Research Article

Crowd Behavior Monitoring and Analysis in Surveillance Applications: A Survey

Ajay Kumar[1], Arunnehru J[2]

[1,2]Department of Computer Science and Engineering,
SRM Institute of Science and Technology, Faculty of Engineering and Technology, Vadapalani Campus, Chennai, India
E-mail ID : ajayku87@gmail.com[1], arunnehru.aucse@gmail.com[2]

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

Abstract: In the research field of computer vision, crowd monitoring and analyzing the behavior is an open topic for researchers due to its importance. Over the last decade many methodology has been proposed to do these task. These methodologies supposed to perform various tasks for the crowd which includes finding the strength of crowd in number for the proper crowd management in time or for the security reasons, prediction of future behavior of the crowd etc. Although many complex methodologies have been implemented for analyzing the crowd but there is still open scope for the methodologies which analyze the crowd in real-time, especially for the non-organized crowd. This paper presents a literature survey of the methodologies, proposed for the crowd monitoring and behavior analyzing for the both organized crowd and non-organized crowd. We also included the dataset details, used for those proposed methods with advantages and disadvantages. We included the methodologies based on traditional approaches as well as modern deep learning concept. We have a faith with a motive that this paper will help in the research community to understand about the various state of art methodologies used for crowd monitoring and analyzing.

Keywords : Crowd Behavior, Deep Learning, Computer Vision, Crowd Management.

1. Introduction:

In our daily life, almost every individual face the crowd, gathered for various reasons such as cultural events, sports event, or one can face the crowd at airport or railway stations while going for the journey. In some cases like sports event one can avoid the crowd but in some situations like in cultural events, it is very hard to avoid to getting the crowd experienced. So in the crowded area there is always a risk of life of individuals due to unwanted activity, stamped. So it is important to have some mechanism formonitoring the crowd and analyzing their behavior over the time period for the safety of an individual in the crowd. Analyzing the big crowd manually is a very complex task and time consuming. So we can have a computer vision based system that can perform these tasks. Over the last 10-15 years, in the field of crowd behavior analysis, many researches have been done with low level success. So there is a broad scope in the research field in this area. This paper includes a survey of the work done in the field of crowd behavior monitoring and analyzing System (CBMAS).

In broad perspective, CBMAS performs managing and monitoring the crowd, analyzing the crowd behavior, have number of possibilities in different applications. It means CBMAS not limited to safety domain only but it also includes counting the total number of individuals in crowd, on traffic signal, emergency services etc. So such broad field of monitoring and analyzing the crowd motivate the research and development in this field for further development. The article [1] explained the three approaches for counting the individuals in the crowd as object based, regression based and clustered counting.

The dynamic CBMAS has got tremendous attention over the past 10 years. Behind this attention mainly two reasons can be identified. First, the task in the field of image analysis can get benefit from the management system which manage the huge crown like Hajj and Kumbh[2-5].

Second, although some development on low level has been done so far but still there is lot of challenges, specifically video taken from unmanaged crowd, to be solve.

A group can increment in a matter of seconds and controlling the group can turn out to be very difficult for the management. In such scenario, any abnormal activity[6] can cause the damage on large scale like stampedes. The crowd monitoring can be done using drones with cameras i.e. Unmanned Arial Vehicle(UAV) and CCTV. But with the CCTV, there are a number of limitations like video capturing area, static location etc. These limitations can be overcome using UAV. The UAV proved some high resolution images and can cover larger area and can change its location in real time.The authors in [7] explained a framework for crowd monitoring with the help of UAV and face recognition.Likewise the authors in [8] suggested a different model with UAV for monitoring the crowd. The authors adopted color segmentation approach for the identification of the pedestrian. Theauthors' approach was capable for the identification of the Hajj crowd correctly. The element of feature vector is back or white or any one or two with mainly white and black. This makes the task easy to identify for the system.

The authors in [9] suggested a deep CNN based technique for counting the crowd with the help of Internet of Thing (IoT) concept. The model was developed particularly for Arabians country and was capable to count the people in crowd in low and high density crowd.

2. Applications

A number of applications dependent crowd monitoring systems. There are many applications used in crowd monitoring activity and some of them summarized in the following figure-1

Figure 1. Applications of crowd monitoring and management system.

The task depending on CBMAS summarized following

- Counting the people in high density crowed area: Over the time, the population growth rate in the world has in increased very rapidly. And hence we are getting more crowds in public places like airport, railway stations, mall etc. So monitoring the crowd has become an essential task to maintain the predefined orders on such places. In CBMAS system finding the strength of crowd in number is a major task. Specifically in smaller region, the increment of density of crowd may create many major problems like physical injury, suffocation, even fatalities sometimes etc. So getting the count of people in small area in beginning can help to avoid such incident. In this, based upon the number of people present in that area, some mechanism like blockage can be implemented to stop the people from entering in that area. Despite of large number of research in this area, various crowd counting techniques encountering number of challenges like lighting situations, occlusion challenges etc. Due to continues development in CBMAS architecture, problem in crowd counting has reduced to somewhat extend. A number of approached are suggested in [10–12], that solve the challenges in crowd counting using an efficient CBMAS.

- Public gathering management: Events like music concert, election rallies, and sports meet etc., where the huge number of people gathers together. The CBMAS is beneficial on such places for analyzing the crowd behavior and to avoid some dangerous situations. This system is benefited particularly in analyzing the space capacity for the crowd and in the movement of the crowd [12-15]. Like that managing the crowd in religious festival like Kumbh or Hajj is a tedious task. In Kumbh there is a special ritual to take bath on some specific day and time, so there is more probability for huge crowd. In these situations this system is very useful to track the crowd movement for avoiding any kind of disaster.
- Defense Application: During any war, for making a strategy, officials need the strength of opponent soldiers. So this system can be filleted in fighter jet or in drone and one can get the strength of opponent soldiers and also can get the information of opponent movement.
- Disaster Management: We have number of disaster example in past which caused due to huge crowd in music concert and sports meet etc. In such places some time some group of people in crowd try to change their direction and this reason causes some life threatening situations. Even some time it causes death in huge number because of suffocation in crowd. So with the help of proper crowd management, this can be avoided [16-18].
- Detection of abnormal activity: CBMAS can be used to detect any suspicious activity on public places to reduce the terrorist attack. The existing machine learning algorithms not perform well in such situations. For this particular scenario where poster of an individual required, some approaches a explained in [19-20].
- Safety Monitoring: We can provide a better crowd monitoring system by installing in large number of closed circuit television (CCTV) at crowded places like airports, railway stations, etc. for public safety purpose. For example, the author [21] provides a model which tracks the crowd strength, movement and decides the time slot for each group to get entry in main area. Likewise the author [22] proposed a model which inspects the suspicious activity based on the crowd density.

3. Motivations

The effective CBMAS contributes in many applications and more over having further extension with Computer Vision concept. However managing the live crowd is completely different from what has been solved so far specially with non-organized crowd. The systems encounter number of challenges when it comes to work with a non-organized crowd. So far published literatures suggest some success stories in this field and some research also has been done specially with organized crowd. But still there is number of challenges to be solved in this field especially with non-organized crowd. Analyzing the non-organized crowd is still an open challenge in the field of research due to its unpredictable nature. The CBMS depends on many factors such as light condition, occlusion, different forms of noises, head positions etc. Moreover for developing a new model in research field, the model must be trained with large dataset to improve the efficiency of that model but in crowd management field very less number of data set available publicly. So following are the some challenges we encounter in this field:

- At the point when at least two items approach one another and therefore converge, in such situations, it is difficult to perceive each object separately. And hence, observing and estimating the precision of the framework gets troublesome.
- A non-uniform kind of course of action of different items that are near one another is looked at by these frameworks. This course of action is called Clutter. This Clutter is firmly identified with picture commotion which makes acknowledgment and checking all the more testing [23].

- In non-organized crowd the density of crowd is not equally distributed everywhere, which make the challenging situation for the system.
- Aspect ratio is another main challenge, encounter during the crowd analysis. For the system implementation, camera fixed with drone and drone fly over the crowd. To fix the aspect ratio issue, drone fly on some particular height from the ground level and camera lens is focused towards the ground. This provides the top view images and videos of the crowd under surveillance. And this top view content in the form of image and video creates the aspect ratio problem.

Any task that belong to machine learning field is solved by developing a model after proper training and testing. So size of data set is an important factor for these models to train and test and to get the success in that task. For CBMAS development, the availability of data set is crucial factor for its success. For finding the strength of crowd in number, datasets are in stock in public domain but for the behavior analysis type dataset are not available in plenty. Additionally over the decade many number of researches have been done in the field of crowd monitoring system and many methodology, models have been proposed and developed but still there is large scope left for the further enhancement for more accuracy and efficiency in the existing models. These challenges, issues, and cutting edges variables motivates to a researcher for development of the more accurate, efficient model for analyzing the behavior of organizedor non-organized crowd.

4. Contributions:

This paper is a detailed literature review of crowd behavior analysis and managing system and includes the methods for organized and non-organized crowd. We present the benefits and negatives of state of art approach by looking in on original work, and afterward finishing state of art techniques that belongs todeep learning structures. A comparison task in the previous proposed models gives us a road map for the future research topic. We have a faith that a review paper in particular field provides a lot of idea among that particular domain and provides a concrete concept of the ongoing research progress which helps to others who doing the research in that particular domain.

Our proposed paper includes the literature review of crowd behavior analysis and managing field over a decade. We focused mainly on state of art crowd monitoring systems which have been in discussion over a decade. We also tried to point out the shifting paradigm in crowd behavior analysis and managing system from a traditional to deep learning approaches.

Here after our proposed review paper includes detailed information of available dataset for CBMAS in section 5. The section 6 includes the existing proposed methodology for CBMAS. The comparison between the existing states of art methods included in section 7. And at last in section 8 we concluded our discussion and included the future scope in the CBMAS.

5. Databases

The exhibition of the CBMAS is assessed with accessible crowd datasets. The crowd behavior analysis and monitoring filed is comparatively very less discussed filed with very less accessible dataset. Due to less number of scenes in most of the available dataset, cannot utilize to understand the general behavior of the crowd. The following section includes details of the available datasets that can be used for the development of crowd behavior monitoring and analysis system.

- Mecca [24]: The data set Mecca captured during the Hajj at Saudi Arabia holy city Makkah. Hajj is a very important and mandatory duty in Muslim religion. In very large number of Muslim pilgrims pray during the each year of Hajj time near the central Kabba. The available video clip of this dataset is of 10 minutes long. The video cut records explicit period when travelers enter the Kaaba and possess the spot. The video capturing cameras are placed in three directions includes north, west and east. The output video gets synchronized since this is captured by three different cameras from three different directions. The date, time is also recorded besides of other information. This data set includes 480 frames.
- Kumbh Mela [25]: Kumbh is a very huge crowded Hindu festival which comes in every 12 years. Taking bath in Ganga River is so important during the Kumbh time. Hindu religion people across the

country come to take bath near to Allahabad where three rivers meet together Ganga, Yamuna and invisible Saraswati. In the last Kumbh in 2013, approximately 0.12 billion took part in the Kumbh holy festival. During this festival top view videos and images were captured with the help of drone cameras. The size of this dataset is very large and the captured video is of length approximately 6 hours in which 600K frames exists. The Kumbh dataset can be used for finding the crowd strength in number as well as for analyzing the crowd behavior.

- NWPU-crowd [26]:In current data set scenario, many data sets are of very small sizes and they do not satisfy the requirements of the DCNN methods. Recently a new dataset introduced in public domain which is of large size and can also satisfy the requirements of DCNN methodology. The dataset NWPU-crowd includes approximately 5000 images in which 2133375 individuals are identified. The crowd density in this dataset in very high with so many illumination variations with respect to other available datasets. Content of this dataset has collected from internet as well as self-shooting. A very different approaches has adopted to collect the data like images from railway stations, malls, resorts etc. are available in this dataset.
- Beijing Bus Rapid Transit (BRT) [27]: The available 1280 images in this dataset are captured by surveillance cameras from different bus stations in Beijing. Out of 1280 images the authors used 720 images for training the system and the remaining images are used for the testing by the authors. For the complexity in the dataset some images with different effects like shadows, images with low illumination, high illumination added.
- UCF-QNRF [28]: This is one of the recently introduced dataset having 1535 images. These images are of huge density variationwith resolution 400X300 to 9000X6000 from state of art datasets. This is one of the largest dataset that preferred by the researchers for implementing the model for crowd counting, localization etc. and it is well suited with the DCNN methods. The biggest source of the content in this dataset is Hajj pilgrims captured footage and web searches. This data set has the maximum number of images with crowd compare to other dataset and moreover object in the images are annotated. There is a huge variety in quality like crowd density, illumination, etc. in the scene. It includes images of all type of buildings, highway, pedestrian etc. in non-organized manner. Because of these many diversities the dataset is very complex and realistic.
- The Shanghai Tech. [29]: This dataset the Shanghai Tech was introduced for the developing a model for counting the strength of crowd. This includes 1198 images whereas 330165 heads are annotated among the given images. Because of large number of annotated objects in this dataset, it is nicely suited for the training and testing of the model. Basically this dataset can categories in two parts. The source of the available 482 images is internet whereas the second part of the dataset contains 716 images which are captured from the major crowded places of Shanghai. The training and testing data ratio is set by the user. The dataset owner tried to give all possible variation in images quality on the scale of crowd density and illumination condition to make this dataset more challenging.The preparation and testing stages are extremely one-sided in the Shanghai dataset as the pictures are of different thickness levels and are not consistent in nature.
- WorldExpo [30]: This is mainly used in cross scene scenario for the crowd analysis. The WorldExpo dataset includes 3980 frames with 576X720 size. In this dataset 199923 persons are annotated. For collecting the videos in this dataset total 180 surveillance cameras were used. For ensuring the diversity in scene the videos captured by the cameras in disjoint bird view. For training the model there are total 1127 videos are available of total length one minute for each whereas for testing the system 5 videos are there of each one length is one hour. Since the data is very less comparatively, this not very efficient for the developing a model for dense crowd behavior analysis.
- WWW [31]: The dataset WWW (Who Do What at Some Where) is mainly designed for analyzing the high density crowd scene. The many different locations like mall, gardens, local roads, airports scenes are used for this dataset. This dataset having 10000 videos taken from 8257 different locations and total number of frames is 8 million. The cited reference paper author suggested 94 features for the better

utilization of the data. Explicit catchphrases are utilized to look for recordings from changed web indexes including YouTube, Pond, and Getty Images.

- UCF_CC_50 [23]: With respect to other dataset this is a complex dataset and there are so many variations in crowd density in this dataset. The data in this dataset captured from many different events like political election campaign, protest and sports. There is total 1279 images are available among them 50 images are annotated. Problem in this dataset is very less number of images can be used for testing the model. The number of individuals in each frame varies in large scale from 94- 4542. Since the data is less for system evaluation, the author used cross-validation methodology. Because of complexity of the data,the results obtained from deep learning methodology are not closed to optimal.
- Mall [32]: As the name suggest, this dataset collected from the shopping mall after installing number of cameras in the mall at different locations. Variations are there in density and activity on large scale in the frame data in this dataset. Static and dynamic nature of crowd video and images are recorded in this dataset.Extreme viewpoint mutilations are available in the recordings, bringing about varieties both in appearance and sizes of the articles. Few obstructions object in the scene are also available like small trees, shopping stalls etc. For the model evaluation the data ratio is pre-defined. For the training purpose 800 frames can be used and the other 1200 frames can be utilized for testing purpose.
- PETS [33]: Compare to all available dataset, this is one of the oldest dataset, whereas on the usability factor this is still remains so important just because of its diversity and complexity in nature. Total 8 cameras were used for capturing the video for this dataset in the premises. This data set can be used for the developing the application for surveillance as well as finding the strength of crowd in number. All the video frames are labeled in this dataset. The movement in this dataset can be categorized of three types, each one having 221 frames equally.
- UCSD [34]: This is the first dataset used for calculating the strength of the crowd in number. This dataset contains mainly video of pedestrians captured by a camera. Each 5th frame is annotated in this dataset. The other frames are annotated using linear interpolations. To avoid the obstacle object like trees, vehicles etc. a ROI is also mentioned. In the UCSD dataset total 2000 frames are available and a total 49885 pedestrians has captured. For evaluating the system, the data ration is defined in which for the training purpose indices starts from 600 to 1399 and for testing it contains 1200 frames. Compare to other available dataset, this is simplest one because of captured from only one location. On average there is 15 persons in a single video. No diversity in the scene viewpoint across the recordings can be taken note.

6. Approaches

The concept of crowd counting gives an idea about the available number of person and/or any objects in the crowd. It does not tell about anything regarding the crowd location. Even density map approach does not provide any large information regarding the objects locations. So for finding the location of person in image or video we can go with localization approach, although it is a complex task. So instead of performing crowd counting, behavior analysis and localization separately, we can simulate the task together with assumption that all are correlated with each other.

In the following section we tried to list out some methodology used for CBMAS, which has been proposed by the some authors in their paper with the references. We also included pros and cons of those methods I our paper.

6.1 Localization

The idea behind the localization process is to indicate the most visible object of an image. In the paper [35], the author used density map approach for the localization process. For getting the optimized result the authors enhanced the objective function which works on density map of the object, obtained from a particular location [36]. A better accuracy and review esteems are gotten with this methodology. For density map generation a Gaussian filter is convolved with the location. In another approach the authors [37] obtain the density map by using a sliding window concept on the image [36]. After getting the density map the authors used the integer programming concept for object localization. On the other hand the author [23] proposed a methodology for analyzing the crowd with the help of crowd strength in number, density calculation, and localization as a variable for the composition loss method. There is an assumption in development in [23] that the above said three factors crowd counting, behavior analysis and localization are correlated. Since the localization process works better on high resolution images, so author used a new UCF-QNRF dataset. Few papers recently published and author of those papers has introduced some methods for detecting the anomaly [38-40].

6.2 Crowd Behavior Detection:

For conducting a crowded event in peaceful manner, now days it is essential to know the behavior of the crowd [41]. So now days analyzing the crowd behavior, and object identification is an important task in the field of video processing [6]. Over the time period many number of the methodology has been proposed for doing the same. Some authors proposed to utilize optical flow approach for detecting the crowd behavior [6, 42]. Another authors [43] proposed a method in which they modified the optical flow approach with SVM. Here SVM stands for support vector machine. Apart from the above discussed methods, some more methodology like Spatial-temporal texture [30], Isometric Mapping [44], and spatio- temporal [45] has been proposed for crowd behavior analysis and monitoring.

6.3 Counting

Coming together of people in large number with some different motives like religious festivals, political rallies, or watching the sports events is known as crowd. Crowd counting is a process in which we estimate the strength of the crowd in number. This counting process can be categorized of two types; one is supervised approach whereas another is unsupervised approach. In the first approach the given dataset content is labeled and then we apply machine learning algorithms for the count prediction.In the second approach data is unknown and labels are not given. The first approach i.e. supervised approach can also be categorized as following:-

- Methods based on Supervised learning:
  - Counting using object detection approach: A frame of proper dimension roll over the complete scene of images or videos for detecting the people. After identification of interested object in frames many researchers proposed various methodology for counting the crowd. The authors [46] advise a histogram based approach using oriented gradients, in short HOG, some another authors [47-49] used shape of object, Haar features, edges boundary of the objects. Authors in [50, 51] utilized many machine learning approaches but many of them got failed on the scale of huge crowded scenario. Zhao et. al. used a 3 dimensional shape methodology for getting the better result over state of arts methods [51-52]. Some more paper proposed crowd counting methods based on object detection is listed in references [53-55]. Almost all the discussed methods in this section get failed when the crowd density is high and for a non-organized crowd the performance of methodology, counting using object detections, fall down.
  - Regression based approach: This approach based methods handle the challenges like highly dense or cluttered crowd very nicely. The working of methods based on the regression classify in 2 steps: Attributes extraction and then define a regression model. The attributes extraction includes the process of eliminating the background details and considering only the foreground content. Binary large objects can also be used as an attribute for getting the better performance [34, 56]. Nearby attributes incorporate removing edge and surface information from the given sample dataset.Examples of nearby attributes are GLCMs, LBP, and HoG. After getting the local attributes, the next is to perform the mapping into extracted attributes with the help of regression approaches. This regression can be linear regression, Gaussian regression and ridge regression [57]. For performing the mapping step the authors [23] used the mixture of Fourier Transform and scale-invariant feature transform. Likewise the authors [58] used sparse images for extracting the attributes and cumulative attribute space for mapping. Few more proposed methodology used for crowd counting is listed in [11-12, 58, 59].

    The regression based approaches solve the occlusion and non-organized crowd challenges, but there is still some challenges left in finding the spatial information.

- Estimation: The authors [36] proposed a method to find the spatial information using linear mapping with local attributes. These methods utilize the local attributes for mapping with density of object. For developing the density map the author used convolved the quadratic optimization with the help of some optimization algorithm.

  Algorithms based on density level can be categories of following types:
  Low level: Low dense estimation contains methods like optical flow, tracking, and background segmentation [60, 61]. Motion elements are the key factor of these methods. Implementing the modeling strategy on frame to frame, gives these motion elements, which provides a important way in object detection.

  Middle Level: In estimation of density at mid-level, the class data becomes the child of classification methods.

  High level: In these estimation approaches of density, non-static texture modeling is used [62]. These techniques are prevailing crowd modeling strategies.

- Methods based on Deep Learning: The newly proposed deep learning methods have brought big differences in performance over the traditional machine learning approaches for the task of recognition [63-67]. The traditional machine learning approaches depends on manual calculation for features but in deep learning approaches we use the different technologies for the features calculations. These deep learning approaches are also used by the researchers for developing the crowd counting task. The authors [26] used AlexNet architecture of deep learning for high density crowd image in their experiments. Likewise the authors [68], with the help of density, categorize the images in 5 different levels. These 5 categories are very high, high, medium, low and very low density. Similarly the authors [69] a cross counting methodology. They added the layered boosting convolutional neural network in model for the calculation of residual error in proposed model. For reducing the effects of low resolution, selective sampling is used. The authors [29] proposed a multi column deep convolution neural network based approach for calculating the crowd strength in number. To cater to different head sizes, 3 segments with different channel sizes are utilized. The authors [70] used dilated deep convolutional neural networks for getting the better knowledge of congested scenario. The authors [51] use scale adaptive deep convolution neural network for calculating the number of person in the crowd.

- Methods based on Unsupervised learning:
- Clustering: These techniques depend on the suspicion that some visual attributes and movement are uniform. In clustering approaches objects with similar attributes are categorize in different categories. The authors [71] extracted low level attributes by using KLT trackers. Bayesian clustering is used to calculate the total number of person in crowd, after getting the features [72]. Such sorts of calculation constitute appearance based attributes. These approaches give a false estimation when persons are in non- dynamic position.Few more methods are listed in [72-74].

  Finding the strength of crowd and detecting the abnormality in crowd behavior is the most talked topic in research field. In the state of art, numbers of researches have published for detection of abnormality in crowd response. The authors proposed 2 novelties approach for detecting the unusual behavior of crowd [75]. In the first approach the features matrix developed using spatial temporal concept. The second approaches detecting the abnormality in the crowd using the crowd motion directions. All are named as signatures. A modified 8-bits matrix developed for the signatures. The results of experiments done by the authors, showed the better performance over the other approaches. The authors in [76] considered both motion of the crowd and their appearance for detecting the abnormal behavior. The authors proposed a Swarm based novelty theoretical approach Histograms of Oriented Swarms (HOSs). In crowded conditions this HOS provides a signature. The appearance of crowd and attributes of motion are used only for the reduction of local noise, performance enhancement for non-predominant discovery of nearby abnormalities, and reducing the cost of implementation. In that capacity, the methodology gets an expanded exactness for pixel-based occasion acknowledgment in the group. The authors in [77] proposed a point trajectory based approach Histograms of Optical flow (PT-HOF) for detecting the abnormality in crowd behavior. For calculating the point trajectory of the crowded scene, the PT-HOF uses the spatial information and temporal information. For encoding the relevant attributes it use the deep learning approach. The authors in [11] define another model for abnormality detection using space time and they named their model as Markov Random Field (MRF). The MRF contains a graph in which the nodes of graph store the local small area of the video frames. Any two nodes of the MRF graph connected through an edge store the neighboring local region.

The crowd behavior monitoring and analysis is not only limited to calculate the crowd strength and abnormalities detection in crowd behavior, but it also includes the challenges like saliency detection, detection of congestions etc. In the field of computer vision, the saliency detection is a process to detect and segment salient objects from natural scenario. For finding the saliency in people gathering the authors in [78] used knowledge based scale with the help of visual system of humans by implementing the convolutional neural network with self-attention approach. Likewise the authors in [79] detected the salient motion in the crowd using selective directional information and a network with repulsive force. The frame sequence of crowded video examined through optimal flow approach. After this the vector for the crowd motion is calculated. The authors used 3 video frames scene from datasets railway station, marathon and Hajj. The authors in [80] explained about the use of temporal changes of crowd flowfor the identification of salient regions. These salient regions occur in many scenarios like occlusion, evacuation arrangements at passage and leave points etc. In a non-regular flow of crowd, the natures of motions vary from person to person. The authors in [81] used unsupervised methodology to detect the salient region of the crowd.

7. Result Analysis

7.1 Quantification of Task

- Counting:

Here the variables i and C_i used to denote the strength in number of crowded image. The assumption of this single evaluation matrix does not give any useful information for the person in crowd distribution or its location either in the image or in the video, but this is even useful in the many model development used for crowd counting of a scattered crowd over a big area. The authors in [82] discussed one method that splits the whole crowded area in a different small region and find the min value of the number of people for each region and then calculates the min density for whole area. Although using this methodit is absolutely complex task to find the number of the person in crowd for many images of different locations, so integration of a particular region value based on density is also permitted. Because of complexity, the two vector matrices mean square error (MSE) and mean absolute error (MAE) are used to evaluate the performance of the methods used for crowd counting.

The MSE and MAE are defined as follows:

$$\text{MSE}= \sqrt{\frac{1}{N}\sum_{i=1}^{N}|X_i - X'_i|^2} \text{------------------------------(1)}$$

$$\text{MAE}= \frac{1}{N}\sum_{i=1}^{N}|X_i - X'_i| \text{------------------------------------(2)}$$

where,

N=Number of test sample

X_i = Count on ground level

X'_i= Estimated count of ith sample.

- Localization: In numerous applications, the exact location of individuals is needed, for instance, instating a tracking technique in high thickness swarmed scene.In any case, to compute the limitation mistake, anticipated area is related with ground truth area by performing 1-1 coordinating. And this achieved using greedy association technique, followed by precision calculation, F-measurement and Recall.However the precision Recall curvature can utilize for the computing of complete performance. This is called as L-AUC.

We contend here, exact crowd confinement is relatively less investigated zone. For the localization problem there is no any proper evaluation matric given by the researchers. There is single attempt which discussed about 1-1 matching, listed in [23]. But we can notice that the evaluation matricdiscussed in [23] caused an optimistic problem in some situations. No penalty has been characterized in over location cases. For example, if there is one real object matching with many other objects, then the closest case will be considered with ignoring the other remaining objects without any considerations. We accept that for a reasonable correlation, the talked about measurement neglects to be recognized broadly. The evaluation matrices define as follows:

Precision = True_Positive/(True_Positive+False_Positive) ------- (3)

Recall = True_Positive/(True_Positive +False_Negetive) ---------(4)

F- measures = 2* (Precision * Recall) / (Precision + Recall) -------(5)

- Density estimation: When we calculate the density/pixel for a specific region of an image then this process is known as density estimation. Density estimation is not quite the same as considering a picture may include tallies inside specific safe cutoff points, though containing a few areas which will have a nearly higher thickness.This can be seen in some empty area of a scene like roads, sky, and walls etc., captured through aerial cameras. The counting estimation matrices can be used for density estimation also, though the matrices MAE, MSE were calculated based on pixel per unit.

7.2 Data Annotation

Tools: Adding some extra information i.e. metadata with dataset is called annotation. This makes the task easy for machine learning approaches. The content of a dataset may be images, videos, text etc. A computer use the observe data to discover a pattern class in a given new unseen data. An online technique was developed for annotation with the help of Java, Python and HTML. This technique used the observed data for the labeling starting points. This technique provides two types of labeling, one is point labeling and another is bound box labeling.

Basically, making ground truth marks were generally delivered through a manual cycle.This naming was performed without programmed marking device.A particular sort of naming was absolutely reliant on the emotional impression of a solitary person who was associated with this naming errand.Hence providing an accurate ground truth label in the image was very difficult and a time-consuming task.

7.3 Comparative Analysis

- Over the time period number of significant work has done in the field of crowd behavior monitoring and analysis. Number of new dataset has been utilized in the research. Although many of these available dataset, are good in crowd counting problem. The localization problem and behavior analysis problem has yet to be explored broadly and the only datasets UCF-QNRF, NWPU can be used for this. Even more, we have a large dataset for CBMAS, but we still require some more dataset in public domain.
- Most labeling task for observed data was done manually. Some tools are used to label the observed data based on user perception and in that case perception of one user may be different with another user. So there is always a chance for error.
- Comparatively less work has done for localization problem with respect to two other challenges crowd counting and analyzing the crowd behavior. Still there is a lack of real performance matric for the behavior analysis.
- Some authors proposed deep learning based approaches for counting, localization and behavior analysis, whose performance was comparatively better than the traditional approaches. The traditional machine learning approaches was good for the organized crowd, however, when the complex in nature type dataset used for these approaches, the performance dropped significantly.

- We observed that there is a scope of improvement in complex architecture of deep convolutional neural network for multi-scale challenges. Also, the current strategies have more spotlights on the framework exactness, while the accuracy of density distribution is overlooked.
- Customary AI techniques have adequate execution in controlled lab conditions.In any case, when these techniques were applied to datasets with unconstrained and un-controlled conditions, huge drop in execution is taken note.Notwithstanding, profound learning based techniques show much better execution in the wild conditions.
- Analyzing the crowd is the hot topic in current research field. A lot of development has been done in this filed over a decade. The experiments result show the enhance performance of evaluation matrices.

8. Summary and Concluding remakes:

In current scenario where the safety of the people matters most specially in the crowd gathering, the monitoring and analyzing the behavior of crowd is play a vital role in this aspect. These tasks giveinformation like crowd strength in number, individual location, predict behavior etc. Although these task become very difficult when we use the non-organized crowd dataset for the model development. One major issue in this field is unviability of publicly accessible dataset, particularly in the case of localization problem and crowd behavior problem.

References:

1. Khan, A., Ali Shah, J., Kadir, K., Albattah, W. and Khan, F., 2020. Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review. Applied Sciences, 10(14), p.4781.
2. Al-Salhie, L., Al-Zuhair, M. and Al-Wabil, A., 2014, June. Multimedia surveillance in event detection: crowd analytics in Hajj. In International Conference of Design, User Experience, and Usability (pp. 383-392). Springer, Cham.
3. Alabdulkarim, L., Alrajhi, W. and Aloboud, E., 2016, July. Urban Analytics in Crowd Management in the Context of Hajj. In International Conference on Social Computing and Social Media (pp. 249-257). Springer, Cham.
4. Mohamed, S.A.E. and Parvez, M.T., 2019, February. Crowd Modeling Based Auto Activated Barriers for Management of Pilgrims in Mataf. In 2019 International Conference on Innovative Trends in Computer Engineering (ITCE) (pp. 260-265). IEEE.
5. Rahim, M.S.M., Fata, A.Z.A., Basori, A.H., Rosman, A.S., Nizar, T.J. and Yusof, F.W.M., 2011, November. Development of 3D Tawaf simulation for hajj training application using virtual environment. In International Visual Informatics Conference (pp. 67-76). Springer, Berlin, Heidelberg.
6. Rohit, K., Mistree, K. and Lavji, J., 2017, March. A review on abnormal crowd behavior detection. In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS) (pp. 1-3). IEEE.
7. Motlagh, N.H., Bagaa, M. and Taleb, T., 2017. UAV-based IoT platform: A crowd surveillance use case. IEEE Communications Magazine, 55(2), pp.128-134.
8. Al-Sheary, A. and Almagbile, A., 2017. Crowd monitoring system using unmanned aerial vehicle (UAV). Journal of Civil Engineering and Architecture, 11(11), pp.1014-1024.
9. Alotibi, M.H., Jarraya, S.K., Ali, M.S. and Moria, K., 2019. CNN-Based Crowd Counting Through IoT: Application For Saudi Public Places. Procedia Computer Science, 163, pp.134-144.
10. Khan, K., Albattah, W., Khan, R.U., Qamar, A.M. and Nayab, D., 2020. Advances and trends in real time visual crowd analysis. Sensors, 20(18), p.5073.
11. Fiaschi, L., Köthe, U., Nair, R. and Hamprecht, F.A., 2012, November. Learning to count with regression forest and structured labels. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012) (pp. 2685-2688). IEEE.
12. Giuffrida, M.V., Minervini, M. and Tsaftaris, S.A., 2016. Learning to count leaves in rosette plants.
13. Boulos, M.N.K., Resch, B., Crowley, D.N., Breslin, J.G., Sohn, G., Burtner, R., Pike, W.A., Jezierski, E. and Chuang, K.Y.S., 2011. Crowdsourcing, citizen sensing and sensor web technologies for public

and environmental health surveillance and crisis management: trends, OGC standards and application examples. International journal of health geographics, 10(1), pp.1-29.

14. Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F.Y., 2014. Traffic flow prediction with big data: a deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 16(2), pp.865-873.
15. Sadeghian, A., Alahi, A. and Savarese, S., 2017. Tracking the untrackable: Learning to track multiple cues with long-term dependencies. In Proceedings of the IEEE International Conference on Computer Vision (pp. 300-311).
16. Perez, H., Hernandez, B., Rudomin, I. and Ayguade, E., 2016. Task-based crowd simulation for heterogeneous architectures. In Innovative Research and Applications in Next-Generation High Performance Computing (pp. 194-219). IGI Global.
17. Martani, C., Stent, S., Acikgoz, S., Soga, K., Bain, D. and Jin, Y., 2017. Pedestrian monitoring techniques for crowd-flow prediction. Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction, 170(2), pp.17-27.
18. Khouj, M., López, C., Sarkaria, S. and Marti, J., 2011, May. Disaster management in real time simulation using machine learning. In 2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE) (pp. 001507-001510). IEEE.
19. Barr, J.R., Bowyer, K.W. and Flynn, P.J., 2014, March. The effectiveness of face detection algorithms in unconstrained crowd scenes. In IEEE Winter Conference on Applications of Computer Vision (pp. 1020-1027). IEEE.
20. Ng, H.W., Nguyen, V.D., Vonikakis, V. and Winkler, S., 2015, November. Deep learning for emotion recognition on small datasets using transfer learning. In Proceedings of the 2015 ACM on international conference on multimodal interaction (pp. 443-449).
21. Zhou, B., Tang, X. and Wang, X., 2015. Learning collective crowd behaviors with dynamic pedestrian-agents. International Journal of Computer Vision, 111(1), pp.50-68.
22. Danilkina, A., Allard, G., Baccelli, E., Bartl, G., Gendry, F., Hahm, O., Hege, G., Kriegel, U., Palkow, M., Petersen, H. and Schmidt, T., 2015, February. Multi-Camera Crowd Monitoring: The SAFEST Approach.
23. Idrees, H., Saleemi, I., Seibert, C. and Shah, M., 2013. Multi-source multi-scale counting in extremely dense crowd images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2547-2554).
24. Li, Y., Sarvi, M., Khoshelham, K. and Haghani, M., 2020. Multi-view crowd congestion monitoring system based on an ensemble of convolutional neural network classifiers. Journal of Intelligent Transportation Systems, pp.1-12.
25. Pandey, A., Pandey, M., Singh, N. and Trivedi, A., 2020. KUMBH MELA: a case study for dense crowd counting and modeling. Multimedia Tools and Applications, pp.1-22.
26. Wang, Q., Gao, J., Lin, W. and Li, X., 2020. Nwpu-crowd: A large-scale benchmark for crowd counting. arXiv preprint arXiv:2001.03360.
27. Ding, X., Lin, Z., He, F., Wang, Y. and Huang, Y., 2018, April. A deeply-recursive convolutional network for crowd counting. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1942-1946). IEEE.
28. Idrees, H., Tayyab, M., Athrey, K., Zhang, D., Al-Maadeed, S., Rajpoot, N. and Shah, M., 2018. Composition loss for counting, density map estimation and localization in dense crowds. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 532-546).
29. Zhang, Y., Zhou, D., Chen, S., Gao, S. and Ma, Y., 2016. Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 589-597).
30. Zhang, C., Li, H., Wang, X. and Yang, X., 2015. Cross-scene crowd counting via deep convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 833-841).

31. Shao, J., Kang, K., Change Loy, C. and Wang, X., 2015. Deeply learned attributes for crowded scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4657-4666).
32. Chen, K., Gong, S., Xiang, T. and Change Loy, C., 2013. Cumulative attribute space for age and crowd density estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2467-2474).
33. Ferryman, J. and Shahrokni, A., 2009, December. Pets2009: Dataset and challenge. In 2009 Twelfth IEEE international workshop on performance evaluation of tracking and surveillance (pp. 1-6). IEEE.
34. Chan, A.B., Liang, Z.S.J. and Vasconcelos, N., 2008, June. Privacy preserving crowd monitoring: Counting people without people models or tracking. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-7). IEEE.
35. Rodriguez, M., Laptev, I., Sivic, J. and Audibert, J.Y., 2011, November. Density-aware person detection and tracking in crowds. In 2011 International Conference on Computer Vision (pp. 2423-2430). IEEE.
36. Lempitsky, V. and Zisserman, A., 2010. Learning to count objects in images. Advances in neural information processing systems, 23, pp.1324-1332.
37. Ma, Z., Yu, L. and Chan, A.B., 2015. Small instance detection by integer programming on object density maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3689-3697).
38. Sánchez, F.L., Hupont, I., Tabik, S. and Herrera, F., 2020. Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects. Information Fusion.
39. Zhang, X., Ma, D., Yu, H., Huang, Y., Howell, P. and Stevens, B., 2020. Scene perception guided crowd anomaly detection. Neurocomputing, 414, pp.291-302.
40. Sikdar, A. and Chowdhury, A.S., 2020. An adaptive training-less framework for anomaly detection in crowd scenes. Neurocomputing, 415, pp.317-331.
41. Tripathi, G., Singh, K. and Vishwakarma, D.K., 2019. Convolutional neural networks for crowd behaviour analysis: a survey. The Visual Computer, 35(5), pp.753-776.
42. Lahiri, S., Jyoti, N., Pyati, S. and Dewan, J., 2018, August. Abnormal Crowd Behavior Detection Using Image Processing. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.
43. Yimin, D.O.U., Fudong, C., Jinping, L.I. and Wei, C., 2019, June. Abnormal Behavior Detection Based on Optical Flow Trajectory of Human Joint Points. In 2019 Chinese Control And Decision Conference (CCDC) (pp. 653-658). IEEE.
44. Fradi, H., Luvison, B. and Pham, Q.C., 2016. Crowd behavior analysis using local mid-level visual descriptors. IEEE Transactions on Circuits and Systems for Video Technology, 27(3), pp.589-602.
45. Rao, A.S., Gubbi, J. and Palaniswami, M., 2016. Anomalous Crowd Event Analysis Using Isometric Mapping. In Advances in Signal Processing and Intelligent Recognition Systems (pp. 407-418). Springer, Cham.
46. Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.
47. Sabzmeydani, P. and Mori, G., 2007, June. Detecting pedestrians by learning shapelet features. In 2007 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE.
48. Viola, P. and Jones, M., 2001, July. Robust real-time face detection. In null (p. 747). IEEE.
49. Wu, B. and Nevatia, R., 2005, October. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 (Vol. 1, pp. 90-97). IEEE.
50. Viola, P., Jones, M.J. and Snow, D., 2005. Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision, 63(2), pp.153-161.

51. Zhao, T., Nevatia, R. and Wu, B., 2008. Segmentation and tracking of multiple humans in crowded environments. IEEE transactions on pattern analysis and machine intelligence, 30(7), pp.1198-1211.
52. Ge, W. and Collins, R.T., 2009, June. Marked point processes for crowd counting. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2913-2920). IEEE.
53. Wang, H., Roa, A.C., Basavanhally, A.N., Gilmore, H.L., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F. and Madabhushi, A., 2014. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging, 1(3), p.034003.
54. Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F. and Madabhushi, A., 2014, March. Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection. In Medical Imaging 2014: Digital Pathology (Vol. 9041, p. 90410B). International Society for Optics and Photonics.
55. Dollar, P., Wojek, C., Schiele, B. and Perona, P., 2011. Pedestrian detection: An evaluation of the state of the art. IEEE transactions on pattern analysis and machine intelligence, 34(4), pp.743-761.
56. Ryan, D., Denman, S., Fookes, C. and Sridharan, S., 2009, December. Crowd counting using multiple local features. In 2009 Digital Image Computing: Techniques and Applications (pp. 81-88). IEEE.
57. Paragios, N. and Ramesh, V., 2001, December. A MRF-based approach for real-time subway monitoring. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 (Vol. 1, pp. I-I). IEEE.
58. Chen, K., Loy, C.C., Gong, S. and Xiang, T., 2012, September. Feature mining for localised crowd counting. In BMVC (Vol. 1, No. 2, p. 3).
59. Cheng, Z., Qin, L., Huang, Q., Yan, S. and Tian, Q., 2014. Recognizing human group action by layered model with multiple cues. Neurocomputing, 136, pp.124-135.
60. McIvor, A.M., 2000. Background subtraction techniques. Proc. of Image and Vision Computing, 4, pp.3099-3104.
61. Black, M.J. and Fleet, D.J., 2000. Probabilistic detection and tracking of motion boundaries. International Journal of Computer Vision, 38(3), pp.231-245.
62. Chan, A.B. and Vasconcelos, N., 2008. Modeling, clustering, and segmenting video with mixtures of dynamic textures. IEEE transactions on pattern analysis and machine intelligence, 30(5), pp.909-926.
63. Khan, K., Ahmad, N., Khan, F. and Syed, I., 2020. A framework for head pose estimation and face segmentation through conditional random fields. Signal, Image and Video Processing, 14(1), pp.159-166.
64. Khan, K., Attique, M., Khan, R.U., Syed, I. and Chung, T.S., 2020. A multi-task framework for facial attributes classification through end-to-end face parsing and deep convolutional neural networks. Sensors, 20(2), p.328.
65. Khan, K., Attique, M., Syed, I. and Gul, A., 2019. Automatic gender classification through face segmentation. Symmetry, 11(6), p.770.
66. Ullah, F., Zhang, B., Khan, R.U., Chung, T.S., Attique, M., Khan, K., El Khediri, S. and Jan, S., 2020. Deep Edu: A Deep Neural Collaborative Filtering for Educational Services Recommendation. IEEE Access, 8, pp.110915-110928.
67. Ahmad, K., Khan, K. and Al-Fuqaha, A., 2020. Intelligent Fusion of Deep Features for Improved Waste Classification. IEEE Access.
68. Fu, M., Xu, P., Li, X., Liu, Q., Ye, M. and Zhu, C., 2015. Fast crowd density estimation with convolutional neural networks. Engineering Applications of Artificial Intelligence, 43, pp.81-88.
69. Walach, E. and Wolf, L., 2016, October. Learning to count with cnn boosting. In European conference on computer vision (pp. 660-676). Springer, Cham.
70. Li, Y., Zhang, X. and Chen, D., 2018. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1091-1100).

71. Rabaud, V. and Belongie, S., 2006, June. Counting crowded moving objects. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 1, pp. 705-711). IEEE.
72. Brostow, G.J. and Cipolla, R., 2006, June. Unsupervised bayesian detection of independent motion in crowds. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 1, pp. 594-601). IEEE.
73. Duygulu, P., Barnard, K., de Freitas, J.F. and Forsyth, D.A., 2002, May. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In European conference on computer vision (pp. 97-112). Springer, Berlin, Heidelberg.
74. Moosmann, F., Triggs, B. and Jurie, F., 2007. Fast discriminative visual codebooks using randomized clustering forests. In Advances in neural information processing systems (pp. 985-992).
75. Hao, Y., Xu, Z.J., Liu, Y., Wang, J. and Fan, J.L., 2019. Effective crowd anomaly detection through spatio-temporal texture analysis. International Journal of Automation and Computing, 16(1), pp.27-39.
76. Kaltsa, V., Briassouli, A., Kompatsiaris, I., Hadjileontiadis, L.J. and Strintzis, M.G., 2015. Swarm intelligence for detecting interesting events in crowded environments. IEEE transactions on image processing, 24(7), pp.2153-2166.
77. Zhao, K., Liu, B., Li, W., Yu, N. and Liu, Z., 2018, July. Anomaly Detection and Localization: A Novel Two-Phase Framework Based on Trajectory-Level Characteristics. In 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-6). IEEE.
78. Nguyen, M.T., Siritanawan, P. and Kotani, K., 2020. Saliency detection in human crowd images of different density levels using attention mechanism. Signal Processing: Image Communication, 88, p.115976.
79. Zhang, X., Lin, D., Zheng, J., Tang, X., Fang, Y. and Yu, H., 2019. Detection of salient crowd motion based on repulsive force network and direction entropy. Entropy, 21(6), p.608.
80. Lim, M.K., Chan, C.S., Monekosso, D. and Remagnino, P., 2014. Detection of salient regions in crowded scenes. Electronics letters, 50(5), pp.363-365.
81. Lim, M.K., Kok, V.J., Loy, C.C. and Chan, C.S., 2014, August. Crowd saliency detection via global similarity structure. In 2014 22nd International Conference on Pattern Recognition (pp. 3957-3962). IEEE.
82. Jacobs, H., 1967. To count a crowd. Columbia Journalism Review, 6(1), p.37.