

***CrowdVAS-Net: A Deep-CNN Based Framework to Detect
Abnormal Crowd-Motion Behavior in Videos for
Predicting Crowd Disaster***

A DISSERTATION

*Submitted in partial fulfillment of the
requirements for the award of degree*

of

MASTER OF TECHNOLOGY

in

DISASTER MITIGATION AND MANAGEMENT

By

TANU GUPTA

(Enrollment No.: 17552010)



Centre of Excellence in Disaster Mitigation and Management

Indian Institute of Technology Roorkee

Roorkee - 247667 (INDIA)

JUNE, 2019

CANDIDATE'S DECLARATION

I declare that the work presented in this dissertation with title “*CrowdVAS-Net: A Deep-CNN Based Framework to Detect Abnormal Crowd-Motion Behavior in Videos for Predicting Crowd Disaster*” towards fulfillment of the requirement for the award of the degree of **Master of Technology in Disaster Mitigation and Management** submitted in the **Centre of Excellence in Disaster Mitigation and Management, Indian Institute of Technology Roorkee, India** is an authentic record of my own work carried out during the period of **May, 2018 to June, 2019** under the supervision of **Dr. Sudip Roy** and **Dr. Josodhir Das**, Associated Faculty, Centre of Excellence in Disaster Mitigation and Management, Indian Institute of Technology Roorkee, India.

The content of this dissertation has not been submitted by me for the award of any other degree of this or any other institute.

Date:

Place: Roorkee

(**Tanu Gupta**)

CERTIFICATE

This is to certify that Thesis Report entitled “***CrowdVAS-Net: A Deep-CNN Based Framework to Detect Abnormal Crowd-Motion Behavior in Videos for Predicting Crowd Disaster***” which is submitted by **Tanu Gupta (Enrollment No.: 17552010)**, towards the fulfillment of the requirements for the award of the degree of **Master of Technology in Disaster Mitigation and Management** submitted in the **Centre of Excellence in Disaster Mitigation and Management, Indian Institute of Technology Roorkee, India** is carried out by her under our esteemed supervision and the statement made by the candidate in declaration is correct to the best of our knowledge and belief.

Date:

Place: Roorkee

(Dr. Sudip Roy)

Associated Faculty

Centre of Excellence in Disaster Mitigation and Management

Indian Institute of Technology Roorkee

(Dr. Josodhir Das)

Associated Faculty

Centre of Excellence in Disaster Mitigation and Management

Indian Institute of Technology Roorkee

ABSTRACT

With the increased occurrences of crowd disasters like human stampedes, crowd management and their safety during mass gathering events like concerts, congregation or political rally, etc., are vital tasks for the security personnel. In this research work, we propose a framework named as *CrowdVAS-Net* for crowd-motion analysis that considers velocity, acceleration and saliency features in the video frames of a moving crowd. *CrowdVAS-Net* relies on a deep convolutional neural network (DCNN) for extracting motion and appearance feature representations from the video frames that helps us in classifying the crowd-motion behaviour as abnormal or normal from a short video clip. These feature representations are then trained with a random forest classifier. Furthermore, a dataset having 704 video clips having dense crowded scenes has been created for performance evaluation of the proposed method. Simulation results confirm that the proposed *CrowdVAS-Net* achieves the classification accuracy of 77.8% outperforming the state-of-the-art machine learning models. Moreover, this framework also reduced the video processing and analyzing time up to 96.8% compared to the traditional method of video processing. Based on our results, we believe that our work will help security personnel and crowd managers in ensuring the public safety during mass gatherings with better accuracy.

ACKNOWLEDGMENTS

I would first like to thank my thesis advisors **Dr. Sudip Roy** and **Dr. Josodhir Das** for guiding me throughout my thesis work, helping me whenever needed and being a constant source of motivation. I am also grateful to the **Centre of Excellence in Disaster Mitigation and Management** and **Computing and Design Automation (CoDA) Lab, Department of Computer Science and Engineering, IIT Roorkee** for providing the valuable resources to aid my research.

I would like to thank my friend **Mr. Naveen Gupta** and all the **lab colleagues** who were always willing to help and give their best suggestions during my thesis work.

Last but not the least, I would like to thank **my family** for their blessings and support without which I would not have reached this stage of my life.

Tanu Gupta

CONTENTS

1	INTRODUCTION	1
1.1	Basics of Disaster Management	1
1.2	Machine Learning and Computer Vision for DM	2
1.3	Importance of Crowd Motion Analysis	3
1.4	Contribution of Dissertation	6
1.5	Organization of Dissertation	6
2	LITERATURE SURVEY	7
2.1	Detection of Violent Behavior in Crowded Event	7
2.2	Detection of Abnormal Crowd Behavior	7
2.3	Deep Learning Framework for Crowd Density Estimation	8
2.4	XCS-LBP based Framework for Crowdd Density Estimation	8
2.5	Super-pixel based Crowd Flow Segmentation	8
2.6	Fully Convolutional Neural Networks (FCNN) for Crowd Segmentation	9
2.7	Segmentation in Dense Crowds	10
2.8	A Trajectory Clustering Approach for Crowd Flow Segmentation	10
2.9	Abnormal Crowd Tracking and Motion Analysis	11
2.10	SIFT Flow based Detection of Abnormal Crowd-Motion Behaviour	11
2.11	CNN for Crowd Motion Analysis: An Application in Abnormal Event De- tection	11
2.12	Summary of the Literature Survey	12
3	MOTIVATION AND PROBLEM STATEMENT	14
3.1	Motivation	14
3.2	Problem Statement	15
4	<i>CrowdVAS-Net</i>: A DEEP-CNN BASED FRAMEWORK TO DETECT AB- NORMAL CROWD-MOTION BEHAVIOR	16
4.1	Dataset Creation and Pre-processing	17
4.2	Proposed Framework: <i>CrowdVAS-Net</i>	18
4.2.1	System Architecture	18
4.2.2	Optical Flow	19
4.2.3	Visual Saliency Flow Map	20
4.2.4	<i>DCNN</i> for Representation Extraction	21
4.2.5	Classification: Random Forest Classifier	22

5	SIMULATION RESULTS	23
5.1	Simulation Setup	23
5.2	Simulation Scenarios	23
5.2.1	Scenario 1: Comparison of <i>CrowdVAS-Net</i> with other state-of-the-art methods	23
5.2.2	Scenario 2: Check performance of <i>CrowdVAS-Net</i> on UMN dataset	24
5.2.3	Scenario 3: Comparison of video processing time of <i>CrowdVAS-Net</i> and classical approach	25
6	CONCLUSIONS AND FUTURE SCOPE	27
6.1	Conclusions	27
6.2	Future Scope	27
	DISSEMINATION FROM DISSERTATION	28
	REFERENCES	29




List of Figures

1.1	Disaster management cycle [50].	2
1.2	Crowd type based on the direction of crowd movement.	3
1.3	Some cases of crowd disasters during mass gathering events.	5
2.1	Overview of solution approach [55].	9
2.2	Overview of solution approach [56].	10
3.1	Overview of the problem.	15
4.1	Overview of the proposed framework.	16
4.2	Sample frames of proposed dataset.	17
4.3	Block diagram of the proposed <i>CrowdVAS-Net</i>	18
4.4	Salient features extracted from the crowd video. (a) Input frames of a video, (b) single saliency map, and (c) co-saliency map.	21
5.1	Comparative analysis of feature extraction time, training time and testing time of proposed and classical approach.	26

List of Tables

1.1	Most recent crowd crushes and their details. [51].	4
2.1	Summary of the literature survey.	12
4.1	Descriptions of the proposed dataset.	18
5.1	Statistical comparison of <i>CrowdVAS-Net</i> with other state-of-the art techniques on proposed dataset.	24
5.2	A description of UMN dataset [16].	24
5.3	Statistical comparison of <i>CrowdVAS-Net</i> with previous methods on UMN dataset.	25
5.4	Statistical comparison of the performance of proposed approach with other state-of-art feature extraction techniques with different classifiers on UMN SocialForce [16] dataset.	25

List of Abbreviations



Abbreviation	Description
CCTV	Closed Circuit Television
CNN	Convolutional Neural Network
CV	Computer Vision
DCNN	Deep Convolutional Neural Network
DM	Disaster Management
DRO	Disaster Relief Operation
FCNN	Fully Convolutional Neural Network
HMM	Hidden Markov Model
KNN	k-Nearest Neighbors
RAM	Random Access Memory
RF	Random Forest
SVM	Support Vector Machine
TCP	Temporal CNN Pattern
XCS-LBP	eXtended Center-Symmetric Local Binary Pattern

Chapter 1

INTRODUCTION

Disaster is a dreadful event and can cause heavy destruction on nature and community. It can occur anytime, anywhere with varying intensity [1]. In 2015, total 344 disasters occurred worldwide, out of which 160 disasters had occurred on the Asia-Pacific region [2]. According to [2] the destruction occurred in 2015 is more than twice the destruction occurred in 2014 due to disaster. There are basically two types of disasters which are as follows:

1. Natural Disaster: physical process that may cause loss of life or property due to some geological event.
e.g., earthquakes, landslides, tsunamis, floods, cyclones, etc.
2. Man-made Disaster: events that occur due to human settlements which cause risk to human life.
e.g., industrial accidents, acts of terrorism, stampede, etc.

Altay *et al.* [3] had stated that data analysis and coordination are crucial tasks for decision making during disaster relief operation (DRO). Akhtar *et al.* [4] and Li *et al.* [5] also revealed the facts that there is an absence of coordination and collaboration in DRO teams which adversely impacts on the rescue operation. In order to reduce the potential loss of life, property and provide appropriate assistance to victims, disaster management (DM) is necessary for proper planning and preparedness before the disaster occurs. Effective disaster response requires well-organized teamwork and coordination by all various teams working in parallel to deal with the situation.

1.1 Basics of Disaster Management

Disaster management is a process of planning and managing resources for dealing with all humanitarian needs of emergencies. DM is helpful to mitigate the causes of disaster. DM is an active process that goes through all preparedness, response and recovery phases of the disaster management cycle. The DM cycle shows the ongoing process by which government organizations and communities plan for and reduce the effect of disaster, take actions during and after the disaster has occurred for immediate response and recovery.

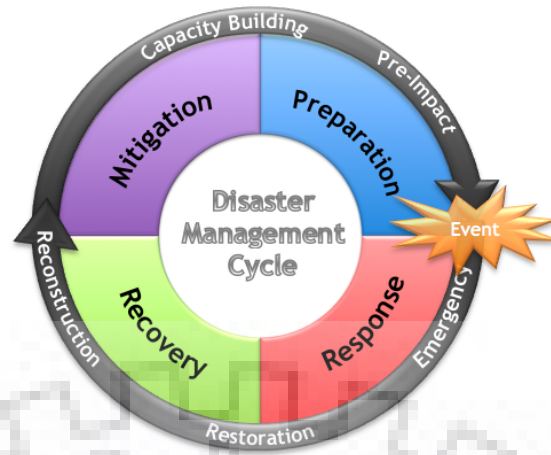


Figure 1.1: Disaster management cycle [50].

Phases of disaster management cycle has been presented in Figure 1.1 and explained in the following points.

1. Mitigation: actions taken pre-disaster in order to lessen the result of disaster
2. Preparedness: includes proper planning and training to tackle event when it occurs
3. Response: actions to reduce the damage done after disaster
4. Recovery: recovery of community to get back to previous condition

1.2 Machine Learning and Computer Vision for DM

Disaster management not only relates to the prediction of the course and consequences of disasters, but also to mitigate their unwanted consequences. This is undoubtedly a very challenging task and even more due to time constraint.

During natural or human-induced emergencies a series of carefully chosen decisions are important to mitigate the causalities caused by natural catastrophe. In such situations, computer vision (CV) algorithms can be helpful to solve various problems of disaster mitigation and management (like quick access of real-time disaster information, automatic video analysis, Analysis the crowd behavior and crowd motion and others).

Now-a-days, crowded scenes have been quite common due to increase in the human population and its diverse activities. Whether the crowd comes together to protest or just to entertain themselves at a party, suddenly it can turns to catastrophic. Occasionally it can catalyze fatal accidents. Hence, it is essential to analyze the crowd to nullify such deadly disasters.

The application of CV for crowd behavior analysis is appealing because it leads to a range of benefits. For example:

1. Save time and improve effectiveness

2. Specific events or scenes can be analyzed and tested more precisely
3. Minimizing human interventions

The aforementioned benefits motivate the application of CV algorithms to analyse the crowd motion behavior system that lead to crowd disaster.

1.3 Importance of Crowd Motion Analysis

The analysis of crowd disasters is an area of agile research that covered detection and tracking of crowd movement, analysis of crowd behavior, and density level of crowd. A lot of problems occurred due to the poor crowd management, such as crowd crushes, lack in co-ordination, which increase the requirement of automatic framework that can observe crowd-motion behavior by using a short video feeds from surveillance cameras. Table 1.1 shows the most recent crowd crushes with their details. It is clear from the table that crowd management and early detection of abnormal crowd-motion is an important task. It should be done in real time.

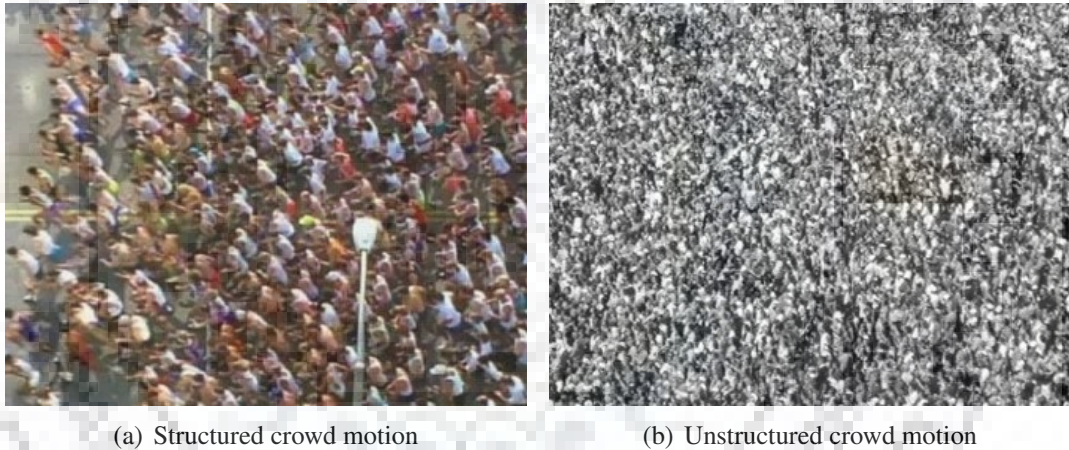


Figure 1.2: Crowd type based on the direction of crowd movement.

According to the crowd-motion movement pattern the crowd can be categorized into two classes [6]: structured and unstructured. When crowd moves in well defined common direction then that motion of the crowd is said to be structured motion, whereas if individual in the crowd is moving in any direction then the crowd type is unstructured [60]. Fig. 1.2(a) and Fig. 1.2(b), shows the structured and unstructured crowd-motion respectively. Fig. 1.3, shows some cases of crowd disasters during mass gathering events. A mass gathering can be of two types, planned and unplanned. Planned events are held at particular place for certain time period and that make every effort for adequate planning and response process. Conversely, unplanned events are those where the density of crowd is unpredictable and unhampered. Risk associated with crowd disasters in public gatherings can be reduced by considering a series of safety measures. Firstly, the infrastructure must be sufficient for public gatherings with sufficient capacity and no bottlenecks or compression points. Secondly, there should be emergency plans, like evacuation plan in case of any mis-happening.

Table 1.1: Most recent crowd crushes and their details. [51].

Year	Place	Casualties	Type of stampede	Crowd Type	Density level
2017	Mumbai, India	23	Stampede	Unstructured	High
2015	Mina, Saudi Arabia	769	Pilgrim Stampede	Structured	high
2015	Cairo, Egypt	28	Stadium Stampede	Unstructured	high
2015	France	130	Terrorist attack	Unstructured	Medium
2015	Rajahmundry Pushkar ghat, India	27	Festival stampede	Unstructured	High
2015	Shanghai, China	36	New year stampede	Unstructured	High
2014	Patna, India	32	Festival stampede	Unstructured	High
2013	Uttar Pradesh, India	36	Festival stampede	Unstructured	High
2013	Madhya Pradesh, India	115	Festival stampede	Structured	High
2013	Kumbh mela, Allahabad, India	42	Pilgrim stampede	Unstructured	High
2011	Kerala, India	102	Pilgrim stampede	Structured	High
2010	Kunda, India	71	Festival stampede	Structured	High
2010	Phnom Penh, Cambodia	347	Festival Stampede	Unstructured	High
2008	Jodhpur, India	224	Festival stampede	Structured	High
2008	Madhya Pradesh, India	8	Pilgrim stampede	Unstructured	High

Before any public gathering careful identification of all the factors that provoke the risks that lead to a miss-happening in a crowd is necessary.

In order to monitor and understand the complex crowd-motion behavior, an automated video surveillance systems are highly in demand. Manual crowd monitoring includes intensive human participation which is more prone to human error and also it is time consuming task. It is often impossible to detect and monitor individual, due to poor camera views or occlusions because of large numbers of people. This work particularly focuses on monitoring the crowd-motion behavior in dense crowded scene.

The previous studies on crowd-motion behavior are conducted on less dense to dense



(a) Hillsborough stadium stampede, 1989, England



(b) Love parade stampede, 2010, Germany



(c) Pushkar festival stampede, 2015, India



(d) Mumbai stampede, 2017, India



(e) Black Friday stampede, 2018, USA

Figure 1.3: Some cases of crowd disasters during mass gathering events.

crowded scenes. Also, these studies have been conducted on relatively small dataset. In this paper, we develop a *CrowdVAS-Net* framework that rely on motion and appearance features of crowd-motion to analyse the crowd-motion behavior. Furthermore, it also reduced the video processing time up to 96.8% and also increase the accuracy in terms of recognition of unusual events. In addition, in this work we also introduce a novel dataset comprising of 704 videos of crowd-motion. This dataset includes different scenarios such as pilgrimage, mall, stadium, mass gathering in constraint environment and in open places, etc. The proposed research aim is to develop a framework that can recognize abnormal motion of crowd by using only few contiguous frames of video. The main advantage of such framework are listed as follows:

- The framework can be deployed in such places which are mostly crowded such as mall, railway station, airport, public workplaces, etc. The deployment of framework in such places can help to automatic recognition of abnormal crowd-motion behavior such as chaos, panic, stampede, etc.
- The framework can be helpful to develop a crowd disaster early warning system that can help concerned authority to receive relevant and timely information before a disaster.
- The framework can be used to monitor mass gathering events such as contest, rally, etc.

1.4 Contribution of Dissertation

Our study focuses on the following aspects:

- Propose a dataset comprises of large number of high to very high dense crowd videos.
- Propose a new hybrid CNN with random forest classifier in order to recognize the abnormal crowd-motion behavior using deep features.
- Compare and deduce the best optimal classifier model that gives best classification result when trained on various deep representations.
- Compare the efficiency of proposed framework with other state-of-art algorithms.

1.5 Organization of Dissertation

This dissertation has been divided in the following chapters.

Chapter 1 provides the introduction of DM and explains how CV algorithms are helpful in crowd management.

Chapter 2 presents different methods/models for crowd-motion behavior analysis proposed till now.

Chapter 3 describes the motivation behind this dissertation work and the problem statement.

Chapter 4 presents proposed *CrowdVAS-Net* framework and the simulation scenarios along with the results are given in Chapter 5.

Chapter 6 concludes the dissertation and provides future scope in this work.

Chapter 2

LITERATURE SURVEY

This chapter includes various research works which has been previously done to avoid the crowd disaster. Different sections include the problem with their solution which was mentioned by author in their research work. At the end, summary is presented which compares the solution approaches for the studied research.

2.1 Detection of Violent Behavior in Crowded Event

Problem Statement: Hassner *et al.* [52] proposed a framework to monitor crowded scene to detect violent activity.

Input: Surveillance Videos.

Output: Classified scene as violent and non-violent.

Solution Approach: A framework named as Violent Flows (ViF) descriptor is proposed to detect violent and non-violent scenes. It considered change in flow vector with respect to time and then used support vector machine as classifier to classify the scene among two classes.

2.2 Detection of Abnormal Crowd Behavior

Problem Statement: Mousavi *et al.* [35] developed a improved video descriptor for recognizing abnormal situation in crowded scenes.

Input: Author had used three video datasets named as UCSD, Violence-in- Crowds and UMN in order to recognize the abnormal crowd behavior.

Output: Classify video frames as normal and abnormal.

Solution approach: A video descriptor named as Histogram of Oriented Tracklets (HOT)

is proposed. This descriptor used Latent Dirichlet Allocation and Support Vector Machines for classification of frames as normal and abnormal. This framework worked on sparse crowded scene and give false alarm in case of dense crowded scene.

2.3 Deep Learning Framework for Crowd Density Estimation

Problem Statement: Boominathan *et al.* [53] proposed a deep learning framework that estimates the crowd density of highly dense crowds from static images.

Input: Author used UCF CC 50 [54] dataset for experiment.

Output: Generates crowd density map and estimate the total count of people in the crowded scene.

Solution approach: A framework based on a combination of deep and shallow fully convolutional neural network is proposed. This framework estimate the density map for a given crowded scene. In case of overflow situation the warning message is not generated.

2.4 XCS-LBP based Framework for Crowdd Density Estimation

Problem Statement: Nagananthini *et al.* [55] proposed an algorithm that generates warning message when the total people count crosses a particular limit.

Input: Worked on three datasets named as PETS2009, UCSD and UFC_CC_50 for experiments.

Output: Warning message if people count exceeds certain limit.

Solution approach: Fig. 2.1 shows the schematic view of [55] approach in which texture based feature using XCS-LBP are extracted and then used to train the deep Convolutional Neural Network (CNN) to estimate the total count of people in crowded scene. In proposed method K is the count limit.

2.5 Super-pixel based Crowd Flow Segmentation

Problem Statement: Biswas *et al.* [56] proposed an algorithm for segmentation of high density crowd flows based on super-pixels.

Input: H.264 compressed videos.

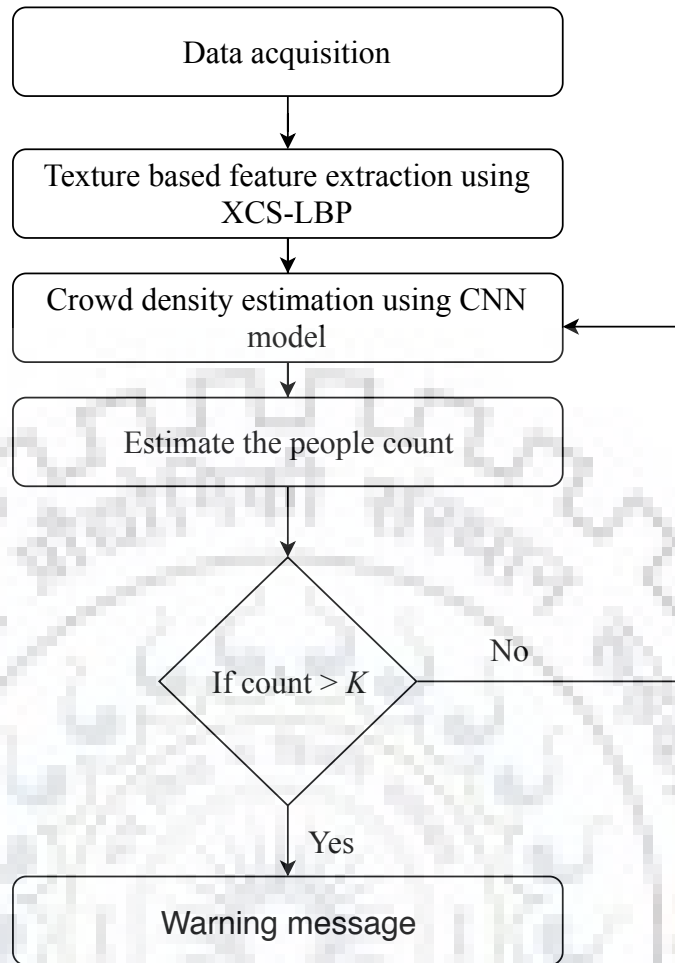


Figure 2.1: Overview of solution approach [55].

Output: Flow segmented image.

Solution approach: A framework for crowd flow segmentation is proposed. After pre-process the motion vector super-pixel based clustering has done. Crowd flow is segmented by detecting the significant edges. Fig. 2.2 represents the overview of the proposed approach of [56].

2.6 Fully Convolutional Neural Networks (FCNN) for Crowd Segmentation

Problem Statement: Kang *et al.* [57] proposed a fast fully convolutional neural network (FCNN) for crowd segmentation. The fully connected layers in CNN are replaced with 1×1 convolution kernels.

Input: Crowded scene images.

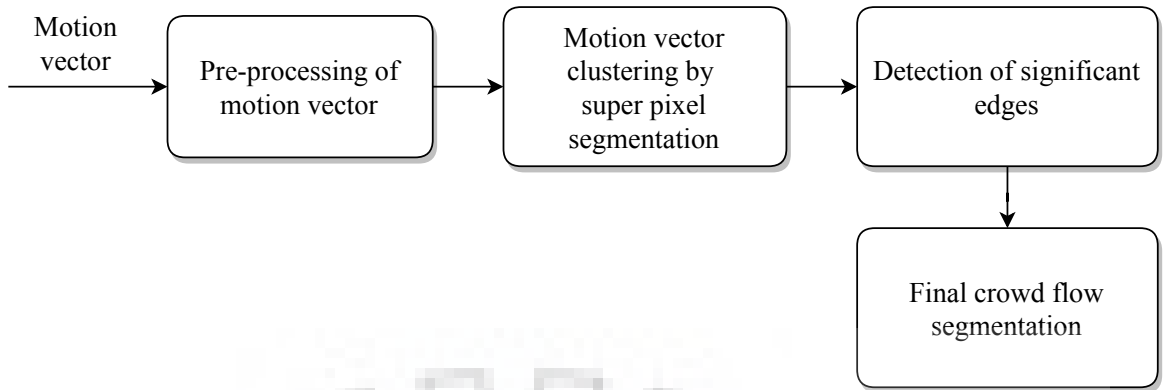


Figure 2.2: Overview of solution approach [56].

Output: FCNN feature map of input images.

Solution approach: A framework to segment the crowd is proposed. In this framework the fully connected layers in CNN are replaced with 1×1 convolution kernel. The framework takes whole image as input and directly outputs segmentation maps by one pass of forward propagation.

2.7 Segmentation in Dense Crowds

Problem Statement: Nazir *et al.* [58] proposed a framework to segment the flow of dense crowds.

Input: Video frames from UCF crowd dataset.

Output: Segmented image represents the crowd flow segmentation.

Solution approach: An approach based on flow fields is proposed. The flow fields between two consecutive frames are calculated. Average of the flow fields value is used. The watershed algorithm is used for segmenting the region.

2.8 A Trajectory Clustering Approach for Crowd Flow Segmentation

Problem Statement: Sharma *et al.* [59] proposed a framework to segment flow of dense crowds.

Input: Video frames from UCF crowd dataset.

Output: Pixel-wise segmentation of crowd flow.

Solution approach: A trajectory clustering-based approach for segmenting flow patterns in highly dense crowd videos is proposed. All trajectories of the scene are extracted and then divide in clusters by using trajectory algorithm. On the basis of cluster groups the region is segmented.

2.9 Abnormal Crowd Tracking and Motion Analysis

Problem Statement: Santhiya et al. [42] proposed a framework to detect the abnormal behavior of crowd motion.

Input: Video frames from UMN crowd dataset.

Output: Classify scene as normal or abnormal based on the crowd motion.

Solution approach: A framework consists of four modules named as: background/foreground modelling, blob analysis, crowd detection and abnormal crowd tracking is proposed. The framework relies on blob analysis to detect crowd and abnormal crowd-motion tracking.

2.10 SIFT Flow based Detection of Abnormal Crowd-Motion Behaviour

Problem Statement: Zhang *et al.* [41] focuses on the detection of the abnormal crowd-motion behaviour.

Input: Video frames from UMN crowd dataset.

Output: Observation value which define frame as abnormal or normal.

Solution approach: A method which consists three steps: SIFT flow, weighted orientation histogram and Hidden Markov Model(HMM) is proposed. Motion features are extracted by using SIFT flow. A pre-trained HMM model has been used to detect the abnormal crowd-motion.

2.11 CNN for Crowd Motion Analysis: An Application in Abnormal Event Detection

Problem Statement: Ravanbakhsh *et al.* [40] focuses on the detection of local anomalies in crowded scenarios.

Input: Video frames from UCSD crowd dataset.

Output: Localized the abnormal event.

Solution approach: A CNN based approach is proposed. The approach can be divided into three steps:

1. Extract CNN-based binary maps from a sequence of input frames.
2. Compute the Temporal CNN Pattern (TCP) measure using the extracted CNN-binary maps.
3. The TCP measure fuse with low-level motion features (optical-flow) to find the refined motion segments

2.12 Summary of the Literature Survey

The summary of the whole literature survey discussed above is presented chronologically in Table 2.1 with the problem statement and its solution approach.

Table 2.1: Summary of the literature survey.

PAPER REFERENCE	REFER-ENCE	PROBLEM STATEMENT	SOLUTION APPROACH
Hassner 2012 [52]	<i>et al.</i> ,	To provide a framework to monitor crowd scene to detect violent activity	ViF descriptor and SVM
Biswas <i>et al.</i> , 2014 [56]		To provide a framework to segment the crowd flow	Super-pixel based segmentation
Kang <i>et al.</i> , 2014 [57]		To provide a framework to segment the image	Fully convolutional neural network
Santhiya 2014 [42]	<i>et al.</i> ,	To provide a framework for abnormal crowd tracking and motion analysis	Background modelling, foreground detection and blob analysis
Mousavi 2015 [35]	<i>et al.</i> ,	To provide an improved video descriptor for recognizing abnormal situation in crowded scenes	Histogram of Oriented Tracklets(HOT) descriptor and SVM
Nazir <i>et al.</i> , 2015 [58]		To provide a framework for flow segmentation in dense crowd	Watershed algorithm

PAPER REFERENCE	PROBLEM STATEMENT	SOLUTION APPROACH
Boominathan <i>et al.</i> , 2016 [53]	To provide a deep learning framework that estimates the crowd density of highly dense crowds from static images	Deep and shallow neural network
Sharma <i>et al.</i> , 2016 [59]	To provide a framework for crowd flow segmentation	Trajectory clustering algorithm
Zhang <i>et al.</i> , 2016 [41]	To provide a framework for abnormal mass movement detection	SIFT flow, hidden markov model
Nagananthini <i>et al.</i> , 2017 [55]	To provide an algorithm that generates alert message when the people count exceeds a particular limit.	XCS-LBP descriptor and CNN
Ravanbakhsh <i>et al.</i> , 2018 [40]	To provide a framework for crowd analysis	Optical flow, CNN

Chapter 3

MOTIVATION AND PROBLEM STATEMENT

This chapter describes the motivation behind the proposed work by describing the scenarios where proposed framework can be implemented along with the problem definition that we are handle in this research work.

3.1 Motivation

Consider a situation where a huge number of people are gathered at some place (either constrained or unconstrained environment) for some purpose. In such places a CCTV cameras are installed by the concerned authority for monitoring and surveillance purpose. This continuous video of the crowd can be save onto the server. A software should be implemented on the server which analyse the video automatically and generates alarm in case of abnormal behaviour of crowd and gives alert message to the related security personnel.

Let us assume one more situation, in which a person stuck at some tragic site. He can record the event from his mobile and upload it onto the social network platform. These type of platforms allow users to interact with each other via chatting, messages, etc. With the help of these type of platform one can alert others by sharing the video of tragic event and so that others can arrange some help. Thus, self-activating software can be deployed on such platform also that analyse video clip and issue warning message in case of abnormal crowd-motion behavior. It reduces the need of expert being active on social network, to watch and analysis the video all the time.

With the help of in-time notification of any abnormal crowd-motion activity during public gathering, the process of outlet planning, crowd control planning and evacuation of victims can be start on time without any delay as time plays a crucial role in case of any disaster.

In such scenarios the proposed framework is beneficial for the detection and recognition process inside the software mentioned above.

3.2 Problem Statement

The overview of the crowd analysis problem is represent in Fig. 3.1. The problem that we are handling in this work is as follows: given the video clip of the crowded scene, analysing the crowd-motion behavior in real time. After analysing the crowd-motion generates the warning alarm message in case of abnormal behavior. The proposed framework extract the motion and appearance features of the scene and use these features to detect the abnormal behavior of crowd-motion. This task follows the following steps:

1. Supervised learning of motion and appearance features by using labelled videos (training data)
2. Predict the class of new unseen video by using the trained model

For feature extraction and selection we rely on deep convolutional neural network. Two types of motion features are considered: velocity and acceleration. For capture the change in brightness between two contagious frame visual saliency features are considered.

Based on the crowd-motion behavior in public gathering, crowd can be classified between two classes [26]: *first* is normal and *second* is abnormal. So, the training data is labeled in one of these two classes. Accordingly the proposed framework categorize the video clip in one of the classes. We will discuss about the training data and categorization in detail in the later section.

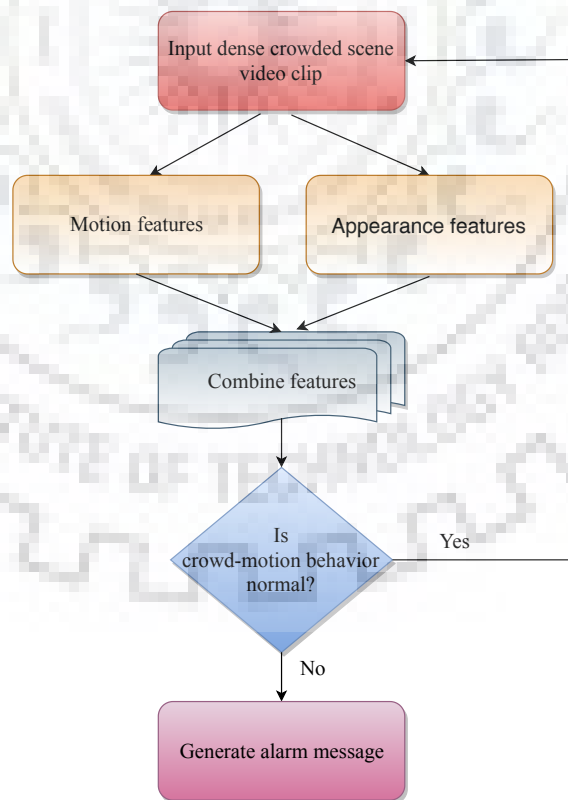


Figure 3.1: Overview of the problem.

Chapter 4

***CrowdVAS-Net*: A DEEP-CNN BASED FRAMEWORK TO DETECT ABNORMAL CROWD-MOTION BEHAVIOR**

In our study we proposed a Crowd Velocity, Acceleration and Saliency based CNN NETWORK. *CrowdVAS-Net* analyse the crowd-motion behavior. On the basis of the crowd-motion behavior it classify the crowd behavior as normal or an abnormal. A detailed description of the proposed *CrowdVAS-Net* framework along with the dataset creation and pre-processing has been provided in this chapter. The overview of the proposed framework are shown in Fig. 4.1. *CrowddVAS-Net* framework considered both motion and appearance features of the crowd flow. The detail description of the proposed framework are discussed in later section of this chapter.

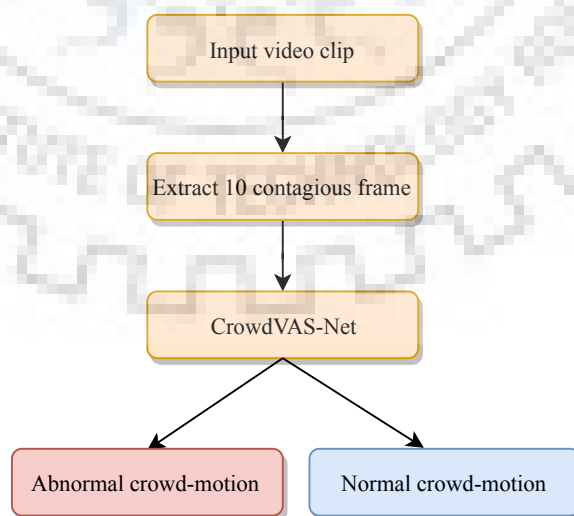


Figure 4.1: Overview of the proposed framework.

4.1 Dataset Creation and Pre-processing

A large number of labeled videos which captured dense crowded scenes are required to conduct the experiments. The available datasets: [16–21] either consist of very less number of videos or captured low dense crowded scene or partially labeled data. To the best of our knowledge such data which consists large number of labeled videos and also focuses on the dense crowded scenes only are not available till date.



Figure 4.2: Sample frames of proposed dataset.

Therefore, the dataset for this experiments has been constructed on our own and collected videos for our dataset from YouTube. During preparation of dataset, we considered various historical stampedes such as, Black Friday stampede, Mumbai stampede in India, rave party stampede in Los Angeles etc. In addition, we also focused on covering different scenarios such as indoor, outdoor, stadium, and night view of dense crowded scenes in order to keep variations in the dataset.

Fig. 4.2 shows the sample frames of different scenes of prepared dataset. We got the videos in the form of CCTV camera recordings, individually captured videos by various people and news coverage recordings of the scenes. These gathered videos are then trimmed into small clips of length up to 10-30 seconds covering only the dense crowd-motion. After the whole process, we left with 704 video clips capturing the dense crowded scene. On the basis of crowd behavior, we then classified the video clips into two classes named as *normal* and *abnormal*, which represent the crowd-motion behavior. The videos that came under abnormal category cover the situation where the motion pattern of the people shows some unusual behavior such as abrupt change in the speed of crowd movement or overcrowded situation or congested situation, etc. The normal category consists all clips where the motion of people did not depict any unusual behaviour. Table 4.1 provides the description of the proposed dataset.

The proposed dataset made up of total 704 video clips. Based on the crowd-motion behavior the dataset is divided into two classes, named as, *abnormal* and *normal*. The

Table 4.1: Descriptions of the proposed dataset.

Category	Number of Videos
Abnormal	372
Normal	332
Total	704

abnormal class having total 372 video clips of dense crowd motion and the *normal* class having 332 video clips.

4.2 Proposed Framework: *CrowdVAS-Net*

A detailed description of the proposed *CrowdVAS-Net* for the recognition of abnormal crowd-motion behaviour has been provided in this section.

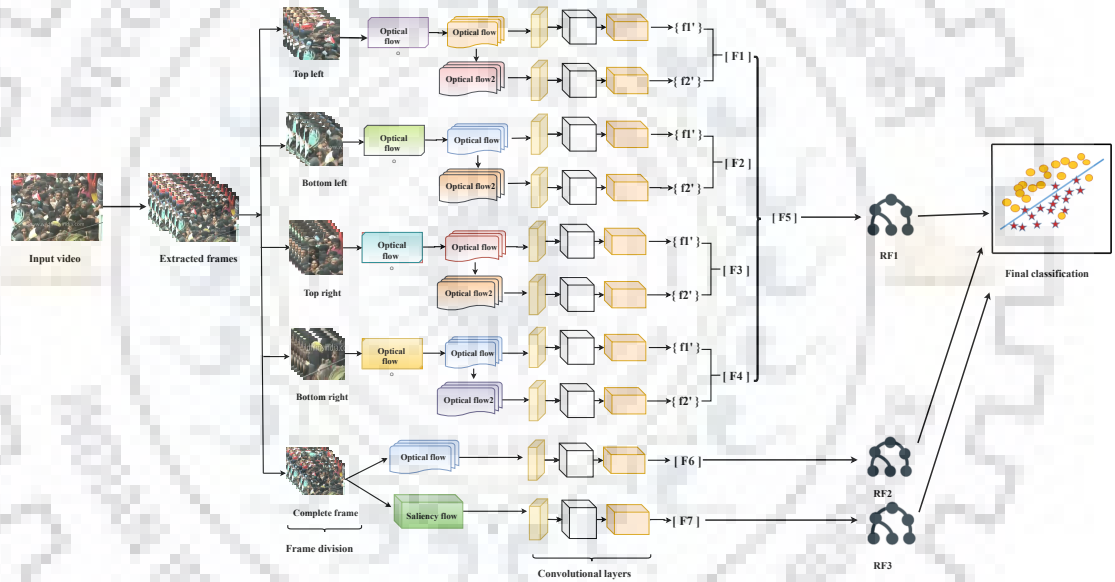


Figure 4.3: Block diagram of the proposed *CrowdVAS-Net*.

4.2.1 System Architecture

The main objective of our work is to keep track of any abnormal changes in crowd-motion behaviour with time. One of the abnormal changes types can be congestion i.e., either halt or moving slower, sudden increase or decrease in crowd-motion speed, panic behavior of people or overcrowded scene. In order to prevent any calamity, the detection of abnormal motion behavior should be done in real-time.

In order to increase the video processing speed, we split each video into ten equal length of video clips and then from each clip we took one frame. So for each video we consider only ten frames. This approach not only reduces the video processing time but also increases the framework capability to recognize the abnormal behavior of crowd-motion.

In addition to incorporate the local and global features of the scene, we divided each frame into four equal parts. For all these four frames along with complete frame, the optical flow has been calculated. A second round of an optical flow also applied to all these frames to track the change in velocity of each pixel. Optical flow and change in speed, together define the motion characteristics [13] of moving objects. So we called these features as motion features.

Furthermore, to include appearance features, we calculated the visual saliency for each complete frame. All these features are then fed into the different CNNs for extracting the vital representation of the crowded scene. These representations are then grouped to form a single representation vector $F5$ such that,

$$F5 = F1 + F2 + F3 + F4 \quad (4.1)$$

Now the three representation vector, $F5$, the extracted representation of the motion feature $F6$ and appearance feature $F7$ that has been calculated for complete frame, are trained separately on a random forest classifier. The dimension of representation vectors $F5$, $F6$ and $F7$ are 32768×1 , 4096×1 and 4096×1 , respectively. A block diagram of proposed the framework is shown in Fig. 4.3.

4.2.2 Optical Flow

In 1981, Horn and Schunck [12] proposed a theory for optical flow estimation. The purpose of the algorithm was to estimate the distribution of velocities of pixels between two frames. Let $I(x, y, t)$ is the brightness at pixel $P(x, y)$ in a frame at time t . Considering the brightness constancy assumption (shift of location but brightness stays same), we can write Eqn. (4.2).

$$I\left(x + \frac{dx}{dt} \partial t, y + \frac{dy}{dt} \partial t, t + \partial t\right) = I(x, y, t) \quad (4.2)$$

According to the optical flow constraint (movement of the pixel is very small or negligible) equation,

$$\frac{dI}{dt} = \frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \quad (4.3)$$

or,

$$I_x u + I_y v + I_t = 0 \quad (4.4)$$

where, u and v are the orthogonal components of an image along x coordinate and y coordinate, respectively and can be defined by Eqn. (4.5).

$$u = \frac{dx}{dt}, v = \frac{dy}{dt} \quad (4.5)$$

Eqn. (4.4) can be rewritten as shown in Eqn. (4.6).

$$\nabla I \cdot (u, v) + I_t = 0 \quad (4.6)$$

In order to solve Eqn. (4.6), Horn and Schunck [12] considered that the velocity of the objects do not change drastically and minimized the global energy function Eqn. (4.7) to get the values of u and v .

$$E = \iint \left((I_x u + I_y v + I_t)^2 + \alpha^2 (|\nabla u|^2 + |\nabla v|^2) \right) dx dy \quad (4.7)$$

where, α is a regularization constant parameter. By minimizing the value of E , Horn and Schunck [12] derived an iterative equation Eqn. (4.8) and (4.9) to get the values of u and v in each iterations.

$$u^{j+1} = u^{-j} - \frac{I_x (I_x u^{-j} + I_y v^{-j} + I_t)}{\alpha^2 + I_x^2 + I_y^2} \quad (4.8)$$

$$v^{j+1} = v^{-j} - \frac{I_y (I_x u^{-j} + I_y v^{-j} + I_t)}{\alpha^2 + I_x^2 + I_y^2} \quad (4.9)$$

4.2.3 Visual Saliency Flow Map

In order to fetch the saliency features of crowded scene, cluster based co-saliency method [22] has been used. The aim of the framework is to detect co-salient regions in a single image and a multi image using three cluster-based cues. In our work, we used all three cues, contrast cue, spatial cue and corresponding cue, that are extracted through cluster-based pipeline.

Let the i^{th} pixel of the image I^j is represented as p_i^j and the N_j represents the j^{th} image lattice. Let there are K clusters denoted as $\{C_k\}_{k=1}^K$ and the center of the cluster is denoted by $\{\mu^k\}_{k=1}^K$. The normalized location of the pixel p_i^j is represented by $z_i^{jN_j}$. The function $b: R^2 \rightarrow \{1 \dots K\}$ associates pixel p_i^j and cluster index $b(p_i^j)$.

Contrast Cues: Distinctive visual features of an image is calculated by the contrast cue. This is one of the most important cues for recognizing the salient parts of an image. The contrast cue for a cluster $w^c(k)$ is defined by Eqn. (4.10).

$$w^c(k) = \sum_{i=1: i \neq k}^K \left(n^i \|\mu^k - \mu^i\|_2 \frac{1}{N} \right) \quad (4.10)$$

where, N represents the total number of pixels in the image and n^i shows the pixel number of cluster C^i . The contrast cue is extremely useful for images whose foreground is sufficiently different from the background.

Spatial Cues: In order to recognize salient regions in an image the spatial cue focuses extra on objects that are close to the center. In addition, the spatial cue takes into account their proximity with the center of an image. For a given cluster C^K the spatial cue is given by Eqn. (4.11).

$$w^*(k) = \frac{1}{n^k} \sum_{i=1}^M \sum_{j=1}^N \left(\mathcal{N}(\|z^j - o^j\|^2 | 0, \sigma^2) \delta[b(p^i - c^k)] \right) \quad (4.11)$$

where $\delta(\cdot)$ represents the Kronecker Delta Function. Euclidean distance between the center of the image o^j and the given pixel z^j is calculated by Gaussian kernel $\mathcal{N}(\cdot)$. The pixel number in a cluster C^k is denoted by n^k . Compare to the contrast cue, the spatial cue is mainly useful for complex backgrounds.

Corresponding Cues: The corresponding cue is calculating for measuring the cluster distribution on multiple images. It finds the recurrent objects distributed among multiple images. The corresponding cue $w^d(k)$ is calculated by Eqn. (4.12).

$$w^d(k) = \frac{1}{\text{var}(q^k) + 1} \quad (4.12)$$

where, q^k is the M -bin histogram and calculated Eqn. (4.13).

$$q^k = \frac{1}{n^k} \sum_{i=j}^{N_j} \delta[b(p_i^j) - C^k] \quad (4.13)$$

where, $j = 1, \dots, M$ and q^k computes the distribution of cluster C^k among M images. Finally, single image saliency map and co-saliency map are created from these cues. Fig. 4.4 shows single saliency map and co-saliency map corresponding to the input video frames generated with help of above three cues.

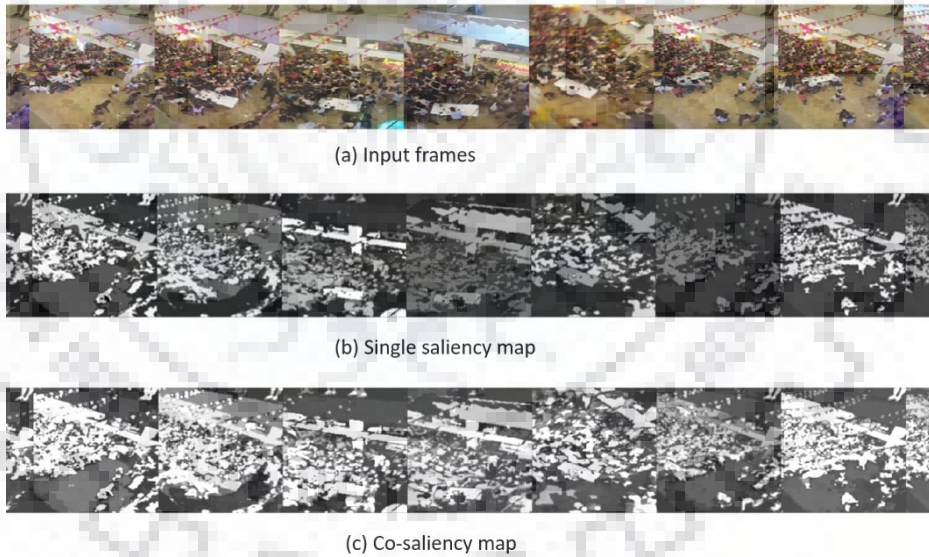


Figure 4.4: Salient features extracted from the crowd video. (a) Input frames of a video, (b) single saliency map, and (c) co-saliency map.

4.2.4 DCNN for Representation Extraction

Inspired by the superiority of CNN in video classification [23], it is used to extract the frame level representations from videos in order to infer the useful information about the crowded scene. To train the model, CNN requires a large amount of data. so we considered pre-trained CNN model. The pre-trained Alex-net [24] model has been used. For Alex-net model, the input image size should be 227×227 . So, we resized each input frame to

227×227 . The model has been made up of multiple layers of different sizes. The input layer of size $227 \times 227 \times 3$ is the first layer of the model and then the following layers have been piled up one after another $96 \times 11 \times 11$ conv, max pooling 3×3 , conv $128 \times 5 \times 5$, again max pooling 3×3 , conv $256 \times 3 \times 3$, $192 \times 3 \times 3$ conv, 3×3 pooling, fc 4096×1 , fc 4096×1 and fc 1000×1 layers.

In the proposed system architecture, the representation of each video frames from each CNN layer were extracted from *fc7* fully connected layer. The dimension of each representation vector is 4096×1 . Single feature vector has been constructed by concatenating the representation vector of each 10 frames of a video.

4.2.5 Classification: Random Forest Classifier

In this work, we used random forest (RF) classifier as it is outperform compare to other classifiers. Random forest is first introduced by Breiman [25] in 2001. Random Forest creates several decision trees and combine them to obtain a more accurate and stable prediction. As the tree grows, it increases the randomness of the model. Thus, it considers a random subset of features to train the model which prevent from over fitting of the model.

Three RF classifiers have been used in the proposed framework. Each classifier results in score. The final score, say *fScore*, of each class is calculated by Eqn. (4.14).

$$fScore = w_{t1} \times score_1 + w_{t2} \times score_2 + w_{t3} \times score_3 \quad (4.14)$$

where, w_{t1} , w_{t2} and w_{t3} are the weights associated with RF1, RF2 and RF3, respectively. The values of w_{t1} , w_{t2} and w_{t3} are set to be 0.35, 0.75 and 0.10, respectively, which are estimated empirically. For each test example, the class label is assigned by using the Eqn. (4.14). The label corresponding to the highest *fScore* value are assigned to the corresponding test example.

Chapter 5

SIMULATION RESULTS

A detail description of simulation setup and simulation results are enlisted in this section.

5.1 Simulation Setup

The experiments were conducted on Intel(R) Xeon(R) Gold 5118 processor with 96GB RAM having Quadro P2000 NVIDIA graphics card. MATLAB 2019a was used as a software for carry out the experiments.

As discussed in the Sec. 4.1, the proposed dataset consists of total 704 video clips having 372 videos of abnormal class and 332 videos of normal class. For conducting the experiments, the dataset split into training and testing data in the ratio of 7:3.

5.2 Simulation Scenarios

In order to assess the power of the proposed framework, we considered three different experimental scenarios. The scenarios with their outcomes have been discussed below.

5.2.1 Scenario 1: Comparison of *CrowdVAS-Net* with other state-of-the-art methods

In the first scenario, the main focus was on comparing the performance of proposed framework with the existing state-of-the-art methods and classifiers. Table 5.1 represents the statistical comparison of the results of various hand crafted features with the proposed framework. These accuracies have been achieved after five-cross fold validation.

Result analysis: After observing the simulation results shown in Table 5.1, we conclude the following:

- Hand crafted representation extraction methods such as optical flow [12], optical flow of optical flow and saliency flow [22] did not perform well.
- Adding optical flow features with state of the art deep models improve the recognition accuracy. However, this achieved recognition accuracy is not greater than 72%.

- Instead of increasing the recognition capability of the model, the combination of the motion feature representations and appearance feature representations with CNN, decreased the capability of the model.
- The proposed framework achieved significant higher accuracy (approximately 6% more) as compared to other hand crafted as well as deep feature representation extraction techniques.

Table 5.1: Statistical comparison of *CrowdVAS-Net* with other state-of-the art techniques on proposed dataset.

Method	Accuracies in % for different classifiers			
	KNN	SVM	Decision Tree	RF
Optical flow	55.71	56.67	50.48	72.38
Optical flow2	54.29	55.24	53.33	62.38
Optical flow + Saliency flow	57.62	48.10	52.76	51.43
Optical flow + CNN	70.95	65.71	61.43	72.02
Optical flow2 + CNN	54.76	52.86	56.67	60.48
Optical flow + Optical flow2 + Saliency + CNN	59.29	56.90	53.81	62.38
Proposed Method				77.8

5.2.2 Scenario 2: Check performance of *CrowdVAS-Net* on UMN dataset

In the second set of experiments, we checked the power of proposed framework on well known UMN dataset [16]. This dataset consists of total 20 videos of *abnormal* and *normal* classes. A detailed description of this dataset is shown in Table 5.2. The performance of the proposed framework compare to the previous method on UMN dataset is shown in Table 5.3 and the performance of the proposed framework and other state-of-art models on this dataset are shown in Table 5.4.

Table 5.2: A description of UMN dataset [16].

Category	Number of videos
Abnormal	8
Normal	12
Total	20

Result analysis: After observing the simulation results shown in Table 5.3 and Table 5.4, we conclude the following:

- It is clearly concluded from Table 5.3 that the proposed framework outperform on UMN dataset compared to the previous methods. The values of previous methods are taken from [40].
- It is also concluded from Table 5.4 that the proposed framework outperform on UMN [16] dataset. These accuracies had been achieved after five-cross fold validation.

Table 5.3: Statistical comparison of *CrowdVAS-Net* with previous methods on UMN dataset.

Method	Area Under Curve (AUC)
Optical flow [12]	0.84
Social Force [16]	0.96
STVF [11]	0.873
Sparse reconstruction [10]	0.976
Sparse Dictionaries [9]	0.972
Cem et al. [8]	0.964
Tracklet-based Commotion [7]	0.98
Temporal CNN [40]	0.98
Proposed Method	0.999

Table 5.4: Statistical comparison of the performance of proposed approach with other state-of-art feature extraction techniques with different classifiers on UMN SocialForce [16] dataset.

Method	Accuracies in % for different classifiers			
	KNN	SVM	Decision Tree	RF
Optical flow	60	80	20	48
Optical flow2	80	80	40	60
Optical flow + Saliency flow	60	68	20	64
Optical flow + CNN	80	48	60	60
Optical flow2 + CNN	76	56	60	64
Optical flow + Optical flow2 + Saliency + CNN	80	48	60	52
Proposed Method				99.8

5.2.3 Scenario 3: Comparison of video processing time of *CrowdVAS-Net* and classical approach

The aim of this set of experiment was to compare the video processing and analysis time of proposed approach versus classical approach (where all frames are considered) of video

processing. Feature extraction time, model training time and model predicting time are considered for analyzing. Fig. 5.1 represents the comparative analysis of all these three times.

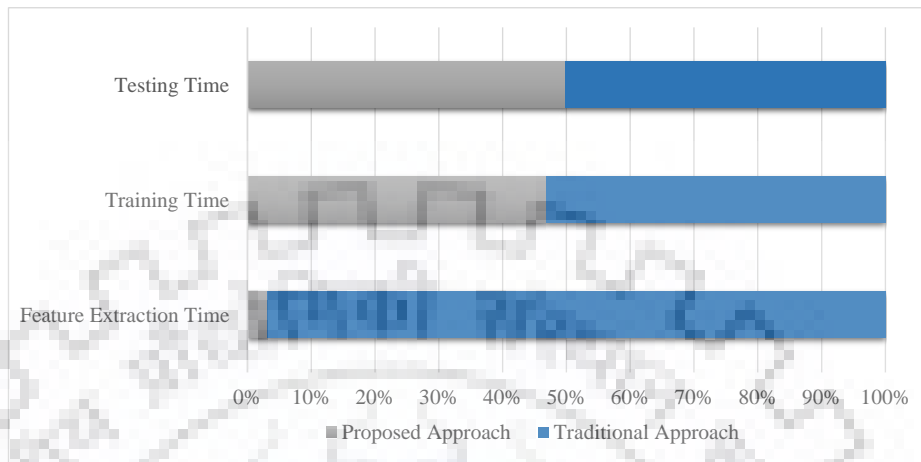


Figure 5.1: Comparative analysis of feature extraction time, training time and testing time of proposed and classical approach.

Result analysis: After analysing the simulation results, we infer the following:

- It is observed by the results that the feature extraction time of classical approach is more dominant than its training and testing time.
- It is also inferred from results that the proposed approach reduces the video analyses time up to 96.8%. Therefore, via our proposed approach, the process of recognition of abnormal crowd-motion behavior is fastened, which can help to control the disastrous situation that may emerge due to abnormal crowd behavior.

Chapter 6

CONCLUSIONS AND FUTURE SCOPE

6.1 Conclusions

In this research work, a deep-CNN based framework namely *CrowdVAS-Net* has been proposed. This framework is able to detect the abnormal crowd-motion behavior from short video clip. *CrowdVAS-Net* relies on DCNN to extract both motion and appearance features. In this work a new dataset that consists of 704 video clips of dense crowd behavior also introduced. In order to evaluate the performance of the *CrowdVAS-Net* various simulation experiments have been performed. Moreover, *CrowdVAS-Net* can reduce the video processing time up to 96.8% compared to state-of-the-art technique. The simulation results confirm that *CrowdVAS-Net* can detect the abnormal crowd-motion efficiently and such a model is required for predicting crowd disaster in video surveillance systems.

6.2 Future Scope

- As a future work, one can focus on multi-view analysis of crowded area using multiple cameras for more effective system in managing crowd disasters.
- Incorporate more classes like crowd split, crowd merge, etc., of crowd-motion behavior in the proposed dataset.
- Classify crowd-motion behavior based on trajectory pattern of there movement.

DISSEMINATION FROM THE DISSERTATION

Tanu Gupta, Vimala Nunavath and Sudip Roy, “*CrowdVAS-Net: A Deep-CNN Based Framework to Detect Abnormal Crowd-Motion Behavior in Videos for Predicting Crowd Disaster*”, submitted to the IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC), October, 2019 to be held in Bari, Italy.



REFERENCES

- [1] M. K. Starr and L. N. Van Wassenhove, "Introduction to the special issue on humanitarian operations and crisis management," *Production and Operations Management*, vol. 23, no. 6, pp. 925–937, 2014.
- [2] U. Escape, "Disasters in asia and the pacific: 2015 year in review," *United Nations report. Economic and social commission for Asia and the Pacific Google Scholar*, 2015.
- [3] N. Altay and M. Labonte, "Challenges in humanitarian information management and exchange: evidence from haiti," *Disasters*, vol. 38, no. s1, pp. S50–S72, 2014.
- [4] P. Akhtar, N. Marr, and E. Garnevska, "Coordination in humanitarian relief chains: chain coordinators," *Journal of Humanitarian Logistics and Supply Chain Management*, vol. 2, no. 1, pp. 85–103, 2012.
- [5] L. Li, T. Chi, T. Hao, and T. Yu, "Customer demand analysis of the electronic commerce supply chain using big data," *Annals of Operations Research*, pp. 1–16, 2016.
- [6] M. Rodriguez, S. Ali, and T. Kanade, "Tracking in unstructured crowded scenes," in *Proc. of the Computer Vision*, 2009, pp. 1389–1396.
- [7] H. Mousavi, M. Nabi, H. Kiani, A. Perina, and V. Murino, "Crowd Motion Monitoring using Tracklet-based Commotion Measure," in *Proc. of the ICIP*, 2015, pp. 2354–2358.
- [8] C. Direkoglu, M. Sah, and N. E. O'Connor, "Abnormal crowd behavior detection using novel optical flow-based features," in *Proc. of the AVSS*, 2017, pp. 1–6.
- [9] Y. Yu, W. Shen, H. Huang, and Z. Zhang, "Abnormal Event Detection in Crowded Scenes using two Sparse Dictionaries with Saliency," *Journal of Electronic Imaging*, vol. 26, no. 3, p. 033013, 2017.
- [10] Y. Cong, J. Yuan, and J. Liu, "Sparse Reconstruction Cost for Abnormal Event Detection," in *CVPR 2011*, 2011, pp. 3449–3456.
- [11] H. Su, H. Yang, S. Zheng, Y. Fan, and S. Wei, "The Large-scale Crowd Behavior Perception based on Spatio-temporal Viscous Fluid Field," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 10, pp. 1575–1589, 2013.
- [12] B. K. Horn and B. G. Schunck, "Determining Optical Flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [13] C. S. Royden and K. D. Moore, "Use of Speed Cues in the Detection of Moving Objects by Moving Observers," *Vision research*, vol. 59, pp. 17–24, 2012.

REFERENCES

- [14] A. Bisht, R. Bora, G. Saini, and P. Shukla, "Indian Dance Form Recognition from Videos," in *Proc. of the SITIS*, 2017, pp. 123–128.
- [15] T. Gupta, S. Saini, A. Saini, S. Aggarwal, and A. Mittal, "A Deep Learning Framework for Recognition of Various Skin Lesions due to Diabetes," in *Proc. of the ICACCI*, 2018, pp. 92–98.
- [16] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection Using Social Force model," in *Proc. of the CVPR*, 2009, pp. 935–942.
- [17] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly Detection in Crowded Scenes," in *Proc. of the CVPR*, 2010, pp. 1975–1981.
- [18] J. Ferryman and A. Shahrokni, "Pets2009: Dataset and Challenge," in *Proc. of the IWPEITS*, 2009, pp. 1–6.
- [19] S. Ali and M. Shah, "A lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis," in *Proc. of the CVPR*, 2007, pp. 1–6.
- [20] D.-Y. Chen and P.-C. Huang, "Motion-based Unusual Event Detection in Human Crowds," *Journal of Visual Communication and Image Representation*, vol. 22, no. 2, pp. 178–186, 2011.
- [21] M. Rodriguez, J. Sivic, I. Laptev, and J.-Y. Audibert, "Data-driven Crowd Analysis in Videos," in *Proc. of the ICCV*, 2011, pp. 1235–1242.
- [22] H. Fu, X. Cao, and Z. Tu, "Cluster-based Co-saliency Detection," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3766–3778, 2013.
- [23] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale Video Classification with Convolutional Neural Networks," in *Proc. of the CVPR*, 2014, pp. 1725–1732.
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in *Proc. of the ANIPS*, 2012, pp. 1097–1105.
- [25] L. Breiman, "Random Forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [26] T. Cao, X. Wu, J. Guo, S. Yu, and Y. Xu, "Abnormal Crowd Motion Analysis," in *Proc. of the ROBIO*, 2009, pp. 1709–1714.
- [27] S. Amraee, A. Vafaei, K. Jamshidi, and P. Adibi, "Anomaly detection and localization in crowded scenes using connected component analysis," *Multimedia Tools and Applications*, vol. 77, no. 12, pp. 14 767–14 782, 2018.
- [28] R. Chaker, Z. Aghbari, and I. N. Junejo, "Social Network Model for Crowd Anomaly Detection and Localization," *Pattern Recognition*, vol. 61, pp. 266–281, 2017.
- [29] A. D. Giorno, J. A. Bagnell, and M. Hebert, "A Discriminative Framework for Anomaly Detection in Large Videos," in *Proc. of the ECCV*, 2016, pp. 334–349.
- [30] R. Emonet, J. Varadarajan, and J.-M. Odobez, "Multi-camera Open Space Human Activity Discovery for Anomaly Detection," in *Proc. of the AVSS*, 2011, pp. 218–223.

REFERENCES

- [31] H. Mousavi, M. Nabi, H. K. Galoogahi, A. Perina, and V. Murino, "Abnormality Detection with Improved Histogram of Oriented Tracklets," in *Proc. of the ICIAP*, 2015, pp. 722–732.
- [32] H. Rabiee, J. Haddadnia, and H. Mousavi, "Crowd Behavior Representation: An Attribute-based Approach," *SpringerPlus*, vol. 5, no. 1, p. 1179, 2016.
- [33] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting Anomalous Events in Videos by Learning Deep Representations of Appearance and Motion," *Computer Vision and Image Understanding*, vol. 156, pp. 117–127, 2017.
- [34] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly Detection in Crowded Scenes," in *Proc. of the CSCCVPR*, 2010, pp. 1975–1981.
- [35] H. Mousavi, S. Mohammadi, A. Perina, R. Chellali, and V. Murino, "Analyzing Tracklets for the Detection of Abnormal Crowd Behavior," in *Proc. of the WACV*, 2015, pp. 148–155.
- [36] V. Saligrama and Z. Chen, "Video Anomaly Detection based on Local Statistical Aggregates," in *Proc. of the CVPR*, 2012, pp. 2112–2119.
- [37] K. Lohumi and S. Roy, "Automatic Detection of Flood Severity Level from Flood Videos using Deep Learning Models," in *Proc. of the ICT-DM*, 2018, pp. 1–7.
- [38] B. Zhai, "Identification of Abnormal Human Behavior in Intelligent Video Surveillance System," in *Proc. of the WARTIA*, 2018.
- [39] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal Event Detection in Videos using Generative Adversarial NSets," in *Proc. of the ICIP*, 2017, pp. 1577–1581.
- [40] M. Ravanbakhsh, M. Nabi, H. Mousavi, E. Sangineto, and N. Sebe, "Plug-and-play CNN for Crowd Motion Analysis: An Application in Abnormal Event Detection," in *Proc. of the WACV*, 2018, pp. 1689–1698.
- [41] D. Zhang, K. Xu, H. Peng, and Y. Shen, "Abnormal Crowd Motion Behaviour Detection based on SIFT Flow," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 1, pp. 289–302, 2016.
- [42] G. Santhiya, K. Sankaragomathi, S. Selvarani, and A. N. Kumar, "Abnormal Crowd Tracking and Motion Analysis," in *Proc. of the ICACCCT*, 2014, pp. 1300–1304.
- [43] M. Gao, J. Jiang, J. Shen, G. Zou, and G. Fu, "Crowd Motion Segmentation and Behavior Recognition fusing streak flow and collectiveness," *Optical Engineering*, vol. 57, no. 4, p. 043109, 2018.
- [44] R. Raghavendra, M. Cristani, A. Del Bue, E. Sangineto, and V. Murino, "Anomaly detection in crowded scenes: A novel framework based on swarm optimization and social force modeling," in *Modeling, Simulation and Visual Analysis of Crowds*, 2013, pp. 383–411.
- [45] J. Kim and K. Grauman, "Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates," in *Proc. of the CVPR*, 2009, pp. 2921–2928.

REFERENCES

- [46] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. of the CVPR*, 2010, pp. 1975–1981.
- [47] P. Shukla, T. Gupta, A. Saini, P. Singh, and R. Balasubramanian, "A deep learning frame-work for recognizing developmental disorders," in *Proc. of the WACV*, 2017, pp. 705–714.
- [48] M. Marsden, K. McGuinness, S. Little, and N. E. O'Connor, "Resnetcrowd: A Residual Deep Learning Architecture for Crowd Counting, Violent Behaviour Detection and Crowd Density Level Classification," in *Proc. of the AVSS*, 2017, pp. 1–7.
- [49] H. Rabiee, J. Haddadnia, H. Mousavi, M. Nabi, V. Murino, and N. Sebe, "Emotion-based Crowd Representation for Abnormality Detection," *arXiv preprint arXiv:1607.07646*, 2016.
- [50] *Phases of Disaster Management*, November, 2018, <http://www.avondaleaz.gov/government/departments/fire-medical/emergency-management>.
- [51] B. Yogameena and C. Nagananthini, "Computer vision based crowd disaster avoidance system: A survey," *International Journal of Disaster Risk Reduction*, vol. 22, pp. 95–129, 2017.
- [52] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in *Proc. of the CVPRW*, 2012, pp. 1–6.
- [53] L. Boominathan, S. S. Kruthiventi, and R. V. Babu, "Crowdnet: A deep convolutional network for dense crowd counting," in *Proc. of the Multimedia Conference*, 2016, pp. 640–644.
- [54] H. Idrees, I. Saleemi, C. Seibert, and M. Shah, "Multi-source multi-scale counting in extremely dense crowd images," in *Proc. of the CVPR*, 2013, pp. 2547–2554.
- [55] C. Nagananthini and B. Yogameena, "Crowd disaster avoidance system by deep learning using extended center symmetric local binary pattern texture features," in *Proc. of the CVIP*, 2017, pp. 487–498.
- [56] S. Biswas, R. G. Praveen, and R. V. Babu, "Super-pixel based crowd flow segmentation in h. 264 compressed videos," in *Proc. of the ICIP*, 2014, pp. 2319–2323.
- [57] Kang, Kai and Wang, Xiaogang, "Fully convolutional neural networks for crowd segmentation," *arXiv preprint arXiv:1411.4464*, 2014.
- [58] J. Nazir and M. Sirshar, "Flow segmentation in dense crowds," *arXiv preprint arXiv:1506.04608*, 2015.
- [59] R. Sharma and T. Guha, "A trajectory clustering approach to crowd flow segmentation in videos," in *Proc. of the ICIP*, 2016, pp. 1200–1204.
- [60] N. N. A. Sjarif, S. M. Shamsuddin, S. Z. M. Hashim, and S. S. Yuhaziz, "Crowd analysis and its applications," in *Proc. of the SECS*, 2011, pp. 687–697.
- [61] A. Bansal and K. Venkatesh, "People counting in high density crowds from still images," *arXiv preprint arXiv:1507.08445*, 2015.