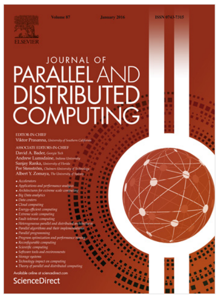
Accepted Manuscript



An intelligent decision computing paradigm for crowd monitoring in the smart city

Santosh Kumar, Deepanwita Datta, Sanjay Kumar Singh, Arun Kumar Sangaiah

PII: S0743-7315(17)30082-5
DOI: http://dx.doi.org/10.1016/j.jpdc.2017.03.002
Reference: YJPDC 3644

To appear in: *J. Parallel Distrib. Comput.*

Received date : 4 December 2016
Revised date : 7 February 2017
Accepted date : 1 March 2017

An Intelligent Decision Computing Paradigm for Crowd Monitoring in the Smart City

Santosh Kumar[1], Deepanwita Datta[2], Sanjay Kumar Singh[3], Arun Kumar Sangaiah[4]

[1, 2, 3]Department of Computer Science and Engineering, Indian Institute of Technology (Banaras Hindu University)
Varanasi-221005, India
[4]School of Computing Science and Engineering, VIT University, Vellore-632014, Tamil Nadu, India
[1]santosh.rs.cse12@iitbhu.ac.in, [2]ddatta.rs.cse13@iitbhu.ac.in, [3]sks.cse@iitbhu.ac.in,
[4]arunkumarsangaiah@gmail.com

Abstract

The ever-expanding urbanization and the advent of smart cities need better crowd management and security surveillance systems. Advanced systems are required to improve and automate the crowd management system. The aim of the closed circuit television and visual monitoring systems using multiple cameras faces many challenges like illumination variance, occlusion and small spatial-temporal resolution, person in sleep, shadows, dynamic backgrounds, and noises. Therefore, the crowd monitoring, prevention of stampedes and crowd-related emergencies in the smart cities are major challenging problems. In this paper, we propose an intelligent decision computing based paradigm for crowd monitoring in the smart city. In the intelligent computing based framework, the optimization algorithm is applied to compute the feature of crowd motion and measure the correlation between agents based motion model and the crowd data using extended Kalman filtering approach and KL-divergence technique. The proposed framework measures the correlation measure based on extracted novel distinctive feature, and holistic feature of crowd data represent and to classify the crowd motion of individual. Our experimental results demonstrate that the proposed approach yields 96.20% average precision in classifying real-world highly dense crowd scenes.

Keywords: Crowd motion, Crowd monitoring, Computer vision, Smart city, SIFT, Agent Motion Model, K-NN, KL-divergence.

1. Introduction

Crowd management in the smart city is getting more proliferation due to its widespread of application and usage recently. Pedestrian crowd are the essential part of smart cities. To provide better solutions and services in the smart cities, crowd monitoring, planning, and crowd management are necessary [1]. Therefore, various mathematical simulations, theoretically based models, and efficient simulation tools as well as various intelligent support systems using computer vision approaches, pattern recognition and image processing for the crowd management plays a vital role in monitoring and tracking of crowd motion [2]. In the smart cities, different surveillance systems are deployed to monitor the various activities and control and monitor the traffic in the several crowds at several places, such as shopping mall, traffic signals, roads, railways, and airport platforms, etc. The control and monitoring of crowd are important task and major changeling problems in the smart cities recently throughout the world [3].

Nowadays, crowd management and advance visual surveillance system of moving crowds in the smart cities have achieved much attention by scientific interests, research communities, and multidisciplinary researchers to improve crowd planning, management and monitoring of crowd and crowd safety in public facilities at the different places in the smart cities [3] [4]. Deep perspicacity into crowd dynamics and its management have contributed a better platform to several researchers and crowd control researcher groups for the understanding of individual behaviour, tracking of crowd motions in the smart cities. Based on the current approaches crowd consists of discrete individuals able to react with their surroundings [5].

In current years, numerous multidisciplinary researchers, scientists, and research committees have begun to understand and study the management of crowd by detection and tracking of individual behaviours in the crowd [7] [8]. The crowd management system monitors the crowd of smart cities by tracking and identifies the person in the crowds which is already started to prevent the catastrophic events. Therefore, it needs to design and develop automatic crowd management systems and better prediction of the crowded traffic flow in the smart cities across the world. The study of crowd motion is an NP-hard research problem in the computer vision because the individual trajectories of crowd movement are tough to predict the exact individual motions and different complex motions of individual are involved in it. The problem of crowd analysis becomes complicated due to changing nature of crowd density and background scenarios. In computer vision, some of the previous works have been done for the management of crowds as event recognition and anomaly detection [1] [2] [4] [5] [8] [9] in the smart city. However, these monitoring approaches failed to perform the correct prognostication and matching of events with corresponding stored crowd datasets. Other related works based on learning models are applied for evaluating of interactions amongst a small number of persons to analysis the action [7] [10].

Similarly, the author proposed a system to distinguish the anomaly behaviours of individual in the dense crowd video in the paper [10]. However, the proposed approach was inadequate to detect the anomaly behaviours for classifying the individual. The traditional crowd management systems and tracking based crowd management system fail to provide a competent level of monitoring and tracking of individual in the crowds [11]. These approaches are also applied to solve the significant problems, such as pedestrian tracking, understanding, anomaly behaviour detection, and unattended baggage detection using computer vision methods [12]. However, these classical methods did not provide the satisfactory solutions to solve these major problems of crowd and management of groups in the smart cities [11] [13].

The development of various crowd monitoring approaches and machine learning models can be applied to track the trajectory of crowd motions of individual and this process provides a controlling mechanism of crowd. However, these methods are inadequate to perform the computation of extracted crowd motions for crowd surveillance in the crowds, because these approaches are not suitable for crowd detection or management crowd under various constraint scenarios, such as massive flow of data, complex background, and noises.

The crowd monitoring methods and machine learning models require a lot of effort to eliminate these adverse consequences of the vision-related factors [12] [13] [14]. The involved significant factors are essentially noises, occlusions, alignment and transformation of captured crowds or scenes and learning the scene-specific based features for the training the crowd motion models. Therefore, it is the requirement to design and develop an automatic crowd management based system for better crowd planning, crowd monitoring, event tracking, and estimating the crowd flows [14].

In this paper, we focus on crowd management in the smart city by extracting the individual features as well as holistic features of the crowd motion for the detection and tracking of people in the dense crowds by motion detection in the crowd video or scenes. The primary objective is to learn the discriminatory features of crowd-motions from the crowd trajectories. In this paper, it is done using agent-based motion model. The model extracts discriminatory features from the captured crowd trajectory for better representation of features. The proposed system caters the robust solutions for crowd management [15].

We have considered the trajectories of crowd motions of the individual in the given surveillance video or scenes. The primary motivation of own model is to classify the individual based on the extracted discriminatory features from the crowd video or scenes. For crowds tracking, we have applied the optical-flow-based methods to for localizing for track the individuals in the crowds. In this paper, proposed system extracts the holistic and individual features from the crowd video for the localization and detection of crowd people in crowd video. We have applied the Scale Invariant Feature Transform (SIFT) [42] [43] descriptor-based technique for the detection and localization of important key points in the crowd video or crowd scene in the smart cities. To bridge the gap between models and trajectories of crowd motions, we have also applied the iterative crowd optimization algorithm. The Kalman-filtering based technique [36] and KL-divergence approach [35] [46] are used to evaluate how well the models are learned by particular discriminant features of crowd motions at both the individual feature level and the holistic feature level. The objective of this learning mechanism to achieve the different state transitions of crowd-motions. We have applied KL-divergence technique [46] to determine the distances matrices between these two distributions (the state transition based Gaussian distribution and the crowd states based Gaussian distribution).

1.1 Major contributions of the research work:

1. In this paper, we propose a crowd management system for monitoring and tracking the crowds in the smart cites. The proposed system extracts the individual features and holistic features from the crowd data using multiple agent motion models based learning methods. These methods are applied for the extraction not only for global information but also for the individual information of crowd motion.
2. The localization and detection of crowd people based on individual and holistic discriminatory feature of crowd scene are done using scale invariant feature transform technique.
3. A better analysis of different Gaussian distributions of crowd state transition and sequence crowd states in a given time is done. The models are learned with particular discriminant features of crowd motion at both the individual feature level and the holistic feature level using Kullback–Leibler divergence (KL-divergence) technique [35] [46]between the crowd trajectories and any of agent motion models.
4. The proposed system is learned by a holistic feature and an individual characteristic of crowd video for tracking the individual. Based on the extracted features, proposed approach performs the classification of crowd motions by applying multi-label classification technique.

The rest of this paper is organized as follows. Section 2 summarizes the related works and background in the field of crowd management and its various approaches. Section 3 demonstrated the proposed system for the monitoring the crowd. It also illustrates how to evaluate the individual and holistic discriminatory features for the tracking the individuals in the crowd scenes. Section 4 presents the formulation of multi-classification of extracted features. The experimental results and discussion of proposed approach is illustrated. Finally, Section 6 briefly concludes this work.

2. Related work

In this section, a comprehensive review of crowd monitoring and tracking of individuals in the crowd video or crowd scenes are illustrated in detail. In the holistic feature based methods, crowd scenes are represented as a complete data for group analysis [1] [2] [3] [15] [16]. In [1], authors have proposed a method for the analysis of crowd video. The proposed method tracks the trajectory of crowd motion in the trajectory of moving targets to their nearby regions of the crowded crowd scenes using map based crowd monitoring approaches. The map-based crowd monitoring methods are efficient for the group analysis. In the map-based crowd analysis methods, geometrical features of crowd scene are extracted from the set of images (set of frames) of crowd surveillance video [3]. The detection of image foreground [4], [5], extraction of foreground edges features [6], extraction of foreground corner features [7, 8], and other image texture features [9] are necessary for crowd analysis [10], [11]. These features are highly correlated for detection and counting of crowd people. Under such scenarios, the map-based approach also caters fairly accurate human count estimate in the given massive videos. However, the performance of the applied map-based learning models is typical scene dependent, and the actual trajectory of crowd location of each is unavailable.

In the dense crowd scene, different objects are overlapped in the same region of interest due to occlusion problems. Therefore, these models fail to monitor the crowds in smart cities because the applied group models are unable to isolate the individuals in the crowd scene. The tracking of pedestrians is a well-studied problem in computer vision research. However, these approaches fail because they rely on either background modelling [1], [2], [3], detection of different body part [3], [4], or shape models of various body parts [5], [6] [7]. In the similar direction, Wang et al. [10] proposed a learning model for motion path. The model provides the technique to cluster the different crowd's scene into different clusters for semantic region detection. The model provides a way to investigate and analysis the crowd scene or video based on the movement patterns of individuals′ interactions. But the learned models are not able to generalizable for more occluded parts of different crowd scenes. Therefore, crowd tracking models have not performed the better analysis of highly dense crowd in the more complex motion and trajectory in the crowd video.

For the estimation of crowd density, particle optical flow [8] and tracklet approaches are applied to analyse the high-density of crowd video or crowd scenes [9] [10] [11] [12]. The particle optical flow based methods are applied to calculate the pixel-wise motion of individuals between consecutive video frames [13] [14] [15].

The crowd detection-based techniques provide a method to count the number of crowd members in the group video or scene by identifying each human target. In this direction, research has been emphasized on human detection [12, 13], and localization of different body parts and detection of crowd member [14], [15], [16], [17]. Furthermore, these methods applied for joint consideration of detection and tracking of crowd people in the given crowd video [15, 18, 19]. These approaches provide more accurate detection and counting on lightly crowded scenarios. Since the location of each is also available, local crowd density measurement can also be computed. The major challenges of these crowd analysis techniques are higher computational cost, viewpoint dependent based training the crowd analysis model for learning the purpose and massive crowd data are also required (larger image size) to cater sufficient pixel intensity values on target for detection. Mahadevan et al. [16] proposed a framework using texture feature extraction technique for detection of the anomaly in the crowd scenes. For the detection of abnormality, the proposed system utilized the low-level visual features of crowd scenes. However, the proposed framework based systems are unable to perform the monitoring and detection of the anomaly in the massive crowd video.

In the similar direction, the authors [17] proposed a system for crowd analysis based on the extraction of features using the dynamic texture feature techniques. Dynamic texture techniques are a spatiotemporal generative based model for crowd video. The proposed system represents the extracted features for identification of different objects using video sequences as observations. These observations are performed using learned hidden Markov model from a linear dynamic system. The proposed system achieved better performances in the crowd motion segmentation over the traditional crowd representations methods based on the optical flow and other localized motion based representations the mixture of dynamic textures achieves superior performance in the problems of clustering and segmenting video of such processes. In this paper, one of the major shortcomings is that proposed system has not provided methods to estimate crowd density using non-iterative subspace techniques efficiently.

In [18], the author proposed a method using particle-based approach. The proposed method explicitly utilizes the interactions among crowd members and crowd agents. The tracklet-based crowd analysis methods are efficient for high-density video using tracklets generated by KLT trackers.

The classical crowd representation technique such as optical flow method is applied to propose a crowd monitoring based framework system [19]. The authors have utilized the using multiple instance learning based models to avoids these problems of tracking a crowded scene in a surveillance video and extract the crowd features using discriminative classifiers for classification of motion patterns in crowd scenes [19]. The proposed approach results in the most robust tracking of individual in the crowd motion. However, the proposed system is unable to completely avoid the types of major problems that adaptive appearance-based trackers suffer from. The proposed system is also unable to track the object if it is completely occluded for a long period of time. In [19], the authors proposed a system for understanding collective the crowd behaviours of individual using learning a mixture model of dynamic pedestrian-Agents. The proposed system can model the beliefs of pedestrians and the missing states of observations. The system can able to learn well from massive crowd trajectories caused by frequent tracking failures. Therefore the system fails to perform basic simulate and predict techniques for analysis of collective crowd behaviours. In [26], the authors proposed a group monitoring system to learn the nonlinear motion patterns to assist the association of tracks from detection responses [20] [21].

In the similar direction, Zhou et al. [22] proposed a mixture model-based system to learn the motion patterns from the crowd scenes. After that, they performed the prediction of the anomaly behaviours of pedestrians in the group videos. The enhanced model can be used to cross-scene crowd video retrieval. But the proposed crowd systems could not be well quantitatively measured the extracted features. In [23], the authors applied the floor-fields based approach to generate the probability distribution moving individual in the crowd video. The proposed method used the segmentation technique to find the regions in the extracted video frames. Antonini et al. [24] introduced a method to discretize the speed, space of a pedestrian into many areas and models the probability of determining velocities for tracking of individuals in the crowd video or crowd scene. While [26] learns the nonlinear motion patterns to assist the association of trackless of detection responses. Associated work is trackers using agent-based motion models. A trajectory-based crowd motion system utilizes a crowd data as a collection of individuals and models the interactions among the people. They use the machine learning approaches to analyze the complete motion trajectories in the given crowd video [7] [6] [10], [11]. Group tracking techniques apply the complex motion models and learning strategy to improve monitoring of crowd member in the huge crowd video [1], [20] [23], [24], [25] [30] [35].

In [26], the authors proposed a hierarchical activity-prediction model to recognize and classify the interaction between individual trajectories and cooperative activities. Bazzani et al. [27] proposed a framework to analysis the joint individual-crowd state space in two dependent subspaces. The major shortcomings of this proposed framework is that it fails to provide a complete analysis of crowd state space based on individual features. Benfold et al. [28] proposed a system to tack the multi-target in real-time surveillance video. The system extracts the Histogram of gradient features from crowd video using simultaneous KLT tracking technique and Markov-Chain Monte-Carlo Data Association (MCM-CDA) approached to cater a real-time tracking of crowd people in video. The system achieved 82.0% recognition accuracy to identify individual in the crowd video. The major short coming of proposed system is that it fails to monitor and identify individual in occluded scenes or crowd video.

3. Proposed system

In this section, we discuss our proposed crowd management system. Fig.1 depicts the workflow. The individual and global features are extracted from the crowd video frames (set of images), and these features are classified the crowd trajectory, shown in Fig. 1 (as shown in red and gray colours).

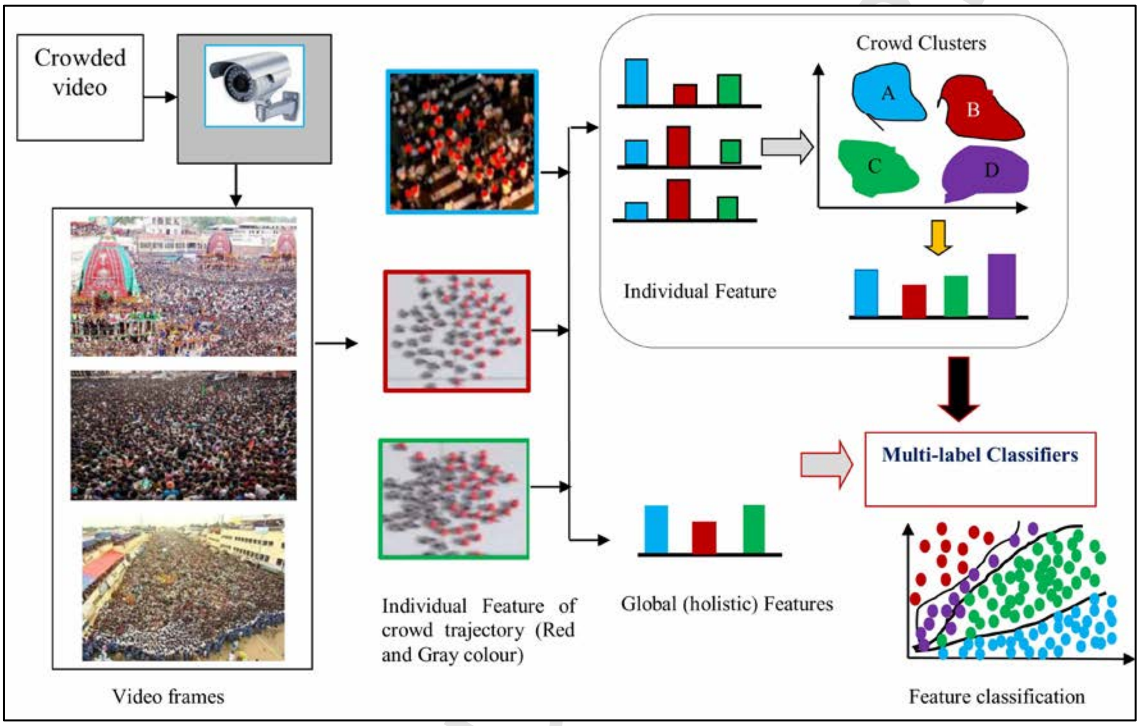


Fig. 1: Proposed Crowd Monitoring System

The trajectories are marked in red dots and gray dots. The red or gray colour depicts the identified and the red/gray colour reflects the tracked crowd people in the crowd video. Multiple agent-based motion learning models are also applied. The proposed crowd management system is composed of different steps. These steps are illustrated briefly as follows:

3.1 Acquisition of crowd scene database

In the acquisition phase, surveillance cameras are applied to capture the crowd video. The set of video frames is extracted from the captured crowd scenes for the pre-processing and enhancement of extracted video frames or images. After that, salient sets of discriminatory features are extracted from the pre-processed crowd images.

For training and testing, database of crowd video is prepared with surveillance cameras using 60 mm×120 mm lens from the biggest festival of puri rath yatra, in the Orissa state, India. To monitor the crowd during rath yatra, more than 500 surveillance cameras have been deployed at various locations or temple. The captured crowd videos contain considerable variations in term of view angle, scale, crowd motions, trajectory and density of crowd videos. The deployed surveillance cameras kept static in all the video. We have chosen 5000crowd scenes(e.g., video frames set of image sequence) from the captured crowd video. The size of crowd database consists of 5000 crowd scenes (video frames) with counts ranging between 120 and4543 with an average of 1440 individuals per image. Muchlike the range of counts, the scenes in these images also belong to a diverse set of events: concerts, protests, stadiums, marathons, and pilgrimages. The training and testing sample size of captured crowd database is 500×500 pixels. In this paper, two prominent features, such as individual feature and holistic features are extracted from the images for the analysis and monitoring. The captured crowd images are stored in the crowd database for the matching and classification of crowd motions.

3.2 Pre-processing and Enhancement

Pre-processing and enhancement are fundamental steps of feature extraction and matching. The video frames (e.g., set of crowd image sequences) are extracted from the captured video. The extracted sets of video frames are applied for the pre-processing and enhancement of crowd images. The quality of the captured videos (data) by the surveillance camera is initially assessed in order to determine its better quality and suitability of extracted set of video frames of crowd scenes for further processing. Generally, crowd data are captured using surveillance camera in the unconstrained environments, such as low-illumination, highly occluded, poor image quality, and image blurriness. The captured set of video frames can be more defective, blurred, and of poor quality and low contrast [49]. Therefore, the acquired sets of image of crowd database are subjected to pre-processing and enhancement algorithms in order to improve the quality of crowd database. The extracted video frames of crowd database is processed and enhanced by using Contrast Limited Adaptive Histogram Equalization (CLAHE) image processing technique [29] [51] [52]. The contrast limited adaptive histogram equalization technique enhances the contrast and quality of the gray scale crowd scenes or video frames (set of image sequence) by transforming the pixel intensity values by using contrast-limited adaptive histogram equalization (CLAHE) [49] [50] [51] technique. The pre-processed and enhanced image of crowd database is shown in Fig. 3.

3.3 Feature Extraction and Matching

Efficient features are crucial for encoding the given crowd scenes patterns into the set of measurable discriminatory information. In this section, we adopt the agent motion-based learning model (AMM) [30] for the learning the extracted individual and global (holistic) features of crowd motion. The AMM learning models [30] play a vital role for the tracking anyone in crowd scenes. The movement of the crowd is very complex by nature. It is hard to directly perform the comparison of feature vectors of group movements with stored crowd scenes based on trajectories of individuals [41].

The primary objective of own proposed approach is to measure the similarity scores among the extracted discriminatory features. Usage of AMM motion models [30] provides us a better learning model to capture the crowd motions. For in-depth level analysis, we use different types of crowd motion features and their interactions in crowd scenes. Although in the current literature, different systems or models are available. AMM learning based techniques [30] [41] are good at modelling the crowd motion and analysis of the collective behaviour of individual in the crowd scenes [41]. Also AMM based learning model avoids several collisions in the crowd video [11], [12].

For each AMM based learning model, we have computed an entropy descriptor to measure the similarity between the given (test) crowd motion data and stored crowd images. Also, based on the correlation measure of features, the particular features in combination with the holistic feature are utilized to describe group motions. For calculation of entropy information, entropy descriptor is used. All the entropy descriptors are calculated from the multiple AMM learning models. At last, all computed, discriminatory entropy information is fused. The entropy information of crowd motions is shown in Fig. 1.

3.4 Localization and detection of crowds

In this subsection, we have demonstrated the detection and the tracking of trajectory path of individuals in the crowd scene or video. We have applied the scale invariant feature transforms (SIFT) [43] feature descriptor for the localization and detection of key points in the crowd scenes [31].

Fig. 2: Illustrates (a) normal crowd scene and (b) presents the localization and detection of SIFT key-point features of crowd scene.

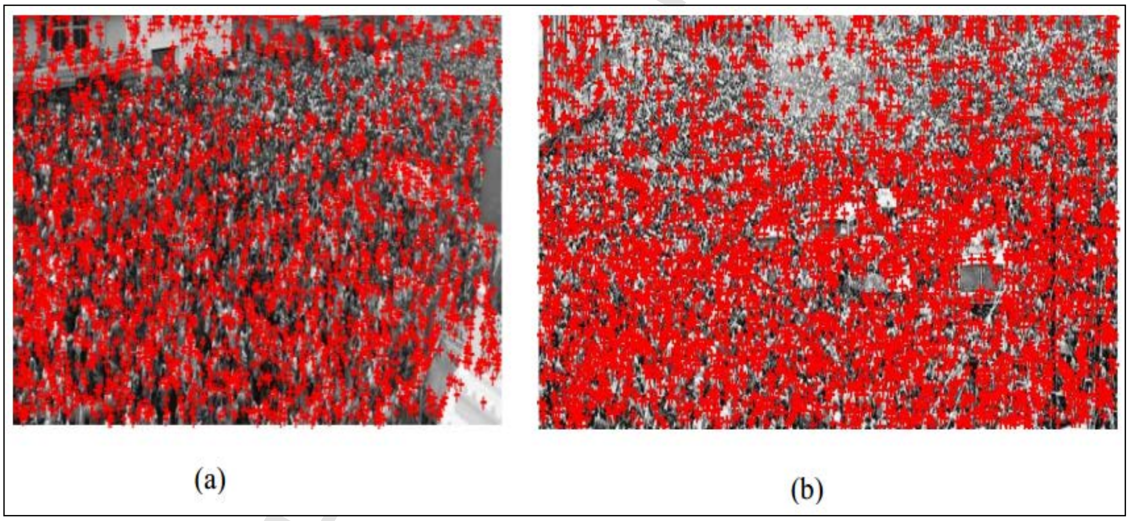


Fig 3: shows (a) low occluded crowd scene and (b) illustrates the localization of individuals in the highly dense crowd scenes for the computation of key-points features using SIFT descriptor based technique.

The SIFT detector [42] extracts the key points from the crowd images from the crowd video. The SIFT detector [42] detects the similar consistent with (some) variations of the illumination, viewpoint, and other viewing conditions. The localization and detection of crowd key-point features are shown in Figs. 2 and 3, respectively. The similarity matching scores are calculated based detected key-points using SIFT feature descriptor technique and calculated matching scores (%) are shown in Table 1 and Table 2, respectively.

Table 1: Illustrates the matching scores of low occluded and highly dense crowd scene from the crowd database using SIFT descriptor based technique.

S. N.	Crowd images	Detected key-points	Matched key-points	Matching scores
1		1021	766	75%
2		882	705	84%
3		989	860	87%
4		1345	1278	95%

Table 2: Illustrates the detection and similarity matching scores of highly occluded and highly crowded scenes from the crowd database using SIFT descriptor based technique.

Type of crowded scenes	Crowd images	Detected key-points	Matched key-points	Matching scores
Crowded scene		684	474	69.29%
Highly occluded crowd scene		882	745	84.46%
Highly occluded and dense crowded scene		754	560	74.27%

Highly occluded and highly crowded scene		685	498	72.70%

3.4 Agent Motion Model Based Learning of Crowd Data

In the last few decades, machine learning based agent motion modelling techniques (AMM) have been proposed for the analysis of crowd motion and interactions in computer vision. We have utilized multiple AMMs to extract individual and as well as motion features for representing crowd from the captured trajectories.

AMM model combines the social force patterns [32], heuristic models [33] and geometry modelling based algorithms [33, 34] for representing the crowds. Moreover, these learning models also applied to analyze the spatial location of each crowd people or the agent in the smart cities. This model locates the crowd agents by circles in $(2-D)$ feature space. Various parameters, such as number of the nearest neighbours, maximum speed, and radius, are required for training purpose.

The AMM is formulated as a non-linear function (f) which is controlled by a parameter (β). The function estimates the crowds at the next time step $(t+1)$ from the current crowd state (X_t). Let the computation error is (B_t) then the prediction of the next step is given in Equations (1-3) as follows:

$$X_t = f(X_t) \tag{1}$$

$$X_{t+1} = f(X_{t+1}) \forall \mathrm{t} \in (1, 2, 3...., \mathrm{N}) \tag{2}$$

$$X_{t+1} = f(X_t) + \mathrm{B}_t \tag{3}$$

Algorithm 1: pre-processing and enhancement of crowd database

1. **Input**: set of video frames $F = (F_1, F_2, ..., F_N)$ of captured crowd video, N is total number of crowd video frames.
2. Set of video frames is taken as samples from stored crowd database.
3. Pre-processing of extracted frames transforming the colour image into gray scale images is done.
4. Cropping and resizing of the extracted set of video frames (F) into 500×500pixels.
5. Pixel intensity values of cropped and resized images are transferred into gray scale image using contrast-limited adaptive histogram equalization (CLAHE) [49] [50] [51] technique.
6. Apply SIFT descriptor technique for detection and extraction the discriminatory set of features from enhanced images as shown in Fig. 2 and Fig. 3.
7. **Result:** perform the similarity matching of detected key-points of test crowd scene with stored database of crowd scene $(F_1, F_2, ..., F_N)$

4. Computation of crowd using AMM model

As mentioned in the previous section, crowd motion analysis is an NP-hard research problem. Therefore, learning an individual, robust model for recognizing crowd scenes is tough. To perform this task, we utilize the multiple agent motion model based frameworks to leverage the various agent motions to produce a robust crowd.

In this section, we introduce our algorithm to measure the differences between an AMM and the trajectories. After pre-processing, the extracted feature vectors are transformed and saved individually as well as a holistic set.

Next, the query set of images are matched with stored templates of the crowd database and the correlation between then is calculated. The similarity scores are generated from that matched with stored database. Then scores are compared with multiple trained AMMs learning models by iteratively and optimized algorithm. The AMMs learning model is used to describe the crowd movement in as efficient fashion. Mainly, the iterative optimization algorithm is applied to mitigate specific artefacts and to filter the unknown noises and remove different types of objects from the crowd images. The similarity scores of AMM models [35] are calculated to generate a crowd motion feature [30]. Along with the global information, this method also captures the individual information which is used for in depth analysis later.

4.1 States declaration and AMMs

Let us say for an AMM learning model, a given moment is (t) and present state is (x_t). The non-linear function of AMM is (f) that propagates present state (x_t) to the next time step (x_{t+1}) as in (Equation 4):

$$x_{t+1} = f(x_t) \qquad (4)$$

4.1.1 Computation of velocity of trajectory motion

The captured crowd motion based trajectories hold only the temporal-spatial information. Let us say a set of feature is describes as a sequence of two dimension $(2-D)$ vectors $(V_1, V_2, ...V_N)$. It gives a clamorous projection of the actual (true) crowd state $\mathrm{T} \in (T_1, T_2, ..., T_N)$ at the similar time step: $Z_t = H(\mathrm{x}_t) + S_t$ where H projects the crowd state $Z_t \in R^{6N}$ to $Z_t \in R^{2N}$ only to keep the position information. Here (S_t) represents the observation noise and $t \in (1, 2, 3, ..., N)$ represents the time interval. We assume that it is follows the zero-mean Gaussian distributions.

4.1.2 State transition and distributions

In this subsection, we have illustrated the statistical analysis of extracted features (individual feature and holistic features) from the crowd motion. For a given an AMM learning model, we first consider the prediction of extracted features that are subjects to Gaussian distributions (as shown in Equation 5):

$$R_t = N(f(x_{t-1}), \sum\nolimits_t^{\wedge}) \ (5)$$

Where $(\sum\nolimits_t^{\wedge})$ indicates the scale of the Gaussian distribution at time step (t) .By the measuring scale of (R), we quantify the similarity between the given AMM and the data. A larger scale implies a larger divergence between the AMM and the data.

4.1.3 Computation and distribution of crowd states

As we mentioned in above section, the real crowd states $T \in (T_1, T_2, ..., T_N)$ are not known in the crowd motion the observed data is often compounded with noise [47]. To properly model the true crowd states, we describe the crowd states as computed distributions, rather than some fixed values. Therefore, the difference between any AMM and the crowd states is equal to estimating the distance between the state transition distributions and the crowd state distributions [48] [53].

In this paper, we applied the Kullback–Leibler divergence (KL-divergence) technique [35] [46] to estimate the difference between two given probability distributions [35]. Particularly, we formulate the KL-divergence approach for the two Gaussian distributions [35]. These distributions are (1) the distributions of state transition $R_t \in (t \in (1,2,3,.....N))$ and (2) the distributions of estimated crowd states $P_t \in (t \in (1,2,3,.....N))$. Our goal is to find the minimum value of their KL-divergence. So, our objective function is formulated as follows (shown in Equation 6):

$$X_{1:P1}, \mathrm{R}_{1:\mathrm{P}1} = \operatorname{argmin} \sum_{X_{1:}:\mathrm{P}_1, \mathrm{R}_1:P_1} DivergenceKL(X_{1:P_1}, \mathrm{R}_{1:\mathrm{P}_1}) \tag{6}$$

To simplify the problem, we assume that the scales of all state transition distributions are the same at any time, which means that the prediction capability of (f) always remains stable in the same crowd motion. In this paper, primary objective is to compute the minimum value of Kullback–Leibler divergence (KL-divergence). Therefore, the objective function is defined as follows (Eqs. (7)- (9)):

$$X_{1:P1}, R_{1:P1} = \arg\min \mathrm{D}_{KL}(X_{1:P1}, R_{1:P1}) = \arg\min \mathrm{D}_{KL}^{t}(X_{1:P1}, R_{1:P1}) \tag{7}$$

Resultantly, all the covariance matrices $(\sum_t)$ of $((R_{1:P1}), P_t \in (t \in (1,2,3,.....N)))$ are equal, i.e. $(\sum_t)=(\sum_0)$, where $(\sum_0)$ is a defined as specific co-variance matrix of crowds. Furthermore, the computation of KL-divergence for two probability distributions $(X_1 : P_1,)$ and $(R_1 : P_1)$ are elaborated given in Equations ((8)-(9)).

$$D_{KL}^{t} = \frac{1}{2}\left[trce(\sum\nolimits_t^{-1} \sum\nolimits_t) + (f(x_{t-1}) - x_t)^{\mathrm{T}} \sum\nolimits_t^{-1} (f(x_{t-1}) - x_t) + \ln \frac{\left| \hat{\sum_t}(x_t) \right|}{\left| \sum_t (x_t) \right|} \right] + \text{costant} \tag{8}$$

$$= \frac{1}{2}\left[trce(\sum\nolimits_0^{'-1} \sum\nolimits_t) + (f(x_{t-1}) - x_t)^{\mathrm{T}} \sum\nolimits_0^{'-1} (f(x_{t-1}) - x_t) + \ln \frac{\left| \hat{\sum_t}(x_t) \right|}{\left| \sum_t (x_t) \right|} \right] + \text{costant}$$

In the above Eq. (8), $\left|\hat{\sum}\right|$ is the determinant of covariance matrix $\left(\sum\right)$ of crowd database.

Hence, the objective function in Eq. (7) is expressed as follows:

$$X_{1:P1}, R_{1:P1} = \arg\min \sum_{t=1}^{T} tr(\sum\nolimits_0^{-1} \sum\nolimits_t) + (f(x_{t-1}) - \mathrm{x_t})^T \sum\nolimits_t^{-1} (f(x_{t-1}) - \mathrm{x_t}) + \ln \frac{(\sum_0)}{(\sum_t)} \tag{9}$$

5. Computation of Entropy Descriptor

In this section, entropy descriptor is used to compute the discriminatory features from the crowd video or crowd scenes. To introduce the entropy descriptor, we first assume that the prediction error $\left(B_t\right)$ in Eq. 7 is the subject to a zero-mean Gaussian $D = N\left(0, \sum\right)$. Measuring the scale of $\left(D\right) = N\left(0, \sum\right)$, we can quantify the similarity between the given AMM learning model and the data. A larger scale implies a larger divergence between the AMM and the data. As mentioned above, the noisy observations $\left(B_t\right)(t \in \left(1, 2, 3, \ldots \mathrm{T}\right))$ contain the observed positions of crowd member in crowd scene only. But the true crowd state $\left(x_t\right)$ (which contains all positions, velocities, and desired velocities of people moment) is often uncertain.

To address this uncertainty, crowd state is consisted as a Gaussian distribution $X_t = N(x_t, .), (\mathrm{t} \in \{1, 2, 3 \ldots . \mathrm{T}\})$ in this paper. With regard to the prediction error $\left(\sum\right)$ and crowd motion data, primary objective is to maximize the log-likelihood of $\left(D\right)$ with a given AMM learning model:

$$\left[X_t, \sum\right] = \arg\max_{x_t, \Sigma} \text{ loglikehood(D)} \quad (10)$$

5.1 Optimization of Entropy Descriptor

We have performed the optimization of objective function of Equation10. The objective function of (Eq. 10) has some decisive variables. The decisive variables are prediction error $\left(\sum\right)$ and the crowd state distribution $\left(X_t\right)$. We adopt an optimization strategy similar to the Expectation Maximization (EM) algorithm [53]. We first optimize $\left(X_t\right)$ based on observations and then maximize the likelihood of $\left(\sum\right)$. We have applied the expectation maximization (EM) algorithm to optimize the objective function which is given in Eq. (9) with defined variables $\left[\left[X_{1:P1}\right], \left[R_{1:P1}\right]\right]$. We estimate the optimal value of two variables by iteratively performing steps (1) and (2).

- **Step 1: Optimize $\left(X_t\right)$ based on co-variance matrix $\left(\sum\right)$ and known $\left(\mathbf{D}\right)$:** The trajectories of crowd motions are often consisted of an unknown noises at different scales. For that, we need to estimate the crowd state distribution. This problem is solved by the Kalman-filter based technique [36] $\left(K_T\right)$ where the crowd trajectories are given as observation $\left[R_{1:P1}\right]$. The crowd trajectory data serve as the state transition information to train the proposed system and agent motion models [36] [37].

- We have also applied the extended Kalman smoother filtering technique [36] to provide the smoothness of trajectory path of crowd motions for every of individuals in the crowd. It optimizes and smoothes the discrete temporal states for non-linear models.
- Further, transition states of crowd motion are considered, and it is subject to Gaussian distribution. Specifically, we have applied the extended Kalman smoother based filter technique [36] to mitigate the redundant data by the filtration and smoothing processes. The filtering process is further divided into phases, such as prediction phase and phase correction for better computation of features.
- **In the predictor phase:** It is stated earlier that agent-motion based learning model are trained using the extracted individual feature as well as holistic features of crowd trajectory for the better prediction of the test (query) crowd scene. The working principle of the prediction is as follows (as shown in Equations (11) and (12)):

$$\bar{x} = f(x_{t-1}) \tag{11}$$

$$\sum_t = J_f(x_{t-1}) \times \sum_t J_t^T (x_{t-1}) + (\sum_t{}') \tag{12}$$

Where (x_t) refers to the predicted crowd state which is provided by using agent-based motion models (AMM) (f). $\left(\sum_t\right)$ is the covariance matrix associated with the crowd state distribution at each timestamp (t). $\left(\sum_t{}'\right)$ is the state transition covariance matrix. The term $J_f(.)$ refers to the Jacobean matrix of AMM learning model (f). $\left(\bar{x}_t\right)$ and $(\sum_t{}'')$ are the predicted state and covariance (shown in equation (13)).

In the corrector phase:

$$x_t \approx \bar{x}_t + K_T(Z_t - h(x_t)) \quad (13)$$

Where $K_T = \sum_t J_h^T(\bar{x}_t) Jh(\mathrm{x}_t) \sum_t J_h^T(\bar{x}_t)$ and $\sum_t (1 - K_T J_h(x_t))(\sum_t{}')$

- In the step (2), we have estimated the crowd state distribution for the computation of crowd trajectory. For state distribution, $\left[X_{1:T}\right] = (N(x_t \mid T, \sum_{t|T})_{1:T}$ is computed. After that, expectation maximization is applied for optimization of the probability distribution of state transitions for AMM based motion models [52]. The optimization algorithm caters a near-optimal solution. The optimal evaluation of crowd states is computed from the observed trajectories using the Extended Kalman smoother [36]. We later mitigate the KL-divergence [35] between the estimated crowd states and the AMM-predicted states. Instinctively, the objective (Eq. (9)) is to update continuously. We complete these two steps iteratively.

The expectation-maximization (EM) optimization technique [53] [54] is applied to the computation of crowd motion based features which are presented in Algorithm 2. In the expectation-maximization method, expectation step (E-step) performs the evaluation based on the Kalman filter [36] to compute 1) the expectations of the hidden state variables. After expectation step, maximization step (M-step) is used to calculate the maximum-likelihood values for each crowd motion component by averaging over all sequences of crowd states [54].

Algorithm 2: Expectation Maximization (EM) algorithm for optimization of extracted features

1. **Input:** video sequences $N \in (Y^{(\mathrm{i})})_{i=1}^{N}$, number of components (K), state variables (X_t), where, $X_t = N(x_t, .), (\mathrm{t} \in \{1,2,3....\mathrm{T}\})$, Gaussian distribution $(D) = N\left(0, \sum\right)$ of crowd data,
2. **Initialize:** $\left[X_t, \sum_t\right] = \arg\max_{X_t, \Sigma_t} \text{ loglikehood}(D) \text{ for } \mathrm{t} \in \{1,2,3....\mathrm{T}\}$
3. **Repeat**
4. **Expectation Step**
5. **for** $i = \{1,2,3,.......N\}$ and $j = \{1,2,3,.......N\}$ **do**
6. Compute the expectations equations ((14)-(16)) with the Kalman smoothing filtering technique on $(Y^{(\mathrm{i})})_{i=1}^{N}$ and (X_t)

$$X_{t|j} = E_{X^{(i)}|\mathrm{Y}^{(i)},\mathrm{Z}^{(i)}} P(X_t^{(i)}) \tag{14}$$

$$P_{t,t|j}^{(i)} = E_{X^{(i)}|\mathrm{Y}^{(i)},\mathrm{Z}^{(i)}=j} P(X_{t|t}^{(i)}) \tag{15}$$

$$P_{t,t-1|j}^{(i)} = E_{X^{(i)}|\mathrm{Y}^{(i)},\mathrm{Z}^{(i)}=j} P(X_{t,t-1}^{(i)}) \tag{16}$$

8. **end for**
9. Maximization Step
10. for $j = \{1,2,3,4,.....,K\}$ **do**
11. Compute aggregate expectations

$$\varphi_j = \sum_j Z_{i,j} \sum_{t=2}^{\tau} P(X_{t,t-1}^{(i)}) \tag{17}$$

$$Z_{i,j} = p(\mathrm{z}^{(\mathrm{i})} = \mathrm{j} \mid \mathrm{y}^{(\mathrm{i})}) = \frac{\alpha_j \, p(y^{(i)} \mid \mathrm{z}^{(\mathrm{i})} = \mathrm{j})}{\sum_{k=1}^{K} p(y^{(i)} \mid \mathrm{z}^{(\mathrm{i})} = k)} \tag{18}$$

12. Compute new parameters for next crowd states and co-variance matrix $\left[X_t, \sum_t\right]$

end for

13. **Output:** $\left[X_t, \sum_t\right]_{j=1}^{K}$

5.2 Computation of individual Entropy

In this subsection, we have applied the entropy descriptor for the computation of the entropy information of individual features of the crowd motion data. The computed entropy $\left(\left[\hat{\sum}_0\right]\right)$ is composed of the following individual entropy (shown in Equation (19)):

$$\hat{\sum}_0 = \begin{pmatrix} \sum^1, \sum^{1,2} \ldots, \sum^{1,N} \\ \sum^{1,2}, \sum^{1,2} \ldots, \sum^{1,N} \\ \vdots \quad \vdots \quad \vdots \\ \sum^{1,\mathrm{N}}, \sum^{2.N} \ldots, \sum^{N} \end{pmatrix}_{U \times V} \tag{19}$$

Where each $\left(\sum^i\right)$ $i \in (1,2,3,\ldots N)$, and $\left(\sum^{i,j}\right)$ $(i, j \in (1,2,3,\ldots N), \mathrm{i} \neq \mathrm{j})$ are a 4×4 feature matrix and N is a total number of crowd people in the smart city. In the above matrix, $\left[\hat{\sum}_0\right]$ contains the internal entropy information about how well an agent motion model (learning model) [52] fits the trajectories of crowd motion for better computation of features. In the above entropy feature matrix (Eq. 14), the diagonal sub-matrix $\left(\sum^i\right)$ illustrates how well the given AMM learning model [52]fits the motion of agent (i), while the non-diagonal sub-matrix $\left(\sum^{i,j}\right)$ presents the mutual impact between agents (i) and (j) respectively. We state to quantize action of the internal information to compute the individual correlation between the AMM [52] and the individual crowd motion. The individual correlation, denoted as the individual entropy (F), is formulated in three different ways. The CA, CB and CC are individual features calculated from the above entropy matrix (shown in Equations (20)-(22)).

- (CA) **:Using the diagonal sub-matrix** $\left(\sum^i\right)$**:**We have performed the computation of the diagonal sub-matrix $\left(\sum^i\right)$ as follows:

$$CA\left(\sum^i\right) = \sum_i^4 \sum_{j=1}^4 \left|a_{ij}\right| \tag{20}$$

In above equation (14-15) it can be observed that fitness between specific motions of individual and the defined AMM learning based model [42] . The smaller (CA) is, the better the AMM [52] fits into the trajectories.

- (CB)**: Using the non-diagonal sub-matrix on the** i^{th} **row**: we have extracted the individual entropy and computed as follows:

$$CB\left(\sum^{ij} j \neq i\right) = C_A\left(\frac{1}{N-1} \sum_{j \neq i} \sum^{ij}\right) \tag{21}$$

Equation (16) indicates the motion correlation for other crowd members. The larger (CB) is, the bigger the motion dissimilarity between crowd agents (i) and other crowd members.

- (CC): For the computation of features (CC), diagonal and non-diagonal sub-matrix values are combined (Equation 17)

$$CB\left(\sum^{i}, \sum^{ij}\right) = \left(C_A\left(\frac{1}{N}\left(\sum^{i} + \sum_{j} \sum^{ij}\right)\right)\right) \tag{22}$$

In practice, we usually choose (CB) as the individual entropy for discriminatory features.

5.3 Computation of holistic features from the crowd data

In this subsection, we illustrate the computation of holistic correlation between an agent-motion based models [52] and crowd trajectories using holistic features from the crowd scene. The computation of holistic entropy $F(Ent)$ of crowd motion is given as follows:

$$F(Ent) = \frac{1}{2}\log(2\pi e)\left|\sum_{0}\right|$$

5.4 Multi-label classification

Multi-label classification technique is machine learning approach [38]. In this approach, multiple target labels $Y = (Y_1, Y_2, Y_3, ...Y_N)$ are assigned to each instance of input data (X). The dynamic changes occurs due to low illumination, complex movement of an individual, and background in the crowd data, therefore, captured massive crowd data suffers from assignment of a single label crowd video [39] [40] [41].

In this paper, we treat the identification of the crowd motion problem as a multi-label classification problem. Specifically, let $X = [X_1, X_2, X_3, ..., X_N]$ be the feature vector of an instance (e.g., a crowd motion sequence) and (Y) be a finite set of labels (L) where $[L \in (1, 2, 3, ...Q)]$ .Given a training set $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), ...(X_m, Y_m)$, the primary objective in this research work is to find out output using a multi-label classifier $\left(H : X \rightarrow 2^{Y}\right)$.

The multi-label classifier is usually formed as a real-value function (G) $[G : X \times Y \rightarrow \mathbb{R}]$. The function $G(X, Y)$ is appliedas ranking function to measure the matching scores between extracted individual features. If ranking function $G(X_i, Y_i)$ is greater than $G(X_j, Y_j)$ $\left(G(X_i, Y_i) > G(X_j, Y_j)\right)$ then raking score is decided based on the rank scores such as $\mathrm{Rank}(X_i, Y_i) > Rank(X_j, Y_j)$. The multi-label classifiers (H) classify the different extracted feature vector using ranking values as follows:

$$H = \left(y \mid G(x_i, y) > threshold, y \in Y\right)$$

In this work, multi-label based K-Nearest Neighbour (K-NN) algorithm is applied to evaluate performance of our proposed approach based on individual and holistic features of crowd motion. The multi-label based K-Nearest Neighbour (K-NN) [38] approach identifies the K-nearest neighbours in training sets of extracted features for matching of each test instance feature [38]. Based on the neighbour information, maximum likelihood a posterior technique is applied to determine the label set of the test instances. The multi-label based

K-Nearest Neighbour (K-NN) technique [38] classifies the feature vector of crowd motion by a majority vote classification principle.

6. Experimental results and Discussion

In this section, we present the experimental results. We first introduce our prepared crowd dataset and how we choose the agent motion model based learning approaches to jointly compute the individual and holistic features from the input crowd video frames (e.g., set of crowd image sequence). Then, we evaluate the individual feature and the global (holistic) features for the tracking of individuals in crowded video or scene.

6.1 Dataset preparation and description

The database of crowd video is prepared with surveillance cameras using 60mm×120 mm lens from the biggest festival of Puri rath yatra, in the Orissa state, India. During the rath-yatra, millions of pilgrims are joined to pull the cart of Lord Jagaranath to celebrate the festival.

To monitor the crowd during rath yatra, more than 500 surveillance cameras have been deployed at various locations or temple. The crowd videos are captured. The crowded frames (e.g., set of image sequence) are extracted from the crowd video dataset. It includes an average of 1190 individuals per image, ranging from 94 to 4543 people. We obtained 63705 annotations in which the coordinates of each individual one present. Some sample images of the crowd are shown in Fig. 4 and 5, respectively.

Fig. 4: Some sample images from the crowd database

Fig. 5: Illustrates the occluded crowd images from prepared database

The crowd video dataset is applied to train the SIFT descriptor [42] [43] based interest point matching model [31]. For training of the cascade head detector, we have created our database in the following ways:

- By extracting video frame of people on railway stations, or walking on busy roads.
- By taking all positions of head varying from -90 degree to +90 degree [22].
- By using 6684 negative images including those of empty roads, malls, buildings, gardens and stadiums.

While our proposed approach is based on detection of people and analyzing crowd trajectories, we have applied multi-person trackers to capture the crowd trajectories from the videos. We manually cater the initial positions of each people. After that, classical tracker techniques are applied to track the positions of individual crowd people [31]. The position consists of full information of trajectories, which are used as input. Sometimes, the applied tracking models do not perform well because of occlusion problem (covering and non-covering problem).

The occlusion problem is overcome by reinitializing the tracker model if it drifts too far. Finally, we have applied perspective transformation technique to transform the labelled trajectories from the image-space to the ground-space. So that, crowd database contains 524 short crowd videos of puri rath-yatra. All of the group (crowd) video sequences are surveillance videos captured from over headed cameras. In these images sequences, the average crowd density is 1.10 persons/m^2. The maximum density of captured surveillance video is 4.50 persons/m^2.

6.2 Selection strategy of Agent-based Motion Models

In this paper, we have applied 20 agent-based motion based learning models for the training of extracted discriminatory individual features and holistic features by generating different motion parameters which create different motions. Further, we choose six parameters and tune them for the better analysis of crowd in the smart city. These tune parameters are applied for the prediction of crowd trajectory. The list of tuned parameters is (1) the maximum distance to succeed as neighboring agents, (2) the number of neighboring agents (3) the effective radius, (4) how long the agents take to react to an upcoming collision, (5) speed and (6) how smoothly agents change velocities. These parameters are chosen by multi-label classification models to predict and classify the extracted discriminatory features (e.g., CA, CB and CC features). We manually fix a threshold for each parameter, sample the key values and add small amount of white noise in this experiment.

6.3. Classification of crowd features

We compare the different individual features denoted as (e.g., CA, CB and CC features). We applied a multi-label classification model to validate the K-means clustering [45] for the individual features. The individual entropies are computed from AMMs [53]to describe trajectory pattern of individual crowd motions [39] [40]. Thus, by calculating the features for different individuals in various crowd scenes, we are able to categorize individual motions.

We assume that the distribution of the different kinds of individuals in a crowd scene determines the characteristics of the crowd motion. For instance, a crowd scene where most of the individuals belong to the same motion category should be under coherent motion, while a crowd scene with a lot of different people should probably be a random motion. Hence, we would like to use the proper categorization to recognize crowd scenes.

We compute the distribution of various kinds of individuals in each crowd motion to identify the crowd scene's class. The entire process is illustrated in Fig. 1. The idea is similar to the Bag-of-words model [29] [41]. In Fig. 1, the clusters of the individual features serve are as ‘visual words’ and the last computation of histogram is used to summarize the crowd motion.

- **Holistic feature and individual feature of crowd database**

The holistic entropies computed from multiple agent based motion models are combined to form holistic features for the efficient analysis of crowd motions.

To calculate the real crowd density, both the number and the location of each human target need to be used. Thus the accuracy of both has to be measured. In the experimentation, the area of interest is defined for each test scenario. We manually count the number of individual targets inside each area of interest as required the ground truth. The comparison of different trained agent-based motion model using crowd database is illustrated in Table 3.

Table 3: Comparison of different AMM learning model based on individual entropy of crowd data

Types of AMM Model	Accuracy	Precision	Recall	F-measures
Agent motion model-1	0.461	0.327	0.416	0.340
Agent motion model-2	0.617	0.543	0.575	0.528
Agent motion model -3	0.685	0.651	0.672	0.633
Agent motion model -4	0.729	0.704	0.712	0.681
Agent motion model -5	0.834	0.794	0.796	0.766
Agent motion model -6	0.881	0.875	0.882	0.864
Agent motion model -7	0.924	0.928	0.934	0.959
Proposed approach	**0.949**	**0.958**	**0.944**	**0.969**

The detection process defined as the total number of detected individual objectives in the crowd trajectory using global information. Thus, tracking of individuals and the location inaccuracy will improve the overall performance of the system. The performance measure of different AMM learning based models using individual and holistic features of crowd motion is shown in Figs. 6, 7 and 8 respectively. The experimental results are also summarized in Table3and Table 4, respectively. For the performance analysis, we use multi-label classification technique to validate the K-means clustering [44] [45] of the individual features (shown in Fig. 6). First, we apply the extracted individual features for clustering and generate types of the crowd person. Then, we use 70% of the data for training and rest 30% of the data for testing. This set of experiments has been conducted 20 times. The ML-KNN [30] classifier is used to predict the labels of the real-world crowd scenes using different crowd motion featuresThe default settings of its two parameters are (1) the number of neighbours K =10 and (2) the smoothing parameter = 1. The number of clusters is assigned as K = 2 to 12 to evaluate their average precision in the classification tasks.

We have performed the comparison of among three individual features (shown in Fig. 7). It is observed that extracted feature is better among the three features for the classification of crowd scenes, even if the feature (CB) does not include the information contained in the diagonal sub-matrix.

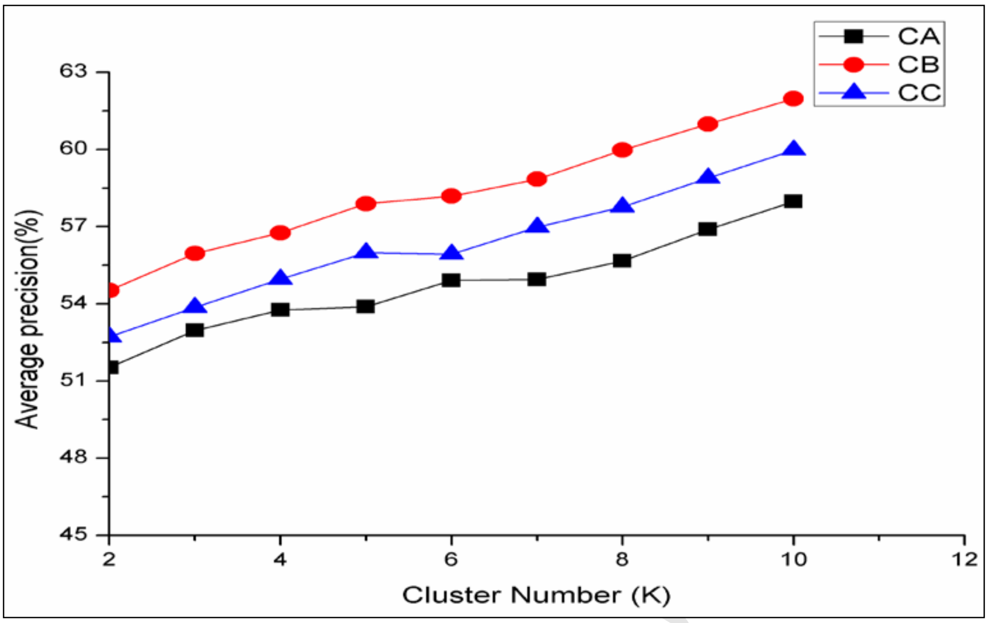


Fig. 6: Classification accuracy based on clustering of individual features

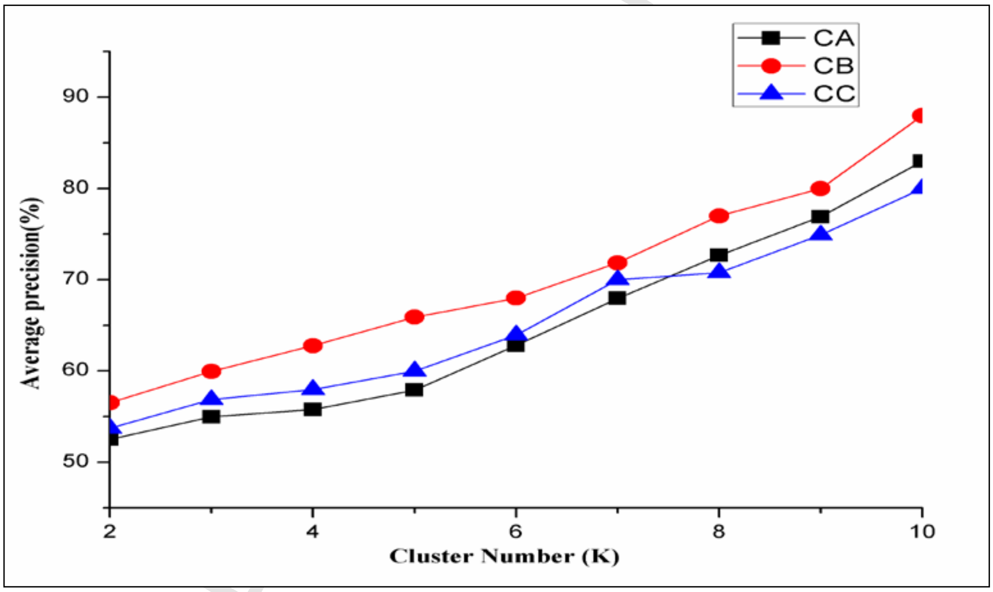


Fig. 7: Classification accuracy based on clustering the trained holistic feature

It supersedes the others. We can conclude from this fact that the information provided in the diagonal sub-matrix is an independent motion pattern of the individuals. It is not reliable for classifying the crowd motion reflects the independent motion patterns of the Individuals in the crowd trajectory, is not reliable for classifying the crowd scenes. This is because of the independent of pedestrian motion pattern. From Fig. 7, it can be observed that each of the three curves reaches to its approximate optimal value as K = 2 to 10 of individual features of crowd motion and individual feature (CB) still performs better than the others.

Table 3 illustrates the classification accuracy of different trained AMM based learning models for classification of crowd motion based on clusters of individual features of crowd motion. The salient sets of individual feature of crowd database are clustered in different clusters. Each cluster is grouped for the better analysis of crowd motion. In Table 3, it can be observed that classification accuracy of agent based motion model (AMM) increases as increasing the number of individual feature set of crowd database because the selected features from each cluster are discriminatory for better analysis by classify of individual people in crowd motion. The trained AMM model yields 98.91% classification accuracy to classify the crowd motion based on the salient set of individual features.

Table 4: Comparative study of different Agent motion model (AMM) (AMM-1 to AMM -7) learning approaches based on individual features of crowd data

	Classification accuracy (%) of different individual feature clusters		
Set of Features	**CA**	**CB**	**CC**
50	65.79	68.39	67.91
100	67.17	67.89	68.78
150	69.76	70.76	70.95
200	70.79	72.99	71.29
250	71.57	74.87	74.94
300	72.76	77.76	75.85
350	76.78	81.78	77.98
400	82.85	86.95	84.95
450	84.89	88.89	86.49
500	87.98	94.28	89.99
550	88.91	96.94	90.98
600	89.95	97.95	92.91
650	90.93	98.91	94.67

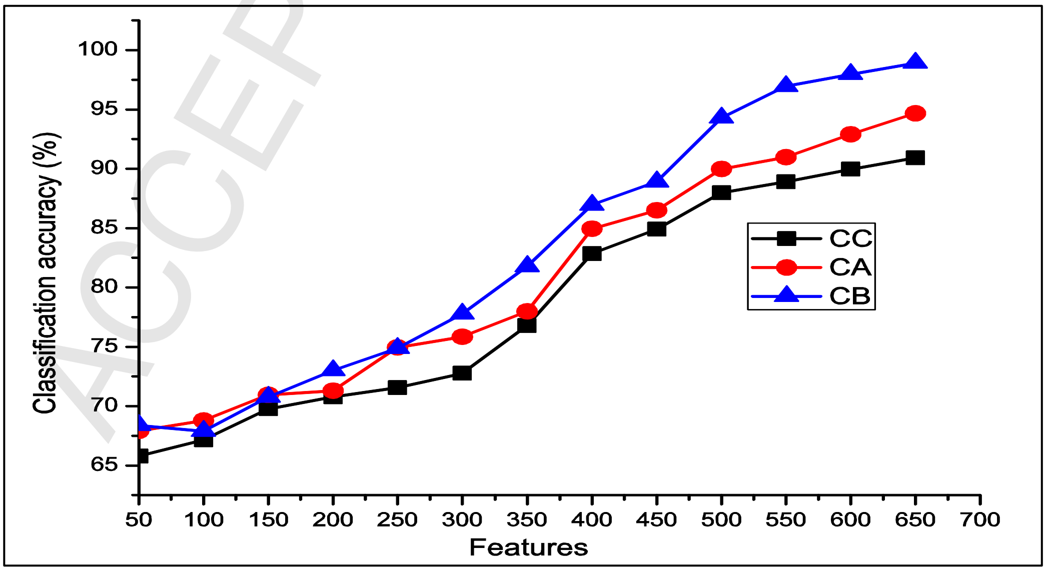


Fig. 8: Classification accuracy of different AMM learning based models using clustering of individual features of crowd data

7. Comparison to the state-of-the-art

In this section, we have performed the detailed analysis of experimental analysis of proposed g approach for analysis of crowd motion of individual based on information of the crowd trajectory. We now compare the performance of learning framework based crowd system against current state-of-art approaches for computation of features and their analysis based on crowd data.As a baseline comparison, we have applied the motion boundary histograms (MBH) [46]based descriptor technique for the computation of histogram of crowd motion for the classification purpose. Also, we also compared our proposed method with the latest method [28]. The proposed approach performs the computation of feature descriptors on the crowd motion and then it combines the crowd descriptors as feature descriptor for classification of crowd movements [46]. For the fair comparison, we also performed the comparative analysis of both KLT [36] tracklets and complete crowd trajectories as input to their algorithms. Besides these approaches, Genetic algorithm (GA) [21] is also applied to train the models using an optimal RVO [51] [58] parameter for each instance and implement the learned parameters as the feature.

In this experiment, leave-one-out based cross-validation strategy is applied to evaluate the performance of the multi-label classification technique. For the performance evaluation of crowd motion based on individual feature and holistic features, we have chosen the different metrics like (1) average precision. The average precision measures positive predictive value , (2) hamming distance based error estimation. It measures the how many times an instance label pairs are dissimilar with the target label in the crowd database, (3) one error based measure metrics. The one error based technique measures the how many top-ranked paired labels are dissimilar in the proper labelled datasets, (4) coverage based measure parameters – which measures how many labels in the crowd database are required to full coverage of all the instances in given sets, and (5) rank score based loss – rank score based technique measures the how many extracted feature vector pairs are reversely ordered to match instances.

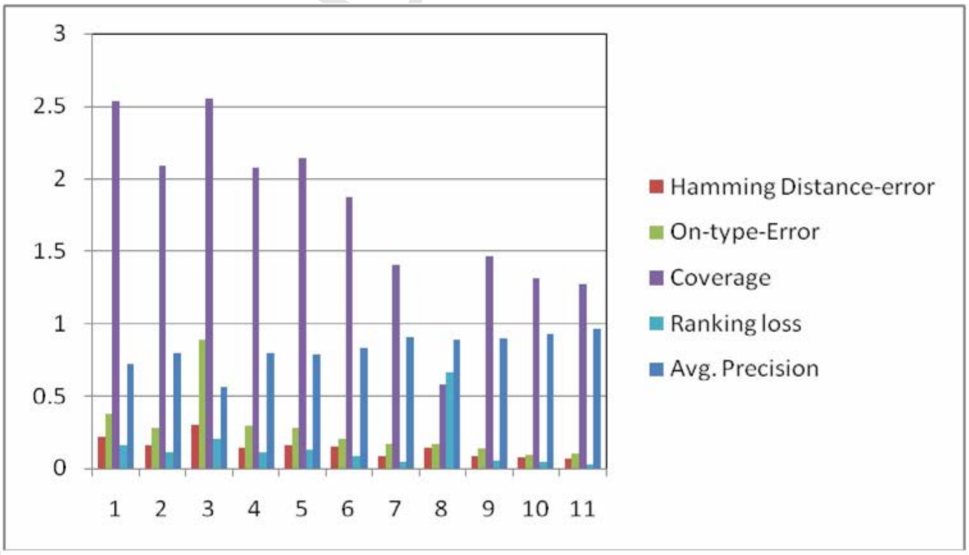


Fig. 9: illustrates the comparative study of different crowd approaches based on table 5.

Based on mentioned above parameters, experimental results are summarized in Table 5. Table 5 illustrates the classification accuracy of various current state-of-the-art based approaches in depth analysis of crowd motion. The comparative study and comparison to the state-of-the-art approaches for tracking and detection of crowd motion is shown in Fig. 9. In this experiment, we have applied the KLT- based [28] , MBH [46] , GA [21] and different agent based motion models [53] based learning methods for the comparative study of experimental results. In Table 3, it can be observed that the trajectory-based method provides better performance than the motion boundary histogram (MBH) [46] based descriptor technique. The KLT-based descriptor method [28] yields accurate crowd movement information. In this paper, MBH [46]and KLT-based descriptor method [28] cater holistic features for better evaluation of crowd motion. These approaches are not robust to illustrate tracking of crowd movement.

Table 5: Comparative study of experimental results of multi-label classification for crowd features

S.No.	Approaches	Avg. Precision	Hamming Distance-error	On-type-Error	Coverage	Ranking loss
1	Shan et al. [28] KLT-based approach	0.719	0.214	0.379	2.528	0.162
2	Shao et al. [25]	0.795	0.157	0.272	2.089	0.112
3	MBH [46]	0.558	0.304	0.887	2.552	0.203
4	GA [21]	0.790	0.142	0.292	2.071	0.112
5	Agent motion model-1 [53]	0.789	0.156	0.275	2.141	0.122
6	Agent motion model -2[53]	0.825	0.147	0.201	1.875	0.087
7	Agent motion model-3[53]	0.902	0.082	0.166	1.403	0.040
8	Agent motion model -6[53]	0.887	0.141	0.168	0.578	0.660
9	Agent motion model -7[53]	0.897	0.080	0.134	1.462	0.049
10	Liu et al. [30]	0.925	0.072	0.091	1.313	0.041
11	**Proposed approach**	**0.962**	**0.067**	**0.0981**	**1.268**	**0.021**

Based on overall observations, we conclude that our proposed method extracts the individual and holistic features from the prepared crowd database and achieves the significant improvements over the entire existing feature learning based models, crowd analysis detection approaches, and baselines algorithms which are available in the literature.

8. Conclusion and Future Directions

In this paper, we have proposed a crowd management system for monitoring the crowd of the smart city. Agent motion-based learning model is applied for discriminatory features extraction of crowd motions of every individual from the crowd trajectories. An extended Kalman-filter based technique [36] and a KL-divergence based approach [35] [46] are used to evaluate the discrimination of the individual feature and the holistic features of crowd scenes. The extended Kalman-filter based approach mitigates the noises and provides the smoothed feature sets. We achieve the precision rate of 96.20% for the monitoring of crowd. Own experimental results show that the proposed crowd management system performs well in recognizing the individuals from the real world crowd scene. The shortcomings of this work are discussed below. We plan to address these issues in our future work.

- The proposed crowd management system is limited for the massive trajectory of several crowd motion inputs. It is a major issue to correlate between any crowd videos automatically. Manual estimations are required for perspective transformation, and in this case, we cannot entirely rely on multi-target tracking.
- The proposed system can be speeded if we use parallel programming paradigm. Even it can be implemented for some real-time application like visual camera surveillance. Finally, we plan to apply the deep learning based framework for crowd management and crowd tracking.

References

1. Franke, T., Lukowicz, P. and Blanke, U. ‘Smart crowds in smart cities: real life, city scale deployments of a smartphone based participatory crowd management platform’, Journal of Internet Services and Applications, 2015, 6(1), pp.1-19.
2. Curry, E., Dustdar, S., Sheng, Q.Z. and Sheath, A., ‘Smart cities–enabling services and applications’, Journal of Internet Services and Applications, 2016, 7(1), pp.1-6.
3. X. Zhu, J. Liu, J. Wang, C. Li, H. Lu, ‘Sparse representation for robust abnormality detection in crowded scenes’, Pattern Recognition, 2014, 47 (5), pp.1791–1799.
4. Choi W., Savarese S., ‘A unified framework for multi-target tracking and collective activity recognition’, in Proceedings of European Conference on Computer Vision, 2012.
5. Li, R. Chellappa R., Zhou S. K., ‘Recognizing interagent-based group activities using temporal interaction matrices and their Riemannian statistics’, International Journal of Computer Vision, 2013, 101 (2), pp. 305–328.
6. Franke, T., Lukowicz, P. & Blanke, U. J, ‘Smart crowds in smart cities: real life, city scale deployments of a smartphone based participatory crowd management platform’, Internet Serv Appl 2015, 6: pp. 27.
7. Curry, E., Dustdar, S., Sheng, Q.Z. and Sheath, A., ‘Smart cities–enabling services and applications’, Journal of Internet Services and Applications, 2016, 7(1), pp.1-6.

8. Mehran, R., Oyama, A. and Shah, M., ‘Abnormal crowd behaviour detection using social force model’, in Proceedings of IEEE international conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 935-942.
9. Li, T., Chang, H., Wang, M., Ni, B., Hong, R. and Yan, S., ‘Crowded scene analysis: A survey’, IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(3), pp.367-386.
10. Wang, X., Ma, X. and Grimson, W.E.L., Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models’, IEEE Transactions on pattern analysis and machine intelligence, 2009, 31(3), pp.539-555.
11. Rahman M. A., ‘A framework to Support Massive Crowd: A Smart City Perspective’, In Proceedings of IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA, USA, 2016, pp. 1-6.
12. Asimakopoulou E. and Bessis N., ‘Buildings and Crowds: Forming Smart Cities for More Effective Disaster Management’, in Proceedings of 5th IEEE International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), Seoul, 2011, pp. 229-234.
13. Ali S., Shah M., ‘Floor fields for tracking in high density crowd scenes’, in Proceedings of European Conference on Computer Vision, volume 5303 of the series Lecture Notes in Computer Science, pp. 1-14.
14. Pellegrini S., Ess A., Schindler K., van Gool L., ‘You'll never walk alone: modelling social behaviour formulation target tracking’, in Proceedings of IEEE International Conference on Computer Vision, 2009, pp. 261-268.
15. Liu W., Lau, R.W. and Manocha, D., ‘Robust individual and holistic features for crowd scene classification. Pattern Recognition’,2016, 58, pp.110-120.
16. Mahadevan V., Li W.,Bhalodia V., Vasconcelos N., ‘Anomaly detection in crowded scenes’, in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010, 10(5), pp. 374-381.
17. Chan A. B. and Vasconcelos, N., ‘Modeling, clustering, and segmenting video with mixtures of dynamic textures’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(5), pp.909-926.
18. Babenko B., Yang M. H. and Belongie S., ‘Robust object tracking with online multiple instance learning’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8), pp.1619-1632.
19. Zhou B., Wang X., Tang X., ‘Understanding collective crowd behaviours: learning a mixture model of dynamic pedestrian-agents’, in Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 2871-2878.
20. McPhail C., ‘The Myth of the Madding Crowd’, Aldine de Gruyter, 1991.
21. McPhail C., and Wohlstein R., ‘Using Film to Analyze Pedestrian Behavior’, Sociological Methods and Research, 1982, 10, pp. 347-375.
22. Zhou B., Tang X., Wang X.,‘Measuring Crowd Collectiveness’,IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 36(8), pp. 1586-1599.

23. Kratz L. and Nishino K., 'Tracking Pedestrians Using Local Spatio-Temporal Motion Patterns in Extremely Crowded Scenes',IEEE Transactions on Pattern Analysis and Machine Intelligence,2012, 34(5), pp. 987–1002.
24. Rodriguez M., Ali S., and Kanade T., 'Tracking in unstructured crowded scenes', in Proceedings of 12th IEEE International Conference on Computer Vision, Kyoto, 2009, pp. 1389-1396.
25. Song X., Shao X., Zhao H., Cui J., Shibasaki R., and Zha H., 'An online approach: Learning-semantic-scene-by-tracking and trackingby-learning-semantic-scene', in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 739–746.
26. Yang B. and Nevatia R., 'Multi-target tracking by online learning of non-linear motion patterns and robust appearance models', in Proceedings ofIEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1918–1925.
27. Bazzani L., M. Cristani, and V. Murino, 'Decentralized particle filter for joint individual-group tracking,'in Proceedings ofIEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1886–1893.
28. Benfold B. and Reid I., 'Stable multi-target tracking in real-time surveillance video,' inProceedings ofIEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, Providence, RI, pp. 3457-3464.
29. Pisano, E.D., Zong, S., Hemminger, B.M., DeLuca, M., Johnston, R.E., Muller, K., Braeuning, M.P. and Pizer, S.M., 'Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms', Journal of Digital imaging, 1998, 11(4), pp.193-200.
30. Liu W., Lau R., Wang X., and Manocha D., 'Exemplar-AMMs: Recognizing crowd movements from pedestrian trajectories', IEEE Transactions on Multimedia, no.99, pp.1-1.
31. Lowe, D.G., 'Object recognition from local scale-invariant features', in Proceedings of the 7thIEEE International Conference on Computer vision, 1999, 2, pp. 1150-1157.
32. Helbing D., and Molnar P., 'Social force model for pedestrian dynamics', Physical review E, 1995, 51(5), pp.42-82.
33. Reynolds C., 'Flocks, herds and schools: a distributed behavioral model', in Proceedings of the 14th annual conference on Computer graphics and interagent-based techniques (SIGGRAPH '87), 21(4), July 1987,pp. 25-34
34. Van den Berg, J., Lin, M. and Manocha, D., 'Reciprocal velocity obstacles for real-time multi-agent navigation', in Proceedings of IEEE International Conference on Robotics and Automation, (ICRA), pp. 1928-1935.
35. Hershey J. R. and Olsen P. A., 'Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models', IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07), 2007, Honolulu, HI, 2007, pp. IV-317-IV-320.
36. Fahrmeir, L., 'Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models', Journal of the American Statistical Association, 1992, 87(418), pp.501-509.
37. Julier, S.J. and Uhlmann, J.K., 'Unscented filtering and nonlinear estimation', Proceedings of the IEEE, 92(3), 2004, pp. 401-422

38. Zhang M., Zhou Z., ‘ML-KNN: a lazy learning approach to multi-label learning’, Pattern Recognition, 2007, 40(7),pp. 2038–2048.
39. Akata, Z., Perronnin, F., Harchaoui, Z. and Schmid, C., ‘Good practice in large-scale learning for image classification’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3), pp.507-520.
40. Wang, H. and Wang, J., ‘An effective image representation method using kernel classification’, In proceedings of IEEE 26th International Conference on Tools with Artificial Intelligence, pp. 853-858, 2014.
41. Ciregan, D., Meier, U. and Schmidhuber, J. ‘Multi-column deep neural networks for image classification’, In proceedings of IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2012 , pp. 3642-3649
42. Lowe, D.G. ‘Distinctive image features from scale-invariant keypoints’, International journal of computer vision, 2004, 60(2), pp.91-110.
43. Lowe, D.G. ‘Object recognition from local scale-invariant features’, In proceedings of the 7thIEEE international conference onComputer vision, 1991, vol. 2, pp. 1150-1157.
44. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R. and Wu, A.Y. ‘An efficient k-means clustering algorithm: Analysis and implementation’, IEEE transaction on pattern analysis and machine intelligence, 2002, 24(7), pp.881-892.
45. Kumar, S., Datta, D. and Singh, S.K. ‘Black hole algorithm and its applications’, In Computational Intelligence Applications in Modeling and Control, 2015, pp. 147-170.
46. Kumar, S. and Singh, S.K. ‘Monitoring of pet animal in smart cities using animal biometrics’, Future Generation Computer Systems, 2016, http://dx.doi.org/10.1016/j.future.2016.12.006
47. Liu, W., Lau, R.W. and Manocha, D. ‘Robust individual and holistic features for crowd scene classification’, Pattern Recognition, 2016, 58, pp.110-120.
48. Rangel-Huerta, A., Ballinas-Hernández, A.L. and Muñoz-Meléndez, A. ‘An entropy model to measure heterogeneity of pedestrian crowds using self-propelled agents’, Physica A: Statistical Mechanics and its Applications,’2016,http://dx.doi.org/10.1016/j.physa.2016.12.090
49. Kumar, S., Tiwari, S. and Singh, S.K., Face Recognition of Cattle: Can it be Done?. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 2016, 86(2), pp.137-148.
50. Kumar, S., Singh, S.K., Datta, T. and Gupta, H.P. ‘A Fast Cattle Recognition System using Smart devices’, In Proceedings of the 2016 ACM on Multimedia Conference, 2016, pp. 742-743.
51. Kumar, S., Singh, S.K., Dutta, T. and Gupta, H.P. ‘Poster: a real-time cattle recognition system using wireless multimedia networks’, In Proceedings of the 14th Annual ACM International Conference on Mobile Systems, Applications, and Services (Mobisys-2016), 2016, pp. 48-48.
52. Kumar, S., Singh, S.K., and Singh A. K., ‘A Muzzle Point Pattern based Techniques for Individual Cattle Identification’,IET Image Processing, 2017, pp. 1-21, DOI: 10.1049/iet-ipr.2016.0799
53. Liu, W., Lau, R.W., Wang, X. and Manocha, D. ‘Exemplar-AMMs: Recognizing Crowd Movements From Pedestrian Trajectories,’IEEE Transactions on Multimedia, 2016, 18(12), pp. 2398-2406, doi: 10.1109/TMM.2016.2598091

54. Zhang, Y., Brady, M. and Smith, S., 'Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm', IEEE transactions on medical imaging, 2001, 20(1), pp.45-57.
55. Moon, T.K., 'The expectation-maximization algorithm', IEEE Signal processing magazine, 1996, 13(6), pp.47-60.

Author Biography

Santosh Kumar received the B. Tech. degree in Computer Science & Engineering, from Uttar Pradesh Technical University (U.P.T.U.), in 2008, and M. Tech. in Computer Science & Engineering from Birla Institute of Technology, Mesra, Ranchi (Jharkhand, India) in 2012.

Currently he is pursuing Ph. D. from the Department of Computer Science & Engineering, Indian Institute of Technology (B. H. U.) Varanasi, India. His research interests span a wide range of spectrum including computer vision, biometrics, pattern recognition, image analysis, swarm intelligence, metaheuristics, bio-inspired computing, wireless sensor, multimedia and artificial intelligence. He was a recipient of best poster awards at Institute Day, Indian Institute of Technology (B.H.U.) Varanasi-221005. He is a member of the Computer Society and the Association for Computing Machinery.

Deepanwita Datta is currently pursuing her PhD in Computer Science and Engineering at Indian Institute of Technology (BHU), Varanasi, India. She completed her M.Tech in Information Security from National Institute of Technology, Durgapur and B.Tech in Electronics and Communication Engineering from West Bengal University of Technology. Her areas of interest include Multimodal Retrieval, Image Processing, Image Encryption and Security.

Dr. Sanjay Kumar Singh completed his B. Tech. in Computer Engg., M. Tech. in Computer Applications and Ph.D. in Computer Science and Engineering. Currently he is an associated professor at the Department of Computer Science and Engineering, IIT BHU, Varanasi. He is a Certified Novell Engineer (CNE) from Novell Netware, USA and a Certified Novell Administrator (CNA) from Novell Netware, USA. He is a member of LIMSTE, IEE, International Association of Engineers and ISCE. His research areas include Biometrics, Computer Vision, Image Processing, Video Processing, Pattern Recognition and Artificial Intelligence. He has over 50 national and international journal publications, book chapters and conference papers. He is also a Guest Editorial Board Member of Multimedia Application and Tools, Springer, and the EURASIP Journal of Image and Vision Processing (Springer). Currently, Dr. Singh has served as Guest editor, TPC member and reviewers for various conferences and reputed journals. He is a member of the Computer Society and the Association for Computing Machinery.

Dr.Arun Kumar Sangaiah has received his Master of Engineering (ME) degree in Computer Science and Engineering from the Government College of Engineering, Tirunelveli, Anna University, India. He had received his Doctor of Philosophy (PhD) degree in Computer Science and Engineering from the VIT University, Vellore, India. He is presently working as an Associate Professor in School of Computer Science and Engineering, VIT University, India. His area of interest includes software engineering, computational intelligence, wireless networks, bio-informatics, and embedded systems. He has authored more than 100 publications in different journals and conference of national and international repute. His current research work includes global software development, wireless ad hoc and sensor networks, machine learning, cognitive networks and advances in mobile computing and communications. Also, he was registered a one Indian patent in the area of Computational Intelligence. Besides, Prof. Sangaiah is responsible for Editorial Board Member/Associate Editor of various international journals.

Santosh Kumar

Deepanwita Datta

Dr. Sanjay Kumar Singh

Dr. Arun Kumar Sangaiah

Highlights

- Presents an intelligent decision computing based paradigm for crowd monitoring in the smart city.
- The proposed monitoring system is learned by a holistic feature and performs the classification of crowd motions by applying multi-label classification technique.
- Evaluated the performance of the proposed system for dense crowds in the smart city.
- The proposed system provides 92.20% average precision accuracy.