Journal of Artificial Intelligence Research & Advances
ISSN: 2395-6720 (Online)
Volume 6, Issue 2
www.stmjournals.com

A Review on Crowd Analysis Techniques

Kinjal V. Joshi[1,*], *Narendra M. Patel*[2], *Mehul B. Shah*[3]
[1]Department of Computer Engineering, Gujarat Technological University, Ahmedabad, Gujarat, India
[2]Department of Computer Engineering, Birla Vishvakarma Mahavidyalaya Engineering College, Vallabh Vidyanagar, Gujarat, India
[3]Department of Electronics and Communication, Gujarat Technological University, Ahmedabad, Gujarat, India

Abstract

Nowadays crowd analysis is an active open research problem in computer vision field. Its various applications like crowd management, public space design, virtual environments, visual surveillance, and intelligent environments attract researchers. Today, security is a major concern, and any abnormal activity can automatically be detected by crowd analysis, so this research topic has got good attention of different researchers. Abnormal activity detection in the crowd is a very challenging task because of occlusion. In this study, a survey of different crowd analysis techniques is presented. Crowd analysis can be divided among three major modules: people counting, people tracking and behavior understanding. Various methods for each module are discussed here.

Keywords: *Crowd analysis, people counting, people tracking, abnormal activity detection*

****Author for Correspondence*** E-mail: kinjaljoshi@gcet.ac.in

INTRODUCTION

Crowd is a unique group of individuals or something which involves community or society. Crowd scene can be divided into two types: structured crowd scene and unstructured crowd scene [1]. The term structured crowd scene can be described as crowd moves coherently in common direction, motion direction does not vary with time, each spatial location of the scene supports only one dominant crowd behavior over the time. For example, marathon race, queues of people event and traffic on the road. Meanwhile the term unstructured crowd scene represents the random crowd motion; different participants move in different directions at different time, each spatial location supports multi model crowd behavior. For example, people walking on a zebra crossing in opposite direction, exhibition, railway station and airport [1]. The phenomena of the crowd are very familiar in a variety of research disciplines such as sociology, civil and physic. Nowadays, it becomes the most active-oriented research and trendy topic in computer vision. Traditionally, three processing steps are involved in crowd analysis namely people counting, people tracking and behavior understanding. People detection or identification is a very challenging task in crowd analysis because person could be occluded in many ways. In a high-level analysis, one could segment their body parts (e.g., head, face, hands, and arms) that could be used in gesture recognition or machine-human interaction. It is also possible to deal with group behavior, once people can be part of high-level structures, such as groups or crowds [2, 3]. In particular, the behavioral analysis of crowd scenes is of great interest in a large number of applications such as [4]:

- *Crowd Management*: It can be used for developing crowd management strategies, to avoid crowd related disasters and ensure public safety.
- *Public Space Design*: To provide guidelines for the design of public spaces.
- *Virtual Environments*: It can be used to validate or increase the performance of the mathematical models used in crowd simulations.
- *Visual Surveillance*: It can be used for automatic detection of alarms.
- *Intelligent Environments*: It can be used to take a decision on how to split a crowd in a museum, based on their behavior.

Despite the potential of crowd analysis applications using computer vision algorithms, most of the existing work for detection/identification of people, groups of people, or even for the estimation of body parts (in a high-level analysis) have been focused on no crowded situations. As related by Zhan *et al*., when very crowded video sequences are analyzed, conventional computer vision methods are not appropriate [4]. Although one may think that a straightforward extension of techniques designed for no crowded scenes could be suitable for dealing with crowded situations, that is not true. First of all, a crowd is something beyond a simple sum of individuals. The crowd can assume different and complex behaviors as those expected by their individuals. The behavior of crowd is widely understood to have collective characteristics that can be described in general terms. For example, descriptions such as "an angry crowd" or a "peaceful crowd" are well accepted [5].

CROWD ANALYSIS

Three important problems that have been reported in the literature regarding automated analysis of crowded scenes are (Figure 1):

i) People counting/density estimation models,
ii) Tracking in crowded scenes, and
iii) Crowd behavior understanding models. Figure1 shows schematic representation of crowd analysis.

People Counting

An important problem in crowd analysis is people counting/density estimation (either in images or video sequences). For instance, crowd density analysis could be used to measure the comfort level in public spaces, or to detect potentially dangerous situations. There are several models developed to estimate the number of people in crowded scenarios using computer vision techniques. In this work, these models are divided into three categories: i) pixel-based analysis, ii) texture level analysis, and iii) object level analysis.

Pixel Based Analysis

Pixel-based methods rely on very local features (such as individual pixel analysis obtained through background subtraction models or edge detection) to estimate the number of people in a scene. Since very low-level features are used, this class is mostly focused on density estimation rather than precise people counting.

a) Method based on Background Subtraction and Edge Detection: Davies *et al*. have proposed an approach to estimate the density of the crowd using pixel level information [5]. They combined a background subtraction technique and detected edges to estimate the density. They assumed linear models to map foreground pixels or edges to the number of people, which were integrated using a Kalman filtering approach to improve the results. Their method also includes geometrical correction due to the perspective of the camera, to account for the fact that the dimension of the same person (in pixels) may change at different locations (the size in pixels decreases as the distance from the camera increases). Clearly, the linear function that maps pixel counts to the number of people in the scene fails when strong occlusions occur.

Regazzoni *et al*. proposed an approach to estimate the crowd density in images [8]. In their work, features extracted from each acquired image (basically the results of an edge detection algorithm and finding vertical edges) are related to the number of people present in the monitored scene by using nonlinear models obtained by means of dynamic programming in an offline training phase. In the operation phase, the detected features are integrated across time through an extended Kalman filter to improve the results. This method was mostly focused on indoor scenes with a limited number of people (up to around 30). Furthermore, the offline training procedure is adjusted to a given camera setup and scenario, and changes require a new training phase.

b) Method based on Neural Network: This method is proposed by Cho *et al.* [9]. Here, the objective was to detect high-density situations and to provide statistical analysis about the flow of people across time for activity planning. They proposed a simple edge detector based on binary image thresholding and explored the length of detected edges as a feature for people counting through a single hidden layer neural network. Their approach was implemented using a hybrid method for global learning that combines the least-squares method with three types of global optimization approaches, defined as follows: random search simulated annealing, and genetic algorithms. Their results presented approximately 90% of correctness over all tested methods. They also pointed out that the combination of least-square and random search algorithm was the fastest, opposed to the combination of the least-square and simulated annealing (about 100 times slower).

Kong and collaborators presented a method based on learning to estimate the number of people in crowd [10]. Edge orientation and the histogram of the object areas (extracted from foreground objects through a background subtraction algorithm) are used as image features. A normalization procedure is performed to account for camera perspective, and the training model used to relate the detected features with the number of people was based on a feed-forward neural network. An interesting characteristic of this approach is the training of normalized features, so that changes in the camera setup do not require a new training phase.

c) Crowd Density Estimation based on Pixel Count: Ma *et al*. proposed a crowd density estimation algorithm for surveillance purposes [11]. In their work, an approach based on pixel counting of foreground objects is used to estimate the crowd density, combined with a projective correction using a calibrated camera. A linear relation between the number of pixels and number of persons was then derived by applying the geometric correction. The density over time is also monitored, aiming to detect some unusual behavior. Their approach suffers the same occlusion problems faced by Davies *et al.*, since a linear model is used to estimate the people count [5].

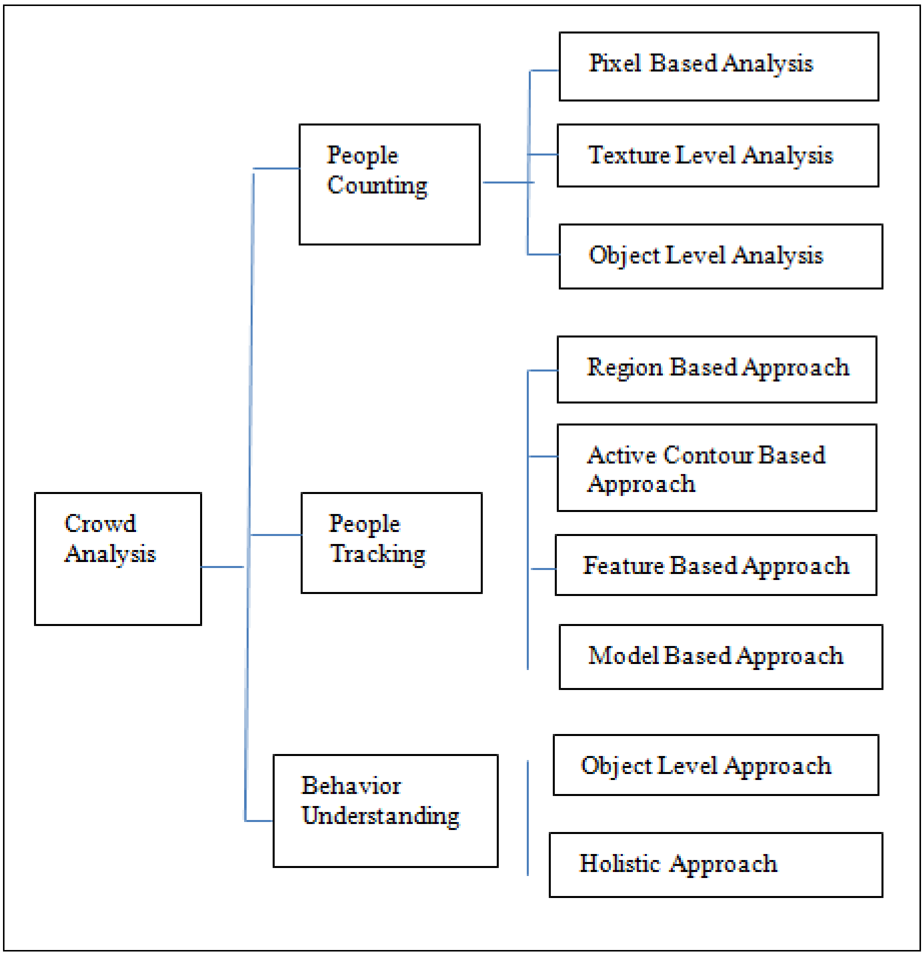

Fig. 1. *Schematic Representation of Crowd Analysis [6, 7].*

d) Method Based on Convolutional Neural Network: Sindagia *et al*. had proposed convolutional neural network to learn two related sub-tasks: crowd count classification (high-level prior) and density map estimation in a cascaded fashion [12]. The network takes an image of arbitrary size, and outputs crowd density map. The cascaded network has two stages corresponding to the two sub-tasks, with the first stage learning high-level prior and the second stage preforming density map estimation. Both stages share a set of convolutional features. The first stage consists of a set of convolutional layers and spatial pyramid pooling to handle arbitrarily sized images followed by a set of fully connected layers. The second stage consists of a set of convolutional layers followed by fractionally stridden convolutional layers for up sampling the previous layer's output to account for the loss of details due to earlier pooling layers. Two different sets of loss layers are used at the end of the two stages, however, the loss of the second layer is dependent on the output of the earlier stage. Results are compared with ground truth observations and this method represents very less Mean Squared Error and Mean Absolute Error.

Zhang *et al.* had also used CNN for crowd counting and density estimation [13]. For training model, each pedestrian's spatial position is labeled and the density map is created based on pedestrians' spatial location, human body shape and perspective distortion of images. Due to severe occlusions, heads and shoulders can be the main cues to judge whether there exists a pedestrian at each position. The body parts of pedestrians are not reliable for human annotation. Taking these characteristics into account, the crowd density map is created by the combination of several distributions with perspective normalization. This method is tested on very large dataset and provides very good accuracy.

Texture Level Analysis

Algorithms that rely on texture analysis explore a coarser grain if compared to pixel-based methods as texture modeling requires the analysis of image patches. Although this class explores higher-level features when compared to pixel-based approaches, it is also mostly used to estimate the number of people in a scene rather than identifying individuals.

Different methods are available for texture level analysis like (i) gray-level dependence matrix, (ii) straight lines segments, (iii) Fourier analysis (iv) fractal dimension, (v) support vector machine and (vi) crowd counting algorithm. Gray level dependence matrix-based method combined with a Bayesian classifier provides better contrast and homogeneity as texture features.

a) Method Based On Support Vector Machine: Wu *et al*. proposed an approach to estimate the crowd density using support vector machines (SVMs) and texture analysis [14]. In their work, a perspective projection model is adopted to generate a series of multiresolution cells, and the gray-level dependence matrix method is used to extract textural information within these cells. A multiscale texture vector is built, and an SVM is trained to relate the textural features with the actual density of the scene. The authors reported a maximum estimation error for each cell below 5% and proposed as future work the possibility of including a background subtraction method in the feature extraction stage. One drawback of their approach, however, is the need of retraining the SVM for scenarios with different camera setups, since the density cells are highly dependent on camera parameters.

b) Crowd Counting Algorithm: Chan *et al*. developed a crowd counting algorithm based on a texture-based motion segmentation technique and Gaussian process regression [15]. The authors initially segment the crowd into different motion directions using the mixture of dynamic textures, and for each motion cluster, they extract segment features (area, perimeter, perimeter edge orientation, perimeter-area ratio), internal edge features (total edge pixels, edge orientation, Minkowski dimension) and texture features (homogeneity, energy, and entropy). These features are normalized to account for camera perspective, and a Gaussian process regression is used to relate the number of people per segment.

Object Level Analysis

Methods that rely on object-level analysis try to identify individual objects in a scene. They tend to produce more accurate results when compared to pixel-level analysis or texture-based approaches but identifying individuals in a single image or a video sequence is mostly feasible in lower density crowds. In denser crowds, clutter and severe occlusions make the individual counting problem almost impossible to solve, despite the recent advances of computer vision and pattern recognition techniques.

a) Method Based on HaaR Wavelet Transform

Lin *et al*. proposed an algorithm to estimate the crowd density in three stages [16]. In the first one, they searched for objects with head-like contour in the image space, using the Haar wavelet transform. In a second stage, the features of the object are analyzed, using an SVM, aiming to classify it as a head or not. Finally, a perspective transformation is done aiming better estimate the density of the entire crowd.

b) Segmentation Approach: Leibe *et al*. proposed a pedestrian detection scheme using a top-down segmentation approach [17]. In fact, they explored a combination of local features (a scale-invariant version of the implicit shape model) and global features (Chamfer distance) to obtain the probability of a person being present, which is measured by comparing small learned image patches of the appearance of humans and their occurrence distribution. Their algorithm can reliably detect and localize pedestrians in relatively crowded scenes and with severe overlaps.

People Tracking

It is interesting to note that the problems of people counting and tracking are related, since both of them have the goal of identifying the participants of a crowd. However, the counting problem usually requires only an estimate of the number of people, regardless of their position. The tracking problem, on the other hand, involves the determination of the position of each person in the scene as a function of time. Categories of people tracking approach include region based approach, active contour based approach, feature based approach and model based approach.

Region Based Approach

In region based approach, crowd information such as density, direction and velocity is extracted using optical flow technique. For example, Weighted Maximum Cardinality Matching scheme with disparity estimation technique are presented by Kellly and colleagues to evaluate both environment condition indoor and outdoor [18].

Active Contour Based Approach

In active contour based approach, initial position of counter is very important. Iteration will start at the position of maximum probability of interested object and it will stop when the exact object is found. Khansari *et al*. has used this approach for people tracking [19]. Here, color histogram of head is used as feature. But the color histogram of head of different people may not have noticeable variation, so tracking accuracy is less.

Feature Based Approach

Peng *et al*. has proposed pixel wise event representation method to construct feature images, in which each blob corresponds to a visual event [20]. Then the original blob-level features are transformed into subspaces to model probabilistic appearance manifolds for each event-class. With the probability of an observation associated with each event-class (or state) derived from probabilistic manifolds and state transitional probability, the prior and posterior state distributions can be estimated.

Rittsche *et al*. presented approach for tracking. In their approach, each track is modeled by a color signature, an appearance template and a probabilistic target mask that is an autoregressive estimate of the foreground information [21]. People walking close to each other are clustered into “group tracks,” and individual tracking within the same group are smoothed using a constant velocity Kalman filter. Their approach can handle short term occlusions between isolated tracks, but it tends to fail when the crowd density gets very high.

Model Based Approach

a) Crowd Modeling: Ali and Shah proposed an approach for people tracking in structured high-density scenarios [22]. In their approach, each frame of a video sequence is divided into cells, each cell presenting just one particle. A person consists of a set of particles, and each person is affected by the layout of the scene (obstacles and barriers, which are learned automatically), as well as the motion of other people. An interesting aspect of this work is the use of concepts related to crowd modeling (obstacles and relationship with neighbors). On the other hand, since a manual identification of the individuals is required to initialize each track, this approach is not adequate for automatic people counting.

b) Correlated Topic Model: Rodriguez *et al*. presented a tracking approach to deal with unstructured environments [23]. They employed the correlated topic model (CTM), which allows each location of the scene to have various crowd behaviors. In their approach, the video sequence is divided into non-overlapping clips. For each of these clips, the optical flow is computed, and both position and velocity vector are quantized to generate a word in a codebook needed for the CTM. The motion words are assumed to arise from a generative process, whose parameters are estimated using a collection of training video sequences.

Crowd Behavior Understanding

There are two main approaches for crowd behavior analysis. In the "object-based" approach, a crowd is treated as a collection of individuals. On the other hand, "holistic" approach treats the crowd as a single entity, without the need of segmenting each individual. Clearly, in denser scenes it is very difficult to track individual components in the crowd, and hence the second approach tends to be more appropriate.

Object Based Approach

Here a crowd is treated as a collection of individuals and crowd behavior understanding is performed through some kind of segmentation or detection of individuals to analyze group behaviors.

a) *Method Based on Motion Pattern:* This method is proposed by Cheriyadat *et al*. [24]. This is the approach for clustering a set of low-level motion features into trajectories such as the dominant movements are computed based on the longest common subsequences. Here the main goal is to extract dominant motion patterns in a crowded scene. Then, individual motions not coherent with dominant flows could be highlighted and marked as potential unusual behavior. For example, their approach could detect a person making a U-turn in a scenario where most people move along the same direction.

Holistic Approach

Instead of tracking individual objects, this top-down view treats the crowd as a single entity, which directly tackles the problem of dense occluded crowds in contrast to the object-based methods. Holistic approaches usually try to obtain coarser-level (global) information, such as main crowd flows, and they tend to disregard local information (e.g., a single person moving against the flow).

a) *Method based on Discrete Fourier Transform:* Davies *et al*. proposed an approach based on discrete Fourier transform (faster moving objects are associated with high-frequency and nonmoving objects are associated with constant levels), combined with a linear area transform algorithm to distinguish static from moving crowds (a useful indicator of congestion or potential danger) [5]. The approach proposed in their work is to measure motion features at pixel or pixel-neighborhood level that are then aggregated to obtain motion properties for larger regions in an image. The aggregated results can then be used to establish overall preferential crowd velocities (direction and magnitude).

b) *Method Based on Hidden Markov Models:* Andrade *et al*. characterized a "usual behavior" of a crowd based on the analysis of their optical flow, using hidden Markov models (HMMs) [25]. In their work, an unsupervised algorithm is used for feature extraction and spectral clustering. They employed a background subtraction model to extract foreground objects, and used them as a mask to compute the optical flow. They validated the model with synthetic and real situations. The simulated situations are: unidirectional flow, people evacuating an environment with the exits obstructed, and a crowd moving around a fallen person.

c) *Method Based on Social Force Model:* Mehran *et al*. developed an approach for abnormal crowd behavior detection exploring the social force model [26], which was originally introduced by Helbing and Molnár to model crowds based on socio-psychological studies [27]. In their method, a set of particles is overlaid to the image, and they are moved according to the computed optical flow. The motion of the particles is then used to estimate the social forces. To detect normal patterns of forces over time, the magnitude of the interaction force vectors is mapped to the image plane, obtaining a force flow. The likelihood force flow is then estimated using the bag of words approach (for normal videos), and a fixed threshold on the estimated likelihood is used to classify each frame as normal or abnormal. An interesting aspect of this work is the exploration of socio-psychological to detect abnormal behavior.

CONCLUSION

In this study, we present the state of the art of crowd analysis and the classification of common approaches of the crowd analysis that could be useful to researchers. Crowd analysis is the most important concept for understanding the behavior, especially in analyzing the abnormal behavior in a crowded scene. Unusual behavior detection in structured crowd with limited number of people is also challenging task, for example, monitoring examination hall. Therefore, abnormal activity detection in crowd is open research area.

Crowd analysis can be done by three important modules i.e. people counting, people tracking and behavior understanding. Every module has different analysis techniques under different categories such as for people counting module the analysis for pixel level analysis, texture level analysis and object level analysis is different. While for people tracking module, it includes region based, active contour based, feature based and model-based approaches, the performances are measured differently. Finally, for behavior understanding module, the object level approach and holistic approach work differently. Most of the researchers used these procedures to analyze, detect and recognize the anomalies in crowded scene.

REFERENCES

1. Rodriguez M, Ali S, Kanade T. Tracking in Unstructured Crowded Scenes. In: *IEEE 12th International Conference of Computer Vision*. 2009.
2. Jacques JCS Jr, Braun A, Soldera J, *et al.* Understanding People Motion in Video Sequences Using Voronoi Diagrams. *Pattern Anal Applicat.* 2007; 10(4): 321–332p.
3. Kilambi P, Ribnick E, Joshi AJ, *et al.* Estimating Pedestrian Counts in Groups. *Comput Vis Image Understand.* 2008; 110(1): 43–59p.
4. Zhan B, Monekosso DN, Remagnino P, *et al.* Crowd Analysis: A Survey. *Machine Vis Applicat.* 2008; 19(2): 345–357p.
5. Davies AC, Yin JH, Velastin SA. Crowd Monitoring Using Image Processing. *IEE Electron Commun Eng J.* 1995; 7(1): 37–47p.
6. Julio Cezar Silveira Jacques Junior, Soraia Raupp Musse, Cláudio Rosito Jung. Crowd Analysis Using Computer Vision Techniques. *IEEE Signal Process Mag.* Sep 2010; 27(5): 66-77p.
7. Nilam Nur Amir Sjarif, Siti Mariyam Shamsuddin, Siti Zaiton Mohd Hashim, *et al.* Crowd Analysis and its Applications. *International Conference on Software Engineering and Computer Systems ICSECS*. 2011; 687-697p.
8. Regazzoni CS, Tesei A, Murino V. A Real-Time Vision System for Crowding Monitoring. In *Proc Int Conf Industrial Electronics, Control, and Instrumentation*. 1993; 3: 1860–1864p.
9. Cho S-Y, Chow TWS, Leung C-T. A Neural-Based Crowd Estimation by Hybrid Global Learning Algorithm. *IEEE Trans Syst Man Cybern.* 1999; 29(4): 535–541p.

10. Kong D, Gray D, Tao H. A Viewpoint Invariant Approach for Crowd Counting. In *Proc Int Conf Pattern Recognition*. 2006; 3: 1187–1190p.
11. Ma R, Li L, Huang W, *et al*. On Pixel Count Based Crowd Density Estimation for Visual Surveillance. In *Proc IEEE Conf Cybernetics and Intelligent Systems*. 2004; 1: 170–173p.
12. Sindagia Vishwanath A, Patelb Vishal M. CNN-based Cascaded Multi-task Learning of High-level Prior and Density Estimation for Crowd Counting. *International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 2017.
13. Cong Zhang, Hongsheng Li, Xiaogang Wang, *et al.* Cross-scene Crowd Counting via Deep Convolutional Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition.* 2015; 833-8p.
14. Wu X, Liang G, Lee KK, *et al*. Crowd Density Estimation Using Texture Analysis and Learning. In *Proc IEEE Int Conf Robotics and Biomimetics*. 2006; 214–219p.
15. Chan A, Liang Z, Vasconcelos N. Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking. In *Proc IEEE Conf Computer Vision and Pattern Recognition*. 2008; 1–7p.
16. Lin S-F, Chen J-Y, Chao H-X. Estimation of Number of People in Crowded Scenes Using Perspective Transformation. *IEEE Trans Syst Man Cybern A*. 2001; 31(6): 645–654p.
17. Leibe E, Seemann B, Schiele B. Pedestrian Detection in Crowded Scenes. In *Proc IEEE Conf Computer Vision and Pattern Recognition*, Washington, DC. 2005; 878–885p.
18. Kelly P, O'Connor NE, Smeaton AF. Robust Pedestrian Detection and Tracking in Crowded Scenes. *Image Vis Comput*. 2009; 27(10): 1445–1458p.
19. Khansari M, Rabiee HR, Asadi M, *et al.* Crowded Scene Object Tracking in Presence of Gaussian White Noise Using Undecimated Wavelet Features. In: *9th International Symposium Signal Processing and Its Applications, ISSPA.* 2007.
20. Peng C, Li-Feng S, Zhi-Qiang L, *et al.* A Sequential Monte Carlo Approach to Anomaly Detection in Tracking Visual Events. *IEEE Conference Computer Vision and Pattern Recognition, CVPR.* 2007.
21. Rittsche J, Tu PH, Krahnstoeve N. Simultaneous Estimation of Segmentation and Shape. In *Proc Computer Vision and Pattern Recognition*, Washington, DC. 2005; 2: 486–493p.
22. Ali S, Shah M. Floor Fields for Tracking in High Density Crowd Scenes. In *Proc European Conf Computer Vision*. 2008; II: 1–14p.
23. Rodriguez M, Ali S, Kanade T. Tracking in Unstructured Crowded Scenes. In *Proc IEEE Int Conf Computer Vision*, Kyoto, Japan. 2009; 1389–1396p.
24. Cheriyadat AM, Radke R. Detecting Dominant Motions in Dense Crowds. *IEEE J Select Topics Signal Process.* 2008; 2(4): 568–581p.
25. Andrade EL, Blunsden S, Fisher RB. Modelling Crowd Scenes for Event Detection. In *Proc Int Conf Pattern Recognition*, Washington, DC. 2006; 175–178p.
26. Mehran R, Oyama A, Shah M. Abnormal Crowd Behavior Detection Using Social Force Model. In *Proc IEEE Conf Computer Vision and Pattern Recognition*. 2009; 935–942p.
27. Helbing D, Molnár P. Social Force Model for Pedestrian Dynamics. *Phys Rev E*. 1995; 51(5): 4282–4286p.

Cite this Article
Joshi Kinjal V, Patel Narendra M, Shah Mehul B. A Review on Crowd Analysis Techniques. *Journal of Artificial Intelligence Research & Advances.* 2019; 6(2): 119–126p.