J. Comp. Sci. Comm., 1(1)1-9, 2015

ISSN 2356-8844
Journal of Computer Sciences and Communication
www.textroad.com

Survey of Holistic Crowd Analysis Models

Muhammad Taimoor Khan[1], Armughan Ali[2], Mehr Yahya Durrani[3], Imran Siddiqui[4]

[1]Department of Computer Science, Bahria University, Islamabad, Pakistan
[2,3]Department of Computer Science, COMSATS Institute of Information Technology, Attock, Pakistan
[4]Department of Computer and Software Engineering, Bahria University, Islamabad, Pakistan
Received: February 7, 2015
Accepted: April 1, 2015

ABSTRACT

The behavior analysis techniques used in computer vision, are mostly targeting individual's behavior. The study on crowd analysis is more focused on counting individuals and devising a management plan for load balancing for vehicles and pedestrians. The recent advancements in vision based techniques have allowed the study of collective behavior of the crowd for valuable information. It targets the application areas where crowds are dense making it impossible to segment individuals separately due to severe occlusion. Since object detection and identification techniques only works in low density crowds, therefore, holistic crowd analysis techniques are considered for this study. Holistic approach does not attempt to separate out the crowd and rather consider it as a single entity. It studies the behavior of the crowd, instead of the individual's behavior. Recently available techniques for holistic crowd analysis are considered for this study. The working of these techniques is described and a comparative analysis is given to highlight their strengths and weaknesses.
KEYWORDS: crowd analysis, crowd behavior, global behavior, escape and non-escape activity

I. INTRODUCTION

Crowd analysis is an area of interest to scientists in the past. Most of the work in this area is targeted towards people counting in a scenario or management of pedestrians at train stations and platforms in peak hours. Traditional anomaly detection techniques were aimed at a single person or a moving object such as belongings dropping, loitering and jumping over fence etc. These techniques assume only one or few separable individuals in the image view and therefore, fail to produce reasonable results in crowded scenes due to severe occlusion. To detect the behavior of a single person, one or multiple cameras are installed at different locations. Image frames from the video are analyzed collectively, to identify and segment human objects from them. Single person behavior has various applications in Human computer intersection (HCI). Since one person can be occluded in different ways, person identification is a concern. An extension of the same techniques, used for a single person, is not suitable for dealing with a crowd situation. Zhan et al. in [1] supported this point in his work, claiming that conventional computer vision techniques are not appropriate, when dealing with a crowd situation. This task is more complex when behavior of a crowd is analyzed, as the objects are more occluded in a crowd scene. Detection or identification of crowd is of key importance towards analyzing their behavior.

Overcrowding of vehicles have been studied more as compared to crowded scenarios for pedestrians. This situation may arise, when buildings are evacuated. The aim in such situation is to avoid trapping and get everyone out peacefully by distributing the load on multiple exits. Pedestrian facilities may also exceed the number of people outside stadiums, creating the situation of a crowd. Daamen et al. in [4] has worked on crowd analysis using parameters that drive the behavior of a crowd. They include age, number of disabled people, illumination conditions and the level of panic. Krausz in [5] has discussed two different types of crowds. It can be due to the inflow exceeding the spatial capacity of the people or two opposing streams of people clashing into each other. They have shown through a graph drawn between velocity and number of the people in a crowd. It reveals that in crowd situation velocity drops down to less than 0.5m/s while the average velocity is 1.34m/s.
Crowd situations has severe occlusion, therefore, it is very hard to identify and segment out individuals in a scene, to study their behavior. The individual identification and behavior analysis techniques did not produced better results in crowd scenarios. From the study of Psychology it is believed that when people group together, they lose their personal traits to some extent and adopt that of the group as a whole. Due to this reason, the later research work in crowd behavior analysis were all focused on considering the global crowd or groups of local crowds as individual entities. Applications of crowd behavior are crowd management (in an event) at shopping centers, stadiums and other crowded places, to identify strangle points and balance the flow of people by splitting the crowd to various exit paths, identifying disasters etc. Since computer vision techniques are more sophisticated, more indept study of the crowd is carried out, focusing on the understanding of crowd behavior and attitude. It can be used for security purposes to identify unusual behavior or anomalies in the crowd behavior that can be alarming. It can predict an understanding of how peaceful or aggressive a crowd is.

[*]**Corresponding Author:** Muhammad Taimoor Khan, Department of Computer Science, Bahria University, Islamabad, Pakistan. taimoor.muhammad@gmail.com

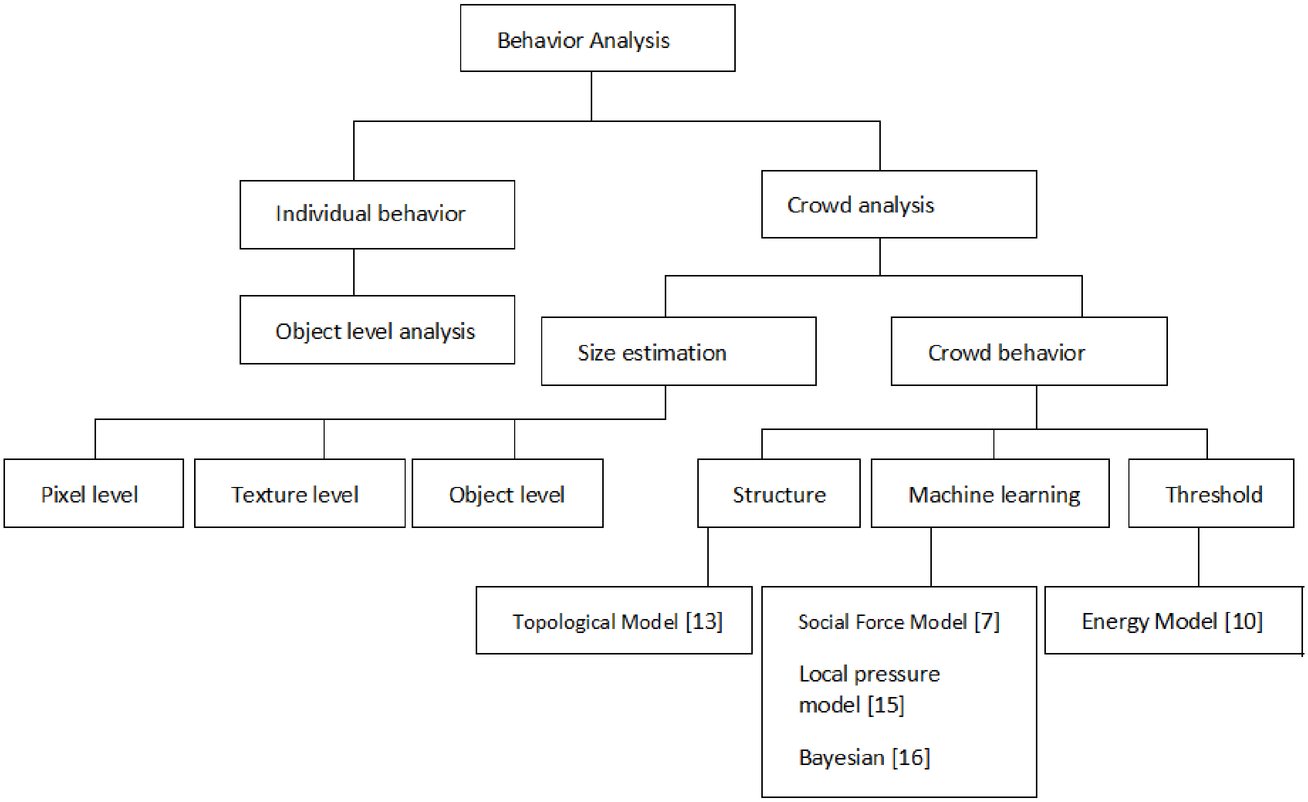


Figure 1: hierarchical distribution of techniques included in the study

The behavior of a crowd is very complex to deal with, as it is not a simple sum of people. Due to severe occlusion, a crowd situation is hard to segment, where occlusion increases as the number of people in the crowd increases. This limitation complied researches to treat crowd as single entity and not as a collection of individuals. The goal of crowd analysis techniques is to make use of computer vision techniques to extract some kind of information from a crowded video sequence that can be used to benefit a large number of applications. There is so much variation in the real world scenarios, which makes it very hard to compare it with any pre-known situation. Some researchers from the computer graphics domain are more interested in simulating such crowds, which brings the two areas closer to each other. Boghossian in [2] in his work has referred to crowd as a situation, in which individual and group of people are hard to identify. Silveira et al in [3] has mentioned some sociological and psychological aspects of group behavior that are following the least effort path and moving in lanes.

Two types of approaches are used for crowd analysis, that are machine learning based techniques and threshold based methods. Machine learning based methods using classifiers perform consistently in situations where normal activities are well defined and constrained. Unknown normal behavior will also be classified as abnormal behavior if does not exist. Threshold based methods are applied on activity sequence data, where the activity exceeds a preset threshold value. Different local monitors are placed on the image frame to perform statistical computations. Threshold based methods are easy to implement but it is hard to set a realistic threshold value. False alarm rate is usually high. Figure 1, explains logical distribution of the techniques involved in the study.

This review paper is organized as follows. In section 2, datasets that are commonly used for the crowd behavior analysis are discussed. Section 3 has a discussion on the recent crowd analysis techniques. Conclusion and future work is placed in section 4.

II. DATA SOURCES

The UMN dateset in [9] has been majorly used as a dataset for crowd videos to test their models. Mehrean et al. in [7] used a dataset from UMN database containing 11 videos of different escape events. These events has samples from both indoor and out door scenes. The activities provided in these videos starts with a normal activity in the initial frames of the video, whereas, a sequence of abnormal behavior is shown towards the end. Social force model is tested on web datasets of Getty Images and thoughtEquity.com. They contain high quality documentaries. These documentaries has more realistic crowd environments that is people walking, running marathon as normal activities along with the abnormal behaviors shown through fights and protests. Social force model showed promising results in identifying group behavior. Li and Zhang in [13] also used UMN dataset for training their model. They also employed another dataset PETS 2009 in [14]. Yang et al. in [15] used UMN dataset for training their model, called pressure model. Wu et al. in [16] tested their model on UMN dataset.

III. Holistic crowd analysis models

Recent models of crowd analysis follow a holistic approach in which the crowd is considered as a single entity. These models focus on identifying the behavior of the crowd as a whole. It is currently considered as a binary classification problem, in which the behavior can either be normal or abnormal. These models are discussed in detail.

a) Social Force Model

Helbing et al. in [6] introduce a model to detect abnormal crowd behavior based on the interaction between the individuals. Social force model considers abnormal crowd behavior as a state of interactions force in a crowd. Likely behaviors in the crowd are modeled, which help to separate abnormal social forces from within the crowd. In order to learn the interaction of an individual in social force, they have to be detected and localized. This technique works well in identifying abnormal behavior in crowd, where there is less clutter and occlusion among individuals. As the density of the crowd increases, the process of distinguishing individuals gets less conclusive.

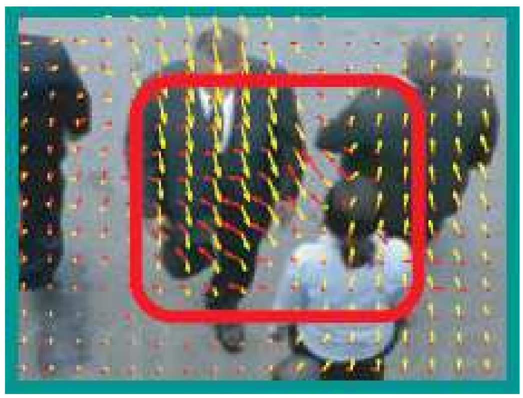


Figure 2: Grid of particles placed over the image showing social interaction [7]

Mehran et al. in [7] proposed a modification of the social force model in which they avoided to keep track of individuals and rather considered a holistic approach towards the crowd. Particle advection method is incorporated to detect the orientation of the crowd group as a whole. In this technique a grid of particles are placed on the image and they are made to move with the flow of the underlying field as shown in Figure 2. The interaction force that makes the basis of social force model is calculated between these moving particles. The interaction forces are mapped to image frames and any change in the interaction of the particles, provide information about the crowd behavior. Force flow vectors are calculated for the particles in frames and are labeled as per their behavior orientation. They are added to the bag of words, to help classifying unknown scenarios. Ali et al. in [8] also made use of particle advection technique to study the behavior of the crowd, however, their approach was based on the segmentation of high density crowds. Segmentation was carried out on the principles of coherent structures in fluid dynamics. The social force model is being depicted in Equation 1, where m_i the mass and v_i is the velocity of an individual i.

$$m_i \frac{dv_i}{dt} = F_a = F_p + F_{\text{int}} \qquad \textbf{Equation 1}$$

The force F_a that an individual i uses to maintain velocity v_i. This force consists of the two overlapping forces that are personal desire F_p to move in a particular direction with a certain velocity and F_{int} the force applied through the interaction. The personal force is being limited by the interaction force in the crowd situation due to obstruction by crowd and other environmental factors.

Figure 3: Original image along with social forces mapped over [7]

This compulsion by the environment and neighborhood bound individuals to move with the average velocity of the group. The author has resembled it with the motion of leaves over flowing water. The forces calculated through the grid of particles on a single frame is inconclusive towards identifying the behavior of a group of people in the crowd. However, estimating them continuously for a sequence of frames gives more in-dept information about the social forces and interaction by groups of people within the crowd as shown in Figure 3. It has the original image on the top left with the social forces calculated for the particles over a series of frames. The red color shows higher values of interaction force while the blue one represents lower values.

Yang et al. in [15] proposed local pressure model that uses social force model to calculate local pressure from local crowd characteristics. A Histogram of oriented pressure (HOP) technique is proposed, that works on the local pressure. SVM classifier and median filter are used to detect abnormal behavior in the crowd.

b) ENERGY MODEL

Xiong et al. in [10] performed crowd behavior analysis based on the energy model. The model works on two types of energy that are kinetic energy and potential energy. Potential energy is used to estimate the density of the crowd while kinetic energy is used to measure flow of the crowd. Histrogram is built for the foregrounds on the X, Y plane to get the probability distribution of the histogram. This distribution helps to calculate the crowd distribution index (CDI) to quantify dispersion in the crowd. An abnormal behavior is considered to be the one that has kinetic energy above the threshold that is more people are running or moving fast in an image scene. A complete working framework of energy model is given in Figure 4.

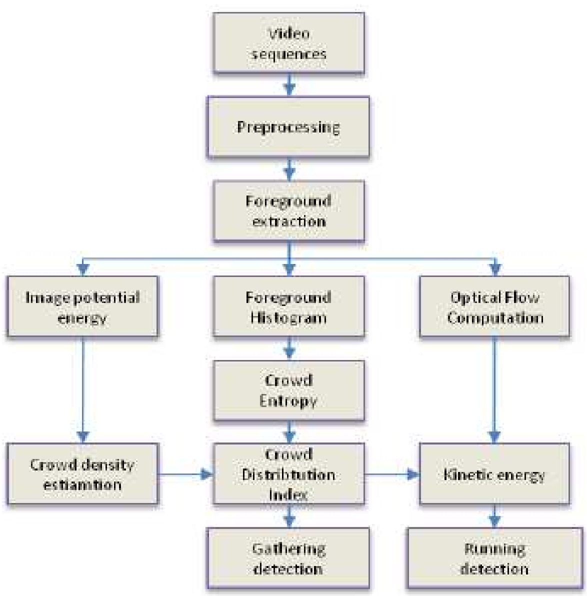


Figure 4: working framework of Energy Model [10]

In order to avoid the problem of occlusion, as is the case in a crowd situation, individuals are not tracked. The system works in 3 steps, that are, to estimate crowd density, crowd distribution index and then its kinetic energy. They used Gaussian mixture modeling (GMM) to separate foreground from the background and is exposed to binary morphology for noise removal. Potential energy model proposed by Xiong et al in [11] is used to calculate the crowd density as shown in Equation 2. It has m_{ij} representing mass of the pixel which can be either 0 or 1 since it is applied on the binarized image. The constant g_{img} is the gravitational potential energy. H is rough measure of the closest distance from the image, Y is height of the frame while y_{ij} are the coordinates of the image. Histograms are build for the foreground and crowd distribution index is calculated through a combined mechanism of crowd entropy and crowd density. The kinetic energy based technique, used Zhong et al. in [12] is employed with modification to be more effective towards abnormal activity detection.

$$E_p = \sum_{i=1}^{X}\sum_{j=1}^{Y} m_{ij} g_{img} (H + Y - y_{ij})$$ **Equation 2**

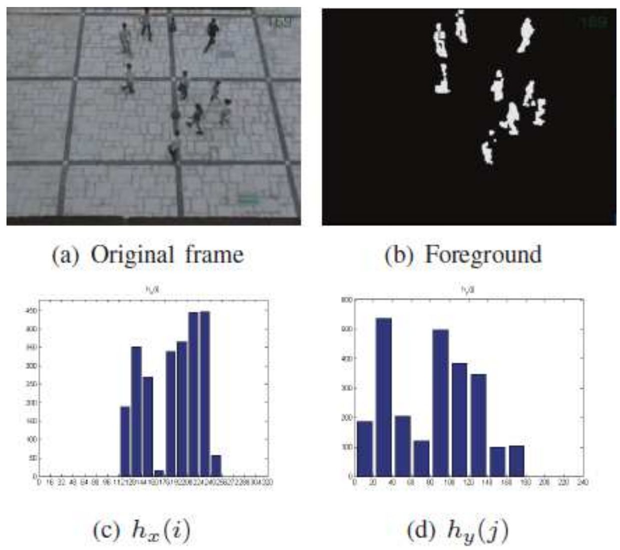


Figure 5: Original Image, Binarized and X,Y histograms [10]

In order to calculate crowd distribution index (CDI), histogram of foreground is built on X and Y axis separatly, to get histogram bins for horizontal and vertical axis. The size of these bins depends on the distance of the camera from the crowd scene and the number of people involved.

The stepwise activity is shown in Figure 5, where the original image is binarized and the X,Y histograms are calculated. The image has more activity towards the center of the X-axis and start of top of the Y-axis, as shown by the histogram. The foreground information is fed into the crowd entropy technique to estimate dispersion along x and y axis. Crowd Entropy is calculated along x and y axis, whose product is equal to Crowd dispersion (D) as shown in equation 3.

$$D = H(X) * H(Y)$$ **Equation 3**

Crowd dispersion helps to calculate Crowd distribution index (CDI) as shown in equation 4. N represents the number of people in the crowd, which is inversely proportional to the dispersion of crowd.

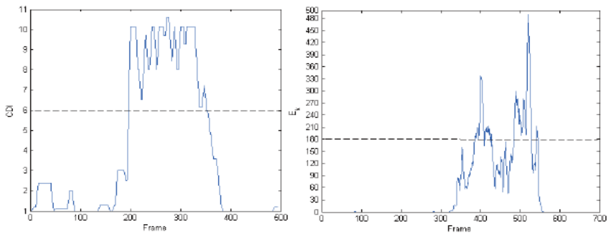
$$CDI = \frac{N^2}{D^3}$$ **Equation 4**

Motion vectors are extracted of a series of images by tracking the image features. Kinetic energy is calculated for each frame using the formula given in equation 5.

$$E_{kn} = CDI * \sum_{i=1}^{m} v_i^{\,2}$$ **Equation 5**

A threshold is applied on the values of CDI and E_{kn} to detect abnormal activity, which shows either too many people gathering or people having high kinetic energy that is moving fast. This crowd analysis model is very efficient and does not need any training data.

This technique make use of clustering groups of individuals in a crowd scene and then apply thresholding on the activity based on statistical calculations. Activity above threshold is considered to be an abnormal activity. It does not require any training set. However, choosing a suitable value for theshold is critical. It only detects two types of abnormal situations that are, either too many people gathering (CDI goes above threshold) or too much movement (Kinetic energy goes above threshold) as shown in Figure 6.

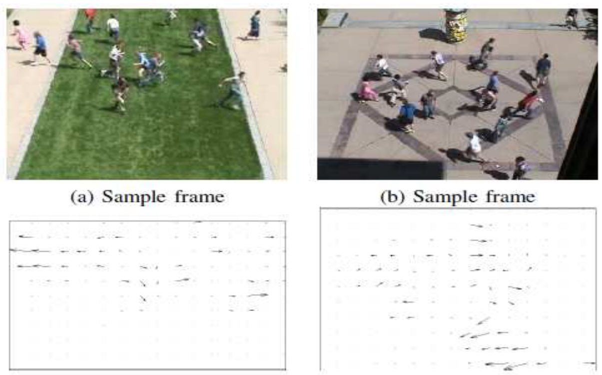


(b) CDI curve (b) E_k curve

Figure 6: Thresholding CDI and Kinetic Energy [10]

c) TOPOLOGY METHOD

This method is modeled over global behavior of the crowd motion. A topological structure is obtained from the particle motion field which is monitored for change. A change in the topological structure of the crowd is marked as an anomoly in the crowd behavior and is considered as abnormal behavior. Since the decision of considering an activity as abnormal is dependent on change in the topological structure which is extracted from the video, therefore, it does not require any classifiers and training data. Li and Zhang introduced the topology based behavior detection in a crowd in [13].

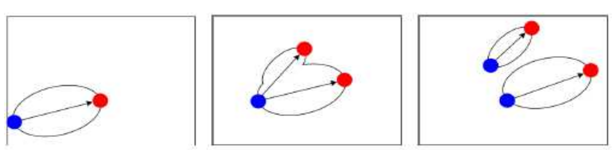


(a) Sample frame (b) Sample frame

(c) The optical flow field (d) The optical flow field

Figure 7: Same frames and their optical flow fields [13]

Topological simplification method is used to simplify and visualize 2D and 3D velocity vector fields. Fluid mechanics does not have too much activity at the boundaries and are ignored. This is not the case with crowd analysis where boundaries are not stationary and could be potential critical points. Points that have magnitude of the corresponding vector fields vanishing, are critical points, also called singular point. Figure 7 shows image frames from the video and the calculated vectors fields for them. Virtual critical points are identified from the boundary points through limit sets. Limit sets are obtained through numerical integration of the particles in the motion field. These critical points, which are the spatial points in the image frame makes the building blocks of the topological structure. All the critical points identified are not potentially critical, therefore, a threshold is applied that remove critical points with velocity close to zero. They are neglected as stationary critical points. Then, the image domain is partitioned into grids using separatrices, where each section has separate distinguishable flow fields.



(a) At time t_1 (b) At time t_2 (c) At time t_3

Figure 8: Source and Sink mappings [13]

The critical points in the partitioned image is concluded to be either sources or sinks. In the crowd a gathering point is considered to be a sink, while the dispersion point is labeled as source. The sources and sinks are mapped through curves, based on the trajectories formed between them through crowd motion, as shown in Figure 8. A topological structure is formed from the events in the crowd video using these parameters placed on the image frames.

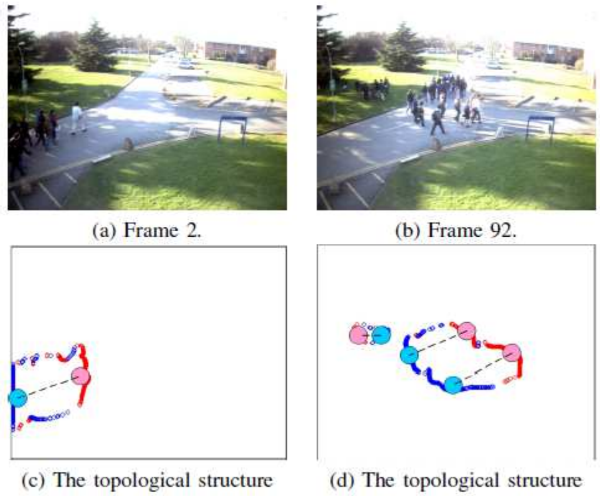


Figure 9: Image frames with their topological structures [13]

The topological structure obtained is monitored for a change, which is considered to be an abnormal behavior on part of the crowd. This technique captures the group activity, therefore, it calculates global behavior of the crowd. It is invariant to continuous transformation and is therefore, more in-sensitive to noise. This technique outperforms other techniques used in the experiment which are optical flow and social force while falling behind chaotic invariants. However, it is very efficient computationally. Figure 9 shows two different image frames and their corresponding topological structures formed.

d) BAYSIAN MODEL

Wu et al. in [16] has used bayesian model to detect abnormal behavior in crowd situation. Since people behave in abnormal way in panic situation, therefore, this information can be used for the timely detection of disasters like fire, explosion or transportation disasters. The model considers two types of behaviors that are non-escape and escape. In normal circumstances people tends to follow the group that are heading towards their desired destination, as non-escape. Anything that deviates from the normal behavior is considered to be abnormal behavior as in disaster situation, causing crowd stampede. Bayesian model considers crowd as a single entity and studies its behavior. It is based on the detection of an escape situation in the crowd to distinguish between normal and abnormal behavior.

In order to identify the crowd behavior in a video, the sequence of frames in the video are used to identify the optical flow. The frames are divided into multiple patches, where some patches represent the foreground optical flow activity. Movement in each patch is considered as a single moving object, since they are treated as separate entities. The class conditional probability density functions are used to estimate the position, magnitude and direction for both escape and non-escape situations. The working mechanism of bayesian model is shown in Figure 10. The two key steps in the process are motion direction model and parameter model. It estimates potential distination locations in different directions outside the optical field of view. The probabilities of these potential destinations are calculated for the training data. People in crowd are considered to deviate from these destination points in escape situations. In equation 6, gives a relationship between foreground positions, optical flow field and their respective potential destinations. In order to identify the flow of direction, divergent points are placed based on the potential destinations. Most of the patches close to the divergent centers is considered to be a non-escape situation where as a larger distance from them shows an escape scenario. The relation used to find distance of the foreground patches, their position and movement from the divergent centers is given in equation 7.

$$\hat{d}_j = \| x + v - \hat{x}_j \| - \| x - \hat{x}_j \| \qquad \textbf{Equation 6}$$

Since it is not possible to train for all or most of the escape and non-escape scenarios for classification. Therefore a probability based bayesian unsupervised approach is used for calculating the key parameters that are flow direction, flow magnitude and flow position. The magnitude at a position shows its probability as a foreground patch. Since the optical flow is unstable, therefore k-mean clustering is employed with k=2 for separating the escape and non-escape flows. Only 10% of the field vectors are randomly selected in training dataset for clustering. Most of the parameters are constant in the model that are frame resolution (320x240), patch size, potential destinations and divergent centers.

$$\tilde{d}_l = \| x + v - \tilde{x}_l \| - \| x - \tilde{x}_l \| \qquad \textbf{Equation 7}$$

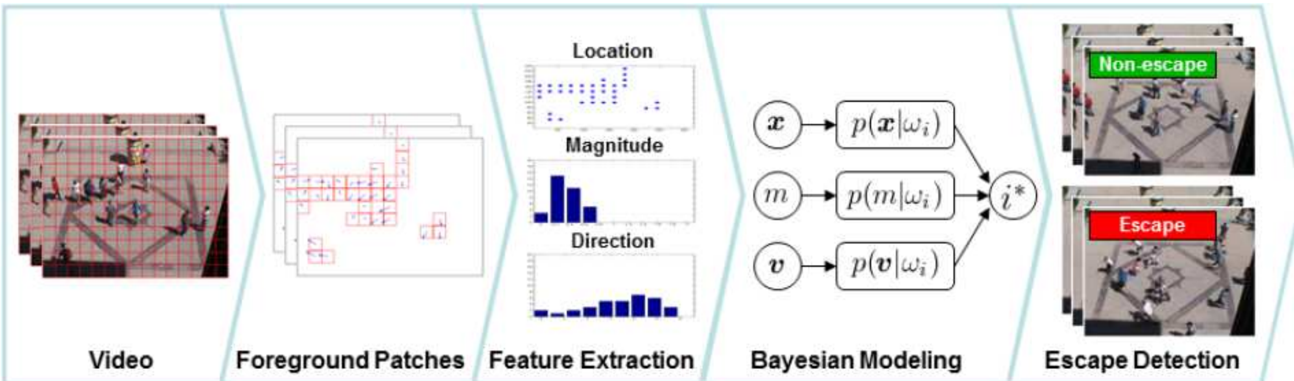


Figure 10: working model of bayesian model [16]

IV. Conclusion and future work

The models included in the study considers a holistic perspective of the crowd to detect its behavior. Social force model requires training through which the model learns the difference between the two types of behaviors. A trained model can distinguish between the normal and abnormal behaviors even from different scenes for which it is not trained. Its computation cost is high as it places particles over the image frame for the sequence of frames throughout the video. As compared to it, the energy model is computationally simple. It can also successfully differentiate between normal and abnormal behviors for unseen scenarios. However, its accuracy drops as the number of people in the crowd increases. It shows promising results for low and medium level crowds only. Energy model make use of two parameters to detect abnormal activity which are crowd distribution index and their kinetic energy. These parameters are not enough to accommodate global abnormal activities and are only focused on few relevant cases.

The topological model uses a combination of threshold based and machine learning based techniques for predicting abnormal crowd behavior. It does not require the model to train as the structure is extracted out of the video frames. Bayesian model on the other hand works on probabilities. The probabilities are calculated to get constant values that form the basis of identifying escape scenarios. The probabilities are performed on the videos provided in the training dataset. The training dataset used for bayesian model contained only a very limited range of scenarios. K-mean clustering is applied with k=2. The parameters used for differentiating between these behaviors is distance of crowd from divergent centers and their movement towards the potential destinations. They are calculated through probabilities.

For future work more parameters should be considered to improve the performance of these techniques. The definition of abnormal behaviors also need to be extended to accommodate more types of behaviors.

Table 1: Comparison of the techniques included in the study

	Social Force Model	Energy Model	Topological Model	Bayesian Model
Technique	Classification	Threshold-based	Structure	Clustering / Probability
Dataset used	UMN Dataset	Web Datasets GrettyImages thoughtEquity	UMN Dataset PETS 2009	UMN Dataset
Driving factor	Interaction force	Density / Movement	Change is structure	Distance from potential destinations / divergent centers
Abnormal behavior	Close encounter	High density / High movement	Too much regrouping	Moving away from PD/ Greater distance from DC
Computation	High	Low	-	-
Require Training	Yes	No	No	No

REFERENCES

[1] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, "Crowd analysis: A survey," Machine Vis. Applicat., vol. 19, no. 2, pp. 345–357, 2008.

[2] B. A. Boghossian and S. A. Velastin, "Motion-based machine vision technique for the management of large crowds," in Proc. 6th IEEE Int. Conf. Electronics, Circuits and Systems, 1999, vol. 2, pp. 961–964.

[3] J. C. Silveira, S. R. Musse and C. R. Jung, "Crowd analysis using computer vision techniques", IEEE Signal processing magazine, 2010.

[4] W. Daamen and S. Hoogendoorn, "Capacity of doors during evacuation conditions", Procedia Engineering, 3:53 – 66, 2010.

[5] B. Krausz and C. Baukhage, "Analyzing pedestrian behavior in crowds for automatic detection of congestions", IEEE International Conference on Computer vison workshops, 2011.

[6] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics". *Physical Review E*, 51:4282, 1995.

[7] R. Mehran, A. Oyama and M. Shah, "Abnormal crowd behavior detection using social force model", IEEE 2009.

[8] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis", *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–6, June 2007.

[9] Unusual crowd activity dataset of University of Minnesota, available e from http://mha.cs.umn.edu/movies/crowdactivity-all.avi.

[10] G. Xiong*, X. Wu*, Y. Chen* and Y. Ou*, "Abnormal crowd behavior detection based on the energy model", IEEE, 2011.

[11] G. G. Xiong, X. Y. Wu, Y. L. Chen, Y. S. Ou, Y. Liu and Y. S. Xu,"Crowd Density Estimation Based on Image Potential Energy Model", submitted to IROS2011.

[12] Z. Zhong, W. Z. Ye, S. S.Wang, M. Yang and Y. S. Xu, "Crowd Energy and Feature Analysis", in Proceedings of the 2007 IEEE International Conference on Integration Technology, 2007, pp. 144-150.

[13] N. Li and Z. Zhang, "Abnormal crowd behavior detection using topological methods", IEEE Computer Society, 2011.

[14] J. Ferryman and A. Shahrokni, "Pets2009: Dataset and challenge," dec. 2009, pp. 1 –6.

[15] H. Yang, Y. Cao, S. Wu, Weiyao Lin, S. Zheng, "Abnormal crowd behavior detection based on local pressure model," Asia Pacific Signal and Processing Association Annual Summit and Conference (APSIPA ASC), 2012

[16] S. Wu, H. Wong and Z. Yu, "A bayesian model for crowd escape behavior detection", IEEE Transaction on circuits and systems for video technology, 2013.