouette extraction, a distance algorithm is applied, and non-human filtering is performed on all extracted labeled objects. At this stage, we administered our work into two streams: the first was for multi-person counting and tracking, where we first performed a human silhouette verification step by converting extracted objects into particles and a robust particles force model was introduced for human silhouette verification. In the next step, after verification of human silhouettes, as all verified human silhouettes are a collection of particles, by treating each group of particles as a cluster we performed labeling and cluster estimation using a K-nearest neighbors searching algorithm for multi-person counting. After that, for multi-person tracking, the position of each detected human silhouette was then located and locked by assigning an integer ID for temporally fixing each human silhouette in the full video, and detected fixed humans were tracked using a Jaccard Similarity Index. However, in the second facet, for crowd behavior detection, the extracted foreground objects were passed through a feature extraction step and multiple distinguishable global and local features were extracted from every frame. After that, all the extracted features were standardized using the bat optimization algorithm. Lastly, in the classification phase, an improved entropy classifier was proposed for detection of crowd behavior. Figure 1 depicts the synoptic schematics of our proposed system.

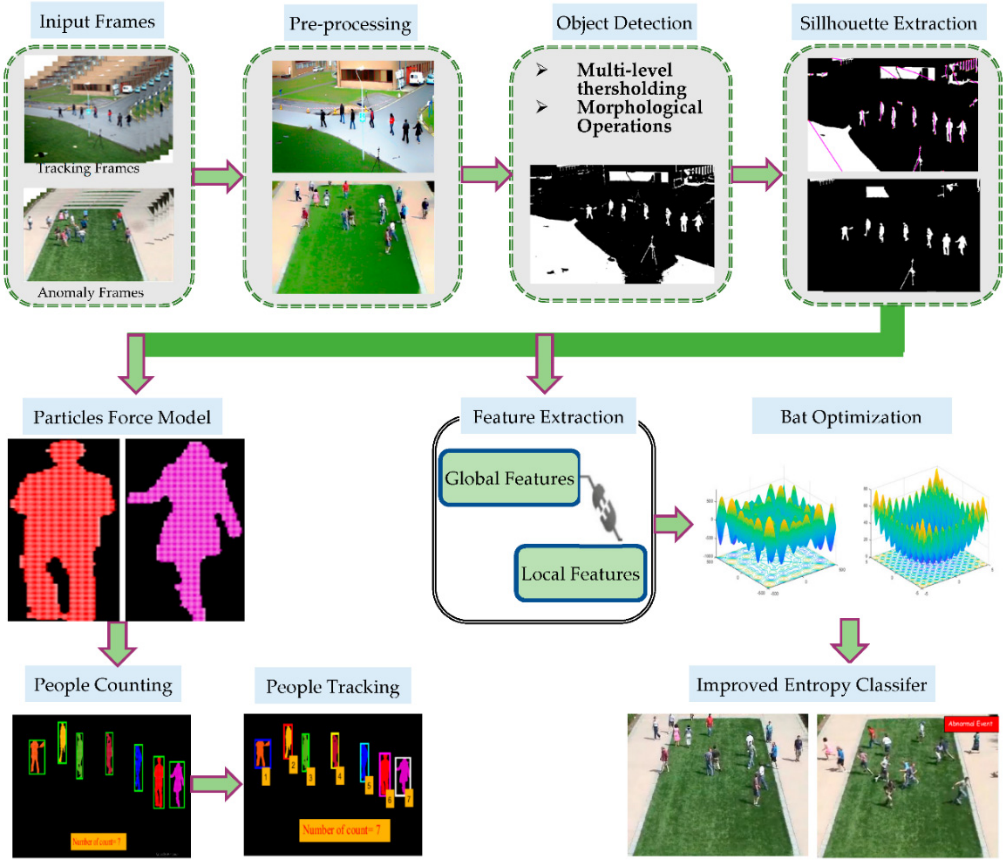


Figure 1. Synoptic schematics of the proposed Multi-Person Tracking and Crowd Behavior Detection system.

3.1. Pre-Processing

During image pre-processing, color frames were extracted from a static video camera $E = [f_1, f_2, f_3, \ldots, f_Z]$, where Z is the total number of frames. These color images were then passed through a Laplacian filter to reduce the noise and sharpen the edges. A Laplacian filter was applied using Equation (1):

$$\nabla^2 f = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y} \tag{1}$$

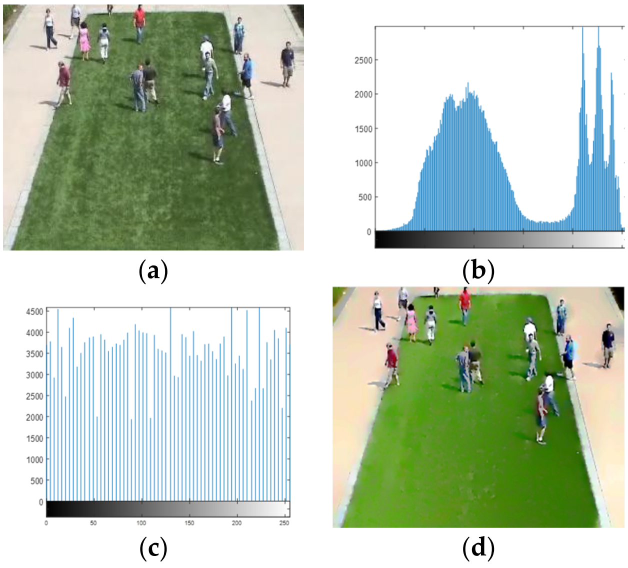
where $\nabla^2 f$ is the 2nd order derivative for obtaining the filtered mask. However, a pure Laplacian filter did not produce an enhanced image, thus, to achieve the sharpened enhanced image, we subtracted the Laplacian outcome from the original image using Equation (2):

$$g(x,y) = f(x,y) - \left[\nabla^2 f\right] \tag{2}$$

where the $g(x,y)$ is the sharpened image and $f(x,y)$ is the input image. After obtaining the sharpened image $g(x,y)$, histogram equalization was performed on the sharpened image in order to adjust the contrast of an image using Equation (3):

$$s_k = T(r_k) = (L-1)\sum\nolimits_{j=0}^{k} p_r(r_j) \quad k = 0,\ 1,\ 2,\ \ldots,\ L-1 \tag{3}$$

where variable r denotes the intensities of an input image to be processed. As usual, we assumed that r is in the range $[0\ L-1]$, with r = 0 representing black and $r = L - 1$ representing white, while s represents the output intensity level after intensity mapping for every pixel in the input image, having intensity r. However, $p_r(r)$ is the probability density function (PDF) of r, where the subscript on p were used to indicate that it was a PDF of r. Thus, a processed (output) image was achieved using Equation (3) by mapping each pixel in the input image with intensity r_k into a corresponding pixel with level s_k in the output image, as shown in Figure 2.



(**a**)

(**b**)

(**c**)

(**d**)

Figure 2. Preprocessing steps. (**a**) Original color frame of a video, (**b**) histogram of original image, (**c**) histogram of enhanced image, and (**d**) enhanced image.

3.2. *Human Silhouettes Extraction*

After obtaining the preprocessed frames, we performed human/non-human detection by performing multi-level thresholding using Equation (4), as depicted in Figure 3c.

$$th(x,y) = \begin{cases} 1 & \text{if} \quad I(x,y) > t_1, t_2, t_3 \\ 0 & \qquad\qquad \text{otherwise} \end{cases} \tag{4}$$

where $th(x, y)$ is the threshold image and t_1, t_2, t_3 are the applied thresholds that are defined by Otsu's procedure. In order to extract more useful information, the resultant binary image was inverted using a point processing operation that subtracts every pixel of an image from the maximum level of the image, as shown in Equation (5).

$$C(x, y) = 1 - th(x, y) \tag{5}$$

where $C(x, y)$ is the inverted image, as shown in Figure 3d, and $th(x, y)$ is the binary image with a maximum level of 1. After obtaining the human/non-human binary foreground frames, we performed morphological operations to remove imperfections in the inverted image C. For the removal of small unwanted objects, erosion was performed, and then, to fill small holes while preserving the size and shape of objects, morphological closing was performed. Every object in image C was first eroded using erosion as represented in Equation (6) and then dilated using Equation (7), after which the dilated image was eroded again using the disk-shaped structuring element, as shown in Equation (8).

$$m(x, y) = \begin{cases} 1 & \text{if} \quad S \text{ fits C} \\ 0 & \quad \text{otherwise} \end{cases} \tag{6}$$

$$m(x, y) = \begin{cases} 1 & \text{if} \quad S \text{ hits C} \\ 0 & \quad \text{otherwise} \end{cases} \tag{7}$$

$$Mo = (C \ominus S)((C \oplus S) \ominus S) \tag{8}$$

where C represents the input inverted image and S is the disk-shaped structuring element used for erosion and dilation, while Mo is the resultant image. The erosion of C by S is denoted as $(C \ominus S)$; however, the dilation of C by S is denoted as $(C \oplus S)$. After morphological operations, all the objects in the image were grouped and labeled, which helped in extracting and uniquely analyzing every object that was required for human silhouette extraction.

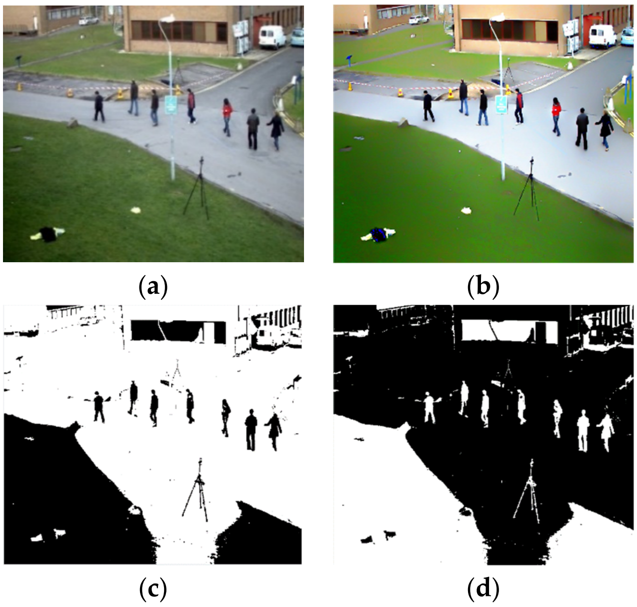
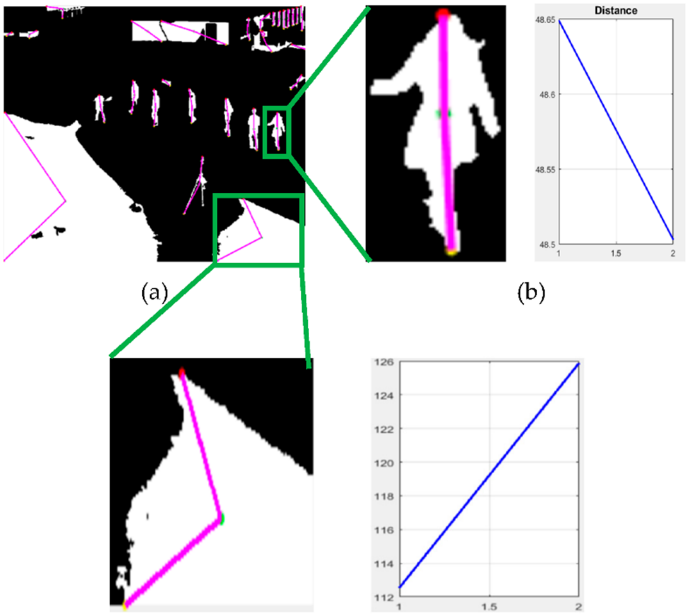


Figure 3. Object detection steps. (**a**) Original color frame of a video, (**b**) enhanced image, (**c**) binary image after multi-level thresholding, and (**d**) inverse of a threshold image.

After human/non-human detection, for human silhouette extraction, we calculated the center and extreme points of each of the labeled objects of M_o, then we extracted each object one by one, and the distance from center to two extreme points was calculated for every object for non-human filtering, as shown in Figure 4. The same procedure was adopted for the frames from frame 1 to frame Z.



(a) (b)

(c)

Figure 4. Human silhouette extraction. (**a**) Distance algorithm from the center to two extreme points for every object, (**b**) single silhouette extracted uniquely through labeling, along with its distance graph, and (**c**) a single non-silhouette, along with its distance graph.

After calculating the distances, those objects whose distances were greater than the set threshold were discarded using Equation (9), and only silhouettes resembling humans were retained.

$$E_h = \begin{cases} 0 & \text{if } d_1 > T \cap d_2 > \text{ T} \\ 1 & \text{otherwise} \end{cases} \tag{9}$$

where the distance from the center to one extreme point is denoted by d_1, the center to the other extreme point distance is represented by d_2, T is the set threshold and E_h is the resultant image. After human silhouette extraction, most of the non-human objects were discarded by the distance algorithm; however, some non-human objects that resembled human objects remained.

3.3. *Multi-Person Tracking*

For accurate human tracking, the extraction of the true foreground, i.e., human pixels only, is a primary step. Thus, after application of the distance algorithm (mentioned in Section 3.2) for multi-person tracking, we performed the human silhouette verification step using the particles force model, and then the multi-person counting and tracking steps were executed.

3.3.1. Human Silhouettes Verification: Particles Force Model

We present a robust particles force model for human silhouette verification. First of all, every extracted labeled silhouette was converted into particles, as shown in Figure 5a. We treated all pixels as fluid particles, thus, every extracted silhouette was a collection of many particles, as depicted in the magnified view in Figure 5b. Therefore, in our designed method, each silhouette was represented by a set of particles Q = [p_1, p_2, p_3, … , p_N], where N is the total number of particles in one silhouette.

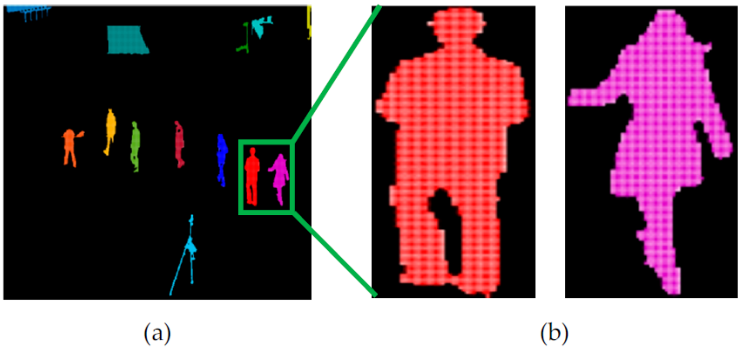


Figure 5. The particles force model. (**a**) Particle conversion of every extracted silhouette and (**b**) magnified view of particle conversion.

We know from physics that, in solids, particles do not have enough kinetic energy to overcome the strong forces of attraction, called bonds, which attract the particles toward each other. Using this physics phenomenon, we found the force of attraction between particles of every extracted silhouette, as shown in Figure 6:

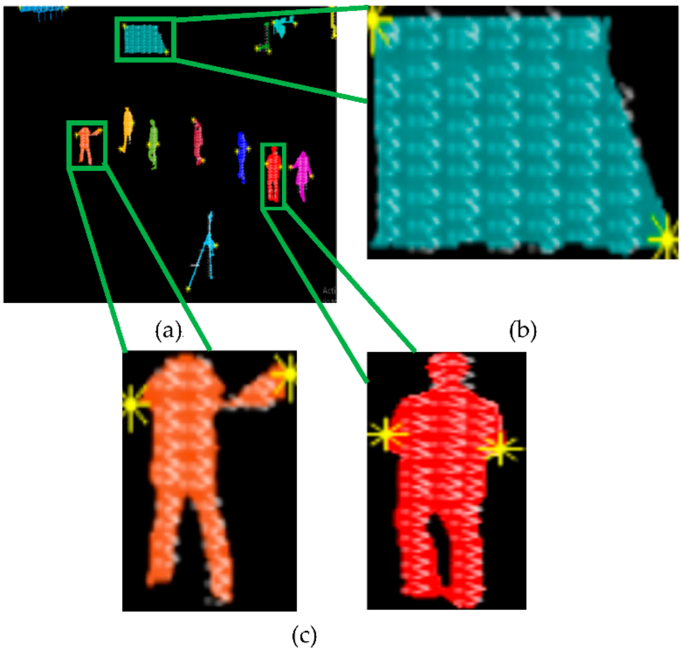


Figure 6. Particles force model. (**a**) Interacting force between two particles (**b**) for non-human silhouettes and (**c**) human silhouettes.

For simplicity, we found the force of attraction between only two mutually interacting particles using Equation (10) in all frames from 1 to Z.

$$F_i = \frac{p_1 p_2}{r^2} \tag{10}$$

where i is in the range [1 E] with E, representing the maximum number of silhouettes per frame, while F_i is the force of attraction between particle p_1 and p_2 of the ith silhouette and r^2 is the square of Euclidian distance between particles p_1 and p_2. After calculating the force between particles of every silhouette in all video frames, we discarded those silhouettes whose force of attraction was static in frame t and frame t + 1 using Equation (11):

$$H_s = \begin{cases} 1 & \text{if } \frac{dF_i}{dt} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $\frac{dF_i}{dt}$ represents the change in attraction force between particles of every *i*th silhouette, with respect to time between frames *t* to *t* + 1. After application of the particles force model, we only retained human silhouettes in each frame, as depicted in Figure 7:

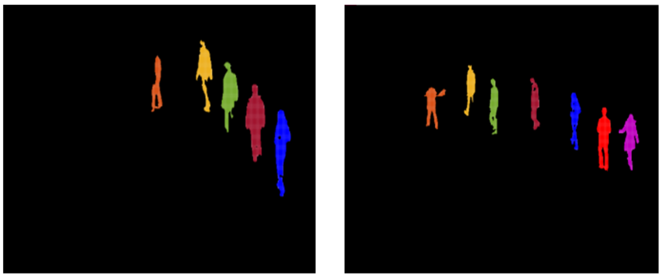


Figure 7. A few examples of verified multi-human silhouettes.

3.3.2. Multi-Person Counting

After extraction of the verified human silhouettes, to count these detected humans silhouettes, which consist of a set of particles, we performed cluster estimation. Since every silhouette is a collection of particles, the group of particles that makes one silhouette was treated as one cluster, and, by using the K-nearest neighbor search algorithm, cluster estimation was performed on every frame, as depicted in Figure 8:

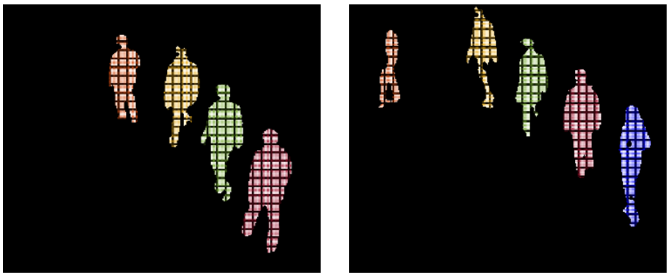


Figure 8. Human contours for cluster estimations.

After that, we labeled clusters in all frames, as shown in Equation (12), and, to make them appear visually, we drew green bounding boxes around each cluster. Thus, by performing cluster estimation and labeling, we counted all the extracted human silhouettes, as shown in Figure 9:

$$I_c = L_m p_N \tag{12}$$

where p_N is the total number of particles in one cluster (the total number of particles in each cluster varies from cluster to cluster and the number of clusters in each frame varies from frame to frame), while L_m represents the label of cluster m and I_c is the resultant extracted labeled cluster that was treated as one silhouette and was considered in counting.

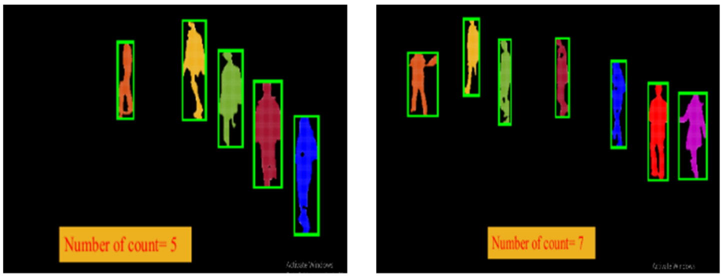


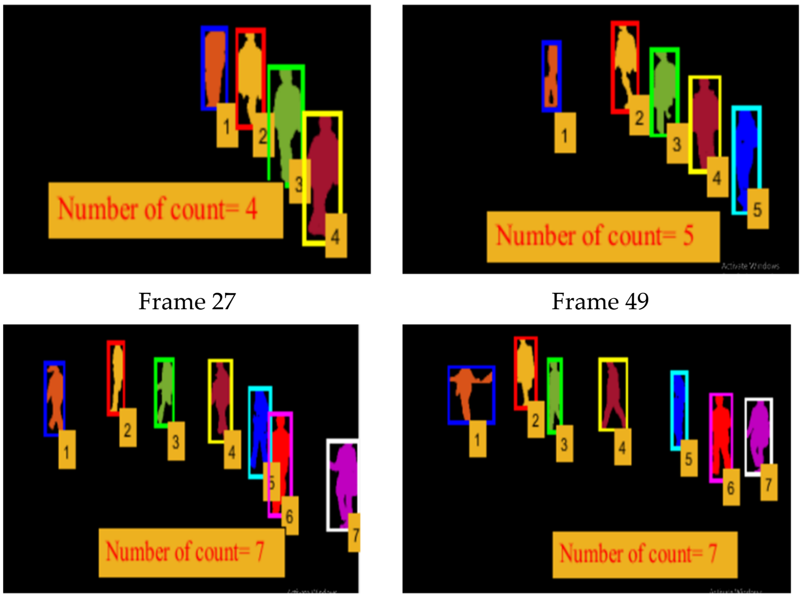
Figure 9. Sample frames of multi-person counts at different time intervals.

3.3.3. Multi-Person Tracking

The goal of person tracking is to establish correspondence between individuals across frames. Thus, to establish correspondence between persons in frame t and frame $t + 1$, we calculated the position and velocity of every detected human silhouette in all frames. In our model, we assumed that people can enter or leave the scene, thus, for temporally fixing of all humans across frames, the position of each human silhouette was located and locked by assigning a unique integer ID, which was fixed to that particular silhouette in all frames. The states of all the predicted persons in frame F_t were stored in a structure and matched with the states of frame $F_{t\,+1}$, while the detected fixed human silhouettes were tracked using the Jaccard similarity index.

$$S_t = \sum\nolimits_{i=1}^{n} I_{Li} \tag{13}$$

While using data association and cross-correlation as a cost function, detected and predicted persons were associated in consecutive frames, as represented in Figure 10. The root steps involved in multi-person tracking are illustrated in Figure 11.



Frame 27 Frame 49

Frame 67 Frame 81

Figure 10. Sample frames of multiple human silhouette-fixing and tracking at different time intervals.

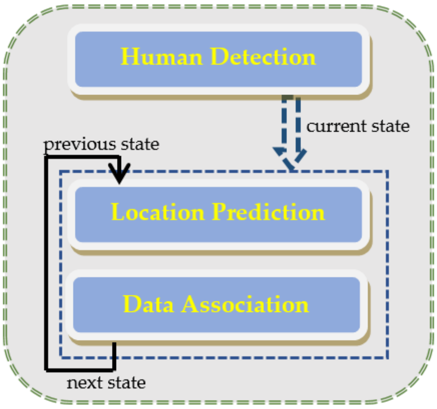


Figure 11. Key steps involved in multi-person tracking.

3.4. Crowd Behavior Detection

Understanding that accurate crowd behavior requires robust global and local feature extraction [101–103], along with a potent decision-making classifier, for crowd behavior detection after applying the distance algorithm (mentioned in Section 3.3), the extracted silhouettes were passed through the feature extraction step and multiple distinguishable global and local features were extracted for every frame. Next, bat optimization was applied for optimal feature extraction and decisions were made by the improved entropy classifier.

3.4.1. Global-Local Descriptors

For the global-local descriptor, we used a fusion of global and local image properties. In global features, we described the visual content of the whole image and we had the ability to represent an image with a single vector. Here, we extracted the crowd contour as a global feature. For local features, we used our newly proposed particles gradient motion features, geometric features, and speeded up robust feature (SURF) [104]. For local features, we extracted interest points and represented them as a set of vectors that respond more vigorously to clutter and occlusions.

Initially, in global features, we found the center of each human and considered all the humans in the scene as a vertex; this can be denoted as $P = \{P1, P2, \ldots, Pn \mid Pi = (Xi, Yi)\}$, where P represents the whole human crowd in the scene, considered as a set of vertices, and (Xi, Yi) are the coordinates of the ith human. We considered only those humans that were at the extreme points and joined them with a line, forming the biggest graph, covering all extreme vertices, as shown in Figure 12. The graph represented the human crowd contour, and thus, the variations in the shape of a graph threw a flash on variations in the outer area of the human crowd, i.e., on global changes. To measure the variations in the crowd contour, we compared the contour temporally by integrating over all of the pixels of the contour. In general, we defined the (p, q) moment of a contour as in Equation (14):

$$m_{p,q} = \sum_{x}^{n} \sum_{y}^{n} I(x,y) x^{p} y^{q} \tag{14}$$

where $I(x,y)$ is the intensity of the pixels in coordinate (x,y). Here, p is the x-order and q is the y-order, whereby, order means the power to which the corresponding component is taken in the sum just displayed. The summation is over all of the pixels of the contour boundary (denoted by n in the equation). It then follows immediately that, if p and q are both equal to 0, then the $m_{0,0}$ moment is actually just the length in pixels of the contour. The moment computation just described gives some rudimentary characteristics of a contour that can be used to compare two contours.

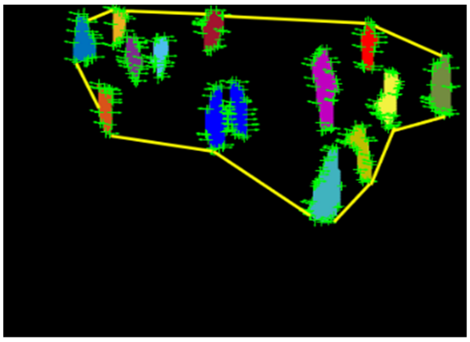


Figure 12. Extraction of human crowd contour as a global feature.

In the SURF descriptor [105], we computed distinctive invariant local features, which detected the interest points and elaborate features that depict some invariance to image noise, rotation, direction, scaling, and changes in illumination. Using SURF, we computed 75 local points for every human silhouette in an image, and thus, for every frame, we had 1050 SURF descriptors in a set of vectors, as shown in Figure 13:

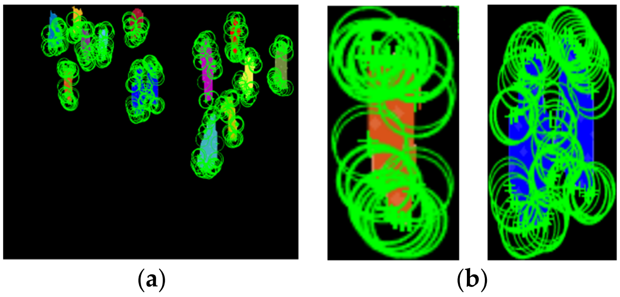


Figure 13. (**a**) SURF features for all human silhouettes and (**b**) magnified view of SURF features for two human silhouettes.

In geometric local features, we first identified the skeleton joints of every human silhouette in each frame using a skeleton model, and then, by considering skeleton joints as vertices, we drew poly-shapes and triangles with three or four vertices. By using the left arm, neck, left shoulder, and torso, a left polygon wing was drawn and filled with a color. Similarly, a right polygon wing was drawn and filled with different colors using the right arm, neck, torso, and right shoulder. Additionally, the torso area, lower area, left shoulder triangles, and right shoulder triangles were drawn, as depicted in Figure 14. The areas enclosed under these polygons were analyzed frame by frame, and on the basis of angle differences and area size, normal and abnormal behaviors of human crowds were detected. Algorithm 1 depicts the overall procedure used for the extraction of the strongest body points for human silhouettes.

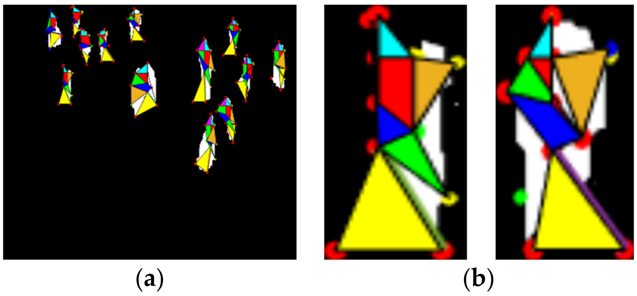


Figure 14. (**a**) Geometric features for all human silhouettes. (**b**) Magnified view of geometric features for two human silhouettes.

In particles gradient motion (PGM), we first converted every human silhouette into particles and then only those particles that were on the human contour were considered, and their interaction force was calculated. Generally, every pedestrian in a crowd has a desired direction and velocity $v_i{}^d$, calculated using Equation (16). However, in crowded scenes, because of the presence of multiple persons, individual movements are limited, and the actual velocity of each pedestrian v_i is different from their respective expected motion. The actual velocity of particles is calculated using Equation (15).

$$v_i = F_{avg}(x_i,\ y_i) \tag{15}$$

where $F_{avg}(x_i,\, y_i)$ is the ith particle average optical flow in the coordinate $(x_i,\, y_i)$. We calculated the desired velocity $v_i{}^d$ of particles as:

$$v_i{}^d = (1 - w_i)\; F(x_i,\, y_i) + w_i F_{avg}(x_i,\, y_i) \tag{16}$$

where $F(x_i,\, y_i)$ represents ith particle optical flow with coordinates $(x_i,\, y_i)$ and w_i is the panic weight parameter. The pedestrian i displays vanity behaviors as $w_i \to 0$ and collective behaviors as $w_i \to 1$. Linear interpolation was used for the enumeration of efficient optical flow and the adequate average flow field of particles. Thus, on the basis of the actual velocity and the desired velocity, we can calculate the interaction force using Equation (17):

$$F_{\text{int}} = \frac{1}{T}\left(v_i{}^d\; - v_i\right) - \frac{dv_i}{dt} \tag{17}$$

where F_{int} is the resultant interaction force, as represented in Figure 15 and T is the relaxation parameter. When the interaction force of particles was greater than the set threshold, it was detected as an abnormal event; otherwise, it was considered to be normal.

Algorithm 1 Extract strongest body points for human silhouettes

```
Input: I: Extracted Human Silhouettes
Output: Strongest body points, i.e., head, shoulders, legs, arms, hips
/* for each connected component, extract body points.
B = bwboundaries(binary_image);
lbl = bwlabel(binary_image);
CC2 = bwconncomp(lbl);
L52 = labelmatrix(CC2);
for objectidx2 = 1:CC2.NumObjects
    individualsilheouts2 = bsxfun(@times, closezn, L52 == objectidx2);
    [labeledImage2,numberofBlobs2] = bwlabel(individualsilheouts2,4);
end
Aa = individualsilheouts2;
/* Defining a upper, middlle and lower portion for each individual silheouts */
th = thershold;
rps = regionprops(Aa,'Boundingbox', 'Area');
for k = 1 to length(rps) do
    w = rps(k). Boundingbox
    if height > th and width > th then
        upper_region = struct('x',w(1), 'y', w(2), 'width',w(3), 'height', w(4)/5); /* head */
        middle_region = struct('x',w(1), 'y', w(2) + w(4)/4, 'width',w(3), 'height', w(4)/4); /* arms */
        lower_region = struct('x',w(1), 'y', w(2) + w(4)/2, 'width',w(3), 'height', w(4)/2); /* legs */
        j = j+1;
        s(j) = w;
    end
end
top = [x,max_y]:left = [min_x,y]:bottom = [x,min_y]:right = [max_x,y];
% label the head region%
Head =top pixels of upper region
Right Shoulder = Bottom right pixels of upper region
Left Shoulder = Bottom left pixels of upper region
Right arm = Right Pixels of middle region
Left arm = Left Pixels of middle region
Right foot = Bottom right pixels of lower region
Left foot = Bottom left pixels of lower region
return Head, Shoulders, arms, foots
```

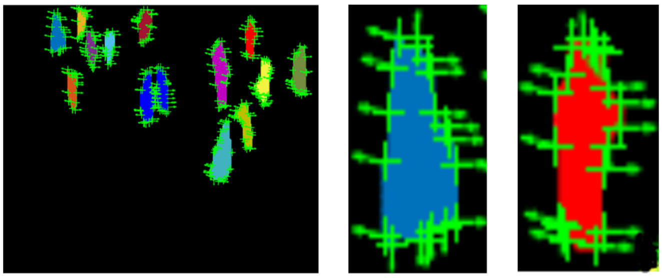


Figure 15. (**a**) Particles gradient motion descriptors for all human silhouettes and (**b**) magnified view of PGM for two human silhouettes.

3.4.2. Event Optimization: Bat Optimization

Optimization is a process by which the optimal solutions of a problem that satisfies and objective function are accessed [106–109]. Yang, in [110], introduced an optimization algorithm inspired by a property of bats, known as echolocation. Echolocation is a type of sonar that enables bats to fly and hunt in the dark. The bat optimization (BO) algorithm is composed of multiple variables of a given problem. Using the echolocation capability, bats can detect obstacles in the way and the distance, orientation, type, size, and even the speed of their prey.

BO has multiple agents depicting the parameters of the layout dilemma, as any other metaheuristic mechanism. From real-valued vectors, the initial population is randomly generated with number N and dimension d by considering lower and upper boundaries using Equation (18):

$$X_{ij} = X_{min} + \varphi(X_{max} - X_{min}) \tag{18}$$

where X_{max} and X_{min} are higher and lesser boundaries for dimension j, respectively, $j = 1,\ 2,\ \ldots,\ d,\ i$ = 1, 2, . . . , and N and φ ranged from 0 to 1 is a randomly generated value. After population initialization, we calculated the fitness function and stored the local and the global best. We evaluated the fitness values of all humans, and, on the basis of their movements, new local and global best solutions were obtained; all the humans had velocity Vi^t affected by a predefined frequency f_i, and finally, their new position Xi^t was located temporally, as described in the following Equations:

$$f_i = f_{min} + \beta(f_{max} - f_{min}) \tag{19}$$

$$Vi^t = Vi^{t-1} + (Xi^t - X*)f_i \tag{20}$$

$$Xi^t = Xi^{t-1} + Vi^t \tag{21}$$

where f_i is the frequency of the ith human, f_{min} and f_{max} are lower and higher frequency values, respectively, β represents a randomly generated value, and, after comparison of all solutions, $X*$ illustrates achieved global best location (solution). Figure 16 depicts the flow chart of the algorithm and Figure 17 represents optimization results.

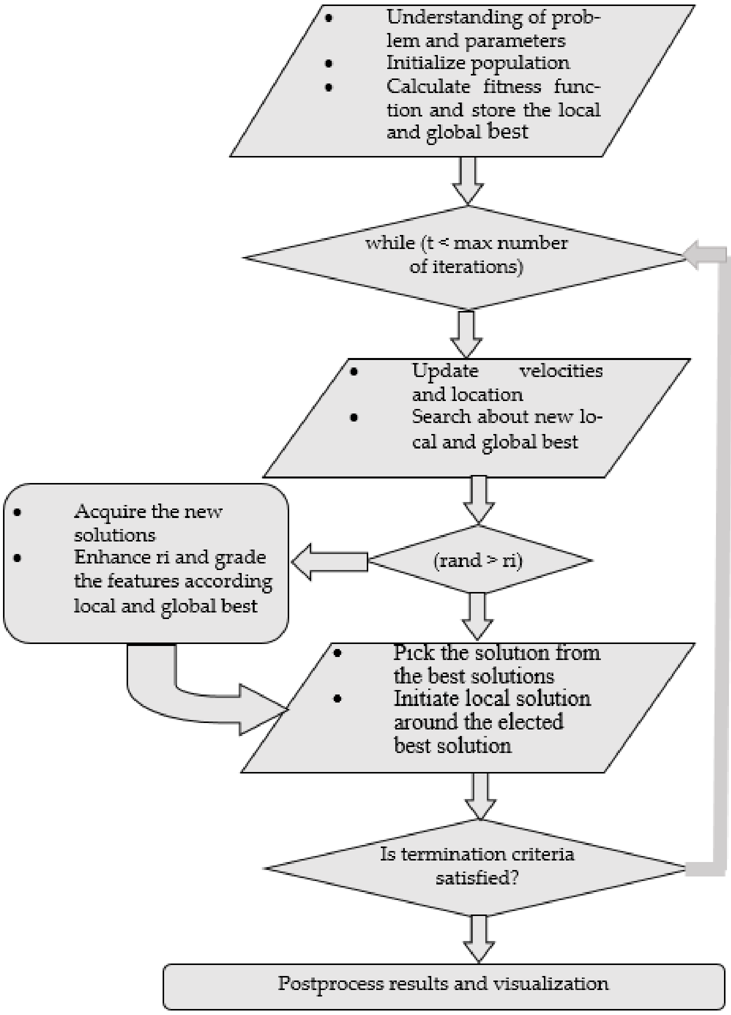
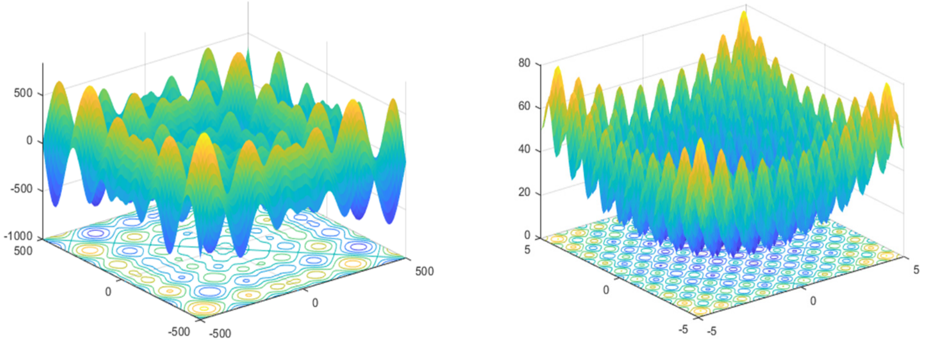

Understanding of problem and parameters
Initialize population
Calculate fitness function and store the local and global best
while (t < max number of iterations)
Update velocities and location
Search about new local and global best
Acquire the new solutions
Enhance ri and grade the features according local and global best
(rand > ri)
Pick the solution from the best solutions
Initiate local solution around the elected best solution
Is termination criteria satisfied?
Postprocess results and visualization

Figure 16. Bat optimization flow chart.



(**a**) (**b**)

Figure 17. Bat optimization results. (**a**) Normal optimal features; (**b**) abnormal optimal features.

3.4.3. Improved Entropy Classifier

Using Shannon's information entropy theory [53] to describe the degree of uncertainty, we proposed an improved entropy classifier for the detection of human crowd behavior. First of all, we standardized all the features using Equation (22):

$$X_{ij}{}^{*} = \frac{X_{ij} - min\{X_j\}}{max\{X_j\} - min\{X_j\}} \tag{22}$$

where $X_{ij}{}^{*}$ is the value of the j-th feature for i-th human. $j = 1, 2, \ldots, m$, $i = 1, 2, \ldots, n$, while n is the count of humans and m represents the count of features. After that, the weight of j-th feature for i-th human was calculated using Equation (23):

$$q_{ij} = \frac{X_{ij}{}^{*}}{\sum_{i=1}^{n} X_{ij}{}^{*}} \tag{23}$$

Thus, the information entropy of each feature was calculated using Equation (24):

$$e_j = -k \sum_{i=1}^{n} \left(q_{ij} \times ln q_{ij}\right) \tag{24}$$

where $k = \frac{1}{\ln m}$. After calculating the information entropy, we then calculated the difference coefficient and maximum ratio of the difference coefficient using Equations (25) and (26):

$$d_j = 1 - e_j \tag{25}$$

$$D = \frac{\max(d_j)}{\min(d_j)}, \qquad (j = 1, 2, \ldots, m) \tag{26}$$

After calculating D, we then built up the scale ratio chart 1–9 using Equation (27):

$$R = \sqrt[a-1]{\frac{D}{a}} \tag{27}$$

where a depicts the highest scale-value worked as an adjustment coefficient by calculating the power ($a - 1$). The D is allocated to the mapping values from 1 to 9 in the above Equation. After that, from scale 1–9, mapped values were calculated, and judgment matrix R was established with elements r_{ij}, respectively, using Equation (28):

$$r_{ij} = \frac{d_i}{d_j}, \ \left(d_i > d_j\right) \tag{28}$$

The obtained judgment matrix satisfied the consistency test because the elements r_{ij} demonstrated the ratio of difference coefficient of two features.

Thus, the consistent weights W_j for each feature were then calculated using an analytical hierarchy process. After that, information entropy was again calculated for each feature, using these weights by utilizing Equation (24). The crowd behavior entropy of the whole system was the summary of all entropies. In this way, for every frame, the entropy value was calculated and utilized as a template. For a small entropy value less than the defined threshold, the behavior was predicted as normal; however, for entropy values higher than the set threshold, the behavior was presumed to be abnormal. A flow chart of the proposed improved entropy classifier is shown in Figure 18. Figure 19 depicts results over event classes.

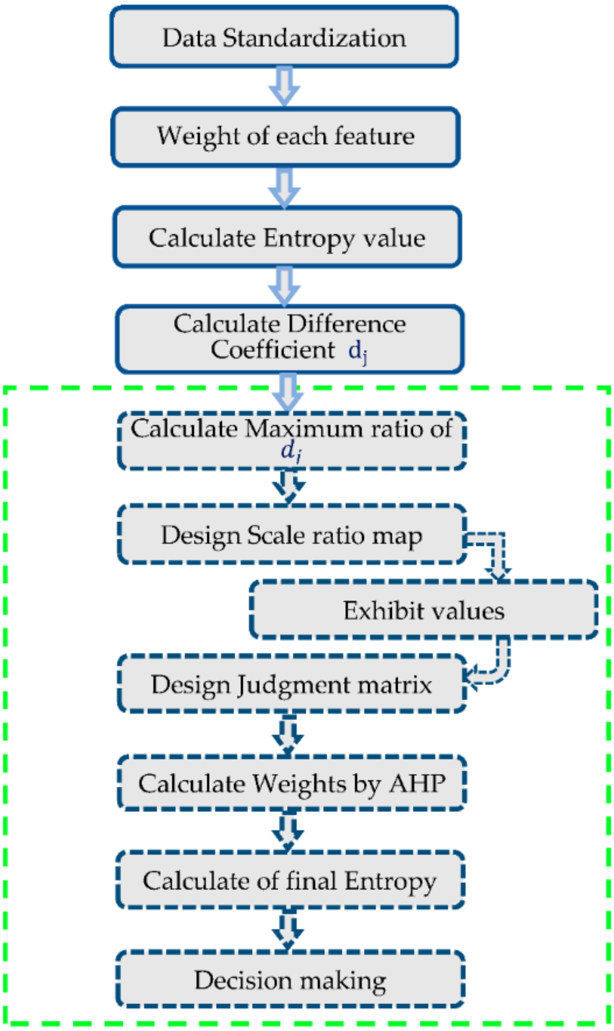


Figure 18. Flow chart of the improved entropy classifier.

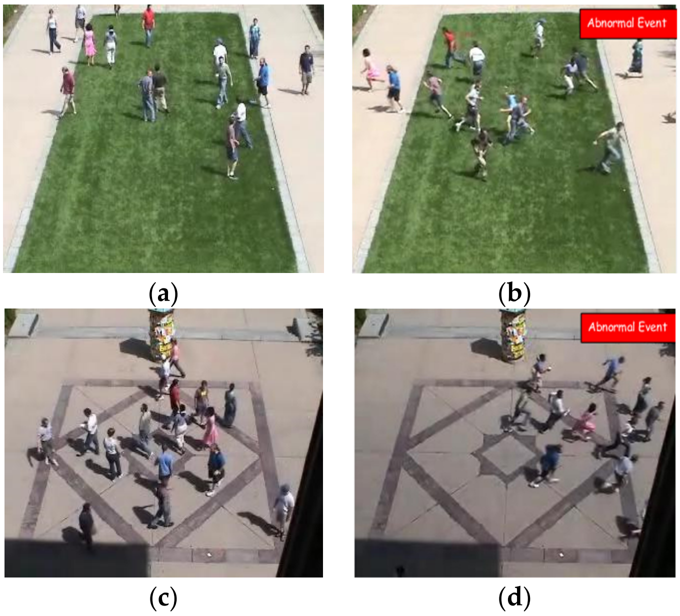


Figure 19. Crowd behavior detection. (**a**,**c**) Normal frames and (**b**,**d**) abnormal frames.

4. Performance Evaluation

In this section, we evaluated the performance of our proposed system. We conducted experiments on two publicly available benchmark datasets to evaluate the accuracy and performance of our proposed model. The PETS2009 dataset was used to evaluate the accuracy of multi-person tracking and the UMN dataset was used to evaluate the accuracy of crowd behavior detection. We started by briefly describing the datasets used, and then the experimental results were discussed. Finally, we showed the mean accuracy of our proposed system. We also compared our proposed model with other state-of-the-art multi-person tracking and crowd behavior detection systems.

4.1. Datasets Description

4.1.1. PETS2009 Dataset

To evaluate different video surveillance challenges, we used PETS2009, one of the publicly available benchmark datasets. The challenges included the S1 dataset for counting persons in a low-density crowd, the S2 dataset for detecting and tracking persons in medium-density crowds, and the S3 dataset for tracking and estimating the number of persons in a high-density crowd. Some sample frames of different synchronized views from PETS2009 dataset are depicted in Figure 20.

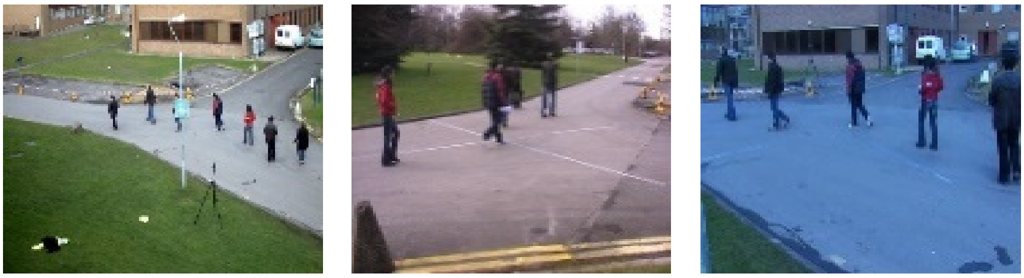


Figure 20. Sample frames of different synchronized views from the PETS2009 dataset.

4.1.2. UMN Dataset

To evaluate different video surveillance challenges for crowd behavior detection, UMN is one of the publicly available benchmark datasets. The UMN dataset consists of three different scenes, specifically, two outdoor and one indoor, with videos of 11 various panic scenarios. For the detection of abnormal behavior of a crowd, the UMN dataset is one of the best datasets that is publicly available. There were two outdoor scenes: the lawn scene, consisting of two scenarios with 1453 frames, and the Plaza scene, with three scenarios that had 2142 frames. There were six scenarios in the indoor scene, with 4144 frames. Sample frames of different scenarios of the UMN dataset are shown in Figure 21.

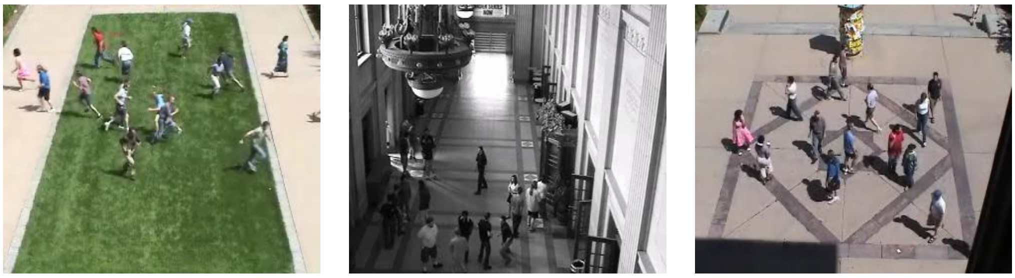


Figure 21. Sample frames of different scenarios of the UMN dataset.

4.2. Experimental Settings and Results

We performed all the experiments on MATLAB, and the hardware system had a 64-bit intel core-i3 2.5 GHz CPU and 6 GB of RAM. Three experimental measures were used to evaluate the performance of the system: (1) mean accuracy of multi-person tracking, (2) the accuracy of human crowd behavior detection, and (3) comparisons between our proposed new system with other current and well-known systems. Experimental results showed that our proposed system produces a higher accuracy rate over existing systems.

4.2.1. Experiment 1: Multi-Person Tracking over the PETS2009 Dataset

Experimental results and mean accuracy of our proposed multi-person counting and tracking model on a publicly available PETS2009 dataset are shown in Tables 1 and 2. The ground truth was obtained by counting the number of persons in every sequence, where one sequence contained 20 frames. Table 1 depicts the mean accuracy of our proposed multi-person counting system on the first 30 sequences. As shown, the mean accuracy of our proposed model was 89.80%.

Table 1. Multi-person counting accuracy over the PETS2009 dataset.

Sequence No (Frame = 20)	Actual Count	Predicted Count	Accuracy
6	3	3	100
12	4	4	100
18	5	4	80
24	6	5	83.33
30	7	6	85.71
Mean Accuracy = 89.80%			

Table 2. Multi-person tracking accuracy over PETS2009 dataset.

Sequence No (Frame = 20)	Successful	Failure	Accuracy
6	3	0	100
12	4	0	100
18	4	1	80
24	5	1	83.33
30	5	2	71.43
Mean Accuracy = 86.95%			

Table 2 presents the mean accuracy of our proposed multi-person tracking system. The actual number of humans is the same as for Table 1, while column 2 represents the successful tracking rate of our proposed particles force model and column 3 depicts the failure case. The mean accuracy of our proposed model for multiple person tracking was 86.95%.

4.2.2. Experiment 2: Human Crowd Behavior Detection over the UMN Dataset

Experimental results using the confusion matrix and the mean accuracy of our proposed HCB model on the publicly available UMN dataset are shown in Table 3. The way to evaluate algorithms is to run them throughout a test sequence with initialization from the ground truth position in the first frame.

Table 3. Confusion matrix, showing mean accuracy for human crowd behavior detection on the UMN dataset.

Events	Normal	Abnormal
Normal	88	12
Abnormal	16	84
Mean Accuracy of Event Detection = 86.06%		

4.2.3. Experiment 3: Multi-Person Tracking and HCB Detection Comparisons with State-of-the-Art Methods

We compared our proposed system with other well-known multi-person tracking and human crowd behavior detection methods. As depicted, our system performed better compared to existing well-known state-of-the-art methods. Table 4 shows that, in comparison to other state-of-the-art methods, our proposed system achieved an admirable accuracy rate of 86.06% for crowd behavior detection, which is higher than the accuracy of the force field model (FF) (81.04%) and the social force model (SF) (85.09%). The accuracy of other methods under the same evaluation settings was taken from [77,79].

Table 4. Comparison of the proposed approach with other state-of-the-art methods for human crowd behavior detection on the UMN dataset.

Indoor/Outdoor Scenes	Force Field Model	Social Force Model	Proposed Method
Scene 1	88.69	84.41	87.43
Scene 2	80.00	82.35	83.21
Scene 3	77.92	90.83	90.63
Overall accuracy	81.04%	85.09%	86.06%

Table 5 presents the comparison of our proposed system with other state-of-the-art systems for multi-person counting. Experiment results show that our proposed system achieved a higher accuracy rate of 89.8% over existing methods.

Table 5. Comparison of proposed approach with state-of-the-art multi-person counting methods.

Methods	Counting Accuracy (%)
Pixel-map based algorithm [94]	83.6
Sparsity-driven [111]	86.3
Head Shoulder based detection [100]	86.7
Skin Detection [81]	88.7
Proposed method	89.8

In Table 6, comparisons of multi-person tracking with other state-of-the-art methods show that our proposed system achieved a higher accuracy rate of 86.9% over existing methods.

Table 6. Comparison of the proposed approach with state-of-the-art multi-person tracking methods.

Methods	Tracking Accuracy (%)
Flow Linear Programming [112]	78.8
DDPMO [113]	81.3
Appearance model [114]	83.0
Proposed method	86.9

5. Conclusions

In this paper, we proposed a new robust approach for crowd counting. We introduced and tested tracking and human behavior detection using the idea of a mutually interacting particles force model and an improved entropy classifier with spatio-temporal and particles gradient motion descriptors. Through detailed experiments, we proved the ability of the method to efficiently count, track, and detect the behavior of multiple persons efficiently in crowded scenes. The performance of our new tracking system decreases marginally with increasing numbers of persons in the scene. This is mainly due to full occlusions that occur in the test videos. We achieved promising results on the publicly available benchmark PETS2009 dataset, with an accuracy of 89.80% for multi-person counting and 86.95% for person tracking, as shown in Tables 1 and 2. However, for HCB detection, we achieved promising results on the publicly available benchmark UMN dataset, with an accuracy of 86.06%, as shown in Table 3. Our future work will focus on some occlusion reasoning methods to further tackle the occlusion problems. We will also extend our work to multiple scene detection. We are interested in recognition of different scenes, such as sport scenes, fight scenes, robbery scenes, traffic scenes, and action scenes.

Author Contributions: Conceptualization, F.A., Y.Y.G. and K.K. methodology, F.A. and A.J.; software, F.A.; validation, F.A., Y.Y.G. and K.K.; formal analysis, M.G. and K.K.; resources, Y.Y.G., M.G. and K.K.; writing—review and editing, F.A., A.J. and K.K.; funding acquisition, Y.Y.G. and K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) under Grant 2018R1D1A1A02085645, by the Korea Medical Device Development Fund grant funded by the Korean government (the Ministry of Science and ICT; the Ministry of Trade, Industry and Energy; the Ministry of Health and Welfare; and the Ministry of Food and Drug Safety) (202012D05-02), and by Hanyang University Grant No. 201800000000647.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fu, Z.; Feng, P.; Angelini, F.; Chambers, J.; Naqvi, S.M. Particle PHD filter based multiple human tracking using online group-structured dictionary learning. *IEEE Access* **2018**, *6*, 14764–14778. [CrossRef]
2. Wen, L.; Lei, Z.; Lyu, S.; Li, S.Z.; Yang, M.H. Exploiting hierarchical dense structures on hypergraphs for multi-object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 1983–1996. [CrossRef] [PubMed]
3. Maggio, E.; Taj, M.; Cavallaro, A. Efficient multitarget visual tracking using random finite sets. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1016–1027. [CrossRef]
4. Yilmaz, A.; Javed, O.; Shah, M. Object tracking: A survey. *Acm Comput. Surv. (CSUR)* **2006**, *38*, 13-es. [CrossRef]
5. Marcenaro, L.; Marchesotti, L.; Regazzoni, C.S. Tracking and counting multiple interacting people in indoor scenes. In Proceedings of the Third IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Copenhagen, Denmark, 1 June 2002; pp. 56–61.
6. Ali, S.; Shah, M. Floor fields for tracking in high density crowd scenes. In Proceedings of the European Conference on Computer Vision, Marseille, France, 12–18 October 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–14.
7. Jalal, A.; Kim, Y. Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data. In Proceedings of the 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Seoul, Korea, 26–29 August 2014; pp. 119–124.
8. Yao, R.; Lin, G.; Xia, S.; Zhao, J.; Zhou, Y. Video object segmentation and tracking: A survey. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*, 1–47. [CrossRef]
9. Pervaiz, M.; Jalal, A.; Kim, K. Hybrid Algorithm for Multi People Counting and Tracking for Smart Surveillance. In Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, 12–16 January 2021; pp. 530–535.
10. Topkaya, I.S.; Erdogan, H.; Porikli, F. Counting people by clustering person detector outputs. In Proceedings of the IEEE International Conference on AVSS, Seoul, Korea, 26–29 August 2014; pp. 313–318.
11. Ren, W.; Wang, X.; Tian, J.; Tang, Y.; Chan, A.B. Tracking-by-Counting: Using Network Flows on Crowd Density Maps for Tracking Multiple Targets. *IEEE Trans. Image Process.* **2020**, *30*, 1439–1452. [CrossRef]
12. Loy, C.C.; Chen, K.; Gong, S.; Xiang, T. Crowd counting and profiling: Methodology and evaluation. In *Modeling, Simulation and Visual Analysis of Crowds*; Springer: New York, NY, USA, 2013; pp. 347–382.
13. Mahmood, M.; Jalal, A.; Kim, K. WHITE STAG model: Wise human interaction tracking and estimation (WHITE) using spatio-temporal and angular-geometric (STAG) descriptors. *Multimed. Tools Appl.* **2010**, *79*, 6919–6950. [CrossRef]
14. Jalal, A.; Kim, K. Wearable inertial sensors for daily activity analysis based on adam optimization and the maximum entropy Markov model. *Entropy* **2020**, *22*, 579.
15. Ryan, D.; Denman, S.; Sridharan, S.; Fookes, C. An evaluation of crowd counting methods, features and regression models. *Comput. Vis. Image Underst.* **2015**, *130*, 1–17. [CrossRef]
16. Idrees, H.; Saleemi, I.; Seibert, C.; Shah, M. Multi-source multi-scale counting in extremely dense crowd images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2547–2554.
17. Ekinci, M.; Gedikli, E. Silhouette based human motion detection and analysis for real-time automated video surveillance. *Turk. J. Electr. Eng. Comput. Sci.* **2005**, *13*, 199–229.
18. Younsi, M.; Diaf, M.; Siarry, P. Automatic multiple moving humans detection and tracking in image sequences taken from a stationary thermal infrared camera. *Expert Syst. Appl.* **2020**, *146*, 113171. [CrossRef]

19. Lempitsky, V.; Zisserman, A. Learning to count objects in images. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 1324–1332.
20. Farooq, M.U.; Saad, M.N.B.; Malik, A.S.; Salih Ali, Y.; Khan, S.D. Motion estimation of high density crowd using fluid dynamics. *Imaging Sci. J.* **2020**, *68*, 141–155. [CrossRef]
21. Sajid, M.; Hassan, A.; Khan, S.A. Crowd counting using adaptive segmentation in a congregation. In Proceedings of the 2016 IEEE International Conference on Signal and Image Processing (ICSIP), Beijing, China, 13–15 August 2016; pp. 745–749.
22. Fehr, D.; Sivalingam, R.; Morellas, V.; Papanikolopoulos, N.; Lotfallah, O.; Park, Y. Counting people in groups. In Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy, 2–4 September 2009; pp. 152–157.
23. Jalal, A.; Uddin, M.Z.; Kim, T.S. Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. *IEEE Trans. Consum. Electron.* **2012**, *58*, 863–871. [CrossRef]
24. Jalal, A.; Mahmood, M.; Hasan, A.S. Multi-features descriptors for human activity tracking and recognition in Indoor-outdoor environments. In Proceedings of the 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 8–12 January 2019; pp. 371–376.
25. Albiol, A.; Silla, M.J.; Albiol, A.; Mossi, J.M. Video analysis using corner motion statistics. In Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Miami, FL, USA, 25 June 2009; pp. 31–38.
26. Li, N.; Wu, X.; Guo, H.; Xu, D.; Ou, Y.; Chen, Y.L. Anomaly detection in video surveillance via gaussian process. *Int. J. Pattern Recognit. Artif. Intell.* **2015**, *29*, 1555011. [CrossRef]
27. Chan, A.B.; Morrow, M.; Vasconcelos, N. Analysis of crowded scenes using holistic properties. In Proceedings of the Performance Evaluation of Tracking and Surveillance Workshop at CVPR, Miami, FL, USA, 25 June 2009; pp. 101–108.
28. Sharif, M.H.; Djeraba, C. An entropy approach for abnormal activities detection in video streams. *Pattern Recognit.* **2012**, *45*, 2543–2561. [CrossRef]
29. Jalal, A.; Sarif, N.; Kim, J.T.; Kim, T.S. Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home. *Indoor Built Environ.* **2013**, *22*, 271–279. [CrossRef]
30. Jalal, A.; Kamal, S. Real-time life logging via a depth silhouette-based human activity recognition system for smart home services. In Proceedings of the 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Seoul, Korea, 26–29 August 2014; pp. 74–80.
31. Raghavendra, R.; Del Bue, A.; Cristani, M.; Murino, V. Abnormal crowd behavior detection by social force optimization. In Proceedings of the International Workshop on Human Behavior Understanding, Amsterdam, The Netherlands, 16 November 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 134–145.
32. Javeed, M.; Gochoo, M.; Jalal, A.; Kim, K. HF-SPHR: Hybrid Features for Sustainable Physical Healthcare Pattern Recognition Using Deep Belief Networks. *Sustainability* **2021**, *13*, 1699. [CrossRef]
33. Chan, A.B.; Liang, Z.S.J.; Vasconcelos, N. Privacy preserving crowd monitoring: Counting people without people models or tracking. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–7.
34. Jalal, A.; Mahmood, M. Students' behavior mining in e-learning environment using cognitive processes with information technologies. *Educ. Inf. Technol.* **2019**, *24*, 2797–2821. [CrossRef]
35. Dey, P.; Roberts, D. A conceptual framework for modelling crowd behavior. In Proceedings of the 11th IEEE International Symposium on Distributed Simulation and Real-Time Applications, Chania, Greece, 22–26 October 2007; pp. 193–200.
36. Bellomo, N.; Dogbe, C. On the modelling crowd dynamics from scaling to hyperbolic macroscopic models. *Math. Models Methods Appl. Sci.* **2008**, *18* (Suppl. 1), 1317–1345. [CrossRef]
37. Roy, S.; Annunziato, M.; Borzì, A. A Fokker–Planck feedback control-constrained approach for modelling crowd motion. *J. Comput. Theor. Transp.* **2016**, *45*, 442–458. [CrossRef]
38. Ansar, H.; Jalal, A.; Gochoo, M.; Kim, K. Hand Gesture Recognition Based on Auto-Landmark Localization and Reweighted Genetic Algorithm for Healthcare Muscle Activities. *Sustainability* **2021**, *13*, 2961. [CrossRef]
39. Javeed, M.; Jalal, A.; Kim, K. Wearable Sensors based Exertion Recognition using Statistical Features and Random Forest for Physical Healthcare Monitoring. In Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, 12–16 January 2021; pp. 512–517.
40. Kremer, M.; Haworth, B.; Kapadia, M.; Faloutsos, P. Modelling distracted agents in crowd simulations. *Vis. Comput.* **2021**, *37*, 107–118. [CrossRef]
41. Khalid, N.; Gochoo, M.; Jalal, A.; Kim, K. Modeling Two-Person Segmentation and Locomotion for Stereoscopic Action Identification: A Sustainable Video Surveillance System. *Sustainability* **2021**, *13*, 970. [CrossRef]
42. Jalal, A.; Kamal, S. Improved Behavior Monitoring and Classification Using Cues Parameters Extraction from Camera Array Images. *Int. J. Interact. Multimed. Artif. Intell.* **2019**, *5*, 71–78. [CrossRef]
43. Elaiw, A.; Al-Turki, Y.; Alghamdi, M. A critical analysis of behavioral crowd dynamics—From a modelling strategy to kinetic theory methods. *Symmetry* **2019**, *11*, 851. [CrossRef]
44. Nady, A.; Atia, A.; Abutabl, A. Real-time abnormal event detection in crowded scenes. *J. Theory Appl. Inf. Technol.* **2018**, *96*, 6064–6075.
45. He, F.; Xiang, Y.; Zhao, X.; Wang, H. Informative scene decomposition for crowd analysis, comparison and simulation guidance. *ACM Trans. Graph. (TOG)* **2020**, *39*, 50–51. [CrossRef]

46. Jalal, A.; Batool, M.; Kim, K. Sustainable Wearable System: Human Behavior Modeling for Life-Logging Activities Using K-Ary Tree Hashing Classifier. *Sustainability* **2020**, *12*, 10324. [CrossRef]
47. Pennisi, A.; Bloisi, D.D.; Iocchi, L. Online real-time crowd behavior detection in video sequences. *Comput. Vis. Image Underst.* **2016**, *144*, 166–176. [CrossRef]
48. Jalal, A.; Kamal, S.; Kim, D. A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors* **2014**, *14*, 11735–11759. [CrossRef] [PubMed]
49. Jalal, A.; Quaid, M.A.K.; Kim, K. A Study of Accelerometer and Gyroscope Measurements in Physical Life-Log Activities Detection Systems. *Sensors* **2020**, *20*, 6670. [CrossRef]
50. Bellomo, N.; Clarke, D.; Gibelli, L.; Townsend, P.; Vreugdenhil, B.J. Human behaviours in evacuation crowd dynamics: From modelling to "big data" toward crisis management. *Phys. Life Rev.* **2016**, *18*, 1–21. [CrossRef]
51. Nadeem, A.; Jalal, A.; Kim, K. Human actions tracking and recognition based on body parts detection via Artificial neural network. In Proceedings of the 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 17–19 February 2020; pp. 1–6.
52. Fakhar, B.; Kanan, H.R.; Behrad, A. Event detection in soccer videos using unsupervised learning of Spatio-temporal features based on pooled spatial pyramid model. *Multimed. Tools Appl.* **2019**, *78*, 16995–17025. [CrossRef]
53. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
54. Zhang, X.; Yu, Q.; Yu, H. Physics inspired methods for crowd video surveillance and analysis: A survey. *IEEE Access* **2018**, *6*, 66816–66830. [CrossRef]
55. Jalal, A.; Khalid, N.; Kim, K. Automatic recognition of human interaction via hybrid descriptors and maximum entropy markov model using depth sensors. *Entropy* **2020**, *22*, 817. [CrossRef] [PubMed]
56. Ryan, D.; Denman, S.; Fookes, C.; Sridharan, S. Crowd counting using group tracking and local features. In Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, Boston, MA, USA, 29 August–1 September 2010; pp. 218–224.
57. Cong, Y.; Gong, H.; Zhu, S.C.; Tang, Y. Flow mosaicking: Real-time pedestrian counting without scene-specific learning. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1093–1100.
58. Nadeem, A.; Jalal, A.; Kim, K. Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model. *Multimed. Tools Appl.* **2021**, *80*, 1–34.
59. Jalal, A.; Akhtar, I.; Kim, K. Human Posture Estimation and Sustainable Events Classification via Pseudo-2D Stick Model and K-ary Tree Hashing. *Sustainability* **2020**, *12*, 9814. [CrossRef]
60. Wu, T.; Toet, A. Color-to-grayscale conversion through weighted multiresolution channel fusion. *J. Electron. Imaging* **2014**, *23*, 043004. [CrossRef]
61. Jalal, A.; Kim, Y.; Kim, D. Ridge body parts features for human pose estimation and recognition from RGB-D video data. In Proceedings of the Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Hefei, China, 11–13 July 2014; pp. 1–6.
62. Wang, J.; Xu, Z. Spatio-temporal texture modelling for real-time crowd anomaly detection. *Comput. Vis. Image Underst.* **2016**, *144*, 177–187. [CrossRef]
63. Zitouni, M.S.; Bhaskar, H.; Dias, J.; Al-Mualla, M.E. Advances and trends in visual crowd analysis: A systematic survey and evaluation of crowd modelling techniques. *Neurocomputing* **2016**, *186*, 139–159. [CrossRef]
64. Pretorius, M.; Gwynne, S.; Galea, E.R. Large crowd modelling: An analysis of the Duisburg Love Parade disaster. *Fire Mater.* **2015**, *39*, 301–322. [CrossRef]
65. Quaid, M.A.K.; Jalal, A. Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm. *Multimed. Tools Appl.* **2020**, *79*, 6061–6083. [CrossRef]
66. Jalal, A.; Quaid, M.A.K.; Kim, K. A wrist worn acceleration based human motion analysis and classification for ambient smart home system. *J. Electr. Eng. Technol.* **2019**, *14*, 1733–1739. [CrossRef]
67. Jalal, A.; Lee, S.; Kim, J.T.; Kim, T.S. Human activity recognition via the features of labeled depth body parts. In Proceedings of the International Conference on Smart Homes and Health Telematics, Artiminio, Italy, 12–15 June 202; Springer: Berlin/Heidelberg, Germany, 2012; pp. 246–249.
68. Chen, D.; Wang, L.; Wu, X.; Chen, J.; Khan, S.U.; Kołodziej, J.; Tian, M.; Huang, F.; Liu, W. Hybrid modelling and simulation of huge crowd over a hierarchical grid architecture. *Future Gener. Comput. Syst.* **2013**, *29*, 1309–1317. [CrossRef]
69. Jalal, A.; Mahmood, M.; Sidduqi, M.A. Robust spatio-temporal features for human interaction recognition via artificial neural network. In Proceedings of the IEEE International Conference on Frontiers of Information Technology, Islamabad, Pakistan, 17–19 December 2018.
70. Alguliyev, R.M.; Aliguliyev, R.M.; Sukhostat, L.V. Efficient algorithm for big data clustering on single machine. *CAAI Trans. Intell. Technol.* **2020**, *5*, 9–14. [CrossRef]
71. Jalal, A.; Batool, M.; Kim, K. Stochastic recognition of physical activity and healthcare using tri-axial inertial wearable sensors. *Appl. Sci.* **2020**, *10*, 7122. [CrossRef]
72. Jalal, A.; Kim, Y.H.; Kim, Y.J.; Kamal, S.; Kim, D. Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern Recognit.* **2017**, *61*, 295–308. [CrossRef]

73. Basavegowda, H.S.; Dagnew, G. Deep learning approach for microarray cancer data classification. *CAAI Trans. Intell. Technol.* **2020**, *5*, 22–33. [CrossRef]
74. Aylaj, B.; Bellomo, N.; Gibelli, L.; Knopoff, D. Crowd Dynamics by Kinetic Theory Modeling: Complexity, Modeling, Simulations, and Safety. *Synth. Lect. Math. Stat.* **2020**, *12*, 1–98.
75. Gochoo, M.; Akhter, I.; Jalal, A.; Kim, K. Stochastic Remote Sensing Event Classification over Adaptive Posture Estimation via Multifused Data and Deep Belief Network. *Remote Sens.* **2021**, *13*, 912. [CrossRef]
76. Rizwan, S.A.; Jalal, A.; Gochoo, M.; Kim, K. Robust Active Shape Model via Hierarchical Feature Extraction with SFS-Optimized Convolution Neural Network for Invariant Human Age Classification. *Electronics* **2021**, *10*, 465. [CrossRef]
77. Wu, S.; Wong, H.S.; Yu, Z. A Bayesian model for crowd escape behavior detection. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *24*, 85–98. [CrossRef]
78. Choudhary, S.; Ojha, N.; Singh, V. Real-time crowd behavior detection using SIFT feature extraction technique in video sequences. In Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 15–16 June 2017; pp. 936–940.
79. Direkoglu, C.; Sah, M.; O'Connor, N.E. Abnormal crowd behavior detection using novel optical flow-based features. In Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 29 August–1 September 2017; pp. 1–6.
80. Aguilar, W.G.; Luna, M.A.; Ruiz, H.; Moya, J.F.; Luna, M.P.; Abad, V.; Parra, H. Statistical abnormal crowd behavior detection and simulation for real-time applications. In Proceedings of the International Conference on Intelligent Robotics and Applications, Wuhan, China, 16–18 August 2017; Springer: Cham, Switzerland, 2017; pp. 671–682.
81. Shehzed, A.; Jalal, A.; Kim, K. Multi-person tracking in smart surveillance system for crowd counting and normal/abnormal events detection. In Proceedings of the 2019 International Conference on Applied and Engineering Mathematics (ICAEM), Taxila, Pakistan, 27–29 August 2017; pp. 163–168.
82. Ren, W.Y.; Li, G.H.; Chen, J.; Liang, H.Z. Abnormal crowd behavior detection using behavior entropy model. In Proceedings of the 2012 International Conference on Wavelet Analysis and Pattern Recognition, Xi'an, China, 15–17 July 2012; pp. 212–221.
83. Wang, G.; Fu, H.; Liu, Y. Real time abnormal crowd behavior detection based on adjacent flow location estimation. In Proceedings of the 2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), Beijing, China, 17–19 August 2016; pp. 476–479.
84. Zhao, Z.; Fu, S.; Wang, Y. Eye Tracking Based on the Template Matching and the Pyramidal Lucas-Kanade Algorithm. In Proceedings of the 2012 International Conference on Computer Science and Service System, Nanjing, China, 11–13 August 2012; pp. 2277–2280.
85. Mehran, R.; Oyama, A.; Shah, M. Abnormal crowd behavior detection using social force model. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 935–942.
86. Bellomo, N.; Bellouquid, A. On multiscale models of pedestrian crowds from mesoscopic to macroscopic. *Commun. Math. Sci.* **2015**, *13*, 1649–1664. [CrossRef]
87. Colombo, R.M.; Rossi, E. Modelling crowd movements in domains with boundaries. *IMA J. Appl. Math.* **2019**, *84*, 833–853. [CrossRef]
88. Khan, S.D.; Basalamah, S. Multi-Scale Person Localization with Multi-Stage Deep Sequential Framework. *Int. J. Comput. Intell. Syst.* **2021**, *14*, 1217–1228. [CrossRef]
89. Zhang, A.; Jiang, X.; Zhang, B.; Cao, X. Multi-scale Supervised Attentive Encoder-Decoder Network for Crowd Counting. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2020**, *16*, 1–20. [CrossRef]
90. Jiang, R.; Mou, X.; Shi, S.; Zhou, Y.; Wang, Q.; Dong, M.; Chen, S. Object tracking on event cameras with offline–online learning. *CAAI Trans. Intell. Technol.* **2020**, *5*, 165–171. [CrossRef]
91. Jalal, A.; Kamal, S.; Kim, D. Shape and motion features approach for activity tracking and recognition from kinect video camera. In Proceedings of the 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, Gwangju, Korea, 24–27 March 2015; pp. 445–450.
92. Keshtegar, B.; Nehdi, M.L. Machine learning model for dynamical response of nano-composite pipe conveying fluid under seismic loading. *Int. J. Hydromechatron.* **2020**, *3*, 38–50. [CrossRef]
93. Antonini, G.; Martinez, S.V.; Bierlaire, M.; Thiran, J.P. Behavioral priors for detection and tracking of pedestrians in video sequences. *Int. J. Comput. Vis.* **2006**, *69*, 159–180. [CrossRef]
94. Choudri, S.; Ferryman, J.M.; Badii, A. Robust background model for pixel based people counting using a single uncalibrated camera. In Proceedings of the 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Snowbird, UT, USA, 7–9 December 2009; pp. 1–8.
95. Chen, H.W.; McGurr, M. Improved color and intensity patch segmentation for human full-body and body-parts detection and tracking. In Proceedings of the 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Seoul, Korea, 26–29 August 2014; pp. 361–368.
96. García, J.; Gardel, A.; Bravo, I.; Lázaro, J.L.; Martínez, M.; Rodríguez, D. Directional people counter based on head tracking. *IEEE Trans. Ind. Electron.* **2012**, *60*, 3991–4000. [CrossRef]
97. Vinod, M.; Sravanthi, T.; Reddy, B. An adaptive algorithm for object tracking and counting. *Int. J. Eng. Innov. Technol.* **2012**, *2*, 64–69.

98. Liu, G.; Liu, S.; Muhammad, K.; Sangaiah, A.K.; Doctor, F. Object tracking in vary lighting conditions for fog based intelligent surveillance of public spaces. *IEEE Access* **2018**, *6*, 29283–29296. [CrossRef]
99. Ristani, E.; Tomasi, C. Features for multi-target multi-camera tracking and re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6036–6046.
100. Xu, H.; Lv, P.; Meng, L. A people counting system based on head-shoulder detection and tracking in surveillance video. In Proceedings of the 2010 International Conference on Computer Design and Applications, Qinhuangdao, China, 25–27 June 2010; Volume 1, pp. 394–398.
101. Kabbai, L.; Abdellaoui, M.; Douik, A. Image classification by combining local and global features. *Vis. Comput.* **2019**, *35*, 679–693. [CrossRef]
102. Lisin, D.A.; Mattar, M.A.; Blaschko, M.B.; Learned-Miller, E.G.; Benfield, M.C. Combining local and global image features for object class recognition. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops, San Diego, CA, USA, 21–23 September 2005; p. 47.
103. Pirsiavash, H.; Ramanan, D.; Fowlkes, C.C. Globally-optimal greedy algorithms for tracking a variable number of objects. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 1201–1208.
104. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417.
105. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
106. Führ, G.; Jung, C.R. Camera self-calibration based on nonlinear optimization and applications in surveillance systems. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *27*, 1132–1142. [CrossRef]
107. Gandomi, A.H.; Yang, X.S.; Alavi, A.H.; Talatahari, S. Bat algorithm for constrained optimization tasks. *Neural Comput. Appl.* **2013**, *22*, 1239–1255. [CrossRef]
108. Murlidhar, B.R.; Sinha, R.K.; Mohamad, E.T.; Sonkar, R.; Khorami, M. The effects of particle swarm optimisation and genetic algorithm on ANN results in predicting pile bearing capacity. *Int. J. Hydromechatron.* **2020**, *3*, 69–87. [CrossRef]
109. Shahgoli, A.F.; Zandi, Y.; Heirati, A.; Khorami, M.; Mehrabi, P.; Petkovic, D. Optimisation of propylene conversion response by neuro-fuzzy approach. *Int. J. Hydromechatron.* **2020**, *3*, 228–237. [CrossRef]
110. Yang, X.S. A new metaheuristic bat-inspired algorithm. In *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 65–74.
111. Alahi, A.; Jacques, L.; Boursier, Y.; Vandergheynst, P. Sparsity-driven people localization algorithm: Evaluation in crowded scenes environments. In Proceedings of the 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Snowbird, UT, USA, 7–9 December 2009; pp. 1–8.
112. Berclaz, J.; Fleuret, F.; Fua, P. Multiple object tracking using flow linear programming. In Proceedings of the 2009 Twelfth IEEE international workshop on performance evaluation of tracking and surveillance, Snowbird, UT, USA, 7–9 December 2009; pp. 1–8.
113. Neiswanger, W.; Wood, F.; Xing, E. The dependent Dirichlet process mixture of objects for detection-free tracking and object modeling. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, PMLR, Reykjavik, Iceland, 22–25 April 2014; Volume 33, pp. 660–668.
114. Conte, D.; Foggia, P.; Percannella, G.; Vento, M. Performance evaluation of a people tracking system on pets2009 database. In Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, Boston, MA, USA, 29 August–1 September 2010; pp. 119–126.